Molecular Mechanism of Amyloid Inhibition By Inositol

by

Grace Li

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Biochemistry
University of Toronto

© Copyright 2014 by Grace Li
Abstract

Molecular Mechanism of Amyloid Inhibition By Inositol

Grace Li

Doctor of Philosophy

Graduate Department of Biochemistry

University of Toronto

2014

Alzheimer’s disease (AD) is a progressive neurodegenerative disease characterized pathologically by the presence of extracellular fibrillar deposits of β-amyloid (Aβ), a 38 to 42 residue protein. Scyllo-inositol is a promising potential therapeutic compound for AD treatment that blocks the accumulation of Aβ oligomeric assemblies and reduce AD-like symptoms in a mouse model of AD. In vitro, scyllo-inositol and its stereoisomer chiro-inositol prevent the formation of Aβ(1-42) fibrils in a stereochemistry-dependent manner.

Understanding the molecular basis of the effect of inositol on Aβ aggregation will aid in the development of inhibitors of amyloid aggregation. Aβ amyloid formation follows a complex aggregation pathway, where intermediate aggregate species are implicated in the disease. To this end, extensive molecular dynamics simulations of scyllo-inositol and chiro-inositol were carried out to characterize the binding of inositol with different self-aggregated peptide states: single peptides, small aggregates, and ordered β-sheet aggregates.

First, I characterized the binding equilibria of inositol with model amyloid peptides, alanine dipeptide and (Gly-Ala)₄. Inositol binds predominantly to the backbone with affinities in the millimolar to molar range, indicating that backbone binding is unlikely to inhibit amyloid fibrillation. Next, I characterized the binding mechanism of inositol to monomer and aggregates of Aβ(16-22). Ordered β-sheet-like aggregates of Aβ were
identified as binding partners of inositol, where scyllo-inositol displays higher binding specificity than chiro-inositol for the grooves at the surface of protofibrillar oligomers. Finally, I characterized the binding of inositol to a protofibril of Aβ(1-42). Scyllo-inositol displays the highest binding specificity for the residues in the central hydrophobic core of Aβ42. Together, these results suggest a molecular mechanism where scyllo-inositol inhibits Aβ amyloid formation by coating the surface of protofibrillar aggregates and disrupting their lateral stacking into fibrils.

In Chapter 6, extensive simulations of PgaB, a key protein in the export of polysaccharides in the formation of bacterial biofilm, were conducted successively in the presence of two monosaccharides that are components of the polymeric substrate, PNAG. Understanding the molecular basis of PgaB-PNAG binding is required for the rational design of inhibitors of biofilm production by bacteria. Here, the putative binding modes and sites of PNAG predicted, in combination with experimental studies, led to a mechanism for the export of PNAG by PgaB. Significantly, the results here demonstrate that the central methodology of this thesis is broadly applicable for characterizing protein-carbohydrate binding.
Dedication

I dedicate this thesis to my mom and grandparents.
Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Régis Pomès, for his insight and guidance. My journey in scientific research begun in his lab in 2003 where I was a summer student. I am grateful that I became older, wiser, and more confident in his lab under his guidance.

I thank my committee members Drs. JoAnne McLaurin, Mark Nitz and Avi Chakrabartty for their advice and for pushing me forward.

I thank the members of my lab, both past and present, for helping me with my work and for providing extensive moral support. Specifically, I would like to thank the early Ph.D. students of the lab, Drs. Chris Madill, Chris Neale, Sarah Rauscher, Tomas Rodinger and Marty Kurylowicz for setting examples of excellence. I especially thank Dr. Nilu Chakrabarti, Dr. John Holyoake, Dr. Chris Neale, David Caplan, Dr. Loan Huynh, Kethika K., and Aditi R. for their help, advice and support in life and in the lab. All of you have contributed substantially in helping me become more mature.

I give my sincerest thanks to Mr. Larry Zimmerman, my high school math teacher and math team coach, and Dr. Nasir Memon for inspiring me and instilling in me a love of mathematics and computer science at a crucial age. You’ve also taught me the appropriate amount of discipline needed to study and master subjects in these fields.

I thank my family and friends both far and near for their support and encouragement (a.k.a had the patience to endure my venting over the years). A special thanks to Suraj C. - you met me at the end of this very important journey, but you’ve been one of the biggest source of encouragement and support thus far. Thank you Ken, Omar and Yelena for all the good times I’ve had whenever I’m back in NYC and for your support and encouragement during stressful times.

Finally, I give my deepest thanks to my mom and my grandparents for their unconditional love, support, encouragement, and understanding.
Contents

List of Symbols and Acronyms xv

1 Introduction 1

1.1 The amyloid state of proteins .................................. 3
  1.1.1 Fibrils ................................................. 4
  1.1.2 Non-fibrillar oligomers ................................. 8

1.2 Amyloid Involvement in Diseases ............................. 9

1.3 Alzheimer’s Disease ............................................ 10

1.4 Amyloid Inhibition by Small Molecules: a promising method of treatment for AD ............................................. 13
  1.4.1 Dye-based molecules .................................... 14
  1.4.2 Polyphenols ............................................. 15
  1.4.3 Inositol ............................................... 16

1.5 Protein-ligand binding equilibria ............................. 18

1.6 Intermolecular forces in biomolecular interactions ........ 20

1.7 Protein-carbohydrate interactions ........................... 22

1.8 Structure-based rational drug design ........................ 25

1.9 Molecular dynamics simulations ............................. 25
  1.9.1 Challenges and limitations of MD simulations .......... 26
1.10 Recent progress in elucidating the molecular mechanism of small-molecule binders and inhibitors of amyloidogenic species .................................. 27

1.11 Thesis objectives and organization ............................................. 33
  1.11.1 Organization of thesis ...................................................... 36

Bibliography .................................................................................. 36

2 Methods ..................................................................................... 53
  2.1 Molecular Mechanics ............................................................... 53
  2.2 Force Fields for Biomolecules ................................................... 54
    2.2.1 Bonded Interactions ......................................................... 55
    2.2.2 Non-bonded Interactions .................................................... 56
    2.2.3 Electronic Polarizability .................................................... 58
    2.2.4 The AMBER and OPLS Force Fields ................................. 58
  2.3 Molecular Dynamics Simulation ............................................... 59
    2.3.1 Temperature and Pressure ................................................ 61
    2.3.2 Periodic boundary conditions .......................................... 61
    2.3.3 Connection to Statistical Mechanics ................................. 62
    2.3.4 Enhanced Sampling Techniques ....................................... 63
  2.4 Preparing a Biomolecular System for MD Simulation .................. 68

Bibliography .................................................................................. 69

3 Binding of Inositol Stereoisomers to Model Peptides ..................... 77
  3.1 Summary ................................................................................ 78
  3.2 Introduction ........................................................................... 78
  3.3 Methods ................................................................................ 82
    3.3.1 Simulation Parameters and Protocol ................................. 82
    3.3.2 Analysis Protocol ............................................................ 85
  3.4 Results .................................................................................. 86
5.2.1 Analysis Protocol ............................................ 145
5.3 Results and Discussion ........................................ 146
  5.3.1 Protofibrillar structure ................................... 146
  5.3.2 Ligand binding ........................................... 147
  5.3.3 Molecular basis of amyloid inhibition by scyllo-inositol 154
  5.3.4 Comparisons of binding to the protofibril of Aβ(16-22) ....... 155
5.4 Conclusions ..................................................... 156

Bibliography ......................................................... 156

6 MD simulations of PgaB-glucosamine binding .......................... 161
  6.1 Summary ...................................................... 162
  6.2 Introduction .................................................. 162
  6.3 Material and Methods .......................................... 163
  6.4 Results and Discussion ......................................... 166
    6.4.1 Protein Dynamics .......................................... 166
    6.4.2 Comparisons of the binding modes of GlcNAc and β-D-GlcNH₃⁺ .... 167
  6.5 Acknowledgements ............................................. 174

Bibliography ......................................................... 174

7 Conclusions and Future Directions ..................................... 178
  7.1 Conclusions .................................................. 178
  7.2 Future directions ............................................... 181
    7.2.1 Effect of inositol on the growth of amyloid fibrils .............. 181
    7.2.2 Effect of inositol on amyloid aggregation in the presence of lipid membranes .................................................. 182
    7.2.3 Effect of inositol on monomers of longer amyloidogenic peptides .... 182
    7.2.4 Possible experimental studies .................................. 183

Bibliography ......................................................... 183
Appendices 185

Appendix A Supplementary Information for Chapter 4 186

A.1 Monomer ......................................................... 187
A.2 Disordered Oligomer ........................................... 190
A.3 $\beta$-oligomers .................................................. 195
List of Tables

1.1 Amyloid-forming peptides and proteins and their associated diseases . . 9
3.1 Summary of simulation systems. ................................................. 83
3.2 Summary of equilibrium dissociation constants ($K_d$) for each system in the study. ................................................................. 86
4.1 Summary of simulation systems .................................................. 111
4.2 Summary of equilibrium constants ($K_{eq}$) and number of reversible binding events ($N_{binding}$) ................................................... 121
A.1.1 Fraction of inositol molecules (in %) bound to nonpolar and polar groups of the monomer. ......................................................... 188
A.2.1 Fraction of inositol molecules (in %) bound to nonpolar and polar groups of the disordered oligomer. ........................................... 190
A.3.1 Fraction of inositol molecules (in %) bound to nonpolar and polar groups of the $\beta$-oligomer. ...................................................... 195
# List of Figures

1.1 Characteristic cross-β spacings from X-ray fibre diffraction studies of amyloid fibrils ........................................ 5
1.2 EM images of amyloid fibrils and oligomers ........................................ 6
1.3 X-ray crystal structure of an amyloid fibril ........................................ 6
1.4 AD tissue pathology ................................................................. 11
1.5 APP processing ................................................................. 11
1.6 Amyloid-binding dyes ............................................................. 14
1.7 Polyphenol molecules ............................................................... 15
1.8 Inositol stereoisomers ............................................................... 17
1.9 Example of protein-carbohydrate binding ........................................ 22
1.10 A schematic of monosaccharide-protein binding mode ................. 24
1.11 Thesis study design and rationale ............................................. 35

2.1 Schematic diagram illustrating the energetic terms of the potential energy function ........................................ 55
2.2 An example snapshot of a MD simulation system ............................ 70

3.1 Inositol stereoisomers most commonly found in nature .................. 80
3.2 Binding of scyllo-inositol to the backbone of alanine dipeptide ....... 87
3.3 Binding of scyllo-inositol to the monomer of (GA)$_4$ .......................... 89
3.4 Binding of scyllo-inositol to the disordered oligomer of (GA)$_4$ ......... 90
3.5 Time evolution of peptide-peptide nonpolar and hydrogen bonding contacts in disordered aggregates in presence and absence of inositol (control). 91

3.6 Binding of scyllo- and chiro-inositol to the fibrillar aggregate of \((GA)\_4\). 93

3.7 Example snapshots of scyllo- and chiro-inositol binding to the fibril of \((GA)\_4\). 95

3.8 Binding mode and orientation of scyllo- and chiro-inositol to the fibril of \((GA)\_4\). 96

4.1 Binding of inositol to an \(A\_\beta(16-22)\) monomer 116

4.2 Binding of inositol to Glu and Phe dipeptides 119

4.3 Binding of inositol to a disordered oligomer of \(A\_\beta(16-22)\) 120

4.4 Inositol binding to a \(\beta\)-oligomer of \(A\_\beta16-22\) 124

4.5 Binding propensity of inositol to nonpolar and polar groups of the \(\beta\)-oligomer 125

5.1 Molecular structures of inositol stereoisomers, glucose, and glycerol 146

5.2 Protofibrillar morphology 147

5.3 Cartoon representation of the model of the protofibril of \(A\_\beta42\) 148

5.4 Spatial probability distribution of bound solutes 150

5.5 Binding contact map 151

5.6 Example binding modes of scyllo-inositol, chiro-inositol, and glucose to the protofibril 152

5.7 Clustered binding of solutes at the surface of the protofibril 153

6.1 Molecular structures of GlcNAc and GlcNH\textsubscript{2} 166

6.2 The structure of PgaB 167

6.3 RMSF of residues in PgaB 168

6.4 Example of glucosamine molecules binding in the C-terminal groove 169

6.5 Spatial probability densities of GlcNAc and \(\beta\)-D-GlcNH\textsubscript{3} in the C-terminal groove 170

6.6 Spatial probability densities of bound GlcNAc and \(\beta\)-D-GlcNH\textsubscript{3} 171
6.7 Ionic distribution ......................................................... 172

A.1.1 The end-to-end distance distribution of monomeric Aβ(16-22) conformations separated by the number of intra-peptide hydrogen bonds, \( n_{hb} \). . . . . . . . 187

A.1.2 Snapshots of representative monomer conformations at each peak of the end-to-end distribution. ................................................................. 188

A.1.3 Time-averaged number of hydrogen bonds and nonpolar made by inositol to each residue of the monomer at an inositol:peptide molar ratio of 2:1. . 189

A.2.1 The equilibrium number of intermolecular contacts per peptide in the disordered oligomer ................................................................. 191

A.2.2 Effect of inositol on the aggregation of Aβ(16-22) ............................................. 192

A.2.3 Time-evolution of peptide self-aggregation to form the disordered oligomer 193

A.2.4 Time-averaged number of hydrogen bonds and nonpolar made by inositol to each residue of the disordered oligomer ......................... 194

A.3.1 Initial model of the β-oligomer ................................................................. 196

A.3.2 Time-averaged number of hydrogen bonds and nonpolar made by inositol to each residue of the β-oligomer ................................. 197
List of Symbols and Acronyms

Å  Angstrom

$\epsilon_0$  permittivity of free space

$\mu_s$  microsecond

$K_d$  equilibrium drug binding constant

$(GA)_4$  a peptide with Glycine and Alanine repeated four times

$A\beta$  $\beta$-Amyloid

$A\beta(16-22)$  A fragment of the $A\beta$ peptide consisting of the residues Lys16, Leu17, Val18, Phe19, Phe20, Ala21, Glu22

$A\beta 40$  $\beta$-Amyloid peptide of 40 residues in length

$A\beta 42$  $\beta$-Amyloid peptide of 42 residues in length

AD  Alzheimer’s disease

ADP  alanine dipeptide

AFM  atomic force microscopy

BBB  blood-brain barrier

CHC  central hydrophobic core of $A\beta$
CNS  central nervous system
CR   Congo red

\text{dPNAG}  \text{de-N-acetylated poly-}\beta\text{-1,6-N-acetyl-D- glucosamine}

DR   distributed replica sampling
DRPE distributed replica potential energy
DSSP dictionary of protein secondary structure
EGCG -epigallocatechin-3-gallate
EM   electron microscopy
fs   femtosecond

GlcNAc \beta\text{-1,6-N-acetyl-D-glucosamine}
GROMACS Groningen machine for chemical simulations
K    Kelvin

MD   molecular dynamics
mM   millimolar

NFT  neurofibrillary tangles
nm   nanometers

NMR  nuclear magnetic resonance
NpT  isothermal-isobaric ensemble with moles, pressure and temperature held constant
ns   nanosecond
NVT  canonical ensemble with moles, volume and temperature held constant
OPLS-AA/L optimized potentials for liquid simulations-all atom / LMP2 calculations

PBC  periodic boundary condition

PiB  Pittsburgh compound B

PME  particle mesh ewald

PMF  potential of mean force

PNAG  poly-β-1,6-N-acetyl-D-glucosamine

SDS  sodium dodecyl sulfate

SSNMR  solid-state nuclear magnetic resonance

ST  simulated tempering

STDR  simulated tempering distributed replica sampling

T  temperature

ThT  thioflavin T

TMAO  trimethylamine N-oxide

VMD  Visual Molecular Dynamics
Chapter 1

Introduction

One of nature’s most remarkable phenomena is the ability of proteins to fold from linear polypeptide chains into structures which function as molecular machines that enable life.\textsuperscript{1,2} Proteins play a key role in many aspects of life. They act as structural scaffolds,\textsuperscript{3} catalyze biochemical reactions,\textsuperscript{4} regulate the cell cycle,\textsuperscript{5,6} and are crucial components of many signal transduction pathways.\textsuperscript{7}

Much of the critical regulation of biological activity within a cell is mediated by binding interactions between a protein receptor and a molecular ligand.\textsuperscript{8} For example, about 40\% of modern drugs target G protein-coupled receptors,\textsuperscript{8} a family of proteins found in eukaryotes that sense molecules outside of the cell and induce cellular responses by activating intracellular signal transduction pathways.\textsuperscript{9} Hence, it is not surprising that many diseases are caused by the improper functioning of proteins due to mutations, denaturation, or misfolding (failure to adopt their native functional state).\textsuperscript{2,10}

A seminal study performed by Anfinsen and colleagues demonstrated that the structure of a folded protein is encoded in its amino acid sequence and solvent environment.\textsuperscript{11} Since then, much progress has been made in understanding protein structure, function, and mechanism of folding.\textsuperscript{12} Structure determination techniques have gained much attention in the fields of biochemistry and biophysics, and have provided valuable insights
into macromolecular structure. However, protein structures determined from nuclear magnetic resonance (NMR) spectroscopy, X-ray crystallography, and homology modeling only provide a static picture, whereas molecular recognition and drug binding are dynamic processes. When a substrate approaches its receptor in solution, it encounters not a single, frozen structure, but rather a macromolecule that is in constant motion. Understanding both protein structure and dynamics enables us to elucidate the molecular basis of disease pathways, which will ultimately contribute to the discovery of novel therapeutics.

Because molecular motions associated with ligand binding are events which take place on the time scale of millionths of a second or less, current experimental techniques are not able to elucidate the complete atomistic energetics and mechanics of binding.\textsuperscript{13} In recent years, molecular dynamics (MD) simulation of biomolecules, a physics-based computer simulation technique, has become a tool of choice to investigate protein dynamics and function, and in particular ligand-binding.\textsuperscript{13,14} Currently, MD simulation is the most accurate computational method for probing small-molecule binding, and is useful for filling in the details where experimental methods cannot.\textsuperscript{14} In the past few years, computational power has increased to the point where all-atom MD simulation studies now routinely attain hundreds of nanoseconds of sampling time.\textsuperscript{15–18} With the ever increasing availability of computing power and data storage, simulations are a promising technique for aiding in the structure-based drug discovery process.\textsuperscript{14}

Not all proteins fold into a unique, compact native state. Intrinsically disordered proteins (IDPs), a class of proteins without a uniquely folded state, have recently gained attention because of their involvement in a multitude of physiological pathways and diseases.\textsuperscript{19} For example, proteins associated with cell signalling and cancer in humans are predicted to be enriched in protein disorder.\textsuperscript{20} A detailed review of disordered proteins and their roles in biology is beyond the scope of this thesis and is provided elsewhere.\textsuperscript{21,22} A number of peptides and proteins (some of which are also IDPs) are able to self-aggregate
to form amyloid, and are associated with incurable diseases such as prion disorders, neurodegenerative diseases, Type II diabetes, and systemic amyloidosis.\textsuperscript{23}

In the rest of this chapter, I first define and review the amyloid state of proteins from a biophysical point of view (Section 1.1-1.2). Following these sections, amyloid disorders is discussed with a central focus on Alzheimer’s disease (Section 1.3). Central to the results in chapters 2-4 of this thesis is scyllo-inositol, a small molecule that has been demonstrated to inhibit amyloid formation. In Section 1.4, I review the pharmacological properties of scyllo-inositol and, briefly, other small molecule inhibitors of amyloid formation \textit{in vitro}. I introduce the computational technique molecular dynamics simulations and the rationale for utilizing it to investigate amyloid inhibition. Lastly, I provide a review of molecular simulation studies of small-molecule inhibitors currently published in literature.

\section{1.1 The amyloid state of proteins}

Amyloids were discovered 150 years ago when tissue deposits of extracellular filaments were observed.\textsuperscript{24,25} These microscopically visible deposits were found on various organs in many seemingly unrelated diseases. Although numerous diseases involve amyloid formation of distinct aggregation-prone proteins or peptides, the ability to form amyloid is not only restricted to these disease-associated proteins. Amyloid fibrils may be formed from proteins that can also fold into well-defined tertiary structures (e.g. myoglobin and lysozyme), suggesting that the ability to form amyloid fibrils may be a generic property of polypeptides.\textsuperscript{23} However, the propensity for a given protein or peptide to form amyloid fibrils is highly dependent on the combination of solution conditions and peptide sequence. This is because, for a globular protein to adopt amyloid states, the protein must first be partly unfolded before conversion into amyloid fibrils is possible.\textsuperscript{23}

The pathway by which amyloid fibrils are formed \textit{in vivo} is not understood. Much
of what we know about amyloid formation currently comes from biochemical and biophysical analysis of synthetic amyloid-forming peptides \textit{in vitro}, which is thought to be analogous to the \textit{in vivo} pathway. Prior to the appearance of amyloid fibrils, a variety of intermediate species may be formed.\textsuperscript{23} Monomers self-assemble into oligomers of different morphologies and sizes, which exist in equilibrium with amyloid fibrils, a visible endpoint of aggregation.\textsuperscript{23}

Kinetically, the mechanism of amyloid formation is akin to those of nucleation-polymerization processes such as crystallization and micelle formation.\textsuperscript{26} During nucleation, a lag phase occurs in which the energetic barriers of aggregation must be overcome by the monomers to form the initial aggregation nucleus.\textsuperscript{26} Following this lag phase, free monomers may bind to the nucleated aggregates, which elongate into mature fibrils.\textsuperscript{26} Seeding, a process where preformed aggregates are introduced into the solution, eliminates the lag phase.\textsuperscript{27,28} In the following sections, the current biophysical and structural data on amyloid fibrils and non-fibrillar oligomers is reviewed, and implications for amyloid disease are discussed.

\subsection*{1.1.1 Fibrils}

Fibrillar amyloid deposits have several physical properties in common. Fibrils are protease resistant, and insoluble in the presence of the detergent sodium dodecyl sulfate.\textsuperscript{29} Importantly, they exhibit specific optical behavior when bound to certain dye molecules. After staining with Congo red, fibrils exhibit bright green birefringence under polarized light.\textsuperscript{30} However, the use of Congo red to detect the presence of amyloid formation is often a laborious process, and only provides a qualitative measurement of the amount of amyloid present.\textsuperscript{30} Thioflavin-T (ThT), a benzothiole fluorescent dye, is more commonly used to detect the presence of amyloid fibrils in post-mortem brain tissue samples, and to monitor fibril formation \textit{in vitro}. Upon binding to fibrils, ThT exhibits both a dramatic enhancement of its emission and a shift in the maximum of its excitation spectrum,
Amyloid fibrils formed from different polypeptides are thought to share a similar morphology known as the cross-\(\beta\) structure.\(^{23}\) To date, independent measurements of fibrillar structure from different instruments have all confirmed the cross-\(\beta\) structural core of amyloid fibrils. X-ray fiber diffraction studies showed that the diffraction pattern of fibrils is characterized by major orthogonal reflections along the meridional and equatorial directions, which correspond to a 4.8 \(\AA\) interpeptide separation, and a 10 \(\AA\) intersheet separation, respectively (Figure 1.1).\(^{25,32,33}\) The inter-peptide and inter-sheet separations are respectively parallel and perpendicular to the long-axis of the fibril. This diffraction pattern is now considered as indicative of the presence of cross-\(\beta\) structure, and hence, of amyloid fibrils.\(^{23}\) When the fibrils are stained, the macromolecular morphology of fibrils can be determined using the transmission electron microscope (TEM): fibrillar structures are long, unbranched, and ribbon-like structures with diameters between 50 - 100 nm (Figure 1.2).\(^{23}\)
Figure 1.2: (A) TEM image of negatively-stained mature amyloid fibrils. Copyright 2002 National Academy of Sciences, USA. (B) SSNMR model proposed by Petkova et al. EM images of oligomers of (C) Aβ42 and (D) Aβ40. (C) Copyright 2003 National Academy of Sciences, USA. (D) This research was originally published in Journal of Biological Chemistry. D M Walsh, D M Hartley, Y Kusumoto, Y Fezoui, M M Condron, A Lomakin, G B Benedek, D J Selkoe, and D B Teplow. Amyloid β-Protein Fibrillogenesis. Journal of Biological Chemistry. 1999; 274:25945-25952. © the American Society for Biochemistry and Molecular Biology.

Figure 1.3: A schematic of the X-ray crystal structure of fibrils formed from short amyloidogenic peptide fragments. Reprinted from Cell, 148, D. Eisenberg, M. Jucker, The Amyloid State of Proteins in Human Diseases, 1188 - 1203., Copyright 2012, with permission from Elsevier.
Chapter 1. Introduction

Advances in solid-state NMR (SSNMR) and X-ray crystallography in the last decade have elucidated the molecular details of amyloid fibrils. One of the first SSNMR models of an amyloid fibril was that of Aβ40, a peptide implicated in Alzheimer’s Disease.\textsuperscript{35} Its core fibril unit consists of a parallel in-register $\beta$-sheet, where each strand is a $\beta$-hairpin with peptide-peptide backbone hydrogen bonds running parallel to the long-axis of the fibril (Figure 1.2).\textsuperscript{34,35} In these models, residues 12 to 24 and 30 to 40 of the Aβ40 peptide were found to be in the $\beta$-sheet core of the fibril. Moreover, mutagenesis of residues that disrupts $\beta$-sheet formation in the region of the peptide spanning residues 17 to 23 (Leu-Val-Phe-Phe-Ala-Glu-Asp) can lead to the disruption of the fibrillation of Aβ, suggesting that the nonpolar residues Leu-Val-Phe-Phe-Ala constitute the central hydrophobic core of Aβ fibrils.\textsuperscript{37,38} Furthermore, smaller fragments of Aβ have been shown to form fibrils that are morphologically similar to those of the full length peptide. For example, Aβ(16-22) (or KLVFFAE) have been shown to form amyloid fibrils.\textsuperscript{39} SSNMR studies of the fibrils of the peptide Aβ(16-22) indicated that they are composed of stacked antiparallel $\beta$-sheets.\textsuperscript{39} Furthermore, fibrils of certain amyloid-forming peptide fragments formed crystals that were amenable to single crystal X-ray diffraction analysis.\textsuperscript{29} In agreement with SSNMR, the crystal structures of these fibrils revealed a structure composed of multiple layers of $\beta$-sheet with a dehydrated (“dry”) stacking interface (Figure 1.3).\textsuperscript{29,40}

The particular packing arrangement of polypeptides in amyloid fibrils can vary with changes in the experimental conditions under which the fibrils are formed. Specific structural polymorphisms include the length of the $\beta$-strands, side chain orientations and inter-protofilament packing.\textsuperscript{41} Fibril polymorphism may have important implications in amyloid diseases because different morphologies exhibit differing toxicities that depend on which residues are exposed at the surface. For example, \textit{in vitro}, quiescently formed fibrils of Aβ(1-40) have been shown to be more toxic than agitated fibrils.\textsuperscript{42}
1.1.2 Non-fibrillar oligomers

Because of their structural disorder and transient nature, it is difficult to obtain high-resolution structural details of amyloid oligomers using traditional structural determination techniques. Studies using low-resolution techniques transmission electron microscopy (TEM) and atomic force microscopy (AFM) have shown that transient, unstable particles may appear prior to the formation of fibrils (“on-pathway” oligomers). These protein aggregates are referred to in literature as amyloid protofibrils. Those that do not progress to form fibrils are considered off-pathway. However, off-pathway oligomers formed in the presence of detergents, lipids, and certain small molecules are typically not considered to be biologically-relevant.

Unlike fibrils, amyloid oligomers do not possess a generic structural element and instead, adopt a wide spectrum of sizes and morphologies. Size exclusion chromatography (SEC) studies of Aβ40 oligomers (isolated in vitro and from the brains of deceased individuals with Alzheimer’s Disease) revealed the existence of oligomers that ranged in size from dimers to large oligomers of hundreds of peptides. Oligomeric assemblies that are annular, spherical, or curvilinear in shape have been reported in literature.

Although there are large variations in morphologies, oligomers formed from different polypeptide sequences can display similar activities in cell metabolic assays. Importantly, many oligomers of different sizes share the ability to interact with a single oligomer-specific antibody. Several studies have indicated that oligomers may possess high β-sheet content. Moreover, some non-fibrillar oligomers may contain common structural elements: high-resolution structural studies of non-fibrillar oligomers of Aβ42 and prion-like peptides suggest that they may contain cross-β like fragments.
Table 1.1: Amyloid-forming peptides and proteins and their associated diseases

<table>
<thead>
<tr>
<th>Disease</th>
<th>Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer’s</td>
<td>Aβ40 and Aβ42</td>
</tr>
<tr>
<td>Parkinson’s</td>
<td>α-synuclein</td>
</tr>
<tr>
<td>Huntington’s</td>
<td>poly-glutamine</td>
</tr>
<tr>
<td>Amyotrophic lateral sclerosis</td>
<td>Superoxide dismutase 1</td>
</tr>
<tr>
<td>Spongiform encephalopathies</td>
<td>Prion protein or fragments thereof</td>
</tr>
<tr>
<td>Type II Diabetes</td>
<td>islet amyloid polypeptide (or Amylin)</td>
</tr>
</tbody>
</table>

1.2 Amyloid Involvement in Diseases

Because numerous diseases share the amyloid plaque pathology, fibrils were initially hypothesized to be the toxic species in these diseases. However, recent research has implicated non-fibrillar oligomers as the more likely toxic agent in the cause of several neurodegenerative diseases (such as Parkinson’s, Alzheimer’s, Huntington’s diseases and spongiform encephalopathies) and type II diabetes. Currently, the mechanism of toxicity of amyloid oligomers has not been determined, and is an area under intensive research. Oligomers formed from a variety of peptides, including those not implicated in amyloid disorders (e.g. lysozyme, β2-microglobulin, transthyretin) all exhibited toxicity, suggesting that the toxicity of amyloid oligomers may be independent of the peptide sequence. Experimental evidence widely supports the hypothesis that amyloid toxicity is based on a generic mechanism that involves the interactions of oligomers with cellular membranes. Specifically, it is hypothesized that oligomeric aggregates may ultimately induce cell death by interacting with and disrupting the integrity of the cellular membrane. Moreover, the aggregation of amyloidogenic peptides was found to occur more rapidly in the presence of membrane surfaces, leading to the claim that membrane-catalyzed fibril formation may induce cellular toxicity.
1.3 Alzheimer’s Disease

Alzheimer’s disease (AD) is a devastating neurodegenerative disease that is the most common cause of dementia in persons of age 65 or older. Upon examination, the post-mortem brains of AD patients show significant neuronal dystrophy. Pathologically, AD is characterized by the presence of extracellular deposits of senile plaques and neurofibrillary tangles (NFTs), both of which appear as lesions on stained neuronal tissue under visible-light microscopy (Figure 1.4).

In 1985, the amyloid-β protein or Aβ was identified as the largest component of these plaques. Monomeric Aβ is an approximately 4 kDa peptide produced by the intramembrane proteolytic cleavage of a larger protein, the amyloid-β precursor protein (APP). APP is sequentially processed by the aspartyl proteases β-secretase and γ-secretase, producing a pool of Aβ peptides of lengths varying from 38 to 43 residues depending on the position of the cleavage by γ-secretase (Figure 1.5). The peptides spanning residues 1-40 (Aβ40) or 1-42 (Aβ42) are predominantly found in AD-associated plaques. Although plaques contain different isoforms of the Aβ peptide, Aβ42 is likely to be the more deleterious form of Aβ. In vitro, Aβ42 displays significantly higher propensity for aggregation than Aβ40, despite differing by only two amino acids. Furthermore, genetic mutations within presenilin 1 and 2 (genes encoding enzymes that cleave APP to produce Aβ peptides) cause an aggressive early-onset form of AD which also lead to an increase in the ratio of Aβ42 to Aβ40 peptides produced.

Although it has been more than one hundred years since Alois Alzheimer first associated the presence of neuronal plaques with the clinical symptoms of Alzheimer’s disease, the exact relationship between the two is still under much contention. The ubiquitous presence of amyloid plaque deposits found in the brains of deceased dementia patients led to the formulation of the long-standing amyloid cascade hypothesis: the amyloidogenesis of Aβ plays a key role in the initiation of AD, which ultimately leads to the clinical symptoms of dementia. Genetic evidence provided strong support for the amyloid hy-
Figure 1.4: Lesions formed from amyloid plaques and NFT tangles in the cerebral cortex tissue of an AD brain. Reprinted from The Lancet, Vol. 368, Kaj Blennow; Mony J de Leon; Henrik Zetterberg, Alzheimer’s Disease, 387-403., Copyright 2006, with permission from Elsevier.

Figure 1.5: A schematic of the production of $\text{A} \beta$ via the proteolytic processing of amyloid precursor protein is depicted in (A). The peptide sequence of $\text{A} \beta 42$ is shown in (B).
Pothesis: In those with trisomy 21 (occurring in Down’s syndrome), the chromosome responsible for encoding APP, the overproduction of Aβ leads to early-onset of dementia with AD-like plaque load. Furthermore, in persons with early-onset familial AD, genetic mutations on the APP lead to the production of Aβ peptides with increased aggregation propensities.

In addition to the presence of amyloid plaques, another hallmark of AD is the intracellular deposition of neurofibrillary tangles (NFTs) composed of aggregated hyperphosphorylated forms of the microtubule-associated protein tau. The role of the tau protein and its interactions with amyloid in the pathogenesis of AD are still being established. Studies with mouse models currently suggest that the role of NFTs in AD may be downstream to that of Aβ because Aβ plaque pathology was not developed in a tau transgenic mouse model, whereas Aβ formation in APP transgenic mice was found to induce hyperphosphorylation of tau, which led to the formation of NFTs.

A puzzling aspect of AD is that the plaque load in the brain of dementia patients is often not correlated with their disease progression and severity. Instead, multiple lines of evidence indicate that synaptic loss and the severity of cognitive impairment are correlated with the concentration of soluble Aβ oligomers in the brain. For example, oligomers extracted from AD brain can impair synapse structure and function. Moreover, when injected into the brains of animal models of AD, Aβ oligomers decreased the number of synapses and impaired learning performance. Furthermore, cellular models of toxicity displayed characteristic symptoms of neurotoxicity that lead to eventual apoptosis upon the addition of Aβ oligomers prepared either in vitro or extracted from cell cultures. Taken together, current experimental evidence indicates that preventing the formation of oligomeric forms of Aβ may be a promising method of treatment for AD.
1.4 Amyloid Inhibition by Small Molecules: a promising method of treatment for AD

With the increasing longevity of our population, AD is approaching epidemic proportions and no cure or preventative therapy is available. In 2010, it was estimated that 36 million people in the world were suffering from AD, and this number is projected to grow to 115 million people by the year 2050. Furthermore, there are no drugs which may target the underlying disease: approved treatments today such as donepezil (a cholinesterase inhibitor), and memantine (a N-methyl-D-aspartate antagonist) only mitigate cognitive symptoms.

Although there are currently no therapeutics for AD, progress is being made. Intensive structural and biochemical studies of amyloid structure have led to the development of potential therapeutics for treating the underlying disease. A detailed review of treatment methods which target the underlying disease is provided elsewhere. In recent years, small-molecule compounds with the ability to reduce the formation, deposition and accumulation of Aβ amyloid aggregates have emerged as a promising method of treatment. In vitro screenings led to the discovery of a large number of small molecules that may affect the amyloid aggregation pathway. Some of these drug-like molecules inhibit the formation of amyloid fibrils, whereas others arrest or reduce non-fibrillar oligomer formation.

A key pharmacological requirement of drugs that target AD and other neurodegenerative diseases is their ability to penetrate the blood-brain barrier (BBB) in sufficient concentrations in the brain to achieve their therapeutic effects (i.e., to inhibit amyloid formation). Although many small molecules appear to be effective in preventing amyloid formation in vitro and attenuating amyloid toxicity in cell cultures, many of these small molecules display poor BBB penetration and are highly toxic, making them unsuitable for immediate use as therapeutics. Below, we provide an overview of small
molecules that are \textit{in vitro} and \textit{in vivo} inhibitors of amyloid fibrillation. Clinical trial data is mentioned where available.

### 1.4.1 Dye-based molecules

Among the first compounds discovered to bind amyloid fibrils were the dye molecules used to identify amyloids (Figure 1.6). Congo red (CR) was initially used in the histological detection of amyloid binding, where, upon binding with CR, fibrils exhibit red-green birefringence when viewed with polarized light.\textsuperscript{30} The binding affinity of CR with various fibrils is in the range of 0.1 - 1.5 $\mu$M.\textsuperscript{104-106} Like CR, ThT displays $K_d$’s in the low $\mu$M range, with values ranging from 0.033 to 23 $\mu$M reported in the literature.\textsuperscript{107} However, unlike CR, although ThT binds tightly to amyloid fibrils, its binding has not been observed to affect amyloid aggregation.

Early amyloid detection using CR revealed that CR not only binds to fibrils, but can also affect the amyloid aggregation pathway by interacting with one or more amyloidogenic species.\textsuperscript{108} Fibril formation of amyloidogenic A$_\beta$ fragments,\textsuperscript{109} prion proteins,\textsuperscript{110} and the immunoglobulin light chain variable domain (SMA)\textsuperscript{111} were found to be promoted
by the presence of CR at low molar ratios, and inhibited at high molar ratios.

Although CR exhibits anti-amyloidogenic and anti-prion properties, its carcinogenic properties make CR a poor therapeutic candidate. Therefore, efforts were applied to find CR-based analogues that maintain their anti-fibrillar aggregation activity but have improved toxicity profile and BBB bioavailability. Chrysamine G (CG) is one such analogue of CR (Figure 1.6). CG has higher lipophilicity and lower toxicity than CR, and is capable of inhibiting aggregation and amyloidogenic toxicity both \textit{in vitro} and \textit{in vivo}.\textsuperscript{112–115}

### 1.4.2 Polyphenols

Polyphenols form a class of molecules found naturally in plants, and are composed of one or more aromatic phenolic rings with multiple hydroxyl groups. Because of their antioxidant properties, the consumption of polyphenols has been reported to be beneficial for health. For example, resveratrol, a polyphenol found in red wine and \textit{epigallocatechin-3-gallate} (EGCG), a major phenolic component of green tea, were found to have cancer-preventative properties.\textsuperscript{116,117} Currently, several clinical trials which ex-
amine their efficacy in cancer treatments are underway. In recent years, polyphenols have gained additional attention due to their potential for treating AD. Here we provide an overview of the compounds resveratrol, EGCG, and curcumin, which have been well-characterized for their ability to inhibit amyloid formation.

*In vitro*, resveratrol inhibits the fibril formation of Aβ and the islet amyloid polypeptide (involved in type II diabetes), and attenuates amyloid-induced cellular toxicity. Similarly, EGCG molecules promote self-assembly of amyloidogenic peptides Aβ and α-synuclein into “off-pathway” oligomers and inhibit the formation of mature fibrils by directly binding to monomeric forms of these peptides. Cell culture experiments indicate that micromolar concentrations of EGCG are protective against Aβ-induced cell death. Curcumin, the main constituent of the spice turmeric, was reported to inhibit Aβ aggregation with IC₅₀ values between 0.1-1 µM.

Because of their anti-amyloidogenic activity, low toxicity, and ability to cross the BBB, polyphenols display therapeutic potential for the treatment of AD and related neurodegenerative diseases. Although it is well-known that ECGC, curcumin, and resveratrol inhibit amyloid aggregation *in vitro*, their mechanisms of action are still unknown. A key disadvantage of these polyphenol molecules is their high metabolic activity in the gastrointestinal system, which leads to poor absorption when administered orally. Clinical trials to measure their efficacy in the treatment of AD are currently underway.

### 1.4.3 Inositol

Inositol or cyclohexane-1,2,3,4,5,6-hexol has the molecular formula C₆H₁₂O₆. Inositol is a simple polyol with nine naturally occurring stereoisomers (Figure 1.8). Out of these nine isomers, seven are optically inactive, and the remaining two (L- and D-chiro-inositol) are chiral enantiomers (Figure 1.8). The stereoisomers of inositol differ in the arrangement of their hydroxyl groups. In particular, scyllo-inositol, with all equatorial hydroxyl groups, is the only isomer that presents two hydrophobic faces. By contrast,
its diastereoisomer, chiro-inositol, with two adjacent axial hydroxyl groups, only has two non-planar faces that are partially hydrophobic. myo-Inositol, the most abundant isomer, is ubiquitous in all eukaryotes and is a physiologically important osmolyte. Furthermore, myo-inositol is a precursor for the synthesis of phosphatidylinositol, an important phospholipid in membranes and second messenger systems. Once phosphorylated, myo-inositol phosphatides act as second messengers in intracellular signal transduction pathways.

Inositol is found in high concentrations in tissues of the human central nervous system (CNS): myo- and scyllo-inositol have approximate concentrations of 5 and 0.1 - 0.5 mM in the CNS, respectively. Accordingly, inositols also function as osmolytes in the CNS, where alterations in their concentrations are known to be associated with neuropathological conditions. For example, the pathogenesis of Down’s syndrome has been linked to abnormal levels of inositol in the cerebral spinal fluid (CSF).

In recent years, scyllo-inositol has been identified as a promising therapeutic candidate for the treatment of Alzheimer’s disease. Inositol was discovered as a possible amyloid inhibitor in a study where lipid bilayers composed of acidic phospholipids were found to induce β-sheet structure in monomeric Aβ42. Upon closer examination, myo-inositol, the headgroup of phosphatidylinositol, was found to be responsible for the in-
duction of non-fibrillar β-sheet structure. This finding led to *in vitro* studies which demonstrated stereochemistry-specific effects of inositol on Aβ fibril inhibition and cytotoxicity: *scyllo*--, *myo*--, and *epi*--, but not *chiro*-inositol inhibit Aβ42 fibril assembly, stabilize an oligomeric complex of Aβ42, and attenuate Aβ-oligomer-induced neurotoxicity *in vitro*. Studies with a transgenic mouse model of AD demonstrated that the decrease in their AD-like symptoms after *scyllo*-inositol treatment was correlated with a decrease in the levels of soluble Aβ oligomers, suggesting that its beneficial effects are attributed to a reduction in the amount of Aβ oligomers.

Presently, *scyllo*-inositol has completed both phase I and II of human clinical trials for the treatment of AD. In phase I trials, based on indicators such as brain plasma and CSF concentration, *scyllo*-inositol was found to be non-toxic to healthy individuals at concentrations effective for amyloid inhibition. From 2007 to 2011, phase II studies of *scyllo*-inositol were conducted with three dosage groups, where subjects in each group (comprised of 84 to 91 people) were administered 250 mg, 1000 mg or 2000 mg of *scyllo*-inositol (ELN005) orally twice a day. However, because of greater incidences of serious adverse events at the two higher dosages, only the lower dose was continued. Due to the decrease in the statistical power of the study after the removal of the high dosage groups, the efficacy of *scyllo*-inositol was not conclusively evaluated. Taken together, these results suggest that *scyllo*-inositol may be a promising therapeutic for the treatment of AD, but that further improvements to enhance its therapeutic properties may be required.

### 1.5 Protein-ligand binding equilibria

Characterizing the binding equilibria between proteins and their ligand substrates is important for understanding the mechanism of action of drug candidates. Binding kinetics describe the rate constants of ligand association ($k_{on}$) and dissociation ($k_{off}$) for the
binding reaction given by

\[
\text{Protein} \xleftrightarrow{\frac{k_{off}}{k_{on}}} \text{Protein} + \text{Ligand} \tag{1.1}
\]

The ratio of the dissociation to the association rate constants establishes the equilibrium dissociation constant of the ligand \((K_d = k_{off}/k_{on})\), which determines the fraction of receptor occupancy at specific ligand concentrations. This equilibrium constant (or the dissociation constant) is

\[
K_d = \frac{[\text{Protein}][\text{Ligand}]}{[\text{Protein} \cdot \text{Ligand}]}.
\tag{1.2}
\]

\(K_d\) is a measurement of the affinity of a ligand for its binding site on the host protein and has units of concentration. Pharmacologically, it is interpreted as the concentration of ligand at which 50\% of the drug is bound to the protein. \(K_d\) is often used as a quantitative indication of drug potency when screening for potential lead compounds. \(\Delta G\), the binding free energy of a ligand (with its receptor) is directly related to its \(K_d\) by

\[
\Delta G = -RT \ln K_d,
\tag{1.3}
\]

where \(R=1.9858775 \times 10^{-3} \text{ kcal} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}\) is the gas constant, and \(T\) is the temperature in the SI unit of Kelvin, and \(K_d\) is the dissociation constant normalized by the standard molar concentration of 1 mol \(\cdot\) L\(^{-1}\).

Hence, a small value of \(K_d\) indicates that the ligand is tightly bound (i.e., it possesses high-affinity binding) to its binding site on the protein, and inversely, a high value indicates weak binding. Enzymes and their inhibitors typically have high binding affinities (or inhibition constants) in the nano- to micromolar range. For example, the nonsteroidal anti-inflammatory drug (NSAID) ibuprofen, which inhibits the enzyme COX-2, displays a \(K_d\) of \(\sim 10 \mu M\).\(^{141}\)

In some cases, a ligand may be capable of interacting with its receptor at multiple
binding sites. Such ligands may possess weak interactions with several receptor sites, but are able to achieve binding specificity by forming many such weak interactions. An important example of this type of binding mechanism occurs in carbohydrate-protein binding, where multivalency plays an important role in ligand recognition. Binding avidity, distinct from binding affinity, is a measurement that accounts for all possible binding modes of a ligand with its receptor, and is well-suited for characterizing multivalent interactions such as protein-carbohydrate interactions.\textsuperscript{142}

### 1.6 Intermolecular forces in biomolecular interactions

Protein-ligand interactions predominantly involve an array of non-covalent interactions. Van der Waals forces are weak intermolecular forces that are either attractive or repulsive. In particular, London dispersion forces are weak attractive forces between all molecules, which arises from induced correlations of instantaneous dipole moments based on the electronic motions of nearby atoms. In addition, all molecules experience steric effects due to the volume of space that atoms in the molecule occupy.

Electrostatic effects also play an essential role in biomolecular recognition and structural stability. Fundamental to biomolecular interactions are hydrogen bonding and charge-charge interactions.\textsuperscript{143} A hydrogen bond is formed from electrostatic interactions arising between two dipoles: an acceptor and a donor. The acceptor group is comprised of an electronegative heavy atom (e.g. F, Cl, O), with exposed lone pairs of electrons, and the donor group is comprised of an electropositive atom bound to a hydrogen atom. Furthermore, unlike dispersion forces, hydrogen bonds are directional due to the interaction of a positively-polarized hydrogen and an electronegative atom. Hydrogen bonds formed between peptide backbones involve N-H and O=O groups as the donor and acceptors, respectively. In liquid and ice water, water molecules form a hydrogen bonding network where a single water molecule may form hydrogen bond with up to four neighboring water
molecules (ice). In this arrangement, the oxygen atom and the two hydrogen atoms of the central water molecule can both form hydrogen bonds with two nearby water molecules.

It has been estimated that the dissociation energies of hydrogen bonds range from weak (0.2 kcal/mol) to strong (40 kcal/mol).\textsuperscript{144} A single water-water hydrogen bond is estimated to have a dissociation energy of approximately 3.2 kcal/mol.\textsuperscript{144} The energy of a N-H $\cdots$ C=O hydrogen bond is estimated to be -4 kcal/mol in the gas phase,\textsuperscript{145} and about -0.5 to -1.5 kcal/mol in solution.\textsuperscript{146} However, the free energy contribution of hydrogen bonding involved in the binding of a ligand and in protein folding is difficult to measure because the energetics of hydrogen bonds are sensitive to the environment in which they are formed.\textsuperscript{145}

Notably, hydrogen bonds are ubiquitous in water, a medium essential to living organisms. A single water molecule is able to tetrahedrally coordinate to four other water molecules to form a network of hydrogen bonds.\textsuperscript{147} This hydrogen-bonded network of water molecules is known to impart to water its uniquely-high heat capacity.\textsuperscript{147} Furthermore, hydrogen bonding plays a key role in stabilizing structures of important biomolecules such as DNA, proteins, and carbohydrates.\textsuperscript{147} In the double-helix structure of DNA,\textsuperscript{148} the complementary base pairing of DNA strands is held together via hydrogen bonding.\textsuperscript{147}

In proteins, intra- and intermolecular hydrogen bonds can be formed between the peptidic backbone, and polar or charged amino acid side chains of polypeptides. Specific hydrogen-bonding patterns define protein secondary structure (such as helices and $\beta$-sheets),\textsuperscript{149,150} and impart stability to the tertiary structures that are formed by the association of secondary structure elements.\textsuperscript{151} Thus, hydrogen bonding patterns determine the arrangement of peptide strands into either parallel or anti-parallel $\beta$-sheets.\textsuperscript{148} In particular, peptides with their peptidic backbones hydrogen-bonded to each other to form in-register $\beta$-sheets are at the core of the cross-$\beta$ structure of amyloid fibrils (see Section 1.1.1).

Salt bridges are considered to be the strongest type of hydrogen bonds in biomolecules.
Figure 1.9: Example of protein-carbohydrate binding: glucose bound to a galactose chemoreceptor protein (Ref: Vyas, N. et al, 1991; Rendered from PDB ID: 2GP). Hydrogen bonds between the protein and glucose are depicted in dotted yellow lines.

In proteins, salt bridges are a type of hydrogen bond formed between two oppositely-charged amino acid side chains (e.g. arginine and aspartic acid), and are thought to contribute to the overall stability of the protein.\textsuperscript{143,152}

1.7 Protein-carbohydrate interactions

Carbohydrates are the most abundant organic compounds in nature and are an important component of living cells. They are sources of energy, building blocks, and recognition elements for both plants and animals. Proteins capable of binding carbohydrates mediate processes such as antigen recognition by the host immune system,\textsuperscript{153,154} signal transduction,\textsuperscript{155} bacterial infections,\textsuperscript{156} and cell-cell adhesion.\textsuperscript{157} Because protein-carbohydrate interactions are involved in many fundamental cellular processes, a great deal of research in the past four decades has been devoted to understanding the molecular basis of their binding mechanism.
Monosaccharides, or simple sugars, are carbohydrates that cannot be hydrolyzed into simpler sugars. They are building blocks of polysaccharides, which are naturally occurring polymers of monosaccharides. Polysaccharides can be hydrolyzed to form many monosaccharide units. Smaller polysaccharides of about three to ten units long are sometimes called oligosaccharides. Furthermore, sugars are chiral compounds: a sugar molecule can either be in the D or L configuration. Most naturally occurring sugars are of the D series.

In the past 20 years, the structures of many sugar-binding proteins (or lectins) have been resolved by X-ray crystallography. Lectins may have monosaccharides or polysaccharides as substrates. An example of a bound glucose in the active site of its protein receptor is depicted in Figure 1.9. Carbohydrate-protein interactions constitute a complex subject. Our discussion here will be focused on the general features of binding sites and binding modes of monosaccharides.

Protein-carbohydrate interactions are relatively weak compared to certain protein-protein interactions. For example, dissociation constants for most lectin-monosaccharide interactions are in the millimolar range. Furthermore, although the affinity increases with oligosaccharides, the dissociation constants are still in the micromolar range. However, the ability of carbohydrates to form multivalent interactions with their receptors, in part, helps lectins achieve their binding specificity. This mechanism is more generally known as the avidity effect and has been well-documented for several protein-carbohydrate systems.

Sugar binding sites are distinct from other ligand binding sites in that they mostly occur at the surface of proteins, and form cavities or grooves. Although any residue may be involved in binding carbohydrates, the most frequently occurring residues are aromatic (Trp, Tyr, and Phe), polar and charged residues (Glu and Asp). Carbohydrate binding sites have been generally classified into two groups: deep, solvent-inaccessible binding pockets and shallow grooves, displaying high and low sugar-binding affinities,
Atomic features of protein-carbohydrate interactions are generally characterized by the concomitant formation of hydrogen bonds and nonpolar interactions with aromatic moieties. Often, a delicate balance between hydrophobic and hydrogen bonding interactions is required for protein-carbohydrate binding, where a slight change in the stereochemistry of the carbohydrate substrate can alter its binding specificity.

Because sugars have numerous hydroxyl groups (-OH), hydrogen bonds are ubiquitous in sugar-protein interactions. The hydroxyl groups can interact with both polar and charged groups of amino acids, and the peptidic backbone. A monosaccharide is often found to form planar bidentate hydrogen bonds with the carboxylate groups of aspartic and glutamic acids (Figure 1.10).

Another characteristic binding mode is the nonpolar stacking between hydrophobic faces of the sugar ring and the side chains of aromatic residues (Figure 1.10). This stacking mode, often referred to as CH-π interactions, is thought to arise from the entropically favorable packing of the sugar’s hydrophobic faces with the protein’s aromatic rings, involving enthalpically favorable interactions with π orbitals.
1.8 Structure-based rational drug design

Very broadly, structure-based rational drug design (SBDD) is a process of developing drug candidates by utilizing the structural knowledge of the target protein. With the molecular structure of the target protein, putative enzymatic active sites and other locations that may bind small molecules may be identified. In SBDD, a target is first identified and its role in the relevant disease pathway is characterized. Then, the molecular structure of the target protein is determined using techniques such as NMR or X-ray crystallography. If the structure of the target is determined in its ligand-bound state, protein-ligand interactions can be characterized at the molecular-level. Importantly, structural characterization of binding sites facilitates selection and construction of the chemical library for high-throughput screening. Screening of a chemical compound library can be performed to identify ligands (e.g. inhibitors) with high binding affinities (low $K_d$) to the binding site. In vitro experiments may be conducted to assess the structure-activity relationship of these compounds. The information obtained from these experiments can either feed back into the design cycle to find better inhibitors, or be used to guide further experiments. SBDD has yielded success in the discovery of drug candidates.\textsuperscript{103,163} Moreover, rational drug design is often applied to optimize ligand binding specificity in an effort to increase the efficacy of the drug candidate, and decrease adverse side effects (toxicity) in the human body.

1.9 Molecular dynamics simulations

Molecular dynamics is a computer simulation technique which employs an empirical mathematical function to describe the atomic interactions in a molecular system, and, together with classical laws of Newtonian mechanics, predicts the atomic trajectory of motion of a molecular system. Thermodynamic and kinetic properties can then be extracted as time averages from these trajectories and used to make a number of predictions.
that are often experimentally challenging to observe or measure.

In general, MD simulations are a useful tool to study the structure, dynamics, and interaction of biomolecules (see Chapter 2 for methodological details). MD simulation studies have been useful in studying many existing fundamental problems of biology and biochemistry, including protein folding, biomolecular self-aggregation, and protein-ligand binding.\textsuperscript{13} With increasingly faster and cheaper computer hardware, and better algorithms, structure-based computer modelling and simulations of protein-ligand interactions are becoming a key component of the modern drug discovery process. For example, MD simulations can predict protein-ligand binding free energies, a quantity that can be used to evaluate how well a ligand binds.\textsuperscript{164–166}

Alternative \textit{in silico} methods such as computational docking, where the energetics of binding is typically estimated without accounting for either ligand or protein flexibility, can provide a crude estimate of ligand binding affinity.\textsuperscript{167,168} Although docking is fast, its inaccuracy often leads to many false positives. In MD simulations, the protein and its putative ligand are allowed to relax and freely move about in the system, allowing a more realistic estimate of the binding free energy. Simulation trajectories of protein-ligand binding can be used to quantitatively assess whether a chemical change to a compound will produce a more potent drug candidate (e.g. residues may be “mutated” in silico).\textsuperscript{169,170}

\subsection{1.9.1 Challenges and limitations of MD simulations}

For molecular simulations to reliably predict and guide experiments, they need to be sufficiently accurate, include a correct representation of the experimental conditions, and adequately sample the relevant biomolecular motions.\textsuperscript{171} Ideally, a simulation should be at least 10 times longer than the slowest important timescale in a system.\textsuperscript{172} However, it is often difficult to sample events on the timescales relevant to many important biologically-relevant phenomena (e.g. protein folding, amyloid formation, ligand binding
and conformational isomerization) because their timescales (typically greater than 1 ms) are often not attainable via brute-force MD simulations using currently-available computing resources. Although modern simulations studies routinely approach microseconds in sampling, only a few studies to date were able to reach timescales of milliseconds.\textsuperscript{173–175}

Consequently, running a single continuous MD simulation alone is unlikely to achieve sufficient sampling of the important states of many biologically relevant systems. For this reason, computational algorithms which enhance sampling of the energy landscape of biomolecular systems are often employed to overcome some of the limitations of conventional MD simulations.\textsuperscript{176}

1.10 Recent progress in elucidating the molecular mechanism of small-molecule binders and inhibitors of amyloidogenic species

Amyloid fibrillation is a multi-stage process involving different species at each stage. Due to the heterogeneous nature of prefibrillar species, experimental determination of the molecular structures of these amyloid species remains a challenge. Further compounding experimental challenges, these small molecules may interact with amyloidogenic species at different stages of aggregation, which are not known \textit{a priori}. Moreover, the binding mechanism of small-molecule inhibitors of amyloid aggregation is not described completely by the classical enzyme-inhibition model. An inhibitor concentration in the micromolar to millimolar range is often required to observe amyloid inhibition,\textsuperscript{102} suggesting that these small-molecule inhibitors are non-specific binders. By contrast, substrates of folded enzymes exhibit higher binding specificity, with binding affinities frequently in the nanomolar range.\textsuperscript{177} Taken together, the above challenges have significantly impeded the determination of the molecular basis of amyloid inhibition by small-molecule inhibitors.
using existing experimental methods.

By contrast, computer simulations are not limited by these experimental challenges and can provide the atomistic level of detail needed to elucidate the action of small-molecules on the inhibition of amyloid formation. As a result, MD simulations have played a key role in advancing the understanding of the binding mechanism of these small-molecule inhibitors. However, elucidating the molecular basis of these small-molecule inhibitors poses a number of difficulties for simulation studies. The structural disorder of amyloid-forming peptides makes it difficult to obtain statistically-meaningful properties from MD simulations. Furthermore, because it is not known whether small-molecule inhibitors may interact with amyloidogenic monomers or aggregates, it is often necessary to examining their binding with several amyloidogenic species in order to gain a complete understanding of their mechanism of action.

A large number of MD simulation studies were published at the time of research reported in this thesis. Below I will provide a review of the recent progress in elucidating the binding mechanism of small molecule inhibitors of amyloid formation. While the central focus of the review is on MD simulation studies, experimental results are discussed if they are available. Both in silico methods and experimentally-tractable model self-assembly systems have provided significant insight into the binding modes of small molecules with amyloid fibrillar aggregates. MD simulation studies examined the binding of Congo red (CR) with the protofibril of the amyloidogenic fragment GNNQNY of the yeast prion Sup35, and most recently, with full-length Aβ40. Although protofibrils of Aβ40 were found to contain more binding sites for CR than GNNQNY, CR shared similar binding modes with protofibrils of both peptides: CR molecules bound in the fibrillar grooves, between peptide strands, in parallel to and perpendicular to the long-axis of the fibril, respectively. These results led the authors to hypothesize that the observed β-sheet surface binding mode disrupts β-sheet stacking, whereas edge binding (found exclusively with Aβ40 fibrils) blocks strand-to-sheet extension. Moreover, based
on these binding modes, a model for explaining the birefringence displayed by Aβ fibrils upon binding CR was proposed.\textsuperscript{179}

To identify the minimal requirements for ThT binding, Biancalana \textit{et. al.} designed a set of novel ThT-binding proteins to create a minimalist binding site for ThT that recapitulated all of its fibril binding properties (i.e. fluorescence emission upon binding).\textsuperscript{180} The X-ray crystal structure of the β-sheet protein containing high-affinity binding sites for ThT showed that its surface contained repetitive grooves formed by tyrosine (Tyr) and leucine (Leu) arranged in a side-by-side manner (“ladders”).\textsuperscript{181} Based on this structure, the authors hypothesized that ThT preferentially binds to aromatic grooves at fibrillar surfaces. Wu \textit{et. al.} later conducted a complementary MD simulation study of this model ThT-binding protein which corroborated this hypothesis.\textsuperscript{182}

In several MD simulation studies, Wu \textit{et. al.} examined the binding of ThT and its analogs to protofibrils of Aβ(16-22)\textsuperscript{183} and full-length Aβ40 peptides.\textsuperscript{184} These studies revealed that similar to the binding modes of CR, ThT was bound to several binding sites located on the surface of fibrils. Furthermore, the results of their studies suggest that although CR and ThT share similar binding site motifs, they are not likely to share similar binding sites because of differences in their chemical structure and variations in the surface properties of fibrils.

Taken together, the above studies suggest a consensus for the molecular mechanism of dye binding to amyloid fibrils: dye molecules adopt specific binding modes in the grooves at the surface of amyloid fibrils, which give rise to the physical properties exhibited by dye-bound fibrils. However, it is still not understood why CR binding leads to amyloid inhibition, but ThT binding does not. Further investigation of the differences in their binding mechanisms will be required.

Using a combination of isothermal titration calorimetry (ITC), NMR, and MD simulations, the interaction of -epigallocatechin-3-gallate (EGCG) molecules with monomeric Aβ42 as modulated by temperature, pH, salt concentration, and ligand:protein molar
ratio was investigated by Wang et. al\textsuperscript{185} The simulations were performed in the presence of EGCG, at increasing molar ratios, with the Aβ peptide in α-helical conformation. Results of this study indicated that both hydrogen bonding and hydrophobic (aromatic) interactions are important for binding of EGCG, and that the balance of these interactions is particularly sensitive to ligand:protein stoichiometry.

The molecular details of EGCG binding to Aβ monomers were further explored using MD simulations, exclusively, in a follow-up study\textsuperscript{186} using a similar simulation protocol as in the previous study.\textsuperscript{185} Using MM-PBSA,\textsuperscript{187} a technique that employs continuum solvent models to estimate binding free energies, the authors found that nonpolar interactions rather than hydrogen bonding contributed more to EGCG binding modes. Furthermore, based on secondary structure analysis of the peptide conformations, EGCG, at ligand:peptide molar ratio of 10:1, was found to prevent β-strand formation.

Binding modes of morin, a polyphenol molecule with the ability to inhibit amyloid formation of Aβ and IAPP peptides \textit{in vitro},\textsuperscript{127,188} were investigated with protofibrils of Aβ\textsubscript{42} using MD simulations by Lemkul et. al.\textsuperscript{189} Based on contact analysis and snapshots from simulations, morin predominantly bound at the edges of the protofibril, and partially penetrated the hydrophobic interior of the protofibril. Furthermore, in a significant fraction of the simulations, morin molecules hydrogen-bonded to and intercalated between a pair of residues which formed salt bridges within the fibril.

In a follow-up MD study, the same authors investigated the effect of morin binding on the aggregation and conformational equilibria of monomers and dimers of Aβ\textsubscript{42}.\textsuperscript{190} Morin was found to bind to the residues flanking the central hydrophobic core region (CHC) of the Aβ peptide (residues 16 to 21), but not with the CHC itself. Morin-peptide interactions were found to impede the formation of intra- and inter-peptide hydrophobic contacts, but did not affect the secondary structure of the peptide. On the basis of these results, the authors proposed that morin prevents fibril formation of Aβ\textsubscript{42} by preventing and disrupting the hydrophobic association to form an initial aggregation nuclei.
NSAIDs, typically administered for pain relief, are also found to inhibit Aβ fibrils in vitro. In a series of MD simulation studies, Klimov et al. examined the binding mechanism of ibuprofen and naproxen with monomers, oligomers and protofibrils of Aβ40. In each of these studies, replica-exchange MD simulations, an enhanced sampling simulation methodology, were conducted to increase the likelihood of NSAID-Aβ binding events. To reduce time required for their simulations, an implicit solvent model was used in place of representing water molecules explicitly.

The authors found that ibuprofen preferentially binds to the protofibril rather than to the monomer of Aβ40, predominantly in the hydrophobic grooves at the edges of the protofibrils. This protofibrillar binding mode of NSAIDs was speculated to inhibit fibril formation by blocking fibril elongation. Furthermore, a comparative study of naproxen and ibuprofen was carried out to identify the molecular basis of naproxen’s stronger in vitro binding affinity. Although both NSAID molecules share binding sites on the fibril, naproxen was found to have higher binding energy than ibuprofen because of its preference for self-interaction.

In summary, the above simulation studies suggest that both hydrophobic and hydrogen bonding interactions are involved in the binding mechanism of known in vitro small molecule inhibitors. Based on MD simulations, several mechanistic hypotheses for the molecular mechanism of amyloid inhibition by small molecules have been put forth. First, it has been hypothesized that fibril formation may be prevented by binding to surfaces of β-sheet aggregates. Furthermore, many small-molecule inhibitors are thought to interact with multiple species along the amyloid formation pathway. For this reason, simulation studies were conducted to examine the interaction of small molecule inhibitors with peptides and protein monomers and non-fibrillar amyloid aggregates. These studies suggest that small molecules prevent peptide self-aggregation by interacting with amyloidogenic aggregates and displacing their β-sheet-forming propensity.

Despite the continued increase in computational power, modern MD simulations of
disordered proteins and peptides remain computationally challenging.\textsuperscript{21} For this reason, the conclusions drawn from these simulations may suffer from systematic or statistical sampling errors that are difficult to identify. Thus far, studies of small molecule binding lack statistically significant numbers of binding events, due to the lack of sufficient sampling time. Hence, dissociation constants ($K_d$), which are an important metric of a compound’s efficacy in the rational drug design process, have not been estimated from simulation studies. Without an estimate for the $K_d$, the relevance of the binding sites identified in simulation studies are more difficult to assess because direct comparisons with \textit{in vitro} measurements cannot be made.

An implicit solvation model may be used in place of explicit solvent representation in order to reduce computational complexity of the simulations.\textsuperscript{197} Some of the studies referred to above have employed this methodology to examine small-molecule binding mechanisms.\textsuperscript{193–195} However, the lack of explicit solvent, an important contributor to the free energy of ligand binding, can lead to quantitative errors in binding mode predictions. Certain models of implicit solvation introduce approximations to account for these effects, but are much less accurate in reproducing them.

Few studies conducted currently are systematic comparative studies of monomers, disordered oligomers, and protofibrillar aggregates. Because it is often not known which amyloidogenic species these small molecules may act on, the lack of adequate comparisons may leave a study susceptible to biased mechanistic hypotheses. Moreover, experimental studies have suggested that the activity of these small molecules is modulated by their concentration and molar ratio. For example, the self-aggregation of small-molecule binders at high concentrations has been reported in both experimental\textsuperscript{198–200} and simulation studies.\textsuperscript{186,193,195} However, the mechanistic link of this effect to their mechanism of amyloid inhibition has not been addressed by most studies.

Amyloidogenic monomers and oligomers can adopt a large number of conformational states, and requires a large amount of computational sampling in order to converge their
conformational properties in order to determine the effect of small-molecule binding on their conformational equilibria. Hence, to reduce simulation time, several MD studies have utilized smaller fragments of the Aβ peptide.\textsuperscript{201–203} However, many existing studies in the literature, particularly those which probe the effect of ligand binding on conformation of monomers and oligomers of the full-length Aβ peptides, draw conclusions based on non-equilibrium simulations where the conformational property of interest likely remains unconverged.\textsuperscript{186,190,202} Due to the lack of assessment of convergence in these simulations, researchers have been unable to assess the statistical significance of their data, which could lead to significant systematic biases.

1.11 Thesis objectives and organization

One out of eight people aged 65 or older has Alzheimer’s disease (AD).\textsuperscript{204} With the increasing longevity of our population, AD is approaching epidemic proportions with no cure or preventative therapy available. A pathological hallmark of AD is the extracellular deposition of amyloid in the brain. These fibrillar deposits (plaques) are formed from the self-aggregation of the β-amyloid (Aβ) peptide, a 38 to 42 residue protein that is produced normally as part of the cellular metabolism. Similarly, amyloid composed of other peptides or proteins are also found in other neurodegenerative diseases such as Parkinson’s, Huntington’s, and prion-related diseases.

One therapeutic approach is the development of small-molecule inhibitors of Aβ aggregation. Recently, scyllo-inositol has emerged as a promising compound for the treatment of AD, which has currently completed phase two of clinical trials. It is one of eight stereoisomers of inositol found in nature and exhibits stereochemistry-dependent inhibition of Aβ fibrillation.\textsuperscript{136} Scyllo-inositol is effective at reversing the established disease state as well as preventing the onset of AD-like symptoms in a transgenic mouse model of AD, whereas chiro-inositol is inactive.\textsuperscript{137} Currently, scyllo-inositol has completed phase
two of clinical trials, which evaluates its dose-related safety and efficacy in participants with mild to moderate AD.

Although scyllo-inositol raises hope for the development of a cure for AD, it is likely that effective therapies for patients with AD will require rational modification of this compound. Understanding the molecular basis for the action of inositol and in particular, its effect on Aβ aggregation will aid in the effective development of inhibitors of amyloid aggregation. At present, however, experimental approaches lack the ability to determine the precise mechanistic modes of action of inositol as the molecular structures of various intermediates in the Aβ aggregation pathway are not known. Moreover, intermediate states in the fibrillation process are very difficult to detect and isolate by experimental methods.

The primary objective of my research is to use MD simulations to elucidate the molecular basis for the activity of inositol by determining its effect on the structure and thermodynamics of Aβ aggregation. Our central hypothesis is that inositol acts by binding to one or more of the aggregated form of Aβ. The formation of amyloid follows a complex aggregation pathway, where different intermediate Aβ aggregate species have been implicated in the disease. Small molecule inhibitors such as inositol may interact with species at various stages of aggregation. Therefore, a meaningful study requires examining inositol binding with different self-aggregated peptide states, from monomeric to fibrillar aggregates. To this end, I have carried out systematic comparative studies of amyloidogenic peptides and their aggregates of increasing sequence and composition complexity and characterized the respective role of specific interactions of inositol with the backbone and sidechains (Figure 1.11). In all of my studies, I have comparatively examined chiro- and scyllo-inositol with each of the aggregates in the pathway to determine the stereochemical basis of the activity of inositol. In addition, I also performed control simulations in the absence of inositol.
Figure 1.11: A schematic of the different peptide sequences and aggregation states which form the basis of my studies involving inositol, arranged in order of structural and sequence complexity along the X and Y axes, respectively.
1.11.1 Organization of thesis

In Chapter 2, I review the details of molecular dynamics simulations, the central methodology throughout my dissertation.

In Chapter 3, as a first step to investigate the stereochemistry-dependent effect on amyloidogenic peptide aggregation and morphology, I examined the effect of backbone interactions in polypeptide self-aggregation by systematically characterizing the binding equilibria of inositol with model peptides, alanine dipeptide (ADP), an amyloid-forming peptide (Gly-Ala)$_4$.

In Chapter 4, I continue my investigation by characterizing the binding mechanism of inositol with peptide and aggregates of the amyloid-forming peptide fragment Aβ(16-22) (or KLVFFAE), a fibril-forming fragment thought to initiate amyloid formation in the full-length Aβ peptide. Binding to grooves at the surface of β-sheet aggregates is found to play a central role in the activity of inositol. Taken together, the results presented in these chapters suggest that inositol is likely to act as a drug on protofibrillar-like aggregates. Accordingly, in Chapter 5, I characterize the binding mechanism of inositol with protofibrils of Aβ42.

On the basis of my work involving inositol, a notable result is that the methodology utilized in these studies may be useful for investigating protein-carbohydrate binding mechanism in general. In Chapter 6, applying the general methodology developed in this thesis, I characterize the binding modes and sites of monosaccharides glucosamine and GlcNac with PgaB, a key protein responsible for the export of polysaccharides important for the formation of biofilms. Using conventional sampling methods, I was able to map out a binding surface that predicted possible binding modes for the biologically-relevant polymer substrate of PgaB. The study presented in this chapter demonstrates that the methodology I have developed and utilized throughout this thesis can be successfully applied to characterize carbohydrate-binding proteins.
Bibliography


Chapter 2

Molecular Modelling and Simulation

2.1 Molecular Mechanics

In principle, the time-evolution of a molecular system can be predicted using quantum mechanics by solving the time-dependent Schrödinger’s equation, which provides a complete description of the motion of atomic nuclei and electrons from first principles.\textsuperscript{38} Notably, this equation can only be solved analytically for a hydrogen atom, and must be solved numerically for more complex systems. Although many quantum-mechanics-based computational methods have been developed in recent years,\textsuperscript{12} they are still computationally prohibitive for probing the dynamics of most protein systems on a biologically-meaningful timescale because these systems, with solvent molecules included, typically possess on the order of $10^4$ to $10^5$ atoms.

Molecular mechanics is an approach in which a molecular system is treated using the principles of classical mechanics. In molecular mechanics, a molecule is modelled by a set of masses (atoms) connected by springs (bonds). An empirical potential energy function (PEF) that only depends on the positions of the atoms is employed to describe the internal energy and interactions of molecules within a system.\textsuperscript{6-8,22} The theoretical basis for molecular mechanics is rooted in quantum mechanics via the Born-Oppenheimer
approximation.\textsuperscript{4} Under this approximation, the electronic and nuclear motion of atoms are considered to be uncoupled, and therefore can be treated separately. Because nuclei are much heavier than electrons, they are practically stationary on the time scales of electronic vibration. For this reason, the total energy (or stationary Hamiltonian, $H$) of a system can be considered to depend only on the electronic motion. As a consequence, the nuclear motion of a system is then governed by a potential energy function under the influence of a single electronic state.

The PEF employed in molecular mechanics corresponds to the potential energy surface of the molecular system in its electronic ground state. In molecular mechanics, atomic nuclei are treated as point charges, and their electronic distribution is implicitly accounted for in the PEF. Because electronics are not treated, no chemical reactions (i.e., bond formation) can occur.

\section*{2.2 Force Fields for Biomolecules}

The functional form of the PEF together with its associated parameters define a \textit{force field}. A typical empirical force field has many parameters to be determined. These parameters are determined by fitting to quantum mechanical calculations, and to physical measurements (e.g., using data from X-ray crystallography, neutron diffraction, vibrational raman spectroscopy) of small organic compounds, building blocks of peptides and nucleic acids. Over time, force field parameters are adjusted to improve the agreement of predicted quantities. In the past four decades, a variety of force fields have been developed and applied toward the modelling of biomolecular systems involving proteins, nucleic acids and carbohydrates. The widely adopted force fields include AMBER,\textsuperscript{8} CHARMM,\textsuperscript{6} GROMOS,\textsuperscript{7} and OPLS.\textsuperscript{21,22,25} Reviews and comparisons of these force fields are provided elsewhere.\textsuperscript{2,26,33} Although force fields may adopt different parameterization
strategies, they share a similar general form of the PEF:

\[ V = V(bonds) + V(angles) + V(dihedrals) + V(impropers) + V(non\text{-}bonded). \quad (2.1) \]

### 2.2.1 Bonded Interactions

The first four functional terms of Equation 2.1, \( V(bonds) \), \( V(angles) \), \( V(dihedral) \), and \( V(improper) \), define the through-bond interactions, covalent bond stretching, angle-bending, bond torsion angles, and planar dihedral angles terms, respectively. Covalent bond stretching can be described using Hooke’s Law for harmonic motion,

\[ V(bonds) = \sum_{bonds} k_b(b - b_0)^2. \quad (2.2) \]
Likewise, angle-bending motions can also be described harmonically as

\[ V(\text{angles}) = \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2. \]  

(2.3)

In Equations 2.2 and 2.3, \( b \) and \( \theta \) are variables representing the bond length and bond angle. The constants \( k_b \) and \( k_\theta \) are the equilibrium force constants for bonds and angles, respectively, and the constants \( b_0 \) and \( \theta_0 \) are the reference bond length and angle, respectively. The total energetic contribution from bond stretching and angle bending are computed by summing over all bonds and angles in the system. In addition, the torsion (or dihedral) angle of bonds is also important for describing molecular conformations. These energetic terms are commonly accounted for by using a periodic function of the form,

\[ V(\text{dihedrals}) = \sum_{\text{dihedrals}} \sum_n k_\chi (1 + \cos(n\chi - \delta)). \]  

(2.4)

where \( \chi \) is the dihedral angle, \( k_\chi \) and \( n \) are the height and the periodicity of the rotational barrier, respectively. \( \delta \) is the reference angle (or phase). The value of \( n \) is usually between 1 and 4, and depends on the specific force field parameterization procedure. Functions of this form are also included the PEF to ensure the planarity of planar chemical groups, in which case \( \chi \) corresponds to a so-called improper dihedral angle.

2.2.2 Non-bonded Interactions

Non-bonded interaction potentials predominantly account for van der Waals and electrostatic interactions. Accurate modeling of these interactions is particularly important for the transferability of a force field, which is the assumption that a force field developed based on a small set of molecules can predict properties of a broader set of molecules composed of similar chemical groups. The non-bonded forces are only applied to atom pairs from different molecules or separated by at least three bonds, and are pair-wise additive, that is, the total contribution to the potential energy from the non-bonded forces...
is the sum total over all interacting atomic pairs in a system. Van der Waals interactions are represented by the Lennard-Jones potential,

\[ U_{LJ}(r) = \left( \frac{r_{\text{min}}}{r} \right)^{12} - 2 \left( \frac{r_{\text{min}}}{r} \right)^6. \] (2.5)

The first term of Equation 2.5 represents a repulsive force at distances where atomic shells overlap. The second term represents the attractive force between atoms which is most favorable at a distance of \( r_{\text{min}} \), the preferred interaction distance for a pair of atoms. \( U_{LJ} \) is dependent on the distance between a pair of atoms, \( r \), and decays rapidly to 0 as \( r \) increases.

The modeling of electrostatics is paramount for describing interactions of polar and charged groups in biomolecules. To do so, the Coulombic potential,

\[ U_{\text{coulombic}}(r) = \frac{q_i q_j}{\epsilon r}, \] (2.6)

is calculated for all atom pairs that carry partial charges or full charges at their nuclear centers. In Equation 2.6, \( q_i \) and \( q_j \) are the point charges assigned to atoms \( i \) and \( j \), respectively, and \( \epsilon \) is the dielectric constant. In contrast to van der Waals interactions, electrostatic interactions can be significant for charged pairs that are separated by relatively long distances. Accordingly, in comparison to the Lennard-Jones potential, the Coulombic potential decays to zero at a much slower rate.

As described in Chapter 1, hydrogen bonding is an important type of electrostatic interaction in biomolecular systems. For this reason, a proper representation of hydrogen bonds should be captured by a force field. Some early versions of force fields have employed an explicit term solely for describing hydrogen bonding.\(^{26}\) For example, the original AMBER force field introduced in 1984 added a function of the form

\[ \sum_{\text{hbonds}} C_{ij}/R_{ij}^{12} - D_{ij}/R_{ij}^{10} \]

to account for hydrogen bonds.\(^{43}\) However, it was later demonstrated that hydrogen bonding interactions are most accurately accounted for by appro-
appropriate parametrization of van der Waals and electrostatic potential terms.\textsuperscript{8,26} Hence, most modern force fields for protein simulations, including the latest AMBER force fields, no longer utilize an explicit hydrogen bonding term.\textsuperscript{26}

### 2.2.3 Electronic Polarizability

Electronic polarizability is the measure of the change in a molecule’s electron distribution in response to an applied electric field, and is a property of matter which can be induced by electrostatic interactions with solvents or ionic reagents. As mentioned above, most current empirical force fields use a fixed-charge model, that is, each atom in the system is assigned a single value that remains unchanged. Polarizable force fields, which account for variations in charge distribution with atomic motion dielectric environment, have been introduced.\textsuperscript{13,42} However, because of the associated high computational cost of using these force fields, they are not routinely used for biomolecular simulations.

### 2.2.4 The AMBER and OPLS Force Fields

In this section, I briefly describe two widely used force fields for protein simulations, AMBER and OPLS. OPLS-AA/L\textsuperscript{23} was employed in the studies presented in Chapters 3 to 5, and AMBER99\textsuperscript{41} was employed in the study in Chapter 6. More extensive and detailed reviews of empirical force fields for biomolecular simulations have been provided elsewhere.\textsuperscript{26,33}

The AMBER force field and the associated molecular mechanics simulation package for the simulation of biomolecules were developed by the Kollman group.\textsuperscript{33} The central ideas for parameterization were to fit atomic point charges to the quantum electrostatic potential (ESP) using \textit{ab initio} quantum mechanical calculations, and develop parameters for the bonded and van der Waals potential terms using experimental structural data and liquid-state simulations.\textsuperscript{8,43} The initial parameters were improved upon via the reproduction of experimental measurements for model compounds. The early versions of the 1983
Chapter 2. Methods

AMBER force field accounted for only non-aliphatic hydrogen atoms, and an all-atom version of AMBER was later introduced in 1995. Since then, many iterations of improvements upon the original all-atom AMBER force field have been developed. In particular, AMBER99 (or ff99) is one such revision with the peptide group dihedral angles re-parameterized to improve the reproduction of peptide secondary structure.

The OPLS (Optimized Potentials for Liquid Simulations) force field, as its name suggests, was originally developed to reproduce the liquid-state properties (e.g., densities and heats of vaporization) of compounds. Parameters were first developed for water, ions, hydrocarbons, and were later extended to amides, peptides, and carbohydrates. As in AMBER, in early OPLS development, a partial united-atom model was used whereby nonbonded interaction sites were placed only on non-hydrogen atoms and polar hydrogens. All-atom versions of the OPLS force field (OPLS-AA) for organic molecules and peptide simulations were subsequently developed. In OPLS, the parameters for non-bonded interactions were derived by comparisons to liquid-state thermodynamics, and the atom types and parameters for bonded interactions (bonds, angles, and dihedrals) were largely adopted from the AMBER all-atom force field.

2.3 Molecular Dynamics Simulation

MD is an algorithm that simulates the motion of a system under the influence of a specific force field by following molecular configurations in time according to Newton’s equation of motion.

In a system composed of $N$ interacting particles, the force acting on each particle is determined by taking the spatial derivative of the potential energy of the system, $V(r_1, r_2, ..., r_N)$. In MD simulations, this potential energy is approximated by a molecular mechanics force field as described above. The force on each particle $i$, at time $t$, is given
by
\[ \vec{F}_i = -\frac{\partial V}{\partial \vec{r}_i}, i = 1, 2, 3, ..., N. \tag{2.7} \]

The variable \( r_i \) represents the spatial coordinates of the particle \( i \). The calculated force vectors are summed together to yield the net force vector for each atom in the system. By Newton’s second law of motion, the acceleration, \( a_i \), of each atom is given by
\[ \vec{a}_i(t) = \frac{\vec{F}_i}{m_i}, \tag{2.8} \]
where \( m_i \) is the mass of the \( i \)th particle. Hence, the velocity at a subsequent time step, \( \vec{v}_i(t + \delta t) \), is then given by,
\[ \vec{v}_i(t + \delta t) = \vec{v}_i(t) + \int_t^{t+\delta t} \vec{a}_i(t)dt = \vec{v}_i(t) + \vec{a}_i \cdot \delta t. \tag{2.9} \]

The position of atom \( i \) at time \( t + \delta t \) is then,
\[ \vec{x}_i(t + \delta t) = \vec{x}_i(t) + \vec{v}_i(t)\delta t + \frac{\vec{a}_i(t)\delta t^2}{2}. \tag{2.10} \]

As shown in Equations 2.9 and 2.10, the predicted positions of each of the \( N \) particles are approximated numerically by finite difference in MD simulations using a time-integration algorithm. For example, the leap-frog and velocity-verlet algorithms are commonly used in MD simulations.\(^{16,40}\) To ensure numerical stability, a small integration time step \( \delta t \) in the range of 1 - 4 femtoseconds (fs) is chosen. In MD simulations of biomolecular systems, when the bond lengths involving hydrogens are frozen, a 2 fs timestep is commonly chosen because it is twice the time period for the fastest vibrational motion (existing in bonds involving hydrogen). Once the position of each particle is computed, subsequent interatomic forces are updated based on these new coordinates, and the entire process is repeated to obtain the coordinates of the system at future time steps.
2.3.1 Temperature and Pressure

Temperature is a measure of the kinetic energy of the system, and the instantaneous temperature of a macroscopic system can be determined from the kinetic energy of every atom. The theorem of equipartition of energy states that in a (classical) many-body system at thermal equilibrium, the average kinetic energy (K) per degree of freedom is related to the thermodynamic temperature, $T$, via

$$K = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2} m_i v_i^2 = \frac{3}{2} k_B T,$$  \hspace{1cm} (2.11)

where $m_i$ is the mass of the atom $i$, $v_i$ is the velocity of $i$ along a single dimension, and $k_B$ is the Boltzmann constant. In MD simulations, the temperature of the system is the total kinetic energy of the system normalized by the number of degrees of freedom in the system, that is,

$$T(t) = \frac{1}{3N} \sum_{i=1}^{N} \frac{m_i v_i^2(t)}{k_B},$$  \hspace{1cm} (2.12)

where $N$ is the number of atoms and $v_i(t)$ is the velocity of atom $i$. To simulate a system close to laboratory conditions, a thermostat and a barostat are employed to control both the temperature and pressure of a simulation system, respectively. Many algorithms exist for the implementation of thermostats and barostats, and are available in modern molecular dynamics simulation software packages.\textsuperscript{5,15} In this thesis, the Berendsen\textsuperscript{3} and Nosé-Hoover algorithms,\textsuperscript{17,31} and the Parrinello-Rahman\textsuperscript{30,32} algorithm were employed for temperature- and pressure-coupling, respectively.

2.3.2 Periodic boundary conditions

Periodic boundary conditions (PBC) are a technique by which the central simulation cell (or “box”) is replicated along each of its dimensions to form a system with an infinite number of images. It is used to alleviate artificial phase boundaries imposed by a finite
simulation cell. Furthermore, PBC allow the simulation system to be treated as a bulk system, akin to *in vitro* experimental conditions. To avoid counting duplicate interactions from multiple adjacent cells, a minimum image convention is applied – an atom $i$ interacts with another atom $j$ only in its closest periodic image.

### 2.3.3 Connection to Statistical Mechanics

MD simulations yield information about the microscopic states of a macroscopic system characterized by thermodynamic variables such as the temperature, pressure, and number of particles. Statistical mechanics is a body of theory that provides a rigorous connection between the microscopic arrangements of a system (determined by MD simulations) to its thermodynamic properties.

MD simulations can be carried out in the canonical (NVT) ensemble, where the number of particles ($N$), volume ($V$) and temperature ($T$) are held constant, or, alternatively, in the isothermal-isobaric (NPT) ensemble, where number of particles and temperature are held constant, but the average pressure, instead of the volume, is held constant by allowing the volume of the simulation box to fluctuate. Because the NPT ensemble is the closest to typical experimental conditions, simulations are often carried out in this ensemble.

For a canonical ensemble of microstates, the probability of finding a system in the microstate $i$, with an energy $E_i$ is given by the Boltzmann distribution,

$$p_i = \frac{1}{Z} e^{-\frac{E_i}{kT}}. \quad (2.13)$$

$Z = \sum_i e^{-\frac{E_i}{kT}}$ is the normalization factor (or the *partition function* of a system) such that $\sum_i p_i = 1$. Any thermodynamic variable can then be derived from the partition function. For example, free energy, an important thermodynamic quantity, measures the useful work obtainable from a system under a certain set of conditions. It is called the
Gibbs free energy \( (G) \) for a system in the NPT ensemble, and the Helmholtz free energy \( (A) \) for a system in the NVT ensemble. In biochemistry, a central quantity of interest is the change in free energy between two states of interest, which measures whether a reversible biochemical process will occur spontaneously. The relative free energy \( \Delta G \) of any two states, 0 and 1, is related to their respective probabilities \( p_0 \) and \( p_1 \) via

\[
\Delta G = -k_B T \ln \frac{p_0}{p_1}.
\]  

(2.14)

Performing a MD simulation yields a time-trajectory of the positions of atoms in the system from which we can obtain time-averaged thermodynamic properties of the system. However, in an experiment, the macromolecular observables measured are a result of an average over all possible microstates of a system, that is, an *ensemble average*. This apparent conundrum is reconciled by one of the fundamental axioms of statistical mechanics, the ergodic hypothesis, which states that if a system is given long enough time to evolve, the time average of any of its macromolecular properties is equivalent to its ensemble average.\(^{28}\) In practice, this means that if a simulation sufficiently samples the relevant regions of phase space (i.e., the collection of all possible configurations of a given system specified by its coordinates and momenta), then statistically-meaningful experimental properties can be extrapolated from such a simulation.

### 2.3.4 Enhanced Sampling Techniques

As described above, in an MD simulation, atomic motion at a given temperature is dictated by a potential energy function, and states are sampled with probabilities dictated by the Boltzmann distribution. Given an infinite amount of time, a simulation will sample all states of a system’s potential energy landscape. However, infinitely long simulations are not possible to achieve computationally. Thus, because of the rugged energy landscapes common to biomolecular systems, many important states may not be sampled
on the timescales of a typical simulation. Furthermore, it is even more challenging to adequately sample the complex conformational landscape of disordered proteins because of the large number of conformational states that may need to be sampled in order to fully characterize their structural ensemble.\textsuperscript{35}

Generalized-ensemble algorithms are a class of algorithms that can be used with molecular simulations to enhance conformational sampling. Simulated tempering (ST) and replica exchange (RE) are two well-known such algorithms that can be used to enhance sampling by allowing simulations to diffuse in temperature-space, while obeying Boltzmann statistics at any of the temperatures visited (Equation 2.13).\textsuperscript{27,29} These algorithms rely on the fact that at higher temperatures the free energy landscape of a biomolecular system becomes less rugged, and, hence, the crossing of energetic barriers is facilitated. Simulations are carried out at different temperatures using a temperature range that encompasses the temperature of interest (typically, room temperature). The optimal temperature range depends on the size and composition of the system.\textsuperscript{34,35} Although generalized-ensemble methods, when combined with MD simulations, can significantly enhance conformational sampling of biomolecular systems,\textsuperscript{34,44} they are computationally expensive, and often require a parallel supercomputing platform to carry out a single study in a feasible time frame.\textsuperscript{36}

**Simulated Tempering (ST)**

The ST algorithm utilizes a generalized Hamiltonian that depends on all configurational degrees of freedom and temperature\textsuperscript{27}:

\[
H(X, m) = \beta_m H(X) - a_m,
\]

where $X$ is the configuration of the system, $\beta_m$ is the inverse temperature, $H(X)$ is the system’s original Hamiltonian, and $a_m$ is a constant which depends on temperature. Hence, the probability of sampling a given temperature, $T_m$ is
\[ P(T_m) = e^{a_m(T)} e^{-\beta_m H(X)}. \] (2.16)

The weight factors, \( a_m(T) \), are dimensionless Helmholtz free energies given by

\[ a_m(T) = \beta_m A_m. \] (2.17)

where \( A_m \) is the free energy of the system at temperature \( T_m \). \( a_m(T) \) must be chosen such that an uniform probability distribution in temperature space is achieved. Allowing a uniform sampling of temperature in turn induces a random walk in PE space which allows a system to escape from local minima.

In a typical ST simulation, a short canonical MD simulation is performed at temperature \( T_i \) followed by an exchange attempt to a neighbouring temperature, \( T_j \). An exchange attempt has the probability

\[ p(T_i \rightarrow T_j) = \min \begin{cases} 1 \\ e^{-(\beta_j - \beta_i) E + (a_j - a_i)} \end{cases}. \] (2.18)

The weight factors \( (a_m) \) are \emph{a priori} unknown and must be determined by a iterative procedure. A main disadvantage of ST is that determining an accurate set of weight factors for complex biomolecular systems can be very challenging to obtain.\(^{34}\)

**Replica Exchange (RE)**

Replica exchange (RE) bypasses the problem of determining unknown weight factors while still allowing random walks in temperature space.\(^{14,29}\) RE molecular dynamics simulation is briefly review below. Detailed description and derivation for the algorithm can be found elsewhere.\(^{29}\)

In RE, a system of multiple non-interacting copies or replicas of a system where each replica has a different temperature are considered.
Let $X = \{x_i^{T_i}\}_{i=1}^M$, where $x_i^{T_i}$ is a replica at temperature $T_i$, with coordinates $q_i$ and momenta $p_i$. Recall that the Boltzmann weight factor of $x$ in the generalized ensemble is $W(x; T) = e^{-\beta H(q)}$. Because the replicas are assumed to be independent, the weight factor for $X$ is

$$W_{REM}(X) = \prod_{i=1}^M e^{-\beta_{T_i} H(q_i^{T_i})}$$  \hspace{1cm} (2.19)

Periodically, a pair of replicas $i$ and $j$ in $X$ exchange their temperatures,

$$X = (\ldots, x_i^{T_i}, \ldots, x_j^{T_j}, \ldots) \rightarrow X' = (\ldots, x_j^{T_j}, \ldots, x_i^{T_i}, \ldots)$$  \hspace{1cm} (2.20)

In order to have the $M$-replica system under a set of exchanges to converge to the equilibrium (Boltzmann) distribution, the detailed balance condition should hold. That is,

$$W_{REM}(X)w(X \rightarrow X') = W_{REM}(X')w(X' \rightarrow X),$$  \hspace{1cm} (2.21)

where $w(X \rightarrow X')$ is the transition probability to go from state $X$ to state $X'$. This directly leads to,

$$\frac{w(X \rightarrow X')}{w(X' \rightarrow X)} = \frac{W_{REM}(X')}{W_{REM}(X)} = e^{-(\beta_{T_j} - \beta_{T_i})(E(q^i) - E(q^j))}$$  \hspace{1cm} (2.22)

Let $\Delta = (\beta_{T_j} - \beta_{T_i})(E(q^i) - E(q^j))$, then the Metropolis acceptance criterion which allows the exchanges to take place is given by,

$$w(X \rightarrow X') = w(x_i^{T_i} | x_j^{T_j}) = \min\{1, e^{-\Delta}\}.$$  \hspace{1cm} (2.23)

In summary, a replica exchange MD simulation is implemented as follows:
Without loss of generality, assume that $T_1 > T_2 > \ldots > T_M$ so that $\beta_1 < \beta_2 < \ldots < \beta_M$, then

1. Simulate each replica $i$ at fixed $T_i$ in parallel and independently for some number
of steps using MC or MD

2. A pair of replicas $i, j$ with neighboring temperatures $T_i, T_j$ are exchanged with probability $w(x^i_{T_i}|x^j_{T_j})$

In the RE scheme, through exchange of replicas, states trapped at local minima are allowed to cross barriers by jumping to higher temperatures. Therefore, a random walk in the temperature space (and thus, PE space) is realized for each replica.

An advantage of RE over ST\textsuperscript{27} is that there is no need for weight factor determination.\textsuperscript{29} Furthermore, RE is suited for implementation on a parallel computing system. Each replica can be assigned to run on a node since the amount of information exchanged between replica (nodes) is minimal. It should be noted here that, unlike conventional MD, dynamic information is lost upon using generalized ensemble sampling methods due to unphysical transitions and exchanges that occur in each simulation.

Simulated Tempering Distributed Replica (STDR)

STDR is a generalized-ensemble simulation method based on ST. We briefly review the methodology below. An in-depth derivation of the algorithm is out of the scope this thesis and is provided elsewhere.\textsuperscript{34,37}

In STDR, individual replicas are coupled through a generalized Hamiltonian containing a potential energy term, the distributed replica potential energy (DRPE), which depends on the distribution of all replicas in the following form:

$$DRPE = c_1 \sum_{m=1}^{M} \sum_{n=1}^{M} [\lambda_{m,\text{linear}} - \lambda_{n,\text{linear}} - \omega(m-n)^2] + c_2 \sum_{n=1}^{M} \lambda_{m,\text{linear}} - \omega \sum_{n=1}^{M} m^2$$ \hspace{1cm} (2.24)

The DRPE enforces a desired sampling distribution of the reaction coordinate. $M$ is the total number of replicas. $m$ and $n$ are the indices representing replicas. The values denoted by $\lambda$ represent a linearly-spaced temperature coordinate. The constant $\omega$ is the
ratio of the number of temperatures to the number of replicas. $c_1$ and $c_2$ are constants and are used to modulate the influence of the DRPE, which enforces the homogeneous temperature sampling of replicas.

The probability of accepting a move from a temperature $T_i$ to a neighbouring temperature $T_j$ is

$$p(T_i \rightarrow T_j) = \min \left\{ \frac{1}{e^{-(\beta_j-\beta_i)E+(a_j-a_i)-(DRPE_j-DRPE_i)}} \right\}$$ (2.25)

If the weight factors are inaccurate, ST results in uneven sampling of the temperature coordinate. However, a key advantage of STDR over ST is that the introduction of the DRPE recovers homogenous sampling and greatly reduces the requirement for accuracy in the dimensionless Helmholtz free energies. Furthermore, due to limitations in the implementations of RE, STDR has been demonstrated to be more computationally-efficient than REMD for sampling the conformational landscape of shorter peptides such as octapeptides.

In Chapters 3 and 4 of this thesis, ST and STDR algorithms were utilized, respectively, to generate the structural ensembles of monomeric $(GA)_4$ and $A\beta(16-22)$ peptides.

### 2.4 Preparing a Biomolecular System for MD Simulation

Coordinates for the initial input into a simulation need to be prepared carefully and appropriately before the MD simulation is performed. In this section, I provide an overview of this process. First, the structure of the protein of interest needs to be generated. If the structure of the protein of interest is known (either from crystallography or NMR), then its spatial coordinates can be extracted from the Protein Databank (PDB). At this point, missing structural information, such as missing residues (usually those participating in
highly-dynamic loops), are modelled into the protein. Furthermore, because a crystal structure does not contain information about hydrogen atoms, they need to be added to the model; the protonation state of basic and acidic residues in the protein needs to be assigned. This can be done by predicting the pKa for each residue or by visually examining the structure for the presence of hydrogen bonds.

After the protein structure is prepared, the simulation environment is prepared to closely mimic \textit{in vitro} experimental conditions: the protein is solvated and other compounds of interest such as ligands, cosolutes, and salt are added to the simulation box. For more complex simulation systems, such as those of membrane proteins which require the presence of a lipid bilayer, additional steps need to be taken before the system is ready for simulating.\textsuperscript{1,24}

Energy minimization is usually performed on the system to remove any bad contacts that could have been introduced during system preparation. An example of a starting state of a simulation is shown in Figure 2.2. Prior to performing production dynamics, the system is first equilibrated to remove any bias inherent in the initial configuration. This equilibration step is usually a short simulation that is performed to allow the solvent molecules to relax around the protein and for the system reach thermal equilibrium. During this period, the system’s temperature, pressure, and volume are closely monitored to ensure that they converge to their intended values. The length of the equilibration simulation is highly dependent on the system of interest. A straightforward method for delineating the equilibration period for a simulation of a protein with well-defined structure is to measure the time-evolution of the root-mean-square deviation (RMSD) of the protein with respect to its crystal structure. Typically, only the data beyond the time at which the RMSD is observed to plateau is included in the final data analysis.
Figure 2.2: An example snapshot of a MD simulation system. The protein (yellow) is shown in a cartoon representation. Cosolute molecules (scyllo-inositol) are rendered in a space-filling representation. Water molecules are shown as blue lines.
Bibliography


[6] B R Brooks, R E Bruccoleri, and B D Olafson. CHARMM: A program for macro-
molecular energy, minimization, and dynamics calculations. *Journal of Computa-

Bürgi, Daan P Geerke, Tim N Heinz, Mika A Kastenholz, Vincent Kräutler, Chris
Oostenbrink, Christine Peter, Daniel Trzesniak, and Wilfred F van Gunsteren. The
GROMOS software for biomolecular simulation: GROMOS05. *Journal of Compu-

Merz, David M Ferguson, David C Spellmeyer, Thomas Fox, James W Caldwell, and
Peter A Kollman. A Second Generation Force Field for the Simulation of Proteins,

1997.

mechanics simulations of proteins based on condensed–phase quantum mechanical

shielding of backbone hydrogen bonds. *Proc Natl Acad Sci USA*, 99(5):2782–2787,
March 2002.

April 2001.


Chapter 3

Binding of Inositol Stereoisomers to Model Peptides

The contents of this section were adapted from an article published in the Journal of Physical Chemistry.


Contributions: Grace Li conducted the research and wrote the section. Régis Pomès provided editorial input and guidance.
Chapter 3. Binding of Inositol Stereoisomers to Model Peptides

3.1 Summary

The self-aggregation of proteins into amyloid fibrils is a pathological hallmark of numerous incurable diseases such as Alzheimer’s disease. Scyllo-inositol is a stereochemistry-dependent in vitro inhibitor of amyloid formation. As the first step to elucidate its mechanism of action, we present molecular dynamics simulations of scyllo-inositol and its inactive stereoisomer, chiro-inositol, with simple peptide models, alanine dipeptide (ADP) and (Gly-Ala)$_4$. We characterize molecular interactions and compute equilibrium binding constants between inositol and ADP as well as, successively, monomers, amorphous aggregates, and fibril-like β-sheet aggregates of (Gly-Ala)$_4$. Inositol interacts weakly with all peptide systems considered, with mM to M affinities, and displaces the conformational equilibria of ADP but not of the (Gly-Ala)$_4$ systems. However, scyllo- and chiro-inositol adopt different binding modes on the surface of β-sheet aggregates. These results suggest that inositol does not inhibit amyloid formation by breaking up preformed aggregates, but rather by binding to the surface of pre-fibrillar aggregates.

3.2 Introduction

Amyloid fibrils formed by various peptides and proteins are known to be associated with neurodegenerative diseases, type II diabetes, and prion-related disorders. In particular, amyloid fibrils of Aβ peptides are found in the extracellular deposits of neuronal plaques and are thought to be central to the pathogenesis of Alzheimer’s Disease (AD), a common and incurable neurodegenerative disease causing dementia and eventual death.

In recent years, amyloid fibril formation was discovered to be a common phenomenon among many proteins in vitro; that is, under certain and denaturing conditions, proteins can self-aggregate to form amyloid fibrils. When viewed with negatively-stained transmission electron microscopy, amyloid fibrils appear as elongated, rope-like structures that are often 100 nm in length. The core structure of all amyloid fibrils is the
cross-β-sheet.\textsuperscript{2,4} At the molecular level, NMR\textsuperscript{1,5} and X-ray crystallography\textsuperscript{6} studies have revealed that the cross-β-structure is comprised of extended polypeptides organized in highly-ordered, in-register β-sheets. Although amyloid fibrils are a pathological hallmark of amyloid-based diseases, smaller nonfibrillar oligomers as little as three or four peptides in size have been demonstrated to display higher cytotoxicity than mature fibrils.\textsuperscript{7–13}

An important strategy to finding a cure to AD and other amyloid diseases is to derive new therapeutic candidates through the rational design of effective small-molecule inhibitors of amyloid formation. In recent years, a number of small molecules capable of preventing aggregation and/or fibril formation have been discovered and have emerged as potential therapeutic approaches for protein misfolding diseases.\textsuperscript{14–19} Interestingly, many of these small molecules share common chemical structural features, such as aromaticity and the presence of multiple hydrogen-bonding groups.\textsuperscript{20–23} However, the molecular basis of the structure-activity relationship of these small molecules is not understood, thus hindering drug development efforts for amyloid-based diseases.

Recently, one such small molecule, scyllo-inositol, has shown promise as a therapeutic for AD.\textsuperscript{24,25} Scyllo–inositol is one of nine stereoisomers that belongs to a class of cyclic polyols called cyclohexanehexols. Four stereoisomers, myo-, epi-, scyllo- and chiro-inositol (Figure 3.1) are physiologically active.\textsuperscript{26} Myo-inositol, the most abundant stereoisomer, plays an important role in signal transduction as precursor of phospholipid headgroups: once phosphorylated, myo-inositol phosphatides act as second messengers in intracellular signal transduction pathways.\textsuperscript{26} Importantly for its therapeutic potential, inositol readily crosses the blood-brain barrier. Myo- and scyllo-inositol are found in tissues of the human central nervous system (CNS), with approximate concentrations of 5 mM and 0.1 to 0.5 mM, respectively.\textsuperscript{27} Accordingly, they are also important osmolytes in the CNS, where alterations in their concentration have been associated with neuropathological conditions.\textsuperscript{26,28}

\textit{In vitro}, inositol stereoisomers stabilize nonfibrillar β-structure and prevent the for-
Figure 3.1: Inositol stereoisomers most commonly found in nature. Stick figures of the stereoisomers were drawn using the ChemDraw software.

The binding of inositol stereoisomers to model peptides is of particular interest because it is known that inositol stereoisomers can affect the aggregation of amyloid fibrils in a stereochemistry-dependent manner: scyllo-, epi- and myo-inositol inhibit Aβ fibril formation, but not chiro-inositol. Moreover, scyllo-inositol was also demonstrated to be the most effective stereoisomer in preventing and reversing AD-like symptoms in transgenic mice while reducing their brain plaque load. Despite this progress, the molecular basis of amyloid inhibition by inositol is not understood. In vitro studies suggest that inositol stereoisomers affects aggregation through direct interaction with Aβ peptides. However, it is not known whether inositol acts on monomeric peptides, non-fibrillar oligomers, or fibrillar aggregates.

Some small molecule inhibitors, including the osmolytes glycerol and trimethylamine N-oxide (TMAO), are known to interfere with in vitro aggregation of amyloidogenic peptides with different sequences, suggesting that generic interactions common to all amyloid-forming peptides and proteins may play a role in the inhibition of amyloid formation. Indeed, small organic osmolytes are hypothesized to modulate protein folding equilibria by interacting with the peptidic backbone, the chemical group common to all polypeptides. Accordingly, the role of backbone solvation in the modulation of pro-
tein folding and aggregation equilibria has recently been highlighted. Furthermore, several studies have suggested that N-methylation of the backbone of amyloidogenic peptides can abolish the formation of amyloid fibrils by preventing intermolecular backbone hydrogen bonding.

Experimental efforts to characterize the molecular interactions of small molecules with amyloid oligomers and fibrils are often impeded due to the non-crystalline, transient, and disordered nature of the aggregates involved. By contrast, molecular simulations are well-suited for studies of proteins involving disorder. Although several molecular dynamics (MD) simulation studies have begun to examine the effect of small molecules on aggregation and fibril formation, the role of backbone binding has not been considered systematically.

In this chapter, we present an MD simulation study of the interaction of inositol with simple model peptides to investigate its stereochemistry-dependent effect on amyloidogenic peptide aggregation and morphology. In a systematic approach, we first characterize the binding equilibria of myo-, epi-, scyllo- and chiro-inositol with alanine dipeptide, a model of the peptidic backbone. Next, to probe the stereochemistry-dependent effect of inositol binding on amyloid aggregation, we study the interaction of scyllo- and chiro-inositol, respectively active and inactive stereoisomers in Aβ amyloid inhibition, successively with monomer, disordered, and fibrillar aggregates of (Gly-Ala)₄ or (GA)₄. (GA)₄ is one of the simplest and shortest amyloidogenic peptides that is known to adopt an extended β-sheet structure both synthetically, as a metallocopolymer, and in nature, in crystalline domains of spider silks. The repetitiveness and simplicity of the peptide sequence allow us to achieve statistically-significant estimates of the binding equilibrium from conventional sampling methods while focusing on the effect of backbone interactions in polypeptide self-aggregation.
3.3 Methods

3.3.1 Simulation Parameters and Protocol

Alanine dipeptide (ADP) was methylated at both the N- and C-terminii. The \((GA)_4\) peptide was acetylated and amidated at the N- and C-termini, respectively. The peptides were built using PyMol\(^\text{53}\) and modelled using the OPLS-AA/L force field\(^\text{54}\). The extended OPLS-AA force field for carbohydrates\(^\text{55}\) was used to model inositol stereoisomers and the TIP3P water model\(^\text{56}\) was used to represent the solvent. Versions 3.3.1 and 3.3.3 of the GROMACS software package\(^\text{57}\) were used to perform unrestrained all-atom MD simulations with the leap frog algorithm using an integration timestep of 2 femtoseconds. Unless otherwise noted, the following parameters were used for all simulations in this study. Electrostatic interactions were calculated using Particle Mesh Ewald (PME) summation with a grid size of 0.15 nm and a real-space cutoff of 1.45 nm.\(^\text{58}\) The Lennard-Jones potential was computed up to 1.3 nm and was switched to zero at 1.4 nm using the GROMACS switch function. The temperature and pressure were controlled at 300 K and at 1 atm using the Berendsen thermostat and pressure coupling scheme, respectively.\(^\text{59}\) Covalent bonds involving hydrogens were constrained using the SHAKE algorithm.\(^\text{60}\) All resultant simulation systems were first subjected to energy minimization and equilibration with isotropic pressure coupling. Replicas of every system were generated with different random seeds for the choice of initial velocities. A trajectory frame was written to disk every picosecond and all frames were used in the final data analysis. Additional details of simulation setup and total sampling time for all systems performed in this study are listed in Table 3.1.

Five initial starting conformations of ADP were obtained by taking a frame every 20 ns from a 100-ns-long simulation of ADP in water. Sets of five independent simulations were carried out successively in the presence of \(\text{myo-}, \text{epi-}, \text{chiro-} \) and \(\text{scyllo-}\)-inositol. The initial conformations of monomeric \((GA)_4\) were taken from an ensemble of monomeric structures.
Table 3.1: Summary of simulation systems.

<table>
<thead>
<tr>
<th>System</th>
<th>$N_{\text{peptides}}$</th>
<th>$N_{\text{inositol}}$</th>
<th>$c_{\text{peptide}}$ (mM)</th>
<th>$c_{\text{inositol}}$ (mM)</th>
<th>$N_{\text{replicas}}$</th>
<th>Time per replica ($\mu$s)</th>
<th>Total time ($\mu$s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>alanine dipeptide with myo-, epi-, chiro- or scyllo-inositol</td>
<td>1</td>
<td>0</td>
<td>61.5</td>
<td>0</td>
<td>5</td>
<td>0.1</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>4</td>
<td>61.5</td>
<td>246</td>
<td>5</td>
<td>0.1</td>
<td>0.5</td>
</tr>
<tr>
<td>(GA)$_4$ monomer with chiro- or scyllo-inositol</td>
<td>1</td>
<td>0</td>
<td>61.5</td>
<td>0</td>
<td>1117</td>
<td>0.005</td>
<td>5.585</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>61.5</td>
<td>123</td>
<td>1117</td>
<td>0.005</td>
<td>5.585</td>
</tr>
<tr>
<td>(GA)$_4$ disordered aggregate (preformed) with chiro- or scyllo-inositol</td>
<td>4</td>
<td>0</td>
<td>246</td>
<td>0</td>
<td>5</td>
<td>0.1</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2</td>
<td>246</td>
<td>123</td>
<td>5</td>
<td>0.08</td>
<td>0.4</td>
</tr>
<tr>
<td>(GA)$_4$ disordered aggregate (dispersed) with scyllo- with chiro-</td>
<td>4</td>
<td>0</td>
<td>246</td>
<td>0</td>
<td>5</td>
<td>0.1</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2</td>
<td>246</td>
<td>123</td>
<td>4</td>
<td>0.08</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(GA)$_4$ fibrillar aggregate with chiro- or scyllo-inositol</td>
<td>16</td>
<td>0</td>
<td>437</td>
<td>0</td>
<td>3</td>
<td>0.1</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>4</td>
<td>437</td>
<td>109</td>
<td>3</td>
<td>0.1</td>
<td>0.3</td>
</tr>
</tbody>
</table>
generated in water at 296 K by simulated tempering distributed replica sampling (STDR) from a previous study.\textsuperscript{61} STDR is a generalized-ensemble simulation method developed in our laboratory, which allows each replica in the simulation to undergo a random walk in temperature to enhance conformational sampling.\textsuperscript{62} The STDR algorithm and implementation are described elsewhere.\textsuperscript{63} A representative set of 1117 structures were chosen from the STDR ensemble at 296 K such that the end-to-end distance probability distribution of this selected subset is similar to the distribution of the entire STDR ensemble of structures (about 12,000 structures in total). These conformations were then used as starting points for simulations at $T = 300$ K in the presence of two molecules of either scyllo- or chiro-inositol. A total of 5 $\mu$s of simulation time was generated for the monomeric systems with either scyllo- or chiro-inositol (Table 3.1). The initial peptide conformations of disordered oligomeric systems were either dispersed monomers drawn from the STDR ensemble at 296K or a preformed $\beta$-sheet oligomer of (GA)$_4$ composed of four peptides.

The (GA)$_4$ peptide in the extended conformation was constructed using PyMol and was used to create the fibril-like $\beta$-sheet model. An eight-stranded antiparallel $\beta$-sheet was constructed by first creating an antiparallel dimer of (GA)$_4$. The principal axis of the dimer was then aligned with the x-dimension of the box and translated along the y-axis to form a single 8-stranded $\beta$-sheet. Two of these 8-stranded sheets were stacked in parallel in a “face-to-back” manner (with all Ala methyl groups facing up in the z direction) and placed in the simulation box such that the first strand at the edge of the $\beta$-sheets was hydrogen-bonded in-register to the nearest periodic image of the eighth strand. The fibril-like systems were first subjected to energy minimization and a 500 ps equilibration stage. Production simulations were performed in the NVT ensemble with final box dimensions of 4 nm x 3.8 nm x 4 nm. Three independent simulations of the (GA)$_4$ fibril-like systems were performed for 100 ns each, successively in the presence and absence of scyllo- and chiro-inositol (see Table 3.1).
3.3.2 Analysis Protocol

The DSSP geometry criteria were used to determine the presence of a hydrogen bond: (1) the distance between donor (D) and acceptor (A) atoms is less than 0.35 nm; (2) the distance between the hydrogen and A is less than 0.25 nm; and (3) the angle formed by D-H-A is greater than 120°. Nonpolar contacts between inositol and peptide were defined by a separation between the center of mass of inositol and the Cβ atom of alanine less than 0.45 nm. The same cut-off was used to compute protein-protein nonpolar contacts between Cβ atoms.

All of the dissociation constants for inositol were calculated based on the presence of intermolecular contacts as defined above. Then, assuming that the binding equilibrium of inositol is

\[
[\text{Protein} \cdot \text{Inositol}] \rightleftharpoons [\text{Protein}] + [\text{Inositol}]
\]

(3.1)

the dissociation constant is the equilibrium constant of this reaction and is given by

\[
K_d = \frac{[\text{Protein}][\text{Inositol}]}{[\text{Protein} \cdot \text{Inositol}]} = f_{ub}[\text{inositol}],
\]

(3.2)

where \(f_{ub}\) is the fraction of unbound over bound peptide states.

The DSSP algorithm was used for the analysis of secondary structure of the disordered oligomer with N- and C-termini of the peptides excluded. The end-to-end distance for a \((\text{GA})_4\) peptide was calculated as the distance between Cα atoms of the N and C-terminus of the peptide. The spatial probability density of inositol is the average spatial occupancy of the atoms of inositol and was computed using the VolMap tool from the Visual Molecular Dynamics (VMD) software package. The planar angle between inositol and the fibrillar model of \((\text{GA})_4\) was computed using the g_angle program from GROMACS analysis tools. All planar angles were corrected to a value between 0° and 90°, using the rule \(\alpha = f(\theta) = 180 - \theta\), if \(\theta > 90°\), otherwise, \(f(\theta) = \theta\). The probabil-
Table 3.2: Summary of equilibrium dissociation constants ($K_d$) for each system in the study.

<table>
<thead>
<tr>
<th>System</th>
<th>Scyllo-$^a$</th>
<th>Chiro-$^a$</th>
<th>$N_{groups, ADP} / N_{groups}$</th>
<th>Scyllo-$^b$</th>
<th>Chiro-$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>alanine dipeptide</td>
<td>1100 (100)</td>
<td>1000 (100)</td>
<td>1.00</td>
<td>1100</td>
<td>1000</td>
</tr>
<tr>
<td>(GA)$_4$ monomer</td>
<td>376 (10)</td>
<td>362 (16)</td>
<td>0.25</td>
<td>275</td>
<td>250</td>
</tr>
<tr>
<td>(GA)$_4$ preformed</td>
<td>85 (12)</td>
<td>89 (8)</td>
<td>0.06</td>
<td>69</td>
<td>63</td>
</tr>
<tr>
<td>(GA)$_4$ dispersed</td>
<td>87 (21)</td>
<td>86 (10)</td>
<td>0.06</td>
<td>69</td>
<td>63</td>
</tr>
<tr>
<td>(GA)$_4$ fibrillar</td>
<td>51 (3)</td>
<td>36 (15)</td>
<td>0.02</td>
<td>17</td>
<td>16</td>
</tr>
</tbody>
</table>

$^a$Each dissociation constant is in units of mM. The standard error is shown within parentheses.

$^b$K$_d$ in units of mM, estimated by scaling the K$_d$ of ADP by the ratio of the number of peptide groups in ADP to the (GA)$_4$ system.

Inositol was found to bind weakly and reversibly to all the peptidic systems considered in our simulations, allowing us to characterize binding equilibria from unbiased sampling.

### 3.4 Results

Inositol was found to bind weakly and reversibly to all the peptidic systems considered in our simulations, allowing us to characterize binding equilibria from unbiased sampling.

#### 3.4.1 Alanine dipeptide

Inositol stereoisomers bound weakly and reversibly to alanine dipeptide with a molar $K_d$. A list of computed dissociation constants for each stereoisomer is shown in Table 3.2. Because all of the results for myo- (1.0 ± 0.1 M), epi- (1.2 ± 0.2 M), chiro- (1.0 ± 0.1 M) and scyllo-inositol (1.1 ± 0.1 M) were within error bars of one another, in this section we provide detailed descriptions and data only for scyllo-inositol. A single molecule of scyllo-inositol can bind the peptidic backbone in either monodentate or bidentate fashion, as
defined by the number of hydrogen bonds between hydroxyl groups of inositol and peptide groups. At a concentration of 250 mM, scyllo-inositol was bound in the monodentate and bidentate modes about 14 ± 1% and 1.1 ± 0.1% of the time, respectively. The dominant bidentate binding modes of scyllo-inositol involved hydrogen bonds formed with the peptide main chain either in the mean plane of the inositol ring (Figure 3.2, panels I-III) or in a “face-to-edge” fashion, where the mean plane of the inositol ring is perpendicular to the plane of the peptide groups (Figure 3.2A, panel IV).

Figure 3.2: Binding of scyllo-inositol to the backbone of alanine dipeptide. A) Main bidentate binding modes. Ramanchandran maps of the conformations of alanine dipeptide sampled in absence of inositol are shown as contours in green. \((\phi, \psi)\) of alanine dipeptide conformers bound by inositol are shown on the map as red crosses. \(\beta\)-conformers (top panels) are bound by inositol through adjacent or non-adjacent (CO,NH) groups; Helical conformers (bottom panels) involve mostly (CO,CO) and (NH,NH) groups. B) Comparisons of the conformational equilibrium of alanine dipeptide for different inositol backbone-bound states.

The Ramanchandran map of ADP is characterized by four dominant basins repre-
senting the α-helical, polyproline II (PPII), and β-sheet conformations. As shown in Figure 3.2A, bidentate-bound ADP adopts backbone dihedral angles that fall within the dominant basins on the ramachandran map, demonstrating that scyllo-inositol is able to bind both helical and β-sheet conformations. Notably, the conformational equilibrium of ADP is shifted in favor of the β-conformer when scyllo-inositol is bound to the peptide backbone in bidentate fashion (Figure 3.2B); in contrast, the relative populations of dominant conformers remained unchanged when inositol is unbound or bound in monodentate form. Taken together, our results show that although binding is weak, inositol may influence peptide conformations by binding to the peptidic backbone. In the next sections, we examine the binding of inositol to monomer and aggregates of a simple β-sheet forming peptide, (GA)₄.

3.4.2 (GA)₄ peptide

In this section, we characterize the binding modes and binding affinity of inositol systematically, first with a peptide monomer, and then with oligomeric and fibrillar aggregates of (GA)₄. Here we examine only the active and inactive stereoisomers scyllo- and chiro-inositol. A summary of the equilibrium binding constants computed for all (GA)₄ systems is shown in Table 3.2.

Monomer

In its monomeric state in solution, (GA)₄ is an intrinsically disordered peptide. An example of scyllo-inositol binding to a monomer is shown in Figure 3.3A. Similar to ADP, binding is weak: the computed dissociation constants were $K_{d,chiro} \approx 362 \pm 16$ mM and $K_{d,scyllo} \approx 376 \pm 10$ mM. Most bound states, 95% for scyllo-inositol and 94.4% for chiro-inositol, involved hydrogen bonds to the backbone (Figure 3.3B). At a concentration of 123 mM, inositol molecules formed a single hydrogen bond about 9% of the time, whereas two or more hydrogen bonds were formed about 4 to 5% of the time. In
Figure 3.3: Binding of *scyllo*-inositol to the monomer of $(GA)_4$. A) Representative snapshot of *scyllo*-inositol forming three hydrogen bonds to a monomer of $(GA)_4$. B) Distribution of the fraction of bound *scyllo*-inositol to polar and nonpolar groups of the monomer. C-D) Conformational equilibrium of $(GA)_4$ as measured by the peptide end-to-end distance distribution. The distributions are all within error bars of each other and are plotted separately by the number of hydrogen bonds for *scyllo*- in C) and *chiro*-inositol in D). For clarity, error is only shown for the Peed curve where $n=1$. 
contrast, nonpolar contacts are less frequent and, alone, account for only 3% of bound scyllo- and chiro-inositol (Figure 3.3B). In total, the peptide monomer is bound to at least one molecule of inositol approximately 25% of the time, 23% of the time with a inositol:peptide stoichiometry of 1:1 and only \( \sim 2\% \) of the time with a 2:1 stoichiometry. Contrary to ADP, the presence of inositol did not have a significant effect on the conformational equilibrium of monomeric (GA)\(_4\) (Figure 3.3C-D).

### Disordered Oligomer

![Disordered Oligomer](image)

**Figure 3.4:** Binding of scyllo-inositol to the disordered oligomer of (GA)\(_4\). A) Snapshot of scyllo-inositol simultaneously hydrogen bonding two peptides in a disordered (GA)\(_4\) aggregate. Hydrogen bonds to the backbone are drawn as red lines. B) Distribution of secondary structure content, as classified by the DSSP algorithm, for preformed and dispersed starting states of the oligomer. Hydrogen bonds to the backbone are drawn as red lines. C) Fraction of scyllo-inositol interacting with the disordered oligomer via hydrogen bonds (HBs), nonpolar contacts, or both. The fractions for chiro-inositol are similar (data not shown).

To probe whether inositol affects the structure and aggregation of small oligomers of (GA)\(_4\) in solution, we performed sets of simulations involving two distinct starting states
of four (GA)₄ peptides: (1) initially monodispersed peptides and (2) a preformed β-sheet aggregate. In the initially dispersed systems, the peptides rapidly aggregated to form a disordered oligomer (Figure 3.4A) in which the majority of the residues (≈60%) retained a coil conformation (Figure 3.4B). Similarly, systems initiated with a preformed 4-stranded β-sheet also evolved into a disordered oligomer over the course of the simulation. Only about 5% and 10% of the residues participated in a β-sheet or in a β-bridge, respectively (Figure 3.4B).

Figure 3.5: Time evolution of peptide-peptide nonpolar and hydrogen bonding contacts in disordered aggregates in presence and absence of inositol (control). Each curve is smoothed using a running average over a window with a length of 500 ps. Results for the dispersed monomer aggregates are shown on the left and the preformed β-sheet aggregates on the right. Top: the total number of inter- and intra-molecular hydrogen bonding contacts; Bottom: Number of intermolecular nonpolar contacts.

Despite different initial conditions and independently of the presence of inositol, all
aggregates evolved to a similar morphology. The total number of peptide-peptide non-
polar and polar contacts formed within the oligomer converged to similar values for both
the dispersed and preformed oligomers and did not change with time (Figure 3.5). As
shown in Figure 3.5 (top panels), the average total number of intermolecular hydrogen
bonds (~8 ± 1) was consistently higher than the number of intramolecular hydrogen
bonds (~2.1 ± 0.3). On average, about 4.3 ± 0.4 nonpolar contacts were formed upon
aggregation in the absence of inositol compared to 4.4 ± 0.5 contacts for scyllo-, and
5.0 ± 0.4 for chiro-inositol (data not shown for the preformed oligomer). When taken
together, the above results show that the presence of scyllo- and chiro-inositol neither
prevented aggregation nor disrupted the preformed oligomer.

Dissociation constants of about 80 mM to aggregates of type 1 and 2 were obtained
for both scyllo- and chiro-inositol. The $K_d$ calculated for each aggregate type is shown
in Table 3.2. In the presence of multiple aggregated chains, a single molecule of inositol
was found to cross-link multiple peptides by simultaneously hydrogen bonding to their
backbones (Figure 3.4A). Similar to monomers, at a concentration of 123 mM, chiro- and
scyllo-inositol were bound predominantly to the backbone: 96% of bound scyllo-inositols
formed only hydrogen bonding contacts (~94% for chiro-inositol), whereas 2% (3% for
chiro-inositol) were involved in nonpolar contacts only (Figure 3.4C).

**Fibril-like oligomer**

In order to probe the binding modes of inositol with a fibril-like aggregate of (GA)$_4$,
we constructed an infinite $\beta$-sheet, where the $\beta$-strands at the edge of an octameric $\beta$-
sheet are hydrogen-bonded to each others nearest periodic image. A single unit of this
periodic model consisted of a stack of two antiparallel and in-register $\beta$-sheets, with eight
strands per sheet (Figure 3.6A,B). Although some of the hydrogen bonds defining the
$\beta$-sheet structure occasionally broke, in the absence of inositol the protofibril remained
approximately planar and aggregated as an infinite fibril throughout the simulation.
Figure 3.6: Binding of scyllo- and chiro-inositol to the fibrillar aggregate of (GA)$_4$. Different views of the initial starting structure of the fibril-like model. Top and bottom sheets are colored in grey and in cyan, respectively. A top down view is depicted in A) showing the backside of the top (GA)$_4$ sheet. A side view of the protofibril is shown in B). The spatial probability density of bound scyllo-inositol (yellow) C) and chiro-inositol (orange) D) are shown overlapping with the fibril. The density is shown at an occupancy isosurface value of 3% for both stereoisomers. E) is the percentage of bound scyllo- and chiro-inositol to polar and nonpolar groups on the β-sheet.
The spatial distribution of bound inositol molecules shows that both chiro- and scyllo-inositol bind at the surface of the fibril (Figures 3.6C, D). Chiro- and scyllo-inositol bound fibrillar aggregates of (GA)$_4$ with a $K_d$ of 36 ± 15 mM and 51 ± 3 mM, respectively. The apparent increase in affinity compared to amorphous aggregates can be attributed to the following reasons. First, the fibrillar aggregate presents a much larger effective surface area than both the monomer and the disordered oligomer (Figures 3.6C, D). Moreover, a larger fraction of the alanine side chains are completely solvent-exposed in the fibril-like aggregate, increasing the fraction of bound conformations involving only nonpolar contacts by nearly an order of magnitude, from 2% in the disordered oligomer to 12% for scyllo- and 18% for chiro-inositol in the fibrillar aggregate in the presence of 109 mM of inositol (Figures 3.3B and 3.6E). Accordingly, a higher fraction of scyllo-inositol, 83 ± 1%, versus 77 ± 1% for chiro-inositol, was found to form hydrogen bonds, where the 6% drop in the hydrogen-bonded-only population of chiro-inositol was compensated by a commensurate increase in the nonpolar-bound-only population of chiro-inositol (Figure 3.6E).

Thus, although chiro- and scyllo-inositol have similar binding constants, they have different binding modes to fibrillar aggregates, a feature not previously observed for the monomer and the disordered oligomer of (GA)$_4$. Both scyllo- and chiro-inositol form nonpolar contacts and backbone hydrogen bonds in poses where the mean plane of the inositol ring lies parallel, at an angle, or perpendicular to the plane of the fibril (Figure 3.7). Furthermore, two or more molecules of inositol may cluster together and bind at the surface of the sheet (Figures 3.7C,D). However, as shown in Figure 3.8A, scyllo-inositol adopts specific binding orientations whereas chiro-inositol does not: scyllo- preferentially binds in either nearly flat ($\alpha = 20^\circ$) or upright ($\alpha = 65^\circ$) to the sheet, whereas chiro-inositol does not have such a bimodal preference and binds the fibril at an average angle of $\alpha = 45^\circ$. This stereochemistry-modulated difference in binding specificity explains the somewhat higher fraction of nonpolar binding by chiro-inositol (Figure 3.6E): chiro- is
more likely than scyllo-inositol to bind at angles of $30^\circ < \alpha \leq 60^\circ$ (Figure 3.8A), where 24% of bound chiro- (versus 16% for scyllo-inositol) is bound by nonpolar contacts only (Figure 3.8B). For $\alpha \leq 30^\circ$, the distributions of scyllo- and chiro-inositol bound to polar and nonpolar groups are similar (data not shown). Moreover, because chiro-inositol has a partially-nonpolar edge whereas scyllo-inositol does not, binding in the upright position also involves more nonpolar interactions for chiro- than for scyllo-inositol (Figure 3.8B). Finally, although inositol was observed to bind at the surface, binding did not change the morphology of the fibrillar aggregate.

### 3.5 Discussion

In the above analysis, we have systematically characterized the association of stereoisomers scyllo-, epi-, myo- and chiro-inositol with alanine dipeptide, a simple model of the peptidic backbone. Furthermore, we examined the binding of scyllo- and chiro-inositol
Chapter 3. Binding of Inositol Stereoisomers to Model Peptides

Figure 3.8: Binding mode and orientation of scyllo- and chiro-inositol to the fibril of (GA)$_4$. The distribution of inositol to sheet planar angles is depicted in A). B) Inositol binding to nonpolar and polar groups as classified by $\alpha$, the angle at which inositol molecules bind at the surface of the fibrillar (GA)$_4$ (see Methods).

to various aggregated states of (GA)$_4$ to probe the role of backbone binding in amyloid inhibition. Our results show that inositol exhibits weak binding with dissociation constants in the range of 0.04 M to 1 M to the different peptides and aggregation states considered.

Furthermore, the $K_d$ of inositol increases linearly with the number of peptide groups in the system (Table 3.2), indicating that inositol does not bind cooperatively to the monomer and aggregate states of (GA)$_4$ considered. As expected, inositol binds most weakly to alanine dipeptide, with a value about 4 times smaller than the $K_d$ of urea to N-acetyl alanine reported recently in the literature (0.3 M for urea$^{66}$ vs 1.1 M for scyllo-inositol). Taken together, our results indicate that the activity of inositol stereoisomers is similar to that of osmolytes, which typically have binding constants in the millimolar to molar range.$^{38,67}$

Moreover, our results are consistent with the hypothesis that osmolytes influence protein and peptide folding and stability through direct binding rather than by modifying
The spacing of consecutive OH groups of inositol is well-suited to bidentate interactions with adjacent groups of the polypeptide backbone (Figure 3.2A). Our findings shown in Figure 3.2A are consistent with similar binding modes observed in a recent ab initio simulation and IR spectroscopic study of the binding of glucose epimers to the phenylalanine dipeptide backbone. Furthermore, inositol stereoisomers displace the backbone conformation of alanine dipeptide towards extended β-strand conformations (Figure 3.2B). However, neither scyllo- nor chiro-inositol had a significant effect on the conformational equilibrium of the (GA)₄ monomer. Taken together, these results indicate that inositol may not act as a drug by directly influencing monomer conformations. However, our results do not preclude the possibility that inositol may block fibril elongation by preferentially binding to monomers that are constrained to extended conformations, such as those at exposed edges of β-sheets.

Independently of the presence of inositol, both the preformed β-sheet oligomer and the monodisperse solution of (GA)₄ evolved into a similar morphology (Figure 3.4B and 3.5) with only a small amount of β-structure (Figure 3.4B), indicating that small aggregates of (GA)₄ are likely to be disordered. Unlike the hydrophobic core of the Aβ peptide, (GA)₄ is a shorter and more polar peptide that is capable of forming more hydrogen bonds than nonpolar contacts. Our results show that peptide-peptide hydrogen bonding play an important role in the aggregation of (GA)₄ peptides in solution (Figure 3.5). Because neither stereoisomer disrupted the aggregates of (GA)₄, our results indicate that inositol is unlikely to inhibit fibril formation by breaking up preformed aggregates. Therefore, we conclude that inositol is unlikely to inhibit fibril formation by binding monomers and small disordered oligomers since binding appears to be weak, non-cooperative, and stereochemistry-independent.

By contrast, although the dissociation constants were similar for both scyllo- and chiro-inositol, binding specificity and binding modes involving nonpolar groups of the fibrillar aggregate of (GA)₄ were modulated by the stereochemistry of inositol. A signifi-
significantly higher fraction of chiro-inositol than scyllo-inositol was bound to nonpolar groups of the fibrillar aggregate (Figure 3.6E). Moreover, scyllo-inositol exhibited a bimodal distribution of binding orientations, with a significant preference for orientations in which the ring of inositol is either parallel or perpendicular to the mean surface of the β-sheet over chiro-inositol (Figure 3.7). As a direct consequence of the presence of axial hydroxyl groups, chiro-inositol is more likely to bind at angles that promote contact with nonpolar groups at the surface of the fibrillar aggregate, whereas the more specific binding modes of scyllo-inositol favors backbone binding. Since this is the only stereochemistry-dependent result of our study, we speculate that scyllo-inositol acts on ordered β-sheet aggregates (as opposed to disordered oligomers or monomers). Moreover, these findings suggest a possible mechanism of action whereby a significant binding affinity to specific side chains on the surface of fibrillar aggregates could lead to the inhibition of β-sheet stacking (and therefore, amyloid fibril growth or maturation) by scyllo-inositol. Similarly, different binding modes observed in MD simulations of Aβ42 fibrils have been proposed to explain differences in binding affinities between Thioflavin T, a well-known amyloid-binding dye, and its chemical analogs.70,71

A factor that we have not considered in this study is the influence of inositol:peptide molar ratio on binding and inhibition. In vitro, the inhibition activity of scyllo-inositol was observed at an inositol:peptide ratio of 25:1, where inositol stereoisomers were present in excess of Aβ at concentrations of 0.25 mM to 5 mM.25 Although our simulations had effective concentration of inositol an order of magnitude higher than in these experiments, it is possible that we have precluded cooperative inositol binding modes by limiting the number of inositol molecules present in the small simulation cell. Furthermore, $K_d$ values obtained from our simulations of (GA)$_4$ were approximately two orders of magnitude higher than measured for Aβ. Based on our results, the predicted $K_d$ of (GA)$_{21}$, a Gly-Ala repeat peptide similar in length to Aβ, would be 1200 mM/21 = 57 mM, which is still an order of magnitude greater than in vitro inhibitory concentrations. This indi-
cates that scyllo-inositol is unlikely to inhibit $\beta$-sheet formation by $(GA)_4$ peptides, and more importantly, that backbone binding by small molecules may not be sufficient for inhibition of amyloid formation. In future studies, elucidating the relationship of binding cooperativity and amyloid inhibition by approaching experimental drug:protein molar ratios, as well as elucidating the sequence specificity of inositol binding to amyloid fibrils, will provide further insight that may be used in the rational drug design of improved inhibitors.

### 3.6 Conclusions

We have performed systematic simulations of simple amyloidogenic peptide models with both active and inactive stereoisomers of inositol to examine the molecular basis of amyloid inhibition. Our results indicate that although peptide backbone dominates the interaction with inositol, the binding affinity is low and remains in the millimolar range. Moreover, this property is independent of stereochemistry and does not appear to be sufficient to impede peptide dimerization through intermolecular backbone hydrogen bonding. Taken together, our results suggest that amyloid inhibition by inositol cannot be accounted for by generic binding to the peptidic backbone alone and is likely to involve sequence-specific interactions with amino-acid side chains as well as binding to specific aggregate morphologies. Accordingly, although the formation of intermolecular hydrogen bonds is the predominant interaction in protein aggregates composed of $(GA)_4$, amyloidogenic peptides involved in amyloid diseases are often more hydrophobic and in general, self-aggregation is driven largely by the hydrophobic effect.\(^2\) In forthcoming studies, we will examine the role of sequence-specific interactions between inositol and aggregates of pathogenic peptides.
Bibliography


[53] The pymol molecular graphics system, version 1.2, Schrodinger Inc.


Chapter 4

Binding Mechanism of Inositol Stereoisomers to Monomers and Aggregates of Aβ(16-22)

The contents of this section were adapted from an article published in the Journal of Physical Chemistry.


Contributions: Grace Li conducted the research and wrote the section. Régis Pomès provided editorial input and guidance.
Chapter 4. Binding of Inositol Stereoisomers to Aβ(16-22)

4.1 Summary

Alzheimer’s disease (AD) is a severe neurodegenerative disease with no cure. A potential therapeutic approach is to prevent or reverse the amyloid formation of Aβ42, a key pathological hallmark of AD. We examine the molecular basis for stereochemistry-dependent inhibition of the formation of Aβ fibrils in vitro by a polyol, scylo-inositol. We present molecular dynamics simulations of the monomeric, disordered aggregate, and protofibrillar states of Aβ(16-22), an amyloid-forming peptide fragment of full-length Aβ, successively with and without scylo-inositol and its inactive stereoisomer chiro-inositol. Both stereoisomers bind monomers and disordered aggregates with similar affinities of 10–120 mM, whereas binding to β-sheet-containing protofibrils yields affinities of 0.2–0.5 mM commensurate with in vitro inhibitory concentrations of scylo-inositol. Moreover, scylo-inositol displays a higher binding specificity for phenylalanine-lined grooves on the protofibril surface, suggesting that scylo-inositol coats the surface of Aβ protofibrils and disrupts their lateral stacking into amyloid fibrils.

4.2 Introduction

One in eight people over the age of 65 has Alzheimer’s Disease (AD), a progressive neurodegenerative disease that currently has no cure.\(^1\) The amyloid cascade hypothesis states that the extracellular neuronal deposition of Aβ amyloid plaque plays a central role in the pathogenesis of AD.\(^2\) Aβ is a peptide proteolytically cleaved from the amyloid precursor protein and is produced as two common alloforms, Aβ40 or Aβ42, which are 40 and 42 residues in length, respectively. In the diseased state, Aβ42 levels are elevated and the peptides deposit as extracellular Aβ plaques.\(^3\)\(^4\)

Aβ40 and Aβ42 are intrinsically-disordered peptides that self-aggregate in vitro to form amyloid fibrils. Amyloid fibrils are protein aggregates with a characteristic cross-β
structure, which consists of in-register $\beta$-sheets with backbone hydrogen bonds running parallel to the long axis of the fibril.\textsuperscript{5} Moreover, smaller fragments of the full-length A$\beta$ sequence are also found to form amyloid \textit{in vitro}.\textsuperscript{6,7} In particular, one of the shortest amyloid-forming peptides structurally characterized using solid-state NMR is KLVFFAE or A$\beta$(16-22).\textsuperscript{6} The residues LVFFA are believed to form the central hydrophobic core critical for the initiation of aggregation and fibril formation in the full-length A$\beta$ peptide.\textsuperscript{8} Furthermore, single-point mutations in this region greatly affect the aggregation propensity of A$\beta$: familial mutations E22Q, E22K, and E22G, known as “Dutch”, “Italian”, and “Arctic” mutations, respectively, significantly accelerate fibril formation,\textsuperscript{9} whereas the mutation F19T abolishes the formation of fibrils \textit{in vitro}.\textsuperscript{10}

Amyloid fibril formation follows a complex pathway: prior to the appearance of fibrils \textit{in vitro}, amyloidogenic monomers self-aggregate into a variety of pre-fibrillar intermediate morphologies. While the fibril is an important state implicated in AD, recent research has shown that soluble oligomers as small as dimers and tetramers play a role in neurotoxicity.\textsuperscript{11} In recent years, drug development and research efforts have been directed toward the development of therapeutic agents to prevent the self-aggregation and amyloid formation of A$\beta$, a promising treatment approach to target the underlying disease.\textsuperscript{1,12,13} As a result, many different types of \textit{in vitro} amyloid inhibitors have been discovered, including peptides,\textsuperscript{14–18} immunotherapies,\textsuperscript{2,19} polyphenolic molecules,\textsuperscript{20–22} and other small molecules.\textsuperscript{20,23–25} These approaches have been reviewed in detail elsewhere.\textsuperscript{1,13}

\textit{Scyllo}-Inositol is a small-molecule inhibitor of A$\beta$-fibrillation developed for the treatment of AD.\textsuperscript{26–29} Inositol is a class of cyclohexylpolyols, of which eight out of nine stereoisomers are commonly found in nature. \textit{Scyllo}-Inositol, with all hydroxyl groups equatorial, is the only isomer with two planar hydrophobic faces. By contrast, its diastereoisomer, \textit{chiro}-inositol, with two adjacent axial hydroxyl groups, has two nonplanar hydrophobic faces. \textit{myo}-Inositol, the most common inositol stereoisomer, is found at high concentrations ($\sim$5 mM) in the tissues of the human central nervous system (CNS).\textsuperscript{30}
Like myo-inositol, scyllo-inositol is present in the brain and can be passively and actively transported across the blood-brain barrier.\textsuperscript{28} Importantly, scyllo-inositol was demonstrated to prevent and reverse AD-like symptoms in a transgenic mouse model of AD.\textsuperscript{26} Because of the positive CNS bioavailability and favorable \textit{in vivo} toxicity profile of inositol, both of which are rare and essential properties of putative AD drug candidates, inositol-based therapies represent a unique and promising approach for the treatment of AD. Phase II of clinical trials for scyllo-inositol (ELN0005) in North America was fast-tracked in 2007 by the United States Food and Drug Administration and was completed in 2011.\textsuperscript{29,31}

\textit{In vitro}, inositol displays stereochemistry-dependent inhibition of A\textsubscript{β}42 fibrils: myo-, epi- and scyllo-inositol were shown to inhibit A\textsubscript{β}42 fibrillation at concentrations of 1–5 mM,\textsuperscript{27} whereas chiro-inositol is inactive below molar concentrations.\textsuperscript{19} Moreover, upon incubation of monomeric A\textsubscript{β}42 with scyllo-inositol, circular dichroism spectroscopy indicated the formation of β-sheet structure at an inositol:peptide molar ratio of 25:1.\textsuperscript{27} Although inositol stereoisomers have been proposed to inhibit amyloid formation by directly interacting with either monomers or non-fibrillar aggregates to “cap off” fibril growth,\textsuperscript{19} the molecular basis of the effect of scyllo-inositol and its stereoisomers on A\textsubscript{β} amyloid formation is currently unknown.

Thus far, experimental efforts to characterize the molecular structure of non-fibrillar oligomers have been impeded by their transient and disordered nature. In turn, the lack of information on the molecular structure of amyloid oligomers hampers experimental determination of the modes of action of inositol. Molecular dynamics (MD) simulations, by contrast, are well-suited for studies of disordered proteins and can provide atomic-level insight into the mechanism of peptide self-aggregation.\textsuperscript{32–38}

MD simulations were previously employed to examine the binding mechanism of other small-molecule inhibitors such as polyphenols,\textsuperscript{39,40} non-steroidal anti-inflammatory drugs\textsuperscript{41,42}, and the well-known amyloid dye thioflavin T\textsuperscript{43,44} to monomers and/or fibrillar
forms of Aβ. Because of the existence of multiple aggregation states, small-molecule inhibitors may have multiple modes of action and may act by binding either to monomers or to non-fibrillar or fibrillar oligomers in the fibrillation pathway. Furthermore, their inhibitory activity may also be affected both by the concentration of the ligand and by the ligand:peptide molar ratio. For example, it has been suggested that the ability of small molecules (−)-epigallochatechin gallate (EGCG) and ibuprofen to inhibit amyloid fibrillation is modulated by the ligand:peptide molar ratio. However, thus far, few MD simulation studies have examined the effect of ligand concentration on different relevant aggregation states along the amyloid fibrillation pathway.

In a previous study, we investigated the stereochemistry-dependent binding of inositol with alanine dipeptide, a model of the peptide backbone, and (GA), a simple β-sheet-forming peptide. Weak binding, with equilibrium constants (0.04 – 1 M) commensurate with those of osmolytes, was found for inositol with both peptides and all aggregation states considered, indicating that backbone binding alone is likely to be insufficient for amyloid inhibition. However, that study uncovered stereochemistry-dependent binding modes between inositol and nonpolar groups on the surface of (GA) fibril-like aggregates, which suggests that both aggregate morphology and sequence-specific interactions could play an important role in Aβ-aggregation inhibition by inositol.

In this chapter, we consider the role of sequence-specific interactions of inositol by examining its binding to Aβ(16-22), an amyloidogenic peptide that is part of the central hydrophobic core of fibrillar Aβ. Because the amyloidogenic species with which inositol may interact are not known, we successively examine its binding to three different morphologies: monomer, disordered oligomer, and protofibrillar-like aggregate (β-oligomer). Using a systematic comparative approach, MD simulation studies of each of the aforementioned states are successively carried out in the presence and absence of scyllo-inositol and its inactive stereoisomer, chiro-inositol. Moreover, we examine the effect of varying inositol:peptide molar ratios on the binding equilibria of inositol to monomers and aggre-
Table 4.1: Summary of simulation systems

<table>
<thead>
<tr>
<th>System</th>
<th>$N_{\text{peptides}}$</th>
<th>$N_{\text{inositol}}$</th>
<th>$c_{\text{peptide}}$ (mM)</th>
<th>$c_{\text{Inositol}}$ (mM)</th>
<th>molar ratio</th>
<th>$N_{\text{replicas}}$</th>
<th>Total time ($\mu$s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glu dipeptide</td>
<td>1</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>5</td>
<td>0.5</td>
</tr>
<tr>
<td>with chiro- or scyllo-inositol</td>
<td>1</td>
<td>4</td>
<td>-</td>
<td>246</td>
<td>4:1</td>
<td>5</td>
<td>0.5</td>
</tr>
<tr>
<td>Phe dipeptide</td>
<td>1</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>5</td>
<td>0.5</td>
</tr>
<tr>
<td>with chiro- or scyllo-inositol</td>
<td>1</td>
<td>4</td>
<td>-</td>
<td>246</td>
<td>4:1</td>
<td>5</td>
<td>0.5</td>
</tr>
<tr>
<td>$\text{A}\beta$(16-22) monomer (STDR)</td>
<td>1</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>33</td>
<td>3.56</td>
</tr>
<tr>
<td>$\text{A}\beta$(16-22) monomer</td>
<td>1</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>1117</td>
<td>5.59</td>
</tr>
<tr>
<td>with chiro- or scyllo-inositol</td>
<td>1</td>
<td>2</td>
<td>-</td>
<td>123</td>
<td>2:1</td>
<td>1117</td>
<td>5.59</td>
</tr>
<tr>
<td>disordered aggregate</td>
<td>4</td>
<td>0</td>
<td>104</td>
<td>0</td>
<td>-</td>
<td>8</td>
<td>1.44</td>
</tr>
<tr>
<td>with chiro- or scyllo-inositol</td>
<td>4</td>
<td>2</td>
<td>104</td>
<td>52</td>
<td>2:4</td>
<td>5</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>15</td>
<td>18</td>
<td>70</td>
<td>15:4</td>
<td>8</td>
<td>1.44</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>45</td>
<td>18</td>
<td>209</td>
<td>45:4</td>
<td>5</td>
<td>1.00</td>
</tr>
<tr>
<td>$\beta$-oligomer</td>
<td>16</td>
<td>0</td>
<td>148</td>
<td>0</td>
<td>-</td>
<td>1</td>
<td>0.13</td>
</tr>
<tr>
<td>with chiro- or scyllo-inositol</td>
<td>16</td>
<td>4</td>
<td>148</td>
<td>37</td>
<td>4:16</td>
<td>18</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>64</td>
<td>15</td>
<td>62</td>
<td>64:16</td>
<td>6</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>64</td>
<td>52</td>
<td>208</td>
<td>64:16</td>
<td>6</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Gates of $\text{A}\beta$(16-22). From a total of 24.5 $\mu$s of simulation, we compute binding constants ($K_{eq}$) and characterize the binding modes of inositol to the different peptide aggregation states considered. The results of our study have implications for the mechanism of amyloid inhibition by small molecules and for the rational design of more efficacious putative therapeutics for AD and related amyloid disorders.

4.3 Materials and Methods

4.3.1 Simulation Parameters and Protocol

To eliminate terminal charge effects, the $\text{A}\beta$(16-22) peptide was acetylated and amidated at the N- and C-termini, respectively. The peptide was represented by the OPLS-AA/L force field. The extended OPLS-AA force field for carbohydrates was used to model...
inorganic stereoisomers. The TIP3P water model\textsuperscript{51} was used to represent the solvent. To mimic \textit{in vitro} fibrillation conditions used in the study by Balbach \textit{et al.}\textsuperscript{6}, no salt was added to the aqueous solution. All MD simulations were performed in the \textit{NpT} ensemble using the GROMACS simulation package,\textsuperscript{52,53} versions 3.3.x and 4.0.x. Unless otherwise noted, the following parameters were used for all simulations in this study. The leapfrog Verlet integration algorithm was used with an integration time step of 2 fs. Long-range electrostatic interactions were calculated using Particle Mesh Ewald summation with a Fourier grid spacing of 0.15 nm and a real-space cutoff of 1.3 nm.\textsuperscript{54} The short-range nonbonded van der Waals interactions were switched to zero from 1.1 to 1.2 nm. The temperature was controlled at 300 K using the Berendsen barostat.\textsuperscript{55} Pressure was controlled by the Berendsen thermostat at 1 atm with a coupling constant of 1.0 ps.\textsuperscript{55} The SHAKE algorithm was used to constrain covalent bonds containing hydrogens.\textsuperscript{56} In all simulations, a cubic box was used with periodic boundary conditions. Prior to data collection, 500 steps of energy minimization were first performed using the conjugate-gradient algorithm, followed by equilibration with isotropic pressure coupling. The center of mass (COM) rotation and translation were removed at every step.

Molecular simulations of monomeric A\textsubscript{β}(16-22) in water were performed in the absence of inositol using the simulated tempering distributed replica sampling algorithm (STDR).\textsuperscript{57} STDR is a generalized-ensemble simulation method that allows each replica in the simulation to undergo a random walk in temperature to enhance conformational sampling.\textsuperscript{57,58} The STDR simulation was performed using 33 replicas undergoing canonical sampling (\textit{NVT} ensemble) at exponentially-spaced temperatures ranging from 280 to 694 K. A total of 108 ns of simulation at each temperature were generated using Langevin dynamics (implemented by the stochastic dynamics integrator in GROMACS 3.3.x), for a total simulation time of 3.564 \textit{µs}.

A set of 1117 structures was drawn randomly from STDR simulations such that the probability distribution of the end-to-end distance of these peptides closely approximated
that of the equilibrium ensemble of KLVFFAE at 296 K. These structures were used as starting points for constant-temperature MD simulations \((NpT\) ensemble) in the presence of 123 mM inositol at an inositol:peptide molar ratio of 2:1. Short, 5-ns MD simulations were performed for each structure in the presence and absence of inositol at \(T = 300\) K. In addition, 15 ns of simulation in the presence of \(scyllo\)- or \(chiro\)-inositol molecules at inositol:peptide molar ratios of 15:1 were performed using each of 550 structures drawn randomly from the larger set of 1117 structures.

Total sampling times of 1.44 and 1.5 \(\mu\)s were generated for disordered aggregates and \(\beta\)-oligomers of \(A\beta(16-22)\), respectively. Each of the disordered aggregate simulations was initiated from four peptide conformations drawn at random from the pool of structures obtained at \(T = 296\) K from the STDR simulation of the monomer. Peptides were initially monodisperse and placed approximately equidistant from each other in the simulation box. Successively 2, 15, and 45 molecules of inositol were added at inositol:peptide molar ratios of 1:2, 15:4, and 45:4, respectively.

The \(A\beta(16-22)\) \(\beta\)-oligomer consists of two eight-stranded antiparallel \(\beta\)-sheets stacked in a “face-to-face” and antiparallel manner and was constructed based on solid-state NMR evidence\(^6\) using a method similar to that described in our previous study.\(^3^4\) Consistent with the experimental study, the \(\beta\)-sheets were stacked so that charged side chains, lysine (Lys) and glutamate (Glu), are located on the solvent-exposed faces of the \(\beta\)-oligomer.

Simulations of the \(\beta\)-oligomer were performed successively in the presence of \(scyllo\)- and \(chiro\)-inositol at inositol:peptide molar ratios of 4:16 and 64:16 using six \(A\beta(16-22)\) \(\beta\)-oligomer structures each taken from every 10 ns of a 100-ns long trajectory in the absence of inositol. Simulations at the lower molar ratio of 4:16 were performed at a concentration of 37 mM. For the higher molar ratio, two separate sets of simulations were performed, one at an inositol concentration of 62 mM (approximately the concentration of the low-molar-ratio simulations) and the other at 208 mM, corresponding, respectively to 15 and 64 molecules of inositol in the simulation cell (Table 4.1). A summary of the
production runs used for the analysis of all systems investigated in this study is provided in Table 4.1.

4.3.2 Analysis Protocol

The binding reaction of inositol defined by

$$[\text{Protein} \cdot \text{Inositol}] \rightleftharpoons [\text{Protein}] + [\text{Inositol}],$$

has an associated equilibrium constant of

$$K_{eq} = \frac{[\text{Protein}][\text{Inositol}]}{[\text{Protein} \cdot \text{Inositol}]}.$$

The equilibrium constant for inositol binding, $K_{eq}$, was calculated based on the presence of intermolecular contacts (either hydrogen bonding or nonpolar) as defined below. The DSSP hydrogen-bonding criteria were used to determine the presence of a hydrogen bond: (1) the distance between donor and acceptor atoms is less than 0.35 nm; (2) the distance between the hydrogen and the acceptor is less than 0.25 nm; and (3) the angle formed by the donor, hydrogen, and acceptor is greater than 120°.59

Nonpolar contacts between inositol and peptide were calculated by considering all nonpolar carbon atoms of amino-acid side chains and carbon atoms of inositol within 0.45 nm and were normalized by the number of peptides present in the system. The total number of intermolecular peptide-peptide nonpolar contacts was calculated by considering all side chain carbon atom pairs within 0.45 nm.

The potential of mean force (PMF) for the binding of scyllo-inositol and chiro-inositol to phenylalanine (Phe) side chains was computed using two reaction coordinates: (1) the distance between the center of mass (COM) of inositol and the COM of the Phe side chain (excluding the $C_\beta$ atom), $r$; and (2) the angle between the mean plane of the cyclohexane ring of inositol and that of the benzene ring of Phe, $\theta$. A molecule of scyllo-
Inositol is considered to be stacked to Phe if $\theta < 20^\circ$ and $r < 0.45$ nm. The PMF is given by $W(r, \theta) = -RT \ln \rho(r, \theta)$, where $R$ is the gas constant, $T$ is the temperature, and $\rho(r, \theta)$ is the probability distribution of $r$ and $\theta$. All error bars were estimated using block averaging or by computing the standard deviation in the mean of the property of interest over all independent simulations.

Inositol clusters were computed using the g_clustsize analysis tool from the GROMACS software package using an atomic cutoff of 0.35 nm. The DSSP algorithm was used for the analysis of the secondary structure of the disordered oligomer with the N- and C-termini of the peptides excluded. The distance between the first and last $C_\alpha$ atoms of the peptide chain defines the end-to-end distance. The spatial probability density of inositol was computed using the VolMap tool from the Visual Molecular Dynamics (VMD) software package.

### 4.4 Results

In the sections below, we successively characterize the binding equilibrium of inositol and its effect on the morphology of monomers and of disordered and protofibrillar oligomers of Aβ(16-22).

#### 4.4.1 Monomer

We performed simulations of an Aβ(16-22) monomer successively in pure water and in the presence of scyllo- and chiro-inositol at inositol:peptide molar ratios of 2:1 and 15:1. These molar ratios were chosen such that the corresponding inositol:residue ratios are above (2:1) and below (<1:1) the inositol:residue molar ratio at which inhibition of Aβ42 fibrils was observed *in vitro*.\(^{27}\)

Independent of the presence of inositol, Aβ(16-22) is a disordered peptide in solution (Figure 4.1A,B) and is able to adopt both collapsed and extended states over the time
Figure 4.1: Binding of inositol to an Aβ(16-22) monomer. End-to-end probability distribution of Aβ(16-22) successively in pure water and in presence of scyllo- and chiro-inositol at inositol:peptide molar ratios of (A) 2:1 and (B) 15:1. (C) Representative snapshots of the different binding modes of scyllo- (left) and chiro-inositol (right) to the peptide monomer. Hydrogen bonds between inositol and backbone NH (blue) and CO (red) groups are shown as solid lines. (D) Percent of bound inositol molecules in contact with nonpolar and polar groups at an inositol:peptide molar ratio of 15:1. (E) Time-averaged number of nonpolar contacts (top), and hydrogen bonds (bottom) made by inositol (at a molar ratio of 15:1) to each residue.
scales of our simulation. The conformational equilibrium of Aβ(16-22), as measured by peptide end-to-end distance distributions, was unaffected by the presence of inositol at both inositol:peptide molar ratios considered (Figure 4.1A,B). The three peaks correspond to different intramolecular hydrogen-bonding arrangements (see Figures A.1.1 and A.1.2 in Appendix A).

Inositol molecules bound weakly and reversibly to the monomer of Aβ(16-22). Representative examples of scyllo- and chiro-inositol binding are depicted in Figure 4.1C. Dissociation constants $K_{eq}(\text{scyllo}) = 127 \pm 3$ mM, $K_{eq}(\text{chiro}) = 104 \pm 1$ mM were obtained at a molar ratio of 2:1, and $K_{eq}(\text{scyllo}) = 120 \pm 2$ mM, $K_{eq}(\text{chiro}) = 93 \pm 2$ mM at a molar ratio of 15:1. Increasing the molar ratio of inositol:peptide by more than 7-fold did not decrease the $K_{eq}$ significantly, suggesting that inositol does not bind cooperatively to the peptide monomer.

Nonpolar contacts played a significant role in inositol binding, with chiro-inositol more likely than scyllo-inositol to form nonpolar contacts: as shown in Figure 4.1D, $\sim 36\%$ of bound scyllo- vs $\sim 45\%$ of chiro-inositol molecules formed nonpolar contacts with the monomer. Both stereoisomers were preferentially bound to the nonpolar group of Phe over the nonpolar groups of the other residues (Figure 4.1E).

To characterize the binding geometry of inositol to Phe in detail, we performed simulations of a Phe dipeptide in the presence of scyllo- or chiro-inositol. Specifically, scyllo- but not chiro-inositol displays a face-to-face stacking mode with the aromatic side chain of Phe (Figure 4.2C). This mode has an approximate binding free energy of $-0.5$ kcal/mol and appears on the potential of mean force (PMF) for scyllo-inositol as a free energy minimum at a distance between the center of inositol and phenyl rings, $r = 0.45$ nm, and an angle between the planes of the rings, $\theta = 12^\circ$ (Figure 4.2D, left panel). By contrast, this stacked binding mode was not observed for chiro-inositol, which lacks planar nonpolar faces because of its adjacent axial hydroxyl groups (Figure 4.2D, right panel).

scyllo-Inositol is more likely than chiro-inositol to bind via hydrogen-bonding inter-
actions: \(\sim 28\%\) of bound scyllo-inositol versus \(\sim 21\%\) of bound chiro-inositol molecules formed at least one hydrogen bond with the monomer. Inositol bound not only to the peptidic backbone of A\(\beta\)\((16-22)\) but also to the charged side chains of glutamic acid (Glu) and lysine (Lys) residues. Both stereoisomers of inositol display similar hydrogen-bonding propensities to each of the residues in the peptide. Inositol molecules bound most favorably to Glu, where their interaction was dominated by hydrogen bonding to the carboxylate group (Figure 4.1E). Both nonpolar and hydrogen bonding propensities were independent of molar ratio (Figure 4.1 and Figure A.1.3 in Appendix A). Furthermore, we found an equal fraction of monodentate and bidentate binding (Figure 4.2A) to the carboxylate group of Glu (Figure 4.2B). In contrast, less than 1% of inositol molecules bound to Lys involved multiple hydrogen bonds to the ammonium group (Figure 4.2B).

### 4.4.2 Disordered Oligomer

To probe the effect of inositol on the early aggregation stages of A\(\beta\)\((16-22)\), we performed multiple sets of independent MD simulations with four initially disperse A\(\beta\)\((16-22)\) monomers with inositol:peptide molar ratios of 2:4, 15:4, and 45:4, corresponding to inositol concentrations of 52 mM, 70 mM, and 209 mM, respectively (see Table 4.1). In each of our simulation studies, the peptides spontaneously aggregated with one another over the course of approximately 40 ns, through both hydrogen bonding and nonpolar contacts, to form a disordered oligomer (Figure 4.3A). A significant fraction of the residues in the aggregate was in the coil conformation, with only a small fraction of \(\beta\)-sheet residues occurring in some of the 180-ns simulations (Figure 4.3B). Importantly, the distribution of the overall secondary structure of the oligomer was not affected by the presence of inositol, regardless of inositol:peptide molar ratio and inositol concentration (Figure 4.3B).

We further characterized the molecular organization of the aggregate by quantifying peptide inter- and intramolecular hydrogen-bonding and nonpolar contacts as measures
Chapter 4. Binding of Inositol Stereoisomers to Aβ(16-22)  

Figure 4.2: Binding of inositol to Glu and Phe dipeptides. Data for scyllo- and chiro-inositol are shown on the left and right panels, respectively. (A) Examples of snapshots of inositol bound to the carboxylate group of Glu. (B) Comparisons of the probability of inositol hydrogen bonding to the side chains of Lys and Glu. (C) Examples of nonpolar association between Phe and inositol. (D) Potential of mean force (PMF) of inositol with the phenyl ring of Phe relating $r$, the distance between the centers of geometry of the phenyl and inositol rings, to $\theta$, the planar angle between the rings. Contours are drawn at 0.1 kcal-mol$^{-1}$ intervals. Face-to-face stacking for scyllo-inositol appears on the PMF at $r = 0.45$ nm and $\theta = 12^\circ$. 

Figure 4.3: Binding of inositol to a disordered oligomer of Aβ(16-22). (A) Example snapshots of scyllo- (left) and chiro-inositol (right) involving both nonpolar contacts and hydrogen bonding. (B) Fraction of residues in coil, β-sheet/bridge, bend, and turn conformations as classified by the DSSP algorithm. (C) Time-averaged number of nonpolar contacts (top) and hydrogen bonds (bottom) made by inositol to each residue (per peptide). Inset: Percent of inositol molecules bound to nonpolar and polar groups of the peptide oligomer at an inositol:peptide molar ratio of 45:4 (inositol concentration of 209 mM).
Table 4.2: Summary of equilibrium constants ($K_{eq}$) and number of reversible binding events ($N_{binding}$)

<table>
<thead>
<tr>
<th>System</th>
<th>Molar ratio</th>
<th>$K_{eq, scyllo}^{a}$</th>
<th>$K_{eq, chiro}^{b}$</th>
<th>$N_{binding, scyllo}^{c}$</th>
<th>$N_{binding, chiro}^{c}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monomer</td>
<td>2:1</td>
<td>127 (3)</td>
<td>104 (1)</td>
<td>150991</td>
<td>185454</td>
</tr>
<tr>
<td></td>
<td>15:1</td>
<td>120 (2)</td>
<td>90 (2)</td>
<td>186948</td>
<td>250922</td>
</tr>
<tr>
<td>Disordered oligomer</td>
<td>2:4</td>
<td>28 (4)</td>
<td>16 (2)</td>
<td>21882</td>
<td>24584</td>
</tr>
<tr>
<td></td>
<td>15:4</td>
<td>18 (2)</td>
<td>11 (1)</td>
<td>78483</td>
<td>102351</td>
</tr>
<tr>
<td>$\beta$-oligomer</td>
<td>4:16</td>
<td>15 (2)</td>
<td>11 (2)</td>
<td>20381</td>
<td>24616</td>
</tr>
<tr>
<td></td>
<td>64:16</td>
<td>0.5 (0.3)</td>
<td>0.18 (0.11)</td>
<td>50135</td>
<td>56842</td>
</tr>
</tbody>
</table>

$^{a}$ Inositol:peptide molar ratios
$^{b}$ $K_{eq}$ is in units of mM. The standard error is shown within parentheses.
$^{c}$ $N_{binding}$ is the total number of reversible inositol binding events.

of the extent of aggregation. Hydrophobic packing was not affected by the presence of inositol: the equilibrium number of inter-peptide hydrophobic contacts formed per peptide remained approximately 15 regardless of the molar ratio (Figure A.2.1 in Appendix A). The average number of intermolecular peptide-peptide hydrogen bonds per chain was approximately the same as the number of intramolecular hydrogen bonds (1.5 vs 1) (Figure A.2.1 in Appendix A). Overall, the presence of inositol had no significant effect on the aggregation kinetics or on the morphology of $\alpha\beta(16-22)$ oligomers as measured by intermolecular and intramolecular contacts (Figures A.2.2 and A.2.3 in Appendix A).

The equilibrium constant ($K_{eq}$) of inositol with the disordered oligomer at molar ratios 2:4 and 15:4 ranged from 10 to 30 mM (see Table 4.2). Similar proportions of bound inositol to nonpolar and polar groups were found at both lower and higher molar ratios (Figure A.2.1 in Appendix A). Consistent with our results for the monomer, *chiro*-inositol was more likely than *scyllo*-inositol to bind disordered oligomers exclusively via nonpolar contacts: $\sim$26% and $\sim$36% for *scyllo*- and *chiro*-inositol, respectively (Figure 4.3C inset and Table A.2.1). Inversely, *scyllo*-inositol was more likely to bind by hydrogen bonding only ($\sim$23% vs. $\sim$17%). Although the number of hydrogen bonds formed along the peptide sequence were independent of inositol concentration (Figure 4.3, A.2.4), the number of nonpolar contacts per peptide approximately doubled upon increasing inositol concentration from 70 to 209 mM.
4.4.3 \(\beta\)-oligomer

Finally, we examine the binding of inositol to an ordered protofibrillar-like aggregate henceforth referred to as the \(\beta\)-oligomer. In the absence of inositol, rectangularly-stacked sheets (Figures A.3.1A-B) spontaneously evolved into a twisted \(\beta\)-sheet structure with significant inter-strand twisting along the long-axis of the fibril and an inter-sheet twist (Figures 4.4A-B). The resulting structure has an average inter-strand twist angle of approximately 25° for the top sheet and 15° for the bottom sheet. Furthermore, the \(\beta\)-oligomer is comprised of two faces and four edges (Figure A.3.1), each of which contains a shallow hydrophobic groove surrounded by polar or charged groups. In particular, the grooves on the faces are formed by solvent-exposed Phe, Val, and Ala residues and are surrounded on either side by charged side chains of Lys and Glu (Figure A.3.1C).

The spatial probability densities of bound inositol depicted in Figures 4.4C-D show that inositol predominantly binds at the faces of the \(\beta\)-oligomer. Both stereoisomers have similar affinities with \(K_{eq} = 15 \pm 2\) mM and \(11 \pm 2\) mM for scyllo- and chiro- inositol, respectively, at a concentration of 37 mM (inositol:peptide molar ratio of 4:16), and \(K_{eq} = 0.5 \pm 0.3\) mM and \(0.18 \pm 0.11\) mM at a concentration of 62 mM (molar ratio of 64:16) (Table 4.2).

Consistent with the binding densities depicted in Figure 4.4, inositol molecules display the highest binding propensity to the nonpolar groups of Phe and Lys and the charged groups of Lys and Glu, all of which are located on the faces of the \(\beta\)-oligomer (Figure A.3.1C). Inositol molecules did not penetrate the \(\beta\)-sheet core of the oligomer: the fraction of hydrogen bonds to each residue depicted in Figure 4.5 (bottom panel) show that, relative to side chains, little or no hydrogen bonds were made with the backbone of residues Leu, Val, Phe, and Ala. Although inositol molecules sometimes intercalated between \(\beta\)-strands, these rare events did not lead to the disaggregation of the preformed \(\beta\)-oligomer in any of our simulations.

Independently of inositol concentration, a higher fraction of scyllo-inositol than chiro-
Figure 4.4: Inositol binding to a β-oligomer of Aβ16-22. Schematic depiction of β-oligomer twisting: (A) the initial rectangular dual-stacked β-sheet, evolved into (B) a twisted morphology. Spatial probability density maps of (C) scyllo-inositol and (D) chiro-inositol are shown in yellow and orange, respectively. Concentration of inositol is 208 mM and surfaces shown correspond to 7% inositol occupancy. (E) An example of cooperatively-bound scyllo-inositol molecules (yellow) at the surface of the β-oligomer (grey). Size distribution of bound and unbound clusters of scyllo-inositol with the β-oligomer at inositol concentrations of (F) 62 mM and (G) 208 mM.

Inositol formed hydrogen bonds with the β-oligomer (inset of Figure 4.5) and Table A.3.1): ~23% versus ~15%, respectively. Concomitantly, the fraction of chiro-inositol molecules forming nonpolar contacts (~29%) was higher than that of scyllo-inositol (~19%) (inset of Figure 4.5) and Table A.3.1).

At the higher inositol:peptide molar ratio of 64:16, bound inositol molecules were significantly more likely to be clustered than free inositol: 20% versus 8% (Figure 4.4F-G) at 62 mM. Moreover, the size of bound clusters increased with concentration. Inositol molecules within a cluster were usually hydrogen bonded to each other via their free hydroxyl groups, while simultaneously forming hydrogen bonds and/or nonpolar contacts with the peptide. Such a binding mode is depicted in Figure 4.4E, where a hydrogen-bonded chain of four scyllo-inositol molecules occupies a shallow groove on the β-oligomer surface. There was no difference in distribution of cluster size between chiro- and scyllo-inositol.

Furthermore, the binding propensity of scyllo- and chiro-inositol for hydrophobic groups increased with inositol concentration (Figure 4.5). At lower concentrations (~30 - 60 mM), both nonpolar and hydrogen bonding propensities increased with molar ratio, suggesting that single molecules and dimers of inositol have similar binding propensities, and binding to the protofibril involves forming both nonpolar contacts and hydrogen bonds. However, at a concentration of 208 mM, the nonpolar binding propensity of inositol increased whereas the hydrogen bonding propensity remained the same. As a result, inositol molecules in large clusters (of size three or more) form, on average, more
Figure 4.5: Binding propensity of inositol to nonpolar and polar groups of the $\beta$-oligomer. Average number of nonpolar contacts (top) and hydrogen bonds (bottom), per peptide, made by inositol to each residue of the $\beta$-oligomer. Inset: Percent of scyllo- and chiro-inositol molecules bound to nonpolar and polar groups of the $\beta$-oligomer. Inositol is present at a concentration of 62 mM in part A and 208 mM in part B.
nonpolar contacts with the $\beta$-oligomer than their singly-bound counterparts.

## 4.5 Discussion

In the above analysis, we have systematically characterized the binding of *scyllo*-inositol and its inactive stereoisomer, *chiro*-inositol, with monomer and aggregates of A$\beta$(16-22). Below, we consider the implications of our findings for the activity of inositol in the A$\beta$42 amyloid aggregation pathway.

### 4.5.1 Comparison of inositol binding to monomers and aggregates of A$\beta$(16-22)

Consistent with our results on the binding equilibrium of inositol with model amyloidogenic peptides, both *scyllo*- and *chiro*-inositol bound weakly, and with similar binding constants, to the monomeric and aggregated states of A$\beta$(16-22) considered. However, the equilibrium constants ($K_{eq}$) computed in this study are about an order of magnitude smaller than those obtained in our previous study, namely, in the range of 0.2 - 120 mM for A$\beta$(16-22) versus 40 - 1000 mM for a model peptide of similar length, (GA)$_4$.

Because all observed binding sites and modes are accounted for in the calculation of $K_{eq}$ in our study, this quantity should be interpreted as an estimate of the binding avidity rather than the binding affinity of inositol. This decrease in $K_{eq}$, and hence, an increase in binding avidity, is due to the presence of sequence-specific binding sites and modes in A$\beta$(16 − 22).

Both *scyllo*- and *chiro*-inositol bound most weakly to monomers, with $K_{eq} = 120 \pm 2$ mM and $90 \pm 2$ mM, respectively, at the highest molar ratio (Table 4.2). Because inositol binding to the peptide monomer is not cooperative, a predicted $K_{eq}$ of monomeric A$\beta$42 can be obtained by linearly scaling the $K_{eq}$ of inositol for monomeric A$\beta$(16-22) with the ratio of peptide lengths of A$\beta$(16-22) to A$\beta$(1-42). On the basis of the value of $K_{eq}$ of
inositol at the highest molar ratio, this value would be 120 mM/6 = 20 mM, which is an order of magnitude higher than the concentration (1 mM) at which inhibition was observed in vitro. Moreover, our results indicate that the conformational equilibrium of monomeric Aβ(16-22) is not displaced in the presence of inositol (Figure 4.1A). Taken together, these results suggest that inositol is unlikely to act as a drug by binding to and displacing the conformational equilibrium of monomers of Aβ42.

Likewise, inositol bound only weakly to small disordered oligomers, with $K_{eq} \sim 10$ - 30 mM at a concentration of 70 mM for both scyllo- and chiro-inositol (Table 4.2). Independently of the presence of inositol, Aβ(16-22) peptides formed amorphous aggregates with only a small amount of secondary structure. These aggregates predominantly involved intermolecular nonpolar contacts (Figure A.2.2), indicating that hydrophobic association is the primary driving force for the self-assembly of Aβ(16-22) peptides in solution. Inositol molecules were found to bind both monomers and small oligomers of Aβ(16-22) predominantly via nonpolar interactions (Figures 4.1 and 4.3), suggesting that they may disrupt the hydrophobic association of nonpolar groups. However, due to weak binding, we speculate that inositol is unlikely to prevent early oligomer formation in the Aβ42 fibrillation pathway by binding to Aβ(16-22).

In contrast, inositol displays a much higher binding avidity for β-oligomers, with $K_{eq} = 0.5 \pm 0.3$ mM and $0.18 \pm 0.11$ mM for scyllo- and chiro-inositol (at a concentration of 62 mM), respectively. Notably, these $K_{eq}$ values are in quantitative agreement with experimental concentrations (0.5 - 1 mM) sufficient for the inhibition of Aβ42 fibrillation in vitro, suggesting that β-oligomers may be an in vitro binding partner of inositol.

A key finding of this study is that the stereospecificity of binding by inositol stereoisomers is not due to different $K_{eq}$’s, but rather to different binding modes with nonpolar groups of side chains with specific geometries. In particular, due to the presence of planar hydrophobic faces, scyllo-inositol, unlike chiro-inositol, can bind Phe side chains (or other side chains with planar geometries) in a planar face-to-face stacking mode (Fig-
Chapter 4. Binding of Inositol Stereoisomers to Aβ(16-22) 128

ure 4.2D). In all of our systems considered, this stacking mode accounts for ~9% of Phe bound by scyllo-inositol, and increases to about 10 - 12% at higher concentrations of scyllo-inositol.

Furthermore, the binding probability densities of scyllo-inositol are more localized to the grooves of β-oligomers, whereas those of chiro- are spread more widely (Figure 4.4). This difference in spatial distribution is consistent with the difference in binding avidity, which was higher for chiro-inositol than for scyllo-inositol in all of the systems considered. Moreover, scyllo-inositol displays a higher hydrogen-bonding propensity than chiro-inositol, which is likely to contribute to its higher binding specificity. Taken together, our results suggest that scyllo-inositol binds with more specificity than chiro-inositol to β-sheet aggregates of Aβ(16-22).

Furthermore, binding modes of inositol involving nonpolar groups are modulated by aggregate morphology: the change in morphology from monomers to oligomers resulted in a significant decrease in the population of stereoisomers bound exclusively by nonpolar contacts, concomitant with an increase in the population of stereoisomers forming both nonpolar contacts and hydrogen bonds with the peptides. This difference is more pronounced for the β-oligomer (Figure 4.5). Our results indicate that both hydrogen bonding and nonpolar interactions are important for the binding of inositol to Aβ(16-22), and that the balance of these interactions is modulated by both aggregate morphology and inhibitor stereochemistry.

4.5.2 Binding cooperativity with β-oligomers

The binding avidity of both scyllo- and chiro-inositol decreased significantly (corresponding to an increase in binding) with an increase of inositol:peptide molar ratio: $K_{eq} = 10$ - 15 mM and 0.1 - 0.5 mM at 4:16 and 64:16 molar ratios, respectively. This finding is consistent with the existence of cooperative binding involving clusters of multiple inositol molecules at higher molar ratios (Figure 4.4E). Here we refer to cooperative binding as the
propensity of ligand binding at one or more sites to increase the affinity for the binding of additional ligands at other sites. Furthermore, the size of these clusters is modulated by inositol concentration (Figure 4.4F-G). Taken together, our results suggest that inositol binding is cooperative at sufficiently high molar ratios and concentrations. In support of our findings, a recent combined simulation and biophysical study on the polyphenolic inhibitor EGCG indicated that binding modes of other small-molecule inhibitors may be modulated by ligand:peptide molar ratio: an increase of EGCG:Aβ42 molar ratio shifted the predominant binding interaction of EGCG from hydrogen-bonding to hydrophobic interactions.40

Our results suggest that the binding cooperativity of inositol results from a combination of favorable intermolecular interactions with peptide and inositol groups: by exposing multiple hydrogen bonding groups in proximity to nonpolar groups of the protein, bound inositol molecules promote the binding of additional inositol molecules at adjacent binding sites. By contrast, a linear dependence of binding avidity upon inositol concentration was observed for the monomers and the disordered oligomers (Table 4.2), presumably because these morphologies cannot accommodate this type of multivalent interaction.

The increase in the binding avidity of inositol for the β-oligomer of Aβ(16-22) relative to monomeric and disordered oligomeric forms may be explained by structural features present in the former, but not in the latter species. First, the β-oligomer has a much larger effective surface area, which can accommodate multiple bound inositol molecules (Figure 4.4E). Second, as a direct consequence of its morphology, the β-oligomer presents grooves on its surface that collocate the residues (i.e. Phe and Glu) capable of high-affinity interactions with inositol.

Taken together, these findings suggest that the clustering of inositol molecules may be important for increasing the local concentration of inositol in the vicinity of the peptide, and could play a role in overcoming the weak binding affinity of individual inosi-
tol molecules in order to achieve drug-like activity. Similarly, recent simulation studies of $\mathrm{A}\beta 40$ fibrillar fragments and non-steroidal anti-inflammatory drugs, ibuprofen and naproxen, suggested that their inhibitory activities may be related to their ability to bind cooperatively and to form clusters on the surface of $\mathrm{A}\beta 40$ fibrillar aggregates.$^{41,42}$

### 4.5.3 Proposed mechanism of $\mathrm{A}\beta$ amyloid inhibition

Mature amyloid fibrils are thought to form either by $\beta$-strand addition along the long axis of the fiber (elongation) or by lateral face-to-face association with other protofibrils.$^{61}$ It follows that small molecules that can disrupt either of these interactions may inhibit fibrillation. In addition, multiple experimental studies have shown that $\mathrm{A}\beta(16-22)$ is part of the $\beta$-sheet core of fibrils of $\mathrm{A}\beta42^{62-66}$ and is central to the fibrillation of $\mathrm{A}\beta40/42^{67,68}$ Furthermore, the structures of several fibril polymorphs suggest that residues 16-22 mediate the stacking of constituent protofilaments in mature fibrils.$^{5,66,69,70}$ Consistent with these observations, the fibrillar structure of full-length $\mathrm{A}\beta42$ from a solid-state NMR study shows that the protofibril has two different $\beta$-sheet faces, one of which is formed by the $\mathrm{A}\beta(17-22)$ peptide segment.$^{69}$

Our results suggest that $\mathrm{A}\beta(16-22)$, the peptide sequence forming the fibrillar core of full-length $\mathrm{A}\beta42$, is a likely binding site for inositol. Furthermore, our results indicate that scyllo-inositol binds to this region more specifically than chiro-inositol. We hypothesize that scyllo-inositol, by binding to and coating the $\beta$-sheet surfaces of protofibrils involving $\mathrm{A}\beta(16-22)$, disrupts the lateral stacking of these oligomers, which ultimately leads to the inhibition of fibril formation. Consistent with this hypothesis, the amyloid dye Congo red has been suggested by previous studies to disrupt amyloid formation in a similar manner - i.e. by binding to grooves on the surface of extended $\beta$-sheets.$^{71,72}$

Furthermore, based on our results, we hypothesize that planar nonpolar faces with multiple hydroxyl groups in equatorial positions around the ring confer binding specificity to small molecules for the amyloidogenic core of $\mathrm{A}\beta$, and thus are key features for their
activity. Consistent with this hypothesis, *in vitro* studies of small-molecule derivatives of scyllo-inositol showed that the substitution of a single hydroxyl by a ketone group resulted in loss of activity (i.e., fibrils were formed).\textsuperscript{25,27,73} Moreover, polyphenols, many of which are strong *in vitro* inhibitors of amyloid formation, all possess planar nonpolar faces with hydroxyl groups arranged equatorially. A similar hypothesis was recently put forth based on structure-activity relationships of polyphenols as a possible explanation for their effectiveness in inhibiting amyloid formation.\textsuperscript{74}

Many differences exist between the β-oligomer of Aβ(16-22) and protofibrils of the full-length Aβ peptides. Our results indicate that inositol binding depends on both the fibrillar morphology and the surface physico-chemical properties of the peptide aggregate. Thus, alternative binding modes and binding sites of inositol may exist on aggregate forms of the full-length Aβ42 peptide, which cannot be deduced from the results of this study. As part of our future directions, we will perform comparative simulation studies of scyllo- and chiro-inositol binding to aggregates of full-length Aβ.

### 4.5.4 Similarity to Carbohydrate Binding

A striking result of our study is the characteristic sugar-like binding affinities and binding modes of inositol. Similar to inositol, monosaccharides exhibit millimolar binding affinities for lectins, a class of sugar-binding proteins.\textsuperscript{76,77} Furthermore, sugar binding usually involves a combination of hydrogen bonds between hydroxyl groups and charged side chains (Asp or Glu) and nonpolar stacking of aromatic moieties, which are important for the recognition and selectivity of sugar enantiomers by lectins.\textsuperscript{78} Consistent with these observations, our results indicate that inositol displays higher binding propensities to Phe and Glu compared to Lys, Leu, Val, and Ala. Moreover, from our simulations, the binding free energy of stacking to the phenyl ring of Phe is approximately -0.5 kcal/mol, in agreement with that of glucose binding to the indole group of tryptophan obtained from recent MD simulation\textsuperscript{76} and NMR studies.\textsuperscript{79}
Finally, the shallow amphiphilic grooves found at the surface of β-oligomers (Figure 4.4E) are analogous to binding sites located at the surface of carbohydrate-binding domains. Akin to the multivalency in binding often exhibited by carbohydrates, by forming multiple weak affinity interactions, inositol molecules cluster in these shallow grooves, which results in a higher overall binding avidity for the β-oligomer. These cooperative binding modes suggest that linearly-linked inositol stereoisomers (e.g. dimers, trimers, or tetramers using scyllo-inositol subunits) may be one possibility for designing putative inhibitors with higher affinities. An improvement in drug affinity is advantageous because patients may be administered smaller dosages so that the risk of side effects is lowered while drug efficacy is retained. Taken together, the above results suggest that inositol binds in carbohydrate-like binding sites on β-sheet surfaces involving Aβ(16-22), and that carbohydrates may be used as a template for the design of AD inhibitors.

4.6 Conclusions

In this study, we have examined the binding of a small molecule inhibitor scyllo-inositol, and its inactive stereoisomer, chiro-inositol, successively to monomers, disordered oligomers, and β-sheet aggregates of Aβ(16-22), whose sequence is thought to be the core aggregation region in the Aβ42 peptide. Notably, the $K_{eq}$ of inositol (∼0.2 - 0.5 mM) for the β-oligomer is commensurate with the concentration at which inhibition of amyloid formation by Aβ42 is observed in vitro. Although both scyllo- and chiro-inositol exhibit similar binding affinities with all peptide states considered, we have uncovered a stereo-specific face-to-face stacking stacking mode of scyllo-inositol with the Phe side chains and a higher propensity for hydrogen bonding, which together suggests a molecular basis for measured differences in activity. Cooperative binding modes of inositol at grooves on the surface of the β-oligomer of Aβ(16-22) suggest a possible mechanism of fibril inhibition whereby inositol prevents the lateral association or stacking of protofibrillar...
\(\beta\)-sheet oligomers. Furthermore, our results suggest that the fibril core of A\(\beta\) amyloid aggregates contains carbohydrate-like binding sites. As such, carbohydrate-based small-molecule derivatives may be a promising avenue to explore for the rational design of novel therapeutics for AD.
Bibliography


Chapter 5

Molecular Mechanism of Aβ42 fibril inhibition by inositol

Contributions: Grace Li conducted the research and wrote the section. Régis Pomès provided editorial input and guidance.
5.1 Introduction

A\(\beta\)42 is the pathological hallmark of Alzheimer’s disease (AD) and forms the largest proteinaceous component of plaques in the brain of AD patients. A\(\beta\) peptides are produced from the cleavage of amyloid precursor protein (APP) in isoforms with lengths of 33 to 42 residues. Although A\(\beta\)42 and A\(\beta\)40 peptides differ in length only by two residues, A\(\beta\)42 is found to display significantly higher aggregation propensity\(^1\)\(^–\)\(^3\) and cellular toxicity than A\(\beta\)40 peptides.\(^4\)\(^,\)\(^5\) Multiple studies have probed the aggregation properties of both A\(\beta\)40 and A\(\beta\)42. Fibril models of the A\(\beta\)40 peptide derived from solid-state NMR (SSNMR) indicate that protofilaments of A\(\beta\)40 contain 2 to 3 layers.\(^6\)\(^,\)\(^7\) By contrast, less is known about the fibril structure of A\(\beta\)42. A SSNMR-based model of the core of the fibril of A\(\beta\)42 was recently proposed by Lührs et al.\(^8\) In that model, the cross-\(\beta\) core of the fibril consisted of residues 17 to 42, with the N-terminus of the peptide residues 1 to 16 unstructured in the fibril.\(^9\)

In recent years, small-molecule inhibitors of A\(\beta\) amyloid formation have emerged as promising candidates for the treatment of Alzheimer’s disease. One such molecule is scyllo-Inositol, an inhibitor of A\(\beta\)42 fibrillation.\(^10\)\(^–\)\(^13\) Inositol is a class of cyclohexylpolyols, of which eight out of nine stereoisomers are commonly found in nature. scyllo-Inositol, in which all hydroxyl groups are equatorial, is the only isomer with two planar hydrophobic faces. By contrast, its diastereoisomer, chiro-inositol, with two adjacent axial hydroxyl groups, has two nonplanar faces with mixed polarity.

In vitro, inositol displays stereochemistry-dependent inhibition of A\(\beta\)42 fibrils: scyllo-inositol was shown to inhibit A\(\beta\)42 fibrillation at concentrations of 1 - 5 mM,\(^11\) whereas chiro-inositol is inactive below molar concentrations.\(^11\) Moreover, upon incubation of monomeric A\(\beta\)42 with scyllo-inositol, circular dichroism spectroscopy indicated the formation of \(\beta\)-sheet structure at an inositol:peptide molar ratio of 25:1.\(^11\)\(^,\)\(^14\)

Importantly, scyllo-inositol was demonstrated to prevent and reverse AD-like symptoms in a transgenic mouse model of AD.\(^10\) Phase I and II of clinical trials for scyllo-
inositol (ELN0005) in North America has been completed.\textsuperscript{13,15} These clinical trials have demonstrated that \textit{scyllo}-inositol possesses positive CNS bioavailability and favorable \textit{in vivo} toxicity profile. Although clinical trials have successfully demonstrated the safety and tolerance profile of \textit{scyllo}-inositol, further chemical modification is likely required to improve its efficacy. Understanding the structural determinants of the effect of inositol stereoisomers on A\textsubscript{β}42 will aid in the rational design of more efficacious derivatives of inositol. However, the molecular basis of the stereospecific effect of \textit{scyllo-} and \textit{chiro-}inositol on A\textsubscript{β}42 amyloid inhibition is not understood.

In two previous studies (Chapters 3 and 4), we have successively examined the binding mechanism of \textit{scyllo}-inositol and \textit{chiro}-inositol with peptide and aggregate states of model amyloid-forming peptides\textsuperscript{16} and A\textsubscript{β}(16-22).\textsuperscript{17} Weak and stereochemistry-independent binding of inositol with the peptidic backbone were found, with binding constants in the range of 0.1 - 1 M, indicating that inositol is unlikely to inhibit amyloid formation by binding the peptidic backbone alone (see Chapter 2).\textsuperscript{16} In that initial study, inositol was found to bind to the surface of \textit{β}-sheet oligomers with the highest binding affinity, but only more weakly monomeric and disordered morphologies the weakest, suggesting that \textit{β}-sheet oligomers are the most likely binding partner of inositol. Upon further investigation with monomeric peptides and aggregates of A\textsubscript{β}(16-22), inositol was found to adopt cooperative, high-avidity binding modes with \textit{β}-oligomers characterized by binding constants commensurate with \textit{in vitro} inhibitory concentrations.\textsuperscript{17} Taken together, the results from our previous studies indicate that inositol may disrupt amyloid fibrillation by binding to \textit{β}-sheet oligomers of A\textsubscript{β}.

In this study, we examine successively the binding of inositol stereoisomers, \textit{scyllo-} and \textit{chiro-}inositol, and osmolytes glucose and glycerol with the SSNMR model of the protofibril of full-length A\textsubscript{β}42. Glucose and glycerol are organic osmolytes that can stabilize the folded structures of globular proteins. Osmolytes are thought to stabilize the folded state via the preferential exclusion mechanism.\textsuperscript{18} In recent years, the osmolytes
TMAO,\textsuperscript{19} betaine,\textsuperscript{20} glucose,\textsuperscript{21} trehalose\textsuperscript{22} and glycerol\textsuperscript{22} have been found to modulate amyloid formation and peptide aggregation.\textsuperscript{21,23–25} Here, we compare the binding of inositol stereoisomers, \textit{scy}llo- and \textit{chiro}-inositol with glucose and glycerol, osmolytes which do not inhibit the amyloid formation of A\textsubscript{\beta}42 at sub-osmolar concentration (in the milimolar range or less). Results of this study elucidate the mechanism of action of inositol in the inhibition of A\textsubscript{\beta} amyloid formation, and shed light on the role of stereochemistry in amyloid inhibition by small molecules in general.

5.2 Material and Methods

The pentameric solid-state NMR model of A\textsubscript{\beta}(17-42) from Lührs \textit{et al.} (PDB code: 2BEG) was taken as the starting structure in our simulations.\textsuperscript{8} In the PDB structure, residues 1 to 16 were truncated in the model because they were found to be disordered.\textsuperscript{8} The peptides were capped by acetyl groups at the N-terminal end. The acetyl groups were modelled onto the fibril structure using the PyMol software.\textsuperscript{26} Titratable amino acids were assigned their charged states at the physiological pH. Ten sodium ions were added to neutralize the remaining charges in the system. To mimic experimental conditions,\textsuperscript{8} 0.15 M (46 ions) of NaCl salt was added. Protein and solvent were represented by the OPLS-AA/L force field\textsuperscript{27} and the TIP3P water model,\textsuperscript{28} respectively. The extended OPLS-AA force field for carbohydrates\textsuperscript{29} was used to model inositol and glucose molecules.

MD simulations were performed in the NpT ensemble using version 4.0.x of the GROMACS simulation package.\textsuperscript{30} The leapfrog Verlet integration algorithm was used with an integration timestep of 2 femtoseconds. Long-range electrostatic interactions were calculated using Particle Mesh Ewald (PME) summation with a Fourier grid spacing of 0.15 nm and a real-space cutoff of 1.3 nm.\textsuperscript{31} The short-range nonbonded van der Waals interactions were switched to zero from 1.1 nm to 1.2 nm. The temperature was controlled at 300 K using the Nose-Hoover thermostat.\textsuperscript{32} Pressure was controlled by the
Chapter 5. Aβ42

Parrinello-Rahman barostat at 1 atm with a coupling constant of 4.0 ps. The SHAKE algorithm was used to constrain covalent bonds involving hydrogen atoms. Center of mass (COM) rotation and translation were removed at every step.

In all simulations, a cubic box geometry was used with periodic boundary conditions. Prior to data collection, 500 steps of energy minimization using the conjugate gradient algorithm was first performed. Simulation systems were equilibrated using a two-step procedure. First, a 200 ps equilibration was conducted in the NVT ensemble to further relax the initial configuration of the system. Next, equilibration in the NpT ensemble (with isotropic pressure coupling) was conducted for 2 ns to relax the solvent around the protein.

In total, a set of ten 0.250 µs simulations of the protofibril of Aβ42 were performed for each of 5 systems, successively in pure water and in the presence of scyllo-inositol, chiro-inositol, glycerol and glucose. The ligands were present at a ligand:peptide molar ratio of either 15:5 or 64:5, yielding a total sampling time of 12.5 µs.

5.2.1 Analysis Protocol

The GROMACS analysis utilites g_rmsd and g_rmsf were used to calculate the root mean square deviation and root mean square fluctuation of the fibril backbone, respectively. The spatial probability density of bound ligands was computed using the VolMap analysis tool implemented in the Visual Molecular Dynamics (VMD) software package.

The number of hydrogen bonds between each pair of peptides in the protofibril was computed using g_hbond. The following geometry criteria used to define a hydrogen bond: (1) the distance between acceptor (A) and donor (D) heavy atoms is less than 0.35 nm, and (2) the angle formed by H-D-A is less than 30°. Nonpolar contacts between inositol and fibril were defined by a carbon-to-carbon distance cutoff of 0.45 nm. The DSSP algorithm was used to analyze the secondary structure of the protofibril.
Chapter 5. Aβ42

5.3 Results and Discussion

We performed simulations of the protofibril of Aβ42 successively in the presence of scyllo-inositol, chiro-inositol, glucose, and glycerol (Figure 5.1) at molar ratios of 15:5 and 64:5. In addition, simulations of the protofibril in pure water were also carried out as a control. Because the results were independent of molar ratio, in the sections below, we comparatively examine the binding of the ligands (Figure 5.1) at the higher molar ratio.

![Molecular structures](image)

Figure 5.1: Molecular structures of (A) glycerol, (B) glucose, (C) chiro-inositol and (D) scyllo-inositol

5.3.1 Protofibrillar structure

We first examine the global structural properties of the fibril fragment by examining the time evolution of the RMSD and RMSF (root mean squared fluctuations) of the protofibril. Both in the absence and presence of the ligands (scyllo-, chiro-inositol, glucose and glycerol) the RMSD of the protofibril over the course of the simulation reached a plateau at 0.5 nm (Figure 5.2A). The RMSF values depicted in Figure 5.2B, shows the average deviation of the position of each residue from its average position. The average RMSF of a residue in the protofibril vary between 0.1 nm and 0.5 nm. The time evolution of both RMSD and RMSF of the protofibril was independent of the presence and identity of ligands. Although chains at the edges of the fibril were observed to detach partially from the core of the protofibril in some of our trajectories, the β-sheet core of the peptide aggregate stayed intact throughout the simulation despite the absence of
structural restraints. The secondary structure of the protofibril was predominantly β-sheet. Taken together, these results indicate that the fibril morphology is not significantly affected by the presence of the ligands on the time scale of our simulations.

Figure 5.2: Fibril structure dynamics in pure water (A and C), and in the presence of scyllo-inositol (B and D).

5.3.2 Ligand binding

The protofibril presents two chemically different β-sheet faces. The β1 face contains residues in the central hydrophobic core of Aβ, LVFFA, and possesses a negatively charged region formed by a row of solvent-exposed Glu22 (residue 22 in the full-length peptide). In contrast, the β2 face presents a predominantly hydrophobic surface containing the residues at the C-terminal end of peptide, IIGLMVGVVIA (Figure 5.3E). Scyllo-inositol, chiro-inositol, and glucose molecules all bound to the surface of the protofibril (Figure 5.4). Binding involved one or more of four main binding sites on
the protofibril: (1) the $\beta_1$ face, (2) the $\beta_2$ face, (3) the edges, and (4) a channel-like cavity formed between the $\beta$-sheets (Figure 5.3). The binding properties of each of the three solutes in each of these sites and their implication for the mechanism of amyloid inhibition are examined in detail below.

![Figure 5.3: The model of the A$\beta$42 protofibril with its major ligand binding sites labeled.](image)

Scyllo-inositol predominantly bound to the $\beta_1$ face rather than the $\beta_2$ face (Figure 5.4A). By contrast, both chiro-inositol and glucose show significantly less binding density on the $\beta_1$ face, and instead were bound predominantly to the $\beta_2$ face. Contact maps of ligand-protofibril binding indicate that binding on the $\beta_1$ face is localized to residues Phe20 and Glu22 (Figure 5.5). As depicted in Figure 5.6, all three molecules adopt binding modes to the $\beta_1$ face via a combination of hydrogen bonding to Glu22 and nonpolar contacts with Phe20. However, scyllo-inositol is the only ligand which can adopt a face-to-face binding mode with the side chain of Phe20 while simultaneously hydrogen bonding with Glu22. This binding mode has been characterized in our previous study in Chapter 4.\(^{17}\)

Contrary to both inositol and glucose, glycerol did not display significant binding density on either face (Figure 5.4C). Because glycerol did not bind to the solvent-exposed
surface of the protofibril, unless otherwise specified, the term ligands will be used to
collectively refer to scyllo-, chiro-inositol and glucose for clarity in the sections below.

Another binding site with significant binding density is found in the grooves located
at the edges which runs parallel to the \( \beta \)-strands in the protofibril. Again, scyllo-inositol
displayed a narrower and more localized density than those of chiro-inositol (Figure 5.4,
5.5), which bound non-specifically to residues along the entire edge. By contrast to the
inositol stereoisomers, glucose did not bind to the edges significantly. Based on these
results, we speculate that binding at the edges may slow down fibril elongation, but is
unlikely to lead to arrest of the growth of a protofibril into amyloid fibrils.

Furthermore, the protofibril exhibits an intersheet channel-like cavity formed by
residues 8 - 19 of the peptide (Asp8, Val9, Gly10, Ser11, Asn12, Lys13, Gly14, Ala15,
Ile16, Ile17, Gly18, and Leu19), which together adopt a turn conformation (Figure 5.3).
Of these residues, only the side chains of Asp23, Lys28, Ile32, and Leu34 line the inside
of the cavity. Chiro-inositol, glucose, and glycerol, unlike scyllo-inositol displayed sig-
nificant binding densities here (Figure 5.4). Both chiro-inositol and glucose bound on
the inside of the channel by forming linear-hydrogen-bonded chains (Figure 5.6) that si-
multaneously formed hydrogen bonds and nonpolar contacts with side chains of residues
inside the channel. Although chiro-inositol and glucose penetrated this cavity, scyllo-
inositol instead was mostly trapped near its entrances (Figures 5.4, 5.6). In support of
our results, several previous simulation studies utilizing the same model of the A\( \beta \)42
protofibril have observed similar binding modes for amyloid dye molecules (ThT and
PiB).\(^{37}\) and a polyphenol morin.\(^{38}\) Differential binding in this cavity exhibited by the
inositol stereoisomers and glucose indicates how the stereochemistry of small molecules
can modulate their binding to amyloid morphologies. We hypothesize that these binding
modes are unlikely to lead to amyloid inhibition, as these binding sites are not located at
an area where fibril growth may occur. Furthermore, we speculate that binding in this
region may be an unproductive binding mode, which may serve to decrease the inhibitory
Figure 5.4: Comparison of the spatial binding probability densities and binding modes for (A) scyllo-inositol, (B) chiro-inositol, (C) glycerol and (D) glucose.
Figure 5.5: Contact maps for ligand-protofibril binding involving hydrogen bonding (top) and nonpolar contact maps (bottom) for (A) scyllo-inositol, (B) glucose, (C) chiro-inositol, and (D) glycerol. Each row and column of the matrix represents the average number of contacts for a peptide and a residue in the protofibril, respectively.
activity of small molecules. Accordingly, the non-inhibitors, glucose and chiro-inositol, were more prone to participating in these binding modes.

Figure 5.6: Binding modes of scyllo-inositol, chiro-inositol and glucose to the β1 face (top row) and channel-like groove (bottom row). Residues that are within 4 Å of a bound solute are represented in stick form. Protein is represented as a cartoon in gray. Residues Phe, Glu, Asp, Lys and Ile are shown in purple, pink, red, green, and light green, respectively.

In addition to partitioning to different binding sites, both ligand binding modes and propensities also were modulated by ligand stereochemistry. In particular, scyllo-inositol has the highest percentage of molecules bound exclusively via hydrogen bonding (≈26%). Furthermore, only scyllo-inositol molecules formed nonpolar contacts (24%) and hydrogen bonds (26%) with the fibril in approximately equal proportions (Figure 5.4). By contrast, a higher fraction of chiro-inositol, glucose, and glycerol molecules binding involved nonpolar contacts (Figure 5.4). Furthermore, the fraction of molecules bound by exclusively nonpolar contacts increased concomitantly with a decrease in the fraction bound by exclusively hydrogen bonds in the following order: scyllo-inositol (26%, 22%), chiro-inositol (33%, 13%), glucose (37%, 9%), and glycerol (50%, 11%) (Figure 5.4).
Scyllo-inositol with its planar hydrophobic faces and all-equatorial hydroxyl groups, has the stereochemistry most prone to form hydrogen bonds with the protofibril. The loss of planar nonpolar faces in chiro-inositol reduced the hydrogen-bonded-only population by 9%. Glucose, with one less hydroxyl group than the inositol stereoisomers, saw a 13% reduction in the population bound by exclusively via hydrogen bonds. Glycerol, which lacks a carbon ring and has the smallest number of hydroxyl groups, was most likely to interact via non-specific nonpolar contacts. Corroborating the results of our previous study, the results of this study suggest that the ability to form both hydrogen bonds and nonpolar contacts equally favorably is an important property of small molecules (to target the CHC) and therefore, ultimately, for the inhibition of amyloid formation.

Figure 5.7: Probability distribution of bound vs. unbound cluster sizes for (A) scyllo-inositol and (B) glycerol.

Another important aspect of binding is the ability to bind in clusters (Figures 5.6, 5.7). Scyllo-, chiro-inositol, and glucose each bound to the protofibril in a supermolecular form with no difference in their cluster size distributions (Figure 5.7). By contrast, the distribution of cluster sizes for glycerol indicates that it is less likely than the other ligands to form clusters when bound. Ligand molecules within a cluster form self-interactions via
intermolecular hydrogen bonds and nonpolar contacts. Examples of such binding modes are shown in Figure 5.6.

Our previous study on the binding mechanism of inositol with the protofibril of Aβ(16-22)\textsuperscript{17} suggested that the ability of inositol to bind in clusters is a mechanism by which to increase the local concentration of bound ligands at the surface. Our result here is consistent with that study and we speculate that clustered binding modes of small molecules also play a role in the amyloid inhibition of full-length Aβ. In support of these results, amyloid dye molecules in recent simulations\textsuperscript{39} and Congo red in experiments,\textsuperscript{40} were also observed to adopt similar binding modes.

5.3.3 Molecular basis of amyloid inhibition by \textit{scyllo}-inositol

Experimental evidence suggests that residues 17 to 21 (LVFFA) forming the central hydrophobic core of full-length Aβ peptide are responsible for the initiation of Aβ aggregation and β-sheet formation.\textsuperscript{41} Our study in chapter 3 indicated that the likely mode of action of inositol is to bind to exposed surfaces of β-sheets of Aβ(16-22).\textsuperscript{17} The spatial binding probability distributions, depicted in Figure 5.4, indicate that \textit{scyllo}-inositol has the highest preference for the face of the protofibril containing the residues found in the central hydrophobic core (CHC) of Aβ, particularly to the groove formed by the residues F-A-E. By contrast, the inactive molecules, \textit{chiro}-inositol and glucose, were both found to predominantly partition to the β2 face, taking part in nonspecific binding. This result is in accordance with our previous study (Chapter 4)\textsuperscript{17} where we showed that \textit{scyllo}-inositol displays higher binding specificity to phenylalanine: \textit{scyllo}-, unlike \textit{chiro}-inositol, adopts a face-to-face binding mode with phenylalanine. By contrast, due to stereochemistry differences, \textit{chiro}-inositol and glucose molecules do not possess such a binding mode. Consistent with this result, both \textit{chiro}-inositol and glucose displayed less binding specificity for the CHC and were predominantly bound via nonspecific nonpolar contacts.
On the basis of our results, we hypothesize that the CHC, particularly residues Phe20 and Glu21, is a critical binding site of scyllo-inositol for preventing the formation of Aβ42 fibrils. Scyllo-inositol, unlike chiro-inositol, is able to inhibit fibril formation due to its binding specificity for the amyloidogenic core of Aβ42. Our results support the hypothesis that selectively targeting the CHC segment of the Aβ may be a viable approach for the inhibition of Aβ fibrillation.

This binding specificity of an in vitro small molecule inhibitor is modulated by both the stereochemistry of the small molecule and physicochemical properties of the fibril surface that is available for binding. For example, similar to scyllo-inositol, anthraquinone was also suggested to act by binding to residues LVFFA in a recent simulation study by Convertino et al. By contrast, the polyphenol morin, a predominantly hydrophobic compound, was found to not interact with residues in the CHC, but rather bound to residues at the C-terminal end of the protofibril.

Because of its binding specificity, scyllo-inositol may prevent the lateral association of fibrillar aggregates by binding to the β1 face, which has been suggested to be involved in the lateral stacking of protofilaments which ultimately become mature amyloid fibrils. Without the ability to stabilize the cross-β structure by stacking laterally, single-layered protofibrillar structures are unlikely to propagate into mature fibrils.

5.3.4 Comparisons of binding to the protofibril of Aβ(16-22)

In a previous study (see Chapter 3), we investigated the binding mechanism of inositol stereoisomers with protofibrillar oligomers of Aβ(16-22). Unlike Aβ42, the protofibril of Aβ(16-22) presents two chemically identical faces consisting of the same polar, charged and hydrophobic grooves. For this reason, in that study, scyllo-inositol and chiro-inositol were found to bind in similar binding sites on the protofibril of Aβ(16-22). Although binding modes of a small molecule may remain consistent across amyloid morphologies of different peptide sequences, its binding site and binding specificity can be modulated
by the fibril surface properties. As a consequence, comparatively examining the binding modes of several small molecules with differing activities is crucial for determining the molecular mechanism of a small molecule amyloid inhibitor.

Taken together, our study suggests that examining binding to model amyloidogenic peptides is useful for eliminating the large number of possible binding modes and binding sites associated with amyloid inhibitors. However, because the activity of small-molecule inhibitors is specific to both sequence and morphology, it is necessary to ultimately examine their binding with the fibrillar aggregate of the full-length A\textbeta.

5.4 Conclusions

In this study, we have performed extensive atomistic molecular dynamics simulations of scyllo-inositol and its inactive stereoisomer, chiro-inositol, and osmolytes glucose and glycerol, with the protofibril of the full-length A\textbeta. From our simulations, we characterized the stereochemistry-dependent ligand binding modes and their effect on the morphology of A\textbeta protofibrils. Although no difference in the fibril conformation was found in the presence of the ligands, stereospecific binding modes and binding sites were uncovered. Most notably, scyllo-inositol displays the highest binding specificity for residues in the central hydrophobic core of the fibrils of A\textbeta, suggesting that small molecule inhibitors which target this region may be effective in preventing A\textbeta amyloid formation.
Bibliography


[26] The pymol molecular graphics system, version 1.2, Schrodinger Inc.


Chapter 6

Molecular Dynamics simulations of PgaB and monosaccharides of N-acetyl-glucosamine and glucosamine

Contributions: Grace Li conducted the MD simulation part of the research and wrote the section. Dustin Little conducted and interpreted the experimental results. Chris Ing parameterized the partial charges for glucosamine. Régis Pomès, Lynne Howell, Mark Nitz provided editorial input and guidance.
6.1 Summary

Production and de-N-acetylation of the exopolysaccharide poly-\(\beta\)-1,6-N-acetyl-D-glucosamine (PNAG) is important for biofilm formation by Escherichia coli. PgaB is required for the de-N-acetylation and export of PNAG, however the mechanistic details in this process are poorly understood. In this study we perform MD simulations of PgaB in the presence of monosaccharides N-acetylglucosamine (GlcNAc) and glucosammonium (GlcNH\(_3^+\)). Notably the predicted binding sites for GlcNAc and GlcN is comparable to crystal structures of C-terminal domain (PgaB\(_{310-672}\)) of PgaB in the apo-form and complexes of GlcNAc and GlcN. Our results show that GlcNAc preferentially binds the N-terminal domain, whereas glucosammonium binds the C-terminal domain. Taken together, the binding density from our results suggest that a sugar polymer can bind by extending from the active site in the N-terminal domain into the C-terminal domain, wrapping around the protein, by binding in grooves at the surface of PgaB. Our study demonstrates that the methodology employed throughout this thesis is generally applicable for probing and understanding protein-carbohydrate binding.

6.2 Introduction

PgaB is a two-domain outer membrane lipoprotein that is required for the partial de-N-acetylation and export of poly-\(\beta\)-1,6-N-acetylglucosamine (PNAG) into dPNAG, the functionally-relevant form of the exopolysaccharide essential for biofilm formation in a variety of pathogenic bacteria.\(^1\) The structure of a stable core of PgaB has recently been determined and its de-N-acetylation activity characterized.\(^1,2\) Structural and functional characterization show that PgaB is composed of two domains, an N-terminal domain de-N-acetylase, and a C-terminal domain with structural similarity to glycoside hydrolases.\(^1\)

The N-terminal domain of PgaB belongs to the family four carbohydrate esterases (CE4s)\(^3\) and displays length and metal-dependent de-N-acetylation of PNAG oligomers.\(^1\)
PNAG oligomers are de-N-acetylates with low catalytic efficiency,\textsuperscript{1,4} which leads to low levels of PNAG de-N-acetylation (\textasciitilde 3-5\%) observed \textit{in vivo}.\textsuperscript{5,6} The C-terminal domain of PgaB is required for de-N-acetylation and export of PNAG \textit{in vivo}, however the role it plays in both of these processes is unknown.\textsuperscript{5}

Although PgaB functions by binding to PNAG, the molecular basis of the mechanism of export of PNAG by PgaB is currently not determined. Moreover, the binding mode and the length of its substrate polymer are not known. Thus far, efforts to crystallize PgaB with oligosaccharides of varying lengths have been impeded by the weak binding of short sugar polymers (e.g. di- and tri-saccharides), and the insolubility of long sugar polymer chains (e.g. those longer than a hexamer). However, molecular dynamics (MD) simulations are not impeded by these challenges and are well-suited for characterizing carbohydrate-protein interactions.\textsuperscript{7}

In this study, we perform atomistic MD simulations of PgaB with the monosaccharide components of the functionally-relevant polymer dPNAG, \(\beta\)-N-acetyl-glucosamine (Glc-NAc) and protonated \(\beta\)-glucosamine (\(\beta\)-D-GlcNH\textsubscript{3}\textsuperscript{+}), to examine the putative polymeric binding site and binding mode. The weak binding affinity and high dissociation rates of these monosaccharides allows us to use a fragment-based methodology to probe for carbohydrate binding sites by employing large-scale brute force sampling. This approach has been employed successfully in our previous studies to examine the binding mechanism of inositol, a carbohydrate-like amyloid inhibitor, with amyloidogenic peptides and their aggregates.\textsuperscript{8,9}

### 6.3 Material and Methods

A composite structure of PgaB containing residues 43-667 (PgaB\textsubscript{43–667}) was used in our simulations. Residue 62 and all missing side chains in their most favourable non-clashing rotamer to PgaB\textsubscript{42–655}, and residues 610-620 and 647-667 from the PgaB\textsubscript{310–672} structure
were added manually using the software COOT.\textsuperscript{10}

The initial crystal structure has Ni\textsuperscript{2+} bound at its enzymatic active site. Ni\textsuperscript{2+} is octahedrally coordinated with surrounding residues and ligands. All ligands and water molecules from the crystal structures were removed from the composite structure. Histidine protonation states were assigned based on predicted pKa values using the web software PROPKA,\textsuperscript{11–13} and histidine hydrogen-bonding geometries in the initial crystal structures.

Protein and ions were modeled using the AMBER99 force field.\textsuperscript{14} Parameters for Ni\textsuperscript{2+} were approximated using those of Mg\textsuperscript{2+}. The net charge of the protein was -11e. Molecules of sugar were positioned randomly inside the simulation box. The final simulation system comprised of 45 molecules of free monosaccharide (either β-D-GlcNAc or β-D-GlcNH\textsubscript{3}\textsuperscript{+}) at an effective concentration of 100 mM, and 19533 and 19991 water molecules for the β-D-GlcNAc and β-D-GlcNH\textsubscript{3}\textsuperscript{+} simulations, respectively (Figure 6.1). Na\textsuperscript{+} counter-ions were used to neutralize the net charge on the simulation system. The initial volume of the simulation box was 713.6 nm\textsuperscript{3}. To mimic experimental conditions, 100 mM of NaCl was added to the aqueous solution containing β-D-GlcNH\textsubscript{3}\textsuperscript{+}.

The structure of β-D-GlcNAc was generated using the web-based Glycam Biomolecule Builder.\textsuperscript{15} β-D-GlcNH\textsubscript{3}\textsuperscript{+} was obtained from the ZINC database.\textsuperscript{16} The GLYCAM06 force field for carbohydrates\textsuperscript{17} was used to model both β-D-GlcNAc and β-D-GlcNH\textsubscript{3}\textsuperscript{+}. Energy minimization of the β-D-GlcNH\textsubscript{3}\textsuperscript{+} molecule was performed using the software GAUSSIAN-09.\textsuperscript{18} The energy-minimized β-D-GlcNH\textsubscript{3}\textsuperscript{+} structure was consistent with the GLYCAM force field, and new RESP-derived partial atomic charges were computed for β-D-GlcNH\textsubscript{3}\textsuperscript{+} (a net charge of +1e) by fitting to a single HF/6-31G* molecular electrostatic potential (MEP) with a restraint weight of 0.01. MEPs were computed using the CHELPG methodology\textsuperscript{19} with the R.E.D. III software package.\textsuperscript{20} The partial charges were assigned so that the HCNH\textsubscript{3} group summed to a net charge of +1.164e, and the rest of the molecule summed to a net charge of -0.164e. Aliphatic hydrogen atoms were fitted with a zero
partial charge for compatibility with GLYCAM06.

The TIP3P water model\textsuperscript{21} was used to represent the solvent. Version 4.5.5 of the GROMACS software package\textsuperscript{22,23} was used to perform unrestrained all-atom MD simulations with the leapfrog stochastic dynamics algorithm using an integration time-step of 2 femtoseconds and an inverse friction coefficient of 2 picoseconds. Electrostatic interactions were calculated using Particle Mesh Ewald (PME) summation\textsuperscript{24} with a grid size of 0.12 nm and a Coulombic real-space cutoff of 1.1 nm. The Lennard-Jones potential was computed up to 1.2 nm using the GROMACS twin-range cutoff function with a short-range cutoff of 1.1 nm. Covalent bonds involving hydrogen atoms were constrained using the LINCS algorithm.\textsuperscript{25}

The simulation system was first subjected to energy minimization followed by a 1 ns equilibration in the canonical (NVT) ensemble using Berendsen temperature coupling\textsuperscript{26} at 300 K with a coupling constant of 2.0. A second equilibration was performed for 1 ns in the isothermal-isobaric (NpT) ensemble using Berendsen temperature coupling\textsuperscript{26} and isotropic pressure coupling\textsuperscript{26} controlled at 300 K and 1 atm, respectively. Production simulations were performed using the Parrinello-Rahman barostat for pressure coupling.\textsuperscript{27} For each of GlcNAc and $\beta$-D-GlcNH$_3^+$, 13 independent MD simulations of 130 ns were performed, yielding a total of 3.38 $\mu$s of sampling time.

**Analysis Protocol**

To compute spatial binding probability densities of GlcNAc and $\beta$-D-GlcNH$_3^+$, simulation frames were first fitted via root mean square deviation (RMSD) alignment of the protein backbone atoms to an energy-minimized and MD-equilibrated structure. The density maps correspond to the fractional atomic occupancy of GlcNAc or $\beta$-D-GlcNH$_3^+$ molecules binned using a grid with 1 Å resolution. The density map for each of GlcNAc and $\beta$-D-GlcNH$_3^+$ was computed using 165,000 time frames. The Visual Molecular Dynamics (VMD) software package\textsuperscript{28} was used to calculate and graphically render the densities
depicted in Figures 6.6, 6.5, and 6.7.

6.4 Results and Discussion

6.4.1 Protein Dynamics

Extensive molecular dynamics simulations were performed on a composite structure of PgaB in the presence of β-D-GlcNAc and β-D-GlcNH$_3^+$ (Figure 6.1). In all of the simulations, PgaB did not exhibit large structural perturbations (Figure 6.2). The average root mean squared deviation of the backbone atoms of PgaB (determined by excluding loop residues 610 to 623) from the X-ray crystal structure is ∼0.2 nm (Figure 6.2).

An important structural feature of PgaB is the presence of a β-hairpin loop, spanning residues 610 to 623, which is located on top of the C-terminal domain groove and pinches off the groove in the crystal structure (Figure 6.2A). This loop retains its secondary structure in all of our simulations. The loop is highly mobile: residues 610 to 623 possess the highest root-mean-square fluctuations (Figure 6.3), indicating that there is significant loop movement. Moreover, the distribution of atomic distances between C$_\alpha$ atoms of ASN616 and GLY361 shows that the loop adopts a wide range of positions throughout the simulation, from about 0.7 to 2.3 nm relative to the position of the C-terminal groove.
Figure 6.2: (A) Cartoon representation of the structure of chimeric PgaB. The N-terminal and C-terminal domain is represented in Magenta and Cyan, respectively. Ni(II)+ is shown in a ball representation in yellow. The residues 610-623 in the flexible loop region which caps the groove in the C-terminal domain is represented in red. (B) Average RMSD of the protein from the crystal structure, in the presence of GlcNAc, with the loop region spanning residues 610-623 removed. Error in the mean is represented by the pink shaded region.

(Figure 6.3). We hypothesize that the loop and its flexibility are involved in the export process whereby its movement over the binding groove facilitates the binding of the substrate polymer in the C-terminal domain.

6.4.2 Comparisons of the binding modes of GlcNAc and β-D-GlcNH$_3^+$

Both sugar molecules adopt similar binding modes involving a combination of hydrogen bonds between the sugar and charged side chains and nonpolar stacking of aromatic moieties. As depicted in Figure 6.4, glucosamine molecules form nonpolar contacts and hydrogen bonds with the side chains of polar (i.e. GLN) and acidic residues (i.e. ASP and GLU) via its hydroxyl and charged groups in both the N-terminal and C-terminal domain of the protein.

Notably, consistent with the crystal structure of the C-terminal domain of PgaB with bound glucosamine, our simulation results indicate that both GlcNAc and β-D-
Figure 6.3: Structural dynamics of PgaB. (A) Average root-mean-square fluctuation of the backbone residues of PgaB in the presence of GlcNAc and glucosamine. (B) Structures of PgaB from different simulation time frames indicating fluctuations in the protein. PgaB residues are coloured by their secondary structure.
Figure 6.4: Example binding modes of glucosamine in the C-terminal groove. β-D-GlcNH$_3^+$ molecules bind in the C-terminal domain carbohydrate-binding groove by forming intermolecular hydrogen bonds with polar and charged residues and forming nonpolar contacts by stacking face-to-face with aromatic moieties.

GlcNH$_3^+$ possess binding sites at Trp613 and Trp552 (Figure 6.5), which are conserved residues located in this groove. Furthermore, both GlcNAc and β-D-GlcNH$_3^+$ molecules were observed to bind in clusters, where the molecules in the cluster form intermolecular hydrogen bonds.

Although GlcNAc and β-D-GlcNH$_3^+$ have similar binding modes, they differ in their binding sites. The spatial binding probability density of GlcNAc, depicted in Figure 6.6, suggests that binding sites for GlcNAc are predominantly located in proximity to the active site (in the N-terminal domain), consistent with the function of PgaB as a PNAG de-acetylase. In particular, the densities in the inter-domain crevice beneath the loop spanning residue numbers 309 to 314 suggest that this region may be involved in substrate binding. This result corroborates the hypothesis that the binding of PNAG to PgaB may initiate in this region, prior to binding at the active site.

By contrast, the spatial distribution of β-D-GlcNH$_3^+$ molecules depicted in Figure 6.6 indicates that they preferentially bind to the C-terminal domain. In particular, a significant binding density was found in a groove heavily lined with acidic and aromatic residues, which are residues frequently found in carbohydrate-binding sites (Figure 6.5) and carbohydrate binding modules. Because this C-terminal groove contains a large
Figure 6.5: Spatial probability densities of GlcNAc and $\beta$-D-GlcNH$_3^+$ in the C-terminal groove. Densities are shown at an occupancy of 0.15 for (A) $\beta$-D-GlcNH$_3^+$ and (B) GlcNAc in the binding groove of the C-terminal domain. The residues in the groove are shown in stick representations.
Figure 6.6: Spatial probability densities of bound GlcNAc (purple) and \( \beta\text{-D-GlcNH}_3^+ \) (pink). Binding densities overlapped with a cartoon representation of the energy minimized crystal structure of PgaB. The protein is shown facing the active site in (A) and (C), and shown facing the C-terminal domain in (B) and (D). In each view, binding densities are depicted at occupancies of 0.15 in (A)-(B), and 0.25 in (C) - (D). In our coloring scheme, residue numbers 43 to 310 and numbers 311 to 667 represent N- (green) and C-terminal (cyan) domains, respectively.
number of acidic residues, positively charged glucosamine preferentially binds in this groove. Our results indicate that the entire length of the groove can bind \( \beta\)-D-GlcNH\(_3^+\) molecules, suggesting that this groove is likely to accommodate a linear polymer of \( \beta\)-D-GlcNH\(_3^+\) (Figure 6.5). Our above results suggest that this C-terminal groove of PgaB is a binding site for the de-acetylated PNAG (dPNAG). In contrast to the spatial distributions of GlcNAc and \( \beta\)-D-GlcNH\(_3^+\), salt ions, which do not possess the sugar ring, do not significantly bind the protein (Figure 6.7).

![Figure 6.7: Spatial probability distribution of sodium (blue) and chloride (grey) ions depicted at an isovalue of 0.005. The protein is depicted in a cartoon representation, with N- and C-terminal domains colored in green and cyan, respectively.](image)

Together, the morphology of the overall binding density of both GlcNAc and \( \beta\)-D-GlcNH\(_3^+\) paints a molecular picture of the putative binding mechanism and export of PNAG. As PNAG is being shuttled through PgaB on to the next protein in the system, it initially binds at the cleft between the N- and C-terminal domains to reach the active site in order to undergo de-N-acetylation. Our simulations suggest that as the polymer is
de-acetylated, the binding affinity of the resulting dPNAG to the C-terminal domain is increased. This hypothesis is currently being validated experimentally. We propose that PgaB-CT facilitates the export of the substrate by favorably binding its de-acetylated region.

This study demonstrates the utility of large-scale MD simulations in predicting carbohydrate-protein binding modes. On the basis of the weak binding affinities and high dissociation rates of monosaccharides, (∼1-5 mM binding affinity), we observed spontaneous binding events on the nanosecond time scale at 300 K. By using a high concentration of molecules, we were able to effectively sample the binding of monosaccharides to the protein surface by performing MD simulations using a conventional sampling approach and performing independent MD simulations.

From our simulations, we were able to accurately predict binding modes and binding sites of GlcNAc and glucosamine for PgaB. The results of this study, in combination with experimental studies, have lead to a mechanistic hypothesis for the role of PgaB in the export of PNAG. Our study demonstrates the utility of large-scale bruteforce MD simulations in predicting carbohydrate-protein binding modes when combined with a force field that can reproduce sugar-protein interactions with accuracy (for example, GLYCAM\textsuperscript{17}). Force fields for molecular simulations of carbohydrates have been recently reviewed here.\textsuperscript{7} Furthermore, we demonstrated the utility of computing spatial probability binding densities using monosaccharides to map carbohydrate binding sites. A recent study suggests that by building 3D density binding maps of carbohydrates and using this data with additional computational algorithms can increase the accuracy of identifying putative carbohydrate binding sites.\textsuperscript{29} As part of our future work, it may be interesting to employ a similar approach to investigate the binding of oligosaccharides to PgaB.
6.5 Acknowledgements

This work was done in collaboration with Dustin Little, Dr. Lynne Howell and Dr. Mark Nitz. The author is grateful for Dr. Nilu Chakrabarti and Chris Ing for assisting in the calculation of the partial charges for glucosamine.
Bibliography


Chapter 7

Conclusions and Future Directions

7.1 Conclusions

The primary objective of my research was to elucidate the molecular basis of the activity of scyllo-inositol, a putative therapeutic for the treatment of Alzheimer’s disease (AD). Specifically, I have investigated the effect of inositol on the structure and thermodynamics of amyloid aggregation. The differential effects on the aggregation equilibrium were determined from microsecond simulations of peptide monomers as well as small and large aggregates. To the best of my knowledge, this thesis represents the most complete molecular dynamics simulation study of a small-molecule amyloid inhibitor (scyllo-inositol) to date. My work provides insight that could be used for designing novel and higher-efficacy derivatives of putative drugs which may prevent the onset and progression of AD and amyloid-related disorders.

Beginning in Chapter 3, I performed systematic simulations of simple amyloidogenic peptide models with scyllo- and chiro-inositol to examine the role of backbone binding on amyloid inhibition. My results indicated that although inositol predominately interacts with the peptide backbone, the binding affinity is low and remains in the millimolar range. Moreover, backbone binding was independent of stereochemistry and did not appear to
be sufficient to impede peptide dimerization. Taken together, the results in this chapter suggest that amyloid inhibition by inositol cannot be accounted for by generic binding to the peptidic backbone alone. In this study, it was hypothesized that amyloid inhibition by inositol is likely to involve sequence-specific interactions with amino-acid side chains as well as binding to specific aggregate morphologies.

To investigate the role of sequence-specific interactions between inositol and aggregates of pathogenic peptides, I examined the binding of inositol stereoisomers, successively, to monomers, disordered oligomers, and \( \beta \)-sheet aggregates of A\( \beta \)(16-22) (Chapter 4). A key result of this study was that the \( K_{eq} \) of inositol (\( \sim 0.2 \) - 0.5 mM) for the \( \beta \)-oligomer is commensurate with the concentration at which inhibition of amyloid formation by A\( \beta \)42 is observed \textit{in vitro}. Both \textit{scyllo}- and \textit{chiro}-inositol exhibited similar binding affinities with all peptide states considered. However, \textit{scyllo}-inositol was found to possess a stereospecific face-to-face stacking stacking mode with the Phe side chains and a higher propensity for hydrogen bonding, which together suggest a molecular basis for measured differences in activity. Specifically, cooperative binding modes of inositol at grooves on the surface of the \( \beta \)-oligomer of A\( \beta \)(16-22) suggest a possible mechanism of fibril inhibition whereby inositol prevents the lateral association or stacking of protofibrillar \( \beta \)-sheet oligomers. Because inositol was found to adopt carbohydrate-like binding modes at the fibril core of A\( \beta \) amyloid, carbohydrate-based small-molecule derivatives may be promising for the development of novel therapeutics for AD.

The above results (presented in Chapters 3 and 4) have led me to hypothesize that \textit{scyllo}-inositol is likely to act on the protofibrillar form of A\( \beta \)42. In Chapter 5, I examined the binding of \textit{scyllo}-inositol to the protofibrillar form of A\( \beta \)42. \textit{Chiro}-inositol, glucose and glycerol, which were shown not to inhibit A\( \beta \)42 amyloid formation, were taken as negative controls in my study. There was no difference in the protofibril conformation in the presence of ligands. However, differences in stereochemistry between the molecules led to differential binding modes and propensities. Notably, we found that \textit{scyllo}-inositol
displays the highest binding specificity to the region of the surface containing the sequence Leu-Val-Phe-Phe-Ala-Glu, the central hydrophobic core (CHC) of Aβ fibrils. The binding specificity of scyllo-inositol for the CHC of Aβ42 suggest a mechanism of inhibition whereby scyllo-inositol prevents the growth of protofibril oligomers into mature amyloid fibrils by disrupting the lateral stacking of these oligomers.

Taken together, the results in Chapters 3 to 5 have demonstrated the applicability of MD simulations in providing insight into how small-molecule amyloid inhibitors may bind to amyloidogenic species and intrinsically disordered peptides.

More generally, my work suggests that MD simulations can be effectively applied to probe weak and transient molecular interactions, which can be difficult to probe using experimental techniques. Specifically, understanding the molecular basis of weak interactions is important for protein-carbohydrate binding.\textsuperscript{1,2} In Chapter 6, we performed MD simulations of PgaB, a key protein in the biofilm formation pathway, in the presence of monosaccharides, GlcNAc and glucosamine (monomeric components of the polymeric substrate). A binding surface and binding mode for the polymeric substrate (PNAG) of PgaB were predicted from my simulations. Notably, results were consistent with electron densities derived from X-ray crystallography studies. This work indicates that MD simulations combined with the systematic methodology employed this thesis is applicable for examining carbohydrate-protein interactions.

Throughout this thesis, a systematic comparative approach was utilized as a central methodological framework in each of my studies. Simulations of amyloidogenic peptides and their aggregates of increasing sequence length and size were successively carried out in the presence and absence of scyllo-inositol, and its inactive isomer, chiro-inositol (see Section 1.11 for rationale and study design details). This approach was instrumental in the investigation of the binding mechanism of inositol. Because inositol displays relatively low binding affinity, a statistically significant number of spontaneous binding and unbinding events were observed using conventional (unbiased) MD simulations. Fur-
Therefore, the presence of inositol at a high concentration in the simulations was both relevant for understanding its activity and for increasing the likelihood of binding events. Hence, an important result of this work is that conventional (unbiased) molecular dynamics simulations in combination with massively-repeated sampling is well-suited for quantitative studies of systems involving weak protein-ligand binding. Moreover, utilizing such a comparative methodology will be important for understanding the molecular mechanism of action of ligands with \textit{a priori} unknown binding sites and binding modes. In the future, with the increasing availability of cheaper and faster computers, we predict that this approach can be readily applied towards studies of related systems, such as those involving small molecules which binding to intrinsically disordered proteins and protein-carbohydrate interactions.

7.2 Future directions

7.2.1 Effect of inositol on the growth of amyloid fibrils

My work led to the hypothesis that binding to the surface of $\beta$-sheet oligomers may lead to the inhibition of amyloid fibrils. MD simulations of protofibrils with free monomers can be carried out in the presence and absence of inositol to directly investigate the effect of inositol on the binding equilibrium of free monomers to pre-existing oligomers. Analysis of the spatial distribution of the bound monomers around the protofibril will reveal how peptides may partition around a protofibril and whether the presence of scyllo-inositol will affect the binding equilibrium of free monomers to pre-formed $\beta$-sheets.
7.2.2 Effect of inositol on amyloid aggregation in the presence of lipid membranes

As previously stated in the introduction (Chapter 1), it is currently hypothesized that the interaction of amyloid with cellular membranes may cause cellular toxicity by disrupting their integrity.\(^3\) A previous MD simulation study from my lab indicated that $\beta$-sheet formation of amyloidogenic peptides is catalyzed by the presence of a hydrophobic surface.\(^4\) An interesting question to probe is how do amyloid inhibitors such as inositol affect the aggregation of $\text{A} \beta$ peptides in the presence of lipid membranes? In previous studies, inositol has been shown to neutralize the toxicity of $\text{A} \beta$ oligomers.\(^5\) It is possible that inositol may prevent toxicity by disfavoring the binding of monomers and small aggregates onto the membrane interface, or impeding the formation of toxic $\beta$-sheets at the interface.

Using a similar systematic approach that was carried out in this thesis, simulations which examine the effect of inositol on the binding of model amyloidogenic peptide aggregates can be carried out in the presence of (1) a membrane-mimetic octane slab\(^4\) or (2) lipid bilayers of differing composition. With the addition of inositol, molecular simulations of these systems will provide insight into the molecular basis of inhibition of amyloid toxicity. These studies can be combined with large-scale sampling and enhanced sampling methods\(^6\) to accelerate attaining the convergence of properties of interest.\(^7\)

7.2.3 Effect of inositol on monomers of longer amyloidogenic peptides

A limitation in my study is that the monomeric peptides examined in this study were either model peptides or shorter fragments of longer amyloidogenic peptides. Small molecule inhibitors may have differing effects on the peptide conformation and aggregation of longer amyloidogenic peptides such as the full-length $\text{A} \beta 40$ or $\text{A} \beta 42$ peptides. To address this possibility, it will be useful to perform simulations to examine the binding
mechanism of chiro- and scyllo-inositol with the full-length Aβ. However, in order to determine the binding equilibria of inositol with the full-length Aβ peptide, enhanced sampling may be needed to attain convergence in the statistical properties of interest at equilibrium so that the studies may be completed in a feasible time frame.

### 7.2.4 Possible experimental studies

My work has currently led to several hypotheses which may be beneficial to cross-validate experimentally. Most notably, our studies led to the conclusion that scyllo-inositol possesses the stereospecificity which may affect aggregation by targeting the central hydrophobic core of Aβ. Moreover, inositol molecules were found to bind to surfaces of protofibrils in clusters, suggesting that derivatives of inositol where multiple monomers of scyllo-inositol are linked together may be a promising avenue to explore. In the future, synergistic studies combining simulations and experimental characterization will be beneficial for the progress towards elucidating the molecular mechanism of amyloid inhibition, and ultimately, a cure for amyloid disorders.
Bibliography


Appendices
Appendix A

Supplementary Information for Chapter 4: Binding Mechanism of Inositol Stereoisomers to Monomers and Aggregates of $\text{A}\beta(16-22)$

The analysis presented in this appendix was originally published online as supplementary material for our article in the Journal of Physical Chemistry B.


Contributions: Grace Li wrote and performed all analysis in this appendix.
A.1 Monomer

Figure A.1.1: The end-to-end distance distribution of monomeric Aβ(16-22) conformations separated by the number of intra-peptide hydrogen bonds, $n_{hb}$. The curve in yellow is the end-to-end distribution over all peptide conformations. The peaks at (A) 0.55 nm, (B) 0.9 nm, and (C) 1.4 nm generally correspond to peptide conformations with a β-bridge between residues Leu17 and Phe20, with 1 or more intermolecular hydrogen bonds, and with no intra-peptide hydrogen bonds, respectively.
Figure A.1.2: Snapshots of representative monomer conformations at each peak of the end-to-end distribution. (A) Compact conformations with end-to-end distances of \( \sim 0.59 \) nm correspond to an ensemble of peptides with a \( \beta \)-bridge formed between Leu and Phe. (B) Representative conformations of the peak at 0.8-0.9 nm are predominantly peptides with a single hydrogen-bonded turn formed between residues Val and Phe20. Furthermore, the end-to-end distance of this peak also captures conformers with salt bridges, which account for the populations with \( n_{hb} = 3 \) or \( n_{hb} = 4 \). (C) Extended peptide conformations have no intra-peptide hydrogen bonds, and have end-to-end distances corresponding to the peak in the distribution at \( \sim 1.4 \) nm.

<table>
<thead>
<tr>
<th></th>
<th>2:1 (inositol:peptide)</th>
<th>15:1 (inositol:peptide)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( scyllo )</td>
<td>( chiro )</td>
</tr>
<tr>
<td>Nonpolar</td>
<td>37</td>
<td>45</td>
</tr>
<tr>
<td>Nonpolar and Hbonds</td>
<td>35</td>
<td>34</td>
</tr>
<tr>
<td>Hbonds</td>
<td>28</td>
<td>21</td>
</tr>
</tbody>
</table>

Table A.1.1: Fraction of inositol molecules (in %) bound to nonpolar and polar groups of the monomer.
Figure A.1.3: Time-averaged number of hydrogen bonds (top), and nonpolar (bottom) made by inositol to each residue of the monomer at an inositol:peptide molar ratio of 2:1.
## A.2 Disordered Oligomer

<table>
<thead>
<tr>
<th>2:4 (inositol:peptide)</th>
<th>scyllo</th>
<th>chiro</th>
<th>$s_{\pi,scyllo}$</th>
<th>$s_{\pi,chiro}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonpolar</td>
<td>36</td>
<td>51</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Nonpolar and Hbonds</td>
<td>20</td>
<td>23</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Hbonds</td>
<td>44</td>
<td>26</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>15:4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scyllo</td>
<td>26</td>
<td>36</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>chiro</td>
<td>51</td>
<td>48</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>$s_{\pi,scyllo}$</td>
<td>23</td>
<td>16</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>$s_{\pi,chiro}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45:4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scyllo</td>
<td>30</td>
<td>36</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>chiro</td>
<td>48</td>
<td>47</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$s_{\pi,scyllo}$</td>
<td>23</td>
<td>17</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>$s_{\pi,chiro}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table A.2.1: Fraction of inositol molecules (in %) bound to nonpolar and polar groups of the disordered oligomer.
Figure A.2.1: The equilibrium number of intermolecular (A) hydrogen bonding, and (B) nonpolar contacts per peptide in the disordered oligomer.
Figure A.2.2: Effect of inositol on the aggregation of Aβ(16-22). Data for inositol:peptide molar ratios 15:4 and 45:4 are shown on the left and right panel, respectively. (A)-(B) Time evolution of the peptide-peptide intermolecular nonpolar contacts per peptide. (C)-(D) Intermolecular peptide-peptide hydrogen bonds per peptide, and (E)-(F) intramolecular peptide-peptide hydrogen bonds per peptide. The time series were smoothed using a running average with a window of length 500 ps.
Figure A.2.3: Time-evolution of peptide self-aggregation to form the disordered oligomer at inositol:peptide molar ratios of (A) 15:4 and (B) 45:4.
Figure A.2.4: Time-averaged number of hydrogen bonds (top), and nonpolar (bottom) made by inositol to each residue of the disordered oligomer at an inositol:peptide molar ratios of (A) 2:4 and (B) 15:4.
A.3 \( \beta \)-oligomers

<table>
<thead>
<tr>
<th></th>
<th>scyllo</th>
<th>chiro</th>
<th>( s_{\text{scyllo}} )</th>
<th>( s_{\text{chiro}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonpolar</td>
<td>23</td>
<td>33</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Nonpolar and Hbonds</td>
<td>52</td>
<td>48</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Hbonds</td>
<td>25</td>
<td>19</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>scyllo</th>
<th>chiro</th>
<th>( s_{\text{scyllo}} )</th>
<th>( s_{\text{chiro}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonpolar</td>
<td>19</td>
<td>29</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Nonpolar and Hbonds</td>
<td>58</td>
<td>56</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Hbonds</td>
<td>23</td>
<td>15</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>scyllo</th>
<th>chiro</th>
<th>( s_{\text{scyllo}} )</th>
<th>( s_{\text{chiro}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonpolar</td>
<td>19.7</td>
<td>30</td>
<td>0.4</td>
<td>1</td>
</tr>
<tr>
<td>Nonpolar and Hbonds</td>
<td>56.7</td>
<td>54</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>Hbonds</td>
<td>23.5</td>
<td>16</td>
<td>0.3</td>
<td>1</td>
</tr>
</tbody>
</table>

Table A.3.1: Fraction of inositol molecules (in %) bound to nonpolar and polar groups of the \( \beta \)-oligomer.
Figure A.3.1: Initial model of the $\beta$-oligomer. (A) Side view; (B) down the fibril axis, and (C) looking down at the face of the sheet. The initial starting structure of the oligomer has dimensions of 2.2 nm x 2.5 nm x 3.65 nm. (D) Snapshot of the $\beta$-oligomer from a simulation without inositol ($t = 80$ ns).
Figure A.3.2: Time-averaged number of hydrogen bonds (top), and nonpolar (bottom) made by inositol to each residue of the β-oligomer at inositol:peptide molar ratios of (A) below 1:1, and (B) 64:16 (effective inositol concentration of 208 mM).