

SATURATION TRANSFER DIFFERENCE NMR STUDIES OF PROTEIN-LIGAND INTERACTIONS

by

Monica G. Szczepina

B.Sc. (Honours), Chemistry, Simon Fraser University 2002
M.Sc., Chemistry, Simon Fraser University 2003

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

In the
Department of Chemistry

© Monica G. Szczepina 2011
SIMON FRASER UNIVERSITY
Spring 2011

All rights reserved. However, in accordance with the *Copyright Act of Canada*, this work may be reproduced, without authorization, under the conditions for *Fair Dealing*. Therefore, limited reproduction of this work for the purposes of private study, research, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.

Approval

Name: **Monica G. Szczepina**
Degree: **Doctor of Philosophy**
Title of Thesis: **Saturation Transfer Difference NMR Studies of Protein-Ligand Interactions**

Examining Committee:

Chair: **Dr. Paul C.H. Li**
Professor - Department of Chemistry

Dr. B. Mario Pinto
Senior Supervisor
Professor and Vice President-Research
Department of Chemistry

Dr. Erika Plettner
Internal Examiner
Associate Professor and Graduate Chair
Department of Chemistry

Dr. Gary Leach
Committee Member
Associate Professor - Department of Chemistry

Dr. Peter D. Wilson
Committee Member
Associate Professor - Department of Chemistry

Dr. Lawrence P. McIntosh
External Examiner
Professor - Department of Chemistry
University of British Columbia

Date Defended/Approved: April 18, 2011



SIMON FRASER UNIVERSITY
LIBRARY

Declaration of Partial Copyright Licence

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the "Institutional Repository" link of the SFU Library website <www.lib.sfu.ca> at: <<http://ir.lib.sfu.ca/handle/1892/112>>) and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author's written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

While licensing SFU to permit the above uses, the author retains copyright in the thesis, project or extended essays, including the right to change the work for subsequent purposes, including editing and publishing the work in whole or in part, and licensing other parties, as the author may desire.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, BC, Canada

Abstract

The mycolyl–arabinogalactan–peptidoglycan complex coats the surface of *Mycobacterium tuberculosis*. It is an impressive structure composed of approximately 30 galactofuranosyl (Gal f) residues attached via alternating β -(1 \rightarrow 6) and β -(1 \rightarrow 5) linkages synthesized by bifunctional galactofuranosyltransferases, GlfT1 and GlfT2. We have used Saturation Transfer Difference (STD) NMR spectroscopy to examine the active site architecture of GlfT2 using two trisaccharide acceptor substrates, β -D-Gal f -(1 \rightarrow 6)- β -D-Gal f -(1 \rightarrow 5)- β -D-Gal f -O(CH $_2$) $_7$ CH $_3$ and β -D-Gal f -(1 \rightarrow 5)- β -D-Gal f -(1 \rightarrow 6)- β -D-Gal f -O(CH $_2$) $_7$ CH $_3$. The STD NMR epitope maps demonstrated a greater enhancement toward the “reducing” ends of both trisaccharides, and that UDP-galactofuranose (UDP-Gal f) made more intimate contacts through its nucleotide moiety. This observation is consistent with the greater flexibility required within the active site of the reaction between the growing polymer acceptor and the UDP-Gal f donor. Competition STD NMR titration experiments with the trisaccharide acceptor substrates demonstrated that they bind competitively at the same site, suggesting that GlfT2 has one active site pocket capable of catalyzing both β -(1 \rightarrow 5) and β -(1 \rightarrow 6)-galactofuranosyl transfer reactions.

STD NMR spectroscopy was also used to probe the bioactive conformation of the carbohydrate mimic MDWNMHAA of the O-polysaccharide of

the *Shigella flexneri* Y bacterium when bound to its complementary antibody, mAb SYA/J6. The dynamic ligand epitope was mapped with the CORCEMA-ST (COmplete Relaxation and Conformational Exchange Matrix Analysis of Saturation Transfer) program that calculates STD-NMR intensities. Comparison of these predicted STD enhancements with experimental data was used to select a representative binding mode. The bound conformation was further refined with a simulated annealing refinement protocol known as STD-NMR Intensity-restrained CORCEMA Optimization (SICO) to give a more accurate representation of the bound peptide epitope.

X-ray crystallographic data of MDWNMHAA when bound to mAb SYA/J6 indicated the immobilization of water molecules, i.e. the presence of “bound” water molecules, in the combining site. Water Ligand Observed via Gradient Spectroscopy (WaterLOGSY) was used in conjunction with STD NMR spectroscopy to provide insight into the presence of water molecules that exist at the interstitial sites between the peptide and the antibody. Molecular dynamics calculations have also provided a more accurate picture of the possibilities for bound-ligand conformations, and water molecules involved in providing complementarity between the peptide and SYA-J6.

Keywords: STD-NMR spectroscopy; WaterLOGSY spectroscopy; NOE; Galactofuranosyl transferase; molecular dynamics; *Shigella flexneri* Y; *Mycobacterium tuberculosis*; peptide-carbohydrate mimicry

Dedication

This thesis is dedicated, with much love, to my mother Lidia Szczepina.

Acknowledgements

I am extremely grateful and thankful to my supervisor Dr. B. Mario Pinto for all his support and guidance throughout my PhD program. It was a privilege to work in his group. Thank-you Mario!

I would like to thank my supervisory committee members Dr. Peter D. Wilson and Dr. Gary Leach for all their support and encouragement throughout my PhD program. I would also like to thank my internal examiner Dr. Erika Plettner and my external examiner Dr. Lawrence P. McIntosh for all their helpful feedback. It was a pleasure to discuss science with all of you!

Thank-you to my collaborator at the University of Alberta, Dr. Todd L. Lowary. I loved working on the GfT2 project. Thank-you so much for the opportunity. Thank-you to Mr. Ruixiang B. Zheng for providing me with GfT2 and Dr. Gladys C. Completo for the synthesis of the trisaccharides.

Thank-you to my collaborators at Zymeworks Inc. especially Mr. Dustin Bleile and the late Dr. Johannes "Hannes" Müllegger. It was a pleasure to work with both of you. Hannes, you were an amazing supervisor! Dustin, you are the best rock-climbing partner in the world. Also, you are one of the smartest human beings I have ever met.

A very big thank-you to Dr. Andrew Lewis for all his NMR expertise and his willingness to drop everything to help me with my NMR experiments. I have learnt so much from you.

Thank-you to my Chem 481 supervisor, Dr. Eberhard Kiehlmann.

I would like to thank my lab mates. In particular, Dr. Yue Yuan for teaching me NMR! Thank-you to Dr. Silvia Borrelli for providing me with antibody. Thank-you also to Dr. Rehana Hossany for synthesizing the octapeptide. Thank-you ladies!

I am extremely thankful to my friends who have supported me along the way. In particular: Erica Louie, Adewale & Bukky Eniade, Kelly Lee, Natalie Nguyen and Tara Hampton. You are all the wind beneath my wings.

A very big thank-you to Charles “Czarles” Walsby. You rock my world.

I would like to thank my mother, Lidia Szczepina, for her love and for all the sacrifices she made for me. All that you endured brings tears to my eyes. Your courage and tenacity has moved mountains. I love you so much.

Table of Contents

Approval.....	ii
Abstract.....	iii
Dedication.....	v
Acknowledgements.....	vi
Table of Contents.....	viii
List of Figures.....	xiv
Abbreviations.....	xxx
CHAPTER 1: General Introduction.....	1
1.1 Carbohydrates.....	3
1.1.1 D-Galf containing glyconconjugates.....	4
1.1.2 <i>Mycobacterium tuberculosis</i>	4
1.1.3 Galactofuranosyl transferases.....	7
1.2 Peptide-Carbohydrate mimicry.....	11
1.2.1 Immune response against invading bacteria and viruses.....	11
1.2.2 Carbohydrate mimetic peptides as leads for vaccines that target the bacterial cell wall.....	13
1.3 NMR spectroscopy.....	14
1.3.1 The nuclear Overhauser effect (NOE).....	14
1.3.2 Saturation Transfer Difference NMR spectroscopy.....	21
1.3.3 1D and 2D STD-TOCSY NMR spectroscopy.....	27

1.3.4	Quantifying STD-NMR signals with CORCEMA-ST	30
1.4	Molecular dynamics	32
1.5	Thesis overview	35
1.6	References.....	37

CHAPTER 2: STD-NMR Studies of Two Acceptor Substrates of GlfT2, a Galactofuranosyltransferase from *Mycobacterium tuberculosis*: Epitope

Mapping Studies	44	
2.1	Keywords	45
2.2	Abstract.....	46
2.3	Introduction	47
2.4	Experimental.....	50
2.4.1	Materials.....	50
2.4.2	Preparation of GlfT2 for NMR studies	50
2.4.3	NMR spectroscopy	51
2.5	Results.....	54
2.5.1	Epitope mapping of donor substrate UDP-Galf.....	54
2.5.2	Epitope mapping of the acceptor substrate 2	55
2.5.3	Epitope mapping of the acceptor substrate 3	60
2.6	Discussion.....	62
2.7	Acknowledgements	64
2.8	Summary.....	65
2.9	References.....	66
2.10	Supporting Information.....	70

CHAPTER 3: STD-NMR Studies Suggest that Two Acceptor Substrates for GfT2, a Bifunctional Galactofuranosyltransferase Required for the Biosynthesis of *Mycobacterium tuberculosis* Arabinogalactan, Compete for the Same Binding Site 85

3.1 Keywords 86

3.2 Abstract..... 87

3.3 Introduction 88

3.4 Experimental 93

 3.4.1 Materials..... 93

 3.4.2 NMR spectroscopy 93

3.5 Results 97

 3.5.1 STD NMR competition studies between **2** and **3**..... 97

3.6 Discussion..... 109

3.7 Acknowledgements 111

3.8 Summary..... 112

3.9 References..... 113

3.10 Supporting Information 120

CHAPTER 4: WaterLOGSY NMR Experiments Detect Immobilized Water Molecules that Bridge Peptide Mimic MDWNMHAA to Anti-carbohydrate Antibody SYA/J6 126

4.1 Keywords 127

4.2 Abstract..... 128

4.3 Introduction 129

4.4	Experimental.....	135
4.4.1	Materials.....	135
4.4.2	NMR spectroscopy of peptide 1	135
4.4.3	Molecular dynamics simulations	136
4.5	Results and Discussion.....	137
4.5.1	Molecular dynamics studies of peptide 1 in the Fv portion of monoclonal antibody SYA/J6.....	137
4.5.2	WaterLOGSY experiments of peptide 1 in the combining site of monoclonal antibody SYA/J6.....	139
4.5.3	Analysis of WaterLOGSY and STD-NMR experiments of peptide 1 in the combining site of monoclonal antibody SYA/J6.....	146
4.6	Conclusions	151
4.7	Acknowledgements.....	151
4.8	Summary.....	152
4.9	References.....	153
4.10	Supporting Information.....	157

CHAPTER 5: Investigation of the Binding of a Carbohydrate-mimetic

Peptide to its Complementary Anti-carbohydrate Antibody by STD-NMR

Intensity-restrained CORCEMA Optimization (SICO) and Molecular

Dynamics Simulations..... 160

5.1	Keywords	161
5.2	Abstract.....	162
5.3	Introduction	163

5.4	Experimental.....	169
5.4.1	Materials.....	169
5.4.2	NMR spectroscopy of peptide 1	169
5.4.3	NMR spectroscopy of pentasaccharide 2	170
5.4.4	CORCEMA-ST calculations.....	172
5.4.5	Molecular dynamics simulations.....	174
5.5	Results and Discussion.....	175
5.5.1	Probing the secondary structure of peptide 1 free in solution.....	175
5.5.2	Binding of peptide 1 to monoclonal antibody SYA/J6 by Saturation Transfer Difference NMR Spectroscopy.....	177
5.5.3	Quantifying the binding of peptide 1 to monoclonal antibody SYA/J6 with CORCEMA-ST Calculations.....	180
5.5.4	Molecular dynamics studies of peptide 1 in the Fv portion of monoclonal antibody SYA/J6.....	182
5.5.5	Simulated annealing refinement of the binding mode of peptide 1 in the Fab of monoclonal antibody SYA/J6 coupled with CORCEMA-ST optimization.....	187
5.5.6	Binding of the pentasaccharide hapten 2 to monoclonal antibody SYA/J6 by saturation transfer difference NMR spectroscopy.....	191
5.6	Conclusions.....	195
5.7	Acknowledgements.....	196
5.8	Summary.....	197
5.9	References.....	198
5.10	Supporting Information.....	203

CHAPTER 6: Conclusions	224
6.1 Thesis summary.....	224
6.2 Appendix - Computational design of peptide mutants.....	227
6.3 References.....	235

List of Figures

- Figure 1.1.** Scanning Electron Micrograph of *Mycobacterium tuberculosis*. (Photo credit: Janice Carr, Content Providers(s): CDC/Dr. Ray Butler; Janice Carr, http://en.wikipedia.org/wiki/File:Mycobacterium_tuberculosis.jpg)5
- Figure 1.2.** Pictorial representation of Mycobacterial mycolyl-arabinogalactan-peptidoglycan (mAGP) complex. Modified from a figure published by Y tambe, http://en.wikipedia.org/wiki/File:Mycobacterial_cell_wall_diagram.png.6
- Figure 1.3.** Putative transition state for a glycosyltransferase-mediated reaction. (Reproduced with permission from *Org. Biomol. Chem.* **2004**, 2, 2418-2420. Copyright © 2004 Royal Chemical Society. All rights reserved).....8
- Figure 1.4.** Crystal structure of SpsA (PDB: 1QGS): A GT-A glycosyltransferase involved in the synthesis of the spore coat of *Bacillus subtilis*.9
- Figure 1.5.** Proposed mechanism for an inverting glycosyl transferase. (Reproduced with permission from *Annu. Rev. Biochem.* **2008**, 77, 521–555. Copyright © 2008 by Annual Reviews. All rights reserved). 10
- Figure 1.6.** B cell activation can be either helper T-cell independent (a) or helper T-cell dependent (b). (Reproduced with permission from Charles A Janeway, Jr, Paul Travers, Mark Walport, and Mark J Shlomchik.

Immunobiology, 5th edition, Garland Science, New York, NY.

Copyright © 2001 by Garland Science. All rights reserved)..... 12

Figure 1.7. American physicist Albert W. Overhauser born August 17, 1925 in San Diego, California. Photo credit: AIP Emilio Segre Visual Archives..... 15

Figure 1.8. 2011 Ford Mustang. <http://www.ford.com/cars/mustang/gallery/photos/> 16

Figure 1.9. 2D NOESY spectrum of 2,3,4-tri-*O*-benzyl-1,5-dideoxy-1,5-[[[(2*R*,3*R*)-2,4-*O*-benzylidene-2,4-dihydroxy-3-(sulfooxy)butyl]-(*R*)-episelenoniumylidene]xylitol inner salt..... 19

Figure 1.10. A two spin energy level diagram. Spins I and S share a dipolar coupling. Modified from a figure published by T. D. W Claridge, *High-resolution NMR techniques in organic chemistry*, Pergamon, Amsterdam, New York, **1999**. 20

Figure 1.11. The STD-NMR experiment. The protein signal is selectively saturated, typically with a soft pulse, in a region devoid of ligand resonances. Magnetization spreads throughout the protein via spin-diffusion (^1H - ^1H intramolecular cross relaxation). This magnetization then travels to the bound ligand via ^1H - ^1H intermolecular cross relaxation. The “hot” ligand departs the binding site and takes with it information about its bound state. Modified from a figure by Margaret A. Johnson, PhD thesis..... 23

Figure 1.12. The STD-NMR pulse sequence. (Reproduced with permission Bruker Biospin pulse program catalogue, TOPSPIN v2.0. NMR Guide. Copyright © 2006 by Bruker Biospin GmbH. All rights reserved) 24

Figure 1.13 a) STD-NMR spectrum of a sample containing BSA, tryptophan and glucose. b) ¹ H NMR spectrum of tryptophan and c) ¹ H NMR spectrum of glucose. Only tryptophan gives STD-NMR signals.	25
Figure 1.14. 2D TOCSY experiment for the pentasaccharide (shown in Figure 5.10) utilizing the mlevsgpph Bruker pulse sequence at 600 MHz and 298 K. Each hexose has a distinct spin system; thus, 5 spin systems are observed.	29
Figure 1.15. The MD algorithm. (Reproduced with permission from GROMACS user manual version 4.5. Copyright © 2001-2010 by The GROMACS development teams at the Royal Institute of Technology and Uppsala University, Sweden. All rights reserved)	34
Figure 2.1. Structure of the mAGP complex, with the galactan region highlighted; x ~10. *The three arabinan chains have been proposed to be linked via the eighth, tenth and twelfth Galf residues of the galactan core. ^[6]	48
Figure 2.2. Trisaccharides 2 and 3 , reported acceptor substrates for GlfT2.	49
Figure 2.3. Expansion of 1D ¹ H NMR (lower trace) and STD NMR (upper trace) spectra of UDP-Galf in the presence of GlfT2, at 600 MHz and 285K. Labels g, r and u refer to Galf, ribose, and uracil, respectively.	54
Figure 2.4. Epitope mapping of UDP-Galf in the active site of GlfT2. Enhancements are referenced to the H-1 resonance of the ribofuranosyl moiety.	55
Figure 2.5. Expansion of (A) 1D ¹ H NMR and (B) STD NMR spectra of 2 at 600 MHz and 298K in the presence of GlfT2.	56

Figure 2.6. Epitope mapping of 2 in the active site of GlfT2. Enhancements are referenced to the methyl resonance of the octyl aglycon. Estimated enhancements from STD-2D-TOCSY NMR data are shown in blue.....	57
Figure 2.7. Epitope mapping of 3 in the active site of GlfT2. Enhancements are referenced to the methyl resonance of the octyl aglycon. Estimated enhancements from STD-2D-TOCSY NMR data are shown in blue.....	60
Figure 3.1. 1D-STD NMR spectra at different ratios of the acceptor substrates 2 to 3 . The upper trace corresponds to the acceptor substrate 2 in the presence of GlfT2. Lower traces indicate spectra obtained with titration of increasing amounts of the acceptor substrate 3 ; the 2:3 ratio is indicated with each trace.	98
Figure 3.2. 1D-STDD NMR spectra at different ratios of the acceptor substrates 2 to 3 . The upper trace, corresponding to a 1D-STD NMR spectrum of the acceptor substrate 2 in the presence of GlfT2, was subtracted from the traces in Figure 3.1.	99
Figure 3.3. Titration of a sample of trisaccharide, 2 (4 mM) containing GlfT2 (51 μ M) with trisaccharide 3 . STD amplification factors were obtained from 1D-STD NMR spectra. The STD amplification factor for resonances H-1A and H-1C from 2 decreased whereas those for resonances H-1B and H-1A from 3 increased.	101
Figure 3.4. STD amplification factors calculated from six cross peak intensities originating from 2D-STD-TOCSY NMR experiments. Trisaccharide 2 to trisaccharide 3 ratios of 1:1, 1:2 and 1:4 correspond to trisaccharide 3 concentrations of 3.2 mM, 5.8 mM and 11.3 mM. These six cross	

peaks comprised the F1 traces from the H-1A, H-1B, and H-1C resonances of acceptor **2** as well as F1 traces from the H-1A, H-1B and H-1C resonances of the acceptor **3**. STD amplification factors for trisaccharide **2** decreased while STD amplification factors for trisaccharide **3** increased as a function of increasing amounts of trisaccharide **3**..... 102

Figure 3.5. 1D-STD NMR spectra at different ratios of the acceptor substrates **3** to **2**. The upper trace corresponds to the acceptor substrate **3** in the presence of GlfT2. Lower traces indicate spectra obtained with titration of increasing amounts of the acceptor substrate **2**; the **3:2** ratio is indicated with each trace. 105

Figure 3.6. Titration of trisaccharide, **3** (4 mM) containing GlfT2 (74 μM) with trisaccharide **2**. STD amplification factors were obtained from 1D-STD NMR spectra. The STD amplification factor for resonances H-1A and H-1B from **3** decreased whereas that for resonance H-1C and H-1A from **2** increased. 106

Figure 3.7. STD amplification factors calculated from cross peak intensities originating from 2D-STD-TOCSY NMR experiments. Trisaccharide **3** to trisaccharide **2** ratios of 1:1, 1:2 and 1:6, correspond to trisaccharide **2** concentrations of 3.9 mM, 7.4 mM and 20.0 mM. These four cross peaks comprised the F1 traces from the H-1A and H-1C resonances of acceptor **2** as well as F1 traces from the H-1A and H-1B resonances of the acceptor **3**. STD amplification factors for trisaccharide **3** decreased while STD amplification factors for trisaccharide **2** increased as a function of increasing amounts of **2**..... 108

Figure 4.1. Structure of the <i>Shigella flexneri</i> Y O-antigen polysaccharide.....	129
Figure 4.2. Structure of the carbohydrate-mimetic peptide MDWNMHAA 1 (peptide 1).....	130
Figure 4.3. X-ray crystal structures of a Fab fragment of the monoclonal antibody SYA/J6 (IgG3, κ), directed against the O-antigen of <i>Shigella flexneri</i> Y, complexed with a) peptide 1 (green), b) pentasaccharide hapten [- α -L- Rhap-(1 \rightarrow 2)- α -L-Rhap-(1 \rightarrow 3)- α -L-Rhap-(1 \rightarrow 3)- β -D-GlcpNAc-(1 \rightarrow 2)- α -L-Rhap-(1 \rightarrow OMe)] (yellow), and c) superposition of the two ligands in the combining site. H and L refer to heavy and light chains, respectively.....	134
Figure 4.4. Resident water molecules identified from MD simulation of peptide 1 bound to Fv. The x-axis refers to time (one frame = 0.0075ns. 1400 total frames = 10.5 ns trajectory) and the y-axis is arbitrary. These water molecules were mapped onto the X-Ray crystal structure (shown on the right). The water molecules found during the MD simulation were present for at least 5 ns (shown on the left) and those that were present for the entire trajectory (10.5 ns) were mapped onto the X-Ray crystal structure and correspond to water molecules 1, 2, 6 and 9 (shown on the right).	138
Figure 4.5. 1D WaterLOGSY NMR spectra of peptide 1 only (upper trace) and peptide 1 in the presence of mAb SYA/J6 (lower trace), at 600 MHz and 298 K.	144

Figure 4.6. 1D WaterLOGSY NMR spectrum of peptide 1 in the presence of mAb SYA/J6, at 600 MHz and 298 K, corrected for ligand-only effects by the subtraction of the upper trace from the lower trace in Figure 4.5.	145
Figure 4.7. Expansion of the 1D WaterLOGSY NMR spectrum of peptide 1 in the presence of mAb SYA/J6, at 600 MHz and 298 K, corrected for ligand-only effects (upper trace), and expansion of the 1D STD-NMR spectrum of the same sample under identical conditions (lower trace).....	149
Figure 4.8. Mapping of the water molecules 1, 2, 6, 8, and 9, identified from combined WaterLOGSY and STD-NMR experiments and molecular dynamics simulations, onto the X-Ray structure of peptide 1 , showing the proposed water lattice that mediates contacts between the peptide and the antibody.	150
Figure 5.1. Structure of the <i>Shigella flexneri</i> Y O-antigen polysaccharide.....	164
Figure 5.2. Structure of the carbohydrate-mimetic peptide MDWNMHAA 1 (peptide 1).....	165
Figure 5.3. X-ray crystal structure of a Fab fragment of the murine monoclonal antibody SYA/J6 (IgG3, κ), directed against the polysaccharide O-antigen of the <i>Shigella flexneri</i> variant Y lipopolysaccharide, complexed with a) peptide 1 (green), b) the pentasaccharide hapten [α -L-Rhap-(1 \rightarrow 2)- α -L-Rhap-(1 \rightarrow 3)- α -L-Rhap-(1 \rightarrow 3)- β -D-GlcpNAc-(1 \rightarrow 2)- α -L-Rhap-(1 \rightarrow OMe) 2 (yellow).	167
Figure 5.4. Comparison of experimental (blue), predicted by the CORCEMA-ST protocol based on the crystal structure of the peptide 1 -SYA/J6 Fab complex (pink), simulated annealing-refined (green) STD values from	

the CORCEMA-ST protocol, and CORCEMA-ST values for the minimized-SICO structure (fluorescent yellow) for peptide **1** in the presence of mAb SYA/J6. Enhancements are shown for a saturation time of 5s. 181

Figure 5.5. a) R-factor as a function of MD simulation time (one frame = 0.0075ns. 1400 total frames = 10.5 ns trajectory) for peptide **1** bound to Fv. The R-factor is represented in pink. The effective R-factor, R_{eff} , is represented in blue. b) The R-factor is broken down per proton. Thus, STD (fit) is shown per proton, averaged over the MD simulation. 186

Figure 5.6. Torsion angles of peptide **1** optimized to reflect the bound conformation in the solution state. 188

Figure 5.7. Alignment of X-ray structure (green) and SICO structure (pink) showing key contacts. The alignment used Pymol's "align" function which performs a sequence alignment (to capture the antibody in addition to the peptide), followed by a structural alignment. 190

Figure 5.8. Alignment of SICO structure (pink) and minimized SICO structure (blue) showing key contacts. The alignment used Pymol's "align" function which performs a sequence alignment (to capture the antibody in addition to the peptide), followed by a structural alignment. 190

Figure 5.9. a) STD NMR spectrum of the octapeptide **1** in the presence of mAb SYA/J6, b) ^1H NMR spectrum of the pentasaccharide **2**, c) STD NMR spectrum of the pentasaccharide **2** in the presence of SYA/J6 at 298K, with a spin-lock, after 4K scans shown with 40 times magnification, and d) STD NMR spectrum of the pentasaccharide **2** in the presence

of SYA/J6, at 298K with water suppression and a spin-lock, after 20K scans. 193

Figure 5.10. Epitope map of pentasaccharide bound to mAb SYA/J6 derived from the STD-NMR spectrum at 298K, recorded with 20K scans and a saturation time of 2s. STD enhancements were measured relative to the signals in an unsaturated reference spectrum i.e $(I_0 - I_{sat})/ I_0$ to correct for a single resonance that may encompass three protons (e.g. OMe or NHAc) or two protons (e.g. methylene CH₂). The methyl signal of the major NHAc isomer is at 2.06 ppm and that of the minor NHAc isomer is at 1.91 ppm. The minor isomer is most enhanced (100%); all enhancements were measured relative to this signal. 194

Figure 6.1. Bound water molecules (red spheres) are harnessed by the wild type peptide (green) to provide contacts to antibody's heavy chain Glu-35, Glu-50, Ala-106 and Met-107 residues. These contacts to the antibody are made by the pentasaccharide hapten α -L-Rhap-(1→2)- α -L-Rhap-(1→3)- α -L-Rhap-(1→3)- β -D-GlcpNAc-(1→2)- α -L-Rhap-(1→OMe). 231

Figure 6.2. Double mutant, MAW(d-TYR)MHAA, makes contacts to the antibody's heavy chain Glu-35, Glu-50, Ala-106 and Met-107 residues and displaces the three water molecules harnessed by the wild type peptide (green) to provide shape complementarity to the antibody. 232

Figure 6.3. Poisson-Boltzmann equation. (Reproduced with permission from Nathan Baker, Computational electrostatics for biomolecular systems. Workshop Lecture for *Collaborative Computational Project for Biomolecular Simulation (CCPB)*. Copyright © Nathan Baker, Washington University in St. Louis, Dept. of Biochemistry and

Molecular Biophysics, Center for Computational Biology,
baker@ccb.wustl.edu. All rights reserved).233

List of supporting information figures

- Figure S2.1.** Expansion of 1D ^1H NMR (upper trace) and STD NMR (lower trace) spectra of a mixture of UDP, UMP, galactose-1-phosphate and galactose following hydrolysis of UDP-Galf at 600 MHz and 285 K in the presence of GIfT2..... 78
- Figure S2.2.** Expansion of 1D ^1H NMR (lower and middle trace) and STD NMR (upper trace) spectra of the time course of the UDP-Galf hydrolysis reaction at 600 MHz and 285K in the presence of GIfT2. The bottom trace is the 1D ^1H NMR spectrum of UDP-Galf. The middle trace is the 1D ^1H NMR spectrum of the sample after 7 h. The upper trace is the STD NMR spectrum of the sample after 7h. UDP and UMP bind preferentially over UDP-Galf. 79
- Figure S2.3.** Expansion of 1D ^1H NMR spectrum of a mixture of trisaccharide **2** and UDP-Galf at 600 MHz and 298 K in the presence of GIfT2 over the course of 24 h. The asterisk denotes an anomeric signal of the formed β -(1 \rightarrow 5), β -(1 \rightarrow 6), β -(1 \rightarrow 5) tetrasaccharide (*J. Org. Chem.* **2008**, *73*, 4513–4525)..... 80
- Figure S2.4.** Expansion of (A) 1D ^1H NMR and (B) STD NMR spectra of **3** at 600 MHz and 298 K in the presence of GIfT2..... 81
- Figure S2.5.** Expansion of 1D ^1H NMR spectrum of a mixture of trisaccharide **3** and UDP-Galf at 600 MHz and 310 K in the presence of GIfT2 over the

course of 24h. The asterisks denote anomeric signals of the formed β -(1 \rightarrow 6), β -(1 \rightarrow 5), β -(1 \rightarrow 6) tetrasaccharide (*J. Org. Chem.* **2008**, 73, 4513–4525.)..... 82

Figure S2.6. Deconvolution of STD-NMR spectrum (3.2-4.2 ppm region) of **2** in the presence of GlfT2 at 298 K. Observed spectrum (black), calculated spectrum (red), individual peaks from fit (green), and difference between calculated and observed spectrum (blue)..... 83

Figure S2.7. STD-2D-TOCSY intensity relative to the 2D-TOCSY intensity for the selectively excited resonance H-6B of **2** in the presence of GlfT2. Row number 190 was extracted from the STD-2D-TOCSY spectrum (upper trace) and corresponds to H-6B and its spin-coupled neighbours, H-6'B, H-5B, H-4B and H-3B. The same row, row number 190, was extracted from the 2D-TOCSY spectrum (lower trace) and also corresponds to H-6B and its spin-coupled neighbours, H-6'B, H-5B, H-4B and H-3B. The STD enhancement of the STD-2D-TOCSY trace (upper trace) was measured relative to the 2D-TOCSY trace (lower trace) and was normalized to the selectively excited resonance, H-6B. The enhancement was quantified in Table S2.4 (second table from the top). 84

Figure S3.1. Expansion of 1D ^1H NMR spectrum of a mixture of trisaccharide **2** and UDP-Galf at 600 MHz and 298 K in the presence of GlfT2 over the course of 24 h. The asterisk denotes an anomeric signal of the formed β -(1 \rightarrow 5), β -(1 \rightarrow 6), β -(1 \rightarrow 5) tetrasaccharide (*J. Org. Chem.* **2008**, 73, 4513–4525.)..... 122

Figure S3.2. Expansion of 1D ¹ H NMR spectrum of a mixture of trisaccharide 3 and UDP-Galf at 600 MHz and 310 K in the presence of GlfT2 over the course of 24h. The asterisks denote anomeric signals of the formed β-(1→6), β-(1→5), β-(1→6) tetrasaccharide (<i>J. Org. Chem.</i> 2008 , 73, 4513–4525.).....	123
Figure S3.3. 1D-STDD NMR spectra at different ratios of the acceptor substrates 3 to 2 . The upper trace corresponds to the acceptor substrate 3 in the presence of GlfT2. Lower traces indicate spectra obtained with titration of increasing amounts of the acceptor substrate 2	124
Figure S3.4. <i>K_D</i> calculations using the Cheng-Prusoff equation. Titration curves were fitted using Excel by a modified version of eq. (1) using the STD amplification factor:	125
Figure S4.1. 1D WaterLOGSY NMR spectrum of peptide 1 in the presence of mAb SYA/J6 at 600 MHz and 298 K corrected for ligand only effects (upper trace) and 1D STD-NMR spectrum of the same sample under identical conditions (lower trace).	159
Figure S5.1. a) Number of hydrogen bonds existing within free MDWNMHAA as a function of simulation time. b) RMSD of free MDWNMHAA as a function of simulation time.....	209
Figure S5.2. Ramachandran plots of free MDWNMHAA during the course of the MD simulation. Each panel refers to phi and psi angles of a particular residue. Panels from a) to f) refer to aspartic acid, tryptophan, asparagine, methionine, histidine, and alanine, respectively.	210

- Figure S5.3.** 2D RMSD plot of free MDWNMHAA MD trajectories. RMSD is calculated at a time point and compared to every other time point. Yellow regions indicate low RMSD values ($< 5\text{\AA}$).211
- Figure S5.4.** Expansion of 1D ^1H NMR (upper trace) and STD-NMR (lower trace) spectra of peptide **1** at 600 MHz and 282 K in the presence of mAb SYA/J6.....212
- Figure S5.5.** Expansion of 1D ^1H NMR (upper trace) and STD-NMR (lower trace) spectra of peptide **1** at 600 MHz and 282 K in the presence of mAb SYA/J6.....213
- Figure S5.6.** STD build-up curves for the protons in MDWNMHAA. The maximal STD intensity, STD_{max} , and the observed saturation rate constant k_{sat} were obtained from fitting the saturation time data to the monoexponential equation: $\text{STD} = \text{STD}_{\text{max}} (1 - e^{-k_{\text{sat}}t}) + c$, as described by Mayer *et al.* The point (0,0) was not used in the fit.214
- Figure S5.7.** a) Number of hydrogen bonds existing between MDWNMHAA and SYA/J6 Fab as a function of simulation time, with MDWNMHAA acting as a hydrogen bond donor (red) and as a hydrogen bond acceptor (blue). The number of hydrogen bonds on the vertical axis refers to a cumulative average over the MD trajectory. b) RMSD of SYA/J6 Fab (black) MDWNMHAA (red) as a function of simulation time.....217
- Figure S5.8.** a) Number of hydrogen bonds existing between MDWNMHAA and SYA/J6 Fv as a function of simulation time, with MDWNMHAA acting as a hydrogen bond donor (red) and as a hydrogen bond acceptor (blue). The number of hydrogen bonds on the vertical axis refers to a

cumulative average over the MD trajectory. b) RMSD of SYA/J6 Fv (black) MDWNMHAA (red) as a function of simulation time.....218

Figure S5.9. a) RMSD per residue for the light chain of Fab (black) bound to MDWNMHAA and the light chain of Fv (blue) bound to MDWNMHAA. b) RMSD per residue for the heavy chain of Fab (black) bound to MDWNMHAA and the heavy chain of Fv (red) bound to MDWNMHAA.219

Figure S5.10. Ramachandran plots of MDWNMHAA bound to Fv during the course of the MD simulation. There are α -helical structural elements in the MHAA region upon binding. Each panel refers to phi and psi angles of a particular residue. Panels from a) to f) refer to aspartic acid, tryptophan, asparagine, methionine, histidine, and alanine, respectively.....220

Figure S5.11. Comparison of B values from the MD simulation of MDWNMHAA bound to Fv with B values from the X-ray crystal structure. a) The light chain is shown in blue and b) the heavy chain is shown in red. B values from the X-ray crystal structure are shown in black.....221

Figure S5.12. a) RMSD per residue for the light chain of Fv bound to MDWNMHAA (blue) and in the absence of MDWNMHAA (black). b) RMSD per residue for the heavy chain of Fv bound to MDWNMHAA (red) and in the absence of MDWNMHAA (black).222

Figure S5.13. Radius of gyration and van der Waals contacts. Upper panels show the radius of gyration of MDWNMHAA and Fv, left and right, respectively, during the course of the MD simulation. Lower panel

shows that van der Waals contacts are maintained throughout the
simulation.....223

Abbreviations


APBS	Adaptive Poisson-Boltzmann Solver
APC	Antigen Presenting Cell
Ala	Alanine
Asn	Asparagine
Asp	Aspartic acid
Ax	Axial
B cell	Bone marrow cell
BSA	Bovine serum albumin
CORCEMA-ST	Complete relaxation and conformational exchange matrix analysis of saturation transfer
COSY	Correlated spectroscopy
CSI	Chemical shift index
Da	Dalton
DEE	Dead end elimination
Eq	Equatorial
Fab	Antigen binding fragment
fs	Femto seconds
Galf	D-Galactofuranose
D-Galf	D-Galactofuranose

Glc _p	Glucopyranose
Glc _p NAc	N-Acetyl glucopyranose
GlfT1	Galactofuranosyl transferase 1
GlfT2	Galactofuranosyl transferase 2
GPI	Glycosylphosphatidylinositol
GT	Glycosyl transferase
His	Histidine
IgG	Immunoglobulin G
LPG	Lipophosphoglycan
LPPG	Lipopeptidophosphoglycan
LPS	Lipopolysaccharide
mAb	monoclonal antibody
mAGP	mycoyl-arabinogalactan-peptidoglycan complex
MD	Molecular dynamics
Met	Methionine
mM	milli Molar
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
NPT	Isothermal-Isobaric ensemble
ns	nanoseconds
PDB	Protein data bank

pPGM	peptidophosphogalactomannan
r	Ribosyl
Rha	Rhamnopyranose
Rhap	Rhamnopyranose
RMSD	Root mean square deviation
SA	Simulated annealing
SALMON	Solvent accessibility, ligand binding, and mapping of ligand orientation by NMR spectroscopy
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SICO	STD-NMR intensity-restrained CORCEMA optimization
STD NMR	Saturation transfer difference nuclear magnetic resonance
STDD NMR	Saturation Transfer Double Difference Nuclear Magnetic Resonance
τ_c	correlation time
T cell	Thymus cell
TB	Tuberculosis
TOCSY	Total correlation spectroscopy
Trp	Tryptophan
U	Uracil
μM	Micromolar

UDP	Uridine 5'-diphosphate
UDP-Galf	Uridine 5'-diphosphate galactofuranose
UMP	Uridine 5'-monophosphate
uv	Ultraviolet
WaterLOGSY	Water-ligand observed via gradient spectroscopy
WHO	World Health Organization
XDR-TB	Extensively drug resistant TB

Critical note: Never delete the section break below !!!!! This break enables the differential page numbering between the preliminary Roman numeral section and the main body of your document, in Arabic numbering.

If you cannot see the section break line, turn on the "Show/Hide button"  on your menu bar.
Do NOT delete this reminder until just before printing final document.
DO delete these paragraphs of instructional text from your final document.

CHAPTER 1: General Introduction

This thesis examines, at the molecular level, the interaction of small molecules bound to protein receptors. Covalent interactions, which involve atoms sharing electrons, and non-covalent interactions, such as hydrogen bonding and van der Waals, which do not involve the sharing of electrons, form the basis of these bonds. Human diseases such as tuberculosis and shigellosis (bacillary dysentery) are caused when carbohydrate coated bacteria enter the body and are recognized and bound by carbohydrate receptors on host cells.

The first target of this thesis work is GlfT2, an enzyme involved in the biosynthesis of the cell wall of tuberculosis-causing *Mycobacterium tuberculosis*. The second target of this thesis work is monoclonal antibody SYA/J6, an antibody specific for the O-polysaccharide on the surface of shigellosis-causing *Shigella flexneri* Y. Our hypothesis is that by understanding the molecular interactions of these two targets with their ligands we can contribute to the design of vaccines and treatments against these diseases.

In the case of GlfT2, we have shown that the enzyme employs one active site when processing its substrates and building the D-Galf containing cell wall of *M. tuberculosis*. The work presented in this thesis has contributed to the current model of the mechanism of bifunctional galactofuranosyltransferase GlfT2, and could facilitate the design and synthesis of an inhibitor against the enzyme.

We also examined *Shigella flexneri* Y antibody SYA-J6, and found that the

antibody interacts with a carbohydrate mimetic peptide using water molecules to fill the binding site. The work in this thesis has increased our understanding of the bound conformation of the mimetic peptide and could influence future structural modifications to the peptide to improve its efficacy as a vaccine therapeutic.

We have used several different techniques to gain insight into the molecular interactions of galactofuranosyl transferase GlfT2 with its substrates and *Shigella flexneri* Y antibody SYA-J6 with the peptide immunogen. These techniques include Nuclear Overhauser Effect (NOE) Spectroscopy, Saturation Transfer Difference (STD) NMR spectroscopy, 1D and 2D-TOCSY NMR spectroscopy, CORCEMA-ST and molecular dynamics (MD). A background to the systems studied and the techniques used is presented.

1.1 Carbohydrates

Carbohydrates, at the molecular level, are classified as polyhydroxy-aldehydes and polyhydroxy-ketones or compounds that can be hydrolyzed to these substances. ^[1-3] They contain carbon, hydrogen and oxygen and have the general formula $C_nH_{2n}O_n$. ^[1-3] Carbohydrates are vital to our existence and they have an assortment of functions. They serve as precursors for energy metabolism (e.g. glucose, starch), and provide the structural integrity in plants (e.g. cellulose) and in bacterial cell walls (e.g. as alternating polymers of *N*-acetylglucosamine and *N*-acetylmuramic acid cross-linked to peptides). ^[1-3] Furthermore, carbohydrates are also found as components of nucleic acids (e.g. D-ribose and 2-deoxy-D-ribose). ^[1-3]

In addition to these roles, carbohydrates are also involved in a variety of biological processes. ^[4-8] Glycoconjugates such as glycolipids, glycoproteins and D-Galactofuranose (D-Galf) containing glycoconjugates, whereby carbohydrates are attached to lipids and proteins, and the carbohydrate portions of these glycoconjugates, known as glycans, are vital to life. ^[4-8] Beyond their structural roles, glycans also serve to mediate cell adhesion as well as cellular signalling processes. ^[4-8] The inflammation response, for example, relies on the binding of selectins, which are expressed on endothelial cells following tissue damage, to carbohydrate ligands, specifically sialyl Lewis X, on circulating leukocytes. ^[4-8]

1.1.1 D-Galf containing glycoconjugates

D-Galactofuranose (D-Galf) glycoconjugates are found in protozoa, fungi and mycobacteria.^[9] In protozoa, D-Galf can exist as part of the glycophosphosphingolipid of *Endotrypanum schaudinni*^[10], the lipophosphoglycan (LPG) of *Leishmania* spp,^[11] as well as the lipopeptidophosphoglycan (LPPG) of *Trypanosoma cruzi* ^[12-15] forming crucial components of Glycosylphosphatidylinositol (GPI)-like structures. ^[9] In fungi, D-Galf can exist as a portion of the peptidophosphogalactomannan (pPGM) secreted by *Penicillium charlesii* ^[16] and as a component of hyphae of *Aspergillus niger* and other *Aspergillus* spp.^[17] Of particular interest to this thesis work, D-Galf-containing glycoconjugates are also found in the mycobacterial cell walls of *Mycobacterium leprae* and *Mycobacterium tuberculosis*. ^[18, 19]

1.1.2 Mycobacterium tuberculosis

According to the World Health Organization (WHO), one third of the World's population is infected with *Mycobacterium tuberculosis* (Figure 1.1). Three million deaths per year are attributed to tuberculosis (TB).^[20, 21] It is the most lethal bacterial infection characterized by prolonged coughing, chest pains and the coughing up of blood. The emergence of extensively drug resistant TB (XDR-TB) strains, resistant to second-line antibiotic treatment, prompted the WHO to declare TB a global health emergency. The cell wall that encapsulates tuberculosis (TB)-causing *Mycobacterium tuberculosis* is a beautiful structure known as the mycobacterial mycoyl-arabinogalactan-peptidoglycan (mAGP) complex (Figure 1.2). ^[22, 23] The polysaccharide moiety is composed of

approximately 30 D-Galf units. Since the mAGP is vital for the survival of the organism, ^[24] and D-Galf is absent in mammalian systems, interruptions in its biosynthesis can be considered a viable therapeutic point of attack. Current antibiotic strategies include isoniazid,^[25] which interrupts mycolic acid biosynthesis and ethambutol,^[26] which inhibits arabinosyl transferases. Recently, Lowary *et al.* ^[27] have proposed and shown that two galactofuranosyl transferases, GlfT1 and GlfT2, are responsible for the biosynthesis of the galactan portion of mAGP. An understanding of the mode of action of GlfT2 is one of the goals of this thesis.

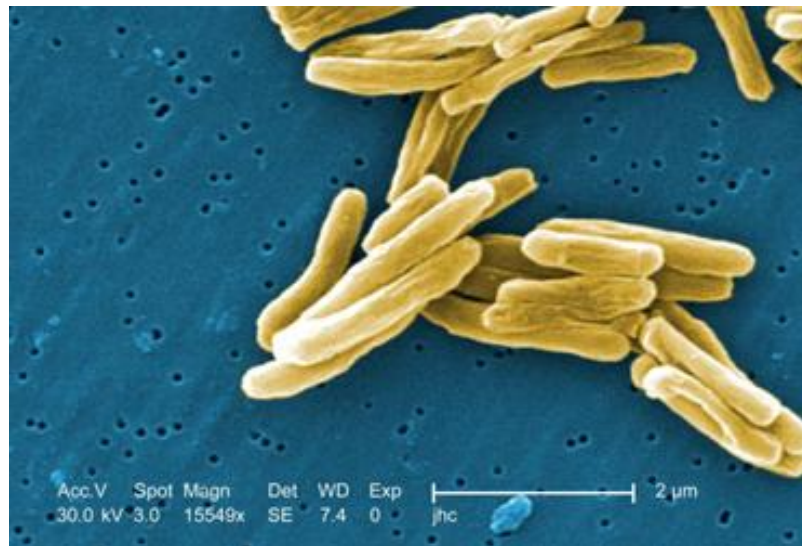


Figure 1.1. Scanning Electron Micrograph of *Mycobacterium tuberculosis*. (Photo credit: Janice Carr, Content Providers(s): CDC/Dr. Ray Butler; Janice Carr, http://en.wikipedia.org/wiki/File:Mycobacterium_tuberculosis.jpg)

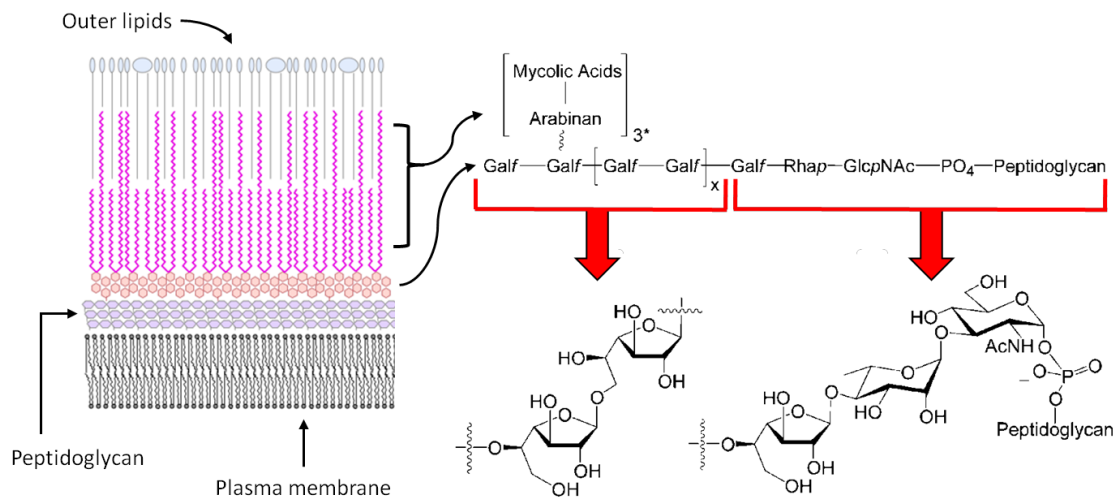


Figure 1.2. Pictorial representation of Mycobacterial mycolyl-arabinogalactan-peptidoglycan (mAGP) complex. Modified from a figure published by Y tambe, http://en.wikipedia.org/wiki/File:Mycobacterial_cell_wall_diagram.png.

1.1.3 Galactofuranosyl transferases

According to CAZy (<http://www.cazy.org>), a database of Carbohydrate-Active enZymes, glycosyl transferases (GTs) (EC 2.4.x.y) are enzymes that catalyze the formation of a glycosidic bond. The two partners involved are a glycosyl donor, typically a UDP sugar such as UDP-Galf, and a nucleophilic acceptor, usually the alcohol group of a growing biopolymer (Figure 1.3) ^[28]. In the database, these carbohydrate-acting enzymes are classified according to amino acid sequence. ^[29, 30] Currently, there are no crystal structures for GltT1 and GltT2; however, based on sequence comparisons available for other glycosyltransferases GltT2 and GltT1 have been assigned to the GT-2 family. The GT-2 family is characterized by DXD motifs that are involved in binding to the sugar nucleotide's phosphodiester moiety via divalent metal ions^[31](Figure 1.4). This family of enzymes also utilizes an inverting mechanism that leads to a β -linked glycosidic bond (Figure 1.5.). ^[31] Further classification based on 3D folds gives rise to the GT-A, GT-B and GT-C superfamilies.^[28, 31] GltT1 and GltT2 belong to the GT-A structural superfamily characterized by a mixed $\alpha/\beta/\alpha$ Rossmann type fold (Figure 1.4). ^[28, 31]

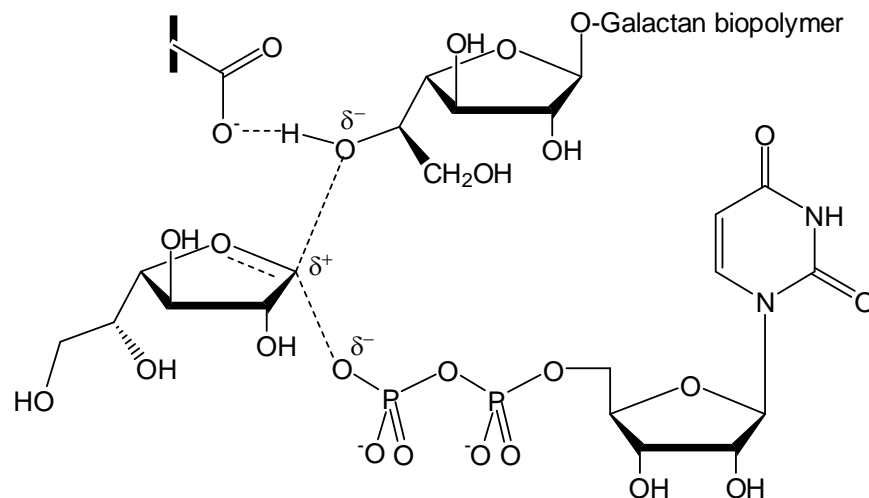


Figure 1.3. Putative transition state for a glycosyltransferase-mediated reaction.

(Reproduced with permission from *Org. Biomol. Chem.* **2004**, 2, 2418-2420.

Copyright © 2004 Royal Chemical Society. All rights reserved).

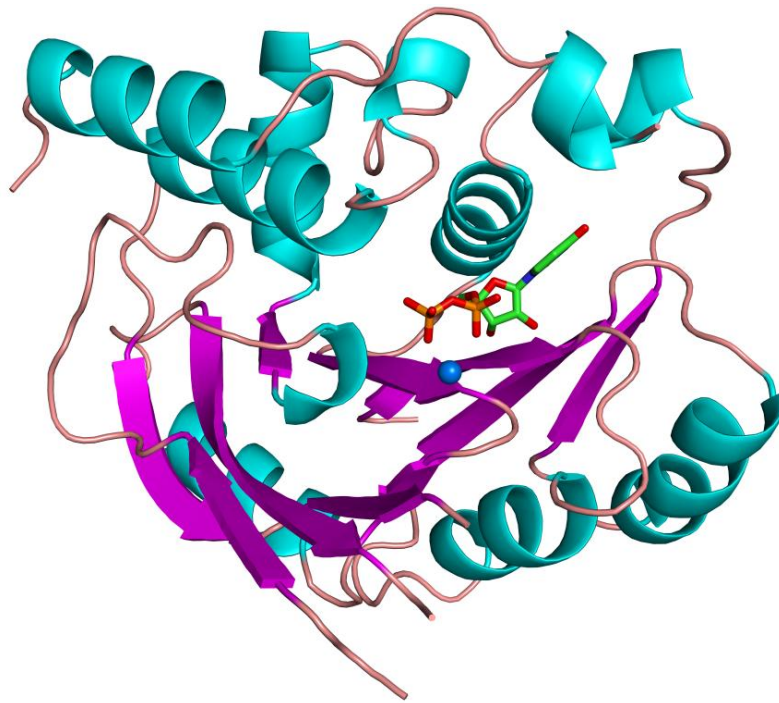


Figure 1.4. Crystal structure of SpsA (PDB: 1QGS): A GT-A glycosyltransferase involved in the synthesis of the spore coat of *Bacillus subtilis*.

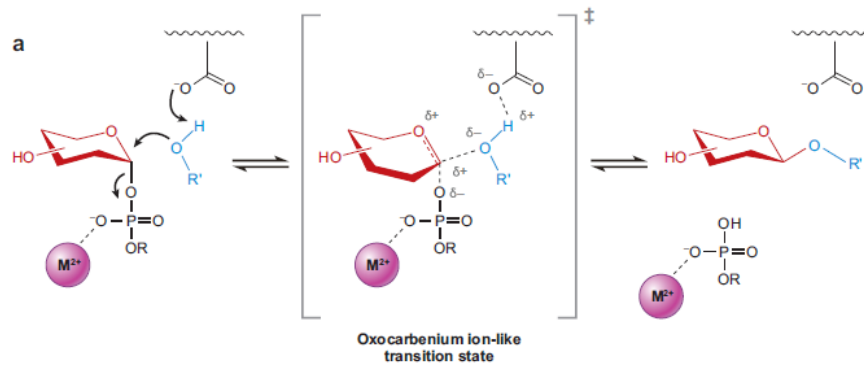


Figure 1.5. Proposed mechanism for an inverting glycosyl transferase. (Reproduced with permission from *Annu. Rev. Biochem.* **2008**, 77, 521–555. Copyright © 2008 by Annual Reviews. All rights reserved).

1.2 Peptide-Carbohydrate mimicry

1.2.1 Immune response against invading bacteria and viruses

When faced with foreign invaders such as viruses or the bacteria that are the focus of this thesis work, namely *M. tuberculosis* and *Shigella flexneri* Y, the human body calls upon white blood cells to mount a response. Lymphocytes are agranulocytes, meaning they lack specific granules in their cytoplasm, of which there are two major classes: B (bone marrow) cells and T (thymus) cells. ^[52] The major function of B cells, upon recognition of an antigen, is to produce antibodies and develop into memory B cells in a process known as humoral mediated acquired immunity, whilst the major function of the T cells is in the direct destruction of the invader (via natural killer cells, macrophages etc), in a process known as cell-mediated acquired immunity. ^[52] B cell immunity is activated when the antigen is protein or polysaccharide, such as the bacterial cell wall of *Staphylococcus aureus*, *Streptococcus pneumoniae* or *Clostridium tetani*. ^[52, 53] T cell immunity is frequently observed in the case of slow-acting bacteria such as *Mycobacterium leprae* and *Mycobacterium tuberculosis*, protozoa (*Leishmania donovani*), fungi (*Pneumocystis carinii*), cancer cells and transplanted tissue. ^[52, 53] B cells recognize soluble and native antigens whereas T cells require a processed, fragmented antigen that is presented by the major histocompatibility complex (MHC) residing on the surface of a professional antigen presenting cell (APC) to the T-helper cell. ^[53] The humoral response can be further classified as either thymus dependent or thymus independent B cell activation (Figure 1.6a). ^[52, 53] In the case of thymus-dependent activation, cytokines are released from

helper T cells, which cause the B cell to differentiate into antibody forming plasma cells that have immunological memory (Figure 1.6b).^[53] Certain antigens, such as the lipopolysaccharide (LPS) on the surface of *Shigella flexneri* Y, initiate a thymus-independent B cell response that is generally weaker than the thymus-dependent B cell response.^[53]

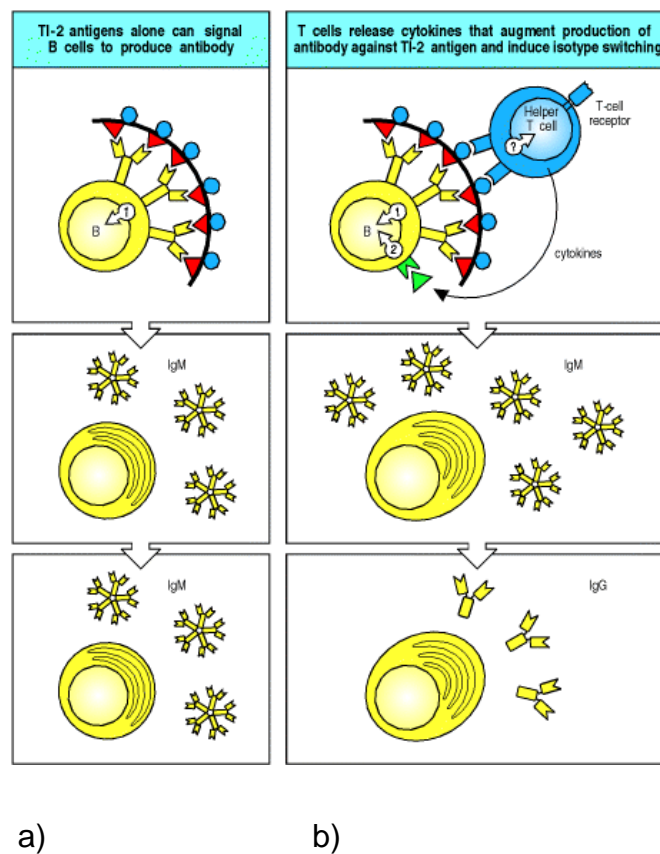


Figure 1.6. B cell activation can be either helper T-cell independent (a) or helper T-cell dependent (b). (Reproduced with permission from Charles A Janeway, Jr, Paul Travers, Mark Walport, and Mark J Shlomchik. *Immunobiology*, 5th edition, Garland Science, New York, NY. Copyright © 2001 by Garland Science. All rights reserved).

1.2.2 Carbohydrate mimetic peptides as leads for vaccines that target the bacterial cell wall

It is well known that carbohydrates in the form of polysaccharides on the surface of bacterial pathogens are critical for the organism to induce disease in its host.^[54] A strategy that involves the use of the bacterial carbohydrate coat as antigen in the development of a vaccine holds promise.^[54] However, carbohydrate-based therapeutics pose several challenges which hinder this goal. First, polysaccharides are difficult to synthesize. Second, polysaccharides induce a thymus independent immune response which does not produce memory B-cells. Third, polysaccharides fail to elicit an immunological response in children under 2 years of age. Peptides that mimic carbohydrates show promise as surrogates for carbohydrate-based vaccines in cases where the peptide demonstrates cross-reactivity, that is, the mimetic peptide can induce the production of antibodies that bind to the original carbohydrate antigen.^[55, 56] In this context, phage-displayed peptide libraries are commonly used in vaccine development for the discovery of new mimetic peptide leads.^[57-59] Phalipon et al.^[60] identified 19 nonapeptides by screening with monoclonal antibodies (mAbs) that bound the *Shigella flexneri* 5a lipopolysaccharide O-antigen. In our group^[61], we were successful in identifying an octapeptide with the sequence MDWNMHAA by phage library screening. We also found that the peptide is cross reactive with the complementary anti-carbohydrate antibody SYA-J6.^[62] An understanding of the interaction of SYA-J6 with the peptide MDWNMHAA is a goal of this thesis.

1.3 NMR Spectroscopy

The major research theme of this thesis is to examine the molecular interaction of small molecules binding to protein receptors. Nuclear magnetic resonance (NMR) spectroscopy is one of the key tools used in this work to study these interactions. An introduction to the nuclear Overhauser effect (NOE) is followed by a description of saturation transfer difference (STD) NMR and 1D- and 2D-STD-TOCSY NMR.

1.3.1 The nuclear Overhauser effect (NOE)

American physicist Albert W. Overhauser (Figure 1.7) developed the theory of dynamic nuclear polarization which became known as the nuclear Overhauser effect whilst a post doctoral fellow at the University of Illinois. ^[32, 33] Most notably he spent 25 years as a researcher at the Ford Motor company (Figure 1.8). Albert Overhauser is currently a faculty member at Purdue University.



Figure 1.7. American physicist Albert W. Overhauser born August 17, 1925 in San Diego, California. Photo credit: AIP Emilio Segre Visual Archives.



Figure1.8.2011 Ford Mustang. <http://www.ford.com/cars/mustang/gallery/photos/>

The nuclear Overhauser effect refers to a change in intensity of spin I resonance, after population of spin S (close in space) is inverted or saturated (equalize α and β populations). The population inversion forms the basis of 1D NOESY and 2D NOESY experiments. Whereas scalar couplings are transmitted through electrons in bonds and serve to identify groups and connectivity, dipolar couplings are transmitted through space and give insight into 3D molecular geometry. An example of a 2D NOESY spectrum is shown in Figure 1.9. ^[34] In this case, the stereochemistry at the stereogenic selenium center of the selenoxylitol compound was confirmed by NOE contacts between H-1a' and H-1equatorial as well as NOE contacts between H-1b' and H-5equatorial. ^[34] We can think of the NOE as a measurement of dipolar interactions whereby the intensity of one resonance changes when the transitions of another are perturbed from their equilibrium populations: ^[35]

$$\eta_I\{S\} = \frac{I - I_0}{I_0} \times 100 (\%) = \% \text{ enhancement}$$

nOe for spin I when spin S is perturbed

The origin of the NOE can be best represented in terms of a schematic energy level diagram for two spins, I and S that share a dipolar coupling (Figure 1.10). Upon saturation of resonance S, the population differences giving rise to resonance S are equalized such that there are no net transitions. The system attempts to return to equilibrium via relaxation processes W_0 or W_2 .

Quantitatively,

$$W_0 \propto \gamma_I^2 \gamma_S^2 \left[\frac{2\tau_C}{r^6 (1 + (\omega_I - \omega_S)^2 \tau_C^2)} \right]$$

and,

$$W_2 \propto \gamma_I^2 \gamma_S^2 \left[\frac{12\tau_C}{r^6 (1 + (\omega_I + \omega_S)^2 \tau_C^2)} \right]$$

For a small molecule, its rotational frequency matches the energy difference brought about by W_2 transitions and so that is the preferred pathway. For large molecules, relaxation is brought about by W_0 transitions.

The population inversion technique is typically used in the study of NOE kinetics, where the initial growth rate stage of the NOE is linear and dependent on r^{-6} .^[35] This transient NOE regime is experimentally achieved by the use of short mixing times prior to sampling the FID.^[35] During the mixing time, spins interact and growth of the NOE occurs.^[35]

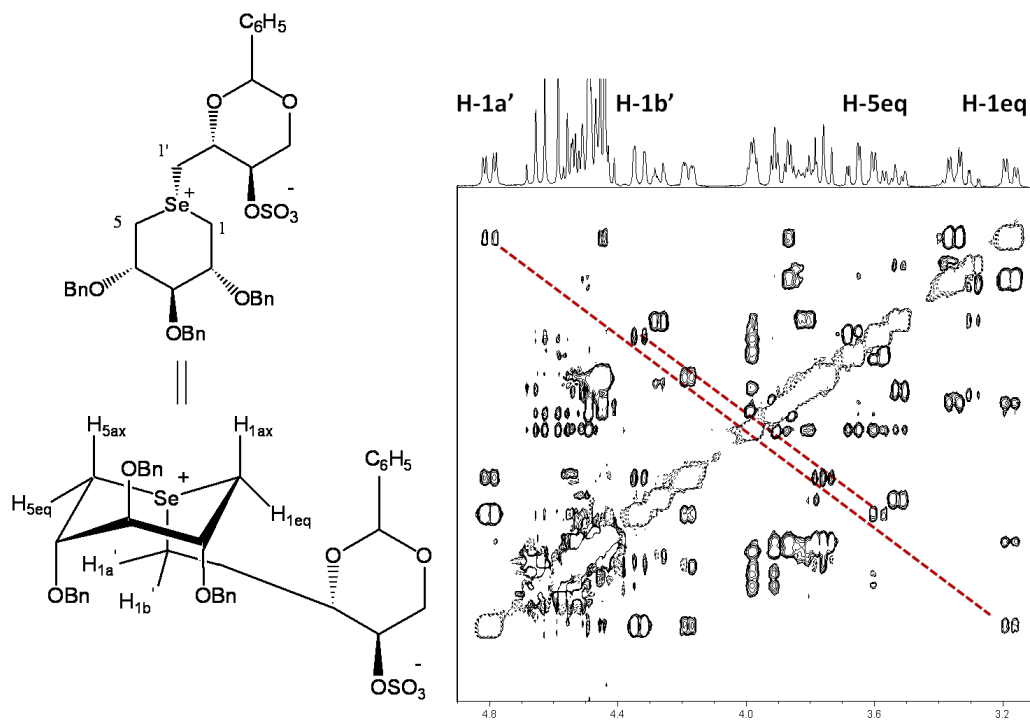


Figure 1.9. 2D NOESY spectrum of 2,3,4-tri-*O*-benzyl-1,5-dideoxy-1,5-[[*(2R,3R)*-2,4-*O*-benzylidene-2,4-dihydroxy-3-(sulfoxy)butyl]-*(R)*-episelenoniumylidene] xylitol inner salt.

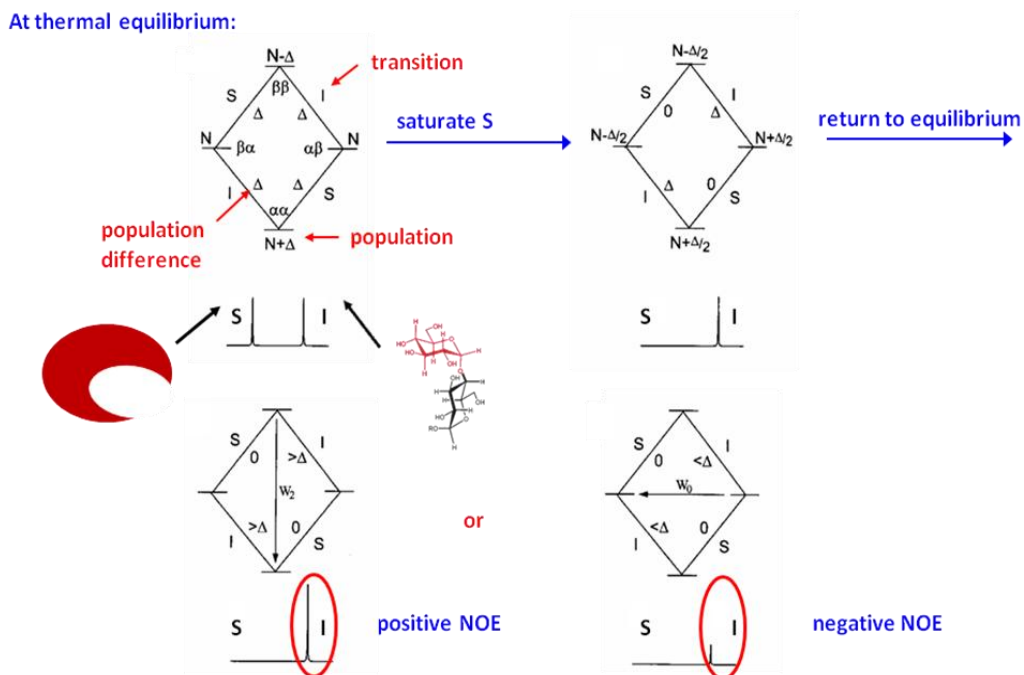


Figure 1.10. A two spin energy level diagram. Spins I and S share a dipolar coupling. Modified from a figure published by T. D. W Claridge, *High-resolution NMR techniques in organic chemistry*, Pergamon, Amsterdam, New York, 1999.

1.3.2 Saturation Transfer Difference NMR spectroscopy

Saturation Transfer Difference NMR (STD NMR) spectroscopy^[36, 37] is a method in Nuclear Magnetic Resonance spectroscopy used to examine the interaction of ligands with protein receptors (Figures 1.11,1.12). This method relies on the principle that small molecules exhibit short correlation times τ_C (time required for molecule to rotate through 1 radian) i.e. fast motion which gives positive NOE effects (Figure 1.10).^[38, 39] Larger molecules, such as proteins typically show long τ_C (slower motion) and exhibit negative NOE effects (Figure 1.10).^[38, 39] On binding, a small molecule ligand can be expected to show properties reflective of the large protein,^[38, 39] namely, negative NOEs that reflect the ligand's bound conformation in the protein receptor.^[38, 39]

Experimentally, an NMR sample is prepared with the ligand in approximately 100 fold molar excess over the protein. This allows one to use the protein in μM concentrations. The STD-NMR pulse sequence from Bruker operates in an interleaved manner, alternating between on-resonance and off-resonance saturation. Specifically, the protein signal is saturated and via spin-diffusion this magnetization spreads throughout the protein (Figure 1.11). Typically, the protein is saturated in a region devoid of ligand resonances (e.g. -2ppm). This is referred to as the *on-resonance* experiment. Ligands which bind to the protein also become saturated.^[38, 39] The degree to which this occurs is dependent on the time the ligand spends in contact with the protein. The saturated ligand undergoes a loss of signal intensity due to negative NOE build-up, as the system relaxes.^[38, 39] One also obtains an NMR spectrum of the same

sample with *off-resonance* irradiation. ^[38, 39] Experimentally, this is obtained by saturation at 30ppm or at any region devoid of protein and ligand NMR signals. From these two NMR experiments it is possible to subtract the on-resonance spectrum from the off-resonance spectrum. ^[38, 39] The difference gives an NMR spectrum of the bound ligand only hence the name, Saturation Transfer Difference NMR spectroscopy. ^[38, 39] Ligands which do not bind do not give signals. ^[38, 39] This is demonstrated with a sample of bovine serum albumin (BSA) containing tryptophan and glucose (Figure 1.13). Only the ligand that binds, i.e. tryptophan, appears in the STD-NMR spectrum.

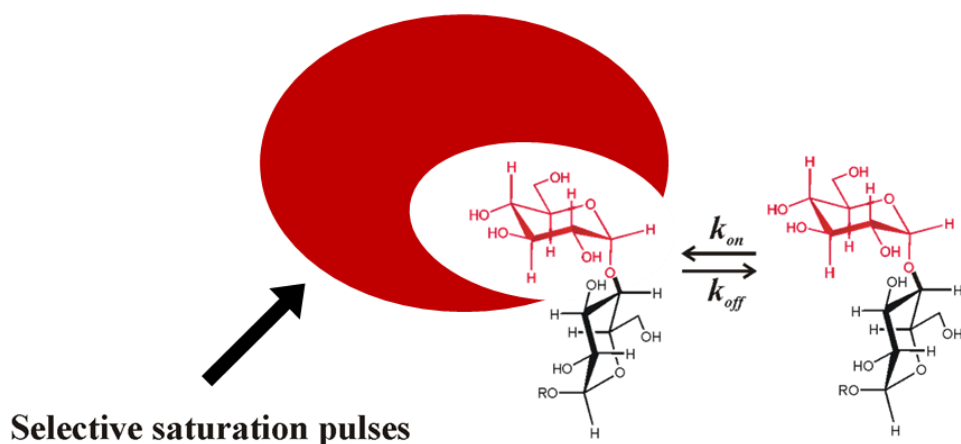


Figure 1.11. The STD-NMR experiment. The protein signal is selectively saturated, typically with a soft pulse, in a region devoid of ligand resonances. Magnetization spreads throughout the protein via spin-diffusion (^1H - ^1H intramolecular cross relaxation). This magnetization then travels to the bound ligand via ^1H - ^1H intermolecular cross relaxation. The “hot” ligand departs the binding site and takes with it information about its bound state. Modified from a figure by Margaret A. Johnson, PhD thesis.

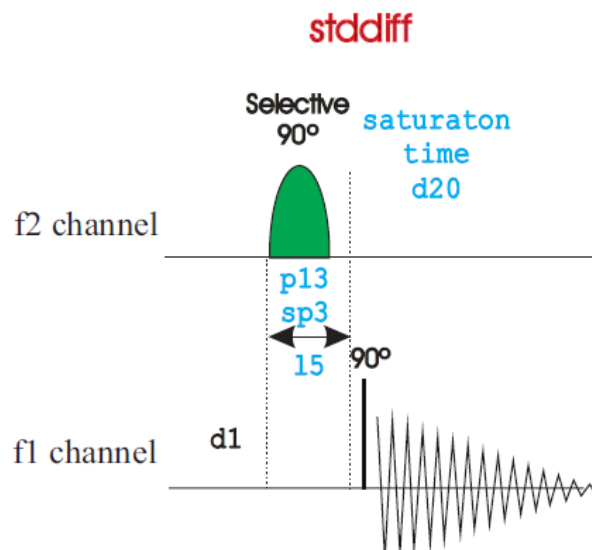


Figure 1.12. The STD-NMR pulse sequence. (Reproduced with permission Bruker Biospin pulse program catalogue, TOPSPIN v2.0. NMR Guide. Copyright © 2006 by Bruker Biospin GmbH. All rights reserved)

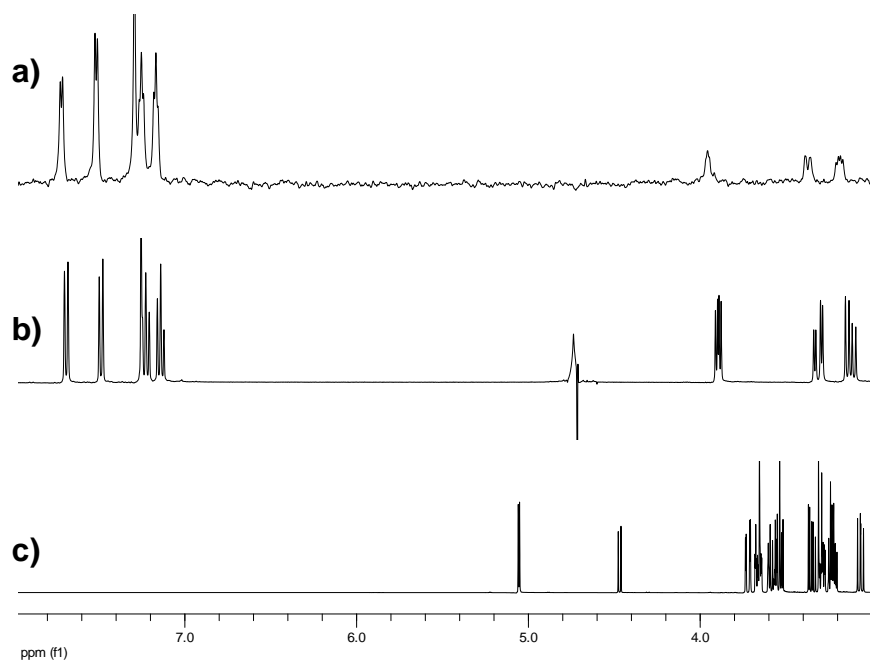


Figure 1.13 a) STD-NMR spectrum of a sample containing BSA, tryptophan and glucose. b) ¹H NMR spectrum of tryptophan and c) ¹H NMR spectrum of glucose. Only tryptophan gives STD-NMR signals.

The use of STD NMR experiments for epitope mapping to determine which residues are essential for binding is critically dependent on the binding kinetics, particularly, k_{off} , of the system in question (see Table 1.1). Extremely tight binding ligands ($K_D < 100\text{nM}$), undergo relaxation at a faster rate than “fresh” ligand being shuttled into the binding site to receive saturation. [38, 39] Extremely weak binding ligands ($K_D > 10\text{mM}$) never enter the binding site. [38, 39] In both cases the STD amplification is *poor* or *non-existent*. Assuming a diffusion controlled on rate, k_{on} , only binding ligands with a K_D in the range of 10^{-3} to 10^{-8} M are able to exit the binding site multiple times which leads to an amplification in STD NMR signal. [38, 39] Furthermore, these ligands relax via W_0 cross-relaxation pathways (Figure 1.10), which are dependent on r^{-6} . Our group has recently generated STD-NMR build-up curves in the study of UDP-Galactopyranose mutase (UGM), a flavoenzyme that catalyzes interconversion of UDP-galactopyranose (UDP-Galp) and UDP-galactofuranose (UDP-Galf), by varying the saturation time (c.f mixing time). [40] At short saturation times, the information obtained is analogous to a transient NOE, which gives distance information between spins. This distance information forms the basis of the *STD-NMR epitope map*. The acquisition of the STD-NMR epitope map is a goal in the two research targets of this thesis: GlfT2, as it binds two trisaccharide substrates and SYA/J6, as it binds a carbohydrate mimetic peptide.

Table 1.1. Kinetic and thermodynamic parameters and suitability for study via STD-NMR spectroscopy.

Binding Range	K_D	k_{off} (s^{-1})	k_{on} ($M^{-1}s^{-1}$)	suitability for STD-NMR
		$< 10^2$ if k_{on} is	10^8	not acceptable
strong	< 100 nM	< 10	10^7	
		< 1	10^6	
		< $10^3 - 1$	10^8	acceptable
medium	> 0.1 μM	< $10^4 - 10$	10^7	
		< $10^5 - 10^2$	10^6	
		> 10^6	10^8	not acceptable
weak	> 100 mM	> 10^5	10^7	
		> 10^4	10^6	

1.3.3 1D and 2D STD-TOCSY NMR spectroscopy

The TOCSY experiment is frequently used to aid in the assignment of multicomponent spin systems. Total Correlation Spectroscopy NMR provides correlations between all spins of a single spin system not simply between neighbouring nuclei. The key feature of this experiment is the application of a spin lock, essentially a series of continuous 180° pulses, which cause all the spins to be locked along one direction. During the spin-lock period, typically 40ms-100ms in duration, the magnetization is transferred between all coupled nuclei in a spin system via strong scalar coupling, even if the nuclei are not directly coupled. The TOCSY NMR pulse sequence is incredibly useful when

examining carbohydrates, with repeating units separated by a glycosidic bond. We fully characterized by 1D and 2D NMR spectroscopy, a synthetic pentasaccharide antigen, which binds to the antibody SYA/J6 which was directed against the O-antigen of *Shigella flexneri* Y. This work is described in Chapter 5. The 2D TOCSY experiment was used to aid in the assignment of proton resonances (Figure 1.14).

In the STD variant of this experiment, if the selectively excited resonance has its intensity decreased due to intermolecular NOE transfer, and this magnetization is propagated along the spin system, protons in the same spin system will also have their intensity decreased. Previous work in our laboratory included the implementation of a novel pulse sequence known as STD-1D-TOCSY.^[41]

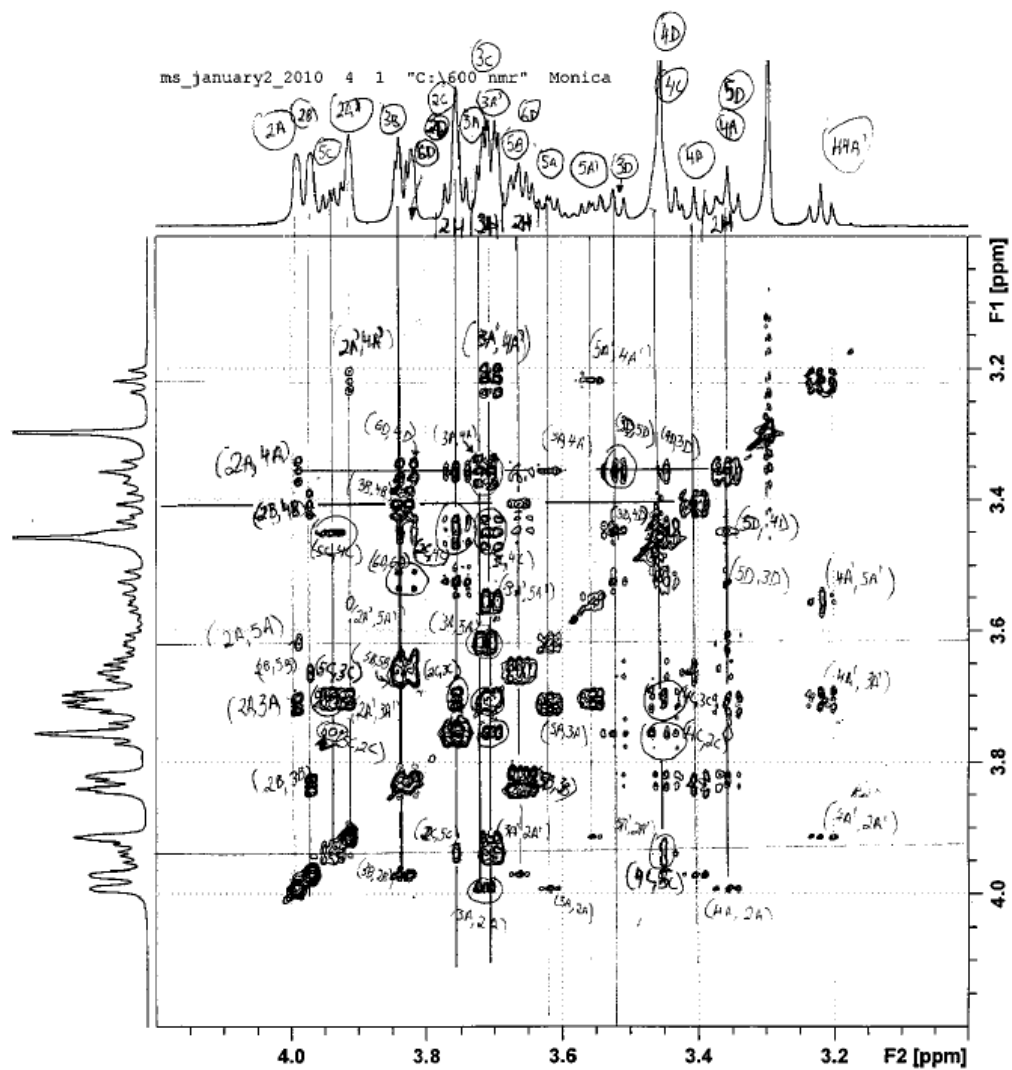


Figure 1.14. 2D TOCSY experiment for the pentasaccharide (shown in Figure 5.10) utilizing the mlevesgpph Bruker pulse sequence at 600 MHz and 298 K. Each hexose has a distinct spin system; thus, 5 spin systems are observed.

1.3.4 Quantifying STD-NMR signals with CORCEMA-ST

In STD-NMR spectroscopy, only ligand protons in close proximity to the protein are preferentially relaxed; these ligands dissociate from the protein-ligand complex, and take with them the information of the bioactive, bound conformation. The average conformation, typically in PDB format, can then be used to calculate the magnetization transfer effects of the ligand protons in close proximity to the protein protons using CORCEMA-ST (complete relaxation and conformational exchange matrix analysis of saturation transfer), ^[42-49] a program that performs a total matrix analysis of relaxation and exchange effects to generate predicted STD-NMR intensities. A comparison of the predicted STD-NMR intensities and the experimental STD-NMR intensities can be used to discriminate between binding models and validate the preferred ligand binding mode. Our laboratory has considerable expertise with this protocol and we have validated binding modes of several substrates and inhibitors in enzyme active sites. ^[40, 50, 51]

The CORCEMA-ST program is based on the observation that the NOE response depends on a competition between cross relaxation and auto relaxation. ^[38] Methylene protons are likely to exhibit smaller STD responses due to strong mutual dipolar relaxation between proximal, geminal protons. ^[38] The CORCEMA-ST program uses as input the PDB coordinates of the ligand and receptor and in addition to parameters such as free-ligand and bound-ligand correlation times τ_C , knowledge of saturated protein protons, exchange rates and

spectrometer frequency. The output is predicted STD-NMR intensities for the ligand.

A good correlation between predicted and experimental STD-NMR intensities indicates an agreement between the solution NMR data and the solid state crystallographic image of the bound structure. A comparison of the predicted STD-NMR intensities and the experimental STD-NMR intensities, as in the case of *Shigella flexneri* Y mAb SYA/J6 complexed to its carbohydrate-mimetic peptide, is used in this thesis to provide insight into the mode of binding of these two partners.

1.4 Molecular Dynamics

In 1977, J. Andrew McCammon et al. ^[63], were the first to publish a study on protein dynamics by solving equations of motions for the atoms of bovine pancreatic trypsin inhibitor, a 58 amino acid protein consisting of approximately 500 atoms. This was the first published study using molecular dynamics. Since then, much larger systems have been examined including the satellite tobacco mosaic virus, a beast measuring 1×10^6 atoms. ^[64-66] The authors were interested in the assembly of the viral capsid. We have also used MD to elucidate the dynamics of UDP-Galactopyranose mutase (UGM).^[40] Molecular dynamics calculations are ideal to study the motions and conformations of biological molecules. In this thesis, we are particularly interested in the molecular interactions of monoclonal antibody SYA/J6 as it binds its peptide ligand. Therefore, molecular dynamics is a tool that can give us information about the hydrophobic interactions, hydrogen bonds and van der Waals forces that exist between SYA/J6 and MDWNMHAA.

Molecular dynamics works by solving Newton's law of motion:

$$F = ma$$

Integration of Newton's law of motion leads to a time and position dependent equation. At each time point, the potential energy, and hence the force, of the atoms are calculated. The force propels the atoms to a new position based on the acceleration generated by the force. As such, a time dependent trajectory is generated. The MD algorithm is shown in Figure 1.15. The potential energy V is defined as:

$$V_{TOTAL} = E_{BOND} + E_{ANGLE} + E_{TORSIONS} + E_{ELECTROSTATIC} + E_{VDW}$$

Where E_{BOND} is the bond length potential energy, E_{ANGLE} is the bond angle potential energy, $E_{TORSIONS}$ is the torsion angle potential energy and the remaining non bonding potential energy contributions, $E_{ELECTROSTATIC}$ and E_{VDW} are electrostatic and van der Waals, respectively.

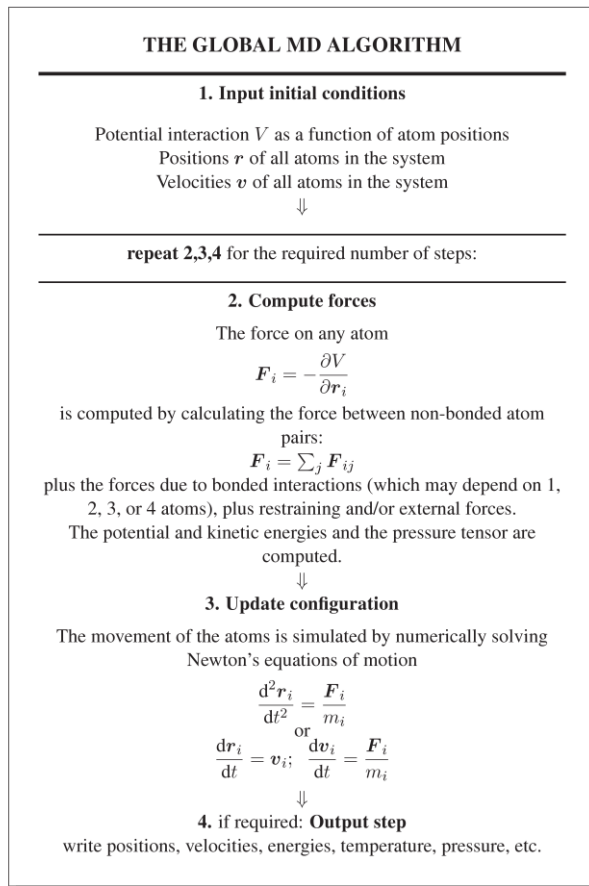


Figure 1.15. The MD algorithm. (Reproduced with permission from GROMACS user manual version 4.5. Copyright © 2001-2010 by The GROMACS development teams at the Royal Institute of Technology and Uppsala University, Sweden. All rights reserved)

1.5 Thesis Overview

This thesis work is presented primarily in journal article style, with **Chapter 1** as a general introduction, followed by **Chapters 2** and **3** as journal articles, **Chapters 4** and **5** as manuscripts submitted, and **Chapter 6** as closing remarks.

In **Chapter 1**, a background to the research carried out in this thesis, as well as a general introduction to the techniques used in this thesis are described.

In **Chapter 2**, the manuscript (Monica G. Szczepina, Ruixiang B. Zheng, Gladys C. Completo, Todd L. Lowary, and B. Mario Pinto; *Bioorg. Med. Chem.* **2010**, *18*, 5123-5128) is presented. It describes STD-NMR epitope studies on two substrates for an enzyme, GlfT2, involved in the biosynthesis of the galactan portion of the *Mycobacterium tuberculosis* cell wall.

In **Chapter 3**, the manuscript (Monica G. Szczepina, Ruixiang B. Zheng, Gladys C. Completo, Todd L. Lowary, and B. Mario Pinto; *ChemBioChem* **2009**, *10*, 2052-2059) is presented. It describes the use of STD-NMR and STDD-NMR competition experiments to answer how many active sites are present in the enzyme GlfT2.

In **Chapter 4**, we present a manuscript (Monica G. Szczepina, Dustin W. Bleile, Johannes Müllegger, Andrew R. Lewis, and B. Mario Pinto) that was submitted to *Biochemistry* entitled “WaterLOGSY NMR Experiments Detect Immobilized Water Molecules that Bridge Peptide Mimic MDWNMHAA to Anti-carbohydrate Antibody SYA/J6”. We describe the use of WaterLOGSY NMR spectroscopy, which when used in combination with STD NMR spectroscopy, allowed us to detect immobilized water molecules sandwiched between the

SYA/J6 antibody and the peptide antigen. We confirm the presence of bridging water molecules by molecular dynamics simulations.

In **Chapter 5**, the manuscript (Monica G. Szczepina, Dustin W. Bleile, and B. Mario Pinto; *Chem. Eur. J.* **ASAP**) is presented. We perform full STD NMR build up curves on the carbohydrate mimetic peptide MDWNMHAA to obtain an epitope map. We quantify our data with CORCEMA-ST and find that the solution structure of bound MWDNMHAA differs from the X-Ray structure. To reconcile this difference we perform molecular dynamics simulations in addition to performing modifications to the peptide backbone via simulated annealing. Finally, we examine the epitope map of the synthetic pentasaccharide antigen.

In **Chapter 6**, a thesis summary is presented.

1.6 References

- [1] W. H. Brown, C. S. Foote, *Organic chemistry*, Saunders College Pub., Fort Worth, **1998**.
- [2] C. K. Mathews, K. E. Van Holde, K. G. Ahern, *Biochemistry*, Benjamin Cummings, San Francisco, Calif., **2000**.
- [3] T. W. G. Solomons, *Organic chemistry*, Wiley, New York, **1992**.
- [4] R. A. Dwek, *Chem. Rev.* **1996**, *96*, 683-720.
- [5] A. Helenius, M. Aebi, *Science* **2001**, *291*, 2364-2369.
- [6] K. W. Moremen, R. B. Trimble, A. Herscovics, *Glycobiology* **1994**, *4*, 113-125.
- [7] M. E. Taylor, K. Drickamer, *Introduction to Glycobiology*, Oxford University Press, Oxford, **2003**.
- [8] A. Varki, *Essentials of glycobiology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, **1999**.
- [9] R. M. de Lederkremer, W. Colli, *Glycobiology* **1995**, *5*, 547-552.
- [10] J. O. Previato, L. Mendoncapreviato, C. Jones, R. Wait, *Glycoconjugate J.* **1993**, *10*, 340-340.
- [11] S. J. Turco, A. Descoteaux, *Annu. Rev. Microbiol.* **1992**, *46*, 65-94.

- [12] R. M. de Lederkremer, M. J. M. Alves, G. C. Fonseca, W. Colli, *Biochim. Biophys. Acta* **1976**, *444*, 85-96.
- [13] R. M. de Lederkremer, O. L. Casal, M. J. M. Alves, W. Colli, *FEBS Lett.* **1980**, *116*, 25-29.
- [14] R. M. de Lederkremer, C. Lima, M. I. Ramirez, M. A. J. Ferguson, S. W. Homans, J. Thomasoates, *J. Biol. Chem.* **1991**, *266*, 23670-23675.
- [15] J. O. Previato, P. A. J. Gorin, M. Mazurek, M. T. Xavier, B. Fournet, J. M. Wieruszesk, L. Mendoncapreviato, *J. Biol. Chem.* **1990**, *265*, 2518-2526.
- [16] C. J. Unkefer, J. E. Gander, *J. Biol. Chem.* **1990**, *265*, 685-689.
- [17] E. M. Barretobergter, L. R. Travassos, *Carbohydr. Res.* **1980**, *86*, 273-285.
- [18] M. Daffe, P. J. Brennan, M. McNeil, *J. Biol. Chem.* **1990**, *265*, 6734-6743.
- [19] M. McNeil, S. J. Wallner, S. W. Hunter, P. J. Brennan, *Carbohydr. Res.* **1987**, *166*, 299-308.
- [20] <http://www.who.int/tb/en/>.
- [21] T. M. Daniel, *Respiratory Med.* **2006**, *100*, 1862-1870.

- [22] P. J. Brennan, *Tuberculosis* **2003**, *83*, 91-97.
- [23] T. L. Lowary, in *Glycoscience: Chemistry and Chemical Biology*, Vol. 3 (Eds.: B. O. Fraser-Reid, K. Tatsuta, J. Thiem), Springer-Verlag, Berlin, **2001**, pp. 2005-2080.
- [24] F. Pan, M. Jackson, Y. F. Ma, M. McNeil, *J. Bacteriol.* **2001**, *183*, 3991-3998.
- [25] E. K. Schroeder, N. DeSouza, D. S. Santos, J. S. Blanchard, L. A. Basso, *Curr. Pharm. Biotech.* **2002**, *3*, 197-225.
- [26] A. E. Belanger, G. S. Besra, M. E. Ford, K. Mikusova, J. T. Belisle, P. J. Brennan, J. M. Inamine, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 11919-11924.
- [27] M. Belanova, P. Dianiskova, P. J. Brennan, G. C. Completo, N. L. Rose, T. L. Lowary, K. Mikusova, *J. Bacteriol.* **2008**, *190*, 1141-1145.
- [28] L. L. Lairson, B. Henrissat, G. J. Davies, S. G. Withers, *Annu. Rev. Biochem.* **2008**, *77*, 521-555.
- [29] J. A. Campbell, G. J. Davies, V. Bulone, B. Henrissat, *Biochem. J.* **1997**, *326*, 929-939.
- [30] P. M. Coutinho, E. Deleury, G. J. Davies, B. Henrissat, *J. Mol. Biol.* **2003**, *328*, 307-317.

- [31] S. Berg, D. Kaur, M. Jackson, P. J. Brennan, *Glycobiology* **2007**, *17*, 35R-56R.
- [32] T. R. Carver, C. P. Slichter, *Phys. Rev.* **1953**, *92*, 212-213.
- [33] A. W. Overhauser, *Phys. Rev.* **1953**, *92*, 411-415.
- [34] M. G. Szczepina, B. D. Johnston, Y. Yuan, B. Svensson, B. M. Pinto, *J. Am. Chem. Soc.* **2004**, *126*, 12458-12469.
- [35] T. D. W. Claridge, *High-resolution NMR techniques in organic chemistry*, Pergamon, Amsterdam ; New York, **1999**.
- [36] M. Mayer, B. Meyer, *Angew. Chem., Int. Ed.* **1999**, *38*, 1784-1788.
- [37] M. Mayer, B. Meyer, *J. Am. Chem. Soc.* **2001**, *123*, 6108-6117.
- [38] C. A. Lepre, J. M. Moore, J. W. Peng, *Chem. Rev.* **2004**, *104*, 3641-3675.
- [39] B. Meyer, T. Peters, *Angew. Chem., Int. Ed.* **2003**, *42*, 864-890.
- [40] Y. Yuan, D. W. Bleile, X. Wen, D. A. R. Sanders, K. Itoh, H. W. Liu, B. M. Pinto, *J. Am. Chem. Soc.* **2008**, *130*, 3157-3168.
- [41] M. A. Johnson, B. M. Pinto, *J. Am. Chem. Soc.* **2002**, *124*, 15368-15374.
- [42] N. R. Krishna, V. Jayalakshmi, in *Bioactive Conformation II, Vol. 273*, **2008**, pp. 15-54.

- [43] N. R. Krishna, V. Jayalakshmi, *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *49*, 1-25.
- [44] V. Jayalakshmi, N. R. Krishna, *J. Am. Chem. Soc.* **2005**, *127*, 14080-14084.
- [45] V. Jayalakshmi, T. Biet, T. Peters, N. R. Krishna, *J. Am. Chem. Soc.* **2005**, *127*, 7261-7261.
- [46] V. Jayalakshmi, T. Biet, T. Peters, N. R. Krishna, *J. Am. Chem. Soc.* **2004**, *126*, 8610-8611.
- [47] V. Jayalakshmi, N. R. Krishna, *J. Magn. Reson.* **2004**, *168*, 36-45.
- [48] A. Bhunia, V. Jayalakshmi, A. J. Benie, O. Schuster, S. Kelm, N. R. Krishna, T. Peters, *Carbohydr. Res.* **2004**, *339*, 259-267.
- [49] V. Jayalakshmi, N. R. Krishna, *J. Magn. Reson.* **2002**, *155*, 106-118.
- [50] X. Wen, Y. Yuan, D. A. Kuntz, D. R. Rose, B. M. Pinto, *Biochemistry* **2005**, *44*, 6729-6737.
- [51] Y. Yuan, X. Wen, D. A. R. Sanders, B. M. Pinto, *Biochemistry* **2005**, *44*, 14080-14089.
- [52] W. Kapit, R. I. Macey, E. Meisami, *The physiology coloring book*, Addison Wesley Longman Inc., San Francisco, **2000**.

- [53] C. Janeway, *Immunobiology : the immune system in health and disease*, Garland Science, New York, **2005**.
- [54] H. J. Jennings, *Adv. Exp. Med. Biol.* **1988**, 228, 495-550.
- [55] W. A. Hutchins, T. Kieber-Emmons, G. M. Carlone, M. A. J. Westerink, *Hybridoma* **1999**, 18, 121-129.
- [56] M. A. J. Westerink, P. C. Giardina, M. A. Apicella, T. Kieberemmons, *Proc. Natl. Acad. Sci. U. S. A.* **1995**, 92, 4021-4025.
- [57] G. Cunto-Amesty, P. Luo, B. Monzavi-Karbassi, T. Kieber-Emmons, *Int. Rev. Immunol.* **2001**, 20, 157-180.
- [58] B. Monzavi-Karbassi, G. Cunto-Amesty, P. Luo, T. Kieber-Emmons, *Trends Biotechnol.* **2002**, 20, 207-214.
- [59] M. B. Zwick, J. Q. Shen, J. Scott, *Curr. Opin. Biotechnol.* **1998**, 9, 427-436.
- [60] A. Phalipon, A. Folgori, J. Arondel, G. Sgaramella, P. Fortugno, R. Cortese, P. J. Sansonetti, F. Felici, *Eur. J. Immunol.* **1997**, 27, 2620-2625.
- [61] S. L. Harris, L. Craig, J. S. Mehroke, M. Rashed, M. B. Zwick, K. Kenar, E. J. Toone, N. Greenspan, F. I. Auzanneau, J. R. Marino-

- Albernas, B. M. Pinto, J. K. Scott, *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 2454-2459.
- [62] S. Borrelli, R. B. Hossany, B. M. Pinto, *Clin. Vaccine Immunol.* **2008**, *15*, 1106-1114.
- [63] J. A. McCammon, B. R. Gelin, M. Karplus, *Nature* **1977**, *267*, 585-590.
- [64] P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, K. Schulten, *Structure* **2006**, *14*, 437-449.
- [65] A. Arkhipov, P. L. Freddolino, K. Schulten, *Structure* **2006**, *14*, 1767-1777.
- [66] A. Arkhipov, W. H. Roos, G. J. L. Wuite, K. Schulten, *Biophys. J.* **2009**, *97*, 2061-2069.

CHAPTER 2: STD-NMR Studies of Two Acceptor Substrates of GlfT2, a Galactofuranosyltransferase from *Mycobacterium tuberculosis*: Epitope Mapping Studies

This chapter comprises the manuscript “***STD-NMR Studies of Two Acceptor Substrates of GlfT2, a Galactofuranosyltransferase from *Mycobacterium tuberculosis*: Epitope Mapping Studies***” which was published in *Bioorganic & Medicinal Chemistry* (2010, 18, 5123-5128).

Monica G. Szczepina,^a Ruixiang B. Zheng,^b Gladys C. Completo,^b

Todd L. Lowary,^b and B. Mario Pinto^a

^aDepartment of Chemistry, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada

^bDepartment of Chemistry and Alberta Ingenuity Centre for Carbohydrate Science, Gunning-Lemieux Chemistry Centre, University of Alberta, Edmonton, Alberta, T6G 2G2, Canada

As a prelude to STD-NMR competition experiments described in Chapter 3, we were interested in examining the epitopes of two trisaccharide acceptor substrates for the enzyme GlfT2, a bifunctional galactofuranosyltransferase involved in the construction of the mycobacterial cell wall, the mycolyl–arabinogalactan–peptidoglycan (mAGP) complex.

The preparation of GlfT2 was carried out by Mr. Ruixiang B. Zheng. The synthesis of the trisaccharides was carried out by Dr. Gladys C. Completo. Dr. A. R. Lewis performed the NMR deconvolutions. The thesis author performed all the NMR experiments and NMR data interpretation.

2.1 Keywords

STD-NMR, Epitope mapping, Galactofuranosyltransferase, GlfT2,
Trisaccharide Substrates

2.2 Abstract

The major structural component of the mycobacterial cell wall, the mycolyl–arabinogalactan–peptidoglycan complex, possesses a galactan core composed of approximately 30 galactofuranosyl (*Gal**f*) residues attached via alternating β -(1→6) and β -(1→5) linkages. Recent studies have shown that the entire galactan is synthesized by two bifunctional galactofuranosyltransferases, GlfT1 and GlfT2. We report here saturation-transfer difference (STD) NMR studies of GlfT2 using two trisaccharide acceptor substrates, β -D-Gal*f*-(1→6)- β -D-Gal*f*-(1→5)- β -D-Gal*f*-O(CH₂)₇CH₃ (**2**) and β -D-Gal*f*-(1→5)- β -D-Gal*f*-(1→6)- β -D-Gal*f*-O(CH₂)₇CH₃ (**3**), as well as the donor substrate for the enzyme, UDP-Gal*f*. Epitope mapping demonstrated a greater enhancement toward the “reducing” ends of both trisaccharides, and that UDP-galactofuranose (UDP-Gal*f*) made more intimate contacts through its nucleotide moiety. This observation is consistent with the greater flexibility required within the active site of the reaction between the growing polymer acceptor and the UDP-Gal*f* donor. The addition of UDP-Gal*f* to either **2** or **3** in the presence of GlfT2 generated a tetrasaccharide product, indicating that the enzyme was catalytically active.

2.3 Introduction

According to the World Health Organization, one third of the World's population is infected with *Mycobacterium tuberculosis*. Three million deaths per year are attributed to tuberculosis (TB), the most lethal bacterial infection. [1, 2] The pathogenicity of the TB-causing bacterium, *Mycobacterium tuberculosis*, is attributed to the mycolyl–arabinogalactan–peptidoglycan (mAGP) complex that not only encapsulates this organism but also other mycobacterial species. [3-6] Galactofuranose (Gal f) residues are present in a number of protozoal, fungal, and bacterial organisms, [7-9] but absent in mammals. [10] Inhibitors of the enzymes involved in Gal f metabolism are therefore of interest as novel therapeutic agents. [11] The mAGP is a major structural component of the mycobacterial cell wall and serves as an essential permeability barrier, thus protecting the organism from its environment. [12]

The core of the mAGP is composed of a polymer of D-Gal f residues connected via alternating β -(1 \rightarrow 5) and β -(1 \rightarrow 6) linkages. This galactan is attached to peptidoglycan through an α -L-rhamnopyranosyl-(1 \rightarrow 3)-2-acetamido-2-deoxy- β -D-glucopyranosyl-phosphate linkage (the linker disaccharide), and also bears three mycolylated arabinan domains (1, Figure 2.1). [5, 13] Investigation of the biosynthesis of the mAGP has identified many of the enzymes involved in this process. [12, 14] The galactan domain is constructed by only two enzymes, both of which are bifunctional. [15] One transferase, GlfT1, [16] adds the first and second Gal f residues to the rhamnopyranose residue in the linker disaccharide while the remaining residues are added by a second transferase, GlfT2. [17-19] Therefore,

GltT1 possesses dual $\text{Gal}f\text{-}\beta\text{-(1}\rightarrow\text{4)-Rhap}$ and $\text{Gal}f\text{-}\beta\text{-(1}\rightarrow\text{5)-Gal}f$ transferase activity while GltT2 affects both $\text{Gal}f\text{-}\beta\text{-(1}\rightarrow\text{5)-Gal}f$ and $\text{Gal}f\text{-}\beta\text{-(1}\rightarrow\text{6)-Gal}f$ transferase activities. As the donor species, both enzymes use UDP-galactofuranose (UDP-Gal*f*), and the acceptor is the nascent biopolymer.

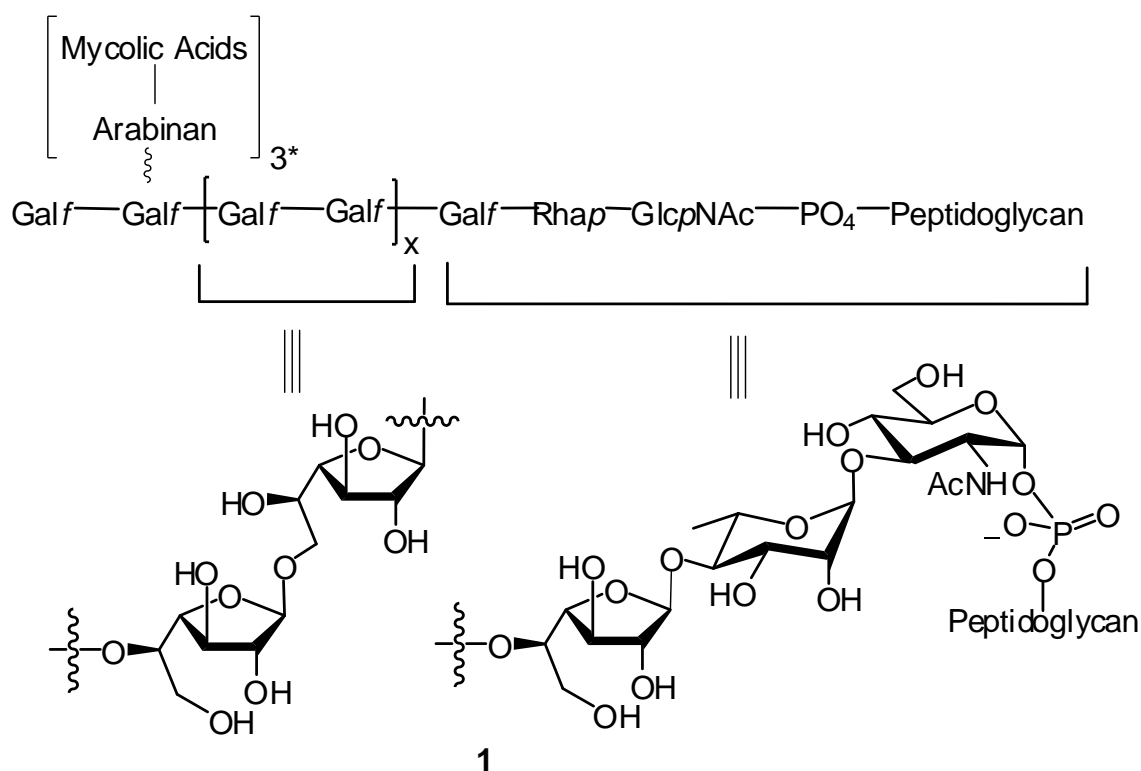


Figure 2.1. Structure of the mAGP complex, with the galactan region highlighted; $x \sim 10$. *The three arabinan chains have been proposed to be linked via the eighth, tenth and twelfth Gal*f* residues of the galactan core.^[6]

Studies to date indicate that the $\beta\text{-(1}\rightarrow\text{6)}$, $\beta\text{-(1}\rightarrow\text{5)}$ - and $\beta\text{-(1}\rightarrow\text{5)}$, $\beta\text{-(1}\rightarrow\text{6)}$ -linked trisaccharides **2** and **3**, respectively (Figure 2.2), are efficiently galactosylated by GltT2.^[20] The former is initially a substrate for the $\beta\text{-(1}\rightarrow\text{5)}$ -

transferase activity of the enzyme while the latter is a substrate for the β -(1 \rightarrow 6)-transferase activity. Because the tetrasaccharide products formed from **2** and **3** are also substrates for the enzyme, subsequent additions of Galf via the UDP-Galf donor^[20] lead to a contiguous series of products. In the organism, the entire galactan structure is assembled on a polyprenol carrier before it is arabinosylated, mycolylated and finally transferred to peptidoglycan. A recent paper by Kiessling and coworkers has proposed that the ability of GlfT2 to efficiently elongate the galactan depends upon the identity of the lipid carrier. ^[21]

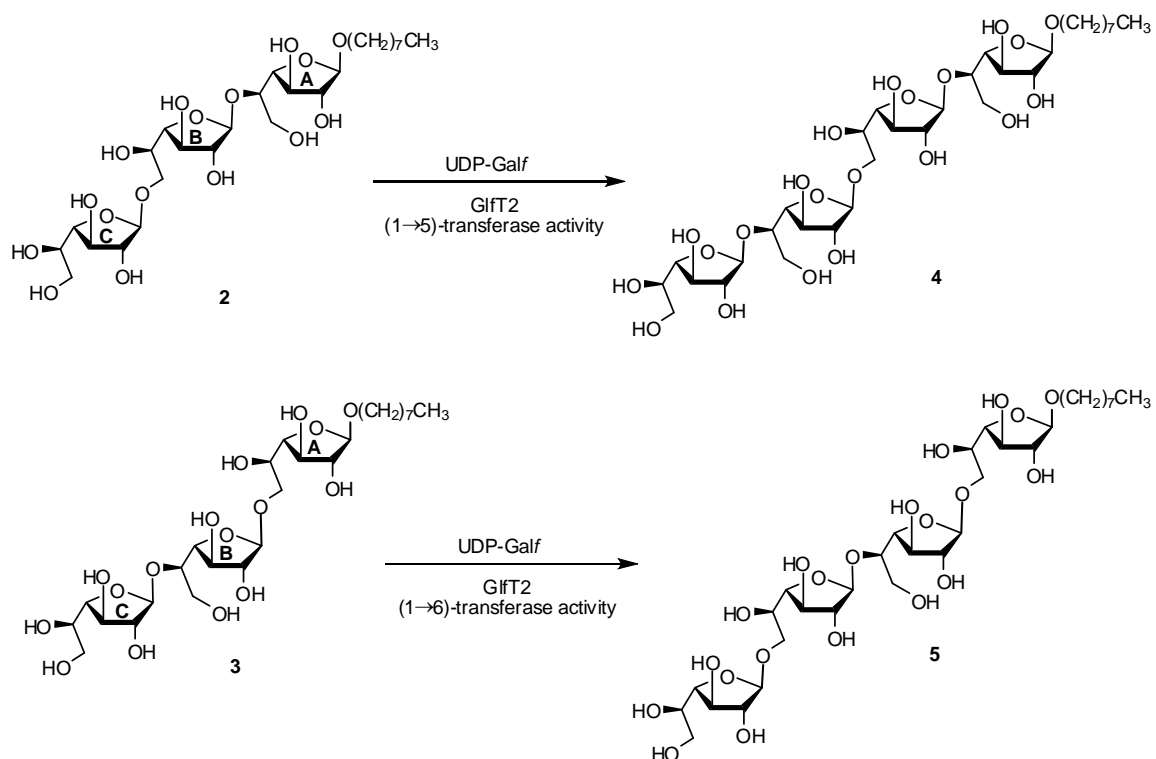


Figure 2.2. Trisaccharides **2** and **3**, reported acceptor substrates for GlfT2.

We have recently explored the bifunctionality of GlfT2,^[22] in particular, we determined that one active site appears to be responsible for the assembly of β -

(1→5)- and β -(1→6)-Gal β linkages. In order to gain further insight into the process by which GlfT2 polymerizes the galactan core of the mAG complex, we report here the use of saturation transfer difference (STD) NMR experiments with acceptor substrates **2** and **3** (Figure 2.2) to obtain their epitope maps, i.e., the nature of the contacts within the enzyme active site. No crystal structures of the enzyme, either unliganded or in complex with a substrate or substrate analogue, have been reported to date.

2.4 Experimental

2.4.1 Materials

Trisaccharides **2** and **3** were synthesized as described elsewhere.^[23]

2.4.2 Preparation of GlfT2 for NMR studies

Expression and purification of GlfT2 has been previously described.^[20] Fractions from a Sephacryl S-100 HR column were evaluated by both SDS-PAGE and a radiochemical activity assay¹⁸ (typical specific activities were 4–6 U/mg). Fractions that were both 90% pure and active were pooled. The concentration (mg/mL) of pooled fractions was determined by UV absorption spectroscopy (A_{280} divided by 1.6, the extinction coefficient). GlfT2 (2 mg) was concentrated to ~0.5 mL using an Amicon Ultra-15 centrifuge tube with a molecular weight cut off of 30,000 Da (Millipore, Billerica, MA). To this solution was added 5 mL of phosphate buffered saline solution (50 mM Na_2HPO_4 – NaH_2PO_4 , 0.1M NaCl, 10 mM MgCl_2 , 99% D_2O ; the pH of the starting buffer was 7.6) and the solution was concentrated as described above to ~0.5 mL. This

procedure was then repeated five times after which the concentration, activity and purity were then again confirmed as outlined above.

2.4.3 NMR Spectroscopy

To a sample of GlfT2 (1.6 mg) in phosphate buffered saline solution (50 mM Na₂HPO₄–NaH₂PO₄, 0.1 M NaCl, 10 mM MgCl₂ and 99% D₂O (pH 7.6)) was added either β-(1→6), β-(1→5) trisaccharide **2** (1.3 mg) or β-(1→5), β-(1→6) trisaccharide **3** (1.2 mg). Like many glycosyltransferases, GlfT2 is a metal-dependent enzyme and manganese Mn⁺² is likely the natural metal ion. The metal stabilizes the binding of the sugar nucleotide donor to the protein. The enzyme is equally efficient with Mg⁺² and that is the metal ion we have used in all our assays. The final ligand concentration was 4 mM at a ligand–protein ratio of 100:1. STD NMR spectra with UDP-Galf were acquired with 0.5 mg of UDP-Galf and 1.2 mg GlfT2. Incubation of GlfT2 (1.25 mg) with UDP-Galf (1.0 mg) and **2** (1.0 mg) generated the β-(1→5), β-(1→6), β-(1→5) tetrasaccharide product. The enzyme was recycled using Centricon preparation and incubated with UDP-Galf (1.0 mg) and **3** (1.0 mg) to generate the β-(1→6), β-(1→5), β-(1→6) tetrasaccharide product. Ligand resonances were assigned using ¹H–¹H COSY, ¹H–¹H TOCSY and ¹H–¹H NOESY NMR spectroscopy. Linkages between Galf residues were assigned on the basis of ¹H–¹H NOESY NMR spectra. Thus, interglycosidic NOEs were observed between H-1A and OCH₂, H-1C and H-6B as well as between H-1B and H-5A, confirming the Galf ring connectivities. Water suppression using presaturation was utilized in experiments with ligand only. ³¹P NMR spectra (at 162 MHz) were recorded on a Bruker AMX-400 NMR

spectrometer. STD-NMR spectra were performed on a Bruker AMX-600 NMR spectrometer at 285 K (to slow hydrolysis) in the case of UDP-Galf. STD-NMR spectra were performed on a Bruker Avance 600 NMR spectrometer, equipped with a TCI cryoprobe at 298 K in the case of the trisaccharides **2** and **3**.

Spectra were recorded with 1024 or 2048 scans and selective saturation of protein resonances at 10 ppm (30 ppm for off resonance spectra) using a series of 40 Gaussian-shaped pulses (50 ms, 1 ms delay between pulses, $\gamma B_1/2\pi = 110$ Hz), for a total saturation time of 2.04 s.²⁴ The protein resonances were broad and had significant intensity in the region downfield from 10 ppm. Irradiation at 10 ppm will result in saturation of aromatic protein resonances and via rapid spin diffusion; this saturation will also spread to aliphatic protein resonances. Irradiation at 10 ppm was also considered prudent in achieving selective saturation of the protein resonances since a ligand resonance was present at 0.8 ppm. Subtraction of saturated spectra from reference spectra was performed by phase cycling, on a Bruker AMX-600 NMR spectrometer. Measurement of enhancement intensities was performed by direct comparison of STD NMR spectra and reference one-dimensional ¹H NMR spectra. In the case of STD NMR spectra acquired on the Bruker Avance 600 NMR spectrometer, the saturated and reference spectra were acquired simultaneously by creating a pseudo-2D experiment. The STD spectrum was obtained by subtraction of saturated spectra from reference spectra after identical processing and phasing. In all cases, the fractional STD effect was calculated by $(I_0 - I_{\text{sat}})/I_0$, where $(I_0 - I_{\text{sat}})$ is the peak intensity in the STD spectrum and I_0 is the peak intensity of an

unsaturated reference spectrum. 1D-STD NMR spectra of **2** and **3** were deconvoluted to obtain peak intensities. Carefully phased and baseline-corrected spectra were deconvoluted using a program written by D. Brouwer.^[25] On- and off-resonance STD spectra were fit using mixed Lorentzian-Gaussian peaks with peak positions and widths held constant, and only peak intensity varied (see, for example, Figure S2.6). In all cases, STD spectra were acquired without water suppression.

STD-2D-TOCSY spectra were recorded with 16 or 32 scans per t_1 increment. A total of 256 t_1 increments were collected in an interlaced mode for the on- and off- resonance spectra. An MLEV-17 spin-lock sequence with a 10 kHz rf field strength and a mixing time of 80 ms was utilized. Rows were extracted from the STD-2D-TOCSY spectra and from the 2D-TOCSY spectra corresponding to spin systems from rings A, B, and C of trisaccharides **2** and **3** in the presence of GlfT2. Intensities ($I_{STD-2D-TOCSY}$) from the rows of the STD-2D-TOCSY spectra were referenced to the intensities ($I_{2D-TOCSY}$) from the rows of the 2D-TOCSY spectra and normalized to the selectively excited resonance. All enhancements were measured in triplicate, and the average values are shown. The error is estimated at $\pm 10\%$.

Longitudinal relaxation times T1s of **2** and **3** in the presence of GlfT2 were determined with the inversion recovery pulse sequence.

Data processing was performed using XWINNMR (Bruker), TopSpin (Bruker), and MestReC.

2.5 Results

2.5.1 Epitope mapping of donor substrate UDP-Galf

The binding of uridine 5'-diphosphate α -D-galactofuranose (UDP-Galf) to GlfT2 was investigated by the use of one-dimensional STD NMR experiments (Figure 2.3). The line broadening observed is caused by ligand resonances being in close contact with the protein; the slow tumbling rate of the protein-ligand complex increases the rate of relaxation, as indicated in our previous work.^[26] All of the protons of UDP-Galf show some degree of enhancement, demonstrating that the entire molecule is bound in the active site of GlfT2. The largest amount of saturation transfer was observed for H-1r of the ribosyl (r) moiety and H-5u from the uracil (u) moiety, indicating that both the uracil and ribose units are making more intimate contacts with groups on the protein (Figure 2.4).

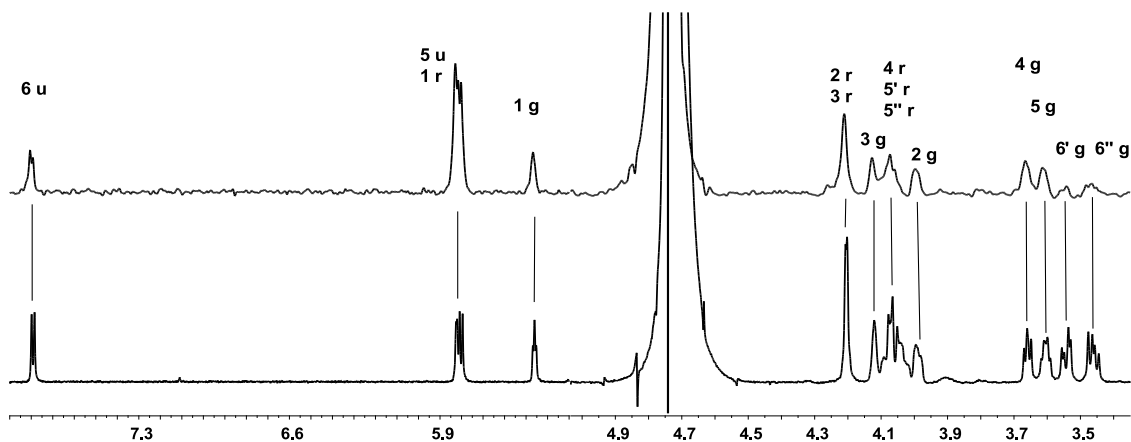


Figure 2.3. Expansion of 1D ^1H NMR (lower trace) and STD NMR (upper trace) spectra of UDP-Galf in the presence of GlfT2, at 600 MHz and 285K. Labels g, r and u refer to Galf, ribose, and uracil, respectively.

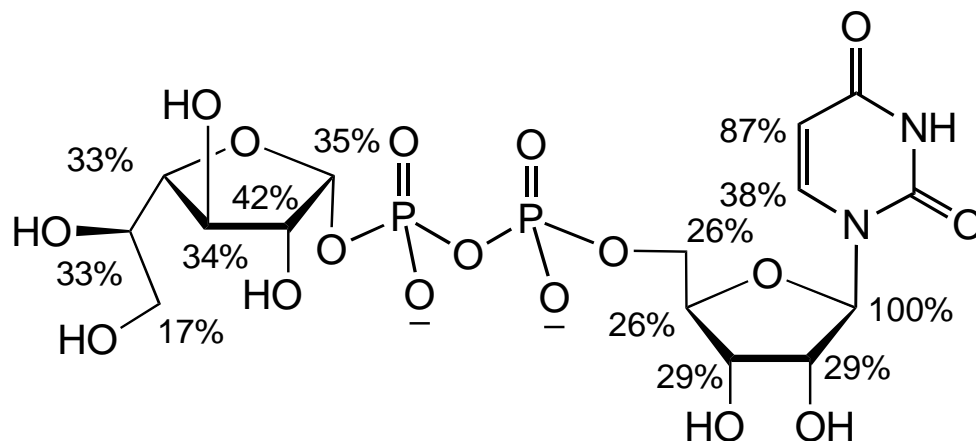


Figure 2.4. Epitope mapping of UDP-Galf in the active site of GIfT2. Enhancements are referenced to the H-1 resonance of the ribofuranosyl moiety.

Within 5–6 hours, the UDP-Galf began to hydrolyze to uridine 5'-monophosphate (UMP), uridine 5'-diphosphate (UDP), Galf-1-phosphate, and galactopyranose. ^{31}P NMR spectroscopy of the reaction products showed signals from UDP and UMP, as well as a signal from inorganic phosphate (from UDP hydrolysis). UDP and UMP were also observed to bind to GIfT2 (See Figures S2.1, S2.2 in Supporting Information).

2.5.2 Epitope mapping of the acceptor substrate 2

The acceptor substrates **2** and **3** were examined individually for binding to GIfT2. One-dimensional STD NMR spectra were obtained for **2** (Figure 2.5). Epitope mapping revealed that the entire molecule was bound in the active site of

GlfT2 and that the three Galf monosaccharide units showed varying degrees of enhancement (Figure 2.6).

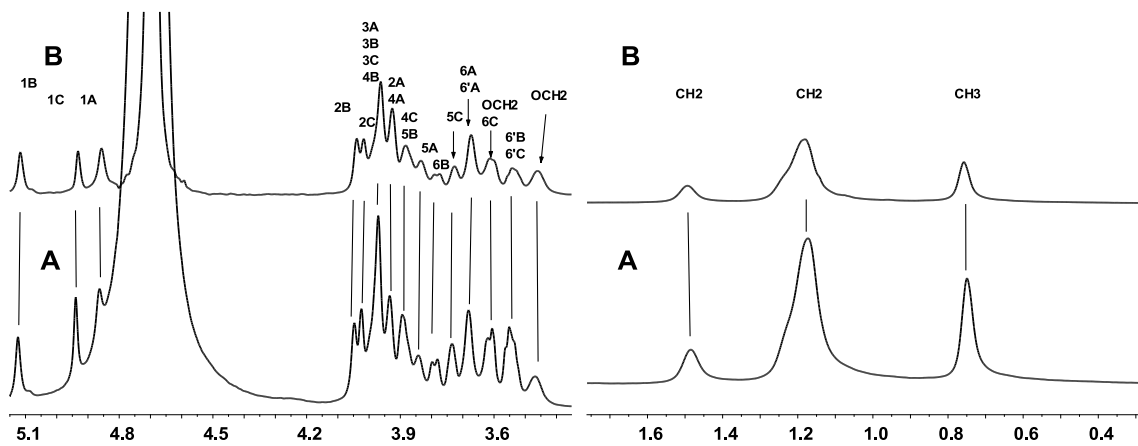


Figure 2.5. Expansion of (A) 1D ¹H NMR and (B) STD NMR spectra of **2** at 600 MHz and 298K in the presence of GlfT2.

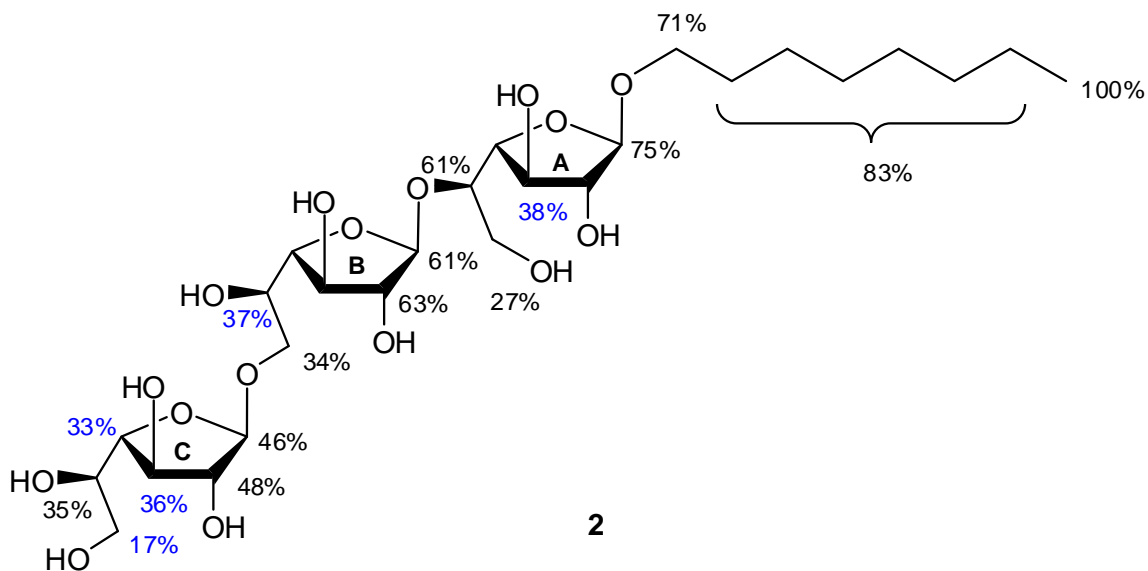


Figure 2.6. Epitope mapping of **2** in the active site of GlfT2. Enhancements are referenced to the methyl resonance of the octyl aglycon. Estimated enhancements from STD-2D-TOCSY NMR data are shown in blue.

Because only one type of monosaccharide residue (D-Galf) is present in these oligosaccharides, spectral overlap was significant and this complicated analysis. For example, the overlapping resonance at 3.96 ppm comprised the H-3B, H-3C, H-3A and H-4B signals. We attempted to estimate the contributions for each proton from STD-2D-TOCSY NMR spectra as this afforded increased spectral dispersion. (see, for example, Figure S2.7).

Accordingly, a row was extracted from the STD-2D-TOCSY spectra corresponding to H-1A and its spin coupled partners, H-2A, H-3A, H-4A etc. This row was referenced to the same row from the 2D-TOCSY spectra and normalized to the selectively excited resonance (H-1A in this case) indicated in

bold (See Table S2.4). In this manner, the STD contribution from H-3A relative to H-1A could be estimated from measuring the enhancement of H-3A (52%) relative to the enhancement of H-1A (100%) in that row (See Table S2.4). Since the STD enhancement of H-1A is known from the 1D-STD NMR spectrum (74.5%) (See Table S2.1), we can estimate the enhancement of H-3A to be 38%. We cannot assign contributions from H-2A and H-4A as these resonances are isochronous and inseparable even in the 2D spectra.

Next, we examined the contribution from H-5B as this forms part of the multiplet at 3.88 ppm (H-4C/5B). The STD contribution from H-5B relative to H-6B was estimated from a row of the STD-2D-TOCSY spectrum referenced to the same row in the 2D-TOCSY spectrum which contained H-6B and its coupled partners, H-6'B, H-5B, H-4B etc. From this row, the enhancement of H-5B (109%) relative to the enhancement of H-6B (100%) was measured (See Table S2.4). Since the STD enhancement of H-6B is known from the 1D-STD NMR spectrum (34%) (see Table S2.1), we can estimate the enhancement of H-5B to be 37%. We cannot assign contributions from H-3B and H-4B as these resonances are isochronous and inseparable even in the 2D spectra.

Lastly, we examined the contribution from H-3C, H-4C and H-6C. Once again we referred to a row from both the STD-2D-TOCSY data and 2D-TOCSY data (See Table S2.4). Since the STD enhancement of H-5C (35%) is known from the 1D-STD NMR spectrum (See Table S2.1), we can estimate the enhancements of H-3C, H-4C and H-6C to be 36%, 33%, and 17%, respectively from their ratios relative to H-5C in the STD-2D spectral data.

The epitope map (Figure 2.6) indicates that the *Galf A* residue makes the most intimate contacts with the protein, residue *Galf B* makes lesser contacts, and *Galf C* has least contacts with the protein.

An examination of the binding epitope of **2** in the presence of the donor, UDP-*Galf* is also of interest. Thus, a sample containing UDP-*Galf* and GlfT2 was treated with trisaccharide **2**, and the STD amplification factors for UDP-*Galf* were calculated. The resonance from UDP-*Galf* corresponding to H-6 uracil was not dramatically affected by the addition of trisaccharide **2**, consistent with the epitope map (Figure 2.4), which indicates that it makes sparse contact with GlfT2. The STD amplification factor for the H-1r/H-5u resonance of ribose decreased by approximately 58% and that of the H-1g resonance of galactose decreased by approximately 61%. The resonance corresponding to H-1 ribose/H-5 uracil from UDP-*Galf* was most affected because H-1 of ribose and H-5 of uracil are critical for binding to GlfT2. During these experiments, we observed the formation of the β -(1 \rightarrow 5), β -(1 \rightarrow 6), β -(1 \rightarrow 5) tetrasaccharide product, indicating that the enzyme was active and turning over (Figure S2.3 in Supporting Information); however, UDP-*Galf* hydrolysis was also observed. The STD amplification factors for **2** were not significantly affected in the presence of UDP-*Galf*. The lack of rigidification of the acceptor deserves comment. Presumably, similar STD amplification factors were obtained because product formation is fast relative to the NMR time scale. In addition, non-ideal substrates are being used, the natural substrates being large, membrane-associated glycolipids.

2.5.3 Epitope mapping of the acceptor substrate **3**

Having established the manner in which GlfT2 binds **2**, the trisaccharide **3** was investigated. As with the acceptor substrate **2**, the enhancements observed in the one-dimensional STD NMR experiments for **3** (Figure S2.4 in Supporting Information) indicated that the entire molecule was bound in the active site of GlfT2. The methodology used to estimate enhancements for acceptor **2** was also used for the case of **3** (See Table S2.5). The epitope map for **3** is shown in Figure 2.7.

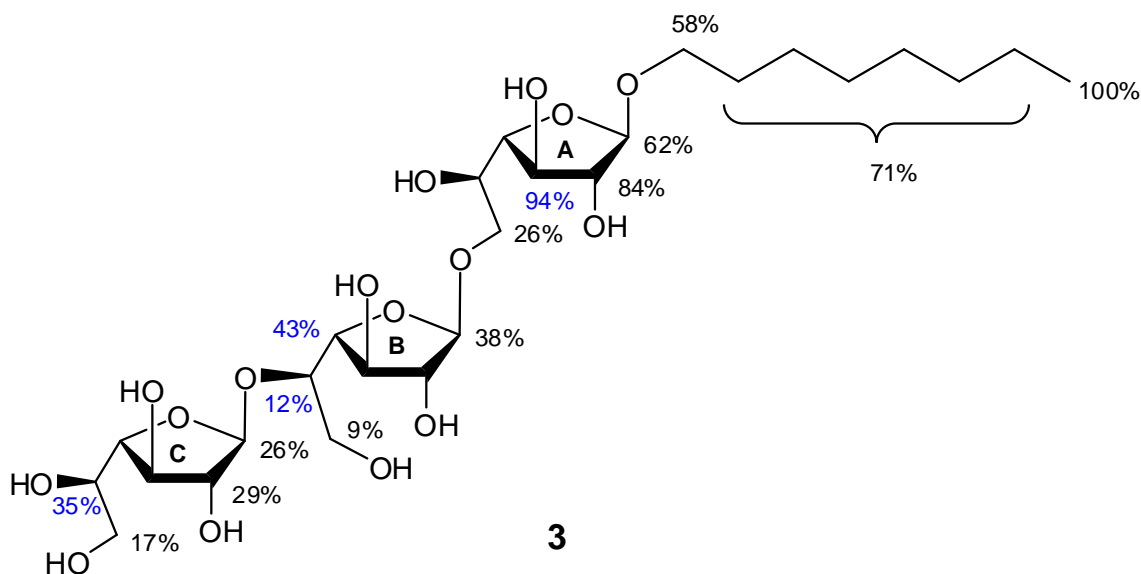


Figure 2.7. Epitope mapping of **3** in the active site of GlfT2. Enhancements are referenced to the methyl resonance of the octyl aglycon. Estimated enhancements from STD-2D-TOCSY NMR data are shown in blue.

Taken together, these epitope mapping studies demonstrate that acceptor substrates **2** and **3** both experienced greater enhancement toward ring A, indicating that this is a critical portion of the binding epitope. These findings are consistent with a recent STD-NMR study on *N*-acetylglucosaminyl transferase V with its trisaccharide acceptor substrate.^[27] In these investigations, the monosaccharide residue of the acceptor undergoing glycosylation showed lower STD effects than the other monosaccharide residues. This is presumably due to the need for this region of the substrate to be only weakly bound so that the conformational changes necessary in the glycosylation transition state can occur. Hence, it is not expected that this portion of the molecule will be tightly bound and this feature is reflected in the reduced STD enhancement.

We also examined the binding epitope of **3** in the presence of the donor, UDP-Galf. As was observed with trisaccharide **2**, it was found that the addition of UDP-Galf had little effect on the binding epitope of trisaccharide **3**. As seen with trisaccharide **2**, the formation of the β -(1 \rightarrow 6), β -(1 \rightarrow 5), β -(1 \rightarrow 6) tetrasaccharide product was observed, indicating that the enzyme was active and turning over (Figure S2.5 in Supporting Information). However, we also observed UDP-Galf hydrolysis.

2.6 Discussion

Significant insight into the process by which the galactan portion of mycobacterial mAGP is biosynthesized has been acquired in recent years.^[3-5, 12, 13, 15] It appears that the ~30-residue long glycan is assembled by the two bifunctional galactofuranosyltransferases, GlfT1 and GlfT2.^[15] Because galactan synthesis is essential for mycobacterial viability,^[28] and *Galf* appears to be absent in mammalian systems, inhibitors of these transferases could be selective therapeutic agents for the treatment of diseases such as tuberculosis and HIV-associated *M. avium* infections.^[29] It is therefore highly desirable that one gain an understanding of the structures and mechanisms of these enzymes. In this paper, we have focused on recombinant *M. tuberculosis* GlfT2.^[20, 30] In the absence of X-ray structural information on GlfT2 and its complexes, we used STD-NMR spectroscopy to investigate GlfT2 substrate interactions.

Epitope mapping studies with the nucleotide donor for the enzyme, UDP-*Galf*, revealed that the strongest interactions with the protein involve the nucleotide base and H-1 of the ribose portions of the molecule. All of the hydrogens from the *Galf* moiety also receive saturation transfer from the protein, but less than the nucleotide moiety. These trends are consistent with previous studies on the blood group B galactosyltransferase,^[31] as well as *N*-acetylglucosaminyl transferase V^[27] which also showed that the donor (UDP-*Galp* and UDP-*GalpNAc*, respectively) is bound most strongly through the nucleotide portion. This can be rationalized based on the need for the sugar that

is to be transferred to be relatively loosely bound, thus facilitating molecular reorganization near the site of reaction.

Similar epitope mapping studies with two known trisaccharide substrates for the enzyme, **2** and **3**, revealed that in both cases the non-reducing end of the molecule showed lower STD effects, indicating reduced interaction with the protein. This observation is in agreement with previous studies on the binding of carbohydrate acceptor molecules by glycosyltransferases,^[27, 31] and, as described above for the binding of UDP-Galf, is consistent with the need for the substrate to have the flexibility to participate in the enzyme-catalyzed glycosylation. If the entire acceptor and donor were tightly bound to the enzyme, it could be envisioned that the substrates would not be able to undergo the conformational distortion needed in the transition state. When these epitope mapping studies were carried out in the presence of the donor UDP-Galf, the binding epitopes were not substantially different, and under these conditions the formation of a tetrasaccharide product was observed, thus demonstrating that the protein was capable of catalysis under the conditions of the experiment.

It is also of interest to note that the octyl aglycon receives substantial saturation transfer, indicating that it is interacting strongly with the protein. This is consistent with the earlier^[32] observation that mycobacterial glycosyltransferases recognize oligosaccharides of long-chain alcohols better than those with shorter aglycons. In nature, GlfT2 normally recognizes, as a minimum acceptor substrate, a tetrasaccharide, the linker disaccharide plus two Galf residues (See Figure 2.1). Thus, the octyl group may interact with portions of the enzyme that

normally bind to these additional carbohydrate residues. This proposal is consistent with the previously mentioned work from the Kiessling group, in which the nature of the aglycon of the sugar acceptor was shown to influence the outcome of GlfT2-mediated galactan elongation.^[21]

In summary, these STD-NMR studies have provided additional information on substrate binding by GlfT2. When taken together with our earlier work,^[22] the picture that emerges is one in which this bifunctional enzyme carries out both glycosylation reactions within a single active site, and binds its substrates in a manner similar to other glycosyltransferases.^[27, 31]

2.7 Acknowledgements

This work was supported by The Alberta Ingenuity Centre for Carbohydrate Science and The Natural Sciences and Engineering Research Council of Canada. We thank A. R. Lewis for performing the deconvolutions, R. Zhou for technical assistance, and D. Bleile for helpful discussions.

2.8 Summary

We have described saturation-transfer difference (STD) NMR studies of GlfT2 using two trisaccharide acceptor substrates, β -D-Galf-(1 \rightarrow 6)- β -D-Galf-(1 \rightarrow 5)- β -D-Galf-O(CH₂)₇CH₃ (**2**) and β -D-Galf-(1 \rightarrow 5)- β -D-Galf-(1 \rightarrow 6)- β -D-Galf-O(CH₂)₇CH₃ (**3**), as well as the donor substrate for the enzyme, UDP-Galf. We also described how epitope mapping demonstrated a greater enhancement toward the “reducing” ends of both trisaccharides, and that UDP-galactofuranose (UDP-Galf) made more contacts to GlfT2 through its nucleotide moiety. With this information in hand, we were now in a position to investigate whether GlfT2 catalyzes the trisaccharide substrates using one or two active sites, which forms the basis of the studies described in Chapter 3.

2.9 References

- [1] T. M. Daniel, *Respiratory Med.* **2006**, *100*, 1862-1870.
- [2] <http://www.who.int/tb/en/>
- [3] S. Mahapatra, J. Basu, P. J. Brennan, D. C. Crick, in *Tuberculosis and the Tubercle Bacillus* (Eds.: S. T. Cole, K. D. Eisenach, D. N. McMurray, J. Jacobs, W. R.), American Society for Microbiology, Washington, DC, **2005**, pp. 275-285.
- [4] T. L. Lowary, in *Glycoscience: Chemistry and Chemical Biology, Vol. 3* (Eds.: B. O. Fraser-Reid, K. Tatsuta, J. Thiem), Springer-Verlag, Berlin, **2001**, pp. 2005-2080.
- [5] G. S. Besra, K. H. Khoo, M. R. McNeil, A. Dell, H. R. Morris, P. J. Brennan, *Biochemistry* **1995**, *34*, 4257-4266.
- [6] L. J. Alderwick, E. Radmacher, M. Seidel, R. Gande, P. G. Hitchen, H. R. Morris, A. Dell, H. Sahm, L. Eggeling, G. S. Besra, *J. Biol. Chem.* **2005**, *280*, 32362-32371.
- [7] R. M. de Lederkremer, W. Colli, *Glycobiology* **1995**, *5*, 547-552.

- [8] S. M. Beverley, K. L. Owens, M. Showalter, C. L. Griffith, T. L. Doering, V. C. Jones, M. R. McNeil, *Eukaryot. Cell* **2005**, *4*, 1147-1154.
- [9] W. Morelle, M. Bernard, J. P. Debeaupuis, M. Buitrago, M. Tabouret, J. P. Latge, *Eukaryot. Cell* **2005**, *4*, 1308-1316.
- [10] J. B. Houseknecht, T. L. Lowary, *Curr. Opin. Chem. Biol.* **2001**, *5*, 677-682.
- [11] L. L. Pedersen, S. J. Turco, *Cell. Molec. Life Sci.* **2003**, *60*, 259-266.
- [12] S. Berg, D. Kaur, M. Jackson, P. J. Brennan, *Glycobiology* **2007**, *17*, 35R-56R.
- [13] S. Bhamidi, M. S. Scherman, C. D. Rithner, J. E. Prenni, D. Chatterjee, K. H. Khoo, M. R. McNeil, *J. Biol. Chem.* **2008**, *283*, 12992-13000.
- [14] P. H. Tam, T. L. Lowary, *Curr. Opin. Chem. Biol.* **2009**, *13*, 618-625.
- [15] M. Belanova, P. Dianiskova, P. J. Brennan, G. C. Completo, N. L. Rose, T. L. Lowary, K. Mikusova, *J. Bacteriol.* **2008**, *190*, 1141-1145.

- [16] K. Mikusova, M. Belanova, J. Kordulakova, K. Honda, M. R. McNeil, S. Mahapatra, D. C. Crick, P. J. Brennan, *J. Bacteriol.* **2006**, *188*, 6592-6598.
- [17] K. Mikusova, T. Yagi, R. Stern, M. R. McNeil, G. S. Besra, D. C. Crick, P. J. Brennan, *J. Biol. Chem.* **2000**, *275*, 33890-33897.
- [18] L. Kremer, L. G. Dover, C. Morehouse, P. Hitchin, M. Everett, H. R. Morris, A. Dell, P. J. Brennan, M. R. McNeil, C. Flaherty, K. Duncan, G. S. Besra, *J. Biol. Chem.* **2001**, *276*, 26430-26440.
- [19] L. J. Alderwick, L. G. Dover, N. Veerapen, S. S. Gurcha, L. Kremer, D. L. Roper, A. K. Pathak, R. C. Reynolds, G. S. Besra, *Prot. Expr. Purif.* **2008**, *58*, 332-341.
- [20] N. L. Rose, G. C. Completo, S. J. Lin, M. McNeil, M. M. Palcic, T. L. Lowary, *J. Am. Chem. Soc.* **2006**, *128*, 6721-6729.
- [21] J. F. May, R. A. Splain, C. Brotschi, L. L. Kiessling, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 11851-11856.
- [22] M. G. Szczepina, R. B. Zheng, G. C. Completo, T. L. Lowary, B. M. Pinto, *ChemBioChem* **2009**, *10*, 2052-2059.
- [23] G. C. Completo, T. L. Lowary, *J. Org. Chem.* **2008**, *73*, 4513-4525.
- [24] M. Mayer, B. Meyer, *Angew. Chem., Int. Ed.* **1999**, *38*, 1784-1788.

- [25] C. A. Fyfe, D. H. Brouwer, *J. Am. Chem. Soc.* **2006**, *128*, 11860-11871.
- [26] T. Weimar, B. Stoffer, B. Svensson, B. M. Pinto, *Biochemistry* **2000**, *39*, 300-306.
- [27] M. A. Macnaughtan, M. Kamar, G. Alvarez-Manilla, A. Venot, J. Glushka, J. M. Pierce, J. H. Prestegard, *J. Mol. Biol.* **2007**, *366*, 1266-1281.
- [28] F. Pan, M. Jackson, Y. F. Ma, M. McNeil, *J. Bacteriol.* **2001**, *183*, 3991-3998.
- [29] B. C. de Jong, D. M. Israelski, E. L. Corbett, P. M. Small, *Annu. Rev. Med.* **2004**, *55*, 283-301.
- [30] N. L. Rose, R. B. Zheng, J. Pearcey, R. Zhou, G. C. Completo, T. L. Lowary, *Carbohydr. Res.* **2008**, *343*, 2130-2139.
- [31] J. Angulo, B. Langpap, A. Blume, T. Biet, B. Meyer, N. R. Krishna, H. Peters, M. M. Palcic, T. Peters, *J. Am. Chem. Soc.* **2006**, *128*, 13529-13538.
- [32] R. E. Lee, P. J. Brennan, G. S. Besra, *Glycobiology* **1997**, *7*, 1121-1128.

2.10 Supporting Information

STD-NMR Studies of Two Acceptor Substrates of GltT2, a Galactofuranosyltransferase from Mycobacterium tuberculosis: Epitope Mapping Studies

Monica G. Szczepina,^a Ruixiang B. Zheng,^b Gladys C. Completo,^b

Todd L. Lowary,^b and B. Mario Pinto^a

^aDepartment of Chemistry, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada

^bDepartment of Chemistry and Alberta Ingenuity Centre for Carbohydrate Science, Gunning-Lemieux Chemistry Centre, University of Alberta, Edmonton, Alberta, T6G 2G2, Canada

Table of Contents

Table S2.1.	Deconvoluted STD-NMR peak areas (at 298 K) and T1 values (at 285 K and 298 K) of 2 in the presence of G1fT2.	73
Table S2.2.	Deconvoluted STD-NMR peak areas (at 298 K) and T1 values (at 285 K and 298 K) of 3 in the presence of G1fT2	74
Table S2.3.	STD-NMR intensities of UDP-Galf at 285 K in the presence of G1fT2.	75
Table S2.4.	STD-2D-TOCSY intensities relative to 2D-TOCSY intensities of 2 in the presence of G1fT2.	76
Table S2.5.	STD-2D-TOCSY intensities relative to 2D-TOCSY intensities of 3 in the presence of G1fT2.	77
Figure S2.1.	Expansion of 1D ¹ H NMR and STD NMR spectra of a mixture of UDP, UMP, galactose-1-phosphate and galactose following hydrolysis of UDP-Galf in the presence of G1fT2.	78
Figure S2.2.	Expansion of 1D ¹ H NMR and STD NMR spectra of the time course of the UDP-Galf hydrolysis reaction in the presence of G1fT2.	79
Figure S2.3.	Expansion of 1D ¹ H NMR spectrum of a mixture of trisaccharide 2 and UDP-Galf in the presence of G1fT2.	80
Figure S2.4.	Expansion of 1D ¹ H NMR and STD NMR spectra of trisaccharide 3 in the presence of G1fT2.	81
Figure S2.5.	Expansion of 1D ¹ H NMR spectrum of a mixture of trisaccharide 3 and UDP-Galf in the presence of G1fT2.	82

Figure S2.6.	Deconvolution of the STD-NMR spectrum (3.2-4.2 ppm region) of 2 in the presence of GlfT2 at 298 K.	83
Figure S2.7.	STD-2D-TOCSY intensity relative to the 2D-TOCSY intensity for the selectively excited resonance H-6B of 2 in the presence of GlfT2.	84

Table S2.1. Deconvoluted STD-NMR peak areas (298 K) and T1s (285 K and 298 K) of **2** in the presence of GlfT2.^a

Chemical Shift δ	Deconvoluted Area (Relative Percentage)		T1(s) ^{298K}	T1(s) ^{285K}
5.12 (H-1B)	0.144	(61%)	1.1	1.2
4.93 (H-1C)	0.109	(46%)	1.2	—
4.85 (H-1A)	0.176	(75%)	1.0	—
4.04 (H-2B)	0.148	(63%)	1.6	1.3
4.02 (H-2C)	0.114	(48%)	1.8	1.5
3.96 (H-3A/3B/3C/4B)	0.122	(52%)	—	—
3.93 (H-2A/4A)	0.183	(77%)	—	—
3.88 (H-4C/5B)	0.102	(43%)	—	—
3.84 (H-5A)	0.145	(61%)	1.1	0.94
3.79 (H-6B)	0.080	(34%)	0.74	0.70
3.73 (H-5C)	0.083	(35%)	1.3	1.1
3.67 (H-6A/6'A)	0.127	(54%)	—	—
3.61 (H-6C/octyl-OCH)	0.099	(42%)	—	—
3.54 (H-6'C/6'B)	0.066	(28%)	—	—
3.46 (octyl-OCH)	0.169	(71%)	0.94	0.71
1.48 (octyl-CH ₂)	0.175	(74%)	—	—
1.18 (octyl-CH ₂)	0.202	(85%)	—	—
0.75 (octyl-CH ₃)	0.237	(100%)	1.1	0.91

^aDeconvoluted peak areas from the STD-NMR spectrum (A_{STD}) were referenced to the deconvoluted peak areas from the off-resonance spectrum (A_0) and normalized to the methyl resonance of the octyl unit. The deconvoluted areas are quoted as percentages. Irradiation is at $\delta = 10$ ppm. Error is $\pm 10\%$.

Table S2.2. Deconvoluted STD-NMR peak areas (298 K) and T1s (285 K and 298 K) of **3** in the presence of GlfT2.^a

Chemical Shift δ	Deconvoluted Area (Relative Percentage)		T1(s) ^{298K}	T1(s) ^{285K}
5.10 (H-1C)	0.049	(26%)	1.3	1.3
4.89 (H-1B)	0.073	(38%)	1.2	—
4.87 (H-1A)	0.117	(62%)	1.2	—
4.03 (H-2C)	0.055	(29%)	1.8	1.7
4.00 (H-2B/3B)	0.061	(32%)	1.5	1.4
3.96 (H-3A/3C/4B/4C)	0.062	(33%)	—	—
3.92 (H-2A)	0.159	(84%)	1.5	1.2
3.85 (H-4A/5A/5B)	0.075	(40%)	—	—
3.73 (H-6'A/5C)	0.045	(24%)	—	—
3.68 (H-6B/6'B)	0.033	(17%)	—	—
3.62 (H-6'C/octyl-OCH)	0.064	(34%)	—	—
3.57 (H-6C)	0.032	(17%)	0.73	0.68
3.49 (H-6A)	0.050	(26%)	0.68	0.73
3.46 (octyl-OCH)	0.110	(58%)	0.82	0.78
1.48 (octyl-CH ₂)	0.108	(58%)	—	—
1.18 (octyl-CH ₂)	0.140	(73%)	—	—
0.75 (octyl-CH ₃)	0.190	(100%)	1.1	0.98

^aDeconvoluted peak areas from the STD-NMR spectrum (A_{STD}) were referenced to the deconvoluted peak areas from the off-resonance spectrum (A_0) and normalized to the methyl resonance of the octyl unit. The deconvoluted areas are quoted as percentages. Irradiation is at $\delta = 10$ ppm. Error is $\pm 10\%$.

Table S2.3. STD-NMR intensities of UDP-Galf at 285 K in the presence of GfT2.^a

<u>Chemical Shift δ</u>	<u>Enhancement (Relative Percentage)</u>
7.80 (H-6u)	0.246 (38%)
5.83 (H-1r)	0.656 (100%)
5.81 (H-5u)	0.5675 (87%)
5.47 (H-1g)	0.229 (35%)
4.21 (H-2r/H-3r)	0.189 (29%)
4.13 (H-3g)	0.224 (34%)
4.07 (H-4r/H-5'r/H-5''r)	0.169 (26%)
4.00 (H-2g)	0.275 (42%)
3.67 (H-4g)	0.214 (33%)
3.61 (H-5g)	0.215 (33%)
3.55 (H-6'g)	0.112 (17%)
<u>3.47 (H-6''g)</u>	<u>0.114 (17%)</u>

^aIntensities (I_{STD}) were referenced to a normal spectrum (I_0) and normalized to the H-1 resonance of the ribosyl unit. The intensities are quoted as percentages. Irradiation is at $\delta = 0$ ppm. Enhancements were measured in triplicates and the average is shown. Error is $\pm 10\%$.

Table S2.4. STD-2D-TOCSY intensities relative to 2D-TOCSY intensities of **2** in the presence of GlfT2.^a Each table refers to a specific spin system (i.e ring A, ring B and ring C) and corresponds to rows extracted from the 2D spectra.

<u>Chemical Shift δ</u>	<u>Enhancement (Relative Percentage)</u>
4.85 (H-1A)	(100%)
3.96 (H-3A)	(52%)
<u>3.93 (H-2A/4A)</u>	<u>(95%)</u>

<u>Chemical Shift δ</u>	<u>Enhancement (Relative Percentage)</u>
3.96 (H-3B/4B)	(111%)
3.88 (H-5B)	(109%)
3.79 (H-6B)	(100%)
<u>3.54 (H-6'B)</u>	<u>(115%)</u>

<u>Chemical Shift δ</u>	<u>Enhancement (Relative Percentage)</u>
4.02 (H-2C)	(96%)
3.96 (H-3C)	(103%)
3.88 (H-4C)	(96%)
3.73 (H-5C)	(100%)
<u>3.61-3.54 (H-6C/6'C)</u>	<u>(101%)</u>

^aIntensities ($I_{STD-2D-TOCSY}$) were referenced to a 2D TOCSY spectrum ($I_{2D-TOCSY}$) and normalized to the selectively excited resonance indicated in bold. Enhancements were measured in triplicates and the average is shown. Error is \pm 10%.

Table S2.5. STD-2D-TOCSY intensities relative to 2D-TOCSY intensities of **3** in the presence of GlfT2.^a Each table refers to a specific spin system (i.e ring A, ring B and ring C) and corresponds to rows extracted from the 2D spectra.

<u>Chemical Shift δ</u>	<u>Enhancement (Relative Percentage)</u>
4.87 (H-1A)	(100%)
3.96 (H-3A)	(152%)
<u>3.92 (H-2A)</u>	<u>(115%)</u>

<u>Chemical Shift δ</u>	<u>Enhancement (Relative Percentage)</u>
4.00 (H-2B/3B)	(149%)
3.96 (H-4B)	(253%)
3.68 (H-6B/6'B)	(100%)
<u>3.85 (H-5B)</u>	<u>(72%)</u>

<u>Chemical Shift δ</u>	<u>Enhancement (Relative Percentage)</u>
5.10 (H-1C)	(114%)
4.03 (H-2C)	(100%)
3.96 (H-3C/4C)	(135%)
<u>3.73 (H-5C)</u>	<u>(122%)</u>

^aIntensities ($I_{STD-2D-TOCSY}$) were referenced to a 2D TOCSY spectrum ($I_{2D-TOCSY}$) and normalized to the selectively excited resonance indicated in bold. Enhancements were measured in triplicates and the average is shown. Error is \pm 10%.

Figure S2.1. Expansion of 1D ^1H NMR (upper trace) and STD NMR (lower trace) spectra of a mixture of UDP, UMP, galactose-1-phosphate and galactose following hydrolysis of UDP-Galf at 600 MHz and 285 K in the presence of Glt2.

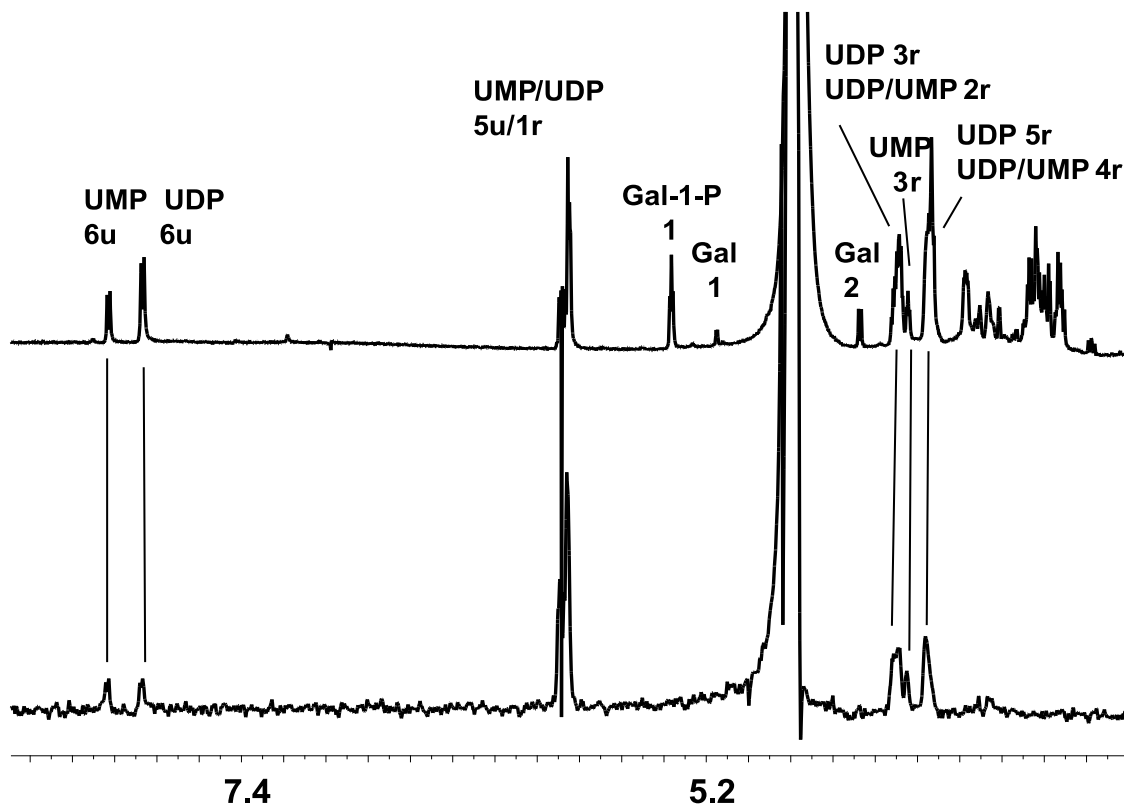


Figure S2.2. Expansion of 1D ^1H NMR (lower and middle trace) and STD NMR (upper trace) spectra of the time course of the UDP-Galf hydrolysis reaction at 600 MHz and 285K in the presence of GlfT2. The bottom trace is the 1D ^1H NMR spectrum of UDP-Galf. The middle trace is the 1D ^1H NMR spectrum of the sample after 7 h. The upper trace is the STD NMR spectrum of the sample after 7h. UDP and UMP bind preferentially over UDP-Galf.

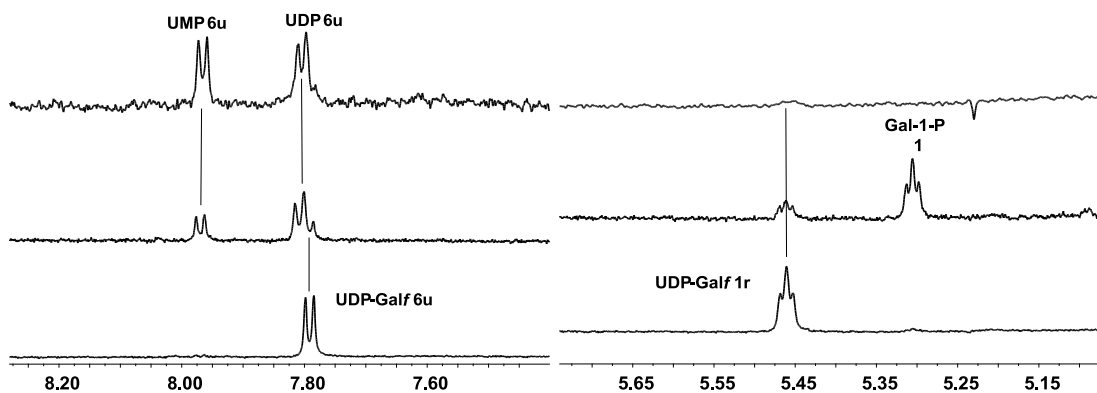


Figure S2.3. Expansion of 1D ^1H NMR spectrum of a mixture of trisaccharide **2** and UDP-Galf at 600 MHz and 298 K in the presence of GlfT2 over the course of 24 h. The asterisk denotes an anomeric signal of the formed β -(1 \rightarrow 5), β -(1 \rightarrow 6), β -(1 \rightarrow 5) tetrasaccharide (*J. Org. Chem.* **2008**, 73, 4513–4525).

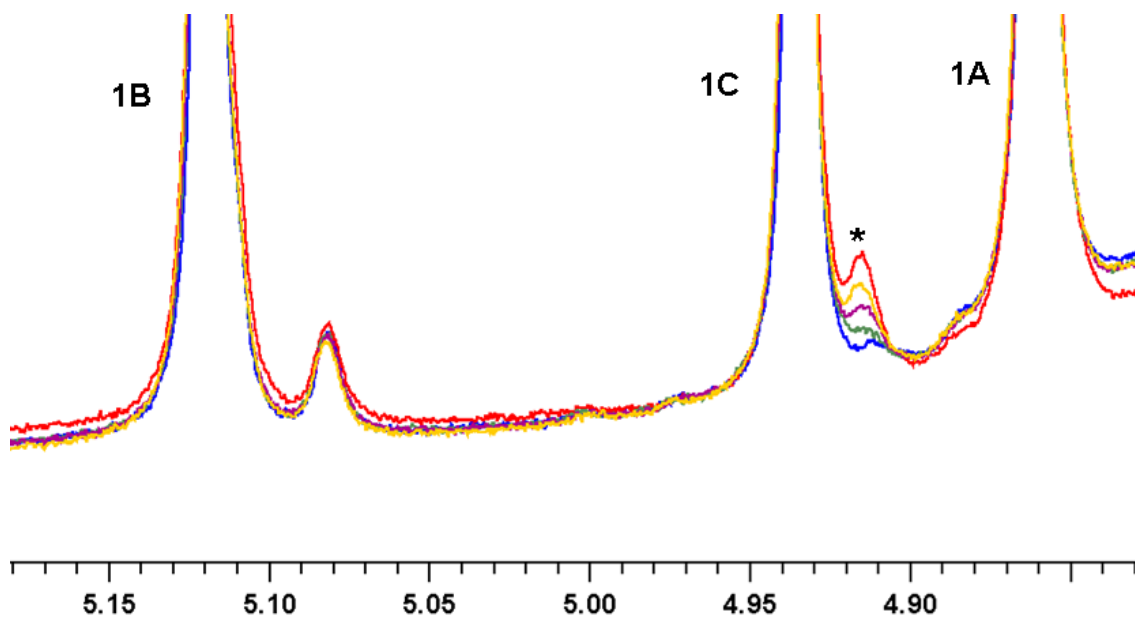


Figure S2.4. Expansion of (A) 1D ^1H NMR and (B) STD NMR spectra of **3** at 600 MHz and 298 K in the presence of GltT2.

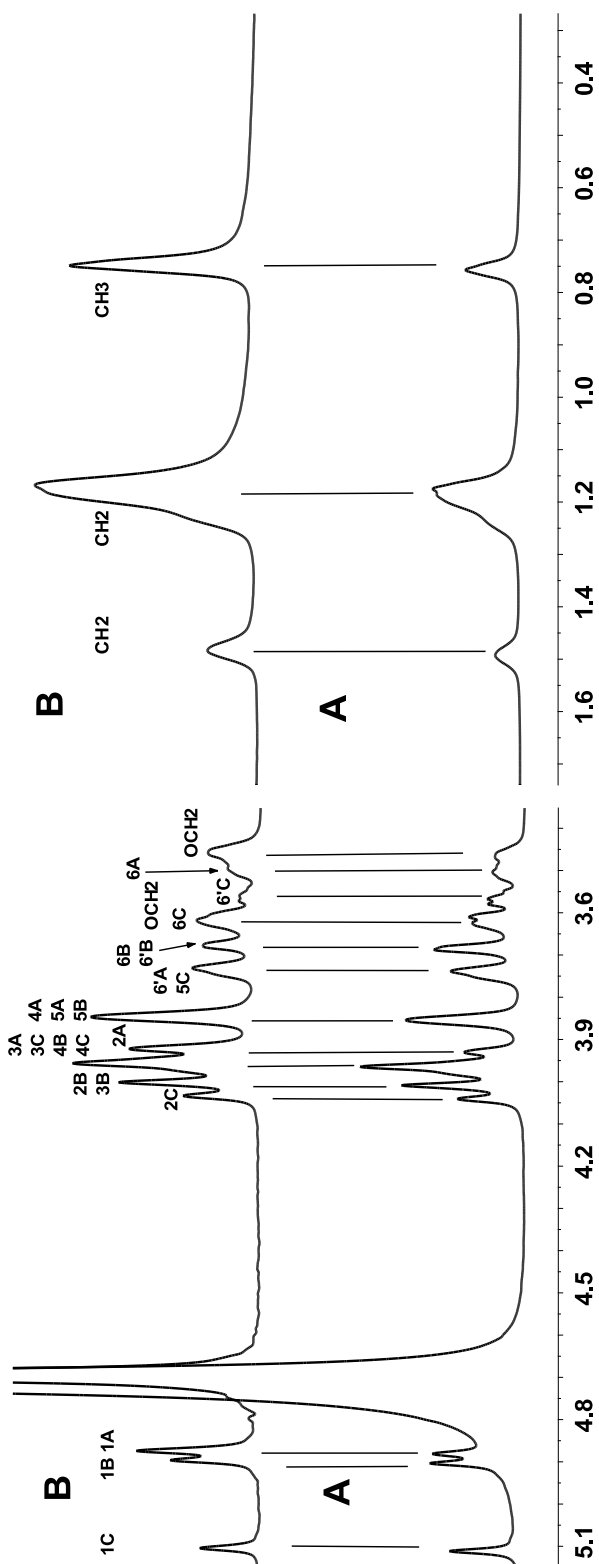


Figure S2.5. Expansion of 1D ^1H NMR spectrum of a mixture of trisaccharide **3** and UDP-Galf at 600 MHz and 310 K in the presence of GlfT2 over the course of 24h. The asterisks denote anomeric signals of the formed β -(1 \rightarrow 6), β -(1 \rightarrow 5), β -(1 \rightarrow 6) tetrasaccharide (*J. Org. Chem.* **2008**, 73, 4513–4525.)

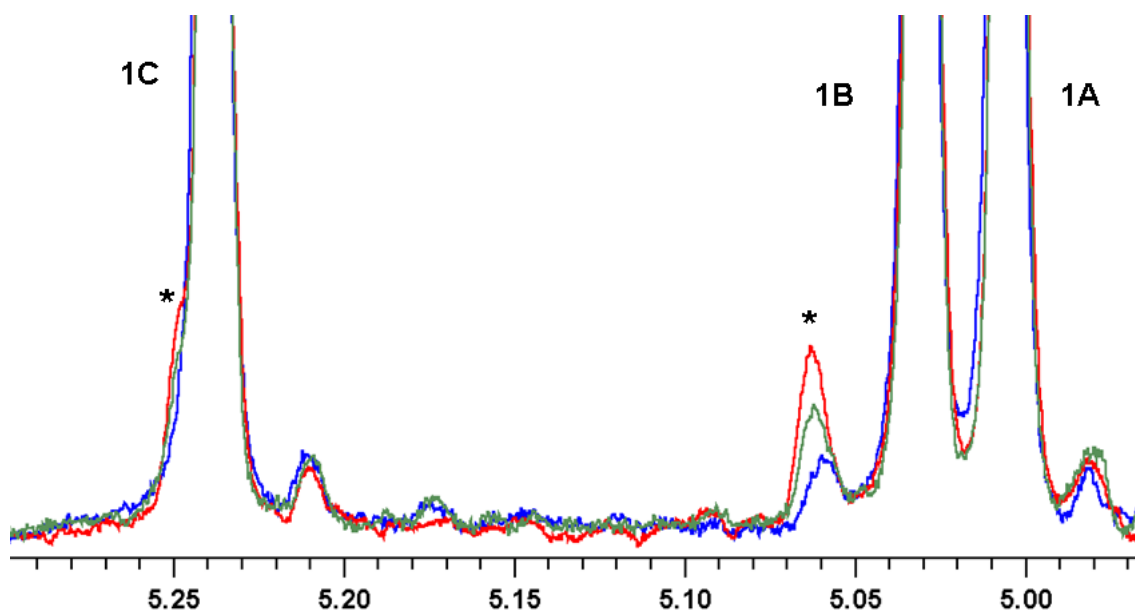


Figure S2.6. Deconvolution of STD-NMR spectrum (3.2-4.2 ppm region) of **2** in the presence of GlfT2 at 298 K. Observed spectrum (black), calculated spectrum (red), individual peaks from fit (green), and difference between calculated and observed spectrum (blue).

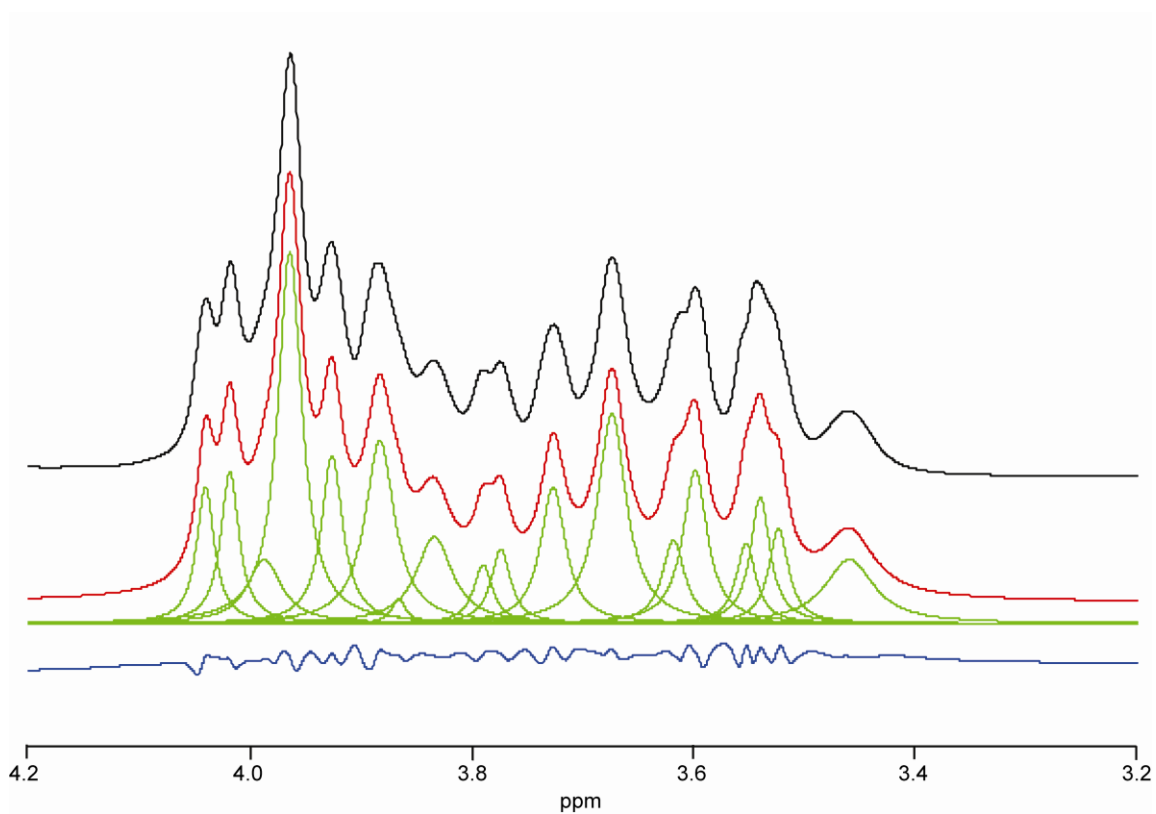
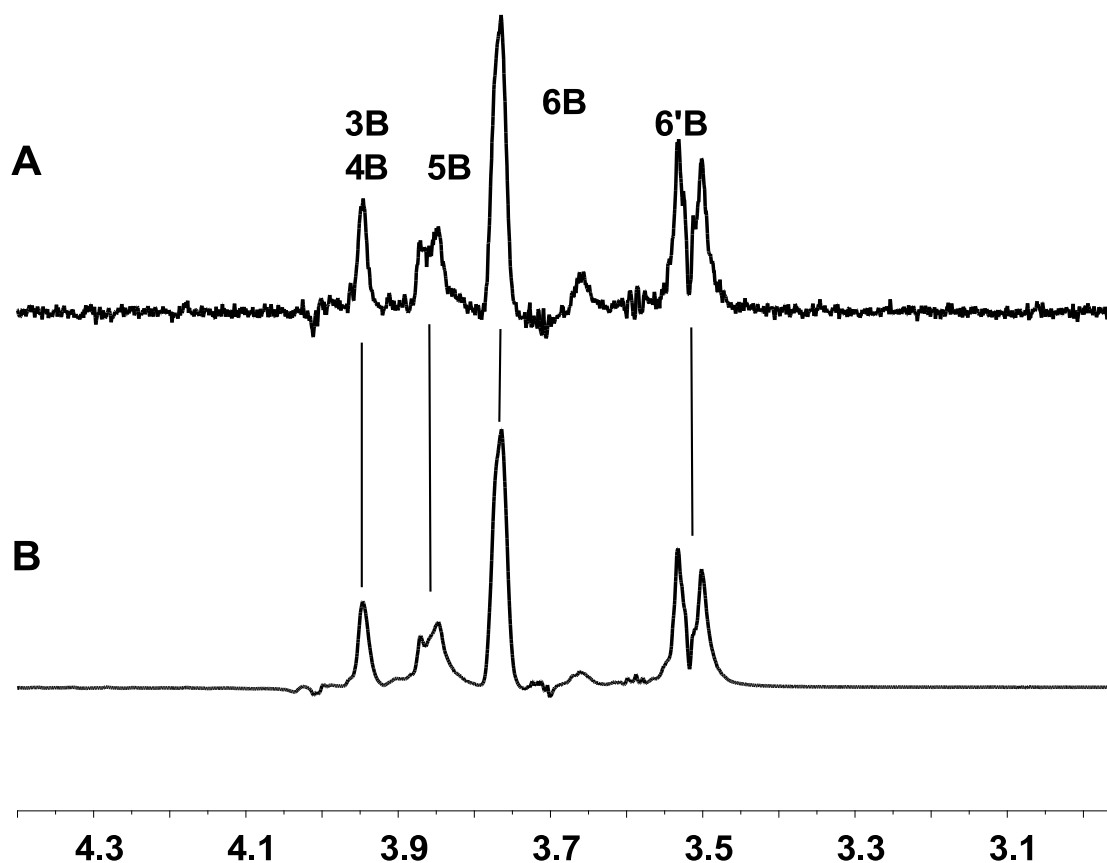


Figure S2.7. STD-2D-TOCSY intensity relative to the 2D-TOCSY intensity for the selectively excited resonance H-6B of **2** in the presence of GlfT2. Row number 190 was extracted from the STD-2D-TOCSY spectrum (upper trace) and corresponds to H-6B and its spin-coupled neighbours, H-6'B, H-5B, H-4B and H-3B. The same row, row number 190, was extracted from the 2D-TOCSY spectrum (lower trace) and also corresponds to H-6B and its spin-coupled neighbours, H-6'B, H-5B, H-4B and H-3B. The STD enhancement of the STD-2D-TOCSY trace (upper trace) was measured relative to the 2D-TOCSY trace (lower trace) and was normalized to the selectively excited resonance, H-6B. The enhancement was quantified in Table S2.4 (second table from the top).



CHAPTER 3: STD-NMR Studies Suggest that Two Acceptor Substrates for GIfT2, a Bifunctional Galactofuranosyltransferase Required for the Biosynthesis of *Mycobacterium tuberculosis* Arabinogalactan, Compete for the Same Binding Site

This chapter comprises the manuscript “***STD-NMR Studies Suggest that Two Acceptor Substrates for GIfT2, a Bifunctional Galactofuranosyltransferase Required for the Biosynthesis of Mycobacterium tuberculosis Arabinogalactan, Compete for the Same Binding Site***” which was published in *ChemBioChem* (2009, 10, 2052-2059).

Monica G. Szczepina,^a Ruixiang B. Zheng,^b Gladys C. Completo,^b

Todd L. Lowary,^b and B. Mario Pinto^a

^aDepartment of Chemistry, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada

^bDepartment of Chemistry and Alberta Ingenuity Centre for Carbohydrate Science, Gunning-Lemieux Chemistry Centre, University of Alberta, Edmonton, Alberta, T6G 2G2, Canada

In Chapter 2, we described that we had established the binding epitope for two trisaccharide acceptor substrates for the bifunctional galactofuranosyl transferase GlfT2. In Chapter 3 we describe an interest in investigating whether GlfT2 catalyzed the substrates using one or two active sites. In the absence of X-Ray crystallographic data, the STD NMR studies carried out in this chapter indicate that one active site is responsible for the formation of both β -(1 \rightarrow 5)- and β -(1 \rightarrow 6)-Gal f linkages.

The preparation of GlfT2 was carried out by Mr. Ruixiang B. Zheng. The synthesis of the trisaccharides was carried out by Dr. Gladys C. Completo. The thesis author performed all the NMR experiments and NMR data interpretation.

3.1 Keywords

Enzymes, Glycosylation, Mycobacteria, NMR Spectroscopy,
Oligosaccharides

3.2 Abstract

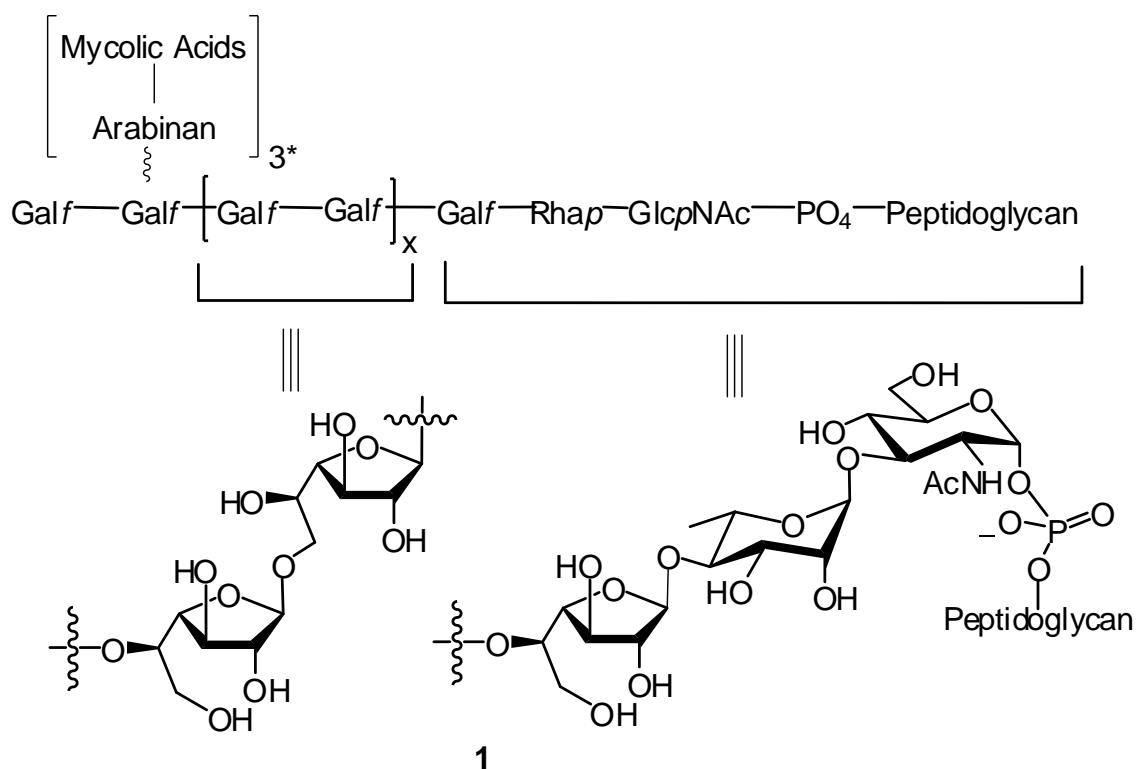
The mycobacterial cell wall is complex architecture, which has, as its major structural component, a lipidated polysaccharide covalently bound to peptidoglycan. This structure, termed the mycolyl–arabinogalactan–peptidoglycan complex, possesses a core galactan moiety composed of approximately 30 galactofuranosyl (Gal f) residues attached via alternating β -(1→6) and β -(1→5) linkages. Recent studies have shown that the entire galactan is synthesized by the action of only two bifunctional galactofuranosyltransferases, GlfT1 and GlfT2. We report here saturation-transfer difference (STD) NMR studies with GlfT2 using two trisaccharide acceptor substrates, β -D-Gal f -(1→6)- β -D-Gal f -(1→5)- β -D-Gal f -O(CH₂)₇CH₃ (**2**) and β -D-Gal f -(1→5)- β -D-Gal f -(1→6)- β -D-Gal f -O(CH₂)₇CH₃ (**3**), as well as the donor substrate for the enzyme, UDP-Gal f . Competition STD NMR titration experiments and saturation transfer double difference (STDD) experiments with **2** and **3** were undertaken to explore the bifunctionality of this enzyme, in particular to answer whether one or two active sites are responsible for the formation of both β -(1→5)- and β -(1→6)-Gal f linkages. It was demonstrated that **2** and **3** bind competitively at the same site, suggesting that GlfT2 has one active site pocket capable of catalyzing both β -(1→5) and β -(1→6)-galactofuranosyl transfer reactions. The addition of UDP-Gal f to either **2** or **3** in the presence of GlfT2 generated a tetrasaccharide product, indicating that the enzyme was catalytically active under the conditions at which the STD-NMR experiments were carried out.

3.3 Introduction

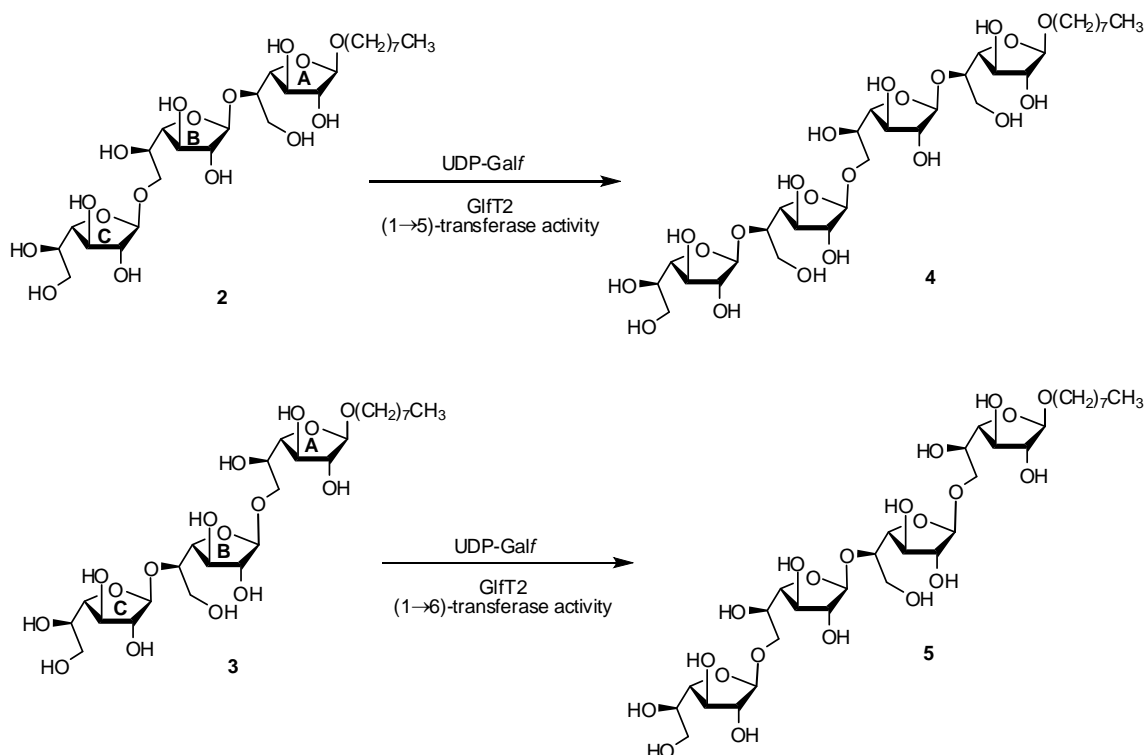
Galactofuranose (Gal_f) residues are constituents of essential glycoconjugates present in a number of microorganisms,^[1-3] but are absent in mammals.^[4] Inhibitors of the enzymes involved in Gal_f metabolism are therefore of interest as novel therapeutic agents.^[5] A Gal_f-containing glycoconjugate receiving increasing scrutiny is the mycolyl–arabinogalactan–peptidoglycan (mAGP) complex, a structure that encapsulates not only the tuberculosis-causing organism *Mycobacterium tuberculosis* but also other mycobacterial species.^[6-8] The mAGP is a major structural component of the mycobacterial cell wall and serves as an essential permeability barrier thus protecting the organism from its environment.^[9]

The mAGP is a galactan moiety composed of D-Gal_f residues connected via alternating β -(1→5) and β -(1→6) linkages. This galactan is attached to peptidoglycan through an α -L-rhamnopyranosyl-(1→3)-2-acetamido-2-deoxy- β -D-glucopyranosyl-phosphate linkage, and is further elaborated by the addition of three mycolylated arabinan domains (1, Scheme 3.1).^[8] mAGP biosynthesis involves the sequential attachment of monosaccharides to a growing polyprenol-linked intermediate, mediated by a series of glycosyltransferases that use either sugar nucleotide or glycosyl phosphoprenol donors.^[9] The process is terminated by subsequent transfer of the lipid-linked arabinogalactan to peptidoglycan and esterification with mycolic acids. A recent report has established that only two enzymes, both bifunctional, are responsible for the assembly of the galactan.^[10] One enzyme, GlfT1,^[11] adds the first and second Gal_f residues while the

remaining residues are added by another transferase, GlfT2.^[12, 13] Therefore, GlfT1 possesses dual β -(1→4) and β -(1→5) transferase activity and GlfT2 harbors both β -(1→5) and β -(1→6) transferase activities. As the donor species, both enzymes use UDP-galactofuranose (UDP-Galf), and the acceptor is the nascent biopolymer.



Scheme 3.1. Structure of the mAGP complex, with the galactan and linker regions highlighted; $x \sim 13-15$. *The three arabinan chains have been proposed to be linked via the eighth, tenth and twelfth Galf residues of the galactan core.^[58]



Scheme 3.2. Trisaccharides **2** and **3**, acceptor substrates for GlfT2, and the corresponding tetrasaccharide products (**4** and **5**) initially formed upon reaction with UDP-Galf and the enzyme.

Previously, oligosaccharides containing *Galf* residues have been tested as acceptor substrates for both transferases,^[14-17] and the screening of potential inhibitors has also been pursued,^[18, 19] with GlfT2 receiving the most attention. Among the substrates tested to date, the trisaccharides **2** and **3** (Scheme 3.2), were shown to be the most efficiently galactosylated by GlfT2.^[20] The former is initially a substrate for the β -(1→5)-transferase activity of the enzyme while the latter is a substrate for the β -(1→6)-transferase activity. As the products formed from **2** and **3** (tetrasaccharides **4** and **5**) are also substrates for the enzyme, as

expected, a ladder of products is formed upon incubation of these compounds with the enzyme and UDP-Galf.^[20]

It is of interest to explore the bifunctionality of GlfT2, in particular, to determine whether one or two active sites are responsible for the assembly of β -(1 \rightarrow 5)- and β -(1 \rightarrow 6)-Galf linkages. Although bifunctional glycosyltransferases other than GlfT1 and GlfT2 are known,^[21-40] they are relatively uncommon. Among those receiving the most study are enzymes involved in proteoglycan biosynthesis. In particular, bifunctional type-II hyaluronan synthases have been shown to possess two active sites.^[24, 27, 28, 39] These enzymes differ from GlfT1 and GlfT2 in that glycosidic linkages involving different sugars are produced by each transferase activity, and thus the proteins bind to two different sugar nucleotides (i.e., UDP-*N*-acetylglucosamine and UDP-glucuronic acid). In this respect, GlfT1 and GlfT2 are more similar to the enzymes involved in cellulose biosynthesis,^[21, 23, 29, 33, 34] as well as the sialyltransferases^[22, 25, 32, 37] responsible for the assembly of alternating α -(2 \rightarrow 8), α -(2 \rightarrow 9)-linked, and α -(2 \rightarrow 3), α -(2 \rightarrow 8)-linked polysialic acid. In all these cases, the enzyme uses a single sugar nucleotide (UDP-Galf, UDP-Glcp, or CMP-NeupNAc) to produce the polymer. Although in the case of cellulose, a single glycosidic linkage (α -Glcp-(1 \rightarrow 4)- α -Glcp) is produced, it has been suggested that the structure of the polysaccharide, in which adjacent glucopyranose residues are oriented 180° relative to each other, requires a bifunctional enzyme.^[21, 23, 29, 33, 34] The details of cellulose biosynthesis remain unclear, and although it has been proposed that two active sites are involved,^[29] more recent investigations have called this into question.^{[21,}

^{23, 34]} In contrast, the mechanism of the α -(2→8), α -(2→9), and α -(2→3), α -(2→8)-linked polysialic acid transferases is less controversial, and recent studies suggest that these enzymes possess a single active site.^[22, 25, 32, 37]

Given the relatively low sequence homology between GIfT1/GIfT2 and these more thoroughly investigated bifunctional glycosyltransferases, it is difficult to predict if catalysis by these enzymes involves one or two active sites. However, based on sequence comparisons available for other glycosyltransferases not of mycobacterial origin, both enzymes have been assigned to the GT-2 family of glycosyltransferases, which is part of the GT-A superfamily.^[9] GT-A glycosyltransferases generally possess DXD motifs that are involved in binding to sugar nucleotides through their phosphodiester moiety via complexes formed with divalent metal ions.^[41] Both GIfT1 and GIfT2 have a number of such motifs,^[9, 12, 13] which may suggest the presence of two active sites, each responsible for catalysis of a single glycosylation event.

Recombinant *M. tuberculosis* GIfT2 and GIfT1 have recently been expressed in *E. coli*.^[42, 43] To gain insight into the process by which GIfT2 polymerizes the galactan core of the mAG complex, we report here the use of competition saturation transfer difference (STD) NMR titration experiments with acceptor substrates **2** and **3** (Scheme 3.2).

The use of STD NMR intensity makes it possible to differentiate if **2** and **3** bind to one or two different active sites (competitive vs noncompetitive binding, respectively).^[44] If the substrate binds to the same active site, titration of the protein with one substrate (e.g., **2**) followed by the second (e.g., **3**) will result in a

loss of STD signals for **2** with the appearance of signals for **3**. Alternatively, if the substrates bind at different sites, the STD NMR intensity of one substrate will not be affected as one titrates with the other substrate, unless the second substrate binds in an allosteric manner. It is also possible to use STD NMR intensity to differentiate between allosteric and noncompetitive binding. If ligand binding is noncompetitive, then the STD NMR signal intensity of ligand A will change if ligand B is titrated in and binds at an allosteric site; however, if one begins with ligand B in the allosteric site, then titrates in ligand A, the STD NMR signal intensity of ligand B will not change upon addition of ligand A. This reasoning assumes that there is no mutual allosteric effect of the binding of the two ligands. Therefore, by performing the STD NMR competition experiments in both directions, it is possible to determine whether the nature of the binding of ligands **2** and **3** to GlfT2 is competitive or noncompetitive.

3.4 Experimental

3.4.1 Materials

Trisaccharides **2** and **3** were synthesized as described elsewhere.^[45]
GlfT2 was expressed and purified as reported.^[20]

3.4.2 NMR spectroscopy

To a sample of GlfT2 (1.6 mg) in phosphate buffered saline solution ($\text{Na}_2\text{HPO}_4\text{-NaH}_2\text{PO}_4$ (50mM), NaCl (0.1 M), MgCl_2 (10 mM) and 99% D_2O (the pH of the starting buffer was 7.6)) was added either β -(1 \rightarrow 6), β -(1 \rightarrow 5) trisaccharide **2** (1.3 mg) or β -(1 \rightarrow 5), β -(1 \rightarrow 6) trisaccharide **3** (1.2 mg). Mg^{2+} was

used because of paramagnetic line broadening caused by Mn^{2+} . Because turnover is observed, the chemistry does not appear to be affected by this change in metal ion. The final ligand concentration was 4 mM at a ligand–protein ratio of 100:1. STD NMR spectra with UDP-Galf were acquired with UDP-Galf (0.5 mg) and GlfT2 (1.2 mg). STD NMR competition studies were acquired at ratios 1:0.2, 1:0.5, 1:1, 1:2 and 1:5 for the ligand combination β -(1→6), β -(1→5) trisaccharide **2** to β -(1→5), β -(1→6) trisaccharide **3**, as well as β -(1→5), β -(1→6) trisaccharide **3** to β -(1→6), β -(1→5) trisaccharide **2**. Incubation of GlfT2 (1.25 mg, 33 μ M) with UDP-Galf (1.0 mg, 3 mM) and **2** (1.0 mg, 3 mM) generated the β -(1→5), β -(1→6), β -(1→5) tetrasaccharide product, **4**. The enzyme was recycled using Centricon preparation and incubated with UDP-Galf (1.0 mg, 3 mM) and **3** (1.0 mg, 3 mM) to generate the β -(1→6), β -(1→5), β -(1→6) tetrasaccharide product, **5**. Ligand resonances were assigned using 1H – 1H COSY, 1H – 1H TOCSY and 1H – 1H NOESY NMR spectroscopy. Linkages between galactose rings were assigned on the basis of 1H – 1H NOESY NMR spectra. Thus, interglycosidic NOEs were observed between H-1A and OCH₂, H-1C and H-6B as well as between H-1B and H-5A, confirming the galactose ring connectivities. Water suppression using presaturation was utilized in experiments with ligand only. ^{31}P NMR spectra (at 162 MHz) were recorded on a Bruker AMX-400 NMR spectrometer.

STD-NMR spectra and competition experiments for the titration of **3** into **2** were performed on a Bruker AMX-600 NMR spectrometer. Competition studies for the titration of **2** into **3** were recorded on a Bruker Avance 600 NMR

spectrometer, equipped with a TCI cryoprobe. All spectra were recorded with 1024 or 2048 scans and selective saturation of protein resonances at 10 ppm (30 ppm for off resonance spectra) using a series of 40 Gaussian-shaped pulses (50 ms, 1 ms delay between pulses, $\gamma B_1/2\pi = 110$ Hz), for a total saturation time of 2.04 s.^[46] The protein resonances were broad and had significant intensity in the region downfield from 10 ppm. Thus, irradiation at 10 ppm was expected to result in saturation of protein resonances, from the aromatic to the aliphatic. Irradiation at 10 ppm was also considered prudent in achieving selective saturation of the protein resonances because a ligand resonance was present at 0.8 ppm. Subtraction of saturated spectra from reference spectra was performed by phase cycling, on the Bruker AMX-600 NMR spectrometer. Measurement of enhancement intensities was performed by direct comparison of STD NMR spectra and reference one-dimensional ^1H NMR spectra. In the case of STD NMR spectra acquired on the Bruker Avance 600 NMR spectrometer, the saturated and reference spectra were acquired simultaneously by creating a pseudo-2D experiment. The STD spectrum was obtained by subtraction of saturated spectra from reference spectra after identical processing and phasing. In all cases, the fractional STD effect was calculated by $(I_0 - I_{\text{sat}})/I_0$, where $(I_0 - I_{\text{sat}})$ is the peak intensity in the STD spectrum and I_0 is the peak intensity of an unsaturated reference spectrum. All STD spectra were acquired without water suppression.

STD-1D-TOCSY spectra were recorded on a Bruker AMX-600 NMR spectrometer with 4096 scans, and selective saturation of the protein using a

series of 40 Gaussian-shaped pulses (50 ms, 1 ms delay between pulses, $\gamma B_1/2\pi = 110$ Hz), for a total saturation time of 2.04 s as described previously.^[47] The MLEV-16 composite pulse was used for isotropic mixing with a power level of $\gamma B_1/2\pi = 11$ kHz. Selective inversion of the resonance of interest was achieved using an 80 ms Gaussian pulse, with 70-71 dB attenuation ($\gamma B_1/2\pi = 11$ Hz). Gradient pulses were of 2 ms duration and strength ratio 7:-3:-10 (where 10% strength = 6.6 G/cm), and were followed by 100 μ s ringdown delays. The additional relaxation delay was minimal (1 ms), and the rephasing delay d_3 was set equal to the length of the last gradient and surrounding power-switching and ring-down delays (2.16 ms). 2D-STD-TOCSY spectra were recorded with 16 or 32 scans per t_1 increment. A total of 256 t_1 increments were collected in an interlaced mode for the on- and off- resonance spectra. An MLEV-17 spin-lock sequence with a 10kHz rf field strength and a mixing time of 80ms was utilized. Enhancement intensities were referenced to the methyl group of the octyl moiety. All enhancements (including cross peaks from 2D-STD-TOCSY experiments) were measured in triplicate, and the average values are shown. The error is estimated at $\pm 10\%$.

Longitudinal relaxation times T_1 s of **2** and **3** in the presence of GlfT2 were determined with the inversion recovery pulse sequence. Data processing was performed using XWINNMR (Bruker), TopSpin (Bruker), and MestReC.

3.5 Results

3.5.1 STD NMR competition studies between **2** and **3**

Incubation of GIfT2 with UDP-Galp and trisaccharide **2** generated tetrasaccharide **4**; similarly, trisaccharide **3** produced **5** (Scheme 3.2), thus indicating that the enzyme was active under the conditions of the NMR experiments (See Figures S3.1 and S3.2 in Supporting Information).

Previous work in our laboratory utilized competition STD NMR experiments to deduce relative binding affinities of ligands for the active site in the enzyme UDP-Galp mutase.^[48, 49] Thus, to examine whether **2** and **3** compete for the same binding site, trisaccharide **3** was titrated into a sample containing GIfT2 (51 μ M) and acceptor substrate **2** (4 mM) at 285 K. The resulting 1D-STD NMR spectra for various ratios of **2** to **3** are shown in Figure 3.1, which indicate that at a **2:3** ratio of 1:0.3, the anomeric resonances from **3** are clearly visible. Upon further addition of **3**, the anomeric resonances of **2** decreased in intensity, while those of **3** increased.

To illustrate this point with greater clarity, we used a modified version of Meyer's Saturation Transfer Double Difference (STDD) NMR method.^[50] This method relies on acquiring the STD NMR spectrum of the protein receptor in addition to acquiring the STD NMR spectrum of the protein receptor with ligand. The modified version of this method entailed subtracting the STD spectrum of **2** in the presence of GIfT2, from the STD spectra of the mixtures at various ratios of **2** to **3** (Figure 3.2). The resonances of **3** increased in intensity, while those of

2 decreased until they became negative, presumably because the sites initially occupied by **2** were now almost exclusively occupied by **3**.

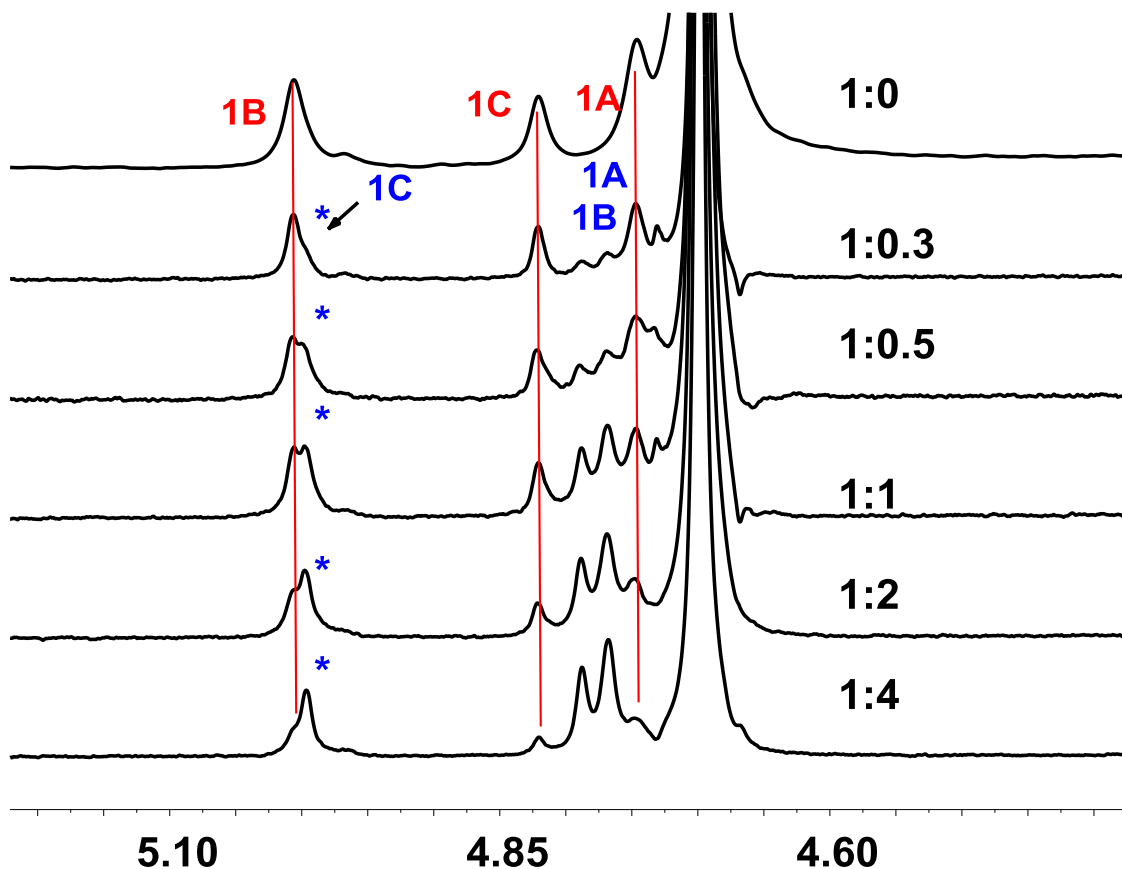


Figure 3.1. 1D-STD NMR spectra at different ratios of the acceptor substrates **2** to **3**. The upper trace corresponds to the acceptor substrate **2** in the presence of GlfT2. Lower traces indicate spectra obtained with titration of increasing amounts of the acceptor substrate **3**; the **2:3** ratio is indicated with each trace.

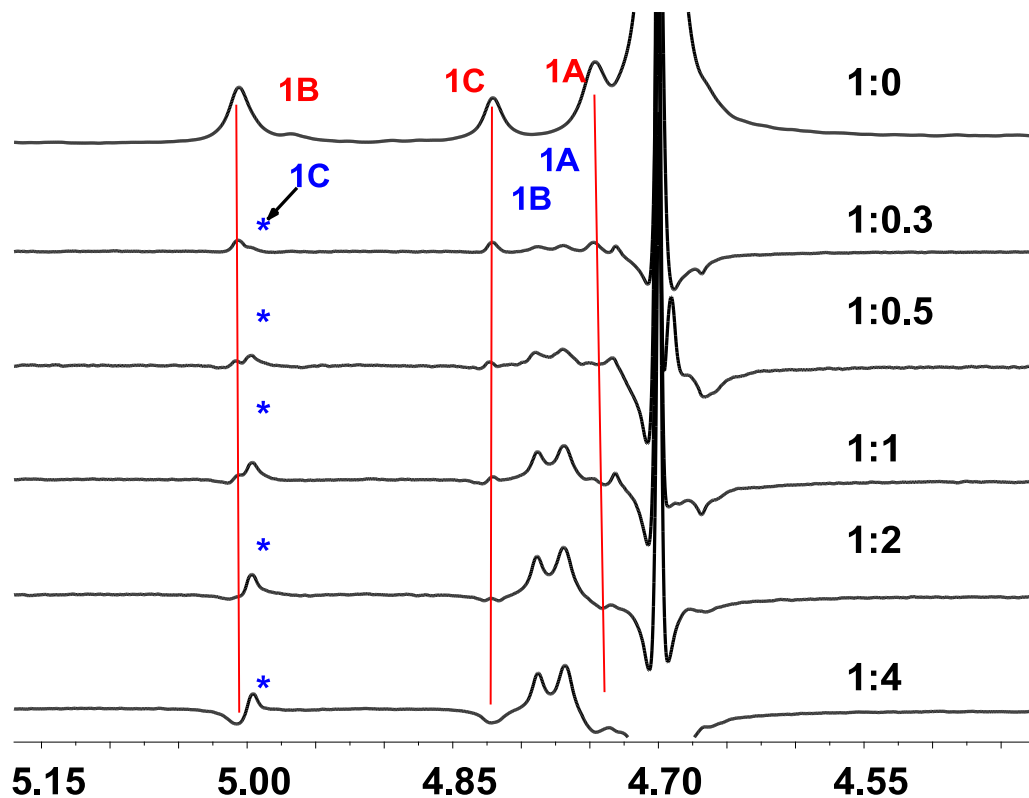


Figure 3.2. 1D-STDD NMR spectra at different ratios of the acceptor substrates **2** to **3**. The upper trace, corresponding to a 1D-STD NMR spectrum of the acceptor substrate **2** in the presence of GlfT2, was subtracted from the traces in Figure 3.1.

Mayer and Meyer have defined^[44] the STD amplification factor in the following manner:

$$\left(\frac{I_0 - I_{sat}}{I_0} \right) \times \text{ligand excess}$$

The STD amplification factor relates the STD signal of a ligand to the population of the ligand relative to the protein. In titration and competition experiments in which ligand concentrations and/or protein concentrations change, the STD amplification factor reflects the STD enhancement under these dynamic conditions. To better quantify the results observed, STD amplification factors for H-1B and H-1A of **3** as well as H-1A and H-1C of **2** were measured at each titration point from 1D-STD NMR spectra (Figure 3.3). The STD amplification factors for **2** decreased with a concomitant increase in the STD amplification factors for **3**. This result suggests that **2** and **3** are either binding competitively at the same site or binding noncompetitively at different sites. Use of the Cheng–Prusoff equation^[51] gave estimates of K_D for **2** of $85 \mu\text{M} \pm 15 \mu\text{M}$ (c.f. $204 \mu\text{M}$ from kinetic experiments^[20]).

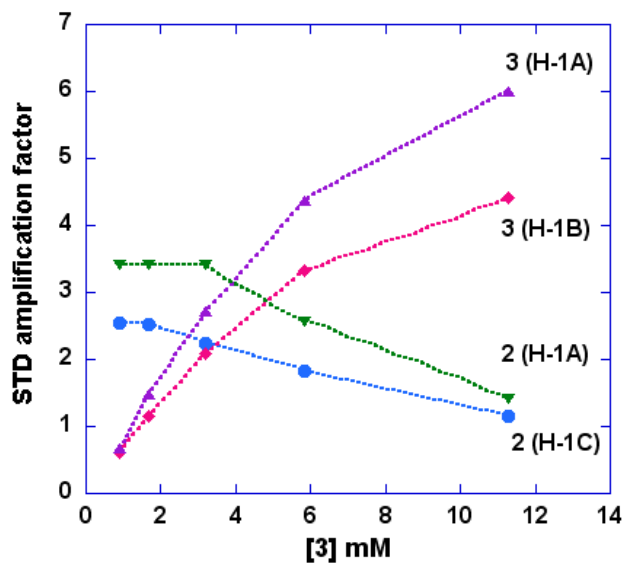


Figure 3.3. Titration of a sample of trisaccharide, **2** (4 mM) containing GlfT2 (51 μ M) with trisaccharide **3**. STD amplification factors were obtained from 1D-STD NMR spectra. The STD amplification factor for resonances H-1A and H-1C from **2** decreased whereas those for resonances H-1B and H-1A from **3** increased.

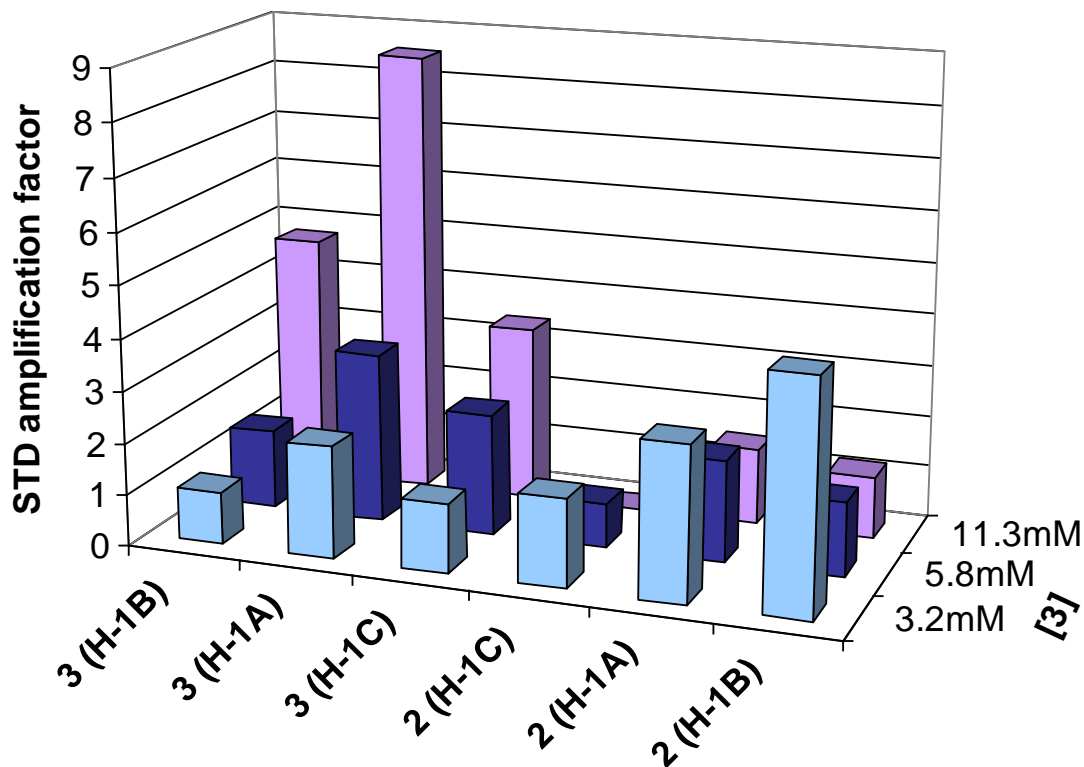


Figure 3.4. STD amplification factors calculated from six cross peak intensities originating from 2D-STD-TOCSY NMR experiments. Trisaccharide **2** to trisaccharide **3** ratios of 1:1, 1:2 and 1:4 correspond to trisaccharide **3** concentrations of 3.2 mM, 5.8 mM and 11.3 mM. These six cross peaks comprised the F1 traces from the H-1A, H-1B, and H-1C resonances of acceptor **2** as well as F1 traces from the H-1A, H-1B and H-1C resonances of the acceptor **3**. STD amplification factors for trisaccharide **2** decreased while STD amplification factors for trisaccharide **3** increased as a function of increasing amounts of trisaccharide **3**.

Due to resonance overlap, 2D-STD-TOCSY NMR spectra were acquired. At acceptor **3** concentrations of 3.2 mM, 5.8 mM and 11.3 mM, corresponding to ratios of **2** to **3** of 1:1, 1:2 and 1:4, six cross peaks were selected from the 2D-STD-TOCSY NMR spectra. These peaks comprised the F1 traces from the H-1A, H-1B, H-1C resonances of acceptor **2** as well as F1 traces from the H-1A, H-1B and H-1C resonances of acceptor **3**. The cross peaks referring to between the H-1 and H-2 resonances were integrated and the corresponding STD amplification factors were calculated (Figure 3.4). Signals from trisaccharide **2** decreased while those from trisaccharide **3** increased; this corroborates the results from the 1D-STD NMR experiments (see Figure 3.1). Application of the Cheng–Prusoff equation^[51] to these selected cross-peaks gave estimates of a K_D for **2** of $115 \mu\text{M} \pm 45 \mu\text{M}$ (c.f. $204 \mu\text{M}$ from kinetic experiments^[20]).

To discriminate between competitive binding or noncompetitive binding of substrates **2** and **3** to GlfT2, the experiment was repeated, but in the reverse direction. Thus, acceptor substrate **2** was titrated into a sample containing GlfT2 ($74 \mu\text{M}$) and trisaccharide **3** (4 mM) at 298 K. The resulting 1D-STD NMR spectra for various ratios of **3** to **2** are shown in Figure 3.5: at a **3:2** ratio of 1:0.2, anomeric resonances from **2** are visible. Upon further addition of **2**, the anomeric resonances of **3** decreased in intensity, while those of **2** increased. The corresponding 1D-STDD NMR spectra are shown in the Supporting Information (Figure S3.3).

STD amplification factors for H-1B and H-1A of **3** as well as H-1C and H-1A of **2** were measured at each titration point from 1D-STD NMR spectra and

these are shown in Figure 3.6. The STD amplification factors for acceptor **3** decreased with a concomitant increase in the STD amplification factors for acceptor **2**. This is significant because it indicates that substrates **2** and **3** are binding competitively. If **3** were binding at an allosteric site, then by reversing the order of addition, the STD signal intensity of **2** would not have been affected. The caveat in this analysis is that there can be no mutual allosteric effect of the binding of the two ligands. In the simplest interpretation, these series of experiments demonstrate that **2** and **3** bind competitively, and, in turn, suggest that the enzyme possesses one active site pocket. Use of the Cheng–Prusoff equation^[51] (see Figure S3.4) gave estimates of K_D for **3** of $50 \mu\text{M} \pm 10 \mu\text{M}$ (c.f. $208 \mu\text{M}$ from kinetic experiments.^[20]

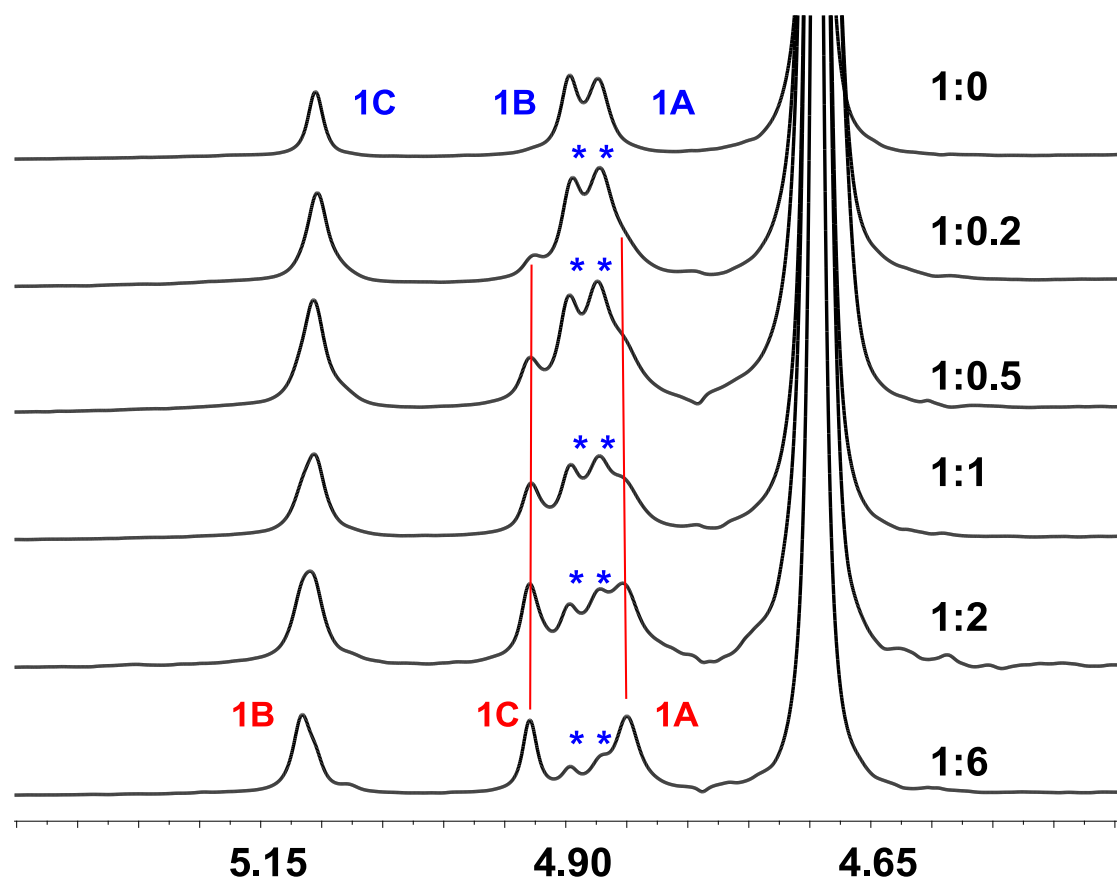


Figure 3.5. 1D-STD NMR spectra at different ratios of the acceptor substrates **3** to **2**. The upper trace corresponds to the acceptor substrate **3** in the presence of GlfT2. Lower traces indicate spectra obtained with titration of increasing amounts of the acceptor substrate **2**; the **3:2** ratio is indicated with each trace.

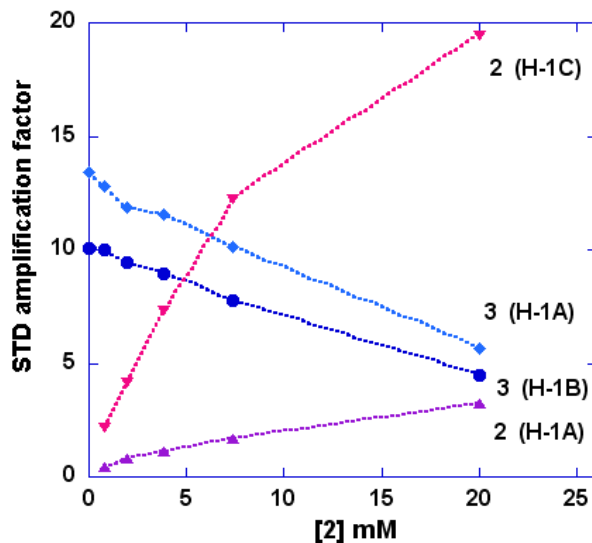


Figure 3.6. Titration of trisaccharide, **3** (4 mM) containing GlfT2 (74 μ M) with trisaccharide **2**. STD amplification factors were obtained from 1D-STD NMR spectra. The STD amplification factor for resonances H-1A and H-1B from **3** decreased whereas that for resonance H-1C and H-1A from **2** increased.

It should be noted that at a **2:3** ratio of 1:1, the STD amplification factor for proton H-1C of **2** at 285 K is 3.4, whereas at 298 K the STD amplification factor is 7.3. Similarly, the STD amplification factor for H-1B of **3** at 285 K is 2.0 but at 298 K it is 8.9. This discrepancy may be due to different off-rates as a result of the temperature change and not due to differences in T1s. As in the previous set of experiments 2D-STD-TOCSY NMR spectra were acquired. At acceptor **2** concentrations of 3.9 mM, 7.4 mM and 20.0 mM, corresponding to ratios of **3** to **2** of 1:1, 1:2 and 1:6, four cross peaks were selected from the 2D-STD-TOCSY NMR spectra. These peaks comprised the F1 traces from the H-1A and H-1C

resonances of acceptor **2** as well as F1 traces from the H-1A and H-1B resonances of acceptor **3**. The cross peaks were integrated and the corresponding STD amplification factors were calculated (Figure 3.7). Signals from the trisaccharide **2** increased while signals from the trisaccharide **3** decreased, which is consistent with the two compounds binding to GlfT2 in a competitive manner. Application of the Cheng–Prusoff equation^[51] to these selected cross-peaks gave estimates of K_D for **3** of $50 \mu\text{M} \pm 10 \mu\text{M}$ (c.f. $208 \mu\text{M}$ from kinetic experiments^[20]).

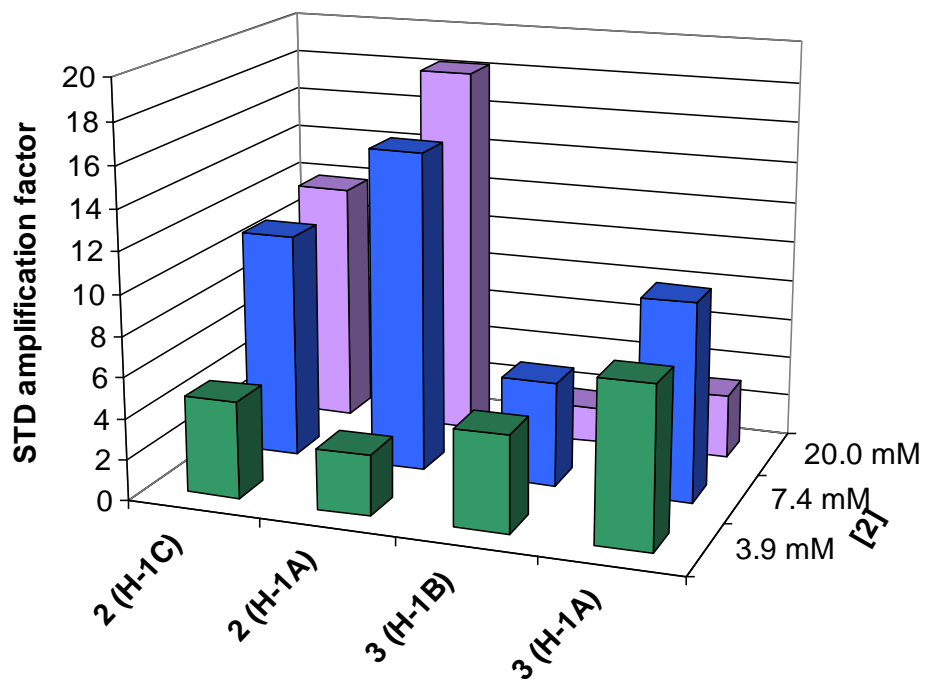


Figure 3.7. STD amplification factors calculated from cross peak intensities originating from 2D-STD-TOCSY NMR experiments. Trisaccharide **3** to trisaccharide **2** ratios of 1:1, 1:2 and 1:6, correspond to trisaccharide **2** concentrations of 3.9 mM, 7.4 mM and 20.0 mM. These four cross peaks comprised the F1 traces from the H-1A and H-1C resonances of acceptor **2** as well as F1 traces from the H-1A and H-1B resonances of the acceptor **3**. STD amplification factors for trisaccharide **3** decreased while STD amplification factors for trisaccharide **2** increased as a function of increasing amounts of **2**.

3.6 Discussion

Recent studies have provided significant insight into the biosynthesis of the galactan portion of mycobacterial mAGP.^[6-10, 52] The current model is that the ~30-residue glycan is assembled by two bifunctional galactofuranosyltransferases, GlfT1 and GlfT2.^[10] Galactan synthesis is essential for mycobacterial viability,^[53] and thus inhibitors of these glycosyltransferases would be novel drug candidates for the treatment of diseases including tuberculosis, and HIV-associated *M. avium* infections.^[54] An understanding of the structures and mechanisms of these enzymes would facilitate the design and synthesis of such inhibitors. In this paper, we have focused on GlfT2.^[20, 42] Our efforts to obtain an X-ray crystal structure of GlfT2 have not been successful to date. Therefore, we have used STD-NMR spectroscopy to investigate protein–substrate interactions and, in particular, to probe the bifunctionality of GlfT2, which synthesizes both β -(1→5) and β -(1→6) Galf-linkages.

The studies described here demonstrate that trisaccharides **2** and **3** bind competitively to GlfT2, in turn suggesting that the protein has a single active site pocket capable of installing β -(1→5)- and β -(1→6)-linked Galf residues. Thus, the enzyme appears to be similar to other bifunctional glycosyltransferases that use a single sugar nucleotide donor. For example, the bifunctional sialyltransferases that synthesize α -(2→8), α -(2→9)-linked and α -(2→3), α -

(2→8)-linked polysialic acid^[22, 25, 32, 37] have been proposed to possess a single active site that can carry out both glycosylation events.

As mentioned previously, there are several DXD motifs in GlfT2, which could indicate more than one active site; however, based on the data presented here it appears that only one of these motifs is involved in sugar nucleotide binding and thus catalysis. In this regard, it should be noted that a previous analysis of the GlfT2 sequence suggested that the location of many of these DXD motifs within the protein made them unlikely candidates for binding of UDP-Galf.^[13] Indeed, this analysis identified only one, a DDD sequence spanning amino acids 256 and 258 that was part of a classic DXD motif in which the *N*-terminal end is flanked with hydrophobic amino acids.^[55] Thus, the results reported here are in agreement with these earlier genomic analyses.

We have also estimated the order of magnitude of K_D values for both acceptors **2** and **3** using both 1D STD NMR and 2D-STD-TOCSY NMR data, and internal consistency is observed. Comparison of K_D values to K_M values previously determined for these oligosaccharides^[21] indicates a similar order of magnitude (i.e., μM affinity). It should be noted that the K_D values are expected to be smaller than K_M values, by a factor of k_2/k_1 , as per Michaelis–Menten kinetics where k_1 and k_2 are the rate constants for the formation of the enzyme–substrate complex and the product, respectively.^[56] Indeed, that is what we observed and this discrepancy is typical for STD-NMR versus kinetic measurements.^[44, 57]

In conclusion, STD-NMR effects have been used to probe the competitive or noncompetitive nature of substrate binding by GlfT2, in a case where two substrates bind to, and are processed by, the same enzyme. In the case of GlfT2, it has been shown that the trisaccharide acceptor substrates **2** and **3** bind competitively at the same binding site rather than binding noncompetitively at different sites. Additional support for the presence of a single active site in GlfT2 would come from an X-ray structure of the protein and such studies are currently underway.

3.7 Acknowledgements

This work was supported by The Alberta Ingenuity for Carbohydrate Science and The Natural Sciences and Engineering Research Council of Canada. We thank Mr. Ruokun Zhou for technical assistance, and Mr. Dustin Bleile for helpful discussions.

3.8 Summary

We have described competition STD NMR titration experiments and saturation transfer double difference (STDD) experiments with trisaccharide acceptor substrates, β -D-Galf-(1 \rightarrow 6)- β -D-Galf-(1 \rightarrow 5)- β -D-Galf-O(CH₂)₇CH₃ (**2**) and β -D-Galf-(1 \rightarrow 5)- β -D-Galf-(1 \rightarrow 6)- β -D-Galf-O(CH₂)₇CH₃ (**3**), for the enzyme GlfT2. These studies were undertaken to answer whether one or two active sites are responsible for the formation of both β -(1 \rightarrow 5)- and β -(1 \rightarrow 6)-Galf linkages. We have described our findings that **2** and **3** bind competitively at the same site, suggesting that GlfT2 has one active site pocket capable of catalyzing both β -(1 \rightarrow 5) and β -(1 \rightarrow 6)-galactofuranosyl transfer reactions.

3.9 References

- [1] R. M. de Lederkremer, W. Colli, *Glycobiology* **1995**, *5*, 547-552.
- [2] S. M. Beverley, K. L. Owens, M. Showalter, C. L. Griffith, T. L. Doering, V. C. Jones, M. R. McNeil, *Eukaryotic Cell* **2005**, *4*, 1147-1154.
- [3] W. Morelle, M. Bernard, J. P. Debeaupuis, M. Buitrago, M. Tabouret, J. P. Latge, *Eukaryotic Cell* **2005**, *4*, 1308-1316.
- [4] J. B. Houseknecht, T. L. Lowary, *Curr. Opin. Chem. Biol.* **2001**, *5*, 677-682.
- [5] L. L. Pedersen, S. J. Turco, *Cell. Molec. Life Sci.* **2003**, *60*, 259-266.
- [6] S. Mahapatra, J. Basu, P. J. Brennan, D. C. Crick, in *Tuberculosis and the Tubercle Bacillus* (Eds.: S. T. Cole, K. D. Eisenach, D. N. McMurray, J. Jacobs, W. R.), American Society for Microbiology, Washington, DC, **2005**, pp. 275-285.
- [7] T. L. Lowary, in *Glycoscience: Chemistry and Chemical Biology*, Vol. 3 (Eds.: B. O. Fraser-Reid, K. Tatsuta, J. Thiem), Springer-Verlag, Berlin, **2001**, pp. 2005-2080.

- [8] G. S. Besra, K. H. Khoo, M. R. McNeil, A. Dell, H. R. Morris, P. J. Brennan, *Biochemistry* **1995**, *34*, 4257-4266.
- [9] S. Berg, D. Kaur, M. Jackson, P. J. Brennan, *Glycobiology* **2007**, *17*, 35R-56R.
- [10] M. Belanova, P. Dianiskova, P. J. Brennan, G. C. Completo, N. L. Rose, T. L. Lowary, K. Mikusova, *J. Bacteriol.* **2008**, *190*, 1141-1145.
- [11] K. Mikusova, M. Belanova, J. Kordulakova, K. Honda, M. R. McNeil, S. Mahapatra, D. C. Crick, P. J. Brennan, *J. Bacteriol.* **2006**, *188*, 6592-6598.
- [12] K. Mikusova, T. Yagi, R. Stern, M. R. McNeil, G. S. Besra, D. C. Crick, P. J. Brennan, *J. Biol. Chem.* **2000**, *275*, 33890-33897.
- [13] L. Kremer, L. G. Dover, C. Morehouse, P. Hitchin, M. Everett, H. R. Morris, A. Dell, P. J. Brennan, M. R. McNeil, C. Flaherty, K. Duncan, G. S. Besra, *J. Biol. Chem.* **2001**, *276*, 26430-26440.
- [14] A. K. Pathak, V. Pathaka, L. Seitz, S. S. Gurcha, G. S. Besra, J. M. Riordan, R. C. Reynolds, *Bioorg. Med. Chem.* **2007**, *15*, 5629-5650.
- [15] X. H. Wen, D. C. Crick, P. J. Brennan, P. G. Hultin, *Bioorg. Med. Chem.* **2003**, *11*, 3579-3587.

- [16] A. K. Pathak, G. S. Besra, D. Crick, J. A. Maddry, C. B. Morehouse, W. J. Suling, R. C. Reynolds, *Bioorg. Med. Chem.* **1999**, *7*, 2407-2413.
- [17] A. K. Pathak, V. Pathak, L. Seitz, J. A. Maddry, S. S. Gurcha, G. S. Besra, W. J. Suling, R. C. Reynolds, *Bioorg. Med. Chem.* **2001**, *9*, 3129-3143.
- [18] S. Cren, S. S. Gurcha, A. J. Blake, G. S. Besra, N. R. Thomas, *Org. Biomol. Chem.* **2004**, *2*, 2418-2420.
- [19] S. Cren, C. Wilson, N. R. Thomas, *Org. Lett.* **2005**, *7*, 3521-3523.
- [20] N. L. Rose, G. C. Completo, S. J. Lin, M. McNeil, M. M. Palcic, T. L. Lowary, *J. Am. Chem. Soc.* **2006**, *128*, 6721-6729.
- [21] S. J. Charnock, B. Henrissat, G. J. Davies, *Plant Physiol.* **2001**, *125*, 527-531.
- [22] C. P. C. Chiu, A. G. Watts, L. L. Lairson, M. Gilbert, D. Lim, W. W. Wakarchuk, S. G. Withers, N. C. J. Strynadka, *Nat. Struct. Mol. Biol.* **2004**, *11*, 163-170.
- [23] G. J. Davies, S. J. Charnock, B. Henrissat, *Trends Glycosci. Glucotechnol.* **2001**, *13*, 105-120.
- [24] P. L. DeAngelis, *Glycobiology* **2002**, *12*, 9R-16R.

- [25] M. Gilbert, J. R. Brisson, M. F. Karwaski, J. Michniewicz, A. M. Cunningham, Y. Y. Wu, N. M. Young, W. W. Wakarchuk, *J. Biol. Chem.* **2000**, *275*, 3896-3906.
- [26] S. Guan, A. J. Clarke, C. Whitfield, *J. Bacteriol.* **2001**, *183*, 3318-3327.
- [27] W. Jing, P. L. DeAngelis, *Glycobiology* **2000**, *10*, 883-889.
- [28] W. Jing, P. L. DeAngelis, *Glycobiology* **2003**, *13*, 661-671.
- [29] M. Koyama, W. Helbert, T. Imai, J. Sugiyama, B. Henrissat, *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 9091-9095.
- [30] D. Llull, E. García, R. López, *J. Biol. Chem.* **2001**, *276*, 21053-21061.
- [31] S. Lobau, U. Mamat, W. Brabetz, H. Brade, *Mol. Microbiol.* **1995**, *18*, 391-399.
- [32] M. M. McGowen, J. Vionnet, W. F. Vann, *Glycobiology* **2001**, *11*, 613-620.
- [33] I. M. Saxena, R. M. Brown, *Ann. Bot.* **2005**, *96*, 9-21.
- [34] C. Somerville, *Annu. Rev. Cell Dev. Biol.* **2006**, *22*, 53-78.
- [35] S. Toivonen, O. Aitio, O. Renkonen, *J. Biol. Chem.* **2001**, *276*, 37141-37148.

- [36] H. van der Wel, S. Z. Fisher, C. M. West, *J. Biol. Chem.* **2002**, *277*, 46527-46534.
- [37] J. Vionnet, W. F. Vann, *Glycobiology* **2007**, *17*, 735-743.
- [38] W. W. Wakarchuk, D. Watson, F. St Michael, J. J. Li, Y. Y. Wu, J. R. Brisson, N. M. Young, M. Gilbert, *J. Biol. Chem.* **2001**, *276*, 12785-12790.
- [39] P. H. Weigel, P. L. DeAngelis, *J. Biol. Chem.* **2007**, *282*, 36777-36781.
- [40] T. Yada, M. Gotoh, T. Sato, M. Shionyu, M. Go, H. Kaseyama, H. Iwasaki, N. Kikuchi, Y. D. Kwon, A. Togayachi, T. Kudo, H. Watanabe, H. Narimatsu, K. Kimata, *J. Biol. Chem.* **2003**, *278*, 30235-30247.
- [41] C. Breton, L. Snajdrová, C. Jeanneau, J. Koca, A. Imberty, *Glycobiology* **2006**, *16*, 29R-37R.
- [42] N. L. Rose, R. B. Zheng, J. Pearcey, R. Zhou, G. C. Completo, T. L. Lowary, *Carbohydr. Res.* **2008**, *343*, 2130-2139.
- [43] L. J. Alderwick, L. G. Dover, N. Veerapen, S. S. Gurcha, L. Kremer, D. L. Roper, A. K. Pathak, R. C. Reynolds, G. S. Besra, *Protein Expression Purif.* **2008**, *58*, 332-341.
- [44] M. Mayer, B. Meyer, *J. Am. Chem. Soc.* **2001**, *123*, 6108-6117.

- [45] G. C. Completo, T. L. Lowary, *J. Org. Chem.* **2008**, 73, 4513-4525.
- [46] M. Mayer, B. Meyer, *Angew. Chem., Int. Ed.* **1999**, 38, 1784-1788.
- [47] M. A. Johnson, B. M. Pinto, *J. Am. Chem. Soc.* **2002**, 124, 15368-15374.
- [48] Y. Yuan, D. W. Bleile, X. Wen, D. A. R. Sanders, K. Itoh, H. W. Liu, B. M. Pinto, *J. Am. Chem. Soc.* **2008**, 130, 3157-3168.
- [49] Y. Yuan, X. Wen, D. A. R. Sanders, B. M. Pinto, *Biochemistry* **2005**, 44, 14080-14089.
- [50] B. Claasen, M. Axmann, R. Meinecke, B. Meyer, *J. Am. Chem. Soc.* **2005**, 127, 916-919.
- [51] Y. Cheng, W. H. Prusoff, *Biochem. Pharmacol.* **1973**, 22, 3099-3108.
- [52] S. Bhamidi, M. S. Scherman, C. D. Rithner, J. E. Prenni, D. Chatterjee, K. H. Khoo, M. R. McNeil, *J. Biol. Chem.* **2008**, 283, 12992-13000.
- [53] F. Pan, M. Jackson, Y. F. Ma, M. McNeil, *J. Bacteriol.* **2001**, 183, 3991-3998.
- [54] B. C. de Jong, D. M. Israelski, E. L. Corbett, P. M. Small, *Annu. Rev. Med.* **2004**, 55, 283-301.

- [55] C. A. R. Wiggins, S. Munro, *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 7945-7950.
- [56] J. C. Mathews, *Fundamentals of receptor, enzyme, and transport kinetics*, CRC press, Boca Raton, **1993**, p. 108.
- [57] A. J. Benie, R. Moser, E. Bauml, D. Blaas, T. Peters, *J. Am. Chem. Soc.* **2003**, *125*, 14-15.
- [58] L. J. Alderwick, E. Radmacher, M. Seidel, R. Gande, P. G. Hitchen, H. R. Morris, A. Dell, H. Sahm, L. Eggeling, G. S. Besra, *J. Biol. Chem.* **2005**, *280*, 32362-32371.

3.10 Supporting Information

STD-NMR Studies Suggest that Two Acceptor Substrates for GltT2, a Bifunctional Galactofuranosyltransferase Required for the Biosynthesis of Mycobacterium tuberculosis Arabinogalactan, Compete for the Same Binding Site

Monica G. Szczepina,^a Ruixiang B. Zheng,^b Gladys C. Completo,^b

Todd L. Lowary,^b and B. Mario Pinto^a

^aDepartment of Chemistry, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada

^bDepartment of Chemistry and Alberta Ingenuity Centre for Carbohydrate Science, Gunning-Lemieux Chemistry Centre, University of Alberta, Edmonton, Alberta, T6G 2G2, Canada

Table of Contents

Figure S3.1.	Expansion of 1D ^1H NMR spectrum of a mixture of trisaccharide 2 and UDP-Galf in the presence of GfT2.	122
Figure S3.2.	Expansion of 1D ^1H NMR spectrum of a mixture of trisaccharide 3 and UDP-Galf in the presence of GfT2.	123
Figure S3.3.	1D-STDD NMR spectra at different ratios of the acceptor substrates 2 to 3 .	124
Figure S3.4.	K_D calculations using the Cheng-Prusoff equation.	125

Figure S3.1. Expansion of 1D ^1H NMR spectrum of a mixture of trisaccharide **2** and UDP-Galf at 600 MHz and 298 K in the presence of GlfT2 over the course of 24 h. The asterisk denotes an anomeric signal of the formed β -(1 \rightarrow 5), β -(1 \rightarrow 6), β -(1 \rightarrow 5) tetrasaccharide (*J. Org. Chem.* **2008**, 73, 4513–4525).

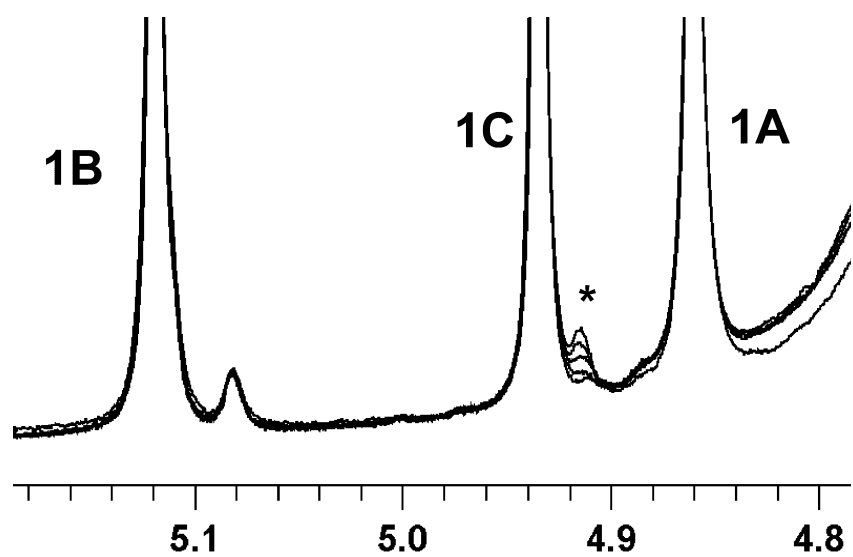


Figure S3.2. Expansion of 1D ^1H NMR spectrum of a mixture of trisaccharide **3** and UDP-Galf at 600 MHz and 310 K in the presence of GlfT2 over the course of 24h. The asterisks denote anomeric signals of the formed β -(1 \rightarrow 6), β -(1 \rightarrow 5), β -(1 \rightarrow 6) tetrasaccharide (*J. Org. Chem.* **2008**, 73, 4513–4525.)

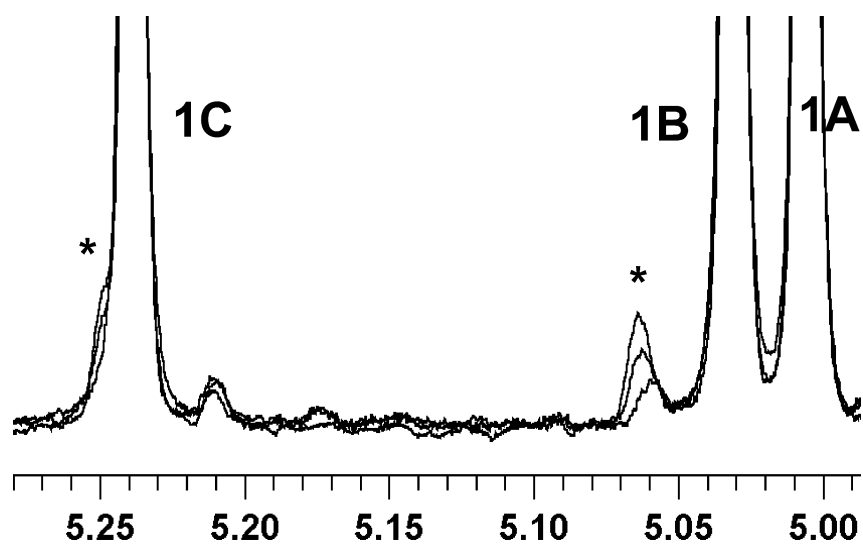


Figure S3.3. 1D-STDD NMR spectra at different ratios of the acceptor substrates **3** to **2**. The upper trace corresponds to the acceptor substrate **3** in the presence of GlfT2. Lower traces indicate spectra obtained with titration of increasing amounts of the acceptor substrate **2**.

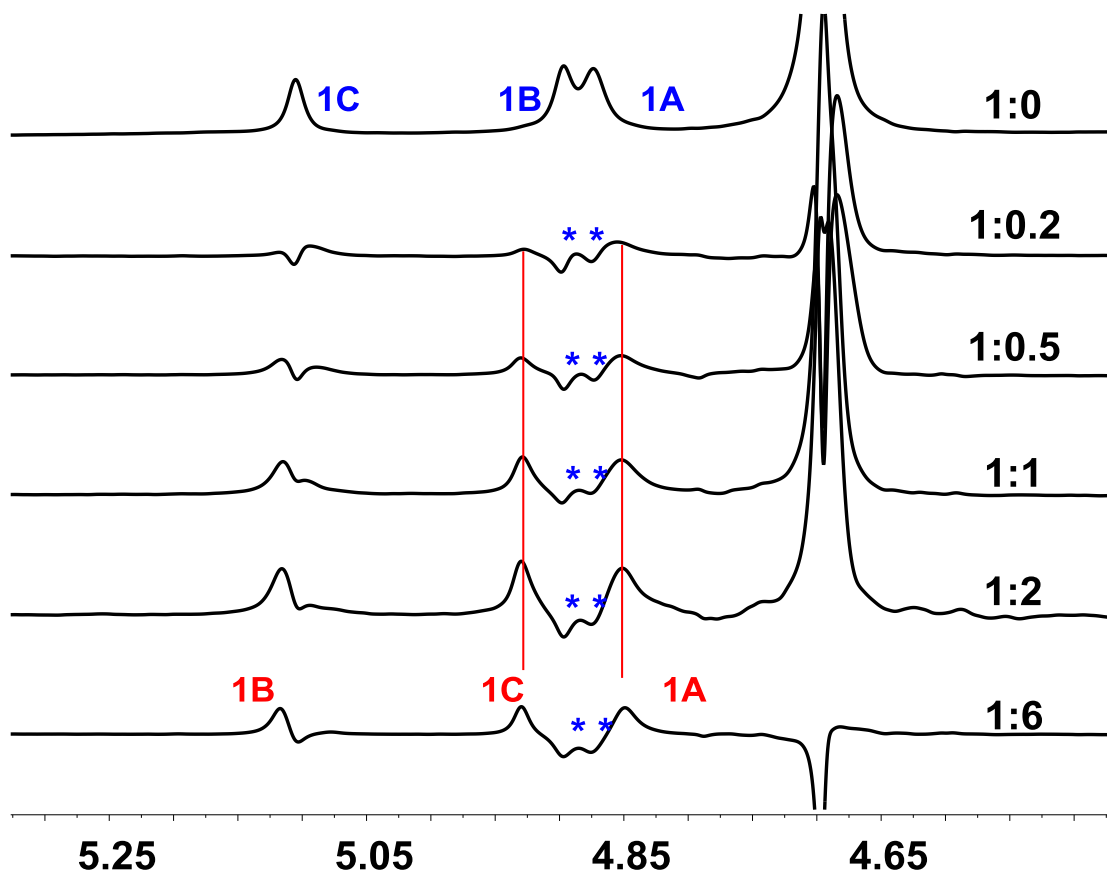


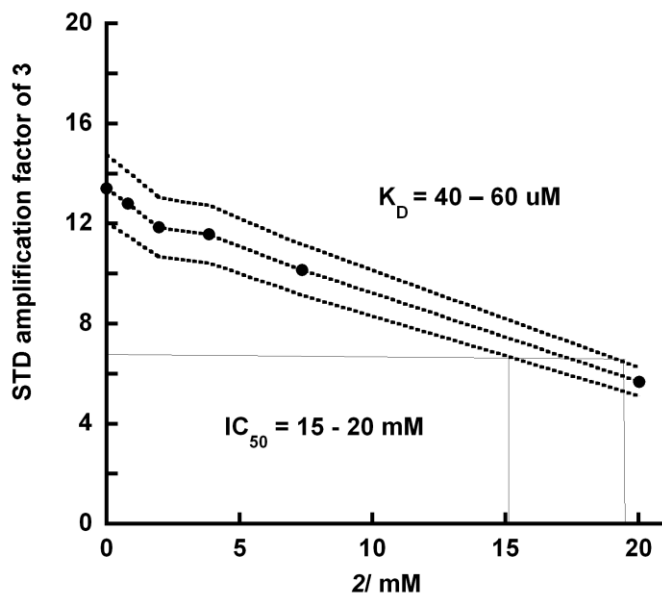
Figure S3.4. K_D calculations using the Cheng-Prusoff equation. Titration curves were fitted using Excel by a modified version of eq. (1) using the STD amplification factor:

$$I_{STD} = -100 \times I_0 / (I_0 + IC_{50}) + 100 \quad \text{eq. (1)}$$

with I_0 being the total concentration of **3**. Assuming an error of $\pm 10\%$, the dashed curves (without data points) are obtained as boundaries. This allows an estimate of IC_{50} to be in the range of 50-20 mM (see below) for **2**. The K_D values of **3** can then be calculated based on the eq. (2) (Cheng, Y.; Prusoff, W. H. *Biochem. Pharmacol.* **1973**, *22*, 3099):

$$K_D = L_0 \times K_i / (IC_{50} - K_i) \quad \text{eq. (2)}$$

L_0 , the concentration of **3** was 4mM, a K_i of 204 μM was used for **2** (K_D from kinetic experiments) and a range of 15-20 mM for the IC_{50} , gave a K_D of 40-60 μM for **3**.



All other titration data as well as cross-peaks from 2D-STD-TOCSY data were treated in an analogous manner.

CHAPTER 4: WaterLOGSY NMR Experiments Detect Immobilized Water Molecules that Bridge Peptide Mimic MDWNMHAA to Anti-carbohydrate Antibody SYA/J6

This chapter comprises the manuscript “*WaterLOGSY NMR Experiments Detect Immobilized Water Molecules that Bridge Peptide Mimic MDWNMHAA to Anti-carbohydrate Antibody SYA/J6*” which was submitted to *Biochemistry*.

Monica G. Szczepina,^a Dustin W. Bleile,^b Johannes Müllegger,^{†b}

Andrew R. Lewis,^a and B. Mario Pinto^a

^aDepartment of Chemistry, Simon Fraser University, Burnaby, British Columbia, Canada, V5A 1S6

^bZymeworks Inc., 540-1385 West 8th Ave, Vancouver, British Columbia, Canada, V6H 3V9

[†] Deceased April 23, 2009. This article is dedicated, with respect, to the memory of Johannes Müllegger.

In this chapter, we describe our attempt to capture functional water molecules that exist at the interface of the carbohydrate mimetic peptide, MDWNMHAA, and the antibody SYA-J6 against which it was raised. We use a combination of STD-NMR and WaterLOGSY spectroscopies combined with molecular dynamics simulations to detect these bound water molecules in the solution state.

Mr Dustin W. Bleile wrote python scripts to analyze the MD trajectory. Dr. Johannes Müllegger conceived and wrote scripts to find the water molecules in the MD trajectory. Dr. Andrew R. Lewis installed the WaterLOGSY pulse sequence and helped to optimize the pulse sequence. The thesis author performed all the MD simulations, NMR experiments and the data analysis.

4.1 Keywords

Saturation Transfer Difference (STD)-NMR, Molecular Dynamics, Water-Ligand Observed *via* Gradient Spectroscopy (WaterLOGSY), Peptide Carbohydrate Mimicry, *Shigella flexneri* Y

4.2 Abstract

X-ray crystallographic data of the carbohydrate mimic MDWNMHAA when bound to an anti-*Shigella flexneri* Y mAb SYA/J6 indicate the immobilization of water molecules, i.e. the presence of “bound” water molecules, in the active site. Water Ligand Observed via Gradient Spectroscopy (WaterLOGSY) was used in conjunction with saturation transfer difference (STD)-NMR spectroscopy to probe the existence of immobilized water molecules in the complex of MDWNMHAA **1** bound to mAb SYA/J6. Molecular dynamics simulations using the ZymeCAD™ Molecular Dynamics platform were then used to specify the likely locations of these water molecules. Of note, those water molecules involved in providing complementarity between the peptide and mAb SYA/J6 remained throughout the course of the simulation. Together, the experimental and computational protocols have been used to identify the bound water molecules present in the antibody-peptide complex.

4.3 Introduction

The Gram negative bacterium *Shigella flexneri* Y causes shigellosis.^[1] The O-polysaccharide on the surface of the bacterium is an extensive $[\rightarrow 2)\text{-}\alpha\text{-L-Rhap-(1}\rightarrow 2)\text{-}\alpha\text{-L-Rhap-(1}\rightarrow 3)\text{-}\alpha\text{-L-Rhap-(1}\rightarrow 3)\text{-}\beta\text{-D-GlcpNAc-(1}\rightarrow]$ biopolymer (Figure 4.1)^[2-6] that can serve as an antigen for vaccine development. An immunization strategy that relies on carbohydrate-mimetic peptide vaccines has been ongoing in our laboratory.^[7, 8] Screening of a panel of phage-displayed peptide libraries with the anti-*S. flexneri* Y monoclonal antibody, SYA/J6, directed against the O-polysaccharide, isolated the carbohydrate mimic MDWNMHAA 1 (peptide 1) (Figure 4.2).^[9]

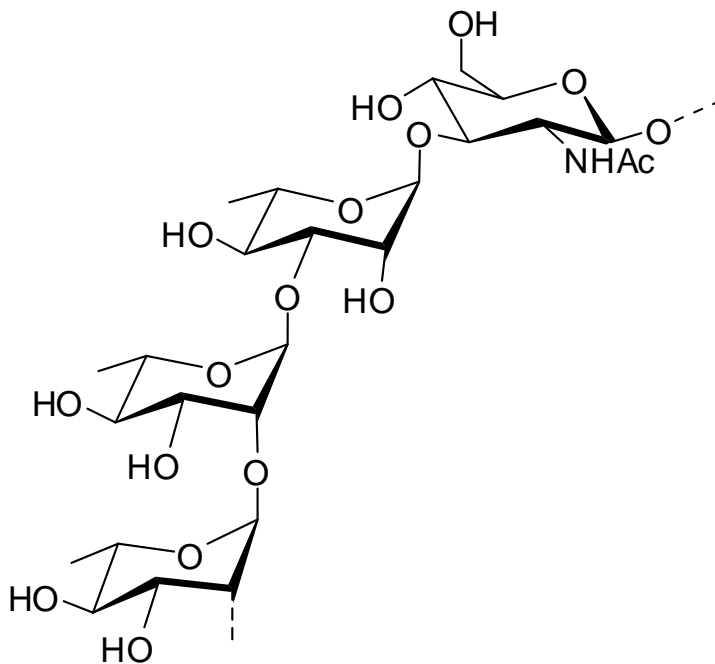


Figure 4.1. Structure of the *Shigella flexneri* Y O-antigen polysaccharide.

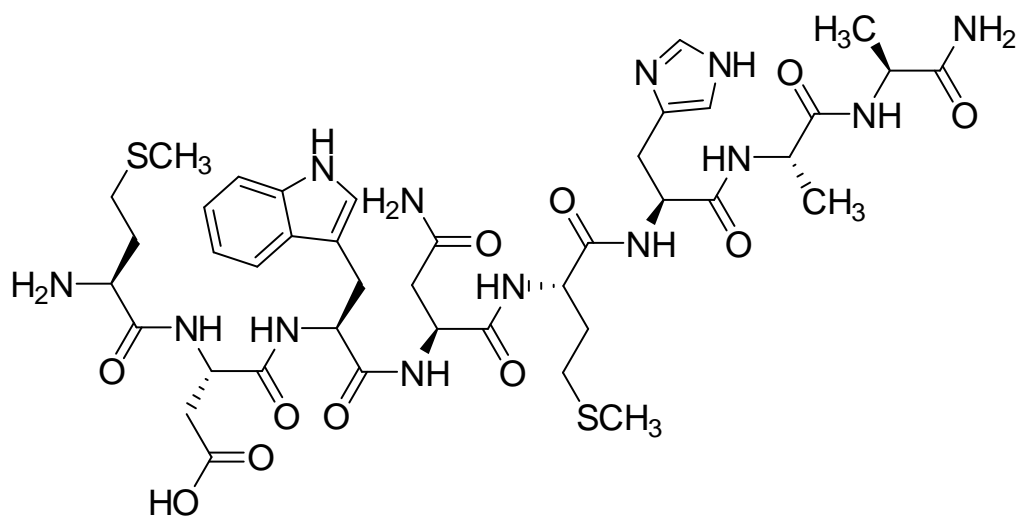


Figure 4.2. Structure of the carbohydrate-mimetic peptide MDWNMHAA 1 (peptide 1).

High resolution X-ray crystal structures of peptide 1 and the pentasaccharide hapten α -L-Rhap-(1 \rightarrow 2)- α -L-Rhap-(1 \rightarrow 3)- α -L-Rhap-(1 \rightarrow 3)- β -D-GlcpNAc-(1 \rightarrow 2)- α -L-Rhap-(1 \rightarrow OME)] bound by the Fab fragment of SYA/J6 (Figure 4.3),^[10, 11] showed that the binding modes of the two ligands differ significantly, with MDWNMHAA 1 displaying elements of functional mimicry, as reviewed recently.^[12] Briefly, whereas the peptide contacts some of the same areas of the binding site as the oligosaccharide, it also extends into other areas of the site. The first four residues, Met P1–Asp P2–Trp P3–Asn P4, adopt an extended conformation, and the last four residues, Met P5–His P6–Ala P7–Ala P8, form one turn of α -helix. The α -helical turn of residues Met P5–Ala P8 exposes the last four peptide side chains for specific interactions, most of which differ from those in the pentasaccharide complex. The peptide residues Trp

P3–Asn P4–Met P5 lie in the center of the groove, roughly coincident with the sugar residues Rha C and GlcNAc D. However, the peptide does not directly contact the central, deepest part of the groove, where Rha C binds but rather, three water molecules occupy this cavity (Figure 4.3). The indole ring of Trp P3 lies in a hydrophobic pocket formed by residues from CDR L3 (Val L94, Pro L95); CDR H2 (His H58) and a framework residue (Trp H47, FR2); the pentasaccharide does not enter this pocket. Overall, the peptide molds itself more readily to the topography of the combining site, making 126 intermolecular contacts, compared to the pentasaccharide's 74. It is noteworthy that despite the differences in peptide versus oligosaccharide binding, a protein conjugate of the peptide ligand is immunogenic and elicits an immune response that is cross-reactive with the bacterial polysaccharide.^[8] Thus, the peptide *is* an antigenic mimic of the polysaccharide.

The binding of the pentasaccharide was characterized thermodynamically by $\Delta H = -1.5$ kcal/mol and $-T\Delta S = -5.9$ kcal/mol. On the other hand, the mimetic peptide binding was strongly enthalpy driven $\Delta H = -16.9$ kcal/mol, but was offset by an entropy term of $-T\Delta S = +9.1$ kcal/mol.^[10, 11] Consequently, the octapeptide binds with approximately equal affinity ($K_A = 5.7 \times 10^5 \text{ M}^{-1}$) as the pentasaccharide ($K_A = 2.5 \times 10^5 \text{ M}^{-1}$). The NMR spectrum of the octapeptide in the presence of the SYA/J6 antibody has been assigned previously.^[13] The quest for higher affinity ligands relies on a detailed understanding of the nature of mimicry. The octapeptide is a functional mimic of the natural pentasaccharide as it binds to the antibody SYA/J6 (and generates a cross-reactive immune

response), yet this results in a severe entropic penalty, presumably caused by (1) the reduction of conformational dynamics upon binding, (2) the introduction of an α -helical turn, and (3) the immobilization of water molecules, i.e. the presence of “bound” water molecules.

In order to probe the immobilization of water molecules at the peptide-antibody interface upon binding in solution, and to provide a more complete understanding of the complementarity between peptide and antibody, we report herein a detailed study of the portions of the octapeptide that have direct contact to immobilized water by WaterLOGSY^[14-18] and STD-NMR spectroscopy,^[19] in conjunction with molecular dynamics simulations.

To date, the use of WaterLOGSY spectroscopy has been targeted towards screening and understanding ligand-receptor interactions.^[14-18, 20-29] In addition, recent work reported in the laboratory of Günther *et al.* has shown that WaterLOGSY can be used to map bound ligand orientations by using solvent accessibility in a new method known as SALMON (Solvent Accessibility and protein Ligand binding studied by NMR Spectroscopy).^[25, 26, 30] Short mixing times are employed to minimize magnetization transfer from the protein to ligand, to allow solvent-exposed protons to retain properties of their unbound state, and thus create a solvent accessibility epitope.^[25, 26, 30] In these studies, the authors compared the solvent accessibility epitope maps to epitope maps obtained from STD-NMR studies which utilize longer mixing times, and therefore result in spin diffusion throughout the protein and subsequent magnetization transfer to ligand.

We present here a study in which we use identical (long) mixing times for both WaterLOGSY and STD-NMR experiments.

The use of the WaterLOGSY experiment alone is not enough to get information about the number of the water-molecules but coupled with STD NMR experiments it is possible to distinguish between direct and indirect transfer of magnetization. In the case of the WaterLOGSY experiment, the transfer of magnetization can be either i) water \rightarrow protein \rightarrow ligand (indirect) or ii) bound water \rightarrow ligand (direct). The STD-NMR experiment detects protein \rightarrow ligand. Pathways of the type protein \rightarrow bound water \rightarrow ligand are not detected in the STD NMR experiment because water molecules have off times k_{off} ($>10^{10} \text{ s}^{-1}$) that are much faster than ligand k_{off} ($>100 \text{ s}^{-1}$). Thus, the absence of a resonance in the STD-NMR experiment coupled with the presence of the same resonance in the WaterLOGSY experiment would indicate that magnetization transfer had been the result of a bound water molecule. The residence times of buried water molecules have lifetimes on the order of 10 nanoseconds to 0.01 seconds ^[31] which makes these molecules amenable to investigation by NMR spectroscopy (timescale on the order of μs to ms) and MD simulation (ns). This information can then be combined with molecular dynamics simulations to identify the likely locations of the water molecules in the complex through internal consistency.

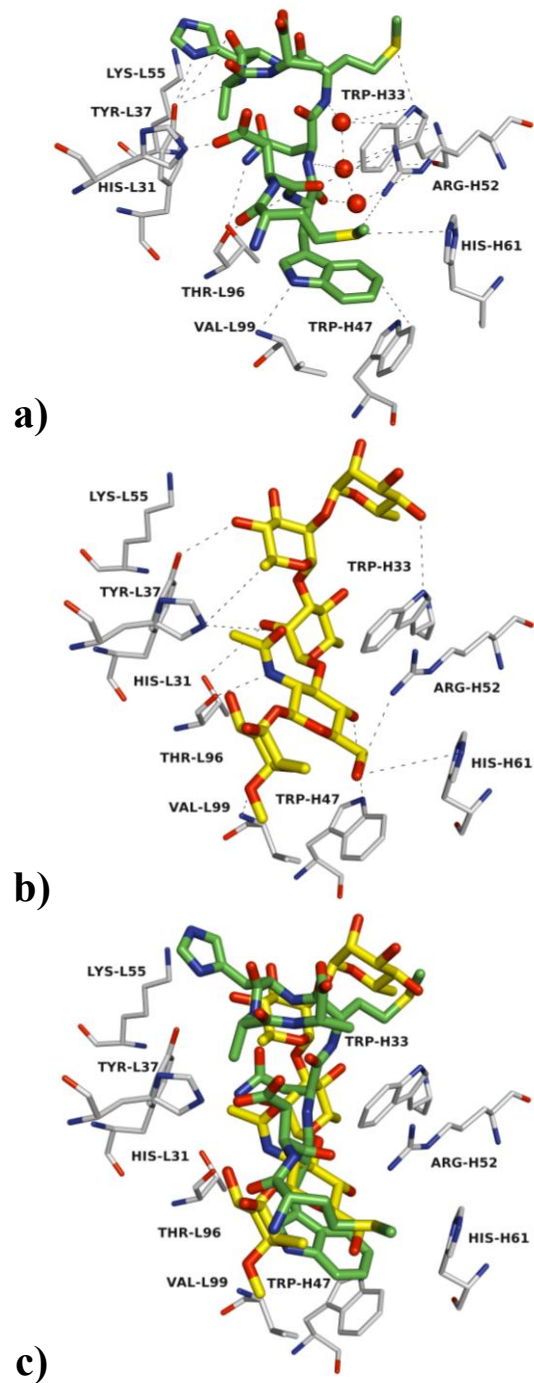


Figure 4.3. X-ray crystal structures of a Fab fragment of the monoclonal antibody SYA/J6 (IgG3, κ), directed against the O-antigen of *Shigella flexneri* Y, complexed with a) peptide 1 (green), b) pentasaccharide hapten [-α-L-Rhap-(1→2)-α-L-Rhap-(1→3)-α-L-Rhap-(1→3)-β-D-GlcpNAc-(1→2)-α-L-Rhap-(1→OMe)] (yellow), and c) superposition of the two ligands in the combining site. H and L refer to heavy and light chains, respectively.

4.4 Experimental

4.4.1 Materials

MDWNMHAA **1** was synthesized according to the method of Hossany *et al.* ^[32] The mAb SYA/J6 antibody was a generous gift from D.R. Bundle.

4.4.2 NMR Spectroscopy of peptide **1**

An NMR sample of the mAb SYA/J6 (13.5 μ M, 27 μ M binding sites) was prepared in phosphate buffered saline (12 mM $\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$, 137 mM NaCl, 3 mM KCl, 0.02% NaN_3 , pH 6.1) in a Shigemi tube containing 10% D_2O and peptide **1** (0.45 mg) was added to the sample for a final peptide concentration of 1.5 mM and a ratio of 100:1 peptide:antibody. The STD-NMR spectra were recorded at 298 K with 4096 scans and selective saturation of protein resonances at -2 ppm (30 ppm for off resonance spectra) using a series of 40 Gaussian-shaped pulses (50 ms, 1 ms delay between pulses, $\gamma B_1/2\pi = 110$ Hz), for a total saturation time of 2.0 s. The protein resonances were broad and had significant intensity in the region downfield from 10 ppm and even at negative parts per million values. Thus, irradiation at -2 ppm was expected to result in saturation of aliphatic protein resonances, and via rapid spin diffusion, to aromatic protein resonances. A 10 ms spin-lock pulse ($\gamma B_1/2\pi = 11$ kHz) was applied after excitation to reduce the intensity of broad protein resonances. The saturated and reference spectra were acquired in the same dataset by creating a pseudo-2D experiment. The STD spectrum was obtained by subtraction of saturated spectra from reference spectra after identical processing and phasing. Water suppression was achieved with the excitation sculpting sequence.

WaterLOGSY NMR experiments employed a 36 ms selective square 180° pulse at the water signal frequency and an NOE mixing time of 2s. The WaterLOGSY spectra were recorded at 298 K with 2048 scans and were corrected by running a peptide-only WaterLOGSY control spectrum. All spectra were acquired on the Bruker Avance 600 NMR spectrometer equipped with a TCI cryoprobe.

4.4.3 Molecular Dynamics Simulations

Crystal-structure coordinates of the protein receptor and ligand were used as a starting structure for the complex (PDB entry 1PZ5). Molecular dynamics simulations were performed on a custom MD engine (ZymeCAD™) using the AMBER force field version 99sb in explicit solvent composed of TP3 water molecules. A 1.5 fs time step was used for integration, with hydrogen bond lengths constrained by the RATTLE algorithm^[33]. The SETTLE algorithm^[34] was used to constrain water molecules to their equilibrium geometry. Non-bonded interactions were treated with a shifting function to scale interaction energies to zero at 12 Angstroms. Simulations were run under NPT conditions of 300 Kelvin and 1 atm pressure. The temperature of the simulation and simulation pressure were controlled by a Berendsen thermostat^[35]. Water molecules were examined for simultaneous contact to both peptide **1** and antibody. These water molecules were ranked according to how long they maintained their bound, low RMSD state. Only those maintaining contact to both peptide **1** and antibody for least 5ns were inspected visually and are shown in Figure 4.4.

4.5 Results and Discussion

4.5.1 Molecular dynamics studies of peptide 1 in the Fv portion of monoclonal antibody SYA/J6

Molecular dynamics studies of the peptide 1 bound to the Fv portion of mAb SYA/J6 began with the crystal-structure coordinates for the complex (PDB entry 1PZ5). An inspection of the MD simulation of peptide 1 bound to the Fv portion of the antibody uncovered the presence of bound water molecules. Water molecules were examined for simultaneous contact to both peptide 1 and antibody. These water molecules were ranked according to how long they maintained their bound, low RMSD state. Only those maintaining contact to both peptide 1 and antibody for least 5 ns were inspected visually and are shown in the left hand side of Figure 4.4. Four of these “resident” water molecules maintained contact to both peptide 1 and antibody for the whole 10.5 ns duration of the MD simulation, and occupied the same position in the X-ray crystal structure (shown on the right hand side of Figure 4.4). In particular, three water molecules remained in a cavity at the bottom of the combining site groove throughout the simulation. Of note, this pocket is not penetrated by the peptide. In the case of the carbohydrate, this pocket is filled by Rha C of the pentasaccharide. These simulations provide further evidence that these three water molecules, along with part of the peptide backbone (from C=O of Trp-3 to NH of Met-5), constitute elements of shape mimicry i.e. they mimic the shape of Rha C. ^[10]

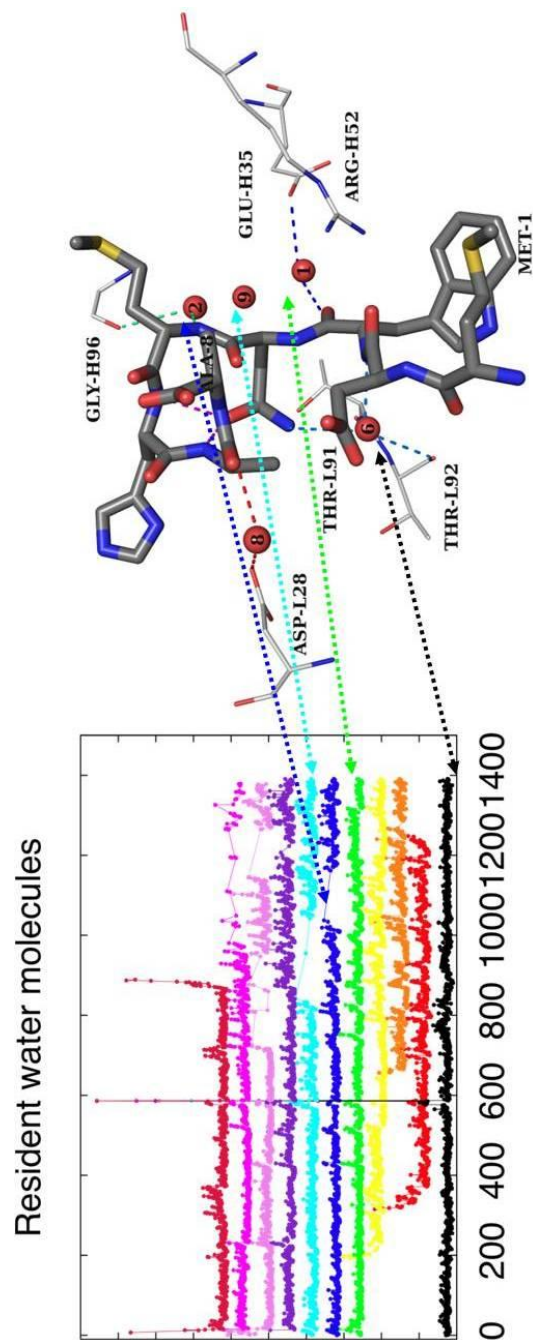


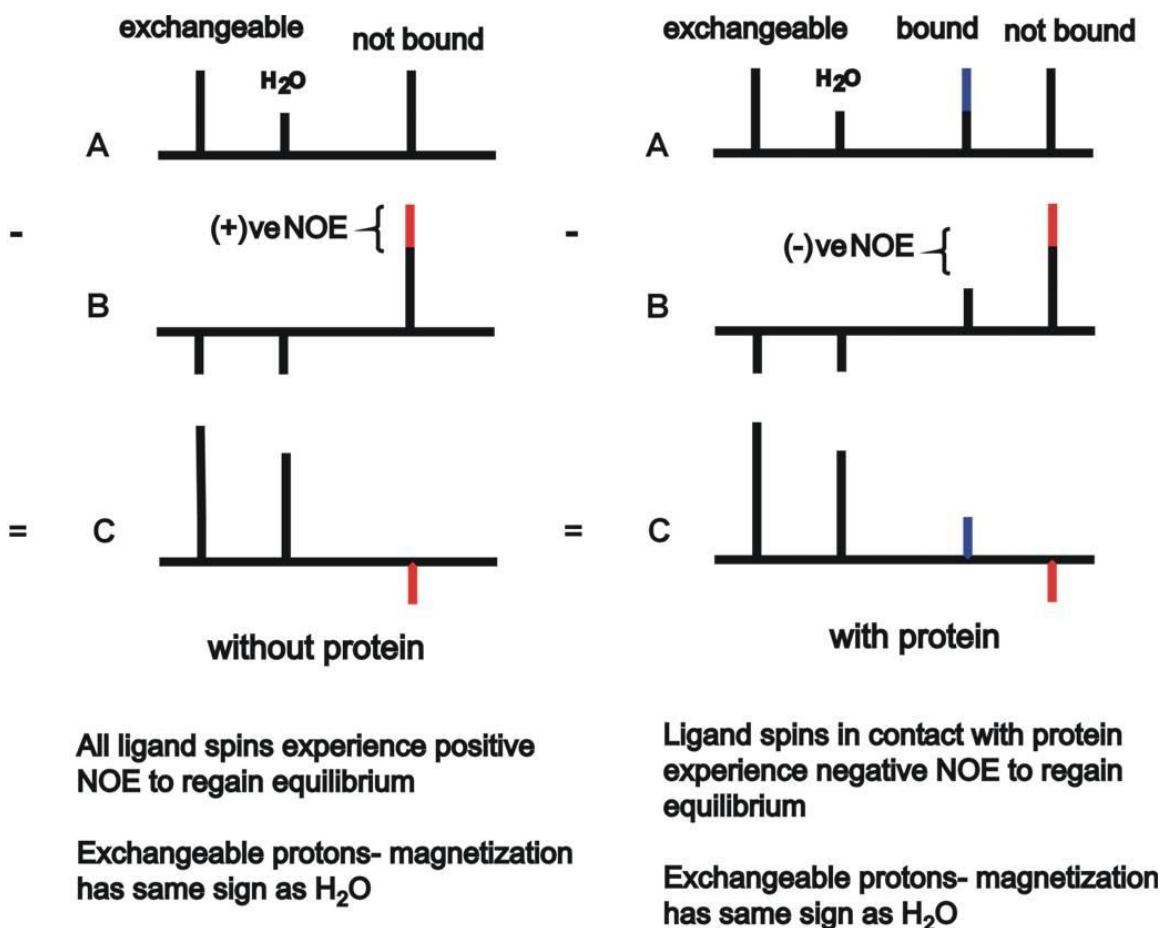
Figure 4.4. Resident water molecules identified from MD simulation of peptide **1** bound to Fv. The x-axis refers to time (one frame = 0.0075ns. 1400 total frames = 10.5 ns trajectory) and the y-axis is arbitrary. These water molecules were mapped onto the X-Ray crystal structure (shown on the right). The water molecules found during the MD simulation were present for at least 5 ns (shown

on the left) and those that were present for the entire trajectory (10.5 ns) were mapped onto the X-Ray crystal structure and correspond to water molecules 1, 2, 6 and 9 (shown on the right).

4.5.2 WaterLOGSY experiments of peptide 1 in the combining site of monoclonal antibody SYA/J6

The WaterLOGSY (Water Ligand Observed via Gradient SpectroscopY) ^[14-18, 20-30] experiment relies on the selective inversion of bulk water; the magnetization is then transferred to the bound ligand via the ligand-protein complex. Magnetization transfer can occur via different pathways. ^[36] These include: (i) Magnetization transfer via “bound” water molecules (lifetimes on the order of ns to μ s) existing at interstitial sites between ligand and protein; the NOE resulting from this transfer is large and negative. (ii) Chemical exchange between bulk water and the protein receptor’s exchangeable NH and OH groups within the binding site as a conduit for magnetization transfer; the protein receptor then transfers this magnetization to the ligand via negative cross-relaxation, again resulting in a negative NOE. (iii) Chemical exchange between bulk water and remote (away from the binding site) exchangeable NH and OH groups on the protein followed by spin diffusion transfers negative NOE throughout the protein and to the bound ligand; exchangeable protons on the ligand as well as non binding components experience positive cross-relaxation (positive NOE) with water. The WaterLOGSY experiment can be thought of as a difference experiment with the net effect on a ligand’s magnetization, as depicted in

Scheme 1. The left panel depicts the net effect on the free ligand magnetization and the right depicts the net effect on the bound ligand's magnetization. In both cases, exchangeable protons have positive WaterLOGSY peaks, as shown in spectrum C. In the case of binding to a receptor, ligand resonances are also positive, as shown in spectrum C. If there is no binding or in the absence of protein, ligand WaterLOGSY peaks are negative.



Scheme 4.1. Schematic representations of WaterLOGSY spectra of a hypothetical ligand in the absence (left) and presence (right) of a protein receptor. State A refers to the equilibrium state, or immediately after off-resonance irradiation. State B refers to irradiation at the water frequency. State C is the WaterLOGSY spectrum and is the difference between states A and B.

As described above, when used in combination with STD-NMR data, the WaterLOGSY experiment can provide insight into the presence of water molecules existing at interstitial sites between the protein and the ligand. STD-NMR spectroscopy relies on irradiation of the protein resonance followed by spin diffusion to the ligand. Since water molecules have off times k_{off} ($>10^{10} \text{ s}^{-1}$) that are much faster than ligand k_{off} ($>100 \text{ s}^{-1}$), magnetization transfer is essentially from protein to ligand and magnetization transfers from protein to bound water to ligand are not detected i.e. in STD-NMR spectroscopy one observes protein \rightarrow ligand pathways and *not* protein \rightarrow bound water \rightarrow ligand pathways. In the case of the WaterLOGSY experiment, water molecules are irradiated and the transfer of magnetization can be either i) water \rightarrow protein \rightarrow ligand or in the case of bound water molecules, ii) bound water \rightarrow ligand. In this manner, the absence of a resonance in the STD-NMR experiment coupled with the presence of the same resonance in the WaterLOGSY experiment would indicate that magnetization transfer had been the result of a bound water molecule. Direct proton exchange from irradiated water molecules to the ligand's exchangeable protons is corrected by the subtraction of a WaterLOGSY spectrum of *ligand only*. The caveat to this argument is that the environment within the antibody combining site may affect the pKas of the NH protons of the ligands and that subtraction of the ligand-only effects will not be completely representative.

The binding of peptide **1** to mAb SYA/J6 was therefore investigated using WaterLOGSY NMR spectroscopy. The WaterLOGSY NMR spectra are shown in Figure 4.5. The upper trace corresponds to the peptide **1**-only control

experiment. The resonances that correspond to exchangeable protons exhibit a positive phase. All other resonances show negative phase, indicative of non-bound, as described in the left hand side of Scheme 4.1. The lower trace in Figure 4.5 corresponds to **1** in the presence of the antibody. Once again, resonances that arise from exchangeable protons exhibit a positive phase (protons on the peptide **1** that are bound by the antibody will show a positive phase, whereas protons that are non-binders will be of opposite sign). It is of interest to note that the methyl resonances of Ala-7 and Ala-8 (at 1.25ppm), show a false negative (they are phased negatively) result. This is observed typically in a ligand excess regime where the free ligand cross correlation time, σ_{free} , can overwhelm the bound ligand cross correlation time.^[36] This is easily corrected by subtracting the ligand-only WaterLOGSY experimental results from those of the bound state, as shown in Figure 4.6. The two methyl signals corresponding to Ala-7 and Ala-8 are now of positive phase, indicative of binding, in keeping with the right-hand panel of Scheme 4.1.

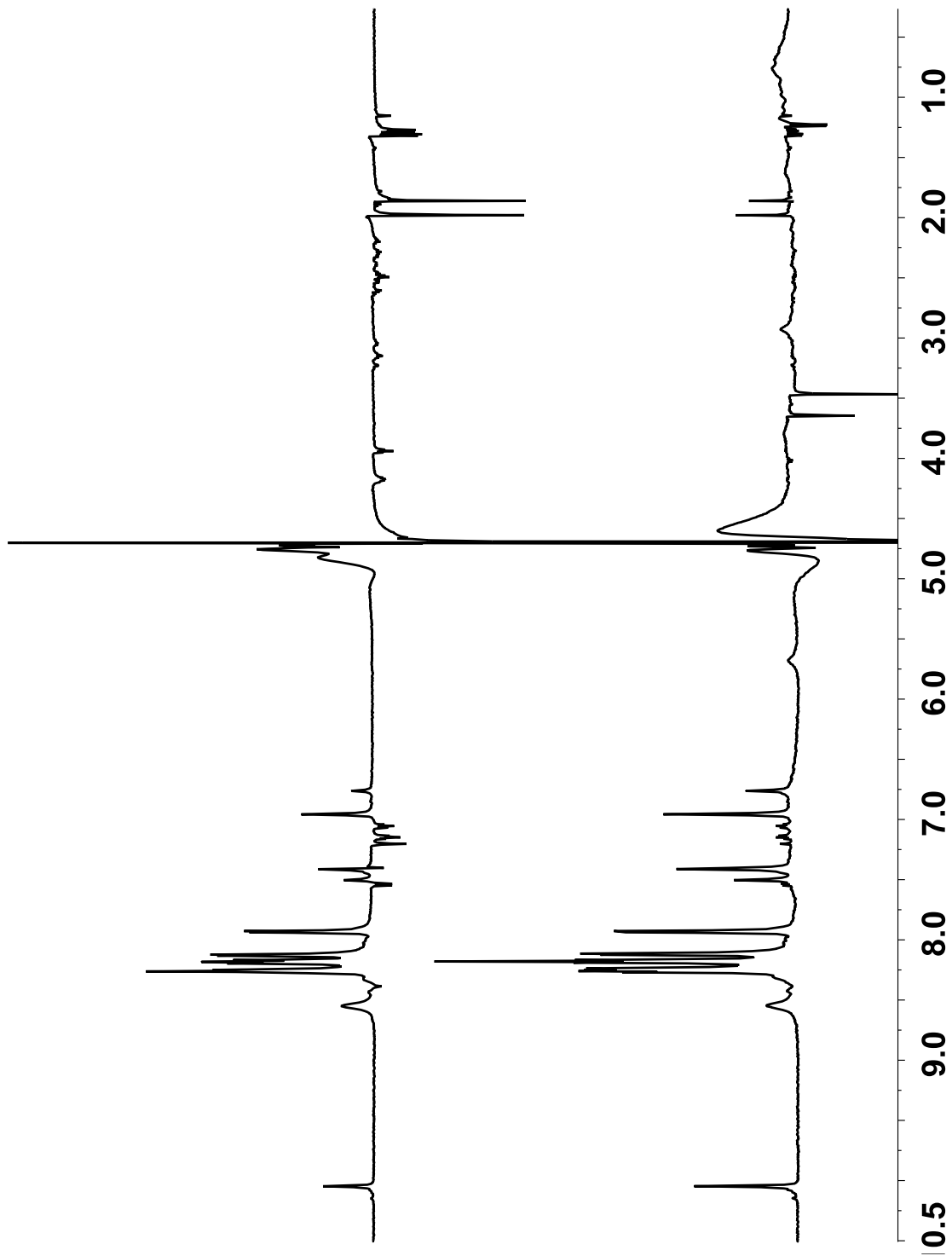


Figure 4.5. 1D WaterLOGSY NMR spectra of peptide **1** only (upper trace) and peptide **1** in the presence of mAb SYA/J6 (lower trace), at 600 MHz and 298 K.

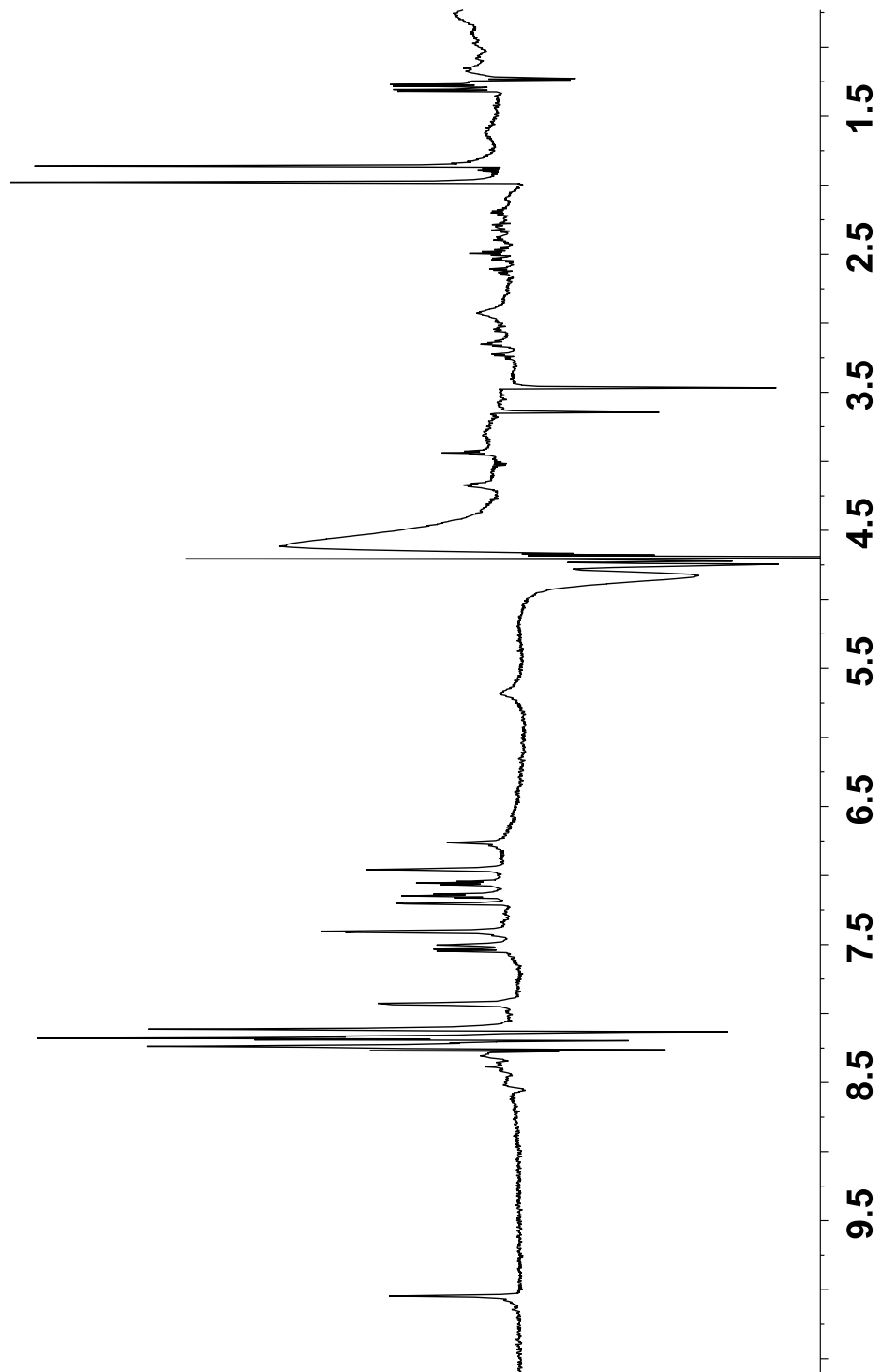


Figure 4.6. 1D WaterLOGSY NMR spectrum of peptide **1** in the presence of mAb SYA/J6, at 600 MHz and 298 K, corrected for ligand-only effects by the subtraction of the upper trace from the lower trace in Figure 4.5.

4.5.3 Analysis of WaterLOGSY and STD-NMR experiments of peptide 1 in the combining site of monoclonal antibody SYA/J6

It was of interest to compare the WaterLOGSY data to the data acquired by STD-NMR spectroscopy. In Figure 4.7, we show an expansion of the region between 6.7 ppm and 8.7 ppm (the full spectrum is shown in Figure S4.1.) that depicts the STD-NMR spectrum (lower trace) and the corrected WaterLOGSY spectrum (upper trace) for peptide 1 in the presence of mAb SYA/J6. Firstly, enhancements are observed in both Asn-4 H δ 2 proton resonances in the WaterLOGSY experiment while being absent in the STD-NMR experiment. This result indicates that the Asn-4 δ 2 hydrogens experience magnetization transfer that originated from a bound water molecule. This enhancement can be explained by examination of both the X-ray crystal structure and the median structure from the molecular dynamics simulation that indicates Asn-4 contacts with L91 Thr and L92 Thr via a water molecule (see blue contacts to water number 6 in Figure 4.8). In the NH region, enhancements are observed for Met-5 NH in the WaterLOGSY spectrum while absent in the corresponding STD NMR spectrum. The Met-5 NH proton makes contacts with H96 Gly via a water molecule (X-ray structure) (see green contacts to water number 2 in Figure 4.8) and to H35 Glu via three water molecules (see water numbers 1, 2 and 9 in Figure 4.8). In the median structure, this triad has loosened up but the water molecule is still present to bridge between Met-5 NH and the glycine (Figure 4.8). We also observe an enhancement of the Trp-3 NH in the WaterLOGSY NMR spectrum and note its absence in the STD NMR spectrum. Again, this is not surprising because that proton is involved in contacts to both L91 Thr and L92

Thr residues via a water molecule (X-ray structure) (see blue contacts to water number 6 in Figure 4.8). These interactions are maintained in the median structure from MD. We note also that His-6 NH and Ala-7 NH are enhanced only in the WaterLOGSY experiment and yet they are not involved in direct contacts between a bound water molecule and the antibody. However, it is interesting to note that they are both involved in *intra*-molecular hydrogen bonds to Asn-4 $\delta 2$ protons which are in direct contact with L91 Thr and L92 Thr residues via water molecule number 6 in both the X-ray structure and the MD median structure (see magenta and blue contacts in Figure 4.8). The Asp-2 NH resonance is absent in both experiments (the proton spectrum shows a very weak Asp-2 NH peak which is possibly the result of water suppression). The resonance corresponding to Asn-4 NH and the Ala-8 NH resonance are both enhanced in the WaterLOGSY experiment, whilst silent in the STD-NMR experiment, but yet they do not make contacts to the antibody via bound water molecules. However, their corresponding peptide bond partners, the carbonyl CO oxygens, are involved in interacting with the antibody via a water molecule. Thus, Ala-7 O interacts with L28 Asp through a water molecule both in the X-ray structure and in the median structure (see red contacts to water number 8 in Figure 4.8). Similarly, Trp-3 O interacts with H35 Glu through a water molecule, as observed in X-ray and MD structures (see dark blue contacts to water number 1 in Figure 4.8) and to H96 Gly via three water molecules (see water numbers 1, 2 and 9 in Figure 4.8) . Thus, it appears that contacts via either the nitrogen or oxygen atom of a peptide bond to a water molecule that bridges to the antibody can be detected by these

experiments. With the last two examples, it is possible, however, that the effects observed may be due to direct NH exchange on the ligand which has been modulated upon binding to antibody and therefore not fully evaluated in the ligand-only WaterLOGSY experiment or that spin diffusion from one proton to another on the ligand leads to indirect transfer from water. Enhancements that are common to both the WaterLOGSY and STD NMR experiments, such as Ala-7 CH₃ and Trp-3 H δ 1, cannot be attributed to solely arising from bound water magnetization transfer. We attribute the slight “antiphase” behaviour in the NH region of the control-corrected WaterLOGSY spectrum to chemical shift differences of NH resonances in the Ab and peptide experiment relative to the peptide-only experiment due to the slightly different environments in the two media.

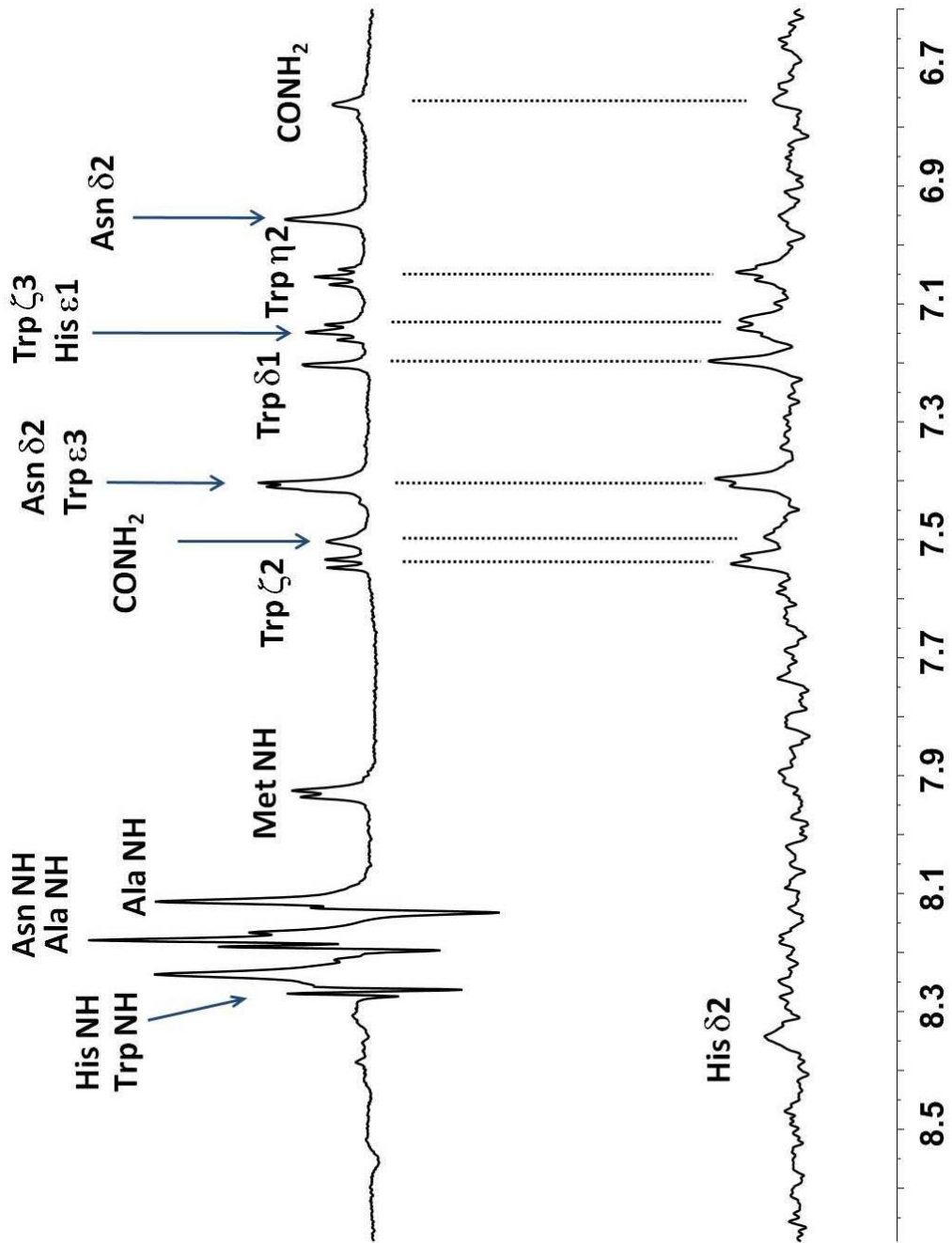


Figure 4.7. Expansion of the 1D WaterLOGSY NMR spectrum of peptide 1 in the presence of mAb SYA/J6, at 600 MHz and 298 K, corrected for ligand-only effects (upper trace), and expansion of the 1D STD-NMR spectrum of the same sample under identical conditions (lower trace).

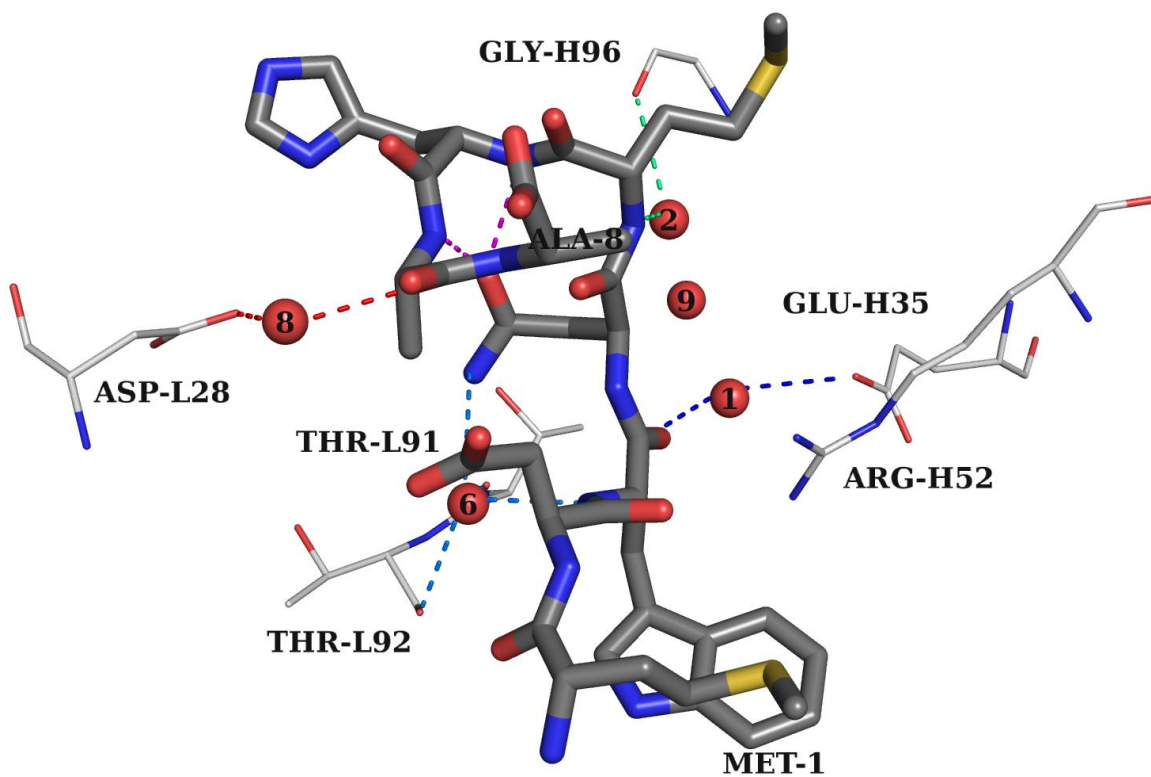


Figure 4.8. Mapping of the water molecules 1, 2, 6, 8, and 9, identified from combined WaterLOGSY and STD-NMR experiments and molecular dynamics simulations, onto the X-Ray structure of peptide **1**, showing the proposed water lattice that mediates contacts between the peptide and the antibody.

4.6 Conclusions

We have shown that it is possible to detect bound water molecules in the complex of MDWNMHAA **1** and the anti-*Shigella flexneri* Y mAb SYA/J6. Three water molecules in particular were first detected by X-ray crystallography and appear to confer shape mimicry to the peptide ligand for a sugar moiety. A severe entropic penalty, $-T\Delta S = +9.1$ kcal/mol, for this system hinted that water molecules were immobilized on binding. This prompted us to begin a search for these “bound” water molecules in the solution state.

¹D WaterLOGSY NMR experiments, when used in conjunction with STD-NMR experiments and molecular dynamics simulations, led to the observation of magnetization which had *originated* from bound water molecules and been subsequently transferred to the bound peptide **1**, thus identifying ligand protons that form hydrogen bonds with water molecules which are *themselves* hydrogen bonded to a receptor. Five unique water molecules that served as a conduit for magnetization transfer upon irradiation at the water frequency were identified.

4.7 Acknowledgements

We are grateful to the Natural Sciences and Engineering Council of Canada for financial support and to the Accelerate BC-MITACS program and Zymeworks Inc. for internships to MGS. We are also grateful to D.R. Bundle for the generous gift of the SYA/J6 antibody.

4.8 Summary

In this chapter we have described the use of Water Ligand Observed via Gradient Spectroscopy (WaterLOGSY) in conjunction with saturation transfer difference (STD)-NMR spectroscopy to experimentally probe the immobilized water molecules that exist at the interstitial sites between MDWNMHAA 1 and mAb SYA/J6. This chapter has also illustrated how molecular dynamics simulations of the peptide bound to Fv uncovered the presence of bound water molecules. In both methodologies our findings that some of these “resident” water molecules occupy the same position in the X-ray crystal structure are described. In the next chapter, the binding of MDWNMHAA 1 to SYA/J6 is examined using full STD-NMR build-up curves to capture the epitope of the peptide. CORCEMA-ST calculations to quantify these STD NMR data are also described. We also highlight the use of molecular dynamics to capture an accurate picture of the conformational flexibility and the possibilities for bound-ligand conformations in the binding of MDWNMHAA 1 to SYA/J6.

4.9 References

- [1] T. L. Hale, in *Topley and Wilson's Microbiology and Microbial Infections, Vol. 3* (Eds.: W. J. Hansler, M. Shuman), Hodder Arnold, London, **1998**, p. 479.
- [2] L. Kenne, B. Lindberg, K. Petersson, E. Katzenellenbogen, E. Romanowska, *Eur. J. Biochem.* **1977**, *76*, 327-330.
- [3] L. Kenne, B. Lindberg, K. Petersson, E. Katzenellenbogen, E. Romanowska, *Eur. J. Biochem.* **1978**, *91*, 279-284.
- [4] L. Kenne, B. Lindberg, K. Petersson, E. Romanowska, *Carbohydr. Res.* **1977**, *56*, 363-370.
- [5] D. R. Bundle, M. A. J. Gidney, S. Josephson, H. P. Wessel, *ACS Symp. Ser.* **1983**, *231*, 49-63.
- [6] N. I. A. Carlin, M. A. J. Gidney, A. A. Lindberg, D. R. Bundle, *J. Immunol.* **1986**, *137*, 2361-2366.
- [7] S. Borrelli, R. B. Hossany, S. Findlay, B. M. Pinto, *Am. J. Immunol.* **2006**, *2*, 73-83.
- [8] S. Borrelli, R. B. Hossany, B. M. Pinto, *Clin. Vaccine Immunol.* **2008**, *15*, 1106-1114.

- [9] S. L. Harris, L. Craig, J. S. Mehroke, M. Rashed, M. B. Zwick, K. Kenar, E. J. Toone, N. Greenspan, F. I. Auzanneau, J. R. Marino-Albernas, B. M. Pinto, J. K. Scott, *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 2454-2459.
- [10] N. K. Vyas, M. N. Vyas, M. C. Chervenak, D. R. Bundle, B. M. Pinto, F. A. Quioco, *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 15023-15028.
- [11] N. K. Vyas, M. N. Vyas, M. C. Chervenak, M. A. Johnson, B. M. Pinto, D. R. Bundle, F. A. Quioco, *Biochemistry* **2002**, *41*, 13575-13586.
- [12] M. A. Johnson, B. M. Pinto, in *Topics in Current Chemistry: Bioactive Conformation II, Vol. 273* (Ed.: T. Peters), Springer-Verlag, Heidelberg, Germany, **2008**, pp. 55-116.
- [13] M. A. Johnson, B. M. Pinto, *Bioorg. Med. Chem.* **2004**, *12*, 295-300.
- [14] C. Dalvit, *J. Magn. Reson., Ser B* **1996**, *112*, 282-288.
- [15] C. Dalvit, *J. Biomol. NMR* **1998**, *11*, 437-444.
- [16] C. Dalvit, *Concepts Magn. Reson. A* **2008**, *32A*, 341-372.
- [17] C. Dalvit, G. Fogliatto, A. Stewart, M. Veronesi, B. Stockman, *J. Biomol. NMR* **2001**, *21*, 349-359.

- [18] C. Dalvit, P. Pevarello, M. Tato, M. Veronesi, A. Vulpetti, M. Sundstrom, *J. Biomol. NMR* **2000**, *18*, 65-68.
- [19] M. Mayer, B. Meyer, *Angew. Chem., Int. Ed.* **1999**, *38*, 1784-1788.
- [20] A. D. Gossert, C. Henry, M. J. J. Blommers, W. Jahnke, C. Fernandez, *J. Biomol. NMR* **2009**, *43*, 211-217.
- [21] M. Conner, M. R. Hicks, T. Dafforn, T. J. Knowles, C. Ludwig, S. Staddon, M. Overduin, U. L. Guenther, J. Thome, M. Wheatley, D. R. Poyner, A. C. Conner, *Biochemistry* **2008**, *47*, 8434-8444.
- [22] K. Furihata, S. Shimotakahara, Y. Shibusawa, M. Tashiro, *Anal. Sci.* **2010**, *26*, 1107-1110.
- [23] K. Furihata, S. Shimotakahara, M. Tashiro, *Magn. Reson. Chem.* **2008**, *46*, 799-802.
- [24] I. M. Figueiredo, A. J. Marsaioli, *Quim. Nova* **2007**, *30*, 1597-1605.
- [25] C. Ludwig, U. L. Guenther, *Front Biosci.* **2009**, *14*, 4565-4574.
- [26] C. Ludwig, P. J. A. Michiels, X. Wu, K. L. Kavanagh, E. Pilka, A. Jansson, U. Oppermann, U. L. Gunther, *J. Med. Chem.* **2008**, *51*, 1-3.
- [27] M. Assfalg, I. Bertini, R. Del Conte, A. Giachetti, P. Turano, *Biochemistry* **2007**, *46*, 6232-6238.

- [28] A. Ciulli, G. Williams, A. G. Smith, T. L. Blundell, C. Abell, *J. Med. Chem.* **2006**, *49*, 4992-5000.
- [29] S. Shimotakahara, K. Furihata, M. Tashiro, *Magn. Reson. Chem.* **2005**, *43*, 69-72.
- [30] C. Ludwig, P. J. A. Michiels, A. Lodi, J. Ride, C. Bunce, U. L. Gunther, *ChemMedChem* **2008**, *3*, 1371-1376.
- [31] M. Levitt, B. H. Park, *Structure* **1993**, *1*, 223-226.
- [32] R. B. Hossany, M. A. Johnson, A. A. Eniade, B. M. Pinto, *Biorg. Med. Chem.* **2004**, *12*, 3743-3754.
- [33] H. C. Andersen, *J. Comput. Phys.* **1983**, *52*, 24-34.
- [34] S. Miyamoto, P. A. Kollman, *J. Comput. Chem.* **1992**, *13*, 952-962.
- [35] H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684-3690.
- [36] C. A. Lepre, J. M. Moore, J. W. Peng, *Chem. Rev.* **2004**, *104*, 3641-3675.

4.10 Supporting Information

***WaterLOGSY NMR Experiments Detect Immobilized Water Molecules
that Bridge Peptide Mimic MDWNMHAA to Anti-carbohydrate Antibody
SYA/J6***

Monica G. Szczepina,^a Dustin W. Bleile,^b Johannes Müllegger,^{†b}

Andrew R. Lewis,^a and B. Mario Pinto^a

^aDepartment of Chemistry, Simon Fraser University, Burnaby,
British Columbia, Canada, V5A 1S6

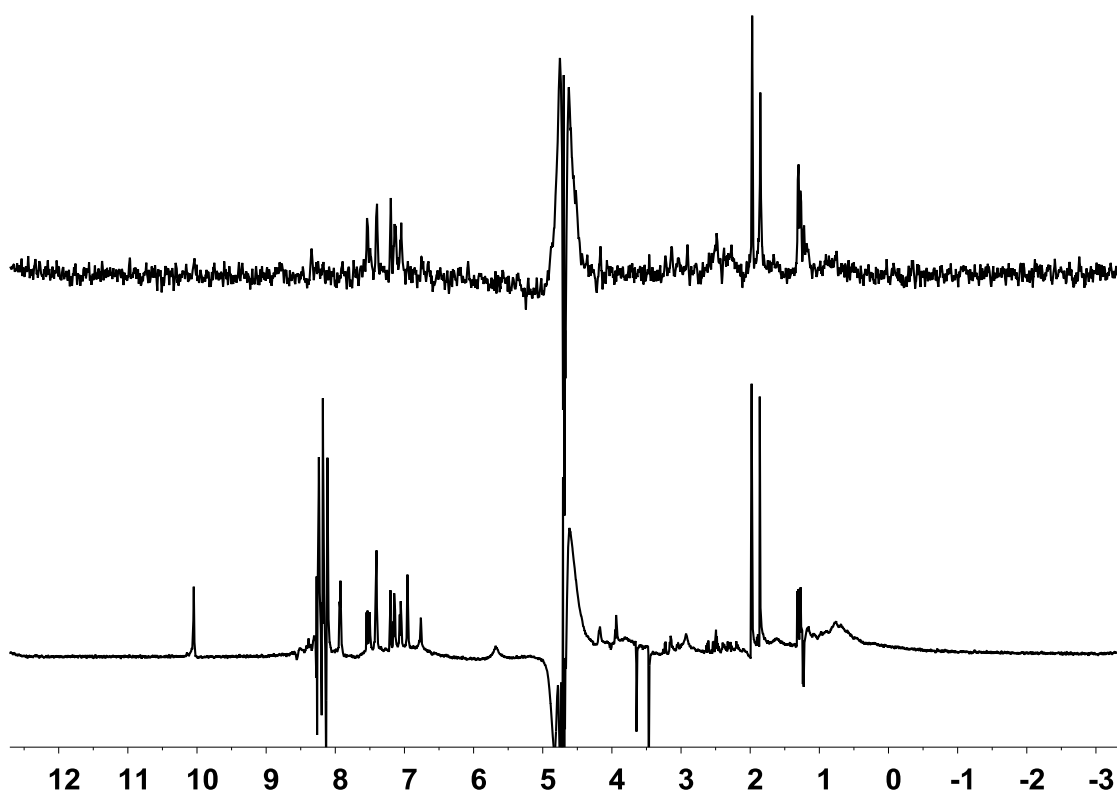
^bZymeworks Inc., 540-1385 West 8th Ave, Vancouver, British Columbia,
Canada, V6H 3V9

[†] Deceased April 23, 2009. This article is dedicated, with respect, to the memory
of Johannes Müllegger.

Table of Contents

Figure S4.1.	1D WaterLOGSY NMR spectrum of peptide 1 in the presence of mAb SYA/J6 at 600 MHz and 298 K corrected for ligand only effects and 1D STD-NMR spectrum of the same sample under identical conditions.	159
---------------------	--	-----

Figure S4.1. 1D WaterLOGSY NMR spectrum of peptide 1 in the presence of mAb SYA/J6 at 600 MHz and 298 K corrected for ligand only effects (upper trace) and 1D STD-NMR spectrum of the same sample under identical conditions (lower trace).



CHAPTER 5: Investigation of the Binding of a Carbohydrate-mimetic Peptide to its Complementary Anti-carbohydrate Antibody by STD-NMR Intensity-restrained CORCEMA Optimization (SICO) and Molecular Dynamics Simulations

This chapter comprises the manuscript “*Investigation of the Binding of a Carbohydrate-mimetic Peptide to its Complementary Anti-carbohydrate Antibody by STD-NMR Intensity-restrained CORCEMA Optimization (SICO) and Molecular Dynamics Simulations*” which was accepted in *Chem. Eur. J.*

Monica G. Szczepina,^a Dustin W. Bleile,^b and B. Mario Pinto^a

^aDepartment of Chemistry, Simon Fraser University, Burnaby, British Columbia, Canada, V5A 1S6

^bZymeworks Inc., 540-1385 West 8th Ave, Vancouver, British Columbia, Canada, V6H 3V9

In Chapter 5, a more detailed study of the carbohydrate mimetic peptide, MDWNMHAA, when bound to the antibody, SYA-J6 is described. Therein, full STD-NMR build-up curves to capture the epitope of the peptide and CORCEMA-ST calculations to quantify these STD NMR data are also described. In particular, we attempt to answer whether the solution NMR data are in agreement with the solid state crystallographic image of the bound structure of MDWNMHAA. We found that the agreement is poor and this lead us to perform molecular dynamics simulations on the peptide while bound to the antibody. As a way to sample different backbone conformations of the peptide, we also turned to simulated annealing. Last, we also describe our attempt to fully characterize the synthetic pentasaccharide antigen while bound to SYA-J6.

Mr Dustin W. Bleile wrote python scripts to analyze the MD trajectory. The thesis author performed the NMR experiments, the MD simulations, the CORCEMA-ST calculations, simulated annealing refinement and data analysis.

5.1 Keywords

Saturation Transfer Difference (STD)-NMR, CORCEMA-ST (Complete Relaxation and Conformational Exchange Matrix Analysis of Saturation Transfer), SICO (STD-NMR Intensity-restrained CORCEMA Optimization), Simulated Annealing (SA), Peptide-Carbohydrate Mimicry, Carbohydrate-mimetic Peptide, *Shigella flexneri* Y

5.2 Abstract

Saturation transfer difference (STD)-NMR spectroscopy was used to probe experimentally the bioactive solution conformation of the carbohydrate mimic MDWNMHAA 1 of the O-polysaccharide of *Shigella flexneri* Y when bound to its complementary antibody, mAb SYA/J6. Molecular dynamics simulations using the ZymeCAD™ Molecular Dynamics platform were also undertaken to give a more accurate picture of the conformational flexibility and the possibilities for bound ligand conformations. The ligand topology, or the dynamic epitope, was mapped with the CORCEMA-ST (COmplete Relaxation and Conformational Exchange Matrix Analysis of Saturation Transfer) program that calculates a total matrix analysis of relaxation and exchange effects to generate predicted STD-NMR intensities from simulation. The comparison of these predicted STD enhancements with experimental data was used to select a representative binding mode. A protocol that employed theoretical STD effects calculated at snapshots during the entire course of a molecular dynamics (MD) trajectory of the peptide bound to the Fv portion of the antibody, and not the averaged atomic positions of receptor-ligand complexes, was also examined. In addition, the R-factor was calculated on the basis of STD (fit) to avoid T1 bias, and an effective R factor was defined such that if the calculated STD (fit) for proton k was within error of the experimental STD (fit) for proton k, then that calculated STD (fit) for proton k was not included in the calculation of the R-factor. This protocol was effective in deriving the antibody-bound solution conformation of the peptide which also differed from the bound conformation determined by X-ray

crystallography; however, several discrepancies between experimental and calculated STD (fit) values were observed. The bound conformation was therefore further refined with a simulated annealing refinement protocol known as STD-NMR Intensity-restrained CORCEMA Optimization (SICO) to give a more accurate representation of the bound peptide epitope. Further optimization was required in this case, but a satisfactory correlation between experimental and calculated STD values was obtained. Attempts were also made to obtain STD enhancements with a synthetic pentasaccharide hapten, corresponding to the O-polysaccharide, while bound to the antibody. However, unfavorable kinetics of binding in this system prevented sufficient STD build-up, which, in turn, hindered a rigorous analysis via full STD build-up curves.

5.3 Introduction

According to the World Health Organization (WHO), one million people are estimated to die from *Shigellosis* ^[1] each year, with additional 165 million cases of severe dysentery, where children under five years of age are most severely affected. The *Shigella flexneri* strains, which are endemic in most developing countries, are a major contributor to the *Shigella* infection rate. ^[2]

The *Shigella flexneri* Y strain has a cell-surface O-linked polysaccharide (LPS) made up of a linear chain $[\rightarrow 2)\text{-}\alpha\text{-L-Rha-(1}\rightarrow 2)\text{-}\alpha\text{-L-Rha-(1}\rightarrow 3)\text{-}\alpha\text{-L-Rha-(1}\rightarrow 3)\text{-}\beta\text{-D-GlcNAc-(1}\rightarrow 2)\text{-}\alpha\text{-L-Rha-(1}\rightarrow]$ (Figure 5.1). ^[3-7] When used to immunize mice, an IgG3 monoclonal antibody (SYA/J6) was identified which is specific for the polysaccharide, therefore opening the possibility of developing a model vaccine against *Shigella flexneri* Y. ^[3-7] Traditionally, carbohydrate-based

vaccines are weakly immunogenic, as determined by the thymus dependent immune response; therefore, studies were undertaken to identify carbohydrate-mimetic peptides as surrogate ligands. [8-10] Using phage display, a weakly immunogenic [11] carbohydrate-mimetic peptide of the *Shigella flexneri* Y O-linked polysaccharide, with the amino acid sequence MDWNMHAA 1 (Figure 5.2), was identified to be cross-reactive with the SYA/J6 monoclonal antibody with 19 μM binding. [12]

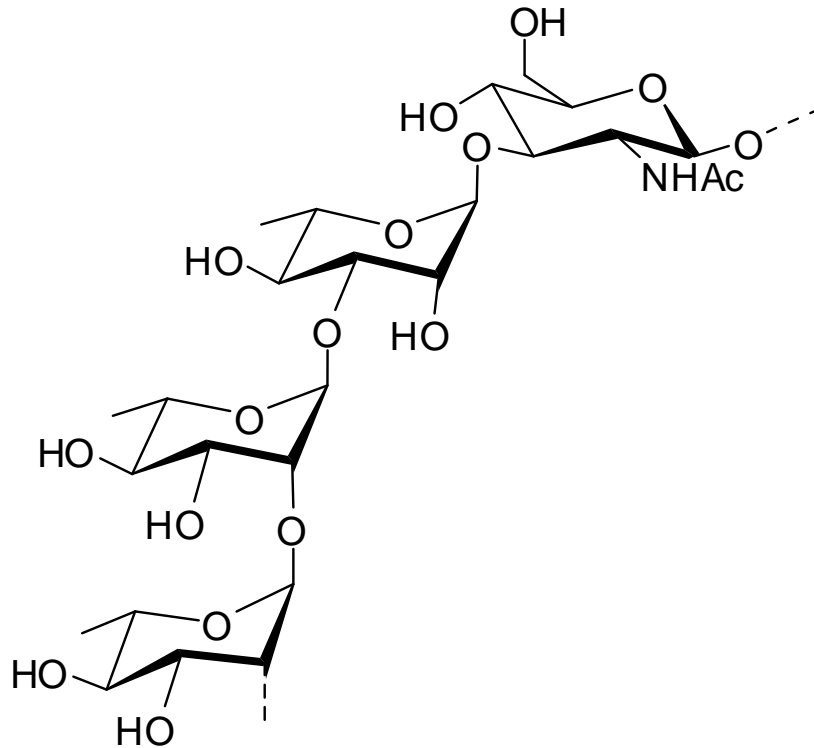


Figure 5.1. Structure of the *Shigella flexneri* Y O-antigen polysaccharide.

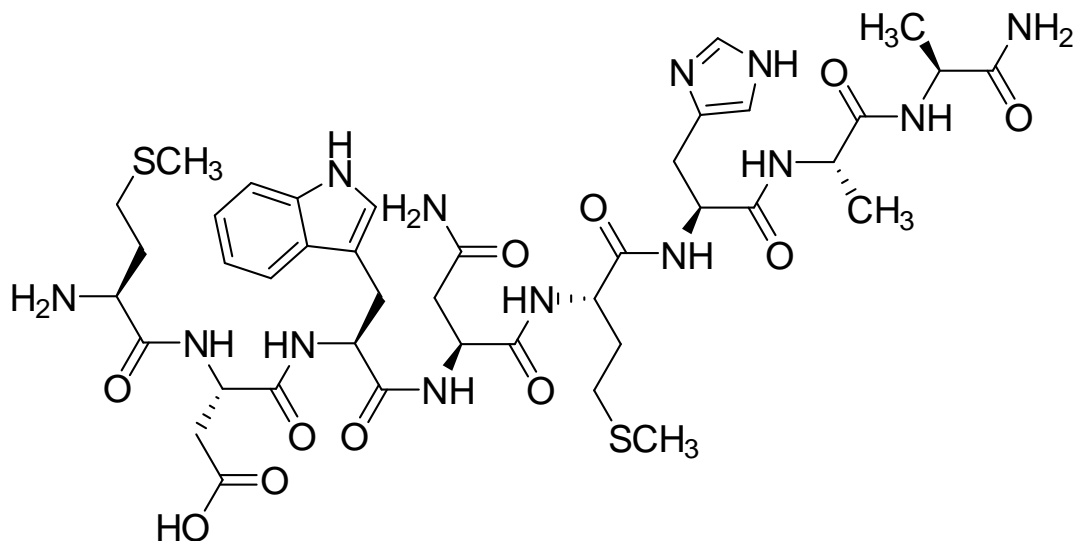


Figure 5.2. Structure of the carbohydrate-mimetic peptide MDWNMHAA **1** (peptide **1**).

The modes of binding between the octapeptide **1** and a pentasaccharide hapten, α -L-Rha-(1 \rightarrow 2)- α -L-Rha-(1 \rightarrow 3)- α -L-Rha-(1 \rightarrow 3)- β -D-GlcNAc-(1 \rightarrow 2)- α -L-Rha-(1 \rightarrow OME **2**, corresponding to the O-polysaccharide, with the SYA/J6 antibody differ, as revealed by the X-ray crystallographic structures of the two ligands complexed with SYA/J6 Fab fragment (Figure 5.3).^[13, 14] Briefly, whereas the peptide contacts some of the same areas of the binding site as the oligosaccharide, it also extends into other areas of the site.^[15] The most pronounced difference in pentasaccharide vs, peptide binding is that the peptide does not directly contact the central, deepest part of the groove, where Rha C binds. Rather, three water molecules occupy this cavity (Figure 5.3a). Overall, the peptide is more highly complementary to the combining site, making 126

intermolecular contacts, compared to the pentasaccharide's 74. It is noteworthy that despite the differences in peptide versus oligosaccharide binding, a protein conjugate of the peptide ligand is immunogenic and elicits an immune response that is cross-reactive with the bacterial polysaccharide.^[11] Thus, the peptide is an antigenic mimic of the polysaccharide.

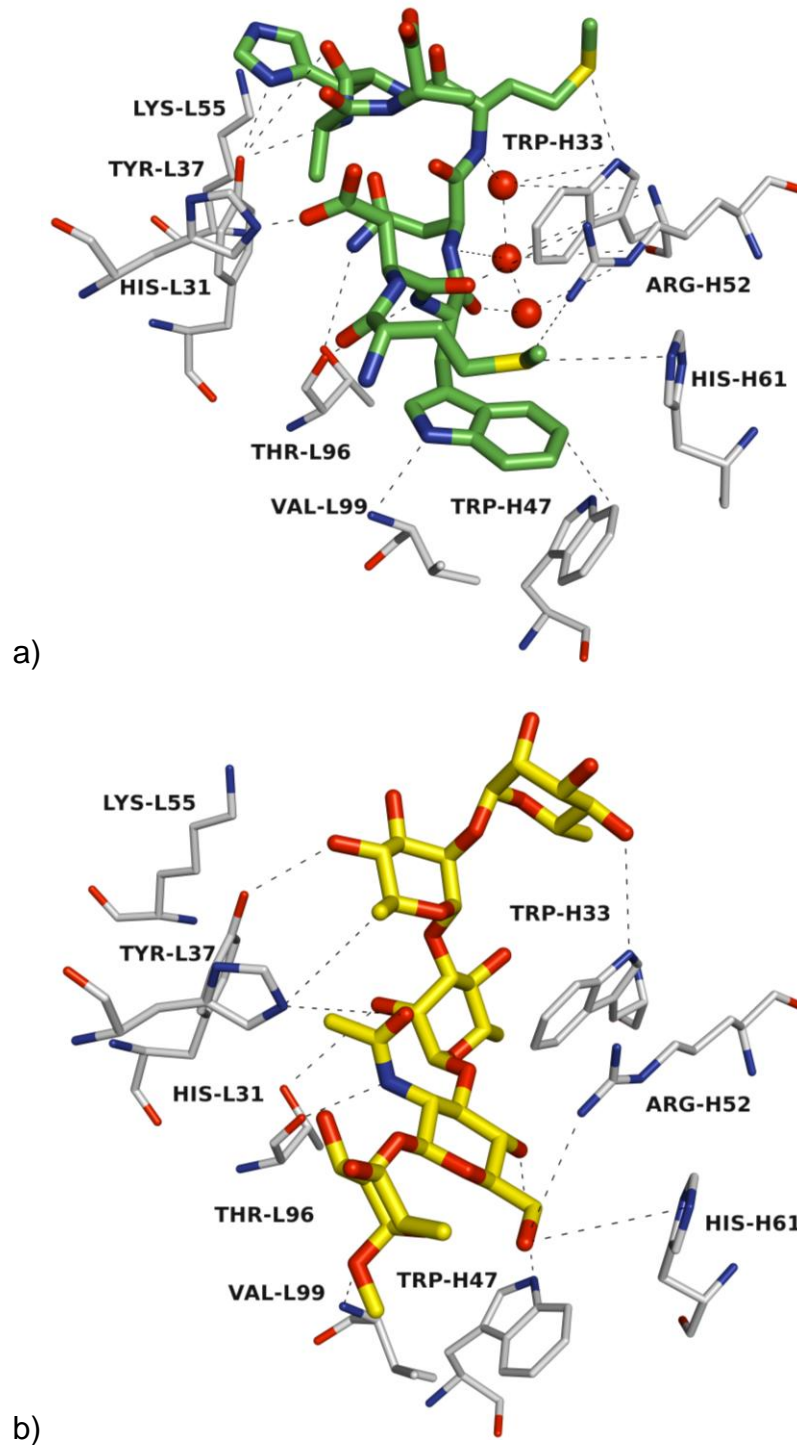


Figure 5.3. X-ray crystal structure of a Fab fragment of the murine monoclonal antibody SYA/J6 (IgG3, κ), directed against the polysaccharide O-antigen of the *Shigella flexneri* variant Y lipopolysaccharide, complexed with a) peptide 1

(green), b) the pentasaccharide hapten [α -L-Rhap-(1 \rightarrow 2)- α -L-Rhap-(1 \rightarrow 3)- α -L-Rhap-(1 \rightarrow 3)- β -D-GlcNAc-(1 \rightarrow 2)- α -L-Rhap-(1 \rightarrow OMe)] **2** (yellow).

Isothermal titration calorimetry found the binding of the pentasaccharide **2** to be $\Delta H = -1.5$ kcal/mol and $-T\Delta S = -5.9$ kcal/mol. On the other hand, the mimetic peptide **1** binding was strongly enthalpy driven $\Delta H = -16.9$ kcal/mol, but this was offset by an entropy term of $-T\Delta S = +9.1$ kcal/mol. Consequently, the octapeptide **1** ($K_A = 5.7 \times 10^5 \text{ M}^{-1}$) and the pentasaccharide **2** ($K_A = 2.5 \times 10^5 \text{ M}^{-1}$) bind with approximately equal affinity. The quest for higher affinity ligands relies on a detailed understanding of the nature of mimicry. The octapeptide is a functional mimic of the natural pentasaccharide as it binds to the antibody SYA/J6 and yet this results in a severe entropic penalty. This penalty is caused by (1) the reduction of conformational dynamics upon binding, (2) the introduction of an α -helical turn, and (3) the immobilization of water molecules, i.e. the presence of “bound” water molecules. In order to probe the reduction of conformational dynamics upon binding, we report herein a detailed study of the bound solution conformation of the octapeptide by STD-NMR spectroscopy. We have described a preliminary study in which the observed STD-NMR enhancements were mapped onto the crystal structure of the peptide-antibody complex in order to define a molecular surface,^[16] but now present a more rigorous study using two approaches. First, molecular dynamics of the peptide-antibody complex, in conjunction with calculation of relaxation and exchange effects using CORCEMA-ST were used to derive a more accurate model.

Second, STD-NMR Intensity-restrained CORCEMA-ST Optimization (SICO), a novel hybrid structure refinement protocol that attempts to identify the global minimum of a bioactive, bound conformation via simulated annealing (SA), was employed. [17-23]

5.4 Experimental

5.4.1 Materials

MDWNMHAA **1** was synthesized according to the method of Hossany *et al.* [8] The SYA/J6 antibody was a gift from D.R. Bundle.

5.4.2 NMR Spectroscopy of peptide **1**

¹H NMR chemical shifts of peptide **1** in D₂O were acquired on the Varian Inova 500 MHz spectrometer. ¹³C NMR spectra were acquired on the Bruker 400 MHz spectrometer. An NMR sample of the mAb SYA/J6 (30 μM, 60 μM binding sites) was prepared in phosphate buffered saline (12mM Na₂DPO₄/NaD₂PO₄, 137mM NaCl, 3mM KCl, 0.02% NaN₃, 100% D₂O, pD 6.1) and peptide **1** (2.5mg) was added to the sample for a final peptide concentration of 6.4mM and a ratio of 100:1 peptide:antibody. The STD-NMR spectra were recorded at 282 K (better signal/noise at this temperature than at 298 K) with 4096 scans and selective saturation of protein resonances at -2 ppm (30 ppm for off resonance spectra) using a series of 40 Gaussian-shaped pulses (50 ms, 1 ms delay between pulses, $\gamma B_1/2\pi = 110$ Hz), for total saturation times ranging from 0.6 s to 5.0 s. The protein resonances were broad and had significant intensity in the region

downfield from 10 ppm and even at negative parts per million values. Thus, irradiation at -2 ppm was expected to result in saturation of protein resonances, from the aromatic to the aliphatic. A 10 ms spin-lock pulse ($\gamma B_1/2\pi = 11$ kHz) was applied after excitation to reduce the intensity of broad protein resonances. STD-NMR spectra were acquired on the Bruker Avance 600 NMR spectrometer equipped with a TCI cryoprobe. The saturated and reference spectra were acquired in the same dataset by creating a pseudo-2D experiment. The STD spectrum was obtained by subtraction of saturated spectra from reference spectra after identical processing and phasing. In all cases, the fractional STD effect was calculated by $(I_0 - I_{\text{sat}})/I_0$, where $(I_0 - I_{\text{sat}})$ is the peak intensity in the STD spectrum and I_0 is the peak intensity of an unsaturated reference spectrum. In all cases, STD spectra were acquired without water suppression because of the proximity of the H- α resonances.

5.4.3 NMR Spectroscopy of pentasaccharide 2

^1H NMR spectra of pentasaccharide **2** in D_2O were acquired on the Bruker Avance 600 NMR spectrometer equipped with a TCI cryoprobe. An NMR sample of the mAb SYA/J6 (15 μM , 30 μM binding sites) was prepared in phosphate buffered saline (12mM $\text{Na}_2\text{DPO}_4/\text{NaD}_2\text{PO}_4$, 137mM NaCl, 3mM KCl, 0.02% NaN_3 , 100% D_2O , pD 6.1) and pentasaccharide (1.0mg) was added to the sample, for a final pentasaccharide concentration of 3.0mM and a ratio of 100:1 pentasaccharide:antibody. The STD-NMR spectra were recorded at 298 K (also at 282 K and 310 K to determine which temperature gave highest signal to noise) with 4096 scans and selective saturation of protein resonances at 10 ppm (30

ppm for off resonance spectra) using a series of 40 Gaussian-shaped pulses (50 ms, 1 ms delay between pulses, $\gamma B_1/2\pi = 110$ Hz), for a total saturation time of 2.04 s. The protein resonances were broad and had significant intensity in the region downfield from 10 ppm and even at negative parts per million values. Thus, irradiation at 10 ppm was expected to result in saturation of protein resonances, from the aromatic to the aliphatic. A 10 ms spin-lock pulse ($\gamma B_1/2\pi = 11$ kHz) was applied after excitation to reduce the intensity of broad protein resonances. STD-NMR spectra were acquired on the Bruker Avance 600 NMR spectrometer equipped with a TCI cryoprobe. The saturated and reference spectra were acquired in the same dataset by creating a pseudo-2D experiment. The STD spectrum was obtained by subtraction of saturated spectra from reference spectra after identical processing and phasing. In all cases, the fractional STD effect was calculated by $(I_0 - I_{\text{sat}})/I_0$, where $(I_0 - I_{\text{sat}})$ is the peak intensity in the STD spectrum and I_0 is the peak intensity of an unsaturated reference spectrum. The methyl signal of the major NHAc isomer is at 2.06 ppm and that of the minor NHAc isomer is at 1.91 ppm. The minor isomer is most enhanced (100%); all enhancements were measured relative to this signal. Due to the extremely weak STD-NMR signal, STD spectra were then acquired with water suppression in addition to a spin-lock. Those spectra were acquired for a running time of 72h corresponding to 20480 scans. However, even under those conditions, the signal to noise ranged from 30 to 80 and this made the acquisition of build-up curves impossible since the error in the signal reading was comparable to the signal itself.

5.4.4 CORCEMA-ST calculations

The theory of CORCEMA-ST and the details of executing the CORCEMA-ST protocol have been described previously.^[18, 21, 23-27]

Briefly, for the ideal case where there is infinite delay between each scan, the magnetizations in the STD experiment are given by

$$\mathbf{I}(t) = \mathbf{I}_0 + [1 - e^{-Dt}] \mathbf{D}^{-1} \mathbf{Q}$$

Where $\mathbf{I}(t)$ is a column matrix containing the magnetizations for the ligand and for those protein protons that do not experience a direct radio frequency (rf) saturation.^[19, 20, 24]

\mathbf{Q} is a column matrix containing cross-relaxation terms between the protein protons that experience a direct rf saturation and the rest of the protons.^[19, 20, 24] The number of protons in peptide **1** and the number of protons of the amino acid residues within the mAb SYA/J6 Fab combining site, as well as the number of protein protons that experience direct rf irradiation and their identities were read into the program on the basis of the PDB coordinates of the peptide **1**-SYA/J6 Fab complex (PDB entry 1PZ5). To speed up the computation of the matrix, spectral densities were calculated for only those proton pairs having a distance of 5 Å or less.^[19, 20, 24] In the calculations, peptide **1** and the 27 amino acid residues (from chain A: His27D, Asp28, Asn30, Tyr32, His34 Lys50, Thr91, Thr92, His93, Val94, Pro95 and from Chain B: Asn31, Tyr32, Trp33, Glu35, Trp47, Glu50, Arg52, Leu52A, His58, Tyr59, Gly95, Gly96, Ala97, Val98, Gly99 and Ala100) were included. These residues were selected because they were

within 5 Å of the ligand. In calculations with MD data, proton pairs within 12Å were considered.

The dynamic matrix **D** is a square matrix and is a sum of the relaxation rate matrix **R** and the exchange matrix **K**. “*t*” is the time period for which the protein proton(s) experience rf irradiation.^[19, 20, 24] The CORCEMA-ST program also has a provision for taking into account the effect of finite delays (*t_d*) between scans in calculating the STD effects, and this finite delay was taken into account.^[19, 20, 24] The order parameter was set to 0.25 for the methyl group while for methyl-X relaxation, *S*² was generally kept in the range of 0.85.^[19, 20, 24] To account for the effect of internal motions of the methyl groups, the corresponding spectral densities were calculated using the model free formalism.^[19, 20, 24] For Tyr and Phe, a simple $\langle 1/r^6 \rangle$ average was used for the dipolar relaxation between aromatic and other protons.^[19, 20, 24] The protein concentration was kept fixed at 60µM and the ratio of ligand: protein was kept at 100:1, as in our NMR experiments. The correlation times (*τ_c*) and equilibrium constants (*K_{eq}*) were tested to find the best fit between experimental data and predicted STD values. Although the correlation time was not determined experimentally, it was kept in the range of molecular weight × 10⁻¹², which is a rough estimate of correlation time.^[28] Since peptide **1** binds to the anti-*Shigella* Y monoclonal antibody SYA/J6 Fab with an affinity of *K_A* = 5.7 × 10⁵ M⁻¹, *K_{eq}* values were set in the range of 10⁶ M⁻¹ because of the lower temperature (282 K) used in the STD-NMR experiments compared to the microcalorimetry experiments; *k_{on}* was set to 10⁷ s⁻¹ M⁻¹ to reflect the fact that conformational reordering would be required upon binding

and that k_{on} would be less than diffusion controlled ($\sim 10^8$ - 10^9). In fact, a best fit with experiment was observed with an on- rate of $10^7 \text{ s}^{-1} \text{ M}^{-1}$.

From the intensity matrix $I(t)$, the fractional intensity changes $[(I_{0(k)} - I(t)_{(k)}) \times 100 / I_{0(k)}]$ for different ligand protons were calculated, and compared to the experimental STD values using an NOE R-factor defined as: [19, 20, 22, 24]

$$\text{R - factor} = \sqrt{\frac{\sum (S_{\text{expt},k} - S_{\text{calc},k})^2}{\sum (S_{\text{expt},k})^2}}$$

In these equations, $S_{\text{expt},k}$ and $S_{\text{calc},k}$ refer to experimental and calculated STD values for proton k . [19, 20, 24] Since the protein signals at -2 ppm were irradiated for the STD experiment, we made the reasonable assumption that the methyl protons in Ile, Ala, Leu and Val were instantaneously saturated and that magnetization would take a finite time to spread to other protein and ligand protons (bound and free) through dipolar networks and chemical exchange. [19, 20, 24]

5.4.5 Molecular Dynamics Simulations

Crystal-structure coordinates of the protein receptor and ligand were used as a starting structure for the complex (PDB entry 1PZ5). Molecular dynamics simulations were performed on a custom MD engine (ZymeCADTM) using the AMBER force field version 99sb^[29] in explicit solvent composed of TP3 water molecules. A 1.5 fs time step was used for integration, with hydrogen bond

lengths constrained by the RATTLE algorithm ^[30]. The SETTLE ^[31] algorithm was used to constrain water molecules to their equilibrium geometry. Non-bonded interactions were treated with a shifting function to scale interaction energies to zero at 12 Angstroms. Simulations were run under NPT conditions of 300 Kelvin and 1 atm pressure. The temperature of the simulation was controlled by a Berendsen thermostat ^[32] and the simulation pressure was controlled by a Berendsen barostat ^[32].

5.5 Results and Discussion

5.5.1 Probing the secondary structure of peptide 1 free in solution

NMR parameters such as $^1\text{H}_\alpha$, $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ chemical shifts are used to identify protein secondary structural elements such as: a) random coil, b) α -helix and c) β -sheet. Wishart *et al.* have developed the chemical shift index method which relies on the observation that $^1\text{H}_\alpha$ protons, relative to a random coil value, experience an upfield shift when in an α -helix and a downfield shift when in a β -sheet.^{[33],[34]} The authors extended this concept to include changes in carbon chemical shifts when secondary structural elements are present. This method has been computerized and uses as input the chemical shifts of amino acids that make up a peptide. These data are then processed and the NMR parameters of each amino acid are compared to random coil values. Deviations from random coil are noted, and as output, the program recognizes and identifies α -helical or β -sheet regions. This method can also be applied to peptides, as illustrated by Schwarzingler *et al.* for the case of a pentapeptide.^[35]

Chemical shifts for peptide **1** are summarized in Tables S5.1 and S5.2. These values were used as input for the Chemical Shift Index (CSI) program and the output indicated that peptide **1** exhibits a random coil arrangement when free in solution. Specifically, the NMR chemical shift values for the peptide are in close agreement with what is currently accepted as representative of a random coil. The output file is shown in Table S5.3. The lack of inter residue NOEs or ROEs also indicates that peptide **1** is a random coil free in solution.

The results are corroborated by molecular dynamics studies of peptide **1** free in solution. Thus, peptide **1** was built with a variety of secondary structural elements: 3:10 helix, antistrand, helix, left-handed helix, pi helix, polyproline and strand. These conformations, along with peptide **1** from the X-ray crystal structure and from a simulated annealing procedure were all used as starting points for MD simulations acquired for 15ns. These trajectories were analyzed by Ramachandran plots and 2D RMSD plots. Ramachandran plots, shown in Figure S5.2, indicate that the free peptide **1** taken from the crystal structure deviates from its starting conformation and does not retain α -helical character in the absence of the antibody. 2D RMSD plots (RMSD of one conformation to every other conformation at every time point throughout the simulation) indicate that the free ligand does not fall into a preferred conformation (Figure S5.3). These results are consistent with the NMR random coil solution structure of free peptide **1**. Thus, when peptide **1** binds to SYA/J6, the reduction of its conformational dynamics should result in an unfavorable entropy of binding.

5.5.2 Binding of peptide 1 to monoclonal antibody SYA/J6 by Saturation Transfer Difference NMR Spectroscopy

Previous attempts to analyze the bound conformation of the peptide by trNOESY experiments were unsuccessful, and transferred NOE effects for the peptide-antibody complex were not observed. This is likely due to unfavorable binding kinetics; if the off-rate for the complex is not significantly greater than the NOE buildup rate in the complex, efficient transfer of the NOEs representative of the bound conformation to the free state of the ligand will not occur. [36-40] However, saturation times employed in saturation transfer difference NMR (STD-NMR) experiments are longer than the mixing times employed in transferred NOE experiments (typical values, and those employed in our previous study^[16], are 2 s and ≤ 300 ms, respectively), and this extra time for NOE buildup may allow the observation of STD-NMR effects even when trNOEs are not observable.

The binding of peptide 1 to mAb SYA/J6 was therefore investigated using STD-NMR spectroscopy. The STD-NMR spectra at a saturation time of 2s are shown in Figures S5.4 and S5.5. STD build-up curves were determined at seven different saturation times ranging from 0.6s to 5s (Figure S5.6). The maximal STD intensity, STD_{max} , and the observed saturation rate constant k_{sat} were obtained from fitting the saturation time data to the monoexponential equation: $STD = STD_{max} (1 - e^{(-k_{sat}t)}) + c$, slightly modified from Mayer and James. [41] The slope of the STD build-up curve at a saturation time of 0, $STD(fit)$ (shown in Table 5.1), was obtained by the multiplication of STD_{max} and k_{sat} . $STD(fit)$ is believed to correspond to the STD intensity in the absence of T1 bias and

depends on the proximity of the ligand proton to the protein.^[41] The STD(fit) values were normalized for ease of comparison to the largest STD (fit) value, namely that for the resonance corresponding to Trp-3H ζ 2, to give STD epitopes (fit).

STD enhancements were greatest for the side chain of residue Trp-3. The X-ray crystal structure shows that the indole ring of Trp-3 is deeply buried in a hydrophobic cavity formed by Val L94, Pro L95, His H58 and Trp H47 and this is reflected in the bound solution structure. STD enhancements were observed for the side chain of residue Met-1 but this is in contradiction to the X-ray crystal structure which indicates that the Met-1 residue is largely exposed to solvent. Likewise, STD enhancements were observed for Asp-2, but the bound X-ray crystal structure shows very few van der Waals contacts with the Fab fragment.^[13, 15] Thus, some structural discrepancies in the bound, bioactive conformation of peptide **1** exist between the STD-NMR spectroscopic data and the crystallographic data.

Table 5.1. Experimental STD intensities of peptide **1** at different saturation times. STD_{max} and k_{sat} were calculated by fitting the data to a rising exponential equation with an offset and STD(fit) and epitopes (fit) were derived.

(A)	Experimental STD ^a			STD _{max} ^b	k_{sat} ^b	STD(fit) ^c	STD epitopes (fit) ^a
	0.6 s	2 s	5 s				
Met-1H α	36.6	15.1	14.6	$(0.78 \pm 0.44) \times 10^{-2}$	0.179 \pm 0.176	$(0.14 \pm 0.16) \times 10^{-2}$	8.5
Met-1H γ	24.4	15.1	12.0	$(0.51 \pm 0.07) \times 10^{-2}$	0.374 \pm 0.180	$(0.19 \pm 0.09) \times 10^{-2}$	11.5
Met-1H ϵ	19.5	23.4	21.0	$(0.99 \pm 0.03) \times 10^{-2}$	0.458 \pm 0.059	$(0.45 \pm 0.06) \times 10^{-2}$	27.5

Asp-2H α	--- ^d	--- ^d	--- ^d	--- ^d	--- ^d	--- ^d	--- ^d
Asp-2H β	46.3	19.3	15.1	$(0.54 \pm 0.02) \times 10^{-2}$	0.438 ± 0.056	$(0.24 \pm 0.03) \times 10^{-2}$	14.5
Trp-3H α	--- ^d	--- ^d	--- ^d	--- ^d	--- ^d	--- ^d	--- ^d
Trp-3H β	46.3	34.9	30.8	$(1.42 \pm 0.06) \times 10^{-2}$	0.432 ± 0.079	$(0.61 \pm 0.12) \times 10^{-2}$	37.3
Trp-3H δ 1	48.8	57.3	53.5	$(2.59 \pm 0.06) \times 10^{-2}$	0.391 ± 0.035	$(1.01 \pm 0.09) \times 10^{-2}$	61.5
Trp-3H ϵ 3	53.7	61.5	62.7	$(3.14 \pm 0.13) \times 10^{-2}$	0.336 ± 0.043	$(1.05 \pm 0.14) \times 10^{-2}$	64.1
Trp-3H ζ 2	100	100	100	$(4.95 \pm 0.14) \times 10^{-2}$	0.333 ± 0.029	$(1.65 \pm 0.15) \times 10^{-2}$	100.0
Trp-3H ζ 3	95.1	89.1	89.4	$(4.45 \pm 0.23) \times 10^{-2}$	0.299 ± 0.043	$(1.33 \pm 0.20) \times 10^{-2}$	80.8
Trp-3H η 2	90.2	85.9	84.9	$(4.15 \pm 0.21) \times 10^{-2}$	0.313 ± 0.047	$(1.30 \pm 0.20) \times 10^{-2}$	79.0
Asn-4H α	--- ^d	--- ^d	--- ^d	--- ^d	--- ^d	--- ^d	--- ^d
Met-5H γ	36.6	18.2	14.8	$(0.56 \pm 0.05) \times 10^{-2}$	0.394 ± 0.141	$(0.22 \pm 0.08) \times 10^{-2}$	13.4
Met-5H ϵ	19.5	32.8	31.7	$(1.56 \pm 0.06) \times 10^{-2}$	0.408 ± 0.065	$(0.64 \pm 0.10) \times 10^{-2}$	38.6
His-6H α	--- ^d	--- ^d	--- ^d	--- ^d	--- ^d	--- ^d	--- ^d
His-6H β	39.0	20.3	19.0	$(1.10 \pm 0.37) \times 10^{-2}$	0.161 ± 0.091	$(0.18 \pm 0.12) \times 10^{-2}$	10.8
His-6H ϵ 1	17.1	29.7	27.7	$(1.35 \pm 0.06) \times 10^{-2}$	0.407 ± 0.076	$(0.55 \pm 0.11) \times 10^{-2}$	33.4
His-6H δ 2	14.6	27.1	28.3	$(1.52 \pm 0.05) \times 10^{-2}$	0.357 ± 0.040	$(0.54 \pm 0.06) \times 10^{-2}$	33.0
Ala-7H β	36.6	13.5	10.4	$(0.31 \pm 0.03) \times 10^{-2}$	0.472 ± 0.198	$(0.15 \pm 0.06) \times 10^{-2}$	9.0
Ala-8H β	29.3	13.5	9.5	$(0.34 \pm 0.01) \times 10^{-2}$	0.627 ± 0.034	$(0.22 \pm 0.01) \times 10^{-2}$	13.1

- For ease of comparison, the STD effects were normalized to the Trp-3H ζ 2 intensity, which was set to 100.
- From fitting the saturation time data to the monoexponential equation:

$$\text{STD} = \text{STD}_{\text{max}} (1 - e^{(-k_{\text{sat}}t)}) + c.$$

- c. Slope of the STD build-up curve at a saturation time of 0.
- d. Not determined because of peak overlap.

5.5.3 Quantifying the binding of peptide 1 to monoclonal antibody SYA/J6 with CORCEMA-ST Calculations

Theoretical STD effects calculated with CORCEMA-ST using the crystal structure coordinates are shown in Figure 5.4. An overall R-factor of 0.47 was obtained using STD data at 5s and antibody residues within 5Å of the peptide. Use of STD data at 5s and antibody residues within 12Å of the peptide gave an R-factor of 0.54. A lower R-factor indicates a good match between predicted value and experimental data. The predicted STD values for the side chains of residues Met-1, Asp-2, Trp-3 and Asn-4 were smaller than the observed values, indicating again that the crystallographic data were not entirely representative of the bound solution structure of peptide 1. In contrast, the predicted STD values for the side chains of residues Met-5, His-6, Ala-7 and Ala-8 agreed with the observed STD enhancements. More significantly, the α -helical turn observed in the X-ray crystal structure was retained in the bound solution structure of peptide 1.

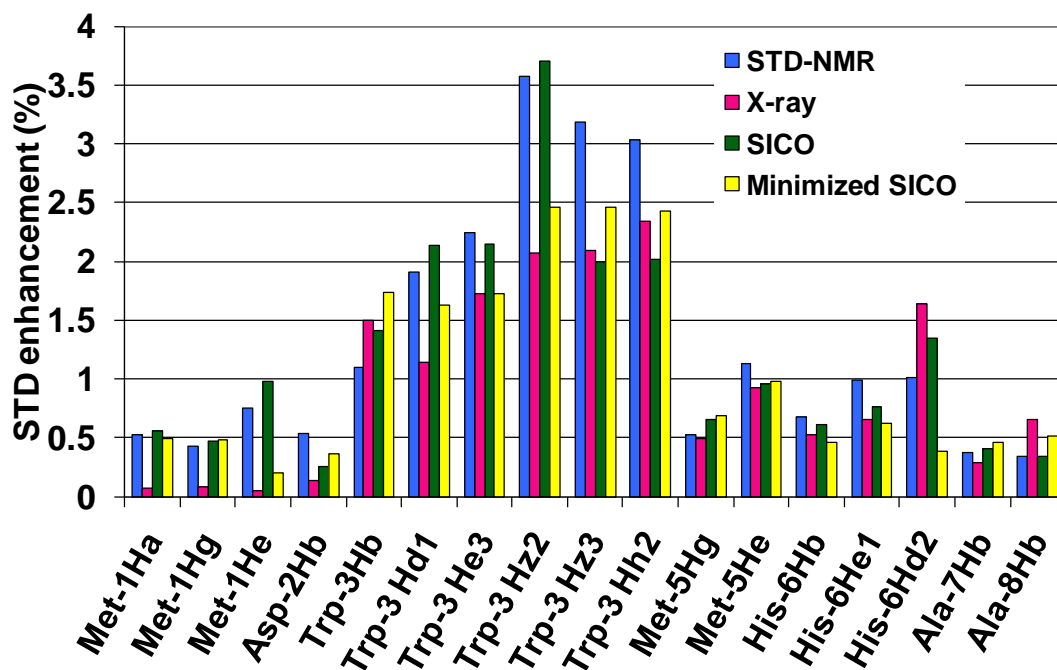


Figure 5.4. Comparison of experimental (blue), predicted by the CORCEMA-ST protocol based on the crystal structure of the peptide 1-SYA/J6 Fab complex (pink), simulated annealing-refined (green) STD values from the CORCEMA-ST protocol, and CORCEMA-ST values for the minimized-SICO structure (fluorescent yellow) for peptide 1 in the presence of mAb SYA/J6. Enhancements are shown for a saturation time of 5s.

We turned next to further optimization of our modeling protocol with CORCEMA-ST coupled with molecular dynamics simulations, and examined theoretical STD effects *during* the entire course of the trajectory, in contrast to our previous work^[26] and that described above which examined the use of averaged atomic positions of receptor-ligand complexes from MD simulations. A protocol that sampled snapshots throughout the MD trajectory, with the aim of obtaining an average R-factor based on the initial slope of the STD build-up curve, i.e. based on STD(fit) and not based on a single saturation time, was expected to give a more accurate representation of the bound ligand conformation. It was also of interest to minimize computational time and we chose, therefore, to test the feasibility of using the smallest binding domain, the variable region, Fv, for further simulations of peptide binding.

5.5.4 Molecular dynamics studies of peptide 1 in the Fv portion of monoclonal antibody SYA/J6

Before embarking on simulation of the conformations of peptide 1 bound to Fv, the Fab-bound molecular dynamics simulations (10.5ns) of peptide 1 (Figure S5.7) were examined. The number of hydrogen bonds between the peptide and Fab is maintained throughout the simulation, and the RMSD over the course of the simulation suggests that the system is stable. For the Fv-bound MD simulations, the number of hydrogen bonds between peptide 1 and Fv as well as the RMSD over the course of the simulation are shown in Figure S5.8.

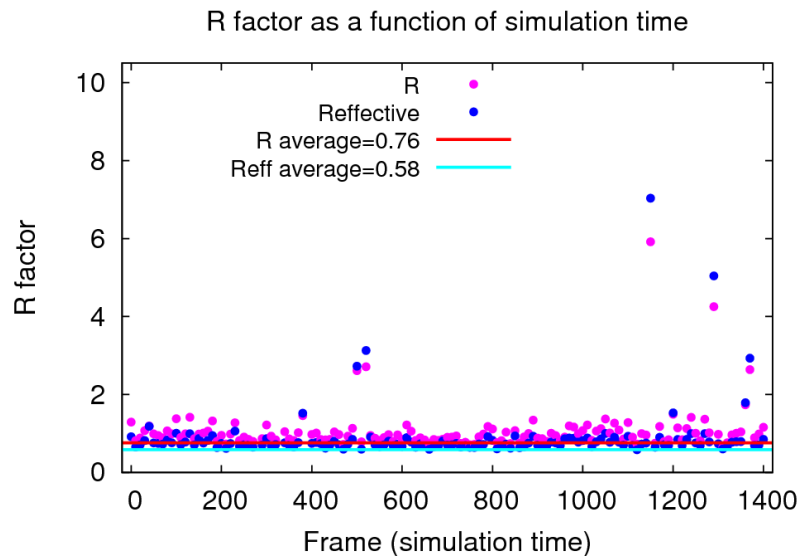
Comparison of the RMSD per residue for Fab and Fv indicates similar behavior (Figure S5.9). In both cases, it appears that peptide **1** samples similar conformational space. Ramachandran plots of the bound MDWNMHAA **1**, indicate that the ligand adopts α -helical structural elements in the MHAA region, which are induced upon binding to the antibody (Figure S5.10). This result is consistent with X-ray crystallographic data.^[13] We have hypothesized that this α -helical component constitutes an immunodominant feature of the peptide mimic.^[15] Indeed, the B-values from the crystal structure show similar behavior to B-values calculated from the trajectory; however, the latter are higher in amplitude, as seen in Figure S5.11. This phenomenon is typically observed due to crystallographic refinement procedures which tend to underestimate B factors.^[42] We also examined the RMSD per residue for chains A (light chain) and B (heavy chain) of the antibody in the presence and absence of peptide **1** which indicates that the antibody is slightly more flexible in the presence of the octapeptide (Figure S5.12). Examination of the radius of gyration (which examines the compactness of a molecule) indicates neither peptide nor antibody unfolds and van der Waals contacts between peptide and antibody are maintained (Figure S5.13).

A closer inspection of the MD trajectory shows that many interactions observed in the crystal structure are maintained in the simulation. Asp-2 swings out but swings back to maintain contacts with His L27D, and Asn-4 maintains interactions with Thr L91. A closer inspection of Trp-3 reveals that this residue experiences the most STD enhancement because the indole ring is buried in a

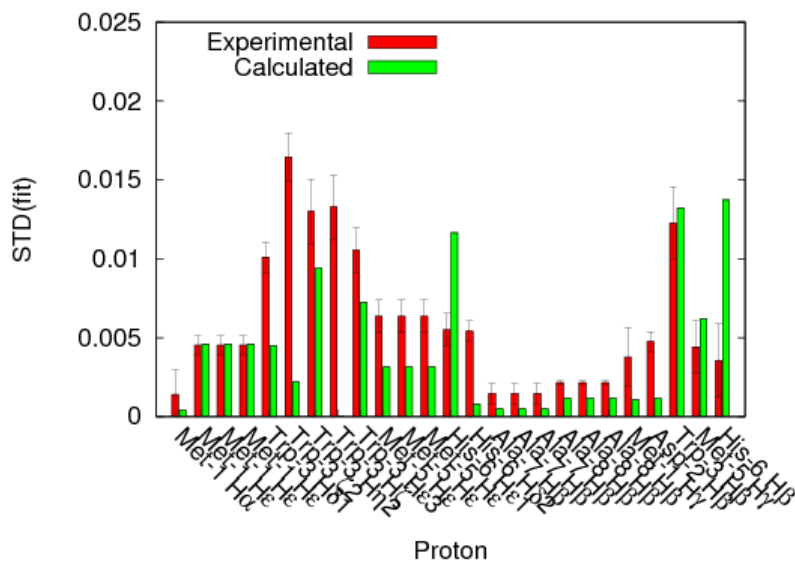
shallow cavity where it interacts with Val L94, Pro L95, His H58 and Trp H47. It is also interesting to note that in the simulation, His H58 swings in and engages in a π -edge interaction with Trp H47 which is maintained throughout the simulation. As in the crystal structure, the side chain of Met-5 interacts with Trp H33, which, in turn, maintains a cation- π contact with Arg H52. A contact between His-6 and Lys L50 is also maintained. This particular lysine is the first residue of a short loop CDR L2 but no movement of the loop is observed. Residue Ala-7 interacts with Tyr L32 and His L27D throughout the simulation. It is also apparent why enhancements for Met-1 are quite weak since this side chain is quite flexible and does not maintain prolonged contacts with the antibody during the course of the simulation.

The theoretical STD effects during the entire course of the trajectory were examined next with CORCEMA-ST. We used antibody residues within 12Å of the peptide. The R-factor was calculated on the basis of STD (fit) every 105ps over the course of the 10.5ns MD trajectory; an average value of 0.76 was obtained. We also introduced an effective R-factor, R_{eff} . If the calculated STD (fit) for proton k was within error of the experimental STD (fit) for proton k, then that calculated STD (fit) for proton k was not included in the calculation of the R-factor; an average R_{eff} of 0.58 was obtained (Figure 5.5a). We also note that the similar calculation, R factor based on STD (fit), on the static 12Å crystal structure gives a value of 0.64 and an R_{eff} of 0.50. Figure 5.5b shows a breakdown of the R-factor, namely STD (fit) per proton averaged over the MD trajectory. One observes poor fit for the tryptophan protons between calculated and experimental

STD (fit). More STD enhancement is observed for these protons experimentally, indicating that the tryptophan residue could be making more contacts with the antibody. We note also the discrepancy between experimental and calculated STD (fit) values for some of the histidine protons. At this point, we turned to STD-NMR Intensity-restrained CORCEMA Optimization (SICO) to explore different peptide backbone conformations.



a)



b)

Figure 5.5. a) R-factor as a function of MD simulation time (one frame = 0.0075ns. 1400 total frames = 10.5 ns trajectory) for peptide **1** bound to Fv. The R-factor is represented in pink. The effective R-factor, Reff, is represented in blue. b) The R-factor is broken down per proton. Thus, STD (fit) is shown per proton, averaged over the MD simulation.

5.5.5 Simulated annealing refinement of the binding mode of peptide 1 in the Fab of monoclonal antibody SYA/J6 coupled with CORCEMA-ST optimization

STD-NMR Intensity-restrained CORCEMA-ST Optimization (SICO) is a novel hybrid structure refinement protocol that progresses towards identifying the global minimum of the bioactive, bound conformation.^[17, 20, 21, 23] The simulated annealing (SA) refinement, based on the version by Alotto *et al.*,^[43] was used for optimizing selected torsion angles in peptide 1 to obtain the best fit between experimental and predicted STD-NMR intensities. The PDB coordinates used were from the peptide 1-SYA/J6 Fab complex. Incremental random changes in torsion angles were accepted or rejected according to the Metropolis criterion. In these calculations, the NOE *R*-factor is minimized. Torsion angles comprising residues Met-1, Asp-2, Trp-3, Asn-4 were optimized as dipeptide pairs to reduce computational time (Figure 5.6). Thus, torsion angles ϕ_1 (Met-1 H–Met-1 N–Met-1 C $_{\alpha}$ –Met-1 C), ψ_1 (Met-1 N–Met-1 C $_{\alpha}$ –Met-1 C–Met-1 O), ϕ_2 (Asp-2 H–Asp-2 N–Asp-2 C $_{\alpha}$ –Asp-2 C) and ψ_2 (Asp-2 N–Asp-2 C $_{\alpha}$ –Asp-2 C–Asp-2 O) were optimized to obtain the best bound conformation for residues Met-1 and Asp-2. Torsion angles ϕ_3 (Trp-3 H–Trp-3 N–Trp-3 C $_{\alpha}$ –Trp-3 C), ψ_3 (Trp-3 N–Trp-3 C $_{\alpha}$ –Trp-3 C–Trp-3 O), ϕ_4 (Asn-4 H–Asn-4 N–Asn-4 C $_{\alpha}$ –Asn-4 C) and ψ_4 (Asn-4 N–Asn-4 C $_{\alpha}$ –Asn-4 C–Asn-4 O) were optimized to obtain the best bound conformation for residues Trp-3 and Asn-4. Optimization of residues Met-5, His-6, Ala-7 and Ala-8 did not result in significant improvement of the NOE *R*-factor. This indicates that residues Met-5, His-6, Ala-7 and Ala-8 exhibit good correlation with the X-ray crystal structure of the peptide 1-SYA/J6 Fab complex. The final

structure was minimized without STD restraints. Bond rotation was sampled from -60° to $+60^\circ$. Optimized torsion angles for the peptide 1-SYA/J6 Fab complex are shown in Table 5.2 and STD values calculated for the SICO structure are shown in Figure 5.4.

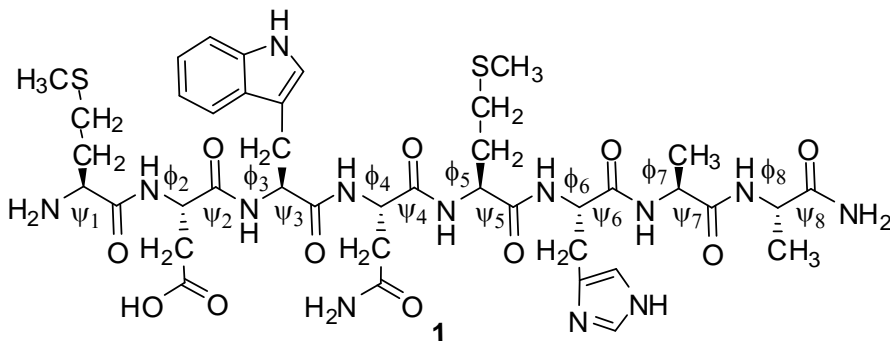


Figure 5.6. Torsion angles of peptide 1 optimized to reflect the bound conformation in the solution state.

In the SA-refined, bound solution conformation of peptide 1, residue Met-1 makes many more contacts with the antibody combining site, notably to Arg H52, relative to the X-ray crystal structure (Figure 5.7). This binding mode is more representative of the experimental STD-NMR enhancements observed for Met-1. Similarly, residue Asp-2 makes more contacts with the mAb combining site after refinement and this is a better reflection of the experimental data. Many more contacts were also observed for residue Trp-3 after optimization (Figure 5.7) to account for the strong enhancements observed. The R-factor for the SA-refined structure of peptide 1 bound to mAb SYA/J6 improved to 0.26. However, this

structure suffered from 40 steric clashes. We performed an Amber minimization on the peptide only, and we were able to obtain a virtually steric clash-free structure with an R-factor of 0.37 (Figure 5.8). The theoretical STD effects calculated with CORCEMA, shown in Figure 5.4, show reasonable agreement with experimental values, especially for the Trp moiety.

Table 5.2. Comparison of the torsion angles from the X-ray crystal structure and the SA- refined structure of the bound conformation of peptide **1** in the combining site of SYA/J6.

Torsion angle	X-Ray Crystal Structure	SICO Structure	Minimized SICO
Met-1 ψ_1	-162.6	-162.7	-159.9
Asp-2 ϕ_2	-107.3	-107.4	-104.0
Asp-2 ψ_2	142.4	142.2	146.3
Trp-3 ϕ_3	-115.9	-115.9	-104.6
Trp-3 ψ_3	25.1	25.3	7.2
Asn-4 ϕ_4	-161.5	-161.5	-170.4
Asn-4 ψ_4	153.3	153.4	163.0
Met-5 ϕ_5	-59.9	-59.9	-67.3
Met-5 ψ_5	-30.0	-30.1	-32.7
His-6 ϕ_6	-67.5	-67.4	-64.0
His-6 ψ_6	-43.1	-42.9	-45.9
Ala-7 ϕ_7	-75.0	-75.2	-70.9
Ala-7 ψ_7	-32.2	-32.1	-35.4
Ala-8 ϕ_8	-60.6	-60.6	-54.8
Ala-8 ψ_8	-51.0	-51.1	-60.1

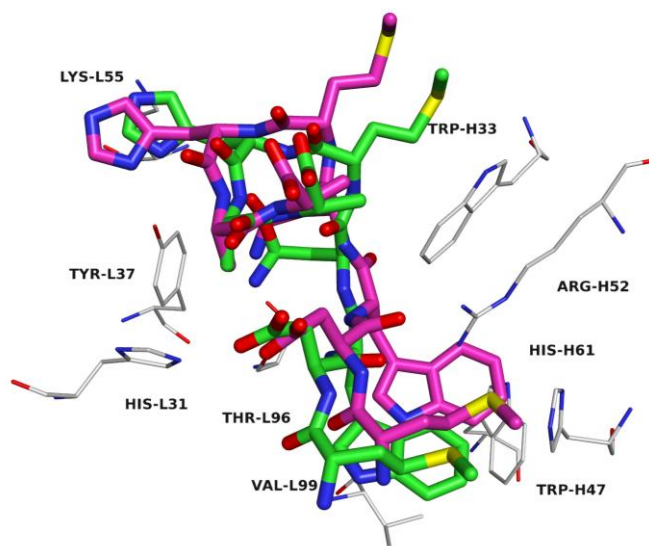


Figure 5.7. Alignment of X-ray structure (green) and SICO structure (pink) showing key contacts. The alignment used Pymol’s “align” function which performs a sequence alignment (to capture the antibody in addition to the peptide), followed by a structural alignment.

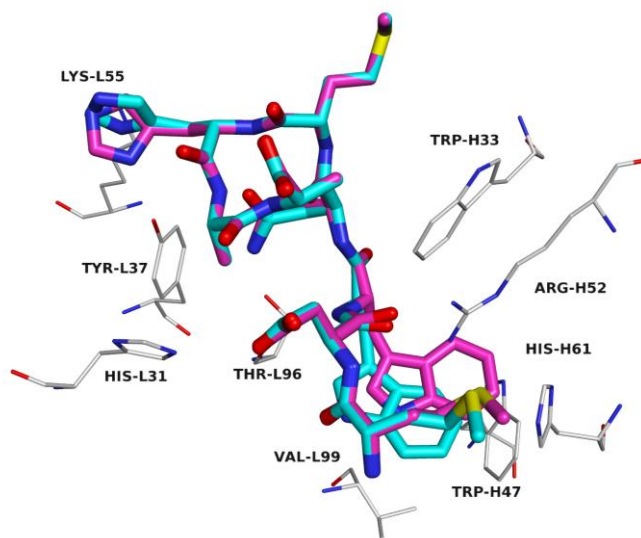


Figure 5.8. Alignment of SICO structure (pink) and minimized SICO structure (blue) showing key contacts. The alignment used Pymol’s “align” function which performs a sequence alignment (to capture the antibody in addition to the peptide), followed by a structural alignment.

5.5.6 Binding of the pentasaccharide hapten 2 to monoclonal antibody SYA/J6 by saturation transfer difference NMR spectroscopy

Attempts were made to investigate the binding of the pentasaccharide hapten **2** to mAb SYA/J6 using STD-NMR spectroscopy. Initially, an STD-NMR spectrum at 298K with 4096 scans using a spin lock to remove the broad antibody signal was obtained. The signal was extremely weak (Figure 5.9c). Attempts to improve the signal strength by varying the temperature (310K or 282K) led to no visible improvement. At that point, a modified version of the pulse sequence that incorporates water suppression was used. Unfortunately, the poor signal to noise, even after 20480 scans, made quantitation via full STD build-up curves impossible (Figure 5.9d). A comparison of the STD-NMR spectrum for the pentasaccharide bound to mAb SYA/J6 and the STD-NMR spectrum of the peptide bound to mAb SYA/J6 is shown in Figure 5.9, highlighting the difference in signal intensity between the two haptens. We propose that the weak STD-NMR signal in the spectrum of the pentasaccharide **2** is due to conformational reordering that must occur in the pentasaccharide to ensure optimal binding to the antibody. We have shown, previously, that penetration of residue C into a deep pocket of the combining site of the Fab mAb SYA/J6 involves changes in the glycosidic linkages of ABCDA' relative to the unbound state. ^[14] We contend that this reordering indicates that the binding of the pentasaccharide to the antibody is *not* diffusion controlled, i.e. k_{on} is *less than* 10^7 - 10^8 $M^{-1} s^{-1}$; from the equilibrium constant, we can then estimate k_{off} as being *less than* $1.0 s^{-1}$. Thus, even though magnetization transfer

from the antibody to the ligand occurs, that information is not being relayed efficiently to the free ligand in solution where detection of magnetization occurs because dissociation of the the ligand-antibody complex is slow relative to the relaxation times of the ligand protons.

Nonetheless, we attempted a qualitative analysis of the STD-NMR data at a saturation time of 2s. The epitope map is shown in Figure 5.10. We found that the greatest enhancements were observed for the NHAc group of ring D and the OMe group of ring A'. This is not surprising as we have shown that the D sugar is involved in the largest number of contacts to the antibody.^[14] The NHAc group, in particular the NH makes hydrogen-bonding interactions with Thr L91. The same NHAc group makes van der Waals contacts to His L27D and Tyr L32. Ring A' makes hydrogen bonding interactions to His L27D in addition to van der Waals contacts to His L27D, Thr L92 and His L93. The OMe group, in particular, interacts with Val-L99. The 6-methyl group of Rha C is also enhanced, consistent with its placement inside the deep pocket (occupied by water molecules in the case of the peptide) where it makes hydrophobic interactions with Met H100A and Thr L91. These interactions are consistent with the STD enhancements observed. It is interesting to note that there are two signals for the NHAc group due to restricted rotation about the amide bond and both of these are enhanced. This result could signify that the antibody binds to both conformational isomers, or alternatively, that one isomer binds and transfers magnetization to the other through exchange, which is still significant on the relaxation time scale. (Figures 5.9c,d)

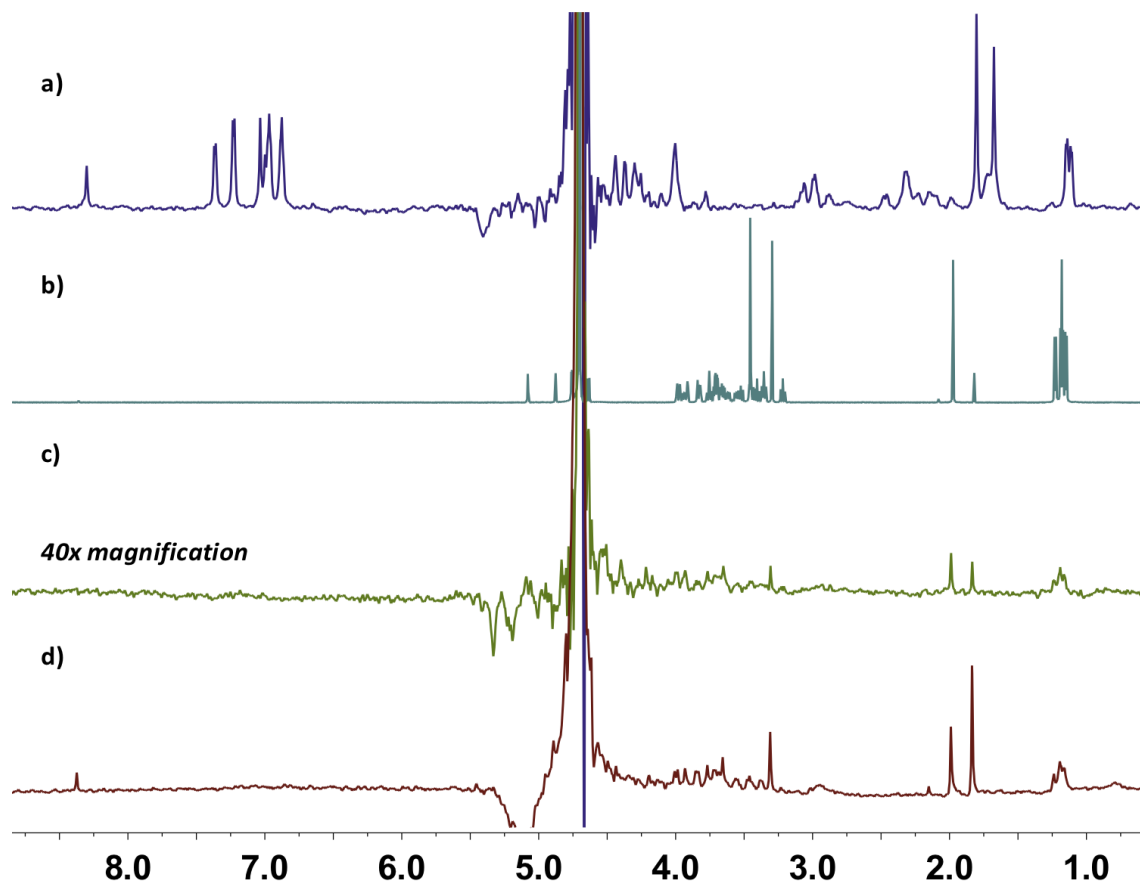


Figure 5.9. a) STD NMR spectrum of the octapeptide **1** in the presence of mAb SYA/J6, b) ¹H NMR spectrum of the pentasaccharide **2**, c) STD NMR spectrum of the pentasaccharide **2** in the presence of SYA/J6 at 298K, with a spin-lock, after 4K scans shown with 40 times magnification, and d) STD NMR spectrum of the pentasaccharide **2** in the presence of SYA/J6, at 298K with water suppression and a spin-lock, after 20K scans.

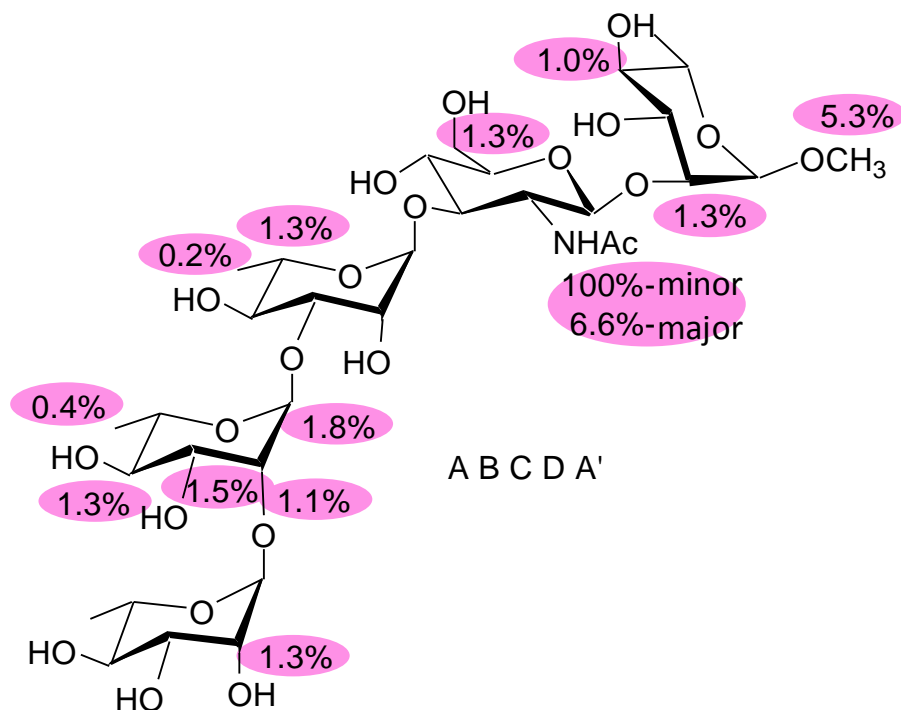


Figure 5.10. Epitope map of pentasaccharide bound to mAb SYA/J6 derived from the STD-NMR spectrum at 298K, recorded with 20K scans and a saturation time of 2s. STD enhancements were measured relative to the signals in an unsaturated reference spectrum i.e. $(I_0 - I_{\text{sat}}) / I_0$ to correct for a single resonance that may encompass three protons (e.g. OMe or NHAc) or two protons (e.g. methylene CH_2). The methyl signal of the major NHAc isomer is at 2.06 ppm and that of the minor NHAc isomer is at 1.91 ppm. The minor isomer is most enhanced (100%); all enhancements were measured relative to this signal.

5.6 Conclusions

We have used two approaches to obtain insight into the solution structure of octapeptide **1** bound to the mAB, SYA/J6. In our first approach we have shown that CORCEMA-ST is ideally suited to interface with MD simulations, whereby STD-NMR enhancements arising from more than one conformation can be estimated. Theoretical STD effects calculated at snapshots during the entire course of the trajectory and not the averaged atomic positions of receptor-ligand complexes have been calculated using CORCEMA-ST. These values have been used to calculate a time-averaged R-factor that is based on STD (fit); this R-factor is a better representation of the dynamics occurring in the solution state. This method has several advantages. It allows for the use of STD(fit) which correspond to the STD intensity in the absence of T1 bias and depends on the proximity of the ligand proton to the protein.^[41] In essence one captures the initial “build-up phase” akin to an NOE effect. Also, this method allows for the movement and dynamics of the *ligand and receptor* to be monitored and quantified.

In our second approach, we have shown that the model of the bound peptide conformation derived from the SICO protocol is superior to that obtained from previous studies, as reflected in better fit with experimental data. It is apparent that peptide geometry displayed in the crystal structure does not completely reflect the interactions that exist in the solution state. Modifications to the octapeptide backbone via simulated annealing resulted in a greater number of interactions between the peptide and the heavy chain. This method

differs from the previous one in that it relies on a single saturation time, in this case 5s, to calculate an R-factor which is used to refine *solely the ligand conformation* from a static crystal structure. Based on the R-factor, this method appears to be superior.

The combined data provide insight into the binding of the peptide mimic MDWNMHAA 1 by the monoclonal antibody SYA/J6, and lend credence to the model of binding which can now be used to design the next-generation, higher-affinity binders for use as vaccine candidates.

5.7 Acknowledgements

We are grateful to the Natural Sciences and Engineering Council of Canada for financial support and to the Accelerate BC-MITACS program and Zymeworks Inc. for internships to MGS. We are also grateful to D.R. Bundle for the generous gift of the SYA/J6 antibody, and to R. N. Krishna for helpful discussions.

5.8 Summary

In this chapter, we have described a detailed study of MDWNMHAA **1** when bound to antibody SYA/J6. The acquisition of full STD-NMR build-up curves to determine the binding components of the peptide are also described. This experimental data was compared to a theoretical treatment by CORCEMA-ST, a program that calculates cross-relaxation effects, and a poor agreement of solution NMR data to the solid state crystallographic image of the bound structure of MDWNMHAA was found. This poor agreement prompted us to turn to molecular dynamics in an effort to obtain an accurate picture of the bound-ligand conformations when MDWNMHAA **1** bind to SYA/J6.

5.9 References

- [1] T. L. Hale, in *Topley and Wilson's Microbiology and Microbial Infections, Vol. 3* (Eds.: W. J. Hansler, M. Shuman), Hodder Arnold, London, **1998**, p. 479.
- [2] K. L. Kotloff, J. P. Winickoff, B. Ivanoff, J. D. Clemens, D. L. Swerdlow, P. J. Sansonetti, G. K. Adak, M. M. Levine, *Bull. World Health Organ.* **1999**, *77*, 651-666.
- [3] D. R. Bundle, M. A. J. Gidney, S. Josephson, H. P. Wessel, *ACS Symp. Ser.* **1983**, *231*, 49-63.
- [4] N. I. A. Carlin, M. A. J. Gidney, A. A. Lindberg, D. R. Bundle, *J. Immunol.* **1986**, *137*, 2361-2366.
- [5] L. Kenne, B. Lindberg, K. Petersson, E. Katzenellenbogen, E. Romanowska, *Eur. J. Biochem.* **1977**, *76*, 327-330.
- [6] L. Kenne, B. Lindberg, K. Petersson, E. Katzenellenbogen, E. Romanowska, *Eur. J. Biochem.* **1978**, *91*, 279-284.
- [7] L. Kenne, B. Lindberg, K. Petersson, E. Romanowska, *Carbohydr. Res.* **1977**, *56*, 363-370.
- [8] R. B. Hossany, M. A. Johnson, A. A. Eniade, B. M. Pinto, *Biorg. Med. Chem.* **2004**, *12*, 3743-3754.

- [9] M. A. Johnson, B. M. Pinto, *Aust. J. Chem.* **2002**, *55*, 13-25.
- [10] B. Monzavi-Karbassi, G. Cunto-Amesty, P. Luo, T. Kieber-Emmons, *Trends Biotechnol.* **2002**, *20*, 207-214.
- [11] S. Borrelli, R. B. Hossany, B. M. Pinto, *Clin. Vaccine Immunol.* **2008**, *15*, 1106-1114.
- [12] S. L. Harris, L. Craig, J. S. Mehroke, M. Rashed, M. B. Zwick, K. Kenar, E. J. Toone, N. Greenspan, F. I. Auzanneau, J. R. Marino-Albernas, B. M. Pinto, J. K. Scott, *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 2454-2459.
- [13] N. K. Vyas, M. N. Vyas, M. C. Chervenak, D. R. Bundle, B. M. Pinto, F. A. Quiocho, *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 15023-15028.
- [14] N. K. Vyas, M. N. Vyas, M. C. Chervenak, M. A. Johnson, B. M. Pinto, D. R. Bundle, F. A. Quiocho, *Biochemistry* **2002**, *41*, 13575-13586.
- [15] M. A. Johnson, B. M. Pinto, in *Topics in Current Chemistry: Bioactive Conformation II, Vol. 273* (Ed.: T. Peters), Springer-Verlag, Heidelberg, Germany, **2008**, pp. 55-116.
- [16] M. A. Johnson, B. M. Pinto, *Bioorg. Med. Chem.* **2004**, *12*, 295-300.

- [17] V. Jayalakshmi, T. Biet, T. Peters, N. R. Krishna, *J. Am. Chem. Soc.* **2004**, *126*, 8610-8611.
- [18] V. Jayalakshmi, T. Biet, T. Peters, N. R. Krishna, *J. Am. Chem. Soc.* **2005**, *127*, 7261-7261.
- [19] V. Jayalakshmi, N. R. Krishna, *J. Magn. Reson.* **2002**, *155*, 106-118.
- [20] V. Jayalakshmi, N. R. Krishna, *J. Magn. Reson.* **2004**, *168*, 36-45.
- [21] V. Jayalakshmi, N. R. Krishna, *J. Am. Chem. Soc.* **2005**, *127*, 14080-14084.
- [22] N. R. Krishna, D. G. Agresti, J. D. Glickson, R. Walter, *Biophys. J.* **1978**, *24*, 791-814.
- [23] N. R. Krishna, V. Jayalakshmi, *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *49*, 1-25.
- [24] X. Wen, Y. Yuan, D. A. Kuntz, D. R. Rose, B. M. Pinto, *Biochemistry* **2005**, *44*, 6729-6737.
- [25] Y. Yuan, X. Wen, D. A. R. Sanders, B. M. Pinto, *Biochemistry* **2005**, *44*, 14080-14089.
- [26] Y. Yuan, D. W. Bleile, X. Wen, D. A. R. Sanders, K. Itoh, H. W. Liu, B. M. Pinto, *J. Am. Chem. Soc.* **2008**, *130*, 3157-3168.

- [27] A. Bhunia, V. Jayalakshmi, A. J. Benie, O. Schuster, S. Kelm, N. R. Krishna, T. Peters, *Carbohydr. Res.* **2004**, *339*, 259-267.
- [28] T. D. W. Claridge, in *Tetrahedron Organic Chemistry Series, Vol. 19* (Eds.: J. E. Baldwin, R. M. Williams), Elsevier Science Ltd, Kidlington, **1999**, p. 283.
- [29] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712-725.
- [30] H. C. Andersen, *J. Comput. Phys.* **1983**, *52*, 24-34.
- [31] S. Miyamoto, P. A. Kollman, *J. Comput. Chem.* **1992**, *13*, 952-962.
- [32] H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684-3690.
- [33] D. S. Wishart, B. D. Sykes, F. M. Richards, *Biochemistry* **1992**, *31*, 1647-1651.
- [34] D. S. Wishart, B. D. Sykes, *J. Biomol. NMR* **1994**, *4*, 171-180.
- [35] S. Schwarzinger, G. J. A. Kroon, T. R. Foss, P. E. Wright, H. J. Dyson, *J. Biomol. NMR* **2000**, *18*, 43-48.
- [36] G. M. Clore, A. M. Gronenborn, *J. Magn. Reson.* **1982**, *48*, 402-417.

- [37] G. M. Clore, A. M. Gronenborn, *J. Magn. Reson.* **1983**, *53*, 423-442.
- [38] J. P. Albrand, B. Birdsall, J. Feeney, G. C. K. Roberts, A. S. V. Burgen, *Int. J. Biol. Macromol.* **1979**, *1*, 37-41.
- [39] P. Balaram, A. A. Bothner-By, E. Breslow, *J. Am. Chem. Soc.* **1972**, *94*, 4017-4018.
- [40] P. Balaram, A. A. Bothner-By, J. Dadok, *J. Am. Chem. Soc.* **1972**, *94*, 4015-4017.
- [41] M. Mayer, T. L. James, *J. Am. Chem. Soc.* **2004**, *126*, 4453-4460.
- [42] J. Kuriyan, G. A. Petsko, R. M. Levy, M. Karplus, *J. Mol. Biol.* **1986**, *190*, 227-254.
- [43] P. G. Alotto, C. Eranda, B. Brandstatter, G. Furnstratt, C. Magele, G. Molinari, M. Nervi, K. Preis, M. Repetto, K. R. Richter, *IEEE Trans. Magn.* **1998**, *34*, 3674-3684.

5.10 Supporting Information

Investigation of the Binding of a Carbohydrate-mimetic Peptide to its Complementary Anti-carbohydrate Antibody by STD-NMR Intensity-restrained CORCEMA Optimization (SICO) and Molecular Dynamics Simulations

Monica G. Szczepina,^a Dustin W. Bleile,^b and B. Mario Pinto^a

^aDepartment of Chemistry, Simon Fraser University, Burnaby, British Columbia, Canada, V5A 1S6

^bZymeworks Inc., 540-1385 West 8th Ave, Vancouver, British Columbia, Canada, V6H 3V9

Table of Contents

Table S5.1	¹ H NMR chemical shifts of MDWNMHAA 1 in D ₂ O.	206
Table S5.2	¹³ C NMR chemical shifts of MDWNMHAA 1 in D ₂ O.	207
Table S5.3	CSI output for peptide MDWNMHAA 1 , indicating a random coil arrangement.	208
Figure S5.1	Number of hydrogen bonds and RMSD of free MDWNMHAA as function of simulation time.	209
Figure S5.2	Ramachandran plots of free MDWNMHAA during the course of the MD simulation.	210
Figure S5.3	2D RMSD plot of free MDWNMHAA.	211
Figure S5.4	Expansion of 1D ¹ H NMR (upper trace) and STD-NMR (lower trace) spectra of peptide 1 at 600 MHz and 282 K in the presence of mAb SYA/J6	212
Figure S5.5	Expansion of 1D ¹ H NMR (upper trace) and STD-NMR (lower trace) spectra of peptide 1 at 600 MHz and 282 K in the presence of mAb SYA/J6.	213
Figure S5.6	STD build-up curves for the protons in MDWNMHAA 1 . Experimental data were fit to a rising exponential with an offset.	214
Figure S5.7	Number of hydrogen bonds and RMSD of SYA/J6 Fab bound to MDWNMHAA 1 as a function of simulation time.	217
Figure S5.8	Number of hydrogen bonds and RMSD of SYA/J6 Fv bound to MDWNMHAA 1 as a function of simulation time.	218
Figure S5.9	a) RMSD per residue for the light chain of Fab bound to MDWNMHAA 1 and the light chain of Fv bound to MDWNMHAA 1 . b) RMSD per residue for the heavy chain of Fab bound to MDWNMHAA 1 and the heavy chain of Fv bound to MDWNMHAA 1 .	219
Figure S5.10	Ramachandran plots of MDWNMHAA 1 bound to Fv during the course of the MD simulation.	220

Figure S5.11	Comparison of B values from the MD simulation of MDWNMHAA 1 bound to Fv with B values from the X-ray crystal structure.	221
Figure S5.12	a) RMSD per residue for the light chain of Fv bound to MDWNMHAA 1 and in the absence of MDWNMHAA 1. b) RMSD per residue for the heavy chain of Fv bound to MDWNMHAA 1 and in the absence of MDWNMHAA 1.	222
Figure S5.13	Radius of gyration and van der Waals contacts.	223

Table S5.1. ^1H NMR chemical shifts of MDWNMHAA **1** in D_2O . Values in the right-hand column refer to literature values for the alpha protons.

<u>Resonance</u>	<u>δ (ppm)</u>	<u>δ (ppm)^{25,26}</u>
Ala-7H β	1.36	
Ala-8H β	1.39	
Met-1/Met-5 H β	1.86-2.00	
Met-1H ϵ	1.94	
Met-5H ϵ	2.05	
Met-1 H γ	2.24	
Met-1 H γ	2.36	
Met-5 H γ	2.39	
Met-5 H γ	2.47	
Asn-4H β	2.55	
Asn-4H β	2.63	
Asp-2H β	2.71	
Asp-2H β	2.85	
His-6H β	3.13	
Trp-3H β	3.21	
His-6H β	3.25	
Trp-3H β	3.30	
Met-1H α	4.00	4.52 \pm 0.1
Ala-7H α	4.24	4.35 \pm 0.1
Met-5H α	4.26	4.52 \pm 0.1
Ala-8H α	4.27	4.35 \pm 0.1
Asn-4H α	4.54	4.75 \pm 0.1
Trp-3H α	4.60	4.7 \pm 0.1
His-6H α	4.65	4.63 \pm 0.1
Asp-2H α	4.74	4.76 \pm 0.1
Trp-3H η 2	7.13	
Trp-3H ζ 3	7.22	
Trp-3H δ 1/ His-6H ϵ 1	7.27	
Trp-3H ζ 2	7.47	
Trp-3H ϵ 3	7.62	
His-6H δ 2	8.58	

Table S5.2. ^{13}C NMR chemical shifts of MDWNMHAA **1** in D_2O .

<u>Resonance</u>	<u>$\delta^{13}\text{C}^\alpha$ (ppm)</u>	<u>$\delta^{13}\text{C}^\beta$ (ppm)</u>
Met-1	54.6	32.3
Asp-2	52.7	38.4
Trp-3	54.9	28.6
Asn-4	52.9	38.2
Met-5	55.6	32.3
His-6	57.4	29.3
Ala-7	52.2	19.2
Ala-8	52.1	19.0

Table S5.3. CSI output for peptide MDWNMHAA 1, indicating a random coil arrangement. The column labeled consensus, takes into account the H- α , C- α and C- β chemical shifts and predicts a C for each residue. The abbreviation C means random coil.

```

#
#
#####
# Program...:          CSI (c)
# Version...:         2.0
# Location...:         University of Alberta
#                   Protein Engineering Network of
#                   Centres of Excellence
# Input.....:         csi.OP
# Date.....:          Wed Aug 23 15:34:50 2006
#####
#
#
#   A      HA      CA      CO      CB      Consensus
#
1   M      0 C     0 C     NA     0 C     0 C
2   D      0 C    -1 B     NA     0 C     0 C
3   W     -1 H    -1 B     NA     0 C     0 C
4   N     -1 H    -1 B     NA     0 C     0 C
5   M     -1 H    -1 B     NA     0 C     0 C
6   H     -1 H     0 C     NA     0 C     0 C
7   A      0 C     0 C     NA     0 C     0 C
8   A      0 C     0 C     NA     0 C     0 C
#
#
#####
#
#                   Secondary Structure Summary
#
#
#
#   HA      |      CA      |      CO      |      CB      |      Consensus      |
#-----|-----|-----|-----|-----|
# C  1 - 2 | C  1 - 1 |      NA      | C  1 - 8 | C  0 - 0 |
# H  3 - 6 | B  2 - 5 |      NA      |          |          |
# C  7 - 8 | C  6 - 8 |      NA      |          |          |
#

```

Figure S5.1. a) Number of hydrogen bonds existing within free MDWNMHAA as a function of simulation time. b) RMSD of free MDWNMHAA as a function of simulation time.

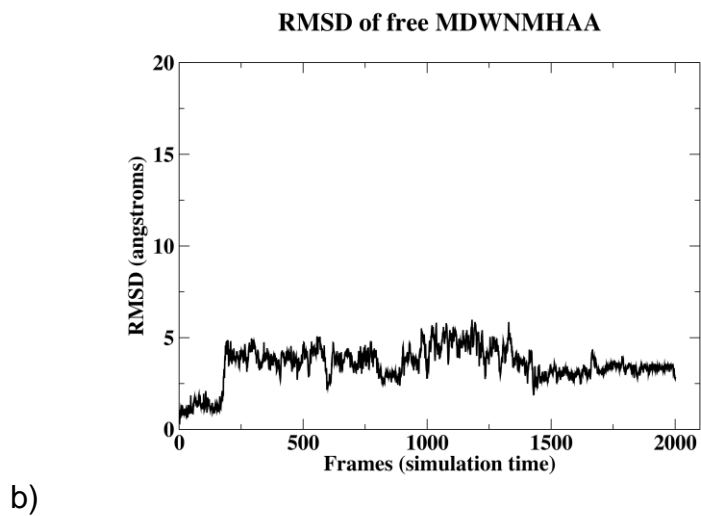
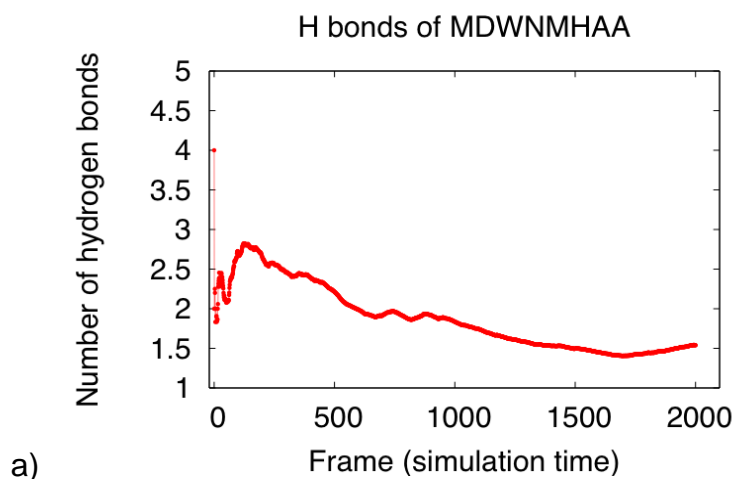


Figure S5.2. Ramachandran plots of free MDWNMHAA during the course of the MD simulation. Each panel refers to phi and psi angles of a particular residue. Panels from a) to f) refer to aspartic acid, tryptophan, asparagine, methionine, histidine, and alanine, respectively.

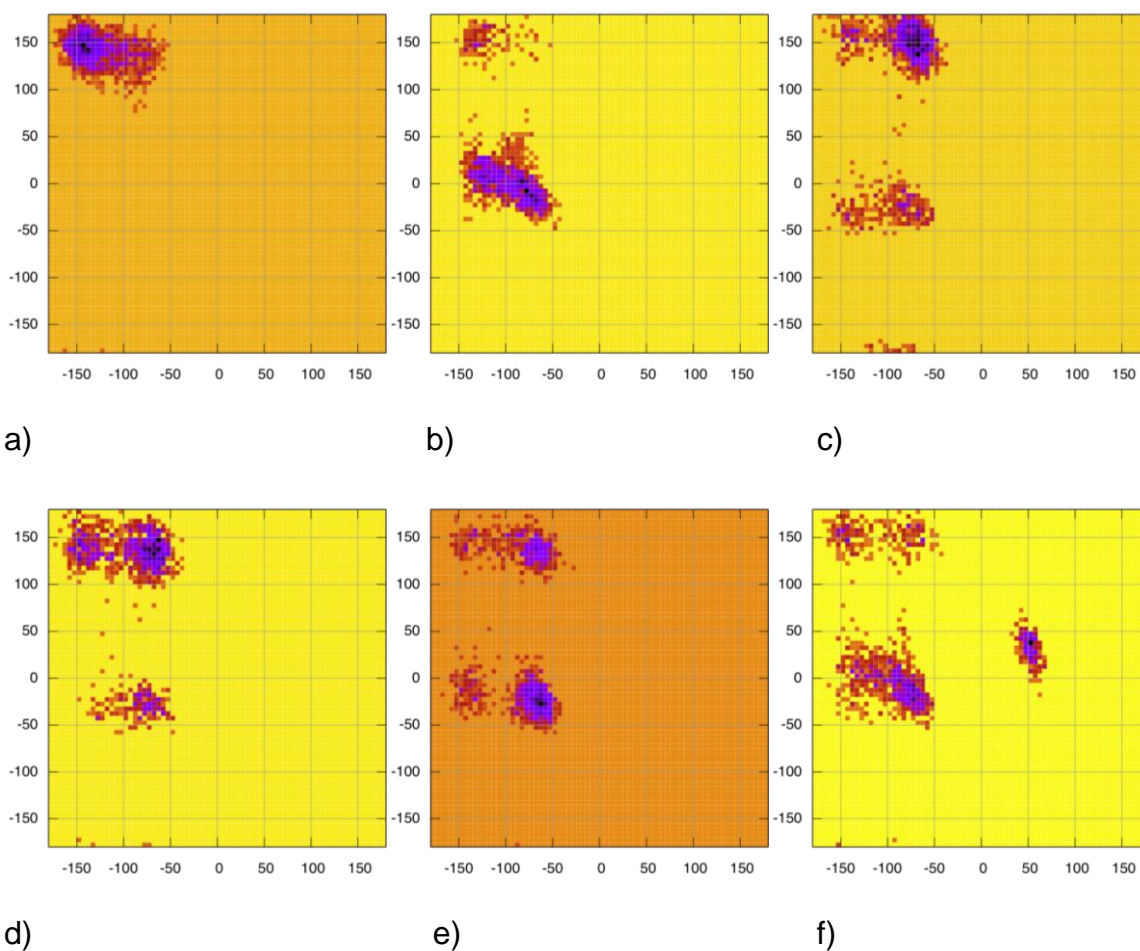


Figure S5.3. 2D RMSD plot of free MDWNMHAA MD trajectories. RMSD is calculated at a time point and compared to every other time point. Yellow regions indicate low RMSD values ($< 5\text{\AA}$).

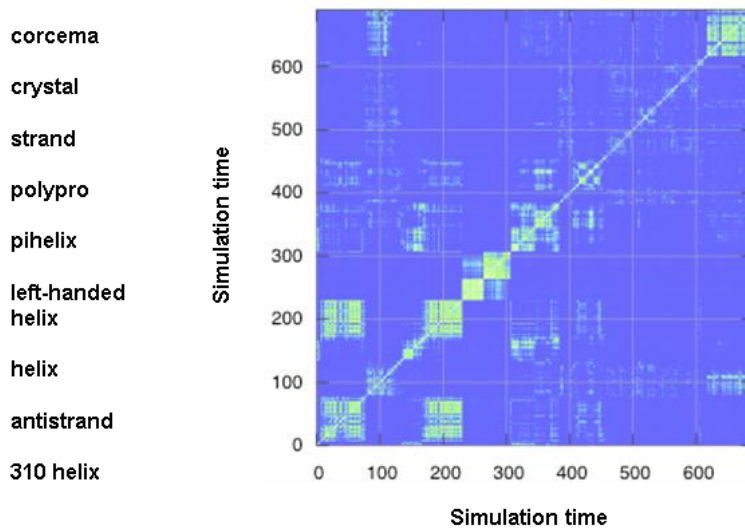


Figure S5.4. Expansion of 1D ^1H NMR (upper trace) and STD-NMR (lower trace) spectra of peptide 1 at 600 MHz and 282 K in the presence of mAb SYA/J6.

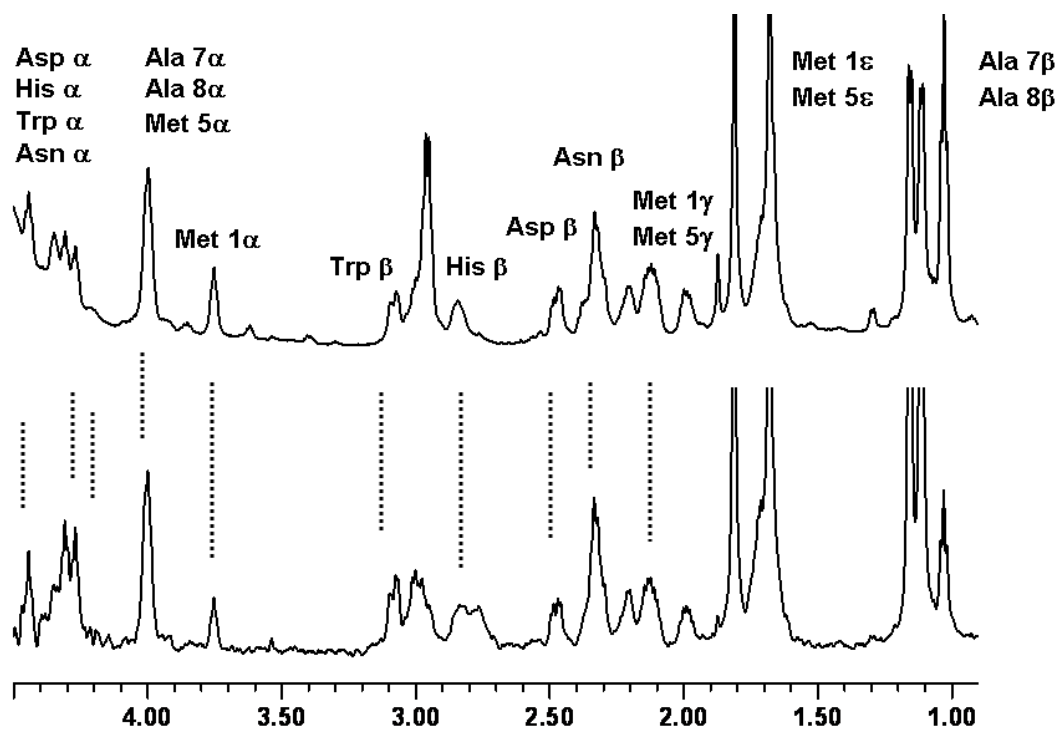


Figure S5.5. Expansion of 1D ^1H NMR (upper trace) and STD-NMR (lower trace) spectra of peptide **1** at 600 MHz and 282 K in the presence of mAb SYA/J6.

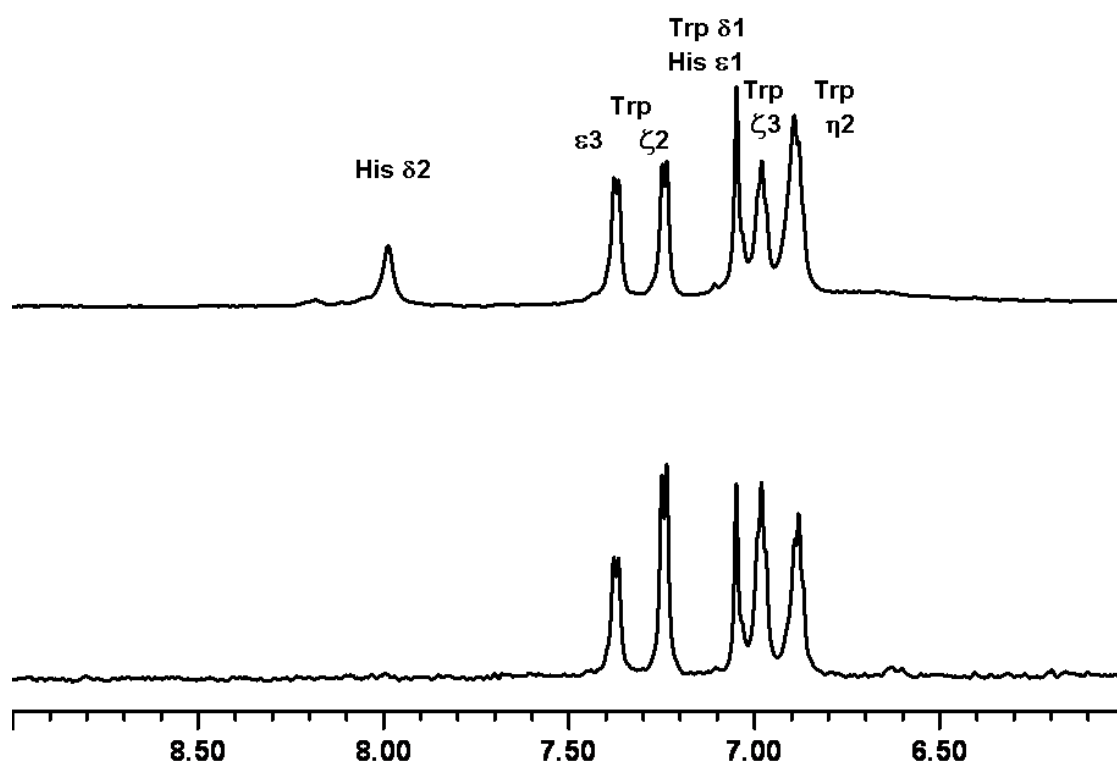
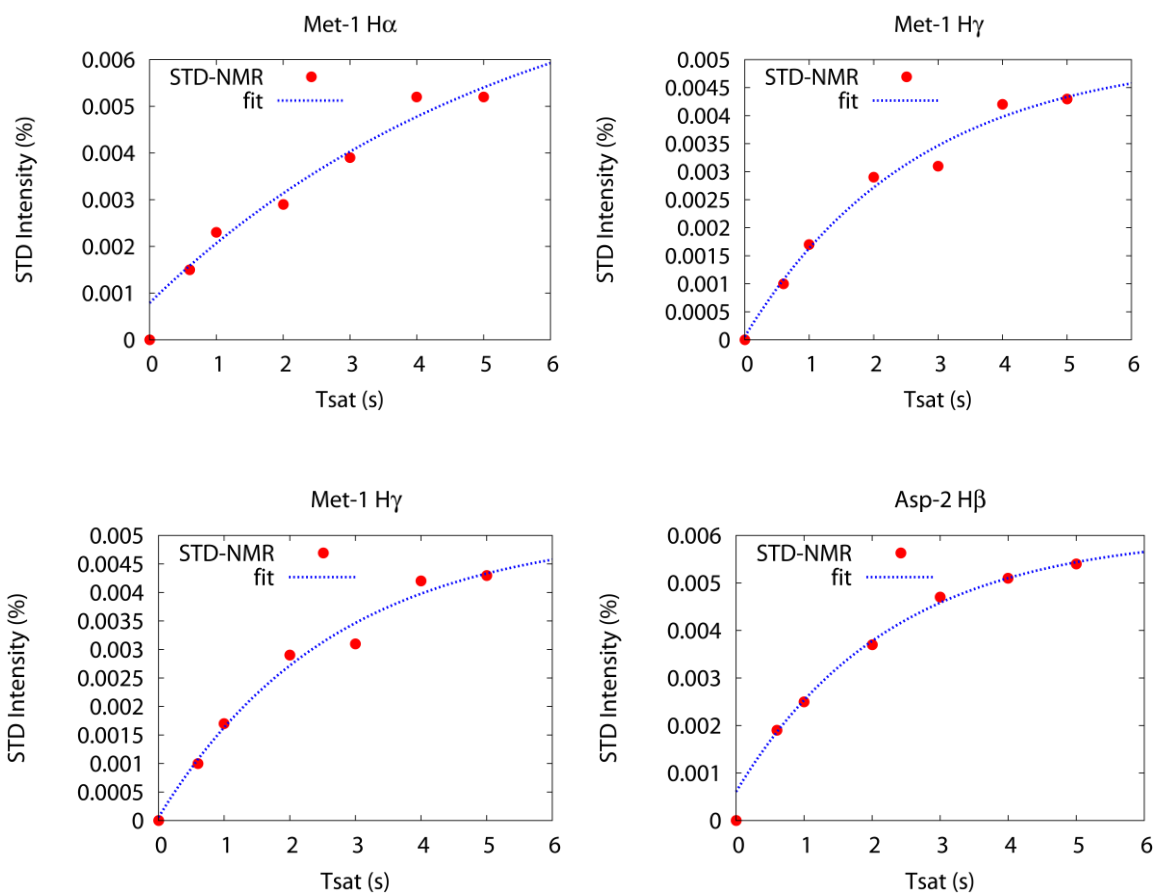
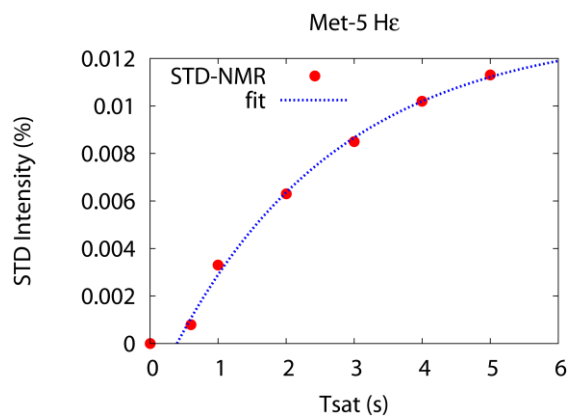
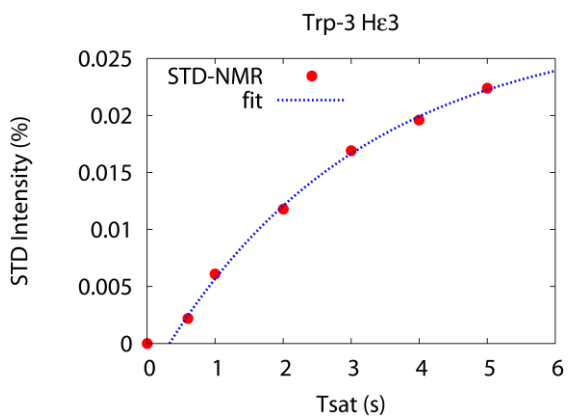
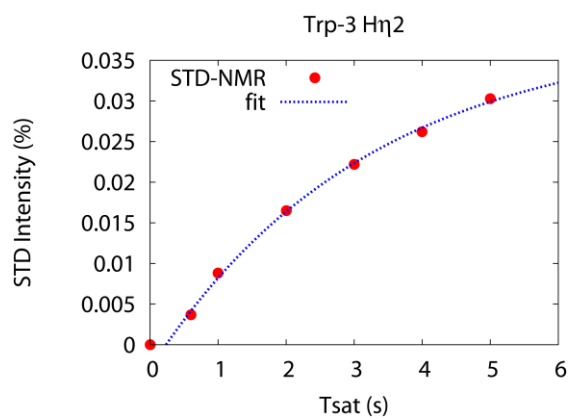
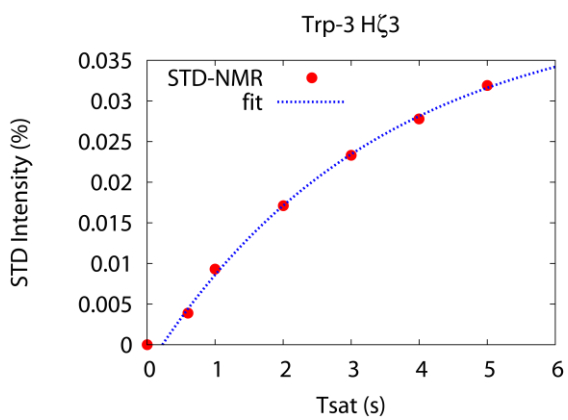
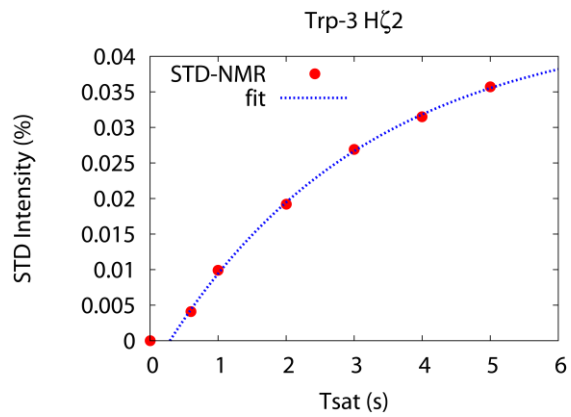
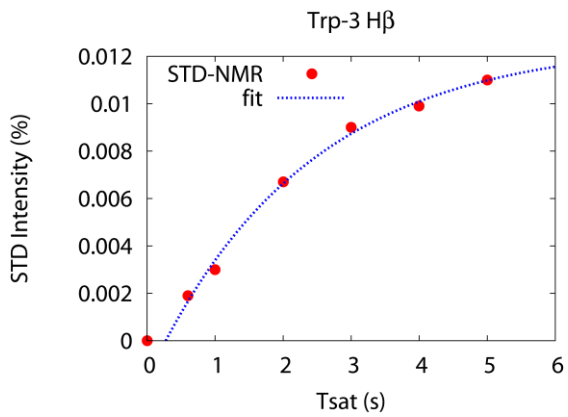


Figure S5.6. STD build-up curves for the protons in MDWNMHA. The maximal STD intensity, STD_{max} , and the observed saturation rate constant k_{sat} were obtained from fitting the saturation time data to the monoexponential equation: $STD = STD_{max} (1 - e^{-k_{sat}t}) + c$, as described by Mayer *et al.* The point (0,0) was not used in the fit.





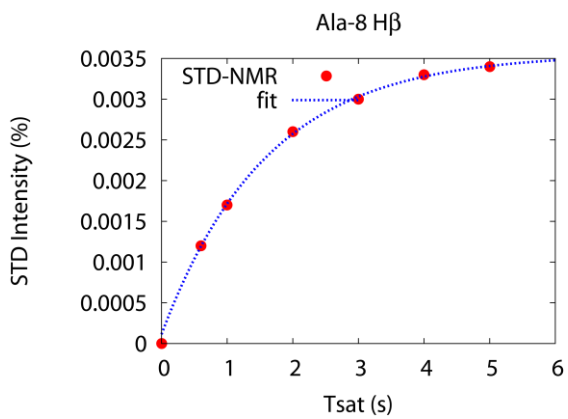
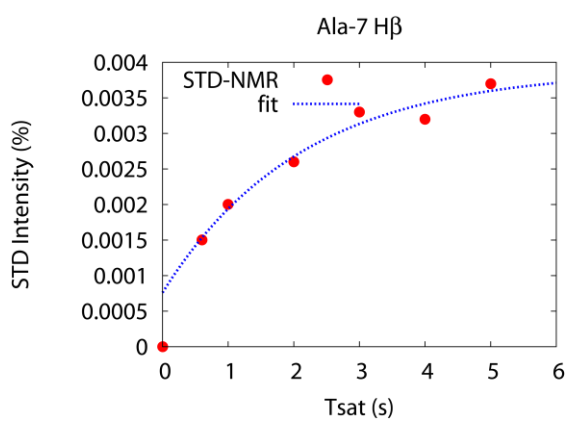
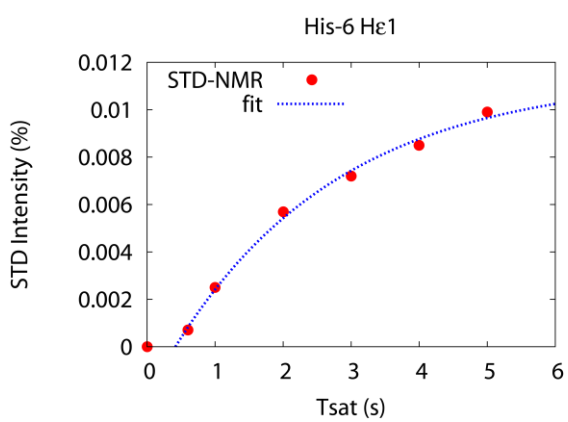
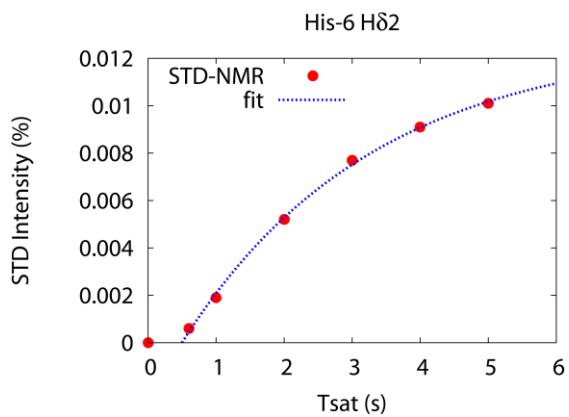
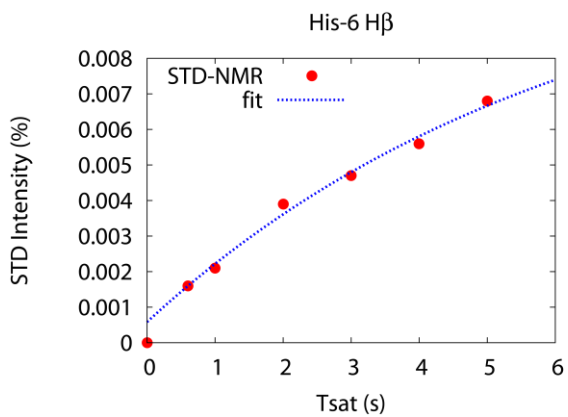


Figure S5.7. a) Number of hydrogen bonds existing between MDWNMHAA and SYA/J6 Fab as a function of simulation time, with MDWNMHAA acting as a hydrogen bond donor (red) and as a hydrogen bond acceptor (blue). The number of hydrogen bonds on the vertical axis refers to a cumulative average over the MD trajectory. b) RMSD of SYA/J6 Fab (black) MDWNMHAA (red) as a function of simulation time.

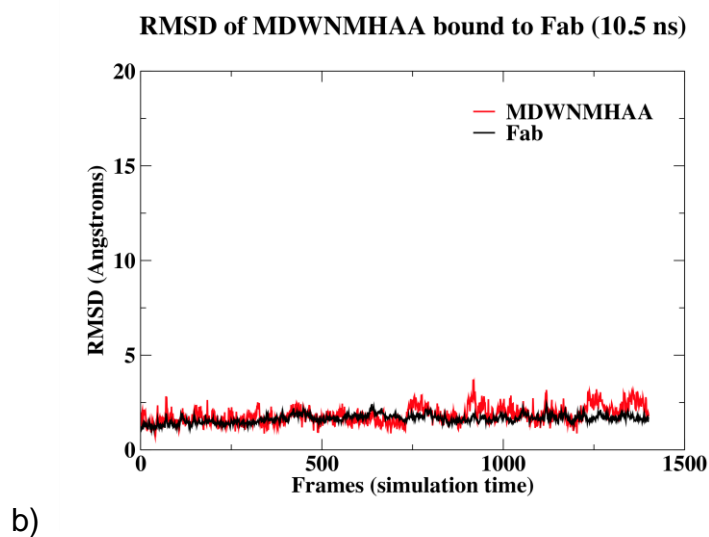
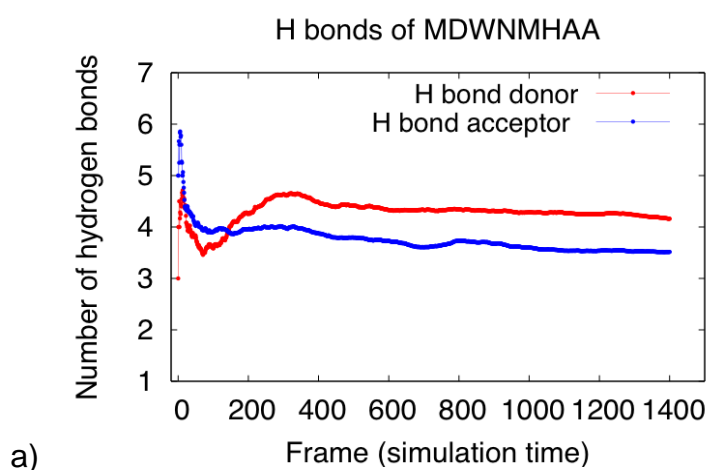


Figure S5.8. a) Number of hydrogen bonds existing between MDWNMHAA and SYA/J6 Fv as a function of simulation time, with MDWNMHAA acting as a hydrogen bond donor (red) and as a hydrogen bond acceptor (blue). The number of hydrogen bonds on the vertical axis refers to a cumulative average over the MD trajectory. b) RMSD of SYA/J6 Fv (black) MDWNMHAA (red) as a function of simulation time.

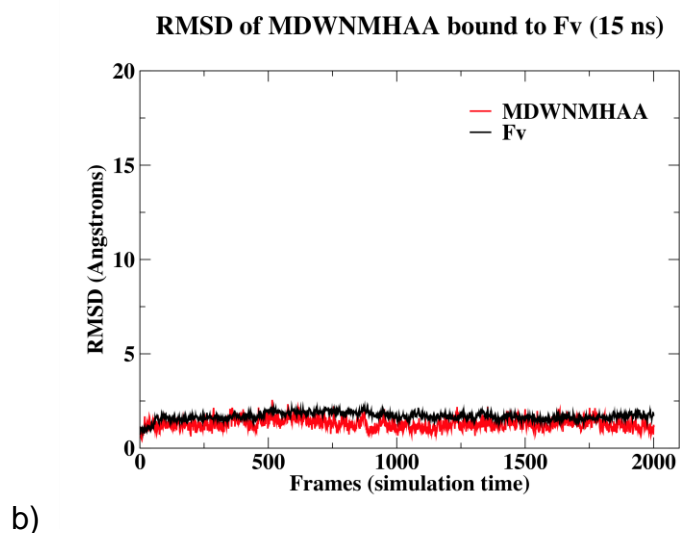
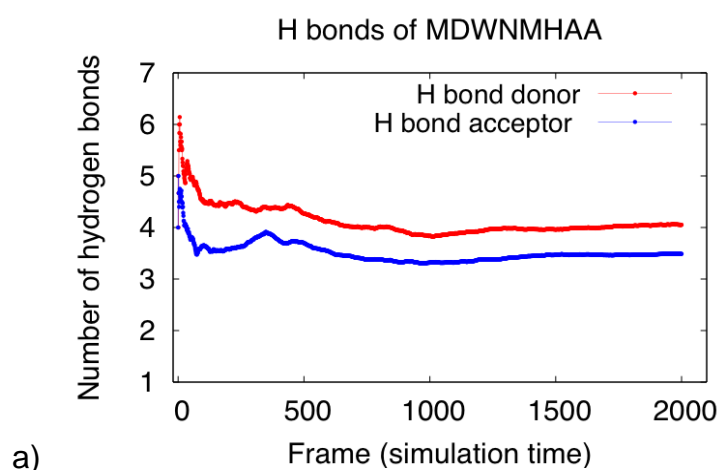


Figure S5.9. a) RMSD per residue for the light chain of Fab (black) bound to MDWNMHAA and the light chain of Fv (blue) bound to MDWNMHAA. b) RMSD per residue for the heavy chain of Fab (black) bound to MDWNMHAA and the heavy chain of Fv (red) bound to MDWNMHAA.

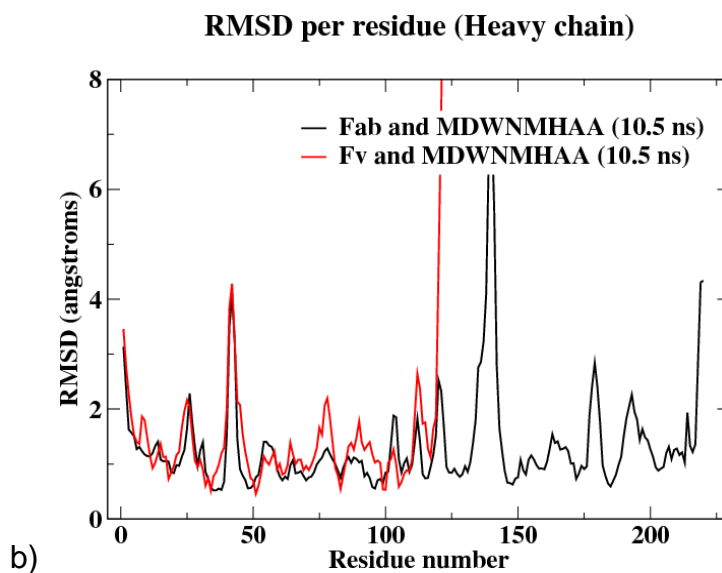
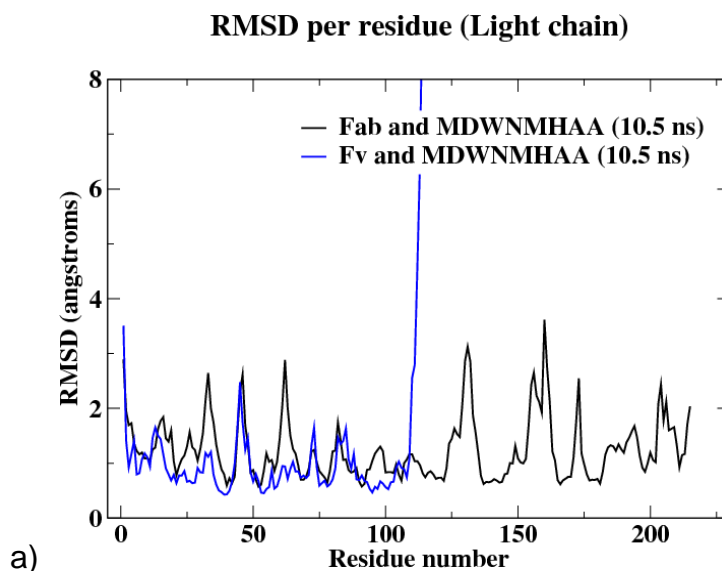


Figure S5.10. Ramachandran plots of MDWNMHAA bound to Fv during the course of the MD simulation. There are α -helical structural elements in the MHAA region upon binding. Each panel refers to phi and psi angles of a particular residue. Panels from a) to f) refer to aspartic acid, tryptophan, asparagine, methionine, histidine, and alanine, respectively.

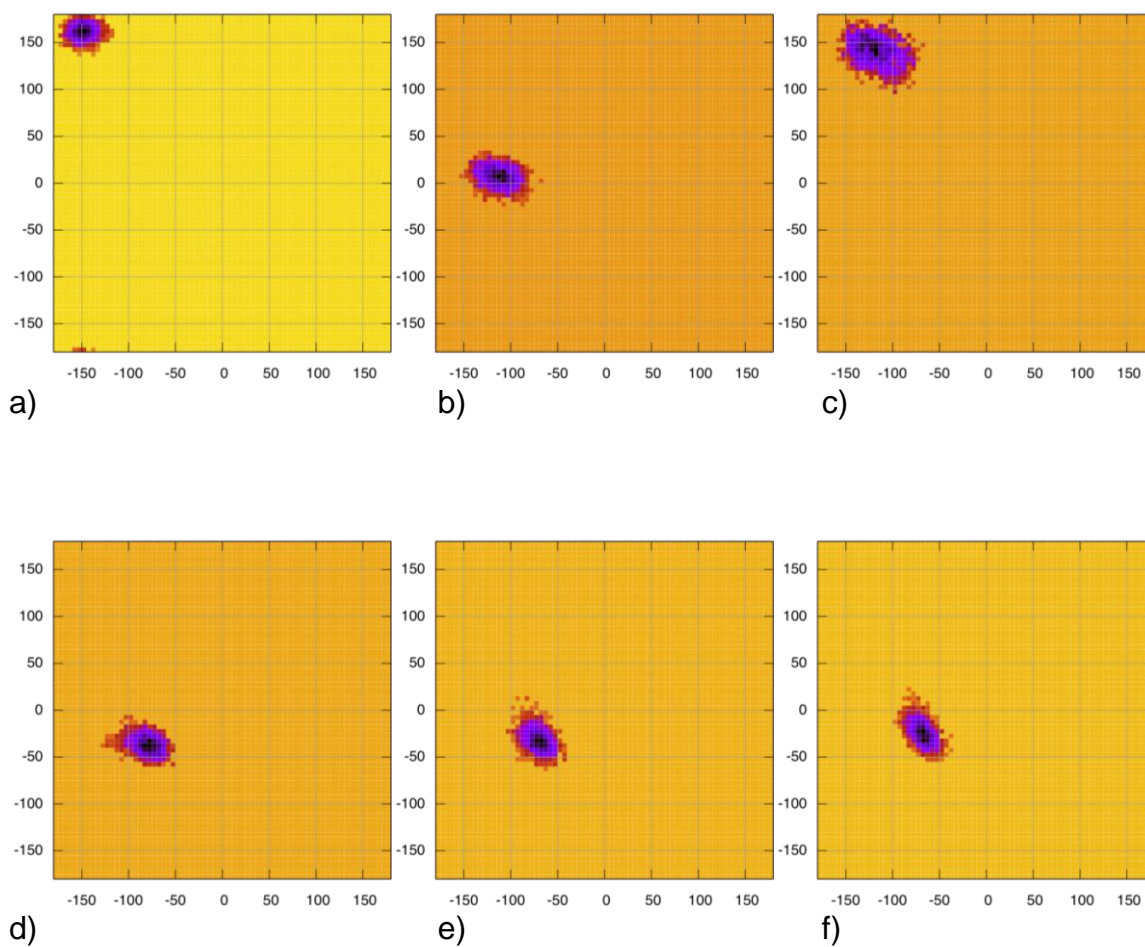
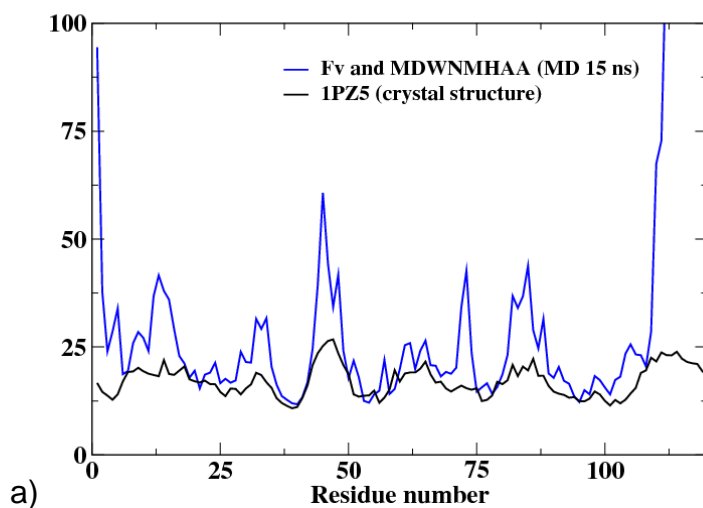


Figure S5.11. Comparison of B values from the MD simulation of MDWNMHAA bound to Fv with B values from the X-ray crystal structure. a) The light chain is shown in blue and b) the heavy chain is shown in red. B values from the X-ray crystal structure are shown in black.

B values Light Chain (MD vs crystal structure)



B values Heavy Chain (MD vs crystal structure)

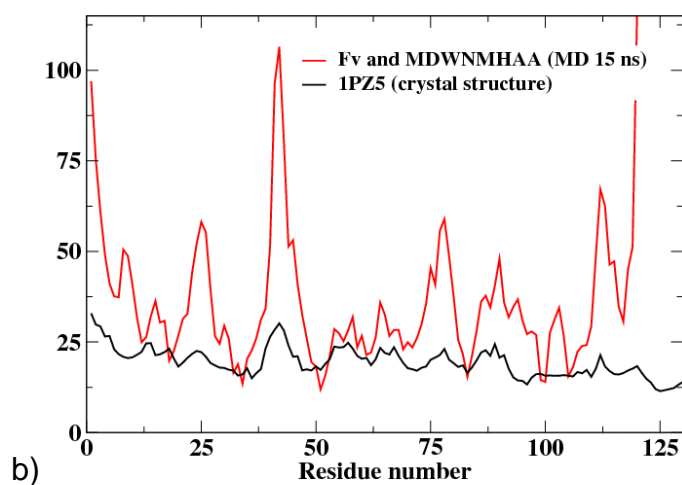


Figure S5.12. a) RMSD per residue for the light chain of Fv bound to MDWNMHAA (blue) and in the absence of MDWNMHAA (black). b) RMSD per residue for the heavy chain of Fv bound to MDWNMHAA (red) and in the absence of MDWNMHAA (black).

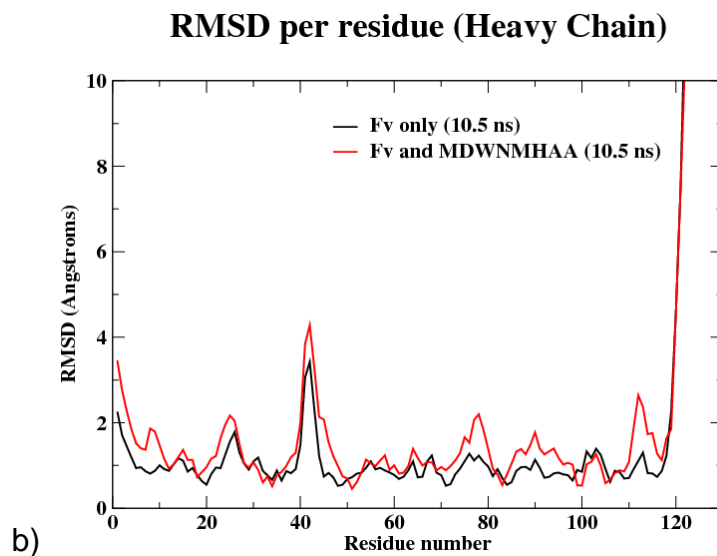
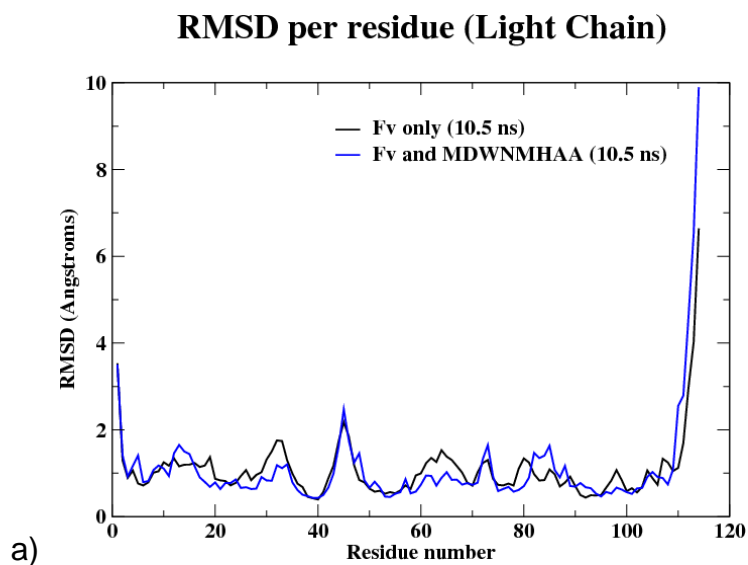
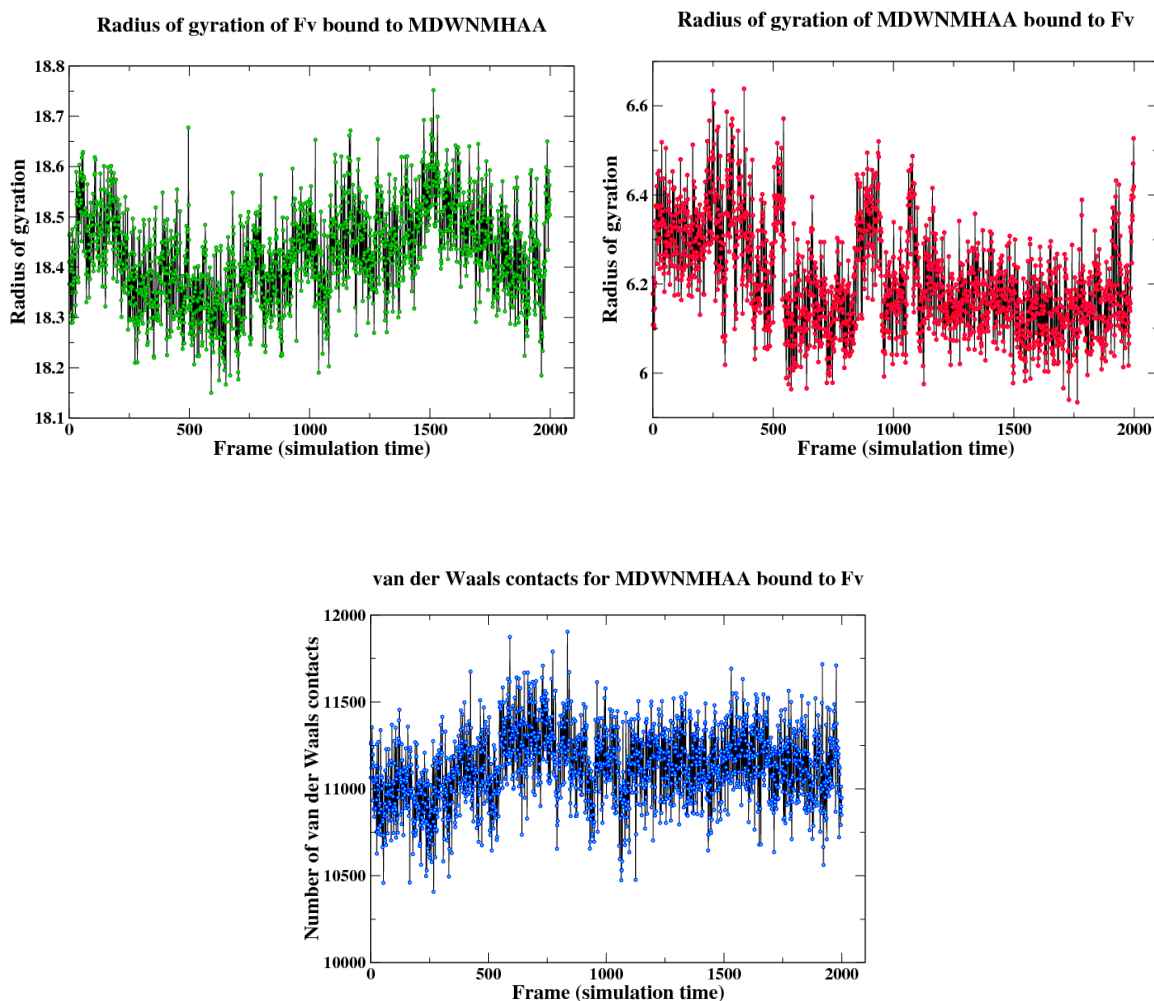


Figure S5.13. Radius of gyration and van der Waals contacts. Upper panels show the radius of gyration of MDWNMHAA and Fv, left and right, respectively, during the course of the MD simulation. Lower panel shows that van der Waals contacts are maintained throughout the simulation.



CHAPTER 6: Conclusions

6.1 Thesis summary

The research discussed in this thesis has focused on nuclear magnetic resonance (NMR) spectroscopy and its application to understanding complex biological problems.

The mycolyl–arabinogalactan–peptidoglycan complex that coats the surface of *Mycobacterium tuberculosis* is vital to the organism’s survival. In addition to the long fatty acids that contribute to the low permeability of the complex, there are approximately 30 galactofuranosyl (Gal f) residues attached via alternating β -(1→6) and β -(1→5) linkages synthesized by bifunctional galactofuranosyltransferases, GlfT1 and GlfT2. We have used Saturation Transfer Difference (STD) NMR spectroscopy to examine the active site architecture of GlfT2 using two trisaccharide acceptor substrates, β -D-Gal f -(1→6)- β -D-Gal f -(1→5)- β -D-Gal f -O(CH $_2$) $_7$ CH $_3$ and β -D-Gal f -(1→5)- β -D-Gal f -(1→6)- β -D-Gal f -O(CH $_2$) $_7$ CH $_3$. The STD NMR epitope maps demonstrated a greater enhancement toward the “reducing” ends of both trisaccharides, and that UDP-galactofuranose (UDP-Gal f) made more intimate contacts through its nucleotide moiety. This observation is consistent with the greater flexibility required within the active site of the reaction between the growing polymer acceptor and the UDP-Gal f donor. Competition STD NMR titration experiments with the trisaccharide acceptor substrates demonstrated that they bind

competitively at the same site, suggesting that GlfT2 has one active site pocket capable of catalyzing both β -(1 \rightarrow 5) and β -(1 \rightarrow 6)-galactofuranosyl transfer reactions. The addition of UDP-Galf to either trisaccharide acceptor substrates in the presence of GlfT2 generated a tetrasaccharide product, indicating that the enzyme was catalytically active under the conditions at which the STD-NMR experiments were carried out. Thus, the work presented in this thesis has contributed to the current model of the mechanism of GlfT2, and could facilitate the design and synthesis of an inhibitor against the enzyme.

STD NMR spectroscopy was also used to probe the bioactive conformations of the carbohydrate mimic MDWNMHAA of the O-polysaccharide of the *Shigella flexneri* Y bacterium when bound to its complementary antibody, mAb SYA/J6. The dynamic ligand epitope, was mapped with the CORCEMA-ST (COmplete Relaxation and Conformational Exchange Matrix Analysis of Saturation Transfer) program that calculates STD-NMR intensities. Comparison of these predicted STD enhancements with experimental data was used to select a representative binding mode. The bound conformation was further refined with a simulated annealing refinement protocol known as STD-NMR Intensity-restrained CORCEMA Optimization (SICO) to give a more accurate representation of the bound peptide epitope. Thus, the work described in this thesis has increased our understanding of 1) the bound conformation of the mimetic peptide and 2) the molecular interactions of the peptide with an antibody raised against shigellosis-causing *Shigella flexneri* Y. This information could contribute to the design of vaccine against shigellosis.

X-ray crystallographic data of MDWNMHAA when bound to mAb SYA/J6 indicate the immobilization of water molecules, i.e. the presence of “bound” water molecules, in the combining site. Water Ligand Observed via Gradient Spectroscopy (WaterLOGSY) was used in conjunction with STD NMR spectroscopy to provide insight into the presence of water molecules that exist at the interstitial sites between the peptide and the antibody. Molecular dynamics calculations have also provided a more accurate picture of the possibilities for bound-ligand conformations, and water molecules involved in providing complementarity between the peptide and SYA-J6. Thus, based on the research carried out in this thesis, a strategy that involves displacing these water molecules can now be used to design the next-generation, higher-affinity binders for use as vaccine candidates.

6.2 Appendix - Computational Design of Peptide Mutants

Under the excellent guidance of Dr. Johannes “Hannes” Müllegger (prior to his passing away) and Mr. Dustin Bleile, of Zymeworks Inc., computational approaches were used to identify twenty MDWNMHAA octapeptide mutants that we proposed to have improved binding affinity to the *Shigella flexneri* Y SYA/J6 antibody over the MDWNMHAA peptide. Our goal was the rational design of a peptide that could produce a stronger immune response to the carbohydrate-binding antibody. We hypothesized that this could be achieved by tightening the interactions of the peptide variants to the anti-carbohydrate antibody in a manner that preserves carbohydrate mimicry. Our aim was to replace the bound water molecules, which create large entropic binding penalties (see Chapter 5) by the use of groups designed to preserve carbohydrate contacts. The criteria that we used to examine the peptide variants included: i) the relative electrostatic contribution to the binding free energy^[1], $\Delta\Delta G_{\text{electrostatic}}$, ii) the number of hydrogen bonds with the antibody, iii) the van der Waals contacts with the antibody, and iv) the number of displaced water molecules in the binding interface, which is indicative of the entropic costs of binding.

The key feature in the design and evaluation of the peptide mutants is to structurally mimic the carbohydrate binding mode to the antibody while optimizing the energetics of binding. This can be seen in Figure 6.1, where MDWNMHAA has interactions missing to the antibody’s heavy chain Ala-106, Met-107, Glu-35 and Glu-50 residues. These interactions to the antibody are present in the case

of the synthetic pentasaccharide antigen. Thus, the peptide has harnessed water molecules to provide these contacts to the antibody (Figure 6.1). Hence, a mutant that is capable of displacing water molecules and preserving these contacts would be of interest to us. We found that the incorporation of unnatural amino acids i.e. D-amino acids, into the amino acid sequence displaced those bound water molecules and recaptured the missing interactions to the antibody's Ala-106, Met-107, Glu-35 and Glu-50 residues (Figure 6.2).

To begin this research project, amino acids were interchanged at each position into the wild type MDWNMHAA structure to generate single mutants. These single mutants were subjected to protein side-chain packing. The side-chain packing protocol involves the discretization of dihedral angles of each amino acid side chain. Dihedral angles of each amino acid residue type are defined according to the Dunbrack library ^[2]. In the *Dead-End Elimination* method ^[3], rotamers are selected from the Dunbrack library that lower the system energy and, in effect, the amino acid side chain geometry is predicted (referred to as *side chain packing*). As part of the optimization routine, if an amino acid side chain was able to provide a lower system energy by occupying a place previously occupied by an interstitial bound water molecule, then that water molecule was removed (Figure 6.2). With these single mutants in hand, the structures were minimized.

Four key scoring parameters were employed to evaluate potential peptide mutants. First, the solvent electrostatic contribution to the binding free energy ($\Delta\Delta G_{\text{electrostatic}}$) ^[1] of the mutants was examined. It is known that free energy is

directly correlated with binding affinity and the relative free energy was computed relative to the wild type MDWNMHAA peptide. This relative free energy change, $\Delta\Delta G_{\text{electrostatic}}$, is based primarily on the APBS (Adaptive Poisson-Boltzmann Solver) program ^[1], which performs Poisson-Boltzmann electrostatics calculations and utilizes an implicit solvent model, to describe electrostatic interactions that exist between molecules in a solvated environment. The Poisson-Boltzmann equation is shown in Figure 6.3 ^[1]. We used APBS to evaluate the solvent electrostatic contribution of the mutants to the binding free energy. Second, we examined hydrogen bond formation. The increase in hydrogen bonds relative to MDWNMHAA indicates increased binding interactions to the antibody. This increase in hydrogen bonds is also implicated in the propensity of the peptides to form stable secondary structures that are important for structural recognition. We have hypothesized that the α -helical component constitutes an immunodominant feature of the peptide ^[4]. Third, we looked for the displacement of bound water molecules. In the binding of the wild type peptide, three water molecules are trapped at the interface with the antibody. These bound water molecules have an associated entropic penalty that can be avoided if the water-mediated interactions are replaced with direct peptide-antibody contacts (Figure 6.2). Finally, we looked at van der Waals interactions. Aside from mimicking the existing carbohydrate-antibody interactions, any additional van der Waals contacts that formed with the mutant peptides, which could further stabilize the bound complex, were identified.

Single amino acid mutants that ranked favourably were used as a stepping-stone for access to double mutants and also to triple mutants. At positions 1, 2 and 4 through to 8, the top single mutants were selected and the amino acid residues of those mutants were used to generate double mutant and triple mutant combinations. We chose to retain tryptophan at position 3 since the research conducted in this thesis has shown this residue makes significant contacts to the antibody. This was followed by another round of side-chain packing, minimizing and scoring to generate potential double and triple mutants.

The mutants were synthesized and tested for binding in experiments using surface plasmon resonance (SPR) at the National Research Council in Montreal but unfortunately, only two of the mutants bound to the antibody SYA-J6. The amino acid sequence of the mutants and their binding data is shown in Table 6.1. There could be several reasons for the lack of binding. First, the side chain packing approach relies on a rigid peptide backbone and does not allow the peptide to respond to a change in amino acid. This issue can be addressed by the use of molecular dynamics which could allow the peptide mutants to find the optimum backbone geometry. Second, the mutants were scored on the basis of an electrostatics-based contribution to the total binding energy. Entropic contributions to the total binding energy were not evaluated. The only entropic contribution considered was that of the bound water molecules. These were estimated at 2 kcal/mol^[5]. Thus if a residue displaced a single water molecule, a total of 2 kcal/mol was subtracted from the calculated binding energy. Third, it is

also important to consider the accuracy of the force field. Thus, a difference in 1 kcal/mol corresponds to an approximate 5 fold difference in binding affinity.

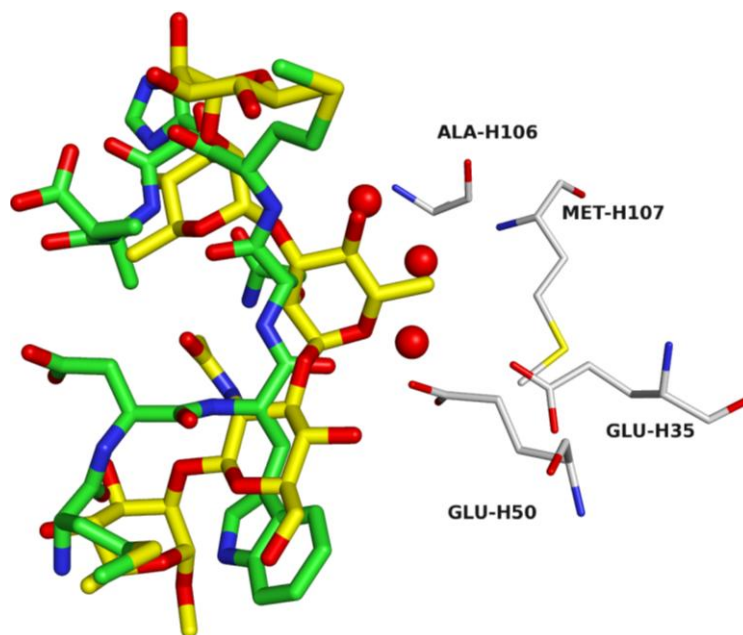


Figure 6.1. Bound water molecules (red spheres) are harnessed by the wild type peptide (green) to provide contacts to antibody's heavy chain Glu-35, Glu-50, Ala-106 and Met-107 residues. These contacts to the antibody are made by the pentasaccharide hapten α -L-Rhap-(1 \rightarrow 2)- α -L-Rhap-(1 \rightarrow 3)- α -L-Rhap-(1 \rightarrow 3)- β -D-GlcpNAc-(1 \rightarrow 2)- α -L-Rhap-(1 \rightarrow OMe).

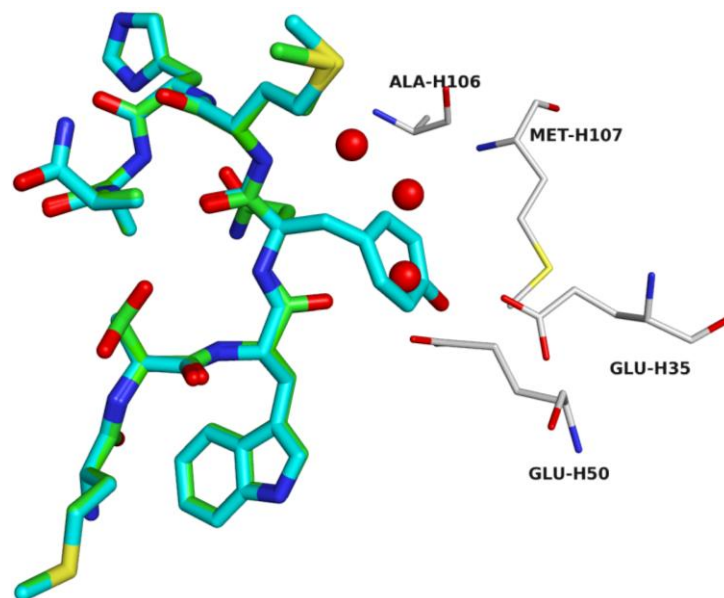


Figure 6.2. Double mutant, MAW(d-TYR)MHAA, makes contacts to the antibody's heavy chain Glu-35, Glu-50, Ala-106 and Met-107 residues and displaces the three water molecules harnessed by the wild type peptide (green) to provide shape complementarity to the antibody.

Poisson-Boltzmann equation

$$-\nabla \cdot \epsilon(\mathbf{x}) \nabla \phi(\mathbf{x}) = \rho_f(\mathbf{x}) + \sum_m q_m c_m e^{-\beta[q_m \phi(\mathbf{x}) + V_m(\mathbf{x})]}$$

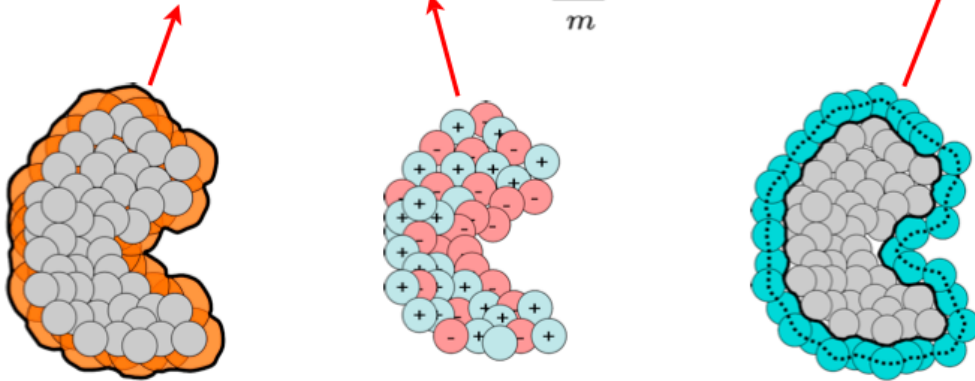


Figure 6.3. Poisson-Boltzmann equation. (Reproduced with permission from Nathan Baker, Computational electrostatics for biomolecular systems. Workshop Lecture for *Collaborative Computational Project for Biomolecular Simulation (CCPB)*. Copyright © Nathan Baker, Washington University in St. Louis, Dept. of Biochemistry and Molecular Biophysics, Center for Computational Biology, baker@ccb.wustl.edu. All rights reserved).

Table 6.1. IC50 data for peptide mutants as well as computed $\Delta\Delta G_{\text{electrostatic}}$, computed Lennard-Jones potential, computed number of hydrogen bonds between peptide and antibody and number of water molecules displaced.

Sequence	IC50 (μM)	$\Delta\Delta G_{\text{electrostatic}}$ (kcal/mol)	LJ (kcal/mol)	# HB	#waters displaced
MDWNMHAA	6.5	0	-50	4	0
M(E)WNMHA(P)	-	-5.63	-54.4	7	0
M(E)WNMH(N)A	-	-4.92	-48.7	7	0
M(E)WNMH(I)A	-	-4.7	-50.5	6	0
M(E)WNM(P)AA	-	-4.65	-43.5	4	0
M(T)WNMH(N)A	-	-4.3	-45.9	5	0
M(E)WNMHA(G)	>87	-4.1	-56.5	7	0
M(M)WNMH(N)A	-	-4.04	-57.2	6	0
M(E)W-(dC)-MHA(P)	-	-2.79	-48.1	4	1
M(E)WNMHAA	7.6	-2.37	-50.6	6	0
M(E)W-(dS)-MHA(P)	-	-2.04	-47.8	4	1
M(E)W-(dS)-MH(N)A	-	-1.58	-36.3	6	0
M(E)W-(dQ)-MHAA	-	-1.15	-47.8	6	2
M(E)W-(dM)-MH(N)A	-	-0.95	-43.7	6	2
M(E)W-(dN)-MH(N)A	-	-0.71	-38.7	6	0
(P)(E)W-(dN)-MHAA	-	-0.31	-53.3	6	0
M(E)W-(dS)-MHA(E)	-	3.34	-51.7	7	0
MDW-(dY)-MHA(P)	-	7.85	-46.3	4	3
M(A)W-(dY)-MHAA	-	7.97	-45.8	5	3
MDW-(dY)-MHA(E)	-	8.8	-42.2	6	3

6.3 References

- [1] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, 98, 10037-10041.
- [2] R. L. Dunbrack, *Curr. Opin. Struct. Biol.* **2002**, 12, 431-440.
- [3] J. Desmet, M. Demaeyer, B. Hazes, I. Lasters, *Nature* **1992**, 356, 539-542.
- [4] M. A. Johnson, B. M. Pinto, in *Bioactive Conformation II*, Vol. 273, **2008**, pp. 55-116.
- [5] J. D. Dunitz, *Science* **1994**, 264, 670.