



THE UNIVERSITY OF QUEENSLAND  
A U S T R A L I A

# Development of Computational Tools for the Rational Design of Glycosaminoglycan Mimetics

by Eric David Boittier

Supervisor: Associate Professor Vito Ferro

Co-supervisor: Doctor Neha S. Gandhi

*A Research report submitted for the degree of Bachelor of Advanced Science (Honours) at  
The University of Queensland in May 2018  
School of Chemistry and Molecular Biosciences*

---

**Declaration by the author**

This research report is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my report.

I have clearly stated the contribution of others to my research report as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my report. The content of my report is the result of work I have carried out since the commencement of my honours research project and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my research report, if any, have been submitted to qualify for another award.

I acknowledge that copyright of all material contained in my research report resides with the copyright holder(s) of that material.

**Statement of Contributions to Jointly Authored Works Contained in the Research Report**

No jointly-authored works.

**Statement of Contributions by Others to the Research Report as a Whole**

No contributions by others.

**Statement of Parts of the Research Report or Submitted to Qualify for the Award of Another Degree**

None

**Published Works by the Author Incorporated into the Research**

None

**Additional Published Works by the Author Relevant to the Research Report or but not Forming Part of it**

Pathway Bifurcation in the (4 + 3)/(5 + 2)-Cycloaddition of Butadiene and Oxidopyrylium Ylides: The Significance of Molecular Orbital Isosymmetry Jed M. Burns and Eric D. Boittier The Journal of Organic Chemistry Article ASAP DOI: 10.1021/acs.joc.8b03236

**Signature of Author:** \_\_\_\_\_ **Date:** \_\_\_\_\_

**Principal Supervisor Agreement**

I have read the final report and agree with the student's declaration.

**Signature of Principal Supervisor:** \_\_\_\_\_ **Date:** \_\_\_\_\_

# Acknowledgements

# Table of Contents

<b>Acknowledgements</b>	i
<b>List of Figures</b>	iii
<b>List of Tables</b>	iii
<b>Acronyms</b>	iv
<b>Abstract</b>	1
<b>1 Computer Aided Drug Design &amp; Glycosaminoglycans</b>	2
1.1 Adrift in Chemical Space . . . . .	2
1.2 Glycosaminoglycans in Biology and Medicine . . . . .	2
1.2.1 Structure and Function . . . . .	2
1.2.2 Case Study 1: Antithrombin and Fondarinux . . . . .	2
1.2.3 Case Study 2: Heparanase and PG545 . . . . .	2
1.3 Computational Chemistry (in a Nutshell) . . . . .	3
<b>2 Methods Development in Protein-Ligand Docking</b>	4
2.1 Theory . . . . .	4
2.1.1 Search Functions . . . . .	4
2.1.2 Scoring Functions . . . . .	4
2.2 Research Aims . . . . .	5
<b>3 Data Mining the Protein Data Bank for Glycosaminoglycan Interactions</b>	6
3.1 Methods . . . . .	6
3.1.1 The GAG70 Dataset: Creation and Curation . . . . .	6
3.1.2 Carbohydrate Structure Analysis . . . . .	6
3.1.3 Intermolecular Interactions . . . . .	7
3.2 Results and Discussion . . . . .	7
<b>4 Development of a Glycosaminoglycan Specific Scoring Function</b>	11
4.1 Glycosidic Torsion Energy Penalty Functions . . . . .	11
4.1.1 Methods . . . . .	11
4.1.2 Results and Discussion . . . . .	12
4.2 Saltbridge Energy Scoring Function . . . . .	13
4.2.1 Methods . . . . .	13

4.2.2	Results and Discussion . . . . .	13
<b>5</b>	<b>Benchmarking Docking Software for Glycosaminoglycans</b>	<b>15</b>
5.1	Methods . . . . .	15
5.1.1	Docking . . . . .	15
5.1.2	Analysis . . . . .	15
5.2	Results and Discussion . . . . .	15
5.2.1	Comparing the Performance of Autodock Vina and Vina-Carb . . . . .	15
<b>6</b>	<b>GlycoTorch Web Server</b>	<b>17</b>
<b>7</b>	<b>Appendix A</b>	<b>18</b>

## List of Figures

2.1	A suggested paradigm for scientific programming. . . . .	5
3.1	. . . . .	7
3.2	. . . . .	7
3.3	. . . . .	8
3.4	. . . . .	8
3.5	. . . . .	8
3.6	. . . . .	9
3.7	. . . . .	9
3.8	. . . . .	9
3.9	. . . . .	10
4.1	. . . . .	12
4.2	. . . . .	13
4.3	. . . . .	14
4.4	. . . . .	14
5.1	Caption . . . . .	16

## List of Tables

# Acronyms

**ADV** Autodock Vina

**GAG** glycosaminoglycan

**MD** molecular dynamics

**PDB** Protein Data Bank

**QM** quantum mechanics

**RMSD** root-mean-squared deviation

**VC** Vina–Carb

# **Abstract**

# 1 | Computer Aided Drug Design & Glycosaminoglycans

## 1.1 Adrift in Chemical Space

## 1.2 Glycosaminoglycans in Biology and Medicine

### 1.2.1 Structure and Function

### 1.2.2 Case Study 1: Antithrombin and Fondarinux

### 1.2.3 Case Study 2: Heparanase and PG545

### 1.3 Computational Chemistry (in a Nutshell)

# 2 | Methods Development in Protein-Ligand Docking

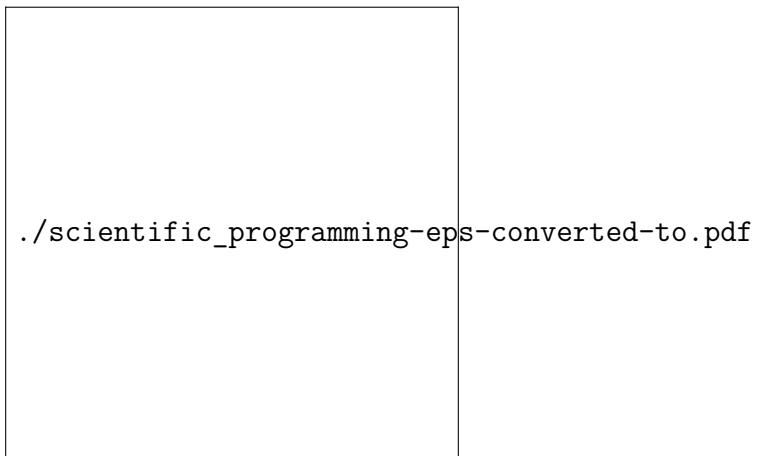
## 2.1 Theory

### 2.1.1 Search Functions

### 2.1.2 Scoring Functions

## 2.2 Research Aims

The aim of this project was to increase the accuracy of *Vina-Carb* when predicting the structure of GAG-Protein complexes by optimizing and developing new and existing scoring functions. Alongside this aim, analyses of crystallographic data of GAG-Protein complexes available in the Protein Data Bank (PDB) were undertaken. Towards these goals, bespoke analytical tools necessary for this task were developed from new and existing software. This code was made entirely open-source. Tools from this code base considered particularly serviceable to the computational glycoscience community were hosted online as the Web-accessible server, *GlycoTorch*, which aims to ‘shine a light on carbohydrate–protein interactions’.



**Figure 2.1:** A suggested paradigm for scientific programming.

# 3 | Data Mining the Protein Data Bank for Glycosaminoglycan Interactions

## 3.1 Methods

### 3.1.1 The GAG70 Dataset: Creation and Curation

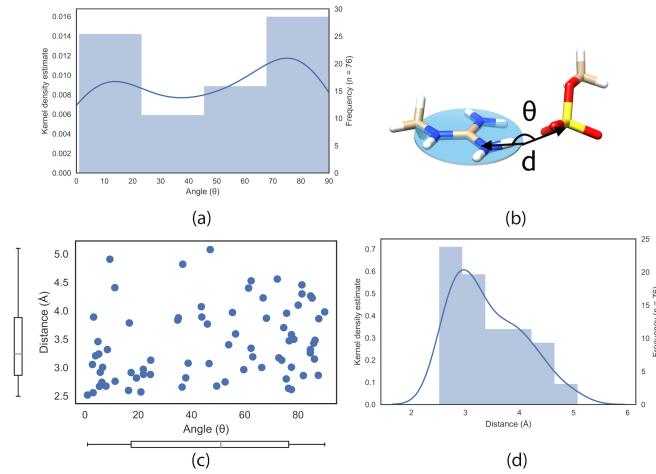
Crystal structures in the PDB containing glycosaminoglycans were identified. Previous work by Samsonov *et. al.* provided 84 structures.<sup>Samsonov2016</sup> From this dataset, the NMR structures (PDB codes: 2LVZ, 1HPN) were discarded. Structures from PDB codes 5W1O and 5DNF were also

### 3.1.2 Carbohydrate Structure Analysis

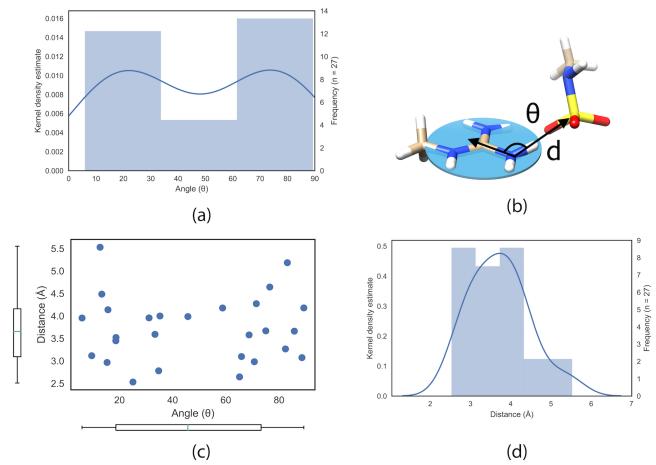
#### Ring Conformations

For N-membered rings, crystallographic, NMR, molecular dynamics (MD) or quantum mechanics (QM) theoretical mechanistic studies, among other fields, all require a standardized method of describing ring conformations. Historically, for rings composed of 6 atoms, six classical ring shapes are commonly assigned: chair (C), skew boat (S), boat (B), twist (T), envelope (E) and half boat (H). For sugars, depending on which atoms occupy the apex and/or bottom of the structure, the atom name is given with a superscript or subscript, respectively. This provides the basis of the 38 IUPAC standard ring conformations, which include, for example, the two most common and the typically low energy conformers,  $^4C_1$  and  $^1C_4$ . These assignments are discrete and easy to visualize; however, they are highly idealized and unsuitable for describing dynamic structures, among other tasks. For this reason, a mathematically “continuous” description of ring conformations, ideally as system that can compliment the standard IUPAC descriptions, is also required.

#### Glycosidic Torsions



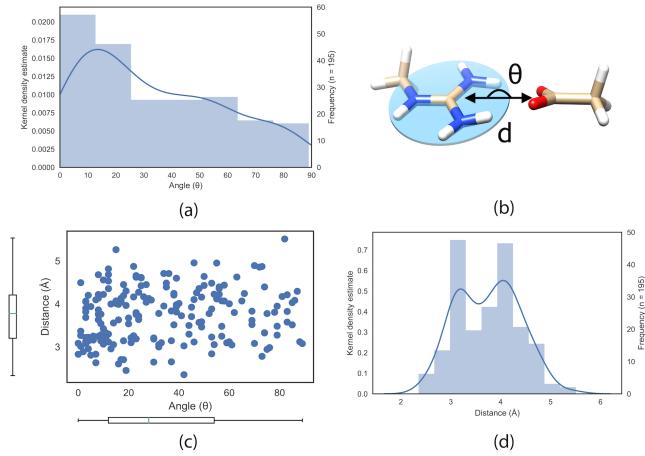
**Figure 3.1**



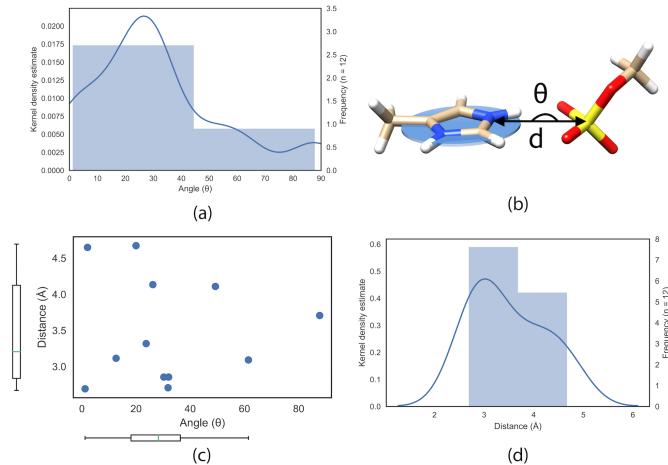
**Figure 3.2**

### 3.1.3 Intermolecular Interactions

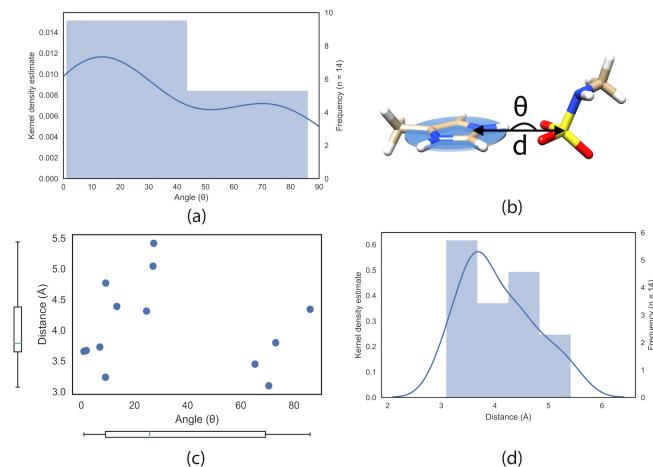
## 3.2 Results and Discussion



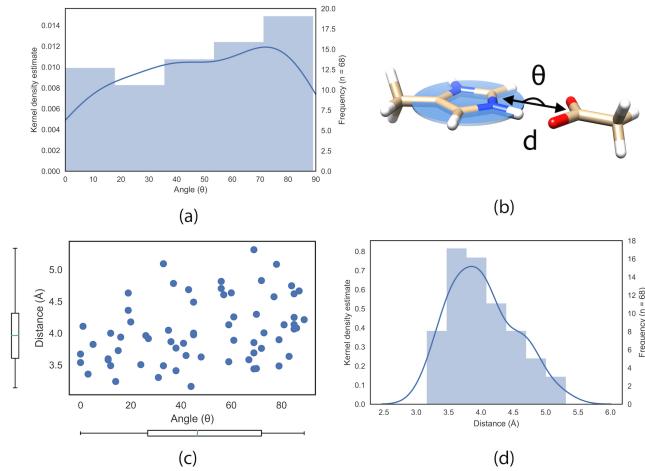
**Figure 3.3**



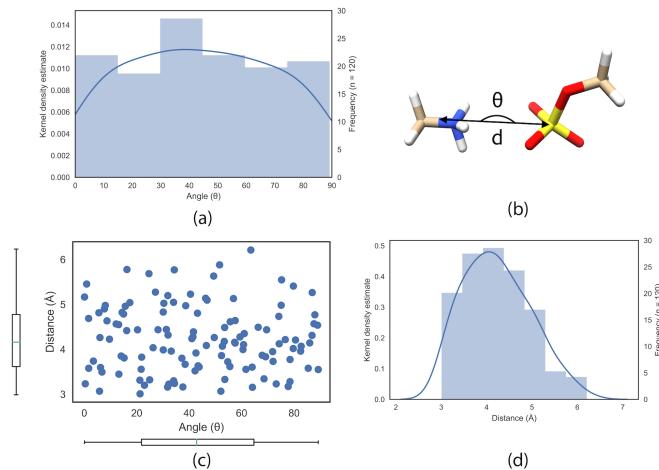
**Figure 3.4**



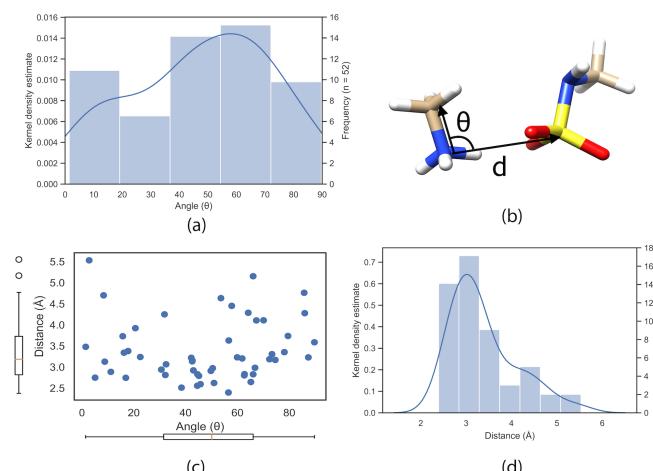
**Figure 3.5**



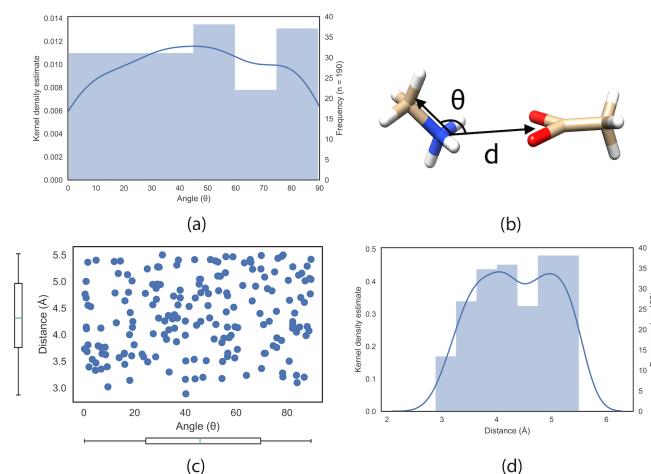
**Figure 3.6**



**Figure 3.7**



**Figure 3.8**



**Figure 3.9**

# 4 | Development of a Glycosaminoglycan Specific Scoring Function

## 4.1 Glycosidic Torsion Energy Penalty Functions

### 4.1.1 Methods

#### Preparing the Model Structures

Glycosidic linkages between disaccharides were modelled using two molecules of tetrahydropyran (THP). THP has been used extensively as a minimal model system for investigating the rotational parameters of the glycosidic torsion. Two reasons for this practice are commonly cited: (1) it is less computationally expensive than using the entire disaccharide structure (THP only requires a fraction of the basis functions for QM calculations compared to the full structure) and (2) it disambiguates other interaction energies (e.g. intramolecular hydrogen bonds between hydroxyl groups) from the energy reported. In regards to the latter, it is argued that, even without the confounding energies from intramolecular hydrogen bonding, the electronic free energy of the glycosidic torsion is undoubtedly dependent on the substituents of the ring and their orientation in space. However, while this model may be an oversimplification, its predictive power to locate ranges of low energy glycosidic torsions has been demonstrated.<sup>Nivedha2015</sup>

Starting ring conformations and linkages of the THP molecules were taken from exemplary geometries from crystal structures in the PDB. All other substituents were removed and replaced with hydrogens.

#### Quantum Mechanics Calculations

These structures were minimized at the HF/6-31+G(d) level of theory. The dihedrals  $\varphi$  and  $\psi$  were scanned using Gaussian 16's *opt=ModRedundant* keyword, in increments of  $15^\circ$  over 24 steps. Frequency and optimization calculations on each of the minimized structures across the scans was performed at the M062X functional and 6-311++G(2d,2p) basis set. The SCRF solvent model for water was also included. During optimization, the glycosidic torsional angle was kept fixed, which allowed for a relaxed potential energy scan.

## 4 Development of a Glycosaminoglycan Specific Scoring Function

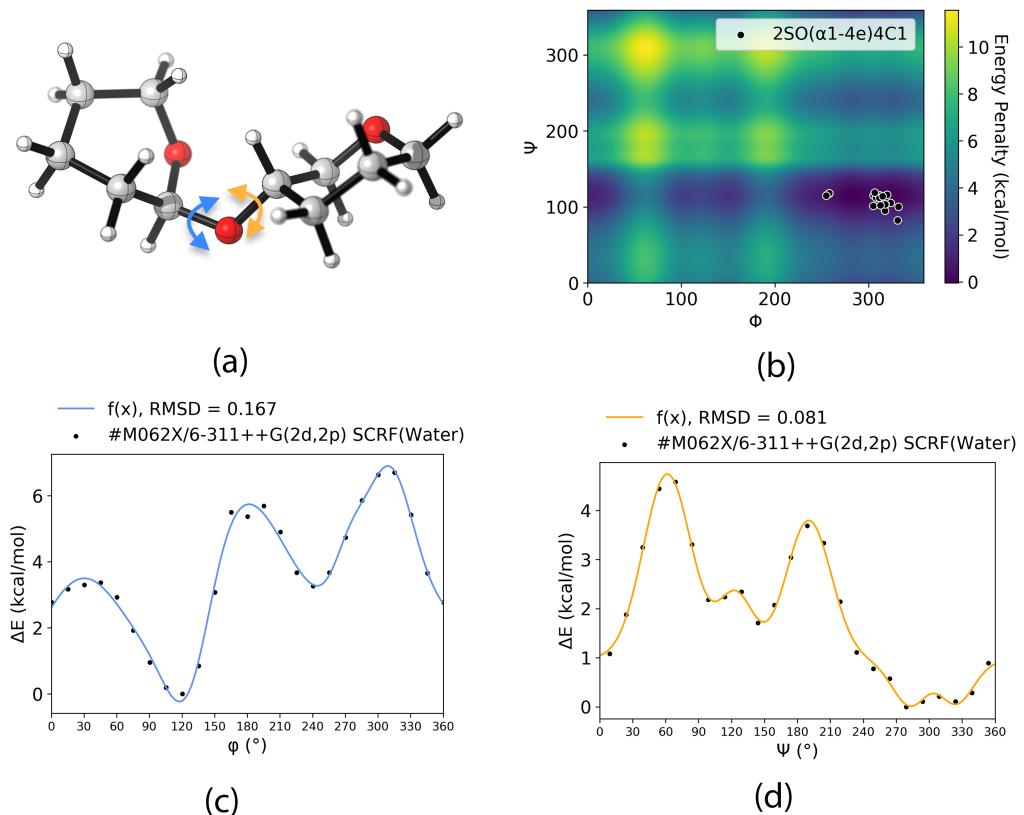
### Fitting the Model

For each glycosidic torsional energy profile, a series of Gaussian curves (equation 4.1) were used to model the data, where  $x$  is the glycosidic torsion and  $a_i$ ,  $b_i$ , and  $c_i$  correspond to the magnitude, width, and mid-point of each individual Gaussian distribution, respectively.

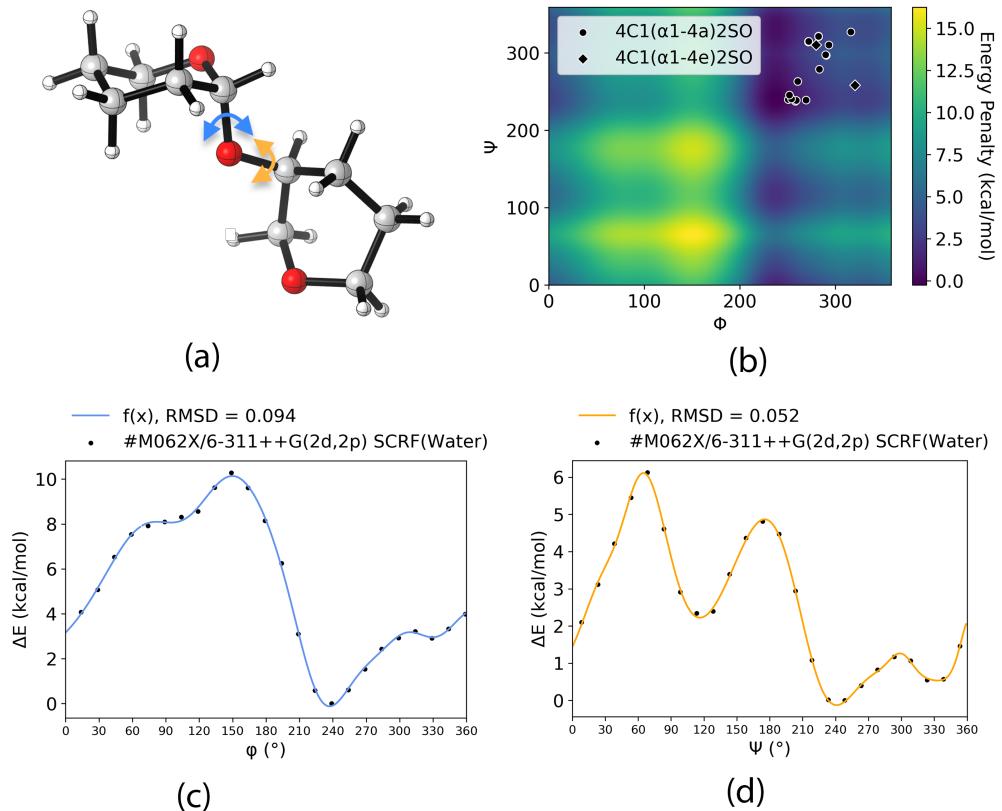
(4.1)

This function was fit to the data using a least squares optimization, available in the Scipy library in the Python programming language, with use of the Levenberg-Marquardt Algorithm.

### 4.1.2 Results and Discussion



**Figure 4.1**

**Figure 4.2**

## 4.2 Saltbridge Energy Scoring Function

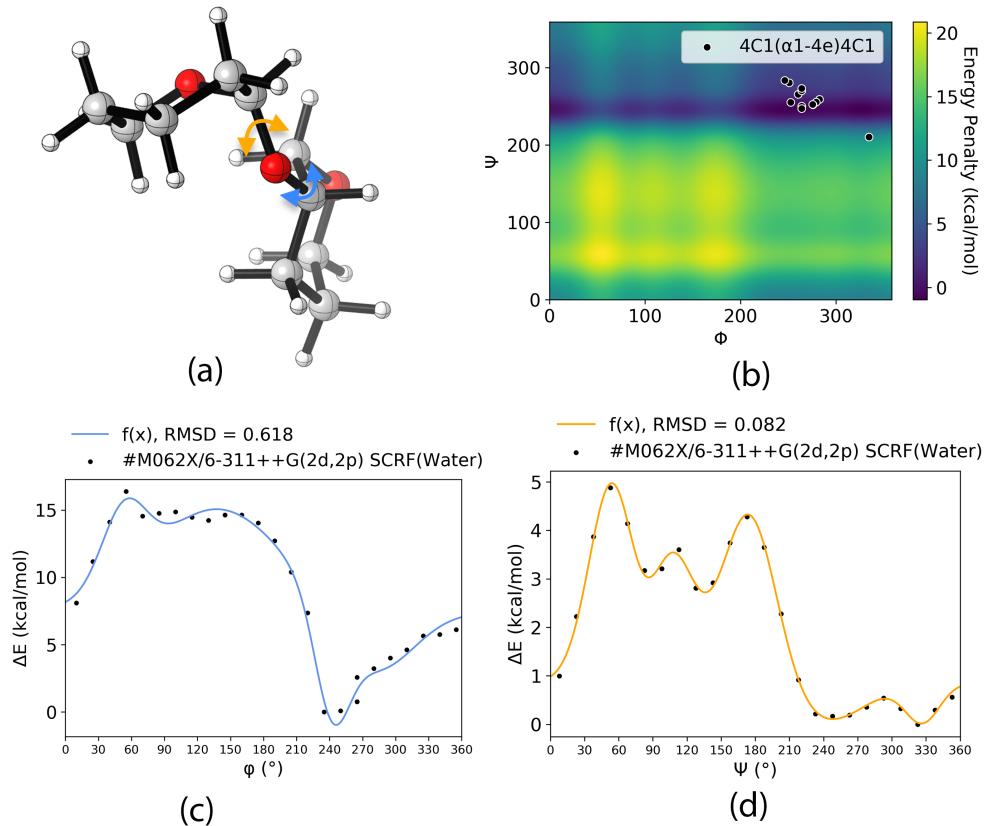
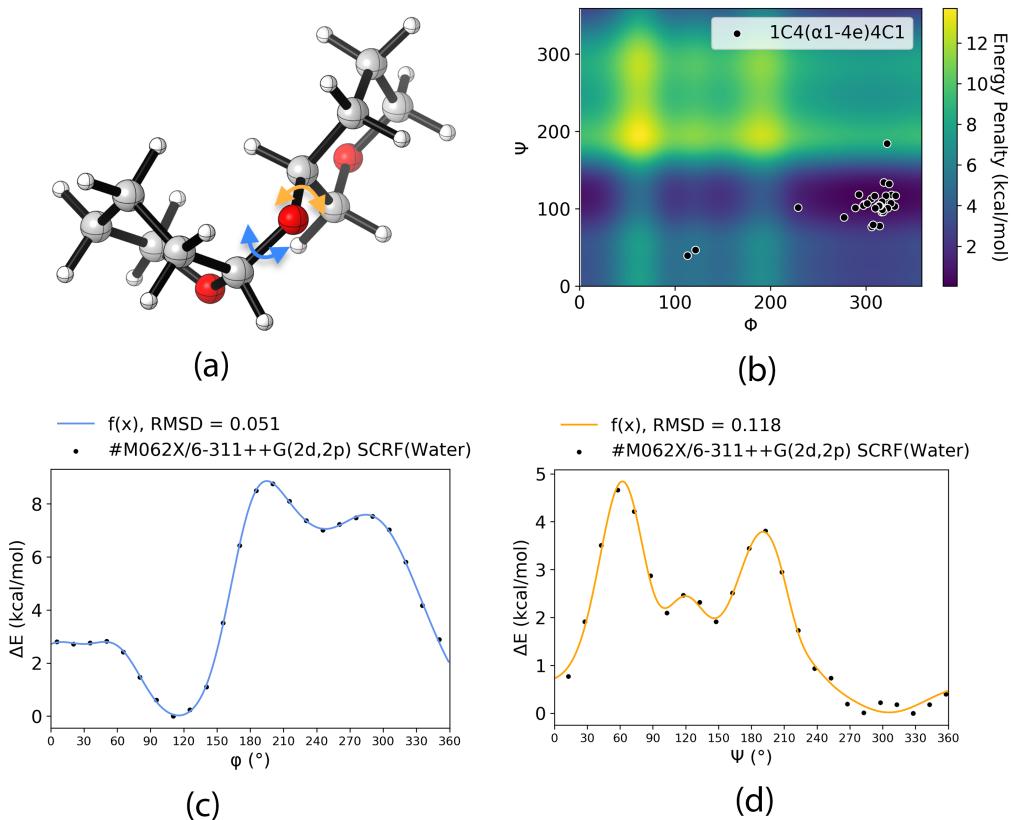
### 4.2.1 Methods

Preparing the model structures

Quantum mechanics calculations

Fitting the Model

### 4.2.2 Results and Discussion

**Figure 4.3****Figure 4.4**

# 5 | Benchmarking Docking Software for Glycosaminoglycans

## 5.1 Methods

### 5.1.1 Docking

**Structure Preparation**

**Docking Protocol**

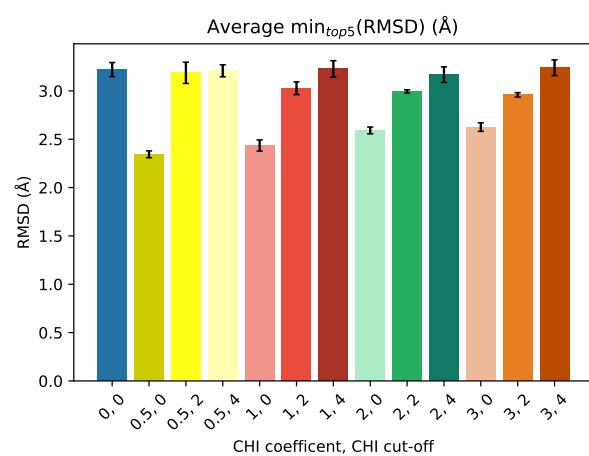
*Statistical Considerations*

In an effort to determine a statistically significant sample size, an initial trial of docking, using the default Vina–Carb (VC) parameters ( $CHI\ cutoff = 0$ ,  $CHI\ coeff = 1$ ) for a subset of structures containing sugars in a  $^1C_4$  or  $^4C_1$  ring conformation, exclusively. Docking was repeated 36 times, with Autodock Vina (ADV) parameters *exhaustiveness* and *energy range* set to 80 and 12, respectively. The top five docked poses produced by VC were scored by root-mean-squared deviation (RMSD) from the native crystal pose. From these docking calculations, a mean standard error ( $E$ ) of 0.32 RMSD and mean standard deviation ( $\sigma$ ) of 0.21 RMSD were observed. To produce a statistically significant result (i.e.  $p < 0.5$ ,  $Z = 1.96$ ), the minimum sample size required can be calculated by:

### 5.1.2 Analysis

## 5.2 Results and Discussion

### 5.2.1 Comparing the Performance of Autodock Vina and Vina-Carb



**Figure 5.1:** Caption

## 6 | GlycoTorch Web Server

## 7 | Appendix A

$$f(x) = 13.3370e^{-\frac{(x-31.3775)^2}{16552.6689}} 1.3170e^{-\frac{(x-135.6499)^2}{391.0353}} + 2.8463e^{-\frac{(x-160.7422)^2}{226.0367}} \\ + 0.5915e^{-\frac{(x-314.4487)^2}{328.4058}} + 8.8900e^{-\frac{(x-189.8429)^2}{1773.6483}} + 16.0838e^{-\frac{(x-305.3791)^2}{8985.1546}} \\ + -9.9028 \quad 0 \leq x < 360 \quad (7.1)$$