



University of HUDDERSFIELD

University of Huddersfield Repository

Swift, Simone

Molecular Modelling of the Complex Polysaccharide Heparan Sulphate

Original Citation

Swift, Simone (2011) Molecular Modelling of the Complex Polysaccharide Heparan Sulphate. Doctoral thesis, University of Huddersfield.

This version is available at <http://eprints.hud.ac.uk/id/eprint/10739/>

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

<http://eprints.hud.ac.uk/>

Molecular Modelling of the Complex Polysaccharide Heparan Sulphate

Simone Karoline Swift

**A thesis submitted to The University of Huddersfield in partial fulfilment of
the requirements for the degree of Doctor of Philosophy.**



University of
HUDDERSFIELD

**Department of Chemical and Biological Sciences
School of Applied Sciences**

March 2011

COPYRIGHT STATEMENT

- i. The author of this thesis (including any appendices and/or schedules to this thesis) owns any copyright in it (the "Copyright") and s/he has given The University of Huddersfield the right to use such Copyright for any administrative, promotional, educational and/or teaching purposes.
- ii. Copies of this thesis, either in full or in extracts, may be made only in accordance with the regulations of the University Library. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.
- iii. The ownership of any patents, designs, trademarks and any and all other intellectual property rights except for the Copyright (the "Intellectual Property Rights") and any reproductions of copyright works, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.

ABSTRACT

Heparan Sulphate plays an important role in many life processes and so an understanding of its role as a universal co-receptor is of great importance. Traditionally oligosaccharides derived from the related molecule heparin have been at the forefront of molecular drug design, due to its similarity in structure and function. To obtain a more complex and detailed picture of the role of HS in structural biology further complex work must be undertaken on HS. Here in this study a number of HS derived octasaccharides have been purified. Alongside this is work carried on HS derived decasaccharides, of which all have undergone extensive molecular modelling simulations. The role of the Iduronates in the HS structure, at monosaccharide level has indicated a major role for these structures in biological activity. Further work has indicated the level of sulphation is also a requirement which in turn influences conformational behaviour. Up until now this has only really been studied at the monosaccharide level and so this study has generated a number of different HS models which can confirm the importance of iduronate conformation in biological activity. Not only that but it has also been identified that certain torsional geometries within the glycosidic bonds between monosaccharides also has a major influence on conformation. Local deviations in the molecular modelling data suggest there is a slight difference between active and inactive oligosaccharides with the ability to bind and activate the HS:FGF2:FGFR IIIc complex.

The implications of these and other structural insights are discussed with the implication heavily towards a combination of both torsional geometry and iduronate conformation in biological activity. To further aid our understanding HS dp10 oligosaccharides were docked into a FGF2 which was in a complex with a heparin hexasaccharide. Docking experiments were carried in order to attain structural information on the binding of these molecules. The Goodger oligosaccharides were docked into the X-ray crystal structure and were chosen specifically for their iduronate conformation or the torsional geometry.

ACKNOWLEDGEMENTS

First and foremost I would like to thank my supervisor, Dr David Pye for all his help, guidance and encouragement, not just over the last four years but also during my undergraduate and MSc studies. Without his belief in me this project would never have happened. Secondly; I would like to thank Dr Kevin Murphy for his help during the first few months of starting this PhD and also for my MSc dissertation. He introduced me to the world of AMBER and really helped me along the way. Without the help of either of these two I would not be in the position that I am today. Alongside this I would like to thank Dr David Cooke, for all his help when everything modelling started to go wrong. We had some helpful discussions and he gave me some valuable pointers. I would like to thank all those who helped me get onto the Science Foundation course a decade ago, as this is really where my journey began. Those people know who they are and I am really grateful to them. I would like to thank The University of Huddersfield for their funding throughout this project. I would also like to thank Dr Neil McLay who expertly recorded all the NMR spectra reported.

I would like to dedicate this thesis to my parents who have been there for me right from day one. Without their support this would have been very difficult. I would also like to thank my other half Fred for all his help, support and encouragement. He has spent many nights encouraging, praising and pushing me along.

My final thanks go to three special people who were there at the beginning of this PhD and are still there now. We all started this journey together and hopefully we can all end it together. Dr Cara Sutton, Mrs Kate Clayton and Dr Hasina Motara thank you for all your help. Without it I would have made it this far.

LIST OF ABBRIVIATIONS

1BFC	File 1BFC from the Protein Data Bank (FGF2 complexed with a hexasaccharide heparin fragment)
1FQ9	File 1FQ9 from the Protein Data Bank (a 2:2:2 complex of ternary FGF2:FGFR1:heparin complex)
2OSTs	Uronosyl-2-O-sulphotransferase enzymes
3OSTs	Glucosaminyl 3-O-sulphotransferase enzymes
6OSTs	Glucosaminyl 6-O-sulphotransferase enzymes
aMan	Anhydromannose
α Me glycoside	Methyl glycoside of IdoUA(2S)
<i>A. crassispina</i>	<i>Anthocidaris crassispina</i> (sea urchin)
AchR	Acetylcholine receptors
AT III	Antithrombin III
BM	Basement membrane
Bnl	Branchless
COL	Collagenous domain
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>
CS	Chondroitin sulphate
<i>D. melanogaster</i>	<i>Drosophila melanogaster</i>
dp	Degrees of polymerisation
DS	Dermatan sulphate
ECM	Extracellular matrix
<i>E.coli</i>	<i>Escherichia coli</i>
EGF	Endothelial Growth Factor
ER	Endoplasmic reticulum
FS	Follistatin like
EXT	Exostoses
EXTL	Exostoses like
FGFs	Fibroblast Growth Factors
FGFRs	Fibroblast Growth Factor receptors (IIIc isoform unless stated otherwise)
FMDV – O ₁ BFS	A foot and mouth disease virus – subtype O ₁ BFS
GAG	Glycosaminoglycan
Gal	Galactose
GalNAc	N-acetyl galactosamine
GalT-I	Galactosamine transferase I
GalT-II	Galactosamine transferase II
GalT-III	Galactosamine transferase III
Glc	Glucosamine
GlcUA	Glucuronic acid
GlcNAc	N-acetyl glucosamine
GlcNS	Glucosamine N-sulphate
GlcNS(6S)	Glucosamine N-sulphate, 6-sulphate
GlcNS (6S, 3S)	Glucosamine N-sulphate, 6-sulphate, 3-sulphate
Gly	Glycine

GPC	Glypican
GPI	Glycosyl-phosphatidylinositol
HA	Hyaluronic acid
HBGF	Heparan sulphate binding Growth Factor
Hep	Heparin
HexUA	Hexuronic acid (either IdoUA or GlcUA).
HGF	Hepatocytes Growth Factor
hh	Hedgehog
HS	Heparan Sulphate
HSPG	Heparan Sulphate Proteoglycan
IdoUA	Iduronic acid
IdoUA(2S)	2-O-sulphated Iduronic acid
Ig	Immunoglobulin
IL-8	Interleukin 8
IUPAC	International Union of Pure and Applied Chemistry
KDa	Kilo Dalton
KS	Keratin sulphate
LMWH	Low molecular weight heparin
MD	Molecular dynamic
mU	micro units
NA-domain	N-acetylated domain
NC	Non-collagenous domain
N-CAM	Neural cell adhesion protein
NDSTs	N-deacetylase/N- sulphotransferase enzymes
NS/NA-domain	N-sulphated/N-acetylated alternating domain
OST	O-sulphotransferase
PAGE	Polyacrylamide gel electrophoresis
PAPS	3'-phosphoadenosine 5'-phosphosulphate
PBS	Phosphate-buffered
PF4	Platelet factor 4
Phi	Torsion angle/Puckering parameter angle
PG	Proteoglycan
Psi	Torsion angle
Rmsd	Root means squared deviation
SAX	Strong anion exchange
S-domain	Sulphated domain
SDS	Sodium dodecyl sulphate
SEA	Sperm, agrin and enterokinase homology
Ser	Serine residue
ST	Serine/threonine rich domains
Theta	Puckering parameter angle
Tyr	Tyrosine
UA/ Δ UA	Δ 4, 5 unsaturated Uronic Acid
UDP-sugar	Uridine diphosphate
Wg	Wingless
Xyl	Xylose
XylT	Xylose transferase enzyme

TABLE OF CONTENTS

	Page
CHAPTER 1: INTRODUCTION	1
1.1 Heparan Sulphate	2
1.2 Heparan Sulphate vs. Heparin: A Brief History	4
1.3 Proteoglycans.....	5
1.3.1 Heparan sulphate proteoglycans (HSPG's).....	5
1.3.1.1 Syndecan.....	6
1.3.1.2 Glypicans	9
1.3.1.3 Perlecan.....	14
1.3.1.4 Collagen XVIII/Endostatin.....	19
1.3.1.5 Agrin	22
1.4 GAG biosynthesis	26
1.4.1 Chain initiation	27
1.4.2 HS/Hep Chain Polymerization	29
1.4.3 HS/Hep Chain Modifications.....	31
1.4.3.1 2-O-Sulphotransferase	36
1.4.3.2 6-O-Sulphotransferase	38
1.4.3.3 3-O-Sulphotransferase	40
1.4.3.4 The Role of the Sulfs.....	44
1.4.4 CS/DS Chain Polymerization.....	45
1.4.4.1 CS/DS O - Sulphotransferase.....	48
1.5 De - Polymerisation Techniques	49
1.5.1 Chemical de - polymerisation	49
1.5.2 Enzymatic de - polymerisation	50
1.6 Domain Structure	51
1.6.1 Action of K5 Lyase	55
1.7 Structural conformation	57
1.7.1 Δ UA/UA	58
1.7.2 The N-glucosamines and The Gauche Effect.....	60
1.7.3 Uronic Acids.....	62

1.7.3.1 GlcUA	62
1.7.3.2 IdoUA.....	63
1.8 Interactions with proteins	67
1.8.1 HS Binding – Consensus sequence.....	67
1.8.2 FGF Biology.....	68
1.8.3 FGF Receptors – The FGFR's.....	68
1.8.3.1 FGF Isoforms and diversity.....	70
1.8.4 Signalling pathways of the FGF's	71
1.8.5 The Role of HS in FGF Signalling.....	72
1.8.6 HS/FGF/FGFR Complexes.....	73
1.8.6.1 Complex Ratios.....	75
1.8.6.2 The Symmetrical Complex	77
1.8.6.3 The Asymmetrical Complex.....	79
1.8.7 Interaction of HS with FGF1 and FGF2.....	81
1.8.7.1 FGF1	81
1.8.7.2 HS interacting groups with FGF1.....	84
1.8.7.3 FGF2	86
1.8.7.4 HS interacting groups with FGF2.....	87
1.8.8 The importance of specific groups on the HS chain and chain length – interactions for both FGF1 and FGF2	91
1.9 Modelling of HS dp10	99
1.9.1 Cremer-Pople puckering parameters.....	105
1.10 Molecular Modelling	105
1.10.1 AMBER.....	105
1.10.1.1 Torsion angles Φ and Ψ vs. Conformational Energy	108
1.10.1.2 The Exo and Endo - Anomeric Effect.....	110
1.10.1.3 Preparing the model in explicit solvent.....	112
1.10.1.4 The Use of a Periodic Boundary	114
 CHAPTER 2: EXPERIMENTAL METHODS AND PROCEDURES	 117
 2.1 Experimental Materials	 118
 2.2 Experimental Procedures	 118

2.2.1 Enzymatic Digestion of Heparan Sulphate	118
2.2.2 Gel Filtration Chromatography Analysis of Heparinase III Digestion	119
2.2.2.1 Analytical	119
2.2.2.2 Preparative	119
2.2.3 SAX-HPLC (Strong Anion Exchange Chromatography)	119
2.2.4 NMR Spectroscopy	120
2.2.5 Molecular Modelling	121
2.2.5.1 Calculation of the Lowest Energy Minima across Glycosidic Linkages	121
2.2.5.2 Preparation of the Model	121
2.2.5.3 Minimisation 1	122
2.2.5.4 Minimisation 2	122
2.2.5.5 Molecular Dynamics 1	123
2.2.5.6 Molecular Dynamics 2	123
2.2.6 Molecular Docking	124

CHAPTER 3: RESULTS127

3.1 Enzymatic digestion of heparan sulphate	128
3.1.2 Preparative gel filtration chromatography of HS digested to completion with heparinase III	133
3.1.3 SAX-HPLC of Octasaccharide (dp8) sized Oligosaccharides	135
3.1.4 PAGE analysis	138
3.1.5 NMR	140
3.1.5.1 Preliminary NMR analysis of dp8 samples	140
3.1.5.2 Partial Spectral Assignment of HS dp8(2)	143
3.2 Molecular Modelling	156
3.2.1 Calculation of potential energy surfaces	156
3.2.2 Molecular dynamic trajectories of the glycosidic Φ° and Ψ°	159
3.2.3 Puckering-Torsion Data	160
3.2.4 Decasaccharide models	168
3.2.4.1 The Reducing End Linkage – Linkage 9 for the eight Decasaccharide models dp10(1) - dp10(8)	170
3.2.4.2 Linkage 8	177
3.2.4.3 Linkage 7	182
3.2.4.4 Linkage 6	188

3.2.4.5 Linkage 5	193
3.2.4.6 Linkage 4	199
3.2.4.7 Linkage 3	203
3.2.4.8 Linkage 2	207
3.2.4.9 Linkage 1	211
3.3 Long 10 ns MD Simulations	216
3.3.1 The Reducing End Linkage – Linkage 9 for the four Decasaccharide models dp10(1), dp10(3), dp10(5) and dp10(7)	217
3.3.2 Linkage 8	220
3.3.3 Linkage 7	223
3.3.4 Linkage 6	227
3.3.5 Linkage 5	230
3.3.6 Linkage 4	234
3.3.7 Oligosaccharide sequences after 10 ns MD simulation	237
3.4 Docking Results of HS Decasaccharides	238
3.4.1 Preliminary Results of 1BFC and 1FQ9	238
3.4.1.1 1BFC	238
3.4.1.2 1FQ9	243
3.4.2 Docking of active and inactive dp10 structures	246
3.4.3 The 1 ns Docking Simulations	249
3.4.3.1 Oligosaccharide dp10(1)	249
3.4.3.2 Oligosaccharide dp10(3)	255
3.4.3.3 Oligosaccharide dp10(5)	259
3.4.3.4 Oligosaccharide dp10(7)	264
3.4.4 The 10 ns Docking Simulation	269
3.4.4.1 Oligosaccharide dp10(1)	269
3.4.4.2 Oligosaccharide dp10(3)	274
3.4.4.3 Oligosaccharide dp10(5)	276
3.4.4.4 Oligosaccharide dp10(7)	281
3.4.5 Flexible Dockings	285
3.4.5.1 The 1 ns Flexible Docking Analysis	285
3.4.5.2 The 10 ns Flexible Docking Analysis	289
3.4.6 Docking Implications	292

CHAPTER 4: DISCUSSION	294
4.1 The Digestion of HS by Heparinase III	295
4.2 Gel Filtration and SAX-HPLC chromatography of HS Octasaccharides	295
4.3 Analysis of HS dp8 NMR Profiles	296
4.4 The Requirement for Binding	296
4.4.1 The Minimum Length Required for Binding	297
4.4.2 The Effect of Temperature and Cations on the HS Chain	302
4.5 The Implications of Iduronate Conformation and Sulphation Pattern within HS chains	302
4.5.1 Linkage 7 – Conformations	308
4.5.2 Linkage 5 – Conformations	310
4.5.3 The Binding According to Literature	313
4.5.3.1 The Proposal for Iduronic Acid ¹ C ₄ : Protein Interactions	313
4.5.3.2 The Proposal for ² S ₀ :	317
4.5.3.3 The Importance of both ¹ C ₄ and ² S ₀ conformations: The FGF Interactions.....	318
4.5.3.4 The Importance of both ¹ C ₄ and ² S ₀ conformations: Other Protein Interactions	321
4.5.3.5 Biological Implications	324
4.7 The Contribution from the ⁴C₁ conformation	324
4.8 The Potential Role for Terminal ΔUA residues	327
4.9 The 3-Dimensional Structure of heparin and the Influence it has on the Binding Site	327
4.10 Molecular Dynamic Studies on Glycosidic Bonds	330
4.11 The Biological Implications of the Docking Experiments	333
CHAPTER 5: GENERAL CONCLUSION.....	336

CHAPTER 6: REFERENCES.....339

CHAPTER 7: APPENDIX353

TABLE OF FIGURES

	Page
Figure 1. 1 The extracellular domains of syndecan.	8
Figure 1. 2 GPI anchor of a glypican core protein attached to the plasma membrane.	11
Figure 1. 3 GAG attachment to the core protein glypican.	13
Figure 1. 4 The domain structure of Perlecan.	15
Figure 1. 5 The release of the endostatin peptide from collagen XVIII.	21
Figure 1. 6 The domain structure of Agrin.	24
Figure 1. 7 Initial GAG biosynthesis for all GAG structure – The GAG-Protein linker region.	28
Figure 1. 8 Basic mechanism for the C-5 epimerization reaction	34
Figure 1. 9 Composition of S-domains with and without 2-O-sulphotransferase.	37
Figure 1. 10 GAG biosynthesis of HS/hep chains.	42
Figure 1. 11 Biosynthesis of Chondroitin/Dermatan sulphate chains.	47
Figure 1. 12 HS/Heparin domain structure.	54
Figure 1. 13 A refined model of the domain structure of HS after treatment with KS lyase enzyme.	56
Figure 1. 14 Structural conformation of the terminal Uronic acid.	59
Figure 1. 15 The ω -angle identified in all N-glucosamines and the possible gauche orientations.	61
Figure 1. 16 Structural conformation of the Iduronates.	64
Figure 1. 17 Interactions of HS with FGF and FGFR.	76
Figure 1. 18 The X-ray crystal structure of a symmetrical complex of a 2:2:2 FGF2:FGFR1:heparin ternary complex.	78
Figure 1. 19 The FGF1:FGFR2:heparin complex.	80

Figure 1. 20 The specific binding interaction between FGF1 and a heparin hexasaccharide.....	83
Figure 1. 21 FGF2 binding site with a heparin hexasaccharide fragment.....	90
Figure 1.22 An example of the difference between active and inactive sequences due to the position of a single IdoUA(2S) residue.....	92
Figure 1.23 An example of a Heparin sequence with both high and low affinity binding sites	95
Figure 1.24 HS sequences identified with binding properties	98
Figure 1.25 HS sequences active and inactive in BaF-32 cells transfected with FGFr1 IIIc	101
Figure 1.26 HS dp10 sequences	102
Figure 1.27 Information flow in the AMBER suite of programmes	107
Figure 1.28 Definition of the phi and psi torsion angles	109
Figure 1.29 The Exo-anomeric effect	111
Figure 1.30 A simple water model showing the TIP3P water model	113
Figure 1.31 An example of a periodic boundary	115
Figure 3.1 Analytical TSK3000PW gel filtration profile of HS after 48 h incubation with 10mU Hep III	129
Figure 3.2 Comparison of large sample digest carried out on an analytical TSK3000PW gel filtration profile of HS after 120 h incubation with 10mU Hep III	130
Figure 3.3 Comparison of the analytical gel-filtration profiles of HS digested with 140 mU (total) of heparinase III over a 7 day period, re-dissolved in 30 ml hep buffer	132
Figure 3.4 Preparative Bio-Gel P10 gel filtration chromatography profiles of HS digested to completion with heparinase III	134
Figure 3.5 SAX-HPLC chromatography of sized HS oligosaccharides	136

Figure 3.6 An expanded SAX-HPLC chromatography of sized HS oligosaccharides	137
Figure 3.7 An example of two different HS hexasaccharides oligosaccharides identified using PAGE analysis.....	139
Figure 3.8 One-dimensional 500 MHz ¹ H-NMR spectra of HS Octasaccharide dp8(1) and dp8(3) oligosaccharides.....	141
Figure 3.9 Assignment of the non-reducing terminal uronic acid, residue “a” of oligosaccharide HS dp8(2)	144
Figure 3.10 Assignment of glucosamine, residue “b” of oligosaccharide HS dp8(2).....	146
Figure 3.11 Spin System B	148
Figure 3.12 Spin System C	150
Figure 3.13 Spin System D	152
Figure 3.14 A summary of the partial assignment of the ¹ H chemical shifts for the oligosaccharide HS dp8(2)	154
Figure 3.15 Cremer-Pople puckering parameters for the HS oligosaccharide iduronate residues alongside the ΔUA residue on the reducing terminal during a 1000 ps molecular dynamic simulation.....	164
Figure 3.16 Cremer-Pople puckering parameters for the internal HS oligosaccharide Iduronate-2-O sulphate residues during a 1000 ps molecular dynamic simulation.....	166
Figure 3.17 Puckering-torsion data for linkage 9.....	174
Figure 3.18 Puckering-torsion data for linkage 8.....	179
Figure 3.19 Puckering-torsion data for linkage 7.....	185
Figure 3.20 Puckering-torsion data for linkage 6.....	190

Figure 3.21 Puckering-torsion data for linkage 5.....	196
Figure 3.22 Puckering-torsion data for linkage 4.....	200
Figure 3.23 Puckering-torsion data for linkage 3.....	204
Figure 3.24 Puckering-torsion data for linkage 2.....	208
Figure 3.25 Puckering-torsion data for linkage 1.....	213
Figure 3.26 Puckering-torsion data for linkage 9 of the 10 ns MD simulation	218
Figure 3.27 Puckering-torsion data for linkage 8 of the 10 ns MD simulation	221
Figure 3.28 Puckering-torsion data for linkage 7 of the 10 ns MD simulation.....	225
Figure 3.29 Puckering-torsion data for linkage 6 of the 10 ns MD simulation.....	228
Figure 3.30 Puckering-torsion data for linkage 5 of the 10 ns MD simulation.....	232
Figure 3.31 Puckering-torsion data for linkage 4 of the 10 ns MD simulation	235
Figure 3.32 A comparison of the X-ray crystal structure of a bFGF (FGF2) complexed with a hexamer heparin fragment alongside the Vina predicted model.....	240
Figure 3.323The important binding residues identified between FGF2 and a heparin hexasaccharide.....	242
Figure 3.34 A comparison of the X-ray crystal structure of a 2:2:2 complex of ternary bFGF-FGFR1-heparin complex alongside the Vina predicted model.....	244
Figure 3.35 The important binding site residues identified in a 2:2:2 complex of ternary bFGF-FGFR1-heparin complex.....	245
Figure 3.36 The highest affinity binding structure as predicted by Vina for dp10(1), frame 45 from the rigid 1 ns MD simulation	252

Figure 3.37 The highest affinity binding structure as predicted by Vina for dp10(3), frame 1300 from the rigid 1 ns MD simulation.	257
Figure 3.38 The highest affinity binding structure as predicted by Vina for dp10(5), frame 1381, from the rigid 1 ns MD simulation	261
Figure 3.39 The highest affinity binding structure as predicted by Vina for dp10(7) frame 200, from the rigid 1 ns MD Simulation.....	266
Figure 3.40 The highest affinity binding structure as predicted by Vina for dp10(1) frame 8300, with the best fit for the X-ray crystal structure for the rigid 10 ns MD simulation.....	271
Figure 3.41 The highest affinity binding structure as predicted by Vina for dp10(5) frame 6000 with the best fit for the X-ray crystal structure for the rigid 10 ns MD simulation	279
Figure 3.42 The highest affinity binding structure as predicted by Vina for dp10(7) frame 250, with the best fit for the X-ray crystal structure for the rigid 10 ns MD simulation.....	282
Figure 3.43 The highest affinity binding structure as predicted by Vina for dp10(1) frame 45 and dp10(5) frame 1381, with the best fit for the X-ray crystal structure for the flexible 1 ns MD simulation	288
Figure 3.44 The highest affinity binding structure as predicted by Vina for dp10(1) frame 200, with the best fit for the X-ray crystal structure for the flexible 10 ns MD simulation.....	290
Figure 4.1 The minimum identified HS oligosaccharide which have been identified to bind to FGF2	299
Figure 4.2 Two HS oligosaccharides which have been identified to interact with FGF1 and FGF2 respectfully	301

Figure 4.3 The conformation of the iduronates before undergoing Cremer-Pople analysis	305
Figure 4.4 The conformation of the iduronates at linkage 7 after the MD simulation	309
Figure 4.5 The conformation of the iduronates at positions 5 and 7 in the HS chain	312
Figure 4.6 The monosaccharide, disaccharide and pentasaccharide in which the 1C_4 conformation was thought to dominate	316
Figure 4.7 Two HS/heparin oligosaccharides which indicate the requirement for both the 1C_4 and 2S_0 conformations present for binding	320
Figure 4.8 The two heparin oligosaccharides; which contain the rare 3-O-sulphate modification and have been identified in both the 1C_4 and 2S_0 conformations.....	323
Figure 4.9 Conformational data for iduronate internal structures, identified in DS	326
Figure 4.10 Three dimensional models of two heparin dodecasaccharides.....	329

LIST OF TABLES

	Page
Table 1.1 Disaccharide content of five different GAG structures.....	2
Table 1.2 High and low affinity binding sites in FGF2.....	96
Table 3.1 Minimum selected Φ° and Ψ° dihedral angles across each glycosidic linkage of HS disaccharides.....	157
Table 3.2 The known binding site identified in the known X-ray crystal structure	247
Table 3.3 Frame numbers chosen for docking analysis with Vina.....	248
Table 3.4 Conformations of all the ring structures for dp10(1) in the frames chosen for docking using Vina	250
Table 3.5 The conformation of the different iduronates in dp10(1) for frame 975	253
Table 3.6 Energy values for each Vina predicted model for dp10(1), frame 45.....	254
Table 3.7 Energy values for each Vina predicted model for dp10(3), frame 1300.....	256
Table 3.8 Conformations of all the ring structures for dp10(3) in the frames chosen for docking using Vina	258
Table 3.9 The Vina predicted models for dp10(5) for frame 1381	260
Table 3.10 Conformations of all the ring structures for dp10(5) in the frames chosen for docking using Vina	262
Table 3.11 Energy values for each Vina predicted model for dp10(5), frame 1381.....	263
Table 3.12 The best fit Vina predicted model for dp10(7) for frame 200	265
Table 3.13 Conformations of all the ring structures for dp10(7) in the frames chosen for docking using Vina	267

Table 3.14 Energy values for each Vina predicted model for dp10(7), frame 250.....	268
Table 3.15 Conformations of the iduronic acid ring structures for dp10(1) for the 10 ns MD simulation.....	272
Table 3.16 Energy values for each Vina predicted model for dp10(1), frame 8300	273
Table 3.17 Conformations of the iduronic acid ring structures for dp10(3) for the 10 ns MD simulation.....	275
Table 3.18 Energy values for each Vina predicted model for dp10(5), frame 6000	278
Table 3.19 Conformations of the iduronic acid ring structures for dp10(5) for the 10 ns MD simulation.....	280
Table 3.20 Conformations of the Iduronic acid ring structures for dp10(7) for the 10 ns MD simulation.....	283
Table 3.21 Energy values for each Vina predicted model for dp10(7), frame 250	284
Table 3.22 Energy values for each Vina predicted model for the flexible dockings for dp10(1), frame 45 in the 1 ns MD simulation	287
Table 3.23 Energy values for each Vina predicted model for the flexible dockings for dp10(1), frame 200 in the 10 ns MD simulation	291
Table 4.1 A comparison of the dihedral angles for dp10(1) set at the start of the MD simulation and those observed as an average value at the end of the MD simulation	331

Chapter 1: Introduction

1.1 Heparan Sulphate

Heparan sulphate (HS), along with the chemically related heparin, are ubiquitous glycosaminoglycans (GAG's) containing alternating sequences of repeating C1-C4 linked disaccharide units (Rabenstein, 2002). HS is ubiquitous in animal cells and found to be located on the surface of cells, but also in the extracellular matrix (ECM). HS is a member of a class of polysaccharides known as the GAG's. It is classified according to the hexosamine component found to make up the chain (Lindahl *et al.* 1994). Both HS/heparin (hep) are classed as glucosaminoglycans because of their glucosamine content; where as other members of the family such as Chondroitin sulphate (CS) and Dermatan sulphate (DS) are classed as galactosaminoglycan because of their galactosamine content (Gallagher and Walker, 1985). HS contains the alternating sequence glucosamine and uronic acid. The glucosamine can be either α -D-N-acetylated (NAc) or α -D-N-sulphated (NS). The uronic acid can be either β -D-glucuronic acid (GlcUA) or the modified α -L-iduronic acid (IdoUA). The glucosamino part of the chain can also be referred to as an aminosugar (Powell *et al.*, 2004), but for the purpose of this report, to maintain continuity, it will be referred to as a glucosamine.

There are a number of structures which fall under the GAG family name. The table below (Table 1.1) shows the names and what they are composed of, thus distinguishing them from each other.

Table 1.1

GAG	Uronic Acid	Galactose	Hexosamine
HS/Hep	GlcUA/IdoUA		GlcNAc
CS	GlcUA		GlcNAc
DS	GlcUA/IdoUA		GalNAc
KS		Gal	GalNAc

Table 1. 1 Disaccharide content of five different GAG structures.

Keratin Sulphate (KS) is the only chain which contains a galactose unit instead of uronic acid, as seen in table 1.1. The subtle differences in these give rise to completely different properties amongst the five chains shown.

Both HS/hep contain negatively charged groups but it is the extent of the modification which separates the two chains. Initial biosynthesis of HS/hep chains leads to the production of a non-sulphated polymer known as heparin or N-acetylheparosan, consisting of:



This structure (above), is then converted into N-sulphated derivatives (Gallagher and Walker, 1985). Additionally to this, many of the disaccharides contain one or more O-sulphate substituents. These can occur on C2 of the IdoUA forming α -L iduronic acid 2-O-sulphate, C6 of the glucosamine forming α -D-N-sulpho-glucosamine 6-O-sulphate (GlcNS(6S)) and α -D-N-acetyl-glucosamine 6-O-sulphate (GlcNAc(6S)). These are covered in more detail in the section GAG biosynthesis-chain modification.

1.2 Heparan Sulphate vs. Heparin: A Brief History

HS is a component found on the surface of all mammalian cells. It has been structurally maintained throughout evolution and the amount of variability comes from the amount of modified sequences found on the chain (Nader *et al.*, 1999). HS is rarely seen as a free entity but as part of what is known as a proteoglycan (PG) where it is involved with a protein core (Rabenstein, 2002). It is this protein core which anchors the chains to the surface of the cell. The GAG chains of HS are associated with a variety of core proteins, the most prominent one thought to be perlecan. Heparin on the other hand, is only found in a few tissues amongst a few species and is identified in the form of granules confined to connective tissue type mast cells. Heparin is only involved with one core protein – serglycin (Lindahl *et al.*, 1994). Heparin was discovered in 1916 and has been in clinical use as an anti-coagulant since the 1930's. It is second only to insulin as a therapeutic agent (Rabenstein, 2002). HS was discovered as an impurity of heparin but was then discovered to be just as critical in its own right. It is biologically important in cell-cell recognition and the control of growth (Nader *et al.*, 1999).

In terms of structural features, the differences between HS and heparin are less easy to identify. This is due to the somewhat 'incomplete' biosynthetic modification pathway the two chains undergo. The modification reactions are incomplete and therefore only a very small fraction of the potential modification targets are attacked by enzymes (Lindahl *et al.*, 1994). These modifications are the main reason why polymeric heparan is converted into HS (Gallagher and Walker, 1985).

1.3 Proteoglycans

The term proteoglycan (PG) refers to the complex in which a protein is attached to one or more GAG side chains. They are specifically made of a core protein to which unbranched polysaccharides are attached, either HS/hep, CS/DS, KS and HA. The proteoglycans represent a family of glycoprotein's which are highly glycosylated and can contain one or a number of GAG side chains. There are approximately 30 proteins which interact and are post-translationally modified as a PG. Below are those which are mostly associated with HS chains.

1.3.1 Heparan sulphate proteoglycans (HSPG's)

Those proteins which are associated with HS are collectively known as HSPG's. These can be classed in a number of categories, as summarised below:

- The Syndecans – type I transmembrane proteins.
- Glypicans – bound by a glycosylphosphatidylinositol (GPI) protein linkage to the cell membrane.
- Perlecan and Collagen XVIII and Agrin – present in the extracellular matrix, especially secreted into basement membranes.

Both the protein core and the GAG side chains play important roles in many biological processes such as control of growth, angiogenesis, differentiation, tissue repair and tumour metastasis. The HS chains are presented at the surface of the cell for potential ligand binding. The following section will discuss the four proteins which interact with HS to make a HSPG.

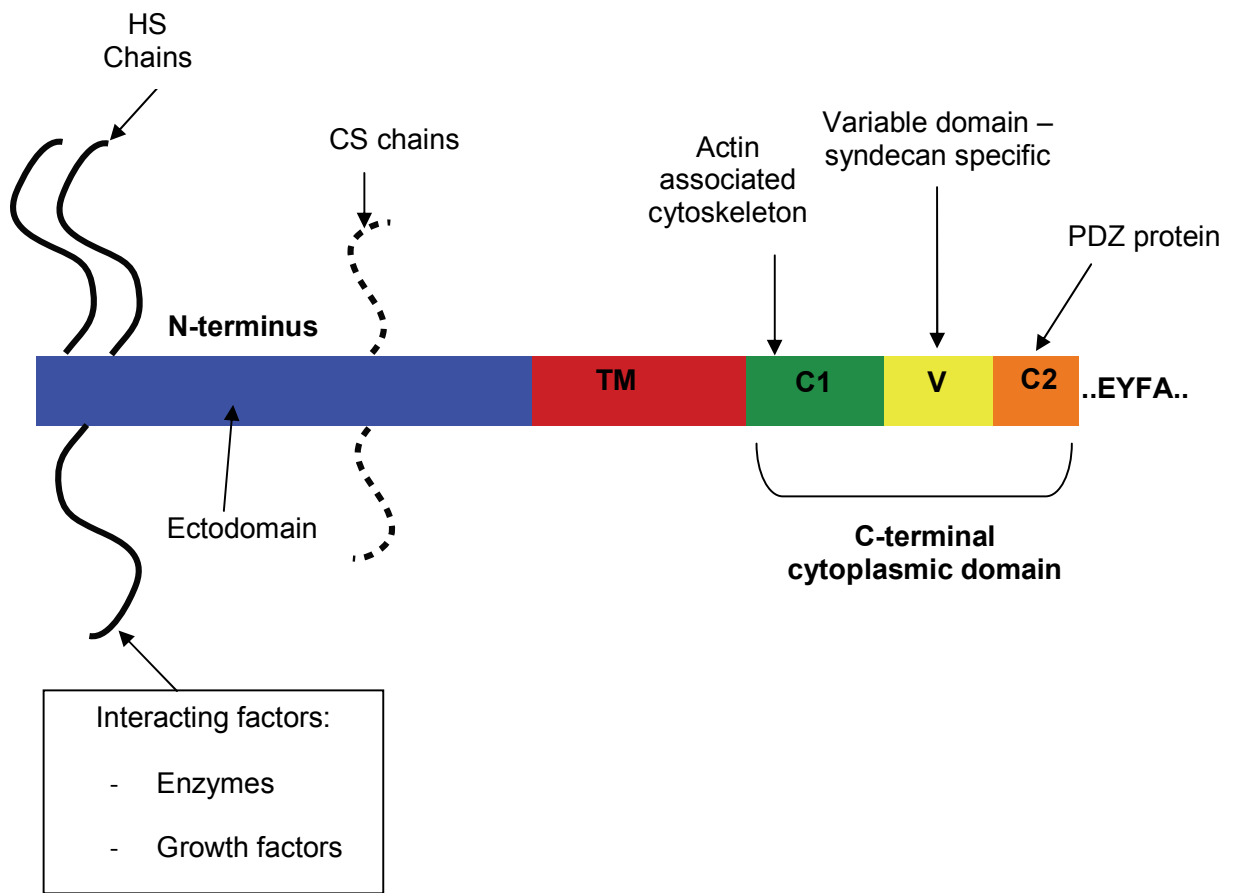
1.3.1.1 Syndecan

The syndecan protein family is made up of four members termed syndecans 1 – 4 in mammals and vertebrate; and are classed as type I (alpha-helical) transmembrane proteoglycans. Syndecan 1 was the first in the family to be cloned and since then the three other members have also been cloned, including one type in each of the following; *D. melanogaster*, *C. elegans* and *A. crassispina*. The numbering system found in the naming of the syndecan family is due to the order in which the cDNAs for each family member were first cloned. These are single transmembrane domain proteins which act as co-receptors for G-protein coupled receptors. Ligands which bind these are such like hormones, neurotransmitters and pheromones and act by activating a signal transduction pathway and ultimately a cellular response inside the cell. They are a large protein family of transmembrane receptors that sense molecules outside the cell and activate signal transduction pathways inside the cell. These proteins can carry between 3 - 5 long, unbranched HS/CS chains which are able to interact with a variety of ligands and play important roles in cell-matrix and cell-cell adhesion. The protein cores are 20 - 45 kDa in size and contain small cytoplasmic domains. These domains are found to be highly conserved and the relatively small size of these domains suggests that they lack any kind of intrinsic activity (Couchman, 2003).

Syndecans 1 and 3 and syndecans 2 and 4 comprise of separate subfamilies arising from gene duplication and divergent evolution from a single ancestral gene. All of the type I transmembrane domains consist of an N-terminal signal peptide of 3 - 60 amino acids long which directs the post-translational transport of a protein. The syndecans consist of an ectodomain which is found to extend into the extracellular space and initiates contact with surface to signal proteins (Figure 1.1). It has several consensus sequences for GAG attachment, but this seems to be the only function of this domain (Carey, 1997). It also consists of a single hydrophobic transmembrane domain and a short C-terminal cytoplasmic domain, which is split into three sections. The first section of the cytoplasmic domain is the C1 domain which is a region of conserved sequence and

is found proximal to the cell membrane. This houses the actin associated cytoskeleton and is followed by the V domain which is the variable region of this cytoplasmic domain, and is specific in all syndecans. The variability here refers to the comparison of one syndecan within family members; meaning the functions are unique to each member of that particular family. The most distal part of the protein is the C2 domain which is another region of conserved sequence. It contains the PDZ protein which helps to anchor the protein to the cytoskeleton of another cell (Couchman, 2003). The attachment site found in syndecans 1 and 3 occur in two distinct clusters. These attachment sites mostly contain HS chains but can also include CS/DS GAG chains. The first is found near to the N-terminus and the other is found near to the membrane attachment site. These sites are separated by a proline and threonine rich 'spacer' region. There is a tetrapeptide sequence found at the end of the C-terminus which is identical in all of the syndecan family.

Figure 1.1



TM = the transmembrane domain

C1, V and **C2** = the cytoplasmic domain

EYFA = amino acids glutamate, tyrosine, phenylalanine and alanine.

Figure 1. 1 The extracellular domains of syndecan.

Here the four syndecans consist of several extracellular domains for HS attachment. The HS chains from here then interact with enzymes and a variety of different growth factors (EGF, FGF, HGF and HBGF).

The C2 part of the cytoplasmic domain contains an EYFA sequence, which is where PDZ proteins mentioned above bind, such as CASK and syntenin.

Each syndecan is presented on the surface of the cell and is presented at specific times in development processes. All syndecans are referred to as “full time” proteoglycans except for those which have yet to undergo post translational modification and they each perform different functions (Couchman, 2003):

Syndecan 1: found in early development on cells associated with tissue morphogenesis.

Syndecan 2: (fibroglycan) is distributed in mesenchymal tissues and is abundant in the liver and neuronal cells.

Syndecan 3: (N-syndecan) is mostly found in neural tissue but can also be found in musculoskeletal tissue.

Syndecan 4: (amphiglycan/ryudocan) found in development and many other cell types such as the liver and kidney (Kim *et al.*, 1994).

1.3.1.2 Glypicans

The glypicans (GPC) are a protein family which are substituted to HS side chains to make a HSPG. They are anchored to the cell surface via a covalent linkage to glycosylphosphatidylinositol (GPI) lipid anchor (De Cat and David, 2001). The GPI anchor is made up of a phosphatidylinositol group linked through a carbohydrate containing linker to the C-terminal amino acid of a mature protein (Figure 1.2). The size of the glypican core protein is between 60 - 70 kDa. They all display an N-terminal secretory signal peptide and a hydrophobic domain which is required for the addition of the GPI anchor to the C-terminus. This is involved in the formation of the GPI structure. The homology amongst most glypicans is moderate but there

is a positioning of 14 cysteine residues contained in a globular domain which are very highly conserved. This suggests the three-dimensional structure is very similar in all the glypicans (Filmus and Selleck, 2001). The positioning of the HS chains is a shared characteristic amongst all glypicans. Here the insertion site is restricted to the last 50 amino acids in the C-terminus. It is this GAG attachment site which is extended further than any of the other domains which places the HS chains close to the cell membrane, allowing for interactions with proteins close to the surface of the cell (De Cat and David, 2001).

There are six members of the glypicans family identified so far in vertebrates named glypicans 1 - 6 (Perrimon and Bernfield, 2000). Two have also been identified in *D. melanogaster* and one in *C. elegans*. Their role is to modify cell signalling pathways and to contribute to cellular proliferation and tissue growth. There has also been evidence linking glypicans with tumour progression (Filmus and Selleck, 2001). GPC-1 has been found to be expressed in significantly higher amounts in pancreatic cancer. It also inhibits the mitogenic response of cultured pancreatic cells to FGF2 and heparin-binding EGF-like growth factors. The mechanism by which this occurs still remains unclear.

The sequence homology of the glypican core protein is 17-63% identical and it is at its highest level of similarity in the N-terminal domain. When the individual structures are compared in glypicans from different vertebrate species the homology is around 90% (De Cat and David, 2001).

Figure 1.2

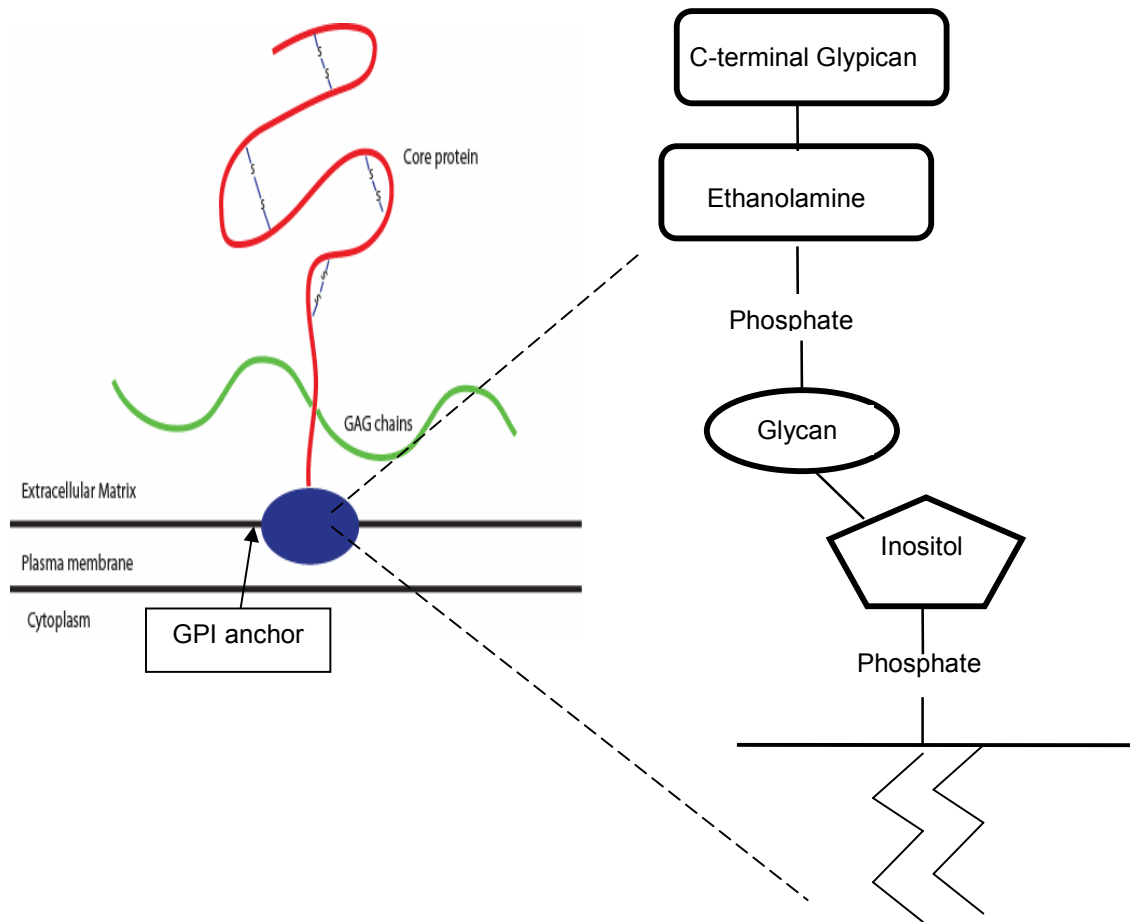
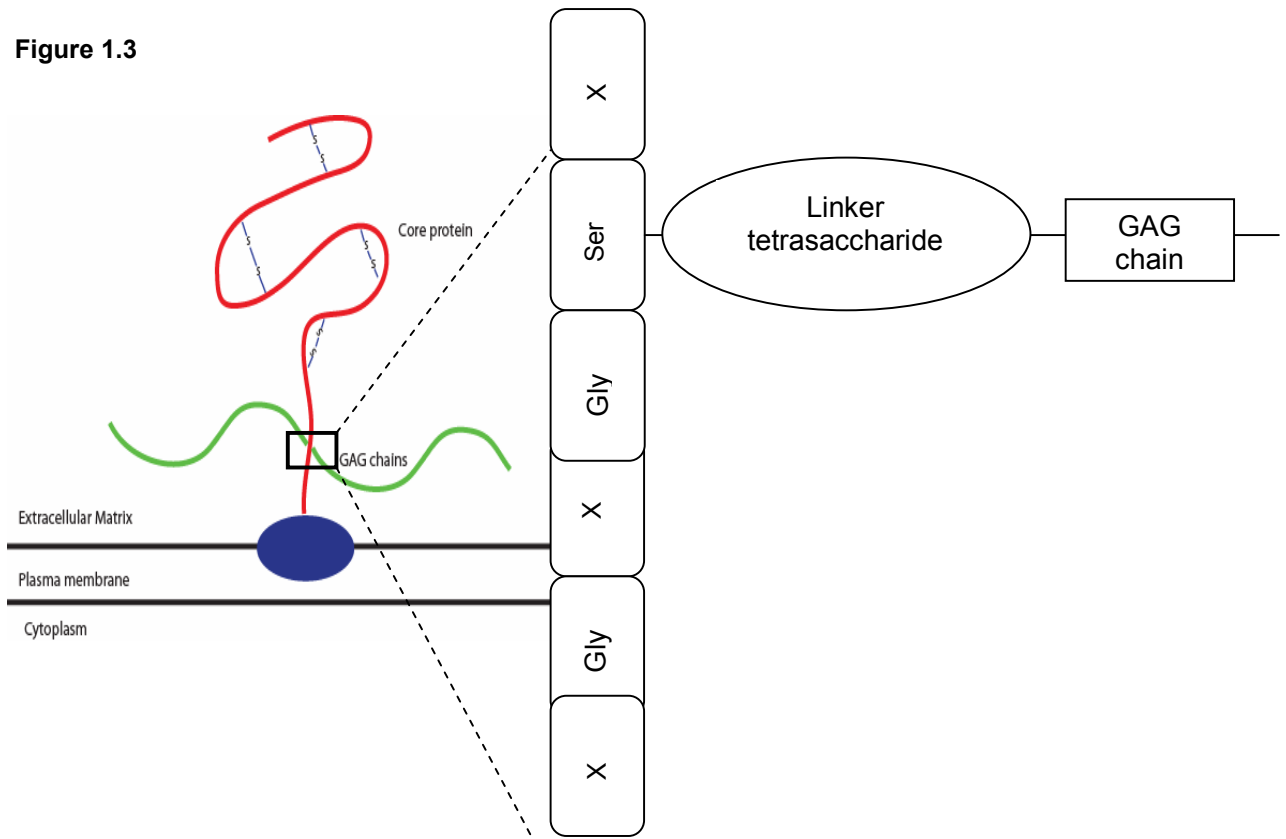


Figure 1. 2 GPI anchor of a glypican core protein attached to the plasma membrane.

The glypican protein is covalently linked to the plasma membrane of the cell by the C-terminal end via a GPI anchor. The presence of the disulphide bridges provides strength to the highly compact structure.

The disulfide bridges found in the core protein are there to give the structure the highly compact globular domain. The GAG side chains are attached to the protein core by the specific serine molecules found in the linker tetrasaccharide (Figure 1.2). The protein is attached to the plasma membrane at its carboxyterminus via the GPI anchor which is shown in the diagram below (Figure 1.3).

Figure 1.3



Adapted from (De Cat and David, 2001)

Figure 1.3 GAG attachment to the core protein glypican.

The GAG chains are linked to the protein via specific serine residues found in the linker tetrasaccharide. Here X can be classed as any other amino acid.

1.3.1.3 Perlecan

Perlecan is a large HSPG which is expressed as an intrinsic constituent of basement membranes; as well as cell surface and peri-cellular proteoglycans (Mongiat *et al.*, 2000). It is secreted into basement membranes and binds to and cross-links many extracellular matrix and cell surface molecules (Perrimon and Bernfield, 2000). Perlecan is thought to be the largest single chain polypeptide found in both vertebrates and invertebrate animals (Iozzo, 2005). It is synthesised by vascular endothelial and smooth muscle cells which are then deposited in the extracellular matrix. It has high affinity interactions with the Integrin family of proteins and is involved in a number of pathological processes. It has evolved from ancient ancestry by gene duplication and exon shuffling and is therefore very highly conserved. Perlecan is made of a 400 - 470 kDa core protein (Iozzo *et al.*, 1994) to which numerous O-linked oligosaccharides and up to four HS chains of 70 - 100 kDa can be attached, making the molecular weight of this protein in excess of 800 kDa (Iozzo, 2005). Antibodies for perlecan identify just one single precursor product suggesting it is encoded by one single gene (Iozzo *et al.*, 1994). One gene identified in *D. melanogaster* and *C.elegans* shows this PG has evolved from ancestors by exon shuffling and gene duplication (Friedrich *et al.*, 1999). The core protein comprises of five distinct structural domains (Figure 1.4).

Figure 1.4

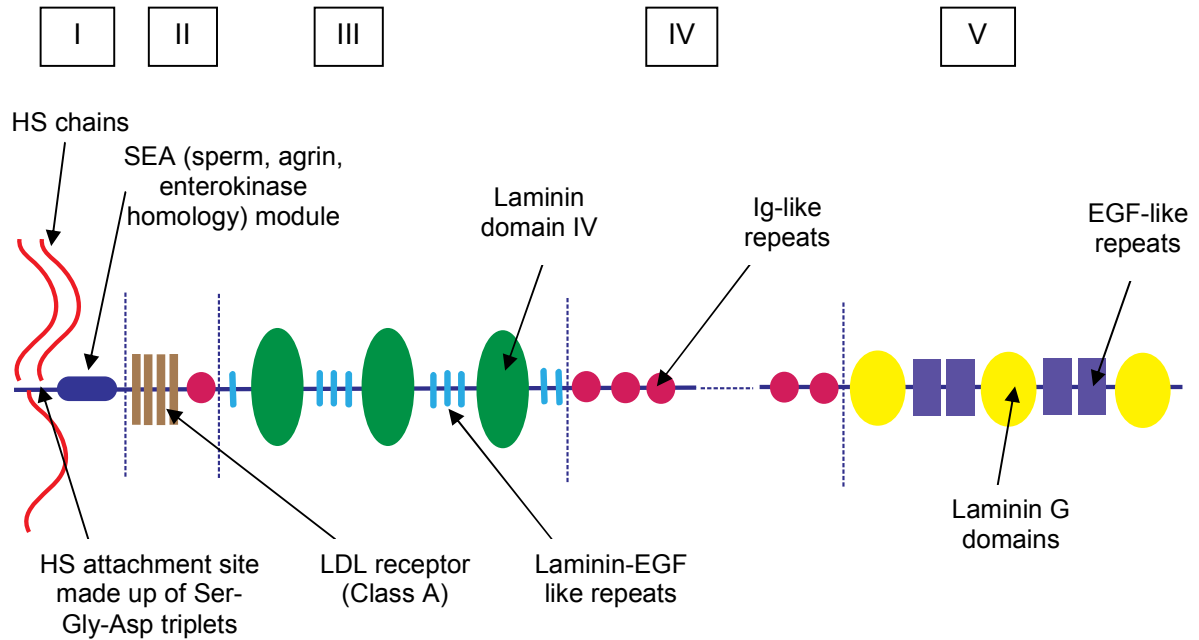


Figure 1.4 The domain structure of Perlecan.

The structure of perlecan is made up of five domains with a molecular mass of 470 kDa. In each domain exists a number of repeats, with the HS binding sites identified only in the first of the five domains.

The protein core is made up of a 'collage' of proteins and repeats which are shared by several ECM proteins involved in the uptake of nutrients and cell control. A major feature of perlecan is the presence of many internal repeats and the assortment of protein domains it is made up of. The different domains are involved in a variety of functions similar to other known polysaccharides including cellular growth, binding and the delivery of nutrients and lipids and many signalling pathways (Iozzo *et al.*, 1994). The domains run from the N-terminus to the C-terminus as shown in the above diagram. (Figure 1.4)

Domain I: This domain is unique to perlecan and is the proposed site for GAG chain attachments. Specific sites have been found on heparin sulphate (Noonan and Hassell, 1993). This end of the structure is known as the N-terminus and begins with a signal peptide, followed by a small domain of amino acids containing the three ser-gly-asp (SGD consensus sequence) tripeptides (or triplets). This section is 172 amino acids in length and the SGD consensus sequences are the predicted attachment sites for HS chains (Noonan and Hassell, 1993; Iozzo, *et al.*, 1994). The SGD sequence has shared features with other glycosylated sites found in other proteoglycans containing HS chains. What makes it different from other domains is the lack of any repeats and the lack of cysteine residues. It has an abundance of amino acids in the domain which is quite an unusual feature (Iozzo *et al.*, 1994). This domain also contains an SEA module which has homology for sperm, enterokinase and agrin and it is this section of the domain which favours glycosylation (Iozzo, 2005).

Domain II: This is made up of four repeats which are homologous to the LDL binding domain present in the LDL receptor gene (Noonan and Hassell, 1993). It contains four motifs which are cysteine rich and are found in a number of binding molecules (Iozzo *et al.*, 1994). These are then followed by immunoglobulin (Ig) -like repeats (Iozzo, 2005). The LDL repeats are receptor class A repeats and are organized in a similar fashion to the LDL receptor. The first two of the four repeats are encoded by one exon of near identical size to that in the LDL receptor. The other repeats are encoded by a highly conserved exon. These exons are all interrupted by phase I introns which allows for the possibility of alternate splicing (Iozzo *et al.*, 1994).

Domain III: Here the domain is made up of several different repeats, the first being homologous to internal globule rod-like regions found in laminin, particularly the laminin A chain (Noonan and Hassell, 1993; Iozzo *et al.*, 1994). These are globular domains which are flanked on either side by nine laminin endothelial growth factor (EGF) repeats. These repeats interrupt the globular domains (Iozzo, 2005). The segments homologous to the EGF repeats are made up of a series of repeats, typically eight highly conserved cysteine residues and two glycine residues (Iozzo *et al.*, 1994).

Domain IV: This has a similarity to the immunoglobulin super family and to N-CAM (neural cell adhesion protein), (Noonan and Hassell, 1993; Iozzo, 2005). This domain is the largest in human perlecan and contains 21 consecutive repeats of Ig-like repeats and is over 200 kDa in size. Mouse and the nematode worm have only 14 Ig-like repeats. 40 exons encode this region with an amino acid sequence of 70 - 100 residues. This part of the structure has three potential roles:

- To stabilize protein-protein interactions.
- To promote dimerization of molecules within the basement membrane.
- To enhance adhesion of neighbouring cells carrying perlecan at the cell surface.

(Iozzo *et al.*, 1994)

One N-CAM repeat separates domains two and three from each other and is 90% conserved with the same repeat found in mice (Noonan and Hassell, 1993).

Domain V: The C-terminal end is encoded by 16 exons which encode three globular repeats which are very similar to domain G repeats found in the laminin A chain (Iozzo *et al.*, 1994); (Noonan and Hassell, 1993). These are separated by two lots of two repeats of EGF-like repeats which are found to be cysteine rich (Noonan and Hassell, 1993). These EGF-like repeats are made of 40 amino acids and several glycine residues which are involved in the folding of this region (Iozzo *et al.*, 1994). Also found in this domain (not shown in the diagram) is another

potential binding site for GAG chains, which consist of two SGXG tetrapeptides. These may be substituted for other GAG chains, not just HS chains.

Perlecan is involved in many developmental processes at early and late stage embryogenesis. It also plays important roles in many cancers, such as breast carcinoma, colorectal carcinomas and cutaneous melanoma where increased amounts of perlecan are present. It binds to specific growth factors promoting activation and mitogenesis. It specifically binds to FGF2 forming a complex, causing blood vessel formation at higher rates than the HS-FGF complex. When suppressed it blocks autocrine and paracrine activities between FGF2 and human melanoma cells (Mongiat *et al.*, 2000). This results in the decrease in cell proliferation and the invasion into other tissues.

The many repeated domains, structures and modular proteins indicate this protein is involved with binding to, and the delivering of nutrients and lipids. It is also involved in many signalling pathways and cellular growth (Iozzo *et al.*, 1994). It is expressed in a variety of tissues, specifically by the sinusoidal epithelial cells found in the liver, but it is not expressed by the hepatocytes in the liver. This changes when the cells are cultured in vitro and they gain the ability to express perlecan. Surprisingly, it is found in normal articular cartilage, which is a type of tissue that does not contain blood vessels, and therefore does not contain a basement membrane. How the expression in this kind of tissue could well be due to the perlecan protein escaping the post-translational modification step. This is the process in which HS chains are added onto the protein. These would then be released as a glycoprotein leading to a different role for the protein (Iozzo *et al.*, 1994).

This information leads to the conclusion that perlecan is more abundant than was previously thought. It is not just restricted to epithelial surfaces; and the quantitative and qualitative expression is highly specific. Under certain conditions it can escape the modification process, allowing for it to have a different role.

1.3.1.4 Collagen XVIII/Endostatin

Collagen XVIII is a protein belonging to the sub family designated the multiplexins (multiple triple-helix domains and interruptions). It is a constituent of almost all embryonic and adult basal lamina (Halfter *et al.*, 1998). Collagen XV is also a member of this family and these are classed differently to other collagen proteins; where the domain structure is very much different to other collagen proteins (Zatterstrom *et al.*, 2000). It is a component of the basement membrane and is quite unique in that it has structural properties of both collagen and a proteoglycan (Marneros and Olsen, 2005). Collagen XVIII along with collagen IV are the only members of the collagen family that are conserved from *D. melanogaster* to *C. elegans* to humans and so suggests an important role of the protein in evolution (Dong *et al.*, 2003). Collagen XVIII is the third basal lamina HSPG, after perlecan and agrin, and is the first of the collagen family to contain HS chains which determine it as a HSPG (Halfter *et al.*, 1998). It also has the characteristics of a HSPG, in that it is a long heparitinase sensitive carbohydrate chain which is highly negatively charged.

Collagen XVIII is made up of 10 collagenous (COL) domains or repeats which are flanked on either side and interrupted by non-collagenous (NC) domains (Halfter *et al.*, 1998; Marneros and Olsen, 2005). It is structurally rigid, inextensible, and is made of a large triple-helical region, which the collagenous domains are found. The non-collagenous domains introduce flexibility into the structure between these helical regions (Zatterstrom *et al.*, 2000). It has typical features as a member of the collagen family existing as a non-covalently linked oligomer under non-denaturing conditions and the core protein of collagen XVIII is sensitive to enzymatic digestion (Halfter *et al.*, 1998). The NC domains contain the Ser-Gly (SG) sequences which are the attachment sites for the GAG chains (Zatterstrom *et al.*, 2000). These are similar to the SGD consensus sequences identified in perlecan, just without the presence of aspartic acid. The presence of a number of amino acids either before or after the SG sequences promotes the attachment of the GAG chains. There are eight potential GAG sites with three of them confirmed as carrying GAG's; thus confirming the protein as a PG (Dong *et al.*, 2003). Western blot studies of intact chick collagen XVIII carried out with chondroitinase had no affect and resulted in no shift in molecular weight,

which would suggest that the GAG chains present are not CS. The same study was carried out with the use of heparitinase and lead to the reduction in size of the protein from 300 kDa to 180 kDa showing that the GAG chains are in fact HS chains. Treatment with collagenase reduced the protein to nothing (Halfter *et al.*, 1998).

The GAG attachment sites were found using the following criteria:

- Peptides containing GAG's run on SDS-PAGE shown as a smear rather than a clear band.
- The digestion of heparitinase or chondroitinase results in a major drop in molecular weight; this can clearly be seen in PAGE.
- GAG-glycosylated peptides bind to anion exchange beads. Elution involves the use of NaCl greater than 1M.

The C-terminal part of the protein is the most conserved region, known as NC1. This part of the protein contains an 18 kDa peptide which has been shown to have anti-angiogenic properties and is known as Endostatin. This also has tumour suppressing activity (Halfter *et al.*, 1998). It was first identified as an endogenous angiogenic inhibitor and was derived from a larger protein isolated from conditioned media from a murine endothelial tumour cell line. The Endostatin peptide is released from collagen XVIII by proteolysis (Figure 1.5). This releases the NC1 terminal as a whole, which consists of three major segments: a 5 kDa N-terminal trimerization domain, a central protease sensitive hinge region and the endostatin domain (Zatterstrom *et al.*, 2000). The role of endostatin has evolved as a sensor for the release of proteolytic activities associated with angiogenesis. It also acts as a negative control by releasing angiogenic fragments. However, the role of endostatin is not thought to play a major role in angiogenesis, but is just one component of many which acts as a balance between pro and anti-angiogenic regulators (Marneros and Olsen, 2005).

Figure 1.5

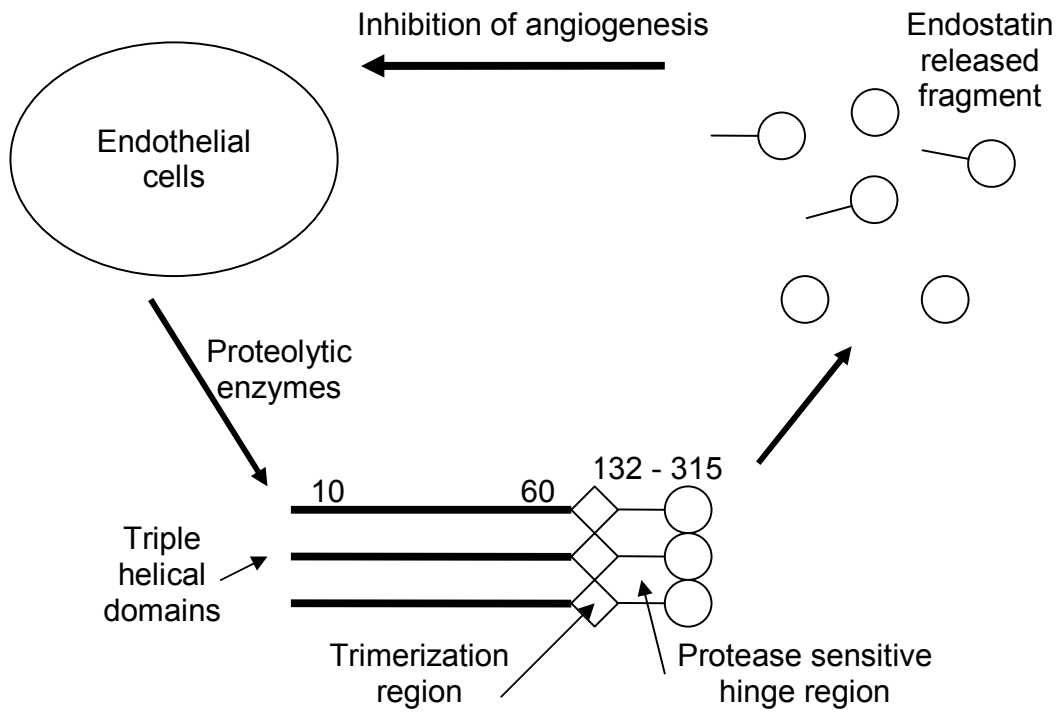


Figure 1.5 The release of the endostatin peptide from collagen XVIII.

The endostatin fragment is released from the C-terminal domain of collagen XVIII via proteolytic cleavage. It is the only known collagen that carries HS chains.

Collagen XVIII has been shown to play a critical role in the maintenance of ocular structures identified in Knobloch syndrome. This is an autosomal recessive disorder which is characterised by macular degeneration, high myopia and many other abnormalities. These lead to blindness at a very young age in patients with this condition. These results come from studies from a Brazilian family who showed a homozygous mutation in one particular gene; COL18A1. This has a particular mutation in the AG sequence of the gene. This affects the short form of collagen and not the long isoform, (collagen is made up of both long and short isoforms) the short form being identified in all vascular and epithelial basement membranes and the long isoform in the liver. This seems to suggest a requirement for collagen XVIII in the eye (Marneros and Olsen, 2005).

1.3.1.5 Agrin

The HSPG agrin is a 220 kDa protein of basement membranes which is highly concentrated in nervous tissue, the brain (Tsen *et al.*, 1995), and in the synaptic basal lamina at the neuromuscular junction. Agrin is a large, multidomain PG composed of domains which are homologous to others found in other ECM components. The agrin gene has been shown to encode multiple isoforms of the protein, through alternative splicing at three different sites in the gene. These spliced variants lead to different roles for the proposed protein (Groffen *et al.*, 2001). The use of SDS-PAGE gives a diffused band at around 400 kDa and after treatment with heparitinase or nitrous acid the band shifts to 220-250 kDa confirming that this is in fact a heparan sulphate proteoglycan (Tsen *et al.*, 1995).

The domain structure is composed of an N-terminal found in domain I, which comprises of a laminin binding domain and a signal peptide. This is known as the NtA domain (Denzer *et al.*, 1998) (Figure 1.6). Domain II then contains nine follistatin-like (FS) domains and within these the first potential binding site for GAG attachment is identified. The FS domains share homology to Kazal-type protease inhibitors, these include follistatin, elastase inhibitors and pancreatic trypsin to name a few (Groffen *et al.*, 2001). Two laminin (LF) EGF repeats separate FS nine from the

other repeats (Winzen *et al.*, 2003). Also in domain II, based in the central part of the domain are two serine/threonine (ST) rich domains. These have extremely high levels of serine, threonine and proline (Groffen *et al.*, 2001). The second of the potential GAG sites is found on the first of the ST domains. These adopt a rod-like shape and are extensively O-linked glycosylated which gives them an elongated, flexible structure and this is encoded by 365 amino acids (Denzer *et al.*, 1998). The two ST regions are interrupted by an SEA module which is also thought to be heavily involved in O-glycosylation. This module is found in sperm protein, enterokinase, agrin and perlecan. The final part of the domain is the EGF-like repeats. Domain III, IV and V each contain a laminin G-like domain which are separated by four EGF-like domains. In III and V there are alternatively spliced sites known as spliced sites X and Y. Another of these sites is found between the SEA module and the second ST site found in domain II. This has been designated as spliced site X. These variations lead to functions which differ greatly (Groffen *et al.*, 2001).

Figure 1.6

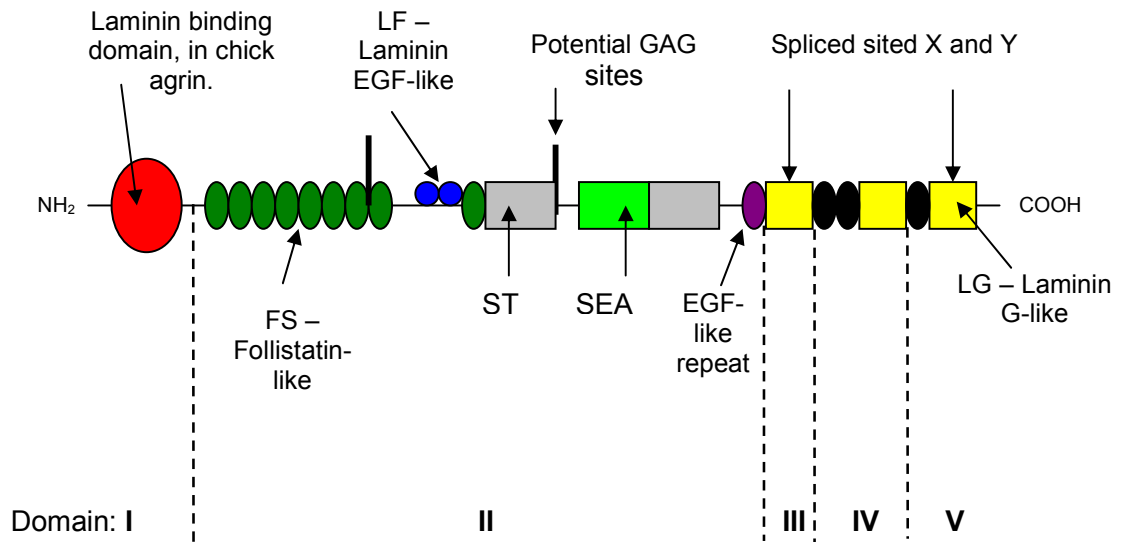


Figure 1. 6 The domain structure of Agrin.

Like perlecan, agrin comprises five domains each of which consist of many repeats. It is a 220 kDa protein found in basement membranes. This figure has been adapted from Groffen *et al* (Groffen *et al.*, 2001).

The first of the two HS attachment sites located in domain II between the seventh and eighth FS domains contains three SG sequences. The second site found in domain II is more centrally located in the ST rich domain and this contains four SG sequences (Winzen *et al.*, 2003).

Agrin is known to play an important role in the organization of the neuromuscular junction during embryonic development (Winzen *et al.*, 2003). Studies have shown agrin mediates the induction of AchR (acetylcholine receptors) clustering around the neuromuscular junction. This process is inhibited by the action of Hep/HS, showing that this protein interacts with GAG chains at the cell surface (Halfter *et al.*, 1998). The elimination of agrin in studies on mice leads to neonatal death due to the inability of the mouse to breathe (Winzen *et al.*, 2003). There is a high presence of agrin transcription in the human lung and kidney. Lung localization was confined primarily to capillary and alveolar basement membranes (BM), where as in the kidney it is confined to the glomerular BM but has also been found in tubular BM's (Groffen *et al.*, 2001).

An unusual role for agrin is as a major component in senile plaques found in patients who suffer with dementia (particularly Alzheimer's). Agrin interacts with an amyloidogenic peptide A β , found in a fibril state. This action is GAG dependant and this process accelerates the formation of A β fibrils. The presence of the HS chains possibly shield agrin from the action of proteolysis meaning that these senile plaques remain for a longer period of time in the body (Winzen *et al.*, 2003).

1.4 GAG biosynthesis

Structural diversity of all GAG's is generated in the Golgi apparatus of each and every cell and this is where each GAG chain undergoes biosynthesis. Chain biosynthesis occurs in three different phases, each one playing a major role in the biosynthesis of proteoglycans. The three steps involved are:

- Chain Initiation
- Polymerization
- Polymer Modification

The high complexity found in the structure of all GAG chains depends highly on the regulated expression of multiple enzymes, leading to a set of biosynthetic reactions which allows for the backbone of the polysaccharide to be modified (Wei *et al.*, 1993). These enzymes include sulphotransferases, glycosyltransferases and an epimerase which are found in the lumen of the Golgi apparatus. Another set of enzymes, the cytoplasmic enzymes are needed to transfer the nucleotide sugars to the forming chain. For this to occur, membrane transporters are required to transport these nucleotides from the cytosol to the lumen of the Golgi apparatus. Many of the sugars and sulphates needed are activated by nucleotide consumption in the cytosol of the Golgi apparatus to form their specific UDP-sugars. All of these reactions require the need of a sulphate donor in the form of 3'-phosphoadenosine 5'-phosphosulphate (PAPS) (Salmivirta *et al.*, 1996). These are then transported into the endoplasmic reticulum (ER) and the Golgi lumen. The presence of PAPS is essential as it is the only sulphur donor to all of the sulphotransferase enzymes (Prydz and Dalen, 2000), thus allowing this process to continue.

1.4.1 Chain initiation

The determining step in the process of chain initiation is the presence of an O-serine residue. This is the start of what is known as the common GAG-protein linker region for all GAG species (Gallagher and Turnbull, 1992; Kim *et al.*, 2001). To this O-serine is the addition of a Xylose molecule which is added to the hydroxyl group of the O-Serine by the action of the UDP-Xylo-transferase enzyme (xylT). Two D-galactose monosaccharides are added to this. The first being added by the enzyme UDP-Gal transferase I and the latter with the UDP-Gal transferase II enzyme (Carey, 1997). The final step in the initial biosynthesis of this GAG - protein linkage region is the addition of a D - Glucuronic acid residue by the enzyme UDP - GlcA transferase I. This reaction is catalyzed by the transfer of the glucuronic acid from the UDP - GlcA transferase I onto the O-3 of the terminal gal 2 (Figure 1.7). This occurs with an attack by the 3-hydroxyl group on the growing chain on the glucuronic acid at the C-1 position (Pedersen *et al.*, 2000). It is at this point in the chain that the GAG and its specific proteins are covalently attached to each other (Perrimon and Bernfield, 2000; Prydz and Dalen, 2000). Here the linker region serves as an acceptor for the first of the monosaccharide units to be attached which will commit the process towards the outcome of either a galactosaminoglycan chain or a glucosaminoglycan chain (Salmivirta *et al.*, 1996). The formation of the linker tetrasaccharide is thought to occur either in the lumen of the Golgi apparatus, or slightly earlier in the secretory pathway. It is thought that in chicken (*Gallus gallus domesticus*), chondrocytes xylosylation appears to occur in a pre-Golgi compartment of the cell in the ER and then this process continues into the Golgi apparatus thus allowing for galactose, glucuronic acid and sulphate groups to be added (Vertel *et al.*, 1993).

Figure 1.7

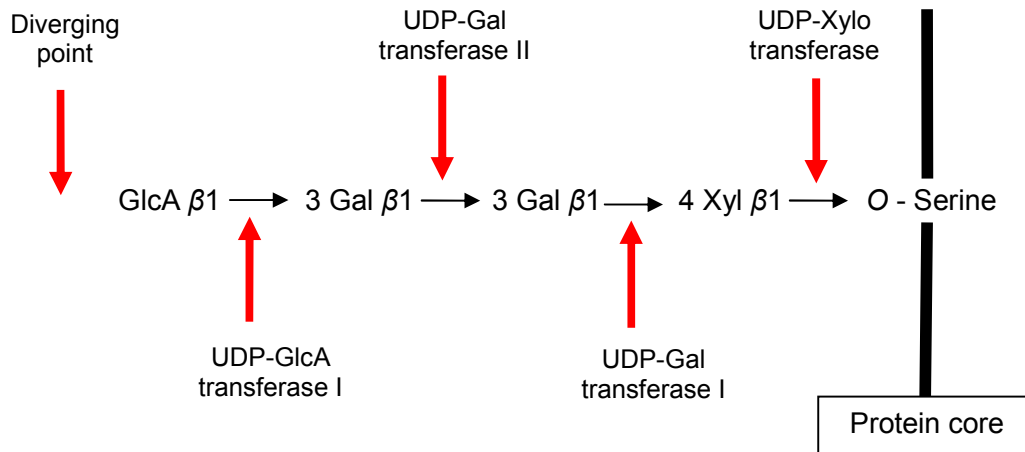


Figure 1. 7 Initial GAG biosynthesis for all GAG structure – The GAG-Protein linker region.

The biosynthesis of HS and CD/DS diverge after the formation of this GAG-Protein linker region. The four enzymes shown are solely responsible for the formation of this region with the next addition determining the generation of the GAG chain. This has been adapted from Sugahara and Kitagawa, (Sugahara and Kitawaga, 2000).

The pathways for HS and CS/DS biosynthesis diverge after the formation of a common core protein linkage tetrasaccharide (see text). Four separate enzymes XylT, GalT-I, GalT-II and GlcAT-I are responsible for the formation of this tetrasaccharide. The next enzyme to act either GlcNAcT-I or GalNAcT-I directs synthesis to HS/heparin or CS/DS (see text).

Once the GAG-protein linker region has been synthesised, the next addition determines which generation of chain it is to become; a HS/heparin chain (glucosaminoglycan) or a CS/DS chain (galactosaminoglycan). The following determines the steps for the biosynthesis of the GAG chains, discussing firstly the polymerization of HS/Hep, followed by CS/DS polymerization.

1.4.2 HS/Hep Chain Polymerization

Chain polymerization continues with the addition of an α - GlcNAc by the enzyme α - GlcNAc transferase I (Kim *et al.*, 2001). Once this specific addition has been made, the chain is then classed as a HS/Hep chain. The same enzyme can also add GalNAc, which is the defining unit in the biosynthesis of the CS/DS chain pathway. The α - GlcNAc transferase I enzyme belongs to the family of enzymes known as hexosaminyl-transferases (Salmivirta *et al.*, 1996). This enzyme also goes by the name of EXTL2. This is a member of the exostosin family (EXT); which are a family of tumour suppressor genes. EXTL1 - 3 all have high homology to the EXT family. EXTL2 exhibits activity which will transfer only the first of the $\alpha,1 - 4$ GlcNAc residues (Kim *et al.*, 2001).

Following on from this is the stepwise addition of two alternating monosaccharides, starting with GlcUA and then GlcNAc by the action of EXT1 and EXT2 respectively. They are transferred from their UDP - sugar nucleotide and are added to the non-reducing end of the forming chain (Salmivirta *et al.*, 1996; Esko and Lindhahl, 2001). EXT1 and EXT2 are known as HS polymerases and act as a copolymer exhibiting dual catalytic activities (Merry and Wilson, 2002). They act by elongating the HS chain by the alternative addition of both GlcUA and GlcNAc to the HS/Hep backbone (Kim *et al.*, 2001). It is thought that the GlcNAc/GalNAc which are added onto

the elongating chain are different from those added onto the GAG-protein linker region of the chain (Salmivirta *et al.*, 1996). These enzymes mentioned above catalyze these reactions and execute this role with incredibly high substrate specificity. The role of the HS polymerases can add in the region of 50 - 200 sugar units to the growing chain (Merry and Wilson 2002). Human EXT2 has been found to have a high sequence homology to human EXT1 (31% identity, 69.5% similarity), (Sugahara and Kitagawa, 2000). The EXT1 and EXT2 proteins join together to form a hetero-oligomeric complex leading to the accumulation of both of these proteins in the Golgi apparatus (Lamanna *et al.*, 2007). This complex contains both the GalNAc and GlcA transferase enzymes also known as HS polymerase. The enzyme complex contains two independent domains, one of which is able to transfer the hexosamine unit to the growing chain and the other domain adds the hexuronic part (Kolset, *et al.*, 2004).

The specificities of the EXT family 1 - 3 overlap amongst the family members but they still have a distinct role from each other (Kim *et al.*, 2001). EXTL3 is thought to have a role which is similar to EXTL2; in that it too transfers the first GlcNAc residue which initiates HS/Hep chains. It is possible that although these both have the same role they will act on different core proteins which contain HS chains. This would take place because the enzymes would be able to discriminate between the different amino acid sequences found in front of the binding sites (the SG consensus sequence) (Kim *et al.*, 2001).

HME (hereditary multiple exostosis), also known as HMO (hereditary multiple osteochondromas), is an autosomal dominant disorder which is characterized by benign cartilage-capped skeletal excrescences (bone tumours). This is found around areas of active bone growth leading to skeletal abnormalities. A genetic linkage of this disorder has been identified on three independent loci on chromosomes 8q24.1 (EXT1), 11p11-12 (EXT2) and also 19p (EXT3). The exostoses are characteristically benign bone tumours, however 2% of HME patients have been found to develop malignant transformations into chondrosarcoma or osteosarcoma (Sugaya *et al.*, 2008). This indicates that these genes responsible for HME may well encode tumour suppressor genes.

1.4.3 HS/Hep Chain Modifications

As chain polymerization occurs, chain modification takes place. Initially the polymer undergoes N - deacetylation/N-sulphation to create N-sulphated products. All N-acetyl-D-glucosamine residues (GlcNAc) will need to undergo the action of N - deacetylase in order for N - sulphation reactions to take place (Wei *et al.*, 1993). Only selected GlcNAc residues will go through the modification process, thought to be in the region of 40-50% of those residues (Merry and Wilson, 2002). This action occurs by the Golgi enzyme glucosaminyl N-deacetylase/N-sulphotransferase (the NDST's) (Perrimon and Bernfield, 2000). All sulphotransferases require PAPS to be the donor for all sulphate groups (Salmivirta *et al.*, 1996; Kakuta *et al.*, 1999). However; it should be noted that the presence of PAPS is not a requirement for N-acetylase activity as shown by Wei *et al.* (Wei *et al.*, 1993). They were also able to show that the activity of the NDST's is compromised by the presence of NaCl with the loss of 90% activity at physiological conditions. The NDST enzymes contain the active sites for both N-deacetylation and N-sulphation. Currently there are four NDST's which have been identified and in vitro studies suggest that each enzyme catalyzes the same reaction just in a different chemical context. The analysis of the four enzymes show a large overlapping distribution of two of the four enzymes, NDST isoforms 1 and 2 (Pikas *et al.*, 2000). This is indicative that both isoforms are needed and both play a role in HS modification. The occurrences of multiple isoforms are the products of different genes or spliced variants of the same gene.

The action of the NDST enzymes result in modified residues from the original GlcNAc residue. The enzymes replace the N - bound acetyl group found on this residue and replaces it with a sulphate group (Lindahl *et al.*, 1998). The possible outcomes of this are:

GlcNAc \longrightarrow GlcNAc – some residues will remain unmodified and are referred to as unmodified domains (see section on Domain structure).

- GlcN – the residue is N - deacetylated but is not N-sulphated (thought to be very rare)
- GlcNS – the residue is N - deacetylated with the removal of the acetyl group and is then N - sulphated.

The addition of these sulphates can alter the functional specificity and diversity of HS (Kakuta *et al.*, 1999).

The GlcN residues are more varied in HS than in heparin, of which they are predominantly N-sulphated. In HS the chain shows a more varied pattern with appreciable proportions of N-acetylated and N-sulphated residues and a very small amount of N-unsubstituted GlcN units (Lindahl *et al.*, 1998). Ultimately it is the NDST's which have a key role in the determination of the final structure of the polysaccharide chain (Ringvall *et al.*, 2000).

The N-deacetylation/N-sulphation step is a prerequisite for GlcUA to be converted to IdoUA by the enzyme GlcA (Glucuronyl) C5 epimerase. This is sometimes also referred to as Uronosyl C5 epimerase, but for the purpose of this report it will be referred to as the former. The epimerase converts GlcUA residues into IdoUA and are typically found in the Golgi apparatus (Pinhal *et al.*, 2001). This epimerase belongs to a family of enzymes known collectively as isomerases. These catalyze the stereochemical inversion of the configuration about an asymmetric carbon atom in a substrate which has more than one centre of asymmetry. This involves the abstraction and re-addition of the C5 proton onto the pyranose ring via a carbanion intermediate (Figure 1.8). This results in an inversion of configuration so that the carboxy group is shifted across the plane of the pyranose ring. This reaction is freely reversible in a solubilised enzyme system in the presence of $^3\text{H}_2\text{O}$ and will result in the incorporation of ^3H into both of the HexA units in the chain. The work carried out by Hager-McWhirter *et al.* (Hager-McWhirter *et al.*, 2000) shows the reversible reaction with regards to two different substrates: N - sulphated K5 polysaccharide (which contains only GlcA components) and O-desulphated heparin (predominantly made up of IdoUA components). After incubation with C5 epimerase with increasing amounts of time, the polysaccharides were

then recovered and subjected to nitrous acid. Results from the N - sulphated K5 polysaccharide showed that the proportion of IdoUA units reached a maximum level after approximately 1 hour. This maximum level did not increase further regardless of the amount of time left to incubate with the addition of fresh enzyme to the reaction. The opposite effect was shown with the O-desulphated heparin polysaccharide. Here an excess of IdoUA decreased over time due to the action of the epimerase. Equilibrium was once again reached after approximately 1 hour. It was also shown that the action of the epimerase did not have an effect on any HexA units located adjacent to any 6-O-sulphate glucosamine.

Figure 1.8

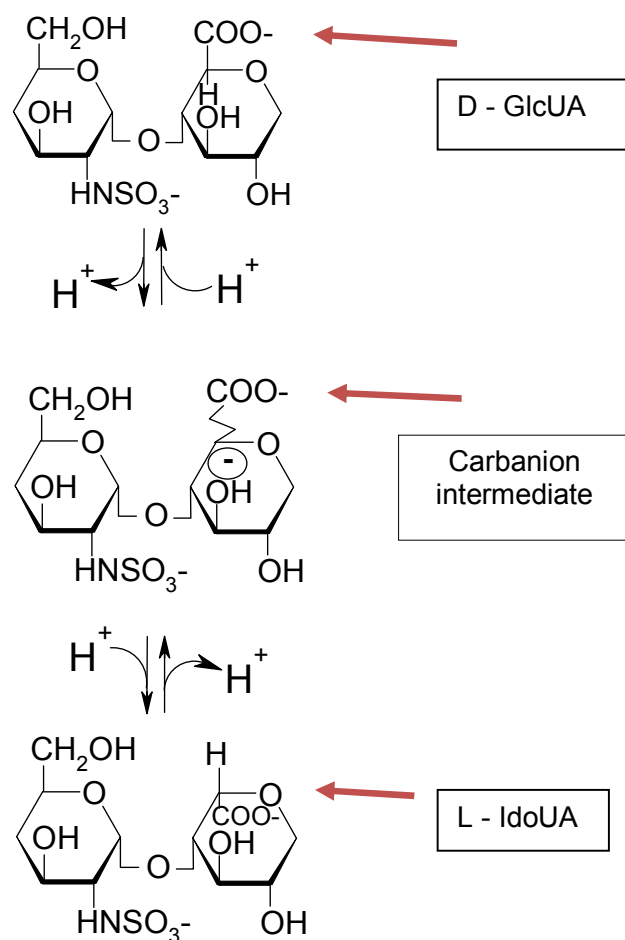
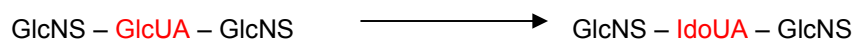


Figure 1.8 Basic mechanism for the C-5 epimerization reaction

The figure gives a step by step process of the conversion of the GlcUA to IdoUA with the use of the enzyme C5 epimerase. This figure has been adapted from Hagner-McWhirter *et al* (Hagner-McWhirter *et al.*, 2000)

A glucuronic acid residue is likely to be a substrate for D-glucuronosyl 5-epimerase if it is located between two N - sulphated glucosamine residues as shown below:



(Jacobsson *et al.*, 1984; Salmivirta *et al.*, 1996)

The role of the epimerase is to introduce flexibility into the GAG chain (the IdoUA residues are highly flexible, see Structural Conformation), (Prydz and Dalen, 2000). The epimerization reaction is thought to be a freely reversible reaction and is also thought to be promoted by the presence of an O-sulphate group found on the C2 position of an IdoUA residue (Lindahl *et al.*, 1998). This process produces the most common abundant unit found in HS/Hep chains, the IdoUA residue (Salmivirta *et al.*, 1996). Studies undertaken by Pinhal *et al.* (Pinhal *et al.*, 2001), suggest that C5 epimerase and 2-O-sulphotransferase form a complex and so therefore may work in co-operation with each other.

The final step in chain modification is the addition of sulphate groups carried out by the O-sulphotransferases. These catalyze the transfer of sulphate groups to various sites in the HS chain from a donor molecule, the most common one being 3'-phosphoadenosine 5'-phosphosulphate (PAPS). O - Sulphation can then take place, specifically at the C2 position of the IdoUA and the C6 position of both the IdoUA and all glucosamine structures (Merry *et al.*, 1999; Kreuger *et al.*, 2006). In rare cases O - sulphation also occurs at the C3 position of the glucosamines (Stringer and Gallagher, 1997; Sasisekharan and Venkataraman, 2000).

1.4.3.1 2-O-Sulphotransferase

Heparan sulphate 2-O-Sulphotransferase (HS2ST) is a critical enzyme which is responsible for the biosynthesis of HS (Merry and Wilson, 2002). Its role is to catalyze the transfer of sulphate groups from the molecule PAPS to the C2 position on the L-iduronic acid (IdoUA), (Kobayashi *et al.*, 1997). This enzyme produces 2-sulphated L-iduronic acid (IdoUA(2S)) which is a universal component in HS chains. It is found in varying amounts throughout the chain. This enzyme predominantly acts on IdoUA and so is generally only thought to act in the S-domain region of the chain (Merry and Wilson, 2002). HS2ST is thought to play an important role in HS degradation (Habuchi, 2000), and is known to share a common subunit with the 6-O-sulphotransferase (Kobayashi *et al.*, 1997). This could be a possibility when considering the following structure:

IdoUA(2S) – GlcNS(6S)

This could be synthesised by both enzymes, with the HS2ST acting on the substrate first followed by HS6ST (6-O-sulphotransferase) (Habuchi *et al.*, 2000). It has also been shown that, if there is a loss of HS2ST leading to an overall decrease in charge on the HS chain; there is an increase in the expression of both N and 6-O-sulphotransferase (Figure 1.9). This leads to more of a balance in charge on the whole chain. This is generally found in the S-domain region of the chain (Merry and Wilson, 2002).

Figure 1.9



Figure 1. 9 Composition of S-domains with and without 2-O-sulphotransferase.

A crude example of the difference in structure between the unmodified domains and those which undergo structural modification.

Panel A. Both HS2ST and HS6ST enzymes have acted on the substrate giving it a variety of sulphate groups.

Panel B. A loss of 2-O-sulphotransferase has lead to an increase in 6-O-sulphotransferase and also N-sulphotransferase (not shown).

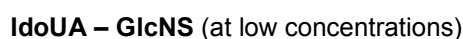
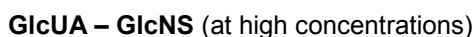
1.4.3.2 6-O-Sulphotransferase

Heparan sulphate 6-O-sulphotransferase (HS6ST) catalyzes the transfer of a sulphate group from 3' - phosphoadenyl sulphate to the C6 position of a glucosamine (Habuchi *et al.*, 2000; Habuchi, 2000). This is one of the key enzymes in the formation of HS domain structure (see Domain structure). There is a similarity between human and mouse HS6ST; with mouse HS6ST having three isoforms. The enzyme found in *D. melanogaster* has similar structural features to that found in vertebrates, but there is only one isoform dHS6ST. There is also only one isoform found in *C. elegans* (Kamimura *et al.*, 2001). HS6ST-1 has been shown to transfer sulphate groups to GlcNS units in the disaccharide sequence:



(Nakato and Kimata, 2002; Sedita *et al.*, 2004).

This study also showed that the HS6ST-2 isoform will transfer sulphate groups to different substrates at different concentrations:



(Sedita *et al.*, 2004).

The third isoform HS6ST-3 acts on both substrates (Habuchi *et al.*, 2000; Nakato and Kimata, 2002; Sedita *et al.*, 2004). However, it is thought that if a GlcNS residue is flanked on either side by IdoUA(2S) giving a structure of:



then this may escape modification (Salmivirta *et al.*, 1996), possibly due to the already heavily modified residues flanking each side of the GlcNS residue. Other studies by (Sedita *et al.*, 2004) have shown that 6 - O-sulphation occurs on specific sequences such as:

IdoUA – GlcNAc

HexA – GlcNS

HexA – GlcNS(3S) (found very rarely)

Here the HexA can be either of the uronic acids IdoUA or GlcUA. According to Harbuchi (Harbuchi, 2000) a good acceptor for HS6ST was GlcNS in either of the following sequences:

IdoUA – GlcNS

IdoUA(2S) – GlcNS

All of these sequences show that HS6ST does not transfer O-sulphate to any GlcNAc residues. This suggests that the process is catalyzed by a different enzyme, which is consistent with results from (Habuchi *et al.*, 1995).

It is thought that the regulation of the sulphation of IdoUA-GlcNS by HS6ST-2 is due to the GlcUA C5 epimerase. This is due to the product of this enzyme which inhibits the sulphation of IdoUA – GlcNS (Habuchi *et al.*, 2000). There is a pH dependence of HS6ST, which is thought to be between pH 6.2 – 6.4, and the activity of this enzyme is decreased by the addition of NaCl. (Habuchi *et al.*, 1995).

1.4.3.3 3-O-Sulphotransferase

The action of the 3-O-sulphotransferase enzyme (3-OST-1) acts on the C3 of N-sulphated GlcN residues (Razi and Lindahl, 1995). This is the final important modification during the biosynthesis of HS (Liu *et al.*, 1996). 3-OST-1 is not a transmembrane protein like the other modification enzymes, but is thought to be an intra-luminal protein (Habuchi, 2000). It is a key enzyme which is involved in converting non-anticoagulant HS into anticoagulant HS (Liu *et al.*, 1999); with the formation of the antithrombin III binding region found in HS (Razi and Lindahl, 1995; Habuchi 2000). This shows that 3-O-sulphated GlcNS(3S) is one residue required for the generation of the antithrombin binding site (Liu *et al.*, 1996). The inhibition of the 3-OST seems to be due to the action of both IdoUA(2S) and GlcNS(6S) which have been found in selectively desulphated heparin (Razi and Lindahl, 1995). The antithrombin III binding domain does seem to require the action of the other enzymes in order for its generation. These include the action of the NDST's, HS2ST, HS6ST and the C5 epimerase (Habuchi, 2000).

3-OST-1 is expressed in brain, kidney and in the heart (Habuchi, 2000). 3-OST-2, 3-OST-3a and 3-OST-3b are all isoforms of the enzyme but these lack the anticoagulant activity. However they are still able to transfer sulphate groups to the C3 position of all glucosamine residues (Liu *et al.*, 1999). 3-OST-2 is expressed strongly in the brain, whereas 3-OST-3a and b are expressed in various tissues throughout the body (Habuchi, 2000).

3-OST-2 has specific sequences to which it can transfer sulphate groups to. These are transferred to the GlcNS residue on:

GlcUA – GlcNS

IdoUA(2S) – GlcNS

The specific sequence for 3-OST-3a and b is just:

IdoUA(2S) – GlcNS

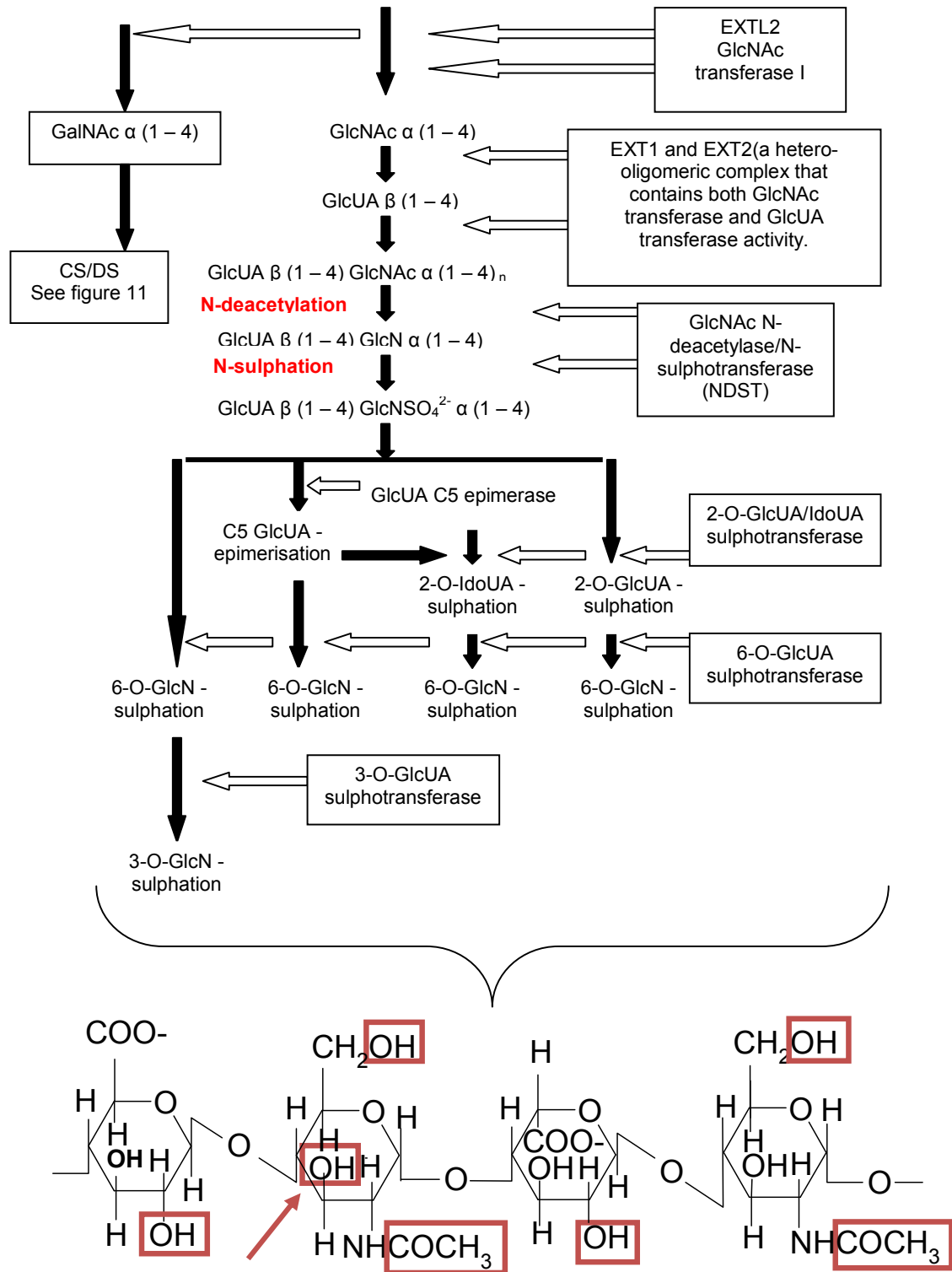
This shows that the sulphation of glucosamine is dependant on different structures around the glucosamine residue (Liu *et al.*, 1999). Figure 1.10 contains a detailed description of the GAG biosynthesis processes for a HS/Hep chain. Those positions which are modified by the O-sulphation process are shown highlighted. The rare 3-O-sulphation is shown by the red arrow on the second monosaccharide in the diagram.

Figure 1.10 GAG biosynthesis of HS/hep chains.

The modification process is quantitatively incomplete and so results in a HS/hep chain with a very highly complex structure. Most of the modifications that occur are non-random and are thought to be interdependent.

Mature HS comprises 40-50% GlcNS residues compared to heparin which has a higher level, approximately 85%. GAG biosynthesis occurs in the Golgi apparatus of all cells. This figure has been adapted from Prydz and Dalen (Prydz and Dalen, 2000).

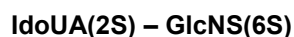
Figure 1.10



1.4.3.4 The Role of the Sulfs

Once synthesis and expression of HS is complete then the chain can undergo further modification to its structure and function. This occurs in extracellular compartments and the two enzymes presently known to take on this role are heparanase and heparan sulphate 6-endosulfatase (Dai *et al.*, 2005). The sulfs are a recent discovery, they are extracellular HS 6-O-endosulfatases which work by removing the sulphate groups from the C6 of specific groups within S-domains (Langsdorf *et al.*, 2007; Kiran-Safran *et al.*, 2009). The sulfs are secreted from the Golgi apparatus and become localised at the cell surface, or are released into the extracellular matrix (ECM) (Dai *et al.*, 2005). Their enzymatic activity and ability to dock in the correct place on the cell surface is due to the conserved sequences found contained in the members of the sulfatase family (Dhoot *et al.*, 2001).

The first family member to be identified was Qsulf - 1; sulfatase - 1 from quail (Dhoot *et al.*, 2001; Dai *et al.*, 2005). This particular enzyme has preference towards a specific substrate sequence, the trisulphated:



(Ai *et al.*, 2003).

Then came the cloning of other sulfs identified in humans (Hsulf -1 and Hsulf -2) and murine (Msulf). Qsulf-2 was also identified, which is very similar to Qsulf-1 but differs in gene sequence (Dai *et al.*, 2005). Qsulf-1 has less of an effect on the following sequence:



With no effect whatsoever on the sequence:



Extended reaction times showed no change from this result (Ai *et al.*, 2003).

Hsulf-1 is present in tumour cell lines and has been shown to be down regulated in these, whereas it is expressed at higher levels in normal tissue (Dai *et al.*, 2005). Sulfs also play important roles in muscle regeneration, specifically involving quiescent and activated satellite cells (self renewing muscle cells) and involves Msulf-1 and Msulf-2 (Langsdorf *et al.*, 2007). Mutations in these enzymes introduce the loss of function in specific pathways. Sulphation in itself is a process which is incomplete and so this alone creates a very complex structure seen in the form of S - domains and non - modified domains (Ai *et al.*, 2007). When the action of the sulfs is added into the equation this adds even more complexity into the structure, bearing in mind the first round of modifications already created earlier on in the biosynthetic pathway.

1.4.4 CS/DS Chain Polymerization

The biosynthesis of the CS/DS is slightly different to that of HS/hep chains. The enzymes responsible for the biosynthesis of the CS/DS chains are localised in different sub-domains of the Golgi apparatus specifically the trans-golgi network which is the most distal part of the Golgi complex. The enzymes responsible for the biosynthesis of HS/hep chains are located further in the Golgi network, specifically in the lower end of the trans-golgi network and also the medial-golgi network (Prydz and Dalen, 2000). These enzymes are a completely different set of enzymes to those found in the biosynthesis of HS chains. Once the formation of the linker tetrasaccharide has been completed the diverging point for CS/DS chains is the addition of GalNAc by the enzyme EXTL2. This is the same enzyme which would add GlcNAc to HS chains. It is also known as GlcNAcT1 but is a different enzyme to that which adds GalNAc to the linker tetrasaccharide in chain initiation. This addition to the chain leads to the synthesis of the repeating disaccharide -4 GlcUA β 1-3 GalNAc β 1-. This is done by the action of GlcUA transferase II and GalNAc transferase II, which are specific enzymes for the backbone synthesis of CS/DS (Sugahara and Kitagawa 2003). They are thought to exist either as a complex or they reside in a single

polypeptide. Their results show that the enzymes co-elute together by gel filtration chromatography and further chromatographic separation processes were unable to separate the two enzymes. This can lead to the addition of 40 - 100 disaccharide units in length for each of the chains (Sugahara and Kitagawa, 2003). The following diagram (Figure 1.11) highlights the CS/DS biosynthesis pathway. Again O - sulphation positions are shown by the red boxes.

Figure 1.11

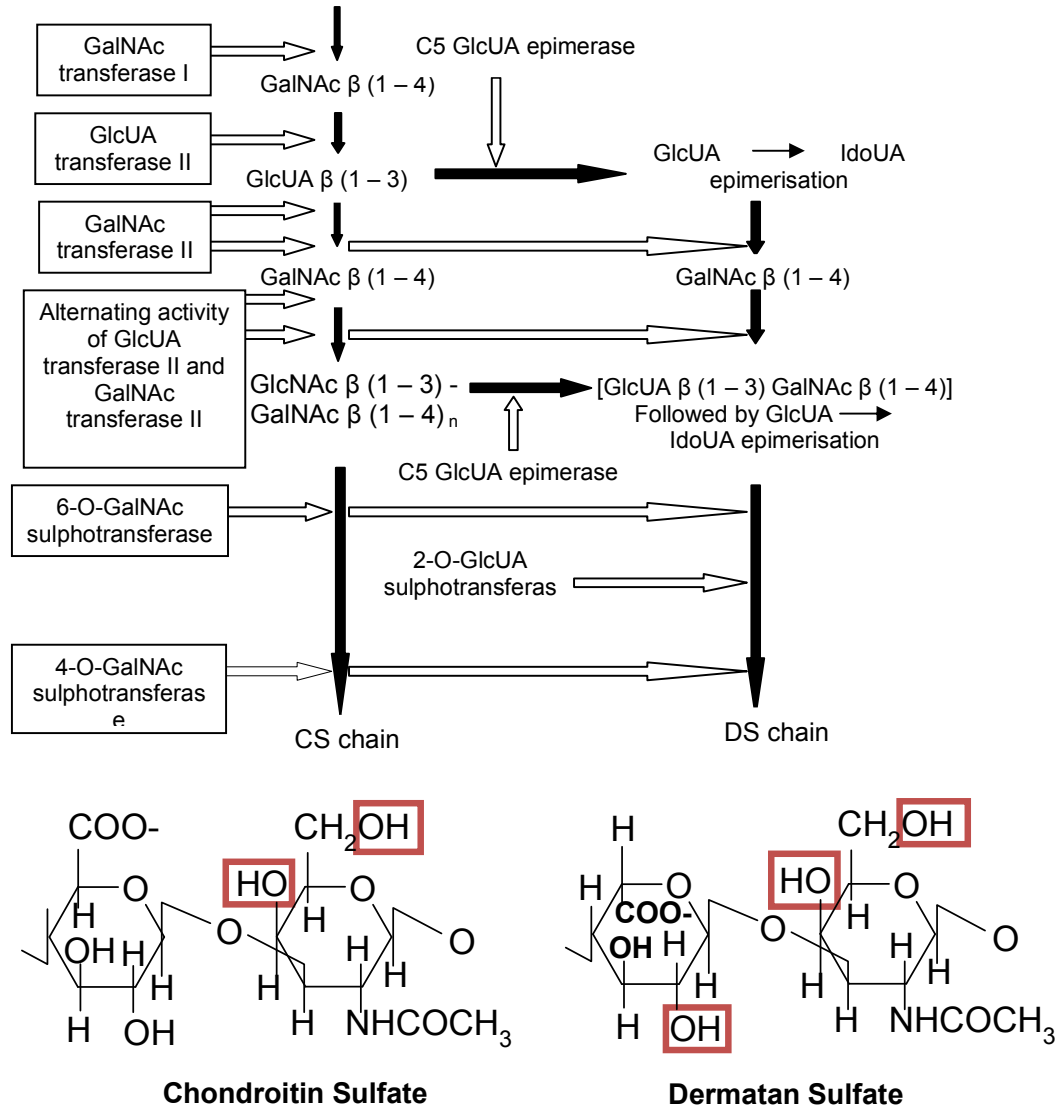


Figure 1. 11 Biosynthesis of Chondroitin/Dermatan sulphate chains.

Biosynthesis of CS/DS chains occurs in different compartments of the Golgi apparatus, specifically the trans-golgi network. This figure has been adapted also from Prydz and Dalen (Prydz and Dalen, 2000).

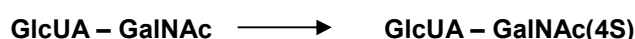
After backbone synthesis epimerization of GlcUA to IdoUA occurs with the same C5 epimerase enzyme used in HS/hep chains. If epimerization occurs then this chain will be classified as a DS chain as IdoUA residues are not found in CS chains. The only hexuronic acid present in CS is D-glucuronic acid, whereas in DS there are both D-glucuronic acid and L-Iduronic acid (Prydz and Dalen, 2000). The alternating action of GlcUA and GalNAc transferase II continues along with the action of the sulphotransferase family of enzymes. Here; like in the HS/hep biosynthesis, the chain can undergo 2, and 6-O-sulphotransferase but it can also undergo 4-O-sulphation too.

1.4.4.1 CS/DS O - Sulphotransferase

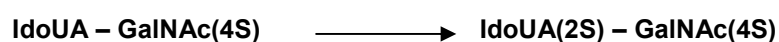
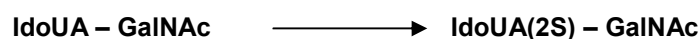
CS/DS can undergo 4 and 6-O-sulphation. Chondroitin-6-sulphotransferase, also known as C6ST, is secreted into the extracellular space of the trans Golgi network. It catalyzes the transfer of sulphate groups to position 6 of GalNAc residues. The product of this modification is Chondroitin-6-sulphate. The accepted structure for C6ST is the following:

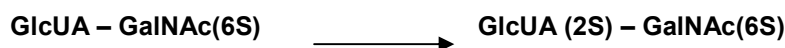


Chondroitin-4-sulphotransferase catalyzes the transfer of sulphate groups to position 4 of GalNAc residues. The acceptor for this is the same sequence as above:



Both of these enzymes are type II transmembrane proteins (Habuchi, 2000). The action of Uronosyl 2 - O - sulphotransferase acts only on DS chains. This adds the sulphate group onto position 2 of GlcUA/IdoUA in the following:





During this process the second C5 epimerase comes into action, again converting GlcUA to IdoUA (Prydz and Dalen, 2000). However; because of the location of the enzymes more of the CS/DS chain remains unmodified, unlike in HS/hep chains. CS/DS chains are often found as a co-polymeric structure (Sugahara and Kitagawa, 2003).

1.5 De-Polymerisation Techniques

There are two favoured methods of de - polymerization of the HS chain. These are chemical and enzymatic and can either be used separately a combined to give different products.

1.5.1 Chemical de-polymerisation

The chemical technique involves the use of the chemical nitrous acid to depolymerise HS/hep chains. This acts on the following sequence and does so selectively and quantitatively (Rabenstein 2002) and releases inorganic SO_4^{2-} :

GlcNS – HexA

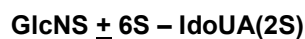
It also converts GlcNS, GlcNS(6S), GlcNS(3S) and GlcNS(3,6S) into 2, 5 - anhydro D - mannose (aMan). With this technique it avoids the production of the ΔUA (formed in enzymatic depolymerisation and not a naturally occurring ring structure) and can be used at a high pH (3 - 4) or a low pH (1.5). The reducing end of the HS chain seems to be more resistant to this treatment, possibly due to the local chemical environment around this area making it resistant (Merry *et al.*, 1999).

1.5.2 Enzymatic de-polymerisation

The enzymes used for the enzymatic depolymerisation of HS chains are naturally produced from the soil bacterium, *Flavobacterium heparinum*. It is able to use heparin as its main source of carbon, nitrogen and sulphur. For this to occur it must produce a range of enzymes which include glucuronidases, heparitinases (heparin monosulphate lyase), sulphotransferase and sulphamidases (Galliher *et al.*, 1981). It is the heparitinase enzyme which has become the enzyme of use for HS/hep depolymerisation (Dietrich *et al.*, 1973). This enzyme is an α 1 - 4 eliminase which acts specifically on glucuronic acid linkages between GlcNS-HexA; resulting in the formation of polysaccharides with a reducing end and a α - β unsaturated acid at the non-reducing terminal (Linker 1979). This process produces an unsaturated double bond between C4 and C5 on the non-reducing end of the chain which is termed Δ UA or just simply UA. This Δ UA is a sensitive UV chromophore with a maximum absorbance at 232 nm. This allows for the enzyme digestion process to be followed to completion and also allows for detection of the fragments produced via this method (Murphy *et al.*, 2008).

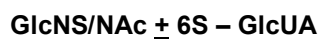
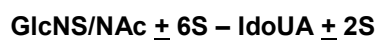
The useful activity of this enzyme allows for NS domains to be released from the chain (with the use of a specific enzyme – heparanase III). There are three types of lyase produced by the bacterium used to study HS/hep structures. Each acts on a different substrate in the chain, allowing for different sections of the chain to be released. These enzymes are:

Heparanase I (hep I): Acts on the sequence:



This releases di/trisulphated disaccharides and therefore releases the modified NS domains (Murphy *et al.*, 2004). This is also found in heparin in invertebrates (Nader *et al.*, 1999).

Heparanase II (hep II): Acts on the sequences:



This is the least specific of the three enzymes and cleaves a wide range of sequences (Murphy *et al.*, 2004). Heparin from invertebrates is also susceptible to the action of hep II (Nader *et al.*, 1999).

Heparanase III (hep III): Acts on the sequence:



This enzyme has a strong preference for GlcUA (Murphy *et al.*, 2004).

Generally even-numbered oligosaccharides are generated and are the most studied oligosaccharides throughout nature. However; odd numbered oligosaccharides may be generated by the mercuric ion-catalyzed removal of the unsaturated hexuronic acid from the non-reducing end. There is currently no method available to remove saccharides from the reducing end (Deakin *et al.*, 2008).

1.6 Domain Structure

Heparin is made up of a structure where >80% of the GlcN residues are N-sulphated (Gallagher and Walker, 1985) leading to the conclusion that it can also be heavily O - sulphated. It also contains a high portion of both GlcUA/IdoUA. In contrast to this, HS has a lower proportion of N - sulphated residues, thought to be closer to 50% of the GlcN residues being N - sulphated. It also has a lower amount of IdoUA and therefore a lower amount of N-sulphation (Gallagher and

Walker, 1985; Lindahl *et al.*, 1994). Complete modification of the chain leads to the trisulphated disaccharide being the most common sequence identified in the structure:

IdoUA(2S) – GlcNS(6S)

The incomplete modification of HS suggests it contains less of these trisulphated disaccharides and also fewer N - and O-sulphate groups (Gallagher and Walker, 1985). It is the difference in these sulphated and unsulphated structures which allows for the identification between HS and heparin. The most common disaccharide found in heparin is the trisulphated disaccharide mentioned above. The most common disaccharide found in HS is the unsulphated GlcUA-GlcNAc. These types of structures help to give HS its well known domain structure (Rabenstein, 2002; Powell *et al.*, 2004) and will make it easier to identify it from a typical heparin structure. The unsulphated disaccharide makes up the NA domain identified in the chains; which are known as the unmodified sections of the chain (also seen as NAc domains). Here no O-sulphation is identified and it is this area which is identified as being N-acetyl rich (Gallagher, 2001). These domains are thought to act as spacers between the more active and modified domains (Rabenstein, 2002). The NA domains make up a significant amount of the chain and are thought to play no role in the interaction with protein ligands (Murphy *et al.*, 2004). Segregated from this domain are the active sequences/domains known as the S domains or sulphated domains (Vives *et al.*, 1999). This section of the chain undergoes the most modification is found to be distributed evenly along the chain (Gallagher, 2001). These are also identified as NS domains and are typically 3 - 8 disaccharides in length (Rabenstein, 2002). This leads to a domain which is N-sulf rich and these are found completely separate from the N-acetyl rich domains (Gallagher and Walker, 1985). N-sulf is an incomplete process and it is this that helps to define the NS domains as this type of modification occurs only in defined regions (Vives *et al.*, 1999). The amount of variation available in the NA and NS domains also has great variety between different sources in the human body and also in different species (Powell *et al.*, 2004). It is the NS domain which bind to protein ligands but also the varying sulphation and sequence patterns, which play a key role in

this interaction between HS chains and different protein ligands (Merry *et al.*, 1999). It is this specific interaction that has led to the study of NS domains in great detail. HS is thought to be the mediator of not just binding to but also the activation of these ligands (Vives *et al.*, 1999).

There has been identified a third type of domain which brings together the NS and NA domains and thus shows that the two domains mentioned are not quite two separate domains in their own right, as was initially thought. The unmodified NA domains are typically made of approximately 15 disaccharide units (Gallagher, 2001). Towards the end of this domain; just before the beginning of the NS domain, is what is known as a transition zone (Gallagher, 2001; Rabenstein, 2002; Murphy *et al.*, 2004). This has been termed the NA/NS domain and although not much is known about these with regards to interactions with proteins (Murphy *et al.*, 2004), they are thought to be functionally significant in helping NS domains to interact with specific ligands (Rabenstein, 2002). These are areas of the chain which have undergone slight modification but not to the same extent as the NS domains. For instance the modification 6 - O - sulphate has been found in the transition zone (Maccarana *et al.*, 1996). The previously held domain structure of HS chains can be seen (Figure 1.12), alongside a comparison of the heparin chain.

Figure 1.12

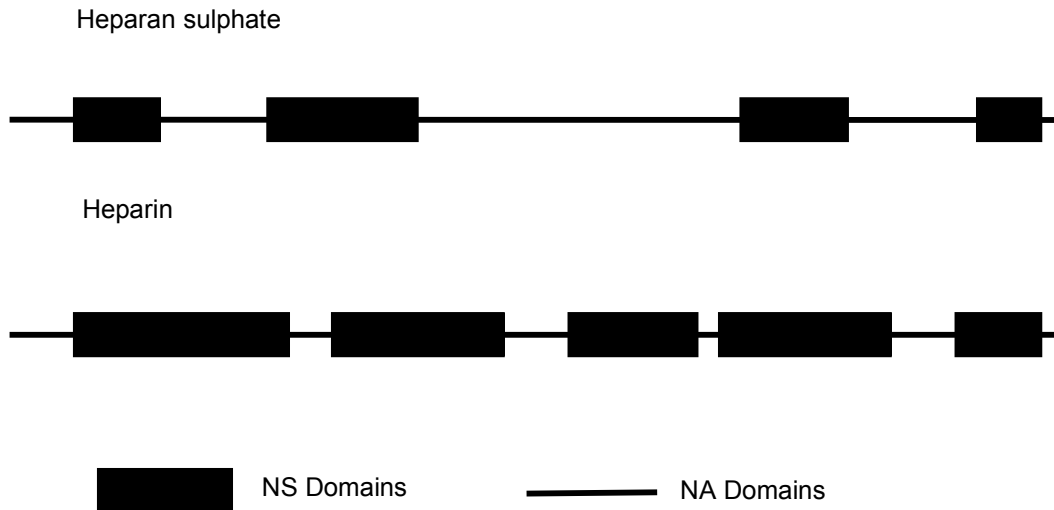


Figure 1. 12 HS/Heparin domain structure.

A representation of the previous held HS domain structure which is made up of highly modified blocks (NS-domains) separated by unmodified sequences (NA domains). The structure of heparin is essentially one continuous modified domain only interrupted by the occasional region of low sulphation.

The introduction of the NA/NS domains changed the view of the domain structure and added more complexity to the chain giving it a unique characteristic (Turnbull and Gallagher 1991). These were first demonstrated by Gallagher and Walker (Gallagher and Walker 1985) and followed up by Murphy *et al* (Murphy *et al.*, 2004) with the use of a different enzyme K5 lyase. The additional characteristic of these NA/NS domains is that they are more dynamic than were first thought. This additional flexibility would enable the NS domains to explore more of the space around it, thus enabling it to adapt more to the number of protein ligands by adapting their shape (Mobli *et al.*, 2008). This process has been shown by Stringer and Gallagher (1997) in HS platelet factor 4 (PF4). Here the NS domains are separated by long NA domains which are able to interact with two binding domains found on a PF4 tetramer, one identified on each side of the protein.

1.6.1 Action of K5 Lyase

The enzyme coliphage K5 lyase cleaves within non-sulphated sequences of four or more GlcNAc disaccharides. Meaning it cleaves within sequences that are found outside the NS domains; therefore concentrating solely within the NA domains (Figure 1.12). Further analysis showed these NA/NS domains were in fact in conjunction with NS domains or at least in close proximity to them (Murphy *et al.*, 2004). However, there is some confusion as to whether they flank both sides of the NS domains equally (Figure 1.13, Panel A), or are found on one side more than the other (Figure 1.13, Panel B).

Figure 1.13

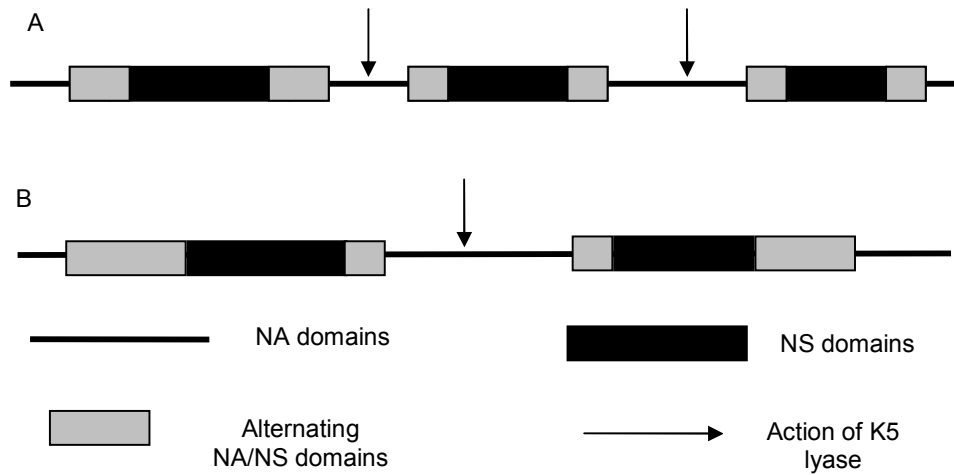


Figure 1. 13 A refined model of the domain structure of HS after treatment with KS lyase enzyme.

The arrows identify areas of the chain where K5 lyase enzymatic action takes place on HS and heparin. However it is unknown as to whether the NA/NS domains equally flank both sides of the NS domain (Panel A), or it occupies one side more then the other (Panel B).

Interestingly, the cleavage sites for both Heparanase I and K5 lyase appear to be around the same area, approximately 16 - 18 disaccharides away from the NS domains (Murphy *et al.*, 2004). The NA/NS were frequently found to contain unsulphated IdoUA residues with half the glucosamines containing 6-O-sulphate groups (Gallagher and Walker, 1985; Maccarana *et al.*, 1996). Some sequences were found to be resistant to the action of K5 lyase, namely the NS domains and the alternating sequences of GlcNAc – GlcNS. It would also only work on substrates with an α 1 - 4 linkage and not a β 1 - 4 linkage (Murphy *et al.*, 2004). With this devised new structure it is even more important in determining the position and placement of both the NA/NS and NS domains because these will both play major factors in the specificity and the local dynamics of the chain (Mobli *et al.*, 2008).

1.7 Structural conformation

It is a well known fact that carbohydrates do not contain secondary and tertiary structures but rather they populate multi conformations (Kirschner and Woods, 2001). These conformations are found in all HS ring structures and fall into three configurations; 1C_4 , 4C_1 and 2S_0 (the two chair conformations and one skew boat conformation respectively). These conformations play a major role in the complexity of the HS chain. Each residue in the chain can exist in its own conformation and these can have a specific influence on other residues upstream or downstream. Each residue will also have the ability to influence not only the binding but also the activity on specific protein ligands. Each of the different conformations will be discussed.

1.7.1 Δ UA/UA

The terminal uronic acid, produced by the action of the heparanase enzymes from the soil bacterium *Flavobacterium heparinum*, is characterised by the unsaturated double bond found between C4 and C5 on the non-reducing end of the chain. This terminal residue exists in the form of equilibrium between the 1H_2 and 2H_1 half chair conformations (Figure 1.14). This equilibrium is thought to be due to the substitution pattern or controlled by it (Bazin *et al.*, 1998). This ring structure is thought to be conformationally flexible (Ragazzi *et al.*, 1993).

Figure 1.14

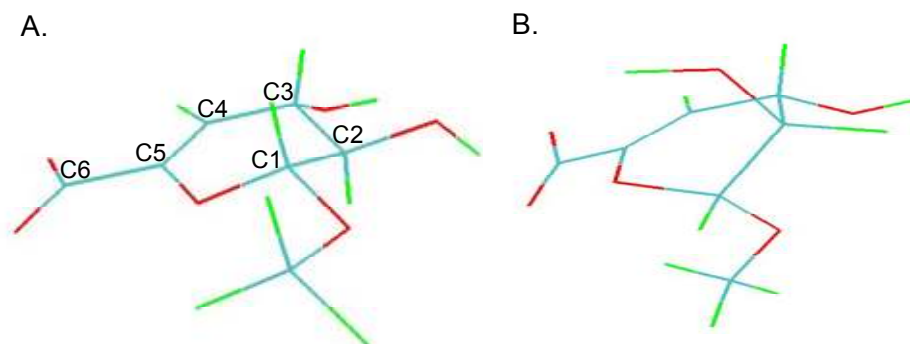


Figure 1. 14 Structural conformation of the terminal Uronic acid.

The pyranose ring of the terminal UA is thought to exist in equilibrium between the two low energy structures identified.

Panel A. The UA in the 2H_1 conformation.

Panel B. The UA in the 1H_2 conformation.

In the 1H_2 conformation all the bonds are found in the axial conformation. For 2H_1 the structure is almost flat in comparison, showing the equatorial conformation (Ragazzi *et al.*, 1993). As this structure does not exist in nature, it is not likely to have any role in biological interactions.

1.7.2 The N-glucosamines and The Gauche Effect

The collective term for this group of monosaccharides is the N-glucosamines and is made up of GlcNS, GlcNS(6S), GlcNAc and GlcNAc(6S). All of these are found in the conformationally locked 4C_1 conformation (Rabenstein, 2002; Powell *et al.*, 2004). This conformation is characterised by all of the non - hydrogen groups attached to the residue being held in an equatorial conformation, apart from the anomeric OH group (Rabenstein, 2002). There is an added feature with these residues which adds flexibility to the ring structure. This is the relative motion, or degrees of freedom, seen in the C5 - C6 bond shown by Hricovini *et al* (Hricovini *et al.*, 1995) on heparin epoxide (a derivative of heparin). This motion can be seen in all glucopyranosides and has been identified as the ω - angle, characterised by the O6 - C6 - C5 - O5 torsion angle (Figure 1.15) (Kirschner and Woods, 2001).

There are three possible conformations known as gauche orientations or rotamer populations which are known as:

- gg – gauche - gauche
- gt – gauche - trans
- tg – trans – gauche

Figure 1.15

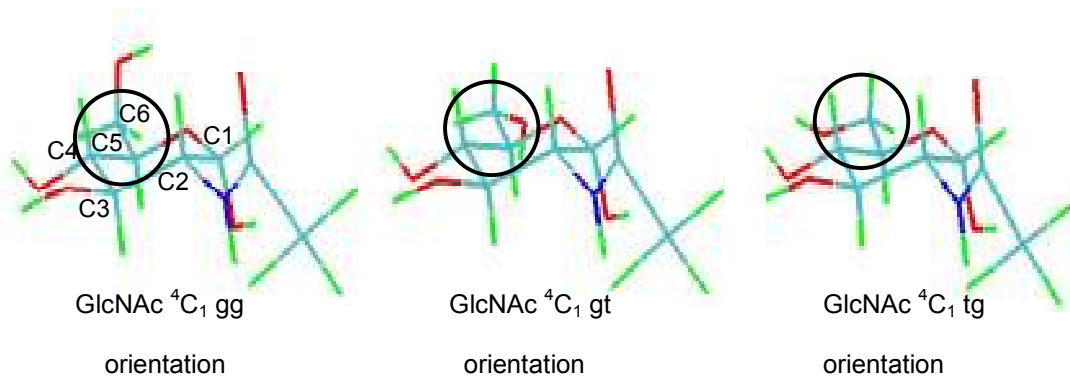


Figure 1. 15 The ω -angle identified in all N-glucosamines and the possible gauche orientations.

The GlcNAc residues is stable in the 4C_1 conformation exists in three possible rotamer conformations around the C5-C6 bond. The three rotamers are termed gg, gt and tg.

The conformations are thought to be solvent dependant as H₂O allows the disruption of the H bonds, thus allowing the rotamer populations to be determined by electronic and steric repulsion (Kirschner and Woods, 2001).

Studies by Hricovini *et al* (Hricovini *et al.*, 1995) have been unable to confirm the relative abundances of these orientations. However other studies have concluded that the rotation around the C5 - C6 bond is slightly hindered and that the gg rotamer is favoured (Mulloy *et al.*, 1994). Further work carried out by Kirschner and Woods (Kirschner and Woods, 2001), using molecular modelling have been able to show actual populations for these rotamers:

$$gt/tg/gg = 40/6/54$$

It has also been shown that excluding the O6 attachment shows no real difference in the rotamer populations, showing that the O6 attachment has no overall effect on the conformation of these residues. These studies mentioned at least agree in that the gg rotamer is the most favoured above the other conformations.

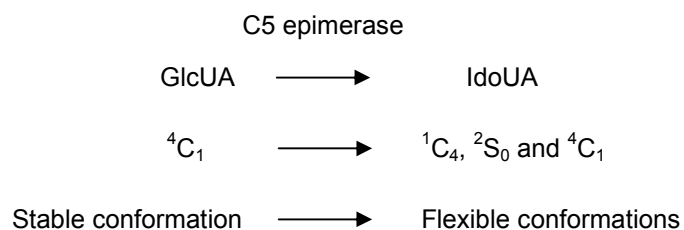
These populations are attributed to the gauche effect, which is defined as the tendency for a particular molecule to adopt a structure which has the maximum number of gauche interactions between adjacent electron pairs and their polar bonds (Clayden *et al.*, 2001).

1.7.3 Uronic Acids

1.7.3.1 GlcUA

The conformation of the GlcUA like the N-glucosamines is fixed in the ⁴C₁ configuration (Rabenstein, 2002; Powell *et al.*, 2004). This residue is highly rigid in comparison to its modified iduronate counterpart (Figure 1.16); which have undergone C5 epimerase (see HS/hep chain

modifications). It is this conversion from the rigid β -D-glucuronate to the α -L-iduronate which alters the conformational and dynamic properties of the chain.



Adapted from (Mulloy and Forster, 2000).

1.7.3.2 IdoUA

The majority of the flexibility in HS chains comes from the presence of three different conformations. These are the common chair forms, 1C_4 and 4C_1 and then the more unusual 2S_0 skew boat form (Figure 1.16), (Murphy *et al.*, 2008). Both the 1C_4 and 2S_0 conformations are found on internal iduronates, whereas the 4C_1 form is thought to only occur on the terminal iduronate residues (Ferro *et al.*, 1990).

Figure 1.16

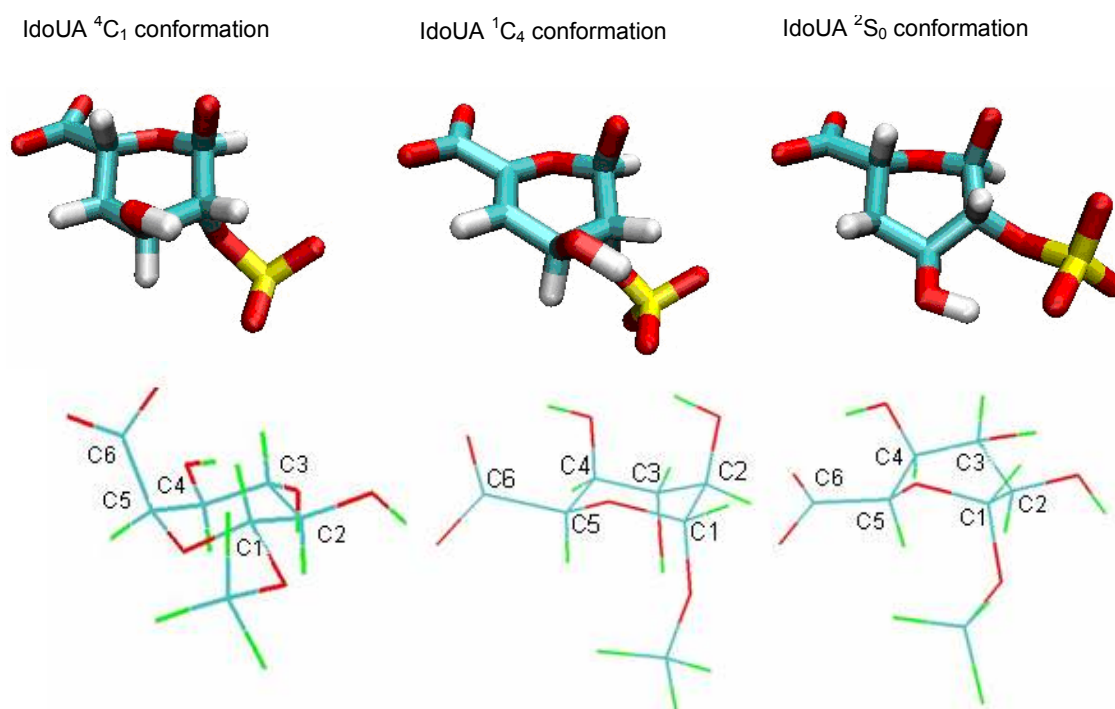
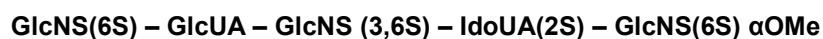


Figure 1. 16 Structural conformation of the Iduronates.

The methyl- α -L-Iduronic acid exists in equilibrium between three low energy conformations. It is thought that the 4C_1 conformation is only found to form the terminal unit in an oligosaccharide. Cremer-Pople puckering parameters are characterised as $\theta =$ approx 160° (1C_4) and $\theta =$ approx 90° (2S_0).

These conformations have been confirmed by the use of NMR with 3 - bond proton - proton coupling constants ($^3J_{\text{H-C-C-H}}$) along with molecular modelling studies (Hricovíni and Bízík, 2007). This change in configuration had also been shown by previous work carried out by Hricovini *et al* (Hricovini *et al.*, 1995) on heparin epoxide which converts from the 'normal' 4C_1 to a half chair configuration not previously identified in any HS structures.

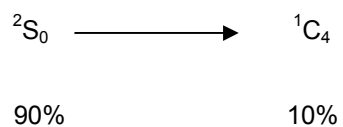
When interactions between specific proteins are involved then the conformation of a particular IdoUA seems to fall into one of the two favoured conformations. This has been identified in heparin when it binds to the molecule histidine. Here it was found to bind in the 1C_4 conformation (Chuang *et al.*, 2000). Another study with heparin which involved binding to antithrombin III (used in preventing blood coagulation) is shown to bind to a particular sequence in the HS chain:



α OMe is the α -methyl glycoside, (Das *et al.*, 2001).

Results from the crystal structure showed both 1C_4 and 2S_0 conformations present in the structure. All the other residues were identified in the rigid 4C_1 conformation. However the resulting NMR data identified only the 2S_0 conformation. It is thought that the presence of the neighbouring sulphate groups, in particular the 3- and 6-O-sulphate group were the main driving force in the resulting configuration (Murphy *et al.*, 2008).

It is thought that the main reasoning behind these conflicting findings is that when in a bound complex the 2S_0 conformation is shown as the 'active' conformation. When unbound both of the configurations are seen but with the equilibrium firmly towards the 2S_0 conformation:



(Hricovini *et al.* 2001)

This belief is also shared by Das *et al* (Das *et al.*, 2001) who also found the iduronate, when bound to antithrombin III, adopts the 2S_0 skew boat form.

1.8 Interactions with proteins

1.8.1 HS Binding – Consensus sequence

HS has been shown to bind to a number of proteins, in fact over 200 proteins have so far been identified (Kreuger *et al.*, 2001; Forsten-Williams *et al.*, 2008). The most well known proteins are antithrombin III, IL - 8 and the FGF family (Bitomsky and Wade, 1999). It is the ionic interactions which are important for the promotion of the binding sites found in specific HS binding proteins. Consensus sequences found on the surface of proteins are made up of positively charged amino acids (Hileman *et al.*, 1998) and these specifically interact with negative groups on the HS chain. These negative groups are specifically the sulphate and the carboxylate groups found on all GAG chains (Bitomsky and Wade, 1999), allowing them to participate in electrostatic interactions with proteins and cations (Turnbull and Gallagher, 1991) which are found in the sulphated domains (NS domains) found throughout the HS chain (Kreuger *et al.*, 2006). It is these domains which provide multiple binding sites for protein ligands. The GAG - protein consensus sequences are defined as an array of basic amino acid clusters of lysine, arginine and histidine found in formations such as:

XBBXBX or XBBBXXBX

B = basic residues

X = hydrophobic residues (Hileman *et al.*, 1998).

There is a distance of 20 Å between one set of positively charged residues and the next binding site (Bitomsky and Wade, 1999).

1.8.2 Fibroblast Growth Factor Biology

There are currently 22 identified members of the fibroblast growth factor family (FGF's) in humans and mice (Rieckmann *et al.*, 2007) and are recognised as fundamentally important in embryogenesis (Kreuger *et al.*, 2005). They have also been identified in *C.elegans*, *D. melanogaster*, vertebrates and some anthropod viruses (Harmer *et al.*, 2004). These FGF's are then split into different sub-families. FGF's were initially purified from bovine pituitary glands as a mitogen which could stimulate the growth of NIH3T3 cells. It has been identified in *D. melanogaster* which is known to contain just one type, known as branchless (bnl). *C. elegans* also contain just the one type, known as Egl - 17 (Ornitz, 2000). They are involved in a number of activities within the body such as tumour cell replication, migration, invasion, vascularisation, targeting and attachment by bacteria, protozoa and viruses (Hileman *et al.*, 1998).

The 22 known FGF's are all similar and are related in both sequence and structure, however the most universal of them all is FGF1 which is able to bind to all FGFR's (Wu *et al.*, 2003). The internal core is made up of 28 highly conserved amino acids with 6 invariant amino acids (Ornitz, 2000). They are 15 - 30 kDa proteins with the whole core region having homology to 120 - 130 amino acids. The FGF's have homology for the IL - 8 form of growth factor (Harmer *et al.*, 2004). These amino acids make up the β - trefoil fold which comprises of 12 β strands arranged in three lots of four β sheets (Olsen *et al.*, 2003). B - trefoil structures β structures with a pseudo - 3 fold axis of symmetry (Harmer *et al.*, 2004). Further diversity is created through alternative splicing of the FGF's and also through a variety of post - translational modifications. They all act specifically with different FGFR's, with the most specific being FGF7. This will only bind to the receptor FGFR2 IIIb (a spliced variant) (Duchesne *et al.*, 2006).

1.8.3 Fibroblast Growth Factor Receptors – The FGFR's

The FGF's are known to bind to two types of receptor:

- HSPG's - low affinity cell surface receptors ($Ic_{50} \sim 5-50$ nM) (Thompson *et al.*, 1994).
- Tyrosine kinase receptors (FGFR's) – FGFR 1 - 4 with high affinity (Pye *et al.*, 2000). ($Ic_{50} \sim 10-200$ pM) (Thompson *et al.*, 1994).

The latter are known as fibroblast growth factor receptors and are products of five distinct mammalian genes (Goodger *et al.*, 2008). These FGFR receptors bind to FGF with a varying amount of affinity and are found to be mostly specific:

- FGFR1 – expressed mostly in connective tissue.
- FGFR2 – found primarily in bone.
- FGFR3 – found in cartilage.
- FGFR4 – Identified in muscle.

Diversity is also found in cloned receptors such as FGFR1. This lacks the cellular tyrosine kinase domain which is required for the signal transduction found in muscle and cartilage (Rieckmann *et al.*, 2007). FGFR's are also found in spliced variants A, B and C, of which it is domain III which is spliced. Spliced variant A is secreted and functions as a receptor. Type's B and C display different binding ligand requirements, such as FGFR2 IIIc can bind to both FGF1 and FGF2 equally. FGFR IIIb binds FGF1 increasingly more than FGFR IIIc, but binds to FGF2 1000 fold less (Goodger *et al.*, 2008). The different receptors and their spliced variants result in the binding of a variety of different ligand binding domains (Nugent and Iozzo, 2000) and the fact that it is tissue specific adds to the great amount of diversity (Ornitz, 2000).

The FGFR's are composed of an extracellular region of three Ig - like domains: Ig I, Ig II and Ig III. They have a single transmembrane helix and an intracellular tyrosine kinase domain (Harmer, 2006). In between domain I and II is a stretch of 30 acidic residues (Wu *et al.*, 2003). X-ray crystal structures show domains II and III are separated by an α 4 - linker peptide, removing any kind of interaction between these two domains (Stauber *et al.*, 2000). The three domains here are

membrane proximal IgG domains with domains II and III containing the FGF binding sites. Domain II also contains a second binding site for binding to HS (Harmer *et al.*, 2004). The HS binding site is situated on the loop between β strands 10 and 11 and is composed of several basic amino acids. It is this structure that forms the binding site, particularly for FGF2 (Ornitz, 2000). The domain structure of domain II means there is a canyon or space of positive charge which is the area where HS/hep chains are able to bind or bridge the complex (Schlessinger *et al.*, 2000). It has been shown that domain I is dispensable with II and III showing an equal degree of binding (Wu *et al.*, 2003).

1.8.3.1 FGF Isoforms and diversity

Splicing of the receptor domains leads to a generation of diverse protein isoforms. These in the presence of HS/hep bind with varying affinities to different FGF ligands (Ornitz *et al.*, 1996). The splicing of the receptors increases the intricacy and the diversity of the receptors; which is enhanced by the splicing of the FGF mRNA (Schlessinger *et al.*, 2000). It has been shown that FGF's can bind differently to different ligands (Wu *et al.*, 2003). An example of this is FGF1 can bind to FGFR2 IIIb and FGFR2 IIIc. However FGF2 can not bind to FGFR2 IIIb, but it is able to bind FGFR2 IIIc. The most specific receptor is thought to be FGFR IIIb which can only bind to FGF 1 and 9. (Ornitz *et al.*, 1996). The expression of the FGFR's is thought to be cell specific. This alongside the added degree of diversity in both the FGF's and FGFR's allows for the generation of a wide range of ligand - receptor combinations. They are produced at different times when needed to perform the many functions in a tightly controlled manner (Powell, 2001).

1.8.4 Signalling pathways of the FGF's

FGF signalling is induced by the binding of specific ligands to HS chains and depends on the size of the chain and the sulphate content of the chain. It has been shown that the FGF can interact with four to five residues on a GAG chain, but most result in an inactive complex (Goodger *et al.*, 2008). The signalling is subject to a range of regulatory inputs; those that are either operative at the level of formation of the complex, or are active inside the cell on the active receptor complex (see HS/FGF/FGFR complexes) (Duchesne *et al.*, 2006). Signal transduction requires the association of both the receptor and the HSPG into a complex, found on the surface of the cell (Pellegrini *et al.*, 2000). The FGF works as part of a dual receptor system, interacting with both the FGFR and HSPG complex enabling it to exert its effects on the cell. It is the FGFR which acts as the signal transducing component and the HSPG plays a secondary part in this dual system (Casu *et al.*, 2002). Here the two receptors combine together and are able to transduce the signal to the interior of the cell (Rieckmann *et al.*, 2007). The interaction here leads to receptor dimerization which then has a knock on effect causing subsequent tyrosine autophosphorylation of the receptor (Pye *et al.*, 2000). This leads to a cascade of activation of the different tyrosine receptors (Stauber *et al.*, 2000), leading to multiple - cellular response. This interaction will depend on the initial cell surface binding of an FGF ligand to a particular FGFR (spliced or unspliced) and would depend on the concentration and the binding kinetics of the FGF (Nugent and Iozzo, 2000). FGF's will bind more favourably to patterns of modulation on the backbone of the HS/hep chain. The FGF - FGFR complex has a higher degree of specificity than either of them alone, this is due to the FGF being able to rapidly bind and release a number of different sequences before activation occurs. This allows for a greater sampling of a wider range of sequences in a chain giving it the ability to bind the correct sequence needed for activation (Harmer 2006),

The eventual signalling response will depend on the level of receptors available, the duration of time the receptor is activated for and also the environment of the cell (Powell, 2001). Any

problems in the signal transduction cascade will inevitably lead to many disorders observed in humans (Stauber *et al.*, 2000), such as human metastatic melanoma (Nikitovic *et al.*, 2008).

1.8.5 The Role of HS in FGF Signalling

The specificity of HS was first identified in early work with the protein antithrombin III (AT III). Here it was identified that NS domains played a crucial role in providing multiple sites for protein ligand binding and activation (Kreuger *et al.*, 2006). Work involving FGF binding and mitogenic assays have aided in backing this theory, with NS domains being identified as being of great importance. Work undertaken by Ornitz (Ornitz, 2000) has shown that highly sulphated regions are significantly more active compared to the unsulphated domains. The role of HS is to aid stability and to localise FGF's in the vicinity of the cell surface (Faham *et al.*, 1996). It is thought of as a co - receptor in that it is able to bring together the FGF and the FGFR (Kreuger *et al.*, 2001; Goodger *et al.*, 2008) allowing for a complex formation (Pellegrini *et al.*, 2000; Duchesne *et al.*, 2006). The interactions between HS, FGF's and FGFR's are looked upon as either highly specific 'lock and key' mechanisms, or a non-specific electrostatic association (Maccarana *et al.*, 1993). Another important part of the interaction between HS and FGF is that the HS chain is able to protect the FGF2 protein from proteolysis (Ornitz, 2000) and protect both FGF1 and FGF2 from thermal denaturation (Zhang *et al.*, 2001).

For HS to bind FGF and/or FGFR interactions between many different molecules must first take place. HS is known for its high negative charge and so cations such as Na⁺ have to bind to minimise this effect. When HS therefore needs to bind to a protein these cations must first be displaced in order for the charged amino acids found on the surface of the protein to bind the HS chain. This process is known as the Polyelectrolyte Effect and can be seen in the following equation:

The complexes were firstly heated to 47 °C. At this temperature the HS:FGFR1 binary complex decreased dramatically. It was then heated to 60 °C and at this point both binary complexes were lost (HS:FGFR1 and HS:FGF1), only the tertiary complex was stable at this temperature. The tertiary structure was finally lost when heated to 100 °C. A dual - receptor system is in place in which the HS:FGF binding will allow or facilitate the binding of FGF2 to its corresponding receptor (FGFR). This suggests that HS acts almost as a template which can bring together FGF2 with its required FGFR (Pye and Gallagher, 1999). FGF signalling is brought about by FGFR dimerization. The formation of the signalling complex can occur in one of two ways:

1. FGF1 is oligomerized by HS which then presents it to the FGFR (Ornitz, 2000; Pellegrini *et al.*, 2000)
2. HS acts almost like a bridge between FGF and FGFR bringing the whole complex together (Guimond *et al.*, 1993).

The above theory would agree with HS being a membrane associated co - receptor, coordinating the interaction between FGF and FGFR's (Goodger *et al.*, 2008). Here a 2:1 HS:FGF1 complex was favoured; even with excess HS present. FGF2 has been proposed to have two receptor binding sites, which suggests it is able to bind either two receptors or the same receptor in two different places (Nugent and Iozzo, 2000). This would allow it to bind both the receptor and HS together; however a high level of FGF is able to activate FGFR in the absence of HS. Here a FGF:FGFR complex can be formed but is not a stable complex formation. It does however result in partial activation of the complex (Ornitz, 2000). The addition of HS into the complex gives greater stability and allows for further activation of the complex.

An active dimer has been identified by Digabriele *et al* (Digabriele *et al.*, 1998) which involve two FGF2 molecules interacting with one heparin chain. The FGF molecules cluster along opposite sides of the HS chain and the distance between the two binding sites is approximately 40 Å.

1.8.6.1 Complex Ratios

A symmetrical dimer consisting of 2 X 1:1 FGF:FGFR complex is able to form; however it has already been identified and has shown to be quite unstable (Ornitz, 2000). Stabilisation of this comes from receptor - receptor interactions between the FGF of one complex and the FGFR of the other complex (Schlessinger *et al.*, 2000). Domain II of the receptor contains a positively charged pocket or canyon in which HS/hep chains are able to fit, therefore providing a crosslink which connects the complex together (Zhang *et al.*, 2001). A 1:1 complex can involve two different ways of binding (Figure 1.17):

1. Ligand and receptor both interaction with HS/Hep.
2. Ligand is in contact with HS/hep but the receptor has no contact with the chain (Ornitz, 2000; Pellegrini *et al.*, 2000)

The interactions between FGF, HS and FGFR can interact in three different ways:

- Cis
- Trans
- Mix – contains both cis and trans interactions.

Figure 1.17

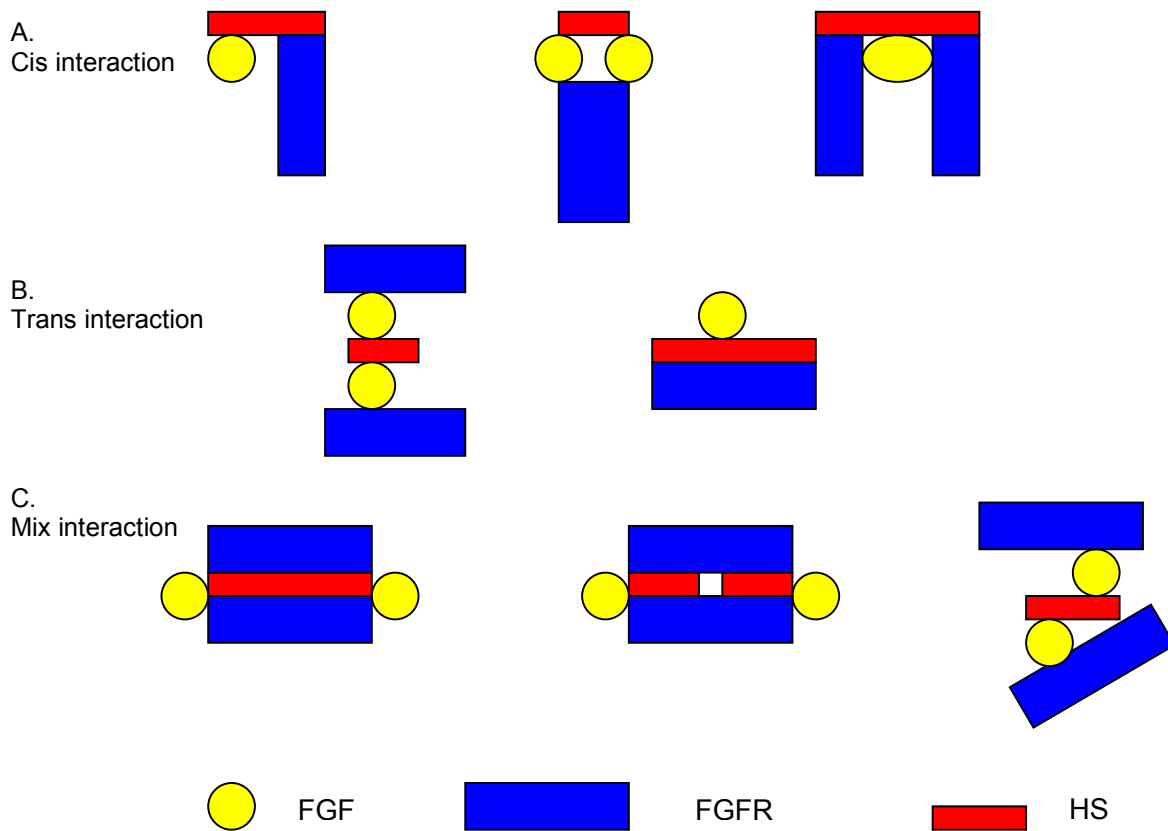


Figure 1.17 Interactions of HS with FGF and FGFR.

HS can interact with FGF and FGFR in a variety of different ways, in a cis, trans or a mixed interaction. This figure has been adapted from Wu *et al* (Wu *et al.*, 2003).

Work carried out by Wu *et al* (Wu *et al.*, 2003) has shown the smallest HS oligosaccharide (dp6) can interact with two FGF1's in a trans interaction. A HS dp12 is the minimum size to form a 1:1 FGF:FGFR in a cis interaction and for the mix interaction, a HS dp16 can fully span a 2:2 FGF1:FGFR2 complex.

The complexes thought to exist in nature are thought to contain a combination of different sized HS/hep chains and different ratios of FGF and FGFR. The complexes identified are:

- 2 x FGF2 + HS + 2 x FGFR
- FGF + HS + FGFR
- FGF + HS + 2 x FGFR (Pye and Gallagher, 1999)
- FGF2 tetramers implicated as an active complex such as cytokines (Interferon γ and PF4) (Moy *et al.*, 1997).

1.8.6.2 The Symmetrical Complex

X-ray crystallography data used to confirm or dismiss these structures has been used and the first structure identified was the symmetrical complex (one of the more well known complexes) (Figure 1.18) (Schlessinger *et al.*, 2000). It involves a 2:2:2 complex and the crystal structure identified a binding cleft in which it was first thought that a hep dp10 would be able to facilitate the complex. However not one, but two hep dp10 chains were identified in the X-ray crystal structure. Here the non-reducing ends were found in the centre of the canyon, with the remainder of the chain then interacting with the high affinity binding sites. The stability of this complex is aided by the hydrogen bonds, which stabilize the configuration of the heparin chains.

Figure 1.18

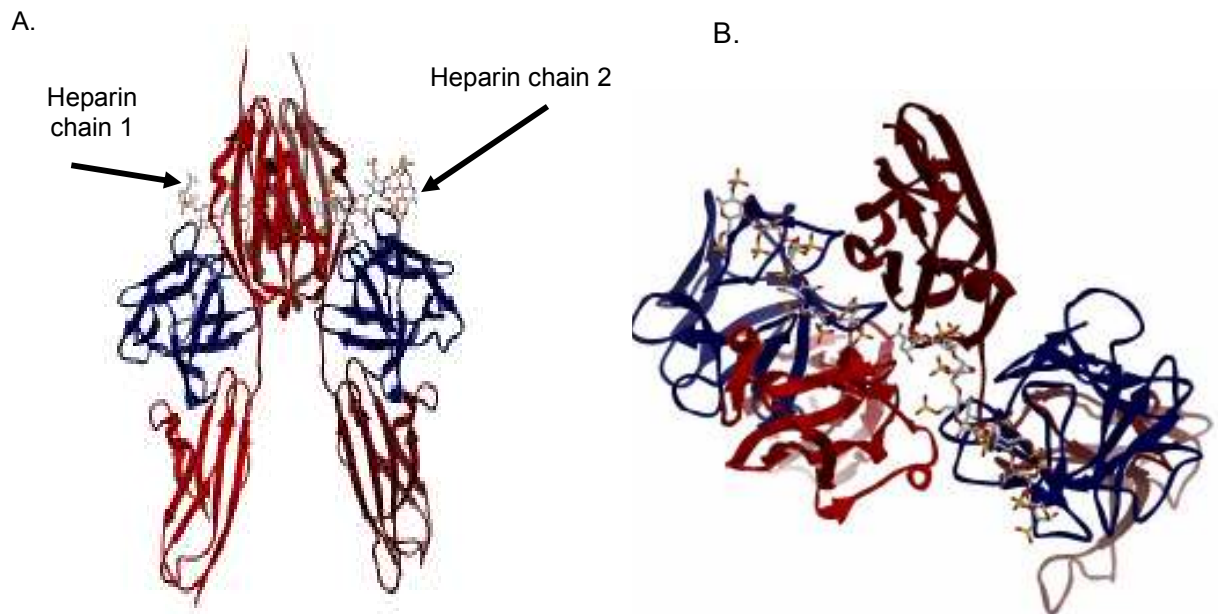


Figure 1.18 The X-ray crystal structure of a symmetrical complex of a 2:2:2 FGF2:FGFR1:heparin ternary complex.

The co-ordinates for this structure were taken from the protein data bank file 1FQ9 (Schlessinger *et al.*, 2000). The FGF molecules are shown in dark blue and the FGFR's are red and dark red respectively.

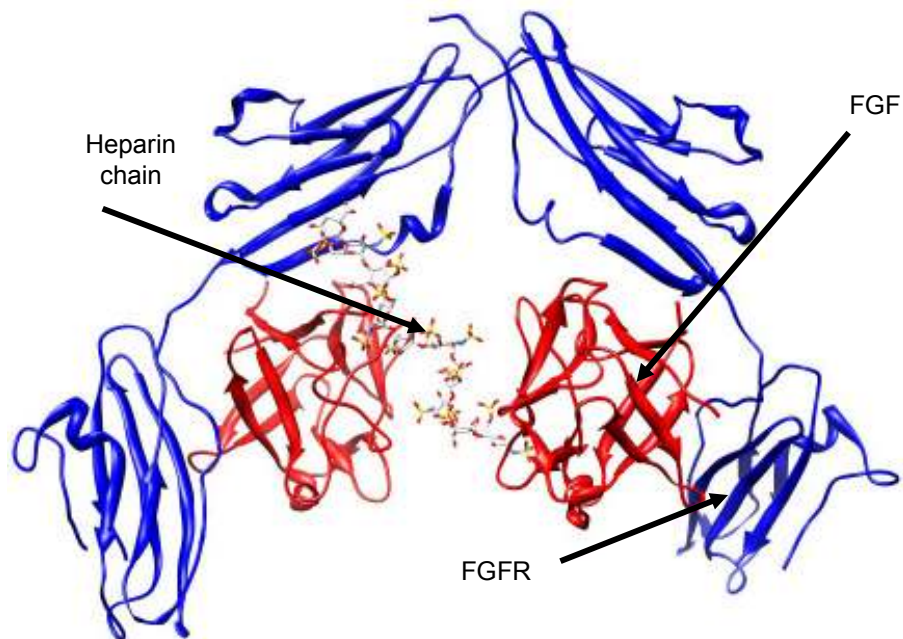
It is not immediately obvious that there are two heparin chains from Panel A. This becomes more apparent in Panel B, where the two chains can be seen running along the top of the complex.

In the crystal structure the GlcNS(6S) was found in the chair form and the IdoUA was found in both the chair conformation and the skew boat conformation. The symmetrical complex is shown to have extensive protein - protein contact and so is thought to be the more stable of the two complexes. This complex is thought to be responsible for mitogenic activity (Goodger *et al.*, 2008). The FGFR plays a role in determining which heparin sequence binds to the complex, with the length of the chain and the sulphation pattern required to elicit a response. The formation and separation of the complex is rapid and as a result the FGF is able to sample many sequences before it has the opportunity to bind to the FGFR (Harmer, 2006); therefore binding many sequences until it ultimately identifies the correct sequence needed to elicit a response.

1.8.6.3 The Asymmetrical Complex

The other well known model is the FGF receptor ectodomain bound to both FGF and heparin. This is known as the asymmetrical complex, where a single chain is involved and can be assembled by itself internally. There is no contact between the FGF and the FGFR so they are reliant on the joining of the hep chain to bring the complex together (Goodger *et al.*, 2008). It identified FGFR2, FGF1 and a hep decasaccharide in a 2:2:1 ratio (Figure 1.19). This involved human FGFR2 IIIc:FGF1:hep decasaccharide and was confirmed by the presence of a single peak using SEC (Pellegrini *et al.*, 2000).

Figure 1.19

**Figure 1.19 The FGF1:FGFR2:heparin complex.**

The coordinates were taken from the protein data bank, file 1E0O (Pellegrini *et al.*, 2000). The FGF molecules are shown in red, the FGFR in blue with the one HS chain spanning the whole complex. A number of electrostatic interactions will be evident between the various negatively charged groups on the HS chain. These will involve the carboxyl and sulphate groups on the chain and the positively charged amino acids present in the FGF and FGFR molecules, which are exposed at the surface of the proteins.

1.8.7 Interaction of HS with FGF1 and FGF2

The two most studied members of the FGF family are FGF1 known also as aFGF and FGF2 known as bFGF. They are both expressed by vascular smooth muscle cells, neurons, fibroblasts and epithelial cells. FGF2 is also released by macrophages in response to inflammation and by platelets in response to blood vessel injury (Goodger *et al.*, 2008). FGF activation is quite specific and is dependant on certain sequences in the HS chain. It is particularly dependant on sufficient chain length, sulphation pattern and the more recently observed phenomenon; the configuration of individual residues in the chain (Guglieri *et al.*, 2008). The involvement of HS is therefore due in some respect to the biological modifications it has undergone during biosynthesis (Wu *et al.*, 2003). The interaction of HS with FGF's is essential for the many signalling events to occur (Harmer, 2006) and the FGF's are implicated in development of almost every major structure and tissue type in the body (Harmer *et al.*, 2004).

1.8.7.1 FGF1

FGF1 is thought to be the universal ligand in the FGF family in that it is able to bind with high affinity to all of the FGFR's. The FGFR's are classed as the signalling receptors so this interaction is very important. However, it is thought to bind to HSPG's with relatively low affinity (Pye *et al.*, 2000). FGF1 is very much similar in terms of structure to all the FGF family (DiGabriele *et al.*, 1998; Harmer, 2006).

It is a small, compact, non - glycosylated protein which adopts a β -trefoil structure or fold (Pellegrini *et al.*, 2000). This consists of 12 anti-parallel β - sheets with a 3 - fold internal symmetry. The β - sheets form a barrel like structure at one end with the use of the amino and carboxyl terminal strands. Inside; the barrel is hydrophobic, whilst the outside is rich in amino acids (Olsen *et al.*, 2003). These charged amino acids play a major role in the interaction with the

sulphate (SO_3^{2-}) and the carboxyl (COO^-) groups found on the HS chain (Bitomsky and Wade, 1999). Docking experiments by Digabriele *et al* (Digabriele *et al.*, 1998) have shown the interaction of FGF1 on specific heparin oligosaccharides and was able to identify a specific binding area within the FGF, especially for binding to heparin (Figure 1.20). The binding site involves amino acids 112 – 128, with 112 – 113 showing the highest probable interaction and was clearly shown when binding FGF1 with a hexasaccharide. Results from the docking with monosaccharides and disaccharides also suggested the importance of this region with the same sequences able to dock with a fully active dimer of FGF promoters. Interactions with each monomer linked back to the binding seen in the decasaccharide with the specific charged groups, with no protein - protein contact observed (Bitomsky and Wade, 1999).

Figure 1.20

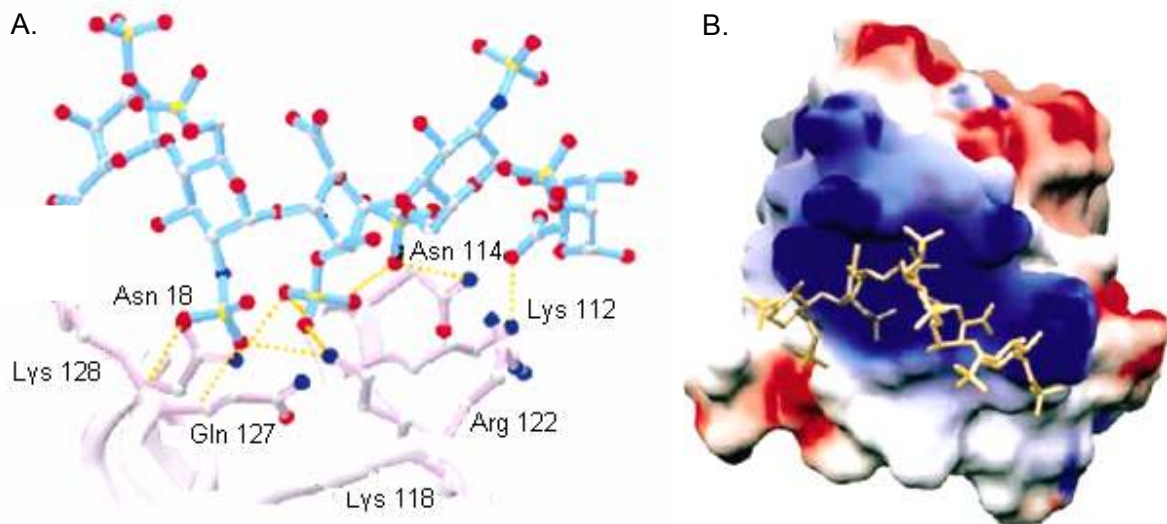


Figure 1. 20 The specific binding interaction between FGF1 and a heparin hexasaccharide.

The specific role of the charged amino acids in the binding site can be seen in Panel A. The binding site involves amino acids 112 – 128 with 112 and 113 showing the highest probable interaction. This figure has been adapted from Digabriele *et al* (Digabriele *et al.*, 1998).

Panel A. The binding site involved in heparin binding. Atoms are colour coded (C = white; O = red; S = yellow; N = blue; with the interactions shown by the yellow dashed line.

Panel B. The electrostatic potential; mapped onto the surface of an FGF1 promoter. The dark blue area is representative of positive electrostatic potential.

As shown by dynamic light scattering, gel filtration and sedimentation velocity ultracentrifugation, the heparin decasaccharide was able to dimerize FGF1 in solution. The complex was also able to promote tyrosine autophosphorylation, activating the receptor and inducing neurite outgrowth in living cells (DiGabriele *et al.*, 1998).

The binding site was also confirmed by work carried out by Pellegrini *et al.* (Pellegrini *et al.*, 2000), who were able to identify residues Asn 118, Lys 112 and 113 (identified as the area with the highest probability of binding a heparin hexasaccharide), Asn 114, Lys 118, Arg 119 and 122, Gln 127 and Lys 128 in the binding site. GAG - protein consensus sequences are known to be rich in lysine, arginine and histidine which all contain polar side chains (Hileman *et al.*, 1998). Lysine and arginine make up a majority proportion of the residues identified in the binding sites mentioned above. It is also interesting to note that arginine is known for its high affinity towards HS; whereas lysine shows the opposite effect and has been known to decrease or lower overall affinity (DiGabriele, Lax *et al.*, 1998). The fact that both of these are present in the binding site contribute towards an intermediate effect (Hileman *et al.*, 1998). Asn 18 has also been identified in the binding of FGF1. All of the mentioned residues make ionic contact with sulphate (SO_3^{2-}) and the carboxyl (COO^-) groups on the HS/heparin chain (DiGabriele *et al.*, 1998).

1.8.7.2 HS interacting groups with FGF1

HSPG's interact with FGF's and the FGF signalling pathway (Guglieri *et al.*, 2008), but the specificity is somewhat unknown. When bound in a complex with HS, it is the interaction of the polar groups from the FGF and the NS groups from the HS chain which are most important. Interestingly the protein - protein interactions (FGF-FGFR interactions) are not thought to be as important, even when HS/heparin bridges the complex (DiGabriele *et al.*, 1998). FGF1 is known to bind with low affinity to HS via the NS domains (Pye *et al.*, 2000). FGF activation has been shown to be dependant on the sulphate pattern; as well as sufficient chain length and the

conformation of specific residues (Guglieri *et al.*, 2008). The sequences identified as important in this binding are:

IdoUA(2S) – GlcNS(6S)

(DiGabriele *et al.*, 1998)

Work carried out by Wu *et al* (Wu *et al.*, 2003) identified 6S to be critical in the formation of a complex involving both FGF and FGFR. 3S was also found to help in binding FGF's, but not in the tertiary complex. HPLC results, using FGF1 and HS, showed the binding of FGF1 occurred with a NaCl concentration greater than 0.5 M; with those sequences containing a greater amount of sulphate groups than those binding with less than 0.5 M NaCl (Kreuger *et al.*, 2001). Experiments on selectively de - sulphated heparin showed there is a requirement for IdoUA(2S) – GlcNS(6S), for binding to FGF1 (Maccarana *et al.*, 1993), which agrees with the work carried out by Digabriele *et al* (Digabriele *et al.*, 1998). This tri - sulphated disaccharide cannot be substituted by any other O-sulphate group. This was shown by comparing two groups with the same amount of 6S groups present, found in different positions. It showed a varying amount of affinity amongst the two groups. By increasing the 6S content the affinity was found to increase (Kreuger *et al.*, 2001). The effect of O-sulphation on FGF1 was also carried out by Jastrebova *et al.*, (Jastrebova *et al.*, 2006); here the amount of both 2S and 6S were investigated by affinity chromatography. This was used to identify the differences in the binding of both FGF1 and FGFR. Results are as follows with the amount of 2S groups increasing stepwise:

2 X 2S groups + variable 6S groups = produced a moderate complex involving FGF1 and FGFR1c and 2c.

3 - 4 X 2S groups + variable 6S groups = produced an increased amount of the complex with FGFR1b and 1c.

8 - 10 X 2S groups + variable 6S groups = formed a strong complex formation with FGFR4.

The complexes were observed and assessed by affinity chromatography. The results indicated that as the amount of NaCl increases, the number of 6S groups found on the chain also increases, therefore the binding increases. This agrees with work carried out by Kreuger *et al* (Kreuger *et al.*, 2001). They were able to show that the amount of 6S groups increased from zero groups up to 4 6S groups as the concentration of NaCl increased. Alongside this the affinity of the complex increased.

1.8.7.3 FGF2

FGF2, like FGF1, has the same common β - trefoil structure and shares approximately 55% homology (Olsen *et al.*, 2003). It is the most studied member of the FGF family because it is involved in a wide range of biological processes including embryonic development; especially in the lungs and limbs of embryos. Here it is the primary inducer of mesoderm formation in embryogenesis (Zhang *et al.*, 2001), is involved in cell growth and tissue repair. FGF2 is a potent angiogenic stimulator which puts it in the forefront for targeted drug design. Angiogenesis is the well known process by which new blood vessels are formed from pre - existing ones. This is a normal process found in wound repair and myocardial infarction; but is also the fundamental step in tumour growth and metastasis. It is an essential process in the transition of a tumour from its dormant to a malignant state (Folkman, 2006). FGF's are secreted by tumours (known as pro - angiogenic molecules) in an autocrine or a paracrine manner. This is used to determine many cancers by their increasing amount of FGF present (Goodger *et al.*, 2008). Human metastatic melanoma has been found to express high levels of FGF2 (Nikitovic *et al.*, 2008). FGF2 is found in an inactive form in normal tissue but has been found to be activated upon injury, inflammation or in tumour invasion (Casu *et al.*, 2002). Interestingly the inactivation of FGF2 or the target of FGF2 with antibodies can lead to the inhibition of cells (Nikitovic *et al.*, 2008).

FGF2 also known as bFGF belongs to the HBGF family along with FGF1 as well as at least two oncogenes, int-2 from stomach tumour cells and also hst/ks from Kaposi sarcoma cells (FGF4 is

also known as Kaposi sarcoma FGF) (Ishihara, 1994). Like FGF1 it is composed of 12 anti-parallel β -strands, each of which is H - bonded to β - strands adjacent to the primary sequence. This forms a barrel structure which is closed by amino and carboxy strands. The core is inaccessible to any solvent and is found to be packed with hydrophobic and aromatic chains (Zhang *et al.*, 1991). It was initially thought that HS binding to FGF was to protect from proteolysis and thermal denaturation. It is also thought to act as a reservoir for growth factors, these being released by enzymes to degrade (Zhang *et al.*, 2001). FGF2 and HS bind either at the surface of the cell or in the extracellular matrix, providing convenient storage for the FGF's (Maccarana *et al.*, 1993). It comprises a 15 KDa protein which is the proteolytic product of an 18 KDa protein and has 90% homology across all species (Nugent and Iozzo, 2000). The activity of FGF2 can be lost completely if binding to HS is hampered in any way. Without the presence of HS, binding would not occur (Guimond *et al.*, 1993). Therefore cells lacking HS, or even unsulphated HS, will not respond to FGF unless this is added into the equation (Maccarana *et al.*, 1993).

1.8.7.4 HS interacting groups with FGF2

The binding site for heparin has been identified as residues 27, 28, 101, 102, 103 and 120 - 137 (Faham *et al.*, 1996) (Figure 1.21). These are similar to the binding residues identified for the FGF1:HS interaction (see pg 79). The binding of heparin to FGF2 is mediated by the same negatively charged groups found on the HS chain (Schlessinger *et al.*, 2000). Further work on the X-ray crystal structure identified 25 H - bonds between FGF and heparin which provided further information on the binding site between these two molecules:

Heparin binding site (- Hydrogen bonds) = Asn 27 found on the β 1 - 2 loop.

= Arg 120 and Thr 121 on the β 9 - 10 loop.

= Lys 125, 129, Gln 134, Lys 135 and Ala 136 found on
the β 11 - 12 loop

(Schlessinger *et al.*, 2000)

All other interactions occur via S - mediated interactions between heparin and FGF with NS and 2S groups. X-ray crystallography structures show the importance of a sequence repeat, (also identified in FGF1); the trisulphated:

IdoUA(2S) – GlcNS(6S)

(Faham *et al.*, 1996).

These are particularly important when considering the sulphate groups found in the identified sequence, and also the amino acid side chains involved in the binding of the protein (Pye *et al.*, 1998). The minimum binding sequence consists of NS and 2S groups found on continuous GlcN - IdoUA sequences. The binding site forms a distorted anti-parallel β - turn on the surface of the protein. Also identified were Tyr 114 and 115 which are thought to be a strong binding determinant and contour maps showed a rather prominent centre of positive charge corresponding to a cluster of lysine amino acids: 119, 120, 125, 129 and 135 (Zhang *et al.* 1991). For a HS chain to be able to bind to more than one molecule a long chain structure is required with the same or similar sequence (Casu *et al.*, 2002). DiGabriele *et al.* (DiGabriele *et al.*, 1998) also identified the same binding sequence but without the attached 6S group:

IdoUA(2S) – GlcNS

(DiGabriele *et al.*, 1998; Harmer, 2006).

These results do not agree on how important the 6S group is in binding. However; the involvement of the 6S group has been identified in the stability of the tertiary complex formation. The presence of the sulphate groups on the NS residues gives the tertiary complex stability (Jastrebova *et al.*, 2006). FGF2 binding is thought to be capable with only one type of O-sulphation, identified as 2-O-sulphation. This is identified in an epitope containing:

IdoUA(2S) - GlcNS - IdoUA(2S)

which FGF2 interacts with strongly. This was identified in HS obtained from pig intestinal mucosa (Kreuger *et al.*, 2005).

Figure 1.21

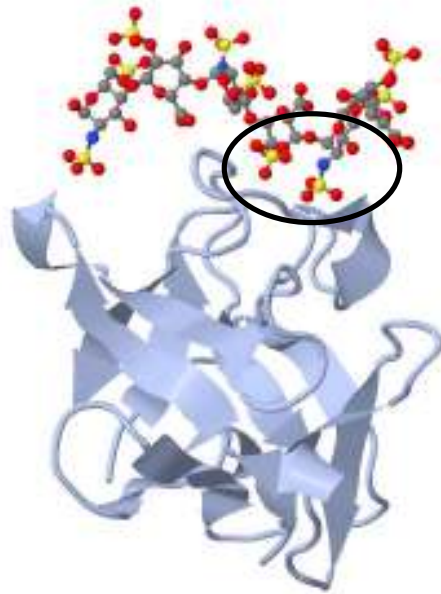


Figure 1.21 FGF2 binding site with a heparin hexasaccharide fragment.

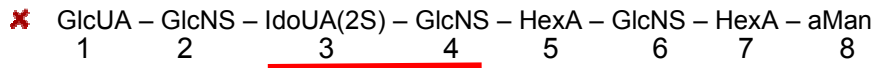
The binding site identified in FGF2, with the heparin hexasaccharide fragment binding to the molecule. The highlighted section shows the identified binding site between the FGF and the heparin chain (see Preliminary results of 1BFC and 1FQ9). This figure has been adapted from (Faham *et al.*, 1996).

1.8.8 The importance of specific groups on the HS chain and chain length – interactions for both FGF1 and FGF2

The binding of HS/heparin chains to FGF has been shown to be highly rich in IdoUA(2S) residues, particularly in binding FGF2. The affinity increases with the addition of more IdoUA(2S) in the chain and also with an increase in chain length (Pye *et al.*, 1998). Binding studies have pinpointed the importance of a single IdoUA(2S) which was determined by investigating bound and unbound fractions of HS digested with Hep I and nitrous acid, alongside investigating the minimum sized chain able to bind the FGF's. Hep I has been shown to cleave sequences between GlcNS-IdoUA(2S) (Murphy *et al.*, 2004), thus allowing for the confirmation of the presence of this important IdoUA(2S) residue. The smallest and least sulphated oligosaccharides capable of binding FGF2 were identified as octasaccharides of which contained essentially all GlcNS with a single IdoUA(2S) residue (Maccarana *et al.*, 1993). The position of this IdoUA(2S) was thought to play a major role in what would determine some sequences to be active when bound to FGF and others to remain inactive but bound (or unbound altogether). The following sequences show the possible importance of the IdoUA(2S) group (Figure 1.22).

Figure 1.22

A. HS dp8



B. HS dp8

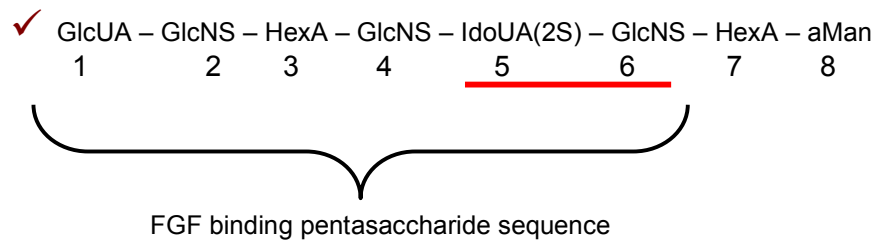


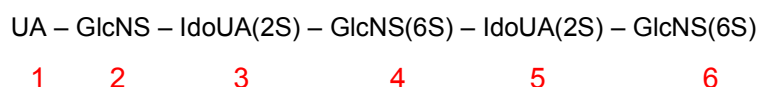
Figure 1.22 An example of the difference between active and inactive sequences due to the position of a single IdoUA(2S) residue.

This figure has been adapted from (Maccarana *et al.*, 1993) and shows that a change in position of only one sulphate group can have an effect on the biological activity of the sequence.

Both these sequences are made up of the same residues but with the IdoUA(2S) found in a different position on the chain. Firstly it is found in position 3 in the inactive sequence (A) and in position 5 of the active sequence (B). The FGF binding pentasaccharide sequence is also identified in the above structure (B). These sequences show the 2-sulphate group as being of great importance in the binding of FGF2 to HS/heparin chains. Studies have also shown that the presence of any 6S group is not a requirement here (Schlessinger *et al.*, 2000), as clearly none are seen in the above structure (Maccarana *et al.*, 1993).

The pentasaccharide binding sequence proposed by Maccarana *et al.* (Maccarana *et al.*, 1993) is thought to be the minimum length of saccharide needed for binding FGF2 but longer is needed for biological activity. The 2S and 6S groups identified on IdoUA(2S)-GlcNS ± (6S) (the identified binding sequence needed for FGF binding (Faham *et al.*, 1996; DiGabriele *et al.*, 1998)), has been shown to neither promote nor impede the interaction with growth factors (Guimond *et al.*, 1993). This could be due to the size of the oligosaccharides used in this study.

An X-ray crystal structure was used to study the length of chain required, but was also used to locate the IdoUA(2S) residues in the chain (Faham *et al.*, 1996). Resulting from this, it was found that heparin itself was too heterogeneous to crystallise and therefore it was decided to use tetrasaccharide and hexasaccharide sized fragments were used in replacement of the heparin chain. The structures were determined and found to be:



Residues 1-4 make up the tetrasaccharide sequence. Interestingly the high affinity sites identified on the surface of FGF2 had a common interaction with residues 2 and 3 on the heparin chain above. This corresponds to the GlcNS-IdoUA(2S) residues both in the tetrasaccharide sequence and the hexasaccharide sequence.

Further down the chain a low affinity site was identified which was occupied by rings 5 and 6, residues IdoUA(2S)-GlcNS(6S). This result is interesting in that the proposed low affinity binding site involves the previously mentioned residues which make up the high affinity site. Also worth nothing are the different conformations identified for the IdoUA(2S) residues in the chain (Figure 1.23).

It is thought that the binding of FGF2 to heparin can occur when both conformations identified above are present (Faham *et al.*, 1996). Binding information regarding the sequence mentioned above can be found in Table 1.2; with information on the high and low affinity binding sites identified when bound to FGF2.

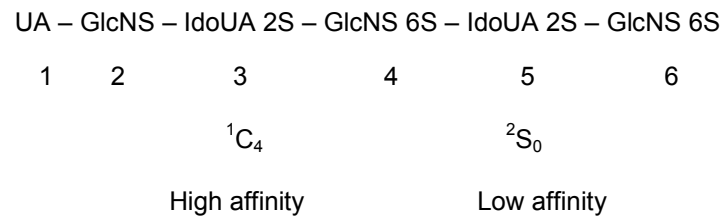
Figure 1.23

Figure 1.23 An example of a heparin sequence with both high and low affinity binding sites.

Here the high affinity site has been identified in the ¹C₄ conformation, with the low affinity binding site in the ²S₀ conformation. This is thought to be the optimum binding conformation required for all active oligosaccharides.

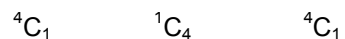
Table 1.2

Residue	Affinity	Interactions with HS	Conformation of IdoUA 2S
Lys 27	Low	5 and 6	2S_0
Asn 28	High	2 and 3	1C_4
Asn 102	Low	5 and 6	2S_0
Arg 121	High	2 and 3	1C_4
Lys 126	High	2 and 3	1C_4
Gln 135	High	2 and 3	1C_4
Lys 136	Low	5 and 6	2S_0

Table 1. 2 High and low affinity binding sites in FGF2.

The main interacting amino acids in the FGF2 molecule and their interaction with specific residues in the heparin chain. This figure has been adapted from (Faham *et al.*, 1996).

Ultimately it is still unknown exactly which conformation or combinations of conformations are required to have a lasting effect on the binding of FGF's and FGFR's. Hileman *et al* (Hileman *et al.*, 1998) have shown that an FGF bound to a hexasaccharide has a tight interaction with one IdoUA(2S) found in the 2S_0 conformation. The other IdoUA(2S) is thought to adopt a weaker interaction and therefore stays in the 1C_4 conformation, unable to flip to the 'active' conformation. This result contradicts results shown in above (Table 1.2), (Faham *et al.* 1996). It has been proven many times that the IdoUA \pm (2S) are a flexible ring structure and can adopt multiple conformations (see Structural Conformation). When a protein such as an FGF molecule interacts with HS/hep, a kink is introduced into the structure. This is known as the kink motif and is identified as the following structure:



The GlcNS groups are found in the fixed ${}^4\text{C}_1$ conformation with the IdoUA(2S) in the ${}^1\text{C}_4$ conformation. Here the kink motif allows for maximum contact with the FGF molecule. However; interesting is the presence of the 6S group (Guglieri *et al.*, 2008) in the kink motif. Up until recently the importance of the 6S group had been dismissed, but the presence of it in this sequence alongside the conformation of the IdoUA \pm (2S) may play an important relationship after all.

The most studied structure is one which contains both the FGF2 binding site (Figure 1.24) and the kink motif, identified above. The IdoUA(2S) plays an important role in both the kink motif and in the binding site of FGF2 and is found in a ${}^1\text{C}_4$ conformation.

Figure 1.24

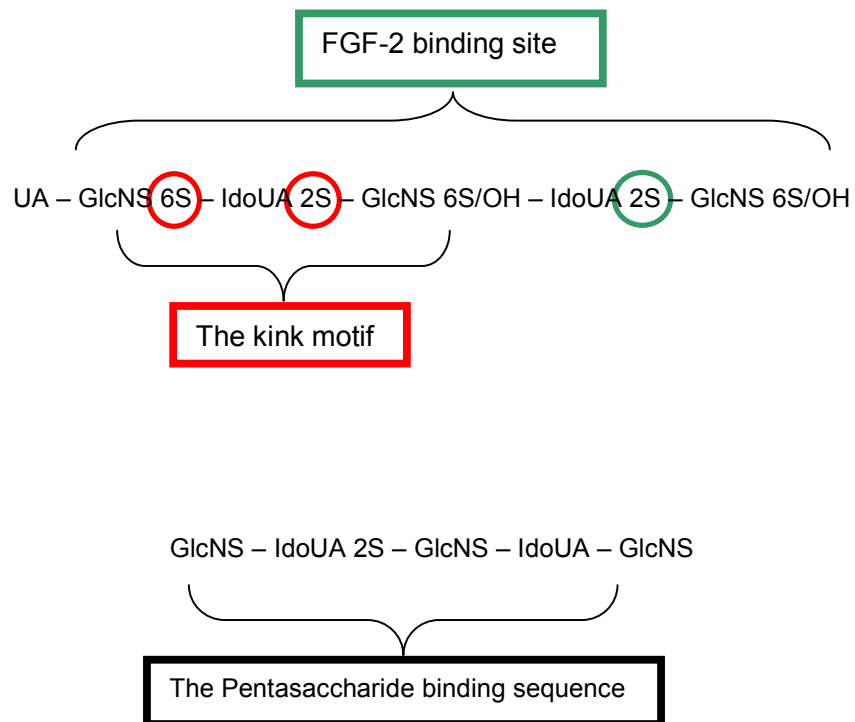


Figure 1.24 HS sequences identified with binding properties

The FGF2 binding motif is highlighted in green, with the Kink motif highlighted in red. The highlighted sulphate groups are identified as playing an important role in the interaction with protein ligands. Also shown is the minimum pentasaccharide binding sequence, which is thought to be the minimum sequence able to activate ligand activity.

A single IdoUA(2S) has been identified which is required for appreciable binding to FGF2 and any additional 2S groups will increase further binding (Kreuger *et al.*, 2001) The minimum pentasaccharide sequence identified contains an IdoUA(2S) in position five which had been previously identified as important for binding, (Figure 1.24, shown in green), and the IdoUA(2S) in position three was thought increase affinity, alongside the 6S group upstream of the iduronic acid residue (Figure 1.24, shown in red). A HS decasaccharide sequence was found to contain a minimum of three IdoUA(2S)-GlcNS repeats substituted with one or two 6-O-sulphate groups increases the biological activity of the FGF2: HS complex (Guimond *et al.*, 1993; Pye *et al.*, 2000).

1.9 Modelling of HS dp10

Work undertaken by Goodger (Goodger, 2003) on HS dp10's showed an inability of certain sequences to activate FGF2 when bound with FGFR1 IIIc. However a slight change in the sequence, such as an extra sulphate group on the chain, or a change in conformation then allowed for some of these sequences to become active when bound to the same complex. This shows that a certain amount of specificity is required when dealing with active and inactive sequences. The following work will look at the issue of specificity and which sequences are essential for promoting the ability of FGF2 with FGFR1 IIIc to bind HS and become active. This was initially undertaken using a cell line devoid of HS, namely BaF-32 cells. The sequences shown (Figure 1.26) are taken from work done by Goodger (Goodger, 2003) who was able to show that some fully sequenced HS dp10 sequences were active in the presence of HS and others were not (Figure 1.25).

These sequences were investigated by using molecular modelling software – AMBER 9; and molecular docking software – Autodock; to try and identify the essential components in the HS chain which allow for some sequences to bind and activate the complex; whilst other HS

sequences are unable to do this. This involved identifying specific sulphate groups in the chain, extracting information for torsion angles (torsional data) and finally conformational changes in the ring structures (from Cremer-Pople puckering parameters).

It should be noted that once sequenced the residue at position 9 in all four sequences was not determined (sequences 1, 3, 5 and 7). Therefore models will be created using one of two different residues at this position, either a GlcUA or an IdoUA.

Figure 1.25

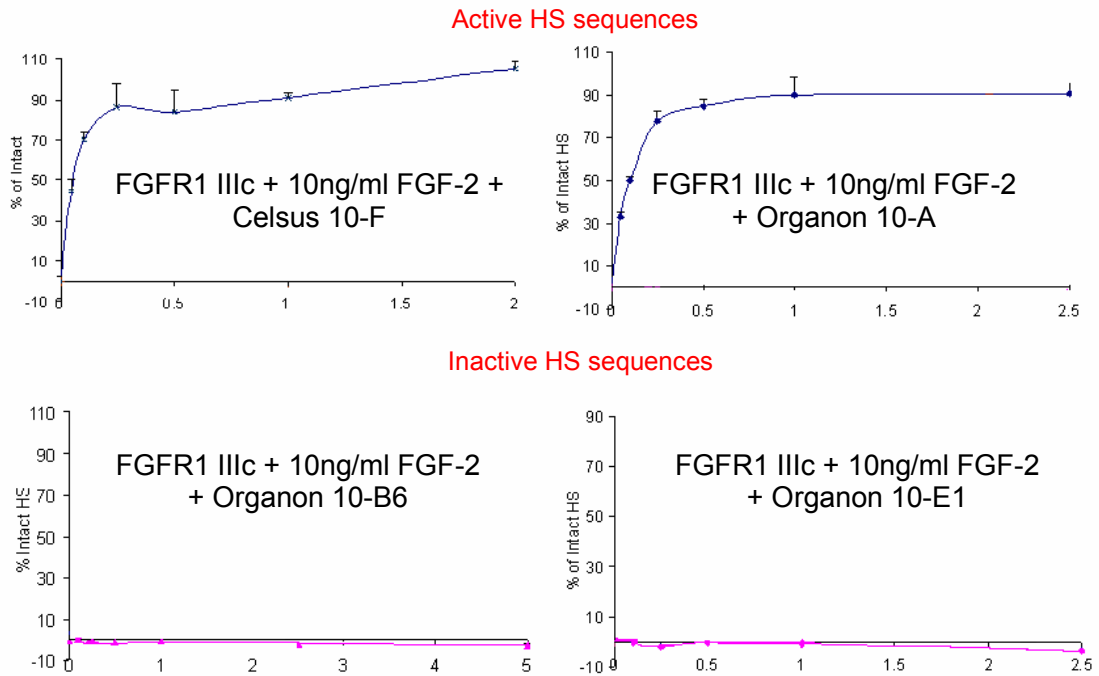


Figure 1.25 HS sequences of active and inactive in BaF-32 cells transfected with FGFR1 IIIc.

Work carried out by Goodger (Goodger, 2003) identified two active and two inactive HS sequences using BaF-32 cells devoid of HS and transfected with FGFR IIIc.

Figure 1.26 HS dp10 sequences.

The Goodger isolated HS sequences isolated from a number of HS sources; namely Celsus HS and Oregon HS. These were purified as in the procedure described (chemical digestion) using nitrous acid.

Panel A. The sequences of the identified Goodger oligosaccharides, with additional sequences where the residue could not be determined (sequences 2, 4, 6 and 8).

Panel B. HS deca-saccharide structures of active and inactive sequences; with FGF2 and FGFR1 IIIc. Sequences 1 and 2 are Celsus 10-F. Sequences 3 and 4 are Org 10-A. Sequences 5 and 6 are Org 10-B6. Finally sequences 7 and 8 are Org E1.

Figure 1.26 A.

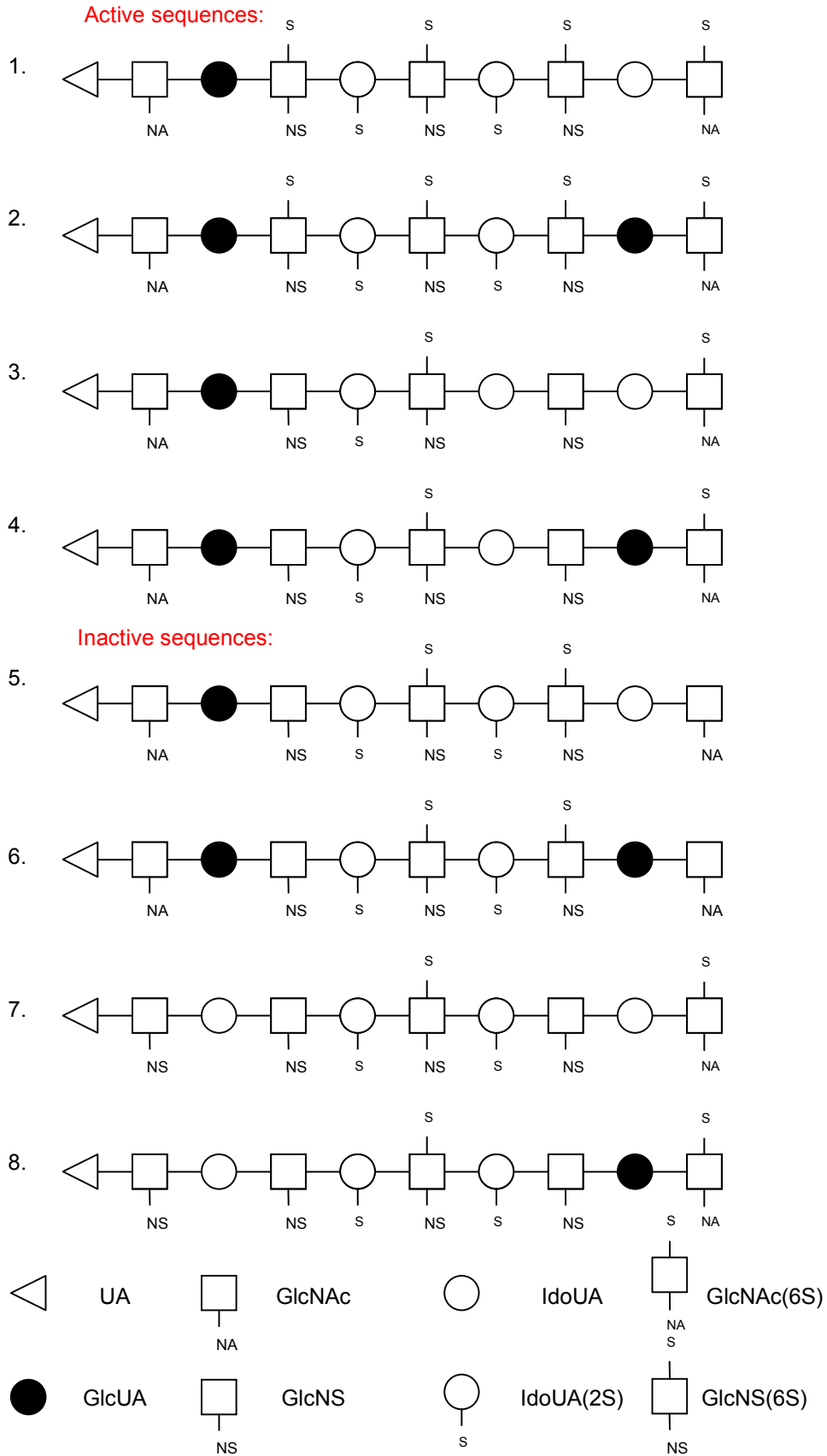
Active Sequences:

1. UA - GlcNAc - GlcUA - GlcNS(6S) - IdoUA(2S) - GlcNS(6S) - IdoUA(2S) - GlcNS(6S) - IdoUA - GlcNAc(6S)
2. UA - GlcNAc - GlcUA - GlcNS(6S) - IdoUA(2S) - GlcNS(6S) - IdoUA(2S) - GlcNS(6S) - GlcUA - GlcNAc(6S)
3. UA - GlcNAc - GlcUA - GlcNS - IdoUA(2S) - GlcNS(6S) - IdoUA - GlcNS - IdoUA - GlcNAc(6S)
4. UA - GlcNAc - GlcUA - GlcNS - IdoUA(2S) - GlcNS(6S) - IdoUA - GlcNS - GlcUA - GlcNAc(6S)

Inactive Sequences:

5. UA - GlcNAc - GlcUA - GlcNS - IdoUA(2S) - GlcNS(6S) - IdoUA(2S) - GlcNS(6S) - IdoUA - GlcNAc
6. UA - GlcNAc - GlcUA - GlcNS - IdoUA(2S) - GlcNS(6S) - IdoUA(2S) - GlcNS(6S) - GlcUA - GlcNAc
7. UA - GlcNS - IdoUA - GlcNS - IdoUA(2S) - GlcNS(6S) - IdoUA(2S) - GlcNS - IdoUA - GlcNAc(6S)
8. UA - GlcNS - IdoUA - GlcNS - IdoUA(2S) - GlcNS(6S) - IdoUA(2S) - GlcNS - GlcUA - GlcNAc(6S)

Figure 1.26 B.



1.9.1 Cremer-Pople puckering parameters.

The ring conformation is a mathematical measurement of the pyranose ring conformation determined by Cremer and Pople (Cremer and Pople 1975; Cremer, 1984). The values are given based on the order that they are selected within the ring system (Forster and Mulloy, 1993). For the purpose of this study they were designated in the order O5-C1-C2-C3-C4-C5 and therefore should be comparable to the majority of the previous work undertaken. The only exception being the work of Regazzi *et al* (Regazzi *et al.*, 1986; Regazi *et al.*, 1993) where the chosen order was C1-C2-C3-C4-C5-O5. The number of the rings was first established by Cremer and Pople (Cremer and Pople, 1974); where the ring oxygen is assigned atom 1 with the carbon linked to the glycosidic bond assigned as atom number 2. Note this numbering is different to those assigned when picking the Cremer-Pople puckering parameters. This is in accordance with IUPAC rules in numbering ring structures.

All types of ring puckering can be mapped on the surface of a globe, with the poles at $\theta = 0^\circ$ or 180° , corresponding to the canonical chair conformations (Haasnoot, 1992).

1.10 Molecular Modelling

1.10.1 AMBER

The name AMBER refers to a suite of programmes which allows the user to perform molecular dynamic simulations on biomolecules. It consists of two parts, AMBER Tools and AMBER 10, which was released as the latest version in April 2008. The information flow of AMBER involves a number of different programmes; the ones used in this report are highlighted (Figure 1.27).

Leap: the primary programme which allows the building of a new system, or to modify an old system.

Antechamber: is responsible for assigning partial charges to the model, assigning atom types when creating a new system, recognizing atom types, generating residue topology files, finding missing parameter force fields and supplying reasonable substitutes.

Sander: is the basic energy minimiser and molecular dynamics programme. This is able to relax the structure by moving the atoms down an energy gradient until a low energy minima has been obtained.

Ptraj: a general purpose programme for analysing and processing trajectory and coordinate files created from molecular dynamic simulations.

Figure 1.27

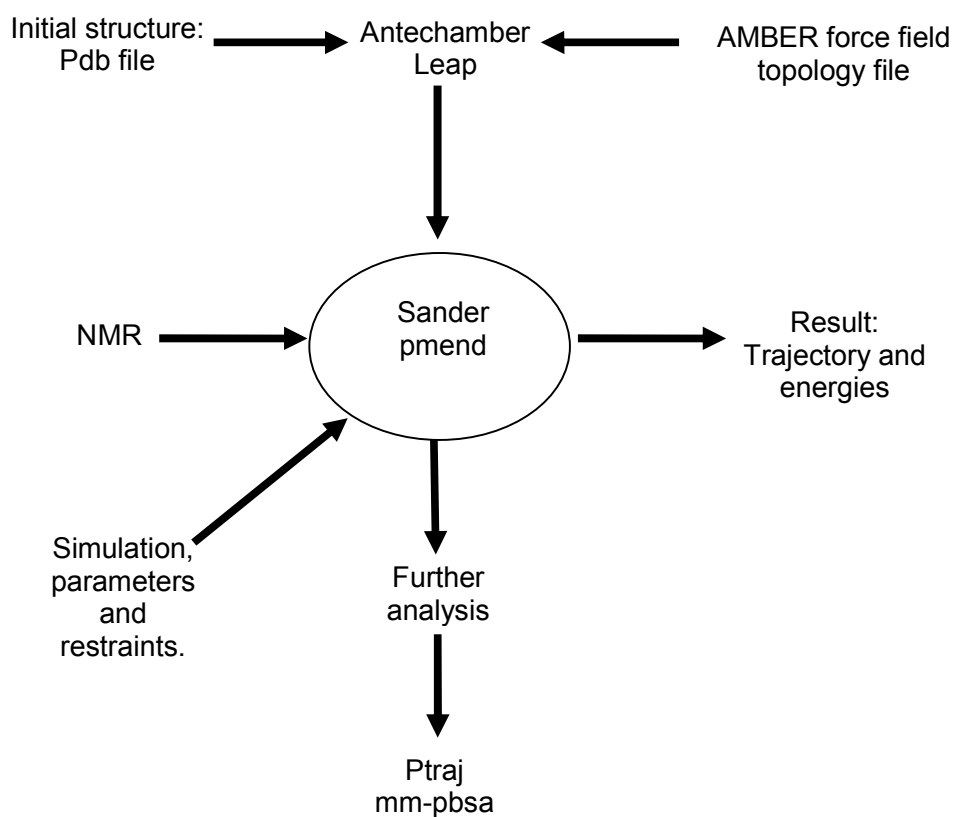


Figure 1.27 Information flow in the AMBER suite of programmes

The flow of information using AMBER 9, which was released from The Scripps Research Institute in March 2006.

1.10.1.1 Torsion angles Φ and Ψ vs. Conformational Energy

To be able to define the orientation of each disaccharide pair found in the deca-saccharide structure, torsion angles play an important role. The torsion angles are used to assign the stereochemistry of a glycosidic linkage. In the case of heparan sulphate, a 1 - 4 linkage occurs joining one ring system to the next. The Φ (phi) torsion angle is determined as H1, C1, O4, C4 in literature (Mulloy *et al.*, 1994) which determines the angle from the non-reducing side monosaccharide (Figure 1.28, Panel A).

The other angle is known as the Ψ (psi) torsion angle. This is defined as C1, O4, C4, and H4 in literature (Mulloy *et al.*, 1994) (Figure 1.28, Panel B).

Figure 1.28

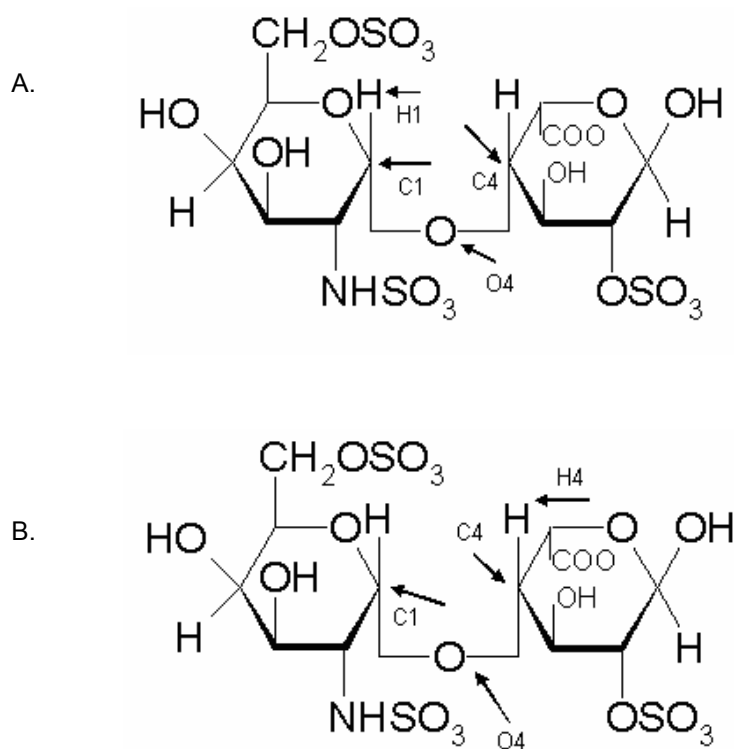


Figure 1.28 Definition of the phi and psi torsion angle

Panel A. The phi torsion angle is defined as H1, C1, O4, and C4.

Panel B. The psi torsion angle is defined as C1, O4, C4 and H4.

Each disaccharide pair within an oligosaccharide will have different torsion angles and the lowest energy conformations will be sought and used to build static models of the oligosaccharides (see Experimental Materials and Procedures). These values can be either positive or negative depending on the position of the lowest energy minima in the energy maps.

1.10.1.2 The Exo and Endo - Anomeric Effect

The exo-anomeric effect is characteristic of ring substituents to reside in the usually energetically favoured position; the equatorial orientation. This is favoured due to steric interactions from the surrounding environment. is not favoured and the axial orientation becomes the preferred position (Figure 1.29).

The opposite of this is the endo - anomeric effect and is characterised by the preference of specific groups, such as the acetyl groups to reside in the axial position, as opposed to the equatorial position (identified in pyranose ring structures). This is due to the lone pairs on the ring oxygen interacting with the non-bonded electrons attached to the O5-C1 bond causing them to move to different rotamer positions. The endo-anomeric effect is thought to give general preference for gauche conformations (Krawczuk, 2005). It is the exo-anomeric effect which is thought to cause a barrier that reduces free rotation around the Φ and Ψ torsion angles.

According to this theory; all the C1 substituents found within HS will have a preference for the axial orientation. This includes the glucuronic acid and N- glucosamine residues which are technically classed as β C1-C4 but only through naming nomenclature (the C1-C4 bonds being classed as β C1-C4). The C1 substituents are actually positioned axially as are all the HS internal units (Murphy, 2007). The energetically 'unfavoured' axial orientation is favoured over the equatorial orientation shown by the larger arrow pointing towards the axial orientation (Figure 1.29, Panel A).

Figure 1.29

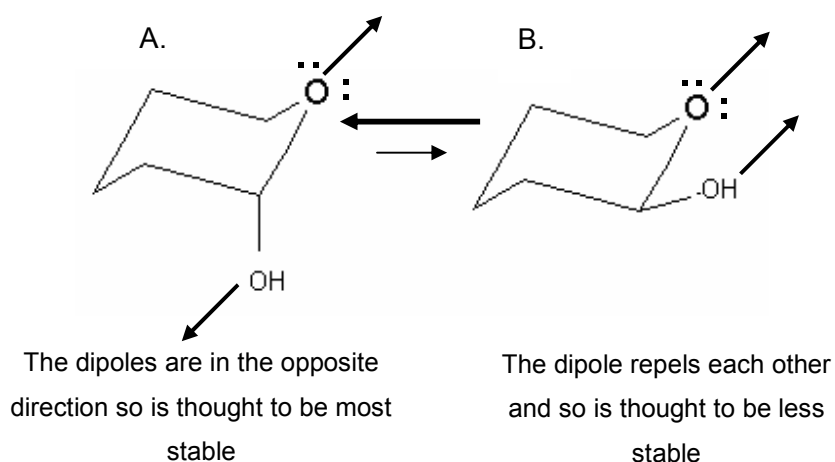


Figure 1.29 The Exo- and endo-anomeric effect

The anomeric effect, also known as the endo-anomeric effect is the preference for certain substituents bonded to the anomeric carbon to reside in the axial position. The energetically favoured position (the equatorial position) is unfavoured due to steric interactions.

Panel A. The axial orientation (α orientation).

Panel B. The equatorial orientation (β orientation) is the 'energetically' preferred position and is known as the exo-anomeric effect.

1.10.1.3 Preparing the model in explicit solvent

All molecular dynamic simulations in this study are run in explicit solvent with periodic boundaries (the effect of the solvent is not included in the calculations). Firstly, the structure needs to be free from all charge. This is achieved by the addition of Na⁺ which is added at positions of high negative electric potential. These electronegative groups have been identified as:

1. The carboxyl group on UA.
2. The N-sulphate on GlcNS.
3. The 2-O-sulphate group on IdoUA(2S).
4. The carboxyl group on IdoUA.
5. The N-acetyl group on GlcNAc.

These are the positions in which the xleap programme will add neutralising counter ions to neutralise the negative charge on the molecule (Walker, 2006).

A truncated octahedral box of pre - equilibrated water is then added into which the structure will be encased. This water is known as TIP3P water because of the three bonds each individual water molecule contains and is made up of triangulated water models which have an extra bond in the structure due to the high bending frequency of the water molecule (Figure 1.30) with no charge on the extra bond.

Figure 1.30

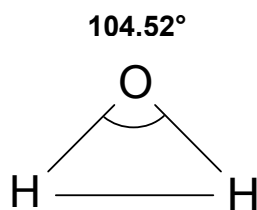


Figure 1.30 A simple water model showing the TIP3P water model

The simulated water used in all molecular models carried out in this study use the TIP3P water model. It is a triangulated water model bonded on all three sides due to the high bending frequency identified in the molecule. This makes the water molecule more effective in terms of computational time.

TIP3P belongs to the simplest water model group as it is rigid and will only rely on non-bonded interactions. It is labelled as TIP3P because of the number of interaction sites available on the molecule, these corresponding to the three atoms of the water molecule. Coulomb's law is used to model electrostatic interactions and the Lennard - Jones potential is used for dispersion and repulsion forces, but this is applied only to the interaction between the oxygen atoms in each of the models (Walker, 2006). TIP3P water is the most common water model used because of its simplicity and its computational efficiency.

1.10.1.4 The use of a Periodic Boundary

A periodic boundary is used to make sure there are no 'holes' between the solvent, the carbohydrate and the solvent and the edges of the box. These gaps can lead to vacuum bubbles forming in the system and lead to the instability during the simulation (Steinbrecher, 2006). The boundary keeps the pressure constant and so allows the volume of the box to change, thus allowing the water to equilibrate around the carbohydrate and attain an equilibrium density. All particles being simulated are enclosed in a specific sized box and are then replicated in all three dimensions which give a periodic array. A 2D array can be seen (Figure 1.31) where movement is replicated across all the boxes available.

The box in grey shows the central box, which houses the main carbohydrate structure and contains all the water molecules. All the other boxes around it are a replica of this central box. As a particle moves out of the box, as shown by the arrow in the central box (in this case particle 1), another particle is able to enter from the box below, therefore being replaced by another atom. This makes the number of particles in the central box constant and this occurs in all boxes (Walker, 2006).

Figure 1.31

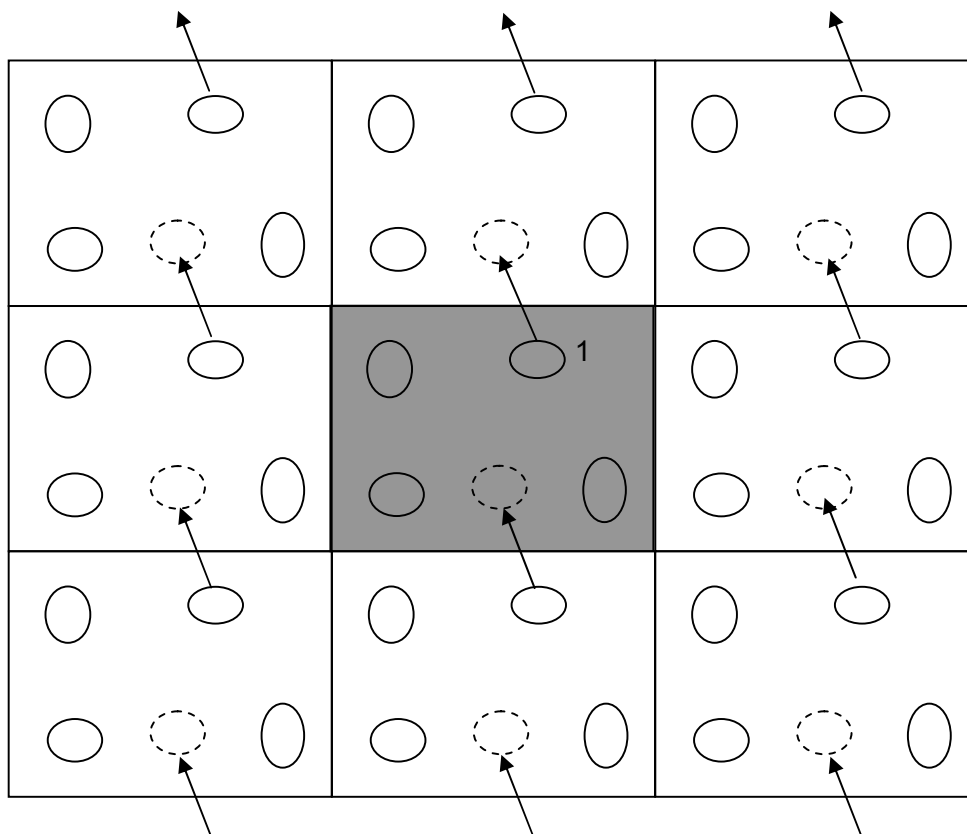


Figure 1.31 An example of a periodic boundary with the box coloured in grey housing the carbohydrate to undergo MD simulation.

A suitable cut off point is needed so that when a particle leaves one side of the box the same image is not seen twice. This can be done by making sure the size of the box is more than twice the cut - off distance, therefore a particle will find it impossible to interact with two images of the same particle simultaneously. For all simulations carried out, the cut - off size has been determined to be 6 Å. The carbohydrate structure from here is now ready to go through minimisation and molecular dynamics (see Molecular Modelling in Experimental Materials and Procedures, pg 117) with any experimental analysis carried out after this process.

Chapter 2: Experimental Materials and Procedures

2.1 Experimental Materials:

Porcine intestinal mucosal heparan sulphate was purchased from Celcus laboratories (Cincinnati, OH, USA). Heparinase I (*Flavobacterium heparinum*; EC 4.2.2.7), Heparinase II (*F. heparinum* no EC number assigned) and Heparinase III (*F. heparinum* EC 4.2.2.7) were all purchased from Grampian enzymes, (Orkney, Scotland, UK). Bio-gel P-10 (fine grade gel) was purchased from Bio - Rad (Hemel Hempstead, UK). PD - 10 prepacked disposable Sephadex G - 25 columns were purchased from GE Healthcare UK Ltd (Buckinghamshire, UK). The ProPac PA - 1 semi - preparative strong anion exchange HPLC column was from Dionex (Camberley, Surrey, UK). The TSK3000PW gel filtration column was from Phenomenex (Macclesfield, Surrey, UK). HPLC grade sodium chloride and calcium acetate were purchased from VWR International (Leicestershire, UK). Ammonium bicarbonate and sodium acetate were from Sigma Aldrich (Dorset, UK). Disodium phosphate was purchased from DH Scientific Ltd (Huddersfield, UK). D₂O was purchased from GOSS scientific (Essex, UK) and the Shigemi NMR micro tube was purchased from Sigma Aldrich (Dorset, UK).

2.2 Experimental Procedures:

2.2.1 Enzymatic Digestion of Heparan sulphate

Porcine Intestinal mucosal HS was depolymerised using the enzyme heparinase III. To 10 g of HS 50 ml is added of 0.1 mM calcium acetate, 0.1 mM sodium acetate (heparinase/hep buffer), pH 7.0, and 20 mU of heparinase III was added. Further additions of 20 mU of enzyme were made every 24 hours until digestion was complete (as detailed by analytical gel filtration). In total, 25 identical digests were carried out from the 10 g of HS starting material. All digests were carried out at 25 °C.

2.2.2 Gel Filtration Chromatography Analysis of Heparinase III

Digestion

2.2.2.1 Analytical

Each 10 μ l aliquot containing 200 μ g of HS was applied to a TSK3000PW gel filtration column. The column was pre - equilibrated with PBS (phosphate buffered saline, 0.15 M NaCl, 20 mM Na₂HPO₄, at pH 7.2). The sample was eluted from the column using the same PBS at a flow rate of 0.4 ml/min and this was monitored in - line for absorbance at 232 nm.

2.2.2.2 Preparative

Each 200 mg digest was applied to a Bio-gel P10 column with a column size of 3 x 120 cm. The samples were then eluted from the column with ammonium bicarbonate (NH₄HCO₃) 0.25 M at a flow rate of 9 ml/hr. 3 ml fractions were collected and their absorbance monitored at 232 nm to create a HS digest profile. The relevant fractions were then pooled together and frozen. These samples were then lyophilised and the concentration of NH₄⁺ ions was reduced by repeatedly re - dissolving the samples in 5 ml of H₂O and again lyophilising.

2.2.3 SAX – HPLC (Strong Anion Exchange Chromatography)

Each pooled fraction corresponds to a different sized HS oligosaccharide. The HS octasaccharides were applied to a ProPac PA - 1 column (9 x 250 mm) pre - equilibrated with distilled water adjusted to pH 3.5 with HCl. Elution of the samples was with HPLC grade NaCl set

at a gradient of 0 - 1.2 M NaCl over 90 minutes. The flow rate was set to 4 ml/min. Those gradients were chosen according to Merry *et al* (Merry *et al.*, 1999). The resulting peaks were measured for absorbance at 232 nm and 4 ml fractions were collected. All relevant, identical fractions were pooled together and concentrated initially by rotary evaporation and then by freeze drying (it was then decided that freeze drying was the better of the two methods and so this was used for the rest of the project). The resulting sample was then diluted in H₂O, desalted on a PD10 desalting column and eluted with H₂O. The sample was then concentrated again and lyophilised.

2.2.4 NMR Spectroscopy

All NMR spectra were recorded by Dr. Neil Mclay (Department of Chemical and Biological Sciences, University of Huddersfield).

For NMR measurements; the total quantity of the sample was dissolved in 250 µl of PBS (0.15 M NaCl, 20 mM NaPO₄ at pH 7.2). It was then repeatedly exchanged in 1 ml ²H₂O with lyophilization immediately between these steps. Analysis was carried out using a 5 mm Shigemi tube, with the sample volume being kept at 250 µl throughout. The ¹H - NMR spectra were acquired using a Bruker NMR spectrometer operating at a ¹H frequency of 500 MHz. The reported chemical shifts are in ppm and are relative to the sample reference tetramethylsilane. TOCSY and NOESY spectra were recorded using the phase sensitive mode using states TPPI. 256 experiments of 2048 data points were recorder with 128 scans over a width of 2741 KHz in both dimensions.

2.2.5 Molecular Modelling

All molecular modelling was carried out using the AMBER 9 molecular modelling software package. (Case *et al.*, 2004).

For all molecular models, all carboxyl, sulphate and sulphamate groups were modelled in their de-protonated form and charges were added to them using the antechamber module of AMBER 9 with the AM1 – BCC charges and GAFF atom types assigned.

2.2.5.1 Calculation of the lowest energy minima across glycosidic bonds

All calculations were carried out using the Glycam04 force field for carbohydrates, provided in the AMBER suite of programmes. Extra parameters needed for the C4 to C5 double bond found on the non-reducing end uronic acid were taken from the General Amber Force Field (GAFF).

For each glycosidic linkage conformational energy maps were created as a function of the Φ and Ψ torsion angles. Each torsion angle was varied over the full possible range of angles (-180° to $+180^\circ$) and the non-bonded energy was calculated at 20° intervals with torsional restraint of 10000 Kcal/mol until the RMS derivative for the whole structure was less than 0.01 (Mulloy *et al.*, 1993).

2.2.5.2 Preparation of the model

All simulations were carried out as explicit solvent models where the effect of the solvent is not included in any of the calculations. Each model built was placed in a truncated octahedral box of explicit TIP3P water molecules. This water has not felt the influence of the charges added to the

structure, neither has it felt the influence of the carbohydrate. Na⁺ ions were added to the structure using the addions method in the xleap programme in AMBER to ensure there is no overall net charge on the molecule. An 8 Å buffer of water was generated around each of the models to the edge of the water box in each dimension. Standard molecular modelling techniques were then used to run minimisation and molecular dynamics in explicit solvent.

2.2.5.3 Minimisation 1

The first part of minimisation involves holding the molecule fixed while the positions of the water and the ions are minimised. This is known as “positional restraints” which works by essentially keeping the atoms fixed in the same position, in this case the starting position. These atoms are restrained to conform to this structure via the use of a constant force. Minimisation is turned on and has 1,000 total steps of minimisation, with molecular dynamics switched off. (500 steps of steepest descent minimisation, followed by 500 steps of conjugate gradient minimisation). The volume is kept constant and it has a cut - off of 6 Å.

2.2.5.4 Minimisation 2

This involves minimising the entire system; including the water and the ions with no restraints on the system. (This involves 1,000 steps steepest descent minimisation, 1,500 steps conjugate gradient minimisation). Again constant volume is used and the cut - off is 6 Å.

2.2.5.5 Molecular Dynamics 1

Phase 1 involves a heating phase with restraints on the molecule. The system is allowed to heat up from 0 K to 300 K. A weak restraint is used to ensure there are no fluctuations in the system. The Langevin temperature equilibration system is used in this instance which maintains and equilibrates the system temperature using a collision frequency of 1.0 ps^{-1} . This is kept at a constant volume for 20 ps. The restraints are then switched off and it changes to constant pressure with a periodic boundary (average pressure of 1 atm) in which time the water is allowed to relax during 100 ps of molecular dynamics. Running at constant temperature over 100 ps closely resembles laboratory conditions.

All hydrogen atoms in the system have no effect on large scale dynamics and therefore can be fixed in position. Using the SHAKE algorithm, all bonds involving hydrogen are constrained, thus removing hydrogen motion which has the highest frequency oscillation in the system. This allows for an increase in time steps from 1 fs (10^{-15}) to 2 fs without introducing any instability in the molecular dynamic simulation, meaning more molecular dynamics can be run in a shorter amount of time.

2.2.5.6 Molecular Dynamics 2

The subsequent MD run involves equilibration of the whole system using constant pressure so the density of the water can relax. The temperature is set at 300 K and the restraints are removed. This is run for anywhere from 100 ps to 10 ns in time with 100 ps running a total of 50,000 molecular dynamic steps with a 2 fs time step. SHAKE is again used in this part of the simulation. The system coordinates were output to a trajectory folder every pico - second (for the 100 ps run) resulting in a final trajectory file for each run consisting of 1,000 frames per structure.

2.2.6 Molecular Docking

Firstly; the receptor oligosaccharide was prepared in AutoDockTools (ADT) version 1.5.4, revision 30. It offers a complete visualization, which is required for the setup of each docking process. Here all polar hydrogen's were added (those not already included in the pdb file) and then Gastiegar charges were automatically calculated. The file was then saved as a PDBQT file ready for input into Vina. PDBQT files are the input file type for both Vina and Autodock 4. These include all of the coordinates needed including atomic coordinates, Autodock atom types and all the partial charges needed. A gridbox; using the gridbox function is used to visibly judge the size of the search area (the area where Vina will calculate the lowest energy conformation for the docked ligand onto the protein). This box needs to be of large enough size to contain the entire oligosaccharide and the binding site on the protein, in general a restricted area of the protein (Seeliger and de Groot, 2010). Here the gridbox was set up using the following steps found in the configuration text (conf.txt):

```
receptor = FGFfahnam_nowater.pdbqt
ligand = frame.pdbqt

out = out.pdbqt

center_x = 1.02
center_y = 0.004
center_z = 58

size_x = 62
size_y = 52
size_z = 42

num_modes = 40
```

This reads in the receptor and the ligand as the names given and prints all of the output files to a file called out.pdbqt. The num-modes (number of modes) will tell it to produce a maximum of 40

low energy conformations. The number of modes did not change when the value was increased to 60 or decreased to 20.

Next the ligand was prepared which involved choosing the number of rotatable bonds in the chain and the positions of them. For all the Goodger decasaccharide structures, 46 rotatable bonds were identified but this was initially reduced to zero rotatable bonds, making the structure completely rigid. This makes them computationally fast in terms of computing time as adding in rotatable bonds increases the amount of conformations available in each space. The Vina software will try to calculate the affinity of all the possible conformations. It is only later on in the process that it will then discard these structures as part of the Genetic Algorithm (GA). This is the only type of search which Vina can perform.

The frames chosen for the Vina docking experiments were taken from data obtained from the Puckering-torsion data. Separate frames from both the 1 ns and the 10 ns MD simulations were taken for docking analysis and these were chosen based on changes in the ring conformation and also the torsional geometry. Each individual frame chosen, was set rigid in the as described above. Through this process, many different frames are chosen and docked to the x-ray crystal structure. The aim; to find the correct conditions in terms of ring conformation, sulphate pattern and torsional geometry required for FGF2 binding.

A multi-model file is created with all of the created conformations ranking in order from lowest to highest energy which is based on Vina's unique scoring function (Trott 2009). The docking process uses random seed when performing the search which will bring a degree of randomness to the search and is used in all docking simulations. The likelihood of not finding the correct binding conformation can be reduced by increasing what is known as 'exhaustiveness' and can be manually changed in the computing process. However increasing this will inevitably have an effect on processing time and so a compromise must be made between the processing time and the likelihood of not finding the correct binding conformation.

The binding output file, logs data on the proposed binding modes rating them on the affinity of binding to the known binding site. The higher up the list the proposed model is found, the higher the affinity for the binding site is. Vina also prints a score based on the distance from the best mode. Here it is split into two sections; the first rmsd l.b (root mean square deviation, lower bound value) and the second being rmsd u.b (from the upper bound value). The model proposed as the best fit binding model (i.e. the lowest energy binding model) is rated as number 1 and this has the highest affinity (kcal/mol) and both rmsd l.b and rmsd u.b are rated as 0.000. The mode (model) number from here increases as the affinity for the binding site decreases. The rmsd values fluctuate in value.

Rmsd – matches each of the atoms in one configuration with the closest atom of the same element type in the other conformation.

Rmsd u.b – matches each atom in the one configuration with itself in the other configuration and it ignores any symmetry (Trott 2009)

Chapter 3: Results

3.1 Enzymatic Digestion of Heparan Sulphate

Studies required the production of a library of HS oligosaccharides that was produced by the digestion of Porcine mucosal HS by the enzyme heparinase III. This was compared to work previously carried out by Murphy (Murphy 2007).

A profile for intact HS which had not undergone enzyme digestion (Figure 3.1, Panel A) was used as a comparison with the digested HS samples. This enabled us to fully attain when complete digestion had taken place (Figure 3.1, Panel B). The digestion process was closely monitored to attain this end point. The undigested HS sample showed a broad range of peaks which is indicative of the diverse structural nature of the HS chain (Figure 3.1, Panel A); whereas the digested sample (Figure 3.1, Panel B) showed a shift in the peak elution times and an amplification of the profile, indicating digestion of the HS chain. Further additions of enzyme and longer incubation times by Hep III showed no further degradation with the digest thought to be complete.

Once conditions for complete heparinase digestion had been determined, a larger digest was set up using 10 g of HS in 100 ml of hep III buffer. Over a course of 5 days 20 mU of enzyme was added over 24 hours. The following elution profile shows the enzyme digest of time 0 hours (Figure 3.2, Panel A) and after 120 hours (Figure 3.2, Panel B).

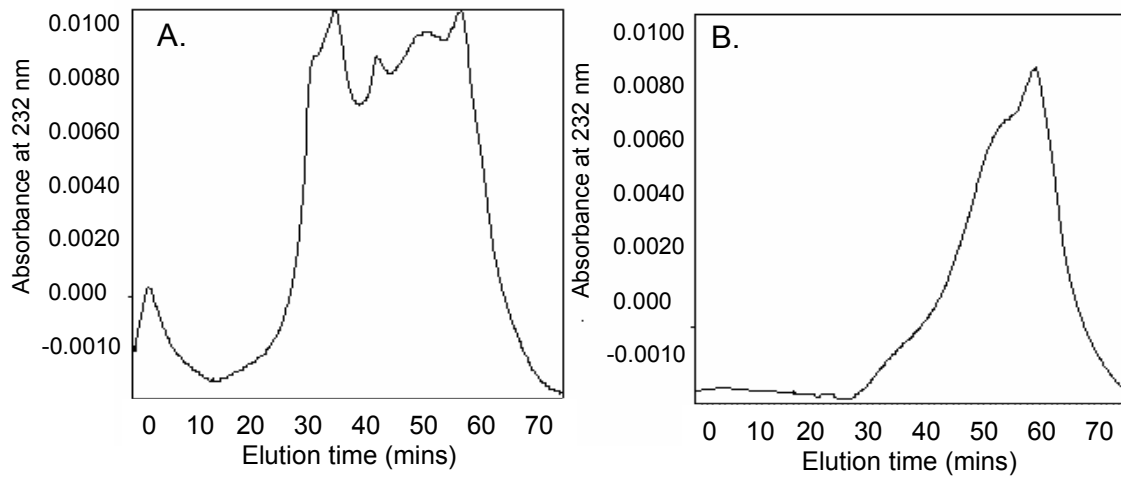
Figure 3.1

Figure 3.1 Analytical TSK3000PW gel filtration profile of HS after 48 h incubation with 10mU Hep III.

Samples of 10 μ g from a total digest of Celsus HS by the heparinase III enzyme were chromatographed on a TSK3000PW analytical gel - filtration column (30 x 0.75 cm).

Panel A. Intact HS profile before the addition of heparinase III.

Panel B. Gel filtration profile after 88 h incubation with 40 mU heparinase III added.

Figure 3.2

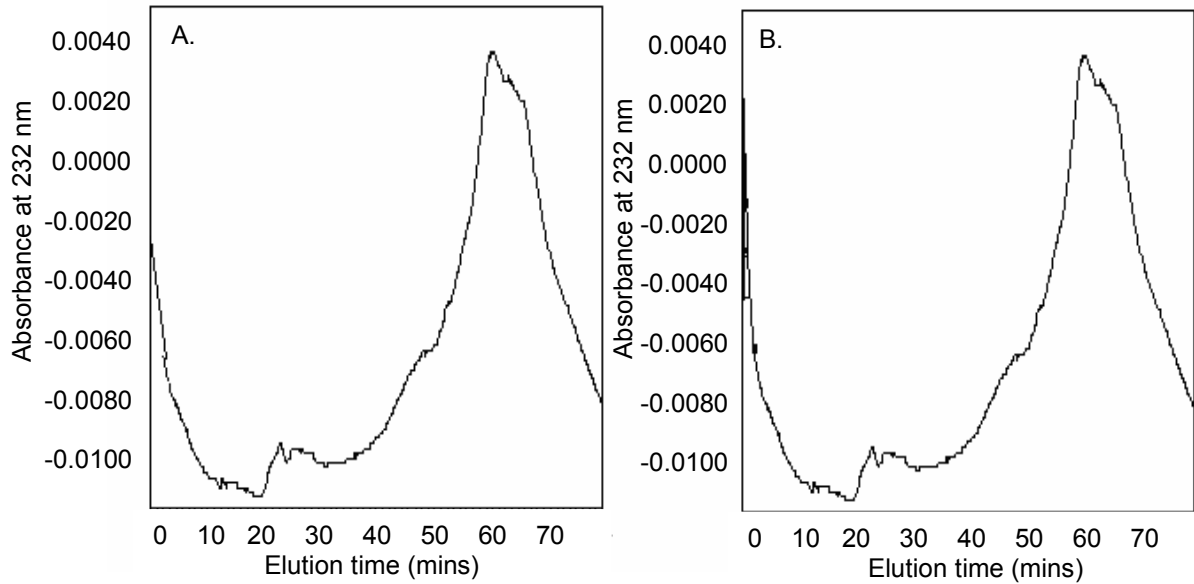


Figure 3.2 Comparison of large sample digest carried out on an analytical TSK3000PW gel filtration profile of HS after 120 h incubation with 10mU Hep III.

Here the two profiles are the same suggesting the enzyme digestion had not gone to completion.

Further enzyme needed to be added.

Panel A. Intact HS profile at 0 hours.

Panel B. After 120 hours and the addition of 100 mU heparinase III.

The two graphs in Figure 3.2 indicate the same digestion profile, suggesting that the enzyme digest has not gone to completion at all. This was thought to be due to the large amount of hep buffer added to the digest. It was decided to freeze dry the samples, and re-suspend in a smaller amount of hep buffer. This time just 30 ml of hep buffer was added to the HS samples and the enzymes were again added in 20 mU aliquots until the digestion went to completion, as determined by analytical gel filtration (Figure 3.3).

These results clearly show that the digestion has gone to completion; which is vital if the HS chain is to be fully digested to the smaller, known oligosaccharide fragments. A poor hep III digestion can lead to problems when identifying the oligosaccharide of choice (dp8) from the size exclusion column. This is because peaks tend to overlap each other making it very difficult to obtain a pure sample. Also confusion in which peak has actually eluted from the column, meaning some of the wrong peaks were collected and pooled together.

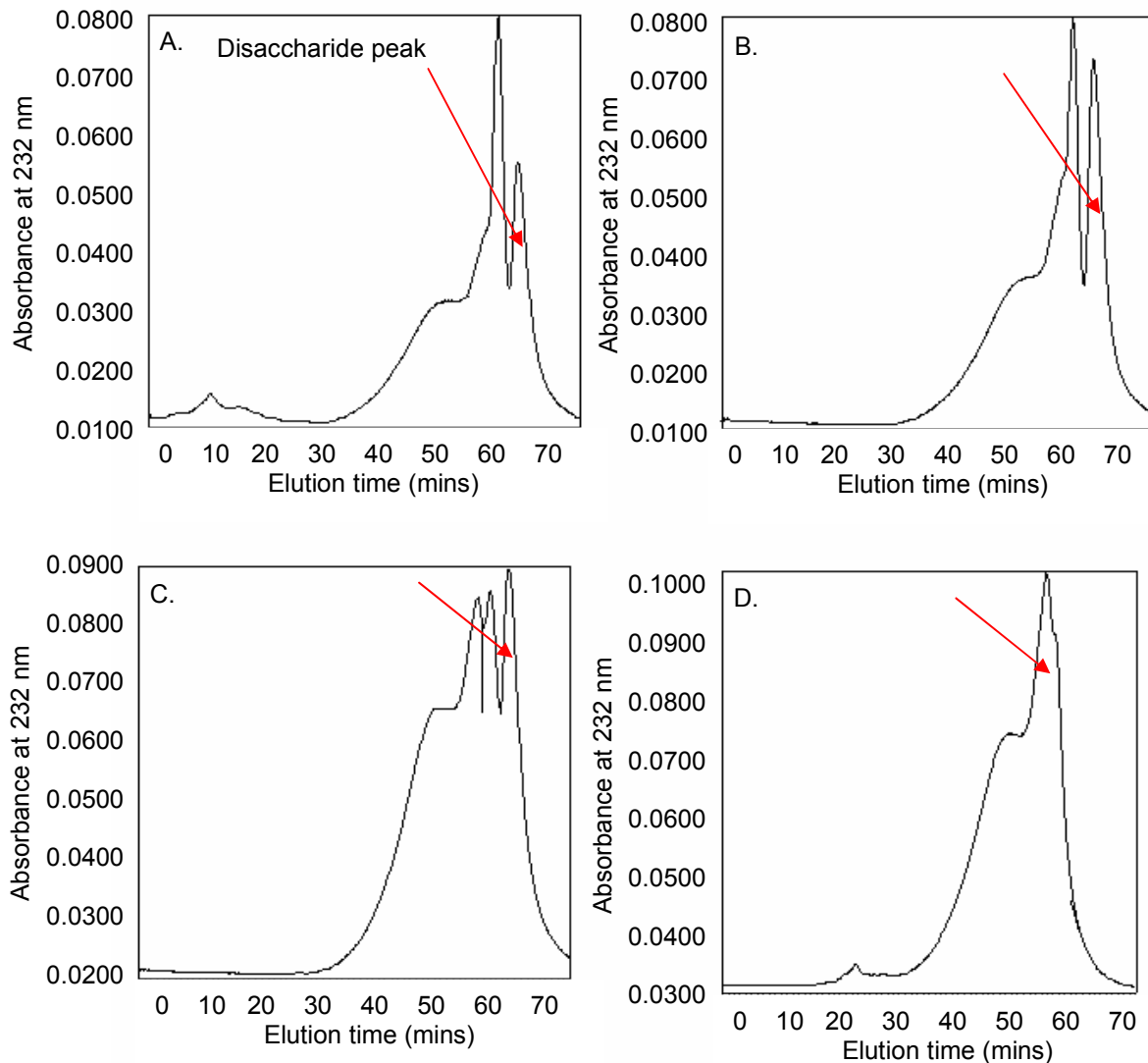
Figure 3.3

Figure 3.3 Comparison of the analytical gel-filtration profiles of HS digested with 140 mU (total) of heparinase III over a 7 day period, re-dissolved in 30 ml hep buffer.

Compare to undigested profile in Figure 3.1, Panel A.

Panel A. 24 hours after the addition of 20 mU of hep III.

Panel B. 48 hours and an additional 20 mU of hep III.

Panel C. 120 hours and a further 60 mU of hep III (20 mU added every 24 hours).

Panel D. 168 hours and an additional 40 mU aliquots of hep III, added every 24 hours.

3.1.2 Preparative Gel Filtration Chromatography of HS Digested to Completion with Heparinase III

Once the enzyme digest has fully gone to completion the HS samples were run on a preparative size exclusion column. Typically 2 ml aliquots of the digested sample was taken from the bulk sample and carefully loaded onto the column (disturbing the gel beads in the column leads to overlapping of peaks). A typical profile is shown in Figure 3.4 (Panels A and B) with the sized oligosaccharides clearly resolved. The two profiles shown are very similar to each other but show that no two HS digests give the same profile. This is indicative of the complex structure of HS oligosaccharides. Disaccharides (dp2) to dodecasaccharides (dp12) were collected and pooled together for further analysis; larger fragments (dp14+) were collected, pooled together and stored at -20 °C.

Figure 3.4 also shows the reproducibility of the separation technique, particularly for oligosaccharides in the range of dp8 - dp2. Larger oligosaccharides were more susceptible to peak overlap due to overloading of the column. However; given this studies' target of isolating octasaccharides, these overloading problems could be ignored and greatly increased the speed by which the digest could be processed. To combat these problems in future studies, a larger column (or two sizeable columns) would mean sample overload would not occur. It also gives an indication of how slight differences in the profile can lead to the collection of the wrong peak, resulting in a sample which is no longer pure.

Figure 3.4

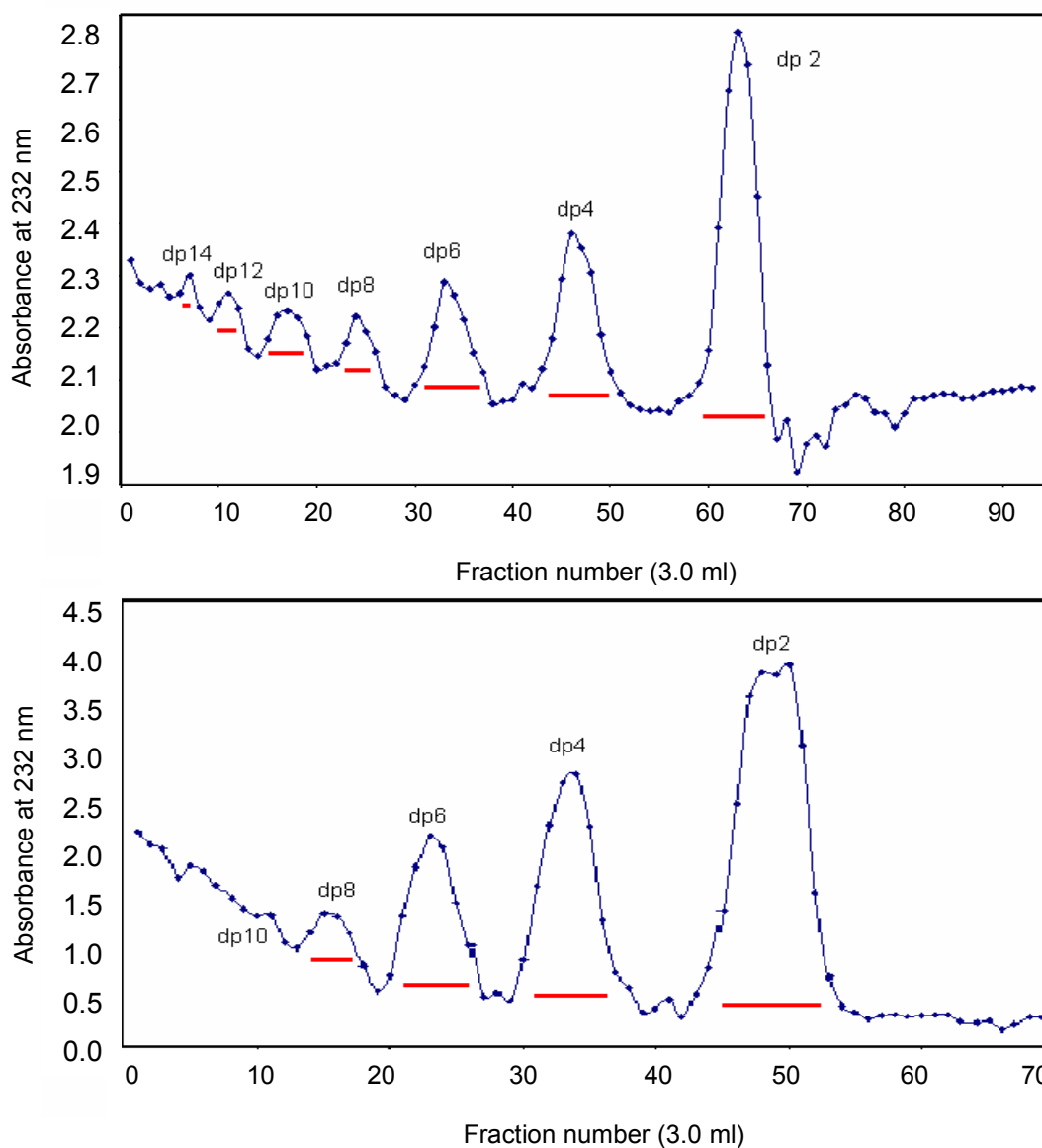


Figure 3. 4 Preparative Bio-Gel P10 gel filtration chromatography profiles of HS digested to completion with heparinase III.

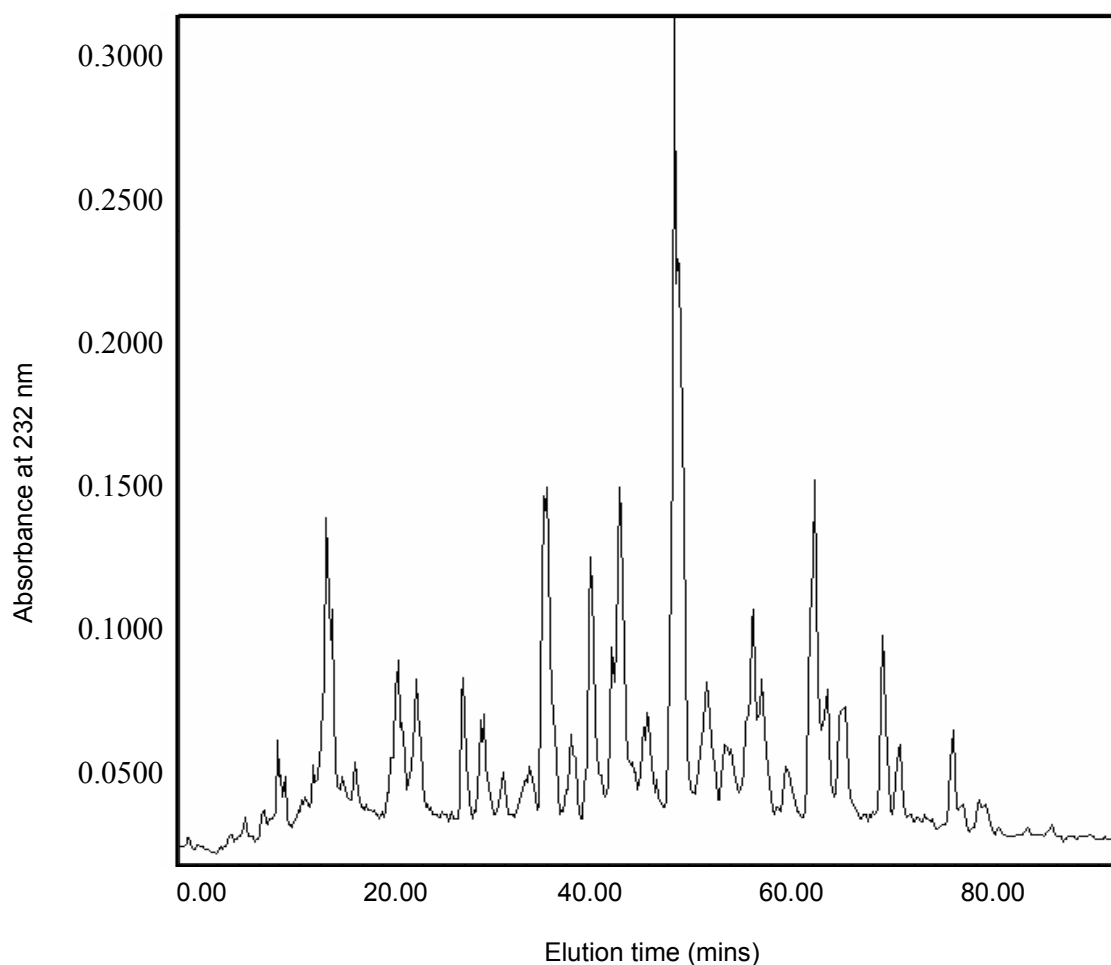
A typical HS gel-filtration profile (Profiles A and B), from a size exclusion column; digested to completion by the addition of 140 mU of heparinase III (which should be far in excess of the amount of enzyme needed). HS was chromatographed on a Bio-Gel P10 gel filtration column set at 9 ml/h, over 72 - 90 hours, (column size 3 x 120 cm) with a buffer of 0.25 M NH_4HCO_3 . Samples were pooled as indicated (red lines on the graph).

3.1.3 SAX-HPLC of Octasaccharide (dp8) sized Oligosaccharides

All octasaccharides samples obtained from the gel-filtration column were applied to a strong anion exchange column to separate on the basis of charge, to identify the number of charged species of oligosaccharide within the defined size group. Size defined oligosaccharide fractions were applied to a Pro-Pac PA-1 SAX-HPLC column (fitted with a 20 µl sample loop), the resulting profiles are shown in the following figures (Figures 3.5 - 3.6).

Previous work undertaken on the smaller dp4 sized oligosaccharides (Hileman *et al.*, 1997) indicated a lack of biological activity and the more recent study on dp6 sized oligosaccharides (Murphy 2007) lead to the analysis of the dp8 sized oligosaccharides. The characterization of larger HS fragments is necessary in building a library of known HS oligosaccharide structures. HS is such a complex polysaccharide; that in reality it would be impossible to purify the billions of potential sequences produced in the modification process of the chain (see Introduction, p32). If NMR and molecular modeling studies enabled the identification of specific patterns in Iduronate conformation, sulphation placement and amount of sulphation; alongside torsional geometries of the glycosidic bonds then this would enable us to successfully predict protein activating sequences. Longer fragments may be one of the fundamental factors in FGF2-HS interactions and so studying the longer oligosaccharides produced by enzymatic digestion could dramatically aid this study. If time constraints were not such a factor, then longer oligosaccharide fragments could also have been studied (ideally dp10 and dp12 sized oligosaccharides).

Many peaks were identified in the SAX-HPLC elution profiles and those large enough to collect were designated peaks 1-19. The number of peaks actually taken on for further purification however was far fewer (5 samples underwent preliminary NMR analysis, with only 1 being partially assigned). This was due to the time constraints of the purification and the requirements for large quantities of oligosaccharide needed for NMR analysis.

Figure 3.5**Figure 3.5 SAX-HPLC chromatography of sized HS oligosaccharides**

A typical SAX-HPLC profile of the dp8 sized oligosaccharides generated by heparinase III digestion, purified by Bio-Gel P10 gel-filtration chromatography. Samples were applied to a Pro-Pac PA1 SAX-HPLC column (250 mm x 9 cm); with a gradient of 0 - 1.2 M NaCl over 90 minutes. Detection was determined by absorbance at 232 nm.

Each single peak or group of peaks collected, were then taken on for further analysis and are numbered in the following figure (Figure 3.6).

Figure 3.6

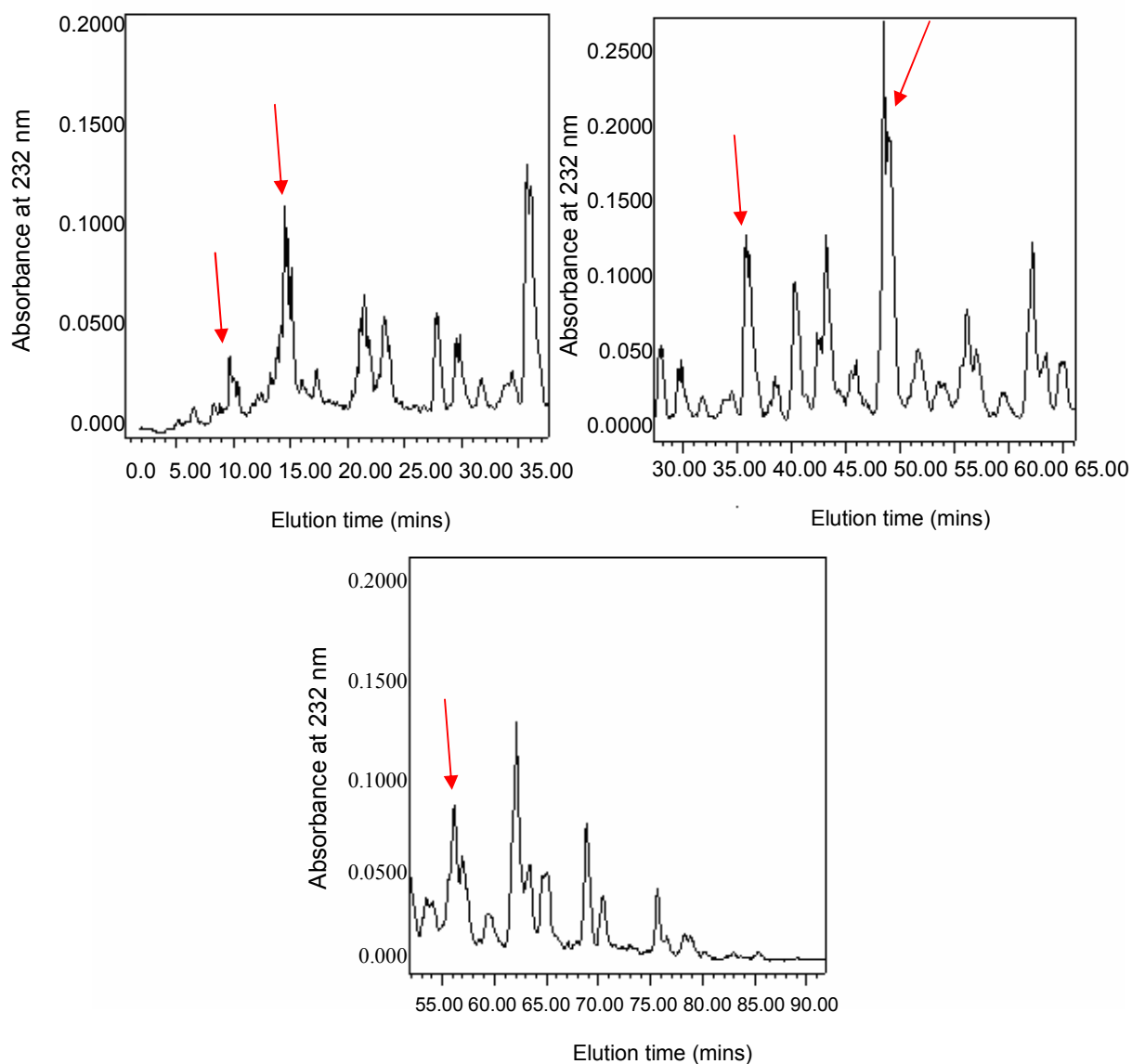


Figure 3.6 An expanded SAX-HPLC chromatography of sized HS oligosaccharides.

A typical SAX-HPLC profile of the dp8 sized oligosaccharides generated by heparinase III digestion, purified by Bio-Gel P10 gel-filtration chromatography. Red arrows indicate peaks collected for NMR analysis.

Panel A. An expanded HPLC profile from Figure 3.5, showing 0-35 minutes and peaks 1-5.

Panel B. An expanded HPLC profile, showing 30-65 minutes and peaks 6-13.

Panel C. An extended HPLC profile, showing 50-90 minutes and peak 14-19.

3.1.4 PAGE Analysis

It has previously been shown that SAX-HPLC purification of Hep III derived HS oligosaccharides may not give a single homogeneous oligosaccharide structure. The use of PAGE is an important step in the purification process of HS oligosaccharides as it separates HS species on the basis of charge as well as on the basis of size. An example of this is the two hexasaccharide samples purified by Vives *et al* (Vives *et al.*, 2001) in which both sequences were identical apart from the positioning of a 6-O-sulphate group in the chain (Figure 3.7). These eluted from the SAX-HPLC column at the same time, however when run through gel electrophoresis these were identified as two differently charged species:

It was decided that this step would be carried out after NMR analysis, to conserve time. A PAGE step, if needed, could be used in the NMR analysis where peaks resonate at the same frequency but appear almost as a shadow of each other. This would create an overlapping of certain peaks where more than one differently charged species was present in the sample.

Figure 3.7

HS-M6B:

Δ HexA - GlcNS - IdoUA(2S) - GlcNS(6S) - IdoUA - GlcNAc

HS-m6B:

Δ HexA - GlcNS - IdoUA(2S) - GlcNS - IdoUA - GlcNAc(6S)

Figure 3.7 An example of two different HS hexasaccharides oligosaccharides identified using PAGE analysis.

This figure has been taken from work by Vives *et al* (Vives *et al.*, 2001). Two different bands were identified in the gel filtration profile, but they were eluted from the SAX-HPLC column at the same time.

3.1.5 NMR

3.1.5.1 Preliminary NMR Analysis of Octasaccharide (dp8) Samples

Initial NMR analysis was carried out on a number of peaks identified in the HPLC profile (Figure 3.6). These were analyzed by ^1H ^1D NMR spectroscopy. (Note not all 19 peaks were submitted for analysis as there was not enough in the sample to get a full COSY, TOCSY and NOESY profile). Chemical shifts are based on values obtained mainly from heparin derived oligosaccharides (Chuangt *et al.* 2001), but also previous work carried on HS dp6 samples (Murphy 2007).

After initial ^1H ^1D NMR runs on several dp8 samples, it was decided that some were pure enough to be classed as single oligosaccharide species. Therefore PAGE analysis was deemed unnecessary, not just because they were thought to be pure samples but also to preserve as much sample as possible. That is not to say that all of the 19 peaks identified would not need this extra purification step. ^1D NMR analysis showed a clean spectrum with no shadowing of any peaks, which is indicative of a heterogeneous sample. The UA (H4) peak is the key in identifying a pure sample. However the major problem with all samples was a lack of material in all the NMR oligosaccharide samples. Sample was lost at every stage of the purification process which means it is especially hard to gain enough sample to enable a full NMR assignment. Even if the purification process was not as demanding, collecting dp8 sized oligosaccharides from gel filtration was not an easy step as that sample size itself was small; making it extremely difficult to collect pure oligosaccharides of the correct size.

Figure 3.8 One-dimensional 500 MHz ^1H -NMR spectra of HS octasaccharide dp8(1) and dp8(3) oligosaccharides.

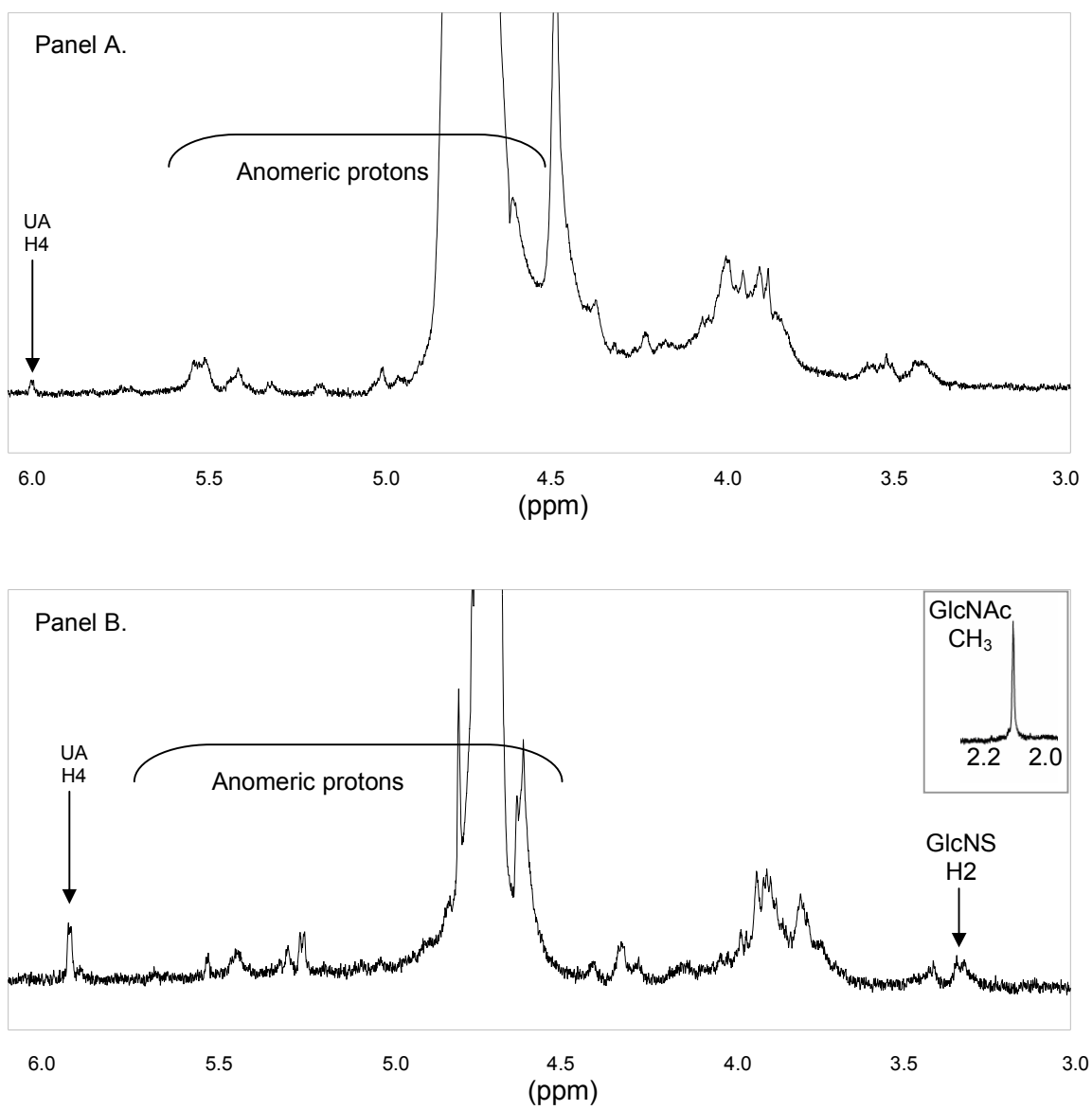
Oligosaccharides dp8(1) and dp8(3) were analysed by ^1H 1-D NMR spectroscopy. However due to the lack of sample assignment of the proton NMR spectra was not possible. Therefore chemical shifts were putatively assigned with the additional information from the literature chemical shift values reported mainly for heparin derived oligosaccharides (Chuang *et al.*, 2001) and also HS purified hexasaccharides (Murphy 2007).

Spectra were recorded in PBS at 25°C.

Panel A. dp8(1).

Panel B. dp8(3).

Figure 3.8



3.5.5.2 Partial Spectral Assignment of HS dp8(2):

Two dimensional COSY, TOCSY and NOESY NMR spectra were recorded for HS dp8(2), although there was not enough sample to assign all chemical shifts. As a demonstration of assignment a detailed breakdown of the proton assignment is given. The recorded spectra can be found in the appendix (see discussion).

COSY spectra

TOCSY spectra

NOESY spectra

The assignment of proton chemical shifts using 2D-NMR is regarded as standard practice, and so proton breakdown is not normally given; however for the purposes of assignment, the non-reducing terminal UA will be designated by the letter "a" with any subsequent assignments made in the chain lettered accordingly (alphabetically).

Figure 3.9 Assignment of the non-reducing terminal uronic acid, residue “a” of oligosaccharide HS dp8(2).

Using the known chemical shift for H4 of the uronic acid this was assigned using the COSY spectrum as 5.970 ppm. Then from here and going anti-clockwise around the Δ UA ring structure, H3 was assigned to 4.42 ppm, H2 to 3.99 ppm and finally H1 to 5.28 ppm.

The coloured arrows in the COSY spectrum (Figure 3.7) trace the path of this proton across the spectrum. Cross peaks formed from this ring system are linked by the arrows:

UA H4 - - - - ->
 H3 - - - - ->
 H2 - - - - ->
 H1 - - - - ->

A single trans-glycosidic cross-peak is then identified in the corresponding NOESY spectrum (Figure 3.9), circled. This is from an unassigned proton of the linking glucosamine residue and was assigned to 5.529 ppm. This cross-peak is circled in panel B.

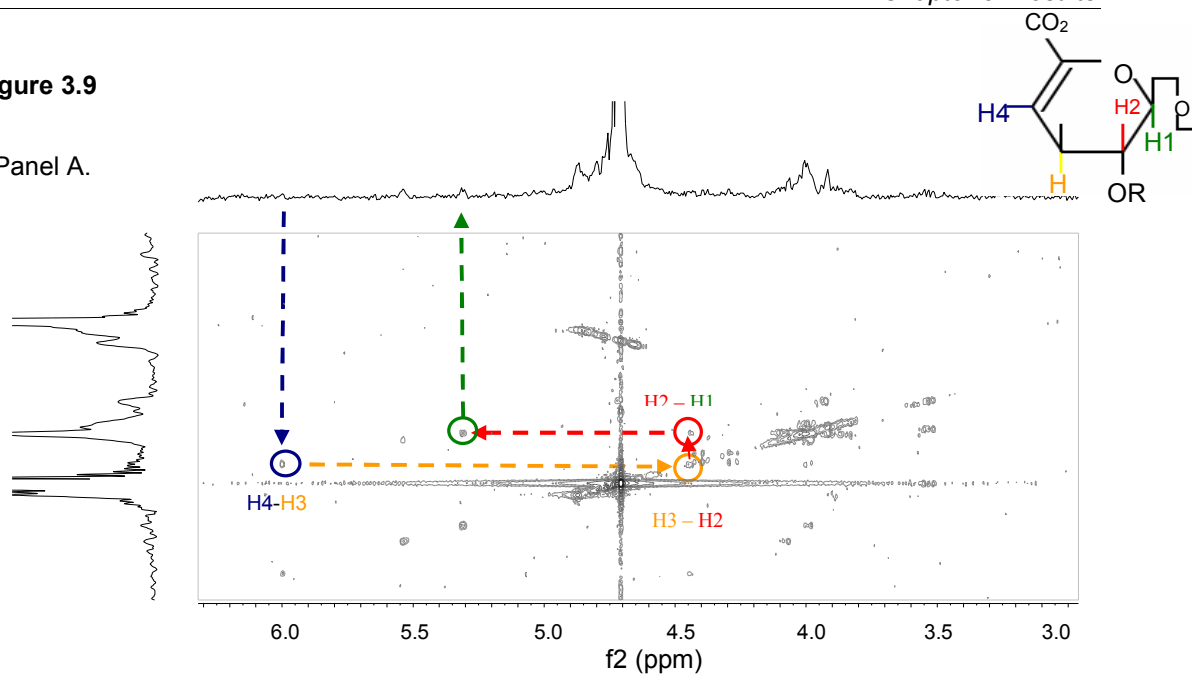
Panel A. COSY spectrum

Panel B. NOESY spectrum

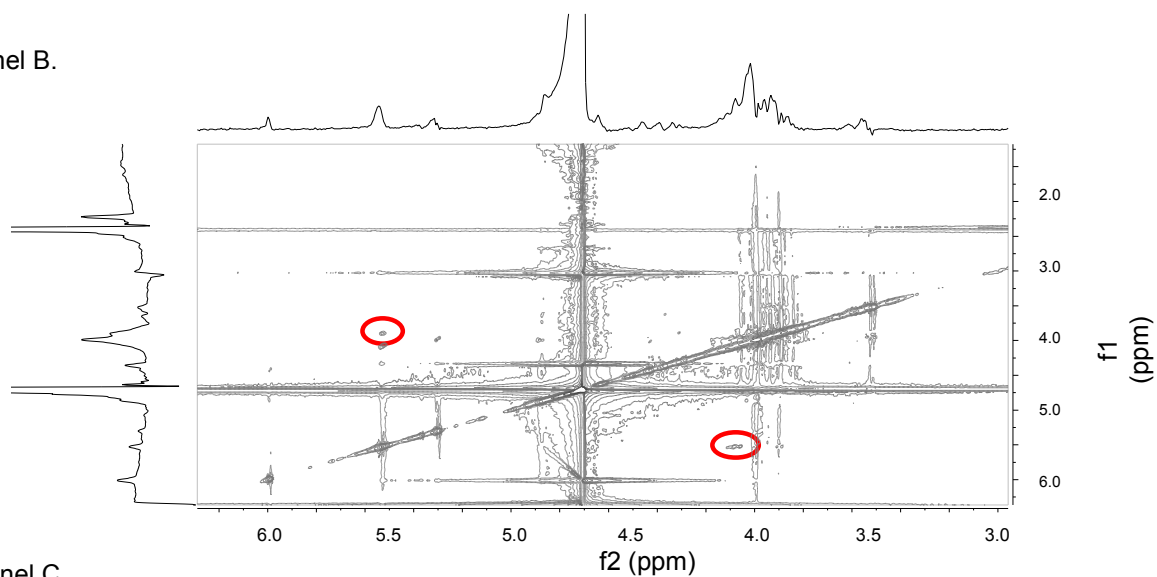
Panel C. ^1H spectrum

Figure 3.9

Panel A.



Panel B.



Panel C.

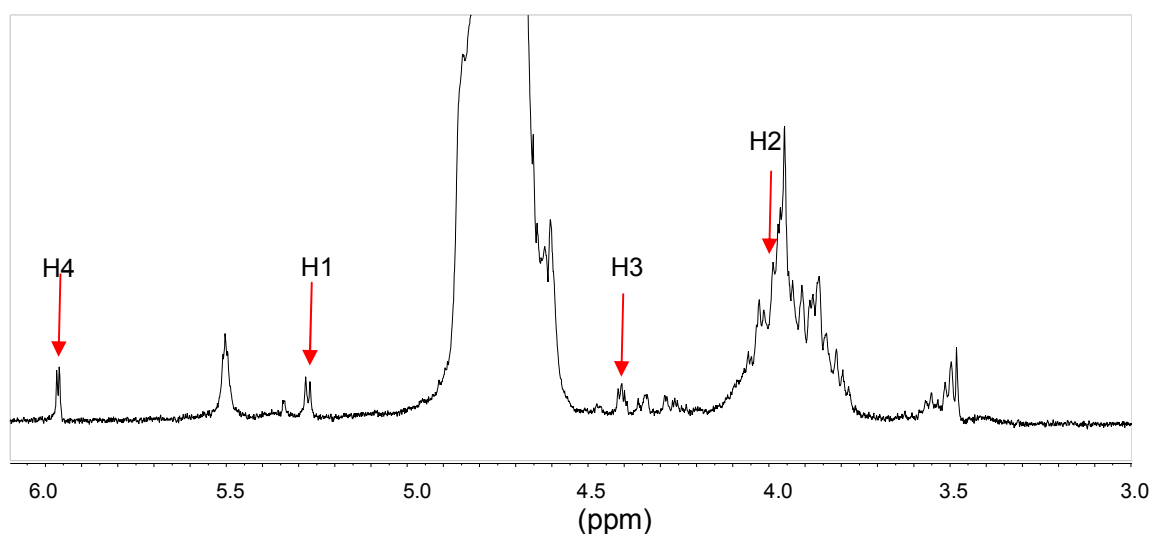


Figure 3.10 Assignment of glucosamine, residue “b” of oligosaccharide HS dp8(2).

Using the TOCSY spectrum (Panel A), the anomeric proton H1 from the ring system containing the unidentified glucosamine proton at 4.085 ppm was assigned. This was identified as the H1 at 5.552 ppm. Again using the COSY spectrum (Panel B) H2 was assigned 4.06 ppm, identifying this as the proton involved in the trans-glycosidic NOE from residue “a”, H3 to 4.15 ppm and H4 to 4.41 ppm. Unfortunately H5 and H6 were not assigned due to the lack of sample (this meant a HSQC spectrum could not be recorded).

The coloured arrows again correspond to the cross-peaks involved in tracking the particular proton of interest across the spectrum. The cross-peaks for protons H1-H4 form part of the glucosamine ring system.

H4 - - - - ->
H3 - - - - ->
H2 - - - - ->
H1 - - - - ->

Unfortunately due to lack of sample we were unable to identify any more cross-peaks in the NOESY spectrum.

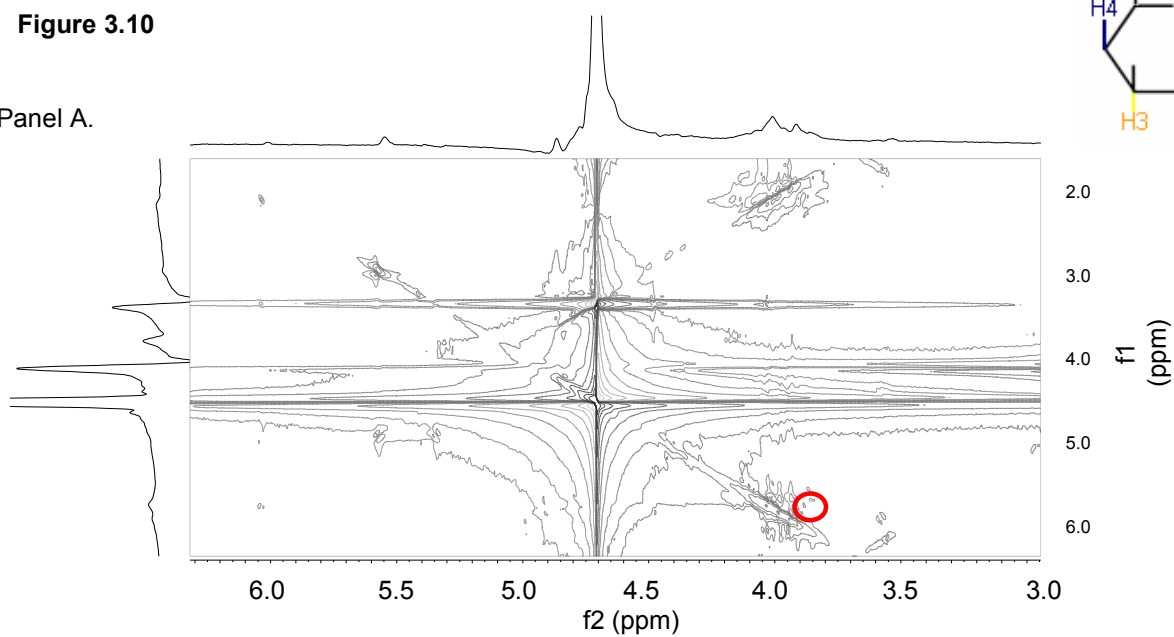
Panel A. TOCSY spectrum

Panel B. COSY spectrum

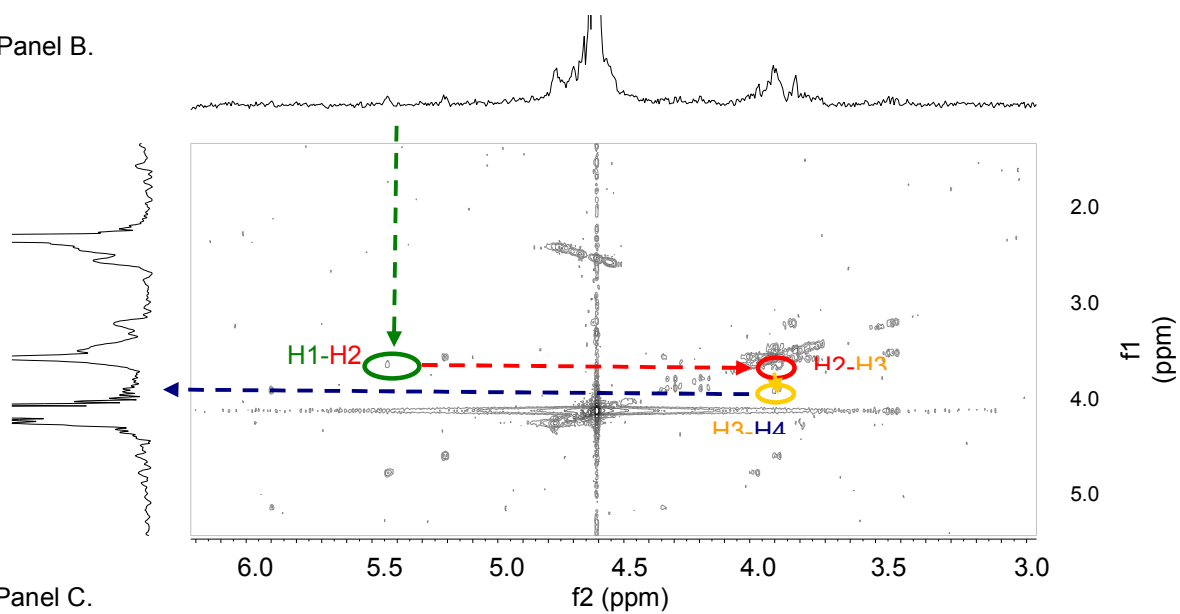
Panel C. ^1H spectrum

Figure 3.10

Panel A.



Panel B.



Panel C.

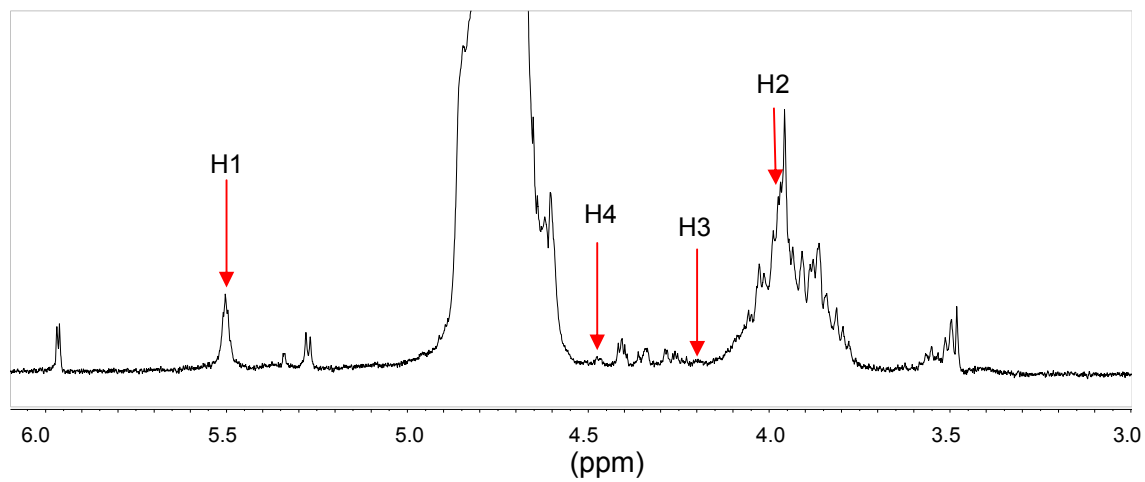


Figure 3.11 Spin System B

Using the COSY spectrum (Panel A) and starting with the chemical shift for the anomeric proton, H1 was assigned as 5.34 ppm; H2 was assigned 4.85 ppm, H3 to 3.92 ppm and H4 to 3.54 ppm. There was also a possible assignment for H5 to 4.75 ppm, but this was unconfirmed as it would be found under the water peak seen in the spectrum.

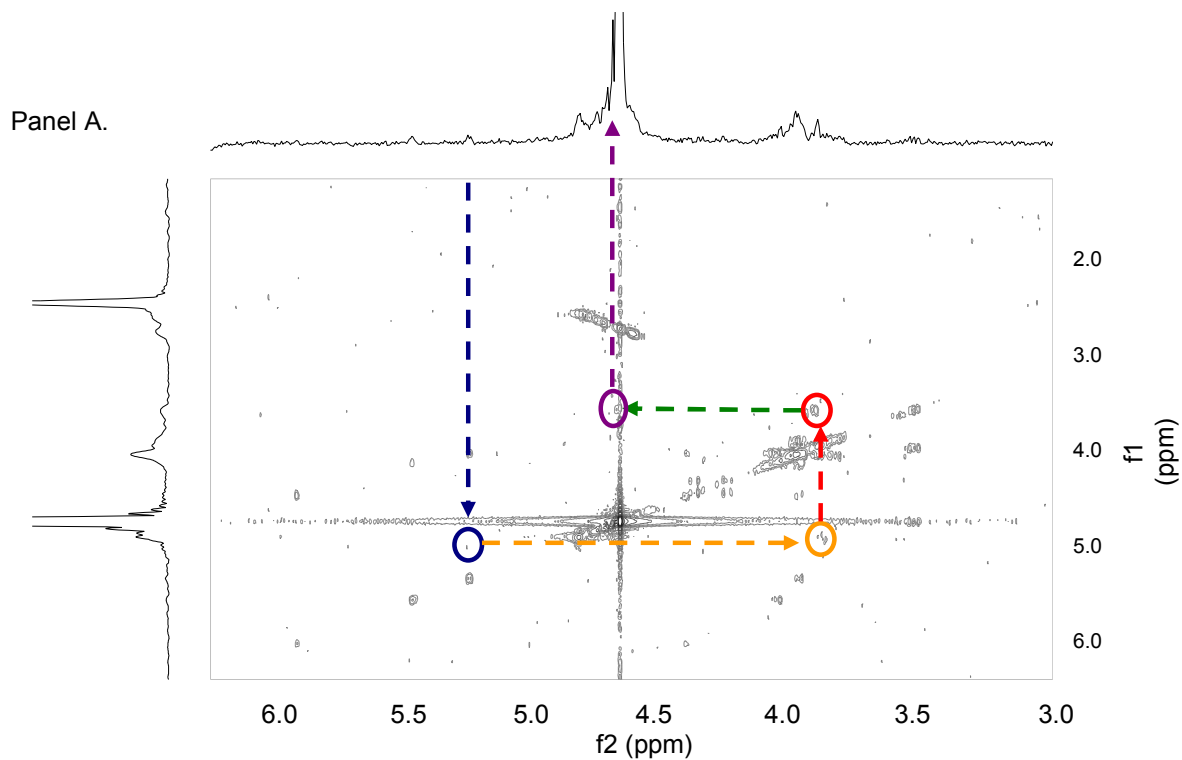
H4 - - - - ->
H3 - - - - ->
H2 - - - - ->
H1 - - - - ->

Possible H5 assignment - - - - ->

Panel A. COSY spectrum

Panel B. ^1H spectrum

Figure 3.11

Panel B.

A 1D ^1H NMR spectrum with the horizontal axis labeled (ppm) ranging from 6.0 to 3.0. Four peaks are labeled with red arrows: H1 at approximately 5.3 ppm, H2 at approximately 5.1 ppm, H3 at approximately 3.8 ppm, and H4 at approximately 3.6 ppm.

Figure 3.12 Spin System C

Using the COSY spectrum (Panel A) and starting with the chemical shift for the anomeric proton, H1 was assigned as 5.02 ppm; H2 was assigned 4.38 ppm, H3 to 4.29 ppm and H4 to 3.57 ppm. Due to the lack of sample for the TOCSY and NOESY analysis, this peak was not identified.

Panel A. COSY spectrum

Panel B. ^1H spectrum

Figure 3.12

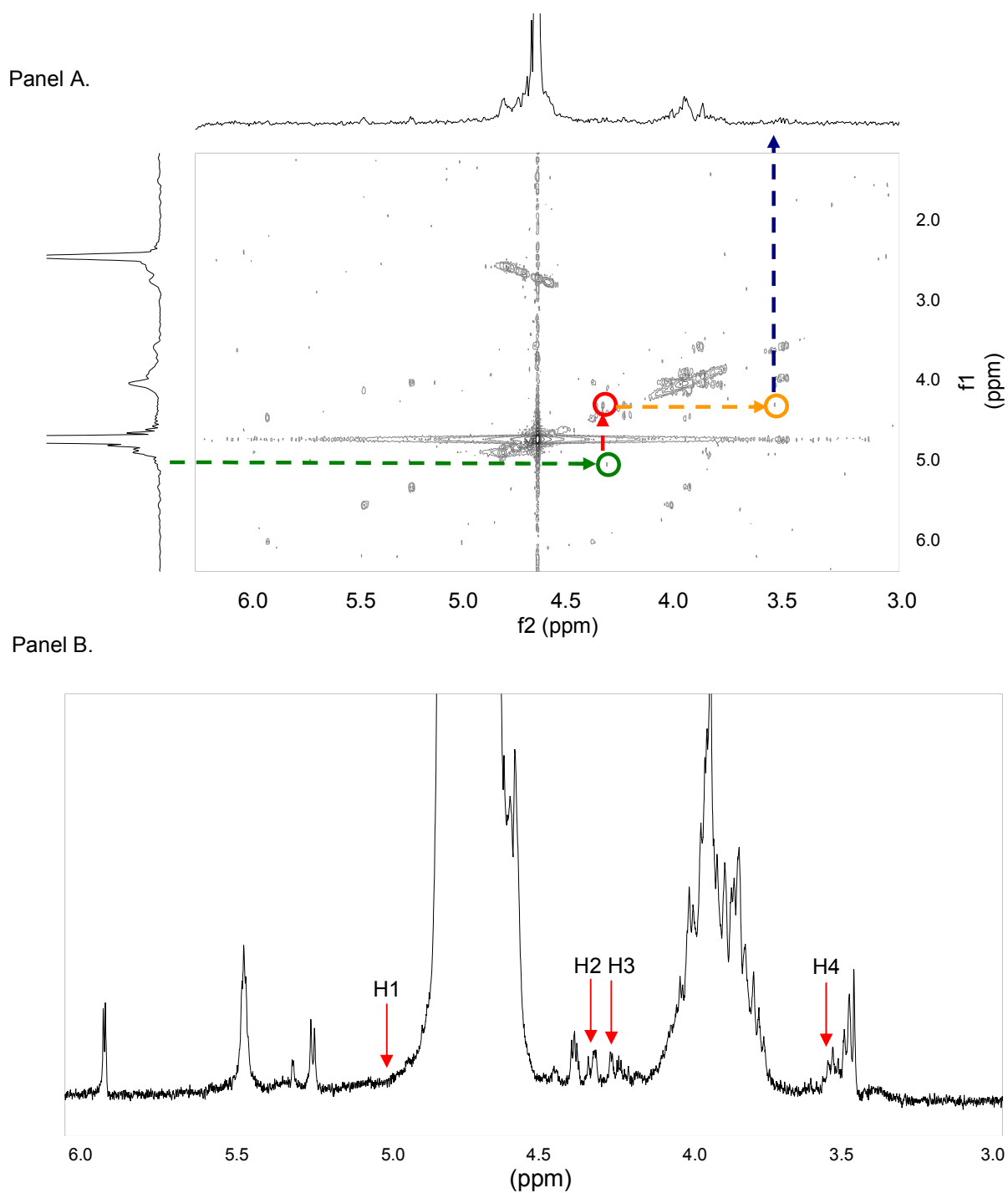


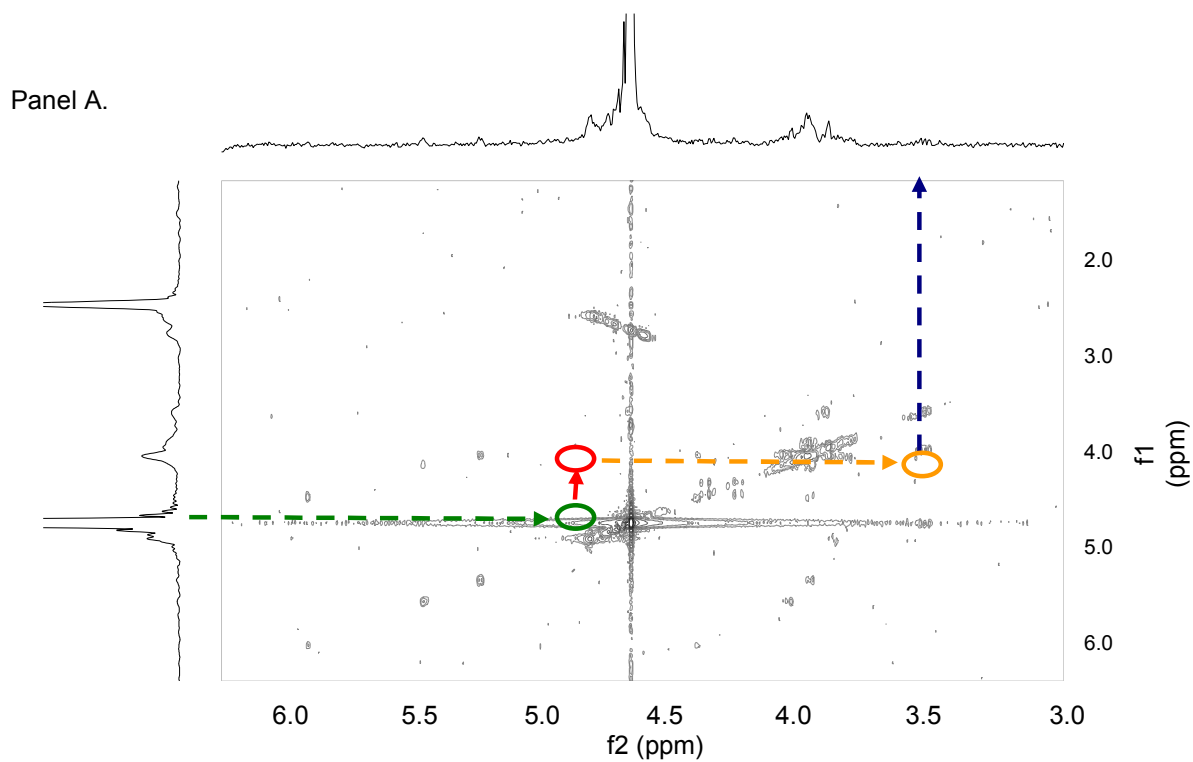
Figure 3.13 Spin System D

Using the COSY spectrum (Panel A) and starting with the chemical shift for the anomeric proton, H1 was assigned as 4.93 ppm; H2 was assigned 4.59 ppm, H3 to 3.96 ppm and H4 to 3.53 ppm. Due to the lack of sample for the TOCSY and NOESY analysis, this peak was again unidentified.

Panel A. COSY spectrum

Panel B. ^1H spectrum

Figure 3.13

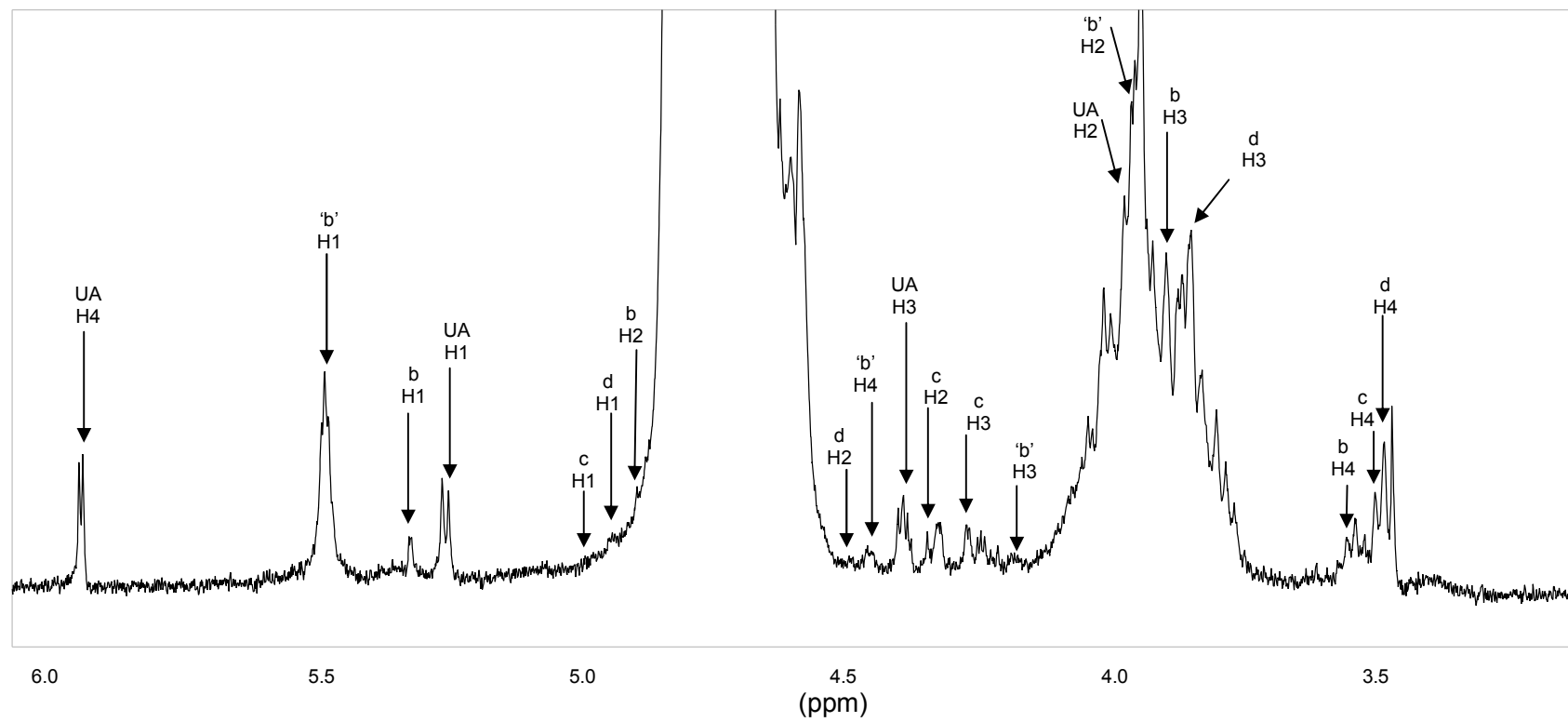
Panel B.

A 1D ^1H NMR spectrum. The horizontal axis is labeled (ppm) and ranges from 6.0 to 3.0. The spectrum shows several peaks. Four peaks are labeled with red arrows: H1 at approximately 5.0 ppm, H2 at approximately 4.6 ppm, H3 at approximately 3.9 ppm, and H4 at approximately 3.6 ppm.

Figure 3.14 A summary of the partial assignment of the ^1H chemical shifts for the oligosaccharide HS dp8(2)

The chemical shift assignments are summarized in Figure 3.14. Only the first two residues in the HS chain, assigned 'UA' (residue 'a') and 'b' are in sequence order. The other partial assignments involve COSY cross-peaks from unknown spin systems, assigned as spin system B, spin system C and spin system D.

Figure 3.14



3.2 Molecular Modelling

3.2.1 Calculation of potential energy surfaces

Potential energy surface maps for each glycosidic linkage were calculated as described under experimental methods and procedures. These are used as a way of identifying low energy linkage geometries, enabling an accurate building of molecular models. Energy surface maps were presented as contour plots, one for each glycosidic linkage in each oligosaccharide (data not shown). Static models were then built using these starting geometries and the minimum energy dihedral angles identified for all of the eight identified HS deca-saccharides (Table 3.1) using the already established strategy established by Mulloy *et al* (Mulloy *et al.*, 1993; Mikhailov *et al.*, 1997). This approach is valid as the glycosidic linkage geometry is the main contributor in determining the overall 3-dimensional shape of an oligosaccharide (see Introduction).

Table 3.1 Minimum selected ϕ° and ψ° dihedral angles across each glycosidic linkage of HS disaccharides

All dihedral angles were selected as a result of low energy minima across each glycosidic linkage. The potential energy was evaluated using the glycam04 force field within the AMBER molecular modelling suite. The conformations shown were used as the starting conformations for each residue in the later 1ns MD simulations for biologically active HS oligosaccharides.








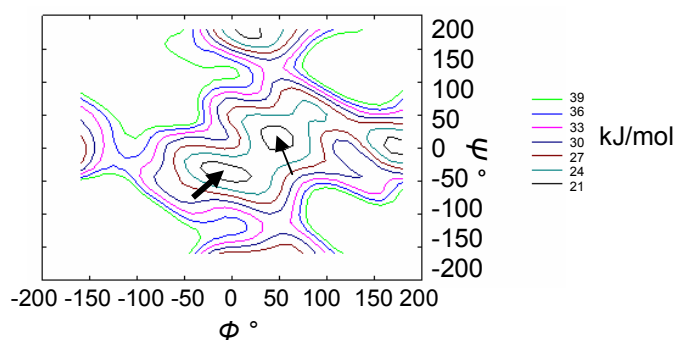
The corresponding contour map is an example of two identified low energy minima areas across the glycosidic bond between GlcUA - GlcNS. For the purpose of building static models the negative area was used. The contour lines present represent the global energy minima () marked with an arrow, and intervals from this point of 3 kJ/mol above this, () +3 kJ/mol above global minima, () +6 kJ/mol above global minima, () +9 kJ/mol above global minima, () +12 kJ/mol above global minima, () +15 kJ/mol above global minima, () +18 kJ/mol above global minima. The potential energy was evaluated using the glycam04 force field within the AMBER molecular modelling suite.

Table 3.1

Linkage and Monosaccharide Conformation Ring conformations indicated in ()	Dihedral angles	
	Φ°	Ψ°
UA (1H_2) – GlcNAc (4C_1)	55	5
UA (1H_2) – GlcNS (4C_1)	-15	-30
GlcUA (4C_1) – GlcNAc (4C_1)	40	5
GlcUA (4C_1) – GlcNAc(6S) (4C_1)	45	15
GlcUA (4C_1) – GlcNS (4C_1)	-10	-35
GlcUA (4C_1) – GlcNS(6S) (4C_1)	60	10
GlcNS (4C_1) – IdoUA (1C_4)	-40	-30
GlcNS(6S) (4C_1) – IdoUA (1C_4)	-75	-25
GlcNS(6S) (4C_1) – IdoUA(2S) (1C_4)	-50	0
GlcNS (4C_1) – IdoUA(2S) (1C_4)	-45	10
GlcNS (4C_1) – GlcUA (4C_1)	-35	-20
GlcNS(6S) (4C_1) – GlcUA (4C_1)	-35	-25
IdoUA (1C_4) – GlcNAc (4C_1)	35	15
IdoUA (1C_4) – GlcNS (4C_1)	-20	-30
IdoUA (1C_4) – GlcNAc(6S) (4C_1)	-30	-10
IdoUA(2S) (1C_4) – GlcNS(6S) (4C_1)	-50	-10
IdoUA(2S) (1C_4) – GlcNS (4C_1)	40	15
GlcNAc (4C_1) – GlcUA (4C_1)	-35	-20



3.2.2 Molecular dynamic trajectories of the glycosidic linkages ϕ° and ψ°

Previous studies on the ability of different HS oligosaccharide structures to activate or inhibit FGF2 signalling showed that it was impossible to make a link between biological activity and HS oligosaccharide structure by analysing the sequence/sulphation pattern of specific HS oligosaccharides (Goodger 2003). It was therefore proposed that conformational changes in HS oligosaccharide structure, which in turn have been shown to be linked to sulphation pattern, may be the source of unexplained FGF2 activatory and inhibitory activities seen with different HS oligosaccharide structures. The two principal variable parameters used in determining the overall conformation of HS oligosaccharides are the conformational flexibility of the IdoUA rings and the rotation around the glycosidic linkages. It must be remembered that static molecular models of HS oligosaccharides alone will not give accurate relationships between structure and biological activities, as the oligosaccharides are in a conformational equilibrium with many different conformations existing. Any of these could represent the biologically active shape. The challenge is to identify these active/inactive conformations from MD simulations of known HS oligosaccharide structures and determine the specific structural/conformational features required for FGF2 activation, which is the main aim of this thesis.

HS oligosaccharides studied by molecular modelling in this thesis were selected from the previously reported FGF2 stimulatory/inhibitory sequences. These structures are shown in Figures 1.23 and Figure 1.24 (see Introduction); along with their relevant activities with FGF2. The static models were built first followed by the MD simulations as described in the experimental methods.

Molecular dynamic trajectories or torsion data were obtained for each HS oligosaccharide studied using the AMBER suite of programmes and the data analysed via the mdxvu software programme (see <http://sourceforge.net/projects/mdxvu>) (Murphy 2007). This software allows the

monitoring and analysis of the individual glycosidic linkage angles (Φ and Ψ) over the 1 ns (1000 ps) explicit water molecular dynamic simulation. The 1000 ps timescale was used since previous studies using this timescale had shown that oligosaccharides would have sampled all possible geometries during this time frame (Murphy 2007). This data was then overlaid onto gnuplot grids which had previously been analysed in the static models and the low energy minima for each of the glycosidic torsion angles identified during the MD simulation.

MD trajectories were produced for each of the selected oligosaccharides using the AMBER suite of programmes and three programs: gnuplot, ptraj and mdxvu were used to analyse these trajectories in order to identify the low energy torsion angles. Both torsion angles across the glycosidic linkages were analysed independently starting with the Φ (phi) angle which has been assigned to atoms H1, C1, O1 and C4 and then the Ψ (psi) angle assigned to atoms C1, O1, C4 and H4 (Mulloy *et al.*, 1994). The resulting MD data was then combined and overlaid onto the static model contour map previously generated.

The results for each HS oligosaccharide are shown as contour maps (see Figures 3.17 through to Figure 3.25) and are found alongside the data obtained from the Cremer-Pople ring puckering parameter data which underwent a similar analysis and will be discussed in the following section. The movement of C1 and C4 is thought to be severely limited due to the addition of flanking residues on either side of the monosaccharide unit in question (Mulloy *et al.*, 1993). This has been shown in iduronate monosaccharides modelled as individual units which undergo more free rotation when modelled as a separate unit, rather than alongside other ring structures (Ferro *et al.*, 1990; Angulo *et al.*, 2005).

3.2.3 Puckering-torsion data

The Cremer-Pople ring puckering parameter is a means of reducing the Cartesian coordinates, of the oligosaccharide sugar rings to form a set of three simple parameters: Q , θ and ϕ_2 . The parameter Q is the maximum puckering amplitude, involving the rms of the displacements of the

six ring atoms in the ring plane. θ and ϕ_2 are angles which are found in the range of $0^\circ \leq \theta \leq 180^\circ$ and $0^\circ \leq \phi_2 \leq 360^\circ$ (Forster and Mulloy, 1993).

Molecular dynamic simulations were again carried out over 1 ns (1000 ps), as described in the experimental section to give an initial indication of any changes to the Cremer-Pople ring puckering parameters θ and where warranted ϕ_2 . Models were built in explicit water and kept at a temperature of 300 K at 1 atm (see Experimental Materials and Procedures). The starting structures of the IdoUA were first held in one of the three low energy conformations 1C_4 , 2S_0 and 4C_1 (see Introduction) to gain insight into the properties of each of the conformations during the length of the MD simulation (Murphy 2007). IdoUA was shown to be the only ring structure with the ability to change conformation and all three of the main conformations adopted by IdoUA (1C_4 , 2S_0 and 4C_1) in solution were shown to be present in the simulations. Variations in these parameters can be clearly seen with each of the IdoUA conformations having very definitive values for each angle. The Cremer-Pople puckering parameters for each IdoUA conformation were averaged out and the following values obtained:

IdoUA 1C_4	$\theta = 164.2^\circ$
IdoUA 2S_0	$\theta = 92.8^\circ$
IdoUA 4C_1	$\theta = 8.98^\circ$
UA 1H_2	$\theta = 131.7^\circ$
UA 2H_1	$\theta = 49.9^\circ$

The above angles were used throughout this thesis as an indicator of changes in ring conformation during the MD simulation. These values were adapted from Murphy (Murphy 2007).

The ideal 4C_1 and 1C_4 structures for hexapyranose rings have average data values of $\theta = 0^\circ$ and 180° respectively, putting them at the north and south poles of the Cremer-Pople sphere (Forster and Mulloy, 1993). All constants were averaged over all 1000 structures produced during the MD

simulation (Figure 3.15). The addition of the 2S group on the IdoUA had no effect on the value obtained, hence the same average angles could be used to identify conformational changes in IdoUA(2S) residues in addition to un-sulphated IdoUA. In some simulations the oligosaccharide structures individual IdoUA rings remained in one conformation (either 1C_4 , 2S_0 and 4C_1) for the duration of the MD simulation (Figure 3.15), others underwent a conformational change as shown (Figure 3.16).

These changes observed in puckering parameters during the MD simulations validated the glycam 04 force field in being able to sufficiently confirm and predict structural changes in HS oligosaccharides with regard to the low energy conformations of the iduronate ring (1C_4 , 2S_0 and 4C_1). This is of great importance as it would not do to predict structures that have not been shown to exist experimentally in solution.

The puckering parameter data once analysed separately was then overlaid onto the initial torsion data in order to understand if the changes identified in the Cremer-Pople ring puckering parameters corresponded to any changes already identified in the torsion data (shown in Molecular dynamic trajectories of the glycosidic linkages Φ and Ψ). This data is known as the puckering-torsion data and has been laid out over the 1000 ps timescale (the timescale run for all initial molecular dynamic simulations).

The initial conformations are shown in Table 3.1 which is then compared to the MD simulation results for all the decasaccharides. Molecular modelling techniques were used in a similar manner to those used by Forster and Mulloy (Forster and Mulloy, 1993).

As will be shown later, a change in ring conformation of just a single IdoUA within HS oligosaccharides can result in a significant change in oligosaccharide shape and hence biological activities. Interestingly, not all possible IdoUA conformations or glycosidic torsion angle conformations are predicted to exist within HS oligosaccharides due to the limits of the MD simulations. This is also true with the measurement of glycosidic torsion angles which were shown in the MD simulations to only exist in a small number of low energy geometries, these

changes between torsion angles will also be shown later to result in large shifts in oligosaccharide shape. This data taken together will be seen to suggest a role for sulphation pattern in controlling conformational dynamics of HS oligosaccharides and their eventual biological activities.

Figure 3.15 Cremer-Pople puckering parameters for the HS oligosaccharide iduronate residues, alongside the Δ UA residue on the reducing terminal during a 1000 ps molecular dynamic simulation.

For the Δ UA, the HS oligosaccharide used was the dp10(1) oligosaccharide, alongside the methyl glycoside of IdoUA which had been previously carried out by Murphy (Murphy 2007). These were all repeated for the purpose of this study.

Panel A. θ° values starting in the 4C_1 conformation for IdoUA.

Panel B. θ° values starting in the 2S_0 conformation for IdoUA.

Panel C. θ° values starting in the 1C_4 conformation for IdoUA.

Panel D. θ° values starting in the 2H_1 conformation for the Δ UA.

Panel E. θ° values starting in the 1H_2 conformation for the Δ UA.

Figure 3.15

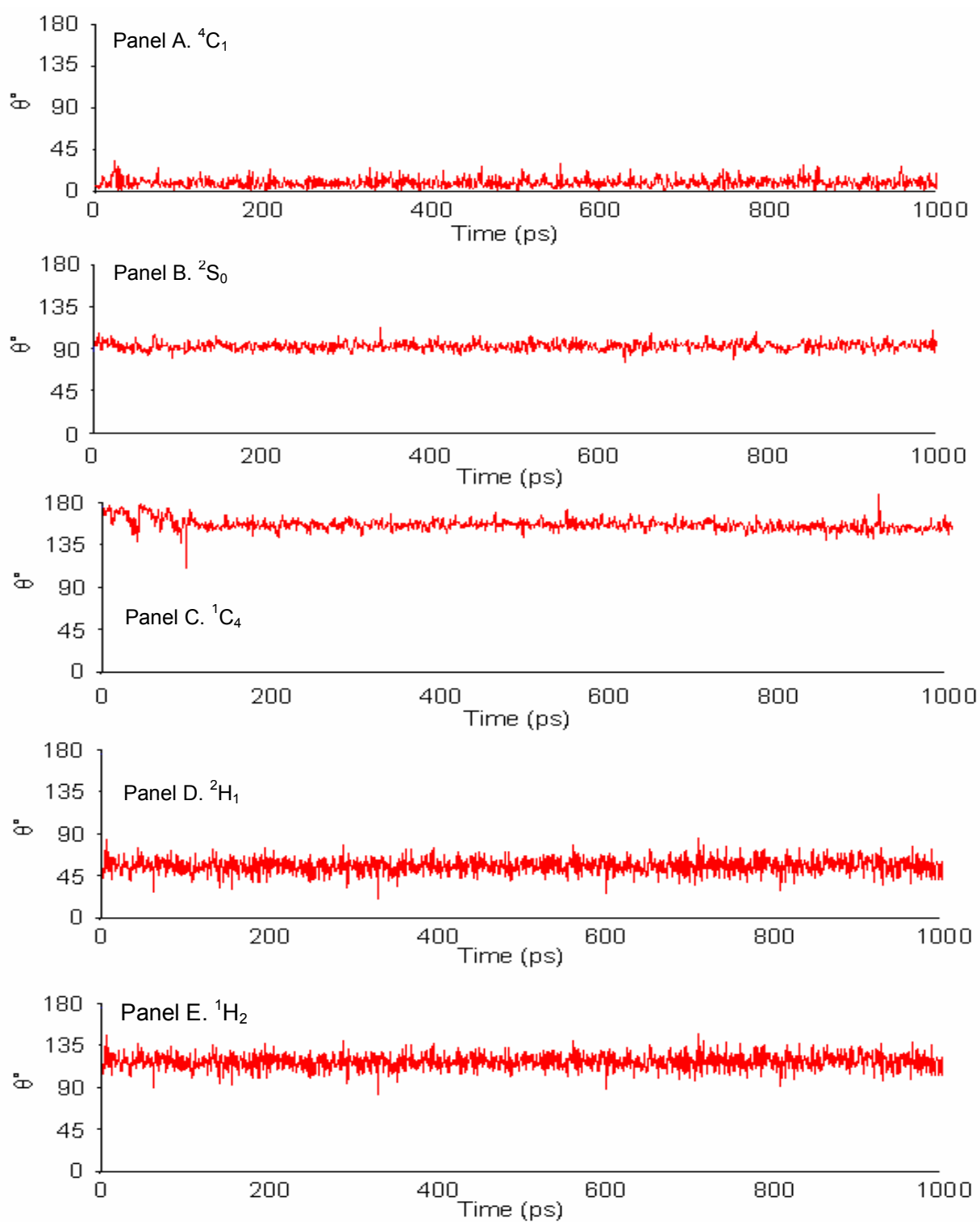
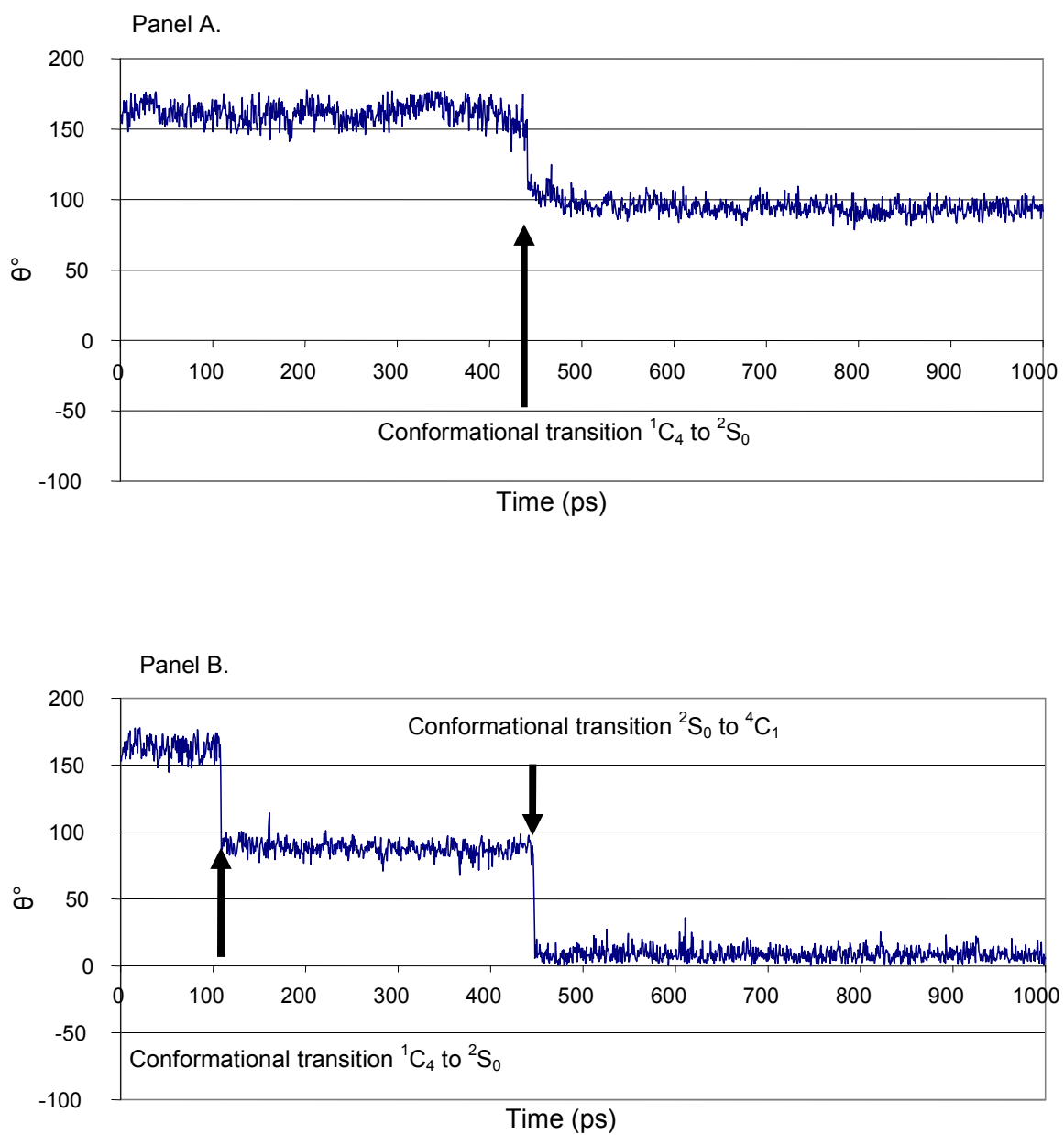


Figure 3.16 Cremer-Pople puckering parameters for the internal HS oligosaccharide Iduronate-2-O sulphate residues during a 1000 ps molecular dynamic simulation.

Panel A. IdoUA(2S) in the 1C_4 starting conformation undergoes a conformational transition to the 2S_0 conformation.

Panel B. IdoUA(2S) in the 1C_4 starting conformation undergoing two conformational transitions; firstly to the 2S_0 conformation and then onto the 4C_1 conformation.

Figure 3.16



All active and inactive decasaccharide 1000 ps MD simulations were analysed using the program mdxvu, to identify the Cremer-Pople ring puckering parameters. It was hoped that differences would be identified between the active and inactive sequences, enabling a greater picture to be built up of the main reason why two groups of oligosaccharides with fairly similar structures have a different ability to bind to FGF2.

Each monosaccharide unit was analysed to identify the conformations it sampled throughout the MD simulation. This was then compared to the starting conformation in which it was initially built in. This was then used to check against the torsion data previously analysed. Any changes identified in the puckering parameters would hopefully correspond to a change in the dihedral angle identified in the torsion data. The results confirmed what was previously known regarding the conformation of the glucosamines i.e. these were all identified in the fixed 4C_1 conformation and remain fixed in this conformation regardless of the length of MD simulation (Ferro *et al.*, 1990). This has been shown in all the corresponding results over the 1000ps timescale and is characterised by an average θ value of 10° . Therefore the major contribution to flexibility within the HS oligosaccharide must come from the iduronates as stated previously (see Introduction, page 64). IdoUA and IdoUA(2S) occur in the three identified conformations (1C_4 , 2S_0 and 4C_1) during our MD simulations and also, as will be shown later, glycosidic torsion angles.

3.2.4 Decasaccharide models

The four decasaccharide structures that were chosen to be modelled were identified in sequencing FGF2 activity work carried out by Goodger (Goodger 2003); who identified a collection of different highly pure HS decasaccharides from a variety of sources and attempted to link HS sequence data with biological activity. Two of the HS oligosaccharide sequences activated FGF2 in biological assays and two were inactive. There was some initial confusion regarding the structure of the decasaccharide sequences at position nine (from the uronic acid terminal position) in the chain. When fully sequenced; using nitrous acid to fully depolymerise the

HS chain (see p50); the residue found at position 9 was never fully identified. This unit was identified as being a uronic acid, either GlcUA or IdoUA, hence two molecular models were built for each oligosaccharide with either GlcUA or IdoUA at position 9. This resulted in eight deca-saccharide models being built designated dp10(1) - dp10(8) (see below).

Any differences due to the change in position 9 would be interesting to note as it is unknown whether the reducing terminal has any influence on ring conformation of IdoUA or the glycosidic linkages further down the chain.

Models were built from the reducing end first as it was easier to add models onto the growing chain as described in the experimental section.

Active Oligosaccharides

dp10(1). UA - GlcNAc - GlcUA - GlcNS(6S) - IdoUA(2S) - GlcNS(6S) - IdoUA(2S) - GlcNS(6S) - IdoUA - GlcNAc(6S)

dp10(2). UA - GlcNAc - GlcUA - GlcNS(6S) - IdoUA(2S) - GlcNS(6S) - IdoUA(2S) - GlcNS(6S) - GlcUA - GlcNAc(6S)

dp10(3). UA - GlcNAc - GlcUA - GlcNS - IdoUA(2S) - GlcNS(6S) - IdoUA - GlcNS - IdoUA - GlcNAc(6S)

dp10(4). UA - GlcNAc - GlcUA - GlcNS - IdoUA(2S) - GlcNS(6S) - IdoUA - GlcNS - GlcUA - GlcNAc(6S)

Inactive Oligosaccharides

dp10(5). UA - GlcNAc - GlcUA - GlcNS - IdoUA(2S) - GlcNS(6S) - IdoUA(2S) - GlcNS(6S) - IdoUA - GlcNAc

dp10(6). UA - GlcNAc - GlcUA - GlcNS - IdoUA(2S) - GlcNS(6S) - IdoUA(2S) - GlcNS(6S) - GlcUA - GlcNAc

dp10(7). UA - GlcNS - IdoUA - GlcNS - IdoUA(2S) - GlcNS(6S) - IdoUA(2S) - GlcNS - IdoUA - GlcNAc(6S)

dp10(8). UA - GlcNS - IdoUA - GlcNS - IdoUA(2S) - GlcNS(6S) - IdoUA(2S) - GlcNS - GlcUA - GlcNAc(6S)

3.2.4.1 The Reducing End Linkage - Linkage 9 for the eight

Decasaccharide models dp10(1) - dp10(8)

The eight decasaccharides had the following structures for linkage 9 between monosaccharide positions 9 and 10:

IdoUA - GlcNAc(6S)	dp10(1, 3 and 7)
GlcUA - GlcNAc(6S)	dp10(2, 4 and 8)
IdoUA - GlcNAc	dp10(5)
GlcUA - GlcNAc	dp10(6)

The GlcUA monosaccharides were all modelled in the 4C_1 conformation with the IdoUA monosaccharides modelled in the 1C_4 conformation prior to the MD simulation.

The reducing end terminal contains the 6-O-sulphated, N-acetylated GlcNAc residue in all decasaccharides apart from dp10(5) and dp10(6). In line with previous studies (Powell *et al.*, 2004) these were all identified only in the 4C_1 conformation and remained in this conformation over the whole 1000 ps MD simulation (data not shown). The average value for this conformation has been identified as $\theta = 9^\circ$ (see puckering-torsion data).

Those oligosaccharides with IdoUA as the chosen residue at position 9 are dp10(1), dp10(3), dp10(5) and dp10(7). The three known conformations identified previously were found to be present in three of these four models; with the exception being dp10(5). Those oligosaccharides which sampled all three conformations show similar profiles in both the torsion data and the puckering-torsion data. In dp10(1) the ring starts in the 1C_4 conformation; (which is the set conformation built at the start of the MD simulation). The corresponding glycosidic torsion angles here are identified in the negative area of the graph (Figure 3.17, Panel A) which has been assigned as geometry B, average value of $\phi = -20^\circ$ $\psi = -30^\circ$. As the ring makes the transition to the 2S_0 conformation after approximately 100 ps and then finally onto the 4C_1 conformation (after

400 ps) the values for the dihedral angle increase up to an average of $\Phi = 30^\circ$ $\Psi = 10^\circ$, identified as geometry A. A similar pattern is identified in dp10(3), with the torsion angle having a starting value of $\Phi = 40^\circ$ $\Psi = 10^\circ$, placing it in the range of positive geometry A, whilst the ring changes from the 1C_4 conformation to the skew boat conformation (this occurs early on in the simulation after approximately 50 ps). As the ring undergoes a conformational change to the 4C_1 conformation the Φ and Ψ values change from geometry A to B ($\Phi = -30^\circ$ $\Psi = -30^\circ$) and then back again (Figure 3.17, Panel C). For dp10(7) the torsion angle begins the MD simulation in the negative quadrant of the graph (geometry B) and remains here for 300 ps until the IdoUA undergoes a conformational transition to the final 4C_1 conformation. As the ring conformation switches the torsion angles change to a positive value as indicated by geometry A ($\Phi = 40^\circ$ $\Psi = 10^\circ$). It remains in this area of the graph for the remainder of the simulation (Figure 3.16, Panel G).

Dp10(5) behaves slightly differently in that at the beginning of the MD simulation the ring structure is already identified in the 2S_0 conformation for a short time before a conformational transition to the 4C_1 conformation after only 50 ps. This shows that the conformational transition from 1C_4 to 2S_0 has already taken place in the equilibrium process leading up to the MD simulation process. The position of the dihedral angle at the point where the ring is in the 2S_0 conformation in dp10(5) is found in the upper right hand quadrant of the contour map, designated as geometry A. This has an average torsion angle of $\Phi = 40^\circ$ $\Psi = 10^\circ$, this area of the graph is sampled many times over the 1000 ps timescale and hence represents the most variable linkage studied so far with respect to the number of torsional changes during the simulation. The negative area of the graph is identified with average values of $\Phi = -20^\circ$ $\Psi = -20^\circ$ and is designated geometry B (Figure 3.17, Panel E).

These residues all end the simulation in the 4C_1 conformation and that is regardless of the presence of the 6S group on the terminal residue. The transition to the apparently more stable conformation occurs early on in the MD simulation with all models identified in the 4C_1 conformation before 500 ps. Longer simulations will be needed in order to confirm that the 4C_1 conformation can return to the 2S_0 or the 1C_4 , however one of the 1 ns MD simulations does show

an increase in the puckering parameter indicating a return shift of the 2S_0 to the 1C_4 so reverting from 4C_1 back to the 2S_0 could reasonably be expected over much longer simulations using our force field parameters.

This realignment of the linkage involves a reorientation of not only the terminal residue but also has a major influence on the rest of the model in terms of its 3-dimensional shape.

It has been suggested that the reducing end conformation is influenced by its environment and particularly the neighbouring residues seem to have some kind of effect on the terminal end conformation. The presence of either the NS or NAc substituent on the flanking amino sugar suggests the dominating conformations are either 1C_4 or the 4C_1 conformations. (Ferro *et al.*, 1990). These conformations identified at the reducing terminal have been identified in a foot and mouth disease virus-oligosaccharide receptor complex (O₁BFS, which binds to HS) where the 1C_4 and 4C_1 conformations are thought to dominate the reducing end conformation (Fry *et al.*, 1999). The presence of the skew boat form is not as common in literature when referring to the reducing end residue; however it has been shown to play a role in the Goodger models studied here.

Those oligosaccharides with the GlcUA modelled in position 9, remained in the 4C_1 throughout the MD simulation and so have shown the previously reported lack of flexibility due to the presence of the rigid 4C_1 conformation (Murphy *et al.*, 2008). This also validates the rule that the rigid conformation is favoured alongside 1C_4 when the connecting amino sugar is either N-sulphated or N-Acetylated (Ferro *et al.*, 1990). The flexibility around this residue will not come from the conformation of the ring structure but from the dihedral angle. The oligosaccharides dp10(6) and dp10(8) show less flexibility around the dihedral angle. They have both geometries present but spend the majority of the 1000 ps timescale in geometry A ($\Phi = 40^\circ$ $\Psi = 10^\circ$ for dp10(6) and $\Phi = 30^\circ$ $\Psi = 10^\circ$ for dp10(8) see Figure 3.16 Panels F and H respectfully). Dp10(2) behaves in a similar way to dp10(1 and 3) with the dihedral angle changing from a positive to a

negative position ($\Phi = 40^\circ$ $\Psi = 10^\circ$, geometry A to $\Phi = -20^\circ$ $\Psi = -20^\circ$, geometry B see Figure 3.17, Panel B).

The only oligosaccharide which acts in a different way is dp10(4) and that oligosaccharide is only present in geometry B. This is characteristic of the angle existing on average in the negative area of the graph and never switching to the positive section (Figure 3.17, Panel D). The observed results go against those obtained by Murphy (Murphy 2007) where it was identified the reducing terminal IdoUA – GlcNAc charged models exist in one geometry which happened to be around the starting geometry. Those with the sequence GlcUA-GlcNAc sampled two alternate geometries of $\Phi = 60^\circ$, $\Psi = 0^\circ$ and $\Phi = 0^\circ$, $\Psi = -20^\circ$. Here it was thought that this sequence had a higher degree of conformational flexibility over the IdoUA-GlcNAc linkage. Murphy (Murphy 2007) also suggested the importance of the 6-O-sulphate group and the fact that they influence the dynamics of adjacent glycosidic linkages. This effect is not seen immediately in the data presented here but may become present in longer MD simulations.



Figure 3.17 Puckering-torsion data for linkage 9

IdoUA - GlcNAc(6S)	dp10(1, 3 and 7)
GlcUA - GlcNAc(6S)	dp10(2, 4 and 8)
IdoUA - GlcNAc	dp10(5)
GlcUA - GlcNAc	dp10(6)

Puckering parameters: all graphs show θ analysis ().

Cremer-Pople ring puckering parameters during a 1000 ps molecular dynamic simulation.

All graphs show θ analysis (ϕ_2 not shown) and changes are compared to the values identified in the torsion data in chronological order.

Torsion data: Φ° dihedral angle (), Ψ° dihedral angle ().

The variation in Φ and Ψ torsion angles over a course of 1000 ps explicit water molecular dynamic simulation (the left hand side figure).


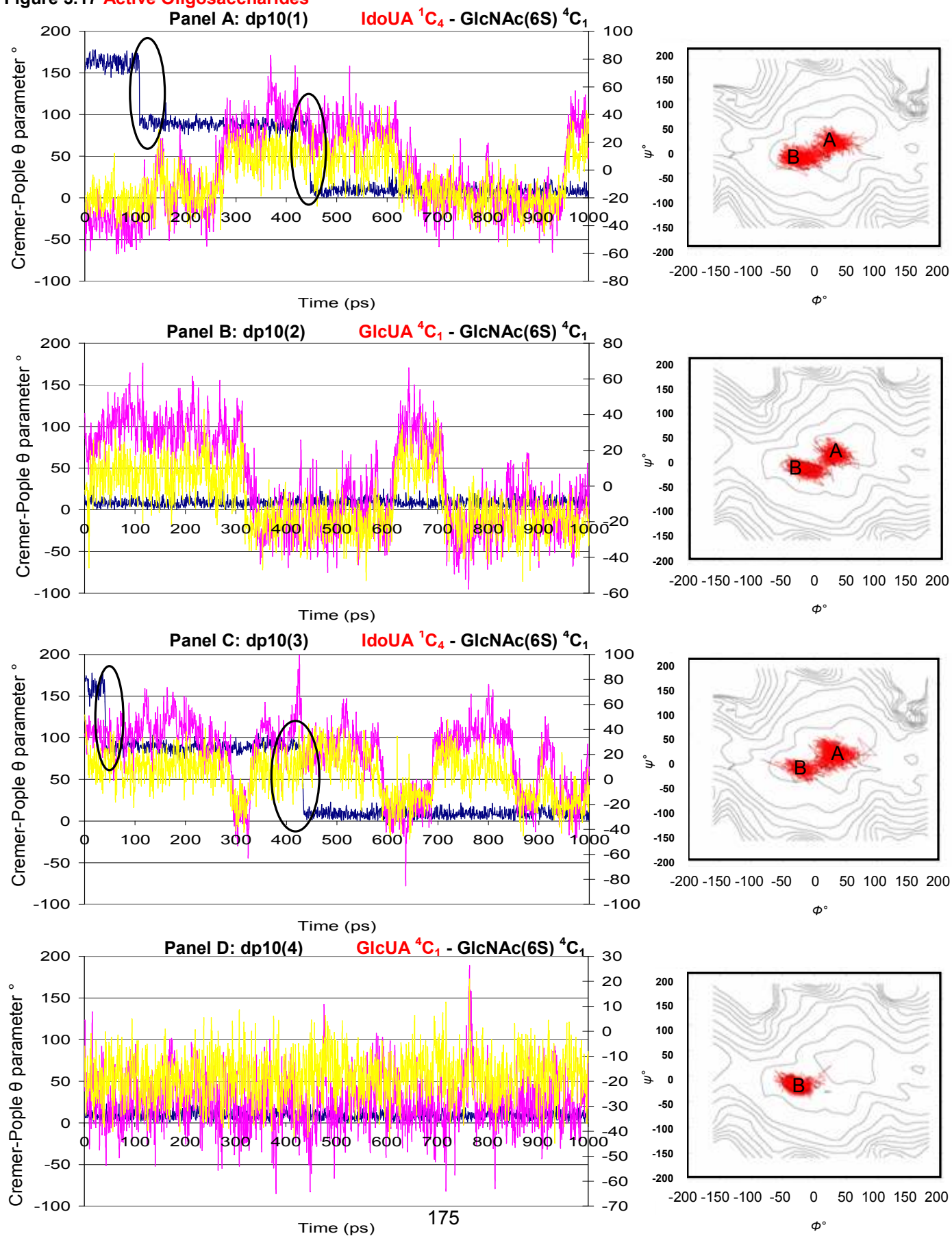
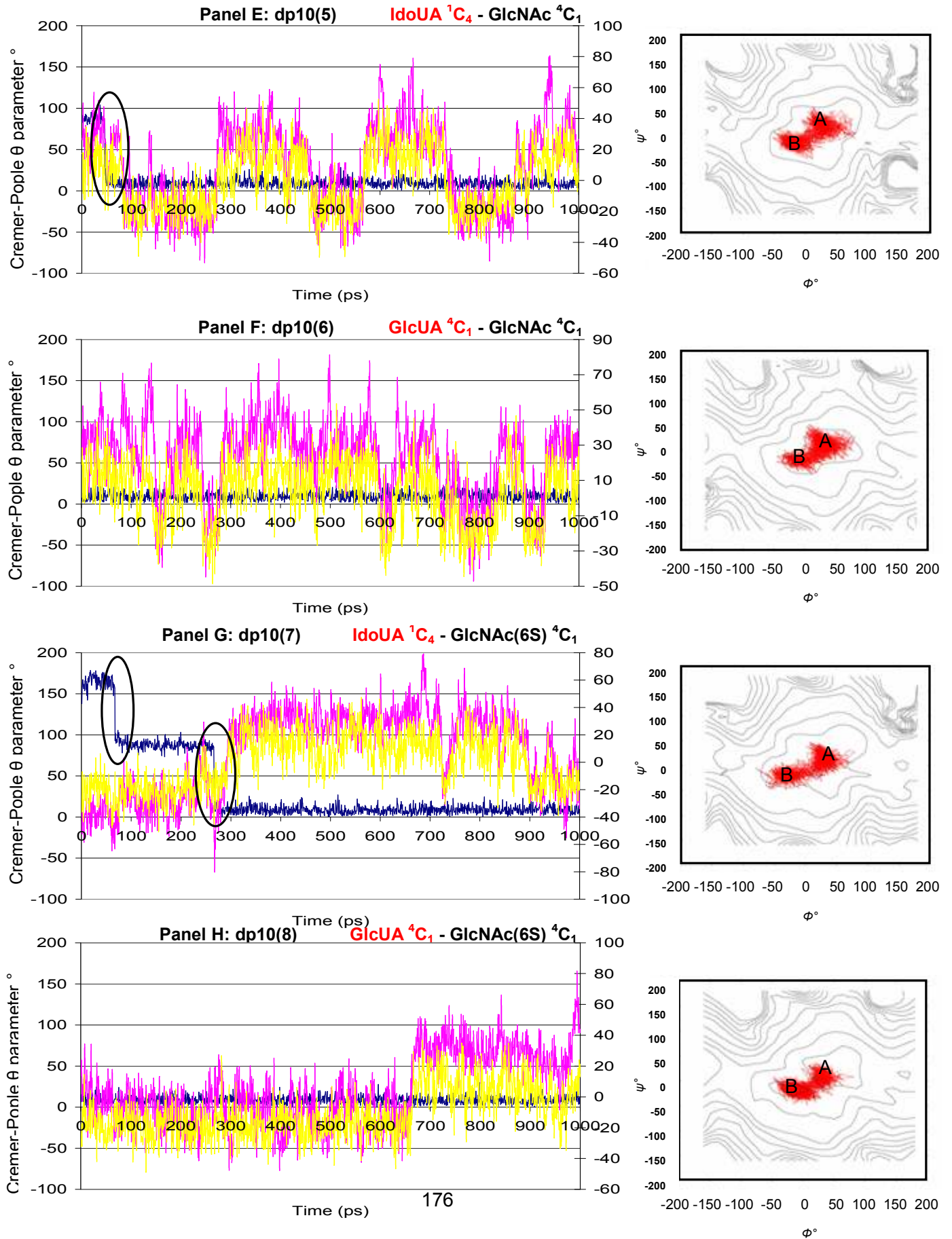
Torsion data: Structures were sampled every picosecond over the course of 1000 ps in TIP3P explicit water. The Φ and Ψ torsion angles across each glycosidic linkage were plotted and connected together in chronological order (). Contour lines are plotted as a guideline reference for each linkage and represent the potential energy surface of each of the glycosidic linkages (the right hand side figure).

Figure 3.17 Active Oligosaccharides



Inactive Oligosaccharides



3.2.4.2 Linkage 8

GlcNS(6S) - IdoUA	dp10(1 and 5)
GlcNS(6S) - GlcUA	dp10(2 and 6)
GlcNS - IdoUA	dp10(3 and 7)
GlcNS - GlcUA	dp10(4 and 8)

The GlcNS \pm (6S) monosaccharides were all modelled in the 4C_1 conformation prior to the 1 ns MD simulation.

All residues here in linkage 8 are glucosamines and so are only found in the 4C_1 conformation as identified recently (Powell *et al.*, 2004). Any flexibility in these linkages will therefore come from changes in torsion angles and not the conformation shifts of the ring structure. There are four different types of disaccharides which make up linkage 8 across the range of oligosaccharides so it is difficult to fully compare them all. It will also be interesting to note any differences with the addition of the 6-O-sulphate group on deca-saccharides making up this linkage.

The main differences in torsion angle geometries of linkage 8 between active and inactive oligosaccharides appears to be with oligosaccharides dp10(3) and dp10(7) which both have the same flanking monosaccharides (GlcNS-IdoUA). In this instance the inactive dp10(7) oligosaccharide exists in a single geometry A whilst the active dp10(3) oligosaccharide has an additional B geometry (Figure 3.18, Panels C and G). This difference is significant and could influence overall biological activity. A minor difference occurs between the active dp10(1) and inactive dp10(5) oligosaccharide, which have the same GlcNS(6S)-IdoUA sequence, there appears to be 2 distinct geometries in the inactive oligosaccharide and a single geometry in the active structure, also the ψ angle of the inactive dp10(5) oligosaccharide (Figure 3.18, Panel E) has an average value in geometry A of around +20 degrees ($\Phi = -20$, $\psi = +20$) whereas the ψ angle for active oligosaccharides dp10(1) remains at a relatively constant angle of -20 degrees ($\Phi = -20$, $\psi = -20$); (Figure 3.18, Panel A). The dp10(5) does briefly spend some time however with the ψ at -20 degrees ($\Phi = -20$, $\psi = -20$), this rules this out as a potential difference between

active and inactive structures. Another difference is between the active dp10(2) and inactive dp10(6) oligosaccharides that they have the same GlcNS(6S)-GlcUA sequence in this linkage and both appear to occupy a similar single geometry on the contour map, however the active dp10(2) has an average value ψ of -30 degrees with the inactive dp10(6) oligosaccharide having a ψ +30 degree average. This is a significant shift and could have a bearing on the biological activity of the samples.



Figure 3.18 Puckering-torsion data for linkage 8

GlcNS(6S) - IdoUA	dp10(1 and 5)
GlcNS(6S) - GlcUA	dp10(2 and 6)
GlcNS - IdoUA	dp10(3 and 7)
GlcNS - GlcUA	dp10(4 and 8)

Puckering parameters: all graphs show θ analysis ().

Cremer-Pople ring puckering parameters during a 1000 ps molecular dynamic simulation.

All graphs show θ analysis (ϕ_2 not shown) and changes are compared to the values identified in the torsion data in chronological order.

Torsion data: Φ° dihedral angle (), Ψ° dihedral angle ().

The variation in Φ and Ψ torsion angles over a course of 1000 ps explicit water molecular dynamic simulation (the left hand side figure).


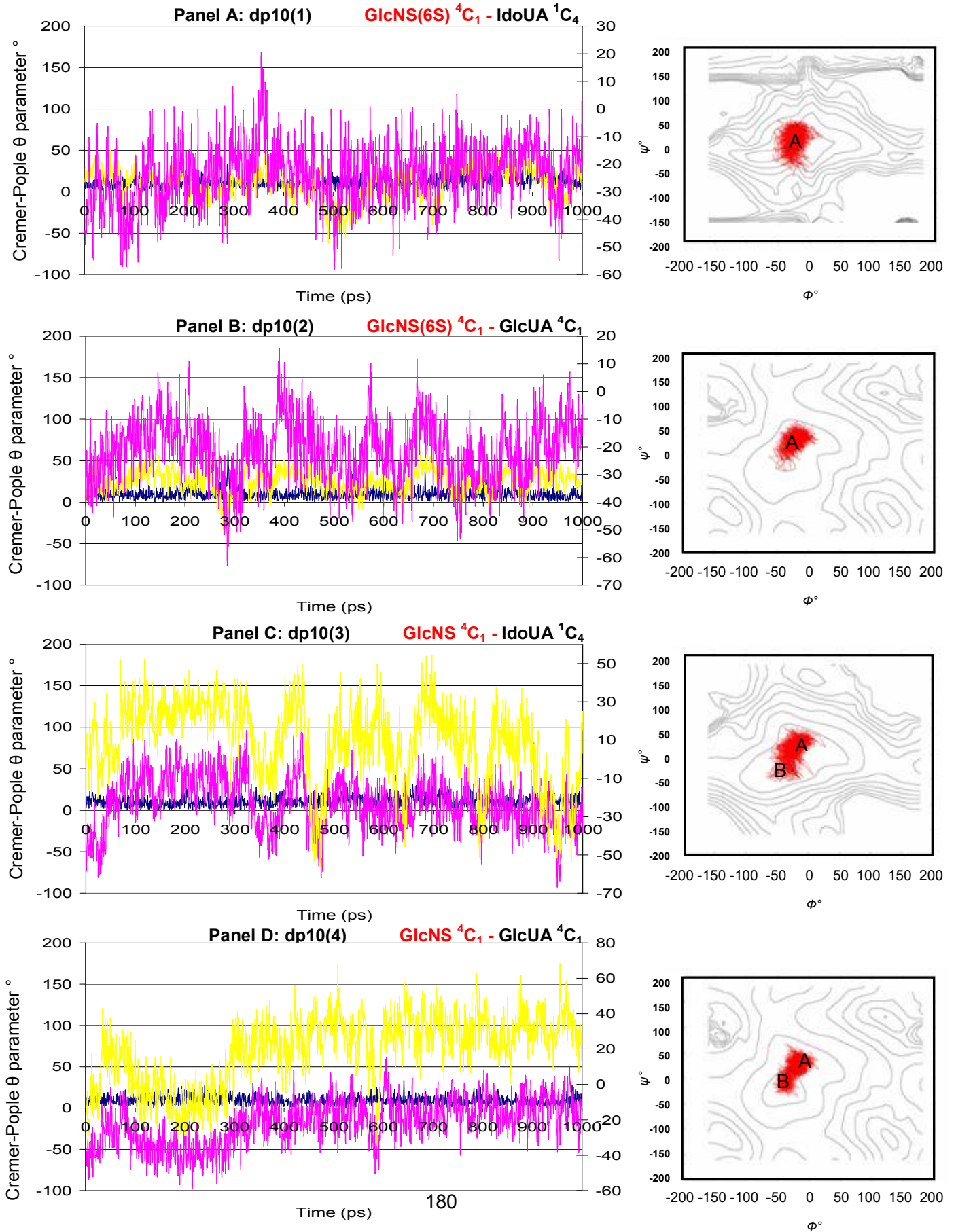
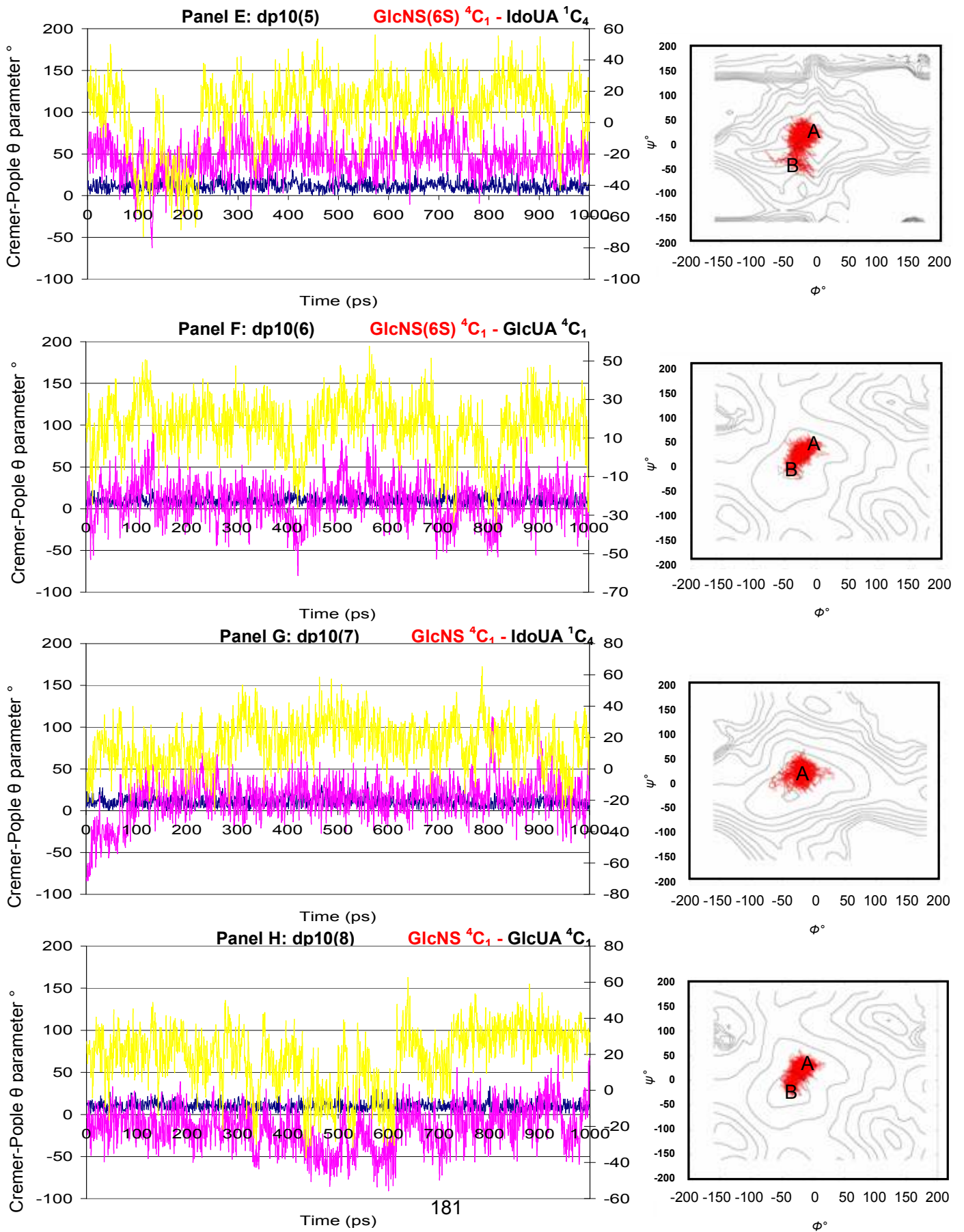
Torsion data: Structures were sampled every picosecond over the course of 1000 ps in TIP3P explicit water. The Φ and Ψ torsion angles across each glycosidic linkage were plotted and connected together in chronological order (). Contour lines are plotted as a guideline reference for each linkage and represent the potential energy surface of each of the glycosidic linkages (the right hand side figure).

Figure 3.18 Active Oligosaccharides



Inactive Oligosaccharides



3.2.4.3 Linkage 7

IdoUA(2S) - GlcNS(6S)	dp10(1 - 2) and dp10(5-6)
IdoUA - GlcNS	dp10(3-4)
IdoUA(2S) - GlcNS	dp10(7-8)

The IdoUA \pm (2S) monosaccharides were all modelled in the 1C_4 conformation prior to the 1 ns MD simulation.

IdoUA(2S) according to literature plays a major role in the binding of the FGF2 protein as well as many other different proteins, the following are just a selection (Ishihara *et al.*, 1994; Kreuger *et al.*, 2001; Maccarana *et al.*, 1993; Pellegrini *et al.*, 2000; Pye *et al.*, 2000; Turnbull *et al.*, 1992; Wu *et al.*, 2003). Many studies have concentrated on the conformation of the internal iduronates and their role in protein binding, the following list will name but a few (Rabenstein, 2002; Hricovini *et al.*, 2001; Guerrini *et al.*, 2006; Raman, 2005; Faham *et al.*, 1996; Fry *et al.*, 1999).

Models dp10(1), dp10(2) dp10(5) and dp10(6) are all comparable because they all contain the same sequence for this specific linkage (IdoUA(2S)-GlcNS(6S)). This forms the kink motif identified by Guglieri *et al* (Guglieri, Hricovini *et al.*, 2008) which forms in the chain when bound to a protein (if taken into account the presence of the GlcNS(6S) residue preceding this sequence):



(Guglieri *et al.*, 2008)

Raman *et al* (Raman *et al.*, 2005) identified this as a kink when working with H – I – H trisaccharides (where H are glucosamines \pm O-sulphation and I is iduronic acid). It is certainly well established that it is the position of the iduronate in a chain which mainly affects the equilibrium of the conformations identified in a chain (Rabenstein, 2002; Faham *et al.*, 1996). 1C_4 and 2S_0 conformations are clearly favoured here in the Goodger models as seven of the eight ring conformations are identified in one or both conformations during this simulation. The 4C_1

conformation is not the favoured conformation as this is only identified in one oligosaccharide - dp10(8).

Both models dp10(1) and dp10(5) are comparable as they have the IdoUA residue for residue 9. They both have a starting conformation of 1C_4 set prior to the MD simulation. Both undergo a conformational transition to the skew boat form. In the active dp10(1) this occurs around 500 ps where as in dp10(5) this happens later on in the simulation, around 800 ps (Figure 3.19, Panels A and E). Here there are clear differences between active and inactive oligosaccharides which could have an influence on the overall biological activity. In this instance the active dp10(1) oligosaccharide exists in two geometries (geometry A and B). This change occurs after the conformational change between 600 – 700 ps. Average values for the angle before this change were $\Phi = -30^\circ$ $\Psi = -30^\circ$ assigned as geometry B. After this transition; typical values increase to $\Phi = +40^\circ$ $\Psi = +10^\circ$ assigned geometry A. In terms of 3-dimensional shape these changes in angle lead to a large change in the overall shape of the structure. The change in the inactive dp10(5) deca-saccharide occurs before the conformational change (it should be noted that the actual time this transition occurs in both models is around the same time, around 600-700 ps in time). Although these models sample both geometries; the active structure undergoes a change in ring conformation first which seems to be the driving force behind the change in torsion angle. In the inactive model this has an opposite effect; it is the change in torsion data which drives the change in ring conformation. Also comparable to these structures are dp10(2) and dp10(6). These contain the same flanking monosaccharides but only exist in one geometry. In this instance the active dp10(2) exists in geometry A, whilst the inactive dp10(6) exists only in geometry B. This could suggest it is the presence of geometry A which influences the skew boat conformation, whilst geometry B influences the 1C_4 conformation. This is a good indication of the difference between the active and inactive structures and could be important in biological activity.

Oligosaccharides dp10(3, 4, 7 and 8) are slightly different in structure, they are not fully sulphated structures. The active dp10(3) and dp10(4) oligosaccharides exist in a single geometry B, with only the 1C_4 conformation present. The ring conformations are identified only in the 1C_4

conformation throughout the whole 1000 ps timescale; possible due to the absence of the 2-O-sulphate group on the iduronate section of the sequence. The angle geometry for dp10(3) remains solely in the geometry labelled B ($\Phi = -40^\circ$ $\Psi = -20^\circ$), with a brief period in a possible transient geometry at the very beginning of the simulation (Figure 3.19, Panel C). The inactive dp10(7) and dp10(8) also exists in one geometry, geometry A. Here it has average values of $\Phi = +60^\circ$ $\Psi = +40^\circ$ and $\Phi = +50^\circ$ $\Psi = +10^\circ$. The 1C_4 conformation undergoes a conformational transition to the 2S_0 conformation after 200 ps.

These results indicate that the presence of either the chair or skew boat conformation has an influence on the corresponding geometry. The chair conformation seems to have an overall influence on geometry B, whilst the skew boat conformation drives geometry A. This could have some bearing on the biological activity of the HS samples. Also there is a slight effect from residue 9 in the chain where the sequences differ in the reducing end terminal. This difference is identified in dp10(2) and dp10(6) which both contain the GlcUA at residue 9 instead of the alternative IdoUA in dp10(1) and dp10(5); here only one conformation in the iduronate at residue 7 is present in the chain and not both conformations which has previously been identified (either 2S_0 or 1C_4 respectfully).



Figure 3.19 Puckering-torsion data for linkage 7

IdoUA(2S) - GlcNS(6S)	dp10(1 - 2) and dp10(5-6)
IdoUA - GlcNS	dp10(3-4)
IdoUA(2S) - GlcNS	dp10(7-8)

Puckering parameters: all graphs show θ analysis ().

Cremer-Pople ring puckering parameters during a 1000 ps molecular dynamic simulation.

All graphs show θ analysis (ϕ_2 not shown) and changes are compared to the values identified in the torsion data in chronological order.

Torsion data: Φ° dihedral angle (), Ψ° dihedral angle ().

The variation in Φ and Ψ torsion angles over a course of 1000 ps explicit water molecular dynamic simulation (the left hand side figure).


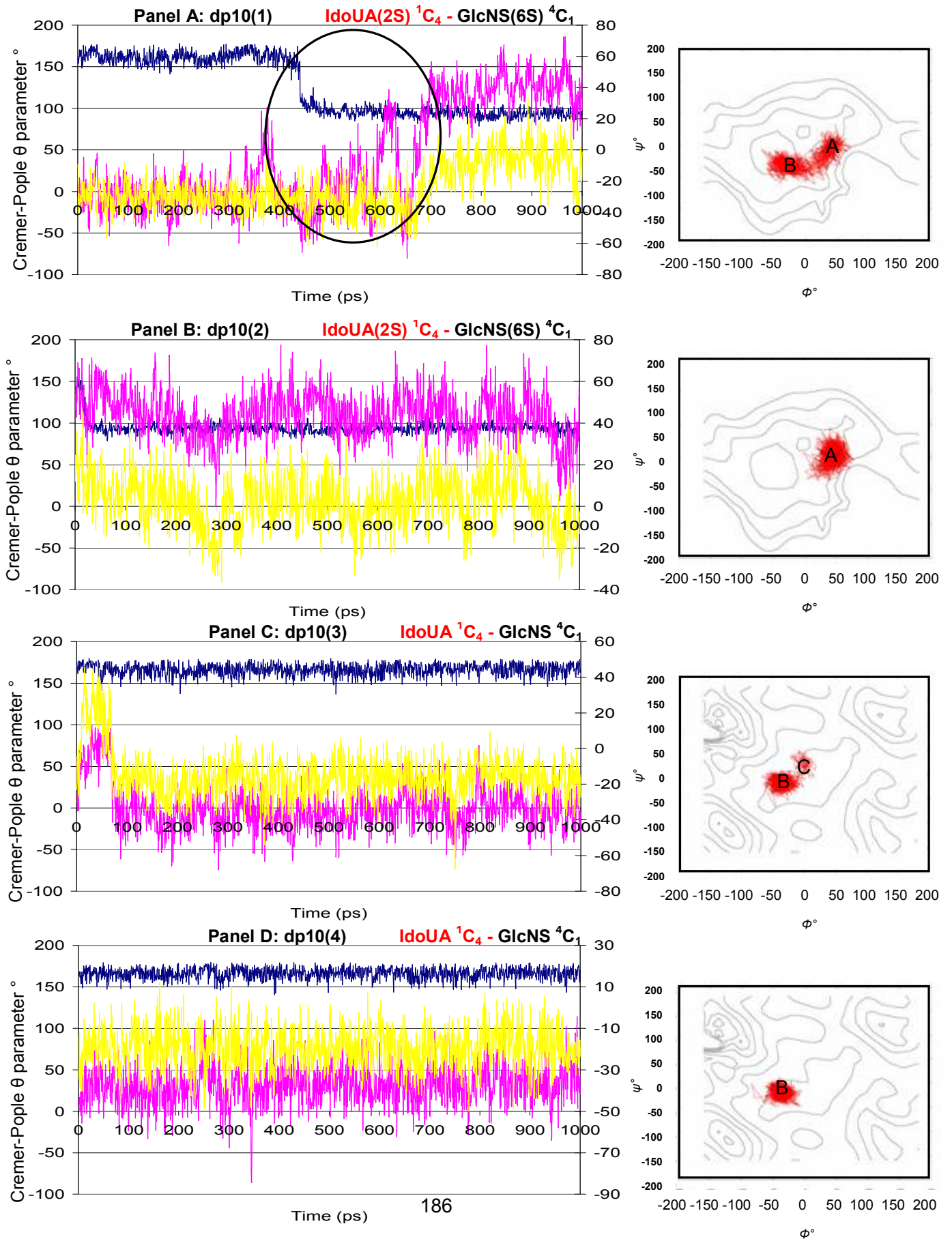
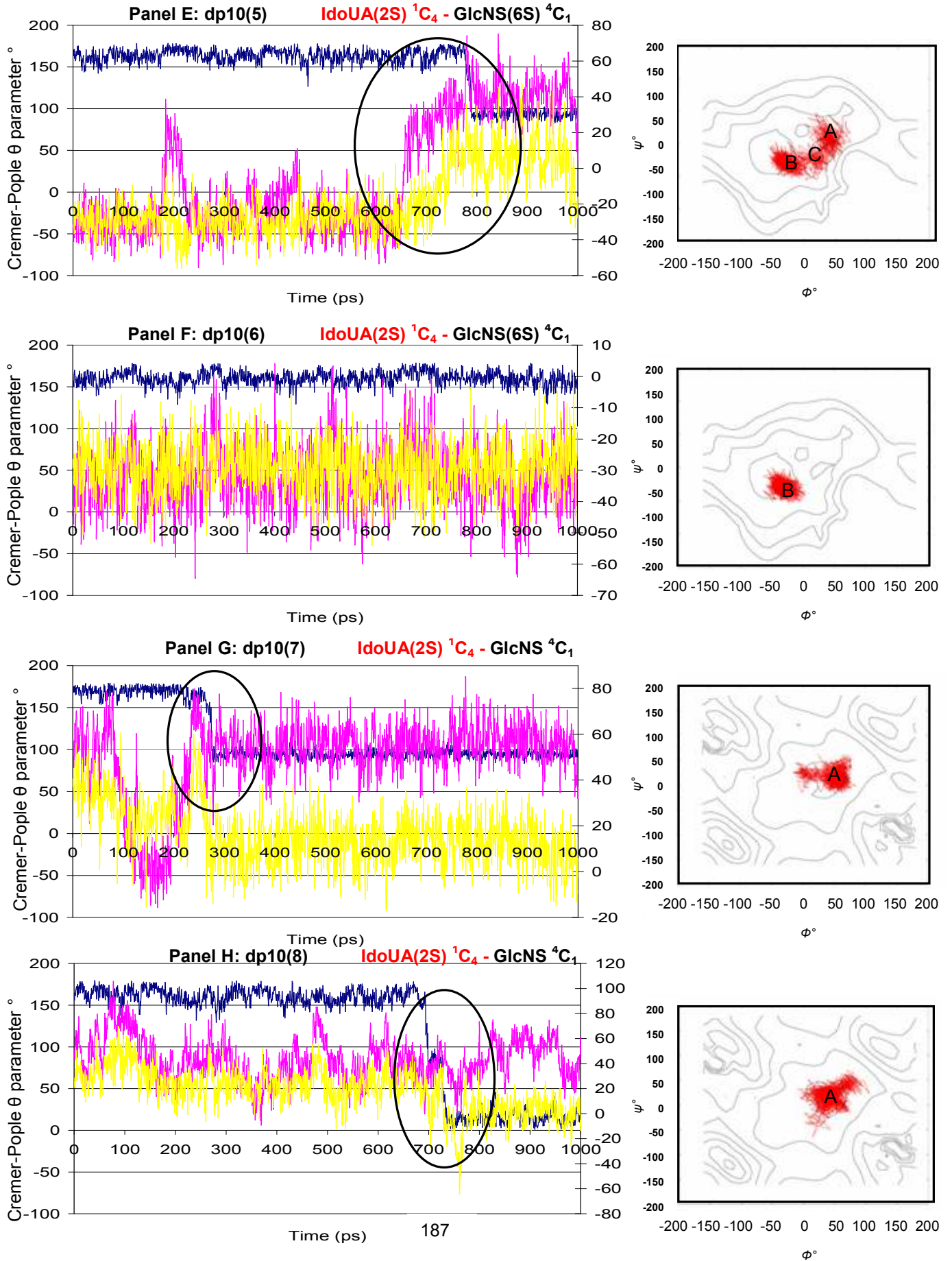
Torsion data: Structures were sampled every picosecond over the course of 1000 ps in TIP3P explicit water. The Φ and Ψ torsion angles across each glycosidic linkage were plotted and connected together in chronological order (). Contour lines are plotted as a guideline reference for each linkage and represent the potential energy surface of each of the glycosidic linkages (the right hand side figure).

Figure 3.19 Active Oligosaccharides



Inactive Oligosaccharides



3.2.4.4 Linkage 6

GlcNS(6S) - IdoUA(2S) dp10(1 - 2) and dp10(5-8)

GlcNS(6S) - IdoUA dp10(3-4)

The GlcNS \pm (6S) monosaccharides were all modelled in the 4C_1 conformation prior to the 1 ns MD simulation.

Like linkage 8, linkage 6 comprises all glucosamines of which all of them contain the additional 6-O-sulphate group. There are two different types of disaccharide which make up linkage 6 across the range of oligosaccharides. For models dp10(1), dp10(2), dp10(5), dp10(6), dp10(7) and dp10(8); this linkage makes up the fully sulphated trisaccharide (GlcNS(6S)-IdoUA(2S)). The IdoUA seen in linkage 7 is unsulphated for dp10(3) and dp10(4).

As previously discussed, the glucosamines here are rigid in the 4C_1 conformation and so are able to dispose of their bulky side groups in an equatorial position (Sanderson *et al.*, 1987). Any change in shape to accommodate binding with a protein will come about from the changes in the torsion angle. Decasaccharide dp10(1) shows the presence of two geometries of which geometry B is the most prominent (Figure 3.20, Panel A) and geometry A which is identified after 800 ps. This corresponds to a similar change in torsion angle in the neighbouring linkage (Figure 3.19, Panel A). The same is identified in the inactive dp10(5) with the change in the torsion angle occurring around the same time (Figure 3.20, Panel E). These both have geometries with average values of $\Phi = -40^\circ$ $\Psi = -20^\circ$ (geometry B) and $\Phi = +30^\circ$ $\Psi = +30^\circ$ (geometry A). This change in the dihedral angle could be due to the presence of the IdoUA at the terminal end (residue 9) as those sequences with the GlcUA do not show this change in geometry and are therefore only identified in geometry B (Figure 3.20, Panels B and F). Again there is a possible influence from the IdoUA at position 9; as oligosaccharides which contain IdoUA instead of GlcUA at position 9 in the chain exist in both geometries A and B. Dp10(7) and dp10(8) show similar Φ and Ψ values as those observed with the same Ψ geometry present (geometry B); dp10(2) and dp10(6). Again there is no obvious influence by the iduronic acid present at position 9 (Figure 3.20, Panels G and H). The presence of and the combination of both the 2-O-sulphate and 6-O-

sulphate groups in this linkage could be a contributing factor to the transient geometry identified in linkage 7.

The main difference in the active and inactive oligosaccharides is the presence of a third geometry in oligosaccharides dp10(3) and dp10(4). This has been identified as geometry C (Figure 3.20, Panels C and D). This occupies a previously un-observed geometry which is found in the top left hand quadrant of the contour map with average values around $\Phi = -20^\circ$ $\Psi = +30^\circ$ for dp10(3) and $\Phi = -30^\circ$ $\Psi = -20^\circ$ for dp10(4). The difference in the angle geometry could be due to the lack of 2-O-sulphate on the neighbouring residue (the IdoUA discussed in linkage 7). The presence or absence of IdoUA here can be ruled out due to both models existing in the same geometry. The presence of this additional geometry does not tie in with those identified in the other active sequences so it is difficult to say whether this has any bearing on the biological activity of these samples.



Figure 3.20 Puckering-torsion data for linkage 6

GlcNS(6S) - IdoUA(2S) dp10(1 - 2) and dp10(5-8)
GlcNS(6S) - IdoUA dp10(3-4)

Puckering parameters: all graphs show θ analysis ().

Cremer-Pople ring puckering parameters during a 1000 ps molecular dynamic simulation.

All graphs show θ analysis (ϕ_2 not shown) and changes are compared to the values identified in the torsion data in chronological order.

Torsion data: Φ° dihedral angle (), Ψ° dihedral angle ().

The variation in Φ and Ψ torsion angles over a course of 1000 ps explicit water molecular dynamic simulation (the left hand side figure).


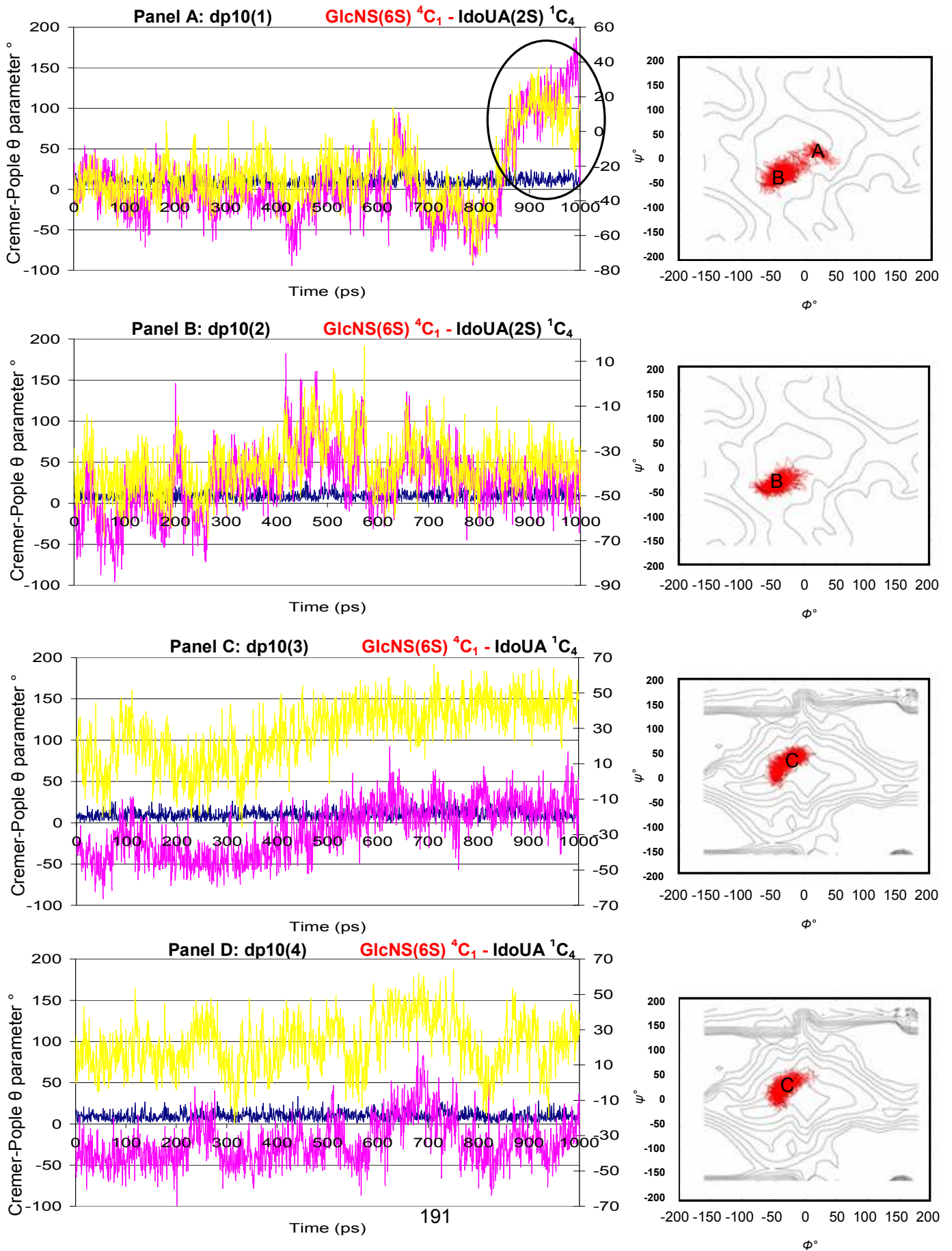
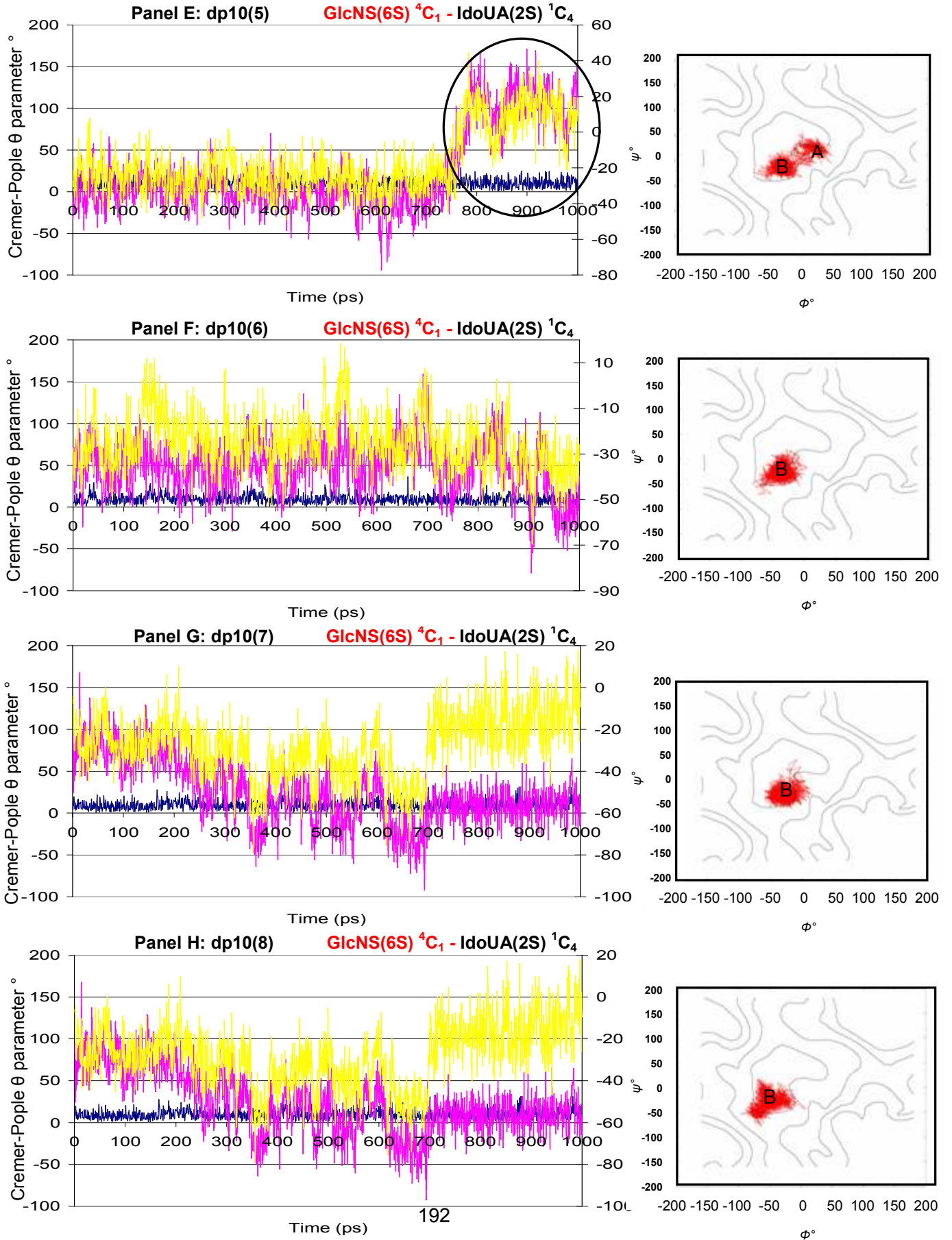
Torsion data: Structures were sampled every picosecond over the course of 1000 ps in TIP3P explicit water. The Φ and Ψ torsion angles across each glycosidic linkage were plotted and connected together in chronological order (). Contour lines are plotted as a guideline reference for each linkage and represent the potential energy surface of each of the glycosidic linkages (the right hand side figure).

Figure 3.20 Active Oligosaccharides



Inactive Oligosaccharides



3.2.4.5 Linkage 5

IdoUA(2S) – GlcNS(6S) dp10(1 - 8)

The IdoUA \pm (2S) monosaccharides were all modelled in the 1C_4 conformation prior to the 1 ns MD simulation.

The trisulphated disaccharide (IdoUA(2S)-GlcNS(6S)) makes up this linkage in all eight Goodger decasaccharide models and so all structures are comparable. It is this sequence, alongside residue 7 which potentially makes up the binding site for FGF2 binding. Here the presence of the 2-O-sulphate and the 6-O-sulphate groups are important in this specific interaction as previously identified in a number of studies (Maccarana *et al.*, 1993; Faham *et al.*, 1996; DiGabriele *et al.*, 1998; Wu *et al.*, 2003).

The main difference between the active and inactive oligosaccharides appears to involve all eight oligosaccharides and is in respect to the different torsion angles. In three out of the four active models, two geometries exist (A and B), with the exception being dp10(4), which exists in only geometry B. The inactive oligosaccharides, with the exception of dp10(8) exist in only one geometry (either geometry A or B). The average data values for geometry A being $\Phi = +25^\circ$ $\Psi = 0^\circ$ and for geometry B $\Phi = -30^\circ$ $\Psi = -20^\circ$. This difference could be significant and have an overall influence on biological activity.

The corresponding Cremer-Pople Puckering parameter data for both dp10(1) and dp10(2) indicates no change in conformation so the ring has not undergone any conformational changes (Figure 3.21 Panels A and B). At the end of the simulation both dp10(1) and dp10(2) show similar conformational patterns to those described by Faham *et al* (Faham *et al.*, 1996). The IdoUA(2S) of position 5 remains in the 1C_4 conformation and the IdoUA(2S) of position 7 is in the 2S_0 conformation. The presence of 2-O-sulphate is thought to stabilize the 1C_4 conformation (van_Boeckel *et al.*, 1987) and this could be one of the reasons why the iduronate is unchanged in this conformation. The 6-O-sulphate on the flanking residues could also play an important role in the conformation of the IdoUA(2S) at this point in the HS chain (Pye *et al.*, 1998).

For dp10(3) the IdoUA(2S) begins this simulation in the 2S_0 conformation (suggesting it had already gone through one conformational transition previous to this MD simulation) with a conformational transition to 4C_1 after 300 ps. Up until this point, the dihedral angle data suggests it exists in geometry B (Figure 3.21, Panel C). The conformational transition and the presence of the skew boat conformation may well be the driving force behind the switch in torsion data to the alternative geometry (geometry A). This is suggested by the greater fluctuation observed in the data values for Φ and Ψ then has been previously identified (average values for these are $\Phi = +130^\circ$ $\Psi = +30^\circ$ at the highest value and $\Phi = -80^\circ$ $\Psi = -50^\circ$ on average at the lowest value). The start of the MD simulation for dp10(4) shows an average puckering parameter data value of $\theta = 141.7^\circ$; which does not fit either the 2S_0 or the 1C_4 conformation values observed in this study and therefore suggests the presence of an intermediate conformation (Figure 3.21, Panel D). A similar angle is also identified in dp10(7) discussed below. An intermediate conformation has previously been identified when a fully sulphated heparin oligosaccharide binds the FMDV- O_1 BFS protein. Here a terminal iduronate was found to adopt a ${}^{2.5}B$ conformation (an intermediate boat conformation) (Fry *et al.*, 1999); which can be mapped on the surface of a globe with the poles at $\theta = 0^\circ$ or 180° , corresponding to the canonical chair conformations (Hasnoot, 1992). Mikhailov *et al* (Mikhailov *et al.*, 1996) also suggests the ${}^{2.5}B$ conformation exists in equilibrium with the three other conformations (1C_4 , 2S_0 and 4C_1). The presence of the 4C_1 conformation in dp10(3) is the second time it has been identified in the Goodger models for an internal residue (the first being for residue 7 in dp10(8); Figure 3.21, Panel H). Initially it was thought to occur only in the external residues (Ferro *et al.*, 1990).

The main differences in the inactive oligosaccharides is the presence of only one torsional geometry and also the presence of more than one ring conformation during the 1 ns MD simulation. The potentially important, active conformation (1C_4) is present but there are further transitions made. Firstly for dp10(5) the ring conformation changes around 200 ps from the starting conformation to the 2S_0 conformation, but only exists in geometry A (which has similar data values to those observed in the active oligosaccharides). This change in conformation is

similar for dp10(7); but occurs over a longer timescale. Again after the 200 ps mark, the θ value slowly decreases until the 2S_0 conformation value is reached ($\theta = 92^\circ$). Here the only geometry observed is geometry B.

The presence of the IdoUA in oligosaccharide dp10(6) (the residue identified at position 9 as apposed to the alternative GlcUA) has an influence on the chain as the IdoUA(2S) at position 5 is initially identified in the skew boat conformation suggesting it has already undergone a conformational transition during the minimisation process (see Experimental Materials and Procedures). After 200 ps there is a further transition to the original 1C_4 conformation. The torsion angle is only observed in geometry B (again with similar data values to those observed in the active oligosaccharides, $\phi = -30^\circ$ $\psi = -20^\circ$). Again further evidence for this influence from the GlcUA (from position 9 in the oligosaccharide) is observed in dp10(8), with not one, but two conformational transitions observed during the 1 ns MD simulation. Here the IdoUA(2S) flips from the starting 1C_4 conformation to the 2S_0 conformation and then back again (Figure 3.21, Panel H). It is this conformational change which influences the change in the dihedral angle which is the only one of the inactive oligosaccharides to exist in both geometries.

Figure 3.21 Puckering-torsion data for linkage 5

IdoUA(2S) – GlcNS(6S) dp10(1 - 8)

Puckering parameters: all graphs show θ analysis ().

Cremer-Pople ring puckering parameters during a 1000 ps molecular dynamic simulation.




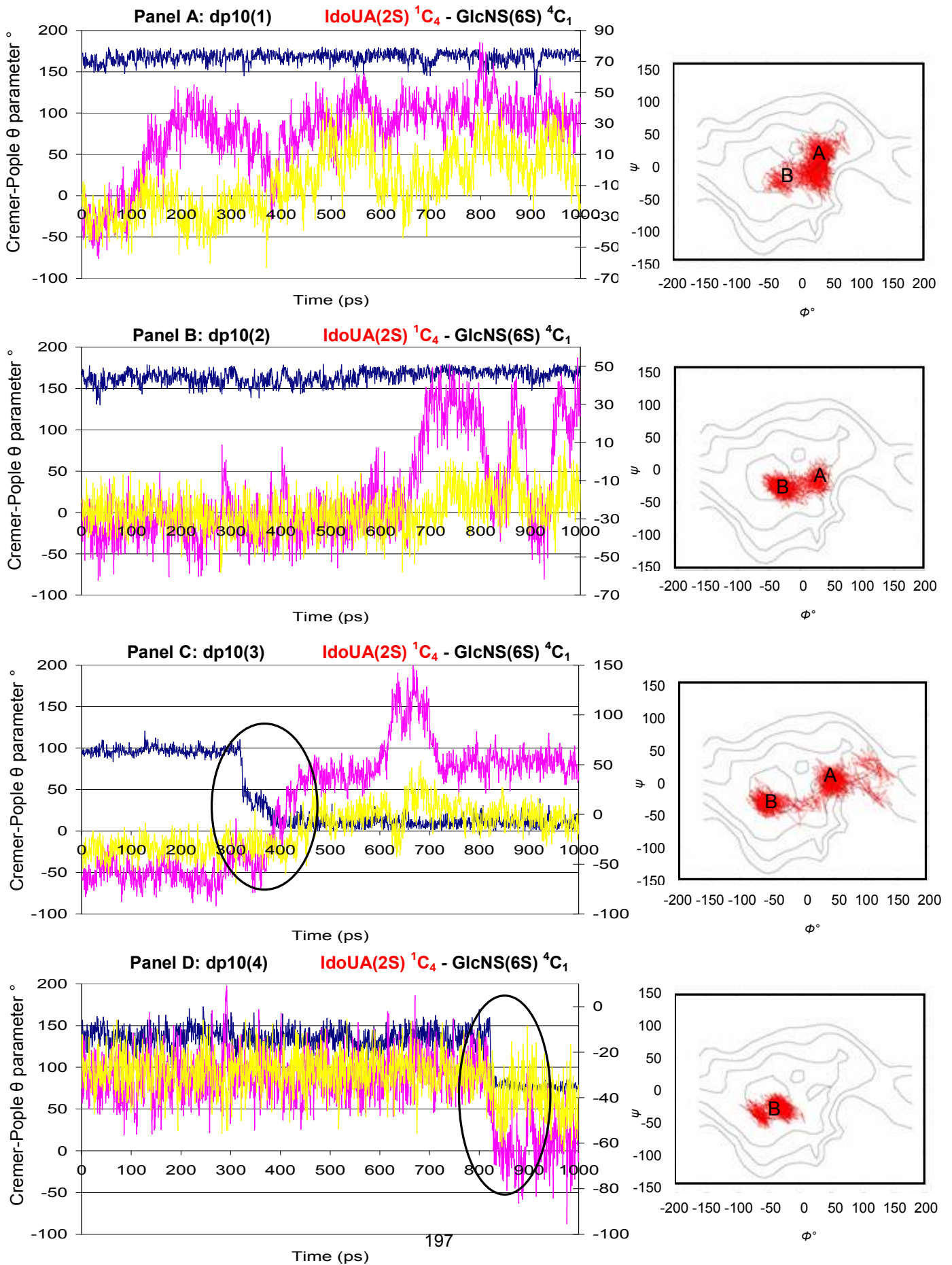
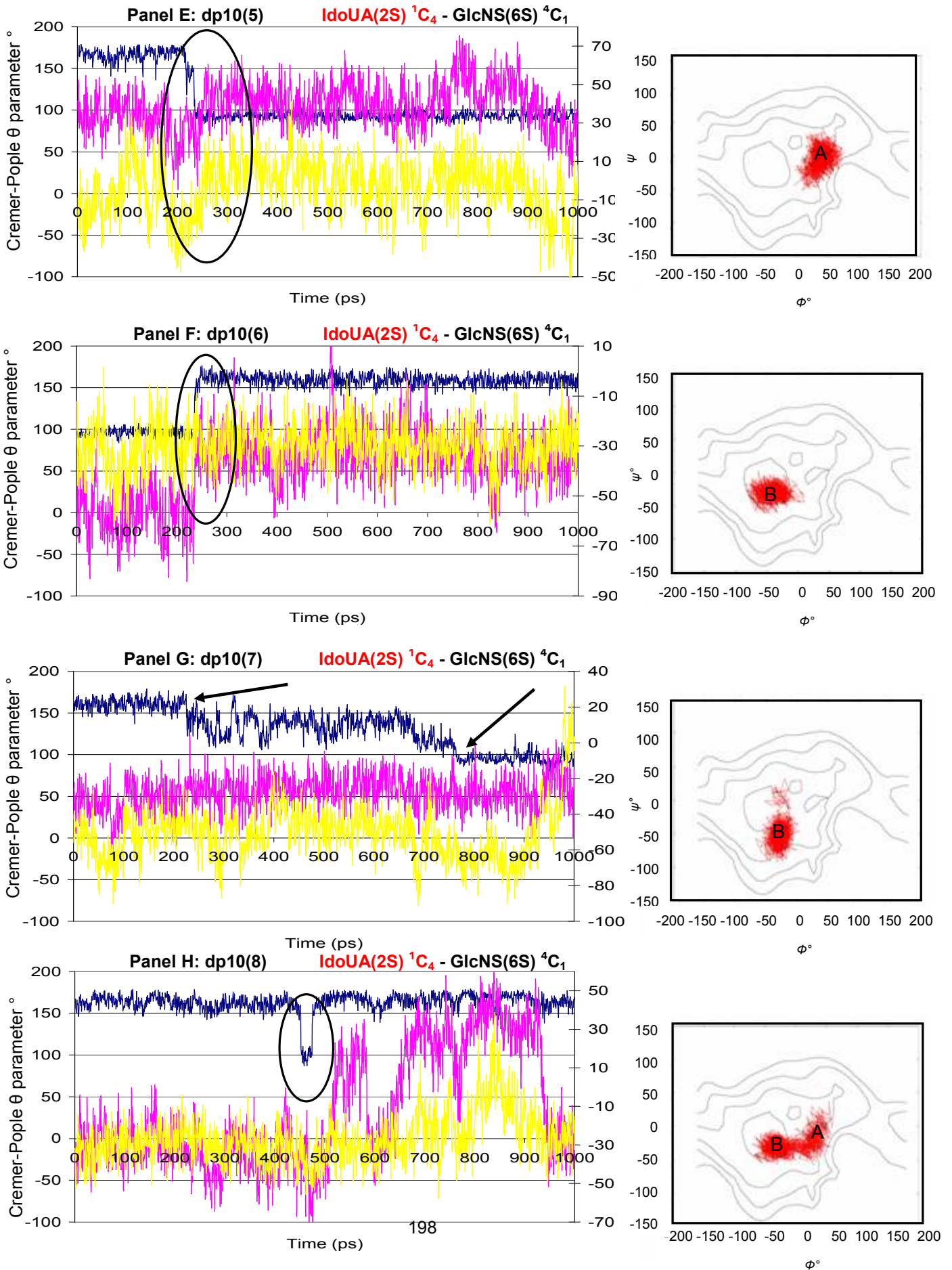
All graphs show θ analysis (ϕ_2 not shown) and changes are compared to the values identified in the torsion data in chronological order.**Torsion data:** Φ° dihedral angle (), Ψ° dihedral angle ().The variation in Φ and Ψ torsion angles over a course of 1000 ps explicit water molecular dynamic simulation (the left hand side figure).**Torsion data:** Structures were sampled every picosecond over the course of 1000 ps in TIP3P explicit water. The Φ and Ψ torsion angles across each glycosidic linkage were plotted and connected together in chronological order (). Contour lines are plotted as a guideline reference for each linkage and represent the potential energy surface of each of the glycosidic linkages (the right hand side figure).

Figure 3.21 Active Oligosaccharides



Inactive Oligosaccharides



3.2.4.6 Linkage 4

GlcNS(6S) - IdoUA(2S)	dp10(1 - 2)
GlcNS - IdoUA(2S)	dp10(3 - 8)

The GlcNS \pm (6S) monosaccharides were all modelled in the 4C_1 conformation prior to the 1 ns MD simulation. Again all glucosamines are found only in this conformation and so any flexibility in these linkages will come from changes in the torsion angles and not the conformational shifts in the ring structure. There are two types of disaccharide which make up linkage 4 across the range of oligosaccharides. Linkage 4 identified in dp10(1) and dp10(2) is not comparable to any of the other linkages at this position as it contains the sequence GlcNS(6S)-IdoUA(2S). Here the only geometry observed in these oligosaccharides is geometry B with average data values of $\Phi = -30^\circ$ $\Psi = -30^\circ$ and $\Phi = -30^\circ$ $\Psi = -20^\circ$ respectfully (Figure 3.22, Panels A and B).

The second of the two disaccharide sequences identified in this linkage is GlcNS-IdoUA(2S); which makes up the remaining six oligosaccharides. Here the only difference between active and inactive oligosaccharides is identified in dp10(7) and dp10(8). Here they behave in a similar way to the active dp10(1) and dp10(2) in that they only exist in geometry B, with similar data values to those mentioned above. This suggests that the presence of the 6S group actually does not have an effect on the geometry of the dihedral angle as it appears to have no immediate effect on the ring structure which has previously been identified in other studies. An important binding site with high affinity for FGF2 involving a 6S group has been identified in this region of the chain in previous studies (Faham *et al.*, 1996). This sequence also forms part of the kink motif, identified previously (Guglieri *et al.*, 2008), so it is possible that there is some effect on biological activity but that its effects might have an influence further down the chain.

Oligosaccharides dp10(3), dp10(4), dp10(5) and dp10(6) all show similar torsion data profiles, with the existence of both geometry A and B (Figure 3.22, Panels C, D, E and F). Geometry B has similar data values to those previously mentioned, and geometry A is characterised with values of $\Phi = +30^\circ$ $\Psi = +10^\circ$. As it stands; these results have shown no significant bearing on the biological activity of the samples in respect to active and inactive oligosaccharides.



Figure 3.22 Puckering-torsion data for linkage 4

GlcNS(6S) - IdoUA(2S)	dp10(1 - 2)
GlcNS - IdoUA(2S)	dp10(3 - 8)

Puckering parameters: all graphs show θ analysis ().

Cremer-Pople ring puckering parameters during a 1000 ps molecular dynamic simulation.

All graphs show θ analysis (ϕ_2 not shown) and changes are compared to the values identified in the torsion data in chronological order.

Torsion data: Φ° dihedral angle (), Ψ° dihedral angle ().

The variation in Φ and Ψ torsion angles over a course of 1000 ps explicit water molecular dynamic simulation (the left hand side figure).


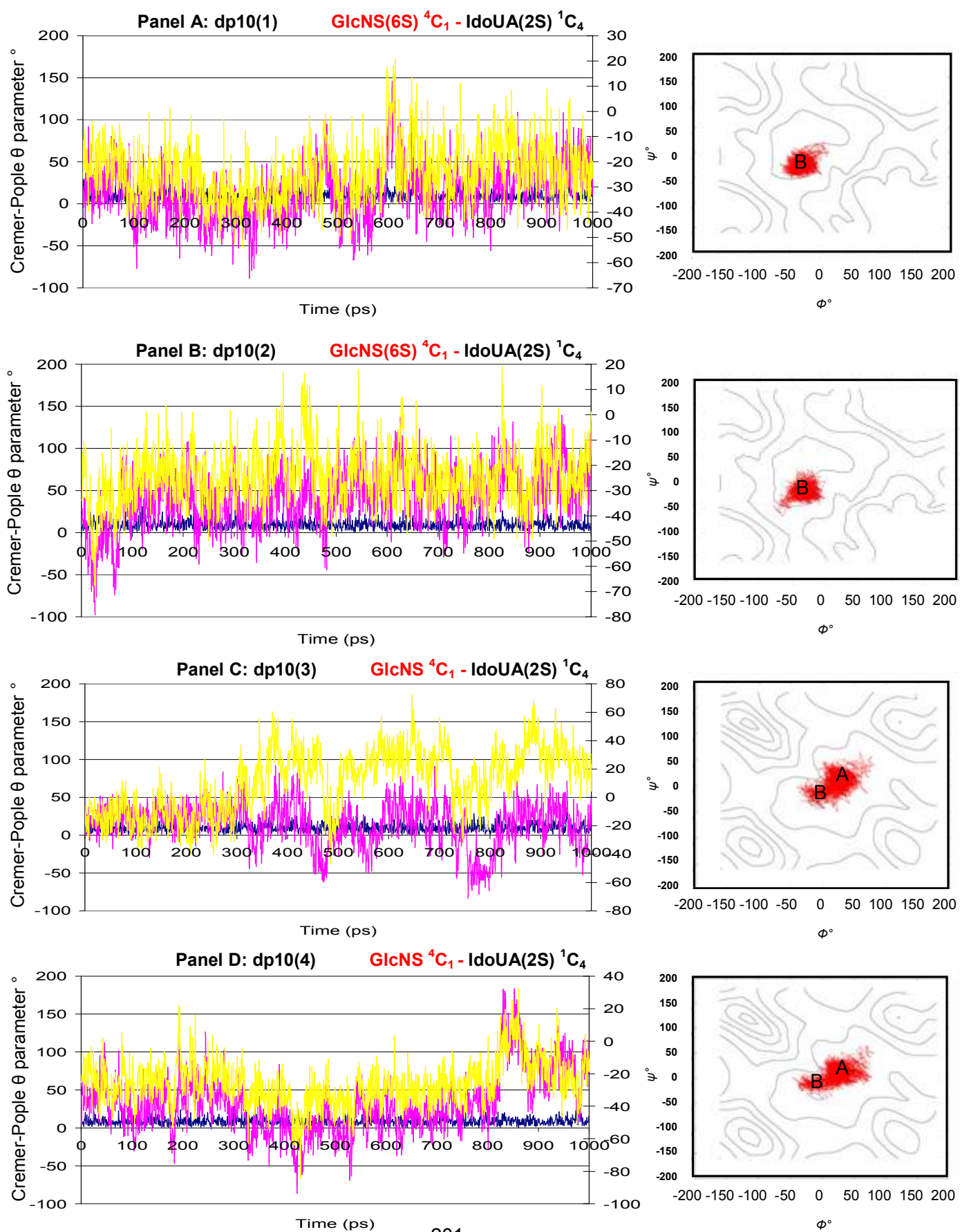
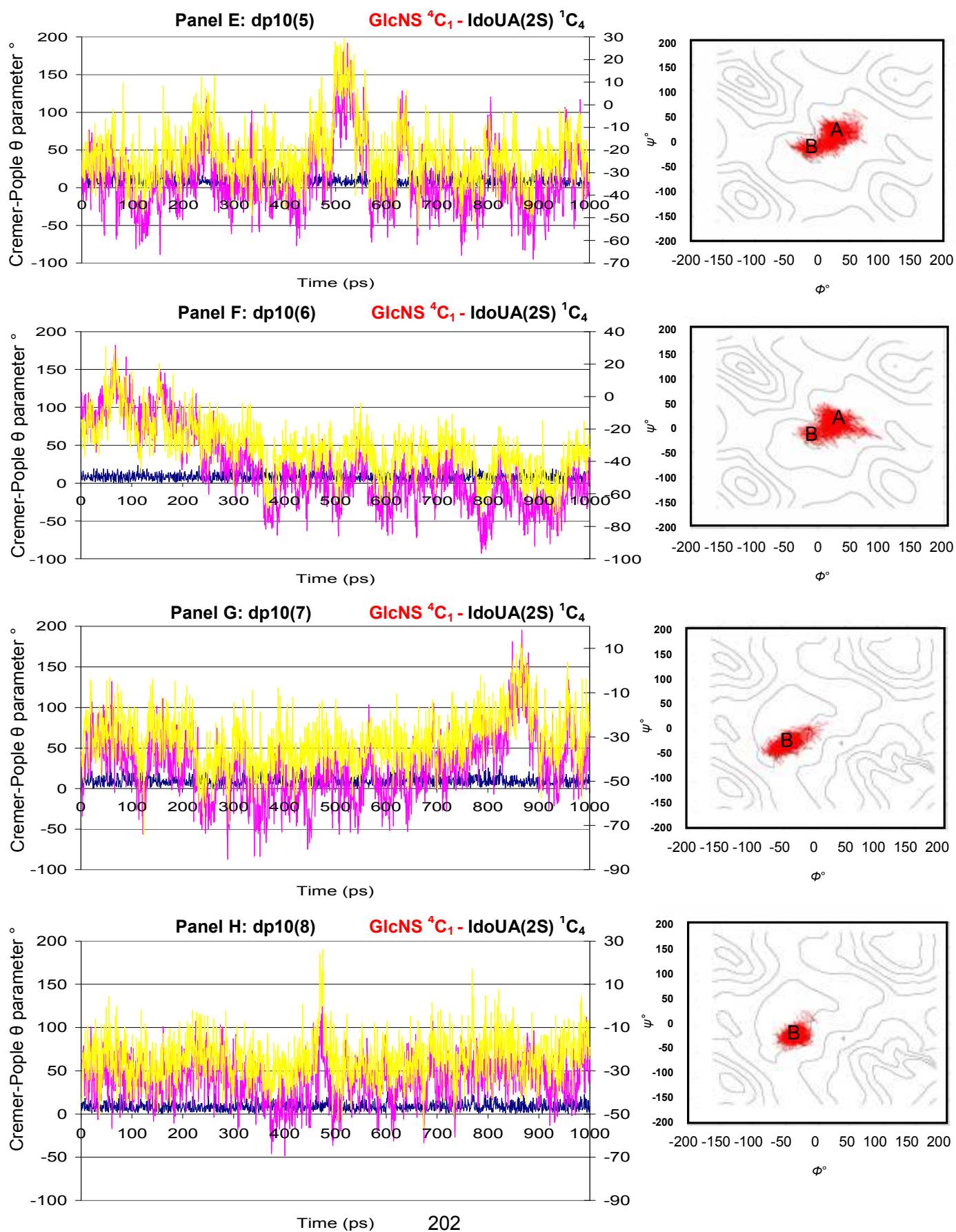
Torsion data: Structures were sampled every picosecond over the course of 1000 ps in TIP3P explicit water. The Φ and Ψ torsion angles across each glycosidic linkage were plotted and connected together in chronological order (). Contour lines are plotted as a guideline reference for each linkage and represent the potential energy surface of each of the glycosidic linkages (the right hand side figure).

Figure 3.22 Active Oligosaccharides



Inactive Oligosaccharides



3.2.4.7 Linkage 3

GlcUA – GlcNS(6S)	dp10(1-2)
GlcUA – GlcNS	dp10(3-6)
IdoUA – GlcNS	dp10(7-8)

The GlcUA monosaccharides were all modelled in the 4C_1 conformation with the IdoUA monosaccharides modelled in the 1C_4 conformation prior to the 1 ns MD simulation. As in previous linkages all flexibility will come from the torsion angle and not from the conformational shifts of the ring structure (with the exception being oligosaccharides dp10(7) and dp10(8), as these contain the iduronate monosaccharide as the flanking monosaccharide).

There are no significant differences between active and inactive oligosaccharides, but this could be due to the diversity of the disaccharides sequences which make up this linkage in all eight oligosaccharides. Those linkages in dp10(1) and dp10(2) exist only in geometry A with average values around $\Phi = +30^\circ$ $\Psi = 0^\circ$. They are the only oligosaccharides which exist solely in this geometry. This could be due to the presence of the 6S group on the preceding residue, but it is unclear if this a determining factor for biological activity.

Those sequences with the sequence GlcUA-GlcNS all exist in two geometries, A and B (Figure 3.23, Panels A – F). Geometry B is characterised with more negative data values than those identified in geometry A ($\Phi = -10^\circ$ $\Psi = -10^\circ$). This rules this out as a potential difference between the active and inactive structures.

The addition of the iduronate at this position, as identified in oligosaccharides dp10(7) and dp10(8) has no major influence on the torsion angle, compared to those sequences mentioned above. Dp10(7) behaves in the same way as dp10(3-6), existing in both geometries (Figure 3.23, Panel G). Dp10(8) exists only in geometry B and is the only oligosaccharides to do so (Figure 3.23, Panel H).



Figure 3.23 Puckering-torsion data for linkage 3

GlcUA – GlcNS(6S)	dp10(1-2)
GlcUA – GlcNS	dp10(3-6)
IdoUA – GlcNS	dp10(7-8)

Puckering parameters: all graphs show θ analysis ().

Cremer-Pople ring puckering parameters during a 1000 ps molecular dynamic simulation.

All graphs show θ analysis (ϕ_2 not shown) and changes are compared to the values identified in the torsion data in chronological order.

Torsion data: Φ° dihedral angle (), Ψ° dihedral angle ().

The variation in Φ and Ψ torsion angles over a course of 1000 ps explicit water molecular dynamic simulation (the left hand side figure).


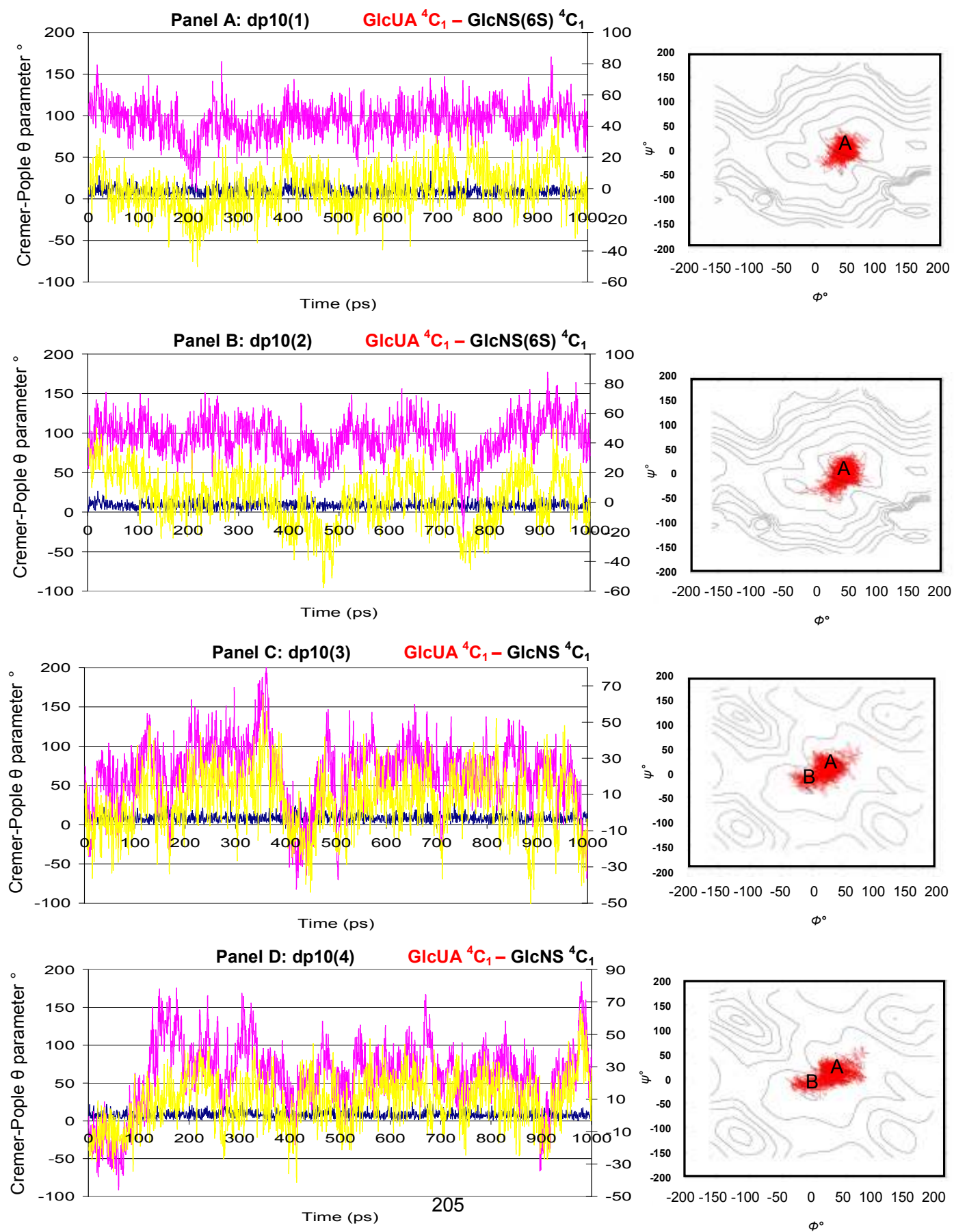
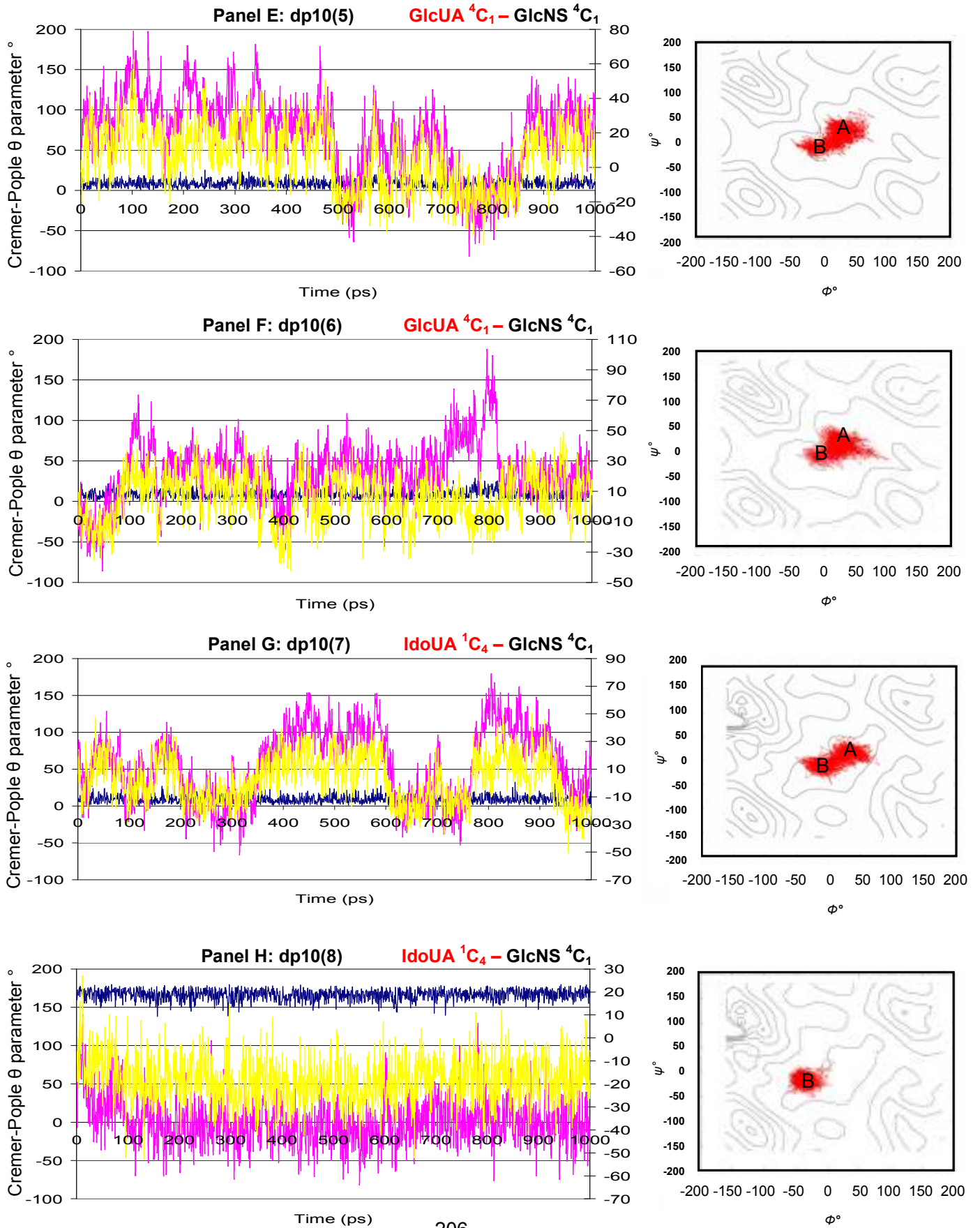
Torsion data: Structures were sampled every picosecond over the course of 1000 ps in TIP3P explicit water. The Φ and Ψ torsion angles across each glycosidic linkage were plotted and connected together in chronological order (). Contour lines are plotted as a guideline reference for each linkage and represent the potential energy surface of each of the glycosidic linkages (the right hand side figure).

Figure 3.23 Active Oligosaccharides



Inactive Oligosaccharides



3.2.4.8 Linkage 2

GlcNAc – GlcUA	dp10(1-6)
GlcNS – IdoUA	dp10(7-8)

The GlcNAc and GlcNS monosaccharides were all modelled in the 1C_4 conformation prior to the 1 ns MD simulation. There are two different types of disaccharide which make up linkage 2 across the range of oligosaccharides. The first is found in six of the eight oligosaccharides (dp10(1-6)) and involves the N-acetylated glucosamine (GlcNAc). Again the Cremer-Pople ring puckering parameters identify all the structures in the rigid 4C_1 conformation. This is the most stable conformation identified at the non-reducing end (Coombe and Kett, 2005). Here all the non-hydrogen ring substituents are found in an equatorial position, apart from the anomeric OH group (Rabenstein, 2002). When GlcNAc is replaced by GlcNS as found in dp10(7) and dp10(8), there are no changes in the ring conformation (Figure 3.24, Panels G and H).

There are no major differences between active and inactive oligosaccharides. All eight exist in geometry A, with constant, average values of $\Phi = -25^\circ$ $\Psi = +20^\circ$. There is no difference between those with the IdoUA at the reducing end and those oligosaccharides with a GlcUA at the reducing end. Oligosaccharides dp10(6) and dp10(7) have an additional B geometry with average values of $\Phi = -50^\circ$ $\Psi = -20^\circ$.

Previous studies have highlighted the lack of flexibility at the non-reducing terminal end and so fit in with the results identified in this study (Coombe and Kett, 2005; Hricovini and Bízík, 2007; Mikhailov *et al.*, 1997; Rabenstein, 2002). Lack of variety in the dihedral angle could be due to the steric interactions that limit the possible Φ° and Ψ° angles available for the glucosamines (Rabenstein, 2002). With a high proportion of glucosamines showing considerably less biological activity (Mikhailov *et al.*, 1997); when compared to the iduronates. This would agree with studies which indicate that the conformation of the IdoUA(2S) influences the geometry of the flanking GlcNS(6S) residues on the non-reducing side of the chain (Coombe and Kett, 2005; Hricovini and Bízík, 2007).



Figure 3.24 Puckering-torsion data for linkage 2

GlcNAc – GlcUA	dp10(1-6)
GlcNS – IdoUA	dp10(7-8)


Puckering parameters: all graphs show θ analysis ().

Cremer-Pople ring puckering parameters during a 1000 ps molecular dynamic simulation.

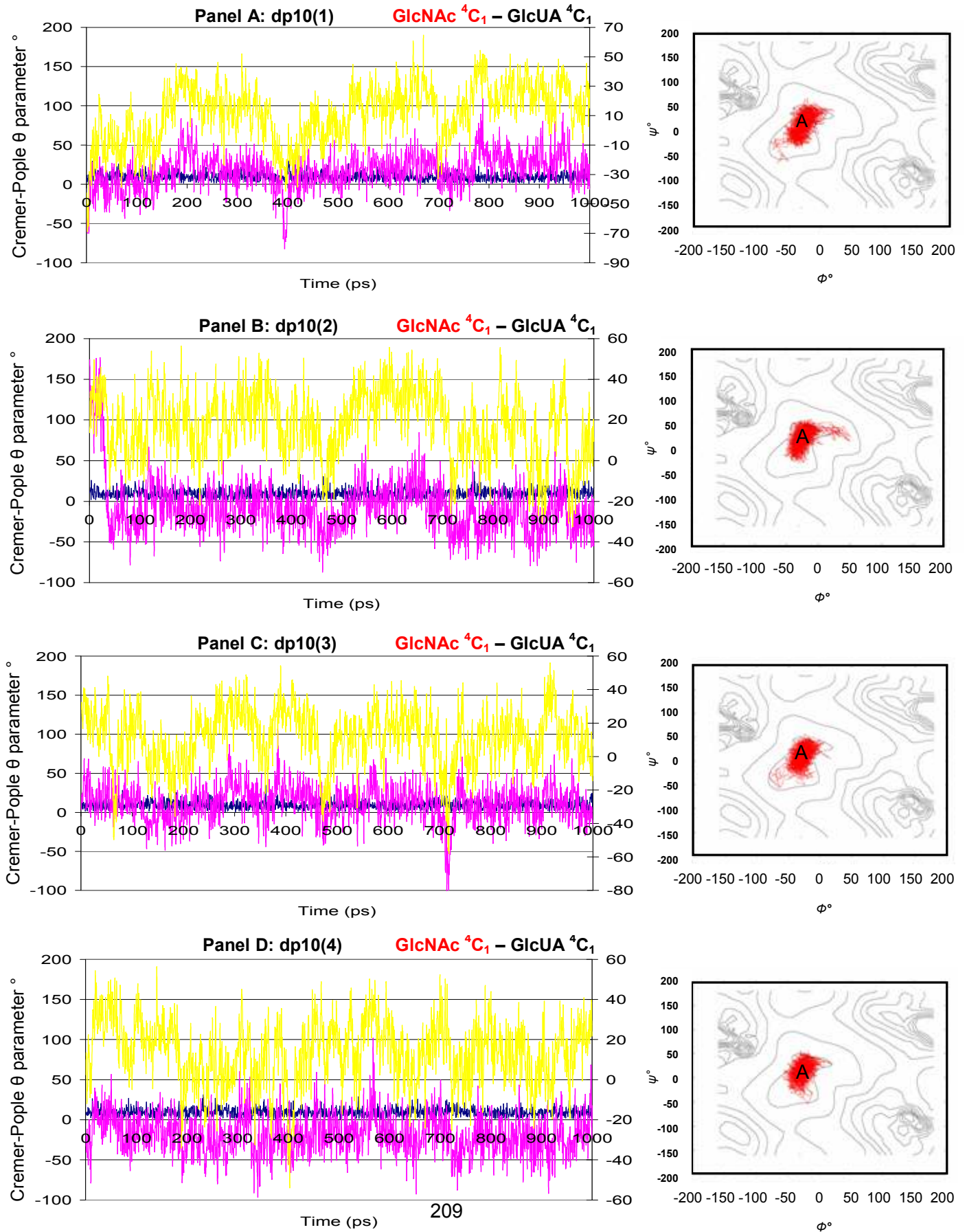
All graphs show θ analysis (ϕ_2 not shown) and changes are compared to the values identified in the torsion data in chronological order.

Torsion data: Φ° dihedral angle (), Ψ° dihedral angle ().

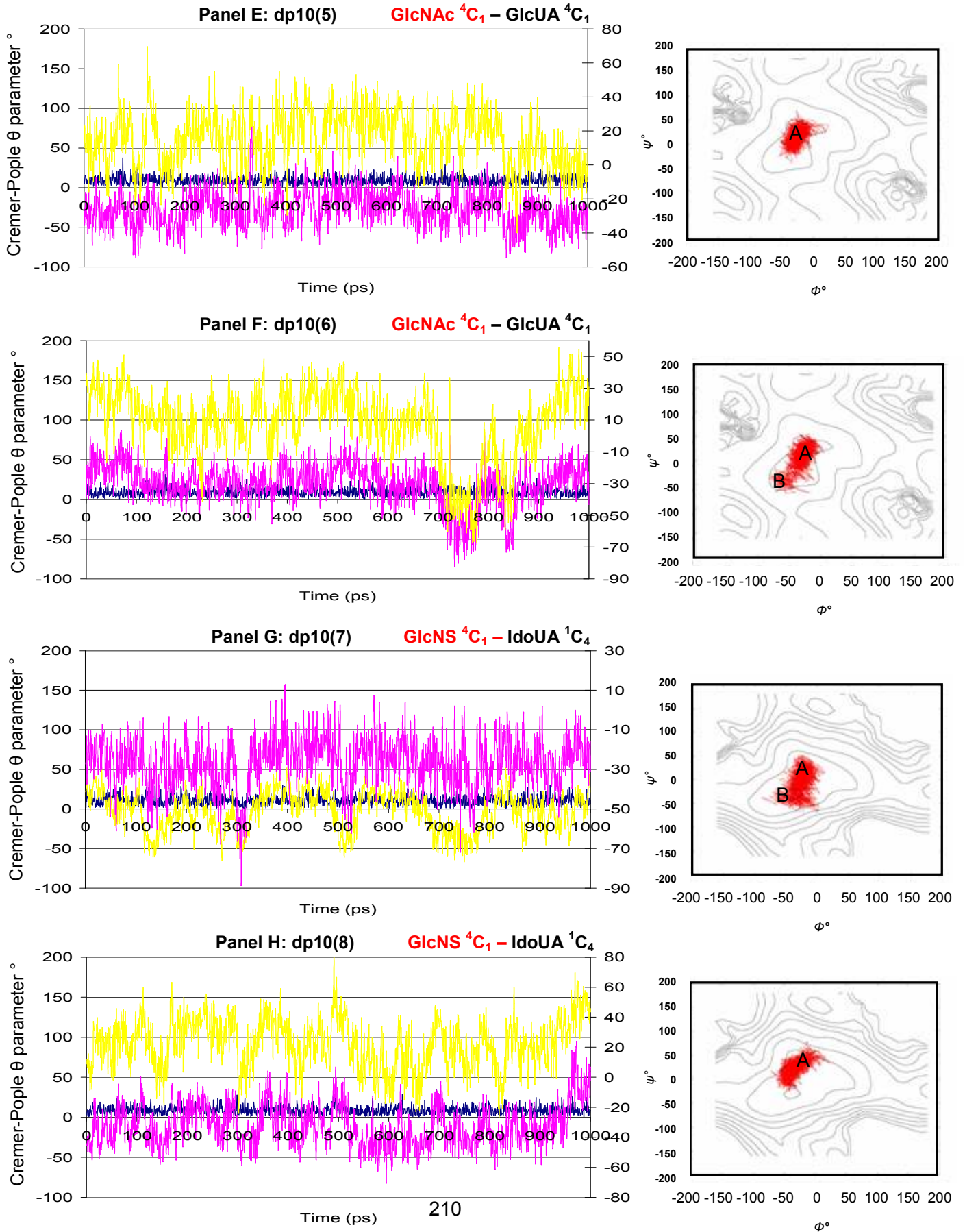
The variation in Φ and Ψ torsion angles over a course of 1000 ps explicit water molecular dynamic simulation (the left hand side figure).

Torsion data: Structures were sampled every picosecond over the course of 1000 ps in TIP3P explicit water. The Φ and Ψ torsion angles across each glycosidic linkage were plotted and connected together in chronological order (). Contour lines are plotted as a guideline reference for each linkage and represent the potential energy surface of each of the glycosidic linkages (the right hand side figure).

Panel 3.24 Active Oligosaccharides



Inactive Oligosaccharides



3.2.4.9 Linkage 1 – the non-reducing end

UA - GlcNAc dp10(1-6)

UA - GlcNS dp10(7-8)

All terminal Δ UA conformations were modelled in the ${}^1\text{H}_2$ starting conformation prior to the 1 ns MD simulation.

UA residues present at the non-reducing end of an oligosaccharide are the result of either chemical or enzymatic degradation of a longer chain. The average Cremer-Pople ring puckering parameter for this conformation was $\theta = 130^\circ$. There are two different types of disaccharide which make up linkage 1 across the range of oligosaccharides.

There are no differences between active and inactive structures with respect to ring conformation and torsional geometry; indicating the non-reducing terminal plays little or no part in biological activity. All ring structures made a conformational transition to the ${}^2\text{H}_1$ conformation either before the start of this MD simulation (dp10(1, 2, 4, 5, 6 and 7) or during the actual run (dp10(3) and dp10(8)). The average Cremer-Pople puckering parameter for the ${}^2\text{H}_1$ conformation was $\theta = 52^\circ$. These positions are differentiated by above-plane and below-plane positions of the C1 and C2 ring atoms (Mikhailov *et al.*, 1997). The Δ UA (unnatural in nature as it is generated through the action of the enzyme heparanase III); contains an unsaturated uronic acid (Linhardt, 1990). The ${}^1\text{H}_2$ and ${}^2\text{H}_1$ conformations of the Δ UA are thought to vary, depending on the neighbouring structure but are thought to be in equilibrium. The ${}^1\text{H}_2$ conformation is thought to be favoured due to the anomeric effect brought on by the neighbouring GlcN residue (Mikhailov *et al.*, 1996). Modelling work carried out by Murphy (Murphy 2007) indicates the equilibrium shifted heavily towards the ${}^2\text{H}_1$ conformation. The torsion data was similar for all eight oligosaccharides. Those with a GlcNAc residue bonded to the UA (dp10(1-6); Figure 3.25, Panels A-F) showed the same profile to those with the alternative GlcNS residue (dp10(7-8); Figure 3.25, Panels G and H). The data suggests that the majority of the Φ° and Ψ° values lie in the positive area of the graph

labelled as geometry A ($\Phi = +50^\circ$ $\Psi = +10^\circ$), with a quarter of the time spent in the negative, lower left hand quadrant of the graph, labelled geometry B $\Phi = -10^\circ$ $\Psi = -10^\circ$ (Figure 3.25).

Most results from linkage 1 up to linkage 4 suggest there are no major differences between active and inactive oligosaccharides in both the conformation of the ring and the torsional geometries. This would suggest that in the case of these oligosaccharides it is not the role of the non-reducing end to interact with the FGF2 protein, although this role has been previously identified in the x-ray crystal structure of a symmetrical complex of a 2:2:2 FGF2:FGFR1:heparin ternary complex (Schlessinger *et al.*, 2000). We can be quite certain with this statement as there are no differences to speak of which would have any effect on the biological activity of the samples.

Figure 3.25 Puckering-torsion data for linkage 1

UA - GlcNAc dp10(1-6)

UA - GlcNS dp10(7-8)

Puckering parameters: all graphs show θ analysis ().

Cremer-Pople ring puckering parameters during a 1000 ps molecular dynamic simulation.




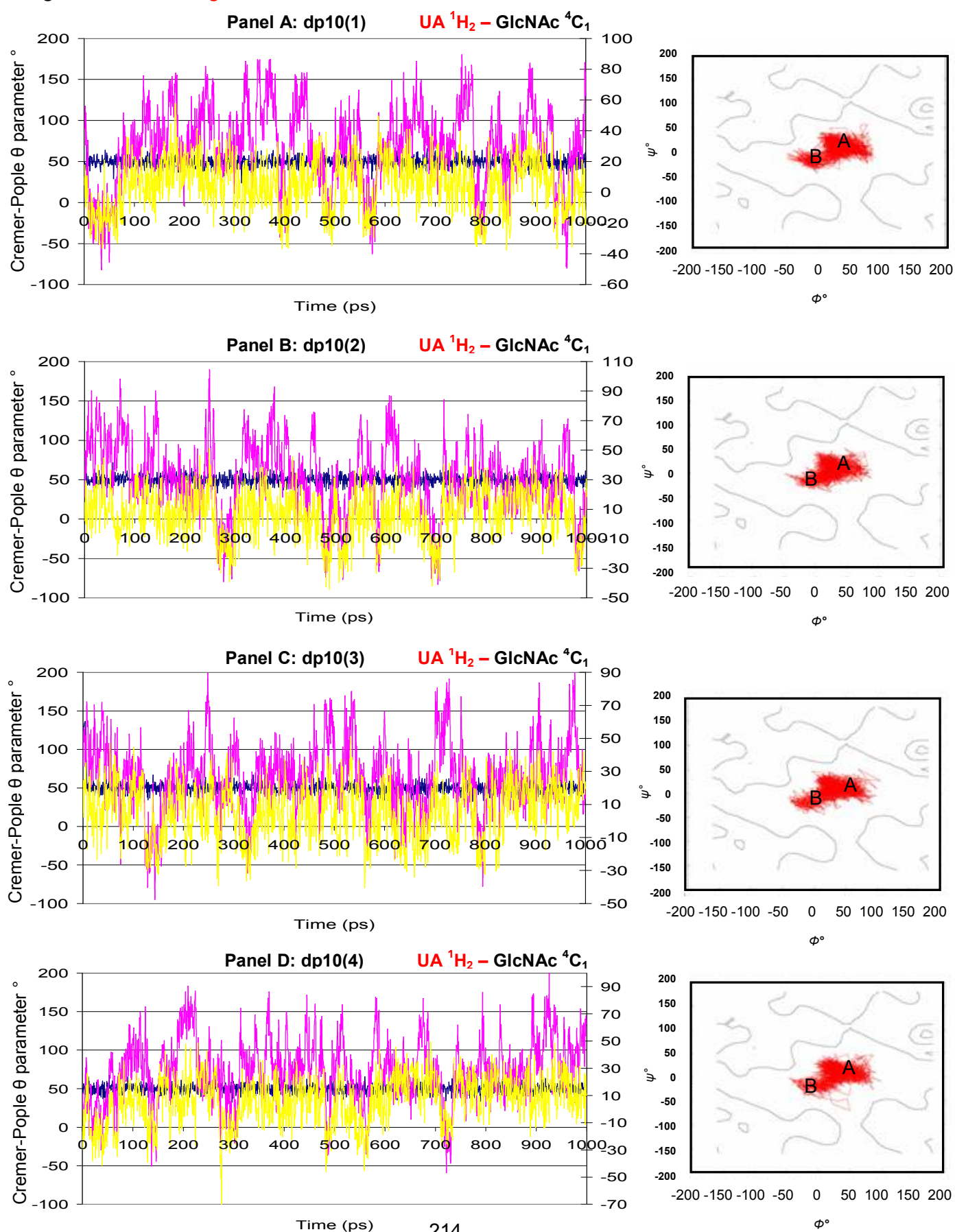
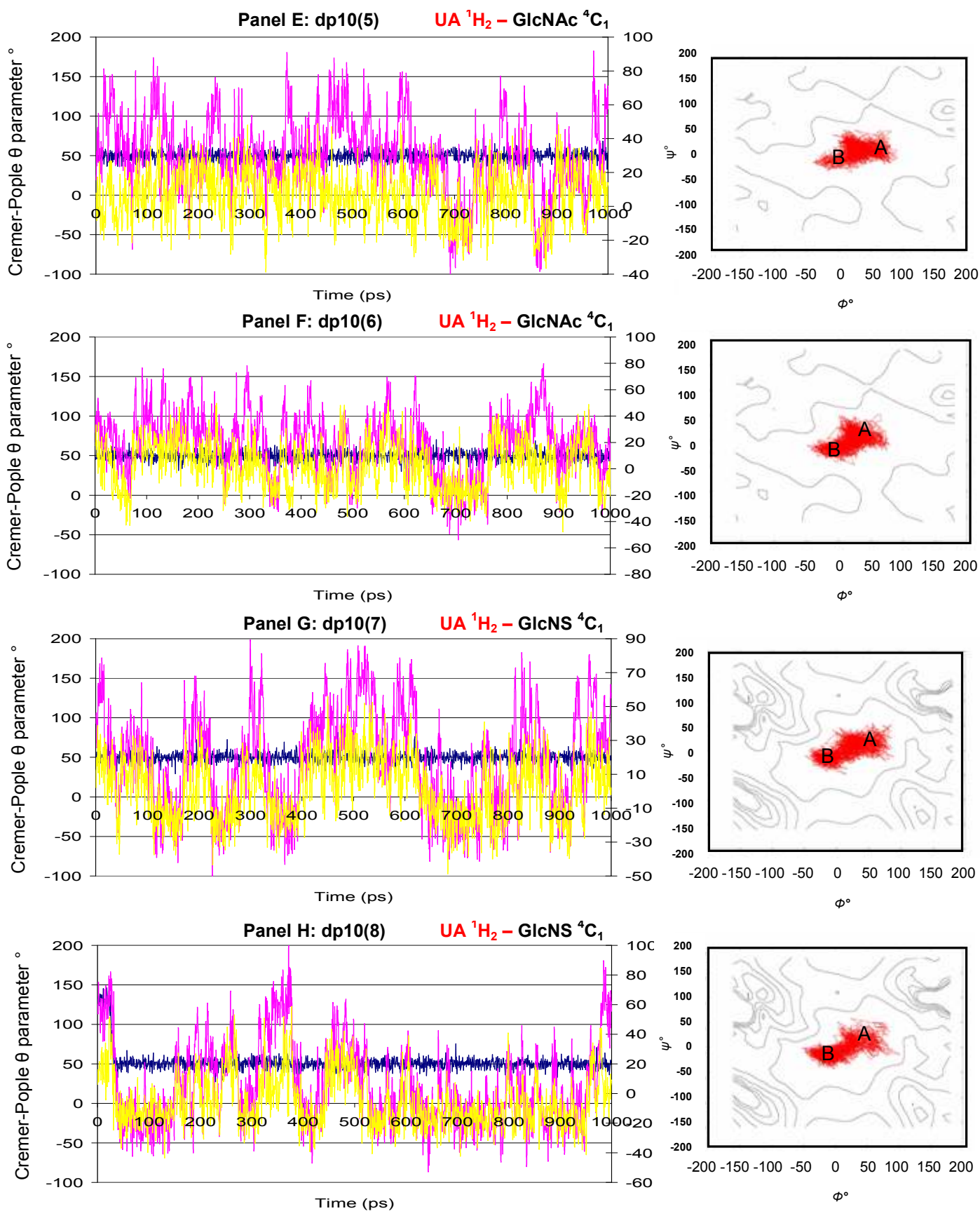
All graphs show θ analysis (ϕ_2 not shown) and changes are compared to the values identified in the torsion data in chronological order.**Torsion data:** Φ° dihedral angle (), Ψ° dihedral angle ().The variation in Φ and Ψ torsion angles over a course of 1000 ps explicit water molecular dynamic simulation (the left hand side figure).**Torsion data:** Structures were sampled every picosecond over the course of 1000 ps in TIP3P explicit water. The Φ and Ψ torsion angles across each glycosidic linkage were plotted and connected together in chronological order (). Contour lines are plotted as a guideline reference for each linkage and represent the potential energy surface of each of the glycosidic linkages (the right hand side figure).

Figure 3.25 Active Oligosaccharides



Inactive oligosaccharides



3.3 Long 10ns molecular dynamic simulations

Longer simulations were undertaken to try to increase the understanding of the differences between the active and the inactive sequences with regards to the puckering parameters and the corresponding torsion data. Not all sequences were studied, those with a GlcUA at position 9 were not taken on for further analysis (decasaccharides dp10(2, 4, 6 and 8)). Those with IdoUA present in this position were taken on for further analysis (decasaccharides dp10(1, 3, 5 and 7)). This consisted of running the same molecular dynamic simulations but for a longer period of time. Instead of just 1000ps (1ns) in time, these were ran for 10ns (10,000ps) and consisted of 5,000,000 steps in the MD run. Mdxvu was again used for the analysis of both the Cremer-Pople ring puckering parameters and the torsion data. Each MD run took approximately 5 weeks to run and the results are as follows.

3.3.1 The Reducing end Linkage – Linkage 9 for the 10 ns MD

simulation

IdoUA - GlcNAc(6S)	dp10(1, 3 and 7)
IdoUA – GlcNAc	dp10(5)

The IdoUA(2S) monosaccharide was modelled in the 1C_4 conformation prior to the 10 ns MD simulation.

There are two different types of disaccharides which make up linkage 9 across the range of species, the difference being the presence of the 6-O-sulphate group on three of the four oligosaccharides.

According to the Cremer-Pople ring puckering parameters there are no differences between active and inactive oligosaccharides for this linkage. For dp10(1), dp10(3) and dp10(7) all start in the 1C_4 starting conformation, which was the set conformation at the start of the MD simulation and all contain the sequence (IdoUA-GlcNAc(6S)). There is a conformational transition to the skew boat conformation briefly, before a further conformational transition to the 4C_1 conformation (Figure 3.26, Panels A, B and D). The inactive dp10(5), which is not comparable as it contains the sequence (IdoUA-GlcNAc) begins the simulation in the skew boat conformation, having already made the transition from the starting conformation prior to the simulation. However the torsion data indicates no differences in geometry, with both A and B geometries present in all four oligosaccharides. All four have average ϕ and ψ values of $\phi = +30$, $\psi = +10$ for geometry A and $\phi = -30$, $\psi = -10$ for geometry B.

There are no differences between the 1 ns MD simulation and the 10 ns MD simulation (compare Figure 3.26 with Figure 3.17).

Figure 3.26 Puckering-torsion data for linkage 9 for the 10 ns MD simulation

IdoUA - GlcNAc(6S) dp10(1, 3 and 7)

IdoUA – GlcNAc dp10(5)

Puckering parameters: all graphs show θ analysis ().

Cremer-Pople ring puckering parameters during a 10 ns molecular dynamic simulation.




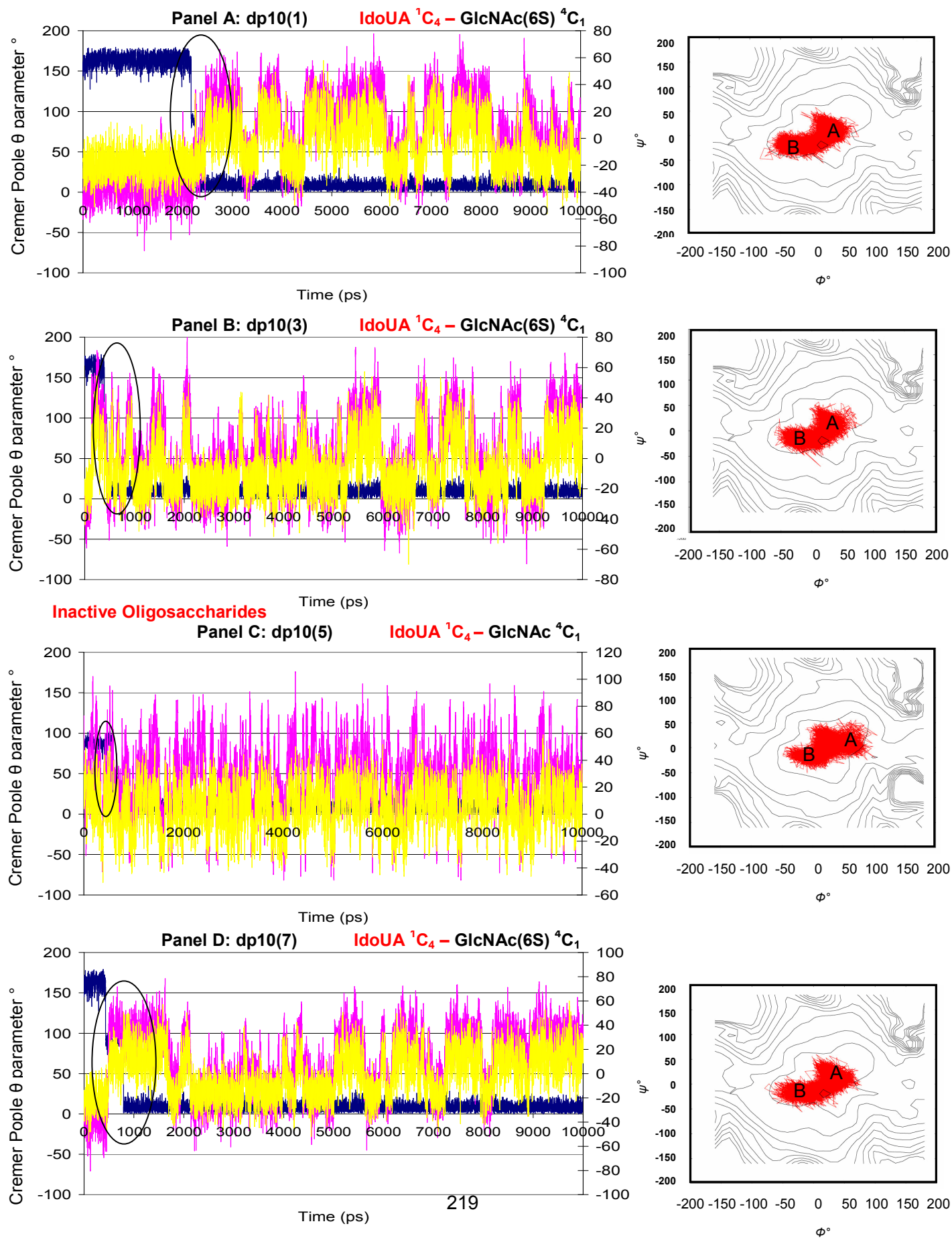
All graphs show θ analysis (ϕ_2 not shown) and changes are compared to the values identified in the torsion data in chronological order.**Torsion data:** Φ° dihedral angle (), Ψ° dihedral angle ().The variation in Φ and Ψ torsion angles over a course of 10 ns explicit water molecular dynamic simulation (the left hand side figure).**Torsion data:** Structures were sampled every picosecond over the course of 10 ns in TIP3P explicit water. The Φ and Ψ torsion angles across each glycosidic linkage were plotted and connected together in chronological order (). Contour lines are plotted as a guideline reference for each linkage and represent the potential energy surface of each of the glycosidic linkages (the right hand side figure).

Figure 3.26 Active Oligosaccharides



3.3.2 Linkage 8

GlcNS(6S) - IdoUA	dp10(1 and 5)
GlcNS - IdoUA	dp10(3 and 7)

The GlcNS \pm (6S) were modelled in the 4C_1 conformation, gg rotamer prior to the 10 ns MD simulation. Again there are two different types of disaccharides which make up this linkage. Here one active oligosaccharide is comparable to one inactive oligosaccharide in both sequences.

As expected there are no major differences between active and inactive oligosaccharides. This was suspected due to the results obtained from the 1 ns MD simulation, where again there were no real differences between active and inactive oligosaccharides. There are also no differences between those with the sequence GlcNS(6S)-IdoUA (dp10(1 and 5)); and those with the alternate sequence GlcNS-IdoUA (dp10(3 and 7)).

The Cremer-Pople puckering parameters indicate all models remain in the 4C_1 starting conformation. The corresponding torsion data indicates geometries A and B are present in all four oligosaccharides. There are no major differences between the 1 ns MD simulation and the 10 ns MD simulation, apart from the presence of geometry B in dp10(7) which was not present in the shorter MD simulation (Figure 3.27). Compare this data set with Figure 3.18.

Figure 3.27 Puckering-torsion data for linkage 8 the 10 ns MD simulation

GlcNS(6S) – IdoUA dp10(1 and 5)

GlcNS – IdoUA dp10(3 and 7)

Puckering parameters: all graphs show θ° analysis ().

Cremer-Pople ring puckering parameters during a 10 ns molecular dynamic simulation.




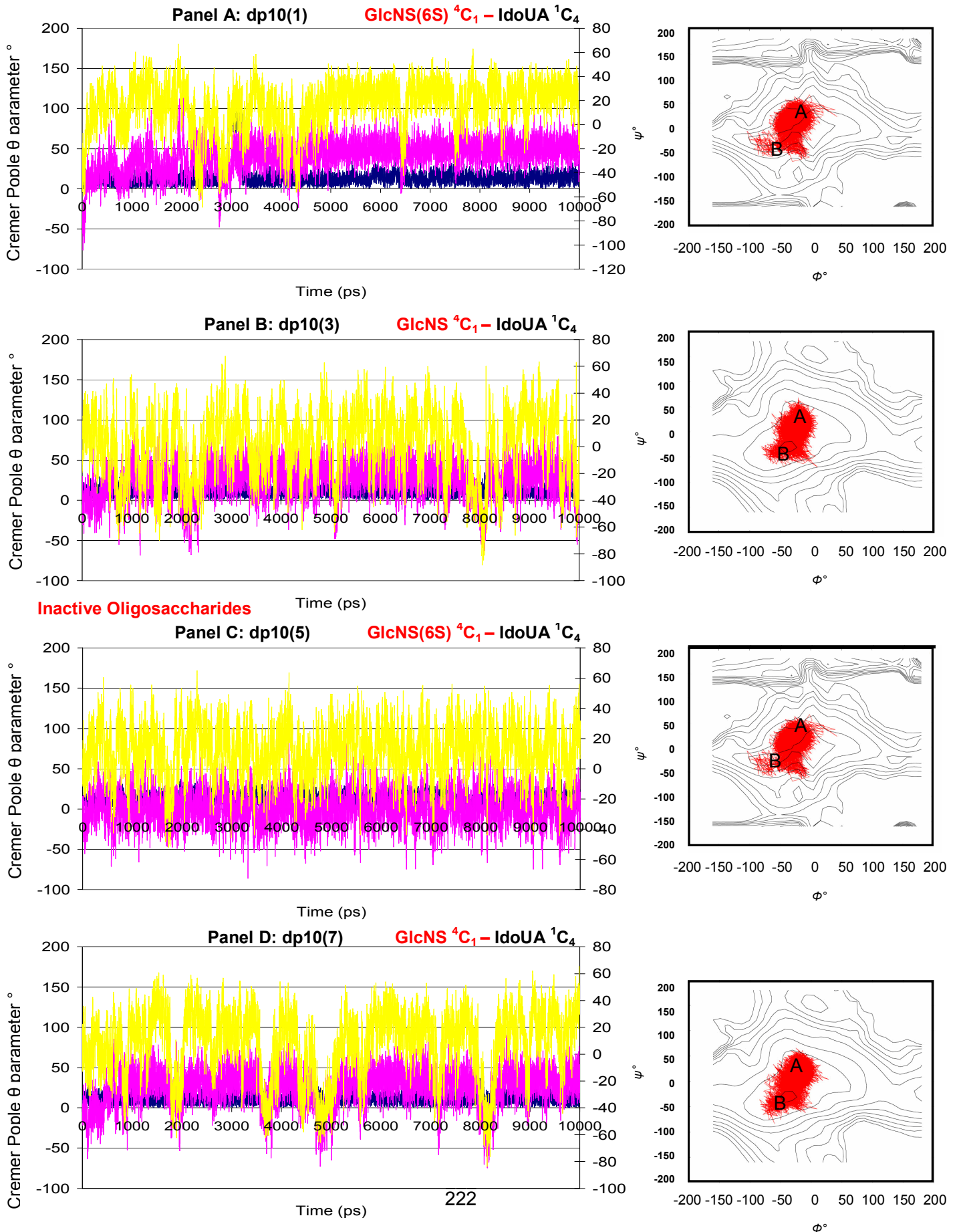
All graphs show θ analysis (ϕ_2 not shown) and changes are compared to the values identified in the torsion data in chronological order.**Torsion data:** Φ° dihedral angle (), Ψ° dihedral angle ().The variation in Φ and Ψ torsion angles over a course of 10 ns explicit water molecular dynamic simulation (the left hand side figure).**Torsion data:** Structures were sampled every picosecond over the course of 10 ns in TIP3P explicit water. The Φ and Ψ torsion angles across each glycosidic linkage were plotted and connected together in chronological order (). Contour lines are plotted as a guideline reference for each linkage and represent the potential energy surface of each of the glycosidic linkages (the right hand side figure).

Figure 3.27 Active Oligosaccharides



3.3.3 Linkage 7

IdoUA(2S) - GlcNS(6S)	dp10(1) and dp10(5)
IdoUA - GlcNS	dp10(3)
IdoUA(2S) - GlcNS	dp10(7)

The IdoUA(2S) monosaccharide was modelled in the 1C_4 conformation prior to the 10 ns MD simulation. There are three different types of disaccharides which make up this linkage.

There are some slight differences between the active and inactive oligosaccharides, the main one being the difference in geometry of the torsion angles. The oligosaccharides dp10(1) and dp10(5) are comparable as they contain the same IdoUA(2S) - GlcNS(6S) sequence. In this instance there is a slight difference in the Ψ angle in the active dp10(1) has average value in geometry A of 0° ($\Phi = +50^\circ$, $\Psi = 0^\circ$); where as the Ψ angle in the inactive dp10(5) has average value in geometry A of $+10^\circ$ ($\Phi = +50$, $\Psi = +10$). Geometry B also shows more of a difference in geometry between active and inactive oligosaccharides. The active dp10(1) has average value in geometry B of $\Phi = -40^\circ$, $\Psi = -25^\circ$; where as the inactive dp10(5) has average value in geometry A of $\Phi = -10$, $\psi = -10$).

Another difference between active and inactive oligosaccharides involves the active dp10(3) oligosaccharide. This has a slightly different sequence (IdoUA-GlcNS) and is not directly comparable to those mentioned previously it behaves in the same way as dp10(1). This contains the same Φ and Ψ data values as the other active structure. This is a significant difference and could cause an influence in the overall biological activity of the samples. The inactive dp10(7) exists in slightly different geometries then those mentioned in all the above oligosaccharides. Geometry A shows the same average data values as dp10(5), where as geometry B exists in a slightly different geometry with the Ψ angle remaining around -10° but a change in the Φ angle ($\Phi = -25$, $\Psi = -10$).

There is also a difference in ring conformation according to the Cremer-Pople ring puckering parameter data, but this is between all oligosaccharides. Dp10(1) undergoes three changes in

conformation from the starting 1C_4 conformation to the 2S_0 conformation after 1 ns. This then undergoes a second conformational transition back to the chair conformation and then a final transition to the skew boat conformation (Figure 3.28, Panel A). Cremer-Pople puckering parameters indicates two conformational changes in dp10(3). Firstly the starting conformation undergoes a conformational change approximately 500 ps into the timescale to the skew boat conformation briefly. It then undergoes a change to the 4C_1 conformation where it remains for the remainder of the simulation (Figure 3.28, Panel B). For the inactive models Cremer-Pople puckering parameters indicate a change from the starting conformation the skew boat conformation after 1 ns for dp10(5). This is the final conformation observed in the 10 ns MD simulation (Figure 3.28, Panel C). Decasaccharide dp10(7) behaves in a similar way to the active dp10(3) oligosaccharide. This makes it difficult to compare a shift in conformation to biological activity between the active and inactive oligosaccharides (compare Figure 3.28 with Figure 3.19).



Figure 3.28 Puckering-torsion data for linkage 7 the 10 ns MD simulation

IdoUA(2S) - GlcNS(6S)	dp10(1 and dp10(5)
IdoUA - GlcNS	dp10(3)
IdoUA(2S) - GlcNS	dp10(7)

Puckering parameters: all graphs show θ analysis ().

Cremer-Pople ring puckering parameters during a 10 ns molecular dynamic simulation.

All graphs show θ analysis (ϕ_2 not shown) and changes are compared to the values identified in the torsion data in chronological order.

Torsion data: Φ° dihedral angle (), Ψ° dihedral angle ().

The variation in Φ and Ψ torsion angles over a course of 10 ns explicit water molecular dynamic simulation (the left hand side figure).


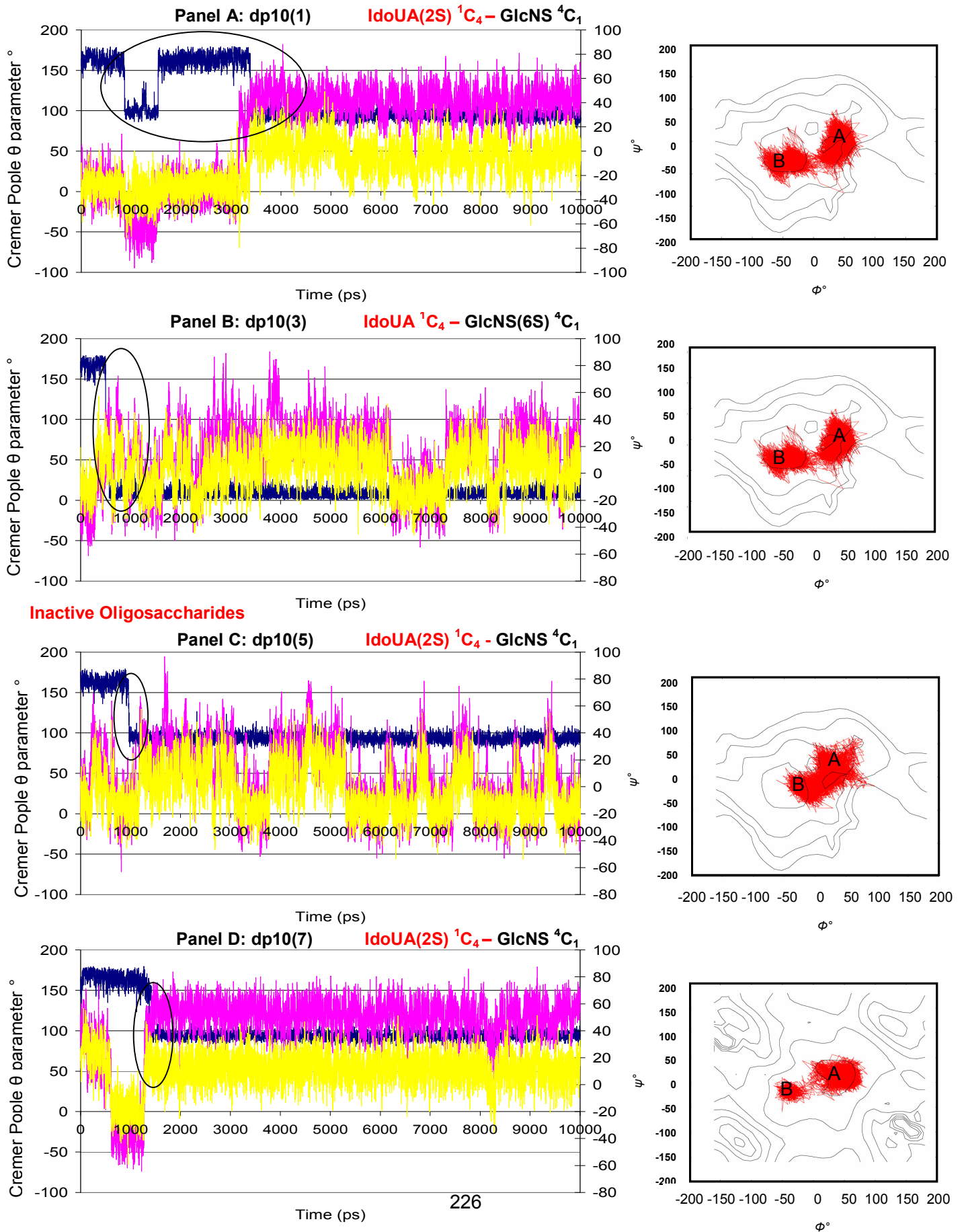
Torsion data: Structures were sampled every picosecond over the course of 10 ns in TIP3P explicit water. The Φ and Ψ torsion angles across each glycosidic linkage were plotted and connected together in chronological order (). Contour lines are plotted as a guideline reference for each linkage and represent the potential energy surface of each of the glycosidic linkages (the right hand side figure).

Figure 3.28 Active Oligosaccharides



3.3.4 Linkage 6

GlcNS(6S) - IdoUA(2S) dp10(1), dp10(5) and dp10(7)

GlcNS(6S) - IdoUA dp10(3)

The GlcNS \pm (6S) were modelled in the 4C_1 conformation, gg rotamer prior to the 10 ns MD simulation.

There are two different types of disaccharides present of which oligosaccharides dp10(1, 5 and 7) are all comparable to each other as they contain the sequence GlcNS(6S)-IdoUA(2S). Interestingly there is the presence of an unknown conformation in oligosaccharides dp10(1) and dp10(7). This conformation is observed for approximately 1 ns in dp10(1) and 2 ns in do10(7) (Figure 3.29, Panels A and D). They have average data values of 81° and 87° respectfully. This has not previously identified in the 10 ns simulation for linkage 6, or for any of the other linkages. This could be an abnormality in the results but as they take over a month in computing time a repeated MD run was not carried out. Another explanation for this unidentified conformation is that it could be a half chair conformation not previously identified. All the other oligosaccharides were identified in the known 4C_1 conformation.

There is also a slight difference in torsion data observed for dp10(3). This had been identified in two different geometries; however they were different to those observed for the other oligosaccharides. Here geometry B had average data values of $\Phi = -50$, $\Psi = -50$, which was the observed geometry in all the oligosaccharides. The additional geometry was identified as geometry C and this was defined with average data values of $\Phi = -25$, $\Psi = 0$ (Figure 3.29, Panel B). All other models were observed in geometry A which had average data values of $\Phi = -10$, $\Psi = -10$ (Figure 3.29, Panels A, C and D).

The presence of the unknown conformation was unexpected as this has not been identified previously in literature and so does indicate a abnormal result. Also there are many differences between oligosaccharides in this linkage but it is difficult to establish a difference between active and inactive structure (compare Figure 3.29 with Figure 3.20).

Figure 3.29 Puckering-torsion data for linkage 6 the 10 ns MD simulation

GlcNS(6S) - IdoUA(2S) dp10(1), dp10(5) and dp10(7)

GlcNS(6S) - IdoUA dp10(3)

Puckering parameters: all graphs show θ analysis ().

Cremer-Pople ring puckering parameters during a 10 ns molecular dynamic simulation.




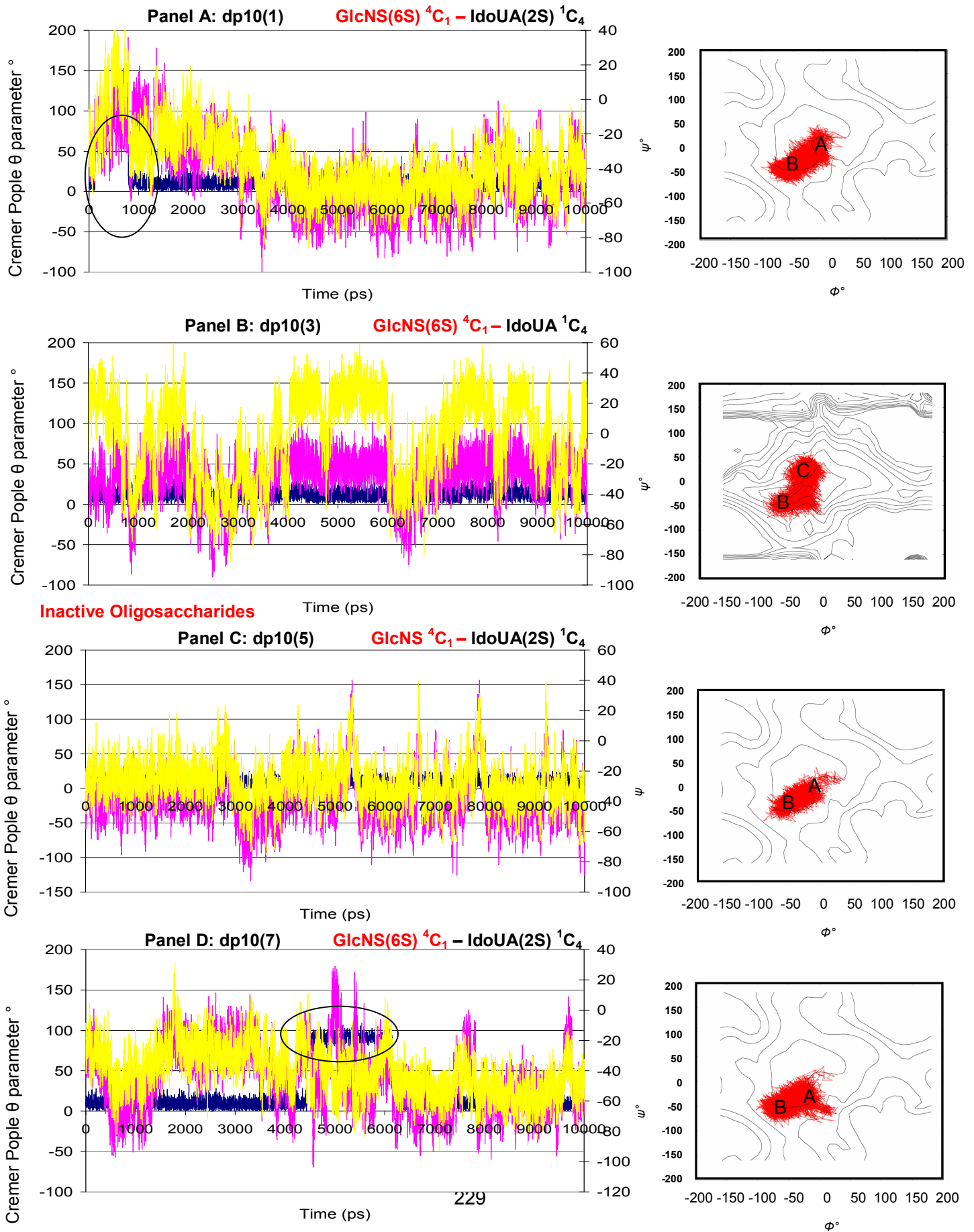
All graphs show θ analysis (ϕ_2 not shown) and changes are compared to the values identified in the torsion data in chronological order.**Torsion data:** Φ° dihedral angle (), Ψ° dihedral angle ().The variation in Φ and torsion angles over a course of 10 ns explicit water molecular dynamic simulation (the left hand side figure).**Torsion data:** Structures were sampled every picosecond over the course of 10 ns in TIP3P explicit water. The Φ and Ψ torsion angles across each glycosidic linkage were plotted and connected together in chronological order (). Contour lines are plotted as a guideline reference for each linkage and represent the potential energy surface of each of the glycosidic linkages (the right hand side figure).

Figure 3.29 Active Oligosaccharides



3.3.5 Linkage 5

IdoUA(2S) – GlcNS(6S)

dp10(1), dp10(3), dp10(5) and dp10(7)

The IdoUA(2S) monosaccharide was modelled in the 1C_4 conformation prior to the 10 ns MD simulation. All oligosaccharides contain the same IdoUA(2S) – GlcNS(6S) structure so all models are comparable for this linkage.

There are major differences between active and inactive structures regarding torsion angle geometries. All exist in two different geometries, with differing torsion angles. The active dp10(1) exists equally in geometry A as it does in geometry B. These geometries have average data values of $\Phi = +50$, $\Psi = -50$ for geometry A and $\Phi = -40$, $\Psi = -60$ for geometry B. Oligosaccharide dp10(3) shows similar results for geometry A with a slight change in the Φ value of -60° ($\Phi = -60$, $\Psi = -60$). When compared to the inactive sequences, although they both contain both geometries (geometry A again shows the same data values as those identified in the active structures); geometry B is very different. Both dp10(5) and dp10(7) only briefly spend time in geometry B and here the Ψ angle remains at a relatively constant -60° , with a major difference in the Φ value ($\Phi = -25$, $\Psi = -60$). This difference is significant and could influence overall biological activity.

Moving on to compare the Cremer-Pople puckering parameters there is a slight difference between active and inactive structures. For dp10(3) the puckering parameter data shows a change in conformation from the skew boat conformation to the 1C_4 conformation and then back again (Figure 3.30, Panel B). This is the same final conformation as that observed in both dp10(1) and dp10(5) (Figure 3.30, Panels A and C). However these only make the one conformational transition during the simulation, from the starting 1C_4 conformation to the final 2S_0 conformation. The fact that the inactive dp10(5) oligosaccharide shows similar changes in conformation suggests the influence for biological activity is also due to the available torsional geometries.

Dp10(7) shows very different results in that, like dp10(3) it undergoes two conformational transitions these occur opposite to those previously described. Here the 1C_4 conformation is

present at the start of the simulation. It undergoes a conformational transformation to the skew boat form after 1 ns remaining at the conformation for approximately 5 ns before returning to the 1C_4 conformation (Figure 3.30, Panel D).

The data for this linkage suggests it is the combination of the torsional geometries alongside the iduronate conformation which is significant for biological activity. Compare this data with Figure 3.21.

Figure 3.30 Puckering-torsion data for linkage 5 the 10 ns MD simulation

IdoUA(2S) – GlcNS(6S)

dp10(1), dp10(3), dp10(5) and dp10(7)

Puckering parameters: all graphs show θ analysis ().

Cremer-Pople ring puckering parameters during a 10 ns molecular dynamic simulation.




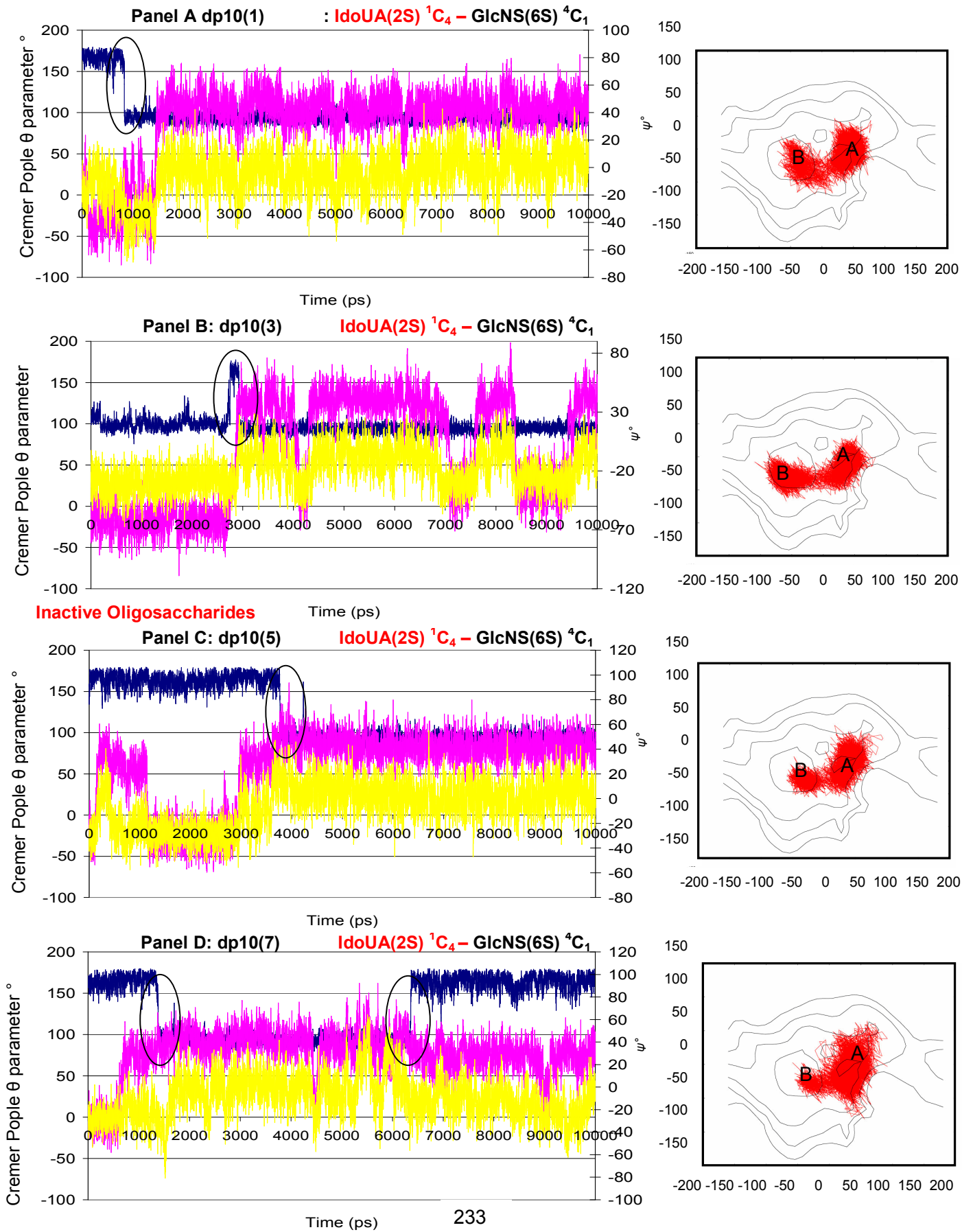
All graphs show θ analysis (ϕ_2 not shown) and changes are compared to the values identified in the torsion data in chronological order.**Torsion data:** Φ° dihedral angle (), Ψ° dihedral angle ().The variation in Φ and Ψ torsion angles over a course of 10 ns explicit water molecular dynamic simulation (the left hand side figure).**Torsion data:** Structures were sampled every picosecond over the course of 10 ns in TIP3P explicit water. The Φ and Ψ torsion angles across each glycosidic linkage were plotted and connected together in chronological order (). Contour lines are plotted as a guideline reference for each linkage and represent the potential energy surface of each of the glycosidic linkages (the right hand side figure).

Figure 3.30 Active Oligosaccharides



3.3.6 Linkage 4

GlcNS(6S) - IdoUA(2S)	dp10(1)
GlcNS - IdoUA(2S)	dp10(3), dp10(5) and dp10(7)

The GlcNS \pm (6S) were modelled in the 4C_1 conformation, gg rotamer prior to the 10 ns MD simulation. The longer simulation for linkage 4 shows no differences in ring conformation between the active and inactive oligosaccharides.

There are slight differences between results from the 1 ns MD simulation and the 10 ns MD simulation. The torsional geometry in the 1 ns MD simulation shows the presence of only geometry B for oligosaccharides dp10(1) and dp10(7). Both geometries A and B existed in dp10(3) and dp10(5) for the 1 ns MD simulation. In the 10 ns MD simulation all four oligosaccharides exist in both A and B geometries (Figure 3.31).

We can assume from the 10 ns MD simulation, which are similar to the 1 ns MD simulation results that there is no role in FGF2 protein interactions for the non-reducing terminal. This has been reached based on observations from both the 1 ns and 10 ns MD simulations and so the final three linkages for the 10 ns simulations will not be discussed here. This is due to there being no differences between the longer simulation and the shorter 1 ns MD simulation (results not shown).

To identify any real differences in active and inactive oligosaccharides longer simulations would need to be performed. Length of computing time required to perform each simulation was a major factor in this study, but to be competing with molecular simulations in industry longer simulations really are needed. A 50 ns MD simulation would be the next step in this kind of study to really determine any major changes and to really compete with other studies. Compare Figure 3.31 to the 1 ns MD simulation data in Figure 3.22.



Figure 3.31 Puckering-torsion data for linkage 4 the 10 ns MD simulation

GlcNS(6S) - IdoUA(2S)	dp10(1)
GlcNS - IdoUA(2S)	dp10(3), dp10(5) and dp10(7)

Puckering parameters: all graphs show θ° analysis ().

Cremer-Pople ring puckering parameters during a 10 ns molecular dynamic simulation.

All graphs show θ analysis (ϕ_2 not shown) and changes are compared to the values identified in the torsion data in chronological order.

Torsion data: Φ° dihedral angle (), Ψ° dihedral angle ().

The variation in Φ and Ψ torsion angles over a course of 10 ns explicit water molecular dynamic simulation (the left hand side figure).


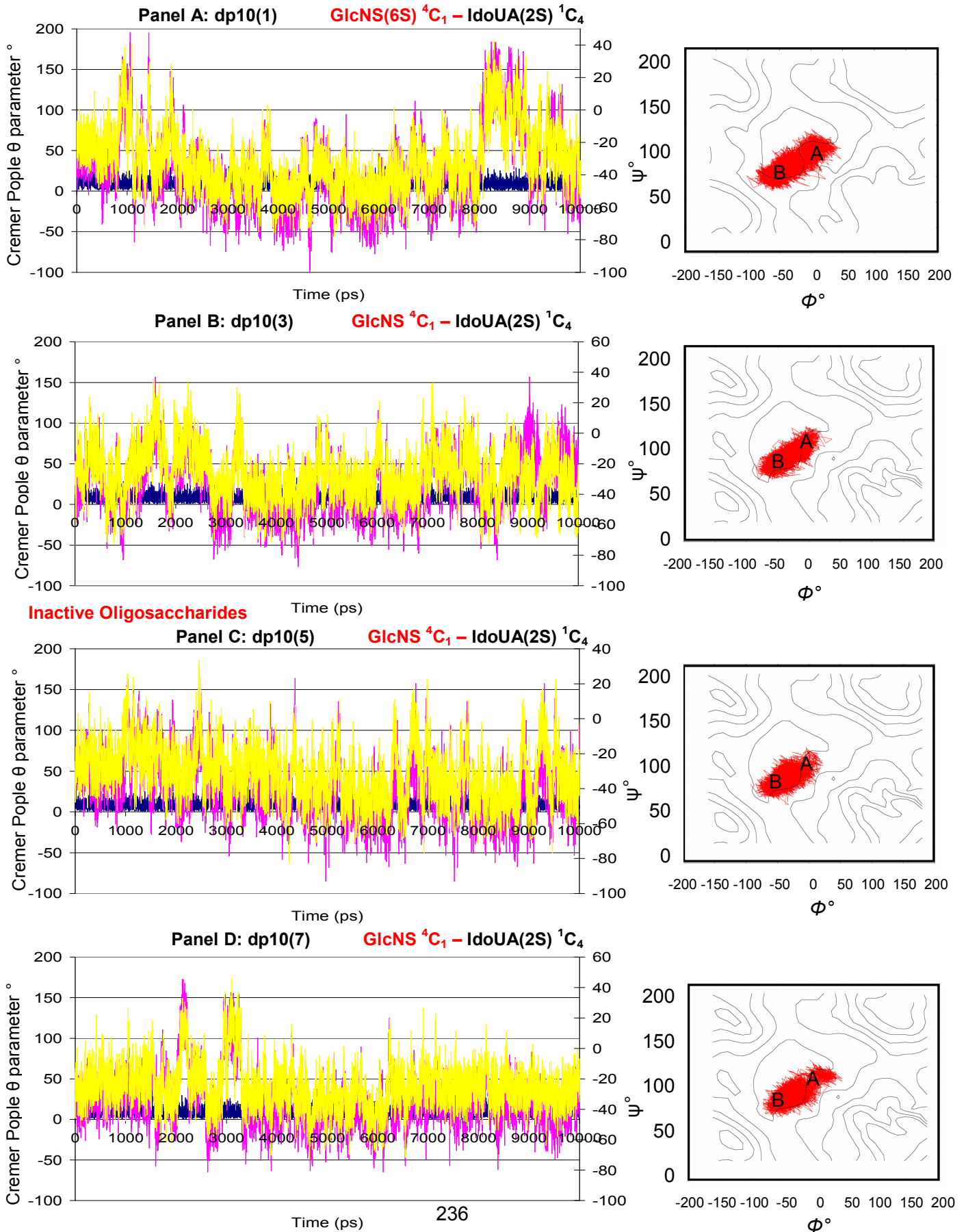
Torsion data: Structures were sampled every picosecond over the course of 10 ns in TIP3P explicit water. The Φ and Ψ torsion angles across each glycosidic linkage were plotted and connected together in chronological order (). Contour lines are plotted as a guideline reference for each linkage and represent the potential energy surface of each of the glycosidic linkages (the right hand side figure).

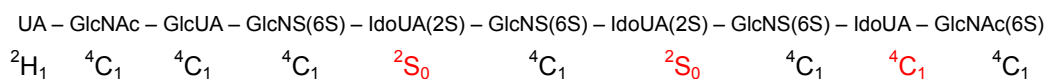
Figure 3.31 Active Oligosaccharides



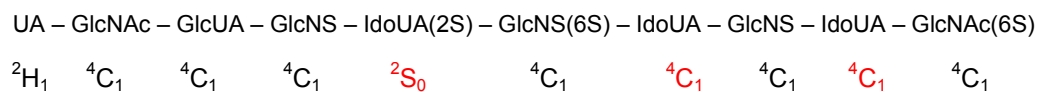
3.3.7 Oligosaccharide sequence after 10 ns MD simulation

The ring conformations for the iduronates after the 10 ns MD simulation are as follows:

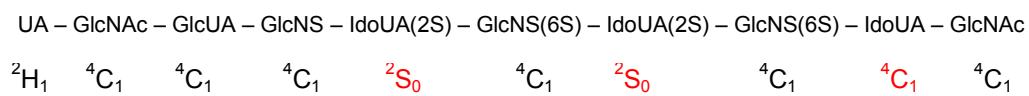
dp10(1) **Active**



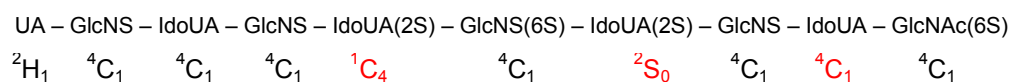
dp10(3) **Active**



dp10(5) **Inactive**



dp10(7) **Inactive**



It is only dp10(7) which contains all three iduronate conformations present after the 10 ns MD simulation. Surprisingly, dp10(1) and dp10(5) show the same ring conformation profiles even though dp10(1) is the active oligosaccharide and dp10(5) is the inactive oligosaccharide.

3.4 Docking results of HS Decasaccharides

3.4.1 Preliminary results of 1BFC and 1FQ9

Docking of molecules and compounds into the binding site of a receptor is an important part of structure based drug design. A thorough understanding of the principles that determine strength of a protein/ligand complex, alongside accurate and fast docking protocol is essential in computer aided drug design. Alongside this is the ability to actually visualise the binding geometries created as visual screening of libraries is fast becoming standard practice in modern drug discovery.

3.4.1.1 1BFC

The initial docking analysis was carried out to assess the viability of using Autodock Vina (here after referred to as Vina) as reliable docking software, alongside UCSF Chimera alpha, version 1.4 (build 29530). Also checked; as a comparison to this were Autodock 4.2 and also Dock 6.2. The first of the two test models was that from the file 1BFC from the Protein Data Bank. This is a bFGF (FGF2) complexed with a hexamer heparin fragment (Faham *et al.*, 1996). The original x-ray structure can be seen on the left of (Figure 3.32, Panel A), with the predicted Vina results on the right hand side of the figure (Figure 3.32, Panel B with the predicted structure shown in green).

Chimera was firstly used to select and create two separate pdb files, the first being the FGF protein file, which was saved as a protein.pdb file; then the heparin oligosaccharide also saved as a .pdb file (Trott and Olsen, 2010). These files were then used to assess all three docking programmes (Vina, Autodock and Dock 6.2). After initial tests it was decided Vina was by far the easiest and the most accurate of the docking programmes and so all dockings from here onwards were carried out using Vina.

Visualization of the predicted geometries was carried out in PyMol molecular graphics system (DeLano 2002).

Figure 3.32

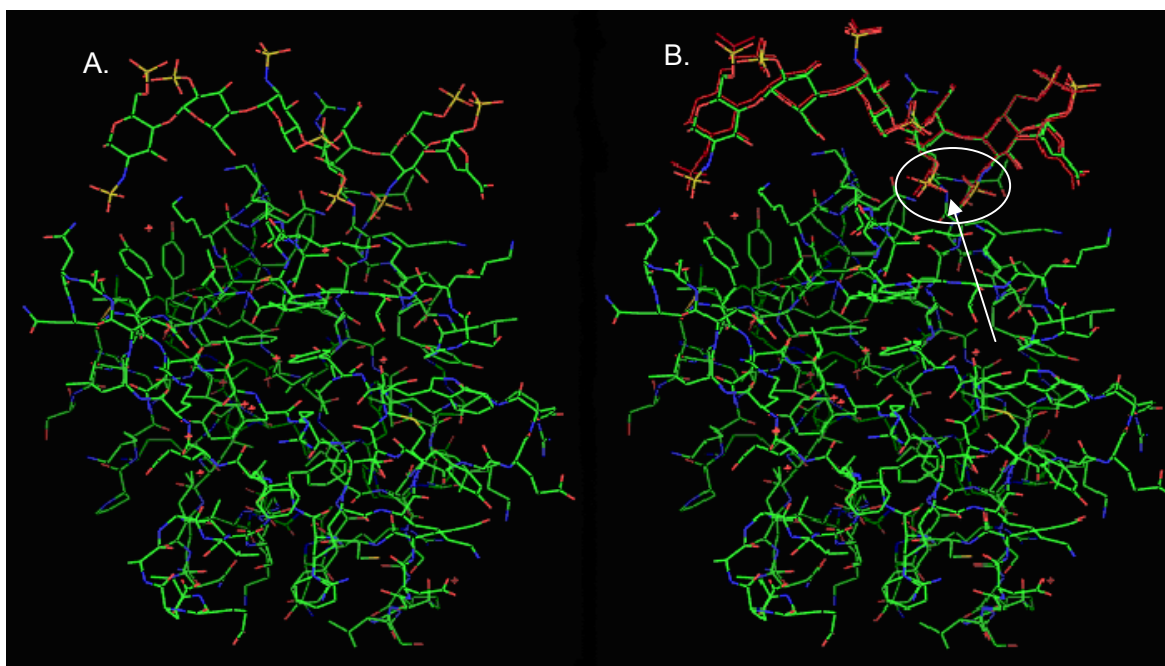


Figure 3.32 A comparison of the X-ray crystal structure of a bFGF (FGF2) complexed with a hexamer heparin fragment alongside the Vina docked model.

Panel A. The original x-ray crystal structure from file 1BFC obtained from the protein data bank and originally identified by work carried out by Faham *et al* (Faham *et al.*, 1996).

Panel B. Shows the original x-ray crystal structure alongside the Vina predicted result shown by the red hexasaccharide. As this shows, Vina was successful in identifying the correct binding position for the heparin hexasaccharide.

The binding site; shown by the arrow on the right hand side of the figure (Figure 3.32, Panel B) involves the two sulphate groups from the heparin fragment interacting with the protein binding site. When referring to the binding site in the Vina results, it is with reference to these two highlighted sulphate groups. An extended version of this binding site can be seen in the following figure (Figure 3.25) which has the binding site highlighted. This binding site highlights those important amino acid residues identified earlier in the report (Figure 1.21, p91). This includes residues 27, 28, 101, 102, 103 and 120-137 (Faham *et al.*, 1996; Schlessinger *et al.*, 2000).

The hydrogen bonding function used in Chimera was used to identify the interactions between the heparin hexasaccharide and the FGF and is clearly shown by the connecting blue lines (Figure 3.33).

Figure 3.33

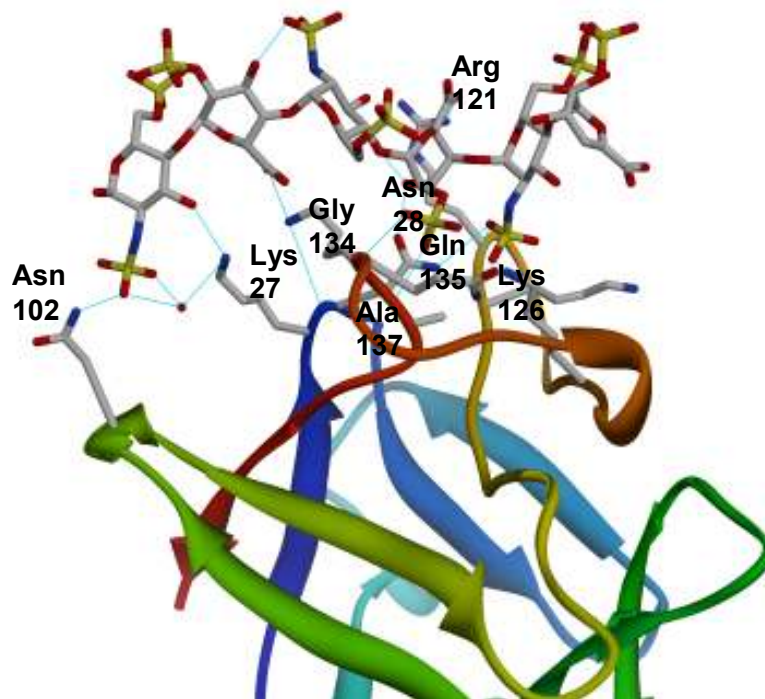


Figure 3.33 The important binding residues identified between FGF2 and a heparin hexasaccharide.

This figure was obtained from the protein data bank (file 1BFC) and has been adapted from the original structure by Faham *et al* (Faham *et al.* 1996). The important binding residues are clearly identified which were previously discussed (Introduction, p91).

3.4.1.2 1FQ9

The file 1FQ9, taken from the Protein Data Bank was also used to check the accuracy of Vina as a relevant molecular docking software programme. The structure is taken from a 2:2:2 complex of ternary bFGF-FGFR1-heparin complex, however for the purposes of the docking only one FGF: heparin chain was used. (Both the FGFR's are removed as are one complex of FGF: heparin). As above the original pdb file was separated and then docked back together again, going through the same process as described above (p124). The image on the left of the diagram (Figure 3.34) shows the original x-ray crystal structure. The image on the right shows the predicted Vina result with the predicted result shown in green. As before; the predicted Vina result is the same as seen in the x-ray crystal structure, thus providing evidence that Vina can be used as a credible docking programme for predicting the binding of other HS/heparin fragments.

An extended version of the binding site can be seen in Figure 3.35 (Schlessinger *et al.* 2000). The find H bond function used in 1BFC was also used here to identify which parts of the chain hydrogen bond to the FGF protein. These bind to similar parts of the protein previously identified in 1BFC.

Figure 3.34

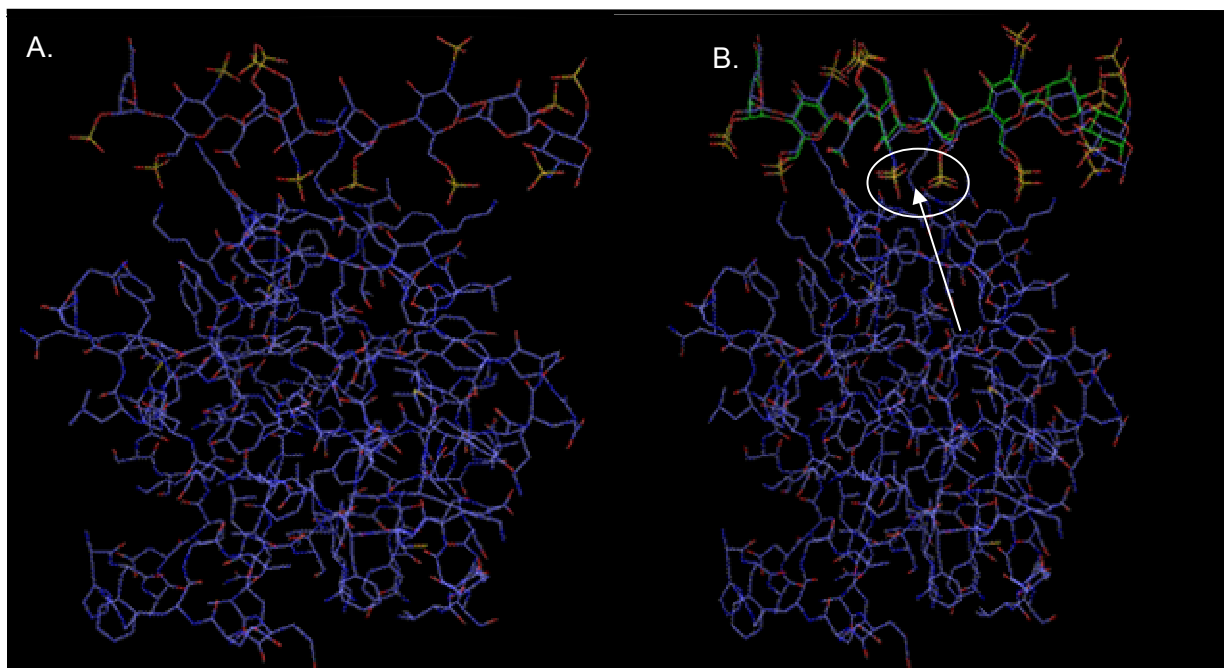


Figure 3.34 A comparison of the x-ray crystal structure of a 2:2:2 complex of ternary bFGF-FGFR1-heparin complex alongside the Vina predicted model.

Panel A. Shows the original x-ray crystal structure from file 1FQ9 obtained from the Protein Data Bank (Schlessinger *et al.* 2000).

Panel B. The Vina predicted docking results compared to the x-ray crystal structure. The binding site was identified by x-ray crystallography; with the binding site identified by the white arrow.

Figure 3.35

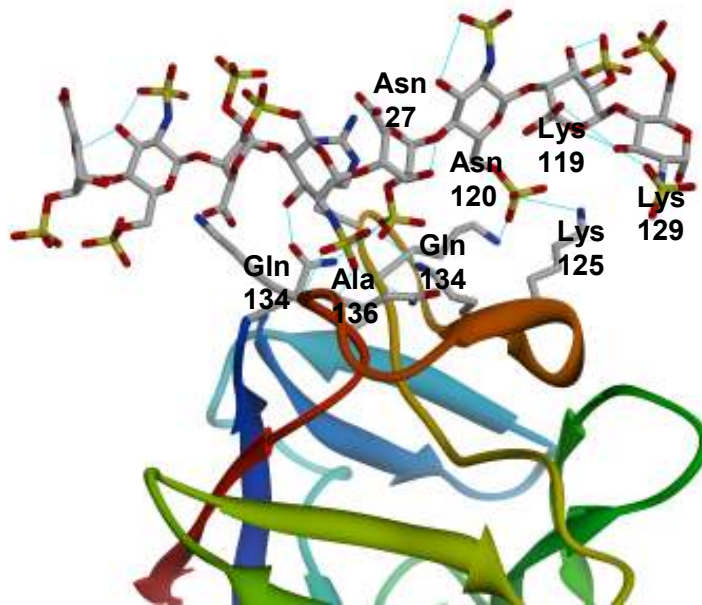


Figure 3.35 The important binding site residues identified in a 2:2:2 complex of ternary bFGF-FGFR1-heparin complex.

This figure has been adapted from Schlessinger *et al* (Schlessinger *et al.*, 2000) obtained from file 1FQ9 from the Protein Data Bank.

Following on from the preliminary docking, further docking work was carried out on the Goodger dp10 oligosaccharides. The purpose of this was to analyse the ability of the Vina docking program to dock the Goodger oligosaccharides to the same binding site as identified in both X-ray structures used in the preliminary dockings (Figures 3.32 and 3.34).

3.4.2 Docking of active and inactive dp10 structures

Those oligosaccharides which were used for docking in this section of the study were dp10(1), dp10(3), dp10(5) and dp10(7). These were docked into the 1BFC x-ray crystal structure only. Time constraints meant only one of the three or four potential x-ray crystal structures could be analysed. Different frames from the initial molecular dynamic simulations were identified as it was impossible to run all frames involved due to the computational time needed for the docking simulations (the longer 10ns simulations each contain 10,000 frames, with each frame taking exceeding more than 48 hours computational docking time). Selected frames were initially docked as rigid structures (all bonds were left as non-rotatable), as shown in the preliminary dockings above and a global search of just the known binding site was performed. Luckily the binding site is known due to the previous work carried out with the resultant x-ray crystal structures (Faham *et al.* 1996; Schlessinger *et al.* 2000). If the same docking calculations were to be carried out on an unknown binding site, then the entire protein would have to be searched, leading to greatly extended computational docking times.

The binding site for the two known x-ray crystal structures both involved the following sequence (Table 3.2).

Table 3.2

IdoUA(2S)	GlcNS(6S)
Residue 4	Residue 5
Binding position X	Binding position Y

Table 3.2 The known binding site identified in the known X-ray crystal structures.

This binding information was identified in a number of studies (Faham, Hileman *et al.* 1996; Schlessinger *et al.* 2000).

All dockings carried out will be in reference to these particular ring structures; with particular attention on the positioning of the sulphate groups highlighted in the sequence.

Models dp10(1), dp10(3), dp10(5) and dp10(7), were chosen because of the presence of the IdoUA at position 9 in the sequence. Those sequences with GlcUA in the same position were not taken on for docking analysis (models dp10(2), dp10(4), dp10(6) and dp10(8)). Table 3.3 contains information on the frames identified for the docking analysis for each model. Each frame corresponds to a change in either the dihedral angle or a conformational change in the iduronate ring structure.

If time constraints had not been an issue then all eight decaaccharide oligosaccharides would have been used in the docking analysis, as according to the puckering-torsion data analysis the presence of the GlcUA may play more of a role in the binding affinity than initially thought.

Table 3.3

Decasaccharide 1ns timescale	Frame number	Decasaccharide 10ns timescale	Frame number
1	45	1	200
	350		1300
	600		3000
	900		8300
3	25	3	250
	200		440
	650		560
	800		2900
5	30	5	6000
	850		500
			1050
			4000
7	50	7	6000
	150		250
	200		750
	900		5000

Table 3.3 Frame numbers chosen for docking analysis with Vina.

These were the frames which were chosen on the basis of either a conformational change and/or torsional angle change, identified in the puckering/torsion data.

3.4.3 The 1 ns Docking simulations

3.4.3.1 Oligosaccharide dp10(1):

Model dp10(1) was docked to the FGF2 protein in the file 1BFC using the identified four frames. The time period chosen for analysis in Vina (with the corresponding frame numbers) were 28ps, 218ps, 600ps and 900ps (frames 45, 315, 975 and 1463) out of a possible 1600 frames (Table 3.4). Each frame counts for 0.625 ps in time from the 1 ns timescale (1000 ps). All frames were chosen mainly based on the conformation of the iduronic acids in the sequence; with these being the most conformationally flexible residues in the chain (see Introduction, p64). Other factors such as the dihedral angle data was also taken into account, which has already been shown to change over the 1 ns timescale (see Results, puckering/torsion data).

The main area of interest for all docking calculations was the identified binding site observed in the x-ray crystal structure of the bFGF complex (file 1BFC). In particular the position of the two sulphate groups identified in the bound heparin fragment were studied in detail (Figure 3.32). For all models these two sulphate groups were used as a point of reference for the binding site since these were the two key positions identified as important for binding and activation of a HS:FGF complex (see Introduction p82 onwards).

Table 3.4

Residue	Frame: 45 Time in ps: 28	Frame: 350 Time in ps: 218	Frame: 975 Time in ps: 600	Frame: 1463 Time in ps: 900
UA	${}^2\text{H}_1$	${}^2\text{H}_1$	${}^2\text{H}_1$	${}^2\text{H}_1$
GlcNAc	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$
GlcUA	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$
GlcNS(6S)	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$
IdoUA(2S)	${}^1\text{C}_4$	${}^1\text{C}_4$	${}^1\text{C}_4$	${}^1\text{C}_4$
GlcNS(6S)	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$
IdoUA(2S)	${}^1\text{C}_4$	${}^1\text{C}_4$	${}^2\text{S}_0$	${}^2\text{S}_0$
GlcNS(6S)	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$
IdoUA	${}^1\text{C}_4$	${}^2\text{S}_0$	${}^4\text{C}_1$	${}^4\text{C}_1$
GlcNAc (6S)	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$

Table 3.4 Conformations of all the ring structures for dp10(1) in the frames chosen for docking using Vina.

The iduronates highlighted in red were of particular interest due to their ability to change conformation during the MD simulation.

Vina successfully predicted 11 high affinity low energy structures as potential docking conformations during the first rigid docking performed. Frame 45, the first structure docked into the 1BFC binding site comprises all iduronates in the 1C_4 conformation. Of the 11 structures, model 1 had the highest binding affinity of -13.0 kcal/mol and rmsd values of 0.000 (for both l.b and u.b values). These values suggest it is the ideal candidate for binding to the protein with both rmsd values at zero and the lowest possible affinity value.

Here the HS chain has the N-sulphated group of the GlcNS(6S) of residue 6 in binding position X with the neighbouring residue, the 2-O-sulphate group of the iduronate in binding position Y (Figure 3.36, Panel A and B). This positions the chain in the opposite direction then that identified in the x-ray crystal structure and although is not an exact fit the sulphate groups are orientated in the exact position identified in the crystal structure.

There were other structures predicted but these did not fit to both binding sites present in the crystal structure. Other structures which had been ranked based on the binding affinity could be dismissed from the search due to the position of the sulphate groups in the HS chain. One such model had been ranked as model number 10 out of 11 and had a binding affinity of -10.1 kcal/mol (Table 3.6). However there was no match to the binding site identified in the crystal structure. The full range of docking results can be found in the appendix section of this study.

Figure 3.36

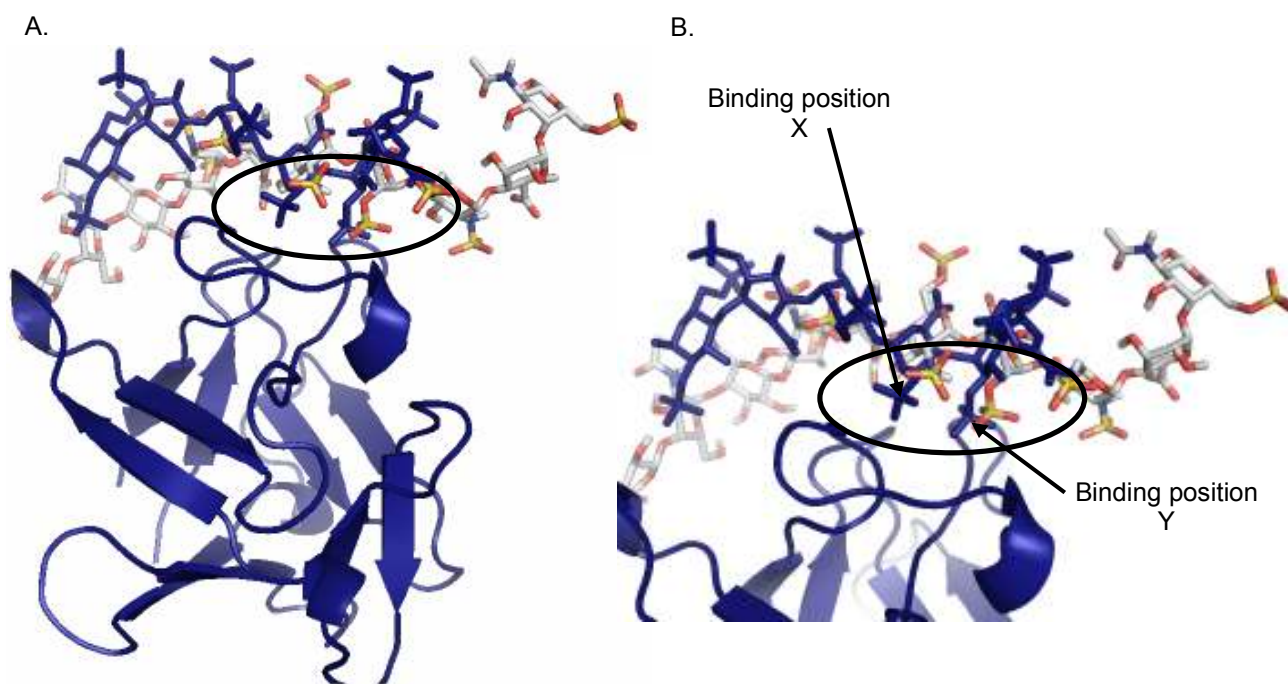


Figure 3.36 The highest affinity binding structure as predicted by Vina for dp10(1) frame 45, from the rigid 1 ns MD simulation

This is the structure with the highest affinity (kcal/mol) and is rated as model number 1 in the Vina search. It has a binding affinity of -13.0 kcal/mol and rmsd values of 0.000 (for both l.b and u.b values). The protein in blue shows the FGF2 protein in the x-ray crystal structure, alongside the heparin fragment from the x-ray crystal structure. The lighter coloured decasaccharide is the Goodger model.

Panel A. Frame 45, model 1 with both binding positions predicted.

Panel B. The magnified image of frame 45, with both binding sites X and Y highlighted.

The presence of the first 2S_0 conformation identified (residue 9) has no influence on the docking, possibly due to its position next to the reducing terminal. 14 low energy structures were created with the best fit model identified as model number 7. The sulphate group from the IdoUA(2S) at residue 7 is in binding site X, however the actual positioning of the HS chain is not the expected conformation for binding in nature as it does not lie planar against the FGF2 molecule. Other models identified were models 10 and 11 and again those do not show a good fit to the x-ray crystal structure (data shown in appendix). For this reason these structures were dismissed as viable structures after the docking process. Four structures out of 19 were identified in frame 975 (600ps). Here in this frame, all three iduronate conformations are present in the HS chain (Table 3.5).

Table 3.5

Residue	IdoUA(2S)	IdoUA(2S)	IdoUA
Position in chain	5	7	9
Conformation	1C_4	2S_0	4C_1

Table 3.5 The conformation of the different iduronates in dp10(1) for frame 975.

At this point in the MD simulation, all three iduronate conformations are present in the chain.

Although the HS chain is found in a more central position to bind the FGF2 molecule; again none were thought to be viable docking structures due to the nature of the sulphate groups. The model identified as having the highest affinity binding according to Vina had no sulphate groups in the either binding site (model 1). The 6-O-sulphate group of residue 6, the GlcNS(6S) appears to be the closest to this binding site (binding site Y), however when the structure is rotated this is further from the binding site than first observed.

All the above frames were chosen because of a conformational change. The final frame, frame 1463 was chosen because it corresponded to a change in the torsion data corresponding to both Φ and ψ angles. 20 models were generated from Vina and of those; model 3 and 15 were thought to be the best fit. Model 15 has possibly the best fit model with the IdoUA(2S) of residue 5 in binding position X and the residue 4 near to the binding position Y (when rotated, again this is not as near as first thought). Model 15 has an affinity of -9.9 kcal/mol with rmsd values of rmsd l.b at 11.727 and the u.b at 19.515. This could be an ideal candidate for forming the tertiary complex due to its position against the FGF protein.

Table 3.6

Frame used	Mode	Affinity (kcal/mol)
45	1	-13.0
	2	-11.1
	3	-11.0
	4	-10.9
	5	-10.6
	6	-10.4
	7	-10.4
	8	-10.2
	9	-10.2
	10	-10.1
	11	-10.0

Table 3.6 Energy values for each Vina predicted model for dp10(1), frame 45

Model number 1 has been highlighted as the best highest affinity binding model for the 1 ns MD simulation. This was based on the predicted position of the sulphate groups in relation to the X-ray crystal structure.

3.4.3.2 Oligosaccharide dp10(3):

Model dp10(3) was docked to the FGF2 protein in the file 1BFC using the identified four frames. The time period chosen for analysis in Vina (with the corresponding frame numbers) were 15 ps, 200 ps, 650 ps and 800 ps (frames 25, 125, 1056 and 1300) out of a possible 1600 frames (Table 3.8).

Vina successfully predicted 20 high affinity low energy structures as potential docking conformations during the rigid docking performed for frame 25. None of the predicted structures were deemed close enough to the X-ray crystal structure and so were dismissed. Another 20 structures were predicted for frame 125, of which model number 3 was identified as the best fit model. Here the iduronates at positions 5 and 9 are in the skew boat conformation with the iduronate at position 7 in the 1C_4 conformation. Here it is just the GlcNS(6S) of residue 6 in binding position X. The predicted binding affinity for this model was -10.6 kcal/mol. Models 10, 12 and 18 all had the N-sulphate group from the GlcNS at position 8 in or around binding position Y with model 10 showing the closest fit model amongst these. All had a similar value for the binding affinity (-9.8, -9.6 and -9.0 kcal/mol respectfully); (Table 3.7).

Vina predicted 20 structures for frame 1056, with model number 11 identified as the best fit model. This has an affinity of -9.9 kcal/mol with the IdoUA(2S) of sequence position 5 and the unsulphated IdoUA of position 9 identified in the 4C_1 conformation. The iduronate at sequence position 7 is in the 1C_4 conformation. Here the iduronate at position 5 is in binding position X with the neighbouring residue, the GlcNS of residue 4 in binding position Y. This positioning puts the residues in the opposite direction to those in the crystal structure. This however is not the best fit structure, this is found in frame 1300, which is the final frame for dp10(3). This frame was chosen for the change in torsion data rather than a conformational change. Of the 12 predicted Vina models, model number 3 was thought to be the best fit (Figure 3.37). It has the 2S of IdoUA(2S) at position 5 in the first binding site (site X), then following on from this is the neighbouring residue 4 the GlcNS in the second binding site (site Y). This positioning/placement reflect that in the crystal structure in that it is an iduronate in the first binding position. However on the Goodger

oligosaccharide the reducing end is found towards the left hand side of the protein, which places it in the opposite direction in the x-ray crystal structure. Here the binding affinity for model 3, frame 1300 is -10.8 kcal/mol and the rmsd are as follows; rmsd l.b 12.712 and rmsd u.b 27.207.

Table 3.7

Frame used	Mode	Affinity (kcal/mol)
1300	1	-12.8
	2	-11.4
	3	-10.8
	4	-10.8
	5	-10.4
	6	-10.2
	7	-10.2
	8	-10.2
	9	-10.1
	10	-10.0
	11	-10.0
	12	-9.9

Table 3.7 Energy values for each Vina predicted model for dp10(3), frame 1300

Model number 3 has been highlighted as the best highest affinity binding model for the 1 ns MD simulation. This was based on the predicted position of the sulphate groups in relation to the X-ray crystal structure.

Figure 3.37

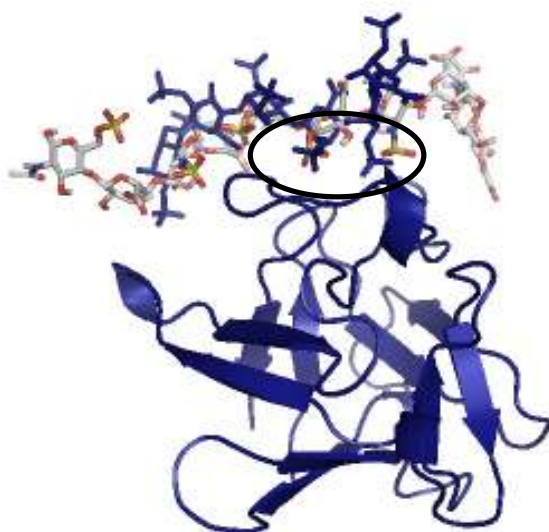


Figure 3.37 The highest affinity binding structure as predicted by Vina for dp10(3) frame 1300, from the rigid 1 ns MD simulation.

The protein molecule in blue shows the FGF2 protein in the x-ray crystal structure, alongside the heparin fragment from the x-ray crystal structure. The lighter coloured decasaccharide is the Goodger model. It has a binding affinity of -10.8 kcal/mol and the rmsd are as follows; rmsd l.b of 12.712 and rmsd u.b of 27.207.

Panel A. Frame 1300, model 3.

Table 3.8

Residue	Frame: 25 Time in ps: 15	Frame: 200 Time in ps: 125	Frame: 1056 Time in ps: 650	Frame: 1300 Time in ps: 800
UA	${}^2\text{H}_1$	${}^2\text{H}_1$	${}^2\text{H}_1$	${}^2\text{H}_1$
GlcNAc	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$
GlcUA	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$
GlcNS	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$
IdoUA(2S)	${}^2\text{S}_0$	${}^2\text{S}_0$	${}^4\text{C}_1$	${}^4\text{C}_1$
GlcNS(6S)	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$
IdoUA	${}^1\text{C}_4$	${}^1\text{C}_4$	${}^1\text{C}_4$	${}^1\text{C}_4$
GlcNS	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$
IdoUA	${}^1\text{C}_4$	${}^2\text{S}_0$	${}^4\text{C}_1$	${}^4\text{C}_1$
GlcNAc (6S)	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$	${}^4\text{C}_1$

Table 3.8 Conformations of all the ring structures for dp10(3) in the frames chosen for docking using Vina.

3.4.3.3 Oligosaccharide dp10(5):

Model dp10(5) was docked to the FGF2 protein in the file 1BFC using the identified two frames. The frames chosen were frame 30 and 1381 (corresponding to 18 ps and 850 ps; Table 3.10). For frame 30, Vina successfully predicted 11 low energy models with model 9 identified as the best fit model (Table 3.11). Here it was residue 4 which was identified with the sulphate group in binding position Y. However there was no identifiable sulphate in binding site X (data in appendix). In this frame the first iduronate at position 5 is in the 1C_4 conformation as in the one at position 7. The position 9 Iduronate is found in the 2S_0 conformation (Table 3.10). Model number 11 was also identified as another best fit model. This has an affinity of -9.9 kcal/mol with the IdoUA(2S) of sequence position 5 and the unsulphated IdoUA of position 9 identified in the 4C_1 conformation.

The data for the two best fit models are identified in Table 3.9. Model 2 for frame 30 has a more negative affinity binding, suggesting it is the best fit model produced from Vina. However; when the structure is rotated using the molecular viewer Pymol, the sulphate groups around the binding position are actually further away from the binding site than first observed.

Frame number 1381 corresponds to 850ps, with Vina successfully predicting 20 low binding structures. Of these, five were deemed as best fit models. Three were identified with the 2S group of the IdoUA(2S) of position 5 in sulphate position X (models 4, 8 and 9, model 9 is shown in Figure 3.38, all others are in the appendix). There are other sulphate groups near to binding site Y, however when rotated these structures are not as close as they first appear (frames 5 and 19; Table 3.11). These other two structures involve firstly the 2S group of residue 7 (the IdoUA(2S) shown in model 5); and residue 6 involving the NS of GlcNS(6S) in the first sulphate position (model 19, Table 3.9).

Table 3.9

Model	4	5	8	9	19
Affinity	-11.0	-11.0	-10.7	-10.7	-9.3
Rmsd l.b	8.241	3.132	7.150	11.497	5.089
Rmsd u.b	21.285	4.550	23.104	24.104	10.658

Table 3.9 The Vina predicted models for dp10(5) frame 1381.

This frame corresponds to 850 ps in time and from 20 predicted models, five were identified based on their positioning near the binding site.

None of these structures discussed above involve two neighbouring sulphate groups occupying the mentioned binding structure identified in the x-ray crystal structure. The occupation of these structures in the binding site involves one site or the other, with no preference to one specific area. The chosen frames do not contain all three mentioned iduronate conformations (1C_4 , 2S_0 and 4C_1), which have been identified in the known active sequences (see Discussion section). The inability of the final iduronic acid to occur in all three conformations may play a major part in this.

All these structures have the same positioning with the reducing terminal found towards the left hand side of the protein, putting it opposite to the known x-ray crystal structure.

Figure 3.38

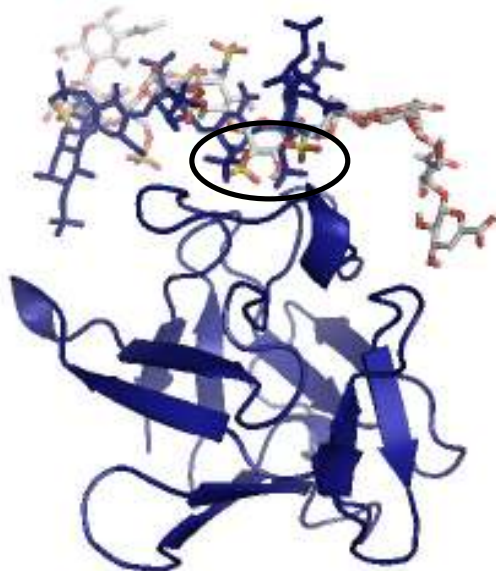


Figure 3.38 The highest affinity binding structure as predicted by Vina for dp10(5) frame 1381, from the rigid 1 ns MD simulation.

The protein molecule in blue shows the FGF2 protein in the x-ray crystal structure, alongside the heparin fragment from the x-ray crystal structure. The lighter coloured decasaccharide is the Goodger model. It has a binding affinity of -10.7 kcal/mol and the rmsd are as follows; rmsd l.b of 11.497 and rmsd u.b of 24.593.

Panel A. Frame 1381, model 9.

Table 3.10

Residue	Frame: 30 Time in ps: 18	Frame: 1381 Time in ps: 850
UA	2H_1	2H_1
GlcNAc	4C_1	4C_1
GlcUA	4C_1	4C_1
GlcNS	4C_1	4C_1
IdoUA(2S)	1C_4	2S_0
GlcNS(6S)	4C_1	4C_1
IdoUA(2S)	1C_4	2S_0
GlcNS(6S)	4C_1	4C_1
IdoUA	2S_0	4C_1
GlcNAc	4C_1	4C_1

Table 3.10 Conformations of all the ring structures for dp10(5) in the frames chosen for docking using Vina

Table 3.11

Frame used	Mode	Affinity (kcal/mol)
1381	1	-11.5
	2	-11.2
	3	-11.2
	4	-11.0
	5	-11.0
	6	-10.9
	7	-10.8
	8	-10.7
	9	-10.7
	10	-10.4
	11	-10.4
	12	-10.3
	13	-10.3
	14	-10.1
	15	-9.9
	16	-9.8
	17	-9.7
	18	-9.5
	19	-9.3
	20	-9.1

Table 3.11 Energy values for each Vina predicted model for dp10(5), frame 1381

Model number 9 has been highlighted as the best highest affinity binding model for the 1 ns MD simulation. This was based on the predicted position of the sulphate groups in relation to the X-ray crystal structure.

3.4.3.4 Oligosaccharide dp10(7):

Finally, dp10(7) was the final sequence chosen for docking analysis and again it was docked to the FGF2 protein in the file 1BFC using the identified three frames. The frames chosen were frames 50, 200 and 1463 (corresponding to 31 ps, 125 ps and 900 ps). This is the only structure with four iduronic acid ring structures in the chain. Two of these are unsulphated and two have the 2S group as a chain modification (see Modifications, p32). The IdoUA at position 3 is the additional residue identified in the chain; in previous sequences this has always been found as a GlcUA. Throughout this 1000 ps time period it did not change from the 4C_1 conformation (Table 3.13). This in itself is something to consider as it was initially thought that the iduronates found in an internal position were found only in the 1C_4 and 2S_0 conformations (Ferro *et al.* 1990; Rabenstein 2002).

The three other iduronates all change conformation at some point during this 1000 ps timeframe, so the first frame chosen for docking was frame 50 which equates to 31 ps in time. Here all the iduronates are found in the 1C_4 conformation apart from the previously mentioned IdoUA at position 3. Of the 20 proposed models it was model 10 which showed similarity to the binding site. The 2S group of residue 7, the IdoUA(2S) interacts around the binding site X (Figure 3.39). This predicted structure has a binding affinity of -10.0 kcal/mol. Model 19 also has residue 7 in a similar position and model 18 has residue 6 in the same binding position (data shown in appendix). This had a lower binding affinity of -9.2 kcal/mol (Table 3.14).

Frame 200 also produced 20 proposed models. Models 2, 7 and 15 have the whole binding site occupied, with the first two models showing the same residues interaction in the same position (the GlcNS of residue 4 and the IdoUA(2S) of residue 5). This is found with the non-reducing terminal towards the left hand side of the protein and all three iduronate conformations are present. Here the iduronate at position 3 was identified in the 4C_1 conformation. Iduronates at positions 5 and 7 were in the 1C_4 conformation and finally the iduronate at position 9 was in the skew boat conformation.

Model 15 shows residues 6 and 7 interacting in the binding site of the protein. The binding data for these models show model number 2 has a greater affinity to bind to the protein than the other models mentioned (Table 3.12).

Table 3.12

Model	2	7	15
Affinity	-10.7	-9.8	-9.0
Rmsd l.b	12.847	12.739	12.684
Rmsd u.b	15.900	15.856	17.995

Table 3.12 The best fit Vina predicted models for dp10(7); frame 200.

Here it was identified that model 15 was the best fit model as it had the best fit according to the X-ray crystal structure.

However model 15 was thought to be the best fit model in terms of binding to the known binding site identified in the x-ray crystal structure (Figure 3.38).

The final frame; number 1463 corresponds to 900 ps and in this frame residues 5, 7 and 9 have all changed conformation. As mentioned previously the IdoUA at position 3 does not change at all during this simulation.

The only Vina predicted models which interact with the binding site are models 5 and 15 and it is the NS group of residue 6, the GlcNS(6S) which is found in binding site Y. Here model 5 is thought to be the best fit model (data shown in appendix).

Figure 3.39

Figure 3.39 The highest affinity binding structure as predicted by Vina for dp10(7) frame 200, from the rigid 1 ns MD Simulation.

The protein molecule in blue shows the FGF2 protein in the x-ray crystal structure, alongside the heparin fragment from the x-ray crystal structure. The lighter coloured decasaccharide is the Goodger model. It has a binding affinity of -9.0 kcal/mol and the rmsd are as follows; rmsd l.b of 12.684 and rmsd u.b of 17.995.

Panel A. Frame 200, model 15.

Table 3.13

Residue	Frame: 50 Time in ps: 31	Frame: 200 Time in ps: 125	Frame: 1463 Time in ps: 900
UA	2H_1	2H_1	2H_1
GlcNS	4C_1	4C_1	4C_1
IdoUA	4C_1	4C_1	4C_1
GlcNS	4C_1	4C_1	4C_1
IdoUA(2S)	1C_4	1C_4	2S_0
GlcNS(6S)	4C_1	4C_1	4C_1
IdoUA(2S)	1C_4	1C_4	2S_0
GlcNS	4C_1	4C_1	4C_1
IdoUA	1C_4	2S_0	4C_1
GlcNAc (6S)	4C_1	4C_1	4C_1

Table 3.13 Conformations of all the ring structures for dp10(7) in the frames chosen for docking using Vina.

Table 3.14

Frame used	Mode	Affinity (kcal/mol)
250	1	-10.9
	2	-10.7
	3	-10.2
	4	-10.0
	5	-9.9
	6	-9.8
	7	-9.8
	8	-9.7
	9	-9.7
	10	-9.6
	11	-9.6
	12	-9.6
	13	-9.4
	14	-9.3
	15	-9.0
	16	-9.0
	17	-9.0
	18	-8.9
	19	-8.9
	20	-8.6

Table 3.14 Energy values for each Vina predicted model for dp10(7), frame 250

Model number 15 has been highlighted as the best highest affinity binding model for the 1 ns MD simulation. This was based on the predicted position of the sulphate groups in relation to the X-ray crystal structure.

3.4.4 The 10 ns Docking simulations

10 ns molecular dynamic simulations were undertaken and so docking analysis was also carried on specific frames within these sequences. In each of these sequences one frame equals 1 ps in time; therefore 10,000 ps (10 ns) equal 10,000 frames. For the longer docking analysis the 1BFC protein was again used as it was in the 1000 ps docking analysis but is shown in a different colour (purple). This was used to signify the different timescales in the docking calculations.

3.4.4.1 Oligosaccharide dp10(1)

The four chosen frames for the rigid docking analysis in dp10(1) were frame numbers 200, 1300, 3000 and 8300 (corresponding to 200 ps, 1.3 ns, 3.0 ns and 8.3 ns, Table 3.15). Again these were based on iduronic acid ring structures and also torsion angle data. Vina produced 11 potential docking structures for frame 200 with models 1, 2, 5 and 7 occupying just one binding position. Models 1 and 2 have only the 2S group of residue 7, the IdoUA(2S) in the binding position. Model 5 has residue 8 the GlcNS(6S), near to the binding site and finally model 7 is identified with the sulphate in binding position X.

There were 19 structures created from frame 1300 and of these it was only model 11 which had any similarity to the x-ray crystal structure. For frame 3000 Vina successfully predicted 20 structures and of those; model 1 was identified as the best fitting, lowest energy structure. This however had only the one sulphate group of residue 6 (the NS part of the GlcNS(6S)) in the correct binding position. The other models identified in Vina do not have sulphate groups in the specific binding site. There were groups identified in and around the vicinity of the binding site, such as model 7; which had the NS part of the GlcNS(6S) of residue 6 approximately nearby; model 20 which had the 2S group of the IdoUA at position 5 near to binding site Y.

Then finally frame 8300, Vina successfully predicted 13 structures. The lowest energy structure identified was model 2, which had only residue 8 (the GlcNS(6S)) in proximity to the binding site.

Models 4, 9, 10 and 11 showed a similar interaction (data not shown). The final model, model 12 was the only frame which when analysed had two residues interacting in the binding site. These interactions were from residues 5 and 4; the 2S and NS groups respectively. Here model 12 for frame 8300 was thought to be the best fit model according to the x-ray crystal structure (Table 3.16). Here there is no presence of the 1C_4 conformation, with residues 5 and 7 in the skew boat conformation and the terminal residue iduronate in the 4C_1 conformation (Figure 3.40).

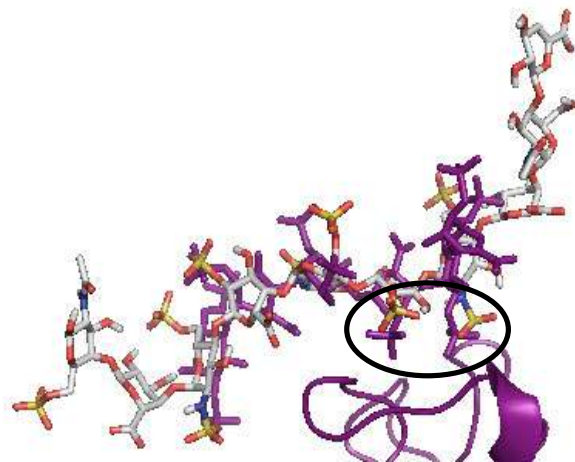
Figure 3.40

Figure 3.40 The highest affinity binding structure as predicted by Vina for dp10(1) frame 8300, with the best fit for the x-ray crystal structure for the rigid 10 ns MD simulation.

The protein molecule in purple shows the FGF2 protein in the x-ray crystal structure, alongside the heparin fragment from the x-ray crystal structure. This is the same structure used in the shorter simulations, but for the purpose of the longer MD simulations is shown in a different colour. The lighter coloured decasaccharide is the Goodger model. It has a binding affinity of -9.2 kcal/mol and the rmsd are as follows; rmsd l.b of 9.778 and rmsd u.b of 16.151.

Panel A. Frame 8300, model 12.

Table 3.15

Residue	Frame: 200 Time in ns: 0.2	Frame: 1300 Time in ns: 1.3	Frame: 3000 Time in ns: 3	Frame: 8300 Time in ns: 8.3
IdoUA(2S) Residue 5	1C_4	2S_0	2S_0	2S_0
IdoUA(2S) Residue 7	1C_4	2S_0	1C_4	2S_0
IdoUA Residue 9	1C_4	1C_4	4C_1	4C_1

Table 3.15 Conformations of the iduronic acid ring structures for dp10(1) for the 10 ns MD simulation.

Table 3.16

Frame used	Mode	Affinity (kcal/mol)
8300	1	-11.9
	2	-10.5
	3	-10.4
	4	-10.1
	5	-9.9
	6	-9.8
	7	-9.8
	8	-9.7
	9	-9.5
	10	-9.4
	11	-9.3
	12	-9.2
	13	-9.1

Table 3.16 Energy values for each Vina predicted model for dp10(1), frame 8300

Model number 12 has been highlighted as the best highest affinity binding model for the 10 ns MD simulation. This was based on the predicted position of the sulphate groups in relation to the X-ray crystal structure.

3.4.4.2 Oligosaccharide dp10(3):

Those frames identified for dp10(3) were frame numbers 300, 430, 550, 2900 and finally 6000 (corresponding to 300 ps, 430 ps, 550 ps, 2.9 ns and finally 6.0 ns). None of the predicted Vina models placed the Goodger structures in the binding site identified in the x-ray crystal structure. This could potentially be due to the lack of all three conformations which had been previously identified in the active oligosaccharides in the shorter simulations (Table 3.17).

Table 3.17

Residue	Frame: 300 Time in ns: 0.3	Frame: 430 Time in ns: 0.43	Frame: 550 Time in ns: 0.55	Frame: 2900 Time in ns: 2.9	Frame: 6000 Time in ns: 6.0
IdoUA(2S) Residue 5	2S_0	2S_0	2S_0	1C_4	2S_0
IdoUA Residue 7	1C_4	1C_4	2S_0	4C_1	4C_1
IdoUA Residue 9	1C_4	2S_0	4C_1	4C_1	4C_1

Table 3.17 Conformations of the iduronic acid ring structure for dp10(3) for the 10 ns MD simulation.

3.4.4.3 Oligosaccharide dp10(5):

Four frames were identified in dp10(5) corresponding to changes in the Cremer-Pople puckering parameters. The chosen frames were frame 500, 1050, 4000 and 6000 (corresponding to 0.5 ns, 1.05 ns, 4.0 ns and 6.0 ns).

Frame 500 had one identified best fit model out of the 13 Vina predicted models. This was in fact model 13, the model with the lowest affinity binding value according to Vina. Here the NS group of residue 6, the GlcNS(6S) was identified in binding position X (shown in appendix). The flanking iduronic acid of residue 7, which is in the 2S_0 conformation, is in close proximity to the binding site. Vina successfully predicted 12 models for frame 1050, of which model 6 was identified as the best fit model. Here residue 5, which is in the 1C_4 conformation is in close proximity to binding site X and the neighbouring residue, the glucosamine in close proximity to binding site Y. During this frame all three conformations are present in each of the iduronates (Table 3.19), with the first IdoUA(2S) at position 5 interacting in that identified binding site.

There are no Vina predicted models which fit the x-ray binding site for frame 4000 (4.0 ns). This could be due to the overall change in shape of the molecule as the iduronic acid at position 5 undergoes its first conformational change to the skew boat conformation. Finally frame 6000 produced 20 Vina predicted models, of which model number 5 was identified as the best fit model. Here the NS group of residue 6, the GlcNS(6S) was identified in binding position X (Figure 3.41). The flanking iduronic acid of residue 7, which is in the 2S_0 conformation, is near to the binding site but not in it (the IdoUA(2S) is identified in the 4C_1 conformation around the second binding position). This is also the best fit structure in terms of the positioning of the HS chain across the protein. The non-reducing terminal is found towards the left of the protein matching the x-ray crystal structure. It then spans the whole of the protein again like that found in the crystal structure. The binding affinity for this model is 9.5 kcal/mol, which is less than those identified in the shorter 1000 ps MD simulation. There was no change in conformation between frame 4000

and 6000 with just the change in dihedral angle taken into account. This slight change has allowed the Vina predicted model for frame 6000 to produce a model which fits the binding site better than those predicted for frame 4000. The model for frame 6000 is the best fit model of all the models predicted by Vina for the whole of the 10 ns MD simulation (Table 3.18).

Table 3.18

Frame used	Mode	Affinity (kcal/mol)
6000	1	-11.0
	2	-10.9
	3	-10.4
	4	-10.2
	5	-10.1
	6	-10.1
	7	-10.0
	8	-9.9
	9	-9.9
	10	-9.8
	11	-9.4
	12	-9.1
	13	-9.0
	14	-8.9
	15	-8.8
	16	-8.6
	17	-8.6
	18	-8.6
	19	-8.6
	20	-8.5

Table 3.18 Energy values for each Vina predicted model for dp10(5), frame 6000

Model number 15 has been highlighted as the best highest affinity binding model for the 10 ns MD simulation. This was based on the predicted position of the sulphate groups in relation to the X-ray crystal structure.

Figure 3.41

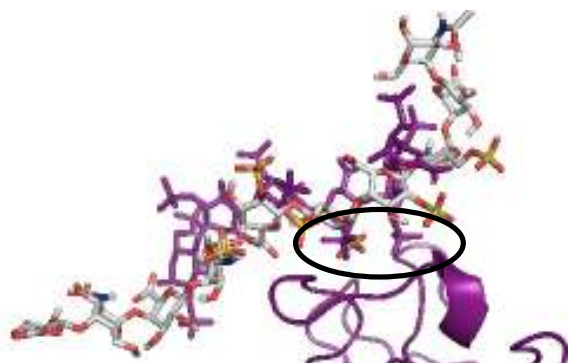


Figure 3.41 The highest affinity binding structure as predicted by Vina for dp10(5) frame 6000, with the best fit for the x-ray crystal structure for the rigid 10 ns MD simulation.

The protein molecule in purple shows the FGF2 protein in the x-ray crystal structure, alongside the heparin fragment from the x-ray crystal structure. The lighter coloured decasaccharide is the Goodger model. It has a binding affinity of -10.1 kcal/mol and the rmsd are as follows; rmsd l.b of 7.375 and rmsd u.b of 22.001.

Panel A. Frame 6000, model 5.

Table 3.19

Residue	Frame: 500 Time in ns: 0.5	Frame: 1050 Time in ns: 1.05	Frame: 4000 Time in ns: 4.0	Frame: 6000 Time in ns: 6.0
IdoUA(2S) Residue 5	1C_4	1C_4	2S_0	2S_0
IdoUA(2S) Residue 7	1C_4	2S_0	2S_0	2S_0
IdoUA Residue 9	2S_0	4C_1	4C_1	4C_1

Table 3.19 Conformations of the iduronic acid ring structures for dp10(5) for the 10 ns MD simulation.

3.4.4.4 Oligosaccharide dp10(7):

Three frames were chosen for the final deca-saccharide used in the docking analysis. These were frames 250, 750 and 5000 (corresponding to 0.25 ns, 0.75 ns and 5.0 ns). This deca-saccharide contains four iduronic acid ring structures in the chain making it potentially the most flexible sequence of all the structures. The first iduronate in this sequence is found at position 3 and does not change conformation throughout the whole 10 ns timescale (Table 3.20). This reflects the result identified in the shorter 1000 ps timescale shown previously (see p182).

For frame 250, Vina successfully predicted 20 models of which model 14 had been identified as the best fit model (the only predicted structure to fit into the known binding structure). Here the sulphate group around the binding position is the 2S group of the IdoUA(2S) of residue 5 (Figure 3.42). This is positioned in the first of the two sulphate positions (binding site X). It has a binding affinity of -9.7 kcal/mol; (Table 3.21). For frame 750 there was only one identifiable frame in the binding site and this was model number 1; the lowest energy binding structure. Here only the NS group of residue 8 (the GlcNS) was identified near to the binding site, however once the model is rotated it becomes clear that the NS group is further away than first identified. The final frame (frame 5000) gave a number of predicted models close to that of the x-ray crystal structure with model number 2 showing only the NS of residue 6 (the GlcNS(6S)) in position Y. Model 8 shows the binding site complete with the 2S group of the IdoUA(2S) at residue 7 in binding position X and the NS group of GlcNS (residue 8) in binding position Y. Vina predicted models 10 and 12 show a similar result to model 8. Other models identify one sulphate group in the binding site, such as the NS of residue 4 in position X for model 12. Model 20, with the highest affinity binding site also mirrors the results identified in models 8, 10 and 12. This shows that it does not have to be the lowest affinity binding model which Vina predicts to bind in the correct binding position and therefore those with a higher affinity binding value should not be so easily dismissed.

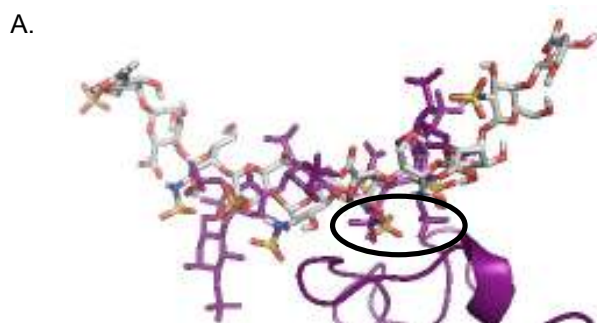
Figure 3.42

Figure 3.42 The highest affinity binding structure as predicted by Vina for dp10(7) frame 250, with the best fit for the x-ray crystal structure for the rigid 10 ns MD simulation.

The protein molecule in purple shows the FGF2 protein in the x-ray crystal structure, alongside the heparin fragment from the x-ray crystal structure. The lighter coloured decasaccharide is the Goodger model. It has a binding affinity of -9.7 kcal/mol and the rmsd are as follows; rmsd l.b of 8.335 and rmsd u.b of 24.010.

Panel A. Frame 250, model 14.

Table 3.20

Residue	Frame: 250 Time in ns: 0.25	Frame: 750 Time in ns: 0.75	Frame: 5000 Time in ns: 5.0
IdoUA Residue 3	4C_1	4C_1	4C_1
IdoUA(2S) Residue 5	1C_4	1C_4	2S_0
IdoUA(2S) Residue 7	1C_4	1C_4	2S_0
IdoUA Residue 9	1C_4	2S_0	4C_1

Table 3.20 Conformations of the Iduronic acid ring structures for dp10(7) for the 10 ns MD simulation.

Table 3.21

Frame used	Mode	Affinity (kcal/mol)
250	1	-11.4
	2	-11.0
	3	-11.0
	4	-10.5
	5	-10.4
	6	-10.4
	7	-10.4
	8	-10.3
	9	-10.3
	10	-10.2
	11	-10.0
	12	-10.0
	13	-9.9
	14	-9.7
	15	-9.6
	16	-9.6
	17	-9.6
	18	-9.5
	19	-9.4
	20	-9.4

Table 3.21 Energy values for each Vina predicted model for dp10(7), frame 250

Model number 14 has been highlighted as the best highest affinity binding model for the 10 ns MD simulation. This was based on the predicted position of the sulphate groups in relation to the X-ray crystal structure.

3.4.5 Flexible Dockings

The chosen frames which were analysed for rigid docking calculations were again analysed with all dihedral angles left rotatable (all other bonds were set as rigid). This was to allow the torsion angles to fall into their own preferred angle and to allow Vina to dock them in that sense. Each frame took an average of 48 hours to run; compared to just 12 hours for the shorter docking simulations. Again the same process was followed, with the number of modes sent to the output file kept at 40.

3.4.5.1 The 1 ns Flexible Docking Analysis

Vina successfully predicted 20 potential binding structures for the first of the flexible docking studies (frame 45) for dp10(1). The best fit binding model was identified as model 1 which was identified with residue 8 (the glucosamine) in binding site Y of the x-ray structure. Due to the nature of the bending in the chain there is also the sulphate group from residue 5 (the IdoUA(2S)) identified in binding site X. However when this is rotated it is actually much further away from the binding site than first thought (Figure 3.43, Panel A). Models 2, 6, 7 and 10 show similar results and model 14 is identified with the 2S group of the IdoUA(2S) of residue 7 in binding site X of the crystal structure (data not shown). Model 1 has the highest affinity of -4.6 kcal/mol (Table 3.22) which; compared to values obtained previously is considered to be a relatively low value for the binding affinity.

Frame 350 produced 20 Vina predicted models of which only model 17 had any interaction with the binding site. This was only from reducing end GlcNAc(6S) residue of which it was the NS group identified near to binding site Y. The same again was identified for frame 975, with only one model interacting with the binding site. This was model 18 and again it showed the reducing end GlcNAc(6S) near to the binding site (data not shown).

The final frame; frame 1463 corresponds to 900 ps in time and Vina successfully predicted 20 models. Of these it was identified that model 1 was the best fit model against the x-ray crystal structure. Again it was the GlcNAc(6S) near the binding position identified in the x-ray crystal

structure. Due to the positioning of the chain along the FGF2 protein and also the interaction of the non-reducing terminal with the binding site, these models can be dismissed as likely docking structures.

Dp10(3) as dp10(1) before it did not produce any best fit models which could be deemed as important when referring to the binding site. Model 6 was identified as the best fit for frame 25 with model 18 identified for both frames 200 and 1056. All had binding affinity values of -7.8 kcal/mol and -7.7 kcal/mol respectively; which were considered quite low compared to the binding affinities identified in their corresponding rigid structures. For frame 200 it was residue 8 (the glucosamine) which was identified in the centre of the binding structure, but was not found in either site X or Y. Residue 6; also the glucosamine residue for frame 1056 was found in the same position as that in frame 200. The final frame, 1300 gave model 1 as the best fit model and again this had a low binding affinity of -8.4 kcal/mol. It was the NS group of residue 6 (GlcNS(6S)) which was identified close to the binding site (data shown in appendix).

Only one frame produced any docking data which somewhat matched the x-ray crystal structure for dp10(5) which was frame 1381. This was a model which Vina had predicted as the last model out of 20 to fit into the binding site due to the binding affinity being the lowest out of all the 20 predicted structures (-7.4 kcal/mol; Table 3.22). Interesting this was the only identified structure which involved the 2S group of the IdoUA(2S) of residue 5 interacting in the binding site (Figure 3.43, Panel B). Although this group is identified in the binding region of the protein it was not identified specifically in either binding site X or Y. Again this structure was dismissed on grounds that it did not fully comply with the x-ray crystal structure.

The final decasaccharide used in the 1000 ps docking analysis was dp10(7). Here like in dp10(5), it was residue 5 which was identified in the FGF binding site in frame 50. Here it was model 11 which showed the 2S group in the binding position; however once the structure was rotated this group is actually nowhere near to the binding site (Figure 3.43, Panel C shown in appendix). This structure had a binding affinity of -9.3 kcal/mol.

Table 3.22

Frame used	Mode	Affinity (kcal/mol)	Frame used	Mode	Affinity (kcal/mol)
45	1	-4.6	1381	1	-8.3
	2	4.6		2	-8.2
	3	-4.5		3	-8.0
	4	-4.4		4	-8.0
	5	-4.4		5	-8.0
	6	-4.3		6	-7.9
	7	4.2		7	-7.8
	8	-4.1		8	-7.8
	9	-4.1		9	-7.8
	10	-4.1		10	-7.7
	11	-4.1		11	-7.7
	12	-4.1		12	-7.7
	13	-4.1		13	-7.7
	14	-4.0		14	-7.7
	15	-4.0		15	-7.6
	16	-4.0		16	-7.6
	17	-3.9		17	-7.6
	18	-3.9		18	-7.5
	19	-3.9		19	-7.4
	20	-3.8		20	-7.4

Table 3.22 Energy values for each Vina predicted model for the flexible dockings for dp10(1), frame 45 and frame 1381 for dp10(5), from the 1 ns MD simulation.

Models 1 and 20 have been highlighted as the best, high affinity binding models for the 1 ns MD simulation. This was based on the predicted position of the sulphate groups in relation to the X-ray crystal structure.

Figure 3.43

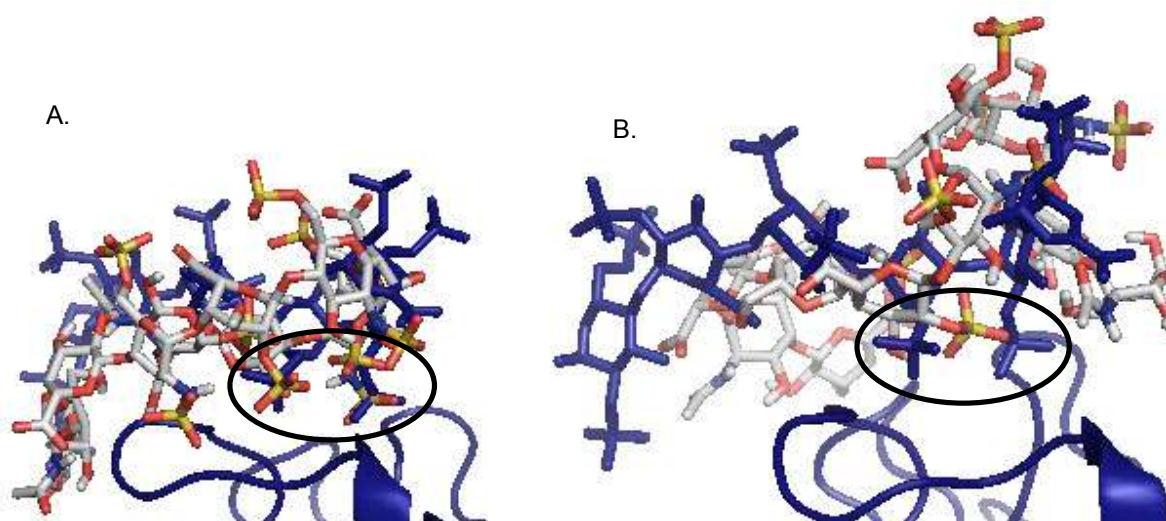


Figure 3.43 The highest affinity binding structure as predicted by Vina for dp10(1) frame 45 and dp10(5) frame 1381, with the best fit for the x-ray crystal structure for the flexible 1 ns MD simulation.

The protein molecule in blue shows the FGF2 protein in the x-ray crystal structure, alongside the heparin fragment. The lighter coloured decasaccharide is the Goodger oligosaccharides. Frame 45 has a binding affinity of -4.6 kcal/mol and the rmsd are as follows; rmsd l.b of 0.000 and rmsd u.b of 0.000. Frame 1381 has a binding affinity of -7.4 kcal/mol and the rmsd are as follows; rmsd l.b of 5.860 and rmsd u.b of 14.899.

Panel A. dp10(1). Frame 45, model 1.

Panel B. dp10(5). Frame 1381, model 20.

3.4.5.2 The 10 ns Flexible Docking Analysis

The longer simulations did not produce any reliable docking results when compared to the 1BFC x-ray crystal structure. As in the 1000 ps docking simulations above, all the dihedral angles were left rotatable with all the other bonds kept rigid. Frame 200 from dp10(1) produced the only Vina predicted structure which in any way resembled the known binding site. This again was identified by the sulphate groups highlighted (from the x-ray crystal structure):

IdoUA(2S) – GlcNS(6S)

Residue 4 Residue 5

This sequence is the same as that identified in dp10(1), with the 2S group of the IdoUA(2S) of residue 5 in binding site X and the NS group of GlcNS(6S) of residue 6 in binding site Y (Figure 3.44). The best fit model here was model 12 out of the 20 models produced by Vina and had a binding affinity of -8.6 kcal/mol; (Table 3.23). Model 8 was also identified as another low energy binding model; however this had only residue 8 identified near to the binding site.

Figure 3.44

A.

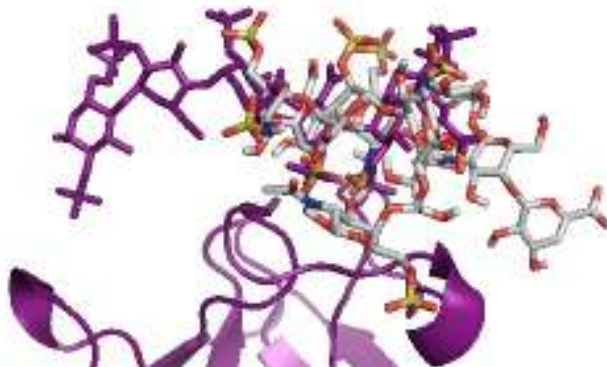


Figure 3.44 The highest affinity binding structure as predicted by Vina for dp10(1) frame 200, with the best fit for the x-ray crystal structure for the flexible 10 ns MD simulation.

The protein molecule in purple shows the FGF2 protein in the x-ray crystal structure, alongside the heparin fragment from the x-ray structure. The lighter coloured decasaccharide is the Goodger oligosaccharide. It has a binding affinity of -8.6 kcal/mol and the rmsd are as follows; rmsd l.b of 8.797 and rmsd u.b of 16.421.

Panel A. Frame 200, model 12.

Table 3.23

Frame used	Mode	Affinity (kcal/mol)
250	1	-9.4
	2	-9.3
	3	-9.0
	4	-9.0
	5	-8.9
	6	-8.9
	7	-8.8
	8	-8.7
	9	-8.6
	10	-8.6
	11	-8.6
	12	-8.6
	13	-8.6
	14	-8.5
	15	-8.5
	16	-8.4
	17	-8.4
	18	-8.4
	19	-8.2
	20	-8.2

Table 3.23 Energy values for each Vina predicted model for the flexible dockings for dp10(1), frame 200 in the 10 ns MD simulation.

Model number 1 has been highlighted as the best highest affinity binding model for the 10 ns MD simulation. This was based on the predicted position of the sulphate groups in relation to the X-ray crystal structure.

3.4.6 Docking Implications

For the amount of work carried out in this study on HS:FGF2 docking interactions it is very difficult to determine the differences between active and inactive oligosaccharides. It can be said that oligosaccharide dp10(1) gave the best docking results in terms of binding affinity and also the amount of Vina predicted models created throughout the docking section of this study. The highest affinity binding identified was in the 1 ns docking for dp10(1) with all bonds set rigid. It was frame 1 out of the 11 Vina predicted models and had a binding affinity of -13.0 kcal/mol. This was by far the best binding affinity for all the active and inactive oligosaccharides. This model had GlcNS(6S)-IdoUA(2S) of residues 6 and 7 in the known binding site. All of the iduronates in this frame were interestingly in the 1C_4 conformation. Oligosaccharide dp10(1) also produced good Vina predicted models for the 1 ns flexible docking and for both flexible and rigid dockings in the 10 ns docking analysis. For the flexible 1 ns docking, it was also model 1 which was deemed as the best fit model with a binding affinity of -4.6 kcal/mol. This value is much lower than those identified in the rigid dockings and could be due to the placement of the HS fragment in the docking position. In fact in all the flexible dockings the HS chain was wrapped around the protein and/or itself and so this suggests an inability for Vina to predict the correct binding, using the default settings, which were set at the start of the process. This led to the assumption that the majority of the flexible dockings can be completely dismissed as they are just not an accurate reflection of the binding seen in many of the x-ray crystal structures (Digabriele *et al.*, 1998; Faham *et al.*, 1996; Pellegrini *et al.*, 2000; Schlessinger *et al.*, 2000). However, with some refinement of the settings, such as a widening of the search area on the protein which was set by the size X, Y, and Z coordinates. Also the centre of the box could be checked and altered if needed (see p124). Another area that could be altered is the 'exhaustiveness', which can be changed manually at the start of the Vina docking process. This was set to 8, which was the default setting but can be increased. This can add considerably large amounts of computing time onto each individual docking process, so for the purpose of this study was kept as the default

setting. However, based on the flexible docking results this would definitely be an area to monitor, changing the setting as necessary.

The other active oligosaccharide; dp10(3) produced the second highest affinity binding value of the best fit models. This was from again from 1 ns MD simulation with the rigid dockings. Here is was frame 1300, model 3 of 12 which had a binding affinity of -10.8 kcal/mol. Interestingly; the iduronate conformations present in this frame were 4C_1 , 1C_4 and 4C_1 respectfully. For both the active models, in the frames chosen as being the best fit models neither contained the skew boat conformation, which the puckering-torsion data indicated as being essential for FGF2 binding.

The inactive oligosaccharides; overall gave lower binding affinities for the rigid structures according to the Vina predicted models. This was even the case when looking at the model Vina predicted as having the highest binding affinity. This could have great implications for binding, between active and inactive sequences.

If these results could be refined and if time was not such a factor then the docking analysis could have provided some invaluable data. Carbohydrate-protein docking interactions are at the forefront of modelling simulations, particularly between HS-FGF interactions. It will be MD simulations, alongside docking simulations which will ultimately allow for carbohydrate-protein interactions to go one step closer, to aid in specific drug design for HS-FGF interaction. With more time and the ability to change some of the default settings, it will be possible to determine the differences between the active and inactive HS oligosaccharides presented in this report.

Chapter 4: Discussion

4.1 The Digestion of HS by Heparinase III

Many studies use the enzyme heparinase III to degrade polymeric HS into the shorter oligosaccharide chain lengths. The quantity of enzyme required is not usually addressed, but is important when determining how much is needed to fully digest the HS sample.

Cleavage of HS occurs through a β -elimination, which generates the un-natural $\Delta^4, 5$ unsaturated uronic acid at the non-reducing end of the oligosaccharide. It absorbs U.V. light at a wavelength of 232 nm, making it ideal for following the progression of the heparinase digestion. Long incubation times and large amounts of buffer lead to difficulties in the spectrometer readings so the minimum amount of both is essential.

The minimum amount of enzyme required for the digestion to be classed as fully completed was found to be 140 mU to digest 500 mg of HS.

4.2 Gel Filtration and SAX-HPLC Chromatography of HS Octasaccharides

Each SAX-HPLC profile from the octasaccharides showed variations in the digested sample. This could have been due to the slight overloading of the gel filtration column, leading to overlapping of the different sized oligosaccharides. This would have lead to possible overlapping of peaks and the presence of contaminants in the sample. Any slight overlap in the sample leads to a increase in sample heterogeneity in terms of size and length; making purification of homogenous size and charge defined oligosaccharides more difficult. The extent of the enzyme digestion should be closely monitored so that all samples are digested completely. Partial digests can lead to a more complex SAX-HPLC profile.

4.3 Analysis of HS dp8 NMR Profiles

The aim of the study was the purification of homogenous size and charged defined dp8 oligosaccharides; with an aim to assign and determine the structure of these samples using NMR structural refinement. Unfortunately due to lack of sample in all those purified this task proved to be too difficult. At each stage of the purification process, material was lost making the purification process more difficult with every step taken. The amounts actually produced by enzyme degradation for these larger oligosaccharides is relatively small in comparison to the smaller fragments previously purified (Murphy 2007). With each oligosaccharide size increase, the complexity of the possible HS sequences available becomes more complex. This makes the number of each possible sequence available increase on quite a large scale. Current estimates for dp8 sized oligosaccharides are that there could potentially be more than 5 billion different combinations and therefore purifying these would be impossible. This massive increase in sequence diversity is due to the nature of the modification process, which is non-random and very highly regulated (Wei *et al.*, 1993). The larger the oligosaccharide fragments are that are purified; the less there is produced by the enzyme digestion, so this also makes purifying larger quantities needed for characterization much more difficult. These issues combined have made it extremely difficult to produce enough samples, purified to the level required for NMR analysis. For a reasonable NMR profile, 200 mg of purified sample is required with an average of 150 mg produced for the HS dp8 oligosaccharides. This was collected over a two and a half year period and gives some indication of the amount of time and energy required to produce these amounts needed (Figure 3.4).

4.4 The Requirement for Binding

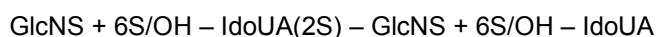
Much of the work carried out in literature has been focused on the requirements for binding between HS/heparin chains and proteins, with most studies being on the minimum length required to form an active signalling complex (specifically for many of the FGF family, IL-8, AT III

and EGF). In actual fact this signalling complex is far more complicated than first thought, as activity is not only due to the length of chain. Many factors need to be taken into account; including sulphation pattern, sulphate content and placement within the chain, iduronate conformation and torsional geometry.

In reality, HS/heparin chains are not rigid and static in shape. In actual fact they sample many different configurations and 3-dimensional shapes over a very small timeframe; constantly oscillating between conformations and torsional geometries, enabling them to sample different conformations until they identify the required one needed for binding (Harmer 2006). It is clear from this study and from many other studies (DiGabriele *et al.*, 1998; Faham *et al.*, 1996; Maccarana *et al.*, 1993) to see that affinity is not dependant on one thing alone (specifically the size of the chain). Instead it takes into account many different requirements including chain length, sulphation pattern and density, the GAG backbone and the internal residue conformation (Guglieri *et al.*, 2008).

4.4.1 The Minimum Length Required for Binding

The suggested minimum chain length required for binding and activating specific proteins will be different for each. A minimum HS chain length for binding FGF2 has been identified as a dp4 sized oligosaccharide (Guglieri *et al.*, 2008), where as other studies have shown larger HS sizes are required, such as a decasaccharide oligosaccharide (Goodger *et al.*, 2008). The minimum tetrasaccharide identified has a sequence of:



(Guglieri *et al.*, 2008).

An oligosaccharide longer than dp4 would then allow for further contacts to be made to the protein ligand via the NS-domains but this contact would likely occur outside of the core binding

site (Deakin *et al.*, 2009). Genuine FGF:FGFR interactions are much more likely to involve longer saccharide chains as those at the physiological level are 100-300 saccharide units long (Harmer *et al.*, 2006); much longer than those studied at any level to date.

A number of HS oligosaccharides have been identified with the ability to display high affinity binding to hepatocyte growth factor/scatter factor (HGF/SF) which is secreted by mesenchymal cells to regulate cell growth, motility and morphogenesis. The minimum sizes identified in this study were trisaccharides with differing affinities; and a tetrasaccharide with high affinity binding (Figure 4.1). Here the final trisaccharide (number 4) is prepared from tetrasaccharide (number 3), and is generated by mercuric ion-catalyzed removal of the unsaturated hexuronate from the non-reducing end). A higher ratio of HS to HGF/SF is needed to illicit the same binding affinity as that identified in the tetrasaccharide.

Figure 4.1

1. GlcNS(6S) – IdoUA(2S) – GlcNS(6S)
2. GlcNAc(6S) – IdoUA – GlcNAc(4S)
3. Δ Hex – GlcNS(6S) – IdoUA(2S) – GlcNS
4. GlcNS(6S) – IdoUA(2S) – GlcNS

Figure 4.1 The minimum identified HS oligosaccharides which have been identified to bind to FGF2.

These sequences have been taken from (Deakin *et al.*, 2009) and contain three identified trisaccharides and one tetrasaccharide which bind to HGF/SF.

Wu *et al* (Wu *et al.*, 2003) identified a dp6 oligosaccharide which was involved in full binding and activation with two FGF1 proteins (see p86). The published x-ray crystal structures of three heparin decasaccharides have identified dp10 sized oligosaccharide as the minimum for binding FGF1 and FGF2 respectfully (Figure 4.2); (Digabriele *et al.*, 1998; Pellegrini *et al.*, 2000; Schlessinger *et al.*, 2000).

It must be taken into account that some studies describe the requirements for binding only the FGF molecule. Other studies describe the length of chain required for activation of the complex (with both the FGF and FGFR).

The most well-known of these complexes are the Symmetrical Complex and the Asymmetrical Complex which look at the complex formation between FGF with both HSPG and FGFR's, leading to receptor dimerization (see Introduction, p74); (Pellegrini, *et al.*, 2000; Schlessinger *et al.*, 2000). Here in both studies the minimum sized heparin chain was a dp10 sized oligosaccharide chain.

Figure 4.2

A. IdoUA(2S) – GlcNS(6S) – IdoUA(2S) – GlcNS(6S) – IdoUA(2S) – GlcNS(6S) – IdoUA(2S) – GlcNS(6S) –
IdoUA(2S) – GlcNS(6S)

B. ΔUA(2S) – GlcNS(6S) – IdoUA – GlcNS(6S) – IdoUA – GlcNS(6S) – IdoUA(2S) – GlcNS(6S) – IdoUA(2S) –
GlcNS(6S)

Figure 4.2 Two HS oligosaccharides which have been identified to interact with FGF1 and FGF2 respectively.

Panel A. Taken from the PDB (file 1E00) and is known as the Asymmetrical Complex (Digabriele *et al.*, 1998; Pellegrini *et al.*, 2000).

Panel B. Taken from the PDB (file 1FQ9) and is known as the Symmetrical Complex (Schlessinger *et al.*, 2000).

Although it has been widely accepted that the smaller sized HS chains can bind to FGF2 (Guglieri *et al.*, 2008), much larger chains are thought to be required for the activation of the protein-ligand complex (Digabriele *et al.*, 1998; Goodger *et al.*, 2008; Pellegrini *et al.*, 2000; Schlessinger *et al.*, 2000). Stability of the HS chain is due to both the amount and positioning of the sulphate groups as well as from the carboxyl groups (Bitomsky and Wade, 1999; Hricovini and Bízik, 2007) with the surrounding solvent/solution having an influence on the configuration of internal residues, due to the change in pH of the solution (Sanderson *et al.*, 1987).

4.4.2 The Effect of Temperature and Cations on the HS Chain

Temperature (Sanderson *et al.* 1987) and the presence of certain cations have also been shown to have an effect on the solution conformational dynamics of internal residues,; specifically within DS (Sanderson *et al.* 1987); heparan sulphate (Coombe and Kett, 2005); and also heparin (Rudd *et al.*, 2007). The addition of certain cations has an influence on the overall ring conformation of the iduronates, with the presence of Zn²⁺ ions showing a preference for the ¹C₄ conformation due to the nature of the monovalent cations having an effect on the radii and coordination geometry with non-charged parts of the molecule. Sulphate groups have the ability to neutralise one of the charges of a divalent cation; which has an effect on orientation and conformation of neighbouring residues and intervening glycosidic linkages (Rudd *et al.*, 2007).

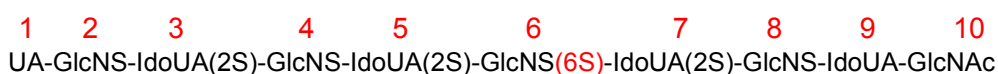
4.5 The Implications for Iduronate Conformation and Sulphation

Patterns within HS chains

This study identifies the conformational equilibrium of the iduronate residues in the FGF2 activatory and inhibitory deca-saccharides sequenced by Goodger (Goodger 2003) as one of the major influences in biological activity. Initial findings suggest the iduronates would be identified in predominately different conformations. Additional modifications; in the form of sulphate groups around the iduronates (in the neighbouring glucosamine residues) has been shown to have an

effect on these conformations. The presence of 6-O-sulphate on flanking glucosamines around the iduronates, have some influence on iduronate conformation in this study (see Linkage 5- 8, p170 onwards). This area of research will play a major role in biological activity in the Goodger oligosaccharides and may well show the differences between the active and inactive oligosaccharides.

The positioning of 6-O-sulphate groups has been suggested as a requirement for promoting FGF1 and FGF2 mitogenic activity (Pye *et al.*, 2000; Kreuger *et al.*, 2001; Wu *et al.*, 2003). This modification is almost certainly under tight control and is thought to be highly regulated; regardless of whether the sulphate group is located centrally in the chain or positioned at the reducing terminal (Merry *et al.*, 1999; Goodger 2003). In the following sequence conformational flexibility is thought to be influenced by the central 6-O-sulphate group:



(Merry *et al.*, 1999)

The central residues are similar in sequence to the Goodger oligosaccharides used in this study. In linkage 8 for the 1 ns MD simulation the 6-O-sulphate group had only a slight influence on only the torsional geometry between GlcNS(6S)-IdoUA (residues 8 and 9 in dp10(1) and dp10(5)); where as it had a larger impact on the torsional geometry for the same decasaccharides in linkage 7. As for linking it to purely conformational changes, this is more complex. Indeed there are changes in iduronate conformation observed in the Goodger oligosaccharides when comparing active and inactive oligosaccharides. However it can not be said if this conformational transition is due to the presence or absence of the 6-O-Sulphate groups present in the HS chain.

The iduronate in position 3 in the oligosaccharide mentioned above (Merry *et al.*, 1999) is found predominately in the 1C_4 conformation (approx 75%). Those at positions 5 and 7 are thought to be

more in equilibrium between two conformations; (the 1C_4 and 2S_0). This sulphate positioning will have some effect on the conformational equilibria; both proceeding and preceding the iduronate residues. This also has an influence on the torsional geometry; which has been observed in this study (Figures 3.17 – 3.21). This kind of modification has been shown to have some effect on the Goodger oligosaccharides, on both the iduronate conformation and the torsional geometries observed. Initial studies carried out by Goodger (Goodger 2003) and continued by Murphy (Murphy 2007); identified the conformations were present in equilibrium and were deduced in the ratio 50/50 (50% 1C_4 , 50% 2S_0). No dominating conformation was determined with no differences shown between active and inactive oligosaccharides (Goodger 2003; Murphy 2007).

The Goodger oligosaccharide sequences, as determined by Goodger (Goodger 2003) along with the deduced conformations in equilibrium (shown in red; Murphy 2007) are identified in Figure 4.3.

Figure 4.3

Active dp10(1)

50:50 50:50 50:50

UA - GlcNAc - GlcUA - GlcNS(6S) - IdoUA(2S) - GlcNS(6S) - IdoUA(2S) - GlcNS(6S) - IdoUA - GlcNAc(6S)

Active dp10(3)

50:50 50:50 50:50

UA - GlcNAc - GlcUA - GlcNS(6S) - IdoUA(2S) - GlcNS(6S) - IdoUA(2S) - GlcNS(6S) - IdoUA - GlcNAc(6S)

Inactive dp10(5)

50:50 50:50 50:50

UA - GlcNAc - GlcUA - GlcNS - IdoUA(2S) - GlcNS(6S) - IdoUA(2S) - GlcNS(6S) - IdoUA - GlcNAc

Inactive dp10(7)

50:50 50:50 50:50 50:50

UA - GlcNS - IdoUA - GlcNS - IdoUA(2S) - GlcNS(6S) - IdoUA(2S) - GlcNS - IdoUA - GlcNAc(6S)

Figure 4.3 The conformation of the iduronates before undergoing Cremer-Pople analysis.

These conformations were initially deduced by Murphy (Murphy 2007), from the Goodger oligosaccharides (Goodger 2003).

The first two oligosaccharides were shown to be able to illicit a mitogenic response from a FGF2-FGFR1 IIIc receptor complex in a Baf cell mitogenesis assay, whereas the second two oligosaccharides were unable to produce any response (Goodger 2003).

These oligosaccharides above which were studied for biological activity (Goodger 2003); were studied used in both the molecular modelling and docking sections of this study. Here the conserved 2-O-sulphate group (found in residue 5 in all active and inactive oligosaccharides) is present, suggesting other structural features are also required for biological activity.

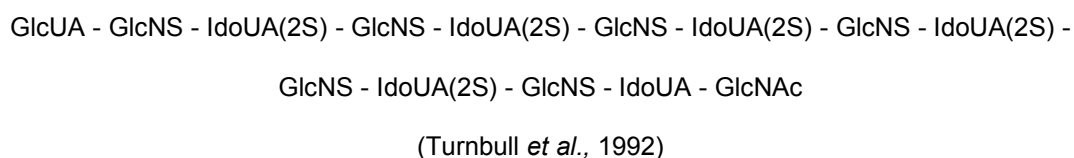
The presence of the 6-O-sulphate group on the reducing end terminal residue could be a determining factor in promoting biological activity. This could be because more 6-O-sulphate is present in the active oligosaccharides than in the inactive oligosaccharides. Also the positioning of this modification in previous reviews (Kreuger *et al.*, 2001; Merry 1999; Wu *et al.*, 2003); suggest roles for this modification when it is placed in the centre of the chain. This is identified in the Goodger oligosaccharides where all the active models contain this additional sulphate group; whereas only two of the inactive groups contain the same modification. This is made even more complex; as in reality dp10(1) and dp10(2) contain all sulphated glucosamines with both 6-O and N-sulphated groups, whereas dp10(3) and dp10(4) contains 6-O-sulphated glucosamines, at position 6 and 10 only. Also dp10(7) and dp10(8) have only one 6-O-sulphate group less than the 'active' dp10(1) and dp10(2); on residue 8 there is an absence of the 6S group. Potentially this additional sulphate group could increase binding in the active oligosaccharides and could knock out the biological activity of what otherwise could be assumed to be an active saccharide.

When the Goodger oligosaccharides are compared to the Schlessinger model (a symmetrical complex of a 2:2:2 FGF2:FGFR1:heparin ternary complex) and the Pellegrini model (an asymmetrical FGF1:FGFR2:heparin complex); there are many sequence similarities. However, the Schlessinger and Pellegrini oligosaccharides are very highly sulphated compared to the Goodger oligosaccharides. Therefore this study indicates a fully sulphated sequence is not essential for binding activity, as both dp10(1) and dp10(3) are not fully sulphated sequences. This

indicates only specific sulphation is required, alongside iduronate flexibility and torsional geometry for biological activity.

The modelling study described in this report suggests specific sulphation is required for biological activity. The 6-O-sulphate group on the reducing terminal of the active Goodger oligosaccharides only a slight influence on torsional geometry, not conformational geometry (data not shown); and so has no major influence on biological activity (Figure 3.17). Torsional differences are also observed in linkage 4 for the 1 ns MD simulation, with reference to the presence and/or absence of the 6S group in the oligosaccharide. Here the presence of the 6-O-sulphate group seems to influence geometry B for the torsion angles observed in the active dp10(1) and dp10(2); (Figure 3.22, Panels A and B). The other active oligosaccharides do10(3) and dp10(4) do not contain the 6S group and are observed in both torsional geometries A and B (Figure 3.22, Panels C and D).

The following studies suggest that the 6-O-sulphate modification is not required in ligand binding. Oligo-H a dp14 oligosaccharide which binds to FGF2, was isolated from human skin fibroblast HS. It also has high affinity to FGF2 with a sequence of:



It is rich in N-sulphation and 2-O-sulphation, but interestingly and quite distinctly has an absence of 6-O-sulphate (Turnbull *et al.*, 1992; Gallagher and Turnbull, 1992). Unfortunately there was no conformer analysis given with this study, but it is an indication of just how complex biological activity is.

Ishihara *et al* (Ishihara *et al.*, 1994) also identified octasaccharides and hexasaccharides rich in IdoUA(2S) - GlcNS(6S) with a high affinity for FGF2. Their study could not determine the requirement for 6-O-sulphate, in particular with the GlcNS residues. Their previous studies

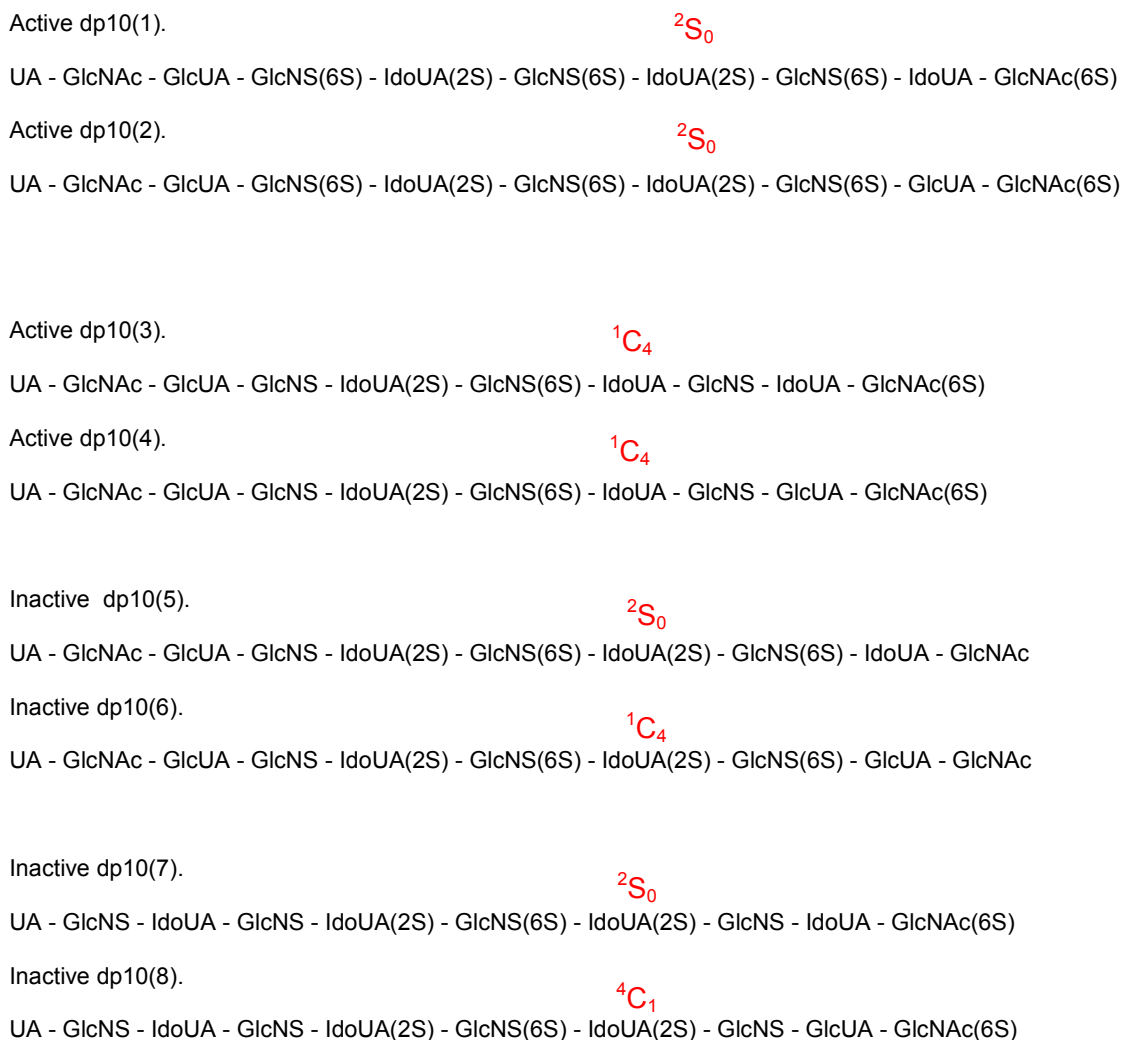
indicated an increase in charge on the heparin correlated to an increase in binding for immobilised FGF2. Their study concluded that the 6S group was a neutral structure which is modified for FGF2 recognition. It is the 2S and carboxyl groups which are required for mitogenic activity of FGF2. Here they were unable to tie in activity with 6-O-sulphation, but the results shown here show this specific modification has an overall effect on torsional geometry. This in turn influences biological activity.

4.5.1 Linkage 7 - Conformations

The internal environment created by the surrounding glucosamines residues, influences the neighbouring iduronate residues. This leads to oscillations in iduronate conformation between the three known conformations (1C_4 , 2S_0 and 4C_1). These changes in environment allow for the different equi-energetic conformations to exist (Coombe, 2005) with changes from the chair to the skew boat form being slower than the overall molecular re-orientation of the structure (Mulloy *et al.*, 2000). The energy barrier between the three forms present in the iduronates is high with a value of 9 Kcal/mol (Ferro *et al.*, 1990). This changing of conformations enables an optimal binding fit for ionic bonding and van der Waals contacts, therefore finding the best fit is imperative.

In our data, the transition between the different conformations is clear. It shows the iduronates in the Goodger oligosaccharides undergo these changes looking for the optimal binding conformation. Some conformational transitions occur fairly rapidly (Figure 3.26, Panel A); whereas some changes in conformation occur over a longer timescale (Figure 3.28, Panel D).

The iduronates which make up linkage 7 are found in the following conformation at the end of the 1 ns MD simulation (Figure 4.4; shown in red and taken from data collected from Figure 3.19):

Figure 4.4**Figure 4.4 The conformation of the iduronates at linkage 7 after the MD simulation.**

The conformations are specifically highlighted in red with an obvious change observed from the starting conformation.

This study suggests that biological activity is not just influenced by iduronate conformation alone. If this was the case then the data above would be very conflicting. Torsional geometry also has to be taken into account and it is this information combined which actually gives more of an indication as to why oligosaccharides dp10(1-4) are active and dp10(5-8) are inactive. The biggest difference between active and inactive oligosaccharides was observed quite clearly in the torsional geometry. The active oligosaccharide dp10(1) was identified in both torsional geometries A and B and this change occurred quite significantly after the conformational change. So far, clear differences have been observed in the Goodger oligosaccharides within torsional geometry. For the first time here, it is the change in iduronate conformation which seems to drive that change in torsional geometry. The inactive dp10(5) undergoes similar changes, but quite significantly this change is observed before the conformational transition (Figure 3.18, Panels A and E). Both oligosaccharides dp10(1) and dp10(5) are sulphated in the same position (IdoUA(2S) - GlcNS(6S)); and so on this occasion sulphation position is shown to have no influence on biological activity of these HS samples.

4.5.2 Linkage 5 - Conformations

A comprehensive analysis of the literature suggests it is the presence of the 1C_4 conformation and/or the presence of the 2S_0 conformation, when considering HS: FGF binding (Pellegrini *et al.*, 2000; van-Boeckel *et al.*, 1987).

It is possible that it is the presence of both conformations in a polymeric HS/heparin chain which is favoured for activity as shown in the literature (Faham *et al.*, 1996; Chuang *et al.*, 2000). For the first time in this study, the results seem to suggest and indeed agree with this statement above. According to the 1 ns MD simulation it can be suggested that it is indeed the presence of both the 1C_4 conformation, alongside the 2S_0 conformation which plays an important role in biological activity. Here we are suggesting both conformations need to be present, in two or more iduronates in any one HS chain to allow for active binding to FGF2. This can clearly be seen in

three out of the four active oligosaccharides. The presence of both conformations alongside the torsional geometry established by the corresponding glycosidic bond, allows for activation and mitogenic response from a FGF2-FGFR1 IIIc receptor complex in a Baf cell mitogenesis assay. For oligosaccharide dp10(1 and 2), the 1C_4 conformation is observed in the iduronate at position 5, with the skew boat conformation observed for the iduronate at position 7. The conformation position is then switched around for dp10(4), with the skew boat conformation observed in residue 5, followed by the 1C_4 conformation. The exception to the rule here is dp10(3), which is initially identified in the 4C_1 conformation for residue 5 and the 1C_4 conformation for the iduronate at position 7 in the HS chain. Excluding dp10(3), the three other active oligosaccharides after 1 ns of MD simulation could show valuable information concerning biological activity.

The iduronate residues which make up linkage 5 and 7 are found in the following conformations at the end of the 1 ns MD simulation (Figure 4.5; shown in red and taken from Figure 3.19 and Figure 3.21):

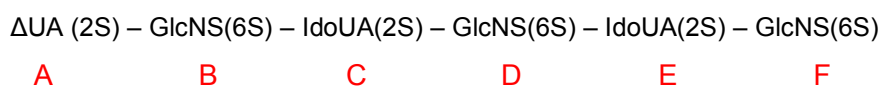
The inactive oligosaccharides are observed in different conformations compared to the active oligosaccharides and this could be a significant indication of biological activity in these samples. Here at the end of the MD simulation the inactive oligosaccharide iduronates at position 5 and 7 in the chain follow a different pattern to that observed in the active oligosaccharides. For dp10(5), both are confirmed in the skew boat conformation. The same conformational data is observed in inactive dp10(7). Decasaccharide dp10(6) is observed in on the 1C_4 conformation, even though it has the same sequence as dp10(5). Finally dp10(8), different again from the other three inactive oligosaccharides shows the iduronate at position 5 is present in the 1C_4 conformation, alongside this is the iduronate at position 7 which is observed in the rigid 4C_1 conformation.

4.5.3 The Binding According to Literature

The following section will provide an overview of information regarding the conformation of iduronates in FGF interactions and also other proteins (EDF/SF, histamine, AT III and O₁BFS) and how they relate to the information obtained in this study from the Goodger oligosaccharides.

4.5.3.1 The Proposal for Iduronic Acid 1C_4 Conformation: Protein Interactions

In histamine-heparin interactions the population of the 1C_4 conformation increased when in the bound form (when the heparin oligosaccharide was bound to histamine). The binding of histamine to the heparin hexasaccharide did initially identify a role for both conformations 1C_4 and 2S_0 (Rabenstein, 2002). However the 1C_4 conformation was identified as the dominating conformation. The bound oligosaccharide sequence was identified from literature as:



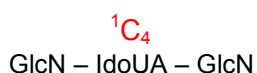
The conformation of the iduronate at residue E was identified with and without histamine bound, and identified a role for both conformations 1C_4 and 2S_0 conformations.

Alongside the conformational changes reported in the literature, were also minor changes in the glycosidic linkage when in the bound form. The final linkage between the IdoUA(2S)-GlcNS(6S); residue E-F changes from $\Phi = 43^\circ$ $\Psi = 12^\circ$ in the unbound form to $\Phi = 49^\circ$ $\Psi = 22^\circ$ when bound. This indicates a decrease in the skew boat conformation when bound to histamine. This same process could be observed in the Goodger deca-saccharides, if the Cremer-Pople puckering parameters were analysed with and without binding FGF2. We have already identified changes in glycosidic geometry in our modelling data (See Results, p172 onwards), which suggests the torsional geometry may change further when bound to FGF2.

One of the earliest studies on iduronate conformation was carried out by van-Boeckel *et al.* (van-Boeckel *et al.*, 1987) on the conformational analysis of synthetic heparin-like di and trisaccharides. The disaccharides indicated equilibrium between all three conformations, with the presence of the 2-O-sulphate group influencing the population of the 1C_4 conformation. When increased to trisaccharides, the skew boat and 1C_4 chair conformation dominated and when increasing the temperature the population of the skew boat conformation was found to increase further (Sanderson *et al.*, 1987; van-Boeckel *et al.*, 1987). Although our study did not look at the effect of temperature, it did show an equilibrium between the two most popular conformations; both 1C_4 and 2S_0 conformations. However, it was not determined if there was in fact one conformation which dominated more than the other. Also identified was a slight influence on certain internal iduronate residues from the 4C_1 conformation (Figure 3.19, Panel H; Figure 3.21, Panel C; Figure 3.28, Panel B).

Studies on a heparin polysaccharides indicate the 1C_4 conformation as the active conformation in which all the sulphate groups are presented on one side of the chain (Mulloy *et al.*, 1993). This sulphation pattern has been identified in these molecular modelling studies, where the sulphate groups are presented in patterns on one side of the HS chain.

Trisaccharide analysis by Raman *et al* (Raman *et al.*, 2005) suggests the 1C_4 conformation dominates due to the position of the glycosidic bond in the axial position. This had been identified in the following sequence:



All of the Goodger molecular modelling sequences used in this study were set in the axial position, at the start of the MD simulation. It is interesting to see that in the structure above they have identified the 1C_4 conformation as the dominating conformation when the bond is in the axial position. This is the only study in which the axial conformation is mentioned in literature, in relation to torsional geometry.

Work carried out by Ferro *et al* (Ferro *et al.*, 1986) on monosaccharides, disaccharides and pentasaccharides, again suggests the 1C_4 conformation dominates over the skew boat form (Figure 4.6). This study suggested monomers of IdoUA(2S) were dominated by the 1C_4 conformation, but only when found as a single residue.

The work carried out on IdoUA(2S) - α Me glycoside; (which was the structure used for the identification of iduronate Cremer-Pople puckering parameter dimensions, Figure 3.15) and was an extension of the work by Murphy (Murphy 2007) identified slight domination from the skew boat ring conformation. Duplicate MD simulations were carried out to identify the correct puckering parameter data values for each conformation. It was determined that those models with a starting conformation of 1C_4 made a conformational transition to the 2S_0 conformation during the MD simulation (1000 ps MD simulation). However further investigation carried in this study indicated a further conformational transition was observed in some MD simulations, resulting in a change to the 4C_1 conformation (Figure 3.15 and Figure 3.16).

Figure 4.6

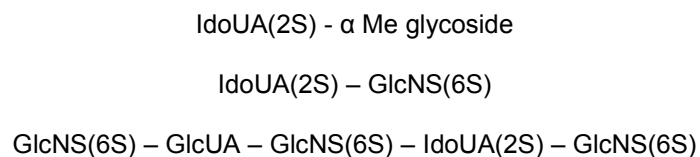


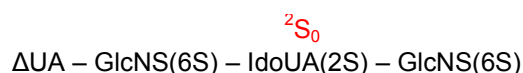
Figure 4.6 The monosaccharide, disaccharide and pentasaccharide in which the 1C_4 conformation was thought to dominate.

This has been taken from (Ferro *et al.*, 1986) with only the iduronate residue studied.

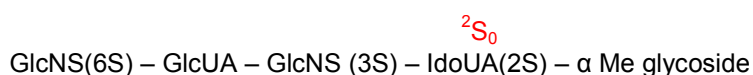
4.5.3.2 The Proposal for 2S_0 :

The skew boat conformation has not been identified in literature with reference to FGF2 interactions. It is also most commonly studied with the related proteoglycan heparin, rather than with HS. A study involving modified heparins, all of which had a high proportion of the 2S_0 conformation were identified (Mulloy *et al.*, 1994). These included native heparin, and other structurally modified heparins such as N-desulphated, re-acetylated heparin, N and O-desulphated, re-N-acetylated heparin, N- and O-desulphated, re-sulphated heparin and N- and 6-O-desulphated, re-N-sulphated heparin. All were shown to have a high proportion of the skew boat conformation amongst the iduronates.

A heparin derived tetrasaccharide containing only one iduronate was identified in the 2S_0 conformation. This was thought to bring all the substituents into a preferred equatorial position (the exo-anomeric effect) and has a sequence of (Mikhailov *et al.*, 1996):



The discovery of the skew boat conformation indicated how important it is in binding when it was identified in the AT III binding sequence (Ferro *et al.*, 1990). Here it had been identified as an important feature in the HS chain as it facilitates electrostatic interactions but is not directly involved in the binding (Coombe and Kett, 2005; Guerrini *et al.*, 2006). It is thought to dominate the equilibrium between the two conformations because the 1C_4 conformation is too bulky. Hricovini *et al.* (Hricovini *et al.*, 2001) studied a synthetic heparin pentasaccharide (the synthetic version of the AT III binding pentasaccharide). It contained the following sequence where again the skew boat conformation was identified:



Das *et al* (Das *et al.*, 2001) also compared a synthetic pentasaccharide differing in its structure due to the presence of a methylated section of the chain (which binds to AT III). Again the iduronate was identified in the 'favourable' 2S_0 form.

The data from both the 1 ns and 10 ns MD simulations shows a role for the skew boat conformation. It is not clear if this is the dominating conformation in our results, but it certainly plays a role in biological activity, alongside the changes in torsional geometry. The skew boat was identified in the majority of the active oligosaccharides but this was alongside the 1C_4 conformation, suggesting a role for both conformations.

4.5.3.3 The Importance of both 1C_4 and 2S_0 Conformations: The FGF Interactions

The most studied and well known heparin structure to date is the work on a heparin hexasaccharide bound to FGF2 with the work resulting in an x-ray crystal structure (file 1BFC from the RCSB Protein Data Bank; Faham *et al.*, 1996). This work is directly comparable to the work carried out in this study as it involved the binding of a heparin hexasaccharide and tetrasaccharide specifically to FGF2 (bFGF). This was also the x-ray crystal structure used in the docking section of this thesis. Here the different dp10 Goodger oligosaccharide structures were docked into the FGF2 protein molecule with the use of AutoDock Vina.

The heparin hexasaccharide and the tetrasaccharide are made up of the same sequence as each other (below) with the tetrasaccharide reducing sugar assigned as ring number 4 (Figure 4.7, Panel A). The high affinity binding site involved rings 2 and 3 and therefore is found in both oligosaccharides. The additional low affinity binding site is identified as rings 5 and 6; therefore the tetrasaccharide will only contain this low affinity binding region. The conformation of the iduronate in the high affinity binding region in the Faham structure has been identified in the 1C_4 conformation. The iduronate in the low affinity binding region again of the Faham structure was identified in the 2S_0 conformation. It is therefore presumed that it the presence of both

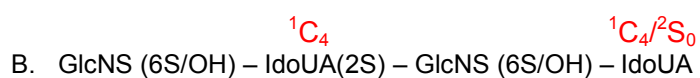
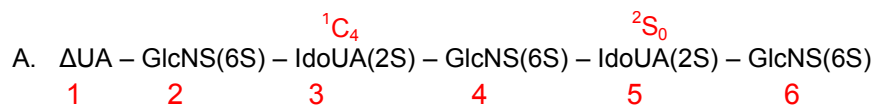
conformations in the hexasaccharide structure which allows for binding of the heparin chain to FGF2.

Residues 4 through to 7 in the active Goodger models; dp10(1) and dp10(3) contain the FGF2 binding motif identified by Guglieri *et al* (Guglieri *et al.*, 2008) which had been identified by NMR spectroscopy. Here both conformations are present when bound to both FGF1 and FGF2 (Figure 4.7, Panel B).

The Faham and Guglieri structures show similarities with the Goodger decasaccharide models built during this study. The IdoUA(2S) in position 5 in all of the Goodger decasaccharide models is identified in the 1C_4 conformation like those models presented (Faham *et al.* 1996; Guglieri *et al.*, 2008); with the second iduronate following on from this in the 2S_0 conformation. For the Goodger oligosaccharides, active dp10(1) and dp10(3) this particular iduronate happens to be positioned in the centre of the HS chain and also happen to be the first iduronate you come across from the non-reducing end of the chain.

Both conformations are present in the active Goodger oligosaccharides, but we have also seen both conformations present in the inactive oligosaccharides; just in varying ratios and presented at different times in the MD simulations. Our results do suggest that both conformations are necessary for biological activity, but this alone does not ultimately decide it. Other factors need to be taken into consideration, such as torsional geometry.

Figure 4.7



(Guglieri, Hricovini *et al.*, 2008)

Figure 4.7 Two HS/heparin oligosaccharides which indicate the requirement for both the $^1\text{C}_4$ and the $^2\text{S}_0$ conformations present for binding.

Panel A. Shows a heparin hexasaccharide and a tetrasaccharide of the same sequence with both conformations present (Faham *et al.*, 1996).

Panel B. The binding motif as identified by Guglieri *et al* (Guglieri *et al.*, 2008) also with both conformations indicated as a requirement for biological activity.

4.5.3.4 The Importance of both 1C_4 and 2S_0 Conformations: Other Protein Interactions

The previously mentioned foot and mouth disease protein (O₁BFS; Fry *et al.*, 1999) behaves in a similar way to FGF2 as identified by Faham *et al.* (Faham *et al.*, 1996). It selects and binds to a fully sulphated heparin of which the three central rings are made up of:



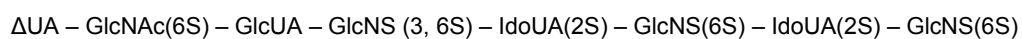
The glucosamine residues as well as the terminal IdoUA adopt chair conformations. The central iduronate appears to adopt the skew boat conformation but also the 1C_4 chair conformation (Fry *et al.*, 1999).

Two heparin octasaccharides Octa A and Octa B were isolated from Lovenox (enoxaparins) and studied. These are low molecular weight heparins obtained by β -eliminative cleavage of a porcine mucosal heparin (Guerrini *et al.*, 2006) and have similar sequences to the Goodger oligosaccharides. Here Octa A and Octa B contain the AT-binding properties previously identified (Hricovini *et al.*, 2001) alongside the rare 3-O-sulphate modification (Figure 4.8). NMR studies suggest that in a free state both 1C_4 and 2S_0 conformations are present in Octa A with the equilibrium driven towards the 2S_0 conformation when bound to antithrombin III (Figure 4.8). For Octa B, the IdoUA(2S) in the AT-binding sequence contains both conformations with the non-sulphated iduronic acid preceding this sequence identified only in the 1C_4 conformation. Guerrini *et al.* (Guerrini *et al.*, 2006) suggests that when in a bound state the conformation of IdoUA(2S) is driven towards the 2S_0 conformation; shown when the oligosaccharide was in solution and the equilibrium was found with ratios (%) towards the 2S_0 conformation: 60: 40 (2S_0 : 1C_4). (This was identified specifically in the AT pentasaccharide binding sequence). The 2S_0 conformation is thought to form part of a high affinity binding site in the heparin chain (Raman *et al.*, 2005) with the low affinity site further down the chain.

Binding studies were not carried out during this study, so it is difficult to compare this data from literature to the modelling data obtained here in this study. It has been identified in this study that a presence of both conformations in the active oligosaccharides which suggest both needs to be present for biological activity. However, it is difficult to compare bound and unbound states as described above. Certainly there is a need for more than one conformation to be present which has been identified in the Goodger oligosaccharides dp10(1, 2 and 4). Oligosaccharide dp10(3) is the exception with both chair forms present at the end of the 1 ns MD simulation, with no sign of the skew boat conformation. This is not observed in the inactive oligosaccharides, where it is the presence of only one conformation after the MD simulation (the 2S_0 conformation in dp10(5) and dp10(7); and the 1C_4 chair conformation for dp10(6)). The exception to this is dp10(8); which shows a similar pattern to dp10(3) with both chair conformations present.

Figure 4.8

Octa A:




AT-binding sequence

Octa B:



Figure 4.8 The two heparin octasaccharides; which contain the rare 3-O-sulphate modification and have been identified in both the ${}^1\text{C}_4$ and the ${}^2\text{S}_0$ conformations.

Octa A and Octa B have both been isolated from Lovenex and are produced by a β -eliminative cleavage of a porcine mucosal heparin (Guerrini *et al.*, 2006).

4.5.3.5 Biological Implications

The information above gives an indication to the complexity of biological activity in HS chains. All views have been put across with data that backs up each argument. According to the data obtained through the MD simulations it could be the presence of both conformations in the chain which leads to a biologically active HS chain. This subtle difference in conformation of the iduronates in the internal environment of the chain would allow for the protein ligand to selectively fit into the conformation it requires. The HS chain in turn would offer both a low and a high affinity binding site with both conformations present. The presence of both conformations could aid in the deviation or kink identified previously (Guglieri *et al.*, 2008). As more conformational transitions occur the effects from the presence of the kink could become more apparent and/or dramatic; therefore having an overall influence on the biological activity of the HS chain.

It can be identified that the conformation of the iduronates plays a major role in biological activity. It is not this alone however, which dictates biological activity. The torsional geometry data can not be ignored as there is a clear indication in this data that this has a major effect on biological activity in the active oligosaccharides. This has been identified in dp10(1) and dp10(3), where the presence or absence of a particular geometry, alongside specific iduronate conformation which seems to have an influence on this activity.

4.7 The Contribution from the 4C_1 Conformation

The presence of the 4C_1 conformation in the 1 ns MD simulation for the iduronate at sequence position 7 in dp10(8), (Figure 3.19, Panel H) and also residue 5 in dp10(3); is the first time this conformation has been identified in any of the Goodger oligosaccharides, for an internal residue. This has previously only been identified in iduronates at terminal ends. There is work in the current literature which suggests the presence of the 4C_1 conformation, for internal structures. This was identified in work specifically with dermatan sulphate (not heparan sulphate). Inoue *et al* (Inoue *et al.*, 1990) looked at conformer populations in a DS tetrasaccharides (Figure 4.9).

The internal IdoUA for sequence 1 has equilibrium of; 54: 42-44: 4 (1C_4 : 2S_0 : 4C_1) with sequence 2 having a ratio of 54: 44: 2 (1C_4 : 2S_0 : 4C_1). Conformer populations were also described in this way by work also carried out on DS by Casu *et al* (Casu *et al.*, 1986). The non-reducing IdoUA in the hexasaccharide (Sequence 3) is in equilibrium with equal populations of all three conformers: 30: 40: 30 (1C_4 : 2S_0 : 4C_1).

Although these are not directly comparable because the work has been carried out on DS, which contains galactosamine in replacement of glucosamine found in HS; it is worthy to note the 4C_1 conformation has been identified as an internal conformation elsewhere. This proves the ability of the iduronates to undergo conformational transitions in order to produce optimal binding conditions which allow for a good structural fit for ionic and van der Waals interaction (Raman *et al.*, 2005).

It is not clear however if this aids in biological activity in the Goodger oligosaccharides as this was identified at residue 7 in an inactive oligosaccharide dp10(8); but also in the active dp10(3) oligosaccharide in residue 5. Because of this it is not clear if this aids in biological activity in these HS samples.

Figure 4.9

Sequence 1: I – H – I – H
Sequence 2: G – H – I – H
Sequence 3: I – H – I – H – I – H

Where I – IdoUA

G – GlcUA

H – GalNAc

Figure 4.9 Conformational data for iduronate internal structures, identified in dermatan sulphate.

The 4C_1 conformation for the first time in literature has been identified for an internal structure iduronate, according to Inoue *et al* (Inoue *et al.*, 1990).

4.8 The Potential Role for Terminal Δ UA Residues

The presence of the Δ UA at the non-reducing end of an oligosaccharide as previously mentioned is a result of the enzymatic degradation process. Previous studies have all but ignored the presence of this residue in protein binding due to its unnatural presence in the chain. A key role has been established for this in the work carried by Schlessinger *et al* (Schlessinger *et al.*, 2000); identified in their x-ray crystal structure of an FGF2:FGFR1:heparin complex. Here the terminal UA residues on both heparin chains are identified in the centre of the canyon, which runs between the two receptor molecules (see introduction, page 78). Here their terminal unit was the IdoUA(2S) and so it would not be directly comparable to the Goodger oligosaccharides. It is difficult to say if this has any effect on biological activity in the Goodger oligosaccharides as there are no conformational or torsional differences between the active and inactive oligosaccharides in either simulation (1ns or 10 ns MD simulation).

4.9 The 3-Dimensional Structure of Heparin and the Influence it has on the Binding Site

There have been many NMR studies on polymeric heparin oligosaccharides and this is due to the relative ease of the purification process for purifying heparin compared to HS (Mulloy *et al.*, 1993; Mulloy *et al.*, 1994; Mikhailov *et al.*, 1996; Mikhailov *et al.*, 1997; Hricovini *et al.*, 2001; Angulo *et al.*, 2005). The purification process for HS is more complex making it more difficult to achieve the levels needed for any analytical process, but especially NMR molecular refinement. The longer oligosaccharides are more difficult to purify because there is not as much produced during the enzyme degradation of the chain (when compared to the shorter oligosaccharides dp2-dp6). The use of molecular modelling, alongside NMR structural refined is important in determining molecular models with accurate structures. Those identified using NMR spectroscopy are an accurate reflection of those identified in nature, whereas molecular models could predict

conformations and structures which are not naturally found in nature. Comparing molecular models with different ring conformations is an important process in determining how these affect the overall 3-dimensional structure. Two 3-D models have previously been produced for a heparin dodecasaccharide (using NMR structural refinement); which identifies the subtle differences between the different iduronate conformations, specifically the 1C_4 and 2S_0 forms (Mulloy *et al.*, 1993). These subtle changes in the overall 3-dimensional structure can be observed in the following figure (Figure 4.10). Here one molecular model has all the iduronates set in the chair conformation with the second of the two models in the skew boat conformation. These differences; in only the conformation of the iduronic acid ring structure, show that when in the 1C_4 conformation the positioning of the iduronic sulphate groups means they are closer together spatially. When converted to the 2S_0 conformation the same sulphate groups are extended further away into the molecular space surrounding the molecule.

It is these subtle changes in the structure (and they are very subtle changes in the overall 3-dimensional shape); which leads to an overall change in the specific way these groups interact with protein ligands (Mulloy and Forster, 2000).

Figure 4.10

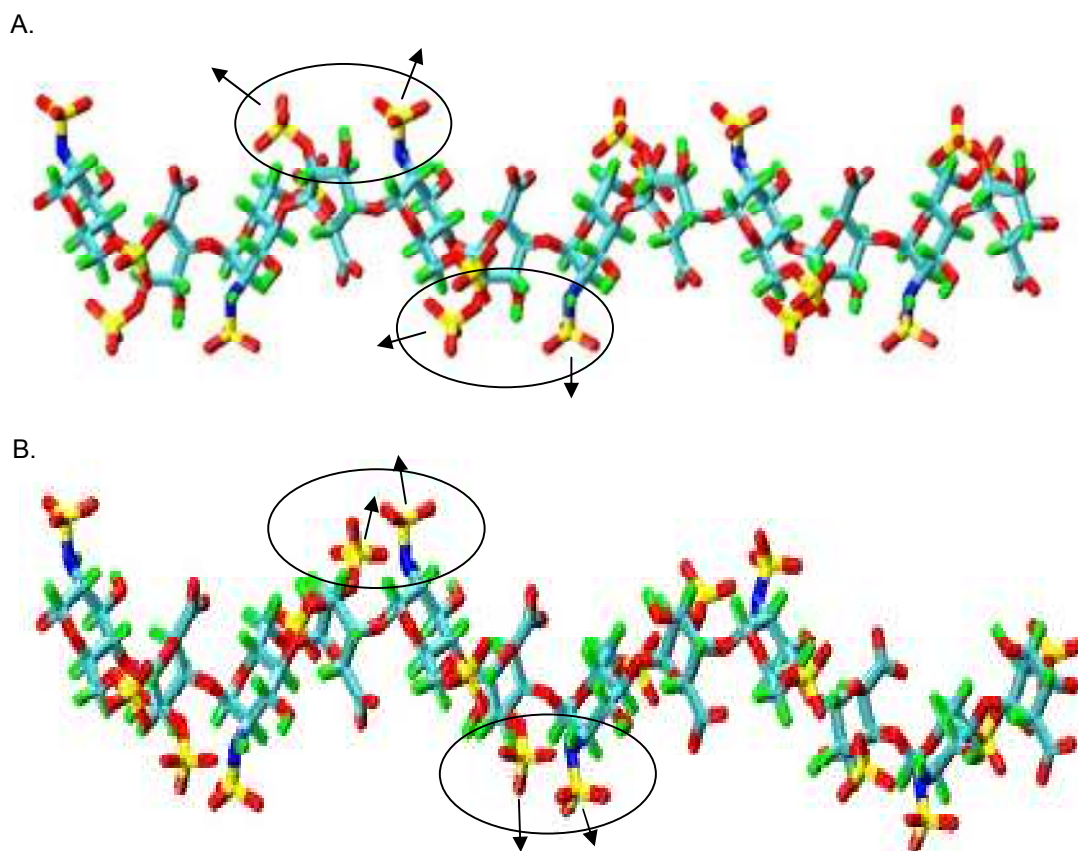


Figure 4. 10 Three dimensional models of two heparin dodecasaccharides.

The coordinates were taken from the Protein Data Base, file 1HPN (Mulloy *et al.*, 1993). They each consist of the same sequence of 6 disaccharide repeats of GlcNS(6S) - IdoUA(2S).

Panel A. All of the IdoUA(2S) units set in the 2S_0 conformation.

Panel B. All of the IdoUA(2S) units set in the 1C_4 conformation.

4.10 Molecular Dynamic Studies on Glycosidic Bonds

The construction of molecular models on the basis of NMR structural data (obtained from NOE interactions) represents only average structures, when in reality a molecule in solution would sample many alternate conformations and 3-dimensional shapes. The important thing to note about these models is that they are able to sample these different geometries as shown in the results section. Most structures actually remained close to their starting geometry as modelled at the start of the simulation (an example of the data values for dp10(1) is shown in Table 4.1). The next step in this process if there were no time constraints would be to obtain NMR structural refinement data for the Goodger oligosaccharides to check that these identified conformations predicted by the MD simulations exist in nature.

All molecular dynamic simulations carried out as part of this thesis work were deliberately set up in a way so that only influences from immediate residues were taken into account. Only the influences from neighbouring residues would have an effect on the ring conformation and the torsional geometry of the glycosidic bonds. This was achieved by setting the influence of charges to a cut off point of 6 Å, therefore no electrostatic interactions from residues further down the chain would have any influence on the chain (see Minimisation 1 in Experimental Materials and Procedures). This would mean only the immediate environment of any residue (the connecting glycosidic bond and the neighbouring residues only) is only influenced by its neighbours. For example, residue 5 in all the oligosaccharides would only be influenced by residue 4 and 6, and the connecting bonds. There would be no further influence any further upstream or downstream from this position.

Table 4.1

Linkage and Monosaccharide Conformation	Dihedral angles set at the star of the MD simulation		Dihedral angles identified at the end of the MD simulation	
	ϕ°	ψ°	ϕ°	ψ°
UA 1H_2 – GlcNAc 4C_1	55	5	-20 30	-20 10
GlcNAc 4C_1 – GlcUA 4C_1	-35	-20	-30	10
GlcUA 4C_1 – GlcNAc(6S) 4C_1	45	15	40	0
GlcNS(6S) 4C_1 – IdoUA(2S) 1C_4	-50	0	-30	-30
IdoUA(2S) 1C_4 – GlcNS(6S) 4C_1	-50	-10	-30 40	-20 10
GlcNS(6S) 4C_1 – IdoUA(2S) 1C_4	-50	0	-40 30	-20 30
IdoUA(2S) 1C_4 – GlcNS(6S) 4C_1	-50	-10	-30 40	-30 -10
GlcNS(6S) 4C_1 – IdoUA 1C_4	-75	-25	-30	-30
IdoUA 1C_4 – GlcNAc(6S) 4C_1	-30	-10	-20 30 -20	-30 10 -20

Table 4.1 A comparison of the dihedral angles for dp10(1) set at the start of the MD simulation and those observed as an average value at the end of the MD simulation.

For those with more than one identified low energy minima, the negative values were used to build the Goodger models. Here the Δ UA for example, was initially set at $\phi = +55^\circ$ $\psi = +5^\circ$. At the end of the MD simulation this geometry was identified as being $\phi = -20^\circ$ $\psi = -20^\circ$ and identified as geometry A and $\phi = +30^\circ$ $\psi = +10^\circ$, identified as geometry A.

Torsional geometries in this study have been identified to play a major part in biological activity in the HS samples. The results presented here indicate an overall influence of the ring conformation of the iduronates on the torsional geometry of the glycosidic bonds. The results have identified that the presence of either the 1C_4 and/or the 2S_0 conformation has an influence on the neighbouring torsional geometry. In linkage 7, of the 1 ns MD simulation the chair conformation seems to influence geometry B which is the negative geometry identified. The skew boat conformation has the opposite influence, on geometry A (the positive geometry) also in linkage 7 (Figure 3.19). Geometry A and the skew boat conformation are identified in the active oligosaccharides dp10(1 and 2); whereas geometry B and the chair conformation is identified mostly in the inactive oligosaccharides dp10(5 and 6). This is identified in both structures with the 2S group attached to the iduronate residue. In those residues without this modification to the HS chain, the results vary from that described above.

Torsional geometries are also identified as a difference between active and inactive oligosaccharides in linkage 6 (involving the GlcNS(6S) identified in all eight oligosaccharides). An additional geometry C was identified in dp10(3 and 4); which had not been previously identified. This was thought to be due to the lack of 2-O-sulphate on the neighbouring iduronate residue. Linkage 5 also shows considerable differences between active and inactive oligosaccharides specifically to do with torsional geometry. Of the active oligosaccharides three out of the four exist in two geometries (A and B). The inactive oligosaccharides exist in only one torsional geometry, which is either geometry A or B.

It does seem to indicate that these subtle changes in torsional geometry actually play a major role in biological activity. These changes go alongside the conformational ring transitions observed throughout both MD simulations. It is fairly accurate to predict that biological activity is most definitely defined by iduronate ring conformation and glycosidic linkage torsion geometries. One or the other is not enough on its own to stand outright and be the dominating factor between inactive and active HS oligosaccharides.

4.11 The Biological Implications for the Docking Experiments

The need for a good fit in the docking experiments would bring this research closer in being able to successfully predict HS structures for molecular drug design. This would put this research one step closer to being able to successfully predict the actual requirements for HS-FGF2 binding. This would involve specific iduronate conformation, sulphation pattern, positioning and type of sulphation needed, glycosidic linkage torsional geometries and also chain length. Up until now; a large amount of information has concentrated on the size of the HS chain in being able to illicit a biological response. The following are just a selection (Digabriele *et al.*, 1998; Pellegrini *et al.*, 2000; Schlessinger *et al.*, 2000; Wu *et al.*, 2003; Guglieri *et al.*, 2008; Goodger *et al.*, 2008; Deakin *et al.*, 2009). There have also been many papers on the sulphation requirements and patterns of sulphation required for biological activity. The following are again just a selection (DiGabriele *et al.*, 1998; Faham *et al.*, 1996; Guglieri *et al.*, 2008; Kreuger *et al.*, 2001; Maccarana *et al.*, 1993; Pye *et al.*, 1998; Schlessinger *et al.*, 2000).

The docking studies carried out here took into account not only the size of the HS chain, but also the conformation of the iduronates and the positioning of the sulphate groups, which were obtained from the MD simulations carried out in this study. Important frames in the MD simulations were taken and analysed using AutoDock Vina, against the x-ray crystal structure 1BFC (from the Protein data bank). Each frame was chosen for a specific change in either ring conformation, or torsional geometry. These were chosen because it has been shown that each of these factors has an influence on biological activity, shown in this study. Unfortunately; due to the nature of the docking simulations and the time constraints in this study, only one FGF signalling complex model was studied during the docking experiments. This was file 1BFC from the protein data bank (an FGF2 protein complexed with a hexamer heparin fragment). Other candidates for this docking were file 1FQ9; a 2:2:2 complex of ternary bFGF-FGFR1-heparin complex, however it was decided that file 1BFC was the most similar in terms of this study.

As discovered in the docking studies; the majority of the Vina predicted models were instantly dismissed due to their predicted docking of the Goodger decasaccharide models sulphate groups in comparison to the hexasaccharide of the 1BFC crystal structure. Many of the rigid (these were set rigid at the start of the docking simulation, dockings were identified wrapped around themselves, rather than stretched out across the binding site on the protein. These kinds of predicted structures were dismissed on the basis of not docking into the correct binding site (Faham *et al.*, 1996; DiGabriele *et al.*, 1998; Schlessinger *et al.*, 2000). The non-reducing terminal of the heparin hexasaccharide in the crystal structure has no visible interaction with the binding with FGF2. However, this is not always the case, as a previous study by Schlessinger *et al.* (Schlessinger *et al.*, 2000) identified the non-reducing ends of both HS chains in the crystal structure, in the centre of the binding canyon. The remainder of the molecule was found to interact with a further high affinity binding site further away. An example of this is shown in Figure 1.21 of the Introduction (page 79). This indicates that any docking structures predicted by Vina which involve either the non-reducing or indeed the reducing ends in the binding region must not be dismissed based on this information alone. As identified by Faham *et al.* (Faham *et al.*, 1996); there could be both a high affinity binding site alongside a low affinity binding site. The size of the Goodger oligosaccharides would indicate that the presence of both binding sites could be a distinct possibility.

It would be wrong to dismiss Vina as a useful docking tool on the basis that most, if not all of the flexible dockings and some of the rigid dockings were dismissed due to them being not correctly positioned in the binding site. Ideally, the highest affinity binding structure would be the best for model according to the crystal structure. In this study, only one model fitted in accurately with the binding site and had been ranked model number 1 in the Vina predicted binding structures. This was for dp10(1), frame 45 of the 1 ns MD simulation. Model number 1 had the highest predicted binding affinity with a value of -13.0 Kcal/mol. The preliminary dockings indicate Vina is a reliable piece of docking software due to its ability to successfully bind two of the known x-ray crystal structures into the correct binding site. These files had previously been taken apart, creating

separate protein and ligand files following the step by step process offered by Trott (Trott 2009). Increasing the exhaustiveness setting in Vina would have increased the likelihood of Vina finding the correct binding conformation. During this study it was set as the default 8, as increasing this would lead to considerable longer docking times. This would be something to reconsider during future docking simulations. It must also be noted that the frames chosen for docking were only 1 ps in time as only one frame was docked at any one time. To gain a better insight into the required docking features; many more frames would need to be studied in order to increase the probability of finding the correct conditions needed for binding.

Chapter 5: General Conclusion

This study has demonstrated that molecular modelling data has enabled significant new insights to be gained into the role of iduronate ring conformation and torsional angle geometry. These have both shown to be significant factors and play a major role in biological function with differences identified between inactive and active HS oligosaccharides. The study has shown that glycosidic linkages do not have to be fixed in a certain position or geometry favoured as a result of the exo-anomeric effect. Static and flexible MD simulations have shown that alternate geometries occur in all linkages, allowing for the movement away from its usual 3-dimensional shape. The data indicates a link to possible biological activity with the presence of these alternate geometries; however HS-FGF2 biological activity is not due to this alone. Our data indicates a significant amount of torsional geometry is due to the presence of certain iduronate conformations. Cremer-Pople puckering parameters indicate the role of both the 1C_4 and 2S_0 conformations are essential for biological activity. The ratio of these conformations is still relatively unknown but the data indicates that certain sulphation patterns on flanking residues in the chain are also responsible for biological activity. These sulphate groups induce a 'kink' in the chain again removing the HS chain away from its original 3-dimensional shape.

Definite differences have been identified between the active and inactive Goodger oligosaccharides, which will become more apparent on longer MD simulations. These differences aid in the ability to determine how FGF2 activation occurs in the active dp10(1) oligosaccharide, as opposed to the inactive dp10(5) oligosaccharide.

The docking studies have indicated that AutoDock Vina is a reliable docking software, however further work is necessary to fully explore all of the functions available in this docking software. The results have shown that it has the ability to accurately predict certain x-ray crystal structures, by correctly docking into the known binding site. The active dp10(1) oligosaccharide gave the most promising results in Vina, which was able to bind in with relative accuracy in the correct binding site. This could be improved with more through knowledge and understanding of the docking software.

Finally the experimental work has shown that some of the experimental difficulties can be overcome when working with heterogeneous HS samples. Large dp8 sized oligosaccharides are

incredibly difficult to purify, with the purification process becoming more complex with every oligosaccharide size increase. Valuable structural data can be gathered for authentic HS derived oligosaccharides through NMR molecular refinement. In this study, only a partial structural assignment was obtained and this was due to the highly complex nature of the purification process. Larger amounts of purified HS fragments could be obtained over a longer timescale.

Virtual screening of HS libraries will become invaluable in modern drug discovery. This would allow for suitable targets to be screened with little or no purification of structures required. Being able to accurately predict binding patterns and required structures would allow for faster drug design without the complications or a difficult purification process.

Chapter 6: References

- Ai, X., Do, A., Lozynska, O., Kusche_Gullberg, M., Lindahl, U. and P., E. J. C. (2003). "QSulf1 remodels the 6-O sulfation states of cell surface heparan sulfate proteoglycans to promote Wnt signaling." *J. Cell. Biol* **162**(2): 341-351.
- Ai, X., Kitazawa, T., Do, A., Kusche-Gullberg, M., Labosky, P. A., and Emerson Jr, C. P., (2007). "SULF1 and SULF2 regulate heparan sulfate-mediated GDNF signaling for esophageal innervation." *Development* **134**(18): 3327.
- Almond, A and J.K. Sheehan (2000). "Glycosaminoglycan conformation; do aqueous molecular dynamics solutions agree with x-ray fibre diffraction." *Glycobiology* **10**(3): 329-338.
- Angulo, J., Hricovini, M., Gairi, M., Guerrini, M., de Paz, J. L., Ojeda, R., Martin-Lomas, M. and Nieto, P. M. (2005). "Dynamic properties of biologically active synthetic heparin-like hexasaccharides." *Glycobiology* **15**(10): 1008-15.
- Bazin, H, Capila, G. I., Linhardt, R. J. (1998). "Conformational study of synthetic Δ^4 -uronate monosaccharides and glycosaminoglycan-derived disaccharides." Carbohydrate Research **309**(2): 135-144.
- Bitomsky, W. and R. C. Wade (1999). "Docking of Glycosaminoglycans to Heparin-Binding proteins: Validation for aFGF, bFGF and Antithrombin and Application to IL-8." *Journal of the American Chemical Society* **121**(13): 3004-3013.
- Carey, D. J. (1997). "Syndecans: multifunctional cell-surface co-receptors." *Biochemical Journal* **327**(Pt 1): 1.
- Casu, B., J. Choay, Ferro, D. R., Gatti, G., Jacquinet, J-C., Petitou, M., Provascoli, A., Ragazzi, M., Sinay, P., and Torrini, G. (1986). "Controversial glycosaminoglycan conformations." *Nature* **322**: 215-216.
- Casu, B., Guerrini, M., Naggi, A., Perez, M., Torri, G., Ribatti, D., Carminati, P., Giannini, G., Penco, S., Pisano, C., Belleri, M., Rusnati, M. and Presta, M. (2002). "Short heparin sequences spaced by glycol-split uronate residues are antagonists of fibroblast growth factor 2 and angiogenesis inhibitors." *Biochemistry* **41**(33): 10519-10528.
- Chuang, W. L., Christ, M. D., Peng, J. and Rabenstein, D. L. (2000). "An NMR and molecular modelling study of the site-specific binding of Histamine by Heparin, chemically modified Heparin and Heparin-derived oligosaccharides." *Biochemistry* **39**: 3542-3555.
- Chuang, W. L., Christ, M. D. and Rabenstein, D. L. (2001). "Determination of the primary structures of heparin and heparan sulfate derived oligosaccharides using band selective homonuclear decoupled two dimensional ^1H NMR experiments." *Analytical. Chemistry*. **73**: 2310-2316.
- Coombe, D. R. and W. C. Kett (2005). "Heparan sulfate-protein interactions: therapeutic potential through structure-function insights." *Cellular and Molecular Life Sciences* **62**(4): 410-424.

- Couchman, J. R. (2003). "Syndecans: proteoglycan regulators of cell-surface microdomains?" *Nature Reviews Molecular Cell Biology* **4**(12): 926-938.
- Cremer, D. and J. A. Pople. (1974). "A General Definition of Ring Puckering Parameters." *Journal of the American Chemical Society* **97**(6): 1354-1357.
- Cremer, D. and Pople, J. A. (1975). "A general definition of ring puckering coordinates." *J. Am. Chem. Soc.* **97**: 1354-1358.
- Cremer, D. (1984). "On the correct usage of the Cremer-Pople puckering parameters as quantitative descriptors of ring shapes - a reply to recent criticism by Petit, Dillen and Geise." *Acta. Cryst.* **B40**: 498-500.
- Dai, Y., Y. Yang, *et al.* (2005). "HSulf-1 and HSulf-2 are potent inhibitors of myeloma tumor growth in vivo." *The Journal of Biological Chemistry* **280**(48): 40066.
- Das, S. K., Mallet, J. M., Esnault, J., Driguez, P. A., Duchaussoy, P., Sizun, P., Herault, J. P., Herbert, J. M., Petitou, M. and Sinay, P. (2001). "Synthesis of conformationally locked carbohydrates: a skew-boat conformation of L-iduronic acid governs the antithrombotic activity of heparin." *Angew Chem Int Ed Engl* **40**: 1670-1673.
- Deakin, J. A., Blaum, B. S., Gallagher, J. T., Uhrin, D. and Lyon, M. (2009). "The binding properties of minimal oligosaccharides reveal a common heparan sulfate/dermatan sulfate-binding site in hepatocyte growth factor/scatter factor that can accommodate a wide variety of sulfation patterns." *Journal of Biological Chemistry* **284**(10): 6311.
- De Cat, B. and G. David (2001). Developmental roles of the glypicans, Elsevier.
- DeLano WL (2002) The PyMol molecular graphics system. <http://www.pymol.org>
- Denzer, A. Schulthess, J., T., Fauser, C., Schumacher, B., Kammerer, R. A., Engel, J and Ruegg, M. A. (1998). "Electron microscopic structure of agrin and mapping of its binding site in laminin-1." *The EMBO journal* **17**: 335-343.
- Dhoot, G. K., Gustafsson, M. K., Ai, X., Sun, W., Standiford, D. M. and Emerson, J. C. P. (2001). Regulation of Wnt signalling and embryo patterning by an extracellular sulfatase. *Science* **293**: 1663-1666.
- Dietrich, C. P., M. E. Silva, Michelacci, Y. M. (1973). "Sequential degradation of heparin in *Flavobacterium heparinum*." *Journal of Biological Chemistry* **248**(18): 6408.
- DiGabriele, A. D., Lax, I., Chan, D. I., Svahn, C. M., Jaye, M., Schlessinger, J. and Hendrickson, W. A. (1998). "Structure of a heparin-linked biologically active dimer of fibroblast growth factor." *Nature* **393**(6687): 812-817.
- Dong, S., Cole, G. J. and Halfter, W. (2003). "Expression of collagen XVIII and localization of its glycosaminoglycan attachment sites." *J. Biological Chemistry*. **278**(3): 1700-7.

- Duchesne, L., Tissot, B., Rudd, T. R., Dell, A. and Fernig, D. G. (2006). "N-glycosylation of fibroblast growth factor receptor 1 regulates ligand and heparan sulfate co-receptor binding." *J. Biological Chemistry* **281**(37): 27178-27189.
- Esko, J. D. and U. Lindhahl (2001). "Molecular diversity of heparan sulfate." *The Journal of Clinical Investigation* **108**(2): 169-173.
- Faham, S., Hileman, R. E., Fromm, J. R., Linhardt, R. J. and Rees, D. C. (1996). "Heparin structure and interactions with basic fibroblast growth factor." *Science* **271**(5252): 1116-20.
- Ferro, D., Provasoli, A., Ragazzi, M., Torri, G., Casu, B., Gatti, G., Claude, J., Pierre, J., Sinay, P., Petitou, M. and Choay, J. (1986). "Evidence for Conformational Equilibrium of the Sulfated L-Iduronate Residue in Heparin and in Synthetic Heparin Mono- and Oligosaccharides: NMR and Forcefield Studies." *American Chemical Society* **108**: 6773-6778.
- Ferro, D. R., Provasoli, A., Ragazzi, M., Casu, B., Torri, G., Bossennec, V., Perly, B. and Sinay, P. (1990). "Conformer populations of L-Iduronic acid residues in glycosaminoglycan sequences." *Carbohydr. Res.* **195**: 157-167.
- Filmus, J. and S. B. Selleck (2001). "Glypicans: proteoglycans with a surprise." *The Journal of Clinical Investigation* **108**(4): 497-501.
- Folkman, J. (2006). "Angiogenesis." *Annual Review of Medicine* **57**: 1-18.
- Forsten-Williams, K., Chu, C. L., Fannon, M., Buczek-Thomas, J-A. and Nugent, M. (2008). "Control of Growth Factor Networks by Heparan Sulfate Proteoglycans." *Annals of Biomedical Engineering* **36**(12): 2134-2148.
- Forster, M. J. and B. Mulloy (1993). "Molecular dynamics study of iduronate ring conformation." *Biopolymers* **33**(4): 575-588.
- Friedrich, M. V. K., Göhring, W., Morgelin, M., Brancaccio, A., David, G. and Timpi, R. (1999). "Structural basis of glycosaminoglycan modification and of heterotypic interactions of perlecan domain V1." *Journal of molecular biology* **294**(1): 259-270.
- Fry, E. E., Lea, S. M., Jackson, T., Newman, J. W. I., Ellard, F. M., Blakemore, W. E., Abu-Ghazaleh, R., Samuel, A., King, A. M. Q. and Stuart, D. I. (1999). "The structure and function of a foot-and-mouth disease virus-oligosaccharide receptor complex." *The EMBO Journal* **18**(3): 543-554.
- Gallagher, J. T. (2001). "Heparan sulfate: growth control with a restricted menu." *The Journal of Clinical Investigation* **108**(357-361).
- Gallagher, J. T. and A. Walker (1985). "Molecular distinctions between heparan sulphate and heparin. Analysis of sulphation patterns indicates that heparan sulphate and heparin are separate families of N-sulphated polysaccharides." *Biochemical Journal* **230**(3): 665.
- Gallagher, J. T. and J. E. Turnbull (1992). "Heparan sulphate in the binding and activation of basic fibroblast growth factor". *Glycobiology* **2**(6): 523-528.

- Gallihier, P. M., Cooney, C. L., Langer, R. and Linhardt, R. J. (1981). "Heparinase production by *Flavobacterium heparinum*." *Appl. Environ. Microbiol.* **41**(2): 360-5.
- Goodger, S. J. (2003). "The Role and Structure of Heparan Sulphate in the Regulation of FGF1 and FGF2." *Faculty of Medicine, Dentistry, Nursing and Pharmacy*. Manchester, The University of Manchester. **PhD**.
- Goodger, S. J., Robinson, C. J., Murphy, K. J., Gasiunas, N., Harmer, N., Blundell, T. L., Pye, D. A. and Gallagher, J. T. (2008). "Evidence that heparin saccharides promote FGF2 mitogenesis through two distinct mechanisms." *Journal of Biological Chemistry* **283**(19): 13001.
- Groffen, A. J. A., Buskens, C. A. F., van Kuppevelt, T. H., Veerkamp, J. H., Monnens, L. A. H. and van der Heuvel, L. P. W. J. (2001). "Primary structure and high expression of human agrin in basement membranes of adult lung and kidney." *European Journal of Biochemistry* **254**(1): 123-128.
- Guerrini, M., Guglieri, S., Beccati, D., Torri, C. and Mourier, P. (2006) "Conformational transitions induced in heparin octasaccharides by binding with antithrombin III." *Biochemical Journal* **399**(Pt 2): 191.
- Guglieri, S., Hricovini, M., Raman, R., Polito, L., Torri, G., Casu, B., Sasisekharan, R. and Guerrini. (2008). "Minimum FGF2 Binding Structural Requirements of Heparin and Heparan Sulfate Oligosaccharides as Determined by NMR Spectroscopy." *Biochemistry* **47**(52): 13862-13869.
- Guimond, S., Maccarana, M., Olwin, B. B., Lindahl, U. and Rapraeger, A. C. (1993). "Activating and inhibitory heparin sequences for FGF-2 (basic FGF). Distinct requirements for FGF-1, FGF-2, and FGF-4." *J. Biol. Chem.* **268**(32): 23906-14.
- Haasnoot, C. A. G. (1992). "The Conformation of six-membered Rings Described by Puckering Coordinates Derived from Endocyclic Torsion Angles". *J. Am. Chem Soc.* **114**: 882-887.
- Habuchi, H., O. Habuchi, *et al.* (1995). "Purification and characterization of heparan sulfate 6-sulfotransferase from the culture medium of Chinese hamster ovary cells." *Journal of Biological Chemistry* **270**(8): 4172.
- Habuchi, H., Tanaka, M., Habuchi, O., Yoshida, K., Suzuki, H., Ban, K. and Kimata, K. (2000). The occurrence of three isoforms of heparan sulfate 6-O-sulfotransferase having different specificities for hexuronic acid adjacent to the targeted N-sulfoglucosamine. *J. Biol. Chem.* **275**(4): 2859-68.
- Habuchi, O. (2000). "Diversity and functions of glycosaminoglycan sulfotransferases." *Biochemica Et Biophysica Acta (BBA)* **1474**(2): 115-127.
- Hagner Mcwhirter, A., Lindahl, U. and Li, J. (2000). "Biosynthesis of heparin/heparan sulphate: mechanism of epimerization of glucuronyl C-5." *Biochem. J.* **347 Pt 1**: 69-75.

- Halfter, W., Dong, S., Schurer, B. and Cole, G. J. (1998). "Collagen XVIII is a basement membrane heparan sulfate proteoglycan." *J. Biol. Chem.* **273**(39): 25404-12.
- Harmer, N. (2006). "Insights into the role of heparan sulphate in fibroblast growth factor signalling." *Biochemical Society Transactions* **34**: 442-445.
- Harmer, N. J., Robinson, C. J., Adam, L. E., Ilag, L. L., Robinson, C. V., Gallagher, J. T. and Blundell, T. L. (2006). "Multimers of the fibroblast growth factor (FGF) - FGFR receptor-saccharide complex are formed on long oligomers of heparin." *Biochemistry* **393**: 741-748.
- Harmer, N. J., Pellegrini, L., Chirgadze, D., Fernandez-Recio, J. and Blundell, T. L. (2004). "The crystal structure of fibroblast growth factor (FGF) 19 reveals novel features of the FGF family and offers a structural basis for its unusual receptor affinity." *Biochemistry* **43**(3): 629-640.
- Harmer, N. J., Ilag, L., Mulloy, B., Pellegrini, L., Robinson, C. V. and Blundell, T. L. (2004). "Towards a resolution of the stoichiometry of the fibroblast growth factor FGF receptor - Heparin complex." *Journal of Molecular Biology* **339**: 821-834
- Hileman, R. E., Smith, A. E., Toida, T. and Linhardt, R. J. (1997). "Preparation and structure of heparin lyase-derived heparan sulfate oligosaccharides." *Glycobiology*. **7**(2): 231-239.
- Hricovini, M., Guerrini, M., Torri, G., Piani, S. and Ungarelli, F. (1995). "Conformational analysis of Heparin epoxidein aqueous solution: An NMR relaxation study." *Carbohydrate Research* **277**(1): 11-23.
- Hricovini, M., Guerrini, M., Bisio, A., Torri, G., Petitou, M. and Casu, B. (2001). Conformation of heparin pentasaccharide bound to antithrombin III. *Biochem. J.* **359**: 265-272
- Hricovini, M. and F. Bízík (2007). "Relationship between structure and three-bond proton-proton coupling constants in glycosaminoglycans." *Carbohydrate Research* **342**(6): 779-783.
- Inoue, Y., Inouye, Y., Nagasawa, K. (1990). "Conformational equilibria of the L-iduronate residue in non-sulphated di-, tetra- and hexa-saccharides and their alditols derived from dermatan sulphate." *Biochemical Journal* **265**(2): 533.
- lozzo, R. V. (2005). "Basement membrane proteoglycans: from cellar to ceiling." *Nature Reviews Molecular Cell Biology* **6**(8): 646-656.
- lozzo, R. V and J. D. San Antonio. (2001) "Heparan sulfate proteoglycans: heavy hitters in the angiogenesis arena." *The journal of Clinical Investigation* **108**(3): 349-355
- lozzo, R. V., Cohen, I. R., Grassel, S. and Murdoch, A. D. (1994). "The biology of perlecan: the multifaceted heparan sulphate proteoglycan of basement membranes and pericellular matrices." *Biochemical Journal* **302**(3): 625.

- Ishihara, M., Shaklee, P. N., Yang, Z., Liang, W., Wei, Z., Stack, R. J. and Holme, K. (1994). "Structural features in heparin which modulate specific biological activities mediated by basic fibroblast growth factor." *Glycobiology* **4**(4): 451-458.
- Ishihara, M. (1994). "Structural requirements in heparin for binding and activation of FGF1 and FGF4 are different from that for FGF2." *Glycobiology* **4**(6):817-824
- Jacobsson, I., Lindahl, U., Jensen, J. W., Roden, L., Prihar, H. and Feingold, D. S. (1984). "Biosynthesis of heparin. Substrate specificity of heparosan N-sulfate D-glucuronosyl 5-epimerase." *J. Biol. Chem.* **259**(2): 1056-63
- Jastrebova, N., Vanwildemeersch, M., Rapraeger, A. C., Gimenez-Gallego, G., Lindahl, U. and Spillmann, D. (2006). "Heparan sulfate-related oligosaccharides in ternary complex formation with fibroblast growth factors 1 and 2 and their receptors." *J. Biol. Chem.* **281**(37): 26884-92.
- Kakuta, Y., Sueyoshi, T., Negishi, M. and Pederson, L. C. (1999). "Crystal structure of the sulfotransferase domain of human heparan sulfate N-deacetylase/N-sulfotransferase 1." *Journal of Biological Chemistry* **274**(16): 10673-10676.
- Kamimura, K., Fujise, M., Villa, F., Izumi, S., Habucni, H., Kimata, K. and Nakato, H. (2001). "Drosophila Heparan Sulfate 6-O-Sulfotransferase (dHS6ST) Gene structure, expression and function in the formation of the tracheal system." *The Journal of Biological Chemistry* **276**(20): 17014-17021.
- Kim, C. W., Goldberger, O. A., Gallo, R. L. and Bernfield, M. (1994) "Members of the syndecan family of heparan sulfate proteoglycans are expressed in distinct cell-, tissue-, and development-specific patterns." *Mol Biol Cell.* **5**(7): 797–805.
- Kim, B. T., Kitagawa, H., Tamura, J., Saito, T., Kusche_Gullberg, M., Lindahl, U. and Sugahara, K. (2001). Human tumor suppressor EXT gene family members EXTL1 and EXTL3 encode alpha 1,4- N-acetylglucosaminyltransferases that likely are involved in heparan sulfate/ heparin biosynthesis. *Proc. Natl. Acad. Sci.* **98**(13): 7176-81.
- Kirn-Safran, C., Farach-Carson, M. C., Carson, D. D. (2009). "Multifunctionality of extracellular and cell surface heparan sulfate proteoglycans." *Cellular and Molecular Life Sciences* **66**: 3421-3434.
- Kirschner, K. N. and R. J. Woods (2001). "Solvent interactions determine carbohydrate conformation." *Proceedings of the National Academy of Sciences* **98**(19): 10541.
- Kobayashi, M., Habuchi, H., Yoneda, M., Habuchi, O. and Kimata.. (1997) "Molecular Cloning and Expression of Chinese Hamster Ovary Cell Heparan-sulfate 2-Sulfotransferase." *J. Biol. Chem.* **272**: 13980-13985
- Kolset, S. O., Prydz, K. and Pejler. (2004). "Intracellular proteoglycans." *Biochemical Journal* **379**(Pt 2): 217.
- Krawczuk, P. (2005) "The Anomeric Effect." Baran Group Meeting.

- Kreuger, J., Salmivirta, M., Sturiale, L., Gimenez_Gallego, G. and Lindahl, U. (2001). "Sequence analysis of heparan sulfate epitopes with graded affinities for fibroblast growth factors 1 and 2." *J. Biol. Chem.* **276**(33): 30744-52.
- Kreuger, J., Jemth, P., Sanders_Lindberg, E., Eliahu, L., Ron, D., Basilico, C., Salmivirta, K. and Lindahl, U. (2005). "Fibroblast growth factors share binding sites in Heparan Sulphate." *Biochem. J.* **389**: 145-150.
- Kreuger, J., Spillmann, D., Li, J.-P. and Lindahl, U. (2006). "Interactions between heparan sulfate and proteins: the concept of specificity." *J. Cell. Biol.* **174**(3): 323-327.
- Lamanna, W. C., Kalus, I., Padva, M., Baldwin, R. J., Merry, C. L. R. and Dierks, T. (2007). "The heparanome — The enigma of encoding and decoding heparan sulfate sulfation." *Journal of biotechnology* **129**(2): 290-307.
- Langsdorf, A., Do, A. T., Kusche-Gullberg, M., Emerson Jr, C. P. and Ai, X. (2007). "Sulfs are regulators of growth factor signaling for satellite cell differentiation and muscle regeneration." *Developmental Biology* **311**(2): 464-477.
- Lindahl, U., Lidholt, K., SPillman, D. and Kjellen, L. (1994). "More to" heparin" than anticoagulation." *Thrombosis research* **75**(1): 1.
- Lindahl, U., Kusche-Gullberg, M. and Kjellen, L. (1998). Regulated diversity of heparan sulfate. *J. Biol. Chem.* **273**: 24979-24982.
- Linhardt, R. J., Turnbull, J. E., Wang, H. M., Loganathan, D. and Gallagher, J. T (1990). "Examination of the substrate specificity of heparin and heparan sulfate lyases. ." *Biochemistry* **29**(10): 2611-7.
- Linker, A. (1979). "Structure of heparan sulphate oligosaccharides and their degradation by exo-enzymes." *Biochemical Journal* **183**(3): 711.
- Liu, J., Shworak, N. W., Fritze, L. M. S., Edelberg, J. M. and Rosenberg. (1996). "Purification of heparan sulfate D-glucosaminyl 3-O-sulfotransferase." *The Journal of Biological Chemistry* **271**(43): 27072.
- Liu, J., Shriver, Z., Blaiklock, P., Yoshida, K., Sasisekharan, R. and Rosenberg, R. D. (1999). "Heparan sulfate D-glucosaminyl 3-O-sulfotransferase-3A sulfates N-unsubstituted glucosamine residues." *J. Biol. Chem.* **274**(53): 38155-62.
- Liu, J., Shworak, N. W., Sinay, P., Schwartz, J. J., Zhang, L., Fritze, L. M. and Rosenberg, R. D. (1999). "Expression of heparan sulfate D-glucosaminyl 3-O-sulfotransferase isoforms reveals novel substrate specificities." *J. Biol. Chem.* **274**(8): 5185-92.
- Maccarana, M., Casu, B. and Lindahl, U. (1993). "Minimal sequence in heparin/heparan sulfate required for binding of basic fibroblast growth factor." *J. Biol. Chem.* **268**(32): 23898-905.
- Maccarana, M., Sakura, Y., Tawada, A., Yoshida, K. and Lindahl, U. (1996). "Domain structure of heparan sulfates from bovine organs." *J. Biol. Chem.* **271**(30): 17804-10.

- Marneros, A. G. and B. R. Olsen (2005). "Physiological role of collagen XVIII and endostatin." *The FASEB Journal* **19**(7): 716-728.
- Merry, C. L., Lyon, M., Deakin, J. A., Hopwood, J. J. and Gallagher, J. T. (1999). "Highly sensitive sequencing of the sulfated domains of heparan sulfate." *J. Biol. Chem.* **274**(26): 18455-62.
- Merry, C. L. R. and V. A. Wilson (2002). "Role of heparan sulfate-2-O-sulfotransferase in the mouse." *Biochemica Et Biophysica Acta (BBA)* **1573**(3): 319-327.
- Mikhailov, D., Mayo, K. H., Pervin, A. and Linhardt, R. J. (1996). "¹³C-NMR relaxation study of heparin-disaccharide interactions with tripeptides GRG and GKG." *Biochem. J.* **315**: 447-454.
- Mikhailov, D., Linhardt, R. J. and Mayo, K. H. (1997). "NMR solution conformation of heparin derived hexasaccharide." *Biochem. J.* **328**: 51-61.
- Mobli, M., Nilsson, M. and Almond, A. (2008). "The structural plasticity of heparan sulfate NA-domains and hence their role in mediating multivalent interactions is confirmed by high-accuracy 15 N-NMR relaxation studies." *Glycoconjugate Journal* **25**(5): 401-414.
- Mongiat, M., Taylor, K., Otto, J., Aho, S., Uitto, J., Whitelock, J. M. and Iozzo, R. V. (2000). "The protein core of the proteoglycan perlecan binds specifically to fibroblast growth factor-7." *Journal of Biological Chemistry* **275**(10): 7095-7100.
- Moy, F. J., Safran, M., Seddon, A. P., Kitchen, D., Bohlen, P., Aviezer, D., Yayon, A. and Powers, R. (1997). "Properly Oriented Heparin- Decasaccharide-Induced Dimers Are the Biologically Active Form of Basic Fibroblast Growth Factor." *Biochemistry* **36**(16): 4782-4791.
- Mulloy, B., Forster, M. J., Jones, C. and Davies, D. B. (1993). "NMR and molecular modelling studies of the solution conformation of heparin." *Biochem. J.* **293**: 849-858.
- Mulloy, B., Forster, M. J., Jones, C., Drake, A. F., Johnson, E. A. and Davies, D. B. (1994). "The effect of variation of substitution on the solution conformation of heparin: a spectroscopic and molecular modelling study." *Carbohydr. Res.* **255**: 1-26.
- Mulloy, B. and Forster, M. J. (2000). "Conformation and dynamics of heparin and heparan sulfate." *Glycobiology* **10**(11): 1147-56.
- Murphy, K. J., Merry, C. L. R., Lyon, M., Thompson, J. E., Roberts, I. S. and Gallagher, J. T. (2004). "A new model for the domain structure of Heparan Sulfate based on the novel specificity of K5 lyase." *J. Biol. Chem.* **279**(26): 27239-27245.
- Murphy, K. J. (2007). "An NMR and molecular modelling study on the solution conformation of Heparin Sulphate: new insights into the relationship between structure and function." *School of Chemical and Biological Sciences*. Huddersfield, The University of Huddersfield. **PhD**.

- Murphy, K. J., Mclay, N. and Pye, D. A. (2008). "Structural studies of Heparan Sulfate hexasaccharides: New Insights into Iduronate Conformational Behaviour." *Journal of the American Chemical Society*. **130**(37): 12435 - 12444.
- Nader, H. B., Chavante, S. F., Dos-Santos, E. A., Oliveria, F. W., de-Paiva, J. F., Jeronimo, S. M. B., Medeiros, G. F., de-Abrell, L. R. D., Leite, E. L., de-Sousa-Filho, J. F., Castro, R. A. B., Toma, L., Tersariol, I. L. S., Porcionatto, M. A. and Dietrich, C. P. (1999). "Heparan sulfates and heparins: similar compounds performing the same functions in vertebrates and invertebrates?" *Brazilian Journal of Medical and Biological Research* **32**: 529-538.
- Nakato, H. and K. Kimata (2002). "Heparan sulfate fine structure and specificity of proteoglycan functions." *BBA-General Subjects* **1573**(3): 312-318.
- Nikitovic, D., Assouti, M., Sifaki, M., Katonis, P., Krasagakis, K., Karamanos, N. K. and Tzanakakis, G. N. (2008). "Chondroitin sulfate and heparan sulfate-containing proteoglycans are both partners and targets of basic fibroblast growth factor-mediated proliferation in human metastatic melanoma cell lines." *International Journal of Biochemistry and Cell Biology* **40**(1): 72-83.
- Noonan, D. M. and J. R. Hassell (1993). "Perlecan, the large low-density proteoglycan of basement membranes: structure and variant forms." *Kidney international* **43**: 53-53.
- Nugent, M. A. and R. V. Iozzo (2000). "Fibroblast growth factor-2." *International Journal of Biochemistry and Cell Biology* **32**(2): 115-120.
- Olsen, S. K., Garbi, M., Zampieri, N., Eliseenkova, A. V., Ornitz, D. M., Goldfarb, M. and Mohammadi, M. (2003). "Fibroblast Growth Factor (FGF) Homologous Factors Share Structural but Not Functional Homology with FGFs." *Journal of Biological Chemistry* **278**(36): 34226-34236.
- Ornitz, D. M., Xu, J., Colvin, J. S., McEwan, D. G., MacArthur, C. A., Coulier, F., Gao, G. and Goldfarb, M. (1996). Receptor specificity of the fibroblast growth factor family. *J. Biol. Chem.* **271**(25): 15292-15297.
- Ornitz, D. M. (2000). "FGFs, heparan sulfate and FGFRs: complex interactions essential for development." *BioEssays* **22**: 108-112.
- Pedersen, L. C., Tsuchida, K., Kitagawa, H., Sugahara, K., Darden, T. A. and Negishi, M. (2000). "Heparan/chondroitin sulfate biosynthesis Structure and mechanism of human glucuronyltransferase I." *The Journal of Biological Chemistry* **275**(44): 34580-34585.
- Pellegrini, L., Burke, D. F., Von_Delft, F., Mulloy, B. and Blundell, T. L. (2000). Crystal structure of fibroblast growth factor receptor 2 ectodomain bound to ligand and heparin. *Nature* **407**: 1029-1034.
- Perez, S., Kouwijzer, M., Mazeau, K. and Engelsen, S. (1996). "Modeling Polysaccharides: Present status and challenges." *Journal of Molecular Graphics* **14**: 307-321.

- Perrimon, N. and M. Bernfield (2000). "Specificities of heparan sulphate proteoglycans in developmental processes." *Nature (London)* **404**(6779): 725-728.
- Pikas, D. S., Eriksson, I. and Kjellen, L. (2000). "Overexpression of Different Isoforms of Glucosaminyl N-Acetylase/N-Sulfotransferase Results in Distinct Heparan Sulfate N-Sulfation Patterns." *Biochemistry* **39**(15): 4552-4558.
- Pinhal, M. A. S., Smith, B., Olsen, S., Aikawa, J., Kimata, K. and Esko, J. D. (2001). "Enzyme interactions in heparan sulfate biosynthesis: Uronosyl 5-epimerase and 2-O-sulfotransferase interact in vivo." *PNAS* **98**(23): 12984.
- Powell, A. K. (2001). Characterization of the Interaction of Fibroblast Growth Factor Receptors with Heparan Sulphate. *School of Biosciences*. Birmingham, The University of Birmingham.
- Powell, A. K., Yates, E. A., Fernig, D. G. and Turnbull, J. E. (2004). "Interactions of heparin/heparan sulfate with proteins: appraisal of structural factors and experimental approaches." *Glycobiology* **14**(4): 17-30.
- Prydz, K. and K. T. Dalen (2000). "Synthesis and sorting of proteoglycans." *Journal of Cell Science* **113**(2): 193-205.
- Pye, D. A., Vives, R. R., Turnbull, J. E., Hyde, P. and Gallagher, J. T. (1998). Heparan sulfate oligosaccharides require 6-O-sulfation for promotion of basic fibroblast growth factor mitogenic activity. *J. Biol. Chem.* **273**(36): 22936-42.
- Pye, D. A. and J. T. Gallagher (1999). "Monomer complexes of basic fibroblast growth factor and heparan sulfate oligosaccharides are the minimal functional unit for cell activation." *Journal of Biological Chemistry* **274**(19): 13456-13461.
- Pye, D. A., Vives, R. R., Hyde, P. and Gallagher, J. T. (2000). Regulation of FGF-1 mitogenic activity by heparan sulfate oligosaccharides is dependent on specific structural features: differential requirements for the modulation of FGF-1 and FGF-2. *Glycobiology*. **10**(11): 1183-92.
- Rabenstein, D. L. (2002). "Heparin and heparan sulfate: structure and function." *Natural Product Reports* **19**(3): 312-331.
- Ragazzi, M., Ferro, D. R. and Provasoli, A. (1986). A force field study of the conformational characteristics of the Iduronate ring. *J. Comput. Chem.* **7**(2): 105-112.
- Ragazzi, M., Ferro, D. R., Provasoli, A., Pumilia, P., Cassinari, A., Torri, G., Guerrini, M., Casu, B., Nader, H. B. and Dietrich, C. P. (1993). Conformation of the unsaturated uronic acid residues of glycosaminoglycan disaccharides. *J. Carbohydr. Chem.* **12**: 523-535
- Raman, R., Sasisekharan, V. and Sasisekharan, R. (2005). "Structural insights into biological roles of protein-glycosaminoglycan interactions." *Chemistry & biology* **12**(3): 267-277.

- Razi, N. and U. Lindahl (1995). "Biosynthesis of heparin/heparan sulfate. The D-glucosaminyl 3-O-sulfotransferase reaction: target and inhibitor saccharides." *The Journal of Biological Chemistry* **270**(19): 11267-11275.
- Rieckmann, T., Kotevic, I. and Trueb, B. (2007). "The cell surface receptor FGFR1 forms constitutive dimers that promote cell adhesion." *Experimental Cell Research* **314**(5): 1071-1081.
- Ringvall, M., Ledin, J., Holmborn, K., van Kuppevelt, T., Ellin, F., Eriksson, I., Olofsson, A., Kjellen, L. and Forsberg, E. (2000). "Defective heparan sulfate biosynthesis and neonatal lethality in mice lacking N-deacetylase/N-sulfotransferase-1." *Journal of Biological Chemistry* **275**(34): 25926-25930.
- Salmivirta, M., Lidholt, K. and Lindahl, U. (1996). "Heparan sulfate: a piece of information." *The FASEB Journal* **10**(11): 1270-1279.
- Sanderson, P. N., Huckerby, T. N. and Nieduszynski, I. A. (1987). "Conformational equilibria of alpha-L-iduronate residues in disaccharides derived from heparin." *Biochemical Journal* **243**(1): 175.
- Sasisekharan, R. and G. Venkataraman (2000). "Heparin and heparan sulfate: biosynthesis, structure and function." *Current opinion in chemical biology* **4**(6): 626-631.
- Schlessinger, J., Plotnikov, A. N., Ibrahimi, O. A., Eliseenkova, A. V., Yeh, B. K., Yayon, A., Linhardt, R. J. and Mohammadi, M. (2000). Crystal structure of a ternary FGF-FGFR-Heparin complex reveals a dual role for heparin in FGFR binding and dimerisation. *Mol. Cell*. **6**: 743-750
- Sedita, J., Izvolsky, K. and Cardoso, W. V. (2004). "Differential expression of heparan sulfate 6-O-sulfotransferase isoforms in the mouse embryo suggests distinctive roles during organogenesis." *Developmental Dynamics* **231**(4): 782-794.
- Seeliger, D and B. do Groot (2010). "Ligand docking and binding site analysis with Pymol and Autodock/Vina." *Journal of Computer-aided Molecular Design* **24**: 417-422
- Stauber, D. J., DiGabriele, A. D. and Hendrickson, W. A. (2000). "Structural interactions of fibroblast growth factor receptor with its ligands." *Proceedings of the National Academy of Sciences of the United States of America* **97**(1): 49-54.
- Steinbrecher, T. (2006). *AMBER Workshop*. AMBER 9, Imperial College, London, UK.
- Stringer, S. E. and Gallagher, J. T. (1997). Specific binding of the chemokine platelet factor 4 to heparan sulfate. *J. Biol. Chem.* **272**(33): 20508-14.
- Stringer, S. E. B., Kandola, B. S., Pye, D. A. and Gallagher, J. T. (2003). "Heparin Sequencing." *Glycobiology* **13**(2): 97-107
- Sugahara, K. and H. Kitagawa (2000). "Recent advances in the study of the biosynthesis and functions of sulfated glycosaminoglycans." *Current Opinion in Structural Biology* **10**(5): 518-527.

- Sugahara, K., Mikami, T., Uyama, T., Mizuguchi, S., Normura, K. and Kitagawa, H. (2003). "Recent advances in the structural biology of Chondroitin Sulfate and Dermatan Sulfate " *Current Opinion in Structural Biology* **13**(5): 612-620.
- Sugaya, N., Habuchi, H., Nagai, N., Ashikari-Hada, S and Kimata, K. (2008). "6-O-Sulfation of Heparan Sulfate Differentially Regulates Various Fibroblast Growth Factor-dependant Signalings in Culture." *The Journal of Biological Chemistry* **283**(16): 10366-10376.
- Thompson, L. D., Pantoliano, M. W. and Springer, B. A. (1994). "Energetic Characterization of the Basic Fibroblast Growth Factor-Heparin Interaction: Identification of the Heparin Binding Domain." *Biochemistry* **33**: 3831-3840.
- Trott, O. D. A. J. Olsen (2010). "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading." *Journal of Computational Chemistry* **31**: 455-461.
- Tsen, G., Halfter, W., Kroger, S. and Cole, G. J. (1995). Agrin is a heparan sulfate proteoglycan, *ASBMB*. **270**: 3392.
- Turnbull, J. E. and J. T. Gallagher (1991). "Distribution of iduronate 2-sulphate residues in heparan sulphate." *Journal of Biochemistry* **273**: 553-559.
- van_Boeckel, C. A. A., van_Aelst, S. F., Wagenaars, G. N., Mellema, J. R., Paulsen, H., Peters, T., Pollex, A. and Sinnwell, V (1987). "Conformational analysis of synthetic heparin-like oligosaccharides containing α -L-idopyranosyluronic acid." *Recueil des Travaux Chimiques des Pays-Bas* **106**(1): 19-29.
- Vertel, B. M., Walters, L. M., Flay, N., Kerns, A. E. and Schwartz, N. B. (1993). "Xylosylation is an endoplasmic reticulum to Golgi event." *Journal of Biological Chemistry* **268**(15): 11105-11112.
- Vives, R. R., Pye, D. A., Salmivirta, M., Hopwood, J. J., Lindahl, U. and Gallagher, J. T. (1999). Sequence analysis of heparan sulphate and heparin oligosaccharides. *Biochem. J.* **339** (3): 767-73.
- Walker, R. (2006). The Amber Workshop. Amber 9, Imperial College, London, UK.
- Wei, Z., Swiedler, S. J., Ishihara, M., Orellana, A. and Hirschberg, C. B. (1993). "A single protein catalyzes both N-deacetylation and N-sulfation during the biosynthesis of heparan sulfate." *PNAS* **90**(9): 3885-3888.
- Whitford, D. (2005). "Proteins, structure and function." Chichester, John Wiley and Sons, Ltd.
- Winzen, U., Cole, G. J. and Halfter, W. (2003). "Agrin Is a Chimeric Proteoglycan with the Attachment Sites for Heparan Sulfate/Chondroitin Sulfate Located in Two Multiple Serine-Glycine Clusters." *ASBMB*. **278**: 30106-30114.

Wu, Z. L., Zhang, L., Yabe, T., Kuberan, B., Beeler, D. L., Love, A. and Rosenberg, R. D. (2003). "The involvement of heparan sulfate (HS) in FGF1/HS/FGFR1 signaling complex." *Journal of Biological Chemistry* **278**(19): 17121-17129.

Zatterstrom, U., K., Felbor, U., Fukal, N. and Olsen, B. R. (2000). "Collagen XVIII/endostatin structure and functional role in angiogenesis." *Cell Structure and Function* **25**(2): 97-101.

Zhang, J., Cousens, L. S., Barr, P. J. and Sprang, S. R. (1991). "Three-dimensional structure of human basic fibroblast growth factor, a structural homolog of interleukin 113." *PNAS* **88**(8): 3446-3450.

Zhang, Z., Coomans, C. and David, G. (2001). "Membrane heparan sulfate proteoglycan-supported FGF2-FGFR1 signaling. Evidence in support of the cooperative end structures" model." *Journal of Biological Chemistry* **276**(45): 41921.

Chapter 7: Appendix

Figure 3.36

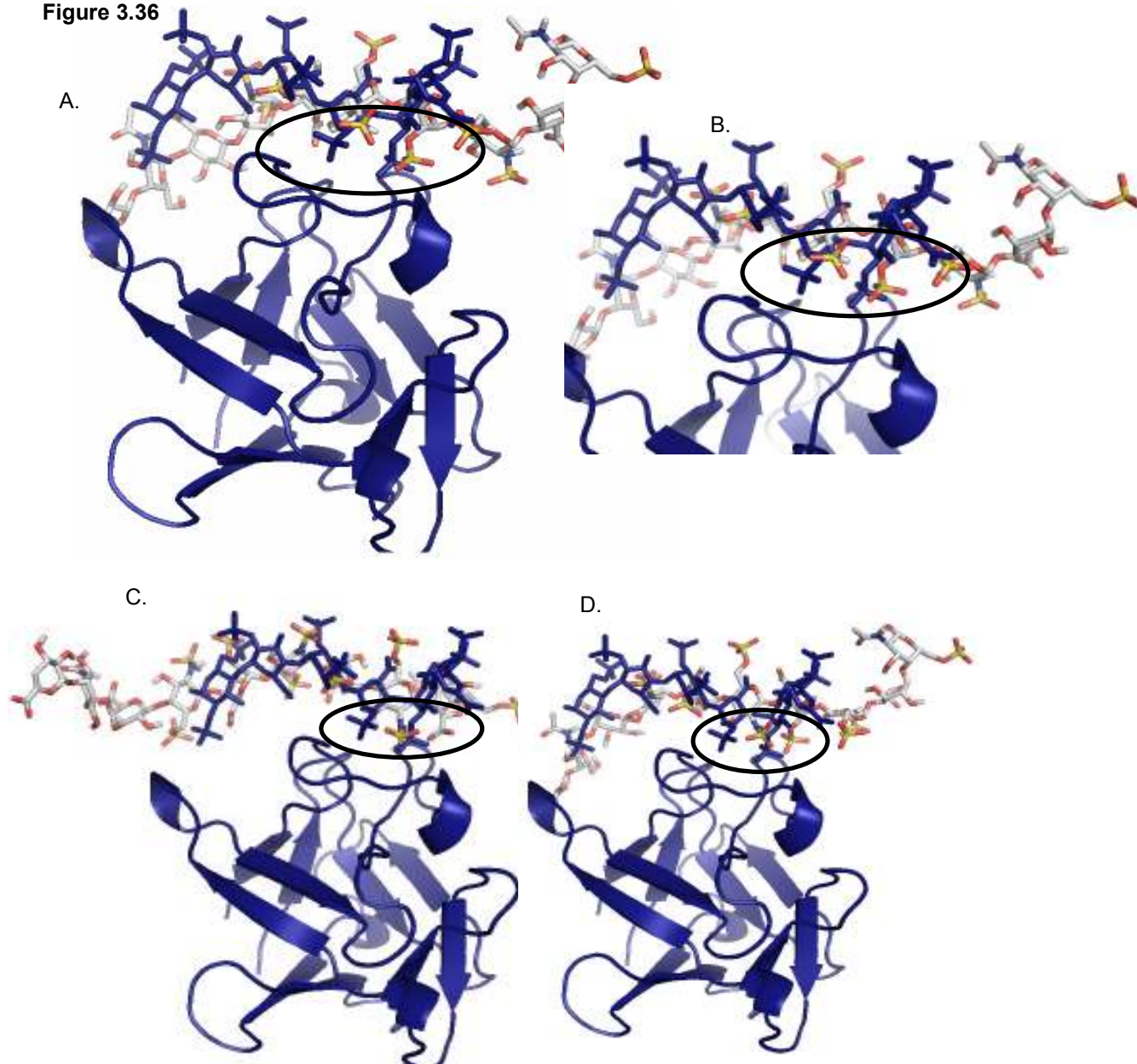


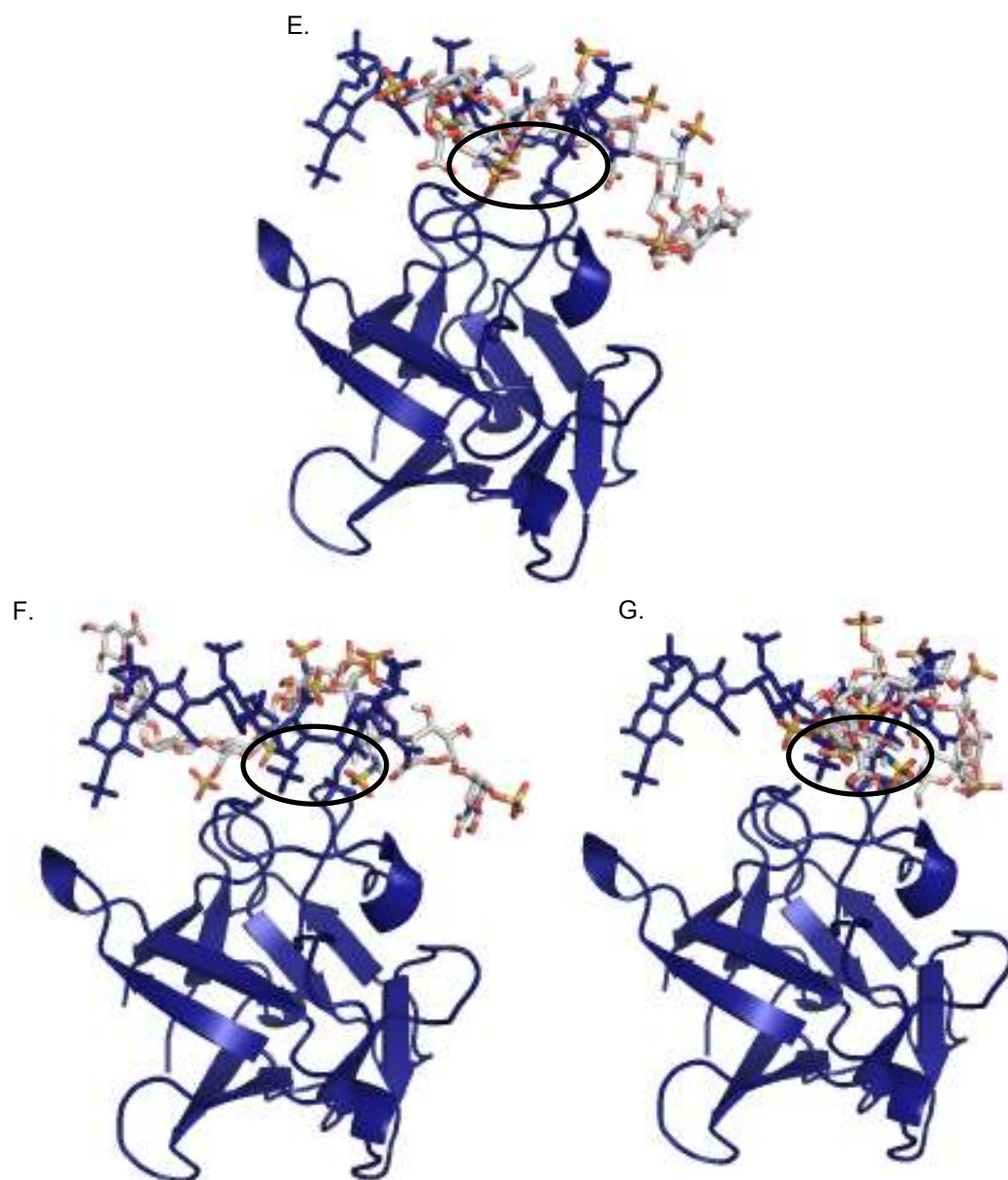
Figure 3.36. The highest affinity binding structure as predicted by Vina for **dp10(1)** for the 1000 ps MD simulation. This is the structure with the highest affinity (kcal/mol) and is rated as mode number 1 in the Vina search.

Panel A. Frame 45, model 1 with both binding positions predicted.

Panel B. The magnified image of frame 45.

Panel C. Frame 45, model 6.

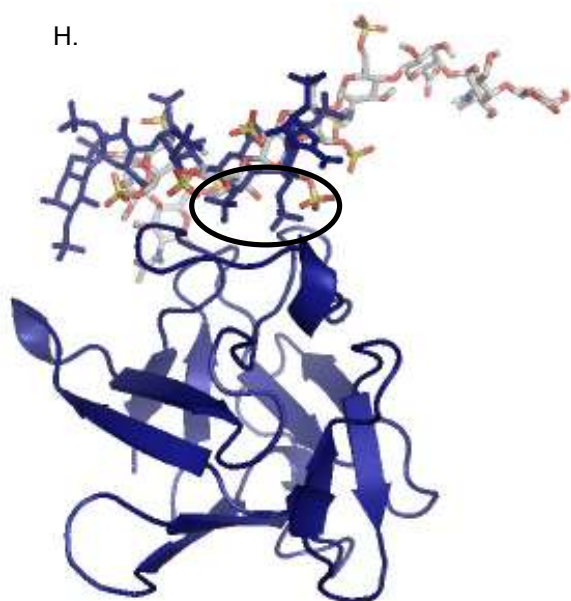
Panel D. Frame 45, model 9.



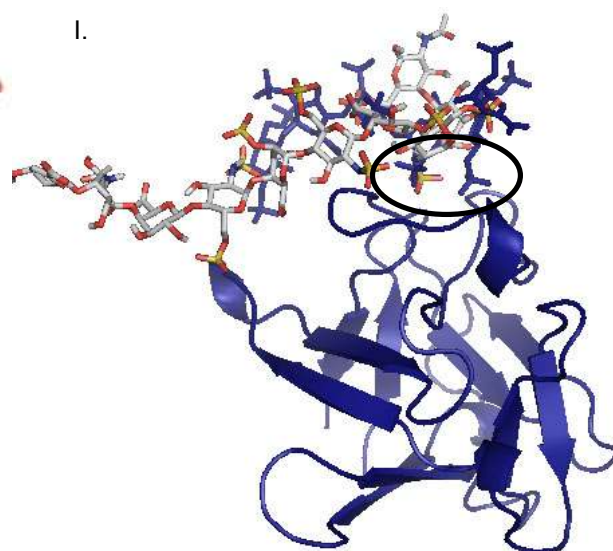
Panel E. Frame 350, model 7.

Panel F. Frame 350, model 10.

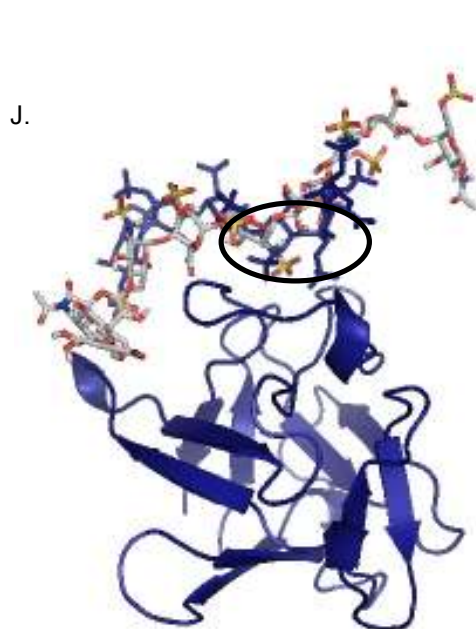
Panel G. Frame 350, model 11.



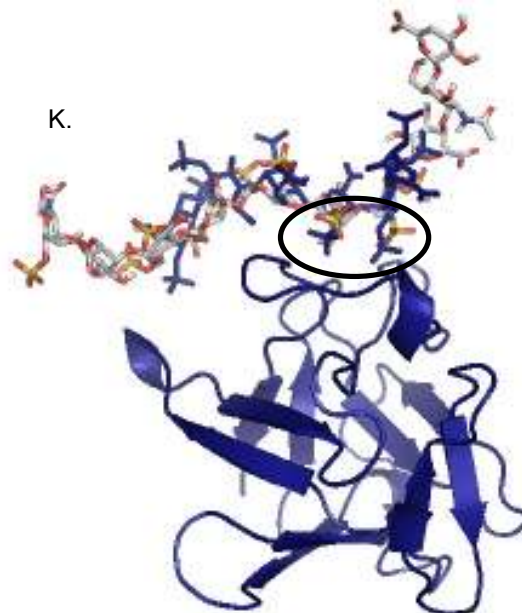
Panel H. Frame 975, model 1.



Panel I. Frame 975, model 3.



Panel J. Frame 1463, model 3.



Panel K. Frame 1463, model 15.

Figure 3.37

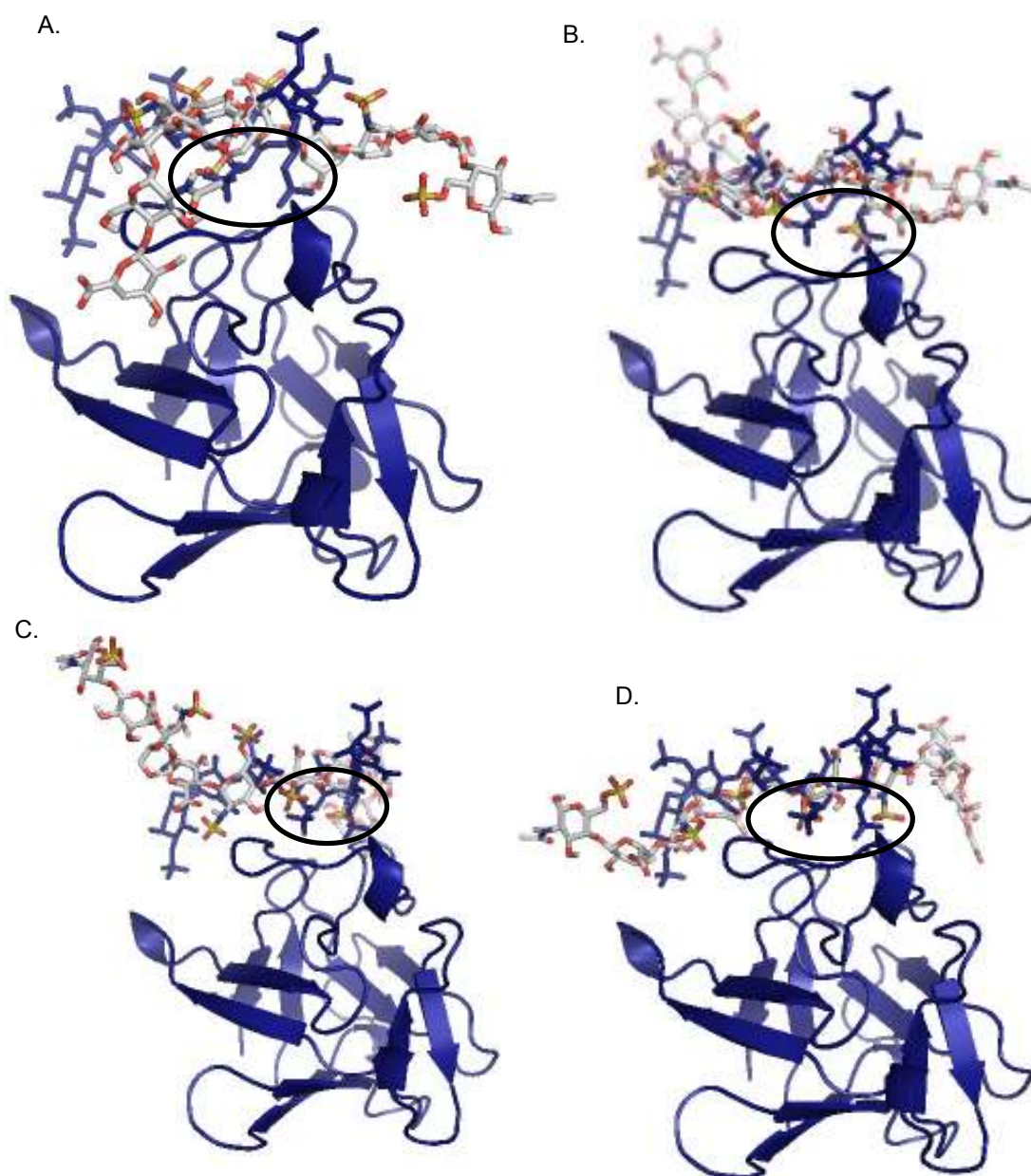


Figure 3.37. The highest affinity binding structure as predicted by Vina for **dp10(3)** with the best fit for the x-ray crystal structure for the 1000 ps MD simulation.

Panel A. Frame 200, model 3.

Panel B. Frame 200, model 10.

Panel C. Frame 1056, number 16.

Panel D. Frame 1300, number 3.

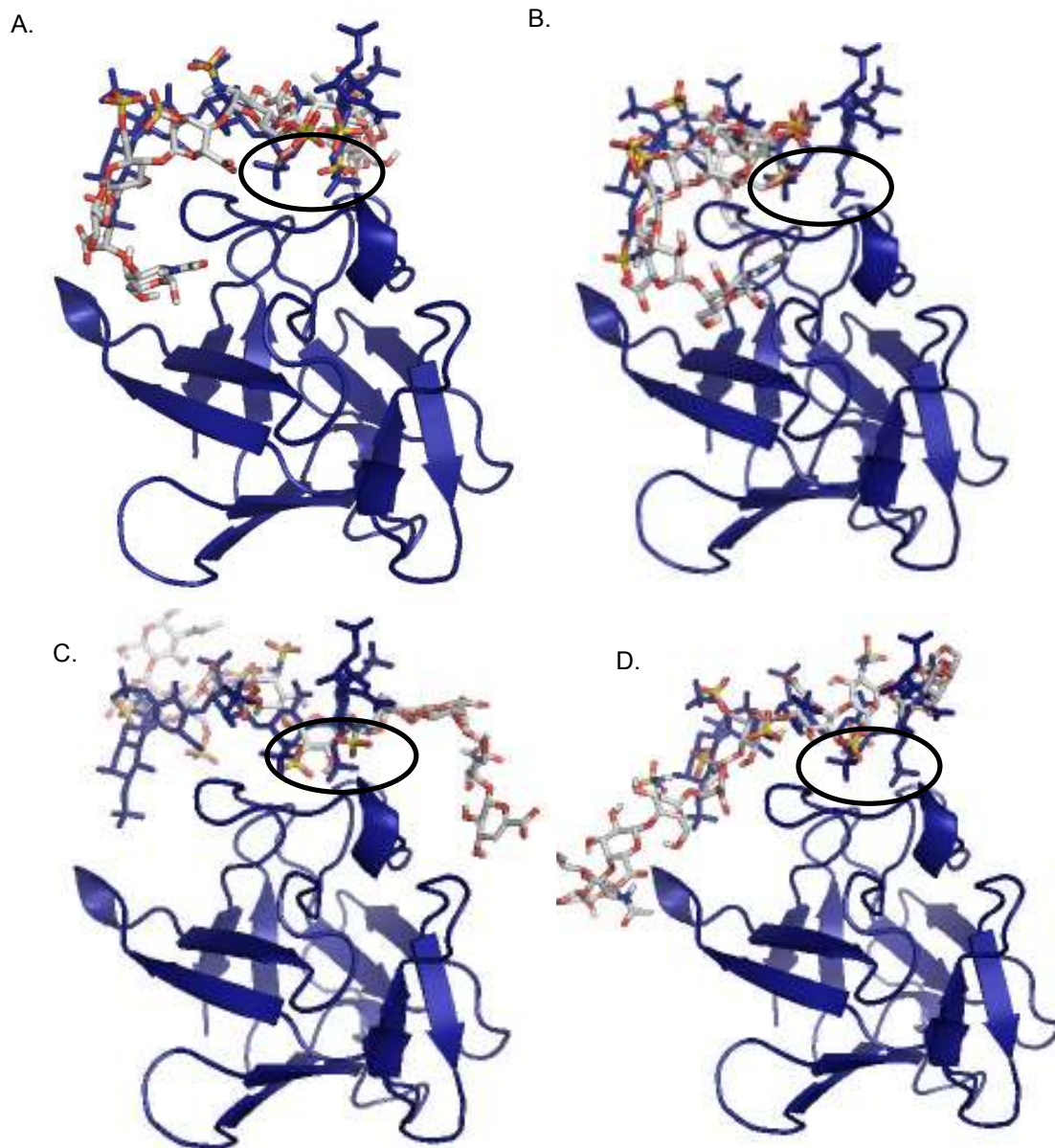
Figure 3.38

Figure 3.38. The highest affinity binding structure as predicted by Vina for **dp10(5)** with the best fit for the x-ray crystal structure for the 1000 ps MD simulation.

Panel A. Frame 30, model 2.

Panel B. Frame 30, model 11.

Panel C. Frame 1381, model 9.

Panel D. Frame 1381, model 5.

Figure 3.39

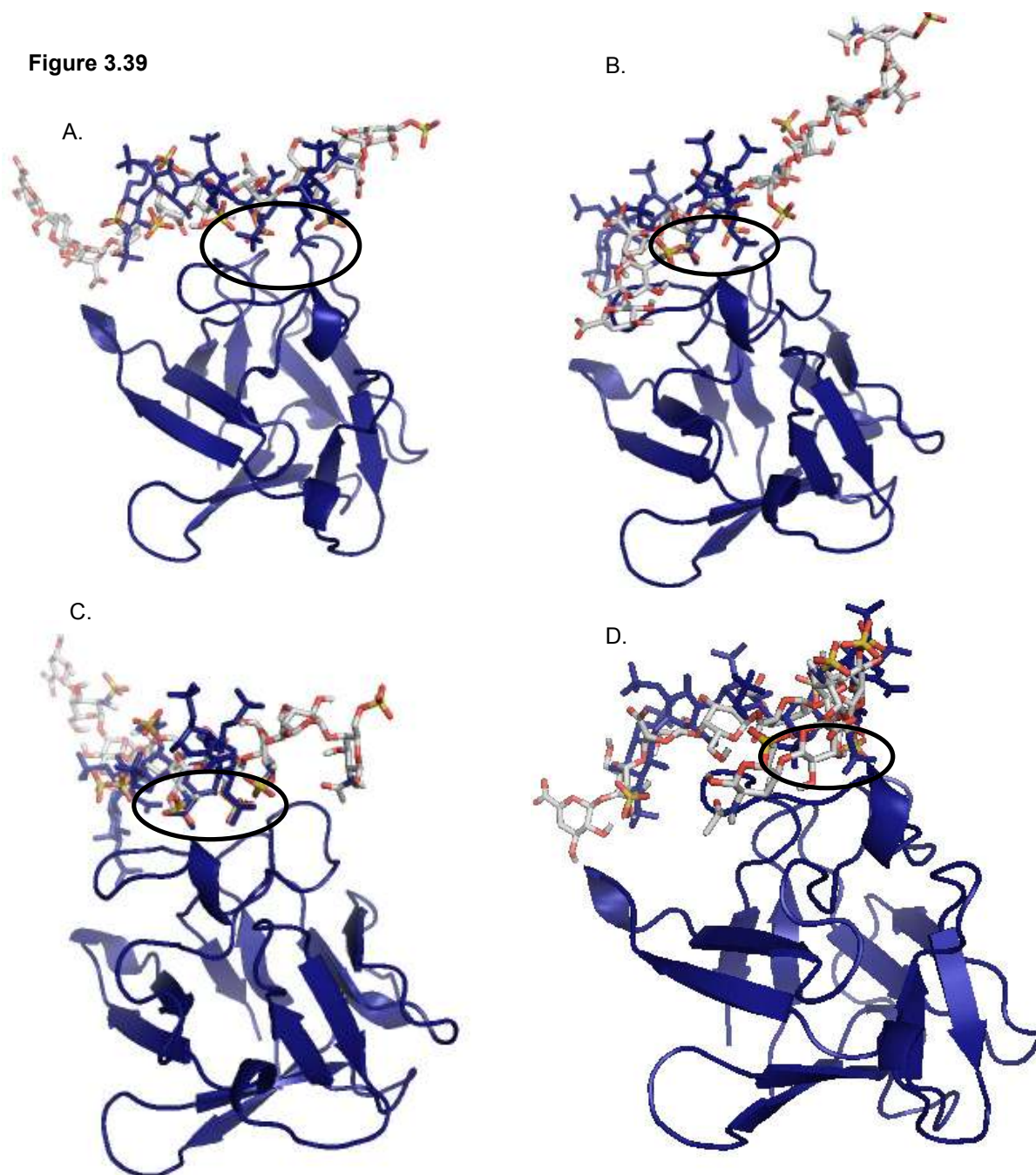


Figure 3.39. The highest affinity binding structure as predicted by Vina for **dp10(7)** with the best fit for the x-ray crystal structure for the 1000 ps MD simulation.

- A. Frame 50, model 10.
- B. Frame 200, model 2.
- C. Frame 200, model 15.
- D. Frame 1463, number 5.

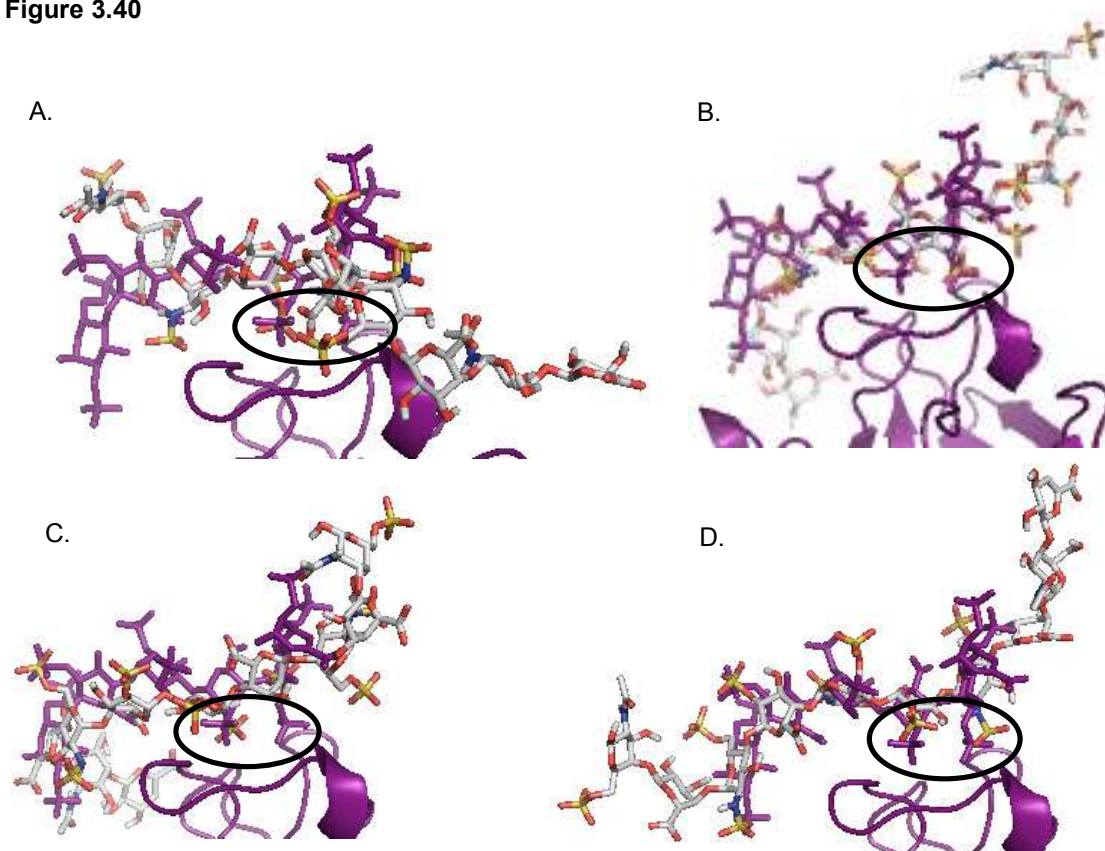
Figure 3.40

Figure 3.40. The highest affinity binding structure as predicted by Vina for **dp10(1)** with the best fit for the x-ray crystal structure for the 10 ns MD simulation.

A. Frame 200, model 1.

B. Frame 1300, model 11.

C. Frame 3000, model 1.

D. Frame 8300, model 12.

Figure 3.41

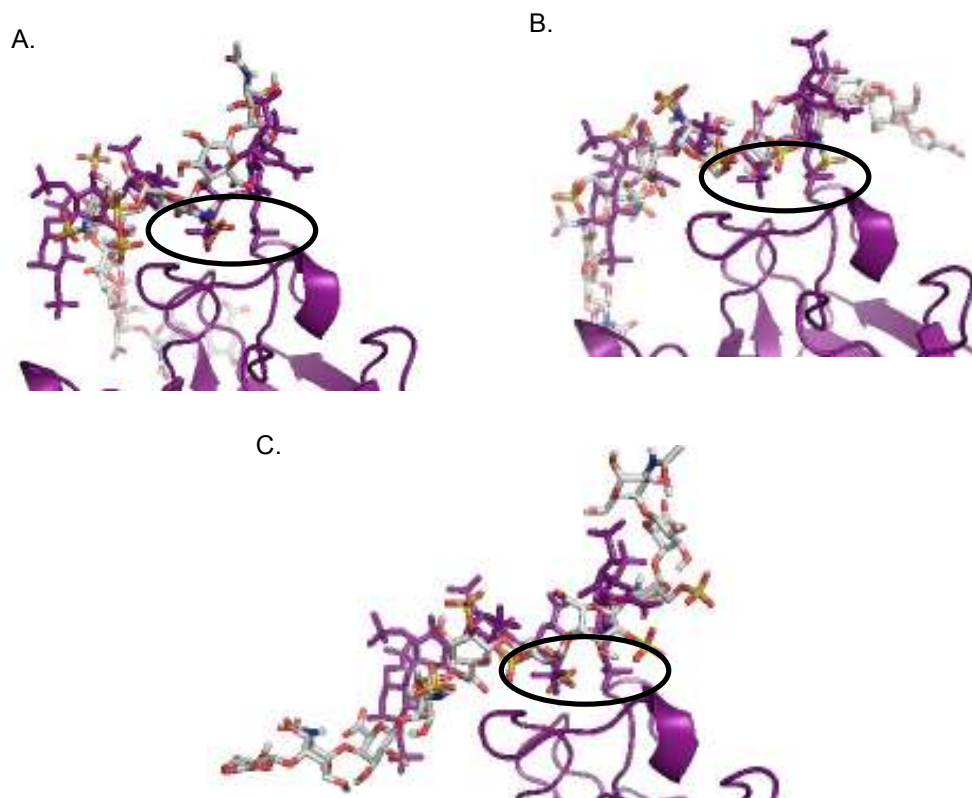


Figure 3.41. The highest affinity binding structure as predicted by Vina for **dp10(5)** with the best fit for the x-ray crystal structure for the 10 ns MD simulation.

Panel A. Frame 500, model 13.

Panel B. Frame 1050, model 6.

Panel C. Frame 6000, model 5.

Figure 3.42.

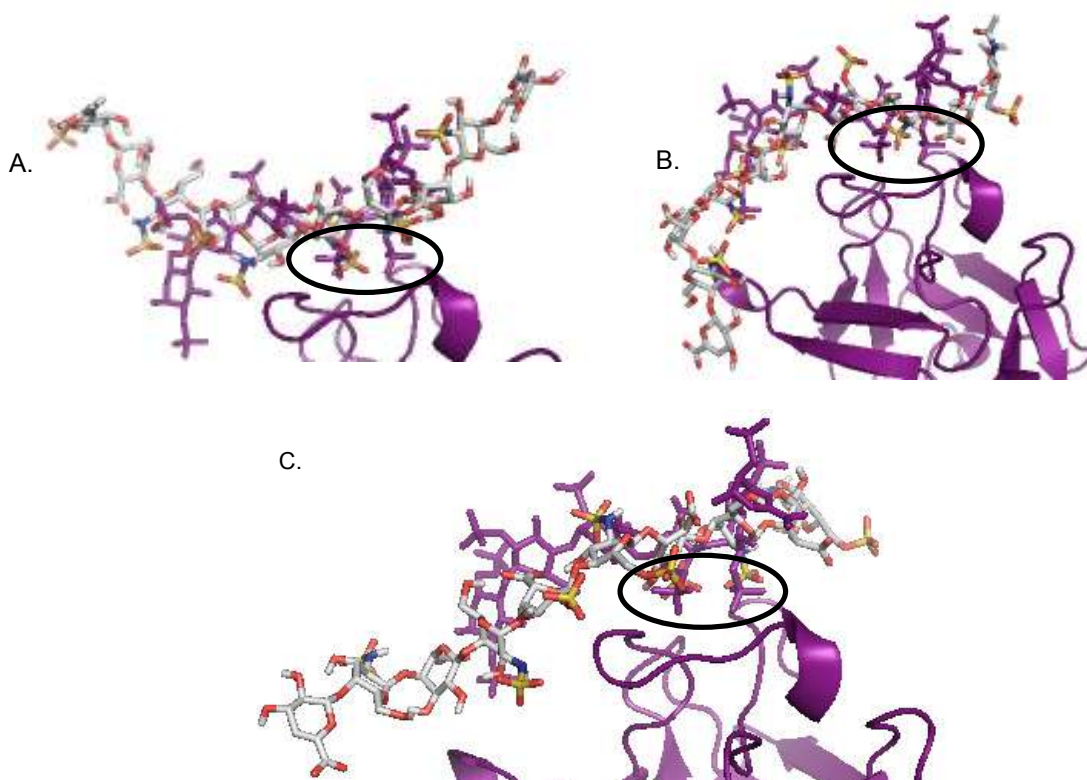


Figure 3.42. The highest affinity binding structure as predicted by Vina for **dp10(7)** with the best fit for the x-ray crystal structure for the 10 ns MD simulation.

Panel A. Frame 250, model 14.

Panel B. Frame 750, model 1.

Panel C. Frame 5000, model 8.

Figure 3.43

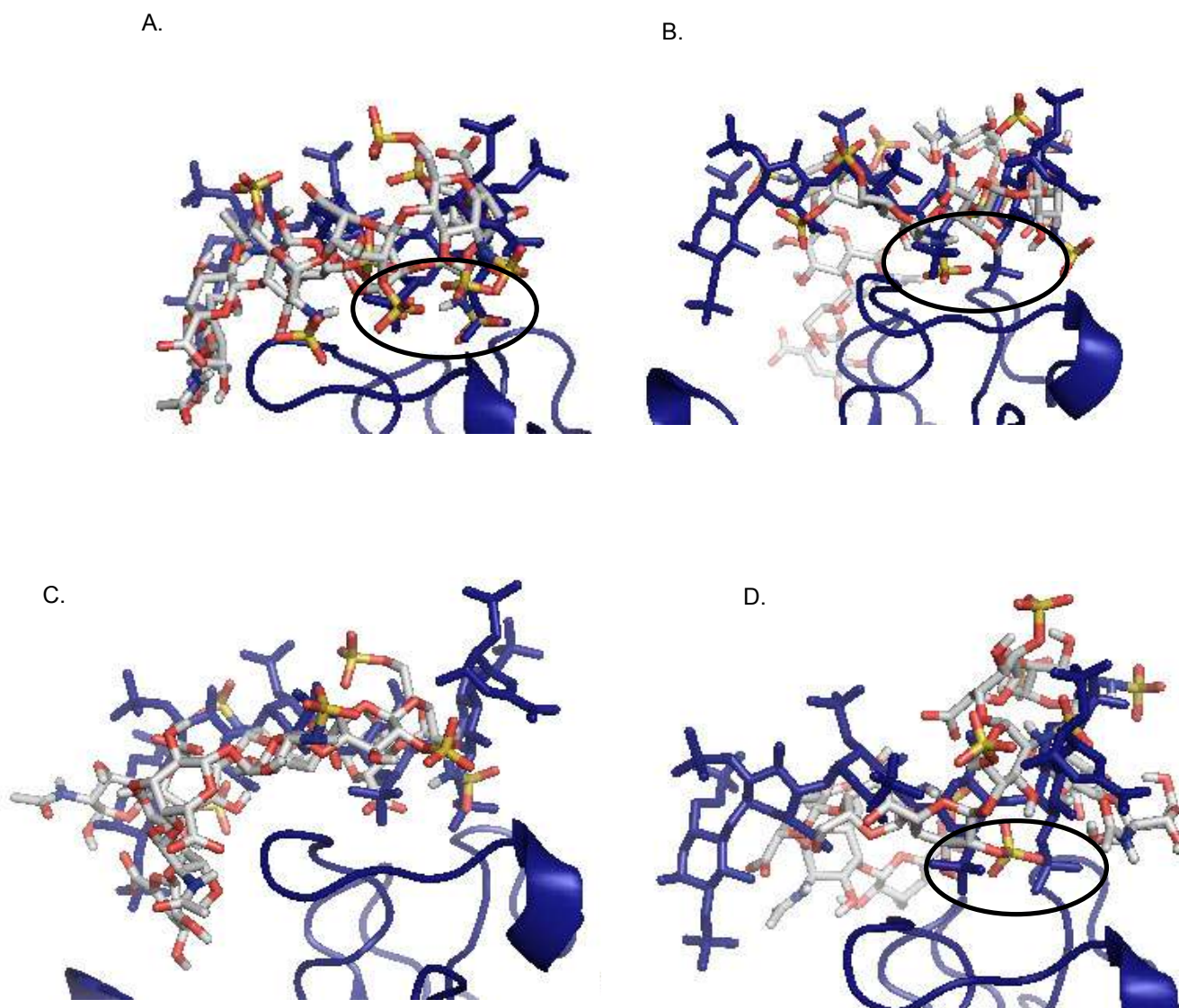


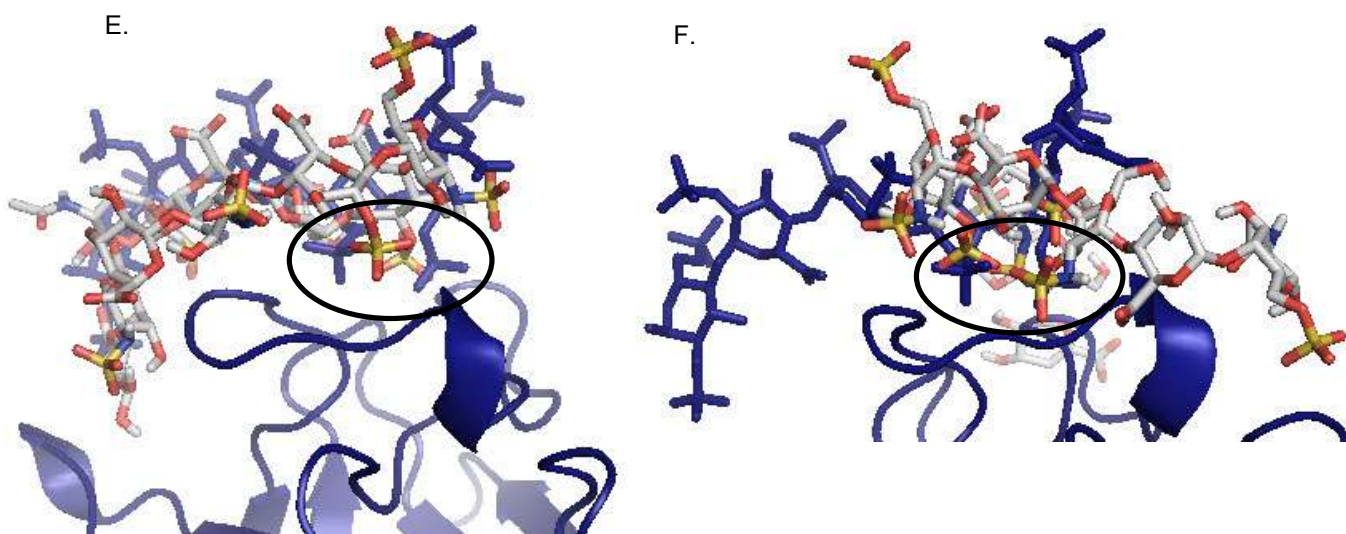
Figure 3.43 The highest affinity binding structure as predicted by Vina for **dp10(1)**, **dp10(3)**, **dp10(5)** and **dp10(7)** with the best fit for the x-ray crystal structure for the 1 ns MD simulation flexible dockings.

Panel A. Decasaccharide 1, Frame 45, model 1.

Panel B. Decasaccharide 1, Frame 45, model 14.

Panel C. Decasaccharide 3, Frame 1300, model 1.

Panel D. Decasaccharide 5, Frame 1381, model 20.



Panel E. Decasaccharide 7, Frame 50, model 11.

Panel F. Decasaccharide 7, Frame 200, model 14.

Figure 3.44.

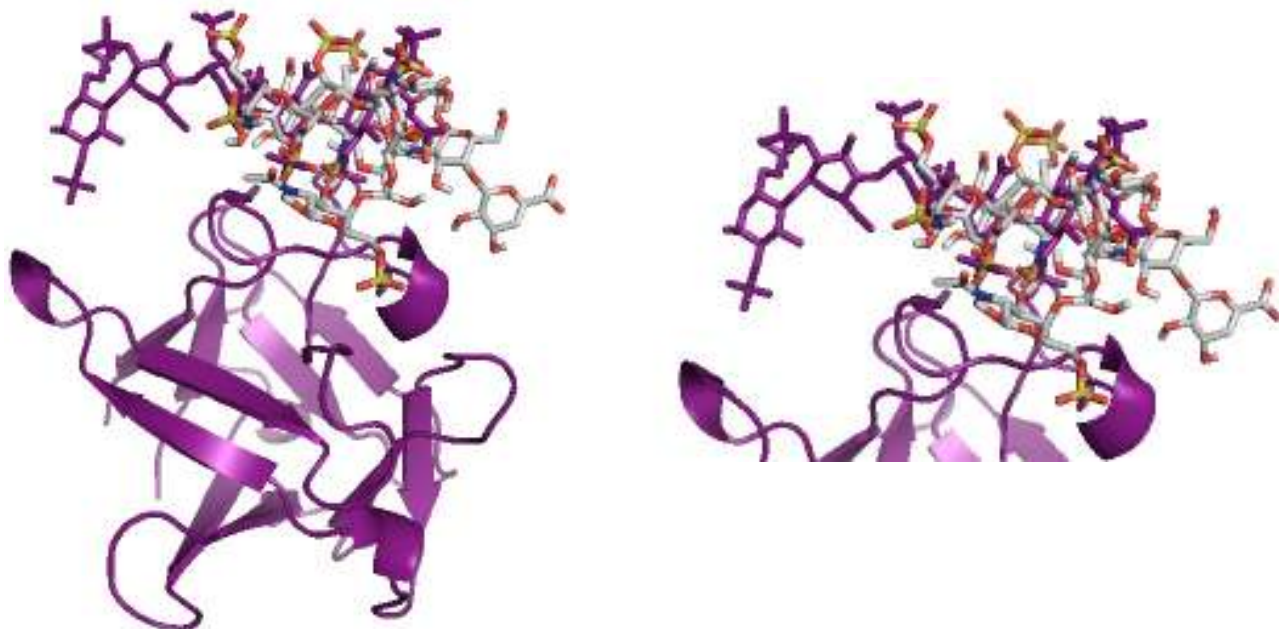


Figure 3.44. The highest affinity binding structure as predicted by Vina for **dp10(1)** with the best fit for the x-ray crystal structure for the 1 ns MD simulation flexible dockings.

Panel A. Decasaccharide 1, Frame 200, model 12.