



University of Kentucky  
UKnowledge

---

Theses and Dissertations--Chemical and  
Materials Engineering

Chemical and Materials Engineering

---

2017

## UNDERSTANDING GLYCOSIDE HYDROLASE PROCESSIVITY FOR IMPROVED BIOMASS CONVERSION

Suvamay Jana

*University of Kentucky*, [jsa236@g.uky.edu](mailto:jsa236@g.uky.edu)

Digital Object Identifier: <https://doi.org/10.13023/ETD.2017.026>

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

---

### Recommended Citation

Jana, Suvamay, "UNDERSTANDING GLYCOSIDE HYDROLASE PROCESSIVITY FOR IMPROVED BIOMASS CONVERSION" (2017). *Theses and Dissertations--Chemical and Materials Engineering*. 71.

[https://uknowledge.uky.edu/cme\\_etds/71](https://uknowledge.uky.edu/cme_etds/71)

This Doctoral Dissertation is brought to you for free and open access by the Chemical and Materials Engineering at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Chemical and Materials Engineering by an authorized administrator of UKnowledge. For more information, please contact [UKnowledge@lsv.uky.edu](mailto:UKnowledge@lsv.uky.edu).

## **STUDENT AGREEMENT:**

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

## **REVIEW, APPROVAL AND ACCEPTANCE**

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Suvamay Jana, Student

Dr. Christina M. Payne, Major Professor

Dr. Thomas D. Dziubla, Director of Graduate Studies

UNDERSTANDING GLYCOSIDE HYDROLASE PROCESSIVITY FOR IMPROVED  
BIOMASS CONVERSION

---

DISSERTATION

---

A dissertation submitted in partial fulfillment of the  
requirements for the degree of Doctor of Philosophy in the  
College of Engineering  
at the University of Kentucky

By

Suvamay Jana

Lexington, Kentucky

Director: Dr. Christina M. Payne, Professor of Chemical and Materials Engineering

Lexington, Kentucky

2017

Copyright © Suvamay Jana 2017

## ABSTRACT OF DISSERTATION

### UNDERSTANDING GLYCOSIDE HYDROLASE PROCESSIVITY FOR IMPROVED BIOMASS CONVERSION

In nature, organisms secrete synergistic enzyme cocktails to deconstruct crystalline polysaccharides, such as cellulose and chitin, to soluble sugars. The cocktails consist of multiple classes of processive and non-processive glycoside hydrolases (GH) that aid in substrate accessibility and reduce product inhibition. Processive GHs attach to chain ends and hydrolyze many glycosidic linkages in sequence to produce disaccharide units before dissociation, and as such, are responsible for the majority of hydrolytic bond cleavages. Accordingly, processive GHs are targets for activity improvements towards efficient and economical biomass conversion. However, the mechanism and factors responsible for processivity are still not understood completely at the molecular level. Specifically, the relationship between processive GH function and the enzyme active site topology and chemical composition has yet to be elucidated. Using molecular simulation and free energy calculations, this work presents a molecular-level understanding of the protein-carbohydrate interactions governing processive GHs, which will facilitate rational design of GHs for enhanced biomass conversion. We hypothesize that processive GHs, having long tunnels or deep active site clefts, will allow more amino acids to interact with the ligand and exhibit strong ligand binding and low substrate dissociation rate constants; whereas non-processive enzymes, having more open tunnels or clefts, will exhibit comparatively weak binding and high dissociation rate constants. Moreover, the ligand binding free energy of a processive enzyme must also be more thermodynamically favorable than the work required to decrystallize a polymer from the substrate matrix. We selected the *Serratia marcescens* Family 18 chitinase model system, including processive chitinases, ChiA and ChiB, and a non-processive chitinase, ChiC, to test our hypotheses. We find that processive ChiA and ChiB exhibit ligand binding free energies that are more thermodynamically favorable than the work to decrystallize a chito-oligosaccharide from the crystalline chitin surface, which is essential for forward processive movement. The non-processive ChiC binds chito-oligosaccharides with a free energy that is significantly less favorable than the work of decrystallization. In general, our findings suggest that processive GH function necessitates tight binding within the enzyme active site. We also observed that aromatic and polar residues close to the catalytic center of ChiA and ChiB have a greater effect on ligand binding and processivity than the residues at the entrance

or exit of the cleft. Mutation of active site aromatic and polar residues generally resulted in reduction in processivity and substantial reduction in substrate binding. Overall, our work demonstrates the existence of a fundamental relationship between ligand binding free energy and processive GH active site characteristics.

KEYWORDS: Biomass, glycoside hydrolase, chitinase, processivity, free energy,  
molecular dynamics

Suvamay Jana

---

February 14<sup>th</sup>, 2017

---

UNDERSTANDING GLYCOSIDE HYDROLASE PROCESSIVITY FOR IMPROVED  
BIOMASS CONVERSION

By

Suvamay Jana

Dr. Christina M. Payne

Director of Dissertation

Dr. Thomas Dziubla

Director of Graduate Studies

February 14<sup>th</sup>, 2017

Dedication

*To my beloved parents*

## **Acknowledgements**

This doctoral research was only possible through the financial supports from the National Science Foundation (NSF), the Research Council of Norway, and the Oak Ridge Associated Universities (ORAU) Ralph E. Powe Junior Faculty Award. The much needed computational time for the research was provided by the NSF Extreme Science and Engineering Discovery Environment (XSEDE), and University of Kentucky.

Apart from the funding agencies, I would like to express my deepest gratitude to my supervisor Dr. Christina M. Payne for her support and guidance throughout my doctoral research. She was ever ready in answering my queries, explaining to me the intricate details of biomolecule simulation and endlessly inspiring me for the quest of excellence. I would also like to thank the members of my dissertation committee: Dr. Thomas Dziubla, Dr. Stephen Rankin, and Dr. Craig Vander Kooi for their useful and constructive recommendations on this research work.

I would like to extend my gratitude to our experimental collaborators from Norwegian University of Life Sciences: Dr. Morten Sørli, and Dr. Anne Grethe Hamre for sharing their thoughts about enzyme processivity. Dr. Hamre especially did most of the processivity and binding affinity measurements I compared with my simulation results in this dissertation and deserves special thanks. I am also thankful to our collaborators from the National Renewable Energy Laboratory: Dr. Brandon Knott, and Dr. Vivek Bharadwaj for teaching me how to model enzyme catalytic reactions and sharing their simulation codes with me. I would also like to thank Dr. Mats Sandgren and Dr. Jerry Ståhlberg from Swedish University of Agricultural Science for hosting me at



Uppsala during the summers. It was a great pleasure to work in your research group and learning more about cellulases.

I also thank my lab colleagues Abhishek, and Yue Yu for being so supportive and nice to me. Specially Abhishek, I really enjoyed conversations with you which undoubtedly helped my graduate research.

Finally I would like to thank my family and friends for their love and encouragement. Especially warm regards to my parents for always believing in me. Last but not the least, a very special thank to my beloved wife Aindrila!! Thank you for your love and motivation that helped me to finish this work.

Suvamay Jana

## Table of Contents

<b>Acknowledgements</b> .....	iii
<b>Table of Contents</b> .....	v
<b>List of Tables</b> .....	xii
<b>List of Figures</b> .....	xiv
<b>Chapter 1 - Introduction</b> .....	1
1.1 Motivation.....	1
1.2 Research background .....	3
1.2.1 Crystalline polysaccharides .....	3
1.2.2 Glycoside hydrolases (GH).....	7
1.2.2.1 GH active site topology .....	9
1.2.2.2 Catalytic mechanism of GHs .....	10
1.2.2.3 GH mode of action.....	12
1.2.2.4 GH processivity .....	12
1.2.2.4.1 Contribution of GH active site in processivity .....	13
1.2.2.4.2 Measurement of processivity .....	15
1.2.2.4.3 Thermodynamics of GH processivity .....	16
1.2.3 <i>Serratia marcescens</i> chitinases as a model system to study processivity.....	18
1.2.3.1 <i>S. marcescens</i> Family 18 chitinases.....	20
1.2.3.2 Catalytic mechanism of Family 18 chitinases .....	23
1.3 Outline of dissertation.....	25
1.3.1 Overall active site contributions to binding and processivity (Chapter 3).....	25

1.3.2 Contributions of aromatic and polar residues to ligand binding and processivity (Chapters 4, 5 and 6).....	26
1.3.3 Rate-limiting step in GH processivity (Chapter 7) .....	26
<b>Chapter 2 - Computational Methodology</b> .....	28
2.1 Introduction.....	28
2.2 Molecular dynamics simulation (MD Simulation) .....	28
2.3 Free energy calculation methods .....	33
2.3.1 Free energy perturbation with replica exchange molecular dynamics (FEP/ $\lambda$ - REMD).....	33
2.3.2 Thermodynamic integration (TI) .....	37
2.3.3 Umbrella sampling.....	41
<b>Chapter 3 - Thermodynamic Relationships with Processivity in <i>Serratia marcescens</i> Family 18 Chitinases</b> .....	45
3.0 Abstract.....	45
3.1 Introduction.....	46
3.2 Materials and methods .....	49
3.2.1 Chemicals.....	49
3.2.2 Enzymes.....	49
3.2.3 Isothermal titration calorimetry experiments.....	52
3.2.4 Analysis of calorimetric data .....	53

3.2.5 Simulation methodology.....	54
3.3 Results.....	58
3.3.1 Binding of (GlcNAc) <sub>6</sub> to ChiA, ChiB, and ChiC determined with FEP/ $\lambda$ -REMD.....	58
3.3.2 Binding of (GlcNAc) <sub>6</sub> to ChiA and ChiC determined with ITC.....	59
3.3.3 Parameterization of the entropic term.....	62
3.3.4 Active site dynamics from MD simulations.....	63
3.4 Discussions.....	68
3.5 Conclusions.....	79
<b>Chapter 4 - Aromatic-mediated Carbohydrate Recognition in Processive <i>Serratia marcescens</i> Chitinases.....</b>	<b>81</b>
4.0 Abstract.....	81
4.1 Introduction.....	82
4.2 Materials and Methods.....	88
4.2.1 Chemicals.....	88
4.2.2 Enzymes.....	88
4.2.2.1 Site directed mutagenesis.....	89
4.2.2.2 Construction of His <sub>10</sub> -tagged double mutants.....	90
4.2.2.3 Protein expression and purification of single mutants.....	91
4.2.2.4 Protein expression and purification of double, inactive mutants.....	91
4.2.3 Degradation of chitin for determination of enzyme processivity.....	92

4.2.4 High performance liquid chromatography (HPLC).....	93
4.2.5 Isothermal titration calorimetry experiments.....	93
4.2.6 Analysis of calorimetric data.....	94
4.2.7 Molecular dynamics simulations.....	95
4.2.8 Thermodynamic integration (TI).....	97
4.3 Results.....	98
4.4 Discussions.....	105
4.4.1 Absolute and relative binding free energies.....	105
4.4.2 ChiA aromatic residues.....	108
4.4.3 ChiB aromatic residues.....	113
4.4.4 Trends across GH Families.....	115
4.5 Conclusions.....	118
<b>Chapter 5 - Processivity, Substrate Positioning, and Binding: The Role of Polar Residues in a Family 18 Glycoside Hydrolase.....</b>	<b>120</b>
5.0 Abstract.....	120
5.1 Introduction.....	121
5.2 Materials and Methods.....	125
5.3 Results.....	135
5.4 Discussion.....	152
5.5 Conclusions.....	162

<b>Chapter 6 - Carbohydrate-Polar Interactions Dictate Substrate Binding and Processivity in a Family 18 Chitinase.....</b>	<b>164</b>
6.1 Introduction.....	164
6.2 Materials and methods .....	169
6.2.1 Chemicals.....	169
6.2.2 Enzymes.....	170
6.2.2.1 Site directed mutagenesis.....	170
6.2.2.2 Protein expression and purification of single, active mutants .....	171
6.2.3 Time course activity assay.....	171
6.2.4 High performance liquid chromatography (HPLC).....	172
6.2.5 Determination of A- and b-values for activity.....	172
6.2.6 Degradation of chitin for determination of enzyme processivity .....	173
6.2.7 Computational methods .....	173
6.2.7.1 Construction of ChiB model systems .....	173
6.2.7.2 MD simulations.....	175
6.2.7.3 TI.....	176
6.2.7.4 Umbrella sampling (US).....	178
6.3 Results.....	179
6.3.1 Initial kinetic characterization.....	179
6.3.2 Degradation of chitin .....	180
6.3.3 MD simulations and free energy calculations.....	181

6.4 Discussions .....	189
6.4.1 Ligand binding free energy versus processive ability of ChiB .....	189
6.4.2 Polar mutations .....	191
6.4.2.1 Asp-316 (-1 site) .....	191
6.4.2.2 Tyr-145 (+1 site) .....	193
6.4.2.3 Arg-294 (+1 site) .....	196
6.4.2.4 Glu-221 (+3 site) .....	198
6.5 Conclusions .....	202
<b>Chapter 7 - The Thermodynamic Relationship of Chitinase Binding, Processive Threading, and Work Required to Decrystallize <math>\beta</math>-Chitin .....</b>	<b>204</b>
7.1 Introduction .....	204
7.2 Computational methodology .....	208
7.2.1 Processive sliding and catalytic activation .....	208
7.2.1.1 Construction of the Michaelis complex .....	209
7.2.1.2 Modeling of the presliding state .....	210
7.2.1.3 Modeling of the sliding state .....	211
7.2.1.4 Free energy barrier estimation for processive sliding and catalytic activation from umbrella sampling .....	212
7.2.2 Decrystallization of $\beta$ -chitin .....	213
7.2.2.1 Modeling of the $\beta$ -chitin substrate .....	214
7.2.2.2 Umbrella sampling .....	215
7.3 Results .....	216

7.3.1 Free energy barrier estimation for processive sliding and catalytic activation .....	216
7.3.2 MD simulations of the presliding, sliding, and the Michaelis complex .....	219
7.3.3 $\beta$ -chitin MD simulation and estimation of work of decrystallization .....	222
7.4 Discussions .....	226
7.4.1 Comparison with cellulases .....	226
7.4.2 Effect of crystalline substrate.....	230
7.5 Conclusions.....	235
<b>Chapter 8 – Conclusions and Future Directions.....</b>	<b>237</b>
<b>Appendix.....</b>	<b>246</b>
<b>References.....</b>	<b>280</b>
<b>Vita.....</b>	<b>300</b>



## List of Tables

<b>Table 3.1</b> Primers used for site-directed mutagenesis and PCR amplification .....	50
<b>Table 3.2</b> Absolute ( $\Delta G^{\circ}_b$ ) ligand binding free energy calculated from FEP/ $\lambda$ -REMD at 300 K and pH 6 .....	59
<b>Table 3.3</b> Thermodynamic parameters for (GlcNAc) <sub>6</sub> binding to ChiA, ChiB, and ChiC of <i>Serratia marcescens</i> at $t = 30$ °C, pH = 6.0 .....	62
<b>Table 3.4</b> Thermodynamic parameters for allosamidin binding to ChiA, ChiB, and ChiC of <i>Serratia marcescens</i> at $t = 30$ °C, pH = 6.0 .....	77
<b>Table 4.1</b> Primers and DNA templates used for site directed mutagenesis .....	90
<b>Table 4.2</b> Primers used to introduce NdeI and XhoI restriction sites into ChiA and ChiB variants .....	91
<b>Table 4.3</b> Absolute and relative free energies of binding (GlcNAc) <sub>6</sub> to ChiA and ChiB wild-type and variants .....	100
<b>Table 5.1</b> Primers used for site-directed mutagenesis .....	126
<b>Table 5.2</b> Kinetic constants of ChiA-WT, ChiA-T276A, and ChiA-R172A for the hydrolysis of (GlcNAc) <sub>4</sub> in 20 mM Sodium Acetate, pH 6.1, and 0.1 mg/mL BSA at 37 °C .....	138
<b>Table 5.3</b> Sequence determination of CHOS of DP = 5 obtained after hydrolysis of chitosan ( $F_A = 0.65$ ) with ChiA-WT, ChiA-T267A, ChiA-R172A to a maximum degree of scission .....	140
<b>Table 5.4</b> Thermodynamic Parameters for (GlcNAc) <sub>6</sub> and allosamidin binding to ChiA-WT, ChiA-T276A, ChiA-R172A from <i>Serratia marcescens</i> at $t = 30$ °C, pH = 6.0 .....	147
<b>Table 6.1</b> Primers used for site directed mutagenesis .....	171

<b>Table 6.2</b> Relative binding free energies determined from thermodynamic integration TI alongside experimentally measured hydrolyzing activity and initial apparent processivity ( $P^{app}$ ) for wild-type (WT) ChiB and selected polar and aromatic variants across the ChiB cleft .....	182
<b>Table 7.1</b> Comparison of the crystal lattice parameters of $\beta$ -chitin between experiment and MD simulation .....	223

## List of Figures

<b>Figure 1.1</b> Chemical structures of (A) cellulose and (B) chitin .....	3
<b>Figure 1.2</b> Polymorphs of (A) cellulose and (B) chitin shown from both side view and top view .....	7
<b>Figure 1.3</b> GH domain constructions .....	8
<b>Figure 1.4</b> The three most common topologies of GH active sites (pocket, cleft, and tunnel), which may dictate mode of action and processive ability .....	9
<b>Figure 1.5</b> Two possible catalytic mechanisms GHs may follow to hydrolyze polysaccharide substrates .....	11
<b>Figure 1.6</b> The active site topology of processive fungal GHs with varying processivity .....	14
<b>Figure 1.7</b> Hypothesized free energy profile of the processive enzyme mechanism for deconstructing crystalline substrates .....	17
<b>Figure 1.8</b> Synergistic function of <i>S. marcescens</i> processive and non-processive chitinases along with accessory enzymes for deconstructing chitin .....	19
<b>Figure 1.9</b> Structural comparison of processive ChiA and ChiB and non-processive ChiC .....	22
<b>Figure 1.10</b> Substrate-assisted catalytic mechanism of Family 18 chitinases .....	24
<b>Figure 2.1</b> Schematic describing the steps involved in MD simulations .....	32
<b>Figure 2.2</b> Thermodynamic cycle for determining $\Delta G^{\circ}_b$ with FEP/ $\lambda$ -REMD .....	34
<b>Figure 2.3</b> FEP/ $\lambda$ -REMD protocol in the context of reversible work staging process to compute absolute binding free energy of the ligand ( $\Delta G^{\circ}_b$ ) .....	36

<b>Figure 2.4</b> A hybrid residue “Trp-Ala” used in dual-topology TI calculations to compute the relative change in binding free energy ( $\Delta \Delta G$ ) for tryptophan (Trp) to alanine (Ala) mutation .....	37
<b>Figure 2.5</b> Thermodynamic cycle for measuring $\Delta \Delta G$ using either ITC or TI. Free and bound refer to the absence and presence of the ligand, respectively .....	39
<b>Figure 2.6</b> TI convergence assessment example from the van der Waals calculations for phenylalanine to alanine mutation in the ligand bound state .....	41
<b>Figure 3.1</b> Schematic of (GlcNAc) <sub>6</sub> occupancy for each of the three chitinases considered in the simulations .....	55
<b>Figure 3.2</b> Binding isotherms with theoretical fits for the binding of (GlcNAc) <sub>6</sub> .....	61
<b>Figure 3.3</b> Average number of water molecules displaced upon binding (GlcNAc) <sub>6</sub> to ChiA, ChiB, and ChiC from MD simulations.....	65
<b>Figure 3.4</b> The RMSF of ChiA (A), ChiB (B) and ChiC (C) protein residue backbones over 100-ns MD simulations for apo and different ligand bound states .....	67
<b>Figure 3.5</b> Upper panel: Chemical structures of (GlcNAc) <sub>6</sub> (left) and allosamidin (right) aligned in their respective subsites. Lower panel: Schematic representation of $-T\Delta S_{\text{solv}}^{\circ}(\text{GlcNAc})_6$ (orange), $-T\Delta S_{\text{solv}}^{\circ}\text{allosamidin}$ (blue), and $-T\Delta S_{\text{solv}}^{\circ}\text{subsite } +1,+2,+3$ (cyan) for ChiA, ChiB, and ChiC .....	78
<b>Figure 4.1</b> Aromatic residues examined in this study near the catalytic centers of <i>S. marcescens</i> processive chitinases ChiA and ChiB .....	87
<b>Figure 4.2</b> Dynamics of the catalytic center and the bound ligand in variants of ChiA and ChiB .....	103

<b>Figure 4.3</b> (A) Crystal structure of ChiA-E315Q in complex with chito-octaose (1EHN; only six sugar moieties are shown), highlighting hydrogen bonding of Thr-276 and Glu-473 with the -3 site pyranose and of Lys-369 with the +2 pyranose .....	112
<b>Figure 4.4</b> Comparison of reported relative binding free energies, $\Delta \Delta G$ , for aromatic to alanine mutations in selected family 6, 7, and 18 GHs: <i>T. reesei</i> Cel6A and Cel7A, and <i>S. marcescens</i> ChiA and ChiB .....	117
<b>Figure 5.1</b> Crystal structures of the active site of ChiA in the presence of (GlcNAc) <sub>6</sub> (A, colored grey) and allosamidin (B, colored yellow) .....	137
<b>Figure 5.2</b> Comparison of initial [(GlcNAc) <sub>2</sub> /(GlcNAc)] ratios for ChiA wt (■), ChiA-R172A (●) and ChiA-T276A (▲) .....	142
<b>Figure 5.3</b> Progress curves for the formation of (GlcNAc) <sub>2</sub> (squares) and GlcNAc (circles) after hydrolysis of $\beta$ -chitin by ChiA WT (black), ChiA-R172A (red), and ChiA-T276A (green) .....	143
<b>Figure 5.4</b> Relationship of the [(GlcNAc) <sub>2</sub> /(GlcNAc)] ratio for ChiA wt (■), ChiA-R172A (●) and ChiA-T276A (▲) with extent of degradation .....	144
<b>Figure 5.5</b> Thermograms and binding isotherms with theoretical fits of allosamidin and (GlcNAc) <sub>6</sub> to chitinases .....	146
<b>Figure 5.6</b> Dynamics of the ChiA-WT, ChiA-R172A, and ChiA-T276A active sites as determined from 250-ns MD simulations .....	151
<b>Figure 5.7</b> Hydrogen bonds formed between ChiA residues and ligand subsites over 250-ns MD simulations .....	158

<b>Figure 6.1</b> The active site of Family 18 <i>S. marcescens</i> processive ChiB, populated with polar and aromatic residues interacting with a hexameric chito-oligosaccharide ligand via hydrogen bonding and hydrophobic stacking.....	169
<b>Figure 6.2</b> Active site dynamics and solvation of ChiB wild-type (WT) and selected polar variants from 250-ns MD simulations.....	185
<b>Figure 6.3</b> Snapshots of WT ChiB bound to a (GlcNAc) <sub>6</sub> ligand at (A) 0 ns and (B) 250 ns, illustrating how the ‘roof’ forming hydrogen bond between Trp-97 and Asp-316 is broken over the course of simulation.....	188
<b>Figure 6.4</b> (A) Simulation snapshots of ChiB wild-type (WT) and the Y145A variant at 250 ns illustrating how catalytic residue Glu-144 compensates for the loss of interaction between residue 145 and the +1 site C6 hydroxyl, following the Tyr-145 to alanine mutation .....	195
<b>Figure 6.5</b> (A) Snapshot of R294A variant MD simulation at 100 ns and 250 ns showing the frequent exchange of positions between OD1 and OD2 atom of Asp-215 while maintaining the hydrogen bond interaction with the -1 site .....	201
<b>Figure 7.1</b> Substrate modes/states in the ChiA processive cycle including: (A) presliding, (B) sliding, and (C) Michaelis complex .....	207
<b>Figure 7.2</b> The free energy profiles obtained from TMD umbrella sampling for the processive sliding and catalytic activation steps in <i>S. marcescens</i> ChiA .....	218
<b>Figure 7.3</b> Dynamical fluctuation and interaction energy contributions evaluated from 150-ns classical MD simulations of the three different states along the processive cycle of <i>S. marcescens</i> ChiA, including presliding, sliding, and the Michaelis complex .....	222

<b>Figure 7.4</b> The decrystallization work ( $W_{\text{decrystallization}}$ ) as a function of native contact ( $\zeta$ ) determined from umbrella sampling simulations of a $\beta$ -chitin microfibril .....	225
<b>Figure 7.5</b> Free energy profiles of the processive GH cycles of (A) <i>T. reesei</i> Cel7A, (B) <i>T. reesei</i> Cel6A, and (C) <i>S. marcescens</i> ChiA .....	233
<b>Figure 7.6</b> Quantitative comparison of ligand binding free energy ( $\Delta G_{\text{binding}}$ ) and work of decrystallization ( $W_{\text{decrystallization}}$ ) for a hexa-oligosaccharide substrate in processive ChiA, ChiB, and non-processive ChiC .....	235

## Chapter 1 - Introduction

### 1.1 Motivation

Increasing consumption of fossil based resources due to growing societal energy demand is driving the quest for alternative sustainable energy source. With the annual production of nearly 1.5 trillion tons of non-grain cellulosic biomass and the 100 billion tons of chitinous waste from marine fishing, biomass represents a large resource of renewable carbon for production of fuels and chemicals.<sup>1-2</sup> Efficient biochemical conversion of biomass can easily meet global energy needs and limit environmental pollution.<sup>3-5</sup> For example, ethanol from cellulosic and chitinous biomass can replace the traditional transportation fuels, such as gasoline and diesel ensuring energy security and reducing green house gas emission.<sup>6-8</sup> The biochemical conversion of cellulosic and chitinous biomass can also produce high value specialty chemicals that have immense applications in food, and packaging industries as well as have medicinal applications.<sup>1-2, 9-</sup>

12

A major challenge associated with efficient and economical biomass conversion arises as a result of the natural recalcitrance of the polysaccharides developed as a mode of protection against microbial and chemical attack.<sup>13</sup> Recalcitrance is the result of strong hydrophobic interactions and complex networks of inter- and intra- chain hydrogen bonds in highly crystalline polysaccharides, such as cellulose and chitin.<sup>14-15</sup> In nature, microorganisms such as bacteria and fungi secrete suites of enzymatic machinery to degrade crystalline polysaccharides (cellulose and chitin) to soluble sugars. These enzyme cocktails are primarily composed of various glycoside hydrolases (GH) with



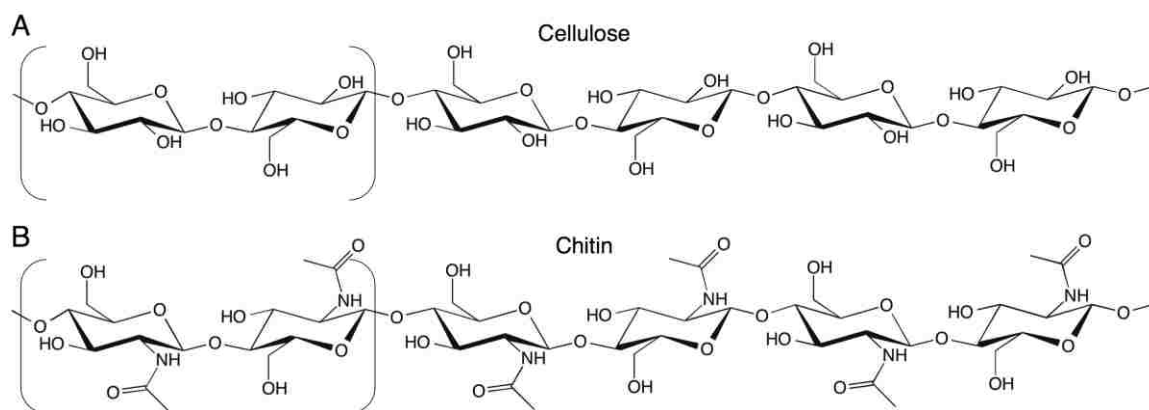
synergistic functions to efficiently cleave the glycosidic linkages between the carbohydrate monomers.<sup>16</sup> However, though these naturally occurring GHs are effective on timescales commensurate with microbial needs, they are not yet sufficiently active for economical production of fuels and chemicals, with timescales for biomass conversion typically on the order of days at ambient conditions. Additionally, the high cost of GHs remains a major factor in the economical conversion of biomass.<sup>17-18</sup> Experimental methods to improve performance, such as directed evolution,<sup>19-21</sup> are limited for cellulases and chitinases with the lack of high throughput crystalline substrate activity assays. Improvements are most likely to arise from rational design efforts requiring mechanistic and molecular level insights into enzyme function.<sup>22-23</sup>

The primary focus of this dissertation is to investigate the origin of recalcitrance of crystalline polysaccharides and how GHs interact with polysaccharides to overcome the natural recalcitrance. Molecular simulation and thermodynamic calculations backed by experimental biochemical characterization has been used to investigate the relationships of protein ligand dynamics, protein substrate interaction to kinetic parameters of GHs,<sup>24</sup> which we anticipate will ultimately form the foundation of a theoretical structure-function relationship for engineering activity improvements in GHs, including cellulases and chitinases.

## 1.2 Research background

### 1.2.1 Crystalline polysaccharides

Cellulose, a  $\beta$ -1, 4-linked homo polymer of glucose (Glc), and chitin, a  $\beta$ -1, 4-linked homo polymer of *N*-acetyl glucosamine (GlcNAc), are the two most abundant renewable carbon resources in the biosphere.<sup>25</sup> Cellulose is the major structural component of plant cell wall, whereas chitin is the key component of fungal cell wall and the exoskeleton of arthropods.<sup>16, 25-26</sup> The monomeric units of cellulose and chitin are dimers of Glc (cellobiose) and GlcNAc (chitobiose), respectively, oriented 180° from each other (Figure 1.1); the polysaccharide chains are made up of tens of thousands of carbohydrate monomers connected via glycosidic bonds, forming insoluble polymer crystals or microfibrils. The array of chains within the microfibril is held together by hydrophobic interactions and a complex network of inter- and intra-chain hydrogen bonds defining the polymorphs (Figure 1.2).



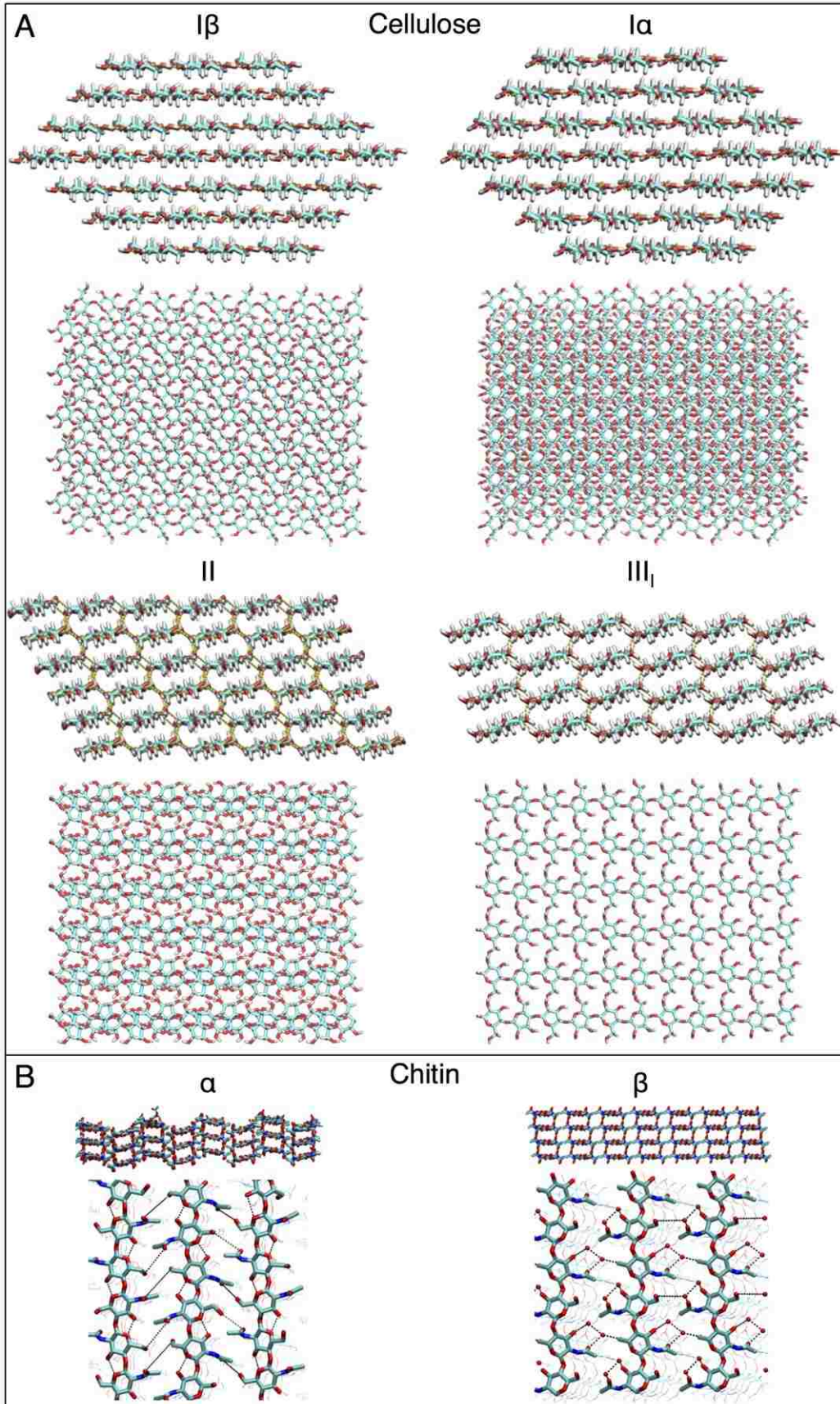
**Figure 1.1** Chemical structures of (A) cellulose and (B) chitin. (A) Cellulose is a linear polymer of  $\beta$ -1, 4-linked glucose (Glc), and (B) chitin is a linear polymer of  $\beta$ -1, 4-linked

*N*-acetyl glucosamine (GlcNAc). The monomeric units of cellulose and chitin are cellobiose and chitobiose, respectively, as denoted by brackets in the figure.

In nature, cellulose is found in two crystalline forms: cellulose I $\alpha$  and I $\beta$ .<sup>16</sup> Cellulose I $\alpha$  is primarily found in bacteria and algae cell walls,<sup>27</sup> whereas cellulose I $\beta$  is available in higher-level plants.<sup>28</sup> Crystal structures of cellulose show that the molecular level packing and arrangements are very different to each other.<sup>29-30</sup> Despite the fact that both naturally occurring cellulose polymorphs are arranged in parallel fashion and interact only through the intra- layer hydrogen bonding, there are primary differences resulting from interlayer chain stacking (Figure 1.2A). Chemical pretreatments, such as treatment with ionic liquids or ammonia, lead to further modification of the cellulose I structures to cellulose II and cellulose III<sub>I</sub>.<sup>31-32</sup> The chains of the cellulose II polymorph are arranged in antiparallel fashion, and in cellulose III<sub>I</sub>, in a parallel fashion, maintaining both inter- and intra- layer hydrogen bonds (Figure 1.2A). The varying molecular arrangement makes these naturally occurring and pretreated celluloses have different recalcitrance. Prior biochemical studies revealed that the enzymes responsible for cellulose degradation are more active on pretreated celluloses than the natural cellulose polymorphs.<sup>33-34</sup> Computational investigation further quantitatively described the energy required to decrystallize a cellobiose from the surface of natural cellulose polymorphs is comparatively higher than for chemically treated celluloses.<sup>14</sup>

Like cellulose, chitin also naturally exists in two crystalline forms:  $\alpha$ -chitin, and  $\beta$ -chitin (Figure 1.2B).<sup>35</sup>  $\alpha$ -chitin is by far the most prevalent form, found in fungal and yeast cell walls, insect cuticles, and exoskeletons of krill, crabs, and lobsters.<sup>26</sup>  $\beta$ -chitin is comparatively rare, typically found in squid pens, diatoms and tubeworms.<sup>26</sup> The crystal

structures from X-ray diffraction show that the polysaccharide chains in the  $\alpha$ -chitin microfibril are arranged in antiparallel parallel fashion like cellulose II.<sup>36-37</sup> The antiparallel orientation in fact, facilitates increased number of intra- and inter- layer hydrogen bonds leading to tight packing of the  $\alpha$ -chitin crystal.<sup>15</sup> The tight packing also eliminates water molecules from the crystal, making  $\alpha$ -chitin anhydrous. On the contrary, the polysaccharide chains in  $\beta$ -chitin polymorph are arranged in parallel orientation and observed mostly in hydrated forms.<sup>38-41</sup> The enzymes responsible for chitin degradation also show varying activity toward  $\alpha$ - and  $\beta$ -chitin like cellulose.<sup>42</sup> Due to the tight packing of the crystal,  $\alpha$ -chitin is comparatively more resistant to biochemical conversions than  $\beta$ -chitin.



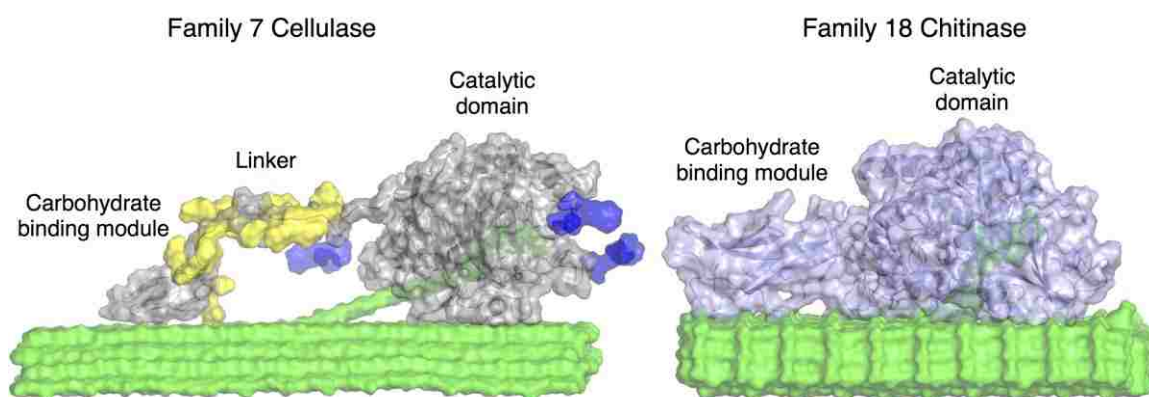
**Figure 1.2** Polymorphs of (A) cellulose and (B) chitin shown from both side view and top view. The major difference between cellulose I $\alpha$  and I $\beta$  arises due to the difference in chain stacking, as can be seen from the top views of Figure A. Similarly, the parallel and antiparallel orientation of cellulose II and III<sub>I</sub> along with different hydrogen bonding and stacking can be identified from both top and side views of Figure A. The yellow lines in cellulose II and III<sub>I</sub> top view structures denote inter- layer hydrogen bonding. Figure (B) top view shows the antiparallel and parallel orientation of  $\alpha$ - and  $\beta$ -chitin. Additionally, the side view of Figure B shows clear difference in intra- layer hydrogen bonding between  $\alpha$ - and  $\beta$ -chitin. The hydrogen bonding in chitin has been shown by black dashed lines. The  $\beta$ -chitin structure is shown in hydrated form. The structures of different polymorphs of cellulose have been taken from Payne et al.<sup>16</sup> This is an adaptation, with permission from Payne et al., of that appeared in an ACS publication.<sup>16</sup>

### 1.2.2 Glycoside hydrolases (GH)

GHs are the naturally occurring enzymes that hydrolyze the glycosidic bonds between the carbohydrates to produce soluble sugars. GHs consist of at least one catalytic domain and can display one or more non-catalytic carbohydrate-binding modules connected by a linker peptide (Figure 1.3).<sup>43</sup> The linker peptide connecting the carbohydrate-binding module and the catalytic domain can be flexible and distinct, or the carbohydrate-binding module can be fused with the catalytic domain without a discrete demarcation (Figure 1.3).<sup>44-45</sup> The carbohydrate-binding module facilitates substrate recognition and binding, positioning the catalytic domain on surface of the crystal.<sup>46-47</sup> The catalytic domain, on the other hand, performs hydrolysis of the acquired polysaccharide chains. Recent computational studies also revealed that the linkers, which

are glycosylated, increase the affinity of the GHs for the crystalline polysaccharides.<sup>48</sup>

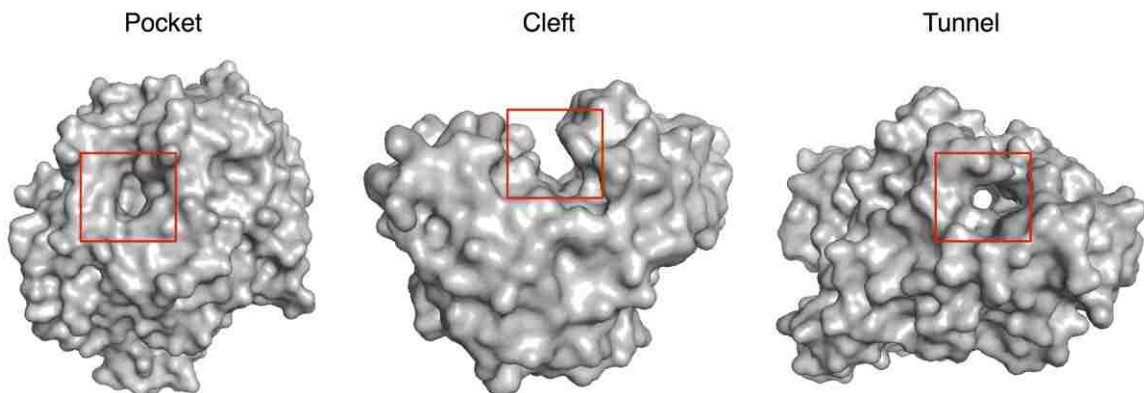
To date, there are 135 GH families identified on the basis on their amino acid sequence and fold similarity ([www.cazy.org](http://www.cazy.org)).<sup>49-52</sup> Classification in this manner provides insights into the structural features, catalytic mechanism, mode of action, and processive behavior. For example, GH Family 7 corresponds to a  $\beta$ -jelly roll fold of the catalytic domain with two-step retaining catalytic mechanism and reducing end specificity. Sequence based classification of the auxiliary carbohydrate-binding modules into CBM families further aids in understanding the binding specificity as well as identifying functional residues in the carbohydrate-binding module ([www.cazy.org](http://www.cazy.org)).<sup>46</sup>



**Figure 1.3** GH domain constructions. The catalytic domain and the carbohydrate binding module can be joined via a flexible linker, as observed in GH Family 7 *Trichoderma reesei* cellulase (left panel), or the catalytic domain and the carbohydrate binding module can be fused, as in GH Family 18 *Serratia marcescens* chitinase A (right panel). The enzymes are shown in gray (left) and ice blue (right) surface representations. The crystalline substrate is shown in green. The GH can also have post-translational modifications such as *N*-linked and *O*-linked glycosylation, shown in blue and yellow, respectively, at different locations on the enzyme.

### 1.2.2.1 GH active site topology

Based on sequence and comparison to available crystal structures from the protein data bank, GH active sites correspond to one of three common topologies, including pockets, clefts, and tunnels (Figure 1.4).<sup>51</sup> The topology of the active site reflects how the subsites participate in substrate binding and dictates substrate specificity. For example, the pocket like topology is typically exhibited by GHs that prefer to attack oligomeric chain ends and are not particularly active in degrading crystalline polysaccharides that have very few exposed chain ends. To degrade crystalline substrates with fewer available chain ends, GHs evolved tunnel-like topologies, wherein a single chain from the crystal is threaded into the active site and hydrolyzed repetitively. The cleft topology, being more open and solvent exposed, facilitates random binding and hydrolysis of the substrate and is mostly present in endo-acting GHs (discussed below).



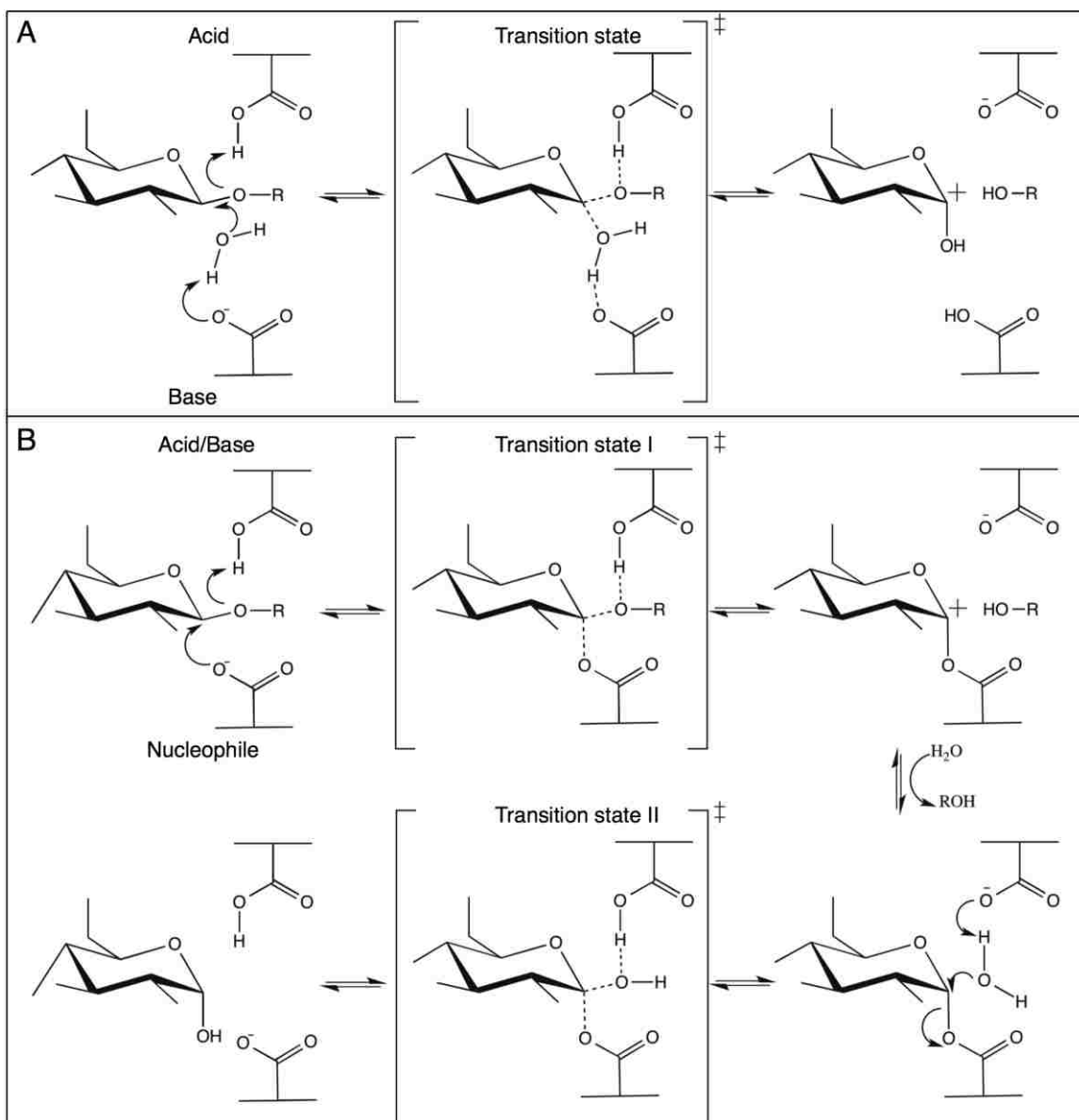
**Figure 1.4** The three most common topologies of GH active sites (pocket, cleft, and tunnel), which may dictate mode of action and processive ability. GHs are shown in gray, whereas different active site topologies are highlighted with red square box. The figure is constructed using Protein databank (PDB) structures 3GLY (pocket),<sup>53</sup> 2BOD (cleft),<sup>54</sup> and 4C4C (tunnel).<sup>44</sup>



### 1.2.2.2 Catalytic mechanism of GHs

GHs hydrolyze the glycosidic bonds of polysaccharides via either a single-step inverting mechanism or a two-step retaining mechanism.<sup>51, 55-56</sup> In the inverting mechanism, the reaction occurs via a single displacement mechanism over a single transition state (Figure 1.5A), where a catalytic acid donates a proton to the glycosidic oxygen and a nearby water molecule is simultaneously deprotonated by a catalytic base. This is followed by nucleophilic attack by the water molecule on the anomeric carbon, completing the hydrolysis. The cellulose-degrading enzymes from GH Family 6 follow the one-step inverting mechanism to hydrolyze cellulosic substrates.<sup>16, 57</sup>

Conversely, the retaining mechanism occurs through a double displacement mechanism via two transition states (Figure 1.5B). In the first step of hydrolysis (glycosylation), a proton is transferred from the catalytic acid to the glycosidic oxygen, and the catalytic acid becomes a base; the catalytic base acts as a nucleophile and attacks the anomeric carbon of the pyranose ring, forming a glycosyl-enzyme intermediate. In the second step (deglycosylation), the deprotonated base acquires a proton from the nearby water molecule and returns to the catalytic acid state. This is followed by nucleophilic attack by the water molecule on the anomeric carbon, completing the reaction. Family 7 cellulases follow a two-step retaining mechanism to hydrolyze cellulosic substrates.<sup>16, 58</sup> Family 18 chitinases also follow two step retaining mechanism to hydrolyze chitinous substrate; however, the *N*-acetyl oxygen of the pyranose itself acts as a nucleophile instead of the catalytic base during the first step of hydrolysis (discussed later).



**Figure 1.5** Two possible catalytic mechanisms GHs may follow to hydrolyze polysaccharide substrates. Figure (A) shows the inverting mechanism over a single transition state, wherein a water molecule acts as the nucleophile during the reaction attacking the anomeric carbon of the pyranose. Figure (B) represents two-step retaining mechanism over two transition states; in the first step, a catalytic base acts as a nucleophile attacking the anomeric carbon of the pyranose, and in the second step, a water molecule acts as a nucleophile. In both of the mechanisms, (A) and (B), a catalytic

acid donates a proton to the glycosidic oxygen to initiate the hydrolysis. The figure is modified from Momeni et al.<sup>59</sup>

### **1.2.2.3 GH mode of action**

GHs can initiate substrate attack via two modes, ‘exo-’ or ‘endo-’ mode.<sup>16</sup> In exo-mode, GHs generally attack the polysaccharide substrate from a free chain end (reducing or non-reducing end). Endo-mode GHs randomly attack an arbitrary location along the polysaccharide surface. Exo-mode initiation is thought to occur at the entrance of the enzyme active site tunnel, where a polysaccharide chain is threaded from the entrance site through the tunnel to initiate hydrolysis. Conversely, endo-initiation mode occurs when the active-site tunnel loops open sufficiently to directly complex with chains along active site tunnels. Endo-initiation mode has been shown to occur more frequently in more open active-site tunnels and clefts.<sup>42</sup> These two modes of complexation are likely the extremes of feasible initiation mechanisms with a distribution of intermediate mechanisms in between; though, a definitive description of initiation mode relative to GH structure remains elusive.

### **1.2.2.4 GH processivity**

Processivity is the ability of a GH to remain associated with the polysaccharide substrate between subsequent hydrolytic reactions and can be quantified by how many times the enzyme performs catalysis before finally dissociating from the substrate.<sup>23</sup> GHs can be non-processive or processive. Non-processive GHs randomly attack the amorphous regions of polymer crystals and cleave glycosidic linkages once or only a few times before dissociating, creating accessible polymeric chain ends. Processive GHs

attach to available polymer chain ends in exo-initiated attack or generate a new chain ends via endo-initiated attack, sequentially hydrolyzing many glycosidic linkages to produce multiple disaccharide products.<sup>16</sup>

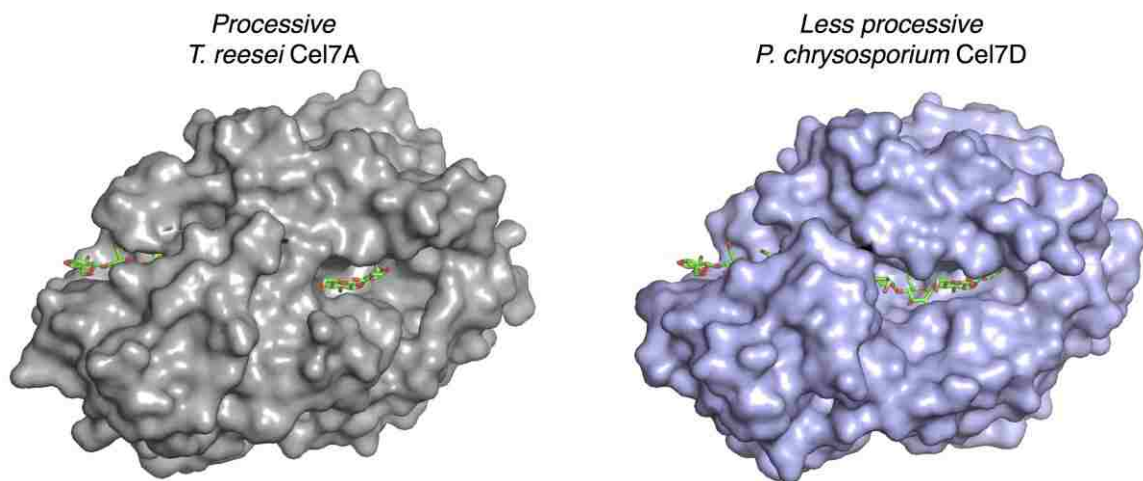
It is thought that processive GHs are responsible for the majority of hydrolytic bond cleavages and, hence, are the most logical targets for activity improvements towards efficient and economical biomass conversion. Unfortunately, the mechanism and factors responsible for processivity are still not understood completely at the molecular-level. It is not clear how processive GHs dynamically bind to the ligand and how this binding influences processivity on crystalline and amorphous polysaccharides. This dissertation's primary objective is to establish the foundations of a general, molecular level theory of processivity in GHs based on enzyme active site topology, chemical composition, dynamics, and binding characteristics.

#### **1.2.2.4.1 Contribution of GH active site in processivity**

It is generally thought that if a GH possesses a long catalytic tunnel or deep active site cleft, it allows more amino acids to interact with the substrate leading to lower probability of dissociation of the GH from the substrate.<sup>51, 60-61</sup> For example, an efficient cellulose degrader from GH Family 7, *Trichoderma reesei* Cel7A, is a processive GH having a 50 Å long catalytic tunnel surrounded by eight flexible loops, important for substrate recruitment and processive movement (Figure 1.6).<sup>16, 62</sup> However, another cellulose degrader from GH Family 7, *Phanerochaete chrysosporium* Cel7D, has a relatively open catalytic tunnel and is comparatively less processive than *T. reesei* Cel7A (Figure 1.6).<sup>63-64</sup> Similarly, one of the best studied enzyme machineries for chitin degradation is GH Family 18 *Serratia marcescens* chitinases, which includes processive

ChiA, processive ChiB, and non-processive ChiC. ChiA and ChiB possess deep tunnel-like active site clefts, and ChiC possesses a more open active site cleft.<sup>65</sup> The generalization regarding processivity only based on active site geometry is always not accurate. For example, a Family 9 cellulase, Cel9A from *Thermobifida fusca*, though having an open shallow active site, can act on polysaccharide substrates processively.<sup>66-67</sup>

It has also been shown that deletion of a single active site loop or deletion of a single amino acid from the GH active site can make processive enzymes less processive.<sup>68-71</sup> For example, point mutation of a tryptophan residue (Trp-38) at the entrance of *T. reesei* Cel7A active site tunnel greatly reduces processive ability,<sup>72</sup> suggesting both the GH active site architecture and the active site chemical composition are important in GH processivity. A recent molecular simulation study further revealed that flexibility of the GH active site, including the flexibility of the ligand and the degree of ligand solvation, greatly contribute to processive function of GHs.<sup>73</sup>



**Figure 1.6** The active site topology of processive fungal GHs with varying processivity. The catalytic domains of the GHs are shown in gray and blue. The ligand is shown in green and red stick. The structural comparison shows both GHs possess a tunnel shaped

active site; though, *P. chrysosporium* Cel7D is comparatively more solvent exposed. The figure was made in PyMOL using crystal structures of *T. reesei* Cel7A (Protein databank (PDB) structure 4C4C) and *P. chrysosporium* Cel7D (PDB structure 1GPI).

#### 1.2.2.4.2 Measurement of processivity

Processivity of GHs can be measured either based on analysis of the product profile or from determination of the catalytic rate coefficient and dissociation rate coefficient.<sup>23, 64, 74</sup> The determination of processivity from product profile is defined as apparent processivity ( $P^{app}$ ), which greatly depends on the nature of the substrate and its heterogeneity. It has been previously shown that the disaccharides to monosaccharides ratio can be used as an estimate of an enzyme's apparent processive ability, since processive GHs produce more disaccharides than non-processive GHs. However, the initial binding mode and the corresponding initial product profile often leads to overestimation of processivity values.<sup>23</sup> Another widely used method to evaluate  $P^{app}$  is simultaneous determination of the ratio of soluble to insoluble reducing ends, assuming processive GHs will produce significantly higher quantities of soluble reducing ends compared to non-processive GHs.<sup>23</sup> However, this approach is also not free from inaccurate  $P^{app}$  estimation, as non-processive GHs also can generate higher amount of soluble reducing ends if the substrate matrix contains an abundance of free chain ends. Recently, new techniques based on substrate labeling, such as fluorescence based labeling and radioactive carbon based labeling ( $^{14}C$ ), have been developed to accurately estimate processivity. Unfortunately, these methods are only applicable to reducing end-specific GHs given the nature of carbohydrate chemistry.

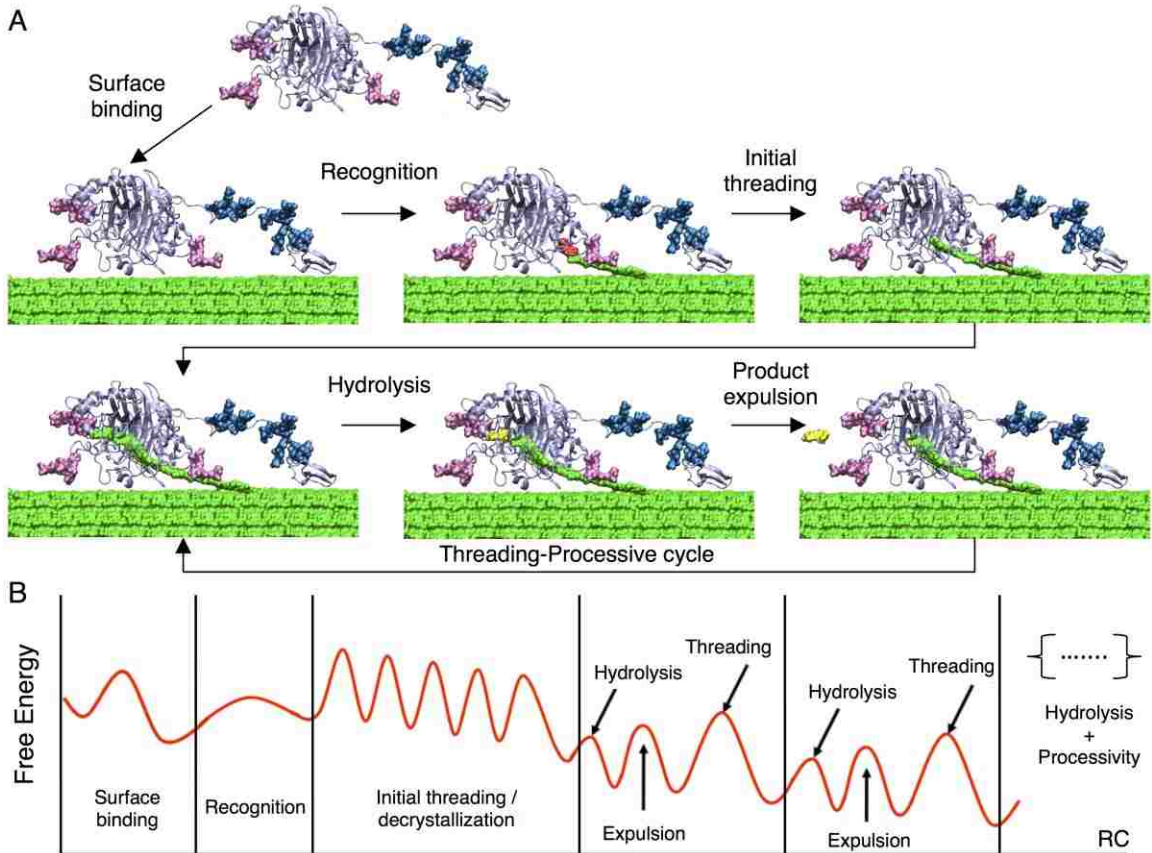
Unlike  $P^{app}$ , which is dependent on the nature of the substrate, intrinsic

processivity,  $P^{\text{Intr}}$ , is defined as the ratio of catalytic rate coefficient ( $k_{\text{cat}}$ ) and the dissociation rate coefficient ( $k_{\text{off}}$ ), as described in Equation 1.1.<sup>74</sup> The Equation assumes that GH has a very low probability of dissociation from the substrate ( $k_{\text{off}}$ ) compared to the catalytic turnover ( $k_{\text{cat}}$ ). Hence,  $P^{\text{Intr}}$  is the theoretical maximum number processive steps a GH may realize on an ideal substrate free from heterogeneity and obstacles.

$$P^{\text{Intr}} \approx \frac{k_{\text{cat}}}{k_{\text{off}}} \because k_{\text{cat}} \gg k_{\text{off}} \quad (1.1)$$

#### 1.2.2.4.3 Thermodynamics of GH processivity

Processive GH turnover is critical for efficient biomass deconstruction and comprises of a number of steps, including surface binding, recognition, decrystallization, threading, hydrolysis, product expulsion, processive sliding and dissociation.<sup>33</sup> Many of these steps, however, are poorly understood due to experimental difficulties in identifying them independently and as a result of the heterogeneity of the substrate. Beckham et al. illustrated this hypothesized processive mechanism in terms of a free energy landscape.<sup>22</sup> They hypothesized that, due to noninvolvement of any cofactors, the entire processive mechanism must be energetically downhill overall for the thermodynamic favorability as shown in Figure 1.7.



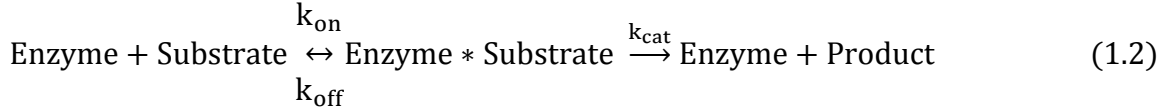
**Figure 1.7** Hypothesized free energy profile of the processive enzyme mechanism for deconstructing crystalline substrates. The figure is taken with permission from Beckham et al.,<sup>22</sup> Copyright (2011) Elsevier.

Recently, Payne et al. developed a complementary mathematical model to describe intrinsic processivity ( $P^{\text{Intr}}$ ) in terms of ligand binding free energy ( $\Delta G_b^\circ$ ).<sup>61</sup> Relating  $\Delta G_b^\circ$  to  $P^{\text{Intr}}$  is advantageous because of its relative simplicity, particularly in quantitatively relating GH structure to function. The detail of the mathematical model proposed by Payne et al. is described below.

When protein substrate complexation reaches equilibrium (Equation 1.2), the dissociation constant  $K_D$ , which is ratio of dissociation ( $k_{\text{off}}$ ) to association ( $k_{\text{on}}$ ) rate



coefficients, can be expressed in terms of  $\Delta G^\circ_b$ , as shown in Equations 1.2 and 1.3.



$$\Delta G^\circ_b = RT \ln K_D \quad \text{where } K_D = \frac{[\text{Enzyme}][\text{Substrate}]}{[\text{Enzyme} * \text{Substrate}]} = \frac{k_{\text{off}}}{k_{\text{on}}} \quad (1.3)$$

R is the ideal gas constant and T is the absolute temperature.

Finally, Equations 1.1 and 1.3 can be combined to establish the fundamental relationship of processivity with ligand binding free energy, as shown in Equation 1.4.

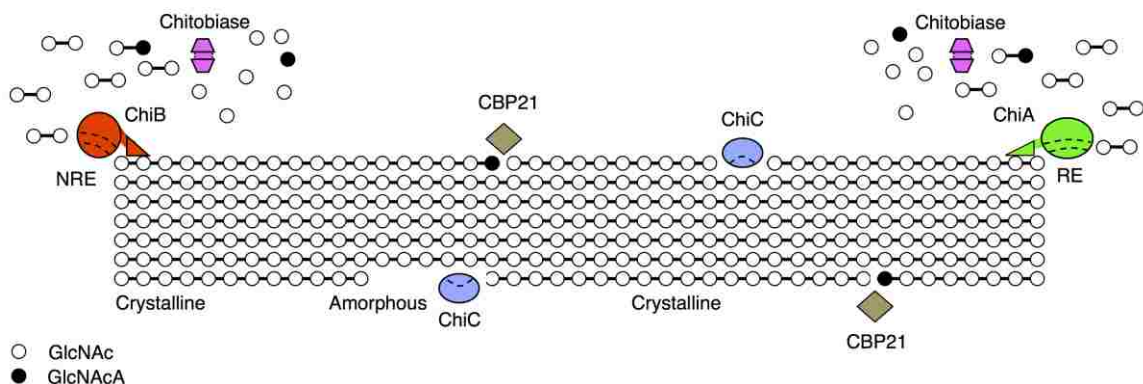
$$-\frac{\Delta G^\circ_b}{RT} = \ln\left(\frac{P^{\text{Intr}}k_{\text{on}}}{k_{\text{cat}}}\right) \quad (1.4)$$

The change in free energy upon ligand binding,  $\Delta G^\circ_b$ , can also be broken down into enthalpic,  $\Delta H$ , and entropic,  $\Delta S$ , contributions. The enthalpic term represents weak enzyme-substrate interactions, and the entropic term includes contributions from loss of translational and conformational freedom as well as changes in solvation. Enthalpic and entropic contributions determine the thermodynamic stability of the complex and arise from changes in the local environment of the active site or the ligand itself. Thus, it is feasible that a relationship between the ligand binding free energy, encompassing active site geometry, dynamics, and chemical composition, and degree of processivity exists.

### 1.2.3 *Serratia marcescens* chitinases as a model system to study processivity

An ideal model system for studying processivity in GHs is the co-evolved chitinolytic system from a gram-negative soil bacterium *Serratia marcescens*.<sup>65</sup> The enzymes from *S. marcescens* include three GH Family 18 chitinases - processive ChiA,

processive ChiB, and non-processive ChiC, as well as two accessory enzymes, a GH Family 20 chitinase and a lytic polysaccharide monooxygenase (CBP21) from Auxiliary Activity Family 10.<sup>75</sup> These enzymes act synergistically to break down crystalline chitin to soluble sugars, as shown in Figure 1.8.



**Figure 1.8** Synergistic function of *S. marcescens* processive and non-processive chitinases along with accessory enzymes for deconstructing chitin. The chitin substrate is shown as polymer chains of GlcNAc (open circle). Non-processive ChiC hydrolyzes glycosidic linkages to create accessible chain ends. Processive ChiA and ChiB attach to chain ends either from reducing (denoted as RE) or non-reducing end (denoted as NRE) and sequentially hydrolyze glycosidic linkages to produce disaccharides. Chitobiase converts disaccharides to monosaccharides. CBP21, like ChiC, attacks a random point of a polymer chain and cleaves the glycosidic bond through copper-mediated catalysis to form chain ends as well as producing aldonic acid (GlcNAcA; filled circles).

There are a number of advantages in studying processivity in *S. marcescens* chitinases over fungal cellulases including: (i) a completed structural suite for direct comparison of co-evolved features and characteristics; (ii) a common catalytic mechanism eliminating mechanistic differences as a variable in processivity; (iii) the ease

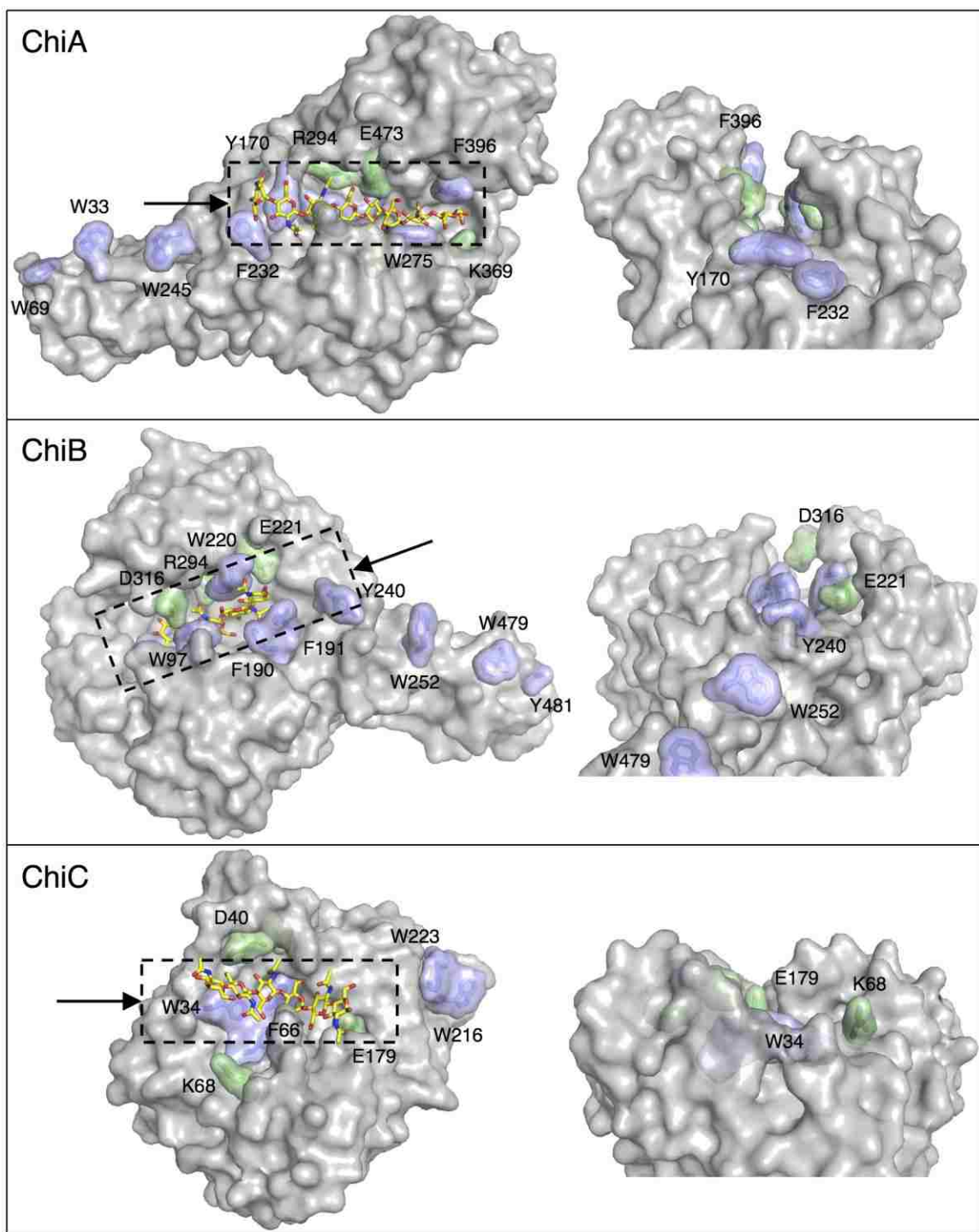
and uniformity of experimental processivity measurements relative to fungal cellulases; and (iv) the existing and developing wealth of experimental data for *S. marcescens* Family 18 chitinases. The insights gained through investigation of chitinases can also be extrapolated to cellulases given the similarities of these GHs in both enzymatic machineries and active site topology and chemical composition.<sup>76</sup> For example, studies of this chitinolytic system recently led to the discovery of cellulolytic counterparts and have been key in developing understanding of processivity and directionality in all GHs.<sup>77</sup>

#### **1.2.3.1 *S. marcescens* Family 18 chitinases**

As mentioned above, Family 18 chitinases from *S. marcescens* include three types – processive ChiA, processive ChiB, and non-processive ChiC.<sup>78</sup> The solved structures from X-ray crystallography show that the catalytic domain of all of the three chitinases have a  $(\beta/\alpha)_8$  TIM barrel fold with key catalytic residues positioned along the fourth  $\beta$  strand.<sup>45, 73, 79</sup> Additionally, ChiA possesses an *N*-terminal carbohydrate-binding module (CBM) with a fibronectin type III (FnIII) fold, ChiB possesses a C-terminal Family 5 CBM, and ChiC possess a C-terminal FnIII-type CBM coupled to a downstream Family 5 and Family 12 CBM.<sup>65, 80</sup> ChiC can also exist without the binding module as a catalytic domain, referred as ChiC2, and can hydrolyze chitin as such.<sup>80</sup>

The structural comparison between ChiA, ChiB, and ChiC shows ChiA and ChiB possess deep tunnel-like active site clefts populated with aromatic and polar residues; whereas, the cleft of ChiC is shallow and more solvent exposed with fewer aromatic and polar residues lining the active site (Figure 1.9), suggesting ChiC is an endo-acting, non-processive GH and ChiA and ChiB are exo-acting, processive GHs.<sup>42, 81</sup> Structural comparison further shows that the CBMs of ChiA and ChiB are positioned at opposite

directions (*N*-terminal for ChiA, and C-terminal from ChiB), suggesting these GHs hydrolyze the substrate from opposite directions; this has been verified by analysis of the reaction products as well as from a recent high-speed atomic force microscopy (HS-AFM), showing ChiA and ChiB are exo-acting processive enzymes with ChiA being reducing-end specific and ChiB non-reducing-end specific.<sup>82-83</sup> However, it has also been shown that the mode of action of ChiA and ChiB changes from exo- to endo- while acting on soluble substrate.<sup>84</sup>



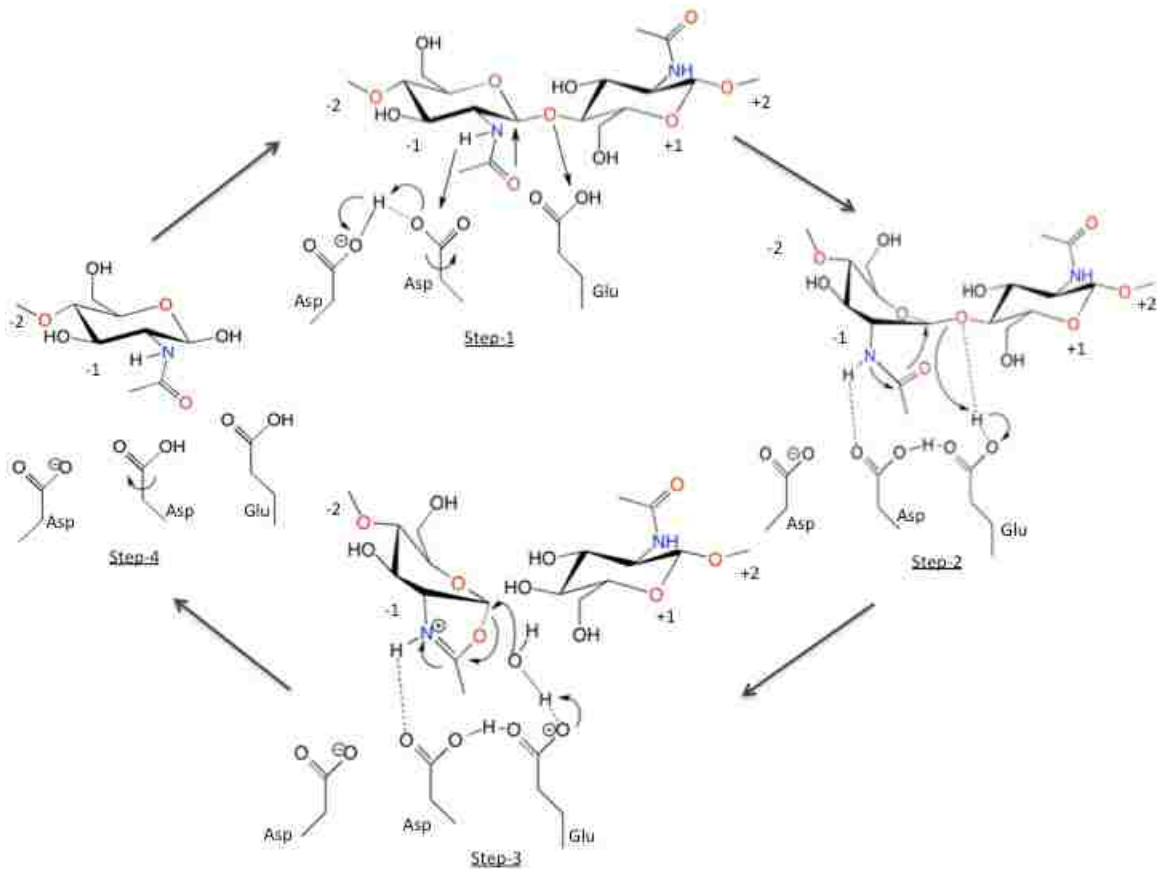
**Figure 1.9** Structural comparison of processive ChiA and ChiB and non-processive ChiC. The figures on the left show the full-length enzymes of ChiA and ChiB; whereas, ChiC is shown only with the catalytic domain. The aromatic and polar residues lining the

active site are shown in blue and green, respectively and the ligand is shown in yellow and red stick. At the right of the figure, the active site architectures of ChiA, ChiB (tunnel like deep clefts), and ChiC (open cleft) are also shown. The figure is made in PyMOL using PDBs 1EHN (ChiA), 1E6N (ChiB), and 4AXN (ChiC).<sup>45, 79, 85-86</sup>

### 1.2.3.2 Catalytic mechanism of Family 18 chitinases

Previous studies suggest that all Family 18 chitinases, irrespective of processivity, hydrolyze chitin through a common substrate-assisted catalytic mechanism that retains the stereochemistry of the anomeric carbon (Figure 1.10).<sup>79, 87-89</sup> These chitinases initiate catalysis via distortion of the chitin substrate in the -1 subsite adjacent to the glycosidic bond. Substrate binding is accompanied by rotation of an aspartic acid, which forms a hydrogen bond with the catalytic glutamic acid and the *N*-acetyl group of the -1 bound sugar. Nucleophilic attack by the *N*-acetyl oxygen on the anomeric carbon leads to scission of the glycosidic bond and generates an oxazolinium ion intermediate, which is subsequently hydrolyzed to complete the reaction. The unique substrate-assisted catalytic mechanism in Family 18 chitinases eliminates mechanistic differences from definition of a general molecular-level theory of processivity and enables observation of processivity more directly and reliably than possible in cellulases.<sup>74</sup> This is possible due to the required presence of this *N*-acetyl group in the -1 subsite for productive binding as described above. The absence of an acetylated sugar in the -1 subsite, as is the case with processive activity on soluble chitosan (partially de-acetylated chitin), leads to high number of even-numbered oligosaccharides during the initial phases of the reaction, as the processive enzymes continue to productively bind the chitosan rather than dissociating. On the contrary, non-processive enzymes will randomly hydrolyze the

substrate producing equal distribution of even and odd-numbered oligosaccharide products. Thus, analysis of the degradation product profiles can accurately delineate processivity in Family 18 chitinases on a variety of substrate.



**Figure 1.10** Substrate-assisted catalytic mechanism of Family 18 chitinases. All Family 18 chitinases have a highly conserved DXDXE motif consisting of two aspartic acids and one glutamic acid, which facilitates hydrolysis between -1 and +1 subsites. Reproduced with permission from Aronson et al., copyright © 2006 Taylor & Francis.<sup>90</sup>

### **1.3 Outline of dissertation**

We hypothesize that ligand binding free energy is correlated to degree of processivity and is a function of active site topology, dynamics, and chemical composition. To test our hypothesis, we have chosen *Serratia marcescens* Family 18 processive chitinases ChiA and ChiB and non-processive ChiC as a model system given the existing wealth of experimental characterization data. We completed the following three objectives to test our hypothesis.

#### **1.3.1 Overall active site contributions to binding and processivity (Chapter 3)**

In the first part of my dissertation in Chapter 3, I will focus on defining the overall active site contributions in substrate binding by computing the absolute ligand binding free energy. This investigation will initially include three wild-type chitinases – processive ChiA and ChiB and non-processive ChiC. The idea behind this approach is to establish ligand binding free energy as a tool to differentiate between processive and non-processive enzymes. We anticipate that processive enzymes, having long tunnels or deep active site clefts, allow more amino acids to interact with the ligand and exhibit strong ligand binding; whereas, non-processive enzymes, having more open tunnels or clefts, will exhibit comparatively weak binding. In this investigation, we employed molecular dynamics simulation coupled with the recently developed free energy perturbation with Hamiltonian replica exchange molecular dynamics to obtain accurate free energies with improved convergence. These computationally obtained values were further validated with experimental isothermal titration calorimetry (ITC) measurements.



### **1.3.2 Contributions of aromatic and polar residues to ligand binding and processivity (Chapters 4, 5 and 6)**

A notable feature in processive GHs is the ubiquity of aromatic and polar residues lining the enzyme tunnels and clefts. It has been thought that these residues are mainly responsible for substrate chain acquisition, binding, and hydrolysis in the catalytic tunnel/cleft via carbohydrate- $\pi$  stacking, hydrogen bonding and salt bridge interactions. These non-bonded interactions are usually assumed to facilitate the processive mechanism whereby the enzyme must maintain attachment to the substrate while still allowing forward processive motion. Chapters 4, 5 and 6 focus on exploring the extent to which these aromatic and polar residues present in the active site contribute to ligand binding and processivity. We used molecular dynamics simulation coupled with thermodynamic integration calculations to explore the relative change in ligand binding free energy resulting from mutation of targeted aromatic and polar residues lining the active site of the two processive *S. marcescens* Family 18 chitinases, ChiA and ChiB. The targeted residues have been selected based on calculated interactions with the bound ligand from molecular dynamics simulations of wild-type enzyme as well as reported experimental studies.

### **1.3.3 Rate-limiting step in GH processivity (Chapter 7)**

In addition to determining the contributions of active site topology and chemical composition in ligand binding, we also seek to identify the rate-limiting step inside the processive cycle of a GH, since the processive cycle is responsible for the majority of disaccharide production in biomass degradation. The proposed processive cycle involves three major steps including hydrolysis, product expulsion, and rethreading of the

substrate into the active site, which are very difficult to characterize experimentally. In Chapter 7, I used molecular dynamics simulation and umbrella sampling free energy calculations to estimate the free energy barriers associated with these elementary steps of the *S. marcescens* ChiA processive cycle. Finally, determining the slowest step inside the processive cycle of *S. marcescens* ChiA and making comparisons to cellulases will enable us to better understand the bottleneck of GH processivity, essential for engineering more efficient biomass converting enzymes.

## Chapter 2 - Computational Methodology

### 2.1 Introduction

To provide insights into how chitinases interact with chito-oligosaccharide substrates and with crystalline chitin, we have performed classical molecular dynamics simulation as well as free energy calculations, which provide information regarding the dynamical behavior and the thermodynamics of the protein-substrate complex. Below, we provide a brief overview of the theoretical background behind molecular dynamics simulation and the free energy calculation techniques utilized in our study.

### 2.2 Molecular dynamics simulation (MD Simulation)

MD is a computer simulation technique used for studying the time-dependent dynamical behavior of biomolecules.<sup>91-93</sup> Since, conformational freedom of the protein and the ligand over the course of time is essential for biological function and cannot be always predicted from static crystal structures, MD simulation is useful in predicting dynamical behavior of the protein-carbohydrate complexes.

In MD simulation, the movement of atoms and molecules are determined by numerically solving Newton's laws of motion. Forces between the atoms and their potential energies are calculated using an interatomic potential energy function,  $U$ , for all  $N$  atoms in a system as shown in Equation 2.1,<sup>94</sup>

$$\vec{F}_i = m_i \vec{a}_i = m_i \frac{d^2 \vec{r}_i}{dt^2} = -\nabla_{\vec{r}_i} U(\vec{r}_1, \vec{r}_2, \vec{r}_3, \dots, \vec{r}_N), \quad i = 1, 2, 3, \dots, N \quad (2.1)$$

where  $m_i$  represents the mass of atom  $i$ ,  $\vec{a}_i$  represents the acceleration of atom  $i$ , and  $\vec{r}_i$  its position.

The potential energy function has two major components defining the interatomic interactions: bonded and non-bonded contributions, as shown in Equation 2.2.<sup>95</sup>

$$\begin{aligned}
U(\vec{r}) &= U_{\text{bonded}} + U_{\text{nonbonded}} \\
&= \left( \sum_{\text{bond}} k_b (b - b_0)^2 + \sum_{\text{angle}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedral}} k_\varphi (1 + \cos(n\varphi - \delta)) \right. \\
&\quad \left. + \sum_{\text{improper}} k_\omega (\omega - \omega_0)^2 + \sum_{\text{Urey-Bradley}} k_{\text{UB}} (S - S_0)^2 + U_{\text{cmap}}(\varphi, \omega) \right) \\
&\quad + \left( \sum_{\text{van der Waals}} \epsilon_{ij}^{\text{min}} \left[ \left( \frac{R_{ij}^{\text{min}}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{ij}^{\text{min}}}{r_{ij}} \right)^6 \right] \right. \\
&\quad \left. + \sum_{\text{coulombic}} \frac{q_i q_j}{4\pi\epsilon_0\epsilon r_{ij}} \right) \tag{2.2}
\end{aligned}$$

The bonded interactions include six terms: bond stretching ( $b$ ), angle bending ( $\theta$ ), dihedral angle rotation ( $\varphi$ ) with a phase shift ( $\delta$ ), improper angle bending ( $\omega$ ), Urey–Bradley vibrational term and the CMAP backbone torsional correction term ( $U_{\text{cmap}}$ ). The non-bonded interactions include two terms: van der Waals and electrostatic interactions. The van der Waals interaction is described by a 12-6 Lennard-Jones potential (LJ), where  $\epsilon_{ij}^{\text{min}}$  represents the depth of the potential well,  $R_{ij}^{\text{min}}$  is the distance at which LJ potential reaches its minimum value, and  $r_{ij}$  is the interatomic distance between two atoms  $i$  and  $j$ . The  $r_{ij}^{-12}$  term represents the short-range repulsive interaction and the  $r_{ij}^{-6}$  term represents the long-range attractive/dispersive interaction of the LJ potential. The electrostatic contribution is defined based on Coulombic interaction, where  $q_i$  and  $q_j$  are the partial charges on atoms  $i$  and  $j$ ,  $\epsilon$  is the relative dielectric constant, and  $\epsilon_0$  is the permittivity of

vacuum. Equation 2.2 is the general form of the CHARMM36 all atom force field used in our study.<sup>95</sup> In Equation 2.2, the “K” values denote the respective force constants, and the variables with subscript “0” represent the equilibrium values of bond length ( $b_0$ ), angle ( $\theta_0$ ), improper angle ( $\omega_0$ ), and Urey-Bradley term ( $S_0$ ), obtained either from experiments or from detailed quantum calculations.

After defining the potential energy function, the energetics of a system can be obtained from MD simulation either in terms of Helmholtz free energy ( $A$ ) in the canonical ensemble ( $NVT$ ) or in terms of Gibb’s free energy ( $G$ ) in the isobaric isothermal ensemble ( $NPT$ ) according to Equations 2.3-2.8, where  $N$  represents the total atoms in the system,  $V$  represents the system volume,  $T$  denotes the system temperature,  $P$  denotes the system pressure,  $H$  represents the enthalpy, and  $S$  the entropy of the system.

$$H = U + PV \quad (2.3)$$

$$A = U - TS \quad (2.4)$$

$$G = H - TS = U + PV - TS = A + PV \quad (2.5)$$

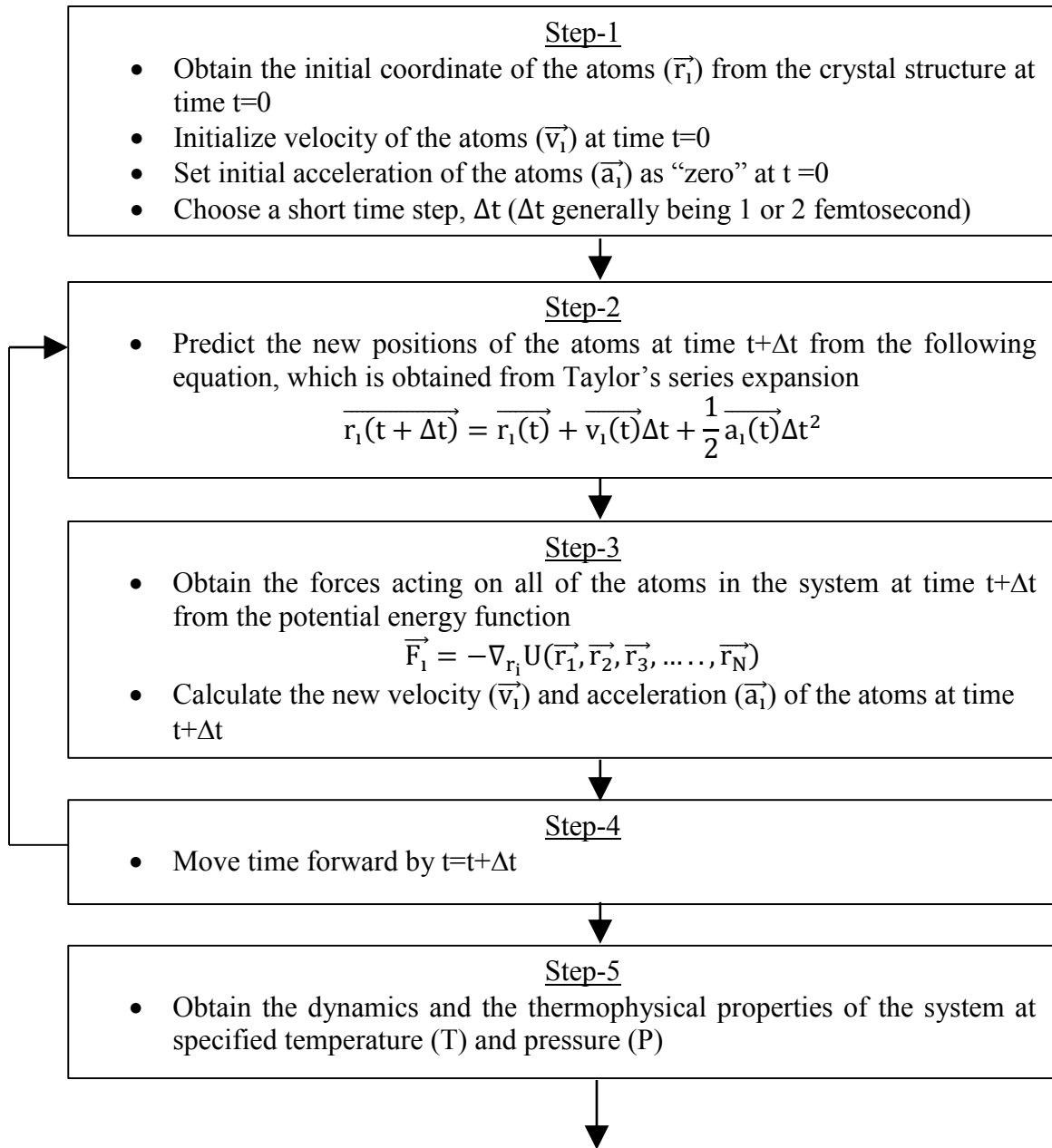
$$dA = dU - TdS \quad (\text{constant } N, V, \text{ and } T) \quad (2.6)$$

$$dG = dH - TdS = dA + d(PV) = dA + PdV \quad (\text{constant } N, P, \text{ and } T) \quad (2.7)$$

$$dG = dA + VdP \quad (\text{constant } N, V, \text{ and } T) \quad (2.8)$$

The Equations from 2.3 to 2.8 represent the relationship between  $G$  and  $A$  for closed systems having a constant number of atoms. We carry out MD simulations in the canonical ensemble to obtain the energetics of the biological complexes assuming  $dA$  and  $dG$  are approximately equal, as our systems of interest are in the condensed phase and are approximately incompressible at room temperature and atmospheric pressure (Equations 2.7 and 2.8).

Finally, we provide a simple schematic summarizing the necessary steps involved in MD simulations to obtain the dynamics and the energetics of the biomolecular complexes (Figure 2.1). The Newton's equations of motion are solved numerically using the verlet, velocity verlet or leapfrog algorithm (modification of Taylor's series expansion) to predict atom position as a function of time.<sup>94</sup> Many software packages are commercially available to perform MD simulation, such as CHARMM,<sup>95</sup> NAMD,<sup>96</sup> AMBER,<sup>97</sup> and GROMACS,<sup>98-102</sup> the first three of which we have used in our study.



**Figure 2.1** Schematic describing the steps involved in MD simulations. Figure adopted with modification from Chikvaidze et al.<sup>103</sup>

## 2.3 Free energy calculation methods

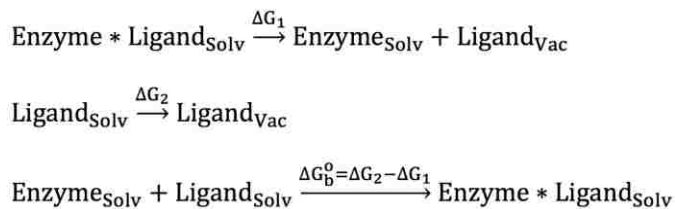
To evaluate the thermodynamics underlying the conformational changes of the protein and the ligand as well as to determine the binding affinities of the ligand to the protein, we have used several free energy calculation techniques, including free energy perturbation, thermodynamic integration, and umbrella sampling, which we briefly describe below.

### 2.3.1 Free energy perturbation with replica exchange molecular dynamics (FEP/ $\lambda$ -REMD)

FEP/ $\lambda$ -REMD is an advanced sampling free energy technique, used to calculate the absolute binding free energy of small molecules to proteins. This approach quantitatively addresses topological contributions of GHs to ligand binding free energy and can be directly compared to measured processivity values for correlation. Moreover, the absolute binding free energy computed from FEP/ $\lambda$ -REMD can be also compared with experimentally measured binding affinity from (ITC).

FEP/ $\lambda$ -REMD is a computational protocol developed by Deng and Roux and further modified by Jiang et al., where free energy perturbation is coupled with replica exchange MD to improve Boltzmann sampling of kinetically trapped conformations.<sup>104-106</sup> Free energy calculations of this type involve two thermodynamic steps: (1) ligand decoupling from the protein-ligand complex in solution and (2) ligand decoupling from the solvated ligand system. The difference in free energy values from these two steps results in the absolute ligand binding free energy ( $\Delta G^{\circ}_b$ ) of the enzyme-ligand complex (Figure 2.2).





**Figure 2.2** Thermodynamic cycle for determining  $\Delta G_b^0$  with FEP/ $\lambda$ -REMD. “Solv” refers to the solvated system and “Vac” refers to the vacuum.

To evaluate  $\Delta G_1$  in the first thermodynamic step, a potential energy function, similar to MD simulation, is used in FEP/ $\lambda$ -REMD, which is expressed in terms of repulsive, dispersive, and electrostatic thermodynamic coupling parameters,  $\lambda_{\text{rep}}$ ,  $\lambda_{\text{disp}}$ ,  $\lambda_{\text{elec}}$ , respectively, and an additional restraint parameter,  $\lambda_{\text{rstr}}$ , as shown in Equation 2.9. The thermodynamic lambda parameters,  $\lambda_{\text{rep}}$ ,  $\lambda_{\text{disp}}$ , and  $\lambda_{\text{elec}}$ , control shifted Weeks-Chandler-Anderson repulsive and dispersive components of the Lennard-Jones potential,  $U_{\text{rep}}$  and  $U_{\text{disp}}$ , and Columbic interactions,  $U_{\text{elec}}$ , respectively.<sup>107</sup> The restraint parameter,  $\lambda_{\text{rstr}}$ , controls translation and rotation in the ligand using a restraint potential,  $U_{\text{rstr}}$ .  $U_0$  represents the potential energy of the system containing the non-interacting ligand.

$$U = U_0 + \lambda_{\text{rep}}U_{\text{rep}} + \lambda_{\text{disp}}U_{\text{disp}} + \lambda_{\text{elec}}U_{\text{elec}} + \lambda_{\text{rstr}}U_{\text{rstr}} \quad (2.9)$$

Each of the thermodynamic lambda parameters ranges from 0 to 1, where “0” represents full interaction of the ligand with the surrounding environment including the protein and the solvent molecules and “1” represents the full decoupling of the ligand from the environment to the vacuum with no interaction with the surrounding. The repulsive contribution ( $\Delta G_{\text{rep}}$ ) evaluated as a function of  $\lambda_{\text{rep}}$  during FEP/ $\lambda$ -REMD is shown in Equation 2.10, wherein other interactions with the environment are kept intact.

$$\begin{aligned}
& U(\lambda_{\text{rep}} = 0, \lambda_{\text{disp}} = 0, \lambda_{\text{elec}} = 0, \lambda_{\text{rstr}} = 1) \\
& \quad \rightarrow U(\lambda_{\text{rep}} = 1, \lambda_{\text{disp}} = 0, \lambda_{\text{elec}} = 0, \lambda_{\text{rstr}} = 1)
\end{aligned} \tag{2.10}$$

Similarly, the  $\Delta G_{\text{disp}}$ ,  $\Delta G_{\text{elec}}$ , and  $\Delta G_{\text{rep}}$  contributions evaluated during FEP/ $\lambda$ -REMD are shown in Equations 2.11, 2.12, and 2.13, respectively.

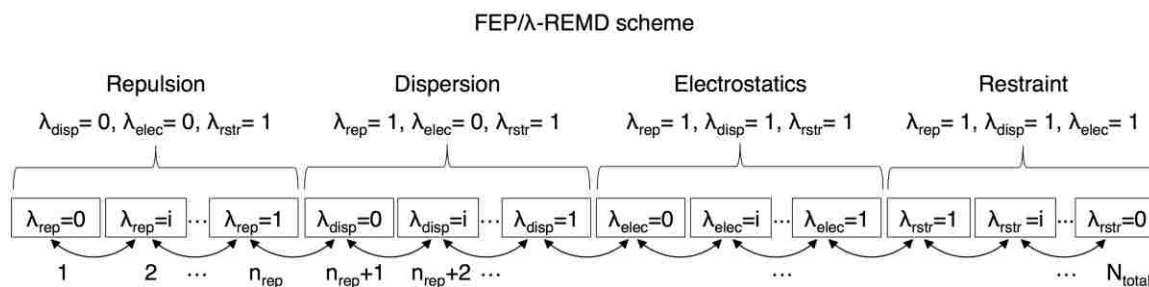
$$\begin{aligned}
& U(\lambda_{\text{rep}} = 1, \lambda_{\text{disp}} = 0, \lambda_{\text{elec}} = 0, \lambda_{\text{rstr}} = 1) \\
& \quad \rightarrow U(\lambda_{\text{rep}} = 1, \lambda_{\text{disp}} = 1, \lambda_{\text{elec}} = 0, \lambda_{\text{rstr}} = 1)
\end{aligned} \tag{2.11}$$

$$\begin{aligned}
& U(\lambda_{\text{rep}} = 0, \lambda_{\text{disp}} = 1, \lambda_{\text{elec}} = 0, \lambda_{\text{rstr}} = 1) \\
& \quad \rightarrow U(\lambda_{\text{rep}} = 1, \lambda_{\text{disp}} = 1, \lambda_{\text{elec}} = 1, \lambda_{\text{rstr}} = 1)
\end{aligned} \tag{2.12}$$

$$\begin{aligned}
& U(\lambda_{\text{rep}} = 1, \lambda_{\text{disp}} = 1, \lambda_{\text{elec}} = 1, \lambda_{\text{rstr}} = 1) \\
& \quad \rightarrow U(\lambda_{\text{rep}} = 1, \lambda_{\text{disp}} = 1, \lambda_{\text{elec}} = 1, \lambda_{\text{rstr}} = 0)
\end{aligned} \tag{2.13}$$

The replica exchange scheme among different  $\lambda$  is illustrated in Figure 2.3. A standard Metropolis Monte Carlo algorithm (Equation 2.14) is used to describe the exchange probability between swaps of systems with different Hamiltonians at different  $\lambda$ , where  $U$  is the Hamiltonian, being a function of coupling parameter  $\lambda$  and position vector  $\vec{r}$  at different values  $m$  and  $n$ .

$$\exp \left[ \frac{1}{k_B T} (U(\lambda_m, \vec{r}_m) + U(\lambda_n, \vec{r}_n) - U(\lambda_m, \vec{r}_n) - U(\lambda_n, \vec{r}_m)) \right] \geq \text{random}(0,1) \tag{2.14}$$



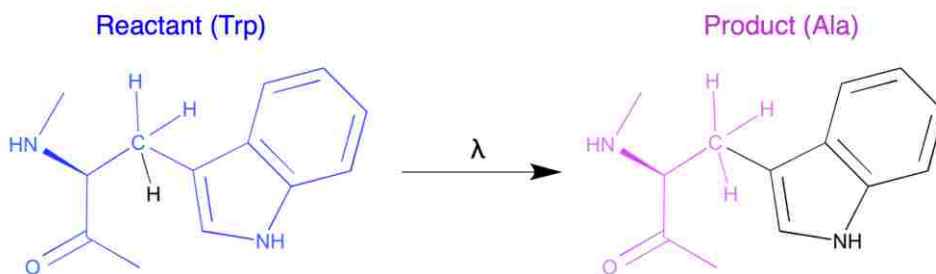
**Figure 2.3** FEP/ $\lambda$ -REMD protocol in the context of reversible work staging process to compute absolute binding free energy of the ligand ( $\Delta G^{\circ}_b$ ). Each rectangular box represents an individual MD simulation with its own input and output under a specified  $\lambda$ . The curly arrows represent the possible exchanges of the Hamiltonians ( $U$ ) between neighboring replicas based on the Metropolis Monte Carlo algorithm (Equation 2.14).  $n_{\text{rep}}$  represents the number of replicas used to calculate the repulsive component of the Lennard-Jones potential whereas  $N_{\text{total}}$  is the total number of replicas used to compute overall  $\Delta G^{\circ}_b$ . Adapted with permission from Jiang et al. 2010, American Chemical Society.<sup>106</sup>

Finally, multistate Bennett acceptance ratio was used to determine the free energies and the statistical uncertainty of the individual repulsive, dispersive, electrostatic and restraining contributions from the energies collected over the production simulation.<sup>108</sup> The ligand solvation energy,  $\Delta G_2$ , in the second thermodynamic step could be calculated in a similar fashion, though without the restraint term. In this dissertation, we implemented FEP/ $\lambda$ -REMD protocol in NAMD to evaluate ligand binding free energy ( $\Delta G^{\circ}_b$ ).<sup>96</sup>

### 2.3.2 Thermodynamic integration (TI)

TI is a computational approach, very similar to FEP/ $\lambda$ -REMD, to calculate the relative change in ligand binding free energy ( $\Delta\Delta G$ ) as a result of mutating a wild-type residue to a mutant residue. This relative change in ligand binding free energy for each variant,  $\Delta\Delta G$ , computed from TI can provide insight into precisely what relationship the residue has with enzyme-substrate binding. We performed TI calculations using NAMD with the dual topology methodology in our study to explore the role of aromatic and polar residues in ligand binding.<sup>96</sup>

The key strategy in the “dual topology” scheme is to build a hybrid residue containing both initial reactant atoms (residue of interest prior to mutation, aromatic and polar amino acids for our study) and final product residue atoms (alanine after mutation in our case), which do not interact with each other during the entire course of the simulation but interact with the rest of the system via bonded and nonbonded interactions during transformation from reactant to product over a scaled coupling parameter  $\lambda$ , ranging from 0 to 1 (Figure 2.4).



**Figure 2.4** A hybrid residue “Trp-Ala” used in dual-topology TI calculations to compute the relative change in binding free energy ( $\Delta\Delta G$ ) for tryptophan (Trp) to alanine (Ala) mutation. The wild-type reactant structure (Trp) is shown in blue and the variant product

structure (Ala) is shown in magenta. During TI calculations, the aromatic ring of the tryptophan and the hydrogen of the alanine, shown in black, do not interact with each other; though, they interact with the rest of the system separately as a function of coupling parameter  $\lambda \in [0, 1]$ .

The free energy changes associated with the reactant to product transformation are then determined simply by integrating the gradient of the mixed potential energy from  $\lambda$  values of 0 to 1, as described in Equations 2.15 and 2.16.

$$U(\lambda) = (1 - \lambda) U_I + \lambda U_F \quad (2.15)$$

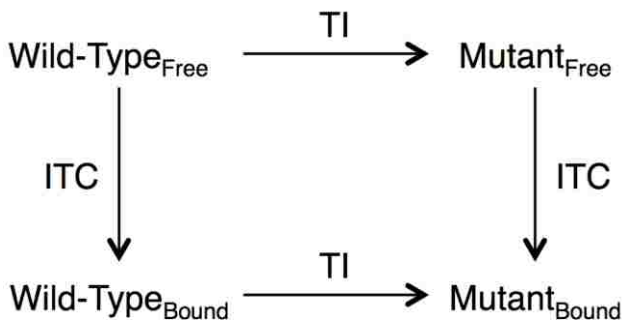
$$\Delta G = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_{\lambda} d\lambda \quad (2.16)$$

Here,  $U_I$  and  $U_F$  are energy functions for the initial reactant and final product states coupled with  $\lambda$  parameter ranging from 0 to 1, and  $\Delta G$  is the change in free energy associated with the transformation of the reactant to the product.

Using the approach described above, the binding free energy of each mutant relative to wild type,  $\Delta\Delta G$ , can be determined from Equation 2.17, where “M” refers to the mutant and “WT” refers to the wild type. The thermodynamic cycle used in these calculations is shown in Figure 2.5. As  $\Delta G$  is a state function and does not depend on path during calculations, TI follows a different thermodynamic path to determine  $\Delta\Delta G$  than ITC. In Equation 2.17,  $\Delta G_{WT(\text{Bound-Free})}$  and  $\Delta G_{M(\text{Bound-Free})}$  are the free energies of binding a ligand to wild-type and mutated enzymes, evaluated from ITC, while  $\Delta G_{\text{Bound}(M-WT)}$  and  $\Delta G_{\text{Free}(M-WT)}$  are free energy changes associated with mutating

wild-type to the variant enzyme with and without a ligand, computed using the TI methodology.

$$\begin{aligned}\Delta\Delta G &= (\Delta G_{M(\text{Bound}-\text{Free})} - \Delta G_{\text{WT}(\text{Bound}-\text{Free})})_{\text{ITC}} \\ &= (\Delta G_{\text{Bound}(M-\text{WT})} - \Delta G_{\text{Free}(M-\text{WT})})_{\text{TI}}\end{aligned}\quad (2.17)$$



**Figure 2.5** Thermodynamic cycle for measuring  $\Delta\Delta G$  using either ITC or TI. Free and bound refer to the absence and presence of the ligand, respectively. ITC and TI labels over arrows represent the thermodynamic legs of the cycle used to determine  $\Delta\Delta G$  from ITC and TI, respectively.

To perform TI calculations, we will evaluate the electrostatic and van der Waals contributions of  $\Delta G$  separately to increase the computational efficiency of our simulations as well as to eliminate the instabilities due to large energy interactions. In addition, we will use soft-core potential functions for both of the electrostatic and the van der Waals terms to overcome the endpoint singularities, which may arise due to high repulsive terms as well as the overlapping of the wild type and the mutant atoms at the very end values of  $\lambda$  at 0 and 1.<sup>109-110</sup> The soft-core potentials will consider the distance between two atoms as  $\sqrt{r^2 + f(\lambda)\delta}$  instead of interatomic distance “ $r$ ”, where  $\delta$  is an adjustable parameter.

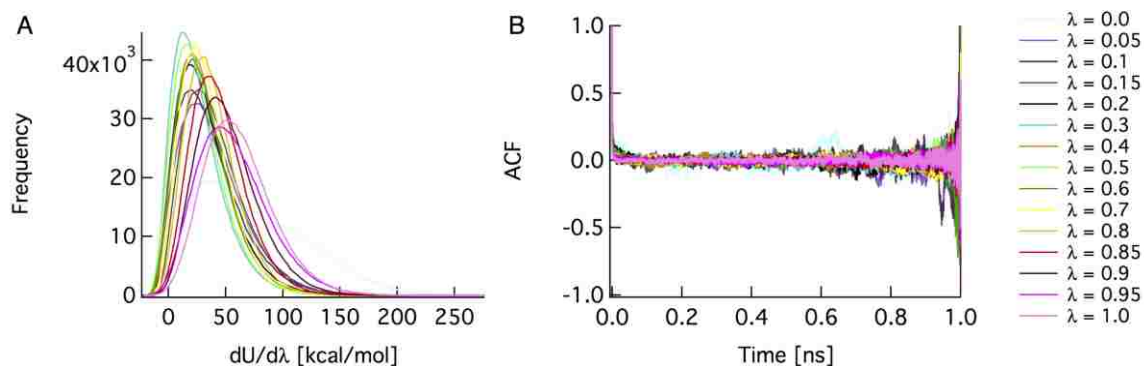
Finally, the simulations comprising the TI calculations were evaluated for appropriate window overlap to ensure convergence of  $dU/d\lambda$  at each  $\lambda$  window, as described by Pohorille et al.<sup>110</sup> Window overlap was considered sufficient when at least 25% of neighboring windows overlapped. The protocol described by Steinbrecher et al. was used further to determine the associated error, according to Equations 2.18 and 2.19.<sup>111</sup> The autocorrelation function (ACF) of each dataset was also determined based on Equation 2.20, as required for error analysis. An example is provided in Figure 2.6, showing the histogram overlap and the ACF determination representative of the trends observed for the thermodynamic data collected in our study.

$$\Delta_i \left( \frac{\partial U}{\partial \lambda} \right) = \frac{\sigma_i}{\sqrt{(t_{MD}/2\tau)}} \quad (2.18)$$

$$\Delta_{total} = \sum_i \frac{1}{2} (\lambda_{i+1} - \lambda_{i-1}) \Delta_i \quad (2.19)$$

$$c(t) = \frac{1}{c_0(N-t)} \sum_i^{N-t} (U_i - \bar{U})(U_{i+t-1} - \bar{U}) \quad (2.20)$$

Here,  $\Delta_i$  and  $\sigma_i$  represent the error as well the standard deviation for window  $i$ , respectively, whereas  $t_{MD}$  represents the total simulation time for each window,  $\tau$  the autocorrelation time and  $\Delta_{total}$  the total error evaluated over all of the  $\lambda$  windows.  $c_0$  and  $c(t)$  are the autocorrelation functions at time 0 and  $t$ , respectively.  $N$  denotes the sample data points, and  $U$  represents the potential energy, which is  $dU/d\lambda$  in our case.



**Figure 2.6** TI convergence assessment example from the van der Waals calculations for phenylalanine to alanine mutation in the ligand bound state. (A) Overlap of the  $dU/d\lambda$  histograms determines if selected windows appropriately sample the change in potential energy. (B) The autocorrelation function (ACF) is used to determine the point at which data is no longer correlated, as required for error analysis. The figure has been taken with permission from Jana et al.,<sup>112</sup> Copyright © 2016, American Chemical Society.

### 2.3.3 Umbrella sampling

Umbrella sampling is a computational technique used to improve the sampling of a system where energy barriers separate multiple configuration states of the system that cannot be accessed easily in classical MD simulation.<sup>113-115</sup> The general idea behind umbrella sampling is to add a biasing potential,  $W(\xi(r))$ , to the unbiased potential function,  $U^{ub}(r)$ , to obtain a modified biased potential function,  $U^b(r)$ , which can connect different energy states of the system by flattening the barrier heights, as shown in Equation 2.21. The sampling of the entire phase space can be achieved either in one simulation or via multiple simulations (windows) along appropriately chosen reaction coordinates,  $\xi(r)$ , which can describe the progress of the transition from one state to another. The biasing potential used to modify the potential energy function is generally in



the form of a harmonic potential restraint that keeps the value of a relevant reaction coordinate to a specified window value,  $\xi_0$ , with the help of a force constant,  $k_{\text{umbrella}}$ , as shown in Equation 2.22.

$$U^b(r) = U^{\text{ub}}(r) + W(\xi(r)) \quad (2.21)$$

$$W(\xi) = k_{\text{umbrella}}(\xi - \xi_0)^2 \quad (2.22)$$

The reaction coordinate chosen for umbrella sampling itself is a function of atom position ( $r$ ), similar to the potential energy functions  $U^{\text{ub}}(r)$  and  $U^b(r)$ . The probability of different conformations ( $P(\xi)$ ) explored across specified  $\xi(r)$  for canonical ( $NVT$ ) or isothermal isobaric ( $NPT$ ) ensemble and the corresponding potential of mean force (PMF) or the free energy change ( $A(\xi)$ ) can be described under biased or unbiased potentials according to Equations 2.23 to 2.27, where the terms superscript with “b” are biased and superscript “ub” represents unbiased terms,  $\beta$  denotes the thermodynamic parameter ( $1/k_b T$ ),  $k_b$  represents the Boltzmann constant and  $T$  the absolute temperature. In addition, the function  $\delta(\xi - \xi(r))$  in Equation 2.23 denotes the conformations considered only along  $\xi(r)$ . The Equations 2.26 and 2.27 represent the relationship of the biased  $A^b(\xi)$  and unbiased  $A^{\text{ub}}(\xi)$  energy states, where  $F$ ,  $F'$ , and  $C'$  are arbitrary constants, and  $W'$  is the additional energy contribution due to the addition of biasing potential  $W(\xi(r))$  and is a function of  $\xi$ .

$$P^{\text{ub}}(\xi) \propto \int \exp(-\beta U^{\text{ub}}(r)) \delta(\xi - \xi(r)) dr \quad (2.23)$$

$$A^{\text{ub}}(\xi) = -\frac{1}{\beta} \ln(P^{\text{ub}}(\xi)) \quad (2.24)$$

$$\begin{aligned}
P^b(\xi) &\propto \int \exp(-\beta U^b(r)) \delta(\xi - \xi(r)) dr \\
&\propto \int \exp(-\beta(U^{ub} + W)) \delta(\xi - \xi(r)) dr \\
&\propto (\exp(-\beta W)) P^{ub}(\xi)
\end{aligned} \tag{2.25}$$

$$\begin{aligned}
A^b(\xi) &= -\frac{1}{\beta} \ln(P^b(\xi)) = -\frac{1}{\beta} \ln(P^{ub}(\xi)(\exp(-\beta W))(\exp(-\beta F))) \\
&= -\frac{1}{\beta} \ln(P^{ub}(\xi)) - \frac{1}{\beta} \ln(\exp(-\beta W)) - \frac{1}{\beta} \ln(\exp(-\beta F)) \\
&= A^{ub}(\xi) + W' + F'
\end{aligned} \tag{2.26}$$

$$A^{ub}(\xi) = A^b(\xi) - W' + C' \tag{2.27}$$

The expression in Equation 2.27 shows the estimation of the unbiased energy state evaluated through a single simulation. However, we used a series of simulations where the reaction coordinate was divided into multiple windows to obtain the unbiased free energy. Accordingly, the Equation 2.27 can be modified to Equation 2.28 for window  $i$ .

$$A_i^{ub}(\xi) = A_i^b(\xi) - W_i' + C_i' \tag{2.28}$$

Finally, weighted histogram analysis method (WHAM) was used to evaluate the unbiased probability distribution and the associated potential of mean force (PMF) by removing the restraint biasing via a self-consistent iteration method, as shown in Equations 2.29 and 2.30.<sup>116-118</sup> In addition, errors were computed via a standard bootstrapping method through block averaging of the sum,<sup>116</sup>

$$P(\xi) = \frac{\sum_{i=1}^N n_i(\xi)}{\sum_{i=1}^N N_i \exp(C_i' - \beta W(\xi))} \tag{2.29}$$

$$C'_i = -\frac{1}{\beta} \ln \left\{ \sum_{\xi_{\text{bins}}} P(\xi) \exp(-\beta W(\xi)) \right\} \quad (2.30)$$

where  $N$  is the total number of windows defining the reaction coordinate, which connects two different energy states,  $n_i$  is the number of counts in each histogram bin associated with  $\xi$ .

## **Chapter 3 - Thermodynamic Relationships with Processivity in *Serratia marcescens***

### **Family 18 Chitinases**

As the title suggests, Chapter 3 provides insights into GH processivity can be related to the thermodynamics of GH substrate binding. This chapter has been adapted with permission from Hamre et al.,<sup>119</sup> Copyright © 2015, American Chemical Society. The experimental free energy of binding chito-oligosaccharide substrate and the inhibitor allosamidin to processive and non-processive chitinases, determined from ITC, were measured by our experimental collaborators (Dr. Anne Grethe Hamre and Prof. Morten Sørli) at the Norwegian University of Life Sciences. Our collaborators were responsible for all experimental steps necessary in producing ITC measurements, including the site directed mutagenesis, protein expression, protein purification, and analysis of the calorimetric data. The author of this dissertation calculated free energy of binding chito-oligosaccharide substrate to wild-type chitinases and used those values to compare with the ITC measurements. Chapter 3 was a collaborative effort from both experimental and computational scientists aimed at addressing the origin of processivity in terms of enzyme substrate interaction.

### **3.0 Abstract**

The enzymatic degradation of recalcitrant polysaccharides is accomplished by synergistic enzyme cocktails of GHs and accessory enzymes. Many GHs are processive which means that they remain attached to the substrate in between subsequent hydrolytic reactions. Chitinases are GHs that catalyze the hydrolysis of chitin ( $\beta$ -1,4-linked *N*-acetyl glucosamine). Previously, a relationship between active site topology and processivity has been suggested while recent computational efforts have suggested a link between the

degree of processivity and ligand binding free energy. We have investigated these relationships by employing computational (MD) and experimental ITC approaches to gain insight into the thermodynamics of substrate binding to *Serratia marcescens* chitinases ChiA, ChiB, and ChiC. We show that increased processive ability indeed corresponds to more favorable binding free energy and that this likely is a general feature of GHs. Moreover, ligand binding in ChiB is entropically driven, in ChiC it is enthalpically driven, and the enthalpic and entropic contributions to ligand binding in ChiA are equal. Furthermore, water is shown to be especially important in ChiA-binding. This work provides new insight into oligosaccharide binding, getting us one step closer to understand how GHs efficiently degrade recalcitrant polysaccharides.

### **3.1 Introduction**

Polymeric carbohydrate constructs, polysaccharides, are the fundamental building blocks for many of nature's most important structures and functions. The enzymatic hydrolysis of glycosidic linkages, the covalent bonds joining carbohydrates, is generally acid-catalyzed by GHs using one of two different mechanisms: either one that results in retention or one that results in inversion of the anomeric configuration.<sup>51</sup> To efficiently degrade the complex and frequently recalcitrant polysaccharide architecture,<sup>13</sup> microorganisms employ synergistic enzyme-cocktails of GHs and other accessory enzymes, each of which has a specific function in the degradation.<sup>16, 65, 120-121</sup> GHs are known to cleave the polymer chains randomly in what is termed endo-initiated hydrolysis, or they may have a preference for hydrolyzing chain ends from either the reducing or non-reducing end of the substrate by what is called exo-initiated hydrolysis. Many GHs capable of the latter also exhibit the ability to conduct endo-initiated

hydrolysis.<sup>122</sup> Furthermore, processive and non-processive GHs work together to optimize polysaccharide degradation. Processive enzymes bind individual polymer chains in long tunnels or deep clefts and repeatedly hydrolyze series of glycosidic linkages along the same chain before dissociation,<sup>51</sup> while non-processive enzymes generate new, free chain ends through rapid association, hydrolysis, and dissociation events.

Chitin, a  $\beta$ -1,4-linked insoluble, linear polymer of GlcNAc is the second most abundant polysaccharide in nature.<sup>2</sup> Chitin owes this abundance to its prevalence as a structural component among many species including the cell wall of most fungi,<sup>123</sup> the microfilarial sheath of parasitic nematodes,<sup>124-125</sup> the exoskeleton of all types of arthropods,<sup>126</sup> and the lining of many insects guts.<sup>127</sup> The GlcNAc units that chitin consists of are rotated 180° relative to each other such that the characteristic *N*-acetyl groups of each pyranose are on opposite sides.<sup>39</sup> The structural unit of chitin is thus a dimer of GlcNAc. Chitin is formed when the long GlcNAc chains orient themselves into well-packed layers mediated by hydrogen bond networks.<sup>36-37, 39, 128</sup>

Chitinases are GHs that catalyze the conversion of chitin into chitobiose units. Based on their characteristic TIM barrel fold and amino acid sequence, they are classified as family 18 GHs (GH18) ([www.cazy.org](http://www.cazy.org)).<sup>52</sup> Family 18 chitinases conduct hydrolysis through a unique, substrate-assisted mechanism in which the *N*-acetyl group of the sugar in subsite -1 acts as the nucleophile. All GH18 chitinases degrade chitin with retention of the stereochemistry at the anomeric carbon.<sup>52, 79, 88, 129-130</sup> The similarity in catalytic mechanism is a result of shared sequence motifs that form the catalytic ( $\beta/\alpha$ )<sub>8</sub>-barrel active site: a characteristic DXXDXDXE motif ending with the catalytic acid and an SXGG motif.<sup>80, 131</sup> In addition to the catalytic residues in negative subsites, all available

structures possess a family-specific hydrophobic platform consisting of one or two highly conserved aromatic acid residues close to subsite -1.<sup>132</sup>

The chitinolytic machinery of the Gram-negative soil bacteria *Serratia marcescens* has often been used as a model system to understand enzymatic degradation of recalcitrant polysaccharides.<sup>65</sup> The *S. marcescens* suite of enzymes consists of three multi-modular chitinases, Chitinase A, B, and C, among a host of other enzymes. Chitinase A (ChiA) is a reducing end-specific processive chitinase, moving towards the non-reducing end as the substrate is hydrolyzed, while Chitinase B (ChiB) is a non-reducing end-specific processive chitinase acting towards the reducing end.<sup>65, 69-70, 82, 133-134</sup> It has previously been suggested that a relationship between the topology of the GH active site and its mode of action exists.<sup>51</sup> Along these lines, ChiA exhibits a relatively open active site cleft, a typical feature of endo-acting enzymes.<sup>51, 133</sup> However, ChiB has a partially closed active site cleft, lending a more tunnel-like active site topology that is frequently observed in exo-acting enzymes.<sup>51, 134</sup> Chitinase C (ChiC) is a non-processive, endo-acting enzyme with a shallow substrate binding cleft.<sup>42, 73</sup> The open cleft is thought to enable the random association/dissociation processes. The full length ChiC, also referred to as ChiC1, tends to be cleaved by endogenous proteases to yield ChiC2, comprising the catalytic domain only.<sup>80, 135-136</sup>

Overall, the three chitinases exhibit distinct differences with regard to topology and processive ability. To this extent, a recent computational study suggests that a positive correlation exists between the degree of processivity and free energy changes.<sup>61</sup> Based on this study, we further hypothesize that the innate differences between the chitinases will manifest in their thermodynamic signatures upon substrate binding. To

investigate this hypothesis, we employ computational and experimental approaches toward determining changes in free energy, enthalpy, and entropy upon substrate binding and the molecular-level contributions to these changes. Free energy changes of binding hexa-*N*-acetyl glucosamine, (GlcNAc)<sub>6</sub>, to ChiA, ChiB, and ChiC were determined using FEP/ $\lambda$ -REMD. The experimental complement to this calculation, ITC, was used as a means of comparison and to identify the enthalpic and entropic contributions to free energies of binding. MD simulations provide additional insight into how the chitinase active sites contribute to ligand binding. The free energy changes are compared with existing apparent processivity measurements to reveal how thermodynamic signatures are related to enzymatic functionality.<sup>78</sup>

## **3.2 Materials and methods**

### **3.2.1 Chemicals**

(GlcNAc)<sub>6</sub> was obtained from Megazyme (Wicklow, Ireland). All other chemicals were of analytical grade.

### **3.2.2 Enzymes**

#### *Site directed mutagenesis*

In order to measure the free energies of the binding between (GlcNAc)<sub>6</sub> and ChiA, ChiB, and ChiC with ITC catalytically inactivated enzymes (mutation of the catalytic acid (Glu to Gln)) must be used. ChiA-E315Q and ChiB-E144Q are previously constructed.<sup>137-138</sup> ChiC-E141Q was prepared using the QuikChange<sup>TM</sup> site directed mutagenesis kit from Stratagene (La Jolla, CA, USA), as described by the manufacturer. The primers used for the mutagenesis are listed in Table 3.1 and were purchased from



Life Technologies (Carlsbad, CA, USA). To confirm that the *chic* gene contained the desired mutations and to check for the occurrence of non-desirable mutations, the mutated gene was sequenced using GATC Biotech's (Constance, Germany) LIGHTrun sequencing service before it was transformed into *Escherichia coli* BL21Star (DE3) cells (Life Technologies).

**Table 3.1** Primers used for site-directed mutagenesis and PCR amplification

Enzyme	Primer	Sequence
<u>Site-directed mutagenesis<sup>a</sup></u>		
ChiC-E141Q	Forward	5'CTGGATATCGATCTGC <b>AG</b> CAGGCGGCGATCGGC 3'
	Reverse	5'GCCGATCGCCGCCTGCT <b>GC</b> AGATCGATATCCAG 3'
<u>PCR-amplification<sup>b</sup></u>		
ChiA-E315Q	Forward	5'TCGAAGGTCGT <i>CATATGGCCGCGCCGGGC</i> 3'
	Reverse	5'CAGCCGGATCCT <i>CGAGTTATTGAACGCCGGCGC</i> 3'
ChiC-E141Q	Forward	5'TCGAAGGTC <i>CATATGAGCACAATAACACTATTAATGC</i> 3'
	Reverse	5' GCAGCCGGATCCT <i>CGAGTTAGGCGATGAGCTGCCA</i> 3'

<sup>a</sup> Mutated nucleotides in bold, <sup>b</sup> Restriction sites in italics

### *Construction of His<sub>10</sub>-ChiA-E315Q and His<sub>10</sub>-ChiC-E141Q*

His<sub>10</sub>-ChiB-E144Q is previously constructed.<sup>138</sup> To subclone ChiA-E315Q and ChiC-E141Q into the vector pET16b (Novagen, Madison, WI, USA), the chitinase fragments were amplified by PCR using primers listed in Table 3.1 (Life Technologies).<sup>137</sup> PCR reactions were conducted with Q5<sup>®</sup> High-Fidelity 2X Master Mix (New England Biolabs, Ipswich, MA, USA). The amplification protocol consisted of an initial denaturation cycle of 30 s at 98 °C, followed by 30 cycles of 5 s at 98 °C, 30 s at 55 °C, and 30 s at 72 °C, and a final step of 2 min at 72 °C. Both PCR-fragments were cloned into a NdeI/XhoI digested pET16b by using the In-Fusion HD Cloning kit (Clontech Laboratories, Kyoto, Japan). The resulting pET16b constructs were sequenced using GATC Biotech's LIGHTrun sequencing service to confirm the correct insert before they were transformed into *E. coli* BL21Star (DE3) cells (Life Technologies).

### *Protein expression*

For protein expression, *E. coli* BL21(DE3) cells containing the appropriate plasmid (His<sub>10</sub>-ChiA-E315Q, His<sub>10</sub>-ChiB-E144Q or His<sub>10</sub>-ChiC-E141Q) were inoculated into 25 ml LB medium containing 115 µg/ml ampicillin and grown at 37 °C and 200 rpm for 16 h. Cell culture were then inoculated into 250 ml LB medium containing 115 µg/ml ampicillin to an OD<sub>600</sub> of 0.1. This culture was cultivated until the OD<sub>600</sub> reached 0.8-1.0. The temperature was decreased to 22°C, and gene expression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 20 h. The cells were then harvested by centrifugation (8000 rpm, 20 min at 4 °C). Periplasmic fractions were prepared by osmotic shocking as described elsewhere.<sup>138</sup> A cytoplasmic protein extraction was also performed by resuspending the spheroplasts in lysis buffer (0.1 mg/ml lysozyme, 50 mM

Tris-HCl, 50 mM NaCl, 4 mM MgCl<sub>2</sub>, 1 mM EDTA, 0.1 mM PMSF pH 8.0) and incubating it at 37 °C for 30 min. Cell debris was removed by centrifugation (8000 rpm, 20 min at 4 °C). The resulting supernatant was used for enzyme purification. Both the periplasmic and cytoplasmic extracts were sterilized by filtration (0.2 μm) prior to protein purification.

### *Protein purification*

Proteins were purified on a column packed with Ni-NTA Agarose matrix (Qiagen, Venlo, Netherlands) (1.5 cm in diameter, 5 ml stationary phase in total). The column was pre-equilibrated in a buffer containing 20 mM Tris-HCl, 20 mM imidazole, and 500 mM NaCl at pH 8.0 before the periplasmic and cytoplasmic extracts were applied. After washing with a buffer containing 20 mM Tris-HCl and 500 mM NaCl at pH 8.0, fractions containing the enzyme were eluted with a buffer containing 20 mM Tris-HCl, 250 mM imidazole, and 500 mM NaCl at pH 8.0. A flow rate of 2.5 ml/min was used at all times. Enzyme purity was verified by SDS-PAGE and fractions containing purified enzyme were concentrated and transferred (Macrosep Advance Centrifugal Device, 10 kDa cutoff, Pall corporation, Port Washington, USA) to 20 mM potassium phosphate buffer pH 6.0. Protein concentrations were determined by using the Bradford Protein Assay from Bio-Rad (Hercules, CA, USA).

### **3.2.3 Isothermal titration calorimetry experiments**

ITC experiments with His<sub>10</sub>-ChiA-E315Q and His<sub>10</sub>-ChiC-E141Q were performed with a VP-ITC system from Microcal, Inc. (Northampton, MA, USA).<sup>139</sup> Solutions were thoroughly degassed prior to experiments to avoid air bubbles in the calorimeter. All

reactions took place in 20 mM potassium phosphate buffer pH 6.0. Standard ITC conditions were 500  $\mu\text{M}$  of  $(\text{GlcNAc})_6$  in the syringe and 15  $\mu\text{M}$  of enzyme in the reaction cell. Normally, 40-60 injections of 4-6  $\mu\text{l}$   $(\text{GlcNAc})_6$  were injected into the reaction cell at 180s intervals at 20, 25, 30, or 37  $^\circ\text{C}$  with a stirring speed of 260 rpm. At least two independent titrations were performed for each binding reaction. The heat of ionization of the buffer is 1.22 kcal/mol.<sup>140</sup> The temperature dependency of His<sub>10</sub>-ChiB-E144Q has previously been determined.<sup>141</sup>

### 3.2.4 Analysis of calorimetric data

ITC data were collected automatically using the Microcal Origin v.7.0 software accompanying the VP-ITC system.<sup>139</sup> Prior to further analysis, data were corrected for heat dilution by subtracting the heat remaining after saturation of binding sites on the enzyme. Data were fitted using a non-linear least-squares algorithm using a single-site binding model employed by the Origin software that accompanies the VP-ITC system. All data from the binding reactions fit well with the single site binding model, yielding the stoichiometry ( $n$ ), equilibrium binding association constant ( $K_a$ ), and the reaction enthalpy change ( $\Delta H_r^\circ$ ) of the reaction. The equilibrium binding dissociation constant ( $K_d$ ), reaction free energy change ( $\Delta G_r^\circ$ ) and the reaction entropy change ( $\Delta S_r^\circ$ ) were calculated from the relation described in Equation 3.1.

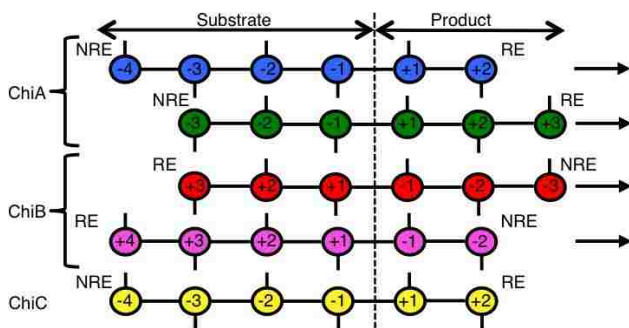
$$\Delta G_r^\circ = -RT \ln K_a = RT \ln K_d = \Delta H_r^\circ - T \Delta S_r^\circ \quad (3.1)$$

Errors are reported as standard deviations of at least two experiments at each temperature. A description of how the entropic term is parameterized has previously been described in detail.<sup>142-143</sup>

### 3.2.5 Simulation methodology

#### *Construction of chitinase models*

The inputs for MD simulations and free energy calculations were built from crystal structures of ChiA, ChiB, and ChiC obtained from Protein Data Bank entries 1EHN,<sup>45</sup> 1OGG,<sup>144</sup> and 4AXN,<sup>73</sup> respectively. In all, 8 chitinase models were constructed including: ChiA, ChiB, and ChiC both with and without ligand states. For ChiA and ChiB, two different ligand-bound simulations were constructed with the hexameric ligand occupying either the -3 to +3 binding subsites or the -4 to +2 (ChiA) and -2 to +4 (ChiB) binding subsites. The consideration of these two cases is based on experimental evidences suggesting that a GlcNAc-moiety may occupy the third product subsite after the catalytic acid (i.e. +3 for ChiA and -3 for ChiB). In the case of ChiB, occupancy of the -3 to +3 subsites is estimated as 20%, while occupancy of the -2 to +4 sites is estimated as 80%.<sup>81</sup> In the case of ChiA, (GlcNAc)<sub>5</sub> has equal probability for productive binding from -3 to +2 as -2 to +3.<sup>145</sup> A cartoon illustrating the ligand bound states for all three chitinases and their positions relative to each other is given in Figure 3.1.



**Figure 3.1** Schematic of (GlcNAc)<sub>6</sub> occupancy for each of the three chitinases considered in the simulations. The circles represent the pyranose sugar rings, and the black sticks attached to the circles correspond to the *N*-acetyl group. The black lines connecting each pyranose ring circle represent  $\beta$ -1,4 glycosidic linkages. The arrows indicate the direction the substrate slides through the active-site cleft (i.e., ChiA is reducing-end (RE) specific, and ChiB is non-reducing end (NRE) specific). ChiC is an endo-acting, non-processive enzyme, thus, no directionality arrow is shown.

Construction of the chitinase models required modification of the PDB structures so as to represent the wild-type, ligand bound systems in catalytically active conformations. For ChiA, the E315Q mutation in the 1EHN PDB was reversed, and the (GlcNAc)<sub>8</sub> bound ligand was shortened to (GlcNAc)<sub>6</sub> bound from subsite -4 to +2. For the -3 to +3 binding sites, the ligand bound structure was prepared by aligning the 1EHN PDB structure with 1E6N, which contains a bound chito-pentaose ligand. The sugar in the -4 binding subsite, from the 1EHN structure was removed, and the +3 sugar from 1E6N was added.<sup>145</sup> Similarly for ChiB, the 1OGG PDB D142N mutation was reverted to the wild-type Asp. The initial coordinates of the ChiB (GlcNAc)<sub>6</sub> ligand docked in the -3 to +3 subsites were obtained through structural alignment of PDB entries 1E6N and 1OGG in PyMOL.<sup>79, 85</sup> The second (GlcNAc)<sub>6</sub> bound structure of ChiB, docked in

subsites -2 to +4, was constructed by removing a pyranose sugar ring from subsite -3 and manually adding a sugar ring at the +4 site. The manual addition of pyranose rings in ChiA and ChiB included additional stepwise minimization gradually releasing restraints on the surrounding system to ensure the addition did not adversely affect dynamics of the remaining 5 pyranose rings. For ChiC, the (GlcNAc)<sub>6</sub> ligand was docked in the -4 to +2 subsites following the procedure described previously by Payne et al.<sup>73</sup> where a combination of ligands obtained from the structurally similar enzymes hevamine from *Hevea brasiliensis* and the NCTU2 chitinase from *Bacillus cereus* (1KQY<sup>146</sup> and 3N18<sup>147</sup>) were aligned. In ChiA and ChiB, complete proteins were considered; whereas for ChiC, only the catalytic domain (ChiC2) was simulated, as described above. The initial apo enzyme structures were constructed by removing the ligand from the active site of the above-described models. H++ was used to calculate the protonation states of all of our enzymes with and without ligand states at pH 6 and an internal and external dielectric constant of 10 and 80, respectively.<sup>148-150</sup> The protonation states are given in Table A1.1 of the Appendix A1. Disulfide bonds were specified according to the structural studies (Table A1.1).

### *MD simulations*

The structures were explicitly solvated with water, minimized, heated, and density equilibrated in CHARMM.<sup>95</sup> The CHARMM36 all-atom force field with CMAP corrections was used to model the protein and carbohydrate interactions.<sup>95, 151-152</sup> Water was modeled with the modified TIP3P force field.<sup>153-154</sup> After equilibration, 100 ns simulations in the NVT ensemble were performed using NAMD.<sup>96</sup> VMD was used for

visualization of all trajectories.<sup>155</sup> Additional details regarding system construction, simulation parameters, and protocol have been provided in the Appendix A1.

### *FEP/ $\lambda$ -REMD*

FEP/ $\lambda$ -REMD was implemented in NAMD to calculate the absolute ligand binding free energy ( $\Delta G^\circ_b$ ) of the enzyme-ligand complex.<sup>104-106</sup> As described in the Method section (Chapter 2), free energy calculations of this type involved two steps: (1) estimation of free energy change ( $\Delta G_1$ ) during ligand decoupling from the protein-ligand complex and (2) evaluation of solvation free energy ( $\Delta G_2$ ) during ligand decoupling from the solvated ligand system. The difference of in free energy values of  $\Delta G_1$  and  $\Delta G_2$  represents  $\Delta G^\circ_b$ .

All of the free energy simulations were constructed from a 25-ns snapshot obtained from the MD simulations and run for a total of 3.5 ns (35 sequential FEP calculations of 0.1 ns each). A set of 128 replicas (72 repulsive, 24 dispersive, and 32 for electrostatics) were used to determine the free energies. The replicas were exchanged with a frequency of 1/100 steps (0.1 ps). Multistate Bennett acceptance ratio (MBAR) was used to determine the free energies and statistical uncertainty of the individual repulsive, dispersive, and electrostatic contributions.<sup>108</sup> Convergence of the free energy values was assessed by monitoring the time evolution of all 35 FEP calculations (Figure A1.3 of Appendix A1). The first 1.5 ns of data were discarded as equilibration, and the last 2 ns of data were used to determine ligand binding free energy and corresponding standard deviation.



### 3.3 Results

#### 3.3.1 Binding of (GlcNAc)<sub>6</sub> to ChiA, ChiB, and ChiC determined with FEP/ $\lambda$ -REMD

Ligand binding free energies of two processive chitinases, ChiA and ChiB, and a non-processive chitinase, ChiC, were determined to understand the contributions of active site architecture to binding free energy (Table 3.2). The binding free energies for ChiA and ChiB were determined with the ligand in two different locations; in the  $-3$  to  $+3$  subsites (ChiA and ChiB) and the  $-4$  to  $+2$  subsites (ChiA) or the  $-2$  to  $+4$  subsites (ChiB). The latter cases were examined to understand the effect product site binding has on binding free energy (i.e., binding a trimer in the product subsites as a result of acetyl positioning rather than the standard dimer product of a processive chitinase). The binding free energy values are provided alongside the corresponding repulsive, dispersive, electrostatic, and restraint components. The solvation free energy of (GlcNAc)<sub>6</sub> is also reported. Errors associated with each contribution to the binding free energy (i.e., repulsive, dispersive, and electrostatic) represent 1 standard deviation over the last 2 ns of collected data, which is more conservative than that from MBAR. The error of the binding free energy was computed by taking the square root of the sum of the squared standard deviations of the chito-oligomer solvation free energy ( $\Delta G_2$ ) and the free energy needed to decouple the ligand from the enzyme-ligand complex ( $\Delta G_1$ ). The binding free energy as a function of time is given in the Appendix A1 (Figure A1.3).

**Table 3.2** Absolute ( $\Delta G_b^\circ$ ) ligand binding free energy calculated from FEP/ $\lambda$ -REMD at 300 K and pH 6.

	$\Delta G_b^\circ$ <sup>a</sup>	$\Delta G_{\text{rep}}^\circ$ <sup>a</sup>	$\Delta G_{\text{dis}}^\circ$ <sup>a</sup>	$\Delta G_{\text{elec}}^\circ$ <sup>a</sup>	$\Delta G_{\text{rstr}}^\circ$ <sup>a</sup>
(GlcNAc) <sub>6</sub>	-----	78.8 ± 1.1	-73.2 ± 0.7	-90.0 ± 0.8	-----
ChiA (-3 to +3)	-13.4 ± 2.2	117.3 ± 2.1	-126.4 ± 1.2	-88.2 ± 1.0	-0.5
ChiA (-4 to +2)	-16.7 ± 1.4	124.9 ± 2.2	-132.3 ± 0.4	-93.4 ± 1.5	-0.2
ChiB (-3 to +3)	-6.2 ± 1.5	119.6 ± 1.5	-124.1 ± 0.3	-85.9 ± 1.2	-0.2
ChiB (-2 to +4)	-15.2 ± 1.3	124.1 ± 1.3	-129.0 ± 0.3	-94.5 ± 1.1	-0.2
ChiC (-4 to +2)	-9.6 ± 1.6	109.4 ± 2.0	-113.4 ± 0.5	-89.5 ± 0.8	-0.5

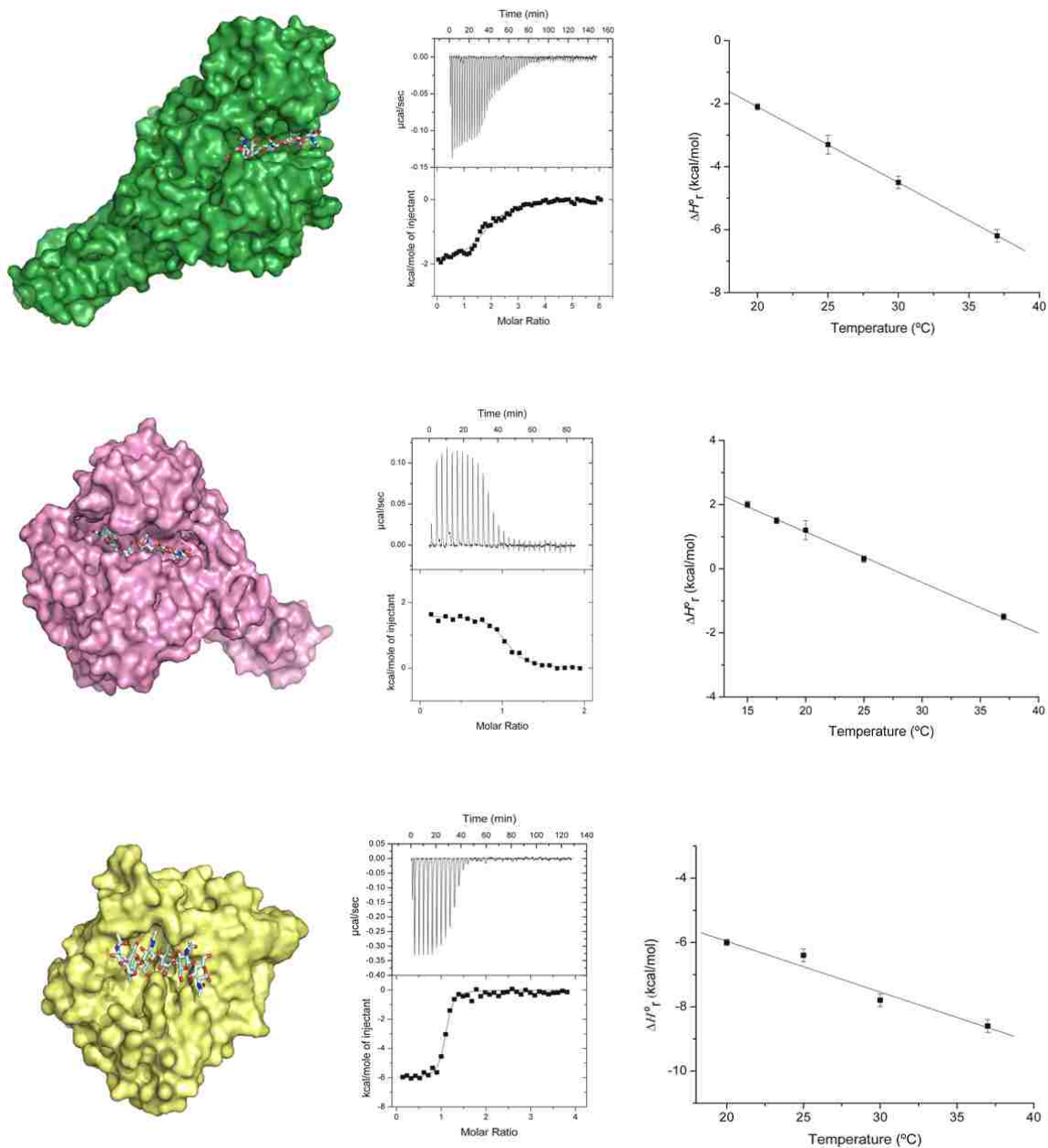
<sup>a</sup> kcal/mol

### 3.3.2 Binding of (GlcNAc)<sub>6</sub> to ChiA and ChiC determined with ITC

The binding of (GlcNAc)<sub>6</sub> to ChiA-E315Q and ChiC-E141Q at pH 6.0 (20 mM potassium phosphate buffer) at different temperatures (20-37 °C) was studied using ITC. Figure 3.2 illustrates a typical ITC thermogram and theoretical fit to the experimental data at  $t = 30$  °C. For ChiA-E315Q, the binding to (GlcNAc)<sub>6</sub> has a  $K_d = 0.56 \pm 0.03$   $\mu\text{M}$  ( $\Delta G_r^\circ = -8.7 \pm 0.1$  kcal/mol, Table 3.3). The reaction is accompanied by an enthalpic change ( $\Delta H_r^\circ$ ) of  $-4.5 \pm 0.2$  kcal/mol and an entropic change ( $\Delta S_r^\circ$ ) of  $13.9 \pm 0.7$  cal/K·mol ( $-T\Delta S_r^\circ = -4.2 \pm 0.2$  kcal/mol). The change in the heat of the reaction, as determined by Equation 3.2, was found to be  $-241 \pm 12$  cal /K·mol.

$$\Delta C_{p,r} = \left( \frac{\partial \Delta H_r^\circ}{\partial T} \right) \quad (3.2)$$

At  $t = 30$  °C, ChiC-E141Q binds (GlcNAc)<sub>6</sub> with  $K_d = 0.10 \pm 0.02$   $\mu\text{M}$  ( $\Delta G_r^\circ = -9.7 \pm 0.1$  kcal/mol, Table 3.3). The enthalpic change of the reaction is  $-7.8 \pm 0.2$  kcal/mol while the entropic change is  $6.3 \pm 0.7$  cal/K·mol ( $-T\Delta S_r^\circ = -1.9 \pm 0.2$  kcal/mol). The change in the heat of the reaction was determined to be  $-158 \pm 12$  cal/K·mol. These values are reported in tabular form alongside previously determined values for ChiB for ease of comparison (Table 3.3).<sup>138</sup>



**Figure 3.2** Binding isotherms with theoretical fits for the binding of  $(\text{GlcNAc})_6$ . Left panel: Crystal structures of  $(\text{GlcNAc})_6$  bound to ChiA, ChiB, and ChiC respectively. Middle panel: Binding isotherms (top) with theoretical fits (bottom) for the binding of  $(\text{GlcNAc})_6$  to ChiA, ChiB, and ChiC at  $t = 20^\circ\text{C}$  in 20 mM potassium phosphate at pH 6.0. Right panel: Temperature dependence of  $(\text{GlcNAc})_6$  to ChiA, ChiB, and ChiC at  $t =$

20 °C in 20 mM potassium phosphate at pH 6.0. The value of  $\Delta C_{p,r}$  is  $-241 \text{ cal/K}\cdot\text{mol}$ ,  $-158 \text{ cal/K}\cdot\text{mol}$ , and  $-158 \text{ cal/K}\cdot\text{mol}$  for ChiA, ChiB, and ChiC respectively.

**Table 3.3** Thermodynamic parameters for (GlcNAc)<sub>6</sub> binding to ChiA, ChiB, and ChiC of *Serratia marcescens* at  $t = 30 \text{ °C}$ , pH = 6.0

Enzyme	$K_d^a$	$\Delta G_r^{\text{ob}}$	$\Delta H_r^{\text{ob}}$	$-T\Delta S_r^{\text{ob}}$	$-T\Delta S_{\text{solv}}^{\text{ob,d}}$	$-T\Delta S_{\text{conf}}^{\text{ob}}$	$-T\Delta S_{\text{mix}}^{\text{ob}}$	$\Delta C_{p,r}^{\text{oc,e}}$
ChiA	$0.56 \pm 0.03$	$-8.7 \pm 0.1$	$-4.5 \pm 0.2$	$-4.2 \pm 0.2$	$-17.5 \pm 1.0$	$10.9 \pm 1.0$	2.4	$-241 \pm 12$
ChiB <sup>f,141</sup>	$0.20 \pm 0.03$	$-9.3 \pm 0.1$	$-0.1 \pm 0.3$	$-9.2 \pm 0.3$	$-11.5 \pm 0.5$	$-0.1 \pm 0.6$	2.4	$-158 \pm 5$
ChiC	$0.10 \pm 0.02$	$-9.7 \pm 0.1$	$-7.8 \pm 0.2$	$-1.9 \pm 0.2$	$-11.5 \pm 1.0$	$7.2 \pm 1.0$	2.4	$-158 \pm 12$

<sup>a</sup>  $\mu\text{M}$ , <sup>b</sup> kcal/mol, <sup>c</sup> cal/K·mol, <sup>d</sup>  $\Delta S_{\text{solv}}^\circ = \Delta C_p \ln(T_{303 \text{ K}}/T_{385 \text{ K}})$  derived using  $\Delta S_r^\circ = \Delta S_{\text{solv}}^\circ + \Delta S_{\text{mix}}^\circ + \Delta S_{\text{conf}}^\circ$  where  $\Delta S_{\text{mix}}^\circ = R \ln(1/55.5) = -8 \text{ cal/K}\cdot\text{mol}$  (“cratic” term), <sup>e</sup> derived from the temperature dependence of  $\Delta H_r^\circ$ , <sup>f</sup> derived from the interpolation of values above and below  $t = 30 \text{ °C}$ .

### 3.3.3 Parameterization of the entropic term

The entropic term,  $\Delta S_r^\circ$ , can be viewed as the sum of translational, solvational and conformational entropic changes,  $\Delta S_{\text{mix}}$ ,  $\Delta S_{\text{solv}}$ , and  $\Delta S_{\text{conf}}$ , respectively, as shown in Equation 3.3.<sup>156</sup>

$$\Delta S_r^\circ = \Delta S_{\text{mix}}^\circ + \Delta S_{\text{solv}}^\circ + \Delta S_{\text{conf}}^\circ \quad (3.3)$$

By recognizing that the entropy of solvation is close to zero for proteins near  $T = 385 \text{ K}$ ,  $\Delta C_{p,r}^\circ$  can be related to the solvation entropy change ( $\Delta S_{\text{solv}}^\circ$ ) of the binding reaction at  $t = 30 \text{ °C}$  as described by Equation 3.4.<sup>156-158</sup>

$$\Delta S_{\text{solv}}^\circ = \Delta C_{p,r} \ln\left(\frac{303 \text{ K}}{385 \text{ K}}\right) \quad (3.4)$$

Using this relationship, a  $\Delta S_{\text{solv}}^\circ$  of  $58 \pm 3$  cal/K·mol and  $38 \pm 3$  cal/K·mol can be calculated for ChiA-E315Q and ChiC-E141Q, respectively. These numbers represent  $-17.5 \pm 1.0$  kcal/mol and  $-11.5 \pm 1.0$  kcal/mol ( $-T\Delta S_{\text{solv}}^\circ$ ) of the total free energy change of  $-8.7 \pm 0.1$  kcal/mol and  $-9.7 \pm 0.1$  kcal/mol for the binding reaction between ChiA-E315Q and ChiC-E141Q and (GlcNAc)<sub>6</sub>, respectively.

Furthermore, the translational entropic change ( $\Delta S_{\text{mix}}^\circ$ ) of the reaction can be calculated as a “cratic” term, a statistical correction that reflects mixing of solute and solvent molecules and the changes in translational/rotational degrees of freedom (Equation 3.5).<sup>156</sup>

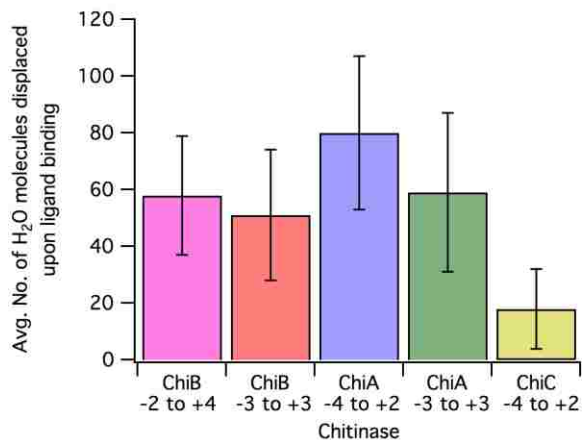
$$\Delta S_{\text{mix}}^\circ = R \ln (1/55.5) \quad (3.5)$$

Using this approach, a  $\Delta S_{\text{mix}}^\circ$  of  $-8$  cal/K·mol can be calculated corresponding to a  $-T\Delta S_{\text{mix}}^\circ$  of 2.4 kcal/mol for both ChiA-E315Q and ChiC-141Q. The conformational entropy change can then be calculated by Equation 3.3, leading to a  $\Delta S_{\text{conf}}^\circ$  of  $-36 \pm 3$  cal/K·mol and  $-24 \pm 3$  cal/K·mol. These numbers correspond to a  $-T\Delta S_{\text{conf}}^\circ$  of  $10.9 \pm 1.0$  kcal/mol and  $7.2 \pm 1.0$  kcal/mol for ChiA-E315Q and ChiC-E141Q, respectively.

### 3.3.4 Active site dynamics from MD simulations

MD simulations were conducted to elucidate molecular contributions to the various components of the thermodynamic signatures of substrate binding. From the simulation trajectories, we determined several quantities that directly connect with enthalpic and entropic changes upon binding. Here, we primarily focus on the average number of water molecules displaced upon ligand binding (Figure 3.3), a quantity that directly relates to  $\Delta S_r^\circ$  through  $\Delta S_{\text{solv}}$ . To determine the number of water molecules

displaced upon ligand binding, we calculated the number of water molecules occupying the active sites of both the ligand-free and bound chitinases over the 100-ns MD trajectories. The difference between the number of water molecules in an empty active site and the bound active site represents the number of water molecules displaced as a result of ligand binding. The number of water molecules occupying a given binding site was determined by considering the ligand-bound chitinases and selecting amino acid residues within 6 Å of the ligand. This set of amino acids was then used as a fixed frame of reference for counting the number of water molecules within the active site over time. For each frame of the eight simulation trajectories, we determined the number of water molecules within 6 Å of the previously defined amino acids, which was averaged as representative of the binding state. The average number of water molecules displaced upon ligand binding is higher for the processive chitinases, ChiA and ChiB, than for the non-processive ChiC. In general, ChiA and ChiB displace between 50 and 75 water molecules. The values for ChiA and ChiB, regardless of binding site occupancy (i.e., how many product binding sites are occupied), are within error of each other. ChiC displaces ~20 water molecules upon binding the (GlcNAc)<sub>6</sub> ligand, which reflects the more open active site topology that leaves the entire face of the (GlcNAc)<sub>6</sub> ligand solvent exposed.

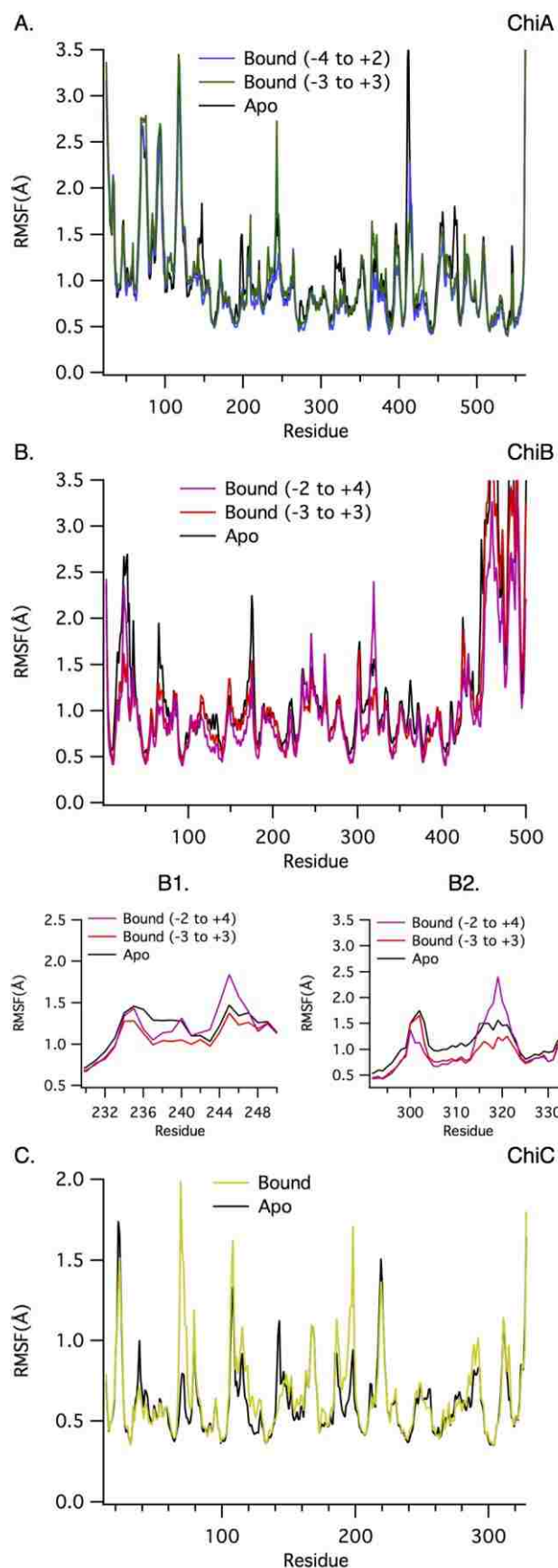


**Figure 3.3** Average number of water molecules displaced upon binding (GlcNAc)<sub>6</sub> to ChiA, ChiB, and ChiC from MD simulations. Active site desolvation upon ligand binding corresponds to experimental estimates of the solvation entropy obtained from parameterization of ITC data. Active site desolvation refers to average number of water molecules displaced upon binding of ligand GlcNAc<sub>6</sub> to all three chitinases. Error bars represent one standard deviation. The comparison shows ChiC experiences the least desolvation upon ligand binding, and desolvation of ChiA and ChiB is statistically similar irrespective of binding site occupancy, suggesting the entropic contribution due to solvation in  $\Delta G_b^\circ$  is least favorable in ChiC.

Moreover, root mean square fluctuations (RMSF) of the protein backbone over the course of 100 ns MD simulations were also undertaken and the result show that ChiA appears to rigidify upon binding (GlcNAc)<sub>6</sub> and exhibits somewhat less fluctuation than apo ChiA (Figure 3.4A). Similarly, the flexibility of ChiB with the ligand bound in the -3 to +3 binding sites is virtually unchanged upon ligand binding (Figure 3.4B). The relationship of ChiC flexibility to conformational entropy contributions is not as obvious as for ChiA and ChiB (Figure 3.4C). Interestingly, a slight increase in protein flexibility



is observed when ChiB binds (GlcNAc)<sub>6</sub> in the -2 to +4 sites. Most regions of ChiC are unaffected by ligand binding, but some regions slightly increase in flexibility. By observing higher RMSF of ligand over all binding sites in ChiC in comparison to processive ChiA and ChiB, we assume that lower affinity for the ligand results in higher unfavorable enthalpic contribution making overall enthalpy-entropy compensation unfavorable.



**Figure 3.4** The RMSF of ChiA (A), ChiB (B) and ChiC (C) protein residue backbones over 100-ns MD simulations for apo and different ligand bound states are shown. Considering only the catalytic domains, the maximum RMSF values are 3.5, 3.3, and 1.9 Å for ChiA, ChiB, and ChiC, respectively. Accordingly, the  $y$ -axes have been truncated for clarity. Insets B.1 and B.2 focus on the RMSF of selected ChiB residues displaying interesting behavior while the ligand occupies different binding sites (B.1 for residues 239 to 240 and B.2 for residues 315 to 322).

Additional MD simulation results are described in the Appendix A1. There, we address how fluctuation of the ligand (Figure A1.1 of Appendix A1), as measured by root mean square fluctuation, relates to binding free energy. Analysis of the number of hydrogen bonds formed between the ligand and the substrate provides additional insight into observed differences in binding free energy resulting from binding site occupancy (Figure A1.2 of Appendix A1).

### 3.4 Discussions

#### *The relationship of processive ability with binding free energy*

The relationship between processive ability and ligand binding free energy in GHs was previously hypothesized on the basis of calculated binding free energies for processive family 7 GHs.<sup>61</sup> The study suggested that increasing degree of processive ability was correlated with increasing affinity of the GH7 cellulases for the cellononaose oligomers. Combining a probabilistic description of processive ability, intrinsic processivity ( $P^{\text{Intr}}$ ), with thermodynamics of chemical equilibrium, a mathematical description of the relationship of processivity with free energy was suggested (Equation 3.6):

$$-\frac{\Delta G^{\circ}_b}{RT} = \ln\left(\frac{P^{\text{Intr}}k_{\text{on}}}{k_{\text{cat}}}\right) \quad (3.6)$$

where  $R$  is the universal gas constant,  $T$  is the temperature,  $k_{\text{on}}$  is the association rate coefficient, and  $k_{\text{cat}}$  is the catalytic rate coefficient. Cellulase processivity measurements vary significantly depending on the method and substrates used, making comparisons across laboratories extremely difficult. Accordingly, the Payne *et al.* study was somewhat limited by the availability of only two processivity measurements of differently able

GH7s. The assumptions underlying Equation 3.6, (1) that  $k_{\text{cat}}$  is much higher than  $k_{\text{off}}$  and (2) that the enzyme-substrate association process reaches equilibrium, are thought to be general for all GHs. Thus, the investigation of this relationship in a model GH system, *S. marcescens* chitinases, is expected to yield valuable insight toward both the validity of Equation 3.6 and its generality.

Intrinsic processivity describes the theoretical maximum processive potential of a given enzyme. The value is derived from a statistical estimation of the likelihood of the GH to dissociate from the substrate versus to take a processive, catalytic step forward on the substrate.<sup>64</sup> Experimental measurements of intrinsic processivity proves difficult for a host of reasons,<sup>74</sup> and thus, relatively few determinations of GH intrinsic processivity have been reported. Apparent processivity,  $P^{\text{app}}$ , is the actual measured value of processive ability, which encompasses environmental factors such as substrate accessibility and environmental conditions. We recently observed that  $P^{\text{app}}$  declines over time as the extent of substrate degradation increases, underlying the importance of uniformity in experimental determinations of processivity.<sup>78</sup> However, we also determined that initial  $P^{\text{app}}$  determinations provide the best measure of processive ability for comparative purposes. These initial  $P^{\text{app}}$  values are expected to trend with  $P^{\text{intr}}$  and thus provide the means by which to qualitatively compare chitinase processive ability with ligand binding free energy.

Previously reported  $P^{\text{app}}$  measurements suggest that ChiA is more processive than ChiB, and that ChiC is least processive.<sup>78</sup>  $P^{\text{app}}$  was determined from the initial ratio of dimeric to monomeric products, the  $[(\text{GlcNAc})_2]/[\text{GlcNAc}]$  ratio, of  $\beta$ -chitin degradation.

ChiA appears to be slightly more processive than ChiB, with measured initial  $P^{\text{app}}$  values of  $30.1 \pm 1.5$  and  $24.3 \pm 2.0$ , respectively. ChiC exhibits a  $P^{\text{app}}$  of  $14.3 \pm 1.4$ .

The calculated free energies of binding (GlcNAc)<sub>6</sub> to the three *S. marcescens* chitinases, ChiA, ChiB, and ChiC, reveal that increased processive ability corresponds to more favorable binding free energy, as hypothesized in Equation 3.6. Here, we consider the most relevant binding site occupancy,  $-4$  to  $+2$ , for ChiA and ChiC and  $-2$  to  $+4$  for ChiB (Figure 3.1), which encompasses the ligand binding scenario encountered in determination of  $P^{\text{app}}$  (i.e, two product subsites and four substrate subsites are occupied, to generate a dimer). The calculated binding free energies were  $-16.7 \pm 1.4$  kcal/mol,  $-15.2 \pm 1.3$  kcal/mol, and  $-9.6 \pm 1.6$  kcal/mol for ChiA, ChiB, and ChiC, respectively. ChiA binds the (GlcNAc)<sub>6</sub> with the highest affinity, while ChiC has the lowest affinity. Accordingly, ChiA is reported to be the most processive of the three chitinases, while ChiC is the least processive.

Quantitative validation of Equation 3.6 is not currently feasible, given available experimental data. Although the estimates of  $k_{\text{cat}}$  values for ChiA are available, the  $k_{\text{off}}$  values have not been measured for chitinases so far.<sup>159</sup> Moreover, the relationship of processivity to ligand binding free energy describes a fully-engaged GH, ready to perform a catalytic event. As we will describe, ITC measurements of binding free energy may not always capture the ligand in such a conformation, wherein two product subsites are bound and the distorted Michaelis complex is completely formed.

#### *Effects of binding site occupancy on binding free energy*

Both ITC and free energy calculations suggest that binding (GlcNAc)<sub>6</sub> across the  $-3$  to  $+3$  binding sites of ChiA and ChiB, with three pyranose rings on either side of the

catalytic cleavage site (+1/-1), is less favorable than when two product subsites and four substrate subsites are bound. From calculation, ChiA binds (GlcNAc)<sub>6</sub> in the -3 to +3 subsites with a free energy of  $-13.4 \pm 2.2$  kcal/mol, slightly less than across the -4 to +2 subsites. The difference between binding (GlcNAc)<sub>6</sub> in the -3 to +3 subsites and the -2 to +4 subsites is even more substantial in ChiB. ChiB binds (GlcNAc)<sub>6</sub> in the -3 to +3 subsites with a free energy of  $-6.2 \pm 1.5$  kcal/mol, less than half the free energy of binding to the -2 to +4 subsites.

The variation in ligand binding free energy as a function of binding site occupation suggests that there is significant variation in affinity for the ligand in the product and substrate binding sites. In ChiB, the significant reduction in binding affinity resulting simply from a 1-site shift into the product side indicates that substrate-side binding is much tighter than product site binding, which is consistent with experimental studies suggesting that ChiB is not product inhibited.<sup>42</sup> Thus, one would not expect the product binding sites to contribute an abnormally high degree of affinity. This contrasts the cellulose-active cellobiohydrolase, *Trichoderma reesei* Cel7A, which is known to be product inhibited, and the product binding sites have been shown to contribute significantly to the overall ligand binding free energy.<sup>61, 160</sup> The difference in free energy of binding (GlcNAc)<sub>6</sub> to ChiA in the -3 to +3 and -4 to +2 sites is more subtle than in ChiB, which is again related to the product inhibition.<sup>159</sup> Like *T. reesei* Cel7A, ChiA is somewhat inhibited by its dimeric product, meaning the +1 and +2 binding sites contribute more to binding the ligand than the equivalent sites (-1 and -2) in ChiB. Nevertheless, a clear difference in the binding free energies as a function of bound position exists for both ChiA and ChiB, which is important in interpreting experimental

measurements of free energy. Finally, we note that a recent high-speed atomic force microscopy study connected binding affinity on the substrate-binding side of ChiA and ChiB active site tunnels with processive ability, suggesting more available substrate binding sites in an enzyme active site correlates with higher processive ability.<sup>83</sup>

Experimental measurements of the free energies of binding (GlcNAc)<sub>6</sub> to ChiA and ChiB are less favorable than calculated values. Experimental data for the ligand binding free energy and accompanying thermodynamic parameters in ChiA and ChiC were obtained by performing temperature dependency measurements by ITC (Table 3.3). These values have previously been reported for ChiB.<sup>141</sup> For simplicity in discussion, we refer to the his-tagged catalytically inactive variants used in the ITC experiments, His<sub>10</sub>-ChiA-E315Q, His<sub>10</sub>-ChiB-E144Q, and His<sub>10</sub>-ChiC-E141Q, as ChiA, ChiB, and ChiC, respectively. The ITC measurements of affinity for ChiA, ChiB, and ChiC for (GlcNAc)<sub>6</sub> indicate each of these enzymes bind the hexamer ligand with approximately equal binding strength ( $-8.7 \pm 0.1$  kcal/mol,  $-9.3 \pm 0.1$  kcal/mol, and  $-9.7 \pm 0.1$  kcal/mol for ChiA, ChiB, and ChiC, respectively). The calculated value for ChiC ( $-9.6 \pm 1.6$ ) is in near perfect agreement with the experimentally obtained value. However, computational assessments of ChiA and ChiB (GlcNAc)<sub>6</sub> affinity in the -4 to +2 and -2 to +4 sites are markedly more favorable than the ITC determinations.

Rather than suggesting inaccuracy in the calculations, we posit that ITC measurements capture a mean of the possible binding conformations. Calculations generally tend to overestimate binding favorability. However, the 5-6 kcal/mol difference observed here is suggestive of a more substantial physical issue rather than computational inaccuracy, particularly with the apparent accuracy of the ChiC calculation. Based on the

above comparison of ChiA and ChiB binding the ligand with either a dimer or trimer in the product binding sites, we suggest that the ITC measurements of ChiA and ChiB describe the mean of at least two different hexameric ligand binding scenarios. It has previously been reported that ChiA equally favors productive binding of (GlcNAc)<sub>5</sub> in the -2 to +3 sites and the -3 to +2 sites suggesting that the +3 subsite may interact with (GlcNAc)<sub>6</sub> binding as well.<sup>145</sup> Similarly, Horn et al. estimate from product profiles that (GlcNAc)<sub>6</sub> may bind in the -3 to +3 binding sites of ChiB approximately 20% of the time.<sup>81</sup> Confirmation of this hypothesized variation in binding mode is difficult and likely not possible by ITC alone, as the enzymes must be catalytically inactivated to assess binding free energy. Moreover, it is possible that the catalytically active wild-type enzymes may not bind hexamer in the same fashion as inactivated enzymes, so product profiles may not accurately represent binding site occupation.

Related to the catalytically inactivated variants used in ITC, we anticipate that the formation of the distorted intermediate ligand conformation, part of the GH reaction coordinate, also may contribute to the observed differences between ITC and calculated free energies. Enzymatic reactions go through at least two distinct steps. First, the substrate reversibly binds to the enzyme forming an enzyme-substrate complex, called the Michaelis complex. Thereafter, the enzyme catalyzes the chemical step and releases the product. In GHs, the formation of the Michaelis complex generally reflects a significant distortion of the -1 binding site pyranose ring (e.g., boat, skew, envelope, etc.) away from the energetically favorable <sup>4</sup>C<sub>1</sub> conformation.<sup>51, 58</sup> Certainly, this is true for GH18 chitinases along their hydrolytic reaction coordinate,<sup>79, 88</sup> which adopt a <sup>1,4</sup>B boat conformation just prior to hydrolysis. Multiple studies have computed the free energy



landscape of all accessible  $\beta$ -glucopyranose conformers, including approximate energetic barriers to transformation between ring states as represented by the Stoddart's diagram.<sup>161-162</sup> Based on these free energy landscapes, the energetic barrier associated with traversing from the  ${}^4C_1$  conformation, through intermediate conformations, to the  ${}^1,4B$  distorted conformation, as required of GH18 ChiA and ChiB, is approximately 8 kcal/mol. Alternative conformational routes for ChiA and ChiB pass through areas with even higher energetic barriers, mainly in the area 10-15 kcal/mol. It is therefore possible to imagine that even small deviations from a complete formation of the Michaelis-complex will have great impact on the binding free energy values. While we can ensure that the complete, distorted Michaelis complex was formed when performing FEP/ $\lambda$ -REMD calculations, we are unable to directly ascertain whether this is the case when determining free energies from ITC.

#### *Thermodynamic signatures of ligand binding*

Water plays a critical role in protein function; it aids in formation of correct protein folds, for flexibility in carrying out biological functions, and is responsible for mediating protein-protein, protein-ligand, and protein-DNA interactions. Binding between biomolecules is usually accompanied by the displacement of bound water molecules from the binding sites and formation of direct interactions between the binding molecules, even though it has been observed that the water molecules sometimes are not completely removed from the binding interface.<sup>163</sup> The balance between the direct interactions gained and the solvation interactions lost determines whether such interactions are favorable, neutral, or unfavorable to binding affinity. This means that favorability is dependent on whether the strong interactions between protein and ligand

can compensate for the loss of beneficial interactions between water molecules and the protein-ligand complex.<sup>163-164</sup> It has been observed that even if the entropy increases when a water molecule is released to solution, the enthalpy can favor the bound water molecule because this will stabilize binding between protein and ligand.<sup>164</sup> Relative to *S. marcescens* chitinases, we previously observed that the average number of water molecules in the +1/-1 binding sites of non-processive chitinases tended to be higher than in processive chitinases.<sup>73</sup> There also appeared to be a relationship between which side of the +1/-1 cleavage location water molecules congregated and end-specificity. Accordingly, we anticipated the solvational entropy change upon ligand binding would yield additional insight into the role of water molecules in chito-oligomeric binding, beyond examination of only the ligand-bound state.

The solvational entropic change, determined from change in heat capacity ( $\Delta C_{p,r}^{\circ}$ ) measured by ITC, and MD simulations indicate that water molecules play a significant role in ChiA substrate binding, but less so in ChiB and ChiC. The change in heat capacity when (GlcNAc)<sub>6</sub> binds with ChiA was nearly 100 cal/K·mol larger (-241 cal/K mol) than for either ChiB and ChiC (both -158 cal/K mol).  $\Delta C_{p,r}^{\circ}$  is directly proportional to  $\Delta S_{\text{solv}}^{\circ}$  (Equation 3.4); hence, the term  $-T\Delta S_{\text{solv}}^{\circ}$  was markedly more negative and energetically beneficial for ChiA than ChiB or ChiC. The favorability of  $-T\Delta S_{\text{solv}}^{\circ}$  physically corresponds to a greater degree of desolvation upon (GlcNAc)<sub>6</sub> binding (Table 3.3). Similarly, the average number of water molecules displaced upon binding (GlcNAc)<sub>6</sub> to the chitinases, determined from MD simulations, indicates more water molecules tend to be displaced in binding to ChiA, followed by ChiB and ChiC (Figure 3.3). Though, the number of water molecules displaced by ChiC is substantially lower than either ChiA or

ChiB, reflective of the more open, shallow binding cleft. The position of the ligand within the active site appears to minimally affect number of water molecules displaced, which means that the ITC determinations of thermodynamic signatures are representative of the driving forces behind binding, even if the measurements do represent a mean rather than a single occupancy.

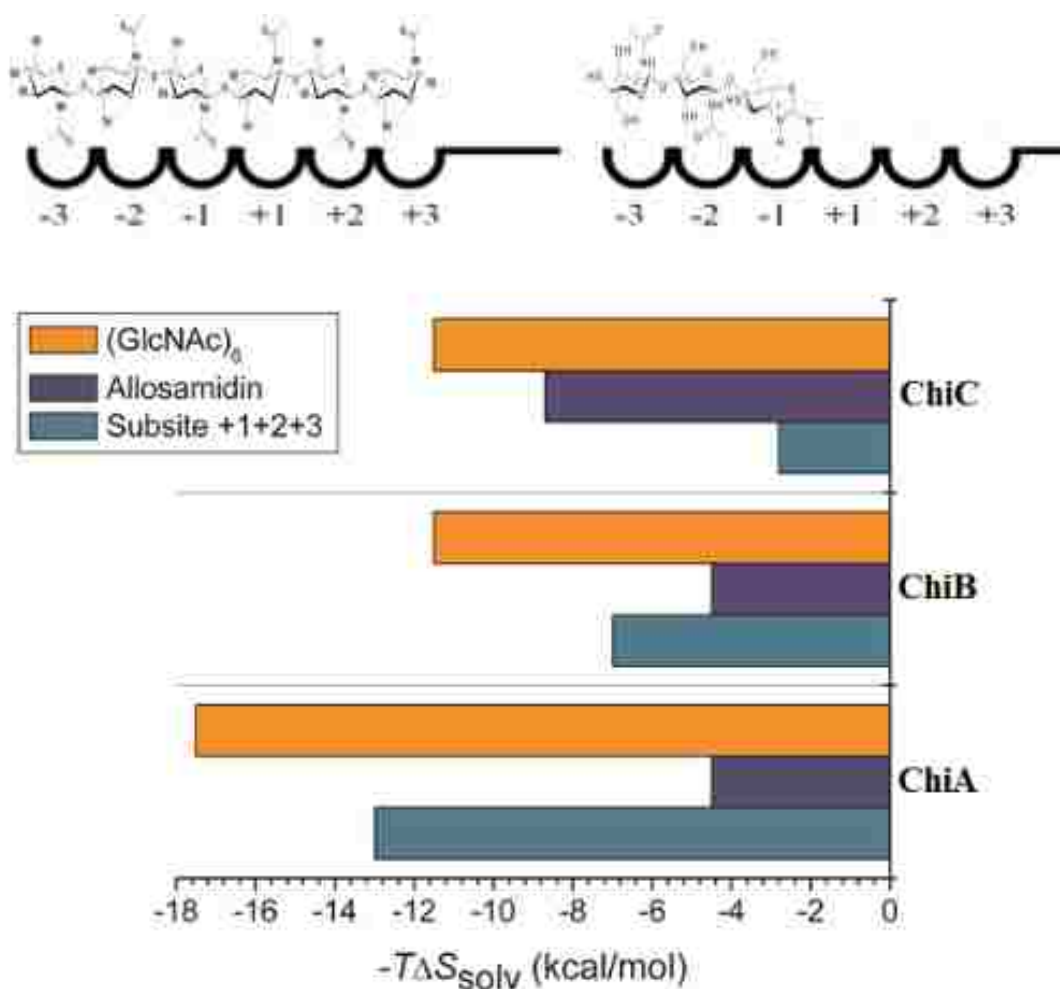
The best known family 18 inhibitor is the pseudotrisaccharide allosamidin, produced by *Streptomyces* sp.<sup>165</sup> It binds from subsite -3 to -1 for all chitinases, where the allosamizoline group of allosamidin resembles the oxazolinium ion intermediate structure formed in the retaining substrate assisted catalysis.<sup>130</sup> Binding of ChiA, ChiB, and ChiC to allosamidin has previously been reported (Table 3.4). We revisit the findings here, as when combined with the current study, allosamidin binding provides a unique perspective into contributions of the various binding subsites.<sup>142, 166</sup> In the case of ChiA, ChiB, and ChiC binding to allosamidin,  $\Delta C_{p,r}^{\circ}$  is twice as large for ChiC ( $-120 \pm 15$  cal/K·mol) as for ChiA and ChiB ( $-61 \pm 13$  cal/K·mol and  $-63 \pm 4$  cal/K·mol, respectively).<sup>166</sup>

**Table 3.4** Thermodynamic parameters for allosamidin binding to ChiA, ChiB, and ChiC of *Serratia marcescens* at  $t = 30\text{ }^{\circ}\text{C}$ ,  $\text{pH} = 6.0$

Enzyme	$K_d^a$	$\Delta G_r^{ob}$	$\Delta H_r^{ob}$	$-T\Delta S_r^{ob}$	$-T\Delta S_{solv}^{ob,d}$	$-T\Delta S_{conf}^{ob}$	$-T\Delta S_{mix}^{ob}$	$\Delta C_{p,r}^{oc,e}$
ChiA <sup>166</sup>	$0.17 \pm 0.06$	$-9.4 \pm 0.2$	$-6.2 \pm 0.2$	$-3.2 \pm 0.3$	$-4.5 \pm 1.3$	$-1.1 \pm 1.3$	2.4	$-61 \pm 13$
ChiB <sup>167</sup>	$0.16 \pm 0.04$	$-9.4 \pm 0.1$	$3.8 \pm 0.2$	$-13.2 \pm 0.2$	$-4.5 \pm 0.5$	$-11.1 \pm 0.6$	2.4	$-63 \pm 4$
ChiC <sup>f</sup>	$2.0 \pm 0.2$	$-7.9 \pm 0.1$	$-0.6 \pm 0.1$	$-7.3 \pm 0.1$	$-8.7 \pm 1.3$	$-1.0 \pm 1.3$	2.4	$-120 \pm 15$

<sup>a</sup>  $\mu\text{M}$ , <sup>b</sup>  $\text{kcal/mol}$ , <sup>c</sup>  $\text{cal/K}\cdot\text{mol}$ , <sup>d</sup>  $\Delta S_{solv}^{\circ} = \Delta C_p \ln(T_{303\text{ K}}/T_{385\text{ K}})$  derived using  $\Delta S_r^{\circ} = \Delta S_{solv}^{\circ} + \Delta S_{mix}^{\circ} + \Delta S_{conf}^{\circ}$  where  $\Delta S_{mix}^{\circ} = R \ln(1/55.5) = -8\text{ cal/K mol}$  (“cratic” term), <sup>e</sup> derived from the temperature dependence of  $\Delta H_r^{\circ}$ .

By comparing the thermodynamic signatures of (GlcNAc)<sub>6</sub> binding, allosamidin binding and the difference between the two, one gains insight into contributions over the whole active site. The difference between (GlcNAc)<sub>6</sub> and allosamidin solvational entropy change,  $(-T\Delta S_{solv}^{\circ}(\text{GlcNAc})_6) - (-T\Delta S_{solv}^{\circ}\text{allosamidin})$ , yields an estimate of the remaining contributions from the +1, +2, and +3 binding subsites,  $-T\Delta S_{solv}^{\circ}\text{subsite } +1,+2,+3$  (Figure 3.5). Such a difference implies that ChiC is most desolvated in negative subsites (substrate), ChiA in positive subsites (product), and equally across ChiB (Figure 3.5). ChiA studies have shown that it has a powerful affinity in subsite +3 (unpublished results;<sup>81, 145</sup>). The enzyme active site is more solvent accessible in subsite +3 than, for example, in subsite -3. Along with the thermodynamical data presented here, we offer this as explanation of why ChiA is more desolvated in positive subsites, bearing in mind this is a simplification for the purposes of discussion. Beneficial changes in the solvation entropy may also be caused by release of entropically constrained water molecules and is not necessarily a measure of the number of released water molecules on the surface of the protein.



**Figure 3.5** Upper panel: Chemical structures of (GlcNAc)<sub>6</sub> (left) and allosamidin (right) aligned in their respective subsites. Lower panel: Schematic representation of  $-T\Delta S_{\text{solv}}^{\circ}(\text{GlcNAc})_6$  (orange),  $-T\Delta S_{\text{solv}}^{\circ}\text{allosamidin}$  (blue), and  $-T\Delta S_{\text{solv}}^{\circ}\text{subsite } +1,+2,+3$  (cyan) for ChiA, ChiB, and ChiC.

Parameterization of the entropic term from the ITC experiments also yielded estimates for the conformational entropy change ( $-T\Delta S_{\text{conf}}^{\circ}$ ) upon (GlcNAc)<sub>6</sub> binding. Interestingly, these values are unfavorable for ChiA (10.9 kcal/mol) and “neutral” for ChiB (-0.1 kcal/mol) in line with the observations from the RMSF MD simulations of the protein backbone (Figure 3.4A & B). For ChiC, the value is also unfavorable (7.2

kcal/mol) and the relationship of ChiC flexibility to conformational entropy contributions is not as obvious as for ChiA and ChiB (Figure 3.4C).

The slight increase in protein flexibility when ChiB binds (GlcNAc)<sub>6</sub> in the -2 to +4 sites offers an intriguing explanation for the apparent enhancement in binding over the -3 to +3 binding sites (Figure 3.4B). The enhanced flexibility would have the effect of increasing the favorability of conformational entropy change. Along with this, we observe enhanced hydrogen bonding as a result of the ligand to the -2 to +4 binding sites (Figure A1.2B of Appendix A1), which would also serve to improve the favorability of the enthalpic term. Together, these variations in molecular behavior would improve the overall affinity of ChiB for (GlcNAc)<sub>6</sub> in the -2 to +4 binding site.

Finally, in the case of each chitinase,  $-T\Delta S_{\text{conf}}^{\circ}$  is less favorable for allosamidin binding than (GlcNAc)<sub>6</sub>, where the increase in  $-T\Delta S_{\text{conf}}^{\circ}$  is roughly equal for all chitinases ( $9 \pm 1$  kcal/mol). This behavior arises from the general flexibility of longer ligands, such as (GlcNAc)<sub>6</sub>, over shorter ones, such as allosamidin. Additionally, larger portions of the proteins will bind (GlcNAc)<sub>6</sub> compared to allosamidin, resulting in a loss of flexibility in these parts of the proteins.

### 3.5 Conclusions

Comparison of  $P^{\text{app}}$  measurements for ChiA, ChiB, and ChiC with calculated ligand binding free energies suggests the hypothesized relationship between the two (Equation 3.6) is general to GHs. Despite the current inability to quantitatively verify the relationship, this is an important finding, as it appears to generally describe an entire class of carbohydrate active enzymes. Of course, developing enzymes that bind too tightly will

eventually result in inhibition, along the lines of the Sabatier principle; thus, quantitative investigations are critical to establishing the limitations of processive GH function and the hypothesized relationship in modeling real behavior. Comparison of the free energies of binding (GlcNAc)<sub>6</sub> to ChiA, ChiB, and ChiC from simulation and ITC reveal potential limitations in comparing thermodynamic properties where the conformational state is unknown. We suspect that the very large energetic penalty associated with formation of the distorted Michaelis complex significantly contributes to differences between simulation and experiment. We also anticipate that the enzymes bind the hexameric ligand in several possible manners, which yields an average evolution of heat when using the ITC methodology. Despite the differences in free energies, both MD simulations and ITC suggest water plays a significant role in (GlcNAc)<sub>6</sub> binding to ChiA. Estimates of desolvation, through comparison with allosamidin binding, suggest ChiA product sites experience significant desolvation upon ligand binding; whereas, ChiC substrate sites are desolvated. ChiB appears to be equally desolvated across the length of the active site. The change in conformational entropy upon (GlcNAc)<sub>6</sub> binding in ChiA, ChiB, and ChiC is generally unfavorable or neutral at best, arising from the protein's need to stabilize a large flexible ligand. In general, ligand binding in ChiB is entropically driven, ChiC is enthalpically driven, and the enthalpic and entropic contributions to ligand binding in ChiA are equal. Overall, this study provides new insights into GH oligosaccharide binding that serve as the foundation for future GH protein engineering efforts through more "rational design" approaches.

## **Chapter 4 - Aromatic-mediated Carbohydrate Recognition in Processive *Serratia marcescens* Chitinases**

This chapter has been reprinted with permission from Jana et al.,<sup>112</sup> Copyright © 2016, American Chemical Society. The experimental binding free energy from ITC and the processivity for the wild-type chitinase and aromatic variants, reported in the current chapter, were measured by our experimental collaborators at the Norwegian University of Life Sciences. The author of this dissertation performed MD simulations and free energy calculations and used those to compare with the experimental measurements. The aromatic residues investigated in the current chapter were chosen by the author of this dissertation on the basis of the interaction energy calculations from MD simulations.

### **4.0 Abstract**

Microorganisms use a host of enzymes, including processive GHs, to deconstruct recalcitrant polysaccharides to sugars. Processive GHs closely associate with polymer chains and repeatedly cleave glycosidic linkages without dissociating from the crystalline surface after each hydrolytic step, and they are typically the most abundant enzymes in both natural secretomes and industrial cocktails by virtue of their significant hydrolytic potential. The ubiquity of aromatic residues lining the enzyme catalytic tunnels and clefts is a notable feature of processive GHs. We hypothesized these aromatic residues have uniquely defined roles, such as substrate chain acquisition and binding in the catalytic tunnel, that are defined by their local environment and position relative to the substrate and the catalytic center. Here, we investigated this hypothesis with variants of *Serratia marcescens* Family 18 processive chitinases ChiA and ChiB. We applied molecular simulation and free energy calculations to assess active site dynamics and ligand binding



free energies. ITC provided further insight into enthalpic and entropic contributions to ligand binding free energy. Thus, the roles of six aromatic residues, Trp-167, Trp-275, and Phe-396 in ChiA and Trp-97, Trp-220, and Phe-190 in ChiB, have been examined. We observed that point mutation of the tryptophan residues to alanine results in unfavorable changes in the free energy of binding relative to wild-type. The most drastic effects were observed for residues positioned at the “entrances” of the deep substrate-binding clefts and known to be important for processivity. Interestingly, phenylalanine mutations in ChiA and ChiB had little to no effect on chito-oligomer binding, in accordance with the limited effects of their removal on chitinase functionality.

#### **4.1 Introduction**

Crystalline homo-polysaccharides, such as cellulose and chitin, are vast networks of covalently bonded carbohydrates secured in well-packed layers through hydrogen bonding interactions. These molecular interactions, contributing both strength and insolubility, slow or preclude microbial and chemical attack on some of nature’s most important structures.<sup>13</sup> Since the monomeric carbohydrate units are ideal sources of carbon, microorganisms secrete enzyme cocktails capable of degrading these recalcitrant polysaccharides to monomeric and dimeric sugars.<sup>16</sup> The enzyme cocktails consist of multiple classes of processive and non-processive GHs and accessory enzymes such as  $\beta$ -glucosidases and lytic polysaccharide monooxygenases (LPMO).<sup>77</sup> Non-processive GHs attack the amorphous regions of polymer crystals and cleave glycosidic linkages once or sparingly few times, creating accessible polymeric chain ends. Conversely, processive enzymes attach to polymer chain ends (in exo-mode) or generate a new end (in endo-mode) and sequentially hydrolyze many glycosidic linkages, producing multiple

disaccharide products.<sup>168-169</sup> LPMOs improve substrate accessibility, while  $\beta$ -glucosidases reduce product inhibition by further degrading disaccharides to monosaccharides. By virtue of their ability to remain associated with the polysaccharide substrate in between subsequent catalytic cycles, processive enzymes are responsible for the majority of hydrolytic bond cleavage, and hence, are of great interest as targets for activity improvements towards efficient and economical biomass conversion.<sup>22, 69-70, 170</sup>

Historically, the ability of a GH to be processive has been attributed to global structural features of the active site geometry.<sup>51</sup> For example, the structures of processive cellulases exhibit tunnel-like active sites with loops on either side; these loops are flexible, exhibiting both an open and closed conformation allowing for endo-initiated hydrolysis and processive behavior.<sup>171</sup> Homologous, non-processive cellulases lack the loops forming the active site tunnels resulting in a more open cleft.<sup>172</sup> Similarly, processive chitinases possess deep substrate binding clefts, while non-processive chitinases have a shallow, open binding cleft.<sup>45, 73, 79</sup> Critically, biochemical studies have demonstrated that processivity can be changed dramatically by deletions in active loops as well as point mutations of key aromatic residues near the catalytic center.<sup>69-70, 68, 173</sup>

Critical examination of the chemical and dynamic composition of the active sites of processive GHs is needed to elucidate the features governing processive function. A notable feature of processive GHs is the ubiquity of aromatic residues lining the substrate-binding clefts. It is thought that these aromatic residues primarily interact with carbohydrate substrates in the catalytic tunnel via carbohydrate- $\pi$  stacking interactions.<sup>174-</sup>  
<sup>180</sup> The non-specific nature and the large interaction surface of the hydrophobic stacking interactions likely facilitate the processive mechanism, whereby the enzyme must

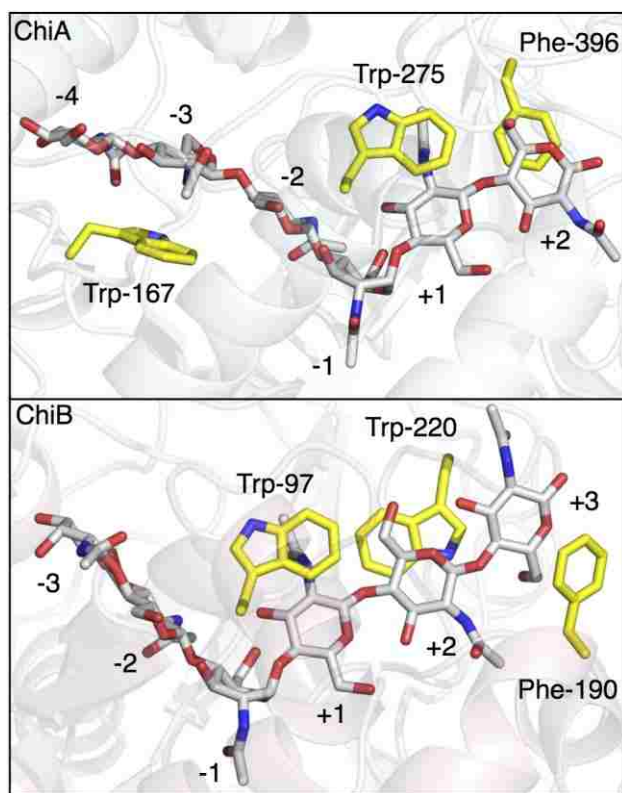
maintain attachment to the substrate while still allowing forward processive motion.<sup>44</sup> Given the prevalence of aromatic residues along GH active sites, for cellulases and chitinases alike, we set out to study the aromatic/chito-oligosaccharide interactions that occur in processive chitinases from *Serratia marcescens*, which are among the best-studied processive GHs and thus prove a useful model system.<sup>65</sup> *Serratia marcescens* family 18 chitinases include processive chitinase A (ChiA), processive chitinase B (ChiB), and non-processive chitinase C (ChiC),<sup>42, 81</sup> which act synergistically to degrade crystalline chitin to GlcNAc dimers. It is well established that ChiA and ChiB are largely exo-acting, processive enzymes (ChiA reducing end specific and ChiB non-reducing end specific) while ChiC is an endo-acting, non-processive enzyme.<sup>42, 81-84</sup> It has also been shown that when acting on soluble chitosan, ChiA and ChiB act as endo-acting processive enzymes.<sup>84</sup>

To date, experimental studies on the function of aromatic residues in GH tunnels have primarily addressed the extent to which the residue affects either activity or processive ability on crystalline or amorphous substrates. For example, early work on ChiA showed that mutation of aromatic residues at the entrance of the substrate-binding cleft (the -6 subsite)<sup>45</sup> and on the surface of ChiA's *N*-terminal chitin-binding domain reduced activity on chitin, whereas activity on soluble substrates remained unaltered.<sup>181</sup> In a later study, it was shown that aromatic residues close to the catalytic center of ChiB, Trp-97 and Trp-220 in subsites +1 and +2, respectively, are of major importance for processivity. Upon mutating each of these residues to alanine, processivity was greatly reduced; this reduction was accompanied by reduced activity on crystalline chitin and increased activity of soluble substrates.<sup>69</sup> Similar results were obtained by Zakariassen et

al. for ChiA, where an aromatic residue in the -3 subsite is crucial for processivity, consistent with the opposite directionalities of ChiA and ChiB.<sup>70</sup> The increased activity of less processive mutants on soluble substrates is thought to be related to the “stickiness” needed for gaining and maintaining access to an insoluble substrate during processive action, which becomes a disadvantage when acting on easily accessible soluble substrates where product release may become rate-limiting.<sup>71, 182</sup>

Recent computational investigations of the roles aromatic residues play in processive cellulase carbohydrate binding predict the function of an aromatic residue is specific to its position within the tunnel/cleft. For example, for *Trichoderma reesei* Cel6A, we demonstrated that mutation of aromatic residues near the catalytic center had little impact on ligand binding free energy, but mutation of aromatic residues at the entrance and exit of the Cel6A tunnel had considerable impact on binding affinity.<sup>86</sup> In a later study investigating four conserved aromatic residues in the substrate-binding clefts of processive *T. reesei* Cel7A and non-processive *T. reesei* Cel7B, Taylor et al. calculated that removal of aromatic residues near the catalytic center resulted in much less favorable relative binding free energies than tunnel entrance residues.<sup>183</sup> Similarly, Nakamura et al. illustrated that the tryptophan residue located at the entrance of the *T. reesei* Cel7A binding tunnel is responsible for substrate acquisition on crystalline cellulose substrates.<sup>184</sup> Taken together, these prior studies underpin the importance of aromatic residues, but also show variation in their possible roles. Moreover, these prior studies reveal that aromatic residues in similar binding sites in different members of the same GH family may have different functional roles.

Here, we explore the aromatic-mediated interactions of processive chitinases ChiA and ChiB with chito-oligosaccharides with the intent of defining roles for each aromatic residue. Given the frequent comparisons made between chitinases and cellulases toward understanding the broader themes of GH function, it is critical we understand the similarities and differences between the two from a molecular perspective. Further, we anticipate that detailed studies such as this will ultimately lead to a general mechanism for aromatic-carbohydrate interactions across various carbohydrate active enzyme families.<sup>23</sup> We integrate computational and experimental methods to define the roles of Trp-167 (-3 subsite), Trp-275 (+1), and Phe-396 (+2) in the cleft of ChiA, and Trp-97 (+1), Trp-220 (+2), and Phe-190 (+3) in the cleft of ChiB (Figure 4.1), which were selected on the basis of interaction energies from wild-type MD simulations conducted here. Notably, the residues in subsites +1 and +2 are structurally conserved in ChiA and ChiB, despite the enzymes' opposite directionalities. As we will discuss below, neither of the phenylalanines are completely conserved. MD simulations of wild-type and variant ChiA and ChiB enzymes bound to a hexameric oligomer of *N*-acetyl-D-glucosamine, (GlcNAc)<sub>6</sub>, were performed to understand dynamic contributions to carbohydrate binding and processive ability. Free energy calculations, using TI, and experimental determinations of ligand binding free energy, from ITC, revealed the various thermodynamic contributions. Apparent processivity of wild-type ChiA and ChiB and several of the variants have been previously reported in the literature.<sup>69-70,78</sup> We build on this literature, defining the roles of each active site aromatic residue.



**Figure 4.1** Aromatic residues examined in this study near the catalytic centers of *S. marcescens* processive chitinases ChiA and ChiB. ChiA is shown in transparent cyan cartoon, and ChiB is shown in transparent pink cartoon. The (GlcNAc)<sub>6</sub> ligand is shown in stick with gray carbons. The aromatic residues are shown in stick with yellow carbons. The ChiA and ChiB binding sites are labeled according to standard nomenclature from -4 to +2 and -3 to +3, respectively. Hydrolysis occurs between the +1 and -1 binding sites in all GHs. The two enzymes utilize identical catalytic machineries (not shown), though in opposite directions (i.e., toward the reducing end vs. toward the non-reducing end) on the substrate. ChiB lacks aromatic residues capable of stacking with the substrate in the -1 to -3 subsite.

## 4.2 Materials and Methods

### 4.2.1 Chemicals

(GlcNAc)<sub>6</sub> was purchased from Megazyme (Wicklow, Ireland). All other chemicals were of analytical grade and purchased from standard manufacturers.

### 4.2.2 Enzymes

In this study, six different residues were investigated: three in ChiA (Trp-167, Trp-275, and Phe-396) and three in ChiB (Trp-97, Trp-220, and Phe-190). Single, alanine mutants have previously been made for all ChiA residues.<sup>70</sup> Double mutants with the respective aromatic residue mutated to alanine and the catalytic acid, glutamate, mutated to glutamine were made. All double mutant genes were cloned into the vector pET16b (Novagen, Madison, WI, USA). This vector contains an N-terminal His<sub>10</sub>-tag, allowing one-step protein purification with Ni-NTA Agarose. Depending on the plasmid available, either plasmids containing the respective genes in the original pMay20-1 (ChiA) or pMay2-10 (ChiB) vector or the new pET16b vector were used as templates in the mutagenesis of the double mutants (Table 4.1).<sup>185-186</sup> For practical reasons only, two different protocols were used. The wild-type enzymes (i.e. variants containing the deactivating Glu to Gln in the catalytic center) were not produced for this study since the thermodynamic signatures of (GlcNAc)<sub>6</sub> binding to these enzymes have been determined previously.<sup>119, 138</sup>

#### 4.2.2.1 Site directed mutagenesis

All ChiA mutants, using ChiA E315Q as a template, and the F190A mutant of ChiB, using ChiB E144Q as a template, were made using the QuikChange™ site directed mutagenesis kit from Stratagene (La Jolla, CA, USA) as described by the manufacturer. The primers and templates used for the mutagenesis are listed in Table 4.1. Primers were purchased from Life Technologies (Carlsbad, CA, USA). To confirm that the genes contained the desired mutations and to check for the occurrence of undesirable mutations, the mutated genes were sequenced using GATC Biotech's (Constance, Germany) LIGHTrun sequencing service before they were transformed into *Escherichia coli* BL21Star (DE3) cells (Life Technologies, Carlsbad, CA, USA).

The ChiB mutations W220A (His<sub>10</sub>-ChiB E144Q template) and E144Q (His<sub>10</sub>-ChiB W97A template) were introduced using a two-stage PCR protocol.<sup>187</sup> The primers and templates used for the mutagenesis are listed in Table 4.1 and were purchased from Life Technologies (Carlsbad, CA, USA). In the first step of the two-stage protocol, two separate PCR reactions were completed containing only the forward or the reverse primer, respectively. These reactions consisted of a preheating step at 98 °C for 30 s followed by 6 reaction cycles of 15 s at 98 °C, 20 s at 49.8 °C, and 4 min at 72 °C, and a final extension step at 72 °C for 7 min. In the second step, the two PCR reactions were combined and continued for 30 reaction cycles of 15 s at 98 °C, 20 s at 63 °C, and 4 min at 72 °C. The amplification products were subjected to parental template digest by DpnI and transformed into *E. coli* BL21Star (DE3) cells (Life Technologies, Carlsbad, CA). *Phusion* High-Fidelity Polymerase (Thermo Scientific, Waltham, USA) was used for amplification. The mutated genes were sequenced using GATC Biotech's (Constance,



Germany) LIGHTrun sequencing service in order to verify the introduction of the desired mutations.

**Table 4.1** Primers and DNA templates used for site directed mutagenesis

Mutant	DNA template	Primer	Sequence
ChiA E315Q/W167A	ChiA E315Q <sup>137</sup>	W167Afw	5'TTCTTATTTTCGTCGAGGCGGGCGTTTACGG'3
		W167Arev	5'CGCCCGTAAACGCCCGCCTCGACGAAATA'3
ChiA E315Q/W275A	ChiA E315Q <sup>137</sup>	W275Afw	5'GTCGATCGGCGGCGCGACGCTGTCCGAC'3
		W275Arev	5'GTCGGACAGCGTCGCGCCGCCGATCGAC'3
ChiA E315Q/F396A	His <sub>10</sub> -ChiA E315Q	F396Afw	5'ACGACTTCTATGGCGGCCGCCGATCTGAAGAAC CTGG'3
		F396Arev	5'CCAGGTTCTTCAGATCGGCGGCGCCATTAGAAG TCGT'3
ChiB E144Q/W220A	His <sub>10</sub> -ChiB E144Q <sup>138</sup>	W220Afw	5'TGGCCGCCCCCGCGGAGAAG-'3
		W220Arev	5'CTTCTCCGCGGGGCCGCA-'3
ChiB E144Q/F190A	ChiB E144Q <sup>188</sup>	F190Afw	5'GCCGCGGCGCCGCTTCTGTCCGCG'3
		F190Arev	5'CGCGACAGGAAGCGGCGCCGCCGCG'3
ChiB E144Q/W97A	His <sub>10</sub> -ChiB W97A	E144Qfw	5'GGACATCGACTGGCAGTACCCGCAAGC'3
		E144Qrev	5'GCTTGCGGGTACTGCCAGTCGATGTCC'3

#### 4.2.2.2 Construction of His<sub>10</sub>-tagged double mutants

To subclone genes from their original vectors to pET16b the chitinase fragments were amplified by PCR using primers (Life Technologies) listed in Table 4.2. PCR reactions were conducted with Q5<sup>®</sup> High-Fidelity 2X Master Mix (New England Biolabs, Ipswich, MA). The amplification protocol consisted of an initial denaturation cycle of 30 s at 98 °C, followed by 30 cycles of 5 s at 98 °C, 30 s at 55 °C, and 30 s at 72 °C, and a final step of 2 min at 72 °C. The amplified inserts were cloned into a NdeI/XhoI (New England Biolabs, Ipswich, MA, USA) digested pET16b vector using the In-Fusion HD Cloning kit (Clontech Laboratories, Kyoto, Japan). The resulting pET16b constructs were sequenced using GATC Biotech's LIGHTrun sequencing service to confirm the sequence before they were transformed into *E. coli* BL21Star (DE3) cells (ChiB) or *E. coli* Rosetta 2(DE3) cells (Novagen). By this strategy, all chitinase variants were produced with identical N-terminal His<sub>10</sub> tags.

**Table 4.2** Primers used to introduce NdeI and XhoI restriction sites into ChiA and ChiB variants

Chitinase	Primer	Sequence
ChiA	Forward	5' TCGAAGGTCGTCATATGGCCGCGCCGGGC '3
	Reverse	5' CAGCCGGATCCTCGAGTTATTGAACGCCGGCGC '3
ChiB	Forward	5' TCGAAGGTCGTCATATGTCCACACGCAAAGCCGTT '3
	Reverse	5' AGCCGGATCCTCGAGTTACGCTACGCGGCCCA '3

#### 4.2.2.3 Protein expression and purification of single mutants

ChiA-F396A, ChiA-W167A, and ChiA-W275A genes were expressed in *E. coli* as described previously.<sup>185,189</sup> Periplasmic extracts were purified as described by Hamre et al.<sup>190</sup>

#### 4.2.2.4 Protein expression and purification of double, inactive mutants

For protein expression of ChiB-mutants, *E. coli* BL21Star (DE3) cells containing the appropriate plasmid were inoculated into 25 mL Luria broth (LB)-medium containing 115 µg / mL ampicillin (Amp) and grown at 37 °C and 200 rpm for 16 h. This culture was then used to inoculate 250 mL LB-Amp medium containing 115 µg / mL ampicillin to a starting OD<sub>600</sub> of 0.1. After growing this culture until the OD<sub>600</sub> reached 0.8-1.0, the temperature was decreased to 22°C, and gene expression was induced by adding isopropyl-β-D-thiogalactopyranoside to a final concentration of 1 mM, followed by cultivation for another 20 h. The cells were harvested by centrifugation (8000 rpm, 20 min at 4 °C). Periplasmic fractions were prepared by osmotic shocking as described elsewhere.<sup>191</sup> A cytoplasmic protein extraction was also performed by re-suspending the spheroplasts in lysis buffer (0.1 mg / mL lysozyme, 50 mM Tris-HCl, 50 mM NaCl, 4 mM MgCl<sub>2</sub>, 1 mM EDTA, 0.1 mM PMSF pH 8.0), followed by incubation at 37 °C for 30 min. Cell debris was removed by centrifugation (8000 rpm, 20 min at 4 °C), and the

resulting supernatant was used for further enzyme purification. Both the periplasmic and cytoplasmic extracts were sterilized by filtration (0.2  $\mu\text{m}$ ) and used immediately for protein purification. ChiA mutants were expressed as described previously.<sup>190</sup>

All proteins were purified by using Ni-NTA Agarose (Qiagen, Venlo, Netherlands) as described by Hamre et al.<sup>190</sup>

#### **4.2.3 Degradation of chitin for determination of enzyme processivity**

Hydrolysis of chitin was carried out as described previously.<sup>78</sup> The extent of degradation is defined as the percentage of number of moles of solubilized GlcNAc-units with respect to number of moles GlcNAc-units in solid form (chitin) used in the experiments. Processivity data have previously been obtained for ChiA wild-type (WT), ChiB-WT, and ChiB-W97A.<sup>78</sup>  $[(\text{GlcNAc})_2]/[\text{GlcNAc}]$  production ratios from initial degradation of  $\beta$ -chitin by ChiA-W167A, ChiA-W275A, and ChiA-F396A are given in the Appendix A2; however, the inability to extricate the effects of substrate positioning and ChiA endo-activity from the measurements make the ratios inconclusive as a measure of processive ability. Thus, discussion of these results is confined to the Appendix A2.

#### **4.2.4 High performance liquid chromatography (HPLC)**

Concentrations of mono- and disaccharides were determined using HPLC with a Rezex Fast fruit H<sup>+</sup> column (100 mm length and 7.8 mm inner diameter, Phenomenex). An 8  $\mu$ l sample was injected on the column, and the mono- and disaccharides were eluted isocratically at 1 mL/min with 5 mM H<sub>2</sub>SO<sub>4</sub> at 85 °C. The mono- and disaccharides were monitored by measuring absorbance at 210 nm, and the amounts were quantified by measuring peak areas. Peak areas were compared with peak areas obtained with standard samples with known concentrations of mono- and disaccharides.

#### **4.2.5 Isothermal titration calorimetry experiments**

ITC experiments were performed with a VP-ITC system from Microcal, Inc. (Northampton, MA, USA).<sup>139</sup> Solutions were thoroughly degassed prior to experiments to avoid air bubbles in the calorimeter. For experiments with ChiA-W275A and ChiA-F396A, 500  $\mu$ M of (GlcNAc)<sub>6</sub> was placed in the syringe, and 15  $\mu$ M of enzyme was placed in the reaction cell with a volume of 1.42 ml. For ChiA-W167A, 3 mM (GlcNAc)<sub>6</sub> and 15  $\mu$ M enzyme were used; for ChiB-W220A, 2 mM of (GlcNAc)<sub>6</sub> and 20  $\mu$ M of enzyme were used; for ChiB-F190A, 500  $\mu$ M of (GlcNAc)<sub>6</sub> and 20  $\mu$ M of enzyme were used; and for ChiB-W97A, 2 mM of (GlcNAc)<sub>6</sub> and 80  $\mu$ M of enzyme were used. Both enzyme and ligand were diluted in 20 mM potassium phosphate buffer pH 6.0. The heat of ionization of this buffer is 1.22 kcal/mol.<sup>140</sup> Typically, 40 to 60 injections of 4-8  $\mu$ l (GlcNAc)<sub>6</sub> were injected into the reaction cell at 180 s intervals, at 30 °C with a stirring speed of 260 rpm. Due to observed instability of ChiB-W97A in the ITC experiments at such relative high concentrations at 30 °C, experiments at 10, 12.5, 15, 17.5, and 20 °C were performed, and the values for 30 °C were extrapolated from the data obtained at

lower temperatures. At least two independent titrations were performed for each binding reaction. In all ITC experiments, catalytically inactivated double mutants were used; for simplicity, we refer to them by the aromatic mutation only.

#### 4.2.6 Analysis of calorimetric data

The shape of the ITC binding curve is determined by the Wiseman  $c$ -value as expressed in Equation 4.1,<sup>139</sup>

$$c = nK_a[M]_t \quad (4.1)$$

where  $n$  is the stoichiometry of the reaction,  $K_a$  is the equilibrium binding association constant, and  $[M]_t$  is the protein concentration. When the  $c$ -value is in the range of  $10 < c < 1000$ ,  $K_a$  can be determined from the Wiseman binding isotherm. When the  $c$ -value is in the range of  $0.01 < c < 10$ , the binding thermodynamics can be determined if a sufficient portion of the binding isotherm is used for analysis.<sup>192</sup> This can be achieved by ensuring a high molar ratio of ligand to protein at the end of the titration, accurate knowledge of the concentrations of both ligand and receptor, an adequate level of signal-to-noise in the data, and known stoichiometry.

ITC data were collected automatically using the Microcal Origin v.7.0 software accompanying the VP-ITC system.<sup>139</sup> Prior to further analysis, data were corrected for heat of dilution by subtracting the heat remaining after saturation of binding sites on the enzyme. Data were fit to a single-site binding model using a non-linear least-squares algorithm available in Origin. The binding reactions for ChiA-W275A, ChiA-F396A, ChiB-W97A, and ChiB-F190A yielded a  $c$ -value in the range of  $10 < c < 1000$ . Here, all data were well-represented by the single site binding model yielding stoichiometry ( $n$ ) of

1 and allowing for the determination of the equilibrium binding association constant ( $K_a$ ), and the reaction enthalpy change ( $\Delta H_r^\circ$ ) of the reaction. Binding to ChiA-W167A and ChiB-W220A yielded  $c$ -values  $< 10$  and  $n$  was set to be 1 before fitting of theoretical data to the experimental. The equilibrium binding dissociation constant ( $K_d$ ), reaction free energy change ( $\Delta G_r^\circ$ ) and the reaction entropy change ( $\Delta S_r^\circ$ ) were calculated as in Equation 4.2.

$$\Delta G_r^\circ = -RT \ln K_a = RT \ln K_d = \Delta H_r^\circ - T \Delta S_r^\circ \quad (4.2)$$

Errors are reported as standard deviations of at least two experiments at each temperature.

#### 4.2.7 Molecular dynamics simulations

Models of ChiA and ChiB were constructed from crystal structures of enzyme-ligand complexes obtained from the Protein Data Bank (PDB), PDB 1EHN for ChiA and PDBs 1E6N and 1OGG for ChiB.<sup>45, 79,144</sup> Hexameric chito-oligosaccharides were constructed such that they occupied subsites -4 to +2 in ChiA and -3 to +3 in ChiB. As ChiA is a reducing end-specific enzyme,<sup>82-83</sup> substrate bound in the ChiA active site is numbered -4 to +2, from cleft entrance to exit (Figure 4.1).<sup>193</sup> ChiB is a non-reducing end-specific enzyme in which the binding sites are numbered +3 to -3, from entrance to exit (Figure 4.1).<sup>81,134</sup> Hydrolysis occurs between the -1 and +1 site, with a glutamic acid acting as catalytic acid. For both ChiA and ChiB, substrate-binding induces a rotation of an aspartic acid in the -1 binding site from a position pointing away from the catalytic glutamate to a position where it forms a hydrogen bond with the catalytic glutamate and the *N*-acetyl group of the sugar bound in the -1 subsite. This re-orientation in the active site promotes nucleophilic attack by the *N*-acetyl oxygen on the anomeric carbon of the

pyranose in the -1 subsite.<sup>79,188,87-88,90</sup> Model construction included manual manipulation of the glutamic and aspartic acids and *N*-acetyl side chains from their initial crystal structure orientations, such that the models mimicked the catalytically competent Michaelis complexes. Additionally, the mutations in the PDB structures (Glu to Gln, leading to enzyme inactivation) were reverted to the wild type amino acid. A total of 10 MD simulations were constructed including: a ligand-free and ligand-bound version of wild-type ChiA and ChiB, ligand-bound simulations of ChiA variants W167A, W275A, and F396A, and ligand-bound simulations of ChiB variants W97A, W220A, and F190A. The ligand-free ChiA and ChiB simulations were constructed for the purposes of conducting free energy calculations only. The (GlcNAc)<sub>6</sub> ligand was bound as shown in Figure 4.1. Explicit details of the system construction, including the steps involved in manipulation of the catalytic residues and the construction of a hexameric ligand in the active sites of ChiA and ChiB, are provided in the Appendix A2.

After the systems were constructed, the enzymes and ligands were solvated with water and sodium ions were added using CHARMM.<sup>95</sup> The solvated ChiA and ChiB systems contained approximately 175,000 atoms and 52,000 atoms, respectively. ChiA was simulated with the fused *N*-terminal chitin-binding domain, as we observed the chitin-binding domain contributes to protein stability. ChiB does not suffer from this issue and was thus simulated without its chitin-binding domain for computational efficiency. Using CHARMM, the solvated systems were (1) minimized using an extensive step-wise procedure, (2) heated from 100 K to 300 K in 50 K increments over 20 ps, and (3) density equilibrated in the *NPT* ensemble for 0.1 ns.<sup>95</sup> After equilibration, the data collection MD simulations were performed using NAMD in the *NVT* ensemble

for 250 ns.<sup>96</sup> For all MD simulations and the free energy calculations, the CHARMM36 all atom force field with CMAP<sup>151-152</sup> corrections was used to model the enzyme. The CHARMM 36 carbohydrate force field was used to model the (GlcNAc)<sub>6</sub> ligands,<sup>194-195</sup> and water was modeled with the modified TIP3P force field.<sup>153-154</sup> VMD was used for visualization of trajectories and calculation of hydrogen bonding behavior.<sup>155</sup> Explicit details regarding simulation minimization procedure and parameters are given in the Appendix A2.

#### 4.2.8 Thermodynamic integration (TI)

Relative ligand binding free energies ( $\Delta\Delta G_{TI}$ ) were determined from TI calculations performed using NAMD with the dual-topology methodology.<sup>96,94, 196-197</sup> The theoretical background of TI dual-topology has been described earlier in Chapter 2.

The simulations were started from an equilibrated 25 ns snapshot obtained from the wild-type MD simulations. The TI methodology required further equilibration of the system, as a “hybrid” residue was introduced at the mutation site. The hybrid residue contained atoms from both the wild-type aromatic residue and the mutant alanine residue; the wild-type and mutant atoms did not interact with one another. The hybrid residue atoms interacted with the rest of the system via standard bonded and nonbonded interactions scaled by  $\lambda$  from the reactant (wild-type) to the product state (mutant) in windows over  $\lambda$ . Electrostatic and van der Waals interactions were calculated separately, and soft-core potentials were used to overcome endpoint singularities.<sup>110</sup> The electrostatic and van der Waals calculations included 15 windows ranging from  $\lambda$  values of 0 to 1 for a total of 30 simulations per mutation. The windows were divided so as to use more



closely coupled windows near the endpoints, improving accuracy. The electrostatics and van der Waals calculations were equilibrated for 0.5 ns before collecting 14.5 ns of TI data. The potential energy was calculated over the range of the coupling parameter. In prior studies, we have found this approach yielded sufficient sampling and well converged simulations.<sup>86, 183</sup>

Sufficient sampling was visually confirmed by plotting the  $dU/d\lambda$  histograms from each  $\lambda$  window. The autocorrelation time was determined for each window, which was then used as input for error analysis. Error analysis was performed following the method of Steinbrecher et al.<sup>111</sup>

### 4.3 Results

Ligand binding free energies obtained experimentally from ITC and computationally from TI are reported in Table 4.3. The ITC experiments yielded dissociation constants,  $K_d$ , from which the free energy of binding,  $\Delta G_{ITC}$ , was calculated as in Equation 4.2. Enthalpic and entropic contributions to the binding free energy are also reported (Table 4.3). The binding isotherms obtained from ITC are provided in Figure A2.1 of Appendix A2. For direct comparison with computational results, the ligand binding free energy relative to wild-type,  $\Delta\Delta G_{ITC}$ , was determined by subtracting the free energy of binding (GlcNAc)<sub>6</sub> to the wild-type from the free energy of binding (GlcNAc)<sub>6</sub> to the variant. Errors were propagated using standard rules for combining errors. The calculated relative ligand binding free energy,  $\Delta\Delta G_{TI}$ , was determined from the sum of the electrostatic and van der Waals components, as described in the Computational Methods. The individual components are reported in Table A2.1 of Appendix A2.

The binding free energies determined from ITC,  $\Delta G_{ITC}$ , reveal that binding to wild-type ChiA involves nearly equal enthalpic and entropic contributions.<sup>119</sup> Deletion of Trp-167 results in significant reduction in entropic favorability, without a concomitant increase in enthalpy. Deletion of Trp-275 also reduces entropic favorability, but the enthalpic term increases considerably. In the case of the Phe-396 mutation, entropic and enthalpic contributions are approximately the same as in the wild-type and the overall effect on the  $K_d$  is only a factor two. Binding to ChiB wild-type and the ChiB-F190A variant is almost entirely due to entropic contributions. Upon mutation of ChiB Trp-97 and Trp-220, some entropic favorability is lost with marginal gains in enthalpic contributions; the loss in entropic favorability is clearly greater for the W220A mutation. Notably, mutations with the most drastic effects on binding (GlcNAc)<sub>6</sub> are ChiA-W167A (-3 subsite) and ChiB-W220A (+2 subsite).

**Table 4.3** Absolute and relative free energies of binding (GlcNAc)<sub>6</sub> to ChiA and ChiB wild-type and variants. The dissociation constants and free energies denoted with the “ITC” subscript were obtained experimentally from ITC with catalytically inactivated enzymes. The calculated relative ligand binding free energies are denoted with the “TI” subscript. All ITC data refer to experiments at  $t = 30$  °C and pH = 6.0; see Methods section (Section 4.2) for more details.

Chitinase	Mutation	Binding site	Isothermal Titration Calorimetry					Thermodynamic Integration
			$K_d^a$	$\Delta G_{ITC}^{\circ b}$	$\Delta H_T^{\circ b}$	$-T\Delta S_T^{\circ b}$	$\Delta\Delta G_{ITC}^b$	$\Delta\Delta G_{TI}^b$
ChiA	WT <sup>c</sup>	-----	0.56 ± 0.03	-8.7 ± 0.1	-4.5 ± 0.2	-4.2 ± 0.2	-----	-----
	W167A	-3	134 ± 13	-5.4 ± 0.1	-4.5 ± 0.3	-0.9 ± 0.2	3.3 ± 0.1	5.0 ± 0.4
	W275A	+1	2.1 ± 0.3	-7.9 ± 0.1	-6.7 ± 0.7	-1.2 ± 0.6	0.8 ± 0.1	3.6 ± 0.4
	F396A	+2	1.2 ± 0.1	-8.2 ± 0.1	-3.3 ± 0.1	-4.9 ± 0.1	0.5 ± 0.1	-0.2 ± 0.2
ChiB	WT <sup>c</sup>	-----	0.20 ± 0.03	-9.3 ± 0.1	-0.1 ± 0.3	-9.2 ± 0.3	-----	-----
	W97A	+1	0.87 ± 0.14 <sup>d</sup>	-8.4 ± 0.1	-1.2 ± 0.1	-7.2 ± 0.2	0.9 ± 0.1	2.8 ± 0.4
	W220A	+2	44 ± 2	-6.0 ± 0.1	-3.0 ± 0.2	-3.0 ± 0.2	2.7 ± 0.1	3.5 ± 0.3
	F190A	+3	0.55 ± 0.12	-8.7 ± 0.2	-0.6 ± 0.1	-8.1 ± 0.2	0.6 ± 0.1	-0.3 ± 0.2

<sup>a</sup> μM, <sup>b</sup> kcal/mol, <sup>c</sup> interpolated from values above and below  $t = 30$  °C as described by Norberg et al.<sup>141</sup>, <sup>d</sup> extrapolated from values below  $t = 30$  °C, <sup>e</sup> Wild-type data originally appeared in Hamre et al.<sup>119</sup>

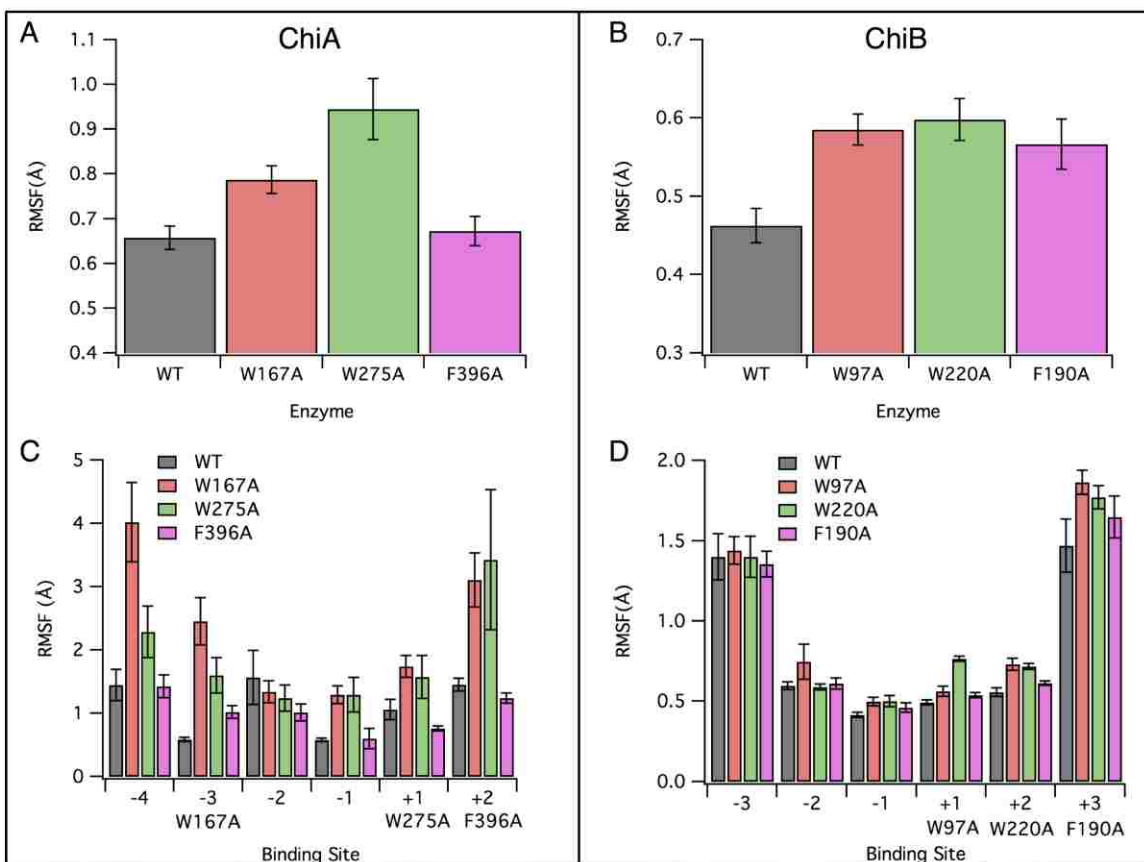
The calculated relative binding free energies ( $\Delta\Delta G_{TI}$ ) for ChiA-W167A (-3) and ChiA-W275A (+1) are quite unfavorable, suggesting aromatic residues in these positions play important roles in tight binding. The free energy change associated with mutating Phe-396 (+2), on the product side of the ChiA cleft, is relatively inconsequential. It is thus unlikely this residue has much direct influence in ligand binding. The calculated relative ligand binding free energies for the ChiB-W97A (+1) and ChiB-W220A (+2) mutants, are also clearly unfavorable, indicating a preference for aromatic residues in these positions. The change in binding free energy associated with mutating Phe-190 (+3) of ChiB to alanine is negligible by comparison to the effects of the W97A and W220A mutations. The residues mutated in this study include two pairs of residues that are

structurally equivalent (Figure 4.1): Trp-97 and Trp-220 in ChiB align with Trp-275 and Phe-396 in ChiA, respectively. Interestingly, both experiment and calculations indicate similarity in the mutational effects of W97A in ChiB and W275A in ChiA (subsite +1); on the other hand, the calculated and observed effects of W220A in ChiB differ strongly from the effects of F396A in ChiA (subsite +2), reflecting differences in substrate directionality.

Direct comparison of  $\Delta\Delta G_{TI}$  and  $\Delta\Delta G_{ITC}$  reveals that calculations and experiments are in reasonable agreement for nearly all variants, except the +1 binding site variants, ChiA-W275A and ChiB-W97A. The calculated and experimental  $\Delta\Delta G$  of these two variants differ by  $\sim 2$  kcal/mol or more. As we will discuss below, we expect this is neither coincidental nor reflective of inaccuracy in either technique, but rather, a result of the uncertainty involved in experimental determination of ligand binding free energies. In other words, experimentally, we cannot be certain that the ligand binds in precisely the same binding sites as those investigated computationally. Moreover, these +1 mutations, combined with the catalytic Glu mutations, may have direct effects on the catalytic center that would not be captured in a computational assessment of the Trp point mutation. This illustrates the importance of using computational methods to probe the molecular details of ligand binding in GHs.

To understand the role of aromatic residues in the catalytic clefts of processive chitinases, we performed 250-ns MD simulations of wild-type ChiA and ChiB and the ChiA and ChiB aromatic to alanine variants bound with hexameric chito-oligosaccharides. From these simulations, we have calculated the root mean square fluctuation (RMSF) of four key catalytic residues (Figure 4.2A & B) and the RMSF of

the ligand on a subsite basis (Figure 4.2C & D). To measure RMSF, the time-averaged position of a residue or group of residues was computed as an initial point of reference. Variance from this time-averaged position was then determined over the course of the 250-ns simulations. The RMSF of a given residue or selection of residues was then determined from the square root of the variance, physically corresponding to the average fluctuation from the time averaged position. Associated error was determined by dividing the 250-ns simulation into 100 time windows and determining the RMSF for each 2.5 ns window; the standard deviation was then determined from this set of RMSF values. These dynamic measures provide additional insights as to how point mutations affect stabilization of the ligand in the clefts and potentially how this affects processivity.



**Figure 4.2** Dynamics of the catalytic center and the bound ligand in variants of ChiA and ChiB. All simulation analyses were obtained from (GlcNAc)<sub>6</sub> ligand-bound 250-ns MD simulations. (A) Root mean square fluctuation (RMSF) of the catalytic tetrad (Asp-313, Glu-315, Tyr-390, and Asp-391) in wild-type ChiA (WT) and the three aromatic mutants. (B) Root mean square fluctuation (RMSF) of the catalytic tetrad (Asp-142, Glu-144, Tyr-214, and Asp-215) in ChiB wild-type and variants. (C and D) RMSF of the (GlcNAc)<sub>6</sub> ligand across the binding sites of (C) ChiA wild-type and aromatic variants and (D) ChiB wild-type and aromatic variants. Labels below the *x*-axis of panels C and D indicate subsites and the positions of the mutated aromatic residues relative to the catalytic center. Error bars have been obtained through 2.5 ns block averaging.

In both ChiA and ChiB, four residues have been implicated as direct participants in the substrate-assisted catalytic mechanism.<sup>79,87-88,90</sup> In ChiA, this tetrad includes Asp-313, Glu-315, Tyr-390, and Asp-391.<sup>87,90</sup> The ChiB catalytic tetrad includes Asp-142, Glu-144, Tyr-214, and Asp-215.<sup>188, 198-199</sup> On the basis of a comparative analysis of chitinase structures and MD simulations, we have previously suggested that the flexibility of the catalytic center, as measured by the RMSF of the catalytic tetrad, may be associated with reduced substrate chain association ( $k_{on}$ ) and likely reduced processivity.<sup>73</sup> Thus, we examined this measure of active site dynamics and compared it with both binding free energy and literature assessments of processive ability, where available. In general, the RMSF values for the mutant chitinases are higher than those for wild-type ChiA and ChiB (Figures 4.2A and 4.2B). The exception to this is the F396A mutant of ChiA variant, whose catalytic center is unaffected by the mutation. We have previously shown this residue has little impact on hydrolysis of crystalline chitin despite its location in the active site (Figure 4.1) and only moderate impact on processive ability on chitosan.<sup>70</sup>

The RMSF of carbohydrate ligands bound in the active sites of GHs has also been linked to processive function. To understand the relationship of this measure of active site dynamics with binding free energy and processive ability, we have determined the RMSF of the (GlcNAc)<sub>6</sub> ligands of wild-type ChiA and ChiB, as well as the six variants over the course of the 250-ns MD simulations. As with the RMSF of catalytic residues, the RMSF of the (GlcNAc)<sub>6</sub> ligand is a measure of the degree to which the ligand fluctuates about an average position; higher RMSF values correspond to ligands with a higher degree of freedom. Compared to wild-type ChiA, the ligands bound to both the W167A and

W275A variants fluctuate significantly more, across the length of the binding cleft, whereas the F396A mutation does not affect ligand flexibility (Figure 4.2C). In ChiB, the effect of mutations on (GlcNAc)<sub>6</sub> flexibility is more modest and localized, primarily affecting substrate binding sites +1, +2, and +3; in these three sites, the flexibility of (GlcNAc)<sub>6</sub> is slightly higher for all three mutants (Figure 4.2D).

## **4.4 Discussions**

### **4.4.1 Absolute and relative binding free energies**

The reduction in the entropic favorability of substrate binding upon mutating Trp-167 and Trp-275 in ChiA and Trp-97 and Trp-220 in ChiB appears to be attributable to unfavorable effects on solvation entropies. The deletion of large, hydrophobic aromatic residues removes unfavorable solvation in the apo-enzyme that would normally be alleviated by substrate-binding. Furthermore, these deletions create cavities in the enzyme active site that may accommodate entropically unfavorable water molecules in the enzyme-ligand complexes. Finally, small changes in the substrate-binding mode and increased substrate flexibility may lead to more water molecules being restrained in the enzyme-substrate-complexes. We have previously shown that desolvation of the wild-type ChiA active site upon ligand binding greatly affects solvation entropy.<sup>119</sup> For ChiA, the MD simulations of enzyme-ligand complexes indeed show higher active site solvation for the W167A and W275A mutants (Figure A2.2 of Appendix A2). For the F396A mutant, solvation is similar to wild-type across all six binding sites, supporting the notion that Phe-396 does not play a strong role in binding of (GlcNAc)<sub>6</sub>. On the other hand, we did not observe significant differences in active site solvation for the ChiB



variants, which is in line with previous observations that solvation plays a more dominant role in ligand binding in ChiA than ChiB.<sup>119</sup> In ChiB, the loss in entropic favorability must be attributed to other factors such as the conformational entropy of the protein; this is not entirely unexpected, as studies examining the thermodynamic signature of ChiB binding an allosamidin inhibitor show large, positive conformational entropy changes.<sup>142,119</sup>

The experimental and computational determinations of relative binding free energies (Table 4.3) are in good agreement with each other for the two mutations with the most drastic effects on the  $K_d$  (ChiA-W167A in the -3 subsite, and ChiB-W220A in the +2 subsite) and for the two mutations that, by all criteria, seem to have little effect on binding of (GlcNAc)<sub>6</sub>, ChiA-F396A and ChiB-F190A. Larger discrepancies are observed for the mutations in the +1 subsite, ChiB-W97A, and even more so with ChiA-W275A. A number of factors may contribute to such discrepancies, which we believe have more to do with variations in binding mode and the conformation of enzyme and substrate than errors in either method. As with any computational assessment of binding free energy, the starting configuration of the bound ligand relies heavily on structural input from crystallographic studies or docking calculations. Here, we have initiated simulations using crystallographic structures representative of thermodynamically stable configurations of the ligand within the ChiA and ChiB active sites.<sup>45, 79,144</sup> The enzymes in these crystallographic structures carry the single point mutation of the catalytic acid, such that a non-hydrolyzed ligand could be observed across the length of the cleft. In each case, a distorted -1 pyranose conformation was captured in the crystallographic structure. However, the catalytic residues' side chains did not reflect the presumed

catalytically active distorted Michaelis complex. To construct the simulations, these side chains were rotated so as to represent the proposed Michaelis complex. Accordingly, free energy calculations representing the fully-formed Michaelis complex may incorporate an additional free energy penalty not captured in ITC experiments. This penalty can, in fact, be quite large, as the free energy of forming the complete distorted complex, with the -1 sugar in a  ${}^1,4B$  conformation, can be up to 8 kcal/mol.<sup>161</sup> Thus, small variations in ligand conformation may translate to considerable deviations in the calculated and experimental  $\Delta\Delta G$  values.

Similarly, it is noteworthy that the largest deviations between  $\Delta\Delta G_{ITC}$  and  $\Delta\Delta G_{TI}$  were observed for mutation of the conserved tryptophan residues in the +1 subsites, ChiA-W275 and ChiB-W97A. Both of these tryptophan residues are at the heart of the enzyme binding clefts, immediately adjacent to the catalytic tetrads. In ChiA, Trp-275 sits only 3.2 Å from the catalytic Glu-315, and Trp-97 of ChiB is only 3.4 Å from the catalytic Glu-144. It is thus quite conceivable that mutation of both the tryptophan and glutamate affects substrate-binding interactions beyond the mere reduction of aromatic stacking interactions, which could help explain the observed discrepancies in computational and experimental assessments of  $\Delta\Delta G$  in these two variants.

Additionally, ChiA and ChiB have multiple binding subsites extending beyond those examined here. Kinetic studies have shown both ChiA and ChiB are capable of binding shorter oligomers, i.e., (GlcNAc)<sub>5</sub> and (GlcNAc)<sub>6</sub>, in various locations along the cleft.<sup>81,145</sup> These prior studies suggest that ChiA can bind (GlcNAc)<sub>5</sub> in either the -3 to +2 binding sites or the -2 to +3 binding sites; one study suggests the two occupancies are equal favorability and a second suggests the -2 to +3 occupancy is favored 70% of the

time.<sup>145,190</sup> ChiB appears to favor binding (GlcNAc)<sub>6</sub> in the -2 to +4 binding sites over the -3 to +3 binding sites by 80%.<sup>81</sup> These kinetic studies were conducted on wild-type ChiA and ChiB, and it is unknown what effect mutation of active site residues has on substrate-binding preferences. Assuming that these preferences do not change, only a percentage of binding events observed by ITC measurements will exactly match the computationally analyzed binding events. Due to lack of structural information, our computational investigation for ChiB considered the presumably less favorable binding mode for (GlcNAc)<sub>6</sub>, as a structure with a resolved +4 pyranose in the ChiB active site does not yet exist. All in all, the ITC values more likely represent a mean of ligand binding conformations, while the computational results examine a single, catalytically active conformation. Thus, for the purposes of examining roles of aromatic residues in the chitinase active sites, we primarily consider the computational relative binding free energy results, as we can control both the position of the ligand across the active site as well as ensure the Michaelis complex is fully formed at the start of the thermodynamic assessment.

#### **4.4.2 ChiA aromatic residues**

Aromatic residues Trp-167 and Trp-275 of ChiA appear to play significant roles in ligand binding (Table 4.3). This is illustrated dynamically by both the enhanced flexibility of the ChiA-W167A and ChiA-W275A active sites and the flexibility of the (GlcNAc)<sub>6</sub> ligands across the substrate-binding clefts of the mutants (Figures 4.2A and 4.2C). Trp-167 is largely responsible for maintaining the shape of the substrate-side of the ChiA cleft, since the (GlcNAc)<sub>6</sub> ligand face opposite Trp-167 is quite solvent exposed (Figure 4.3A). Replacement of Trp-167 with alanine greatly reduces ligand contact with

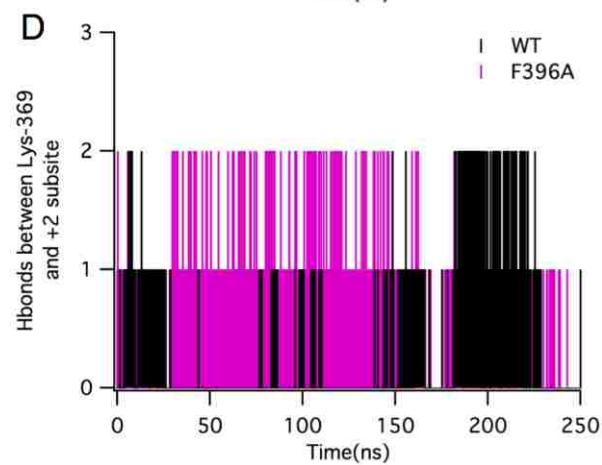
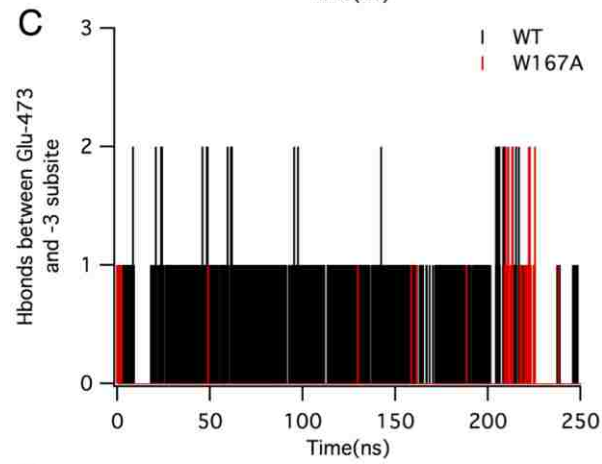
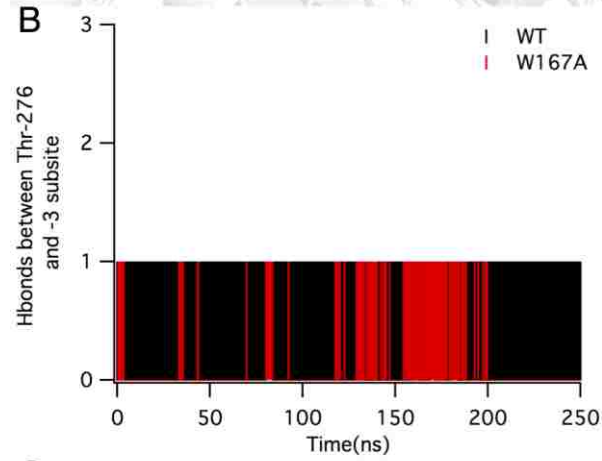
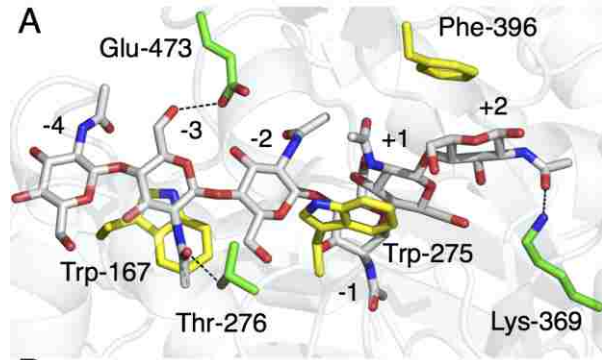
the protein, and the normally strained ligand relaxes into the newly opened space to maintain protein interactions. As a result, the ligand has a significantly greater degree of freedom, as observed in the MD simulations (Figure 4.2).

Figure 4.3A shows that Thr-276 and Glu-473 form hydrogen bonds with the carbonyl and hydroxyl oxygens of the sugar in the -3 subsite (PDB 1EHN). We suggest that the stacking interaction of Trp-167 with the -3 sugar promotes these additional favorable interactions, as the MD simulations show that the W167A mutation significantly reduces hydrogen bond formation between the -3 sugar and Thr-276 and Glu-473 (Figure 4.3B and C). Thus, Trp-167 is responsible for maintaining the -3 protein-carbohydrate interactions by both direct stacking and facilitating polar residue hydrogen bonding. Interestingly, W167A is known to induce transglycosylation in ChiA in lieu of hydrolysis.<sup>90,200</sup> We hypothesize the added flexibility and reduced affinity may be responsible, in part, for the observed transglycosylation activity.

Trp-275 and Phe-396 are the only two aromatic residues in the product side of the ChiA binding cleft with the ability to form stacking interactions with (GlcNAc)<sub>6</sub>. The significant reduction in affinity by mutation to alanine indicates the binding platform created by the bulk of the Trp-275 hydrophobic side chain plays an important role in ligand binding. As with W167A, removal of Trp-275 destabilizes the ligand across the length of the active site and leads to increased active site fluctuations. In addition to overall ligand stability, Trp-275 occupies an important position in the +1 product binding site. The large change in affinity and destabilization of the ligand implicates Trp-275 in product stability. After hydrolysis of the ligand, we hypothesize that the favorability of Trp-275 for the dimeric product may even contribute to inhibition. ChiA is moderately

inhibited by its products,<sup>159</sup> and mutation of Trp-275 to a slightly less favorable aromatic amino acid, such as phenylalanine, might alleviate inhibition, while maintaining overall ligand binding affinity.

The data suggest that the product side phenylalanine, Phe-396, in subsite +2, has little or no role in ligand binding. Aronson et al. previously suggested that mutation of Phe-396 would initiate a local shift of the entire ligand towards the non-reducing end to make up for lost interactions.<sup>193</sup> However, we find that nearby polar residue Lys-369 maintains ligand interactions when Phe-396 is mutated, as illustrated by hydrogen bond formation over the course of the MD simulations (Figure 4.3D). Compensating hydrogen bonds could explain why both experimental and computational analyses show that mutating Phe-396 to alanine has only minor consequences, despite what appears to be an important hydrophobic stacking interaction between Phe-396 and the +2 pyranose ring (Figure 4.3A). It is noteworthy that it has been experimentally shown that mutation of Phe-396 to Trp most likely increases substrate affinity.<sup>200</sup>



**Figure 4.3** (A) Crystal structure of ChiA-E315Q in complex with chito-octaose (1EHN; only six sugar moieties are shown), highlighting hydrogen bonding of Thr-276 and Glu-473 with the -3 site pyranose and of Lys-369 with the +2 pyranose. Panels B through D illustrate hydrogen bonding (Hbond) over the course of 250-ns MD simulations between: (B) Thr-276 and the -3 pyranose in the wild-type (WT) and the W167A mutant; (C) Glu-473 and the -3 pyranose in WT and W167A; (D) Lys-369 and the +2 pyranose in WT and F396A. In each panel, WT hydrogen bonds with the ligand are illustrated in black. The discrete value of a hydrogen bond is shown as a line drawn from either 1 or 2 to the origin. A distance cutoff of 3.4 Å and a 60° angle from linear were used to determine formation of hydrogen bonds, considering only those formed with polar atoms (N and O).

Finally, both changes in binding free energy ( $\Delta\Delta G_{TI}$ , Table 4.3) and dynamic measurements of fluctuation from MD simulation (Figure 4.2A & C) correlate with what we currently know about the effect of these aromatic mutations on processivity; specifically, large reductions in binding affinity in increased active site flexibility correspond to reduced processive ability. Kurašin et al. recently reported that ChiA-W167A and ChiA-W275A significantly impact the ability of the enzymes to processively degrade crystalline  $\alpha$ -chitin, reducing apparent processivity by more than half.<sup>71</sup> We find that W167A and W275A mutations markedly reduce binding affinity, in qualitative agreement with our recent hypothesis that ligand binding free energy and processive ability are connected through transition state theory and probability of dissociation.<sup>61</sup> Similarly, the RMSF of the catalytic tetrad and the RMSF of the ligand across the ChiA-W167A and ChiA-W275A active sites fluctuate more than that of wild-type ChiA, which is almost certainly related to the reduction in binding affinity. A comparable

determination of ChiA-F396A processive ability on a crystalline substrate does not yet exist, yet we do know that ChiA-F396A has only a moderate effect on processivity of the soluble chitosan substrate and practically no effect on hydrolysis of crystalline  $\beta$ -chitin,<sup>70</sup> which suggests that the F396A mutation will have only marginal effects on processive degradation of crystalline chitin. The minimal impact of F396A on both binding affinity and enzyme dynamics corresponds with this extrapolation of processive ability. Overall, the two tryptophans of ChiA appear to play more significant roles in both binding affinity and processive ability than the product-side phenylalanine.

#### **4.4.3 ChiB aromatic residues**

At first glance, each of the ChiB aromatic residues investigated here, Trp-97, Trp-220, and Phe-190, appear responsible for formation of a hydrophobic stacking platform that would interact with the polymeric substrate during processive motion. Structurally, the large side chains nicely interact with the +1, +2, and +3 substrate-side pyranose rings (Figure 4.1).<sup>79</sup> However, the data presented above indicate that only Trp-97 and Trp-220 are key to tight binding. These two residues are closer to the catalytic center, at the +1 and +2 binding sites, respectively, and mutation to alanine reduces ligand affinity by nearly 3 kcal/mol (Table 4.3). Mutation of these residues yields minor destabilization of the ligand, primarily affecting stability in the substrate binding sites (Figure 4.2D); this difference, relative to ChiA, may be due to the fact that ChiB has a more confined (“tunnel-like”) substrate-binding cleft than ChiA.<sup>45, 79</sup> The destabilization also manifests in flexibility of the catalytic site residues (Figure 4.2B). Horn et al. previously reported that the W97A and W220A mutations reduce processivity, and the current findings thus correspond with our prior hypothesis that increased flexibility of the catalytic residues



correlates with decreased processivity.<sup>69,73</sup> The presumed role of Trp-97 and Trp-220 in ChiB is to tightly bind with the substrate, enabling the enzyme to maintain its association with the substrate throughout processive action. Several studies have concluded that binding affinity, in particular in substrate-binding sites, is important for processivity.<sup>83</sup>

Despite what appears to be an important stacking interaction according to structural studies, Phe-190 seems to play a minor role in ligand binding. Both experimental and computational measures of the binding free energy change associated with mutating Phe-190 to alanine are low (Table 4.3). Unexpectedly, the F190A mutation in ChiB resulted in increased RMSF of the catalytic residues (Figure 4.2B), which corresponds to conformational changes in the Asp-142 and Glu-144 side chains. In wild-type ChiB, Glu-144 initially forms a hydrogen bond with both Asp-142 and the glycosidic oxygen linking the +1/-1 sugars, but Glu-144 quickly rotates such that only the Asp-142 is maintained (Figure A2.3 of Appendix A2). In ChiB-F190A, Glu-144 makes this same conformational change, but the transition takes approximately twice as long, which while not likely statistically significant, appears as increased fluctuation in the RMSF of the ChiB-F190A catalytic residues. After the conformational change, the ChiB-F190A active site mimics that of wild-type.

The RMSF of the ligand in ChiB-F190A was the same as for the wild-type (Figure 4.2D), as would be expected considering the mutation's minimal effect of binding affinity. Indeed, Katouno et al. observed that mutation of Phe-190 has little impact on chitin binding and chitin hydrolysis, which implies processive ability is similar to wild-type.<sup>201</sup> Interestingly, a sequence alignment with closely related GH18 chitinases (Figure A2.4 of Appendix A2) shows that Phe-190 is not conserved, and that polar or charged

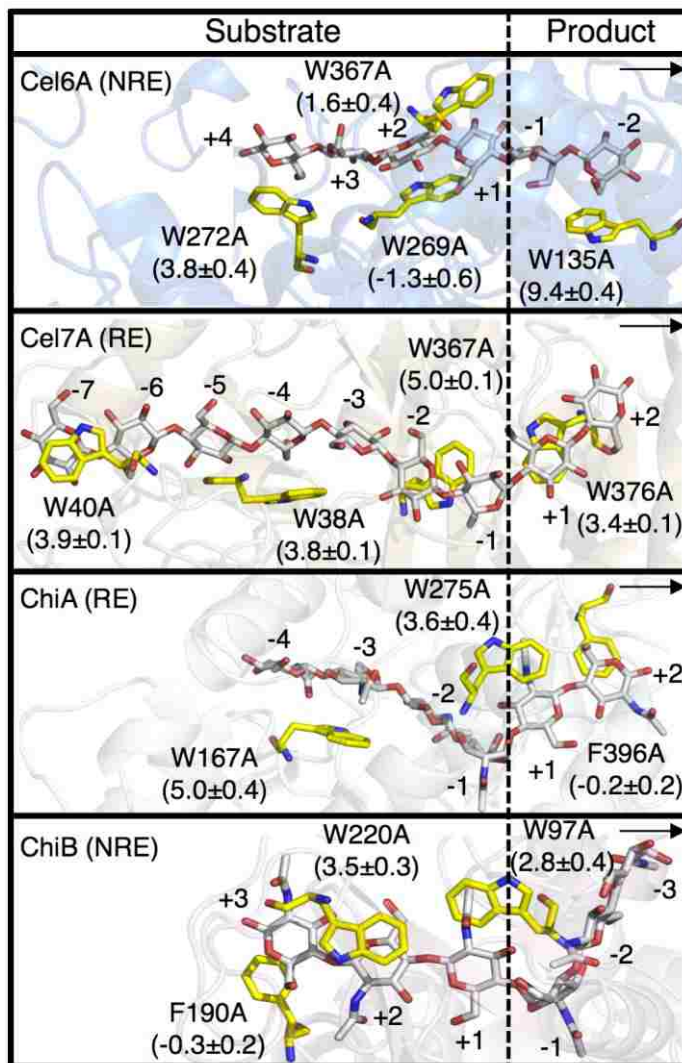
residues (Arg, Thr, Ser, or Glu) frequently appear in this position. In contrast, Trp-97 and Trp-220 are highly conserved.

#### 4.4.4 Trends across GH Families

Here, we find that Phe-190 at the ChiB substrate entrance (+3 subsite) and Phe-396 at the ChiA product exit (+2 subsite) have the least direct involvement in ligand binding interactions and binding affinity. These observations differ from observations made for analogous cellulases (Figure 4.4). Mutation of Trp-272 in the +3 subsite of non-reducing end specific Cel6A from *T. reesei*, a ChiB analogue, yielded a 3.8 kcal/mol change in  $\Delta\Delta G$ ,<sup>86</sup> and experimental data have shown that this mutation reduces activity on insoluble cellulosic substrates.<sup>202</sup> The cellulase analogue of ChiA, Cel7A from *T. reesei*, does not have an aromatic residue in the +2 product subsite. Interestingly, the relative binding free energies of mutating tryptophans in the +1 (product) subsites of these two reducing end specific processive GHs (i.e., W275A in ChiA and W367A in Cel7A) are approximately equal (Figure 4.4).<sup>183</sup> The product sites of Cel6A contain a tryptophan in the -2 subsite (Trp-135), which may be considered functionally analogous to Phe-396 in the +2 subsite of ChiA. Mutation of Trp-135 results in a remarkably large, unfavorable change in binding affinity amounting to  $9.4 \pm 0.4$  kcal/mole, suggesting this residue plays a critical role in product stabilization.<sup>86</sup>

Considering the data presented here, the above discussion, and the literature data summarized in Figure 4.4, one general conclusion appears to be that tryptophans are generally much more important for carbohydrate substrate interactions in processive GHs than phenylalanines. Furthermore, the largest impacts of mutation, judged on the basis of both computational and experimental observations, concern one tryptophan residue in

each enzyme that defines the “entrance” of the deep active site clefts, namely Trp-167 in ChiA and Trp-220 in ChiB. These key tryptophans are important for crystalline chitin processivity and their mutation causes weakening of ligand binding and activity on chitin.<sup>69-70</sup> These ChiA and ChiB tryptophans are analogous to the substrate acquisition tryptophans in processive cellulases *T. reesei* Cel7A and Cel6A, Trp-40 and Trp-272, respectively. As with ChiA and ChiB, mutation of the substrate entrance tryptophans drastically decreases the enzymes’ ability to turnover crystalline cellulose substrates. Taken together, it is tempting to speculate that a substrate-side tryptophan at or near the entrance of the active site is an identifying feature of a processive GH.



**Figure 4.4** Comparison of reported relative binding free energies,  $\Delta\Delta G$ , for aromatic to alanine mutations in selected family 6, 7, and 18 GHs: *T. reesei* Cel6A and Cel7A, and *S. marcescens* ChiA and ChiB. ‘RE’ and ‘NRE’ are used to label reducing end and non-reducing end specific processive enzymes. The arrows indicate the direction of substrate sliding through the active sites. The free energies shown in the figure below each mutation are in kcal/mol. The free energy values shown have all been obtained using the same computational method described here. Of the cellulase variants illustrated,

experimental data is available only for *T. reesei* Cel7A W38A ( $\Delta\Delta G$ : 0.8 kcal/mol for the intact protein and 0.4 kcal/mol for the catalytic domain only).<sup>72</sup>

Finally, binding free energy alone is not sufficient to entirely describe aromatic residue function. A great example of this is the residues at the catalytic center in *T. reesei* Cel6A, Trp-269 and Trp-376.<sup>86</sup> Each of these residues had relatively little impact on ligand binding affinity upon mutation, but their mutation was detrimental to hydrolytic function, as illustrated in a homologous Family 6 GH.<sup>203</sup> MD simulations have revealed roles of these residues in physically maintaining the -1 ring distortion through steric hindrance. Thus, generalization of aromatic residue function from binding free energies alone only partially describes a given residue. MD simulation coupled with free energy calculations or experimental data enables the description of additional factors that help define the roles of aromatic residues. With these approaches, we may capture dominant effects, including hydrophobic stacking interactions, active site architectures and dynamics, substrate conformational variation, and the flexibility of tunnel forming loops.<sup>83,204</sup> Further detailed evaluations of GH active sites, such as in this study, will provide the necessary information to catalogue variations in behavior and enable generalization of function in GH active sites.

#### **4.5 Conclusions**

Using experimental binding free energy techniques alongside MD simulations and calculated binding free energies, we have explored aromatic-mediated carbohydrate interactions in Family 18 *S. marcescens* processive chitinases ChiA and ChiB, ascribing roles to each of the aromatic residues directly interacting with a bound (GlcNAc)<sub>6</sub> ligand.

We find that ChiA aromatic residues Trp-167 and Trp-275 contribute greatly to ligand binding and stability of the active site by virtue of their location near the catalytic center. The ChiA residue at the exterior product-binding site, Phe-396, appears to have little to no role in ligand binding. ChiB residues near the catalytic center, Trp-97 and Trp-220, play critical roles in stabilizing ligand binding, while Phe-190 at the substrate entrance plays a minimal role. Generalization across GH families is difficult, in part due to the relatively limited amount of available data explicitly describing aromatic residue function. However, one general trend stands out; processive chitinases and cellulases seem to have a crucial tryptophan residue at what may be considered the “entrance” of their active site cleft and tunnels. The position of this residue relative to the catalytic center (i.e. which subsite it is part of) may vary, but in all cases, the residue makes a major contribution to substrate binding and recruitment. Similar studies across a variety of GH families will ultimately enable generalization of aromatic residue function in carbohydrate-active enzymes. With the abundance of carbohydrate active enzymes across all kingdoms of life, such a generalization will greatly benefit a broad variety of scientific fields.

## **Chapter 5 - Processivity, Substrate Positioning, and Binding: The Role of Polar Residues in a Family 18 Glycoside Hydrolase**

This chapter has been reprinted with permission from Hamre et al.,<sup>190</sup> Copyright © 2015, American Chemical Society. The experimental binding free energy, activity, and the processivity for the wild-type and the polar variants of ChiA, reported here (Chapter 5), were measured by our experimental collaborators at the Norwegian University of Life Sciences. The author of this dissertation performed MD simulations to provide insights into the polar amino acid interaction with the ligand in the active site of ChiA.

### **5.0 Abstract**

The enzymatic degradation of recalcitrant polysaccharides such as cellulose (Glc) and chitin (GlcNAc) by GHs is of significant biological and economical importance. In nature, depolymerization is primarily accomplished by processive GHs, which remain attached to the substrate in between subsequent hydrolytic reactions. Recent computational efforts have suggested that the processive ability of a GH is directly linked to ligand binding free energy. The contribution of individual aromatic residues in the active site of these enzymes has been extensively studied. In the present study, we offer the first experimental evidence confirming correlation of binding free energy and degree of processivity and that polar residues are essential for maintaining processive ability. Exchanging Thr<sup>276</sup> with Ala in substrate binding subsite -2 in the processive ChiA of *Serratia marcescens* results in a reduction of both enthalpy (2.6 and 3.8 kcal/mol) and free energy (0.5 kcal/mol and 2.2 kcal/mol) for the binding to the substrate (GlcNAc)<sub>6</sub> and the inhibitor allosamidin, respectively, compared to that of the wild type. Moreover,

initial apparent processivity as measured by  $[(\text{GlcNAc})_2]/[\text{GlcNAc}]$  ratios ( $17.1 \pm 0.4$  vs.  $30.1 \pm 1.5$ ) and chitin degradation efficiency (20 % vs. 75 %) are greatly reduced for ChiA-T276A vs. the wild type. Mutation of Arg<sup>172</sup> to Ala, reduces the recognition and positioning of the substrate into the active site. MD simulations indicate ChiA-R172A behaves similarly to the wild type, but dynamics of ChiA-T276A are greatly influenced by mutation, which is reflective of their influence on processivity.

## 5.1 Introduction

Chitin and cellulose are insoluble, linear homopolymers consisting of  $\beta$ -1,4 linked *N*-acetyl-glucosamine (GlcNAc; A-unit) and  $\beta$ -1,4 linked glucose units, respectively. The individual sugar moieties are rotated 180° relative to each other, yielding a disaccharide structural unit.<sup>39</sup> Depolymerization of chitin and cellulose to soluble, dimeric units is typically accomplished through the action of GHs known as cellulases and chitinases, respectively. Chitin and cellulose polysaccharides are the two most abundant biopolymers in nature with an annual production amounting to 100 billion and one trillion tons, respectively.<sup>2, 205</sup> With such massive availability, these polysaccharides represent a nearly unlimited source of raw material for the production of fuels and specialty chemicals through enzymatic approaches. However, efficient enzymatic degradation of these materials is critical to the economic viability of any commercial conversion process. Accordingly, engineering enzymes for new and more efficient conversion requires development of fundamental knowledge of both catalytic mechanisms and the means by which the enzyme interacts with the polysaccharide substrate.



GHs catalyze the hydrolysis of *O*-glycosidic bonds between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety (www.cazy.org<sup>52</sup>). In general, hydrolysis occurs via acid catalysis that requires two critical residues: one proton donor and a nucleophile/base.<sup>206</sup> Each enzyme has a customized mode of action, either by random cleavage of polymer chains (endo-acting enzymes) or by preferential cleavage of the reducing or non-reducing ends of chains (exo-acting enzymes). Both endo- and exo- mechanisms can be combined with processive action, meaning that the enzymes bind individual polymer chains in long tunnels or deep clefts and hydrolyze a series of glycosidic linkages along the same chain before dissociation.<sup>51</sup> This range of potential functions has historically made characterization of these enzymes, particularly cellulases, quite difficult.

In nature, processive GHs are the primary enzymes responsible for polysaccharide depolymerization. It is thought that processivity enhances catalytic efficiency by keeping the enzyme closely associated to the substrate in between subsequent hydrolytic reactions and keeping once-detached single chains from re-associating with the insoluble material.<sup>68, 207</sup> Hydrolysis of recalcitrant polysaccharides by exo-processive enzymes can be divided into at least 4 putative steps: (i) binding to the polymer surface; (ii) recognition and capture of the chain end; (iii) formation of the productive complex and processive hydrolysis of the polymer chain, and; (iv) dissociation.<sup>23</sup> For processes ii) and iii), it is vital that individual binding subsites recognize and orient their specific substrates through formation of intermolecular bonds. In protein-carbohydrate complexes, the dominant interactions are hydrogen bonding and carbohydrate-aromatic stacking interactions, which also happen to be the primary determinants of ligand binding free

energy. Accordingly, it has been hypothesized that ligand binding free energy may be directly correlated with processive ability. Using free energy calculations, Payne et al. qualitatively illustrated this relationship for Family 7 GHs,<sup>61</sup> though, this study was inherently limited given the dearth of experimental measurements of processivity and binding free energy. Experiments quantifying the relationship of binding free energy with processivity do not yet exist.

Stacking or hydrophobic interactions can be formed between aromatic residues, in particular tryptophans, at the binding site of the enzyme and one or both sides of the sugar ring.<sup>208-209</sup> Such interactions have been explored in many studies showing that many processive GHs have a path of conserved solvent exposed aromatic residues leading into the active site cleft.<sup>43, 210</sup> The hydrophobic interactions create a flexible sheath allowing the polymeric substrate to slide through the active site as well as playing a central role in the binding and guidance of the insoluble substrate into the active site cleft.<sup>181, 210-211</sup> Mutagenesis of these aromatic residues nearly abolishes processivity.<sup>69-70, 212</sup> Hydrogen bonding can occur between polar residues and sugar-hydroxyl groups that have the potential to be involved in as many as three hydrogen bonds; one as a donor and two as an acceptor. Proctor *et al.* showed that the removal of the steric block mediated by the side chains of the polar Gln-316 and Asp-53 at the -3 subsite of the exo-active CjArb43A from *Cellvibrio japonicus* changed the mode of action.<sup>213</sup> A computational study of a processive *Trichoderma reesei* cellulase implicates polar residues in product inhibition, having the effect of reduced overall turnover.<sup>214</sup> Moreover, in a maltooligosaccharide – maltoporin model, it has been shown that the combination of hydrogen bonding and hydrophobic interactions makes a smoother energy profile with

regard to processivity than the two interactions alone.<sup>215</sup> The latter study highlights the importance of both polar-mediated hydrogen bonding and aromatic-mediated hydrophobic stacking. Nevertheless, the role of polar residues in mediating GH processivity is poorly understood.

The chitinolytic machinery of the gram-negative soil bacteria *Serratia marcescens* offers several advantages toward the study of processive GH action, and thus, is often used as model system for enzymatic degradation of recalcitrant polysaccharides.<sup>65, 83, 216-217</sup> The suite of *S. marcescens* chitinases includes two processive enzymes, one of which is chitinase A (ChiA), a Family 18 GH that preferentially acts from the reducing end of the sugar chain. The active site of ChiA has a deep, cleft-like architecture, where the catalytic domain contains 4 substrate (-4 to -1) and 3 product (+1 to +3) binding subsites. The carbohydrate binding module, fused with the catalytic domain, exhibits additional substrate binding sites extending toward the catalytic domain.<sup>42, 45, 70, 82</sup> In the catalytic domain of ChiA, the roles of three aromatic residues (Trp<sup>167</sup>, Trp<sup>275</sup>, and Phe<sup>396</sup>) in substrate binding, processivity, and activity have been characterized. Of the three, Trp<sup>167</sup>, situated in the -3 substrate binding subsite, was shown to be most important to processivity, efficiency of chitin degradation, and for the recognition and positioning of the substrate before hydrolysis.<sup>70, 145</sup> A comparable study of the ChiA polar residues will provide a more complete description of the protein-carbohydrate interactions governing processive ability, and further, can shed light on the relationship of binding free energy with processive ability.

In the present study, we investigate the role of two polar residues, Arg<sup>172</sup> and Thr<sup>276</sup>, in catalysis of glycosidic linkages, substrate binding, and processivity. These two

polar residues have been selected on the basis of their position in the active site, avoiding residues that will clearly abolish activity (as part of the reaction mechanism). Available crystal structures suggest Arg<sup>172</sup> and Thr<sup>276</sup> participate in hydrogen bonding with the substrate, and as such, may play key roles. We apply experimental and computational approaches to uncover the roles of these specific residues. Further, our investigation yields the experimental evidence toward validation of the relationship of binding free energy with processivity. Site directed mutagenesis has been used to obtain R172A and T276A variants. We have characterized the effects of these mutations with respect to wild-type using standard enzymological methods to determine apparent processivity (HPLC), the equilibrium binding association constant ( $K_a$ ) and  $\Delta H^\circ_r$  (ITC), and the preference of acetylated and deacetylated units in the different subsites (mass spectrometry (MS)). To understand the molecular-scale effects these residues have on substrate binding, MD simulations of the wild-type and variants have been conducted.

## 5.2 Materials and Methods

### Chemicals

Chito-oligosaccharides (CHOS) were obtained from Megazyme (Wicklow, Ireland). Squid pen  $\beta$ -chitin was purchased from France Chitin (180 $\mu$ m microparticulate, Marseille, France). Allosamidin was isolated from *Streptomyces sp.*, and the purity was controlled by <sup>1</sup>H NMR as described elsewhere.<sup>165</sup> Previously, the structure of allosamidin has been verified by both NMR and crystallography.<sup>218</sup> All other chemicals were of analytical grade.

## Site directed mutagenesis

Mutagenesis of ChiA-R172A, ChiA-T276A, ChiA-E315Q-R172A, and ChiA-E315Q-T276A was performed using the QuikChange™ site directed mutagenesis kit from Stratagene (La Jolla, CA, USA), as described by the manufacturer. To concentrate the DNA, the Pellet Paint® Co-precipitant kit from Novagen (Madison, WI, USA) was used as described in the product manual. The primers and templates used for the mutagenesis are listed in Table 5.1 and were purchased from Life Technologies (Carlsbad, CA, USA). To confirm that the genes contained the desired mutations and to check for the occurrence of nondesirable mutations, the mutated genes were sequenced using GATC Biotech's (Constance, Germany) LIGHTrun Sequencing service before they were transformed into *Escherichia coli* BL21Star (DE3) cells (Life Technologies, Carlsbad, CA, USA).

**Table 5.1** Primers used for site-directed mutagenesis

Mutant	DNA template	Primer	Sequence
ChiA-R172A	ChiA-WT	Forward	5'- GGGGCGTTTACGGGGCCAATTCACCGTCG-3'
		Reverse	5'- CGACGGTGAATTGGCCCCTAAACGCCCC-3'
ChiA-T276A	ChiA-WT	Forward	5'- TCGGCGGCTGGGCGCTGTCCGAC-3'
		Reverse	5'- GTCGGACAGCGCCCAGCCGCCGA-3'
ChiA-E315Q-R172A	ChiA-E315Q <sup>a</sup>	Forward	5'- GGGGCGTTTACGGGGCCAATTCACCGTCG-3'
		Reverse	5'- CGACGGTGAATTGGCCCCTAAACGCCCC-3'
ChiA-E315Q-T276A	ChiA-E315Q <sup>a</sup>	Forward	5'- TCGGCGGCTGGGCGCTGTCCGAC-3'
		Reverse	5'- GTCGGACAGCGCCCAGCCGCCGA-3'

<sup>a</sup> From Dybvik et al.<sup>137</sup>

## Construction of His<sub>10</sub>-ChiA-E315Q-R172A and His<sub>10</sub>-ChiA-E315Q-T276A

In order to subclone the inactive mutants ChiA-E315Q-R172A and ChiA-E315Q-T276A from pMay2-10 to pET16b (Novagen, Madison, WI, USA), the chitinase

fragments were amplified by PCR to introduce NdeI and XhoI restriction sites. PCR reactions were conducted with Q5<sup>®</sup> High-fidelity 2X Master Mix (New England Biolabs, Ipswich, MA). The amplification protocol consisted of an initial denaturation cycle of 30 s at 98 °C, followed by 30 cycles of 5 s at 98 °C, 30 s at 55 °C, 30 s at 72 °C, and a final step of 2 min at 72 °C. The following primers, purchased from Life Technologies (Carlsbad, CA, USA), were used in the PCR reaction: 5'TCGAAGGTCGTCATATGGCCGCGCCGGGC'3 (forward) and 5'CAGCCGGATCCTCGAGTTATTGAACGCCGGCGC'3 (reverse). The amplified insert was subcloned via NdeI and XhoI (New England Biolabs, Ipswich, MA, USA) restriction sites into pET16b by using the In-Fusion HD Cloning kit (Clontech Laboratories, Kyoto, Japan). The resulting pET16b constructs were sequenced using GATC Biotech's (Constance, Germany) LIGHTrun sequencing service to confirm the correct insert before they were transformed into *E. coli* Rosetta 2(DE3) cells (Life Technologies, Carlsbad, CA, USA).

### **Protein expression and purification of single mutants**

ChiA-R172A and ChiA-T276A genes were expressed in *E. coli* as described previously.<sup>185, 189</sup> For protein purification, the periplasmic extracts were loaded on a column packed with chitin beads (New England Biolabs) (1.5 cm in diameter, 10 ml stationary phase in total) and equilibrated in 50 mM Tris-HCl pH 8.0. After washing the column with the same buffer, the enzymes were eluted with 20 mM acetic acid. The buffer was then changed to 100 mM Tris-HCl (pH 8.0) using Macrosep Advance Centrifugal Device (30 kDa cutoff, Pall corporation, Port Washington, USA). Enzyme

purity was verified by SDS-PAGE, and protein concentrations were determined by using the Bradford protein assay from Bio-Rad (Hercules, CA, USA).

### **Protein expression and purification of double, inactive mutants**

For protein expression, *E. coli* Rosetta 2(DE3) cells containing the appropriate plasmid (ChiA-E315Q-R172A, ChiA-E315Q-T276A) were inoculated into 25 mL Terrific Broth (TB) medium containing 115 µg/mL ampicillin and 50 µg/mL chloramphenicol and grown at 37 °C and 200 rpm for 16 h. Cell culture were then inoculated into 250 mL TB medium containing 115 µg/mL ampicillin and 50 µg/mL chloramphenicol to an OD<sub>600</sub> of 0.1. This culture was cultivated until the OD<sub>600</sub> reached 0.8-1.0. The temperature was decreased to 22°C, and gene expression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside for 20 h. The cells were harvested by centrifugation (8000 rpm, 20 min at 4 °C). Periplasmic fractions were prepared by osmotic shocking as described elsewhere.<sup>138</sup> A cytoplasmic protein extraction was also performed by re-suspending the spheroplasts in lysis buffer (0.1 mg/mL lysozyme, 50 mM Tris-HCl, 50 mM NaCl, 4 mM MgCl<sub>2</sub>, 1 mM EDTA, 0.1 mM PMSF; pH 8.0) before incubating it at 37 °C for 30 min. Cell debris was removed by centrifugation (8000 rpm, 20 min at 4 °C). The resulting supernatant was used for further enzyme purification. Both the periplasmic and cytoplasmic extracts were sterilized by filtration (0.2 µm) prior to protein purification.

Proteins were purified on a column packed with Ni-NTA Agarose matrix (Qiagen, Venlo, Netherlands) (1.5 cm in diameter, 5 ml stationary phase in total). The column was pre-equilibrated in buffer A (20 mM Tris-HCl, 20 mM imidazole, 500 mM NaCl, pH 8.0) before the periplasmic and cytoplasmic extracts were applied. After washing with buffer

B (20 mM Tris-HCl, 500 mM NaCl, pH 8.0), fractions containing the enzyme were eluted with buffer C (20 mM Tris-HCl, 250 mM imidazole, 500 mM NaCl, pH 8.0). A flow rate of 2.5 mL/min was used at all times. Enzyme purity was verified by SDS-PAGE, and fractions containing purified enzyme were concentrated and transferred (Macrosep Advance Centrifugal Device, 30 kDa cutoff, Pall corporation, Port Washington, USA) to 20 mM potassium phosphate buffer pH 6.0. Enzyme purity was verified by SDS-PAGE while protein concentrations were determined by using the Bradford protein assay from Bio-Rad (Hercules, CA, USA).

### **Degradation of chitosan**

Chitosan was dissolved in 80 mM sodium acetate buffer (pH 5.5) to a final concentration of 10 mg/mL as described previously.<sup>219</sup> Chitosan with a fraction of *N*-acetylated units ( $F_A$ ) = 0.65 was depolymerized by adding 2.5  $\mu$ g enzyme per 1 mg chitosan. The reactions were run to completion (maximum degree of scission ( $\alpha$ )) before enzyme activity was stopped by lowering the pH with 150  $\mu$ L 1 M HCl and 2 minutes boiling.<sup>219</sup>

### **2-aminoacridone derivatization and sequence determination of chito-oligosaccharides**

In order to determine the sequence of chitosan oligomers, the oligosaccharides were derivatized by reductive amination of the reducing end with 2-aminoacridone (AMAC) as described previously.<sup>167, 220</sup>



### **Matrix assisted laser desorption/ionization mass spectrometry**

Sequencing of the pentameric chitosan oligomers (degree of polymerization (DP) = 5) was performed using matrix assisted laser desorption/ionization time-of-flight/time-of-flight mass spectrometry/mass spectrometry (MALDI-TOF/TOF-MS/MS) as described earlier.<sup>167</sup> MS spectra were acquired using an Ultraflex<sup>TM</sup> TOF/TOF mass spectrometer (Bruker Daltonik GmbH, Bremen, Germany) with gridless ion optics under control of Flexcontrol. For sample preparation, 1  $\mu$ l of the reaction products was mixed with 1  $\mu$ l 10 % 2,5-dihydroxybenzoic acid (DHB) in 30 % ethanol and spotted onto a MALDI target plate. The MS experiments were conducted using an accelerating potential of 20 kV in the reflectron mode.

### **Degradation of chitin**

Hydrolysis of chitin was carried out as described previously.<sup>78</sup> The extent of degradation is defined as the percentage of number of moles of solubilized GlcNAc-units with respect to number of moles GlcNAc-units in solid form (chitin) used in the experiments.

### **High performance liquid chromatography (HPLC)**

Concentrations of mono- and disaccharides were determined using HPLC with a Rezex Fast fruit H<sup>+</sup> column (100 mm length and 7.8 mm inner diameter) (Phenomenex). An 8  $\mu$ l sample was injected on the column, and the mono- and disaccharides were eluted isocratically at 1 mL/min with 5 mM H<sub>2</sub>SO<sub>4</sub> at 85 °C. The mono- and disaccharides were monitored by measuring absorbance at 210 nm, and the amounts were quantified by

measuring peak areas. Peak areas were compared with peak areas obtained with standard samples with known concentrations of mono- and disaccharides.

### **Isothermal titration calorimetry (ITC)**

ITC experiments were performed with a VP-ITC system from Microcal, Inc. (Northampton, MA).<sup>139</sup> Solutions were thoroughly degassed by vacuum pump prior to experiments to avoid air bubbles in the calorimeter. Standard ITC conditions were 250  $\mu\text{M}$  of allosamidin or 500  $\mu\text{M}$  of  $(\text{GlcNAc})_6$  in the syringe and 15  $\mu\text{M}$  of enzyme in the reaction cell in 20 mM potassium phosphate buffer of pH 6.0. The only exception was for ChiA-E315Q-T276A against  $(\text{GlcNAc})_6$ . To ensure a  $c$ -value between 10 and 1000, which is a prerequisite for meaningful calculations of  $K_a$ ,<sup>139</sup> 1 mM  $(\text{GlcNAc})_6$  and 30  $\mu\text{M}$  enzyme were used yielding a  $c$ -value of 25. Aliquots of 4-8  $\mu\text{l}$  were injected into the reaction cell at 180s intervals at 30 °C with a stirring speed of 260 rpm. 45 injections were performed. At least three independent titrations were performed for each binding reaction.

### **Analysis of calorimetric data**

ITC data were collected automatically using the Microcal Origin v.7.0 software accompanying the VP-ITC system.<sup>139</sup> Prior to further analysis, data were corrected for heat dilution by subtracting the heat remaining after saturation of binding sites on the enzyme. Data were fitted using a non-linear least-squares algorithm using a single-site binding model employed by the Origin software that accompanies the VP-ITC system. All data from the binding reactions fit well with the single site binding model yielding the stoichiometry ( $n$ ), equilibrium binding association constant ( $K_a$ ), and the reaction enthalpy change ( $\Delta H_r^\circ$ ) of the reaction. The equilibrium binding dissociation constant

( $K_d$ ), reaction free energy change ( $\Delta G_r^\circ$ ) and the reaction entropy change ( $\Delta S_r^\circ$ ) were calculated from the relation described in Equation 5.1.

$$\Delta G_r^\circ = -RT \ln K_a = RT \ln K_d = \Delta H_r^\circ - T \Delta S_r^\circ \quad (5.1)$$

Errors are reported as standard deviations of at least three experiments.

### **Kinetic Analysis**

The kinetic constants,  $k_{cat}$  and  $K_m$ , of the ChiA-T276A and ChiA-R172A mutants were determined using (GlcNAc)<sub>4</sub> substrate,<sup>200, 221</sup> which at the substrate concentrations and in the time frames used for kinetic analysis, is hydrolyzed into two dimers by both enzyme variants in this study. Reaction mixtures, with 10 different (GlcNAc)<sub>4</sub> concentrations varying from 2 to 100  $\mu$ M in 20 mM sodium acetate buffer, pH 6.1 and 0.1 mg/mL BSA (final concentrations), were pre-incubated in a 37 °C water bath for 10 min prior to starting the reactions. The reactions were started by adding purified enzyme to reach an enzyme concentration of 0.2 nM (ChiA-R172A) and 1 nM (ChiA-T276A) in a total reaction volume of 1.0 mL. To determine the rate of (GlcNAc)<sub>4</sub> hydrolysis at a specific concentration, seven 75- $\mu$ L samples, including one at time equal zero, were withdrawn at regular time intervals over a total period of 1 – 10 min, and the enzyme was inactivated by adding 75  $\mu$ L of 20 mM H<sub>2</sub>SO<sub>4</sub>. All of the reactions were run in duplicate, and all of the samples were stored at -20 °C until HPLC analysis. Reaction conditions were such that the rate of hydrolysis of (GlcNAc)<sub>4</sub> was essentially constant over time, with the (GlcNAc)<sub>4</sub> concentration always staying above 80% of the starting concentration. Data points were only discarded when hydrolysis had not taken place or more than 20% of the initial (GlcNAc)<sub>4</sub> were hydrolyzed (to ensure initial rates only). If more than two of the seven data points had to be removed due to the reasons described

above, the whole set was discarded. The slopes of plots of 0.5 times the (GlcNAc)<sub>2</sub> concentration versus time were taken as the hydrolysis rate. The rates were plotted versus substrate concentration in a Michaelis–Menten plot, and the experimental data were fitted to the Michaelis–Menten equation by nonlinear regression using Origin v7.0 (OriginLab Corp., Northampton, MA).

### **Initial hydrolysis of (GlcNAc)<sub>5</sub> in H<sub>2</sub><sup>18</sup>O**

To determine the preferred positioning of the substrate in the active site, initial hydrolysis of (GlcNAc)<sub>5</sub> in H<sub>2</sub><sup>18</sup>O (Larodan Fine Chemicals, Malmö, Sweden) was performed, as described previously.<sup>222-223</sup> The hydrolysis was completed at 37 °C and 600 rpm in H<sub>2</sub><sup>18</sup>O containing 20 mM Ammonium acetate pH 6.1, 100 μM (GlcNAc)<sub>5</sub> and 5, 15, and 20 nM ChiA-WT, ChiA-R172, and ChiA-T276A, respectively. The hydrolysis products were analyzed by MALDI-TOF-MS as described by Eide et al.<sup>223</sup>

### **MD simulations**

Classical MD simulations were constructed for the three ChiA systems: ChiA-WT, ChiA-R172A, and ChiA-T276A. The initial coordinates for all three MD simulation sets were from the protein data bank entry 1EHN.<sup>45</sup> The E315Q mutation of the 1EHN structure was reversed in each case using the PyMOL mutagenesis tool. Similarly, the R172A and T276A variants were constructed from the wild-type coordinates. The (GlcNAc)<sub>6</sub> ligand was also obtained from 1EHN. The deposited structure exhibits a (GlcNAc)<sub>8</sub> ligand bound from the –6 to +2 binding subsites. In accordance with experimental protocol, we manually shortened this ligand by deleting the atoms in the –6 and –5 subsites, leaving a bound (GlcNAc)<sub>6</sub> in the –4 to +2 sites. The acetyl group of the –1 pyranose ring and the side chain of Asp<sup>313</sup> were manually rotated around their range

of dihedrals, such that the catalytic residues and the -1 sugar reflected the catalytically active conformation of a Family 18 chitinase.<sup>79</sup> The protonation states for each simulation were determined using H++ at a pH of 6.0 and an internal and external dielectric constant of 10 and 80, respectively.<sup>148-150</sup> Two disulfide bridges observed in the structure also covalently bonded, between Cys<sup>115</sup> and Cys<sup>120</sup> and Cys<sup>195</sup> and Cys<sup>218</sup>. Sodium ions were added to each system to ensure the total system charge equaled zero.

The constructed systems were then minimized in vacuum, solvated, and re-minimized. The vacuum minimizations were performed to remove any initial bad contacts between overlapping atoms as a result of the addition of hydrogens. The initial minimization protocol included 1000 iterations of steepest descent (SD) followed by 1000 iterations of adopted basis Newton Raphson (ABNR), each applied to the entire protein-ligand complex. The minimized systems were then solvated in explicit TIP3P water. The periodic boundary conditions were 120 Å x 120 Å x 120 Å. This box size was selected such that the protein had a minimum of 10 Å solvent buffer on each side. The total system size for each simulation is approximately 175,000 atoms. The solvated systems were then minimized again according to a stepwise protocol: (1) the water molecules were minimized for 10,000 SD steps, keeping the protein and ligand rigid; (2) the protein and water were then minimized for 10,000 SD steps keeping the ligand fixed; and (3) the entire system was minimized for 10,000 steps using SD.

The completely constructed systems were then heated and equilibrated prior to collection of production MD simulation data. In all cases, the CHARMM36 all-atom force field with CMAP corrections was used to model the protein and carbohydrate interactions,<sup>95, 151-152</sup> and water was modeled with the modified TIP3P force field.<sup>153-154</sup>

The systems were heated from 100 K to 300 K in 50 K increments over the span of 4 ps. The system density was equilibrated in the NPT ensemble for 100 ps. The Nosè-Hoover thermostat and barostat were used for pressure control.<sup>224-225</sup> The systems were constructed, minimized, heated, and equilibrated using CHARMM.<sup>95</sup>

After equilibration, the simulations were moved into the NAMD simulation package for efficient integration.<sup>96</sup> In the NVT ensemble, 250-ns simulations were performed of each of the three systems at 300 K. The periodic volumes were based on the final values from the NPT equilibrations in CHARMM. Each simulation used a 2-fs time step for the integration scheme. The Particle Mesh Ewald method was used to describe long-range electrostatics with a sixth order b-spline, a Gaussian distribution width of 0.312 Å, and a 1 Å grid spacing.<sup>226</sup> A non-bonded cutoff distance of 10 Å, a switching distance of 9 Å, and a non-bonded pair list distance of 12 Å were used. All hydrogen atom distances were fixed using the SHAKE algorithm.<sup>227</sup> VMD was used for visualization of all the trajectories and hydrogen bond analysis.<sup>155</sup> Hydrogen bond formation was determined using a 3.4 Å distance cutoff along with an angle cutoff of 60°.

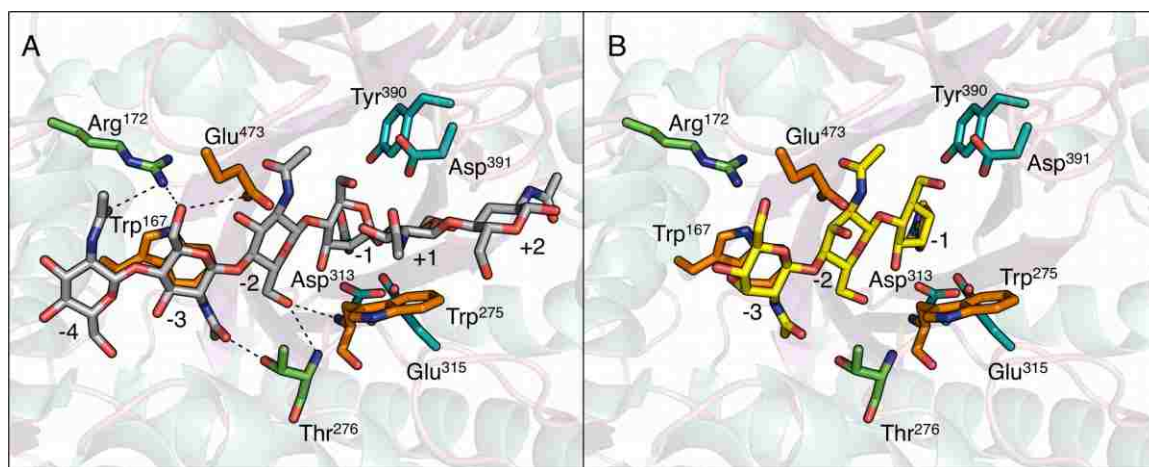
## 5.3 Results

### Mutant design and initial mutant characterization

ChiA is exo-processive acting from the reducing end of the substrate. According to accepted naming convention, ChiA substrate binding subsites are identified by negative numbers, with hydrolysis occurring between binding sites -1 and +1.<sup>228</sup> Several important intermolecular interactions between the chito-oligosaccharide substrate and the enzyme occur in substrate binding subsites -2, -3, and -4 (Figure 5.1), as discussed by

Norberg and co-workers.<sup>145</sup> Arg<sup>172</sup> is notable in that it appears to interact with the substrate in three subsites, -2, -3, and -4, rather than through a single proximal interaction. From examination of the crystal structure, Arg<sup>172</sup> interacts with the acetyl group carbonyl oxygen in the -4 subsite. Though, the -2 subsite GlcNAc is too distant to form an interaction with Arg<sup>172</sup>, the charged protein side chain is thought to be a part of an electrostatic-dipole interaction with the C3 secondary alcohol or the acetyl group. In subsite -3, Arg<sup>172</sup> appears to participate in a bifurcated hydrogen bond to the GlcNAc primary alcohol, in conjunction with Glu<sup>473</sup>.

Similarly, the crystal structure indicates Thr<sup>276</sup> interacts with both the -2 and -3 GlcNAc moieties. In the -3 subsite, the threonine OH group forms a hydrogen bond with the substrate carbonyl oxygen. Thr<sup>276</sup>, together with Trp<sup>275</sup>, forms hydrogen bonds with the -2 GlcNAc primary alcohol through the protein backbone. Given the ability of Arg<sup>172</sup> and Thr<sup>276</sup> to form strong electrostatic – dipole interactions and their proximity to multiple hydrogen bonding partners along the length of the active site cleft, we anticipate each residue plays a role in substrate binding, processivity, and possibly catalytic turnover. Mutating these residues to alanine effectively abolishes the electrostatic potential of the residue and negates the ability to hydrogen bond except through backbone-mediated interactions. This latter point is particularly important for examination of the effect of mutating Thr<sup>276</sup> to alanine, where loss in binding or processive ability may then be attributed to interaction of Thr<sup>276</sup> with the carbonyl oxygen in the -3 subsite.



**Figure 5.1** Crystal structures of the active site of ChiA in the presence of (GlcNAc)<sub>6</sub> (A, colored grey) and allosamidin (B, colored yellow). Highlighted in green are the two residues investigated in this study (Thr<sup>276</sup> and Arg<sup>172</sup>). Other important substrate binding residues (Trp<sup>167</sup>, Glu<sup>473</sup>, Asp<sup>313</sup> and Glu<sup>315</sup>) are highlighted in orange while product binding residues (Trp<sup>275</sup>, Tyr<sup>390</sup> and Asp<sup>391</sup>) are shown in cyan.

To assess activity of the constructed mutants, attempts were made to determine  $k_{\text{cat}}$  and  $K_{\text{m}}$  using the natural substrate (GlcNAc)<sub>6</sub>, which is also the substrate used in the ITC experiments (see below). However, we experienced a non-linear decrease in (GlcNAc)<sub>6</sub> concentration with time using the state-of-the-art HPLC based method for chito-oligosaccharide quantification.<sup>221</sup> The (GlcNAc)<sub>6</sub> substrate productively binds to ChiA in at least three different occupancies: i) from the -3 to the +3 subsite to yield two (GlcNAc)<sub>3</sub> as products, ii) from the -4 to the +2 subsite to yield (GlcNAc)<sub>4</sub> and (GlcNAc)<sub>2</sub> as products, and iii) from the -4 subsite to the +2 subsite with concomitant processive hydrolysis to yield three (GlcNAc)<sub>2</sub> as products.<sup>81</sup> In addition, all three products cause product inhibition,<sup>159</sup> and (GlcNAc)<sub>4</sub> is an especially strong competitive substrate.<sup>221</sup> Accordingly, the relative low precision (for this purpose) of the HPLC based



method for chito-oligosaccharide quantification does not allow for fitting of the data to individual kinetic parameters. Then, as a simplification, the natural substrate (GlcNAc)<sub>4</sub>, which productively binds from -2 to + 2 without concomitant processive hydrolysis and production of new substrates, was used to obtain Michaelis-Menten kinetics as an initial characterization of the mutants.<sup>221</sup> Moreover, previous work has shown that effects of mutations in the -2 and -3 subsites can be manifested in the kinetics of (GlcNAc)<sub>4</sub> hydrolysis.<sup>145, 200</sup> In the substrate concentration range studied, both mutants showed straightforward Michaelis-Menten kinetics without substrate inhibition, as also previously shown for ChiA-WT. The kinetic parameters are listed in Table 5.2.

**Table 5.2** Kinetic constants of ChiA-WT, ChiA-T276A, and ChiA-R172A for the hydrolysis of (GlcNAc)<sub>4</sub> in 20 mM Sodium Acetate, pH 6.1, and 0.1 mg/mL BSA at 37 °C.

Enzyme	$k_{\text{cat}}$ (s <sup>-1</sup> )	$K_{\text{m}}$ (μM)	$k_{\text{cat}}/K_{\text{m}}$ (μM <sup>-1</sup> s <sup>-1</sup> )
ChiA-WT <sup>a)</sup>	33 ± 1	9 ± 1	4
ChiA-T276A	12 ± 2	7 ± 1	2
ChiA-R172A	39 ± 4	88 ± 12	0.4

<sup>a)</sup> From Krokeide et al.<sup>221</sup>

### Degradation of chitosan

Deacetylation of chitin yields the water-soluble, heterologous de-*N*-acetylated analog, chitosan, that consists of linear β-1,4-linked *N*-glucosamine (GlcN; **D**-unit) and GlcNAc units.<sup>229-230</sup> It can be prepared with varying amount and pattern of *N*-acetylated units as well as varying length of the polymer chain.<sup>231-232</sup> Characteristic of Family 18 GHs, chitinases degrade chitin with retention of the stereochemistry at the anomeric carbon (www.cazy.org<sup>52</sup>). They also employ a specialized substrate-assisted mechanism in which the *N*-acetyl group of the sugar in subsite -1 acts as the nucleophile.<sup>79, 88, 130, 188</sup>

As a result, Family 18 chitinases have an absolute preference for acetylated units in this subsite. For this reason, degradation of chitosan by chitinases has proven to be a useful tool in the determination of both processivity as well as substrate positioning.<sup>70, 145, 219</sup> In this study, chitosan with a degree of acetylation of 65 % was degraded by ChiA-R172A and ChiA-T276A to a maximum degree of scission. The pentameric products obtained were labeled at the reducing end and analyzed with respect to sequence of acetylated and deacetylated units by reducing end labeling and MALDI-TOF-TOF-MS/MS.<sup>220</sup> Such sequences detail the preferences each individual subsite has for acetylated vs. deacetylated units, and hence, show its importance in recognizing and positioning of the substrate before hydrolysis.<sup>145, 167</sup> Beyond the absolute preference for an **A** in subsite -1 for each enzyme, the sequences (Table 5.3) illustrate a preference for an acetylated unit in subsites -3 and -4 for ChiA-T276A while ChiA-R172A has no strong preferences as seen by the many different sequences present. In comparison, ChiA-WT has a preference for an acetylated unit in subsites -1, -2, and -4.<sup>145</sup>

**Table 5.3** Sequence determination of CHOS of DP = 5 obtained after hydrolysis of chitosan ( $F_A = 0.65$ ) with ChiA-WT, ChiA-T267A, ChiA-R172A to a maximum degree of scission.

<i>m/z</i>	CHOS	Sequence WT <sup>a)</sup>	T276A	R172A
1124.252		-	-	DDADA-▲
1166.265	D2A3	DADAA-▲	-	DDAAA-▲ DAADA-▲ DADAA-▲ ADADA-▲ ADDAA-▲
1208.269	D1A4	AADAA-▲	DAAAA-▲ AAADA-▲	AAADA-▲ ADAAA-▲

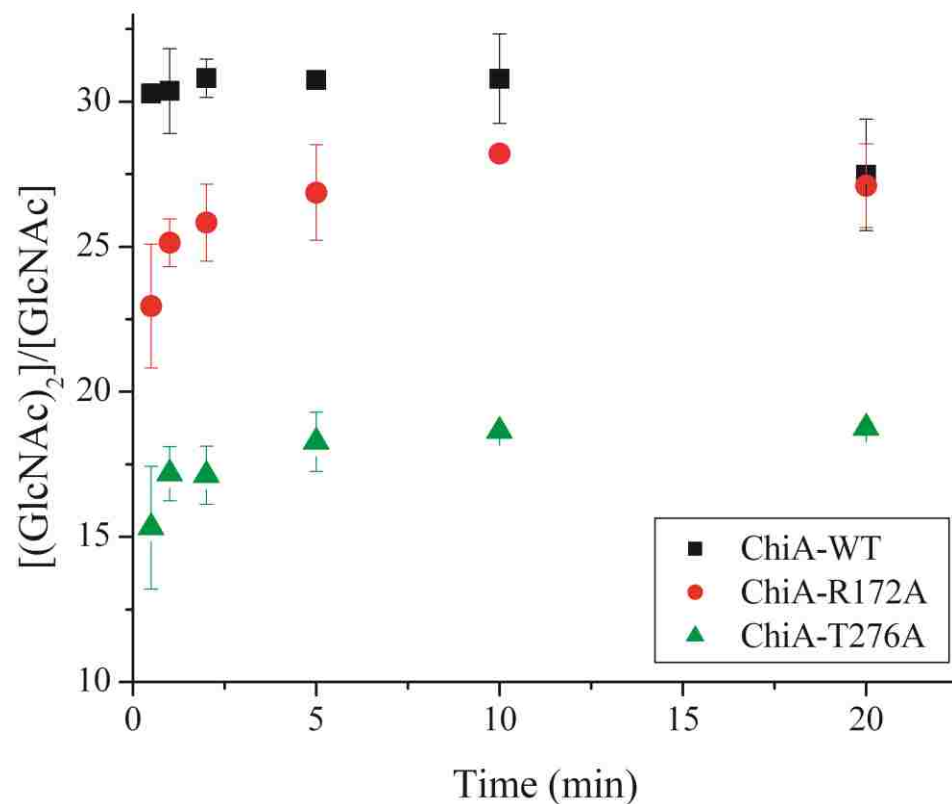
▲ represents the reducing end tag. <sup>a)</sup> From Norberg et al.<sup>145</sup>

### Degradation of chitin

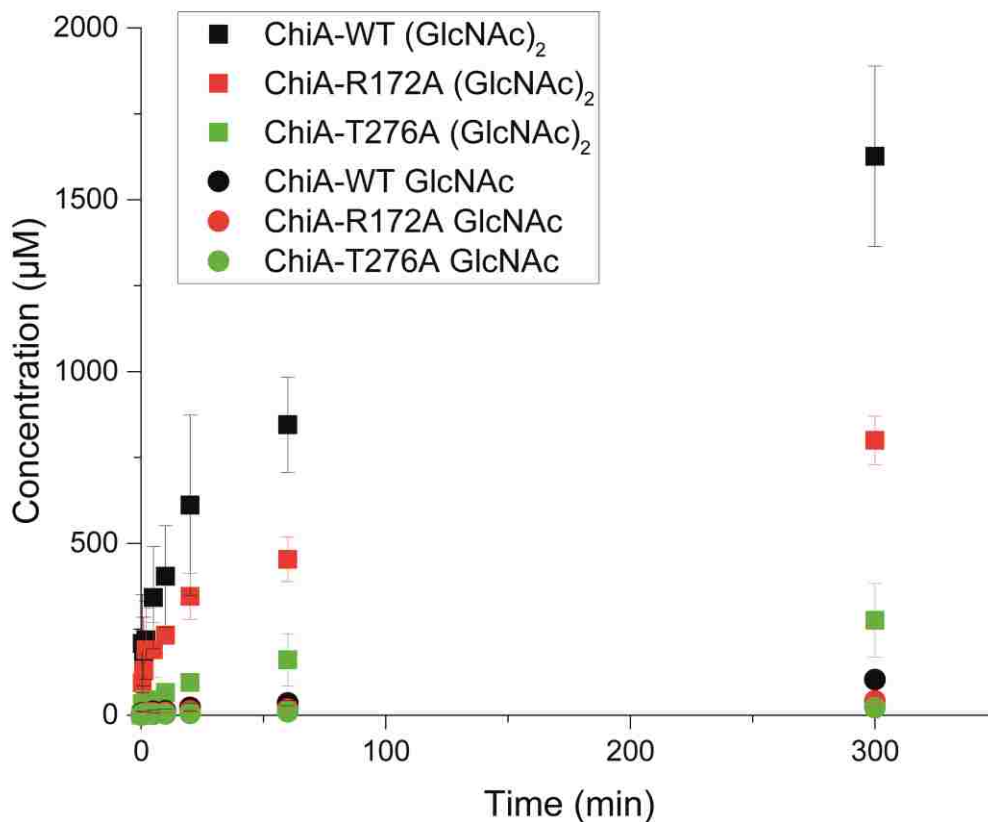
Apparent processivity ( $P^{app}$ ) is defined as the average number of consecutive catalytic cycles performed per initiated processive run (either through endo- or exo-mode of attachment) along the crystalline substrate. This value can be measured by a number of different methods, depending on the substrate.<sup>64, 74, 78</sup> A common means of measuring  $P^{app}$  in chitinases is by determining the  $[(\text{GlcNAc})_2/(\text{GlcNAc})]$  ratio. Given the requirement for the *N*-acetyl group in hydrolysis, this approach yields consistent results in determining processive degradation of chitin polysaccharides. In the case of a processive enzyme, the first cleavage from a polymer chain end will result in the release of an odd-numbered oligosaccharide (e.g., mono- or trisaccharide), whereas all subsequent cleavages result in the production of disaccharides because of the 180° rotation of the GlcNAc-units. For non-processive enzymes, the same measurement will result in a random distribution of oligosaccharide lengths. In each case, the steady-state ratio of dimers to monomers represents the relative processive ability; neither processive

nor non-processive chitinases are capable of hydrolyzing (GlcNAc)<sub>2</sub>. Lower ratios are indicative of more initiated runs associated with the inability to maintain prolonged substrate association. The [(GlcNAc)<sub>2</sub>/(GlcNAc)] ratio is valid if it is assumed that the first product is a trisaccharide that subsequently is hydrolyzed to a mono- and disaccharide.<sup>74, 78</sup>

$P^{app}$  tends to decrease as the substrate is consumed, most likely because the substrate becomes enriched with recalcitrant regions where there are fewer obstacle-free paths for processive enzyme attachment.<sup>64, 78, 233</sup> Without the addition of accessory enzymes and glucosidases, the enzymes eventually encounter obstacles or fail to release from the substrate causing traffic jams of unproductively bound enzymes. It is thus important to assess processivity during the early stages of the reaction.<sup>78</sup> Here, initial degradation of  $\beta$ -chitin yielded [(GlcNAc)<sub>2</sub>/(GlcNAc)] ratios of  $25.9 \pm 0.9$  and  $17.1 \pm 0.4$  for ChiA-R172A and ChiA-T276A, respectively.<sup>78</sup> The value for ChiA-WT has previously been determined to be  $30.1 \pm 1.5$  (Figure 5.2). Furthermore, earlier studies demonstrate a positive correlation between processive ability and rate of GH-catalyzed polysaccharide hydrolysis.<sup>64, 78, 83, 234</sup> By examining initial rates of chitin degradation as determined by product formation vs. time (Figure 5.3), it is clear that ChiA-WT, the most processive, is faster than ChiA-T276A, the least processive, with ChiA-R172A being intermediate.

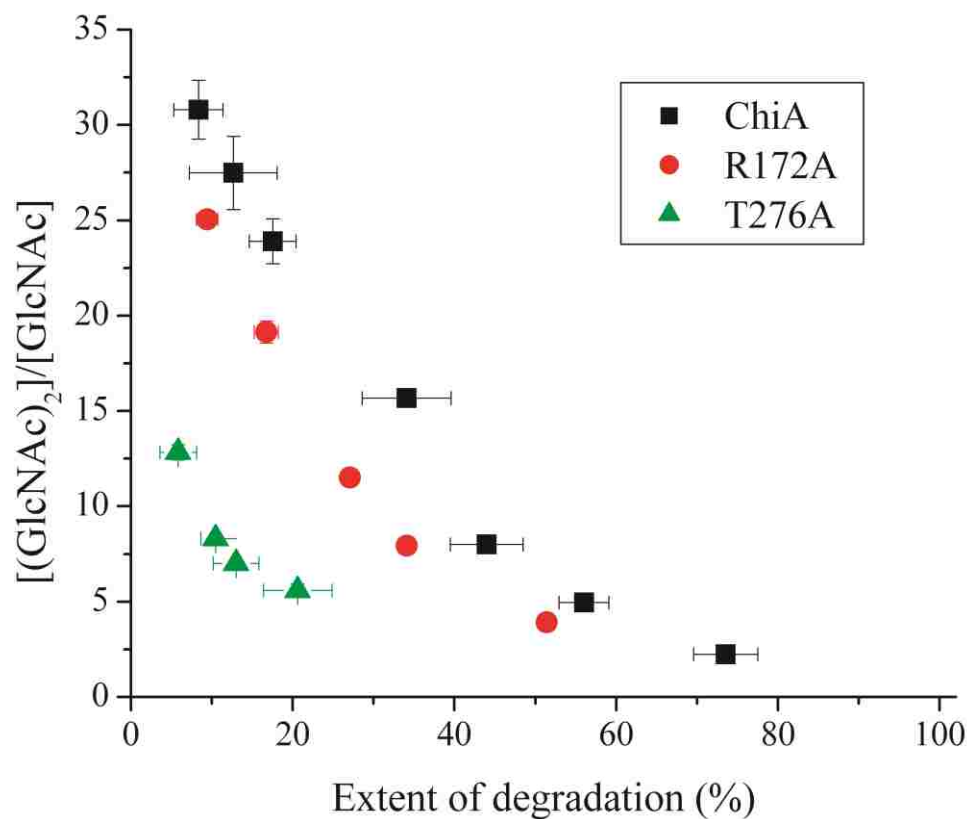


**Figure 5.2** Comparison of initial  $[(\text{GlcNAc})_2]/[\text{GlcNAc}]$  ratios for ChiA wt (■), ChiA-R172A (●) and ChiA-T276A (▲). Hydrolysis was undertaken with 2.5  $\mu\text{M}$  enzyme, pH 6.1 sodium acetate buffer at  $t = 37^\circ\text{C}$  with 2.0 mg/ml chitin. Error bars represent standard deviation of duplicate experiments. WT control experiments were conducted at the same time as the variant experiments and were in agreement with previously published values, which have been provided here for comparison.<sup>78</sup>



**Figure 5.3** Progress curves for the formation of (GlcNAc)<sub>2</sub> (squares) and GlcNAc (circles) after hydrolysis of  $\beta$ -chitin by ChiA WT (black), ChiA-R172A (red), and ChiA-T276A (green). (GlcNAc)<sub>2</sub> and GlcNAc were the only products detected. A clear correlation between initial rates and  $P^{app}$  is observed. WT control experiments were conducted at the same time as the variant experiments and were in agreement with previously published values, which have been provided here for comparison.<sup>78</sup>

Finally, distinct differences between the three enzymes can also be seen with regard to the degradation efficiency. ChiA-WT has the ability to degrade 75 % of the  $\beta$ -chitin while 50 % is degraded by ChiA-R172A. Only 20 %  $\beta$ -chitin is degraded when ChiA-T276A is used (Figure 5.4).



**Figure 5.4** Relationship of the  $[(\text{GlcNAc})_2]/[\text{GlcNAc}]$  ratio for ChiA wt (■), ChiA-R172A (●) and ChiA-T276A (▲) with extent of degradation. The figure shows chitinases having higher initial processivity degrades  $\beta$ -chitin more efficiently. Hydrolysis was undertaken with 2.5  $\mu\text{M}$  enzyme, pH 6.1 sodium acetate buffer at  $t = 37$  °C with 2.0 mg/ml chitin. Error bars represent standard deviation of duplicate experiments. WT control experiments were conducted at the same time as the variant experiments and were in agreement with previously published values, which have been provided here for comparison.<sup>78</sup>

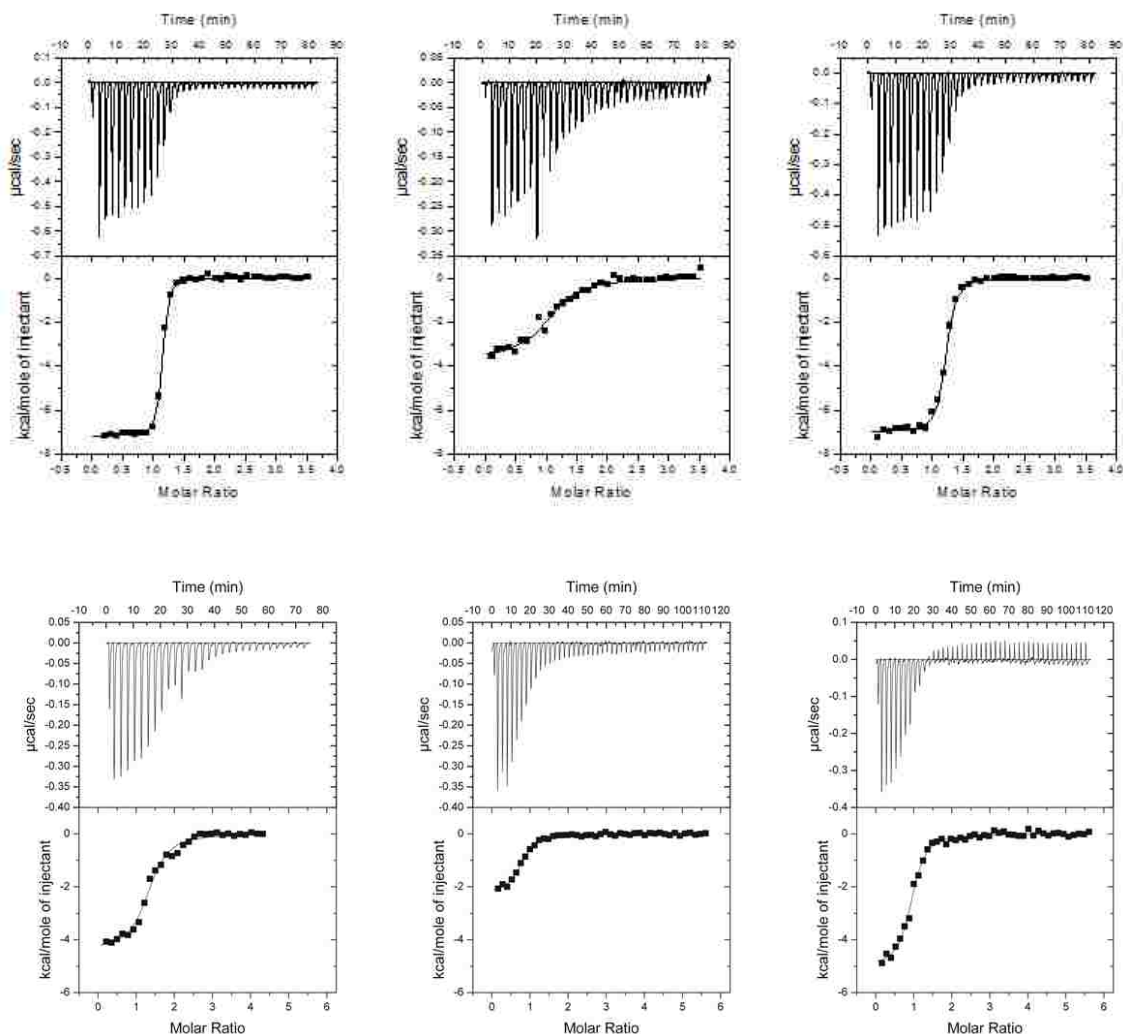
## Thermodynamics

To assess the contribution Arg<sup>172</sup> and Thr<sup>276</sup> have on the binding free energy between ChiA and substrate, ITC measurements were taken for the individual mutants binding the soluble substrate (GlcNAc)<sub>6</sub> and the well-known inhibitor allosamidin. In addition, a subsite occupancy study with (GlcNAc)<sub>5</sub> was undertaken. All ligands are similar to the natural substrate and span the important catalytic subsites -3 to -1 (Figure 5.1). When allosamidin is the ligand, catalytically active enzymes can be applied (less perturbation of the system), in contrast to when (GlcNAc)<sub>6</sub> is the ligand and catalytically inactivated enzymes (mutation of the catalytic Glu to Gln) must be used.

The binding of (GlcNAc)<sub>6</sub> to ChiA-WT was undertaken at pH 6.0 (20 mM potassium phosphate buffer) and 30 °C. To determine the (GlcNAc)<sub>6</sub> binding thermodynamics, an inactive variant of the enzyme was used. Thus, the free energy values of (GlcNAc)<sub>6</sub> were determined with a single point mutation (E315Q) representing wild-type and double mutations for ChiA-R172A and ChiA-T276A. A typical thermogram and theoretical fit to the experimental data is given in Figure 5.5. ChiA binds (GlcNAc)<sub>6</sub> with a  $K_d$  of  $0.56 \pm 0.03 \mu\text{M}$ , corresponding to a free energy change ( $\Delta G_r^\circ$ ) of  $-8.7 \pm 0.1 \text{ kcal/mol}$ , an enthalpic change ( $\Delta H_r^\circ$ ) of  $-4.5 \pm 0.2 \text{ kcal/mol}$ , and an entropic change ( $\Delta S_r^\circ$ ) of  $13.9 \pm 0.7 \text{ cal/K mol}$  ( $-T\Delta S_r^\circ = -4.2 \pm 0.2 \text{ kcal/mol}$ ) (Table 5.4). The  $K_d$  for the binding between ChiA-R172A and (GlcNAc)<sub>6</sub> equals  $0.61 \pm 0.02 \mu\text{M}$  with a  $\Delta G_r^\circ = -8.6 \pm 0.1 \text{ kcal/mol}$ , Table 5.4). The reaction is accompanied by a  $\Delta H_r^\circ = -4.8 \pm 0.2 \text{ kcal/mol}$  and a  $\Delta S_r^\circ = 12.5 \pm 0.7 \text{ cal/K mol}$  ( $-T\Delta S_r^\circ = -3.8 \pm 0.2 \text{ kcal/mol}$ ). For ChiA-T276A, the binding has a  $K_d = 1.2 \pm 0.2 \mu\text{M}$  ( $\Delta G_r^\circ = -8.2 \pm 0.1 \text{ kcal/mol}$ ) and a  $\Delta H_r^\circ$  and  $\Delta S_r^\circ$  of  $-1.9 \pm 0.2 \text{ kcal/mol}$  and  $20.8 \pm 0.7 \text{ cal/K mol}$  ( $-T\Delta S_r^\circ = -6.3 \pm 0.2 \text{ kcal/mol}$ ),



respectively.



**Figure 5.5** Thermograms and binding isotherms with theoretical fits of allosamidin and (GlcNAc)<sub>6</sub> to chitinases. Top: Thermograms (upper panels) and binding isotherms with theoretical fits (lower panels) obtained for the binding of allosamidin to ChiA-WT (left), ChiA-T276A (middle), and ChiA-R172A (right) at  $t = 30$  °C in 20 mM potassium phosphate at pH 6.0. Bottom: Thermograms (upper panels) and binding isotherms with theoretical fits (lower panels) obtained for the binding of (GlcNAc)<sub>6</sub> to ChiA-WT (left), ChiA-T276A (middle), and ChiA-R172A (right) at  $t = 30$  °C in 20 mM potassium phosphate at pH 6.0.

**Table 5.4** Thermodynamic Parameters for (GlcNAc)<sub>6</sub> and allosamidin binding to ChiA-WT, ChiA-T276A, ChiA-R172A from *Serratia marcescens* at  $t = 30$  °C, pH = 6.0.

Enzyme	$K_d^a$	$\Delta G_r^{ob}$	$\Delta H_r^{ob}$	$-T\Delta S_r^{ob}$
<u>(GlcNAc)<sub>6</sub></u>				
ChiA-WT <sup>c</sup>	$0.56 \pm 0.03$	$-8.7 \pm 0.1$	$-4.5 \pm 0.2$	$-4.2 \pm 0.2$
ChiA-T276A <sup>c</sup>	$1.2 \pm 0.2$	$-8.2 \pm 0.1$	$-1.9 \pm 0.2$	$-6.3 \pm 0.2$
ChiA-R172A <sup>c</sup>	$0.61 \pm 0.02$	$-8.6 \pm 0.1$	$-4.8 \pm 0.2$	$-3.8 \pm 0.2$
<u>Allosamidin</u>				
ChiA-WT <sup>d</sup>	$0.027 \pm 0.002$	$-10.5 \pm 0.1$	$-7.5 \pm 0.3$	$-2.8 \pm 0.3$
ChiA-T276A	$1.0 \pm 0.1$	$-8.3 \pm 0.1$	$-3.7 \pm 0.2$	$-4.6 \pm 0.2$
ChiA-R172A	$0.067 \pm 0.008$	$-9.9 \pm 0.1$	$-7.1 \pm 0.1$	$-2.8 \pm 0.1$

<sup>a</sup>  $\mu$ M, <sup>b</sup> kcal/mol, <sup>c</sup> the catalytic acid Glu<sup>315</sup> have been exchanged to Gln, <sup>d</sup> From Baban et al.<sup>166</sup>

Since (GlcNAc)<sub>6</sub> binding to the wild type and the individual mutants in the ITC experiments takes place with inactive mutants, it would be valuable to design an experiment that would elucidate the difference in binding strengths using active enzymes. To do so, initial productive binding analysis of (GlcNAc)<sub>5</sub> in the presence of H<sub>2</sub><sup>18</sup>O labeled water was undertaken to determine the occupancy of the -3 to +2 subsites versus that of the -2 to +3 subsites, and hence, the relative strength of Arg<sup>172</sup>/Thr<sup>276</sup>, Ala<sup>172</sup>/Thr<sup>276</sup>, and Arg<sup>172</sup>/Ala<sup>276</sup> in binding a sugar moiety.<sup>222-223</sup> In short, productive binding from -3 to + 2 will yield a (GlcNAc)<sub>3</sub> moiety with an <sup>18</sup>OH at the reducing end and a (GlcNAc)<sub>2</sub> moiety with <sup>16</sup>OH at the reducing end. Productive binding from -2 to +3 will yield a (GlcNAc)<sub>3</sub> moiety with an <sup>16</sup>OH at the reducing end and a (GlcNAc)<sub>2</sub> moiety with <sup>18</sup>OH at the reducing end. The different occupancies can be determined using MALDI-TOF-MS.<sup>222-223</sup> ChiA-WT has a (GlcNAc)<sub>5</sub> occupancy of  $71 \pm 2$  % in the -2 to +3 subsites. The +3 subsite is a strong binding subsite,<sup>145</sup> and a shift to a higher -2 to +3 occupancy would indicate reduced binding affinity in the -3 subsite. This is exactly what

we observed for ChiA-T276A ( $99 \pm 4$  % occupancy in the  $-2$  to  $+3$  subsites), while ChiA-R172A was identical to that of the wild type ( $68 \pm 2$  %) within experimental error.

Moreover, the binding of allosamidin to ChiA-WT has previously been measured at pH 6.0 (20 mM potassium phosphate buffer) and 30 °C using ITC (Table 5.4).<sup>166</sup> In our study, the binding to ChiA-R172A and ChiA-T276A was studied under the same conditions for direct comparison. Figure 5.5 shows typical ITC thermograms and theoretical fits to the experimental data for each enzyme. For ChiA-R172A, the binding has a  $K_d = 0.067 \pm 0.008$   $\mu\text{M}$  ( $\Delta G_r^\circ = -9.9 \pm 0.1$  kcal/mol, Table 5.4). The reaction is accompanied by a  $\Delta H_r^\circ$  of  $-7.1 \pm 0.1$  kcal/mol and a  $\Delta S_r^\circ$  of  $9.2 \pm 0.3$  cal/K mol ( $-T\Delta S_r^\circ = -2.8 \pm 0.1$  kcal/mol). The binding between ChiA-T276A and (GlcNAc)<sub>6</sub> has a  $K_d = 1.0 \pm 0.1$   $\mu\text{M}$  ( $\Delta G_r^\circ = -8.3 \pm 0.1$  kcal/mol) and a  $\Delta H_r^\circ$  and  $\Delta S_r^\circ$  of  $-3.7 \pm 0.2$  kcal/mol and  $15.2 \pm 0.7$  cal/K mol ( $-T\Delta S_r^\circ = -4.6 \pm 0.2$  kcal/mol), respectively (Table 5.4).

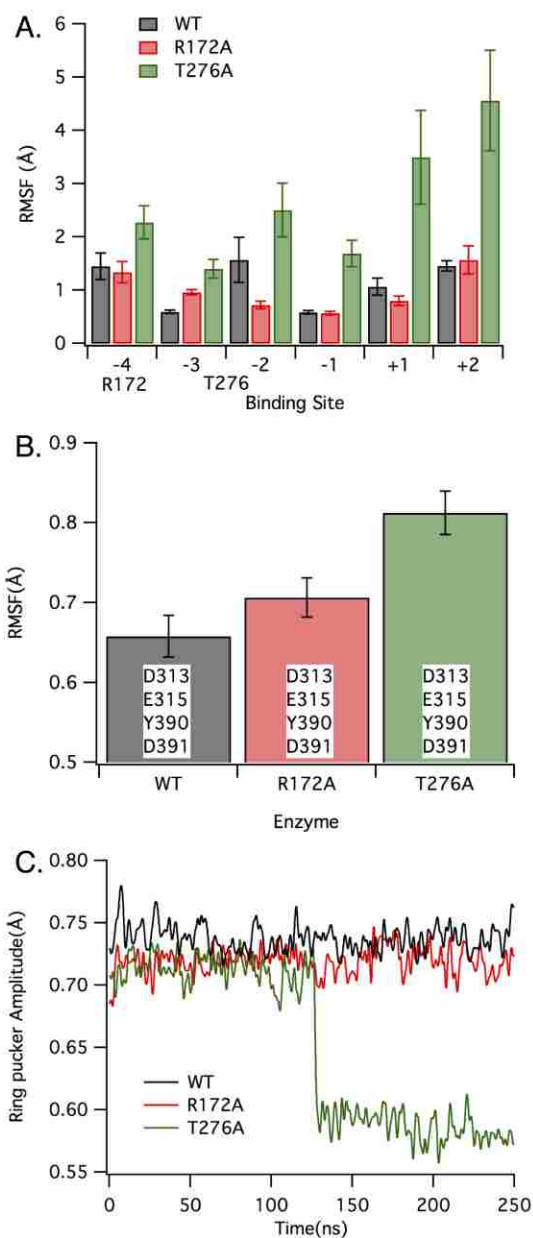
Interestingly, the binding affinity of both (GlcNAc)<sub>6</sub> and allosamidin to ChiA-WT and ChiA-R172A is virtually identical, while ligand binding affinity to ChiA-T276A is significantly weaker. Moreover, the reduced affinity of (GlcNAc)<sub>6</sub> and allosamidin for ChiA-T276A is due to less favorable enthalpy changes (2.6 and 3.8 kcal/mol, respectively), signifying a weaker binding interaction between the protein and the ligand when a threonine is substituted with an alanine.

## MD Simulations

MD simulations were performed to obtain molecular-level insight into the experimental observations of binding free energy and apparent processivity measurements. From the 250-ns MD trajectories of ChiA-WT, ChiA-R172A, and ChiA-T276A, we calculated three different quantities related to physical behavior in the active site. First, we calculated the root mean square fluctuation (RMSF) of the hexamer ligand on a per-binding-site basis (Figure 5.6A). This value represents the degree to which the ligand fluctuates in its binding site, as referenced against the average structure. Uncertainty was estimated using block averaging; the standard deviation of 2.5 ns blocks of data is shown in the figure. The ligand bound to ChiA-R172A behaves similarly to ChiA-WT. By comparison, the ChiA-T276A ligand fluctuates significantly more than either ChiA-WT or ChiA-R172A along the length of the active site.

In a similar fashion, the RMSF of the catalytic tetrad, Asp<sup>313</sup>, Glu<sup>315</sup>, Tyr<sup>390</sup>, and Asp<sup>391</sup>, reveals that the ChiA-T276A catalytic center exhibits a greater degree of freedom (Figure 5.6B). The tetrad of residues selected are either directly involved in catalysis or are known to play a key role in stabilization of the catalytically active conformation of the ligand, in which the  $\alpha$ -1 pyranose ring displays a <sup>1,4</sup>B boat conformation.<sup>45</sup> The RMSF value here again represents the average degree of fluctuation of these four residues alone, as compared to the average structure. Block averaging of 2.5 ns data blocks was used to determine uncertainty. The ChiA-R172A catalytic center fluctuates to the same extent as ChiA-WT. As we will discuss below, the prior two simulation observations are thought to be indicative of processive ability or lack thereof.<sup>73</sup>

The effect of the R172A and T276A mutations on the conformation of the ligand in the active site was also evaluated. Specifically, we were interested in the effect each mutation had on the ability of the enzyme to maintain the energetically unfavorable skew conformation in the -1 binding site. Ring distortion in the -1 binding site is a notable requirement of GHs, where the catalytic itinerary invariably passes through a skew, boat, envelope, or otherwise distorted conformation to effect catalysis.<sup>235-236</sup> ChiA-WT initially exhibits a <sup>1,4</sup>B conformation in the -1 binding site. Unless catalysis takes place, relieving distortion, the <sup>1,4</sup>B conformation should be maintained throughout the simulations. Periodic jumps to other conformations can be expected, but prolonged occupation of the relaxed chair conformation is suggestive of a fundamental change in the active site behavior, though not necessarily inactivation. To determine this conformation, we measured the Cremer-Pople ring pucker amplitude, which is a geometric measure of a ring's conformation, over the course of the 250-ns simulation (Figure 5.6C).<sup>237-238</sup> A boat conformation will have an amplitude of 0.73 Å, while a chair conformation will have an amplitude of 0.57 Å.<sup>238</sup> Both ChiA-WT and ChiA-R172A maintain the <sup>1,4</sup>B conformation of the -1 pyranose ring over the entire 250-ns simulation. At approximately 130 ns, the ChiA-T276A ligand relaxes to the chair conformation and never recovers the distortion. The latter behavior suggests that the T276A mutation affords the ligand a lower degree of affinity and more flexibility for relaxation; the length of time required to relax the -1 sugar conformation reflects the distance of the T276A mutation from the catalytic center (-1/+1 binding sites).



**Figure 5.6** Dynamics of the ChiA-WT, ChiA-R172A, and ChiA-T276A active sites as determined from 250-ns MD simulations. (A) Root mean square fluctuation (RMSF) of the chito-oligosaccharide (GlcNAc)<sub>6</sub> given on a per-binding-site basis. (B) RMSF of the four residues implicated in catalysis, either directly or indirectly: Asp<sup>313</sup>, Glu<sup>315</sup>, Tyr<sup>390</sup>, and Asp<sup>391</sup>. In both panels (A) and (B) uncertainty of the RMSF value was obtained through block averaging (2.5 ns blocks). (C) Cremer-Pople ring pucker amplitude of the -

1 site pyranose sugar ring over the entire 250-ns simulation. The boat conformation is represented by 0.73 Å, and chair conformation is represented by 0.57 Å.<sup>238</sup>

## 5.4 Discussion

The possibility that GH binding affinity and degree of processivity is correlated was recently described in a computational study of five processive Family 7 cellulases.<sup>61</sup> Using thermodynamics of chemical equilibrium and a previously defined statistical definition of processivity,<sup>64</sup> intrinsic processivity ( $P^{\text{Intr}}$ ), Payne *et al.* defined a mathematical relationship between ligand binding free energy and intrinsic processivity (Equation 5.2),

$$-\Delta G_{\text{b}}^{\circ} / RT = \ln (P^{\text{Intr}} * k_{\text{on}} / k_{\text{cat}}) \quad (5.2)$$

where  $R$  is the universal gas constant,  $T$  is the temperature,  $P^{\text{Intr}}$  is intrinsic processivity,  $k_{\text{on}}$  is the association rate coefficient, and  $k_{\text{cat}}$  is the catalytic rate coefficient. The relationship in Equation 5.2 is thought to be general to processive GHs; though, this has not been explicitly demonstrated. Further, the difficulties associated with validating the relationship in cellulases are multitude. Degree of processivity is difficult to accurately quantify, particularly for cellulases, as they do not make use of a substrate-assisted mechanism,<sup>74</sup> and the values vary significantly by laboratory. Until recently, the free energy of binding a cello-oligomer to the entire processive cellulase active site had also not been reported.<sup>160</sup> Even now, this information is available for only a single *T. reesei* cellobiohydrolase. Accordingly, Payne *et al.* used an enhanced sampling free energy methodology to calculate the binding free energies of five Family 7 GHs. Of these five enzymes, self-consistent processivity measurements were available for two cellulases,

which served as the basis for substantiating a link between calculated binding free energies and processivity.<sup>64</sup> Thus far, additional experimental evidence has not been made available by which to quantitatively or qualitatively confirm the relationship in Equation 5.2.

To describe how our results pertain to Equation 5.2, we must address the terms ‘apparent’ and ‘intrinsic’ processivity. The degree of processivity reported here are measurements of apparent processivity, a value that includes contributions from variables such as substrate heterogeneity and environmental conditions. Intrinsic processivity can be thought of as the theoretical limit of processive ability, under ideal conditions. While Equation 5.2 was developed with respect to intrinsic processivity, apparent processivity can be used to confirm a qualitative relationship. The experiments performed here have been conducted under the same conditions, and thus, confounding environmental variables will contribute equally to each measurement.

The ligand binding free energies of ChiA-WT, ChiA-R172A, and ChiA-T276A alongside measurements of apparent processivity provide the first complete experimental data set illustrating qualitative agreement with Equation 5.2. This relationship implies that the stronger an enzyme binds to the substrate, the more processive ability it will have, up to a maximum value, after which the enzyme becomes inhibited by the substrate. We therefore set out to examine the relationship between binding free energy and processivity using a model chitinase system, in which measurements of binding affinity and processivity are more straightforward than for cellulase systems. The mutations R172A and T276A were selected such that the effects of polar residues on binding free energy and processivity could also be assessed. Initial



[(GlcNAc)<sub>2</sub>/(GlcNAc)] ratios show that exchange of Arg<sup>172</sup> to Ala only slightly alters the initial degree of processivity ( $P^{\text{app}}$  of  $25.9 \pm 0.9$  vs.  $30.1 \pm 1.5$ ). The impact is significantly larger when Thr<sup>276</sup> is exchanged to Ala with a reduction in  $P^{\text{app}}$  to  $17.1 \pm 0.4$ . The T276A mutation clearly reduces the degree of processivity. As a comparison, the non-processive endo-chitinase ChiC, also from *S. marcescens*, display a  $P^{\text{app}}$  of  $14.3 \pm 1.4$ .<sup>78</sup> The free energy changes for binding (GlcNAc)<sub>6</sub> to ChiA-WT, ChiA-R172A, and ChiA-T276A indeed reveal a correlation between the degree of processivity and binding strength. ChiA-WT and ChiA-R172A bind (GlcNAc)<sub>6</sub> with approximately the same affinity, and the processive ability of ChiA-R172A is only moderately lower than the wild-type. The reduced affinity of ChiA-T276A is reflected in its reduced processive ability. This relationship is even clearer in examining the subsite occupancy determination of (GlcNAc)<sub>5</sub> and the binding affinity of the allosamidin inhibitor to the three enzymes. The subsite occupancy determination clearly shows that the affinity the -3 subsite has for the sugar moiety is greatly weakened when Thr<sup>276</sup> is exchanged to Ala compared to the Arg<sup>172</sup> to Ala exchange and the wild type. Furthermore, the slightly reduced affinity of ChiA-R172A for allosamidin, compared to wild-type, reflects the slightly lower processive ability. This phenomenon, inhibitor binding more closely trending with processivity, is likely related to the use of inactive variants in the (GlcNAc)<sub>6</sub> ITC measurements. It is suggestive of an experimental design approach, should binding affinity-based predictions of processive ability become a standard tool.

The changes in enthalpic and entropic components of binding free energy resulting from the T276A mutation indicate Thr<sup>276</sup> plays a significant role in the weak interactions, such as van der Waals interactions and hydrogen bonding, that maintain the

rigidity of the bound ligand. The less processive ChiA-T276A exhibits significantly reduced binding free energy change towards both (GlcNAc)<sub>6</sub> and allosamidin compared to the wild-type (0.5 kcal/mol and 2.2 kcal/mol, respectively). In general,  $\Delta H_f^\circ$  reflect changes in weak interactions (i.e. hydrogen bonds, electrostatic and polar interactions) between the ligand and the enzyme compared to those with the solvent. The differences in  $\Delta H_f^\circ$  for (GlcNAc)<sub>6</sub> and allosamidin binding to ChiA-T276A vs. the wild type are even more pronounced than the  $\Delta G_f^\circ$  values with a decrease of 2.6 and 3.8 kcal/mol respectively. Enthalpy-entropy compensation, where removal of a strong binding interaction may allow for a gain in entropy through more flexibility in the protein-substrate interactions, negates a full conversion of the reduced enthalpy change into free energy change.<sup>239</sup>

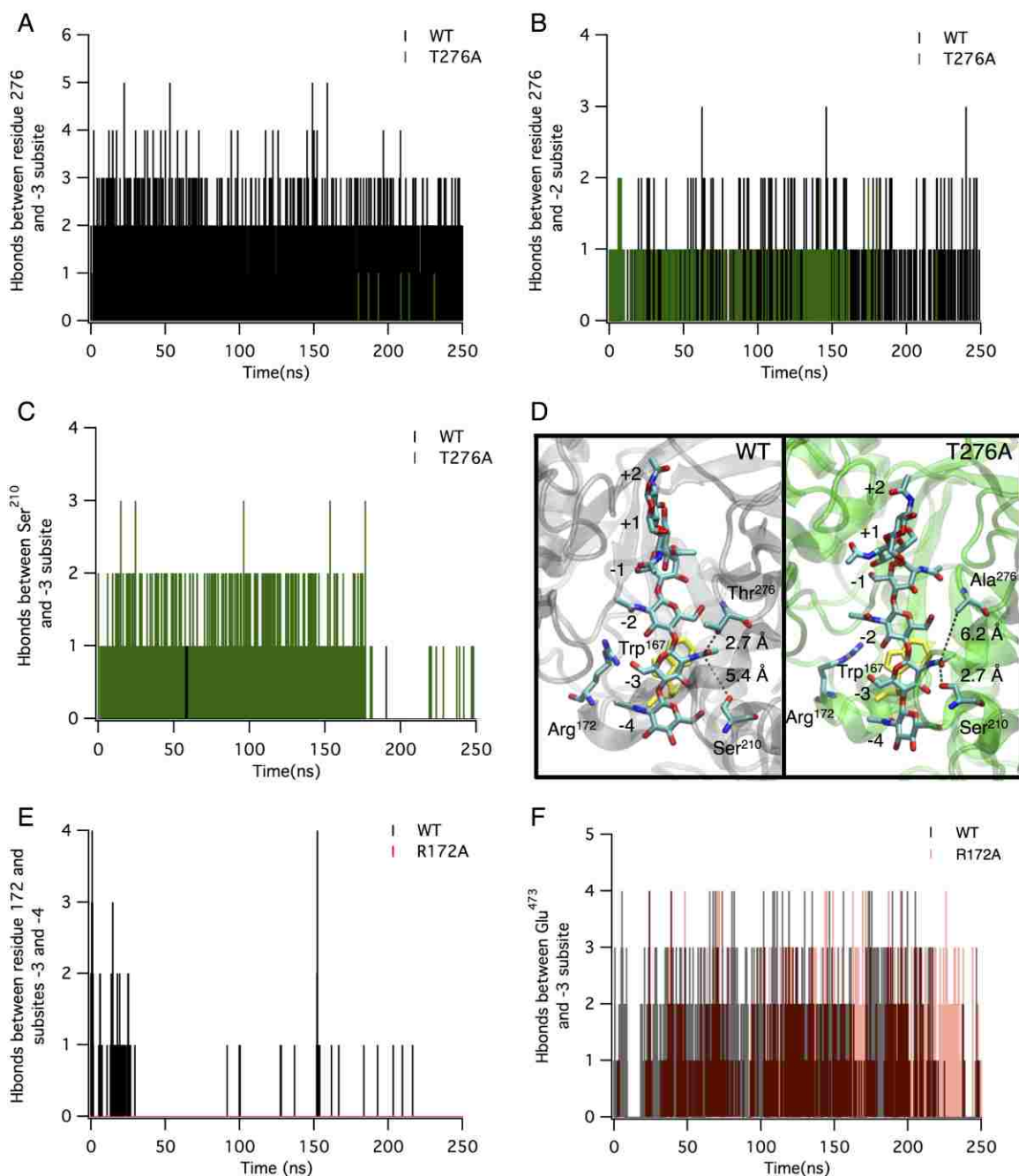
MD simulations support the observation that the active sites of ChiA-WT and ChiA-R172A behave similarly and that ChiA-T276A is a clear outlier. Four characteristic measurements of active site dynamics were determined from 250-ns MD simulations of the three enzymes: hydrogen bonding patterns, the RMSF of the bound (GlcNAc)<sub>6</sub> ligand, the RMSF of the protein catalytic center, and the -1 pyranose ring pucker amplitude. Each of these measurements indirectly reflects binding affinity and was expected to illustrate the above-described trend, namely that enzymes with more flexible/dynamic active sites will have a lower binding affinity and lower processive ability.

As previously mentioned, the ligand-bound ChiA crystal structure depicts an interaction of the Thr<sup>276</sup> OH functional group with the carbonyl oxygen of the -3 pyranose moiety. Along with Trp<sup>275</sup>, Thr<sup>276</sup> also mediates hydrogen bonding with the primary alcohol of the -2 site pyranose ring through the backbone.<sup>145</sup> Mutating Thr<sup>276</sup> to

alanine effectively abolished hydrogen bonding with the -3 sugar (Figure 5.7A). Backbone mediated hydrogen bonding with the -2 site was still feasible, though slightly reduced in number of hydrogen bonds formed (Figure 5.7B). The loss of interaction with the -3 site in the T276A ChiA variant greatly affects ligand stability in the -2 to +2 binding sites (Figure 5.6A). The lost hydrogen bond between the Thr<sup>276</sup> side chain and the -3 acetyl group is compensated, in part, by a new hydrogen bond between Ser<sup>210</sup> and the -3 side chain (Figure 5.7C). As Ser<sup>210</sup> is somewhat further from the -3 binding site (Figure 5.7D), the ligand relaxes to accommodate the new interaction. After the ligand fully relaxes (~175 ns), the hydrogen bond to the -3 acetyl group is almost completely lost. Additionally, the large aromatic residue, Trp<sup>167</sup>, directly opposite Thr<sup>276</sup> in the cleft (Figure 5.7D), maintains contact with the substrate to some extent. Both of these interactions, with Ser<sup>210</sup> and Trp<sup>167</sup>, allows the variant to bind (GlcNAc)<sub>6</sub> with slightly reduced affinity. As we will describe, the loss of the single hydrogen bond between the -3 carbonyl oxygen and the protein propagates to the dynamics along the binding cleft and greatly influences processive ability.

Similarly, the ligand-bound ChiA crystal structure depicts Arg<sup>172</sup> participating in a bifurcated hydrogen bond with the -3 pyranose and Glu<sup>473</sup>.<sup>145</sup> The structure also illustrates Arg<sup>172</sup> forming a hydrogen bond with the -4 pyranose, along with residues Gly<sup>171</sup>, Glu<sup>208</sup>, Ser<sup>210</sup>, and His<sup>229</sup>. On the basis of these structural observations, we hypothesized Arg<sup>172</sup> likely plays a role in substrate binding and that mutation to alanine would substantially reduce ligand interactions and affinity. However, MD simulation reveals that the hydrogen bond formed between Arg<sup>172</sup> and the GlcNAc ligand in the -3 binding site in wild-type ChiA is not persistent (Figure 5.7E). While mutation of Arg<sup>172</sup>

to alanine completely abolishes any hydrogen bonds between this residue and the -3 pyranose, the intermittent nature of hydrogen bonding in wild-type indicates the hydrogen bond rarely exists and is not critical to ligand binding. Rather, Glu<sup>473</sup> appears to be responsible for the bulk of hydrogen bonding with the -3 pyranose, with hydrogen bonding unaffected by the R172A mutation (Figure 5.7F). However, we anticipate the presence of Trp<sup>167</sup> alone, residing directly next to both Arg<sup>172</sup> and Glu<sup>473</sup>, is sufficient to maintain ligand affinity with the loss of hydrogen bonding from either residue. No significant difference in hydrogen bond patterns between subsite -4 and ChiA residues Gly<sup>171</sup>, Glu<sup>208</sup>, Ser<sup>210</sup>, and His<sup>229</sup> was observed upon the R172A mutation.



**Figure 5.7** Hydrogen bonds formed between ChiA residues and ligand subsites over 250-ns MD simulations. (A) Number of hydrogen bonds (hbonds) formed between wild-type ChiA Thr<sup>276</sup> and ChiA-T276A Ala<sup>276</sup> with the -3 subsite. (B) Number of hydrogen bonds formed between wild-type ChiA Thr<sup>276</sup> and ChiA-T276A Ala<sup>276</sup> with the -2 subsite. (C) Number of hydrogen bonds formed between wild-type ChiA Ser<sup>210</sup> and ChiA-T276A

Ser<sup>210</sup> with the -3 subsite. (D) 100-ns snapshots from wild-type ChiA and ChiA-T276A MD simulations illustrating the loss of the wild-type Thr<sup>276</sup>/<sub>-3</sub> acetyl group hydrogen bond and formation of the Ser<sup>210</sup>/<sub>-3</sub> acetyl group hydrogen bond upon mutation to alanine. The ChiA protein is shown in cartoon. The chito-oligomer and highlighted residues are shown in cyan and red stick. Trp<sup>167</sup> is highlighted in yellow stick. Distances are labeled in units of Å and indicated by dashed lines. (E) Number of hydrogen bonds formed between wild-type ChiA Arg<sup>172</sup> and ChiA-R172A Ala<sup>172</sup> with subsites -3 and -4. (F) Number of hydrogen bonds formed between wild-type ChiA and ChiA-R172A Glu<sup>473</sup> and the -3 subsite. In this final panel, the data sets are shown as translucent overlays to indicate the abundance of hydrogen bonding regardless of mutation.

The RMSF of the ligand is perhaps most closely related to the binding affinity; a ligand able to fluctuate to any significant degree in an enzyme binding site is unlikely to be strongly bound. In the case of the T276A variant, the (GlcNAc)<sub>6</sub> ligand fluctuates significantly more across the length of the cleft than ChiA-WT, while the ChiA-R172A ligand behaves roughly the same as ChiA-WT (Figure 5.6A). The results align with experimental measurements of binding affinity. The R172A mutation has little effect on the overall stability of the ligand, again indicating that the hydrogen bonds formed with the substrate in the -3 and -4 binding sites are not critical to binding. Rather, the remaining hydrogen bond with Glu<sup>473</sup> is sufficient to maintain stable binding. On the other hand, the broken hydrogen bond in subsite -3 resulting from the T276A mutation cannot be compensated by the surrounding protein, and the localized range of freedom translates across the length of the active site.

The RMSF of the protein catalytic center, residues Asp<sup>313</sup>, Glu<sup>315</sup>, Tyr<sup>390</sup>, and Asp<sup>391</sup>, also indicates that protein fluctuation correlates with binding affinity, wherein increased fluctuation corresponds to lower binding affinity. These catalytic residues are responsible for maintaining a suitable reactive ligand conformation in the  $-1/+1$  binding site.<sup>87,90</sup> We observe that the fluctuation of the catalytic tetrad in ChiA-R172A is within error of ChiA-WT, but that ChiA-T276A fluctuates significantly more in this critical region of the active site (Figure 5.6B). Previously, we hypothesized that both RMSF of the ligand and the catalytic center is a molecular “hallmark” of processivity.<sup>73</sup> This prior study focused on the delineation between processive and non-processive chitinases rather than variations as a result of mutagenesis. The results we present here confirm the observed relationship between active site dynamics and processive ability and extend its relevance to mutations of the same enzyme. The link between catalytic center flexibility and processivity is likely a result of the need, either evolutionarily or engineered, to associate and dissociate quickly from the substrate. Flexibility in the active site affords the enzyme with momentum to escape chemical attraction.

Related to the dynamics of the catalytic center, the conformation of the pyranose sugar in the  $-1$  binding site illustrates the new dynamics imposed by the R172A and T276A mutations. The Cremer-Pople ring puckering amplitude of the  $-1$  site pyranose ring was calculated over the 250-ns MD simulations for all three enzymes (Figure 5.6C).<sup>237</sup> Structural studies indicate that the ChiA ligand forms an approximate boat conformation along the catalytic itinerary of the substrate-assisted mechanism.<sup>87</sup> Both ChiA-WT and ChiA-R172A maintain the structural conformation of the ring in the  $-1$  binding site over the length of the simulation. However, ChiA-T276A allows the  $-1$

pyranose to relax to the chair conformation after approximately 130 ns. The catalytic conformation is never recovered, indicating that the active site does not maintain as close an association with the substrate in the  $-1$  binding site. This is particularly suggestive of the role T276A indirectly plays in catalysis. As previously mentioned, T276A primarily interacts with the substrate in the  $-2$  and  $-3$  binding subsites, yet clearly affects the dynamics of the entire binding site.

Chitosan degradation experiments suggest that Arg<sup>172</sup> is responsible in part for the recognition and positioning of the oligomer in the active site. In the case of ChiA-R172A, the variety in oligomeric reaction products produced by the variant implies the enzyme had no strong preferences for either acetylated or deacetylated units in subsites other than  $-1$ . In comparison, wild-type ChiA requires an acetylated moiety in subsites  $-1$ ,  $-2$ , and  $-4$ . This result is particularly interesting, as Arg<sup>172</sup> appears to interact with as many as three different sugar moieties ( $-2$ ,  $-3$ , and  $-4$ ). Thus, despite having no role in processive activity, Arg<sup>172</sup> is central in the recognition and positioning of the substrate into the active site.

Similarly, chitosan degradation experiments reveal that the strong interacting ChiA-T276 that mainly binds with the  $-2$  sugar does not appear to significantly affect the positioning of the substrate. Rather, the only prominent change in the chitosan degradation profile resulting from the T276A mutation is the more stringent requirement for an acetylated unit in the  $-3$  binding site. This arises as a result of compensatory interactions made within the ChiA-T276A binding site that enables the ligand to remain bound. Naturally, one might think that deletion of the Thr<sup>276</sup> side chain, which forms a hydrogen bond with the  $-3$  *N*-acetyl group would randomize the acceptance of acetylated



or deacetylated groups in the mutant binding site; however, to maintain sufficient contact with the enzyme active site, a new hydrogen bond is formed between the -3 acetyl group and the nearby Ser<sup>210</sup> (Figure 5.7C and 5.7D). Until the relaxation of the entire ligand (following alleviation of ring distortion at ~130 ns), Ser<sup>210</sup> takes over ligand stabilization, allowing a modicum of remaining activity.

Finally, processive ability correlates with the rate and extent of chitin degradation. The enzymes with a higher initial degree of processivity are both faster and more efficient degraders of  $\beta$ -chitin (Figures 5.3 and 5.4). Drastic reduction in rate of hydrolysis and efficiency combined with reduction in processivity has previously been shown for aromatic residues in both ChiA and ChiB from *S. marcescens*.<sup>69-70</sup> This result supports the notion that overall substrate turnover is improved by processivity, which results from the ability to maintain proximity with the substrate after each catalytic event.

## 5.5 Conclusions

From both simulation results and experimental evidence, we suggest that T276 is critical to formation of a stable, processive chitinase, while R172 does not play a significant role in processive action. Prior studies have suggested that both hydrophobic stacking interactions and hydrogen bonding are important for a smooth processive cycle,<sup>215</sup> and our results indicate that beyond that canonical carbohydrate-aromatic stacking interactions, polar residues can play a key role in this process. In particular, Thr<sup>276</sup> in the ChiA active site is important to processive function. The fact that Arg<sup>172</sup> does not participate in processive function indicates that replacement could be beneficial for enhanced processive action. Furthermore, our findings provide the first qualitative experimental characterization of the hypothesized relationship of binding affinity to GH

processive action.<sup>61</sup> The free energy of binding chito-oligosaccharides and allosamidin are indicative of processive ability. Perhaps the most interesting finding was that the inhibitor binding provided a more sensitive characterization of the relationship between binding and processivity. On the whole, our results provide new and valuable insight into the role polar residues along the length of a GH active site contributes to molecular interactions, substrate binding, and processivity in chitinases. More broadly, we suggest our results represent an important first step toward validating a hypothesized relationship that potentially describes the action of an entire class of GHs and greatly compliments our understanding of how GHs are able to depolymerize recalcitrant polysaccharides.

## **Chapter 6 - Carbohydrate-Polar Interactions Dictate Substrate Binding and Processivity in a Family 18 Chitinase**

The activity and processivity for the wild-type chitinase and the aromatic and polar variants, reported in Chapter 6, were measured by our experimental collaborators at the Norwegian University of Life Sciences. The author of this dissertation performed MD simulation as well as free energy calculations and related those to the experimental processivity data. The polar residues investigated here were selected by the author of this dissertation with the help of his dissertation advisor, Dr. Christina M. Payne, based on interaction energy calculations from MD simulations. Copyright © Suvamay Jana 2017.

### **6.1 Introduction**

Cellulose, an insoluble  $\beta$ -1, 4-linked homo polymer of Glc and Chitin, an insoluble  $\beta$ -1, 4-linked homo polymer of GlcNAc are the two most abundant resources of renewable carbon in biosphere. Cellulose is the major structural component of plant cell wall, whereas chitin is the key component of fungi cell wall and the exoskeleton of arthropods.<sup>25, 168</sup> These polysaccharides are highly crystalline in nature owing to their complex networks of intra- and inter- molecular hydrogen bonds and stacking interactions; this provides strong resistance to microbial and chemical attack.<sup>13</sup> Nature evolves cocktails of enzymes, including GHs, to break down these recalcitrant polysaccharides to soluble sugars.<sup>16</sup> Since, the end products of the enzymatic depolymerization of these polysaccharides are of high value in chemical and clinical applications,<sup>3, 26, 240</sup> understanding GH catalytic mechanism and GH carbohydrate interactions are essential to design GHs for efficient biomass conversion.

The GHs responsible for cellulose and chitin degradation can be classified either based on GH mode of action or based on GH processivity.<sup>65, 77</sup> Mode of action depends on how GHs attack the polysaccharide substrate prior to acquisition of the substrate chain into the active site, which includes either random cleavage of the polymeric substrate (endo mode) or cleavage from the free chain ends available in the substrate matrix (exo mode). On the contrary, GH processivity is defined as upon substrate binding how many times GHs perform catalysis before finally dissociating from the substrate. Based on processivity, GHs are either processive (hydrolyze glycosidic bonds many times in series) or non-processive (hydrolyze glycosidic bonds once or for very few times). Non-processive GHs usually follow endo mode of action to acquire a substrate chain into the active site prior to hydrolysis, whereas processive GHs can accommodate a substrate chain via endo or exo fashion.<sup>42</sup> Overall, processive GHs remain attached to the polysaccharide substrate between subsequent hydrolytic steps, and hence, provide the majority of hydrolytic potential for efficient conversion of crystalline polysaccharides.<sup>23,</sup>

69

A characteristic feature of processive GHs is the abundance of aromatic residues in the active site, usually favored over other hydrophobic residues (aliphatic) in that position.<sup>241</sup> These residues in general are highly conserved and their interaction with the carbohydrate in the active site is well defined. Experimental and computational studies previously showed that the aromatic residues interact with the carbohydrate substrate via hydrophobic stacking and strong electrostatic CH- $\pi$  interactions that facilitate substrate recognition and binding.<sup>174-175, 177-179, 241-242</sup> Additionally, the large surface area of the aromatic residues provides a tight binding, yet fluid-like, platform that is essential for

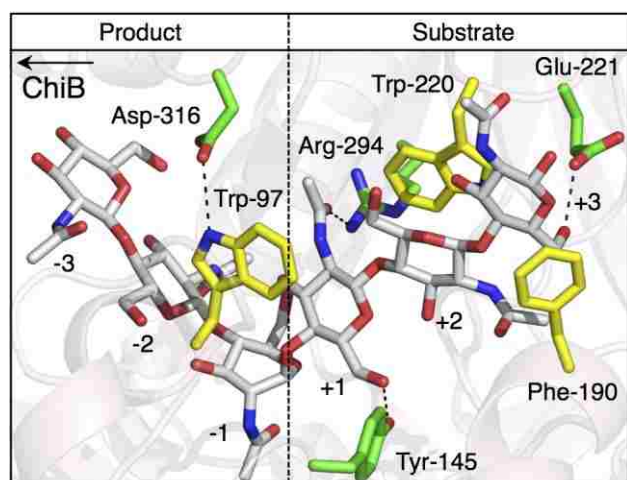
processive sliding of the substrate in the GH active site.<sup>44, 210</sup> Deletion of aromatic residues in the GH active site often leads to lower activity towards crystalline substrate, lower substrate binding, change in mode of action, and reduction in processivity.<sup>69-70, 183-184, 201-202</sup>

Another noteworthy feature of GH active sites is the abundance of polar residues alongside the conserved aromatic residues. Several of these polar residues participate in catalysis,<sup>87-88, 198-199</sup> while others may be involved in protein stabilization and binding. Previously, Bu *et al.* identified three polar residues – Arg-251, Asp-259, and Asp-262 in cellulose degrading processive cellulase *Trichoderma reesei* Cel7A, having significant contributions in product inhibition.<sup>214</sup> These residues were identified as high-affinity binding residues in the product-binding site, the removal of which is anticipated to ease product expulsion. In another study of different a Family 43 GH, Proctor *et al.* found that the polar residues Gln-316 and Asp-53, at the -3 subsite of *Cellvibrio japonicus* Arb43A, play an important role in dictating the GH mode of action.<sup>213</sup> Recently, we demonstrated that in Family 18 *Serratia marcescens* processive chitinase ChiA, Thr-276, at the -3 site, is a key residue in substrate binding and processivity;<sup>190</sup> the deletion of this residue lowers substrate affinity and processivity in comparison to the wild-type ChiA. Outside these studies, knowledge regarding the contribution of polar residues in binding, mode of GH action and processivity is severely limited, and further experimental and computational analyses are needed to explore their roles. Here, we examine the function of polar residues in carbohydrate recognition, binding, activity, and processivity of *Serratia marcescens* Family 18 chitinase B (ChiB).

ChiB is an exo-acting, non-reducing end specific enzyme that acts on crystalline chitin.<sup>82-83</sup> However, the mode of action changes from exo- to endo- in presence of soluble chitosan substrate.<sup>84</sup> The deep catalytic cleft of ChiB has four substrate subsites (+4 to +1) and three product subsites (-1 to -3) in the catalytic domain (CD).<sup>81, 138</sup> The carbohydrate binding module (CBM), fused with the CD, provides additional substrate binding sites beyond the four substrate subsites in the CD. The role of several aromatic residues across different binding sites of ChiB, including in the CD and CBM, has previously been examined. For example, in the ChiB CD, Trp-97 (+1 site) and Trp-220 (+2 site) were found to be critical for substrate binding, while Phe-190 at the +3 site seems to have little role in ligand binding (Figure 6.1).<sup>69, 112</sup> Similarly, several tryptophan and tyrosine residues in the ChiB CBM were found to be crucial in substrate recruitment and hydrolysis.<sup>201</sup> However as mentioned above, the information regarding polar residues in the ChiB active site is ill defined, unlike aromatic residues. Accordingly, information learnt from this current study can be applied toward understanding and generalizing the role of polar residues in other families of GH.

Here, we focus on four polar residues across the ChiB cleft, including Asp-316 (-1 site), Tyr-145 (+1 site), Arg-294 (+1 site), and Glu-221 (+3 site), as shown in Figure 6.1. The selection of the residues was based on their position and interaction with the substrate and the adjacent amino acids in the crystal structure. In the crystal structure (PDB 1E6N),<sup>79</sup> Asp-316 does not interact with the substrate directly, but rather, it interacts with Trp-97 opposite the ChiB binding cleft via a hydrogen bond (Figure 6.1). It has long been thought that this putative roof formation over the -1 site is mediated by hydrogen bonding between Asp-316 and Trp-97 facilitates tight binding and processive

sliding of the substrate in the GH active site.<sup>65</sup> Other polar residues included in this study appear to interact directly with the chito-oligosaccharide substrate in the ChiB binding cleft. Tyr-145 and Arg-294 form a hydrogen bond with the +1 site sugar, while Glu-221 forms a hydrogen bond with the +3 site sugar (Figure 6.1). We hypothesize that the hydrogen bonds and strong electrostatic interactions from the polar residues stabilize the ligand during hydrolysis to facilitate binding and processivity. To test our hypothesis, we examined interactions of wild-type (WT) and polar variants of ChiB bound to a hexameric chito-oligosaccharide ligand, (GlcNAc)<sub>6</sub> occupying the -3 to +3 subsites using MD simulations and computational free energy calculations. In addition, we conducted biochemical studies to measure activity and apparent processivity of the variants of the selected polar residues on crystalline  $\beta$ -chitin substrate. In addition to the four polar residues investigated in this study, we measured activity and apparent processivity of three aromatic variants: W97A (+1 site), W220A (+2 site), and F190A (+3 site), whose contribution in (GlcNAc)<sub>6</sub> binding has previously been determined.<sup>112</sup> The overall characterization of polar residues in the ChiB cleft along with the aromatic residues will demonstrate how specific polar hydrogen bonding and nonspecific aromatic stacking facilitate substrate recruitment, binding and processivity.



**Figure 6.1** The active site of Family 18 *S. marcescens* processive ChiB, populated with polar and aromatic residues interacting with a hexameric chito-oligosaccharide ligand via hydrogen bonding and hydrophobic stacking. The ChiB protein is shown in transparent pink cartoon. The (GlcNAc)<sub>6</sub> ligand is shown in gray sticks. The polar residues investigated in this study (Asp-316, Tyr-145, Arg-294, and Glu-221) are highlighted in green sticks. Aromatic residues are shown in yellow sticks. The dashed lines represent the hydrogen bonding between the polar residues with the ligand and the other amino acids of ChiB. The arrow indicates the direction of substrate sliding through the ChiB active site. The coordinates of the protein were obtained from the 1E6N PDB structure, and the coordinates of the hexameric ligand were obtained from structural alignments of 1E6N and 1OGG in PyMOL.<sup>79, 85, 144</sup> Hydrolysis occurs between -1 and +1 site.

## 6.2 Materials and methods

### 6.2.1 Chemicals

Chito-oligosaccharides (CHOS) were obtained from Megazyme (Wicklow, Ireland). Squid pen  $\beta$ -chitin was purchased from France Chitine (180  $\mu$ m



microparticulate, Marseille, France). All other chemicals were of analytical grade and purchased from standard manufacturers.

## **6.2.2 Enzymes**

In this study, the following residues in the active site of ChiB from *S. marcescens* were studied: Asp-316, Trp-97, Tyr-145, Arg-294, Trp-220, Phe-190, and Glu-221. All residues were mutated to alanine single mutants. Previously, ChiB-W97A and ChiB-W220A were made.<sup>69</sup> Initial characterization of ChiB-Y145A, close to the catalytic acid, resulted in inactivation.

### **6.2.2.1 Site directed mutagenesis**

Mutation of ChiB-D316A, ChiB-R294A, ChiB-Y145A, ChiB-F190A, and ChiB-E221A were performed using the QuikChange<sup>TM</sup> site directed mutagenesis kit from Stratagene (La Jolla, CA, USA), as previously described.<sup>119</sup> The mutated genes were sequenced using GATC Biotech's (Constance, Germany) LIGHTrun Sequencing service to verify the introduction of the desired mutations. All primers and templates used for the mutagenesis are listed in Table 6.1. Primers were purchased from Life Technologies (Carlsbad, CA, USA).

**Table 6.1** Primers used for site directed mutagenesis

Mutant	DNA template	Primer	Sequence
ChiB-D316A	ChiB-WT <sup>186</sup>	D316A fw	5'-CGCCGGGCGAAGCGCCGTATCCGAG-3'
		D316A rev	5'-CTCGGATACGGCGCTTCGCCCCGGCG-3'
ChiB-R294A	ChiB-WT <sup>186</sup>	R294A fw	5'-GCCCTTCTATGGCGCCGCCTTCAAGGGC
		R294A rev	5'-CGGGAACCTCCGCCCGCGTATCTTCCCG-3'
ChiB-Y145A	ChiB-WT <sup>186</sup>	Y145A fw	5'-CATCGACTGGGAGGCCCCGCAAGCGGGCG-3'
		Y145A rev	5'-CGCCGCTTGCGGGGCCTCCAGTCGATG-3'
ChiB-F190A	ChiB-WT <sup>186</sup>	F190A fw	5'GCCGGCGGCGCCGCCTTCCTGTTCGCG'3
		F190A rev	5'CGCGACAGGAAGGCGGCGCCGCGCCGGC'3
ChiB-E221A	ChiB-WT <sup>186</sup>	E221A fw	5'-GCCGGCCCCTGGGCGAAGGTAACCAAC-3'
		E221A rev	5'-GTTGGTTACCTTCGCCCAGGGGCCGGC-3'

### 6.2.2.2 Protein expression and purification of single, active mutants

The active mutants were expressed in *E. coli* as described previously.<sup>186</sup> Purification was performed as described elsewhere.<sup>78</sup>

### 6.2.3 Time course activity assay

Hydrolysis of chitin (20 mg/ml) was carried out in 50 mM sodium acetate at pH 6.1. Previously, we have shown that 20 mg/ml chitin gives substrate saturating conditions.<sup>243</sup> The chitin samples were sonicated for 20 min in a sonication bath (Transsonic Elma) to increase the surface of the substrate and thereby increase the availability of chitin ends for the enzymes.<sup>244</sup> The reaction tubes were incubated at 37 °C in an Eppendorf thermomixer at 800 rpm to avoid settling of the chitin particles. The enzyme concentrations were 100 nM in all experiments. Aliquots of 75 µl were withdrawn every hour for 7 hours, and the enzymes were inactivated by adding 75 µl 20 mM H<sub>2</sub>SO<sub>4</sub>. Prior to further HPLC analysis, all samples were filtered through a 0.45 Duapore membrane (Millipore) to remove denatured protein and chitin particles. All reactions were run in 3-6 replicates, and all samples were stored at -20 °C until HPLC analysis. The degree of degradation is defined as the percentage of number of moles

solubilized GlcNAc units with respect to number of moles GlcNAc units in solid form (chitin) used in the experiments.

#### **6.2.4 High performance liquid chromatography (HPLC)**

Concentrations of mono- and disaccharides were determined using HPLC with a Rezex Fast fruit H<sup>+</sup> column (100 mm length and 7.8 mm inner diameter) (Phenomenex). An 8 µl sample was injected on the column, and the mono- and disaccharides were eluted isocratically at 1 mL/min with 5 mM H<sub>2</sub>SO<sub>4</sub> at 85 °C. The mono- and disaccharides were monitored by measuring absorbance at 210 nm, and the amounts were quantified by measuring peak areas. Peak areas were compared with peak areas obtained with standard samples with known concentrations of mono- and disaccharides.

#### **6.2.5 Determination of A- and b-values for activity**

According to Kostylev and Wilson, the following two-parameter kinetic model can be used to determine a time-dependent activity constant:

$$X = At^b \tag{6.1}$$

where  $t$  is time (h),  $X$  is % digestion,  $A$  is net activity of the added enzyme, and  $b$  is an intrinsic constant that quantifies the curvature of the time course profile.<sup>245</sup>  $A$  and  $b$  values were determined after fitting the data to Equation 6.1 by use of Origin 7. The data were combined in one plot before fitting.

## **6.2.6 Degradation of chitin for determination of enzyme processivity**

Hydrolysis of chitin was carried out as described previously.<sup>78</sup> The extent of degradation is defined as described in Section 6.2.3, and the concentration of mono- and disaccharides were determined as described in Section 6.2.4.

## **6.2.7 Computational methods**

MD simulations were performed along with free energy calculations to determine the role of selected polar amino acid residues present in the catalytic cleft of ChiB in protein ligand dynamics and binding affinity. The detailed protocols used in this study including system construction, MD simulations, and free energy calculations are described below.

### **6.2.7.1 Construction of ChiB model systems**

The coordinates for MD simulations and free energy calculations were obtained from a combination of ChiB crystal structures (PDB entries 1E6N and 1OGG).<sup>79, 144</sup> Nine ChiB simulation sets were constructed in total in this study. Five of them included models of ChiB wild-type and four polar variants (D316A, Y145A, R294A, and E221A) all bound to a hexameric chito-oligosaccharide ligand, (GlcNAc)<sub>6</sub>, occupying the +3 (entrance) to -3 (exit) sites in the catalytic cleft (Figure 6.1). A sixth model of ChiB wild-type was also constructed bound to an  $\alpha$ -chitin microfibril (Figure A3.1B of Appendix A3). Differences in dynamical behaviors of the bound protein in solution and in the presence of crystalline substrate will be considered. The remaining three simulation sets included: ChiB wild-type with no ligand, ChiB wild-type complexed with a (GlcNAc)<sub>6</sub> ligand in solution with an harmonic restraint applied to residues Trp-97 and Asp-316 to

maintain a fixed distance, and a model of ChiB wild-type bound to an  $\alpha$ -chitin microfibril with applied harmonic restraints on residues Trp-97 and Asp-316. The last three simulations were constructed for free energy calculation setup and will be discussed further below. In all cases, the coordinates for the protein came from the 1E6N structure, the coordinates for the hexameric ligand were obtained from the aligned structures of 1E6N and 1OGG, and the coordinates for the  $\alpha$ -chitin microfibril came from the crystal structure replicated along x, y, and z directions.<sup>37, 79, 85, 144, 246</sup>

To construct the wild-type model, the E144Q mutation in the 1E6N structure was reverted to the wild-type Glu, followed by manual rotation of the catalytic residues Asp-142 and Glu-144 and the *N*-acetyl side chain of the -1 site of the ligand to reflect the catalytically competent Michaelis complex.<sup>45</sup> Minimization was performed to ensure the rotation did not affect the stability of the protein or the ligand (100 steps of steepest descent followed by 100 steps adopted basis Newton Raphson in vacuum). The variant simulations were constructed by point mutating the selected wild-type polar residue to alanine. The wild-type apo model was constructed by removing the ligand from the ChiB active site. In all cases, only the catalytic domain (up to Leu-448) was modeled instead of the full-length enzyme to reduce the computational time, which has been successful in prior simulation studies.<sup>73</sup>

After construction of the initial models, protonation states were determined at pH 6.0 using the H++ software package.<sup>148-150</sup> All of the systems were then solvated with explicit water molecules, and sodium ions were added to ensure charge neutrality. The overall system size for each solvated enzyme without the microfibril was approximately

80 Å x 80 Å x 80 Å (52,000 atoms); with the microfibril, the solvated systems were approximately 120 Å x 80 Å x 80 Å (75,000 atoms).

### 6.2.7.2 MD simulations

After solvation, a three step minimization protocol, in CHARMM, was used to eliminate initial bad contacts between the overlapping atoms:<sup>95</sup> (1) 10,000 steps of steepest decent (SD) minimization were performed on the entire system while keeping the protein and ligand (if present) fixed allowing the water molecules to move freely; (2) 10,000 SD minimization steps were then performed, allowing the protein and the water to move freely and keeping the ligand fixed; and (3) 10,000 steps of SD minimization were performed on the entire system with no imposed restraints. The CHARMM36 all-atom force field with CMAP corrections was used to model the proteins,<sup>95, 151-152</sup> the CHARMM36 carbohydrate force field was used to model the (GlcNAc)<sub>6</sub> ligand,<sup>194-195</sup> and the modified TIP3P force field was used to model water.<sup>153-154</sup> After minimization, the systems were heated from 100 K to 300 K in the *NVE* ensemble for 20 ps, followed by density equilibration in the *NPT* ensemble for 100 ps in CHARMM. After density equilibration, 250-ns production simulations were performed at 300 K in the *NVT* ensemble using the NAMD software package.<sup>96</sup> However, the three simulations wherein a harmonic restraint was applied between Trp-97 and Asp-316, the apo wild-type and wild-type ChiB complexed with the ligand and the microfibril, were run for only 25 ns in NAMD following *NPT* equilibration in CHARMM. In all MD simulations, the nonbonded cutoff distance used was 10 Å with a switching distance of 9 Å and a nonbonded pair list distance of 12 Å. SHAKE was used to fix all the hydrogen bond distances. Finally, VMD was used to visualize all the simulation trajectories.<sup>155</sup>

### 6.2.7.3 TI

To measure the relative change in binding free energy ( $\Delta\Delta G_{TI}$ ) due to mutation, a 25-ns equilibrated snapshot from the wild-type MD simulations (apo and ligand bound states), free from any restraint, was used to initiate the TI calculations in NAMD. The “dual-topology” approach was applied in NAMD to complete the TI calculations (discussed in details in Chapter 2).<sup>96,94, 196-197</sup>

Previously we demonstrated the successful implementation of this method to calculate  $\Delta\Delta G_{TI}$  for aromatic to alanine mutations in the catalytic cleft of ChiA and ChiB (Chapter 5).<sup>112</sup> However, unlike aromatic residues, implementation of the TI scheme toward charged polar residues to measure  $\Delta\Delta G_{TI}$  is not straightforward. Computational challenges arise as a result of the transformation of charged polar residue side chains to that of a neutral alanine, resulting in non-neutrality of the overall system. In non-neutral system, the periodic images repel each other electrostatically creating an unwanted self-energy artifact that leads to inaccurate free energy estimation.<sup>247-248</sup> Hence, to evaluate the free energy change accurately for charged amino acid to neutral alanine mutations, the self-energy term must be removed.<sup>249-250</sup>

The approach often used for charge mutations is to avoid the self-energy problem by performing two simultaneous mutations, ensuring system neutrality in both initial and final states.<sup>249, 251-254</sup> Either a residue far from the area of interest or a neutral dummy atom in the system can be mutated to compensate for the change in charge.<sup>249, 251</sup> However, Zhou *et al.* found that counter-mutating a distant residue to the opposite charge is probably more suitable than the incorporation of ion mutations, as ion mutations usually result in a significantly larger standard deviation due to solvation free energy

effects of the ions.<sup>251</sup> To determine  $\Delta\Delta G_{TI}$  for charged residue mutations here (D316A, R294A, and E221A), we used the dual counter-mutation approach, where a distant polar residue was protonated or deprotonated simultaneously with the charged polar residue to alanine mutation (two hybrid residues were incorporated instead of one). For charged polar residue mutations, D316A and E221A, residue Lys-42 was deprotonated simultaneously. For the R294A mutation, Asp-426 was protonated simultaneously. The distant residues, Lys-42 and Asp-426, were chosen for counter-mutation, as these residues were far away from the ChiB active site cleft with little correlation to the region of interest. A traditional TI approach was used for the neutral polar residue mutation Y145A, similar to that of tryptophan/phenylalanine to alanine mutations we conducted previously.<sup>86, 112, 183</sup>

To calculate free energy changes associated with ligand bound ( $\Delta G_{Bound}$ ) and ligand free states ( $\Delta G_{Free}$ ), 15  $\lambda$  windows were chosen ranging from 0 to 1 for smooth transformation of the reactant (wild-type polar residue) to the product (mutant alanine). Electrostatic and van der Waals components of the free energy were calculated separately for computational efficiency. For each  $\lambda$  window, 1000 minimization steps were performed first, followed by 15-ns production simulations in the *NVT* ensemble. The simulation parameters used for the TI calculations were same as the MD simulations described above. After production simulations, first 0.5 ns data was discarded before collecting 14.5 ns of TI data. TI data was further integrated numerically using the cubic spline Gaussian quadrature method over 15  $\lambda$  windows to obtain  $\Delta G_{Bound}$  and  $\Delta G_{Free}$ . Error was calculated following the method of Steinbrecher et al.<sup>111</sup> The detailed free energy components, including electrostatics and van der Waals contributions at each



state, are provided in the Table A3.1 of Appendix A3.

#### 6.2.7.4 Umbrella sampling (US)

Residues Trp-97 and Asp-316 in the ChiB 1E6N crystal structure appear to interact via hydrogen bonding (the nitrogen of the indole ring of Trp-97 with the carbonyl oxygen of Asp-316), which results in the formation of a putative roof over the -1 site of the ChiB cleft (Figure 6.1).<sup>79</sup> Historically, it is believed that the roof over -1 site forces ChiB to act in exo-mode to attack the substrate. In addition, closure of the ChiB cleft is assumed to facilitate processive movement of ChiB. However, this roof is partially or completely absent in several crystal structures, where ChiB was bound to an inhibitor allosamidin (PDB 1OGG) or in the ChiB apo structure (PDB 1E15).<sup>134, 144</sup> Additionally, biochemical studies revealed that ChiB could act in endo-mode in the presence of soluble chitosan substrate,<sup>84</sup> implying the roof interaction may persist. To quantitatively determine whether the formation of the roof over the -1 site is thermodynamically favorable, we conducted umbrella sampling simulations in NAMD using the distance between the hydrogen (HE1) of Trp-97 and the Asp-316 oxygen (OD2) as our reaction coordinate. The initial coordinates for umbrella sampling were two 25-ns MD snapshots of ChiB wild-type models with and without the  $\alpha$ -chitin microfibril, where harmonic restraint was applied on residues Trp-97 and Asp-316 to maintain the active site roof conformation throughout an equilibration period.

To perform umbrella sampling simulations, we used 15 windows, ranging from 1.8 Å to 8.5 Å, with an interval of 0.5 Å along the reaction coordinate. A harmonic potential with a force constant of 10 kcal/mol/Å<sup>2</sup> was used to maintain the reaction coordinate in each window. Each window was run for 10 ns using the same simulation

parameters used in the MD simulation and TI calculations. The potential of mean force (PMF) was constructed using the weighted histogram analysis method (WHAM) code of Alan Grossfield.<sup>116</sup> Associated error was also calculated with the method of bootstrapping. The histograms and auto correlation function (ACF) used to generate the PMF and to calculate the errors are provided in Figure A3.1 of Appendix A3.

## 6.3 Results

### 6.3.1 Initial kinetic characterization

Previously, methods with short oligosaccharides or non-natural substrates have been used to characterize the activity of chitinases.<sup>189, 221</sup> In this study, the two-parameter kinetic model proposed by Kostylev and Wilson has been adapted to chitin in order to initially characterize the mutants with respect to activity on a natural substrate.<sup>245</sup> Substrate saturating conditions, and thus, low enzyme-to-substrate ratios were used, as described previously.<sup>243</sup> In line with Kostylev and Wilson, samples from the turnover reactions were taken every hour until 7 hours was reached. The results show that ChiB-WT and R294A have an approximately 10-fold higher activity constant ( $A$ -value) than the other mutants (Table 6.2). The  $b$ -values, which quantify the curvature of the time course profile, are close to 1 for W97A, W220A, and D316A and just below 1 for F190A, E221A, and D316A, while WT and R294A gave a considerably lower value (Table 6.2). This is likely indication that these enzymes are still in the linear phase of degradation due to their low activity.

### 6.3.2 Degradation of chitin

Apparent processivity ( $P^{app}$ ) is mathematically defined as the number of catalytic events an enzyme performs divided by the number of times the enzyme acquires a chain end.<sup>74</sup> However, in practice  $P^{app}$  can be regarded as the actual processive ability of enzyme acting on a particular substrate under a given set of conditions.<sup>23</sup> This processive ability can be measured by a number of different methods, depending on the substrate.<sup>64, 71, 74, 78</sup> For chitinases, one way of measuring  $P^{app}$  is to determine the  $[(\text{GlcNAc})_2] / [\text{GlcNAc}]$  ratio. The  $(\text{GlcNAc})_2$  dimer originates from repeated hydrolysis of crystalline chitin inside a processive cycle or can be obtained outside the processive cycle during the initial binding event, wherein the threaded substrate in the active site occupies even number of product subsites. However, during the initial binding event,  $(\text{GlcNAc})$  is also produced when the substrate occupies an odd number of product sites. Hence, to eliminate the effect of initial binding,  $P^{app}$  is defined as the ratio of  $[(\text{GlcNAc})_2] / [\text{GlcNAc}]$  instead of  $[(\text{GlcNAc})_2]$  only. A high value of  $[(\text{GlcNAc})_2] / [\text{GlcNAc}]$  usually represents a high  $P^{app}$ .

It has been shown that  $P^{app}$  tends to decrease as the substrate degrades.<sup>78</sup> This has been explained by the increase in prevalence of more recalcitrant regions over time, where there are fewer obstacle-free paths for the processive enzymes to move along.<sup>64, 78, 233</sup> Without the addition of accessory enzymes and glucosidase, the enzymes eventually encounter obstacles or fail to release from the substrate causing traffic jams of unproductively bound enzymes. It is thus important to assess processivity during the early stages of the reaction.<sup>78</sup> Here, initial degradation of  $\beta$ -chitin yielded  $[(\text{GlcNAc})_2]/[\text{GlcNAc}]$  ratios of  $23.0 \pm 0.6$ ,  $23.9 \pm 0.6$ ,  $9.8 \pm 0.4$ ,  $11.6 \pm 0.4$ , and  $16.3 \pm$

0.6 for ChiB-D316A, ChiB-R294A, ChiB-W220A, ChiB-F190A, and ChiB-E221A, respectively. The values for ChiB-WT and ChiB-W97A have previously been determined to be  $24.3 \pm 2.0$  and  $11.6 \pm 0.5$ , respectively.<sup>78</sup> The apparent processivity numbers ( $P^{app}$ ) of ChiB-WT, and aromatic and polar variants are summarized in Table 6.2. Finally, differences can also be seen with respect to the degradation efficiency (Table 6.2). The wild-type and R294A degrade 40-45 % of the  $\beta$ -chitin, while W220A, E221A, and D316A degrade between 20 and 25 %. Only 11 and 17 %  $\beta$ -chitin are degraded when F190A and W97A, respectively.

### 6.3.3 MD simulations and free energy calculations

Along with the experimentally determined activity and apparent processivity, we calculated the relative change in ligand binding free energy ( $\Delta\Delta G_{Tl}$ ) for each polar variant (Table 6.2). We reiterate the previously determined relative binding free energies associated with the aromatic to alanine mutations in the ChiB cleft to demonstrate how the overall ChiB cleft chemical composition, including aromatic and polar residues, contribute to ligand stabilization and binding. We observed that the E221A (+3 site), W220A (+2 site), and W97A (+1 site) mutations significantly reduce binding affinity, which corresponds to the reduction in processive ability of ChiB determined experimentally. Neither F190A (+3 site) nor D316A (-1 subsite) affects binding affinity. Overall, the comparison of the free energy values with apparent processivity (Table 6.2), demonstrated that ligand binding free energy is often proportional to the processive ability of GHs, as hypothesized previously.<sup>61</sup> However, the R294A (+1 site) and F190A (+3 site) mutations appear to be outliers, as the arginine mutation, in spite of having unfavorable effect on ligand binding, does not reduce processivity and the phenylalanine

mutation does not affect binding but greatly reduces hydrolyzing activity and processivity. Additionally, our TI calculations reveal the mutation of Tyr-145 to alanine (+1 site) reduces binding free energy by  $3.1 \pm 0.5$  kcal/mol; experimentally this mutation inactivated ChiB and cannot be compared with the experimental processivity in the current scenario.

**Table 6.2** Relative binding free energies determined from thermodynamic integration TI alongside experimentally measured hydrolyzing activity and initial apparent processivity ( $P^{app}$ ) for wild-type (WT) ChiB and selected polar and aromatic variants across the ChiB cleft.

ChiB	Binding site	A <sup>a</sup>	b <sup>a</sup>	[(GlcNAc) <sub>2</sub> ]/[GlcNAc] <sup>b</sup>	Degradation (%) <sup>b</sup>	$\Delta\Delta G_{TI}$ <sup>c</sup>
WT	-----	$0.27 \pm 0.03$	$0.86 \pm 0.06$	$24.3 \pm 2.0^d$	45 <sup>d</sup>	-----
D316A	-1	$0.04 \pm 0.01$	$0.91 \pm 0.16$	$23.0 \pm 0.6$	23	$-0.3 \pm 0.3$
W97A	+1	$0.05 \pm 0.01$	$0.96 \pm 0.10$	$11.6 \pm 0.5^d$	17 <sup>d</sup>	$2.8 \pm 0.4^e$
Y145A	+1	-----	-----	-----	-----	$3.1 \pm 0.5$
R294A	+1	$0.23 \pm 0.09$	$0.74 \pm 0.24$	$23.9 \pm 0.6$	42	$4.6 \pm 0.5$
W220A	+2	$0.04 \pm 0.01$	$1.16 \pm 0.20$	$9.8 \pm 0.4$	23	$3.5 \pm 0.3^e$
F190A	+3	$0.06 \pm 0.01$	$0.84 \pm 0.14$	$11.6 \pm 0.4$	11	$-0.3 \pm 0.2^e$
E221A	+3	$0.07 \pm 0.02$	$0.81 \pm 0.14$	$16.3 \pm 0.6$	24	$1.8 \pm 0.4$

<sup>a</sup> 20 mg/ml chitin, 100 nM enzyme, <sup>b</sup> 2 mg/ml chitin, 2.5  $\mu$ M enzyme, <sup>c</sup> kcal/mol, <sup>d</sup>

Reported previously by Hamre et al.,<sup>78</sup> <sup>e</sup> Reported recently by Jana et al.<sup>112</sup>

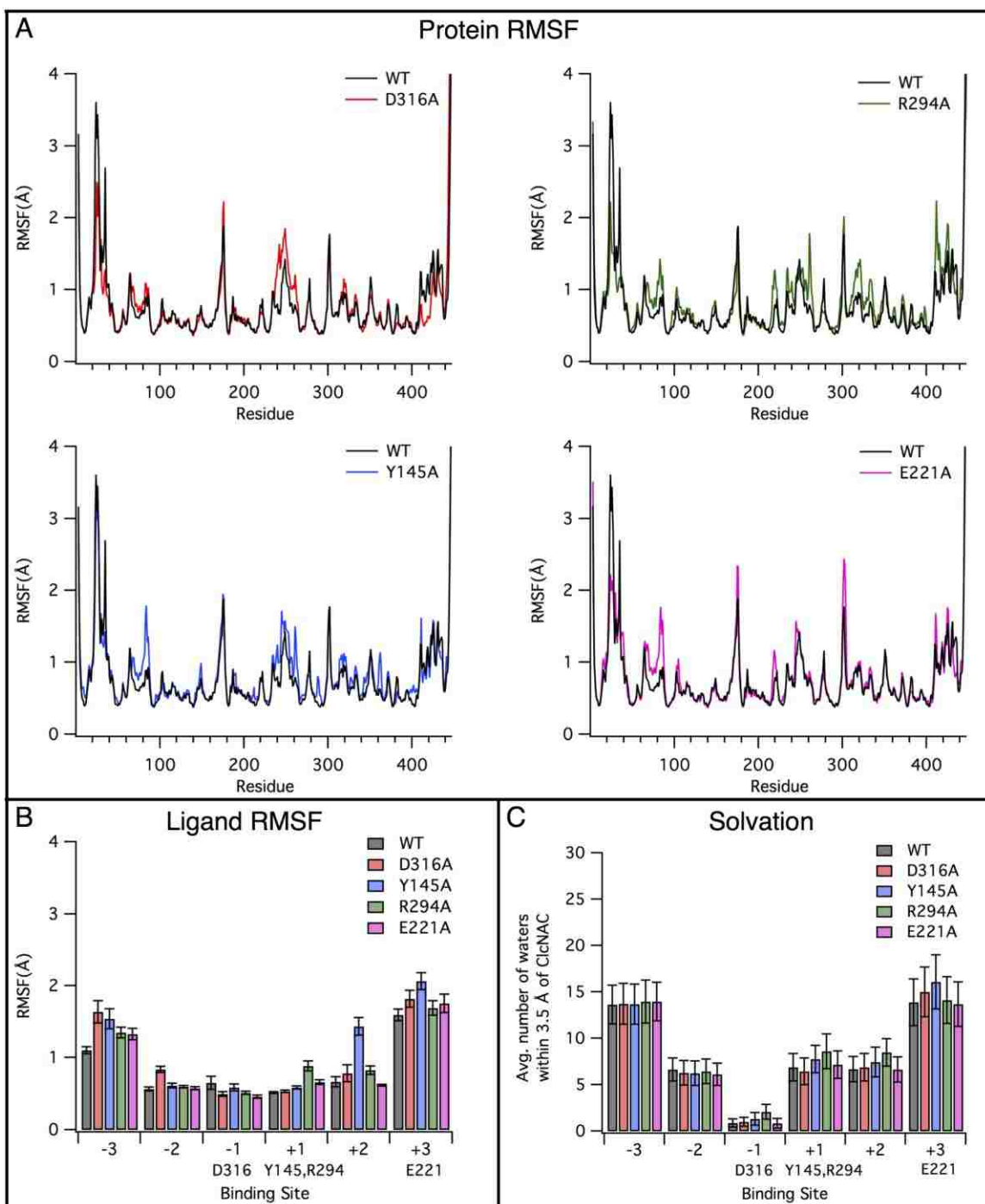
In addition to conducting free energy calculations, we also carried out classical MD simulations to provide further insights into the role of polar residues in the protein ligand dynamics and binding affinity. From 250-ns MD simulation trajectories, we

computed the root mean square fluctuation (RMSF) of the protein backbone and the RMSF of the ligand (Figure 6.2). We also counted the average number of water molecules present in the ChiB cleft per ligand binding site, which can be correlated to the change in solvation entropy (Figure 6.2C).

The RMSF represents the fluctuation of a residue or group of residues from the time averaged position over the simulation. Higher RMSF in the variant with respect to the wild-type is expected to correspond to lower ligand binding affinity due to mutation. The calculated RMSF of the protein backbone from 250-ns MD simulations shows the mutants, in general, fluctuate more than the wild-type (Figure 6.2A), particularly in the area surrounding the mutation.

The RMSF of the ligand in a given binding site can also be related to the binding affinity of the enzyme for the ligand. The RMSF of the ligand shows that the Y145A mutation has the highest impact on ligand stability, destabilizing the substrate binding sites (+2 and +3) in comparison to the wild-type (Figure 6.2B). The R294A mutation also destabilized the ligand, though the effect is local, affecting only the +1 substrate site. In contrast, the D316A mutation destabilizes the two product subsites (-2 and -3) but does not affect the substrate sites. We previously showed that the substrate sites in the ChiB cleft have more dominant contribution in ligand binding than the product subsites,<sup>119</sup> which is in accordance to our  $\Delta\Delta G_{TI}$  values (Table 6.2) obtained from TI calculations. Though the measured RMSF of the ligand in E221A was quite similar to the wild-type, the higher flexibility of the protein backbone at certain locations in the variant might result in the unfavorable relative free energy change (Figure 6.2).

In addition to the RMSF of the protein and the ligand, we also determined active site solvation in the cleft of ChiB wild-type and all polar variants bound to the (GlcNAc)<sub>6</sub> ligand (Figure 6.2C). Higher active site solvation corresponds to an unfavorable contribution to entropy due to solvation. To determine active site solvation, we counted the average number of water molecules within 3.5 Å of each ligand binding site over 250 ns. The active site solvation across all six binding sites in the wild-type ChiB and all of the polar variants are similar (Figure 6.2C), suggesting the entropic contribution due to solvation in ligand binding is very similar across the wild-type and the polar variants.



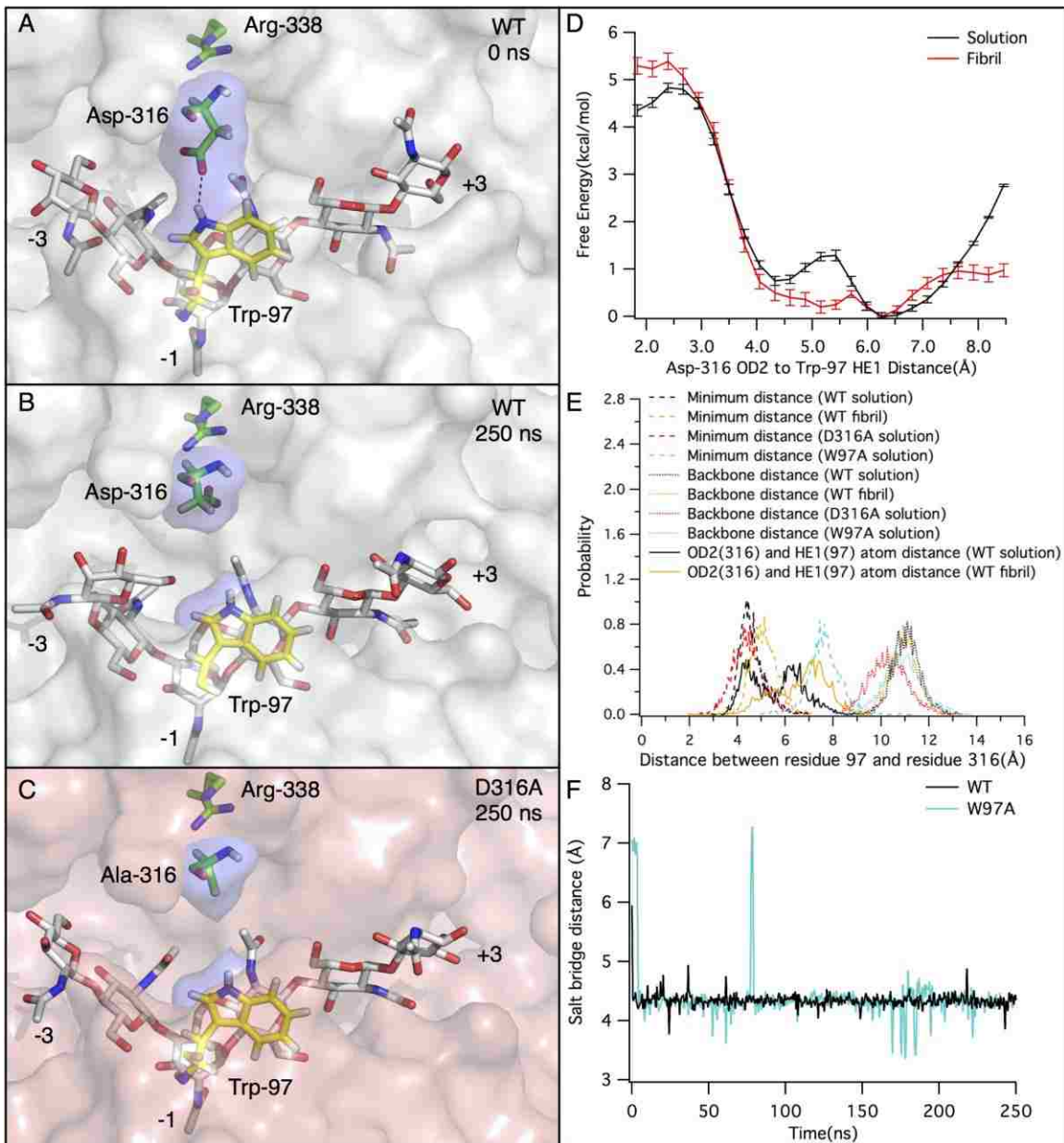
**Figure 6.2** Active site dynamics and solvation of ChiB wild-type (WT) and selected polar variants from 250-ns MD simulations. All simulation sets analyzed here, contained a hexameric chito-oligosaccharide ligand (GlcNAc)<sub>6</sub> in the ChiB cleft. Figure (A) illustrates the root mean square fluctuation (RMSF) of the ChiB protein backbone



determined over the 250-ns simulation trajectories. Figure (B) illustrates the RMSF of the ligand across all six binding sites of ChiB WT and selected polar variants. Figure (C) illustrates the average number of water molecules within 3.5 Å of the ligand on a per-binding-site basis. In all cases, associated errors were obtained from 2.5 ns block averaging.

As described above, hydrogen bonding between Trp-97 and Asp-316 is thought to lead to formation of a roof over the -1 site of the ChiB cleft (Figure 6.3A), which may play a role in the exo-activity of ChiB. However, from the MD simulations, we observed the Trp-97/Asp-316 hydrogen bond breaking during the initial phase of simulation and never reforming throughout the remainder of the simulation (Figure 6.3B). To determine quantitatively whether the formation of the roof over -1 site is thermodynamically favorable, we conducted umbrella sampling in presence and absence of the  $\alpha$ -chitin microfibril (Figure A3.1 of Appendix A3). A harmonic restraint was applied during equilibration MD simulations to maintain the hydrogen-bonding distance between Trp-97 and Asp-316; the harmonic restraint was removed prior to introduction of the umbrella sampling biasing force to increase the distance between Trp-97 and Asp-316 in a stepwise manner. The free energy required to do so is shown in Figure 6.3D, where we see that the free energy reaches zero when the two residues are approximately 6.4 Å apart from each other, both in absence and presence of the microfibril. The free energy profile reveals that roof formation is thermodynamically unfavorable (5.3 kcal/mol in presence of the microfibril and 4.4 kcal/mol with the bound oligomer). Additionally, we observed that the two residues prefer to remain at a distance of approximately 6.4 Å from each other, as measured from the Trp-97 HE1 atom and Asp-316 OD2 atom, which corresponds to the

statistical frequency of this atom pair distance from unrestrained MD simulation (Figure 6.3E). Moreover, the side chain of Asp-316 forms a salt bridge interaction with Arg-338 instead of Trp-97 in the wild-type MD simulation, which is favored over the roof formation (Figure 6.3F).



**Figure 6.3** Snapshots of WT ChiB bound to a (GlcNAc)<sub>6</sub> ligand at (A) 0 ns and (B) 250 ns, illustrating how the ‘roof’ forming hydrogen bond between Trp-97 and Asp-316 is broken over the course of simulation. (C) Similarly, the hydrogen bond is broken in the D316A variant simulation, as illustrated by the snapshot at 250 ns. (D) The potential of mean force (PMF) generated from umbrella sampling simulations using weighted histogram analysis method (WHAM). The PMF shows ‘roof’ formation both in the

absence and presence of the  $\alpha$ -chitin microfibril is energetically unfavorable. The Asp-316 oxygen (OD2) and Trp-97 hydrogen (HE1) prefer to lie approximately 6.4 Å apart. (E) Histograms of the minimum distance between Trp-97 and Asp-316 calculated over 250-ns simulations show the distance between Trp-97 HE1 and Asp-316 OD2 is most frequent near 4.5 Å and 6.5Å for the (GlcNAc)<sub>6</sub> and microfibril bound enzymes, respectively, as also observed in umbrella sampling. The W97A simulation used here for distance comparison was published previously.<sup>112</sup> (F) Comparison of salt bridge distance between residue 316 and Arg-338 in WT, and W97A.

## 6.4 Discussions

### 6.4.1 Ligand binding free energy versus processive ability of ChiB

The relationship of ligand binding free energy to the processive ability of a GH (Equation 6.2) was described previously by Payne et al., where it was suggested that more favorable binding affinity correlates with increased cellulase processivity.<sup>61</sup>

$$-\frac{\Delta G_b^0}{RT} = \ln\left(\frac{P^{Intr} k_{on}}{k_{cat}}\right) \quad (6.2)$$

$\Delta G_b^0$  is the ligand binding free energy, T is the temperature, R is the universal gas constant,  $P^{Intr}$  is the intrinsic processivity,  $k_{on}$  is the enzyme substrate association rate constant, and  $k_{cat}$  is the catalytic rate constant. The relationship has not yet been quantitatively validated, however, as processivity data for only two cellulases was available and experimental determination of the kinetic parameter,  $k_{on}$ , is exceptionally challenging. The formulation of Equation 6.2 is thought to be true of all families of GHs acting processively. Accordingly, we have investigated this relationship with respect to

Family 18 chitinase processive ChiB function, as chitinases processivity data is well documented and kinetic parameters are easier to measure due to the substrate-assisted catalytic mechanism.<sup>74</sup>

We recently demonstrated that a positive correlation exists between processivity and binding free energy in Family 18 chitinase ChiA.<sup>190</sup> In that study, we found that the mutation of a polar residue Thr-276 to alanine in processive ChiA active site cleft reduces binding affinity as well as processivity, and mutation of another polar residue Arg-172 to alanine had little effect on processivity and ligand binding free energy.<sup>190</sup> In another recent study of ours, we observed the mutation of tryptophan residues to alanine in both processive ChiA and ChiB catalytic clefts (Trp-167, Trp-275 for ChiA and Trp-97, Trp-220 for ChiB) significantly reduced binding affinity, and based on available processivity data in literature, we suggested that the mutation of those residues would also affect processivity.<sup>112</sup> We also observed the phenylalanine residues in the ChiA and ChiB active site (Phe-396 in ChiA and Phe-190 in ChiB) did not contribute much to the ligand binding, and we expected that mutation of those residues would not affect processivity either.

To further corroborate the relationship in Equation 6.2, we measured processivity for selected ChiB aromatic and polar variants and compared with both existing relative change in binding free energies data in literature and those determined here (Table 6.2). Additionally, we explore the role of polar residues in addition to the canonical aromatic residues in the ChiB cleft and describe how the polar residues contribute to binding, activity, and processivity. In general, all mutations resulting in reduced favorability in binding, including both aromatic and polar mutations, decreased processive substrate

turnover; mutations that did not contribute to ligand binding did not affect processivity. ChiB-Y145A, ChiB-R294A, and ChiB-F190A variants do not strictly align with the trend described in Equation 6.2, the reasons for which we will describe in detail below. Sequence alignment across homologous chitinases shows that the residues having a greater role in ligand binding are more conserved in the ChiB cleft than the others (Figure A3.2 of Appendix A3).

## **6.4.2 Polar mutations**

### **6.4.2.1 Asp-316 (-1 site)**

As described earlier, the hydrogen bond between residues Trp-97 and Asp-316 appears to form a roof over -1 site of the ChiB cleft while the enzyme is complexed with a (GlcNAc)<sub>5</sub> ligand, as observed in the 1E6N crystal structure (Figure 6.3A).<sup>79</sup> However, the hydrogen bond interaction is partially or completely absent in other crystal structures of *S. marcescens* ChiB, where the enzyme is either bound to an inhibitor or is in the apo state,<sup>134, 144</sup> implying hydrogen bonding between Asp-316 and Trp-97 arises due to conformational change in the protein backbone upon binding to a chito-oligosaccharide ligand. The conformation change in the protein upon ligand binding is a common phenomenon and has been previously reported for several GH families, where it was observed that protein loops come closer to the -1 site of the substrate to facilitate further interactions between the protein and the ligand.<sup>255</sup> However in the ChiB 1E6N crystal structure, no direct interaction is observed between Asp-316 and the (GlcNAc)<sub>5</sub> ligand; though, Trp-97 is found to stack with the +1 pyranose ring (Figure 6.1). From the structural information, it was primarily believed that the formation of roof over -1 site would narrow the active site cleft of ChiB facilitating interaction between protein and the

ligand and would be beneficial for processive movement of ChiB by lowering the probability of detachment of the substrate from the enzyme.

Mutation of either Asp-316 or Trp-97 to alanine results in breaking of the hydrogen bond and disruption of the roof; Trp-97 affects binding severely, but Asp-316 has no role in binding. As described above, classical MD simulations reveal the hydrogen bond between Trp-97 and Asp-316 is not stable and is broken quite early in the simulations and never reforms (Figure 6.3B & E). Moreover, we found that the side chain of Asp-316 interacts with the adjacent Arg-338 via a salt bridge instead of interacting with Trp-97 through hydrogen bonding (Figure 6.3F). As a result, Asp-316 moves far away from the -1 site, and thus, the D316A mutation has no contribution in ligand binding ( $\Delta\Delta G_{TT}$ :  $-0.3 \pm 0.3$  kcal/mol), though slightly increased RMSF of -2 and -3 product sites were observed in the D316A variant. Trp-97 continued to stack with the +1 site of the substrate. Previously, we showed the W97A mutation reduces ligand binding affinity by  $2.8 \pm 0.4$  kcal/mol, suggesting Trp-97 has certainly a significant role in substrate binding and processive ability.<sup>112</sup> Biochemical studies revealed that D316A affects neither activity nor processivity on crystalline chitin, and enzyme efficiency was not increased on soluble chitosan substrate, which corresponds to our determination that D316A has no effect on ligand binding.<sup>69</sup> Combined biochemical and computational data strongly suggest the ‘roof’ over the -1 site may be a crystallographic artifact and has no role in the processive mechanism of ChiB. It was also evident from prior experimental studies that ChiB can accommodate substrate in endo mode, which is not possible if the roof is strongly favorable.

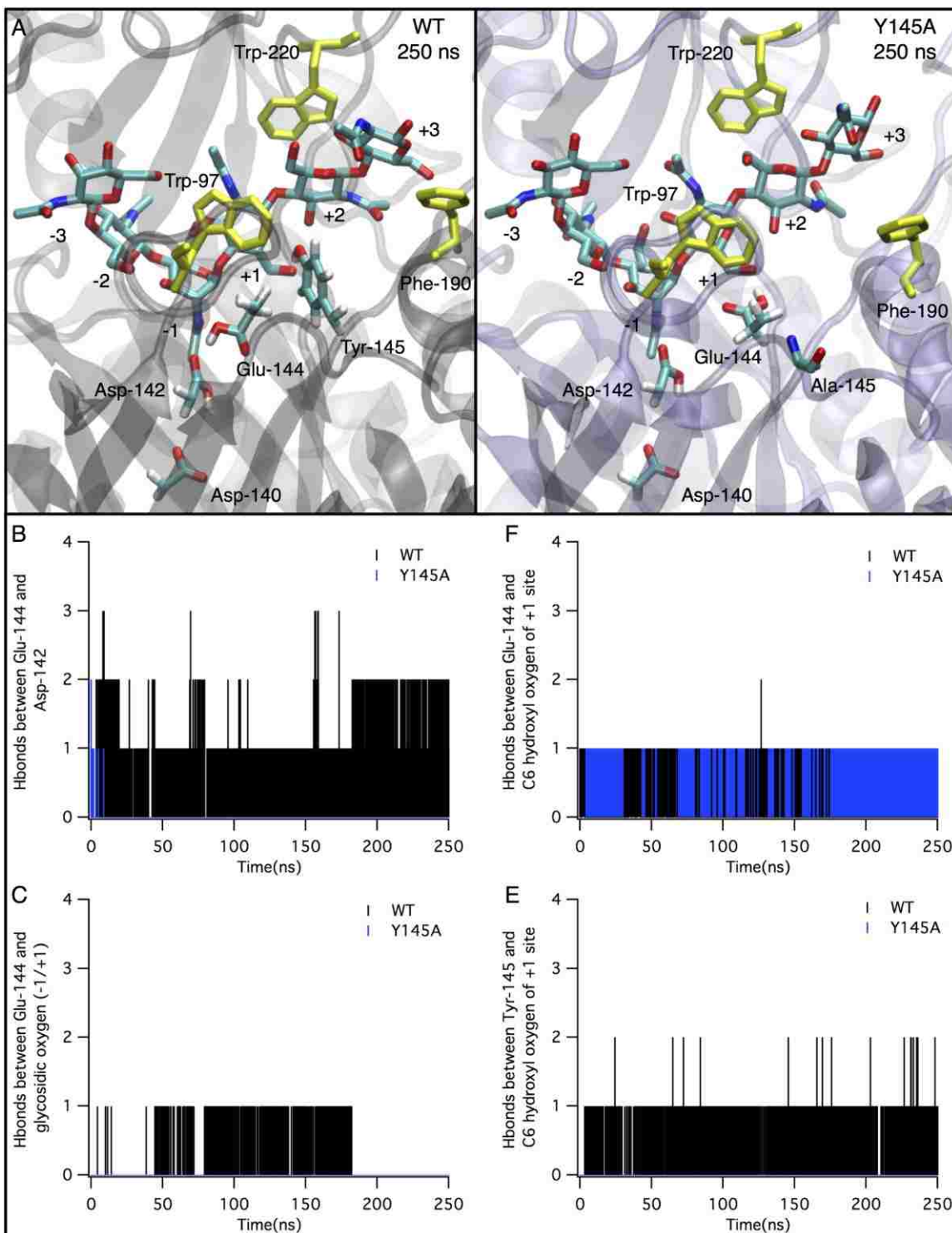
To examine the thermodynamic feasibility of forming a roof over -1 site due to hydrogen bonding, we conducted umbrella sampling free energy calculations in which we observed that the formation of the roof is not thermodynamically favorable, requiring a free energy of approximately 4.5 to 5.5 kcal/mol, based on the nature of the substrate, to close the loops (Figure 6.3D). In addition, irrespective of the nature of the substrate, the most thermodynamically favorable distance between Asp-316 OD2 atom and Trp-97 HE1 is approximately 6.5 Å, significantly longer than the distance between two hydrogen bonding atoms. From sequence alignment of homologous species of ChiB, we further found that Asp-316 is not a conserved residue though Trp-97 is highly conserved, suggesting residues in these positions circumstantially hydrogen bond (Figure A3.2 of Appendix A3). It is the aromatic carbohydrate stacking interaction between Trp-97 and the +1 GlcNAc, rather, that appears to be important in maintaining enzyme substrate association and the processive movement of ChiB.

#### **6.4.2.2 Tyr-145 (+1 site)**

In the crystal structure of ChiB (PDB 1E6N), the hydroxyl tip of Tyr-145 (-OH) is found to interact with the +1 site pyranose C6 hydroxyl group via hydrogen bonding (Figure 6.1), where Tyr-145 acts as a hydrogen bond donor and the +1 site acts as an acceptor. The position of this residue next to the catalytic Glu-144 is highly conserved (Figure A3.2 of Appendix A3). However, no study has explored the function of this residue either in catalysis or binding. Deletion of the Tyr-145 side chain completely inactivates the enzyme, implying this residue is critical for hydrolysis (Table 6.2). MD simulations and TI calculations further revealed that the mutation of Tyr-145 destabilizes the ligand at +2 and +3 sites (Figure 6.2B), resulting in a reduction of favorable binding



free energy by  $3.1 \pm 0.5$  kcal/mol in comparison to the wild-type (Table 6.2). However, due to inactivity of the Y145A variant towards chitin,  $P^{app}$  was not possible to measure; hence, the comparison of  $P^{app}$  with  $\Delta\Delta G_{TI}$  was also not possible. Nevertheless, to explore why the mutation of Tyr-145 to alanine would make the enzyme inactive, the MD simulation trajectories of wild-type and the Y145A variant were further analyzed. From the simulation trajectories, we observed the catalytic Glu-144 side chain in wild-type is very flexible, interacting primarily with catalytic residue Asp-142 and the glycosidic oxygen between the -1/+1 pyranose and seldom with +1 site C6 hydroxyl oxygen (Figure 6.4). Eventually, Glu-144 interacts only with Asp-142. In the ChiB-Y145A variant, the -OH functional group of Glu-144 interacts only with the +1 site C6 -OH via hydrogen bonding instead of interacting with Asp-142 or the glycosidic oxygen between the -1 and +1 site (Figure 6.4). Based on these observations, we suggest the Glu-144 hydroxyl interacts with +1 site C6 -OH group to account for the lost interactions resulting from the Tyr-145 mutation. As a result, Glu-144 is never available for proton transfer to the -1/+1 glycosidic oxygen. In addition, hydrogen bonding between catalytic residues Glu-144 and Asp-142 was low in the Y145A variant with respect to the wild-type (Figure 6.4B). Detailed quantum mechanical studies are required to further speculate a mechanism, however. Being next to the catalytic residue Glu-144, Tyr-145 is crucial for stability of the Glu-144 side chain, forcing Glu-144 to interact with glycosidic oxygen (-1/+1) and Asp-142 to facilitate hydrolysis. Along with conserved Tyr-145, a proline residue (Pro-146) adjacent to Tyr-145 is also highly conserved and may also be important for stabilizing the loop containing the DXDXE catalytic motif.



**Figure 6.4** (A) Simulation snapshots of ChiB wild-type (WT) and the Y145A variant at 250 ns illustrating how catalytic residue Glu-144 compensates for the loss of interaction between residue 145 and the +1 site C6 hydroxyl, following the Tyr-145 to alanine

mutation. Figures (B) through (E) compare hydrogen bond (Hbonds) formation in WT ChiB and the Y145A variant over 250 ns MD simulation between: (B) catalytic residues Glu-144 and Asp-142; (C) Glu-144 and the glycosidic oxygen between -1 and +1 sites; (D) Glu-144 and the +1 site C6 hydroxyl oxygen; and (E) residue 145 and the +1 site C6 hydroxyl oxygen. To determine the number of hydrogen bonds, a distance cutoff of 3.4 Å and an angle cut of 60° from linear were used.

### 6.4.2.3 Arg-294 (+1 site)

In the crystal structure of ChiB, Arg-294 appears to interact with the +1 site substrate via hydrogen bonding between the arginine side chain nitrogen and the +1 *N*-acetyl side chain carbonyl oxygen (Figure 6.1). Additionally, Arg-294 was found to form salt bridge with Asp-215 (Asp-215 atom OD1 and Arg-294 Atom N).<sup>79</sup> The formation of a salt bridge lowers the  $pK_a$  of Asp-215, which plays a major role in catalysis by stabilizing the -1 site via hydrogen bonding and increasing the  $pK_a$  of catalytic residues Asp-142 and Glu-144.<sup>188</sup> A previous density functional theory (DFT) study of the ChiB mechanism ascribed no role to Arg-294 in catalysis, though experimental evidence was unavailable.<sup>198</sup> Our experimental activity data validated the DFT study findings, showing that the deletion of Arg-294 at +1 site does not affect the hydrolyzing ability of ChiB toward crystalline  $\beta$ -chitin. From MD simulations, we observe the R294A mutation indeed disrupts the salt bridge interaction with Asp-215, leading to higher flexibility of the Asp-215 side chain (Figure 6.5A & C). In the wild-type simulation, the Asp-215 OD1 maintains a salt bridge interaction with Arg-294, and the Asp-215 OD2 atom maintains interactions with the -1 site C6 hydroxyl through hydrogen bonding, as seen in the crystal structure. In the R294A variant simulation, Asp-215 OD1 and OD2 atoms rotate back and

forth frequently, altering their positions over the simulation, though both OD1 and OD2 atoms alternatively maintained interaction with the -1 site (Figure 6.5A, D, and E). This reveals stability of -1 site was not affected by deletion of the Arg-294 side chain, as is evident from the RMSF of the ligand (Figure 6.2B). However, we are unable to determine from MD simulation if the frequent exchange of position between OD1 and OD2 atom of Asp-215 side chain in R294A variant affects the  $pK_a$  of Asp-142 and Glu-144, though the activity data clearly showed that the variant behaved same as the wild-type. Surprisingly, sequence alignment among homologous species (Figure A3.2 of Appendix A3) shows Arg-294 is highly conserved, like other catalytic residues (Asp-140, Asp-142, Glu-144, Tyr-214, Asp-215), indicating a positively charged residue is essential at this position to stabilize the Asp-215.

In addition to exploring the role of Arg-294 in catalysis, we also investigated how Arg-294 contributes to ligand binding. The deletion of the arginine residue destabilizes the +1 site, resulting in an unfavorable change in free energy of  $4.6 \pm 0.5$  kcal/mol. From simulation trajectories, we also observed that the adjacent Trp-220 gains flexibility (Figure 6.2A) and attempts to stack with the +2 and +3 sites instead of stacking only with the +2 site (Figure 6.5A). Accordingly, it is conceivable that the destabilization of Trp-220 contributes to unfavorable change in  $\Delta\Delta G_{TI}$  following the R294A mutation. In general, Arg-294 appears to play several roles; though it is not a direct participant in catalysis, Arg-294 stabilizes Asp-215 and the +1 pyranose prior to catalysis and acts as a space filler forcing Trp-220 to stack with +2 site.

Finally, we have considered only the catalytically competent Michaelis state of ChiB for MD simulation and free energy calculations and did not consider other substrate

orientations along the steps of the processive cycle prior to hydrolysis (i.e., processive sliding of the substrate in the ChiB active site) and polar residues may interact differently with the substrate at different points along the cycle. For example, in the crystal structure of the reducing-end specific processive chitinase A from *Vibrio harveyi*, a (GlcNAc)<sub>5</sub> ligand in the intermediate sliding state was captured in a completely opposite orientation at the -2 subsite (analogous to non-reducing ChiB +1 site) along with the other substrate sites of the cleft, demonstrating a different pattern of polar interaction at the substrate side in comparison to the Michaelis complex.<sup>256</sup> If we assume the (GlcNAc)<sub>6</sub> ligand during the processive sliding forms similar intermediate state in ChiB, the strong specificity of Arg-294 towards the *N*-acetyl side chain (+1 site) may be crucial in correctly positioning the substrate in the cleft to ultimately form the Michaelis complex state.

#### 6.4.2.4 Glu-221 (+3 site)

As shown in Figure 6.1, the -OH side chain of Glu-221 forms a hydrogen bond with the +3 site C6 hydroxyl oxygen and is believed to be critical for substrate recognition and binding. Deletion of the side chain reduced ligand binding free energy by nearly 2 kcal/mol and reduced  $P^{app}$  (Table 6.2). Though the RMSF of the E221A ligand was similar to the wild-type, the RMSF of the protein backbone was increased in various loop regions: residues 174-176, 300-305, and adjacent to the mutation (Figure 6.2), suggesting fluctuation in the protein backbone due to mutation is responsible for the loss in favorability in  $\Delta\Delta G_{TI}$ .

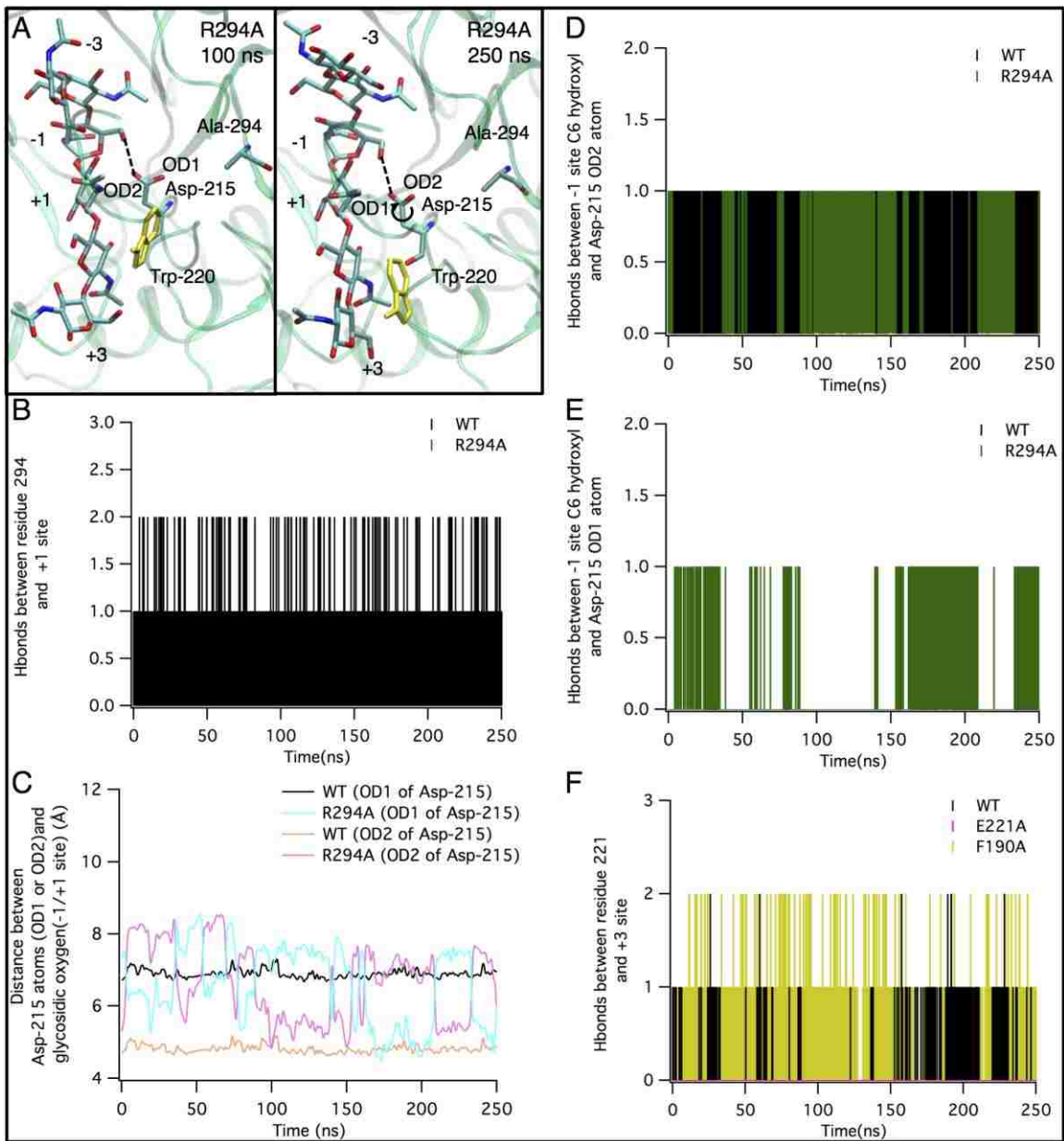
A hydrophobic phenylalanine (Phe-190) opposite Glu-221 in the ChiB cleft stacks with the +3 site (Figure 6.1). Previously, we showed the contribution of Phe-190 in chito-oligosaccharide binding was nearly zero.<sup>112</sup> Here, we find the ChiB-F190A has reduced

$P^{app}$  and activity in comparison to the wild-type (Table 6.2); though, this is in stark contrast to activity measurements previously published by Katouno, *et al.* on the same enzyme.<sup>201</sup> Processivity measurement obtained from the dimer-to-monomer ratio format are known to be inconsistent across laboratories by virtue of the difficulties associated with measuring GH processive ability.<sup>74</sup> Accordingly, the conflicting F190A processivity and activity measurements leave many remaining questions.

Phe-190 is not a conserved residue and polar residues frequently appear in that location (Figure A3.2 of Appendix A3), suggesting the residue should not significantly impact activity or processivity. In contrast, Glu-221 is a relatively conserved residue interacting with the +3 site. The comparison of hydrogen bonding between Glu-221 and the +3 site in ChiB wild-type, the ChiB-E221A variant, and the ChiB-F190A further shows the hydrogen bond between Glu-221 and the +3 site is quite frequent in the F190A variant, indicating the stacking interaction lost due to the Phe-190 mutation is compensated by Glu-221 (Figure 6.5F). Similar observations were made with respect to the *S. marcescens* ChiA-F396A variant, where a nearby charged polar residue, Lys-369, was found to interact with the +2 product site, compensating for the stacking interaction loss due to a Phe-396 to alanine mutation.<sup>112</sup> Taken together, it appears that to stabilize the carbohydrate-phenylalanine stacking in chitinases, a charged polar amino acid is essential in an adjacent position.

Despite of having no effect on oligomer binding,<sup>112</sup> our study also shows that the F190A variant drastically reduces activity and processivity in comparison to the wild-type ChiB (Table 6.2). This observation completely contradicts a previous finding where Phe-190 mutation was reported to have no effect on binding or hydrolyzing ability.<sup>201</sup>

Nevertheless, our MD simulations showed that in solution, the Phe-190 side chain did not continue to stack with +3 pyranose, unlike what has been observed in the crystal structure, and rotates its hydrophobic face toward the solution (Figure A3.3 of Appendix A3). Additionally, in presence of the  $\alpha$ -chitin microfibril, Phe-190 stacks with the +4 subsite instead of stacking with the +3 pyranose (Figure A3.3). These observations agree well with our previous findings, where an extra pyranose modeled at the tunnel entrance of ChiB at +4 site significantly increased ligand binding free energy.<sup>119</sup> This suggests our computational modeling of ChiB oligomer binding in solution underestimates the binding contribution of the tunnel entrance residues in presence of the crystalline substrate, which may account for the discrepancy between experiments conducted as part of this study and our calculations. Free energy calculations exploring the role of these residues in the presence of the crystalline substrate would be beneficial, though computationally intractable. Based on the different conformations of Phe-190 observed in both the crystal structure and over MD simulation, it is conceivable that Phe-190 facilitates guiding of the threaded polymer chain to the catalytic center for hydrolysis, as activity and processivity were severely affected due to mutation of Phe-190.



**Figure 6.5** (A) Snapshot of R294A variant MD simulation at 100 ns and 250 ns showing the frequent exchange of positions between OD1 and OD2 atom of Asp-215 while maintaining the hydrogen bond interaction with the -1 site. The snapshot also illustrates the flexibility of Trp-220 in stacking with the +2 and +3 site instead of only with +2 site. Panel (B) shows hydrogen bonding over the 250-ns simulation between Arg-294 and the +1 site in WT and the R294A variant. Panel (C) shows the distance between the



glycosidic oxygen (-1/+1 site) and atoms OD1 and OD2 of Asp-215, illustrating how OD1 and OD2 alternate their positions in the R294A variant, while in WT, the positions of the atoms are nearly fixed due to salt bridge formation between Arg-294 (N) and Asp-215 (OD1). In this analysis, the glycosidic oxygen between -1 and +1 site has been used as the reference to show the flexibility of Asp-215 atoms. Panels (D) and (E) show hydrogen bonds over 250 ns between the -1 site and atoms OD1 and OD2 of Asp-215. Panel (F) shows hydrogen bonds over 250 ns between residue 221 and the +3 site in WT, and variants E221A and F190A. The F190A simulation used here for hydrogen bond analysis was published previously.<sup>112</sup>

## 6.5 Conclusions

From combined MD simulations, free energy calculations, and experimentally assessed activity and processivity data we have characterized the function of polar residues in the ChiB active site. In general, we observed that the amino acids in the Family 18 *S. marcescens* ChiB cleft, including aromatic and polar residues, that significantly contribute to substrate binding are critical for processivity, though outliers exist as a result of compensating interactions that occur as a result of mutation. In addition, we observed that Asp-316, which was believed to be critical in forming a narrow active site with a roof over -1 site, has no role in binding and processivity. MD simulations and umbrella sampling calculations further revealed that the formation of roof over -1 site is thermodynamically unfavorable, opposing the hypothesis of a conformational change in the active site loops of ChiB upon substrate binding to provide additional interactions. From MD simulations and free energy calculations, we also observed that Arg-294 and Glu-221, at different locations of ChiB cleft, are important in

stabilizing the aromatic stacking interaction of Trp-220 and Phe-190 in the cleft. Previous studies showed that the combination of hydrogen bonding and hydrophobic interactions facilitates processivity, which is in line with the observations made here.<sup>215</sup> In addition, we also suggest Arg-294 is crucial for substrate positioning at +1 site.

## **Chapter 7 - The Thermodynamic Relationship of Chitinase Binding, Processive Threading, and Work Required to Decrystallize $\beta$ -Chitin**

Copyright © Suvamay Jana 2017

### **7.1 Introduction**

Crystalline polysaccharides, such as cellulose and chitin, play many functions in cell biology including energy storage, metabolism, cellular interactions and structures of cell walls.<sup>168, 257-259</sup> To utilize these polysaccharides for food and energy, nature evolves cocktail of enzymes including GHs to depolymerize these polysaccharides to soluble sugars.<sup>16</sup> The GHs responsible for cellulose and chitin degradation usually attack the substrate either from a free chain end (exo-mechanism) or from a random point along a polymer chain (endo-mechanism) in the crystal.<sup>42, 84</sup> Each of the two mechanisms is further accompanied by a processive mode of action, wherein after the acquisition of the substrate in the active site, GHs hydrolyze the substrate repetitively to produce many disaccharide units before dissociation from the substrate. GH processivity is the workhorse of enzymatic polysaccharide deconstruction, yet the details of processive GH action are not understood completely in molecular level.

The community's working hypothesis is that GH processivity occurs over several steps.<sup>22-23</sup> The enzyme must first find the substrate in solution, bind to the crystalline surface, and complex with a single chain prior to the initial hydrolytic event. Then, the GH embarks on a processive cycle of hydrolysis, product expulsion, and rethreading the chain, which is repeated until the enzyme reaches the end of a chain or a barrier and dissociates. Each step likely involves overcoming a free energy barrier, but the entire process must be energetically downhill, being a spontaneous process. It has been

previously demonstrated that the cellulose-degrading processive GH *Trichoderma reesei* Cel7A conforms to the proposed downhill mechanism, and the hydrolytic step is rate-limiting inside the processive cycle;<sup>44</sup> though, dissociation is the overall rate-limiting step outside the processive cycle.<sup>61, 72</sup> Similarly, it has been shown that another cellulose degrading processive enzyme, *T. reesei* Cel6A, also follows the proposed downhill free energy profile with the single-step inverting reaction being the rate-limiting step inside the processive cycle.<sup>57</sup> Nevertheless, these are the only two such studies of processive mechanism free energy profiles, and it is unclear if the hypothesized mechanism holds for processive GHs across different families working on various crystalline polysaccharides. Accordingly, we will test the hypothesis that the free energy profile of processive GHs is downhill for multiple GH families by elucidating the free energy barriers associated with processive chitin degradation by *S. marcescens* chitinase A (ChiA). ChiA has been selected for comparison based on the availability of experimental kinetic rate data,<sup>190, 221</sup> it is also widely considered an advantageous model enzyme system for studying GH mechanisms by virtue of its ease of experimental characterization.<sup>74</sup>

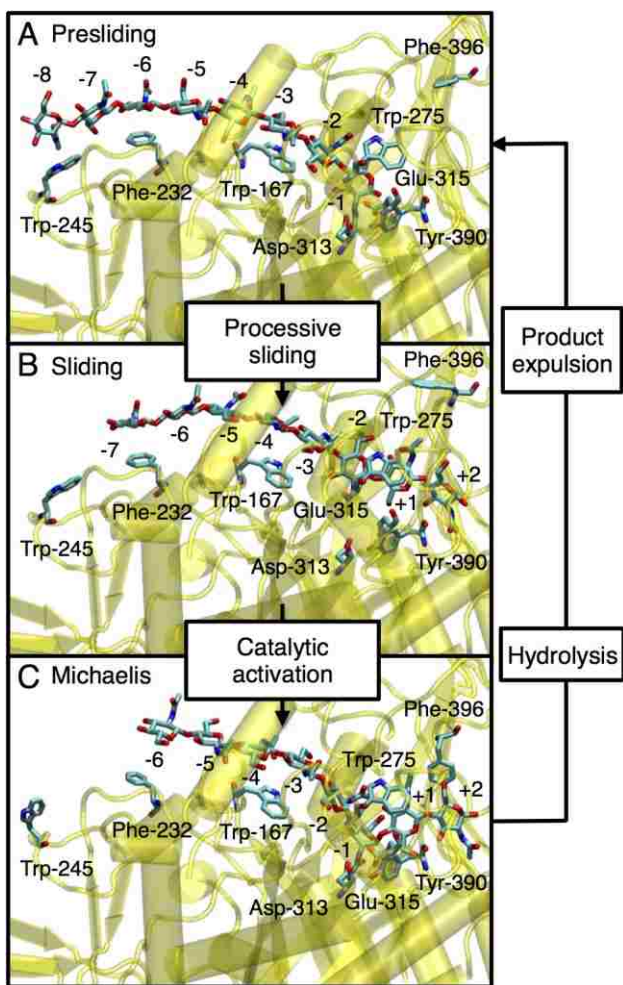
*S. marcescens* ChiA is a reducing-end specific processive chitinase from GH family 18.<sup>81-82</sup> The deep catalytic cleft of ChiA contains nine binding sites in total, six at the substrate site and the three at the product site.<sup>45, 119</sup> Additionally, the carbohydrate-binding module (CBM) fused with the catalytic domain (CD), provides additional binding sites outside the tunnel entrance.<sup>181</sup> From crystal structure, it has been observed that the active site of ChiA, including the CD and the CBM, is lined with many aromatic and polar residues, providing strong hydrophobic stacking and electrostatic interaction to

the substrate and facilitating substrate binding, hydrolysis, and processivity.<sup>45, 70, 112, 181, 190</sup>

Here, we compute free energy barriers for each of the steps in ChiA processive cycle allowing us to identify the rate-limiting step in the processive cycle of ChiA, which is not necessarily the same as for Cel7A and Cel6A. The intermediate steps along the processive ChiA mechanism have been captured in crystal structures and through classical MD simulations conducted previously and as part of this study.<sup>190, 256</sup> The “presliding mode” is the initial state of the GH enzyme, following initial chain acquisition (Figure 7.1A). In presliding mode, the +1/+2 binding sites of ChiA are vacant. The chitin chain then slides through the active site by one chitobiose unit filling the vacant sites, though remaining in a relaxed, non-catalytic conformation (i.e., “sliding mode”, A to B in Figure 7.1B). Following sliding, the activated “Michaelis complex” is formed, wherein the -1 site pyranose ring is distorted into a skew-boat conformation positioning the glycosidic bond for attack (i.e., catalytic activation, B to C in Figure 7.1C). Substrate assisted hydrolysis, via a two-step glycosylation/deglycosylation mechanism, follows (Figure 7.1). The dimeric product is expelled (Figure 7.1), and the processive cycle begins anew. Experimentally it was found that this cycle in ChiA occurs at a velocity of  $70.5 \pm 25.2$  nm/s with a half-life of processivity of 21 reactions on crystalline  $\beta$ -chitin.<sup>83</sup>

In the current study, we will focus on steps A to C of the ChiA processive cycle to determine whether ChiA follows a downhill mechanism prior to hydrolysis and product expulsion. Since understanding each of the processive steps experimentally is intractable, we use MD simulations and computational free energy calculations to estimate the free

energy barriers for the sliding and the catalytic activation step and compare this with the existing kinetic rate data to identify the rate-limiting step inside the processive cycle. Finally, identifying the slowest step inside the processive cycle of *S. marcescens* ChiA and making comparisons to *T. reesei* Cel7A and Cel6A enables us to understand and generalize the bottlenecks of GH processivity, which is essential for engineering proteins for efficient biomass conversion.



**Figure 7.1** Substrate modes/states in the ChiA processive cycle including: (A) presliding, (B) sliding, and (C) Michaelis complex. The figures were made from snapshots obtained from 150-ns classical MD simulations. In all cases, the protein is shown in yellow

cartoon. The octamer ligand ((GlcNAc)<sub>8</sub>) and critical aromatic and polar residues along the active site cleft are shown in cyan stick. In the presliding state, the ligand occupied the -8 to -1 sites with vacant product sites (+1/+2). In the sliding state, the ligand occupies the -7 to +2 sites. In the Michaelis complex, the ligand occupies the -6 to +2 sites. Free energies associated with the three changes of state in A through C will be determined here. Hydrolysis and product expulsion will be examined in future studies.

## **7.2 Computational methodology**

### **7.2.1 Processive sliding and catalytic activation**

Structures of the chitinases processive cycle “transition states” are not readily available for all steps, as is the case for cellulases. For example, *T. reesei* cellulase Cel7A has been captured with a bound cello-oligomer substrate in both the intermediate sliding state and in the catalytically active Michaelis complex;<sup>58, 62</sup> whereas, *S. marcescens* ChiA has only been captured with a bound octamer substrate in the Michaelis complex.<sup>45</sup> Hence, to understand the thermodynamics and the free energy barriers of discrete processive steps along ChiA’s processive cycle, we modeled the presliding mode and the sliding mode structures in addition to the existing Michaelis complex structure. Here, we provide detailed descriptions of the computational methodology to model each of the transition states, including presliding mode, sliding mode, and the Michaelis complex to evaluate the free energy barriers associated with the processive sliding and the catalytic activation steps in ChiA.

### 7.2.1.1 Construction of the Michaelis complex

To model the Michaelis complex, the coordinates of the protein and the octamer ligand were obtained from the crystal structure of ChiA with PDB entry 1EHN.<sup>45</sup> The catalytically inactivating mutation, Gln-315, was reversed to the wild-type Glu-315. Key catalytic residues, Asp-313 and Glu-315, and the *N*-acetyl side chain of the -1 pyranose of the ligand were rotated manually in PyMOL to mimic the catalytically competent Michaelis complex.<sup>85</sup> H<sup>++</sup> was used to determine the protonation state of the protein at a pH of 6 and at inner and outer dielectric of 10 and 80, respectively.<sup>148-150</sup> Two disulfide bonds were specified between Cys-115: Cys-120 and Cys-195: Cys-218 based on the structural studies. Crystal waters were retained during the modeling of the Michaelis complex structure. In the final structure, the ligand contained 8 pyranose rings connected via glycosidic bonds occupying the ChiA active site from -6 to +2 binding sites with the -1 pyranose in the boat conformation (Figure 7.1C). The CHARMM36 force field with CMAP correction was used to model the protein and the ligand,<sup>151-152</sup> and modified TIP3P was used to model the water molecules.<sup>153-154</sup>

After construction of the initial model of the Michaelis complex, a two-step vacuum minimization was carried out in CHARMM to remove the initial bad contacts between the overlapping atoms.<sup>95</sup> In the first minimization step, the four catalytic residues, Asp-313, Glu-315, Tyr-390, and Asp-391, and the -1 pyranose of the ligand were minimized for 100 steps of steepest descent (SD) followed by 100 steps of adopted basis Newton-Raphson (ABNR), keeping the rest of the system fixed. At the second step, the harmonic restraints were removed, and the entire protein-ligand complex and the crystal waters were minimized for 1000 steps of SD, followed by 1000 ABNR steps. The



entire protein-ligand complex was then solvated with water, and sodium ions were added to ensure the charge neutrality of the system. The final solvated protein-ligand complex contained nearly 220,000 atoms with approximate periodic cell dimensions of 130 Å x 130 Å x 130 Å. After solvation, three-step minimization was performed in CHARMM, wherein the first step included minimization of the water molecules keeping the rest of the system fixed, followed by water and ligand minimization keeping the protein fixed, and finally the minimization of the overall system where everything is allowed to move freely; each minimization step included 1000 SD steps, followed by 1000 steps of ABNR.

After minimization, the entire system was heated from 100 K to 300 K in CHARMM in the *NVE* ensemble for 20 ps with a temperature increment of 50 K after every 4 ps. The system was then density equilibrated in CHARMM for 100 ps in the *NPT* ensemble, followed by 150 ns of classical, unrestrained MD simulation at 300 K in the *NVT* ensemble with a time step of 2 fs using the NAMD software package.<sup>96</sup> The simulation parameters used for density equilibration and MD simulation included a non-bonded cutoff distance of 10 Å, a switching distance of 9 Å, and a non-bonded pair list distance of 12 Å. In addition, the SHAKE algorithm was used to fix all hydrogen distances for computational efficiency.<sup>227</sup> Additional details of the protonation states and the simulation methodology can be found in our previous publications.<sup>112, 119</sup>

#### **7.2.1.2 Modeling of the presliding state**

The coordinates of the presliding mode structure came from the modeled structure of the wild type Michaelis complex of ChiA prior to vacuum minimization. To construct the presliding state structure, the glycosidic bond between -1 and +1 site of the ligand in

the Michaelis complex was hydrolyzed manually, followed by vacuum minimization, solvation, three-step minimization, heating, and *NPT* equilibration in CHARMM. This was further followed by 20 ns unrestrained MD simulation in NAMD in the *NVT* ensemble at 300 K. During the course of the MD simulation, the -1 pyranose conformation relaxed to chair from boat, mimicking the state after hydrolytic reaction, as observed in analogous chitinase ChiB from quantum mechanics molecular mechanics (QM/MM) simulation (Figure 1.10 of Chapter 1).<sup>89</sup> Finally, a 20-ns snapshot from the unrestrained MD simulation was used to model the presliding state structure. The +1 and +2 pyranose were manually removed from the product site, and the two additional pyranose were modeled at the entrance of the ChiA cleft at the -7 and -8 sites in PyMOL. The final modeled structure of the presliding state contained an octamer ligand occupying the -8 to -1 binding sites, with two product subsites vacant (Figure 7.1A). Each of the eight pyranose rings were in the chair conformation connected via glycosidic linkages. The modeled enzyme-substrate complex was then re-solvated, keeping the existing water molecules in the system, followed by minimization, heating at 300 K, and *NPT* equilibration in CHARMM. The equilibrated presliding mode structure was then simulated for 150 ns in the *NVT* ensemble at 300 K using NAMD. The time step and the simulation parameters used for presliding state were identical to the Michaelis complex simulation.

### 7.2.1.3 Modeling of the sliding state

The intermediate sliding mode structure was constructed from two 100-ns MD snapshots of the presliding mode and the Michaelis complex structures via a pulling procedure implemented in AMBER,<sup>97</sup> where a biasing force of 7.5 kcal/mol/Å<sup>2</sup> was

applied on the -1 and -2 pyranose of the presliding mode structure pulling the ligand to the +1 and +2 product sites of the Michaelis complex. This was accomplished using the targeted MD (TMD) utility in the Sander module of AMBER,<sup>260</sup> where the root mean squared deviation (RMSD) between the -1 and -2 site of the presliding mode and the +1 and +2 site of the Michaelis complex was decreased in a stepwise fashion from 9.0 Å to 0 Å in 0.5 Å increments every 10 ps. During the sliding of the substrate from the presliding to the Michaelis complex, a harmonic restraint (5 kcal/mol/Å<sup>2</sup>) was applied on the C $\alpha$  of the protein residues within 8 Å of the ChiA active site cleft to ensure protein did not move along with the ligand. Finally, 150 ns unrestrained MD simulation was performed in the pmemd module of AMBER on the final output of the TMD process (RMSD of 0 Å), where the +1/+2 subsites were already filled with two pyranose rings (Figure 7.1B). The time step and the simulation parameters used here were same as the presliding state and the Michaelis complex simulation. We note that the sliding mode structure was modeled with the CHARMM force field, and the CHAMBER module in AMBER was used to convert the CHARMM structure, coordinate and parameter files to the AMBER compatible topology and coordinate files.<sup>97</sup>

#### **7.2.1.4 Free energy barrier estimation for processive sliding and catalytic activation from umbrella sampling**

The 100-ns unrestrained MD simulation snapshots from the presliding, sliding, and the Michaelis complex states served as input to the free energy calculations. To evaluate the free energy barriers for the threading of the substrate from presliding to sliding mode (state A to B) and catalytic deactivation of the -1 pyranose ring conformation from boat to chair (state C to B), we used TMD in AMBER. As described

previously, TMD is a type of umbrella sampling that imposes harmonic restraints on the root mean squared deviation (RMSD) between endpoints. To evaluate the free energy barriers from the presliding to sliding mode transition, we applied TMD with a force constant of  $7.5 \text{ kcal/mol/\AA}^2$  on the pyranose ring carbons and oxygen as well as the glycosidic oxygen of the -1 and -2 sites of the presliding mode structure to align with the product chitobiose (the +1/+2 sites) of the sliding mode structure. Umbrella sampling windows were collected in  $0.25 \text{ \AA}$  intervals over an RMSD of  $8.25 \text{ \AA}$  to  $0 \text{ \AA}$  (34 windows); each window was run for 15 ns. For transition from the catalytically-active state to sliding mode, TMD ( $7.5 \text{ kcal/mol/\AA}^2$ ) was applied to the -1 pyranose of the Michaelis complex structure to allow overlapping with the -2 site of the intermediate sliding mode structure, encouraging the conformational change of the sugar from boat to chair. In this step, windows were collected in  $0.25 \text{ \AA}$  intervals over a RMSD of  $5.30 \text{ \AA}$  to  $0 \text{ \AA}$  (23 windows), with 15 ns of data collection. The simulation parameters used for TMD umbrella sampling were same as for classical MD simulation described earlier. To obtain the free energy profile in each of the processive steps, the last 9 ns of the 15 ns collected per window was used to construct the potential of mean force (PMF) using the weighted histogram analysis method (WHAM).<sup>116</sup> Errors were estimated from the standard bootstrapping method. Convergence and histogram overlap were assessed.

### **7.2.2 Decrystallization of $\beta$ -chitin**

In conjunction with determining the free energy barriers associated with the processive steps, we also modeled a  $\beta$ -chitin substrate to evaluate the work of decrystallization performed by ChiA prior to processive threading of the substrate in the active site. We modeled a  $\beta$ -chitin microfibril and performed 5 ns classical MD

simulation in CHARMM to obtain an equilibrated structure, which was used to estimate the decrystallization work by peeling a chito-oligosaccharide from the  $\beta$ -chitin substrate. Classical MD simulation and free energy calculations were conducted in CHARMM, as described below. The CHARMM36 force field was used to model the substrate and the TIP3P was used to model the water molecules.<sup>153-154, 194-195</sup>

### 7.2.2.1 Modeling of the $\beta$ -chitin substrate

The coordinate of a di-hydrate  $\beta$ -chitin monomer ((GlcNAc)<sub>2</sub>) was replicated along the x, y, and z direction in Mercury to model a  $\beta$ -chitin crystal.<sup>41, 261</sup> The modeled  $\beta$ -chitin contained 840 GlcNAc units in total over 4 layers; each layer contained 7 chains, and each chain was 30 GlcNAc units long (Figure 7.4A).

After construction of the initial model, the substrate matrix was minimized in vacuum for 100 steps of SD, followed by 100 ABNR steps. The substrate was then solvated in a water box of an approximately 170 Å x 100 Å x 60 Å in dimension. Solvation was followed by a two-step minimization in CHARMM, wherein water was minimized at the first step for 1000 SD and 1000 ABNR steps, keeping the rest of the system fixed. In the second minimization step, the harmonic restraint was removed and the whole system was minimized for 1000 steps of SD, followed by 1000 ABNR steps. The system was then heated from 100 K to 300 K over 20 ps in the *NVE* ensemble, followed by 100 ps density equilibration in the *NPT* ensemble in CHARMM. Finally, an unrestrained MD simulation was conducted in CHARMM in the *NVT* ensemble for 5 ns.

### 7.2.2.2 Umbrella sampling

A 5-ns MD simulation snapshot of the  $\beta$ -chitin substrate was used for umbrella sampling to evaluate the work to decrystallize a decamer (10 pyranose) substrate from the  $\beta$ -chitin matrix. The decrystallization was carried out from the edge chain as well as the middle chain of the surface of the substrate matrix to understand the dependence of the substrate morphology on decrystallization work. The decamer substrate was chosen to match prior computational work evaluating work of decrystallization for  $\alpha$ -chitin, to which we will directly compare our results.<sup>15</sup>

The umbrella sampling was performed in CHARMM using fraction of native contact ( $\zeta$ ) as the reaction coordinate, as defined in protein folding.<sup>15, 262</sup> Normalized  $\zeta$  ranged from 0 to 1, where  $\zeta = 0$  represents the chain of interest fully in contact with the polymer crystal and  $\zeta = 1$  represents the chain of interest completely peeled away from the substrate. The cutoff distance considered for native contacts was 12 Å, corresponding to the non-bonded cutoff distance. In total, 41 umbrella sampling windows were constructed for each morphology (edge and middle chain) with a window interval of 0.025, ranging from 0 to 1. During umbrella sampling, the bottom layer of  $\beta$ -chitin was kept fixed using a harmonic potential of 5 kcal/mol. Finally, each umbrella sampling window was run for 13 ns at 300 K with a biasing force of 7,500 kcal/mol to keep the each window in the specified  $\zeta$  value. The last 8 ns data of the 13 ns simulation was used to construct the PMF using WHAM. Error was determined using the standard bootstrapping method. Overlapping of the histograms of the simulation were used to assess convergence.

## 7.3 Results

### 7.3.1 Free energy barrier estimation for processive sliding and catalytic activation

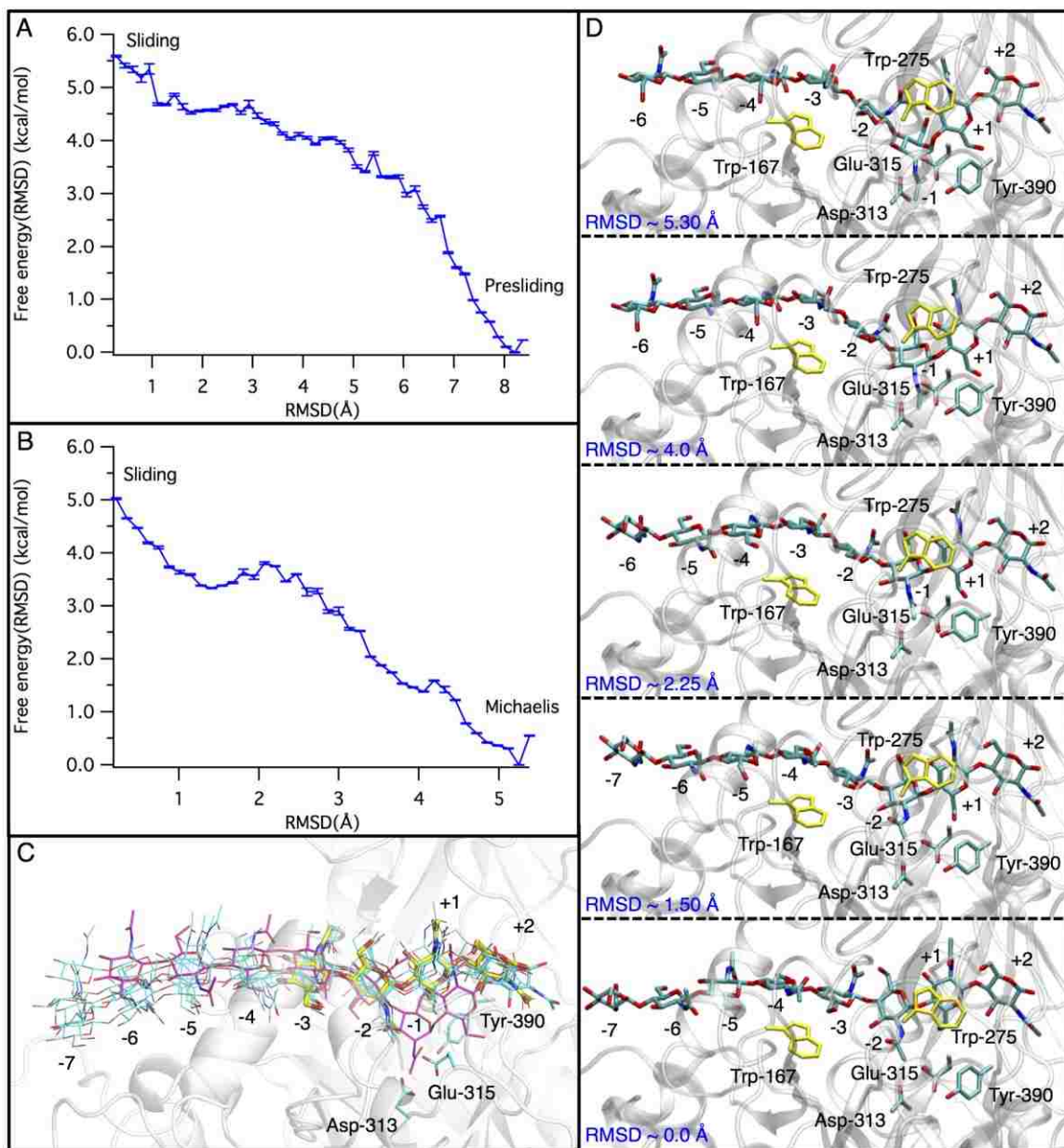
The free energy barrier estimated from TMD umbrella sampling shows that the threading of the substrate from presliding to sliding state is thermodynamically unfavorable with a change in free energy of roughly  $5.5 \pm 0.1$  kcal/mol (Figure 7.2A). Though processive threading filled the vacant product sites (+1/+2) in the ChiA cleft, the slack remaining in the oligomer, not yet in the distorted Michaelis complex in the -1 site, left a pyranose ring outside the ChiA cleft in solution (-7 site) (Figure 7.2C). Moreover, the ligand positioned at the substrate-binding sites (-7 to -2) of the sliding state (RMSD  $\sim 0\text{\AA}$ ) is in a flipped orientation, opposite the presliding state ligand conformation, resulting in a different pattern of interactions with the protein, arising primarily from significant changes in polar interaction (Figure 7.2C). Accordingly, the PMF constructed from the umbrella sampling during processive threading suggests the presliding state is more energetically stable than the sliding state.

The PMF describing the free energy of catalytic activation step shows that the Michaelis complex is more energetically favorable than the sliding mode by approximately  $5.0 \pm 0.1$  kcal/mol (Figure 7.2B). Threading is completed during the catalytic activation step when the -7 and -2 pyranose of the sliding mode structure shift to the -6 and -1 position along with a distortion of the -1 pyranose from chair with 1,4 B boat (Figure 7.2C & D). The approximately 0.5 kcal/mol difference between the two sliding mode states (Figure 7.2C & D) is a result of applying the TMD biasing force to two different structures (the presliding and the Michaelis complex) to encourage transition, and is not significant given the limitations of the computational method.

Finally, the overall free energy profile shows that presliding and the Michaelis complex states are approximately energetically equal; whereas, the sliding state is unfavorable by  $\sim 5$  kcal/mol (Figure 7.2A & B).

Since, the PMFs constructed from umbrella sampling for the processive threading and the catalytic activation steps greatly depend on the target sliding mode structure, it was essential to justify that the intermediate sliding mode structure we modeled from simulation is plausible. Though there are not currently any structures of the intermediate sliding state in the *S. marcescens* chitinases to compare to, the homologous family 18 *Vibrio harveyi* chitinase ChiA (PDB 3B9D) has been captured in an intermediate sliding state.<sup>256</sup> *S. marcescens* ChiA and *V. harveyi* ChiA have a sequence similarity of approximately 53 %. Given the similarity of the two chitinases and the good alignment of the *V. harveyi* ChiA ligand with our model intermediate state ligand obtained from unrestrained MD simulation (Figure 7.2C), we have confidence in the accuracy of the structure and free energy values obtained from umbrella sampling.





**Figure 7.2** The free energy profiles obtained from TMD umbrella sampling for the processive sliding and catalytic activation steps in *S. marcescens* ChiA. Panel (A) shows the potential of mean force (PMF) constructed from the umbrella sampling simulation while the octamer substrate is sliding from the presliding state to the sliding state (state A to B transition, as shown in Figure 7.1). Panel (B) shows the PMF obtained during conformational change of the -1 pyranose from boat to chair conformation while the

substrate transitioned from the Michaelis complex to the sliding state (state C to B transition in Figure 7.1). Panel (C) shows clusters of sliding states obtained from MD simulations compared to the crystal structure sliding state from a homologous chitinase, *Vibrio harveyi* ChiA (PDB 3B9D),<sup>256</sup> and the Michaelis complex of *S. marcescens* ChiA. The pyranose rings in cyan stick are the sliding states obtained from 150-ns MD simulations. The pyranose rings in magenta are the 150-ns snapshot from the Michaelis complex MD simulation. The pyranose rings in yellow stick are the crystal structure of the sliding state from *V. harveyi* ChiA. Panel (D) shows the gradual change of conformation of -1 pyranose from boat to chair at RMSD values along the umbrella sampling pathway, from the Michaelis complex to the sliding state.

### **7.3.2 MD simulations of the presliding, sliding, and the Michaelis complex**

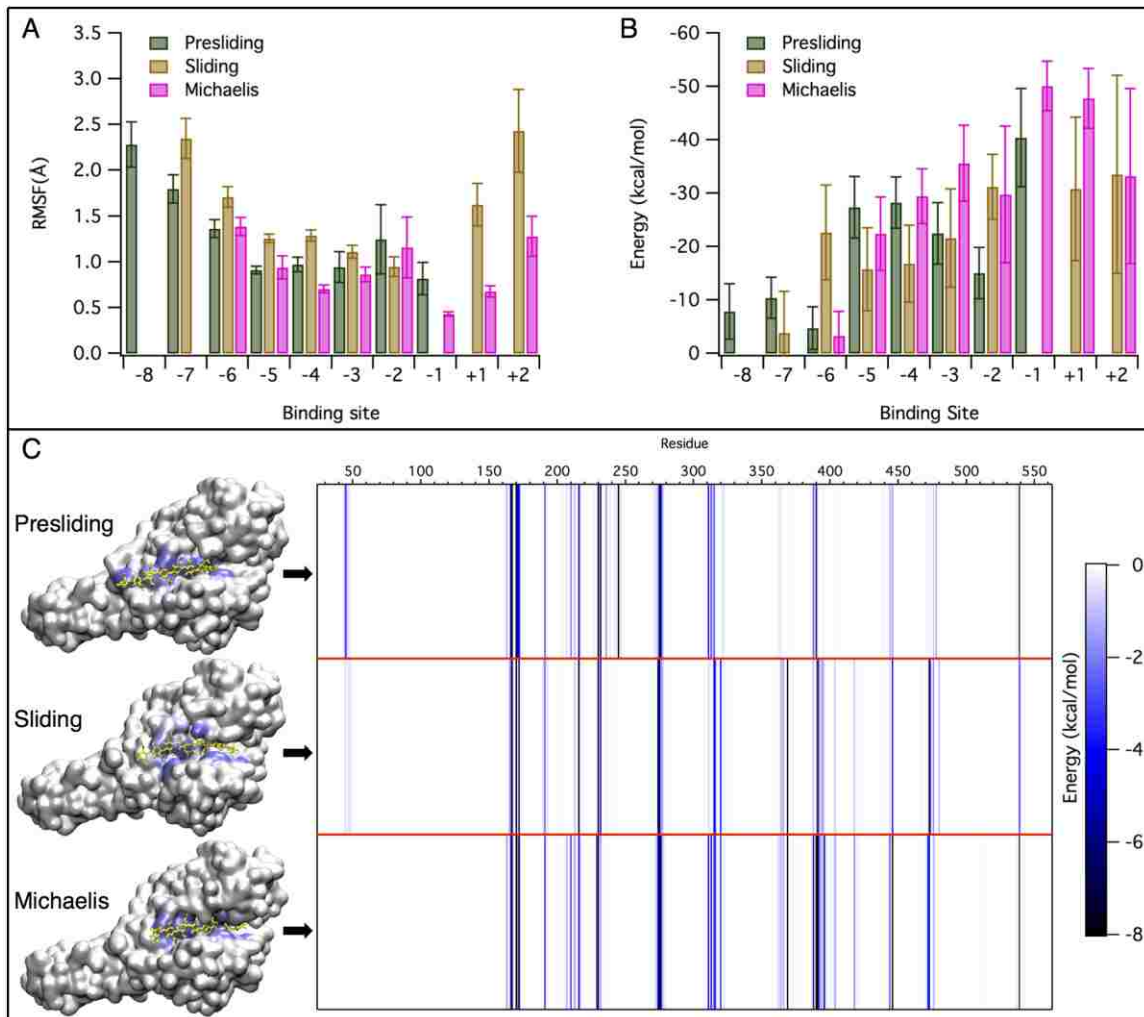
To provide further insights of the dynamic interactions of the protein with the chito-oligosaccharide ligand in the ChiA cleft, 150-ns classical MD simulations were conducted on the presliding, sliding, and the Michaelis complex structures, out of which first 25 ns of simulation data were discarded for equilibration purposes, and the last 125 ns was used to evaluate the dynamic properties of the protein and the ligand. From MD simulation trajectories, the root mean squared fluctuation (RMSF) and the interaction energy of the ligand per binding site were evaluated (Figure 7.3A and 7.3B). In addition, the interaction energy contribution as a function of protein residue was computed for the three different ligand-bound states (Figure 7.3C).

The RMSF of the ligand per binding site represents the average fluctuation of each pyranose from the time-averaged position over the course of the last 125 ns of the MD simulation. Higher RMSF of the ligand is anticipated to reflect lower affinity of the

enzyme toward the substrate in the active site. The calculated RMSF from MD simulation shows the sliding mode ligand in general exhibits slightly higher RMSF than the presliding and the Michaelis complex states, indicating the sliding mode is relatively unstable compared to the other two states (Figure 7.3A). This observation correlates well with the PMFs suggesting the sliding mode is thermodynamically unfavorable.

The interaction energy evaluated per ligand binding site shows the average interaction energies of the ligand in the Michaelis state with the protein is more favorable than that of the presliding and the sliding ligand states, which corresponds with the thermodynamic driving force encouraging the sliding mode state to move toward the distorted catalytically-active state. We also explicitly evaluated interaction energy contributions from each of the amino acid residues with the octamer ligand in the ChiA active site for each of the three modeled ligand states of the processive cycle. We observed a significantly different interaction energy contribution from amino acids in the cleft at the presliding, sliding, and the Michaelis complex states. In the presliding state, though the product sites were vacant, it was observed that the -8 pyranose overlapped with Trp-245 outside the cleft entrance (Figure 7.2). However, in the sliding state, the vacant product sites were filled with a chitobiose with additional interaction between Trp-275 and +1 site pyranose via carbohydrate- $\pi$  stacking and strong electrostatic interaction between Lys-369 and the +2 *N*-acetyl side chain. However, Phe-396 was not found to stack with the +2 pyranose unlike in the crystal structure. During the sliding of the oligomer in the cleft, it was also observed that the stacking interaction with aromatic residue Trp-245 at the entrance of the cleft was lost, and the sliding mode ligand made different polar interactions in the substrate side of the cleft due to its flipped orientation,

not seen in the presliding and the Michaelis complex. In the final Michaelis state, however, the interaction with the entrance aromatic residues were lost, and the polar interactions in the cleft were regained; in addition, the catalytic residues, Asp-313, Glu-315 and Tyr-390, formed strong electrostatic interactions with the -1 pyranose; whereas, Phe-396 was found to stack with the +2 pyranose, providing additional stabilization to the Michaelis complex. We previously showed that Phe-396 did not have significant role in binding affinity,<sup>112</sup> though, the multiple conformations of Phe-396 indicates that Phe-396 is essential in stabilizing the Michaelis complex prior to hydrolysis.



**Figure 7.3** Dynamical fluctuation and interaction energy contributions evaluated from 150-ns classical MD simulations of the three different states along the processive cycle of *S. marcescens* ChiA, including presliding, sliding, and the Michaelis complex. Panels (A) and (B) show the RMSF and the interaction energy of the ligand on a per binding site basis, respectively, for each of the ligand-bound states. Panel (C) shows the interaction energy as a function of protein residue. The ligand occupied the -8 to -1 binding sites in the presliding state, the -7 to +2 binding sites, where the -1 site distortion is not yet formed in the sliding state, and the -6 to +2 binding sites in the Michaelis state. In all cases, last 125 ns out of 150 ns simulation trajectories were used for data analysis. The errors in Panels (A) and (B) were determined from 2.5 ns block averaging.

### 7.3.3 $\beta$ -chitin MD simulation and estimation of work of decrystallization

Prior to evaluating the work of decrystallization, a 5-ns classical MD simulation was conducted in CHARMM at 300 K in the *NVT* ensemble. The crystal lattice parameters of the  $\beta$ -chitin substrate averaged over the simulation were in good agreement with the X-ray diffraction data. The gamma ( $\gamma$ ) angle is slightly higher than experiment but remains within error (Table 7.1).

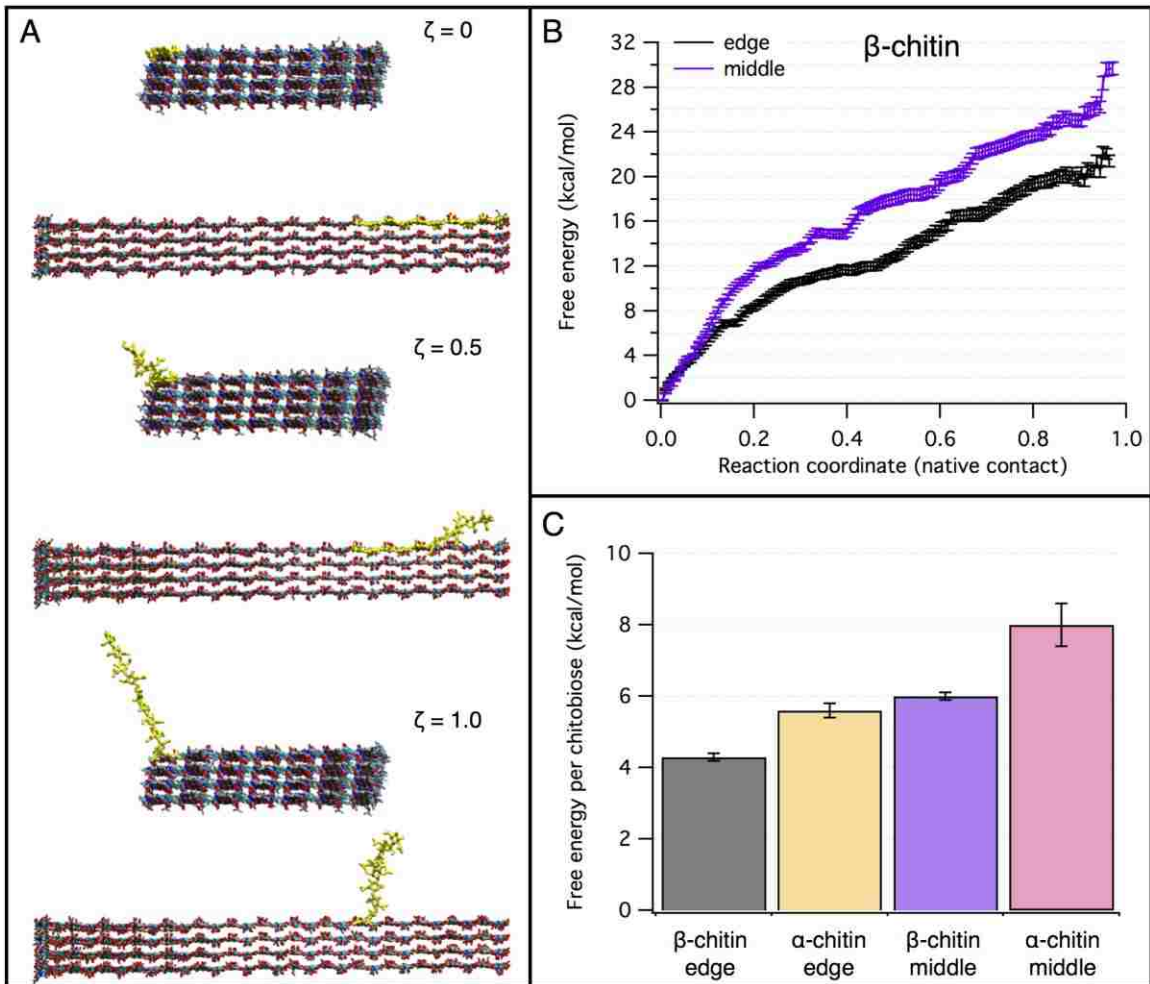
**Table 7.1** Comparison of the crystal lattice parameters of  $\beta$ -chitin between experiment and MD simulation. The errors reported here represent 1 standard deviation from the average lattice parameters evaluated over 5 ns simulation.

Lattice parameters	Experiment	MD simulation
a (Å)	4.814	4.755 $\pm$ 0.226
b (Å)	11.167	11.313 $\pm$ 0.268
c (Å)	10.423	10.401 $\pm$ 0.105
$\alpha$ (°)	90	89.89 $\pm$ 2.70
$\beta$ (°)	90	89.92 $\pm$ 1.83
$\gamma$ (°)	96.45	100.12 $\pm$ 3.63

In addition to obtaining the time averaged lattice parameters, we performed umbrella sampling free energy calculations to evaluate the free energy barriers to be overcome by exo-processive chitinases in threading a polymer chain into the enzyme active site. The PMF constructed from umbrella sampling simulation shows that the middle chain is comparatively more difficult to decrystallize than the edge chain from the surface of the  $\beta$ -chitin crystal (Figure 7.4B). To peel off a decamer (10 pyranose) from the edge of the crystal surface, approximately 21.4  $\pm$  0.5 kcal/mol energy was required, and 29.8  $\pm$  0.5 kcal/mol was needed for the middle chain decrystallization. The free energy values for the edge and the middle chain were also determined on a per chitobiose basis by dividing the final free energy value by five, as two pyranose represent one  $\beta$ -chitin monomer. Thus, the free energy required to decrystallize a chitobiose from the edge and the middle chain of the  $\beta$ -chitin surface was 4.3  $\pm$  0.1 kcal/mol and 6.0  $\pm$  0.1

kcal/mol, respectively. The free energy calculation illustrates how substrate morphology is important to enzymes in degrading crystalline  $\beta$ -chitin, as has been observed previously in work of decrystallizing cellulose polymorphs.<sup>14-15</sup>

To understand the dependence of substrate morphology on enzyme activity, we also examined how substrate polymorph impacts where the enzyme will attack the substrate by quantitatively comparing the decrystallization work of two chitin polymorphs:  $\alpha$ -chitin and  $\beta$ -chitin. Prior calculations showed that the free energy to decrystallize an edge chain and a middle chain from the  $\alpha$ -chitin crystal surface was  $5.6 \pm 0.2$  kcal/mol per chitobiose and  $8.0 \pm 0.6$  kcal/mol per chitobiose, respectively.<sup>15</sup> The quantitative comparison of free energy of decrystallization per chitobiose shows that  $\beta$ -chitin is easier to hydrolyze than the  $\alpha$ -chitin, both from the middle and the edge chain of the substrate (Figure 7.4C). The free energy values are in agreement with experiments showing that  $\alpha$ -chitin is more difficult to hydrolyze than  $\beta$ -chitin.<sup>42</sup> The quantitative comparison also illustrates how crucial a substrate polymorph is on threading of the substrate in the active site.



**Figure 7.4** The decrystallization work ( $W_{\text{decrystallization}}$ ) as a function of native contact ( $\zeta$ ) determined from umbrella sampling simulations of a  $\beta$ -chitin microfibril. Panel (A) shows snapshots of decamer (10 pyranose) decrystallization from the edge of the  $\beta$ -chitin crystal surface at  $\zeta = 0, 0.5,$  and  $1.0$ . For each  $\zeta$ , the front and the side view of the crystal is shown. The chain of interest decrystallized from the crystal surface is shown in yellow. Panel (B) shows the potential of mean force constructed from umbrella sampling simulations by weighted histogram analysis. The PMF shows the edge chain of the crystal surface is easier to decrystallize than the middle chain. Panel (C) compares the  $W_{\text{decrystallization}}$  per chitobiose across  $\alpha$ -chitin and  $\beta$ -chitin polymorphs at the edge and



middle morphologies. The data for  $\alpha$ -chitin decrystallization was taken from Beckham et al.<sup>15</sup>

## 7.4 Discussions

Processivity is a critical characteristic of GHs for efficient biomass deconstruction and comprises of a number of steps, which includes surface binding, recognition, decrystallization, initial threading, and dissociation outside the processive cycle and catalytic activation, hydrolysis, product expulsion, and processive sliding of a dimer inside the processive cycle. Many of these steps, however, are poorly understood due to experimental difficulties in identifying them independently as a result of the heterogeneity of the substrate.<sup>74</sup> In the present work, we used molecular simulation and free energy calculations to examine the processive sliding and the catalytic activation steps of the processive cycle of *S. marcescens* ChiA that occur in between product expulsion and hydrolysis step. We also measured free energy barrier chitinases need to overcome to thread a polymer chain into the active site from the crystalline  $\beta$ -chitin surface, which enables us to understand the relationship of threading/catalytic activation of the chito-oligomer substrate with decrystallization work for a chitobiose inside the processive cycle.

### 7.4.1 Comparison with cellulases

The free energies associated with the processive threading and the catalytic activation steps in ChiA show that the preslide and the Michaelis complex states are approximately energetically equal ( $\Delta G \sim 0$ ) with an energy barrier ( $\Delta G^{\ddagger}_{\text{sliding}}$ ) between the two states of over 5 kcal/mol, suggesting the combined processive threading and the

catalytic activation is a thermodynamically spontaneous process (Figure 7.2 and 7.5). We note that the entire ChiA processive threading/catalytic activation study is carried out considering ChiA hydrolysis of an octamer substrate in solution and does not account the effect of crystalline substrate interactions with ChiA. The case we studied here mimics an ideal situation, where after dissociation of the enzyme from the substrate at the very end of a processive cycle (no glycosidic linkage between the polymer in the active site with the crystalline substrate), an oligomer remains in the active site in the preslide mode, which ChiA can hydrolyze spontaneously to generate dimers via processive mechanism. Prior work in understanding processive threading and catalytic activation of a polymer chain in *T. reesei* Cel7A was also modeled in solution while Cel7A was interacting with a cellononaose ligand.<sup>44</sup> The authors observed a favorable gain in free energy of -8.0 kcal/mol between the presliding and the Michaelis complex states, suggesting processive sliding/catalytic activation a spontaneous process in Cel7A (Figure 7.5A).

The overall free energy profile, however, is significantly different between ChiA and Cel7A; though, both of them are reducing end specific GHs. In ChiA, processive threading is unfavorable by 5.5 kcal/mol whereas the catalytic activation is energetically favorable by 5 kcal/mol (Figure 7.5C). In contrast, the threading of the substrate in the *T. reesei* Cel7A active site was energetically favorable by nearly 8 kcal/mol, and the catalytic activation step had virtually no barrier (Figure 7.4). This comparison reveals a substantial difference in the threading mechanism of the chito/cello-oligomers in the ChiA and Cel7A active site, which is not particularly surprising given the structural differences in mechanism. As described above, during the threading of the substrate in ChiA, though the vacant product sites were filled with chitobiase, one pyranose remained

outside of the ChiA cleft since the chitinase cleft significantly distorts the ligand in the catalytically activated complex, which leaves a great deal of “slack” in the ligand (Figure 7.2). In addition, the sliding state ligand was “flipped” in orientation with respect to the Michaelis complex ligand, leading to unfavorable interactions with the polar residues. During catalytic activation, the ligand is completely threaded with the restoration of the favorable contacts with the protein through -1 pyranose distortion. In contrast, during the threading of cellulose in Cel7A, the sliding state accommodates all nine pyranose rings inside the active site tunnel of Cel7A, with favorable interactions similar to that of the Michaelis complex state.<sup>44</sup>

The combined threading/catalytic activation in *T. reesei* Cel7A is more favorable than that of *S. marcescens* ChiA (by nearly 8 kcal/mol), which we believe arises from the way both of the enzymes were modeled (i.e., full-length vs. truncated). In studying the threading mechanism in Cel7A, all of the states, including the preslide, slide, and the Michaelis complex, considered only the catalytic domain (CD) and did not consider the linker and the carbohydrate binding module (CBM). Here, we considered a full-length ChiA, as deletion of the fused CBM from the CD leads to structural instability of ChiA during simulation. This results in substantially different preslide modes modeled between Cel7A and ChiA. Accordingly, in Cel7A, the two pyranose rings outside the tunnel entrance were in solution in the preslide mode and did not interact with any protein residues. However, for ChiA, the last two pyranose residues at the cleft entrance interacted with the protein residues, mainly with Trp-245 and Phe-232, in the preslide mode. The study of the sliding mechanism in Cel7A showed a significant gain in free energy (favorable) from preslide to slide mode due to strong interaction of the amino acid

residues with the cellobiose at the product site,<sup>44</sup> which may be overly favorable if the oligomer were to interact with the CBM-linker construct. Biochemical studies and simulations have illustrated that the linker and the CBM of Cel7A contribute to substrate binding,<sup>48, 263</sup> hence, we suggest that the preslide mode in the full-length Cel7A would be energetically more stable/favorable than the energy state reported for preslide mode Cel7A CD only. In full-length ChiA, the favorable stacking interaction between the -8 pyranose and Trp-245 needed to be broken to allow sliding of the oligomer from the presliding state. Biochemical studies have previously acknowledged Trp-245, along with other cleft entrance residues, plays a role in binding and hydrolyzing activity, suggesting Trp-245 may substantially hinder the sliding of the ligand in the active site of ChiA.<sup>181</sup> Accordingly, the free energy barrier in the processive sliding/catalytic activation of ChiA is at its maximum between the preslide and Michaelis complex states, with a barrier height of 5.5 kcal/mol; this values would be more favorable ( $\Delta G$  being more negative) if only the CD had been considered. However, we should not forget that both Cel7A and ChiA are different enzymes with different active site topology and chemical composition interacting with different substrates. In addition, the length of the active site tunnel or cleft is also different in both of the enzymes accommodating different number of pyranose rings in the active site (Cel7A: 9 pyranose, and ChiA: 8 pyranose). Finally, despite differences in threading mechanism as well as the modeling methods adopted for ChiA and Cel7A, the combined threading/catalytic activation step is thermodynamically spontaneous, suggesting both of the enzymes are able to hydrolyze the soluble substrate processively.

In addition to measuring the activation energy barrier associated with the processive sliding and the catalytic activation of the oligomer substrate in ChiA, we integrate the sliding/catalytic activation step with the hydrolysis step to identify the rate-limiting step inside the processive cycle prior to product expulsion. Previous biochemical studies revealed that ChiA hydrolyzes a tetramer with a  $k_{\text{cat}}$  of  $33 \text{ s}^{-1}$  at  $37 \text{ }^{\circ}\text{C}$ , corresponding to an energy barrier of  $(\Delta G^{\ddagger}_{\text{Rxn}})$   $15.7 \text{ kcal/mol}$ .<sup>190, 221</sup> The comparison between processive sliding/activation ( $5.5 \text{ kcal/mol}$ ) with the hydrolysis step shows that, inside a processive cycle, reaction is the rate-limiting step in ChiA, similar to that of the *T. reesei* Cel7A and Cel6A though the product expulsion step in ChiA is omitted (Figure 7.5).

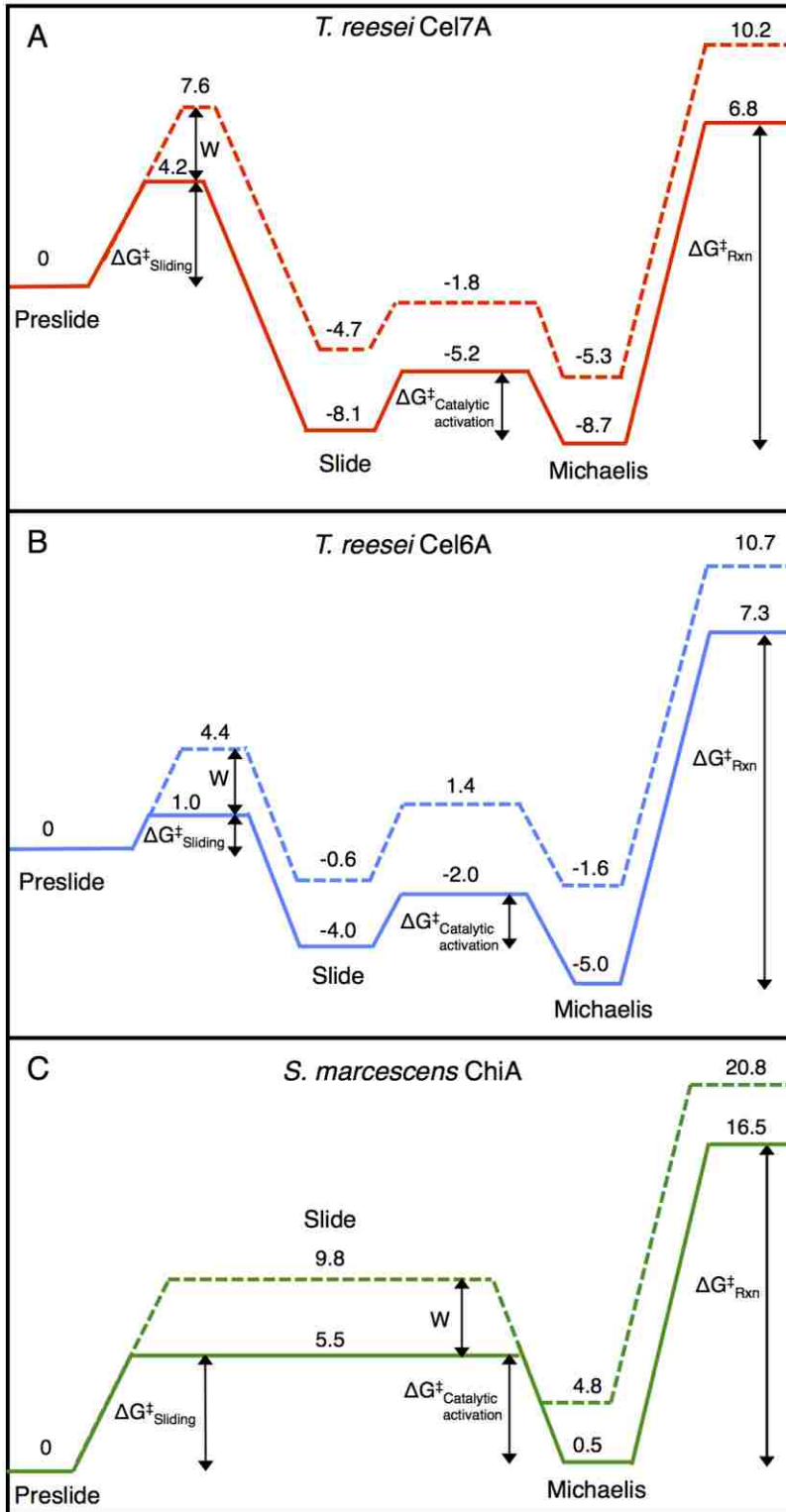
#### 7.4.2 Effect of crystalline substrate

We explicitly calculated the work of decrystallization that chitinases need to perform for each processive sliding step to fill the vacant product sites after hydrolysis and product expulsion. We observed that the work of decrystallization per chitobiose ranges from  $4.3$  to  $8 \text{ kcal/mol}$ , depending on crystalline chitin polymorph and morphology. Adding the work of decrystallization per chitobiose to the free energy barriers of the processive sliding/catalytic activation step, the overall activation energy barrier height reaches about  $9.8$  to  $13.5 \text{ kcal/mol}$  for ChiA, which are still below the activation energy barrier of the reaction ( $15.7 \text{ kcal/mol}$ ), suggesting the reaction would still be the rate-limiting step inside the processive cycle for crystalline chitin substrate (Figure 7.5C).

Interestingly, both in *T. reesei* Cel7A and Cel6A, the Michaelis complex state is more thermodynamically stable than the presliding state by roughly  $-8 \text{ kcal/mol}$  and  $-5$

kcal/mol, respectively.<sup>44, 57</sup> Adding the decrystallization work per cellobiose to that free energy value, the overall energy gain is still favorable by -5.3 kcal/mol for Cel7A and -1.6 kcal/mol for Cel6A, indicating sliding/catalytic activation is a thermodynamically favorable process for the cellulases while simultaneously decrystallizing a cellobiose from the crystal during processive sliding (Figure 7.5A & B). However, in ChiA, the presliding and the Michaelis complex being equal energy states, chitobiose decrystallization would not be favorable thermodynamically (Figure 7.5C). However, all of the previous experimental studies showed that ChiA is processive GH on crystalline chitin with a processive velocity of 70 nm/s with processivity ranging from 21 to 35,<sup>78, 83</sup> indicating combined chitobiose decrystallization and the processive sliding/catalytic activation must be a spontaneous process. We argue that the preslide mode in ChiA should have two pyranose modeled in solution with no interaction with the protein, which certainly did not happen in our current study. The modeling of such a hypothesized preslide mode in ChiA is only possible if the fused CBM is also completely filled with the substrate, allowing two additional pyranose to be modeled in solution. In other words, the free energy profile would have changed its shape to accommodate the chitobiose decrystallization term if the full contribution of the CBM along with the CD was accounted for. Unfortunately, no crystal structure exists to illustrate substrate binding to the ChiA CBM; though, biochemical studies demonstrated that the many aromatic residues in the CBM including, Trp-69 and Trp-33, strongly contribute to substrate binding and the hydrolyzing activity.<sup>181</sup> Moreover, it was also shown that CD mediated binding in ChiA was slow in comparison to Cel7A, and the substrate binding was mostly dominated by the CBM for ChiA.<sup>159</sup> This demonstrates how critical the role of the CBM

is in substrate acquisition and binding in ChiA, suggesting processive sliding of the substrate in ChiA is both mediated by the product sites as well as the fused CBM.

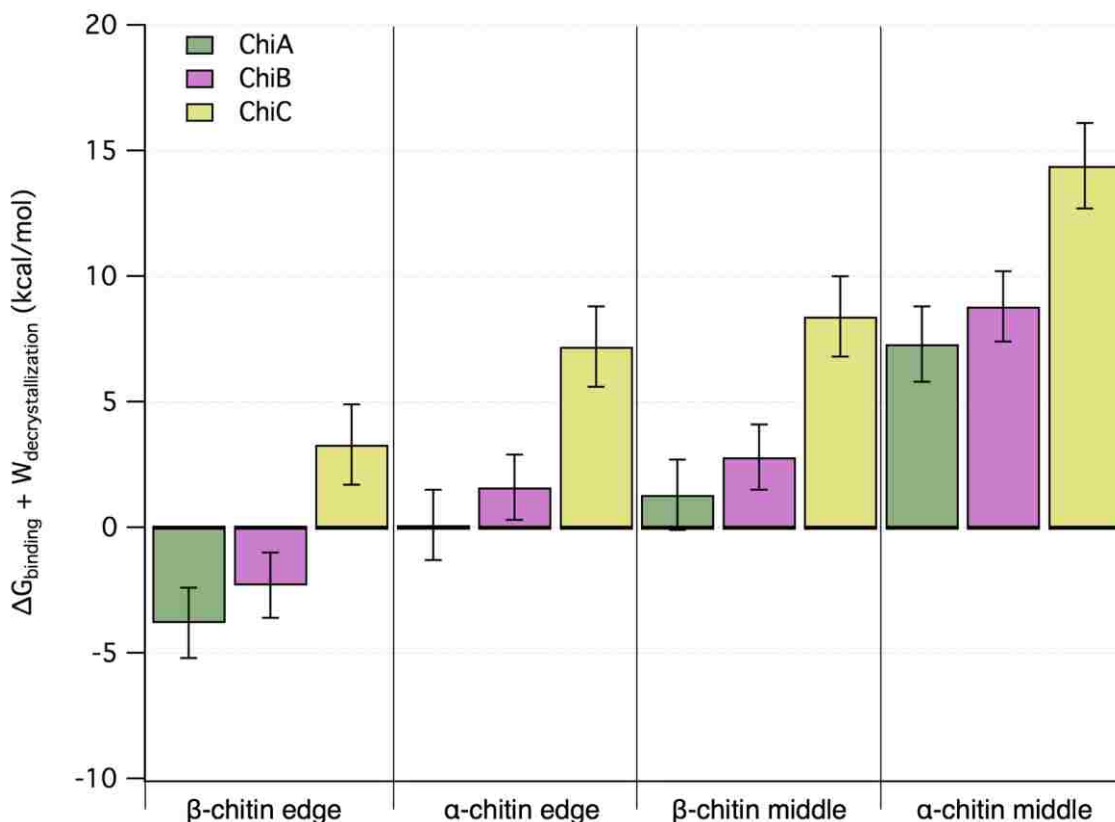


**Figure 7.5** Free energy profiles of the processive GH cycles of (A) *T. reesei* Cel7A, (B) *T. reesei* Cel6A, and (C) *S. marcescens* ChiA. The solid lines represent the sliding, catalytic activation and hydrolysis steps studied in solution. The dashed lines represent the free energy profiles predicted while GHs act on crystalline polysaccharides, which were obtained by adding decrystallization work ( $W_{decrystallization}$ ) per dimer to the solution free energy profiles (3.4 kcal/mol for cellobiose for a corner chain decrystallization from cellulose I $\beta$  surface and 4.3 kcal/mol for chitobiose for an edge chain decrystallization from  $\beta$ -chitin surface).<sup>14-15</sup> The  $\Delta G^{\ddagger}_{\text{Sliding}}$ ,  $\Delta G^{\ddagger}_{\text{Catalytic activation}}$ , and  $\Delta G^{\ddagger}_{\text{Rxn}}$  represent the activation energy barriers for processive sliding, catalytic activation, and hydrolysis steps in solution. The free energy barriers for cellulases were previously determined by Knott et al. and Mayes et al. from computational studies.<sup>44, 57-58</sup> The  $\Delta G^{\ddagger}_{\text{Rxn}}$  for ChiA has been taken from experiments.<sup>190, 221</sup> All of the free energy values in the figure are in kcal/mol. Overall, the free energy profiles show that hydrolysis is the rate-limiting step inside the processive cycle for all of the GHs examined, both in solution and in the presence of crystalline substrates. Here, we note that the product expulsion step has not been considered for comparisons.

Finally, we hypothesize that for forward processive movement GHs should bind the substrate tight enough in the active site to prevent the detached polymer chain from re-associating with the crystalline substrate. In other words, the combined  $\Delta G_{\text{binding}}$  and  $W_{\text{decrystallization}}$  should be either zero or negative for thermodynamic favorability of forward processive movement. To examine the hypothesis, we compare the binding free energies of two processive chitinases ChiA and ChiB and a non-processive chitinase ChiC to an oligomer substrate in the active site with the work of decrystallization. From



our previous calculations, ChiA, ChiB and ChiC bind a hexameric chito-oligosaccharide substrate in the active site with binding free energies of  $-16.7 \pm 1.4$ ,  $-15.2 \pm 1.3$ , and  $-9.6 \pm 1.6$  kcal/mol respectively, with the ligand occupying four substrate sites and two products sites of the cleft and the -1 pyranose in the boat conformation (Michaelis complex).<sup>119</sup> Estimated decrystallization work per chitobiose for  $\alpha$ - and  $\beta$ -chitin from different locations of the crystal surface, from which we calculated work of decrystallization per hexose for different chitin polymorphs and morphologies, have been determined here and in previous studies.<sup>15</sup> Combining the work of decrystallization for a hexameric substrate with the protein-ligand binding free energies, we see that processive ChiA and ChiB exhibit a combined  $\Delta G_{\text{binding}}$  and  $W_{\text{decrystallization}}$  that is either negative or very close to zero, when considering hydrolysis from the edge of the crystal, hence favorable for forward processive movement (Figure 7.6). In contrast, ChiC, which is non-processive, demonstrates unfavorable  $\Delta G_{\text{binding}} + W_{\text{decrystallization}}$ , both from the middle and the edge chain of each chitin polymorph. If we extend this finding in ChiA, where we observed that the Michaelis complex and the presliding state bind to the octamer substrate with almost the same affinity, that means the ligand will be unlikely to reanneal to the microfibril during productive binding, hence will be able to hydrolyze the crystalline chitin processively.



**Figure 7.6** Quantitative comparison of ligand binding free energy ( $\Delta G_{\text{binding}}$ ) and work of decrystallization ( $W_{\text{decrystallization}}$ ) for a hexa-oligosaccharide substrate in processive ChiA, ChiB, and non-processive ChiC. The  $\Delta G_{\text{binding}}$  for all of the three chitinases have been taken from our earlier work.<sup>119</sup>

## 7.5 Conclusions

In the present study we used classical MD simulations and umbrella sampling free energy calculations to determine the free energy barriers inside the processive cycle of *S. marcescens* ChiA. We observed that both in solution and in presence of the crystalline substrate, the hydrolytic step is rate-limiting inside the processive cycle of ChiA. Comparison to *T. reesei* processive Cel7A and Cel6A reveals that hydrolysis being the

slowest step inside the processive cycle is a common phenomenon across various families of processive GHs. Interestingly, we observe that mirroring the reaction mechanism, the processive sliding/catalytic activation mechanism is also very different in cellulase and chitinase. In cellulases during the processive sliding step, the ligand completely enters into the active site filling the vacant product sites. However, in ChiA the sliding of the substrate occurs during both the sliding and the catalytic activation steps. Moreover, the product sites of cellulases provide the majority of driving force for the sliding/catalytic activation of the substrate along the active site. In contrast, in ChiA, both of the product site residues along with the CBM provide the driving force for processive sliding of the substrate. In addition, we find that the comparisons of decrystallization work with the oligomer binding in the active site across processive ChiA, ChiB and non-processive ChiC show tight binding is essential for GHs to remain associated with the substrate for thermodynamic favorability of forward processive movement.

## Chapter 8 – Conclusions and Future Directions

Processive GHs thread individual chains from the polysaccharide into the active site and successively cleave glycosidic linkages producing disaccharides; though this mechanism is the workhorse of enzymatic biomass deconstruction, the molecular-level mechanism responsible for processive hydrolysis remains elusive. Here, I hypothesized that after protein substrate complexation reaches equilibrium, processivity can be positively correlated to ligand binding free energy. I attempted to validate this hypothesis in Family 18 *Serratia marcescens* chitinases, which have been identified as an ideal model system for studying processivity due to its unique substrate assisted catalytic mechanism.

In Chapter 3, I first studied how varying active site topology of processive and non-processive chitinases affected substrate binding. My computational calculations showed that processive ChiA and ChiB bound a hexa-oligosaccharide ligand more tightly than the non-processive ChiC. However, experimentally measured (ITC) and computationally assessed ligand binding free energies showed some significant differences, which we attributed to limitations of the experimental techniques. However, both of the experimental and simulation methodologies showed that entropic contributions from solvation in the ligand binding free energy was less favorable in ChiC than processive ChiA and ChiB. Overall, this work validated the hypothesized relationship of ligand binding free energy and processivity and showed the tunnel-like cleft of processive chitinases favor tight binding of the substrate.

As mentioned above, in Chapter 3, we observed the ligand binding affinity from simulation differed from the experimental ITC measurements for several reasons. In my

concluding remarks, I will expand upon potential reasons for these differences and propose future experiments that may provide additional insight. We anticipate differences primarily arise from the catalytic inactivation of chitinases as part of the ITC protocol. ITC experiments cannot use catalytically-active, wild-type enzymes to measure binding affinity for natural chito-oligosaccharide substrate, as the heat of reaction would also evolve due to hydrolysis; hence, free energy contribution solely from binding cannot be determined. Instead, a catalytically inactive enzyme (point mutation of a key catalytic residue) was introduced to significantly reduce hydrolytic turnover during the ITC experiment, which, unfortunately, limits us to comparing the binding free energies assessed from ITC directly to the enzyme processivity. For the same reason, crystal structures of chitinases complexed with the chito-oligomers are only found in the protein data bank with a point mutation of a key catalytic residue in the enzyme (for example, wild-type glutamic acid to mutant glutamine); otherwise, the natural substrate is unlikely to appear in the bound ligand state across the +1/-1 site (discussed in detail in Chapter 3). In contrast, our MD simulations modeled wild-type enzyme structures bound to the chito-oligomer ligand, reverting the mutated catalytic residue to the wild-type residue, as classical MD simulation is unable to model the catalytic reactions (i.e., not suitable to study the breakage and formation of bonds due to the usage of predefined classical force fields). Thus, the ligand binding free energy for wild-type chitinases could be computed easily from classical MD simulation, with no heat of reaction term, and was compared directly to the experimentally measured enzyme processivity, which is the ultimate objective of my dissertation (i.e., to establish a molecular level relationship of processivity with the ligand binding affinity).

However, the accurate prediction of binding affinity and comparison to ITC values is also essential, not only to convince the broader science community to rely on the simulation data generated by in our study, but also to identify the limitations of computer simulations in accurately predicting free energy values comparable to the experimental ITC measurements. In that line, I propose a few alternative approaches that can be implemented in future studies to generate additional binding free energy data sets that can be further validated by ITC measurements, without losing sight of the overall objective of the dissertation. With that, I first propose calculating ligand binding affinities for catalytically inactive mutant chitinases for direct comparison to ITC binding affinity measurements. The benefits of choosing catalytically inactive mutant to measure binding affinity computationally are multitude. First, calculating ligand binding affinity for an inactive mutant and comparing that with the wild-type chitinases from simulation will allow us to explore whether the catalytic residue that has been mutated for ITC experiments is also responsible for ligand binding or not, and if so, that free energy contribution might be neglected in ITC due to usage of the mutant, potentially underestimating the experimentally measured binding free energies. Second, comparison of the free energy barriers associated with the conformational change of -1 pyranose from chair ( ${}^4C_1$ ) to boat ( ${}^1,4B$ ), essential for hydrolysis, will allow us to understand the kinetics of the substrate activation in the presence and absence of the catalytic residue. Finally, side-by-side comparison of the ligand binding free energy evaluated for the inactive chitinases from simulation and ITC experiments will enable us to pinpoint the additional reasons behind the discrepancies, beyond the different binding site occupancies and the conformational change of the ligand, if arises.

Another alternative approach I propose, in addition to the usage of the catalytically inactive mutants, is implement a chemically modified ligand, an inhibitor very similar to the chito-oligosaccharide substrates (e.g., allosamidin), for both simulation and experimental binding free energy assessments, where the wild-type enzyme can be studied directly without the introduction of point mutations. Allosamidin has a very similar chemical structure to the chito-oligomer substrate occupying from -3 to -1 binding sites in all three chitinases (ChiA, ChiB, and ChiC). The usage of the inhibitor allosamidin for binding affinity measurements also has many advantages, including no heat of reaction evolved due to hydrolysis, occupancy of similar binding sites as the natural chito-oligomer substrate, abundance of crystal structures available in the protein data bank complexed with the allosamidin, and existing literature values for binding free energy data from ITC. We, in fact, used allosamidin (Chapter 5) while exploring the contribution of polar residues Thr-276 and Arg-172 of ChiA in binding and observed consistent results both for natural substrate and the inhibitor. Hence, allosamidin chitinase complexes are likely a good model system to study binding affinity using computer simulation.

In Chapters 4, 5 and 6, I explored the contribution of aromatic and polar residues in ligand binding in processive ChiA and ChiB, as processive chitinase active sites are populated with these types of residues. I chose those aromatic and polar residues in the active site whose activity and processivity were previously determined experimentally or structural studies showed they were interacting with the ligand directly via hydrophobic stacking or electrostatic interaction. I applied molecular simulation and ligand binding free energy calculations to describe the functional roles of these selected aromatic and

polar residues in terms of dynamic measurements and energetic contributions to ligand binding. Most of them showed unfavorable change in free energy due to mutation to alanine, indicating they likely negatively affect processivity, which was also validated through experimental processivity measurements. In particular, I found tryptophan residues at the entrance of the active site or near the catalytic center of processive chitinases (Trp-167 of ChiA and Trp-220 and Trp-97 of ChiB) greatly contributed to ligand binding and processivity. Comparison to processive cellulases further revealed that tryptophan at the entrance of the tunnel is a typical characteristic of processive GHs, providing a tight but fluid-like platform for processive sliding of the substrate. Additionally, from MD simulation and free energy calculations, I observed that polar residues in the active site of chitinases not only contribute to favorable ligand binding by directly interacting with the substrate but also facilitates stabilization of the nearby aromatic carbohydrate stacking interactions. Interestingly, I also observed that some of the aromatic and polar residues in the active site contradicted the hypothesized relationship of binding affinity and processivity (Chapter 6). For example, point mutation of Phe-190 to alanine at the entrance of ChiB cleft though did not affect ligand binding, but greatly reduced activity and processivity toward crystalline  $\beta$ -chitin. This essentially points out that binding free energy alone only partially describes the role of a given residue in processivity. In particular, after noticing multiple conformations of Phe-190 during MD simulation, it is imaginable that Phe-190 guides the substrate chain to the catalytic center and facilitates hydrolysis, which was not reflected in binding free energy value. Overall, the study carried out in Chapters 4, 5 and 6 showed that ligand-binding affinity is positively related to processivity, and nonspecific aromatic stacking and



specific polar hydrogen bonding mostly contribute to ligand binding affinity. However, the additional parameters such as dynamics of the protein and the ligand in solution and in presence of crystalline substrate are also crucial in governing processivity.

Based on the findings in Chapters 4, 5 and 6, I further recommend potential mutations that can be incorporated to benefit GH engineering efforts for efficient degradation of the crystalline polysaccharides. For example, to facilitate more favorable binding and processivity, hydrophobic residues with aliphatic side chains (e.g., alanine, leucine, isoleucine, and valine), particularly at the entrance of the active site, can be mutated to tryptophan. Bulky aliphatic hydrophobic residues, such as isoleucine and valine (have branching at  $\beta$  carbon), should be targeted first for mutation to minimize the steric clashing if sufficient space is not available in the hydrophobic core to accommodate the bulky indole side chain of tryptophan. However, at the entrance of the cleft, tryptophan would have more space to occupy, in comparison to the catalytic center (near the -1 site pyranose); thus, small residues such as alanine can also be considered for mutation. Along that line, I suggest two residues in the processive ChiA cleft suitable for mutation to tryptophan for future studies, Ile-207 (-3 substrate site) and Val-169 (-4/-5 substrate sites), which are located at the entrance of the ChiA cleft and are in suitable proximity to the ligand. Here, we note that Ile-207 and Val-169, both in the active site are conserved and considered to be critical for native folding of the enzyme. Hence, individual role of these residues in activity and processivity must also be determined experimentally. Additionally, crystal structures of the I207W and V169W should be solved in parallel, in both ligand free and ligand bound states, to gain more insight into the enzyme substrate interactions and the protein folding. In the processive ChiB cleft,

we do not see many leucine, isoleucine, or valine residues suitably positioned to interact with the ligand in the active site. However, we observe an alanine (Ala-186) at the +3 substrate site, not a conserved residue, which could also be mutated to tryptophan to increase the favorable binding. In addition to the aliphatic hydrophobic residue mutations, I suggest mutation of the phenylalanine at the substrate sides of chitinases, Phe-232 in ChiA and Phe-190 in ChiB, to tryptophan. As mentioned earlier, Phe-190 in ChiB, though not having a role in binding, can guide the substrate chain toward the catalytic center to facilitate hydrolysis. Accordingly, we are interested in investigating how the phenylalanine to tryptophan mutation at the entrance of the cleft can improve substrate recognition, binding, and processivity. Another important point to which I want to draw attention is that most of the hydrophobic residues (aliphatic) in the processive ChiA and ChiB active sites recommended for tryptophan mutation are on the substrate side and not on the product side, as our results suggest tighter binding at the product site will lead to product inhibition and is not suitable for processive degradation. Moreover, previous studies have pointed out that tryptophans at the product sites enable transglycosylation over processive hydrolysis, and hence, are not recommended for mutation.

The polar residues in the chitinase active sites are also critical for hydrolysis, ligand binding, and processivity. In particular, polar residues close to the catalytic center (near -1 subsite) were found to play significant roles in hydrolysis and are not recommended for alteration. Polar residues also improve aromatic carbohydrate stacking interactions by stabilizing the nearby aromatic residues (Glu-473 and Thr-276 stabilizes Trp-167 in ChiA, Arg-294 and Glu-221 stabilizes Trp-220 in ChiB), and hence, are not

recommended for mutation to any other residue. Polar residues involved in salt bridge interactions are also not suggested for alteration, as the salt bridge interaction is essential for structural stability. Overall, it is highly recommended that polar residues should not be removed from the active site unless to be modified to a different polar residue. Finally, alanine in the active sites (close to the ligand) is recommended for mutation to polar residues (serine, threonine, asparagine) if the more bulky tryptophan or phenylalanine is not suitable at those locations due to steric hindrance.

Finally, I simulated only the Michaelis complex state (state just prior to hydrolysis where the chito-oligosaccharide ligand is in the  ${}^1,4\text{B}$  conformation) so far to explore the role of active site topology and chemical composition in ligand binding (Chapters 3, 4, 5, and 6) and did not account the overall processive cycle, which includes many other steps including hydrolysis, product expulsion, and processive sliding of the substrate into the active site. In other words, binding affinity as well as the dynamical behavior of the protein and the ligand was assessed only for the Michaelis complex state and it remains unclear how enzyme substrate interaction changes during the other processive steps. This sets the future directions of my research, where instead of measuring binding affinity at a single Michaelis complex state, I intended to explore the entire free energy profile of the processive mechanism, which is very difficult to determine experimentally. Accordingly, in Chapter 7, I first explored the free energy profile of the processive sliding step of the substrate in the cleft of *S. marcescens* ChiA and observed that free energy barrier of the sliding of the substrate is significantly lower than the experimentally measured hydrolysis barrier. However, I did not compute the free energy barrier related to the product expulsion step from the ChiA active site, relying

instead on experimentally measured inhibition constants for comparison, which is an anticipated next step in this project. Calculating the free energy barrier corresponding to the product release and further comparison to the barrier heights of the processive sliding and the hydrolytic step will not only help to precisely identify the rate-limiting step of the processive cycle of ChiA but will also provide insights how ChiA is inhibited by chitobiose product.

Another immediate future direction of my research would be to follow the similar computational methodology to explore the entire free energy profile of the processive cycle of ChiB. The free energy profile comparison between two processive chitinases will eventually help us to recognize the driving potential of processivity. Finally, the same methodology can also be implemented to explore the free energy profile of the aromatic and the polar variants of ChiA and ChiB examined here. The comparison of free energy profile between the wild-type and the mutant chitinases will further explore discrepancies in both observed simulation data and contradictory experimental efforts.

Overall, my work demonstrated a straightforward and fundamental relationship of ligand binding free energy and active site characteristics – binding site location, chemical nature of the residues, and steric effects with processivity in GHs. Application of this information will be directly applicable to GH engineering efforts for economical biofuels production.

## Appendix

### **A1 Supporting Information for Thermodynamic Relationships with Processivity in *Serratia marcescens* Family 18 Chitinases**

Appendix section A1 has been adapted with permission from Hamre et al.,<sup>119</sup> Copyright © 2015, American Chemical Society.

#### **A1.1 Computational methods**

##### *System preparation and MD simulations*

After building initial systems from crystal structures as described in the manuscript, vacuum minimization was performed in CHARMM to overcome initial bad contacts between overlapping atoms.<sup>95</sup> In the cases of ChiA, ChiB, and ChiC where the ligand occupies the -4 to +2, -3 to +3, and -4 to +2 binding sites, respectively, 100 steps of steepest descent minimization (SD) were performed on the entire protein ligand complex followed by 100 steps of adopted basis Newton Raphson minimization (ABNR). In case of ChiA and ChiB where the ligand binds from the -3 to +3 site and -2 to +4 site, a two-step vacuum minimization protocol was followed. For ChiA, the entire protein and ligand bound from the -3 to +2 site was fixed, and the +3 site sugar was allowed to move freely for 100 SD steps followed by 100 ABNR steps. The harmonic restraints were then removed from the system, and the entire protein-ligand complex was minimized for 100 SD and 100 ABNR steps. Similarly, for ChiB, the entire protein and the ligand bound from the -2 to +3 site was fixed allowing the +4 site sugar to move freely for 100 SD steps followed by 100 ABNR steps. All restraints were removed and the entire ChiB protein-ligand complex was minimized for 100 SD and 100 ABNR steps. After initial

minimization, all simulation sets were solvated with water molecules, and sodium ions were added for charge neutrality. Complete setup details are provided in Table A1.1.

For all MD simulations following solvation, the water molecules were minimized for 10,000 steps of SD keeping the protein and ligand (if present) rigid. The protein and water were then minimized for 10,000 SD steps, and finally, the entire system was minimized for 10,000 steps using SD. The system was then heated from 100 K to 300 K in steps of 4 ps with 50 K increments, and the density was equilibrated in the NPT ensemble with the Nosè-Hoover thermostat and barostat for 100 ps.<sup>224-225</sup> Following density equilibrations in CHARMM, MD simulations were conducted in NAMD for 100 ns in the NVT ensemble at 300K with a 2 fs time step.<sup>96</sup> Long-range electrostatic interactions were described using the Particle Mesh Ewald method with a sixth order b-spline, a Gaussian distribution width of 0.312 Å, and a 1 Å grid spacing.<sup>226</sup> The non-bonded cutoff distance was 10 Å with a switching distance of 9 Å and a non-bonded pair list distance of 12 Å. The SHAKE algorithm was also used to fix all hydrogen distances for computational efficiency.<sup>227</sup>

### *Free energy simulations*

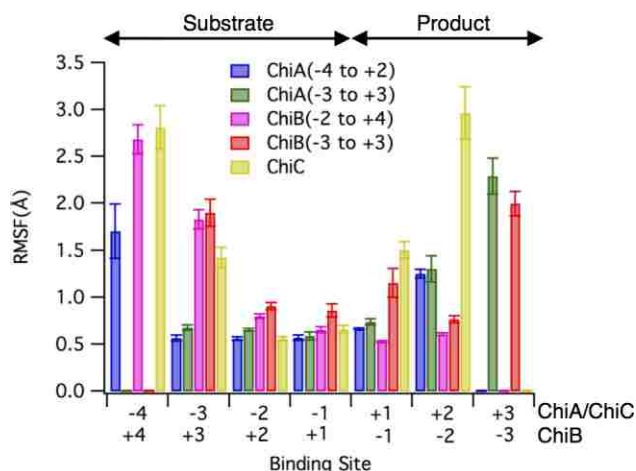
For both solvation and binding site free energy calculations, 128 FEP windows ran concurrently with an acceptance ratio of > 70% along the alchemical path. The simulations were performed in the NVT ensemble at 300 K with a 1 fs time step. The force field parameters for FEP/ $\lambda$ -REMD were, for consistency, as described in the manuscript for the MD simulations. To determine enzyme-ligand free energy ( $\Delta G_1$ ), a positional restraint (spring constant 10 kcal/mol/Å<sup>2</sup>) was used to maintain a fixed distance between the initial center of mass of the ligand and the protein.

**Table A1.1** Simulation details for all eight MD simulations considered in this study.

	ChiA			ChiB			ChiC	
	Apo	Bound	Bound	Apo	Bound	Bound	Apo	Bound
Crystal Structures used to build systems	1EHN <sup>45</sup>	1EHN <sup>45</sup> (-4 to +2)	1EHN <sup>45</sup> & 1E6N <sup>79</sup> (-3 to +3)	1OGG <sup>144</sup>	1OGG <sup>144</sup> & 1E6N <sup>79</sup> (-3 to +3)	1OGG <sup>144</sup> & 1E6N <sup>79</sup> (-2 to +4)	4AXN <sup>73</sup>	4AXN, <sup>73</sup> 1KQY <sup>146</sup> & 3N18 <sup>147</sup> (-4 to +2)
Periodic Boundary Conditions [ $\text{\AA}$ ] <sup>3</sup>	120x120x120			120x120x120			80x80x80	
Atoms	174569	174725	174728	174583	174583	174547	51869	51926
Sodium Ions	2			3			2	
Protonated Residues	Asp313; Glu315; Glu540			Asp142; Glu144			Asp139; Glu141	
Disulfide Bonds	Cys115- Cys120; Cys195- Cys218			Cys328- Cys331			----	

**A1.2 MD simulation results***Root mean square fluctuation (RMSF) of the ligand*

From the 100 ns trajectories, the RMSF of the ligand on a per binding site basis was calculated (Figure A1.1). The pyranose residues in the product side of non-processive ChiC (+1 and +2 ligand binding site) fluctuate significantly more than those of processive chitinases ChiA and ChiB, in line with previous observations illustrating that the product side residues of ChiC have comparatively lower affinity towards substrate than processive chitinases allowing ease of product expulsion.<sup>73</sup>



**Figure A1.1** Root mean square fluctuation (RMSF) of the (GlcNAc)<sub>6</sub> oligomer by binding site. Error bars for the RMSF values were obtained through 2.5 ns block averaging.

### *Hydrogen bond analysis*

Along with the RMSF of the ChiB protein backbone, we also performed hydrogen bond analysis to provide additional insights into the significant change in free energy associated with binding site occupation (i.e., when the ligand occupies either binding sites -3 to +3 or -2 to +4). Here, we anticipate the difference in protein backbone fluctuation as the ligand occupies different binding sites in ChiB (Figure 3.4B of Chapter 3) results from addition of pyranose rings at the +4 and -3 sites, which can be illustrated in terms of hydrogen bond analysis (Figure A1.2).

From the MD simulation trajectories, we determined the number of hydrogen bonds formed between Asp316 to Pro317 of the ChiB product side and the last product side pyranose ring of the two different ligand binding site locations (Figure A1.2A & C). We also examined hydrogen bonds between the tunnel entrance residues of ChiB (Phe239 and Tyr240) and the +3 site pyranose in the -3 to +3 ligand-bound case and the +4 site pyranose in the -2 site to +4 ligand-bound case (Figure A1.2B & D). The protein

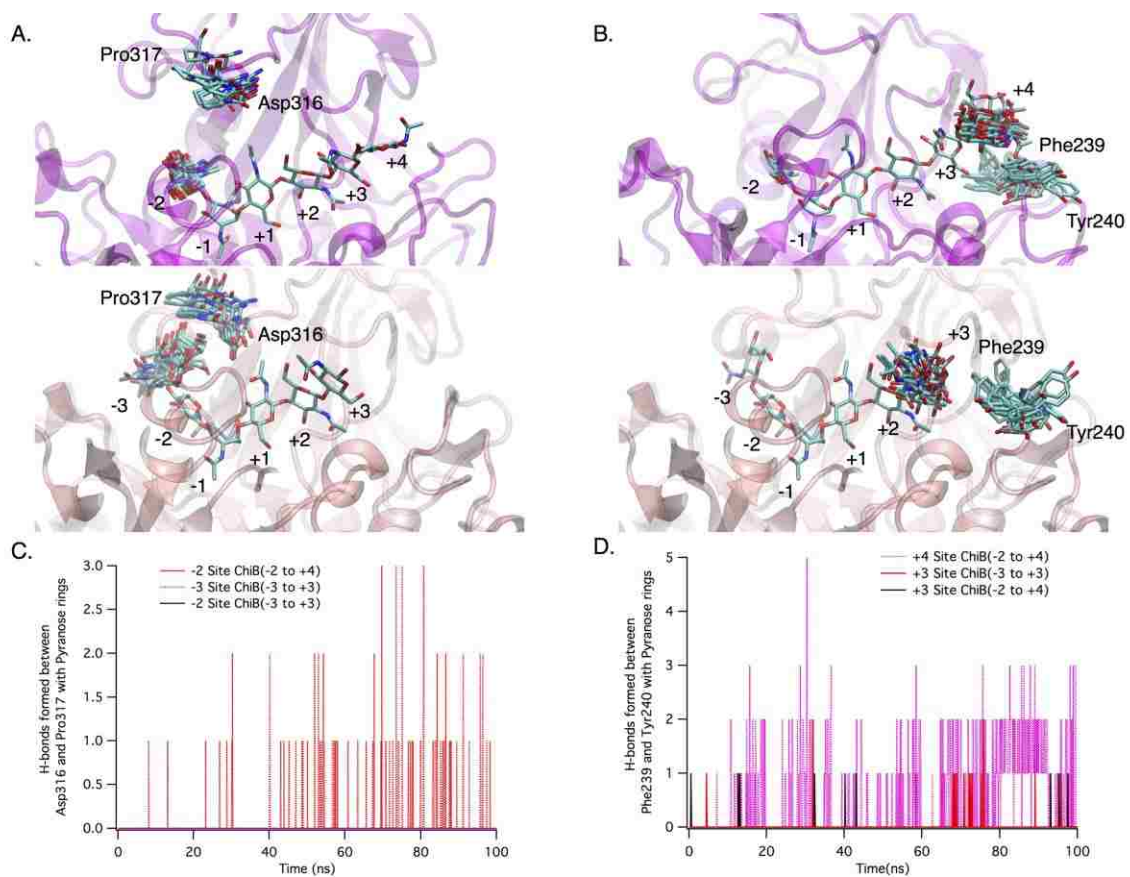


residues were selected based on observed differences in fluctuation within the ChiB backbone as shown in Figure 3.4B of Chapter 3. Here, we define a hydrogen bond as within 3.4 Å of the donor or acceptor and 60° from linear.

Figure A1.2C illustrates that in the presence of 3 product subsites, ChiB residues Asp316 and Pro317 maintain hydrogen bonds with the ligand most of the 100-ns MD simulations. However, no hydrogen bonding was observed when only 2 product subsites were available. Further, while the ligand is bound in the -3 to +3 binding sites, Asp316 and Pro317 does not appear to hydrogen bond with the -2 pyranose ring, which suggests the availability of third pyranose ring on the product side is responsible for enhanced hydrogen bonding with Asp316 and Pro317. This ultimately results in rigidification of ChiB residues from 315 to 322. We anticipate binding a pyranose ring in the -3 product site enhances favorable enthalpic contribution but unfavorable entropic contribution.

Figure A1.2D illustrates that throughout most of the 100-ns simulations, Phe239 and Tyr240 form hydrogen bonds with a pyranose in the +4 binding site. Very sparse hydrogen bonding between Phe239 and Tyr240 was observed in the -3 to +3 binding occupancy. Here, we suspect that favorable enthalpic contributions dominate in the presence of a +4 site pyranose sugar.

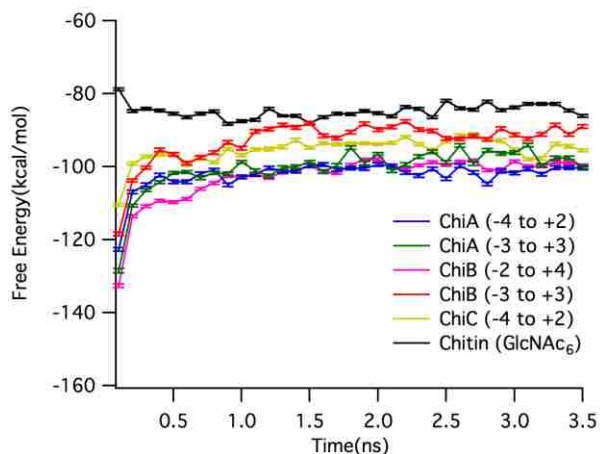
We also note that the binding of a +4 substrate pyranose facilitates interaction with the tunnel entrance residues, Phe239 and Tyr240, strengthening the observations that substrate-binding sites contribute more in binding than product sites. This is in line with a previous experimental study indicating that Tyr240 is critical to binding and hydrolysis.<sup>201</sup>



**Figure A1.2** Time-resolved hydrogen bonding of key ChiB active site regions and the bound (GlcNAc)<sub>6</sub> ligand. (A) Illustration of hydrogen bonds formed between ChiB product-side residues Asp316 and Pro317 and the last product side pyranose ring while the (GlcNAc)<sub>6</sub> ligand occupies -2 to +4 binding sites and -3 to +3 binding sites. (B) Illustration of the hydrogen bonds formed between ChiB substrate-side entrance residues Phe239 and Tyr240 to the first tunnel entrance pyranose ring while ligand binds from -2 to +4 sites and -3 to +3 sites, respectively. (C) Hydrogen bonds formed between selected product-side pyranose rings and residues Asp316 and Pro317. (D) Hydrogen bonds formed between selected substrate-side pyranose rings and residues Phe239 and Tyr240.

### A1.3 Free energy results

As described in the manuscript, the absolute binding free energies were obtained from 35 sequential 0.1 ns calculations, wherein the sequential calculation was restarted from the previous calculation's final configuration. Figure A1.3 illustrates the time progression of the calculated Gibbs free energy of each of these 0.1 ns blocks. The final 2 ns of each free energy calculation were used to determine the average absolute binding free energy.



**Figure A1.3** Gibbs free energy calculated from 35 sequential 0.1 ns FEP/  $\lambda$ -REMD calculations. The statistical certainty of each 0.1 ns calculation was estimated using MBAR.<sup>108</sup>

## **A2 Supporting Information for Aromatic-mediated Carbohydrate Recognition in Processive *Serratia marcescens* Chitinases**

Appendix section A2 has been reprinted with permission from Jana et al.,<sup>112</sup> Copyright © 2016, American Chemical Society.

### **A2.1 Computational methods**

#### *Initial system preparation*

The crystal structures from PDB entries 1EHN and 1E6N were used as input for construction of the ChiA and ChiB models, respectively.<sup>45, 79</sup> The overall quality of the crystal structures was assessed by carefully looking into the global quality parameters such as the resolution,  $R_{\text{free}}$  value, and the normalized real space R value (RSRZ). Only crystal structures of chitinases having resolution below 2.5 Å,  $R_{\text{free}}$  value below 0.25, and RSRZ below 2 were considered for our simulation. Additionally, the quality of the ligand positioning in the chitinase active site was evaluated by checking local structure parameters such as the real space correlation coefficient (RSCC) and the local ligand density fit (LLDF) scores that generally correspond to the calculated electron density fit of the modeled ligand to the observed electron density from the diffraction. The ligand (each GlcNAc unit) in the active site whose RSCC were above 0.8 and LLDF were below or close to 2 were used for our simulation.

Using the crystal structures, in total, 10 separate MD simulations were constructed: ChiA and ChiB wild-type with and without a bound (GlcNAc)<sub>6</sub> ligand, and the chitinase variants ChiA-W167A, ChiA-W275A, ChiA-F396A, ChiB-W97A, ChiB-W220A, and ChiB-F190A with a bound (GlcNAc)<sub>6</sub> ligand state.

The ChiA wild-type simulations were constructed from the 1EHN entry in the PDB (ChiA E315Q complexed with chitooctase) by reversing the E315Q mutation. The entire catalytic domain and fused *N*-terminal chitin binding domain ChiA structure was used in these models, as cleaving the chitin binding domain from the catalytic domain adversely impacted the stability of the catalytic domain fold during the simulations. Aromatic to alanine variants, W167A, W275A, and F396A, were constructed by simply removing the aromatic residue side chain atoms. In the simulations where the (GlcNAc)<sub>6</sub> ligand was bound to ChiA, the pyranose rings bound to the -4 through +2 subsites of 1EHN were retained, and the remaining two pyranose rings (in the solvent exposed -5 and -6 sites) were deleted. The ligand-free ChiA models were constructed by removing all the 1EHN pyranose rings from the cleft.

The wild-type ChiB model was constructed from the 1E6N entry in the PDB (ChiB E144Q in complex with chitopentaose) by reversing the E144Q mutation. Here, we cleaved the chitin binding domain (up to Leu-448) from the ChiB catalytic domain to minimize computational expense. The stability of the catalytic domain in the absence of the chitin binding domain was maintained over the course of simulations. Again, ChiB variants, W97A, W220A, and F190A, were constructed from the wild-type by removing the aromatic side chain atoms. The (GlcNAc)<sub>6</sub> ligand bound in the -3 to +3 subsites of the ChiB cleft was modeled from the ligands of two different chitinase structures. The protein backbones of PDBs 1E6N and 1OGG were aligned in PyMOL.<sup>85, 144</sup> The pyranose ring in the -2 through +3 binding sites were retained from the 1E6N structure, while the -3 binding site sugar originated from the 1OGG structure (ChiB D142N in complex with allosamidin).

In both ChiA and ChiB ligand-bound simulations, the catalytically competent active site conformation was constructed by manually rotating the *N*-acetyl group of the -1 site pyranose ring and the side chains of the catalytic residues (Asp-313 and Glu-315 in ChiA and Asp-142 and Glu-144 in ChiB) in PyMOL to ensure that the catalytic residues and the -1 sugar reflected the distorted Michaelis complex of Family 18 chitinases.<sup>79</sup> The manual rotation of the -1 pyranose and catalytic residues was followed by additional stepwise minimization (100 steps of steepest descent minimization followed by 100 steps of adopted basis Newton Raphson minimization) to make sure the rotation did not adversely affect stability of the protein and the remaining 5 pyranose rings.

CHARMM was used to build and solvate the wild-type and aromatic variants of ChiA and ChiB.<sup>95</sup> The input protonation states of the systems were determined using H++ at pH 6 and internal and external dielectric constants of 10 and 80, respectively.<sup>148-150</sup> The disulfide bonds, between Cys115-Cys120 and Cys195-Cys218 for ChiA, and Cys328-Cys331 for ChiB, were specified based on crystal structures. The constructed systems were solvated with TIP3P water molecules, and sodium counterions were used to make the systems charge neutral. For ChiA, the final system size was approximately 120 Å x 120 Å x 120 Å, totaling roughly 175,000 atoms. The ChiB final system was 80 Å x 80 Å x 80 Å, for a total of approximately 52,000 atoms.

#### *MD simulations protocols*

The solvated systems were minimized, heated, and equilibrated using CHARMM. The minimization of the systems was conducted in a step-wise manner. First, the water molecules were minimized for 10,000 steps with the protein and ligand (if present) held rigid. The harmonic restraints on the ligand were then released, and the protein and water

molecules were minimized for 10,000 steps. Finally, all restraints were removed, and the entire system was minimized for 10,000 steps. The minimized system was heated from 100 K to 300 K in steps of 4 ps with 50 K increments. The system density was equilibrated in the *NPT* ensemble at 1 atm and 300K with the Nosé-Hoover thermostat and barostat for 100 ps using a 2-fs time step.<sup>224-225</sup> For all MD simulations conducted as part of this study, the CHARMM force field with the CMAP correction was used to describe the protein.<sup>95,151-152</sup> The chitin oligomer was described with the CHARMM C36 carbohydrate force field,<sup>194-195</sup> and water was described with the modified TIP3P model.<sup>153-154</sup>

The equilibrated systems were then simulated for 250 ns in the *NVT* ensemble using NAMD.<sup>96</sup> A 2-fs time step was used in the velocity Verlet integration scheme. The Langevin thermostat was used for temperature control in the 250-ns MD simulations.<sup>264</sup> The SHAKE algorithm was used to fix all hydrogen distances for computational efficiency.<sup>227</sup> Non-bonded interactions used the following cutoffs: a non-bonded cutoff distance of 10 Å, a switching distance of 9 Å, and a non-bonded pair list distance of 12 Å. The Particle Mesh Ewald method was used to calculate the long-range electrostatics.<sup>226</sup> The PME grid spacing was 1 Å, and a 6<sup>th</sup> order b-spline and Gaussian distribution width of 0.320 Å was used.

### *TI protocols*

An equilibrated 25-ns snapshot from MD simulations was used as the input coordinates for the TI calculations. We used the dual-topology method with NAMD to determine the relative change in free energy.<sup>94, 196-197</sup> The electrostatic and van der Waals calculations were decoupled in separate processes, each of which included 15  $\lambda$ -windows

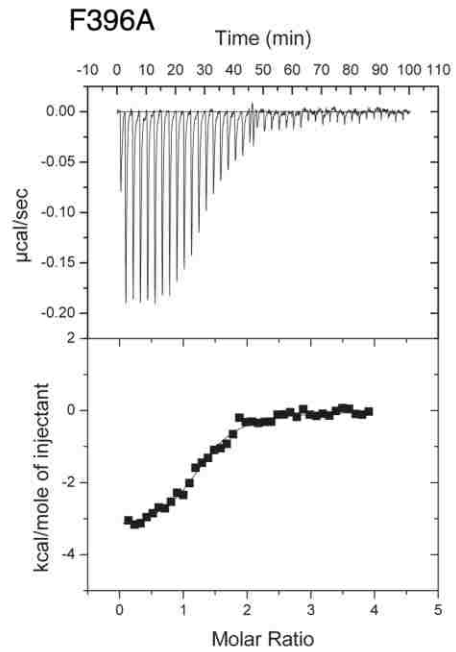
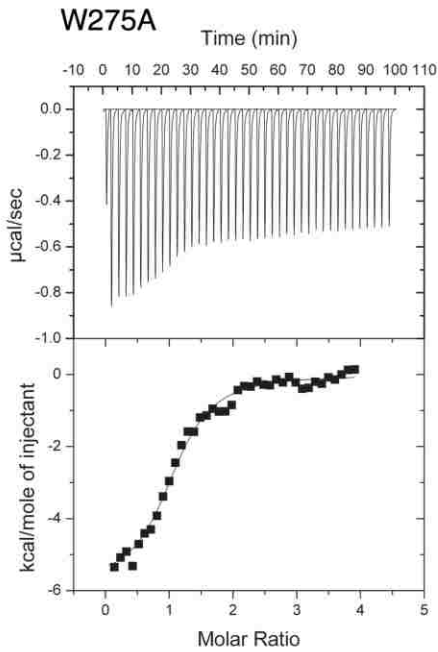
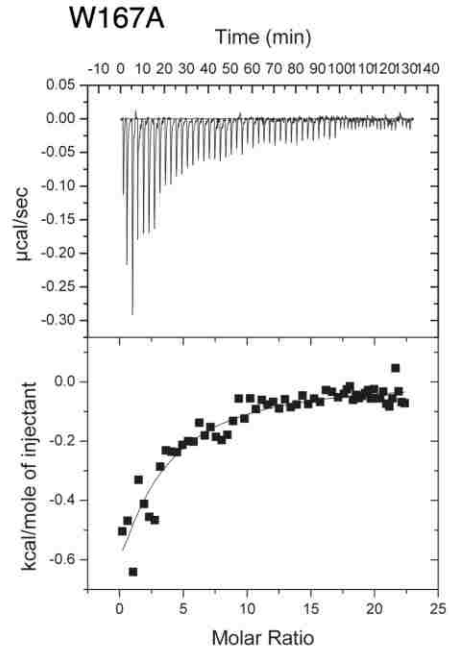
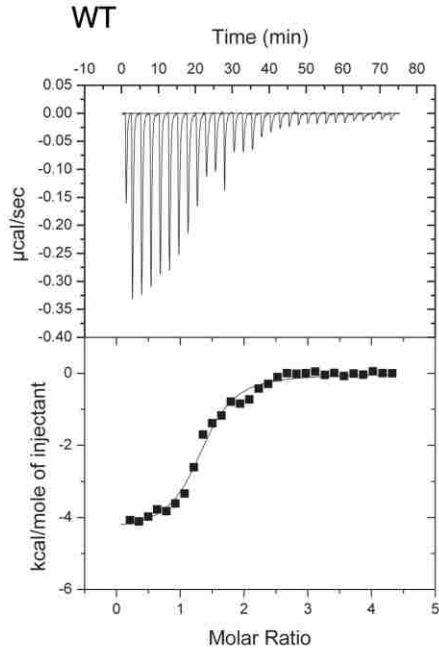
ranging from  $\lambda$  values of 0 to 1 (total of 30 simulations per mutation). The windows were divided as follows: 0.0, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85, 0.9, 0.95 and 1.0. More closely coupled windows were selected near the endpoints to improve the accuracy of the calculations.<sup>110</sup> The electrostatics and van der Waals calculations were equilibrated (0.5 ns) before collecting 14.5 ns of TI data. The change in free energy,  $\Delta G$ , for each set of simulations was evaluated using cubic spline Gaussian quadrature numerical method to integrate  $dU/d\lambda$  over  $\lambda = 0$  to 1. The protocol described by Steinbrecher et al. was followed to determine associated error.<sup>111</sup>



## A2.2 Binding isotherms from isothermal titration calorimetry

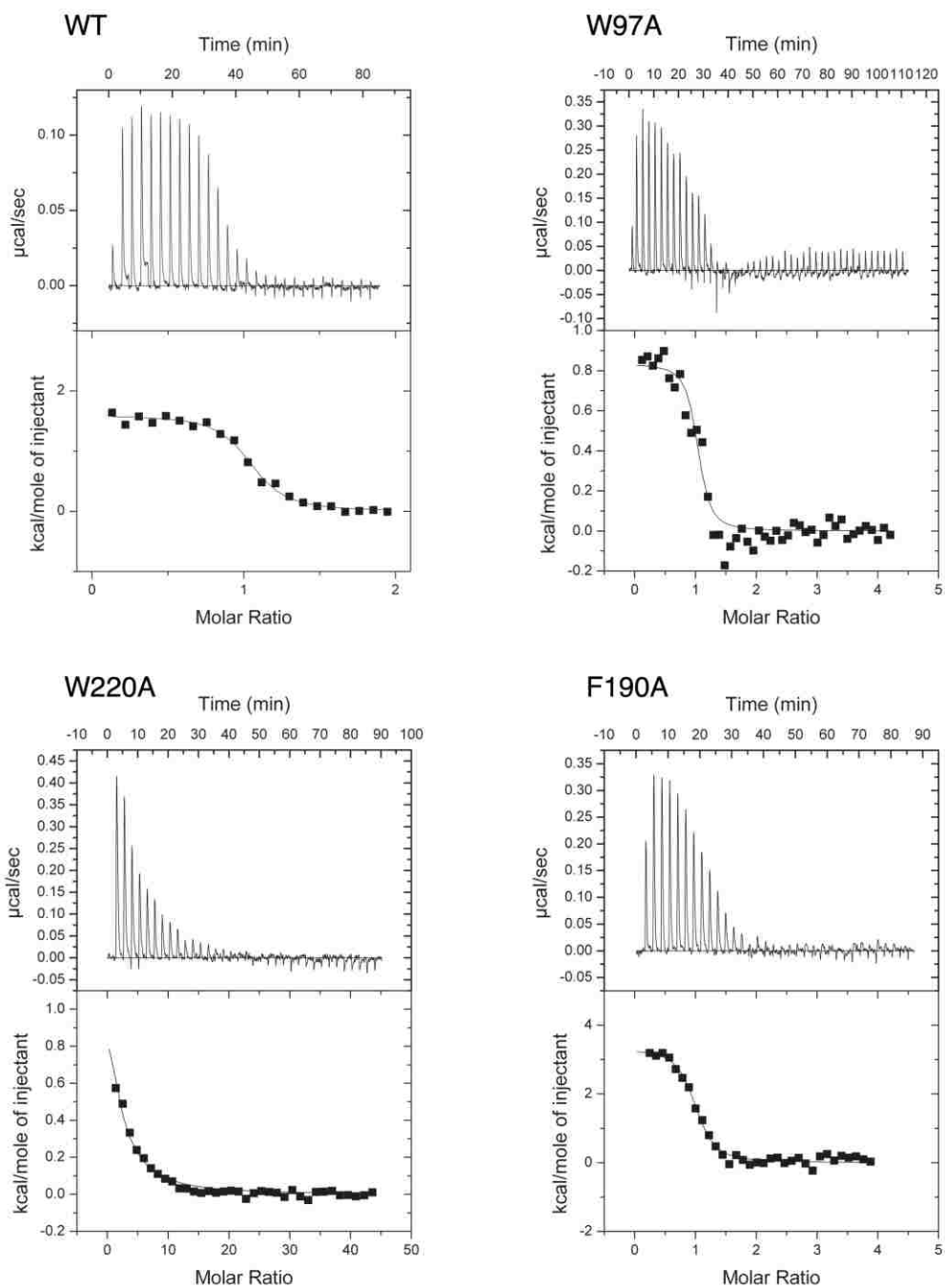
A

ChiA



B

ChiB



**Figure A2.1** Binding isotherms with theoretical fits for (A) ChiA wild-type<sup>119</sup> and its aromatic variants at  $t = 30$  °C and pH = 6.0 as well as for (B) ChiB wild-type<sup>138</sup> and its aromatic variants at  $t = 20$  °C and pH = 6.0. In each case, the ligand was (GlcNAc)<sub>6</sub>.

### A2.3 Experimental determination of [(GlcNAc)<sub>2</sub>]/[GlcNAc] for ChiA

Apparent processivity ( $P^{app}$ ) is defined as the average number of consecutive catalytic cycles performed per initiated processive run along the crystalline substrate. It has previously been shown that the [(GlcNAc)<sub>2</sub>]/[GlcNAc] ratio can be used as a measure of an enzyme's apparent processive ability.<sup>74, 78</sup> However, this approach has several pitfalls, including the assumption of the exclusive formation of odd numbered oligosaccharides from the first cleavage. This assumption may not hold in the case of all chitinases, as different enzymes may have different preferences for the orientation of the chain end relative to the polymer surface or a different probability of endo-mode initiation.<sup>74</sup> Additionally, the [(GlcNAc)<sub>2</sub>]/[GlcNAc] ratio as a measure of processive ability requires 50% / 50% occupancy of the -3 to +3 and +2 to -4 or -2 to +4 subsites, which we know may not absolutely be true and likely changes with mutations in the active site.<sup>74, 145</sup> Moreover,  $P^{app}$  tends to decrease as the substrate is consumed, most likely because the substrate becomes enriched with recalcitrant regions where there are fewer obstacle-free paths for processive enzyme attachment.<sup>64, 78, 233</sup> It is thus important to assess processivity during the early stages of the reaction.

Unlike  $P^{app}$ , which is dependent on the substrate, intrinsic processivity ( $P^{intr}$ ) is the average number of successive catalytic events before dissociation of an “ideal” substrate and is thus the upper limit of  $P^{app}$ .  $P^{intr}$  was recently determined for ChiA-WT, ChiA-W167A, and ChiA-W275A on  $\alpha$ -chitin, showing the wild type to have an approximately 3.5 and 1.5 fold higher intrinsic processivity than ChiA-W167A and ChiA-W275A, respectively. It was also shown that these three enzymes have a relatively high probability of endo-mode initiation with the values for the two mutants being even

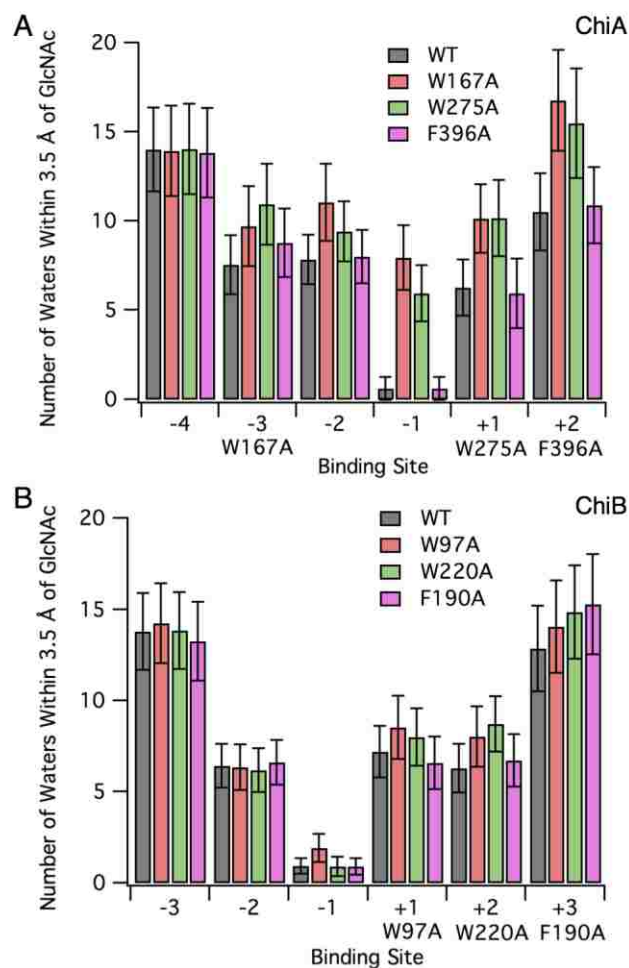
higher than that for the wild type.<sup>71</sup>

Here, we have determined from initial degradation of  $\beta$ -chitin  $[(\text{GlcNAc})_2]/[\text{GlcNAc}]$  ratios of  $27.2 \pm 1.8$ ,  $36.9 \pm 4.9$ , and  $15.3 \pm 0.6$  for ChiA-W167A, ChiA-W275A, and ChiA-F396A, respectively. The value for ChiA-WT has previously been determined to be  $30.1 \pm 1.5$ .<sup>78</sup> Taken at face value, this means that ChiA-W167A and W275A show a similar, initial processive ability as the wild type, while ChiA-F396A is less processive. However, as described above, we cannot extricate the effects of substrate occupancy and endo-initiated activity from the  $[(\text{GlcNAc})_2]/[\text{GlcNAc}]$  values, and thus,  $[(\text{GlcNAc})_2]/[\text{GlcNAc}]$  ratios are imperfect measures of processive ability in ChiA. Accordingly, we are unable to determine the processive ability of ChiA-F396A on  $\beta$ -chitin using  $[(\text{GlcNAc})_2]/[\text{GlcNAc}]$  ratios.

#### **A2.4 Additional results from MD simulations and TI calculations**

##### *Active site solvation*

Active site solvation has been determined from MD simulations for wild-type ChiA, ChiB, and the aromatic variants for comparison to the entropic energy changes determined by ITC. To determine active site solvation, we counted the number of water molecules present within 3.5 Å of the ligand on a per-binding-site basis in VMD every 0.1 ns. The value for a given binding site was then averaged over the 250 ns simulations. A higher number of water molecules is generally indicative of unfavorable contributions to entropy due to solvation effects.



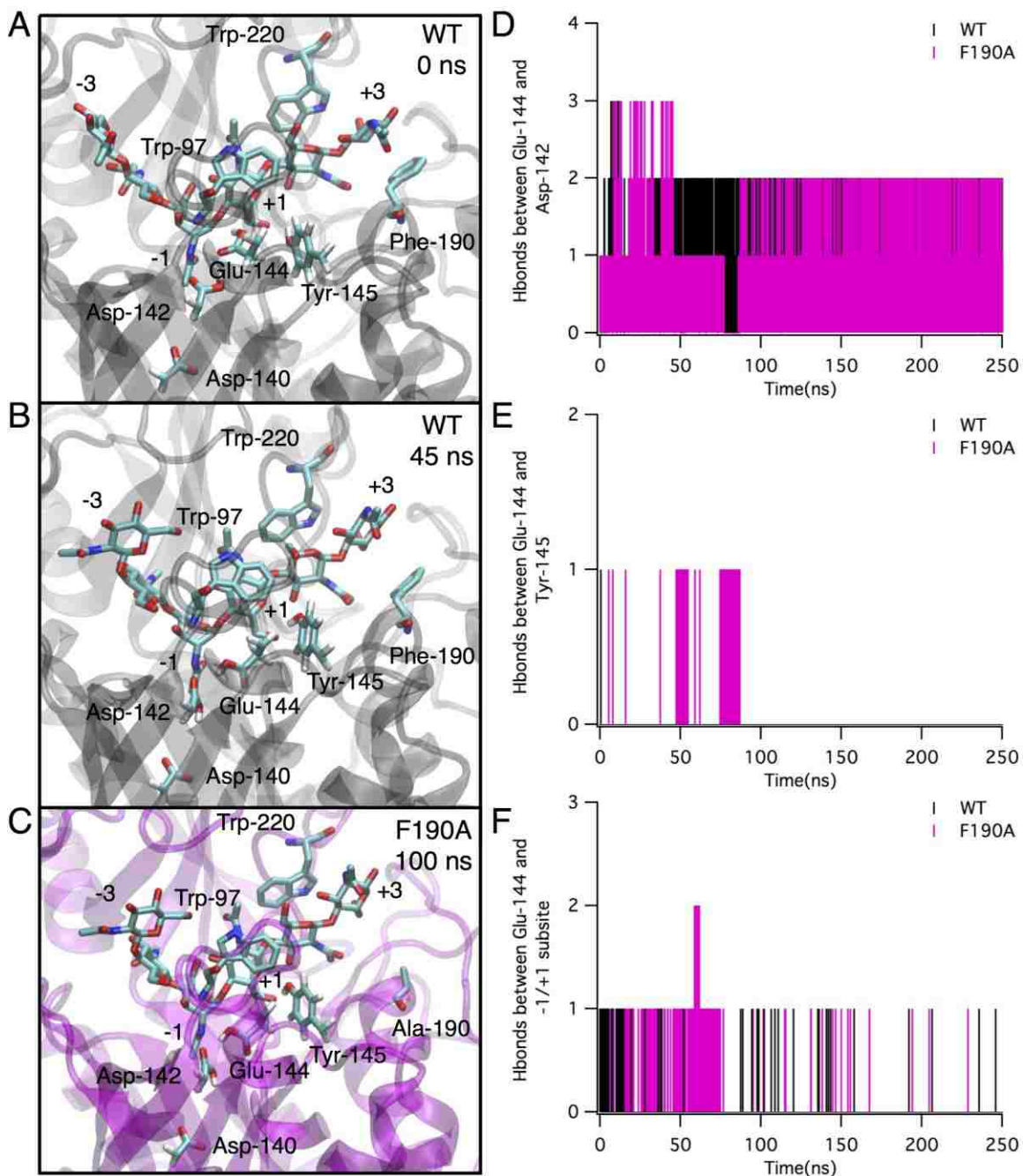
**Figure A2.2** Active site solvation is represented by the average number of water molecules within 3.5 Å of the ligand in each of the six binding sites in wild-type and aromatic variants of (A) ChiA and (B) ChiB. Labels below the  $x$ -axis indicate the position of the aromatic residues relative to the binding site. Error bars were obtained through 2.5 ns block averaging.

*RMSF of ChiB catalytic residues*

The MD simulations were initialized such that the Glu-144 side chain formed a hydrogen bond with both Asp-142 and the glycosidic linkage between the +1 and -1 pyranose rings; however, in each case, the Glu-144 side chain eventually rotated such

that only the Asp-142 hydrogen bond persisted. In wild-type ChiB, this conformational change occurred after 45 ns (Figure A2.3A & B). In both ChiB-W220A (not shown) and ChiB-F190A (Figure A2.3C), the reorientation of the Glu-144 side chain occurred after ~100 ns. In W97A, the side chain of Asp-142 rotated almost immediately after the simulation started (within 10 ns) to interact with Asp-140 instead of Glu-144 and remained in that state for the remainder of the simulation (not shown). Though the catalytic side chain of Glu-144 and Asp-142 behave similarly in W220A and F190A after 100 ns, the W220A mutation still results in higher fluctuation in the catalytic residues, as the mutation also affects the flexibility of Tyr-214 and Asp-215.

The ChiB-F190A conformational change proceeds via a slightly different route than wild-type, which also adds to the early-stage increase in RMSF of the catalytic residues. The ChiB-F190A Glu-144 side chain hydrogen bonds with the Asp-142 side chain (Figure A2.3D) throughout the simulation and the glycosidic linkage intermittently (Figure A2.3F), as does wild-type. However, the reduced interaction of Glu-144 in ChiB-F190A with the glycosidic linkage encourages hydrogen bond formation with Tyr-145 (Figure A2.3E), which does not occur in wild-type. This behavior subsides after the simulation reaches equilibrium.



**Figure A2.3** (A) Snapshot from the ChiB wild-type (WT) MD simulation at 0 ns, where the Glu-144 forms a hydrogen bond with Asp-142 and the glycosidic oxygen between -1 and +1 sites. (B) Snapshot from the ChiB WT simulation at 45 ns, where the Glu-144 side chain reorients to interact with only Asp-142. This state is maintained through the remainder of the simulation. (C) Snapshot of ChiB-F190A at 100 ns, where the Glu-144

side chain reorients in the same fashion as ChiB-WT Glu-144, interacting with only Asp-142, which persisted through the remainder of the simulation. Panels (D) through (E) illustrate hydrogen bond formation (Hbonds) over the 250 ns simulation between: (D) Glu-144 and Asp-142 in ChiB-WT and ChiB-F190A; (E) Glu-144 and Tyr-145 in ChiB-WT and ChiB-F190A; and (F) Glu-144 and the glycosidic oxygen between the -1/+1 subsite in ChiB-WT and ChiB-F190A. A distance cut off of 3.4 Å and an angle cutoff of 60° from linear was used to determine the number of hydrogen bonds formed between the selections.<sup>155</sup>

### *TI data*

As described in Chapter-2 of the dissertation (Figure 2.5), the relative change in binding free energy due to mutation ( $\Delta\Delta G_{TI}$ ) is determined by calculating individual electrostatic and van der Waals components in the presence of a ligand (bound) and in the absence of a ligand (free) (Table A2.1). The sum of the electrostatic and van der Waals components describe the free energy of a given state ( $\Delta G$ ), where  $\Delta G_{Bound(M-WT)}$  and  $\Delta G_{Free(M-WT)}$  describe the bound and free ligand states, respectively. The difference of  $\Delta G_{Bound(M-WT)}$  and  $\Delta G_{Free(M-WT)}$ , as described in Equation 2.17 (Chapter-2) of the dissertation, is the relative change in binding free energy,  $\Delta\Delta G_{TI}$ . As an example, in the ligand bound state of ChiA, the mutation of residue Trp-167 to alanine causes a free energy change,  $\Delta G_{Bound(M-WT)}$ , of  $12.8 \pm 0.4$  kcal/mol, which is the sum of the  $9.3 \pm 0.1$  kcal/mol from electrostatic interactions and the  $3.5 \pm 0.4$  kcal/mol from van der Waals (Table A2.1). Similarly in the ligand free state, the mutation changes free energy,  $\Delta G_{Free(M-WT)}$ , by  $7.8 \pm 0.1$  kcal/mol (electrostatic:  $5.7 \pm 0.0$  kcal/mol and van der Waals:



$2.1 \pm 0.1$  kcal/mol). The difference of  $\Delta G_{Bound(M-WT)}$  and  $\Delta G_{Free(M-WT)}$  gives a  $\Delta\Delta G_{TI}$  of  $5.0 \pm 0.4$  kcal/mol, which is the overall effect of mutating Trp-167 to alanine upon ligand binding.

**Table A2.1** Detailed ligand binding free energies calculated from TI

ChiA	Energy	Bound		Free	
		Energy (kcal/mol)	Error (kcal/mol)	Energy (kcal/mol)	Error (kcal/mol)
W167A	Electrostatics	9.3	$\pm 0.1$	5.7	$\pm 0.0$
	VDW	3.5	$\pm 0.4$	2.1	$\pm 0.1$
	$\Delta G$	12.8	$\pm 0.4$	7.8	$\pm 0.1$
	$\Delta\Delta G_{TI}$ (kcal/mol)	$5.0 \pm 0.4$			
W275A	Electrostatics	8.5	$\pm 0.1$	5.2	$\pm 0.0$
	VDW	2.5	$\pm 0.3$	2.2	$\pm 0.1$
	$\Delta G$	11.0	$\pm 0.3$	7.4	$\pm 0.1$
	$\Delta\Delta G_{TI}$ (kcal/mol)	$3.6 \pm 0.4$			
F396A	Electrostatics	4.3	$\pm 0.0$	3.5	$\pm 0.0$
	VDW	-2.8	$\pm 0.1$	-1.8	$\pm 0.1$
	$\Delta G$	1.5	$\pm 0.1$	1.7	$\pm 0.1$
	$\Delta\Delta G_{TI}$ (kcal/mol)	$-0.2 \pm 0.2$			
ChiB	Energy	Bound		Free	
		Energy (kcal/mol)	Error (kcal/mol)	Energy (kcal/mol)	Error (kcal/mol)
W97A	Electrostatics	8.4	$\pm 0.1$	6.8	$\pm 0.1$
	VDW	2.4	$\pm 0.4$	1.2	$\pm 0.2$
	$\Delta G$	10.8	$\pm 0.4$	8.0	$\pm 0.2$
	$\Delta\Delta G_{TI}$ (kcal/mol)	$2.8 \pm 0.4$			
W220A	Electrostatics	9.6	$\pm 0.1$	7.8	$\pm 0.0$
	VDW	2.9	$\pm 0.3$	1.2	$\pm 0.1$
	$\Delta G$	12.5	$\pm 0.3$	9.0	$\pm 0.1$
	$\Delta\Delta G_{TI}$ (kcal/mol)	$3.5 \pm 0.3$			
F190A	Electrostatics	-8.4	$\pm 0.0$	-7.7	$\pm 0.0$
	VDW	6.3	$\pm 0.1$	5.9	$\pm 0.1$
	$\Delta G$	-2.1	$\pm 0.1$	-1.8	$\pm 0.1$
	$\Delta\Delta G_{TI}$ (kcal/mol)	$-0.3 \pm 0.2$			

## A2.5 Multiple sequence alignment

```

>gi|48425130|SmarChiB|      58 .....LECAWDPATNDA.....K
>gi|497397203|DspChiB|    269 .....ANGEVVVGDVYADTDKAFPGD..CWEFG...C
>gi|573461615|LputChiB|   225 .....TNGEIVVGDYADTDKFFPGD..CWDPG...C
>gi|550993861|LsacChiB|   221 .....ANGEIVVGDYADTDKAFPGD..CWDPG...C
>gi|557622580|BtoyChiB|     1 MQNKEVPTNGTLVLGEPWADVTKSYPSGTTWEDCDKYA
>gi|163938376|BweiChiB|   105 LQNKEVPTNGTLVLGEPWADVTKSYPSGTTWEDCDKYA
>gi|376264400|BcerChiB|   105 LQNKEVPTNGTLVLGEPWADVTKSYPSGTTWEDCDKYA
>gi|345548065|GsteChiB|    66 LQNKEVPTNGTLVLGEPWADVTKSYPSGTTWEDCDKYA
>gi|259558661|BthuChiB|    73 LQNKEVPTNGTLVLGEPWADVTKSYPSGTTWEDCDKYA
>gi|56962420|BclaChiB|   101 G.VIDVPTNGSIVMGDPWIDAQKSNPGD..TWDE....P
>gi|15613479|BhalChiB|   102 G.VIDVPTNGSIVMGDPWIDAQKSNPGD..TWDE....P
>gi|498185052|PelgChiB|  332 G.NINVPTNGSIVQGDWADTGMSPYGD..TWDQ....P
>gi|517582257|PsanChiB|  105 GQTIHVPTNGTIVLGDWIDTGTKQFPGD..VWDQ....P
>gi|927653|KzopChiB|     106 SQTINVPTNGTIVLGDWIDTGTKTFAGD..TWDQ....P
>gi|20150306|BcirChiB|    74 SQTINVPTNGTIVLGDWIDTGTKTFAGD..TWDQ....P
>gi|90023510|SdegChiB|   262 .....LECFIDVEKAD.....A
>gi|321173846|PspChiB|    58 .....LECAWDPATNDA.....K
>gi|518650018|SplyChiB|   58 .....LECAWDPATNDA.....K
>gi|22652066|SliqChiB|    58 .....LECAWDPATNDA.....K
>gi|157371713|SproChiB|   58 .....LECAWDPATNDA.....K
consensus>70
.....ng..vlgd.w.d..k...g....wd.....

```

```

>gi|48425130|SmarChiB|      71 ARDVVNRRLTALKAHNPSLRIMFSIGGWYYSNDLGVSHA
>gi|497397203|DspChiB|    296 KRGNFNQLNKLKQKHPHLKTLISVGGWTWSNN.....
>gi|573461615|LputChiB|   252 KRGNFNQLQKLKAAYPHLKTLISVGGWTWSGG.....
>gi|550993861|LsacChiB|   248 KRGNFNQLTKLKQKYPHLKTLISVGGWTWSGG.....
>gi|557622580|BtoyChiB|     39 RCGNFGELKRLKAKYPHLKTIISVGGWTWSNR.....
>gi|163938376|BweiChiB|   143 RCGNFGELKRLKAKYPHLKTIISVGGWTWSNR.....
>gi|376264400|BcerChiB|   143 RCGNFGELKRLKAKYPHLKTIISVGGWTWSNR.....
>gi|345548065|GsteChiB|   104 RCGNFGELKRLKAKYPHLKTIISVGGWTWSNR.....
>gi|259558661|BthuChiB|   111 RCGNFGELKRLKAKYPHLKTIISVGGWTWSNR.....
>gi|56962420|BclaChiB|   132 IRGNFKQLQKLKEEHPHLKTLISVGGWSWSNR.....
>gi|15613479|BhalChiB|   133 LRGNFKQLNKLKEEHPHLKTLISVGGWTWSNR.....
>gi|498185052|PelgChiB|  363 LKGSFNQLIKLKKANPNLKTLSVGGWSWSNR.....
>gi|517582257|PsanChiB|  137 YAGNINQLNKLKQANPHLKTIISVGGWTWSNR.....
>gi|927653|KzopChiB|     138 IAGNINQLNKLKQINPNLKTIIISVGGWTWSNR.....
>gi|20150306|BcirChiB|   106 IAGNINQLNKLKQTNPNLKTIIISVGGWTWSNR.....
>gi|90023510|SdegChiB|   275 ETQIIAELQALKNWNADLKILFVGGWAESNDAAETVS
>gi|321173846|PspChiB|    71 ARDVVNRRLTALKAHNPSLRIMFSIGGWYYSNDLGVSHA
>gi|518650018|SplyChiB|   71 ARDVVGRRLTALKAHNPSLRIMFSIGGWYYSNDLGVSHA
>gi|22652066|SliqChiB|    71 ARDVVGRRLTALKAHNPSLRIMFSIGGWYYSNDLGVSHA
>gi|157371713|SproChiB|   71 ARDVVSRRLTALKAHNPNLRIMFSIGGWYYSNDLGVSHA
consensus>70
..g...qL..LK...p.Lkt.is!GGW.wSn.....

```

```

>gi|48425130|SmarChiB| 175 QALPYQLTIAGAGGAFFLSRYYSKLLAQIVAPLDYINLM
>gi|497397203|DspChiB| 401 E...YLLTIASGANPNYV..KNTELDQVQAQLDWINIM
>gi|573461615|LputChiB| 357 K...YLLTIASGASPNYV..QNTELGQVKDYVDWINIM
>gi|550993861|LsacChiB| 353 H...YLLTIAGASPSYY..NDNTELIDKVKNYVDWINIM
>gi|557622580|BtoyChiB| 146 Q...YLLTIASGASQRYA..DHTELKKISQILDWINIM
>gi|163938376|BweiChiB| 250 Q...YLLTIASGASQRYA..DHTELKKISQILDWINIM
>gi|376264400|BcerChiB| 250 Q...YLLTIASGASQRYA..DHTELKKISQILDWINIM
>gi|345548065|GsteChiB| 211 Q...YLLTIASGASQRYA..DHTELKKISQILDWINIM
>gi|259558661|BthuChiB| 218 Q...YLLTIASGASQRYA..DHTELKKISQILDWINIM
>gi|56962420|BclaChiB| 239 E...YLVTIASGASSEYV..ENNKLAEIAEVVDWINIM
>gi|15613479|BhalChiB| 240 D...YLLTIASGASPGYV..ENNKLNEIAEIVDWINIM
>gi|498185052|PelgChiB| 470 Q...YLLTIASGAGPNYV..NNTELSKIAQTLDWINIM
>gi|517582257|PsanChiB| 244 H...YLLTIASGASPAFV..NNTELGNIASIVDWINIM
>gi|927653|KzopChiB| 245 K...YLLTIASGASTTYA..ANTELAKIASIVDWINIM
>gi|20150306|BcirChiB| 213 K...YLLTIASGASATYA..ANTELAKIAAIVDWINIM
>gi|90023510|SdegChiB| 373 ..NGELVTIAGAGGAFFLSRYYSKLLAAIVEQLDFINLM
>gi|321173846|PspChiB| 175 QALPYQLTIAGAGGAFFLSRYYSKLLPQIVASLDYINLM
>gi|518650018|SplyChiB| 175 QALPYQLTIAGAGGAFFLSRYYSKLLPQIVASLDYINLM
>gi|22652066|SliqChiB| 175 QALPYQLTIAGAGGAFFLSRYYSKLLPQIVASLDYINLM
>gi|157371713|SproChiB| 175 QALPYQLTIAGAGGAFFLSRYYSKLLPQIVASLDYINLM
consensus>70 q...y11TIA.ga...%.....1...!....DwINiM

```

```

>gi|48425130|SmarChiB| 213 TYDLAGPWEKVT..NHQAALFGDAAGPTFYNALREANL
>gi|497397203|DspChiB| 434 TYDFHGGWENVS..GHNAPLYFDPKDPS.....
>gi|573461615|LputChiB| 390 TYDFHGGWENVS..GHNAPLYVDPSDPS.....
>gi|550993861|LsacChiB| 387 TYDFRGAWENAN..GHNAPLYADPSDPH.....
>gi|557622580|BtoyChiB| 179 TYDFHGGWEATS..NHNAALYKDPNDPA.....
>gi|163938376|BweiChiB| 283 TYDFHGGWEATS..NHNAALYKDPNDPA.....
>gi|376264400|BcerChiB| 283 TYDFHGGWEATS..NHNAALYKDPNDPT.....
>gi|345548065|GsteChiB| 244 TYDFHGGWEATS..NHNAALYKDPNDPA.....
>gi|259558661|BthuChiB| 251 TYDFHGGWEATS..NHNAALYKDPNDPA.....
>gi|56962420|BclaChiB| 272 TYDFNGGWQNTS..GHNAPLYFDPATAD.....
>gi|15613479|BhalChiB| 273 TYDFNGGWQNIS..GHNAPLYYDPATAN.....
>gi|498185052|PelgChiB| 503 TYDFHGGWESKS..GHNAPLYFDPADNS.....
>gi|517582257|PsanChiB| 277 TYDFNGGWQMTS..AHNAPLYFDPAAEA.....
>gi|927653|KzopChiB| 278 TYDFNGAWQKVS..AHNALLNYDPAASA.....
>gi|20150306|BcirChiB| 246 TYDFNGAWQKIS..AHNAPLNYDPAASA.....
>gi|90023510|SdegChiB| 409 TYDLNGPWNGVTKINFHAHLYGNNQEPRFYNALREADL
>gi|321173846|PspChiB| 213 TYDLAGPWEKVT..NHQAALFGDAAGPTFYNALREANL
>gi|518650018|SplyChiB| 213 TYDLAGPWEKIT..NHQAALFGDSAGPTFYNALREANL
>gi|22652066|SliqChiB| 213 TYDLAGPWEKIT..NHQAGLFGDSAGPTFYNALREANL
>gi|157371713|SproChiB| 213 TYDLAGPWEKIT..NHQAGLFGDSAGPTFYNALREANL
consensus>70 TYDf..GW#.....hNA.Ly.#p..p.....

```

**Figure A2.4** Multiple sequence alignment of species phylogenetically related to *S. marcescens* ChiB. The multiple sequence alignment was generated with Clustal Omega<sup>265-267</sup> and prepared for publication using the ESPrift 3.0 server.<sup>268</sup> The cutoff consensus used was 70%. Fully conserved residues are shown in bold white characters on a red background. Chemically similar residues are shown in bold black characters boxed in yellow. In the consensus line, uppercase letters represent full conservation of the

residues at a given sequence location (100%), and lowercase letters represent similarity greater than 70%. The symbol “.” represents no conservation at a given location. Additional symbols include: “!” representing Ile or Val, “\$” representing Leu or Met, “%” representing Phe or Tyr, and “#” representing Asn, Asp, Gln, or Glu. ChiB residues relevant for the present study are indicated by black arrows. The alignment shows that Trp-97 and Trp-220 of *S. marcescens* ChiB are well-conserved aromatic residues, while Phe-190 is quite variable. The short abbreviations of the species used for sequence comparisons are: Smar = *Serratia marcescens*, Dsp = *Desmospora sp.*, Lput = *Laceyella putida*, Lsac = *Laceyella sacchari*, Btoy = *Bacillus toyonensis*, Bwei = *Bacillus weihenstephanensis*, Bcer = *Bacillus cereus*, Gste = *Geobacillus stearothermophilus*, Bthu = *Bacillus thuringiensis serovar kurstaki*, Bela = *Bacillus clausii*, Bhal = *Bacillus halodurans*, Pelg = *Paenibacillus elgii*, Psan = *Paenibacillus sanguinis*, Kzop = *Kurthia zopfii*, Bcir = *Bacillus circulans*, Sdeg = *Saccharophagus degradans*, Psp = *Pseudomonas sp.*, Sply = *Serratia plymuthica*, Sliq = *Serratia liquefaciens*, Spro = *Serratia proteamaculans*.

## **A3 Supporting Information for Carbohydrate-Polar Interactions dictate Substrate Binding and Processivity in a Family 18 Chitinase**

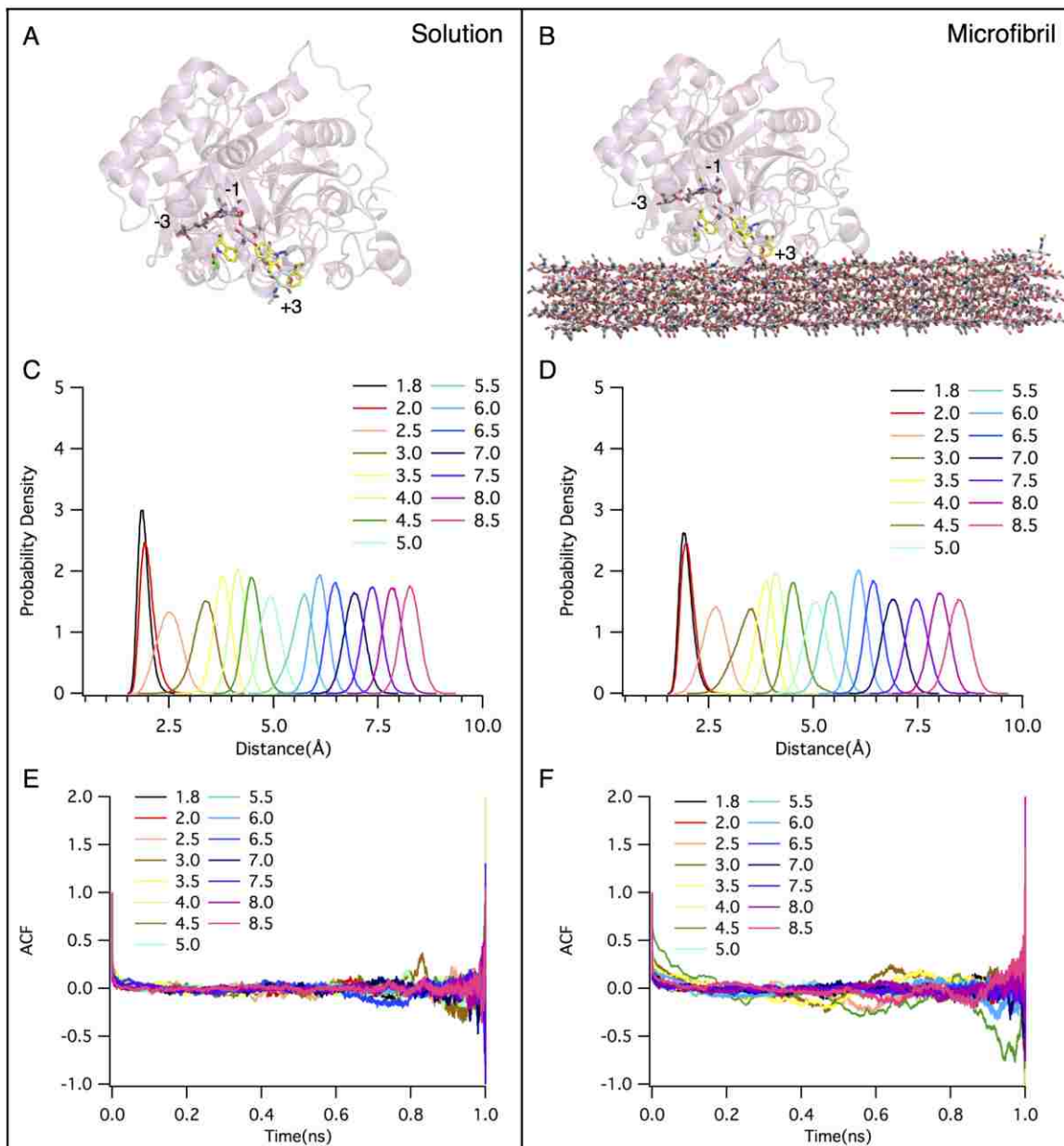
### **A3.1 TI data**

As described in the manuscript, the relative change in ligand binding free energy ( $\Delta\Delta G_{TI}$ ) is defined as the difference of  $\Delta G_{Bound}$  and  $\Delta G_{Free}$ , which are computed in ligand bound and ligand free states.  $\Delta G$  at each state is the sum of electrostatic and van der Waals components of free energy, determined separately from TI. For example,  $\Delta G_{Bound}$  for the D316A variant is  $132.07 \pm 0.3$  kcal/mol (Table A3.1), which is a combination of both electrostatic and van der Waals free energy contributions (electrostatic:  $135.1 \pm 0.2$  kcal/mol; VDW:  $-2.4 \pm 0.2$  kcal/mol). Similarly,  $\Delta G_{Free}$  for the D316A variant is  $133.0 \pm 0.1$  kcal/mol (electrostatic:  $135.9 \pm 0.1$  kcal/mol; VDW:  $-2.9 \pm 0.1$  kcal/mol). The difference of  $\Delta G_{Bound}$  and  $\Delta G_{Free}$  results in  $\Delta\Delta G_{TI}$ , which is  $-0.3 \pm 0.3$  kcal/mol for the D316A mutation.

**Table A3.1** Detailed electrostatic and VDW components of  $\Delta G$ , determined in the presence and absence of the ligand using the TI methodology.

ChiB	Energy	Protein ligand complex Bound		Protein without the ligand Free	
		Energy (kcal/mol)	Error (kcal/mol)	Energy (kcal/mol)	Error (kcal/mol)
D316A	Electrostatics	135.1	$\pm 0.2$	135.9	$\pm 0.1$
	VDW	-2.4	$\pm 0.2$	-2.9	$\pm 0.1$
	$\Delta G$	132.7	$\pm 0.3$	133.0	$\pm 0.1$
	$\Delta\Delta G_{TI}$ (kcal/mol)	$-0.3 \pm 0.3$			
Y145A	Electrostatics	-3.0	$\pm 0.1$	-4.4	$\pm 0.1$
	VDW	6.7	$\pm 0.3$	5.0	$\pm 0.4$
	$\Delta G$	3.7	$\pm 0.3$	0.6	$\pm 0.4$
	$\Delta\Delta G_{TI}$ (kcal/mol)	$3.1 \pm 0.5$			
R294A	Electrostatics	144.7	$\pm 0.3$	139.6	$\pm 0.2$
	VDW	-1.3	$\pm 0.1$	-0.8	$\pm 0.2$
	$\Delta G$	143.4	$\pm 0.3$	138.8	$\pm 0.3$
	$\Delta\Delta G_{TI}$ (kcal/mol)	$4.6 \pm 0.5$			
E221A	Electrostatics	127.6	$\pm 0.3$	127.2	$\pm 0.3$
	VDW	2.5	$\pm 0.1$	1.1	$\pm 0.1$
	$\Delta G$	130.1	$\pm 0.3$	128.3	$\pm 0.3$
	$\Delta\Delta G_{TI}$ (kcal/mol)	$1.8 \pm 0.4$			

### A3.2 Convergence analysis of umbrella sampling



**Figure A3.1** Window overlap and autocorrelation function (ACF) evaluated from umbrella sampling simulations. Cases considered for in the umbrella sampling simulations included: (A) a ChiB-hexasaccharide complex in solution, and (B) a ChiB- $\alpha$ -chitin microfibril complex in solution. The reaction coordinate used for umbrella sampling was the distance between the Trp-97 HE1 atom and the Asp-316 OD2 atom,

divided into 15 windows. The windows were as follows: 1.8 Å, 2.0 Å, 2.5 Å, 3.0 Å, 3.5 Å, 4.0 Å, 4.5 Å, 5.0 Å, 5.5 Å, 6.0 Å, 6.5 Å, 7.0 Å, 7.5 Å, 8.0 Å, and 8.5 Å, all under a harmonic bias potential of 10 kcal/mol/Å<sup>2</sup>. Panels (C) and (D) show the histogram each window, which was used to check for sufficient overlap between the adjacent windows before constructing the potential of mean force (PMF). Panels (E) and (F) show the ACF corresponding to each window in the absence and presence of the microfibril. The autocorrelation time (the time at which the ACF reach  $e^{-1}$ ) was used for error analysis in WHAM.



### A3.3 Multiple sequence alignment

```

>gi|48425130|SmarChiB|      58  .....LECAWDPATNDA.....K
>gi|497397203|DspChiB|    269  .....ANGEVVVGDVYADTDKAFP GD..CWE PG...C
>gi|573461615|LputChiB|   225  .....TNGEIVVGDYADTDKAFP GD..CWD PG...C
>gi|550993861|LsacChiB|   221  .....ANGEIVVGDYADTDKAFP GD..CWD PG...C
>gi|557622580|BtoyChiB|     1  MQNKEVPNGTLVLGEPWADVTKSYPGSGTTWEDCDKYA
>gi|163938376|BweiChiB|   105  LQNKEVPNGTLVLGEPWADVTKSYPGSGTTWEDCDKYA
>gi|376264400|BcerChiB|   105  LQNKEVPNGTLVLGEPWADVTKSYPGSGTTWEDCDKYA
>gi|345548065|GsteChiB|    66  LQNKEVPNGTLVLGEPWADVTKSYPGSGTTWEDCDKYA
>gi|259558661|BthuChiB|    73  LQNKEVPNGTLVLGEPWADVTKSYPGSGTTWEDCDKYA
>gi|56962420|BclaChiB|   101  G.VIDVPSNGSIVMGDPWIDAQKSNP GD..TWE....P
>gi|15613479|BhalChiB|   102  G.VIDVPSNGSIVMGDPWIDAQKSNP GD..TWE....P
>gi|498185052|PelgChiB|   332  G.NINVPNGSIVQGDYADTGMSYPGD..TWDQ....P
>gi|517582257|PsanChiB|   105  GQTIHVPNGTIVLGDYADTGKQFP GD..VWDQ....P
>gi|927653|KzopChiB|      106  SQTINVPNGTIVLGDYADTGKTFAGD..TWDQ....P
>gi|20150306|BcirChiB|    74  SQTINVPNGTIVLGDYADTGKTFAGD..TWDQ....P
>gi|90023510|SdegChiB|   262  .....LECFIDVEKAD.....A
>gi|321173846|PspChiB|     58  .....LECAWDPATNDA.....K
>gi|518650018|SplyChiB|    58  .....LECAWDPATIDA.....K
>gi|22652066|SliqChiB|    58  .....LECAWDPATNDA.....K
>gi|157371713|SproChiB|   58  .....LECAWDPATNDA.....K
consensus>70
.....ng..vlgd.w.d..k...g....wd.....

```

```

>gi|48425130|SmarChiB|      71  ARDVVNRLLTALKAHNPSLRIMFSIGGWYYSNDLGVSHA
>gi|497397203|DspChiB|    296  KRGNFNQLNKLKQKHPHLKTLISVGGWTWSNN.....
>gi|573461615|LputChiB|   252  KRGNFNQLQKLLKAAAYPHLKTLISVGGWTWSGG.....
>gi|550993861|LsacChiB|   248  KRGNFNQLTKLQKQYPHLKTLISVGGWTWSGG.....
>gi|557622580|BtoyChiB|    39  RCGNFGE LKRLKAKYPHLKTLISVGGWTWSNR.....
>gi|163938376|BweiChiB|   143  RCGNFGE LKRLKAKYPHLKTLISVGGWTWSNR.....
>gi|376264400|BcerChiB|   143  RCGNFGE LKRLKAKYPHLKTLISVGGWTWSNR.....
>gi|345548065|GsteChiB|   104  RCGNFGE LKRLKAKYPHLKTLISVGGWTWSNR.....
>gi|259558661|BthuChiB|   111  RCGNFGE LKRLKAKYPHLKTLISVGGWTWSNR.....
>gi|56962420|BclaChiB|   132  IRGNFKQLQKLLKKEEHPHLKTLISVGGWSWSNR.....
>gi|15613479|BhalChiB|   133  LRGNFKQLNKLKKEEHPHLKTLISVGGWTWSNR.....
>gi|498185052|PelgChiB|   363  LKGSFNQLIKLKKANPNLKTLISVGGWSWSNR.....
>gi|517582257|PsanChiB|   137  YAGNIHQ LNKLKQANPHLKTLISVGGWTWSNR.....
>gi|927653|KzopChiB|      138  IAGNIHQ LNKLKQINPNLKTLISVGGWTWSNR.....
>gi|20150306|BcirChiB|   106  IAGNIHQ LNKLKQTNPNLKTLISVGGWTWSNR.....
>gi|90023510|SdegChiB|   275  ETQIIAELQALKNWNADLKILFSVGGWAE SNDAAETVS
>gi|321173846|PspChiB|     71  ARDVVNRLLTALKAHNPSLRIMFSIGGWYYSNDLGVSHA
>gi|518650018|SplyChiB|    71  ARDVVGRLLTALKAHNPSLRIMFSIGGWYYSNDLGVSHA
>gi|22652066|SliqChiB|    71  ARDVVGRLLTALKAHNPSLRIMFSIGGWYYSNDLGVSHA
>gi|157371713|SproChiB|   71  ARDVVSRLLTALKAHNPNLRIMFSIGGWYYSNDLGVSHA
consensus>70
..g...qL..LK...p.Lkt.is!GGW.wSn.....

```

```

>gi | 48425130 | SmarChiB | 109
>gi | 497397203 | DspChiB | 328
>gi | 573461615 | LputChiB | 284
>gi | 550993861 | LsacChiB | 280
>gi | 557622580 | BtoyChiB | 71
>gi | 163938376 | BweiChiB | 175
>gi | 376264400 | BcerChiB | 175
>gi | 345548065 | GsteChiB | 136
>gi | 259558661 | BthuChiB | 143
>gi | 56962420 | BclaChiB | 164
>gi | 15613479 | BhalChiB | 165
>gi | 498185052 | PelgChiB | 395
>gi | 517582257 | PsanChiB | 169
>gi | 927653 | KzopChiB | 170
>gi | 20150306 | BcirChiB | 138
>gi | 90023510 | SdegChiB | 313
>gi | 321173846 | PspChiB | 109
>gi | 518650018 | SplyChiB | 109
>gi | 22652066 | SliqChiB | 109
>gi | 157371713 | SproChiB | 109
consensus>70

```

```

      . . . . .
NYVNAVKTPASRAKFAQSCVRIMKDYGFDGVDINWEYP
      . FSDAALTDASRSKFAADSAVRFIREWGF DGVLDWEYP
      . FSDAALTDASRTKFAADSAVRFVRYGFDGVDLDWEYP
      . FSDAALTDASRTRFAADSAVRFIRQYGF DGVLDWEYP
      . FSDMAADEKTRKVF AESTVDFLRAYGFDGVDLDWEYP
      . FSDMAADEKTRKVF AESTVAF LREYGF DGVLDWEYP
      . FSDMAADEKTRKI F AESTVAF LREYGF DGVLDWEYP
      . FSDMAADEKTRKVF AESTVDFLREYGF DGVLDWEYP
      . FSDMAADEKTRKVF AESTVDFLREYGF DGVLDWEYP
      . FSDVAAATKE TRENFA NSAVDFVRQYGF DGVLDWEYP
      . FSDVAAATKE TRENFA NSAVEFIRKYGF DGVLDWEYP
      . FSDVAAADPATRSNFAKSAVD FIRKYQFDGVLDWEYP
      . FSDVAASQV TRETFA NSAVDFLRKFNFDGVLDWEYP
      . FSDVAAATAA TREVFA NSAVDFLRKYNFDGVLDWEYP
      . FSDVAAATAA TREVFA NSAVDFLRKYNFDGVLDWEYP
RYRDA.FAPANREHFVSCVAFMQQHGF DGVLDWEYP
NYVNAVKTPASRAKFAQSCVRIMKDYGF DGVLDWEYP
NYVNAVKTPAARTKFAQSCVRIMKDYGF DGVLDWEYP
NYVNAVKTPAARTKFAQSCVRIMKDYGF DGVLDWEYP
NYVNAVKTPAARTKFAQSCVRIMKDYGF DGVLDWEYP
      .%s#.a.....R..FaqS.V.flrdygfDg!D.#WEYP

```

```

>gi | 48425130 | SmarChiB | 147
>gi | 497397203 | DspChiB | 365
>gi | 573461615 | LputChiB | 321
>gi | 550993861 | LsacChiB | 317
>gi | 557622580 | BtoyChiB | 108
>gi | 163938376 | BweiChiB | 212
>gi | 376264400 | BcerChiB | 212
>gi | 345548065 | GsteChiB | 173
>gi | 259558661 | BthuChiB | 180
>gi | 56962420 | BclaChiB | 201
>gi | 15613479 | BhalChiB | 202
>gi | 498185052 | PelgChiB | 432
>gi | 517582257 | PsanChiB | 206
>gi | 927653 | KzopChiB | 207
>gi | 20150306 | BcirChiB | 175
>gi | 90023510 | SdegChiB | 350
>gi | 321173846 | PspChiB | 147
>gi | 518650018 | SplyChiB | 147
>gi | 22652066 | SliqChiB | 147
>gi | 157371713 | SproChiB | 147
consensus>70

```

```

      . . . . .
QA.....AEVDGFIAALQEI RTLLNQQTITDGR
VEGGLV..DGRPEDKRNFTLL LQKVRQKLDEAGKQDGGK
VGGGLN..PGRPEDKQNF TLL LQKMRKLNAGQDGGK
VGGGLN..PGRPEDKQNF TLL LQKMRKLNAGQDGGK
GVETIPGGSYRPE DKQNF TLL LQDVRNALNKAGTEDGK
GVETIPGGSYRPE DKQNF TLL LQDVRNALTKAGAEDGK
GVETIPGGSYRPE DKQNF TLL LQDVRNALTKAGAEDGK
GVETIPGGSYRPE DKQNF TLL LQDVRNALTKAGAEDGK
GVETIPGGSYRPE DKQNF TLL LQDVRNALNKAGAEDGK
VSGGLPGNSRRPE DKENHLL LQEV RNKLNAGQEDGK
VSGGLPGNSRRPE DKENHLL LQEV RDKLDEAGQEDGK
VSGGLAGNSKRPE DKQNF TLL LQAVRSELNTAGSADGK
VSGGLPGNSYRAEDKQNYTKLLQEI RTKLDAAGAADGK
VSGGLDGNNSKRSGDKQNYTLL LSKIREKLDAGA VDGK
VSGGLDGNNSKRPE DKQNYTLL LSKIREKLDAGA VDGK
RA.....EDVDNFIAGLAAMRNQLDARG.....
QA.....AEVDGFIAALQEI RTLLNQQTITDGR
QS.....SEVDGFVAALQEI RALLNQQTITDGR
QS.....SEVDGFVAALQEI RTLLNQQTILADGR
QS.....SEVDGFVAALQEI RTLLNQQTILADGR
      .....r.e#kqnf..lLq.vR..Ln.ag..dgg

```

>gi|48425130|SmarChiB| 175  
 >gi|497397203|DspChiB| 401  
 >gi|573461615|LputChiB| 357  
 >gi|550993861|LsacChiB| 353  
 >gi|557622580|BtoyChiB| 146  
 >gi|163938376|BweiChiB| 250  
 >gi|376264400|BcerChiB| 250  
 >gi|345548065|GsteChiB| 211  
 >gi|259558661|BthuChiB| 218  
 >gi|56962420|BclaChiB| 239  
 >gi|15613479|BhalChiB| 240  
 >gi|498185052|PelgChiB| 470  
 >gi|517582257|PsanChiB| 244  
 >gi|927653|KzopChiB| 245  
 >gi|20150306|BcirChiB| 213  
 >gi|90023510|SdegChiB| 373  
 >gi|321173846|PspChiB| 175  
 >gi|518650018|SplyChiB| 175  
 >gi|22652066|SliqChiB| 175  
 >gi|157371713|SproChiB| 175  
 consensus>70

↓

```

175 QALP YQLTIAAGAGGAFFLSRYYSKLAQIVAPLDYINLM
401 E... YLLTIASGANPNYV..KNTELDQVQAQLDWINIM
357 K... YLLTIASGASPNIYV..QNTTELGVKDYVDWINIM
353 H... YLLTIAAGASPSYY..NDNTEIDKVKNYVDWINIM
146 Q... YLLTIASGASQRYA..DHTELEKKSISQILDWINIM
250 Q... YLLTIASGASQRYA..DHTELEKKSISQILDWINIM
250 Q... YLLTIASGASQRYA..DHTELEKKSISQILDWINIM
211 Q... YLLTIASGASQRYA..DHTELEKKSISQILDWINIM
218 Q... YLLTIASGASQRYA..DHTELEKKSISQILDWINIM
239 E... YLVTIASGASSEYV..ENNKLAETAEVVDWINIM
240 D... YLLTIASGASPGYV..ENNKLEIAEIVDWINIM
470 Q... YLLTIASGAGPNYV..NNTTELSKIAQTLDWINIM
244 H... YLLTIASGASPAFV..NNTTELGNIASIVDWINIM
245 K... YLLTIASGASTTYA..ANTELANIASIVDWINIM
213 K... YLLTIASGASATYA..ANTELAKIAAIVDWINIM
373 . .NGELVTIAGAGGAFFLSRYYSKLAALIVEQLDFINLM
175 QALP YQLTIAAGAGGAFFLSRYYSKLAQIVAPLDYINLM
175 QALP YQLTIAAGAGGAFFLSRYYSKLPQIVASLDYINLM
175 QALP YQLTIAAGAGGAFFLSRYYSKLPQIVASLDYINLM
175 QALP YQLTIAAGAGGAFFLSRYYSKLPQIVASLDYINLM
q...y1lTIA.ga...%.....l...!....DwINiM
  
```

>gi|48425130|SmarChiB| 213  
 >gi|497397203|DspChiB| 434  
 >gi|573461615|LputChiB| 390  
 >gi|550993861|LsacChiB| 387  
 >gi|557622580|BtoyChiB| 179  
 >gi|163938376|BweiChiB| 283  
 >gi|376264400|BcerChiB| 283  
 >gi|345548065|GsteChiB| 244  
 >gi|259558661|BthuChiB| 251  
 >gi|56962420|BclaChiB| 272  
 >gi|15613479|BhalChiB| 273  
 >gi|498185052|PelgChiB| 503  
 >gi|517582257|PsanChiB| 277  
 >gi|927653|KzopChiB| 278  
 >gi|20150306|BcirChiB| 246  
 >gi|90023510|SdegChiB| 409  
 >gi|321173846|PspChiB| 213  
 >gi|518650018|SplyChiB| 213  
 >gi|22652066|SliqChiB| 213  
 >gi|157371713|SproChiB| 213  
 consensus>70

⇓

```

213 TYDLAGPWEKVT..NHQAALFGDAAGPTFYNALREANL
434 TYDFHGGWENV S..GNNAPLYFDPKDP S.....
390 TYDFHGGWENV S..GHNAPLYVDP SDP S.....
387 TYDFRGAWENAN..GHNAPLYADP SDP H.....
179 TYDFHGGWEAT S..NHNAALYKDPNDP A.....
283 TYDFHGGWEAT S..NHNAALYKDPNDP A.....
283 TYDFHGGWEAT S..NHNAALYKDPNDP T.....
244 TYDFHGGWEAT S..NHNAALYKDPNDP A.....
251 TYDFHGGWEAT S..NHNAALYKDPNDP A.....
272 TYDFNGGWQNT S..GHNAPLYFDPATAD.....
273 TYDFNGGWQNI S..GHNAPLYYDPATAN.....
503 TYDFHGGWESKS..GHNAPLYFDPADNS.....
277 TYDFNGGWQMT..AHNAPLYFDPAAEA.....
278 TYDFNGAWQKV S..AHNALNLYDPAASA.....
246 TYDFNGAWQKI S..AHNAPLNLYDPAASA.....
409 TYDLNGPWNQVTKTNFHAHLYGNNQEP R FYNALREADL
213 TYDLAGPWEKVT..NHQAALFGDAAGPTFYNALREANL
213 TYDLAGPWEKI T..NHQAALFGDSAGPTFYNALREANL
213 TYDLAGPWEKI T..NHQAALFGDSAGPTFYNALREANL
213 TYDLAGPWEKI T..NHQAALFGDSAGPTFYNALREANL
TYDf.G.W#.....hnA.Ly.#p..p.....
  
```

```

>gi|48425130|SmarChiB| 249 GWSWEELTRAFPSPFSLTVDAAVQQHLMMEGVP SAKIV
>gi|497397203|DspChiB| 460 .....PGATSFYVDAAVEGHLK.AGVPKKNLI
>gi|573461615|LputChiB| 416 .....PNKDNFNVA AAVQNHIN.AGVPASKIV
>gi|550993861|LsacChiB| 413 .....PKRDTYNIDA AVNGFLN.KGVPASKLV
>gi|557622580|BtoyChiB| 205 .....A...DTK F YVDG A I D V Y K N . E G V P A D K L V
>gi|163938376|BweiChiB| 309 .....A...DTK F YVDG A I D I Y T N . E G V P A D K L V
>gi|376264400|BcerChiB| 309 .....A...DTK F YVDG A I D I Y T N . E G V P A D K L V
>gi|345548065|GsteChiB| 270 .....A...DTN F YVDG A I N I Y T N . E G V P A D K L V
>gi|259558661|BthuChiB| 277 .....A...DTK F YVDG A I D I Y T N . E G V P A D K L V
>gi|56962420|BclaChiB| 298 .....SELP T P E Q F N V E S A V E G H L Q . A G V P G N K L V
>gi|15613479|BhalChiB| 299 .....TELP T P E H F N V E S A V E G H L Q . A G V P E H K L V
>gi|498185052|PelgChiB| 529 .....EDP V N F N V D K A V A N H L N . A G V P A S K L V
>gi|517582257|PsanChiB| 303 .....AGV P S A S T F N G D V A V Q G H L N . A G V P A S K L V
>gi|927653|KzopChiB| 304 .....AGV P D A N T F N V A A G A Q G H L N . A G V P A A K L V
>gi|20150306|BcirChiB| 272 .....AGV P D A N T F N V A A G A Q G H L D . A G V P A A K L V
>gi|90023510|SdegChiB| 447 GLTWEEIVERFPSPFELTVDAAIKQHLMMDD.IPREKIV
>gi|321173846|PspChiB| 249 GWSWEELTRAFPSPFSLTVDAAVQQHLMMEGVP SAKIV
>gi|518650018|SplyChiB| 249 GWSWEELTRAFPSPFSLTVDAAVQQHLMLEGVP SSKIV
>gi|22652066|SliqChiB| 249 GWSWEELTRAFPSPFSLTVDAAGVQQHLMLEGVP SNKIV
>gi|157371713|SproChiB| 249 GWSWEELTRAFPSPFSLTVDAAVQQHLMLEGVP SNKIV
consensus>70
.....p....f.vd.avq.hl...g!P..K.!

```

```

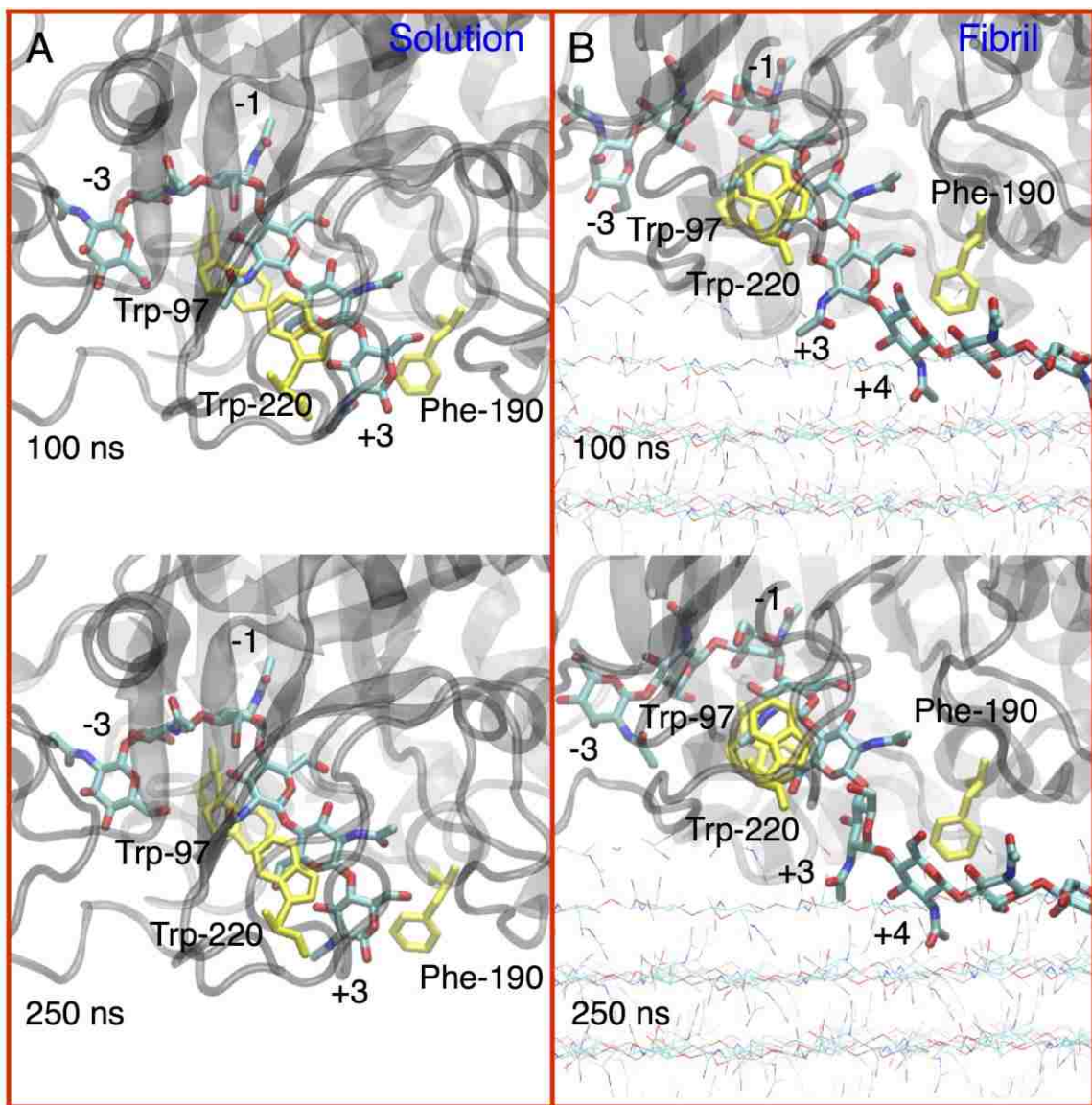
                ↓                ↓
>gi|48425130|SmarChiB| 287 MGVPFYCRAFKGVSGGNGCQYSSHSTPGEDPYPSTIDYW
>gi|497397203|DspChiB| 486 LCMPFYCRGW TGCDGTDHCLYQSCSGP.....SEGTW
>gi|573461615|LputChiB| 442 MCMPFYCRGWQGCKSTNNGEYQSCGTG.....PTGTW
>gi|550993861|LsacChiB| 439 LCMPFYCRGWKDCPATNNGQFQTCNGP.....SQGTW
>gi|557622580|BtoyChiB| 230 LGVPFYCRGWKSCGKDNNGQYQPCKPGSDGKLASKGTW
>gi|163938376|BweiChiB| 334 LGVPFYCRGWKSCGKENNGQYQPCKPGSDGKLASKGTW
>gi|376264400|BcerChiB| 334 LGVPFYCRGWKSCGKENNGQYQPCKPGSDGKLASKGTW
>gi|345548065|GsteChiB| 295 LGVPFYCRGWKSCGKENNGQYQPCKPGSDGKLASKGTW
>gi|259558661|BthuChiB| 302 LGVPFYCRGWKSCGKENNGQYQPCKPGSDGKLASKGTW
>gi|56962420|BclaChiB| 327 VGM PFYCRGWMNCEEANYGEYQH CSP.P.....QDGTW
>gi|15613479|BhalChiB| 328 LCM PFYCRGWSNCDGANQGEYQRCAP.P.....REGTW
>gi|498185052|PelgChiB| 555 LGLPFYCRGWQGCASANNGQYQCSGVS.....QKGTW
>gi|517582257|PsanChiB| 332 LGVPFYCRGWDGVT.GGD CQYQGASGAS.....SSGTW
>gi|927653|KzopChiB| 333 LGVPFYCRGWDGCAQAGNGQYQTC TGG S.....SVGTW
>gi|20150306|BcirChiB| 301 LGVPFYCRGWDGCAQAGNGQYQTC TGG S.....SVGTW
>gi|90023510|SdegChiB| 484 MGV PFYCRAF FNTGSSNTGLYQTFNTPNGDPYVGDASL
>gi|321173846|PspChiB| 287 MGV PFYCRAF KGVSGGNGCQYSSHSTPGEDPYPSTIDYW
>gi|518650018|SplyChiB| 287 MGV PFYCRAF KGVSGSNGCQYSSHSTPGEDPFP GTDYW
>gi|22652066|SliqChiB| 287 MGV PFYCRAF KGVSSSNGCQYSSHSTPGEDPFP GTDYW
>gi|157371713|SproChiB| 287 MGV PFYCRAF KGVSSSNGCQYSSHSTPGEDPFP GTDYW
consensus>70
lGvP%YGrgw.....n.Gq%q.....gtw

```

**Figure A3.2** Multiple sequence alignment of species phylogenetically related to *S. marcescens* ChiB. The aromatic and polar residues included in this study are indicated by black arrows. The abbreviations of the species used for sequence comparisons are: Smar = *Serratia marcescens*, Dsp = *Desmospora sp.*, Lput = *Laceyella putida*, Lsac = *Laceyella sacchari*, Btoy = *Bacillus toyonensis*, Bwei = *Bacillus weihenstephanensis*,

Bcer = *Bacillus cereus*, Gste = *Geobacillus stearothermophilus*, Bthu = *Bacillus thuringiensis serovar kurstaki*, Bcla = *Bacillus clausii*, Bhal = *Bacillus halodurans*, Pelg = *Paenibacillus elgii*, Psan = *Paenibacillus sanguinis*, Kzop = *Kurthia zopfii*, Bcir = *Bacillus circulans*, Sdeg = *Saccharophagus degradans*, Psp = *Pseudomonas sp.*, Sply = *Serratia plymuthica*, Sliq = *Serratia liquefaciens*, Spro = *Serratia proteamaculans*. The sequence alignment comparison was appeared originally in Jana et al.<sup>112</sup>

### A3.4 Snapshots of wild-type MD simulation in solution and in the presence of $\alpha$ -chitin microfibril



**Figure A3.3** The orientation of Phe-190 at the cleft entrance during MD simulation (A) in solution and (B) in presence of crystalline  $\alpha$ -chitin. In solution, Phe-190 faces toward solution; whereas in the presence of chitin, Phe-190 stacks with the +4 pyranose.

## References

1. Klemm, D.; Heublein, B.; Fink, H. P.; Bohn, A., Cellulose: Fascinating Biopolymer and Sustainable Raw Material. *Angewandte Chemie International Edition* **2005**, *44* (22), 3358-3393.
2. Tharanathan, R. N.; Kittur, F. S., Chitin — The Undisputed Biomolecule of Great Potential. *Crit. Rev. Food Sci. Nutr.* **2003**, *43* (1), 61-87.
3. Ragauskas, A. J.; Williams, C. K.; Davison, B. H.; Britovsek, G.; Cairney, J.; Eckert, C. A.; Frederick, W. J.; Hallett, J. P.; Leak, D. J.; Liotta, C. L., et al., The Path Forward for Biofuels and Biomaterials. *Science* **2006**, *311* (5760), 484-489.
4. Huber, G. W.; Chheda, J. N.; Barrett, C. J.; Dumesic, J. A., Production of Liquid Alkanes by Aqueous-Phase Processing of Biomass-Derived Carbohydrates. *Science* **2005**, *308* (5727), 1446-1450.
5. Steen, E. J.; Kang, Y. S.; Bokinsky, G.; Hu, Z. H.; Schirmer, A.; McClure, A.; del Cardayre, S. B.; Keasling, J. D., Microbial Production of Fatty-Acid-Derived Fuels and Chemicals From Plant Biomass. *Nature* **2010**, *463* (7280), 559-U182.
6. M. Wang, C. S., and D. Santini, Effects of Fuel Ethanol Use on Fuel-Cycle Energy and Greenhouse Gas Emissions. Energy, D. o., Ed. Argonne National Laboratory: 1999.
7. Inokuma, K.; Takano, M.; Hoshino, K., Direct Ethanol Production from *N*-Acetylglucosamine and Chitin Substrates by *Mucor species*. *Biochem. Eng. J.* **2013**, *72*, 24-32.
8. Inokuma, K.; Hasunuma, T.; Kondo, A., Ethanol production from *N*-Acetyl-D-Glucosamine by *Scheffersomyces stipitis* strains. *AMB Express* **2016**, *6* (1), 83.
9. Jonas, R.; Farah, L. F., Production and Application of Microbial Cellulose. *Polym. Degrad. Stab.* **1998**, *59* (1), 101-106.
10. Kamel, S.; Ali, N.; Jahangir, K.; Shah, S. M.; El-Glendy, A. A., Pharmaceutical Significance of Cellulose: A Review. *eXPRESS Polymer Letters* **2008**, *2* (11), 758-778.
11. Harish Prashanth, K. V.; Tharanathan, R. N., Chitin/Chitosan: Modifications and Their Unlimited Application Potential—An Overview. *Trends Food Sci. Technol.* **2007**, *18* (3), 117-131.
12. Köping-Höggård, M.; Mel'nikova, Y. S.; Vårum, K. M.; Lindman, B.; Artursson, P., Relationship Between the Physical Shape and the Efficiency of Oligomeric Chitosan as a Gene Delivery System in Vitro and in Vivo. *J. Gene Med* **2003**, *5* (2), 130-141.
13. Himmel, M. E.; Ding, S. Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D., Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production. *Science* **2007**, *315* (5813), 804-807.
14. Beckham, G. T.; Matthews, J. F.; Peters, B.; Bomble, Y. J.; Himmel, M. E.; Crowley, M. F., Molecular-Level Origins of Biomass Recalcitrance: Decrystallization

- Free Energies for Four Common Cellulose Polymorphs. *J. Phys. Chem. B* **2011**, *115* (14), 4118-4127.
15. Beckham, G. T.; Crowley, M. F., Examination of the  $\alpha$ -Chitin Structure and Decrystallization Thermodynamics at the Nanoscale. *J. Phys. Chem. B* **2011**, *115* (15), 4516-4522.
  16. Payne, C. M.; Knott, B. C.; Mayes, H. B.; Hansson, H.; Himmel, M. E.; Sandgren, M.; Ståhlberg, J.; Beckham, G. T., Fungal Cellulases. *Chem. Rev.* **2015**, *115* (3), 1308-1448.
  17. Merino, S. T.; Cherry, J., Progress and Challenges in Enzyme Development for Biomass Utilization. *Biofuels* **2007**, *108*, 95-120.
  18. Wyman, C. E., What is (and is Not) Vital to Advancing Cellulosic Ethanol. *Trends Biotechnol.* **2007**, *25* (4), 153-157.
  19. Arnold, F. H., Combinatorial and Computational Challenges for Biocatalyst Design. *Nature* **2001**, *409*, 253-257.
  20. Arnold, F. H., Design by Directed Evolution. *Acc. Chem. Res.* **1998**, *31*, 125-131.
  21. Jäckel, C.; Kast, P.; Hilvert, D., Protein Design by Directed Evolution. *Annual Review of Biophysics* **2008**, *37*, 153-173.
  22. Beckham, G. T.; Bomble, Y. J.; Bayer, E. A.; Himmel, M. E.; Crowley, M. F., Applications of Computational Science for Understanding Enzymatic Deconstruction of Cellulose. *Curr. Opin. Biotechnol.* **2011**, *22* (2), 231-238.
  23. Beckham, G. T.; Ståhlberg, J.; Knott, B. C.; Himmel, M. E.; Crowley, M. F.; Sandgren, M.; Sørli, M.; Payne, C. M., Towards a Molecular-Level Theory of Carbohydrate Processivity in Glycoside Hydrolases. *Curr. Opin. Biotechnol.* **2014**, *27*, 96-106.
  24. Hamre, A. G. Thermodynamic Aspects of Processive Enzymatic Degradation of Recalcitrant Polysaccharides. Norwegian University of Life Sciences, 2015.
  25. Salmon, S.; Hudson, S. M., Crystal Morphology, Biosynthesis, and Physical Assembly of Cellulose, Chitin, and Chitosan. *Journal of Macromolecular Science-Reviews in Macromolecular Chemistry and Physics* **1997**, *C37* (2), 199-276.
  26. Kurita, K., Chitin and Chitosan: Functional Biopolymers from Marine Crustaceans. *Mar. Biotechnol.* **2006**, *8* (3), 203-226.
  27. Imai, T.; Sugiyama, J.; Itoh, T.; Horii, F., Almost Pure I $\alpha$  Cellulose in the Cell Wall of *Glaucocestis*. *J. Struct. Biol.* **1999**, *127* (3), 248-257.
  28. Horii, F.; Hirai, A.; Kitamaru, R., CP/MAS Carbon-13 NMR Spectra of the Crystalline Components of Native Celluloses. *Macromolecules* **1987**, *20* (9), 2117-2120.
  29. Nishiyama, Y.; Langan, P.; Chanzy, H., Crystal Structure and Hydrogen-Bonding System in Cellulose I $\beta$  From Synchrotron X-ray and Neutron Fiber Diffraction. *J. Am. Chem. Soc.* **2002**, *124* (31), 9074-9082.



30. Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P., Crystal Structure and Hydrogen Bonding System in Cellulose I $\alpha$  From Synchrotron X-ray and Neutron Fiber Diffraction. *J. Am. Chem. Soc.* **2003**, *125* (47), 14300-14306.
31. Langan, P.; Nishiyama, Y.; Chanzy, H., X-ray Structure of Mercerized Cellulose II at 1 Å Resolution. *Biomacromolecules* **2001**, *2* (2), 410-416.
32. Wada, M.; Chanzy, H.; Nishiyama, Y.; Langan, P., Cellulose III Crystal Structure and Hydrogen Bonding by Synchrotron X-ray and Neutron Fiber Diffraction. *Macromolecules* **2004**, *37* (23), 8548-8555.
33. Chundawat, S. P. S.; Beckham, G. T.; Himmel, M. E.; Dale, B. E., Deconstruction of Lignocellulosic Biomass to Fuels and Chemicals. *Annual Review of Chemical and Biomolecular Engineering, Vol 2* **2011**, *2*, 121-145.
34. Chundawat, S. P. S.; Bellesia, G.; Uppugundla, N.; da Costa Sousa, L.; Gao, D.; Cheh, A. M.; Agarwal, U. P.; Bianchetti, C. M.; Phillips, G. N. J.; Langan, P., et al., Restructuring the Crystalline Cellulose Hydrogen Bond Network Enhances its Depolymerization Rate. *J. Am. Chem. Soc.* **2011**, *133* (29), 11163-11174.
35. Rinaudo, M., Chitin and Chitosan: Properties and Applications. **2006**, *31* (7), 603-632.
36. Minke, R.; Blackwell, J., The Structure of  $\alpha$ -Chitin. *J. Mol. Biol.* **1978**, *120* (2), 167-181.
37. Sikorski, P.; Hori, R.; Wada, M., Revisit of  $\alpha$ -Chitin Crystal Structure Using High Resolution X-ray Diffraction Data. *Biomacromolecules* **2009**, *10* (5), 1100-1105.
38. Ogawa, Y.; Kimura, S.; Wada, M., Electron Diffraction and High-Resolution Imaging on Highly-Crystalline  $\beta$ -Chitin Microfibril. *J. Struct. Biol.* **2011**, *176*, 83-90.
39. Gardner, K. H.; Blackwell, J., Refinement of Structure of  $\beta$ -Chitin. *Biopolymers* **1975**, *14* (8), 1581-1595.
40. Kobayashi, K.; Kimura, S.; Togawa, E.; Wada, M., Crystal Transition Between Hydrate and Anhydrous  $\beta$ -Chitin Monitored by Synchrotron X-ray Fiber Diffraction. *Carbohydr. Polym.* **2010**, *79* (4), 882-889.
41. Sawada, D.; Nishiyama, Y.; Langan, P.; Forsyth, V. T.; Kimura, S.; Wada, M., Water in Crystalline Fibers of Dihydrate  $\beta$ -Chitin Results in Unexpected Absence of Intramolecular Hydrogen Bonding. *PLoS One* **2012**, *7* (6), e39376.
42. Horn, S. J.; Sørbotten, A.; Synstad, B.; Sikorski, P.; Sørlie, M.; Vårum, K. M.; Eijsink, V. G. H., Endo/Exo Mechanism and Processivity of Family 18 Chitinases Produced by *Serratia marcescens*. *FEBS J.* **2006**, *273* (3), 491-503.
43. Rouvinen, J.; Bergfors, T.; Teeri, T.; Knowles, J. K. C.; Jones, T. A., 3-Dimensional Structure of Cellobiohydrolase-II from *Trichoderma reesei*. *Science* **1990**, *249* (4967), 380-386.
44. Knott, B. C.; Crowley, M. F.; Himmel, M. E.; Ståhlberg, J.; Beckham, G. T., Carbohydrate-Protein Interactions That Drive Processive Polysaccharide Translocation in

- Enzymes Revealed from a Computational Study of Cellobiohydrolase Processivity. *J. Am. Chem. Soc.* **2014**, *136* (24), 8810-8819.
45. Papanikolau, Y.; Prag, G.; Tavlas, G.; Vorgias, C. E.; Oppenheim, A. B.; Petratos, K., High Resolution Structural Analyses of Mutant Chitinase A Complexes with Substrates Provide New Insight into the Mechanism of Catalysis. *Biochemistry* **2001**, *40* (38), 11338-11343.
46. Boraston, A. B.; Bolam, D. N.; Gilbert, H. J.; Davies, G. J., Carbohydrate-Binding Modules: Fine-tuning Polysaccharide Recognition. *Biochem. J.* **2004**, *382* (Pt 3), 769-781.
47. Nimlos, M. R.; Beckham, G. T.; Matthews, J. F.; Bu, L.; Himmel, M. E.; Crowley, M. F., Binding Preferences, Surface Attachment, Diffusivity, and Orientation of a Family 1 Carbohydrate-Binding Module on Cellulose. *J. Biol. Chem.* **2012**, *287* (24), 20603-20612.
48. Payne, C. M.; Resch, M. G.; Chen, L.; Crowley, M. F.; Himmel, M. E.; Taylor, L. E., 2nd; Sandgren, M.; Ståhlberg, J.; Stals, I.; Tan, Z., et al., Glycosylated Linkers in Multimodular Lignocellulose Degrading Enzymes Dynamically Bind to Cellulose. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (36), 14646-51.
49. Henrissat, B., A Classification of Glycosyl Hydrolases Based on Amino Acid Sequence Similarities. *Biochem. J.* **1991**, *280* (2), 309-316.
50. Henrissat, B.; Bairoch, A., New Families in the Classification of Glycosyl Hydrolases Based on Amino Acid Sequence Similarities. *Biochem. J.* **1993**, *293* (3), 781-788.
51. Davies, G.; Henrissat, B., Structures and Mechanisms of Glycosyl Hydrolases. *Structure* **1995**, *3* (9), 853-859.
52. Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B., The Carbohydrate-Active Enzymes Database (CAZy) in 2013. *Nucleic Acids Res.* **2014**, *42* (D1), D490-D495.
53. Aleshin, A. E.; Hoffman, C.; Firsov, L. M.; Honzatko, R. B., Refined Crystal Structures of Glucoamylase from *Aspergillus awamori* var. X100. *J. Mol. Biol.* **1994**, *238* (4), 575-591.
54. Larsson, A. M.; Bergfors, T.; Dultz, E.; Irwin, D. C.; Roos, A.; Driguez, H.; Wilson, D. B.; Jones, T. A., Crystal Structure of *Thermobifida fusca* Endoglucanase Cel6A in Complex with Substrate and Inhibitor: The Role of Tyrosine Y73 in Substrate Ring Distortion. *Biochemistry* **2005**, *44* (39), 12915-12922.
55. Koshland, D. E., Stereochemistry and the Mechanism of Enzymatic Reactions. *Biol. Rev. (Camb.)* **1953**, *28* (4), 416-436.
56. Rye, C. S.; Withers, S. G., Glycosidase Mechanisms. *Curr. Opin. Chem. Biol.* **2000**, *4* (5), 573-580.
57. Mayes, H. B.; Knott, B. C.; Crowley, M. F.; Broadbelt, L. J.; Ståhlberg, J.; Beckham, G. T., Who's on Base? Revealing the Catalytic Mechanism of Inverting Family 6 Glycoside Hydrolases. *Chemical Science* **2016**, *7* (9), 5955-5968.

58. Knott, B. C.; Momeni, M. H.; Crowley, M. F.; Mackenzie, L. F.; Gotz, A. W.; Sandgren, M.; Withers, S. G.; Ståhlberg, J.; Beckham, G. T., The Mechanism of Cellulose Hydrolysis by a Two-Step, Retaining Cellobiohydrolase Elucidated by Structural and Transition Path Sampling Studies. *J. Am. Chem. Soc.* **2014**, *136* (1), 321-329.
59. Momeni, M. H. Structural Insights into the Catalytic Mechanism, Protein Dynamics, Inhibition and Thermostability of GH7 Cellobiohydrolases. Swedish University of Agricultural Sciences, Uppsala, 2014.
60. Breyer, W. A.; Matthews, B. W., A Structural Basis for Processivity. *Protein Sci.* **2001**, *10* (9), 1699-1711.
61. Payne, C. M.; Jiang, W.; Shirts, M. R.; Himmel, M. E.; Crowley, M. F.; Beckham, G. T., Glycoside Hydrolase Processivity Is Directly Related to Oligosaccharide Binding Free Energy. *J. Am. Chem. Soc.* **2013**, *135* (50), 18831-18839.
62. Divne, C.; Ståhlberg, J.; Teeri, T. T.; Jones, T. A., High-Resolution Crystal Structures Reveal How a Cellulose Chain is Bound in the 50 Angstrom Long Tunnel of Cellobiohydrolase I from *Trichoderma reesei*. *J. Mol. Biol.* **1998**, *275* (2), 309-325.
63. Munoz, I. G.; Ubhayasekera, W.; Henriksson, H.; Szabo, I.; Pettersson, G.; Johansson, G.; Mowbray, S. L.; Stahlberg, J., Family 7 Cellobiohydrolases from *Phanerochaete chrysosporium*: Crystal Structure of the Catalytic Module of Cel7D (CBH58) at 1.32 Å Resolution and Homology Models of the Isozymes. *J. Mol. Biol.* **2001**, *314* (5), 1097-1111.
64. Kurasin, M.; Våljamäe, P., Processivity of Cellobiohydrolases Is Limited by the Substrate. *J. Biol. Chem.* **2011**, *286* (1), 169-177.
65. Vaaje-Kolstad, G.; Horn, S. J.; Sørlie, M.; Eijsink, V. G., The Chitinolytic Machinery of *Serratia marcescens* - A Model System for Enzymatic Degradation of Recalcitrant Polysaccharides. *FEBS J.* **2013**, *280* (13), 3028-3049.
66. Sakon, J.; Irwin, D.; Wilson, D. B.; Karplus, P. A., Structure and Mechanism of Endo/Exocellulase E4 from *Thermomonospora fusca*. *Nat. Struct. Biol.* **1997**, *4* (10), 810-818.
67. Li, Y. C.; Irwin, D. C.; Wilson, D. B., Processivity, Substrate Binding, and Mechanism of Cellulose Hydrolysis by *Thermobifida fusca* Cel9A. *Appl. Environ. Microbiol.* **2007**, *73* (10), 3165-3172.
68. von Ossowski, I.; Ståhlberg, J.; Koivula, A.; Piens, K.; Becker, D.; Boer, H.; Harle, R.; Harris, M.; Divne, C.; Mahdi, S., et al., Engineering the Exo-Loop of *Trichoderma reesei* Cellobiohydrolase, Cel7A. A Comparison with *Phanerochaete chrysosporium* Cel7D. *J. Mol. Biol.* **2003**, *333* (4), 817-829.
69. Horn, S. J.; Sikorski, P.; Cederqvist, J. B.; Vaaje-Kolstad, G.; Sørlie, M.; Synstad, B.; Vriend, G.; Vårum, K. M.; Eijsink, V. G. H., Costs and Benefits of Processivity in Enzymatic Degradation of Recalcitrant Polysaccharides. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (48), 18089-18094.

70. Zakariassen, H.; Aam, B. B.; Horn, S. J.; Vårum, K. M.; Sørli, M.; Eijsink, V. G. H., Aromatic Residues in the Catalytic Center of Chitinase A from *Serratia marcescens* Affect Processivity, Enzyme Activity, and Biomass Converting Efficiency. *J. Biol. Chem.* **2009**, *284* (16), 10610-10617.
71. Kurasin, M.; Kuusk, S.; Kuusk, P.; Sørli, M.; Väljamäe, P., Slow Off-rates and Strong Product Binding Are Required for Processivity and Efficient Degradation of Recalcitrant Chitin by Family 18 Chitinases. *J. Biol. Chem.* **2015**, *290* (48), 29074-29085.
72. Kari, J.; Olsen, J.; Borch, K.; Cruys-Bagger, N.; Jensen, K.; Westh, P., Kinetics of Cellobiohydrolase (Cel7A) Variants with Lowered Substrate Affinity. *J. Biol. Chem.* **2014**, *289* (47).
73. Payne, C. M.; Baban, J.; Horn, S. J.; Backe, P. H.; Arvai, A. S.; Dalhus, B.; Bjørås, M.; Eijsink, V. G. H.; Sørli, M.; Beckham, G. T., et al., Hallmarks of Processivity in Glycoside Hydrolases from Crystallographic and Computational Studies of the *Serratia marcescens* Chitinases. *J. Biol. Chem.* **2012**, *287* (43), 36322-36330.
74. Horn, S. J.; Sørli, M.; Vårum, K. M.; Väljamäe, P.; Eijsink, V. G. H., Measuring Processivity. *Cellulases* **2012**, *510*, 69-95.
75. Levasseur, A.; Drula, E.; Lombard, V.; Coutinho, P. M.; Henrissat, B., Expansion of the Enzymatic Repertoire of the CAZy Database to Integrate Auxiliary Redox Enzymes. *Biotechnol. Biofuels* **2013**, *6*, 41.
76. Eijsink, V. G. H.; Vaaje-Kolstad, G.; Vårum, K. M.; Horn, S. J., Towards New Enzymes for Biofuels: Lessons from Chitinase Research. *Trends Biotechnol.* *26* (5), 228-235.
77. Horn, S. J.; Vaaje-Kolstad, G.; Westereng, B.; Eijsink, V. G. H., Novel Enzymes for the Degradation of Cellulose. *Biotechnol. Biofuels* **2012**, *5*, 45.
78. Hamre, A. G.; Lorentzen, S. B.; Väljamäe, P.; Sørli, M., Enzyme Processivity Changes with the Extent of Recalcitrant Polysaccharide Degradation. *FEBS Lett.* **2014**, *588* (24), 4620-4624.
79. van Aalten, D. M. F.; Komander, D.; Synstad, B.; Gaseidnes, S.; Peter, M. G.; Eijsink, V. G. H., Structural Insights into the Catalytic Mechanism of a Family 18 Exo-Chitinase. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (16), 8979-8984.
80. Suzuki, K.; Taiyoji, M.; Sugawara, N.; Nikaidou, N.; Henrissat, B.; Watanabe, T., The Third Chitinase Gene (ChiC) of *Serratia marcescens* 2170 and the Relationship of its Product to other Bacterial Chitinases. *Biochem. J.* **1999**, *343* (Pt 3), 587-596.
81. Horn, S. J.; Sørli, M.; Vaaje-Kolstad, G.; Norberg, A. L.; Synstad, B.; Vårum, K. M.; Eijsink, V. G. H., Comparative Studies of Chitinases A, B and C from *Serratia marcescens*. *Biocatal. Biotransform.* **2006**, *24* (1-2), 39-53.
82. Hult, E. L.; Katouno, F.; Uchiyama, T.; Watanabe, T.; Sugiyama, J., Molecular Directionality in Crystalline  $\beta$ -Chitin: Hydrolysis by Chitinases A and B from *Serratia marcescens* 2170. *Biochem. J.* **2005**, *388* (Pt 3), 851-856.

83. Igarashi, K.; Uchihashi, T.; Uchiyama, T.; Sugimoto, H.; Wada, M.; Suzuki, K.; Sakuda, S.; Ando, T.; Watanabe, T.; Samejima, M., Two-Way Traffic of Glycoside Hydrolase Family 18 Processive Chitinases on Crystalline Chitin. *Nat. Commun.* **2014**, *5*, 3975.
84. Sikorski, P.; Sørbotten, A.; Horn, S. J.; Eijsink, V. G. H.; Vårum, K. M., *Serratia marcescens* Chitinases with Tunnel-Shaped Substrate-Binding Grooves Show Endo Activity and Different Degrees of Processivity during Enzymatic Hydrolysis of Chitosan. *Biochemistry* **2006**, *45* (31), 9566-9574.
85. Schrödinger, L. *The PyMOL Molecular Graphics System*, 1.5.0.4; 2010.
86. Payne, C. M.; Bomble, Y.; Taylor, C. B.; McCabe, C.; Himmel, M. E.; Crowley, M. F.; Beckham, G. T., Multiple Functions of Aromatic-Carbohydrate Interactions in a Processive Cellulase Examined with Molecular Simulation. *J. Biol. Chem.* **2011**, *286* (47), 41028-41035.
87. Brameld, K. A.; Goddard, W. A., Substrate Distortion to a Boat Conformation at Subsite-1 is Critical in the Mechanism of Family 18 Chitinases. *J. Am. Chem. Soc.* **1998**, *120* (15), 3571-3580.
88. Tews, I.; vanScheltinga, A. C. T.; Perrakis, A.; Wilson, K. S.; Dijkstra, B. W., Substrate-Assisted Catalysis Unifies Two Families of Chitinolytic Enzymes. *J. Am. Chem. Soc.* **1997**, *119* (34), 7954-7959.
89. Jitonnorn, J.; Lee, V. S.; Nimmanpipug, P.; Rowlands, H. A.; Mulholland, A. J., Quantum Mechanics/Molecular Mechanics Modeling of Substrate-Assisted Catalysis in Family 18 Chitinases: Conformational Changes and the Role of Asp142 in Catalysis in ChiB. *Biochemistry* **2011**, *50* (21), 4697-4711.
90. Aronson, N. N.; Halloran, B. A.; Alexeyev, M. F.; Zhou, X. E.; Wang, Y. J.; Meehan, E. J.; Chen, L. Q., Mutation of a Conserved Tryptophan in the Chitin-Binding Cleft of *Serratia marcescens* Chitinase A Enhances Transglycosylation. *Biosci. Biotechnol. Biochem.* **2006**, *70* (1), 243-251.
91. Karplus, M.; McCammon, J. A., Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Mol. Biol.* **2002**, *9* (9), 646-652.
92. Dror, R. O.; Dirks, R. M.; Grossman, J. P.; Xu, H.; Shaw, D. E., Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annual Review of Biophysics* **2012**, *41* (1), 429-452.
93. Schlick, T.; Collepardo-Guevara, R.; Halvorsen, L. A.; Jung, S.; Xiao, X., Biomolecular modeling and Simulation: A Field Coming of Age. *Q. Rev. Biophys.* **2011**, *44* (2), 191-228.
94. Frenkel, D.; Smit, B., Understanding Molecular Simulations: From Algorithms to Applications. *Computational sciences series* **2002**, *1*, 1-638.
95. Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S., et al., CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30* (10), 1545-1614.

96. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781-1802.
97. Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N., et al., AMBER 2016. *University of California, San Francisco* **2016**.
98. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., GROMACS: A Message-passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91* (1), 43-56.
99. Lindahl, E.; Hess, B.; van der Spoel, D., GROMACS 3.0: A Package for Molecular Simulation and Trajectory Analysis. *Molecular modeling annual* **2001**, *7* (8), 306-317.
100. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C., GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26* (16), 1701-1718.
101. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4* (3), 435-447.
102. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High Performance Molecular Simulations Through Multi-level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1-2*, 19-25.
103. Chikvaidze, M. Understanding Structure, Dynamics and Function of *Escherichia coli* Alkaline Phosphate by Molecular Modelling and Simulations. Ruperto-Carola University of Heidelberg, Germany, 2013.
104. Deng, Y. Q.; Roux, B., Calculation of Standard Binding Free Energies: Aromatic Molecules in the T4 Lysozyme L99A Mutant. *J. Chem. Theory Comput.* **2006**, *2* (5), 1255-1273.
105. Jiang, W.; Hodoseck, M.; Roux, B., Computation of Absolute Hydration and Binding Free Energy with Free Energy Perturbation Distributed Replica-Exchange Molecular Dynamics. *J. Chem. Theory Comput.* **2009**, *5* (10), 2583-2588.
106. Jiang, W.; Roux, B., Free Energy Perturbation Hamiltonian Replica-Exchange Molecular Dynamics (FEP/H-REMD) for Absolute Ligand Binding Free Energy Calculations. *J. Chem. Theory Comput.* **2010**, *6* (9), 2559-2565.
107. Deng, Y. Q.; Roux, B., Hydration of Amino Acid Side Chains: Nonpolar and Electrostatic Contributions Calculated from Staged Molecular Dynamics Free Energy Simulations with Explicit Water Molecules. *J. Phys. Chem. B* **2004**, *108* (42), 16567-16576.
108. Shirts, M. R.; Chodera, J. D., Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129* (12), 124105.

109. Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F., Avoiding Singularities and Numerical Instabilities in Free Energy Calculations Based on Molecular Simulations. *Chem. Phys. Lett.* **1994**, *222* (6), 529-539.
110. Pohorille, A.; Jarzynski, C.; Chipot, C., Good Practices in Free-Energy Calculations. *J. Phys. Chem. B* **2010**, *114* (32), 10235-10253.
111. Steinbrecher, T.; Mobley, D. L.; Case, D. A., Nonlinear Scaling Schemes for Lennard-Jones Interactions in Free Energy Calculations. *J. Chem. Phys.* **2007**, *127* (21), 214108/1-214108/13.
112. Jana, S.; Hamre, A. G.; Wildberger, P.; Holen, M. M.; Eijsink, V. G.; Beckham, G. T.; Sørli, M.; Payne, C. M., Aromatic-Mediated Carbohydrate Recognition in Processive *Serratia marcescens* Chitinases. *J. Phys. Chem. B* **2016**, *120* (7), 1236-49.
113. Torrie, G. M.; Valleau, J. P., Nonphysical Sampling Distributions in Monte Carlo Free Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23* (2), 187-199.
114. Kästner, J., Umbrella Sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1* (6), 932-942.
115. Mills, M.; Andricioaei, I., An Experimentally Guided Umbrella Sampling Protocol for Biomolecules. *J. Chem. Phys.* **2008**, *129* (11), 114101.
116. Grossfield, A., WHAM: The Weighted Histogram Analysis Method.
117. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A., The Weighted Histogram Analysis Method for Free Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13* (8), 1011-1021.
118. Souaille, M.; Roux, B., Extension to the Weighted Histogram Analysis Method: Combining Umbrella Sampling with Free Energy Calculations. *Comput. Phys. Commun.* **2001**, *135* (1), 40-57.
119. Hamre, A. G.; Jana, S.; Holen, M. M.; Mathiesen, G.; Väljamäe, P.; Payne, C. M.; Sørli, M., Thermodynamic Relationships with Processivity in *Serratia marcescens* Family 18 Chitinases. *J. Phys. Chem. B* **2015**, *119* (30), 9601-9613.
120. Reese, E. T.; Siu, R. G. H.; Levinson, H. S., The Biological Degradation of Soluble Cellulose Derivatives and its Relationship to the Mechanism of Cellulose Hydrolysis. *J. Bacteriol.* **1950**, *59* (4), 485-497.
121. Wilson, D. B., Studies of *Thermobifida fusca* Plant Cell Wall Degrading Enzymes. *The Chemical Record* **2004**, *4* (2), 72-82.
122. Ståhlberg, J.; Johansson, G.; Pettersson, G., *Trichoderma reesei* has no True Exo-cellulase: All Intact and Truncated Cellulases Produce New Reducing End Groups on Cellulose. *Biochim. Biophys. Acta* **1993**, *1157* (1), 107-113.
123. Debono, M.; Gordee, R. S., Antibiotics that Inhibit Fungal Cell Wall Development. *Annu. Rev. Microbiol.* **1994**, *48* (1), 471-497.

124. Araujo, A. C.; Souto-Padrón, T.; de Souza, W., Cytochemical Localization of Carbohydrate Residues in Microfilariae of *Wuchereria bancrofti* and *Brugia malayi*. *J. Histochem. Cytochem.* **1993**, *41* (4), 571-578.
125. Fuhrman, J. A.; Piessens, W. F., Chitin Synthesis and Sheath Morphogenesis in *Brugia malayi* Microfilariae. *Mol. Biochem. Parasitol.* **1985**, *17* (1), 93-104.
126. Neville, A. C.; Parry, D. A.; Woodhead-Galloway, J., The Chitin Crystallite in Arthropod Cuticle. *J. Cell Sci.* **1976**, *21* (1), 73-82.
127. Shahabuddin, M.; Kaslow, D. C., Plasmodium: Parasite Chitinase and Its Role in Malaria Transmission. *Exp. Parasitol.* **1994**, *79* (1), 85-88.
128. Sawada, D.; Nishiyama, Y.; Langan, P.; Forsyth, V. T.; Kimura, S.; Wada, M., Direct Determination of the Hydrogen Bonding Arrangement in Anhydrous  $\beta$ -Chitin by Neutron Fiber Diffraction. *Biomacromolecules* **2012**, *13* (1), 288-291.
129. Armand, S.; Tomita, H.; Heyraud, A.; Gey, C.; Watanabe, T.; Henrissat, B., Stereochemical Course of the Hydrolysis Reaction Catalyzed by Chitinases A1 and D from *Bacillus circulans* WL-12. *FEBS Lett.* **1994**, *343* (2), 177-180.
130. Terwisscha van Scheltinga, A. C.; Armand, S.; Kalk, K. H.; Isogai, A.; Henrissat, B.; Dijkstra, B. W., Stereochemistry of Chitin Hydrolysis by a Plant Chitinase/Lysozyme and X-ray Structure of a Complex with Allosamidin Evidence for Substrate Assisted Catalysis. *Biochemistry* **1995**, *34* (48), 15619-15623.
131. Terwisscha van Scheltinga, A. C.; Hennig, M.; Dijkstra, B. W., The 1.8 Å Resolution Structure of Hevamine, a Plant Chitinase/Lysozyme, and Analysis of the Conserved Sequence and Structure Motifs of Glycosyl Hydrolase Family 18. *J. Mol. Biol.* **1996**, *262* (2), 243-257.
132. Nerinckx, W.; Desmet, T.; Claeysens, M., A Hydrophobic Platform as a Mechanistically Relevant Transition State Stabilising Factor Appears to be Present in the Active Centre of all Glycoside Hydrolases. *FEBS Lett.* **2003**, *538* (1-3), 1-7.
133. Perrakis, A.; Tews, I.; Dauter, Z.; Oppenheim, A. B.; Chet, I.; Wilson, K. S.; Vorgias, C. E., Crystal-Structure of a Bacterial Chitinase at 2.3 Å Resolution. *Structure* **1994**, *2* (12), 1169-1180.
134. van Aalten, D. M. F.; Synstad, B.; Brurberg, M. B.; Hough, E.; Riise, B. W.; Eijsink, V. G. H.; Wierenga, R. K., Structure of a Two-Domain Chitotriosidase From *Serratia marcescens* at 1.9 Å Resolution. *Proceedings of the National Academy of Sciences* **2000**, *97* (11), 5842-5847.
135. Gal, S. W.; Choi, J. Y.; Kim, C. Y.; Cheong, Y. H.; Choi, Y. J.; Lee, S. Y.; Bahk, J. D.; Cho, M. J., Cloning of the 52-kDa Chitinase Gene from *Serratia marcescens* KCTC2172 and its Proteolytic Cleavage into an Active 35-kDa Enzyme. *FEMS Microbiol. Lett.* **1998**, *160* (1), 151-158.
136. Synstad, B.; Vaaje-Kolstad, G.; Cederkvist, F. H.; Saua, S. F.; Horn, S. J.; Eijsink, V. G. H.; SØrlie, M., Expression and Characterization of Endochitinase C from *Serratia marcescens* BJL200 and Its Purification by a One-Step General Chitinase Purification Method. *Biosci. Biotechnol. Biochem.* **2008**, *72* (3), 715-723.



137. Dybvik, A. I.; Norberg, A. L.; Schute, V.; Soltwisch, J.; Peter-Katalinic, J.; Varum, K. M.; Eijsink, V. G. H.; Dreisewerd, K.; Mormann, M.; Sørlie, M., Analysis of Noncovalent Chitinase-Chito-Oligosaccharide Complexes by Infrared-Matrix Assisted Laser Desorption Ionization and Nano-electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2011**, *83* (11), 4030-4036.
138. Norberg, A. L.; Karlsen, V.; Hoell, I. A.; Bakke, I.; Eijsink, V. G. H.; Sørlie, M., Determination of Substrate Binding Energies in Individual Subsites of a Family 18 Chitinase. *FEBS Lett.* **2010**, *584* (22), 4581-4585.
139. Wiseman, T.; Williston, S.; Brandts, J. F.; Lin, L. N., Rapid Measurement of Binding Constants and Heats of Binding Using a New Titration Calorimeter. *Anal. Biochem.* **1989**, *179* (1), 131-137.
140. Fukada, H.; Takahashi, K., Enthalpy and Heat Capacity Changes for the Proton Dissociation of Various Buffer Components in 0.1 M Potassium Chloride. *Proteins-Structure Function and Genetics* **1998**, *33* (2), 159-166.
141. Norberg, A. L.; Eijsink, V. G. H.; Sørlie, M., Dissecting Factors that Contribute to Ligand-Binding Energetics for Family 18 Chitinases. *Thermochim. Acta* **2010**, *511* (1-2), 189-193.
142. Cederkvist, F. H.; Saua, S. F.; Karlsen, V.; Sakuda, S.; Eijsink, V. G. H.; Sørlie, M., Thermodynamic Analysis of Allosamidin Binding to a Family 18 Chitinase. *Biochemistry* **2007**, *46* (43), 12347-12354.
143. Zakariassen, H.; Sørlie, M., Heat Capacity Changes in Heme Protein-Ligand Interactions. *Thermochim. Acta* **2007**, *464* (1-2), 24-28.
144. Vaaje-Kolstad, G.; Houston, D. R.; Rao, F. V.; Peter, M. G.; Synstad, B.; van Aalten, D. M. F.; Eijsink, V. G. H., Structure of the D142N Mutant of the Family 18 Chitinase ChiB from *Serratia marcescens* and its Complex with Allosamidin. *Biochim. Biophys. Acta, Proteins Proteomics* **2004**, *1696* (1), 103-111.
145. Norberg, A. L.; Dybvik, A. I.; Zakariassen, H.; Mormann, M.; Peter-Katalinic, J.; Eijsink, V. G. H.; Sørlie, M., Substrate Positioning in Chitinase A, A Processive Chito-Biohydrolase from *Serratia marcescens*. *FEBS Lett.* **2011**, *585* (14), 2339-2344.
146. Bokma, E.; Rozeboom, H. J.; Sibbald, M.; Dijkstra, B. W.; Beintema, J. J., Expression and characterization of active site mutants of hevamine, a chitinase from the rubber tree *Hevea brasiliensis*. *Eur. J. Biochem.* **2002**, *269* (3), 893-901.
147. Hsieh, Y. C.; Wu, Y. J.; Chiang, T. Y.; Kuo, C. Y.; Shrestha, K. L.; Chao, C. F.; Huang, Y. C.; Chuankhayan, P.; Wu, W. G.; Li, Y. K., et al., Crystal structures of *Bacillus cereus* NCTU2 chitinase complexes with chito-oligomers reveal novel substrate binding for catalysis: a chitinase without chitin binding and insertion domains. *J. Biol. Chem.* **2010**, *285* (41), 31603-31615.
148. Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V., H<sup>++</sup>3.0: Automating pK Prediction and the Preparation of Biomolecular Structures for Atomistic Molecular Modeling and Simulations. *Nucleic Acids Res.* **2012**, *40* (W1), W537-W541.

149. Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A., H<sup>++</sup>: A Server for Estimating pK(a)s and Adding Missing Hydrogens to Macromolecules. *Nucleic Acids Res.* **2005**, *33*, W368-W371.
150. Myers, J.; Grothaus, G.; Narayanan, S.; Onufriev, A., A Simple Clustering Algorithm can be Accurate Enough for Use in Calculations of pKs in Macromolecules. *Proteins: Struct., Funct., Genet.* **2006**, *63* (4), 928-938.
151. Mackerell, A. D.; Feig, M.; Brooks, C. L., Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* **2004**, *25* (11), 1400-1415.
152. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S., et al., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3586-3616.
153. Jørgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926-935.
154. Durell, S. R.; Brooks, B. R.; Bennaïm, A., Solvent-Induced Forces between 2 Hydrophilic Groups. *J. Phys. Chem.* **1994**, *98* (8), 2198-2202.
155. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual Molecular Dynamics. *J. Mol. Graph. Model.* **1996**, *14* (1), 33-38.
156. Baker, B. M.; Murphy, K. P., Dissecting the Energetics of a Protein-Protein Interaction: The Binding of Ovomuroid Third Domain to Elastase1. *J. Mol. Biol.* **1997**, *268* (2), 557-569.
157. Baldwin, R. L., Temperature Dependence of the Hydrophobic Interaction in Protein Folding. *Proceedings of the National Academy of Sciences* **1986**, *83* (21), 8069-8072.
158. Murphy, K. P.; Privalov, P. L.; Gill, S. J., Common Features of Protein Unfolding and Dissolution of Hydrophobic Compounds. *Science* **1990**, *247* (4942), 559-561.
159. Kuusk, S.; Sørlië, M.; Våljamäe, P., The Predominant Molecular State of Bound Enzyme Determines the Strength and Type of Product Inhibition in the Hydrolysis of Recalcitrant Polysaccharides by Processive Enzymes. *J. Biol. Chem.* **2015**, *290* (18), 11678-11691.
160. Colussi, F.; Sorensen, T. H.; Alasepp, K.; Kari, J.; Cruys-Bagger, N.; Windahl, M. S.; Olsen, J. P.; Borch, K.; Westh, P., Probing Substrate Interactions in the Active tunnel of a Catalytically Deficient Cellobiohydrolase (Cel7). *J. Biol. Chem.* **2015**, *290* (4), 2444-54.
161. Mayes, H. B.; Broadbelt, L. J.; Beckham, G. T., How Sugars Pucker: Electronic Structure Calculations Map the Kinetic Landscape of Five Biologically Paramount Monosaccharides and Their Implications for Enzymatic Catalysis. *J. Am. Chem. Soc.* **2014**, *136* (3), 1008-1022.

162. Biarnes, X.; Ardevol, A.; Planas, A.; Rovira, C.; Laio, A.; Parrinello, M., The Conformational Free Energy Landscape of  $\beta$ -D-Glucopyranose. Implications for Substrate Preactivation in  $\beta$ -Glucoside Hydrolases. *J. Am. Chem. Soc.* **2007**, *129* (35), 10686-10693.
163. Ladbury, J. E., Just Add Water! The Effect of Water on the Specificity of Protein-Ligand Binding Sites and its Potential Application to Drug Design. *Chem. Biol.* **1996**, *3* (12), 973-980.
164. Li, Z.; Lazaridis, T., The Effect of Water Displacement on Binding Thermodynamics: Concanavalin A. *The Journal of Physical Chemistry B* **2005**, *109* (1), 662-670.
165. Sakuda, S.; Isogai, A.; Matsumoto, S.; Suzuki, A., Search for Microbial Insect Growth Regulators. II. Allosamidin, a Novel Insect Chitinase Inhibitor. *The Journal of Antibiotics* **1987**, *40* (3), 296-300.
166. Baban, J.; Fjeld, S.; Sakuda, S.; Eijsink, V. G. H.; Sørli, M., The Roles of Three *Serratia marcescens* Chitinases in Chitin Conversion Are Reflected in Different Thermodynamic Signatures of Allosamidin Binding. *The Journal of Physical Chemistry B* **2010**, *114* (18), 6144-6149.
167. Henning Cederkvist, F.; Parmer, M. P.; Vårum, K. M.; Eijsink, V. G. H.; Sørli, M., Inhibition of a Family 18 Chitinase by Chitoooligosaccharides. *Carbohydr. Polym.* **2008**, *74* (1), 41-49.
168. Stern, R.; Jedrzejewski, M. J., Carbohydrate Polymers at the Center of Life's Origins: The Importance of Molecular Processivity. *Chem. Rev.* **2008**, *108* (12), 5061-5085.
169. Divne, C.; Ståhlberg, J.; Reinikainen, T.; Ruohonen, L.; Pettersson, G.; Knowles, J. K. C.; Teeri, T. T.; Jones, T. A., The 3-Dimensional Crystal-Structure of the Catalytic Core of Cellobiohydrolase-I from *Trichoderma reesei*. *Science* **1994**, *265* (5171), 524-528.
170. Levine, S. E.; Fox, J. M.; Blanch, H. W.; Clark, D. S., A Mechanistic Model of the Enzymatic Hydrolysis of Cellulose. *Biotechnol. Bioeng.* **2010**, *107* (1), 37-51.
171. Zou, J. Y.; Kleywegt, G. J.; Ståhlberg, J.; Driguez, H.; Nerinckx, W.; Claeysens, M.; Koivula, A.; Teeri, T. T.; Jones, T. A., Crystallographic Evidence for Substrate Ring Distortion and Protein Conformational Changes During Catalysis in Cellobiohydrolase Cel6A from *Trichoderma reesei*. *Structure Fold. Des.* **1999**, *7* (9), 1035-1045.
172. Spezio, M.; Wilson, D. B.; Karplus, P. A., Crystal-Structure of the Catalytic Domain of a Thermophilic Endocellulase. *Biochemistry* **1993**, *32* (38), 9906-9916.
173. Davies, G. J.; Brzozowski, A. M.; Dauter, M.; Varrot, A.; Schulein, M., Structure and Function of *Humicola insolens* Family 6 Cellulases: Structure of the Endoglucanase, Cel6B, at 1.6 Å Resolution. *Biochem. J.* **2000**, *348*, 201-207.
174. Asensio, J. L.; Arda, A.; Canada, F. J.; Jimenez-Barbero, J., Carbohydrate-Aromatic Interactions. *Acc. Chem. Res.* **2013**, *46* (4), 946-954.
175. McGaughey, G. B.; Gagne, M.; Rappe, A. K.,  $\pi$ -Stacking Interactions - Alive and Well in Proteins. *J. Biol. Chem.* **1998**, *273* (25), 15458-15463.

176. Santana, A. G.; Jimenez-Moreno, E.; Gomez, A. M.; Corzana, F.; Gonzalez, C.; Jimenez-Oses, G.; Jimenez-Barbero, J.; Asensio, J. L., A Dynamic Combinatorial Approach for the Analysis of Weak Carbohydrate/Aromatic Complexes: Dissecting Facial Selectivity in CH/ $\pi$  Stacking Interactions. *J. Am. Chem. Soc.* **2013**, *135* (9), 3347-3350.
177. Ramirez-Gualito, K.; Alonso-Rios, R.; Quiroz-Garcia, B.; Rojas-Aguilar, A.; Diaz, D.; Jimenez-Barbero, J.; Cuevas, G., Enthalpic Nature of the CH/ $\pi$  Interaction Involved in the Recognition of Carbohydrates by Aromatic Compounds, Confirmed by a Novel Interplay of NMR, Calorimetry, and Theoretical Calculations. *J. Am. Chem. Soc.* **2009**, *131* (50), 18129-18138.
178. Pace, C. J.; Kim, D.; Gao, J. M., Experimental Evaluation of CH- $\pi$  Interactions in a Protein Core. *Chemistry-a European Journal* **2012**, *18* (19), 5832-5836.
179. Wimmerova, M.; Kozmon, S.; Necasova, I.; Mishra, S. K.; Komarek, J.; Koca, J., Stacking Interactions between Carbohydrate and Protein Quantified by Combination of Theoretical and Experimental Methods. *PLoS One* **2012**, *7* (10), e46032.
180. Kiehna, S. E.; Laughrey, Z. R.; Waters, M. L., Evaluation of a Carbohydrate- $\pi$  Interaction in a Peptide Model System. *Chem. Commun.* **2007**, (39), 4026-4028.
181. Uchiyama, T.; Katouno, F.; Nikaidou, N.; Nonaka, T.; Sugiyama, J.; Watanabe, T., Roles of the Exposed Aromatic Residues in Crystalline Chitin Hydrolysis by Chitinase A from *Serratia marcescens* 2170. *J. Biol. Chem.* **2001**, *276* (44), 41343-41349.
182. Zakariassen, H.; Eijsink, V. G. H.; Sørlie, M., Signatures of Activation Parameters Reveal Substrate-Dependent Rate Determining Steps in Polysaccharide Turnover by a Family 18 Chitinase. *Carbohydr. Polym.* **2010**, *81* (1), 14-20.
183. Taylor, C. B.; Payne, C. M.; Himmel, M. E.; Crowley, M. F.; McCabe, C.; Beckham, G. T., Binding Site Dynamics and Aromatic-Carbohydrate Interactions in Processive and Non-Processive Family 7 Glycoside Hydrolases. *J. Phys. Chem. B* **2013**, *117* (17), 4924-4933.
184. Nakamura, A.; Tsukada, T.; Auer, S.; Furuta, T.; Wada, M.; Koivula, A.; Igarashi, K.; Samejima, M., The Tryptophan Residue at the Active Site Tunnel Entrance of *Trichoderma reesei* Cellobiohydrolase Cel7A Is Important for Initiation of Degradation of Crystalline Cellulose. *J. Biol. Chem.* **2013**, *288* (19), 13503-13510.
185. Brurberg, M. B.; Eijsink, V. G. H.; Nes, I. F., Characterization of a Chitinase Gene (ChiA) from *Serratia marcescens* B JL200 and One Step Purification of the Gene Product. *FEMS Microbiol. Lett.* **1994**, *124* (3), 399-404.
186. Brurberg, M. B.; Eijsink, V. G. H.; Haandrikman, A. J.; Venema, G.; Nes, I. F., Chitinase B from *Serratia marcescens* B JL200 Is Exported to the Periplasm without Processing. *Microbiology* **1995**, *141*, 123-131.
187. Wang, W. Y.; Malcolm, B. A., Two-Stage PCR Protocol Allowing Introduction of Multiple Mutations, Deletions and Insertions Using QuikChange (TM) Site-Directed Mutagenesis. *BioTechniques* **1999**, *26* (4), 680-682.

188. Synstad, B.; Gaseidnes, S.; van Aalten, D. M. F.; Vriend, G.; Nielsen, J. E.; Eijsink, V. G. H., Mutational and Computational Analysis of the Role of Conserved Residues in the Active Site of a Family 18 Chitinase. *Eur. J. Biochem.* **2004**, *271* (2), 253-262.
189. Brurberg, M. B.; Nes, I. F.; Eijsink, V. G. H., Comparative Studies of Chitinases A and B from *Serratia marcescens*. *Microbiology* **1996**, *142*, 1581-1589.
190. Hamre, A. G.; Jana, S.; Reppert, N. K.; Payne, C. M.; Sørli, M., Processivity, Substrate Positioning, and Binding: The Role of Polar Residues in a Family 18 Glycoside Hydrolase. *Biochemistry* **2015**, *54* (49), 7292-7306.
191. Manoil, C.; Beckwith, J., A Genetic Approach to Analyzing Membrane Protein Topology. *Science* **1986**, *233* (4771), 1403-1408.
192. Turnbull, W. B.; Daranas, A. H., On the Value of C: Can Low Affinity Systems be Studied by Isothermal Titration Calorimetry? *J. Am. Chem. Soc.* **2003**, *125* (48), 14859-14866.
193. Aronson, N. N.; Halloran, B. A.; Alexyev, M. F.; Amable, L.; Madura, J. D.; Pasupulati, L.; Worth, C.; Van Roey, P., Family 18 Chitinase-Oligosaccharide Substrate Interaction: Subsite Preference and Anomer Selectivity of *Serratia marcescens* Chitinase A. *Biochem. J.* **2003**, *376*, 87-95.
194. Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; Mackerell, A. D., Additive Empirical Force Field for Hexopyranose Monosaccharides. *J. Comput. Chem.* **2008**, *29* (15), 2543-2564.
195. Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D., CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *J. Chem. Theory Comput.* **2009**, *5* (9), 2353-2370.
196. Kirkwood, J. G., Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3* (5), 300-313.
197. Straatsma, T. P.; Mccammon, J. A., Multiconfiguration Thermodynamic Integration. *J. Chem. Phys.* **1991**, *95* (2), 1175-1188.
198. Jitonnorn, J.; Sattayanon, C.; Kungwan, N.; Hannongbua, S., A DFT Study of the Unusual Substrate-Assisted Mechanism of *Serratia marcescens* Chitinase B Reveals the Role of Solvent and Mutational Effect on Catalysis. *J. Mol. Graph. Model.* **2015**, *56*, 53-59.
199. Jitonnorn, J.; Limb, M. A. L.; Mulholland, A. J., QM/MM Free-Energy Simulations of Reaction in *Serratia marcescens* Chitinase B Reveal the Protonation State of Asp142 and the Critical Role of Tyr214. *J. Phys. Chem. B* **2014**, *118* (18), 4771-4783.
200. Zakariassen, H.; Hansen, M. C.; Joranli, M.; Eijsink, V. G. H.; Sørli, M., Mutational Effects on Transglycosylating Activity of Family 18 Chitinases and Construction of a Hypertransglycosylating Mutant. *Biochemistry* **2011**, *50* (25), 5693-5703.
201. Katouno, F.; Taguchi, M.; Sakurai, K.; Uchiyama, T.; Nikaidou, N.; Nonaka, T.; Sugiyama, J.; Watanabe, T., Importance of Exposed Aromatic Residues in Chitinase B

- from *Serratia marcescens* 2170 for Crystalline Chitin Hydrolysis. *J. Biochem. (Tokyo)* **2004**, *136* (2), 163-168.
202. Koivula, A.; Kinnari, T.; Harjunpaa, V.; Ruohonen, L.; Teleman, A.; Drakenberg, T.; Rouvinen, J.; Jones, T. A.; Teeri, T. T., Tryptophan 272: An Essential Determinant of Crystalline Cellulose Degradation by *Trichoderma reesei* Cellobiohydrolase Cel6A. *FEBS Lett.* **1998**, *429* (3), 341-346.
203. Vuong, T. V.; Wilson, D. B., Processivity, Synergism, and Substrate Specificity of *Thermobifida fusca* Cel6B. *Appl. Environ. Microbiol.* **2009**, *75* (21), 6655-6661.
204. Nakamura, A.; Watanabe, H.; Ishida, T.; Uchihashi, T.; Wada, M.; Ando, T.; Igarashi, K.; Samejima, M., Trade-off between Processivity and Hydrolytic Velocity of Cellobiohydrolases at the Surface of Crystalline Cellulose. *J. Am. Chem. Soc.* **2014**, *136* (12), 4584-4592.
205. Kim, J.; Yun, S.; Ounaies, Z., Discovery of Cellulose as a Smart Material. *Macromolecules* **2006**, *39* (12), 4202-4206.
206. Sinnott, M. L., Catalytic Mechanism of Enzymic Glycosyl Transfer. *Chem. Rev.* **1990**, *90* (7), 1171-1202.
207. Teeri, T. T., Crystalline Cellulose Degradation: New Insight into the Function of Cellobiohydrolases. *Trends Biotechnol.* **1997**, *15* (5), 160-167.
208. Zolotnitsky, G.; Cogan, U.; Adir, N.; Solomon, V.; Shoham, G.; Shoham, Y., Mapping Glycoside Hydrolase Substrate Subsites by Isothermal Titration Calorimetry. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (31), 11275-11280.
209. Vyas, N. K., Atomic Features of Protein-Carbohydrate Interactions. *Curr. Opin. Struct. Biol.* **1991**, *1* (5), 732-740.
210. Varrot, A.; Frandsen, T. P.; von Ossowski, I.; Boyer, V.; Cottaz, S.; Driguez, H.; Schulein, M.; Davies, G. J., Structural Basis for Ligand Binding and Processivity in Cellobiohydrolase Cel6A from *Humicola insolens*. *Structure* **2003**, *11* (7), 855-864.
211. Watanabe, T.; Ariga, Y.; Sato, U.; Toratani, T.; Hashimoto, M.; Nikaidou, N.; Kezuka, Y.; Nonaka, T.; Sugiyama, J., Aromatic residues within the substrate-binding cleft of *Bacillus circulans* chitinase A1 are essential for hydrolysis of crystalline chitin. *Biochem. J.* **2003**, *376*, 237-244.
212. Zhou, W.; Irwin, D. C.; Escovar-Kousen, J.; Wilson, D. B., Kinetic Studies of *Thermobifida fusca* Cel9A Active Site Mutant Enzymes. *Biochemistry* **2004**, *43* (30), 9655-9663.
213. Proctor, M. R.; Taylor, E. J.; Nurizzo, D.; Turkenburg, J. P.; Lloyd, R. M.; Vardakou, M.; Davies, G. J.; Gilbert, H. J., Tailored Catalysts for Plant Cell-wall Degradation: Redesigning the Exo/Endo Preference of *Cellvibrio japonicus* Arabinanase 43A. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (8), 2697-2702.
214. Bu, L. T.; Beckham, G. T.; Shirts, M. R.; Nimlos, M. R.; Adney, W. S.; Himmel, M. E.; Crowley, M. F., Probing Carbohydrate Product Expulsion from a Processive Cellulase with Multiple Absolute Binding Free Energy Methods. *J. Biol. Chem.* **2011**, *286* (20), 18161-18169.

215. Meyer, J. E. W.; Schulz, G. E., Energy Profile of Maltooligosaccharide Permeation Through Maltoporin as Derived From the Structure and From a Statistical Analysis of Saccharide-Protein Interactions. *Protein Sci.* **1997**, *6* (5), 1084-1091.
216. Fuchs, R. L.; McPherson, S. A.; Drahos, D. J., Cloning of a *Serratia marcescens* Gene Encoding Chitinase. *Appl. Environ. Microbiol.* **1986**, *51* (3), 504-509.
217. Monreal, J.; Reese, E. T., The Chitinase of *Serratia marcescens*. *Can. J. Microbiol.* **1969**, *15* (7), 689-696.
218. Sakuda, S.; Isogai, A.; Matsumoto, S.; Suzuki, A.; Koseki, K., The Structure of Allosamidin, A Novel Insect Chitinase Inhibitor, Produced by *Streptomyces Sp. Tetrahedron Lett.* **1986**, *27* (22), 2475-2478.
219. Sørbotten, A.; Horn, S. J.; Eijsink, V. G.; Vårum, K. M., Degradation of Chitosans with Chitinase B from *Serratia marcescens*. *FEBS J.* **2005**, *272* (2), 538-549.
220. Bahrke, S.; Einarsson, J. M.; Gislason, J.; Haebel, S.; Letzel, M. C.; Peter-Katalinić, J.; Peter, M. G., Sequence Analysis of Chitooligosaccharides by Matrix-Assisted Laser Desorption Ionization Postsource Decay Mass Spectrometry. *Biomacromolecules* **2002**, *3* (4), 696-704.
221. Krokeide, I. M.; Synstad, B.; Gaseidnes, S.; Horn, S. J.; Eijsink, V. G.; Sørli, M., Natural Substrate Assay for Chitinases Using High Performance Liquid Chromatography: A Comparison With Existing Assays. *Anal. Biochem.* **2007**, *363* (1), 128-134.
222. Hekmat, O.; Lo Leggio, L.; Rosengren, A.; Kamarauskaite, J.; Kolenova, K.; Stålbrand, H., Rational Engineering of Mannosyl Binding in the Distal Glycone Subsites of *Cellulomonas fimi* Endo- $\beta$ -1,4-mannanase: Mannosyl Binding Promoted at Subsite -2 and Demoted at Subsite -3. *Biochemistry* **2010**, *49* (23), 4884-4896.
223. Eide, K. B.; Lindbom, A. R.; Eijsink, V. G. H.; Norberg, A. L.; Sørli, M., Analysis of Productive Binding Modes in the Human Chitotriosidase. *FEBS Lett.* **2013**, *587* (21), 3508-3513.
224. Hoover, W. G., Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31* (3), 1695-1697.
225. Nosé, S., and Klein, M. L., Constant Pressure Molecular Dynamics for Molecular Systems. *Mol. Phys.* **1983**, *50* (5), 1055-1076.
226. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577-8593.
227. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical Integration of Cartesian Equations of Motion of a System with Constraints - Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327-341.
228. Davies, G. J.; Wilson, K. S.; Henrissat, B., Nomenclature for Sugar-Binding Subsites in Glycosyl Hydrolases. *Biochem. J.* **1997**, *321* (Pt 2), 557-559.
229. Hackman, R. H., Studies on Chitin. I. Enzymic Degradation of Chitin and Chitin Esters. *Aust. J. Biol. Sci.* **1954**, *7* (2), 168-178.

230. Sannan, T.; Kurita, K.; Iwakura, Y., Studies on Chitin, 2. Effect of Deacetylation on Solubility. *Makromol. Chem.* **1976**, *177* (12), 3589-3600.
231. Vårum, K. M.; Anthonsen, M. W.; Grasdalen, H.; Smidsrød, O., <sup>13</sup>C-N.M.R. Studies of the Acetylation Sequences in Partially N-deacetylated Chitins (Chitosans). *Carbohydr. Res.* **1991**, *217*, 19-27.
232. Vårum, K. M.; Anthonsen, M. W.; Grasdalen, H.; Smidsrød, O., Determination of the Degree of N-acetylation and the Distribution of N-acetyl Groups in Partially N-deacetylated Chitins (Chitosans) by High-Field N.M.R. Spectroscopy. *Carbohydr. Res.* **1991**, *211* (1), 17-23.
233. Igarashi, K.; Uchihashi, T.; Koivula, A.; Wada, M.; Kimura, S.; Okamoto, T.; Penttila, M.; Ando, T.; Samejima, M., Traffic Jams Reduce Hydrolytic Efficiency of Cellulase on Cellulose Surface. *Science* **2011**, *333* (6047), 1279-1282.
234. Kostylev, M.; Alahuhta, M.; Chen, M.; Brunecky, R.; Himmel, M. E.; Lunin, V. V.; Brady, J.; Wilson, D. B., Cel48A from *Thermobifida fusca*: Structure and Site Directed Mutagenesis of Key Residues. *Biotechnol. Bioeng.* **2014**, *111* (4), 664-673.
235. Blake, C. C.; Johnson, L. N.; Mair, G. A.; North, A. C.; Phillips, D. C.; Sarma, V. R., Crystallographic Studies of the Activity of Hen Egg-White Lysozyme. *Proc R Soc Lond B Biol Sci.* **1967**, *167* (1009), 377-388.
236. Phillips, D. C., The Three-Dimensional Structure of an Enzyme Molecule. *Sci. Am.* **1966**, *215* (5), 78-90.
237. Cremer, D.; Pople, J. A., General Definition of Ring Puckering Coordinates. *J. Am. Chem. Soc.* **1975**, *97* (6), 1354-1358.
238. Hill, A. D.; Reilly, P. J., Puckering Coordinates of Monocyclic Rings by Triangular Decomposition. *J. Chem. Inf. Model.* **2007**, *47* (3), 1031-1035.
239. Cooper, A.; Johnson, C. M.; Lakey, J. H.; Nöllmann, M., Heat Does Not Come in Different Colours: Entropy–Enthalpy Compensation, Free Energy Windows, Quantum Confinement, Pressure Perturbation Calorimetry, Solvation and the Multiple Causes of Heat Capacity Effects in Biomolecular Interactions. *Biophys. Chem.* **2001**, *93* (2–3), 215-230.
240. Kumar, M. N. V. R., A Review of Chitin and Chitosan Applications. *React. Funct. Polym.* **2000**, *46* (1), 1-27.
241. Hudson, K. L.; Bartlett, G. J.; Diehl, R. C.; Agirre, J.; Gallagher, T.; Kiessling, L. L.; Woofson, D. N., Carbohydrate-Aromatic Interactions in Proteins. *J. Am. Chem. Soc.* **2015**, *137* (48), 15152-15160.
242. Chen, W. T.; Enck, S.; Price, J. L.; Powers, D. L.; Powers, E. T.; Wong, C. H.; Dyson, H. J.; Kelly, J. W., Structural and Energetic Basis of Carbohydrate-Aromatic Packing Interactions in Proteins. *J. Am. Chem. Soc.* **2013**, *135* (26), 9877-9884.
243. Hamre, A. G.; Eide, K. B.; Wold, H. H.; Sørli, M., Activation of Enzymatic Chitin Degradation By a Lytic Polysaccharide Monooxygenase. *Carbohydr. Res.* **2015**, *407*, 166-9.



244. Fan, Y.; Saito, T.; Isogai, A., Preparation of Chitin Nanofibers From Squid Pen  $\beta$ -Chitin by Simple Mechanical Treatment Under Acid Conditions. *Biomacromolecules* **2008**, *9* (7), 1919-23.
245. Kostylev, M.; Wilson, D., Two-Parameter Kinetic Model Based on a Time Dependent Activity Coefficient Accurately Describes Enzymatic Cellulose Digestion. *Biochemistry* **2013**, *52* (33), 5656-64.
246. Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R., Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J Cheminformatics* **2012**, *4*.
247. Hunenberger, P. H.; McCammon, J. A., Ewald artifacts in computer simulations of ionic solvation and ion-ion interaction: A continuum electrostatics study. *J. Chem. Phys.* **1999**, *110* (4), 1856-1872.
248. Figueirido, F.; Delbuono, G. S.; Levy, R. M., On Finite-Size Effects in Computer-Simulations Using the Ewald Potential. *J. Chem. Phys.* **1995**, *103* (14), 6133-6142.
249. Morgan, B. R.; Massi, F., Accurate Estimates of Free Energy Changes in Charge Mutations. *J. Chem. Theory Comput.* **2010**, *6* (6), 1884-1893.
250. Filippini, G.; Bonal, C.; Malfreyt, P., Methodological approaches for the free energy calculations in electroactive SAMs. *Mol. Phys.* **2012**, *110* (11-12), 1081-1095.
251. Zhou, R. H.; Das, P.; Royyuru, A. K., Single Mutation Induced H3N2 Hemagglutinin Antibody Neutralization: A Free Energy Perturbation Study. *J. Phys. Chem. B* **2008**, *112* (49), 15813-15820.
252. Donnini, S.; Mark, A. E.; Juffer, A. H.; Villa, A., Incorporating the Effect of Ionic Strength in Free Energy Calculations Using Explicit Ions. *J. Comput. Chem.* **2005**, *26* (2), 115-122.
253. Das, P.; Li, J. Y.; Royyuru, A. K.; Zhou, R. H., Free Energy Simulations Reveal a Double Mutant Avian H5N1 Virus Hemagglutinin with Altered Receptor Binding Specificity. *J. Comput. Chem.* **2009**, *30* (11), 1654-1663.
254. Nina, M.; Beglov, D.; Roux, B., Atomic Radii for Continuum Electrostatics Calculations Based on Molecular Dynamics Free Energy Simulations. *J. Phys. Chem. B* **1997**, *101* (26), 5239-5248.
255. Varrot, A.; Schulein, M.; Davies, G. J., Insights into Ligand-Induced Conformational Change in Cel5A from *Bacillus agaradhaerens* Revealed by a Catalytically Active Crystal Form. *J. Mol. Biol.* **2000**, *297* (3), 819-828.
256. Songsiriritthigul, C.; Pantoom, S.; Aguda, A. H.; Robinson, R. C.; Suginta, W., Crystal Structures of *Vibrio harveyi* Chitinase A Complexed with Chitooligosaccharides: Implications for the Catalytic Mechanism. *J. Struct. Biol.* **2008**, *162* (3), 491-499.
257. Carpita, N. C., Structure and Biogenesis of the Cell Walls of Grasses. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1996**, *47*, 445-476.
258. Delmer, D. P., Cellulose Biosynthesis: Exciting Times for A Difficult Field of Study. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1999**, *50*, 245-276.

259. Cosgrove, D. J., Growth of the Plant Cell Wall. *Nat. Rev. Mol. Cell Biol.* **2005**, *6* (11), 850-61.
260. Schlitter, J.; Engels, M.; Kruger, P., Targeted Molecular Dynamics - A New Approach for Searching Pathways of Conformational Transitions. *J. Mol. Graph.* **1994**, *12* (2), 84-89.
261. Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; van de Streek, J.; Wood, P. A., Mercury CSD 2.0 - New Features for the Visualization and Investigation of Crystal Structures. *J. Appl. Crystallogr.* **2008**, *41* (2), 466-470.
262. Sheinerman, F. B.; Brooks, C. L., Calculations on Folding of Segment B1 of Streptococcal Protein G. *J. Mol. Biol.* **1998**, *278* (2), 439-456.
263. Taylor, C. B.; Talib, M. F.; McCabe, C.; Bu, L.; Adney, W. S.; Himmel, M. E.; Crowley, M. F.; Beckham, G. T., Computational Investigation of Glycosylation Effects on a Family 1 Carbohydrate Binding Module. *J. Biol. Chem.* **2012**, *287* (5), 3147-55.
264. Grest, G. S.; Kremer, K., Molecular Dynamics Simulation for Polymers in the Presence of a Heat Bath. *Phys. Rev. A* **1986**, *33* (5), 3628-3631.
265. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W. Z.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J., et al., Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539.
266. Goujon, M.; McWilliam, H.; Li, W. Z.; Valentin, F.; Squizzato, S.; Paern, J.; Lopez, R., A New Bioinformatics Analysis Tools Framework at EMBL-EBI. *Nucleic Acids Res.* **2010**, *38*, W695-W699.
267. McWilliam, H.; Li, W. Z.; Uludag, M.; Squizzato, S.; Park, Y. M.; Buso, N.; Cowley, A. P.; Lopez, R., Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* **2013**, *41* (W1), W597-W600.
268. Robert, X.; Gouet, P., Deciphering Key Features in Protein Structures with the New ENDscript Server. *Nucleic Acids Res.* **2014**, *42* (W1), W320-W324.

## Vita

### Personal Information

Place of Birth Paschim Medinipur, West Bengal, India

### EDUCATION

2012 - present Ph.D., Chemical Engineering, University of Kentucky, United States of America

2010 - 2012 M.Tech., Chemical Engineering, Indian Institute of Technology Guwahati, India

2006 - 2010 B.E., Chemical Engineering, Jadavpur University, India

### Honors & Awards

2016 Outstanding Graduate Student Research Award in Chemical Engineering by University of Kentucky

2012 - present Graduate School Academic Year Fellowship for Doctoral Research from University of Kentucky

2014 - 2015 August T. Larsson Guest Researcher Tag Along Fellowship from Swedish University of Agricultural Sciences

2010 - 2012 Ministry of Human Resources and Development (MHRD) Fellowship for Masters Research from Government of India

## Publications

‡ Equally contributing

**S. Jana**‡, A. G. Hamre‡, V.G.H. Eijsink, M. Sørli, and C. M. Payne, “The Effect of Carbohydrate-Polar Interactions on Substrate Binding and Processivity in a Family 18 *Serratia marcescens* Chitinase”. (In preparation)

**S. Jana**, B.C. Knott, G.T. Beckham, and C. M. Payne, “The Relationship Between Chitinase Binding Affinity, Processivity, and Work Required to Decrystallize  $\beta$ -Chitin”. (In preparation)

F. Goedegebuur, L. Dankmeyer, P. Gualfetti, S. Karkehabadi, H. Hansson, **S. Jana**, V. Huynh, B. R. Kelemen, P. Kruithof, E. A. Larenas, P.J.M. Teunissen, J. Ståhlberg, C.M. Payne, C. Mitchinson, and M. Sandgren, “Improving the Thermal Stability of Cellobiohydrolase Cel7A from *Trichoderma reesei* by site directed-evolution”. (In preparation)

**S. Jana**‡, A. G. Hamre‡, P. Wildberger, M. M. Holen, V.G.H. Eijsink, G.T. Beckham, M. Sørli, and C. M. Payne, “Aromatic-mediated Carbohydrate Recognition in Processive *Serratia marcescens* Chitinases,” *J. Phys. Chem. B*, 2016, 120 (7), 1236–1249.

A. G. Hamre‡, **S. Jana**‡, N. K. Reppert, C. M. Payne, and M. Sørli, “Processivity, Substrate Positioning and Binding; the Role of Polar Residues in a Family 18 Glycoside Hydrolase”, *Biochemistry*, 2015, 54 (49), 7292-7306.

A. G. Hamre‡, **S. Jana**‡, M. M. Holen, G. Mathiesen, P. Väljamäe, C. M. Payne, and M. Sørli, "Thermodynamic Relationships with Processivity in *Serratia marcescens* Family 18 Chitinases," *J. Phys. Chem. B*, 2015, 119 (30), 9601–9613.

A. S. Borisova, E. Eneyskaya, K. Bobrov, S. Jana, A. Logachev, D. Poley, A. Lapidus, F. Ibatullin, U. Saleem, M. Sandgren, C. M. Payne, A. Kulminskaya, and J. Ståhlberg, "Sequencing, Biochemical Characterization, Crystal Structure and Molecular Dynamics of Cellobiohydrolase Cel7A from *Geotrichum candidum* 3C," *FEBS J.*, 2105, 282, 4515-4537.