Spring 4-2021

Toward Improving Understanding of the Structure and Biophysics of Glycosaminoglycans

Elizabeth K. Whitmore
University of New England, elizabeth.k.whitmore@maine.edu

Recommended Citation

Whitmore, Elizabeth K., "Toward Improving Understanding of the Structure and Biophysics of Glycosaminoglycans" (2021). *Electronic Theses and Dissertations*. 3348.
https://digitalcommons.library.umaine.edu/etd/3348

TOWARD IMPROVING UNDERSTANDING OF THE STRUCTURE AND BIOPHYSICS OF GLYCOSAMINOGLYCANS

By

Elizabeth K. Whitmore

B.S. University of New England, 2014

A DISSERTATION

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

(in Biomedical Sciences)

The Graduate School

The University of Maine

May 2021

Advisory Committee:

Olgun Guvench, Professor of Pharmaceutical Science, University of New England, Advisor

Dustin Updike, Professor of Biomedical Science, Mount Desert Island Biological Laboratory

Sergey Ryzhov, Professor of Biomedical Science, Maine Medical Center Research Institute

Lucy Liaw, Professor of Biomedical Science, Maine Medical Center Research Institute

Ronald Hills, Professor of Pharmaceutical Science, University of New England

TOWARD IMPROVING UNDERSTANDING OF THE STRUCTURE AND BIOPHYSICS OF GLYCOSAMINOGLYCANS

By Elizabeth K. Whitmore

Thesis Advisor: Dr. Olgun Guvench

An Abstract of the Dissertation Presented
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
(in Biomedical Sciences)
May 2021

Glycosaminoglycans (GAGs) are the linear carbohydrate components of proteoglycans (PGs) that mediate PG bioactivities, including signal transduction, tissue morphogenesis, and matrix assembly. To understand GAG function, it is important to understand GAG structure and biophysics at atomic resolution. This is a challenge for existing experimental and computational methods because GAGs are heterogeneous, conformationally complex, and polydisperse, containing up to 200 monosaccharides. Molecular dynamics (MD) simulations come close to overcoming this challenge but are only feasible for short GAG polymers. To address this problem, we developed an algorithm that applies conformations from unbiased all-atom explicit-solvent MD simulations of short GAG polymers to rapidly construct 3-D atomic-resolution models of GAGs of arbitrary length.

MD simulations of GAG 10-mers (i.e., polymers containing 10 monosaccharides) and 20-mers were run and conformations of all monosaccharide rings and glycosidic linkages were analyzed and compared to existing experimental data. These analyses demonstrated that (1) MD-generated GAG conformations are in agreement with existing experimental data; (2) MD-generated GAG 10-mer ring and linkage conformations match those in corresponding GAG 20-mers, suggesting that these conformations are representative of those in longer GAG biopolymers; and (3) rings and linkages in GAG 10- and 20-mers behave randomly and independently in MD simulation. Together, these findings indicate that MD-

generated GAG 20-mer ring and linkage conformations can be used to construct thermodynamically-correct models of GAG polymers. Indeed, our findings demonstrate that our algorithm constructs GAG 10- and 20-mer conformational ensembles that accurately represent the backbone flexibility seen in MD simulations. Furthermore, within a day, our algorithm constructs conformational ensembles of GAG 200-mers that we would reasonably expect from MD simulation, demonstrating the efficiency of the algorithm and reduction in its time and computational cost compared to simulation.

While there are other programs that can quickly construct atomic-resolution models of GAGs, those programs use conformations from short GAG subunits in solid state. Our findings suggest that GAG 20-mers are more flexible than short GAG subunits, meaning our program constructs ensembles that more accurately represent GAG polymer backbone flexibility and provide valuable insights toward improving the understanding of the structure and biophysics of GAGs.

ACKNOWLEDGEMENTS

There are many people I want to thank for their help during my graduate career. First, I would like to thank my advisor Dr. Olgun Guvench for giving me a great education, supporting my travels to conferences to further my academic and professional development, and helping to prepare me for post-graduate life and achieve my professional goals. I want to thank my committee members Dr. Dustin Updike, Dr. Sergey Ryzhov, Dr. Lucy Liaw, and Dr. Ronald Hills for all their encouragement and support and for helping me strengthen my skills in communicating scientific information. Special thanks to my rotation advisors Dr. Dustin Updike, Dr. Sergey Ryzhov, and Dr. Douglas Sawyer for their guidance and attentiveness throughout my rotations in their labs.

I appreciate all of the GSBSE faculty and staff who helped advance my education. I would especially like to thank Dr. Clarissa Henry and Dr. David Neivandt, who served as Program Directors during my time at GSBSE, for their incredible strides toward excellence in academics and research for the GSBSE program and improving student life. Thanks to Tammy Crosby and Zhen Zhang, who served as administrative assistants during my time at GSBSE, for devoting so much time and effort to quickly answering my questions, resolving any issues, and making my life easier.

I want to thank my colleagues in the Guvench Lab, especially the co-authors of our publications Gabriel Vesenka, Hanna Sihler, and Devon Martin for their contributions to the work presented in this dissertation.

Thanks to my family and friends, including my fellow GSBSE students, for all their support throughout this entire process.

Finally, I would like to thank the National Science Foundation (grant number MCB-1453529 to O.G.) for funding this research.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

LIST OF TABLES viii

LIST OF FIGURES x

LIST OF ABBREVIATIONS xviii

Chapters

1. INTRODUCTION 1
    - 1.1. Proteoglycans 1
    - 1.2. Glycosaminoglycans 1
        - 1.2.1. Biosynthesis 2
        - 1.2.2. Binding Properties 2
        - 1.2.3. GAG Types 3
            - 1.2.3.1. Chondroitin Sulfate 3
            - 1.2.3.2. Hyaluronan 4
            - 1.2.3.3. Dermatan Sulfate 4
            - 1.2.3.4. Keratan Sulfate 5
            - 1.2.3.5. Heparan Sulfate 5
        - 1.2.4. GAGs in Pathology and Therapeutics 6
        - 1.2.5. Gap in Knowledge 7

2. METHODS 9
    - 2.1. Laboratory Methods 9
    - 2.2. Computational Methods 10
        - 2.2.1. Molecular Dynamics Simulations 10
            - 2.2.1.1. Force Field 11

2.2.1.2. Periodic Boundary Conditions 14

2.2.1.3. All-Atom Explicit-Solvent MD 15

2.2.1.4. Challenges with All-Atom Explicit-Solvent MD 15

2.2.2. Conformational Analysis 16

2.2.2.1. Glycosidic Linkage Dihedral Free Energy Changes $\Delta G(\phi, \psi)$ 16

2.2.2.2. Monosaccharide Ring Cremer-Pople Parameters ϕ, θ, Q 18

2.2.3. Algorithm to Construct 3-D Models of Glycosaminoglycans 19

3. EFFICIENT CONSTRUCTION OF ATOMIC-RESOLUTION MODELS OF NONSULFATED CHONDROITIN GLYCOSAMINOGLYCAN USING MOLECULAR DYNAMICS DATA 21

3.1. Introduction 21

3.2. Materials and Methods 26

3.2.1. Molecular Dynamics 26

3.2.1.1. System Construction 26

3.2.1.2. Energy Minimization and Heating 26

3.2.1.3. Production Simulations 27

3.2.1.4. Conformational Analysis 28

3.2.2. Construction Algorithm to Generate GAG Conformational Ensembles 29

3.3. Results and Discussion 32

3.3.1. Glycosidic Linkage Geometries 32

3.3.2. GlcA Ring Pucker Effects 36

3.3.3. Treating Glycosidic Linkage and Ring Pucker Geometries as Independent Variables 42

3.3.4. Handling Nonphysical Constructed Conformations 43

3.3.5. Internal Validation on 10-mers 45

3.3.6. Application to Longer Chondroitin Polymers 47

3.4. Conclusions 50

4. CONSTRUCTING 3-DIMENSIONAL ATOMIC-RESOLUTION MODELS OF NONSULFATED GLYCOSAMINOGLYCANS WITH ARBITRARY LENGTHS USING CONFORMATIONS FROM MOLECULAR DYNAMICS 52

4.1. Introduction 52

4.2. Materials and Methods 57

4.2.1. Molecular Dynamics: System Construction, Energy Minimization, Heating, and Production Simulations 57

4.2.2. Conformational Analysis 58

4.2.3. Construction Algorithm to Generate GAG Conformational Ensembles 60

4.3. Results and Discussion 61

4.3.1. Hyaluronan 61

4.3.1.1. Molecular Dynamics Simulations: Glycosidic Linkage and Monosaccharide Ring Geometry Effects on Polymer Backbone Flexibility 61

4.3.1.2. Construction Algorithm 71

4.3.2. Nonsulfated Dermatan 74

4.3.2.1. Molecular Dynamics Simulations: Glycosidic Linkage and Monosaccharide Ring Geometry Effects on Polymer Backbone Flexibility 74

4.3.2.2. Construction Algorithm 81

4.3.3. Nonsulfated Keratan 84

4.3.3.1. Molecular Dynamics Simulations: Glycosidic Linkage and Monosaccharide Ring Geometry Effects on Polymer Backbone Flexibility 84

4.3.3.2. Construction Algorithm 90

4.3.4. Nonsulfated Heparan 92

4.3.4.1. Molecular Dynamics Simulations: Glycosidic Linkage and Monosaccharide Ring Geometry Effects on Polymer Backbone Flexibility 92

4.3.4.2. Construction Algorithm 100

4.4. Conclusions 102

REFERENCES 108

APPENDIX A: SUPPORTING INFORMATION FOR "EFFICIENT CONSTRUCTION OF ATOMIC-RESOLUTION MODELS OF NONSULFATED CHONDROITIN GLYCOSAMINOGLYCAN USING MOLECULAR DYNAMICS DATA" 132

APPENDIX B: SUPPORTING INFORMATION FOR "CONSTRUCTING 3-DIMENSIONAL ATOMIC-RESOLUTION MODELS OF NONSULFATED GLYCOSAMINOGLYCANS WITH ARBITRARY LENGTHS USING CONFORMATIONS FROM MOLECULAR DYNAMICS" 146

BIOGRAPHY OF THE AUTHOR 187

LIST OF TABLES

Table 3.1. Glycosidic Linkage Dihedrals (ϕ, ψ) and Free Energy ($\Delta G(\phi, \psi)$ kcal/mol) Minima 35

Table 3.2. Most Probable End-to-End Distances (d) in MD-Generated and Constructed Ensembles 47

Table 4.1. Most Probable End-to-End Distances (d) in MD-Generated and Constructed Hyaluronan Ensembles 62

Table 4.2. Hyaluronan Glycosidic Linkage Dihedrals (ϕ, ψ) and Free Energy ($\Delta G(\phi, \psi)$ kcal/mol) Minima 68

Table 4.3. Most Probable End-to-End Distances (d) in MD-Generated and Constructed Nonsulfated Dermatan Ensembles 75

Table 4.4. Nonsulfated Dermatan Glycosidic Linkage Dihedrals (ϕ, ψ) and Free Energy ($\Delta G(\phi, \psi)$ kcal/mol) Minima 81

Table 4.5. Most Probable End-to-End Distances (d) in MD-Generated and Constructed Nonsulfated Keratan Ensembles 85

Table 4.6. Nonsulfated Keratan Glycosidic Linkage Dihedrals (ϕ, ψ) and Free Energy ($\Delta G(\phi, \psi)$ kcal/mol) Minima 88

Table 4.7. Most Probable End-to-End Distances (d) in MD-Generated and Constructed Nonsulfated Heparan Ensembles 93

Table 4.8. Nonsulfated Heparan Glycosidic Linkage Dihedrals (ϕ, ψ) and Free Energy ($\Delta G(\phi, \psi)$ kcal/mol) Minima 97

Table A.1. Comparison to Observed Literature Values of Glycosidic Linkage Dihedrals (ϕ, ψ) in Nonsulfated Chondroitin 132

Table A.2. Bond Energies of Constructed Chondroitin 20-mer Conformations with Pierced Rings 132

Table B.1. Most Probable End-to-End Distances (*d*) in MD-Generated 20-mer Conformations with Glycosidic Linkage Conformations in Secondary Basins 146

Table B.2. Percent of Occurrences of Different IdoA Ring Puckers in Nonsulfated Dermatan MD Simulations 146

Table B.3. Percent of Occurrences of Different IdoA Ring Puckers in Nonsulfated Heparan MD Simulations 147

LIST OF FIGURES

Figure 2.1. Bond Distortion Energy as a Function of Bond Length 12

Figure 2.2. Snapshots from a MD Trajectory of a GAG 10-mer 13

Figure 2.3. Snapshots from a MD Trajectory of a GAG 10-mer 14

Figure 2.4. GAG Trisaccharide Showing ϕ and ψ Dihedral Rotations........ 17

Figure 2.5. Spherical Map of C-P parameters ϕ, θ, and Q and Atomic-Resolution Molecular Graphic of a Pyranose Ring 19

Figure 3.1. Molecular Graphic of a Compact Nonsulfated Chondroitin 20-mer Conformation........ 23

Figure 3.2. Molecular Graphic of a Nonsulfated Chondroitin Trisaccharide with Linkage Dihedrals and Ring and Linkage Atoms Labeled 28

Figure 3.3. $\Delta G(\phi, \psi)$ Plots for MD-Generated and Constructed (Before and After Energy Minimization) Nonsulfated Chondroitin 20-mer Ensembles 34

Figure 3.4. C-P Parameter (ϕ, θ) Plots for Monosaccharides in MD-Generated and Constructed (Before and After Energy Minimization) Nonsulfated Chondroitin 20-mer Ensembles........ 38

Figure 3.5. Molecular Graphics Showing Different Monosaccharide Ring and Glycosidic Linkage Conformations in MD-Generated Nonsulfated Chondroitin 20-mer Ensemble........ 40

Figure 3.6. End-to-End Distance Probability Distributions of MD-Generated and Constructed Nonsulfated Chondroitin 20-mer Ensembles........ 42

Figure 3.7. $\Delta G(\phi, \psi)$ Plots for Glycosidic Linkages Flanking Non-$^4C_1$ GlcA Conformers in MD-Generated Nonsulfated Chondroitin 20-mer Ensemble........ 43

Figure 3.8. Molecular Graphic of a Constructed Nonsulfated Chondroitin 20-mer Conformation with a Pierced Ring 44

Figure 3.9. $\Delta G(\phi, \psi)$ Plots for MD-Generated Nonsulfated Chondroitin 10-mer Ensemble 45

Figure 3.10. End-to-End Distance Probability Distributions of MD-Generated and Constructed Nonsulfated Chondroitin 10-mer Ensembles 46

Figure 3.11. End-to-End Distance Probability Distribution of Constructed Nonsulfated Chondroitin 100-mer Ensemble and Molecular Graphics of Constructed 100-mer Conformations 48

Figure 3.12. End-to-End Distance Probability Distribution of Constructed Nonsulfated Chondroitin 200-mer Ensemble and Molecular Graphics of Constructed 200-mer Conformations 49

Figure 4.1. Molecular Graphics of Hyaluronan and Nonsulfated Dermatan, Keratan, and Heparan Trisaccharides with Linkage Dihedrals and Ring and Linkage Atoms Labeled 59

Figure 4.2. End-to-End Distance Probability Distribution of MD-Generated Hyaluronan 20-mer Ensemble 62

Figure 4.3. C-P Parameter (ϕ, θ) Plots for Monosaccharides in MD-Generated Hyaluronan 10- and 20-mer Ensembles 63

Figure 4.4. Molecular Graphics Showing Different Monosaccharide Ring Conformations in MD-Generated Hyaluronan 20-mer Ensemble 65

Figure 4.5. $\Delta G(\phi, \psi)$ Plots for MD-Generated Hyaluronan 10- and 20-mer Ensembles 68

Figure 4.6. End-to-End Distance Probability Distributions of MD-Generated and Constructed Hyaluronan 10- and 20-mer Ensembles 72

Figure 4.7. End-to-End Distance Probability Distribution of Constructed Hyaluronan 200-mer Ensemble 74

Figure 4.8. End-to-End Distance Probability Distribution of MD-Generated Nonsulfated Dermatan 20-mer Ensemble 75

Figure 4.9. C-P Parameter (ϕ, θ) Plots for Monosaccharides in MD-Generated Nonsulfated Dermatan 10- and 20-mer Ensembles 76

Figure 4.10. Molecular Graphics Showing Different Monosaccharide Ring Conformations in MD-Generated Nonsulfated Dermatan 20-mer Ensemble 79

Figure 4.11. $\Delta G(\phi, \psi)$ Plots for MD-Generated Nonsulfated Dermatan 10- and 20-mer Ensembles 80

Figure 4.12. End-to-End Distance Probability Distributions of MD-Generated and Constructed Nonsulfated Dermatan 10- and 20-mer Ensembles 82

Figure 4.13. End-to-End Distance Probability Distribution of Constructed Nonsulfated Dermatan 200-mer Ensemble 83

Figure 4.14. End-to-End Distance Probability Distribution of MD-Generated Nonsulfated Keratan 20-mer Ensemble 84

Figure 4.15. C-P Parameter (ϕ, θ) Plots for Monosaccharides in MD-Generated Nonsulfated Keratan 10- and 20-mer Ensembles 86

Figure 4.16. $\Delta G(\phi, \psi)$ Plots for MD-Generated Nonsulfated Keratan 10- and 20-mer Ensembles 87

Figure 4.17. End-to-End Distance Probability Distributions of MD-Generated and Constructed Nonsulfated Keratan 10- and 20-mer Ensembles 91

Figure 4.18. End-to-End Distance Probability Distribution of Constructed Nonsulfated Keratan 200-mer Ensemble 92

Figure 4.19. End-to-End Distance Probability Distribution of MD-Generated Nonsulfated Heparan 20-mer Ensemble 93

Figure 4.20. C-P Parameter (ϕ, θ) Plots for Monosaccharides in MD-Generated Nonsulfated Heparan 10- and 20-mer Ensembles 95

Figure 4.21. ΔG(ϕ, ψ) Plots for MD-Generated Nonsulfated Heparan 10- and 20-mer Ensembles........... 97

Figure 4.22. End-to-End Distance Probability Distributions of MD-Generated and Constructed Nonsulfated Heparan 10- and 20-mer Ensembles 101

Figure 4.23. End-to-End Distance Probability Distribution of Constructed Nonsulfated Heparan 200-mer Ensemble 102

Figure 4.24. Molecular Graphics of Constructed Hyaluronan and Nonsulfated Dermatan, Keratan, and Heparan 200-mer Conformations 107

Figure A.1. End-to-End Distance Probability Distribution of Nonsulfated Chondroitin 20-mer Ensembles Generated by MD and an Early Version of the Construction Algorithm Which Applied Glycosidic Linkage Geometries from MD-Generated 20-mer Ensembles 133

Figure A.2. Scatterplots of Radius of Gyration as a Function of End-to-End Distance in MD-Generated and Constructed Nonsulfated Chondroitin Ensembles 134

Figure A.3. ΔG(ϕ, ψ) Plots for Each Glycosidic Linkage in MD-Generated Nonsulfated Chondroitin 20-mer Ensembles 135

Figure A.4. C-P Plots for Each Monosaccharide Ring in MD-Generated Nonsulfated Chondroitin 20-mer Ensembles 137

Figure A.5. Probability Histograms of Bond Lengths for Each Type of Bond in MD-Generated and Constructed Nonsulfated Chondroitin 20-mer Ensembles 140

Figure A.6. Probability Histogram Showing Changes in Glycosidic Linkage Dihedral Angles During Energy Minimization in Constructed Nonsulfated Chondroitin 20-mer Ensembles 141

Figure A.7. C-P Plots for Monosaccharide Rings in MD-Generated Nonsulfated Chondroitin 10-mer Ensembles 141

Figure A.8. End-to-End Distance Probability Distribution of Nonsulfated Chondroitin 20-mer Ensembles Generated by MD and an Early Version of the Construction Algorithm Which Applied Glycosidic Linkage Geometries from Nonsulfated Chondroitin Disaccharide MD-Generated Ensembles 142

Figure A.9. $\Delta G(\phi, \psi)$ Plots for Glycosidic Linkages in MD-Generated Nonsulfated Chondroitin Disaccharide Ensembles and 20-mer Ensembles Constructed Using Glycosidic Linkage Dihedral Probabilities from These MD-Generated Disaccharide Ensembles 143

Figure A.10. Bond Energy Distribution Probability Histograms from Constructed Nonsulfated Chondroitin 10-, 20-, 100-, and 200-mer Ensembles 144

Figure A.11. Average Bond Energies as a Function of Nonsulfated Chondroitin Polymer Length 145

Figure B.1. Molecular Graphic of Constructed Hyaluronan 20-mer Conformation with a Pierced Ring 147

Figure B.2. Scatterplots of Radius of Gyration as a Function of End-to-End Distance of MD-Generated and Constructed Hyaluronan Ensembles 148

Figure B.3. System Potential Energy Probability Distribution of MD-Generated Hyaluronan 20-mer Ensemble 149

Figure B.4. C-P Plots and C-P Parameter θ Timeseries for Each GlcNAc Monosaccharide Ring in MD-Generated Hyaluronan 20-mer Ensemble 150

Figure B.5. C-P Plots and C-P Parameter θ Timeseries for Each GlcA Monosaccharide Ring in MD-Generated Hyaluronan 20-mer Ensemble 153

Figure B.6. Scatterplot of Radius of Gyration as a Function of End-to-End Distance of MD-Generated Hyaluronan 20-mer Conformations with Non-${}^4C_1$ Ring Puckers 155

Figure B.7. $\Delta G(\phi, \psi)$ Plots for Each Glycosidic Linkage in MD-Generated Hyaluronan 20-mer Ensembles 156

Figure B.8. MD Snapshots of Hyaluronan 20-mer GlcAβ1-3GlcNAc Glycosidic Linkages with Dihedrals in Different Basins and End-to-End Distance Distributions of 20-mer Conformations with Linkages in Secondary Basin 158

Figure B.9. Δ$G(\phi, \psi)$ Plots for Glycosidic Linkages Flanking Non-$^4C_1$ Ring Puckers in MD-Generated Hyaluronan 20-mer Ensemble 159

Figure B.10. MD Snapshots of Hyaluronan 20-mer GlcNAcβ1-4GlcA Glycosidic Linkages with Dihedrals in Different Basins and End-to-End Distance Distributions of 20-mer Conformations with Linkages in Secondary Basin 160

Figure B.11. C-P Plots for Monosaccharides in Constructed Hyaluronan 20-mer Ensemble 161

Figure B.12. Δ$G(\phi, \psi)$ for Aggregated Glycosidic Linkage Data from Constructed Hyaluronan 20-mer Ensemble 161

Figure B.13. Scatterplots of Radius of Gyration as a Function of End-to-End Distance of MD-Generated and Constructed Nonsulfated Dermatan Ensembles 162

Figure B.14. C-P Parameter θ Timeseries for Each GalNAc Monosaccharide Ring in MD-Generated Nonsulfated Dermatan 20-mer Ensemble 163

Figure B.15. End-to-End Distance Distributions of MD-Generated Nonsulfated Dermatan 20-mer Conformations with Boat/Skew-Boat Ring Puckers that Cause a Kink in the Polymer Chain and ^2S_O Conformations 164

Figure B.16. C-P Plots and C-P Parameter θ Timeseries for Each IdoA Monosaccharide Ring in MD-Generated Nonsulfated Dermatan 20-mer Ensemble 165

Figure B.17. C-P Plots and C-P Parameter θ Timeseries for Each IdoA Monosaccharide Ring in MD-Generated Nonsulfated Dermatan 10-mer Ensemble 168

Figure B.18. C-P Plots for Monosaccharides in Constructed Nonsulfated Dermatan 20-mer Ensemble 169

Figure B.19. $\Delta G(\phi, \psi)$ for Aggregated Glycosidic Linkage Data from Constructed Nonsulfated Dermatan 20-mer Ensemble 169

Figure B.20. Scatterplots of Radius of Gyration as a Function of End-to-End Distance of MD-Generated and Constructed Nonsulfated Keratan Ensembles 170

Figure B.21. C-P Plots and C-P Parameter θ Timeseries for Each GlcNAc Monosaccharide Ring in MD-Generated Nonsulfated Keratan 20-mer Ensemble 171

Figure B.22. Scatterplot of Radius of Gyration as a Function of End-to-End Distance in MD-Generated Nonsulfated Keratan 20-mer Conformations with Non-$^4C_1$ Ring Puckers 173

Figure B.23. MD Snapshots of Nonsulfated Keratan 20-mer Galβ1-4GlcNAc Glycosidic Linkages with Dihedrals in Different Basins and End-to-End Distance Distributions of 20-mer Conformations with Linkages in Secondary and Tertiary Basins 174

Figure B.24. MD Snapshots of Nonsulfated Keratan 20-mer GlcNAcβ1-3Gal Glycosidic Linkages with Dihedrals in Different Basins and End-to-End Distance Distributions of 20-mer Conformations with Linkages in Secondary and Tertiary Basins 175

Figure B.25. C-P Plots and C-P Parameter θ Timeseries for Each Gal Monosaccharide Ring in MD-Generated Nonsulfated Keratan 10-mer Ensemble 176

Figure B.26. C-P Plots for GlcNAc and Gal in Constructed Nonsulfated Keratan 20-mer Ensemble 177

Figure B.27. $\Delta G(\phi, \psi)$ Plots for Aggregated Glycosidic Linkage Data from Constructed Nonsulfated Keratan 20-mer Ensemble 177

Figure B.28. Scatterplots of Radius of Gyration as a Function of End-to-End Distance of MD-Generated and Constructed Nonsulfated Heparan Ensembles 178

Figure B.29. C-P Parameter θ Timeseries for Each IdoA Monosaccharide Ring in MD-Generated Nonsulfated Heparan 20-mer Ensemble 179

Figure B.30. End-to-End Distance Distributions of MD-Generated Nonsulfated Heparan 20-mer Conformations with Boat/Skew-Boat Ring Puckers that Cause a Kink in the Polymer Chain and ^2S_O Conformations 180

Figure B.31. MD Snapshot of Nonsulfated Heparan 20-mer Highlighting GlcNAcα1-4IdoA Linkages 181

Figure B.32. End-to-End Distance Distributions of MD-Generated Nonsulfated Heparan 20-mer Conformations with Linkage Dihedrals in Secondary Basins 181

Figure B.33. $\Delta G(\phi, \psi)$ Plots for IdoAα1-4GlcNAc Linkages Flanking Different IdoA Conformations in MD-Generated Nonsulfated Heparan 20-mer Ensemble 182

Figure B.34. C-P Plots for Monosaccharides in Constructed Nonsulfated Heparan 20-mer Ensemble 182

Figure B.35. $\Delta G(\phi, \psi)$ Plots for Aggregated Glycosidic Linkage Data from Constructed Nonsulfated Heparan 20-mer Ensemble 183

Figure B.36. Bond Potential Energy Probability Distributions from Constructed Hyaluronan Ensembles 183

Figure B.37. Bond Potential Energy Probability Distributions from Constructed Nonsulfated Dermatan Ensembles 184

Figure B.38. Bond Potential Energy Probability Distributions from Constructed Nonsulfated Keratan Ensembles 185

Figure B.39. Bond Potential Energy Probability Distributions from Constructed Nonsulfated Heparan Ensembles 186

LIST OF ABBREVIATIONS

PG	Proteoglycan
GAG	Glycosaminoglycan
ECM	Extracellular Matrix
MD	Molecular Dynamics
PBC	Periodic Boundary Condition
C-P	Cremer-Pople
GalNAc	N-Acetylgalactosamine
GlcNAc	N-Acetylglucosamine
GlcA	Glucuronate/Glucuronic Acid
IdoA	Iduronate/Iduronic Acid
UA	Uronic Acid
Gal	Galactose

CHAPTER 1

INTRODUCTION

1.1 Proteoglycans

Proteoglycans (PGs) are fundamental components of tissue structures and regulators of signaling pathways in animals. PGs are a diverse group of protein-carbohydrate conjugates strategically positioned in the extracellular matrix (ECM), on cell surfaces, and inside cells to serve structural and signaling roles at the cell-tissue interface and within the cell [1-12]. Specifically, they are found in fibroblasts, chondrocytes [5-9], immune cells [9], mast cells [9,13], several hematopoietic cells [9,13], endothelial cells [9,14,15], embryonic stem cells [9,16], axons [1-5,11], and many other cell types. PGs are also found in cytoplasmic granules of mast cells, neutrophils, cytotoxic T lymphocytes, eosinophils, endocrine and neuroendocrine cells, and secretory cells in epithelia where they activate other proteins, aid in protein transport after release, mediate storage of other granule components, organize and concentrate other proteins in granules, and protect other proteins from deactivation or destruction [9]. PGs also act as receptors and bind growth factors [1-5,11,17-22] and growth factor-binding proteins [5], enzymes and enzyme inhibitors [5,22], membrane receptors [22], and ECM molecules [5,9,22]. These bioactivities make PGs key modulators of various biological processes including cell growth, differentiation, proliferation, migration, adhesion, spreading, guidance, and development [1-5,11,23,24] as well as homeostasis [22], signal-transduction [9,25], tissue morphogenesis [1-5], chemoattraction [5], matrix assembly [5], lipoprotein uptake [5], and extracellular proteolysis [5,22]. PGs also aid in delivery of certain molecules to the nucleus [9,26], bind to molecules inside the nucleus to facilitate their packaging [9,27], and may influence gene expression [9,28,29].

1.2. Glycosaminoglycans

Glycosylation of PG core proteins plays a key functional role in development and physiological processes. This is because many PG bioactivities are mediated by their covalently linked glycosaminoglycan (GAG) chains, which typically account for more than half the molecular weight of PGs [6,30]. GAGs are linear carbohydrate polymers consisting of variably sulfated repeating disaccharide units which typically contain a hexosamine, i.e., N-acetylglucosamine (GlcNAc) or N-acetylgalactosamine (GalNAc), and an uronic acid (UA), i.e., glucuronic acid (GlcA) or iduronic acid (IdoA).

1.2.1. Biosynthesis

GAG biosynthesis is a multistep process that, for most GAGs, is initiated by glycosylation of a PG core protein serine residue by a tetrasaccharide linker [-GlcA β1-3 galactose β1-3 galactose β1-4 xylose β1-O-serine], a stepwise process which starts in the endoplasmic reticulum (ER) and ends in the Golgi [6,31,32]. Next, alternating hexosamine and UA residues are added successively in the Golgi. During polymerization, individual monosaccharides can be modified. For example, a given GlcA may be epimerized to IdoA and any GlcA, IdoA, or hexosamine may be variably sulfated [6,9,22,32-36]. This process is nontemplate based [37] and the length, UA composition, and sulfation pattern of the GAG chain are dependent upon the cell-specific polymerase, epimerase, and sulfotransferase enzymes, respectively. Once synthesized, fully-glycosylated PGs are transported to the cell surface and either remain embedded in the cell membrane or are released into the ECM [9]. The PG does not necessarily remain intact. ECM enzymes can cleave the protein, releasing the GAG-containing protein into the ECM, or the GAG chain itself, releasing functional free-floating GAGs [38].

1.2.2. Binding Properties

The negatively charged UA residues and sulfate groups give GAGs a high negative charge density that dictates GAG binding of ECM cations to regulate ion mobility [11,39] as well as collagen, elastin, and water to form gels that enable absorption of stress on joints and tissues [35,40]. The negative charge density of GAGs also contributes to formation of covalent and noncovalent complexes of PGs with other biomolecules to activate signaling pathways. GAG binding sites on proteins are determined by protein secondary structure, conformation, charge, and amino acid sequence [22,41]. GAG binding is type-specific [11,42,43] suggesting that GAG structural composition also determines binding affinity and specificity [22]. Additionally, GAG conformational flexibility allows binding to a wide range of biomolecules with varying affinity and specificity as well as multiple interactions with the same GAG chain [22]. Thus, GAG biophysics, and therefore function, is determined by GAG structure and conformational flexibility.

1.2.3. GAG Types

1.2.3.1. Chondroitin Sulfate. The GAG chondroitin sulfate (CS) is the most abundant GAG in the body [33,44-52]. CS consists of repeating units of [-4GlcAβ1-3GalNAcβ1-]$_n$ and can be variably sulfated. CS composition depends on cell type [11,53-56]. For example, nonsulfated chondroitin is necessary for morphogenesis and cell division in embryos of *Caenorhabditis elegans* (*C. elegans*) [2,57,58] and *Drosophila melanogaster* [59] suggesting a possible role in cytokine and growth factor signaling during embryonic development.

CS is found in several different cell types including natural killer cells [60], platelets [61], macrophages [62], leukocytes [63,64], and mast cells [9,12,65-69] and is the major GAG in most hematopoeitic cells [9]. CS is the most abundant GAG in the ECM of the central nervous system (CNS) [10,11] where it is important for formation, development, and maintenance of brain morphology and function [2,70-73] and might be a binding partner or co-receptor for neurotrophic factors [11,74-76]. For example, CS in the CNS interacts with the chemoattractive protein semaphorin 3A [77,78] to guide neuronal migration [79] and modulate synapse morphology and function [80], transmembrane glycoprotein amyloid precursor protein (APP) which is involved in neural growth and repair [11,81], collapsin response mediator protein-4 (CRMP-4) which induces growth cone collapse in the CNS [11,82-85], orthodenticle homeobox protein 2 (Otx2) which plays a role in postnatal development and controls plasticity [11,86], receptor protein tyrosine phosphatase (RPTP) which is involved in neural development [87-89], and neural cell adhesion molecule (N-CAM) which is a cell surface glycoprotein [11,72]. In addition, CS on certain PGs modulates signaling pathways. For example, NRP1 PG containing CS has enhanced ability to bind its co-receptor vascular endothelial growth factor (VEGF) compared to NRP1 core protein without a GAG chain, indicating that CS modulates VEGF signaling [90]. CS is also the most abundant GAG in cartilage ECM [33,51,52,91] where it binds collagen to modulate matrix assembly [47,92,93], inhibits chondrocyte apoptosis, cartilage degradation, and proinflammatory enzyme activation [94,95], and is involved in articular surface integrity [96,97], the calcification process [96,97], cartilage maturation [96-99], and modulation of cartilage growth plate morphogenesis [36].

1.2.3.2. Hyaluronan. Hyaluronan (a.k.a. hyaluronic acid; HA) consists of repeating units of $[\text{-}4\text{GlcA}\beta1\text{-}3\text{GlcNAc}\beta1\text{-}]_n$ and is always nonsulfated. HA biosynthesis is unique from that of other GAGs in that it takes place at the plasma membrane rather than the Golgi and does not involve a core protein [9]. Instead, HA synthases are synthesized in the ER then move through the Golgi to the plasma membrane where they are embedded and activated to initiate HA synthesis [100-102].

HA is anchored to cell surfaces, either by HA synthases or cell surface receptors like CD44 [103,104], where it is hydrated and binds PGs, forming the gel-like perineuronal nets in CNS cells [56,105] and pericellular matrices surrounding cells like chondrocytes [106], fibroblasts [107], and smooth muscle cells [108,109]. These pericellular matrices are necessary for ECM assembly and cell proliferation and migration. HA is also found intracellularly in brain cells [71,110], chondrocytes [111,112], hepatocytes [113-115], and keratinocytes [116-118]. It is thought that this intracellular localization of HA may result from receptor-mediated endocytosis to catabolize HA [111,112,117,118].

1.2.3.3. Dermatan Sulfate. Dermatan sulfate (DS) is structurally similar to CS in that it may contain disaccharide units $[\text{-}4\text{GlcA}\beta1\text{-}3\text{GalNAc}\beta1\text{-}]_n$ (i.e., the same sequence as CS) and can be variably sulfated. However, during synthesis of DS, some or all GlcA residues are epimerized to IdoA. Consequently, DS primarily consists of repeating units of $[\text{-}4\text{IdoA}\alpha1\text{-}3\text{GalNAc}\beta1\text{-}]_n$ but a single DS GAG may contain a mixture of GlcA and IdoA monosaccharides. This is sometimes referred to as a CS/DS hybrid chain which is found, for example, in the CNS, promoting neurite outgrowth [119-121]. DS fragments containing only IdoA as the UA monosaccharide are shown to control coagulation via binding to heparin cofactor II [122-131], collagen matrix assembly via binding to tenascin-X [132], and cellular proliferation and organogenesis via binding to hepatocyte growth factor/scatter factor (HGF/SF) [133,134].

1.2.3.4. Keratan Sulfate. Keratan sulfate (KS) is unique from other GAGs in that it contains galactose (Gal) in place of UA and it can be anchored to a PG core protein via GlcNAc–asparagine or GalNAc–serine/threonine linkages [9]. KS consists of repeating units of [-3Galβ1-4GlcNAcβ1-]$_n$ and can be variably sulfated. KS is found in ECM, on cell surfaces, and intracellularly in corneal [135-138], cartilage [91,139-141], and brain [73,142-146] cells where it regulates development and differentiation. For example, KS regulates corneal development by interacting with a wide variety of ECM, kinase, cytoskeletal, and neuroregulatory proteins in the cornea, including SLIT2 which is a nerve growth signaling protein, EFNA1 which is a tyrosine kinase receptor, nerve growth factor receptor (NGFR), and many others [138].

1.2.3.5. Heparan Sulfate. Heparan sulfate (HS) has the highest degree of sequence variability of all GAGs. Not only do UA composition (i.e., GlcA/IdoA) and sulfation pattern vary, but the glucosamine derivative can be either N-acetylated (i.e., GlcNAc), N-sulfated (i.e., GlcNS), or unsubstituted (GlcN). HS primarily consists of highly sulfated "S-domains" containing mostly 2-O-sulfated IdoA and variably O-sulfated GlcNS ([-4IdoA(2S)α1-4GlcNSα1-]$_n$), low sulfated domains containing nonsulfated UA (primarily IdoA but sometimes GlcA) and alternating GlcNAc/GlcNS ([-4IdoAα1-4GlcNAc/NSα1-]$_n$), and nonsulfated domains [-4GlcAβ1-4GlcNAcα1-]$_n$ [30,147]. The latter domains are closest to the PG core protein and have not been shown to interact with other proteins [147].

HS is a key regulator in the cardiovascular system. For example, HS is thought to immobilize fibroblast growth factor (FGF) in the ECM. Heparanase enzymes from platelets [148], neutrophils [149], and lymphocytes [150,151] degrade HS and consequently release active FGF which stimulates cellular adhesion, migration, proliferation, and differentiation [148-153]. In addition, HS regulates anticoagulant activity by binding antithrombin III, inducing a conformational change that allows antithrombin III to bind its targets thrombin and factor Xa [154,155].

1.2.4. GAGs in Pathology and Therapeutics

GAGs are key effectors of onset, progression, and outcome of infection and disease. Due to the ability of GAGs to interact with a diverse array of both ligands and receptors, GAGs can cause abnormal activation of cell signaling pathways and thus promote viral and bacterial infections [1-5,156,157], inflammation [1-4], and development of cancer [158-161] and many other diseases. For example, CS has been implicated in a number of neuropathological conditions including epilepsy [162], stroke [163], schizophrenia [164], Alzheimer's disease [165], and Parkinson's disease [166]; HA-mediated activation of CD44 has been implicated in tumor growth and metastasis [167,168]; binding of DS to HGF/SF may aid in tumor progression [133,134] and DS deficiency has been implicated in a new form of Ehlers-Danlos Syndrome [35,169,170]; KS is released from damaged articular cartilage thus serum KS is elevated in early stage osteoarthritis [171,172] and in the cerebral cortex of patients with Alzheimer's disease, decreases in overall KS levels and in sulfation of existing KS are seen [142]; and degradation of HS by heparanase triggers an excessive inflammatory response during sepsis, inducing septic shock [173,174], and activates FGF, which is thought to play an important role in tumor metastasis [150,175,176], inflammation [149], and autoimmunity [151].

On the other hand, their diverse capabilities give GAGs diagnostic and therapeutic potential. For example, screening for urinary DS may be used as a diagnostic tool for Ehlers-Danlos Syndrome [35] and KS may be used as a marker for early-stage osteoarthritis [171,172]. Pathogenesis can be inhibited via supplementation of or gene knockdown against enzymes involved in GAG biosynthesis [157] which would alter GAG structure and therefore receptor and ligand binding activities. For example, sodium chlorate, which prevents the normal extent of GAG sulfation, has been shown to inhibit hepatitis B viral infection [177]. Therapeutic effects are also seen when GAGs and GAG-degrading enzymes are targeted.

For example, a peptide interacting with heparanase has been shown to act as an anti-inflammatory agent during sepsis, thus preventing septic shock [173,174]. Removal of GAGs can also have therapeutic effects. For example, enzymatic removal of chondroitin sulfate may aid in recovery after spinal cord injury [178-181] and enhancing plasticity thus improving memory retention in neurodegeneration [182]. Additionally, GAG mimetics have shown promising results for use as viral inhibitors [157] as well as anti-cancer and anti-inflammatory agents [183-187]. For example, HS hexasaccharide HS06 activates p38 MAP kinase, inhibiting self-renewal in colorectal, breast, and pancreatic cancer stem cells [184] and a modified CS inhibits cell proliferation and induces apoptosis in myeloma and breast cancer cells [187]. The biggest challenges currently associated with using GAGs as therapeutics are the diverse structure and functions of GAGs [183]. Alteration of a given GAG structure or supplementation with GAG mimetics may inhibit pathogenesis but activate another pathway causing adverse side effects [157,183]. Uncovering mechanisms behind GAG-mediated pathogenesis will help unlock the therapeutic potential of GAGs.

1.2.5. Gap in Knowledge

The aforementioned studies have provided valuable information about the effects of GAGs in various tissues and pathways throughout the body. However, many of the mechanisms by which GAGs perform their functions have yet to be elucidated. GAG-protein binding is determined by the sequence, structure, charge, and shape (i.e., conformation) of both the GAG and the GAG-binding site on the protein [22,41]. Therefore, examining the three-dimensional structure and conformation of GAGs at atomic resolution is necessary to understand GAG binding properties and, thus, GAG bioactivity. However, this presents an unsolved problem due to the complexity of GAGs. For example, nontemplate based synthesis and enzymatic modifications (e.g., sulfation and UA isomerization) give GAGs variable length, sequence, sulfation, and charge. Furthermore, GAGs are conformationally complex due to flexibility of monosaccharide rings [188] and glycosidic linkages between monosaccharides [189-193].

Toward addressing the problem of studying these heterogeneous, polydisperse, and conformationally flexible biomolecules, this thesis describes the development of a method for rapidly producing three-dimensional atomic-resolution models of GAGs of arbitrary length. This is a key step toward improving understanding of the structure and biophysics of GAGs.

CHAPTER 2

METHODS

This chapter discusses some methods currently used to study GAGs and provides an in-depth explanation of the key methods used in this dissertation project.

2.1. Laboratory Methods

Studying GAG structural and thermodynamic properties presents a challenge for existing experimental methods, which are often labor intensive and require large quantities of GAGs. Each individual method has its own strengths and additional limitations. For example, reversed-phase ion-pair high performance liquid chromatography (HPLC) mass spectrometry has been used for structural analyses of tissue- and stage-specific glycosaminoglycans in *C. elegans* and *Drosophila* and provided direct measures of unsaturated CS disaccharides [58]. This is an effective method for determining GAG composition but not conformational properties and is limited to measuring disaccharide units. Isothermal titration calorimetry (ITC) has been used to determine the GAG binding site on a protein as well as binding specificity and affinity [43]. This method shows the thermodynamics of binding in solution but might not show polymer concentration and there may be cases in which ITC shows only the equilibrium part of the binding process and will not detect dissociation [194] and thus the thermodynamics of the dissociated protein and ligand which may be of functional importance. NMR experiments have proposed stable pyranose ring conformers [195] and are capable of revealing binding sites between a protein and ligand by observing chemical shifts before and after binding [196,197]. However, NMR data provide limited understanding because (1) they vary depending upon sample preparation and molecular composition and length and (2) GAG backbone and monosaccharide ring flexibility in aqueous solution produces various conformations in equilibrium thus NMR data can be explained by multiple different conformations [198] and is not sufficient to predict GAG structure and dynamics [199]. X-ray crystallography has been used to produce atomic-resolution images of GAG-protein complexes but these include only short oligosaccharides corresponding to the binding site of the protein [200-204], which are not necessarily the complete GAG structures participating in those complexes. Crystallization of these complexes is challenging because of the heterogeneity of GAGs and the fact that there may be multiple binding poses due to ionic interactions of GAGs with proteins [205].

2.2. Computational Methods

Computational methods can complement knowledge gained by experimental methods. For example, docking has been used to find potential binding poses of GAGs on proteins [206-209]. One problem with this methodology is that it screens rigid molecules, so while attractive and repulsive forces between fixed conformations of the GAG and protein are shown, it does not take into account the dynamic changes that take place before, during, and after binding. Docking can be used as a tool to help generate hypotheses or help interpret previous observations but it is not sufficient to study the dynamics of an interaction [207].

2.2.1. Molecular Dynamics Simulations

It is important to study molecular dynamics at an atomic level to determine biophysics, and thus bioactivity, of the molecule(s) of interest. This can be accomplished through molecular dynamics (MD) simulations, which capture random, realistic atomic movement with respect to surrounding atoms in a system and reveal thermodynamic properties of molecules in aqueous solution [210-214]. Many studies have validated MD data with experimental methods. For example, results from liquid chromatography-mass spectrometry [215], x-ray diffraction [216], and NMR [216-221] have validated GAG structures and conformations produced by MD, demonstrating the predictive power of MD simulations. Due to the ability to provide additional insights that cannot be obtained using experimental methods and its well-established predictive power, applications of MD for structure and activity prediction have expanded beyond academic research into drug design and pharmacophore modeling [222-225] as well as materials science [226-228].

2.2.1.1. Force Field. MD simulations are performed using a series of equations and parameters based on experimentally determined atomic and molecular properties (e.g., atomic radii, partial charges, bond lengths and angles of carbohydrates, lipids, nucleic acids, amino acids, and water molecules). Collectively, these equations and parameters are referred to as a force field. A force field uses a potential energy function to describe energy changes as a function of atomic movements and interactions and is used to construct a molecular mechanical model. The CHARMM force field [229-232] potential energy function (Equation 2.1) for a system with a particular spatial arrangement of atoms is equal to the sum of the energy costs of bond distortion, angular distortion, dihedral rotation, out-of-plane angle (or improper) distortion, and a nonbonded term which includes the strength of the interaction in units of energy and distance of interaction ($\frac{R_{\min}}{r}$, which is unitless).

$$\begin{aligned} V = & \sum_{\text{bonds}} k_{\text{b}}(b-b_0)^2 + \sum_{\text{angles}} k_{\theta}(\theta-\theta_0)^2 \\ & + \sum_{\text{dihedrals}} k_{\phi}[1+\cos(n\phi-\delta)] + \sum_{\text{impropers}} k_{\omega}(\omega-\omega_0)^2 \\ & + \sum_{\text{nonbonded}} \epsilon_{ij}\left[\left(\frac{R_{\min_{ij}}}{r_{ij}}\right)^{12} - \left(\frac{R_{\min_{ij}}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{r_{ij}} \end{aligned} \qquad (2.1)\ [232]$$

In Equation 2.1: k values are force constants; $b - b_0$ represents the bond distortion (i.e., the difference between the current and most energetically-favorable, or equilibrium, bond length; Figure 2.1.); $\theta - \theta_0$ represents the distortion of the angle between three bonded atoms from equilibrium; n represents the multiplicity of the dihedral function, which corresponds to the number of energy minima as the bond is rotated through 360°; ϕ represents a dihedral angle (i.e., angle of rotation about the central bond between four bonded atoms); δ represents the phase shift of the trigonometric function of the dihedral from that at equilibrium; $\omega - \omega_0$ represents the out-of-plane (improper) angular distortion; i and j represent a nonbonded pair of atoms; ϵ is the well depth of the Lennard-Jones potential (i.e., minimum energy), which is a measure of the strength of interaction between two atoms; r_{ij} is the distance between two nonbonded atoms (i and j); $R_{\min_{ij}}$ is the value of the interatomic distance at the Lennard-Jones potential energy minimum (i.e., the equilibrium distance between two atoms); and q is the partial atomic charge.

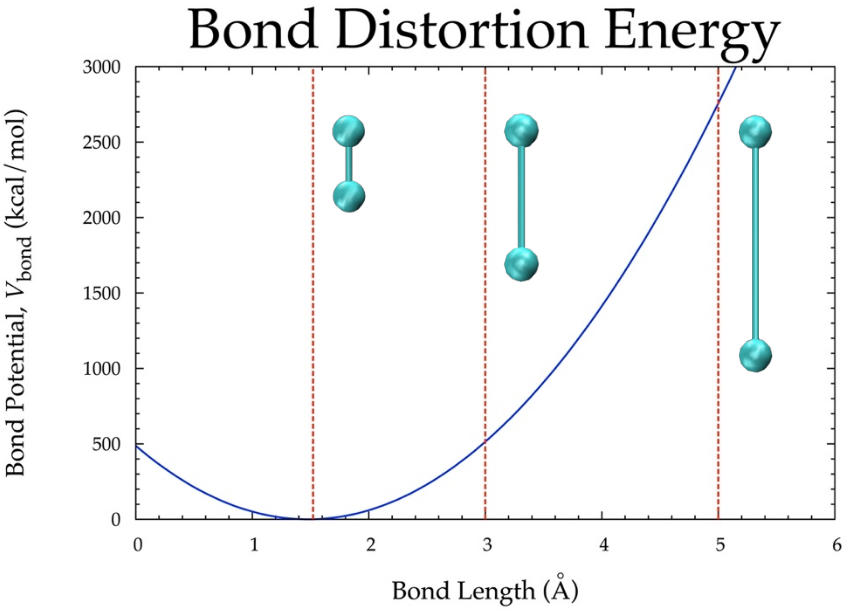


Figure 2.1. Bond distortion energy [$V_{\text{bond}} = k_{\text{b}}(b - b_0)^2$] as a function of bond length b (k_{b} is the bond stretching force constant) and molecular graphics of the same carbon-carbon bond with three different bond lengths (1.52 Å, 3.00 Å, and 5.00 Å). The red dotted lines on the plot mark the bond lengths shown. The shortest bond (1.52 Å) is the equilibrium bond length (i.e., where V_{bond} = 0 kcal/mol).

MD simulations combine the force field equation with a modified Newton's equation to generate the time evolution of molecular coordinates at a given temperature. The resulting ensemble of states (conformations) can be used to create a statistical mechanical model of an atomic system. MD can also be used for modeling atomic and molecular kinetics, which describes the rates of change of the system. It does so by integrating the equations of motion. MD also can provide thermodynamic properties (thermodynamics is a measure of the system's energy and entropy at equilibrium). MD does so by sampling equilibrium states of the statistical mechanical model. All of this allows MD to give atomic position, movement trajectories (Figure 2.2), and free energies as a function of time for a given atomic system [233,234].

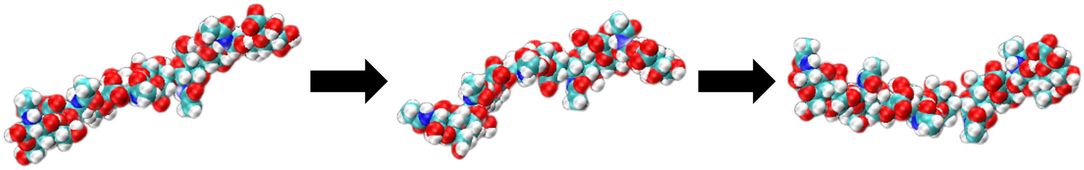


Figure 2.2. Three subsequent snapshots from a MD trajectory of a GAG 10-mer. Solvent atoms are not shown.

Each of the four main types of biomolecules (i.e., carbohydrates, lipids, nucleic acids, and proteins) has its own unique properties. Therefore, there are different parameter sets for different biomolecule types. The most commonly used force fields for carbohydrates include the AMBER [235], GLYCAM [236,237], OPLS [238,239], GROMOS [240-242], and CHARMM [214,243-245] force fields [246]. While each of these has been compared to experimental results and validated to some degree, not all force fields have demonstrated the same predictive power. For example, it has been shown that the CHARMM36 force field for carbohydrates [214,243-245] best reproduces ring conformations observed by NMR, compared to the general AMBER force field (GAFF1) [235] and GLYCAM06 [236,237] force fields for carbohydrates [221]. Furthermore, compared to GROMOS [242] and OPLS [239], CHARMM36 [243] gives more accurate predictions of the rotation about the improper angle ω in carbohydrates (namely gluco- and galactopyranosides) [246-250]. This may, in part, be due to the fact that, unlike other carbohydrate force fields like GAFF1 [235] and OPLS [238,239], CHARMM does not dampen 1—4 nonbonded electrostatic interactions (i.e., between atoms separated by three adjacent bonds) relative to longer range interactions. Historically, 1—4 interactions were scaled in force fields for proteins and kept in force fields for carbohydrates for simplicity. However, dampening 1—4 nonbonded interactions relative to longer-range interactions causes a problem with the energetics for the rotation of improper dihedral (ω) angles (n.b.: 1—4 nonbonded atoms participate in the same dihedral angles), and thus should not be used for carbohydrates [238,251].

2.2.1.2. Periodic Boundary Conditions. Periodic boundary conditions (PBCs) allow simulation of atoms and molecules in an infinite system by defining a unit cell, which is a periodic box with penetrable walls that is constrained so that the number of atoms is constant [233]. When an individual particle (i.e., ion or molecule) exits one side of the unit cell, it reenters the unit cell from the opposite side. This minimizes the number of solvent atoms in a system, making the system feasible to simulate without placing unrealistic restrictions on the movement of the atoms through space. The location of each particle within the unit cell is based on the Cartesian coordinates of the center of mass of the particle. For example, in Figure 2.3.c,d, a portion of the molecule appears to be outside the unit cell but the center of mass of the molecule is within the unit cell in each panel.

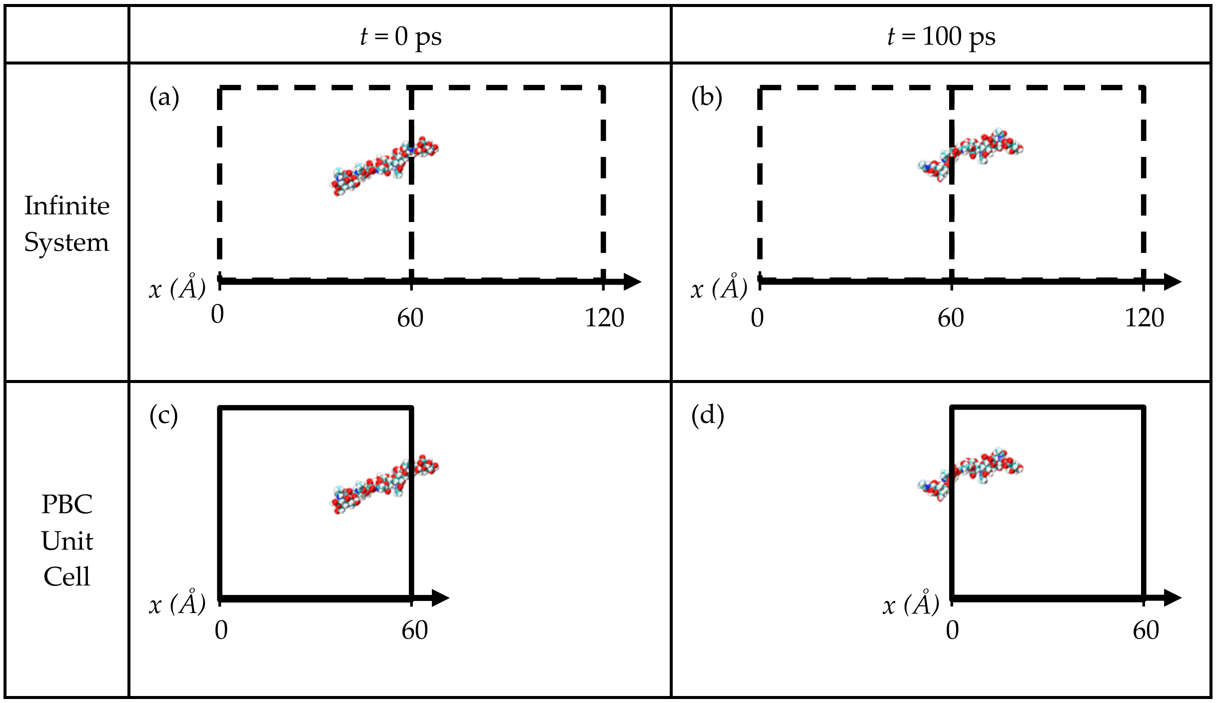


Figure 2.3. Snapshots of a GAG 10-mer at two different time (t) points in (a,b) an "infinite" system and (c,d) a unit cell with periodic boundary conditions; solid borders represent the edges of the unit cell

(edges have the same Cartesian coordinates at each time point in a MD trajectory). Solvent atoms are not shown.

2.2.1.3. All-Atom Explicit-Solvent MD. "All-atom" means that each atom in a biomolecule is treated as a separate particle, allowing full conformational flexibility (as opposed to treating clusters of atoms or an entire molecule as a single particle so that only the *inter*molecular interactions are revealed). Calculating energies of bonded and close-range nonbonded atom pairs within biomolecules reveals full conformational free energy landscapes, which can be used to study covalent conformational properties and noncovalent binding properties.

"Explicit solvent" means the user chooses the solvent conditions in which to simulate their molecule(s) of interest. Specifically, they can define the number of water molecules, ion type and concentration, pH, temperature, and pressure. For example, for simulation of molecules in human ECM, the solvent can be set to include *n* water molecules (based on solute and periodic unit cell dimensions) with 140 mM NaCl and neutralizing Na^+ or Cl^- counterions and have a temperature of 310 K (98.6 °F) and a pressure of 1 atm. As water molecules and ions play a crucial role in GAG conformation and bioactivity, solvent effects must be taken into account [252]. Therefore, all-atom explicit-solvent MD is an appropriate method to simulate bioactivity and improve understanding of conformational thermodynamics of short GAG biopolymers (e.g., up to 20 monosaccharide units) at atomic resolution.

2.2.1.4. Challenges with All-Atom Explicit-Solvent MD. One drawback of all-atom explicit-solvent MD is that atomic-resolution simulations of large systems are computationally taxing so there is a limit to the number of atoms in a system that is feasible to simulate. Due to the linear nature of GAGs, each additional monosaccharide unit leads to an increase in periodic box dimensions and thus an exponential increase in the number of solvent atoms in the system. Therefore, simulating GAGs with biologically-relevant chain lengths presents a challenge for all-atom explicit-solvent MD.

Another form of MD, known as coarse-grained (CG) simulations, can model 3-D structures of GAGs of any given composition but uses implicit solvent [253,254] and treats clusters of atoms as a single particle. This gives much fewer degrees of freedom for the system [218] and makes the simulation less taxing and thus much faster, allowing handling of much larger systems [255]. Therefore, CG MD can generate conformational ensembles of longer GAGs than all-atom explicit-solvent MD. However, since clusters of atoms are treated as a single particle and the solvent is treated as a continuum, CG MD-generated GAG conformational ensembles may not be as accurate as those generated by all-atom explicit-solvent MD.

2.2.2. Conformational Analysis

Conformational thermodynamics of molecules can impact binding properties. A GAG's shape, along with sequence and charge, must be complementary to a biomolecule in order to bind. Studying atomic-resolution conformational thermodynamics of GAGs opens the door to predicting GAGs' binding capabilities with other biomolecules.

2.2.2.1. Glycosidic Linkage Dihedral Free Energy Changes $\Delta G(\phi, \psi)$. Free energy calculations are widely used to examine conformational landscapes and binding properties. For example, MD and free energy analyses were performed to determine binding poses and thermodynamics of experimentally determined protein-GAG complexes [209] and glycosidic linkage free energy calculations are commonly used to study GAG backbone conformation and compare MD predictions to experimental data [216,240,256-259].

Glycosidic linkage flexibility is important because the linkage conformations have the greatest effect on the overall shape of a GAG. For example, rigid glycosidic linkages will favor extended, linear GAG conformations, whereas highly flexible glycosidic linkages will give a wide range of compact (i.e., curved) and extended GAG conformations and be able to conform to a wider variety of protein binding sites.

Glycosidic linkage conformations are quantified by ϕ and ψ, which represent the angles of rotation about glycosidic dihedrals (Figure 2.4). The change in Gibbs free energy as a function of these angles—$\Delta G(\phi, \psi)$—quantifies the stability of a given conformation relative to another at equilibrium. In MD analysis, $\Delta G(\phi, \psi)$ (Equation 2.2) is calculated based on the frequency of a given conformation during simulation, which can be determined by binning ϕ and ψ values into, for example, 2.5° x 2.5° bins and counting the occurrences of a conformation in each bin.

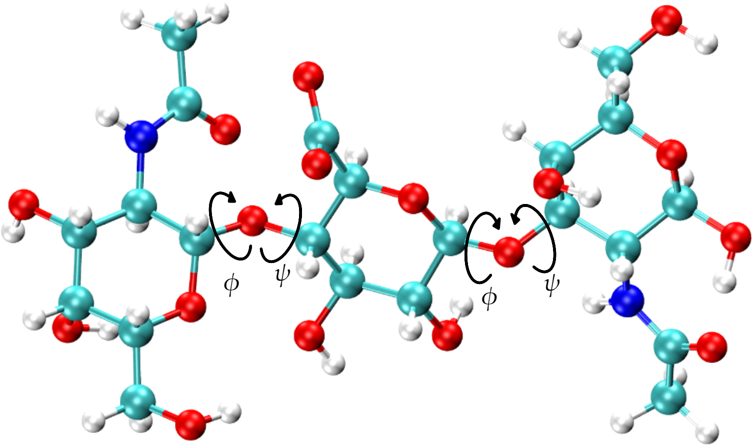


Figure 2.4. GAG trisaccharide showing ϕ and ψ dihedral rotations about glycosidic linkage bonds.

$$\Delta G(\phi_i, \psi_j) = -RT\ln(n_{ij}) - k \tag{2.2}$$

In equation 2.2: R is the universal gas constant; T is the temperature of the MD simulation; n_{ij} is the bin count for the bin corresponding to ϕ_i, ψ_j; and k is a constant chosen so that the global free energy minimum (equilibrium conformation) is located at $\Delta G = 0$ kcal/mol.

In this work, glycosidic linkage $\Delta G(\phi, \psi)$ values were compared both within individual GAG polymers and between GAG polymers of the same type but different lengths to determine (1) glycosidic linkage flexibility and (2) if glycosidic linkage conformations are independent of adjacent glycosidic linkage and monosaccharide ring conformations and GAG polymer length. Glycosidic linkage dihedral free energy analyses from all-atom explicit-solvent MD-simulated GAGs suggest that different glycosidic linkage types have varying degrees of flexibility but all exhibit random behavior when GAGs are free floating in aqueous solution [188,260].

2.2.2.2. Monosaccharide Ring Cremer-Pople Parameters ϕ, θ, Q. Cremer-Pople (C-P) parameters quantify pyranose ring conformations [261]. These parameters are the standard for analyzing pyranose ring puckering and have been used extensively in studies of GAG conformational thermodynamics [216,219,262,263]. To calculate C-P parameters using Cartesian coordinates of the atoms in a ring, the ring is treated as a sphere with polar and equatorial planes [261,264] (Figure 2.5). The C-P parameters ϕ and θ represent the angular distortions of the endocyclic ring atoms from the polar and equatorial planes (n.b.: by historical convention, the glycosidic linkage dihedral ϕ and C-P parameter ϕ are referred to by the same symbol but are completely different geometric properties). Q is the puckering amplitude, which describes the displacement of atoms from the point where the polar and equatorial planes meet. Of note, the name of the ring pucker indicates the type of conformation (e.g., "chair" conformation is represented by "C") and position of carbon atoms with respect to polar and equatorial planes from the standard view (i.e., oriented such that carbons 1-3 are in the front and carbon 1 is on the far right of the plane of view); e.g., in $^{4}C_{1}$ chair conformation, carbon 4 is located above the equatorial plane and to the left of the polar plane (hence "^{4}…") and carbon 1 is located below the equatorial plane and to the right of the polar plane (hence "…$_{1}$").

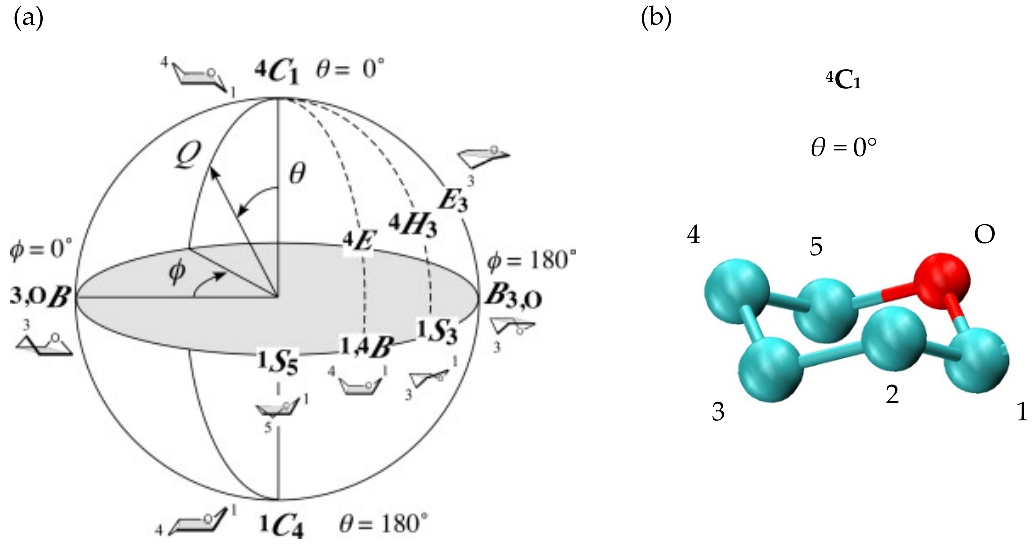


Figure 2.5. (a) Spherical map of pyranose ring conformations represented by C-P parameters ϕ, θ, and Q (image taken from Figure 2 in Reference [264]) and (b) an atomic-resolution molecular graphic of a carbohydrate ring in $^4C_1$ chair conformation (oxygen atom in red; carbon atoms in cyan and numbered according to standard numbering scheme).

In this work, C-P parameters of all monosaccharides of the same type (e.g., all GalNAc monosaccharides) within a GAG polymer and between GAG polymers of the same type but different lengths were compared to determine (1) the ring flexibility of each monosaccharide type and (2) if ring puckering is independent of adjacent glycosidic linkage and monosaccharide ring conformations and GAG polymer length. C-P analyses of all-atom explicit-solvent MD-simulated GAGs indicate that different monosaccharide types have varying degrees of flexibility but all exhibit random behavior when GAGs are free floating in aqueous solution [188,260]. This conclusion, along with that drawn from GAG glycosidic linkage free energy analyses, informed the development of the methodology presented in this dissertation.

2.2.3. Algorithm to Construct 3-D Models of Glycosaminoglycans

As a single GAG may contain up to 200 monosaccharide units, all-atom explicit-solvent MD cannot routinely be applied to generate atomic-resolution conformational ensembles of GAG biopolymers in aqueous solution because of the enormous computational costs. This dissertation describes an algorithm that was developed to overcome this technical barrier presented by MD simulations of GAGs with biologically-relevant chain lengths in fully-solvated systems (containing millions of atoms). This algorithm applies glycosidic linkage and monosaccharide ring conformations from unbiased all-atom explicit-solvent MD simulations of nonsulfated GAG chains containing 20 monosaccharide units (i.e., "20-mers") to construct 3-D atomic-resolution models of GAGs of arbitrary chain length. Due to its efficient computational performance, this algorithm can generate tens of thousands of 3-D atomic-resolution models (representative of snapshots from a MD trajectory) of a given long-chain GAG (e.g., 200-mer) in under a day. For perspective, a MD simulation of a fully-solvated GAG 20-mer system (containing ~191,000 atoms) produced a trajectory with 10,000 snapshots in one month.

CHAPTER 3

EFFICIENT CONSTRUCTION OF ATOMIC-RESOLUTION MODELS OF NONSULFATED CHONDROITIN GLYCOSAMINOGLYCAN USING MOLECULAR DYNAMICS DATA

This chapter has been published as:

Whitmore, E.K.; Vesenka, G.; Sihler, H.; Guvench, O. Efficient Construction of Atomic-Resolution Models of Non-Sulfated Chondroitin Glycosaminoglycan Using Molecular Dynamics Data. *Biomolecules*. **2020**, *10*(4), 537. doi:10.3390/biom10040537

(Note that minor edits have been made for consistency in this dissertation.)

3.1. Introduction

The diverse group of protein–carbohydrate conjugates called proteoglycans (PGs) is a fundamental component of tissue structure in animals and can be found in the extracellular matrix (ECM) as well as on and within cells. PGs bind growth factors [1-3,5,11,17-22,265], enzymes [5,22], membrane receptors [22], and ECM molecules [5,9,22]. By doing so, they modulate signal transduction [9,25], tissue morphogenesis [1-3,5,265], and matrix assembly [5,47,92,93]. PG bioactivity is often dependent on the covalently linked carbohydrate chains called glycosaminoglycans (GAGs), which are linear, highly negatively charged, and structurally diverse carbohydrate polymers. GAGs mediate receptor–ligand complex formation by either forming noncovalent complexes with proteins or inhibiting the formation of complexes with other biomolecules. This makes GAGs key modulators in many diseases, giving them potential therapeutic applications. For example, heparan sulfate (HS) is released during sepsis and induces septic shock [266,267]; the removal of chondroitin sulfate (CS) may enhance memory retention and slow neurodegeneration in patients with Alzheimer’s disease [182,268,269]; and dermatan sulfate (DS) deficiency has been implicated in Ehlers–Danlos syndrome, thus the screening of DS in urine could be used as an early diagnostic tool [169,170].

GAG binding sites on proteins are determined by protein sequence and structure, with requirements for both shape and charge complementarity [22,41]. Thus, GAG function depends on GAG three-dimensional structure and conformation. Even subtle structural differences impact GAG function. For example, while CS and DS have many functional differences, the only structural difference is in the chirality of the uronic acid monosaccharides. While much is known about GAG function, attempting to study GAG conformational thermodynamics at atomic resolution presents a largely unsolved problem for existing experimental methods. This is largely due to the structural and conformational complexities of GAGs. For example, a given GAG consists of a repeating sequence of a particular disaccharide, but conformational complexity is introduced through flexibility in the glycosidic linkages between monosaccharides [189-193] (Figure 3.1). Additional complexity results from nontemplate-based synthesis [37] and variable enzymatic sulfation [270], which means a biological sample of a GAG composed of a specific disaccharide repeat will be polydisperse and heterogeneous owing to the variable length and sulfation of the individual polymer molecules. Liquid chromatography–mass spectrometry (LC-MS) [271-273], X-ray crystallography [200-205], and nuclear magnetic resonance (NMR) [195,215,274,275] are used to study GAGs but are limited in their ability to account for all of these complexities. Additionally, some studies have used results from LC-MS [215], X-ray crystallography [216], and NMR [216-221] to compare and validate conformational data from molecular dynamics (MD) simulations. This suggests that MD simulations can produce results complementary to experimental analysis methods by providing realistic three-dimensional atomic-resolution molecular models of GAG conformational ensembles [210-214].

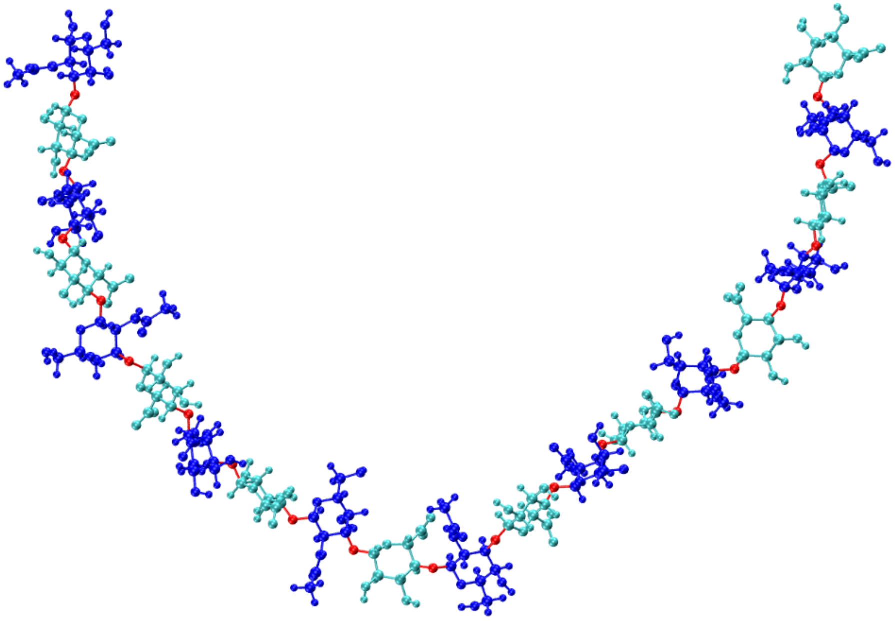


Figure 3.1. Compact nonsulfated chondroitin 20-mer conformation arising from flexible glycosidic linkages (red) between monosaccharide rings (GalNAc in blue and GlcA in cyan). The molecular graphics throughout are produced with the VMD program [276].

A critical challenge with MD simulations of GAGs is that a single biological GAG polymer chain may contain up to 200 monosaccharide units [2]. When fully solvated, the resulting system will have in excess of 10^6 atoms. It is not feasible to routinely simulate such a system using the current graphics processing unit (GPU)-accelerated MD code with a modern GPU and multi-core CPU. This limits the utility of all-atom explicit-solvent MD as a tool for routine conformational analysis of GAGs of this size.

Coarse-grained (CG) MD simulations are the most feasible current alternative to all-atom explicit-solvent MD as they entail fewer degrees of freedom for the solute [218] and often an implicit (continuum) description of the solvent [253,254]. This can make CG MD two to three orders of magnitude faster, thereby allowing for the handling of large systems [255], such as GAG 200-mers. Indeed, a recent CG model using glycosidic linkage and ring pucker energy functions has provided previously-unseen details of the structure–dynamics relationship of GAGs in the context of PGs [218]. An important insight from that study was that GAGs, in contrast to the unique ordered conformations of folded proteins, need to be considered as existing in conformational ensembles containing a large diversity of three-dimensional conformations.

As an alternative approach to using CG MD to generate such conformational ensembles for GAGs, we propose using glycosidic linkage and monosaccharide ring conformations from unbiased all-atom explicit-solvent MD simulations [214,243-245] of short GAG polymers to rapidly construct conformational ensembles for GAGs of an arbitrary length. Toward this end, we studied a nonsulfated chondroitin 20-mer with the sequence [-4 glucuronate β1-3 N-acetylgalactosamine β1-]$_{10}$ for its simplicity and homogeneity. We first ran a microsecond-scale all-atom explicit solvent MD on the 20-mer and used the resulting trajectories to develop a database of conformations. From this database, we randomly selected individual values for the bond lengths, bond angles, and dihedral angles in the glycosidic linkages connecting glucuronate (GlcA) and N-acetylgalactosamine (GalNAc) and in the monosaccharide rings. These values were used to construct a 20-mer conformational ensemble. The comparison of the constructed ensemble with the MD-generated ensemble of 20-mer conformations revealed similar end-to-end distance distributions, with a strong bias toward extended conformations in both cases. Short end-to-end distances associated with more compact conformations were facilitated by the sampling of non-$^4C_1$ ring puckering by GlcA. This change in ring geometry, which occurs rarely on the microsecond timescale, introduced kinks into the polymer, causing it to bend back toward itself. The fact that the MD-generated ensemble had a great deal of variability in both end-to-end distances and radii of gyration demonstrates the inherent flexibility of the chondroitin polymer in aqueous solution. The fact that the constructed ensemble has very similar conformational properties to the MD-generated ensemble suggests that there is little correlation between the individual dihedral angle values that determine the internal geometry of a given conformation. Therefore, on the timescale of the simulations, nonsulfated chondroitin 20-mer does not appear to have any higher-order structure, in contrast to, for example, the secondary and tertiary structure seen in proteins. This lack of higher-order structure was borne out in a comparison of end-to-end distances for constructed vs. MD-generated ensembles of 10-mers, with the constructed ensemble built using the 20-mer database. Finally, we used the methodology to produce conformational ensembles of 100-mers and of 200-mers. The ability to model polymers with biologically-relevant chain lengths (e.g., 100- to 200-mers) will provide insights into GAG binding by other biomolecules. This will be especially useful in understanding the formation of complexes containing multiple biomolecules bound to a single GAG.

Other programs that construct three-dimensional atomic-resolution models of GAG polymers exist, for example, Glycam GAG Builder [277], POLYS Glycan Builder [278], CarbBuilder [279], and MatrixDB GAG Builder [280,281], which allow the user to choose GAG type, length, and sequence and are useful tools for producing an initial structure for MD simulations. Glycam and POLYS Glycan Builder allow the user either to specify particular glycosidic linkage dihedral angle values or use default parameters pulled from their databases. The databases used by Glycam, POLYS Glycan Builder, and Carb Builder include GAG mono- and disaccharide structures determined by molecular mechanics and/or MD. MatrixDB pulls from databases of experimentally determined conformations of GAG disaccharides from crystallized GAG–protein complexes. While the user has the option to choose the GAG length, these tools are intended for shorter GAG polymers. In contrast to these tools, our algorithm pulls from a database of full conformational landscapes of unbound GAG 20-mers. Additionally, our algorithm is intended for modeling long GAG polymers with biologically-relevant chain lengths and can quickly produce large ensembles (e.g., on the order of 10,000 3-D models) of polymer conformations that we would expect to see in simulation. Thus, it eliminates the need for simulation, reducing time and computational cost.

3.2. Materials and Methods

3.2.1. Molecular Dynamics

3.2.1.1. System Construction. Coordinates for all systems were constructed using the CHARMM software [229-231] v. c41b2 with the CHARMM36 (C36) biomolecular force field for carbohydrates [214,243-245]. Of note, it has been shown that MD simulations can reproduce ring puckers observed by NMR [219-221], with one study demonstrating the capacity of the CHARMM36 force field to reproduce NMR data for an iduronate derivative in the context of a heparin analogue [221]. The initial conformation for an MD simulation of nonsulfated chondroitin 20-mer was fully extended, with glycosidic linkage dihedrals ϕ = -83.75° and ψ = -156.25° in all GlcAβ1-3GalNAc linkages and ϕ = -63.75° and ψ = 118.75° in all GalNAcβ1-4GlcA linkages. These glycosidic linkage dihedral angle values were found to be the most energetically favorable in MD-simulated nonsulfated chondroitin disaccharides [212]. All other internal coordinates were taken from the force-field files. In this conformation, the 20-mer had an end-to-end distance of 101.8 Å and was solvated in a cubic periodic unit cell with an edge length of 124.3 Å (~63,000 water molecules). The explicit solvent consisted of the TIP3P water model [282,283], neutralizing Na+ counterions, and 140 mM sodium chloride.

3.2.1.2. Energy Minimization and Heating. The NAMD program [284] v. 2.12 (http://www.ks.uiuc.edu/Research/namd/) was used to minimize the potential energy for 1000 steps using the conjugate gradient method [285,286] then heat the system to the target temperature of 310 K by reassigning velocities from a random distribution at the target temperature every 1000 steps for 20,000 steps with a timestep of 0.002 ps (40 ps). During heating, harmonic positional restraints were placed on nonhydrogen atoms of the solute and constraints [287-289] were applied to maintain equilibrium values for TIP3P geometries and for bond lengths involving hydrogen atoms. The Lennard-Jones (L-J) [290] and electrostatic potential energies had cutoff distances of 10 Å. An energy switching function [291] was applied to L-J interactions between 8 and 10 Å and an isotropic pressure correction accounted for contributions from L-J interactions beyond the cutoff [292]. The particle mesh Ewald (PME) method [293] with fourth order B-spline interpolation for a cubic unit cell and fast Fourier transform (FFT) grid spacing of 1.0 Å along each axis was used to account for electrostatic interactions beyond the cutoff. Consistent with the CHARMM additive force fields for proteins, nucleic acids, lipids, and small molecules [294], carbohydrate 1-4 nonbonded interactions were not scaled (i.e., scaling factor = 1.0). Heating was done under constant pressure with pressure regulated at 1 atm by a Langevin Piston barostat [295]. A 500,000-step (1-ns) unbiased constant particle number/constant pressure/constant temperature (NPT) MD run followed with a temperature of 310 K maintained by a Langevin thermostat [296] and without positional restraints. The average periodic cell parameters from the last half of this NPT ensemble trajectory (123.7 Å) were used as cell basis vectors for the quadruplicate canonical (NVT) ensemble MD simulations detailed below. These were preceded by minimization and heating as detailed above, with the exception of constant volume (i.e., no Langevin piston barostat) with a box edge length of 123.7 Å.

3.2.1.3. Production Simulations. Unbiased canonical (NVT) ensemble MD was run using CHARMM software with the OpenMM GPU acceleration interface [297-302] on CUDA platform and GTX 1080 Ti graphics cards (NVIDIA Corp., Santa Clara, USA.). Nonbonded interaction truncations and energy calculations were performed using the same methods from the heating stage, and Ewald summation of Gaussian electrostatic charge density distributions [303,304] with a width of 0.320 was performed. The SHAKE algorithm [287] was used to constrain all water geometries and bonds involving hydrogen atoms using bond distances from the parameter table and the leapfrog Verlet integration algorithm [305] was used for Langevin dynamics with a friction coefficient of 0.1 ps^{-1}, a constant temperature of 310 K, and a 0.002-ps timestep. Prior to production, each of four replicates was equilibrated for 50,000 steps (100 ps). Simulations were run for 250,000,000 steps (500 ns) and atomic coordinates were saved at 25,000-step (50-ps) intervals for analyses (10,000 snapshots per quadruplicate simulation).

3.2.1.4. Conformational Analysis. Chondroitin conformational properties were quantified by bond length, bond angle, and dihedral angle values. Glycosidic linkage conformational values considered were bond lengths for C_1-O and O-C_n, bond angles defined by O_5-C_1-O and C_1-O-C_n, and ϕ and ψ dihedral angles with IUPAC definitions: $\phi = O_5\text{-}C_1\text{-}O\text{-}C_n$ and $\psi = C_1\text{-}O\text{-}C_n\text{-}C_{(n-1)}$ (Figure 3.2). Glycosidic linkage dihedral free energies $\Delta G(\phi, \psi)$ were analyzed to characterize potential conformational patterns. ϕ, ψ dihedral values from the MD-generated 20-mer ensemble are taken to have uniform probabilities. $\Delta G(\phi, \psi)$ is therefore computed by binning these values into 2.5° x 2.5° bins and then using the relationship $\Delta G(\phi_i,\psi_j) = -RT\ln(n_{ij}) - k$, where n_{ij} is the bin count for the bin corresponding to ϕ_i,ψ_j, R is the universal gas constant, T is the temperature of the MD simulations, and k is chosen so that the global minimum is located at $\Delta G = 0$ kcal/mol.

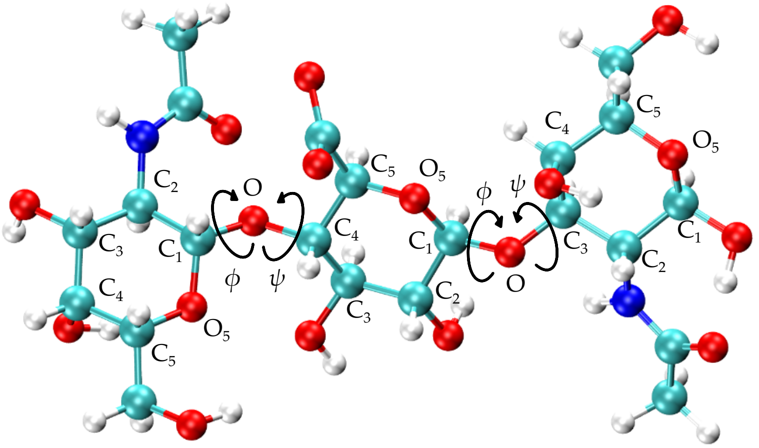


Figure 3.2. Nonsulfated chondroitin trisaccharide with glycosidic linkage dihedral angles, monosaccharide ring atoms, and linker oxygen atoms labeled; glycosidic linkage parameters used in the construction algorithm include C_1-O and O-C_n bond lengths, O_5-C_1-O and C_1-O-C_n bond angles, and $\phi = O_5\text{-}C_1\text{-}O\text{-}C_n$ and $\psi = C_1\text{-}O\text{-}C_n\text{-}C_{(n-1)}$ dihedral angles.

Geometric values defining monosaccharide ring conformations included all bond lengths, bond angles, and dihedral angles within the ring and in exocyclic functional groups that are not part of a glycosidic linkage. To characterize potential conformational patterns in monosaccharide rings, Cremer-

Pople (C-P) ring-puckering parameters (ϕ, θ, Q) of each monosaccharide ring in the MD-simulated 20-mer ensemble were computed. Conformations of each element (i.e., each GlcA monosaccharide, GalNAc monosaccharide, β1-3 linkage, and β1-4 linkage) were extracted separately from each saved snapshot of the 20-mer MD trajectories. Initially, data were separated out by run and residue/linkage number and aggregated across all snapshots in each run to determine if conformational data were the same in different runs and if individual linkage and ring conformations are dependent upon one another. Subsequently, all individual conformations were aggregated across all snapshots in all runs (e.g., 10,000 snapshots * 4 runs * 10 GlcA monosaccharides = 400,000 samples of GlcA monosaccharide conformations) to yield one set of data for each of: (1) GlcA monosaccharide conformation, (2) GalNAc monosaccharide conformation, (3) β1-3 linkage conformation, and (4) β1-4 linkage conformation.

3.2.2. Construction Algorithm to Generate GAG Conformational Ensembles

The conformational data described above served as inputs to an algorithm we developed to generate chondroitin polymer conformational ensembles of user-specified length and with a user-specified number of conformations. The algorithm works as follows:

1. In each constructed polymer conformation, each glycosidic linkage and monosaccharide ring is treated independently, and conformational parameters are randomly selected from the database containing the corresponding linkage or ring conformations from the 20-mer MD trajectories;
2. Two CHARMM stream files are written, one to define the sequence and linkages in the polymer and another to perform the following procedure for each frame: (1) All internal geometry conformation values selected by the algorithm are assigned and used to construct atomic coordinates. (2) End-to-end distance (i.e., distance between C_1 of the reducing end and C_4 of the nonreducing end) and radius of gyration are calculated. (3) A 100-step steepest descent (SD) potential energy minimization followed by a 100-step conjugate gradient minimization, each with intramolecular restraints, is performed to relieve bonded strain and steric clashes. The Lennard–

Jones potential ($E_{\mathrm{L-J}}$) is calculated on an atom–atom pair (i, j) basis using an energy switching function, as implemented in CHARMM with r_{on} = 7.5 Å and r_{off} = 8.5 Å [229]. As there is no solvent and thus no solvent screening of electrostatic interactions, electrostatics are excluded from energy calculations to prevent the nonphysical intramolecular association of charged and polar groups. All glycosidic linkage and endocyclic ring dihedral angles, along with a dihedral angle in each GlcA carboxylate group (C_4-C_5-C_6-O_{61}) and GalNAc N-acetyl group (C-N-C_2-C_3), are restrained to their starting values (i.e., those randomly selected from the database) during minimization so as not to change the conformations observed in simulation. Dihedral restraint energy (E_{rdihe}) is calculated by comparing each restrained dihedral angle's database value (ϕ_0) to its value (ϕ_1) in the current frame of minimization with a force constant (k_{dihe}) of 100.0 kcal/mol/radian/radian (Equation 3.1)

$$E_{\mathrm{rdihe}} = k_{\mathrm{dihe}} * \sum (\phi_1 - \phi_0)^2 \qquad (3.1);$$

3. To ensure conformational ensembles do not contain nonphysical conformations, a bond potential energy (E_{b}) cutoff is applied. This cutoff is the sum of a polymer-length-specific cutoff and a constant independent of polymer length. The length-specific component of the cutoff is the bond potential energy after energy minimization, performed using the same restraints and minimization protocol used for each frame of the constructed ensemble (outlined above), of the polymer constructed in a fully-extended conformation (i.e., with the same glycosidic linkage ϕ and ψ angles as the starting conformation for MD simulations). The constant is added as a buffer to account for slight variations in the energies of other extended conformations. As linkage and ring conformations are treated independently and selected at random, it is possible to have a bond piercing another monosaccharide ring that may not be corrected by minimization. To estimate the ring-piercing bond strain energy for each exocyclic bond not participating in a glycosidic linkage, a system containing two nonbonded monosaccharides (i.e., GlcA and GalNAc,

GlcA and GlcA, or GalNAc and GalNAc) was constructed such that an exocyclic bond of one monosaccharide pierces the ring of the other. To estimate the bond strain energy for each bond participating in a glycosidic linkage, a system containing one disaccharide unit (i.e., GlcAβ1-3GalNAc or GalNAcβ1-4GlcA) and a single monosaccharide (i.e., GlcA or GalNAc) was constructed such that a linkage bond in the disaccharide pierces the ring of the single monosaccharide. Systems containing interlocking rings (i.e., GlcA-GalNAc, GlcA-GlcA, and GalNAc-GalNAc) were also constructed to estimate the bond strain energy of the bonds piercing the opposite ring. The same energy minimization protocol used in the algorithm was performed on this conformation, as well as a conformation in which the nonbonded saccharide units are 20 Å apart, and the post-minimization lengths of the bond piercing the ring in the initial conformation were compared. The pierced bond length (x_2), the nonpierced bond length (x_1), and the equilibrium bond length (x_0) and corresponding force-field bond-stretching constant (k_b) from the CHARMM parameter file were used to estimate a lower bound on the energy (ΔE_b) resulting from the bond distortion (Equation 3.2).

$$\Delta E_\mathrm{b} = k_\mathrm{b} * [(x_2 - x_0)^2 - (x_1 - x_0)^2] \tag{3.2}$$

Of all conformations that still had a bond piercing a ring after minimization, the smallest ΔE_b = 132.3 kcal/mol. Of the conformations in which ring piercing was corrected during minimization, the maximum $\Delta E_\mathrm{b} < 1$ kcal/mol. Thus, a buffer of 100 kcal/mol is added to the post-minimization bond potential of the initial extended conformation for any given polymer length. If the post-minimization bond potential of a given frame is beyond this cutoff, the frame is excluded from the ensemble.

For internal validation of our implementation of the algorithm, bond length probability distributions for each type of bond (i.e., C-C single bond, C-O single bond, C=O double bond, C-O partial double bond of GlcA carboxylate group, C_2-N single bond between GalNAc amide and ring carbon, C-N single bond within GalNAc amide, C-H bond, O-H bond, and N-H bond), free energies $\Delta G(\phi,\psi)$ for β1-3 and β1-4 glycosidic linkages, C-P parameters of GlcA and GalNAc monosaccharide rings, end-to-end distance distributions, and scatterplots of radius of gyration as a function of end-to-end distance from MD-generated ensembles and constructed ensembles both before and after energy minimization were compared. Additionally, bond potential energy distributions from constructed ensembles after energy minimization were plotted to verify that the algorithm calculated an appropriate energy cutoff and gave the expected energy distributions for the given polymer size.

To assess the expediency of application of MD-generated 20-mer conformations to construct chondroitin polymers of variable length, we constructed a nonsulfated chondroitin 10-mer ensemble using the algorithm and compared it to chondroitin 10-mer conformational ensembles generated by MD using the same protocol as the 20-mer simulations. We also constructed conformational ensembles of a nonsulfated chondroitin 100-mer and 200-mer to demonstrate the efficacy and efficiency of our algorithm to construct conformational ensembles of chondroitin polymers with biologically-relevant chain lengths.

3.3. Results and Discussion

3.3.1. Glycosidic Linkage Geometries

In nonsulfated chondroitin 20-mer MD simulations, we found that all ϕ, ψ dihedrals sampled in GlcAβ1-3GalNAc linkages were centered about a global free energy minimum (Min I) while GalNAcβ1-4GlcA linkages showed more flexibility. In addition to a global minimum, $\Delta G(\phi, \psi)$ for GalNAcβ1-4GlcA also had two local minima (Min II and Min II’) (Figure 3.3 and Table 3.1). To validate these observed glycosidic linkage geometries, we looked at the free energy minima of nonsulfated chondroitin glycosidic linkage dihedrals from biased MD simulations of disaccharides (using dihedral definitions $\phi = O_5\text{-}C_1\text{-}O\text{-}C_n$ and $\psi = C_1\text{-}O\text{-}C_n\text{-}C_{(n+1)}$ as opposed to the IUPAC $\psi = C_1\text{-}O\text{-}C_n\text{-}C_{(n-1)}$ used in our study) [212] (Table A.1). We found that at each free energy minimum in β1-3 and β1-4 linkages, our ϕ dihedrals differed by no more than +/-2.5° and our ψ dihedrals differed by no more than +/-127.5°, which is in close agreement if we assume C_1-O-C_3-C_2 = C_1-O-C_3-C_4 + 120° and C_1-O-C_4-C_3 = C_1-O-C_4-C_5 - 120°. Additionally, our data were mostly in agreement with the most energetically-favorable glycosidic linkage dihedrals (i.e., at global minima) in nonsulfated chondroitin hexasaccharides from MD simulations (using dihedral definitions $\phi = O_5\text{-}C_1\text{-}O\text{-}C_n$ and $\psi = C_1\text{-}O\text{-}C_n\text{-}C_{(n+1)}$) and validated by NMR [259] (Table A.1). The biggest difference was in our β1-3 ψ dihedrals, which differed by about +100° (+120° difference expected). This study restrained pyranose rings to $^4C_1$ chair and did not use explicit solvent in simulations. Each of these factors may contribute to interactions between neighboring monosaccharides and thus glycosidic linkage conformation, which would explain the variation from our results.

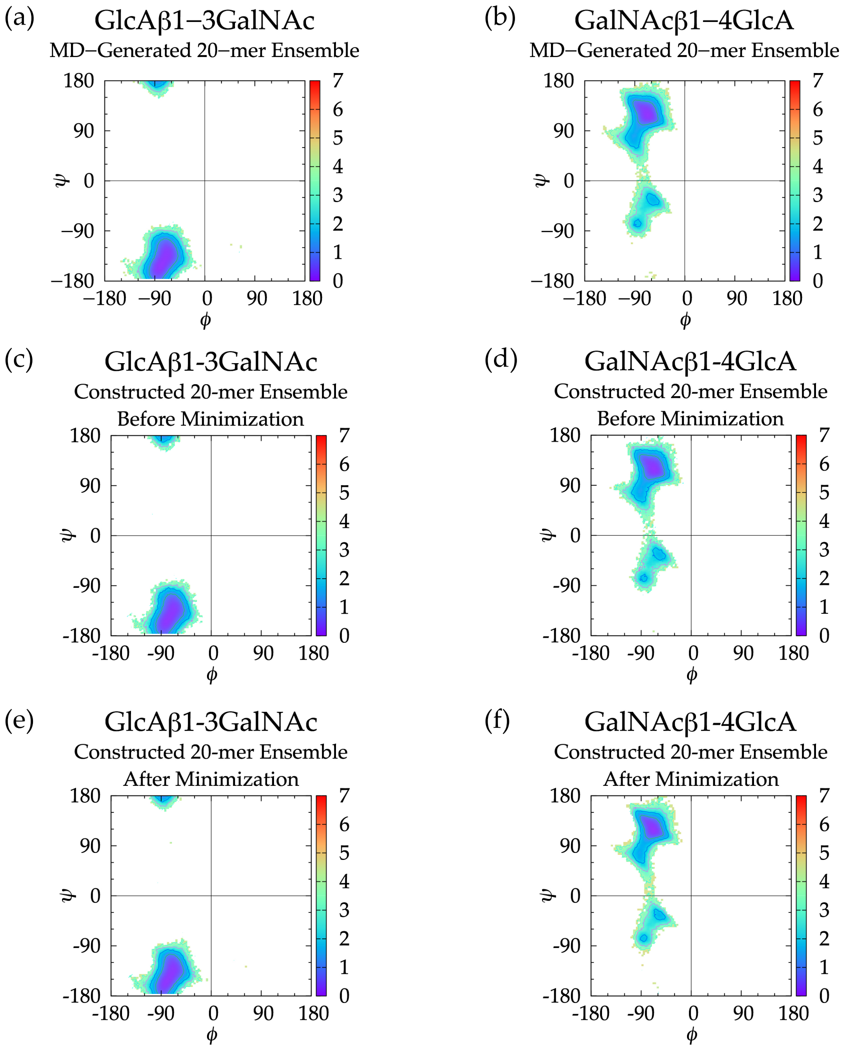


Figure 3.3. ΔG(ϕ, ψ) in nonsulfated chondroitin 20-mer ensembles for aggregated GlcAβ1-3GalNAc and GalNAcβ1-4GlcA glycosidic linkage data (a,b) MD-generated ensembles, (c,d) constructed ensembles before minimization, and (e,f) constructed ensembles after minimization; contour lines every 1 kcal/mol.

Table 3.1. Glycosidic Linkage Dihedrals (ϕ, ψ) [1] and Free Energy ($\Delta G(\phi, \psi)$ kcal/mol) Minima.

	MD-Generated 20-mer Ensemble				Constructed 20-mer Ensemble (Before Energy Minimization)				Constructed 20-mer Ensemble (After Energy Minimization)			
	GlcAβ1-3 GalNAc		GalNAcβ1-4 GlcA		GlcAβ1-3 GalNAc		GalNAcβ1-4 GlcA		GlcAβ1-3 GalNAc		GalNAcβ1-4 GlcA	
Min	ϕ, ψ	ΔG	ϕ, ψ	ΔG	ϕ, ψ	ΔG	ϕ, ψ	ΔG	ϕ, ψ	ΔG	ϕ, ψ	ΔG
I	-81.25°, -153.75°	0.00	-66.25°, 116.25°	0.00	-76.25°, -148.75°	0.00	-66.25°, 116.25°	0.00	-81.25°, -153.75°	0.00	-68.75°, 121.25°	0.00
II			-58.75°, -33.75°	1.57			-61.25°, -33.75°	1.54			-58.75°, -33.75°	1.57
II′			-86.25°, -73.75°	1.80			-86.25°, -73.75°	1.71			-86.25°, -78.75°	1.73

[1] ϕ, ψ dihedral angles were sorted into 2.5° bins.

For internal validation of the construction algorithm, we compared glycosidic linkage input and output data. If the algorithm is performing correctly, $\Delta G(\phi, \psi)$ from the MD ensemble and the constructed ensemble *before minimization* will be nearly identical, and $\Delta G(\phi, \psi)$ from the constructed ensemble *after minimization* will not be substantially different. Performing the comparison between the ensemble of 40,000 20-mer conformations from the MD and a constructed ensemble of the same size confirmed this to be the case. Figure 3.3 demonstrates that $\Delta G(\phi, \psi)$ for GlcAβ1-3GalNAc and for GalNAcβ1-4GlcA glycosidic linkages were qualitatively identical when comparing the MD-generated and constructed ensembles. Quantitative analysis (Table 3.1) showed that the global minima (Min I) for both types of linkages and the secondary local minima (Min II and Min II′) for GalNAcβ1-4GlcA linkages were basically identical with 0° to 5° differences between the MD-generated input data and the constructed ensemble output data before minimization. The minimization in the construction algorithm, used to resolve any steric clashes, resulted in relatively minor changes in the location of the global minima, also ranging from 0° to 5°. Constructed ensemble glycosidic dihedral values after minimization were within 5° of the MD data, and $\Delta G(\phi, \psi)$ values were within 0.1 kcal/mol.

As detailed in Methods and discussed below, the minimization portion of the construction algorithm not only relieves any steric clashes, but also is used to detect bond-strain energies indicative of ring piercing. The close similarity of $\Delta G(\phi, \psi)$ for the constructed ensemble after minimization in comparison to the MD-generated ensemble suggests that few constructed conformations had large steric clashes, resulting in large conformational shifts after minimization. It also suggests that a few constructed conformations had ring-piercing events that necessitated their exclusion from the constructed ensemble altogether. Indeed, this was the case: during the creation of the 40,000-member constructed ensemble, only 18 conformations were excluded because they failed to meet the bond-strain energy criterion.

3.3.2. GlcA Ring Pucker Effects

An initial implementation of the construction algorithm used default force-field geometries for all GalNAc and GlcA rings. The result was that all GalNAc and GlcA rings in an algorithmically-constructed conformation had the same internal geometries. Ensembles constructed using this version of the algorithm had longer average end-to-end distances than MD-generated ensembles (Figure A.1), which meant that, on average, constructed conformations were overly extended. The default force-field ring pucker geometry for both types of monosaccharides was $^4C_1$. With that ring pucker, all βGlcA and all but one βGalNAc exocyclic functional groups are equatorial, and therefore the $^4C_1$ ring pucker is expected to be strongly preferred to other ring pucker geometries. To validate this simple approach to assigning ring pucker geometry, we computed C-P parameters of each monosaccharide ring in the MD-simulated 20-mer ensemble (10 * 40,000 = 400,000 ring conformations for each of the two monosaccharide types). As seen in NMR and force-field studies, the stable $^4C_1$ chair ring pucker was the principal conformer for both GlcA [220,259,306-308] and GalNAc [216,259,307,309] in the MD simulations, with slight deformations ($0° < \text{C-P } \theta < 30°$) (Figure 3.4.a,b). However, a small minority of GlcA ring conformers were skew-boat or boat, namely $^3S_1$, $B_{1,4}$, $^5S_1$, $^{2,5}B$, ^2S_O, $B_{3,O}$, $^1S_3$, $^{1,4}B$, and $^1S_5$ ($60° < \text{C-P } \theta < 120°$) (Figure 3.4.b). Studies that performed unbiased MD simulations with other force fields observed skew-boat and boat ring puckers of nonsulfated GlcA monosaccharides on the microsecond timescale but the occurrences were negligible due to high energy barriers [220,307]. In line with those findings, we observed only occasional GlcA skew-boat and boat pucker transitions in chondroitin 20-mers in our 500-ns unbiased CHARMM simulations. However, the C-P ϕ values in non-$^4C_1$ GlcA conformers in these studies differed from ours. Specifically, one study found ^2S_O, $B_{3,O}$, $^1S_3$, $^{1,4}B$, and $^1S_5$ [307]. Slight differences could be explained by differing ion concentrations which likely impacts pyranose ring puckers [310]. However, it is likely that the differences primarily result from intramolecular interactions. The aforementioned literature data come from simulated GlcA monosaccharides only, whereas our results come from simulated chondroitin 20-mers.

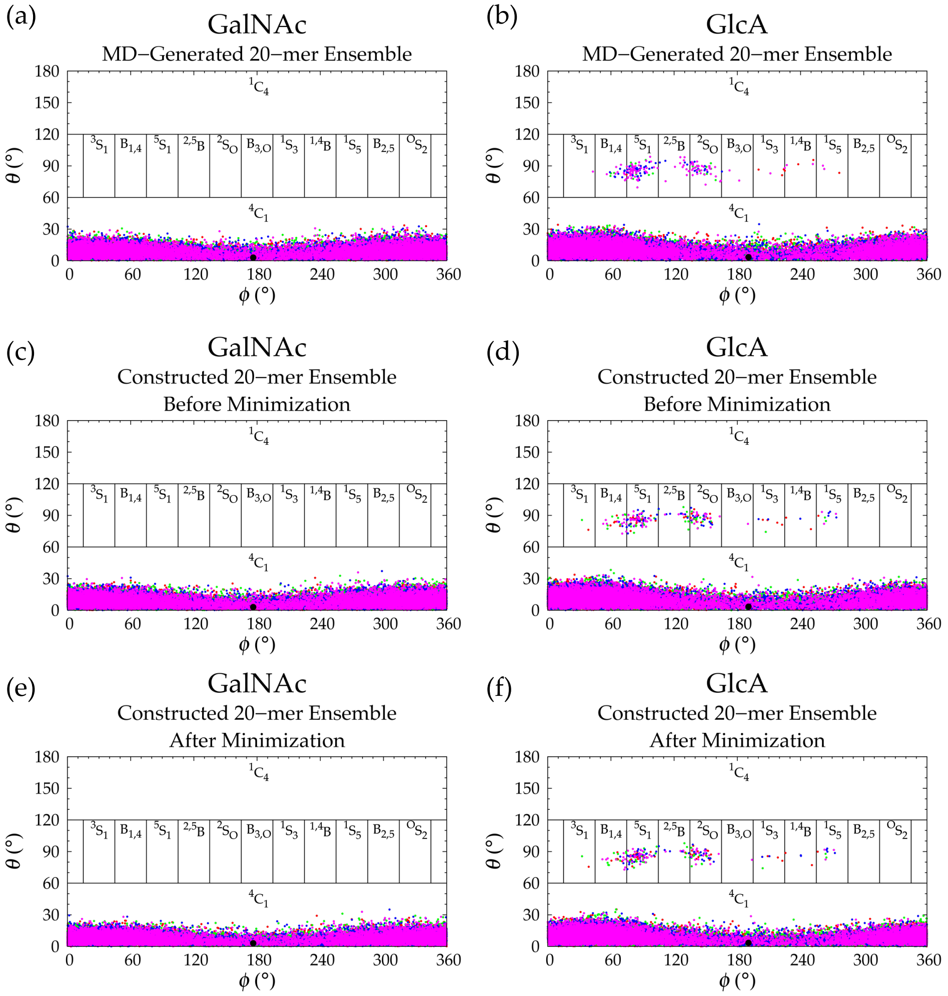


Figure 3.4. Cremer–Pople data for GalNAc and GlcA in (a,b) MD-generated ensembles and constructed ensembles (c,d) before and (e,f) after energy minimization; geometries from the four sets of each type of ensemble are represented by red, green, blue, and magenta dots, respectively and the force-field geometry is represented by a single large black dot. Cremer–Pople parameters (ϕ, θ) for all rings in every tenth snapshot from each ensemble were plotted (i.e., 10 rings * 1,000 snapshots per run * 4 runs = 40,000 parameter sets). As the algorithm reads all ring conformations sampled in MD, not all datapoints in panels c, d, e, and f are seen in panels a and b but the full MD-generated dataset contains all datapoints in the constructed ensembles.

The MD-generated 20-mer GlcA ring conformations can be separated into two broad categories: those that do not introduce a kink into the polymer and those that do. With the inclusion of $^4C_1$, the former category encompasses $^4C_1$ and ^2S_O GlcA ring puckers, both of which place the two glycosidic linkage oxygen atoms, located at opposite ends of the ring, in an equatorial conformation (Figure 3.5.a). As such, the O-C_1 and C_4-O bond vectors therein are approximately parallel and promote extended polymer conformations. The latter category encompasses $^3S_1$, $B_{1,4}$, $^5S_1$, $^{2,5}B$, $B_{3,O}$, $^1S_3$, $^{1,4}B$, and $^1S_5$ GlcA ring puckers (Figure 3.5.b). These ring puckers all place one of these glycosidic linkage oxygen atoms in the equatorial position and the other in the axial position. For these ring puckers, the O-C_1 and C_4-O bond vectors are approximately perpendicular, which results in a kink in the polymer chain, and can reduce end-to-end distance even when the remainder of the polymer is fully extended (Figure 3.5.b).

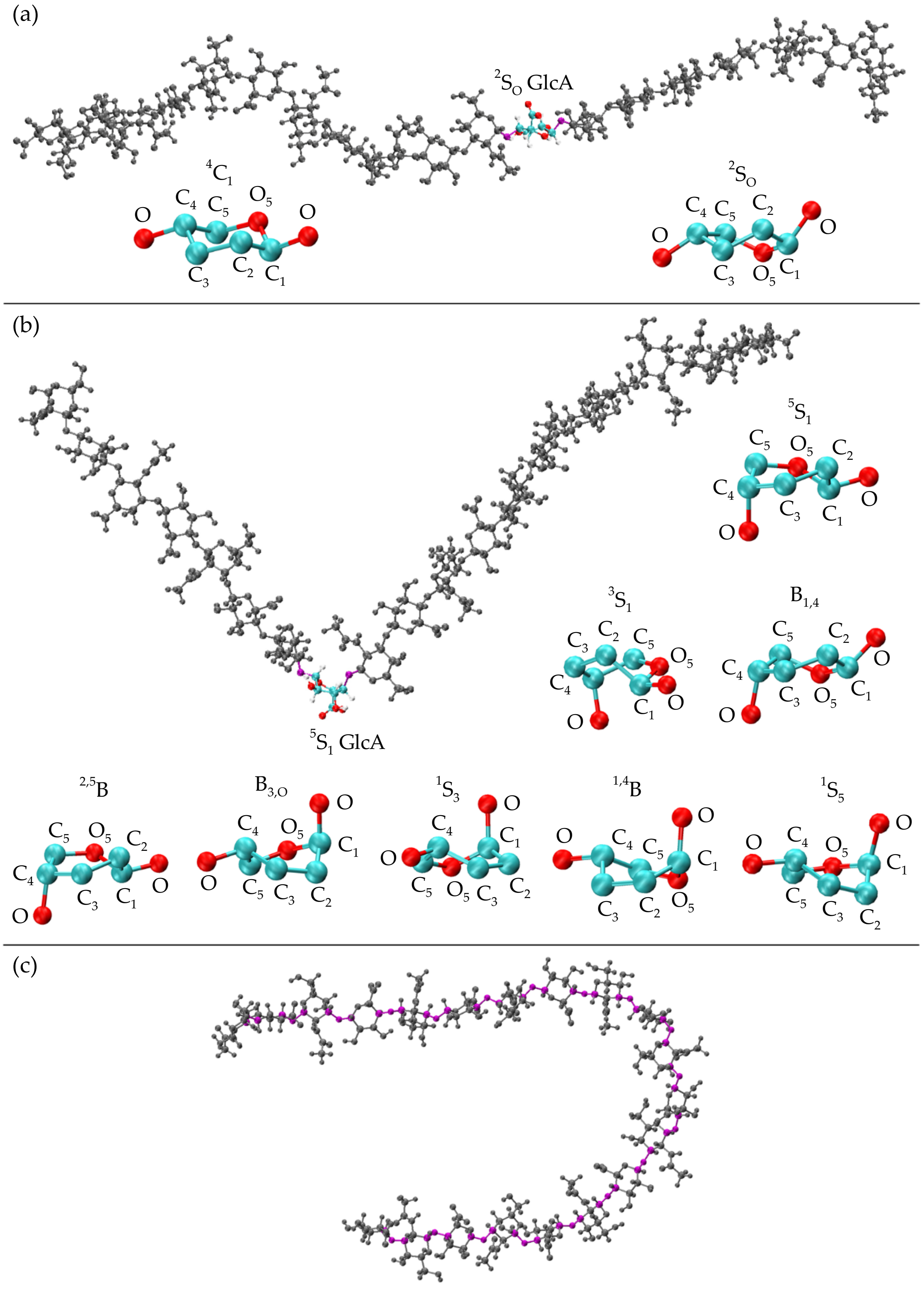


Figure 3.5. (a) 20-mer conformation with a ^2S_O GlcA conformer (colored by atom type with flanking linkage atoms highlighted in purple) and close-ups of GlcA monosaccharide rings in $^4C_1$ and ^2S_O conformations (shows endocyclic ring atoms and linker oxygen atoms only); (b) 20-mer conformation

with a kink at a $^{5}S_{1}$ GlcA conformer (colored by atom type with flanking linkage atoms highlighted in purple) and GlcA monosaccharide rings in $^{5}S_{1}$, $^{3}S_{1}$, $B_{1,4}$, $^{2,5}B$, $B_{3,O}$, $^{1}S_{3}$, $^{1,4}B$, and $^{1}S_{5}$ conformations (shows endocyclic ring atoms and linker oxygen atoms only); (c) 20-mer conformation with a curve caused by flexible glycosidic linkage geometries (highlighted in purple) and all monosaccharides in $^{4}C_{1}$ conformations; all images came from MD-generated ensembles.

As such, the final version of the construction algorithm uses MD-generated ring conformations instead of default force-field topology geometries. As an added benefit, this approach includes not only MD-generated ring dihedral angles but also bond lengths and angles. Using this finalized version of the algorithm, the peak in the end-to-end distance histogram for the constructed ensemble was shifted left compared to that resulting from force-field topology ring geometries (Figure 3.6 vs. Figure A.1) and much more closely matched the reference MD-generated ensemble data. This finding shows the importance of accounting for ring flexibility in constructing chondroitin glycosaminoglycan polymer conformations similar to those sampled in all-atom explicit-solvent MD simulations. Of note, the radius of gyration was also analyzed as a function of end-to-end distance in MD-generated and constructed ensembles after minimization (Figure A.2.a,b). These results showed that the radius of gyration is highly correlated with end-to-end distance in both MD-generated and constructed ensembles.

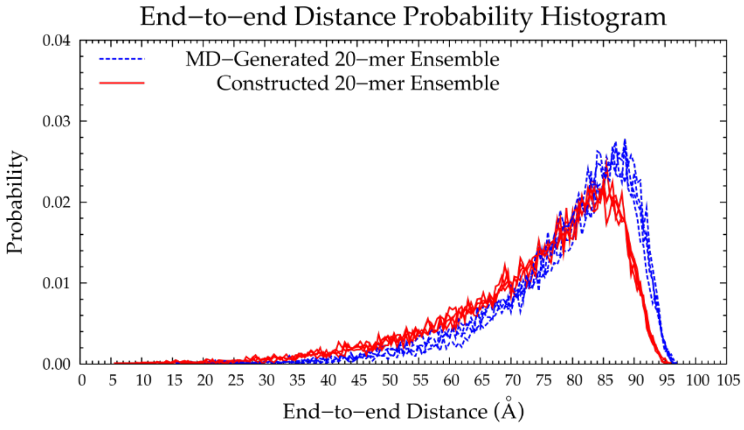


Figure 3.6. End-to-end distance probability distribution of MD-generated (blue dashed lines) and constructed (red solid lines) 20-mer ensembles; each type of ensemble includes four sets of 10,000 conformations; probabilities were calculated for end-to-end distances sorted into 0.5 Å bins.

While polymer kinks from ring puckering can lead to the shortening of polymer end-to-end distances, they are not required to achieve this. For 20-mers, flexibility in the glycosidic linkages even with $^4C_1$ ring puckering in all constituent monosaccharides can be sufficient to produce compact conformations (Figure 3.5.c). Furthermore, because of the flexibility in the glycosidic linkages flanking non-$^4C_1$ ring puckers, polymer kinks from ring puckering do not always lead to compact conformations. Thus, the leftward shift in the end-to-end distance histogram upon the inclusion of non-$^4C_1$ ring puckers supplements glycosidic linkage flexibility in yielding compact conformations.

3.3.3. Treating Glycosidic Linkage and Ring Pucker Geometries as Independent Variables

To determine if linkage geometries and ring deformations are interdependent, individual glycosidic linkage $\Delta G(\phi, \psi)$ plots (Figure A.3) were created (as opposed to the aggregate data $\Delta G(\phi, \psi)$ plots in (Figure 3.3) and no distinguishing patterns emerged. These per-linkage glycosidic linkage data were also examined in the context of C-P plots of adjacent rings (Figure A.4), for which there were also no distinguishing patterns. Additionally, ϕ and ψ values in linkages flanking GlcA rings not in a $^4C_1$ chair conformation were checked. For each linkage type, these conformations were all centered about the global $\Delta G(\phi, \psi)$ minima for the aggregate data (Figure 3.3) and 99.96% of conformations fell within the basin extending to $\Delta G(\phi, \psi) = +2$ kcal/mol (Figure 3.7). Furthermore, different types of non-$^4C_1$ chair conformers did not have unique flanking linkage geometries. As no connection between linkage and ring conformations was observed in this analysis of the MD data, each linkage conformation and ring pucker was treated independently in the construction algorithm.

(a) (b)

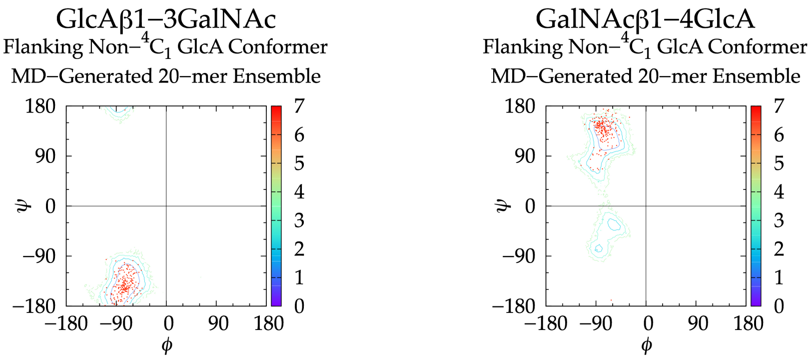


Figure 3.7. ΔG(ϕ, ψ) plots for glycosidic linkages flanking non-$^4C_1$ GlcA conformers in nonsulfated chondroitin 20-mer MD-generated ensembles: (a) GlcAβ1-3GalNAc and (b) GalNAcβ1-4GlcA.

3.3.4. Handling Nonphysical Constructed Conformations

To determine a criterion to exclude nonphysical conformations from constructed ensembles, energy minimization with restrained endocyclic ring and glycosidic linkage dihedrals was performed on each conformation, and post-minimization bond potential energy and bond length probability distributions were analyzed. Conformations with outlying bond energies and abnormally long bonds may point to ring piercing (Figure 3.8) that cannot be fixed by minimization. To confirm this possibility, the post-minimization conformations with bond energies greater than that of the fully-extended 20-mer conformation after minimization were visualized. As anticipated, among these conformations, most with outlying total bond energies contained pierced rings which were not resolved by minimization. For each of the 12 20-mer conformations with a pierced ring, the difference between the post-minimization bond energy and that of the fully-extended 20-mer conformation was greater than the predicted energy change caused by bond distortions of that pierced ring (Table A.2). The six conformations with outlying bond energies that did not contain pierced rings had kinks that resulted in bond length and bond angle distortions in glycosidic linkages that were nearly overlapping even after minimization. Of note, those conformations with bond energies that were not outlying were fully extended. These findings motivated using a bond potential energy cutoff in the construction algorithm. As stated previously, applying this cutoff resulted in 18 conformations being excluded during creation of the 40,000-member constructed ensemble. The resulting constructed ensemble contains no outlying bond lengths (Figure A.5).

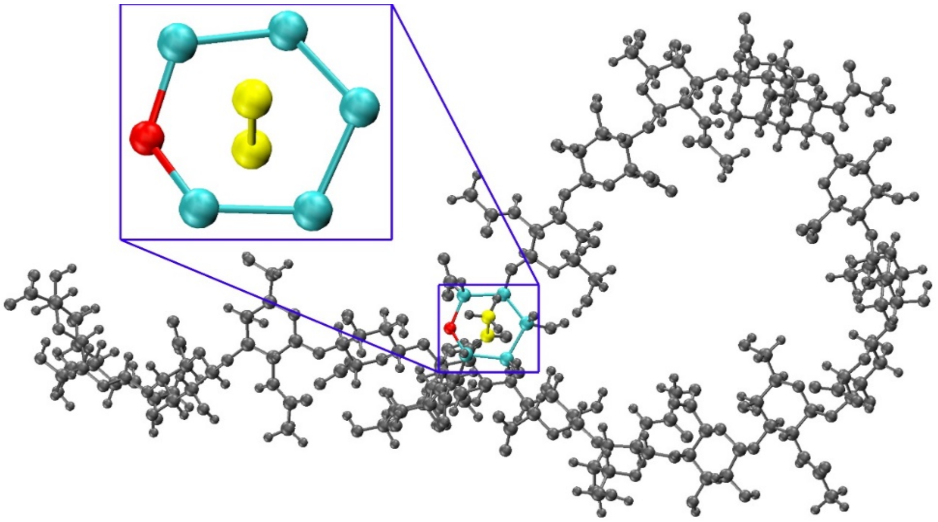


Figure 3.8. Constructed 20-mer conformation with a GlcA ring pierced by a GalNAc C-CT bond and a closeup panel showing atoms involved in the ring pierce; E_b = 787.7 kcal/mol, fully-extended 20-mer post-minimization E_b = 29.6 kcal/mol, ΔE_b = 758.1 kcal/mol (Table A.2).

Dihedral angles before and after energy minimization were compared by analyzing glycosidic linkage $\Delta G(\phi, \psi)$ (Table 3.1 and Figure 3.3.c-f), monosaccharide ring C-P parameters (Figure 3.4.c-f), and the change in ϕ and ψ dihedral angles due to minimization. All $\Delta G(\phi, \psi)$ and C-P plots after energy minimization matched those from before energy minimization and 99.6% of all ϕ and ψ dihedral angle differences before and after minimization were within 4° (Figure A.6). This confirmed that dihedrals did not undergo any major changes during minimization. Additionally, differences in end-to-end distance before and after minimization were calculated and the maximum change was 2.13 Å with 99.9% of changes under 0.5 Å, confirming that overall backbone conformation did not change as a result of minimization. According to these results, the selected bond potential energy cutoff and restraint scheme during minimization gave conformations with little deviation from the initial constructed conformations before minimization.

3.3.5. Internal Validation on 10-mers

To further validate the algorithm and test the use of conformational parameters from MD-generated 20-mer ensembles to construct polymers of variable length, we constructed nonsulfated chondroitin 10-mer ensembles and compared them to MD-generated 10-mer ensembles. All linkage and most ring conformations in MD-generated 10-mer (Figures 3.9 and A.7) and 20-mer ensembles matched (Figures 3.3.a,b and 3.4.a,b), with the exception of non-$^4C_1$ GlcA rings which were not sampled in 10-mer simulations. This finding, combined with the report from NMR and force-field studies suggesting that GlcA skew-boat and boat conformations are negligible in nonsulfated chondroitin mono- and oligosaccharides [220,307] suggests that these GlcA conformations may result from intramolecular interactions in longer GAG polymers.

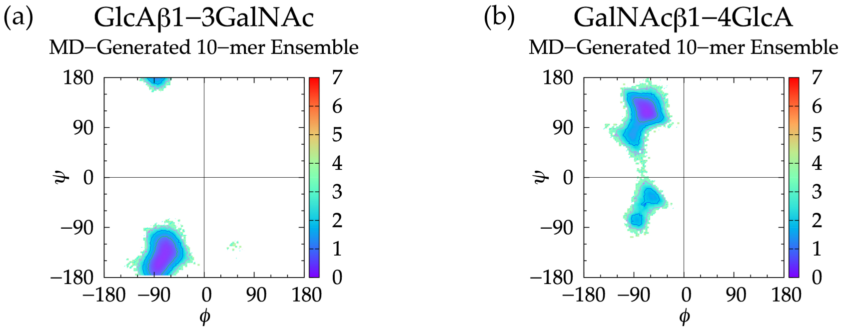


Figure 3.9. ΔG(ϕ, ψ) plots for each glycosidic linkage in nonsulfated chondroitin 10-mer MD-generated ensembles: (a) GlcAβ1-3GalNAc and (b) GalNAcβ1-4GlcA.

Additionally, the end-to-end distance distributions of constructed and MD-generated 10-mer ensembles matched with minimal difference in the most probable end-to-end distance (Figure 3.10 and Table 3.2). Further, the radius of gyration was highly correlated with end-to-end distance in both MD-generated and constructed ensembles (Figure A.2.c,d). Of note, the end-to-end distance distributions of MD-generated 20-mer ensembles more closely matched those of 20-mer ensembles constructed using MD-generated 20-mer conformations (Figure 3.6) than those of 20-mer ensembles constructed using MD-generated disaccharide conformations (Figures A.8 and A.9). Together, these findings suggest that MD-generated 20-mer conformational parameters are ideal for constructing chondroitin polymers of different lengths.

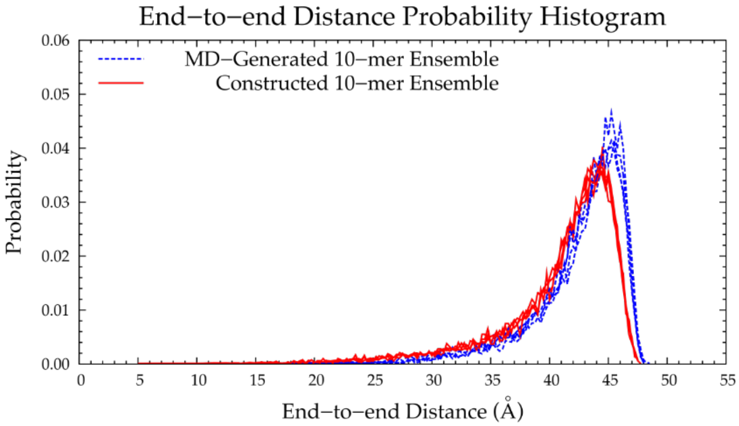


Figure 3.10. End-to-end distance probability distribution of MD-generated (blue dashed lines) and constructed (red solid lines) 10-mer ensembles; each type of ensemble includes four sets of 10,000 conformations; probabilities were calculated for end-to-end distances sorted into 0.25 Å bins.

Table 3.2. Most Probable End-to-End Distances (*d*) in MD-Generated and Constructed Ensembles [1].

	20-mer Ensembles			10-mer Ensembles		
	MD-Generated *d* (Å)	Constructed *d* (Å)	% Difference	MD-Generated *d* (Å)	Constructed *d* (Å)	% Difference
Run 1	88.5	83.0		45.25	44.50	
Run 2	88.5	85.5		45.25	43.50	
Run 3	86.0	85.0		45.50	44.50	
Run 4	86.5	85.0		45.25	44.25	
All [2]	88.5	85.0	4.03%	45.25	44.50	1.671%

[1] Probabilities were calculated for end-to-end distances sorted into 0.5 Å bins for the 20-mer ensembles and 0.25 Å bins for the 10-mer ensembles. [2] All = end-to-end distance distribution aggregated across all four runs.

3.3.6. Application to Longer Chondroitin Polymers

To implement the algorithm in the construction of conformational ensembles of nonsulfated chondroitin polymers of biologically-relevant chain lengths, chondroitin 100-mer and 200-mer ensembles were constructed and the end-to-end distance (Figures 3.11 and 3.12), radius of gyration (Figure A.2.e,f), and bond potential energy distributions (Figure A.10.c,d) were examined. The skewness of the end-to-end distance distributions shifted toward the right with increasing polymer length. This stands to reason, as there is a greater chance of folding with longer chains. This also explains why there were more frames excluded from these ensembles (i.e., 457 and 1407 frames excluded from the 100-mer and 200-mer ensembles, respectively). Bond potential energy distributions had similarly-shaped curves in all polymer lengths (Figure A.10) and energy values increased linearly as a function of atom count (Figure A.11). These results suggest that 100-mer and 200-mer conformational ensembles constructed using our algorithm (Figures 3.11 and 3.12) are reasonable predictions of biological conformations given their high number of degrees of freedom.

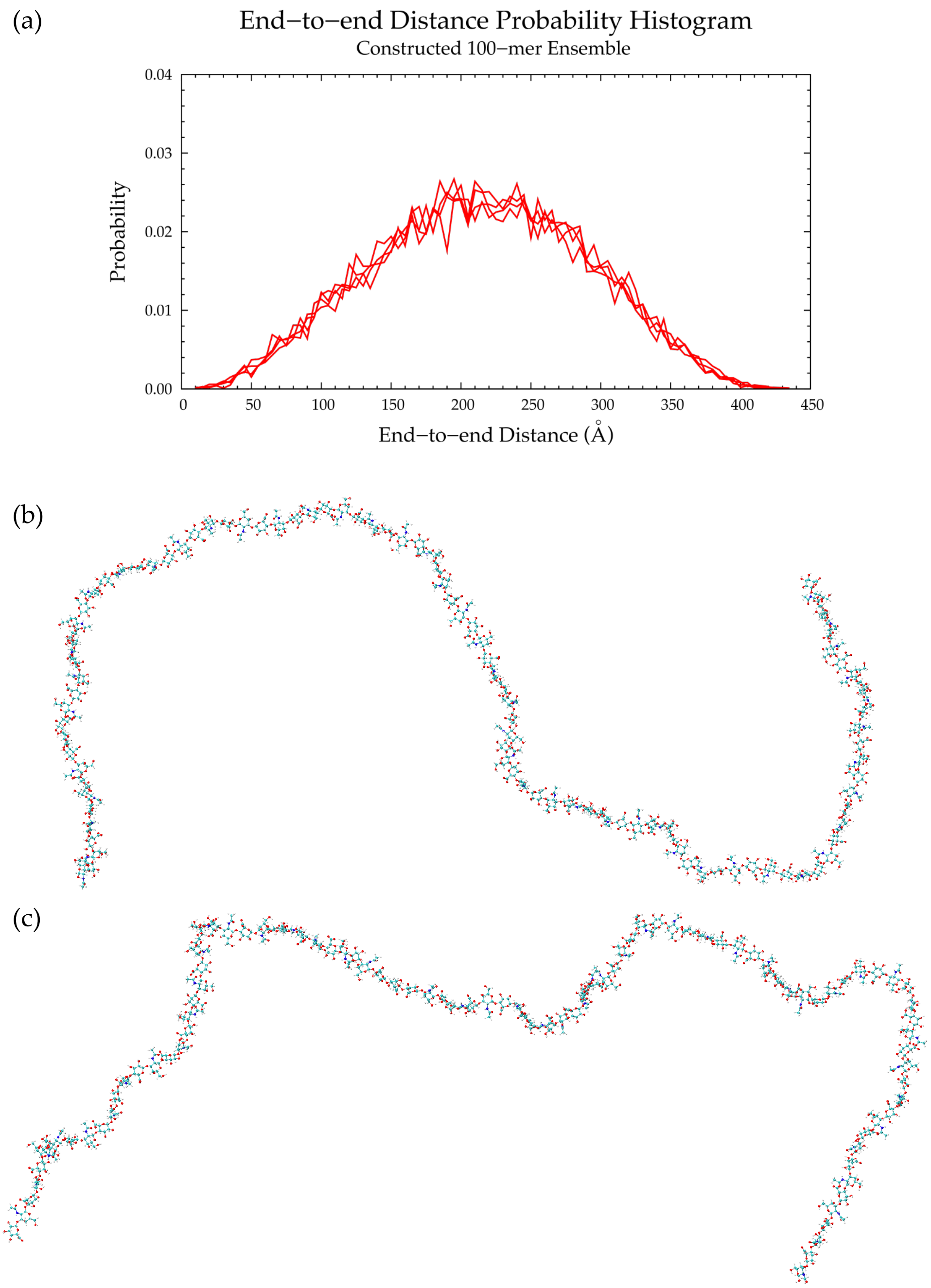


Figure 3.11. (a) End-to-end distance probability distribution of constructed 100-mer ensemble; includes four sets of 10,000 conformations; probabilities were calculated for end-to-end distances sorted into 5 Å bins. (b,c) Snapshots of the nonsulfated chondroitin 100-mer having the most-probable end-to-end distance (225 Å in both snapshots) from constructed ensembles.

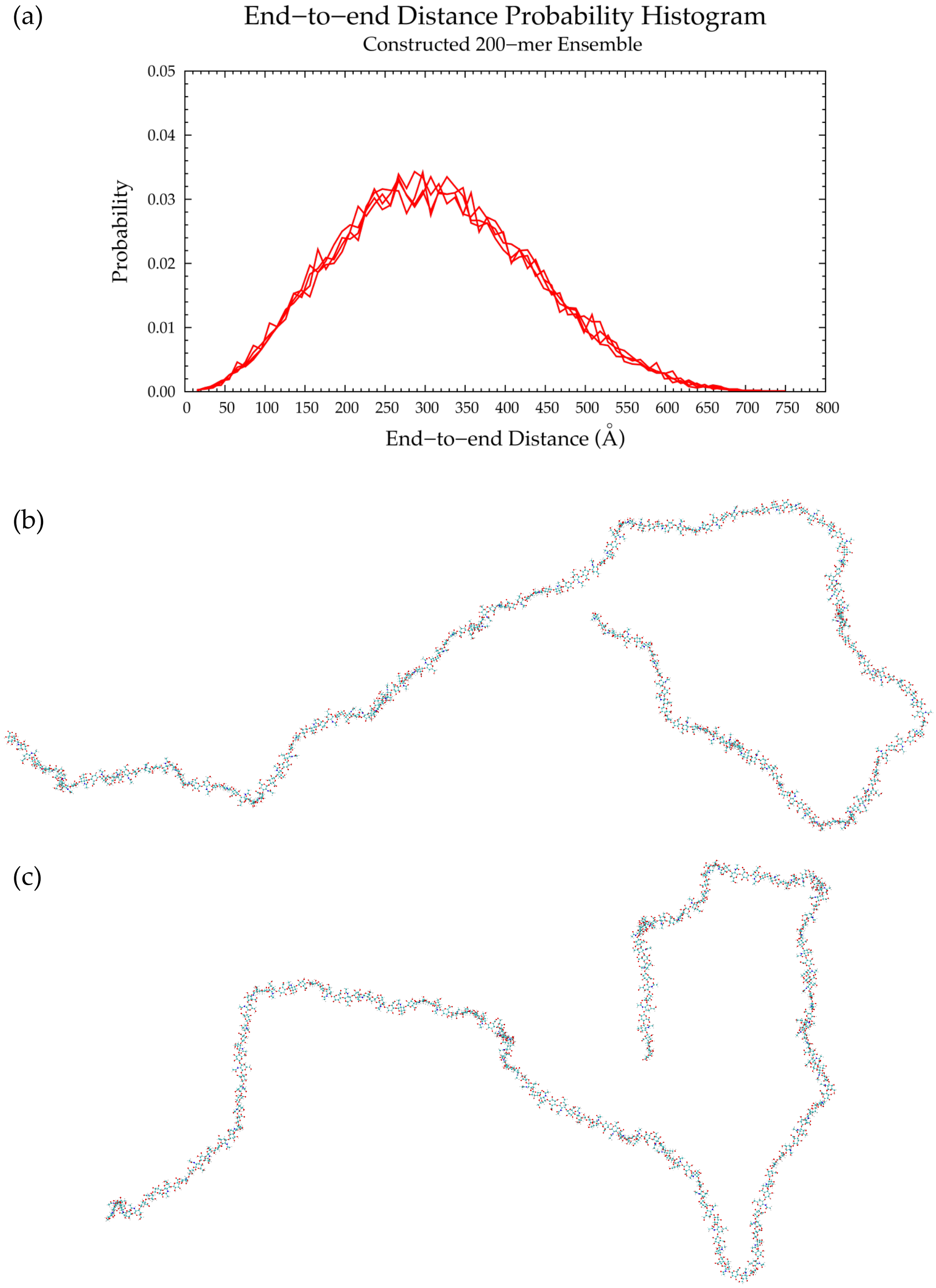


Figure 3.12. (a) End-to-end distance probability distribution of constructed 200-mer ensemble; includes four sets of 10,000 conformations; probabilities were calculated for end-to-end distances sorted into 10 Å bins. (b,c) Snapshots of the nonsulfated chondroitin 200-mer having the most-probable end-to-end distance (300 Å in both snapshots) from constructed ensembles.

3.4. Conclusions

With (1) all bond, bond angle, and dihedral angle conformational parameters from MD incorporated into the algorithm, (2) monosaccharide rings and glycosidic linkages treated independently, (3) energy minimization performed on each constructed conformation, and (4) a bond potential energy cutoff applied, end-to-end distance probability distributions from constructed and MD-generated ensembles matched with minimal differences in most probable end-to-end distances (Table 3.2 and Figures 3.6, 3.10, 3.11, and 3.12) suggesting that our algorithm produces conformational ensembles that mimic the backbone flexibility seen in MD simulations of nonsulfated chondroitin polymers.

Our program is also valuable for its efficiency. For example, the fully-solvated chondroitin 20-mer system contains ~191,000 atoms and took about one month to simulate. It took only about 12.5 h to construct the 20-mer conformational ensembles using our algorithm, in which all end-to-end distances, radii of gyration, bond lengths, dihedral angles, monosaccharide ring PDB files for C-P analysis of every tenth frame, and bond and system potential energies before and after minimization, C-P parameters of every GlcA ring in every frame, and PDBs of all conformations with bond energies greater than that of the fully-extended 20-mer (including conformations with pierced rings) were written. The fully-solvated chondroitin 10-mer system contains ~36,000 atoms and took about five days to simulate. The 10-mer ensembles were constructed using our algorithm in about 40 min and energies, dihedral angles, end-to-end distances, and radii of gyration after minimization, C-P parameters of every GlcA ring in every frame, and PDBs of all conformations with bond energies greater than that of the fully-extended 10-mer were written. Fully-solvated chondroitin 100-mer and 200-mer systems would contain ~3,370,000 atoms and ~75,450,000 atoms, respectively. Systems of this magnitude are not feasible to simulate with current computational resources but if they could be simulated, they would take on the order of years to complete. Construction of the chondroitin 100-mer and 200-mer conformational ensembles using our algorithm took about four and nine hours, respectively, and the algorithm produced the same output data types as for the 10-mer (n.b.: these timings are less than the 20-mer timing above because of all of the additional output written to disk for analysis in the case of the 20-mer ensemble construction).

In conclusion, our algorithm, incorporating glycosidic linkage and monosaccharide ring conformations from the MD simulation of nonsulfated chondroitin 20-mers, can be used to efficiently generate conformational ensembles of nonsulfated chondroitin polymers of arbitrary length. We are investigating the applicability of this approach to various sulfo-forms of CS and different types of GAGs, including hyaluronan (HA), dermatan sulfate (DS), heparan sulfate (HS), and keratan sulfate (KS). Given the variability and complexity of GAGs, as well as existing barriers to the experimental characterization of the three-dimensional conformational properties of GAGs of lengths relevant in the context of PGs, there are currently very few efforts to target GAGs. We anticipate that the presented algorithm, combined with experimental data on PG core proteins and conformational analysis of the linker tetrasaccharide [210,218,311-313], may provide a useful means of generating atomic-resolution three-dimensional models of full PGs. The algorithm could also be used to model full GAG–protein complexes, which may provide insights into potential interactions between multiple biomolecules within a single GAG complex. The ability to model these complex biomolecules would be a key step towards improving understanding of GAG bioactivity, assessing the druggability of GAGs, designing agonists or antagonists to treat disease, and developing diagnostic tools. Thus, this methodology may open a new avenue into disease modulation.

CHAPTER 4

CONSTRUCTING 3-DIMENSIONAL ATOMIC-RESOLUTION MODELS OF NONSULFATED GLYCOSAMINOGLYCANS WITH ARBITRARY LENGTHS USING CONFORMATIONS FROM MOLECULAR DYNAMICS

This chapter has been published as:

Whitmore, E.K.; Martin, D.; Guvench, O. Constructing 3-Dimensional Atomic-Resolution Models of Nonsulfated Glycosaminoglycans with Arbitrary Lengths Using Conformations from Molecular Dynamics. *International Journal of Molecular Sciences*. **2020**, *21*(20), 7699. doi:10.3390/ijms21207699

4.1. Introduction

Proteoglycans (PGs) are a diverse group of proteins in the extracellular matrix (ECM), as well as on and within cells in animal tissue. PGs play key roles in signal transduction [9,25], tissue morphogenesis [1-3,5,265], and matrix assembly [5,47,92,93] by binding growth factors [1-3,5,11,17-22,265], enzymes [5,22], membrane receptors [22], and ECM molecules [5,9,22]. Many of these functions are mediated by glycosaminoglycans (GAGs), which are linear, highly negatively charged, and structurally diverse carbohydrate chains covalently linked to PGs. Specifically, GAGs can form covalent and noncovalent complexes with proteins or inhibit complex formation with other biomolecules. All of these functions allow GAGs to modulate disease. For example, hyaluronan can predict disease outcome and is used as a treatment for osteoarthritis [314,315]; deficiency of dermatan sulfate has been linked to Ehlers-Danlos syndrome, so screening of dermatan sulfate in urine may present a method of diagnosis [169,170]; reduced cerebral cell keratan sulfate levels have been implicated in Alzheimer's disease [142]; and heparan sulfate has been shown to induce septic shock [173,267].

GAG-protein binding is determined by the sequence, structure, shape, and charge of the GAG and GAG-binding region of the protein [22,41]. Therefore, the three-dimensional structure and conformation of GAGs determine their bioactivity, and even subtle structural differences can change their function. For example, while hyaluronan and chondroitin have many functional differences, the only structural difference is in the chirality of carbon four in the amide monosaccharide derivatives (i.e., N-acetylglucosamine and N-acetylgalactosamine, respectively). The structural and conformational complexities of GAGs make studying GAG conformational thermodynamics at atomic resolution a challenge using existing experimental methods. Structural complexities arise from nontemplate-based synthesis [37] and variable sulfation [270], which means a single GAG type can have variable lengths and sulfation patterns, making it polydisperse and heterogeneous. Conformational complexities result from glycosidic linkage flexibility [188-193] and ring puckering [188]. There are several experimental techniques for studying GAGs, including liquid chromatography-mass spectrometry (LC-MS) [271-273], X-ray crystallography [200-205], and nuclear magnetic resonance (NMR) [195,215,274,275], but no single technique is able to account for all of these complexities. While each has its limitations, these experimental techniques provide a critical means to validate molecular dynamics (MD) simulations [215-221]. MD can be used to generate atomic-resolution conformational ensembles of GAGs under physiological conditions, and agreement between MD-generated results and available experimental data builds confidence that MD does indeed generate realistic three-dimensional atomic-resolution GAG models [210-214]. However, all-atom explicit-solvent MD is not a feasible tool for routine simulation of large systems, and GAG biopolymers may consist of up to 200 monosaccharide units [2], which, when fully solvated, would result in a system with over 10^6 atoms. Coarse-grained (CG) MD is a feasible alternative for simulations of long GAG polymers, but may be limited because it lacks the atomic resolution and the accurate treatment of solvation provided by all-atom explicit-solvent MD [218,253-255].

As an alternative approach to obtaining atomic-resolution, three-dimensional, conformational ensembles of long GAG biopolymers, we developed an algorithm that uses glycosidic linkage and monosaccharide ring conformations from unbiased all-atom explicit-solvent MD simulations [214,243-245] of GAG 20-mers (i.e., 20 monosaccharide units) to rapidly generate GAG polymer conformational ensembles of user-specified length and with a user-specified number of conformations [188]. In our previous work, we examined nonsulfated chondroitin GAGs. Here, we aim to expand the application of our algorithm to hyaluronan and to nonsulfated forms of dermatan, keratan, and heparan. Nonsulfated forms were chosen for their simplicity and homogeneity. The algorithm is expected to produce conformational ensembles that mimic the backbone flexibility observed in simulations only if individual ring and linkage conformations of GAG 20-mers behave randomly and independently in simulation. This means that the algorithm will not produce polymer conformations with higher-order structures.

It has been suggested that hyaluronan may take on higher-order structures [316] and while it has been shown that helical structures occur in hyaluronan oligosaccharides [317,318], this behavior may not necessarily be extrapolated to long polysaccharides under physiological conditions. Hyaluronan oligosaccharides have been shown to exhibit rigid structures in aqueous solution, possibly resulting from electrostatic effects and hydrogen-bonding between neighboring monosaccharides [257,309,317,319-323]. MD studies of hyaluronan di- and tri- saccharides in aqueous solution (without ions) revealed that hydrogen-bonds form but do not remain for long periods of time, and do not always occur at the same time in different regions of the oligosaccharide [257,319]. It has been shown that under physiological conditions, the extended conformations and large hydrodynamic radius of hyaluronan can be attributed primarily to hydrogen-bonding between adjacent monosaccharides (and less so to electrostatic effects) [320]. It is important to note that each of these studies examined short hyaluronan oligosaccharides; thus, atomic-resolution conformational properties of long polysaccharides under physiological conditions have not been reported.

We sought to determine (1) if longer polysaccharides of hyaluronan and nonsulfated dermatan, keratan, and heparan took on a higher-order structure during MD simulations in an aqueous solution with physiological ion strength and, consequently, (2) if our algorithm could construct conformational ensembles of GAGs that mimic the backbone flexibility observed in MD simulations. To this end, we ran a microsecond-scale, all-atom, explicit-solvent MD on 20-mers of each GAG type with sequences [-4 glucuronate β1-3 N-acetylglucosamine β1-]$_{10}$ (hyaluronan), [-4 iduronate α1-3 N-acetylgalactosamine β1-]$_{10}$ (dermatan), [-3 galactose β1-4 N-acetylglucosamine β1-]$_{10}$ (keratan), and [-4 iduronate α1-4 N-acetylglucosamine α1-]$_{10}$ (heparan). Analysis of glycosidic linkage and monosaccharide ring conformations revealed no apparent correlation between individual rings and flanking linkages in any of the 20-mers. Glycosidic linkages took on experimentally-observed helical conformations, and some linkages sampled secondary conformations that caused a kink in the polymer. No patterns of interdependence between adjacent linkage conformations were seen. While ring and linkage conformations did not reach equilibrium in hyaluronan simulations on the microsecond timescale, overall hyaluronan 20-mer appeared to behave randomly. End-to-end distance distributions revealed mostly extended conformations (n.b.: we use the term "extended" throughout to describe conformations with high end-to-end distances; this does not necessarily contradict helical structure) with some compact conformations resulting from non-$^4C_1$ ring puckering, which brought the oxygen atoms in sequential glycosidic linkages closer together, and certain glycosidic linkage conformations that brought neighboring monosaccharides closer together. Thus, these ring and linkage conformations both introduce a kink in the polymer. These conformations are independent and occur rarely on the microsecond timescale. The variability in individual ring and linkage conformations, as well as end-to-end distances and the radii of gyration, demonstrate the flexibility and lack of stable higher-order structures of each of these nonsulfated GAG polymers during simulation in aqueous solution.

For each nonsulfated GAG type, monosaccharide ring and glycosidic linkage conformations from MD-generated 20-mers were used as inputs in our algorithm to generate conformational ensembles of 10-, 20-, and 200-mers. For each GAG type, 10- and 20-mer constructed ensembles were compared to 10- and 20-mer MD-generated ensembles. In each case, end-to-end distances in constructed and MD-generated ensembles match, with minimal differences in most probable end-to-end distances. The similarities in backbone conformation of constructed and MD-generated 10- and 20-mer ensembles provide further evidence that there is little correlation between individual glycosidic linkage and ring conformations, and thus, that nonsulfated GAG 10- and 20-mers do not exhibit a higher-order structure. This suggests that our program can be used to construct realistic conformations of long GAG biopolymers (e.g., 200 monosaccharides), and the resulting atomic-resolution models may be used to develop further insights into GAG binding properties, and thus, GAG-protein complex formation. These will be especially useful for studying complexes in which there are multiple biomolecules bound to the same GAG.

Other programs exist which are capable of constructing three-dimensional, atomic-resolution GAG polymer models including Glycam GAG Builder [277], POLYS Glycan Builder [278], CarbBuilder [279], MatrixDB GAG Builder [280,281], and CHARMM-GUI [324-327]. These are useful tools for constructing the initial structures of GAGs having user-specified sequences and lengths for MD simulations. Additional features of Glycam and POLYS Glycan Builder include user specification of glycosidic linkage dihedral angle values or construction using conformations from their databases. Glycam, POLYS Glycan Builder, and Carb Builder databases consist of GAG mono- and di- saccharide structures sampled in MD and/or determined by molecular mechanics calculations. The MatrixDB databases contain GAG disaccharide conformations from GAG-protein complex crystal structures. CHARMM-GUI uses the CHARMM software [229-231] to construct GAGs with user-specified sequences, using default force field parameters for carbohydrates [214,243-245], adding explicit solvents with user-specified water box size and ion concentrations, and generating inputs for MD equilibration and production simulations. Although the user has the option to choose the GAG length, these tools are best used for short GAG oligosaccharides. Our algorithm differs from these tools in that it draws upon our databases containing full MD-generated unbound GAG 20-mer atomic-resolution conformational ensembles. Furthermore, our algorithm is developed for modeling long GAG biopolymers (e.g., 200-mers), and can efficiently produce large conformational ensembles (e.g., 10,000 3-D models) of long GAG biopolymers. Therefore, our algorithm provides an alternative to MD simulation that takes much less time and computational resources.

4.2. Materials and Methods

The following protocol was conducted for each of the hyaluronan and nonsulfated dermatan, keratan, and heparan GAGs.

4.2.1. Molecular Dynamics: System Construction, Energy Minimization, Heating, and Production Simulations

System construction, energy minimization, and heating followed the same protocol we described previously [188]. Briefly, all polymers were constructed with an explicit solvent using the CHARMM software [229-231] v. c41b2 with the CHARMM36 (C36) biomolecular force field for carbohydrates [214,243-245]. The initial conformation of each GAG 10- and 20-mer had glycosidic linkage dihedrals (with IUPAC definitions: $\phi = \mathrm{O}_5\text{-}\mathrm{C}_1\text{-O-}\mathrm{C}_n$ and $\psi = \mathrm{C}_1\text{-O-}\mathrm{C}_n\text{-}\mathrm{C}_{(n-1)}$) set to $\phi = -83.75°$ and $\psi = -156.25°$ in all GlcAβ1-3GlcNAc (hyaluronan), IdoAα1-3GalNAc (dermatan), and Galβ1-4GlcNAc (keratan) linkages and $\phi = -63.75°$ and $\psi = 118.75°$ in all GlcNAcβ1-4GlcA (hyaluronan), GalNAcβ1-4IdoA (dermatan), and GlcNAcβ1-3Gal (keratan) linkages. These dihedral values were the most energetically stable in nonsulfated chondroitin disaccharide glycosidic linkages during MD simulation [212]. In heparan 10- and 20-mers, all IdoAα1-4GlcNAc glycosidic linkage dihedrals were set to $\phi = -69°$ and $\psi = 147°$ and all GlcNAcα1-4IdoA linkage dihedrals were set to $\phi = 77°$ and $\psi = 30°$. These dihedral values were found in an NMR and force field study of heparan disaccharides [191] and gave extended heparan polymer conformations. With these glycosidic linkage conformations, 20-mer end-to-end distances were 99.0 Å, 89.7 Å, 85.8 Å, and 87.5 Å for hyaluronan, dermatan, keratan, and heparan, respectively. GAG 20-mers were solvated in cubic periodic unit cells with edge lengths of 124.3 Å (63,347 water molecules), 115.0 Å (50,066 water molecules), 111.9 Å (46,103 water molecules), and 111.9 Å (46,122 water molecules) for hyaluronan, dermatan, keratan, and heparan, respectively. Each of these edge lengths is at least 10 Å longer than the maximum dimension of the corresponding 20-mer in a fully-extended (i.e., maximum end-to-end distance) conformation. The TIP3 water model [282,283] as implemented in the CHARMM force field, was used for all simulations, and neutralizing Na^+ counterions and 140 mM NaCl [328] were added.

The NAMD program [284] v. 2.12 (http://www.ks.uiuc.edu/Research/namd/) was used to minimize the potential energy and heat the system to the target temperature of 310 K under constant pressure at 1 atm. An unbiased constant particle number/constant pressure/constant temperature (NPT) MD run followed, and average periodic cell parameters from the last half of this trajectory were used as cell basis vectors for the quadruplicate canonical (NVT) ensemble MD simulations. NVT MD simulations were preceded by minimization and heating at constant volume with box edge lengths of 123.7 Å, 114.4 Å, 111.3 Å, and 111.3 Å for hyaluronan, dermatan, keratan, and heparan, respectively.

Production simulations were performed using a similar protocol as previously described [188] but with a longer timescale and lower frequency for saving snapshots. Briefly, using the CHARMM software, each of four replicates (per GAG system) was equilibrated. Then, unbiased canonical (NVT) ensemble MD simulations were run for 500,000,000 steps (1 μs) and atomic coordinates were saved at 50,000-step (100-ps) intervals for analyses (10,000 snapshots per simulation, 40,000 snapshots per GAG).

4.2.2. Conformational Analysis

GAG conformational parameters include bond length, bond angle, and dihedral angle values. The glycosidic linkage conformational parameters considered were bond lengths for C_1-O and O-C_n, bond angles for O_5-C_1-O and C_1-O-C_n, and dihedral angles ϕ and ψ (Figure 4.1). To characterize potential patterns in glycosidic linkage conformations, dihedral angle free energies $\Delta G(\phi, \psi)$ were analyzed. As ϕ, ψ dihedral values from MD-generated ensembles were taken to have uniform probabilities, $\Delta G(\phi, \psi)$ was calculated by sorting dihedral angle values into 2.5° × 2.5° bins and then using the formula $\Delta G(\phi_i, \psi_j) = -RT\ln(n_{ij}) - k$, where n_{ij} is the bin count for the bin corresponding to ϕ_i, ψ_j, R is the universal gas constant, T

is the temperature of the MD simulations, and k was chosen so that the global minimum was located at $\Delta G = 0$ kcal/mol.

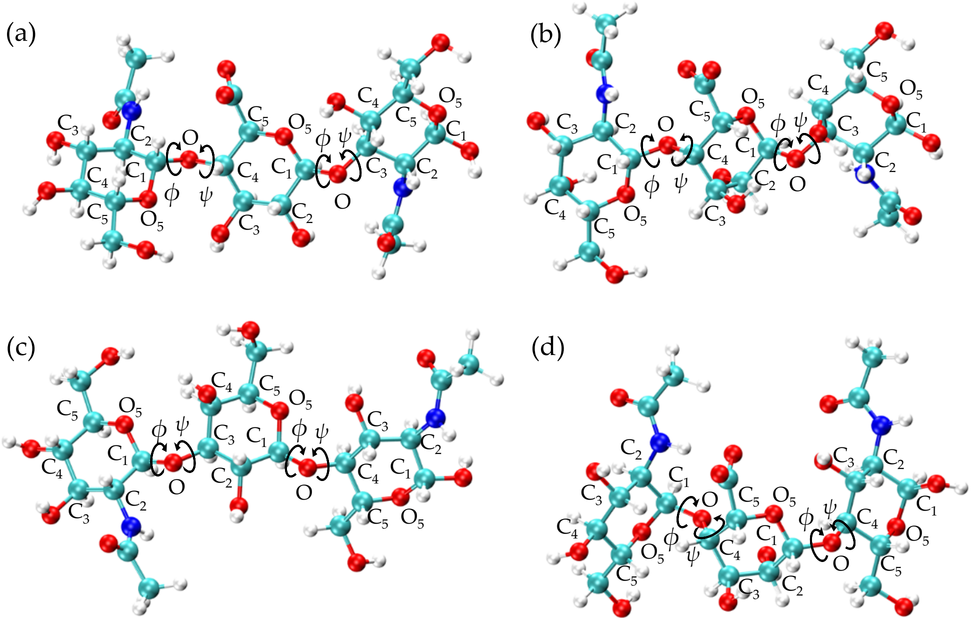


Figure 4.1. GAG trisaccharides: (a) hyaluronan [GlcNAcβ1-4GlcAβ1-3GlcNAc], (b) nonsulfated dermatan [GalNAcβ1-4IdoAα1-3GalNAc], (c) nonsulfated keratan [GlcNAcβ1-3Galβ1-4GlcNAc], (d) nonsulfated heparan [GlcNAcα1-4IdoAα1-4GlcNAc]; endocyclic ring atoms and oxygen atoms linking adjacent rings are labeled; arrows indicate rotations about glycosidic linkage bonds C$_1$-O and O-C$_n$ (quantified by ϕ and ψ dihedral angles, respectively); glycosidic linkage parameters used in the construction algorithm include C$_1$-O and O-C$_n$ bond lengths, O$_5$-C$_1$-O and C$_1$-O-C$_n$ bond angles, and ϕ = O$_5$-C$_1$-O-C$_n$ and ψ = C$_1$-O-C$_n$-C$_{(n-1)}$ dihedral angles. Molecular graphics throughout are produced with the VMD program [276].

Monosaccharide ring geometries are quantified by all bond lengths, bond angles, and dihedral angles within the ring and in exocyclic functional groups that are not part of a glycosidic linkage. Cremer-Pople (C-P) ring puckering parameters (ϕ, θ, Q) [261] of each monosaccharide ring in the MD-generated 20-mer ensembles were computed to characterize potential ring puckering patterns. Conformations of each monosaccharide and each linkage of the 20-mers were extracted separately from each snapshot of the MD trajectories. First, to determine if conformational data in different runs matched, and if there was interdependency in the individual linkage and ring conformations, data were separated by run and monosaccharide/linkage number and aggregated across all snapshots in each MD simulation run. Next, all individual conformations were aggregated across all snapshots in all runs (e.g., 10,000 snapshots * 4 runs * 10 GlcA monosaccharides = 400,000 samples of GlcA monosaccharide conformations in hyaluronan 20-mer) to generate a single dataset for each of the two monosaccharide types and two linkage types in each GAG (e.g., GlcA monosaccharide, GlcNAc monosaccharide, β1-3 glycosidic linkage, and β1-4 glycosidic linkage conformations in hyaluronan 20-mer).

4.2.3. Construction Algorithm to Generate GAG Conformational Ensembles

The MD-generated conformational datasets described above made up the database used for the algorithm we developed to generate GAG polymer conformational ensembles of user-specified length and with a user-specified number of conformations (previously described) [188]. Essentially, our algorithm (1) incorporates all bond, bond angle, and dihedral angle conformational parameters from MD simulation of GAG 20-mers, (2) treats monosaccharide rings and glycosidic linkages independently, (3) performs a restrained energy minimization on each constructed conformation to relieve steric overlap while maintaining polymer conformation, and (4) applies a bond potential energy cutoff to exclude conformations that remain nonphysical after minimization. A nonphysical conformation results from either overlapping bonds or a bond that pierces the center of another monosaccharide ring in the initial constructed conformation (Figure B.1). As dihedral angles are restrained, minimization addresses these issues by stretching the overlapping or ring-piercing bonds to nonphysical lengths, which increases the post-minimization bond potential energy by more than 132 kcal/mol [188]. The energy cutoff is the sum of a 100 kcal/mol buffer and a polymer-length specific cutoff, which is equal to the post-minimization bond potential energy of the fully-extended conformation. This conformation was constructed by assigning energetically-favorable glycosidic linkage dihedrals from the corresponding MD simulations of each GAG type that give a fully-extended conformation (i.e., with the maximum end-to-end distance observed in simulation). We used the algorithm to construct 20-mer conformational ensembles of each nonsulfated GAG.

For internal validation of the algorithm, glycosidic linkage dihedral free energies $\Delta G(\phi,\psi)$, monosaccharide ring C-P parameters, end-to-end distance distributions, and radii of gyration from MD-generated ensembles and constructed ensembles were compared. Additionally, post-minimization bond potential energy distributions from constructed ensembles were analyzed to validate the calculated bond potential energy cutoff and verify that the ensemble had expected energy distributions for the given polymer length.

To evaluate application of MD-generated 20-mer conformations to construct GAG polymers of variable length, we constructed 10-mer ensembles using the algorithm and compared them to 10-mer ensembles generated using the same protocol as 20-mer MD simulations. To assess the efficacy and efficiency of our algorithm to construct conformational ensembles of GAG polymers with biologically-relevant chain lengths, we also constructed 200-mer ensembles.

4.3. Results and Discussion

4.3.1. Hyaluronan

4.3.1.1. Molecular Dynamics Simulations: Glycosidic Linkage and Monosaccharide Ring Geometry Effects on Polymer Backbone Flexibility. To examine the backbone flexibility of hyaluronan 20-mers, end-to-end distances and radii of gyration were analyzed in each of the four MD simulation runs. One of the four runs produced more compact conformations, i.e., with lower end-to-end distances and radii of gyration than the other three (Figures 4.2 and B.2.a and Table 4.1). The system potential energy distributions were identical for all four runs, suggesting that this outlying run was energetically stable (Figure B.3). To uncover the conformational factors that contributed to end-to-end distance and explain the differences in backbone flexibility in the different MD runs, the monosaccharide ring and glycosidic linkage conformations were examined.

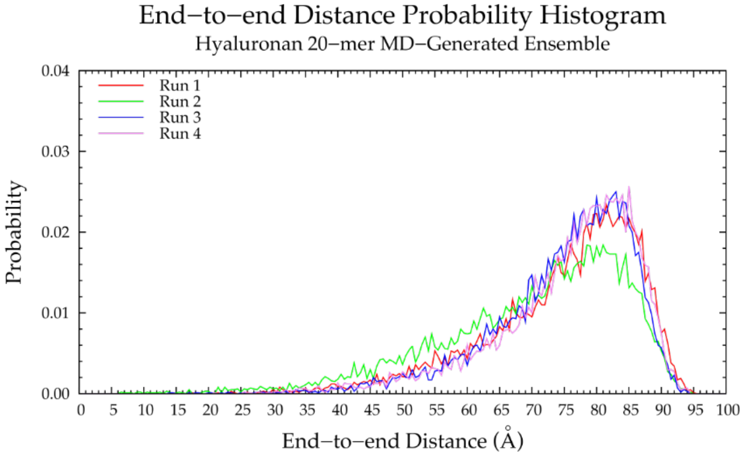


Figure 4.2. End-to-end distance probability distribution of MD-generated hyaluronan 20-mer ensemble; each of the four runs includes 10,000 conformations; probabilities were calculated for end-to-end distances sorted into 0.5 Å bins.

Table 4.1. Most Probable End-to-End Distances (*d*) in MD-Generated and Constructed Hyaluronan Ensembles [1].

	20-mer Ensembles			10-mer Ensembles		
	MD-Generated *d* (Å)	Constructed *d* (Å)	% Difference	MD-Generated *d* (Å)	Constructed *d* (Å)	% Difference
Run 1	81.5	80.5		43.00	42.25	
Run 2	81.0	78.0		42.50	41.75	
Run 3	83.0	80.5		44.00	42.75	
Run 4	85.0	78.0		43.50	42.75	
All [2]	80.0	78.0	2.53%	43.50	41.75	4.11%

[1] Probabilities were calculated for end-to-end distances sorted into 0.5 Å bins for the 20-mer ensembles and 0.25 Å bins for the 10-mer ensembles. [2] All = end-to-end distance distribution aggregated across all four runs.

Analysis of C-P parameters of GlcNAc monosaccharide rings revealed mostly $^4C_1$ chair conformations, as found in crystal structures [329-336] and NMR and force field studies [257,337-339], with a minority of samples in boat and skew-boat conformations, namely $B_{3,O}$, $^1S_3$, $^{1,4}B$, and $^1S_5$ (Figures 4.3.a and 4.4.a). While non-$^4C_1$ conformations of GlcNAc are rare, they have been observed in crystal structures of protein co-complexes [332,340-343]. For example, one GlcNAc ring in the crystal structure of a hyaluronan tetramer in complex with hyaluronidase was found in the $^{1,4}B$ boat conformation [332]. Additionally, GlcNAc boat/skew-boat conformations have been sampled in biased MD [307], 1-μs unbiased MD [344], and 10-μs unbiased MD simulations of GlcNAc monosaccharides [337]. One NMR and force field study suggested that these conformations were important for GAG-protein interactions [337]. Analysis of individual GlcNAc rings in each run revealed that these non-$^4C_1$ conformations were not specific to any particular region of the 20-mer and did not occur simultaneously in multiple rings in the same snapshot (Figure B.4). Furthermore, GlcNAc rings that sampled non-$^4C_1$ puckers returned to $^4C_1$ chair conformations relatively quickly, i.e., in under 25 ns (Figure B.4).

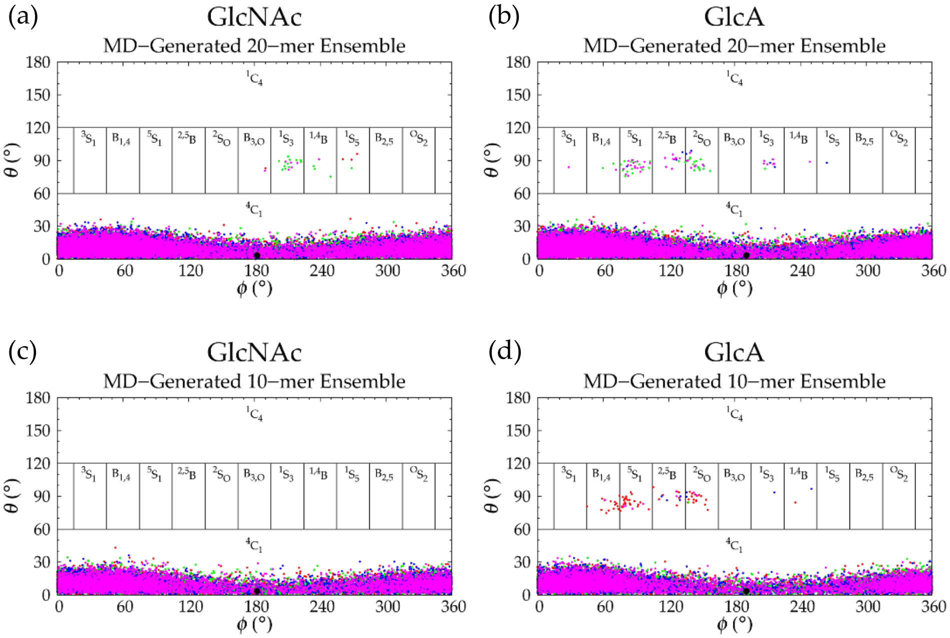


Figure 4.3. Cremer–Pople data for MD-generated hyaluronan ensembles: (a) GlcNAc and (b) GlcA in the 20-mer and (c) GlcNAc and (d) GlcA in the 10-mer; geometries from the four sets of each ensemble are

represented by red, green, blue, and magenta dots, respectively and the force-field geometry is represented by a single large black dot; Cremer–Pople parameters (ϕ, θ) for all rings in every tenth snapshot from each ensemble were plotted (i.e., 10 rings * 1,000 snapshots per run * 4 runs = 40,000 parameter sets for the 20-mer and 20,000 parameter sets for the 10-mer).

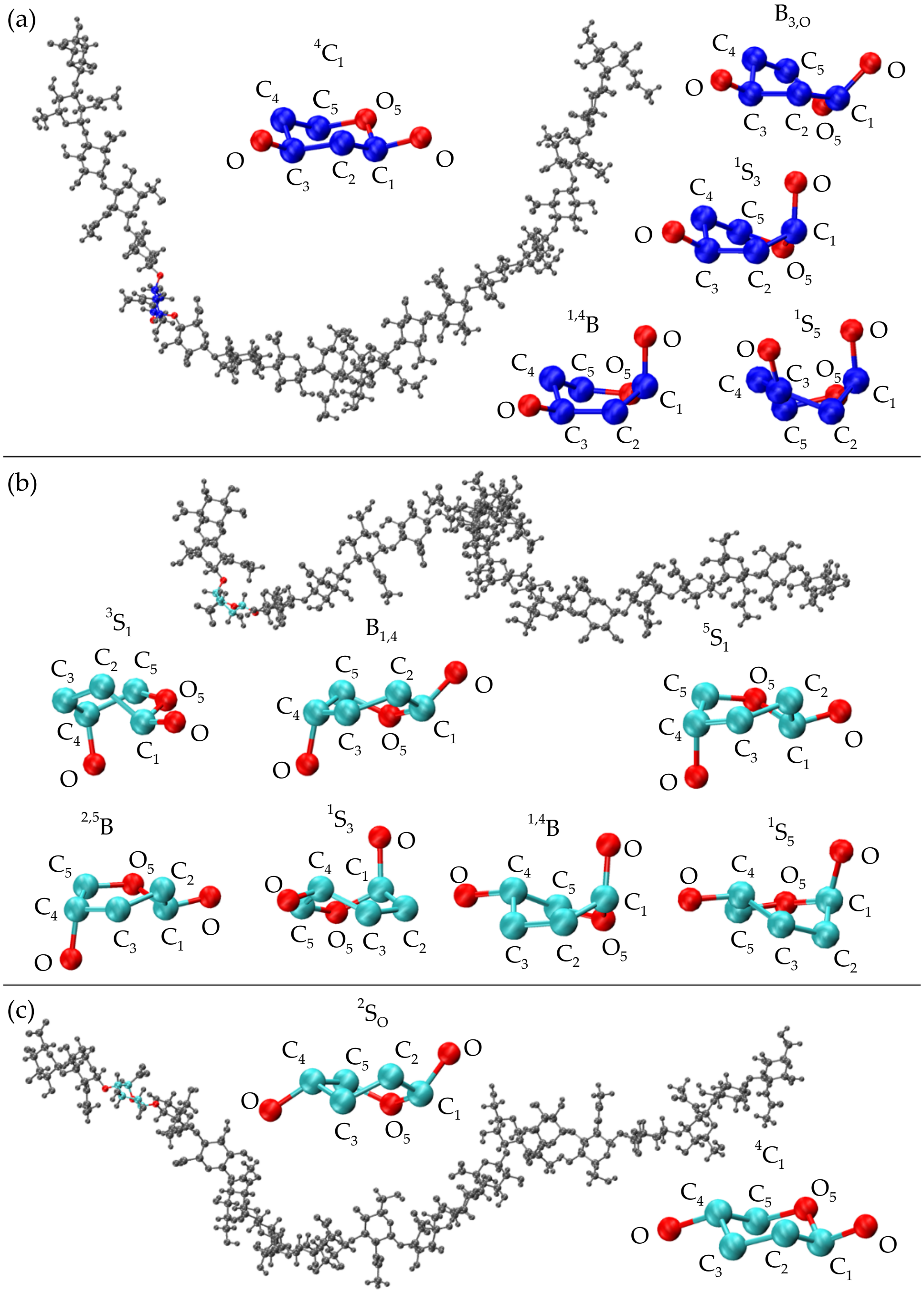


Figure 4.4. MD-generated hyaluronan conformations: (a) 20-mer with a $^{1,4}B$ GlcNAc conformer (ring atoms in blue; linkage atoms in red) and $^4C_1$, $B_{3,O}$, $^1S_3$, $^{1,4}B$, and $^1S_5$ GlcNAc monosaccharide conformations (endocyclic ring atoms and linker oxygen atoms only); (b) 20-mer with a $^5S_1$ GlcA conformer (ring atoms in cyan; linkage atoms in red) and $^3S_1$, $B_{1,4}$, $^5S_1$, $^{2,5}B$, $^1S_3$, $^{1,4}B$, and $^1S_5$ GlcA monosaccharide conformations (endocyclic ring atoms and linker oxygen atoms only); (c) 20-mer with a ^2S_O GlcA conformer (ring atoms

in cyan; linkage atoms in red) and $^4C_1$ and ^2S_O GlcA monosaccharide conformations (endocyclic ring atoms and linker oxygen atoms only); in each 20-mer snapshot, all gray rings are in $^4C_1$ conformations.

C-P analysis of GlcA monosaccharide rings in hyaluronan 20-mer showed that GlcA is found mostly in $^4C_1$ chair conformations (Figure 4.3.b), as seen in crystal structures [329-332] and NMR and force field studies [220,257,307,308,338,339,345], with some boat and skew-boat conformations, including $^3S_1$, $B_{1,4}$, $^5S_1$, $^{2,5}B$, ^2S_O, $^1S_3$, $^{1,4}B$, and $^1S_5$ (Figures 4.3.b and 4.4.b,c), which were also observed in GlcA rings in unbiased MD simulations of nonsulfated chondroitin 20-mers [188]. Some of these conformations were also sampled rarely in unbiased MD simulations of GlcA monosaccharides [220,307]. The slight differences between boat/skew-boat conformations in our simulations and monosaccharide simulations may have arisen from intramolecular interactions in the hyaluronan 20-mer. As with GlcNAc rings, these conformations did not occur consistently in any one region of the 20-mer, did not occur simultaneously in multiple rings in the same snapshot, and returned to $^4C_1$ chair conformations quickly, i.e., within 40 ns (Figure B.5). Except for ^2S_O, each of these GlcA and GlcNAc boat and skew-boat conformations brought the linker oxygen atoms closer together, causing a kink in the polymer chain (Figure 4.4), which may have contributed to more compact polymer conformations. To investigate this, end-to-end distances and radii of gyration of 20-mer conformations with non-$^4C_1$ ring puckers were analyzed (Figure B.6) and both compact and extended conformations were observed. This can be explained by the flexibility in the glycosidic linkages. Of note, MD-generated atomic-resolution snapshots before and after a monosaccharide ring underwent a conformational change between $^4C_1$ and non-$^4C_1$ were visualized in VMD to check for ionic interactions (e.g., between Na^+ and carboxylate O^- atoms and/or partially-charged hydroxyl O atoms in GlcA, between Cl^- and hydroxyl H atoms in GlcA, between Na^+ and partially-charged N-acetyl N or O atoms and/or hydroxyl O atoms in GlcNAc, or between Cl^- and H atoms of methyl and/or hydroxyl groups in GlcNAc). Specifically, we looked for ions within 5 Å of the exocyclic atoms of the monosaccharide of interest and found that ionic interactions were most often absent during, and thus, did not significantly contribute to, ring puckering in the MD simulation of hyaluronan 20-mer.

To analyze glycosidic linkage conformations, free energies of glycosidic linkage dihedrals Δ$G(\phi, \psi)$ were calculated (Figure 4.5.a,b and Table 4.2). If we assume O_5-C_1-O-C_n = H_1-C_1-O-C_n - 120° for ϕ values and C_1-O-C_3-C_2 = C_1-O-C_3-H_3 - 120° and C1-O-C_4-C_3 = C_1-O-C_4-H_4 + 120° for ψ values, our data match well with X-ray diffraction and force field data from hyaluronan tetrasaccharides [216] as well as NMR and molecular mechanics data from hyaluronan tetrasaccharides [346], hexasaccharides [256], and octasaccharides [347]. If we assume C_1-O-C_3-C_2 = C_1-O-C_3-C_4 + 120° and C_1-O-C_4-C_3 = C_1-O-C_4-C_5 - 120°, our data are also in close agreement with X-ray diffraction and quantum molecular modeling data [317,348-350] as well as NMR and MD data [339]. This agreement is with conformations sampled about the global minimum in each linkage in our MD simulations. All of these conformations were helical, so it stands to reason that 2-, 3-, and 8-fold helices were observed in hyaluronan oligosaccharides. This does not necessarily mean that adjacent glycosidic linkage conformations were interdependent. In fact, variation in the types of helices observed in the force field and experimental studies in only short oligosaccharide units suggests a lack of consistent and uniform helical structures in long hyaluronan polymers.

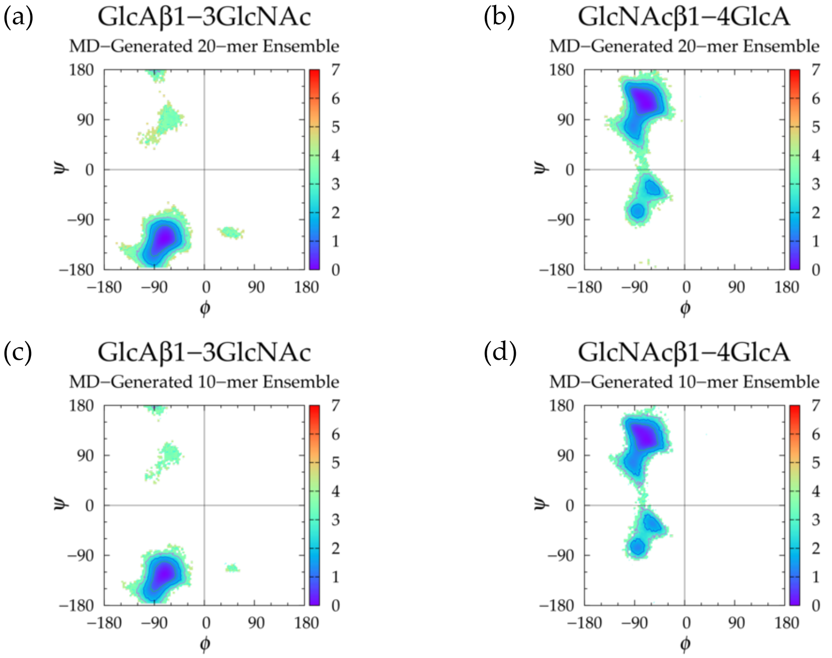


Figure 4.5. $\Delta G(\phi, \psi)$ in the MD-generated hyaluronan ensembles for aggregated GlcAβ1-3GlcNAc and GlcNAcβ1-4GlcA glycosidic linkage data in the (a,b) 20-mer and (c,d) 10-mer, respectively; contour lines every 1 kcal/mol.

Table 4.2. Hyaluronan Glycosidic Linkage Dihedrals (ϕ,ψ) [1] and

Free Energy ($\Delta G(\phi,\psi)$ kcal/mol) Minima.

	MD-Generated 20-mer Ensemble				MD-Generated 10-mer Ensemble				Constructed 20-mer Ensemble			
	GlcAβ1-3 GlcNAc		GlcNAcβ1-4 GlcA		GlcAβ1-3 GlcNAc		GlcNAcβ1-4 GlcA		GlcAβ1-3 GlcNAc		GlcNAcβ1-4 GlcA	
Min	ϕ, ψ	ΔG	ϕ, ψ	ΔG	ϕ, ψ	ΔG	ϕ, ψ	ΔG	ϕ, ψ	ΔG	ϕ, ψ	ΔG
I	−71.25°, −123.75°	0.00	−68.75°, 116.25°	0.00	−71.25°, −123.75°	0.00	−68.75°, 118.75°	0.00	−71.25°, −123.75°	0.00	−71.25°, 116.25°	0.00
II	−53.75°, 91.25°	3.06	−83.75°, −73.75°	1.31	−58.75°, 86.25°	3.08	−81.25°, −76.25°	1.16	−56.25°, 91.25°	3.01	−83.75°, −73.75°	1.29
II′			−58.75°, −33.75°	1.42			−53.75°, −36.25°	1.27			−56.25°, −36.25°	1.40

[1] ϕ, ψ dihedral angles were sorted into 2.5° bins.

Our data also show that a secondary basin was sampled in each linkage type in our MD simulations. In GlcAβ1-3GlcNAc linkages, conformations in the upper left quadrant ($-\phi$, $+\psi$) were sampled. A molecular modeling study that reportedly sampled the full allowable range of glycosidic linkage conformations in agreement with experimental data for several disaccharide units also revealed hyaluronan GlcAβ1-3GlcNAc linkage conformations in a secondary basin similar to ours [351]. Furthermore, crystal structures of hyaluronan oligosaccharides in complexes with proteins [329-332] were examined and we found that most glycosidic linkage conformations were similar to the most energetically-favorable conformations observed in MD simulation, with the exception of one GlcAβ1-3GlcNAc linkage conformation with $-\phi$, $+\psi$ dihedrals, which was seen in a hyaluronan hexasaccharide in complex with hyaluronidase [331]. This suggests that, although rare, this is a physical conformation. Analysis of individual glycosidic linkages in individual runs (Figure B.7) showed that this basin was sampled mostly in two particular GlcAβ1-3GlcNAc linkages in the MD run with the lowest end-to-end distance distribution. Therefore, 20-mer MD snapshots with these glycosidic linkage conformations were analyzed; we found that these linkage conformations were associated with lower end-to-end distances (Figure B.8 and Table B.1). Visual analysis of these snapshots showed that GlcAβ1-3GlcNAc glycosidic linkages with $-\phi$, $+\psi$ dihedrals caused a kink in the polymer, which explains the more compact polymer conformations in comparison to those containing only linkage conformations at the global free energy minimum (Figure B.8). To find out if there was a connection between these glycosidic linkage conformations and non-$^4C_1$ puckers of adjacent monosaccharide rings, glycosidic linkages flanking non-$^4C_1$ ring puckers were plotted, and they were all centered about the global minimum, i.e., $-\phi$, $-\psi$ (Figure B.9). In fact, there did not appear to be any correlation between GlcAβ1-3GlcNAc linkage $-\phi$, $+\psi$ conformations and a non-$^4C_1$ ring pucker in any region of the polymer.

Analysis of GlcNAcβ1-4GlcA glycosidic linkage conformations (Figure 4.5.b and Table 4.2) revealed secondary minima in the lower left quadrant ($-\phi$, $-\psi$), which was also seen in a molecular modeling study of GAG disaccharides with results validated by experimental data [351]. In hyaluronan 20-mer MD simulations, these conformations were sampled more in the GlcNAcβ1-4GlcA linkages neighboring the GlcAβ1-3GlcNAc linkages with $-\phi$, $+\psi$ dihedrals in the MD run with the lowest end-to-end distances. The same analyses to those performed on 20-mer conformations with GlcAβ1-3GlcNAc linkages with $-\phi$, $+\psi$ dihedrals were performed on MD-generated 20-mer snapshots with GlcNAcβ1-4GlcA linkages with $-\phi$, $-\psi$ dihedrals. Similarly, these glycosidic linkage conformations caused a kink in the polymer chain and were associated with lower end-to-end distances (Figure B.10 and Table B.1). To determine if there was a correlation between GlcAβ1-3GlcNAc $-\phi$, $+\psi$ linkage dihedrals and GlcNAcβ1-4GlcA $-\phi$, $-\psi$ linkage dihedrals, snapshots with both of these linkage conformations flanking the same monosaccharide ring were examined. It was found that there were very few snapshots in which this was the case. GlcNAcβ1-4GlcA linkage conformations flanking non-$^4C_1$ monosaccharide rings were also analyzed, and these conformations were centered about the global minimum, i.e., $-\phi$, $+\psi$ (Figure B.9). Additionally, ionic interactions (e.g., between Na^+, O^- in GlcA carboxylate group, and O in C=O of adjacent GlcNAc N-acetyl group, between Cl^-, H in GlcNAc methyl group, and H in GlcA hydroxyl groups, or between Na^+ and O atoms or Cl^- and H atoms of hydroxyls in adjacent monosaccharides) did not appear to play a role in the transition of glycosidic linkage conformations between their primary and secondary ϕ, ψ basins.

Hyaluronan 10-mers were also simulated and showed the same conformations as the 20-mer in MD simulation except for GlcNAc monosaccharide rings, which sampled only $^4C_1$ chair conformations in the 10-mer (Figures 4.3.c,d and 4.5.c,d and Table 4.2). This was expected, as these conformations occurred very few times in 20-mer MD simulations, and there were half as many samples (i.e., monosaccharide rings) in the 10-mer.

Together, these results suggest that (1) realistic helical glycosidic linkage conformations in hyaluronan 10- and 20-mers are seen in MD simulation, (2) non-$^4C_1$ ring puckers, $-\phi$, $+\psi$ dihedrals in

GlcAβ1-3GlcNAc linkages, and $-\phi$, $-\psi$ dihedrals in GlcNAcβ1-4GlcA linkages all contribute to more compact conformations, and (3) monosaccharide rings and glycosidic linkages in hyaluronan 10- and 20-mers behave randomly and independently in MD simulation.

4.3.1.2. Construction Algorithm. Our construction algorithm was used to construct four sets of hyaluronan 20-mer ensembles with 10,000 conformations each (40,000 conformations total). A comparison of end-to-end distance distributions (Figure 4.6.a and Table 4.1) and radii of gyration (Figure B.2.a,b) showed that our construction algorithm produces ensembles that mimic backbone flexibility observed in MD simulations. Furthermore, the C-P parameters (Figure B.11) and glycosidic linkage dihedral free energies $\Delta G(\phi, \psi)$ (Figure B.12) of constructed conformations post minimization matched those of MD-generated ensembles (i.e., construction algorithm input data; Figures 4.3.a,b and 4.5.a,b and Table 4.2), indicating that dihedral angles were properly restrained and that minimization did not change the overall ring and linkage conformation.

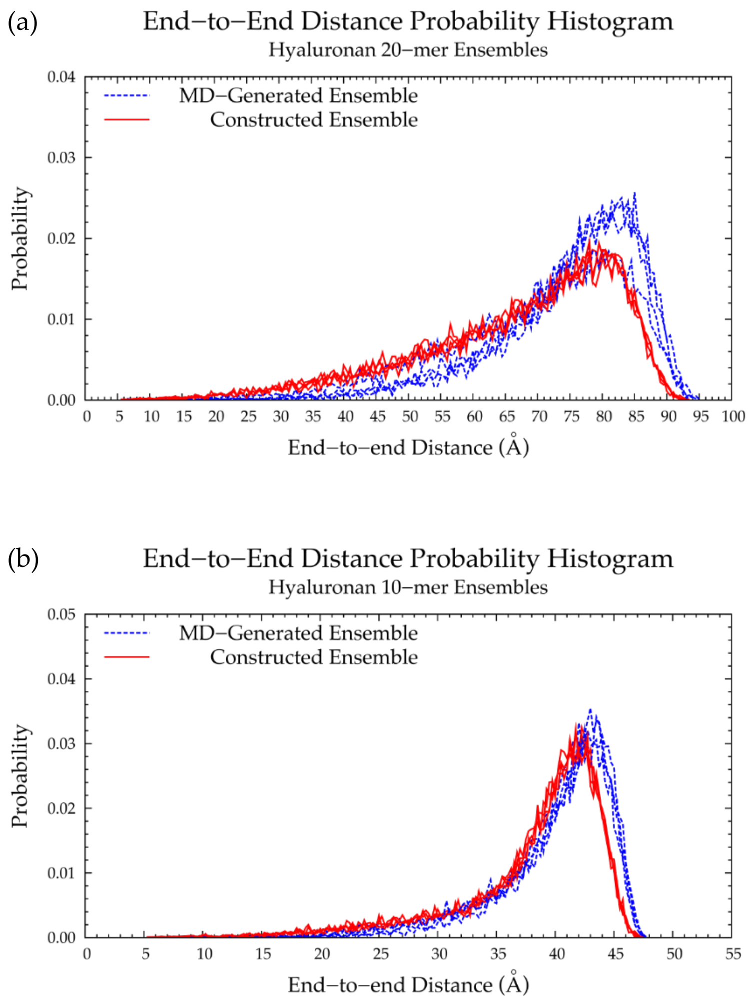


Figure 4.6. End-to-end distance probability distributions of MD-generated (blue dashed lines) and constructed (red solid lines) hyaluronan ensembles: (a) 20-mer and (b) 10-mer; probabilities were calculated for end-to-end distances sorted into 0.5 Å bins for the 20-mer and 0.25 Å bins for the 10-mer; each ensemble includes four sets of 10,000 conformations.

A hyaluronan 10-mer ensemble with 40,000 conformations was also constructed using the algorithm; similarly, the end-to-end distances and radii of gyration of the constructed 10-mer ensemble matched those of the MD-generated 10-mer ensemble (Figures 4.6.b and B.2.c,d and Table 4.1). Both ensembles contained mostly extended conformations, as expected for short oligosaccharides with fewer opportunities for kinks or curves.

Finally, a hyaluronan 200-mer ensemble containing four sets of 10,000 conformations was constructed using the algorithm. The construction procedure produced end-to-end distances, radii of gyration, and bond potential energies of all conformations, as well as PDB files of conformations with most probable end-to-end distances, bond potential energies above that of the fully-extended conformation, and all excluded conformations (for a total of about 1,000 PDBs for each of the four sets) for visualization and validation of the bond potential energy cutoff. As seen in MD-generated hyaluronan 10- and 20-mer ensembles, we expected the end-to-end distance distribution curve skewness to shift toward the right (i.e., lower end-to-end distances indicating more compact conformations) with increasing polymer length, as there were more opportunities for kinks and curves resulting from ring puckering and the flexibility of the glycosidic linkages. This pattern was seen in the end-to-end distance distributions of the hyaluronan 200-mer constructed ensembles (Figure 4.7) and in those of nonsulfated chondroitin 100- and 200-mer constructed ensembles [188]. The relationship between end-to-end distance and radius of gyration was decreasingly linear with increasing polymer length. In other words, conformations with the same end-to-end distance have a wider range of radii of gyration in longer polymers. This was evidenced by comparison of R^2 values of the end-to-end distance vs. radius of gyration regression lines of hyaluronan 20- and 10-mer ensembles, both MD-generated and constructed (Figure B.2.a–d). As expected, the end-to-end distance and radius of gyration relationship was decreasingly linear with increasing polymer length, as shown by comparison of hyaluronan 10-, 20-, and 200-mer constructed ensembles (Figure B.2.b,d,e). This suggests that our algorithm constructs hyaluronan 200-mer ensembles with the backbone conformations that we would expect to see in MD simulations.

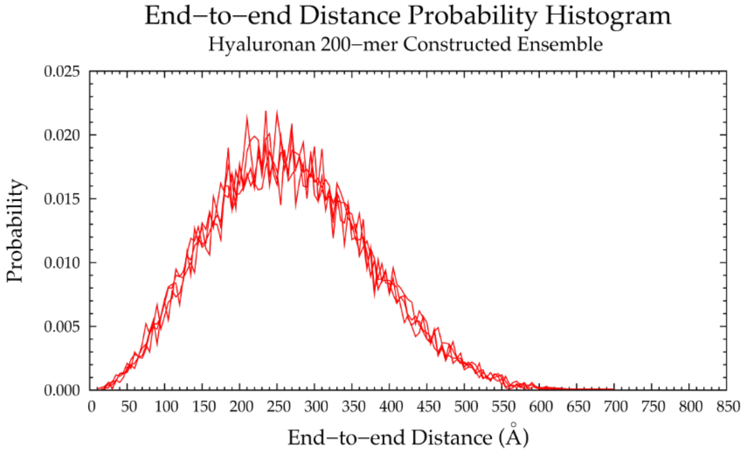


Figure 4.7. End-to-end distance probability distribution of constructed ensemble of hyaluronan 200-mer; most probable end-to-end distance across all four sets is 235 Å; probabilities were calculated for end-to-end distances sorted into 5 Å bins; the ensemble contains four sets of 10,000 conformations.

4.3.2. Nonsulfated Dermatan

4.3.2.1. Molecular Dynamics Simulations: Glycosidic Linkage and Monosaccharide Ring Geometry Effects on Polymer Backbone Flexibility. MD simulations of nonsulfated dermatan 20-mer revealed relatively rigid and linear backbone conformations, as evidenced by the narrow and highly left-skewed end-to-end distance distributions, which match in all four MD runs (Figure 4.8 and Table 4.3), and the linear relationship between end-to-end distance and radius of gyration (Figure B.13.a). To understand the factors contributing to this rigid behavior of the dermatan 20-mer in simulation, conformations of monosaccharide rings and glycosidic linkages were analyzed.

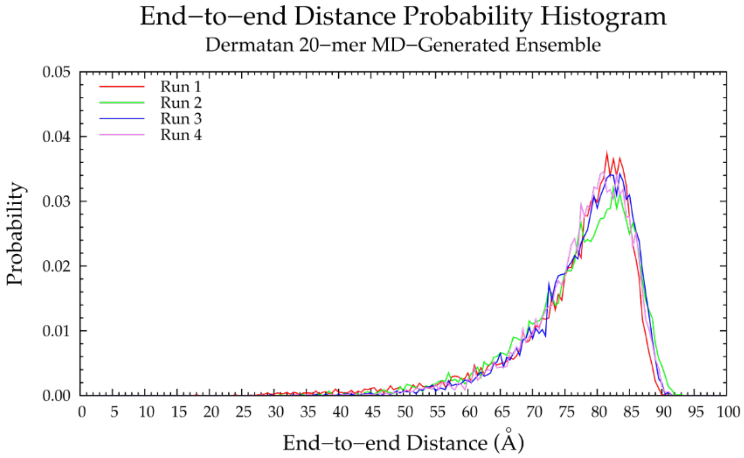


Figure 4.8. End-to-end distance probability distribution of MD-generated nonsulfated dermatan 20-mer ensemble; each of the four runs includes 10,000 conformations; probabilities were calculated for end-to-end distances sorted into 0.5 Å bins.

Table 4.3. Most Probable End-to-End Distances (*d*) in MD-Generated and Constructed Nonsulfated Dermatan Ensembles [1].

	20-mer Ensembles			10-mer Ensembles		
	MD-Generated *d* (Å)	Constructed *d* (Å)	% Difference	MD-Generated *d* (Å)	Constructed *d* (Å)	% Difference
Run 1	81.5	81.5		42.25	42.75	
Run 2	82.5	83.5		42.50	42.00	
Run 3	83.5	84.0		42.00	42.75	
Run 4	81.0	81.5		42.25	42.75	
All [2]	83.5	81.5	2.42%	42.25	42.75	1.18%

[1] Probabilities were calculated for end-to-end distances sorted into 0.5 Å bins for the 20-mer ensembles and 0.25 Å bins for the 10-mer ensembles. [2] All = end-to-end distance distribution aggregated across all four runs.

GalNAc monosaccharide ring C-P parameters (Figure 4.9.a) show mostly $^4C_1$ conformations, as well-established by X-ray diffraction [216,343,352], NMR [217,353], and force field [188,216,217,307] data. Biased MD simulations of β-GalNAc monosaccharides showed that nonsulfated β-GalNAc sampled boat and skew-boat conformations, namely $B_{3,O}$, $^{1}S_{3}$, and $^{1,4}B$, with relatively high free energies [307]. In line with this study, our simulations showed very few boat and skew-boat puckers, namely $^{1}S_{3}$, $^{1,4}B$, $^{1}S_{5}$ (Figure 4.4.a; -3GalNAcβ1- endocyclic ring and linker oxygen atoms were identical to those of -3GlcNAcβ1-), and $B_{2,5}$, on the microsecond timescale. These puckers were sampled by different rings of the 20-mer in different MD runs and returned to $^{4}C_{1}$ within 10 ns (Figure B.14), suggesting that this behavior was random, and confirming that non-$^{4}C_{1}$ GalNAc puckers are not energetically favorable. While these boat and skew-boat conformations caused a kink in the polymer chain, there were so few occurrences of this that they did not impact the overall end-to-end distance distribution.

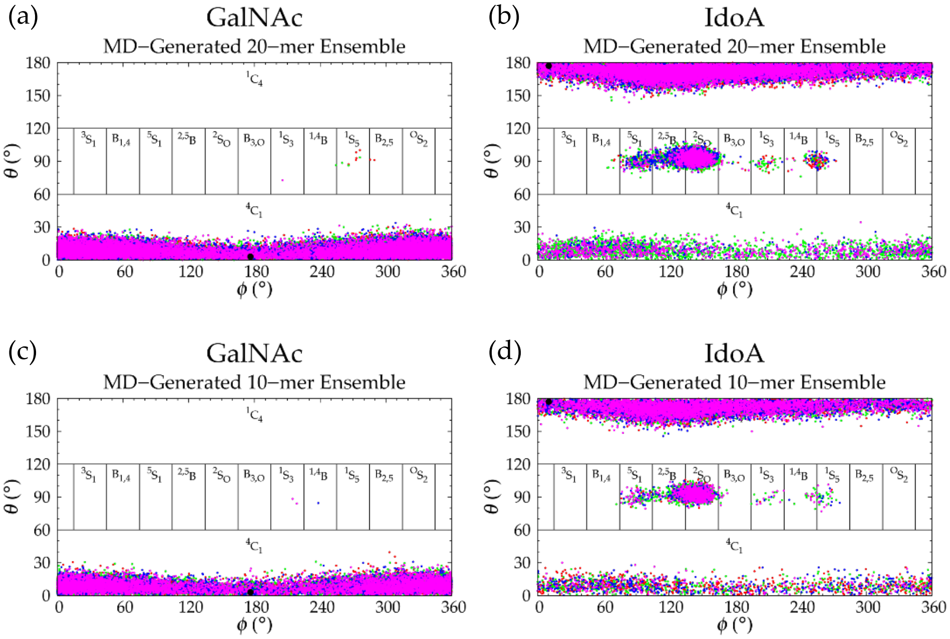


Figure 4.9. Cremer–Pople data for MD-generated nonsulfated dermatan ensembles: (a) GalNAc and (b) IdoA in the 20-mer and (c) GalNAc and (d) IdoA in the 10-mer; geometries from the four sets of each ensemble are represented by red, green, blue, and magenta dots, respectively and the force-field geometry is represented by a single large black dot; Cremer–Pople parameters (ϕ, θ) for all rings in every tenth snapshot from each ensemble were plotted (i.e., 10 rings * 1,000 snapshots per run * 4 runs = 40,000 parameter sets for the 20-mer and 20,000 parameter sets for the 10-mer).

IdoA monosaccharide rings in 20-mer MD sampled a majority of $^1C_4$ and, to a lesser degree, ^2S_O and $^4C_1$ puckers (Figure 4.9.b and Table B.2), as observed in NMR and force field studies [190,217,220,262,275,306,307,310,354-356]. Literature reports of the relative proportions of these three ring puckers in nonsulfated IdoA vary, which may be explained by differences in the structure of neighboring residues (or lack thereof) [275,310,354,355] and differences in ion concentrations [310]. Furthermore, existing studies were conducted on IdoA monosaccharides and short IdoA-containing GAG oligosaccharides (i.e., <20 monosaccharides). NMR data have shown that, when at the nonreducing terminal in GAG oligosaccharides, nonsulfated IdoA is in equilibrium with $^1C_4$, ^2S_O, and $^4C_1$ puckers, and when flanked by nonsulfated N-acetylated sugar derivatives (i.e., GalNAc or GlcNAc), nonsulfated IdoA is primarily in $^1C_4$ and ^2S_O conformations [310] with fewer than 10% of samples in $^4C_1$ [275,355]. Indeed, we found this to be the case in our dermatan 20-mer MD simulations (Table B.2). Importantly, none of the $^1C_4$, ^2S_O, or $^4C_1$ puckers introduced a kink in the polymer chain, so we would not expect variations in the relative proportions of these ring puckers to alter overall polymer backbone conformations. IdoA also sampled other boat and skew-boat conformations in 20-mer MD simulations, specifically $B_{1,4}$, $^5S_1$, $^{2,5}B$, $B_{3,O}$, $^1S_3$, $^{1,4}B$, and $^1S_5$ (Figure 4.9.b and Table B.2), which is in agreement with results from MD simulations of IdoA monosaccharides [263,307]. Another force field study showed that $^{2,5}B$ and $B_{3,O}$ IdoA conformers are also feasible interpretations of existing NMR data [357]. Furthermore, a force field study that mapped X-ray diffraction and NMR data for dermatan sulfate [262] and a molecular modeling study of IdoA monosaccharides [263] showed that interconversion between different boat/skew-boat forms of IdoA is much more common than that between boat/skew-boat and chair forms, as it is less energetically costly, which helps explain the sampling of multiple different boat/skew-boat IdoA conformations. As each of these boat/skew-boat puckers (except for ^2S_O) introduces a kink in the polymer chain (Figures 4.10 and 4.4.b,c; -4IdoAα1- endocyclic ring and linker oxygen atoms are identical to those of -4GlcAcβ1-), end-to-end distances of 20-mer conformations with these puckers were analyzed to determine if they were associated with more compact conformations. The end-to-end distance distribution of conformations with boat/skew-boat ring puckers (other than $^{2}S_{O}$) matched that of the average of the four runs from the full MD-generated 20-mer ensemble (Figure B.15), indicating that these ring puckers do not necessarily give compact polymer conformations. Additionally, the end-to-end distance distribution of 20-mer conformations with $^{2}S_{O}$ IdoA conformations was compared to that of the full MD ensemble (Figure B.15). These distributions were similar, further suggesting that $^{2}S_{O}$ conformations are not associated with more or less compact 20-mer conformations. To determine if IdoA ring puckering in nonsulfated dermatan 20-mer MD was random, an analysis was performed of C-P parameters of individual IdoA monosaccharides in each of the four MD runs (Figure B.16). This revealed that (1) not all IdoA rings overcame the energy barrier to sample a $^{4}C_{1}$ pucker, (2) no single IdoA ring sampled $^{4}C_{1}$ in all four MD runs, (3) each IdoA ring overcame the energy barrier to sample boat/skew-boat puckers (i.e., C-P $\theta \sim 90°$) in at least one MD run, and (4) after sampling boat/skew-boat puckers, some IdoA rings returned to $^{1}C_{4}$ in as little as 165 ns, while others remained in boat/skew-boat conformations for the remainder of the 1-μs trajectory. These observations support the idea that monosaccharide rings in a nonsulfated dermatan 20-mer behave randomly and independently in unbiased MD simulation.

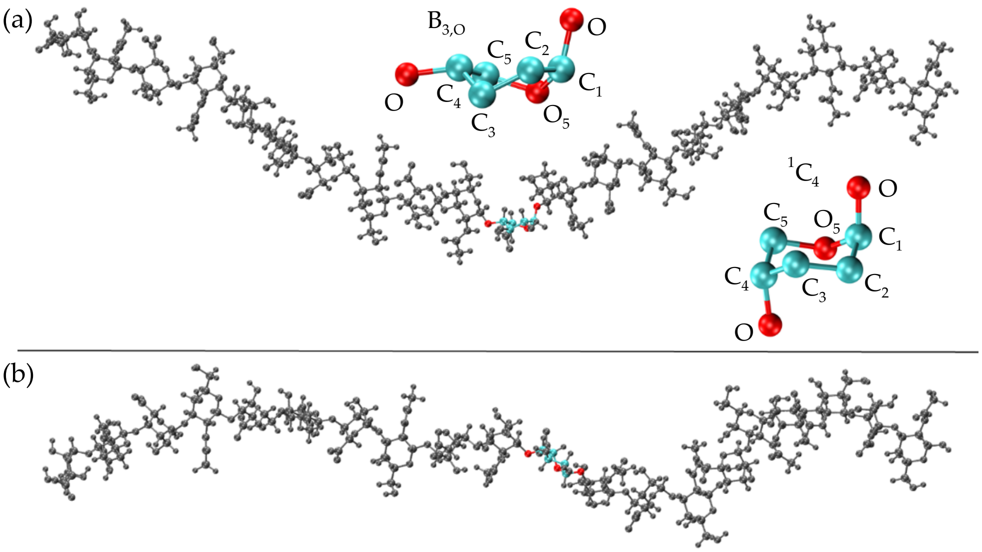


Figure 4.10. MD-generated nonsulfated dermatan conformations: (a) 20-mer with a $B_{3,O}$ IdoA conformer (ring atoms in cyan; linkage atoms in red) and $B_{3,O}$ IdoA and $^1C_4$ IdoA monosaccharide conformations (endocyclic ring atoms and linker oxygen atoms only); (b) 20-mer with a ^2S_O IdoA conformer highlighted (ring atoms in cyan; linkage atoms in red; ^2S_O α-IdoA is identical to ^2S_O β-GlcA in Figure 4.4.c); in both 20-mer conformations, all IdoA monosaccharides in gray are in $^1C_4$ and all GalNAc monosaccharides are in $^4C_1$.

Analysis of dihedral free energies revealed a single $\Delta G(\phi, \psi)$ minimum for each of IdoAα1-3GalNAc and GalNAcβ1-4IdoA glycosidic linkages (Figure 4.11.a,b and Table 4.4). This is consistent across all linkages in all runs for each linkage type. The lack of secondary basins in ϕ, ψ dihedral samples may explain the higher degree of rigidity and tendency toward extended backbone conformations of nonsulfated dermatan 20-mer in comparison to hyaluronan and nonsulfated chondroitin [188] 20-mers in MD simulation. An X-ray diffraction study of sodium dermatan sulfate observed three different helical forms (i.e., right-handed, left-handed, and achiral) [352] and quantified their dihedral angles, which were defined by: $\phi = O_5\text{-}C_1\text{-}O\text{-}C_n$ and $\psi = C_1\text{-}O\text{-}C_n\text{-}C_{(n+1)}$. Assuming $C_1\text{-}O\text{-}C_3\text{-}C_2 = C_1\text{-}O\text{-}C_3\text{-}C_4 + 120°$ and $C1\text{-}O\text{-}C_4\text{-}C_3 = C_1\text{-}O\text{-}C_4\text{-}C_5 - 120°$, we see that the right-handed and achiral helical forms corresponded to high $\Delta G(\phi, \psi)$ values in our nonsulfated dermatan 20-mer MD simulations, and the left-handed helical form corresponded to glycosidic linkage conformations very close to the $\Delta G(\phi, \psi)$ global minimum. In line with our results, force field studies that compared MD results to existing NMR and X-ray diffraction data of dermatan sulfate dismissed right-handed and achiral helical forms and confirmed the presence of left-handed helical structure [216,262]. A quantitative comparison of our MD-generated glycosidic linkage data to their findings and other NMR data for dermatan tetrasaccharides [217] showed close agreement if we assume O_5-C_1-O-C_n = H_1-C_1-O-C_n - 120° for all ϕ values, C_1-O-C_3-C_2 = C_1-O-C_3-H_3 - 120° for α1-3 linkage ψ values, and C1-O-C_4-C_3 = C_1-O-C_4-H_4 + 120° for β1-4 linkage ψ values.

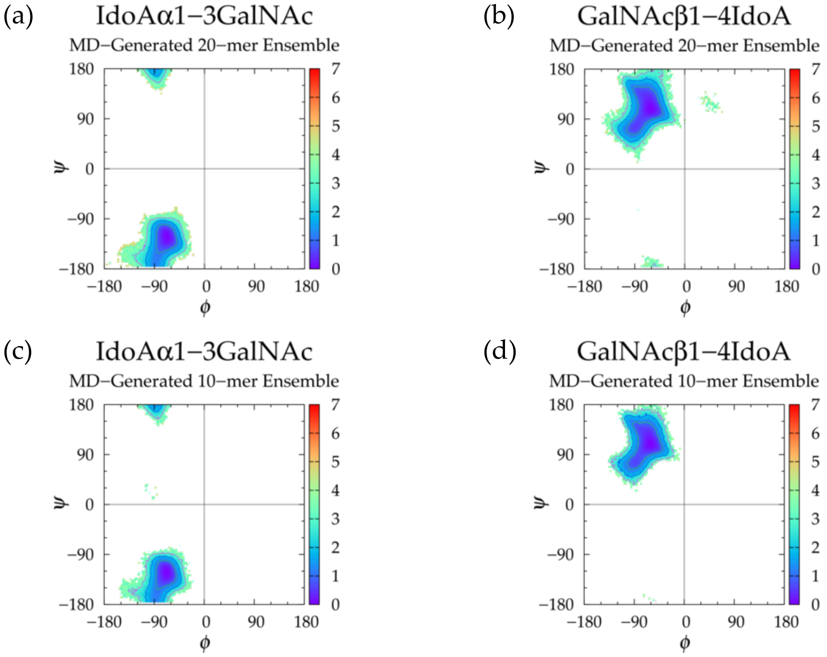


Figure 4.11. $\Delta G(\phi, \psi)$ in the MD-generated nonsulfated dermatan ensembles for aggregated IdoAα1-3GalNAc and GalNAcβ1-4IdoA glycosidic linkage data in the (a,b) 20-mer and (c,d) 10-mer, respectively; contour lines every 1 kcal/mol.

Table 4.4. Nonsulfated Dermatan Glycosidic Linkage Dihedrals (ϕ, ψ) [1] and

Free Energy ($\Delta G(\phi, \psi)$ kcal/mol) Minima.

	MD-Generated 20-mer Ensemble				MD-Generated 10-mer Ensemble				Constructed 20-mer Ensemble			
	IdoAβ1-3 GalNAc		GalNAcβ1-4 IdoA		IdoAβ1-3 GalNAc		GalNAcβ1-4 IdoA		IdoAβ1-3 GalNAc		GalNAcβ1-4 IdoA	
Min	ϕ, ψ	ΔG	ϕ, ψ	ΔG	ϕ, ψ	ΔG	ϕ, ψ	ΔG	ϕ, ψ	ΔG	ϕ, ψ	ΔG
I	−66.25°, −123.75°	0.00	−61.25°, 106.25°	0.00	−68.75°, −121.25°	0.00	−58.75°, 108.75°	0.00	−68.75°, −121.25°	0.00	−61.25°, 108.75°	0.00

[1] ϕ, ψ dihedral angles were sorted into 2.5° bins.

MD simulations of a nonsulfated dermatan 10-mer produced mostly linear backbone conformations (Figure B.13.c) and similar monosaccharide ring (Figure 4.9) and glycosidic linkage (Figure 4.11 and Table 4.4) geometries to nonsulfated dermatan 20-mer. The major difference was that in 10-mer MD simulations, 9% of all IdoA rings in all runs were found in $^4C_1$ form; this was almost exclusively in the nonreducing terminal IdoA (i.e., ring #10; Figure B.17 and Table B.2), which is mostly in line with the aforementioned NMR study of IdoA ring conformations [275]. However, based on the findings from the NMR study and from our dermatan 20-mer MD simulations, we would still expect the nonterminal IdoA rings in a GAG polymer to sample some $^4C_1$ conformations. This supports our position that GAG 20-mers are better-suited for predictions of long-chain backbone conformations than short GAG oligosaccharides.

These findings suggest that (1) monosaccharide rings and glycosidic linkages in nonsulfated dermatan GAG 20-mers behave randomly and independently in MD simulation, (2) nonsulfated dermatan polymers take on rigid left-handed helical structure with a tendency toward linear backbone conformations in unbiased MD simulations, and (3) nonsulfated dermatan 20-mer conformations in MD simulation provide a realistic representation of longer nonsulfated dermatan polymer conformations.

4.3.2.2. Construction Algorithm. Four sets of 10,000 conformations of a nonsulfated dermatan 20-mer were constructed using our algorithm and compared to conformations seen in nonsulfated dermatan 20-

mer MD. As expected, the C-P parameter plots (Figure B.18) and glycosidic linkage free energies (Figure B.19) in the constructed 20-mer ensemble matched those in the MD-generated 20-mer ensemble (Figures 4.9.a,b and 4.11.a,b and Table 4.4). Furthermore, the end-to-end distances (Figure 4.12.a) and radii of gyration (Figure B.13.a,b) in the constructed ensemble matched those in the MD-generated ensemble, with only a 2.42% difference in most probable end-to-end distance (Table 4.3). This demonstrates that our algorithm produces nonsulfated dermatan 20-mer ensembles that mimic 20-mer backbone flexibility seen in MD simulations.

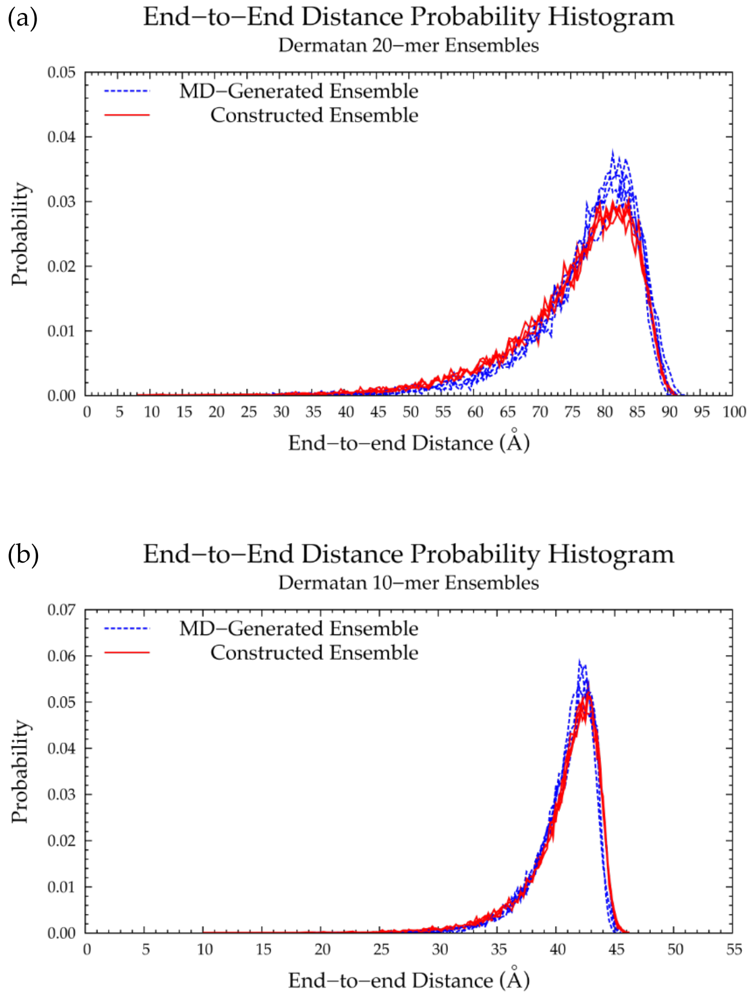


Figure 4.12. End-to-end distance probability distributions of MD-generated (blue dashed lines) and constructed (red solid lines) nonsulfated dermatan ensembles: (a) 20-mer and (b) 10-mer; probabilities

were calculated for end-to-end distances sorted into 0.5 Å bins for the 20-mer and 0.25 Å bins for the 10-mer; each ensemble includes four sets of 10,000 conformations.

The nonsulfated dermatan 10-mer ensemble with 40,000 conformations constructed by the algorithm had very similar end-to-end distance and radius of gyration data compared to the MD-generated ensemble (Figures 4.12.b and B.13.c,d and Table 4.3). The algorithm was also used to create a nonsulfated dermatan 200-mer ensemble. As expected, the 200-mer end-to-end distance distribution (Figure 4.13) and radii of gyration (Figure B.13.e) showed that conformations tended to be more compact as polymer length increased. The shift in skewness of the end-to-end distance distribution with increasing polymer length was more subtle in nonsulfated dermatan than in hyaluronan, as there were fewer linkage and ring conformations causing kinks in the polymer.

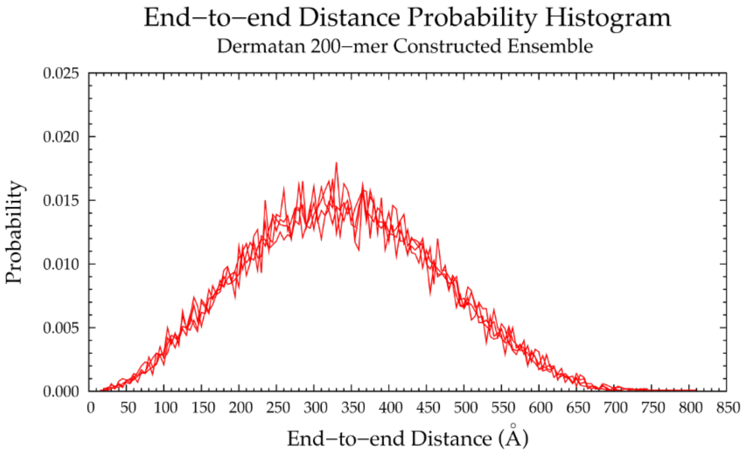


Figure 4.13. End-to-end distance probability distribution of constructed ensemble of nonsulfated dermatan 200-mer; most probable end-to-end distance across all four sets is 365 Å; probabilities were calculated for end-to-end distances sorted into 5 Å bins; the ensemble contains four sets of 10,000 conformations.

4.3.3. Nonsulfated Keratan

4.3.3.1. Molecular Dynamics Simulations: Glycosidic Linkage and Monosaccharide Ring Geometry Effects on Polymer Backbone Flexibility. In MD simulations, nonsulfated keratan 20-mers showed rigid backbone behavior and favored extended backbone conformations, as evidenced by end-to-end distance distributions (Figure 4.14 and Table 4.5) which were identical in all four MD runs, and the high correlation of end-to-end distance to radius of gyration (Figure B.20.a). To understand the causes of this rigid behavior, monosaccharide ring and glycosidic linkage conformations were analyzed.

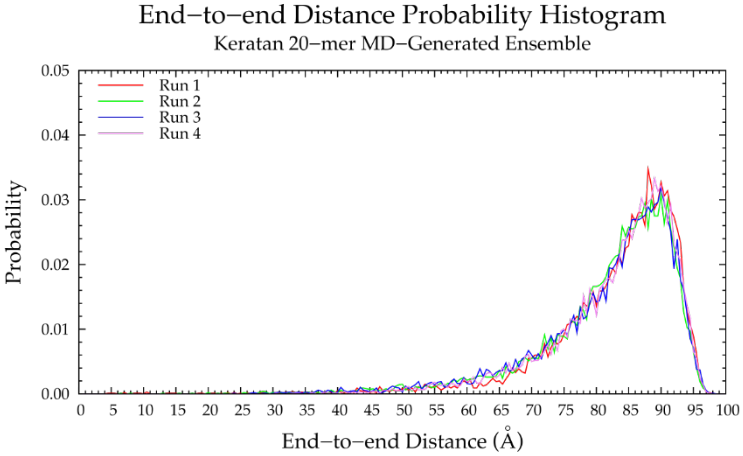


Figure 4.14. End-to-end distance probability distribution of MD-generated nonsulfated keratan 20-mer ensemble; each of the four runs includes 10,000 conformations; probabilities were calculated for end-to-end distances sorted into 0.5 Å bins.

Table 4.5. Most Probable End-to-End Distances (*d*) in MD-Generated and Constructed Nonsulfated Keratan Ensembles [1].

	20-mer Ensembles			10-mer Ensembles		
	MD-Generated *d* (Å)	Constructed *d* (Å)	% Difference	MD-Generated *d* (Å)	Constructed *d* (Å)	% Difference
Run 1	88.0	86.0		45.50	45.00	
Run 2	90.0	87.5		46.00	45.75	
Run 3	90.0	87.0		45.75	45.00	
Run 4	89.0	86.0		45.50	45.00	
All [2]	90.0	87.0	3.39%	45.50	45.00	1.10%

[1] Probabilities were calculated for end-to-end distances sorted into 0.5 Å bins for the 20-mer ensembles and 0.25 Å bins for the 10-mer ensembles. [2] All = end-to-end distance distribution aggregated across all four runs.

C-P parameters of GlcNAc monosaccharides in keratan 20-mer MD (Figure 4.15.a) showed similar conformations to GlcNAc in hyaluronan 20-mer. There were predominantly $^4C_1$ chair conformations, which are also found in a keratan sulfate tetrasaccharide crystal structure [358], with very few transitions (i.e., no more than one per ring or two per run) to boat/skew-boat conformations (i.e., ^2S_O, $^1S_3$, $^{1,4}B$, and $^1S_5$), which were sampled for no more than 20 ns before returning to $^4C_1$ chair (Figure B.21). Analysis of end-to-end distance and radius of gyration of 20-mer conformations containing non-$^4C_1$ ring puckers showed that these ring puckers were not associated with compact 20-mer conformations (Figure B.22). C-P analysis of Gal monosaccharides showed that for the duration of the keratan 20-mer MD simulations, Gal remained in $^4C_1$ chair (Figure 4.15.b), which was the predominant conformation found in NMR [353,359-361] and force field data [307] for Gal in general, and in a keratan sulfate tetrasaccharide crystal structure [358].

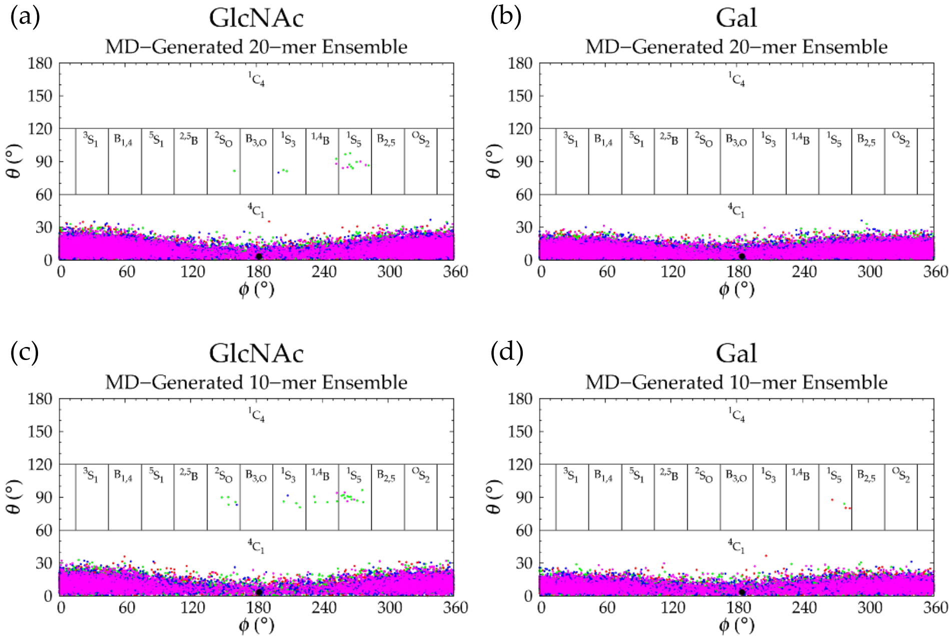


Figure 4.15. Cremer–Pople data for MD-generated nonsulfated keratan ensembles: (a) GlcNAc and (b) Gal in the 20-mer and (c) GlcNAc and (d) Gal in the 10-mer; geometries from the four sets of each ensemble are represented by red, green, blue, and magenta dots, respectively and the force-field geometry is represented by a single large black dot; Cremer–Pople parameters (ϕ, θ) for all rings in every tenth snapshot from each ensemble were plotted (i.e., 10 rings * 1,000 snapshots per run * 4 runs = 40,000 parameter sets for the 20-mer and 20,000 parameter sets for the 10-mer).

Glycosidic linkage free energy $\Delta G(\phi, \psi)$ analysis of Galβ1-4GlcNAc linkages revealed a global minimum in the basin with $-\phi$, $+\psi$ and a secondary minimum in the basin with $-\phi$, $-\psi$ (Figure 4.16.a and Table 4.6), similar to hyaluronan GlcNAcβ1-4GlcA linkages (Figure 4.5.b,d and Table 4.2), but with additional rare conformations in $+\phi$, $+\psi$. A molecular modeling study, which was shown to agree with experimental data, showed that β1-4 linkages in keratan sulfate disaccharides took on the same conformations as β1-4 linkages in hyaluronan disaccharides [351], confirming the primary and secondary basins sampled in our 20-mer MD simulations. Keratan 20-mer conformations with glycosidic linkage conformations in the tertiary basin (i.e., $+\phi$, $+\psi$) were visualized and end-to-end distances were analyzed

(Figure B.23). These conformations formed a slight bend in the polymer chain and were associated with more compact polymer backbone conformations. However, there did not appear to be any close contacts between atoms of adjacent monosaccharides in this conformation or in previous snapshots that could explain these linkage bond rotations. Furthermore, this conformation was not specific to any particular region of the polymer. Therefore, this behavior was likely random. As −*ϕ*, −*ψ* β1-4 linkage conformations are associated with compact hyaluronan polymer conformations, and nonsulfated keratan 20-mer is relatively rigid and favors extended conformations in contrast to hyaluronan, we sought to determine the effects of Galβ1-4GlcNAc linkage conformations in the secondary (i.e., −*ϕ*, −*ψ*) basin on keratan 20-mer backbone flexibility in MD simulations. As in hyaluronan, these conformations caused a kink in the keratan 20-mer chain and were associated with more compact conformations (Figure B.23 and Table B.1). However, these conformations were slightly less stable in keratan 20-mer MD (Table 4.6) than in hyaluronan MD (Table 4.2), and were, therefore, less common.

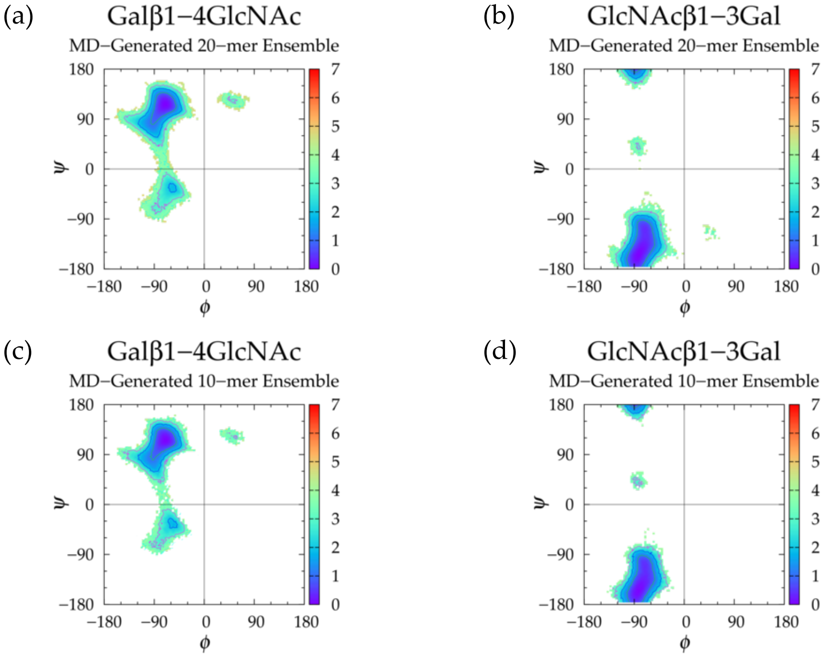


Figure 4.16. Δ*G*(*ϕ*, *ψ*) in the MD-generated nonsulfated keratan ensembles for aggregated Galβ1-4GlcNAc and GlcNAcβ1-3Gal glycosidic linkage data in the (a,b) 20-mer and (c,d) 10-mer, respectively; contour lines every 1 kcal/mol.

Table 4.6. Nonsulfated Keratan Glycosidic Linkage Dihedrals (ϕ, ψ) [1] and

Free Energy ($\Delta G(\phi, \psi)$ kcal/mol) Minima.

	MD-Generated 20-mer Ensemble				MD-Generated 10-mer Ensemble				Constructed 20-mer Ensemble			
	Galβ1-4 GlcNAc		GlcNAcβ1-3 Gal		Galβ1-4 GlcNAc		GlcNAcβ1-3 Gal		Galβ1-4 GlcNAc		GlcNAcβ1-3 Gal	
Min	ϕ, ψ	ΔG	ϕ, ψ	ΔG	ϕ, ψ	ΔG	ϕ, ψ	ΔG	ϕ, ψ	ΔG	ϕ, ψ	ΔG
I	−73.75°, 116.25°	0.00	−81.25°, −156.25°	0.00	−71.25°, 116.25°	0.00	−81.25°, −156.25°	0.00	−71.25°, 118.75°	0.00	−81.25°, −158.75°	0.00
II	−56.25°, −31.25°	1.75			−56.25°, −36.25°	1.67			−53.75°, −33.75°	1.75		

[1] ϕ, ψ dihedral angles were sorted into 2.5° bins.

Analysis of GlcNAcβ1-3Gal glycosidic linkages showed a global free energy minimum in the $-\phi$, $-\psi$ basin (Table 4.6) similar to that of IdoAβ1-3GalNAc linkages in nonsulfated dermatan MD (Figure 4.11.a,c and Table 4.4), but with additional rare occurrences in a secondary basin in $-\phi$, $+\psi$, confirmed by the aforementioned molecular modeling study [351], and a small tertiary basin in $+\phi$, $-\psi$ (Figure 4.16.b). To determine the effects of GlcNAcβ1-3Gal linkage conformations in these secondary and tertiary basins on backbone conformation, we visualized 20-mer conformations containing these glycosidic linkage conformations and analyzed end-to-end distances (Figure B.24). Glycosidic linkage conformations in the tertiary basin caused a kink and were associated with more compact 20-mer conformations. In contrast to hyaluronan, nonsulfated keratan β1-3 glycosidic linkage conformations in the secondary basin (i.e., $-\phi$, $+\psi$) caused only a slight bend in the polymer chain, but were still associated with compact 20-mer conformations (Table B.1). Both secondary and tertiary GlcNAcβ1-3Gal linkage conformations were rare in keratan 20-mer MD, which helps explain why the keratan 20-mer chain was relatively rigid and favored extended conformations in MD simulation.

Backbone conformational analysis of nonsulfated keratan 10-mer MD revealed that the 10-mer chain was rigid and favored extended conformations (Figure B.20.c and Table 4.5), which is similar behavior to that of nonsulfated keratan 20-mer in MD simulation. This stands to reason, as conformations of monosaccharide rings (Figure 4.15) and glycosidic linkages (Figure 4.16 and Table 4.6) in keratan 10-mer MD matched those in keratan 20-mer MD, with the only differences being that (1) Gal briefly (i.e., for < 5 ns) sampled a $^1S_5$ skew-boat conformation once in each of two of the 10-mer MD runs (Figure B.25), and (2) GlcNAcβ1-3Gal linkages in the 10-mer did not sample conformations in the tertiary basin (i.e., $+\phi$, $-\psi$) sampled in the 20-mer. Boat/skew-boat conformations including $^1S_5$ were shown to have a high secondary free energy minimum (i.e., > 4 kcal/mol) in biased MD simulations of Gal monosaccharides [307] indicating that while possible, these conformations are not very stable in Gal monosaccharides. In both biased and unbiased MD simulations of a tetrasaccharide with a central Gal-GlcNAc disaccharide unit, the free energy minimum of Gal boat/skew-boat conformations was much higher (i.e., ~7 kcal/mol) [344], suggesting that non-$^4C_1$ conformations are less stable in Gal linked to other monosaccharides. This could explain why Gal does not sample non-$^4C_1$ conformations in keratan 20-mer MD, and suggests that these conformations would not be seen in long polymers. Although $+\phi$, $-\psi$ conformations in GlcNAcβ1-3Gal linkages are rare, they are physical conformations [351] that minorly contribute to polymer backbone flexibility. This further supports our belief that conformational landscapes from 20-mer MD are better-suited to construct conformational ensembles of GAGs with biologically-relevant chain lengths than those of short GAG oligosaccharides.

The higher degree of rigidity and probability of extended conformations in nonsulfated keratan MD compared to hyaluronan MD can be explained by the facts that: (1) for each glycosidic linkage type, the conformations that caused a kink in the polymer chain were less stable and, thus, rarer in nonsulfated keratan MD than in hyaluronan MD; and (2) GlcNAc and Gal are mostly rigid $^4C_1$ chair conformers in nonsulfated keratan whereas GlcNAc and GlcA took on more boat/skew-boat conformations that caused kinks in hyaluronan. Importantly, nonsulfated keratan 10- and 20-mers behaved randomly, and glycosidic linkages and monosaccharide rings behaved independently in MD simulations.

4.3.3.2. Construction Algorithm. Nonsulfated keratan 20-mer conformational ensembles (four sets of 10,000 conformations) were constructed using our algorithm and compared to MD-generated keratan 20-mer ensembles. The monosaccharide ring (Figure B.26) and glycosidic linkage (Figure B.27) conformations in the constructed keratan 20-mer ensemble were identical to the input data (i.e., MD-generated keratan 20-mer conformations; Figures 4.15.a,b and 4.16.a,b and Table 4.6), as expected. The end-to-end distances (Figure 4.17.a and Table 4.5) and radii of gyration (Figure B.20.a,b) were similar in constructed and MD-generated ensembles, demonstrating that nonsulfated keratan 20-mer conformational ensembles provide an accurate representation of backbone flexibility in nonsulfated keratan 20-mer MD simulations.

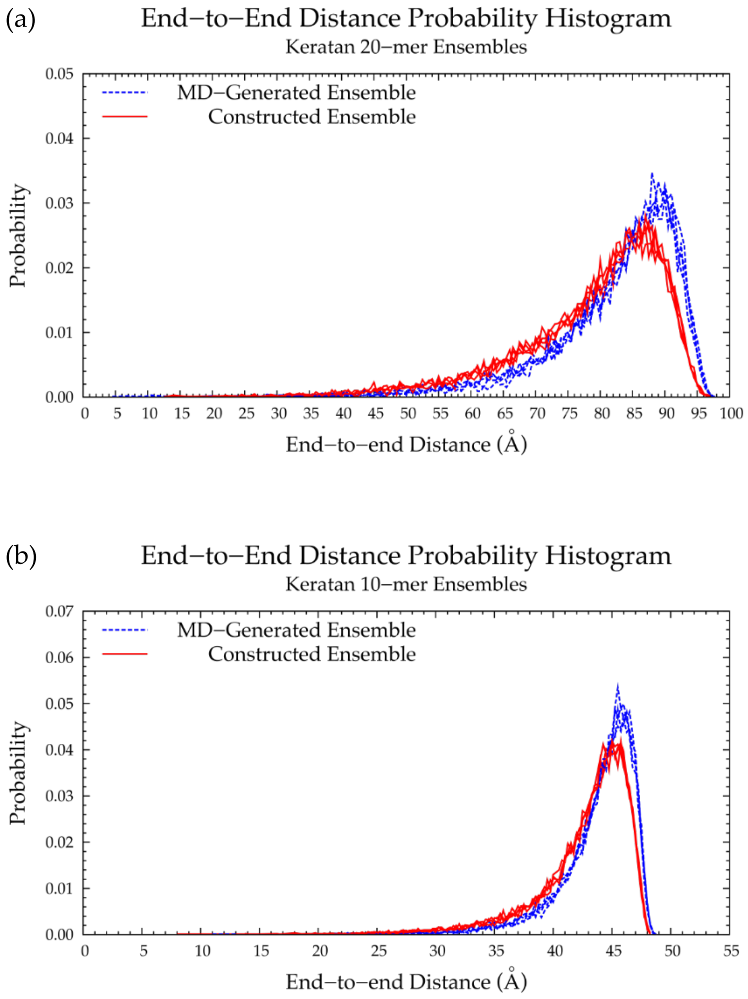


Figure 4.17. End-to-end distance probability distributions of MD-generated (blue dashed lines) and constructed (red solid lines) nonsulfated keratan ensembles: (a) 20-mer and (b) 10-mer; probabilities were calculated for end-to-end distances sorted into 0.5 Å bins for the 20-mer and 0.25 Å bins for the 10-mer; each ensemble includes four sets of 10,000 conformations.

Four sets of nonsulfated keratan 10-mer ensembles with 10,000 conformations each were constructed, and the results were compared to those of nonsulfated keratan 10-mer MD. The end-to-end distances (Figure 4.17.b and Table 4.5) and radii of gyration (Figure B.20.c,d) in constructed 10-mer ensembles matched those of MD-generated 10-mer ensembles, demonstrating that MD-generated 20-mer conformations can be applied to construct ensembles of keratan polymers of different lengths that mimic backbone flexibility seen in MD simulation. A nonsulfated keratan 200-mer ensemble was constructed and, as expected, the end-to-end distance distribution (Figure 4.18) and radii of gyration (Figure B.20.e) showed that as the polymer length increased, conformations tended to be more compact. Notably, this shift in the end-to-end distance distribution curve was more subtle in keratan than in hyaluronan, which stands to reason, as monosaccharide rings and glycosidic linkages are more rigid and tend to be more extended in keratan 20-mer MD than in hyaluronan 20-mer MD.

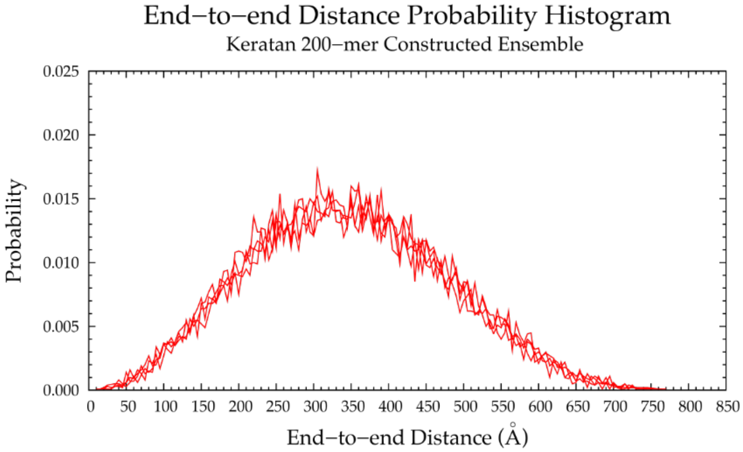


Figure 4.18. End-to-end distance probability distribution of constructed ensemble of nonsulfated keratan 200-mer; most probable end-to-end distance across all four sets is 305 Å; probabilities were calculated for end-to-end distances sorted into 5 Å bins; the ensemble contains four sets of 10,000 conformations.

4.3.4. Nonsulfated Heparan

4.3.4.1. Molecular Dynamics Simulations: Glycosidic Linkage and Monosaccharide Ring Geometry Effects on Polymer Backbone Flexibility. The backbone flexibility of nonsulfated heparan 20-mer, quantified by the end-to-end distances (Figure 4.19 and Table 4.7) and radii of gyration (Figure B.28.a) in MD simulations was analyzed. The wide end-to-end distance distribution curves and tendency toward lower end-to-end distances, relative to those of the extended 20-mer conformation, indicated that nonsulfated heparan 20-mer was highly flexible and tended toward compact conformations in MD simulations. To determine factors contributing to this conformational flexibility, monosaccharide ring and glycosidic linkage conformations were examined for patterns.

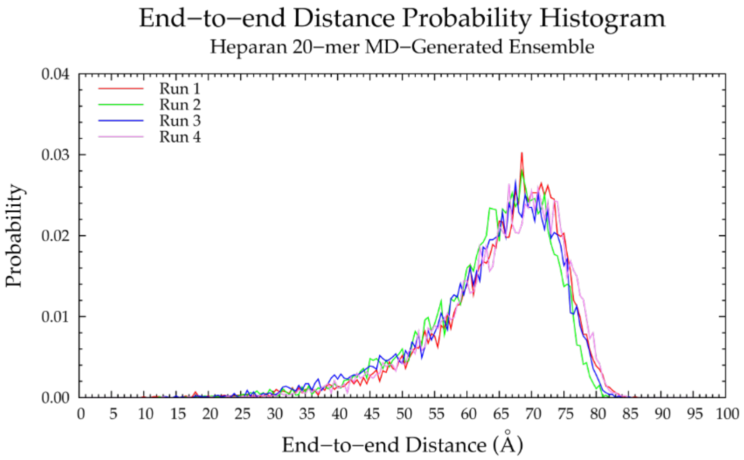


Figure 4.19. End-to-end distance probability distribution of MD-generated nonsulfated heparan 20-mer ensemble; each of the four runs includes 10,000 conformations; probabilities were calculated for end-to-end distances sorted into 0.5 Å bins.

Table 4.7. Most Probable End-to-End Distances (*d*) in MD-Generated and Constructed Nonsulfated Heparan Ensembles [1].

	20-mer Ensembles			10-mer Ensembles		
	MD-Generated *d* (Å)	Constructed *d* (Å)	% Difference	MD-Generated *d* (Å)	Constructed *d* (Å)	% Difference
Run 1	68.5	70.0		37.00	37.25	
Run 2	68.5	72.0		36.25	38.25	
Run 3	67.5	67.5		37.50	37.25	
Run 4	66.5	71.5		36.75	36.75	
All [2]	68.5	72.0	4.98%	37.00	37.25	0.67%

[1] Probabilities were calculated for end-to-end distances sorted into 0.5 Å bins for the 20-mer ensembles and 0.25 Å bins for the 10-mer ensembles. [2] All = end-to-end distance distribution aggregated across all four runs.

C-P parameters of GlcNAc rings (Figure 4.20.a) revealed all $^4C_1$ chair conformations, in line with NMR and force field data for α-GlcNAc [353]. This stands to reason, as we would not expect much difference in ring conformation between α- and β-GlcNAc, because the only structural difference is the orientation of the exocyclic oxygen atom on C_1 (i.e., the linker oxygen on nonterminal GlcNAc monosaccharides in GAG polysaccharides). This structural difference will impact glycosidic linkage conformation but is less likely to impact GlcNAc ring conformation. The IdoA monosaccharide C-P parameters (Figure 4.20.b) from nonsulfated heparan 20-mer MD were similar to those of IdoA in nonsulfated dermatan 20-mer MD (Figure 4.9.b): predominantly $^1C_4$, ^2S_O, and some $^4C_1$, which is also in line with NMR and force field data for nonsulfated IdoA in heparan sulfate oligosaccharides [310,354] and nonsulfated heparan trisaccharides [263], and occasional boat and skew-boat conformations (Figure 4.4.b,c; -4IdoAα1- endocyclic ring and linker oxygen atoms in heparan are identical to those of -4GlcAβ1- in hyaluronan). As with GalNAc in dermatan, the presence of neighboring GlcNAc monosaccharides substantially decreased sampling of $^4C_1$ conformations [310], which is in line with our results. One NMR and force field study reported 19– 49% ^2S_O conformations of nonsulfated IdoA in heparan sulfate hexasaccharides, and found that as the degree of GlcNAc sulfation increased, the percentage of ^2S_O conformations in adjacent nonsulfated IdoA decreased [354]. As we studied only nonsulfated heparan, we would expect to see a higher proportion of IdoA ^2S_O conformations (i.e., close to 49%) if simulated under the same conditions. However, the aforementioned study performed MD in aqueous solution with only neutralizing Na^+ ions, whereas our systems contained an additional 140 mM NaCl. Another NMR and molecular modeling study showed that increasing NaCl salt concentrations caused a shift in equilibrium of $^1C_4$ and ^2S_O toward $^1C_4$ conformations of 2-O-sulfated IdoA flanked by sulfated glucosamine [310]. This may explain the tendency toward $^1C_4$ IdoA conformations, with ~54% of IdoA rings across all four 20-mer MD runs in $^1C_4$, ~25% in ^2S_O, and ~4% in $^4C_1$ (Table B.3). Furthermore, very few IdoA rings sampled $^4C_1$ chair conformations, and one run contained no IdoA $^4C_1$ chair conformations

(Figure B.29), which is in contrast to the more random distribution of IdoA $^4C_1$ chair conformations in nonsulfated dermatan 20-mer MD (Figure B.16). This is likely the result of random kinetic trapping during nonsulfated heparan 20-mer MD simulations, which means IdoA only rarely overcame energy barriers between boat/skew-boat and $^4C_1$ conformations. Importantly, it appears that all relevant IdoA conformations (i.e., those in line with the literature) were sampled in nonsulfated heparan 20-mer MD simulations. Therefore, we believe our database of MD-generated 20-mer conformations contains a full conformational landscape of IdoA rings in nonsulfated heparan.

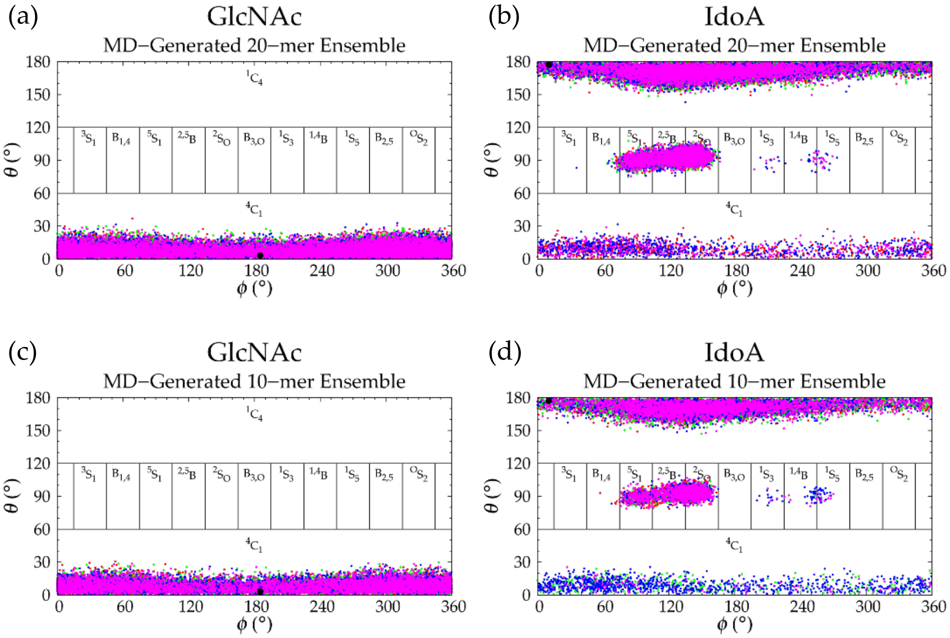


Figure 4.20. Cremer–Pople data for MD-generated nonsulfated heparan ensembles: (a) GlcNAc and (b) IdoA in the 20-mer and (c) GlcNAc and (d) IdoA in the 10-mer; geometries from the four sets of each ensemble are represented by red, green, blue, and magenta dots, respectively and the force-field geometry is represented by a single large black dot; Cremer–Pople parameters (ϕ, θ) for all rings in every tenth snapshot from each ensemble were plotted (i.e., 10 rings * 1,000 snapshots per run * 4 runs = 40,000 parameter sets for the 20-mer and 20,000 parameter sets for the 10-mer).

The primary goal of our algorithm is to predict backbone conformations of long GAG chains, which requires that (1) monosaccharide rings in GAG 20-mers behave independently in MD simulation and (2) contributions of ring puckers to backbone flexibility match those expected in nonsulfated heparan polymers in aqueous solution. There did not appear to be any interdependency between adjacent rings, suggesting that monosaccharide rings behave independently in nonsulfated heparan 20-mer MD simulations. To determine the effects of IdoA ring puckering on polymer backbone conformations, end-to-end distances of 20-mer conformations containing monosaccharides in (1) boat and skew-boat conformations that caused a kink in the polymer chain and (2) ^2S_O conformations, were analyzed (Figure B.30). The end-to-end distance distribution of 20-mer conformations with the ring puckers that caused a kink was similar to that of the average of the four MD runs and the most probable end-to-end distance was only 1 Å lower in conformations with ring puckers that caused a kink, suggesting that boat and skew-boat conformations that introduce a kink do not necessarily give less compact 20-mer conformations, and thus are not major factors contributing to backbone flexibility. The distribution curve of 20-mer conformations with ^2S_O IdoA ring puckers and that of 20-mer conformations with non-^2S_O boat/skew-boat puckers that caused a kink were qualitatively similar to that of the full 20-mer MD ensemble, but had additional peaks in probability of higher end-to-end distances. These findings suggest that IdoA boat/skew-boat conformations are associated with the full range of 20-mer backbone conformations in MD simulations with only a slight tendency toward extended 20-mer conformations. Therefore, the proportion of IdoA ^2S_O conformations likely does not have a major effect on nonsulfated heparan 20-mer backbone flexibility.

Next, glycosidic linkage dihedral free energies $\Delta G(\phi, \psi)$ were examined (Figure 4.21 and Table 4.8). IdoAα1-4GlcNAc glycosidic linkages sampled primarily $-\phi$, $+\psi$ with secondary basins in $-\phi$, $-\psi$ and GlcNAcα1-4IdoA glycosidic linkages sampled conformations in a single basin ($+\phi$, $+\psi$). We compared our results to those of an NMR and force field study that performed two sets of MD on each of nonsulfated

IdoAα1-4GlcNAc and GlcNAcα1-4IdoA disaccharides, one with IdoA restrained to ${}^1C_4$ and the other with IdoA restrained to ${}^2S_O$, each with a biasing potential on glycosidic linkage dihedrals defined by ϕ = H_1-C_1-O-C_4 and ψ = C_1-O-C_4-H_4 [191]. If we assume O_5-C_1-O-C_4 = H_1-C_1-O-C_4 - 120° in IdoAα1-4GlcNAc linkages, O_5-C_1-O-C_4 = H_1-C_1-O-C_4 + 120° in GlcNAcα1-4IdoA linkages, and C_1-O-C_4-C_3 = C_1-O-C_4-H_4 + 120° in both linkage types, our data are in close agreement with theirs.

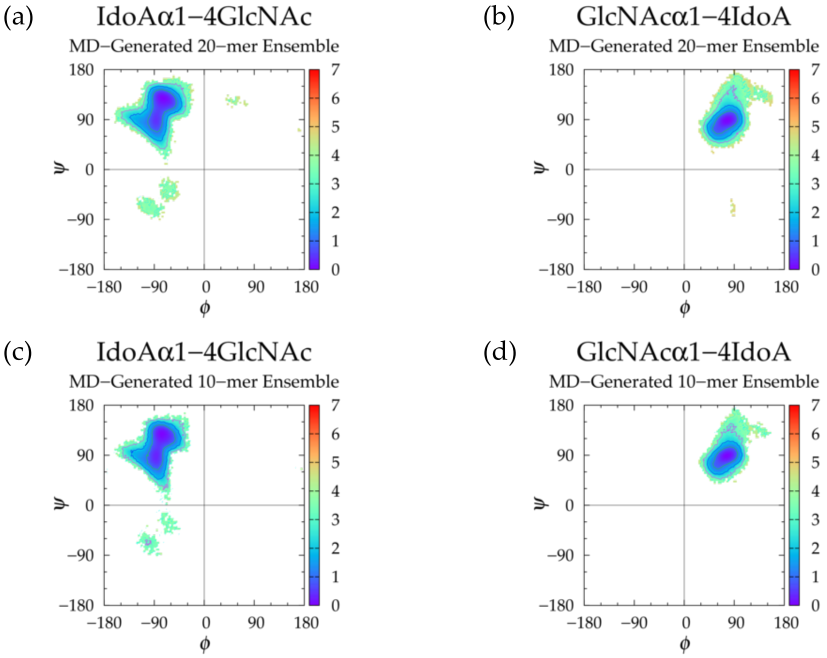


Figure 4.21. $\Delta G(\phi, \psi)$ in the MD-generated nonsulfated heparan ensembles for aggregated IdoAα1-4GlcNAc and GlcNAcα1-4IdoA glycosidic linkage data in the (a,b) 20-mer and (c,d) 10-mer, respectively; contour lines every 1 kcal/mol.

Table 4.8. Nonsulfated Heparan Glycosidic Linkage Dihedrals (ϕ, ψ) [1] and

Free Energy ($\Delta G(\phi, \psi)$ kcal/mol) Minima.

	MD-Generated 20-mer Ensemble				MD-Generated 10-mer Ensemble				Constructed 20-mer Ensemble			
	IdoAα1-4 GlcNAc		GlcNAcα1-4 IdoA		IdoAα1-4 GlcNAc		GlcNAcα1-4 IdoA		IdoAα1-4 GlcNAc		GlcNAcα1-4 IdoA	
Min	ϕ, ψ	ΔG	ϕ, ψ	ΔG	ϕ, ψ	ΔG	ϕ, ψ	ΔG	ϕ, ψ	ΔG	ϕ, ψ	ΔG
I	−73.75°, 128.75°	0.00	76.25°, 88.75°	0.00	−76.25°, 128.75°	0.00	73.75°, 88.75°	0.00	−76.25°, 128.75°	0.00	−71.25°, 116.25°	0.00
II	−103.75°, −68.75°	3.13			−103.75°, −63.75°	2.66			−98.75°, −68.75°	3.23		
II′	−66.25°, −31.25°	3.38			−66.25°, −33.75°	2.99			−61.25°, −36.25°	3.34		

[1] ϕ, ψ dihedral angles were sorted into 2.5° bins.

GlcNAcα1-4IdoA glycosidic linkage conformations differ from any other GAG glycosidic linkage conformation because of the orientation of the linker oxygen atom with respect to GlcNAc. Specifically, the oxygen atom on C_1 of α-GlcNAc is in the opposite orientation as that on C_1 of any other GAG monosaccharide (Figure 4.1). The GlcNAcα1-4IdoA linkages have a coiled conformation (Figure B.31), which helps explain the high tendency toward compact conformations in nonsulfated heparan compared to other GAGs.

As IdoAα1-4GlcNAc linkages have secondary conformations, we sought to determine if their behavior was random and independent. $\Delta G(\phi, \psi)$ was plotted for each individual IdoAα1-4GlcNAc linkage in each MD run and there did not appear to be any connection between adjacent IdoAα1-4GlcNAc linkages or any patterns across different MD runs. To determine the effects of IdoAα1-4GlcNAc linkages with $-\phi$, $-\psi$ dihedrals on 20-mer backbone flexibility, end-to-end distances of 20-mer conformations with these glycosidic linkage conformations were analyzed. Although these linkage conformations caused a kink in the polymer chain (as with hyaluronan GlcNAcβ1-4GlcA linkages), they were not associated with more compact conformations (Figure B.32). In fact, the end-to-end distance distribution of heparan 20-mer conformations with IdoAα1-4GlcNAc linkage $-\phi$, $-\psi$ dihedrals was similar to that of the full MD-generated 20-mer ensemble. This was likely because these linkage conformations occurred infrequently in nonsulfated heparan 20-mer MD, meaning a single heparan 20-mer conformation was not likely to have many kinks resulting from these $-\phi$, $-\psi$ linkage dihedrals. These findings are in line with the literature, which suggests that IdoA ring conformational flexibility in heparin/heparan sulfate oligo- and polysaccharides does not affect glycosidic linkage conformation or overall backbone shape [362].

To determine if there was any interdependency between different IdoA ring puckers and flanking IdoAα1-4GlcNAc linkage geometries in 20-mer MD simulations, conformations of all IdoAα1-4GlcNAc linkages flanking each of ${}^{1}C_{4}$, ${}^{2}S_{O}$, ${}^{4}C_{1}$, and boat/skew-boat (non-${}^{2}S_{O}$) conformations were analyzed separately (Figure B.33). There did not appear to be strong associations between any particular IdoA ring conformation and flanking IdoAα1-4GlcNAc linkage conformation, which is in line with IdoAα1-4GlcNAc disaccharide data from NMR and MD with restrained ${}^{1}C_{4}$ and ${}^{2}S_{O}$ IdoA ring conformations and a biasing potential on glycosidic linkage dihedrals [191]. This supports our hypothesis that glycosidic linkages and monosaccharide rings behave independently in MD simulation of nonsulfated heparan 20-mer.

Backbone conformational analysis of nonsulfated heparan 10-mer in MD simulation revealed flexible, compact conformations (Figure B.28.c and Table 4.7), as seen in nonsulfated heparan 20-mer MD. Monosaccharide ring and glycosidic linkage conformations in nonsulfated heparan 10-mer MD were similar to those of nonsulfated heparan 20-mer MD. These findings indicate that in MD simulations of nonsulfated heparan 10- and 20-mers, (1) IdoAα1-4GlcNAc glycosidic linkages with $-\phi$, $-\psi$ dihedrals cause a kink in the polymer chain but are rare and thus do not contribute to compact polymer conformations and (2) monosaccharide rings and glycosidic linkages behave randomly and independently.

4.3.4.2. Construction Algorithm. A nonsulfated heparan 20-mer ensemble with 40,000 conformations was constructed by the algorithm. Monosaccharide ring C-P parameters (Figure B.34) and glycosidic linkage conformations (Figure B.35) post minimization in the constructed ensemble matched the input data (Figures 4.20.a,b and 4.21.a,b), as expected. Additionally, the end-to-end distance distribution and radii of gyration of the constructed ensemble closely resembled those of the MD-generated 20-mer ensemble (Figures 4.22.a and B.28.a,b and Table 4.7), demonstrating that our algorithm generates nonsulfated heparan 20-mer ensembles with backbone conformations that mimic backbone flexibility seen in 20-mer MD.

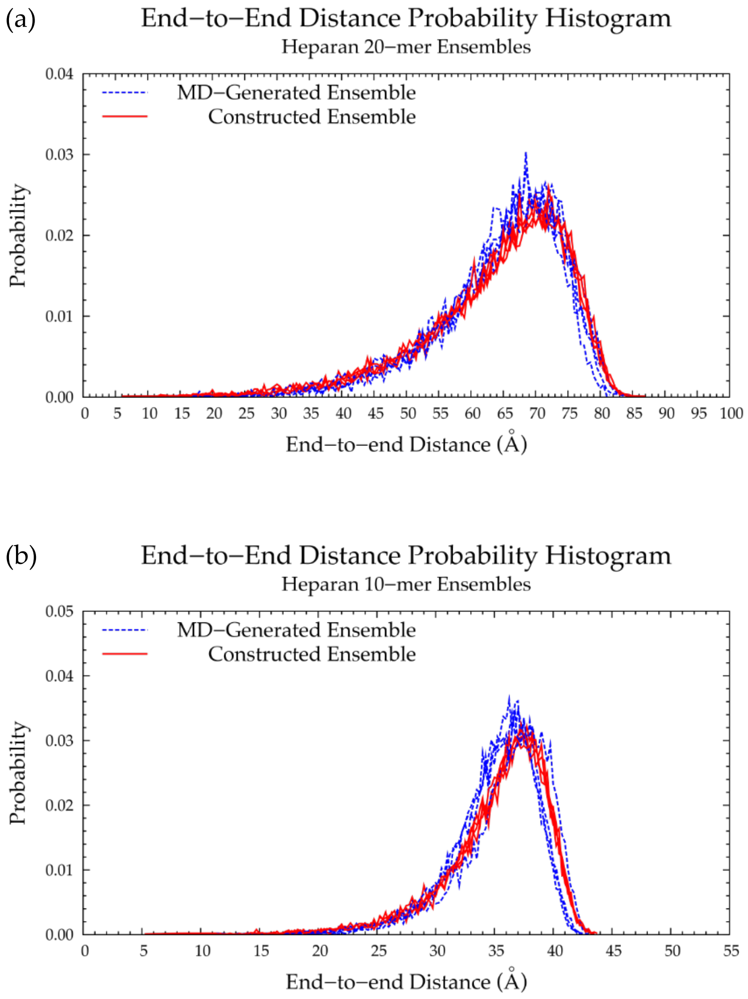


Figure 4.22. End-to-end distance probability distributions of MD-generated (blue dashed lines) and constructed (red solid lines) nonsulfated heparan ensembles: (a) 20-mer and (b) 10-mer; probabilities were calculated for end-to-end distances sorted into 0.5 Å bins for the 20-mer and 0.25 Å bins for the 10-mer; each ensemble includes four sets of 10,000 conformations.

A nonsulfated heparan 10-mer ensemble with 40,000 conformations was also constructed by the algorithm and had end-to-end distances and radii of gyration that matched those of MD-generated nonsulfated heparan 10-mer ensembles (Figures 4.22.b and B.28.c,d and Table 4.7). This demonstrates that nonsulfated heparan 20-mer conformations from MD can be used to construct 10-mer ensembles that mimic the backbone flexibility seen in nonsulfated heparan 10-mer MD.

The constructed nonsulfated heparan 200-mer ensemble had reasonably expected end-to-end distance probability distributions (Figure 4.23), i.e., the skewness of the curves shifted toward the right, indicating more compact conformations with increasing polymer length. The heparan 200-mer end-to-end distance distribution curve was qualitatively similar to that of hyaluronan 200-mer (i.e., the probability peak was of similar magnitude and most probable end-to-end distances were similar), which stands to reason, as ring and linkage conformations in nonsulfated heparan 20-mer MD cause more compact backbone conformations, as in hyaluronan 20-mer MD.

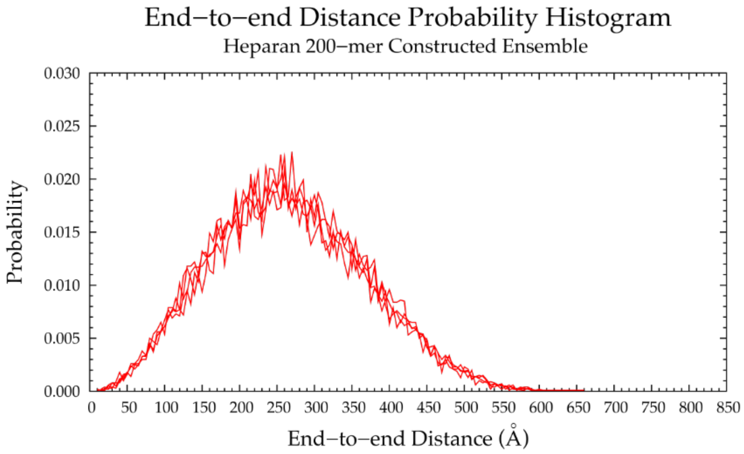


Figure 4.23. End-to-end distance probability distribution of constructed ensemble of nonsulfated heparan 200-mer; most probable end-to-end distance across all four sets is 260 Å; probabilities were calculated for end-to-end distances sorted into 5 Å bins; the ensemble contains four sets of 10,000 conformations.

4.4. Conclusions

Collectively, our findings support our hypotheses that (1) glycosidic linkages and monosaccharide rings in hyaluronan and nonsulfated dermatan, keratan, and heparan GAG 20-mers behave randomly and independently in MD simulation and (2) using a database of conformations from corresponding GAG 20-mer MD simulations and treating glycosidic linkages and monosaccharide rings independently, our algorithm can efficiently construct conformational ensembles of hyaluronan and nonsulfated dermatan, keratan, and heparan GAG 10- and 20-mers that mimic backbone flexibility observed in corresponding MD simulations. Additionally, our algorithm constructed sets of 10,000 molecular conformations of nonsulfated GAG 200-mers with backbone conformations that we would reasonably expect to see in MD simulation and did so within 12 h. This suggests that the algorithm can generate conformational ensembles of nonsulfated heterogeneous GAG polymers of arbitrary length in under a day. For perspective, 1-μs MD simulations, each producing 10,000 snapshots (i.e., 3-D atomic coordinate sets), of GAG 10- and 20-mers were completed in about 1–2 weeks and 1–2 months, respectively, and using modern GPU-accelerated hardware and software. Furthermore, the algorithm's potential energy minimization and bond potential energy cutoff criterion exclude nonphysical conformations from constructed ensembles, leaving only conformations with reasonably expected bond energies (Figures B.36, B.37, B.38, and B.39).

A comparison of the different GAG 20-mer MD simulations provided further insights into the relationship between GAG structure and conformation. For example, no associations between adjacent monosaccharide conformations were seen in any of the GAG 20-mer MD simulations, but differences in conformations of the same monosaccharide ring type in different GAGs were observed. This indicates that the structure of adjacent monosaccharides contributes to ring conformation. Furthermore, IdoA monosaccharides were much more conformationally flexible than GlcA monosaccharides in GAG 10- and 20-mer MD simulations, even though the structural difference between these two monosaccharide types (specifically, the orientation of the carboxylate group on C_5) is subtle. Interestingly, dermatan and heparan are the only GAGs with IdoA monosaccharides, yet hyaluronan and heparan showed the most conformational flexibility, and heparan showed the greatest tendency toward compact conformations compared to other GAGs. This was likely because of variability in glycosidic linkage conformation, which is independent of monosaccharide ring conformation. Independence of glycosidic linkages and monosaccharide rings was further evidenced by the fact that despite differences in flanking monosaccharide structure in different GAG types, there were similarities between all 1–3 linkage conformations and between all 1–4 linkage conformations, with the exception of GlcNAcα1–4IdoA linkages in heparan.

A comparison of MD-generated GAG 20-mer conformations to those of GAG 10-mers and to existing experimentally-determined conformations of monosaccharides and GAG oligosaccharides supported the use of 20-mer data for the construction of longer GAG polymers. For example, certain IdoA conformations (namely ${}^4C_1$ chair) were found more in unbound monosaccharide rings and terminal rings of GAG oligosaccharides than in central rings. Similarly, galactose occasionally sampled non-${}^4C_1$ conformations in monosaccharide rings and short oligosaccharides, including nonsulfated keratan 10-mer in MD simulation, but boat/skew-boat conformations of galactose were decreasingly common in central rings of polymers of increasing length. This is in line with our observation that only ${}^4C_1$ conformations of galactose were sampled in nonsulfated keratan 20-mer MD. Additionally, some nonhelical glycosidic linkage conformations that caused a kink in the polymer chain, and thus contributed to compact GAG backbone conformations, were found more in GAG 20-mers than in short GAG oligosaccharides. Therefore, conformational landscapes from GAG 20-mer MD simulations likely provide a better representation of conformations of long GAG polymers than existing conformational landscapes of monosaccharides and GAG oligosaccharides.

A comparison of backbone conformational analyses of 200-mers of different GAG types provided insights into structural features and conformational behaviors contributing to GAG polymer backbone flexibility. For example, 200-mer end-to-end distance distribution curves were qualitatively similar in nonsulfated dermatan (Figure 4.13), keratan (Figure 4.18), and chondroitin [188], whereas hyaluronan and nonsulfated heparan polymer end-to-end distance distributions (Figures 4.7 and 4.23, respectively) showed an even higher tendency toward compact conformations with increasing polymer length than other nonsulfated GAGs. To find possible explanations for this, we compared observations from MD analyses of all GAG types. In hyaluronan 10- and 20-mer MD simulations, (1) GlcNAc conformations were similar to those in nonsulfated keratan and heparan MD, and (2) GlcA conformations were similar to those in nonsulfated chondroitin MD [188]. Furthermore, IdoA rings showed more flexibility in MD simulation of nonsulfated dermatan and heparan than GlcA rings in hyaluronan and chondroitin MD [188]. However, boat/skew-boat conformations (except for $^{2}S_{O}$) in IdoA did not appear to cause more compact polymer backbone conformations than $^{1}C_{4}$, $^{2}S_{O}$, and $^{4}C_{1}$ IdoA conformations, while boat/skew-boat GlcA ring conformations were more highly associated with more compact polymer backbone conformations in hyaluronan and chondroitin [188]. Additionally, in hyaluronan 10- and 20-mer MD simulations, (1) GlcNAcβ1-4GlcA linkages took on the same conformations as Galβ1-4GlcNAc linkages in nonsulfated keratan MD, IdoAα1-4GlcNAc linkages in nonsulfated heparan MD, and GalNAcβ1-4GlcA in nonsulfated chondroitin MD [188] (i.e., primarily $-\phi$, $+\psi$ with a secondary basin at $-\phi$, $-\psi$) and (2) GlcAβ1-3GlcNAc linkages took on the same conformations as IdoAα1-3GalNAc in nonsulfated dermatan MD, GlcNAcβ1-3Gal in nonsulfated keratan MD, and GlcAβ1-3GalNAc linkages in nonsulfated chondroitin MD [188] (i.e., primarily $-\phi$, $-\psi$), but with more conformations at $-\phi$, $+\psi$. All secondary and tertiary conformations (i.e., $-\phi$, $-\psi$ in 1–4 linkages and $-\phi$, $+\psi$ in 1–3 linkages) were nonhelical and caused a kink or slight bend in the polymer chain, and were, therefore, associated with more compact conformations. The differences in energetic stability of secondary and tertiary linkage conformations between different GAG types can be explained by the adjacent monosaccharide structure. The major observation that is unique to hyaluronan is that both glycosidic linkage types take on more of these nonhelical secondary conformations in 10- and 20-mer MD simulations than any of nonsulfated dermatan, keratan, or chondroitin [188] in 10- and 20-mer MD simulations. Heparan is unique, in that it has all 1–4 linkages, while all other GAGs have alternating 1–3 and 1–4 linkages, and it contains α-GlcNAc, which gives unique conformations in GlcNAcα1-4IdoA linkages. These characteristics give nonsulfated heparan a tendency toward more coiled structures than other GAGs.

Based on these observations and the fact that nonsulfated dermatan, keratan, and chondroitin showed more extended polymer backbone conformations than hyaluronan and nonsulfated heparan in MD simulations, it is likely that: (1) monosaccharide structure and conformational flexibility determine adjacent monosaccharide conformational flexibility; (2) although there is a much higher degree of ring flexibility in IdoA than in GlcA, the flexibility of GlcA rings is more highly associated with GAG polymer backbone flexibility than that of IdoA rings; and (3) nonhelical glycosidic linkage conformations, which cause kinks in GAG polymer chains, contribute to more compact GAG polymer backbone conformations and, thus, a higher degree of GAG polymer backbone flexibility. These are valuable insights that would be difficult to obtain from conformational analyses of only solid-state structures and MD-generated conformational ensembles of short GAG oligosaccharides. Other important insights that can be gained from 3-D atomic-resolution conformational ensembles of GAG polymers include potential binding properties, and thus, predictions of binding poses with other biomolecules of interest. Models of long-chain GAG polymers (Figure 4.24 and [188]) can help characterize complete PGs and GAG-mediated complexes between multiple biomolecules, and consequently, improve our understanding of the bioactivity and function of GAG biopolymers in animal tissue.

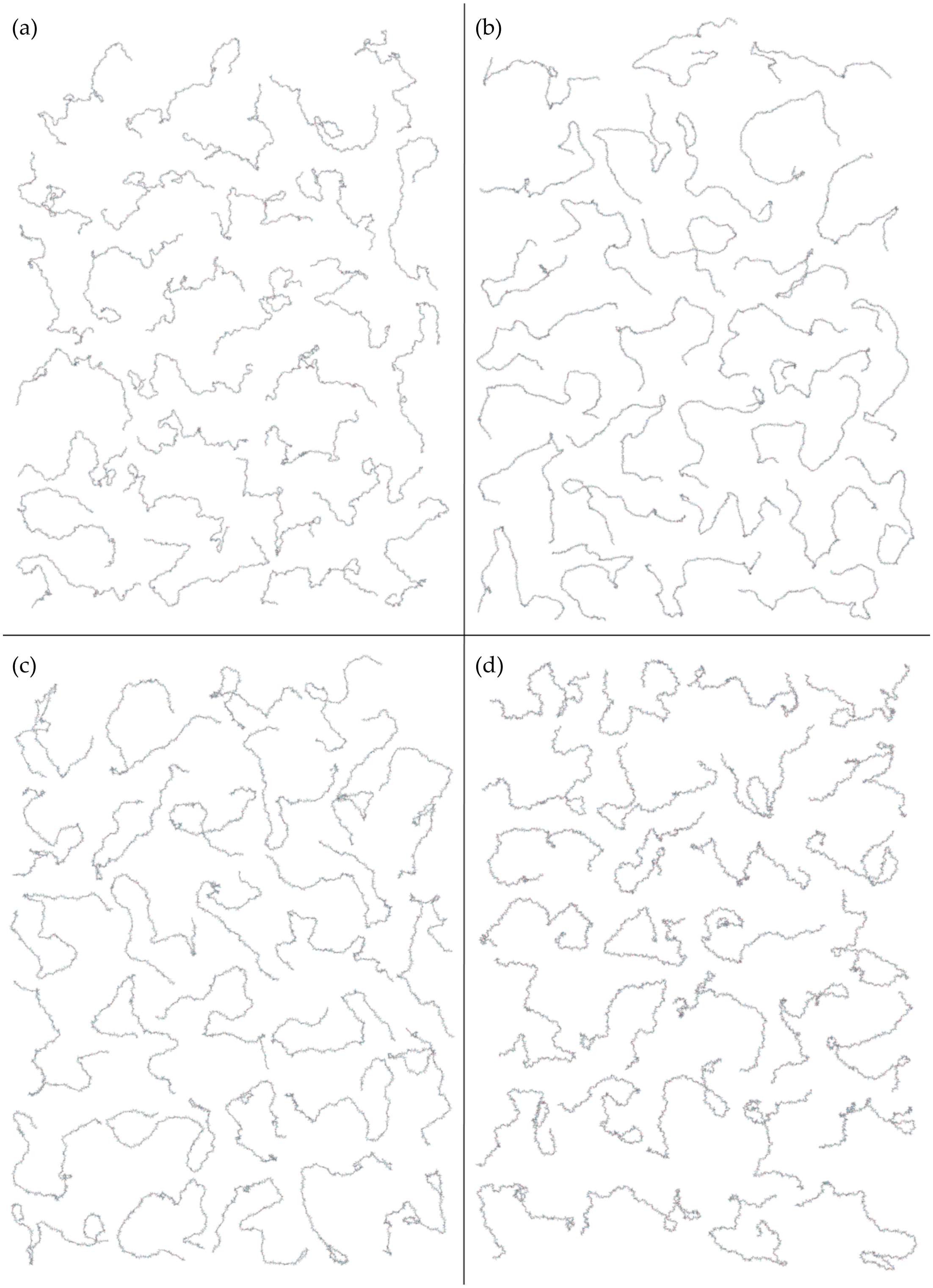


Figure 4.24. Constructed 200-mer conformations with probable end-to-end distances for each GAG type: (a) hyaluronan, (b) nonsulfated dermatan, (c) nonsulfated keratan, (d) nonsulfated heparan; n.b.: depth is not shown, so some conformations appear to have overlap.

REFERENCES

1. Mizumoto, S.; Fongmoon, D.; Sugahara, K. Interaction of chondroitin sulfate and dermatan sulfate from various biological sources with heparin-binding growth factors and cytokines. *Glycoconj. J.* **2012**, *30*, 619-632, doi:10.1007/s10719-012-9463-5.

2. Sugahara, K.; Mikami, T.; Uyama, T.; Mizuguchi, S.; Nomura, K.; Kitagawa, H. Recent advances in the structural biology of chondroitin sulfate and dermatan sulfate. *Curr. Opin. Struct. Biol.* **2003**, *13*, 612-620.

3. Sugahara, K.; Mikami, T. Chondroitin/dermatan sulfate in the central nervous system. *Curr. Opin. Struct. Biol.* **2007**, *17*, 536-545, doi:10.1016/j.sbi.2007.08.015.

4. Kawashima, H.; Atarashi, K.; Hirose, M.; Hirose, J.; Yamada, S.; Sugahara, K.; Miyasaka, M. Oversulfated chondroitin/dermatan sulfates containing GlcAbeta1/IdoAalpha1-3GalNAc(4,6-O-disulfate) interact with L- and P-selectin and chemokines. *J. Biol. Chem.* **2002**, *277*, 12921-12930, doi:10.1074/jbc.M200396200.

5. Lander, A.D.; Selleck, S.B. The elusive functions of proteoglycans: in vivo veritas. *J. Cell Biol.* **2000**, *148*, 227-232.

6. Esko, J.D.; Selleck, S.B. Order out of chaos: assembly of ligand binding sites in heparan sulfate. *Annu. Rev. Biochem.* **2002**, *71*, 435-471, doi:10.1146/annurev.biochem.71.110601.135458.

7. Kjellén, L.; Lindahl, U. Proteoglycans: structures and interactions. *Annu. Rev. Biochem.* **1991**, *60*, 443-475, doi:10.1146/annurev.bi.60.070191.002303.

8. Bernfield, M.; Götte, M.; Park, P.W.; Reizes, O.; Fitzgerald, M.L.; Lincecum, J.; Zako, M. Functions of Cell Surface Heparan Sulfate Proteoglycans. *Annu. Rev. Biochem.* **1999**, *68*, 729-777, doi:10.1146/annurev.biochem.68.1.729.

9. Kolset, S.O.; Prydz, K.; Pejler, G. Intracellular proteoglycans. *Biochem. J.* **2004**, *379*, 217-227, doi:10.1042/BJ20031230.

10. Nicholson, C.; Syková, E. Extracellular space structure revealed by diffusion analysis. *Trends Neurosci.* **1998**, *21*, 207-215.

11. Djerbal, L.; Lortat-Jacob, H.; Kwok, J. Chondroitin sulfates and their binding molecules in the central nervous system. *Glycoconj. J.* **2017**, *34*, 363-376, doi:10.1007/s10719-017-9761-z.

12. Yamada, S.; Sugahara, K.; Ozbek, S. Evolution of glycosaminoglycans: Comparative biochemical study. *Commun. Integr. Biol.* **2011**, *4*, 150-158, doi:10.4161/cib.4.2.14547.

13. Humphries, D.E.; Nicodemus, C.F.; Schiller, V.; Stevens, R.L. The human serglycin gene. Nucleotide sequence and methylation pattern in human promyelocytic leukemia HL-60 cells and T-lymphoblast Molt-4 cells. *J. Biol. Chem.* **1992**, *267*, 13558-13563.

14. Kulseth, M.A.; Kolset, S.O.; Ranheim, T. Stimulation of serglycin and CD44 mRNA expression in endothelial cells exposed to TNF-alpha and IL-1alpha. *Biochim. Biophys. Acta.* **1999**, *1428*, 225-232.

15. Schick, B.P.; Gradowski, J.F.; San Antonio, J.D. Synthesis, secretion, and subcellular localization of serglycin proteoglycan in human endothelial cells. *Blood.* **2001**, *97*, 449-458.

16. Schick, B.P.; Ho, H.-C.K.; Brodbeck, K.C.; Wrigley, C.W.; Klimas, J. Serglycin proteoglycan expression and synthesis in embryonic stem cells. *Biochim. Biophys. Acta.* **2003**, *1593*, 259-267.

17. Forsten-Williams, K.; Chu, C.L.; Fannon, M.; Buczek-Thomas, J.A.; Nugent, M.A. Control of growth factor networks by heparan sulfate proteoglycans. *Ann. Biomed. Eng.* **2008**, *36*, 2134-2148, doi:10.1007/s10439-008-9575-z.

18. Bhavanandan, V.P.; Davidson, E.A. Mucopolysaccharides associated with nuclei of cultured mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **1975**, *72*, 2032-2036.

19. Fedarko, N.S.; Conrad, H.E. A unique heparan sulfate in the nuclei of hepatocytes: structural changes with the growth state of the cells. *J. Cell Biol.* **1986**, *102*, 587-599.

20. Ishihara, M.; Fedarko, N.S.; Conrad, H.E. Transport of heparan sulfate into the nuclei of hepatocytes. *J. Biol. Chem.* **1986**, *261*, 13575-13580.

21. Tumova, S.; Hatch, B.A.; Law, D.J.; Bame, K.J. Basic fibroblast growth factor does not prevent heparan sulphate proteoglycan catabolism in intact cells, but it alters the distribution of the glycosaminoglycan degradation products. *Biochem. J.* **1999**, *337 (Pt 3)*, 471-481.

22. Kjellén, L.; Lindahl, U. Specificity of glycosaminoglycan-protein interactions. *Curr. Opin. Struct. Biol.* **2018**, *50*, 101-108, doi:10.1016/j.sbi.2017.12.011.

23. Gu, W.-L.; Fu, S.-L.; Wang, Y.-X.; Li, Y.; Lü, H.-Z.; Xu, X.-M.; Lu, P.-H. Chondroitin sulfate proteoglycans regulate the growth, differentiation and migration of multipotent neural precursor cells through the integrin signaling pathway. *BMC Neurosci.* **2009**, *10*, 128, doi:10.1186/1471-2202-10-128.

24. Cui, H.; Freeman, C.; Jacobson, G.A.; Small, D.H. Proteoglycans in the central nervous system: role in development, neural repair, and Alzheimer's disease. *IUBMB Life.* **2013**, *65*, 108-120, doi:10.1002/iub.1118.

25. Yoneda, A.; Couchman, J.R. Regulation of cytoskeletal organization by syndecan transmembrane proteoglycans. *Matrix Biol.* **2003**, *22*, 25-33.

26. Sandgren, S.; Cheng, F.; Belting, M. Nuclear targeting of macromolecular polyanions by an HIV-Tat derived peptide. Role for cell-surface proteoglycans. *J. Biol. Chem.* **2002**, *277*, 38877-38883, doi:10.1074/jbc.M205395200.

27. Serafin, W.E.; Katz, H.R.; Austen, K.F.; Stevens, R.L. Complexes of heparin proteoglycans, chondroitin sulfate E proteoglycans, and [3H]diisopropyl fluorophosphate-binding proteins are exocytosed from activated mouse bone marrow-derived mast cells. *J. Biol. Chem.* **1986**, *261*, 15017-15021.

28. Busch, S.J.; Martin, G.A.; Barnhart, R.L.; Mano, M.; Cardin, A.D.; Jackson, R.L. Trans-repressor activity of nuclear glycosaminoglycans on Fos and Jun/AP-1 oncoprotein-mediated transcription. *J. Cell Biol.* **1992**, *116*, 31-42.

29. Grammatikakis, N.; Grammatikakis, A.; Yoneda, M.; Yu, Q.; Banerjee, S.D.; Toole, B.P. A novel glycosaminoglycan-binding protein is the vertebrate homologue of the cell cycle control protein, Cdc37. *J. Biol. Chem.* **1995**, *270*, 16198-16205.

30. Salmivirta, M.; Lidholt, K.; Lindahl, U. Heparan sulfate: a piece of information. *FASEB J.* **1996**, *10*, 1270-1279.

31. Vertel, B.; Walters, L.; Flay, N.; Kearns, A.; Schwartz, N. Xylosylation is an endoplasmic reticulum to Golgi event. *J Biol Chem.* **1993**, *268*, 11105-11112.

32. Victor, X.V.; Nguyen, T.K.; Ethirajan, M.; Tran, V.M.; Nguyen, K.V.; Kuberan, B. Investigating the elusive mechanism of glycosaminoglycan biosynthesis. *J Biol Chem.* **2009**, *284*, 25842-25853.

33. Cheng, F.; Heinegård, D.; Malmström, A.; Schmidtchen, A.; Yoshida, K.; Fransson, L.A. Patterns of uronosyl epimerization and 4-/6-O-sulphation in chondroitin/dermatan sulphate from decorin and biglycan of various bovine tissues. *Glycobiology.* **1994**, *4*, 685-696.

34. Sugahara, K.; Kitagawa, H. Recent advances in the study of the biosynthesis and functions of sulfated glycosaminoglycans. *Curr Opin Struc Biol.* **2000**, *10*, 518-527.

35. Prydz, K.; Dalen, K.T. Synthesis and sorting of proteoglycans. *J Cell Sci.* **2000**, *113*, 193-205.

36. Klüppel, M.; Wight, T.N.; Chan, C.; Hinek, A.; Wrana, J.L. Maintenance of chondroitin sulfation balance by chondroitin-4-sulfotransferase 1 is required for chondrocyte development and growth factor signaling during cartilage morphogenesis. *Development.* **2005**, *132*, 3989-4003, doi:10.1242/dev.01948.

37. Shriver, Z.; Raguram, S.; Sasisekharan, R. Glycomics: a pathway to a class of new and improved therapeutics. *Nat. Rev. Drug Discov.* **2004**, *3*, 863-873, doi:10.1038/nrd1521.

38. Lindahl, U.; Couchman, J.; Kimata, K.; Esko, J.D. Proteoglycans and Sulfated Glycosaminoglycans. In *Essentials of Glycobiology*, 3rd ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY): 2015.

39. Morawski, M.; Reinert, T.; Meyer-Klaucke, W.; Wagner, F.E.; Tröger, W.; Reinert, A.; Jäger, C.; Brückner, G.; Arendt, T. Ion exchanger in the brain: Quantitative analysis of perineuronally fixed anionic binding sites suggests diffusion barriers with ion sorting properties. *Sci. Rep.* **2015**, *5*, 16471, doi:10.1038/srep16471.

40. Buckwalter, J.A.; Rosenberg, L.C. Electron microscopic studies of cartilage proteoglycans. *Electron Microsc Rev.* **1988**, *1*, 87-112.

41. Gallagher, J. Fell-Muir Lecture: Heparan sulphate and the art of cell regulation: a polymer chain conducts the protein orchestra. *Int. J. Exp. Pathol.* **2015**, *96*, 203-231, doi:10.1111/iep.12135.

42. Nandini, C.D.; Mikami, T.; Ohta, M.; Itoh, N.; Akiyama-Nambu, F.; Sugahara, K. Structural and functional characterization of oversulfated chondroitin sulfate/dermatan sulfate hybrid chains from the notochord of hagfish. Neuritogenic and binding activities for growth factors and neurotrophic factors. *J. Biol. Chem.* **2004**, *279*, 50799-50809, doi:10.1074/jbc.M404746200.

43. Beurdeley, M.; Spatazza, J.; Lee, H.H.C.; Sugiyama, S.; Bernard, C.; Di Nardo, A.A.; Hensch, T.K.; Prochiantz, A. Otx2 binding to perineuronal nets persistently regulates plasticity in the mature visual cortex. *J. Neurosci.* **2012**, *32*, 9429-9437, doi:10.1523/JNEUROSCI.0394-12.2012.

44. Nishimura, M.; Yan, W.; Mukudai, Y.; Nakamura, S.; Nakamasu, K.; Kawata, M.; Kawamoto, T.; Noshiro, M.; Hamada, T.; Kato, Y. Role of chondroitin sulfate–hyaluronan interactions in the viscoelastic properties of extracellular matrices and fluids. *BBA-Gen Subjects.* **1998**, *1380*, 1-9.

45. Rogers, C.J.; Hsieh-Wilson, L.C. Microarray Method for the Rapid Detection of Glycosaminoglycan–Protein Interactions. In *Carbohydrate Microarrays*, Chevolot, Y., Ed. Humana Press: Clifton, N.J., 2012; Vol. 808, pp. 321-336.

46. Nakano, T.; Dodd, C.M.; Scott, P.G. Glycosaminoglycans and proteoglycans from different zones of the porcine knee meniscus. *J Orthop Res.* **1997**, *15*, 213-220.

47. van Susante, J.L.C.; Pieper, J.; Buma, P.; van Kuppevelt, T.H.; van Beuningen, H.; van Der Kraan, P.M.; Veerkamp, J.H.; van den Berg, W.B.; Veth, R.P.H. Linkage of chondroitin-sulfate to type I collagen scaffolds stimulates the bioactivity of seeded chondrocytes in vitro. *Biomaterials.* **2001**, *22*, 2359-2369.

48. Vrana, N.E.; Builles, N.; Justin, V.; Bednarz, J.; Pellegrini, G.; Ferrari, B.; Damour, O.; Hulmes, D.J.; Hasirci, V. Development of a reconstructed cornea from collagen–chondroitin sulfate foams and human cell cultures. *Invest Ophth Vis Sci.* **2008**, *49*, 5325-5331.

49. Tovar, A.M.; de Mattos, D.A.; Stelling, M.P.; Sarcinelli-Luz, B.S.; Nazareth, R.A.; Mourão, P.A. Dermatan sulfate is the predominant antithrombotic glycosaminoglycan in vessel walls: implications for a possible physiological function of heparin cofactor II. *BBA-Mol Basis Dis.* **2005**, *1740*, 45-53.

50. Smith Jr, G.N.; Newsome, D.A. The nature and origin of the glycosaminoglycans of the embryonic chick vitreous body. *Dev Biol.* **1978**, *62*, 65-77.

51. Hascall, V.C.; Sajdera, S.W. Physical properties and polydispersity of proteoglycan from bovine nasal cartilage. *J. Biol. Chem.* **1970**, *245*, 4920-4930.

52. Venn, G.; Mason, R.M. Absence of keratan sulphate from skeletal tissues of mouse and rat. *Biochem J.* **1985**, *228*, 443-450.

53. Miyata, S.; Komatsu, Y.; Yoshimura, Y.; Taya, C.; Kitagawa, H. Persistent cortical plasticity by upregulation of chondroitin 6-sulfation. *Nat. Neurosci.* **2012**, *15*, 414-422, S411-412, doi:10.1038/nn.3023.

54. Herndon, M.E.; Lander, A.D. A diverse set of developmentally regulated proteoglycans is expressed in the rat central nervous system. *Neuron.* **1990**, *4*, 949-961.

55. Oohira, A.; Katoh-Semba, R.; Watanabe, E.; Matsui, F. Brain development and multiple molecular species of proteoglycan. *Neurosci. Res.* **1994**, *20*, 195-207.

56. Deepa, S.S.; Carulli, D.; Galtrey, C.; Rhodes, K.; Fukuda, J.; Mikami, T.; Sugahara, K.; Fawcett, J.W. Composition of perineuronal net extracellular matrix in rat brain: a different disaccharide composition for the net-associated proteoglycans. *J. Biol. Chem.* **2006**, *281*, 17789-17800, doi:10.1074/jbc.M600544200.

57. Yamada, S.; Van Die, I.; Van den Eijnden, D.H.; Yokota, A.; Kitagawa, H.; Sugahara, K. Demonstration of glycosaminoglycans in Caenorhabditis elegans. *FEBS Lett.* **1999**, *459*, 327-331.

58. Toyoda, H.; Kinoshita-Toyoda, A.; Selleck, S.B. Structural Analysis of Glycosaminoglycans inDrosophilaandCaenorhabditis elegansand Demonstration Thattout-velu, aDrosophilaGene Related to EXT Tumor Suppressors, Affects Heparan Sulfatein Vivo. *J. Biol. Chem.* **2000**, *275*, 2269-2275, doi:10.1074/jbc.275.4.2269.

59. Toyoda, H.; Kinoshita-Toyoda, A.; Fox, B.; Selleck, S.B. Structural analysis of glycosaminoglycans in animals bearing mutations in sugarless, sulfateless, and tout-velu. Drosophila homologues of vertebrate genes encoding glycosaminoglycan biosynthetic enzymes. *J. Biol. Chem.* **2000**, *275*, 21856-21861, doi:10.1074/jbc.M003540200.

60. MacDermott, R.P.; Schmidt, R.E.; Caulfield, J.P.; Hein, A.; Bartley, G.T.; Ritz, J.; Schlossman, S.F.; Austen, K.F.; Stevens, R.L. Proteoglycans in cell-mediated cytotoxicity. Identification, localization, and exocytosis of a chondroitin sulfate proteoglycan from human cloned natural killer cells during target cell lysis. *J. Exp. Med.* **1985**, *162*, 1771-1787.

61. Hamad, O.A.; Ekdahl, K.N.; Nilsson, P.H.; Andersson, J.; Magotti, P.; Lambris, J.D.; Nilsson, B. Complement activation triggered by chondroitin sulfate released by thrombin receptor-activated platelets. *J. Thromb. Haemost.* **2008**, *6*, 1413-1421, doi:10.1111/j.1538-7836.2008.03034.x.

62. Jones, L.L.; Yamaguchi, Y.; Stallcup, W.B.; Tuszynski, M.H. NG2 is a major chondroitin sulfate proteoglycan produced after spinal cord injury and is expressed by macrophages and oligodendrocyte progenitors. *J. Neurosci.* **2002**, *22*, 2792-2803, doi:20026258.

63. Seldin, D.C.; Austen, K.F.; Stevens, R.L. Purification and characterization of protease-resistant secretory granule proteoglycans containing chondroitin sulfate di-B and heparin-like glycosaminoglycans from rat basophilic leukemia cells. *J. Biol. Chem.* **1985**, *260*, 11131-11139.

64. Levitt, D.; Ho, P.L. Induction of chondroitin sulfate proteoglycan synthesis and secretion in lymphocytes and monocytes. *J. Cell Biol.* **1983**, *97*, 351-358.

65. Kolset, S.O.; Kjellén, L.; Seljelid, R.; Lindahl, U. Changes in glycosaminoglycan biosynthesis during differentiation in vitro of human monocytes. *Biochem. J.* **1983**, *210*, 661-667.

66. Enerbäck, L.; Kolset, S.O.; Kusche, M.; Hjerpe, A.; Lindahl, U. Glycosaminoglycans in rat mucosal mast cells. *Biochem. J.* **1985**, *227*, 661-668.

67. Thompson, H.L.; Schulman, E.S.; Metcalfe, D.D. Identification of chondroitin sulfate E in human lung mast cells. *J. Immunol.* **1988**, *140*, 2708-2713.

68. Stevens, R.L.; Fox, C.C.; Lichtenstein, L.M.; Austen, K.F. Identification of chondroitin sulfate E proteoglycans and heparin proteoglycans in the secretory granules of human lung mast cells. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 2284-2287.

69. Farrugia, B.L.; Whitelock, J.M.; O'Grady, R.; Caterson, B.; Lord, M.S. Mast Cells Produce a Unique Chondroitin Sulfate Epitope. *J. Histochem. Cytochem.* **2016**, *64*, 85-98, doi:10.1369/0022155415620649.

70. Carulli, D.; Laabs, T.; Geller, H.M.; Fawcett, J.W. Chondroitin sulfate proteoglycans in neural development and regeneration. *Curr. Opin. Neurobiol.* **2005**, *15*, 252, doi:10.1016/j.conb.2005.03.018.

71. Ripellino, J.; Margolis, R.; Margolis, R. Immunoelectron microscopic localization of hyaluronic acid-binding region and link protein epitopes in brain. *J Cell Biol.* **1989**, *108*, 1899-1907.

72. Friedlander, D.R.; Milev, P.; Karthikeyan, L.; Margolis, R.K.; Margolis, R.U.; Grumet, M. The neuronal chondroitin sulfate proteoglycan neurocan binds to the neural cell adhesion molecules Ng-CAM/L1/NILE and N-CAM, and inhibits neuronal adhesion and neurite outgrowth. *J. Cell Biol.* **1994**, *125*, 669-680.

73. Margolis, R.K.; Rauch, U.; Maurel, P.; Margolis, R.U. Neurocan and phosphacan: two major nervous tissue-specific chondroitin sulfate proteoglycans. *Perspect Dev Neurobi.* **1996**, *3*, 273-290.

74. Maeda, N.; Nishiwaki, T.; Shintani, T.; Hamanaka, H.; Noda, M. 6B4 proteoglycan/phosphacan, an extracellular variant of receptor-like protein-tyrosine phosphatase zeta/RPTPbeta, binds pleiotrophin/heparin-binding growth-associated molecule (HB-GAM). *J. Biol. Chem.* **1996**, *271*, 21446-21452.

75. Maeda, N.; Noda, M. Involvement of receptor-like protein tyrosine phosphatase zeta/RPTPbeta and its ligand pleiotrophin/heparin-binding growth-associated molecule (HB-GAM) in neuronal migration. *J. Cell Biol.* **1998**, *142*, 203-216.

76. Ueoka, C.; Kaneda, N.; Okazaki, I.; Nadanaka, S.; Muramatsu, T.; Sugahara, K. Neuronal cell adhesion, mediated by the heparin-binding neuroregulatory factor midkine, is specifically inhibited by chondroitin sulfate E. Structural ans functional implications of the over-sulfated chondroitin sulfate. *J. Biol. Chem.* **2000**, *275*, 37407-37413, doi:10.1074/jbc.M002538200.

77. Vo, T.; Carulli, D.; Ehlert, E.M.E.; Kwok, J.C.F.; Dick, G.; Mecollari, V.; Moloney, E.B.; Neufeld, G.; de Winter, F.; Fawcett, J.W., et al. The chemorepulsive axon guidance protein semaphorin3A is a constituent of perineuronal nets in the adult rodent brain. *Mol. Cell. Neurosci.* **2013**, *56*, 186-200, doi:10.1016/j.mcn.2013.04.009.

78. Dick, G.; Tan, C.L.; Alves, J.N.; Ehlert, E.M.E.; Miller, G.M.; Hsieh-Wilson, L.C.; Sugahara, K.; Oosterhof, A.; van Kuppevelt, T.H.; Verhaagen, J., et al. Semaphorin 3A binds to the perineuronal nets via chondroitin sulfate type E motifs in rodent brains. *J. Biol. Chem.* **2013**, *288*, 27384-27395, doi:10.1074/jbc.M111.310029.

79. Chen, G.; Sima, J.; Jin, M.; Wang, K.-Y.; Xue, X.-J.; Zheng, W.; Ding, Y.-Q.; Yuan, X.-B. Semaphorin-3A guides radial migration of cortical neurons during development. *Nat. Neurosci.* **2008**, *11*, 36-44, doi:10.1038/nn2018.

80. Bouzioukh, F.; Daoudal, G.; Falk, J.; Debanne, D.; Rougon, G.; Castellani, V. Semaphorin3A regulates synaptic function of differentiated hippocampal neurons. *Eur. J. Neurosci.* **2006**, *23*, 2247-2254, doi:10.1111/j.1460-9568.2006.04783.x.

81. Williamson, T.G.; Nurcombe, V.; Beyreuther, K.; Masters, C.L.; Small, D.H. Affinity purification of proteoglycans that bind to the amyloid protein precursor of Alzheimer's disease. *J. Neurochem.* **1995**, *65*, 2201-2208.

82. Wang, L.H.; Strittmatter, S.M. A family of rat CRMP genes is differentially expressed in the nervous system. *J. Neurosci.* **1996**, *16*, 6197-6207.

83. Nagai, J.; Kitamura, Y.; Owada, K.; Yamashita, N.; Takei, K.; Goshima, Y.; Ohshima, T. Crmp4 deletion promotes recovery from spinal cord injury by neuroprotection and limited scar formation. *Sci. Rep.* **2015**, *5*, 8269, doi:10.1038/srep08269.

84. Ferrer, I.; Tortosa, A.; Blanco, R.; Martín, F.; Serrano, T.; Planas, A.; Macaya, A. Naturally occurring cell death in the developing cerebral cortex of the rat. Evidence of apoptosis-associated internucleosomal DNA fragmentation. *Neurosci. Lett.* **1994**, *182*, 77-79, doi:10.1016/0304-3940(94)90210-0.

85. Oppenheim, R. Cell Death During Development Of The Nervous System. *Annu. Rev. Neurosci.* **1991**, *14*, 453-501, doi:10.1146/annurev.neuro.14.1.453.

86. Spatazza, J.; Lee, H.H.C.; Di Nardo, A.A.; Tibaldi, L.; Joliot, A.; Hensch, T.K.; Prochiantz, A. Choroid-plexus-derived Otx2 homeoprotein constrains adult cortical plasticity. *Cell Rep.* **2013**, *3*, 1815-1823, doi:10.1016/j.celrep.2013.05.014.

87. Aricescu, A.R.; McKinnell, I.W.; Halfter, W.; Stoker, A.W. Heparan sulfate proteoglycans are ligands for receptor protein tyrosine phosphatase sigma. *Mol. Cell. Biol.* **2002**, *22*, 1881-1892.

88. Coles, C.H.; Shen, Y.; Tenney, A.P.; Siebold, C.; Sutton, G.C.; Lu, W.; Gallagher, J.T.; Jones, E.Y.; Flanagan, J.G.; Aricescu, A.R. Proteoglycan-Specific Molecular Switch for RPTP Clustering and Neuronal Extension. *Science.* **2011**, *332*, 484-488, doi:10.1126/science.1200840.

89. Shen, Y. Traffic lights for axon growth: proteoglycans and their neuronal receptors. *Neural Regeneration Res.* **2014**, *9*, 356-361, doi:10.4103/1673-5374.128236.

90. Shintani, Y.; Takashima, S.; Asano, Y.; Kato, H.; Liao, Y.; Yamazaki, S.; Tsukamoto, O.; Seguchi, O.; Yamamoto, H.; Fukushima, T., et al. Glycosaminoglycan modification of neuropilin-1 modulates VEGFR2 signaling. *EMBO J.* **2006**, *25*, 3045-3055, doi:10.1038/sj.emboj.7601188.

91. Stockwell, R.; Scott, J. Distribution of acid glycosaminoglycans in human articular cartilage. *Nature.* **1967**, *215*, 1376-1378.

92. Stallcup, W.B.; Dahlin, K.; Healy, P. Interaction of the NG2 chondroitin sulfate proteoglycan with type VI collagen. *J. Cell Biol.* **1990**, *111*, 3177-3188.

93. Streit, A.; Nolte, C.; Rásony, T.; Schachner, M. Interaction of astrochondrin with extracellular matrix components and its involvement in astrocyte process formation and cerebellar granule cell migration. *J. Cell Biol.* **1993**, *120*, 799-814.

94. Henrotin, Y.; Mathy, M.; Sanchez, C.; Lambert, C. Chondroitin sulfate in the treatment of osteoarthritis: from in vitro studies to clinical recommendations. *Ther. Adv. Musculoskelet. Dis.* **2010**, *2*, 335-348, doi:10.1177/1759720x10383076.

95. du Souich, P. Comments on "OARSI guidelines for the non-surgical management of knee osteoarthritis". *Osteoarthritis Cartilage.* **2014**, *22*, 888-889, doi:10.1016/j.joca.2014.03.021.

96. Mourão, P.A. Distribution of chondroitin 4-sulfate and chondroitin 6-sulfate in human articular and growth cartilage. *Arthritis Rheum.* **1988**, *31*, 1028-1033.

97. Yokoyama, A.; Sekiya, I.; Miyazaki, K.; Ichinose, S.; Hata, Y.; Muneta, T. In vitro cartilage formation of composites of synovium-derived mesenchymal stem cells with collagen gel. *Cell Tissue Res.* **2005**, *322*, 289-298, doi:10.1007/s00441-005-0010-6.

98. Maeda, S.; Miyabayashi, T.; Yamamoto, J.K.; Roberts, G.D.; Lepine, A.J.; Clemmons, R.M. Temporal dynamic changes in synthesis of chondroitin sulfate isomers in canine articular chondrocyte culture. *J. Vet. Med. Sci.* **2003**, *65*, 1373-1376.

99. Liu, H.; Lee, Y.W.; Dean, M.F. Re-expression of differentiated proteoglycan phenotype by dedifferentiated human chondrocytes during culture in alginate beads. *Biochim. Biophys. Acta.* **1998**, *1425*, 505-515.

100. Prehm, P. Synthesis of hyaluronate in differentiated teratocarcinoma cells. Mechanism of chain growth. *Biochem J.* **1983**, *211*, 191-198.

101. Monzon, M.E.; Casalino-Matsuda, S.M.; Forteza, R.M. Identification of glycosaminoglycans in human airway secretions. *Am J Resp Cell Mol.* **2006**, *34*, 135-141.

102. Hascall, V.C.; Wang, A.; Tammi, M.; Oikari, S.; Tammi, R.; Passi, A.; Vigetti, D.; Hanson, R.W.; Hart, G.W. The dynamic metabolism of hyaluronan regulates the cytosolic concentration of UDP-GlcNAc. *Matrix Biol.* **2014**, *35*, 14-17.

103. Dowthwaite, G.; Ward, A.; Flannely, J.; Suswillo, R.; Flannery, C.; Archer, C.; Pitsillides, A. The effect of mechanical strain on hyaluronan metabolism in embryonic fibrocartilage cells. *Matrix Biol.* **1999**, *18*, 523-532.

104. Dowthwaite, G.; Flannery, C.; Flannelly, J.; Lewthwaite, J.; Archer, C.; Pitsillides, A. A mechanism underlying the movement requirement for synovial joint cavitation. *Matrix Biol.* **2003**, *22*, 311-322.

105. Carulli, D.; Rhodes, K.E.; Brown, D.J.; Bonnert, T.P.; Pollack, S.J.; Oliver, K.; Strata, P.; Fawcett, J.W. Composition of perineuronal nets in the adult rat cerebellum and the cellular origin of their components. *J Comp Neurol.* **2006**, *494*, 559-577.

106. Knudson, C.B. Hyaluronan receptor-directed assembly of chondrocyte pericellular matrix. *J Cell Biol.* **1993**, *120*, 825-834.

107. LeBaron, R.G.; Zimmermann, D.R.; Ruoslahti, E. Hyaluronate Binding Properties of Versican. *J Biol Chem.* **1992**, *267*, 10003-10010.

108. Wight, T.N.; Potter-Perigo, S.; Aulinskas, T. Proteoglycans and vascular cell proliferation. *Am Rev Respir Dis.* **1989**, *140*, 1132-1135.

109. Evanko, S.P.; Angello, J.C.; Wight, T.N. Formation of hyaluronan-and versican-rich pericellular matrix is required for proliferation and migration of vascular smooth muscle cells. *Arterioscl Thromb Vas.* **1999**, *19*, 1004-1013.

110. Margolis, R.; Crockett, C.; Kiang, W.-L.; Margolis, R. Glycosaminoglycans and glycoproteins associated with rat brain nuclei. *BBA-Gen Subjects.* **1976**, *451*, 465-469.

111. Hua, Q.; Knudson, C.B.; Knudson, W. Internalization of hyaluronan by chondrocytes occurs via receptor-mediated endocytosis. *J Cell Sci.* **1993**, *106*, 365-375.

112. Aguiar, D.J.; Knudson, W.; Knudson, C.B. Internalization of the hyaluronan receptor CD44 by chondrocytes. *Exp Cell Res.* **1999**, *252*, 292-302.

113. Furukawa, K.; Terayama, H. Isolation and identification of glycosaminoglycans associated with purified nuclei from rat liver. *BBA-Gen Subjects.* **1977**, *499*, 278-289.

114. Furukawa, K.; Terayama, H. Pattern of glycosaminoglycans and glycoproteins associated with nuclei of regenerating liver of rat. *BBA-Gen Subjects.* **1979**, *585*, 575-588.

115. Frost, S.J.; Raja, R.H.; Weigel, P.H. Characterization of an intracellular hyaluronic acid binding site in isolated rat hepatocytes. *Biochemistry.* **1990**, *29*, 10425-10432.

116. Pienimäki, J.-P.; Rilla, K.; Fülöp, C.; Sironen, R.K.; Karvinen, S.; Pasonen, S.; Lammi, M.J.; Tammi, R.; Hascall, V.C.; Tammi, M.I. Epidermal growth factor activates hyaluronan synthase 2 in epidermal keratinocytes and increases pericellular and intracellular hyaluronan. *J Biol Chem.* **2001**, *276*, 20428-20435.

117. Kaya, G.; Rodriguez, I.; Jorcano, J.L.; Vassalli, P.; Stamenkovic, I. Selective suppression of CD44 in keratinocytes of mice bearing an antisense CD44 transgene driven by a tissue-specific promoter disrupts hyaluronate metabolism in the skin and impairs keratinocyte proliferation. *Gene Dev.* **1997**, *11*, 996-1007.

118. Tammi, R.; Rilla, K.; Pienimäki, J.-P.; MacCallum, D.K.; Hogg, M.; Luukkonen, M.; Hascall, V.C.; Tammi, M. Hyaluronan enters keratinocytes by a novel endocytic route for catabolism. *J Biol Chem.* **2001**, *276*, 35111-35122.

119. Bao, X.; Mikami, T.; Yamada, S.; Faissner, A.; Muramatsu, T.; Sugahara, K. Heparin-binding growth factor, pleiotrophin, mediates neuritogenic activity of embryonic pig brain-derived chondroitin sulfate/dermatan sulfate hybrid chains. *J Biol Chem.* **2005**, *280*, 9180-9191.

120. Nandini, C.D.; Itoh, N.; Sugahara, K. Novel 70-kDa chondroitin sulfate/dermatan sulfate hybrid chains with a unique heterogenous sulfation pattern from shark skin, which exhibit neuritogenic activity and binding activities for growth factors and neurotrophic factors. *J Biol Chem.* **2005**, *280*, 4058-4069.

121. Mitsunaga, C.; Mikami, T.; Mizumoto, S.; Fukuda, J.; Sugahara, K. Chondroitin sulfate/dermatan sulfate hybrid chains in the development of cerebellum spatiotemporal regulation of the expression of critical disulfated disaccharides by specific sulfotransferases. *J Biol Chem.* **2006**, *281*, 18942-18952.

122. Maimone, M.; Tollefsen, D. Structure of a dermatan sulfate hexasaccharide that binds to heparin cofactor II with high affinity. *J Biol Chem.* **1990**, *265*, 18263-18271.

123. Denti, A.; Sini, P.; Tira, M.E.; Balduini, C. Structural heterogeneity of dermatan sulfate chains: correlation with heparin cofactor II activating properties. *Thromb Res.* **1995**, *79*, 187-198.

124. Mascellani, G.; Liverani, L.; Prete, A.; Bergonzini, G.; Torri, G.; Bisio, A.; Guerrini, M.; Casu, B.; Bianchini, P. Quantitation of dermatan sulfate active site for heparin cofactor II by 1H nuclear magnetic resonance spectroscopy. *Anal Biochem.* **1994**, *223*, 135-141.

125. Sie, P.; Dupouy, D.; Caranobe, C.; Petitou, M.; Boneu, B. Antithrombotic properties of a dermatan sulfate hexadecasaccharide fractionated by affinity for heparin cofactor II. *Blood.* **1993**.

126. Linhardt, R.J.; Al-Hakim, A.; Jian, L.; Hoppensteadt, D.; Mascellani, G.; Bianchini, P.; Fareed, J. Structural features of dermatan sulfates and their relationship to anticoagulant and antithrombotic activities. *Biochem Pharmacol.* **1991**, *42*, 1609-1619.

127. Liaw, P.C.; Austin, R.C.; Fredenburgh, J.C.; Stafford, A.R.; Weitz, J.I. Comparison of heparin-and dermatan sulfate-mediated catalysis of thrombin inactivation by heparin cofactor II. *J Biol Chem.* **1999**, *274*, 27597-27604.

128. Blinder, M.; Andersson, T.; Abildgaard, U.; Tollefsen, D. Heparin cofactor IIOslo. Mutation of Arg-189 to His decreases the affinity for dermatan sulfate. *J Biol Chem.* **1989**, *264*, 5128-5133.

129. Fromm, J.; Hileman, R.; Caldwell, E.; Weiler, J.; Linhardt, R. Differences in the interaction of heparin with arginine and lysine and the importance of these basic amino acids in the binding of heparin to acidic fibroblast growth factor. *Arch Biochem Biophys.* **1995**, *323*, 279-287.

130. Baglin, T.P.; Carrell, R.W.; Church, F.C.; Esmon, C.T.; Huntington, J.A. Crystal structures of native and thrombin-complexed heparin cofactor II reveal a multistep allosteric mechanism. *PNAS.* **2002**, *99*, 11079-11084.

131. Pike, R.N.; Buckle, A.M.; le Bonniec, B.F.; Church, F.C. Control of the coagulation system by serpins: Getting by with a little help from glycosaminoglycans. *FEBS J.* **2005**, *272*, 4842-4851.

132. Elefteriou, F.; Exposito, J.-Y.; Garrone, R.; Lethias, C. Binding of tenascin-X to decorin. *FEBS Lett.* **2001**, *495*, 44-47.

133. Lyon, M.; Deakin, J.A.; Rahmoune, H.; Fernig, D.G.; Nakamura, T.; Gallagher, J.T. Hepatocyte growth factor/scatter factor binds with high affinity to dermatan sulfate. *J Biol Chem.* **1998**, *273*, 271-278.

134. Béchard, D.; Gentina, T.; Delehedde, M.; Scherpereel, A.; Lyon, M.; Aumercier, M.; Vazeux, R.; Richet, C.; Degand, P.; Jude, B. Endocan is a novel chondroitin sulfate/dermatan sulfate proteoglycan that promotes hepatocyte growth factor/scatter factor mitogenic activity. *Journal of Biological Chemistry.* **2001**, *276*, 48341-48349.

135. Suzuki, M. Prosthetic group of cornea mucoid. *J. Biochem.* **1939**, *30*, 185-191.

136. Meyer, K.; Linker, A.; Davidson, E.A.; Weissmann, B. The mucopolysaccharides of bovine cornea. *J Biol Chem.* **1953**, *205*, 611-616.

137. Funderburgh, J.; Caterson, B.; Conrad, G. Distribution of proteoglycans antigenically related to corneal keratan sulfate proteoglycan. *J Biol Chem.* **1987**, *262*, 11634-11640.

138. Conrad, A.H.; Zhang, Y.; Tasheva, E.S.; Conrad, G.W. Proteomic analysis of potential keratan sulfate, chondroitin sulfate A, and hyaluronic acid molecular interactions. *Invest Ophth Vis Sci.* **2010**, *51*, 4500-4515.

139. Stockwell, R.; Scott, J. Observations on the acid glycosaminoglycan (mucopolysaccharide) content of the matrix of aging cartilage. *Ann Rheum Dis.* **1965**, *24*, 341.

140. Caterson, B.; Christner, J.; Baker, J. Identification of a monoclonal antibody that specifically recognizes corneal and skeletal keratan sulfate. Monoclonal antibodies to cartilage proteoglycan. *J Biol Chem.* **1983**, *258*, 8848-8854.

141. Heinegård, D.; Axelsson, I. Distribution of keratan sulfate in cartilage proteoglycans. *J Biol Chem.* **1977**, *252*, 1971-1979.

142. Lindahl, B.; Eriksson, L.; Spillmann, D.; Caterson, B.; Lindahl, U. Selective loss of cerebral keratan sulfate in Alzheimer's disease. *J. Biol. Chem.* **1996**, *271*, 16991-16994.

143. Burg, M.A.; Cole, G.J. Claustrin, an antiadhesive neural keratan sulfate proteoglycan, is structurally related to MAP1B. *J Neurobiol.* **1994**, *25*, 1-22.

144. Krusius, T.; Finne, J.; Margolis, R.; Margolis, R. Identification of an O-glycosidic mannose-linked sialylated tetrasaccharide and keratan sulfate oligosaccharides in the chondroitin sulfate proteoglycan of brain. *J Biol Chem.* **1986**, *261*, 8237-8242.

145. Bertolotto, A.; Manzardo, E.; Iudicello, M.; Guglielmone, R.; Riccio, A. Keratan sulphate is a marker of differentiation of ramified microglia. *Dev Brain Res.* **1995**, *86*, 233-241.

146. Scranton, T.W.; Iwata, M.; Carlson, S.S. The SV2 protein of synaptic vesicles is a keratan sulfate proteoglycan. *J Neurochem.* **1993**, *61*, 29-44.

147. Lindahl, U.; Kusche-Gullberg, M.; Kjellén, L. Regulated diversity of heparan sulfate. *J Biol Chem.* **1998**, *273*, 24979-24982.

148. Yahalom, J.; Eldor, A.; Fuks, Z.; Vlodavsky, I. Degradation of sulfated proteoglycans in the subendothelial extracellular matrix by human platelet heparitinase. *J Clin Invest.* **1984**, *74*, 1842-1849.

149. Matzner, Y.; Bar-Ner, M.; Yahalom, J.; Ishai-Michaeli, R.; Fuks, Z.; Vlodavsky, I. Degradation of heparan sulfate in the subendothelial extracellular matrix by a readily released heparanase from human neutrophils. Possible role in invasion through basement membranes. *J Clin Invest.* **1985**, *76*, 1306-1313.

150. Vlodavsky, I.; Fuks, Z.; Bar-Ner, M.; Ariav, Y.; Schirrmacher, V. Lymphoma cell-mediated degradation of sulfated proteoglycans in the subendothelial extracellular matrix: relationship to tumor cell metastasis. *Cancer Res.* **1983**, *43*, 2704-2711.

151. Naparstek, Y.; Cohen, I.R.; Fuks, Z.; Vlodavsky, I. Activated T lymphocytes produce a matrix-degrading heparan sulphate endoglycosidase. *Nature.* **1984**, *310*, 241-244.

152. Ishai-Michaeli, R.; Eldor, A.; Vlodavsky, I. Heparanase activity expressed by platelets, neutrophils, and lymphoma cells releases active fibroblast growth factor from extracellular matrix. *Cell Regul.* **1990**, *1*, 833-842.

153. Vlodavsky, I.; Folkman, J.; Sullivan, R.; Fridman, R.; Ishai-Michaeli, R.; Sasse, J.; Klagsbrun, M. Endothelial cell-derived basic fibroblast growth factor: synthesis and deposition into subendothelial extracellular matrix. *PNAS.* **1987**, *84*, 2292-2296.

154. Kusche, M.; Torri, G.; Casu, B.; Lindahl, U. Biosynthesis of heparin. Availability of glucosaminyl 3-O-sulfation sites. *J Biol Chem.* **1990**, *265*, 7292-7300.

155. Marcum, J.A.; Atha, D.; Fritze, L.; Nawroth, P.; Stern, D.; Rosenberg, R. Cloned bovine aortic endothelial cells synthesize anticoagulantly active heparan sulfate proteoglycan. *J Biol Chem.* **1986**, *261*, 7507-7517.

156. García, B.; Merayo-Lloves, J.; Martin, C.; Alcalde, I.; Quirós, L.M.; Vazquez, F. Surface proteoglycans as mediators in bacterial pathogens infections. *Front Microbiol.* **2016**, *7*, 220.

157. Bartlett, A.H.; Park, P.W. Proteoglycans in host–pathogen interactions: molecular mechanisms and therapeutic implications. *Expert Rev Mol Med.* **2010**, *12*.

158. Tímár, J.; Lapis, K.; Dudás, J.; Sebestyén, A.; Kopper, L.; Kovalszky, I. Proteoglycans and tumor progression: Janus-faced molecules with contradictory functions in cancer. In Proceedings of Semin Cancer Biol; pp. 173-186.

159. Iozzo, R.V.; Sanderson, R.D. Proteoglycans in cancer biology, tumour microenvironment and angiogenesis. *J Cell Mol Med.* **2011**, *15*, 1013-1031.

160. Blackhall, F.H.; Merry, C.L.; Davies, E.; Jayson, G.C. Heparan sulfate proteoglycans and cancer. *Brit J Cancer.* **2001**, *85*, 1094-1098.

161. Wade, A.; Robinson, A.E.; Engler, J.R.; Petritsch, C.; James, C.D.; Phillips, J.J. Proteoglycans and their roles in brain cancer. *FEBS J.* **2013**, *280*, 2399-2417.

162. Yutsudo, N.; Kitagawa, H. Involvement of chondroitin 6-sulfation in temporal lobe epilepsy. *Exp Neurol.* **2015**, *274*, 126-133.

163. Soleman, S.; Yip, P.K.; Duricki, D.A.; Moon, L.D. Delayed treatment with chondroitinase ABC promotes sensorimotor recovery and plasticity after stroke in aged rats. *Brain.* **2012**, *135*, 1210-1223.

164. Pantazopoulos, H.; Woo, T.-U.W.; Lim, M.P.; Lange, N.; Berretta, S. Extracellular matrix-glial abnormalities in the amygdala and entorhinal cortex of subjects diagnosed with schizophrenia. *Arch Gen Psychiat.* **2010**, *67*, 155-166.

165. Dewitt, D.A.; Silver, J.; Canning, D.R.; Perry, G. Chondroitin sulfate proteoglycans are associated with the lesions of Alzheimer's disease. *Exp Neurol.* **1993**, *121*, 149-152.

166. Dewitt, D.A.; Richey, P.; Praprotnik, D.; Silver, J.; Perry, G. Chondroitin sulfate proteoglycans are a common component of neuronal inclusions and astrocytic reaction in neurodegenerative diseases. *Brain Res.* **1994**, *656*, 205-209.

167. Bartolazzi, A.; Peach, R.; Aruffo, A.; Stamenkovic, I. Interaction between CD44 and hyaluronate is directly implicated in the regulation of tumor development. *J Exp Med.* **1994**, *180*, 53-66.

168. Sy, M.S.; Guo, Y.-J.; Stamenkovic, I. Distinct effects of two CD44 isoforms on tumor growth in vivo. *J Exp Med.* **1991**, *174*, 859-866.

169. Mizumoto, S.; Kosho, T.; Hatamochi, A.; Honda, T.; Yamaguchi, T.; Okamoto, N.; Miyake, N.; Yamada, S.; Sugahara, K. Defect in dermatan sulfate in urine of patients with Ehlers-Danlos syndrome caused by a CHST14/D4ST1 deficiency. *Clin. Biochem.* **2017**, *50*, 670-677.

170. Kosho, T. CHST14/D4ST1 deficiency: New form of Ehlers–Danlos syndrome. *Pediatr. Int.* **2016**, *58*, 88-99.

171. Wakitani, S.; Nawata, M.; Kawaguchi, A.; Okabe, T.; Takaoka, K.; Tsuchiya, T.; Nakaoka, R.; Masuda, H.; Miyazaki, K. Serum keratan sulfate is a promising marker of early articular cartilage breakdown. *Rheumatology.* **2007**, *46*, 1652-1656.

172. Thonar, E.J.M.; Lenz, M.E.; Klintworth, G.K.; Caterson, B.; Pachman, L.M.; Glickman, P.; Katz, R.; Huff, J.; Kuettner, K.E. Quantification of keratan sulfate in blood as a marker of cartilage catabolism. *Arthritis Rheum.* **1985**, *28*, 1367-1376.

173. Martin, L.; Schmitz, S.; De Santis, R.; Doemming, S.; Haase, H.; Hoeger, J.; Heinbockel, L.; Brandenburg, K.; Marx, G.; Schuerholz, T. Peptide 19-2.5 inhibits heparan sulfate-triggered inflammation in murine cardiomyocytes stimulated with human sepsis serum. *PLoS One.* **2015**, *10*.

174. Martin, L.; De Santis, R.; Koczera, P.; Simons, N.; Haase, H.; Heinbockel, L.; Brandenburg, K.; Marx, G.; Schuerholz, T. The synthetic antimicrobial peptide 19-2.5 interacts with heparanase and heparan sulfate in murine and human sepsis. *PloS One.* **2015**, *10*, e0143583.

175. Kramer, R.H.; Vogel, K.; Nicolson, G. Solubilization and degradation of subendothelial matrix glycoproteins and proteoglycans by metastatic tumor cells. *J Biol Chem.* **1982**, *257*, 2678-2686.

176. Nakajima, M.; Irimura, T.; Nicolson, G.L. Heparanases and tumor metastasis. *J Cell Biochem.* **1988**, *36*, 157-167.

177. Schulze, A.; Gripon, P.; Urban, S. Hepatitis B virus infection initiates with a large surface protein–dependent binding to heparan sulfate proteoglycans. *Hepatology.* **2007**, *46*, 1759-1768.

178. Zhao, R.-R.; Fawcett, J.W. Combination treatment with chondroitinase ABC in spinal cord injury—breaking the barrier. *Neurosci Bull.* **2013**, *29*, 477-483.

179. Bradbury, E.J.; Moon, L.D.; Popat, R.J.; King, V.R.; Bennett, G.S.; Patel, P.N.; Fawcett, J.W.; McMahon, S.B. Chondroitinase ABC promotes functional recovery after spinal cord injury. *Nature.* **2002**, *416*, 636-640.

180. Han, Y.; Gray, R.; Zhang, Y.; Burke, D.; Shields, L.; Shields, C. Can chondroitinase ABC promote axonal regeneration after spinal cord injury. *Neuroscience.* **2003**, *63*.

181. Chau, C.; Shum, D.; Li, H.; Pei, J.; Lui, Y.; Wirthlin, L.; Chan, Y.; Xu, X.M. Chondroitinase ABC enhances axonal regrowth through Schwann cell-seeded guidance channels after spinal cord injury. *FASEB J.* **2004**, *18*, 194-196.

182. Yang, S.; Hilton, S.; Alves, J.N.; Saksida, L.M.; Bussey, T.; Matthews, R.T.; Kitagawa, H.; Spillantini, M.G.; Kwok, J.C.; Fawcett, J.W. Antibody recognizing 4-sulfated chondroitin sulfate proteoglycans restores memory in tauopathy-induced neurodegeneration. *Neurobiol. Aging.* **2017**, *59*, 197-209.

183. Morla, S. Glycosaminoglycans and glycosaminoglycan mimetics in cancer and inflammation. *Int J Mol Sci.* **2019**, *20*, 1963.

184. Patel, N.J.; Sharon, C.; Baranwal, S.; Boothello, R.S.; Desai, U.R.; Patel, B.B. Heparan sulfate hexasaccharide selectively inhibits cancer stem cells self-renewal by activating p38 MAP kinase. *Oncotarget.* **2016**, *7*, 84608.

185. Borsig, L.; Wang, L.; Cavalcante, M.C.; Cardilo-Reis, L.; Ferreira, P.L.; Mourão, P.A.; Esko, J.D.; Pavão, M.S. Selectin blocking activity of a fucosylated chondroitin sulfate glycosaminoglycan from sea cucumber effect on tumor metastasis and neutrophil recruitment. *J Biol Chem.* **2007**, *282*, 14984-14991.

186. Liu, Y.; Yang, H.; Otaka, K.; Takatsuki, H.; Sakanishi, A. Effects of vascular endothelial growth factor (VEGF) and chondroitin sulfate A on human monocytic THP-1 cell migration. *Colloid Surface B.* **2005**, *43*, 216-220.

187. Pumphrey, C.Y.; Theus, A.M.; Li, S.; Parrish, R.S.; Sanderson, R.D. Neoglycans, carbodiimide-modified glycosaminoglycans: a new class of anticancer agents that inhibit cancer cell proliferation and induce apoptosis. *Cancer Res.* **2002**, *62*, 3722-3728.

188. Whitmore, E.K.; Vesenka, G.; Sihler, H.; Guvench, O. Efficient Construction of Atomic-Resolution Models of Non-Sulfated Chondroitin Glycosaminoglycan Using Molecular Dynamics Data. *Biomolecules.* **2020**, *10*, 537.

189. Gandhi, N.S.; Mancera, R.L. The Structure of Glycosaminoglycans and their Interactions with Proteins. *Chem. Biol. Drug Des.* **2008**, *72*, 455-482, doi:10.1111/j.1747-0285.2008.00741.x.

190. Casu, B.; Petitou, M.; Provasoli, M.; Sinay, P. Conformational flexibility: a new concept for explaining binding and biological properties of iduronic acid-containing glycosaminoglycans. *Trends Biochem. Sci.* **1988**, *13*, 221-225.

191. Mulloy, B.; Forster, M.J.; Jones, C.; Drake, A.F.; Johnson, E.A.; Davies, D.B. The effect of variation of substitution on the solution conformation of heparin: a spectroscopic and molecular modelling study. *Carbohydr. Res.* **1994**, *255*, 1-26.

192. Zamparo, O.; Comper, W.D. The hydrodynamic frictional coefficient of polysaccharides: the role of the glycosidic linkage. *Carbohydr. Res.* **1991**, *212*, 193-200.

193. Samsonov, S.A.; Gehrcke, J.-P.; Pisabarro, M.T. Flexibility and explicit solvent in molecular-dynamics-based docking of protein–glycosaminoglycan systems. *J. Chem. Inf. Model.* **2014**, *54*, 582-592.

194. Raju, B.B.; Bangar Raju, B.; Winnik, F.M.; Morishima, Y. A Look at the Thermodynamics of the Association of Amphiphilic Polyelectrolytes in Aqueous Solutions: Strengths and Limitations of Isothermal Titration Calorimetry. *Langmuir.* **2001**, *17*, 4416-4421, doi:10.1021/la001554i.

195. Ferro, D.R.; Provasoli, A.; Ragazzi, M.; Torri, G.; Casu, B.; Gatti, G.; Jacquinet, J.C.; Sinay, P.; Petitou, M.; Choay, J. Evidence for conformational equilibrium of the sulfated L-iduronate residue in heparin and in synthetic heparin mono- and oligo-saccharides: NMR and force-field studies. *J. Am. Chem. Soc.* **1986**, *108*, 6773-6778, doi:10.1021/ja00281a052.

196. Shuker, S.B.; Hajduk, P.J.; Meadows, R.P.; Fesik, S.W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science.* **1996**, *274*, 1531-1534, doi:10.1126/science.274.5292.1531.

197. Meyer, B.; Peters, T. NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors. *Angew. Chem. Int. Ed.* **2003**, *42*, 864-890, doi:10.1002/anie.200390233.

198. Venkataraman, G.; Sasisekharan, V.; Cooney, C.L.; Langer, R.; Sasisekharan, R. A stereochemical approach to pyranose ring flexibility: its implications for the conformation of dermatan sulfate. *Proceedings of the National Academy of Sciences.* **1994**, *91*, 6171-6175, doi:10.1073/pnas.91.13.6171.

199. Peters, T.; Pinto, B.M. Structure and dynamics of oligosaccharides: NMR and modeling studies. *Curr. Opin. Struct. Biol.* **1996**, *6*, 710-720.

200. Shao, C.; Zhang, F.; Kemp, M.M.; Linhardt, R.J.; Waisman, D.M.; Head, J.F.; Seaton, B.A. Crystallographic analysis of calcium-dependent heparin binding to annexin A2. *J. Biol. Chem.* **2006**, *281*, 31689-31695, doi:10.1074/jbc.M604502200.

201. Capila, I.; Hernáiz, M.a.J.; Mo, Y.D.; Mealy, T.R.; Campos, B.; Dedman, J.R.; Linhardt, R.J.; Seaton, B.A. Annexin V–Heparin Oligosaccharide Complex Suggests Heparan Sulfate–Mediated Assembly on Cell Surfaces. *Structure.* **2001**, *9*, 57-64, doi:10.1016/s0969-2126(00)00549-9.

202. Dementiev, A.; Petitou, M.; Herbert, J.-M.; Gettins, P.G.W. The ternary complex of antithrombin–anhydrothrombin–heparin reveals the basis of inhibitor specificity. *Nat. Struct. Mol. Biol.* **2004**, *11*, 863-867, doi:10.1038/nsmb810.

203. Li, W.; Johnson, D.J.D.; Esmon, C.T.; Huntington, J.A. Structure of the antithrombin–thrombin–heparin ternary complex reveals the antithrombotic mechanism of heparin. *Nat. Struct. Mol. Biol.* **2004**, *11*, 857-862, doi:10.1038/nsmb811.

204. Li, Z.; Kienetz, M.; Cherney, M.M.; James, M.N.G.; Brömme, D. The crystal and molecular structures of a cathepsin K:chondroitin sulfate complex. *J. Mol. Biol.* **2008**, *383*, 78-91, doi:10.1016/j.jmb.2008.07.038.

205. Imberty, A.; Lortat-Jacob, H.; Pérez, S. Structural view of glycosaminoglycan–protein interactions. *Carbohydr. Res.* **2007**, *342*, 430-439, doi:10.1016/j.carres.2006.12.019.

206. Raghuraman, A.; Mosier, P.D.; Desai, U.R. Understanding Dermatan Sulfate-Heparin Cofactor II Interaction through Virtual Library Screening. *ACS Med. Chem. Lett.* **2010**, *1*, 281-285, doi:10.1021/ml100048y.

207. Forster, M.; Mulloy, B. Computational approaches to the identification of heparin-binding sites on the surfaces of proteins. *Biochem. Soc. Trans.* **2006**, *34*, 431-434.

208. Krieger, E.; Geretti, E.; Brandner, B.; Goger, B.; Wells, T.N.; Kungl, A.J. A structural and dynamic model for the interaction of interleukin-8 and glycosaminoglycans: Support from isothermal fluorescence titrations. *Proteins: Struct., Funct., Bioinf.* **2004**, *54*, 768-775.

209. Samsonov, S.A.; Pisabarro, M.T. Computational analysis of interactions in structurally available protein–glycosaminoglycan complexes. *Glycobiology.* **2016**, *26*, 850-861.

210. Ng, C.; Nandha Premnath, P.; Guvench, O. Rigidity and Flexibility in the Tetrasaccharide Linker of Proteoglycans from Atomic-Resolution Molecular Simulation. *J. Comput. Chem.* **2017**, *38*, 1438-1446, doi:10.1002/jcc.24738.

211. Guvench, O. Revealing the Mechanisms of Protein Disorder and N-Glycosylation in CD44-Hyaluronan Binding Using Molecular Simulation. *Front. Immunol.* **2015**, *6*, 305, doi:10.3389/fimmu.2015.00305.

212. Faller, C.E.; Guvench, O. Sulfation and cation effects on the conformational properties of the glycan backbone of chondroitin sulfate disaccharides. *J. Phys. Chem. B.* **2015**, *119*, 6063-6073, doi:10.1021/jp511431q.

213. Favreau, A.J.; Faller, C.E.; Guvench, O. CD44 receptor unfolding enhances binding by freeing basic amino acids to contact carbohydrate ligand. *Biophys. J.* **2013**, *105*, 1217-1226, doi:10.1016/j.bpj.2013.07.041.

214. Guvench, O.; Mallajosyula, S.S.; Raman, E.P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T.J.; Jamison, F.W., 2nd; Mackerell, A.D., Jr. CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling. *J. Chem. Theory Comput.* **2011**, *7*, 3162-3180, doi:10.1021/ct200328p.

215. Yamada, S.; Yamane, Y.; Sakamoto, K.; Tsuda, H.; Sugahara, K. Structural determination of sulfated tetrasaccharides and hexasaccharides containing a rare disaccharide sequence, -3GalNAc(4,6-disulfate)beta1-4IdoAalpha1-, isolated from porcine intestinal dermatan sulfate. *Eur. J. Biochem.* **1998**, *258*, 775-783.

216. Almond, A.; Sheehan, J.K. Glycosaminoglycan conformation: do aqueous molecular dynamics simulations agree with x-ray fiber diffraction? *Glycobiology.* **2000**, *10*, 329-338.

217. Silipo, A.; Zhang, Z.; Cañada, F.J.; Molinaro, A.; Linhardt, R.J.; Jiménez-Barbero, J. Conformational analysis of a dermatan sulfate-derived tetrasaccharide by NMR, molecular modeling, and residual dipolar couplings. *Chembiochem.* **2008**, *9*, 240-252, doi:10.1002/cbic.200700400.

218. Sattelle, B.M.; Shakeri, J.; Cliff, M.J.; Almond, A. Proteoglycans and their heterogeneous glycosaminoglycans at the atomic scale. *Biomacromolecules.* **2015**, *16*, 951-961, doi:10.1021/bm5018386.

219. Gandhi, N.S.; Mancera, R.L. Can current force fields reproduce ring puckering in 2-O-sulfo-α-l-iduronic acid? A molecular dynamics simulation study. *Carbohydr. Res.* **2010**, *345*, 689-695.

220. Sattelle, B.M.; Hansen, S.U.; Gardiner, J.; Almond, A. Free energy landscapes of iduronic acid and related monosaccharides. *J. Am. Chem. Soc.* **2010**, *132*, 13132-13134.

221. Balogh, G.; Gyöngyösi, T.s.; Timári, I.; Herczeg, M.; Borbás, A.; Fehér, K.; Kövér, K.E. Comparison of carbohydrate force fields using Gaussian Accelerated Molecular Dynamics simulations and development of force field parameters for heparin-analogue pentasaccharides. *J. Chem. Inf. Model.* **2019**, *59*, 4855-4867.

222. Singh, K.D.; Kirubakaran, P.; Nagarajan, S.; Sakkiah, S.; Muthusamy, K.; Velmurgan, D.; Jeyakanthan, J. Homology modeling, molecular dynamics, e-pharmacophore mapping and docking study of Chikungunya virus nsP2 protease. *J Mol Model.* **2012**, *18*, 39-51.

223. Choudhury, C.; Priyakumar, U.D.; Sastry, G.N. Dynamics based pharmacophore models for screening potential inhibitors of mycobacterial cyclopropane synthase. *J Chem Inf Model.* **2015**, *55*, 848-860.

224. Pinto, M.; Perez, J.J.; Rubio-Martinez, J. Molecular dynamics study of peptide segments of the BH3 domain of the proapoptotic proteins Bak, Bax, Bid and Hrk bound to the Bcl-x L and Bcl-2 proteins. *J Comput Aid Mol Des.* **2004**, *18*, 13-22.

225. Boggara, M.B.; Krishnamoorti, R. Partitioning of nonsteroidal antiinflammatory drugs in lipid membranes: a molecular dynamics simulation study. *Biophys J.* **2010**, *98*, 586-595.

226. Afzal, M.A.F.; Browning, A.R.; Goldberg, A.; Halls, M.D.; Gavartin, J.L.; Morisato, T.; Hughes, T.F.; Giesen, D.J.; Goose, J.E. High-Throughput Molecular Dynamics Simulations and Validation of Thermophysical Properties of Polymers for Various Applications. *ACS Appl Polym Mater.* **2020**.

227. Lee, J.Y.; Kim, J.; Kim, H.; Suh, M.C. Molecular Stacking Effect on Small-Molecular Organic Light-Emitting Diodes Prepared with Solution Process. *ACS Appl Mater Interfaces.* **2020**, *12*, 23244-23251.

228. Hossain, M.; Thomas, R.; Mary, Y.S.; Resmi, K.; Armaković, S.; Armaković, S.J.; Nanda, A.K.; Vijayakumar, G.; Van Alsenoy, C. Understanding reactivity of two newly synthetized imidazole derivatives by spectroscopic characterization and computational study. *J Molec Struct.* **2018**, *1158*, 176-196.

229. Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187-217.

230. Brooks, B.R.; Brooks, C.L., III; Mackerell, A.D., Jr.; Nilsson, L.; Petrella, R.J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S., et al. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545-1614, doi:10.1002/jcc.21287.

231. MacKerell Jr, A.D.; Brooks, B.; Brooks III, C.L.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. CHARMM: The Energy Function and Its Parameterization with an Overview of the Program. In *Encyclopedia of Computational Chemistry*, von Ragué Schleyer, P., Allinger, N.L., Clark, T., Gasteiger, J., Kollman, P.A., Schaefer, H.F., Schreiner, P.R., Eds. John Wiley & Sons: Chichester, 1998; Vol. 1, pp 271-277.

232. MacKerell, A.D.; Bashford, D.; Bellott, M.; Dunbrack, R.L.; Evanseck, J.D.; Field, M.J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S., et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B.* **1998**, *102*, 3586-3616, doi:10.1021/jp973084f.

233. Alder, B.J.; Wainwright, T.E. Studies in molecular dynamics. I. General method. *J. Chem. Phys.* **1959**, *31*, 459-466.

234. Karplus, M.; Petsko, G.A. Molecular dynamics simulations in biology. *Nature.* **1990**, *347*, 631-639.

235. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.

236. Kirschner, K.N.; Yongye, A.B.; Tschampel, S.M.; González-Outeiriño, J.; Daniels, C.R.; Foley, B.L.; Woods, R.J. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008**, *29*, 622-655.

237. Singh, A.; Tessier, M.B.; Pederson, K.; Wang, X.; Venot, A.P.; Boons, G.-J.; Prestegard, J.H.; Woods, R.J. Extension and validation of the GLYCAM force field parameters for modeling glycosaminoglycans. *Can. J. Chem.* **2016**, *94*, 927-935.

238. Damm, W.; Frontera, A.; Tirado–Rives, J.; Jorgensen, W.L. OPLS all-atom force field for carbohydrates. *J. Comput. Chem.* **1997**, *18*, 1955-1970.

239. Kony, D.; Damm, W.; Stoll, S.; Van Gunsteren, W.F. An improved OPLS–AA force field for carbohydrates. *J. Comput. Chem.* **2002**, *23*, 1416-1429.

240. Hansen, H.S.; Hünenberger, P.H. A reoptimized GROMOS force field for hexopyranose-based carbohydrates accounting for the relative free energies of ring conformers, anomers, epimers, hydroxymethyl rotamers, and glycosidic linkage conformers. *J. Comput. Chem.* **2011**, *32*, 998-1032.

241. Plazinski, W.; Lonardi, A.; Hünenberger, P.H. Revision of the GROMOS 56A6CARBO force field: improving the description of ring-conformational equilibria in hexopyranose-based carbohydrates chains. *J. Comput. Chem.* **2016**, *37*, 354-365.

242. Lins, R.D.; Hünenberger, P.H. A new GROMOS force field for hexopyranose-based carbohydrates. *J. Comput. Chem.* **2005**, *26*, 1400-1412.

243. Guvench, O.; Greene, S.N.; Kamath, G.; Brady, J.W.; Venable, R.M.; Pastor, R.W.; Mackerell, A.D., Jr. Additive Empirical Force Field for Hexopyranose Monosaccharides. *J. Comput. Chem.* **2008**, *29*, 2543-2564, doi:10.1002/jcc.21004.

244. Guvench, O.; Hatcher, E.R.; Venable, R.M.; Pastor, R.W.; Mackerell, A.D. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *J. Chem. Theory Comput.* **2009**, *5*, 2353-2370, doi:10.1021/ct900242e.

245. Mallajosyula, S.S.; Guvench, O.; Hatcher, E.; Mackerell, A.D., Jr. CHARMM Additive All-Atom Force Field for Phosphate and Sulfate Linked to Carbohydrates. *J. Chem. Theory Comput.* **2012**, *8*, 759-776, doi:10.1021/ct200792v.

246. Fadda, E.; Woods, R.J. Molecular simulations of carbohydrates and protein–carbohydrate interactions: motivation, issues and prospects. *Drug Discov. Today.* **2010**, *15*, 596-609.

247. Tessier, M.B.; DeMarco, M.L.; Yongye, A.B.; Woods, R.J. Extension of the GLYCAM06 biomolecular force field to lipids, lipid bilayers and glycolipids. *Mol. Simul.* **2008**, *34*, 349-364.

248. DeMarco, M.L.; Woods, R.J. Atomic-resolution conformational analysis of the GM3 ganglioside in a lipid bilayer and its implications for ganglioside–protein recognition at membrane surfaces. *Glycobiology.* **2009**, *19*, 344-355.

249. Bock, K.; Duus, J.Ø. A conformational study of hydroxymethyl groups in carbohydrates investigated by 1H NMR spectroscopy. *J. Carbohydr. Chem.* **1994**, *13*, 513-543.

250. Nishida, Y.; Hori, H.; Ohrui, H.; Meguro, H. 1H NMR Analyses of Rotameric Distribution of C5-C6 bonds of D-Glucopyranoses in Solution. *J. Carbohydr. Chem.* **1988**, *7*, 239-250.

251. Kirschner, K.N.; Woods, R.J. Solvent interactions determine carbohydrate conformation. *PNAS.* **2001**, *98*, 10541-10545.

252. Samsonov, S.A.; Teyra, J.; Pisabarro, M.T. Docking glycosaminoglycans to proteins: analysis of solvent inclusion. *J. Comput. Aid. Mol. Des.* **2011**, *25*, 477-489.

253. Brannigan, G.; Lin, L.C.; Brown, F.L. Implicit solvent simulation models for biomembranes. *Eur. Biophys. J.* **2006**, *35*, 104-124, doi:10.1007/s00249-005-0013-y.

254. Srinivas, G.; Cheng, X.; Smith, J.C. A Solvent-Free Coarse Grain Model for Crystalline and Amorphous Cellulose Fibrils. *J. Chem. Theory Comput.* **2011**, *7*, 2539-2548, doi:10.1021/ct200181t.

255. Ingólfsson, H.I.; Lopez, C.A.; Uusitalo, J.J.; de Jong, D.H.; Gopal, S.M.; Periole, X.; Marrink, S.J. The power of coarse graining in biomolecular simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 225-248, doi:10.1002/wcms.1169.

256. Almond, A.; DeAngelis, P.L.; Blundell, C.D. Hyaluronan: the local solution conformation determined by NMR and computer modeling is close to a contracted left-handed 4-fold helix. *J. Mol. Biol.* **2006**, *358*, 1256-1269.

257. Almond, A.; Brass, A.; Sheehan, J.K. Dynamic exchange between stabilized conformations predicted for hyaluronan tetrasaccharides: comparison of molecular dynamics simulations with available NMR data. *Glycobiology.* **1998**, *8*, 973-980.

258. Almond, A.; Brass, A.; Sheehan, J. Deducing polymeric structure from aqueous molecular dynamics simulations of oligosaccharides: predictions from simulations of hyaluronan tetrasaccharides compared with hydrodynamic and X-ray fibre diffraction data. *J. Mol. Biol.* **1998**, *284*, 1425-1437.

259. Sattelle, B.M.; Shakeri, J.; Roberts, I.S.; Almond, A. A 3D-structural model of unsulfated chondroitin from high-field NMR: 4-sulfation has little effect on backbone conformation. *Carbohydr. Res.* **2010**, *345*, 291-302.

260. Whitmore, E.K.; Martin, D.; Guvench, O. Constructing 3-dimensional atomic-resolution models of nonsulfated glycosaminoglycans with arbitrary lengths using conformations from molecular dynamics. *Int J Mol Sci.* **2020**, *21*, 7699.

261. Cremer, D.; Pople, J. General definition of ring puckering coordinates. *J. Am. Chem. Soc.* **1975**, *97*, 1354-1358.

262. Ernst, S.; Venkataraman, G.; Sasisekharan, V.; Langer, R.; Cooney, C.L.; Sasisekharan, R. Pyranose ring flexibility. Mapping of physical data for iduronate in continuous conformational space. *J. Am. Chem. Soc.* **1998**, *120*, 2099-2107.

263. Forster, M.J.; Mulloy, B. Molecular dynamics study of iduronate ring conformation. *Biopolymers.* **1993**, *33*, 575-588.

264. Fushinobu, S.; Mertz, B.; Hill, A.D.; Hidaka, M.; Kitaoka, M.; Reilly, P.J. Computational analyses of the conformational itinerary along the reaction pathway of GH94 cellobiose phosphorylase. *Carbohydr. Res.* **2008**, *343*, 1023-1033.

265. Kawashima, H.; Atarashi, K.; Hirose, M.; Hirose, J.; Yamada, S.; Sugahara, K.; Miyasaka, M. Oversulfated chondroitin/dermatan sulfates containing GlcAβ1/IdoAα1-3GalNAc(4,6-O-disulfate) interact with L- and P-selectin and chemokines. *J. Biol. Chem.* **2002**, *277*, 12921-12930, doi:10.1074/jbc.M200396200.

266. Lukas, M.; Susanne, S.; De Santis Rebecca, D.S.; Hajo, H.; Janine, H.; Lena, H.; Klaus, B.; Gernot, M.; Tobias, S. Peptide 19-2.5 inhibits heparan sulfate-triggered inflammation in murine cardiomyocytes stimulated with human sepsis serum. *PLoS one.* **2015**, *10*.

267. Schmidt, E.P.; Overdier, K.H.; Sun, X.; Lin, L.; Liu, X.; Yang, Y.; Ammons, L.A.; Hiller, T.D.; Suflita, M.A.; Yu, Y. Urinary glycosaminoglycans predict outcomes in septic shock and acute respiratory distress syndrome. *Am. J. Resp. Crit. Care.* **2016**, *194*, 439-449.

268. Fawcett, J.W. The extracellular matrix in plasticity and regeneration after CNS injury and neurodegenerative disease. In *Progress in brain research*, Elsevier: 2015; Vol. 218, pp. 213-226.

269. Galtrey, C.M.; Fawcett, J.W. The role of chondroitin sulfate proteoglycans in regeneration and plasticity in the central nervous system. *Brain Res. Rev.* **2007**, *54*, 1-18.

270. Trottein, F.; Schaffer, L.; Ivanov, S.; Paget, C.; Vendeville, C.; Cazet, A.; Groux-Degroote, S.; Lee, S.; Krzewinski-Recchi, M.-A.; Faveeuw, C., et al. Glycosyltransferase and sulfotransferase gene expression profiles in human monocytes, dendritic cells and macrophages. *Glycoconj. J.* **2009**, *26*, 1259-1274, doi:10.1007/s10719-009-9244-y.

271. Toyoda, H.; Kinoshita-Toyoda, A.; Selleck, S.B. Structural Analysis of Glycosaminoglycans in Drosophila and Caenorhabditis elegans and Demonstration That tout-velu, a Drosophila Gene Related to EXT Tumor Suppressors, Affects Heparan Sulfate in Vivo. *J. Biol. Chem.* **2000**, *275*, 2269-2275, doi:10.1074/jbc.275.4.2269.

272. Yang, H.O.; Gunay, N.S.; Toida, T.; Kuberan, B.; Yu, G.; Kim, Y.S.; Linhardt, R.J. Preparation and structural determination of dermatan sulfate-derived oligosaccharides. *Glycobiology.* **2000**, *10*, 1033-1039, doi:10.1093/glycob/10.10.1033.

273. Zaia, J.; Costello, C.E. Compositional Analysis of Glycosaminoglycans by Electrospray Mass Spectrometry. *Anal. Chem.* **2001**, *73*, 233-239, doi:10.1021/ac000777a.

274. Bossennec, V.; Petitou, M.; Perly, B. 1H-n.m.r. investigation of naturally occurring and chemically oversulphated dermatan sulphates. Identification of minor monosaccharide residues. *Biochem. J.* **1990**, *267*, 625-630.

275. Ferro, D.R.; Provasoli, A.; Ragazzi, M.; Casu, B.; Torri, G.; Bossennec, V.; Perly, B.; Sinaÿ, P.; Petitou, M.; Choay, J. Conformer populations of l-iduronic acid residues in glycosaminoglycan sequences. *Carbohydr. Res.* **1990**, *195*, 157-167, doi:10.1016/0008-6215(90)84164-p.

276. Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics.* **1996**, *14*, 33-38.

277. Singh, A.; Montgomery, D.; Xue, X.; Foley, B.L.; Woods, R.J. GAG Builder: a web-tool for modeling 3D structures of glycosaminoglycans. *Glycobiology.* **2019**, *29*, 515-518, doi:10.1093/glycob/cwz027.

278. Engelsen, S.B.; Hansen, P.I.; Perez, S. POLYS 2.0: An open source software package for building three-dimensional structures of polysaccharides. *Biopolymers.* **2014**, *101*, 733-743.

279. Kuttel, M.M.; Ståhle, J.; Widmalm, G. CarbBuilder: Software for building molecular models of complex oligo-and polysaccharide structures. *J. Comput. Chem.* **2016**, *37*, 2098-2105.

280. Clerc, O.; Deniaud, M.; Vallet, S.D.; Naba, A.; Rivet, A.; Perez, S.; Thierry-Mieg, N.; Ricard-Blum, S. MatrixDB: integration of new data with a focus on glycosaminoglycan interactions. *Nucleic Acids Res.* **2018**, *47*, D376-D381.

281. Clerc, O.; Mariethoz, J.; Rivet, A.; Lisacek, F.; Pérez, S.; Ricard-Blum, S. A pipeline to translate glycosaminoglycan sequences into 3D models. Application to the exploration of glycosaminoglycan conformational space. *Glycobiology.* **2018**, *29*, 36-44.

282. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926-935.

283. Durell, S.R.; Brooks, B.R.; Ben-Naim, A. Solvent-induced forces between two hydrophilic groups. *J. Phys. Chem.* **1994**, *98*, 2198-2202.

284. Phillips, J.C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781-1802, doi:10.1002/jcc.20289.

285. Hestenes, M.R.; Stiefel, E. Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Stand.* **1952**, *49*.

286. Fletcher, R.; Reeves, C.M. Function minimization by conjugate gradients. *Comput. J.* **1964**, *7*, 149-154.

287. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H.J. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341.

288. Andersen, H.C. Rattle: A "Velocity" Version of the Shake Algorithm for Molecular Dynamics Calculations. *J. Comput. Phys.* **1983**, *52*, 24-34.

289. Miyamoto, S.; Kollman, P.A. SETTLE: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952-962.

290. Jones, J.E.; Chapman, S. On the Determination of Molecular Fields. -II. From the Equation of State of a Gas. *P. R. Soc. Lond. A-Conta.* **1924**, *106*, 463-477, doi:doi:10.1098/rspa.1924.0082.

291. Steinbach, P.J.; Brooks, B.R. New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comput. Chem.* **1994**, *15*, 667-683.

292. Allen, M.; Tildesley, D. *Computer Simulation of Liquids*; Clarendon Press: Oxford, 1987.

293. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.

294. Vanommeslaeghe, K.; MacKerell, A.D., Jr. CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochim. Biophys. Acta.* **2015**, *1850*, 861-871, doi:10.1016/j.bbagen.2014.08.004.

295. Feller, S.E.; Zhang, Y.; Pastor, R.W.; Brooks, B.R. Constant pressure molecular dynamics simulation: the Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613-4621.

296. Kubo, R.; Toda, M.; Hashitsume, N. Statistical Physics II: Nonequilibrium Statistical Mechanics. In *Springer Series in Solid-State Sciences*, 2nd ed.; Cardona, M., Fulde, P., von Klitzing, K., Queisser, H.-J., Eds. Springer-Verlag: Berlin Heidelberg New York, 1991; Vol. 31.

297. Eastman, P.; Swails, J.; Chodera, J.D.; McGibbon, R.T.; Zhao, Y.; Beauchamp, K.A.; Wang, L.-P.; Simmonett, A.C.; Harrigan, M.P.; Stern, C.D. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017**, *13*, e1005659.

298. Friedrichs, M.S.; Eastman, P.; Vaidyanathan, V.; Houston, M.; Legrand, S.; Beberg, A.L.; Ensign, D.L.; Bruns, C.M.; Pande, V.S. Accelerating Molecular Dynamic Simulation on Graphics Processing Units. *J. Comput. Chem.* **2009**, *30*, 864-872.

299. Eastman, P.; Pande, V. OpenMM: A Hardware-Independent Framework for Molecular Simulations. *Comput. Sci. Eng.* **2010**, *12*, 34-39.

300. Eastman, P.; Pande, V.S. Constant Constraint Matrix Approximation: A Robust, Parallelizable Constraint Method for Molecular Simulations. *J. Chem. Theory Comput.* **2010**, *6*, 434-437.

301. Eastman, P.; Friedrichs, M.S.; Chodera, J.D.; Radmer, R.J.; Bruns, C.M.; Ku, J.P.; Beauchamp, K.A.; Lane, T.J.; Wang, L.-P.; Shukla, D. OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *J. Chem. Theory Comput.* **2012**, *9*, 461-469.

302. Eastman, P.; Pande, V.S. Efficient Nonbonded Interactions for Molecular Dynamics on a Graphics Processing Unit. *J. Comput. Chem.* **2010**, *31*, 1268-1272.

303. Kiss, P.T.; Sega, M.; Baranyai, A. Efficient Handling of Gaussian Charge Distributions: An Application to Polarizable Molecular Models. *J. Chem. Theory Comput.* **2014**, *10*, 5513-5519, doi:10.1021/ct5009069.

304. Gingrich, T.R.; Wilson, M. On the Ewald summation of Gaussian charges for the simulation of metallic surfaces. *Chem. Phys. Lett.* **2010**, *500*, 178-183.

305. Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159*, 98-103.

306. Gatti, G.; Casu, B.; Torri, G.; Vercellotti, J.R. Resolution-enhanced 1H-nmr spectra of dermatan sulfate and chondroitin sulfates: conformation of the uronic acid residues. *Carbohydr. Res.* **1979**, *68*, C3-C7.

307. Alibay, I.; Bryce, R.A. Ring puckering landscapes of glycosaminoglycan-related monosaccharides from molecular dynamics simulations. *J. Chem. Inf. Model.* **2019**, *59*, 4729-4741.

308. Nyerges, B.; Kovacs, A. Density functional study of the conformational space of 4C1 D-glucuronic acid. *J. Phys. Chem. A.* **2005**, *109*, 892-897, doi:10.1021/jp047451g.

309. Scott, J.E. Supramolecular organization of extracellular matrix glycosaminoglycans, in vitro and in the tissues. *FASEB J.* **1992**, *6*, 2639-2645.

310. Van Boeckel, C.; Van Aelst, S.; Wagenaars, G.; Mellema, J.R.; Paulsen, H.; Peters, T.; Pollex, A.; Sinnwell, V. Conformational analysis of synthetic heparin-like oligosaccharides containing α-L-idopyranosyluronic acid. *Recl. Trav. Chim. Pays-Bas.* **1987**, *106*, 19-29.

311. Agrawal, P.K.; Jacquinet, J.-C.; Krishna, N.R. NMR and molecular modeling studies on two glycopeptides from the carbohydrate-protein linkage region of connective tissue proteoglycans. *Glycobiology.* **1999**, *9*, 669-677.

312. Choe, B.Y.; Ekborg, G.C.; Roden, L.; Harvey, S.C.; Krishna, N.R. High-resolution NMR and molecular modeling studies on complex carbohydrates: Characterization of O-β-D-Gal-(1→3)-O-β-D-Gal-(1→4)-O-β-D-Xyl-(1→ 0)-L-Ser, a carbohydrate-protein linkage region fragment from connective tissue proteoglycans. *J. Amer. Chem. Soc.* **1991**, *113*, 3743-3749.

313. Krishna, N.; Choe, B.; Prabhakaran, M.; Ekborg, G.C.; Rodén, L.; Harvey, S. Nuclear magnetic resonance and molecular modeling studies on O-β-D-galactopyranosyl-(1→4)-O-β-D-xylopyranosyl-(1→0)-L-serine, a carbohydrate-protein linkage region fragment from connective tissue proteoglycans. *J. Biol. Chem.* **1990**, *265*, 18256-18262.

314. Sharif, M.; George, E.; Shepstone, L.; Knudson, W.; Thonar, E.J.M.; Cushnaghan, J.; Dieppe, P. Serum hyaluronic acid level as a predictor of disease progression in osteoarthritis of the knee. *Arthritis Rheum.* **1995**, *38*, 760-767.

315. Ghosh, P. The role of hyaluronic acid (hyaluronan) in health and disease: interactions with cells, cartilage and components of synovial fluid. *Clin. Exp. Rheum.* **1994**, *12*, 75-82.

316. Hascall, V.C.; Heinegård, D. Aggregation of cartilage proteoglycans II. Oligosaccharide competitors of the proteoglycan-hyaluronic acid interaction. *J. Biol. Chem.* **1974**, *249*, 4242-4249.

317. Winter, W.; Smith, P.; Arnott, S. Hyaluronic acid: structure of a fully extended 3-fold helical sodium salt and comparison with the less extended 4-fold helical forms. *J. Mol. Biol.* **1975**, *99*, 219-235.

318. Sheehan, J.; Gardner, K.; Atkins, E. Hyaluronic acid: a double-helical structure in the presence of potassium at low pH and found also with the cations ammonium, rubidium and caesium. *J. Mol. Biol.* **1977**, *117*, 113-135.

319. Almond, A.; Sheehan, J.K.; Brass, A. Molecular dynamics simulations of the two disaccharides of hyaluronan in aqueous solution. *Glycobiology.* **1997**, *7*, 597-604.

320. Gribbon, P.; Heng, B.C.; Hardingham, T.E. The molecular basis of the solution properties of hyaluronan investigated by confocal fluorescence recovery after photobleaching. *Biophys. J.* **1999**, *77*, 2210-2216.

321. Atkins, E.; Meader, D.; Scott, J. Model for hyaluronic acid incorporating four intramolecular hydrogen bonds. *Int. J. Biol. Macromol.* **1980**, *2*, 318-319.

322. Scott, J.E.; Cummings, C.; Brass, A.; Chen, Y. Secondary and tertiary structures of hyaluronan in aqueous solution, investigated by rotary shadowing-electron microscopy and computer simulation. *Biochem. J.* **1991**, *274*, 699-705.

323. Scott, J.; Heatley, F.; Moorcroft, D.; Olavesen, A. Secondary structures of hyaluronate and chondroitin sulphates. A 1H nmr study of NH signals in dimethyl sulphoxide solution. *Biochem. J.* **1981**, *199*, 829.

324. Jo, S.; Kim, T.; Iyer, V.G.; Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859-1865.

325. Jo, S.; Song, K.C.; Desaire, H.; MacKerell Jr, A.D.; Im, W. Glycan Reader: automated sugar identification and simulation preparation for carbohydrates and glycoproteins. *J. Comput. Chem.* **2011**, *32*, 3135-3141.

326. Park, S.-J.; Lee, J.; Patel, D.S.; Ma, H.; Lee, H.S.; Jo, S.; Im, W. Glycan Reader is improved to recognize most sugar types and chemical modifications in the Protein Data Bank. *Bioinformatics.* **2017**, *33*, 3051-3057.

327. Park, S.-J.; Lee, J.; Qi, Y.; Kern, N.R.; Lee, H.S.; Jo, S.; Joung, I.; Joo, K.; Lee, J.; Im, W. CHARMM-GUI glycan modeler for modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology.* **2019**, *29*, 320-331.

328. Beglov, D.; Roux, B. Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J. Chem. Phys.* **1994**, *100*, 9050-9063.

329. Banerji, S.; Wright, A.J.; Noble, M.; Mahoney, D.J.; Campbell, I.D.; Day, A.J.; Jackson, D.G. Structures of the CD44–hyaluronan complex provide insight into a fundamental carbohydrate-protein interaction. *Nat. Struct. & Molec. Biol.* **2007**, *14*, 234.

330. Jedrzejas, M.J.; Mello, L.V.; de Groot, B.L.; Li, S. Mechanism of Hyaluronan Degradation by Streptococcus pneumoniae Hyaluronate Lyase. *J. Biol. Chem.* **2002**, *277*, 28287-28297.

331. Mello, L.V.; de Groot, B.L.; Li, S.; Jedrzejas, M.J. Structure and Flexibility of Streptococcus agalactiae Hyaluronate Lyase Complex with Its Substrate. *J. Biol. Chem.* **2002**, *277*, 36678-36688.

332. Marković-Housley, Z.; Miglierini, G.; Soldatova, L.; Rizkallah, P.J.; Müller, U.; Schirmer, T. Crystal structure of hyaluronidase, a major allergen of bee venom. *Structure.* **2000**, *8*, 1025-1035.

333. Tashima, T.; Nagatoishi, S.; Caaveiro, J.M.; Nakakido, M.; Sagara, H.; Kusano-Arai, O.; Iwanari, H.; Mimuro, H.; Hamakubo, T.; Ohnuma, S.-i. Molecular basis for governing the morphology of type-I collagen fibrils by Osteomodulin. *Commun. Biol.* **2018**, *1*, 1-10.

334. Paracuellos, P.; Kalamajski, S.; Bonna, A.; Bihan, D.; Farndale, R.W.; Hohenester, E. Structural and functional analysis of two small leucine-rich repeat proteoglycans, fibromodulin and chondroadherin. *Matrix Biol.* **2017**, *63*, 106-116.

335. Ohto, U.; Usui, K.; Ochi, T.; Yuki, K.; Satow, Y.; Shimizu, T. Crystal Structure of Human β-Galactosidase STRUCTURAL BASIS OF GM1 GANGLIOSIDOSIS AND MORQUIO B DISEASES. *J. Biol. Chem.* **2012**, *287*, 1801-1812.

336. Scally, S.W.; Petersen, J.; Law, S.C.; Dudek, N.L.; Nel, H.J.; Loh, K.L.; Wijeyewickrema, L.C.; Eckle, S.B.; van Heemst, J.; Pike, R.N. A molecular basis for the association of the HLA-DRB1 locus, citrullination, and rheumatoid arthritis. *J. Exp. Med.* **2013**, *210*, 2569-2582.

337. Sattelle, B.M.; Almond, A. Is N-acetyl-d-glucosamine a rigid 4C1 chair? *Glycobiology.* **2011**, *21*, 1651-1662.

338. Siciňska, W.; Adams, B.; Lerner, L. A detailed 1H and 13C NMR study of a repeating disaccharide of hyaluronan: the effects of temperature and counterion type. *Carbohydr. Res.* **1993**, *242*, 29-51.

339. Gargiulo, V.; Morando, M.A.; Silipo, A.; Nurisso, A.; Pérez, S.; Imberty, A.; Cañada, F.J.; Parrilli, M.; Jiménez-Barbero, J.; De Castro, C. Insights on the conformational properties of hyaluronic acid by using NMR residual dipolar couplings and MD simulations. *Glycobiology.* **2010**, *20*, 1208-1216.

340. Ikemizu, S.; Sparks, L.M.; van der Merwe, P.A.; Harlos, K.; Stuart, D.I.; Jones, E.Y.; Davis, S.J. Crystal structure of the CD2-binding domain of CD58 (lymphocyte function-associated antigen 3) at 1.8-Å resolution. *PNAS.* **1999**, *96*, 4289-4294.

341. Chang, C.; Magracheva, E.; Kozlov, S.; Fong, S.; Tobin, G.; Kotenko, S.; Wlodawer, A.; Zdanov, A. Crystal structure of interleukin-19 defines a new subfamily of helical cytokines. *J. Biol. Chem.* **2003**, *278*, 3308-3313.

342. Xu, R.; McBride, R.; Paulson, J.C.; Basler, C.F.; Wilson, I.A. Structure, receptor binding, and antigenicity of influenza virus hemagglutinins from the 1957 H2N2 pandemic. *J. Virol.* **2010**, *84*, 1715-1721.

343. Maveyraud, L.; Niwa, H.; Guillet, V.; Svergun, D.I.; Konarev, P.V.; Palmer, R.A.; Peumans, W.J.; Rougé, P.; Van Damme, E.J.; Reynolds, C.D. Structural basis for sugar recognition, including the Tn carcinoma antigen, by the lectin SNA-II from Sambucus nigra. *Proteins: Struct., Funct., Bioinf.* **2009**, *75*, 89-103.

344. Alibay, I.; Burusco, K.K.; Bruce, N.J.; Bryce, R.A. Identification of rare Lewis oligosaccharide conformers in aqueous solution using enhanced sampling molecular dynamics. *J. Phys. Chem. B.* **2018**, *122*, 2462-2474.

345. Blundell, C.D.; Roberts, I.S.; Sheehan, J.K.; Almond, A. Investigating the molecular basis for the virulence of Escherichia coli K5 by nuclear magnetic resonance analysis of the capsule polysaccharide. *J. Mol. Microbiol. Biotechnol.* **2009**, *17*, 71-82.

346. Furlan, S.; La Penna, G.; Perico, A.; Cesàro, A. Hyaluronan chain conformation and dynamics. *Carbohydr. Res.* **2005**, *340*, 959-970.

347. Holmbeck, S.M.; Petillo, P.A.; Lerner, L.E. The solution conformation of hyaluronan: a combined NMR and molecular dynamics study. *Biochemistry.* **1994**, *33*, 14246-14255.

348. Adrian-Scotto, M.; Abdallah, K.B.; Mallet, G.; Vasilescu, D. Quantum molecular modeling of free radical saccharides from hyaluronan. *J. Molec. Struct.* **2003**, *636*, 89-113.

349. Moulabbi, M.; Broch, H.; Robert, L.; Vasilescu, D. Quantum molecular modeling of hyaluronan. *J. Molec. Struct.* **1997**, *395*, 477-508.

350. Guss, J.; Hukins, D.; Smith, P.; Winter, W.; Arnott, S.; Moorhouse, R.; Rees, D. Hyaluronic acid: molecular conformations and interactions in two sodium salts. *J. Molec. Biol.* **1975**, *95*, 359-384.

351. Rees, D. Conformational analysis of polysaccharides. Part II. Alternating copolymers of the agar–carrageenan–chondroitin type by model building in the computer with calculation of helical parameters. *J. Chem. Soc. B.* **1969**, 217-226.

352. Mitra, A.; Arnott, S.; Atkins, E.; Isaac, D. Dermatan sulfate: molecular conformations and interactions in the condensed state. *J. Molec. Biol.* **1983**, *169*, 873-901.

353. Virudachalam, R.; Rao, V.S. Theoretical investigations on 2-acetamido-2-deoxy aldohexopyranoses: conformation and the anomeric effect. *Carbohydr. Res.* **1976**, *51*, 135-139.

354. Hsieh, P.-H.; Thieker, D.F.; Guerrini, M.; Woods, R.J.; Liu, J. Uncovering the relationship between sulphation patterns and conformation of iduronic acid in heparan sulphate. *Nature.* **2016**, *6*, 29602.

355. Inoue, Y.; Inouye, Y.; Nagasawa, K. Conformational equilibria of the L-iduronate residue in non-sulphated di-, tetra-and hexa-saccharides and their alditols derived from dermatan sulphate. *Biochem. J.* **1990**, *265*, 533-538.

356. Ragazzi, M.; Provasoli, A.; Ferro, D. Molecular Mechanics and the Structure of Iduronate-Containing Carbohydrates. In *Computer Modeling of Carbohydrate Molecules*, French, A., Ed. ACS Publications: Washington, D.C., 1990; pp. 332-344.

357. Venkataraman, G.; Sasisekharan, V.; Cooney, C.L.; Langer, R.; Sasisekharan, R. A stereochemical approach to pyranose ring flexibility: its implications for the conformation of dermatan sulfate. *Proc Natl Acad Sci U S A.* **1994**, *91*, 6171-6175, doi:10.1073/pnas.91.13.6171.

358. Arnott, S.; Guss, J.; Hukins, D.; Dea, I.; Rees, D. Conformation of keratan sulphate. *J. Mol. Biol.* **1974**, *88*, 175-184.

359. Rudrum, M.; Shaw, D. 10. The structure and conformation of some monosaccharides in solution. *J. Chem. Soc.* **1965**, 52-57.

360. Angyal, S.J. Conformational analysis in carbohydrate chemistry: I. Conformational free energies. The conformations and $\alpha : \beta$ ratios of aldopyranoses in aqueous solution. *Aust. J. Chem.* **1968**, *21*, 2737-2746.

361. Bush, C.A.; Feeney, R.E. Conformation of the glycotripeptide repeating unit of antifreeze glycoprotein of polar fish as determined from the fully assigned proton nmr spectrum. *Int. J. Pept. Prot. Res.* **1986**, *28*, 386-397.

362. Mulloy, B.; Forster, M.J. Conformation and dynamics of heparin and heparan sulfate. *Glycobiology.* **2000**, *10*, 1147-1156.

363. Darve, E.; Rodríguez-Gómez, D.; Pohorille, A. Adaptive biasing force method for scalar and vector free energy calculations. *J. Chem. Phys.* **2008**, *128*, 144120.

364. Henin, J.; Fiorin, G.; Chipot, C.; Klein, M.L. Exploring multidimensional free energy landscapes using time-dependent biases on collective variables. *J. Chem. Theory Comput.* **2009**, *6*, 35-47.

APPENDIX A: SUPPORTING INFORMATION FOR “EFFICIENT CONSTRUCTION OF ATOMIC-RESOLUTION MODELS OF NONSULFATED CHONDROITIN GLYCOSAMINOGLYCAN USING MOLECULAR DYNAMICS DATA”

Table A.1. Comparison to Observed Literature Values of Glycosidic Linkage Dihedrals (ϕ, ψ) in Nonsulfated Chondroitin

		GlcAβ1-3GalNAc				GalNAcβ1-4GlcA			
	Min	ϕ	Diff	ψ	Diff [1]	ϕ	Diff	ψ	Diff [2]
Biased MD-generated 2-mer Ensembles [212] [3]	I	-83.75°	+2.5°	83.75°	+122.5°	-63.75°	-2.5°	-121.25°	-122.5°
	II					-58.75°	0°	93.75°	-127.5°
	II′					-83.75°	-2.5°	48.75°	-122.5°
Unbiased MD-Generated and NMR-Validated 6-mer Ensembles [259] [4]	I	-72°	-9°	108°	+98°	-73°	+7°	-117°	-127°

Diff = (x_{obs} - x) +/- 360° where x_{obs} = ϕ or ψ observed in our 20-mer simulations. [1] Expected difference for β1-3 = +120°. [2] Expected difference for β1-4 = -120°. [3] ϕ, ψ dihedral angles were sorted into 2.5° bins. [4] ϕ, ψ dihedral angles were sorted into 1° bins; only global minima (I) were reported.

Table A.2. Bond Energies (E_b) of Constructed Chondroitin 20-mer Conformations with Pierced Rings.

Piercing Bond	Pierced Ring	Pierce Type [1]	E_b (kcal/mol)	$\Delta E_b = E_b - E_{b,cut}$ [2] (kcal/mol)	Estimated ΔE_b [3] (kcal/mol)
GalNAc C_5-C_6	GalNAc	Exocyclic	721.0	592.2	285.2
GalNAc C_6-O_6	GalNAc	Exocyclic	784.2	655.4	132.3
GalNAc C_6-O_6	GalNAc	Exocyclic	793.0	664.2	132.3
GalNAc C-CT	GlcA	Exocyclic	787.7	658.9	253.4
GlcA C_5-C_6	GlcA	Exocyclic	720.2	591.4	294.6
GalNAc C_3-O_3	GlcA	Linkage	715.8	587.0	224.1
GlcA C_4-O_4	GalNAc	Linkage	714.8	586.0	238.6
GlcA	GalNAc	Interlocking	981.2	852.4	725.2
GlcA	GalNAc	Interlocking	941.6	812.8	725.2
GlcA	GlcA	Interlocking	961.4	832.6	723.7
GalNAc	GalNAc	Interlocking	1053	924.2	520.3
GalNAc	GalNAc	Interlocking	1110	981.2	520.3

[1] Exocyclic: the piercing bond is an exocyclic bond not participating in a glycosidic linkage; Linkage: the piercing bond is part of a glycosidic linkage; Interlocking: the rings are interlocking (i.e. there are two

piercing bonds and the estimated ΔE_b is the sum of the estimated ΔE_b of these two bonds). [2] $E_{b,cut}$ = 128.85 kcal/mol (bond energy cutoff). [3] Bond strain energy of this piercing bond estimated by comparing this bond energy in individual nonbonded saccharide systems with and without ring piercing.

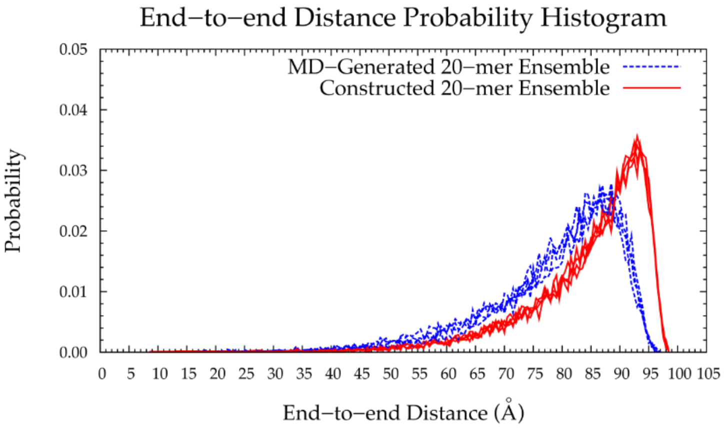


Figure A.1. End-to-end distance probability distribution of 20-mer ensembles generated by MD (blue dashed lines) and an early version of the construction algorithm (red solid lines) which applied glycosidic linkage geometries from MD-generated 20-mer ensembles and standard force field geometries for all monosaccharide rings; each type of ensemble includes four sets of 10,000 conformations; probabilities were calculated for end-to-end distances sorted into 0.5 Å bins.

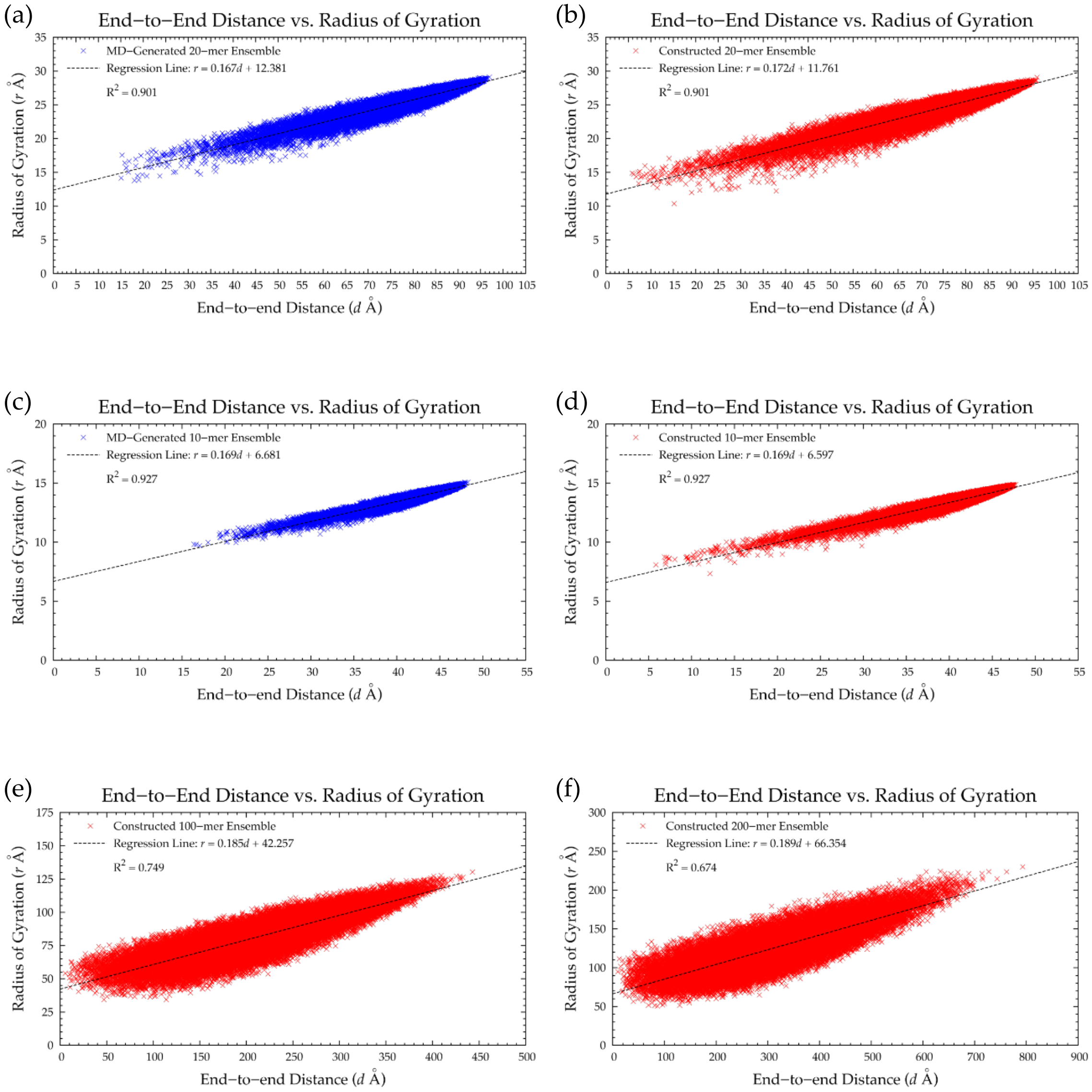
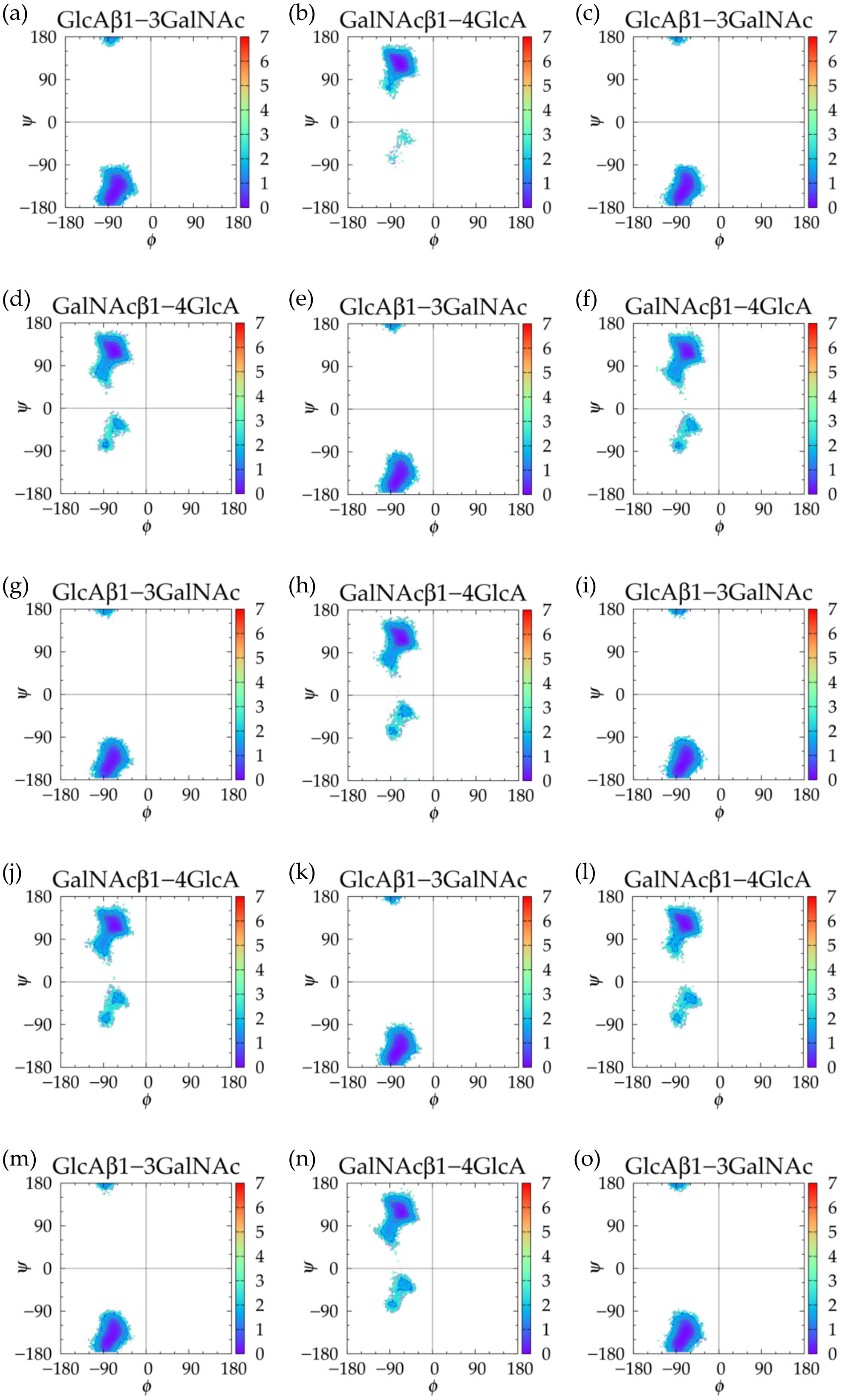


Figure A.2. Scatterplots of radius of gyration as a function of end-to-end distance in MD-generated and constructed ensembles of nonsulfated chondroitin (a,b) 20-mer and (c,d) 10-mer, respectively, and constructed ensembles of the chondroitin (e) 100-mer and (f) 200-mer. Each plot has 40,000 samples and shows linear regression and R^2.


(a) GlcAβ1−3GalNAc
(b) GalNAcβ1−4GlcA
(c) GlcAβ1−3GalNAc
(d) GalNAcβ1−4GlcA
(e) GlcAβ1−3GalNAc
(f) GalNAcβ1−4GlcA
(g) GlcAβ1−3GalNAc
(h) GalNAcβ1−4GlcA
(i) GlcAβ1−3GalNAc
(j) GalNAcβ1−4GlcA
(k) GlcAβ1−3GalNAc
(l) GalNAcβ1−4GlcA
(m) GlcAβ1−3GalNAc
(n) GalNAcβ1−4GlcA
(o) GlcAβ1−3GalNAc
ψ
180
90
0
−90
−180
ϕ
−180
−90
0
90
180
0
1
2
3
4
5
6
7

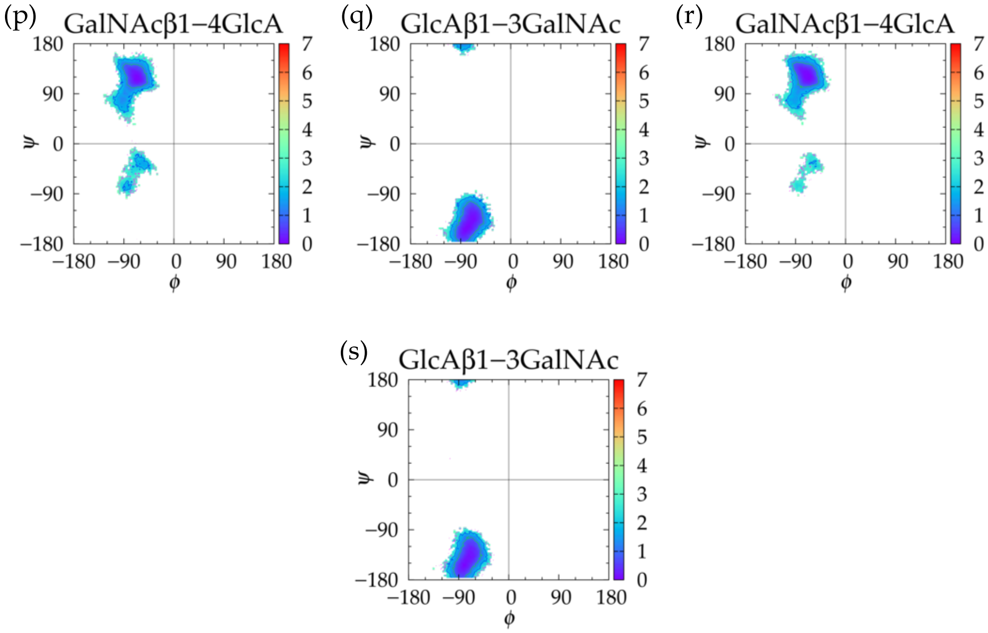
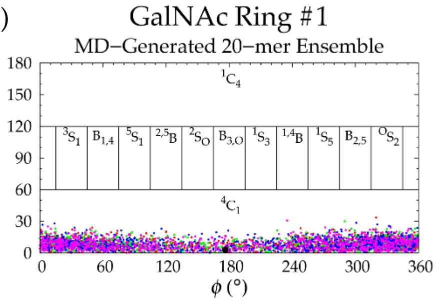
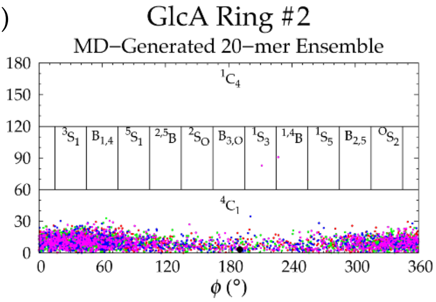


Figure A.3. $\Delta G(\phi, \psi)$ plots for each glycosidic linkage in the chondroitin 20-mer from MD-generated ensembles; (a) GlcA2←GalNAc1, (b) GalNAc3←GlcA2, (c) GlcA4←GalNAc3, (d) GalNAc5←GlcA4, (e) GlcA6←GalNAc5, (f) GalNAc7←GlcA6, (g) GlcA8←GalNAc7, (h) GalNAc9←GlcA8, (i) GlcA10←GalNAc9, (j) GalNAc11←GlcA10, (k) GlcA12←GalNAc11, (l) GalNAc13←GlcA12, (m) GlcA14←GalNAc13, (n) GalNAc15←GlcA14, (o) GlcA16←GalNAc15, (p) GalNAc17←GlcA16, (q) GlcA18←GalNAc17, (r) GalNAc19←GlcA18, and (s) GlcA20←GalNAc19; monosaccharides are numbered from reducing to nonreducing end; ϕ, ψ separated into 2.5° bins.

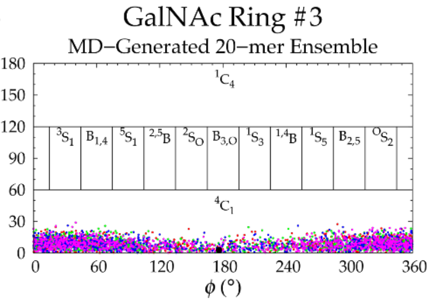
(a)


GalNAc Ring #1
MD−Generated 20−mer Ensemble
$^1C_4$
$^3S_1$ $B_{1,4}$ $^5S_1$ $^{2,5}B$ ^2S_O $B_{3,O}$ $^1S_3$ $^{1,4}B$ $^1S_5$ $B_{2,5}$ $^OS_2$
$^4C_1$
θ (°)
0 30 60 90 120 150 180
ϕ (°)
0 60 120 180 240 300 360

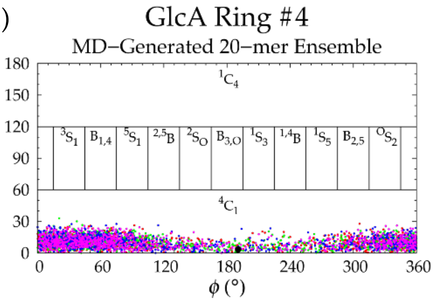
(b)


GlcA Ring #2
MD−Generated 20−mer Ensemble
$^1C_4$
$^3S_1$ $B_{1,4}$ $^5S_1$ $^{2,5}B$ ^2S_O $B_{3,O}$ $^1S_3$ $^{1,4}B$ $^1S_5$ $B_{2,5}$ $^OS_2$
$^4C_1$
θ (°)
0 30 60 90 120 150 180
ϕ (°)
0 60 120 180 240 300 360

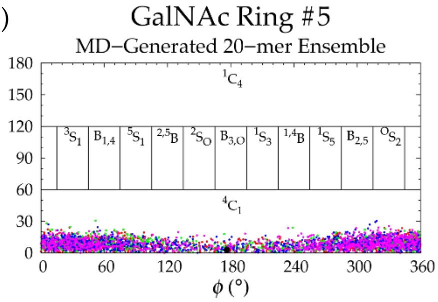
(c)


GalNAc Ring #3
MD−Generated 20−mer Ensemble
$^1C_4$
$^3S_1$ $B_{1,4}$ $^5S_1$ $^{2,5}B$ ^2S_O $B_{3,O}$ $^1S_3$ $^{1,4}B$ $^1S_5$ $B_{2,5}$ $^OS_2$
$^4C_1$
θ (°)
0 30 60 90 120 150 180
ϕ (°)
0 60 120 180 240 300 360

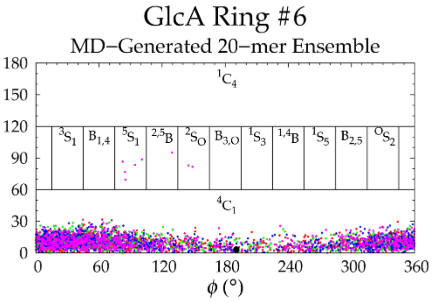
(d)


GlcA Ring #4
MD−Generated 20−mer Ensemble
$^1C_4$
$^3S_1$ $B_{1,4}$ $^5S_1$ $^{2,5}B$ ^2S_O $B_{3,O}$ $^1S_3$ $^{1,4}B$ $^1S_5$ $B_{2,5}$ $^OS_2$
$^4C_1$
θ (°)
0 30 60 90 120 150 180
ϕ (°)
0 60 120 180 240 300 360

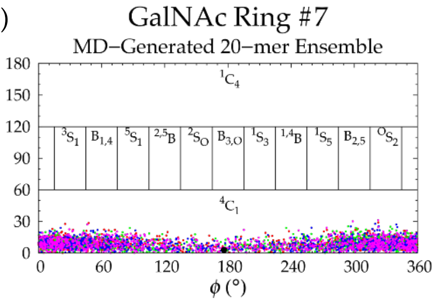
(e)


GalNAc Ring #5
MD−Generated 20−mer Ensemble
$^1C_4$
$^3S_1$ $B_{1,4}$ $^5S_1$ $^{2,5}B$ ^2S_O $B_{3,O}$ $^1S_3$ $^{1,4}B$ $^1S_5$ $B_{2,5}$ $^OS_2$
$^4C_1$
θ (°)
0 30 60 90 120 150 180
ϕ (°)
0 60 120 180 240 300 360

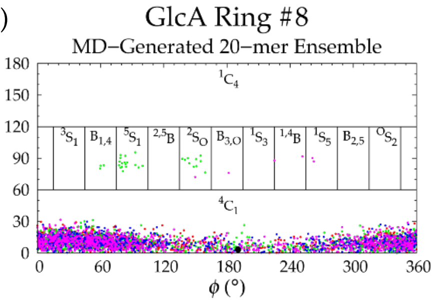
(f)


GlcA Ring #6
MD−Generated 20−mer Ensemble
$^1C_4$
$^3S_1$ $B_{1,4}$ $^5S_1$ $^{2,5}B$ ^2S_O $B_{3,O}$ $^1S_3$ $^{1,4}B$ $^1S_5$ $B_{2,5}$ $^OS_2$
$^4C_1$
θ (°)
0 30 60 90 120 150 180
ϕ (°)
0 60 120 180 240 300 360

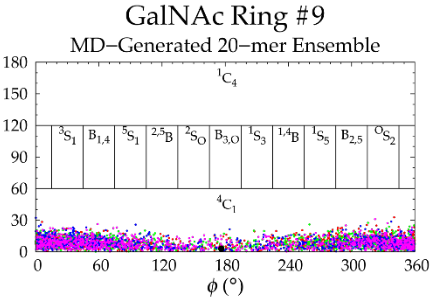
(g)


GalNAc Ring #7
MD−Generated 20−mer Ensemble
$^1C_4$
$^3S_1$ $B_{1,4}$ $^5S_1$ $^{2,5}B$ ^2S_O $B_{3,O}$ $^1S_3$ $^{1,4}B$ $^1S_5$ $B_{2,5}$ $^OS_2$
$^4C_1$
θ (°)
0 30 60 90 120 150 180
ϕ (°)
0 60 120 180 240 300 360

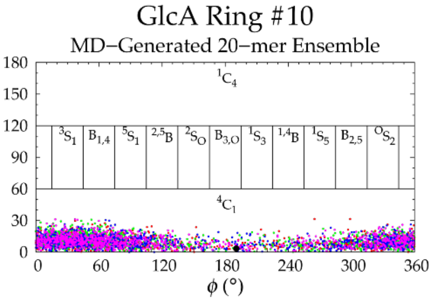
(h)


GlcA Ring #8
MD−Generated 20−mer Ensemble
$^1C_4$
$^3S_1$ $B_{1,4}$ $^5S_1$ $^{2,5}B$ ^2S_O $B_{3,O}$ $^1S_3$ $^{1,4}B$ $^1S_5$ $B_{2,5}$ $^OS_2$
$^4C_1$
θ (°)
0 30 60 90 120 150 180
ϕ (°)
0 60 120 180 240 300 360

(i)

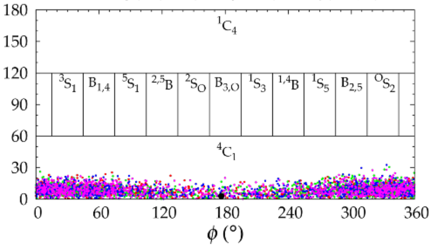

GalNAc Ring #9
MD−Generated 20−mer Ensemble
$^1C_4$
$^3S_1$ $B_{1,4}$ $^5S_1$ $^{2,5}B$ ^2S_O $B_{3,O}$ $^1S_3$ $^{1,4}B$ $^1S_5$ $B_{2,5}$ $^OS_2$
$^4C_1$
θ (°)
0 30 60 90 120 150 180
ϕ (°)
0 60 120 180 240 300 360

(j)


GlcA Ring #10
MD−Generated 20−mer Ensemble
$^1C_4$
$^3S_1$ $B_{1,4}$ $^5S_1$ $^{2,5}B$ ^2S_O $B_{3,O}$ $^1S_3$ $^{1,4}B$ $^1S_5$ $B_{2,5}$ $^OS_2$
$^4C_1$
θ (°)
0 30 60 90 120 150 180
ϕ (°)
0 60 120 180 240 300 360

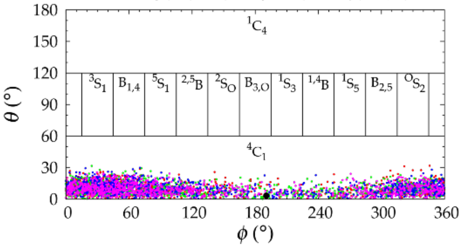
(k) GalNAc Ring #11

MD–Generated 20–mer Ensemble



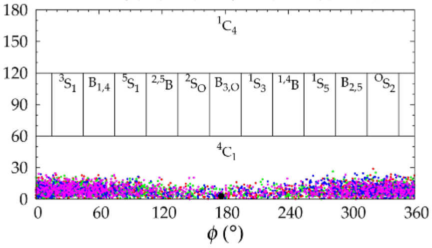
(l) GlcA Ring #12

MD–Generated 20–mer Ensemble



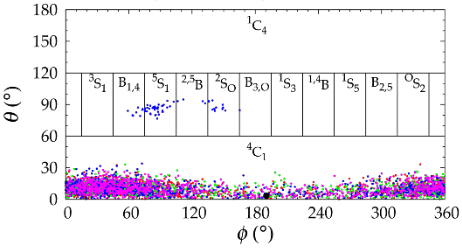
(m) GalNAc Ring #13

MD–Generated 20–mer Ensemble



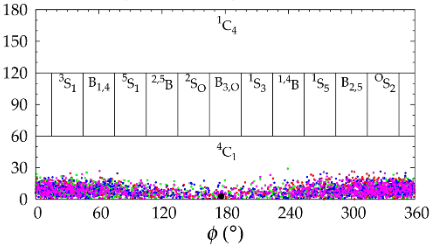
(n) GlcA Ring #14

MD–Generated 20–mer Ensemble



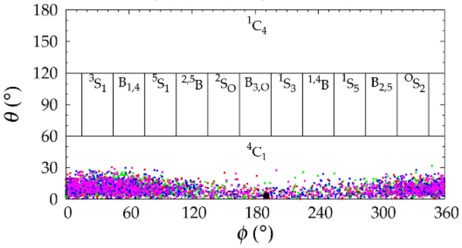
(o) GalNAc Ring #15

MD–Generated 20–mer Ensemble



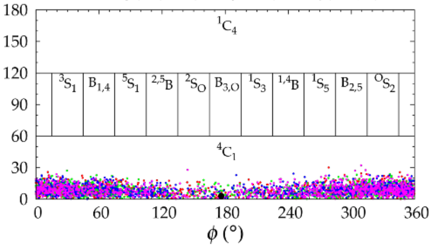
(p) GlcA Ring #16

MD–Generated 20–mer Ensemble



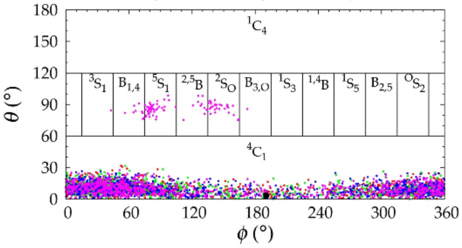
(q) GalNAc Ring #17

MD–Generated 20–mer Ensemble



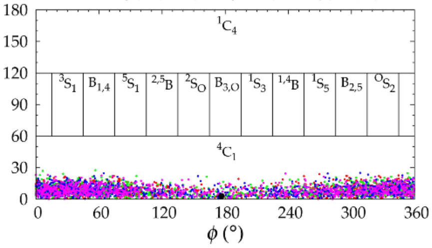
(r) GlcA Ring #18

MD–Generated 20–mer Ensemble



(s) GalNAc Ring #19

MD–Generated 20–mer Ensemble



(t) GlcA Ring #20

MD–Generated 20–mer Ensemble

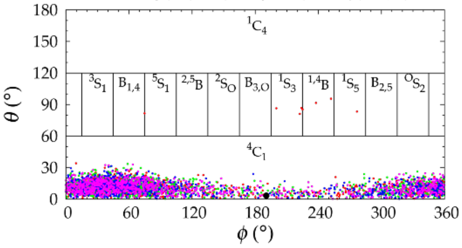


Figure A.4. (a-t) C-P plots for each monosaccharide ring in the chondroitin 20-mer from MD-generated ensembles; monosaccharides are numbered from reducing to nonreducing end; each of the 4 runs is represented by different color.

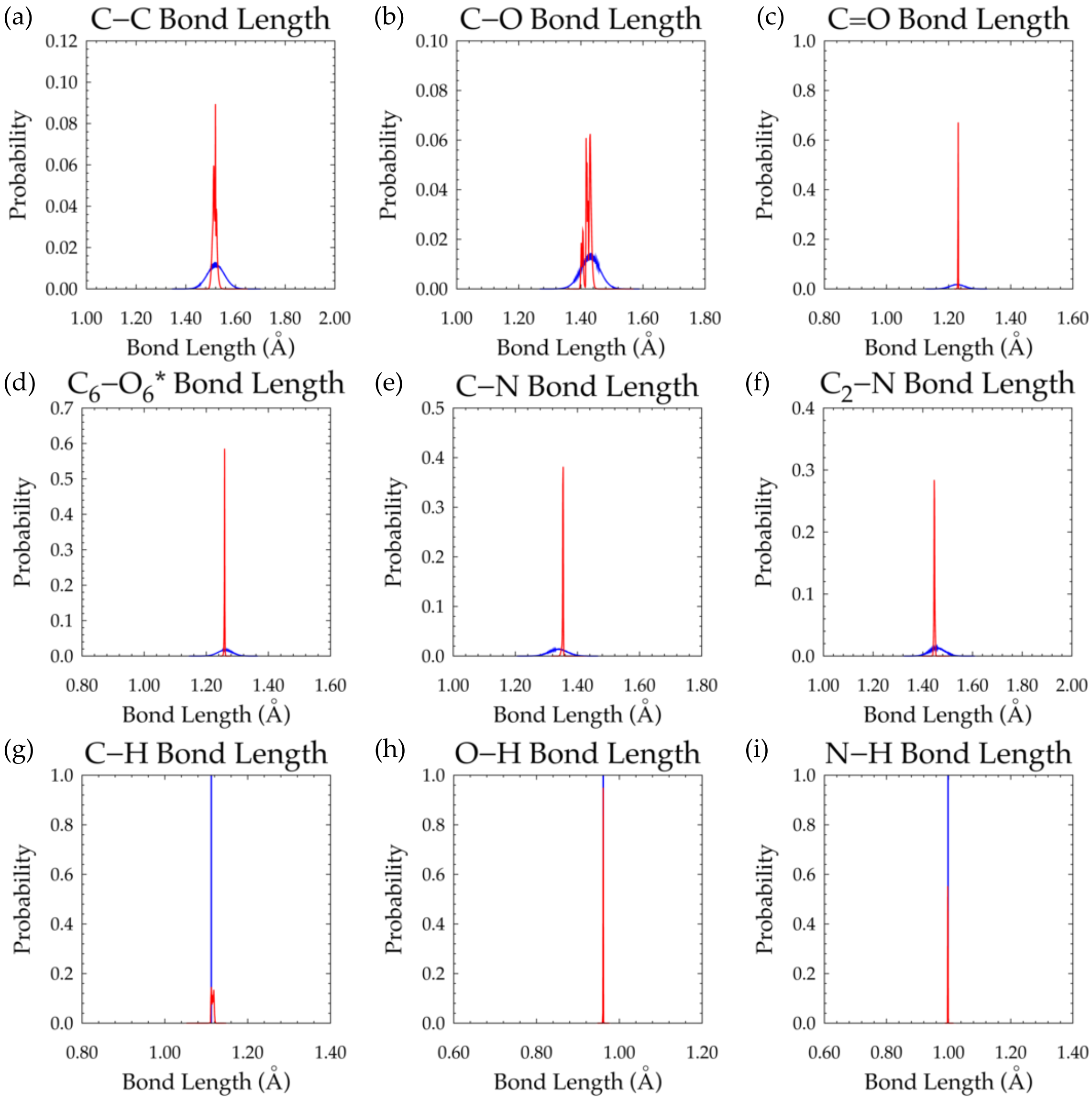


Figure A.5. (a-i) Probability histograms of bond lengths for each type of bond in the chondroitin 20-mer; blue lines represent bond lengths in MD-generated ensembles (which match those in constructed ensembles before minimization) and red lines represent bond lengths in constructed 20-mer ensembles after minimization. (Note: bond lengths involving hydrogen atoms (g-i) are fixed during MD but not during minimization in the algorithm.)

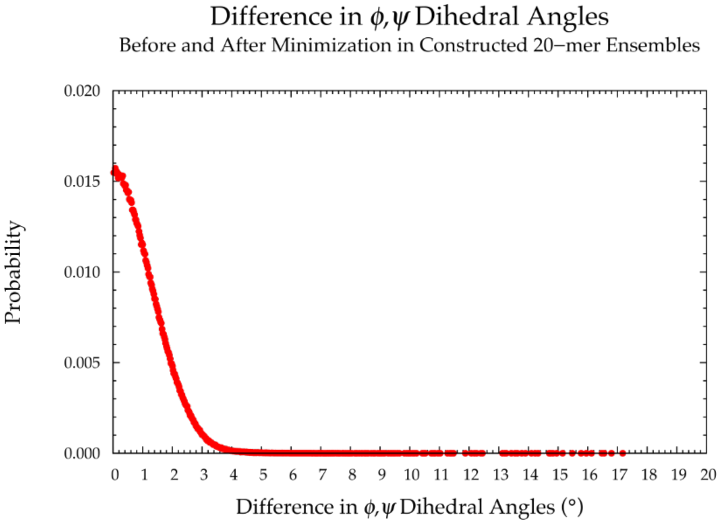


Figure A.6. Probability histogram showing changes in glycosidic linkage ϕ and ψ dihedral angles during energy minimization in constructed 20-mer ensembles; 99.6% of all differences are within 4°.

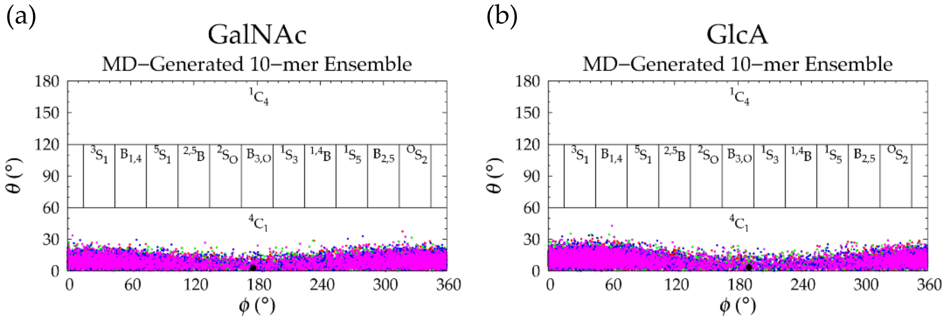


Figure A.7. C-P plots of (a) GalNAc and (b) GlcA in MD-generated chondroitin 10-mer ensembles; geometries from the four sets of each type of ensemble are represented by red, green, blue, and magenta dots, respectively and the force field geometry is represented by a black dot. C-P parameters of all rings in every tenth snapshot from each ensemble were plotted (i.e. 5 rings * 1,000 snapshots per run * 4 runs = 20,000 parameter sets).

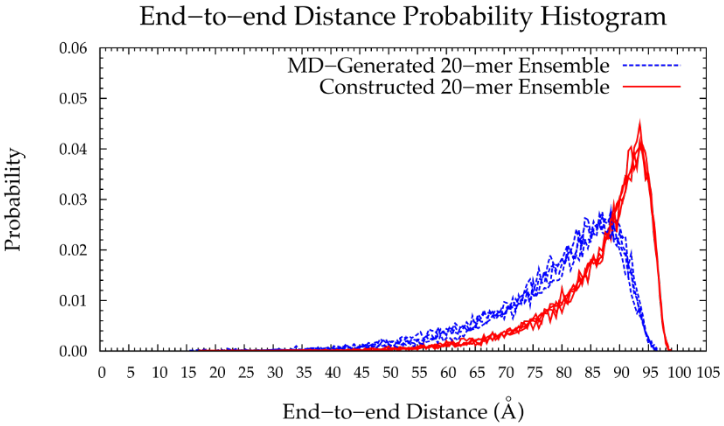


Figure A.8. End-to-end distance probability distribution of chondroitin 20-mer ensembles generated by MD (blue dashed lines) and an early version of the construction algorithm (red solid lines) which applied glycosidic linkage geometries from Adaptive Biasing Force (ABF) [363,364] MD-generated ensembles of nonsulfated chondroitin disaccharides [212] and standard force field geometries for all monosaccharide rings; each type of ensemble includes four sets of 10,000 conformations; probabilities were calculated for end-to-end distances sorted into 0.5 Å bins.

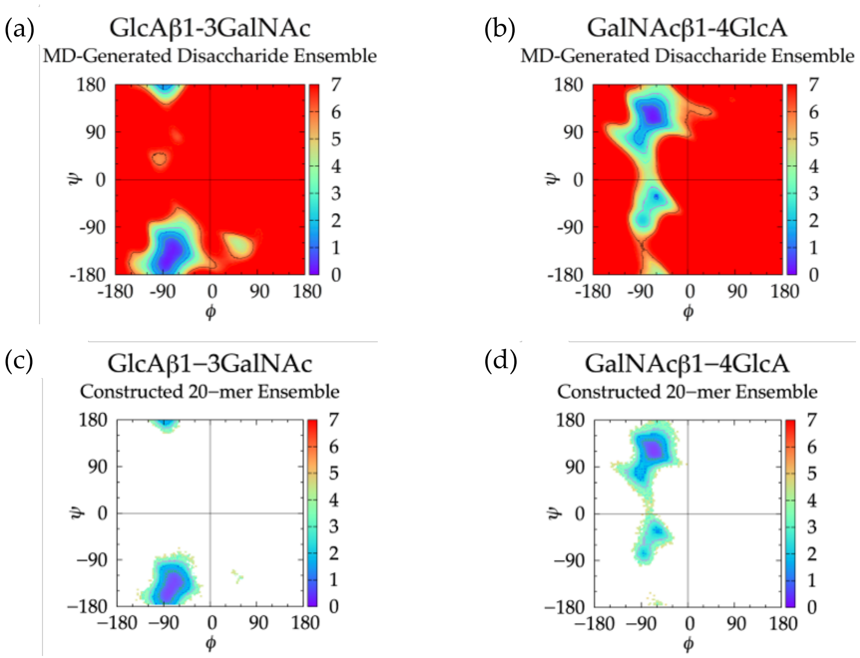


Figure A.9. Δ*G*(ϕ, ψ) plots for GlcAβ1-3GalNAc and GalNAcβ1-4GlcA glycosidic linkages in (a,b) nonsulfated chondroitin disaccharides simulated using MD with ABF on glycosidic linkage dihedrals (ϕ, ψ separated into 1° bins) and (c,d) 20-mer ensembles constructed using glycosidic linkage dihedral probabilities from ABF MD-generated disaccharides and standard force-field monosaccharide ring geometries (ϕ, ψ separated into 2.5° bins); contour lines every 1 kcal/mol; in ABF MD simulations, all values of ϕ, ψ were sampled but contours for data with $\Delta G(\phi, \psi) > 7$ kcal/mol are given values of 7 kcal/mol in plots (a,b) for clarity (shown in red).

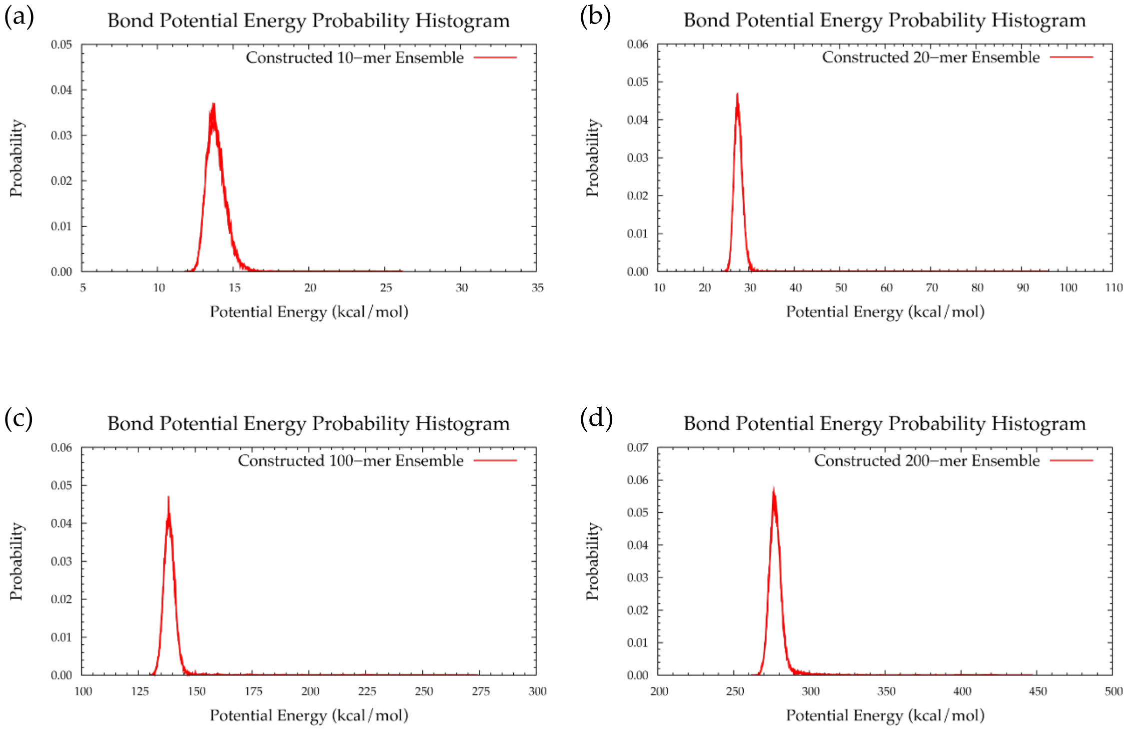


Figure A.10. Bond energy distribution probability histograms from constructed ensembles of the (a) 10-mer (cutoff = 115.49 kcal/mol), (b) 20-mer (cutoff = 128.85 kcal/mol), (c) 100-mer (cutoff = 274.89 kcal/mol), and (d) 200-mer (cutoff = 449.26 kcal/mol).

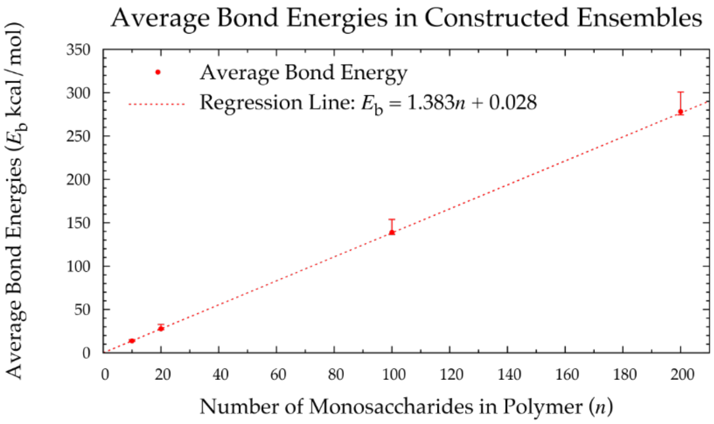


Figure A.11. Average bond energies, standard deviations (calculated by fitting energies to gaussian curve), and regression line as a function of polymer length. Of note, (1) the regression equation calculated using only 10-mer and 20-mer constructed data: $E_b = 1.383n + 0.028$ (regression line plotted) closely matches (2) the regression equation calculated using 10-, 20-, 100-, and 200-mer constructed data: $E_b = 1.391n - 0.139$. Average bond energies of the 100-mer and 200-mer predicted by regression equation (1), $E_{b,predicted}(100) = 138.3$ kcal/mol and $E_{b,predicted}(200) = 276.6$ kcal/mol, are within 2 kcal/mol of the true averages, $E_b(100) = 138.8$ kcal/mol and $E_b(200) = 278.2$ kcal/mol.

APPENDIX B: SUPPORTING INFORMATION FOR "CONSTRUCTING 3-DIMENSIONAL ATOMIC-RESOLUTION MODELS OF NONSULFATED GLYCOSAMINOGLYCANS WITH ARBITRARY LENGTHS USING CONFORMATIONS FROM MOLECULAR DYNAMICS"

Table B.1. Most Probable End-to-End Distances (d) in MD-Generated 20-mer Conformations with Glycosidic Linkage Conformations in Secondary Basins [1].

	Hyaluronan		Nonsulfated Keratan	
	GlcAβ1-3GlcNAc $-\phi, +\psi$ d (Å)	GlcNAcβ1-4GlcA $-\phi, -\psi$ d (Å)	Galβ1-4GlcNAc $-\phi, -\psi$ d (Å)	GlcNAcβ1-3Gal $-\phi, +\psi$ d (Å)
Run 1	81.5	74.0	88.5	79.5
Run 2	73.0	73.5	85.0	84.0
Run 3	61.0	74.0	84.5	88.5
Run 4	77.0	76.5	84.5	73.5
All [2]	66.0	74.0	84.5	84.0

[1] Probabilities were calculated for end-to-end distances sorted into 0.5 Å bins. [2]All = end-to-end distance distribution aggregated across all four runs.

Table B.2. Percent of Occurrences of Different IdoA Ring Puckers in Nonsulfated Dermatan MD Simulations

	20-mer Ensemble	10-mer Ensemble
${}^{1}C_{4}$	68.727%	70.100%
${}^{4}C_{1}$	7.062%	9.025%
$B_{1,4}$	0.018%	0.005%
${}^{5}S_{1}$	0.740%	0.685%
${}^{2,5}B$	3.213%	2.610%
${}^{2}S_{O}$	19.702%	17.065%
$B_{3,O}$	0.080%	0.070%
${}^{1}S_{3}$	0.125%	0.125%
${}^{1,4}B$	0.178%	0.115%
${}^{1}S_{5}$	0.155%	0.200%

Table B.3. Percent of Occurrences of Different IdoA Ring Puckers in Nonsulfated Heparan MD Simulations

	20-mer Ensemble	10-mer Ensemble
$^{1}C_{4}$	54.222%	64.645%
$^{4}C_{1}$	4.410%	6.730%
$^{3}S_{1}$	0.003%	0.000%
$B_{1,4}$	0.085%	0.035%
$^{5}S_{1}$	4.120%	3.110%
$^{2,5}B$	12.510%	8.550%
$^{2}S_{O}$	24.505%	16.520%
$B_{3,O}$	0.008%	0.000%
$^{1}S_{3}$	0.035%	0.085%
$^{1,4}B$	0.047%	0.155%
$^{1}S_{5}$	0.055%	0.170%

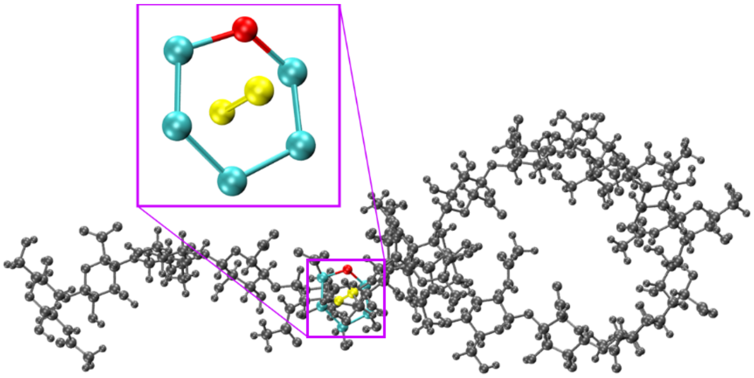


Figure B.1. Excluded constructed hyaluronan 20-mer conformation with GlcA 20 ring pierced by C_1-O_3 bond in GlcA 6 β1-3 GlcNAc 5 linkage post minimization (E_b = 690.5 kcal/mol; closeup shows atoms involved in the ring pierce); $E_{b,\ cutoff}$ = 128.5 kcal/mol.

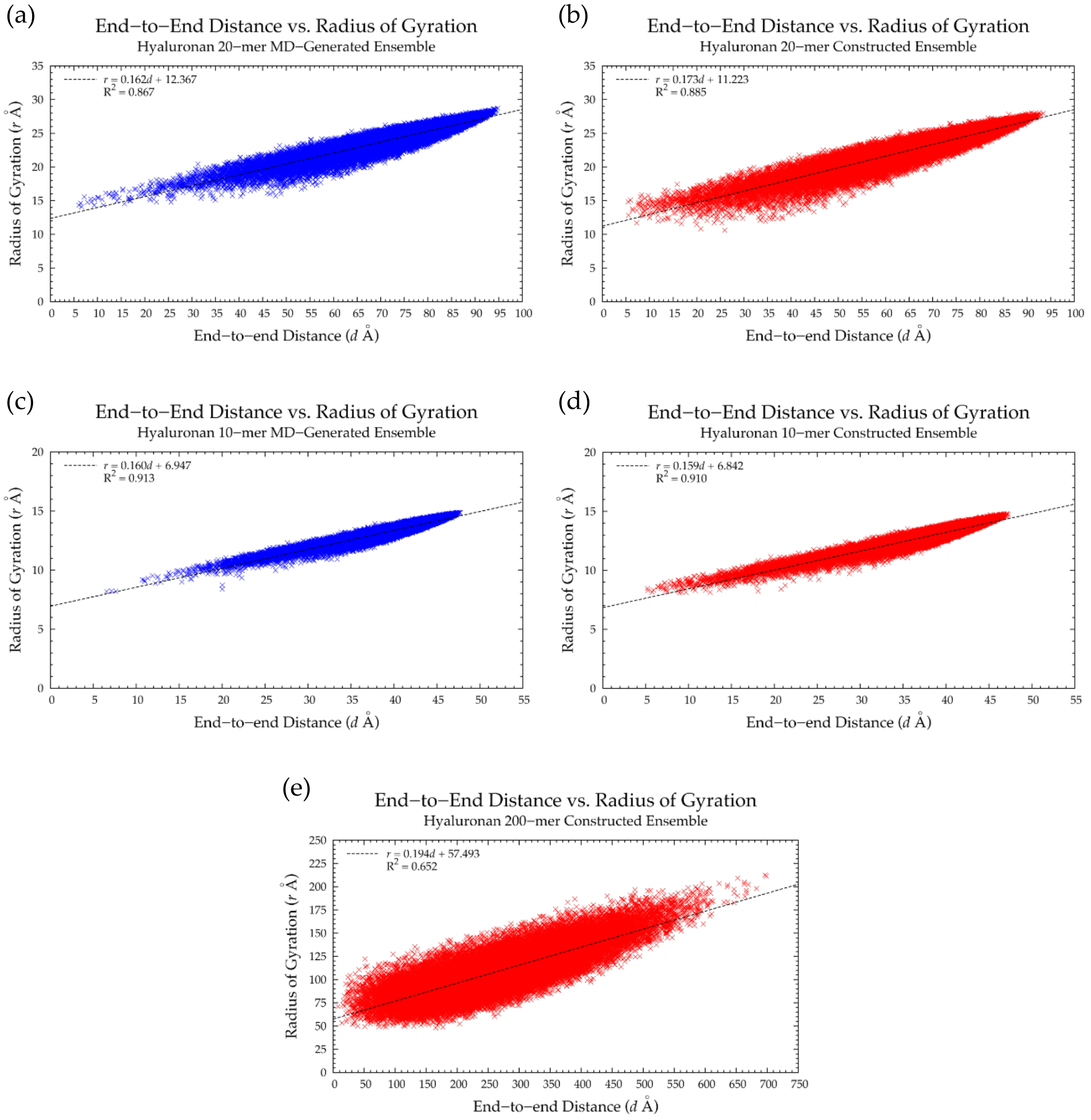


Figure B.2. Scatterplots of radius of gyration as a function of end-to-end distance in MD-generated and constructed ensembles of hyaluronan (a,b) 20-mer and (c,d) 10-mer, respectively, and (e) constructed ensemble of hyaluronan 200-mer; each plot has 40,000 samples and shows linear regression and R^2.

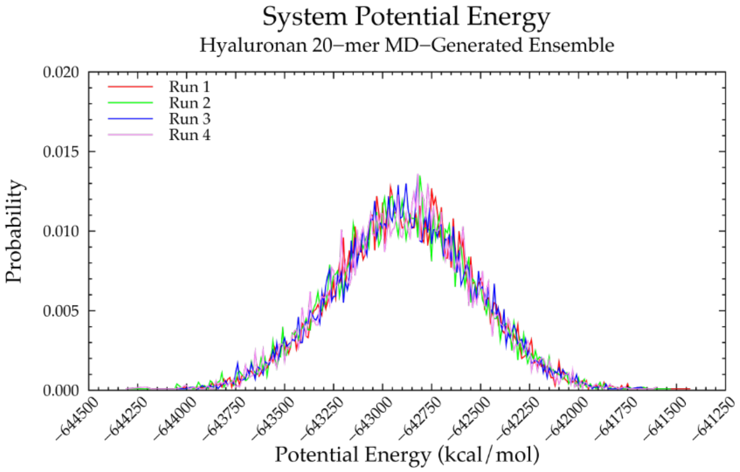
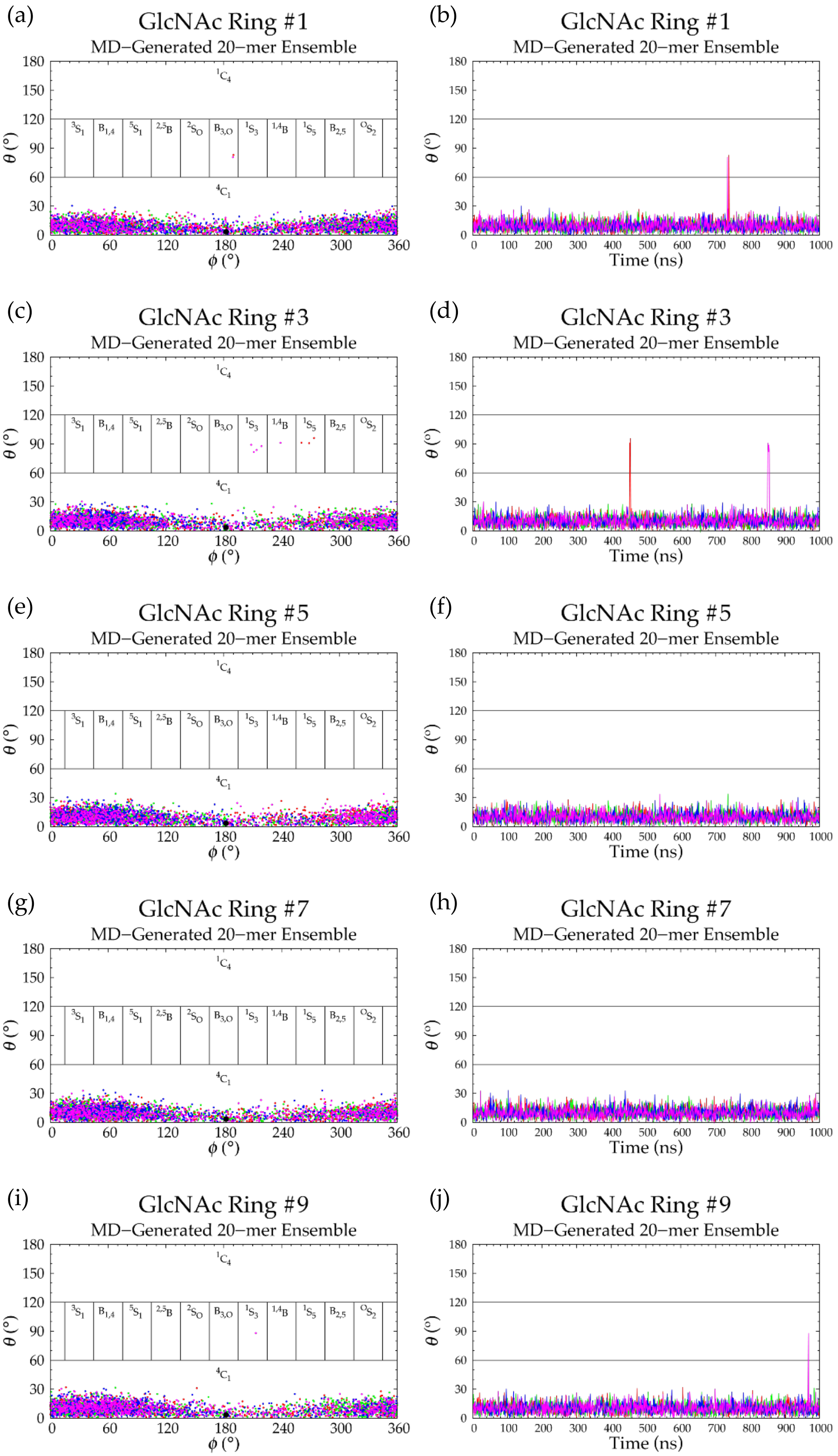
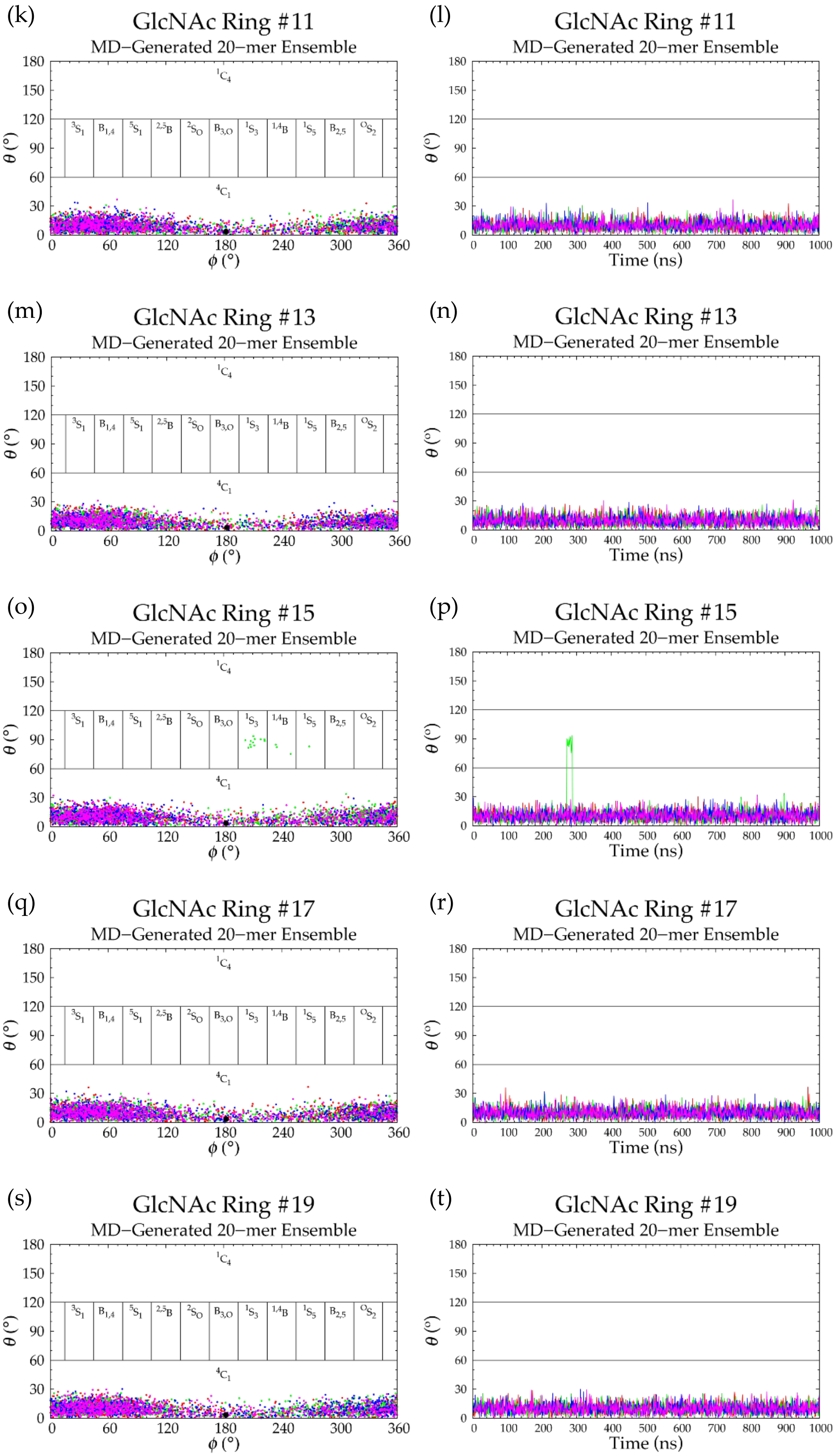


Figure B.3. System potential energy probability distribution of the MD-generated hyaluronan 20-mer ensemble; each MD run is represented by a different color.


(a)
GlcNAc Ring #1
MD−Generated 20−mer Ensemble
(b)
GlcNAc Ring #1
MD−Generated 20−mer Ensemble
(c)
GlcNAc Ring #3
MD−Generated 20−mer Ensemble
(d)
GlcNAc Ring #3
MD−Generated 20−mer Ensemble
(e)
GlcNAc Ring #5
MD−Generated 20−mer Ensemble
(f)
GlcNAc Ring #5
MD−Generated 20−mer Ensemble
(g)
GlcNAc Ring #7
MD−Generated 20−mer Ensemble
(h)
GlcNAc Ring #7
MD−Generated 20−mer Ensemble
(i)
GlcNAc Ring #9
MD−Generated 20−mer Ensemble
(j)
GlcNAc Ring #9
MD−Generated 20−mer Ensemble
θ (°)
0
30
60
90
120
150
180
ϕ (°)
0
60
120
180
240
300
360
Time (ns)
0
100
200
300
400
500
600
700
800
900
1000
$^1C_4$
$^4C_1$
$^3S_1$
$B_{1,4}$
$^5S_1$
$^{2,5}B$
^2S_O
$B_{3,O}$
$^1S_3$
$^{1,4}B$
$^1S_5$
$B_{2,5}$
$^OS_2$


(k) GlcNAc Ring #11 MD–Generated 20–mer Ensemble

(l) GlcNAc Ring #11 MD–Generated 20–mer Ensemble

(m) GlcNAc Ring #13 MD–Generated 20–mer Ensemble

(n) GlcNAc Ring #13 MD–Generated 20–mer Ensemble

(o) GlcNAc Ring #15 MD–Generated 20–mer Ensemble

(p) GlcNAc Ring #15 MD–Generated 20–mer Ensemble

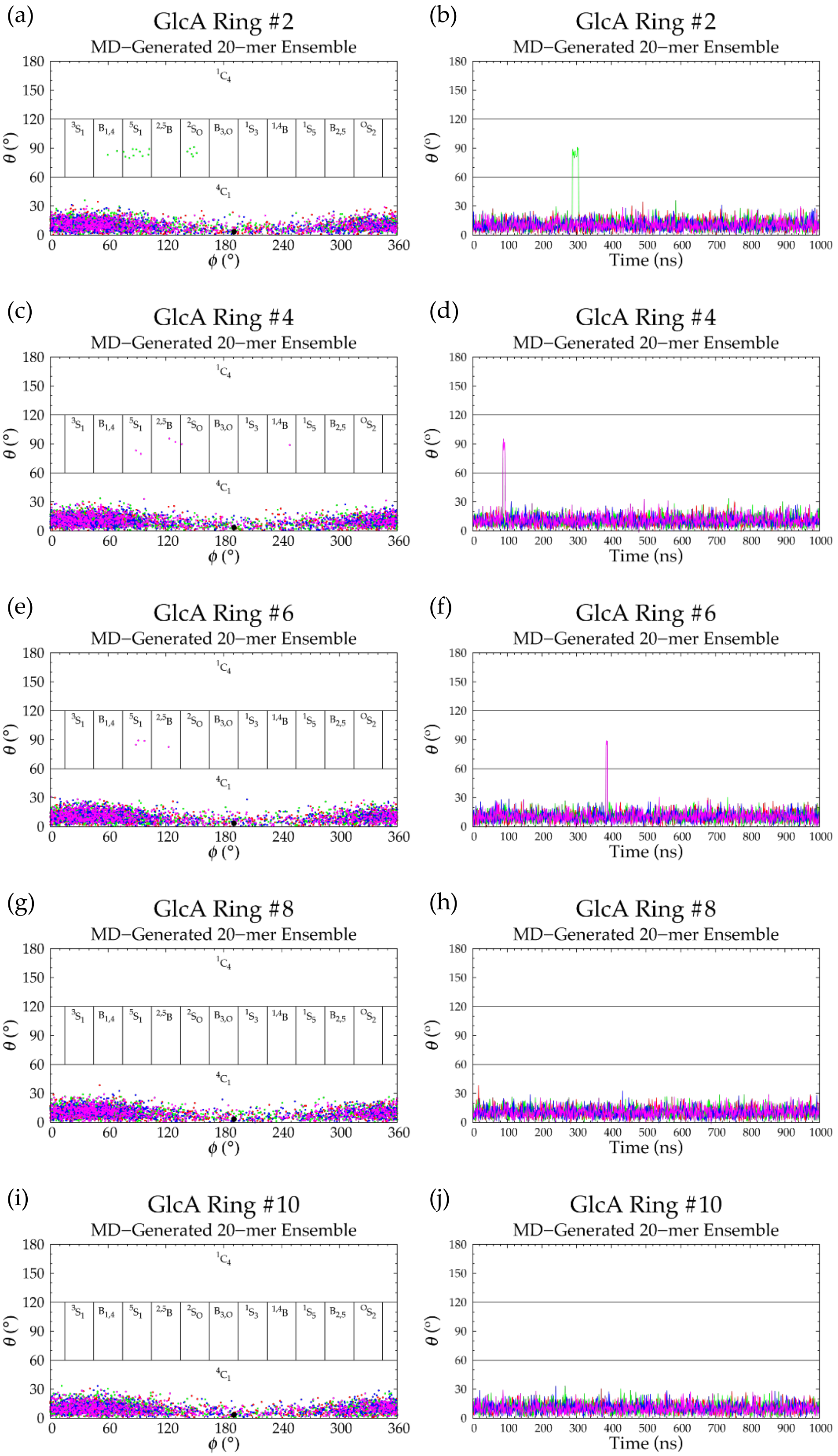
(q) GlcNAc Ring #17 MD–Generated 20–mer Ensemble

(r) GlcNAc Ring #17 MD–Generated 20–mer Ensemble

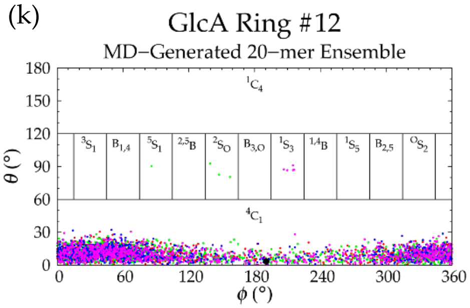
(s) GlcNAc Ring #19 MD–Generated 20–mer Ensemble

(t) GlcNAc Ring #19 MD–Generated 20–mer Ensemble

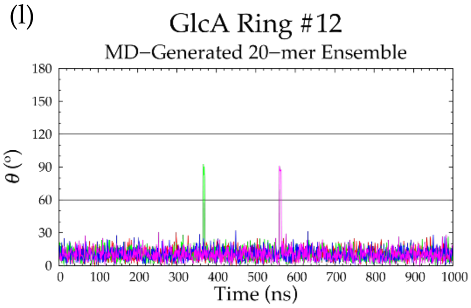
Figure B.4. (a,c,e,g,i,k,m,o,q,s) C-P plots and (b,d,f,h,j,l,n,p,r,t) C-P parameter θ timeseries for each GlcNAc monosaccharide ring in the MD-generated hyaluronan 20-mer ensemble; monosaccharides are numbered from reducing to nonreducing end; each of the 4 runs is represented by different color.


(a)
GlcA Ring #2
MD−Generated 20−mer Ensemble
(b)
GlcA Ring #2
MD−Generated 20−mer Ensemble
(c)
GlcA Ring #4
MD−Generated 20−mer Ensemble
(d)
GlcA Ring #4
MD−Generated 20−mer Ensemble
(e)
GlcA Ring #6
MD−Generated 20−mer Ensemble
(f)
GlcA Ring #6
MD−Generated 20−mer Ensemble
(g)
GlcA Ring #8
MD−Generated 20−mer Ensemble
(h)
GlcA Ring #8
MD−Generated 20−mer Ensemble
(i)
GlcA Ring #10
MD−Generated 20−mer Ensemble
(j)
GlcA Ring #10
MD−Generated 20−mer Ensemble
$^1C_4$
$^3S_1$
$B_{1,4}$
$^5S_1$
$^{2,5}B$
^2S_O
$B_{3,O}$
$^1S_3$
$^{1,4}B$
$^1S_5$
$B_{2,5}$
$^OS_2$
$^4C_1$
θ (°)
0
30
60
90
120
150
180
ϕ (°)
0
60
120
180
240
300
360
Time (ns)
0
100
200
300
400
500
600
700
800
900
1000

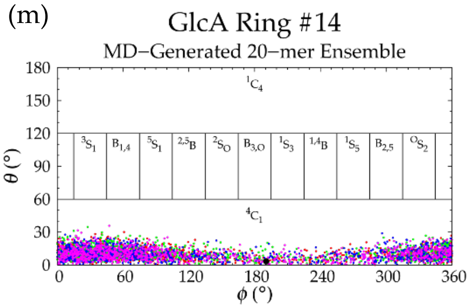
(k)


GlcA Ring #12
MD−Generated 20−mer Ensemble
$^{1}C_{4}$
$^{3}S_{1}$ $B_{1,4}$ $^{5}S_{1}$ $^{2,5}B$ $^{2}S_{O}$ $B_{3,O}$ $^{1}S_{3}$ $^{1,4}B$ $^{1}S_{5}$ $B_{2,5}$ $^{O}S_{2}$
$^{4}C_{1}$
θ (°)
0 30 60 90 120 150 180
0 60 120 180 240 300 360
ϕ (°)

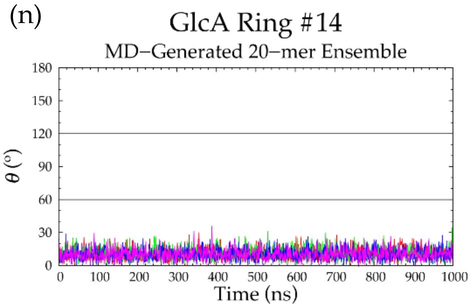
(l)


GlcA Ring #12
MD−Generated 20−mer Ensemble
θ (°)
0 30 60 90 120 150 180
0 100 200 300 400 500 600 700 800 900 1000
Time (ns)

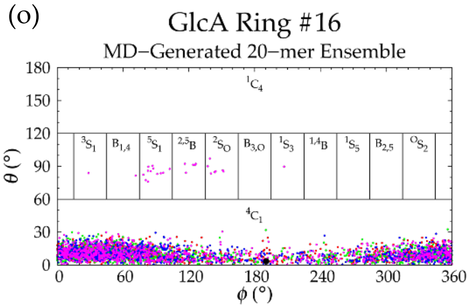
(m)


GlcA Ring #14
MD−Generated 20−mer Ensemble
$^{1}C_{4}$
$^{3}S_{1}$ $B_{1,4}$ $^{5}S_{1}$ $^{2,5}B$ $^{2}S_{O}$ $B_{3,O}$ $^{1}S_{3}$ $^{1,4}B$ $^{1}S_{5}$ $B_{2,5}$ $^{O}S_{2}$
$^{4}C_{1}$
θ (°)
0 30 60 90 120 150 180
0 60 120 180 240 300 360
ϕ (°)

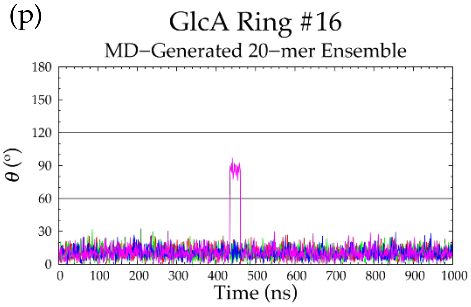
(n)


GlcA Ring #14
MD−Generated 20−mer Ensemble
θ (°)
0 30 60 90 120 150 180
0 100 200 300 400 500 600 700 800 900 1000
Time (ns)

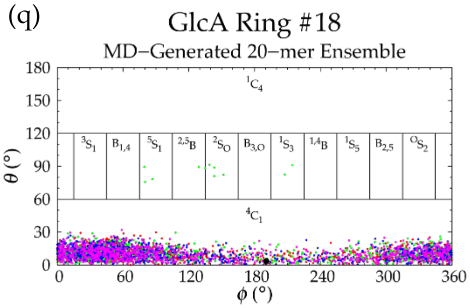
(o)


GlcA Ring #16
MD−Generated 20−mer Ensemble
$^{1}C_{4}$
$^{3}S_{1}$ $B_{1,4}$ $^{5}S_{1}$ $^{2,5}B$ $^{2}S_{O}$ $B_{3,O}$ $^{1}S_{3}$ $^{1,4}B$ $^{1}S_{5}$ $B_{2,5}$ $^{O}S_{2}$
$^{4}C_{1}$
θ (°)
0 30 60 90 120 150 180
0 60 120 180 240 300 360
ϕ (°)

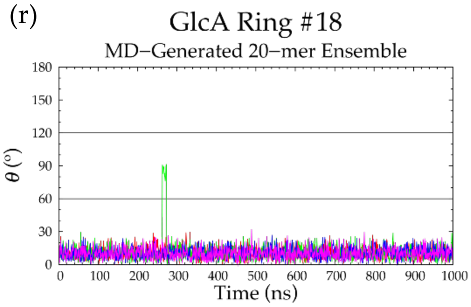
(p)


GlcA Ring #16
MD−Generated 20−mer Ensemble
θ (°)
0 30 60 90 120 150 180
0 100 200 300 400 500 600 700 800 900 1000
Time (ns)

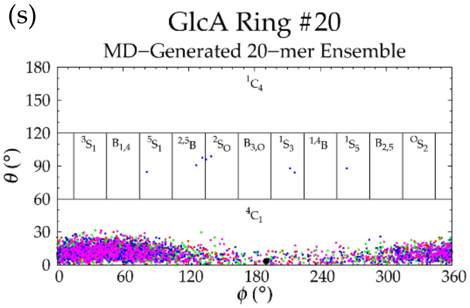
(q)


GlcA Ring #18
MD−Generated 20−mer Ensemble
$^{1}C_{4}$
$^{3}S_{1}$ $B_{1,4}$ $^{5}S_{1}$ $^{2,5}B$ $^{2}S_{O}$ $B_{3,O}$ $^{1}S_{3}$ $^{1,4}B$ $^{1}S_{5}$ $B_{2,5}$ $^{O}S_{2}$
$^{4}C_{1}$
θ (°)
0 30 60 90 120 150 180
0 60 120 180 240 300 360
ϕ (°)

(r)


GlcA Ring #18
MD−Generated 20−mer Ensemble
θ (°)
0 30 60 90 120 150 180
0 100 200 300 400 500 600 700 800 900 1000
Time (ns)

(s)


GlcA Ring #20
MD−Generated 20−mer Ensemble
$^{1}C_{4}$
$^{3}S_{1}$ $B_{1,4}$ $^{5}S_{1}$ $^{2,5}B$ $^{2}S_{O}$ $B_{3,O}$ $^{1}S_{3}$ $^{1,4}B$ $^{1}S_{5}$ $B_{2,5}$ $^{O}S_{2}$
$^{4}C_{1}$
θ (°)
0 30 60 90 120 150 180
0 60 120 180 240 300 360
ϕ (°)

(t)

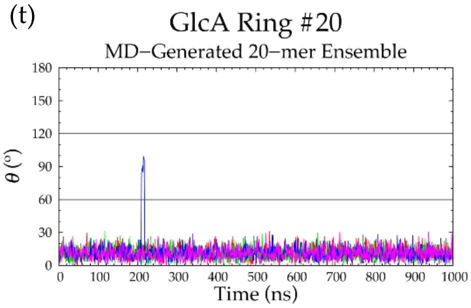

GlcA Ring #20
MD−Generated 20−mer Ensemble
θ (°)
0 30 60 90 120 150 180
0 100 200 300 400 500 600 700 800 900 1000
Time (ns)

Figure B.5. (a,c,e,g,i,k,m,o,q,s) C-P plots and (b,d,f,h,j,l,n,p,r,t) C-P parameter θ timeseries for each GlcA monosaccharide ring in the MD-generated hyaluronan 20-mer ensemble; monosaccharides are numbered from reducing to nonreducing end; each of the 4 runs is represented by different color.

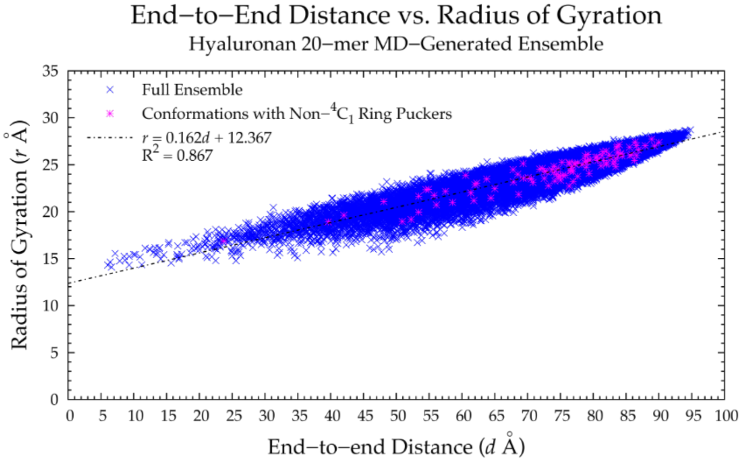
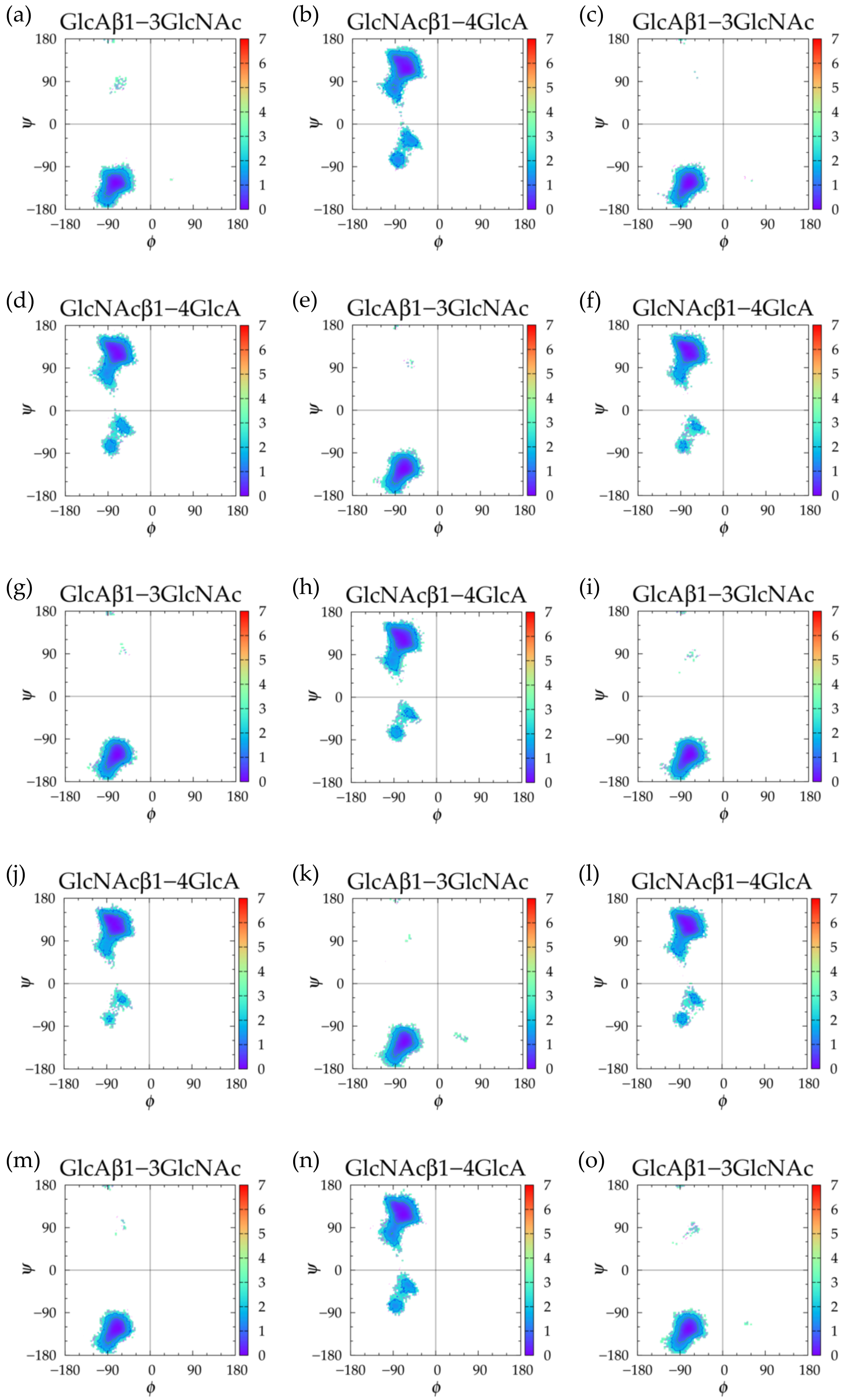


Figure B.6. Scatterplot of radius of gyration as a function of end-to-end distance in MD-generated hyaluronan 20-mer conformations with non-$^4C_1$ ring puckers (pink) overlaid on data for full MD-generated hyaluronan 20-mer ensemble (blue) and corresponding linear regression (Figure B.2.a).


(a) GlcAβ1−3GlcNAc
(b) GlcNAcβ1−4GlcA
(c) GlcAβ1−3GlcNAc
(d) GlcNAcβ1−4GlcA
(e) GlcAβ1−3GlcNAc
(f) GlcNAcβ1−4GlcA
(g) GlcAβ1−3GlcNAc
(h) GlcNAcβ1−4GlcA
(i) GlcAβ1−3GlcNAc
(j) GlcNAcβ1−4GlcA
(k) GlcAβ1−3GlcNAc
(l) GlcNAcβ1−4GlcA
(m) GlcAβ1−3GlcNAc
(n) GlcNAcβ1−4GlcA
(o) GlcAβ1−3GlcNAc
ψ
ϕ
−180
−90
0
90
180
0
1
2
3
4
5
6
7

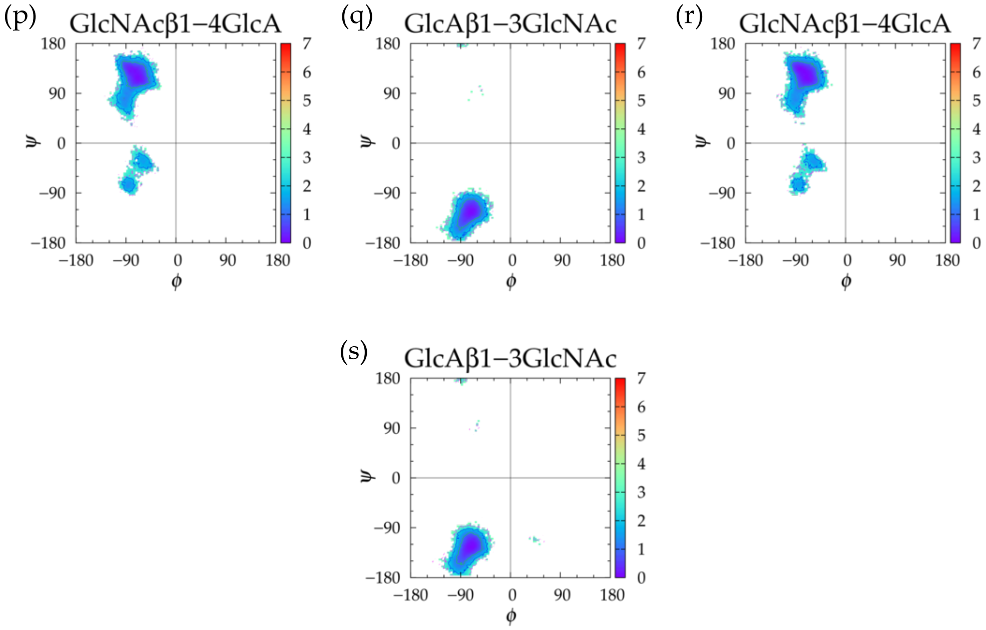


Figure B.7. $\Delta G(\phi, \psi)$ plots for each glycosidic linkage in the MD-generated hyaluronan 20-mer ensemble; (a) GlcA2→GlcNAc1, (b) GlcNAc3→GlcA2, (c) GlcA4→GlcNAc3, (d) GlcNAc5→GlcA4, (e) GlcA6→GlcNAc5, (f) GlcNAc7→GlcA6, (g) GlcA8→GlcNAc7, (h) GlcNAc9→GlcA8, (i) GlcA10→GlcNAc9, (j) GlcNAc11→GlcA10, (k) GlcA12→GlcNAc11, (l) GlcNAc13→GlcA12, (m) GlcA14→GlcNAc13, (n) GlcNAc15→GlcA14, (o) GlcA16→GlcNAc15, (p) GlcNAc17→GlcA16, (q) GlcA18→GlcNAc17, (r) GlcNAc19→GlcA18, and (s) GlcA20→GlcNAc19; monosaccharides are numbered from reducing to nonreducing end; ϕ, ψ separated into 2.5° bins.

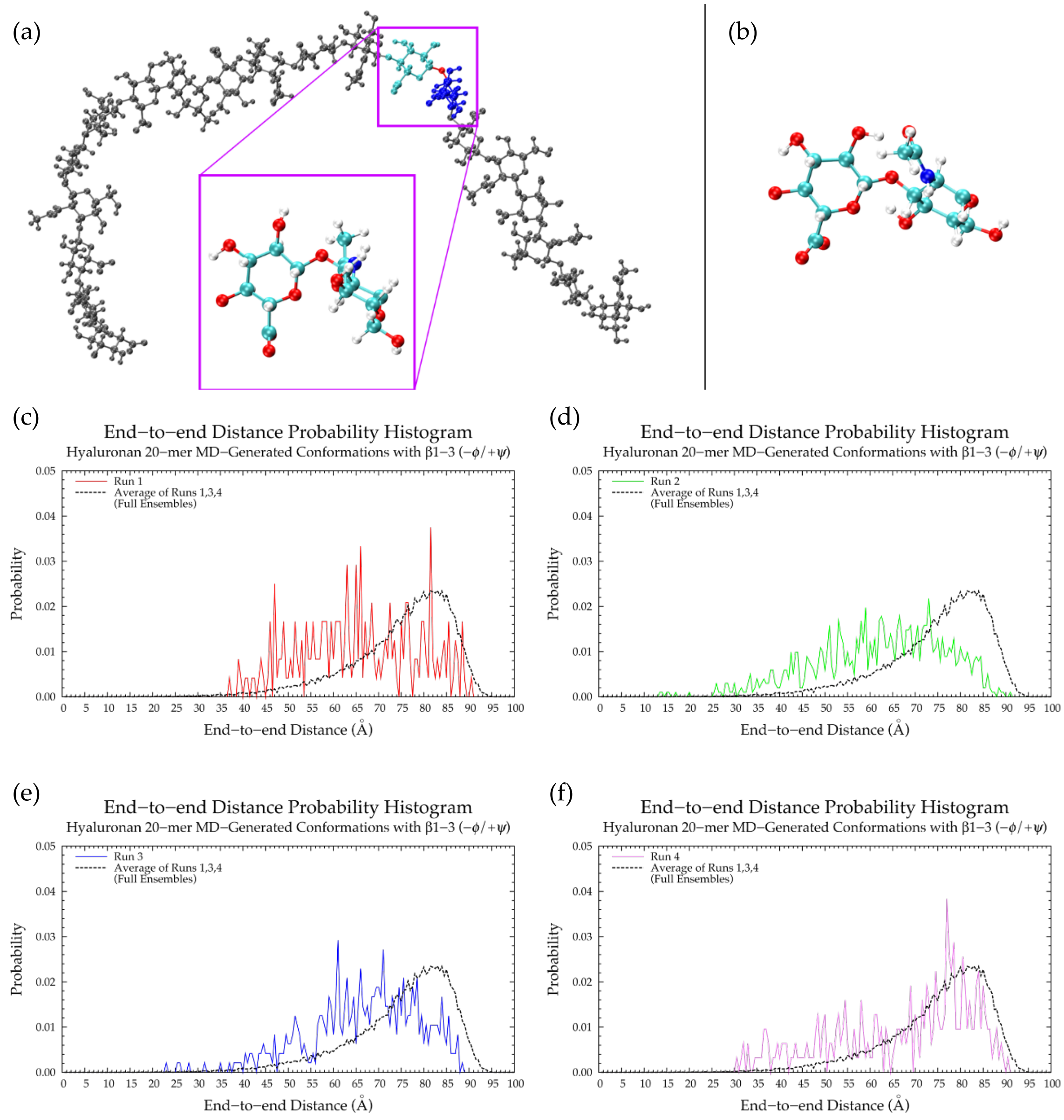


Figure B.8. Snapshots from hyaluronan 20-mer MD: (a) 20-mer conformation with GlcA 8 (cyan) β1-3 GlcNAc 7 (blue) linkage dihedrals near Δ$G(\phi, \psi)$ min II (ϕ = -52.1° and ψ = +91.0°), which causes a kink (linker oxygen is red), and closeup of this disaccharide unit, (b) closeup of the same disaccharide unit with linkage dihedrals near Δ$G(\phi, \psi)$ min I (ϕ = -70.9° and ψ = -119.1°). (c-f) End-to-end distance probability distributions of MD-generated hyaluronan 20-mer conformations with β1-3 linkages with -ϕ, +ψ dihedrals in each of the four runs; as the end-to-end distance distribution from run 2 appeared to be an outlier, these data were compared to the average end-to-end distance distribution of all snapshots in MD runs 1, 3, and 4 (black dashed line).

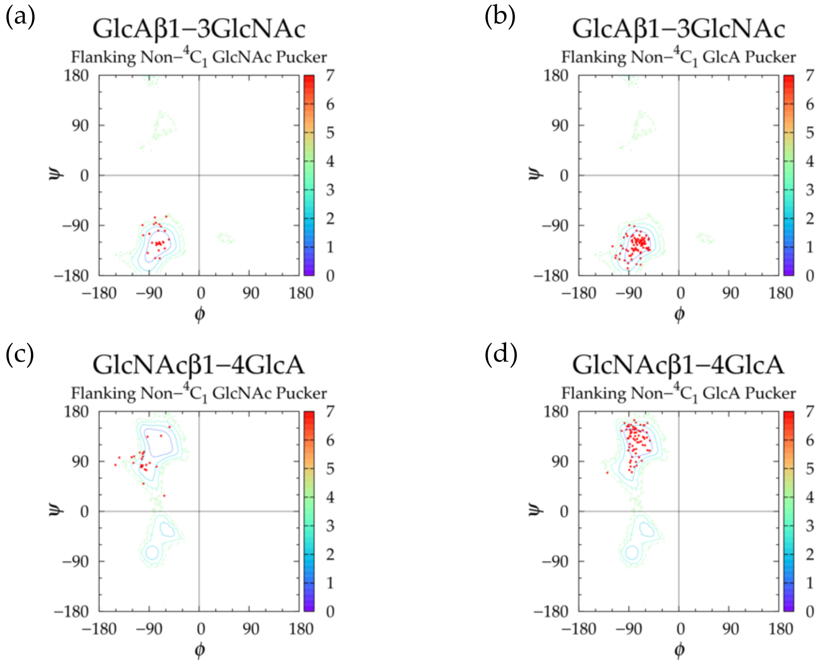


Figure B.9. Scatterplots for dihedrals ϕ and ψ of glycosidic linkages flanking non-$^4C_1$ ring puckers in the MD-generated hyaluronan 20-mer ensemble: (a) GlcAβ1-3GlcNAc flanking GlcNAc, (b) GlcAβ1-3GlcNAc flanking GlcA, (c) GlcNAcβ1-4GlcA flanking GlcNAc, (d) GlcNAcβ1-4GlcA flanking GlcA; contour lines come from corresponding aggregated MD-generated $\Delta G(\phi, \psi)$ data.

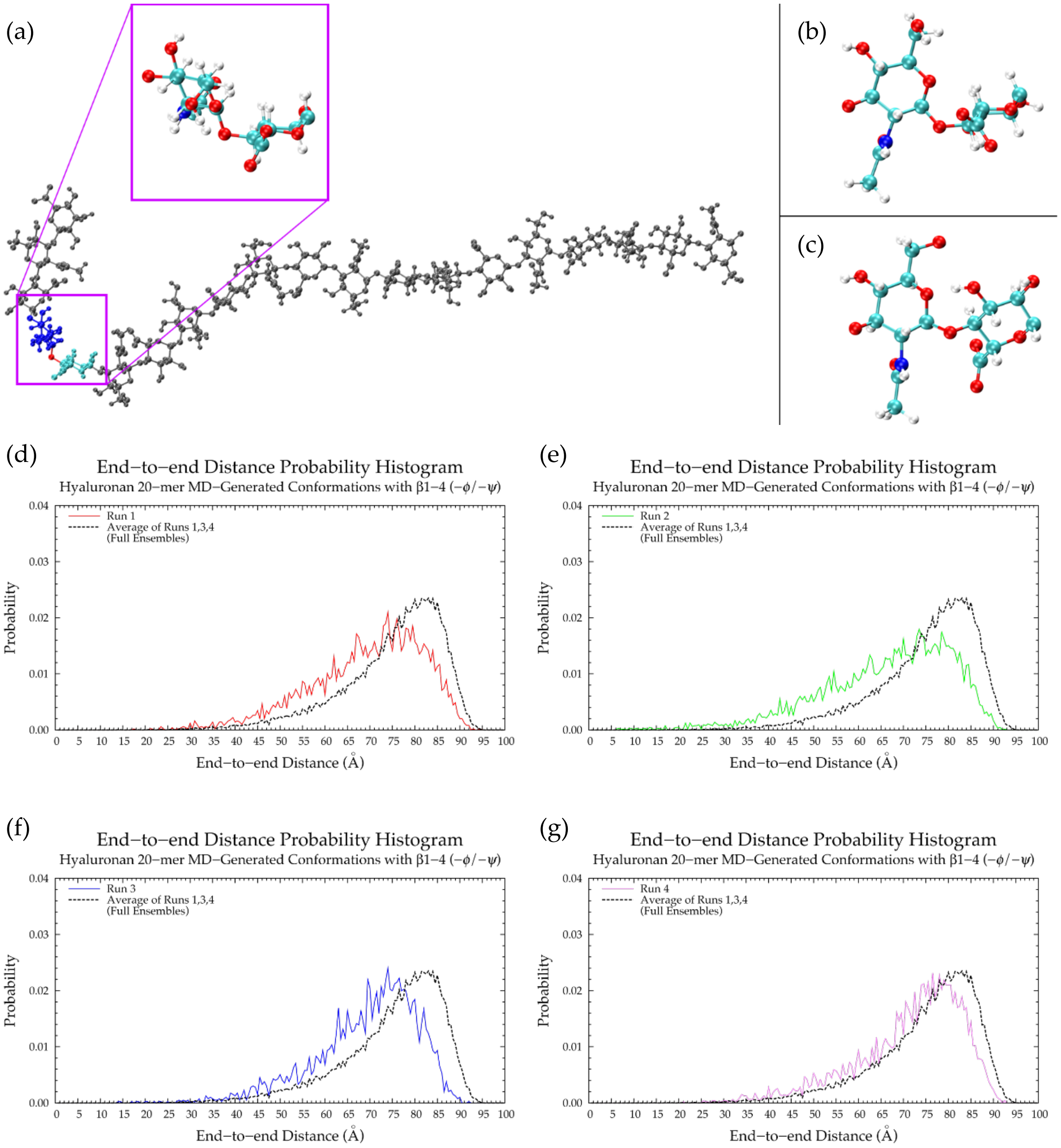


Figure B.10. Snapshots from hyaluronan 20-mer MD: (a) 20-mer conformation with GlcNAc 17 (blue) β1-4 GlcA 16 (cyan) linkage dihedrals near Δ*G*(ϕ, ψ) min II (ϕ = -86.8° and ψ = -78.6°), which causes a kink (linker oxygen is red), and closeup of this disaccharide unit, (b) closeup of this disaccharide unit with linkage dihedrals near Δ*G*(ϕ, ψ) min II′ (ϕ = -55.7° and ψ = -39.7°), (c) closeup of the same disaccharide unit with linkage dihedrals near Δ*G*(ϕ, ψ) min I (ϕ = -68.3° and ψ = +115.2°). (d-g) End-to-end distance probability distributions of MD-generated hyaluronan 20-mer conformations with β1-4 linkages with

-ϕ, -ψ dihedrals in each of the four runs; as the end-to-end distance distribution from run 2 appeared to be an outlier, these data were compared to the average end-to-end distance distribution of all snapshots in MD runs 1, 3, and 4 (black dashed line).

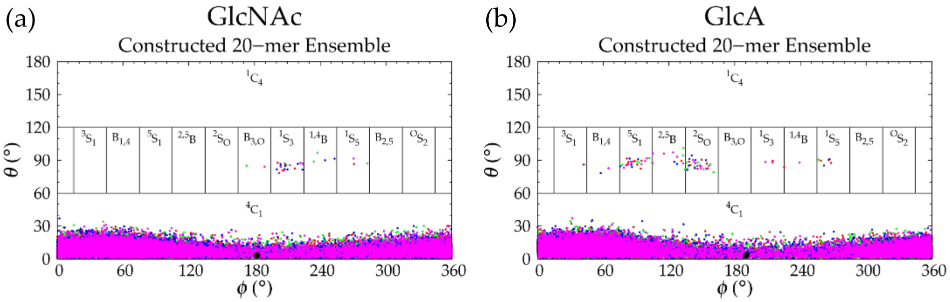


Figure B.11. C–P data for (a) GlcNAc and (b) GlcA in the constructed hyaluronan 20-mer ensemble; each of the 4 runs is represented by different color and the force-field geometry is represented by a single large black dot; each run contains 10,000 parameter sets.

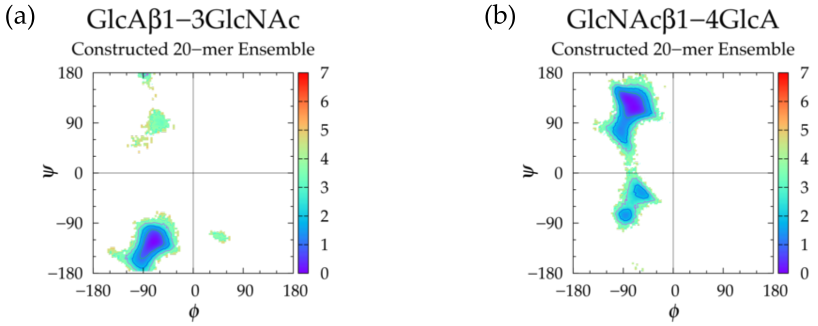


Figure B.12. ΔG(ϕ, ψ) in the constructed hyaluronan 20-mer ensemble for aggregated (a) GlcAβ1-3GlcNAc and (b) GlcNAcβ1-4GlcA glycosidic linkage data; contour lines every 1 kcal/mol.

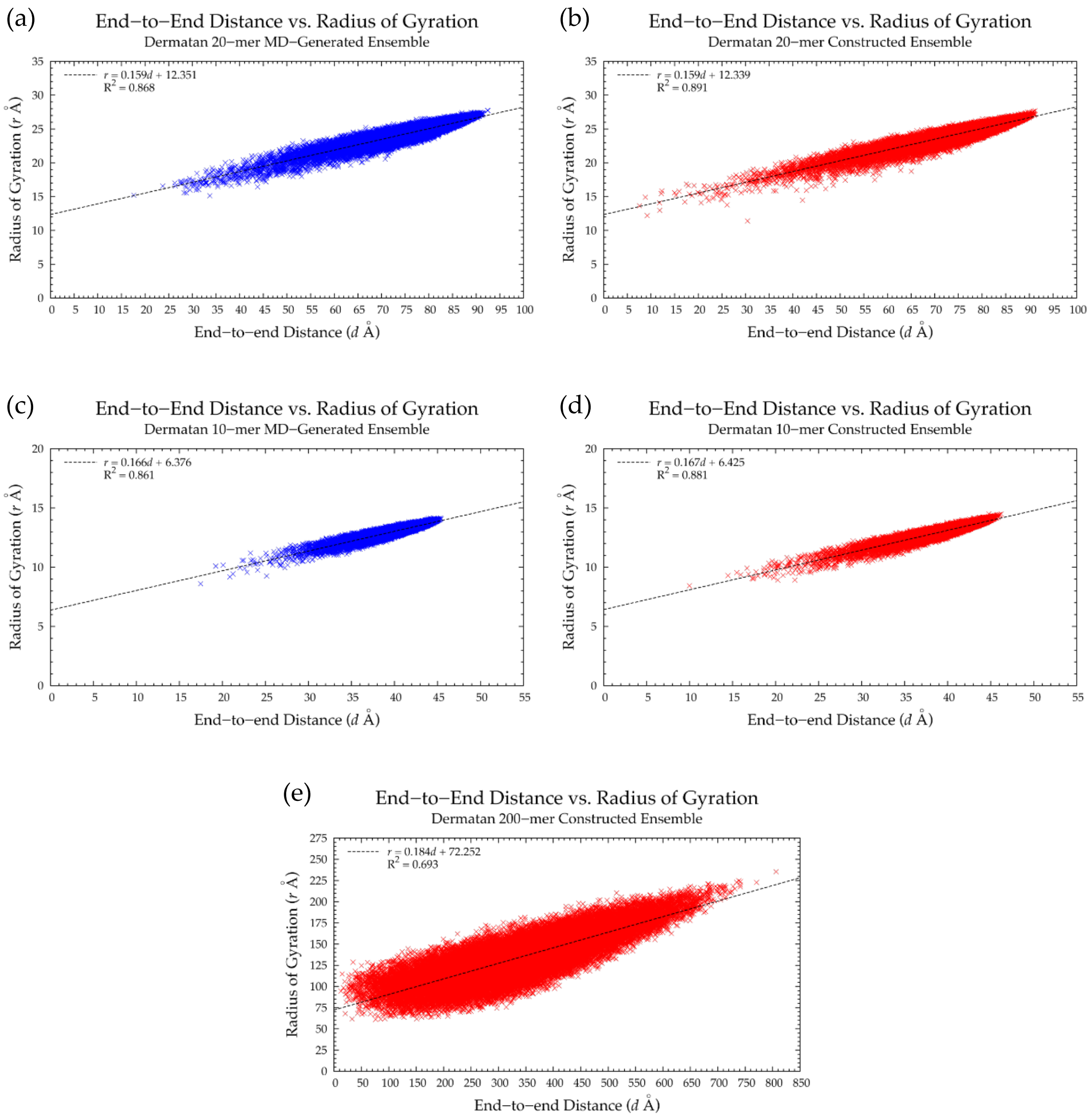


Figure B.13. Scatterplots of radius of gyration as a function of end-to-end distance in MD-generated and constructed ensembles of nonsulfated dermatan (a,b) 20-mer and (c,d) 10-mer, respectively, and (e) constructed ensemble of nonsulfated dermatan 200-mer; each plot has 40,000 samples and shows linear regression and R^2.

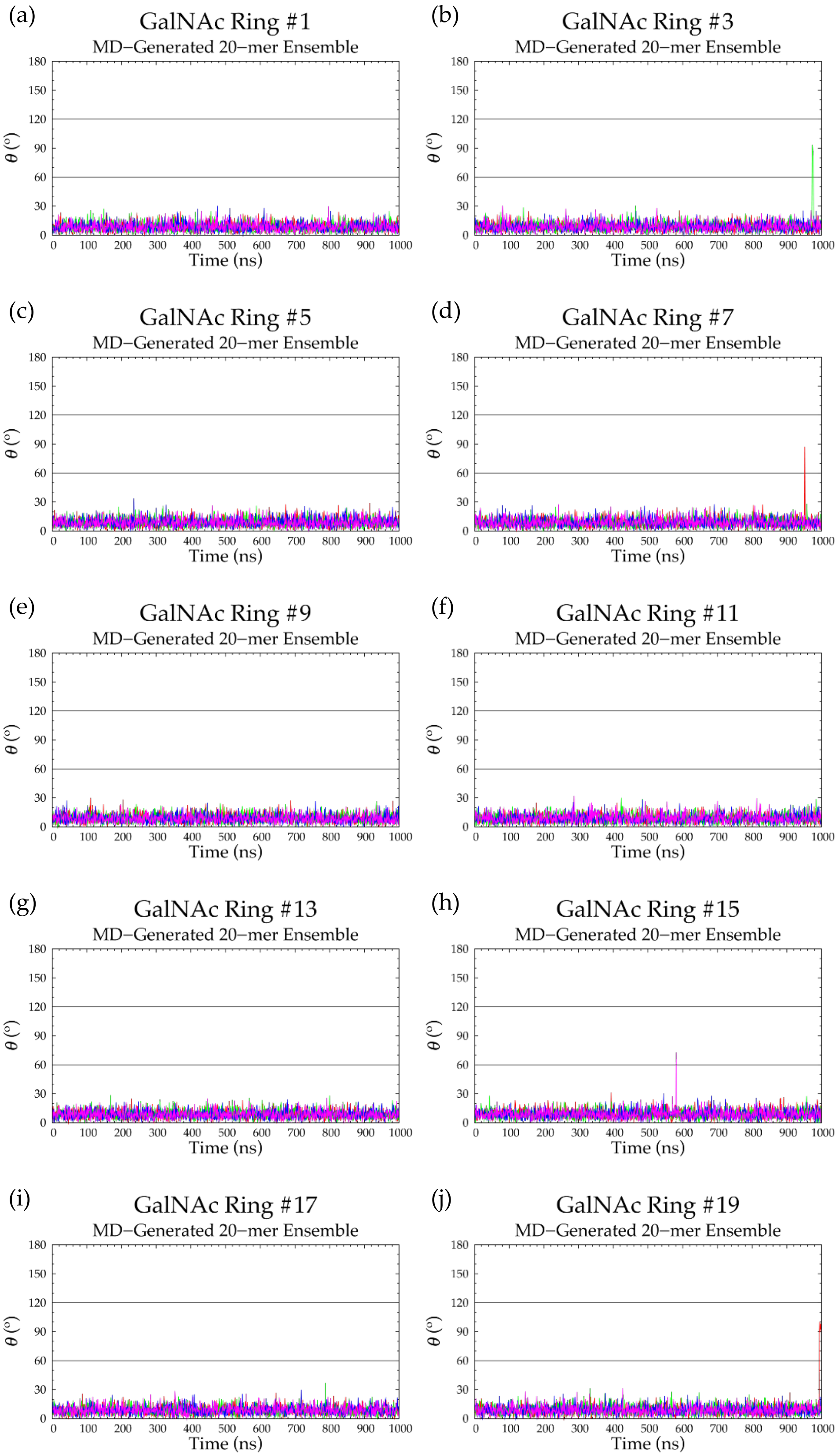

(a)
GalNAc Ring #1
(b)
GalNAc Ring #3
(c)
GalNAc Ring #5
(d)
GalNAc Ring #7
(e)
GalNAc Ring #9
(f)
GalNAc Ring #11
(g)
GalNAc Ring #13
(h)
GalNAc Ring #15
(i)
GalNAc Ring #17
(j)
GalNAc Ring #19
MD−Generated 20−mer Ensemble
θ (°)
0
30
60
90
120
150
180
Time (ns)
0
100
200
300
400
500
600
700
800
900
1000

Figure B.14. (a-j) C-P parameter θ timeseries for each GalNAc monosaccharide ring in the MD-generated nonsulfated dermatan 20-mer ensemble; monosaccharides are numbered from reducing to nonreducing end; each of the 4 runs is represented by different color.

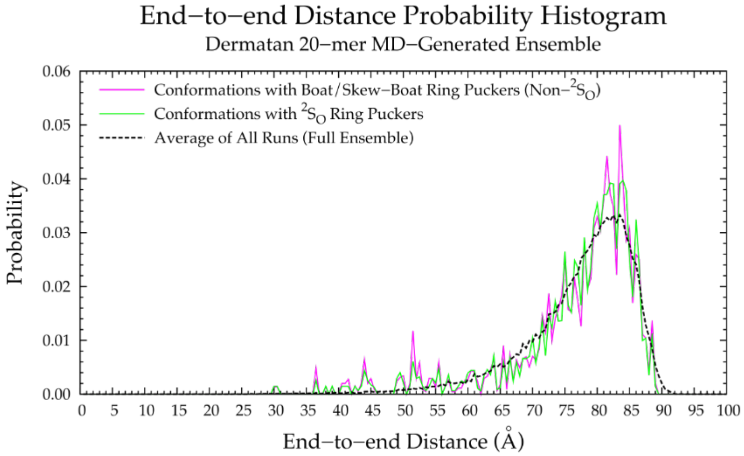
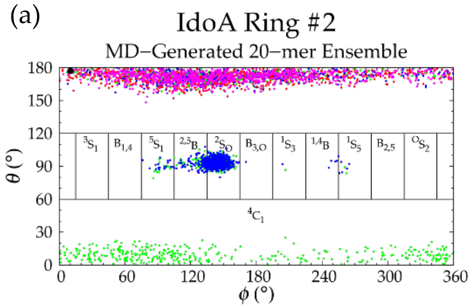
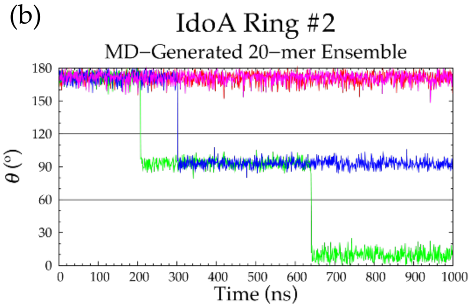


Figure B.15. End-to-end distance distributions of MD-generated nonsulfated dermatan 20-mer conformations with boat/skew-boat ring puckers that cause a kink in the polymer chain, i.e. non-^2S_O (pink solid line; most probable end-to-end distance is 83.5 Å) and ^2S_O conformations (green solid line; most probable end-to-end distance is 84.0 Å) and the average of all four runs in the full MD-generated ensemble (black dashed line; most probable end-to-end distance is 83.5 Å); probabilities were calculated for end-to-end distances sorted into 0.5 Å bins.

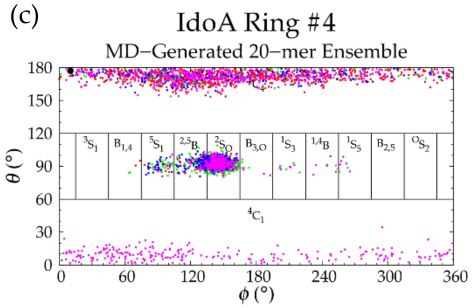
(a)


IdoA Ring #2
MD−Generated 20−mer Ensemble
θ (°)
φ (°)
180
150
120
90
60
30
0
0
60
120
180
240
300
360
$^3S_1$
$B_{1,4}$
$^5S_1$
$^{2,5}B$
^2S_O
$B_{3,O}$
$^1S_3$
$^{1,4}B$
$^1S_5$
$B_{2,5}$
$^OS_2$
$^4C_1$

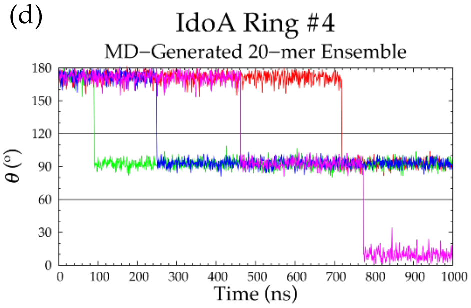
(b)


IdoA Ring #2
MD−Generated 20−mer Ensemble
θ (°)
Time (ns)
180
150
120
90
60
30
0
0
100
200
300
400
500
600
700
800
900
1000

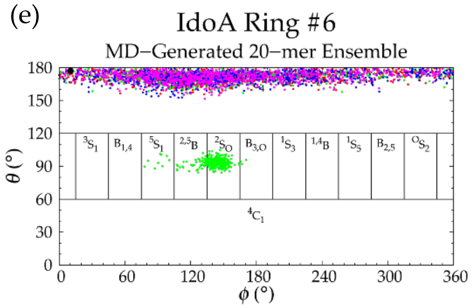
(c)


IdoA Ring #4
MD−Generated 20−mer Ensemble
θ (°)
φ (°)
180
150
120
90
60
30
0
0
60
120
180
240
300
360
$^3S_1$
$B_{1,4}$
$^5S_1$
$^{2,5}B$
^2S_O
$B_{3,O}$
$^1S_3$
$^{1,4}B$
$^1S_5$
$B_{2,5}$
$^OS_2$
$^4C_1$

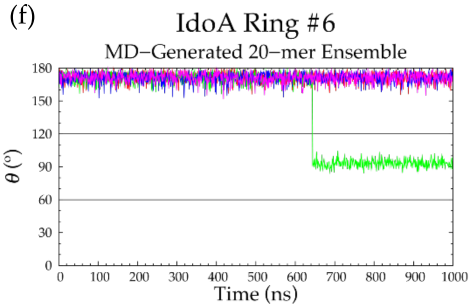
(d)


IdoA Ring #4
MD−Generated 20−mer Ensemble
θ (°)
Time (ns)
180
150
120
90
60
30
0
0
100
200
300
400
500
600
700
800
900
1000

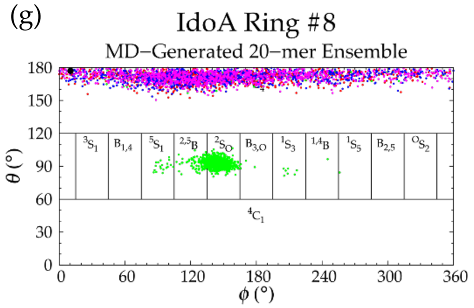
(e)


IdoA Ring #6
MD−Generated 20−mer Ensemble
θ (°)
φ (°)
180
150
120
90
60
30
0
0
60
120
180
240
300
360
$^3S_1$
$B_{1,4}$
$^5S_1$
$^{2,5}B$
^2S_O
$B_{3,O}$
$^1S_3$
$^{1,4}B$
$^1S_5$
$B_{2,5}$
$^OS_2$
$^4C_1$

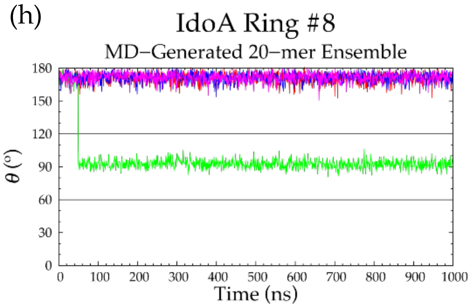
(f)


IdoA Ring #6
MD−Generated 20−mer Ensemble
θ (°)
Time (ns)
180
150
120
90
60
30
0
0
100
200
300
400
500
600
700
800
900
1000

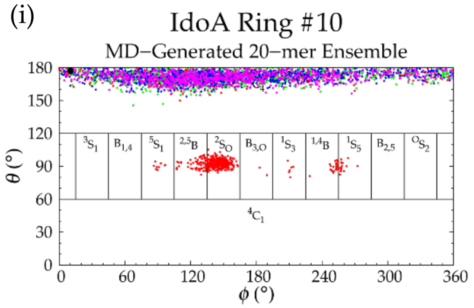
(g)


IdoA Ring #8
MD−Generated 20−mer Ensemble
θ (°)
φ (°)
180
150
120
90
60
30
0
0
60
120
180
240
300
360
$^3S_1$
$B_{1,4}$
$^5S_1$
$^{2,5}B$
^2S_O
$B_{3,O}$
$^1S_3$
$^{1,4}B$
$^1S_5$
$B_{2,5}$
$^OS_2$
$^4C_1$

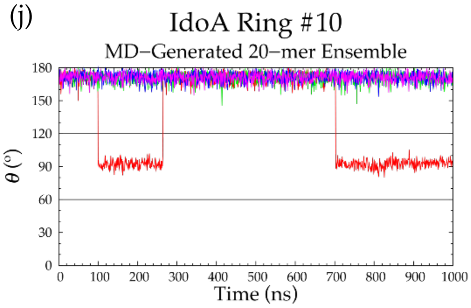
(h)


IdoA Ring #8
MD−Generated 20−mer Ensemble
θ (°)
Time (ns)
180
150
120
90
60
30
0
0
100
200
300
400
500
600
700
800
900
1000

(i)


IdoA Ring #10
MD−Generated 20−mer Ensemble
θ (°)
φ (°)
180
150
120
90
60
30
0
0
60
120
180
240
300
360
$^3S_1$
$B_{1,4}$
$^5S_1$
$^{2,5}B$
^2S_O
$B_{3,O}$
$^1S_3$
$^{1,4}B$
$^1S_5$
$B_{2,5}$
$^OS_2$
$^4C_1$

(j)


IdoA Ring #10
MD−Generated 20−mer Ensemble
θ (°)
Time (ns)
180
150
120
90
60
30
0
0
100
200
300
400
500
600
700
800
900
1000

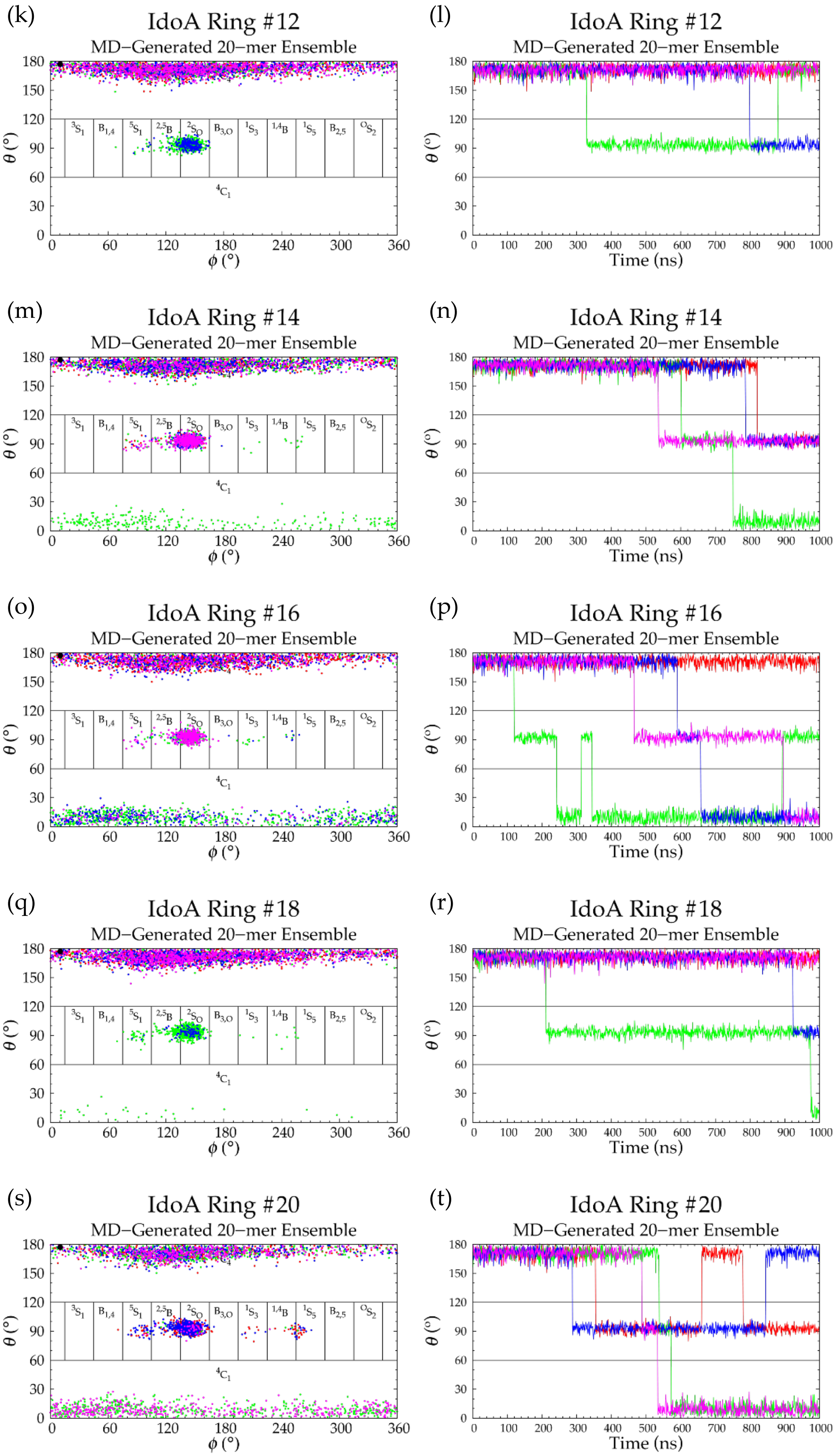
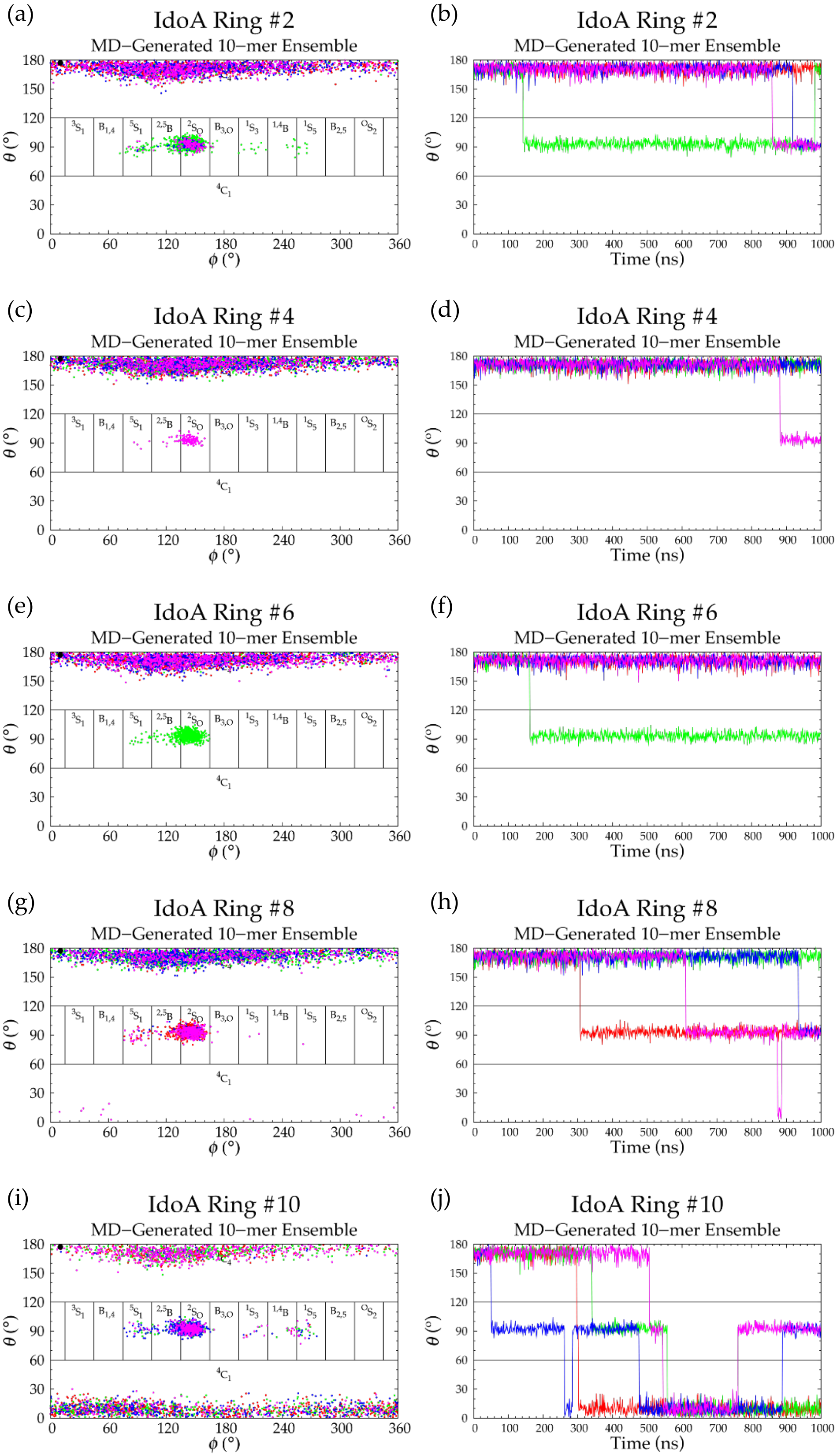

(k)
IdoA Ring #12
MD−Generated 20−mer Ensemble
(l)
IdoA Ring #12
MD−Generated 20−mer Ensemble
(m)
IdoA Ring #14
MD−Generated 20−mer Ensemble
(n)
IdoA Ring #14
MD−Generated 20−mer Ensemble
(o)
IdoA Ring #16
MD−Generated 20−mer Ensemble
(p)
IdoA Ring #16
MD−Generated 20−mer Ensemble
(q)
IdoA Ring #18
MD−Generated 20−mer Ensemble
(r)
IdoA Ring #18
MD−Generated 20−mer Ensemble
(s)
IdoA Ring #20
MD−Generated 20−mer Ensemble
(t)
IdoA Ring #20
MD−Generated 20−mer Ensemble
θ (°)
ϕ (°)
Time (ns)
0
30
60
90
120
150
180
240
300
360
100
200
300
400
500
600
700
800
900
1000
3S1
B1,4
5S1
2,5B
2SO
B3,O
1S3
1,4B
1S5
B2,5
OS2
4C1

Figure B.16. (a,c,e,g,i,k,m,o,q,s) C-P plots and (b,d,f,h,j,l,n,p,r,t) C-P parameter θ timeseries for each IdoA monosaccharide ring in the MD-generated nonsulfated dermatan 20-mer ensemble; monosaccharides are numbered from reducing to nonreducing end; each of the 4 runs is represented by different color.



(a)
IdoA Ring #2
MD−Generated 10−mer Ensemble
(b)
IdoA Ring #2
MD−Generated 10−mer Ensemble
(c)
IdoA Ring #4
MD−Generated 10−mer Ensemble
(d)
IdoA Ring #4
MD−Generated 10−mer Ensemble
(e)
IdoA Ring #6
MD−Generated 10−mer Ensemble
(f)
IdoA Ring #6
MD−Generated 10−mer Ensemble
(g)
IdoA Ring #8
MD−Generated 10−mer Ensemble
(h)
IdoA Ring #8
MD−Generated 10−mer Ensemble
(i)
IdoA Ring #10
MD−Generated 10−mer Ensemble
(j)
IdoA Ring #10
MD−Generated 10−mer Ensemble
θ (°)
ϕ (°)
Time (ns)
0
30
60
90
120
150
180
240
300
360
100
200
300
400
500
600
700
800
900
1000
$^3S_1$
$B_{1,4}$
$^5S_1$
$^{2,5}B$
^2S_O
$B_{3,O}$
$^1S_3$
$^{1,4}B$
$^1S_5$
$B_{2,5}$
$^OS_2$
$^4C_1$

Figure B.17. (a,c,e,g,i) C-P plots and (b,d,f,h,j) C-P parameter θ timeseries for each IdoA monosaccharide ring in the MD-generated nonsulfated dermatan 10-mer ensemble; monosaccharides are numbered from reducing to nonreducing end; each of the 4 runs is represented by different color.

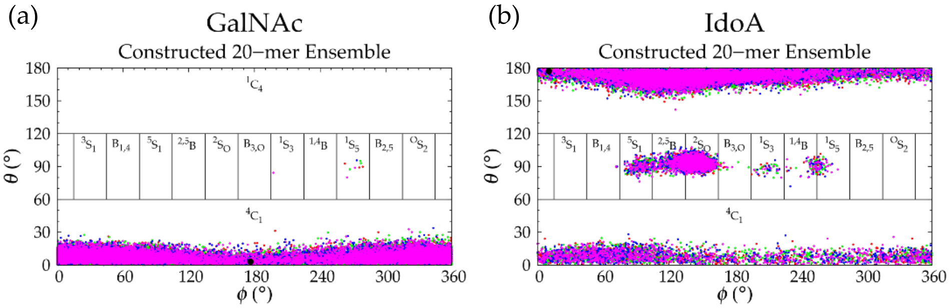


Figure B.18. C–P data for (a) GalNAc and (b) IdoA in the constructed nonsulfated dermatan 20-mer ensemble; each of the 4 runs is represented by different color and the force-field geometry is represented by a single large black dot; each run contains 10,000 parameter sets.

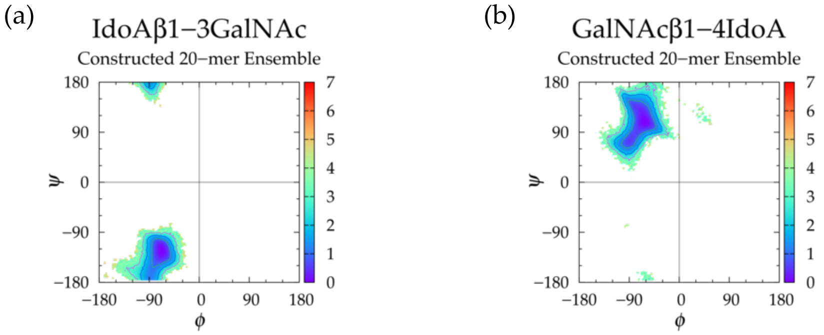


Figure B.19. $\Delta G(\phi, \psi)$ in the constructed nonsulfated dermatan 20-mer ensemble for aggregated (a) IdoAβ1-3GalNAc and (b) GalNAcβ1-4IdoA glycosidic linkage data; contour lines every 1 kcal/mol.

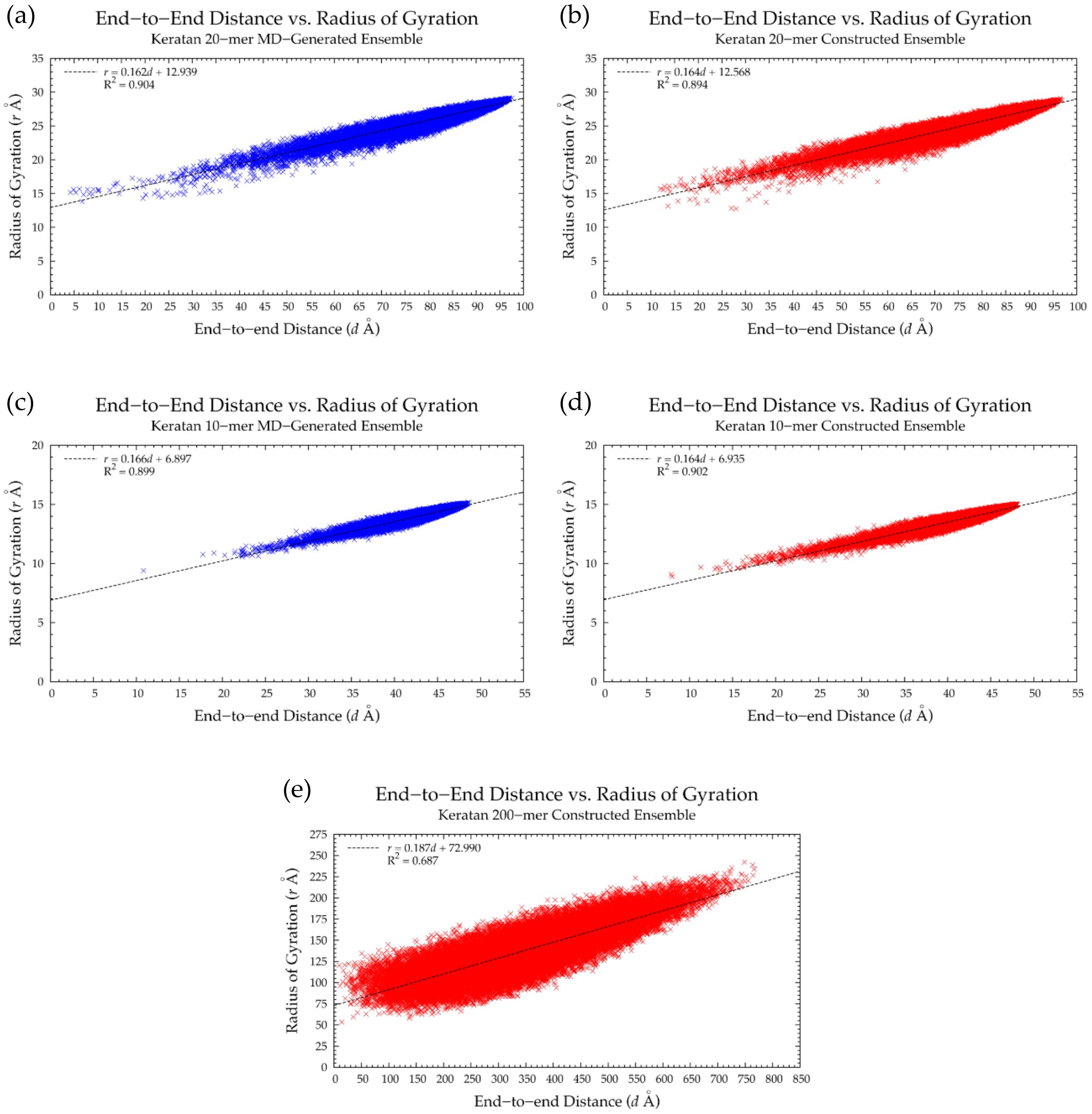
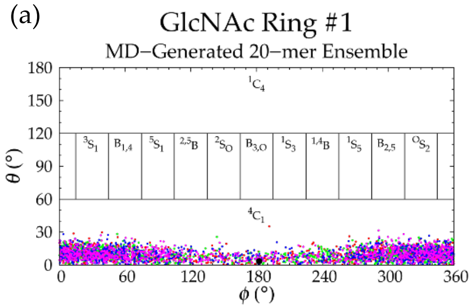


Figure B.20. Scatterplots of radius of gyration as a function of end-to-end distance in MD-generated and constructed ensembles of nonsulfated keratan (a,b) 20-mer and (c,d) 10-mer, respectively, and (e) constructed ensemble of nonsulfated keratan 200-mer; each plot has 40,000 samples and shows linear regression and R^2.

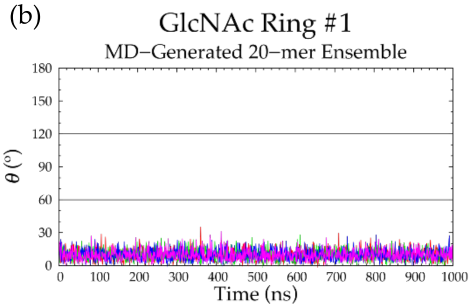
(a)



GlcNAc Ring #1

MD−Generated 20−mer Ensemble

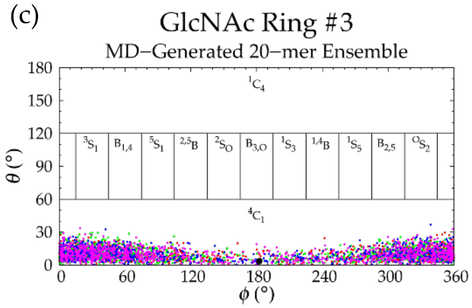
(b)



GlcNAc Ring #1

MD−Generated 20−mer Ensemble

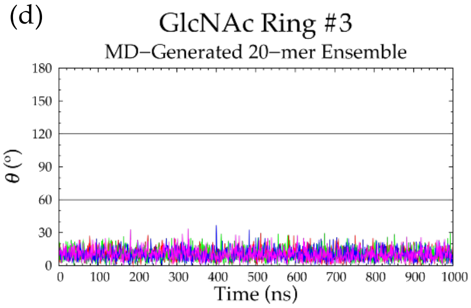
(c)



GlcNAc Ring #3

MD−Generated 20−mer Ensemble

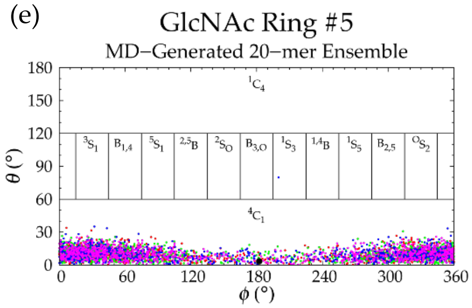
(d)



GlcNAc Ring #3

MD−Generated 20−mer Ensemble

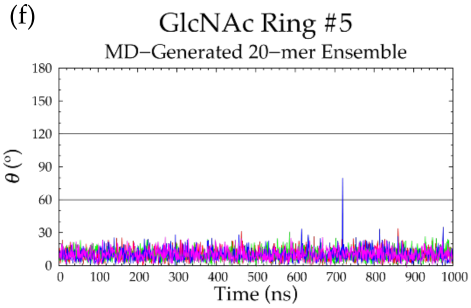
(e)



GlcNAc Ring #5

MD−Generated 20−mer Ensemble

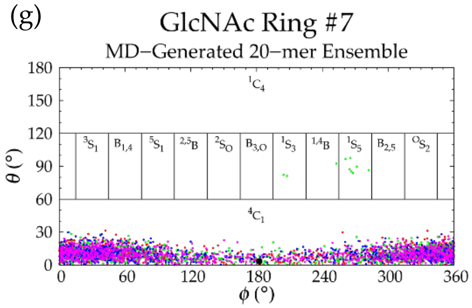
(f)



GlcNAc Ring #5

MD−Generated 20−mer Ensemble

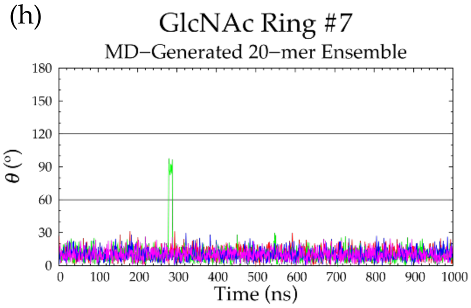
(g)



GlcNAc Ring #7

MD−Generated 20−mer Ensemble

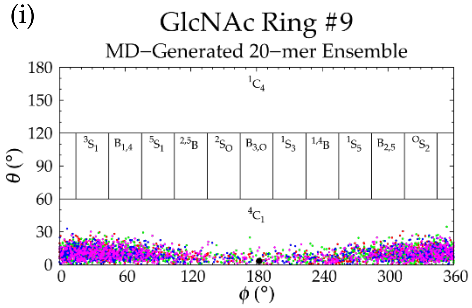
(h)



GlcNAc Ring #7

MD−Generated 20−mer Ensemble

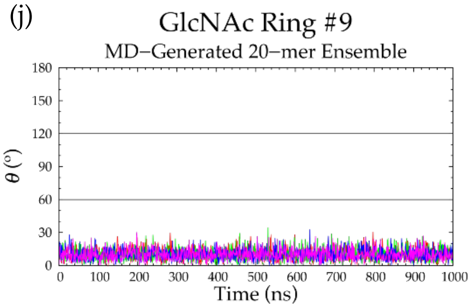
(i)



GlcNAc Ring #9

MD−Generated 20−mer Ensemble

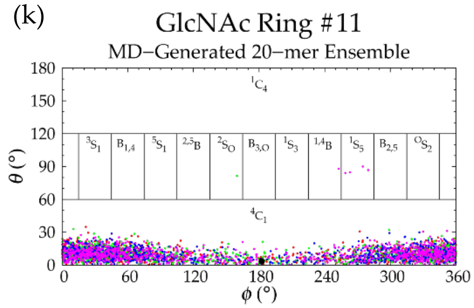
(j)



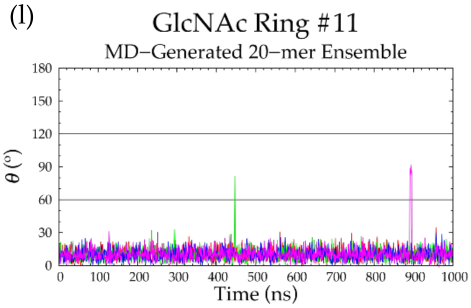
GlcNAc Ring #9

MD−Generated 20−mer Ensemble

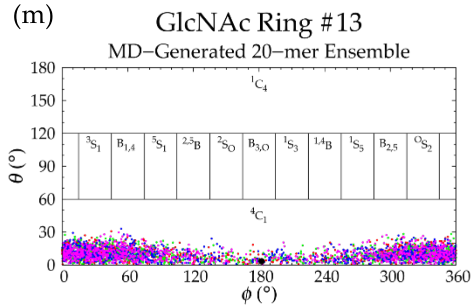
(k)


GlcNAc Ring #11
MD−Generated 20−mer Ensemble

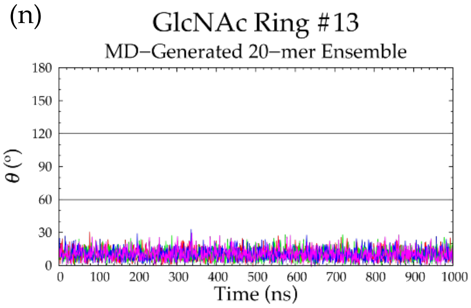
(l)


GlcNAc Ring #11
MD−Generated 20−mer Ensemble

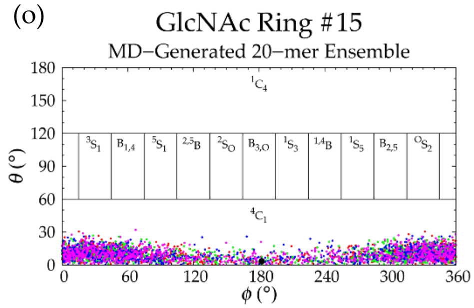
(m)


GlcNAc Ring #13
MD−Generated 20−mer Ensemble

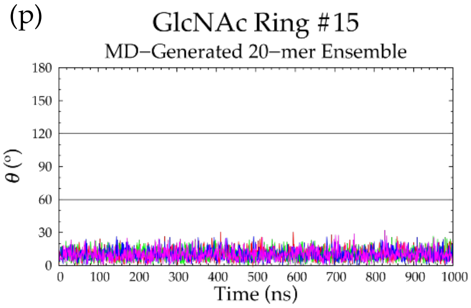
(n)


GlcNAc Ring #13
MD−Generated 20−mer Ensemble

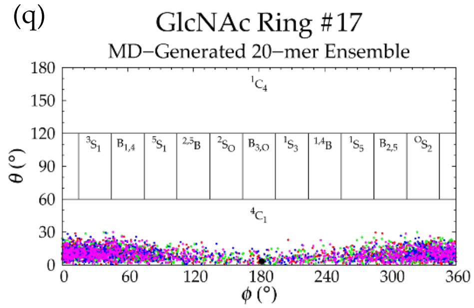
(o)


GlcNAc Ring #15
MD−Generated 20−mer Ensemble

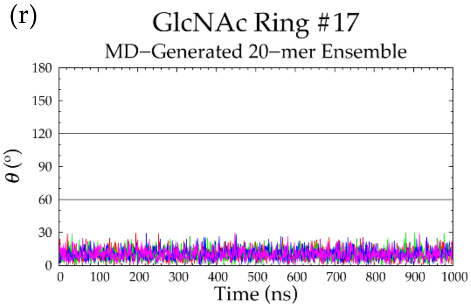
(p)


GlcNAc Ring #15
MD−Generated 20−mer Ensemble

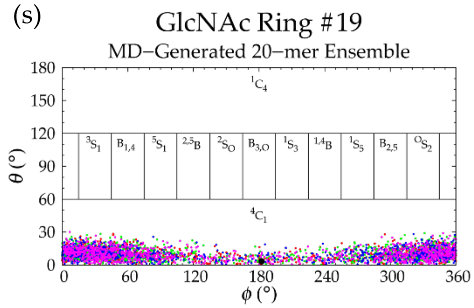
(q)


GlcNAc Ring #17
MD−Generated 20−mer Ensemble

(r)


GlcNAc Ring #17
MD−Generated 20−mer Ensemble

(s)


GlcNAc Ring #19
MD−Generated 20−mer Ensemble

(t)

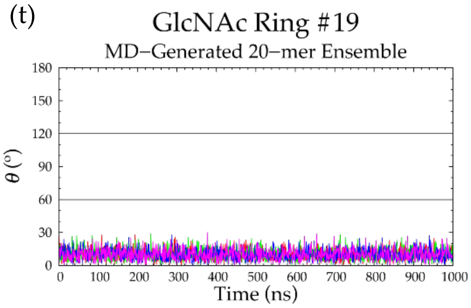

GlcNAc Ring #19
MD−Generated 20−mer Ensemble

Figure B.21. (a,c,e,g,i,k,m,o,q,s) C-P plots and (b,d,f,h,j,l,n,p,r,t) C-P parameter θ timeseries for each GlcNAc monosaccharide ring in the MD-generated nonsulfated keratan 20-mer ensemble; monosaccharides are numbered from reducing to nonreducing end; each of the 4 runs is represented by different color.

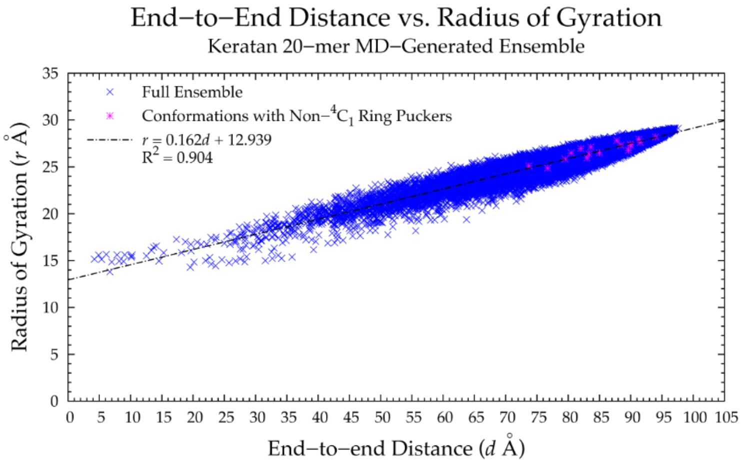


Figure B.22. Scatterplot of radius of gyration as a function of end-to-end distance in MD-generated nonsulfated keratan 20-mer conformations with non-$^4C_1$ ring puckers overlaid on data for full MD-generated nonsulfated keratan 20-mer ensemble and corresponding linear regression (Figure B.20.a).

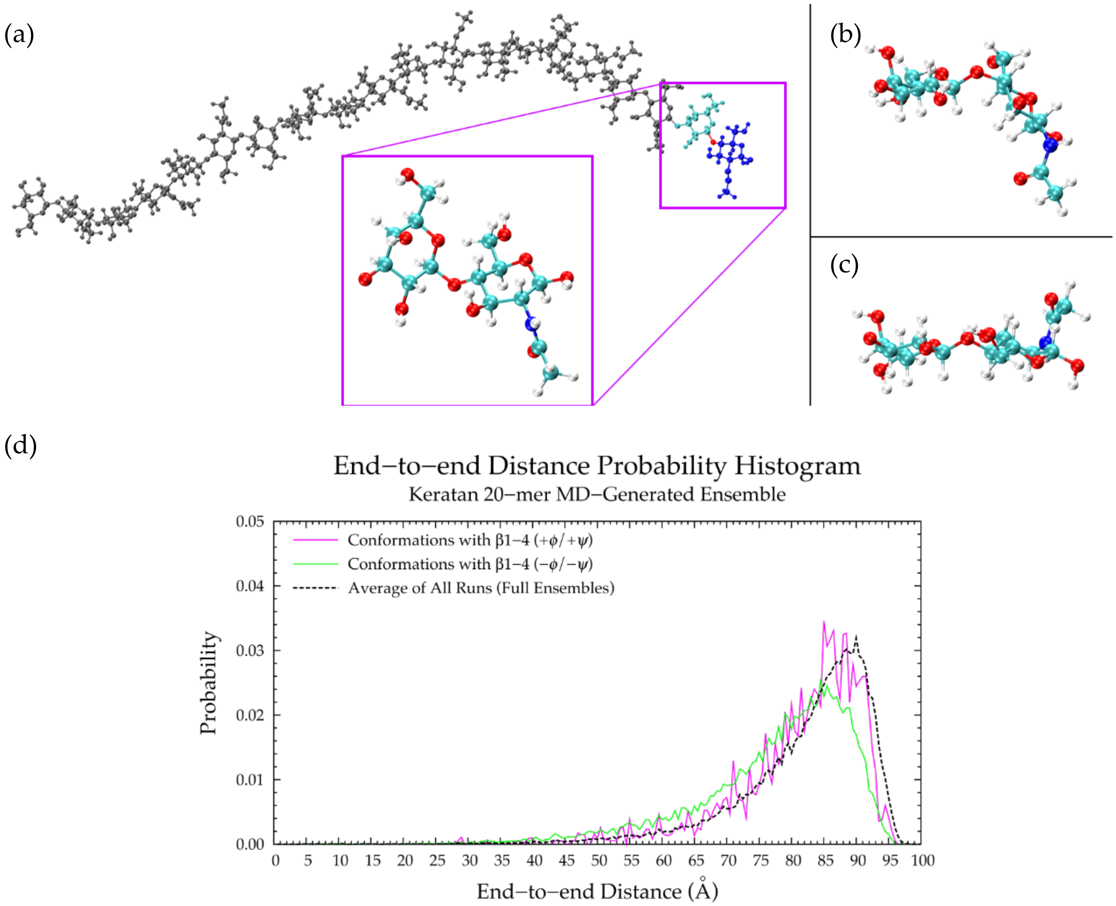


Figure B.23. Snapshots from nonsulfated keratan 20-mer MD: (a) 20-mer conformation with Gal 2 (cyan) β1-4 GlcNAc 1 (blue) linkage dihedrals in tertiary basin (ϕ = +13.8° and ψ = +131.1°), which causes a slight bend (linker oxygen is red), and closeup of this disaccharide unit, (b) closeup of this disaccharide unit with linkage dihedrals in secondary basin, i.e. near $\Delta G(\phi, \psi)$ min II (ϕ = -98.8° and ψ = -71.0°), (c) closeup of the same disaccharide unit with linkage dihedrals in primary basin near $\Delta G(\phi, \psi)$ min I (ϕ = -86.2° and ψ = +100.5°). (d) End-to-end distance probability distributions of MD-generated nonsulfated keratan 20-mer conformations with β1-4 linkages with +ϕ, +ψ (pink solid line; most probable end-to-end distance is 85.0 Å) and -ϕ, -ψ (green solid line; most probable end-to-end distance is 84.5 Å) dihedrals in all four runs; these data were compared to the average end-to-end distance distribution of all snapshots in all four MD runs (black dashed line; most probable end-to-end distance is 90.0 Å).

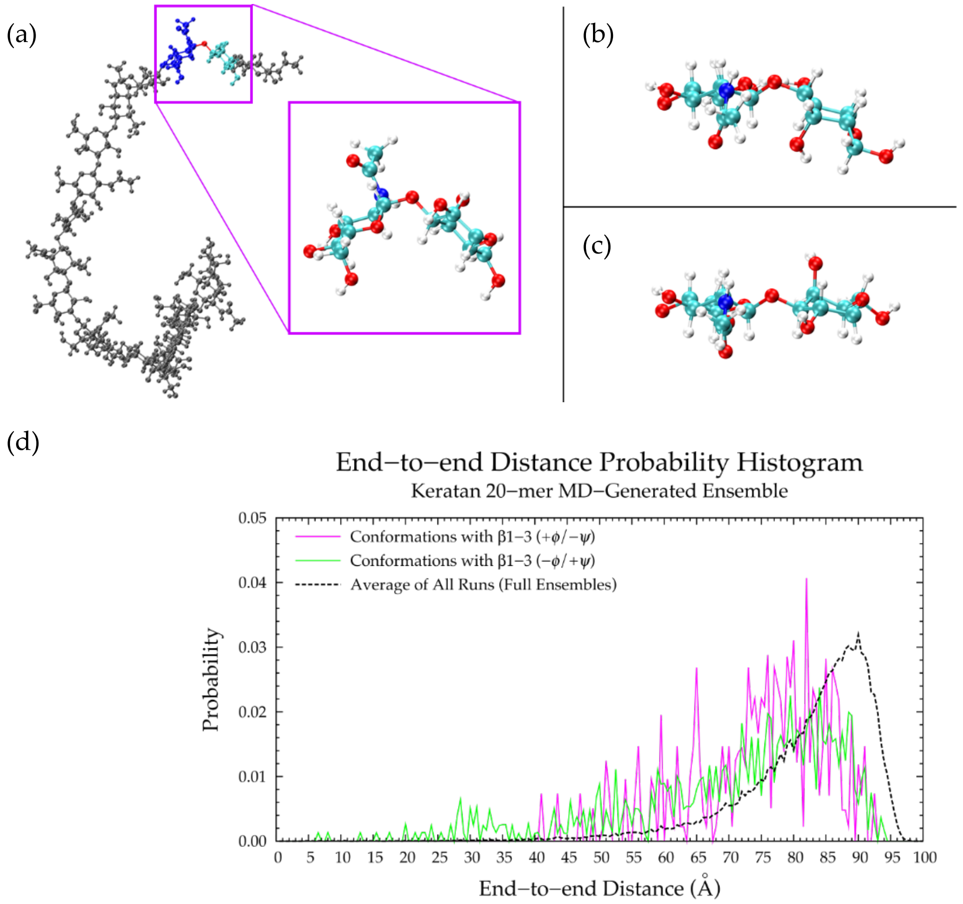
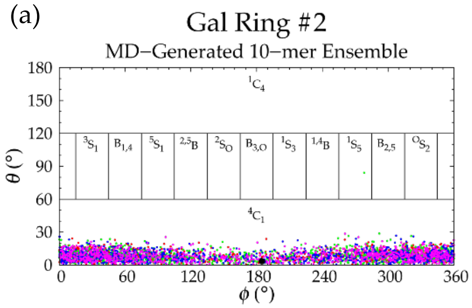
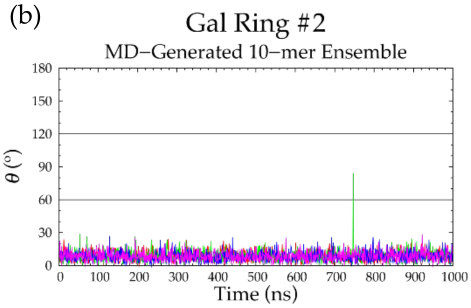


Figure B.24. Snapshots from nonsulfated keratan 20-mer MD: (a) 20-mer conformation with GlcNAc 3 (blue) β1-3 Gal 2 (cyan) linkage dihedrals in tertiary basin (ϕ = +38.8° and ψ = -127.3°), which causes a kink (linker oxygen is red), and closeup of this disaccharide unit, (b) closeup of GlcNAc 5 (blue) β1-3 Gal 4 (cyan) disaccharide unit with linkage dihedrals in secondary basin (ϕ = -95.3° and ψ = +39.9°), (c) closeup of the same disaccharide unit with linkage dihedrals in primary basin (ϕ = -80.7° and ψ = -138.2°). (d) End-to-end distance probability distributions of MD-generated nonsulfated keratan 20-mer conformations with β1-3 linkages with +ϕ, -ψ (pink solid line; most probable end-to-end distance is 82.0 Å) and -ϕ, +ψ (green solid line; most probable end-to-end distance is 84.0 Å) dihedrals in all four runs; these data were compared to the average end-to-end distance distribution of all snapshots in all four MD runs (black dashed line; most probable end-to-end distance is 90.0 Å).

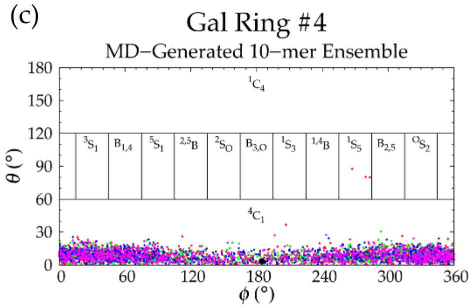
(a)


Gal Ring #2
MD−Generated 10−mer Ensemble
$^{1}C_{4}$
$^{3}S_{1}$ $B_{1,4}$ $^{5}S_{1}$ $^{2,5}B$ $^{2}S_{O}$ $B_{3,O}$ $^{1}S_{3}$ $^{1,4}B$ $^{1}S_{5}$ $B_{2,5}$ $^{O}S_{2}$
$^{4}C_{1}$
θ (°)
0 30 60 90 120 150 180
ϕ (°)
0 60 120 180 240 300 360

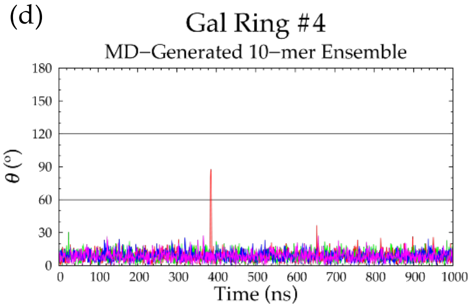
(b)


Gal Ring #2
MD−Generated 10−mer Ensemble
θ (°)
0 30 60 90 120 150 180
Time (ns)
0 100 200 300 400 500 600 700 800 900 1000

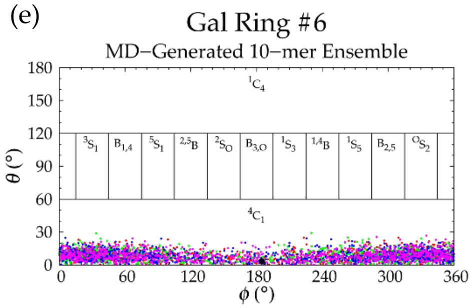
(c)


Gal Ring #4
MD−Generated 10−mer Ensemble
$^{1}C_{4}$
$^{3}S_{1}$ $B_{1,4}$ $^{5}S_{1}$ $^{2,5}B$ $^{2}S_{O}$ $B_{3,O}$ $^{1}S_{3}$ $^{1,4}B$ $^{1}S_{5}$ $B_{2,5}$ $^{O}S_{2}$
$^{4}C_{1}$
θ (°)
0 30 60 90 120 150 180
ϕ (°)
0 60 120 180 240 300 360

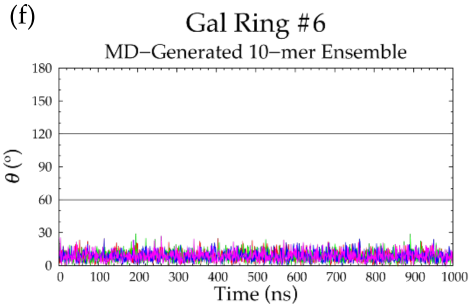
(d)


Gal Ring #4
MD−Generated 10−mer Ensemble
θ (°)
0 30 60 90 120 150 180
Time (ns)
0 100 200 300 400 500 600 700 800 900 1000

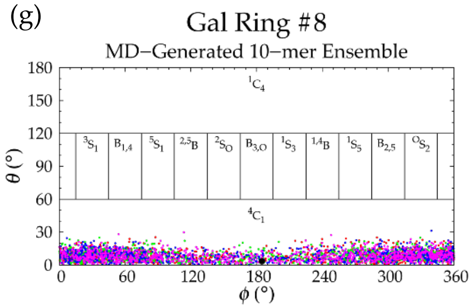
(e)


Gal Ring #6
MD−Generated 10−mer Ensemble
$^{1}C_{4}$
$^{3}S_{1}$ $B_{1,4}$ $^{5}S_{1}$ $^{2,5}B$ $^{2}S_{O}$ $B_{3,O}$ $^{1}S_{3}$ $^{1,4}B$ $^{1}S_{5}$ $B_{2,5}$ $^{O}S_{2}$
$^{4}C_{1}$
θ (°)
0 30 60 90 120 150 180
ϕ (°)
0 60 120 180 240 300 360

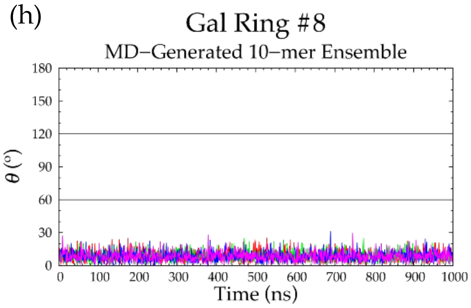
(f)


Gal Ring #6
MD−Generated 10−mer Ensemble
θ (°)
0 30 60 90 120 150 180
Time (ns)
0 100 200 300 400 500 600 700 800 900 1000

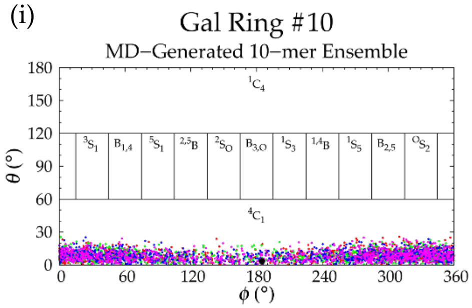
(g)


Gal Ring #8
MD−Generated 10−mer Ensemble
$^{1}C_{4}$
$^{3}S_{1}$ $B_{1,4}$ $^{5}S_{1}$ $^{2,5}B$ $^{2}S_{O}$ $B_{3,O}$ $^{1}S_{3}$ $^{1,4}B$ $^{1}S_{5}$ $B_{2,5}$ $^{O}S_{2}$
$^{4}C_{1}$
θ (°)
0 30 60 90 120 150 180
ϕ (°)
0 60 120 180 240 300 360

(h)


Gal Ring #8
MD−Generated 10−mer Ensemble
θ (°)
0 30 60 90 120 150 180
Time (ns)
0 100 200 300 400 500 600 700 800 900 1000

(i)


Gal Ring #10
MD−Generated 10−mer Ensemble
$^{1}C_{4}$
$^{3}S_{1}$ $B_{1,4}$ $^{5}S_{1}$ $^{2,5}B$ $^{2}S_{O}$ $B_{3,O}$ $^{1}S_{3}$ $^{1,4}B$ $^{1}S_{5}$ $B_{2,5}$ $^{O}S_{2}$
$^{4}C_{1}$
θ (°)
0 30 60 90 120 150 180
ϕ (°)
0 60 120 180 240 300 360

(j)

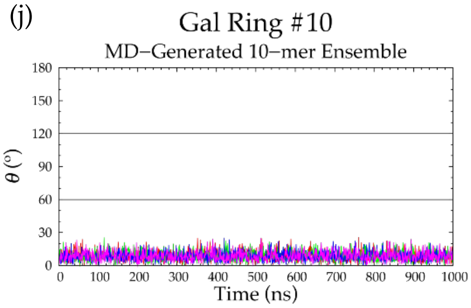

Gal Ring #10
MD−Generated 10−mer Ensemble
θ (°)
0 30 60 90 120 150 180
Time (ns)
0 100 200 300 400 500 600 700 800 900 1000

Figure B.25. (a,c,e,g,i) C-P plots and (b,d,f,h,j) C-P parameter θ timeseries for each Gal monosaccharide ring in the MD-generated nonsulfated keratan 10-mer ensemble; monosaccharides are numbered from reducing to nonreducing end; each of the 4 runs is represented by different color.

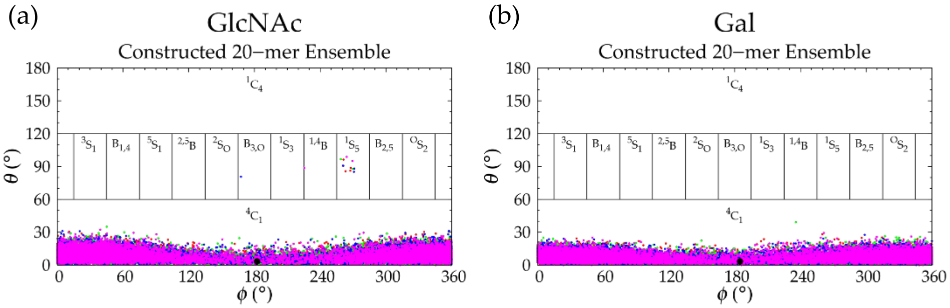


Figure B.26. C–P data for (a) GlcNAc and (b) Gal in the constructed nonsulfated keratan 20-mer ensemble; each of the 4 runs is represented by different color and the force-field geometry is represented by a single large black dot; each run contains 10,000 parameter sets.

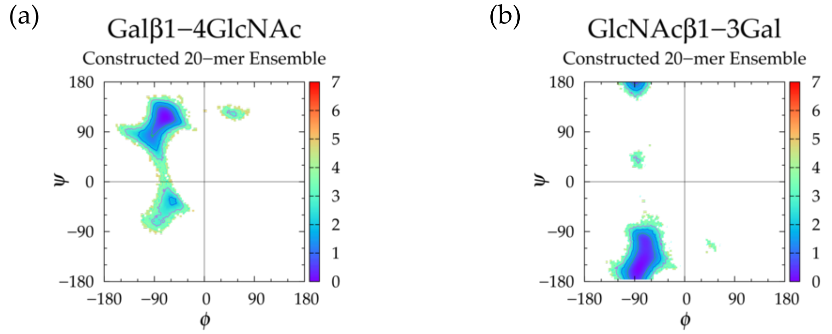


Figure B.27. $\Delta G(\phi, \psi)$ in the constructed nonsulfated keratan 20-mer ensemble for aggregated (a) Galβ1-4GlcNAc and (b) GlcNAcβ1-3Gal glycosidic linkage data; contour lines every 1 kcal/mol.

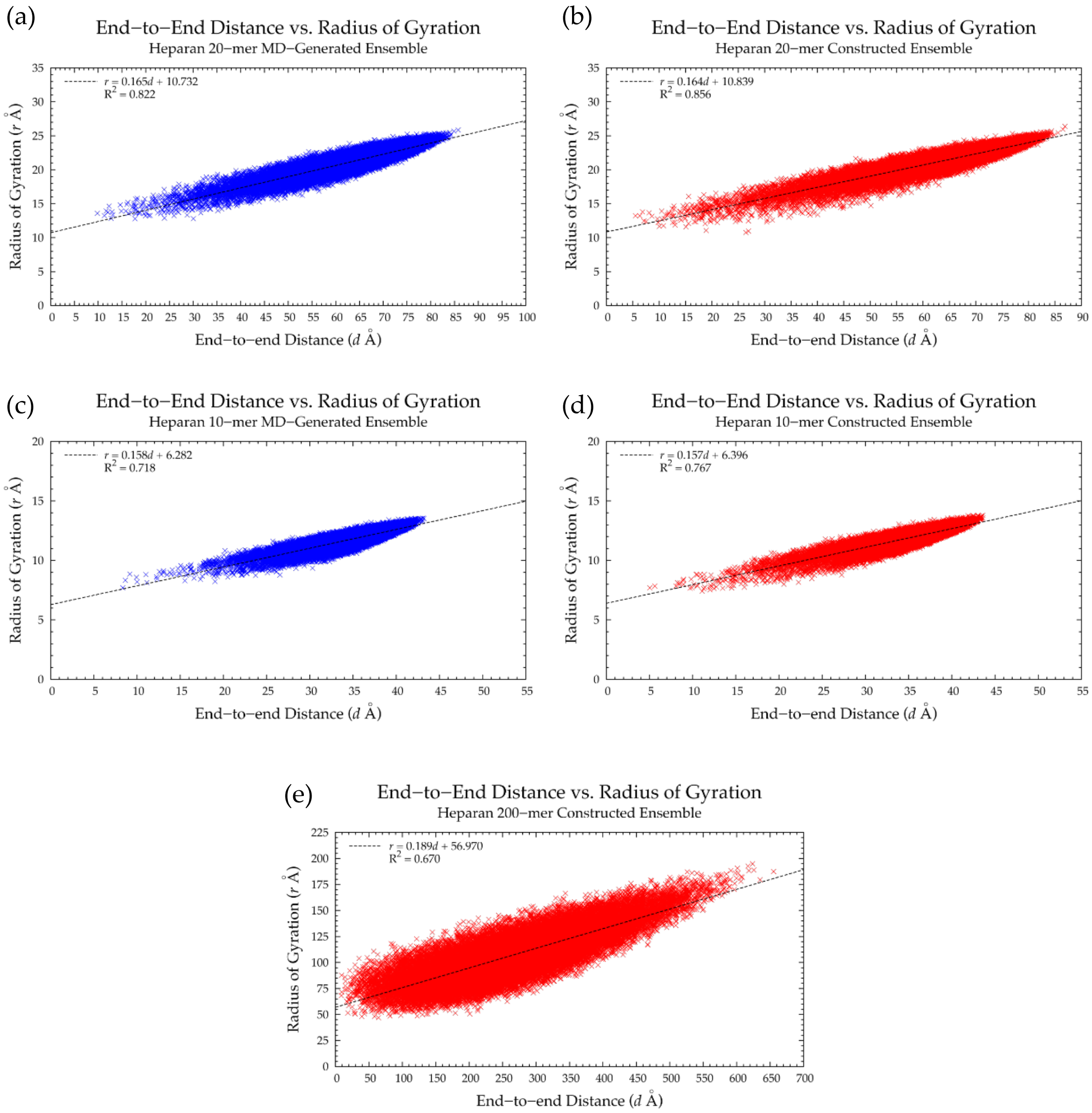
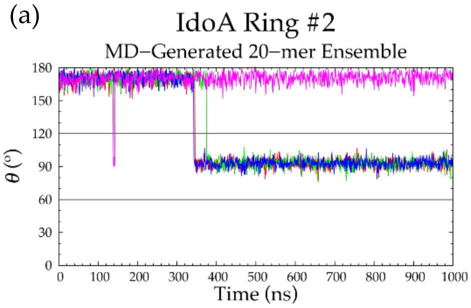
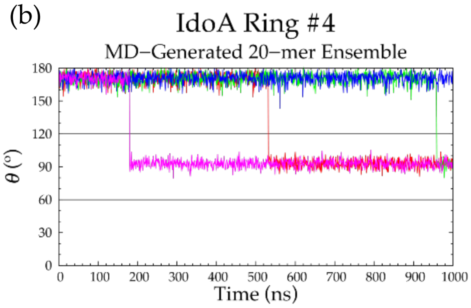


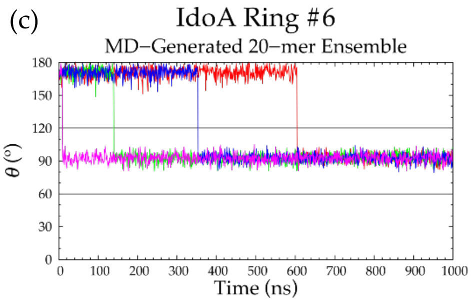
Figure B.28. Scatterplots of radius of gyration as a function of end-to-end distance in MD-generated and constructed ensembles of nonsulfated heparan (a,b) 20-mer and (c,d) 10-mer, respectively, and (e) constructed ensemble of nonsulfated heparan 200-mer; each plot has 40,000 samples and shows linear regression and R^2.


(a) IdoA Ring #2

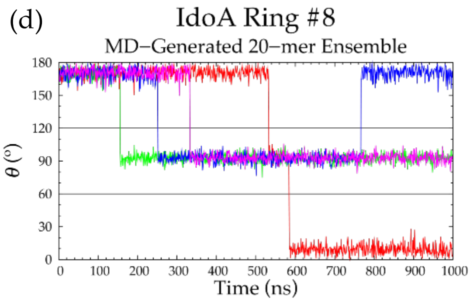
MD−Generated 20−mer Ensemble


(b) IdoA Ring #4

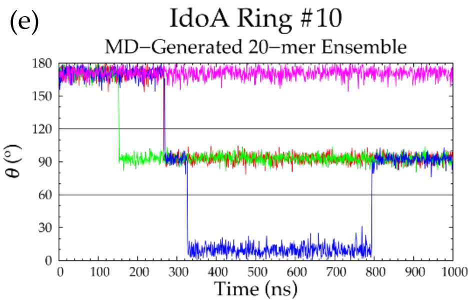
MD−Generated 20−mer Ensemble


(c) IdoA Ring #6

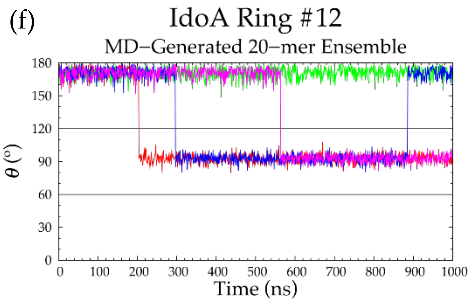
MD−Generated 20−mer Ensemble


(d) IdoA Ring #8

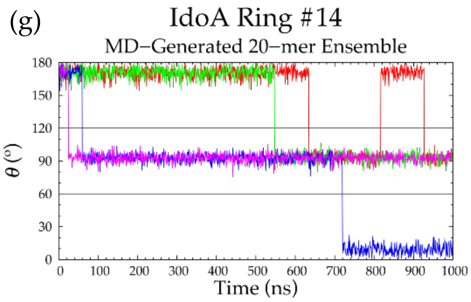
MD−Generated 20−mer Ensemble


(e) IdoA Ring #10

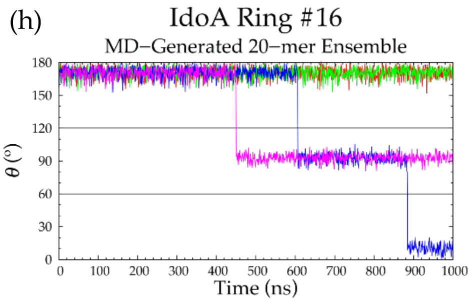
MD−Generated 20−mer Ensemble


(f) IdoA Ring #12

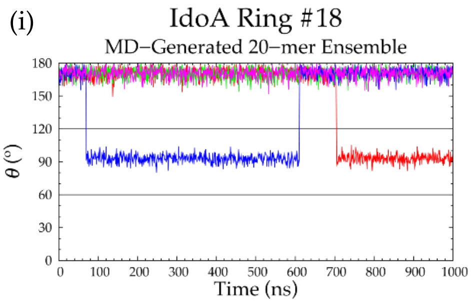
MD−Generated 20−mer Ensemble


(g) IdoA Ring #14

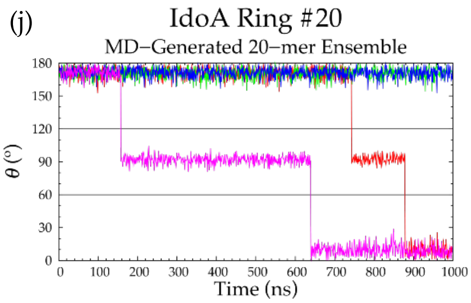
MD−Generated 20−mer Ensemble


(h) IdoA Ring #16

MD−Generated 20−mer Ensemble


(i) IdoA Ring #18

MD−Generated 20−mer Ensemble


(j) IdoA Ring #20

MD−Generated 20−mer Ensemble

Figure B.29. (a-j) C-P parameter θ timeseries for each IdoA monosaccharide ring in the MD-generated nonsulfated heparan 20-mer ensemble; monosaccharides are numbered from reducing to nonreducing end; each of the 4 runs is represented by different color.

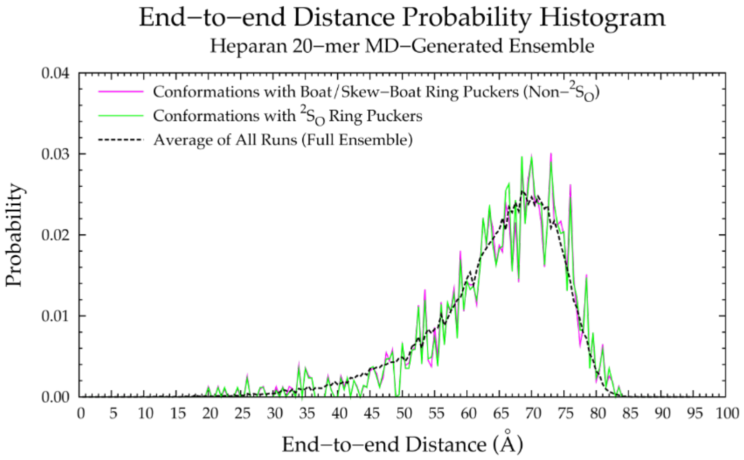


Figure B.30. End-to-end distance distributions of MD-generated nonsulfated heparan 20-mer conformations with boat/skew-boat ring puckers that cause a kink in the polymer chain, i.e. non-^2S_O (pink solid line; most probable end-to-end distance is 73.0 Å) and ^2S_O conformations (green solid line; most probable end-to-end distance is 68.5 Å) and the average of all four runs in the full MD-generated ensemble (black dashed line; most probable end-to-end distance is 68.5 Å); probabilities were calculated for end-to-end distances sorted into 0.5 Å bins.

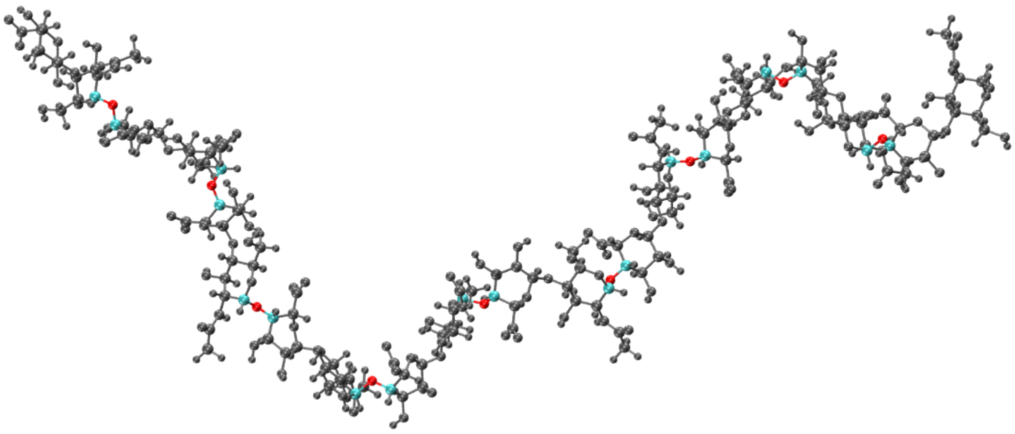


Figure B.31. Snapshot of nonsulfated heparan 20-mer from MD simulation; GlcNAcα1-4IdoA linkages are highlighted; linker oxygen atoms in red and carbon atoms (GlcNAc C_1 and IdoA C_4) in cyan.

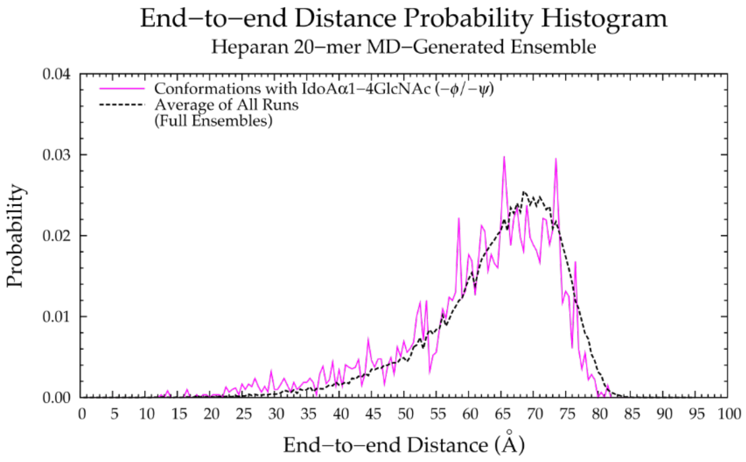


Figure B.32. End-to-end distance probability distribution of MD-generated nonsulfated heparan 20-mer conformations with IdoAα1-4GlcNAc linkages with -ϕ, -ψ dihedrals aggregated across all four MD runs (pink solid line; top two most probable end-to-end distances are 65.5 Å and 73.5 Å); these data were compared to the average end-to-end distance distribution of all snapshots in all four MD runs (black dashed line; most probable end-to-end distance is 68.5 Å).

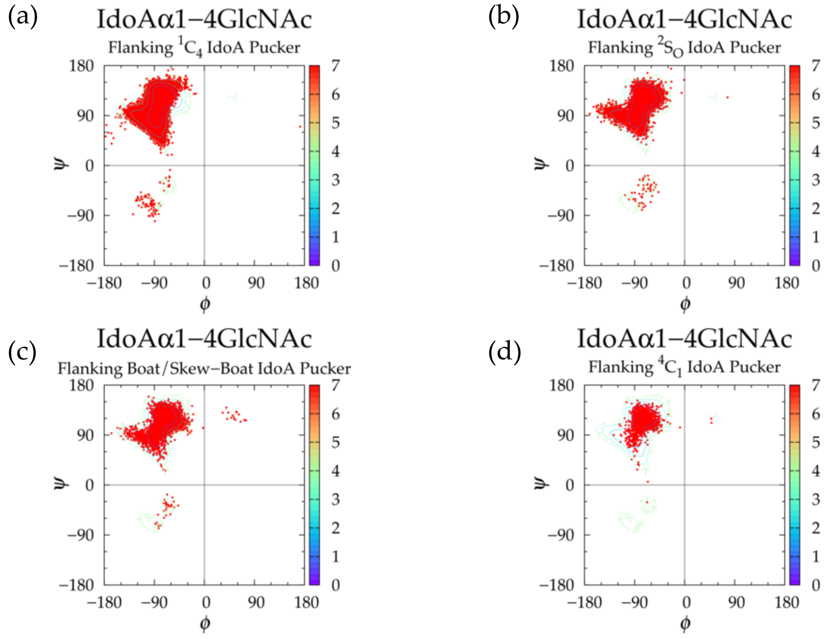


Figure B.33. Scatterplots for dihedrals ϕ and ψ of IdoAα1-4GlcNAc linkages flanking different IdoA conformations in the MD-generated nonsulfated heparan 20-mer ensemble: (a) $^1C_4$, (b) ^2S_O, (c) boat/skew-boat (non-^2S_O), and (d) $^4C_1$; contour lines come from corresponding aggregated MD-generated $\Delta G(\phi, \psi)$ data.

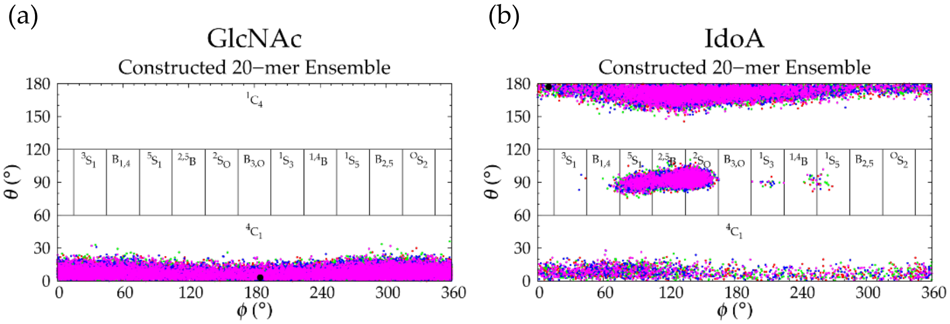


Figure B.34. C–P data for (a) GlcNAc and (b) IdoA in the constructed nonsulfated heparan 20-mer ensemble; each of the 4 runs is represented by different color and the force-field geometry is represented by a single large black dot; each run contains 10,000 parameter sets.

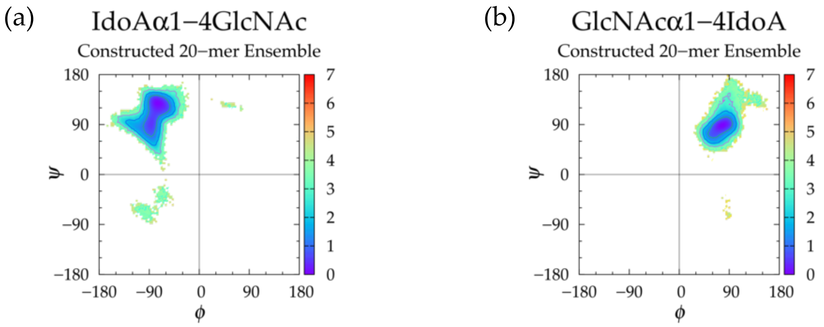


Figure B.35. ΔG(ϕ, ψ) in the constructed nonsulfated heparan 20-mer ensemble for aggregated (a) IdoAα1-4GlcNAc and (b) GlcNAcα1-4IdoA glycosidic linkage data; contour lines every 1 kcal/mol.

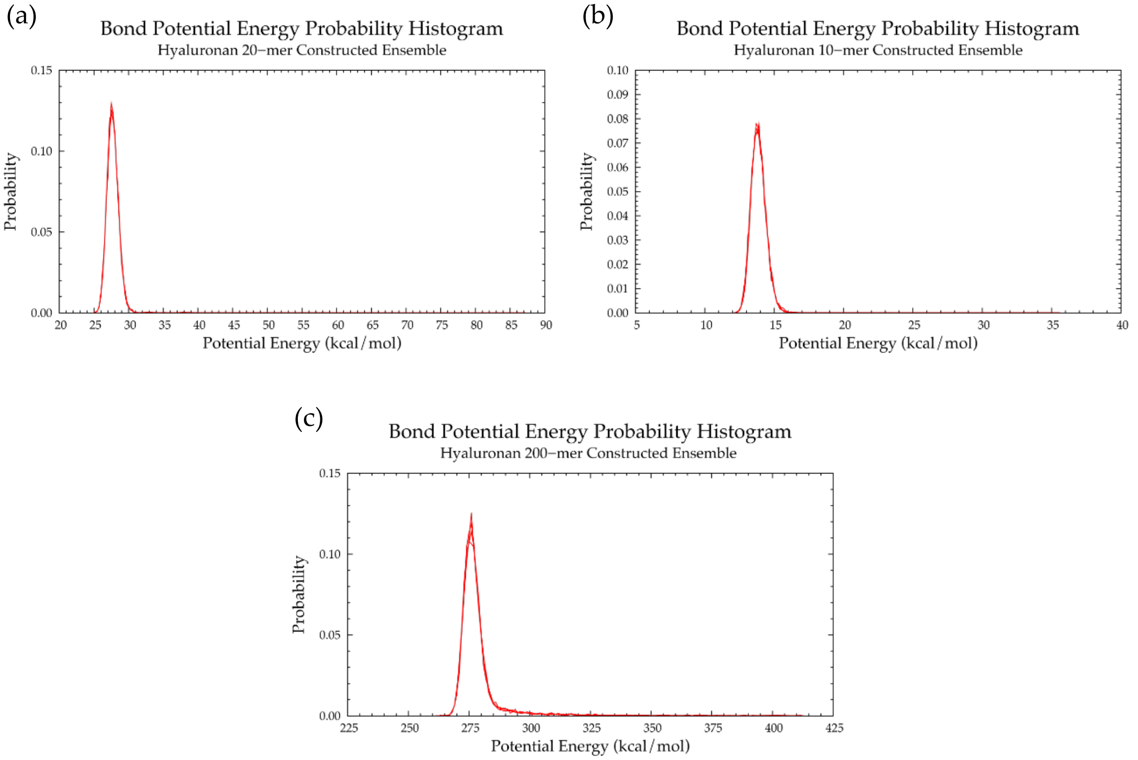


Figure B.36. Bond potential energy probability distributions from constructed ensembles of hyaluronan (a) 20-mer (cutoff = 128.5 kcal/mol), (b) 10-mer (cutoff = 112.8 kcal/mol), and (c) 200-mer (cutoff = 412.2 kcal/mol).

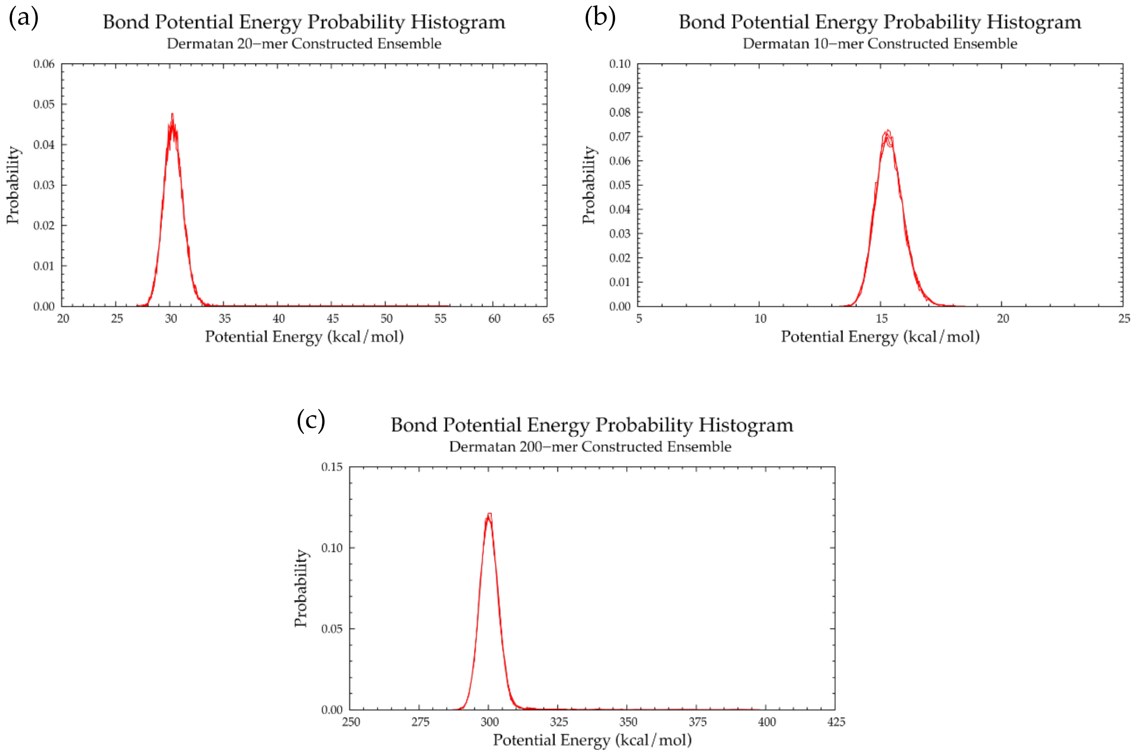


Figure B.37. Bond potential energy probability distributions from constructed ensembles of nonsulfated dermatan (a) 20-mer (cutoff = 131.9 kcal/mol), (b) 10-mer (cutoff = 117.5 kcal/mol), and (c) 200-mer (cutoff = 397.7 kcal/mol).

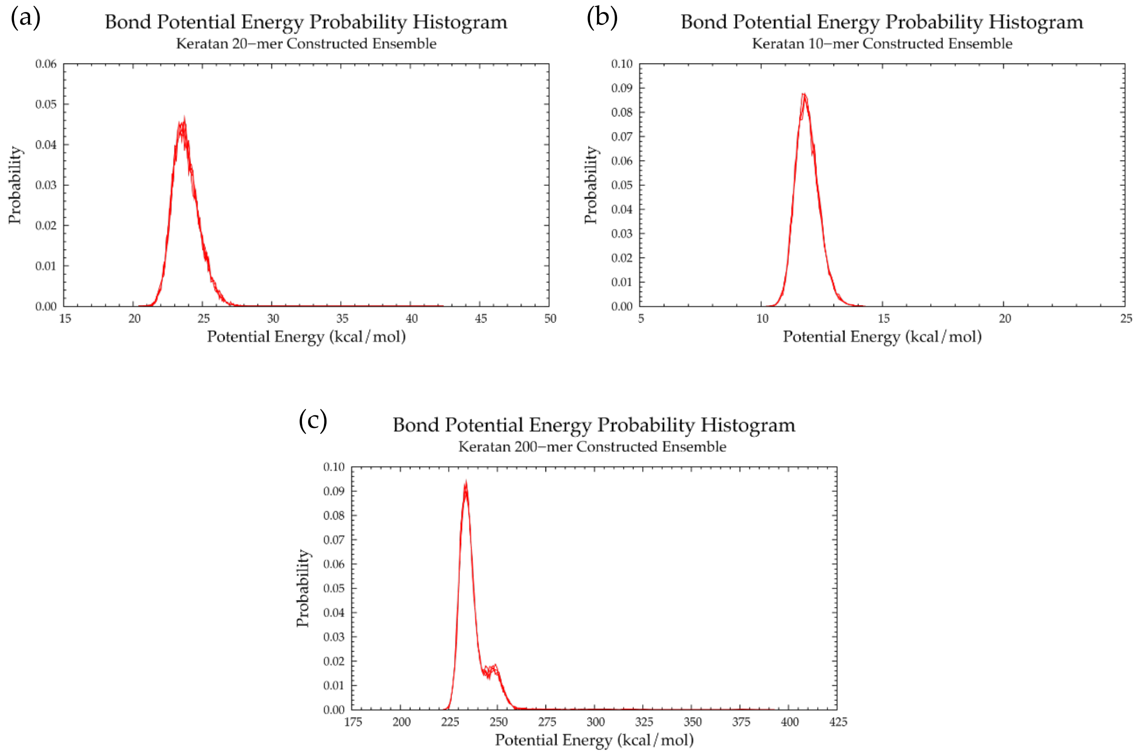


Figure B.38. Bond potential energy probability distributions from constructed ensembles of nonsulfated keratan (a) 20-mer (cutoff = 126.3 kcal/mol), (b) 10-mer (cutoff = 111.5 kcal/mol), and (c) 200-mer (cutoff = 397.3 kcal/mol).

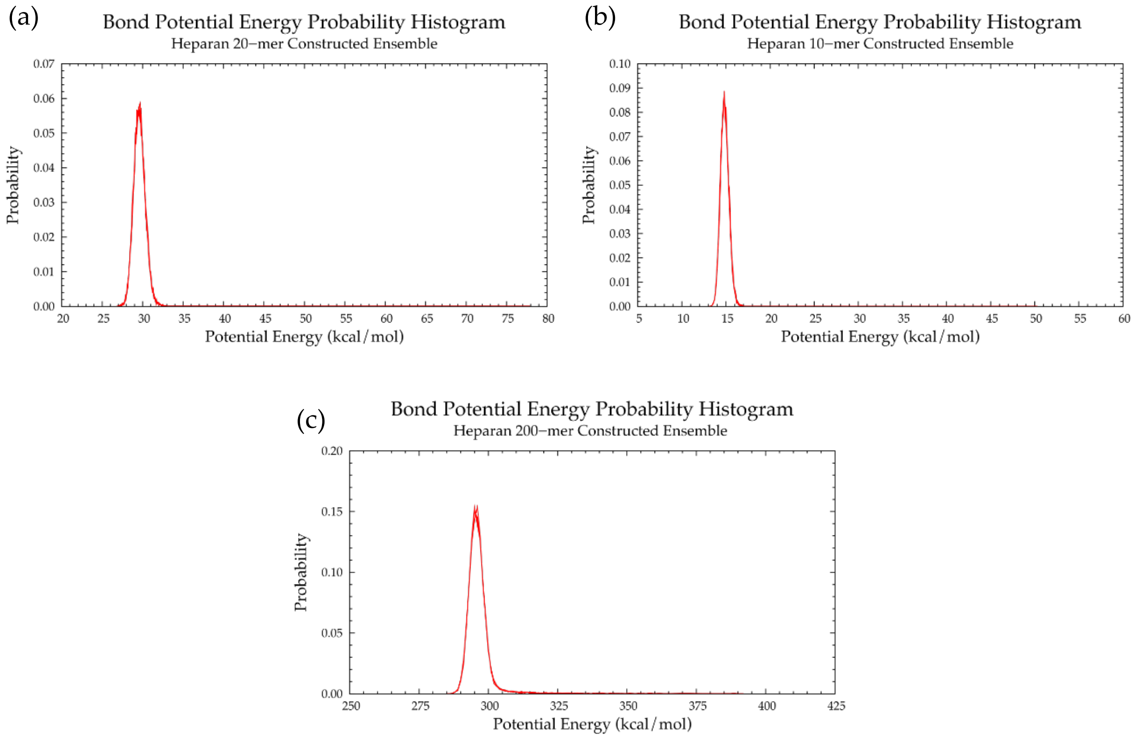


Figure B.39. Bond potential energy probability distributions from constructed ensembles of nonsulfated heparan (a) 20-mer (cutoff = 130.3 kcal/mol), (b) 10-mer (cutoff = 115.8 kcal/mol), and (c) 200-mer (cutoff = 391.5 kcal/mol).

BIOGRAPHY OF THE AUTHOR

Elizabeth K. Whitmore was born in Portland, Maine on 31 May 1993. She was raised in Kennebunk, Maine and graduated from Kennebunk High School in 2011. She attended the University of New England and graduated in 2014 with a Bachelor of Science degree in Medical Biology. She then entered the Graduate School of Biomedical Science and Engineering at The University of Maine in the Fall of 2015. After receiving her degree, Elizabeth will be joining Early Charm Ventures, a venture studio that takes scientific discoveries made at universities and turns them into companies, to begin her career in the fields of business and Biotechnology. Elizabeth is a candidate for the Doctor of Philosophy degree in Biomedical Sciences from the University of Maine in May 2021.