MOLECULAR DYNAMICS SIMULATION OF THE ENDOPOLYGALACTURONASE II—OCTAGALACTURONATE COMPLEX

by

JARROD WESLEY BARNES

(Under the Direction of Robert J. Woods)

ABSTRACT

The *endo*polygalacturonase II enzyme from *A. niger* was studied with an octameric fragment of α-1,4-D-polygalacturonic acid. We present an approach to modeling the complex using MD simulations and offer a structural interpretation for the experimental data given by amide exchange mass spectrometry and site-directed mutagenesis studies. The detailed model predicts a conformational change in the linear form of the bound substrate to an activated bent form, which is supported by the experimental data given. The bent conformation of the substrate partially transverses the primary binding cleft and occupies a newly proposed secondary binding region, thus giving insight into the cleavage pathway for the *endo*polygalacturonase II enzyme.

INDEX WORDS: Molecular Dynamics, Solvent accessible surface area, Amide exchange, Endopolygalacturonase II, MM-GBSA, MM-PBSA, Site-directed mutagenesis and Alanine scanning

MOLECULAR DYNAMICS SIMULATION OF THE ENDOPOLYGALACTURONASE II—OCTAGALACTURONATE COMPLEX

by

JARROD WESLEY BARNES

B.S., The University of Georgia, 2001

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2004

MOLECULAR DYNAMICS SIMULATIONS OF THE ENDOPOLYGALACTURONASE II—OCTAGALACTURONATE COMPLEX

by

JARROD WESLEY BARNES

Major Professor: Robert J. Woods

Committee: Ron Orlando
Carl Bergmann

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2004

ACKNOWLEDGEMENTS

First of all, I would like to thank my family for all the support through this whole experience. I love and owe all my persistence and determination to you… thanks so much. Many thanks to all the people in the lab who have guided me and helped me along the way, especially, Ahamed Pathiaseril, Michael Ford, Sarah Wittkopp, and Maral Basma who were there with me from the very beginning. A BIG thanks to Renu Kadirvelraj and Jorge Gonzalez you both will always have a special place in my heart. I would like to thank my advisory committee Ron Orlando and Carl Bergmann who have supported me and guided me along the way. Thank you so much for all you have done. I want to thank Rob Woods for the guidance and the support without you this would not have been possible. Last but not least, I would like to thank my beautiful wife, Jennifer, who has always had faith in me and given me the support through it all.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS iv

CHAPTER

1 SUMMARY 1

2 LITERATURE REVIEW 3

3 THE ENDOPOLYGALACTURONASE II—OCTAGALACTURONATE COMPLEX: MD SIMULATIONS PREDICT BINDING INTERACTIONS AND GIVE INSIGHT INTO A NOVEL CLEAVAGE PATHWAY 24

ABSTRACT 25

INTRODUCTION 27

EXPERIMENTAL METHODS 31

RESULTS AND DISCUSSION 35

4 CONCLUSION 53

REFERENCES 56

CHAPTER 1

SUMMARY

The plant cell wall is a major barrier against attempted invasions by phytopathogenic fungi, therefore the plant cell wall-degrading enzymes produced by fungi play an important role in their pathogenicity. Many fungi use *endo*polygalacturonases (EPGs) to hydrolyze the cell wall polysaccharide homogalacturonan (HG) as one of the first steps in invasion. The ability to fragment (HG) has led to the use of EPGs in a variety of commercial applications such as fruit juice clarification, texture modification of purees, and release of color in juices and wines. Pectin methylesterases (PMEs) and rhamnogalacturonan acetylesterases (RGAEs) often act in concert with EPGs by converting methyl- and acetyl-esterified regions of homogalacturonan to a substrate which is susceptible to the EPGs. As such, fungal enzymes are not only important in pathogenicity, but figure prominently in industrial applications.

The main part of this thesis focuses on the interactions between *endo*polygalacturonase II (EPG-II) from *Aspergillus Niger* and an octameric fragment of its substrate, 1,4-α-D-polygalacturonic acid $(GalpA)_8$. In the last few years, studies have been done on the binding site of EPG to its homogalacturonan substrate using amide exchange-mass spectrometry and site-directed mutagenesis. Here, we present an approach to modeling the complex and offer a structural interpretation for the experimental data obtained using these two techniques. Molecular Dynamics (MD) simulations were used to study the binding interactions of the EPG-II─$(GalpA)_8$ complex. The results provide a detailed dynamic model for combining amide

exchange mass spectrometry data with a simple model for exchange based on Solvent Accessible Surface Area calculations (SASA). The model also predicts sites of molecular shielding within the enzyme-substrate complex, while identifying residues in the enzyme that play a role in binding the substrate. The conformational change of the linear form of the substrate via a chair-to-distorted boat (^{1,4}B) transition was defined by the conversion of the ground/dynamic state to an activated bent form. The dynamic flexibility of the substrate is supported by the amide exchange MS data and suggests the presence of a previously unidentified secondary binding cleft, thus giving insight into the cleavage pathway of the EPG-II enzyme. The proposed model for the enzyme-substrate complex was further defined by computational alanine scanning studies, which identifies residues that are important for stabilization of the substrate in the initial bound form as well as in the transition conformation by point-mutational screening. The results from the alanine scanning experiments were in good agreement with site-directed mutagenesis studies, while advances were made into depicting other important residues for further investigation.

CHAPTER 2

LITERATURE REVIEW

Pectin is among the most complex carbohydrate structures found in the plant cell wall and is one of three fundamental groups that make up 90 % of the wall (de Vries and Visser 2001). The others, cellulose and hemicellulose (Albersheim, Darvill et al. 1996), are a part of the *cellulose network* that comprises of β-1,4-linked D-glucan residues (cellulose) and xyloglucans/arabinoxylans. The latter are two hemicellulosic polymers that coat the cellulose microfibrils to prevent excessive aggregation (Herron, Benen et al. 2000). Pectin is a part of the *pectate network* that consists of smooth and hairy regions (Carpita and Gibeaut 1993). The smooth regions are made up of polymers, also known as, [homogalacturonan (HG)] α-1, 4-linked D-galacturonic acid (Gal*p*A), which may exist in a partially esterified form (Figure 2.1 A) or as the backbone of the rhamnogalacturonan II (RG-II; Figure 2.1 B)—another complex structure of the cell wall [(Carpita and Gibeaut 1993), (de Vries and Visser 2001), & (Ridley, O'Neill et al. 2001)]. The hairy regions, known as rhamnogalacturonan I (RG-1), consist of repeat units of 1, 4-α-D-galacturonate-1, 2-α-L-rhamnose. The rhamnosyl residues may be substituted with L-arabinose or D-galactose (Figure 2.2), indicating the complexity of the *pectate network* [(Cho, Lee et al. 2001) & (Ridley, O'Neill et al. 2001)]. The hemicellulose, pectin, as well as the aromatic polymer lignin interact with the cellulose mircofibrils to create a very rigid structure, strengthening the plant cell wall, limiting cell growth, and reducing cell wall biodegradability [(Cave 1969) & (de Vries and Visser 2001)].

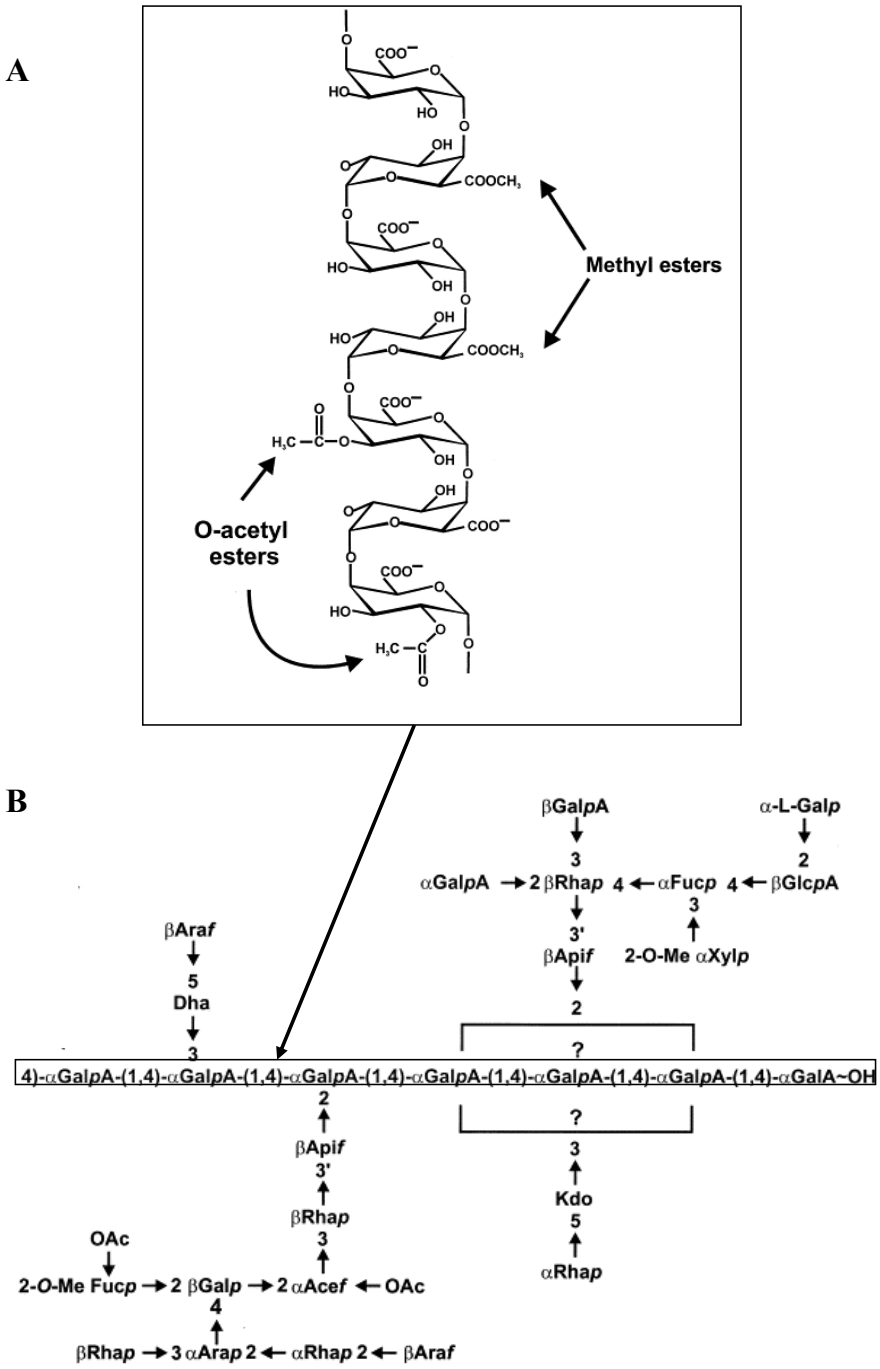

Figure 2.1. **Primary structure of homogalacturan and schematic of Rhamnogalacturonan II.** (**A**) shows the cartoon structure of homogalacturonan (HG) and (**B**) is a schematic of rhamnogalacturonan II (RG-II) with its complex substituents and HG as the backbone. Taken from (Ridley, O'Neill et al. 2001).

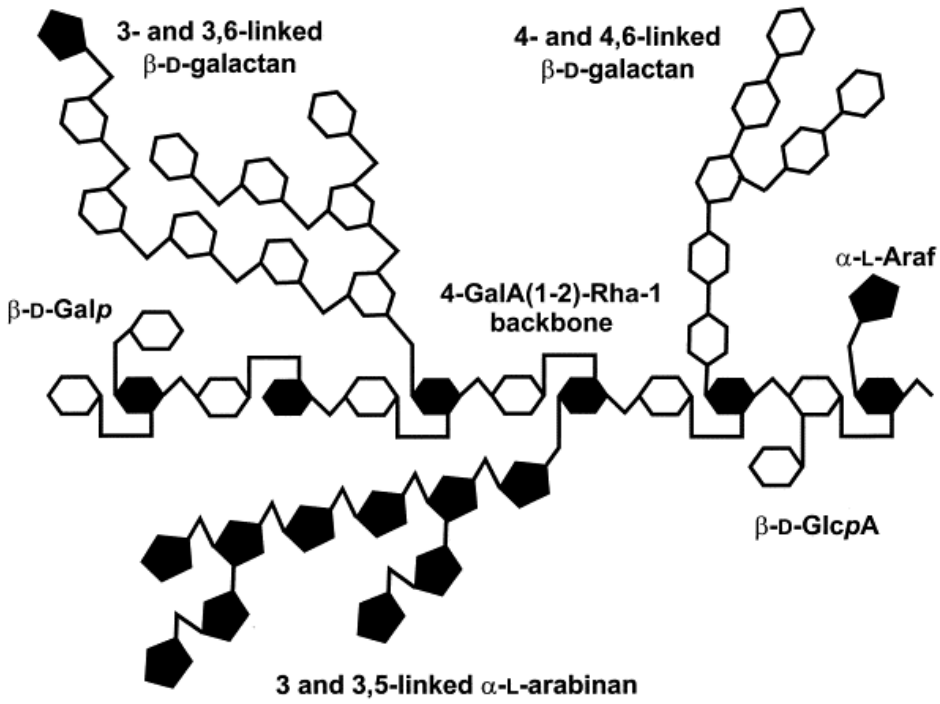

Figure 2.2. **The primary structure of rhamnogalacturan I that is a part of the pectate network**. Taken from (Ridley, O'Neill et al. 2001).

The strong, stable, and rigid structure of the plant cell wall plays an important role as the first line of defense against cell invasion by pathogens. Thus, plant pathogenic fungi and bacteria produce a vast array of enzymes capable of degrading these complex carbohydrate structures in both the *cellulose* network and the *pectate* network. The pectate network has been well characterized [(de Vries and Visser 2001) & (Ridley, O'Neill et al. 2001)] and many classes of pectin-degrading enzymes (PDEs) or pectinases have been studied and determined [(Yoder, Keen et al. 1993), (Lietzke, Scavetta et al. 1996), (Pickersgill, Jenkins et al. 1994), (Mayans, Scott et al. 1997), (Vitali, Schick et al. 1998), (Petersen, Kauppinen et al. 1997) and (van Santen,

Benen et al. 1999)]. Pectinases are classified as either glycoside hydrolases [(Herron, Benen et al. 2000) and (Davies and Henrissat 1995)] or polysaccharide lyases [(Herron, Benen et al. 2000), , and (Sutherland 1995)]. The glycoside hydrolases (EC 3.2.1.-; Figure 2.3) use aspartic or glutamic acids as catalytic residues and display optimum activity at a more acidic pH. These hydrolase mechanisms are classified as either inverting single-displacement (Figure 2.4**a**) or retaining double-displacement (Figure 2.4**b**) depending on the stereochemical outcome of the anomeric configuration about the C-1 adjacent to the scissile glycosidic oxygen (Davies and Henrissat 1995). In hydrolases, the inverting mechanism (Figure 2.4) is initiated by the activation of a water molecule. Hydrolysis occurs with the conjugate base of an acidic residue acting as a general base. The activated water molecule makes a nucleophilic attack on the sugar anomeric carbon, and a different acidic residue acts as a general acid, donating a proton to the glycosidic oxygen of the scissile bond [(Herron, Benen et al. 2000) and (van Santen, Benen et al. 1999)]. On the other hand, polysaccharide lyases (EC 4.2.2.-) hydrolyze the glycosidic bond via a β-elimination reaction mechanism (Figure 2.5) at basic pHs, compared to glycoside hydrolases, using residues that still remain under investigation [(Herron, Benen et al. 2000), and (Sutherland 1995)]. The β-elimination mechanism (Figure 2.5) proposed for these lyases proceeds in three steps: (a) the neutralization of the negative charge on the carboxylate anion of an adjacent sugar moiety, (b) abstraction of a proton from the C-5 position of the polygalacturonic acid residue through ester or salt bridge formation, and (c) proton donation by a general acid or water molecule and concomitant β-elimination of the substrate resulting in an unsaturated $\Delta^{4,5}$ bond in the product [(Gacesa 1987) and (Herron, Benen et al. 2000)].

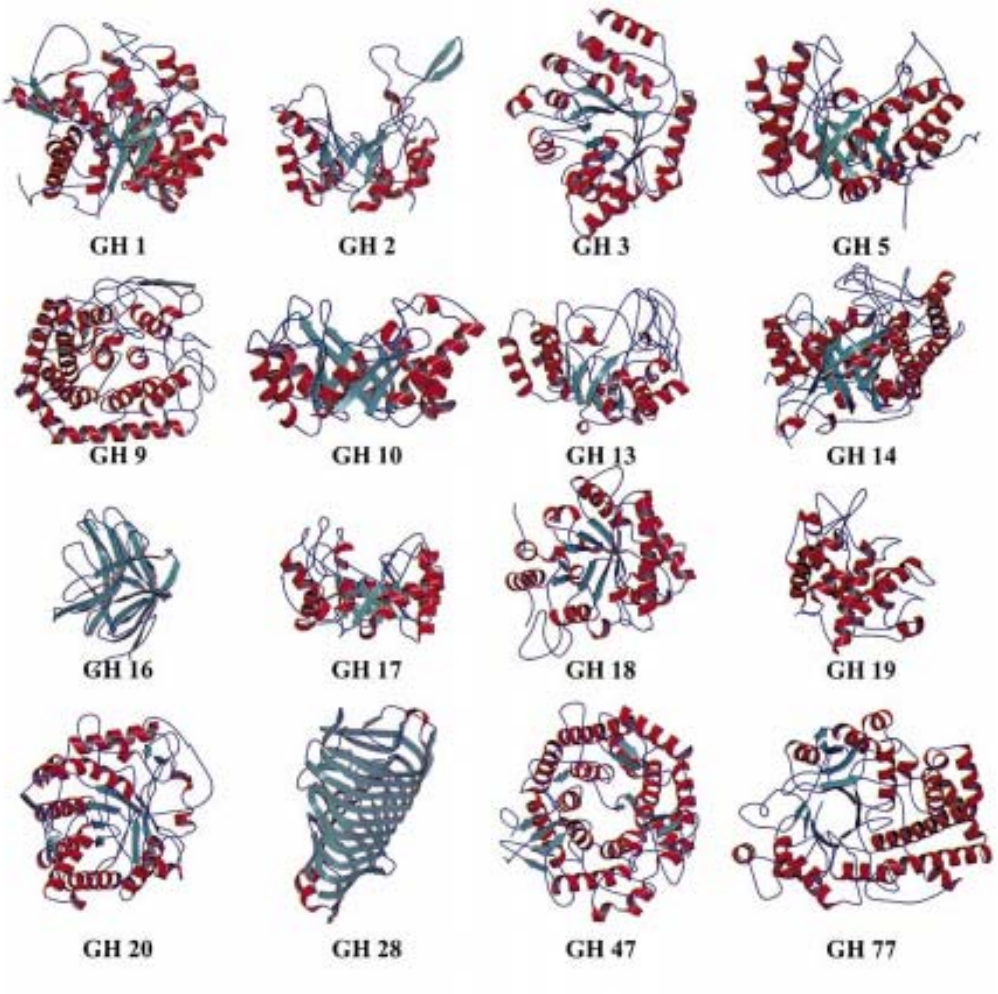

Figure 2.3. **The different families of glycoside hydrolases (GH) are illustrated in their 3D structure** (Henrissat, Coutinho et al. 2001).

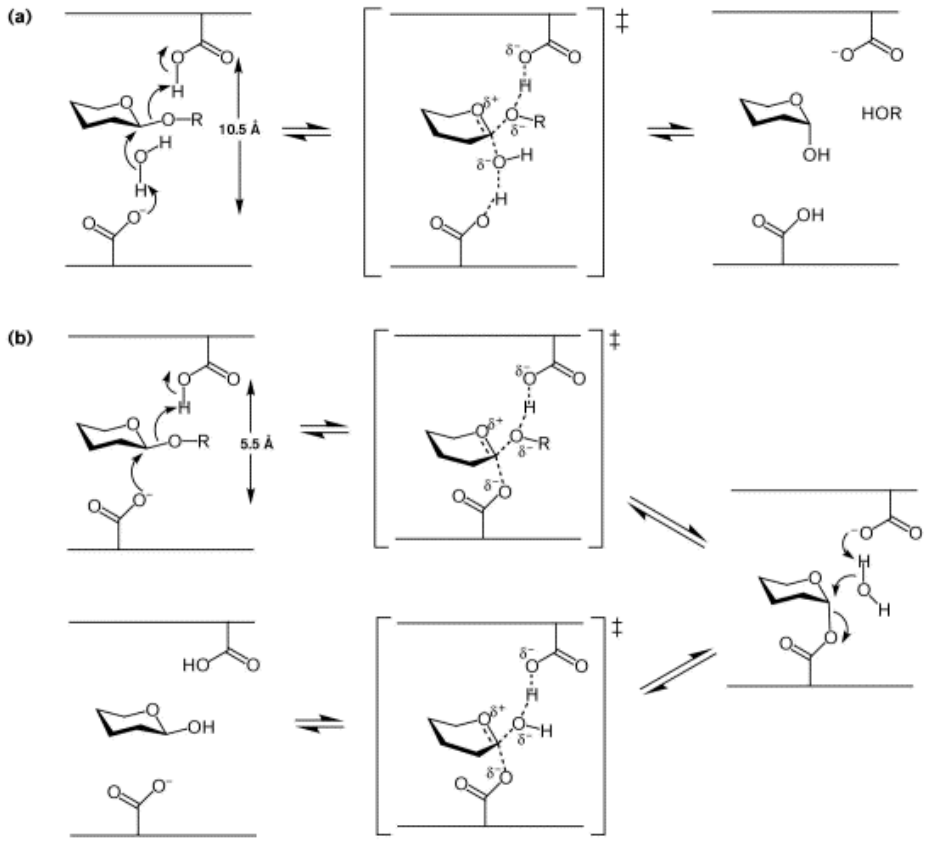

Figure 2.4. **A general glycosidase catalytic mechanism is illustrated.** (**a**) An inverting (single displacement mechanism) and (**b**) a retaining glycosidase (double displacement mechanism). Taken from (Rye and Withers 2000).

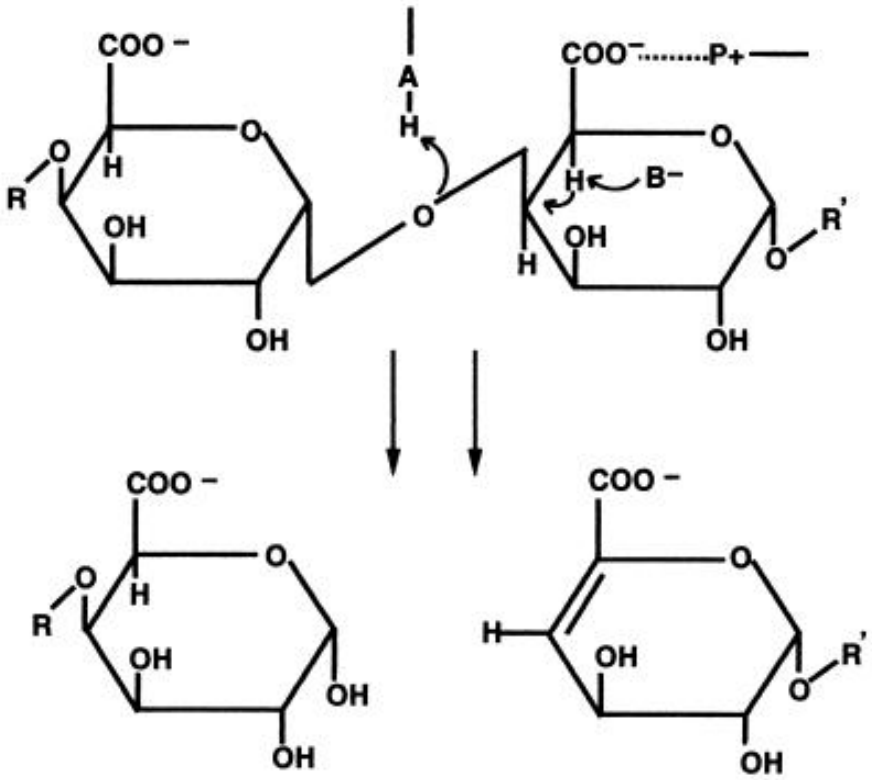

Figure 2.5. **A general polysaccharide lyase mechanism proposed by Gacesa P., 1987**. The schematic diagram of a-1,4-polygalacturonic acid has a P+ which neutralizes charge on carboxylic acid group, a general base (B) which abstracts a proton from the C-5 position and a general acid or water molecule (A) that donates a proton to the glycosidic oxygen, with unsaturated $\Delta^{4,5}$ bond in the product. From (Herron, Benen et al. 2000).

The pectinases that degrade the backbone regions of pectin using the glycosyl hydrolase or polysaccharide lyase mechanisms are pectin, pectate, and rhamnogalacturonan lyases, rhamnogalacturonan hydrolases, and polygalacturonases (Benen, Kester et al. 1999). Pectin methylesterases (PMEs), as well as pectin- and rhamnogalacturonan acetylesterases (RGAEs) degrade the different esterified substituents of the pectin backbone (van Santen, Benen et al. 1999), but are also considered part of the PDEs. These enzymes are important in pathogenicity for the fungi and bacteria (Cooper 1995) that attempt to invade plant cell walls. Due to their primary role in fungal and bacterial secretion as well as plant cell wall invasion, polygalacturonases (EC 3.2.1.15) have ensued widespread recognition.

Many plant pathogens secrete polygalacturonases to hydrolyze the α-1,4-D-polygalacturonate regions ("smooth" regions of pectin) as part of their initial process of invasion (Jones, Albersheim, P. et al. 1972). The polygalacturonases exhibit three types of modes of action known as *endo*-, *exo*-, and *endo/exo*- [(Heldt-Hansen, Kofod et al. 1996) and (Cook, Clay et al. 1999)]. The most well characterized and studied are the *endo*polygalacturonases (EPGs), which hydrolyze the long polygalacturonic acid (PGA) chains at internal glycosidic linkages within the smooth pectin regions. Their ability to hydrolyze homogalacturonan has lead to the widespread use of the EPGs as an industrial tool for fruit juice clarification, texture modification of purees, and release of color in juices and wines [(Grassin and Fauquembergue 1996) and (Heldt-Hansen, Kofod et al. 1996)]. They are also used as an aid in food processing (Bonnin, Le Goff et al. 2001) and determination of pectin structure [(Shea and Hatfield 1993), (Bonnin, Le Goff et al. 2001), and (Ros, Schols et al. 1996)]. EPGs prefer only the PGA form of homogalacturonan as a substrate, therefore other enzymes such as PMEs and RGAEs are required to remove these methyl and acetyl ester substituents on the polygalacturonic acid main-

chain to allow the EPGs to degrade the PGA (Benen, Van Alebeek et al. 2003). After the removal of the ester derivatives from the backbone, the PGA seems relatively simple in nature—consisting of only the α-1,4-galacturonate repeat units. However, the fungal organism *Aspergillus niger* alone has at least 7 EPGs that it secretes to degrade this simple, yet complex polymer (Benen, Van Alebeek et al. 2003). The role of the 7 EPGs in pectin degradation has been investigated and partially documented [(Benen, Van Alebeek et al. 2003), (Parenicova, Benen et al. 1998), (Benen, Kester et al. 1999), & (Parenicova, Benen et al. 2000)], but there remains a lack of information about EPGs and their overall structural architecture in the presence of a bound substrate.

EPGs belong to family 28 hydrolases (Figure 2.3, **GH 28**), which have a characteristic fold known as a parallel β-helix [(Jenkins, Mayans et al. 1998) & (Herron, Benen et al. 2000)]. The parallel β-helix was first discovered in 1993 in the crystal structure of pectate lyase C from *E. chrysanthemi* (Yoder, Keen et al. 1993). This characteristic fold is shared by the majority of pectinolytic enzymes (Figure 2.6), as well as other polysaccharide-degrading enzymes and does not discriminate between hydrolase and lyase mechanisms. Each enzyme shares a similar fold and substrate binding cleft [(Choi, Lee et al. 2004) & (Herron, Benen et al. 2000)]. The parallel β-helix fold has been found in enzymes that degrade oligosaccharides found on cell surface receptors, such as the P22 tailspike *endo*rhamnosidase [(Herron, Benen et al. 2000) & (Steinbacher, Baxa et al. 1996)]. In addition, this fold has been ascertained in enzymes that degrade oligosaccharides located in the mammalian intercellular matrix, such as chondroitinase B (Herron, Benen et al. 2000), (Huang, Matte et al. 1999) & (Michel, Pojasek et al. 2004)]. So, it is not unforeseen that the fold is dominant in the family 28 hydrolases (Figure 2.3 and 2.6).

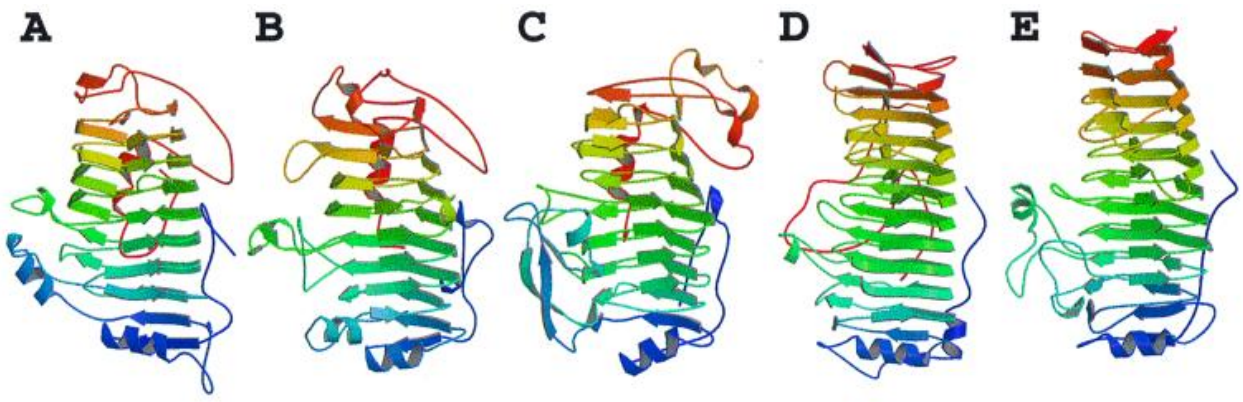

Figure 2.6. **The parallel β-helix motif conserved among plant cell wall degrading enzymes**. The examples shown are (**A**) *E. chrysanthemi* pectate lyase C, (**B**) *E. chrysanthemi* pectate lyase E, (**C**) *A. niger* pectin lyase B, (**D**) *E. caratovora* polygalacturonase, and (**E**) *A. aculeatus* rhamnolgalacturonase. From Herron, S., et al, 2000.

The crystal structure of rhamnogalacturonase A from *Aspergillus aculeatus* (Figure 2.6 **E**) was the first of the family 28 enzymes to be determined (Petersen, Kauppinen et al. 1997). Soon after, five other crystal structures of polygalacturonase (PG) enzymes from various organisms were elucidated and shown to have the right-handed parallel β-helix architecture [(Choi, Lee et al. 2004), (Pickersgill, Smith et al. 1998), (van Santen, Benen et al. 1999), (Cho, Lee et al. 2001), (Shimizu, Nakatsu et al. 2002), & (Federici, Mattei et al. 1999)]. In general, this right-handed parallel β-helix architecture is made up of four twisted parallel β-sheets called PB1, PB1a, PB2, and PB3 (Figure 2.7). PB1 serves as the bottom of the cleft, while PB2 defines the back wall of the cylinder opposite the cleft, whereas PB1a and PB3 constitute the sides of the parallel β-helix [(van Santen, Benen et al. 1999) & (Choi, Lee et al. 2004)]. These parallel β-sheets are separated by turns called T1, T2, T3, and T4 that are considered either loop regions or sharp bends. It has been documented that T2 and T3 show very high sequence and structure conservation within the family of enzymes, whereas turns T1 and T4, which protrude into the binding cleft, display considerable variation in sequence and overall structure [(Choi, Lee et al.

2004), (Markovic and Janecek 2001), (Cho, Lee et al. 2001), (Shimizu, Nakatsu et al. 2002) & (van Santen, Benen et al. 1999)].

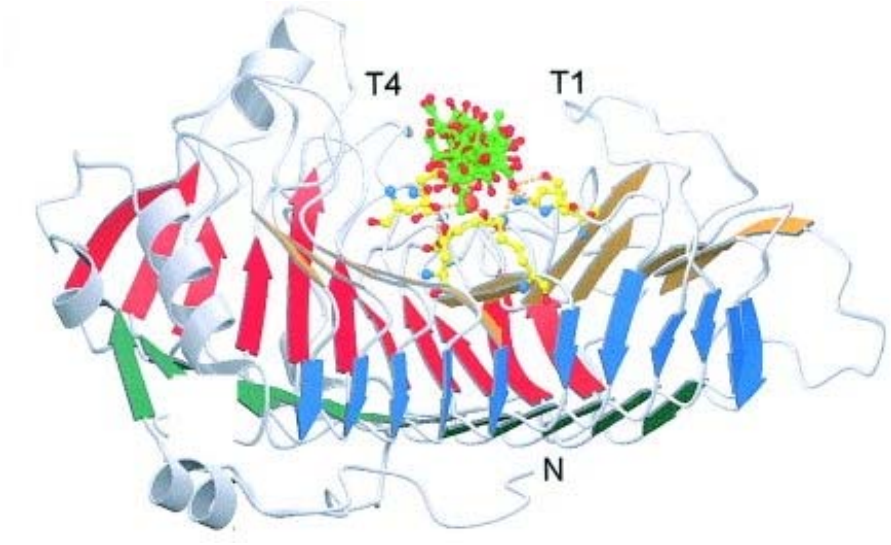

Figure 2.7. **The characteristic fold of the parallel β-helix illustrated in Rhamnogalacturonase A from *Aspergillus aculeatus***. The N terminus is labeled N and the C terminus is on the right away from the viewer. PB1, PB1a, PB2, and PB3 are shown in orange, blue, green, and red, respectively. PB1 makes up the binding cleft where the RG-I substrate is bound. T1 and T4 are the loop regions that are not structurally conserved between enzymes. From (Choi, Lee et al. 2004).

It is important to note that although there are distinct differences between the mechanisms of the hydrolases and the polysaccharide lyases, the overall parallel β-helix structure remains conserved for both families, despite a marginal sequence identity of less than 18 % (Benen, Van Alebeek et al. 2003). The only major structural differences between these families are in the number of strands in the β-sheet region of the parallel β-helix and the length of the loops extending from the helix core.

The parallel β-helix structure folds to form a consistent binding cleft for interactions with a long-chain oligosaccharide substrate. The cleft extends from the N-terminus to the C-terminus in both families of enzymes, with multiple sub-sites spanning the binding region. These sub-sites are based on the sugar-binding sub-site nomenclature proposed by Davies, G. J. et al (1997). In PGs (based on the principles shown in Figure 2.8), each sub-site accommodates a single galacturonic acid unit. Each unit interacts with the residues around that particular sub-site. A variety of experiments have examined the affinity at each individual site as well as identified the overall specificity of the enzyme (Benen, Van Alebeek et al. 2003). These sub-sites were used to map the binding architecture of the enzyme-substrate complex with respect to substrate orientation and position (Pages, Heijne et al. 2000). Active site topologies were studied using sub-site mapping in conjunction with other techniques, such as site-directed mutagenesis and kinetic analysis [(Armand, Wagemaker et al. 2000) & (Pages, Kester et al. 2001)]. Sub-site mapping has proven to be a useful tool in naming, organizing, and determining the binding domain of many polysaccharide-degrading enzymes.

Even though many representative structures of the family 28 hydrolases have been determined and characterized, much remains unknown. Moreover, there is no known crystal structure for a family 28 hydrolase bound to an oligomeric pectin substrate. The only enzyme-oligomeric substrate complex determined for any of the pectinolytic enzymes is the *Erwinia chrysanthemi* pectin lyase C─(Gal*p*A)$_5$ complex (Scavetta, Herron et al. 1999). Also, there has been a crystal structure determined with a (Gal*p*A)$_2$ ligand bound to a PG enzyme from *S. purpureum* (Shimizu, Nakatsu et al. 2002). Due to the lack of information concerning pectin degrading enzyme-substrate complexes, theoretical docking and molecular dynamics (MD)

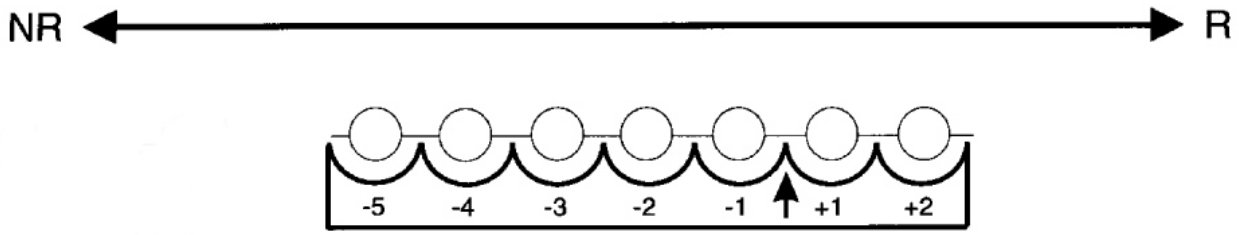

Figure 2.8. **A schematic representation of sugar-binding sub-sites for numerous glycosyl hydrolases**. The conventional scheme for drawing the binding site is with the non-reducing end (NR) of the substrate to the left and the reducing end (R) to the right. This sub-site map uses the *–n*, *+n* system applied to enzymes that cleave dissacharide units from the reducing end of the substrate as proposed for *T.reesi* cellobiohydrolase I (Davies, Wilson et al. 1997). The arrow denotes the cleavage point between residual units. Taken from (Davies, Wilson et al. 1997).

simulations have been used to study the binding interactions of these complexes. In particular, the crystal structure of native polygalacturonase from *A. aculeatus* was solved and an octameric fragment of galacturonic acid (Gal*p*A)$_8$ was modeled with a flexible docking technique, followed by MD simulations, and energy minimization (Cho, Lee et al. 2001). Another recent study employed flexible docking strategies and solvated MD to model a complex of rhamnogalacturonase from *A. aculeatus* and a dodecameric fragment of RG-I (Choi, Lee et al. 2004). These studies have been useful in delineating the sub-sites of PDEs, as well as describing the functions of the sites in catalysis and/or substrate binding. Both studies indicate possible regions for further study of substrate binding via site-directed mutagenesis, and provide conformations for the substrate in its bound form. The modeling suggests that the Gal*p*A residue in the –1 sub-site (Figure 2.8 for definition) prefers a half-chair ($^4H_3$) form, in which it approaches a transition state conformation [(Cho, Lee et al. 2001)& (Choi, Lee et al. 2004)]. However, these studies did not consider experimental data, but focused solely on the modeling of the complex, as the goal of the study was not to validate experimental data, but to give insight into ligand binding as well as enzyme-substrate interactions.

Excluding data from crystal structures, there has not been a model of a pectin-degrading enzyme-substrate complex, which validates experimental data partly because there is little known biochemical data for these enzymes. However, one of the most well documented and characterized enzymes in the family 28 hydrolases is the *Endo*polygalacturonase II (EPG-II) from *Aspergillus niger*. Active site and sub-site mapping by site-directed mutagenesis has been used to predict residues important in substrate binding and catalysis. Specific activity, product progression, and kinetic parameters (K_m and V_{max}) have also been used to determine the residues essential for binding or those that are catalytic in nature [(Armand, Wagemaker et al. 2000) & (Pages, Heijne et al. 2000)]. Bond cleavage frequencies (BCFs) have been calculated to determine the +1/-1 catalytic region (Armand, Wagemaker et al. 2000) and various degrees of pectin esterification have been analyzed in conjunction with hydrolytic cleavage (Pages, Heijne et al. 2000). In this work, it was determined that the substrate binds with the non-reducing end towards the N-terminus (Pages, Heijne et al. 2000) in a linear conformation. Further, it was concluded that Arg-256, Lys-258, and Tyr-291 are important for enzyme and substrate stability, while Asp-180, 201, 202 play a direct role in catalysis with His-223 indirectly stabilizing the complex or possibly helping to ionize a catalytic residue [(Pages, Heijne et al. 2000), (Armand, Wagemaker et al. 2000), & (van Santen, Benen et al. 1999)].

Another technique used in studying the EPG-II—(Gal*p*A)$_8$ complex was hydrogen/deuterium (H_2O/D_2O) exchange mass spectroscopy (amide exchange MS), often used for studying protein three-dimensional structure. Briefly, amide exchange MS is used to monitor the rates of exchange of amide hydrogens in the protein backbone with deuterium in the solvent. Since amide hydrogens are labile and will freely exchange with protons or deuterons in solution, solvent accessible amino acids will incorporate deuterium when a protein is dissolved in D_2O or

a mixture of D_2O and water. The incorporation of one deuteron results in a 1 Dalton overall increase in the mass of the protein, with a maximal increase directly proportional to the number of amino acids in the protein itself. This is easily monitored by mass spectrometers. In theory, all solvent accessible amino acids would freely exchange protons for deuterons, however, interactions between secondary, tertiary and quaternary structures create sites protected from the solvent. This experimentally defines the overall protein architecture [(Zhang and Smith 1993), (Katta and Chait 1993), & (King, Lumpkin et al. 2002)].

Recently, amide exchange MS was used to study protein/ligand complexes. In the unbound protein, solvent accessible regions are exposed to D_2O solvent. In the complex, the ligand shields regions of the protein surface from deuterium exchange, similar to the shielding created by secondary, tertiary and quaternary structures of the protein [(Ehring 1999), (Smith, Liu et al. 1996), (Akashi and Takio 2000), & (Wang, Blanchard et al. 1997)]. The difference in incorporation between the unbound and bound systems identifies the residues that are essential for stabilization of the protein—ligand complex. This type of amide exchange MS investigation was used to predict binding regions for the EPG-II—$(Gal\textit{p}A)_8$ interaction complex (King, Bergmann et al. 2002).

The goal of our research is to use MD simulations to create a detailed model of the complex and offer a structural interpretation for the experimental data provided by mutagenesis and amide exchange MS studies. Our approach consists of two distinct techniques (1) create a simple model for interpreting amide exchange MS based on computing the solvent accessible surface area (SASA) and (2) compute the effects of point mutations in the EPG-II—$(Gal\textit{p}A)_8$ complex using computational alanine scanning. Both of these techniques required the use of explicit solvent MD simulations as well as molecular mechanics and Poisson-

Boltzmann/generalized Born surface area (MM-PB(GB)SA) calculations [(Kollman, Massova et al. 2000) & (Gohlke and Case 2004)]. The SASA model for amide exchange MS is a novel approach discussed in the next chapter, however, free energy calculations and computational alanine scanning methods via MM-PB(GB)SA have been used on various systems [(Kollman, Massova et al. 2000), (Gohlke and Case 2004), (Huo, Massova et al. 2002), (Chong, Duan et al. 1999), (Ford, Weimar et al. 2003) & (Massova and Kollman 2000)]. The first free energy calculations were done in the mid-1980s [(Wong and McCammon 1986), (Warshel, Sussman et al. 1986), (Kollman 1993), & (Jorgensen and Ravimohan 1985)]. The calculation of relative free energies of binding greatly enhanced the ability of theoretical science to exist in synergy with experimental studies (Kollman 1993).

With the combined use of explicit solvent and continuum (implicit) solvent models [(Chong, Duan et al. 1999), (Massova and Kollman 2000), (Kuhn and Kollman 2000), (Gohlke and Case 2004) & (Ford, Weimar et al. 2003)], one can derive both structure and relative binding free energies from molecular dynamics trajectories (Kollman, Massova et al. 2000). Explicit solvent models are needed to derive structure because the continuum models are still under development (Zhou 2003). These explicit models can be used to accurately predict structure as well as calculate relative binding free energies, but the free energy calculations are very intense. Through the course of the simulation, it is difficult to calculate these relative binding free energies due to the dynamics of water molecules involved in the solvent/solute interactions. The removal of water molecules from the trajectory makes the complicated prediction of solvent/solute and solute/solute energy interactions more accessible and efficient with use of continuum solvent models.

The continuum solvent model appears to be promising because it can calculate relative interaction and solvation energies in complex systems (Massova and Kollman 2000). One such program that implements a continuum model in a post-MD analysis is MM-PB(GB)SA, applied and studied by Srinivasan et al, (1998). It has shown promise in combining molecular mechanical gas energies (employing geometries from accurate explicit solvated MD simulations) with a continuum solvation approach to estimate the solvent effect. The continuum solvent model has also been useful in the application of normal mode analysis to evaluate entropy contributions to the total free energy (Massova and Kollman 2000). Poisson-Boltzmann (PB) analysis [(Honig and Nicholls 1995) & (Gilson and Honig 1988)], generalized Born (GB) analysis [(Massova and Kollman 2000), (Feig, Onufriev et al. 2004), & (Gohlke and Case 2004)] and nonpolar solvation free energy calculations to estimate free energies are also combined with this approach (Huo, Massova et al. 2002). The GB analysis is based on a two-dielectric model, where the electrostatic contribution to the free energy of solvation is

$$\Delta G_{pol} = -\tfrac{1}{2}\,(1-1/\varepsilon_{\omega})\sum_{i,j}(q_i q_j / f_{GB}) \qquad (1)$$

where q_i and q_j are partial charges, ε_{ω} is the solvent dielectric constant, and f_{GB} is a function that interpolates between an "effective Born radius" R_i, when the distance r_{ij} between atoms is short, and r_{ij} itself at large distances (Bashford and Case 2000). To obtain the electrostatic potential φ for the GB model in eq. (**1**), it is necessary to solve the Poisson equation,

$$\nabla\,[\varepsilon(\mathbf{r})\ \nabla\,\varphi(\mathbf{r})] = -4\pi\rho(\mathbf{r}) \qquad (2)$$

where ρ is the charge distribution, and the dielectric constant ε takes on the solute molecular dielectric ε_{in} in the solute interior and the exterior dielectric constant ε_{ex} elsewhere (Bashford and Case 2000). If eq. (**2**) is expanded to include ionic salt concentrations, it is called the Poisson-Boltzmann equation. Because of the significant computational expense of the numerical solution to the PB equation, many efforts have focused on equivalent energetic descriptions that describe the same or similar continuum solvent models in a much more efficient manner (Feig, Onufriev et al. 2004). Since GB is significantly faster than PB, MM-GBSA seems more appealing to calculate binding free energy estimates, for use in alanine scanning mutation studies (Gohlke and Case 2004). For this reason, the MM-GBSA approach was employed in this study.

Before using the MM-GBSA method, an MD simulation using the particle mesh ewald (PME) is carried out with a periodic box of explicit water molecules (Huo, Massova et al. 2002). The simulation produces a trajectory file that contains all the coordinates of each atom in its respective Cartesian space over the time course of the simulation. Representative conformations of the macromolecule are extracted from the trajectory. Any ions and water molecules are removed and their effects estimated with the MM-GBSA continuum solvent model (Huo, Massova et al. 2002). The MM-GBSA method calculates the free energy, $G_{molecule}$, by the equation:

$$\mathbf{\langle G_{molecule} \rangle = \langle E_{MM} \rangle + \langle G_{GBSA} \rangle - \langle TS_{MM} \rangle} \qquad (\mathbf{3})$$

where $\langle G_{molecule} \rangle$ is the calculated average free energy, and $\langle E_{MM} \rangle$ is the average molecular mechanical energy,

$$\mathbf{\langle E_{MM} \rangle = \langle E_{bond} + E_{angle} + E_{tors} + E_{vdw} + E_{elec} \rangle} \qquad (\mathbf{4})$$

$$\langle \Delta G_{binding} \rangle = \langle G_{complex} \rangle - (\langle G_{protein} \rangle + \langle G_{ligand} \rangle) \qquad (6)$$

The evaluation of (**3**) can be done by two methods: either (a) run separate simulations of the complex, receptor and ligand then determine the absolute free energies for each or (b) split a single MD trajectory of the complex into trajectories for the complex, receptor, and ligand. Option (b) assumes that the snapshots of the receptor and ligand are comparable to the geometries of those that would be obtained from the separate simulations as defined in the (a) option (Kollman, Massova et al. 2000). Since (b) is more efficient, it has been widely used and is the one implemented here.

An extension of the MM-GBSA approach is a technique called computational alanine scanning mutagenesis (Massova and Kollman 1999). This technique can be used to study the effects of point mutations on binding energy. This approach can be used to scan the sites of interest in the receptor—ligand interface for the 'hotspots' and evaluate what the changes to the binding free energy would be upon modifications/mutations of the residues at the binding region. The most efficient application is to take a complex trajectory and use the same set of snapshots from the wild-type complex to calculate each alanine mutant (Huo, Massova et al. 2002). The relative free energy of binding, $\Delta\Delta G_{binding}$, is calculated as follows:

$$\Delta\Delta G_{binding} = G_{binding\_wild\ type} - G_{binding\_mutant} \qquad (7)$$

where $G_{binding\_wild\ type}$ is the free energy of binding determined from the native structure and the $G_{binding\_mutant}$ is the free energy calculated from each alanine mutation. For the alanine scanning process, each mutant is generated from the wild type structure by replacing the side chain at the mutated position with a methyl group (Huo, Massova et al. 2002). Each mutant binding energy

can then be used in eq. (**7**) to predict residues within the protein that are important for favorable or unfavorable binding of the substrate. A major drawback to this approach is that the changes in binding can only be estimated if a smaller group replaces a larger one, because mutations of a smaller group to a bigger one creates steric clashes that must be dealt with before binding energy calculations can be determined (Gohlke and Klebe 2001). However, this is not a problem for alanine, except in the case of a glycine to alanine mutation.

Additional drawbacks of the MM-GBSA and computational alanine scanning approaches are that most, if not all, studies do not consider charged ligands or carbohydrates (Ford, Weimar et al. 2003). Moreover, the most widely used method of neutralizing charged systems is through the addition of counterions (Ibragimova and Wade 1998). In the post-processing method, these ions are removed with the explicit solvent before the MM-GBSA approach is used. This is one of the major considerations we faced when dealing with the EPG-II—(Gal*p*A)$_8$ complex. We do not know the extent of the effects that the removal of the counterions from the system will have on ΔG. However, we were able to reproduce the results of both site-directed mutagenesis and amide-exchange MS. Furthermore, we were able to offer insight into a new hydrolytic-cleavage pathway and demonstrate new regions in the EPG-II enzyme that are important in binding and stabilizing the (Gal*p*A)$_8$ substrate.

CHAPTER 3

THE ENDOPOLYGALACTURONASE II—OCTAGALACTURONATE COMPLEX: MD SIMULATIONS PREDICT BINDING INTERACTIONS AND GIVE INSIGHT INTO A NOVEL CLEAVAGE PATHWAY [1]

[1] Barnes, J. W., Bergmann, C., Orlando, R., and Woods, R. J. To be submitted to *Journal of Biological Chemistry*

ABSTRACT

Molecular Dynamics simulations were used to study the interactions between the enzyme *Endo*polygalacturonase II (EPG-II) from *Aspergillus niger* and an octameric fragment of its polysaccharide substrate 1,4-α-D-polygalacturonic acid (Gal*p*A)$_8$. The results from the simulations provide a detailed dynamic model that combines amide exchange Mass Spectrometry (MS) data with a simple model for exchange based on solvent accessible surface area (SASA) calculations. These calculations predict sites of molecular shielding as well as indicate residues in the enzyme that play a role in binding the substrate. A conformational change of the linear form of the substrate via a chair-to-distorted boat (^{1,4}B) transition was observed during the MD simulation that corresponded to the conversion of the ground state to an activated bent form. The dynamic flexibility of the substrate is supported by the amide exchange MS data and suggests the presence of a previously unidentified secondary binding cleft, thus giving insight into the cleavage pathway of the EPG-II enzyme. The nature of the enzyme-substrate interaction was further defined by computational alanine scanning studies, which identified residues that are important for stabilization of the substrate in the bound form as well as in the activated conformation by determining the change in free energy of binding when various residues in the wild type enzyme were mutated to alanine. The results from the alanine scanning experiments correlated well with the site-directed mutagenesis studies reported earlier. Moreover, advances were made into identifying key residues that drive the low energy ground state to the activated state. These advances suggest that the chair-to-distorted ^{1,4}B transition

energy is not solely defined by the chemistry of the hydrolytic mechanism but the electrostatic energies also play a role in the enzyme-substrate complex.

Key Words: Molecular Dynamics, Solvent accessible surface area, Amide exchange Mass Spectrometry, *Endo*polygalacturonase II, MM-GBSA, and Computational alanine scanning.

Abbreviations: EPG-II, *endo*polygalacturonase II; $(GalpA)_8$, 1,4-α-D-polygalacturonic acid; amide exchange MS, amide exchange Mass Spectrometry; SASA, solvent accessible surface area; distorted $^{1,4}B$, distorted boat.

[(van Santen, Benen et al. 1999), (Federici, Caprari et al. 2001), (Shimizu, Nakatsu et al. 2002), (Cho, Lee et al. 2001), & (van Pouderoyen, Snijder et al. 2003)] have also been determined. Information is slowly being gathered concerning the carbohydrate binding sites of EPGs. A survey of the EXPASY (EXpert Protein Analysis SYstem, Swiss Institute of Bioinformatics) database reveals that over 50 EPGs have been fully or partially sequenced. In spite of this, the amount of data concerning the substrate-binding site of EPGs is limited [(van Santen, Benen et al. 1999), (Armand, Wagemaker et al. 2000), (Caprari, Mattei et al. 1996), & (Benen, Kester et al. 1999)].

Of the pectin degrading enzymes (PDEs) whose structures have been determined, the pectin hydrolases from bacteria and fungi [(Pickersgill, Smith et al. 1998), (van Santen, Benen et al. 1999), (Federici, Caprari et al. 2001), (Shimizu, Nakatsu et al. 2002), (Cho, Lee et al. 2001), & (Petersen, Kauppinen et al. 1997)] have similar three-dimensional structures and fall into the same family of enzymes (family 28) in the Henrissat classification scheme (Henrissat 1991). Yet, these enzymes exhibit markedly different substrate specificities. RhamnogalacturonaseA (RGA) acts on RG-I, a branched alternating copolymer of 1,2-linked α-L-rhamnose and 1,4-linked α-D-galacturonic acid, while the EPGs act on the unbranched homopolygalacturonan (polyGal*p*A). There are at least two different modes of *endo* cleavage found among fungal EPGs. One of these modes, which has been defined as the *endo*/*exo* mode (Cook, Clay et al. 1999), is demonstrated by a number of fungal EPGs as well as by the EPGs produced by the bacteria *Agrobacterium tumefaciens* and *Pseudomonas solanacearum*. The other mode, defined as the *endo* mode, is demonstrated by certain fungal EPGs as well as by the major EPGs produced in tomato fruit (Cook, Clay et al. 1999).

A series of papers exploring the active site residues of *Aspergillus niger* EPG-II by site-directed mutagenesis [(Armand, Wagemaker et al. 2000), (Pages, Heijne et al. 2000), & (Pages, Kester et al. 2001)] have been of immense value in studying the substrate binding site of the fungal EPGs. Several residues were shown to be involved either directly (Asp-180, 201, and 202) or indirectly (His-223) in the catalytic activity, while other residues were shown to affect the binding stability (Tyr-291, Arg-256, and Lys-258) of the enzyme-substrate complex.

While EPGs that exhibit different modes of cleavage may all have similar gross 3D structures, and in most cases share a number of conserved amino acids, there must be alterations in the binding/active site of the enzymes that account for differences in the modes of action (Pages, Kester et al. 2001). The structure of an isolated EPG, crystallized in the absence of bound substrate, is a useful starting point, but does not itself provide sufficient data to correlate structure to mode of action. Even structures obtained from EPG co-crystallized with dimeric Gal*p*A provide little information on the mode of action since the minimum size required for significant binding is the pentamer form of Gal*p*A and activity is not observed until the trimeric stage. In order to aid in determining the mode of action, information about which amino acid residues come into contact with the substrate for each class of EPG is required to supplement the data obtained from X-ray diffraction studies.

One such supplemental experiment is amide exchange Mass Spectrometry (MS), which was used to predict the incorporation of deuterium in the EPG-II enzyme in the presence and absence of an octameric form of homogalacturonan $(Gal\textit{p}A)_8$ substrate (King, Bergmann et al. 2002). As observed by King et al., the amide exchange MS data showed that the increased levels of deuterium into the backface of the EPG-II—$(Gal\textit{p}A)_8$ complex indicated a conformational change of the enzyme when the substrate was bound, whereas residues in the binding cleft that

were potentially protected by the (Gal*p*A)$_8$ ligand showed less deuterium incorporation. The data of King et al. suggested areas of the EPG-II enzyme that were exposed to exchange as well as regions that are shielded by the substrate.

Experiments, such as X-ray diffraction studies, site-directed mutagenesis, and amide exchange MS have helped to grossly characterize the EPG-II─(Gal*p*A)$_8$ complex, but the many details of catalytic and binding architecture remain unknown. Our goal was to introduce a simple approach to provide a detailed structural model for the enzyme-substrate architecture and generate a complex with a linear octameric oligosaccharide substrate as assumed previously from sub-site mapping (Pages, Heijne et al. 2000). Unexpectedly, our results suggested that the linear oligosaccharide preferred to adopt a distorted boat (^{1,4}B) intermediate, with the distorted ring located near the catalytic site. After further analysis of the substrate, we determined that the (Gal*p*A)$_8$ substrate adopted two different conformations, linear and bent, corresponding to the all chair and chair/boat structures, respectively, during the MD simulation. Therefore, a dynamic model for both the complex with a linear and an activated bent form of the substrate was generated.

Further, we probed the role of individual amino acid side chains in stabilizing the complexes via a computational alanine scanning mutation protocol [(Massova and Kollman 1999), (Kollman, Massova et al. 2000) & (Huo, Massova et al. 2002)]. This was applied to every charged residue in the enzyme, in addition to other key residues as previously determined by site-directed mutagenesis (Armand, Wagemaker et al. 2000) & (Pages, Heijne et al. 2000). Using the data from the MD simulations of the linear and the activated bent complexes, the ΔG of binding for each mutant as well as the wild types (WT) was determined for each system. From these 146 computational mutations (73 mutations for each simulation), we were able to

show that Arg-256, Lys-258, and Tyr-291 were necessary for the overall binding of the EPG-II—(Gal*p*A)$_8$ complex, because the calculated ΔΔG energies (wild type – mutant) where more negative in comparison to the wild type when each residue was mutated to an alanine. This is in agreement with the site-directed mutagenesis studies (Armand, Wagemaker et al. 2000) & (Pages, Heijne et al. 2000). In addition, we demonstrated that a number of other residues are important in binding and suggest these residues for further mutational studies.

We also constructed a simple model for amide exchange based on determining the solvent accessible surface area (SASA) for each residue. Using this approach, we were able to predict residual shielding in the presence and absence of both the linear and activated bent forms of the (Gal*p*A)$_8$ substrate. The results from the SASA model demonstrate good agreement with the previously published amide exchange data (King, Bergmann et al. 2002). A rationalization for the amide exchange data was provided with the exploitation of both conformations of the substrate in the EPG-II binding cleft. Based on the agreement of our MD studies with the available experimental data, we propose a detailed model for the mechanism of binding and a novel cleavage pathway for the EPG-II enzyme.

EXPERIMENTAL METHODS

Construction of the EPG-II—GalpA complex—An initial model for the substrate was constructed as an octameric form of 1,4-α-D-polygalacturonic acid residues with XLEaP from AMBER 7.0 (Case, Pearlman et al. 2002) using the GLYCAM version 2000a parameters for carbohydrates and oligosaccharides [(Woods, Edge et al. 1993) & (Kirschner and Woods 2001)]. Each φ-glycosidic torsion angle was set to 60º, in agreement with the exo-anomeric effect

(Anderson 2000) and ψ was set to 0º, the generally observed experimental conformation (Anderson and Ijeh 1994). The substrate was prepared and subjected to minimization. In the initial solvent minimization, unfavorable contacts were removed from the solvated substrate while the solute coordinates were kept frozen. This was then followed by fully relaxed minimization of the whole system. Each minimization consisted of 20,000 cycles: 5,000 cycles of steepest descent and 15,000 cycles of conjugate gradient minimization. Subsequently, the solvated (Gal*p*A)$_8$ substrate was subjected to 100 ps of heating from 5K to 300K followed by production dynamics at 300K for 1,000 ps. After this equilibration period, the final structure for the oligosaccharide was positioned into the binding cleft of the enzyme with its reducing end toward the C-terminus, using the program INSIGHT (Biosym/MSI 1995) from Accelrys. Placement of the substrate inside the binding cleft of EPG-II was based on sub-site and active site topology mapping [(Pages, Heijne et al. 2000) & (Armand, Wagemaker et al. 2000)]. The sub-site map was modified to accommodate an octameric form of the polygalacturonic acid substrate. Once approximately positioned, a Lamarckian genetic algorithm from AutoDock 3.0.5 [(Goodsell, Morris et al. 1996) & (Morris, Goodsell et al. 1998)] was used to precisely dock the (Gal*p*A)$_8$ substrate into the enzyme binding cleft**.** Ten possible docking strategies of the (Gal*p*A)$_8$ substrate were mapped with AutoDock, and all gave similar low energy conformations. The most stable complex from the docking was then subjected to energy minimization and molecular dynamics, analogous to that performed on the free substrate.

Recognition of a secondary binding site upon transition of the dynamic linear to the activated bent form of the (GalpA)$_8$ substrate—An X-ray crystal structure of the EPG-II enzyme (1CZF) was obtained from the Protein Data Bank. The hydrogen atoms were added using XLEaP module of AMBER 7.0 (Case, Pearlman et al. 2002), while histidine residues that were

not associated with the zinc ions were fully protonated as appropriate for the optimum activity (pH = 4.0 – 4.5) of the EPG-II enzyme. The two histidine residues that coordinate with to the zinc ion were protonated at their respective Nδ or Nε positions as necessary for coordination with the zinc ions. Cysteine residues were also bridged by sulfide bonds where indicated. Solvation of the complex was performed in a spherical droplet of explicit water and a radius of 15 Å about the center of the docked ligand was introduced using the INSIGHT (Biosym/MSI 1995) program. The residues within the 15 Å radius droplet were allowed motional freedom, while residues outside the radius were restrained during the course of the 1000 ps MD trajectory. The droplet MD simulation employed the PARM99 protein parameters [(Case, Pearlman et al. 2002) and GLYCAM version 2000a parameters for carbohydrates and oligosaccharides (Woods, Edge et al. 1993) & (Kirschner and Woods 2001)] with an 8 Å cutoff for non-bonded interactions. Unfavorable contacts were removed from the solvated enzyme by minimization as described above. After minimization, the solvated EPG-II enzyme was subjected to 50 ps of heating from 5 K to 300 K; production dynamics at 300 K were performed for 950 ps for a total of 1000 ps.

Preparation of Molecular Dynamics simulations and equilibration of the EPG-II—$(GalpA)_8$ dynamic linear and activated bent forms of the GalpA substrate—During the droplet simulation, we observed a chair-to-distorted boat $^{1,4}B$ transition of a Gal*p*A residue at the proposed catalytic site, and it was concluded that the substrate does indeed adopt different families over the course of the trajectory. The first family was determined to be a $^4C_1$ chair at the active site and the last conformation was linear combinations of a distorted $^{1,4}B$. Therefore, we divided the complexes into two separate simulations at the point where the transition from chair to distorted $^{1,4}B$ occurred. Based on the different conformations of the substrate, linear and

activated bent complexes were generated from the initial droplet data where the transition was observed. To achieve the best accuracy possible for these systems, both EPG-II—(Gal*p*A)$_8$ complexes were solvated with a 10Å water box, which was minimized as stated above. After minimization, the solvated complexes were subjected to 100 ps of heating from 5 K to 300 K and production dynamics at 300 K for 2 ns. Particle Mesh Ewald (Darden, York et al.) was implemented with a 1.0 Å grid spacing to calculate electrostatic energies, and a fourth order spline for interpolation, while a SHAKE algorithm was applied to constrain all bonds containing hydrogen throughout the trajectory.

SASA—Solvent Accessible Surface Area—Coordinate snapshots were generated from the 2ns trajectories of the EPG-II—(Gal*p*A)$_8$ linear and bent simulations every 10 ps for a total of 200 snapshots per simulation. Residue based SASAs were computed using the MM-GBSA module of AMBER 7, by setting the values *SURFTEN* and *SURFOFF* to 1.0 and 0.0, respectively (Gohlke and Case 2004). These values are used to calculate the nonpolar contribution to desolvation energy, G_{np}.

$$\mathbf{G_{np} = SURFTEN * SASA + SURFOFF}$$

By correct choice of values for these parameters, it is therefore possible to compute SASAs on a per residue basis for an entire trajectory.

Computational Alanine Scanning on the EPG-II/GalpA complex—Computational alanine scanning was performed by using the MM-GBSA [(Sitkoff, Sharp et al. 1994) & (Weiser, Shenkin et al. 1999)] portion of the AMBER 7.0 package (Case, Pearlman et al. 2002). Each coordinate set from the MD trajectory was generated from the 2 ns simulations of both linear and

bent complexes. The 200 snapshots were calculated at a frequency of 10 ps over the course of each 2 ns run. The selection of each mutation was based on the previously determined site-directed mutagenesis studies and extended to include all positive or negative amino acids in the complex. The generalized Born parameters of Jayaram's et al. MGB (1998) were employed, while the correct radii and screening parameters were assigned in the XLEaP program of AMBER 7.0 (Case, Pearlman et al. 2002). The external dielectric constant for the surrounding solvent was set to that of water (80.0 D). In addition, the SURFTEN (Sitkoff, Sharp et al. 1994) and SURFOFF parameters that were used to compute the nonpolar contribution to the desolvation energy were set to 0.0072 and 0.00, respectively (Jayaram, Sprous et al. 1998).

RESULTS AND DISCUSSION

Protonation of histidine residues—It is important to understand underlying factors such as physiological conditions before running computationally intense processes with MD simulations. Even though the majority of the residues have a well defined ionization state at a pH of 7.0, the addition of hydrogen atoms to a protein from a crystal structure usually involves overcoming some ambiguity in terms of the ionization state of the histidine side chains. This ambiguity becomes more certain for the EPG-II enzyme, because its optimal pH activity is ~ 4.5 which alters the protonation state of histidine residues to a + 1 charge. Each residue was protonated at both the Nε and Nδ atoms with the exception of the H177 and H223, which were expected to be neutral and only protonated at their Nε and Nδ positions, because of coordination to a zinc ion. Electrostatic maps of the EPGII enzyme (Figure 3.1) illustrate the differences in the charge state of the protein when its histidines are fully ionized (with the exception of H177

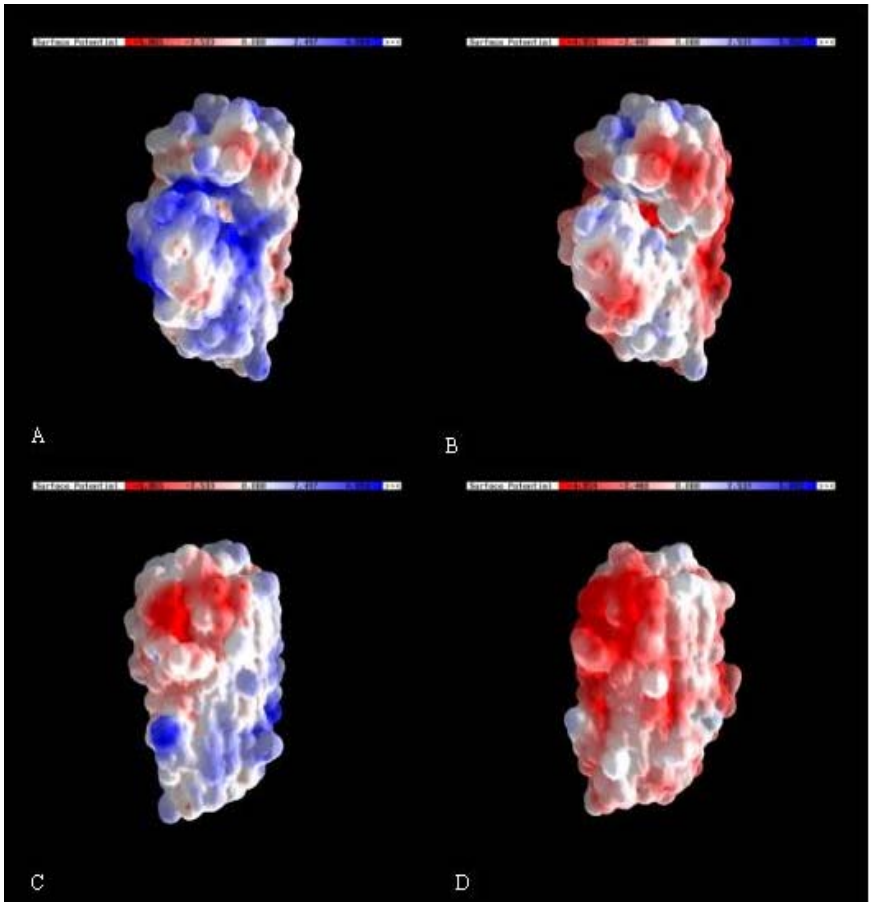

Figure 3.1. **GRASP electrostatic potential maps for the EPG-II enzyme** (Nicholls, Sharp et al. 1991). The values for the potentials shaded blue show the most electropositive areas on the accessible surface and potentials shaded red are the most electronegative regions on the surface of the EPG-II enzyme. (**A**) and (**B**) show the binding cleft of the enzyme (C-terminus at top and N-terminus at the bottom). (**C**) and (**D**) show the back of the enzyme. The potential maps of the fully protonated histidines (except the coordinating H177 and H223 which are bound by zinc) are illustrated in (**A**) and (**C**), while the neutral histidine potential maps are shown in (**B**) and (**D**). The coloring scale for the potential map is –5.065 & 4.994 kT/e for (**A**) and (**C**), while the scale for (**B**) and (**D**) is –4.959 & 5.067, respectively.

and H223) versus neutral protonation at the Nδ atom of the residues. Figure 3.1A shows a more electropositive region in the binding cleft, which suggests better charge complimentarity for a net negatively charged substrate than the more electronegative region shown in Figure 3.1B. The maps also illustrate the electrostatic differences on the backface of the protein (Figure 3.1C and 3.1D). Long-range non-specific electrostatic interactions as well as short-range electrostatics (hydrogen bonds and salt bridges) have proven to be crucial for enzyme/substrate stability. When the histidine ionization states were considered, the complex remained intact throughout the MD simulations, otherwise, the substrate drifted away from the binding site due to the negative electrostatic repulsions between the enzyme and substrate. These details confirm the importance of using appropriate ionization states in MD simulations and lend credence to the model for the EPG-II—(Gal*p*A)$_8$ complex at its pH optimum.

The initial observation of the distorted boat (^{1,4}B) in the EPG-II enzyme—During the initial droplet simulation, we observed a chair-to-distorted boat (^{1,4}B) transition of (Gal*p*A)$_8$ substrate at the catalytic site. A 1D and 2D RMSD (Figure 3.2) was generated over the course of the simulation and it was concluded that the substrate could adopt two distinct conformations. The first conformation (Figure 3.2) was determined to be a $^4C_1$ chair configuration based on the average pyranose ring torsions measured (Table 3.1) for the Gal*p*A residue found at the –1/+1 site (see Figure 3.3 for –1/+1 site). The second distinct family was determined to be linear combinations (Berces, Whitfield et al. 2001) of a distorted ^{1,4}B configuration of the Gal*p*A residue at the –1/+1 catalytic site (Figure 3.2 & 3.3). The linear combinations (Berces, Whitfield et al. 2001) of the pyranose ring are shown in Table 3.1 based on the average populations during the last 750 ps of the simulation.

The presence of a dual binding site—A sub-site topology map of the EPG-II binding cleft is shown in Figures 3.3 A and B. The binding of the octameric substrate corresponds with the sub-site and key residues that interact with the flexible ligand in the binding region as determined by site-directed mutations and charge-charge interactions. Figure 3.3 A shows the

Table 3.1

Characterization of the Gal*p*A residue modeled at the catalytic site with the linear combinations of the most representative ideal conformations over the course of the simulation.

Linear Combinations	τ_1 (C1-C2-C3-C4)†	τ_2 (C2-C3-C4-C5)†	τ_3 (C3-C4-C5-O)†	τ_4 (C4-C5-O-C1)†	τ_5 (C5-O-C1-C2)†	τ_6 (O-C1-C2-C3)†
$^4C_1$	-60.0	60.0	-60.0	60.0	-60.0	60.0
$^1S_3$	-30.0	60.0	-30.0	-30.0	60.0	-30.0
$^{1,4}B$	0.0	60.0	-60.0	0.0	60.0	-60.0
$^1S_5$	30.0	30.0	-60.0	30.0	30.0	-60.0
First 250 ps*	**-54.0**(7.8)	**56.6**(6.1)	**-56.9**(7.4)	**58.1**(11.6)	**-53.0**(13.9)	**49.5**(11.5)
Last 750 ps*	**-8.3**(17.6)	**53.5**(11.0)	**-57.0**(12.1)	**10.8**(21.8)	**38.9**(16.2)	**-39.1**(12.5)
-251-400 ps	**-23.5**(15.5)	**61.3**(9.1)	**-46.2**(11.4)	**-8.8**(21.5)	**49.7**(19.1)	**-30.7**(14.5)
-401-650 ps	**-0.20**(15.5)	**48.7**(11.0)	**-61.0**(9.6)	**19.4**(17.5)	**34.3**(13.3)	**-43.4**(9.9)
-651-1000 ps	**-7.6**(15.8)	**53.7**(9.7)	**-58.8**(11.3)	**12.9**(19.7)	**37.7**(14.8)	**-39.5**(11.5)

* averages determined by CARNAL (Case, Pearlman et al. 2002)
† denotes the measured torsions about the pyranose ring.

dynamic linear form of the $(Gal pA)_8$ substrate bound in the pre-determined sub-site map (Pages, Heijne et al. 2000) showing many more interactions in the binding cleft. The sites along the cleft are numbered from –5 to +3 with the region between the –1 and +1 site prioritized as the location of the catalytic site. The map defines the region where the substrate binds linearly, but does not consider a flexible ligand that distorts and moves through an activated ground state via a chair-

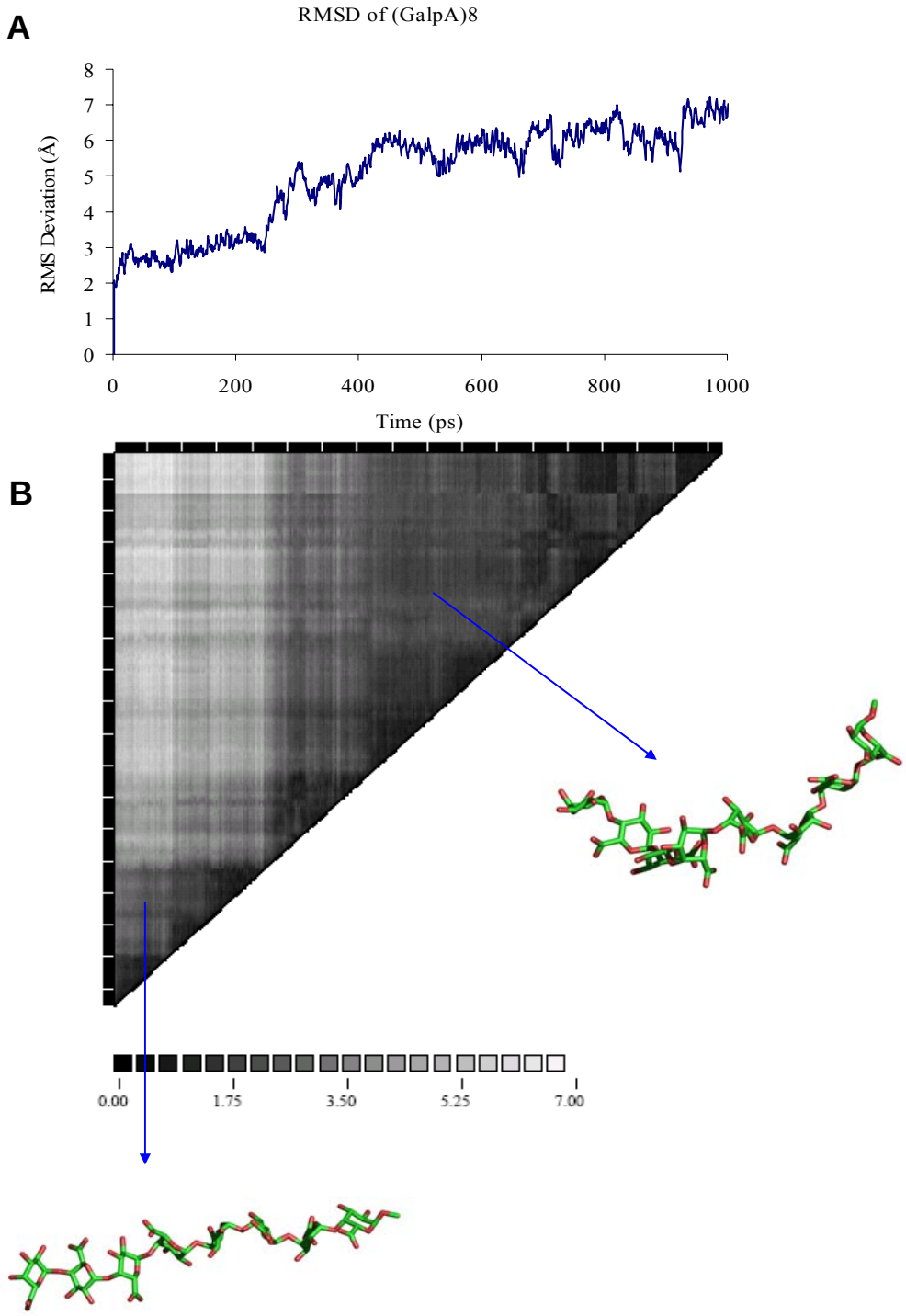

Figure 3.2. **(A) 1D and (B) 2D RMSD of the $(GalpA)_8$ substrate.**

to-boat mechanism. Because it is theorized that the substrate for this enzyme (as well as other hydrolytic enzymes) moves through a transition prior to catalysis to produce enough free energy to drive the process into completion, we used our distorted ^{1,4}B transition MD simulation to derive an ensuing sub-site map as shown in Figure 3.3 B. This map is similar to Figure 3.3 A, except that it shows the neighboring residues about the bent and distorted conformation of the substrate. The residues that flank the maps are those that create hydrogen bond networks or charge-charge interactions throughout each of the MD simulations. These hydrogen bond networks were created with the CARNAL module of the AMBER 7 package suite (Case, Pearlman et al. 2002). Working within a 4.0 Å cutoff, the program found all hydrogen bonds within 3.5 Å or lower range, identifying all the hydrogen bond donors and acceptors for each residue interacting in the enzyme-substrate complex. Each interacting residue is shown with its corresponding sub-site (illustrated in Figure 3.3 A & B). The linear model sub-site map (Fig. 3.3 A) is in good agreement with the previous map created by Pages et al., and shows the suggested residues about the site map with new predicted residues that stabilize the enzyme-substrate complex. However, our ensuing map (Fig. 3.3 B) is novel—illustrating new contacts and interactions as the substrate moves through the transition before hydrolysis. Notice the distorted glycosidic ring at the -1 site in Figure 3.3 B. This distortion causes a change in substrate conformation, which requires the EPG-II enzyme to have an adjacent cleft to accommodate a dynamically flexible ligand. Figure 3.4 shows that the enzyme has the ability to bind the substrate in both linear and bent forms. The cleft has two deep pockets to bind the ground-state linear form (Fig 3.4 shown in blue) as well as the activated bent form (shown in red) during hydrolysis of the substrate. Major changes occur in the interactions between the EPG-II and the (Gal*p*A)$_8$ complex once the substrate has moved into the activated bent form. There are twelve

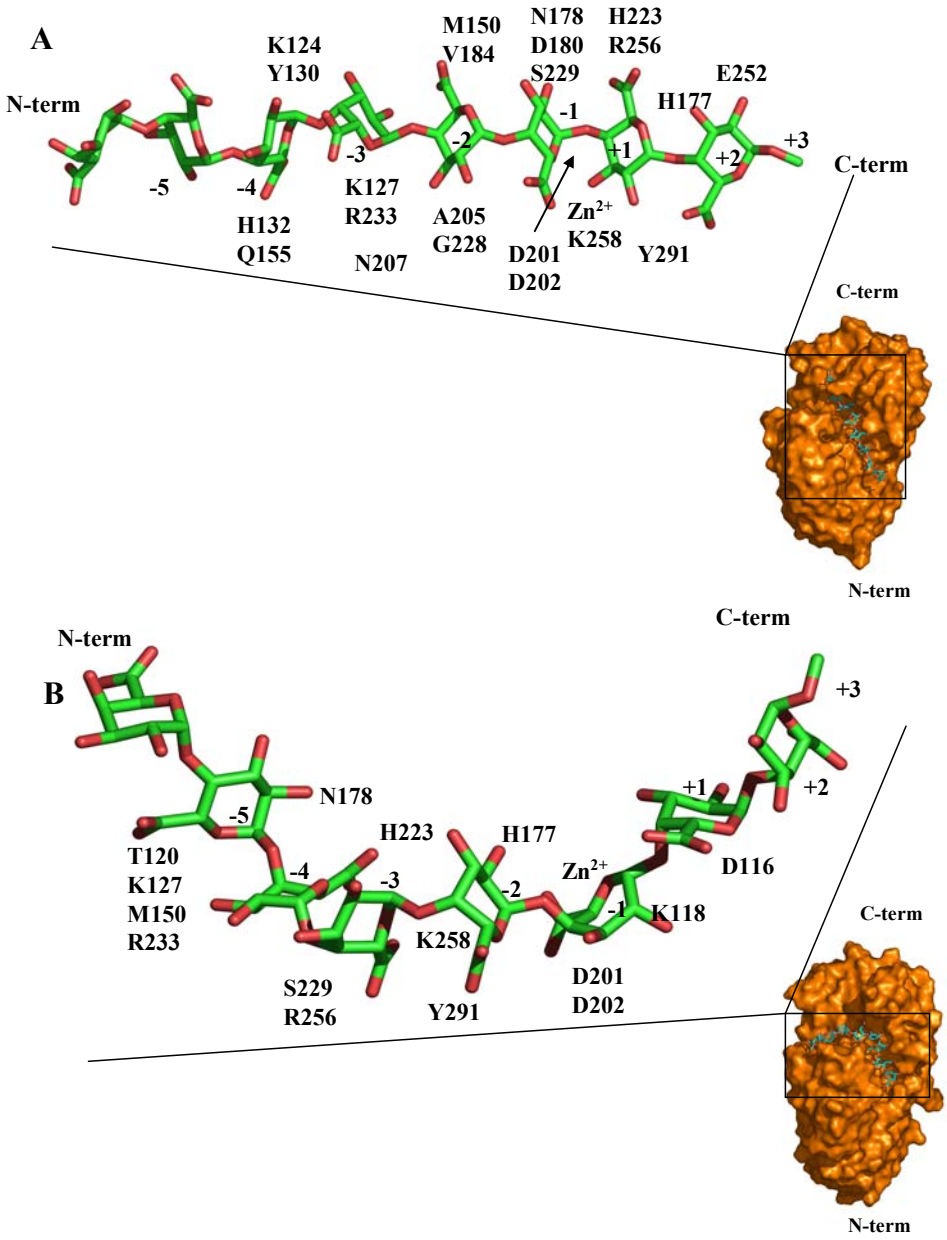

Figure 3.3. **An illustration of the sub-site maps for both the linear and bent conformations of the GalpA substrate**. (**A**) shows the site map for the linear (Gal*p*A)$_8$ substrate. (**B**) shows the site map for the bent (Gal*p*A)$_8$ substrate. Sub-sites are labeled with a numerical value from –5 to +3 together with an octameric fragment of the polygalacturonic acid. Interacting residues with the substrate in their respective sub-site are labeled surrounding the site map. The EPG-II enzyme is shown in orange with the substrate bound in the linear (**A**) and the bent (**B**) configurations. The illustrations were generated using PYMOL (DeLano 2002).

conserved residues that interact with both the linear and the bent form, but most move sub-site location when the substrate partially transverses the primary cleft (Table 3.2 & Figure 3.3 A/B). Also, there are ten distinct residual interactions that are not found in the bent configuration.

Table 3.2

Residues in the binding cleft that are responsible for hydrogen bond interactions.

Residues in Both Linear and Bent Conformations	Residues in Linear Conformation	Residues in Bent Conformation
Lys-127	Lys-124	Asp-116
*Met-150	Tyr-130	Lys-118
His-177	His-132	Thr-120
Asn-178	Gln-155	
*His-223	*Asp-180	
Ser-229	Val-184	
Arg-233	Ala-205	
*Arg-256	Asn-207	
*Lys-258	Gly-228	
*Tyr-291	*Glu-252	
*Asp-201[a]		
*Asp-202[a]		
Zn^{2+} [a]		

[a] Asp-201, 202 & Zn^{2+} are apart of the –1/+1 sub-site in both conformations, but do not show hydrogen bond interactions.
*Denotes previous site-directed mutagenesis studies.
Asp-180, His-223, & Glu-252 were mutated to alanines.

When the transition occurs, the ten distinct interactions in the linear form are disrupted and three new hydrogen bond interactions develop as the bent conformation shifts into the secondary binding site. The shift of the substrate is possibly more important after the hydrolysis than before, because the cleavage of the substrate leaves an intact polygalacturonic acid ligand that is

ready for more cycles of catalysis. This shift may align the substrate to an orientation that allows the linear segment to bind again in the linear sub-sites of the binding cleft and restart the enzymatic cleavage process.

A simple model for Amide exchange MS determined by SASA—After proposing the novel idea of dual modes of binding, we sought to test this theory by comparison with the MS data. The amide-exchange MS study (King, Bergmann et al. 2002) identified direct interaction of the EPG-II enzyme and its substrate (Gal*p*A)$_8$ through a decrease in hydrogen/deuterium exchange in certain residues in the EPG-II enzyme. The results from the amide-exchange experiment

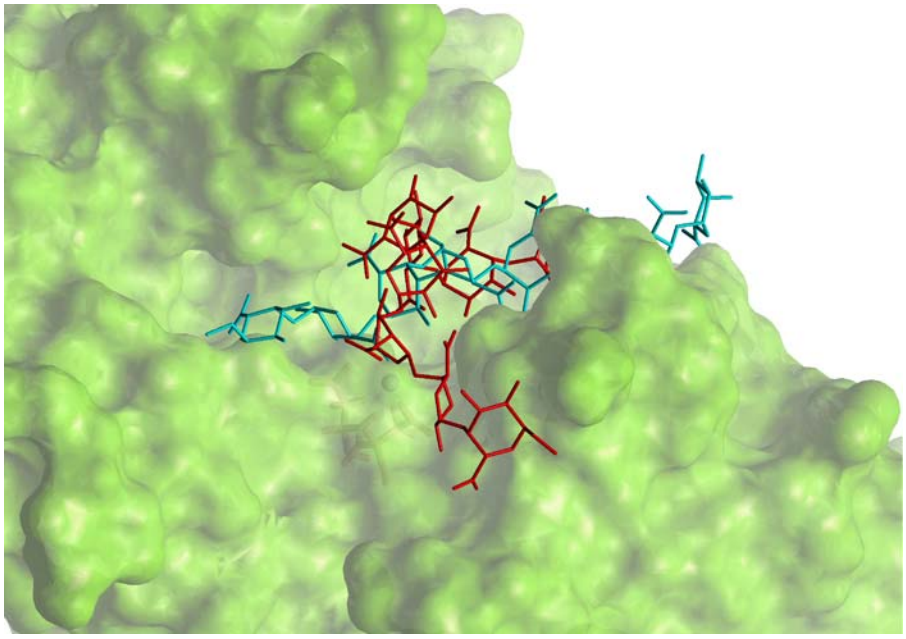

Figure 3.4. **The duality of the EPG-II enzyme binding cleft with both linear and bent conformations of the (GalpA)$_8$ substrate.** The linear conformation is shown in blue and the activated bent form is shown in red. The EPG-II binding cleft is shown in green with the C-terminus facing the viewer. The EPG-II enzyme is illustrated in a Connolly surface generated by INSIGHT (Biosym/MSI 1995).

correlated with local binding regions within the binding cleft observed from earlier point-mutation studies. They also provided some evidence for a conformational change in the enzyme in the presence of the substrate. However, due to the nature of the experiment, the amide exchange studies on EPG-II accounted only for the increase or decrease in deuterium incorporation on a per peptide fragment basis rather than per residue. Moreover, some of the peptide fragments were not accounted for during the experiment, so the extent of deuterium exchange was not identified for the entire EPG surface. Here we propose a simple model for amide exchange MS with a calculation of the solvent accessible surface area (SASA) per residue (Figure 3.5). In this way, we planned to find those areas that were not previously accounted for, and to determine any correlation with the MS data that would support our observation of a dual binding site. The ΔSASA per residue was determined by calculating the difference in solvent accessibility in the absence and presence of the substrate. Figure 3.5 A and B show the difference in molecular shielding upon substrate binding, and the values are given in $Å^3$ for the EPG-II─(Gal*p*A)$_8$ linear (Fig. 3.5 A) and activated bent (Fig. 3.5 B) complexes. Illustrations of the EPG-II─(Gal*p*A)$_8$ complex for both the linear and activated bent forms are shown in Figure 3.6 A and B, while the protected sites observed the amide exchange MS studies proposed by King et al., are shown in Figure 3.6 C. The regions that were shielded in the presence of the bound substrate are displayed (Figure 3.6) and give reference to the shielded solvent accessible surfaces shown in Figure 3.5. The correlation with the amide exchange MS data published by King, D. et al., is also shown in Figure 3.5 and Figure 3.6. The green bars (Figure 3.5) distinguish regions that were studied with the amide exchange MS experiments. The white bars denote peptide sequences that were not calculated or not collected in the amide exchange MS experiment and therefore were not a part of the experimental comparison. Based on peptides that

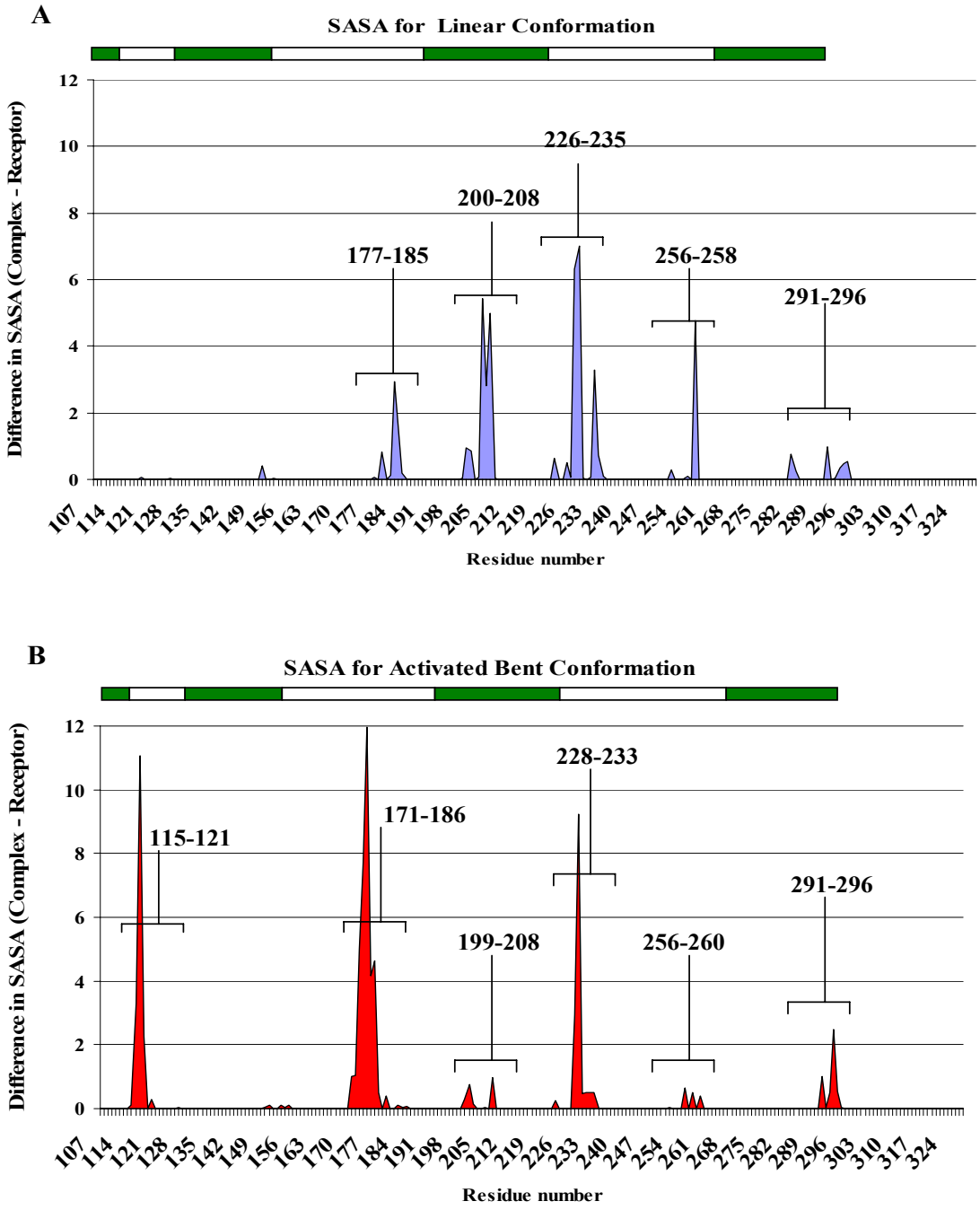

Figure 3.5. **The Solvent Accessible Surface Area simple model for Amide Exchange is shown.** (**A**) illustrates the EPG-II/(Gal*p*A)$_8$ linear conformation with the amount of SASA exposure shown in blue. (**B**) demonstrates the EPG-II/(Gal*p*A)$_8$ activated bent conformation with SASA exposed illustrated in red. Regions before residue 107 and after 324 had no change in SASA, are therefore not shown. The bar graphs show the regions that were previously studied (shown in green) by the amide exchange MS experiments done by King et al. The white bar regions denote areas in the experiment that were (1) not calculated or (2) not collected due to loss of the peptide fragments in the experiment (these regions were given zero values).

were identified, it is not surprising that the binding cleft is more protected from solvent in the presence of the substrate. This agrees with our proposal of a secondary binding cleft in which the linear substrate only partly fills the first binding cleft, but due to its transition into the activated bent form occupies the secondary binding cleft. A α-helical region (residues 110-114; see Figure 3.6 C) that was previously shown to have low levels of deuterium exchange was suggested to shield exchange by its secondary structure (King, Bergmann et al. 2002). Our model determined that residues 115-121 were shielded by the substrate, which is at the end of the helix (Figure 3.5 & 3.6 B). This region is outside the traditional linear binding site is still a part of the binding cleft, but in fact resides in the secondary site and interacts with the bent form of the substrate. The discrepancies between the MS results and our SASA results are in the collection of data at the peptide fragment level rather than on a per residue basis.

We also show that a α-helix and a β-strand (residues 171-186) are shielded by the presence of the bent substrate (Figure 3.5 B and 3.6 B), but this region has not been determined experimentally with the amide exchange studies. However, we saw similar shielding results in the linear complex, in which the B-strand (residues 177,178 and 180-185) was protected from the substrate. A detailed experiment of this α-helix and β-strand region would validate our observations. Still, the simple model for amide exchange suggests residues that are important in both the linear and the activated bent forms of the enzyme-substrate complex and determines regions that are similar between the simple model and the MS data, such as residues 202-208 and 291-296 (Figure 3.5 and 3.6). These regions are protected in the presence of the substrate. On the other hand, the peptide fragment 214-227 was observed to be protected in the MS experiment, which was not shown in the SASA model. However, residues 225-233 were protected by the substrate in the SASA experiment, which overlap with the observed protected

peptide fragment. A more detailed analysis of the residues after 227-279 would help to validate these observations, because this region was not determined for the amide exchange MS experiment.

Regions that were not shielded regardless of the presence of the substrate were also observed for both experiments. Residues 135-156, 194-201, and 270-280 were exposed to solvent in the presence of the substrate for SASA model. This agrees with the previously published MS data, which showed exposure to deuterium incorporation in the presence and absence of the ocatmeric substrate.

Each MD simulation (for the linear and activated bent complexes) was essential in interpreting the MS data, because areas that were shielded from deuterium exchange when the substrate was bound were indicative of both MD simulations. Since the amide exchange MS experiments were done with a mutant form of the EPG-II enzyme (D201E), catalysis was prohibited and the substrate was only allowed to bind and interact. The exact interpretation of the orientation of the substrate in the binding site remains unknown, but here we suggest that the (Gal*p*A)$_8$ substrate binds linearly and moves through the activated bent state via chair-to-distorted ^{1,4}B (at the cleavage site) and occupies a new binding cleft which allows for the exiting of the cleaved galacturonic acid residues.

King et al. observed some regions that were protected from deuterium incorporation that were not validated by the SASA calculations (Figure 3.5). Others were shown to have molecular shielding from the substrate in the SASA model, but were not illustrated in the amide exchange MS studies (Figure 3.5). A number of the discrepancies between the MS data and the SASA calculations were due to the simplicity of our model, i.e. differences in local pH were ignored, all amides were considered equally labile and the exchange was assumed to be proportional only to

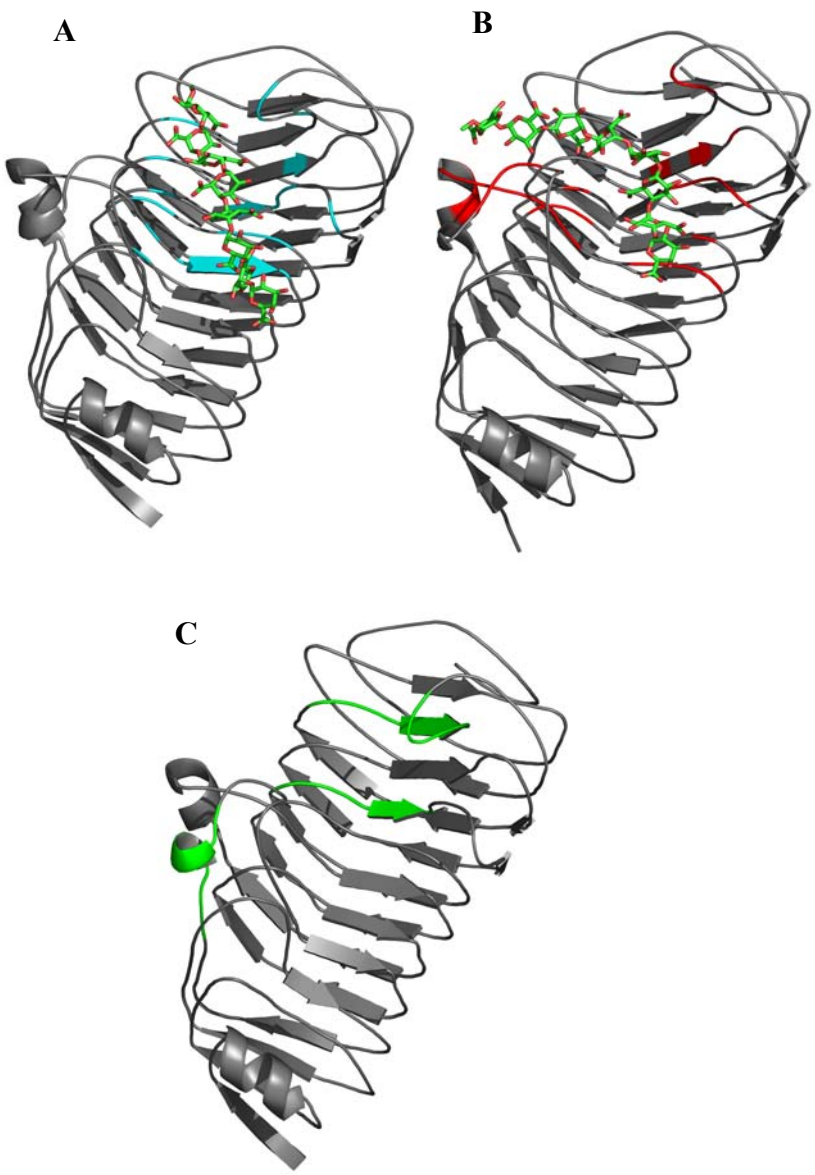

Figure 3.6. **The regions that are shielded by (Gal*p*A)$_8$ substrate from solvent accessibility in the (A) dynamic linear and the (B) activated bent complexes.** The bottom denotes the N-terminus and the top the C-terminus. The EPG-II enzyme is shown in grey cartoon and the (Gal*p*A)$_8$ is shown is CPK generated by PYMOL (DeLano 2002). In (**A**), the cyan illustrates the molecular shielding in the enzyme from the substrate for the linear conformation. In (**B**), the red shows the shielded regions from the activated bent configuration. (**C**) green shows the protected regions shown by King, et al. from MS data.

SASA. A separate unbound simulation of the enzyme was not considered. Therefore, the same coordinate set for both the bound and unbound enzyme was used to predict regions that were in contact with the substrate, and a conformational change analysis between bound and unbound states was not provided. Only regions that were protected by the substrate were considered for the simple SASA model.

Experimental ambiguity was also responsible for some discrepancies, particularly for those peptide sequences that were not observed in the MS data. The SASA calculations were able to measure the relative accessibilities of solvent to the surface area of the backbone. These calculations measured the amount of surface area exposed (from the residual backbone) for solvent interaction. The simple model allowed for a detailed observation of those residues most likely to exchange protons based on surface area exposure. The MS experiment, however, determined exchange based on levels of deuterium incorporation into peptide fragments (digested with pepsin) in the presence and absence of the substrate. The data interpretation was based on a change in the mass of the peptide fragment versus the non-deuterated peptide fragments. The rate of exchange was not calculated, but a percent incorporation per fragment was determined. A few fragments that were lost in the experiment were given a zero percent incorporation value, which did not suggest that exchange or protection occurred, only that they were not found in the collection of experimental results. Moreover, the experimental methods compliment each other, proving that MD simulations and experimental methodology can benefit mutually while providing data that is relevant to the binding interactions of the EPG-II enzyme with its octameric Gal*p*A substrate.

Alanine scanning as an approach to determining binding interaction and stability—A computational alanine scanning approach was used to detect residues that were important in the

overall binding of the enzyme substrate complex. In this study, we investigated the contributions of each charged residue, plus tyrosines and methionines when mutated to binding relative to their replacement by alanine. A generalized Born analysis (MM-GBSA) was performed on both the linear and activated bent forms of the complex, giving rise to total binding energy estimates of –86.4 and –83.4 kcal/mol, respectively. These numbers are not quantitative as they fail to include the large entropic penalty associated with ligand binding, however the relative $\Delta\Delta G_{binding}$ (wild type – mutant) is useful. The site-directed mutagenesis studies (see Table 3.2 for mutants studied) determined previously by Pages, et al (2000) showed that Arg-256, Lys-258, and Tyr-291 had dramatic effects on the binding stability of the substrate to the EPG-II enzyme in the mutagenesis studies. Our alanine scanning data confirms that these residues are important for the stability of the complex in both the linear and bent forms of the complex shown in Figure 3.7 A and B.

Computational mutation of positively charged residues Lys-127, His-132, His-223, and Arg-233 to alanine was shown to have unfavorable effects for in both the linear and the bent configurations of the complex (Figure 3.7). These trends were only observed for the portion of the substrate that maintained a similar conformation in the binding cleft before and after the transition to the activated state. Other residues showed specificity in either the linear or the bent conformation of the enzyme-substrate complex. In the linear complex, mutations K124A was shown to have an unfavorable effect on binding when mutated, while residues K118A and H177A were shown to effect binding unfavorably in the activated bent configuration.

Most of the mutants introduced in the alanine scanning experiment, especially those associated with charged residues, altered the binding of the anionic substrate as expected. The mutation of the negatively charged residues generally had a more favorable effect on binding,

while the replacement of positively charged residues with alanine generally destabilized the complex. However, mutations D180A and E252A were shown to have opposite effects on binding (Figure 3.7 A) than predicted in the linear complex. These two residues were shown to make hydrogen bond interactions with the linear (Gal*p*A)$_8$ substrate during the course of the MD simulation. Asp180 Oδ1-Oδ2 atoms (from the carboxyl group) made hydrogen bond interactions with the polygalacturonic residue's H3O atom at an average distance of 3.2 Å located in the –1 sub-site (Figure 3.3A). On the other hand, the Glu252 Oε1-Oε2 atoms (from the carboxyl group) made hydrogen bond interactions with the H2O and H3O atoms from the polygalacturonic residue located in the +2 sub-site (Figure 3.3A). These H-bond interactions were not found in the MD simulation of the activated bent configuration, instead charge repulsions were seen between the carboxylic acid side chains of the substrate and the Asp180 and Glu252 residues.

Further, in the EPG-II─(Gal*p*A)$_8$ activated bent complex, the D201A mutant showed unfavorable binding in contrast to that seen in the linear complex. This was due to charge repulsions between the carboxylic acid groups of the polygalacturonic acid group located in the –1 sub-site and the side chain carboxylic acid, and was similar to the effects seen for Asp180 and Glu225. However, the reduced affinity for the substrate in the bent form was due to the salt bridge between the polygalacturonic acid carboxylic acid group at the –1 sub-site (Figure 3.3 B), a zinc ion, and the carboxylic acid side chain of the Asp201. Once the aspartic acid was mutated to an alanine, the bridge was disrupted and the binding between the enzyme and substrate became less stable.

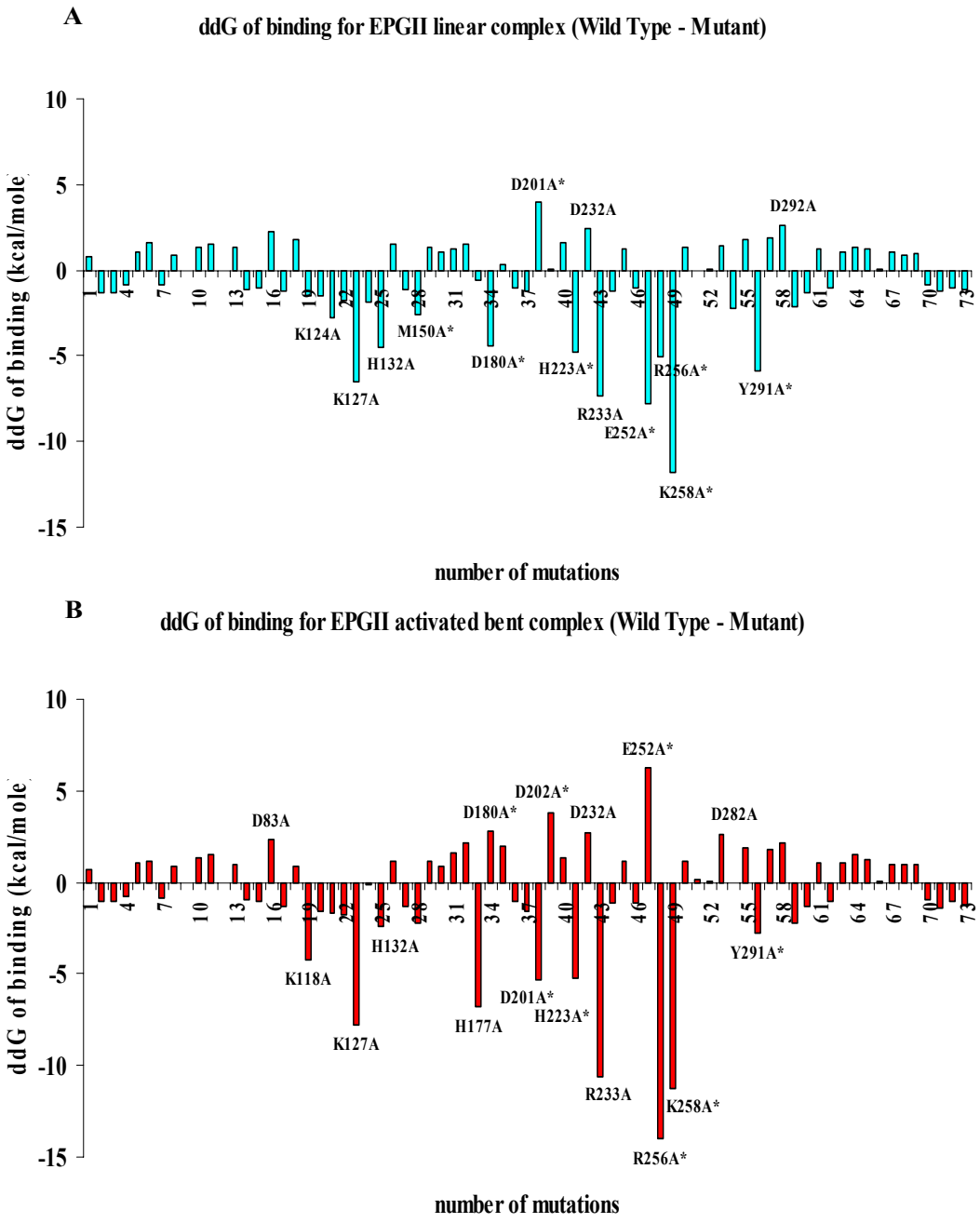

Figure 3.7. **The $\Delta\Delta G_{binding}$ (Wild type – Mutant) for both the dynamic linear and activated bent conformations are given.** Graph (**A**) shows the mutations for the EPG-II─(Gal*p*A)$_8$ dynamic linear configuration and graph (**B**) illustrates the EPG-II─(Gal*p*A)$_8$ activated bent configuration. Each charged residue plus methionines and tyrosines were mutated to an alanine and the $\Delta G_{binding}$ was calculated. The mutations are shown from N-terminus to C-terminus (left to right) in the graph. An asterisk denotes previous studies with site-directed mutagenesis.

regions of the protein surface from deuterium exchange, while the absence of the substrate allows the amide hydrogens to exchange more readily with the deuterium solvent. A change in the amount of deuterium incorporation between the unbound and bound systems identifies areas of importance to the formation of the EPG-II—$(GalpA)_8$ complex. Some regions affected in the presence of the substrate, however, were difficult to interpret due to their location outside the binding region.

Here we presented a detailed model of the EPG-II—$(GalpA)_8$ complex using MD simulations. We combined flexible docking in addition to explicit (solvated MD simulation) and continuum solvent models to offer a structural interpretation for the experimental data provided by site-directed mutagenesis and amide-exchange MS studies. We showed that proper ionization of the EPG-II enzyme through histidine protonation was necessary to keep the enzyme-substrate complex intact over the course of the simulation. After analysis of our initial MD simulation with 2D/1D RMSD, the model was divided into two MD simulations based on an activated chair-to-distorted $^{1,4}B$ transition observed in the MD simulation. The dynamic linear and activated bent MD simulations were subjected to hydrogen bond analysis where a novel cleavage pathway was proposed for the EPG-II enzyme. The hydrogen bond analysis was used to create the secondary binding site-map, which was based on the previous map determined by Pages, et al (2000).

The dynamic linear and activated bent MD simulations were used to construct a simple model for amide-exchange MS based on a SASA per residue calculation with MM-GBSA. The model validated the previous work done by King et al., while it answered questions about the interpretation of the protected regions outside the binding cleft with the proposed activated bent model, which suggests the presence of a secondary binding cleft. We further validated our

model with computational alanine scanning studies, which predicted the residues that were essential for the binding interactions of the EPG-II—$(GalpA)_8$ complex. The point mutation studies showed that Arg-256, Lys-258 and Tyr-291 dramatically affected the binding stability when they were mutated, consistent with the data from the site-directed mutagenesis studies. Lys-127, His-132, His-223, and Arg-233 affected binding when mutated in both the dynamic linear and activated bent conformations, while Lys-124 only affected the dynamic linear and Lys-118 and His-177 only affected the activated bent configuration. These positively charged residues were important in the binding of the ligand because of its negative charge. The negatively charged residues that had an unfavorable effect on binding when mutated such as D180A and E252A for the linear complex were due to hydrogen bond interactions with the $(GalpA)_8$ substrate. For the activated bent complex the mutant D201A showed unfavorable binding due to a salt bridge that was created between a galacturonic acid residue, a zinc ion, and the carboxylic acid side-chain from Asp-201.

We do not know the extent of the effect that removal of counterions from a charged system will have on the ΔG of binding. Future work should be done to investigate these possible effects, especially when considering charged carbohydrates with ions that may be important in binding interactions, neutralizing systems and/or stabilizing transition intermediates. The work presented here is a foundation for more site-directed mutagenesis studies in the primary binding cleft as well as the secondary binding cleft. Also, studies should be done to investigate the catalytic mechanism in more detail to examine the chair-to-distorted $^{1,4}B$ transition and to establish an intermediate for the EPG-II hydrolytic mechanism.

Carpita, N. C. and D. M. Gibeaut (1993). "Structural Models of Primary-Cell Walls in Flowering Plants - Consistency of Molecular-Structure with the Physical-Properties of the Walls During Growth." Plant Journal **3**(1): 1-30.

Case, D. A., D. A. Pearlman, et al. (2002). AMBER 7.0. San Francisco, University of California.

Cave, I. D. (1969). "Longitudinal Youngs Modulus of Pinus Radiata." Wood Science and Technology **3**(1): 40-&.

Centis, S., I. Guillas, et al. (1997). "Endopolygalacturonase genes from Colletotrichum lindemuthianum: Cloning of CLPG2 and comparison of its expression to that of CLPG1 during saprophytic and parasitic growth of the fungus." Molecular Plant-Microbe Interactions **10**(6): 769-775.

Cho, S. W., S. Lee, et al. (2001). "The X-ray structure of Aspergillus aculeatus polygalacturonase and a modeled structure of the polygalacturonase-octagalacturonate complex." Journal of Molecular Biology **311**(4): 863-878.

Choi, J. K., B. H. Lee, et al. (2004). "Computer modeling of the rhamnogalacturonase - "Hairy" pectin complex." Proteins-Structure Function and Bioinformatics **55**(1): 22-33.

Chong, L., Y. Duan, et al. (1999). "Molecular Dynamics and Free energy calculations applied to affinity maturation in anti-body 48G7." Proc. Natl. Acad. Sci. USA **96**: 14330-14335.

Cook, B. J., R. P. Clay, et al. (1999). "Fungal polygalacturonases exhibit different substrate degradation patterns and differ in their susceptibilities to polygalacturonase-inhibiting proteins." Molecular Plant-Microbe Interactions **12**(8): 703-711.

Cooper, R. M. (1995). The mechanisms and significance of enzymic degradation of host cell walls by parasites. Biochemical Plant Pathology. e. J.A. Callow, John Wiley and Sons Ltd**:** 135.

Darden, T., D. York, et al. (1993). "Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems." Journal of Chemical Physics **98**(12): 10089-10092.

Davies, G. and B. Henrissat (1995). "Structures and Mechanisms of Glycosyl Hydrolases." Structure **3**(9): 853-859.

Davies, G. J., K. S. Wilson, et al. (1997). "Nomenclature for sugar-binding subsites in glycosyl hydrolases." Biochemical Journal **321**: 557-559.

de Vries, R. P. and J. Visser (2001). "Aspergillus enzymes involved in degradation of plant cell wall polysaccharides." Microbiology and Molecular Biology Reviews **65**(4): 497-+.

DeLano, W. L. (2002). The PyMOL Molecular Graphics System.

DiPietro, A. and M. I. G. Roncero (1996). "Endopolygalacturonase from Fusarium oxysporum f sp lycopersici: Purification, characterization, and production during infection of tomato plants." Phytopathology **86**(12): 1324-1330.

Ehring, H. (1999). "Hydrogen exchange electrospray ionization mass spectrometry studies of structural features of proteins and protein/protein interactions." Analytical Biochemistry **267**(2): 252-259.

Federici, L., C. Caprari, et al. (2001). "Structural requirements of endopolygalacturonase for the interaction with PGIP (polygalacturonase-inhibiting protein)." Proceedings of the National Academy of Sciences of the United States of America **98**(23): 13425-13430.

Federici, L., B. Mattei, et al. (1999). "Crystallization and preliminary X-ray diffraction study of the endo-polygalacturonase from Fusarium moniliforme." Acta Crystallographica Section D-Biological Crystallography **55**: 1359-1361.

Feig, M., A. Onufriev, et al. (2004). "Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures." Journal of Computational Chemistry **25**(2): 265-284.

Ford, M. G., T. Weimar, et al. (2003). "Molecular dynamics simulations of galectin-1-oligosaccharide complexes reveal the molecular basis for ligand diversity." Proteins-Structure Function and Genetics **53**(2): 229-240.

Gacesa, P. (1987). "Alginate-Modifying Enzymes - a Proposed Unified Mechanism of Action for the Lyases and Epimerases." Febs Letters **212**(2): 199-202.

Gilson, M. K. and B. Honig (1988). "Calculation of the Total Electrostatic Energy of a Macromolecular System - Solvation Energies, Binding-Energies, and Conformational-Analysis." Proteins-Structure Function and Genetics **4**(1): 7-18.

Gohlke, H. and D. A. Case (2004). "Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf." Journal of Computational Chemistry **25**(2): 238-250.

Gohlke, H. and G. Klebe (2001). "Statistical potentials and scoring functions applied to protein-ligand binding." Current Opinion in Structural Biology **11**(2): 231-235.

Goodsell, D. S., G. M. Morris, et al. (1996). "Automated docking of flexible ligands: Applications of AutoDock." Journal of Molecular Recognition **9**(1): 1-5.

Grassin, C. and P. Fauquembergue (1996). Application of pectinases in beverages. Pectins and Pectinases. J. Visser and A. G. J. Voragen, Elsevier Science B. V.**:** 453-462.

Heldt-Hansen, H. P., L. V. Kofod, et al. (1996). Application of Tailor-made Pectinases. Pectins and Pectinases. J. Visser and A. G. J. Voragen, Elsevier Science B. V.**:** 463-474.

Henrissat, B. (1991). "A Classification of Glycosyl Hydrolases Based on Amino-Acid-Sequence Similarities." Biochemical Journal **280**: 309-316.

Henrissat, B., P. M. Coutinho, et al. (2001). "A census of carbohydrate-active enzymes in the genome of Arabidopsis thaliana." Plant Molecular Biology **47**(1-2): 55-72.

Herron, S. R., J. A. E. Benen, et al. (2000). "Structure and function of pectic enzymes: Virulence factors of plant pathogens." Proceedings of the National Academy of Sciences of the United States of America **97**(16): 8762-8769.

Honig, B. and A. Nicholls (1995). "Classical Electrostatics in Biology and Chemistry." Science **268**(5214): 1144-1149.

Huang, W. J., A. Matte, et al. (1999). "Crystal structure of chondroitinase B from Flavobacterium heparinum and its complex with a disaccharide product at 1.7 angstrom resolution." Journal of Molecular Biology **294**(5): 1257-1269.

Huo, S., I. Massova, et al. (2002). "Computational alanine scanning of the 1 : 1 human growth hormone-receptor complex." Journal of Computational Chemistry **23**(1): 15-27.

Ibragimova, G. T. and R. C. Wade (1998). "Importance of explicit salt ions for protein stability in molecular dynamics simulation." Biophysical Journal **74**(6): 2906-2911.

Jayaram, B., D. Sprous, et al. (1998). "Solvation free energy of biomacromolecules: Parameters for a modified generalized born model consistent with the AMBER force field." Journal of Physical Chemistry B **102**(47): 9571-9576.

Jenkins, J., O. Mayans, et al. (1998). "Structure and evolution of parallel beta-helix proteins." Journal of Structural Biology **122**(1-2): 236-246.

Jones, T. M., Albershe.P, et al. (1972). "Host-Pathogen Interactions .4. Studies on Polysaccharide-Degrading Enzymes Secreted by Fusarium Oxysporum F Sp Lycopersici." Physiological Plant Pathology **2**(2): 153-&.

Jones, T. M., P. Albersheim, et al. (1972). "Host-Pathogen Interactions .4. Studies on Polysaccharide-Degrading Enzymes Secreted by Fusarium Oxysporum F Sp Lycopersici." Physiological Plant Pathology **2**(2): 153-&.

Jorgensen, W. L. and C. Ravimohan (1985). "Monte-Carlo Simulation of Differences in Free-Energies of Hydration." Journal of Chemical Physics **83**(6): 3050-3054.

Karr, A. L. and P. Albersheim (1970). "Polysaccharide-Degrading Enzymes Are Unable to Attack Plant Cell Walls without Prior Action by a Wall-Modifying-Enzyme." Plant Physiology **46**(1): 69-&.

Katta, V. and B. T. Chait (1993). "Hydrogen-Deuterium Exchange Electrospray-Ionization Mass-Spectrometry - a Method for Probing Protein Conformational-Changes in Solution." Journal of the American Chemical Society **115**(14): 6317-6321.

King, D., C. Bergmann, et al. (2002). "Use of amide exchange mass spectrometry to study conformational changes within the endopolygalacturonase II-homogalacturonan-polygalacturonase inhibiting protein system." Biochemistry **41**(32): 10225-10233.

King, D., M. Lumpkin, et al. (2002). "Studying protein-carbohydrate interactions by amide hydrogen/deuterium exchange mass spectrometry." Rapid Communications in Mass Spectrometry **16**(16): 1569-1574.

Kirschner, K. N. and R. J. Woods (2001). "Solvent interactions determine carbohydrate conformation." Proceedings of the National Academy of Sciences of the United States of America **98**(19): 10541-10545.

Kollman, P. (1993). "Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena." Chemical Reviews **93**(7): 2395-2417.

Kollman, P. A., I. Massova, et al. (2000). "Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models." Accounts of Chemical Research **33**(12): 889-897.

Kuhn, B. and P. A. Kollman (2000). "Binding of a diverse set of ligands to avidin and streptavidin: An accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models." Journal of Medicinal Chemistry **43**(20): 3786-3791.

Lietzke, S. E., R. D. Scavetta, et al. (1996). "The refined three-dimensional structure of pectate Lyase E from Erwinia chrysanthemi at 2.2 angstrom resolution." Plant Physiology **111**(1): 73-92.

Markovic, O. and S. Janecek (2001). "Pectin degrading glycoside hydrolases of family 28: sequence-structural features, specificities and evolution." Protein Engineering **14**(9): 615-631.

Massova, I. and P. A. Kollman (1999). "Computational alanine scanning to probe protein-protein interactions: A novel approach to evaluate binding free energies." Journal of the American Chemical Society **121**(36): 8133-8143.

Massova, I. and P. A. Kollman (2000). "Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding." Perspectives in Drug Discovery and Design **18**: 113-135.

Mayans, O., M. Scott, et al. (1997). "Two crystal structures of pectin lyase A from Aspergillus reveal a pH driven conformational change and striking divergence in the substrate-binding clefts of pectin and pectate lyases." Structure **5**(5): 677-689.

Michel, G., K. Pojasek, et al. (2004). "The structure of chondroitin B lyase complexed with glycosaminoglycan oligosaccharides unravels a calcium-dependent catalytic machinery." Journal of Biological Chemistry **279**(31): 32882-32896.

Morris, G. M., D. S. Goodsell, et al. (1998). "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function." Journal of Computational Chemistry **19**(14): 1639-1662.

Nicholls, A., K. A. Sharp, et al. (1991). "Protein Folding and Association - Insights from the Interfacial and Thermodynamic Properties of Hydrocarbons." Proteins-Structure Function and Genetics **11**(4): 281-296.

Pages, S., W. H. M. Heijne, et al. (2000). "Subsite mapping of Aspergillus niger endopolygalacturonase II by site-directed mutagenesis." Journal of Biological Chemistry **275**(38): 29348-29353.

Pages, S., H. C. M. Kester, et al. (2001). "Changing a single amino acid residue switches processive and non-processive behavior of Aspergillus niger endopolygalacturonase I and II." Journal of Biological Chemistry **276**(36): 33652-33656.

Parenicova, L., J. A. E. Benen, et al. (1998). "pgaE encodes a fourth member of the endopolygalacturonase gene family from Aspergillus niger." European Journal of Biochemistry **251**(1-2): 72-80.

Parenicova, L., J. A. E. Benen, et al. (2000). "pgaA and pgaB encode two constitutively expressed endopolygalacturonases of Aspergillus niger." Biochemical Journal **345**: 637-644.

Petersen, T. N., S. Kauppinen, et al. (1997). "The crystal structure of rhamnogalacturonase A from Aspergillus aculeatus: A right-handed parallel beta helix." Structure **5**(4): 533-544.

Pickersgill, R., J. Jenkins, et al. (1994). "The structure of Bacillus subtilis pectate lyase in complex with calcium." Nature Structural Biology **1**: 717-723.

Pickersgill, R., D. Smith, et al. (1998). "Crystal structure of polygalacturonase from Erwinia carotovora ssp. carotovora." Journal of Biological Chemistry **273**(38): 24660-24664.

Ridley, B. L., M. A. O'Neill, et al. (2001). "Pectins: structure, biosynthesis, and oligogalacturonide-related signaling." Phytochemistry **57**(6): 929-967.

Ros, J. M., H. A. Schols, et al. (1996). "Extraction, characterisation, and enzymatic degradation of lemon peel pectins." Carbohydrate Research **282**(2): 271-284.

Sanner, M. F., A. J. Olson, et al. (1996). "Reduced surface: An efficient way to compute molecular surfaces." Biopolymers **38**(3): 305-320.

Scavetta, R. D., S. R. Herron, et al. (1999). "Structure of a plant cell wall fragment complexed to pectate lyase C." Plant Cell **11**(6): 1081-1092.

Shea, E. M. and R. D. Hatfield (1993). "Characterization of a Pectic Fraction from Smooth Bromegrass Cell-Walls Using an Endopolygalacturonase." Journal of Agricultural and Food Chemistry **41**(3): 380-387.

Shimizu, T., T. Nakatsu, et al. (2002). "Active-site architecture of endopolygalacturonase I from Stereum purpureum revealed by crystal structures in native and ligand-bound forms at atomic resolution." Biochemistry **41**(21): 6651-6659.

Sitkoff, D., K. A. Sharp, et al. (1994). "Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models." Journal of Physical Chemistry **98**(7): 1978-1988.

Smith, J. B., Y. Q. Liu, et al. (1996). "Identification of possible regions of chaperone activity in lens alpha-crystallin." Experimental Eye Research **63**(1): 125-127.

Srinivasan, J., T. E. Cheatham, et al. (1998). "Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices." Journal of the American Chemical Society **120**(37): 9401-9409.

Steinbacher, S., U. Baxa, et al. (1996). "Crystal Structure of Phage P22 Tailspike Protein Complexed with *Salmonella* Sp. O-Antigen Receptors." Proc. Natl. Acad. Sci. **93**: 10584-10588.

Sutherland, I. W. (1995). "Polysaccharide Lyases." Fems Microbiology Reviews **16**(4): 323-347.

van Pouderoyen, G., H. J. Snijder, et al. (2003). "Structural insights into the processivity of endopolygalacturonase I from Aspergillus niger." Febs Letters **554**(3): 462-466.

van Santen, Y., J. A. E. Benen, et al. (1999). "1.68-angstrom crystal structure of endopolygalacturonase II from Aspergillus niger and identification of active site residues by site-directed mutagenesis." Journal of Biological Chemistry **274**(43): 30474-30480.

Vitali, J., B. Schick, et al. (1998). "The three-dimensional structure of Aspergillus niger pectin lyase B at 1.7-angstrom resolution." Plant Physiology **116**(1): 69-80.

Wang, F., J. S. Blanchard, et al. (1997). "Hydrogen exchange electrospray ionization mass spectrometry studies of substrate and inhibitor binding and conformational changes of Escherichia coli dihydrodipicolinate reductase." Biochemistry **36**(13): 3755-3759.

Warshel, A., F. Sussman, et al. (1986). "Free energy changes in solvated proteins: microscopic calculations using a reversible charging process." Biochemistry **25**: 8368-8372.

Weiser, J., P. S. Shenkin, et al. (1999). "Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO)." Journal of Computational Chemistry **20**(2): 217-230.

Wong, C. F. and J. A. McCammon (1986). "Dynamics and Design of Enzymes and Inhibitors." American Chemical Society **108**: 3830-3832.

Woods, R. J., C. J. Edge, et al. (1993). GLYCAM_93: A Generalized Parameter Set for Molecular Dynamics Simulations of Glycoproteins and Oligosaccharides. Application to the Structure and Dynamics of a Disaccharide Related to Oligomannose. Complex Carbohydrates in Drug Research. K. Bock, H. Clausen, P. Krogsgaard-Larsen and H. Kofod. Copenhagen, Denmark, Munksgaard. **36:** 15-36.

Yoder, M. D., N. T. Keen, et al. (1993). "New Domain Motif - the Structure of Pectate Lyase-C, a Secreted Plant Virulence Factor." Science **260**(5113): 1503-1507.

Zhang, Z. Q. and D. L. Smith (1993). "Determination of Amide Hydrogen-Exchange by Mass-Spectrometry - a New Tool for Protein-Structure Elucidation." Protein Science **2**(4): 522-531.

Zhou, R. H. (2003). "Free energy landscape of protein folding in water: Explicit vs. implicit solvent." Proteins-Structure Function and Genetics **53**(2): 148-161.