



Structure-Based Inhibitor Design Targeting Galectin-8

Mohammad Bohari

Master of Science (Pharmacy)

Institute for Glycomics

Griffith University, Gold coast campus

Submitted in fulfilment of the requirements of the degree of
Doctor of Philosophy

April 2017

Mohammad Bohari

April 2017

Abstract

Galectins are an ancient class of lectins found in all forms of living organisms, responsible for modulating fundamental biological processes. They specifically recognise β -galactose containing glycans through their carbohydrate recognition domain (CRD) and mediate most galectin functions. Many galectins have emerged as novel therapeutic targets due to their association with the progression of various metabolic and disease conditions. Galectin-8, a tandem-repeat type member of the galectin family that contains two CRDs joined by an amino acid linker, is the focus of this research project. Galectin-8 plays a critical role in various biological processes such as cell adhesion and growth, immunomodulation, autoimmunity, inflammation, cancer and bone remodelling. Of interest to this project is the ability of galectin-8 to increase bone resorption factors like RANKL that leads to a decrease in bone mass density. Inhibition of this role of galectin-8 can potentially be of therapeutic importance and therefore could be a novel approach for tackling diseases associated with bone-loss. The individual CRDs of the tandem-repeat lectin exhibit similar functions to that of the full-length galectin-8, however, with lesser potency. The *N*-terminal domain of galectin-8 (galectin-8*N*) preferentially recognises 3'-anionic saccharides, a feature not observed in the *C*-terminal domain of galectin-8 (galectin-8*C*) or any other galectin CRD. This research details the initiation of inhibitor design campaign by targeting the galectin-8*N* CRD of the tandem-repeat.

The main body of research uses structure-based ligand design approaches through a combination of computational and experimental techniques. Molecular dynamics (MD) simulations were employed to analyse galectin-8*N*-glycan interactions and to construct the ligand design hypothesis. The designed ligands were synthesised and subjected to evaluation for binding to galectin-8*N* using various experimental techniques. Subcloning (ligation-dependant cloning) followed by protein expression (in *E. coli*) and affinity-based purification (using Lactosyl-Sepharose column) steps were performed to obtain the purified protein. Crystallographic investigations were conducted on various galectin-8*N*-glycan or galectin-8*N*-ligand complexes to study their binding modes and interactions. Different techniques either in solution or solid state were used to assess the binding strength of the designed ligands towards galectin-8*N*.

Our crystal structures of galectin-8*N* bound to human milk glycans (LNnT and LNT; Chapter 2) not only provided a rationale for the difference in their affinity but also formed the basis for ligand design carried out in the project. The crystal structure of galectin-8*N*-LNT

complex revealed a unique binding mode, wherein, for the first time a non-reducing end disaccharide of the tetrasaccharide was occupying the primary binding site. The presence of unique residue (Tyr141) in the extended galectin-8N binding site was mainly responsible for the observed binding mode. MD simulations including the *in silico* single residue mutations further supported the experimental findings and indicated the flipping of Tyr141 side chain governs the recognition of larger oligosaccharides. The galectin-8N-glycerol complex highlighted the minimum atomic features required for ligand recognition. Taken together, these crystal structures formed the basis to design interaction-based filters that guided the structure-based virtual screening.

A ligand design campaign was initiated by virtual screening a library of non-carbohydrate-small molecules using rigorous interaction-based criteria (Chapter 3). Available structural information including that generated in Chapter 2 and the preference of galectin-8N towards anionic saccharide was at the center of the screening. A library of compounds through iterative docking and molecular dynamics simulations was narrowed down to a small subset of molecules. The top fraction of the *in silico* analysis was purchased and evaluated for binding through saturation transfer difference (STD) nuclear magnetic resonance (NMR) and X-ray crystallography. These compounds did not bind to galectin-8N in our STD and X-ray experiments. However, the simulation outcome of one of the purchased compound provided the basis for exploiting unique amino acid residues in the galectin-8N binding site for ligand design.

Continuing the quest for identifying the ligands against galectin-8N, galactose as the core scaffold was taken forward for ligand design, primarily due to its inherent ability to interact with the conserved amino acid residues in the binding site. Based on our simulation results in Chapter 3 and available literature data, MB46A (compound **6**) was designed. It contains a propionic acid side chain that is ether-linked to the 3'-position of galactose, the idea being to mimic the interactions of the carboxylic acid portion of a high-affinity 3'-*O*-sialylated lactose (3'-SiaLac) (Chapter 4). The designed molecule was synthesised and shown to bind (139 μ M) galectin-8N through isothermal titration calorimetry (ITC) and X-ray crystallography. Our crystal structure confirmed the hypothesis and prompted for furthering the ligand design using compound **6** as a template molecule.

The encouraging outcome from the ligand design exercise so far in combination with molecular dynamics-based examination was employed to develop a library of compounds (Chapter 5). After considering the synthetic feasibility and ligand physicochemical properties, two closely related molecules (MB61B and MB63N) to the library of compounds were

synthesised and experimentally evaluated. MB61B and MB63N contain benzoyl and naphthoyl group respectively that are ester-linked to the 3'-position of the galactose (Chapter 5). The ligands were confirmed to be binding to galectin-8N by SPR (MB61B - 123.6 μ M and MB63N - 124.4 μ M). The obtained structure-activity insights will guide towards progressing the ligand design through ligand optimisation.

Overall the research presented in this thesis, demonstrate the successful rational medicinal chemistry application towards exploring the structure-activity landscape of galectin-8N. The information generated provides insight into the involvement of key binding site residues in recognising natural and synthetic ligands. Of encouragement is that the designed molecules (outlined in Chapter 4 and Chapter 5) inhibited galectin-8 mediated cell migration (preliminary results from our collaborator Prof. Y. Zick, Weizmann Institute, Israel) and are undergoing extensive dose-response analysis and possibly followed by *in vivo* investigations. These exciting preliminary *in vitro* results partially highlights the overall success of our ligand design campaign and further encourages the development of these leads into potent, efficient and selective ligands targeting galectin-8.

Table of Content

STATEMENT OF ORIGINALITY	I
ABSTRACT	III
LIST OF ABBREVIATIONS	VIII
LIST OF FIGURES	X
LIST OF TABLES	XV
ACKNOWLEDGEMENTS	XVI
ACKNOWLEDGEMENTS OF PAPERS INCLUDED IN THE THESIS	XVIII
CHAPTER 1	1
1.1 GALECTINS	2
1.1.1 <i>Historical background</i>	2
1.1.2 <i>General properties</i>	2
1.2 GALECTIN-8 - GENETICS	4
1.3 BIOLOGICAL ROLES OF GALECTIN-8	6
1.3.1 <i>Cell adhesion and growth</i>	6
1.3.2 <i>Immunity</i>	7
1.3.3 <i>Intracellular roles</i>	7
1.3.4 <i>Autoimmunity and inflammation</i>	8
1.3.5 <i>Cancer</i>	9
1.4 STRUCTURAL ANALYSIS	9
1.5 GLYCAN BINDING SPECIFICITY	12
1.6 GALECTIN-8 AND BONE FUNCTION	13
1.6.1 <i>Osteoporosis</i>	13
1.6.2 <i>Galectins and Bone function</i>	14
1.6.3 <i>Galectin-8 and Osteoporosis</i>	15
1.7 STRUCTURE-BASED DRUG DESIGN AND INHIBITORS OF GALECTIN	16
1.8 AIMS AND SCOPE OF THE RESEARCH	17
1.9 APPENDIX	19
STATEMENT OF CONTRIBUTION TO A CO-AUTHORED PUBLICATION	19
CHAPTER 2	39
2.1 FOREWORD	40
2.2 STATEMENT OF CONTRIBUTION TO A CO-AUTHORED PUBLICATION	41
2.3 STRUCTURE-BASED RATIONALE FOR DIFFERENTIAL RECOGNITION OF LACTO- AND NEOLACTO- SERIES GLYCOSPHINGOLIPIDS BY THE N-TERMINAL DOMAIN OF HUMAN GALECTIN-8	42

2.4 APPENDIX.....	55
2.4.1 MD simulation submission script:	58
2.4.2 MD results processing and analysis script.....	59
2.4 FURTHER DISCUSSION.....	60
CHAPTER 3.....	63
3.1 INTRODUCTION	64
3.1.1 Strategy for virtual screening.....	64
3.1.2 Library design.....	67
3.2 METHODS	69
3.2.1 Molecular docking.....	69
3.2.2 Molecular dynamics	70
3.2.3 Saturation Transfer Nuclear Magnetic Resonance spectroscopy.....	71
3.2.4 LIGAND SOAKING.....	71
3.3 RESULTS AND DISCUSSION	73
3.3.1 Efficiency evaluation	73
3.3.2 Docking and filtering	74
3.3.3 Short MD simulations.....	77
3.3.4 Long simulations	82
3.3.5 Binding evaluation through STD NMR.....	86
3.3.6 Binding evaluation through ligand soaking	86
3.4 CONCLUSIONS	87
3.6 APPENDIX.....	89
3.6.1 AutoDock Vina virtual screening master script	89
3.6.2 Docking score analysis	90
CHAPTER 4.....	92
4.1 FOREWORD.....	93
4.2 STATEMENT OF CONTRIBUTION TO A CO-AUTHORED PUBLICATION	94
4.3 STRUCTURE-BASED DESIGN OF A MONOSACCHARIDE LIGAND TARGETING GALECTIN-8.....	95
4.3.1 Table of content	96
4.3.2 Abstract.....	96
4.3.3 Introduction.....	96
4.3.4 Results and Discussion	99
4.3.5 Conclusions.....	106
4.3.5 Acknowledgments.....	106
4.4 SUPPLEMENTARY INFORMATION	107
4.4.1 Experimental section.....	108
4.4.2 Spectral data	115
4.4.3 Saturation Transfer Difference (STD NMR) experiments.....	126

4.4.4 Galectin-8C expression clone preparation.....	128
4.4.5 Preliminary biological evaluation.....	129
4.4.6 Hydrogen bond analysis	130
4.5 FURTHER DISCUSSION.....	131
CHAPTER 5	133
5.1 INTRODUCTION	134
5.1.1 The FDA approved drug.....	134
5.1.2 Galactose and carboxylic acid	135
5.1.3 Library design	136
5.2 METHODS	138
5.2.1 MD simulations	138
5.2.2 Synthesis of MB61B and MB63N.....	138
5.2.3 Isothermal titration calorimetry (ITC)	141
5.2.4 Surface Plasmon Resonance (SPR)	141
5.3 RESULTS AND DISCUSSION	143
5.3.1 MD Simulation of FDA drug.....	143
5.3.2 Simulations of the designed library.....	144
5.3.3 Binding free energy estimation	149
5.3.4 The designed compounds: MB61B and MB63N	151
5.3.5 Isothermal titration calorimetry (ITC)	153
5.3.6 Surface Plasmon Resonance (SPR)	154
5.3.7 Binding mode and interactions	161
5.4 CONCLUSIONS	163
5.5 APPENDIX.....	165
5.5.1 MMPBSA preparation script.....	165
5.5.2 Spectral data	167
5.5.3 FDA Drug fragment	176

List of Abbreviations

ADT	Autodock tools
AGRF	Australian genome research facility
CBP	Carbohydrate binding protein
CCL5	Chemokine ligand-5
CD4	Cluster of differentiation-4
CD44vRA	Cluster of differentiation 44 Rheumatoid arthritis variant
CD8	Cluster of differentiation-8
CG-1A	Chicken galectin-1
CG-8	Chicken galectin-8
CHO	Chinese hamster ovary
COSY	Homonuclear correlation spectroscopy
CRD	Carbohydrate recognition domain
c-Src	Proto-oncogene tyrosine-protein kinase
CXCL1	Chemokine ligand 1
DCM	Dichloromethane
DIPEA	Diisopropyl ethyl amine
DMSO	Dimethyl sulphoxide
<i>E. coli</i>	Escherichia coli
EDTA	Ethylenediamine tetra-acetic acid
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
ERK-1	Extracellular signal-regulated kinase 1
EtOAc	Ethyl acetate
FDA	Food and Drug Administration
FP	Fluorescence anisotropy.
FPP	Farnesyl pyrophosphate
GAFF	General amber force field
Galectin-8C	Galectin-8 C-terminal domain
Galectin-8N	Galectin-8 N-terminal domain
GGPP	Geranylgeranyl pyrophosphate
GlcNAc	N-acetyl glucosamine
GM-CSF	Granulocyte-macrophage colony-stimulating factor
GRIFIN	Galectin related interfibre protein
GRP	Galectin related protein
Hex	Hexane
HMBC	Heteronuclear Multiple Bond Correlation
HNK1	Human natural killer-1
HSQC	Heteronuclear single quantum correlation
IL6	Interleukin-6
IPTG	Isopropyl β -D-1 thiogalactopyranoside
LacNAc	N-acetyl lactosamine
LB	Luria-Bertani
LNFI	Lacto-N-fucopentaose III
LNnT	Lacto-N-neotetraose
LNT	Lacto-N-tetraose
MB46A	Methyl 3-O-[1-carboxyethyl]- β -D-galactopyranoside (compound 6)
MB61B	Methyl 3-O-benzoyl- β -D-galactopyranoside

MB63N	Methyl 3- <i>O</i> -naphthoyl- β -D-galactopyranoside
MD	Molecular dynamics
MSD	Mean square displacement
NDP52	Nuclear dot protein 52
NMR	Nuclear magnetic resonance
NHS	N-hydroxysuccinimide
OD	Optical density
PBS	Phosphate buffer saline
PDB	Protein data bank
PME	Particle mesh Ewald
PMSF	Phenylmethanesulfonyl fluoride
PTH	Parathyroid hormone
RANK	Receptor activator of nuclear factor kappa-B
RANKL	Receptor activator of nuclear factor kappa-B ligand
SDS-PAGE	sodium dodecyl sulfate-polyacrylate gel electrophoresis
SPR	Surface plasmon resonance
STD NMR	Saturation transfer difference NMR
TBS	Tris Buffer Saline
TE	Tris EDTA
Th1	Helper T-cell 1
Th17	Helper T-cell 17
TLC	Thin layer chromatography
Treg cells	Regulatory T-cells
VMD	Visual molecular dynamics
<i>wt</i>	wild-type
3'-SiaLac	3'-O-sialylated lactose

List of Figures

Figure 1.1: Depiction of the domain organisation within the galectins.	3
Figure 1.2: Schematic representation of cell surface glycan recognition	4
Figure 1.3: Amino acid sequence of the two isoforms of galectin-8	5
Figure 1.4: Multiple sequence alignment for CRDs of human galectins.....	11
Figure 1.5: Carbohydrate recognition domain of galectin-8N [3AP5 [76]]	11
Figure 1.6: Galectin-8 as a regulator of differentiation and maturation of osteoblast and osteoclast.....	15
Figure 2.1: Omit electron density maps calculated from refinement with the lactose omitted from the model.	56
Figure 2.2: Omit electron density maps calculated from refinement with the LNT omitted from the model.....	56
Figure 2.3: Omit electron density maps calculated from refinement with the LNT omitted from the model.....	57
Figure 2.4: Omit electron density maps calculated from refinement with the glycerol omitted from the model.....	57
Figure 3.1: A general theme of the virtual screening protocol.	64
Figure 3.2: Overview of the galectin-8N carbohydrate recognition domain.	65
Figure 3.3: Interactions made by glycans with the carbohydrate binding site of galectin-8N.	66
Figure 3.4: Efficiency evaluation through cognate ligand docking.	73
Figure 3.5: The strategy used to perform the structure-based virtual screening.....	75
Figure 3.6: An example of conformation-based (interaction-based) criteria.....	77
Figure 3.7: Chemical structures of the lead-like library of compounds retained during the short 5 ns simulation stage of the virtual screening.....	80
Figure 3.8: Chemical structures of the FDA drugs retained during the short 5 ns simulation stage of in the virtual screening.	81
Figure 3.9: Snapshot from MD simulations depicting binding conformation and hydrogen bonding interactions between the final lead-like library of compounds.....	84
Figure 3.10: Snapshot from MD simulations depicting binding conformation and hydrogen bonding interactions between the final FDA drugs (carbon in green sticks).....	85
Figure 4.1: Overview of the galectin-8N carbohydrate recognition domain..	98
Figure 4.2: The ligand design concept.	98

Figure 4.3: Interaction analysis from MD simulations..	100
Figure 4.4: Isothermal calorimetric analysis.....	102
Figure 4.5: Galectin-8N-6 complex..	103
Figure 4.6: Binding mode comparison of compound 6 towards the galectin-8N (crystal structure [5VWG]; on the left) and the galectin-8C (in silico complex; on the right).....	105
Figure 4.7: Agarose gel electrophoresis analysis of G8C:.....	129
Figure 4.8: Effect of MB46A on RANKL expression.....	130
Figure 5.1: Chemical structure of Pemetrexed (PMT).....	134
Figure 5.2: Galectin-8N-MB46A complex.	136
Figure 5.3: Designed library of compounds subjected to MD simulations.	137
Figure 5.4: Synthetic scheme for MB61B and MB63N.....	139
Figure 5.5: Progression of PMT during MD simulation from its predicted binding conformation to galectin-8N.....	143
Figure 5.6: Simulation of compound 1-6 (3-Atom spacer category) in complex with galectin-8N.....	145
Figure 5.7: Comparison of the initial placement of PMT (carbon in cyan) and compound 5 (carbon in brown) of the designed library.....	146
Figure 5.8: Simulation of compound 6-10 (2-Atom spacer category) in complex with galectin-8N.....	146
Figure 5.9: Simulation of compound 11-15 (1-Atom spacer category) in complex with galectin-8N.....	147
Figure 5.10: Simulation of compound 16-20 (2-Atom aromatic category) in complex with galectin-8N.....	148
Figure 5.11: Chemical structure of the compounds MB61B and MB63N synthesised along with the previously designed compound MB64A.	152
Figure 5.12: Isotherm for comparison of baseline during the titration of MB61B into galectin-8N with 200 s as ligand injection interval.	153
Figure 5.13: Isotherms for comparison of heats generated during protein-ligand (top; maroon) and protein-buffer (bottom; green/blue) titration.....	154
Figure 5.14: a) The pre-concentration scouting results for determination of optimum conditions for coupling galectin-8N to the chip surface.....	155
Figure 5.15: Binding level screen to examine binding of the designed ligands (Lig1-MB61B in red and Lig2-MB63N in green).	156
Figure 5.16: Binding sensogram for MB61B (lig1) and MB63N (lig2) at a wider 2 mM to 7.8 μ M concentration window.	157

Figure 5.17: Binding sensogram for triplicate injection of MB61B (a) and MB63N (b) at a narrow concentration window of 200 μ M to 0.78 μ M.....	158
Figure 5.18: The optimised binding sensogram (in triplicate) for MB61B (a) and MB63N (b) that was fitted into the steady state model to obtain the dissociation constants.	160
Figure 5.19: The snapshots from the MD simulation of the in silico generated galectin-8N-MB61B (on top; a, b) and galectin-8N-MB63N.....	161
Figure 5.20: MD simulation snapshot of galectin-8N-MB61B complex depicting flipping of Tyr141 side chain and its interaction with the aromatic ring of the ligand.	162
Figure 5.21: Structure of FDA drug PMT with yellow highlighted section showing the structure of purchased FDA fragment.....	176

List of Tables

Table 1.1: List of galectin-8 structures either full-length or isolated galectin-8N and galectin-8C deposited in the protein data bank.....	10
Table 1.2: List of FDA approved drugs and their mechanism of action for treatment of osteoporosis.....	13
Table 3.1: Mean square displacement analysis performed for the lead-like library (far left and centre) and FDA approved drugs (on far right).	79
Table 3.2: Mean square displacement values from long simulations for the lead-like compounds (on the left) and FDA drugs (on the right)..	82
Table 4.1: Crystallographic data merging and refinement statistics for galectin-8N-6 complex structure.....	103
Table 5.1: Estimated ligand binding free energies (in kcal/mol) of ligand towards galectin-8N from MMPBSA.....	150

Acknowledgements

I would like to heartedly thank my Principal supervisor Professor Helen Blanchard for allowing me to be a part of her research group. The structural biologist in her identified the nucleating points (of learning) in me (at the beginning) that during the suitable incubation period (my candidature) would result in beautiful [diffracting!] crystals (the present research project). She has always been very encouraging towards developing the independent thinking ability and has provided endless support and guidance throughout the course of my higher research degree. I would also greatly acknowledge my Associate supervisor Dr Xing Yu for the support, advice and uncountable odd-timely discussions during my degree (particularly during the long walks to the carpark!). He has been the shock absorber for the days when going was getting tough and guided me to keep focussing.

I would like to thank the Blanchard group members (both past and present) for sharing the lab and open-office hours with me. Dr Patrick Collins, thanks for helping me get into the foreign world of biochemistry during the early days. Thanks Rahul for sitting next to me in the open-plan and wondering along the beautiful world of science. Matt, thanks for being so welcoming in the chemistry lab and being my go-to friend for various science-related and not-so-science related subjects. I would also like to extend my appreciation to Dr Chandan Kishor for becoming the discussion centre (at the lunch hour) towards the last phase of my PhD. Thanks, Hadieh for the shared goodies and Brijesh for all the odd timely photos. Drinks time is something that would be hugely missed (not because of the quality of beer!) due to enlightening chats on real-life topics.

I would also like to acknowledge the advice and support provided by Dr Darren Grice during my stay in a chemistry lab (and in the G25 common room). Special thanks to Catherine for the performing mass spectrometry samples and Lauren for the advice and support for doing ITC and SPR experiments. Thanks, Pradeep and Sai for various festive feasts and refreshing long journeys. Thanks, Diana and Jian for the thought-provoking and stimulating common room discussions. I would like to thank the Institute for Glycomics for hosting me and Griffith University for the financial support.

I would like to heartedly acknowledge the patience and sacrifice of my parents (Husain and Ajab) without which reaching the highest career milestone was not even thinkable. I am sorry for not being present on so many occasions when you needed me the most, thanks to my

sister Tasneem for being such a lovely person and keeping my goals checked always. Thanks to my wife Fatema for the much-needed warmth, unconditional love, care, support and the unlimited patience. Batul thanks for filling our lives with happiness and being such a great kid. Love you all. Last but not the least I thank Almighty for making everything possible.

Acknowledgement of published papers included in this thesis

THE PAPERS INCLUDED ARE ALL CO-AUTHORED PAPERS

Section 9.1 of the Griffith University Code for the Responsible Conduct of Research (“Criteria for Authorship”), in accordance with Section 5 of the Australian Code for the Responsible Conduct of Research, states:

To be named as an author, a researcher must have made a substantial scholarly contribution to the creative or scholarly work that constitutes the research output, and be able to take public responsibility for at least that part of the work they contributed. Attribution of authorship depends to some extent on the discipline and publisher policies, but in all cases, authorship must be based on substantial contributions in a combination of one or more of:

- conception and design of the research project
- analysis and interpretation of research data
- drafting or making significant parts of the creative or scholarly work or critically revising it so as to contribute significantly to the final output.

Section 9.3 of the Griffith University Code (“Responsibilities of Researchers”), in accordance with Section 5 of the Australian Code, states:

Researchers are expected to:

- Offer authorship to all people, including research trainees, who meet the criteria for authorship listed above, but only those people.
- accept or decline offers of authorship promptly in writing.
- Include in the list of authors only those who have accepted authorship
- Appoint one author to be the executive author to record authorship and manage correspondence about the work with the publisher and other interested parties.
- Acknowledge all those who have contributed to the research, facilities or materials but who do not qualify as authors, such as research assistants, technical staff, and advisors on cultural or community knowledge. Obtain written consent to name individuals.

Included in the thesis are papers in Chapters 1, 2 and 4 are co-authored with other researchers. My contribution to each co-authored paper is outlined at the front of the relevant chapter. The bibliographic details for these papers including all authors, are:

Chapter 1

Helen Blanchard, Khuchtumur Bum-Erdene, **Mohammad Hussaini Bohari** and Xing Yu (2016). Expert Opinion on Therapeutic Patents. 26(5):537-54. doi: 10.1517/13543776.2016.1163338.

Chapter 2

Mohammad H. Bohari, Xing Yu, Yehiel Zick and Helen Blanchard (2016). Scientific Reports. 6:39556. doi: 10.1038/srep39556.

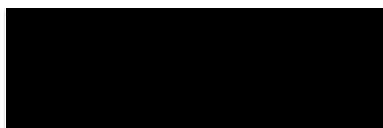
Chapter 4 (manuscript under revision)

Mohammad H. Bohari, Xing Yu, Darren Grice, Yaron Vinik, Yehiel Zick, Helen Blanchard (author list to be finalised). ChemMedChem communications. (manuscript under revision).

Chapter 5 (manuscript in preparation)

Mohammad H. Bohari, Xing Yu, Darren Grice, Yaron Vinik, Yehiel Zick, Helen Blanchard. (author list to be finalised; manuscript in preparation).

Appropriate acknowledgements of those who contributed to the research but did not qualify as authors are included in each paper.



(Signed) _____ (Date) __03/04/2017__

Mohammad Bohari



(Countersigned) _____ (Date) __24/11/2017__

Supervisor: Prof. Helen Blanchard

Chapter 1

Introduction

1.1 Galectins

1.1.1 Historical background

Complex carbohydrates are an important constituent found in all forms of living organisms and are mainly involved in glycan-mediated information transmission. They are responsible for carrying out diverse biological activities, including cell-cell recognition, communication, and host-microbe interactions [1-4]. The decoding of the varied and complex nature of information present on the cell surface is carried out by a family of carbohydrate-binding protein called lectins [5]. The decoding process mostly involves the engagement of either the terminal or the internal units of the complex glycans within the shallow binding pocket of the lectins. The first ever lectin reported in animal tissue, after agglutinins (plant) and discoidin I (amoeba), was asialoglycoprotein that showed calcium ion dependency (C-type lectin) [6]. In 1975, electrolectin was the first lectin isolated from the electric organ of electric eel (*Electrophorus electricus*) that required reducing conditions for retaining its biological activity (S-type lectins, later termed galectins) [7].

Galectins are soluble S-type lectins that show sulfhydryl (-SH) dependency for functioning as evident from the presence of free cysteine residues. All galectins exhibit overall conserved amino acid motifs in their binding site. In 1976, the first mammalian lectin (now termed galectin-1) was isolated from calf heart/lung extracts using its lactose-binding property [8, 9]. In the early 1980s, a 35 kDa carbohydrate binding protein (CBP35, now known as galectin-3) was isolated from mouse fibroblast [10]. Based on the order of discovery, all lectins sharing specific features were grouped and named as galectins. [11]. To date, galectin-1 and galectin-3 are structurally and functionally well characterised amongst the 15 different isolated galectins [11]. The characterisation repertoire for other galectins is increasing with the unravelling of the fundamental roles played by the glycan-mediated signalling system.

1.1.2 General properties

Galectins are a family of soluble animal lectins that specifically recognises β -galactoside-containing glycans found on cell surfaces and modulates critical biological processes. Galectins are the most ancient class of glycan binding proteins found in almost all forms of organisms from protists to invertebrates to vertebrates [12, 13]. Galectins can be found inside the cell, either in the nucleus or the cytoplasm, and they also are secreted into the extracellular space, where it is either associated with the membrane or the cell matrix [3, 13-15]. Controlling cellular behavior in an extracellular manner is a well-characterized function

of galectins, while some studies suggest their presence intracellularly as potential pattern recognition receptors [16, 17].

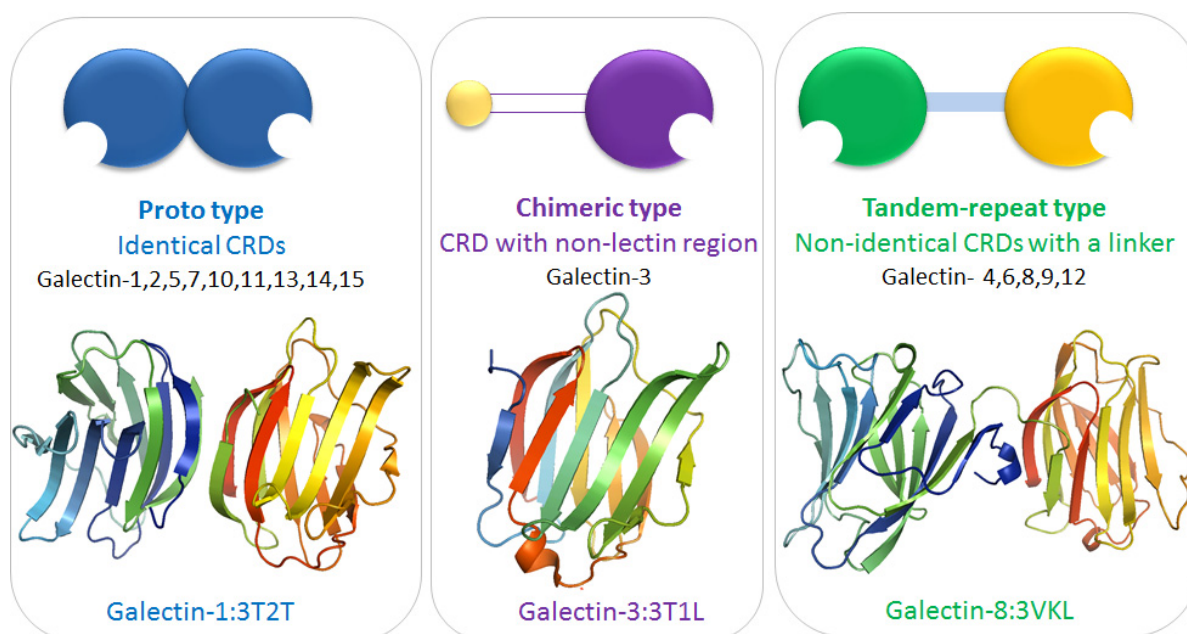


Figure 1.1: Depiction of the domain organisation within the galectins along with one representative structure from the protein data bank. Reference [18] for entry 3T2T and 3T1L; reference [19] for 3VKL.

The galactose recognition by galectins occurs through their carbohydrate recognition domain (CRD). Biologically, the CRDs become associated with each other to form higher order structures. The three groups namely prototype, chimera type and tandem-repeat type are formed based on varying spatial arrangement of these CRDs [7] (Figure 1.1).

Prototype galectins contain a single CRD that self-associates non-covalently to form homodimers. This dimerization occurs on the opposite sides of the carbohydrate binding site [20, 21]. Due to the sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) mobility behaviour, these galectins are also referred as 14 kDa galectins. Prototype galectins include Galectin-1, 2, 5, 7, 10, 11, 13, 14, 15.

Chimera type galectins encompass the single lectin CRD fused with a non-lectin *N*-terminal domain. This non-lectin domain is a sequence of 120 amino acids, rich in proline, glycine, and tyrosine [22, 23]. This *N*-terminal domain undergoes multimerisation via self-aggregation and is found susceptible to proteolytic hydrolysis by some matrix metalloproteases [24]. Chimera-type galectin includes only galectin-3.

Tandem-repeat galectins are comprised of two different CRDs connected via a linker of variable length (5-70 amino acids). The linker shows susceptibility towards proteases, and the cleavage of the tandem-repeat CRDs at the linker would generate a prototype product, that

may function independently [11]. Two different CRDs in tandem extends the possibility of binding to two distinct ligands simultaneously; thereby acting as hetero-bifunctional cross-linking agents [13, 14]. Galectin-4, 6, 8, 9 and 12 are members of this group.

The CRD of galectins share particular binding site residues which are conserved throughout the family, and these motifs are responsible for recognition of galactose. However, the affinity of galectins towards these cell surface glycans is weaker. Fundamentally, though, this increases by several orders of magnitude either upon self-association or with the presence of covalently linked multiple domains [25]. This multivalent nature of galectins makes them a versatile communicator between the glycan present on a cell or between different cells (Figure 1.2). Some outliers that only partly contain the conserved sequence motifs include GRP (galectin related protein) that has no lectin attributes, galectin-11 (GRIFIN-galectin related interfiber protein) which is devoid of the carbohydrate binding affinity, and galectin-10 (Charcot-Ledeyn crystal) that recognises a mannosyl residue over galactose residue [13].

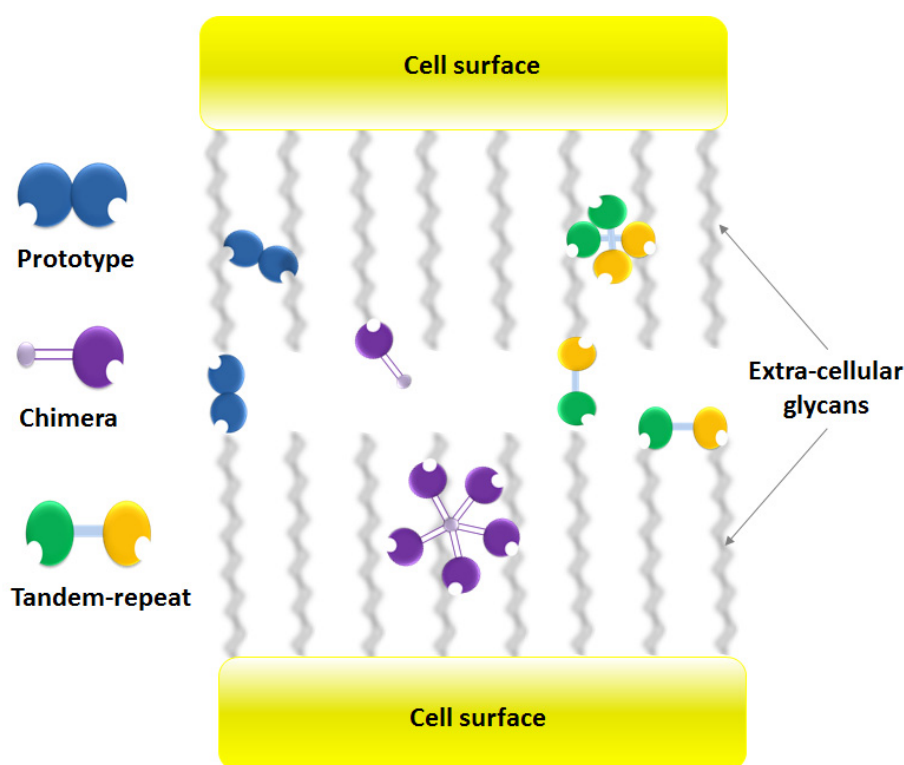


Figure 1.2: Schematic representation of cell surface glycan recognition and multivalent nature of the galectins.

1.2 Galectin-8 - Genetics

During evolution, the duplication and subsequent divergence of relevant exons of the prototype galectins led to the formation of the tandem-repeat type galectins. Galectin-8 was first identified serendipitously from the screening of a rat liver cDNA library with antibodies directed towards Insulin Receptor Substrate-1 [26]. Following which three unique galectin-8

cDNA sequences were identified from the human prostate cancer and lung cancer expression libraries at the messenger RNA level [27, 28]. Galectin-8 is localised on chromosome 1 (1q42.11) and contains 11 exons. The open reading frame of galectin-8 codes for 318 amino acids. Among the eleven constitutive exons coding the galectins, exons 5, 6 and part of 7 and exon 13, 14 and part of 15 marks the coding of the CRDs [14, 19, 28-30].

Galectin-8 Full length variant-1 (G8L; Long isoform)

MMLSLNNLQNIINYNPVIFVGTIPDQLDPGTLIVIRGHVPSDADRFQVDLQNGSSMKPRADVAFHFNPR
FKRAGCIVCNTLINEKWGREETYDTPFKREKSFEIVIMVLKDKFQVAVNGKHTLLYGHRIGPEKIDTL
GIYGKVNHSIGFSFSSDLQSTQASSLELTEISRENVPKSGTPQLPSNRGGDISKIAPRTVYTKSKDST
VNHTLTCTKIPPMNYVSKRLPFAARLNTPMGPGRTVVVKGEVNANAKSFNVDLLAGKSKDIALHLNPR
NIKAFVRNSFLQESWGEEERNITSFPFSPGMYFEMIIYCDVREFKVAVNGVHSLEYKHRFKELSSIDTL
EINGDIHLLEVRWS

Galectin-8 Full length variant-2 (G8M; Major isoform)

MMLSLNNLQNIINYNPVIFVGTIPDQLDPGTLIVIRGHVPSDADRFQVDLQNGSSMKPRADVAFHFNPR
FKRAGCIVCNTLINEKWGREETYDTPFKREKSFEIVIMVLKDKFQVAVNGKHTLLYGHRIGPEKIDTL
GIYGKVNHSIGFSFSSDLQSTQASSLELTEISRENVPKSGTPQLRLPFAARLNTPMGPGRTVVVKGEV
NANAKSFNVDLLAGKSKDIALHLNPRLNKAFVRNSFLQESWGEEERNITSFPFSPGMYFEMIIYCDVR
EFKVAVNGVHSLEYKHRFKELSSIDTLEINGDIHLLEVRWS

Variant-1 / Long isoform			
G8full	1080 bp	360 aa	
G8C	408 bp	136 aa	GREEN - N-terminal domain key recognition site
G8N	465 bp	155 aa	BLUE - C-terminal domain key recognition site
G8full	954 bp	318 aa	LINKER
Variant-2 / Major isoform			

Figure 1.3: Amino acid sequence of the two isoforms of galectin-8 with the conserved sequence motifs highlighted in galectin-8_N (green) and galectin-8_C (blue) and the linker peptide adjoining the two CRDs highlighted in yellow.

All the six isoforms of galectin-8 isolated possess the same open reading frame with or without the insertional sequence [27, 28]. These insertional sequences either increases the size of the amino acid linker (PCTA-1 [14, 27]; Po66-CBP and Po66-CBP-IS1 [28]) or in some isoforms reduces the amino acid sequence by the stop codon that results in prototype isoform with single CRD (Po66-CBP-IS2; Po66-CBP-IS1-IS2; Po66-CBP-IS1-IE-IS2 [28]). The presence of these prototype isoforms of galectin-8, although non-isolated, is a unique feature not observed with other tandem-repeat galectins. The genomic organisation of galectin-8 varies from that of other tandem repeat type (galectin-4, 6 and 9) by these alternatively spliced products [27, 29, 30]. Out of the three tandem-repeat isoforms, PCTA-1 and Po66-CBP (major

isoform) share 98.5 % amino acid sequence identity and Po66-CBP-IS1 (long isoform) has larger linker peptide (Figure 1.3). The long isoform has a 70 residue long linker while the major isoform has 27 residues (Figure 1.3). The presence of a variable length linker connecting the two galectin-8 CRDs allows for recognising different spatially oriented sugars and hence broadening its spectrum of interacting ligands.

1.3 Biological roles of galectin-8

Galectin-8 is involved in modulation of critical biological processes mainly by recognising the cell-surface glycans. The apparent functional spectrum is evident from the growing number of publications reporting galectin-8's structure-function investigations (with last five years' average of ~15 articles per year). Galectin-8 regulates cell growth and adhesion, plays a role in the immune response of the body, shows varied expression profiles in cancer, involved in blood vessel formation and bone remodeling. Most binding partners of galectin-8 are heavily glycosylated, and galectin-8 mediates its functions by interacting with these glycans in a carbohydrate-dependent manner. The biological roles of galectin-8 have been described in the section below except for the bone related functions which are elaborated in section 1.6.

1.3.1 Cell adhesion and growth

Cell-matrix interactions are the key element in regulating cell-to-cell communication; that occurs through integrin-mediated signal transduction. Galectin-8 was shown to exert anti-adhesive effects by selectively interacting with a sub-group of the integrin family, unlike other galectins that sterically blocks the cell-binding sites for adhesion receptor [31]. Galectin-8 interacts with β_1 subunit and α_6 subunit containing integrins. In contrast, immobilised galectin-8 adhere to various cell types and acts equipotent to fibronectin in promoting cell adhesion and cell spreading and migration [32]. Therefore, through its pro- and anti-adhesive functions, galectin-8 is regarded as a member of the adhesion-modulating extracellular matrix protein [33]. Galectin-8 transfected in lung carcinoma cells showed inhibition of colony formation where apart from galectin-8's anti-adhesive functions, it induced apoptosis [31, 33].

The two CRDs of galectin-8 are functionally dependent on each other as the linker peptide determines the orientation of the CRDs for proper functionality [34]. Various mutants of galectin-8 (I90R, R253I, E88Q, E251Q) that were constructed for the structure-function analysis showed altered sugar-binding capacity [34]. Galectin-8 selectively induces reversible adhesion of peripheral blood neutrophils, unlike to galectin-1, -3 and -9 which showed minimal

effects on neutrophil adhesion [35]. Interestingly, galectin-8 isolated from the neutrophil membrane revealed different binding partners. The galectin-8^C bound to integrin and pro-matrix metalloproteinase-9 (proMMP-9) while the galectin-8^N bound only to the proMMP9 [35].

1.3.2 Immunity

Galectin-8 is expressed in thymocytes, thymic epithelial cells, and spleen with two alternative splicing variants isolated from mouse thymus [36, 37]. Exposure of cultured thymocyte with galectin-8 induces apoptosis only in the CD4^{high} and CD8^{high} cells through activation of caspase pathway [36]. In contrast, galectin-1 induces apoptosis in both CD4^{+/-} and CD8^{+/-} thymocytes, while galectin-3 preferentially depleted CD4⁻ and CD8⁻ population [36]. Unlike galectin-1 and galectin-3, galectin-8 induced adhesion in Jurkat T-cells through an integrin-mediated signalling pathway [38]. Galectin-8 is a potent T-cell suppressor that acts as a pro-apoptotic agent in Jurkat T-cells through a unique complex phospholipase D/phosphatidic acid signalling pathway, not involved for other galectins [39]. Galectin-8 exposure to T-cells induced antigen-independent proliferation of CD4⁺ T-cells, through agonistic binding to CD45, and a co-stimulatory response on a given T-cell response [37]. Interestingly, for inducing T-cell proliferation, two CRDs of galectin-8 are required in tandem while co-stimulation can still occur through either CRDs [40].

1.3.3 Intracellular roles

Galectin-8's preferential recognition of sialylated galactosides is a unique feature not observed for other galectins [41, 42]. However, recognition of these galactosides is needed for mediating extracellular binding and cellular activation by galectin-8 [43, 44]. Galectin-8 endocytosed in Chinese Hamster Ovary (CHO) cells (containing sialylated glycans) and Lec2 mutant cells (lacking sialylated glycans) through a non-clathrin and non-cholesterol dependent pathway [45]. However, after endocytosis, galectin-8 localised in the plasma membrane, around the nucleus and small vesicles in CHO cells, while it was found evenly distributed in large vesicles in Lec2 cells [45].

Intracellularly, galectin-8 acts like a danger receptor by initiating and perpetuating inflammatory response. The galectin-8^C specifically recognises and kills human blood group antigen- and α 1-3Gal-expressing *E. coli* while leaving the other *E. coli* (not expressing ABO(H) antigens) and other gram-negative, and positive bacteria unaltered both *in vitro* and *in vivo* [46]. Galectin-8 also restricts the proliferation of Salmonella by recognising glycans on the damaged bacteria-containing vacuoles and kills the bacteria by inducing an autophagic

response [47]. The autophagy is mediated by the interaction of galectin-8 with autophagy receptor NDP52. Interestingly, the crystal structure of galectin-8-NDP52 peptide revealed the involvement of the convex surface of the galectin-8C CRD in recognising the peptide [48, 49]. The selectivity in the interaction of galectin-8 with NDP52 is believed to be due to steric hindrance of the hook-like conformation of the NDP52 peptide which is positioned strategically on the convex surface of the galectin-8C CRD [49].

1.3.4 Autoimmunity and inflammation

Galectin-8 is expressed during plasma cell differentiation where it promotes the formation of plasma cells from mature B-cells and acts as a selective modulator of B cell function [50]. Synovial fluid cells from rheumatoid arthritis patients expressed galectin-8 and had a high affinity for an arthritis-specific CD44 variant (CD44vRA) [51]. Auto-antibodies against galectin-8 were found in the sera of patients with systemic lupus erythematosus, rheumatoid arthritis and septicemia [52, 53]. Sera from primary glomerular nephritis (an altered IgA glycosylation disease) patients exhibited reduced binding to galectin-8N which was ultimately related to the disease severity [54]. Furthermore, a non-synonymous single nucleotide polymorphism in galectin-8, leading to F19Y mutation, is strongly associated with the occurrence of rheumatoid arthritis [55].

Exposure of endothelial cells to galectin-8 increased the adhesion of platelets and stimulated the cells to produce pro-inflammatory molecules such as chemokine ligand 1 (CXCL1), granulocyte-macrophage colony-stimulating-factor (GM-CSF), interleukin-6 (IL-6) and chemokine ligand-5 (CCL5) [56]. CCL5 was the most dramatically upregulated upon galectin-8 treatment, this molecule is involved in recruitment of inflammatory cells during chronic inflammation [56]. The expression of galectin-8 was observed in rejected allografts than accepted allografts further suggesting its role in inflammation and promotes rejection [57]. Therefore, inhibition of galectin-8 could work as a potential therapeutic intervention in enhancing graft survival [57]. In contrast, retinal pathology can be reduced, and photoreceptor cell damage can be prevented by treatment with galectin-8 [58]. The treatment of galectin-8 with mouse uveitis model caused downregulation of pro-inflammatory cytokines (IFN γ and IL-17A) and upregulation of anti-inflammatory cytokines (IL-10) [58]. Therefore, Galectin-8 can be used a therapeutic molecule for uveitis. Galectin-8 increases production of regulatory T-cells and blunted the production of inflammatory cytokines by retinal Th1 and Th17 cells [58]. Galectin-8 promotes differentiation of highly suppressive regulatory T-cells and increases expression of cytotoxic T-lymphocyte-4 and interleukin-10 (IL-10), which has implications for the treatment of autoimmune and inflammatory disease [59].

1.3.5 Cancer

Galectin-8 plays an important role in the development and progression of various cancers and cancer-associated bone remodeling processes. Galectin-8 was first isolated by using a rat liver cDNA library screening [26], and the homologous human counterpart was isolated from a prostate cancer and lung cancer cDNA expression library [14, 27] [28]. It was only observed in neoplastic cells and not in normal cells where galectin-8 modulates cell growth and adhesion, to possibly assist in metastasis [60]. Galectin-8 identified as lung tumor-associated antigen which at RNA level was shown to exist in six different isoforms. Nevertheless, wide distribution of galectin-8 has been reported in normal (brain, breast, colon, retina, kidney, pancreas, placenta, spleen, testis, uterus, vascular, oesophagus, and heart) and tumor (brain, breast, colon, germ cells, head and neck, kidney, muscles, ovary, pancreas, thyroid, placenta, prostate, uterus, lung, stomach and oesophagus) tissues (summarised in ref [61]). Galectin-8 stimulates glioblastoma cell migration suggesting its role in tumor invasion in the brain parenchyma [62]. The expression of galectin-8 was inversely related to the tumor growth of human colon cancer [63]. A correlation between expression of galectin-8 and the degree of differentiation of squamous cell carcinomas and neuro-endocrine tumors was observed [64, 65]. Galectin-8 is strongly expressed in squamous cell carcinoma, very weakly in adenocarcinoma and showed no expression in small cell carcinoma [64]. In addition, expression levels of galectin-8 could be used as a marker for lung and colon cancers and antibodies specifically targeting galectin-8 could subsequently be used in cancer treatment [61]. Galectin-8 was found to express most abundantly in 59 out of 61 different cancer cell lines in comparison to other members of the family [66]. Galectin-8 interacts with CD166 and plays a critical role in vascular angiogenesis [67]. *In vivo* injection of matrix rich in extracellular proteins supplemented with galectin-8 induced angiogenesis [67]. Interestingly, galectin-8 (but not galectin-1, -2, -3, -7) was found to interact with podoplanin, a key molecule in lymphatic endothelial cells, potentially implying roles in lymphangiogenesis [68, 69].

1.4 Structural analysis

There are about 250 galectin structures deposited in the PDB, of which 150 structures are for human galectins. To date, 28 galectin-8 structures deposited in this resource including the isolated domains of the tandem-repeat and truncated full-length protein. All of which were solved using X-ray crystallography except one of these (PDB code 2YRO) which was solved by NMR. That structure represents the first visualization of the isolated galectin-8 domain (Table 1.1). Amongst the X-ray structures, there are 22 structures for the isolated galectin-8N

either in complex with various naturally occurring carbohydrate ligands (17) or unliganded forms (5). Due to the presence of the long protease susceptible linker peptide, attempts to determine structure of the full-length galectin-8 were restricted [70]. Crystallisation of full-length galectin-8 was carried out by replacing the protease susceptible long linker with two amino acid residues (His-Met) [19, 71]. There are four such truncated full-length galectin-8 structures, two in complex with natural carbohydrate ligands, one bound to a peptide and one unliganded. The CRDs of galectins including the CRDs of tandem-repeat galectins share sequence similarity of about 35-40 % (Figure 1.4). However, with the low overall sequence similarity, the majority of galectins share the two sequence motifs H-X-NPR and WG-X-E-X-R that are responsible for galactose recognition. Galectin-8*N* domain is about 20 amino acid longer than galectin-8*C* (Figure 1.4).

Table 1.1: List of galectin-8 structures either full-length or isolated galectin-8*N* and galectin-8*C* deposited in the protein data bank.

S. No.	PDB	Year	Domain	Resolution (Å)	Ligand
1	2YRO	2008	G8C	-	-
2	2YV8	2008	G8N	1.92	-
3	2YXS	2008	G8N	2.13	Lactose
4	3AP4	2011	G8N	2.33	Lactose
5	3AP5	2011	G8N	1.92	-
6	3AP6	2011	G8N	1.58	Lactose-3'-sulfate
7	3AP7	2011	G8N	1.53	Lactose-3'-sialic acid
8	3AP9	2011	G8N	1.33	Lacto-N-fucopentaose III
9	3APB	2011	G8N	1.95	-
10	3OJB	2011	G8N	3.01	-
11	3VKL	2012	G8full	2.55	Lactose in each domain
12	3VKM	2012	G8full	2.98	Lactose-3'-sialic acid (G8N) and lactose(G8C)
13	4FQZ	2012	G8full	2.80	-
14	3VKN	2012	G8N	1.98	-
15	3VKO	2012	G8N	2.08	Lactose-3'-sialic acid
16	4GXL	2013	G8C	2.02	NDP52
17	4HAN	2013	G8full	2.55	NDP52 (G8C) and NAD (G8C)
18	4BMB	2014	G8N	1.35	Lactose
19	4BME	2014	G8N	2.00	Lactose
20	5GZG	2016	G8N	2.00	Lactose
21	5GZF	2016	G8N	2.00	Lactose
22	5GZE	2016	G8N	1.32	Lactose
23	5GZD	2016	G8N	1.19	Lactose
24	5GZC	2016	G8N	1.08	Glycerol
25	5T7U	2016	G8N	1.58	Glycerol
26	5T7S	2016	G8N	1.90	Lactose
27	5T7I	2016	G8N	2.00	Lacto-N-neotetraose
28	5T7T	2016	G8N	1.96	Lacto-N-tetraose

```

Gal-1/1-135 1--MACGLVASNLNLKPGECLRVGEVAPDAKSFVLNLGKDS----NNLCLHFNPRFNAHGDAITVCNSKDGGAWGTEQRE--
Gal-2/1-132 1--MTGELEVKNMMDMKPGSTLKITGSIADGTDGEVINLGGGT----DKLNLHFNPRFSE---STIVCNSLDGSNWGQEQRE--
Gal-3/1-136 1--IVPYNLPLPGGVVPRMLITILGTVKPNANRIALDFQRG----NDVAFHFNPRENENN--RRVIVCNTKLDNNWGREERQ--
Gal-4C/1-130 1--YFGRLLQGGLTARRTIIIGGYVPPTGKSFAINFKVGS----SGDIALHINPRMGNGT---VVRNSLLNGSWGSEEEKI-
Gal-4N/1-132 1--YYGPPIPGGLNVGMSVYIQGVASEHMKRFVNFVVGQ--DPGSDVAFHFNPRFDGW--DKVVFNTLQGGKWSSEERK--
Gal-7/1-136 1MSNVPHKSSLPGLRPGTVLIRGLVPPNASRFHVNLLCGE--EQGSDAALHFNPRLDTS---EVVFNNSKEQGSWGREERG--
Gal-8C/1-131 1-----FAARLNTPMGPRTVVVKGQEVNANAKSFNVDLLAGK---SKDIALHLNPRNLNIKA---FVRNSFLQESWGQEEERNI-
Gal-8N/1-134 1-----FVGITPDQLDPGLTIVIRGHVPSDADRFOVDLQNGSSMKPRADVAFHFNPRFKRA---GCIVCNTLINEKGREEIT--
Gal-9C/1-129 1-----FITITLGLYPSKSIILLSTVLPQAQRHINLCSG-----NHIAFHNLNPRFDENA---VVRNTQIDNSWGSEERSLP-
Gal-9N/1-96 1-----FSGNDIAFHFNPRFEDG---GYVVCNTRQNGSWGPEERK--
Gal-12C/1-125 1-----CSHALPQGLSPGGVIVRGLVLQEPKHFTVSLRDQA---AHAPVTLRASFAVRT---LAWISRWGOKKLI-S-
Gal-12N/1-135 1-----YVTTIFGGHLGAKMVMLOGVVPLDAHREQVDFQCGCSLCPRPDI AFHFNPRFHTTK--PHVICNTLHGGRWQREARW--

Gal-1/1-135 AVFRRFPQPSVAEVCITFDQANLTVKLDPGYEFKFPNRLN-LEAINYMAA DQDFKIKCVAFD-- 135
Gal-2/1-132 DHLCFSPQSEVKFTVTFTESDKFKVKLPDGHETFPNRLG-HSHLSYLSV RQGFNMSSFKLKE- 132
Gal-3/1-136 SVFPFESGKPFKIQVLVEPDHFKVAVNDHAHLQYNHRVKKLNEISKLGISDDIDLTASASYTMI 136
Gal-4C/1-130 THNPFPGQGFDDLSIRCGLDRFKVVYANGQHLDFAHRLSAFORVDTLEI QGDVTLSTVYQI-- 130
Gal-4N/1-132 RSMPPFKKGAAAFELVFIVLAEHYKVVVNGNPFYEYGHRLP-LQMVTHLQV DQDLQLQSI NFI-- 132
Gal-7/1-136 PGVPFQRQPPFEVLIIASDDGFKAVVGDAQYHHFRHRLP-LARVRLVEV GQDVQLDSVRIF-- 136
Gal-8C/1-131 TSFPFSPGMYFEMI IYCDVREFKAVNGVHSLEYKHRFKELSSIDTLEI NODIHLLEVRSW-- 131
Gal-8N/1-134 YDTPFKREKSF EIVIMVLKDKFQAVNGKHTLLYGHRI G-PEKIDTLGI YGKVNHSIGFS-- 134
Gal-9C/1-129 RKMPPVRGQSF SVWILCEAHCLKVAVDGQHLFEYHRLRLNLP T INRLEV GQDIQLTHVQT-- 129
Gal-9N/1-96 THMPFQKGMPFDLCFLVQSSDFKVMVNGILFVQYFHRVP-FHRVDTISVINGSVOLSYISFQ-- 96
Gal-12C/1-125 APFLYPRQRF FEVLLLFQEGGLKLALNGGGLGATSMNQALEQLRELRI SCVOLYCVHS-- 125
Gal-12N/1-135 PHLALRRGSSFLILFLFGNEEVKVSVNGQHLHFHRYRLP-LSHVDTLGI FQDILVEAVGFL-- 135

```

Figure 1.4: Multiple sequence alignment for CRDs of human galectins obtained using EMBL-EBI Clustal omega [72, 73]. The residues are colored based on their conservation using Jalview [74].

A typical CRD structurally exhibits a “jelly-roll” topology made up of the β -sandwich fold with the strands S1-S6 forming the concave face and the strand F0-F5 forming the convex face (Figure 1.5) [14]. Structurally, galectin-8N and galectin-8C exhibit identical topology as is the case with other galectins. The presence of the long S3-S4 loop bearing an arginine (Arg59) is a unique feature to galectin-8N, not observed in any other galectins, including the galectin-8C (Figure 1.5). Another unique difference lies in the presence of isoleucine (Ile91) on S6 in the conserved WG-X-E-X-R motif of galectin-8N in place of arginine. In addition, other differences in the galectin-8N CRD include Gln47 on stran S3, and Tyr141 on strand S2 unlike to that in galectin-8C and other galectins. These differences in the galectin-8N and galectin-8C domains are responsible for differential glycan-binding profile and the altered binding specificity [75, 76].

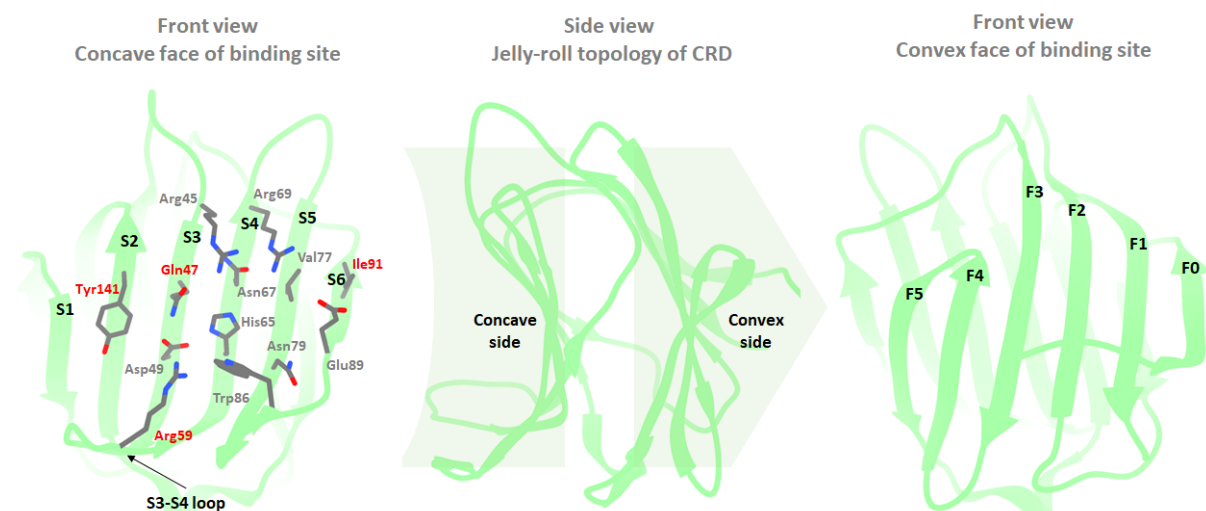


Figure 1.5: Carbohydrate recognition domain of galectin-8N [3AP5 [76]]. In the center is the side-view of the jelly-roll topology in green ribbons. On the left is the carbohydrate binding site of the CRD with the sticks (carbon gray; oxygen red; nitrogen blue) representation of amino acid residues involved in interaction with glycans. On the right is the convex surface of the CRD.

1.5 Glycan binding specificity

The two CRDs of galectin-8 are functionally dependent on each other, and a full-length linker peptide is required to attain proper orientation of the CRDs for efficient lectin function [34]. There exists a dramatic difference in glycan-binding specificity between the two domains of galectin-8, although the correct orientation of the two domains is required for conducting certain galectin-8 functions [42, 44]. Overall, the full-length galectin-8 has low affinity for complex-type glycans while high affinity for repeated-type ligand (*N*-acetyllactosamine) [41]. Substitution of the 3'-position in galactose of *N*-acetyllactosamine with a sulfate or sialic acid moiety increased the binding affinity [42, 77]. Binding of cell surface ligands with Gal β 1-3GlcNAc or Gal β 1-3GalNAc as basic motifs towards galectin-8 was better than that with canonical Gal β 1-4GlcNAc [77, 78].

Galectin-8*N* recognises a broad range of carbohydrates with distinct affinities and shows a unique predominance in binding 3'-SiaLac and 3'-sulfated lactose over galectin-8*C* (K_d values of 0.5–30 μ M for galectin-8*N* and 97–230 μ M for galectin-8*C*) [41, 44]. The presence of Ile90 as a part of a conserved motif in galectin-8*N* instead of arginine suggests its differential binding specificity [14, 26]. Galectin-8*C* shows much lower affinity in recognising ligands compared to galectin-8*N* but shows higher affinity for N-glycan-type branched oligosaccharides (K_d values of 26–52 μ M for galectin-8*C* and 47–290 μ M for galectin-8*N*) [41]. The most potent saccharides recognised by galectin-8*C* are the blood group A determinant and GalNAc β 1,3(Fuc α -1,2) Gal [44]. The presence of Arg59 on the long S3–S4 loop in galectin-8*N*, together with Arg45 and Gln47, located on the S3 strand, through hydrogen bonding interactions help recognise and strongly bind α -2,3-sialylated oligosaccharides [76]. In addition, the structural difference on the S4–S5 loop is thought to be necessary for the recognition of branched oligosaccharides [19].

Apart from its major contribution to the preferential recognition of anionic saccharides by galectin-8, galectin-8*N* exhibits specificity in binding the milk group antigens. These glycans are essential components of human milk that regulate immune response against pathogens in infants [79]. Being rich in lactose-containing glycans, their interactions with the galactose recognising proteins become critical. The two tetrasaccharides lacto-*N*-tetraose (LNT) and lacto-*N*-neotetraose (LNnT), which closely resembles the LacNAc dimer and differ only by the glycosidic linkage at the non-reducing end, have been reported to bind several galectins [41]. However, there exist differences in binding affinities between the two

tetrasaccharides with galectin-8, as it is for galectin-3 [42, 44, 80]. For galectin-3, the affinities are relatively comparable with weaker binding for LNT; but the magnitude of difference in the case of galectin-8N is significant. Chapter 3 provides a structure-based rationale for variation in binding affinities between the two tetrasaccharides [81].

1.6 Galectin-8 and bone function

1.6.1 Osteoporosis

Bone is continuously being formed and destroyed in the body at a constant rate to maintain the overall bone mass. There exists a delicate balance between the activity of bone-forming cells (osteoblasts) and bone-metabolising cells (osteoclasts). This delicate balance in the activity of osteoblast and osteoclast is maintained through interactions with extracellular proteins. Osteoporosis is a condition where excess bone demineralisation causes loss of bone strength thereby making the bone fragile and fracture-prone. Disturbance in the activity of these bone remodeling cells leads to diseases of bone like osteoporosis and bone cancer. Worldwide, one in three women and one in five men over the age of 50 experience osteoporotic fractures (International Osteoporosis Foundation). The current range of treatment approved by the FDA includes either anti-resorptive agents or anabolic agents or dual acting that mainly restores the lost balance between bone formation and bone metabolism (Table 1.2).

Table 1.2: List of FDA approved drugs and their mechanism of action for treatment of osteoporosis.

Table 12: List of FDA approved drugs and their mechanism of action for treatment of osteoporosis.	
Drugs	Mechanism of action [82]
Anti-resorptive agents – slow down the bone metabolism and prevents bone loss and decreases the risk of fracture.	
Bisphosphonates anti-resorptive agents	
Alendronate	These compounds apart from having affinity for bone matrix material hydroxyapatite, inhibits farnesyl pyrophosphate synthase and thereby prevents the biosynthesis of isoprenoid lipids (FPP and GGPP) that are essential for the post-translational farnesylation and geranylgeranylation of small GTPase signalling proteins. This overall inhibits osteoclast activity and reduces bone resorption and turnover.
Risedronate	
Ibandronate	
Zoledronic acid	
Other anti-resorptive agents	
Denosumab	It is a monoclonal antibody specific to receptor activator of nuclear factor kappa-B ligand (RANKL) and thereby prevents RANKL from activating its receptor, RANK, on the surface of osteoclasts. Prevention of this interaction inhibits osteoclast formation, function, and survival, thereby decreasing bone resorption and increasing bone mass and strength.

Raloxifene	It binds to estrogen receptors, resulting in differential expression of multiple estrogen-regulated genes in different tissues, thereby acting as an estrogen agonist in pre-osteoclastic cells, which results in the inhibition of their proliferative capacity and overall bone resorption.
Hormone replacement therapy (HRT)	By restoring the estrogen levels, HRT helps slow down the rate of bone-loss, that ultimately leads to increase in bone mass density and reduced risks of fracture [83].
Anabolic agents - helps to make new bone, increases bone density and can also reduce the risk for a broken bone	
Teriparatide	It is recombinant human parathyroid hormone (PTH) which acts as a regulator of calcium and phosphate metabolism in bone and kidney. Daily injections stimulate new bone formation leading to increased bone mineral density.
Dual action	
Strontium ranelate	The drug is dual acting; increases deposition of new bone by osteoblasts and reduces the resorption of bone by osteoclasts. Bone-formation action involves increased osteoblast differentiation and activity by increasing expression of the master gene <i>Runx2</i> and bone sialoprotein [84], osteoblast survival and regulation of osteoblast-induced osteoclastogenesis. Bone-resorption action involves decreased osteoclast differentiation, activity and increases apoptosis [85].

1.6.2 Galectins and Bone function

Skeletal tissues are mainly composed of extracellular matrix containing collagen; that maintains the structural integrity of the bone. It also contains matricellular protein such as galectins that regulate cellular functions by interacting with integrins and other secreted proteins [86]. Varied expression profiles of galectin-1 and galectin-3 during the osteoblast development stage indicated their role in cell-to-cell and cell-matrix interactions [87, 88]. Galectin-3 is an endogenous substrate for matrix metalloprotease-9 and acts as a downstream regulator of osteoclast recruitment during endochondral bone formation [89]. Galectin-9 induces proliferation of the human osteoblasts through clustering of lipid rafts on the membrane and regulates bone metabolism through differentiation of osteoblasts via the CD44/Smad signalling pathway [90, 91]. Chicken galectin-1 (CG-1A) and galectin-8 (CG-8) are involved in the formation of pre-cartilage mesenchymal cells in developing limbs, although their effects are opposite to each other [92]. CG-1 knockdown inhibits the formation of skeletal elements while knockdown of CG-8 enhances it [92]. Matricellular proteins thus play a critical role in overall development and homeostasis of bone [93].

1.6.3 Galectin-8 and Osteoporosis

Galectin-8 is widely distributed throughout the body, it aids in settlement of metastatic prostate cancer cells in the bone, thereby causing bone lesions. Galectin-8 induces expression of relevant metastatic factors like matrix metalloproteinase-9 (MMP-9), bone morphogenetic protein 2A and urokinase-type plasminogen activator [94]. Adhesion of metastatic myeloma cells to endothelial cells of the vasculature is the first event in plasma cell metastasis in bone marrow. Galectin-8 is overexpressed in plasma of multiple myeloma patients compared to healthy individuals and promotes the myeloma cell adhesion [95].

Of particular interest is the recent report of galectin-8's role in the regulation of bone remodeling process (Figure 1.6) [96]. Exposure of osteoblasts to galectin-8 increased the production of RANKL, a protein responsible for bone metabolism, and those cells differentiated into osteoclasts when grown with bone marrow stem cells. The RANKL transcription was mediated by the ERK signalling pathway, with low-density lipoprotein-related protein 1, mannose receptor C and urokinase plasminogen activator receptor as the galectin-8's receptors. Furthermore, transgenic mice overexpressing galectin-8 resulted in a RANKL-mediated decrease in bone mass density causing apparent bone loss and increased bone turnover [96]. These results not only highlighted a novel mechanism of the bone remodeling process but also showed galectin-8 to be an osteoclastogenic agent. Therefore, galectin-8 inhibition holds potential for the newer approach in tackling bone-loss diseases like osteoporosis.

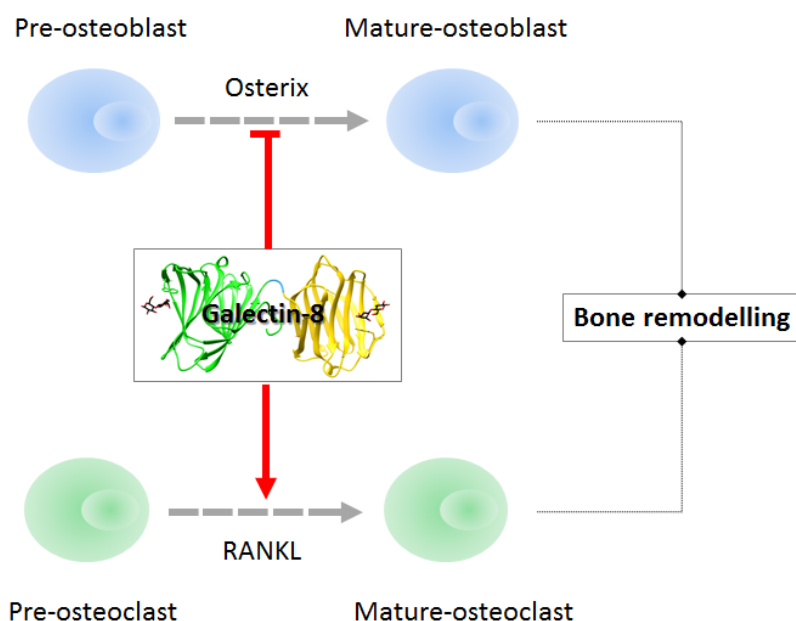


Figure 1.6: Galectin-8 as a regulator of differentiation and maturation of osteoblast and osteoclast.

1.7 Structure-based drug design and inhibitors of galectin

The advancement in macromolecular structure determination methods such as X-ray crystallography, NMR spectroscopy and electron microscopy have largely contributed towards the field of structure-based drug design. The determined protein structures provide valuable insight into the active site of the target protein at the atomic level. This gained information is then utilised to tailor the design of ligands specific to the protein of interest. The initial hits identified are then optimised taking into consideration of the binding site residues and topology, to further modulate the potency, efficacy, safety and progressing them from lead to a drug stage

Carbohydrates provide a diverse arena of ligand space by their varied shape, orientation and composition. This diversity can potentially be employed towards the development of new therapeutics with either pure carbohydrates or in conjugation with peptides (glycopeptide) and proteins (glycoproteins). Examples of carbohydrate-based drugs approved by FDA includes anti-diabetic drugs such as Acarbose (Bayer AG), Voglibose (Takeda/Abbott) and Miglitol (Bayer); antiviral (influenza) drugs such as Zanamavir (GlaxoSmithKline), Oseltamivir (Roche); anti-epileptic: Topiramate (Johnson and Johnsons), etc. [82]. Examples of FDA approved glycopeptides include antibacterial antibiotic: Vancomycin (Mylan), Telavancin (Theravance) anticancer: Bleomycin (Bristol), etc. Examples of FDA approved glycoproteins include erythropoietin (Amgen), interleukin-2 (Chiron), tissue plasminogen activator (Genetech) and various monoclonal antibodies [82].

Galectins have emerged as a therapeutic and prognostic target of interest due to their context-dependent involvement into various metabolic and disease conditions. The therapeutic potential for certain disease lies in inhibiting the functions of galectin, while under certain circumstances the lectin by itself can be employed for the therapeutic application. The CRD of galectins shares evolutionarily conserved amino acid sequence motifs that mainly recognise the galactose/lactose core unit of the glycans. Crystal structures of lactose bound to galectin-1 (2ZKN [97]), galectin-2 (5DG1 [98]), galectin-3 (2NN8 [99]), galectin-4N (5DUV [100]), galectin-4C (4YM3 [101]), galectin-7 (4GAL [102]), galectin-8N (5T7S [81]), galectin-8C (as part of truncated full-length protein; 3VKL [19]), galectin-9N (2EAK [103]) and galectin-9C (as a part of truncated full-length protein; 3WV6), reveal identical placement and interactions of the lactose molecule with the galectin binding site of the galectin, highlighting the fact that ligands designed for one galectin may potentially be recognised by other galectins. A better

understanding of the common and unique interaction features of available inhibitors with galectins may, therefore, contribute to the fine-tuning of inhibitor design.

To this end, a systematic analysis of all the galectin-1 inhibitors reported in the literature and those patented was carried out (Appendix 1.9). Galectin-1 is one of the most exhaustively characterised members of the galectin family. In our review, two aspects were covered one being the inhibitors of galectin-1 and the second being galectin-1 acting as a therapeutic molecule in various diseases. The inhibitors span from synthetic monosaccharide-based carbohydrates (galactose- and talose-based) to complex natural carbohydrates (modified citrus pectin and Davanat) to peptide-based (Anginex and G3-C12) ligands with varying levels of target specificity [104-106]. The taloside-based scaffold identified through structure-based design is a good example where identical scaffolds bind to various galectins nevertheless specificity towards a galectin can, however, be achieved by adding suitable groups at the C2 position. [18, 107, 108]. Some of the lead molecules that are actively pursued galectin inhibitors by the pharmaceutical companies, include GSC100 (La Jolla Pharmaceuticals), GM-CT-01 (Galectin Therapeutics) and GM-MD-02 (Galectin Therapeutics). However, the structural complexity of these molecules and unknown mechanisms of action prompt for designing small molecule inhibitors that can be structurally investigated to obtain the information about the exact site of interaction within the protein. Taken together, information from this galectin-1 review helped us better understand the existing inhibitor pool in the galectin field and will ultimately guide us toward designing galectin-8 specific inhibitors.

1.8 Aims and scope of the research

Osteoporosis, a disturbed bone homeostasis disease is one of the leading causes of bone fractures in people over the age group 50. The treatment regimen approved by FDA either directly or indirectly restores the lost balance between bone formation and bone metabolism. However, with increasing disease burden and widespread distribution of osteoporosis throughout the world, there is a need to employ novel approaches to directly tackle the imbalanced homeostasis. Galectin-8 to this end regulates the bone remodelling process by inducing expression of bone metabolising factors, and its inhibition holds the potential of being a novel approach for tackling diseases associated with bone-loss including osteoporosis. The primary aims of the project are characterising atomic level interactions of galectin-8N with its natural oligosaccharides to gain insight into functional involvement of galectin-8 and identifying hotspots in the galectin-8N binding site that subsequently will guide the ligand design process towards developing potential inhibitors of galectin-8.

A structure-based ligand design campaign was initiated to identify potential lead molecules targeting galectin-8 that were subjected to biological evaluations. Galectin-8 being a tandem-repeat, ligand design towards the galectin-8 N was focussed mainly due to the broad glycan recognition profile and interesting hotspots in the carbohydrate binding site. It is anticipated that interfering with the functionality of one domain would affect the overall functioning of the tandem-repeat. Furthermore, ligand design towards the galectin-8 C is a potential future work. In the present thesis, a combination of theoretical and experimental methods was employed towards practising rational medicinal chemistry. Key interaction hotspots in the galectin-8 N binding site were identified, based on which the interaction filters were designed to virtually screen a library of non-carbohydrate-based compounds. Taking together the information generated, three ligands in total were designed based on the native galactose core. These ligands mainly exploit the evolutionarily conserved and unique amino acid residue in the binding site for interactions. The designed ligands were synthesised, evaluated for binding with a combination of various techniques including STD NMR, ITC, SPR and X-ray crystallography. Biological evaluations (by our collaborators) are underway to investigate the effects of the designed ligands on galectin-8 functions both *in vitro* and *in vivo*.

1.9 Appendix

Statement of contribution to a co-authored publication

This chapter includes a co-authored paper. The bibliographic details of the co-authored paper, including all authors, are:

Helen Blanchard, Khuchtumur Bum-Erdene, **Mohammad Hussaini Bohari** and Xing Yu (2016). Expert Opinion on Therapeutic Patents. 26(5):537-54. doi: 10.1517/13543776.2016.1163338.

My contribution to the review involved:

My contribution to review included curating the patent information from Lens database, contributing to the writing of the abstract, article highlights, introduction, “structure of galectin-1” section, “Galectin-1 as a therapeutic molecule” section and contributed towards writing the conclusion and expert opinion.

(Signed) _____ (Date) 03-04-2017 _____

Mohammad Bohari

(Countersigned) _____ (Date) 24/11/2017 _____

Corresponding author of paper: Professor Helen Blanchard

(Countersigned) _____ (Date) 24/11/2017 _____

Supervisor: Professor Helen Blanchard

REVIEW

Galectin-1 inhibitors and their potential therapeutic applications: a patent review

Helen Blanchard, Khuchtumur Bum-Erdene, Mohammad Hussaini Bohari and Xing Yu

Institute for Glycomics, Griffith University, Gold Coast Campus, Queensland, Australia

ABSTRACT

Introduction: Galectins have affinity for β -galactosides. Human galectin-1 is ubiquitously expressed in the body and its expression level can be a marker in disease. Targeted inhibition of galectin-1 gives potential for treatment of inflammatory disorders and anti-cancer therapeutics.

Areas covered: This review discusses progress in galectin-1 inhibitor discovery and development. Patent applications pertaining to galectin-1 inhibitors are categorised as monovalent- and multivalent-carbohydrate-based inhibitors, peptides- and peptidomimetics. Furthermore, the potential of galectin-1 protein as a therapeutic is discussed along with consideration of the unique challenges that galectin-1 presents, including its monomer-dimer equilibrium and oxidized and reduced forms, with regard to delivering an intact protein to a pathologically relevant site.

Expert opinion: Significant evidence implicates galectin-1's involvement in cancer progression, inflammation, and host-pathogen interactions. Conserved sequence similarity of the carbohydrate-binding sites of different galectins makes design of specific antagonists (blocking agents/inhibitors of function) difficult. Key challenges pertaining to the therapeutic use of galectin-1 are its monomer-dimer equilibrium, its redox state, and delivery of intact galectin-1 to the desired site. Developing modified forms of galectin-1 has resulted in increased stability and functional potency. Gene and protein therapy approaches that deliver the protein toward the target are under exploration as is exploitation of different inhibitor scaffolds.

ARTICLE HISTORY

Received 6 September 2015
Accepted 4 March 2016
Published online
25 March 2016

KEYWORDS

Galectin-1; carbohydrate;
inhibitor design; patents

1. Introduction

Galectins are a family of lectins that specifically recognize β -galactoside containing glycans. Recognition of the carbohydrate moiety, either located terminally or internally within glycan chains, is by the galectin carbohydrate recognition domain (CRD).[1] The galectin family members share conserved amino acid sequence motifs within their carbohydrate-binding sites. Based on the structural organization of their CRDs, galectins are classified into three types. Prototype galectins (galectin-1, -2, -5, -7, -10, -11, -13, -14, and -15) have a single CRD (~130 amino acids) and can form non-covalent homodimers; tandem-repeat galectins (Gal-4, -6-8, -9, and -12) have two non-identical CRDs joined by a linker of variable length, and chimeric galectin (galectin-3 being the only identified member in vertebrates) has one CRD and a non-lectin N-terminal domain.[2] The multivalent nature of glycans and the oligomeric tendencies of galectins correlate well together and can result in a higher binding affinity.[3,4] Glycan binding affinity and exact specificity for larger oligosaccharides varies across the galectin family, including between the two CRDs of a tandem-repeat galectin.[5]

Galectin-1 is ubiquitously expressed throughout the body and is involved in the regulation of cell growth, adhesion, signaling, differentiation, development, immune system and host-pathogen interactions.[6-9] Galectin-1 also plays important roles in the embryonic development of primary sensory neurons and their synaptic connections in the spinal cord for

example shown in the formation of the neural network of the olfactory bulb of mice.[10] Altered expression of galectin-1 has been associated with various neurological diseases with correlation of increased expression with regenerative success following injury, and an implicated protective role with respect to ischemic brain injury.[11,12]

Despite lacking a secretion signal sequence, galectin-1 is secreted into the extracellular matrix via a non-classical pathway.[13,14] Galectins can be found inside the cell, either in the nucleus or in the cytoplasm, or outside the cell either associated with the membrane or the cell matrix.[15] Regulation of cell behavior via extracellular pathways is a well-characterized function of galectins; growing evidence of their interaction with glycans on the surface of bacteria/viruses suggests a potential role as pattern recognition receptor.[16] Expression profiles of galectin-1 in the various stages of cancer progression and its role in the tumor microenvironment have been thoroughly reviewed.[17-24] There is clear evidence of galectin-1 as a diagnostic tumor marker,[25] which along with the direct evidence of galectin-1 involvement in tumor angiogenesis and growth [26-29] further highlights the potential of galectin-1 inhibitors as anti-cancer agents.[30,31] Overexpression of galectin-1 in malignant tumors, in particular glioblastoma, is associated with the development of drug resistance, thus a combination therapy incorporating galectin-1 inhibitors may enhance the efficiency of co-administered drugs.[32,33] In addition, galectin-1 enhances the binding affinity of HIV-1 glycoprotein gp120 (spike protein) to susceptible host cells and

CONTACT Helen Blanchard  h.blanchard@griffith.edu.au

© 2016 Informa UK Limited, trading as Taylor & Francis Group

Article highlights

- The pathological effects of galectin-1 are dependent on the biological context.
- Significant correlation between the over-expression of galectin-1 and certain cancers underscores the relevance of galectin-1 antagonists as prospective anti-cancer reagents.
- Monovalent and multivalent inhibitors of galectin-1 function are underexplored with respect to their specificity and specific mechanisms of action.
- Galectin-1 can induce both pro- and anti-inflammatory effects through the regulation of immune cells.
- The use of galectin-1 as a therapeutic agent faces protein-specific challenges in addition to the poor pharmacokinetics of protein delivery.

This box summarizes key points contained in the article.

increased viral infectivity.[34] This occurs via direct binding of galectin-1 to clustered glycans of gp120 in a carbohydrate-dependent manner. Galectin-1 specific inhibitors have been shown to decrease the viral infectivity by reducing the binding of gp120 to host cells. Therefore, galectin-1 specific inhibitors may be effective in combating HIV-1 infection.[35,36]

Galectin-1 uses the differential glycosylation pattern on T-helper cells to selectively induce apoptosis in activated Th1 and Th17 cells.[37,38] In general, galectin-1 is involved in homeostasis by turning down the T-cell immunity, while galectin-3 is a pro-inflammatory molecule [39–45] and reviewed in the works of Rabinovich and colleagues.[46–48] Galectin-1 has therefore emerged as a potential treatment option for autoimmune and other inflammatory conditions. [49–52] Galectin-1 has mainly been shown to exhibit an anti-

inflammatory effect on polymorphonuclear neutrophils [53–55] though conversely, galectin-1 has also been shown to demonstrate a pro-inflammatory effect on neutrophils.[54] Recognizing the context-dependent multifunctionality of galectin-1 is important in consideration of galectin-1 as a therapeutic target (as reviewed in Smetana et al.[56]).

2. Structure of galectin-1

Structural characterization of galectin-1 has provided the basis for understanding its binding modes toward different glycans and has also facilitated the rational design of inhibitors. Galectin-1 has a β -sandwich fold formed by two anti-parallel β -sheets of five (F1–F5) and six (S1–S6) β -strands, the latter forming the carbohydrate-binding site (Figure 1A).[57] Upon lactose binding, the pyranose ring of galactose stacks against the conserved Trp68 residue and the hydroxyl groups at the C4' and C6' of galactose and the C2 and C3 of glucose form hydrogen bonds with His44, Arg48, Asn61, and Glu71 residues (Figure 1B). His52 of the large S4–S5 loop, unique for galectin-1, is positioned in the vicinity of the lactose molecule.

The carbohydrate binding sites are positioned at the opposite ends of the galectin-1 dimer, with the N- and C-termini of each galectin-1 monomer coming together at the dimer interface and a hydrogen bonding network established between the two monomers (Figure 1A and C). Dimer formation permits bivalent interactions with cell-surface glycans resulting in an overall enhancement of avidity and enables crosslinking and the formation of glycan-galectin-1 lattices.[3,4] There exists a reversible concentration ($K_d \sim 7 \mu\text{M}$) and time ($\sim 20 \text{ h}$)

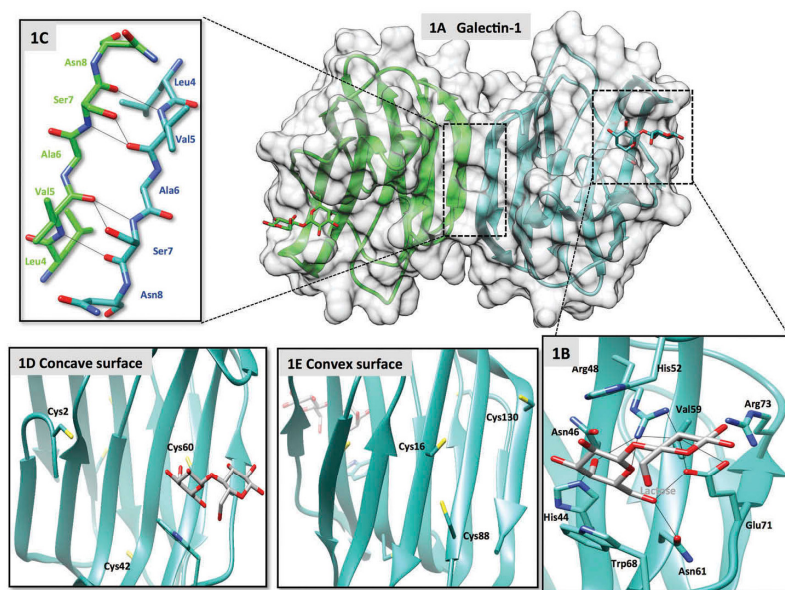


Figure 1. Galectin-1 bound to lactose (PDB ID 1GZW [57]). A) Ribbon representation with transparent surface of the dimer form of galectin-1 along with bound lactose (stick model). B) Amino acids depicted (cyan sticks, nitrogen blue and oxygen red) involved in hydrogen bonding (thin black lines) with lactose (grey sticks). C) Residues involved in hydrogen bonding at the dimeric interface. D) and E) Cysteine residues (sticks) present in galectin-1 on the convex and concave surface of the CRD. Lactose and the conserved tryptophan are displayed for reference.

dependent equilibrium between the monomeric and dimeric forms of galectin-1.[58] An increasing number of reports suggest that the dimeric form of galectin-1 is more prevalent and its dimeric nature contributes to an overall higher affinity for glycans compared to its monomeric form.[59,60]

The galectins were originally referred to as S-type lectins because of galectin-1's dependency on reducing conditions for its lectin activity.[61] This is due to the presence of six cysteine residues in the galectin-1 CRD that could, upon structural reorganization, potentially form disulfide bonds under a non-reducing environment.[62,63] In reduced galectin-1, Cys130, Cys16, Cys88, are present on the convex surface, while Cys2, Cys42, Cys60 are present on concave side (Figure 1D and E). Cysteine-to-serine single mutations did not significantly affect the carbohydrate binding profile of galectin-1, but rather increased stability on storage and prolonged asialofetuin binding activity.[64] Interestingly in another study, single mutants or double mutant of Cys2Ser and Cys60Ser showed reduced binding to lactosyl Sepharose by 40% and 80%, respectively, and the Cys2Ser and Cys130Ser mutant remained susceptible to oxidation.[65] Mutation of all cysteines demonstrated that there was no significant change in carbohydrate-binding affinity for a range of glycans.[66] The redox potential of the environment governs the redox status of galectin-1.[67] The four solvent exposed cysteines (Cys2, Cys16, Cys88, and Cys130) were mutated to serine and analyzed by circular dichroism, which revealed that all the single mutants except Cys130 affected the protein conformation.[67] In general, the pro-apoptotic and immunomodulatory functions are associated with the reduced form of galectin-1, while the proliferative effects, such as the regulation of axonal regeneration after nerve injury, were associated with its oxidized form.[68]

Binding of intracellular reduced galectin-1 to oncogenic Ras, preference is to H-Ras, is paramount to Ras membrane anchorage, a prerequisite for Ras-mediated signal transduction, regulating normal cell growth and malignant transformation.[69–71] A recent study [71] reported that reduced galectin-1 interacts with farnesyl in a carbohydrate-independent manner. Further, the interaction with farnesyl appears to occur at a site away from the galectin-1 carbohydrate-binding site, involving at least some hydrophobic residues at the N-terminus and the amino acid Lys28. This finding was supported by a previous study that proposed a farnesyl-binding pocket in reduced galectin-1, comprising N- and C-terminal hydrophobic residues L9, L11, L17, F30, L32, and I128.[70,72]

Being either a positive or negative regulator of various disease processes, galectin-1 offers an opportunity to be utilized as a therapeutic agent or as a drug target. The involvement of galectin-1 in cancer development, metastasis and angiogenesis allows the utility of galectin-1 as a tumor marker and the development of galectin-1 inhibitors as therapeutics. The role of galectin-1 in autoimmune diseases as a selective inducer of T-cell apoptosis warrants the development of galectin-1 as a therapeutic. This review surveys and highlights the available patents and patent applications regarding the therapeutic potential of galectin-1 inhibitors as well as the development of therapeutics that incorporate the galectin-1 protein.

3. Therapeutic potential of galectin-1 inhibitors

Galectin-1 inhibition, as with galectin inhibition in general, has been pursued for three different types of inhibitors. As galectin-1 binds galactosides, one category is modified mono- and di-saccharides containing galactose or its mimics. The propensity of galectin-1 to multimerize and its ability to recognize large complex glycans have been exploited by design of synthetic glycodendrimers and modified complex glycans. Finally, non-saccharide-based inhibitors, specifically the peptide anginex and its mimetics have been utilized to some success. The discovery of a non-saccharide small molecule inhibitor of galectin-1 has not yet been reported.

3.1. Monovalent carbohydrate-based inhibitors

As galectins recognize galactose as well as di- and oligosaccharides that contain galactose, the prevailing majority of drug discovery efforts have been focused on the synthetic modification of galactosides, lactosides, and their mimics, such as thiodigalactosides as well as talosides (C2 epimer of galactose). Each saccharide scaffold offers a unique advantage over the others with respect to exploiting aspects of affinity, stability, or selectivity.

The monosaccharide galactose (Figure 2 depicts compounds) (1) has an innately weaker affinity (10 mM via fluorescence polarization assay [73]) than that of di- or oligosaccharides, such as lactose (2) (190 μ M [73]) and thiodigalactoside (3) (TDG; 24 μ M [74]). However, monosaccharide-based inhibitors offer higher ligand efficiency and glycolytic stability, which could improve bioavailability and uptake. Due to the nature of the galectin–galactose interaction (Figure 1B), the C1 and C3 hydroxyls of galactose are modified to improve affinity and/or selectivity. Conjugation of decorated aromatic moieties at the C1 position increases the affinity of β -galactosides toward galectin-1 by an appreciable margin, up to 10–40-fold higher affinity,[73–77] but the increase in affinity does not reach the affinity of an unmodified lactose. The only such inhibitor to have higher affinity than lactose is (E)-methyl 2-phenyl-4-(β -D-galactopyranosyl)-but-2-enoate (4) with a reported 313- μ M affinity on a hemagglutination assay (2.5-fold higher than lactose at 800- μ M affinity in the comparable assay).[78] Galactose and lactose C1-linked to a purpurinimide (5) and (6) have been patented as targeted anti-cancer photodynamic therapeutics [79] (Patents listed in Table 1) by Health Research, Inc. The chlorin-conjugated galactose (5) displayed 22- μ M affinity toward galectin-1 and the conjugated lactose (6) displayed 0.54- μ M affinity via an ELISA assay. In addition to the *in vitro* affinity assay, the compounds along with the light therapy induced cytotoxicity against a number of cancer cells and 3 out of 6 mice treated with the therapy cleared the radiation-induced fibrosarcoma implantation. Interestingly, the affinity of the purpurinimide on its own was higher than that of galactose toward galectin-1, possibly suggesting a binding site for the tag despite the *in silico* modelling by the patent applicants suggesting otherwise. The effectiveness of the patented molecules and the accompanying therapy warrants further studies into the specificity of the compounds as well as the exact mechanism of tumor killing.

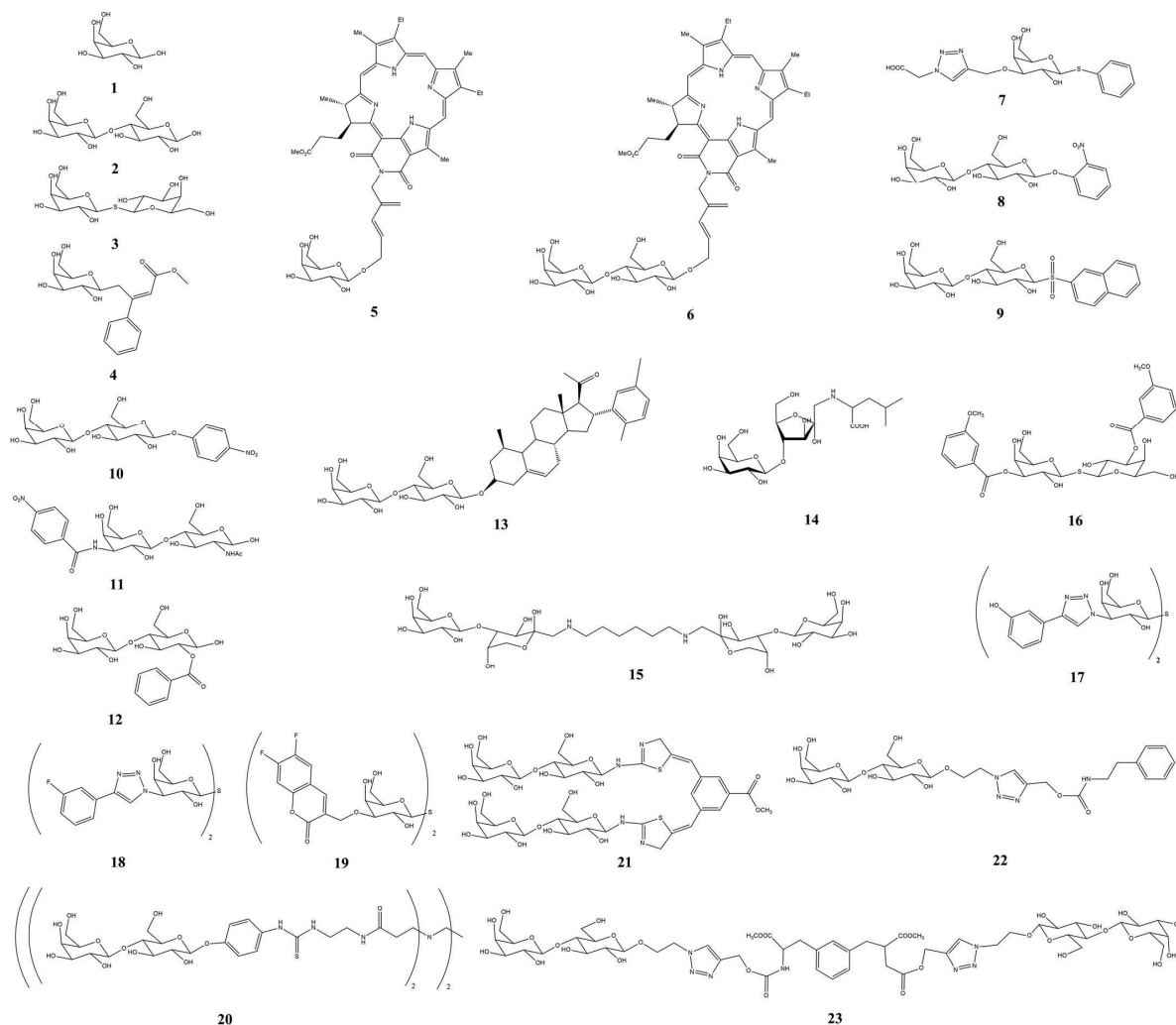


Figure 2. Galectin-1 inhibitors. Structures of known galectin-1 inhibitors.

Although a number of reports explored the potential of C3-modified galactosides with relative success in inhibiting galectin-3,[122–125] inhibition of galectin-1 has not been thoroughly explored. Simultaneous modification of galactose C1 and C3 resulted in an inhibitor (**7**) with 40-fold higher affinity toward galectin-1 than methyl galactoside, almost having the same affinity as methyl lactoside.[76] Considering that the C1 modification of this inhibitor was not optimal and only a limited number of C3 modifications were explored in the study,[76] it seems feasible to synthesize a high affinity and selective inhibitor of galectin-1 via the same fragment-based approach that was used to synthesize such galactoside-based inhibitors for galectin-3.[73]

Compared to galactose, lactose and *N*-acetyllactosamine have inherently higher affinity toward galectin-1. Due to the binding interaction between galectin-1 and the glucose portion of lactose, the hydroxyl moieties at C3, C4', and C6' are engaged, which leaves C1, C2, C6, C2', and C3' hydroxyls free

for modifications. Ether-linked 2-nitrophenyl (**8**) or sulfonate-linked (**9**) β -naphthalene groups at lactose C1 resulted in inhibitors with 10- and 20-fold higher galectin-1 affinity than lactose in a hemagglutination assay.[77] More importantly, these compounds significantly decreased the galectin-1-dependent enhancement of HIV-1 infection *in vitro*. [36] However, when the same group tested these and other compounds by a solid-phase assay, the above-mentioned compounds were not found to have high affinity.[126] In fact, only a single compound with lactose C1 *O*-linked *p*-nitrophenyl modification (**10**) resulted in a twofold affinity enhancement over lactose.[126] This highlights the problem in trying to compare the results of different affinity assays and possibly indicates a problem with the solid-phase assay.

Addition of aromatic groups at the C3' of lacNAc (**11**) resulted in up to 7.5-fold affinity enhancement over lacNAc toward galectin-1, though much higher affinity enhancements were observed against galectin-3.[127] Upon

Table 1. Patents and patent applications concerning various inhibitors and mimicking agents of galectin-1.

Patent	Applicant	General information
US20030109464A1 [80]	Sydney Kimmel Cancer Center; Curators of the University of Missouri The; La Jolla Institute for Allergy and Immunology	Mono- and di-valent lactulose-based glyco-amines as inhibitors of galectin-1/4 for the diagnosis, prognosis and treatment of breast cancer ^a
US9050352B2 [81]	Cancure Limited Acn	Simpler to complex oligosaccharides along with synthetic inhibitors, glycopeptides and glycopolymers – immunomodulators ^a
US6849607B2 [79]	Health Research Inc.	Saccharides conjugated with porphyrin-based photosensitizers for targeted photodynamic therapy – Cancer therapy ^a
US7700763B2 [82]	Galecto Biotech AB	TDG-based inhibitors of galectins. Most potent against galectin-3 ^a
US20140336146A1 [83]		
WO2014067986A1 [84]		
US7012068B2 [85]	Galectin Therapeutics	GM-CT-01, extraction and composition of and as treatment against cancer ^a
US7893252B2 [86]		
US8722645B2 [87]	Galectin Therapeutics	GR-MD-02, extraction and composition of and as treatment against cancer, fibrosis, nephropathy as well as inflammatory, autoimmune, cardiovascular and neurodegenerative diseases with elevated iNOS ^a
US8658787B2 [88]		
US8962824B2 [89]		
US20140086932A1 [90]		
US20140235571A1 [91]		
US20150147338A1 [92]		
US8877263B2 [93]	La Jolla Pharmaceutical	GCS-100, extraction and composition of and as therapeutic agent ^a
US20150133399A1 [94]		
US20030004132A1 [95]	Glycogenesys Inc	GCS-100, treatment for immune diseases, angiogenesis, neurodegenerative diseases and hyper-proliferative diseases ^a
US20040121981A1 [96]		
US20060014719A1 [97]		
US20060074050A1 [98]		
US2007010438A1 [99]	Regents of the University of Minnesota	Anginex in conjunction with radiotherapy as cancer treatment ^b
WO2006128027A1 [100]	Peptx Inc	Recombinant production of anginex and use as anti-angiogenic ^b
US7339023B2 [101]	Regents of the University of Minnesota	6DBF7 and partial mimetics of anginex as anti-angiogenic and anti-tumor agents ^b
US8716343B2 [102]	Regents of the University of Minnesota	OTX008 and other topomimetics of anginex as anti-bacterial, anti-angiogenic and anti-tumor agents ^b
WO2012131079A1 [103]	Oncoethix	OTX008 as galectin-1-targeting compound for the treatment of cancer ^b
WO2014070214A1 [104]	Regents of the University of Minnesota	OTX008 as galectin-1-targeting compound for the treatment of cancer ^b
EP2858681 [105]	Koninkl Philips Nv	Radiolabeled analogue of OTX008 as imaging tracer and to determine whether OTX008 containing pharmaceutical could be useful for a cancer patient ^b
WO2012061395A2 [106]	Regents of the University of Minnesota	PTX013, PTX015, and other improved topomimetics of anginex as anti-angiogenic and anti-tumor agents ^b
US9034325B2 [107]	Abylnx Nv	Nanobodies binding to multiscavenger receptors including that against galectin-1 – Alzheimer's disease, atherosclerosis, diabetes and arthritis ^b
WO2015013388A2 [108]	Dana Farber Cancer Institute, Inc.	Monoclonal antibody, and fragment thereof, directed against galectin-1 ^b
US8968740B2 [109]	Dana Farber Cancer Institute, Inc.; Brigham and Women's Hospital, Inc.; CONICET;	Monoclonal antibody directed against galectin-1 diagnosis/prognosis/ treatment EBV-associated-post-transplant lymphoproliferative disorder (PTLD) ^b
US7662385B2 [110]	Keio University, Advanced industrial science and technology	Anti-galectin-1 and anti-integrin-β1 mAbs as inhibitors of implanted neural stem cells for CNS injury ^b
US7964575B2 [111]	Universite Libre de Bruxelles; Universite Catholique de Louvain	RNAi-based approach to knockdown galectin-1 expression for the treatment of cancer ^b
EP2771367 [112]	Indiana University Research and Technology Corporation	Gigaxonin protein fused to a cell penetrating peptide for the treatment of giant axonal neuropathy, through the regulation of cellular galectin-1 levels ^b
US20070185014A1 [113]	Schepens Eye Research Institute	Modulation of goblet cell proliferation and secretion – dry eyes syndrome ^c
US20100004163A1 [114]	Tufts University	Therapeutic amount of galectin-1 with diluents – Dry eyes syndrome ^c
WO2002089831A1 [115]	Protegene Inc; Wada Jun	Prevention and treatment of nephritis, glomerular diseases with galectin-1, -3, -8 ^c
US5948628A [116]	University of Oklahoma	Galectin-1 and different mutants (C2S-galectin-1, V5D-galectin-1, N-galectin-1) – treatment and modulation of inflammation ^c
US6225071B1 [117]		
US8598323B2 [118]	The Brigham And Women's Hospital, Inc.	Galectin-1 fused with Fc region of human IgG1 – Treatment of immune dysfunction ^c
US6890531B1 [119]	Kirin Beer Kabushiki Kaisha	Oxidized galectin-1 for treating nerve injury, nerve degeneration ^c
US8513208B2 [203]	Argos Therapeutics Inc.	Cells transfected with mRNA encoding galectin-1 for immunomodulation – prevention/treatment of transplant rejection and other allergic reactions ^c
US20040023855A1 [120]	Mandalmed Inc.	Nanoparticles in delivery of galectin as protein/ nucleotides/antisense – anti-cancer ^c
TW201410702A [121]	National Cheng Kung University	Gold nanoparticles coated with galectin-1 – anti-rheumatoid arthritis ^c

For this review, patents covering either the use of galectin-1 or its modified forms or molecules directly targeting galectin-1 for therapeutic applications are covered while those patents on galectin-1 as a diagnostic or prognostic marker are excluded, as are those patents covering galectin-1 use as a cellular targeting and cell penetrating agent for other therapeutic molecules and other such applications. The patent data is collected from Lens database (<https://www.lens.org>).

^aCarbohydrate-based inhibitors.

^bNon-carbohydrate-based inhibitors.

^cProteins/peptides mimicking galectin-1.

changing the C2 *N*-acetyl group of lacNAc with an ester-linked aromatic (**12**), the affinity toward galectin-1 increased by up to 15-fold, though as with the C3' modifications, the enhancements toward galectin-3 were greater.[127,128] However, a recent report by van Hattum et al. demonstrated that it is feasible to switch the selectivity of saccharide-based inhibitors from one galectin to another.[74] The compounds showed increased binding affinity and selectivity toward galectin-3 over galectin-1, but the same structure-based inhibitor design approach could be applied to selectively inhibit galectin-1 over galectin-3. Although galectin-1 inhibition reports are sparse compared to galectin-3, one could already see the potential of combining the C1,[36,77] C2, [128] and C3'-linked [127,128] modifications that have been reported so far on a lactose scaffold to achieve a high affinity synthetic galectin-1 inhibitor.

In addition to the modification of the lactose structure to increase the binding affinity and selectivity, additional moieties with therapeutic potential could be conjugated to lactose, where the saccharide acts as a targeting agent that delivers the payload to galectin-rich cancer cells. An above-mentioned example of chlorin-conjugation works better with lactose compared to galactose, perhaps due to its higher affinity and selectivity toward galectins, and the compound has been patented together with the galactose-chlorin conjugates.[79] Similarly, a steroid-conjugated lactose molecule (**13**) displayed anti-migratory and cytotoxic effects on various cancer cells and also displayed *in vivo* synergy with cisplatin against lymphoma and glioblastoma models.[129]

Lactulosyl-L-leucine (**14**) has been hypothesized to function as a T-antigen mimic and is known to inhibit galectin-3.[130,131] The compound has also shown anti-metastatic [132] and pro-apoptotic [133] effects *in vitro* and *in vivo*. Despite the reported successful results, significant insight into its selectivity and target profile are lacking. In fact, synthetic mono- and divalent lactulose-amines were shown to function through galectin-1 and/or galectin-3 in inducing tumor-cell apoptosis and inhibition of cell aggregation and tubule formation.[134] Furthermore, Huflejt et al. applied for a patent [80] utilizing lactulosyl-L-leucine (**14**) and a synthetic divalent lactulose (**15**) for their use on tumors with high expression of galectin-1 or galectin-4, not galectin-3. The results in the patent application include *in vitro* and *in vivo* effects such as inhibition of galectin-1 interactions, inhibition of galectin-1-induced immunosuppressive effects and inhibition of tumor growth by the lactulosyl amines. An immunomodulatory composition containing molecules that interact with galectin-1, -3, and -9 were patented for prophylaxis/treatment of pathogenic infections, autoimmune diseases, transplant rejection, graft versus host disease, allergies, inflammatory disease, as well as cancers and tumors.[81] Interacting molecules included in the application range from simple disaccharides, such as lactose, lactulose, TDG, and other synthetic derivatives to larger saccharides, such as glycopolymers and polylactosamines.

Perhaps the most successful class of saccharide-based galectin inhibitors have been based on the TDG scaffold (**3**), which innately has higher affinity toward galectins than

galactose (**1**) and lactose (**2**).[127] As a consequence of the binding site arrangement of galectin-1, where the unique His52 of galectin-1 (Figure 1B) can interact with the larger sulfur-linkage,[135] TDG displays up to eightfold higher affinity to galectin-1 compared to that of lactose.[127] The unmodified TDG scaffold has already been shown to have effects *in vitro* and *in vivo*, displaying pro-inflammatory effects through galectin-1 antagonism.[135,136] The unmodified TDG is not selective toward a specific galectin, [127,137] and its effects cannot easily be attributed to the inhibition of one galectin over another. However, in the work of Stannard et al.,[135] there is significant evidence that the measured effects are due to TDG blocking galectin-1. This includes evidence from immunoblot analysis of whole cell protein analyzed using a polyclonal antibody raised against recombinant galectin-1 showing that the major galectin present in the breast cancer model is galectin-1. Galectin-2 is the most similar by amino acid identity but is confined to the gastrointestinal tract and galectin-3 levels secreted by breast cancer cells are reported as very low to almost negligible. Though it is known that TDG can interact with galectin-3 – and we know that *in vitro* this disaccharide binds galectin-3 as we have determined the X-ray crystal structure of the complex (Blanchard, *unpublished*), it is convincing that in the experiments reported by Stannard et al. [135] that it is galectin-1 that is being blocked by TDG and that this has an effect in ultimately reducing tumor progression.

The exploration of modifications on the TDG scaffold has been targeted toward the inhibition of galectin-3,[74,127,138] due to the largely negative effects of galectin-3 in many disease processes. Although the effective and selective inhibition of galectin-1 has not been the prime focus of previous studies, the reported off-target affinity toward galectin-1 [74,127,136] encourages the notion that TDG can be a successful inhibitor scaffold for galectin-1. Simple symmetric addition of a C3/C3' 3-methoxybenzoyl moiety (**16**) has resulted in 12-fold improvement in affinity over TDG and 93-fold improvement over lactose toward galectin-1.[127] In a recent report, modifications of the TDG scaffold with aromatic groups through a triazole-linkage resulted in nanomolar inhibitors (**17**) of galectin-1, with 490 and 1850-fold affinity improvement over TDG.[74] The study was successful in switching the selectivity of the molecules toward galectin-3 over galectin-1, [74] but the same structure-based drug design methodology could be utilized to develop selective and avid galectin-1 inhibitors. A number of patents have been filed that utilize TDG-based inhibitors to inhibit galectin-3 function,[82–84] including the TD139 compound (**18**),[84] which has shown promising *in vivo* effects,[128–130] as well as a new series of coumaryl derivatives (example is **19**) with a more stable O-linkage.[83] Even amongst those compounds that have been already patented against galectin-3, a number of compounds display high off-target effects against galectin-1. [82,83]

In addition to TDG, other saccharide mimics of the natural glycan partners of galectins have been described. Similarly to TDG, Gal β 1-4Fuc was described as a higher affinity alternative disaccharide compared to lactose for binding to galectins. [139] This disaccharide may be a possible alternative to the lactose/lacNAc as it offers higher innate affinity.[139] In

addition, a number of studies have tried to replace the name-sake galactose with alternate sugars such as mannose and talose, and had some success in improving selectivity toward certain galectins.[140–142] However, with respect to galectin-1, these alternative scaffolds have yet to be fully exploited.

3.2. Multivalent carbohydrate-based inhibitors

Due to the dependence of cell–cell recognition and cell signaling on the formation of multiple receptor–ligand complexes, the concept of multivalency has gained interest.[143] A major difficulty in the assessment of glycocluster-based inhibitors is the discrepancies that occur between different assays. The most common assays are the solid-phase competition assay and the hemagglutination assay, while some researchers have utilized fluorescence polarization or surface-plasmon resonance (SPR) assays. Galectin–glycocluster interactions are complex as exemplified by the formation of galectin clusters induced by multivalent binding partners, such as ASF, leading to synergistic affinity enhancements. As a consequence, the type of competition matrix used in the solid-phase assay affects the inhibition capacity of the inhibitors.[144–148] As a consequence, a given multivalent inhibitor may show highly efficient inhibition of galectin-1 in a solid-phase assay in competition with ASF, while displaying very poor competition against laminin, while a SPR or hemagglutination assay might find the compound to be completely inactive.[145,147,149] As with the monovalent saccharide-based inhibitors, a large portion of the multivalent inhibitors have been specifically designed to target galectin-3, which perhaps explains why oligodendrimeric inhibitors with carbohydrate head groups have not been very successful against galectin-1.[78,147,148,150–152]

Earlier studies into lactosylated starburst poly (amidoamine) glycodendrimers and wedge-like glycodendrimers with 3,5-di-(2-aminoethoxy)benzoic acid branching units (**20**) dramatically improved the per-lactose affinity of the inhibitors toward galectin-1, but the absolute affinity of these inhibitors were similar to those toward galectin-3.[144,145] In addition, the affinity of these inhibitors toward any of the studied galectins were consistently lower than toward the mistletoe lectin (VAA) by several orders of magnitude.[144,145] The rigidification of the spacer units using propargylamines (**21**) induced 10-fold per-lactose affinity improvements toward galectin-1, but the inhibitors had become highly selective toward galectin-3 with 1000-fold per-lactose affinity enhancements (bearing in mind the limitations of solid-phase assay data).[146] Polyfunctional unnatural amino acids such as phenyl-bis and -tris-alanine based scaffolds to link carbohydrate moieties were used to explore glycoside cluster effect. The compound containing phenylethyl carbamate not only showed better affinity toward galectin-1 in monovalent (**22**) (24 μ M) and divalent forms (**23**) (3.2 μ M) but also selectivity when compared with other galectins.[153] However, pronounced cluster effect was observed for the divalent lactoside.[153]

Synthetic divalent lactulose amines with 8 and 10 carbon chain linkers were shown to inhibit galectin-mediated homotypic cell aggregation and endothelial cell morphogenesis *in vitro*, but the compounds were shown to be non-selective.[134] A similar divalent lactulosyl amine with a 6-carbon linker

(**15**) was filed for patent as an inhibitor of galectin-1 and galectin-4 to be used as a treatment option for tumors presenting these galectins.[80] The divalent lactulosyl amine was able to inhibit galectin-1 binding to its cell–surface interaction partners *in vitro* and decreased tumor growth *in vivo*. The divalent form was more effective than the monovalent lactulosyl-L-leucine.[80]

Complex polysaccharides extracted and modified from natural sources such as guar, apple and citrus are relatively successful in inhibiting galectins. The evidence for these being galectin-antagonists is generally based on data from cell cultures or animal experiments where indirect effects cannot be ruled out, and therefore though there is indication that galectins could be a target of these plant polysaccharides, this is not conclusive. Importantly, Galectin Therapeutics Inc. has developed two such polysaccharides that have shown significant success in clinical trials, namely GM-CT-01 (Davanat) and GR-MD-02 (extensively reviewed [154–156]). GM-CT-01 was the first discovered and was patented as a cancer therapeutic with the ability to decrease the negative side effects of co-therapeutics such as 5-fluorouracil (5-FU) and adriamycin.[85,86] Phase I clinical trials of Davanat in concert with 5-FU in patients with advanced solid tumors showed no adverse effects and led to improvement in reducing side effects associated with 5-FU. Early results from the phase II clinical trials in patients with colorectal cancer (Clinical trials <http://clinicaltrials.gov> (NCT00110721, NCT00388700)) and gall bladder and bile duct cancer (NCT00386516) were encouraging, with reduction of side effects but the studies were terminated prior to completion, citing financing and restructuring, which may have to do with the discovery of the more potent GR-MD-02. Unlike GM-CT-01, which is a galactomannan polysaccharide (β -1,4-linked mannan chain backbone decorated with α -1,6-linked galactose at regular intervals), GR-MD-02 is a complex polysaccharide with rhamnogalacturonate backbone and branches terminating with galactose and arabinose at varying intervals. As with the GM-CT-01 composition, Galectin Therapeutics Inc. has patented not just the GR-MD-02 extract that was found to be most efficacious, but also all the similar extraction/modification variants. GR-MD-02 is known as a galectin-3 inhibitor and GM-CT-01 is known as a galectin-1 inhibitor, but both have the ability to bind to galectin-1 and to galectin-3 as well as other galectins and potentially some other lectins. GR-MD-02 and similar extracts were patented as anti-fibrotics,[87–90] as immune modulatory treatment against cancer,[91] as treatment against diabetic nephropathy [90] and as therapeutics against inflammatory, autoimmune, cardiovascular and neurodegenerative diseases displaying elevated induced nitric oxide synthase levels.[92] Phase I clinical trials against non-alcoholic steatohepatitis (NASH) and advanced liver fibrosis were successful, demonstrating efficacy and tolerance at 2–4 mg/kg dose (NCT01899859). Phase Ib clinical trials are currently recruiting patients with metastatic melanoma to test the safety and efficacy of GR-MD-02 in conjunction with 3 mg/kg dose of Ipilimumab (Yervoy®) taking advantage of the ability of GR-MD-02 to potentiate immune modulatory treatments against cancer [91]; NCT02117362. Meanwhile, a phase II trial against NASH with cirrhosis (NASH-CX) is also recruiting patients to evaluate the intervention capacity of GR-MD-02 in a dose of 2

or 8 mg/kg doses every fortnight over 52 weeks (NCT02462967). A smaller study with a cohort of 10 patients presenting NASH with advanced fibrosis (NASH-FX) was also planned to start in 2015 (NCT02421094). In addition, a small phase IIa study for the treatment of psoriasis over 12-week treatment is also planned (NCT02407041) as a phase I patient had remission of psoriasis.

Compositionally GR-MD-02 is very similar to another galectin inhibitor, known as modified citrus pectin (MCP) as well as its newer patented formulation GCS-100, currently being developed by La Jolla Pharmaceutical Company. Due to the different origins and the differential extraction methods utilized, minor compositional differences may be present between the extracts, but GCS-100 and GR-MD-02 are both 1,4-linked poly-galacturonic acid chains with 1,2-linked rhamnose interruptions with branches terminating in galactose and arabinose. Although there is no information regarding where on the galectin surface these polysaccharides interact, it might be speculated that the branch-ending galactoses could contribute much of the binding. As with GR-MD-02, GCS-100 is known as a galectin-3 inhibitor, but this product shows binding to galectin-1 and possibly other lectins. It is not clear what happens in a disease where galectin-3 and galectin-1 both function at levels of high importance but with antagonistic disease roles. GCS-100 was first developed and patented by Glycogenesys Inc. and has recently been taken over by La Jolla Pharmaceutical Company.[93,94] A number of applications have been submitted by Glycogenesys for the use of GCS-100 as a treatment for immune diseases,[95] as an anti-angiogenic,[96] as a treatment for neurodegenerative diseases,[97] and as a treatment for hyperproliferative diseases.[98] GCS-100 was well tolerated in phase I clinical trials on patients with chronic kidney disease (NCT01717248), where the major side effect was a rash at the injection site, and phase II trials are underway (NCT02155673, NCT01843790, NCT02312050). A small phase II clinical trials on elderly patients with relapsed chronic lymphocytic leukemia was performed, citing 25% partial response (NCT00514696).

Overall, the complex polysaccharides being developed by Galectin Therapeutics and La Jolla Pharmaceuticals are fairly successful. As these naturally derived polysaccharides are well tolerated and safe in humans, there are a number of diseases that could be targeted, especially considering the capacity of galectin-1 and galectin-3 to affect so many human diseases. It is perhaps rash to attribute the clinical and pre-clinical efficacy of these polysaccharides to just the inhibition of galectin-3, but so far the disease targets have been ones that display negative effects of galectin-3, such as fibrosis and cancer. As the binding capacity of these formulations toward galectin-1 are similar to galectin-3, it is likely feasible to utilize them against diseases where galectin-1 plays a major role; and the clinical safety assessments that are in progress will make the road simpler for these further developments.

3.3. Non-carbohydrate-based inhibitors

Currently, the only type of non-carbohydrate inhibitors designed to target galectin-1 are peptides and peptidomimetics, and thus far there have not been any reports of any

other non-saccharide small-molecule inhibitors. The most successful peptide-based galectin-1 inhibitor is anginex (β pep-25), a peptide of 33 amino acids that displays anti-angiogenic and anti-tumor effects.[157] The peptide, its design and development as well as its *in vitro* and *in vivo* effects have been exhaustively reviewed.[155,158]

Anginex is a synthetic peptide (ANIKLSVQMKLFKRHLKW KIIVKLNDGRELSD), designed using basic folding principles utilizing short sequences from the β -strand regions of anti-angiogenic proteins PF4, IL-8 (note that IL-8 also can behave as a pro-angiogenic agent [159]) and BPI.[157] Depending on its concentration, the peptide switches between a flexible random-coil and a rigid β -sheet structure,[160] but the active form has been elucidated to be the β -sheet.[160,161] Mechanism of action of anginex starts in the plasma, where it hitchhikes on fibronectin to arrive at angiogenic sites,[162] where it can interact with activated endothelial cells (ECs). The cellular target of anginex has been shown to be galectin-1, which can induce cell-cell attachment in activated ECs.[163] The blocking of galectin-1 on EC surface results in anoikis-induced apoptosis and hence a decrease in vascularization of the tumor, which leads to slower growth and decreased oxygenation.[157,163–165] In addition to being a potent inhibitor of tumor growth on its own, anginex has been shown to work synergistically with a number of other therapeutics, including irifolven,[166,167] carboplatin,[168] angiostatin, [168] and radiation.[169,170] The use of anginex along with scheduled radiation therapy was filed for patenting by Mayo et al. citing experimental results that show anginex- sensitizing ECs to radiation therapy and the combination therapy resulting in synergistic effects of decreased angiogenesis and tumor growth *in vitro* and *in vivo*.[99]

Due to the highly potent effects of anginex, Peptx Inc. applied for a patent that covered the recombinant expression methodology of anginex in yeast as well as its use as an antiangiogenic therapeutic targeting galectin-1.[100] In addition to the recombinant expression methodology, the patent application also covered the specific oligonucleotide sequences utilized for the expression. The recombinant protein was less potent compared to the synthetic peptide in inhibiting EC growth, microvessel formation and tumor growth, but the method may offer a cheaper alternative to peptide synthesis. The sequence utilized, the recombinant expression protocol and the potency of the recombinant peptide have since been published.[171,172]

Based on the prior knowledge of the key amino acid residues required for anginex function and knowing its bioactive conformation (anti-parallel β -sheet), a partial peptide mimetic of anginex, 6DBF7, was designed.[160,173] 6DBF7, has six amino acid residues at the *N*-terminus and seven at the C-terminus linked by a dibenzofuran (DBF) moiety (SVQMKL-[DBF]-IIVKLND), and is a better angiostatic and anti-tumor agent than the parent anginex *in vivo*.[173] Screening alanine and unnatural amino acid substitutions at different positions of the 6DBF7 compound to increase its aqueous solubility resulted in a number of derivatives that displayed potent anti-proliferative effects on ECs.[174] DB16 (SVQMKL-[DBF]-AIVKLNA) and DB21 (SVQNvaKL-[DBF]-IIVKLNA) displayed both high solubility as well as potent anti-proliferative effects.

¹⁵N-¹H-HSQC experiment showed that DB16 interacted with galectin-1 allosterically with the residue segments 10–14, 33–42, 70–80, 90–95, and 106–116, toward one edge of the monomer.[174] Interestingly, the mimetics display lower anti-proliferative effects *in vitro*, but the effects were enhanced *in vivo*, which may be due to higher bioavailability of the mimics.[173,174] 6DBF7 and others fitting the same scaffold were patented by ‘Regents of the University of Minnesota’ based on their anti-angiogenic and anti-tumor effects [101] as reported.[173]

Based on the success of the partial peptidomimetics, a series of topomimetics were synthesized, based on a calix[4] arene backbone, which allowed for hydrophobic and hydrophilic faces as found in the anginex β -sheet.[175] Compounds PTX008 (also known as 0118/OTX008) and PTX009 (previously referred to as compound 1097 [175]) were identified as potent inhibitors of angiogenesis in cell proliferation and migration assays and in the mouse models of ovarian carcinoma and melanoma.[163] This set of calixarene-based topomimetics were patented by Regents of the University of Minnesota as anti-bacterial, anti-angiogenic and anti-tumor agents displaying the mentioned activities *in vitro* and *in vivo*. [102] The combination of PTX008 with irrofluen or sunitinib showed synergistic anti-tumor effects *in vitro* and *in vivo*. [166,176] Galectin-1 was identified as the molecular target of PTX008, though binding occurs at an allosteric site with only attenuation of the lactose binding.[166] The allosteric binding site was reportedly formed from residues 6–10, 14–17, and 89–92.[166] Based on galectin-1 being the target of PTX008, 2 patents were filed, both citing the ability of PTX008 to function as anti-galectin-1, anti-angiogenic and anti-tumor agent. [104,174] PTX008 has been found to be safe in toxicological studies following which it has entered Phase I clinical trials for advanced solid tumors (NCT01724320). A radiolabeled analogue of the same compound has been patented as an imaging tracer for tumor diagnosis and for assessing patients’ suitability for PTX008 treatment.[105]

Further modifications of parent PTX008 resulted into PTX013 which effectively inhibited cell proliferation in drug resistant tumors (50-fold better than PTX008), though the new compound resulted in dramatic loss of weight in the mice, displaying toxicity, which has not been reported previously for anginex-based therapies.[177] Based on the toxicity results and the differing effects on cell cycle regulation, the authors postulated that the target of PTX013 is not galectin-1. Nevertheless, PTX013 and similar compounds reported within [177] have been filed for patenting as better anti-angiogenic and anti-tumor agents by Regents of the University of Minnesota [106] displaying the same experimental results found in the literature.[177]

Overall, anginex, and its mimetics are quite successful anti-angiogenic agents, with the first of its kind PTX-008 being approved for phase I clinical trials (NCT01724320). However, the major challenge behind further development of these inhibitors is that their mechanism of action is more elusive than they appear. Anginex itself has been reported to bind several other galectins (galectin-1, -2, -7, -8 N, -9 N), indicating that the inhibitors are not selective.[178] The same report also suggests that anginex does not inhibit, but potentiates the

affinity of galectin-1 toward glycoproteins,[178] despite a number of other reports classifying the compounds as being inhibitors of galectins.[174,179] Furthermore, anginex has been shown to interact with lipid membranes via charge-charge interactions to disrupt cellular membrane, possibly indicating a different mechanism of action other than galectin-1 interaction.[180] Finally, as improvements and changes are made to the calixarene-based topomimetics, the molecules seem to be diverging from the mechanism of action of anginex, as in the case of PTX013,[177] which requires further evaluation of the mimetics.

In a different application of peptidic molecules binding galectin-1, nanobodies (single domain antibodies) developed by Ablynx Inc. were patented to target multiscavenger receptors including galectin-1.[107] The binding affinities observed for binding to the multiscavenger receptor were in nanomolar range, although the binding specificity toward galectin-1 was not addressed. In another patent application, EBV-associated post-transplantation disorder (PTLD) presenting with overexpression of galectin-1 was proposed to be treated by antibodies, fragments thereof, peptides targeting galectin-1. [108,109] This patent claims that targeting galectin-1 not only overcomes cancer immunosuppression but also can circumvent resistance to anti-vasculo-endothelial growth factor (VEGF) treatment, which is an important clinical problem in clinical therapies.[28,181,182] Furthermore, anti-galectin-1 or anti-integrin- β 1 monoclonal antibodies were patented as anti-proliferative targeting neuronal stem cells to control the overgrowth of the stem cells after implantation in patients with central nervous system injury.[110] RNA interference-based approach was used to knock-down galectin-1 expression for the treatment of glioma, pancreatic cancer, head and neck cancer, melanoma, non-small-cell lung cancer and non-Hodgkin’s lymphoma.[111] Micropump or encapsulation were the methods used to deliver the RNAi into animal cells with special attention to thermal stability and resistance to nuclease digestion. Finally, gigaxonin protein linked to a cell penetrating peptide was patented for its ability to normalize the overexpression of galectin-1 and decrease the phosphorylation of vimentin in giant axonal neuropathy cells,[112] but the protein may not necessarily be a specific down-regulator or inhibitor of galectin-1.

4. Galectin-1 as a therapeutic molecule

The presence of galectin-1 in the thymus, the lymph nodes, activated macrophages and T cells along with its capability to induce apoptosis and inhibit T-cell adhesion to extracellular glycoproteins imply its significant role in immune regulation. [37,39,183,184] The first evidence for the potential of galectin-1 protein therapeutic applications came from testing for amelioration of autoimmune myasthenia gravis in rabbits, where galectin-1 acted in the immune system by binding to rabbit lymphocytes and stimulated mitogenesis.[185] Recombinant human galectin-1 mediated the suppression of clinical symptoms of autoimmune encephalomyelitis in rats by blocking the sensitization of myelin basic protein-specific T-cells and inducing suppressor cells.[186] Furthermore, a dramatic decrease in galectin-1 expression was seen in the synovial

tissue of juvenile idiopathic arthritis patients,[187] which correlated with increased anti-galectin-1 antibodies observed in rheumatoid arthritis (RA) patients.[188] Intravenous galectin-1 pre-treatment in concanavalin A-induced hepatitis model in mice resulted in protection from the disease via the apoptosis of activated T-cells and inhibition of the release of pro-inflammatory cytokines.[189] Improvement in the histopathological signs of inflammation were observed in 2,4,6-trinitrobenzene sulfonic acid (TNBS)-induced colitis model of mice, particularly in already established lesions, furthering the therapeutic paradigm of galectin-1 to inflammatory bowel disease.[190] TNBS administration resulted in decreased colonic expression of galectin-1, as seen in the case of RA patients.[190] Furthermore, hapten-activated lamina propria T-cell apoptosis was associated with increased caspase-8 and -9 activities.[190]

A number of patents have already been lodged for the therapeutic use of galectin-1. The ability of galectin-1 to modulate proliferation and secretion of goblet conjunctival cells was patented by Schepens Eye Research Institute.[113] The method of treating Dry Eyes Syndrome using different formulation of galectins was patented by the University of Tufts.[114] Galectin-1 has also been patented for prevention and treatment of nephritis.[115] As demonstrated, the functional features of galectin-1 make it a promising therapeutic molecule for the treatment of autoimmune, inflammatory and other diseases.[191]

Ongoing reports of altered expression profiles of galectin-1 in cancerous tissues and tumor progression have drawn great attention to the diagnostic value of galectin-1. Of importance is that elevated expression level of galectin-1 in serum is frequently correlated with its increased tissue expression, such as in lung cancer,[192] colon cancer, [193] and T-cell lymphoma,[194] which suggests the potential role of serum galectin-1 serving as a diagnostic biomarker. It has to be noted that decreased serum levels of galectin-1 are also observed in patients with cancer, of particular in the reproductive system, which is contradictory to the common phenomenon of increased expressions of galectin-1 in cancerous reproductive tissues.[182,195,196] This discrepancy may relate to differences in patient population or methodology applied in studies. Thus further studies using larger patient cohorts may help to further explore the diagnostic value of galectin-1 expression.

From the pharmaceutical perspective, conventional formulation methods for proteins suffer drawbacks such as poor biopharmaceutical properties (high molecular weight, proteolytic degradation, rapid clearance, limited cellular uptake), structure fragility (physical/chemical inactivation during formulation and/or storage), and intrinsic immunogenicity.[197] Nevertheless, targeted delivery and controlled release approaches achieved by using suitable nanocarriers, such as nanoparticles and liposomes can be used for transporting active molecules to the desired site and the therapeutic efficacy can be enhanced by releasing active molecule only upon stimulation (physical/chemical). In the context of galectin-1, the differing redox states and the existence of a monomer-dimer equilibrium, both with reported differential activity in normal and pathological processes, presents additional challenges in developing galectin-1-based therapeutics.

4.1. Galectin-1 and its modified forms

At concentrations under 10 μ M, galectin-1 exists as a mixture of monomer and dimer, and both of the species exhibit similar binding affinity toward a range of monovalent and divalent saccharides and glycoproteins, primarily displaying monovalent interactions consistent with a monomer.[58] Dimer interface mutants including single mutants Cys2Ser and Val5Asp as well as a multisite mutant (N-Gal-1) with Cys2Ser, Leu4Gln, Val5Asp, and Ala6Ser mutations were generated.[60] The mutants displayed higher tendency to remain as monomers in solution and the K_d of dimerization increased to 250 μ M for the N-Gal-1 mutant. The wild-type and mutant galectin-1 were patented for their use in modulating inflammatory responses, where the dimeric wild-type protein was used to kill activated neutrophils leading to reversal of inflammation while the monomeric mutant galectin-1 were used to block the apoptosis of neutrophils acting as a pro-inflammatory agent.[116]

On the other hand, low *in vivo* efficacy of the monomeric form of galectin-1 due to the existence of monomer-dimer equilibrium, prompted the development of covalently linked galectin-1 dimer.[198] Covalent linkage by a two-glycine linker sequence resulted in the two galectin-1 CRDs maintaining proper contact and the overall orientation of CRDs with respect to each other remaining similar to the native dimer. The engineered product was more potent in a hemagglutination assay and also induced apoptosis in thymocytes and activated T-cells more potently compared to the native non-covalent dimer.[198] Varied linker sequences were studied, including a 14-amino acid long random coil from the linker of galectin-9,[199] a 33-amino acid long flexible linker from galectin-8 [200] and a 34-amino acid long rigid helix from bacterial ribosomal L9 protein,[201] each leading to increased hemagglutination activity as well as increased T-cell apoptosis induction.

To utilize galectin-1 as a therapeutic agent, varying functionality of the oxidized and reduced forms of galectin-1 demands the development of modified galectin-1 that remains stable as one form or the other under different biological environments. One approach employed while dealing with sensitivity of galectin-1 to oxidized environments was the generation of a cysteine-less galectin-1 mutant.[66] This mutant was found to be more stable than the wild-type protein and retained its hemagglutination activity after storage without reducing agents.[66] This is additionally important as the use of chemicals to retain the structural and functional integrity of recombinant galectin-1 could have adverse effects on cell assays as the use of chemicals may sensitize cells and cause interference in the overall outcome of the assay. The cysteine-less mutant showed no significant structural difference compared to the wild-type galectin-1 and was also effective at inhibiting cell growth.[66] In another approach, Dimitroff et al. engineered a fusion product of mouse galectin-1 with the Fc region of human IgG1, which was functionally as active as the native galectin-1 and more stable due to the facilitated dimerization.[202] The fusion protein showed a binding preference for LacNAc bearing glycans and induced apoptosis in inflammatory leukocytes from

RA patients.[202] The engineered product was patented for the treatment of immune dysfunctions.[118]

Unlike the reduced form of galectin-1, its oxidized form is not well understood, though increasing evidence suggests that the oxidized form possesses therapeutic potential.[62,68] As a consensus on the biologically relevant structure of the oxidized form has not been reached, it may be difficult to formulate a perpetually oxidized galectin-1. An oxidized form of galectin-1 with at least one disulfide bond between Cys16–Cys88 was patented for the treatment of nerve injury, nerve degeneration and hypofunction after nerve grafting.[119] The oxidized galectin-1 in the invention is to be covalently modified with soluble polymers, such as polyethylene glycol or granulated along with these polymers to enhance the solubility, stability, release rate, and clearance.

In summary, the protein-specific challenges related to galectin-1 therapeutic formulations are the monomer–dimer equilibrium and the oxidized and reduced forms. Progress made in the field demonstrates that it is feasible to control the monomeric and dimeric forms through mutations and covalent linkages. The cysteine-less mutant and the Fc-galectin fusion proteins demonstrate that the stabilization of the reduced form is possible, though further studies *in vivo* and especially in solid tumor models may be necessary. The oxidized form of galectin-1 acts as a completely different protein to galectins, displaying no lectin activity, while performing opposing roles to reduced galectin-1, such as the induction of cell proliferation, which warrants further exploration.

4.2. Galectin-1 formulation and delivery

In addition to the protein-specific challenges of redox states and monomer–dimer equilibrium, the delivery of intact galectin-1 to the target pathological site is a major hurdle. One approach may be the use of gene therapy. Galectin-1 via gene and protein therapy was applied in collagen-induced arthritis (CIA) murine models, to assess its *in vivo* effects in rheumatoid arthritis.[189] A single dose of engineered fibroblast secreting galectin-1, at disease onset was enough to cease the disease progression. Similar effects were observed with daily administration of the recombinant galectin-1.[41] In another gene therapy approach, galectin-1 was delivered at the site of action using a lentivirus vector. The viral vector expressing galectin-1 and small hairpin RNA galectin-3 were administered intra-articularly just before disease onset in the ankle joints of murine CIA models to successfully ameliorate the disease condition.[203] However, limiting factors associated with gene therapy such as fine-tuning of transgene expression to obtain desired therapeutic effects, synchronization with host transcription, and the design of a switch-off mechanism to terminate expression upon therapy completion need consideration.[204]

In another therapeutic application of galectin-1, significant reduction in morbidity and mortality of graft versus host disease was observed in murine models of allogeneic transplants.[205] In contrast to the gene therapy approach, mRNA approach was used for galectin-1 delivery with advantages such as no insertional mutagenesis in host genome and no use of viral vectors that often cause immunogenic response,

which clears the transfected cells. The method and composition containing galectin-1 nucleotide sequence in combination with other immunomodulators has been patented for the prevention and treatment of undesired immune responses, such as transplant rejection, autoimmune disease and allergic reactions.[206] The composition contains cells that transiently express or are transfected with mRNA encoding galectin-1 along with other immunomodulators that preferably get accumulated in lymphoid tissues in the vicinity of the undesired immune response.

In a parallel approach to gene and mRNA therapy, galectin-1 was formulated as nanoparticles. A patent application covering the formulation of galectin-1 into particles, nanoparticles, nanocapsules, nanocores, or nanospheres that are to be delivered into cells was filed.[120] The formulations also rely on galectin coating to enhance cellular targeting and uptake. Taking the advantage of the anti-angiogenic property of gold nanoparticles, which innately has VEGF inhibitory properties, galectin-1 was conjugated to 13-nm gold nanoparticles through physical adsorption and the nanoparticle was patented as an anti-inflammatory reagent targeting arthritis.[121,207] The multivalent organization of galectin-1 on the nanoparticles resulted in increased binding to cell–surface and enhanced apoptosis of Jurkat cells as a result of facilitated CD45 clustering.[208] Efficiency of the galectin-1-conjugated gold nanoparticles in alleviating rheumatoid arthritis in CIA model was higher compared to gold nanoparticles or galectin-1 alone. This enhanced therapeutic efficiency furthers the application of nanoparticles from increasing the payload at the desired site to regulation of receptor distribution and downstream cell signaling.

Nanocarriers, such as liposomes, polyplexes, polymersomes, dendrimers, etc. appear to be attractive alternatives to the conventional gene and protein delivery methods.[209–211] Protein modification particularly with polyethylene glycol (PEG) results in an improvement of the overall physiochemical drawbacks of polypeptide formulations, prolonging serum half-life, which is one of the limiting factors with unmodified proteins formulations, as well as providing protection from the host immune system.[212,213] Such a formulation could be used in galectin-1 therapeutics as has already been done for galectin-2 [214] and galectin-9.[117]

5. Conclusion

Galectin-1, with its ubiquitous expression throughout the body and involvement in important cell-to-cell communications mediating adhesion, cell growth, migration and regulation processes, especially in pathological conditions, has emerged as a potential therapeutic target for various serious diseases. Anti-migratory, anti-invasive and anti-angiogenic activities of various galectin-1 inhibitors prove their potential as anti-cancer agents. Galectin-1 inhibitors reported to date mainly fall under three categories: monovalent- and multivalent-carbohydrate-based inhibitors and peptidomimetics. Use of mono- and disaccharide scaffolds have resulted in high-affinity inhibitors of galectin-1, but the overall progress is slow compared to the development of galectin-3 inhibitors. Multivalent inhibitors exploit the inherent nature of galectins

as cross-linkers of glycan, utilizing synthetic glycodendrimers and natural polysaccharide extracts. A number of patents have been filed regarding polysaccharide extracts and early clinical trials show limited adverse effects. Both mono- and multivalent inhibitors of galectin-1, and galectins in general, lack selectivity due to the sequence conservation within galectin CRDs, though monovalent saccharide-based inhibitors show improved selectivity.

Galectin-1 mainly acts as negative regulator of immunity by selectively killing activated T-cells and hence can serve as a potential therapeutic molecule for autoimmune and inflammatory diseases. Although intact protein delivery still suffers from pharmacokinetic drawbacks, the use of nanoparticles, gene therapy, and the use of stabilizing conjugates, such as PEGs may help overcome the issues. Protein-specific challenges unique to galectin-1 are the monomer-dimer equilibrium and its redox states with differential functions. Recombinant galectin-1 with cysteine-to-serine mutants as well as those with mutations at the dimer interface or covalent dimer linkers improve the potential of galectin-1 protein as a therapeutic.

6. Expert opinion

Galectin-1 is an important target for inhibition, though perhaps due to its dual capacity in the progression and amelioration of certain diseases. Inhibition of galectin-1 has been neglected compared to galectin-3, the latter having been established more evidently as a disease-promoting protein. Lectins in general are not the ideal inhibition targets having shallow grooves dominated by electrostatic interactions within their binding sites. Galectins present an additional difficulty of their carbohydrate-binding sites being highly conserved with respect to amino acid sequence, which limits the potential of selective inhibitors. Despite these difficulties, there have been major improvements in the affinity and selectivity of monovalent inhibitors based on modified saccharides. A number of TDG-based nanomolar affinity inhibitors have been reported for galectin-1. As previous inhibitor design efforts favored galectin-3 selectivity and in a number of cases intentionally tried to decrease affinity toward galectin-1, it seems more than feasible to develop galectin-1-specific inhibitors, utilizing a number of galactose and lactose modifications that have already been reported, but in new combinations. Due to the general lack of synthetic inhibitors specific to galectin-1, applicable patents only contain general inhibitors such as lactose conjugated to functional groups and lactulosyl amines, both of which could target all galectins, perhaps even other lectins.

Multivalent inhibitors of galectins, especially glycodendrimers possess perhaps the most potential, but its realization is difficult due to a number of factors. Glycodendrimers will have inherently high affinity due to multisite interactions, and since galectins tend to form clusters upon interaction with multivalent ligands and affinity improves synergistically, a low dose of multivalent inhibitors may prove to be highly potent. In addition, use of rigid spacers can improve selectivity due to the homo-dimerization leading to a fixed spacing of carbohydrate-binding sites in proto-type galectins. Furthermore, the

improvements that are being made in the synthetic modification of monovalent inhibitors can be incorporated into glycodendrimers. On the other hand, glycodendrimer inhibitor design has been significantly lagging in galectin inhibition, due to the complexities involved. To realize the full potential of oligodendrimers and to induce selectivity and potency, the glycan profile of the disease state should be known and mimicked by the inhibitors, as can be seen from the current differing results from competition assays when different matrices are used. In addition, the natural glycan recognition pattern of proto and tandem-repeat type galectins needs to be better known in order to select for galectin-1 over others. Furthermore, the commonly used assays to determine ligand efficiency are not infallible, with results not always translating from one assay to the other, and further there is a need for a comprehensive specificity profile reporting for each ligand.

Interestingly, the most advanced inhibition strategy in the field of galectin-1 inhibition is the use of modified polysaccharides, such as GM-CT-01 and GCS-100, which are currently being tested in phase I and II clinical studies. An important benefit regarding these inhibitors is their limited toxicity. It is important to recognize that these polysaccharides might not be specific inhibitors of galectin-3 or galectin-1 since any galectin should be able to bind these extracts, perhaps even other lectins. Despite their current success, with respect to inhibitor development the polysaccharides possess limited potential without a significant breakthrough, as modification of the saccharide chains is synthetically difficult, their selectivity is minimal, they do not cross the cell membrane and are thus unable to inhibit the intracellular functions of galectin-1.

The peptide anginex and its peptidomimetics are certainly promising however they still face challenges as galectin-1 inhibitors, particularly with respect to cost and bioavailability. Unlike the other inhibitory approaches, peptide-based inhibitor development of galectin-1 is ahead of that of galectin-3 with significant effort already put into partial and full mimics. The *in vivo* studies indicate a strong potential for efficacy with limited side effects. The main challenge that remains is the selectivity of these compounds as available reports lack this crucial information. Although the general binding site of the peptide as well as the mimics have been mapped by the use of NMR techniques, detailed structural information of galectin-1-anginex complex may prove useful in further development of the inhibitor.

Comparison of the different kinds of inhibitors highlights considerations and challenges in designing galectin antagonists. Carbohydrate-based monovalent inhibitor design takes advantage of the inherent affinity of galectins for β -galactosides. The key features of these antagonists lie in their relative ease of synthesis, quantification of one-to-one binding with the lectin and availability of structural information of galectins with bound ligands. Considering the key interaction needed for galactose recognition, the C1 and C3 positions have been used to obtain affinity but in general, the specificity remains questionable. The structure- and fragment- based approaches employed for galectin-3 antagonist design also have relevance to development of potent and selective galectin-1 antagonists. Differences between galectins assist in understanding design of selective inhibitors. In galectin-1, for example there

is an advantageous interaction, via its binding site His52, to the sulfur that links two galactose monomers in TDG, enhancing the potential usefulness of this as a scaffold for galectin-1 antagonists.

Other carbohydrate-based antagonists exploit the concept of multivalency to achieve stronger binding. However, synergism arising from multivalent binding partners involved in glycocluster formation cannot be ruled out. The discrepancies in binding affinities between various methods further limits clear appreciation of the biophysical phenomenon while no structural information for the glycocluster hinders the rational design approach. Lactosylated glycodendrimers and other divalent lactulose amine derivatives show an enhanced affinity for galectin-1 but remain non-selective. Despite being actively pursued, plant extracted polysaccharides and derivatives lack a direct mechanism of action for inhibiting galectins.

Regarding the non-carbohydrate-based inhibitors, peptide and peptidemimetics are promising molecules under development. The recombinant production of anginex (peptide inhibitor) offers a cheaper alternative to peptide synthesis considering the discrepancy in binding between recombinantly expressed and synthesized peptide. All the non-carbohydrate-based inhibitors lack precise atomic detail regarding their binding to galectin-1, leaving only the analogue-based approach to further the inhibitor design. Nevertheless, benefits obtained from combination of such inhibitors with chemotherapy agents appears promising, while the RNAi or antibodies-based targeting of galectin-1 is in a nascent stage and requires further exploration.

Use of galectin-1 as a therapeutic option, perhaps targeting inflammatory conditions, is becoming rapidly feasible, with solutions being reported to the redox state and monomer-dimer equilibrium challenges. The development of nanoparticle-, liposome- and PEG-conjugated variants of galectin-1 are already under way. However, caution should be observed when wielding a double-edged sword like galectin-1 as the protein displays concentration, environment, and disease-state-dependent functions. It is advisable to use galectin-1 as a therapeutic in a system, where the function of galectin-1 is well defined.

Declaration of interest

H. Blanchard gratefully acknowledges the financial support from the Cancer Council Queensland. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

References

Papers of special note have been highlighted as either of interest (*) or of considerable interest (**) to readers.

1. Barondes SH, Cooper DN, Gitt MA, et al. Galectins. Structure and function of a large family of animal lectins. *J Biol Chem*. 1994;269:20807–20810.
2. Leffler H, Carlsson S, Hedlund M, et al. Introduction to galectins. *Glycoconj J*. 2004;19:433–440.
3. Sacchettini JC, Baum LG, Brewer CF. Multivalent protein-carbohydrate interactions. A new paradigm for supermolecular assembly and signal transduction. *Biochemistry*. 2001;40:3009–3015.

4. Fred Brewer C. Binding and cross-linking properties of galectins. *Biochim Biophys Acta*. 2002;1572:255–262.
5. Hirabayashi J, Hashidate T, Arata Y, et al. Oligosaccharide specificity of galectins: a search by frontal affinity chromatography. *Biochim Biophys Acta*. 2002;1572:232–254.
6. Camby I, Le Mercier M, Lefranc F, et al. Galectin-1: a small protein with major functions. *Glycobiology*. 2006;16:137R–57R.
7. Liu FT, Patterson RJ, Wang JL. Intracellular functions of galectins. *Biochim Biophys Acta*. 2002;1572:263–273.
8. Patterson RJ, Wang W, Wang JL. Understanding the biochemical activities of galectin-1 and galectin-3 in the nucleus. *Glycoconj J*. 2004;19:499–506.
9. He J, Baum LG. Galectin interactions with extracellular matrix and effects on cellular function. *Methods Enzymol*. 2006;417:247–256.
10. Puche AC, Poirier F, Hair M, et al. Role of galectin-1 in the developing mouse olfactory system. *Dev Biol*. 1996;179:274–287.
11. Wada M, Ono S, Kadoya T, et al. Decreased galectin-1 immunoreactivity of the skin in amyotrophic lateral sclerosis. *J Neurol Sci*. 2003;208:67–70.
12. Wang J, Xia J, Zhang F, et al. Galectin-1-secreting neural stem cells elicit long-term neuroprotection against ischemic brain injury. *Sci Rep*. 2015;5:9621.
13. Hughes RC. Secretion of the galectin family of mammalian carbohydrate-binding proteins. *Biochim Biophys Acta*. 1999;1473:172–185.
14. Cooper DN, Barondes SH. Evidence for export of a muscle lectin from cytosol to extracellular matrix and for a novel secretory mechanism. *J Cell Biol*. 1990;110:1681–1691.
15. Cho M, Cummings RD. Galectin-1, a beta-galactoside-binding lectin in Chinese hamster ovary cells. I. Physical and chemical characterization. *J Biol Chem*. 1995;270:5198–5206.
16. Vasta GR. Galectins as pattern recognition receptors: structure, function, and evolution. *Adv Exp Med Biol*. 2012;946:21–36.
17. Ito K, Stannard K, Gabutero E, et al. Galectin-1 as a potent target for cancer therapy: role in the tumor microenvironment. *Cancer Metastasis Rev*. 2012;31:763–778.
- The first example that TDG binds and blocks galectin-1, promotes survival in a breast cancer mouse model, and reduces tumor growth.
18. Demydenko D, Berest I. Expression of galectin-1 in malignant tumors. *Exp Oncol*. 2009;31:74–79.
19. Braeuer RR, Shoshan E, Kamiya T, et al. The sweet and bitter sides of galectins in melanoma progression. *Pigment Cell Melanoma Res*. 2012;25:592–601.
20. Barrow H, Rhodes JM, Yu LG. The role of galectins in colorectal cancer progression. *Int J Cancer*. 2011;129:1–8.
21. Verschuere T, De Vleeschouwer S, Lefranc F, et al. Galectin-1 and immunotherapy for brain cancer. *Expert Rev Neurother*. 2011;11:533–543.
22. Smetana K Jr, Szabo P, Gal P, et al. Emerging role of tissue lectins as microenvironmental effectors in tumors and wounds. *Histol Histopathol*. 2015;30:293–309.
23. Berois N, Osinaga E. Glycobiology of neuroblastoma: impact on tumor behavior, prognosis, and therapeutic strategies. *Front Oncol*. 2014;4:114–127.
24. Bacigalupo ML, Manzi M, Rabinovich GA, et al. Hierarchical and selective roles of galectins in hepatocarcinogenesis, liver fibrosis and inflammation of hepatocellular carcinoma. *World J Gastroenterol*. 2013;19:8831–8849.
25. Thijssen VL, Heusschen R, Caers J, et al. Galectin expression in cancer diagnosis and prognosis: a systematic review. *Biochim Biophys Acta*. 2015;1855:235–247.
26. Thijssen VLJL, Postel R, Brandwijk RJ, et al. Galectin-1 is essential in tumor angiogenesis and is a target for antiangiogenesis therapy. *Proc Natl Acad Sci*. 2006;103:15975–15980.
27. Thijssen VL, Griffioen AW. Galectin-1 and -9 in angiogenesis: a sweet couple. *Glycobiology*. 2014;24:915–920.
28. Croci DO, Cerliani JP, Pinto NA, et al. Regulatory role of glycans in the control of hypoxia-driven angiogenesis and sensitivity to anti-angiogenic treatment. *Glycobiology*. 2014;24:1283–1290.

29. Sakaguchi M, Okano H. Neural stem cells, adult neurogenesis, and galectin-1: from bench to bedside. *Dev Neurobiol.* **2012**;72:1059–1067.
30. Rabinovich GA. Galectin-1 as a potential cancer target. *Br J Cancer.* **2005**;92:1188–1192.
31. Astorgues-Xerri L, Riveiro ME, Tijeras-Raballand A, et al. Unraveling galectin-1 as a novel therapeutic target for cancer. *Cancer Treat Rev.* **2014**;40:307–319.
32. Strik HM, Kolodziej M, Oertel W, et al. Glycobiology in malignant gliomas: expression and functions of galectins and possible therapeutic options. *Curr Pharm Biotechnol.* **2012**;13:2299–2307.
33. Messaoudi K, Clavreul A, Lagarde F. Toward an effective strategy in glioblastoma treatment. Part I: resistance mechanisms and strategies to overcome resistance of glioblastoma to temozolomide. *Drug Discov Today.* **2015**;20:990–997.
34. Mercier S, St-Pierre C, Pelletier I, et al. Galectin-1 promotes HIV-1 infectivity in macrophages through stabilization of viral adsorption. *Virology.* **2008**;371:121–129.
35. St-Pierre C, Many H, Ouellet M, et al. Host-soluble galectin-1 promotes HIV-1 replication through a direct interaction with glycans of viral gp120 and host CD4. *J Virol.* **2011**;85:11742–11751.
36. St-Pierre C, Ouellet M, Giguere D, et al. Galectin-1-specific inhibitors as a new class of compounds to treat HIV-1 infection. *Antimicrob Agents Chemother.* **2012**;56:154–162.
37. Perillo NL, Pace KE, Seilhamer JJ, et al. Apoptosis of T cells mediated by galectin-1. *Nature.* **1995**;378:736–739.
38. Toscano MA, Bianco GA, Ilarregui JM, et al. Differential glycosylation of TH1, TH2 and TH17 effector cells selectively regulates susceptibility to cell death. *Nat Immunol.* **2007**;8:825–834.
39. Blaser C, Kauffmann M, Muller C, et al. β -galactoside-binding protein secreted by activated T cells inhibits antigen-induced proliferation of T cells. *Eur J Immunol.* **1998**;28:2311–2319.
40. Rabinovich GA, Modesti NM, Castagna LF, et al. Specific inhibition of lymphocyte proliferation and induction of apoptosis by CLL-1, a β -galactoside-binding lectin. *J Biochem.* **1997**;122:365–373.
41. Rabinovich GA, Daly G, Dreja H, et al. Recombinant galectin-1 and its genetic delivery suppress collagen-induced arthritis via T-cell apoptosis. *J Exp Med.* **1999**;190:385–398.
42. Santucci L, Fiorucci S, Cammilleri F, et al. Galectin-1 exerts immunomodulatory and protective effects on concanavalin-A-induced hepatitis in mice. *Hepatology.* **2000**;31:399–406.
43. Hsu DK, Yang RY, Pan Z, et al. Targeted disruption of the galectin-3 gene results in attenuated peritoneal inflammatory responses. *Am J Pathol.* **2000**;156:1073–1083.
44. Cortegano I, Del Pozo V, Cardaba B, et al. Galectin-3 down-regulates IL-5 gene expression on different cell types. *J Immunol.* **1998**;161:385–389.
45. Karlsson A, Follin P, Leffler H, et al. Galectin-3 activates the NADPH oxidase in exudated but not peripheral-blood neutrophils. *Blood.* **1998**;91:3430–3438.
46. Rabinovich GA, Baum LG, Tinari N, et al. Galectins and their ligands: amplifiers, silencers or tuners of the inflammatory response? *Trends Immunol.* **2002**;23:313–320.
47. Almkvist J, Karlsson A. Galectins as inflammatory mediators. *Glycoconj J.* **2004**;19:575–581.
48. Mizoguchi E, Mizoguchi A. Is the sugar always sweet in intestinal inflammation? *Immunol Res.* **2007**;37:47–60.
49. Liu FT. Galectins: novel anti-inflammatory drug targets. *Expert Opin Ther Targets.* **2002**;6:461–468.
50. Cedeno-Laurent F, Dimitroff CJ. Galectin-1 research in T cell immunity: past, present and future. *Clin Immunol.* **2012**;142:107–116.
51. Salatino M, Croci DO, Bianco GA, et al. Galectin-1 as a potential therapeutic target in autoimmune disorders and cancer. *Expert Opin Biol Ther.* **2008**;8:45–57.
52. Hokama A, Mizoguchi E, Mizoguchi A. Roles of galectins in inflammatory bowel disease. *World J Gastroenterol.* **2008**;14:5133–5137.
53. La M, Cao TV, Cerchiaro G, et al. A novel biological activity for galectin-1: inhibition of leukocyte-endothelial cell interactions in experimental inflammation. *Am J Pathol.* **2003**;163:1505–1515.
54. Cooper D, Norling LV, Perretti M. Novel insights into the inhibitory effects of Galectin-1 on neutrophil recruitment under flow. *J Leukoc Biol.* **2008** Jun;83:1459–1466.
55. Rabinovich GA, Sotomayor CE, Riera CM, et al. Evidence of a role for galectin-1 in acute inflammation. *Eur J Immunol.* **2000**;30:1331–1339.
56. Smetana K Jr, André S, Kaltner H, et al. Context-dependent multifunctionality of galectin-1: a challenge for defining the lectin as therapeutic target. *Expert Opin Ther Targets.* **2013**;17:379–392.
57. Lopez-Lucendo MF, Solis D, Andre S, et al. Growth-regulatory human galectin-1: crystallographic characterisation of the structural changes induced by single-site mutations and their impact on the thermodynamics of ligand binding. *J Mol Biol.* **2004**;343:957–970.
58. Salameh E, Leuzy A, Tejler J, et al. Monoclonal interactions of galectin-1. *Biochemistry.* **2010**;49:9518–9532.
59. Leppanen A, Stowell S, Blixt O, et al. Dimeric galectin-1 binds with high affinity to α 2,3-sialylated and non-sialylated terminal N-acetylglucosamine units on surface-bound extended glycans. *J Biol Chem.* **2005**;280:5549–5562.
60. Cho M, Cummings RD. Characterization of monomeric forms of galectin-1 generated by site-directed mutagenesis. *Biochemistry.* **1996**;35:13081–13088.
61. Kasai K-I, Hirabayashi J. Galectins: a family of animal lectins that decipher glycodes. *J Biochem.* **1996**;119:1–8.
62. Kadoya T, Horie H. Structural and functional studies of galectin-1: a novel axonal regeneration-promoting activity for oxidized galectin-1. *Curr Drug Targets.* **2005**;6:375–383.
63. Pande AH, Gupta RK, Sumati, et al. Oxidation of goat hepatic galectin-1 induces change in secondary structure. *Protein Pept Lett.* **2003**;10:265–275.
64. Hirabayashi J, Kasai K. Effect of amino acid substitution by sitedirected mutagenesis on the carbohydrate recognition and stability of human 14-kDa β -galactoside-binding lectin. *J Biol Chem.* **1991**;266:23648–23653.
65. Abbott WM, Feizi T. Soluble 14-kDa β -galactoside-specific bovine lectin. Evidence from mutagenesis and proteolysis that almost the complete polypeptide chain is necessary for integrity of the carbohydrate recognition domain. *J Biol Chem.* **1991**;266:5552–5557.
66. Nishi N, Abe A, Iwaki J, et al. Functional and structural bases of a cysteine-less mutant as a long-lasting substitute for galectin-1. *Glycobiology.* **2008**;18:1065–1073.
67. Guardia CM, Caramelo JJ, Trujillo M, et al. Structural basis of redox-dependent modulation of galectin-1 dynamics and function. *Glycobiology.* **2014**;24:428–441.
68. Horie H, Kadoya T, Hikawa N, et al. Oxidized galectin-1 stimulates macrophages to promote axonal regeneration in peripheral nerves after axotomy. *J Neurosci.* **2004**;24:1873–1880.
69. Paz A, Haklai R, Elad-Sfadia G, et al. Galectin-1 binds oncogenic H-Ras to mediate Ras membrane anchorage and cell transformation. *Oncogene.* **2001**;20:7486–7493.
70. Elad-Sfadia G, Haklai R, Ballan E, et al. Galectin-1 augments Ras activation and diverts Ras signals to Raf-1 at the expense of phosphoinositide 3-kinase. *J Biol Chem.* **2002**;277:37169–37175.
71. Yu X, Scott SA, Pritchard R, et al. Redox state influence on human galectin-1 function. *Biochimie.* **2015**;116:8–16.
72. Rotblat B, Niv H, Andre S, et al. Galectin-1(L11A) predicted from a computed galectin-1 farnesyl-binding pocket selectively inhibits Ras-GTP. *Cancer Res.* **2004**;64:3112–3118.
73. Tejler J, Salameh B, Leffler H, et al. Fragment-based development of triazole-substituted O-galactosyl aldoximes with fragment-induced affinity and selectivity for galectin-3. *Org Biomol Chem.* **2009**;7:3982.
74. Van Hattum H, Branderhorst HM, Moret EE, et al. Tuning the preference of thiodigalactoside- and lactosamine-based ligands to galectin-3 over galectin-1. *J Med Chem.* **2013**;56:1350–1354.
75. Cumpstey I, Carlsson S, Leffler H, et al. Synthesis of a phenyl thio- β -D-galactopyranoside library from 1,5-difluoro-2,4-

- dinitrobenzene: discovery of efficient and selective monosaccharide inhibitors of galectin-7. *Org Biomol Chem*. 2005;3:1922–1932.
76. Giguère D, Patnam R, Bellefleur M, et al. Carbohydrate triazoles and isoxazoles as inhibitors of galectins-1 and -3. *Chem Commun (Camb)*. 2006;22:2379–2381.
 77. Giguère D, Sato S, St-Pierre C, et al. Aryl O- and S-galactosides and lactosides as specific inhibitors of human galectins-1 and -3: role of electrostatic potential at O-3. *Bioorg Med Chem Lett*. 2006;16:1668–1672.
 78. Giguère D, Bonin MA, Cloutier P, et al. Synthesis of stable and selective inhibitors of human galectins-1 and -3. *Bioorg Med Chem*. 2008;16:7811–7823.
 79. Pandey RK, Dougherty TJ. Galectin recognized photosensitizers for photodynamic therapy. *US6849607B2* 2005.
 80. Huflejt M, Mossine V, Croft M. Galectins -1 and -4 in tumor development. *US20030109464A1* 2003.
 81. Ralph SJ. Immunomodulating compositions and uses thereof. *US9050352B2* 2015.
 82. Leffler H, Salameh B, Nilsson U. 3-Triazolyl-galactoside inhibitors of galectins. *US770763B2* 2010.
 83. Nilsson U, Leffler H, Mukhopadhyay B, et al. Novel galectoside inhibitors of galectins. *US20140336146A1* 2014.
 84. Nilsson U, Leffler H, Henderson N, et al. Galactoside inhibitor of galectin-3 and its use for treating pulmonary fibrosis. *WO2014067986A1* 2014.
 - **Galectin-3 inhibitors based on TDG scaffold – TD139 being one that is currently in clinical trials.**
 85. Klyosov A, Platt D. Co-administration of a polysaccharide with a chemotherapeutic agent for the treatment of cancer. *US7012068B2* 2006.
 86. Platt D, Klyosov A. Selectively depolymerized galactomannan polysaccharide. *US7893252B2* 2011.
 87. Platt D, Zomer E, Klyosov A. Galactose-pronged polysaccharides in a formulation for antifibrotic therapies. *US8722645B2* 2014.
 88. Traber PG, Zomer E, Klyosov AA. Galacto-rhamnogalacturonate compositions for the treatment of non-alcoholic steatohepatitis and non-alcoholic fatty liver disease. *US8658787B2* 2014.
 89. Zomer E, Traber PG, Klyosov AA, et al. Composition of novel carbohydrate drug for treatment of human diseases. *US8962824B2* 2015.
 90. Traber PG. Method for treatment of pulmonary fibrosis. *US20140235571A1* 2014.
 91. Traber PG, Redmond WL, Zomer E, et al. Method for enhancing specific immunotherapies in cancer treatment. *US20140086932A1* 2014.
 92. Traber PG, Zomer E, Klyosov AA. Galacto-rhamnogalacturonate compositions for the treatment of diseases associated with elevated inducible nitric oxide synthase. *US20150147338A1* 2015.
 93. Staples M, Rolke J. Modified pectins, compositions and methods related thereto. *US8877263B2* 2014.
 94. Chang Y, Sasak V. Composition and uses of galectin antagonist. *US2015133399A1* 2015.
 95. Chang Y, Sasak V. Method and material for treating immune diseases. *US20030004132A1* 2003.
 96. Chang Y, Sasak V. Method for controlling angiogenesis in animals. *US20040121981A1* 2004.
 97. Chang Y. Method for treating neurodegenerative diseases. *US20060014719A1* 2006.
 98. Chang Y, Cotter F. Composition and method for treating hyperproliferative diseases. *US20060074050A1* 2006.
 99. Mayo K, Dings R, Griffin R. Tumor treatment using beta-sheet peptides and radiotherapy. *US2007010438A1* 2007.
 100. Glucosyl Inc. Functionally active recombinant peptides, methods for producing same and interactions with other peptides. *WO2006128027A1* 2006.
 101. Mayo KH, Hoye TR, Flader-Lavey C. Partial peptide mimetics and methods. *US7339023B2* 2008.
 102. Mayo KH, Hoye TR, Chen X. Calixarene-based peptide conformation mimetics, methods of use, and methods of making. *US8716343B2* 2014.
 103. Astorgues-Xerri L, Cvitkovic E, Faivre S, et al. Compounds inhibiting galectin-1 expression, cancer cell proliferation, invasion, and tumorigenesis. *WO2012131079A1* 2012.
 104. Mayo KH. Anti-tumor agent otx-008 targets human galectin-1. *WO2014070214A1* 2014.
 105. Lappchen T, Gruell H, Robillard MS, et al. Analog (s) radiolabeled (s) of a 0118 and its/their use in connection with PET imaging and/or compound spect to determine if a pharmaceutical product containing the compound 0118 is a candidate anti-cancer treatment for a patient. *EP2858681* 2015.
 106. Dings RP, Hoye TR, Levine JI, et al. Cytotoxic agents against cancer cells and uses thereof. *WO2012061395A2* 2012.
 107. Saraiva MJ, Vanlandschoot P, Dolk E, et al. Amino acid sequences directed against multitarget scavenger receptors and polypeptides. *US9034325B2* 2015.
 108. Shipp MA, Ouyang J, Rodig SJ. Anti-galectin-1 (gal1) monoclonal antibodies and fragments thereof for neutralizing gal1. *WO2015013388A2* 2015.
 109. Shipp MA, Ouyang J, Takeyama K, et al. Compositions, kits, and methods for the diagnosis, prognosis, monitoring, treatment and modulation of post-transplant lymphoproliferative disorders and hypoxia associated angiogenesis disorders using galectin-1. *US8968740B2* 2015.
 110. Okano H, Sawamoto K, Sakaguchi M, et al. Agent for inhibiting proliferation of neural stem cells. *US7662385B2* 2010.
 111. Camby I, Henriot P, Lefranc F, et al. Use of a galectin-1-targeted RNAi-based approach for the treatment of cancer. *US7964575B2* 2011.
 112. Payne RM, Babbey CM, Martin KB, et al. Gigaxonin fusion protein and methods for treating giant axonal neuropathy. *EP2771367* 2014.
 113. Dardt D. Methods and compositions for modulating conjunctival goblet cells. *US20070185014A1* 2007.
 114. Panjwani N. Composition and uses of a galectin for treatment of dry eye syndrome. *US20100004163A1* 2010.
 115. Jun W. Preventives/remedies for nephritis. *WO2002089831A1* 2002.
 116. Cummings RD, Cho MJ. Methods of screening for compounds which mimic galectin-1. *US5948628A* 1999.
 117. Cummings RD, Cho MJ. Methods of screening for compounds which mimic galectin-1. *US6225071B1* 2001.
 118. Dimitroff CJ, Laurent FC, Barthel SR. Galectin-immunoglobulin chimeric molecules. *US8598323B2* 2013.
 119. Horie H, Inagaki Y, Sohma Y, et al. Neuronal growth factor galectin-1. *US6890531B1* 2005.
 120. John C, Unger G. Biologic modulations with nanoparticles. *US20040023855A1* 2004.
 121. Chao-Liang W, Yen-Jang H, Ai-Li S, et al. Galectin-1-gold particle complex and applications thereof. *TW201410702A* 2014.
 122. Salameh BA, Leffler H, Nilsson UJ. 3-(1,2,3-Triazol-1-yl)-1-thio-galactosides as small, efficient, and hydrolytically stable inhibitors of galectin-3. *Bioorg Med Chem Lett*. 2005;15:3344–3346.
 123. Öberg CT, Leffler H, Nilsson UJ. Arginine binding motifs: design and synthesis of galactose-derived arginine tweezers as galectin-3 inhibitors. *J Med Chem*. 2008;51:2297–2301.
 124. Öberg CT, Norell A-L, Leffler H, et al. Arene-anion based arginine-binding motif on a galactose scaffold: structure-activity relationships of interactions with arginine-rich galectins. *Chem Eur J*. 2011;17:8139–8144.
 125. Rajput VK, Leffler H, Nilsson UJ, et al. Synthesis and evaluation of iminocoumarinyl and coumarinyl derivatized glycosides as galectin antagonists. *Bioorg Med Chem Lett*. 2014;24:3516–3520.
 126. André S, Giguère D, Dam TK, et al. Synthesis and screening of a small glycomimetic library for inhibitory activity on medically relevant galactoside-specific lectins in assays of increasing biorelevance. *New J Chem*. 2010;34:2229.
 127. Cumpstey I, Salomonsson E, Sundin A, et al. Double affinity amplification of galectin–ligand interactions through arginine–arene interactions: synthetic, thermodynamic, and computational studies with aromatic diamido thiodigalactosides. *Chem Eur J*. 2008;14:4233–4245.

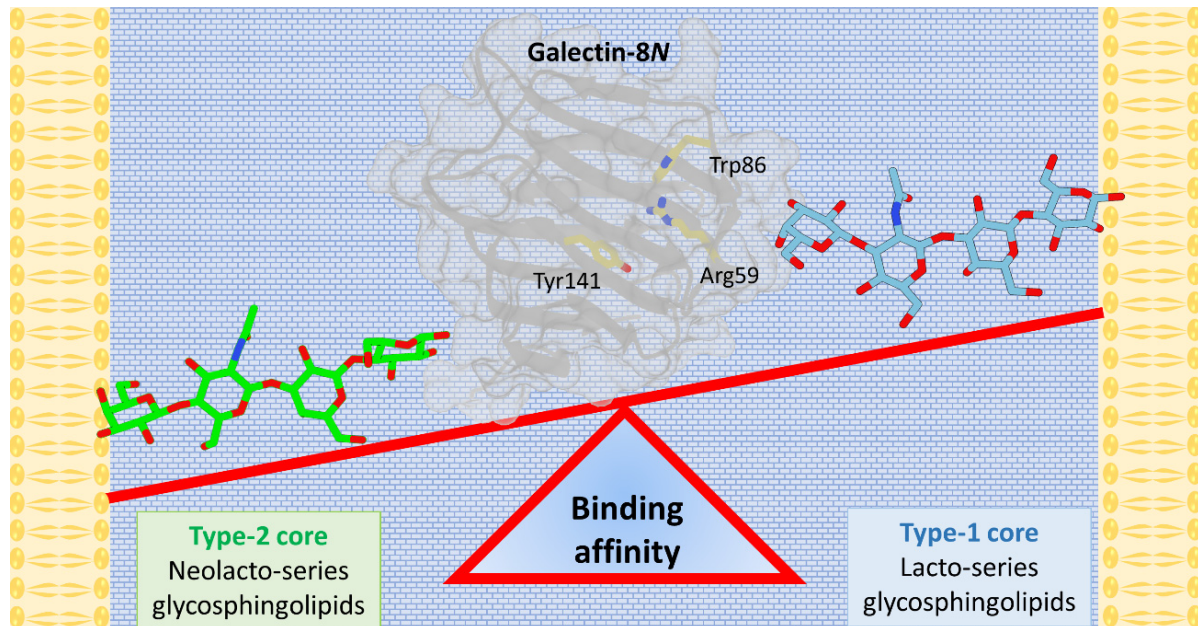
128. Cumpstey I, Salomonsson E, Sundin A, et al. Studies of arginine-arene interactions through synthesis and evaluation of a series of galectin-binding aromatic lactose esters. *ChemBioChem*. 2007;8:1389–1398.
129. Ingrassia L, Nshimyumukiza P, Dewelle J, et al. A lactosylated steroid contributes in vivo therapeutic benefits in experimental models of mouse lymphoma and human glioblastoma. *J Med Chem*. 2006;49:1800–1807.
130. Glinsky GV, Price JE, Glinsky VV, et al. Inhibition of human breast cancer metastasis in nude mice by synthetic glycoamines. *Cancer Res*. 1996;56:5319–5324.
131. Glinsky VV, Glinsky GV, Rittenhouse-Olson K, et al. The role of Thomsen-Friedenreich antigen in adhesion of human breast and prostate cancer cells to the endothelium. *Cancer Res*. 2001;61:4851–4857.
132. Glinskii OV, Sud S, Mossine VV, et al. Inhibition of prostate cancer bone metastasis by synthetic TF antigen mimic/galectin-3 inhibitor lactulose-L-leucine. *Neoplasia*. 2012;14:65–73.
133. Glinsky VV, Kiriakova G, Glinskii OV, et al. Synthetic galectin-3 inhibitor increases metastatic cancer cell sensitivity to taxol-induced apoptosis in vitro and in vivo. *Neoplasia*. 2009;11:901–909.
134. Rabinovich GA, Cumashi A, Bianco G, et al. Synthetic lactulose amines: novel class of anticancer agents that induce tumor-cell apoptosis and inhibit galectin-mediated homotypic cell aggregation and endothelial cell morphogenesis. *Glycobiology*. 2006;16:210–220.
135. Stannard KA, Collins PM, Ito K, et al. Galectin inhibitory disaccharides promote tumour immunity in a breast cancer model. *Cancer Lett*. 2010;299:95–110.
136. Stowell SR, Qian Y, Karmakar S, et al. Differential roles of galectin-1 and galectin-3 in regulating leukocyte viability and cytokine secretion. *J Immunol*. 2008;180:3091–3102.
137. Delaine T, Cumpstey I, Ingrassia L, et al. Galectin-inhibitory thiodigalactoside ester derivatives have antimigratory effects in cultured lung and prostate cancer cells. *J Med Chem*. 2008;51:8109–8114.
138. Van Scherpenzeel M, Moret EE, Ballell L, et al. Synthesis and evaluation of new thiodigalactoside-based chemical probes to label galectin-3. *ChemBioChem*. 2009;10:1724–1733.
139. Takeuchi T, Tamura M, Nishiyama K, et al. Mammalian galectins bind Galactose β 1–4Fucose disaccharide, a unique structural component of protostomal N-type glycoproteins. *Biochem Biophys Res Commun*. 2013;436:509–513.
140. Tejler J, Skogman F, Leffler H, et al. Synthesis of galactose-mimicking 1H-(1,2,3-triazol-1-yl)-mannosides as selective galectin-3 and 9N inhibitors. *Carbohydr Res*. 2007;342:1869–1875.
141. Öberg CT, Norellson A-L, Leffler H, et al. Synthesis of 3-amido-3-deoxy- β -D-talopyranosides: all-cis-substituted pyranosides as lectin inhibitors. *Tetrahedron*. 2011;67:9164–9172.
142. Collins PM, Öberg CT, Leffler H, et al. Taloside inhibitors of galectin-1 and galectin-3. *Chem Biol Drug Des*. 2012;79:339–346.
- **The first atomic structures of a galectin bound with taloside-based carbohydrates.**
143. Kiessling L, Gestwicki JE, Strong L. Synthetic multivalent ligands in the exploration of cell-surface interactions. *Curr Opin Chem Biol*. 2000;4:696–703.
144. André S, Cejas Ortega PJ, Perez MA, et al. Lactose-containing starburst dendrimers: influence of dendrimer generation and binding-site orientation of receptors (plant/animal lectins and immunoglobulins) on binding properties. *Glycobiology*. 1999;9:1253–1261.
145. André S, Pieters RJ, Vrasidas I, et al. Wedgelike glycodendrimers as inhibitors of binding of mammalian galectins to glycoproteins, lactose maxiclusters, and cell surface glycoconjugates. *ChemBioChem*. 2001;2:822–830.
146. Vrasidas I, André S, Valentini P, et al. Rigidified multivalent lactose molecules and their interactions with mammalian galectins: a route to selective inhibitors. *Org Biomol Chem*. 2003;1:803–810.
147. André S, Kaltner H, Furukie T, et al. Persubstituted cyclodextrin-based glycoclusters as inhibitors of protein-carbohydrate recognition using purified plant and mammalian lectins and wild-type and lectin-gene-transfected tumor cells as targets. *Bioconjug Chem*. 2004;15:87–98.
148. André S, Sansone F, Kaltner H, et al. Calix[n]arene-based glycoclusters: bioactivity of thiourea-linked galactose/lactose moieties as inhibitors of binding of medically relevant lectins to a glycoprotein and cell-surface glycoconjugates and selectivity among human adhesion/growth-regulatory galectins. *ChemBioChem*. 2008;9:1649–1661.
149. Cecioni S, Matthews SE, Blanchard H, et al. Synthesis of lactosylated glycoclusters and inhibition studies with plant and human lectins. *Carbohydr Res*. 2012;356:132–141.
150. Gouin SG, García Fernández JM, Vanquelef E, et al. Multimeric lactoside “click clusters” as tools to investigate the effect of linker length in specific interactions with peanut lectin, galectin-1, and -3. *Chem Eur J Chem Bio*. 2010;11:1430–1442.
151. Giguère D, André S, Bonin M-A, et al. Inhibitory potential of chemical substitutions at bioinspired sites of β -D-galactopyranose on neoglycoprotein/cell surface binding of two classes of medically relevant lectins. *Bioorg Med Chem*. 2011;19:3280–3287.
152. Wang G-N, André S, Gabius H-J, et al. Bi- to tetravalent glycoclusters: synthesis, structure-activity profiles as lectin inhibitors and impact of combining both valency and headgroup tailoring on selectivity. *Org Biomol Chem*. 2012;10:6893.
153. Tejler J, Tullberg E, Frejd T, et al. Synthesis of multivalent lactose derivatives by 1,3-dipolar cycloadditions: selective galectin-1 inhibition. *Carbohydr Res*. 2006;341:1353–1362.
154. Klyosov A, Zomer E, Platt D, DAVANAT * (GM-CT-01) and colon cancer: preclinical and clinical (phase I and II) studies. *ACS Symposium Series: American Chemical Society (ACS)*. 2012;89–130.
155. Blanchard H, Bum-Erdene K, Hugo MW. Inhibitors of galectins and implications for structure-based design of galectin-specific therapeutics. *Aust J Chem*. 2014;67:1763.
156. Blanchard H, Yu X, Collins PM, et al. Galectin-3 inhibitors: a patent review (2008-present). *Expert Opin Ther Pat*. 2014;24:1053–1065.
157. Mayo KH, Van Der Schaft DWJ, Griffioen AW. Designed beta-sheet peptides that inhibit proliferation and induce apoptosis in endothelial cells. *Angiogenesis*. 2001;4:45–51.
- **Important design of the peptide inhibitor anginex and example of effect of this antagonist.**
158. Wang JB, Wang MD, Li EX, et al. Advances and prospects of anginex as a promising anti-angiogenesis and anti-tumor agent. *Peptides*. 2012;38:457–462.
159. Mizukami Y, Jo W-S, Duerr E-M, et al. Induction of interleukin-8 preserves the angiogenic response in HIF-1 α -deficient colon cancer cells. *Nat Med*. 2005;11:992–997.
160. Dings RP, Arroyo MM, Lockwood NA, et al. Beta-sheet is the bioactive conformation of the anti-angiogenic anginex peptide. *Biochem J*. 2003;373:281–288.
161. Arroyo MM, Mayo KH. NMR solution structure of the angiostatic peptide anginex. *Biochim Biophys Acta (BBA) Proteins Proteom*. 2007;1774:645–651.
162. Akerman ME, Pilch J, Peters D, et al. Angiostatic peptides use plasma fibronectin to home to angiogenic vasculature. *Proc Natl Acad Sci*. 2005;102:2040–2045.
163. Thijssen VL, Postel R, Brandwijk RJ, et al. Galectin-1 is essential in tumor angiogenesis and is a target for antiangiogenesis therapy. *Proc Natl Acad Sci USA*. 2006;103:15975–15980.
- **Research on the PTX008-potent inhibitors of angiogenesis in cell proliferation and migration assays and in the mouse models of ovarian carcinoma and melanoma.**
164. Griffioen AW, Van Der Schaft DW, Barendsz-Janson AF, et al. Anginex, a designed peptide that inhibits angiogenesis. *Biochem J*. 2001;354:233–242.
165. Van Der Schaft DW, Dings RP, De Lussanet QG, et al. The designer anti-angiogenic peptide anginex targets tumor endothelial cells and inhibits tumor growth in animal models. *Faseb J*. 2002;16:1991–1993.
- **Important example of the effect of the galectin-1 antagonist anginex.**
166. Dings RPM, Van Laar ES, Webber J, et al. Ovarian tumor growth regression using a combination of vascular targeting agents

- anginex or topomimetic 0118 and the chemotherapeutic irifolven. *Cancer Lett.* **2008**;265:270–280.
167. Dings RPM, Van Laar ES, Loren M, et al. Inhibiting tumor growth by targeting tumor vasculature with galectin-1 antagonist anginex conjugated to the cytotoxic acylfulvene, 6-hydroxylpropylacylfulvene. *Bioconjug Chem.* **2010**;21:20–27.
 168. Dings RP, Yokoyama Y, Ramakrishnan S, et al. The designed angiostatic peptide anginex synergistically improves chemotherapy and antiangiogenesis therapy with angiostatin. *Cancer Res.* **2003**;63:382–385.
 169. Amano M, Suzuki M, Andoh S, et al. Antiangiogenesis therapy using a novel angiogenesis inhibitor, anginex, following radiation causes tumor growth delay. *Int J Clin Oncol.* **2007**;12:42–47.
 170. Dings RP, Loren M, Heun H, et al. Scheduling of radiation with angiogenesis inhibitors anginex and Avastin improves therapeutic outcome via vessel normalization. *Clin Cancer Res.* **2007**;13:3395–3402.
 171. Brandwijk RJ, Nesmelova I, Dings RP, et al. Cloning an artificial gene encoding angiostatic anginex: from designed peptide to functional recombinant protein. *Biochem Biophys Res Commun.* **2005**;333:1261–1268.
 172. Brandwijk RJMG, Dings RPM, Van Der Linden E, et al. Anti-angiogenesis and anti-tumor activity of recombinant anginex. *Biochem Biophys Res Commun.* **2006**;349:1073–1078.
 173. Mayo KH, Dings RP, Flader C, et al. Design of a partial peptide mimetic of anginex with antiangiogenic and anticancer activity. *J Biol Chem.* **2003**;278:45746–45752.
 174. Dings RP, Kumar N, Miller MC, et al. Structure-based optimization of angiostatic agent 6DBF7, an allosteric antagonist of galectin-1. *J Pharmacol Exp Ther.* **2013**;344:589–599.
 - **Peptide antagonist with properties improved upon from anginex.**
 175. Dings RP, Chen X, Hellebrekers DM, et al. Design of nonpeptidic topomimetics of antiangiogenic proteins with antitumor activities. *J Natl Cancer Inst.* **2006**;98:932–936.
 - **Design of potent inhibitors of angiogenesis.**
 176. Zucchetti M, Bonezzi K, Frapolli R, et al. Pharmacokinetics and antineoplastic activity of galectin-1-targeting OTX008 in combination with sunitinib. *Cancer Chemother Pharmacol.* **2013**;72:879–887.
 177. Dings RP, Levine JI, Brown SG, et al. Polycationic calixarene PTX013, a potent cytotoxic agent against tumors and drug resistant cancer. *Invest New Drugs.* **2013**;31:1142–1150.
 178. Salomonsson E, Thijssen VL, Griffioen AW, et al. The anti-angiogenic peptide anginex greatly enhances galectin-1 binding affinity for glycoproteins. *J Biol Chem.* **2011**;286:13801–13804.
 179. Dings RPM, Miller MC, Nesmelova I, et al. Antitumor agent calixarene 0118 targets human galectin-1 as an allosteric inhibitor of carbohydrate binding. *J Med Chem.* **2012**;55:5121–5129.
 180. Pilch J, Franzin CM, Knowles LM, et al. The anti-angiogenic peptide anginex disrupts the cell membrane. *J Mol Biol.* **2006**;356:876–885.
 181. Croci DO, Cerliani JP, Dalotto-Moreno T, et al. Glycosylation-dependent lectin- receptor interactions preserve angiogenesis in anti-VEGF refractory tumors. *Cell.* **2014**;156:744–758.
 182. Dalotto-Moreno T, Croci DO, Cerliani JP, et al. Targeting galectin-1 overcomes breast cancer-associated immunosuppression and prevents metastatic disease. *Cancer Res.* **2013**;73:1107–1117.
 183. Baum LG, Pang M, Perillo NL, et al. Human thymic epithelial cells express an endogenous lectin, galectin-1, which binds to core 2 O-glycans on thymocytes and T lymphoblastoid cells. *J Exp Med.* **1995**;181:877–887.
 184. Rabinovich GA, Ariel A, Herschkovitz R, et al. Specific inhibition of T-cell adhesion to extracellular matrix and proinflammatory cytokine secretion by human recombinant galectin-1. *Immunology.* **1999**;97:100–106.
 185. Levi G, Tarrab-Hazdai R, Teichberg VI. Prevention and therapy with electrolectin of experimental autoimmune myasthenia gravis in rabbits. *Eur J Immunol.* **1983**;13:500–507.
 186. Offner H, Celnik B, Bringman TS, et al. Recombinant human beta-galactoside binding lectin suppresses clinical and histological signs of experimental autoimmune encephalomyelitis. *J Neuroimmunol.* **1990**;28:177–184.
 187. Harjacek M, Diaz-Cano S, De Miguel M, et al. Expression of galectins-1 and -3 correlates with defective mononuclear cell apoptosis in patients with juvenile idiopathic arthritis. *J Rheumatol.* **2001**;28:1914–1922.
 188. Xibillé-Friedmann D, Bustos Rivera-Bahena C, Rojas-Serrano J, et al. A decrease in galectin-1 (Gal-1) levels correlates with an increase in anti-Gal-1 antibodies at the synovial level in patients with rheumatoid arthritis. *Scand J Rheumatol.* **2013**;42:102–107.
 189. Santucci L, Fiorucci S, Cammilleri F, et al. Galectin-1 exerts immunomodulatory and protective effects on concanavalin A-induced hepatitis in mice. *Hepatology.* **2000**;31:399–406.
 190. Santucci L, Fiorucci S, Rubinstein N, et al. Galectin-1 suppresses experimental colitis in mice. *Gastroenterology.* **2003**;124:1381–1394.
 191. Li S, Yu Y, Koehn CD, et al. Galectins in the pathogenesis of rheumatoid arthritis. *J Clin Cell Immunol.* **2013**;4:1000164.
 192. Kuo PL, Hung JY, Huang SK, et al. Lung cancer-derived galectin-1 mediates dendritic cell anergy through inhibitor of DNA binding 3/IL-10 signaling pathway. *J Immunol.* **2011**;186:1521–1530.
 193. Watanabe M, Takemasa I, Kaneko N, et al. Clinical significance of circulating galectins as colorectal cancer markers. *Oncol Rep.* **2011**;25:1217–1226.
 194. Saussez S, Glinier D, Chantrain G, et al. Serum galectin-1 and galectin-3 levels in benign and malignant nodular thyroid disease. *Thyroid.* **2008**;18:705–712.
 195. Kim HJ, Jeon HK, Cho YJ, et al. High galectin-1 expression correlates with poor prognosis and is involved in epithelial ovarian cancer proliferation and invasion. *Eur J Cancer.* **2012**;48:1914–1921.
 196. Kim HJ, Do IG, Jeon HK, et al. Galectin-1 expression is associated with tumor invasion and metastasis in stage IB to IIA cervical cancer. *Hum Pathol.* **2013**;44:62–68.
 197. Solaro R, Chiellini F, Battisti A. Targeted delivery of protein drugs by nanocarriers. *Materials.* **2010**;3:1928–1980.
 198. Battig P, Saudan P, Gunde T, et al. Enhanced apoptotic activity of a structurally optimized form of galectin-1. *Mol Immunol.* **2004**;41:9–18.
 199. Bi S, Earl LA, Jacobs L, et al. Structural features of galectin-9 and galectin-1 that determine distinct. *J Biol Chem.* **2008**;283:12248–12258.
 200. Vertesy S, Michalak M, Miller MC, et al. Structural significance of galectin design: impairment of homodimer stability by linker insertion and partial reversion by ligand presence. *Protein Eng Des Sel.* **2015**;28:199–210.
 201. Earl LA, Bi S, Baum LG. Galectin multimerization and lattice formation are regulated by linker region structure. *Glycobiology.* **2011**;21:6–12.
 202. Cedeno-Laurent F, Barthel SR, Opperman MJ, et al. Development of a nascent galectin-1 chimeric molecule for studying the role of leukocyte galectin-1 ligands and immune disease modulation. *J Immunol.* **2010**;185:4659–4672.
 203. Wang CR, Shiau AL, Chen SY, et al. Intra-articular lentivirus-mediated delivery of galectin-3 shRNA and galectin-1 gene ameliorates collagen-induced arthritis. *Gene Ther.* **2010**;17:1225–1233.
 204. Weber W, Fussenegger M. Pharmacologic transgene control systems for gene therapy. *J Gene Med.* **2006**;8:535–556.
 205. Baum LG, Blackall DP, Arias-Magallano S, et al. Amelioration of graft versus host disease by galectin-1. *Clin Immunol.* **2003**;109:295–307.
 206. Nicolette CA, Fathman CG, Creusot R. Transient expression of immunomodulatory polypeptides for the prevention and treatment of autoimmune disease, allergy and transplant rejection. *US8513208B2* **2013**.
 207. Tsai CY, Shiau AL, Chen SY, et al. Amelioration of collagen-induced arthritis in rats by nanogold. *Arthritis Rheum.* **2007**;56:544–554.
 208. Huang YJ, Shiau AL, Chen SY, et al. Multivalent structure of galectin-1-nanogold complex serves as potential therapeutics for rheumatoid arthritis by enhancing receptor clustering. *Eur Cell Mater.* **2012**;23:170–181.
 209. Claudia Oliveira IS, Veiga F, Ribeiro AJ. Recent advances in characterization of nonviral vectors for delivery of nucleic acids: impact

- on their biological performance. *Expert Opin Drug Deliv.* [2015](#);12:27–39.
210. Saie AA, Ray M, Mahmoudi M, et al. Engineering the nanoparticle-protein interface for cancer therapeutics. *Cancer Treat Res.* [2015](#);166:245–273.
 211. Chatin B, Mevel M, Devalliere J, et al. Liposome-based formulation for intracellular delivery of functional proteins. *Mol Ther Nucl Acids.* [2015](#);4:e244.
 212. Harris JM, Chess RB. Effect of pegylation on pharmaceuticals. *Nat Rev Drug Discov.* [2003](#);2:214–221.
 213. Veronese FM, Harris JM. Introduction and overview of peptide and protein pegylation. *Adv Drug Deliv Rev.* [2002](#);54:453–456.
 214. Kopitz J, Fik Z, André S, et al. Single-site mutational engineering and following MonoPEGylation of the human lectin galectin-2: effects on ligand binding, functional aspects, and clearance from serum. *Mol Pharm.* [2013](#);10:2054–2061.

Chapter 2

Interaction of Galectin-8 *N*-terminal Domain with Glycosphingolipids



2.1 Foreword

There are various galectins present in humans each has its distinct role to play apart from certain common biological involvement. Structurally all galectin CRDs share the “*jelly roll*” topology, and the primary binding site is mostly conserved for all the members, that recognises the galactose/lactose core of the oligosaccharides. However, the extended binding site has key variations that are responsible for governing the specificity in recognising larger oligosaccharides. Therefore, understanding the nuances in recognising cell surface glycans interactions with galectins is critical. To this end, the interactions of galectin-8N with lacto- and neolacto-series glycosphingolipids, which are also major components of human milk glycans, were studied. Crystallographic analysis was performed to gain atomic level information about their interactions and provided a structure-based rationale for the differences in the binding affinity. Our molecular dynamics simulations revealed the importance of extended binding site residues in differentially recognising these glycosphingolipids. This study thus provides the atomic details of their interactions and furthers our understanding of galectin-8’s biological functions. In addition, the minimum atomic framework revealed from the glycerol and lactose complex structures hinted towards developing potential ligands.

2.2 Statement of contribution to a co-authored publication

This chapter includes a co-authored paper. The bibliographic details of the co-authored paper, including all authors, are:

Mohammad H. Bohari, Xing Yu, Yehiel Zick and Helen Blanchard (2016). Scientific Reports. 6:39556. doi: 10.1038/srep39556.

My contribution to the paper involved:

Subcloning, expression, purification, crystallisation and structure determination of the galectin-8N protein bound to LNT, LNnT, lactose and glycerol ligands with advice from Dr. Xing Yu and Prof. Helen Blanchard. Molecular dynamics simulations and analysis of galectin-8N-LNT including the *in silico* Tyr141Ala-galectin-8N-LNT mutant and galectin-3-LNT complex. Input into interpretation and analysis of the generated data and contribution to writing the manuscript.


(Signed) _____ (Date) 03/04/2017 _____

Mohammad Bohari


(Countersigned) _____ (Date) 24/11/2017 _____

Corresponding author of paper: Professor Helen Blanchard


(Countersigned) _____ (Date) 24/11/2017 _____

Supervisor: Professor Helen Blanchard

2.3 Structure-based rationale for differential recognition of lacto- and neolacto- series glycosphingolipids by the *N*-terminal domain of human galectin-8.

Reprinted from Scientific Report, Volume 6, Mohammad H. Bohari, Xing Yu, Yehiel Zick and Helen Blanchard, “Structure-based rationale for differential recognition of lacto- and neolacto-series glycosphingolipids by the *N*-terminal domain of human galectin-8”, Article number 39556, Copyright (2016), with permission from Nature Publishing Group. [doi: 10.1038/srep39556]

Published article: <http://www.nature.com/articles/srep39556>

SCIENTIFIC REPORTS

OPEN

Structure-based rationale for differential recognition of lacto- and neolacto- series glycosphingolipids by the *N*-terminal domain of human galectin-8

Mohammad H. Bohari¹, Xing Yu¹, Yehiel Zick² & Helen Blanchard¹

Received: 12 September 2016

Accepted: 23 November 2016

Published: 21 December 2016

Glycosphingolipids are ubiquitous cell surface molecules undertaking fundamental cellular processes. Lacto-*N*-tetraose (LNT) and lacto-*N*-neotetraose (LNnT) are the representative core structures for lacto- and neolacto-series glycosphingolipids. These glycolipids are the carriers to the blood group antigen and human natural killer antigens mainly found on blood cells, and are also principal components in human milk, contributing to infant health. The β -galactoside recognising galectins mediate various cellular functions of these glycosphingolipids. We report crystallographic structures of the galectin-8 *N*-terminal domain (galectin-8N) in complex with LNT and LNnT. We reveal the first example in which the non-reducing end of LNT binds to the primary binding site of a galectin, and provide a structure-based rationale for the significant ten-fold difference in binding affinities of galectin-8N toward LNT compared to LNnT, such a magnitude of difference not being observed for any other galectin. In addition, the LNnT complex showed that the unique Arg59 has ability to adopt a new orientation, and comparison of glycerol- and lactose-bound galectin-8N structures reveals a minimum atomic framework for ligand recognition. Overall, these results enhance our understanding of glycosphingolipids interactions with galectin-8N, and highlight a structure-based rationale for its significantly different affinity for components of biologically relevant glycosphingolipids.

Glycosphingolipids are ubiquitous cell surface molecules containing, at minimum, a monosaccharide joined by a glycosidic linkage to either ceramide or sphingoid¹. These glycolipid molecules are involved in the fundamental cellular processes such as cell adhesion and signal transduction, mediated through protein-protein or protein-carbohydrate interaction². Lacto-*N*-tetraose (LNT) and Lacto-*N*-neotetraose (LNnT) are the tetrasaccharides that form the core structural component of the lacto- and neo-lacto glycosphingolipid series, respectively. LNT and LNnT differ only in the type of glycosidic linkage within the non-reducing end disaccharide (Gal β 1-3/4GlcNAc) (Fig. 1). Structurally, these tetrasaccharides also resemble poly-*N*-acetylglucosamine chains. The upregulation and modification of poly-*N*-acetylglucosamines into tumour-associated antigens is reported to be correlated with cancer progression^{3,4}. The presence of galactose/*N*-acetylglucosamine/fucose linked to the GlcNAc ring of these tetrasaccharide makes them equivalent to the core structural component of the blood group antigens. The lacto-/neolacto-series are also the carriers to some functional antigens such as blood group antigens and HNK-1 (human natural killer-1/CD57) antigens found on the hematopoietic cells and play important roles in the immune system⁵. The cell surface composition and presentation for these glycolipids vary in cancerous cells compared to normal cells, and also regulates the fate of tumour progression⁶.

¹Institute for Glycomics, Griffith University, Gold Coast Campus, 4222, Australia. ²Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel. Correspondence and requests for materials should be addressed to H.B. (email: h.blanchard@griffith.edu.au)

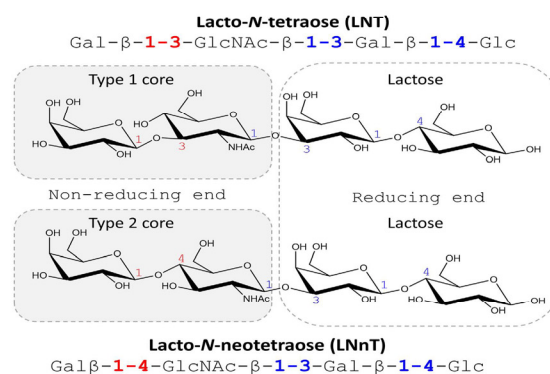


Figure 1. LNT and LNTnT oligosaccharide structures. The LNT type 1 and the LNTnT type 2 cores, as well as the reducing and non-reducing ends of the disaccharides are indicated.

Galectins are a class of lectin that recognise β -galactoside containing glycans, including the glycosphingolipids⁷. Galectin-8 is a member of the tandem-repeat galectin category, having two CRDs in tandem that are joined by an amino acid linker of variable length. Galectin-8 was first identified in prostate⁸ and lung cancer cells⁹, and later found to be widely distributed in normal tissues as well as tumour cells^{10–14}. Similar to other galectins, galectin-8 is present in the nucleus, the cytoplasm and the extracellular space¹⁵. Based on the cellular context, galectin-8 regulates integrin-mediated cell adhesion, growth, and apoptosis^{16–18}. Induction of neutrophil adhesion is a unique feature of galectin-8^{16,18}. Galectin-8 has been shown to regulate T-cell homeostasis, exhibiting immunomodulatory and inflammatory roles with the implication in rheumatoid arthritis and uveitis^{19–23}. Galectin-8 also plays a critical role in the capillary tube formation and endothelial cell migration *in vitro* and angiogenesis *in vivo*²⁴. The ability of this lectin to regulate the bone remodelling process, by increasing expression of RANKL *in vitro* and increased bone turnover *in vivo*, can be exploited for bone-loss diseases²⁵. Further, galectin-8 selectively recognises bacterial expressing blood group antigens²⁶, with galectin-8C being shown to selectively recruit NDP52 to engulf the invading pathogens²⁷.

Intracellularly, galectin-8 interactions are mainly protein-protein however it has significant interactions with glycans on the cell surface. At high concentrations, galectin-8 recognises a broad range of glycans²⁸ whilst specifically recognising the blood group antigens at submicromolar concentrations²⁶. The two CRDs of galectin-8, i.e. galectin-8N and galectin-8C, share ~40% sequence identity and exhibit differential glycan binding specificities. Galectin-8N recognises a broader spectrum of glycans compared to galectin-8C, and notably exhibits a preferential binding towards anionic sugars²⁹. A significant contribution to the preferential binding of galectin-8 towards 3'-O-sulfate/3'-O-sialylated lactose, determined using surface plasmon resonance, was attributed to the unique binding site residues of the galectin-8N²⁹. The majority of galectin-8 structures reported to date are for its N-terminal domain (galectin-8N): in its *apo* form, bound to lactose, and also in complex with 3'-O-sulfated lactose, 3'-O-sialylated lactose and LNFIII³⁰. However, a few galectin-8C structures also have been reported both in the *apo* (unpublished), and bound to NDP52 peptide³¹. Structural characterisation of the full-length galectin-8 has been challenging due to high flexibility and protease susceptibility of the linker. Nevertheless, the structure of a truncated galectin-8 comprising a dipeptide linker joining the CRDs has been solved, and the truncated protein was shown to retain the neutrophil adhesion function as the intact full-length galectin-8³².

The tetrasaccharide LNT and LNTnT are major components of human milk, providing a source of carbohydrates to the infant and also acting as physiological and immunological regulators of the intestinal tract^{33–35}. Being rich in galactose-based glycans (lactose in particular), their interaction with galactose-recognising proteins such as galectins is of special interest. Furthermore, a recent study reports a systematic analysis of interactions of various galectin-human milk glycans, highlighting the significance of their interactions in infant health³⁶. Interestingly, one of the major milk glycans found to be recognised by galectin-8 was essentially LNTnT with an additional disaccharide (LacNAc) joined by a β 1-3 linkage to the non-reducing end galactose. We have reported crystal structures of galectin 3 (4LBM³⁷ and 4LBN³⁷) and also galectin 4C (4YM0³⁸ and 4YLZ³⁸) in complex with LNT and LNTnT, providing atomic details of these protein-receptor interactions. Crystal structures of galectin-9N bound to LacNAc dimers, that structurally resemble LNTnT (differing only by the *N*-acetyl group) have also been reported (2ZHK³⁹ and 2ZHL³⁹). Though these galectins all bind to both these glycosphingolipid core structures, they do show evidence of fine specificity amongst them, as well as between the N- and C-terminal CRD domains in the case of tandem-repeat galectins.

In this study, we have performed crystallographic analysis to gain atomic level information pertaining to the lacto- and neolacto-series glycosphingolipids interactions with galectin-8N. We provide a structure-based rationale for the difference in binding affinities of LNT and LNTnT, based on the observed alternative binding modes. Molecular dynamics (MD) simulations performed on the galectin-8N-LNT complex revealed

	Galectin-8N-LNT	Galectin-8N-LNnT	Galectin-8N-Lactose	Galectin-8N-Glycerol
Crystal system	Orthorhombic	Orthorhombic	Orthorhombic	Orthorhombic
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell	$a = 45.40$,	$a = 47.17$,	$a = 47.61$,	$a = 47.40$,
	$b = 49.61$,	$b = 50.14$,	$b = 50.40$,	$b = 50.30$,
	$c = 80.47^*$	$c = 69.86$	$c = 69.73$	$c = 69.39$
Resolution (Å)	1.96	2.00	1.90	1.58
Total observations	99554 (7554)	96027 (7320)	119349 (7359)	182683 (7925)
Unique observations	12696 (929)	11695 (994)	13910 (934)	23530 (1110)
Multiplicity	7.8 (8.1)	8.2 (8.7)	8.6 (7.9)	7.8 (7.1)
Completeness (%)	93.7 (100)	99.6 (100)	100 (100)	100 (99.6)
I/σ	14.5 (11.8)	26.0 (11.4)	24.0 (8.9)	24.3 (7.3)
R_{merge} (%)	11.2 (15.3)	5.7 (14.7)	5.9 (21.9)	4.6 (19.3)
Refinement				
Resolution	42.23–1.96	40.73–2.00	40.85–1.90	40.73–1.58
R factor (%)	17.6	18.9	14.9	11.7
R_{free} (%)	21.8	21.7	17.9	14.7
Number of atoms				
Protein	1197	1178	1183	1201
Ligand	26	48	23	6
Water molecules	245	178	223	239
Root mean square deviation				
Bond length (Å)	0.0055	0.0086	0.0065	0.0105
Bond angle (°)	1.1742	1.3969	1.2554	1.5338
Ramachandran plot statistics				
Favoured (%)	98.56	97.93	98.61	97.84
Allowed (%)	1.44	2.07	1.39	2.16
Average B-factor (Å ²)				
Protein	12.77	14.99	15.03	13.83
Ligand	20.95	18.47	21.81	27.32
Water	23.34	23.90	29.98	31.86
PDB ID	5T7T	5T7I	5T7S	5T7U

Table 1. Crystallographic data for galectin-8N-ligand complex structures. The values in parenthesis are for the highest-resolution shell. *Note the increase in cell edge length compared to other complexes.

interesting features in the binding site governing the recognition of oligosaccharide by galectin-8N. Comparison of the crystal structures reported in this study with previously reported galectins interacting with either these tetrasaccharides or similar ligands has led to an understanding of the nuances in binding modes and interactions across the galectin family. Crystallographic analysis of galectin-8N-lactose and galectin-8N-glycerol complexes revealed the minimum atomic framework required by galectins for ligand recognition. The atomic details of the binding mode and associated interactions of naturally occurring cell surface oligosaccharides with cell-to-cell communicating agents, such as galectin-8N, provide systematic understanding of the molecular phenomena.

Results and Discussion

X-ray crystallographic structures of human galectin-8N were determined with bound LNT, LNnT, lactose, and glycerol at 1.58–2.00 Å resolution (Table 1).

Galectin-8N-LNT complex. The galectin-8N CRD exhibits a typical β -sandwich comprising two anti-parallel β -sheets with the concave side housing the carbohydrate-binding site. The binding groove is formed from six beta strands labelled S1 to S6 (Fig. 2), with amino acids on strand S4–S6 forming the glycan recognition pocket (referred to as the “primary binding site”). The amino acids on strand S1–S3 form the extended binding site and are involved in recognising oligosaccharides (Fig. 2). The galectin-8N binding site contains a few unique residues which are either different or absent in other galectins. In galectin-8N the unique features are the presence of a long S3–S4 loop bearing an arginine (Arg59), the Gln47 on strand S3, Ile91 on S6, and Tyr141 on S2 (Fig. 2). These unique features potentially contribute to imparting glycan recognition specificity and may thereby affect the overall function of galectin-8. The typical binding pattern of galectin-8N towards disaccharides (and effectively that of typical reducing-end interactions of larger oligosaccharides) is shown by our 1.9 Å resolution lactose-bound structure (Fig. 3a, Supplementary Fig. S1), which we obtained by soaking lactose into an *apo*-galectin-8N-crystal (Table 1). The O4' of lactose engages in hydrogen bonding with His65, Asn67, Arg45, and Arg69; the O6' hydrogen bonds with Asn79, Glu89, and the glucose O3 interacts with Arg69 and Glu89.

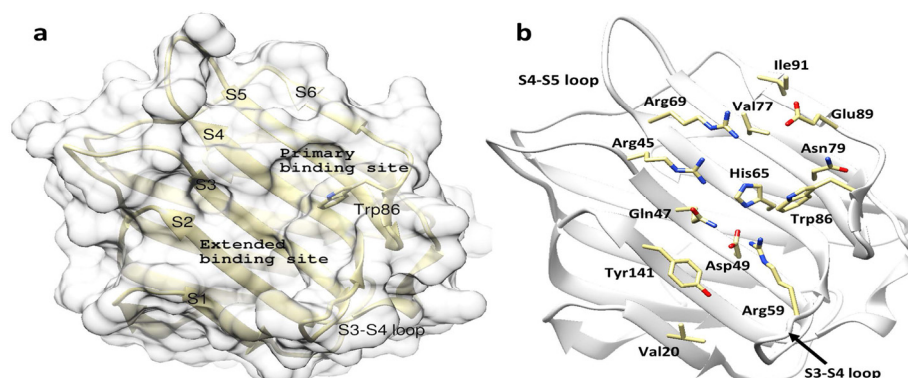


Figure 2. Overview of galectin-8N carbohydrate recognition domain. (a) The CRD (yellow ribbons) showing the carbohydrate binding face of the β -sandwich and the primary and extended binding regions labelled with strand S1–S6. (b) Depicts the amino acid residue (yellow carbon, oxygen red, nitrogen blue; stick representation) involved in glycan binding interactions.

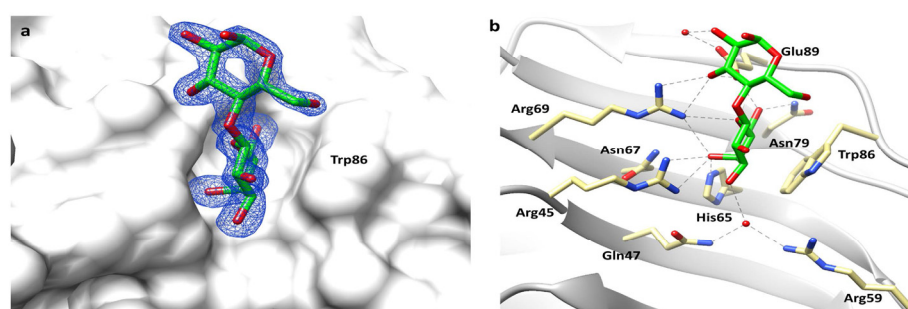


Figure 3. Galectin-8N in complex with lactose. (a) Electron density map (blue mesh) $2[F_o] - [F_c]$ contoured at 1σ , for lactose (carbon green, oxygen red; stick representation) in complex with galectin-8N (surface representation). (b) Hydrogen bonding interactions (grey dashed lines) made between lactose and galectin-8N binding site residues (carbon yellow, oxygen red and nitrogen blue; stick representation).

Water-mediated interactions were also observed between the galactose O3' with the unique Arg59 and glucose O2 with Glu89 (Fig. 3b). The indirect interaction with the unique Arg59 is likely a contributing factor to the galectin-8 domain differences in binding affinity toward carbohydrates, as exemplified by galectin-8N having an affinity for lactose of $79\mu\text{M}$, compared to $440\mu\text{M}$ by galectin-8C²⁹. The binding mode of lactose observed in our structure is identical to previously reported galectin-8N-lactose bound [2YXS (unpublished), 3AP4³⁰] and that of other galectin lactose complexes, as would be expected due to the evolutionarily conserved galactose recognition residues.

The crystal structure of galectin-8N-LNT was obtained by soaking LNT into an *apo*-crystal, and determined at a resolution of 1.96 \AA (Table 1). Electron density associated with the oligosaccharide was clearly visible only to the extent of revealing a disaccharide portion. Based on the structure of LNT, either the reducing end (Gal β 1-4Glc) disaccharide, the non-reducing end (Gal β 1-3GlcNAc; type 1 core) disaccharide or the middle disaccharide portion (GlcNAc β 1-3Gal) (Fig. 1) would have potential to bind at the primary binding site. Of these three possible disaccharides, the non-reducing end was identified to fit the electron density map. The clear bulge of the galactose O4' facing toward the Arg45 and Arg69, confirmed the presence of a galactose ring at the conserved galactose recognition site. Critically, the *N*-acetyl group of the GlcNAc was identified clearly in the electron density, and confirmed the location of this carbohydrate moiety (Fig. 4a, Supplementary Fig. S2). In addition, there is positive difference electron density extending at C1 of GlcNAc that implies the direction for the reducing end

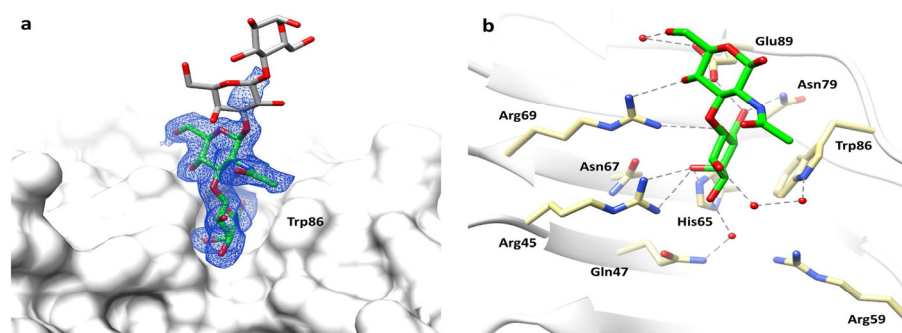


Figure 4. Galectin-8N in complex with LNT. (a) Electron density map (blue mesh) $2[F_o] - [F_c]$ σ_c contoured at 1σ , for the non-reducing end disaccharide (type 1 core) portion of LNT (carbon green, oxygen red; stick representation) in complex with galectin-8N (surface representation). Also represented is the possible position of the reducing end disaccharide of LNT (carbon grey, oxygen red; stick representation) directed into the solvent. (b) Hydrogen bonding interactions (grey dashed lines) made by the non-reducing end disaccharide of LNT (green carbon, oxygen red, nitrogen blue; stick representation) with the galectin-8N binding site residues (yellow carbon, oxygen red, nitrogen blue; stick representation).

disaccharide (Fig. 4a). The binding of LNT to galectin-8N is not influenced by crystal contacts, evident by a large solvent channel alongside the ligand-binding site. Interestingly, the unit cell parameters of the galectin-8N-LNT complex differs from the glycerol-, lactose-, LNnT-bound structures, with the unit cell edge length c increased by ~ 10 Å whereas there is a decrease of ~ 2 Å in the cell edge length a (Table 1).

Thus in summary, the electron density map clearly shows that a galactose ring is stacked against the conserved Trp86, with the β 1-3 linked GlcNAc occupying the adjacent position (that site occupied by glucose in galectin-lactose complexes) (Fig. 4a, Supplementary Fig. S2). This is the first report of the non-reducing end disaccharide occupying the primary binding site, in place of the traditionally observed reducing end disaccharide, for the tetrasaccharide LNT. This mode of type 1 core binding has not been observed previously for any other galectin. The galactose ring portion of the non-reducing end type 1 core itself makes identical interactions to that traditionally exhibited by the galactose of lactose. Significantly though, the difference in the glycosidic linkage (β 1-3) of the type 1 core, as compared to β 1-4 in lactose (and type 2 core), leads to variation in the placement of the GlcNAc ring and differences in overall interaction profile. Surface plasmon resonance has shown a greater binding affinity for lactose ($79 \mu\text{M}$), over the type 1 core ($160 \mu\text{M}$)²⁹. The GlcNAc ring is flipped by $\sim 180^\circ$ causing the N -acetyl group to be solvent exposed and not directly interacting with binding site residues, whilst the O4 directly hydrogen bonds with Arg69 and Glu89 and the O6 interacts with Glu89 through a water molecule (Fig. 4b). The nature of the glycosidic linkage and the placement of N -acetyl group are factors that influence the resulting binding affinity observed for type 1 core, which is weaker than for lactose, but stronger than for LacNAc ($420 \mu\text{M}$)²⁹. Going from lactose to LacNAc results in a 5-fold weaker binding, but then changing from LacNAc to incorporate a β 1,3-linkage results in an affinity that is just 2-fold weaker than lactose and ~ 2.6 -fold stronger than for LacNAc. Conformation of the type 1 core disaccharide (Type 1 N -Acetyl-lactosamine) bound to galectin-1 (4XBL⁴⁰), galectin-3 (4XBN⁴⁰) and galectin-7 (4XBQ⁴⁰), observed in crystal structures, also reveal an identical placement of the galactose ring, whereas the slight variation in the glycosidic torsional angle between the two rings results in different GlcNAc orientation. Given the higher average B-factor for GlcNAc (~ 19 Å²) as compared to the galactose ring (~ 14 Å²) observed in our structure, which is a similar trend to other galectin-glycan complexes, overall the binding conformation of the type 1 core appears to be comparable throughout the galectin family.

Molecular dynamics simulations of galectin-8N-LNT complex. Differences in binding affinities between the two tetrasaccharides with galectin-8 reported using two independent methods, suggested a weaker affinity of LNT compared to that of LNnT, that is particularly pronounced for the galectin-8N domain^{26,41} (Table 2). For galectin-3, the affinities are relatively comparable with just a slight indication of weaker binding for LNT; but the magnitude of difference in the case of galectin-8N is significant (Table 2). In the case of galectin-3, both tetrasaccharides bind by their reducing ends, with the placement of the LNT non-reducing end galactose more exposed to solvent than for LNnT³⁷. The implication for galectin-3 is that the different glycosidic linkages within the non-reducing end is not a dominating factor with respect to the overall binding affinities of these two tetrasaccharides. Our structure of galectin-8N with the LNT positioning its non-reducing end at the primary binding site is the first example of such a binding mode in galectins, and we propose that this alternative binding mode has could be the cause of the 10-fold magnitude of difference in the affinity that is reoproted between LNT and LNnT towards galectin-8N²⁹.

Ligands	Galectin-8N		Galectin-3 ⁴¹
	SPR ²⁹	FA ⁴²	FA ⁴¹
Lactose	79	3.1	2.8
LacNAc	420	9.7	1.8
LNT	140	2.1	0.97
LNnT	13	0.33	0.65

Table 2. Binding affinities (K_d μ M) of oligosaccharides towards galectins^{29,41,61}. ²⁹SPR - Surface Plasmon Resonance; ⁴¹FA - Fluorescence Anisotropy.

Tyrosine 141 on the S2 strand is unique to galectin-8N, and is strikingly different to the amino acid at that position in other galectins, examples being Asn (galectin-8C, galectin-9N), Asp (galectin-1, galectin-4N), Gln (galectin-4C), Gly (galectin-9C) and Ser in galectin-3. Interestingly, in galectin-8N this amino acid influences binding affinity of LNnT. This was demonstrated by a Tyr141Ser mutation that resulted in significant reduction in affinity for LNnT (20 μ M) from that observed for wild-type galectin-8N (0.33 μ M), as determined by fluorescence anisotropy⁴². Given the varied nature of this amino acid, that is positioned within the extended binding site region, coupled with the site-directed mutation which clearly shows it has influence on binding the tetrasaccharide then here we investigated whether the Tyr141 at this position could prove crucial in defining the binding mode and profile of galectin-8N towards larger oligosaccharides, and in particular whether this Tyr141 would be responsible for the alternate binding mode observed for LNT. To investigate the behaviour of LNT bound to galectin-8N in solution, MD simulations were carried out. For all the galectin-8N-LNT simulations, the starting LNT conformation used was from the galectin-3-LNT complex (4LBM³⁷) where the typical behaviour of reducing end occupation of the primary binding site is exhibited (as in all other galectin-LNT structures). This approach would offer insight into why galectin-8N favoured instead binding to the non-reducing end of LNT. Furthermore, to investigate the influence of Tyr141 in the binding of LNT, the Tyr141Ala mutant of galectin-8N, and the galectin-3-LNT (4LBM) structure, were simulated. The analysis of MD results were mainly focused on the position and conformation of the non-reducing end disaccharide of LNT.

All the three systems considered here, specifically: *wt*-galectin-8N-LNT, Tyr141Ala-galectin-8N-LNT and galectin-3-LNT, showed retention of the ligand in the binding site throughout the length of the simulation. Of note is that the hydrogen bonding interactions made by the reducing end disaccharide of LNT with the conserved binding site residues showed almost 100% occupancy. Interactions made by GlcNAc with galectin-8N were transient due to flexibility imparted by the highly fluctuating non-reducing galactose ring (Fig. 5a). When LNT is bound via its reducing end then this most flexible galactose is positioned above the Tyr141. In the case of the *wt*-galectin-8N-LNT simulation the Tyr141 was one of the most flexible residues, it initially stays flat and facing the protein surface but then flips-up by about 70° after approximately 1 ns of simulation (Fig. 5a). The non-reducing end galactose then moves further away from the protein surface to accommodate the flipped Tyr141 and becomes even more flexible, and more solvent exposed. We believe that this flipping of Tyr141 induces an overall shift in the ligand positioning that results in the non-reducing end occupying the primary binding site as we reveal in the galectin-8N-LNnT complex structure, and in contrast to other galectins-LNT complexes.

The results of the galectin-3-LNT complex (4LBM³⁷) MD simulations, that have the reducing end of LNT occupying the primary binding site, supports our hypothesis pertaining to the galectin-8N-LNT complex. In the case of galectin-3, there is a serine (Ser235) in place of Tyr141. As anticipated, the non-reducing end galactose in the galectin-3-LNT complex simulation was less fluctuating when compared with that observed in the *wt*-galectin-8N-LNT simulation (Fig. 5c). This relatively lower flexibility of the non-reducing end galactose ring possibly explains the occurrence of the reducing end galactose in the primary binding site in the galectin-3-LNT complex, unlike the alternative binding mode observed in galectin-8N-LNT complex. The presence of a smaller residue, for example the Ser in galectin-3 in place of Tyr, allows the non-reducing end galactose to relatively stabilise more during simulation and thereby weakly interact with the protein surface which eventually results in the traditional scenario of the reducing end occupying the primary binding site.

To further support the impact of the size of the amino acid at the Tyr141 location, an *in silico* Tyr141Ala mutant was also subjected to MD simulation. The small side chain of alanine should not affect the non-reducing end galactose, and therefore we should see a more stable positioning of the non-reducing end galactose. The overlayed trajectory from simulations clearly show that the galactose ring was less flexible, and occupied just one conformation for the major part of the simulation, in contrast to the situation for the *wt*-galectin-8N-LNT complex (Fig. 5b). Overall, the order of flexibility of the non-reducing end galactose ring, governed by the amino acid residue positioned beneath is: galectin-8N-LNT > galectin-3-LNT > Tyr141Ala-galectin-8N-LNT. We anticipate that this order predicts that the Tyr141Ala-galectin-8N-LNT complex is most likely to witness the binding mode observed for LNT bound to other galectins, where the reducing end of LNT occupies the primary binding site. Thus, the MD simulation analysis supports our novel crystallographic findings and identified a critical residue Tyr141 that may potentially play a key role in determining the alternative binding mode for galectin-8N-LNT complex.

Galectin-8N-LNnT complex. The galectin-8N-LNnT complex was obtained by soaking LNnT into an *apo*-crystal, and the structure was determined at 2.0 Å (Table 1). The electron density maps showed unambiguous electron density for all the four sugars of the tetrasaccharide, and with the reducing end occupying the primary binding site (Fig. 6a, Supplementary Fig. S3). The lactose portion of LNnT retains the interactions observed

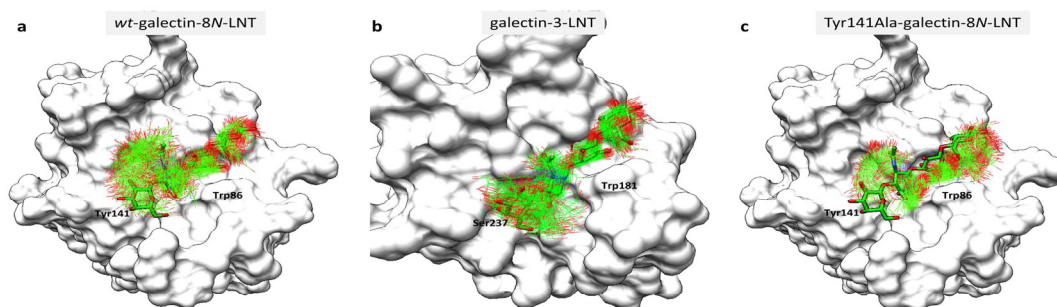


Figure 5. Overlay of trajectories from MD simulations. The coordinates of ligand extracted from simulations (carbon green, oxygen red, nitrogen blue; line representation) were superimposed onto the starting conformation (green carbon, oxygen red, nitrogen blue; stick representation). (a) Simulation of wt-galectin-8N-LNT complex. (b) Simulation of galectin-3-LNT complex. (c) Simulation of Tyr141Ala-galectin-8N-LNT mutant complex.

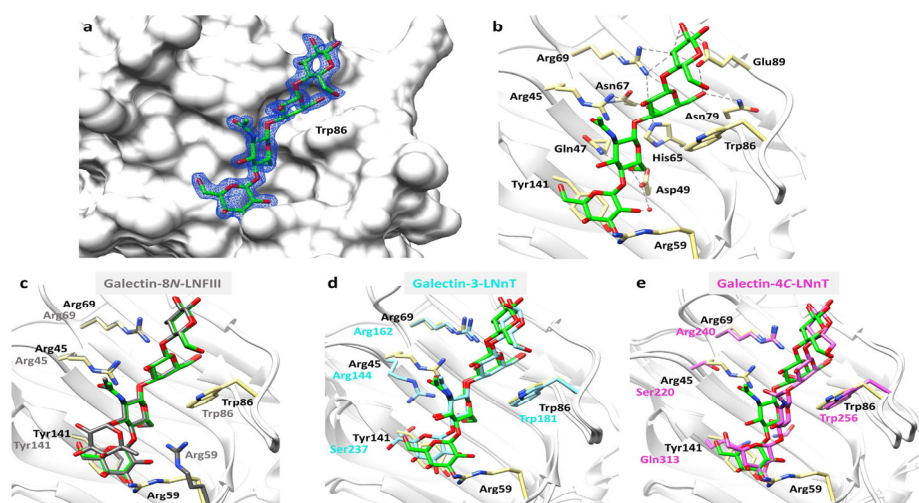


Figure 6. Galectin-8N in complex with LNnT. (a) Electron density map (blue mesh) $2[F_o] - [F_c]$ α_c contoured at 1σ , for LNnT (carbon green, oxygen red; stick representation) in complex with galectin-8N (surface representation). (b) Hydrogen bonding interactions (grey dashed lines) made by LNnT (green carbon; sticks) with the galectin-8N binding site residues (yellow carbon; sticks). (c–e) Superimposed (C_α atoms) conformation of LNnT observed in galectin-8N-LNnT complex with that of previously reported: (c) galectin-8N-LNFI complex (3AP9³⁰; grey carbon; sticks), (d) galectin-3-LNnT complex (4LBN³⁷; cyan carbon; sticks) and (e) galectin-4C-LNnT complex (4YLZ³⁸; magenta carbon; sticks).

previously in the galectin-8N-lactose complex (Fig. 6b). The other two sugars of LNnT extend through a β 1-3 linkage into the extended binding site of galectin-8N contacting residues on strand S3 and S2 (Fig. 3). Due to the β 1-4 linkage within the non-reducing end disaccharide (type 2 core), as opposed to the β 1-3 linkage in LNT, all the four LNnT sugars stay close to the protein surface and interact with both the primary and the extended binding sites. The O6 of GlcNAc interacts with Gln47 and Asp49 (Fig. 6b) and O3 is pointing away from the binding site into the solvent. The non-reducing end galactose ring forms CH- π type interactions with Tyr141 and O2 engages in hydrogen bonding interactions with Asp49 (Fig. 6b). Arg59, is oriented away from the conserved Trp86, and may be involved in water-mediated interactions with the non-reducing galactose.

LNFIH is a branched pentasaccharide that contains LNT as a core structure with an additional α 1-3 linked fucose on the non-reducing end GlcNAc. The binding affinity of galectin-8N is stronger for LNFIH (3.3 μ M) than for LNT (13 μ M)²⁹. The crystal structure of galectin-8N-LNFIH complex (3AP9)³⁰ revealed the structural basis for galectin-8N's greater affinity for LNFIH than for LNT. Essentially, the non-reducing end galactose and the branched fucose ring of LNFIH interact with the Tyr141 through CH- π type of interaction, enhancing binding affinity²⁹. Overall, the conformation of the LNT core of LNFIH in the galectin-8N-LNFIH complex, is identical to that observed in our galectin-8N-LNT complex (Fig. 6c). The identical ligand placement further supports that the decreased binding affinity of galectin-8N toward LNT (13 μ M) compared to LNFIH (3.3 μ M) is due to the lack of a branched fucose ring on the GlcNAc in LNT. Despite having identical ligand conformation, the orientation of unique Arg59 on the long S3-S4 loop is different in our galectin-8N-LNT complex to that seen in all the reported galectin-8N *apo* and ligand-bound structures. In the galectin-8N-LNT complex, Arg59 is directed towards the non-reducing end galactose, whereas in the other structures, including the galectin-8N-LNFIH complex, it faces towards the conserved Trp86 (Fig. 6c). In the galectin-8N *apo* and lactose-bound structures, the conformation of Arg59 appears to be uninfluenced by the presence of the ligand, and consequently stretches forward towards the conserved Trp86, further forming water-mediated interactions with the lactose. For the galectin-8N-LNFIH complex, the incoming ligand displaces the water molecule (located towards the 3' position of lactose) and causes the Arg59 to move slightly away from conserved Trp86, where it then interacts with the O4 of GlcNAc. However, for the galectin-8N-LNT structure, the orientation of Arg59 is unique, and this finding may hold significance in ligand specificity.

To understand the differences in binding conformation within the galectin family, comparison was performed of the LNT conformation observed in our structure with that in galectin-3-LNT and galectin-4C-LNT complexes. In the case of galectin-3 (structure 4LBN³⁷), amino acid differences (*galectin-8N given in brackets*) such as Arg186 (Ile91) on S6, Ala146 (Gln47) on S3, Ser237 (Tyr141) on S2 and importantly, the absence of the long S3-S4 loop, influence the positioning of LNT and thereby cause a slight variation in the conformation to that found in the galectin-8N-LNT complex (Fig. 6d). In contrast, the LNT conformation in galectin-4C-domain (4YLZ³⁸) differs to a significant extent from our galectin-8N-LNT complex, possibly due to greater differences in the nature of the amino acids on strand S2. Gln313 (Tyr141 in galectin-8N, Ser237 in galectin-3) and Glu311 (Gly139 in galectin-8N, Gly235 in galectin-3) are large in size and their aliphatic chain facing the carbohydrate binding site, compared to their counterparts in galectin-3. In contrast, in the case of galectin-8N, Tyr141 (Gln313 in galectin-4C) although being a large amino acid, it has an aromatic side chain that stays parallel and forms CH- π type interactions with the incoming non-reducing end galactose ring. The presence of Gln313 and Glu311 in galectin-4C thus may cause the change in glycosidic torsion angle within the non-reducing end disaccharide, compared to that observed in LNT bound to galectin-8N. These differences cause the shift in the placement of the non-reducing end galactose, leading to the overall difference in the LNT conformation when in complex with galectin-4C. Thus the significance of residues present in the extended binding site governs the positioning of oligosaccharides, and anticipated to affect their overall binding strengths towards galectins.

Galectin-8N-glycerol complex. The cryoprotectant glycerol that was used during the cryo-cooling of the *apo*-galectin-8N crystal, soaked into the crystal and was unambiguously showed to occupy the galactose-binding site (Fig. 7a, Supplementary Fig. S4). Glycerol oxygen atoms engage in hydrogen bonding with His65, Arg69, Arg45, Asn79 and Glu89 whilst the carbon atom of glycerol makes van der Waals' interactions with the conserved Trp86 (Fig. 7b). The glycerol hydroxyls also form water-mediated hydrogen bonds with Arg45 (Fig. 7b). The most interesting revelation from the galectin-8N-glycerol complex comes from alignment of the glycerol conformation observed in our structure to that of the lactose conformation and positioning within the binding site. The alignment shows an exact overlap of three carbon atoms of glycerol onto C4', C5' and C6' atoms of the lactose's galactose ring with the identical positioning of oxygen atoms (Fig. 7c). Interestingly, the position of oxygen atom of glycerol in our structure along with the presence of a water molecule (mimicking O3 of glucose in the lactose-bound structure) matches exactly with the previously highlighted hotspots for ligand recognition from the galectin-3-glycerol structure (2NMO⁴³). The conformation of glycerol in our structure is also identical to that observed in other high-resolution galectin-3 structure (3ZSK⁴⁴) and also in the galectin-4N structure (5DUU⁴⁵). This indicates that the glycerol represents a moiety that exhibits key features desired for interaction by the galectins, though due to the smaller size, presence of only three hydroxyl groups and more importantly the lack of other key interactions (made by galactose) poses challenges in quantifying the glycerol binding affinity⁴⁴. The similarity in the conformation of glycerol or water molecule location within the galactose recognition site of galectins suggests common hotspots required for ligand recognition throughout the family. This basic atomic framework together with other interactions made by galactose can be incorporated into ligand design strategies for identifying efficient binders of galectins.

Concluding Remarks. Overall, the structures reported herein provide insight into the binding mode and interactions of lacto- and neolacto series glycosphingolipids with galectin-8N. The LNT and LNT complex structure are biologically significant as they are principal components in human milk, and also form the core structural component of the blood group antigens. We demonstrate for the first time the occupancy of the primary binding site of galectin-8N by the non-reducing end disaccharide (thus an alternative binding mode) of the tetrasaccharide LNT, contrasting the reducing end binding traditionally observed for galectins. Hence we provide a structure-based rationale for the 10-fold weaker binding affinity of LNT towards galectin-8N, compared to LNT. Amino acid differences in the extended binding site primarily governs the recognition of oligosaccharides and structures in the current study imply a preference of galectin-8N for neolacto-series (LNT) over lacto-series (LNT) glycosphingolipids. MD simulations investigating the possible reasons for the alternative-binding mode for LNT to galectin-8N, highlighted Tyr141 as a critical residue governing the recognition of LNT. In addition,

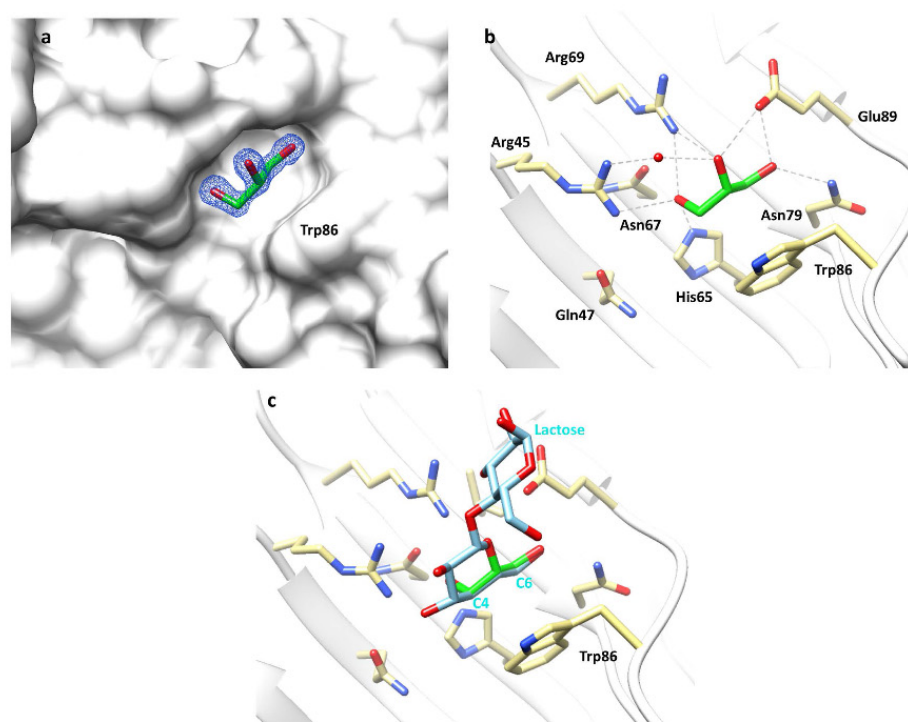


Figure 7. Galectin-8N in complex with glycerol. (a) Electron density map (blue mesh) $2|F_o| - |F_c|$ α_c contoured at 1σ , for glycerol (carbon green, oxygen red; stick representation) in complex with galectin-8N (surface representation). (b) Hydrogen bonding interactions (grey dashed lines) made by glycerol (green carbon; sticks) with the galectin-8N binding site residues (yellow carbon; sticks). (c) Superimposition of the observed glycerol conformation (green carbon; sticks) in the galectin-8N-glycerol complex to that of lactose conformation (cyan carbon; sticks) in the galectin-8N-lactose complex.

we observed a novel orientation of Arg59, which is a unique residue on the S3-S4 loop, in the galectin-8N-LNnT complex that may further hold significance in ligand specificity. The observed binding mode of glycerol that matches key atoms of the defining galectin ligand of galactose, highlights minimal ligand atomic features for recognition by galectins. Unique amino acid residues such as Arg59, Gln47, Ile91 and Tyr141 are other potential hotspots to not only gain affinity but also to explore specificity. Overall, taking into account critical residues in the binding site and the information about minimal atomic features for recognition, more potent and specific ligands could be designed. In all, this study not only highlights preference of galectin-8N towards recognising neolacto-series over lacto-series glycosphingolipids implying specific roles for the galectin-8 over other galectins but also points to structural features that can potentially be exploited for specificity in ligand design.

Materials and Methods

Materials. Oligosaccharides LNT and LNnT were purchased from Carbosynth Limited UK. Lactose was bought from Sigma US.

Sub-cloning, protein expression and purification. The galectin-8N sequence encoded in pQE vector was amplified using a forward primer (5'-G GAA TTC CAT ATG ATG TTG TCC TTA AAC AAC C-3') and a reverse primer (5'-CGC GGA TCC CTA CGA GCT GAA GCT AAA ACC-3') with NdeI and BamHI restriction sites (sequence underlined) at 5' and 3' direction respectively. Double digestion of PCR product by NdeI and BamHI allowed sticky ends ligation into pET-3a vector resulting in pET-3a- galectin-8N encoding the untagged galectin-8N. The integrity of galectin-8N gene sequence inserted into pET-3a was assessed by Australian Genome Research Facility Ltd. (AGRF, Queensland, Australia). The bacterial culture grown in LB medium were induced using 1 mM IPTG (isopropyl 3-D-1 thiogalactopyranoside) at room temperature for 4 h when OD of the medium reached 0.6. Cells were harvested and sonicated in a lysis buffer (10 mM sodium phosphate, 137 mM sodium chloride, 2.7 mM potassium chloride, 1.8 mM potassium phosphate; PBS) containing 1 mM PMSF (phenylmethylsulfonyl fluoride). The released protein was applied onto the lactosyl-Sepharose column and eluted using 50 mM lactose solution.

Crystallisation, X-ray data collection and structure determination. The galectin-8N oligosaccharide complexes were generated by soaking *apo* galectin-8N crystals in the presence of the LNT/LNnT/Lactose oligosaccharides. The *apo* galectin-8N crystals were formed in the phosphate buffer saline at 5 mg/mL (PBS) in the microcentrifuge tube over a period of one month in the refrigerator. These crystals were used for soaking oligosaccharide dissolved in PBS at 20 mM concentration for about 18 hours. Diffraction data for all the complexes were remotely collected at the Australian Synchrotron using Blu-Ice software⁴⁶ at 100 K with a wavelength of 0.9537 Å, and ADSC Quantum detector. The data were integrated using iMOSFLM⁴⁷, and the point group determination and scaling of the data was performed using AIMLESS⁴⁸. The phases were solved using Phaser⁴⁹ with the galectin-8N-*apo* structure (3AP5⁵⁰) as the search model. The model obtained was refined using REFMAC5^{50,51} in CCP4 program suite⁵². Visualisation and model building was done in Coot⁵³. Model validation and analysis were performed by MolProbity and PDB_REDO^{54,55}.

Molecular dynamics simulation. GROMACS version 4.5.6⁵⁶ was used for MD simulations with built-in AMBER99SB-ILDN force field⁵⁷, as used for other galectins simulations^{37,38}. Particle mesh Ewald method⁵⁸ was employed to compute long-range electrostatics. The ligand topology and parameters were generated using acpype⁵⁹ applied with General Amber Force Field⁶⁰ and AM1-BCC charges⁶¹. Initially, the protein-ligand complex was minimized using the steepest-descent method followed by brief simulation at constant volume and then 2-ns constant pressure equilibrations. Subsequently, 100 ns production run was carried out for the complex and analysis of results was carried out using the various script provided with GROMACS.

References

- Chester, M. A. IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN). Nomenclature of glycolipids—recommendations 1997. *European journal of biochemistry/FEBS* **257**, 293–298 (1998).
- Hakomori, S. Structure and function of glycosphingolipids and sphingolipids: Recollections and future trends. *Biochimica et biophysica acta* **1780**, 325–346, doi: 10.1016/j.bbagen.2007.08.015 (2008).
- Demetter, P. *et al.* The galectin family and digestive disease. *The Journal of Pathology* **215**, 1–12, doi: 10.1002/path.2334 (2008).
- Fukuda, M., Hiraoka, N. & Yeh, J.-C. C-Type Lectins and Sialyl Lewis X Oligosaccharides: Versatile Roles in Cell–Cell Interaction. *The Journal of Cell Biology* **147**, 467–470, doi: 10.1083/jcb.147.3.467 (1999).
- Togayachi, A. *et al.* Lack of lacto/neolacto-glycolipids enhances the formation of glycolipid-enriched microdomains, facilitating B cell activation. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 11900–11905, doi: 10.1073/pnas.0914298107 (2010).
- Hakomori, S. & Zhang, Y. Glycosphingolipid antigens and cancer therapy. *Chemistry & biology* **4**, 97–104 (1997).
- Perillo, N. L., Marcus, M. E. & Baum, L. G. Galectins: versatile modulators of cell adhesion, cell proliferation, and cell death. *Journal of molecular medicine (Berlin, Germany)* **76**, 402–412 (1998).
- Su, Z. Z. *et al.* Surface-epitope masking and expression cloning identifies the human prostate carcinoma tumor antigen gene PCTA-1 a member of the galectin gene family. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 7252–7257 (1996).
- Bidon, N. *et al.* Two messenger RNAs and five isoforms for Po66-CBP, a galectin-8 homolog in a human lung carcinoma cell line. *Gene* **274**, 253–262 (2001).
- Gopalkrishnan, R. V. *et al.* Molecular characterization of prostate carcinoma tumor antigen-1, PCTA-1, a human galectin-8 related gene. *Oncogene* **19**, 4405–4416, doi: 10.1038/sj.onc.1203767 (2000).
- Bidon-Wagner, N. & Le Pennec, J. P. Human galectin-8 isoforms and cancer. *Glycoconjugate journal* **19**, 557–563, doi: 10.1023/b:glyc.0000014086.38343.98 (2004).
- Lahm, H. *et al.* Comprehensive galectin fingerprinting in a panel of 61 human tumor cell lines by RT-PCR and its implications for diagnostic and therapeutic procedures. *Journal of cancer research and clinical oncology* **127**, 375–386 (2001).
- Hadari, Y. R. *et al.* Galectin-8. A new rat lectin, related to galectin-4. *The Journal of biological chemistry* **270**, 3447–3453 (1995).
- Nagy, N. *et al.* Galectin-8 expression decreases in cancer compared with normal and dysplastic human colon tissue and acts significantly on human colon cancer cell migration as a suppressor. *Gut* **50**, 392–401 (2002).
- Hadari, Y. R. *et al.* Galectin-8 binding to integrins inhibits cell adhesion and induces apoptosis. *Journal of cell science* **113**(Pt 13), 2385–2397 (2000).
- Nishi, N. *et al.* Galectin-8 modulates neutrophil function via interaction with integrin α M. *Glycobiology* **13**, 755–763, doi: 10.1093/glycob/cwg102 (2003).
- Zick, Y. *et al.* Role of galectin-8 as a modulator of cell adhesion and cell growth. *Glycoconjugate journal* **19**, 517–526, doi: 10.1023/B:GLYC.0000014081.55445.af (2004).
- Yamamoto, H. *et al.* Induction of cell adhesion by galectin-8 and its target molecules in Jurkat T-cells. *Journal of biochemistry* **143**, 311–324, doi: 10.1093/jb/mvm223 (2008).
- Eshkar Sebban, L. *et al.* The involvement of CD44 and its novel ligand galectin-8 in apoptotic regulation of autoimmune inflammation. *Journal of immunology (Baltimore, Md.: 1950)* **179**, 1225–1235 (2007).
- Sampson, J. E. *et al.* Galectin-8 Ameliorates Murine Autoimmune Ocular Pathology and Promotes a Regulatory T Cell Response. *PLoS one* **10**, e0130772, doi: 10.1371/journal.pone.0130772 (2015).
- Sampson, J. E., Suryawanshi, A., Chen, W. S., Rabinovich, G. A. & Panjwani, N. Galectin-8 promotes regulatory T-cell differentiation by modulating IL-2 and TGF β signaling. *Immunology and cell biology*, doi: 10.1038/icb.2015.72 (2015).
- Norambuena, A. *et al.* Galectin-8 induces apoptosis in Jurkat T cells by phosphatidic acid-mediated ERK1/2 activation supported by protein kinase A down-regulation. *The Journal of biological chemistry* **284**, 12670–12679, doi: 10.1074/jbc.M808949200 (2009).
- Tribulatti, M. V., Cattaneo, V., Hellman, U., Mucci, J. & Campetella, O. Galectin-8 provides costimulatory and proliferative signals to T lymphocytes. *Journal of leukocyte biology* **86**, 371–380, doi: 10.1189/jlb.0908529 (2009).
- Delgado, V. M. *et al.* Modulation of endothelial cell migration and angiogenesis: a novel function for the “tandem-repeat” lectin galectin-8. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology* **25**, 242–254, doi: 10.1096/fj.09-144907 (2011).
- Vinik, Y. *et al.* The mammalian lectin galectin-8 induces RANKL expression, osteoclastogenesis, and bone mass reduction in mice. *eLife* **4**, e05914, doi: 10.7554/eLife.05914 (2015).
- Stowell, S. R. *et al.* Innate immune lectins kill bacteria expressing blood group antigen. *Nature medicine* **16**, 295–301, doi: 10.1038/nm.2103 (2010).
- Thurston, T. L., Wandel, M. P., von Muhlinen, N., Foeglein, A. & Randow, F. Galectin 8 targets damaged vesicles for autophagy to defend cells against bacterial invasion. *Nature* **482**, 414–418, doi: 10.1038/nature10744 (2012).
- Stowell, S. R. *et al.* Dimeric Galectin-8 induces phosphatidylserine exposure in leukocytes through polyactosamine recognition by the C-terminal domain. *The Journal of biological chemistry* **283**, 20547–20559, doi: 10.1074/jbc.M802495200 (2008).

29. Ideo, H., Seko, A., Ishizuka, I. & Yamashita, K. The N-terminal carbohydrate recognition domain of galectin-8 recognizes specific glycosphingolipids with high affinity. *Glycobiology* **13**, 713–723, doi: 10.1093/glycob/cwg094 (2003).
30. Ideo, H., Matsuzaka, T., Nonaka, T., Seko, A. & Yamashita, K. Galectin-8 N-domain recognition mechanism for sialylated and sulfated glycans. *The Journal of biological chemistry* **286**, 11346–11355, doi: 10.1074/jbc.M110.195925 (2011).
31. Li, S. *et al.* Sterical hindrance promotes selectivity of the autophagy cargo receptor NDP52 for the danger receptor galectin-8 in antibacterial autophagy. *Science signaling* **6**, ra9, doi: 10.1126/scisignal.2003730 (2013).
32. Nishi, N. *et al.* Development of highly stable galectins: Truncation of the linker peptide confers protease-resistance on tandem-repeat type galectins. *FEBS Letters* **579**, 2058–2064, http://dx.doi.org/10.1016/j.febslet.2005.02.054 (2005).
33. Kunz, C., Rudloff, S., Baier, W., Klein, N. & Strobel, S. Oligosaccharides in human milk: structural, functional, and metabolic aspects. *Annual review of nutrition* **20**, 699–722, doi: 10.1146/annurev.nutr.20.1.699 (2000).
34. Bode, L. Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology* **22**, 1147–1162, doi: 10.1093/glycob/cws074 (2012).
35. Asakuma, S. *et al.* Physiology of consumption of human milk oligosaccharides by infant gut-associated bifidobacteria. *The Journal of biological chemistry* **286**, 34583–34592, doi: 10.1074/jbc.M111.248138 (2011).
36. Noll, A. J. *et al.* Galectins are human milk glycan receptors. *Glycobiology* **26**, 655–669, doi: 10.1093/glycob/cww002 (2016).
37. Collins, P. M., Bum-Erdene, K., Yu, X. & Blanchard, H. Galectin-3 Interactions with Glycosphingolipids. *Journal of Molecular Biology* **426**, 1439–1451, http://dx.doi.org/10.1016/j.jmb.2013.12.004 (2014).
38. Bum-Erdene, K., Leffler, H., Nilsson, U. J. & Blanchard, H. Structural characterization of human galectin-4 C-terminal domain: elucidating the molecular basis for recognition of glycosphingolipids, sulfated saccharides and blood group antigens. *The FEBS journal* **282**, 3348–3367, doi: 10.1111/febs.13348 (2015).
39. Nagae, M. *et al.* Structural analysis of the recognition mechanism of poly-N-acetyllactosamine by the human galectin-9 N-terminal carbohydrate recognition domain. *Glycobiology* **19**, 112–117, doi: 10.1093/glycob/cwn121 (2009).
40. Hsieh, T.-J. *et al.* Structural Basis Underlying the Binding Preference of Human Galectins-1, -3 and -7 for Gal β 1-3/4GlcNAc. *PLoS one* **10**, e0125946, doi: 10.1371/journal.pone.0125946 (2015).
41. Carlsson, S. *et al.* Affinity of galectin-8 and its carbohydrate recognition domains for ligands in solution and at the cell surface. *Glycobiology* **17**, 663–676, doi: 10.1093/glycob/cwm026 (2007).
42. Salomonsson, E. *et al.* Mutational tuning of galectin-3 specificity and biological function. *The Journal of biological chemistry* **285**, 35079–35091, doi: 10.1074/jbc.M109.098160 (2010).
43. Collins, P. M., Hidari, K. I. & Blanchard, H. Slow diffusion of lactose out of galectin-3 crystals monitored by X-ray crystallography: possible implications for ligand-exchange protocols. *Acta crystallographica. Section D, Biological crystallography* **63**, 415–419, doi: 10.1107/s090744490605270x (2007).
44. Saraboji, K. *et al.* The Carbohydrate-Binding Site in Galectin-3 Is Preorganized To Recognize a Sugarlike Framework of Oxygens: Ultra-High-Resolution Structures and Water Dynamics. *Biochemistry* **51**, 296–306, doi: 10.1021/bi201459p (2012).
45. Bum-Erdene, K., Leffler, H., Nilsson, U. J. & Blanchard, H. Structural characterisation of human galectin-4 N-terminal carbohydrate recognition domain in complex with glycerol, lactose, 3'-sulfo-lactose, and 2'-fucosyllactose. *Scientific Reports* **6**, doi: 10.1038/srep20289 (2016).
46. McPhillips, T. M. *et al.* Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. *Journal of synchrotron radiation* **9**, 401–406 (2002).
47. Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta crystallographica. Section D, Biological crystallography* **67**, 271–281, doi: 10.1107/s0907444910048675 (2011).
48. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta crystallographica. Section D, Biological crystallography* **69**, 1204–1214, doi: 10.1107/s0907444913000061 (2013).
49. McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of applied crystallography* **40**, 658–674, doi: 10.1107/s0021889807021206 (2007).
50. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological crystallography* **67**, 355–367, doi: 10.1107/s0907444911001314 (2011).
51. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta crystallographica. Section D, Biological crystallography* **53**, 240–255, doi: 10.1107/s0907444996012255 (1997).
52. The CCP4 suite: programs for protein crystallography. *Acta crystallographica. Section D, Biological crystallography* **50**, 760–763, doi: 10.1107/s0907444994003112 (1994).
53. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* **66**, 486–501, doi: 10.1107/s0907444910007493 (2010).
54. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography* **66**, 12–21, doi: 10.1107/s0907444909042073 (2010).
55. Joosten, R. P. *et al.* PDB_REDO: constructive validation, more than just looking for errors. *Acta crystallographica. Section D, Biological crystallography* **68**, 484–496 (2012).
56. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854, doi: 10.1093/bioinformatics/btt055 (2013).
57. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958, doi: 10.1002/prot.22711 (2010).
58. Essmann, U. *et al.* A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103**, 8577–8593, http://dx.doi.org/10.1063/1.470117 (1995).
59. Sousa da Silva, A. W. & Vranken, W. F. ACPYPE - AnteChamber PYthon Parser interface. *BMC Research Notes* **5**, 1–8, doi: 10.1186/1756-0500-5-367 (2012).
60. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a General Amber Force Field. *J Comput Chem* **25**, doi: 10.1002/jcc.20035 (2004).
61. Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem* **23**, 1623–1641, doi: 10.1002/jcc.10128 (2002).

Acknowledgements

H.B. gratefully acknowledges the financial support from the Cancer Council Queensland (ID1080845). The MX1 beamline scientists at the Australian Synchrotron, Victoria, Australia are acknowledged for their support during X-ray diffraction data collection.

Author Contributions

H.B. directed the research study. M.H.B. performed experimental including protein expression, purification, crystallisation and structure determination, with advice from X.Y. and H.B. Y.Z. generated the clone of galectin-8N. M.H.B., X.Y. and H.B. undertook analysis of the structures and prepared the manuscript. All authors subsequently reviewed and contributed to the manuscript.

Additional Information

Accession codes: Protein Data Bank: Atomic coordinates and structure factors have been deposited with accession codes for galectin-8N CRD with bound lactose (5T7S), LNT (5T7T), LNnT (5T7I) and glycerol (5T7U).

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Bohari, M. H. *et al.* Structure-based rationale for differential recognition of lacto- and neolacto- series glycosphingolipids by the *N*-terminal domain of human galectin-8. *Sci. Rep.* **6**, 39556; doi: 10.1038/srep39556 (2016).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016

2.4 Appendix

SUPPLEMENTARY INFORMATION

Structure-based Rationale for Differential Recognition of Lacto- and Neolacto- series Glycosphingolipids by the *N*-terminal Domain of Human Galectin-8

Mohammad H Bohari¹, Xing Yu¹, Yehiel Zick², Helen Blanchard^{1*}

¹Institute for Glycomics, Griffith University, Gold Coast Campus, 4222, Australia.

²Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel.

* To whom correspondence should be addressed: Associate Professor Helen Blanchard

Institute for Glycomics, Griffith University, Gold Coast Campus, 4222, Australia.

Phone: +61 7 555 27023; Fax: +61 7 555 28098 ; E -mail: h.blanchard@griffith.edu.au

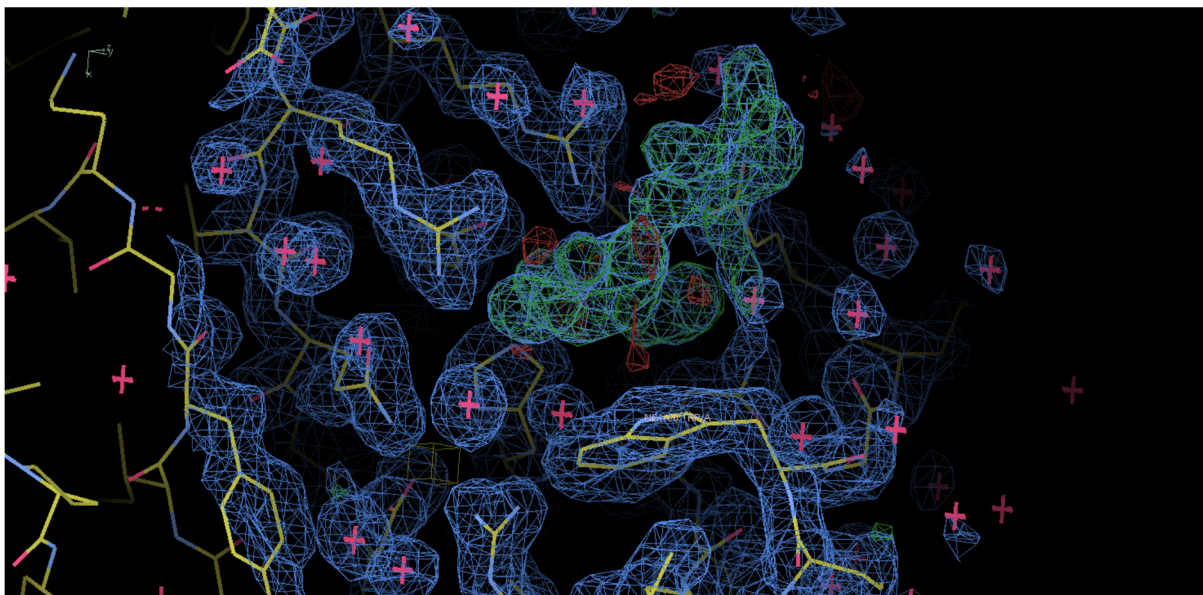


Figure 2.1: Omit electron density maps calculated from refinement with the lactose omitted from the model ($2mF_o - DF_c$: 1.0σ [blue], $mF_o - DF_c$: $\pm 3.2 \sigma$ [green/red]) in the galectin-8N-Lactose complex. Red crosses indicate water molecules.

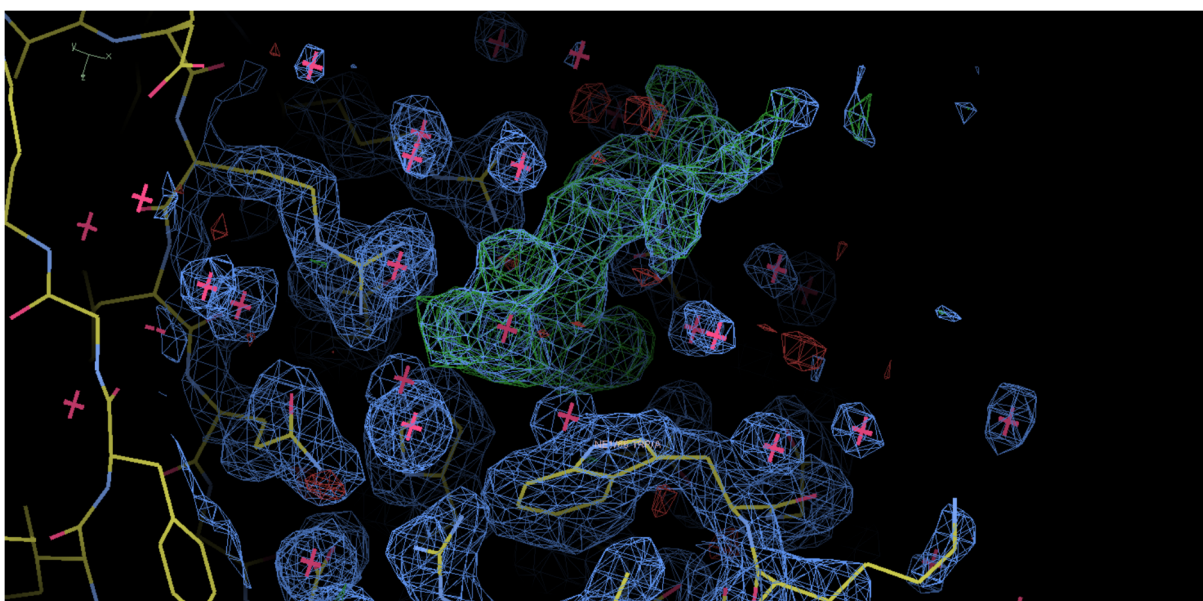


Figure 2.2: Omit electron density maps calculated from refinement with the LNT omitted from the model ($2mF_o - DF_c$: 1.0σ [blue], $mF_o - DF_c$: $\pm 3.2 \sigma$ [green/red]) in the galectin-8N-LNT complex. Red crosses indicate water molecules.

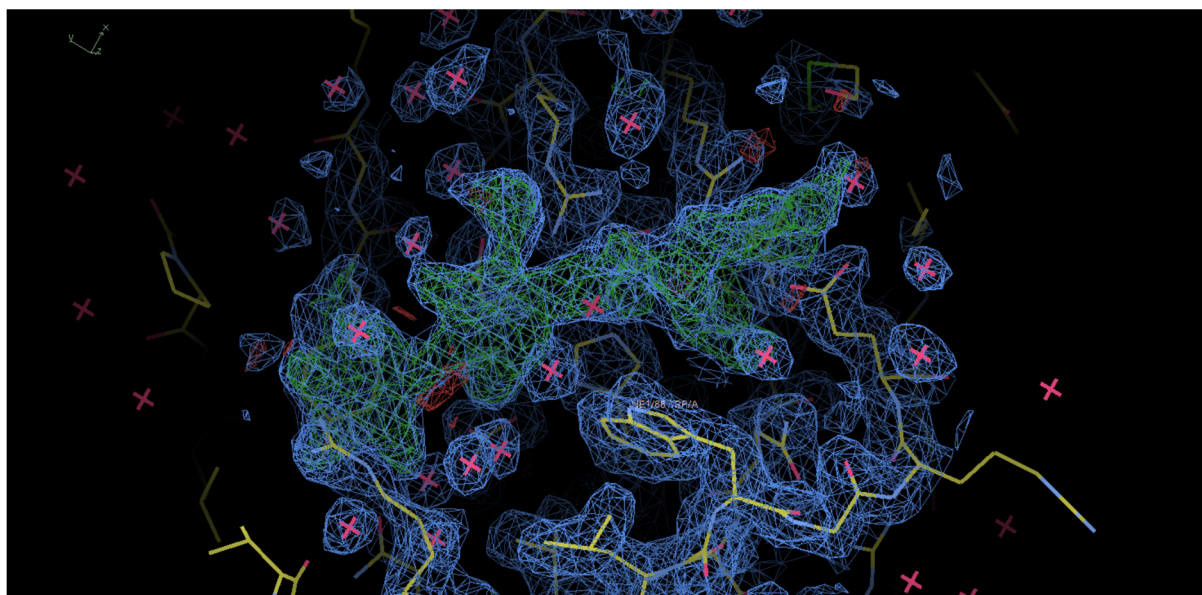


Figure 2.3: Omit electron density maps calculated from refinement with the LNnT omitted from the model ($2mF_o - DF_c$: 1.0σ [blue], $mF_o - DF_c$: $\pm 3.2 \sigma$ [green/red]) in the galectin-8N-LNnT complex. Red crosses indicate water molecules.

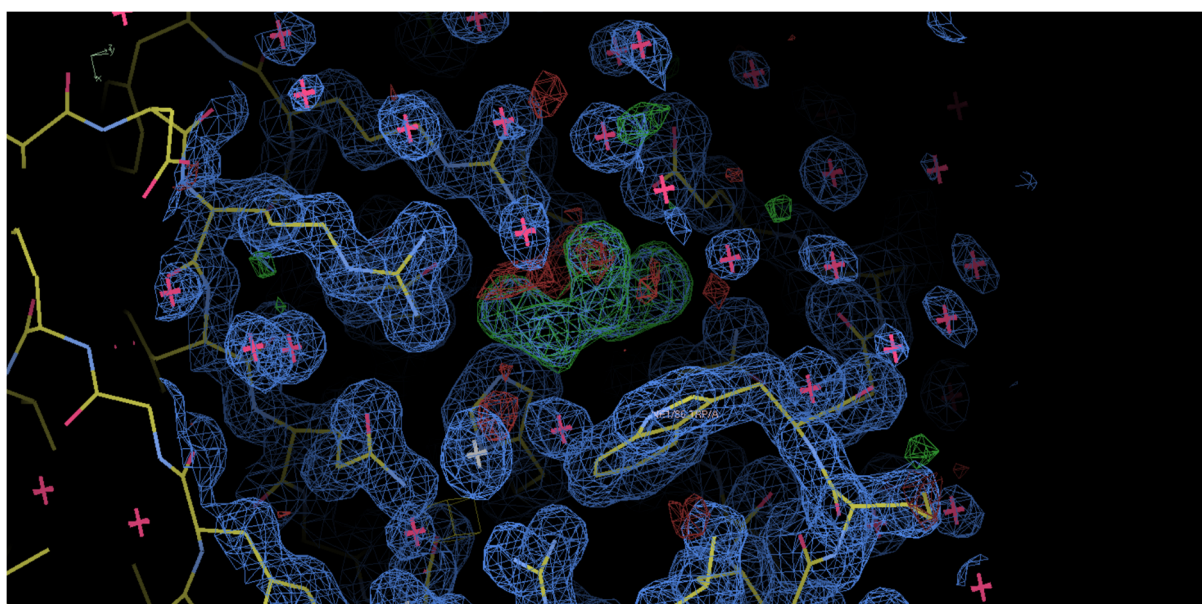


Figure 2.4: Omit electron density maps calculated from refinement with the glycerol omitted from the model ($2mF_o - DF_c$: 1.0σ [blue], $mF_o - DF_c$: $\pm 3.2 \sigma$ [green/red]) in the galectin-8N-glycerol complex. Chloride ion indicated by white cross, water molecules by red crosses.

2.4.1 MD simulation submission script:

```
#!/bin/bash -l
#PBS -N lnt001
#PBS -l walltime=200:00:00
### Number of nodes: Number of CPUs: Number of threads per node
#PBS -l select=4:ncpus=6:mem=200mb:mpiprocs=6
## The number of nodes is given by the select =<NUM > above
NODES=4
##$PBS_NODEFILE is a node-list file created with select and mpiprocs
options by PBS
### The number of MPI processes available is mpiprocs * nodes
NPROCS=24

export I_MPI_PLATFORM=auto
export I_MPI_DEBUG=100
export I_MPI_MPD_RSH=ssh

# This job's working directory
echo "Working directory is $PBS_O_WORKDIR"
cd /export/home/s2874507/scratch/Gal8/manu-simul/lnt/non-hypo

source $HOME/.bashrc
module load gromacs/4.5.5-intel-mpi

echo "Starting job"
echo Running on host `hostname`
echo Time is `date`
echo Directory is `pwd`
#echo This jobs runs on the following processors:
echo `cat $PBS_NODEFILE`

##### PREPare#####
cat << EOF >| his
6
1
0
1
1
1
1
1
EOF
pdb2gmx -f LNT-hypo-prot.pdb -o prot.pdb -p prot.top -i prot.itp -
his -water tip3p < his

sed -i /ENDMDL/d prot.pdb
sed -i /TER/d prot.pdb
sed -i /REMARK/d *_NEW.pdb
cat prot.pdb *_NEW.pdb > prot-lig.pdb
sed -i ".bak" '21i ; Include LNT-hypo topology\n#include "LNT-
hypo_GMX.itp"\n' prot.top
sed -e "\$a LNT-hypo          1" prot.top>>prot-lig.top

editconf -f prot-lig.pdb -o conf.pdb -bt triclinic -d 0.9 -c
genbox -cp conf.pdb -cs spc216.gro -p prot-lig.top -o genbox.pdb

grompp -f em.mdp -c genbox.pdb -p prot-lig.top -o ion.tpr
```

```

echo "15"|genion -s ion.tpr -o ion.pdb -nn 3 -p prot-lig.top

grompp -f em.mdp -c ion.pdb -p prot-lig.top -o em.tpr
mdrun -v -deffnm em

grompp -f pr.mdp -c em.gro -p prot-lig.top -o pr.tpr
mdrun -v -deffnm pr

grompp -f md.mdp -c pr.gro -p prot-lig.top -o md.tpr

##### RUN PRODUCTION #####

mpirun -f $PBS_NODEFILE -n "$NODES" -r ssh -n "$NPROCS" env
PATH=$PATH env LD_LIBRARY_PATH=$LD_LIBRARY_PATH mdrun -np $NPROCS -
deffnm md

echo "Done with job"

```

2.4.2 MD results processing and analysis script

```

#!/bin/bash
cat << EOF >| ndx.txt
1|13
q
EOF
make_ndx -f conf.pdb -o new.ndx < ndx.txt

echo "0"| trjconv -s md.tpr -n new.ndx -f md.xtc -o md_whole.xtc -
pbc whole -ur compact
echo "0"| trjconv -s md.tpr -n new.ndx -f md_whole.xtc -o
md_nojump.xtc -pbc nojump -ur compact
echo "14 14"| trjconv -s md.tpr -n new.ndx -f md_nojump.xtc -o
md_center.xtc -pbc mol -ur compact -center
echo "14 14"| trjconv -s md.tpr -n new.ndx -f md_center.xtc -o
md_fitted.xtc -ur compact -fit rot+trans
rm md_whole.xtc md_nojump.xtc md_center.xtc

echo "1"| g_rmsf -s md.tpr -f md_fitted.xtc -n new.ndx -res
echo "1 1"| g_rms -s md.tpr -f md_fitted.xtc -n new.ndx -o prot-
rms.xvg
echo "4 4"| g_rms -s md.tpr -f md_fitted.xtc -n new.ndx -o bb-
rms.xvg
echo "13 13"| g_rms -s md.tpr -f md_fitted.xtc -n new.ndx -o lig-
rms.xvg

echo "14"| trjconv -s md.tpr -n new.ndx -f md_fitted.xtc -o
traj_all.pdb
echo "14"| trjconv -s md.tpr -n new.ndx -f md_fitted.xtc -o
traj_50.pdb -skip 10

```

2.4 Further discussion

Atomic level understanding of interactions of galectins with their natural binding partners is indispensable towards elucidating the physiological functions of galectins and ligand design. The distinct functionality observed for each galectin despite sharing conserved sequence motifs hints for the specificity in glycan recognition. This specificity in the case of galectin-8N and other galectins is mediated by some unique residues mostly located in the extended carbohydrate binding site. To this end, crystallographic structures of the isolated CRDs of the tandem-repeat galectin-8 in complex with natural ligands have been reported. The galectin-8N was mainly responsible for recognising SM3 (sulfatides) and GM3 (sialic acid-containing) glycosphingolipids, as revealed from its preferential interaction of unique residues with sulfated and sialylated glycans [42, 76]. Galectin-8N-LNFIH complex revealed van der Waal's interactions between the galactose and Tyr141 and the fucose ring attached on the C3 of GlcNAc does not form hydrogen bonds with the protein. However, the galactose ring is stabilized between the Tyr141 and fucose ring via hydrophobic interactions, each ring separated by ~ 4 Å distance [76]. These hydrophobic interactions contribute towards increased binding affinity of LNFIH (3.3 μ M) compared to LNT (13 μ M) [42], whilst, identical interactions were observed for the tetrasaccharide LNT in galectin-8N-LNT complex [81].

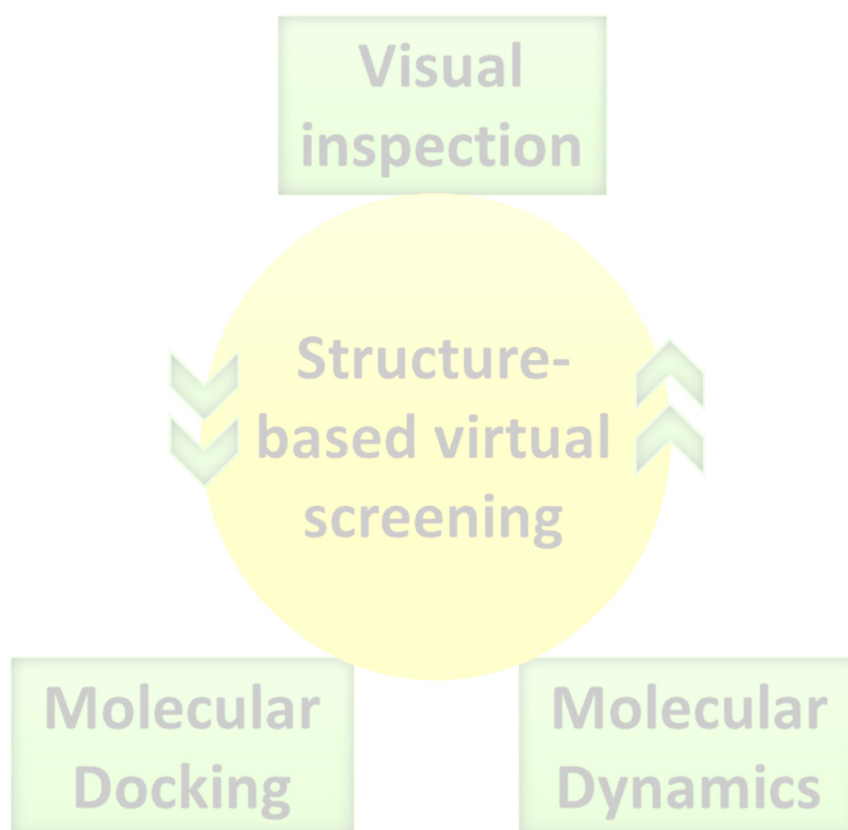
Binding of larger oligosaccharides have been structurally investigated to other tandem-repeat members of galectins such as galectin-4 and galectin-9. Galectin-4 exhibits specificity for sulfated glycans like that being observed with other galectins including galectin-8 [42] but does not recognise sialylated glycosphingolipids unlike to galectin-8 [109]. The galectin-4N-LNT crystal structure (4YM0; [101]) showed outwards pointing of the non-reducing end galactose of LNT as being noted for the galectin-3-LNT complex (4LBM; [110]) which is unlike to the terminal disaccharide binding noted in our galectin-8N-LNT structure (5T7T; [81]). Out of the four monomers in the asymmetric unit of galectin-4C-LNT complex, lactose was found in the monomer C while in the monomer D the non-reducing end disaccharide of LNT was observed. However, the crystal packing resulted in varied occupancy of the binding site in monomer C and D and favoured soaking of LNT only to monomer A and B [101]. The crystal packing, on the other hand, did not affect the soaking of LNT in galectin-8N, where the presence of unique residues such as Tyr141 (in galectin-8N) on strand S2 unlike to Gln313 in galectin-4C was responsible for binding of LNT [81]. In contrast, all the four units of the tetrasaccharide LNT interacted with galectin-4N (4YLZ; [101]) as it was noted for galectin-3 (4LBN; [110]) and galectin-8N (5T7I; [81]) resulting from the relative higher affinity of LNT to galectins as compared to LNT. There are no reports of the interaction of these

tetrasaccharides with galectin-9. The LacNAc dimers that structurally resembles LNnT have been complexed with galectin-9N, where the residue Asn137 equivalent to Tyr141 of galectin-8N was noted to increase the binding affinity poly-LacNAc units towards galectin-9N CRD [111].

The galectin-8N-glycerol complex revealed minimum atomic features required by a ligand to be recognised by a galectin. Importantly, this provides potential clues for a non-galactose (non-carbohydrate) molecule binding to galectin-8N. Taking the minimum atomic feature into account, in chapter 3, structure-based virtual screening to identify non-carbohydrate based molecules as binders of galectin-8 will be carried out. Also, these minimum features will be employed to derive a monosaccharide galactose-based scaffold as potential inhibitor of galectin-8. Our MD simulations performed on various complexes in the study also supports the suitability of parameters and topology generated for simulations, as they corroborate well the experimentally observed conformations and interactions in the crystal structures. Overall, these simulations formed the basic methodology that was employed throughout the thesis for either rank ordering compounds or investigating binding mode and interactions of various protein-ligand complexes.

Chapter 3

Structure-Based Virtual Screening for Identification of Non-Carbohydrate-Based Ligands Targeting Galectin-8



3.1 Introduction

As discussed in Chapter 1 (section 1.6), galectin-8 is involved in the regulation of bone remodelling process where its inhibition could lead to a potential new approach in tackling diseases associated with bone-loss [94, 96]. Bearing this in mind and the preferential recognition of anionic saccharides [42, 76], the search for novel binders of galectin-8 was initiated. In this Chapter, structure-based virtual screening was performed for investigating novel non-carbohydrate-based ligands targeting the galectin-8 N . The non-carbohydrate-based hydrophobic molecules offer some advantages such as metabolic stability and crossing of cellular barriers over the carbohydrate-based molecules. The virtual screening strategy involved performing iterative molecular docking and molecular dynamics simulations to narrow down a customised library of commercially available compounds. The rank-ordering and filtering of compounds were done mainly based on retention of key binding site interactions. The top fraction from the screened library was purchased and evaluated for binding to galectin-8 N by STD NMR and X-ray crystallography.

3.1.1 Strategy for virtual screening

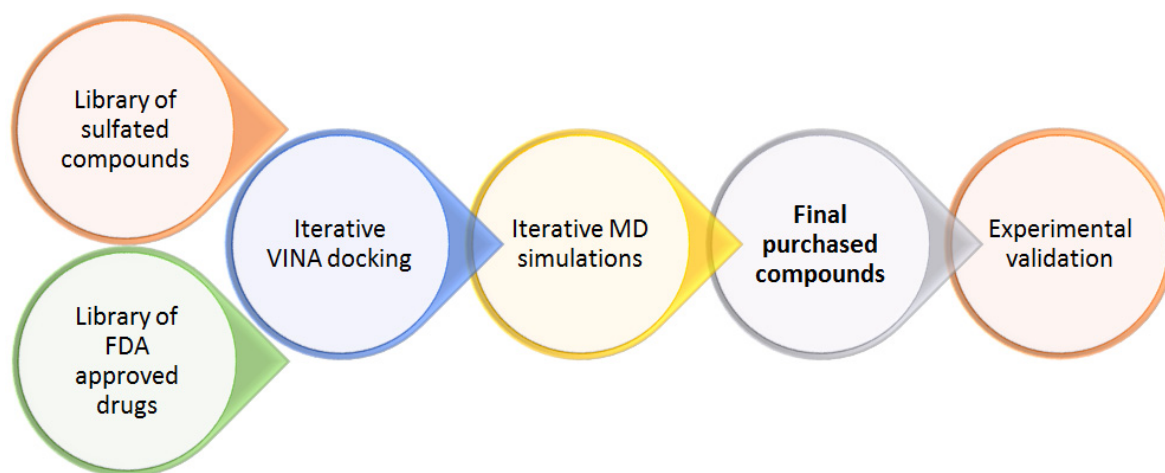


Figure 3.1: A general theme of the virtual screening protocol.

The success of any structure- or ligand-based virtual screening campaign relies primarily on the strategy employed for filtering the compounds in hand. The ligand design area in the field of galectin-8 is at a nascent stage, and there are no specific reports in the literature focussing on designing ligands against galectin-8, particularly any non-carbohydrate-based molecules. With no currently available galectin-8 specific ligands, the use of ligand-based approaches for identifying novel molecules against galectin-8 would be less feasible. There exists structural information of the galectin-8 N , the galectin-8 C isolated domains and the truncated full-length galectin-8 either in *apo* form or bound to their natural glycans in the

protein data bank (Table 1.1, Chapter 1). Furthermore, with the structural biology resources at our disposal, a structure-based approach was followed to identify novel ligands against galectin-8 (Figure 3.1).

From the glycan array and other binding studies, it has become apparent that galectin-8N is a major contributor towards the recognition of glycans by galectin-8 [42, 44]. The preferential recognition of anionic oligosaccharides by galectin-8 is attributed to unique binding site residues of galectin-8N CRD. Structural insights from the reported ligand bound galectin-8N complexes draws us to significant interaction profile that potentially contributes to the ligand affinity and specificity. With the availability structural information, a wide-ranging glycan binding profile and interesting hot-spots in the binding site, galectin-8N was used as the target protein in our screening campaign [42, 44, 76]. A similar approach, though not in the scope of the present work, can be followed for identifying ligands against galectin-8C.

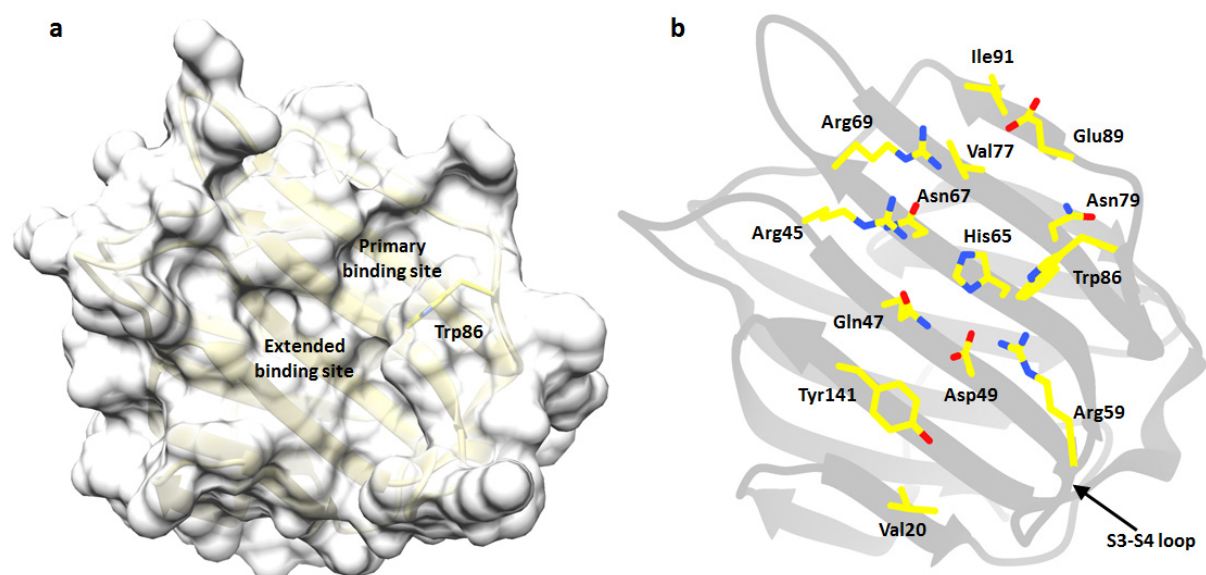


Figure 3.2: Overview of the galectin-8N carbohydrate recognition domain. a: The CRD (yellow ribbons) showing the carbohydrate binding face of the jelly-roll and the primary and extended binding regions on the concave binding surface of the CRD. b: Depicts the amino acid residue (yellow carbon, oxygen red, nitrogen blue; stick representation) involved in glycan binding interactions.

Overall, the topology and features of a galectin CRD are previously discussed in Chapter 2 (Figure 3.2). Briefly, the conserved interactions formed between the galectin-8N-lactose and galectin-8N-glycerol complex reveals basic atomic framework required by any binder of galectin [81]. Interactions include hydrogen bonds with His65, Asn67, Arg45, Arg69, Asn79, Glu89 and the galactose ring partly stacking against the evolutionarily conserved amino acid Trp86 (Figure 3.3). These interactions have been observed for glycans not only with galectin-8 but also with all other galectins. Therefore, it becomes apparent that most of these conserved interactions are required for ligand recognition by a galectin. The first filtering

criteria was designed wherein the occupancy of primary binding site (galactose binding site) would ensure interaction with some of the conserved amino acids including the hydrophobic interactions with Trp86. Those compounds in the library that are occupying the primary binding site will be retained in the screening tier.

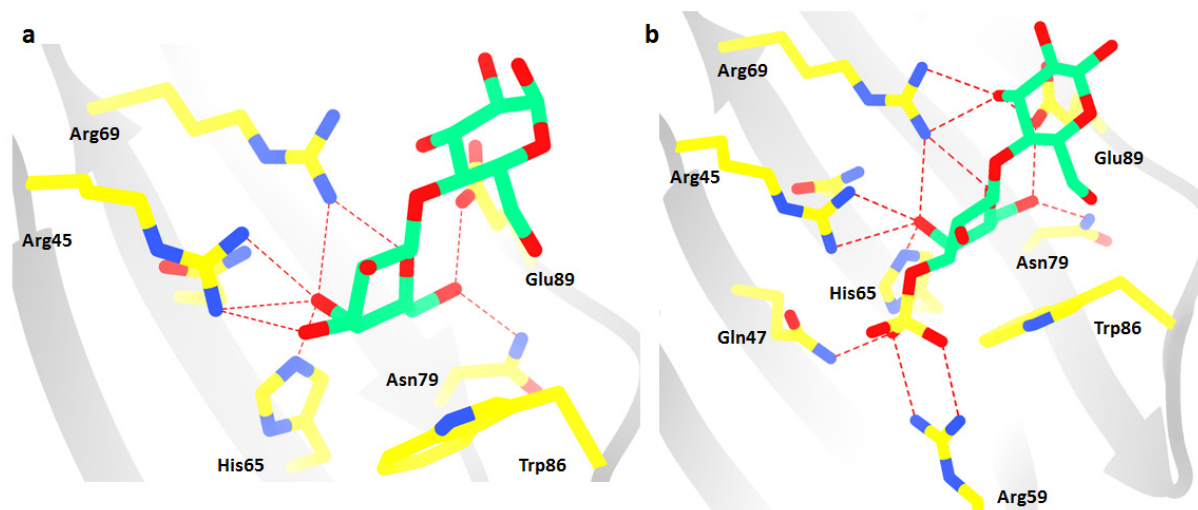


Figure 3.3: Interactions made by glycans with the carbohydrate binding site of galectin-8N. (a) Highlights the hydrogen bonds (red dashed lines) formed upon interaction with lactose. (b) Hydrogen bonding interactions made by 3'-sulfated lactose, of note, are the ionic interactions between the sulfate group and the unique Arg59.

The most important aspect of a ligand design campaign is addressing the target specificity, particularly when various isoforms exist. Given 15 members known so far within the galectin family, some of which have distinctive opposite biological effects, specificity is the main concern. Furthermore, with the conserved galactose recognition site, cross-galectin binding would be very likely. For this purpose, exploring the unique features found only in the galectin-8N CRD was required. The long S3-S4 loop bearing an arginine (Arg59) is one of the unique features not observed in any other galectin (Figure 3.2). This residue along with Gln47 was structurally demonstrated to interact with the anionic group (sulfate/carboxylate group) at the 3'-position of galactose, particularly via strong ionic interactions (Figure 3.3) [76]. The outcome of these ionic interactions was increased binding affinity of galectin-8N towards anionic sugars, a key preferential aspect specific to galectin-8. Bearing the unique residues and the involved ionic interactions in mind, the second filtering criteria was designed (specificity filter), wherein the possibility of exploiting these residues for interactions was assessed.

As for other galectins like galectin-3, the inhibitor design campaigns were more based on the native ligand such as galactose/lactose as the template molecule [112-114]. Notably, a structure-based approach led to the design and development of a novel talose-based scaffold, that explored a region in the galectin-3 binding site not observed in many galectins [18, 108]. During assessing selectivity of inhibitors towards galectin-3, binding of ligands was evaluated

against a panel of galectins including galectin-8N. Since these ligands were designed to better explore the subsites of galectin-3 with some core elements, it showed poor affinities towards galectin-8 and therefore were not suitable templates for initiating ligand design targeting galectin-8.

However, due to high polarity of the carbohydrate-based inhibitors crossing the lipophilic cell-barriers becomes challenging and likely gets metabolised in the body. There are various thiol-based carbohydrates with aromatic rings developed to deal with polarity and stability of the carbohydrate-based inhibitor [115-118]. Nevertheless, the aim for this project was to investigate non-carbohydrate small molecules as binders of galectin-8N. These compounds offer some advantages such as metabolic stability and crossing of cellular barriers over the carbohydrate-based molecules.

3.1.2 Library design

The concept of virtual screening revolves around rationally identifying potential inhibitors against a target protein. The library of compounds used for screening using any design approach plays a critical role in the overall success of a screening campaign. A wealth of information reporting novel compounds and their biological activity, including both *in vitro* and *in vivo* data, is available in the literature. Accordingly, there are various web-based interfaces curating and making it available to download the chemical compound information obtained from the literature. PubChem, for example, is an open database that includes chemical substance information, chemical structures, and their bioactivity in three primary databases viz. Pcsubstance, Pccompound and PCBioactivity respectively, including the BioAssay description with literature reference and assay data points [119]. ChEMBL is a freely available, manually abstracted database containing binding, functional and ADMET information for bioactive compounds [120]. BindingDB contains literature extracted binding affinity information for potential drug targets. [121]. Further, a database like Binding MOAD [122], PDBind [123] provide binding affinity information for macromolecule-ligand complexes reported in the protein data bank. DrugBank [82] provides complete annotation of chemical structure, mechanism of action and other pharmacological targets for all the FDA approved drugs. ZINC [124, 125] is a free database of commercially available and easily accessible compounds in a ready-to-dock 3D format for virtual screening experiments. These databases prove to be a critical guide in decision making while carrying out drug discovery analysis.

The broad glycan binding profile of galectin-8N with a notable preference for anionic saccharides through interaction with the unique Arg59 (on S3-S4 loop, Figure 3.2) was at the center of library design [42]. Therefore, the lead-like (molecular weight range of 200-350 Da)

[126] purchasable sulfur compounds containing a R-SO₂-R group such as sulfates, sulfones and sulfonamides from the ZINC database was chosen as our first library, and the second library being the FDA-approved small molecule drugs. The positive hits from the lead-like library could potentially be considered for optimisation without exceeding the drug likeliness rule of the Lipinski [127, 128]. Another advantage with the lead-like and FDA drug library is its commercial availability that indicates the synthetic feasibility of the core scaffold.

The FDA approved compound library was chosen for the screening process to investigate the idea of drug repositioning. Finding a new application for an approved drug has been successfully applied in some cases and hold great potential as far as the reach to market is concerned [129, 130]. Because the drugs have already gone through pre-clinical and clinical stage toxicity testing in humans, the clinical investigation for the new application would be quicker and cheaper as compared to performing those tests for a new compound. With the biotech evolution, there are macromolecular drugs including monoclonal antibodies and complex natural products that are being approved by FDA, but considering the scope of current work, any drug molecule with a molecular weight of more than 500 was excluded.

The aim was to perform structure-based virtual screening for identification of non-carbohydrate based binders of galectin-8*N*. Iterative molecular docking and MD simulations were performed on a specific library of compounds. The rank ordering and filtering of the compounds were performed based on interaction-based criteria designed to explore affinity and specificity. The final best-fitting compounds were purchased and were evaluated for binding to galectin-8*N* through experimental techniques. These include STD NMR and X-ray crystallographic techniques. A positive outcome from this work would be the first-ever evidence of a non-carbohydrate binder of galectin and will be evaluated biologically.

3.2 Methods

Virtual screening was performed by undertaking iterative molecular docking and molecular dynamics simulations, with visual analysis of protein-ligand interactions. The UCSF Chimera package [131] was used for visualisation of docked poses and interaction analysis throughout the work. A combination of interaction with the unique galectin-8*N* residue (Arg59) and the evolutionarily conserved residue (Trp86) was used for filtering the library of compounds. The assessment of interactions constituted i) the occupancy of the primary binding site by the ligand potentially indicate hydrophobic (CH- π type or π - π type) and/or polar interactions with the evolutionarily conserved Trp86 and other conserved residues in the binding site and ii) hydrogen bonding interaction with the unique Arg59 located on S3-S4 loop. Off note, the interaction between sulfated compounds in the library of Arg59 were more of ionic nature like salt bridge interactions in place of hydrogen bonding interaction in case of other lead-like compounds in the library and were therefore assessed respectively. The interaction made with Trp86 could be of a polar nature such as hydrogen bonds involving the backbone amide group or with the nitrogen of the side chain in pyrrole ring. In addition, non-polar type interactions are also possible such as hydrophobic interaction involving the π -electron cloud of the Trp86 side chain and either aliphatic or aromatic rings of the ligands. Some leverage was given to compounds that occupied the primary binding site, but the sulfate group (or any other ionic group) was not placed in a geometrically favoured position to interact with Arg59. All the computational calculations were performed on High-Performance Computing Cluster facility “Gowonda” at Griffith University with Intel Xeon CPU X5650 @2.67GHz processor, while analysis was undertaken on a local computer.

3.2.1 Molecular docking

AutoDockVina (Vina) [132], a freely available application, was used for performing docking calculations. This program automatically calculates the grid maps and uses a machine learning-based scoring function to rank order the docked poses. It is much faster in performing a calculation than the parent software AutoDock and is, therefore, more suitable for virtual screening applications. Galectin-8*N*-3'-*O*-sulfated lactose crystal structure (3AP6 [76]) was chosen for our screening protocol mainly because the aim was to identify ligands that had a similar topology to one of the best affinity natural ligands of galectin-8*N* (2.7 μ M [42]). Furthermore, the 3'-*O*-sulfated lactose exhibits strong binding affinity towards galectin-8*N* and therefore forms a good starting point to investigate potential inhibitors.

Protein: This protein preparation step was done using AutoDockTools (ADT) implemented in MGLTools package [133]. The protein was stripped of non-protein atoms and

saved into the `pdbqt` format. The efficiency of Vina in predicting the experimentally observed binding conformation was evaluated through cognate ligand docking. The grid for cognate docking was defined using the grid options menu of ADT, centred on the co-crystal ($x = 24.069$, $y = 29.662$, $z = 23.165$) with 25 points in each dimension and 1 Å spacing. The ligand file was converted into `pdbqt` format after defining torsion centers, rotatable bonds and applying gasteiger charges [134]. The docked pose was compared with the X-ray conformation for evaluation purpose.

Ligand library: From the various readily available categories of compounds, lead-like library containing R-SO₂-R group was obtained through substructure search and were downloaded using the batch script files (in `mol2` format). While the already available ready-to-download special subset (Zdd) of Zinc drug database of commercially available approved drugs was directly downloaded (in `mol2` format). The `mol2` ligand files were converted into `pdbqt` using Raccoon [135], while the protein `pdbqt` and the grid dimension used for cognate docking were used for screening. The Raccoon was also used to generate virtual screening master-script that was subsequently modified (`vs_submit.sh`) for performing automated virtual screening in Vina (Appendix 3.6.1).

Analysis: Vina docking scores from the output files were extracted using `analyse.sh` script (Appendix 3.6.2) and added into a master file. Each docking calculation was performed five times and the compounds consistently getting a score over the set cut-off (-6.5 kcal/mol) were then retained for the next tier of analysis. The compounds retained in the score-threshold stage were subjected to interaction-based filtering performed through visual inspection. Compounds satisfying the interaction-criteria have been submitted to the next tier of molecular dynamics simulations.

3.2.2 Molecular dynamics

All simulations were performed using GROMACS package version 4.5.6 [136] with in-built AMBER99SB-ILDN force field [137], similar to that used in Chapter 2 and other galectin-ligand simulations carried out in the Blanchard group [101, 107, 110]. Briefly, after creating a cubic solvent box, water molecules modelled through TIP3P solvent model [138] was arranged 10 Å around the protein surface. Initial 5000 steps of the steepest descent energy minimisation were followed by 15000 steps of position-restrained minimisation before going for the production run. The parameter and topology for ligands were generated using `acpype` [139] employing Generalised Amber Force Field [140] and AM1-BCC charges [141]. A bash

script was used to prepare the input files and generate topology and parameters for the protein, and ligand and for running simulations (Appendix 2.4.1 and 2.4.2).

For filtering compounds, initially, a short 5 ns simulation was performed on the galectin-8*N*-ligand complex where the starting conformations were the corresponding docking output conformation. The production run trajectory was visually analysed using VMD. Here retention of ligand in the binding site throughout the length of simulation was assessed. Mean square displacement analysis (MSD) through gromacs' utility `g_msd` was used for analysing retention of ligand in the binding site. To further overcome the chance outcome, a triplicate of short simulations was performed, and compounds that stayed in the binding site maintaining the interaction criteria were selected for long simulations. Long simulations were run for 100 ns, using the protocol used for short simulations. The ligands that consistently stayed in the binding site maintaining the overall key interactions then formed the final leads, and the best compounds among them were purchased and experimentally evaluated for binding to galectin-8*N*.

3.2.3 Saturation Transfer Nuclear Magnetic Resonance spectroscopy

All STD NMR spectra were acquired on Bruker 600 MHz Avance spectrometer with a conventional $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ gradient cryoprobe system at 298 K. The on-resonance and off-resonance spectra were recorded, and the data analysis was performed using TopSpin 3.5pI5 software package. The purchased ligands were used in varying concentrations, 500 μM - 5 mM with galectin-8*N* (5 μM), in 250 μL of deuterated buffer containing 20 mM phosphate buffer and 20 mM sodium chloride. The protein was saturated (on-resonance) with a cascade of 40 Gaussian-shaped pulses with a duration of 50 ms each at -0.1 ppm (on-resonance frequency) and 33 ppm (off-resonance frequency), totalling to saturation time of ~ 2 s.

3.2.4 Ligand soaking

Galectin-8*N*-*apo* crystals, obtained as previously described in [81], were used for soaking the purchased ligands. Ligand soaking time was varied from 10 minutes to overnight with a range of ligand concentrations containing about 5% DMSO as ligand solubility enhancer. Crystals were cryo-cooled with glycerol as a cryoprotectant, and the X-ray data was remotely collected at Australian Synchrotron (MX2 beamline) using Blu-Ice software [142] at 100 K with a wavelength of 0.9537 Å, and ADSC Quantum detector. The structures were solved using the method specified in [81]. The data was integrated using iMOSFLM [143], and the point group determination and scaling of the data was performed using AIMLESS [144]. The phases were solved using Phaser [145] with the galectin-8*N*-*apo* structure (3AP5 [76]) as

the search model. The model obtained was refined using REFMAC5 [146, 147] in CCP4 program suite [148]. Visualisation and model building was done in Coot [149].

3.3 Results and Discussion

3.3.1 Efficiency evaluation

The performance of a docking software varies considerably with changing macromolecule, ligand or algorithm parameters [150, 151]. As a measure of the ability of Vina's docking algorithm to reproduce the experimental binding conformation, cognate docking was performed. Cognate docking involves extracting the ligand from the crystal structure coordinate file, docking it back again in the same protein structure and comparing the deviation of the docked conformation from the experimentally observed conformation. Correctly predicting experimental binding conformation is critical as the efficiency of docking algorithm decreases with increasing rotatable bonds of ligand and therefore, challenges the conformation sampling efficiency of the docking program [150, 151].

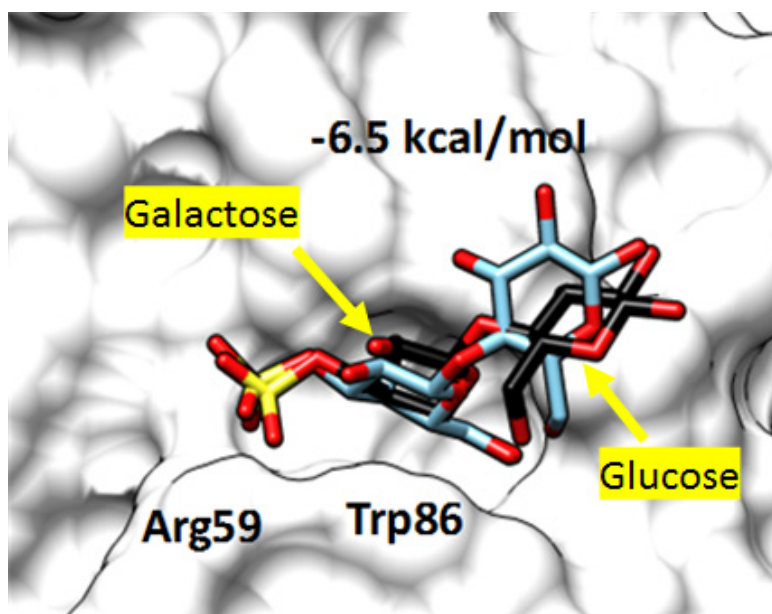


Figure 3.4: Efficiency evaluation through cognate ligand docking. Superimposition of lactose-3'-SO₃ docked pose (carbon colored black) with X-ray conformation (3AP6, carbon colored blue).

Carbohydrates are cyclic aliphatic molecules that exhibit high flexibility stemming from the glycosidic linkage joining the monosaccharide units. The linkage creates an array of orientations possible for example for the two units in disaccharide, however carbohydrates populate into only a specific set of orientation [152]. Therefore, the docking algorithms may not be able to efficiently sample those specific orientation, this in addition to limitation of capturing the hydrophobic interactions are potential bottlenecks in protein-carbohydrate docking [153, 154]. Cognate docking was performed on the galectin-8N-3'-O-sulfated lactose (3AP6 [76]) crystal structure to assess sampling efficiency of Vina towards a carbohydrate molecule. The ligand 3'-O-sulfated lactose was extracted from the coordinate file and docked

back into the galectin-8N binding site (section 3.2.1). The calculation produced the docking output conformation that was assigned a score of -6.5 kcal/mol by the Vina scoring function. Superimposition of the docked conformation on to the experimentally observed conformation revealed identical placement of the galactose ring and the sulfate group of the 3'-*O*-sulfated lactose (Figure 3.4). The placement of the glucose ring of 3'-*O*-sulfated lactose was different in the docked conformation to that experimentally observed, the flexibility of glucose ring is also noted experimentally [81]. Importantly for galectin-8N, recognition of the galactose and preference for the sulfate group is critical, which the docking algorithm correctly positioned and identically placed in the binding site. Based on this cognate docking calculation, the docking score for 3'-*O*-sulfated lactose (-6.5 kcal/mol) was used as an arbitrary cut-off to filter compounds during the initial stage of screening.

3.3.2 Docking and filtering

The screening campaign started with a total of ~9000 compounds that were obtained from the freely available ready-to-dock ZINC database [124]. ZINC database provides ligands that are prepared by the curators and made easily accessible in reasonable geometry, protonation state at the desired pH and tautomeric forms. These molecules are pre-filtered and made available in subsets with various physical properties (such as drug-like, lead-like and fragment-like), removing compounds containing reactive functional groups such as aldehydes and thiols [125].

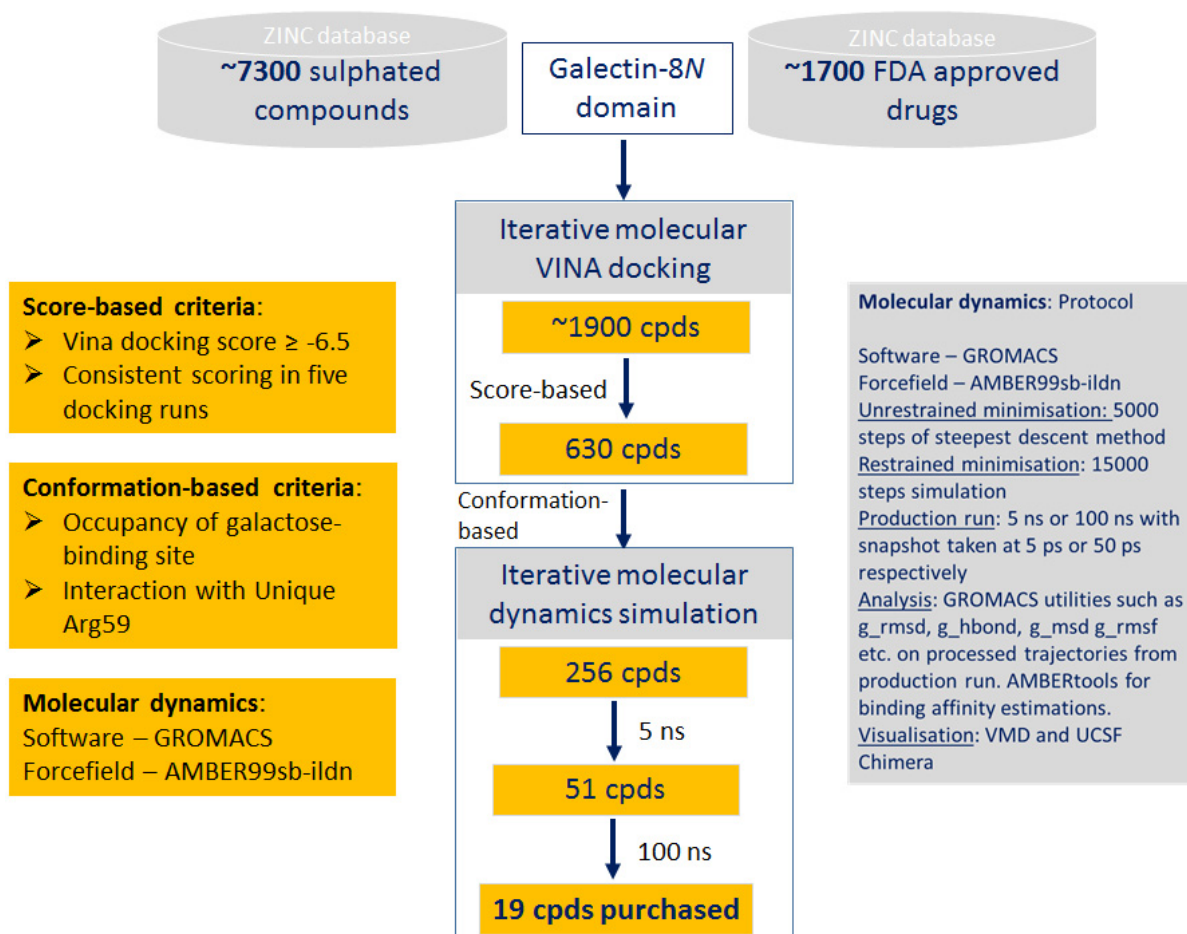


Figure 3.5: The strategy used to perform the structure-based virtual screening.

Herein, a combination of clean lead-like (containing R-SO₂-R group) molecules and FDA-approved drugs library was used for screening. With the lead-like molecules, the advantage being their relatively small-to-medium molecular weight (≤ 350 Da), that would allow some flexibility in accommodating different functional groups at the lead optimisation stage. While any positive hits from the FDA-approved drug library would lead to exploring new uses for old drugs. The general scheme detailing the strategy employed for filtering compounds at different stages of the screening protocol has been illustrated (Figure 3.5).

Docking of the library of compounds was carried out using a similar protocol to that used during cognate ligand docking. Submission, curation and analysing of the docking results were automated using the bash shell scripts (Appendix 3.6.1 and 3.6.2). Compounds with docking-scores less than or equal to that of 3'-O-sulfated lactose (-6.5 kcal/mol) were curated, where the larger negative value (a better score) indicates better binder. This score roughly corresponds to interactions made by one of the strong binders of galectin-8N [42, 44] and hence was arbitrarily used as a threshold for filtering. Of the total starting compounds, score-filtering left with ~1900 compounds. Importantly, to avoid chance correlation, all the docking calculations were carried out five times. In total, 630 compounds that consistently cleared the

score threshold upon iterative docking, and were subjected to the interaction-based filtering tier.

Given that ligands are scored based on van der Waal's and coulombic interactions, employing an arbitrary cut-off would imply retention of some of the key binding site interactions. However, it is not certain that the interactions made by 3'-*O*-sulfated lactose and docked ligands with a score above -6.5 kcal/mol were identical. Therefore, interaction-based (conformation-based) filters were designed which the experimentally observed interactions (in the crystal structure, Figure 3.3) were compared with those made by the docked ligand. As discussed in the previous section, interaction with the conserved Trp86 would roughly assure that the ligand is occupying the primary binding site while interactions with unique Arg59 would allow gaining affinity along with specificity. Visual inspection for the docked pose of 630 complexes was carried out for analysing interactions.

The interaction-based filtering will ensure that compounds will occupy the primary binding site, and suitable functionality will be placed near unique Arg59 for interaction. The compounds from previous tier that cleared the score-threshold screen were subjected to interaction-based filtering. The docked conformation output (pdbqt) from Vina was converted into Chimera readable format (pdb) for analysis. Compounds satisfying both the interaction criteria were retained for the next step of screening, while those not near the conserved Trp86 were excluded as for their unlikeliness to occupy the primary binding site. An example (two compounds from each category) of three possible cases for interaction-based criteria has been demonstrated (Figure 3.6). For certain compounds where they occupied the primary binding site but were not in a geometrically favored position to interact either through salt bridge interaction or hydrogen bonding with Arg59 were also retained for next steps of filtering (in blue rectangle; Figure 3.6). This leverage was given based on the consideration that the protein was held rigid during a docking run, which is not correct in regards to the true biological environment. The interaction-based filtering resulted in a total of 256 compounds that satisfied the interaction criteria and were therefore taken forward to the next tier of screening. Overall, this analysis would ensure the engagement of both conserved and unique binding site residues in interactions with the filtered compounds.

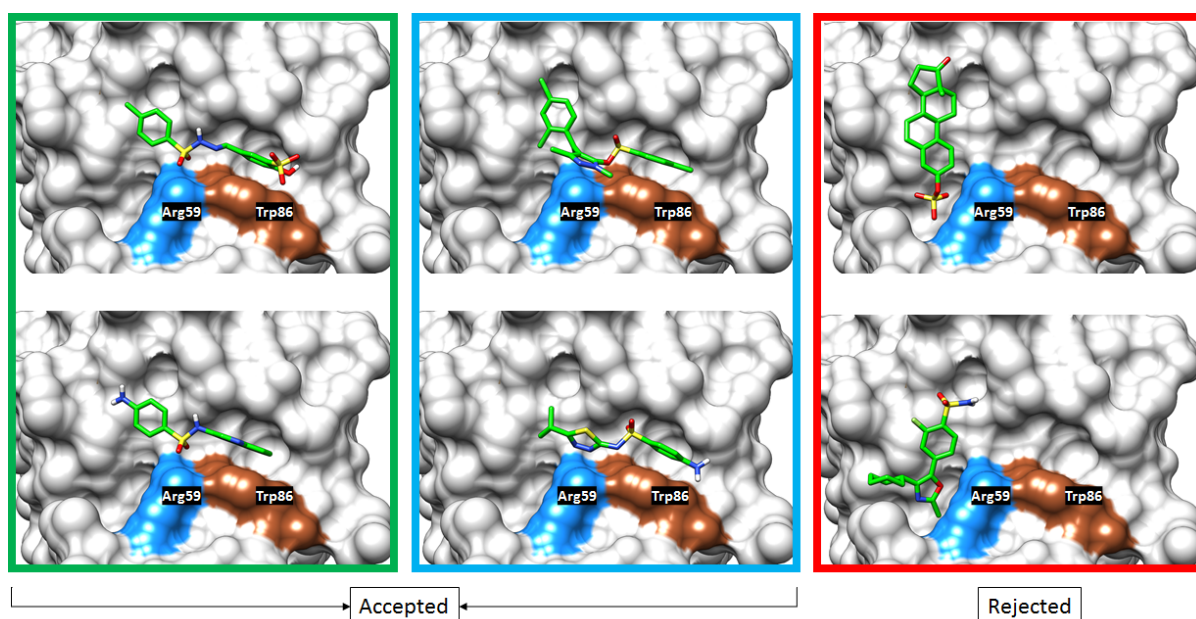


Figure 3.6: An example of conformation-based (interaction-based) criteria. Galectin-8N is represented in white surface and docked ligand in green sticks. For reference, evolutionarily conserved Trp86 is coloured in brown while the unique Arg59 is coloured in blue. Green box (on left) show two compounds that satisfied both interaction criteria set viz. being placed in a primary binding site and were interacting with the unique Arg59. Blue box (in middle) show two compounds that partially followed the criteria where the sulfate group (top) and sulphonamide group (bottom) was not placed towards the unique residue. Red box (on the right) show two compounds that did not follow any of the criteria and therefore were rejected during the filtering process.

3.3.3 Short MD simulations

During Vina docking calculations, flexibility is imparted only to the ligand while the protein is held rigid. Although there is an option for allowing flexibility to certain residue side chains during docking, however, explicit solvent MD simulations on the whole protein-ligand complex was employed. MD simulations were carried out on a total of 256 compounds where the starting conformation was the docking output conformation. Parameter and topology for each ligand were generated using *acpype*, that uses antechamber module of Amber package [155, 156]. Input files for MD simulations were prepared using a bash script file (see Appendix).

For analysis of simulation results, MSD which is the mean displacement of atoms from a set of initial positions was employed. The ligand MSD from simulations provides the magnitude of ligand movement during the simulation. Higher displacement would indicate larger movement of ligand implying possible instability of the interactions between the protein and ligand. In contrast, lower value of MSD would imply lesser fluctuations in the ligand conformation and more stable protein-ligand interactions. For setting up an arbitrary MSD threshold, trajectories of simulations for ten randomly selected complexes were visually inspected using VMD. Based on the MSD analysis of these complexes an arbitrary value of ≤ 0.1 was used as cut-off, as being retained in the binding site and hence acceptable. Although

this threshold is highly dependent on the number of rotatable bonds and some compounds may miss out due to the cut-off. However, a major part of the ligand space from the filtered compounds would still be captured, since numerous molecules were simulated. The simulation run starts from a random velocity and therefore causes the outcome of two simulations to be non-identical. A triplicate of short simulations was therefore performed for each complex, to rule out the chance observation. Short 5 ns short simulations might not be enough to observe any major displacement in the placement or conformation of ligand from its starting position. However, these simulations will help narrow down the compound pool by filtering rather unlikely binders before performing computationally intensive long simulations.

Table 3.1: Mean square displacement analysis performed for the lead-like library (far left and centre) and FDA approved drugs (on far right).

	MSD-1	MSD-2	MSD-3		MSD1	MSD2	MSD3		MSD-1	MSD-2	MSD-3
ZINC01562064	0.0856	0.1908	0.1191	ZINC00002106	0.3089	0.306	0.355	ZINC00537791	1.734	0.0529	0.0778
ZINC01599368	0.1268	1.3857	0.6716	ZINC00005803	0.8549	13.813	0.3351	ZINC00537795	2.124	1.778	0.977
ZINC01621433	0.0978	0.4406	0.2447	ZINC00028349	0.3938	0.362	0.2871	ZINC00537804	0.1205	0.1074	0.0668
ZINC01628079	0.3796	0.3881	0.0677	ZINC00029424	0.024	0.2812	0.4135	ZINC00538312	0.1176	0.2024	0.1703
ZINC01653709	0.198	1.0908	0.5595	ZINC00029866	0.172	0.1558	0.2386	ZINC00601316	0.4982	0.8745	1.243
ZINC01670444	0.3825	0.203	0.1978	ZINC00033018	0.0312	0.1247	0.7051	ZINC00608204	0.0692	1.978	0.1499
ZINC01672834	0.2161	0.5509	0.4372	ZINC00034199	0.0457	0.1307	0.1932	ZINC00608382	0.0837	0.0682	0.0289
ZINC01672854	0.4145	0.5529	0.2246	ZINC00036356	0.3094	0.2181	0.3817	ZINC00643143	0.0866	0.0826	0.1906
ZINC01743739	14.255	0.158	0.179	ZINC00037582	1.2508	0.1457	0.296	ZINC01482077	0.0866	0.1005	0.1929
ZINC01786818	0.8237	0.362	0.2453	ZINC00038954	0.2515	0.8024	0.1681	ZINC01530886	0.2449	0.0362	0.1083
ZINC01795820	1.5623	0.2814	0.4838	ZINC00041740	0.6157	0.6295	0.2338	ZINC01530922	0.0372	0.6185	0.0777
ZINC01802626	0.6422	0.1227	0.1348	ZINC00041825	0.1942	0.6665	0.2782	ZINC01542199	0.0181	0.0166	0.0509
ZINC01802628	0.064	1.154	0.18	ZINC00041957	0.3374	0.4318	0.2992	ZINC01550477	0.1274	0.4127	0.0374
ZINC01802630	0.2245	0.2279	0.525	ZINC000448539	0.3549	0.2305	0.3689	ZINC01554274	0.1042	0.0475	0.0341
ZINC01802637	0.0255	0.6281	0.1643	ZINC000449542	0.2274	0.1653	0.2577	ZINC01851132	0.11	0.066	0.0596
ZINC01805625	0.1485	0.9805	0.1167	ZINC00050152	6.5426	0.105	0.07781	ZINC01851149	0.1574	0.0418	0.0983
ZINC01867107	0.453	0.7232	0.2257	ZINC00051778	0.4544	0.0677	2.734	ZINC02568036	0.0594	0.0376	0.0291
ZINC01909001	0.391	0.2228	0.392	ZINC00054526	0.073	0.1616	0.1358	ZINC03812892	6.3643	0.5012	69.825
ZINC01926014	0.2535	0.2047	0.1639	ZINC00054821	1.0142	0.3828	0.8446	ZINC03824921	0.0162	0.5309	0.0702
ZINC02023751	0.4822	0.4488	0.8134	ZINC00058794	0.7438	0.0732	2.4732	ZINC03831157	0.0288	0.192	0.3128
ZINC02024269	0.0259	0.0239	0.3829	ZINC00059228	0.2186	0.3577	1.109	ZINC03831511	0.0687	0.1335	0.2261
ZINC02024275	0.0343	0.3827	0.0643	ZINC00063634	0.181	0.0943	0.1339	ZINC03860156	0.13506	0.1142	0.0508
ZINC02024285	0.1415	0.0492	0.0271	ZINC00064150	0.5499	0.0353	0.1735	ZINC03872566	0.3263	0.2395	0.1156
ZINC02057647	0.5108	5.2197	59.851	ZINC00065395	0.3883	1.0413	0.0744	ZINC03875334	0.2018	0.9895	0.2371
ZINC02057774	0.7751	0.0447	0.0951	ZINC00065402	0.4687	0.1253	1.1835	ZINC03913937	0.3416	0.347	0.5085
ZINC02059027	0.4351	0.4851	0.3966	ZINC00066256	0.1013	0.0678	0.2329	ZINC03915154	3.5371	0.5494	0.5565
ZINC02060609	0.3968	0.5157		ZINC00068043	0.1313	0.4188	0.1976	ZINC03918453	0.1027	0.1473	0.0406
ZINC02061680	0.3581	0.0413	0.0425	ZINC00068063	0.3887	0.0661	0.1189	ZINC03920266	0.073	0.0853	0.2913
ZINC02061701	0.3711	0.5118	0.399	ZINC00071989	0.18	0.3186	0.2697	ZINC03977942	0.5025	0.4762	0.3227
ZINC02064730	0.2061	0.5428	0.2412	ZINC00073687	0.081	0.1934	0.3011	ZINC03977978	0.4011	0.1115	0.1545
ZINC02064732	0.3016	0.8677	0.4149	ZINC00075950	0.652	1.519	0.0412	ZINC03978005	0.165	0.765	0.1484
ZINC02069538	0.0395	0.0755	0.0393	ZINC00078911	0.4766	0.1736	0.1668	ZINC04097344	0.0405	0.0388	0.0333
ZINC02069618	0.0704	0.6726	0.7592	ZINC00081537	0.0102	0.1896	14.7701	ZINC06716957	0.0278	0.0211	0.1465
ZINC02069619	0.5983	0.4945	0.4577	ZINC00081701	0.4646	0.5869	0.3729	ZINC06745272	0.11602	0.3532	0.0791
ZINC02069629	0.0965	0.2934	0.0641	ZINC00081906	0.8015	0.1426	0.1675	ZINC11671039	0.3323	0.0222	0.0719
ZINC02069636	0.3986	0.1736	0.0793	ZINC00082622	0.0204	0.3384	0.0733	ZINC11677837	0.014	0.0408	0.0161
ZINC02069659	0.7644	0.0247	0.361	ZINC00084759	0.4426	0.4425	0.0367	ZINC13513942	0.0229	0.0294	0.0294
ZINC02069676	3.0564	0.8239	0.2468	ZINC00085793	0.1079	0.2762	1.58	ZINC13540266	0.2855	0.2716	0.512
ZINC02069682	0.2268	0.3973	0.466	ZINC00086070	0.0229	0.04	0.03	ZINC14879969	0.1302	0.1927	0.0296
ZINC02069711	0.4404	0.3348	0.2272	ZINC00086341	0.0942	0.2947	1.3458	ZINC14879999	0.0345	0.1159	0.02
ZINC02077101	0.0287	0.0371	0.0271	ZINC00087423	0.0469	0.474	0.4212	ZINC14880002	0.1968	0.4152	1.089
ZINC02106609	1.5344	0.1042	0.0405	ZINC00089688	0.1672	0.266	0.1978	ZINC15668997	0.0551	0.0695	0.0662
ZINC02142389	0.5471	0.0472	0.053	ZINC00091838	0.2956	0.221	0.1645	ZINC19340795	0.0212	0.0709	0.0974
ZINC02144897	0.6732	0.1824	0.9862	ZINC00094269	0.6628	0.4571	0.2353	ZINC19360739	0.1187	0.4241	0.1833
ZINC02157953	0.0612	0.1938	0.0333	ZINC00094388	0.1154	0.3552	0.5851	ZINC19361042	1.827	0.1293	0.1352
ZINC02182571	0.2588	0.3272	0.1312	ZINC00097010	2.7617	10.8843	0.1998	ZINC19364226	0.3812	2.593	0.1163
ZINC02182628	1.0063	0.8668	0.4155	ZINC00102308	0.0447	0.0826	0.2128	ZINC19364228	0.0193	0.1959	1.8054
ZINC02374395	0.0161	0.0144	1.3621	ZINC00102322	0.2767	0.1708	0.1034	ZINC19594557	2.094	0.2041	0.4663
ZINC02420638	0.0384	0.0164	0.0715	ZINC00102769	0.1455	2.2157	0.3458	ZINC19632614	0.2113	0.3341	0.1723
ZINC02510948	0.1336	0.0391	0.1455	ZINC00103247	0.2138	0.338	0.2943	ZINC19632618	0.0615	0.3451	0.062
ZINC02560378	0.0119	0.0913	0.0461	ZINC00105645	0.0498	0.1061	0.3901	ZINC19632891	0.2665	0.1244	0.1568
				ZINC00106815	0.0179	0.0197	0.0297	ZINC19796080	0.2432	0.0899	0.2713
				ZINC00106879	0.0587	0.1953	0.0748	ZINC19796087	0.0694	0.1087	0.2209
				ZINC00106900	0.0631	0.0243	0.1078	ZINC21297660	0.247	0.3097	0.0777
				ZINC00109321	0.1131	0.3331	0.2501	ZINC21982951	0.0508	0.0446	0.0308
				ZINC00110907	0.2545	0.3474	0.1406	ZINC22448696	0.8445	0.1416	0.8802
				ZINC00111501	0.1075	0.1734	0.2433	ZINC29416466	0.0928	0.1049	2.0448
				ZINC00112294	0.074	0.2451	0.0405	ZINC33359785	0.1011	0.39	0.1985
				ZINC00114132	0.01975	0.2289	0.02601	ZINC36701290	0.1319	0.1127	0.1034
				ZINC00114137	0.0675	0.0423	0.0576	ZINC43207238	0.0895	0.0551	0.1152
				ZINC00114142	0.073	0.2225	0.0643	ZINC49841054	0.1277	0.068	0.301
				ZINC00114145	0.2272	0.4813	0.5357	ZINC49918330	0.1774	0.0715	0.1227
				ZINC00114215	0.1605	0.3271	0.3858	ZINC52509366	0.0268	1.3032	0.3868
				ZINC00114829	6.01	6.0102	0.6783	ZINC52955754	0.4282	0.1921	0.221
				ZINC00114832	0.1332	0.1328	0.4709	ZINC53682927	0.119	0.1549	0.154
				ZINC00116833	0.0361	0.2511	0.3617	ZINC53683151	0.1226	0.5693	0.1451
				ZINC00117486	0.1988	0.3722	0.0668	ZINC64622550	0.1005	0.0593	0.332
				ZINC00118207	0.4162	0.7497	0.8226	ZINC64622551	0.3136	0.1236	0.7173
				ZINC00118374	0.0781	0.0579	0.0611	ZINC64624931	0.0991	0.05578	0.034
				ZINC00118389	0.0906	0.7188	0.2375				
				ZINC00120092	89.354	0.1689	0.8899				
				ZINC00121224	0.1775	0.3148	0.5956				
				ZINC00122426	0.112	1.203	0.2427				
				ZINC00123200	0.8544	0.0383	1.737				
				ZINC00125385	0.2394	0.437	0.1247				
				ZINC00134332	0.0941	0.0709	0.2615				

A total of 256 galectin-8N-ligand complexes in triplicate were simulated (totalling to 768 simulations), and the MSD analysis was carried out. The ligand MSD values for all the simulations performed has been presented (Table 3.1) and the cell highlighted in yellow are ones that were less displaced during the simulation (within the MSD threshold). Any compound having MSD less than 0.1 in two out of three runs of simulation were retained for the next tier of screening. This analysis culminated into 51 compounds whose chemical structures are presented in Figure 3.7 and Figure 3.8. The structural architecture for most of lead-like compounds had two aromatic rings roughly one on each side joined by a linker (Figure 3.7). The FDA drugs on the other hand were structurally diverse, interestingly, some of the drugs had the sulfated linker (like the lead-like library) and one antidiabetic drug Canagliflozin contains sugar moiety in the structure (Figure 3.8).

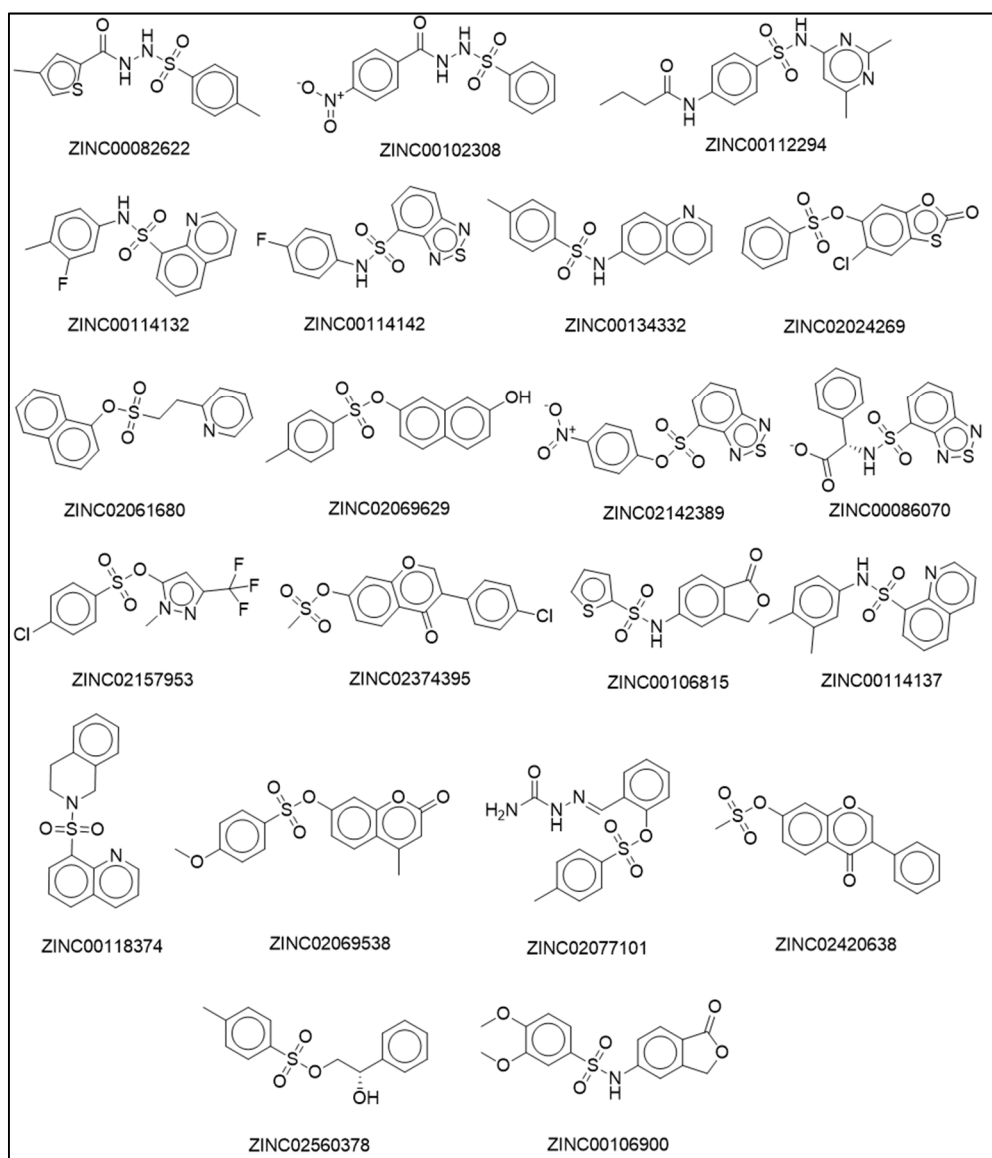


Figure 3.7: Chemical structures of the lead-like library of compounds retained during the short 5 ns simulation stage of the virtual screening.

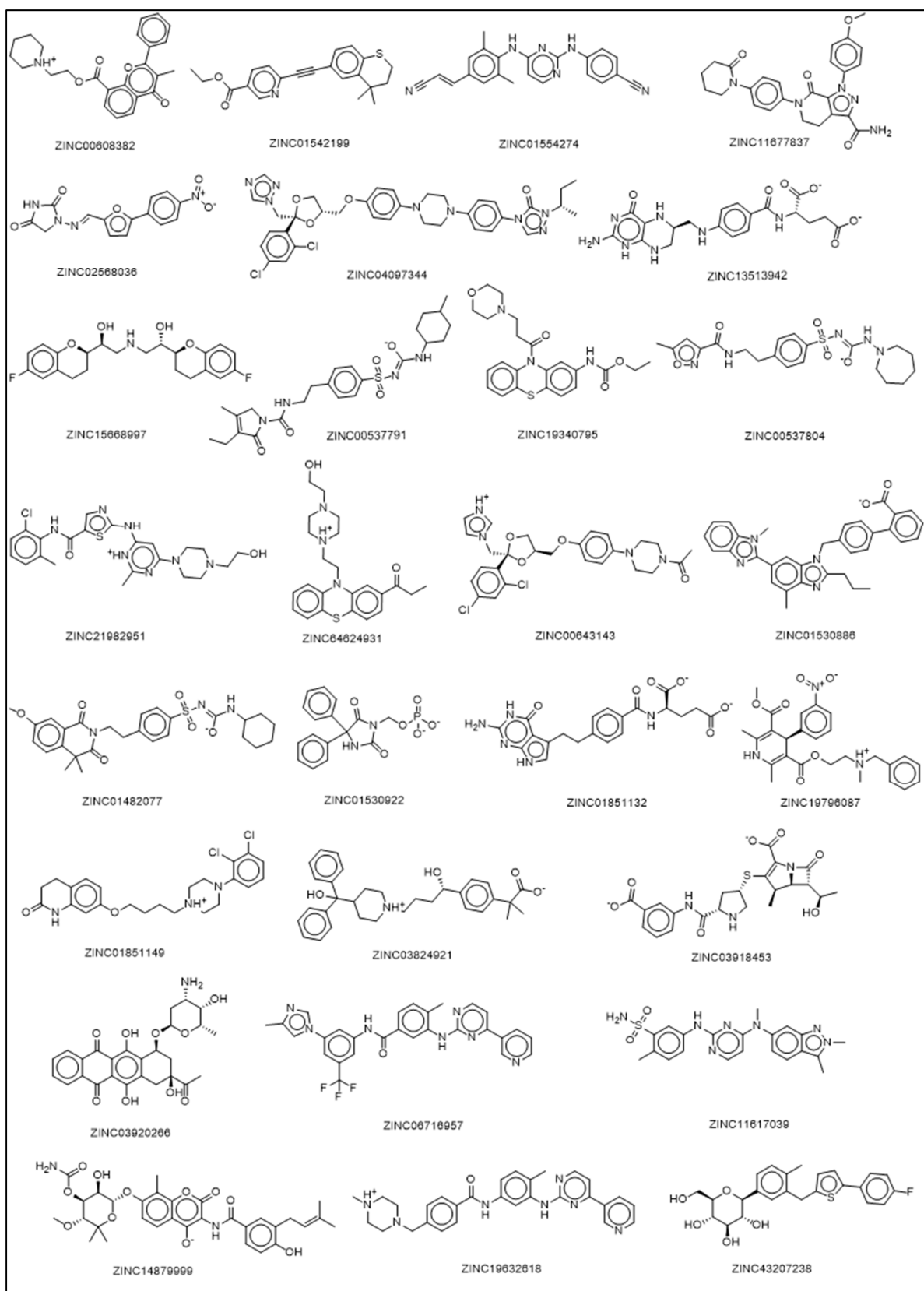


Figure 3.8: Chemical structures of the FDA drugs retained during the short 5 ns simulation stage of in the virtual screening.

3.3.4 Long simulations

A final round of 100 ns simulations were performed on compounds from the previous tier. The starting conformation being used was the docking output conformation. VMD was used for visually analysing the MD trajectories. The primary criteria for analysis were the same as previous, which was looking for compounds that were placed such that they were occupying the primary binding site. Nevertheless, compounds that were forming reasonably stable interactions were also considered for purchase apart from the ones that were showed the lowest fluctuation in conformation and interactions (highlighted yellow in Table 3.2).

Table 3.2: Mean square displacement values from long simulations for the lead-like compounds (on the left) and FDA drugs (on the right). The yellow highlight is the ones that followed the MSD rule set previously and compound ID coloured green are the ones purchased.

Sulfated	MSD	FDA drugs	MSD
ZINC00086070	0.07	ZINC00608382	2.9606
ZINC00106815	0.0427	ZINC01542199	0.2077
ZINC00114137	0.1224	ZINC01554274	0.5794
ZINC00118374	0.0752	ZINC02568036	0.099
ZINC02069538	0.6512	ZINC04097344	0.2077
ZINC02077101	0.8572	ZINC11677837	46.037
ZINC02420638	1.498	ZINC13513942	0.2867
ZINC02560378	0.0174	ZINC15668997	0.1473
ZINC00082622	0.0605	ZINC19340795	0.4317
ZINC00102308	0.552	ZINC21982951	0.0388
ZINC00106900	3.442	ZINC64624931	0.282
ZINC00112294	0.963	ZINC00537791	0.0746
ZINC00114132	1.5413	ZINC00537804	0.157
ZINC00114142	0.375	ZINC00643143	1.374
ZINC00134332	249.32	ZINC01482077	0.0805
ZINC02024269	0.0817	ZINC01530886	1.462
ZINC02061680	0.597	ZINC01530922	0.8437
ZINC02069629	0.404	ZINC01851132	1.0311
ZINC02142389	0.1287	ZINC01851149	0.0453
ZINC02157953	0.1341	ZINC03824921	1.028
ZINC02374395	1.712	ZINC03918453	0.1206
		ZINC03920266	0.2405
		ZINC06716957	0.1621
		ZINC11617039	0.2953
		ZINC14879999	0.8082
		ZINC19632618	1.185
		ZINC19796087	0.1536
		ZINC29416466	2.8045
		ZINC43207238	0.3856
		ZINC64622550	0.1468

From the short (5 ns) simulations of 256 compounds, 51 compounds were retained in the binding site at the minimum of two out of three times (Figure 3.7 and Figure 3.8). The MSD analysis in Table 3.2 shows that 11 compounds (zinc ID coloured in green and MSD value yellow highlighted) were least fluctuating during the length of simulation i.e. making stable

interactions with the binding site residues. While, other 40 compounds showed a varied degree of flexibility as indicated by higher MSD value. Nevertheless, eight additional compounds (Table 3.2; zinc ID coloured in green) were also included considering the chemical structure (Figure 3.7 and 3.8), interactions that were made with the key binding site residues (Figure 3.9 and 3.10) and their availability with vendors.

The binding modes of the final sulfated compounds and FDA drugs with the galectin-8N binding site has been represented (Figure 3.9 and Figure 3.10). Of note, these interaction figures were snapshot from simulations to roughly depict the placement of a compound in the binding site. The conformation presented in the interaction figures particularly is the starting coordinates of long simulations. For all the final compounds, key binding site interactions, that were employed as interaction filters, showed over 60% occupancy of hydrogen bonds as qualitatively assessed in VMD (not shown), whilst the interactions with other residues were transient. All the compounds in the purchase list not only occupied but also were retained in the primary binding site during the simulation. Being passed through the unique residue interaction criteria, all the compounds in sulfated molecules hydrogen bonded with Arg59 including partial occupancy for Gln47 and Arg45. In the case of FDA drugs, due to varied chemical structure, diverse range of residues including Arg59, Arg45, Gln47, Asn79, Asn67, His65 and Trp86 were involved in hydrogen bonding interactions (Figure 3.10).

All the compounds in the final list composed of the ones that were filtered through our screening tier and certain compounds that did not follow the MSD cut-off were also included due to availability from the same vendor (corresponding MSD value highlighted in yellow Table 3.2). A total of 19 compounds were purchased including nine compounds from the lead-like library and ten FDA drugs. The lead-like compounds were obtained through MolPort, and most of the FDA drugs were purchased from Selleckchem. Smallest available package size from vendors was acquired mostly either in 5 mg (lead-like compounds) and 25 mg (FDA drugs) sizes. These compounds were tested for validation by performing one-to-one binding to galectin-8N through STD NMR and X-ray crystallography.

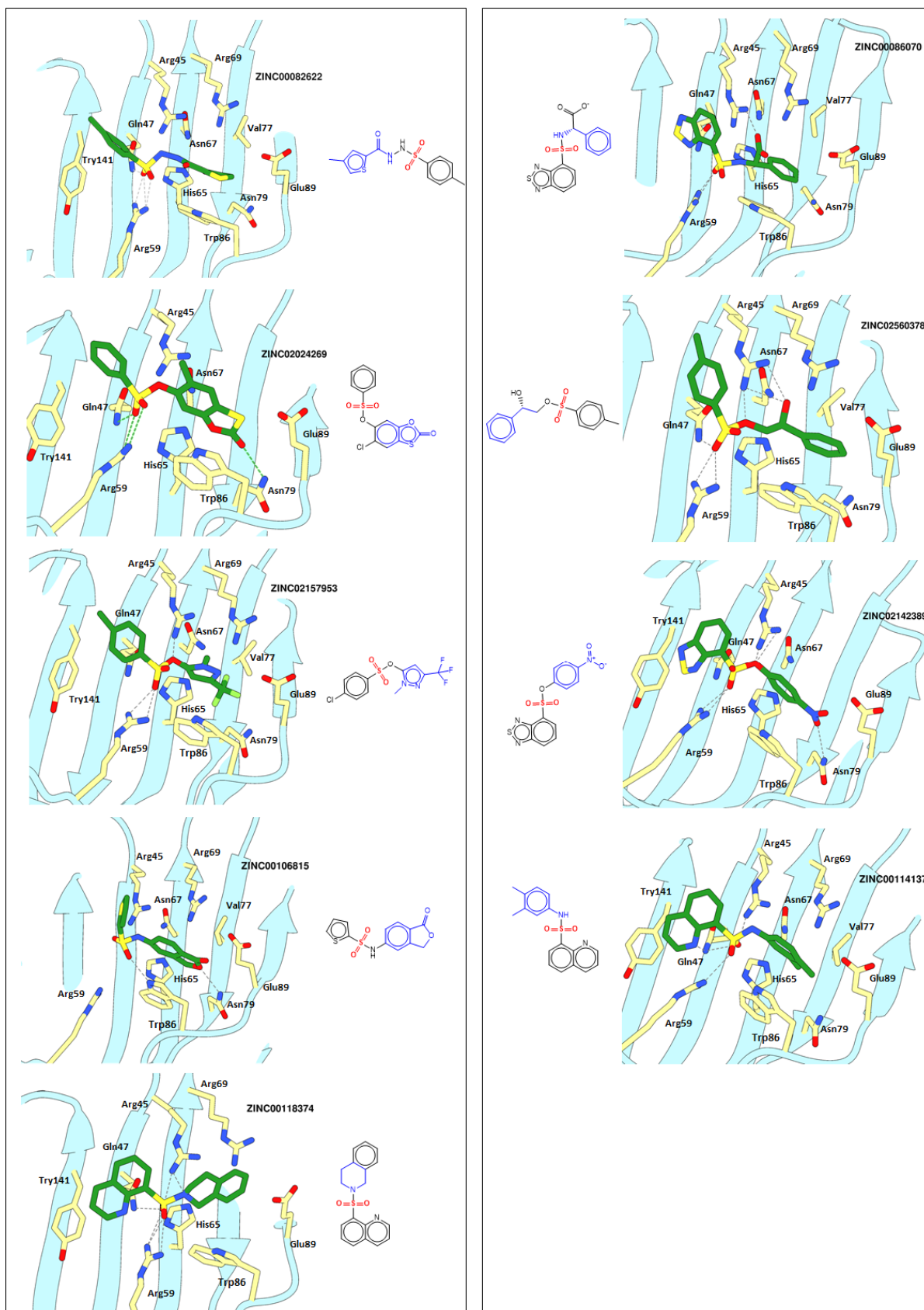


Figure 3.9: Snapshot from MD simulations depicting binding conformation and hydrogen bonding interactions between the final lead-like library of compounds (carbon in green sticks) and galectin-8N (light blue ribbon) binding site (carbon in yellow sticks). The corresponding ligand chemical structure colour coded based on respective part of ligands interacting with the galectin-8N binding site; blue for atom occupying the galactose binding pocket, red for atoms interacting or placed near Arg59 and black for atoms placed on either side of the binding site.

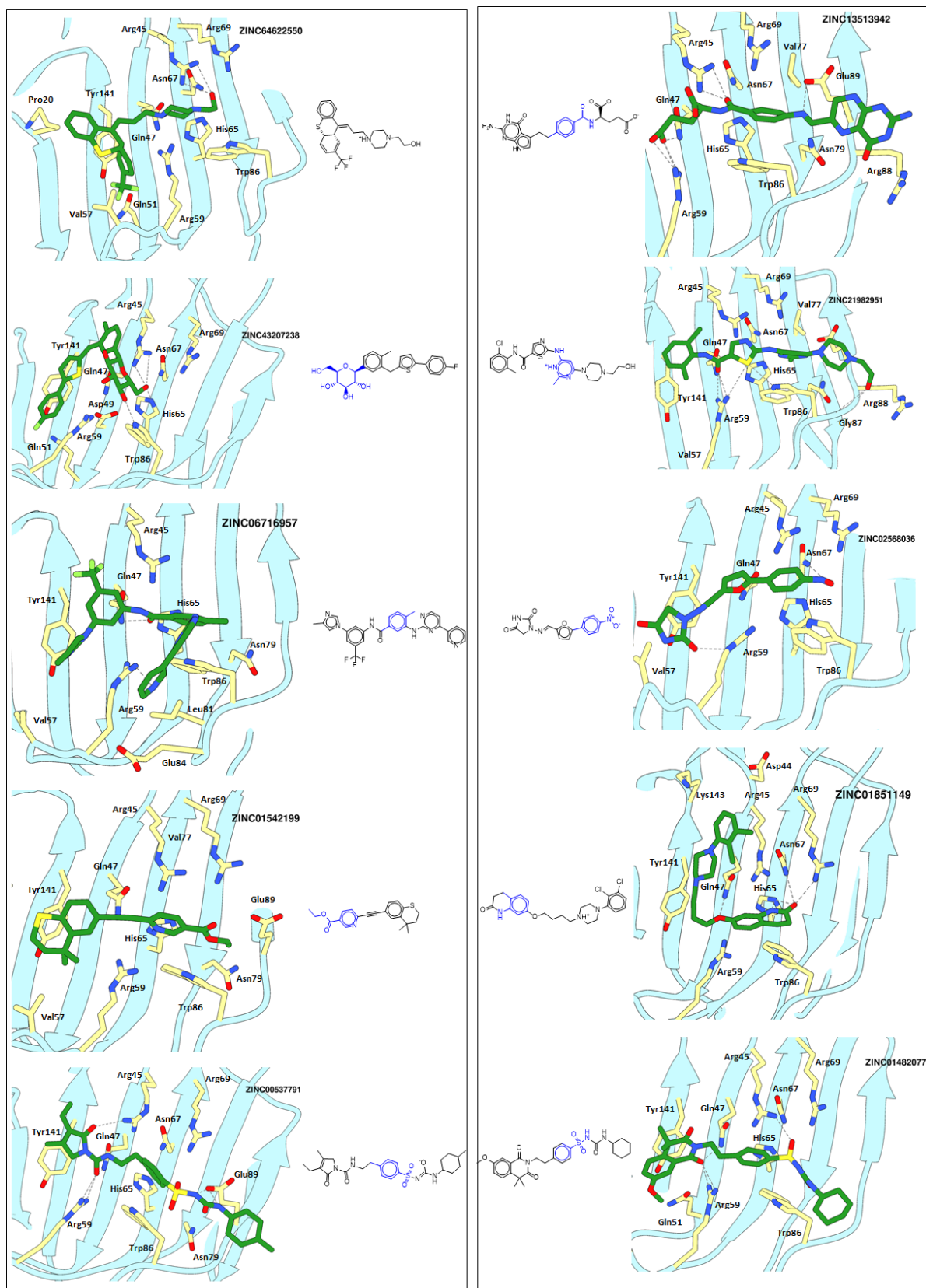


Figure 3.10: Snapshot from MD simulations depicting binding conformation and hydrogen bonding interactions between the final FDA drugs (carbon in green sticks) and galectin-8N (light blue ribbon) binding site (carbon in yellow sticks). The corresponding ligand chemical colour coded based on respective part of ligands interacting with the galectin-8N binding site; blue for atom occupying the galactose binding pocket, red for atoms interacting or placed near Arg59 and black for atoms placed on either side of the binding site.

3.3.5 Binding evaluation through STD NMR

The purchased compounds from the virtual screening campaign were subjected to experimental evaluation for binding using STD NMR. Almost all the compounds were solubilized in phosphate buffer to a final concentration of 500 μ M – 5 mM containing about 5-8 % DMSO. However, compound #19 (ZINC13513942) which is an FDA drug (Pemetrexed) was soluble in water/phosphate buffer due to the presence of two carboxylic acid groups. Sample solution prepared for STD NMR run contained a maximum of 5% DMSO with 5 μ M galectin-8N (previously expressed and purified, see Chapter 3). Ligand concentration in the sample solution was varied from 5 mM to 500 μ M while keeping the protein concentration constant. Some compounds in sample solutions turned cloudy due to their insolubility in the buffer at 5% DMSO. For aiding the solubility of compounds during the experiment, temperature of STD NMR experiments was increased to room temperature as compared to 10 degrees Celsius used otherwise.

During the initial screening, 64 scans were performed which were further increased to 512 scans for investigating any possible weaker interactions for some ligands between protein and ligand. Unfortunately, STD NMR-based screening did not show any of the purchased compound binding to the galectin-8N. For a true positive binder, the STD NMR spectra would reflect signals corresponding to the substructure of ligand that is in close contact with the protein. However, the difference spectra for all the compounds tested through STD NMR did not show any binding.

3.3.6 Binding evaluation through ligand soaking

Ligand soaking experiments were also performed to cross-verify the outcome of STD experiments. For this purpose, a range of ligand concentrations with 5-8 % DMSO was soaked in the *apo* galectin-8N crystals. These apo crystals should have enough solvent channel allowing for soaking of the ligands as these crystals (previously) have been soaked with tetrasaccharides of human milk glycans. The ligand soaking time was varied from overnight to about 30 min. The soaked crystals were cryo-cooled with 20% glycerol as the cryoprotectant, X-ray diffraction datasets were collected at Australian Synchrotron and structures were determined at varied resolution range. All the ligand soaking datasets were processed, and initial model building and refinement revealed no ligand in the binding site i.e. they were galectin-8N bound to glycerol (from the cryo-protectant). The crystallographic information about the processing of the soaked ligand datasets and their subsequent structure solution for each of those ligands has therefore been not presented here. The ligand soaking results further

corroborate the STD NMR results and supports no binding of the purchased compounds to the galectin-8*N*.

3.4 Conclusions

Structure-based virtual screening has its success stories that led design and development of new drug molecules [157, 158]. The performance of docking applications is highly dependent on the macromolecular target under investigation [150, 151]. Docking applications can predict and rank-order the experimentally observed conformation well for some class of protein than for others [150]. This irregularity particularly highlights the dependence of docking outcome on to the nature of the binding site. In addition, given the fact that the outcome of MD simulations is dependent on the initial coordinates, there is a chance that some compounds would have been missed. Furthermore, the arbitrary cut-off employed during filtering of compounds at the simulation stage could also lead to missing true positives. Compounds with more number of rotatable bonds are more flexible during the simulations if no amino acid residues are available in the vicinity for interaction. Even though some of part of that compound would be occupying the binding site maintaining the filtering criteria the remaining flexible part of compound could result in higher MSD value and thereby not being retained during screening.

Given the higher attrition from the virtual screening exercises, the number of compounds purchased for evaluation becomes critical, at the least 100s of compounds are purchased to obtain two or three true positives [159]. Another factor that affects the success of a virtual screening campaign is the strategy employed for filtering and rank ordering the library of the compound. There could be numerous ways to approach a problem at hand. Currently, there is no reported success in literature with galectins and virtual screening for identification of novel ligands. The lack of virtual screening reports in literature may be associated with the solvent exposed galectin binding site that would mainly require displacement of water molecules occupying the binding site. Therefore, the incoming ligand needs more polar functionalities to establish a network of hydrogen bonds to be retained in the binding site, as observed with galectin-8*N*-glycan interactions [99, 160]. Having multiple polar functionalities is a signature of carbohydrate-based molecules (that contains multiple hydroxyl groups) and less commonly observed in synthetic lead-like non-carbohydrates.

As for carbohydrate-based molecules, the more the number of sugar units in the ligand the more complex it is to synthesise. The presence of polar group architecture mimicking only the galactose ring is not enough as particularly evident from low mM affinity for the

monosaccharide (towards galectins) as compared to the μM affinity of lactose (a disaccharide) [41, 42, 44]. The possibility of finding a non-carbohydrate type binder for galectin remains a challenge. Therefore, from another perspective, one can initiate the ligand design through carbohydrate-based molecules and modify the key positions to eventually head to a non-carbohydrate binder, as approached for the development of a non-peptide-based inhibitor of galectin-1 [161-163]. Following chapters (Chapter 4 and 5) will, therefore, explore possible modification on the monosaccharide galactose considering both the conserved and the unique features of the galectin-8*N* binding site.

3.6 Appendix

3.6.1 AutoDock Vina virtual screening master script

```
#!/bin/bash
#
# Generated with Raccoon | AutoDockVS
#

#### PBS jobs parameters CPUT="24:00:00"
WALLT="24:00:00"
#
# There should be no reason
# for changing the following values
NODES=1
PPN=1
MEM=512mb

### CUSTOM VARIABLES
#
# use the following line to set special options (e.g. specific queues)
#OPT="-q MyPriorQueue"
OPT="-q workq"

module load autodock/autodock423

# Paths for executables on the cluster
# Modify them to specify custom executables to be used
QSUB="qsub"
AUTODOCK="/sw/autodock/autodock423/bin/autodock4"

AUTOGRID="/sw/autodock/autodock423/bin/autogrid4"

VINA="/sw/autodock/autodockvina/autodock_vina_1_1_2/bin/vina"

# Special path to move into before running
# the screening. This is very system-specific,
# so unless you're know what are you doing,
# leave it as it is
WORKING_PATH=`pwd`

#####There
should be no need to modify anything below this line
#####

type $AUTODOCK &> /dev/null || {
    echo -e "\nError: the file [$AUTODOCK] doesn't exist or is not
executable\n";
    echo -e "Try to specify the full path to the executable of the
AutoDock binary in the script";
    echo -e "( i.e. AUTODOCK=/usr/bin/autodock4 )\n\n";
    echo -e " [ virtuals screening submission aborted]\n"
    exit 1; }

type $AUTOGRID &> /dev/null || {
    echo -e "\nError: the file [$AUTOGRID] doesn't exist or is not
executable\n";
    echo -e "Try to specify the full path to the executable of the
AutoGrid binary in the script";
    echo -e "( i.e. AUTOGRID=/usr/bin/autogrid4 )\n\n";
```

```

        echo -e " [ virtuals screening submission aborted]\n"
        exit 1; }

type $QSUB &> /dev/null || {
    echo -e "\nError: the file [$QSUB] doesn't exist or is not
executable\n";
    echo -e "Try to specify the full path to the executable of the Qsub
command binary in the script";
    echo -e "( i.e. QSUB=/usr/bin/qsub )\n\n";
    echo -e " [ virtuals screening submission aborted]\n"
    exit 1; }
#####

echo Starting submission...
for NAME in `cat jobs_list`
do
    mkdir $NAME
    cp 3AP6-p.pdbqt config.txt $NAME
    cp $NAME.pdbqt $NAME
    cd $NAME
    echo "#!/bin/bash" > $NAME.job
    echo "cd $WORKING_PATH/$NAME" >> $NAME.job
    echo "$VINA --config config.txt --ligand $NAME.pdbqt --log
$NAME.txt" >> $NAME.job
    chmod +x $NAME.job
    echo -n "Submitting $NAME : "
    $QSUB $OPT -l cput=$CPUT -l nodes=1:ppn=1 -l walltime=$WALLT -l
mem=$MEM $NAME.job
    cd ..
done

```

3.6.2 Docking score analysis

```

#!/bin/bash

for NAME in `cat jobs_list`
do
    cd $NAME
    grep "    1 " $NAME.txt | cut -c12-20,1-12 >>../scores.txt
    cd ..
done

```


Structure-based Design of a Monosaccharide Ligand Targeting Galectin-8



4.1 Foreword

Galectin-8 is a modulator of bone remodelling process where its inhibition could serve as an emerging new approach to tackling diseases with associated bone-loss. Towards this end, our attempt to identify non-carbohydrate-based ligands through structure-based virtual screening (in Chapter 3) did not result in any positive hits. However, the idea of exploiting the unique Arg59 through negatively charged carboxylic acid group was originated. The existing structural information of galectin-8N-glycan complexes provided an ideal template to initiate the ligand design process. In this study, our quest to employ rational approaches towards design and development of a monosaccharide-based ligand of galectin-8N was carried forward. A combination of modelling and experimental techniques was used to generate the ligand design hypothesis and validate it for binding to galectin-8N through ITC and X-ray crystallography. This study will therefore act as a proof-of-concept for the idea of exploring the combination of unique and conserved binding site residues towards the design of efficient ligands.

4.2 Statement of contribution to a co-authored publication

This chapter includes a co-authored paper. The bibliographic details of the co-authored paper, including all authors, are:

Mohammad H. Bohari, Xing Yu, Rob-Marc Go, Yaron Vinik, I. Darren Grice, Yehiel Zick and Helen Blanchard (author list needs to be finalised). ChemMedChem (submitted)

My contribution to the paper involved:

Molecular dynamics simulations and analysis for the galectin-8*N*-3'-SiaLac and galectin-8*N*-**6**. Synthesis of compound **6** with advice from Dr. Darren Grice. Protein expression, purification and crystallisation. Galectin-8*N*-**6** complex structure determination with advice from Dr. Xing Yu and Prof. Helen Blanchard. Binding affinity determination and data analysis by ITC. Input into interpretation and analysis of the generated data and contribution to writing the manuscript.


(Signed) _____ (Date) 03/04/2017 _____

Mohammad Bohari


(Countersigned) _____ (Date) 24/11/2017 _____

Corresponding author of paper: Professor Helen Blanchard


(Countersigned) _____ (Date) 24/11/2017 _____

Supervisor: Professor Helen Blanchard

4.3 Structure-based Design of a Monosaccharide Ligand Targeting Galectin-8

Mohammad H. Bohari¹, Xing Yu¹, Rob-Marc Go¹, Yaron Vinik³, I. Darren Grice^{1,2}, Yehiel
Zick³, and Helen Blanchard^{1*}

¹Institute for Glycomics, Griffith University, Gold Coast Campus, 4222, Australia.

²School of Medical Science, Griffith University, Gold Coast Campus, 4222, Australia

³Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel.

*To whom correspondence should be addressed: Professor Helen Blanchard

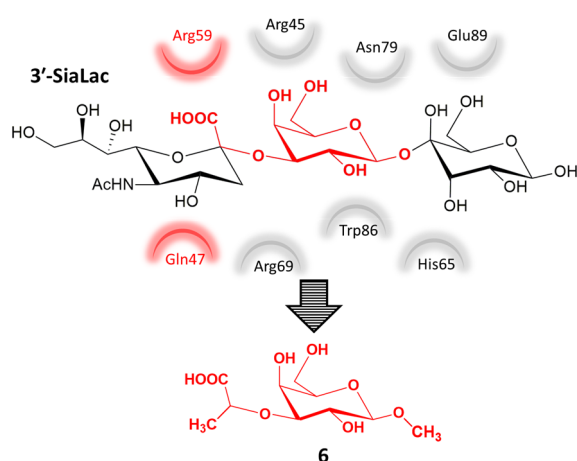
Institute for Glycomics, Griffith University, Gold Coast Campus, 4222, Australia.

Phone: +61 7 555 27023; Fax: +61 7 555 28098; E-mail: h.blanchard@griffith.edu.au

Keywords

Structure-based ligand design, carbohydrate, galectin-8N terminal domain, X-ray crystal
structure

4.3.1 Table of content



Exploiting both evolutionarily conserved and unique binding-site amino acids in ligand design: Designing an efficient ligand is at the crux of drug development. Targeting human galectin-8, we have undertaken in silico structure-based ligand design, chemical synthesis and validation of binding using ITC and X-ray crystallography.

4.3.2 Abstract

Galectin-8 is a β -galactoside-recognising protein that has a role in the regulation of bone remodelling and is an emerging new target for tackling diseases with associated bone-loss. We have designed and synthesised methyl 3-*O*-[1-carboxyethyl]- β -D-galactopyranoside (compound **6**) as a ligand to target the *N*-terminal domain of galectin-8 (galectin-8*N*). Our design involved molecular dynamics (MD) simulations that predicted **6** to mimic the interactions made by the galactose ring as well as the carboxylic acid group of 3'-*O*-sialylated lactose (3'-SiaLac), with galectin-8*N*. The binding affinity of compound **6** to galectin-8*N* was 139 μ M as measured by isothermal titration calorimetry (ITC). The crystal structure of the galectin-8*N*-**6** complex validated the predicted binding conformation and revealed the exact protein-ligand interactions that involve galectin evolutionarily conserved amino acids and also those unique to galectin-8*N* for recognition. Overall, we have initiated and demonstrated a rational ligand design campaign to develop a monosaccharide-based scaffold as a binder of galectin-8.

4.3.3 Introduction

Galectin-8 is a β -galactoside recognising protein that contains two carbohydrate recognition domains (CRD) in tandem, linked by a variable length amino acid linker [27, 28]. It has widespread tissue distribution in both normal and tumour cells. Within the cell, galectin-8 occurs in the nucleus, the cytoplasm, and also it is secreted into the extracellular space [14, 26, 63]. Apart from being involved in cell-to-cell and cell-to-surrounding communication, an increasingly broader functional spectrum of galectin-8 is apparent [33, 61]. Galectin-8 plays

an important role in inflammatory disorders through the regulation of T-cell homeostasis [37, 39, 51, 164] and is critically involved in capillary tube formation and angiogenesis [67]. Antibacterial activity, mediated through induction of selective autophagy, highlights an additional cellular mechanism to combat infection recruiting galectin-8 [46, 47]. It is of interest that galectin-8 has shown *in vivo* regulation of bone remodelling via enhancing expression of bone resorbing factors that attribute to increased bone turnover culminating in reduced bone mass [96]. Inhibition of galectin-8 may thus offer a potential new therapeutic approach in managing diseases with bone-loss.

Available structural information of galectin-8N bound to different biologically relevant oligosaccharides provide insight into various biological processes and highlights key interactions imparting affinity and specificity to a ligand [76, 81]. Typically, the CRD has a β -sandwich “jelly-roll” topology formed from two β -sheets. The concave surface of the roll constitutes the carbohydrate-binding site and comprises six β -strands, S1-S6 (Figure 1a). The galactose recognition site consists of evolutionarily conserved amino acids on strands S4-S6. Galectin-8N preferentially recognises anionic oligosaccharides such as 3'-sulfated lactose and 3'-O-sialylated lactose (3'-SiaLac)[42]. This preferential binding arises from the presence of unique structural features in the galectin-8N CRD such as a long S3-S4 loop bearing an arginine residue (Arg59) and Gln47 on strand S3 [76]. The crystal structure of 3'-SiaLac bound to galectin-8N shows engagement of the carboxylic acid group of sialic acid in salt bridge interactions with Arg59 and hydrogen bonding to Gln47, and the pyranose ring of the galactose formed conserved interactions observed for the counterpart of lactose (Figure 1b) [76]. These interactions between the anionic group of the glycans with Arg59 and Gln47 were attributed to the high affinity towards galectin-8N [76].

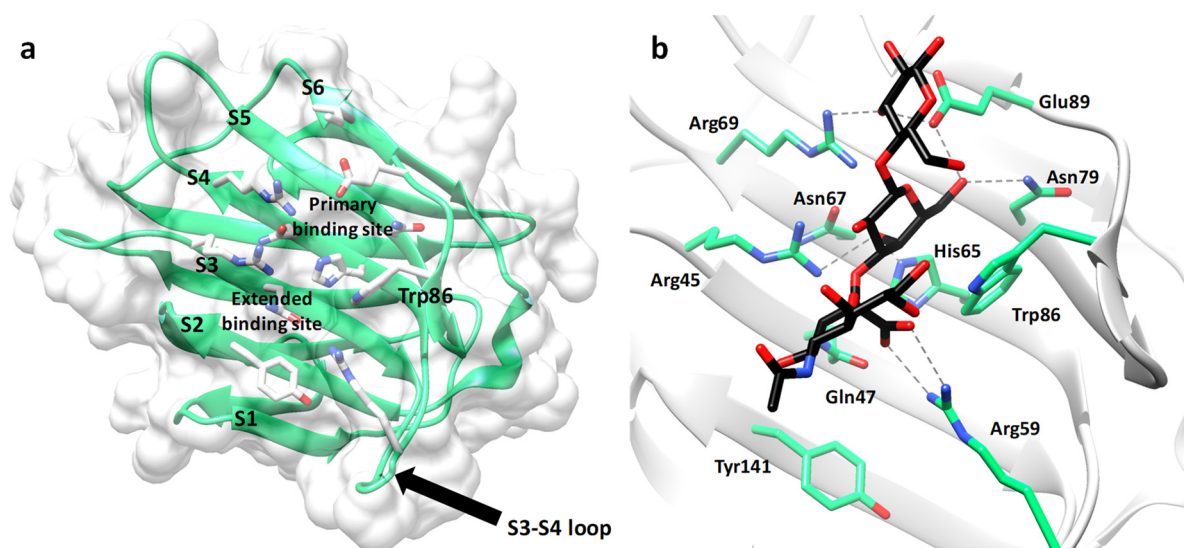


Figure 4.1: Overview of the galectin-8N carbohydrate recognition domain. a) The CRD showing the carbohydrate binding face of the “jelly-roll”. The primary and the extended binding site are indicated. b) Binding conformation and hydrogen bond and salt bridge interactions (in grey dashed lines) made by 3'-SiaLac (in sticks; carbon in black, oxygen in red) upon binding to galectin-8N residues (in sticks; carbon in green, oxygen in red and nitrogen in blue) PDB ID: 3AP7 [76].

Given the inherent nature of galectins to recognise β -galactoside-containing glycans and galectin-8's preferential recognition of anionic sugars, then a monosaccharide-based ligand such as galactose bearing a negatively charged carboxylic acid group at the 3'-position akin to 3'-SiaLac seems promising as a ligand scaffold. The advantageous incorporation of a carboxylic acid group in ligands to engage unique binding site residues is also supported from the outcome of our *in silico* virtual screening study (unpublished). We report the first synthetic ligand (methyl 3-*O*-[1-carboxyethyl]- β -D-galactopyranoside (**6**); Figure 2) developed through a structure-based rational design approach targeting galectin-8.

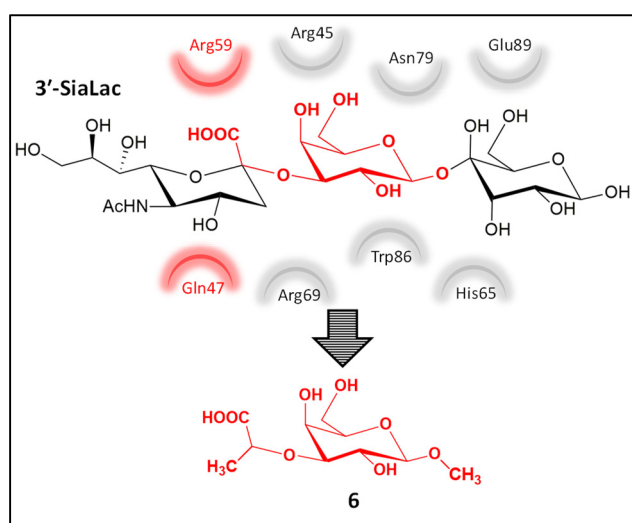


Figure 4.2. The ligand design concept. Depicted are the chemical structures of 3'-SiaLac, the designed compound **6**, and the surrounding amino acid residues in the galectin-8N binding site. The interaction with Arg59 and Gln47 (labelled in red) are the unique residues that **6** is designed to exploit.

4.3.4 Results and Discussion

We performed MD simulations on the galectin-8N-3'-SiaLac crystal complex along with the galectin-8N-lactose complex to compare their interaction profiles. Hydrogen bond occupancy analysis was used to assess the most frequent interactions occurring between the protein and ligand. The hydrogen bonds observed in galectin-8N-lactose and galectin-8N-3'-SiaLac crystal structures were found with occupancy of 60-100 % (total occupancy for the residue) during simulations (Figure 3). An identical interaction profile for the common lactose moiety within both ligands was observed for the conserved residues His65, Glu89, Arg69, and Asn79. The occupancy for Arg45 hydrogen bonds was higher in 3'-SiaLac compared to that in the lactose complex. The carboxylate group of 3'-SiaLac showed 100 % occupancy of salt bridge interactions with Arg59 and over 60 % occupancy of hydrogen bond with Gln47 and Trp86. Our simulation of galectin-8N-3'-SiaLac crystals complex revealed that 60% of the time the Gln47 side chain was re-orientated by 180° interchanging the N^{δ1} and O^{δ1} compared to the deposited PDB coordinates (PDB ID: 3AP7[76]) and thus enabled hydrogen bonding to occur with the carboxylic acid group. Critically, amino acids Arg59 and Gln47 are unique to galectin-8N and are identified as supposed ligand specificity hotspots [76, 81]. Taking advantage of this existing structural information and findings from our *in silico* virtual screening a monosaccharide-based ligand of galectin-8N was conceived to exploit interactions with both unique and conserved residues of the galectin-8N binding site.

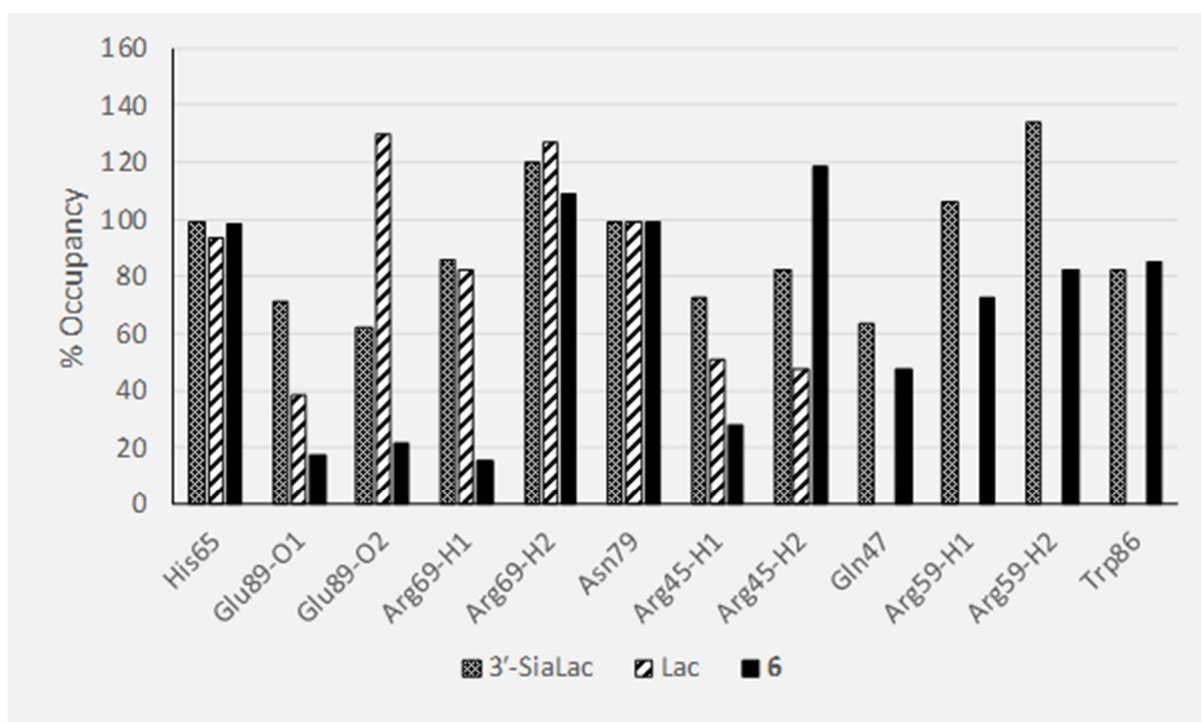
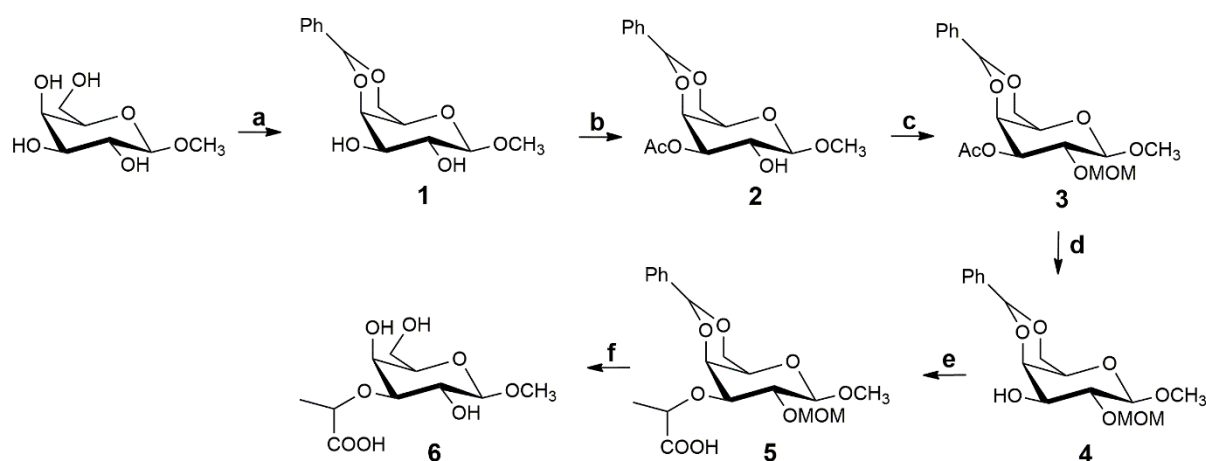


Figure 4.3. Interaction analysis from MD simulations. Hydrogen bond occupancy (in percentage) is analysed from 100 ns MD simulations of 3'-SiaLac, lactose (Lac) and **6** bound to galectin-8*N*. Note that the occupancy represented is for full side chain of the amino acid residue and not for individual donor/acceptor atoms.

We have designed compound **6** to incorporate stable interactions made by the galactose ring and by the carboxylic acid group, such as that shown by 3'-SiaLac, as an initial basis for the ligand design campaign (**Figure 4.2**). The main advantage here would be recognition of the galactose portion by galectin-8*N* with an anticipated increase in binding strength and specificity by the carboxylic acid group at the 3'-position of the galactose. Lactose from the galectin-8*N*-lactose complex (5T7S [81]) was modified to obtain initial placement of the designed ligand **6** in the galectin-8*N* binding site (Supplementary Figure S1). We then employed MD simulations (100 ns) to investigate the feasibility of the interactions observed in the initial model as well as the suitability of the designed ligand in the galectin-8*N* binding site. The ligand stayed in the primary binding site for the duration of the simulation, retaining the typical CH- π interactions with the evolutionarily conserved Trp86, as observed for natural galectin ligands. Hydrogen bond occupancy analysis revealed an almost identical interaction profile of **6** to that observed for the corresponding part of the galectin-8*N*-3'-SiaLac simulation (**Figure 4.3**). The hydrogen bond occupancy, particularly for the unique residues Arg59 and Gln47 and the conserved residue Trp86, noted for **6** is identical to that observed for 3'-SiaLac. The interaction analysis from MD simulations suggests that the galactose ring of **6** would occupy the primary binding site of galectin-8*N* and that the carboxylic acid group at the 3'-position would engage in interactions with the unique Arg59 and Gln47. Based on these predictions, we synthesised **6**,

and its binding affinity, binding mode and interactions to galectin-8N were validated using ITC and X-ray crystallography.

Synthesis of **6** was initiated with β -methyl galactoside as the starting material (**Scheme 4.1**). The β -methyl galactoside was 4,6-benzylidene protected (**1**) utilising benzaldehyde dimethyl acetal with a catalytic amount of camphor sulfonic acid [165]. Following which, selective 3O-acetylation (**2**) was performed using silver oxide and a catalytic amount of KI [166]. The acetylated compound was 2O-methoxymethyl ether (MOM) protected (**3**) using an adapted procedure [167] employing diisopropyl ethyl amine (DIPEA) in dichloromethane. After quantitative deacetylation (**4**) using sodium metal in methanol, the 2-chloropropionic acid side chain was coupled using NaH in 1,4-dioxane to yield compound (**5**) [168]. The final removal of all the protecting groups on the 2- and 4,6-positions was performed using concentrated HCl in methanol to yield a racemic mixture of compound **6** [169].



Scheme 4.1. Synthesis scheme for 6. Reagents and condition: a) Camphorsulfonic acid, benzaldehyde dimethyl acetal, ACN, 60 °C, 2.5 h (yield 88 %); b) Ag_2O (freshly prepared), AcCl , KI, DCM, RT, 16 h (yield 70 %); c) DIPEA, MOMBr, DCM, reflux, 16 h (yield 80 %); d) Na metal, MeOH, RT, 1.5 h (yield 90 %); e) NaH, 2-chloropropionic acid, anhydrous 1,4-dioxane, 50 °C, 36 h (yield 60 %); f) concentrated HCl, MeOH (yield 75 %).

The dissociation constant (K_d) of **6** to galectin-8N was measured through ITC. In the experiment, heat changes occurring from the ligand's titration into a protein solution and the ligand's titration into a buffer (without the protein) were recorded. The ligand exhibited 1:1 stoichiometric binding to galectin-8N with a binding affinity of $139.5 \pm 88 \mu\text{M}$ (**Figure 4.4: Isothermal calorimetric analysis**. Binding isotherm for titration of 1 mM lactose (on the left) and 1 mM compound **6** (on the right) into 200 μM galectin-8N in phosphate buffered saline.). Interestingly, the binding affinity obtained for the monosaccharide-based **6** was thus close to that measured for the disaccharide lactose ($136 \pm 31 \mu\text{M}$). Discrepancy in the absolute K_d value for lactose compared to those previously reported that used surface plasmon resonance (74-79

μM [42, 76]) and fluorescence polarisation (1.7-3.1 μM [44]) relate to the different techniques. Importantly, our results reveal that the binding mechanism noted from the thermograms of lactose and compound **6** was different, enthalpic component dominated the binding affinity of lactose while the entropic component contributed the most in the case of **6** (Figure 4.4).

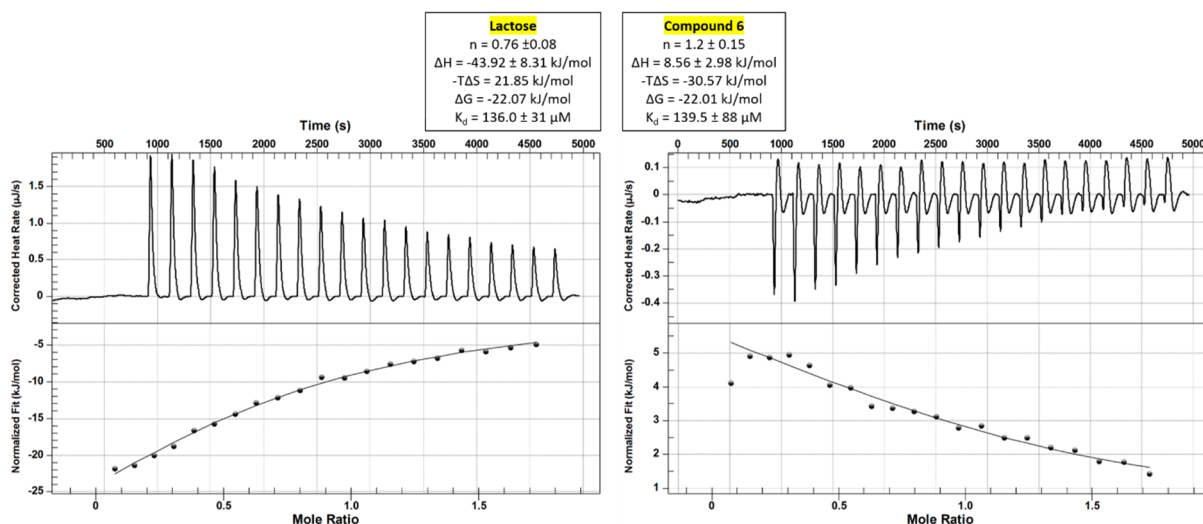


Figure 4.4: Isothermal calorimetric analysis. Binding isotherm for titration of 1 mM lactose (on the left) and 1 mM compound **6** (on the right) into 200 μM galectin-8N in phosphate buffered saline.

We subsequently performed crystallographic analysis to investigate binding mode and interactions of **6** to galectin-8N. The galectin-8N-**6** complex was obtained by soaking compound **6** into the *apo* galectin-8N crystals, and the structure was determined at 2.1 Å resolution. The electron density maps reveal unambiguous placement of **6** in the galectin-8N binding site (Figure 5, Supplementary Figure S2). A clear bump in the difference electron density map pointing towards the conserved Arg69 confirmed the positioning of the O4' hydroxyl of the galactose. Further, the planar topology of the electron density adjacent to the unique Arg59 is consistent with the placement of the carboxylic acid group and thereby confirming the overall placement of compound **6** (Figure 4.5). The electron density for the methyl group of the propionic acid side chain that points toward solvent is weak. However, it is sufficient to give the direction of methyl group that correlates with an R-configuration. The positive difference electron density that appears as an extension from the anomeric methoxy group, is concluded as a water molecule. The structure of the ligand was confirmed with the ^{13}C NMR showing peaks for two methyl groups, one being the anomeric (55.84 ppm) and the other on the propionic acid side chain (17.63 ppm). Additionally, corresponding proton peaks observed in the ^1H NMR and molecular weight from mass spectrometry confirmed the integrity of the ligand.

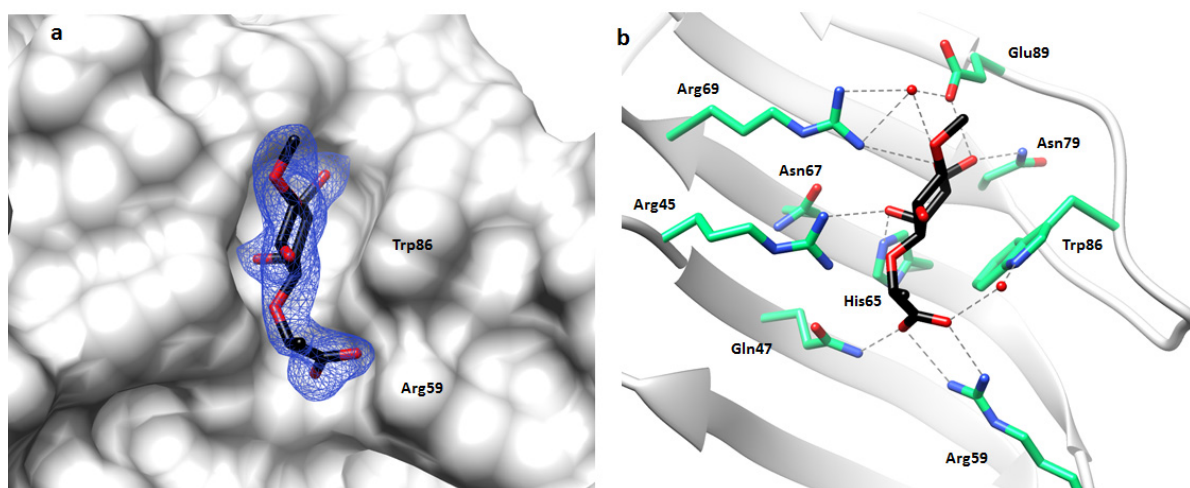


Figure 4.5: Galectin-8N-6 complex. a) Electron density (blue mesh) $2|F_o|-|F_c|$ α_c contoured at 1σ , for **6** (in sticks; carbon in black, oxygen in red) in complex with galectin-8N. b) Hydrogen bonding and salt bridge interactions (in grey dashed line) made between **6** and galectin-8N (in sticks; carbon in green; oxygen in red; nitrogen in blue).

Overall, the binding mode observed for the galactose portion of **6** is identical to that noted for the corresponding part of the galectin-8N-lactose complex [81]. The O4' of the galactose engages in hydrogen bonding with His65, Asn67, Arg45, Arg69 whereas the O6' hydrogen bonds with Asn79, and Glu89, as also noted from our simulations (**Figure 4.5**). Importantly, from the original design concept, the carboxylic acid was found in a geometrically favoured position to make ionic interactions with the unique Arg59 and hydrogen bonding interactions with Gln47. Furthermore, the placement of the anionic group and the galactose ring is identical to the equivalent part of 3'-SiaLac complexed with galectin-8N (Supplementary Figure S3). The carboxylic acid group displaced a water molecule that was observed in the vicinity of the conserved Trp86 and Arg59 in the galectin-8N-lactose complex (5T7S [81]), further supporting the comparatively strong binding seen in our ITC experiments. With the galectin-8N-**6** crystal structure, we validated our design concept and the predicted binding conformation for **6**.

Table 4.1: Crystallographic data merging and refinement statistics for galectin-8N-**6** complex structure.

Data	Galectin-8N-6
Indexing	
Crystal system	Orthorhombic
Space group	$P2_12_12_1$
Unit cell	$a=45.72$, $b=50.32$, $c=80.87$
Merging and scaling	
Resolution (Å)	42.73 – 2.10
Total observations	54110 (4147) [†]
Unique observations	11126 (873)
Multiplicity	4.9 (4.8)
Completeness (%)	98.0 (95.0)
I/ σ	5.9 (1.8)

R _{merge} (%)	14.9 (78.2)
Refinement	
Resolution	42.59-2.10
R factor (%)	22.2
R _{free} (%)	25.1
Number of atoms	
Protein	1196
Ligand	18
Water	103
Root mean square deviation	
Bond length (Å)	0.0072
Bond angle (°)	1.2858
Ramachandran plot statistics	
Favoured (%)	97.18
Allowed (%)	2.82
Average B-factor (Å ²)	
Protein	26.3
Ligand	36.3
Water	33.0
PDB ID	5VWG

[†]The value in parenthesis is for the highest-resolution shell

Since our ultimate aim is to design ligands that can specifically target galectin-8, then our focus has been to exploit the uniqueness of the galectin-8_N binding site. As a first step in assessing the level of specificity toward galectin-8_N compared to the C-terminal domain of galectin-8 (galectin-8_C), we have employed MD simulations to predict the binding mode and interactions of the designed ligand **6** towards galectin-8_C. The primary binding site of the two CRDs of the galectin-8 is mostly conserved. However, amino acid differences in the extended binding site have potential to play a critical role in recognising the compound **6**. We used the crystallographic conformation of **6** from galectin-8_N-**6** complex and generated the *in silico* galectin-8_C-**6** complex through superimposition of the two CRDs. The *in silico* galectin-8_C-**6** complex reveals possible hydrogen bonding interactions of **6** with His271 (His65 in galectin-8_N), Arg275 (Arg69), Glu294 (Glu89), and Asn284 (Asn79) as observed in the galectin-8_N-**6** crystal structure (Figure 4.6). However, the carboxylic acid side chain of **6** lacked interactions with the galectin-8_C binding site due to the absence of Arg59 on S3-S4 loop, the presence of Ser255 in place of Arg45 and Asn257 in place of Gln47 (Figure 4.6). To further investigate the favourability of these interactions in the *in silico* complex (galectin-8_C-**6**), 100 ns MD simulations were conducted and analysed for the stability of ligand in the binding site and binding free energies were estimated. As evident from simulation trajectory analysis compound **6** occupied the primary galectin-8_N binding site with no significant fluctuations in the ligand placement (Figure 4.6). However, the compound **6** showed significant fluctuations in case of

the galectin-8C binding site, possibly due to lack of the long S3-S4 loop bearing unique residue like Arg59 (Figure 4.6). The average estimated ligand binding free energy for the galectin-8C-**6** complex was therefore observed to be only half (-25.5 kcal/mol) as compared to that estimated for the galectin-8N-**6** complex (-60.6 kcal/mol). Overall, our simulation results and binding free energy analysis suggest favourability of compound **6** towards galectin-8N compared to galectin-8C.

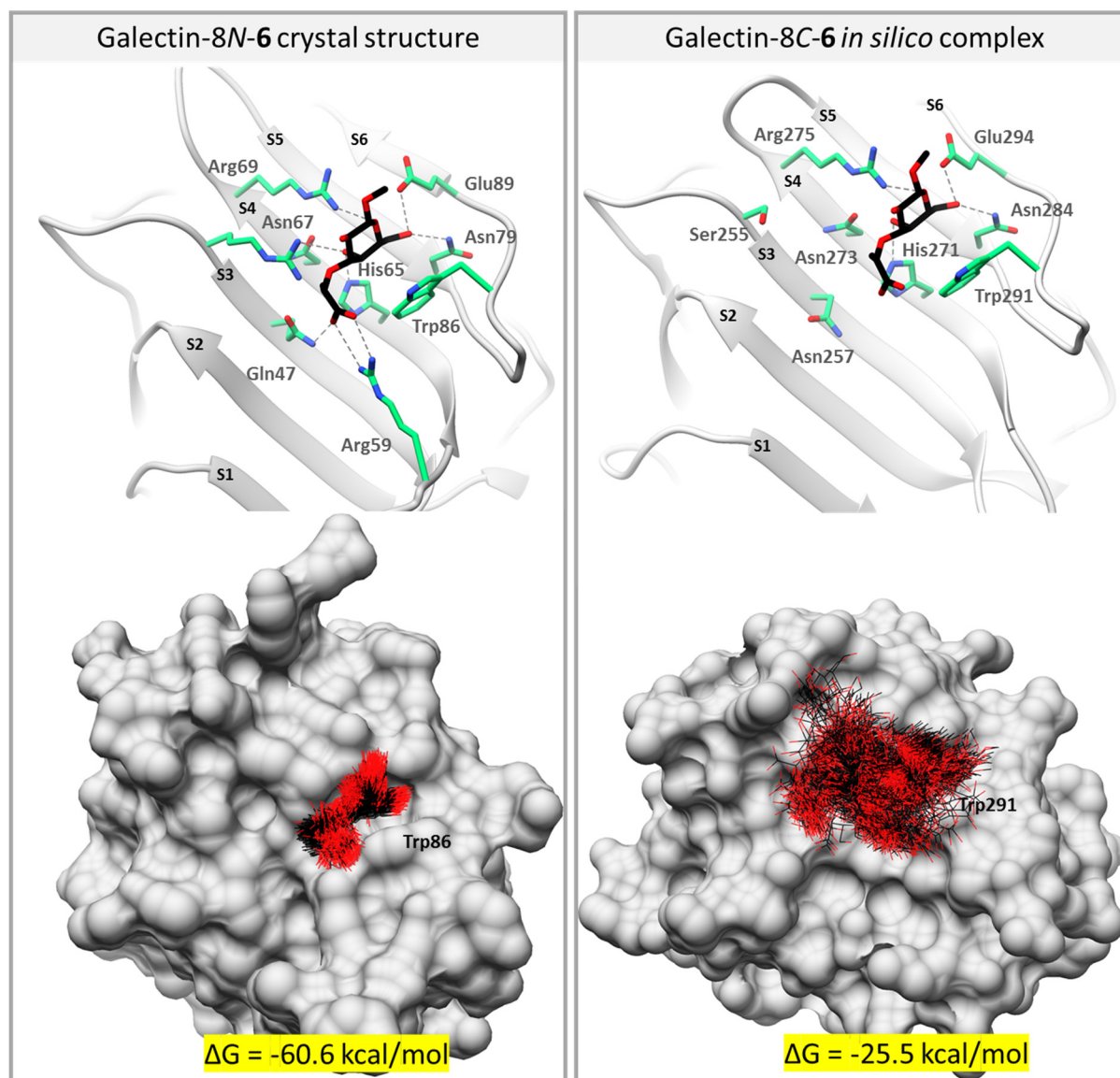


Figure 4.6: Binding mode comparison of compound **6** towards the galectin-8N (crystal structure [5VWG]; on the left) and the galectin-8C (*in silico* complex; on the right). On top, key binding site residues are represented in sticks (carbon in green for protein atoms and carbon in black for ligand atoms) and hydrogen bonds and salt bridge in grey dashed lines. At bottom, overlaid ligand (in the wire; carbon in black and oxygen in red) coordinates extracted from 100 ns simulation bound to galectin-8N (bottom right) and galectin-8C (bottom left). The MMPBSA estimated binding free energy (in kcal/mol) is highlighted in yellow.

4.3.5 Conclusions

In summary, we have employed knowledge from our virtual screening and from existing structural information to rationally design a monosaccharide-based ligand to target galectin-8N. The viability of the anionic moiety in the galectin-8N binding site was one of the main outcomes from our *in silico* virtual screening, along with its preferential recognition of anionic saccharides. Considering those results, the most frequent hydrogen bonding interactions from MD simulations, and the synthetic feasibility, compound **6** was designed. Subsequently, it was synthesised and tested for binding to galectin-8N through ITC and X-ray crystallography. The X-ray structure of galectin-8N-**6** complex validated our predicted conformation, where the ligand explored both the evolutionarily conserved and the unique amino acids of galectin-8N for interaction. The exploitation of unique residues for interaction by compound **6** suggested its favourability in the galectin-8N over the galectin-8C as noted from our MD simulations. The promising binding affinity observed for our monosaccharide-based ligand **6** (K_d 139.5 μ M) prompts further modifications to explore various functionalities on the galactose core and investigation of resulting effects on the binding affinity to galectin-8N. Overall, we have initiated a rational ligand design campaign by employing a combination of *in silico* and experimental approaches to successfully identify a monosaccharide-based scaffold that binds to galectin-8, and that will undergo ligand optimisation aimed at enhancing binding affinity and specificity towards galectin-8.

4.3.5 Acknowledgments

H.B gratefully acknowledges the financial support from the Cancer Council Queensland (ID1080845).

Accession code: Protein Data Bank: Atomic coordinates and structure factors have been deposited with accession code 5VWG for the galectin-8N-Methyl 3-*O*-[1-carboxyethyl]- β -D-galactopyranoside complex.

4.4 Supplementary Information

Structure-based Design of a Monosaccharide Ligand Targeting Galectin-8

Mohammad H. Bohari¹, Xing Yu¹, Rob-Marc Go¹ Yaron Vinik³ I. Darren Grice^{1,2}, Yehiel Zick³ and

Helen Blanchard^{1*}

¹Institute for Glycomics, Griffith University, Gold Coast Campus, 4222, Australia.

²School of Medical Science, Griffith University, Gold Coast Campus, 4222, Australia

³Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel.

*To whom correspondence should be addressed: Professor Helen Blanchard

Institute for Glycomics, Griffith University, Gold Coast Campus, 4222, Australia.

Phone: +61 7 555 27023; Fax: +61 7 555 28098; E -mail: h.blanchard@griffith.edu.au

4.4.1 Experimental section

4.4.1.1 Molecular dynamics (MD) simulations

The initial coordinates of the designed compound **6** bound to galectin-8N were obtained by modifying the lactose from the galectin-8N-lactose complex (5T7S [81]) using the BIOVIA Discovery Studio Visualiser [170]. *In silico* galectin-8C-**6** complex was generated by removing the bound peptide and superimposing the galectin-8C crystal structure (4GXL [49]) on to the galectin-8N-**6** crystal structure using the MatchMaker utility of UCSF Chimera [131]. All MD simulations were performed using the GROMACS 4.5.6 version [171] with AMBER99SB-ILDN force field [137], as employed previously [81, 110]. Long-range electrostatics were handled using Particle Mesh Ewald method [172]. Ligand topology and parameters were generated by applying AM1-BCC charges and Generalised Amber Force Field [140] using a python script Acpye [139] that uses Antechamber module of AMBER [155]. The protein-ligand complexes were initially energy minimised using the steepest descent method, followed by position restrained minimisation and finally the 100 ns production run. Hydrogen bond analysis was performed using the g_hbond utility available with the GROMACS package, and the occupancy analysis was performed using the python script readHBmap, written by R. O. S. Soares. Visualisation of MD trajectories was carried out in VMD [173]. The ligand binding free energy was estimated using molecular mechanics energies with the Poisson-Boltzmann and surface area continuum solvation method (MMPBSA.py) [174, 175] implemented in Amber package [156]. A set of 100 frames periodically extracted from the trajectory file at an interval of 1 ns were subjected to MMPBSA analysis to obtain the ligand binding free energies.

4.4.1.2 Synthesis

General procedure: Thin Layer Chromatography (TLC) on pre-coated aluminium-backed silica plates (60 F₂₅₄; Merck) were used to assess reactions and visualised by charring in 4% sulphuric acid in ethanol. Reaction products were purified using flash chromatography silica gel 60. ¹H and ¹³C NMR spectra were recorded at 298 K using an Avance (400 MHz and 100 MHz, respectively; Bruker Biospin) spectrometer. Two-dimensional COSY (¹H to ¹H correlation), HSQC (¹H to ¹³C correlation) and HMBC (¹H to ¹³C long range correlation) NMR experiments were used to assist in assigning relevant peaks for ¹H and ¹³C NMR spectra. Electrospray ionisation low-resolution mass spectrometry was performed using a Bruker Daltronics esquire 3000 Ion-Trap instrument.

Methyl 4,6-O-benzilidene-β-D-galactopyranoside [165] (**1**): Methyl-β-D-galactopyranoside (1 g, 5.1 mmol) was dried under high vacuum overnight before being dissolved in anhydrous acetonitrile (10 mL) under argon. A catalytic amount of camphor sulfonic acid (52 mg, 0.22 mmol) was added followed by benzaldehyde dimethyl acetal (1.5 mL, 14.8 mmol) added dropwise. The reaction was heated to 60 °C for about 3 hours. The reaction was quenched with Et₃N, purified via flash chromatography using 40:1 Dichloromethane (DCM):MeOH to yield **1** (88%). ¹H NMR (CDCl₃, 400 MHz): δ = 3.52 (1H, br. q), 3.59 (3H, s), 4.01 (1H, q, *J*=2.52, 10.28 Hz), 4.07 (1H, dd, *J*=1.84, 12.48 Hz), 4.29 (1H, d, *J*=7.76 Hz), 4.34 (1H, dd, *J*=1.52, 12.44 Hz), 4.40 (1H, dd, *J*=0.76, 3.64 Hz), 4.85 (1H, dd, *J*=3.68, 10.24 Hz), 5.50 (1H, s), 7.35 (3H, m), 7.49 (2H, m).

Methyl 3-O-acetyl 4,6-O-benzilidene-β-D-galactopyranoside [166] (**2**): Compound **1** was dissolved in anhydrous DCM, cooled to -20 °C using ice-salt mixture. Freshly prepared silver oxide [176] was added and left for 30 minutes, followed by slow addition of acetyl chloride and KI [166]. The reaction was left stirring overnight at room temperature. Silver oxide was filtered off, and the solvent was evaporated. The product was purified by flash chromatography (1:1 hexane:EtOAc) to yield **2** (70% yield). ¹H NMR (CDCl₃, 400 MHz): δ = 2.11 (3H, s), 2.52 (1H, br.), 3.48 (1H, d, *J*=1.1 Hz), 3.56 (3H, s), 3.98 (1H, app. t, *J*= 8.96), 4.06 (1H, dd, *J*=1.84, 6.24 Hz), 4.26 (1H, d, *J*=7.72 Hz), 4.31 (1H, dd, *J*=1.48, 11.98 Hz), 4.36 (1H, dd, *J*=0.64, 3.6 Hz), 4.82 (1H, dd, *J*=3.64, 10.24 Hz), 5.48 (1H, s), 7.34 (3H, m), 7.47 (2H, m).

Methyl 2-O-methoxymethyl-3-O-acetyl-4,6-O-benzilidene-β-D-galactopyranoside (**3**): The methoxy methyl ether protection procedure was adapted from reference [167]. Compound **2** was dissolved in DCM under argon at room temperature, diisopropylethylamine (DIPEA) was added at 0 °C followed by drop-wise addition of bromomethyl methyl ether and refluxed overnight. The reaction was diluted with DCM and washed with water and brine solution, then purified using flash column chromatography (2:1 hexane:EtOAc) to give **3** (80% yield). ¹H NMR (CDCl₃, 400 MHz): δ = 2.10 (3H, s), 3.38 (3H, s), 3.47 (1H, d, *J*=1.08 Hz), 3.55 (3H, s), 3.95 (1H, q, *J*=2.4, 10.16 Hz), 4.04 (1H, dd, *J*=1.76, 12.40 Hz), 4.31 (1H, dd, *J*=1.52 Hz), 4.33 (1H, d, *J*=2.2 Hz), 4.35 (1H, dd, *J*=0.68, 3.68 Hz), 4.68 (1H, d, *J*=6.44 Hz), 4.85 (2H, m), 5.48 (1H, s), 7.35 (3H, m), 7.50 (2H, m). ¹³C NMR (CDCl₃, 100 MHz): δ = 21.10, 55.74, 56.93, 66.14, 68.98, 72.74, 73.50, 97.15, 101.08, 104.10, 126.43, 128.13, 129.03, 137.67, 170.79. MS (ESI): *m/z* calculated for C₁₈H₂₄NaO₈ [M+Na]⁺ 391.2, found 398.2.

Methyl 2-O-methoxymethyl-4,6-O-benzilidene-β-D-galactopyranoside (**4**): Compound **3** was dissolved in methanol and cooled to 0 °C before addition of sodium metal previously suspended in hexane. The reaction was left at room temperature for 1.5 hours, then carefully acidified to

pH 5 using 1 M HCl. Salts were removed by water washing, and the product was extracted with ethyl acetate, then solvent removed to give **4** (90% yield). ^1H NMR (CD_3OD , 400 MHz): δ = 3.41 (3H, s), 3.44 (1H, s), 3.54 (3H, s), 3.56 (1H, s), 3.66 (2H, app., s), 4.06 (1H, d, J =12.16 Hz), 4.20 (1H, s), 4.25 (1H, d, J =6.16 Hz), 4.31 (1H, d, J =12.12 Hz), 4.75 (1H, d, J =6.4 Hz), 4.83 (1H, d, J =6.41 Hz), 5.53 (1H, s), 7.33 (3H, m), 7.49 (2H, m). ^{13}C NMR (CD_3OD , 100 MHz): δ = 54.72, 55.83, 66.54, 68.71, 71.80, 75.63, 76.33, 96.73, 101.13, 104.59, 126, 127, 128.51, 138.28. MS (ESI): m/z calculated for $\text{C}_{16}\text{H}_{22}\text{NaO}_7$ $[\text{M}+\text{Na}]^+$ 349.1, found 349.1.

Methyl 3-O-[1-carboxyehtyl]- β -D-galactopyranoside (6): The propionic acid side chain was installed onto **4** using previously reported conditions [168]. Compound **4** was dissolved in anhydrous 1,4-dioxane under argon and cooled to 0 °C before addition of NaH. 2-Chloropropionic acid was slowly added at 0 °C, then the reaction was stirred at 50 °C for 36 hours to yield the enantiomeric mixture of novel ligand **5** (60% yield). ^1H NMR (CD_3OD , 400 MHz): δ = 1.66 (3H, d, J =6.92 Hz), 3.44 (3H, s), 3.54 (1H, d, J =0.84 Hz), 3.56 (3H, s), 3.61 (1H, dd, J =3.4, 9.8 Hz), 3.71 (1H, q, J =7.72, 9.8 Hz), 4.14 (1H, dd, J =1.8, 12.4 Hz), 4.21 (1H, dd, J =1.56, 12.4 Hz), 4.36 (1H, d, J =7.76 Hz), 4.40 (1H, q, J =6.92, 13.84 Hz), 4.50 (1H, d, J =3.0 Hz), 4.78 (1H, d, J =6.32 Hz), 5.62 (1H, s), 7.36 (3H, br), 7.56 (2H, m). ^{13}C NMR (CD_3OD , 100 MHz): δ = 21.32, 54.98, 56.33, 66.42, 68.80, 74.81, 76.02, 79.34, 97.02, 101.07, 103.95, 126, 127, 128.49, 138.23. MS (ESI): m/z calculated for $\text{C}_{19}\text{H}_{26}\text{NaO}_9$ $[\text{M}+\text{Na}]^+$ 421.2, found 421.3.

The deprotection of **5** was carried out in methanol and concentrated HCl at room temperature. The enantiomeric mixture of novel ligand **6** was purified using reversed-phase chromatography (C_{18} , 4:1 water:methanol). ^1H NMR (CD_3OD , 400 MHz): δ = 1.35 (3H, d, J =6.92 Hz), 3.27 (2H, app. d, J =1.6 Hz), 3.40 (1H, app. q, J =5.4, 6.44 Hz), 3.44 (4H, s), 3.53 (1H, q, J =7.8, 9.56 Hz), 3.66 (3H, m), 3.92 (1H, dd, J =0.8, 3.24 Hz), 4.05 (1H, m), 4.35 (1H, q, J =6.92, 13.84 Hz). ^{13}C NMR (CD_3OD , 100 MHz): δ = 17.63, 55.84, 60.98, 67.18, 70.72, 74.89, 75.15, 82.21, 104.46. MS (ESI): m/z calculated for $\text{C}_{10}\text{H}_{18}\text{KO}_8$ $[\text{M}+\text{K}]^+$ 305.2, found 306.1.

4.4.1.3 Protein expression and purification

The galectin-8N protein was expressed in a untagged form as described previously [81]. Briefly, the bacterial culture was induced with IPTG for 4 hours at room temperature and purified using affinity chromatography on a lactosyl-Sepharose column at 4 °C. The purity of the expressed protein was assessed using SDS-PAGE and was used directly for binding assays and crystallisation.

4.4.1.4 Isothermal Calorimetry (ITC)

Quantitation of the binding affinity was done by measuring the dissociation constant using NanoITC (TA Instruments). A solution of 50 μL of 1 mM ligands (Lactose and **6**) was titrated in aliquots of 2.5 μL into the calorimetric cell, with stirring speed of 150 rpm, containing 300 μL of 200 μM galectin-8N. Injections were performed at an interval of 300 s at 298 K. All the protein and ligand samples were prepared in PBS. A new experiment consisting of ligand being titrated into the buffer (blank experiment) was also performed and subtracted from actual binding experiment before analysing the thermograms. The thermodynamic analysis was conducted using an independent model with NanoAnalyze software.

4.4.1.5 X-ray data collection and structure determination

The galectin-8N-**6** complex structure was obtained by soaking compound **6** into the *apo* galectin-8N crystals, as performed previously [81]. The *apo* galectin-8N crystals were generated in phosphate buffer saline (10 mM sodium phosphate, 137 mM sodium chloride, 2.7 mM potassium chloride, 1.8 mM potassium phosphate; PBS). Compound **6** was dissolved in PBS at a concentration of 10 mM and soaked into the *apo* crystals for 5 mins. X-ray diffraction data were collected at room temperature. The crystals were mounted on the goniometer using the MicroRT capillary system from MiTigen. A Rigaku MicroMaxTM-007 HF rotating anode generator coupled with VariMax optics and shutter-less PILATUS 200 K detector was used to perform the experiments. HKL-3000R[177] was used to control the instrument, and HKL-3000R and iMosflm [143] were used for data processing including indexing, integration and scaling. PHASER [145] implemented in CCP4 [148] was used for molecular replacement with the *apo* galectin-8N (3AP5 [76]) structure used as the search model. The chemical information file for **6** was generated using the PRODRG server [178]. Refinement was carried out using REFMAC5 [146], model building and visualisation done in COOT [149]. Final model validation performed using MolProbity [179]. Molecular graphics and electron density illustrations for figures were performed using the UCSF Chimera package [131, 180].

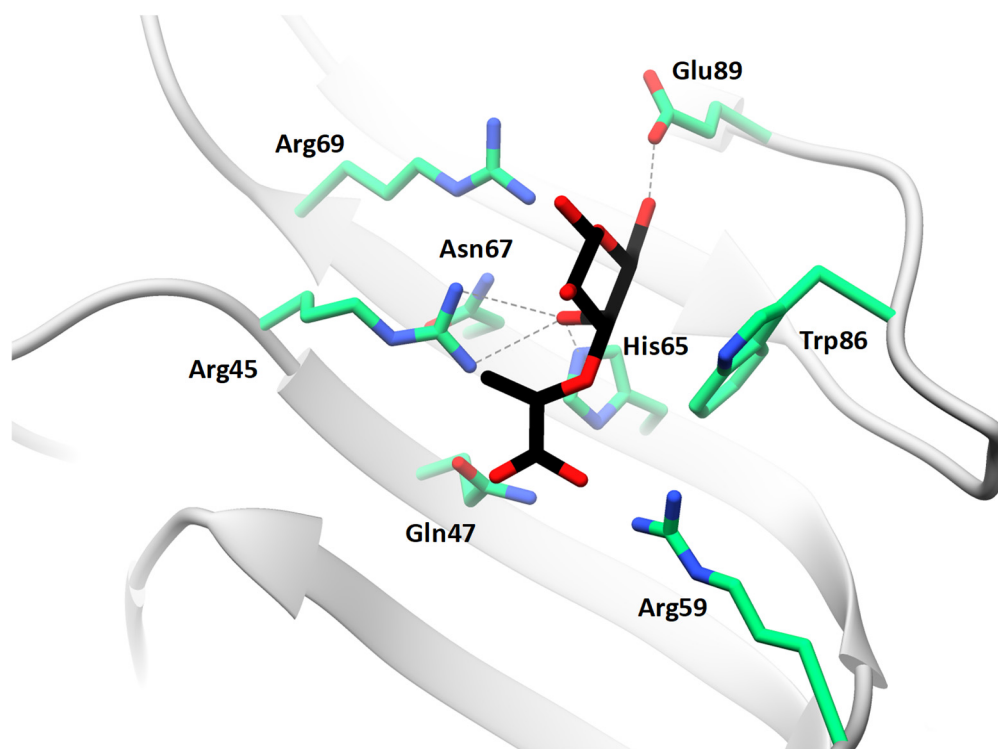


Figure S1: Predicted binding conformation. The coordinates of compound **6** placed in the galectin-8N binding site before performing simulations. Hydrogen bonding interactions are represented in grey dashed lines.

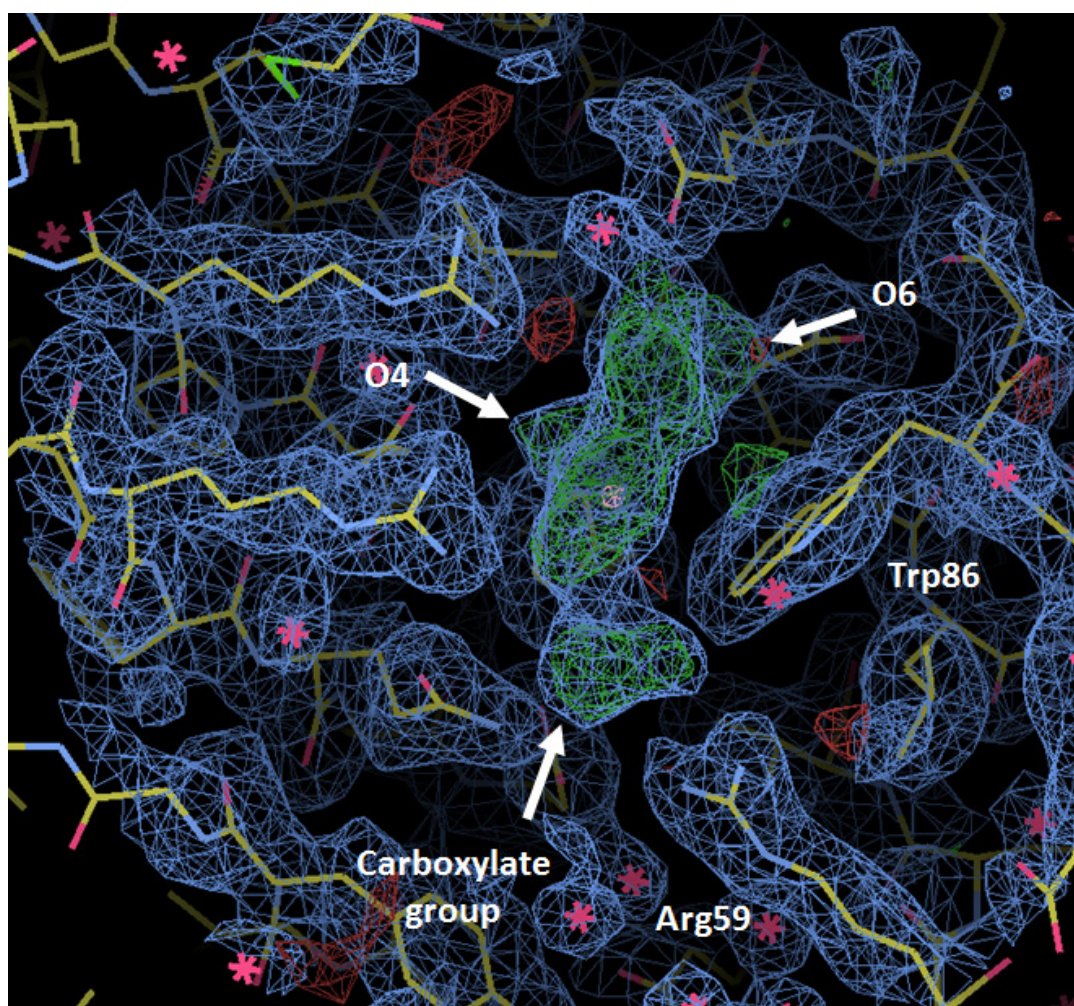


Figure S2: Omit electron density maps calculated from refinement with the compound **6** omitted from the model ($2mFo - DFc$: 1.0σ [blue], $mFo - DFc$: $\pm 3.2 \sigma$ [green/red]) in the galectin-8N-6 complex. Red crosses indicate water molecules.

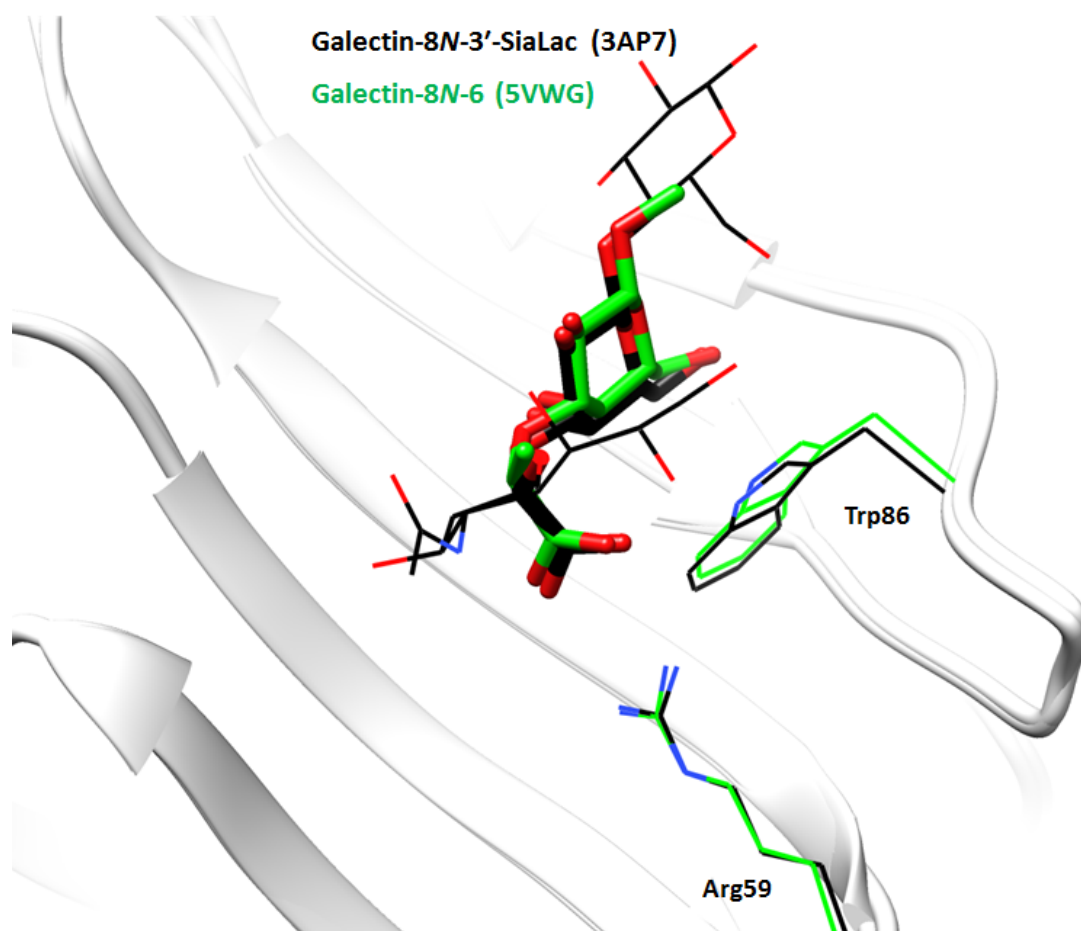
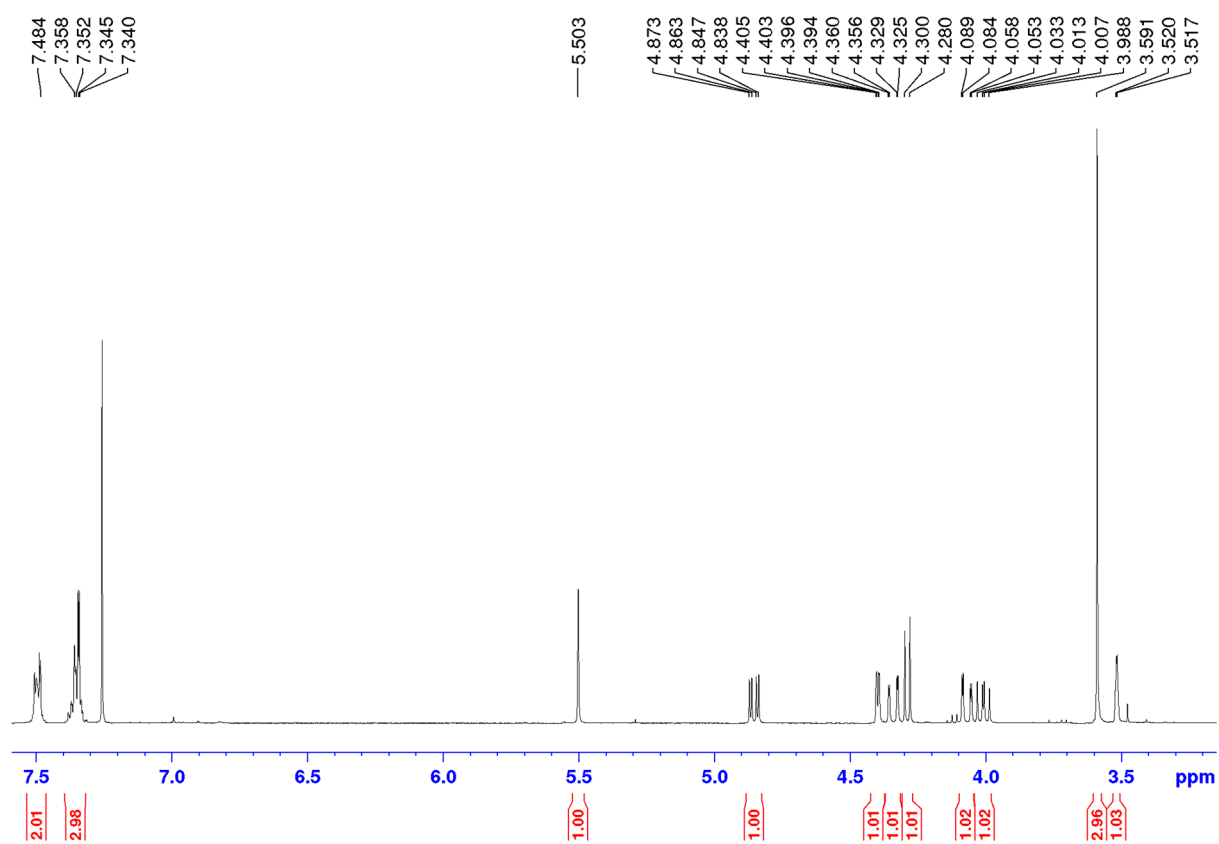


Figure S3: Comparison of the binding conformation of 6 and 3'-SiaLac bound to galectin-8N. Superimposition of the galectin-8N-6 crystal structure (5VWG; carbon in green, oxygen red and nitrogen blue) reported in the current study on the galectin-8N-3'-SiaLac crystal structure (3AP7; carbon in black, oxygen in red and nitrogen in blue). The compound **6** equivalent part of the 3'-SiaLac is represented in black sticks, all other atoms of 3'-SiaLac including the evolutionarily conserved Trp86 and unique Arg59 of the two crystal structures are represented in the wire.

4.4.2 Spectral data

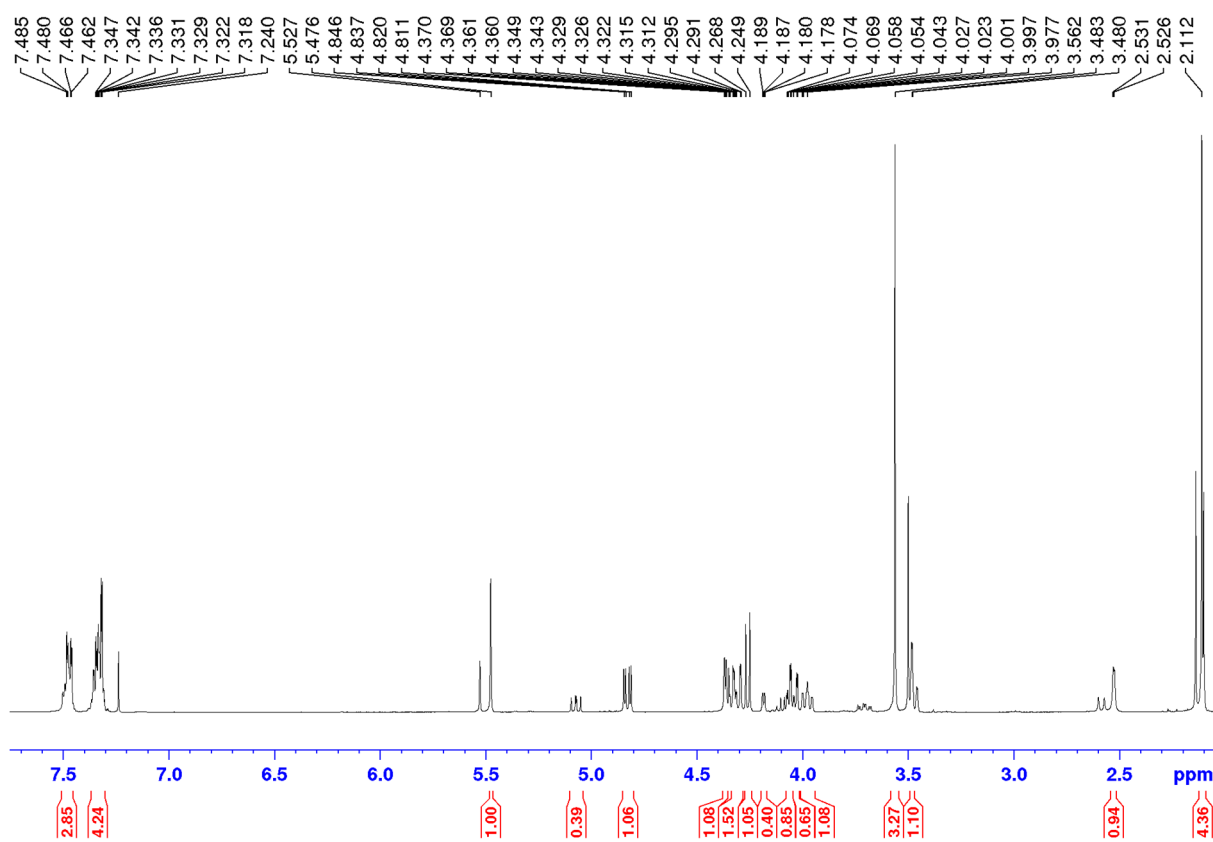
Methyl 4,6-*O*-benzilidene- β -D-galactopyranoside (1)

^1H NMR (CDCl_3 , 400 MHz)



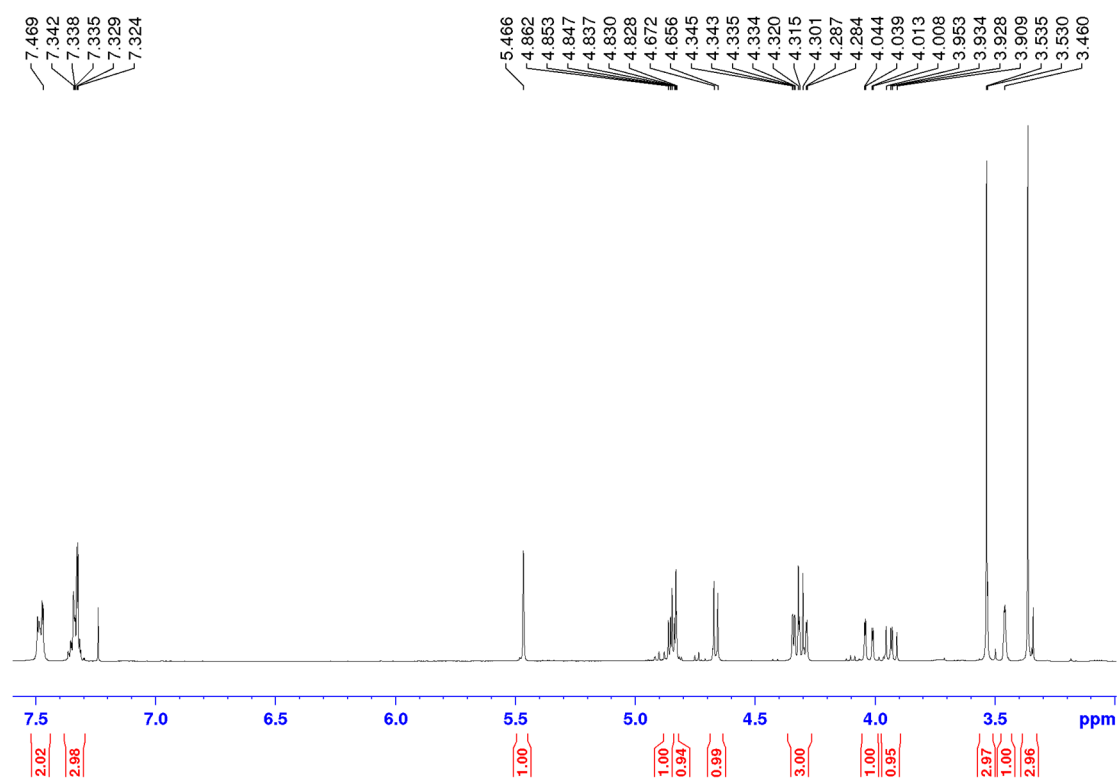
Methyl 3-*O*-acetyl 4,6-*O*-benzilidene- β -D-galactopyranoside (2)

^1H NMR (CDCl_3 , 400 MHz)

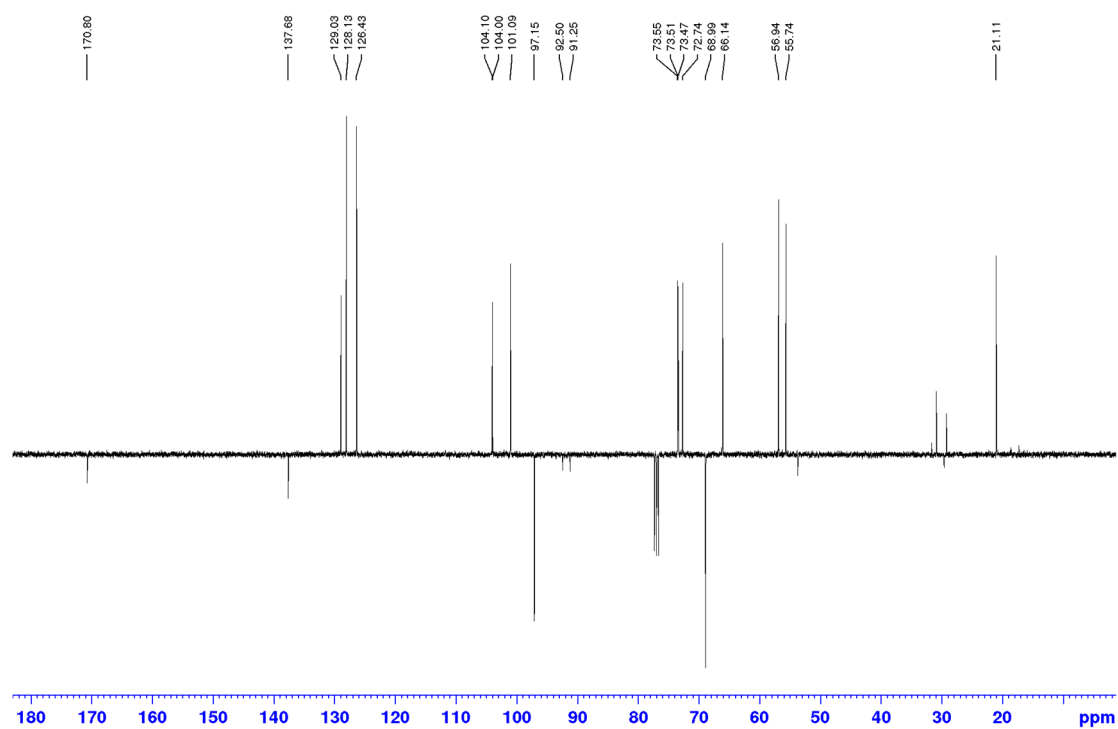


Methyl 2-*O*-methoxymethyl-3-*O*-acetyl-4,6-*O*-benzilidene- β -D-galactopyranoside (3)

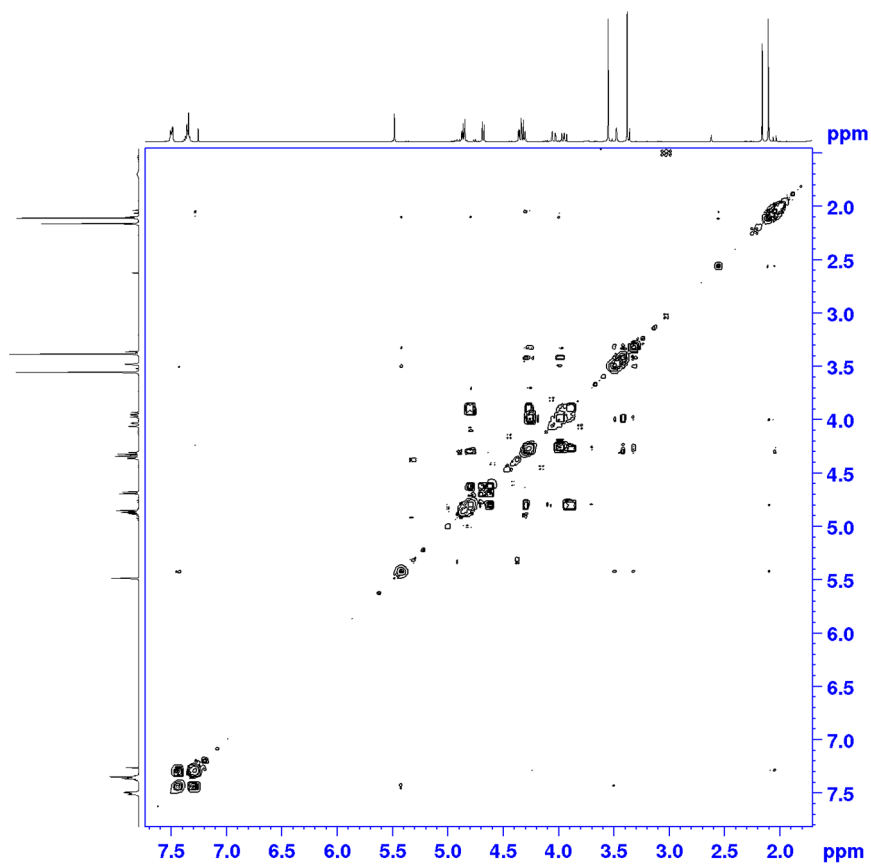
^1H NMR (CDCl_3 , 400 MHz)



^{13}C NMR

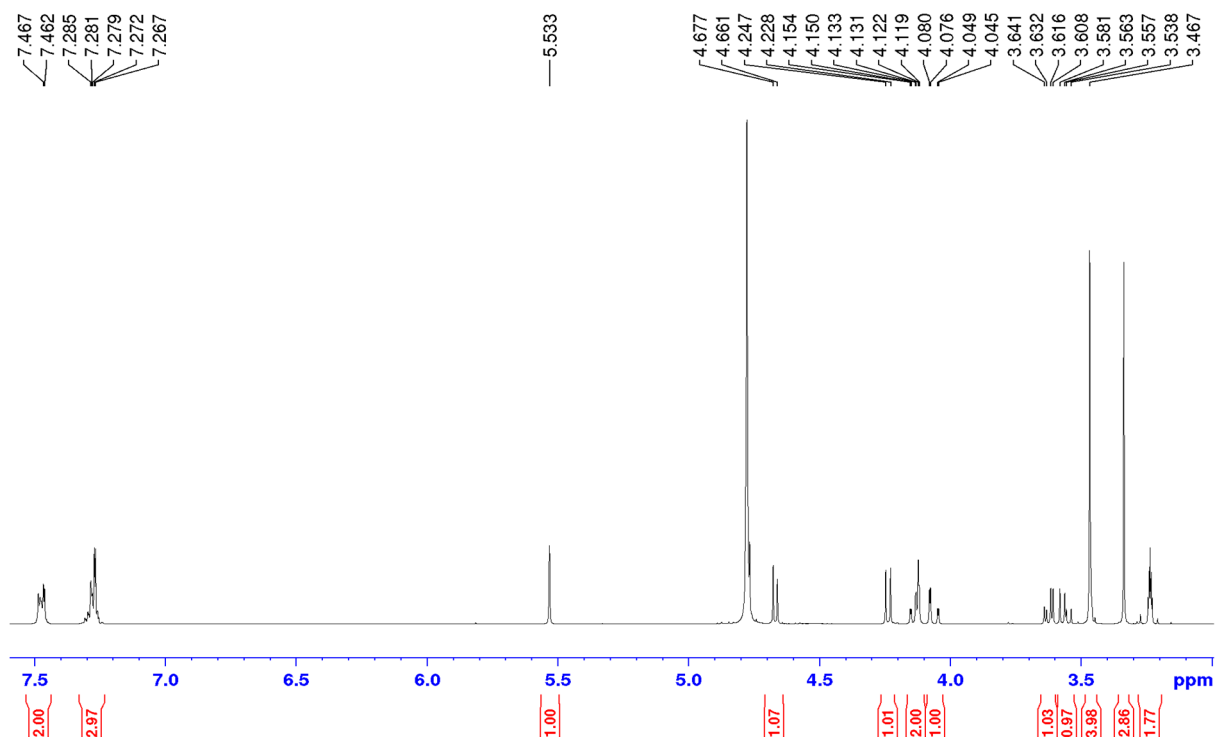


^1H - ^1H COSY

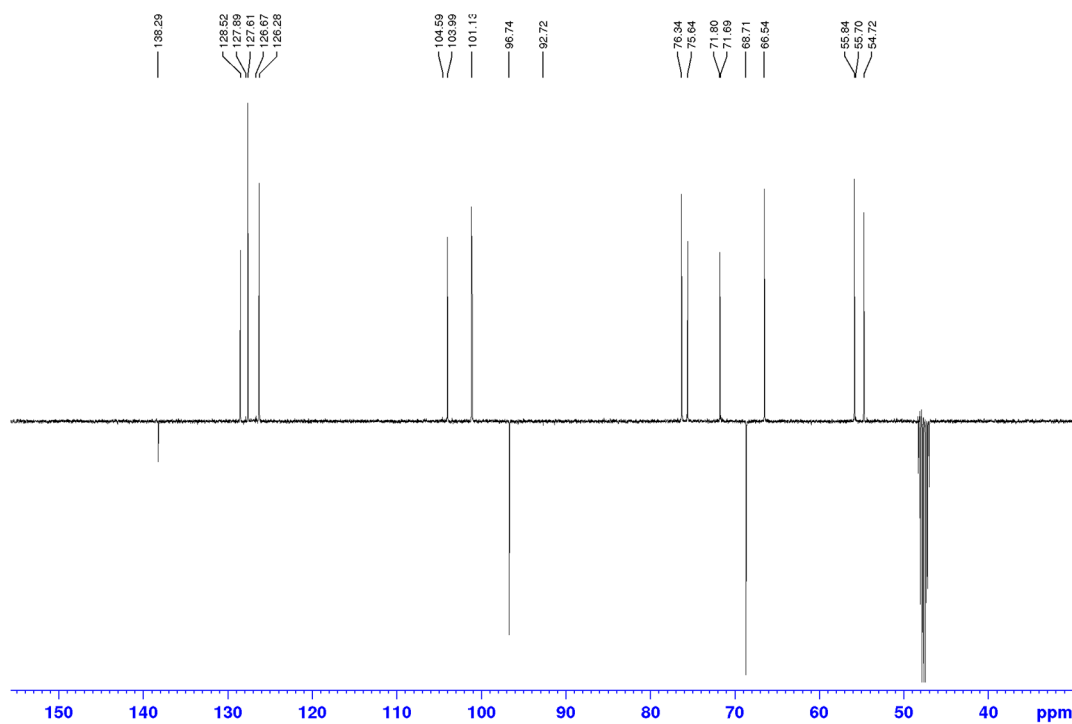


Methyl 2-*O*-methoxymethyl-4,6-*O*-benzilidene- β -D-galactopyranoside (4)

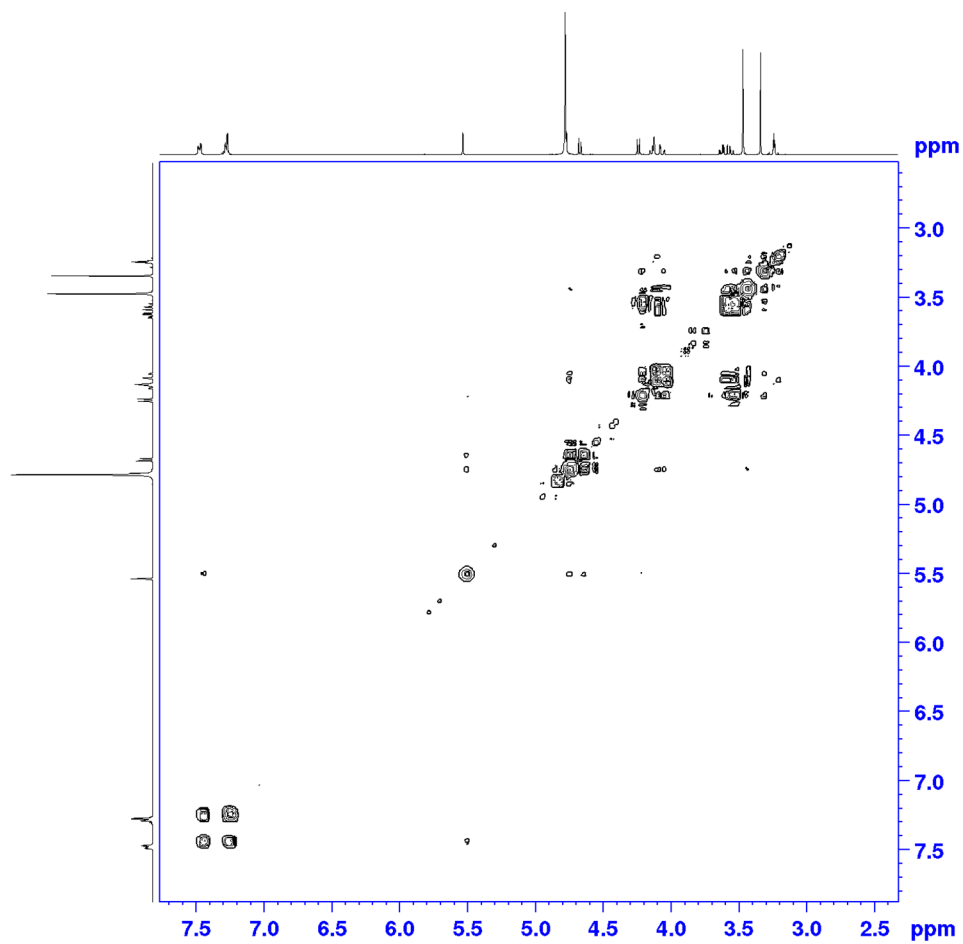
^1H NMR (CDCl_3 , 400 MHz)



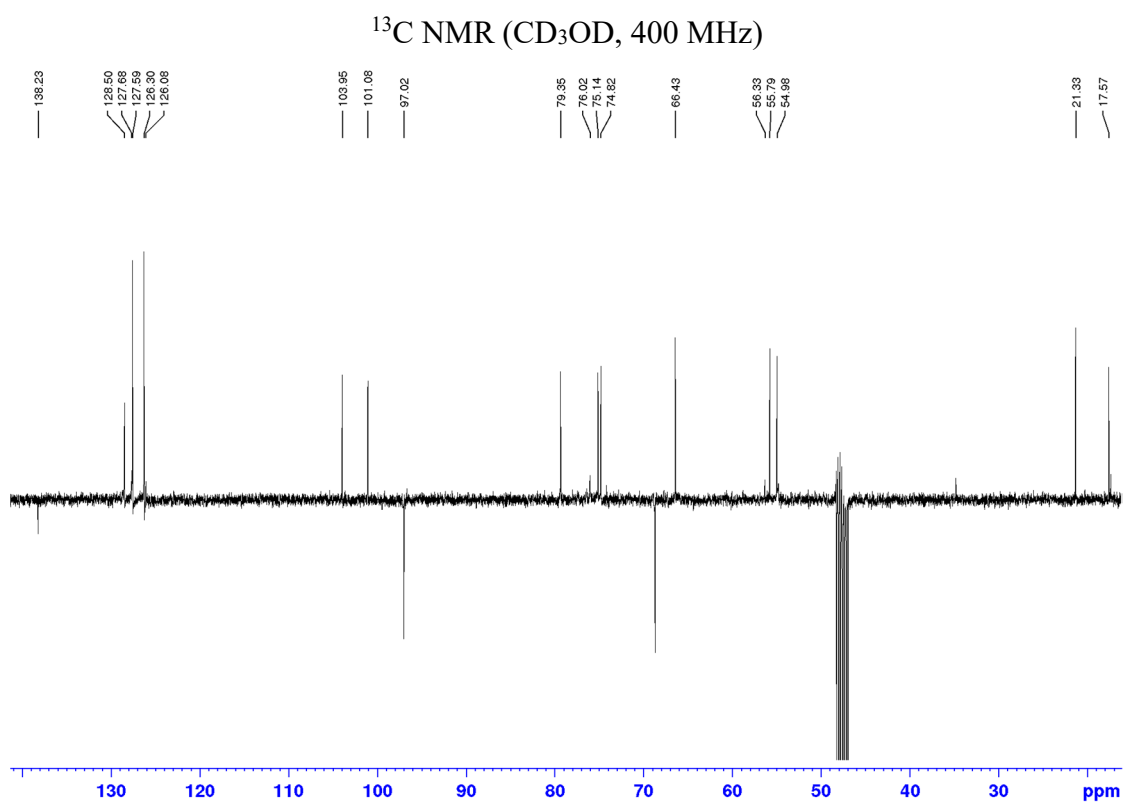
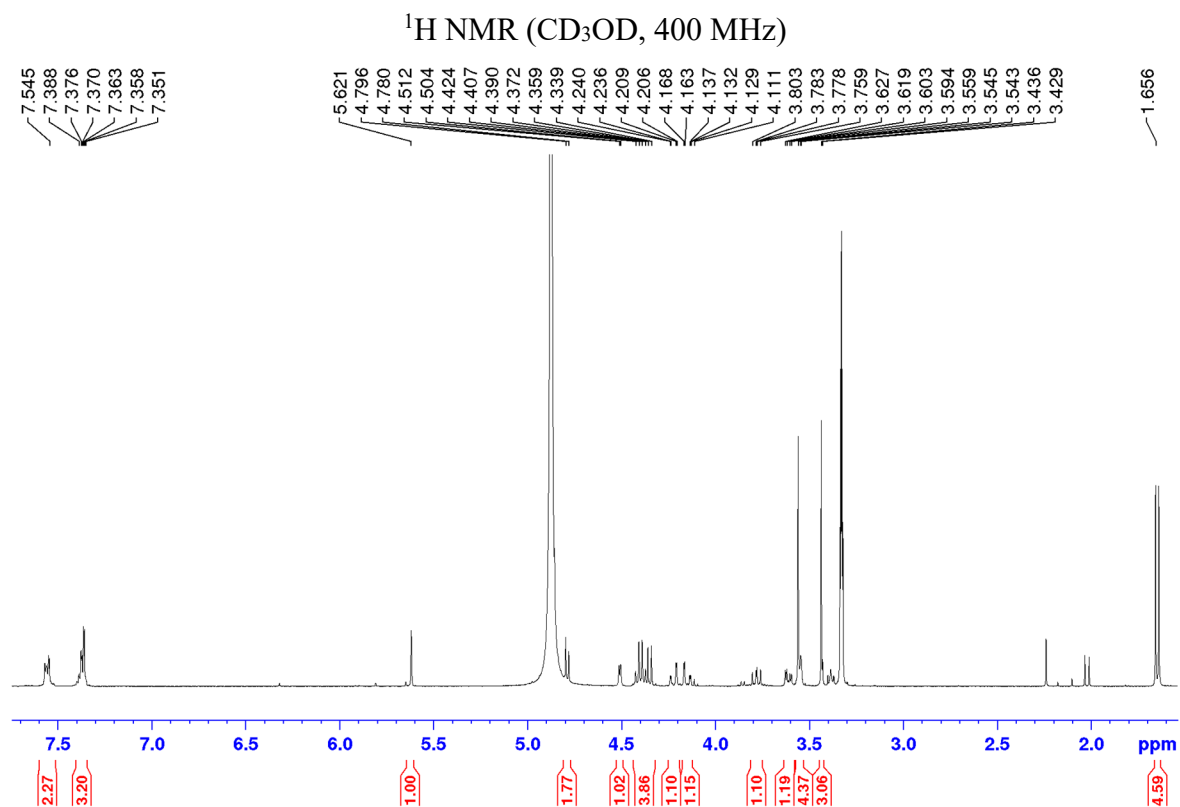
^{13}C NMR



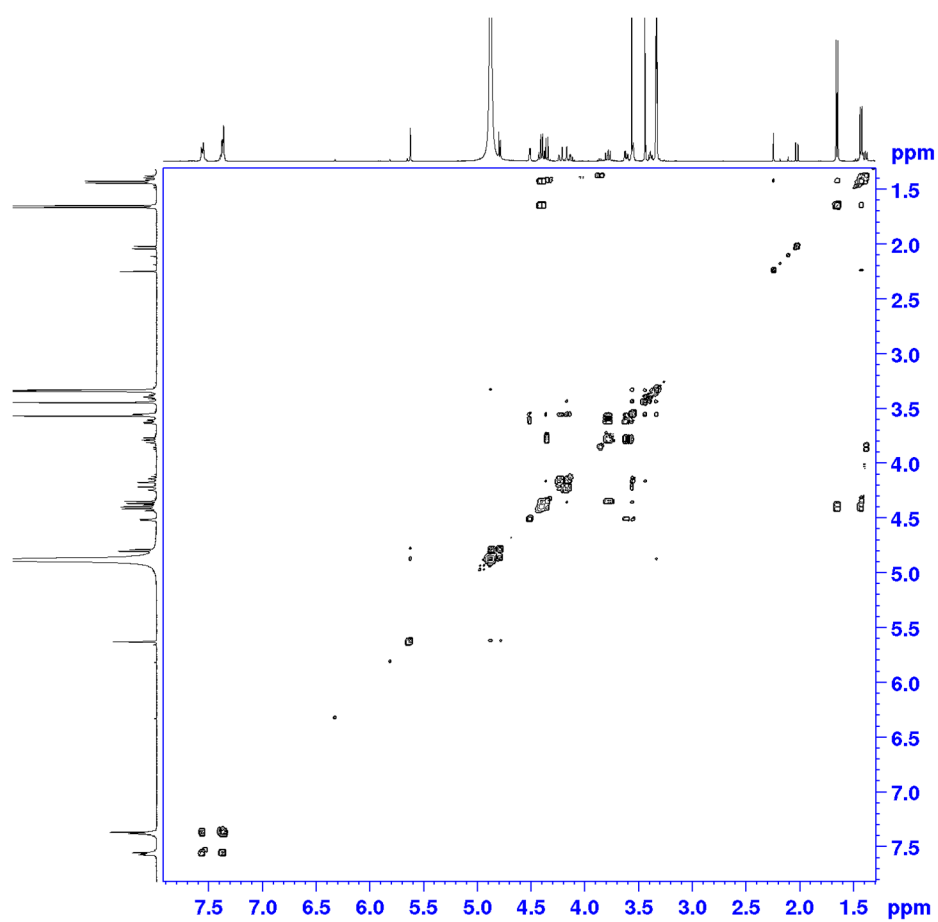
^1H - ^1H COSY



Methyl 2-*O*-methoxymethyl-3-*O*-[1-carboxyethyl]- 4,6-*O*-benzilidene-β-*D*-galactopyranoside (5)

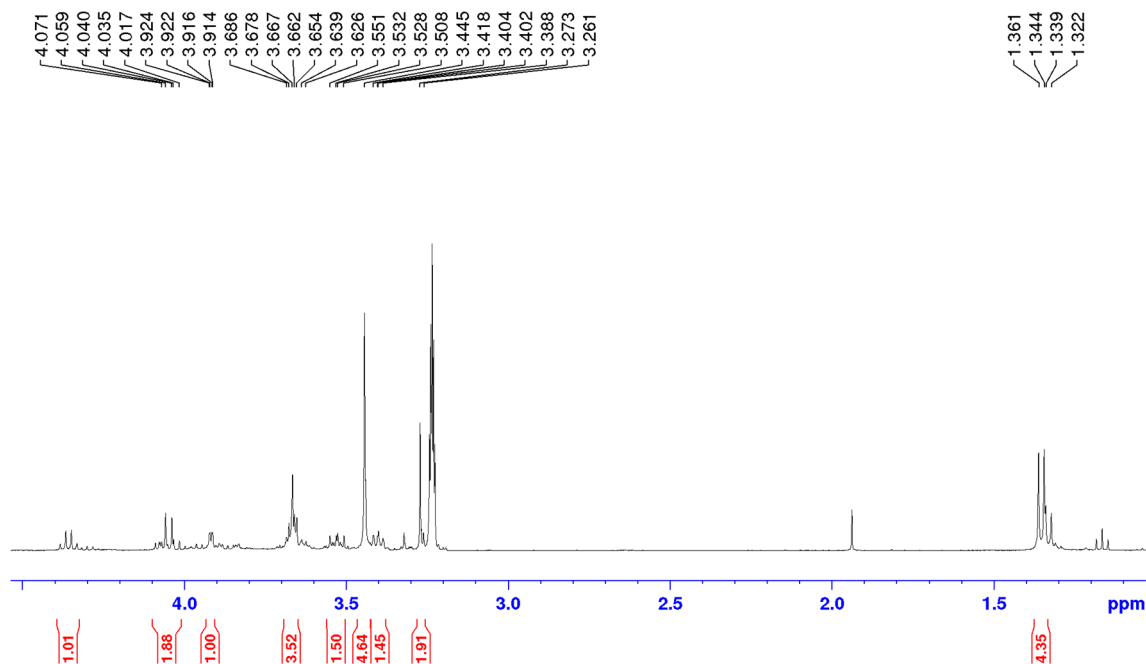


^1H - ^1H COSY



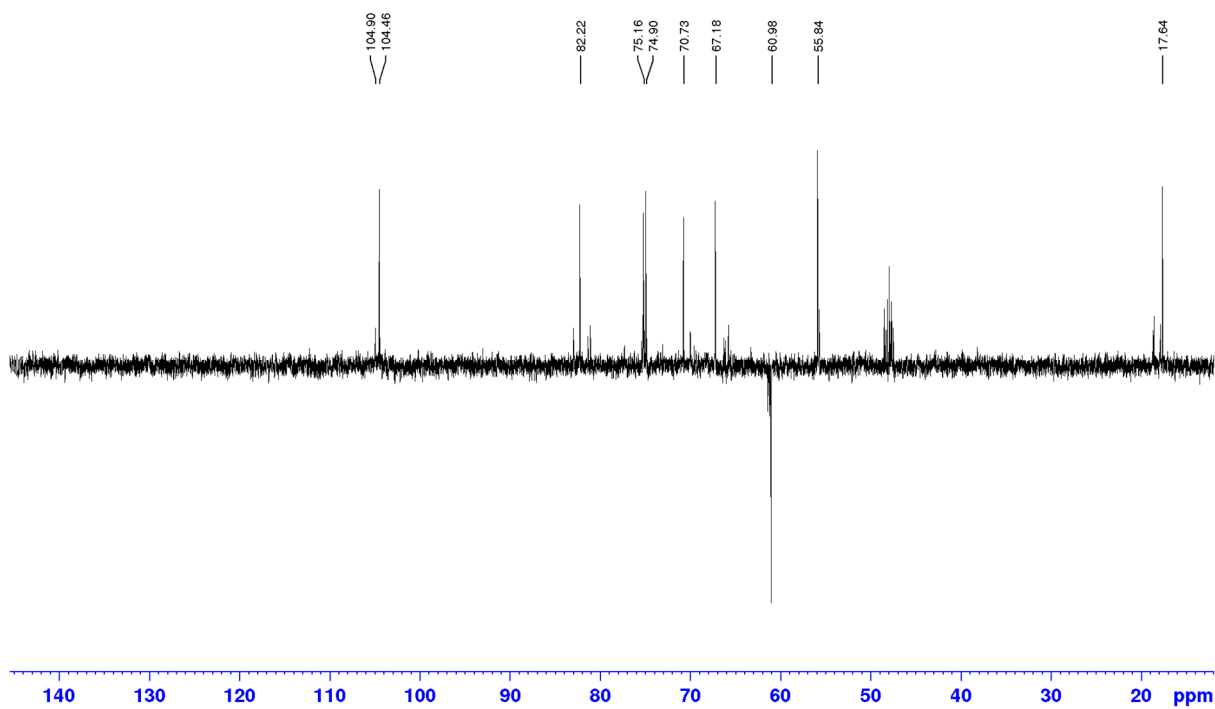
Methyl 3-*O*-[1-carboxyethyl]- β -D-galactopyranoside (6)

^1H NMR (CD_3OD , 400 MHz)

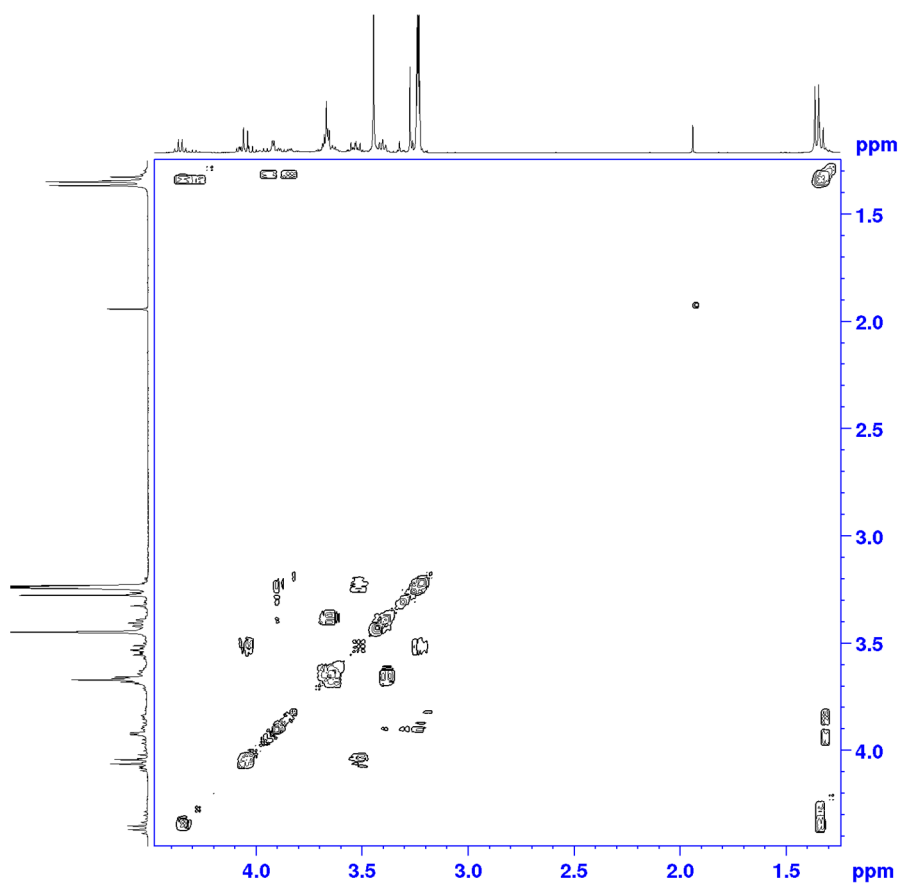


^{13}C NMR (CD_3OD , 400 MHz)

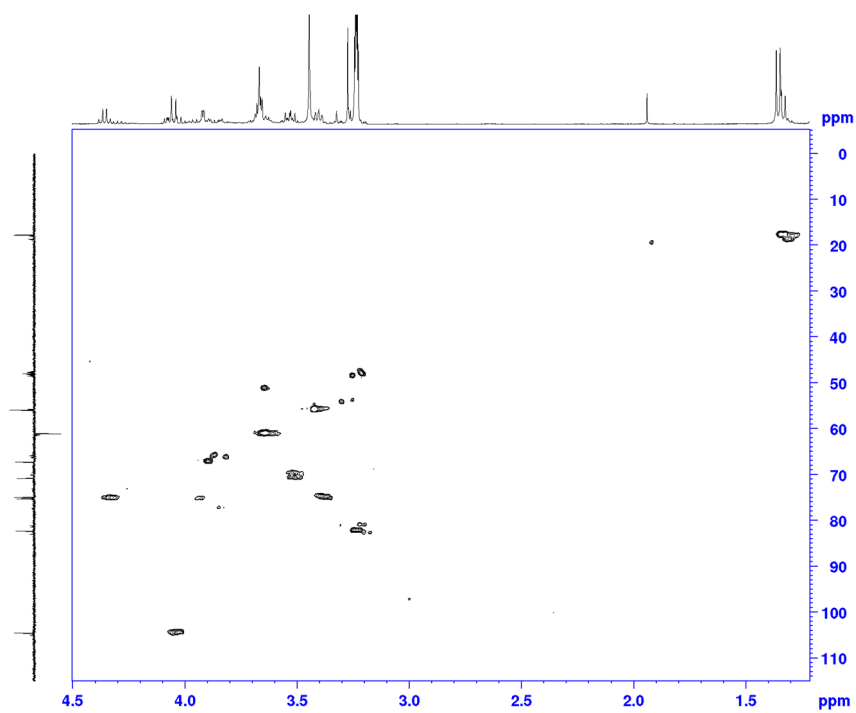
kygggggn



^1H - ^1H COSY



^1H - ^{13}C HSQC [short range]



4.4.3 Saturation Transfer Difference (STD NMR) experiments

Method:

The STD NMR spectra were acquired on Bruker 600 MHz Avance spectrometer with a conventional $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ gradient cryoprobe system at 298 K. Compound **6** in 5 mM and 1 mM concentrations mixed with 5 μM galectin-8N in 250 μL deuterated buffer containing 20 mM phosphate buffer and 20 mM sodium chloride (ligand:protein ratio of 1000:1). The data analysis was performed with the help of TopSpin 3.5p15 software package. The spectra were acquired at -1 ppm and 7.13 ppm on resonance frequency and 33 ppm off-resonance frequency with a total of 512 scans.

Results:

The binding of compound **6** to galectin-8N was examined in solution-state using STD NMR. In the ^1H spectrum of **6**, proton signals for CH and CH_3 groups of the propionic acid side chain appeared at 3.98 ppm and 1.29 ppm, respectively. The signal for the acidic proton was not observed as the carboxylic acid group would likely be deprotonated due to the pH of the buffer being slightly above neutral. The control spectrum of **6** (without the protein) showed background signals of CH and CH_3 groups (spectra in red; Figure S4) when the on-resonance frequency was set to -1.00 ppm. STD signals of CH and CH_3 groups observed in the protein-ligand STD spectra (spectra in green; Figure S4) indicate the closeness of the propionic acid side chain of **6** to galectin-8N. The carboxylic acid group was predicted to engage in interaction with the unique Arg59 (equivalent to 3'-SiaLac). This interaction may cause the chiral CH group to be closer to the protein, as evident from the 3.98 ppm (CH) and 1.29 ppm (CH_3) STD signals (Figure S4). To cross-check the binding of **6** to galectin-8N, the spectra with the on-resonance frequency of 7.13 ppm were also acquired. The control spectra with 1 mM **6** in the second experiment resulted in no background signals of CH and CH_3 groups of ligand; however, the protein-ligand STD showed no significant signals (Figure S5). Due to this discrepancy, the original experiment at -1 ppm on-resonance frequency with a lesser amount of ligand (1 mM) was repeated. Comparable result to that previously achieved (Figure S4 and Figure S6) results were obtained. The same batch of compound **6** was used to perform the experiments which over the storage might have caused a slight alteration in either its chemical purity or integrity. Nevertheless, the detected binding mode from our crystal structure and binding affinity determined through ITC indicate binding of **6** to galectin-8N. Further investigations are needed to better understand the factors responsible for the observed mixed STD NMR results.

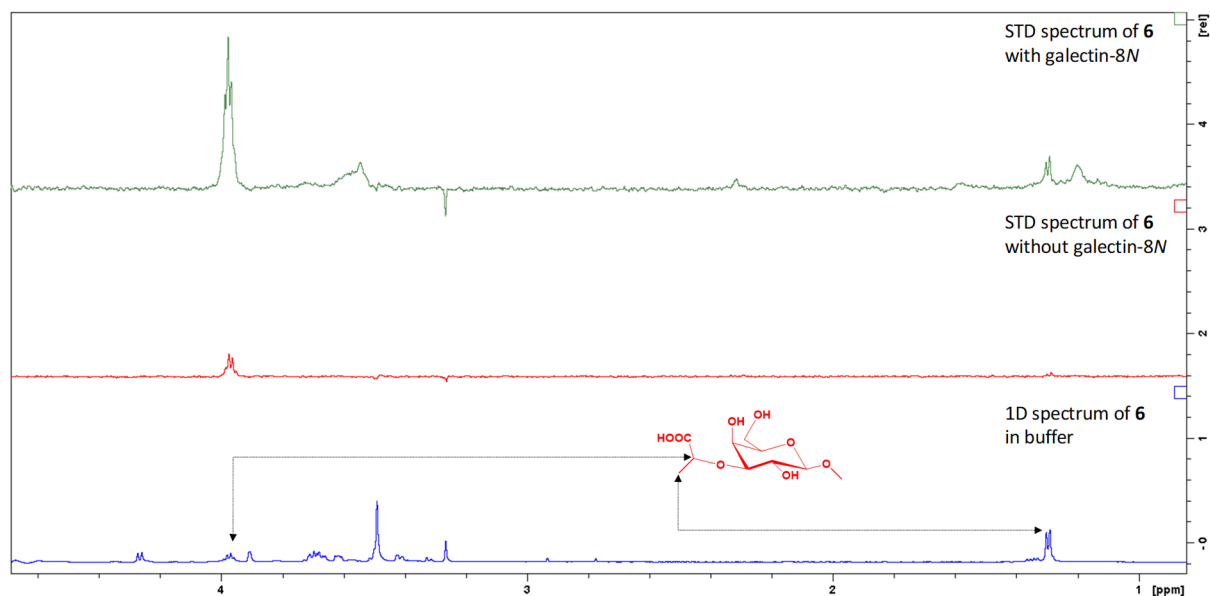


Figure S4. Investigation of ligand binding. An overlaid STD NMR spectrum of ligand with (in green) and without (in red) galectin-8N (5 μ M) over the ^1H spectrum of ligand in buffer (in blue).

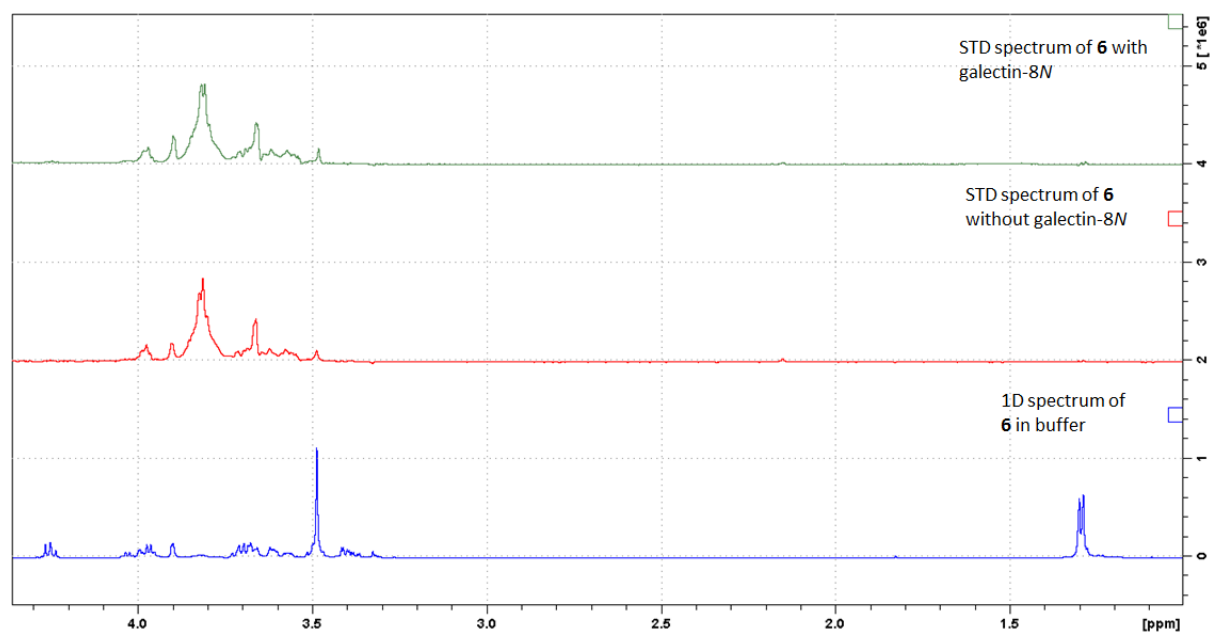


Figure S5. Investigation of ligand binding. An overlaid STD NMR spectrum of ligand with (in green) and without (in red) galectin-8N (5 μ M) over the ^1H spectrum of ligand in buffer (in blue) acquired at 7.13 on-resonance frequency.

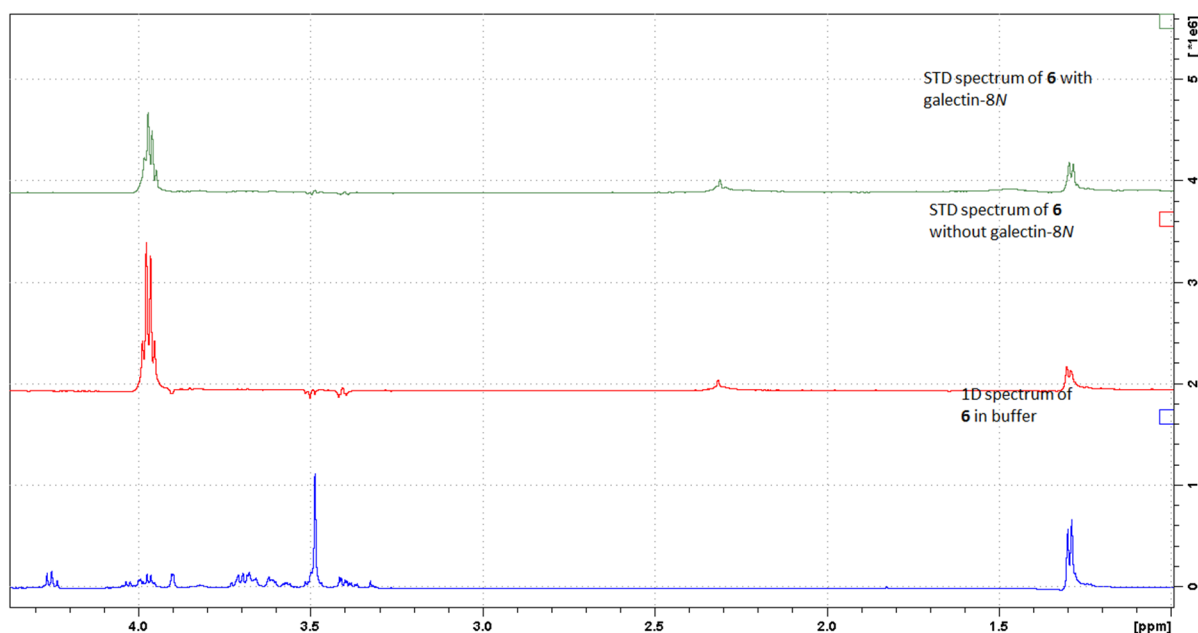


Figure S6. **A repeat experiment of ligand binding investigation.** An overlaid STD NMR spectrum of ligand with (in green) and without (in red) galectin-8N (5 μ M) over the ^1H spectrum of ligand in buffer (in blue) acquired at -1.0 ppm on-resonance frequency.

4.4.4 Galectin-8C expression clone preparation

The design approach practised in the thesis focuses on the *N*-terminal domain of the galectin-8; however, the *C*-terminal could also potentially interact with the designed ligands. Because the ligands designed to explore both the conserved and unique amino acid residues in the binding site, there is a possibility of interaction with the conserved residues of the *C*-terminal domain. However, the affinity of interaction would vary in galectin-8C, due to lack of galectin-8N specific residues. The future experiments would, therefore, explore the interactions of designed ligands and galectin-8C. Towards that direction, subcloning of galectin-8C has been performed from myc tagged pQE vector (provided by our collaborator) to the expression vector pET-3a and confirmed through sequencing (Figure 4.7). The initial trial expression did not show any significant signs of overexpression of galectin-8C, particularly in the soluble fraction on the SDS-PAGE but was rather observed in the insoluble fraction. Expression optimisation by varying parameters like ITPG concentration, expression temperature, bacterial growth temperature needs to be performed to increase the amount of soluble protein. In addition to the galectin-8C clone made, few other expression clones can be prepared including certain parts of the linker peptide as well. The expressed protein would then be utilised for crystallographic analysis and binding affinity determination either by STD NMR, SPR or ITC.

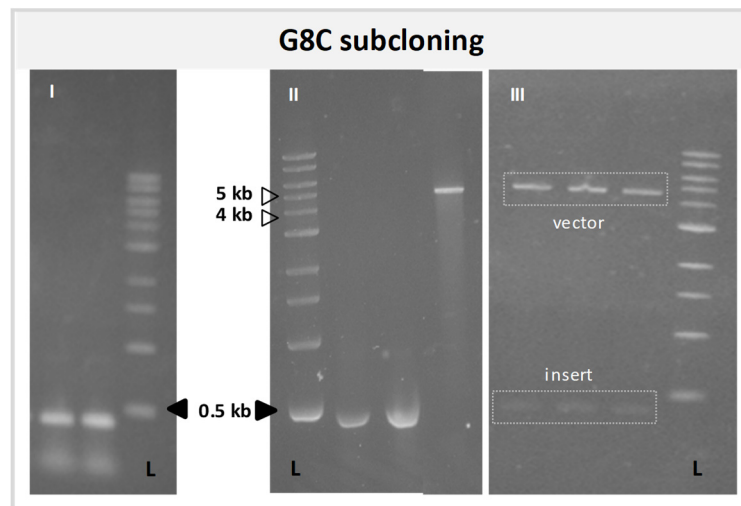


Figure 4.7: Agarose gel electrophoresis analysis of G8C: I) PCR products, II) gel-extracted vector and insert, and III) NdeI-BamHI double digested ligated products from colony screening. L is DNA ladder.

4.4.5 Preliminary biological evaluation

Galectin-8, at the molecular level, was reported to bind low-density lipoprotein receptor-related protein 1 (LRP-1) and the mannose receptor C, type 2 (MRC2) which are part of a multi-protein complex including urokinase plasminogen activator receptor (uPAR). The suggested mechanism was that galectin-8 interacts with a complex of uPAR/LRP-1/MRC2 that binds to cell surface integrins to mediate the downstream ERK signalling pathway to induce the transcription of RANKL. However, the atomic level details about the site of interaction of galectin-8 to this multi-protein complex are not known. Also, it is not clear whether it is the galectin-8C or galectin-8N or both the domains that are involved in mediating RANKL release. Therefore, crystallographic analysis can be performed on this multi-protein complex to derive galectin-8 involvement in the process. The preliminary *in vitro* data (from our collaborators) showed no inhibition of RANKL release by MB46A, the ligand which was designed to bind galectin-8N (Figure 4.8). This may be due to galectin-8C mediating the bone remodelling related functions of galectin-8 while MB46A might preferentially bound to the galectin-8N. Another possibility could be that the full-length protein or isolated domains including the galectin-8N are involved in this function but not through carbohydrate binding, and these functions might be mediated through protein-protein interactions. This, however, needs further investigations into the mechanism of galectin-8-mediated RANKL release and overall bone remodelling process.

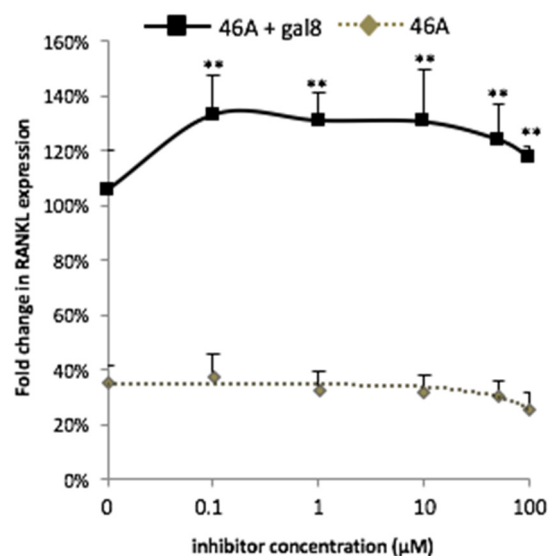


Figure 4.8: Effect of MB46A on RANKL expression. Osteoblasts from newborn CD1 calvarias were transferred to serum free medium for 1 hour before addition of 46A in the indicated concentrations. One hour later galectin-8 (50 nM) was added for 24 hours. RNA was extracted and qRT-PCR was performed in order to measure RANKL expression. Results shown are mean \pm SEM of 2 experiments done in duplicates. **p-value<0.05 of galectin-8 treated cells vs. cells with the same concentration of inhibitor and without galectin-8 [Prof. Y. Zick and Dr. Y. Vinik; Weizmann Institute of Sciences, Israel].

4.4.6 Hydrogen bond analysis

```
g_hbond -f md_fitted.xtc -s md.tpr -n new.ndx -hbn -hbm
```

```
## copy readHBmap.py script in the working directory
```

```
python readHBmap.py -hbm hbmap.xpm -hbn hbond.ndx -f md.gro
```

4.5 Further discussion

In the present study, a combination of computational and experimental approaches was applied towards ligand design. Molecular dynamics simulation using previously generated information (Chapter 3) and existing structural information (Chapter 2 and literature) were employed to design a hypothesis. The designed ligand was subsequently validated through synthesis, X-ray crystallography and ITC. The designed novel ligand methyl 3-*O*-[1-carboxyethyl]- β -D-galactopyranoside (**6**; also referred as MB46A) mimics the key interactions made by the equivalent part of a preferentially binding natural glycan 3'-SiaLac, also reflected from the binding affinity of **6** is very close to that of lactose. The use of a galactose scaffold plays a crucial role as it imparts the inherent ability to the designed ligand to interact with galectin-8N. Furthermore, the galactose core with a carboxylic acid group, compound **6**'s polarity might hinder its cellular penetration ability. However, esterification of the acid group can be carried out to enhance the cellular uptake.

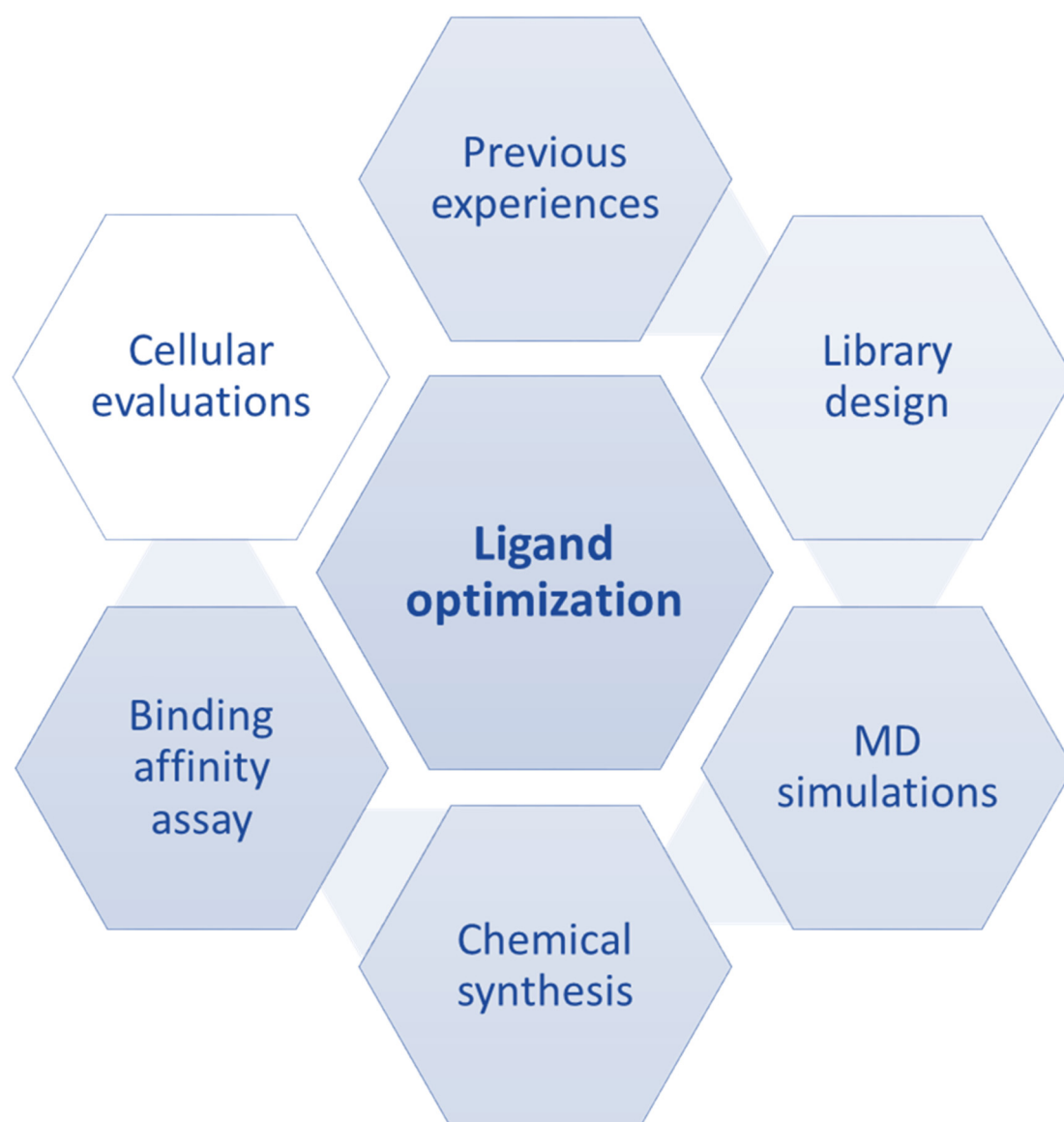
Ligand specificity is a crucial aspect when various members sharing a similar binding site exists. Using the unique binding site residues for interactions is a pragmatic approach towards designing specific ligands. Here the interaction with Arg59, a unique residue not present in any other galectins apart from galectin-8N, was exploited which in turn hints towards possible favourability in galectin-8N over another galectin CRDs. Further supportive evidence from our comparison based on MD simulation results of galectin-8N-**6** and galectin-8C-**6** suggest the favourability of compound **6** towards galectin-8N. Towards this end, galectin-8C gene was subcloned into pET-3a expression vector through ligation-dependant cloning and confirmed by sequencing (section 4.4.4). Initial protein expression trials did not result in protein overexpression particularly in the soluble fraction and therefore more extensive optimisation needs to be carried out to obtain the protein. Subsequently, the galectin-8C then can be employed for binding affinity and crystallographic investigations for determining binding mode and interaction of compound **6**.

The designed ligand is under *in vitro* evaluation for its ability to interfere with galectin-8-mediated bone remodeling (performed by our collaborators). The preliminary results show no inhibition of galectin-8 as assessed by monitoring RANKL expression levels in galectin-8-mediated osteoblasts (section 4.4.5). In addition, hemagglutination assay evaluating ability of compound **6** to interfere in agglutination ability of galectin-8 was inconclusive. However, further experimental repeats and overall experiment redesign is underway to further investigate the galectin-8 inhibitory potential of compound **6** both in general and in bone cells.

Therefore, a successful exploitation of both the evolutionarily conserved amino acids and those unique to galectin-8*N* for interactions by the ligand was achieved. The likely outcome of this study will be carried forward into the next chapter where other possible functional groups on the 3'-position of galactose will be explored. Functionalities that can be added on the 3'-position includes unsubstituted and substituted aromatic rings, sulfate groups or the combination of both groups to explore other important interaction hot spots such as Tyr141 and Arg45 individually or in parallel. Another potential side chain that can be added on to the 3-position is the glutamic acid like side chain containing two carboxylic acid group that can crosslink two arginines in the binding site. Overall, with the positive results, compound **6** will form a potential template for future designed molecules.

Chapter 5

Structure-based Design and Evaluation of a Monosaccharide-based Inhibitor of Galectin-8: A Structure-Activity Relationship Study



5.1 Introduction

Extending further the ligand design campaign targeting galectin-8, the aim being using the information from previously generated results in the thesis to design potential inhibitors of galectin-8. The minimum atomic features, identified in Chapter 2 [81], required for a ligand to bind to galectin-8 N , prompted the design of efficient small molecule ligands as inhibitors of galectin-8. In addition, the experience from the structure-based virtual screening campaign (Chapter 3) for identifying non-carbohydrate based binders of galectin-8 N was also the premise for this work. MB46A was designed based on glycan 3'- O -sialylated lactose (3'-SiaLac) and FDA drug pemetrexed (PMT) to explore the unique Arg59 and Gln47 residues for interactions. Following on from MB46A, an *in silico* library of compounds were generated to explore other potential hotspots in the galectin-8 N binding site. In the present work, residues such as Arg45 which is located across from the unique Arg59, and Tyr141 which is another unique residue in the extended galectin-8 N binding site, were explored for interactions. Various modifications based on the PMT side chain and the 3'-position of galactose were initially investigated to explore the best possible binding mode and interactions. Considering the synthetic feasibility and with the aim of expanding the structure-activity landscape, small-molecule galactose-based derivatives were synthesised. The validation of binding was performed using SPR. These compounds are undergoing biological evaluations (by our collaborator, Prof. Yehiel Zick, Weizmann Institute of Science, Rehovot, Israel) for examining inhibitory effects galectin-8 functions *in vitro* models and any promising ligand may then be subjected to *in vivo* evaluations.

5.1.1 The FDA approved drug

Our virtual screening experiment (Chapter 3) culminated in the identification of several molecules as potential binders of galectin-8 N as amongst them was Pemetrexed (PMT). PMT is an FDA-approved [181] anticancer drug that works as an antifolate agent by disrupting the folate-dependant metabolic processes essential for cell replication [182]. PMT is used as a monotherapy for the treatment of non-small cell lung cancer, and in combination with cisplatin, it is utilised for the treatment of malignant pleural mesothelioma [183].

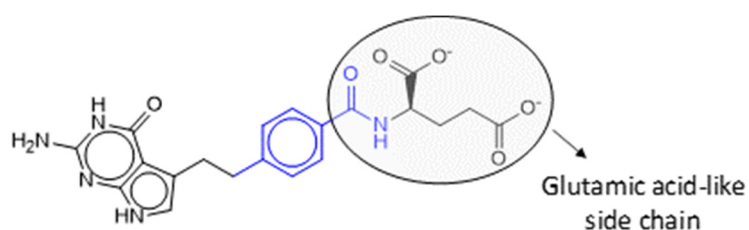


Figure 5.1: Chemical structure of Pemetrexed (PMT).

PMT was purchased for experimental evaluation (see section 3.3) after being identified in the top fraction from the final tier of the virtual screening. Structurally, PMT is an interesting molecule that contains a pyrrolopyrimidine-based ring system with an aliphatic glutamic acid-type side chain (Figure 5.1). During the analysis of the virtual screening results (Chapter 3; section 3.3.3 and section 3.3.4), an interesting placement of PMT in the galectin-8N binding site was noticed. The two carboxylic acid groups of the aliphatic side chain became involved in salt-bridge interactions with the multiple arginine residues in the binding site. Among these arginines, there is the evolutionarily conserved Arg69, Arg59 which is unique to galectin-8N, and Arg45 present on at the beginning of strand S3, straight across the unique residue. The anionic saccharides such as 3'-*O*-sialylated lactose (2.7 μ M [42]) or 3'-*O*-sulfated lactose (1.9 μ M [42]) have been reported to interact with these two residues to produce a unique network of hydrogen bonds that is responsible for imparting high binding affinity to the lactose core [76]. However, in our STD NMR experiments, there was no interaction recorded for PMT with galectin-8N, possibly due to the bulky aromatic rings which may be affecting the placement of the ligand in the primary binding site (Figure 5.1). The non-binding of PMT also suggested that importance of primary binding site residues in recognising any ligand.

5.1.2 Galactose and carboxylic acid

MB46A is methyl 3-*O*-[1-carboxyethyl]- β -D-galactopyranoside and was designed and validated previously (Chapter 4), based on the idea of exploring the galactose and carboxylic acid portion of 3'-SiaLac, that is reported to be preferentially recognised by galectin-8N [42, 44]. In our crystal structure of the galectin-8N-MB46A complex, unambiguous occupancy of the primary binding site by the galactose ring of MB46A was observed and interactions between the carboxylic acid group and unique Arg59 and Gln47 residues were revealed (Figure 5.2) [Bohari et al. manuscript submitted]. The presence of other interaction hot spots in the binding site warrants investigation of other functionalities in addition to the carboxylic acid group that can help gain further affinity and specificity. The binding affinity determined for the monosaccharide-based MB46A was 139 μ M close to 136 μ M affinity determined for the disaccharide lactose [Bohari et al. manuscript in preparation]. The enhanced affinity of MB46A, as compared to the parent galactose and involvement of unique residues in interactions, makes the designed ligand an ideal template for modifications. The modification based on the galactose core also assures the prerequisite binding to galectin-8N and the modified groups are favoured in the binding site. Therefore, the scaffold employed for generating the library of compounds were based on galactose.

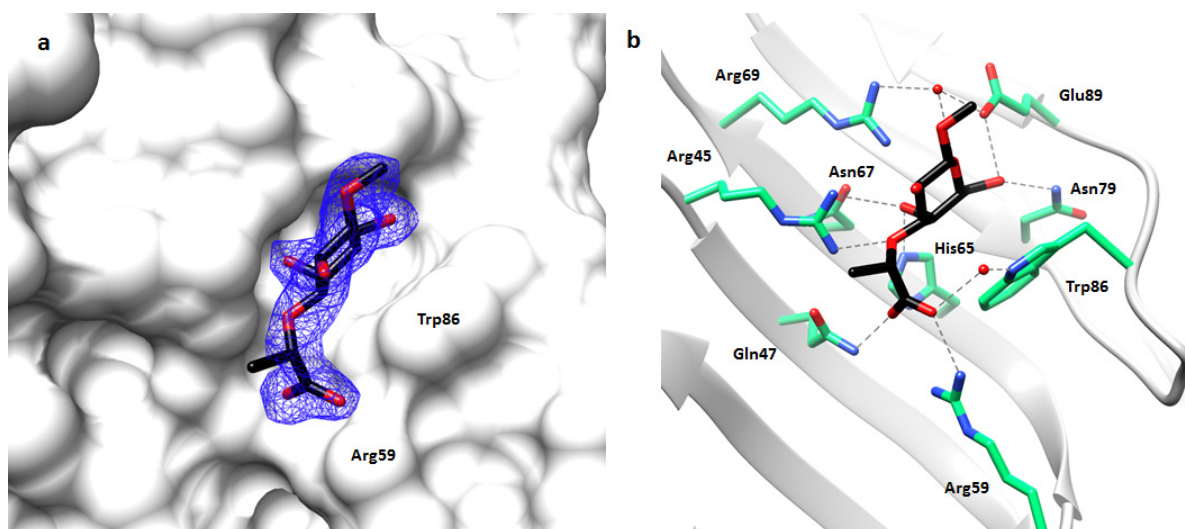


Figure 5.2: Galectin-8N-MB46A complex. (a) Electron density (blue mesh) $2|F_o|-|F_c|$ α c contoured at 1σ , for the ligand (in the sticks; carbon black; oxygen red) in complex with galectin-8N. (b) Hydrogen bonding interactions (in red dashed line) made between **6** (in black sticks) and galectin-8N (in green sticks; carbon green; oxygen red; nitrogen blue).

5.1.3 Library design

The simultaneous engagement of Arg59 and Arg45 residues by the carboxylic acid groups of the glutamic acid side chain portion of PMT was an interesting observation that was employed for designing the library. This observation formed the central part of the hypothesis to explore the unique Arg59 and Arg45 individually and then simultaneously using small monosaccharide-based molecules. Challenges were experienced in our virtual screening experiments to identify non-carbohydrate based binders, particularly due to the nature of the galectin-8N binding site. Furthermore, from Chapter 4 it was clear that using galactose as a core scaffold to design novel binders is a more pragmatic and reliable approach. The biggest advantage with galactose-based binders is that these molecules would have inherent binding to galectin-8N, unlike the non-galactose-based molecules. Although the overall binding affinity of galactose-based binders might not be high, they will form a good template for modifications during the ligand optimisation process.

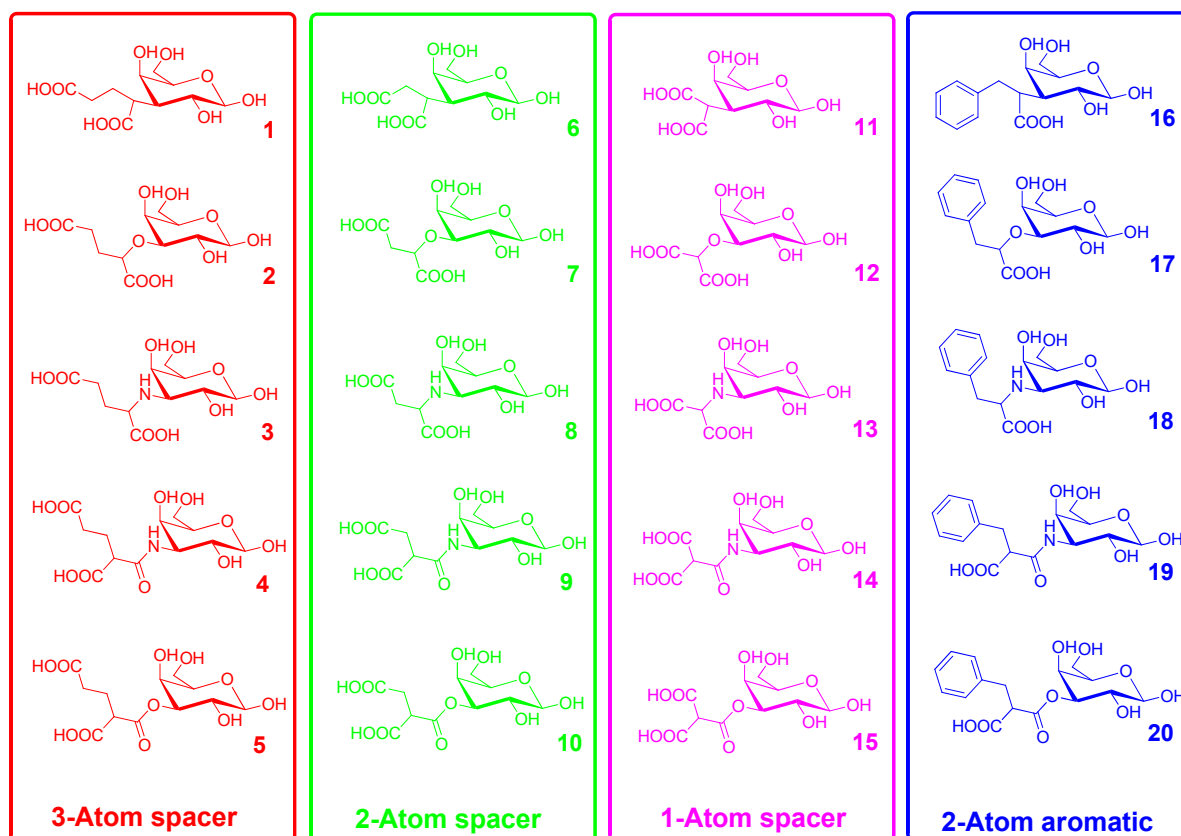


Figure 5.3: Designed library of compounds subjected to MD simulations.

Considering the two carboxylic acid groups on the glutamic acid side chain of PMT and the molecule MB46A, a library of compounds was generated (Figure 5.3). Here, the galactose ring from MB46A was retained and a dicarboxylic acid aliphatic side chain of PMT was attached on to the 3'-position of galactose. Subsequently, the number of atoms separating the two acidic groups were explored including three (as in PMT), two, and one atom spacer to better engage Arg45 and Arg59 in interactions. This aliphatic side chain was linked to the 3'-position of galactose either directly or through an oxygen (O), nitrogen (N), amide (CONH) or ester (COO) linkages. Compounds **1-5** had two COOH groups separated by three carbon atoms and joined to galactose either directly or through O, N, amide (CONH) or ester (COO) linkages. Similar linkages were used for compounds **6-10** that contains two COOH groups separated by two atoms while compounds **11-15** contains one separating atom between the COOH groups. To note, that compound **6** in the existing virtual library is different from the synthesised compound **6** (MB46A) in Chapter 4. From the physicochemical perspective of the ligands, compounds **16-20** were also considered, where one carboxylic acid group was replaced by an aromatic ring.

5.2 Methods

5.2.1 MD simulations

MD simulations were performed using the protocol used in [81], for a duration of 100 ns. The ligands considered here were modified in the Discovery Studio Visualizer [170] based on the MB46A template and were prepared for simulation as described previously in Chapter 4 (section 4.4.1). The ligand binding free energy was estimated using molecular mechanics energies with the Poisson-Boltzmann and surface area continuum solvation method (MMPBSA) [174, 184, 185] implemented in the Amber package [156]. VMD was used to convert the Gromacs format trajectories (`xtc`) into Amber readable format (`mdcrd`). The Amber module, tLeap was employed to generate corresponding parameter and topology for protein, ligand and protein-ligand complex applied with Amber99SB-ILDN force field [137] [`tLeap.in`; Appendix 5.5.1]. A set of 100 frames were periodically extracted from the trajectory file at an interval of 1 ns. The protein `pdb` file was manually edited by replacing the 'His' residue ID to 'Hid' to assign correct delta position protonation.

The binding free energy is estimated from the following equation [174, 184, 186]:

$$G = E_{bond} + E_{angle} + E_{dihedral} + E_{el} + E_{vdW} + G_{pol} + G_{np} - TS$$

$$\nabla\epsilon(r)\nabla\phi(r) + 4\pi\rho(r) = 0$$

$$\Delta G_{bind} = \langle G_{PL} \rangle - \langle G_P \rangle - \langle G_L \rangle$$

Where the molecular mechanics energy terms from bonded (bond, angle and dihedral), electrostatic and van der Waals interactions. G_{pol} is a polar contribution to free energies calculated based on finited difference solution to the PB equation (equation [187] in middle), and G_{np} is the non-polar contribution term estimated from a linear relation to the solvent-accessible surface area (SASA).

$\phi(r)$ is the electrostatic potential, $\epsilon(r)$ is the position dependent dielectric function and $\rho(r)$ is the charge density from solute. ΔG_{bind} is the difference in binding free energy for ligand binding to the protein, G_{PL} is free energy for protein-ligand complex, G_L and G_P is free energy for ligand and protein alone respectively.

5.2.2 Synthesis of MB61B and MB63N

Based on the template molecule MB46A, the propionic acid side chain was replaced with either benzoyl (MB61B) and naphthoyl (MB63N) group. The synthetic scheme until

deprotection of 3'-position (deacetylation product) is identical to that used for the synthesis of MB46A (Chapter 4; section 4.4.1). The 3'-hydroxy per protected sugar was either benzoylated (MB61B) or naphthoylated (MB63N) using benzoyl chloride and naphthoyl chloride (Figure 5.4).

General procedure: Thin Layer Chromatography (TLC) on pre-coated aluminium-backed silica plates (60 F₂₅₄; Merck) were used to assess reactions and visualised by charring in 4% sulphuric acid in ethanol. Reaction products were purified using flash chromatography silica gel 60. ¹H and ¹³C NMR spectra were recorded at 298 K using an Avance (400 MHz and 100 MHz, respectively; Bruker Biospin) spectrometer. Two-dimensional COSY (¹H to ¹H correlation), HSQC (¹H to ¹³C correlation) and HMBC (¹H to ¹³C long range correlation) NMR experiments were used to assist in assigning relevant peaks for ¹H and ¹³C NMR spectra. Electrospray ionisation low-resolution mass spectrometry was performed using a Bruker Daltronics esquire 3000 Ion-Trap instrument.

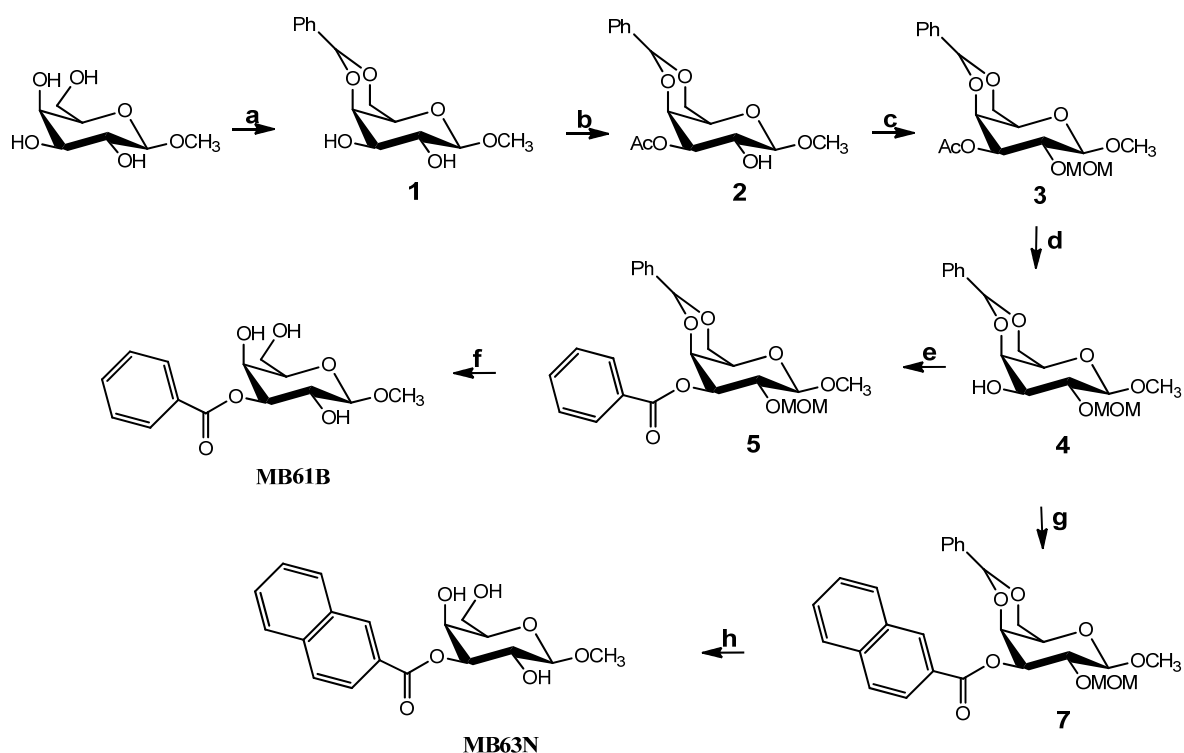


Figure 5.4: Synthetic scheme for MB61B and MB63N. Reagents and conditions: a) Camphorsulfonic acid, benzaldehyde dimethyl acetal, ACN, 60 °C, 2.5 h; b) Ag₂O (freshly prepared), AcCl, KI, DCM, RT, 16 h; c) DIPEA, MOMBr, DCM, reflux, 16 h; d) Na metal, MeOH, RT, 1.5 h; e) Benzoyl chloride, pyridine, RT, 1.5 h; f) 60% aqueous acetic acid, 60 °C 20 h; g) Naphthoyl chloride, pyridine, RT 1.5 h; h) 60% aqueous acetic acid, 60 °C, 48 h.

Methyl 2-O-Methoxymethyl-3-O-benzoyl-4,6-O-benzilidene-β-D-galactopyranoside (5): Compound **4** was dissolved in pyridine under argon at room temperature. Benzoyl chloride was added slowly, and the reaction was left stirring at room temperature for about 1.5 h. Upon completion of the reaction, pyridine was evaporated under vacuum and left on high vacuum

for at least 30 minutes to remove traces of pyridine. Following which the material obtained was washed with brine solution, the product was extracted in the DCM fraction and purified via flash chromatography (2:1 Hexane: EtOAc) to obtain **5** (~85 % yield). ¹H NMR (CDCl₃, 400 MHz): δ = 3.23 (3H, s), 3.58 (1H, d, *J*=1.08 Hz), 3.59 (3H, s), 4.10 (1H, dd, *J*=1.80, 12.40 Hz), 4.16 (1H, q, *J*=2.30, 10.12 Hz), 4.38 (1H, dd, *J*=1.56, 12.40 Hz), 4.43 (1H, d, *J*=7.8 Hz), 4.48 (1H, dd, *J*=0.72, 3.68 Hz), 4.70 (1H, d, *J*=6.60 Hz), 4.90 (1H, d, *J*=6.64 Hz), 5.20 (1H, dd, *J*=3.72, 10.12 Hz), 5.51 (1H, s), 7.34 (3H, m), 7.46 (9H, m), 7.60 (3H, m), 8.12 (6H, m). ¹³C NMR (CDCl₃, 100 MHz): δ = 55.89, 57.02, 66.26, 69.01, 73.06, 73.75, 97.23, 100.81, 104.17, 126.26, 128, 129.93, 130.22, 133.79.

Methyl 3-O-benzoyl-β-D-galactopyranoside (MB61B) [188]: The 4,6-benzylidene and 2-methoxymethyl groups of **5** were deprotected in 60 % aqueous acetic acid. Compound **5** was dissolved in aq. acetic acid and the reaction was left stirring for about 20 h to obtain deprotected MB61B as major product whilst with a minor proportion of partial mono-deprotected products were also observed. Water and acetic acid were removed under vacuum, followed by addition of small amount of toluene to remove traces of acetic acid. The final compound was purified by flash chromatography (1:2 Hexane: EtOAc) to a yield of 72 %. ¹H NMR (MeOD, 400 MHz): δ = 3.47 (3H, s), 3.55 (1H, m), 3.66 (2H, m), 3.80 (1H, q, *J*=2.36, 10.12 Hz), 4.06 (1H, d, *J*=3.08 Hz), 4.20 (1H, d, *J*=7.76 Hz), 4.48 (1H, s), 4.86 (1H, dd, *J*=3.32, 10.16 Hz), 7.37 (2H, t), 7.50 (1H, t), 8.01 (2H, d). ¹³C NMR (CD₃OD, 100 MHz): δ = 55.97, 60.81, 66.53, 68.67, 75.04, 76.63, 94.27, 104.61, 128.04, 129.44, 132.86, 321.1

Methyl 2-O-Methoxymethyl-3-O-naphthoyl-4,6-O-benzilidene-β-D-galactopyranoside (**7**): Compound **7** was prepared with similar method as compound **5**. Compound **4** was dissolved in pyridine followed by addition of 2-Naphthyl chloride and left stirring at room temperature for about 2 h. The purified product was obtained through flash column chromatography with 2:1 Hexane:EtOAc solvent system to a yeild of ~80 %. ¹H NMR (CDCl₃, 400 MHz): δ = 3.18 (3H, s), 3.53 (1H, s), 3.53 (3H, s), 4.03 (1H, m), 4.13 (1H, q, *J*=2.30, 10.08 Hz), 4.32 (1H, dd, *J*=1.48, 12.40 Hz), 4.37 (1H, d, *J*=7.76 Hz), 4.45 (1H, d, *J*=0.6, 3.72 Hz), 4.62 (1H, d, *J*=6.64 Hz), 4.84 (1H, d, *J*=6.64 Hz), 5.17 (1H, dd, *J*=3.76, 10.12 Hz), 5.45 (1H, s), 7.26 (3H, m), 7.47 (4H, m), 7.84 (3H, dd), 8.04 (1H, dd). ¹³C NMR (CDCl₃, 100 MHz): δ = 55.63, 57.04, 66.30, 69.03, 73.05, 73.82, 97.24, 100.82, 104.27, 125, 126, 127.78, 128, 129.45, 131, 123.47, 135.07, 137.69.

Methyl 3-O-naphthoyl-β-D-galactopyranoside (MB63N): 60% aq. acetic acid was used to 4,6 benzylidene and 2 MOM deprotection to obtain MB63N. The reaction was left stirring for 48 h at room temperature to obtain the completely deprotected product. The final purified product

was obtained via flash chromatography in 1:2 Hexane:EtOAc solvent system (70 % yield). ^1H NMR (MeOD, 400 MHz): δ = 3.48 (3H, s), 3.58 (1H, m), 3.66 (2H, m), 3.85 (1H, dd, J =2.4, 10.12 Hz), 4.11 (1H, d, J =2.88 Hz), 4.23 (1H, d, J =7.72 Hz), 4.93 (1H, dd, J =3.32, 10.12 Hz), 7.47 (2H, m), 7.84 (2H, dd), 7.91 (1H, d), 8.01 (1H, dd), 8.63 (1H, s). ^{13}C NMR (CD_3OD , 100 MHz): δ = 55.99, 60.83, 66.60, 68.74, 75.07, 76.79, 104.63, 124.95, 126.46, 127, 128, 130.97. $\text{C}_{18}\text{H}_{20}\text{O}_7$ 371.2

5.2.3 Isothermal titration calorimetry (ITC)

Quantitation of the binding affinity of MB61B and MB63N towards galectin-8N was performed using NanoITC (TA Instruments). A solution of 50 μL of 1 mM MB61B was titrated in aliquots of either 2.5 μL or 2.0 μL into the calorimetric cell, containing 300 μL of 200 μM galectin-8N with stirring speed of 150 rpm. Injections were performed at an interval of 300 s at 298 K. Galectin-8N was expressed and purified using the previously employed protocol [81]. The protein and ligand samples were prepared in either PBS or TBS (50 mM Tris base, and 150 mM sodium chloride). A blank experiment of ligand being titrated into the buffer was also performed and subtracted from actual binding experiment before analysing the thermograms. The thermodynamic analysis was conducted using an independent model with NanoAnalyze software.

5.2.4 Surface Plasmon Resonance (SPR)

The dissociation constants between the galectin-8N and the designed compounds (MB61B and MB63N) were measured using galectin-8N-immobilised sensor chip on a BIAcore T200 instrument (GE Healthcare). The purified and dialysed galectin-8N was covalently immobilised on the CM5 sensor surface according to manufacturer's instructions. To determine the suitable immobilisation conditions, galectin-8N (50-200 $\mu\text{g}/\text{mL}$) was diluted in the coupling buffer (10 mM sodium acetate with pH 4.5 or 5.0) and injected onto the non-activated chip surface for 2 minutes. The immobilisation procedure involved activation of the chip surface with a mixture of freshly prepared 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC)/N-hydroxysuccinimide (NHS) followed by injection of galectin-8N. Ethanolamine was used to block the excess of activated functional groups on the chip surface, that were not involved in coupling with the galectin-8N. The remaining activated functional groups on the chip surface that were not involved in amide bond formation with galectin-8N were blocked by ethanolamine. Lactose was employed as a positive control along with a binding investigation of the designed ligands MB61B and MB63N. Lactose in TBS buffer was introduced onto the chip surface at a flow rate of 30 $\mu\text{L}/\text{min}$, and

the interactions were monitored at 25°C as the change in the surface plasmon resonance response. The non-covalently bound ligand, before injection of another ligand, was cleared by a two-step regeneration involving flushing 30 μ L of 10 mM glycine pH 2 and TE (Tris and EDTA) buffer. The analysis of results was performed using BIA evaluation 3.0 software.

5.3 Results and Discussion

5.3.1 MD Simulation of FDA drug

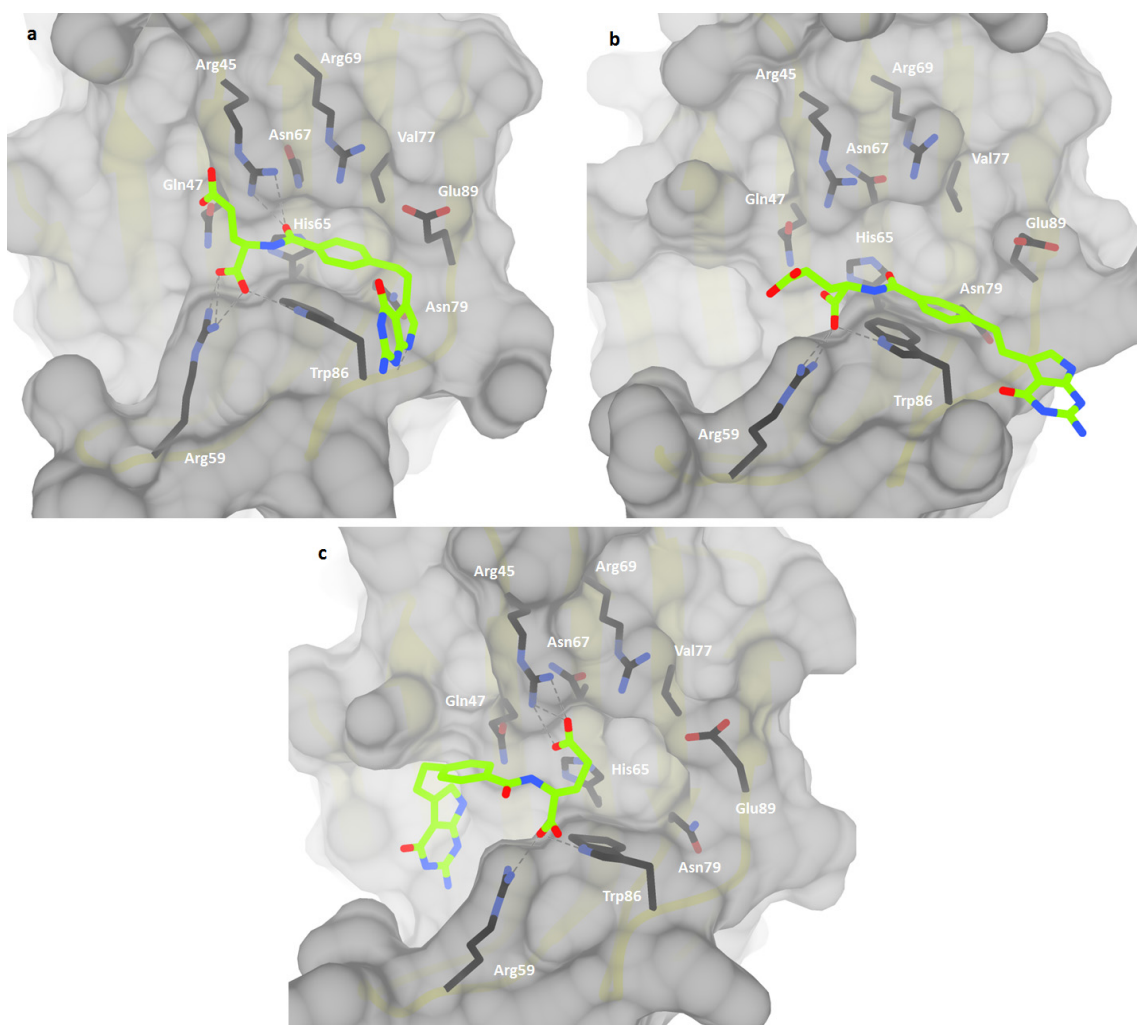


Figure 5.5: Progression of PMT during MD simulation from its predicted binding conformation to galectin-8N. Hydrogen bonding interactions (grey dashed lines) at the start (A), at halfway (B) and the end (B) of simulation.

For MD simulation of the galectin-8N-PMT complex (see section 3.2.2 for methods), the starting conformation used was the docking output conformation. In this conformation, the pyrrolopyrimidine ring is placed over the residues on strand S6, the phenyl ring occupies the galactose binding pocket, and one of the carboxylic acid group of the glutamic acid-type side chain interacts with Arg59, and the second carboxylic acid group was facing towards the solvent (Figure 5.5a). However, as the simulation progressed, the pyrrolopyrimidine ring of PMT that was involved in fewer less probable interactions in the earlier conformation, started fluctuating which affected the placement of remaining substructure of the ligand (Figure 5.5b). The aliphatic side chain is bearing the two carboxylic acids at the beginning of simulation interacted in a stable manner with Arg59 and Trp86. As the simulation progressed, the weak

interaction of the bulky aromatic rings caused a large flip in the ligand conformation, which shifted the aromatic rings from the primary binding site to the extended binding site (Figure 5.5b). Importantly, now the residues Arg45 and Arg59 were effectively cross-linked with the two carboxylic acid groups of the aliphatic side chain (Figure 5.5c).

Interestingly, during this aromatic ring flip-over, the carboxylic acid group interacting with Arg59 remained mostly unperturbed throughout the simulation. Moreover, after the aromatic ring flip, the second carboxylic acid of the aliphatic side chain was positioned such that it interacted with the Arg45. This is a significant observation that highlights potential hot spots in the galectin-8*N* binding site that can be engaged in interactions simultaneously by a ligand. Therefore, the concept of utilising these differences in the binding site can be beneficial in modulating the specificity of a ligand while the use of galactose core would ensure recognition of the ligand by the galectin-8*N*.

5.3.2 Simulations of the designed library

Investigation of alternative novel binders of galectin-8*N* apart from the MB46A (Chapter 4) was undertaken. Our virtual screening results performed in Chapter 3 and the available structural information (from Chapter 4 and the literature) formed the basis for design of a library of galactose-based compounds in the present chapter. Interestingly, the FDA approved PMT, identified from virtual screening cross-linked the unique Arg59 and Arg45 (that lies across the binding site). Hence, the glutamic acid or similar carboxylic acid aliphatic or aromatic side chains were modelled onto the 3'-position of galactose and engage those residues in interaction. Apart from the similarity in this side chain, the chemical structure of PMT differs significantly from the designed library of compounds (Figure 5.1). To predict the possible binding mode, interactions and ligands' occupancy of the galectin-8*N* primary binding site, 100 ns MD simulations were performed on the protein-ligand complex. The library of compounds was divided into four categories depending on the side chain attached on the 3'-position. The side chains were varied in each category based on the number of atoms separating either the two carboxylic acid groups or the carboxylic acid group and an aromatic ring. These side chains were then further divided based on the atom linking the side chain to galactose.

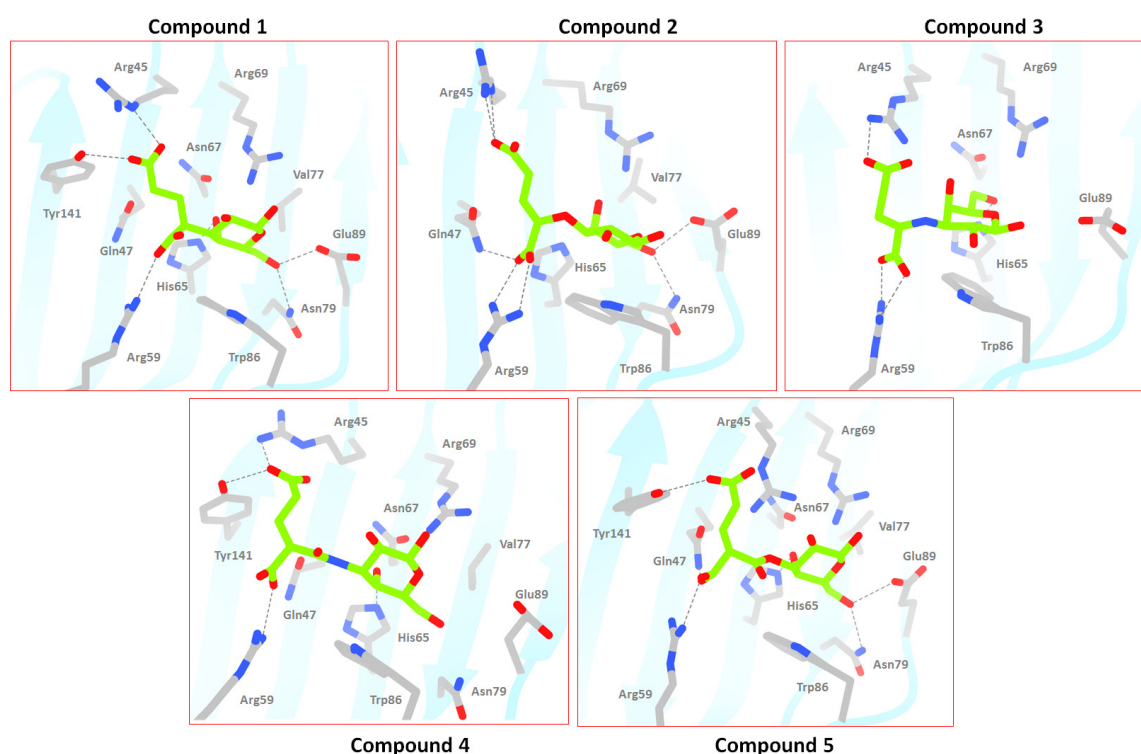


Figure 5.6: Simulation of compound 1-6 (3-Atom spacer category) in complex with galectin-8N. The snapshot depicted here is the final coordinates saved at the last step of 100 ns run. The compound carbon atoms are coloured in green while amino acid residues carbons in grey and the secondary structure in blue ribbons. Hydrogen bonding interaction between the protein and ligand is represented by grey dashed lines.

The 3-Atom spacer category was considered to exactly match the aliphatic side chain of PMT (Figure 5.3). Because no binding for the ligand PMT was detected towards galectin-8N (Chapter 3), the aliphatic side chain was coupled to galactose for investigating the possibility of engaging the two arginines for interactions simultaneously in the galectin-8N binding site. In this category of five compounds, the simulations confirmed that the galactose ring always occupied the primary binding site stacking against the Trp86 throughout the duration of simulation (Figure 5.6 and Figure 5.6). The carboxylic acid group that is close to the galactose ring almost always engages in salt bridge interactions with the Arg59. While, the second carboxylic acid group, only partly formed salt bridge interactions with Arg45, more so for this interaction to occur Arg45 moved farther close to Tyr141 (Figure 5.6). The simultaneous engagement of the two arginines and carboxylic acid group from the glutamic-acid type side chain was only transient in this category of compound (1-5), unlike the more probable engagement noted in the case of PMT. This may be due to the difference in initial placement of the two ligands in the primary binding site (Figure 5.7). Galactose being the native ligand to galectin-8N occupies the primary binding site completely, while the bulky aromatic rings in PMT disallows this complete occupancy (Figure 5.5). As a consequence, the side chain of the designed galactose-based molecules is placed slightly ahead, farther into the binding site, to that observed in the galectin-8N-PMT complex (Figure 5.7).

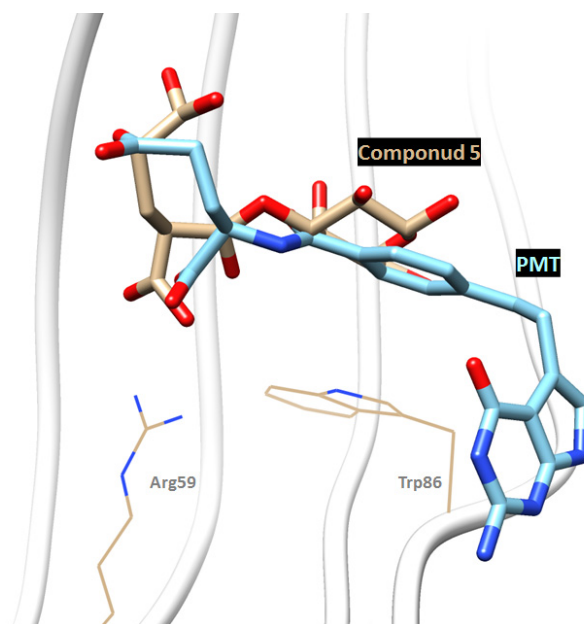


Figure 5.7: Comparison of the initial placement of PMT (carbon in cyan) and compound 5 (carbon in brown) of the designed library.

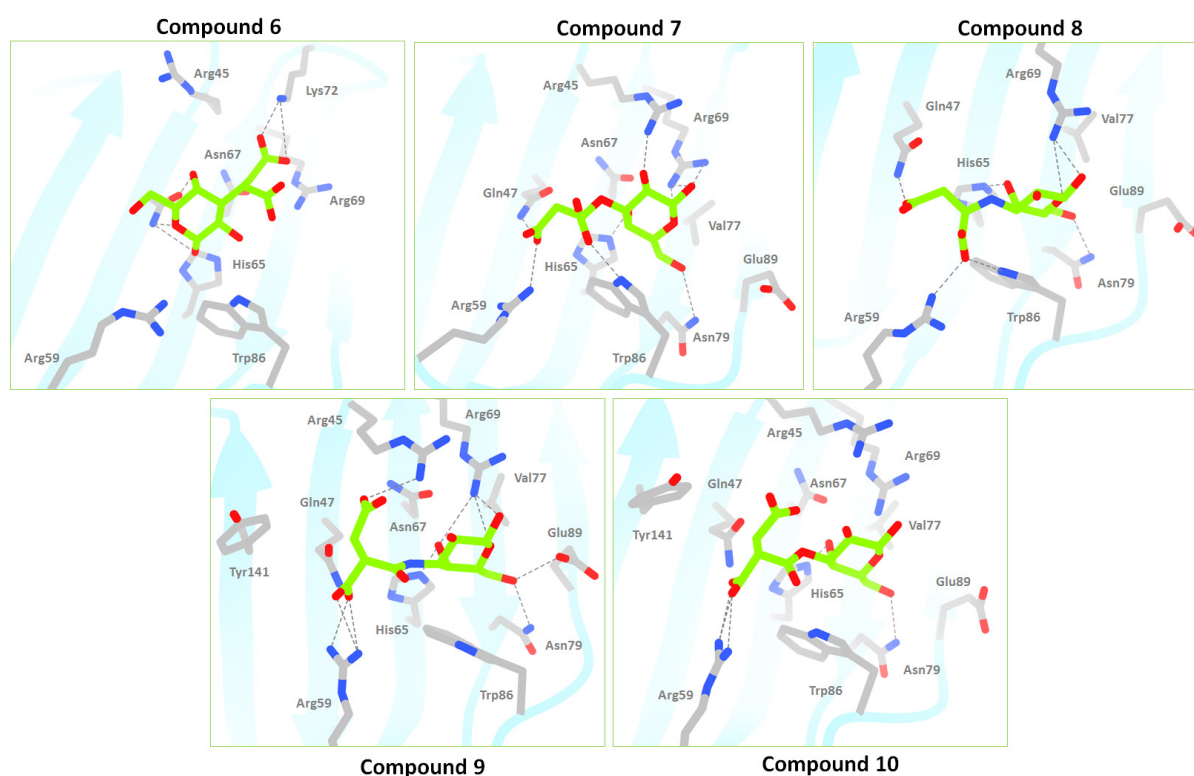


Figure 5.8: Simulation of compound 6-10 (2-Atom spacer category) in complex with galectin-8N. The snapshot depicted here is the final coordinates saved at the last step of 100 ns run. The compound carbon atoms are coloured in green while amino acid residues carbons in grey and the secondary structure in blue ribbons. Hydrogen bonding interaction between the protein and ligand is represented by grey dashed lines.

Based on the structural differences between our library of compounds and PMT, one methylene group from the spacer between the two carboxylic acid groups was removed to form a 2-Atom spacer category (Figure 5.3). Interestingly, for compound 6 wherein the carboxylic

acid side chain is directly linked to the carbon atom at 3'-position of the galactose ring (devoid of any linking atom) was the most unstable in occupying the primary binding site (Figure 5.8). This unsuitability of direct linking possibly reflects the effect of increased rigidity of the side chain causing the placement of the galactose ring in an unfavourable vicinity of the Arg45 and the conserved Arg69 residues. Therefore, compound **6** is flipped over and rotated such that the carboxylic acid side chain is now interacting with Lys72 on S4-S5 loop and is in close vicinity of conserved Arg69 (Figure 5.8). While in the case of compound **7** and **8** where the side chain is linked through O and N, the occupancy of Arg45 hydrogen bond was less and the carboxylic acid groups of the side chain were facing towards Trp86 and Arg59 (Figure 5.8). Compound **9** and **10** with amide and ester linkers respectively had one carboxylic acid group engaged with Arg59 while the other one near Arg45 (Figure 5.8). However, the conformation of Arg45 appears more restrained due to its stacking over the Arg69 (Figure 5.8).

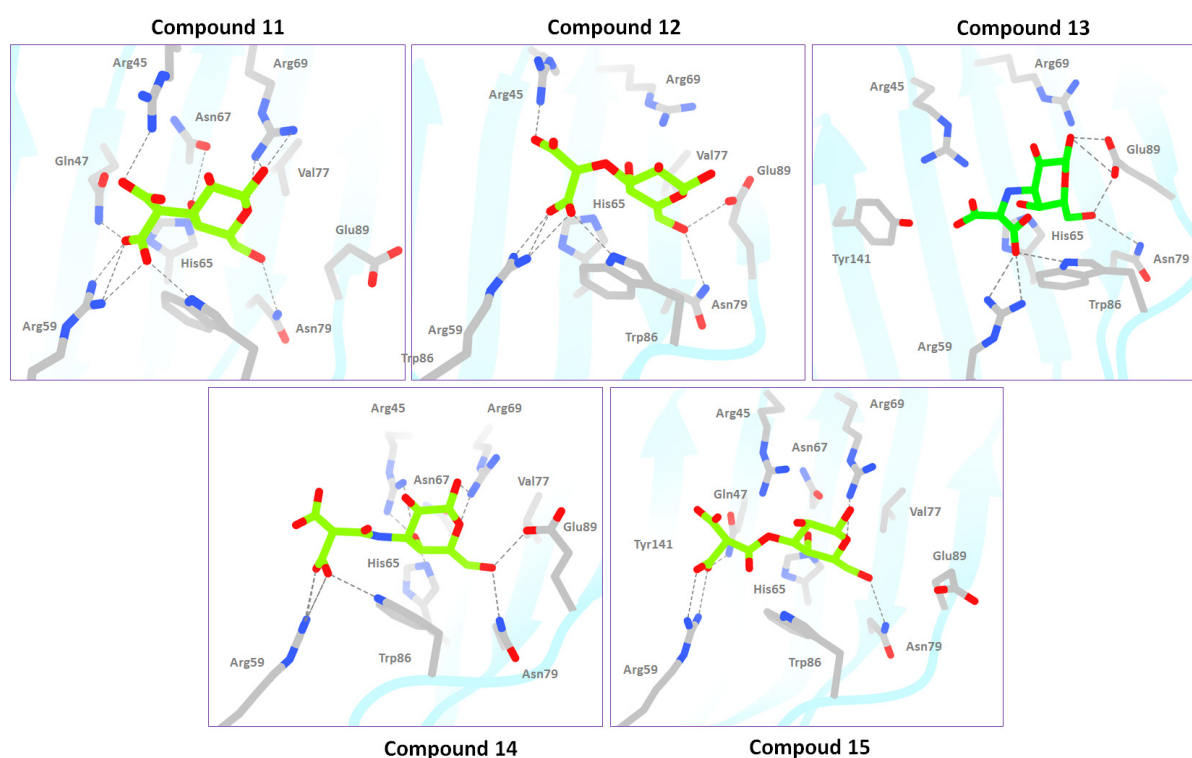


Figure 5.9: Simulation of compound 11-15 (1-Atom spacer category) in complex with galectin-8N. The snapshot depicted here is the final coordinates saved at the last step of 100 ns run. The compound carbon atoms are coloured in green while amino acid residues carbons in grey and the secondary structure in blue ribbons. Hydrogen bonding interaction between the protein and ligand is represented by grey dashed lines.

For 1-Atom spacer category, the two carboxylic acid groups were attached onto a single carbon atom, although synthetically these compounds might be challenging to make (Figure 5.3). Nevertheless, they were considered in the library to have a better comparison of a suitable number of spacer atoms needed between the two carboxylic acid groups. When the side chain was directly linked or linked through one atom (as in compound **11** and **12**), then the carboxylic

acid groups cross-linked the arginines. However, with amide/ester linker (as in compound **14** and **15**) the side chain was placed such that O4 hydroxyl of galactose interacted with Arg45 while the carboxylic acid group was facing towards solvent (Figure 5.9).

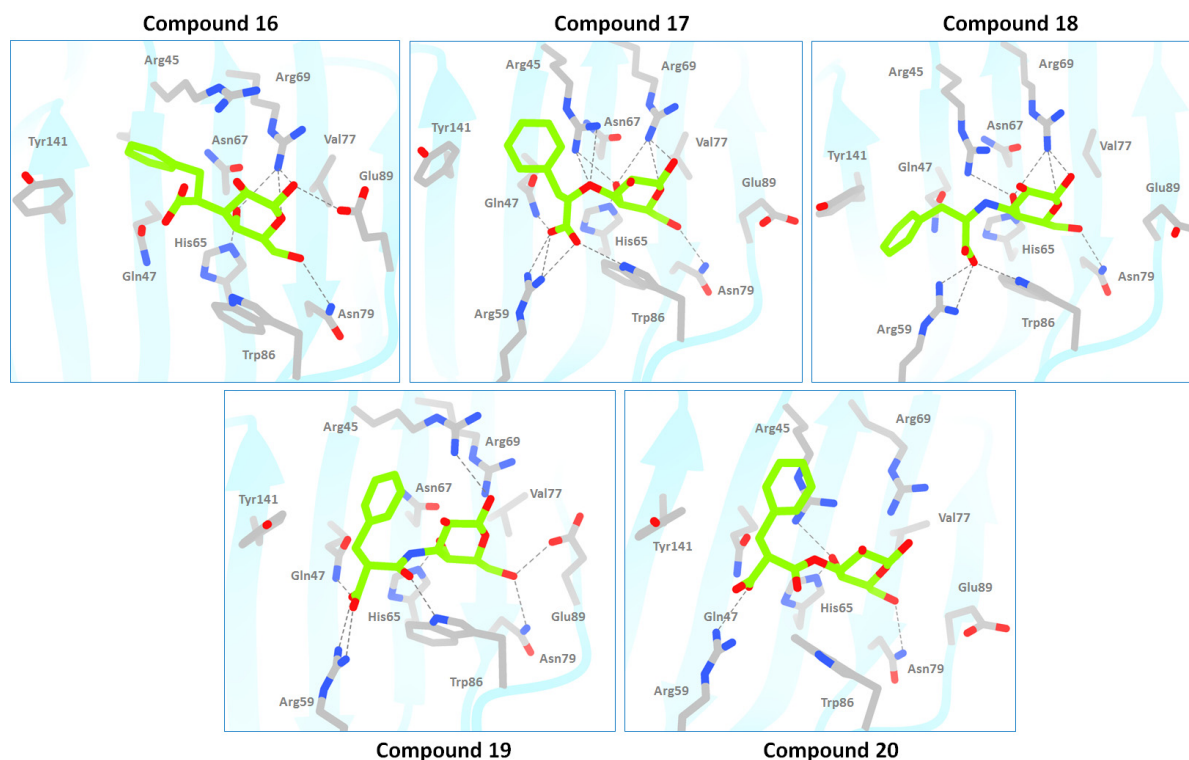


Figure 5.10: Simulation of compound 16-20 (2-Atom aromatic category) in complex with galectin-8N. The snapshot depicted here is the final coordinates saved at the last step of 100 ns run. The compound carbon atoms are coloured in green while amino acid residues carbons in grey and the secondary structure in blue ribbons. Hydrogen bonding interaction between the protein and ligand is represented by grey dashed lines.

As the compounds in the library were based on a galactose core, they possess inherent polarity which is further escalated by each added carboxylic acid groups. The increasing polarity not only interferes with the pharmacokinetic properties of the ligands but also poses synthetic challenges. Therefore, to balance overall polarity of the ligands, one of the carboxylic acid group was replaced with an aromatic ring. Based on the simulation results (discussed above), the carboxylic acid from the 2-Atom spacer category with ester/amide linkers where the arginines were engaged in interaction simultaneously, were chosen for adding an aromatic ring (Figure 5.3). As noted previously during the simulation of compounds where the side chain was directly attached to the C-3 of the galactose, in this group as well compound **16** did not attain the predicted conformation, and Arg45 was noted facing away from the binding site. While in compounds **17-20** the aromatic ring gets partly stacked over the Arg45, exhibiting the cation- π type of interactions and hydrogen bonding interactions between the Arg45 and hydroxyls of the galactose (O4 hydroxyl) (Figure 5.10). Of note, that the snapshots used for generating figures in case of compound **18** and **19** does not exactly reflect the partial stacking

over Arg45, however, during most part of simulation the aromatic ring was in the vicinity of Arg45. The carboxylic acid group almost always was involved in hydrogen bonding with unique Arg59 and partly with Gln47. Having retained the critical interactions that were reported to contribute towards specificity of glycans [76], the compounds with aromatic rings can potentially be taken up for synthesis.

5.3.3 Binding free energy estimation

Quantitative assessment of the simulations was performed through the estimation of binding free energy using the MMPBSA method. This method takes both the bonded and non-bonded interactions into account for a protein-ligand complex and estimates the binding free energy. The comparison of the enthalpy fraction of the binding free energy was performed because of its good correlation with the experimental binding affinity, lesser margin of error and less computation time as compared to the entropy fraction obtained from the normal mode analysis. Furthermore, compounds in the designed library were based on galactose core with small modifications between each other; the overall entropic contribution would be similar and therefore is not considered in the analysis.

Several galectin-8*N*-glycan complexes for which the experimental binding affinity data was available (3AP4 [Lac], 3AP6 [SulLac], 3AP7 [SiaLac] and 3AP9 [LNFIII]) [76] were subjected to binding free energy estimations. These complexes were subjected to 100 ns MD simulation, and 100 frames were extracted at a statistically independent interval of 1 ns from the trajectory file for binding free energy estimations. The estimated enthalpies and various contributing components were averages performed on 100 protein-ligand complexes for one system (Table 5.1). Here more negative enthalpy value indicates favourable polar interaction with the galectin-8*N* binding site which in turn implies overall stronger binding. Trend-wise, the enthalpies from MMPBSA correlated well with the experimental observed binding affinities. However, the bias from the size of the ligand was inevitable, as noticed from the estimated enthalpy value for LNFIII that was about ~15 kcal/mol lower than the estimated enthalpy of SulLac. However, the experimental binding affinity for all the three glycans SulLac, SiaLac and LNFIII determined by SPR were comparable (1.7 – 3.3 μ M [42]). A reasonable correlation was observed here, however, is promising for prioritising different ligands.

Table 5.1: Estimated ligand binding free energies (in kcal/mol) of ligand towards galectin-8N from MMPBSA.

Ligand		VDWAALS	EEL	EPB	ENPOLAR	ΔH
Crystal ligands*	Gal	-14.8	-42.9	40.5	-12.1	-29.3
	Lac	-16.5	-79.7	77.9	-15.9	-34.3
	SulLac	-20.0	-204.7	193.4	-18.7	-50.9
	SiaLac	-24.1	-243.5	220.1	-22.3	-69.8
	LNFIH	-41.8	-138.4	149.0	-35.2	-66.5
3-Atom spacer	1	-17.4	-355.3	335.4	-17.5	-54.7
	2	-18.1	-347.6	320.0	-18.0	-63.8
	3	-13.9	-344.1	316.4	-15.9	-57.4
	4	-15.8	-362.2	328.1	-19.6	-69.5
	5	-21.8	-337.6	308.6	-20.3	-71.1
2-Atom spacer	6	-14.3	-308.7	290.2	-14.5	-47.4
	7	-18.9	-334.3	305.7	-18.3	-65.8
	8	-16.5	-305.8	284.3	-17.2	-55.3
	9	-16.3	-348.7	320.8	-18.6	-62.9
	10	-16.9	-345.0	317.1	-18.7	-63.5
1-Atom spacer	11	-14.2	-378.9	344.3	-15.3	-64.1
	12	-16.3	-357.5	329.9	-15.8	-59.7
	13	-12.2	-304.6	289.4	-12.7	-40.1
	14	-15.4	-319.1	297.7	-15.5	-52.4
	15	-19.2	-342.0	313.7	-18.0	-65.5
2-Atom aromatic	16	-30.6	-157.7	163.5	-21.7	-46.4
	17	-22.2	-213.0	186.7	-19.4	-67.4
	18	-16.3	-171.7	161.2	-15.4	-42.9
	19	-19.6	-189.4	171.6	-19.2	-56.7
	20	-23.9	-205.0	179.0	-20.8	-70.6
Tested	MB61B	-20.9	-32.5	32.9	-16.0	-36.55
	MB63N	-23.9	-25.2	29.1	-16.5	-36.59
	MB46A	-18.0	-205.5	179.5	-16.6	-60.62

* VDWAALS – van der Waals interactions; EEL – electrostatic interactions; EPB – Polar contribution to the free energy calculated from the Poisson-Boltzmann equation; ENPOLAR – Non-polar contribution to the free energy obtained from solvent accessible surface area

Following the crystal complex, the designed ligand in complex with galectin-8N were subjected to binding free energy estimations. This analysis will help gain quantitative insight into the impact of varying the side chains and linker towards the binding affinity. There was an obvious increase in the overall enthalpy for the designed monosaccharide-based molecules as compared to the parent galactose due to at least one added a carboxylic acid group. The

resulting strong contribution from the electrostatic component (Table 5.1) is evident from the hydrogen bonding interaction between the carboxylic acid group and Arg59 and Gln47 (discussed above). Further favourable electrostatic contribution (Table 5.1) to the binding was provided by the second carboxylic acid group that formed additional hydrogen bonds in some cases with Arg45 and Tyr141 (compound **1-15**). In the case of compounds **15-20**, a decrease in electrostatic contribution was noted which was compensated by the gained van der Waal's interactions (Table 5.1), resulting from the replacement of the second carboxylic acid group with an aromatic ring.

Further, it was apparent that those having ester/amide as a linker between the side chain and the galactose ring resulted in relatively better enthalpies as compared to other linker atoms (Table 5.1). The relatively higher flexibility exhibited in the compounds containing amide/ester linkers compared to compounds with the ether linker, results in better placement and interactions of the side chain within the binding site. Compound **5** and **20**, both with the ester linker from the 3-Atom spacer and 2-Atom aromatic categories, produced the best enthalpic contribution with ΔH of -71.1 kcal/mol and -70.6 kcal/mol respectively amongst the designed library of compounds. Implying the linker atom coupling the side chains to the galactose are critical in directing the placement of the side chain into the binding site.

Overall, from interaction analysis, the compounds with galactose and the side chain was directly linked were the most unsuitable ones. While the O/N linked and ester/amide linked side chain interacted well in case of 3-Atom spacer group and 2-atom aromatic as compared to the other two groups. With the structural information of compound MB46A containing a single carboxylic acid group in hand, obtaining information in parallel for compounds containing an aromatic ring was aimed. Especially with balanced contribution from both the electrostatics and the van der Waal's component of binding free energy observed for the compounds in the 2-Atom aromatic group may overall result in an equal gain in the binding affinity as for MB46A. Furthermore, adding an aromatic ring would balance the inherent high polarity of the galactose ring. The high polarity of compounds might interfere with their physiological uptake and thereby the biological effects.

5.3.4 The designed compounds: MB61B and MB63N

Developing the structure-activity relationship from the ligand design perspective is a crucial aspect towards progressing a design campaign. The successful design and experimental validation of its binding to galectin-8N with its binding affinity close to that of a disaccharide is very promising and warrants further structure-activity exploration. Therefore, the

contribution of the carboxylic acid group in MB46A towards the observed binding was investigated. MD simulation results revealed that an aromatic ring could be accommodated well in the binding site onto the 3'-position of the galactose in place of the carboxylic acid group, to retain overall binding strength. The simulation results indicated that ester linker containing molecules interacted with the binding site well, as observed from their estimated binding free energies (Table 5.1). Also, the results of 2-Atom aromatic group prompted for exploring aromatic substitution on to galactose (Figure 5.10). Two galactose-based ligands with either benzoyl (MB61B) or naphthoyl (MB63N) group at the 3'-position (Figure 5.11) were synthesised considering mainly the synthetic feasibility. These aromatic rings are attached to galactose via an ester linkage. These molecules evaluated for binding to galectin-8N by MD simulations, ITC and surface plasmon resonance.

The estimated binding free energy from MMPSA revealed identical enthalpies for MB61B and MB63N, suggesting no particular interactions gained from the additional aromatic ring in MB63N. There was an obvious decrease in the electrostatic component between the aromatic compounds and MB46A due to the lack of carboxylic acid group. However, there is some gain in the van der Waal's component noted for MB61B and MB63N. The availability of 3-hydroxy per protected galactose and relative ease with which bezoyl/naphthoyl chlorides could be coupled at 3-position, MB61B and MB63N were conceived for experimental investigations

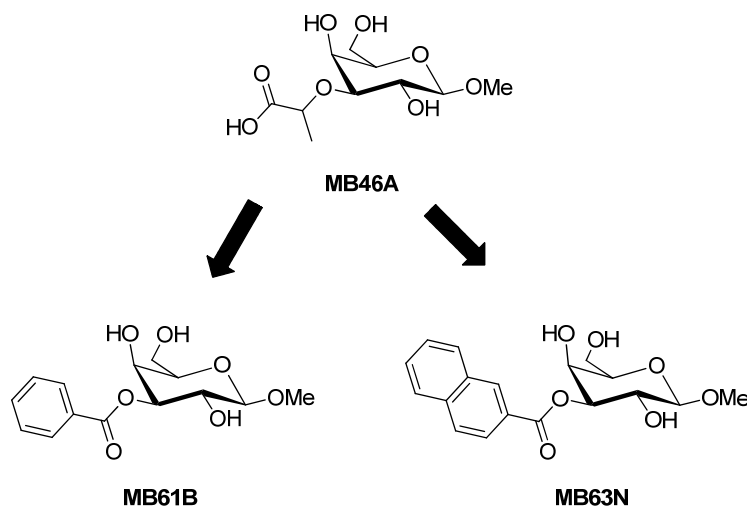


Figure 5.11: Chemical structure of the compounds MB61B and MB63N synthesised along with the previously designed compound MB64A.

5.3.5 Isothermal titration calorimetry (ITC)

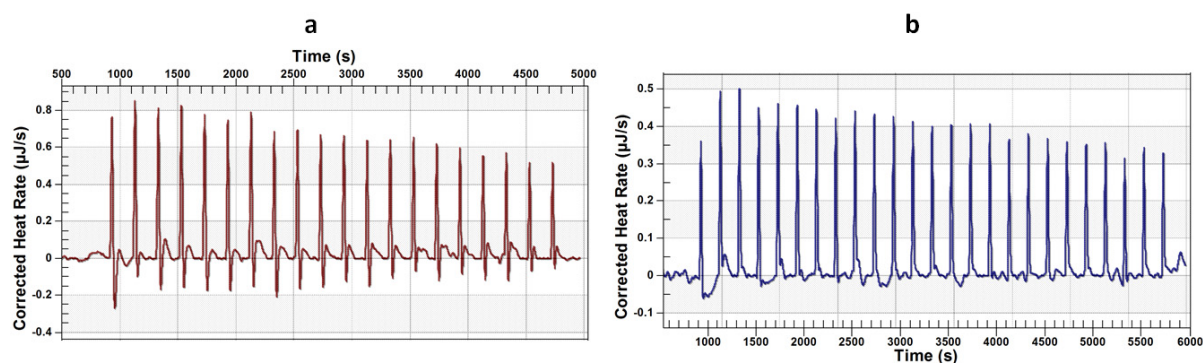


Figure 5.12: Isotherm for comparison of baseline during the titration of MB61B into galectin-8N with 200 s as ligand injection interval. a) 1 mM MB61B titrated in 2.5 µL aliquots into 200 µM protein in PBS; b) 0.8 mM MB61B titrated in 2.0 µL aliquots into 200 µM protein in PBS. Note the variation in the baseline.

ITC was used to determine the binding affinity of MB61B and MB63N towards galectin-8N. Lactose was used as positive control for the experiment. The titration of 1 mM lactose into 200 µM galectin-8N in PBS resulted with a binding affinity of 136 µM (Chapter 4). Due to the less aqueous solubility of MB61B, 100 mM ligand stock was prepared in 100 % DMSO, and 1 mM ligand was titrated containing a total of 5 % DMSO. A similar concentration of DMSO was maintained in the protein sample as well to compensate the heat of dilutions. The titration of MB61B in galectin-8N resulted in heat generation but with a significantly disturbed baseline, due to heavy precipitation upon titration in the calorimeter cell (Figure 5.12). The buffer change to Tris buffer saline (TBS) with similar concentrations of protein and ligand resulted in only fine precipitation and relatively stable baseline (Figure 5.12a). Therefore, various components in phosphate buffer saline with varying water solubility were found to be causing the precipitation. Despite having reasonable titration curves for protein-ligand titration, the buffer change to TBS led to the high heat of dilution observed in the blank titration. This heat was overpowering the heat generated from protein-ligand titration experiment, while the observed heat in the blank titration of lactose was much lower (Figure 5.13). Few things could be tried to resolve the issue, such as reducing the DMSO concentration and varying the protein and ligand concentrations to overall gain the heat changes. However, due to the challenges faced with the binding affinity determination by ITC, another approach to determine the binding affinities of the designed ligands was employed.

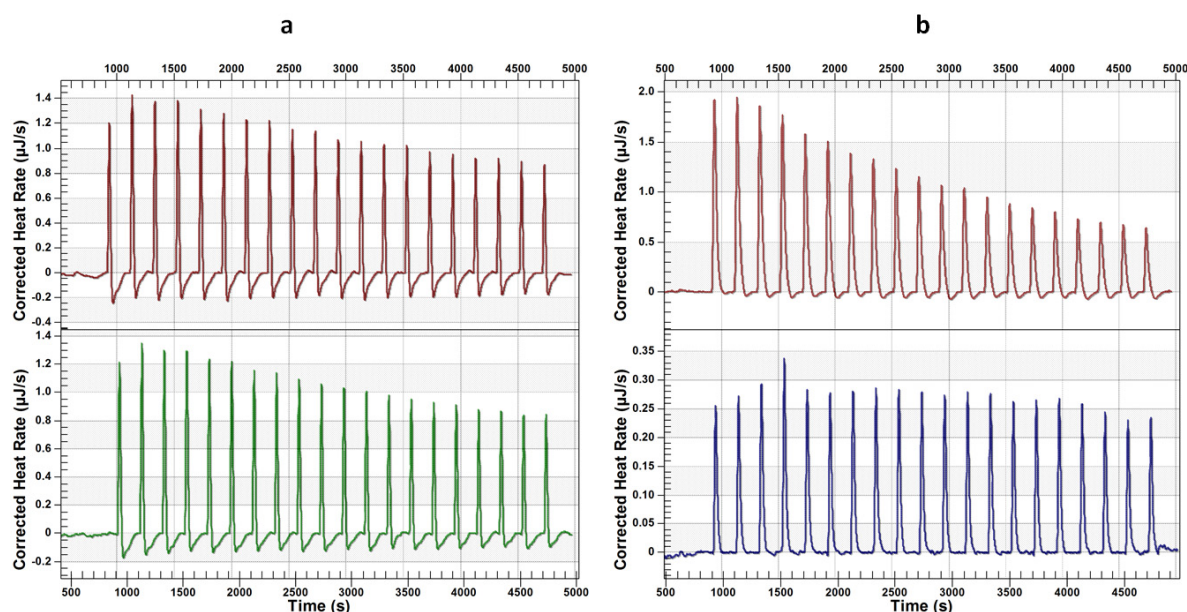


Figure 5.13: Isotherms for comparison of heats generated during protein-ligand (top; maroon) and protein-buffer (bottom; green/blue) titration. a) 1 mM MB61B titrated into 200 μ M galectin-8N and b) titration of 1 mM lactose into 200 μ M galectin-8N. Note the Y-axis for the two blanks (bottom) for a and b.

5.3.6 Surface Plasmon Resonance (SPR)

SPR is a phenomenon that occurs when a polarised light strikes an electrically conducting surface at the junction of two systems. The GE Healthcare SPR system allows for detection of biomolecular interactions. These interactions are measured on a removable sensor chip by the detector. The sensor surface is made up of a glass slide coated with electrically conducting gold film attached to dextran matrix. The protein of interest is covalently attached to the dextran through an amide linkage. The carboxymethyl groups on dextran are activated by 0.2 M EDC and 0.05 M NHS to obtain reactive succinimide ester that spontaneously reacts with the free amine group on the protein surface, thereby causing direct immobilisation. The ligands under investigation are injected over the immobilised protein. SPR causes a reduction in the intensity of light reflected at a specific SPR angle from the glass slide of the sensor surface. The change in the refractive index upon ligand binding causes an alteration in the angle of reflection which is displayed in the sensogram.

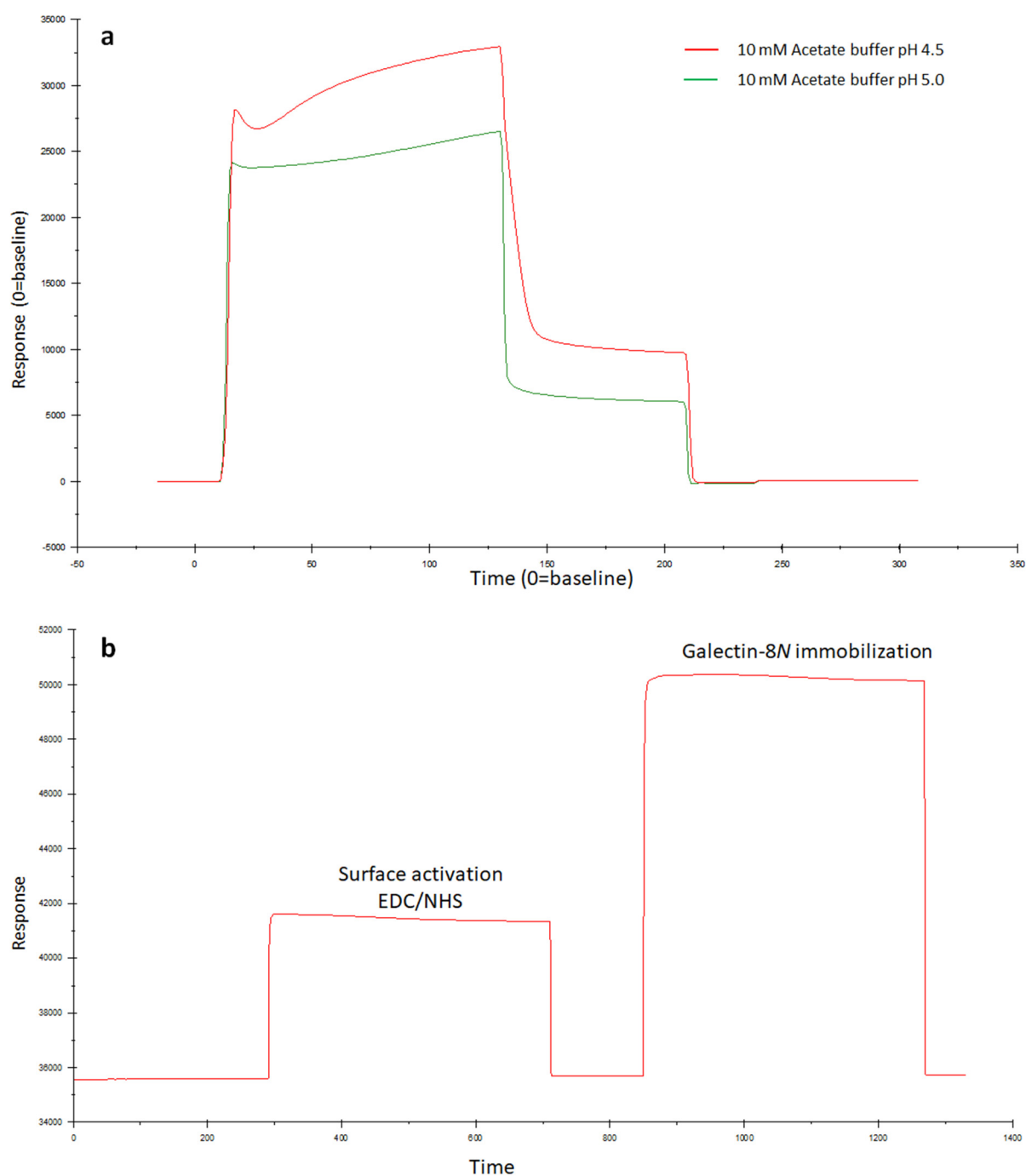


Figure 5.14: a) The pre-concentration scouting results for determination of optimum conditions for coupling galectin-8N to the chip surface. b) Sensogram for immobilisation of galectin-8N through amine coupling, illustrating the response units generated upon activation of carboxymethyl groups of dextran and amine coupling of galectin-8N.

The GE recommended pH for immobilisation of proteins is 5.0. However, the effect of lower pH values on the binding of the protein to chip surface was assessed in the pre-concentration scouting experiment. The magnitude and nature of response were analysed for injection of galectin-8N on the non-activated chip surface. A very low response unit (10-20 RU) was noted for the dilute protein sample ($\sim 20\text{-}40\text{ }\mu\text{g/mL}$) which then prompted for increasing the protein concentration (not shown in sensogram). There were no irregularities in

the sensogram noted for the galectin-8*N* (200 µg/mL) in TBS buffer at pH 4.5 and 5.0 (pI of galectin-8*N* is 8.1), that produced a response of 20,000-30,000 units (Figure 5.14). The pre-concentration condition at pH 4.5 was chosen for immobilisation although pH 5.0 coupling condition could also be used as enough response (suggested by GE) was recorded in both conditions. About 10,000 RU of galectin-8*N* was immobilised on the flow cell 2 (FC2) of the CM5 chip with flow cell 1 (FC1) being the reference cell (Figure 5.14).

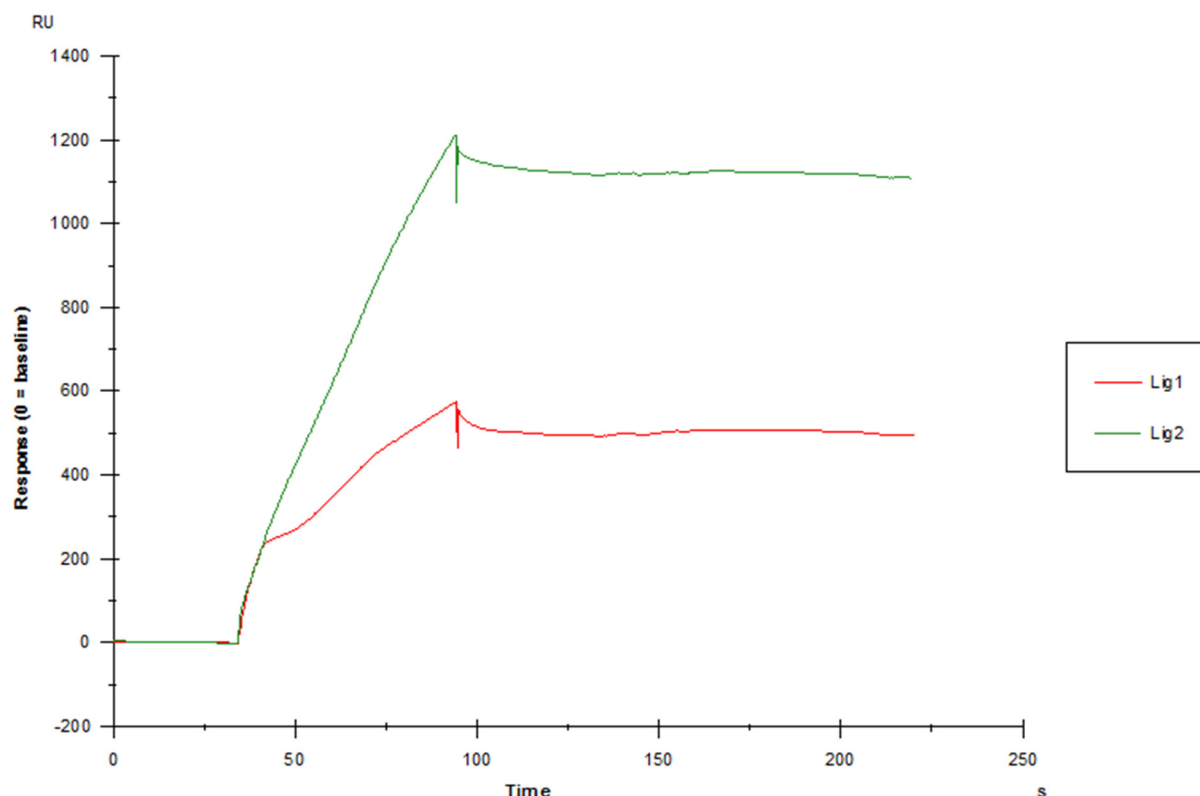


Figure 5.15: Binding level screen to examine binding of the designed ligands (Lig1-MB61B in red and Lig2-MB63N in green) towards immobilised galectin-8*N* at 1 mM concentration and 30 sec dissociation time.

Initially, a quick screening was performed to assess qualitatively whether the designed ligands were binding to galectin-8*N*. This binding screen was carried out at 1 mM ligand concentration and 30 seconds of dissociation time. The resulting curved sensograms (7.8-125 µM) indicated the association and dissociation behaviour and thereby binding of MB61B and MB63N to galectin-8*N* (Figure 5.15). However, the dissociation of the ligand was incomplete as noticed from the sensogram not reaching the baseline after the provided contact time, particularly noted at higher ligand concentrations (0.5 – 2 mM). Furthermore, the RU noted for both ligands were in the range of 200-1000 which is much higher than the calculated theoretical R_{max} of ~120-140 RU (Figure 5.15). This behaviour could be due ligand aggregating onto itself in the binding site or binding at multiple sites on the protein (super stoichiometric binding) or

very high concentrations of ligand or perhaps insufficient dissociation time. During the method development stage, these factors were further assessed.

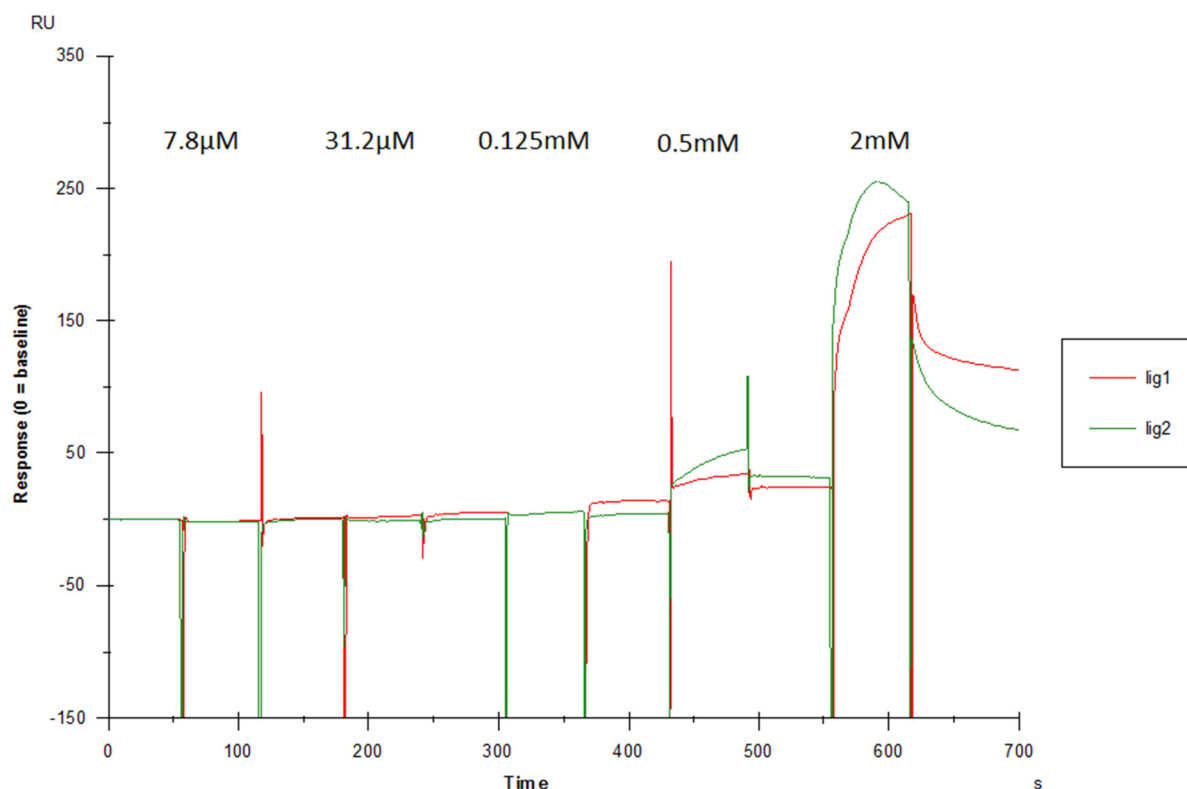


Figure 5.16: Binding sensogram for MB61B (lig1) and MB63N (lig2) at a wider 2 mM to 7.8 μ M concentration window.

The single cycle kinetics experiments were designed to determine the working concentration range and the dissociation constant for the ligands. For all the experiments, the response recorded on FC1 (the reference cell) with no immobilised galectin-8N, was subtracted from the actual response sensogram from FC2 (reference corrected). Further, due to low aqueous solubility, 5 % DMSO was supplied both in the sample preparation and the running buffer PBS. For this reason, a solvent correction curve was prepared by injecting a varying amount of DMSO (3-8% DMSO range) in the buffer and was subtracted from the reference corrected sensogram (double reference corrected) before analysis. Performing a solvent correction is critical as it removes the bias introduced in the sensogram by the DMSO. A broader concentration of 2 mM to 7.8 μ M ligands were injected onto the immobilised galectin-8N. The RUs recorded for the first three injections (7.8, 31.2, 125 μ M) were incremental but fit of data to a steady state affinity model was unsuccessful. This was due to the unusually high response (up to 200 RU) recorded for the last two concentrations where the non-zero baseline was not achieved at the end of sample injection (Figure 5.16). One reason for such a higher response (last two injections) could be due to the higher ligand concentrations that are causing

the ligands to stick to the chip surface. Therefore, the working ligand concentration was lowered to a micromolar concentration range (0.7 - 200 μM) for all the following experiments.

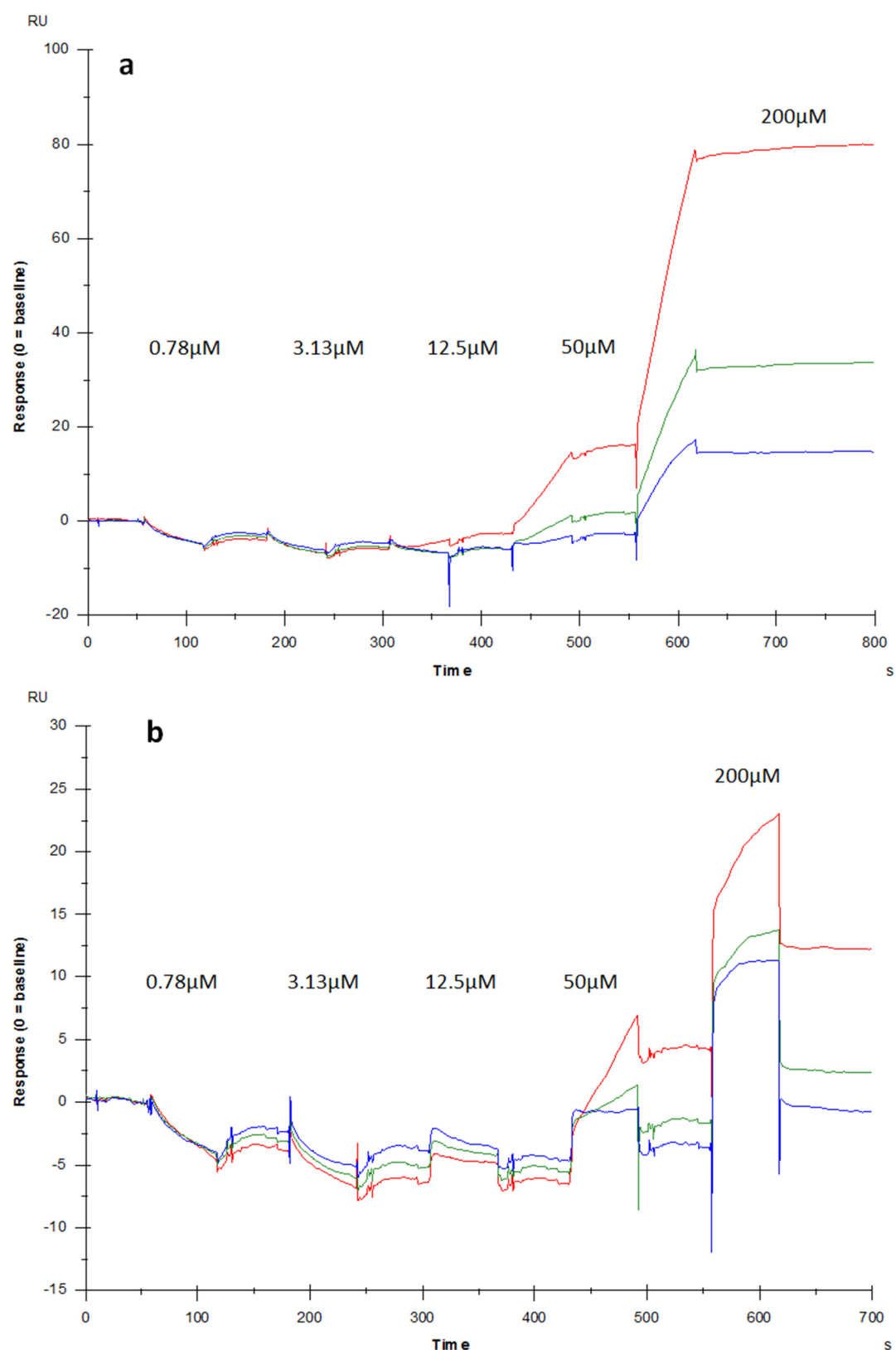


Figure 5.17: Binding sensogram for triplicate injection of MB61B (a) and MB63N (b) at a narrow concentration window of 200 μM to 0.78 μM .

However, the highest concentration injection again leads to a relatively increased response (in comparison with other concentrations) and a similar baseline behaviour to previous SPR run was noted (Figure 5.17). Surprisingly, the binding sensogram of MB61B at 200 μM concentration appeared to be through a two-step binding mechanism, which however is not possible particularly due to smaller ligand size and solvent exposed binding site. The observed unusual response may be associated with the overall loss of activity of galectin-8N upon storage for a few days. Another reason for the loss of activity of protein could be due to the method of immobilisation. The presence of a few positively charged residue in the galectin-8N binding site may also involve in the amide bond formation during immobilisation and result in loss of activity of protein. The storage conditions suggested by the manufacturer might also be not suitable for galectin-8N. This overall would interfere with galectin-8N's ability to interact with ligands leading to unusual peaks in the sensogram.

Therefore, a fresh batch of protein was immobilised on the unused flow cell on the same chip. The flow cell 4 (FC4) was used to immobilise the fresh protein, and the flow cell 3 (FC3) was used as the reference cell. The previously used ligand concentration range was used for this experiment to assess whether the condition of the protein was responsible for the previous results. Interestingly, during this run, the highest ligand concentrations produced expected typical responses (Figure 5.18), unlike to the previously obtained responses with old protein (Figure 5.16). A triplicate of the obtained data was fitted into a steady state affinity model to determine the average dissociation constants. The affinity of lactose was determined to be 76.6 μM which is in agreement with the previously reported affinity (79 μM) by SPR, although the authors have to use GST-tagged galectin-8N as opposed to our analysis on the untagged protein [42]. The variation in binding affinity based on the method employed was also evident from our ITC data determining the affinity of lactose (Chapter 4) to be 136 μM , as compared to 76.6 μM affinity determined by SPR. The average binding affinity determined from a triplicate run for MB61B and MB63N were 123.6 μM and 124.4 μM respectively.

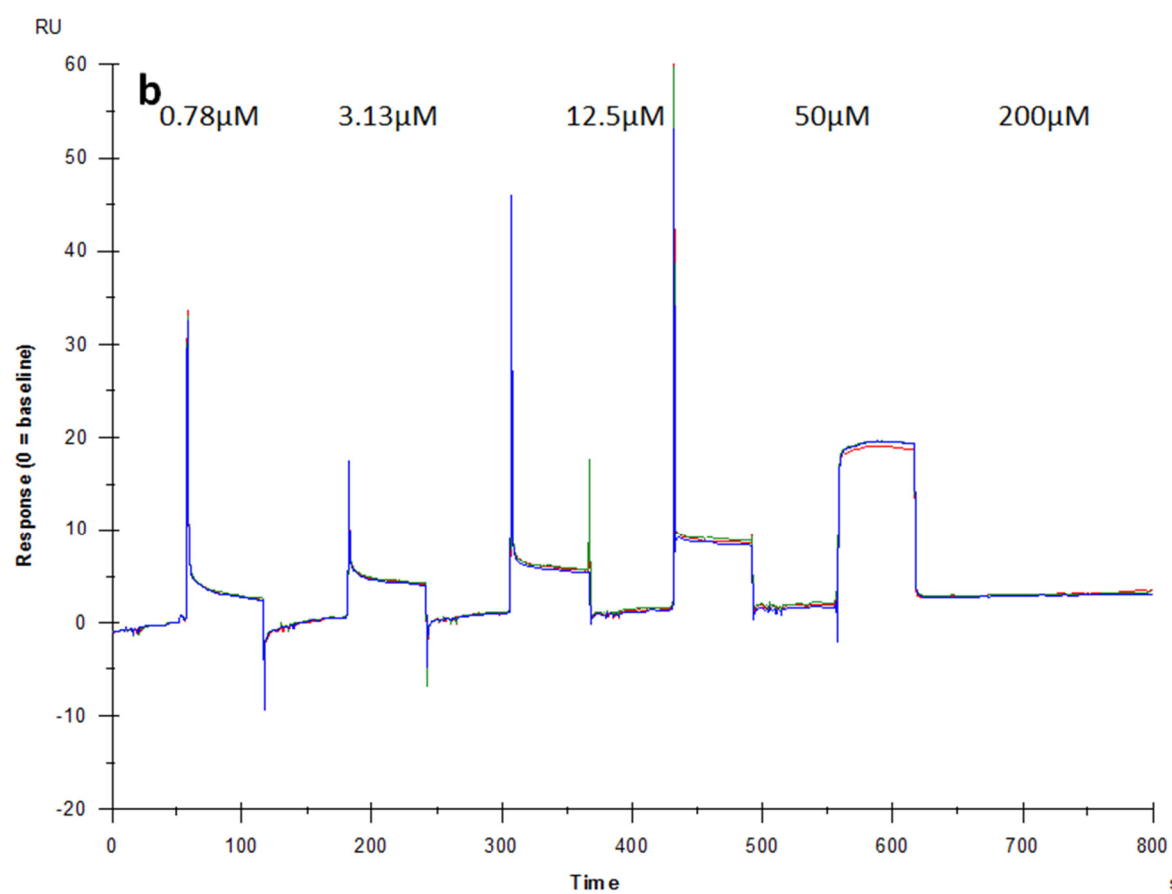
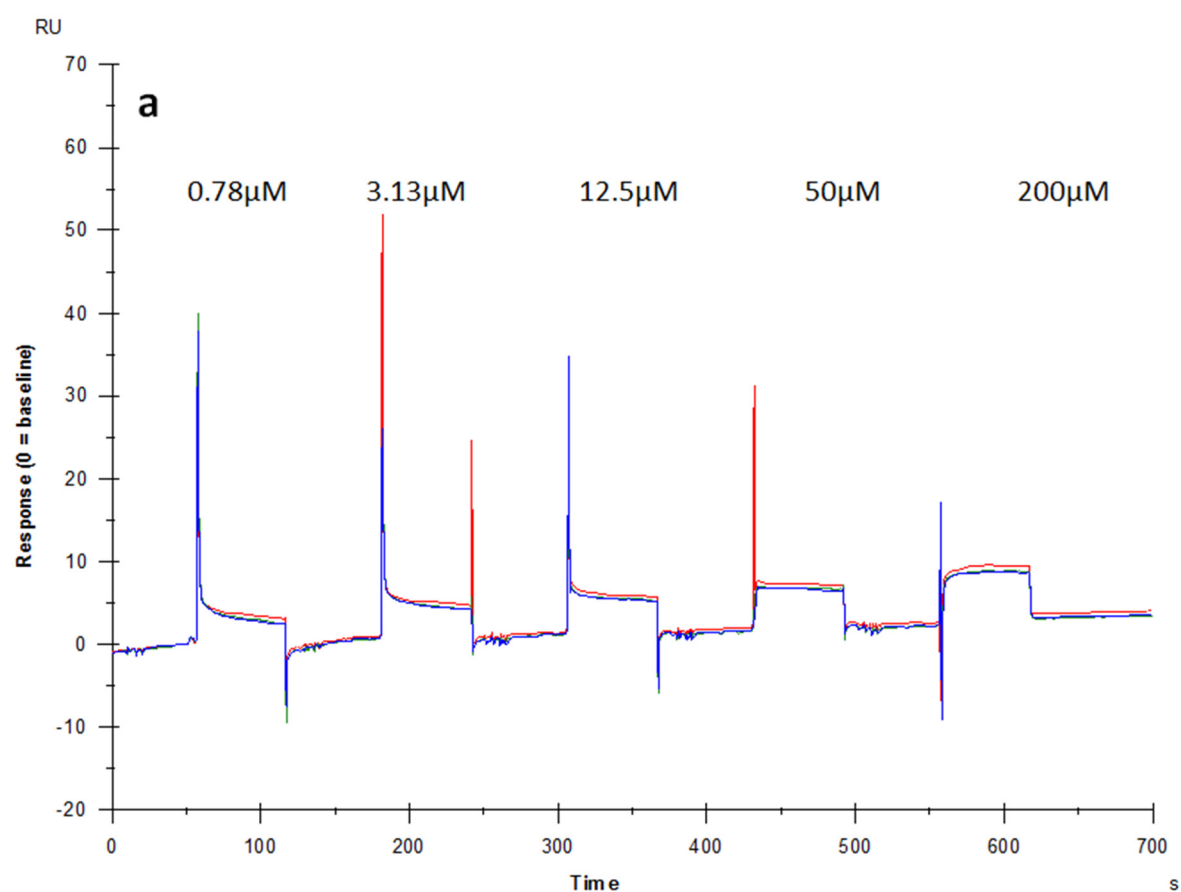


Figure 5.18: The optimised binding sensogram (in triplicate) for MB61B (a) and MB63N (b) that was fitted into the steady state model to obtain the dissociation constants.

5.3.7 Binding mode and interactions

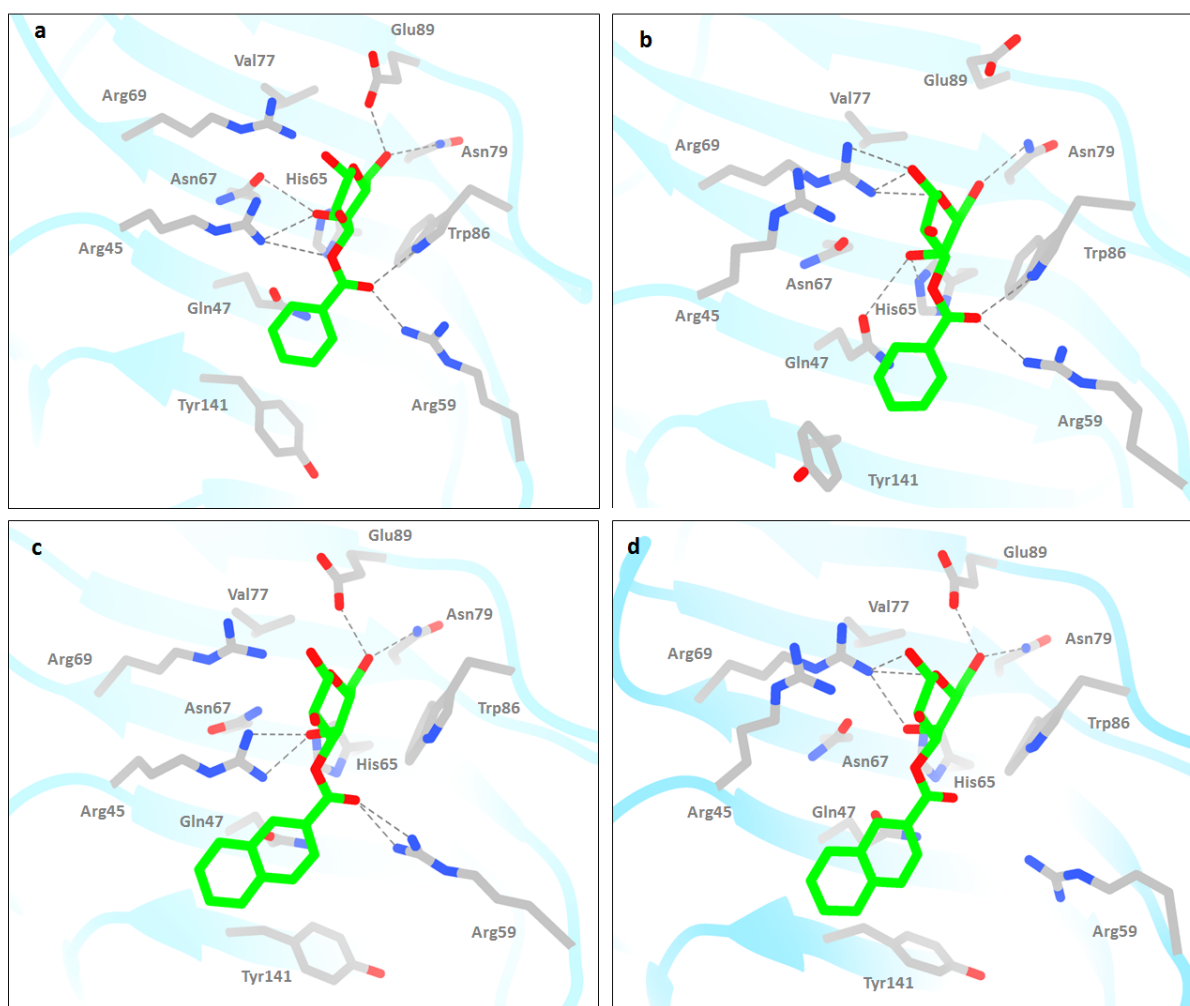


Figure 5.19: The snapshots from the MD simulation of the in silico generated galectin-8N-MB61B (on top; a, b) and galectin-8N-MB63N (at bottom; c and d) complexes at the start (a and c) and end (b and d) of MD simulations

Having determined the affinity values by SPR, the binding mode and interaction of the designed ligands (MB61B and MB63N) were placed in the galectin-8N primary binding site were analysed through 100 ns MD simulation. Overall, both ligands were retained in the binding site for the duration of simulation, where the galactose ring occupied the primary binding site. The hydrogen bonding interactions noted for the galactose ring of both MB61B and MB63N were mostly identical to those made by previously designed MB46A. The interactions include hydrogen bonds between the O4 and Arg45 and His65; O6 and Glu89 and Asn79 and van der Waal's type interactions upon stacking with the evolutionarily conserved Trp86 (Figure 5.17). The additional carbonyl group of the linker in MB61B occasionally engaged in hydrogen bond (30 % occupancy) with the unique Arg59. While in the case of MB63N simulation, the hydrogen bond with Arg59 was not observed. This difference in interaction pattern for galectin-8N-MB63N complex may be due to increased bulkiness (Figure

5.17). The aromatic ring in both the ligands was placed towards the extended binding site. The benzene ring in MB61B is placed over the Gln47 and gets involved partly in π - π type interactions with Tyr141 (Figure 5.18 and Figure 5.20). The side chain of Tyr141 experience a flip during simulation and gets involved in T-type interactions with the aromatic ring of MB61B (Figure 5.17b). However, in case of MB63N the ligand experience about 180° flip after about 1.5 ns. This flip causes the naphthyl ring to be placed over the strand S5 between the residues Arg69, Cys75 and Ile91. This at the end of 90 ns flips back to the initial position, indicating its likeliness to interact with galectin-8N. The bulky aromatic system in MB61N appeared to be not well accommodated like the phenyl ring of MB61B and therefore may experience the flipping. This is also evident from the comparable enthalpy estimated for MB61B (-36.5 kcal/mol) and MB63N (-36.6 kcal/mol) from our MMPBSA analysis (Table 5.1). In addition, no difference in binding affinity determined by SPR for MB61B and MB63N corroborates with the simulation results and estimated binding free energies. Having observed good agreement between our simulation predicted conformation and interactions in Chapter 4, it is anticipated that the results from the MD simulations presented here would potentially reflect the experimentally observed binding conformation and placement.

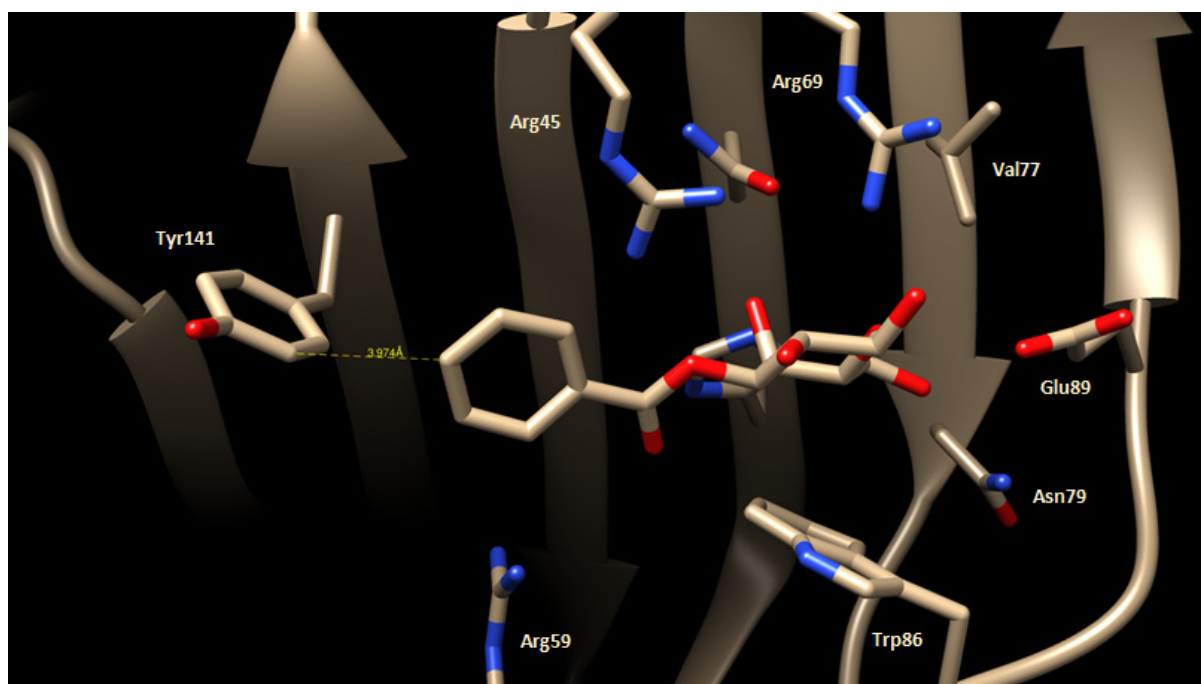


Figure 5.20: MD simulation snapshot of galectin-8N-MB61B complex depicting flipping of Tyr141 side chain and its interaction with the aromatic ring of the ligand.

5.4 Conclusions

Employing a combination of theoretical and experimental methods is a more efficient way when designing ligands through either structure-based or ligand-based approaches. The high-performance computing resources, chemical synthesis and assessment of various biochemical aspects of binding in addition to structural biology proved to be of great value for the current project. Carrying forward the ligand design campaign started against the galectin-8N, in this chapter, design and evaluation of two monosaccharide-based galactose-containing ligands was carried. The ligands in the present work contain an aromatic ring in place of the carboxylic acid group on the methyl galactose core. The wealth of information generated during structure-based virtual screening (in Chapter 3) for identifying non-carbohydrate-based binders of galectin-8N, was used to conceptualise the present work. The monosaccharide-based carboxylic acid-containing ligand designed in Chapter 4 formed the template for building the library of compounds.

The designed ligands (MB61B and MB63N) with hydrophobic groups are much better from a pharmacokinetics perspective compared to MB46A. The replacement of the carboxylic acid group with the aromatic ring balances the overall polarity of the compound. These compounds therefore, stand a better chance to be absorbed from the gut if administered orally or may show better cell-to-cell penetration/cell permeability. However, the pharmacokinetics of MB46A can also be improved by generating a prodrug form of the free carboxylic acid group.

Ligand selectivity is a critical aspect of any design campaign but needs to be modulated such that both the affinity and selectivity are balanced. In Chapter 4, a carboxylic acid ligand that exploited the unique residue Arg59 and Gln47 for interactions was designed. However, with ligands in the present work, a benzene or naphthalene ring was added through an ester linkage. With the ester linkage, the designed ligands are provided with additional flexibility by the carbonyl group and therefore might not interact with the unique Arg59 like the carboxylic acid ligand. Nevertheless, the aromatic ring of these ligands is predicted to stack against the Arg45, which is positioned across from the unique Arg59, and possibly involves in cation- π type interactions. However, our SPR and MD analysis revealed that larger aromatic systems than benzyl at the 3-position through ester linkage might not have much advantage. The benzyl ring was noted to explore another unique residue Tyr141 in the extended binding site, while naphthyl ring at the same position was less likely to interact with Tyr141.

Presently, the ligands designed are towards the galectin-8*N*, but the possibility of their interactions with the galectin-8*C* also exists (future experiments; Chapter 6). Because the designed ligands are based on monosaccharide galactose core with aromatic rings may be accommodated in the galectin-8*C* binding site. The corresponding residues to Arg45 and Tyr141 in galectin-8*C* are Ser and Asn. The investigation of binding of the designed ligands to the galectin-8*C* potentially forms the basis for future work. However, their contribution to overall binding affinity might not be strong as compared to that in galectin-8*N*. Furthermore, the presence of the unique Arg59 in the galectin-8*N* binding site would also play a critical role in binding of these ligands.

In summary, a successful application of theoretical and experimental methods was demonstrated towards the design and development of ligands targeting galectin-8*N*. Through this study, it was demonstrated that computational techniques can be employed to generate the ligand design hypothesis which after rigorous computational analysis led to identification of potential lead molecules. Taking the polarity changes in to prior consideration, more efficient ligands were designed, this furthermore eased the overall handling of these compounds during chemical synthesis and purification. Unlike MB46A, compounds in this project required DMSO for enhancing their aqueous solubility. The binding affinity of MB61B and MB63N was relatively lower in comparison to lactose than almost identical affinity noted for MB46A. However, the relatively decreased binding affinity and lowered aqueous solubility is acceptable given the gained pharmacokinetics benefits through the aromatic ring. Therefore, it is critical to consider all the ligand development aspects and implement during the design process to efficiently carry out a design campaign.

5.5 Appendix

5.5.1 MMPBSA preparation script

5.5.1.1 tleap.in

```
set default PBradii mbondi2
source leaprc.ff99SBildn
source leaprc.gaff
REC = loadpdb 3AP6-md.pdb
lig = loadmol2 5_bcc_gaff.mol2
COM = combine {REC lig}
saveamberparm REC prot.prmtop prot.inpcrd
saveamberparm lig lig.prmtop lig.inpcrd
saveamberparm COM com.prmtop com.inpcrd
quit
```

5.5.1.2 mmpbsa.in [GBSA and normal mode analysis]

```
Input file for running PB and GB in serial
&general
    endframe=100, keep_files=1,
/
&gb
    igb=2, saltcon=0.100,
/
&nmode
    nmstartframe=1, nmendframe=100, nminterval=10
    maxcyc=50000, drms=0.0001
/
```

5.5.1.3 mmpbsa.in [PBSA]

```
Input file for running PB and GB in serial
&general
    endframe=100, keep_files=1,
/
&pb
    istrng=0.100,
```

5.5.1.4 Submit_pbsa.sh

```
#!/bin/bash -l

#PBS -N nm-gal005
#PBS -l walltime=60:00:00
### Number of nodes: Number of CPUs: Number of threads per node
```

```

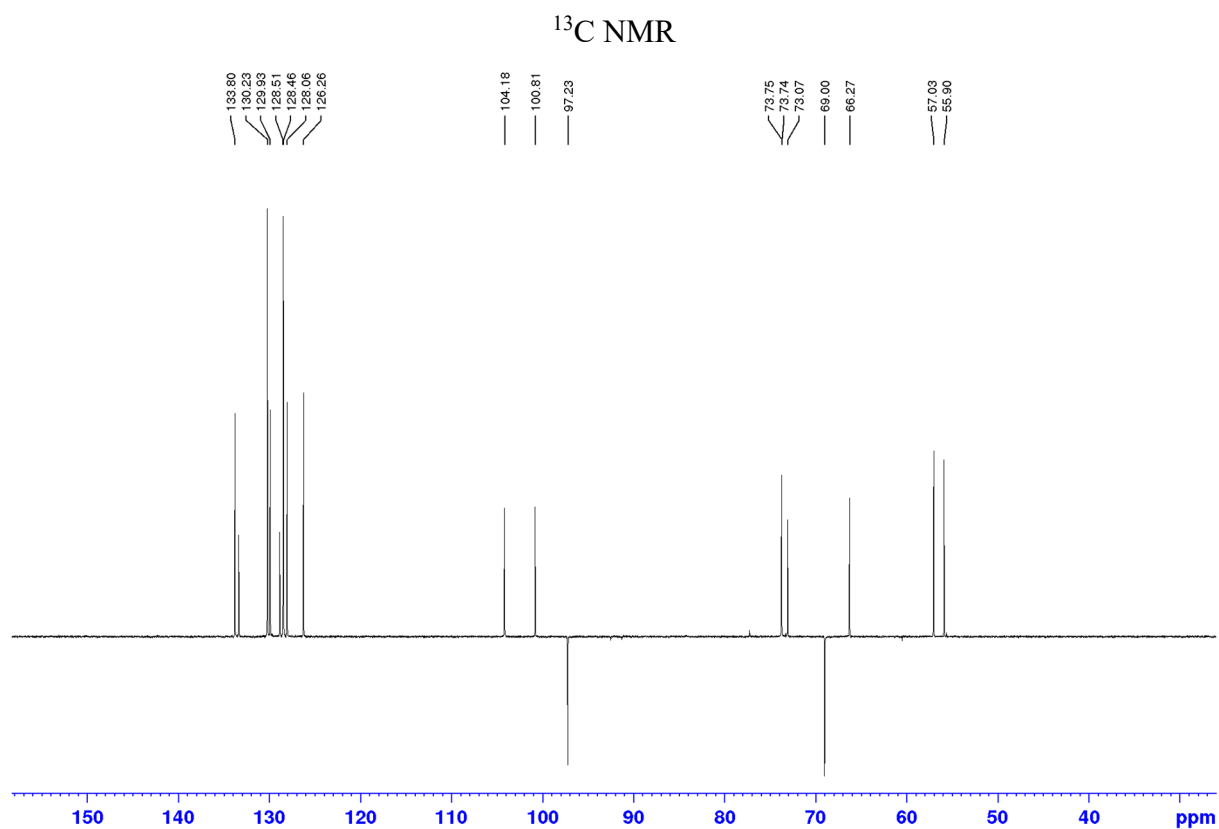
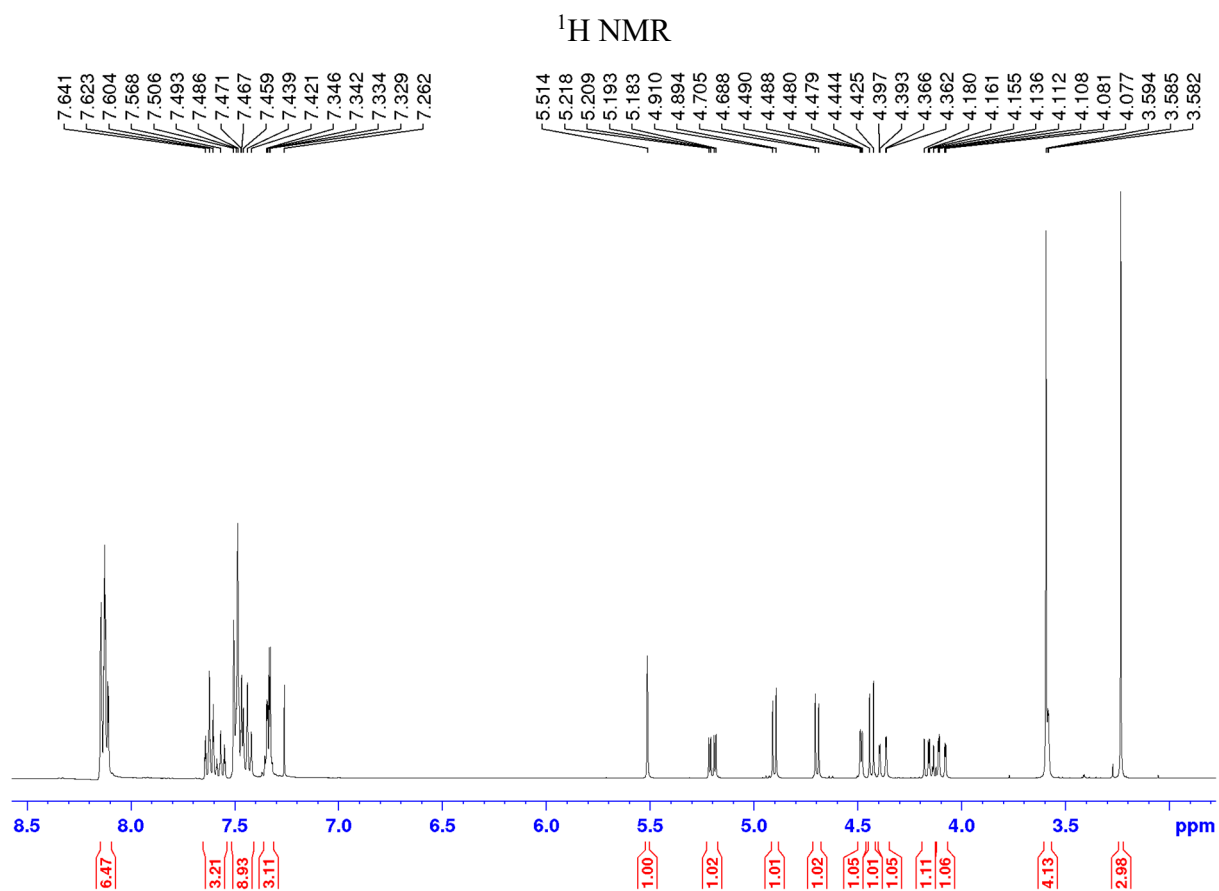
#PBS -l select=1:ncpus=4:mem=1g:mpiprocs=4
## The number of nodes is given by the select =<NUM > above
NODES=1
##$PBS_NODEFILE is a node-list file created with select and mpiprocs
options by PBS
### The number of MPI processes available is mpiprocs * nodes
NPROCS=4
# This job's working directory
echo "Working directory is $PBS_O_WORKDIR"
cd $PBS_O_WORKDIR
source $HOME/.bashrc
module load amber/12-modified
echo "Starting job"
echo Running on host `hostname`
echo Time is `date`
echo Directory is `pwd`
#echo This jobs runs on the following processors:
echo `cat $PBS_NODEFILE`

MMPBSA.py -O -i mmpbsa.in -o FINAL_RESULTS_MMPBSA.dat -cp com.prmtop
-rp prot.prmtop -lp lig.prmtop -y traj_100.mdcrd

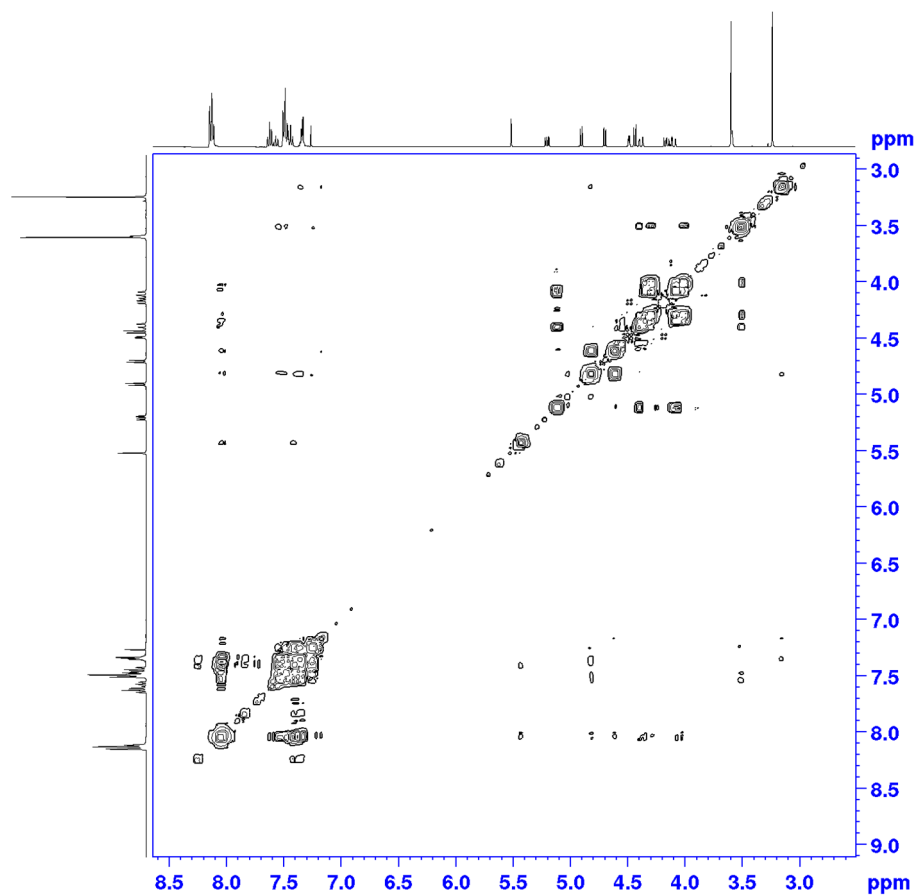
```

5.5.2 Spectral data

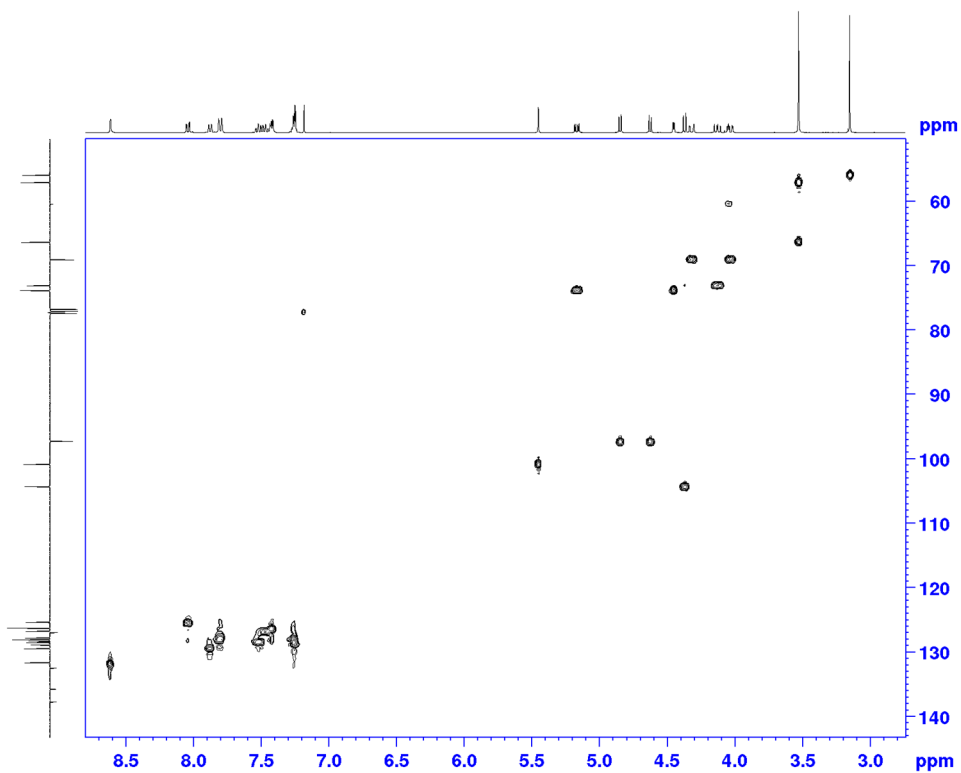
Methyl 2-*O*-Methoxymethyl-3-*O*-benzoyl-4,6-*O*-benzilidene-β-D-galactopyranoside (5)



^1H - ^1H COSY

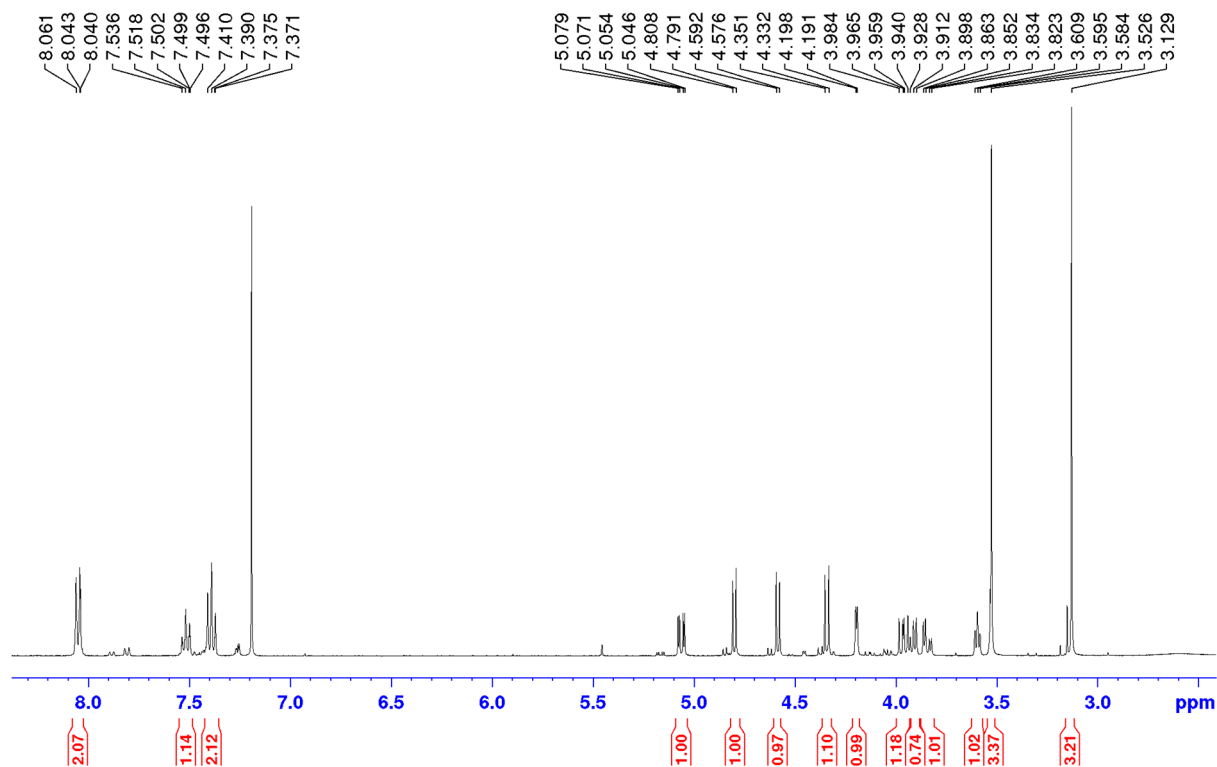


^1H - ^{13}C HSQC [short range]

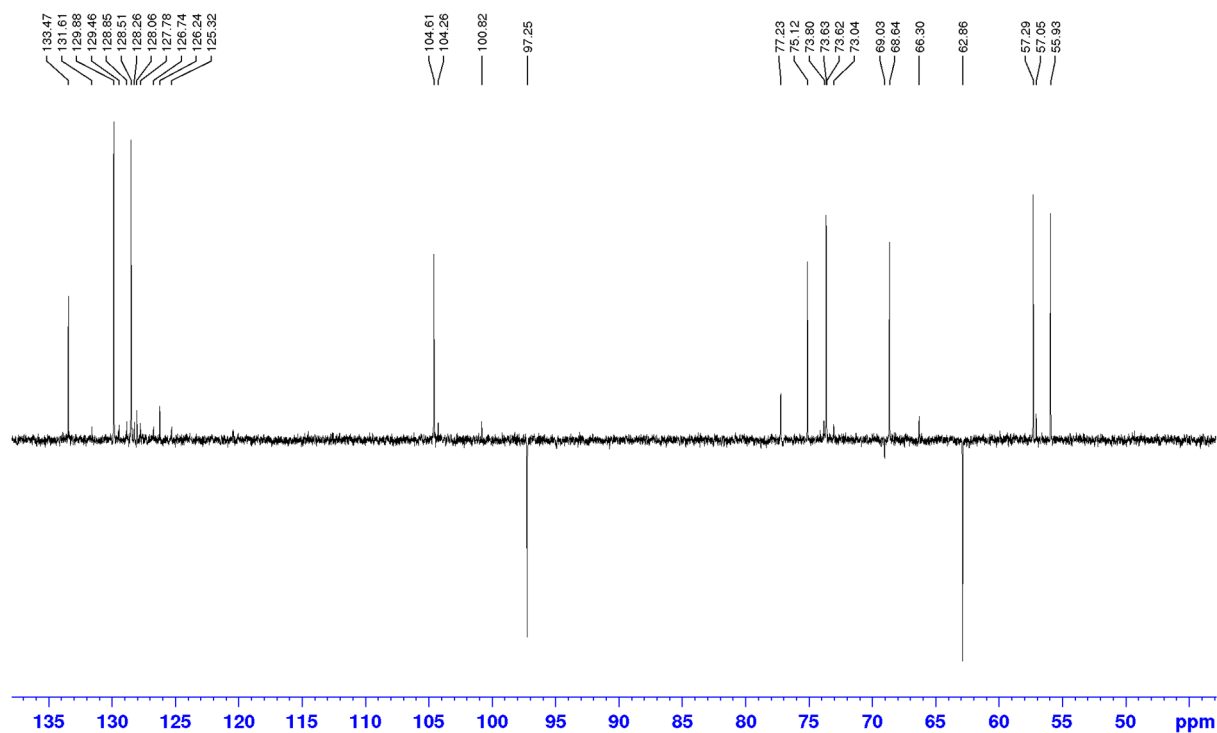


Methyl 3-*O*-benzoyl- β -D-galactopyranoside (MB61B)

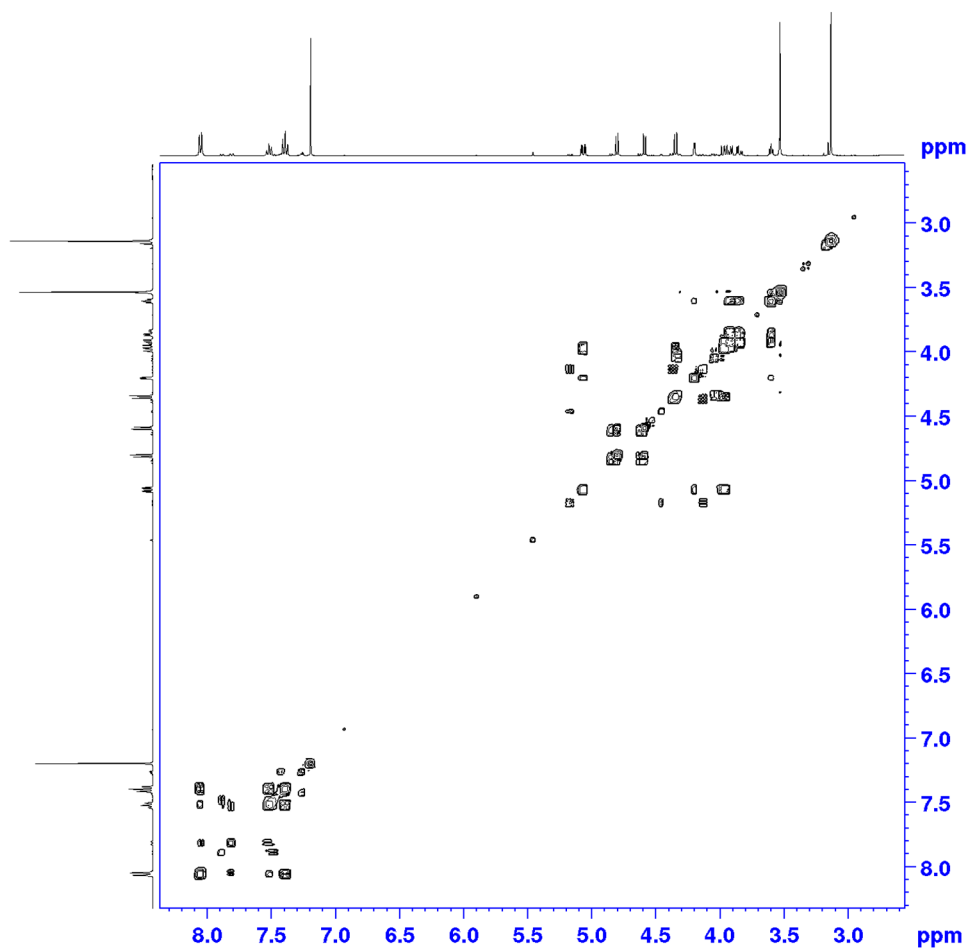
^1H NMR



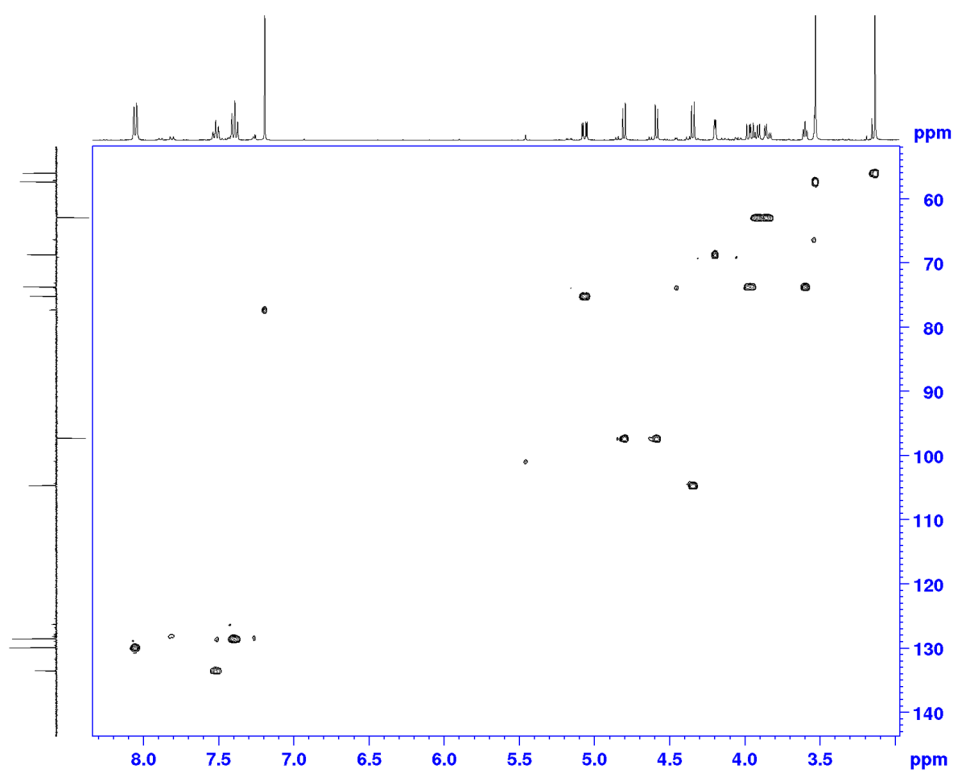
^{13}C NMR



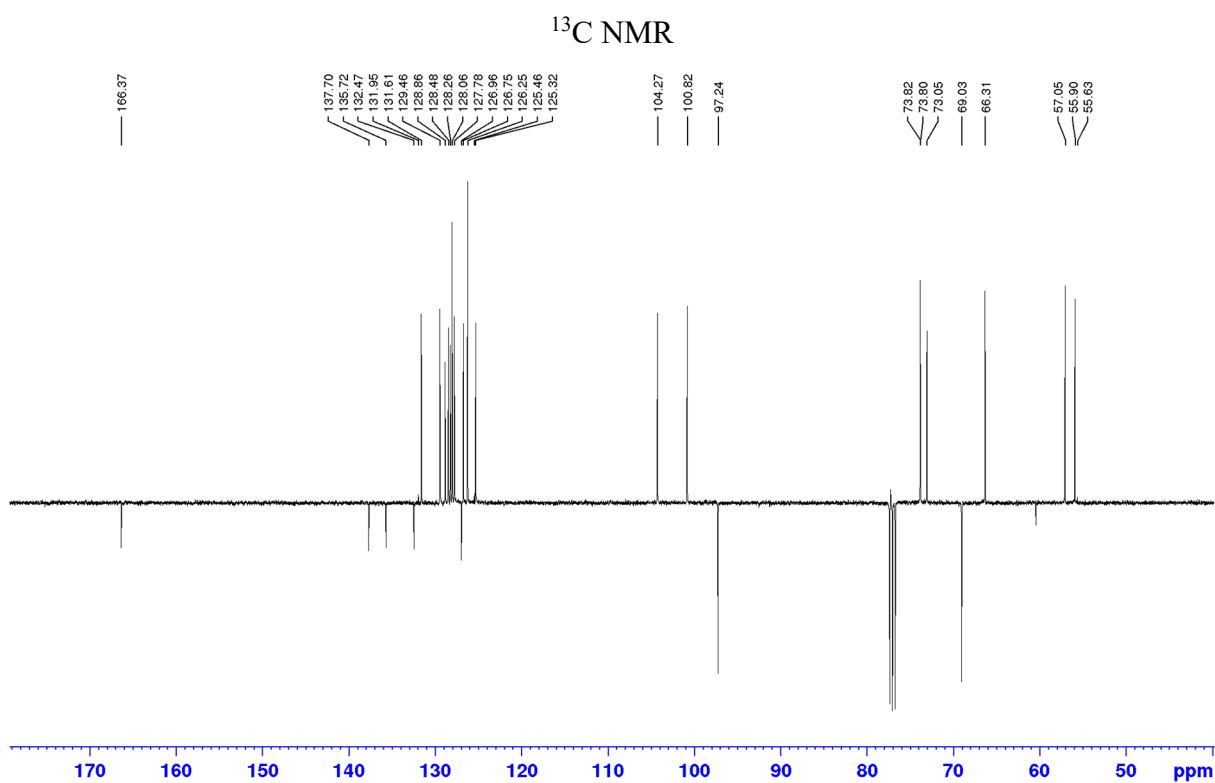
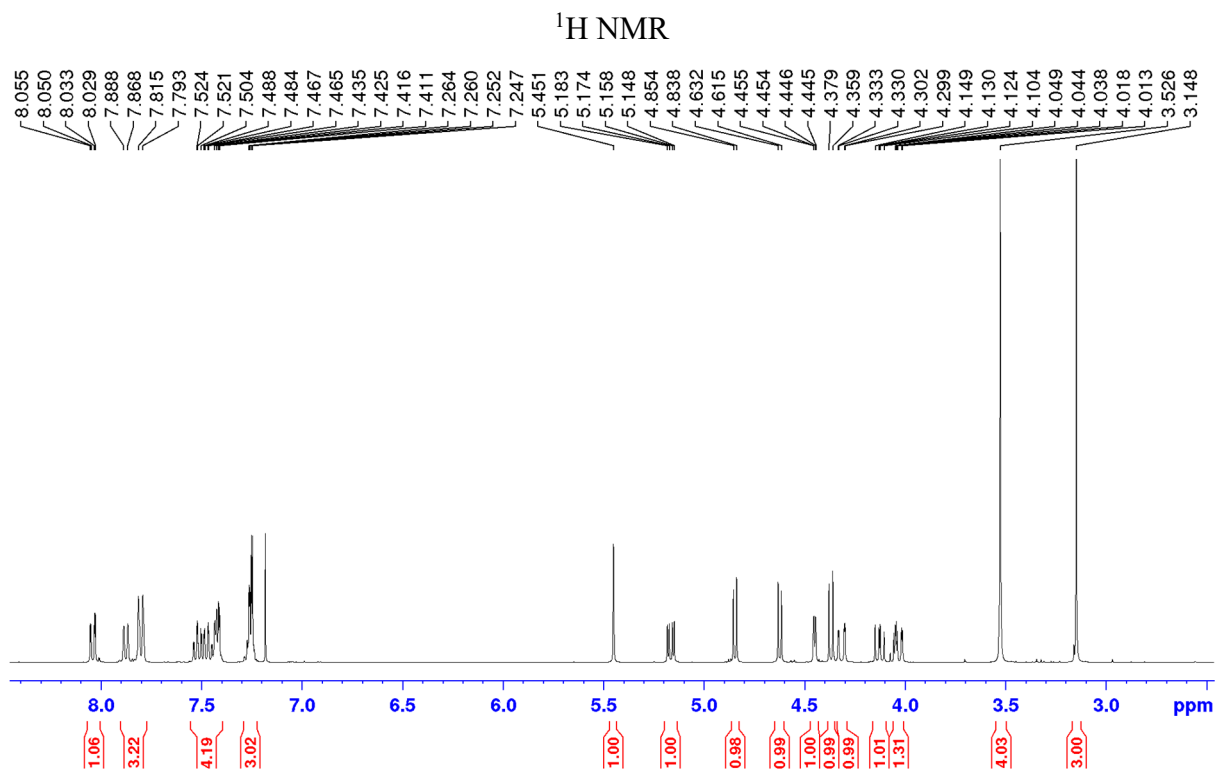
^1H - ^1H COSY



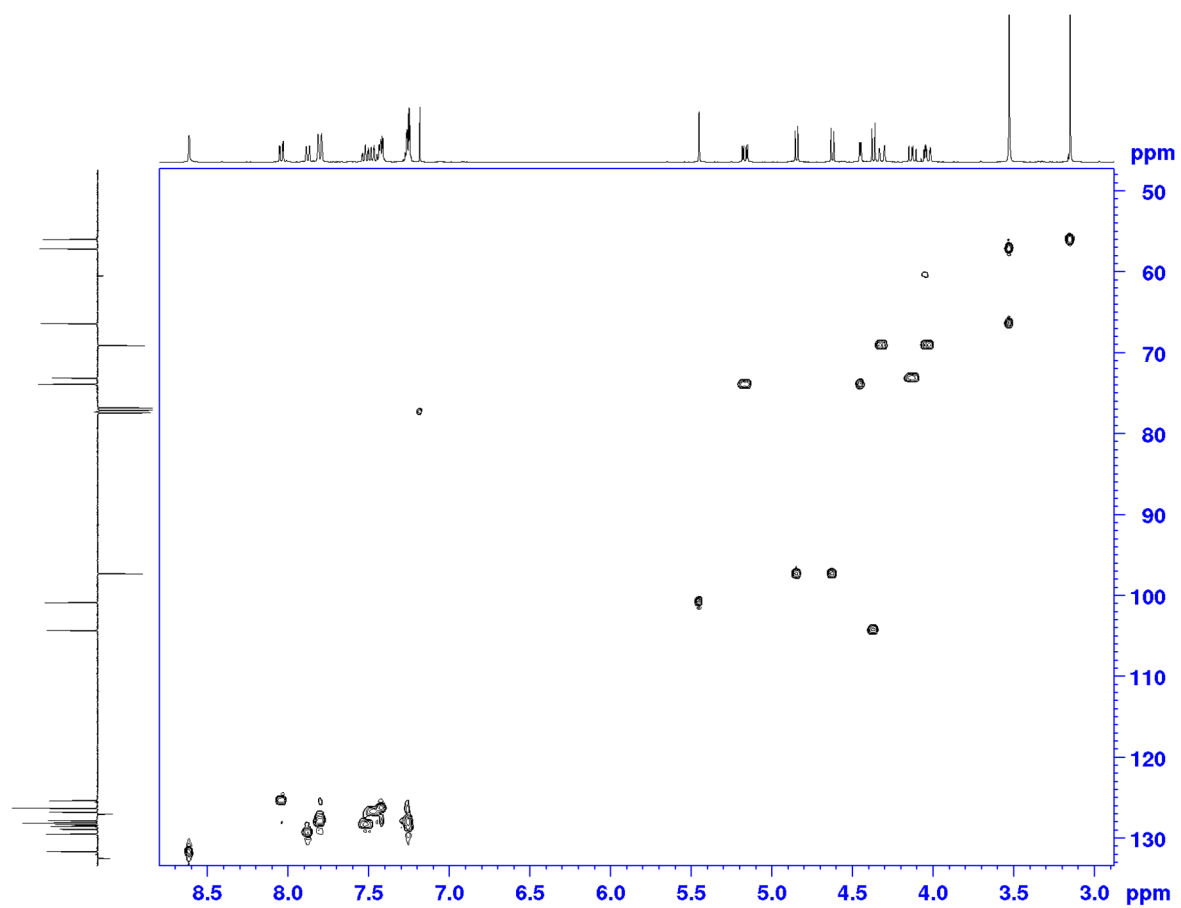
^1H - ^{13}C HSQC [short range]



Methyl 2-*O*-Methoxymethyl-3-*O*-naphthoyl-4,6-*O*-benzilidene- β -D-galactopyranoside (7)

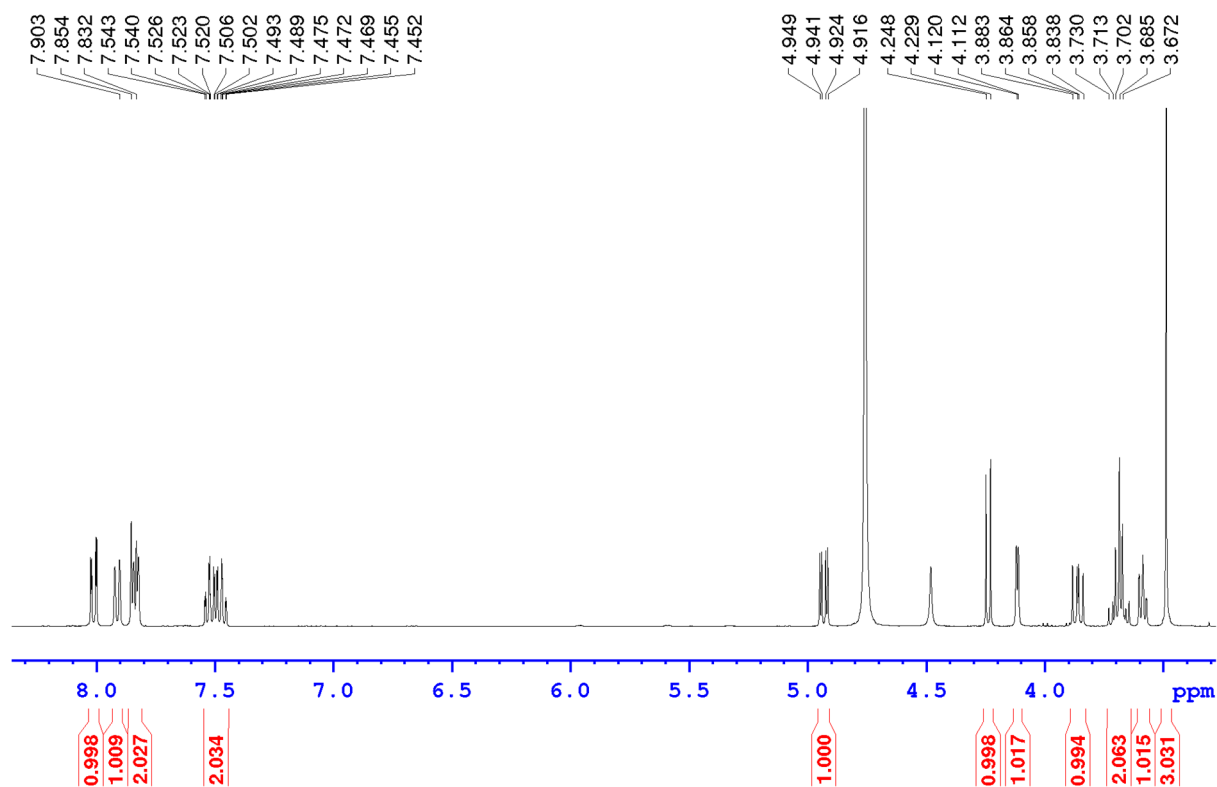


^1H - ^{13}C HSQC [short range]

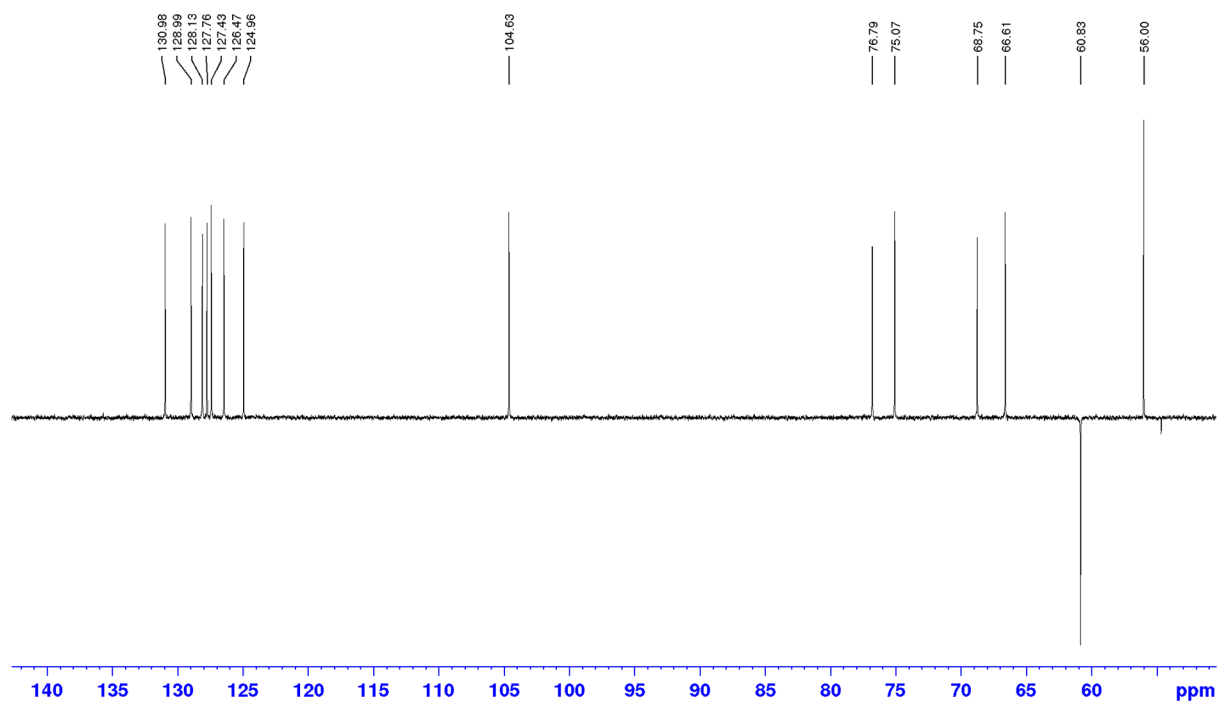


Methyl 3-*O*-naphthoyl- β -D-galactopyranoside (MB63N)

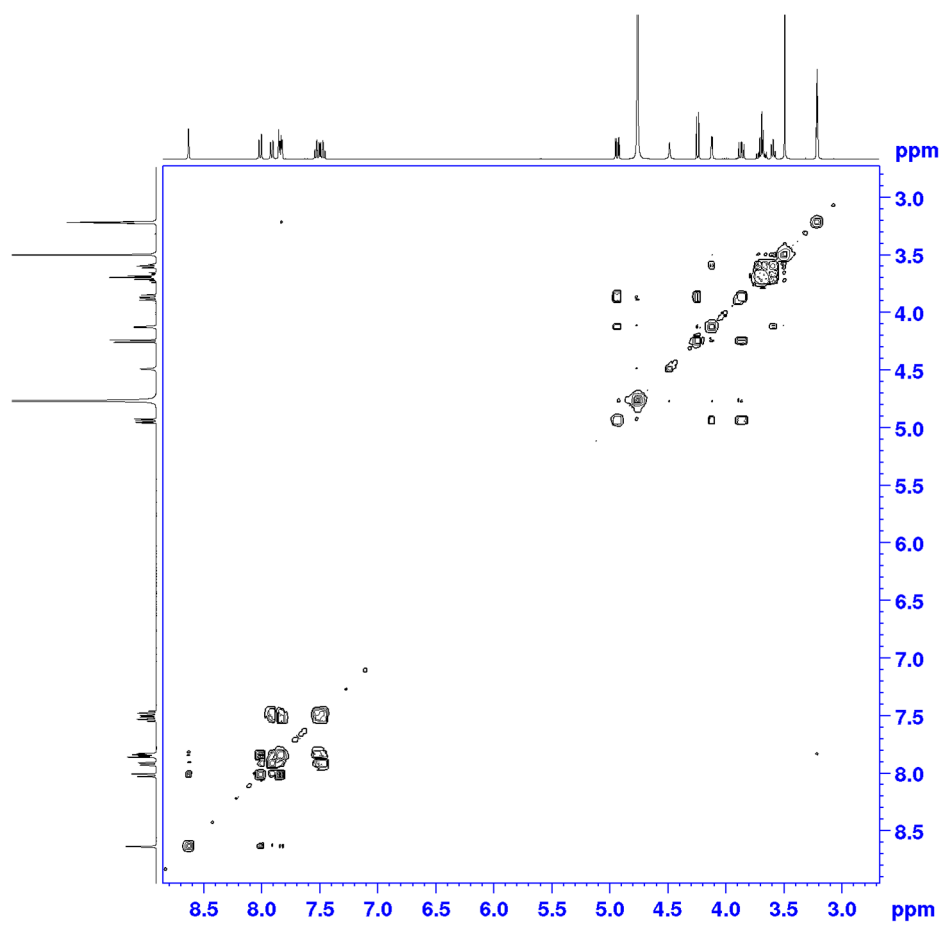
^1H NMR



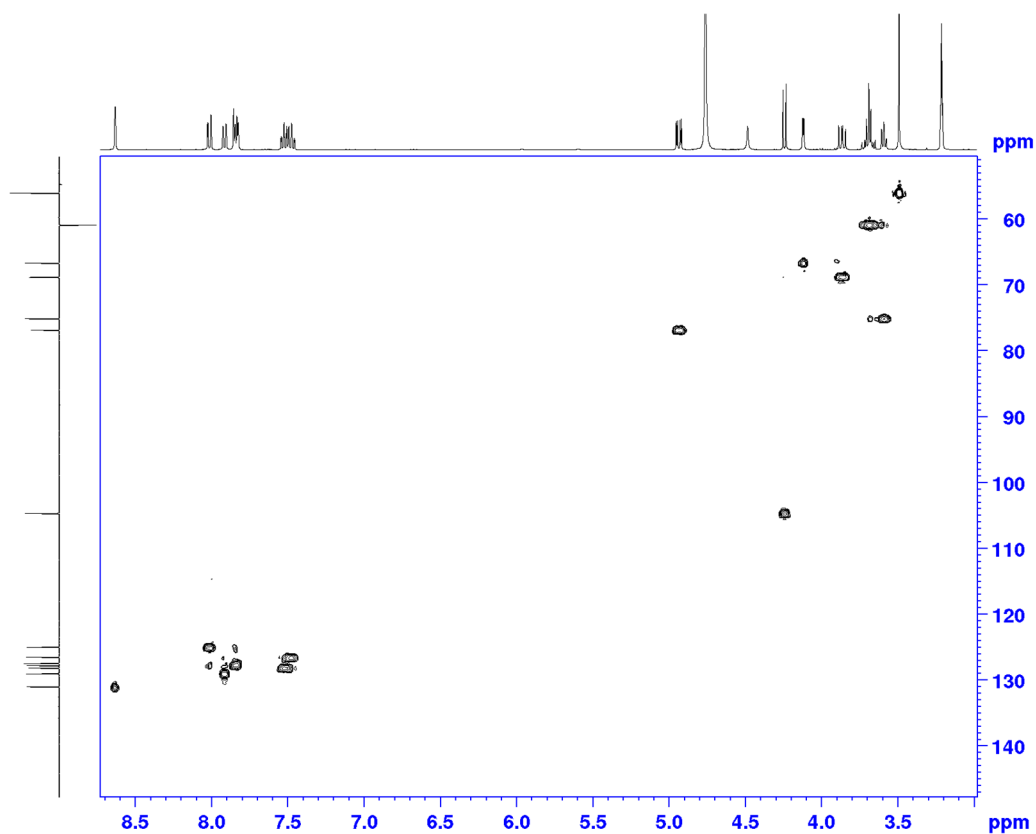
^{13}C NMR



^1H - ^1H COSY



^1H - ^{13}C HSQC [short range]



5.5.3 FDA Drug fragment

The compounds purchased during structure-based virtual screening analysis were examined through STD NMR for binding to galectin-8*N*. However, as different methods show varying results, those compounds can also be tested for binding to galectin-8*N* using a different method such as SPR or ITC. The FDA drug PMT was one of the most promising compounds among all the molecules evaluated, mainly because of its two carboxylic acid groups displayed the ability to cross-link the Arg45 and Arg59 in the galectin-8*N* binding site. However, soaking of PMT into the *apo* galectin-8*N* crystals was unsuccessful, and neither did it show binding in our STD experiments. Potentially the bulky aromatic rings (pyrrolopyrimidine ring system) in PMT could be interfering with binding, and therefore no binding for PMT in our STD experiments was detected.

The fragment of PMT (PMT-frag) only contains the aliphatic carboxylic acid side chain amide linked to the phenyl ring, is commercially available (Figure 6.3). These carboxylic acid groups were of interest as during the simulation of the galectin-8*N*-PMT complex they could cross-link Arg59 and Arg45. During the SPR experiments conducted for evaluating binding of MB61B and MB63N, this fragment was also included to investigate its binding to galectin-8*N*. Interestingly, the curvy binding pattern of the sensogram revealed binding of PMT-frag to galectin-8*N*, however, preliminary data was fitted into steady state model to obtain 0.86 mM affinity. Further optimisation of the assay protocol is required to derive precise binding affinity. Similarly, SPR based binding screening can also be performed, in addition to STD NMR, for the purchased compounds (Chapter 3), to further investigate the binding of non-carbohydrate compounds towards galectin-8*N*.

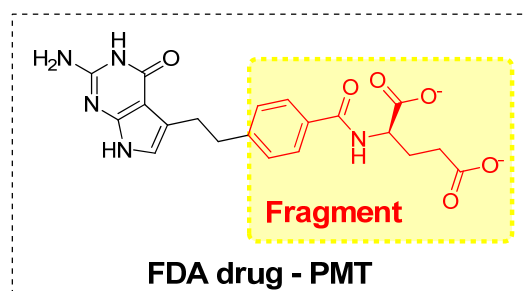


Figure 5.21: Structure of FDA drug PMT with yellow highlighted section showing the structure of purchased FDA fragment.

Summary and Conclusions

In this research thesis, a combination of theoretical and experimental techniques has been employed to perform structure-based ligand design targeting galectin-8. This tandem-repeat lectin is involved in various biological processes such as cell adhesion, cell growth, cell migration, immunomodulation, autoimmunity, inflammation cancer and bone remodelling process. The modulation of bone remodelling process via inducing expression of RANKL mediated through ERK signalling pathway can be a potential new approach to tackle diseases associated with bone-loss. Structure-based approaches including molecular modelling, X-ray crystallography and biophysical techniques were employed to design, develop and evaluate binders of galectin-8. All the biological assays evaluating the inhibitory potential of the ligands towards galectin-8 are performed by our collaborator Professor Yehiel Zick (Weizmann Institute of Science, Israel).

The ligand design campaign was initiated by investigating the galectin-8N binding site residues for governing specificity in recognising human milk glycans. These glycans included the tetrasaccharides LNT and LNnT, which differs only in the glycosidic linkage between the non-reducing end disaccharide. However, regarding the affinity towards galectin-8N, they exhibit a ten-fold difference, a magnitude not observed with any other galectins. Our crystal structures revealed for the first time the non-reducing end disaccharide part of the tetrasaccharide LNT was occupying the primary binding site. In addition to these novel findings, our MD simulations provided the justification towards observing a unique binding mode, and Tyr141 was noted to be governing the specificity in recognising larger oligosaccharides. The complex structure with glycerol further provided insight into minimum atomic feature required for binding to galectin-8N. Overall, the information generated was then employed towards designing novel inhibitors of galectin-8.

Identification of novel non-carbohydrate-based ligands was initially aimed through structure-based virtual screening using a series of molecular docking and molecular dynamics simulations to filter a library of small molecules rationally. The purchased top fraction of compounds did not soak in the *apo* galectin-8N crystals as determined from the crystal structures and did not bind to galectin-8N in our STD NMR experiments. However, the simulation results of a purchased FDA drug attracted our attention due to its two carboxylic acid groups simultaneously engaging Arg45 and Arg59 in interaction. Exploring these arginines for interaction can prove to be very crucial from specificity perspective since Arg59 is the unique residue not found in any galectins while Arg45 is found only in few galectins. Challenges were encountered in dealing with computational methods to work on target like galectin-8N which has solvent exposed shallow binding site, for identifying non-carbohydrate

ligands. Based on these findings, the native ligand core (galactose) was employed to perform modifications, particularly on the 3'-position towards the extended binding site.

Taken together our results and those previously published, mimicking the interaction of the galactose ring and the carboxylic acid part of the 3'-*O*-sialylated lactose, a ligand that is preferentially recognised by the galectin-8N, was aimed. Methyl 3-*O*-[1-carboxyethyl]- β -D-galactopyranoside (**6**; MB46A) was then designed to exploit both the evolutionarily conserved and the unique amino acid residues in the binding site for interactions. MB46A was synthesised and shown to bind with 139 μ M affinity (from ITC). The crystal structure revealed the galactose ring stacked against the evolutionarily conserved tryptophan (Trp86), and the carboxylic acid interacted with the Arg59. Following this successful design, MD simulation based structure-activity relationship study was undertaken to identify other possible modification that can be performed on the galactose core. The simulations suggested that replacing the carboxylic acid with an aromatic ring would also result in a similar affinity gain. Subsequently, the carboxylic acid group of MB46A was replaced with the benzoyl (MB61B) and naphthoyl (MB63N) rings and the observed binding affinities were 123.6 μ M and 124.4 μ M (by SPR). The identical affinity for both the ligand indicated no particular gain in the affinity with additional aromatic ring (in MB63N ligand). These ligands are under *in vitro* investigations by our collaborator (Prof. Yehiel Zick, Weizmann Institute of Science, Rehovot, Israel), who are assessing decrease in the RANKL expression upon ligand treatment in galectin-8 treated osteoblasts. A successful design and evaluation of monosaccharide-based ligands modified with either propionic acid side chain ether linked or ester linked benzoyl or naphthoyl groups on the 3'-position of methyl galactose was thus carried out. Overall, reflecting upon an efficient structure-based campaign employed towards design and developing potential inhibitors of galectin-8.

Galectins are the evolutionarily conserved class of lectin found in all forms of living organism performing various biological functions. Galectin-8 is involved in various metabolic and disease states, of interest was its ability to modulate bone remodeling process. To this end, structure-based approaches were employed to design and develop inhibitors of galectin-8 that can potentially work in *in vitro* and *in vivo* set up. With the apparent increase in the functional spectrum of galectin-8, our inhibitor design project is the first report of molecules specifically designed targeting galectin-8. The computational techniques in combination with X-ray crystallography employed in the project have provided valuable insight into the interaction of the designed inhibitors with galectin-8N. Further, these ligands would also form an excellent template for future ligand optimisations to improve upon the potency and specificity towards

galectin-8. The results generated are hoped to contribute to the ongoing galectin inhibitor design area and more so would usher the way towards finding novel approaches tackling diseases associated with bone-loss.

References

1. Sharon, N. and H. Lis, *Lectins: cell-agglutinating and sugar-specific proteins*. Science, 1972. **177**(4053): p. 949-59.
2. Barondes, S.H., et al., *Galectins. Structure and function of a large family of animal lectins*. J Biol Chem, 1994. **269**(33): p. 20807-10.
3. Hughes, R.C., *The galectin family of mammalian carbohydrate-binding molecules*. Biochem Soc Trans, 1997. **25**(4): p. 1194-8.
4. Gabius, H.-J., *Animal lectins and life: a guided tour into the realm of the sugar code*. Biochimica et Biophysica Acta (BBA) - General Subjects, 2002. **1572**(2-3): p. 163-164.
5. Gabius, H.J., et al., *The sugar code: functional lectinomics*. Biochim Biophys Acta, 2002. **1572**(2-3): p. 165-77.
6. Kilpatrick, D.C., *Animal lectins: a historical introduction and overview*. Biochim Biophys Acta, 2002. **1572**(2-3): p. 187-97.
7. Hirabayashi, J. and K. Kasai, *The family of metazoan metal-independent beta-galactoside-binding lectins: structure, function and molecular evolution*. Glycobiology, 1993. **3**(4): p. 297-304.
8. Teichberg, V.I., et al., *A beta-D-galactoside binding protein from electric organ tissue of Electrophorus electricus*. Proceedings of the National Academy of Sciences, 1975. **72**(4): p. 1383-1387.
9. de Waard, A., S. Hickman, and S. Kornfeld, *Isolation and properties of beta-galactoside binding lectins of calf heart and lung*. Journal of Biological Chemistry, 1976. **251**(23): p. 7581-7587.
10. Roff, C.F., et al., *Identification of carbohydrate-binding proteins from mouse and human fibroblasts*. Biochem J, 1983. **211**(3): p. 625-9.
11. Leffler, H., F.R. Masiarz, and S.H. Barondes, *Soluble lactose-binding vertebrate lectins: a growing family*. Biochemistry, 1989. **28**(23): p. 9222-9229.
12. Cooper, D.N. and S.H. Barondes, *God must love galectins; he made so many of them*. Glycobiology, 1999. **9**(10): p. 979-84.
13. Cooper, D.N., *Galectinomics: finding themes in complexity*. Biochim Biophys Acta, 2002. **1572**.
14. Gopalkrishnan, R.V., et al., *Molecular characterization of prostate carcinoma tumor antigen-1, PCTA-1, a human galectin-8 related gene*. Oncogene, 2000. **19**(38): p. 4405-16.
15. Hughes, R.C., *Mac-2: a versatile galactose-binding protein of mammalian tissues*. Glycobiology, 1994. **4**(1): p. 5-12.
16. Vasta, G.R., *Galectins as pattern recognition receptors: structure, function, and evolution*. Adv Exp Med Biol, 2012. **946**: p. 21-36.
17. Vasta, G.R., *Roles of galectins in infection*. Nat Rev Microbiol, 2009. **7**(6): p. 424-38.
18. Collins, P.M., et al., *Taloside inhibitors of galectin-1 and galectin-3*. Chem Biol Drug Des, 2012. **79**(3): p. 339-46.
19. Yoshida, H., et al., *X-ray structure of a protease-resistant mutant form of human galectin-8 with two carbohydrate recognition domains*. FEBS Journal, 2012. **279**(20): p. 3937-3951.
20. Cho, M. and R.D. Cummings, *Galectin-1, a beta-galactoside-binding lectin in Chinese hamster ovary cells. I. Physical and chemical characterization*. J Biol Chem, 1995. **270**(10): p. 5198-206.
21. Cho, M. and R.D. Cummings, *Galectin-1, a β -Galactoside-binding Lectin in Chinese Hamster Ovary Cells: II. LOCALIZATION AND BIOSYNTHESIS*. Journal of Biological Chemistry, 1995. **270**(10): p. 5207-5212.
22. Albrandt, K., N.K. Orida, and F.T. Liu, *An IgE-binding protein with a distinctive repetitive sequence and homology with an IgG receptor*. Proc Natl Acad Sci U S A, 1987. **84**(19): p. 6859-63.
23. Cherayil, B.J., S.J. Weiner, and S. Pillai, *The Mac-2 antigen is a galactose-specific lectin that binds IgE*. J Exp Med, 1989. **170**(6): p. 1959-72.

24. Massa, S.M., et al., *L-29, an endogenous lectin, binds to glycoconjugate ligands with positive cooperativity*. Biochemistry, 1993. **32**(1): p. 260-267.
25. Fred Brewer, C., *Binding and cross-linking properties of galectins*. Biochim Biophys Acta, 2002. **1572**(2-3): p. 255-62.
26. Hadari, Y.R., et al., *Galectin-8. A new rat lectin, related to galectin-4*. J Biol Chem, 1995. **270**(7): p. 3447-53.
27. Su, Z.Z., et al., *Surface-epitope masking and expression cloning identifies the human prostate carcinoma tumor antigen gene PCTA-1 a member of the galectin gene family*. Proc Natl Acad Sci U S A, 1996. **93**(14): p. 7252-7.
28. Bidon, N., et al., *Two messenger RNAs and five isoforms for Po66-CBP, a galectin-8 homolog in a human lung carcinoma cell line*. Gene, 2001. **274**(1-2): p. 253-62.
29. Lobsanov, Y.D., et al., *X-ray crystal structure of the human dimeric S-Lac lectin, L-14-II, in complex with lactose at 2.9-A resolution*. Journal of Biological Chemistry, 1993. **268**(36): p. 27034-27038.
30. Bidon, N., Brichory, F., Bourguet, P., Le Pennec, J., & Dazord, L., *Galectin-8: a complex sub-family of galectins (Review)*. International Journal of Molecular Medicine, 2001. **8**: p. 245-250.
31. Hadari, Y.R., et al., *Galectin-8 binding to integrins inhibits cell adhesion and induces apoptosis*. J Cell Sci, 2000. **113** (Pt 13): p. 2385-97.
32. Levy, Y., et al., *Galectin-8 functions as a matricellular modulator of cell adhesion*. J Biol Chem, 2001. **276**(33): p. 31285-95.
33. Zick, Y., et al., *Role of galectin-8 as a modulator of cell adhesion and cell growth*. Glycoconj J, 2004. **19**(7-9): p. 517-26.
34. Levy, Y., et al., *It depends on the hinge: a structure-functional analysis of galectin-8, a tandem-repeat type lectin*. Glycobiology, 2006. **16**(6): p. 463-76.
35. Nishi, N., et al., *Galectin-8 modulates neutrophil function via interaction with integrin alphaM*. Glycobiology, 2003. **13**(11): p. 755-63.
36. Tribulatti, M.V., et al., *Galectin-8 induces apoptosis in the CD4(high)CD8(high) thymocyte subpopulation*. Glycobiology, 2007. **17**(12): p. 1404-12.
37. Tribulatti, M.V., et al., *Galectin-8 provides costimulatory and proliferative signals to T lymphocytes*. J Leukoc Biol, 2009. **86**(2): p. 371-80.
38. Yamamoto, H., et al., *Induction of cell adhesion by galectin-8 and its target molecules in Jurkat T-cells*. J Biochem, 2008. **143**(3): p. 311-24.
39. Norambuena, A., et al., *Galectin-8 induces apoptosis in Jurkat T cells by phosphatidic acid-mediated ERK1/2 activation supported by protein kinase A down-regulation*. J Biol Chem, 2009. **284**(19): p. 12670-9.
40. Cattaneo, V., María V. Tribulatti, and O. Campetella, *Galectin-8 tandem-repeat structure is essential for T-cell proliferation but not for co-stimulation*. Biochemical Journal, 2011. **434**: p. 153-160.
41. Hirabayashi, J., et al., *Oligosaccharide specificity of galectins: a search by frontal affinity chromatography*. Biochim Biophys Acta, 2002. **1572**(2-3): p. 232-54.
42. Ideo, H., et al., *The N-terminal carbohydrate recognition domain of galectin-8 recognizes specific glycosphingolipids with high affinity*. Glycobiology, 2003. **13**(10): p. 713-23.
43. Patnaik, S.K., et al., *Complex N-glycans are the major ligands for galectin-1, -3, and -8 on Chinese hamster ovary cells*. Glycobiology, 2006. **16**(4): p. 305-317.
44. Carlsson, S., et al., *Affinity of galectin-8 and its carbohydrate recognition domains for ligands in solution and at the cell surface*. Glycobiology, 2007. **17**(6): p. 663-76.
45. Carlsson, S., M.C. Carlsson, and H. Leffler, *Intracellular sorting of galectin-8 based on carbohydrate fine specificity*. Glycobiology, 2007. **17**(9): p. 906-12.
46. Stowell, S.R., et al., *Innate immune lectins kill bacteria expressing blood group antigen*. Nat Med, 2010. **16**(3): p. 295-301.
47. Thurston, T.L., et al., *Galectin 8 targets damaged vesicles for autophagy to defend cells against bacterial invasion*. Nature, 2012. **482**(7385): p. 414-8.
48. Kim, B.W., et al., *Structural basis for recognition of autophagic receptor NDP52 by the sugar receptor galectin-8*. Nat Commun, 2013. **4**: p. 1613.
49. Li, S., et al., *Sterical hindrance promotes selectivity of the autophagy cargo receptor NDP52 for the danger receptor galectin-8 in antibacterial autophagy*. Sci Signal, 2013. **6**(261): p. ra9.

50. Tsai, C.M., et al., *Galectin-1 and galectin-8 have redundant roles in promoting plasma cell formation*. J Immunol, 2011. **187**(4): p. 1643-52.
51. Eshkar Sebban, L., et al., *The involvement of CD44 and its novel ligand galectin-8 in apoptotic regulation of autoimmune inflammation*. J Immunol, 2007. **179**(2): p. 1225-35.
52. Massardo, L., et al., *Autoantibodies against galectin-8: their specificity, association with lymphopenia in systemic lupus erythematosus and detection in rheumatoid arthritis and acute inflammation*. Lupus, 2009. **18**(6): p. 539-46.
53. Sarter, K., et al., *Autoantibodies against galectins are associated with antiphospholipid syndrome in patients with systemic lupus erythematosus*. Glycobiology, 2013. **23**(1): p. 12-22.
54. Carlsson, M.C., et al., *Galectin-8 in IgA nephritis: decreased binding of IgA by galectin-8 affinity chromatography and associated increased binding in non-IgA serum glycoproteins*. J Clin Immunol, 2012. **32**(2): p. 246-55.
55. Pál, Z., et al., *Non-synonymous single nucleotide polymorphisms in genes for immunoregulatory galectins: Association of galectin-8 (F19Y) occurrence with autoimmune diseases in a Caucasian population*. Biochimica et Biophysica Acta (BBA) - General Subjects, 2012. **1820**(10): p. 1512-1518.
56. Cattaneo, V., et al., *Galectin-8 elicits pro-inflammatory activities in the endothelium*. Glycobiology, 2014. **24**(10): p. 966-73.
57. Sugaya, S., et al., *Comparison of galectin expression signatures in rejected and accepted murine corneal allografts*. Cornea, 2015. **34**(6): p. 675-81.
58. Sampson, J.F., et al., *Galectin-8 Ameliorates Murine Autoimmune Ocular Pathology and Promotes a Regulatory T Cell Response*. PLoS One, 2015. **10**(6): p. e0130772.
59. Sampson, J.F., et al., *Galectin-8 promotes regulatory T cell differentiation by modulating IL-2 and TGF β signaling*. Immunol Cell Biol, 2016. **94**(2): p. 213-9.
60. Compagno, D., et al., *Glycans and galectins in prostate cancer biology, angiogenesis and metastasis*. Glycobiology, 2014. **24**(10): p. 899-906.
61. Bidon-Wagner, N. and J.P. Le Pennec, *Human galectin-8 isoforms and cancer*. Glycoconj J, 2004. **19**(7-9): p. 557-63.
62. Camby, I., et al., *Galectins are differentially expressed in supratentorial pilocytic astrocytomas, astrocytomas, anaplastic astrocytomas and glioblastomas, and significantly modulate tumor astrocyte migration*. Brain Pathol, 2001. **11**(1): p. 12-26.
63. Nagy, N., et al., *Galectin-8 expression decreases in cancer compared with normal and dysplastic human colon tissue and acts significantly on human colon cancer cell migration as a suppressor*. Gut, 2002. **50**(3): p. 392-401.
64. Henno, S., et al., *Expression of Po66-CBP, a galectin-8, in different types of primary and secondary broncho-pulmonary tumors*. Oncol Rep, 2002. **9**(1): p. 177-80.
65. Caulet-Maugendre, S., et al., *Immunohistochemical expression of the intracellular component of galectin-8 in squamous cell metaplasia of the bronchial epithelium in neoplastic and benign processes*. Pathol Res Pract, 2001. **197**(12): p. 797-801.
66. Lahm, H., et al., *Comprehensive galectin fingerprinting in a panel of 61 human tumor cell lines by RT-PCR and its implications for diagnostic and therapeutic procedures*. J Cancer Res Clin Oncol, 2001. **127**(6): p. 375-86.
67. Delgado, V.M., et al., *Modulation of endothelial cell migration and angiogenesis: a novel function for the "tandem-repeat" lectin galectin-8*. Faseb j, 2011. **25**(1): p. 242-54.
68. Cueni, L.N. and M. Detmar, *Galectin-8 interacts with podoplanin and modulates lymphatic endothelial cell functions*. Exp Cell Res, 2009. **315**(10): p. 1715-23.
69. Troncoso, M.F., et al., *Galectin-8: A matricellular lectin with key roles in angiogenesis*. Glycobiology, 2014. **24**(10): p. 907-914.
70. Nishi, N., et al., *Galectin-8 and galectin-9 are novel substrates for thrombin*. Glycobiology, 2006. **16**(11): p. 15C-20C.
71. Yoshida, H., et al., *Crystallization and preliminary X-ray diffraction analysis of a protease-resistant mutant form of human galectin-8*. Acta Crystallographica Section F, 2009. **65**(5): p. 512-514.
72. Sievers, F., et al., *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega*. Molecular Systems Biology, 2011. **7**: p. 539-539.
73. McWilliam, H., et al., *Analysis Tool Web Services from the EMBL-EBI*. Nucleic Acids Research, 2013. **41**(Web Server issue): p. W597-W600.

74. Waterhouse, A.M., et al., *Jalview Version 2--a multiple sequence alignment editor and analysis workbench*. Bioinformatics, 2009. **25**(9): p. 1189-91.
75. Yoshida, H., et al., *X-ray Structures of Human Galectin-9 C-terminal Domain in Complexes with a Biantennary Oligosaccharide and Sialyllactose*. Journal of Biological Chemistry, 2010. **285**(47): p. 36969-36976.
76. Ideo, H., et al., *Galectin-8-N-domain recognition mechanism for sialylated and sulfated glycans*. J Biol Chem, 2011. **286**(13): p. 11346-55.
77. Vokhmyanina, O.A., et al., *Carbohydrate specificity of chicken and human tandem-repeat-type galectins-8 in composition of cells*. Biochemistry (Mosc), 2011. **76**(10): p. 1185-92.
78. Vokhmyanina, O.A., et al., *Comparative study of the glycan specificities of cell-bound human tandem-repeat-type galectin-4, -8 and -9*. Glycobiology, 2012. **22**(9): p. 1207-17.
79. Newburg, D.S., G.M. Ruiz-Palacios, and A.L. Morrow, *Human milk glycans protect infants against enteric pathogens*. Annu Rev Nutr, 2005. **25**: p. 37-58.
80. Salomonsson, E., et al., *Mutational tuning of galectin-3 specificity and biological function*. J Biol Chem, 2010. **285**(45): p. 35079-91.
81. Bohari, M.H., et al., *Structure-based rationale for differential recognition of lacto- and neolacto-series glycosphingolipids by the N-terminal domain of human galectin-8*. Scientific Reports, 2016. **6**: p. 39556.
82. Wishart, D.S., et al., *DrugBank: a knowledgebase for drugs, drug actions and drug targets*. Nucleic Acids Res, 2008. **36**(Database issue): p. D901-6.
83. Hickey, M., J. Elliott, and S.L. Davison, *Hormone replacement therapy*. BMJ, 2012. **344**.
84. Zhu, L.L., et al., *Induction of a program gene expression during osteoblast differentiation with strontium ranelate*. Biochem Biophys Res Commun, 2007. **355**(2): p. 307-11.
85. Fonseca Jã, E. and M.L. Brandi, *Mechanism of action of strontium ranelate: what are the facts?* Clin Cases Miner Bone Metab, 2010. **7**(1): p. 17-8.
86. Bornstein, P. and E.H. Sage, *Matricellular proteins: extracellular modulators of cell function*. Current Opinion in Cell Biology, 2002. **14**(5): p. 608-616.
87. Stock, M., et al., *Expression of Galectin-3 in Skeletal Tissues Is Controlled by Runx2*. Journal of Biological Chemistry, 2003. **278**(19): p. 17360-17367.
88. TÛBEL, J., et al., *Expression of the Tumor Markers Sialyl Lewis A, Sialyl Lewis X, Lewis Y, Thomsen-Friedenreich Antigen, Galectin-1 and Galectin-3 in Human Osteoblasts In Vitro*. Anticancer Research, 2012. **32**(5): p. 2159-2164.
89. Ortega, N., et al., *Galectin-3 Is a Downstream Regulator of Matrix Metalloproteinase-9 Function during Endochondral Bone Formation*. Molecular Biology of the Cell, 2005. **16**(6): p. 3028-3039.
90. Tanikawa, R., et al., *Galectin-9 induces osteoblast differentiation through the CD44/Smad signaling pathway*. Biochemical and Biophysical Research Communications, 2010. **394**(2): p. 317-322.
91. Tanikawa, R., et al., *Interaction of galectin-9 with lipid rafts induces osteoblast proliferation through the c-Src/ERK signaling pathway*. J Bone Miner Res, 2008. **23**(2): p. 278-86.
92. Bhat, R., et al., *A regulatory network of two galectins mediates the earliest steps of avian limb skeletal morphogenesis*. BMC Dev Biol, 2011. **11**: p. 6.
93. Alford, A.I. and K.D. Hankenson, *Matricellular proteins: Extracellular modulators of bone development, remodeling, and regeneration*. Bone, 2006. **38**(6): p. 749-757.
94. Vinik, Y., S. Boura-halfon, and Y. Zick, *Galectin-8 as a novel lectin modulating cancer metastasis to bone*. Bone, 2010. **47**, **Supplement 3**: p. S415.
95. Friedel, M., et al., *Galectin-8 enhances adhesion of multiple myeloma cells to vascular endothelium and is an adverse prognostic factor*. Glycobiology, 2016.
96. Vinik, Y., et al., *The mammalian lectin galectin-8 induces RANKL expression, osteoclastogenesis, and bone mass reduction in mice*. Elife, 2015. **4**: p. e05914.
97. Nishi, N., et al., *Functional and structural bases of a cysteine-less mutant as a long-lasting substitute for galectin-1*. Glycobiology, 2008. **18**(12): p. 1065-73.
98. Si, Y., et al., *Human galectin-2 interacts with carbohydrates and peptides non-classically: new insight from X-ray crystallography and hemagglutination*. Acta Biochim Biophys Sin (Shanghai), 2016. **48**(10): p. 939-947.

99. Collins, P.M., K.I. Hidari, and H. Blanchard, *Slow diffusion of lactose out of galectin-3 crystals monitored by X-ray crystallography: possible implications for ligand-exchange protocols*. Acta Crystallogr D Biol Crystallogr, 2007. **63**(Pt 3): p. 415-9.
100. Bum-Erdene, K., et al., *Structural characterisation of human galectin-4 N-terminal carbohydrate recognition domain in complex with glycerol, lactose, 3'-sulfo-lactose, and 2'-fucosyllactose*. Sci Rep, 2016. **6**.
101. Bum-Erdene, K., et al., *Structural characterization of human galectin-4 C-terminal domain: elucidating the molecular basis for recognition of glycosphingolipids, sulfated saccharides and blood group antigens*. Febs j, 2015. **282**(17): p. 3348-67.
102. Leonidas, D.D., et al., *Structural basis for the recognition of carbohydrates by human galectin-7*. Biochemistry, 1998. **37**(40): p. 13930-40.
103. Nagae, M., et al., *Structural analysis of the human galectin-9 N-terminal carbohydrate recognition domain reveals unexpected properties that differ from the mouse orthologue*. J Mol Biol, 2008. **375**(1): p. 119-35.
104. Blanchard, H., K. Bum-Erdene, and M.W. Hugo, *Inhibitors of Galectins and Implications for Structure-Based Design of Galectin-Specific Therapeutics*. Australian Journal of Chemistry, 2014. **67**(12): p. 1763-1779.
105. Blanchard, H., et al., *Galectin-1 inhibitors and their potential therapeutic applications: a patent review*. Expert Opin Ther Pat, 2016. **26**(5): p. 537-54.
106. Blanchard, H., et al., *Galectin-3 inhibitors: a patent review (2008-present)*. Expert Opin Ther Pat, 2014. **24**(10): p. 1053-65.
107. Bum-Erdene, K., et al., *Investigation into the feasibility of thioditaloside as a novel scaffold for galectin-3-specific inhibitors*. Chembiochem, 2013. **14**(11): p. 1331-42.
108. Öberg, C.T., et al., *Protein subtype-targeting through ligand epimerization: Talose-selectivity of galectin-4 and galectin-8*. Bioorganic & Medicinal Chemistry Letters, 2008. **18**(13): p. 3691-3694.
109. Ideo, H., et al., *High-affinity binding of recombinant human galectin-4 to SO₃→3Galβ1→3GalNAc pyranoside*. Glycobiology, 2002. **12**(3): p. 199-208.
110. Collins, P.M., et al., *Galectin-3 Interactions with Glycosphingolipids*. Journal of Molecular Biology, 2014. **426**(7): p. 1439-1451.
111. Nagae, M., et al., *Structural analysis of the recognition mechanism of poly-N-acetyllactosamine by the human galectin-9 N-terminal carbohydrate recognition domain*. Glycobiology, 2009. **19**(2): p. 112-7.
112. Sörme, P., et al., *Structural and Thermodynamic Studies on Cation-Π Interactions in Lectin-Ligand Complexes: High-Affinity Galectin-3 Inhibitors through Fine-Tuning of an Arginine-Arene Interaction*. Journal of the American Chemical Society, 2005. **127**(6): p. 1737-1743.
113. Sorme, P., et al., *Low micromolar inhibitors of galectin-3 based on 3'-derivatization of N-acetyllactosamine*. Chembiochem, 2002. **3**(2-3): p. 183-9.
114. Salameh, B.A., H. Leffler, and U.J. Nilsson, *3-(1,2,3-Triazol-1-yl)-1-thio-galactosides as small, efficient, and hydrolytically stable inhibitors of galectin-3*. Bioorganic & Medicinal Chemistry Letters, 2005. **15**(14): p. 3344-3346.
115. Cumpstey, I., et al., *Synthesis of a phenyl thio-[small beta]-d-galactopyranoside library from 1,5-difluoro-2,4-dinitrobenzene: discovery of efficient and selective monosaccharide inhibitors of galectin-7*. Organic & Biomolecular Chemistry, 2005. **3**(10): p. 1922-1932.
116. van Hattum, H., et al., *Tuning the Preference of Thiodigalactoside- and Lactosamine-Based Ligands to Galectin-3 over Galectin-1*. Journal of Medicinal Chemistry, 2013. **56**(3): p. 1350-1354.
117. Delaine, T., et al., *Galectin-Inhibitory Thiodigalactoside Ester Derivatives Have Antimigratory Effects in Cultured Lung and Prostate Cancer Cells*. Journal of Medicinal Chemistry, 2008. **51**(24): p. 8109-8114.
118. Cumpstey, I., et al., *Double Affinity Amplification of Galectin-Ligand Interactions through Arginine-Arene Interactions: Synthetic, Thermodynamic, and Computational Studies with Aromatic Diamido Thiodigalactosides*. Chemistry – A European Journal, 2008. **14**(14): p. 4233-4245.
119. Kim, S., et al., *PubChem Substance and Compound databases*. Nucleic Acids Res, 2016. **44**(D1): p. D1202-13.

120. Gaulton, A., et al., *ChEMBL: a large-scale bioactivity database for drug discovery*. Nucleic Acids Res, 2012. **40**(Database issue): p. D1100-7.
121. Liu, T., et al., *BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities*. Nucleic Acids Res, 2007. **35**(Database issue): p. D198-201.
122. Benson, M.L., et al., *Binding MOAD, a high-quality protein-ligand database*. Nucleic Acids Research, 2008. **36**(suppl 1): p. D674-D678.
123. Wang, R., et al., *The PDBbind Database: Methodologies and Updates*. Journal of Medicinal Chemistry, 2005. **48**(12): p. 4111-4119.
124. Irwin, J.J. and B.K. Shoichet, *ZINC – A Free Database of Commercially Available Compounds for Virtual Screening*. J Chem Inf Model, 2005. **45**(1): p. 177-82.
125. Irwin, J.J., et al., *ZINC: A Free Tool to Discover Chemistry for Biology*. Journal of Chemical Information and Modeling, 2012. **52**(7): p. 1757-1768.
126. Teague, S.J., et al., *The Design of Leadlike Combinatorial Libraries*. Angew Chem Int Ed Engl, 1999. **38**(24): p. 3743-3748.
127. Lipinski, C.A., et al., *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings*. Adv Drug Deliv Rev, 2001. **46**(1-3): p. 3-26.
128. Lipinski, C.A., *Lead- and drug-like compounds: the rule-of-five revolution*. Drug Discovery Today: Technologies, 2004. **1**(4): p. 337-341.
129. Ashburn, T.T. and K.B. Thor, *Drug repositioning: identifying and developing new uses for existing drugs*. Nat Rev Drug Discov, 2004. **3**(8): p. 673-83.
130. Chong, C.R. and D.J. Sullivan, Jr., *New uses for old drugs*. Nature, 2007. **448**(7154): p. 645-6.
131. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis*. J Comput Chem, 2004. **25**(13): p. 1605-12.
132. Trott, O. and A.J. Olson, *AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading*. Journal of Computational Chemistry, 2010. **31**(2): p. 455-461.
133. Morris, G.M., et al., *AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility*. J Comput Chem, 2009. **30**(16): p. 2785-91.
134. Gasteiger, J. and M. Marsili, *A new model for calculating atomic charges in molecules*. Tetrahedron Letters, 1978. **19**(34): p. 3181-3184.
135. Forli, S., et al., *Computational protein-ligand docking and virtual drug screening with the AutoDock suite*. Nat. Protocols, 2016. **11**(5): p. 905-919.
136. Pronk, S., et al., *GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit*. Bioinformatics, 2013. **29**(7): p. 845-854.
137. Lindorff-Larsen, K., et al., *Improved side-chain torsion potentials for the Amber ff99SB protein force field*. Proteins, 2010. **78**(8): p. 1950-8.
138. Jorgensen, W.L., *Revised TIPS for simulations of liquid water and aqueous solutions*. The Journal of Chemical Physics, 1982. **77**(8): p. 4156-4163.
139. Sousa da Silva, A.W. and W.F. Vranken, *ACPYPE - AnteChamber PYthon Parser interfacE*. BMC Research Notes, 2012. **5**(1): p. 1-8.
140. Wang, J., et al., *Development and testing of a General Amber Force Field*. J Comput Chem, 2004. **25**.
141. Jakalian, A., D.B. Jack, and C.I. Bayly, *Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation*. J Comput Chem, 2002. **23**(16): p. 1623-41.
142. McPhillips, T.M., et al., *Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines*. J Synchrotron Radiat, 2002. **9**(Pt 6): p. 401-6.
143. Battye, T.G., et al., *iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM*. Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 271-81.
144. Evans, P.R. and G.N. Murshudov, *How good are my data and what is the resolution?* Acta Crystallogr D Biol Crystallogr, 2013. **69**(Pt 7): p. 1204-14.
145. McCoy, A.J., et al., *Phaser crystallographic software*. J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
146. Murshudov, G.N., et al., *REFMAC5 for the refinement of macromolecular crystal structures*. Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 355-67.

147. Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of macromolecular structures by the maximum-likelihood method*. Acta Crystallogr D Biol Crystallogr, 1997. **53**(Pt 3): p. 240-55.
148. *The CCP4 suite: programs for protein crystallography*. Acta Crystallogr D Biol Crystallogr, 1994. **50**(Pt 5): p. 760-3.
149. Emsley, P., et al., *Features and development of Coot*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
150. Bohari, M.H. and G.N. Sastry, *FDA approved drugs complexed to their targets: evaluating pose prediction accuracy of docking protocols*. J Mol Model, 2012. **18**(9): p. 4263-74.
151. Plewczynski, D., et al., *Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database*. J Comput Chem, 2011. **32**(4): p. 742-55.
152. Imberty, A., *Oligosaccharide structures: theory versus experiment*. Curr Opin Struct Biol, 1997. **7**(5): p. 617-23.
153. Agostino, M., et al., *Molecular Docking of Carbohydrate Ligands to Antibodies: Structural Validation against Crystal Structures*. Journal of Chemical Information and Modeling, 2009. **49**(12): p. 2749-2760.
154. Nivedha, A.K., et al., *Vina-Carb: Improving Glycosidic Angles during Carbohydrate Docking*. J Chem Theory Comput, 2016. **12**(2): p. 892-901.
155. Wang, J., et al., *Automatic atom type and bond type perception in molecular mechanical calculations*. J Mol Graphics Model, 2006. **25**.
156. D.A. Case, T.A.D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Götz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman, *AMBER 12*. University of California, San Francisco, 2012.
157. Ou-Yang, S.-s., et al., *Computational drug discovery*. Acta Pharmacol Sin, 2012. **33**(9): p. 1131-1140.
158. Talele, T.T., S.A. Khedkar, and A.C. Rigby, *Successful applications of computer aided drug discovery: moving drugs from concept to the clinic*. Curr Top Med Chem, 2010. **10**(1): p. 127-41.
159. Irwin, J.J. and B.K. Shoichet, *Docking Screens for Novel Ligands Conferring New Biology*. J Med Chem, 2016. **59**(9): p. 4103-20.
160. Saraboji, K., et al., *The Carbohydrate-Binding Site in Galectin-3 Is Preorganized To Recognize a Sugarlike Framework of Oxygens: Ultra-High-Resolution Structures and Water Dynamics*. Biochemistry, 2012. **51**(1): p. 296-306.
161. Mayo, K.H., et al., *Design of a partial peptide mimetic of anginex with antiangiogenic and anticancer activity*. J Biol Chem, 2003. **278**(46): p. 45746-52.
162. Dings, R.P., et al., *Design of nonpeptidic topomimetics of antiangiogenic proteins with antitumor activities*. J Natl Cancer Inst, 2006. **98**(13): p. 932-6.
163. Dings, R.P., et al., *Polycationic calixarene PTX013, a potent cytotoxic agent against tumors and drug resistant cancer*. Invest New Drugs, 2013. **31**(5): p. 1142-50.
164. Sampson, J.F., et al., *Galectin-8 promotes regulatory T-cell differentiation by modulating IL-2 and TGFbeta signaling*. Immunol Cell Biol, 2015.
165. Wong, C.-H., et al., *Small Molecules as Structural and Functional Mimics of Sialyl Lewis X Tetrasaccharide in Selectin Inhibition: A Remarkable Enhancement of Inhibition by Additional Negative Charge and/or Hydrophobic Group*. Journal of the American Chemical Society, 1997. **119**(35): p. 8152-8158.
166. Wang, H., et al., *Silver(I) Oxide Mediated Selective Monoprotection of Diols in Pyranosides*. The Journal of Organic Chemistry, 2004. **69**(17): p. 5774-5777.
167. Sato, K.-i., et al., *The first total synthesis of telephiose A*. Tetrahedron Letters, 2007. **48**(21): p. 3745-3748.
168. Lars Andersson, L.K., *Synthesis and NMR studies of methyl 3-O-[(R)- and (S)-1-carboxyethyl]-D-gluc-, galacto- and manno-pyranosides*. Carbohydrate Research, 1998. **313**: p. 157-164.
169. Hama, N., et al., *Total Synthesis of Broussonetine F: The Orthoamide Overman Rearrangement of an Allylic Diol*. Organic Letters, 2011. **13**(4): p. 616-619.

170. Dassault Systèmes BIOVIA, *Discovery Studio Modeling Environment, Release 2017*. San Diego: Dassault Systèmes, 2016.
171. Hess, B., et al., *GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation*. J Chem Theory Comput, 2008. **4**.
172. Essmann, U., et al., *A smooth particle mesh Ewald method*. The Journal of Chemical Physics, 1995. **103**(19): p. 8577-8593.
173. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual Molecular Dynamics*. J Mol Graphics, 1996. **14**.
174. Kollman, P.A., et al., *Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models*. Acc Chem Res, 2000. **33**(12): p. 889-97.
175. Miller, B.R., 3rd, et al., *MMPBSA.py: An Efficient Program for End-State Free Energy Calculations*. J Chem Theory Comput, 2012. **8**(9): p. 3314-21.
176. Curtin, D.Y. and R.J. Harder, *Stereochemistry and Reactions with Hydroxyl Ion and with Silver Oxide of the 2-Bromo-4-phenylcyclohexanols and the 1-Methyl-2-bromo-4-phenylcyclohexanols I*. Journal of the American Chemical Society, 1960. **82**(9): p. 2357-2368.
177. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode*. Methods Enzymol, 1997. **276**: p. 307-26.
178. Schuttelkopf, A.W. and D.M. van Aalten, *PRODRG: a tool for high-throughput crystallography of protein-ligand complexes*. Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 8): p. 1355-63.
179. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.
180. Goddard, T.D., C.C. Huang, and T.E. Ferrin, *Visualizing density maps with UCSF Chimera*. J Struct Biol, 2007. **157**(1): p. 281-7.
181. Adjei, A.A., *Pemetrexed (Alimta): a novel multitargeted antifolate agent*. Expert Rev Anticancer Ther, 2003. **3**(2): p. 145-56.
182. Rollins, K.D. and C. Lindley, *Pemetrexed: a multitargeted antifolate*. Clin Ther, 2005. **27**(9): p. 1343-82.
183. Vogelzang, N.J., et al., *Phase III study of pemetrexed in combination with cisplatin versus cisplatin alone in patients with malignant pleural mesothelioma*. J Clin Oncol, 2003. **21**(14): p. 2636-44.
184. Foloppe, N. and R. Hubbard, *Towards predictive ligand design with free-energy based computational methods?* Curr Med Chem, 2006. **13**(29): p. 3583-608.
185. Genheden, S. and U. Ryde, *The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities*. Expert Opin Drug Discov, 2015. **10**(5): p. 449-61.
186. Orozco, M. and F.J. Luque, *Theoretical Methods for the Description of the Solvent Effect in Biomolecular Systems*. Chemical Reviews, 2000. **100**(11): p. 4187-4226.
187. Srinivasan, J., et al., *Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices*. Journal of the American Chemical Society, 1998. **120**(37): p. 9401-9409.
188. Demizu, Y., et al., *Regioselective protection of sugars catalyzed by dimethyltin dichloride*. Org Lett, 2008. **10**(21): p. 5075-7.