COMPUTATIONAL INSIGHT INTO THE GLYCOSIDASE ENZYME USING BIOINFORMATICS TOOLS AND MOLECULAR DYNAMIC SIMULATIONS

MAHASIN EBRAHIM

(BPharm)

A thesis submitted to the school of Health Science, University of KwaZulu-Natal, Westville, in fulfilment of the degree of Masters of Pharmacy

Supervisor

Dr. Mahmoud Soliman

2013

Durban

Computational insight into Glycosidase enzymes using bioinformatics tools and molecular dynamic simulations

Mahaseen Ebrahim

2013

A thesis submitted to the School of Pharmacy and Pharmacology, Faculty of Health Science, University of KwaZulu-Natal, Westville, for the degree of Pharmacy.

This is the thesis in which the chapters are written as a set of discrete research publications, with an overall introduction and final summary. Typically these chapters will have been published in internationally recognized, peer-reviewed journals.

This is to certify that the contents of this thesis are the original research work of Mrs. Mahaseen Ebrahim.

As the candidate's supervisor, I have approved this thesis for submission.

Supervisor:

Signed: ---------- Name: **Dr. Mahmoud Soliman** Date: ----------

Abstract

One of the subjects focused on by theoretical and computational studies is that of the molecular systems. Molecular dynamic simulations can provide useful insights into the dynamic, thermodynamic and structural properties of the system. It is also a convenient way to study enzyme reactions.

The glycoside hydrolase enzyme xylanase, a family GH-11 enzyme, hydrolyses xylan with a net retention of anomeric configuration via a double displacement mechanism. Catalytic residues Glu78 acts as a nucleophille and Glu172 as the general acid/base. In its catalytic cycle, the enzyme passes through different phases, three in particular: the free enzyme phase, the covalent enzyme-substrate complex and the non-covalent substrate in the enzyme pocket. All-atom molecular mechanic MD simulations were performed on the enzyme at each of the above phases for a total of 10ns, to study the effect of the substrate binding on the configuration and movement of the thumb-finger region. Various metrics were implemented to estimate the dynamics of this movement. The study has shown that the covalently bound substrate shows a shortening of the thumb-finger distance, assumedly to hold the substrate in the cleft for the reaction to occur, compared to the non-covalent system in which the distance widens at the beginning of the pathway to allow the substrate into the active site cleft and at the end of the process to expel the substrate.

The enzyme was also investigated using binding free energy calculations, MMGBSA and MMPBSA, on 2ns simulation at all-atom MM level as well as AM1 level of description of the QM/MM simulation. Calculations were performed on the wild-type, Glu78Asp, Glu172Asp, Tyr69Phe, Tyr80Phe and Arg112Asn mutants. Results show that the catalytic residues Glu78 and Glu172 do not play a role in substrate binding yet they are crucial for the catalytic mechanism. The Tyr69 residue however, plays an important role in binding, since it is the hydrogen bond that stabilizes the boat conformation that is essential to reach TS. Tyr80 and Arg112 similarly play a role in binding. Future calculations that are in progress are QM/MM simulations at PM3, DFTB and RM1 level of description to verify the accuracy of each method.

The information from this study sheds further light on factors that affect the catalytic process of the glycoside hydrolase enzyme such as the thumb-finger dynamics as well as the role played by some key residues in substrate binding, and can be used to aid the design of potential inhibitors.

Declaration 1- Plagiarism

I, Mahaseen Ebrahim, declare that

1. The research reported in this thesis, except where otherwise indicated, and is my original research.

2. This thesis has not been submitted for any degree or examination at any other university.

3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted,then:

a. Their words have been re-written but the general information attributed to them has been referenced.

b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.

5. This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the references section.

Signed----------

Declaration 2- Publications

Ebrahim,M., Skelton,A, Mhlongo,N.N, Kruger,H.G., Williams,I.H. and. Soliman,M.E.S. (2013), Insight into the conformational changes of glycosidases during the catalytic pathway: molecular dynamics simulations on GH- 11 xylanase.

Contribution:

Ebrahim, M: contributed to the project by performing all the computational work, data analysis and interpretation as well as manuscript writing.

Skelton A: assisted with the interpretation of data and writing of the manuscript.

Mhlongo N N: assisted with the interpretation of data.

Kruger H G and Williams I H (collaborators).

Soliman M E S: (Principal investigator).

Research Output

Publications

Ebrahim,M., Skelton,A, Mhlongo,N.N, Kruger,H.G., Williams,I.H. and. Soliman,M.E.S. (2013), Insight into the conformational changes of glycosidases during the catalytic pathway: molecular dynamics simulations on GH- 11 xylanase, submitted to Journal of Molecular Graphics and Modeling (**accepted with major corrections).**

Acknowledgement

I would like to give thanks to:

- My supervisor: Dr Mahmoud soliman for his support and knowledge.
- My family for their support and encouragement.
- My husband and three kids for their unconditional love, patience and understanding.
- My friends for their help and encouragement.
- My colleagues: Buisile Chibi without whom I would not have got through this, Yushir Maharaj and Ndumiso Mhlongo for all the good, the bad and the ugly times, I am grateful.
- Dr Adam Skelton for his assistance.
- Dr Mahmoud Ebrahim for his technical support.
- The Centre for High Performance Computing (CHPC) for the availability of cluster as well as the support of their help desk members.

- **Computational insight into Glycosidase enzymes using bioinformatics tools and molecular dynamic simulations**

Table of Contents

List of figures

List of abbreviations

AM1	Austin Model 1
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
BCX	*β Circulans* Xylanase
CAZY	Carbohydrate Active enzyme
Cyc	Cysteine
DNPdFXb	Dinitrophenyl deoxy fluro xylobioside
FEP	Free Energy Perturbation
GH	Glycoside hydrolase
Glu	Glutamic acid
HF	Hartree Fock
IPCM	Isodensity Polarized Continuum Model
IUBMB	International Union of Biochemistry and Molecular Biology
KIE	Kinetic Isotope Effect
LCAO	Linear combination of atomic orbitals
LR	Linear Response
MC	Monte Carlo
MD	Molecular Dynamic
MM	Molecular Mechanics
MMGBSA	Molecular Mechanic Generalized Born Surface Area

MMPBSA	Molecular Mechanic Poisson Boltzmann Surface Area
NDDO	Neglect of Diatomic Differential Overlap
NPH	Isoenthalpic-isobaric ensemble (Number of particles N, pressure P and enthalpy H are constant)
NPT	Isothermal-isobaric ensemble (Number of particles N, pressure P and temperature T are constant)
NVE	Microcanonical ensemble (Number of particles N, volume V and energy E are constant)
PDB	Protein Data Base
Phe	Phenylalanine
PMF	Potential Mean Force
PNP	Para Nitro Phenyl
Pro	Proline
PSE	Potential Surface Energy
QM	Quantum Mechanics
RIN	Residue Interaction Network
RMSF	Root Mean Square Fluctuation
RMSD	Root Mean square Deviation
SASA	Solvent Accessible Surface Area
SCF	Self-Consistent Field
SEA	Substrate Envelope Analysis
TI	Thermodynamic Integration
TS	Transition State
TS1	Transition State 1

TS2	Transition State 2
Tyr	Tyrosine
VDW	Van Der Waals

Chapter 1

1.1. Background and rational to this study

Glycoside hydrolases (GH) are key enzymes in carbohydrate metabolism that are found in three major kingdoms archaebacteria, eubacteria and eukaryotes. These are a large group of enzymes with a diverse range of functions in living organisms; fertilization, neuronal development, cell proliferation as well as pathogenic processes. Further study of the catalytic mechanisms and subsequent changes in enzyme conformation at an atomic level, factors which are important in the function of the enzyme, is crucial for the development of new compounds with different applications including new inhibitors for the enzyme that can be used as medical drugs.

The xylanase enzyme, a member of the GH family, is highly homologous with a conserved jelly-roll frame work. The enzyme resembles a partially closed left hand with what is described as a finger and thumb region. Previous work suggested that the movement of the thumb region plays a key role in substrate selectivity and in the catalytic process.

Earlier computational studies into the dynamics of the thumb-like structure of the enzyme have yielded inconsistent and incomplete results, as they have not considered the enzyme at the various stages of glycosylation; namely in the free enzyme stage, the covalent enzyme-substrate complex stage and the non-covalently bound stage, to consider the effect the substrate has on the enzyme dynamics. Simulation time scale used in previous studies was short, and the thumb-finger dynamics were measured based on a single distance parameter.

The work in this thesis is to gain further insight into the factors that affect the catalytic process by providing a more complete and accurate understanding of the dynamics of the thumb-finger region.

Another aspect of interest that we have explored is the contributions of certain key residues in the active site of the xylanase enzyme. In the past mutational studies had shown that the two carboxylic residues Glu78 Glu172 as well as Tyr 60, Tyr80 and Arg112 all play a significant role in the catalytic activity of the enzyme. The extent to which each of these residues

contributes to enzyme-substrate binding warrants further investigation as this knowledge could be pertinent for drug-inhibitor design.

Molecular dynamic (MD) simulations provide detailed information on the fluctuations and conformational changes of enzymes, and are a reliable means of investigating the structure, dynamics and thermodynamics of biological systems. MD also offers a convenient and reliable means to further study and understand binding affinities. In this thesis MD simulations have been used to this end.

1.2. Aim and objective of this study

This study has two aims:

1. The main aim is to investigate the conformational changes in the endo-β-1,4-xylanase enzyme.

   To accomplish this, the following objectives were outlined:

   – To obtain insight into the effect of substrate binding on the thumb-finger movement.
   – To explore different matrices to describe this dynamic motion.

2. Our second aim is to gain insight into the role of the active sight residues Glu78, Glu172, Tyr69, Tyr80 and Arg112 on binding and catalytic activity of the enzyme.

   To accomplish this aim our objectives were:

   – To measure the difference in binding affinities between the enzyme wild type and the mutations.
   – Performing binding using different methods and at different levels of theory to see which level will best validate experimental data.

1.3. Novelty of this study

Previous computational studies have raised a few important questions, which have prompted us to conduct this study: Only short MD simulations of 1ns were performed. Thumb-finger dynamics in these studies were based on a single distance measurement between the thumb and finger regions. The enzyme was considered at only one stage of the mechanism cycle, which begs to the question of the comprehensiveness and completeness of these studies.

In our study we have attempted to expand our understanding of the thumb-finger movement by performing molecular dynamic simulations for the three different phases of the enzyme in the catalytic pathway, that is the free enzyme, the non-covalently bonded enzyme and the enzyme covalently bound to the substrate. We have also increased the simulation time to a total of 10ns and have used various matrices to describe the thumb-finger motion.

We believe that our work presented in this thesis has provided crucial insight into some of factors affecting the GH catalytic process, which can potentially assist with the design of new drug inhibitor that interfere with the natural dynamics of the thumb-finger region.

For our second study in order to quantitatively describe the contribution of each residue, Glu 78, Glu172, Tyr60, Tyr80 and Arg112, we have used free binding calculations performed on an MD simulation of each of the mutated residues Glu78Asp, Glu172Asp, Tyr69Phe, Tyr80Phe and Arg112Asn including the wild-type. This was done at all atom MM level. We have also explored different methods of QM/MM description using AM1 level of description.

1.4. Overview of this thesis

The content of this thesis is divided into six chapters including this one:

Chapter 1: is an introductory chapter to give a background and overview on this study

Chapter 2: provides an introduction to the glycoside hydrolase enzyme. This includes the role of the enzyme, classification, mechanism, aspects such as sugar ring distortion and transition state analysis as well as introducing the xylanase enzyme; the member of the glycoside hydrolases used for computational simulation, analysis and study in this thesis.

Chapter 3: provides a background to the science of computational chemistry and the different molecular modelling and simulation techniques, highlighting computational tools such as molecular dynamic simulations, molecular mechanics and binding free energy calculations.

Chapter 4: (Submitted article- this chapter is presented in the required format of the journal): this is the research paper from this study and is entitled Insight into the conformational changes of glycosidases during the catalytic pathway: molecular dynamics simulations on GH- 11 xylanase. It is our effort to better understand the dynamics of the thumb-finger movement of the xylanase enzyme in its free form as well as the non-covalently

bound enzyme and the enzyme covalently bound to the substrate. It contains the computational methods used, discussion and results obtained.

Chapter 5: consists of the work undertaken to gain insight into the role of some key active site residues in substrate binding using free binding energy calculations, including methods used, results and a brief discussion.

Chapter 6: expound the concluding remarks and future plans for my research.

Chapter 2

2. Glycoside Hydrolases

The following chapter gives a brief overview of the glycoside hydrolase enzyme, its function, classification, mechanisms, transition states and ring distortion, with a brief introduction to the family member endo-1,4-β-xylanase.

2.1. The role of Glycoside Hydrolase

Carbohydrates, also known as saccharides, are organic compounds that are involved in a wide range of biological processes. Monosacchrides combine, via the glycosidic bond to form disaccharides, oligosaccharides and polysaccharides.

Glycoside hydrolases or Glycosidases are the group of enzymes that catalyse the hydrolysis of the glycosidic bond between two or more carbohydrate polymers, or between a carbohydrate and a non-carbohydrate moiety (Figure 1) (*1*).

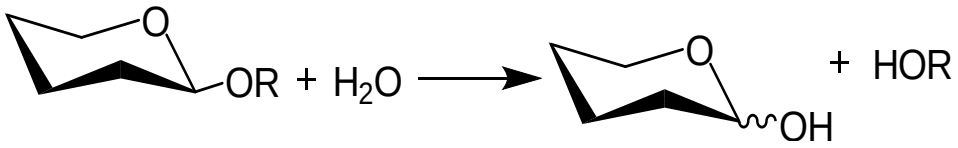

Figure 1: Reaction mechanism of the glycoside hydrolase enzyme (*2*).

2.2. Nomenclature

The IUBMB Enzyme nomenclatures of Glycoside hydrolases are typically based on their substrate specificity. Glycosidase catalyses the hydrolysis of glycosides, Lactase catalyses the cleavage of lactose and Xylanase hydrolyses xylose based polymer xylan.

2.3. Classification

There are a few methods used to classify Glycoside hydrolases. They can be classified as retaining or inverting enzymes depending on the stereochemical outcome of the hydrolysis reaction (*3*). They can also be classified as exo- or endo-acting, depending on whether they act at the non-reducing end or middle of the oligosaccharide or polysaccharide respectively. Another useful method is by sequence and structure based methods (*4*). Using this method, GHs have been classified into more than 100 families which are permanently available on the CAZY (Carbohydrate Active enzyme) database (*5, 6*).

2.4. Active site topology

The overall topology of the active site of the 22 families of known 3D structure, fall into 3 classes (*7*):

Pocket or crater- these enzymes are modified for substrates that have a large number of non-reducing chain ends, that are readily recognised at the surface. It's found in most monosaccarides such as the glycosidases and sialidasesas well as some exo-acting polymerase example, βamylase and glucoamylase.

Cleft or grove- the open structure of these enzymes allow for multiple binding with the polymeric substrates. This topology is commonly seen in endo-acting polymerase, such as lysozymes, chitanases, α amylase and xylanase.

Tunnel- they arise from the cleft topology when the protein develops long loops that covers the cleft forming a tunnel. It is found only in cellobiohydrolases. The tunnel enables the polysaccharide to be threaded through it, thereby allowing the enzyme to discharge the product while remaining bound to the polysaccharide chain.

2.5. Mechanisms

Glycosidase catalysis proceeds via two distinct mechanistic pathways, either inverting or retaining, based on the stereochemical outcome.

2.5.1. Inverting glycosidases

Inverting glycosidases hydrolyze the glycosidic bond with a net inversion of the stereochemical configuration of the anomeric carbon atom. It proceeds via the displacement of the leaving group with the aid of a water molecule, and utilizes two residues, usually, carboxylate residues, which act as an acid and base respectively, See figure 2.

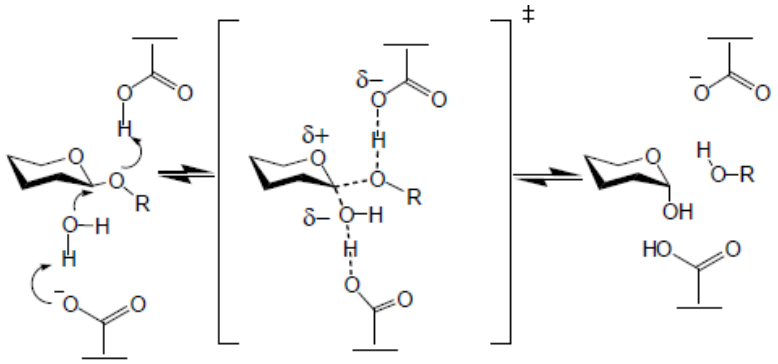

Figure 2: General mechanism for an inverting glycosidase enzyme (*2*).

2.5.2. Retaining glycosidases

Retaining glycosidases proceed in a two step double displacement mechanism with the net retention of the stereochemistry at the anomeric center (*8*). Again the reaction uses two carboxylate residues, one acting as a neucleophille and the other an acid/base.

In the first step, called glycosylation, the reaction involves cleavage of the glycosidic linkage and the formation of a glycosyl-enzyme intermediate via transition state 1 (TS1). The acid/base residue assists cleavage by protonating the glycosidic oxygen.

The second step, referred to as deglycosylation proceeds via transition state 2 (TS2) and involves water hydrolysis of the glycosyl-enzyme intermediate aided by a deprotonated carboxylate that now acts as a base accepting a hydrogen from the water. See figure 3.

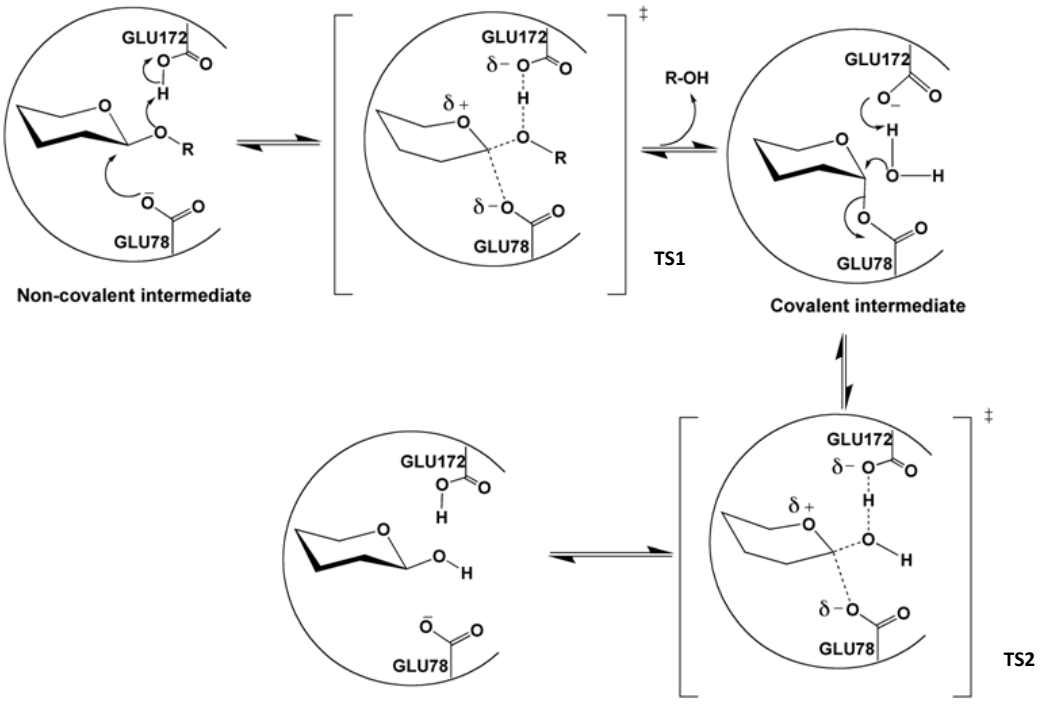

Figure 3: General mechanism for a retaining glycosidase enzyme.

2.6. Transition state analysis

One of the characteristics that the retaining and inverting enzymes share is that they both proceed through a transition state characterized by glycosyl bond cleavage and oxcarbinium-

ion like character. (Figure 2 and 3). The oxcarbinium character is cantered around the anomeric center.

Studies combining isotope effect measurement with computational methods have provided further insight into this transition state structure. The Potential surface energy (PSE) of inosine at transition state shows a distribution of positive charge resulting from oxcarbinium character of the ribosyl ring (*9*).

One important feature of transition state analysis is the measurement of the binding energy contribution of sugar hydroxyl groups in the enzyme-substrate complex at transition state. Kinetic analysis of specifically deoxygenated analogues revealed that the activity if the enzyme is heavily dependent on binding with certain hydroxyl groups compared to others (*10, 11*).

2.7. Covalent intermediate

The first hypothesis proposed on the covalent nature of the transition state intermediate stems from studies using kinetic isotope effect (*12*).

The large α-secondary deuterium kinetic isotope effect for both glycosylation and deglycosylation is easily explained by the formation of a transient glycosyl-enzyme intermediate flanked by oxcarbinium ion like transition states having greater sp^2 hybridisation than the covalent intermediate (*13*).

But perhaps the best insight into its covalent nature has come from x-ray crystallographic analysis of trapped intermediates. This was achieved by adopting a 2-flurosugar strategy. The glycosyl-enzyme intermediate was trapped by treating it with an activated 2-deoxy-2-fluoroglycoside. Since such substituents slow both glycosylation and deglycosylation steps, a good leaving group, such as fluoride or dinitrophenolate, are incorporated into the analogue to speed up glycosylation allowing the intermediate to accumulate (*14, 15*).

2.8. Substrate ring distortion

In order for the enzyme catalytic activity to proceed with transition states having sufficient oxcarbinium ion character, it requires a coplanar geometry in which C5, O5, C1 and C2 adopt a coplanar arrangement (*8*). The mechanism utilized by the enzyme to achieve this is through the distortion of the pyranose ring at 1-subsite. The substrate sugar ring at 1-subsite changes conformation and adopts a distorted conformation, characterised by a raise in the charge at

the anomeric carbon with an increase in distance between the C1 atom and the leaving group (*16*).

This was originally proposed by Phillips and co-workers based on the structure of a lactone derivative of chitotetraose. α –deuterium and β –deuterium KIE experiments showed that planarity limits the number of conformations the pyranose ring would adopt (*8, 17*).

Among the potential pseudo rotational conformational itineraries for the sugar ring interconversions there are four potential conformations where the planarity of C5, O5, C1 and C2 is satisfied ($^{2.5}B$, $B_{2.5}$, $^{4}H_{3}$, $^{3}H_{4}$) (Figure 4). Recent data have shown 3 conformations known in GH families ($^{2.5}B$, $B_{2.5}$, $^{4}H_{3}$) (*18*).

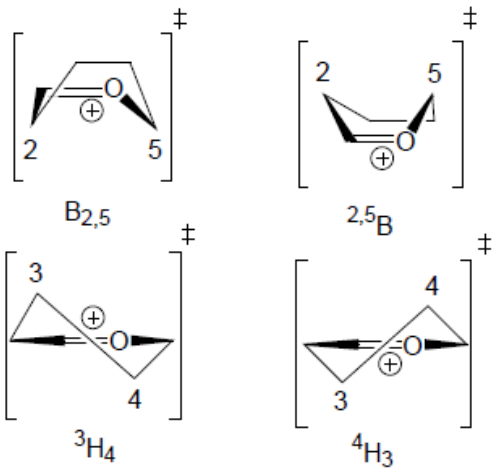

Figure 4: Possible conformations of oxcarbinium ion-like transition states(*2*).

2.9. Xylanase

The endo-β-1,4-xylanase enzyme from *Bacillus circulans* is a member of family G/11 xylanases. Xylanase is a relatively small glycosidase (2kDa), for which the X-ray crystal structure is known (*19-21*).

We have taken a closer look at this member of the glycoside hydrolase family in order to gain a better understanding of certain aspects in the mechanism and function of GH's in general.

2.9.1. Classification, mechanism and active site residues

Xyalanse, is the enzyme that catalysis the hydrolysis of xylan, a polymer of xylose. X-ray crystallographic studies (*19, 20, 22*) indicate that the xylanase enzyme is a retaining enzyme.

The catalytic residues are the conserved glutamic acid residues, Glu78 and Glu172, the lie on opposite sides of each other in the active site. Glu172 acts as an acid/base and Glu78 as the nucleophile. These studies have also identified residues Tyr69 and Tyr80 which are involved in substrate binding and Arg112 which plays a role in catalysis.

Mutational analysis together with X-ray crystallographic studies of the enzyme-substrate complex has revealed some important details of the active site as well as the mechanism of the enzyme (*19*). Mutation of the glutamic acid residues to glutamate (E78Q and E172Q), resulted in no detectable activity in the enzyme; however, when the carboxylate side chain is maintained but the length is shortened in (E78D and E172D), some residual activity is observed. The great decrease in activity observed with the mutants suggests that Glu78 and Glu172 are the catalytic acid residues.

Mutation of the Tyr80 to Phenylalanine (Y80F) resulted in a striking loss of enzyme activity, leaving only 0.03%, residual activity; however mutation of Tyr69 (Y69F) resulted in the total loss of detectable enzyme activity. Mutation of Arg112 to Asparagine (R112N) resulted in a 68% decrease in specific activity.

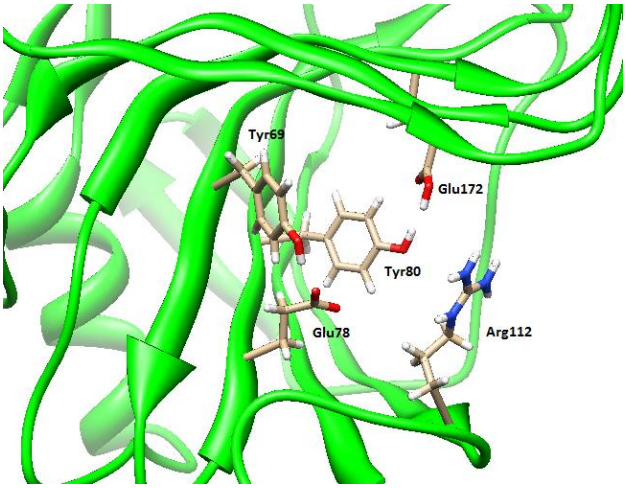

Figure 5: The active site of *β*-xylanase from *Bacillus circulans* PDB code 1BVV.

2.9.2. Conformation of covalently bonded disaccharide

Kinetic and spectroscopic studies were done on the glycosyl-enzyme intermediate trapped, using the mechanism based inhibitor 2', 4'-dinitrophenyl 2-deoxy-2-fluro-β-xylobioside (DNPdFXb) (*23*).

The structure of BCX-2FXb glycosyl-enzyme intermediate (Figure 6) was determined to a resolution of 1,8A^{o} (*24*). The intermediate exhibited excellent stereochemistry.

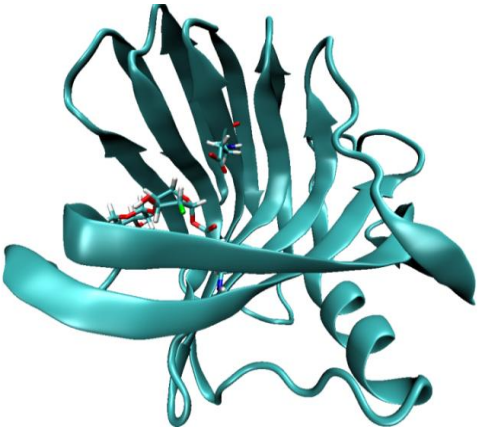

Figure 6: Tree dimentional structure of BCX-2FXb glycosyl-enzyme intermediate along with the two glutamate residues involved in the double displacement mechanism PDB code 1BVV.(*2*).

The X-ray structure of (2FXb) shows the ligand covalently bonded to the nucleophilic residue, Glu78, via an α- anomeric link to the C atom of the proximal sugar ring. The proximal sugar ring occupies the 1-subsite of BCX. The formation of the intermediate occurs through a double displacement mechanism.

2.9.3. Sugar ring distortion

The xylanase moiety, which is covalently bonded to Glu78 shows distortion from the conventional $^{4}C_{3}$ (chair) conformation, and stabilizes in a $^{2,5}B$ (boat) conformation (Figure 7) allowing C5, O5, C1 and C2 atoms of the proximal sugar ring to achieve a coplanar arrangement at the oxcarbinium ion-like transition state (*24*).

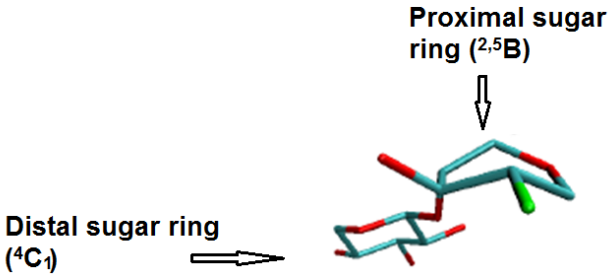

Figure 7: A diagram showing the proximal sugar ring adopting $^{2,5}B$ conformation, while the distal saccharide adopts $^{4}C_{1}$ conformation. Covalent bond is removed for clarity (*2*).

2.9.4. Positioning of the catalytic residues

In retaining GHs including xylanase, the 2 carboxylic residues are ~5.5Å apart in the active site. In inverting glycosidases they are typically further ~10-11Å (*7*).

Modification by mutagenesis of catalytic nucloephile Glu78 by increasing the separation (Glu78Asp) resulted in a large decrease in the rate of glycosylation but had little effect on the extent of bond cleavage or proton donation at TS. A decrease in separation using a selective carboxymethylation of Glu78Cys resulted in no catalytic activity. Mutagenesis of the acid/base residue Glu172 by increasing the distance with Glu172Asp mutant or decreasing it with carboxymethalation of Glu172Cys mutant results on the same little reduction of the rate of hydrolyses (*25*).

2.9.5. Role of Tyr69

More recently, studies have been done to investigate the role of Tyr69 in catalysis and substrate binding (*26*). Molecular dynamic simulations using hybrid QM/MM potential was used to demonstrate that BCX enzyme binds to phenyl β-xylobioside substrate at 1-subsite of the proximal ring in a ^{2,5}B conformation, complimenting the results of previous reports (*24*). Analogous simulations for Tyr69 mutant (Y69F) confirmed the key role of Tyr69 in stabilizing the boat as compared to the chair conformation with a 20 kJ.mol^{-1} difference in relative free energy.

Additionally, 2D Potential mean force or PMF calculations were computed for the BCX wild type as well as Tyr69Phe mutant. Free-energy profiles indicated that Tyr69 lowers the free energy barrier for the glycosylation step by 42 kJ.mol^{-1} (*27*).

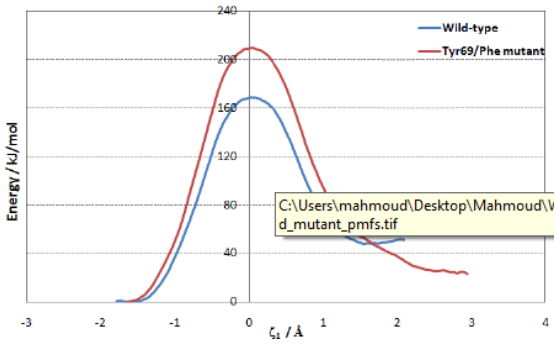

Figure 8: Free energy profiles for formation of the glycosyl-enzyme intermediate in the wild-type BCX(blue) and Y69F mutant(red)(*28*).

2.9.6. Preferred catalytic mechanism

2D PMF calculations based on molecular dynamic simulations for the glycosylation step of wild type BCX by PNP-xylobioside substrate indicated that the pathway proceeds stepwise

for glycosylation. The 1st step involves nucleophilic substitution by Glu78 resulting in the formation of the covalently bonded glycosyl-enzyme intermediate via TS1. The 2nd step proceeds via TS2 involving proton transfer (Figure 9) (*27*).

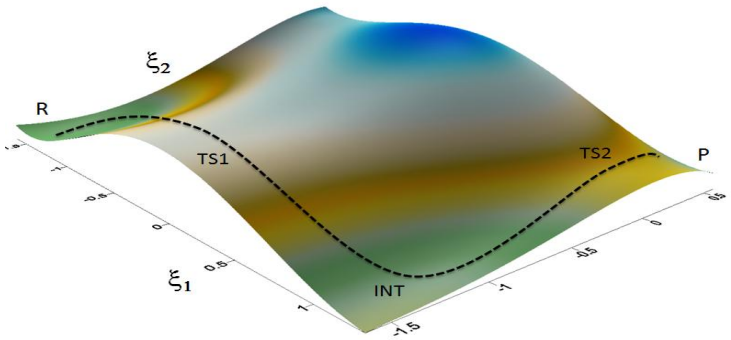

Figure 9: AM1/OPLS calculated 2D-free-energy profile for glycosylation as a function of the coordinates for nucloephilic substitution ξ_1 and proton transfer ξ_2

(relative energies in kJ mol^{-1}) (*28*).

References

1. Meinke, A., Gilkes, N. R., Kilburn, D. G., R. C. Miller, J. R. C., and Warran, R. A. J. (1993) Cellulose-Binding Polypeptides from Cellulomonas Fimi:Endoglucanase D (Cend), a Family a 1-1,4-Glucanase, *J. Bacteriol. 175*, 1910-1918.
2. Soliman, M. E. S. (2009) Computational Modelling of Glycoside Hydrolase Mechanism, In *Department of chemistry*, University of Bath.
3. Henrissat, B., and Davies, G. J. (2000) Glycoside Hydrolases and Glycosyltransferases. Families, Modules, and Implications for Genomics, *Plant Physiol. 124*, 1515-1519.
4. Henrissat, B. (1991) A Classification of Glycosyl Hydrolases Based on Amino Acid Sequence Similarities, *Biochem. J 280*, 309-316.
5. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active Enzymes Database (Cazy): An Expert Resource for Glycogenomics, *Nucleic Acids Res. 37*, 233-238.
6. http://www.cazy.org.
7. Davies, G., and Henrissat, B. (1995) Structures and Mechanisms of Glycosyl Hydrolases, *Structure 3*, 853-859.
8. Sinnott, M. L. (1990) Catalytic Mechanisms of Enzymic Glycosyl Transfer, *Chem. Rev. 90*, 1171-1202.
9. Horenstein, B. A., and Schramm, V. L. (1993) Correlation of the Molecular Electrostatic Potential Surface of an Enzymatic Transition State with Novel Transition-State Inhibitors, *Biochemistry 32*, 9917-9925.
10. Sierks, M. R., Bock, K., Refn, S., and Svensson, B. (1992) Active Site Similarities of Glucose Dehydrogenase, Glucose Oxidase, and Glucoamylase Probed by Deoxygenated Substrates, *Biochemistry 31*, 8972-8977.
11. McCarter, J. D., Adam, M. J., and Withers, S. G. (1992) Binding Energy and Catalysis-Fluorinated and Deoxygenated Glycosides as Mechanistic Probes of *Escherichia coli (Lacz*) beta-galactosidase, *Biochem. J. 286*, 721-727.
12. Sinnott, M. L., and Souchard, I. J. (1973) The Mechanism of Action of B-Galactosidase- Effect of Aglycone Nature and A-Deuterium Substitution on the Hydrolysis of Aryl Galactosidase, *Biochem. J 133*, 89-98.
13. Notenboom, V., Birsan, C., Nitz, M., Rose, D. R., Warren, R. A. J., and Withers, S. G. (1998) Insights into Transition State Stabilization of the Beta-1,4-Glycosidase Cex by Covalent Intermediate Accumulation in Active Site Mutants, *Nat. Struct. Biol. 5*, 812-818.
14. McCarter, J. D., Yeung, W., Chow, J., Dolphin, D., and Withers, S. G. (1997) Design and Synthesis of 2'-Deoxy-2'-Fluorodisaccharides as Mechanism-Based Glycosidase Inhibitors That Exploit Aglycon Specificity, *J. Am. Chem. Soc. 119*, 5792-5797.
15. McCarter, J. D., and Withers, S. G. (1996) 5-Fluoro Glycosides: A New Class of Mechanism-Based Inhibitors of Both Alpha- and Beta-Glucosidases, *J. Am. Chem. Soc. 118*, 241-242.
16. Biarnés, X., Nieto, J., Planas, A., and Rovira, C. (2006) Substrate Distortion in the Michaelis Complex of Bacillus 1,3–1,4-B-Glucanase, *J. Biol. Chem. 281*, 1432-1441.
17. Sulzenbacher, G., Driguez, H., Henrissat, B., Schulein, M., and Davies, G. J. (1996) Structure of the Fusarium Oxysporum Endoglucanase I with a Nonhydrolyzable Substrate Analogue: Substrate Distortion Gives Rise to the Preferred Axial Orientation for the Leaving Group, *Biochemistry 35*, 15280-15287.

18. Davies, G. J., Ducros, V. M. A., Varrot, A., and Zechel, D. L. (2003) Mapping the Conformational Itinerary of Beta-Glycosidases by X-Ray Crystallography, *Biochem. Soc. Trans. 31*, 523-527.
19. Wakarchuk, W. W., Campbell, R., Sung, W., Davoodi, J., and Yaguchi, M. (1994) Mutational and Crystallographic Analyses of the Active Site Residues of the Bacillus Circulans Xylanase, *Protein Sci. 3*, 467-475.
20. Krengel, U., and Dijkstra, B. W. (1996) Three-Dimensional Structure of Endo-1,4-B-Xylanase I from *Aspergillus niger*: Molecular Basis for Its Low Ph Optimum, *J. Mol. Biol. 263*, 70-78.
21. Wakarchuk, W., Methot, N., Lanthier, P., Sung, W., Seligy, V., Yaguchi, M., To, R., Campbell, R., and Rose, D. (1992) *The 20 Kd Xylanase of Bacillus subtilis: A Structure/Function Analysis*.
22. Torronen, A., and Rouvinen, J. (1997) Structural and Functional Properties of Low Molecular Weight Endo-1,4-B-Xylanases, *J. Biotechnol. 57*, 137-149.
23. Miao, S. C., Ziser, L., Aebersold, R., and Withers, S. G. (1994) Identification of Glutamic Acid 78 as the Active Site Nucleophile in *Bacillus Subtilis* Xylanase Using Electrospray Tandem Mass Spectrometry, *Biochemistry 33*, 7027-7032.
24. Sidhu, G., Withers, S. G., Nguyen, N. T., McIntosh, L. P., Ziser, L., and Brayer, G. D. (1999) Sugar Ring Distortion in the Glycosyl-Enzyme Intermediate of a Family G/11 Xylanase, *Biochemistry 38*, 5346-5355.
25. Lawson, S. L., Wakarchuk, W. W., and Withers, S. G. (1997) Positioning the Acid/Base Catalyst in a Glycosidase: Studies with Bacillus Circulans Xylanase, *Biochemistry 36*, 2257-2265.
26. Soliman, M. E. S., Ruggiero, G. D., Pernia, J. J. R., Greig, I. R., and Williams, I. H. (2009) Computational Mutagenesis Reveals the Role of Active-Site Tyrosine in Stabilising a Boat Conformation for the Substrate: Qm/Mm Molecular Dynamics Studies of Wild-Type and Mutant Xylanases, *Org. Biomol. Chem. 7*, 460-468.
27. Soliman, M. E. S., Pernia, J. J. R., Greig, I. R., and Williams, I. H. (2009) Mechanism of Glycoside Hydrolysis: A Comparative Qm/Mm Molecular Dynamics Analysis for Wild Type and Y69f Mutant Retaining Xylanases, *Org. Biomol. Chem. 7*, 5236-5244.
28. Soliman, M. E. S., Pernia, J. J. R., Greig, I. R., and Williams, I. H. (2009) Mechanism of Glycoside Hydrolysis: A Comparative Qm/Mm Molecular Dynamics Analysis for Wild Type and Y69f Mutant Retaining Xylanases, *Org. Biomol. Chem.*, 5236-5244.

Chapter 3

3. Introduction to Computational Chemistry

In this chapter a summary of the basic principles of computational chemistry and methods used in this study has been presented.

3.1. Background

Computational chemistry is a branch of chemistry used to assist in solving chemical problems by incorporating the results of theoretical chemistry, into proficient computer programs, to calculate the structures and properties of molecules and solids. Results generated, normally compliment experimental data; however, they can be particularly useful for determining properties that are inaccessible experimentally, allowing us to make predictions for unobserved chemical happenings.

Computational chemistry methods cover both static and dynamic situations, and range from decidedly accurate to very approximate; accurate, or *ab initio*, methods are based entirely on theoretical principles and are typically feasible for only small systems. Other less accurate methods are called empirical or semi empirical. Their inaccuracy is attributed to the fact that they employ experimental results in their calculations.

Both *ab initio* and semi empirical approaches involve approximations. Most *ab initio* calculations make the Born-Oppenheimer approximation, in order to simplify the underlying Schrödinger equation by assuming that the nuclei remain in place during the calculation. In principle, as the number of approximations is reduced, the calculation converges to the exact solution of the underlying equation. In practice, however, it is impossible to eliminate all approximation, and residual error inevitably remains. The goal of computational chemistry is to minimize this residual error while keeping the calculations acquiescent. It is anticipated, however, that the less severe the approximations making up a particular theoretical model, the closer the results are to experiment.

3.2. History

The first theoretical calculation in chemistry stems from theories steeped in the history of quantum mechanics. Many early works done by Salter, London, Eyring and Pauling were influential, and served as the primary references for the early development of computational quantum chemistry (*1*). Computational chemistry became more of a reality with the development of efficient computer technology.

Distinguished areas within computational chemistry:

-Predicting the molecular structure of molecules using force simulations, or more accurate quantum chemical methods, to find stationary points on the energy surface, as the position of the nuclei is varied.

-Storing and searching for data on chemical entities.

-Identifying correlations between chemical structure and properties.

-Computational approaches to help in efficient synthesis of compounds

-Computational methods for drug design and catalysis.

3.3. Methods

3.3.1. Quantum mechanics

There are two models of atomic structure, the Bohr model and the quantum model. Both models are used to evaluate the energy of a particular system.

3.3.1.1. Bohr model

Bohr suggested that electrons are particles that revolve around the nuclei in orbits. The orbits remain at fixed distances from the nucleus. Electrons are free to move between orbits, using or releasing energy in the process see figure 10 below.

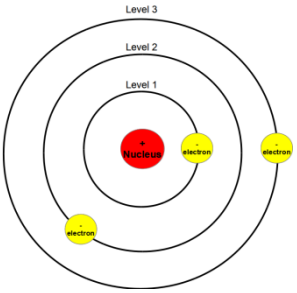

Figure 10: Bohrs Model (*2*)

The energy needed for electrons to move to a different orbital is given in the equation

$$\Delta E = h \times r\left(1/n_2^2 - {}^1/_{n_1^2}\right)$$

Eq. 1

Where,

E=energy

R=radius

h= Plank's constant

n= orbit number

Using the Bohr method we are able to determine exactly where each electron is in its path around the nucleus.

3.3.1.2. Quantum model

The quantum model suggests that electrons are not particles, but have wavelike characteristics. It uses the Schrödinger equation to calculate different properties of the electron (*3*).

If we consider that ψ is in a stationary state we get the time independant schröedinger equation:

$$H\psi = E\psi$$

Eq.2

Where,

ψ= psi, wavefunction

E= the energy of the system

H= the Hamiltonian

The time-dependent Schrödinger for free particles (*4*):bohr model

$$H\psi(x,t)=i\hbar\frac{\partial}{\partial t}\psi(x,t)$$

Eq.3

Where,

H= Hamiltonian operator which corresponds to the total energy of the system.

ħ= ħ/2π (reduced Planks constant)

ψ (x,t)= the wave function at time t and at position x.

i = imaginary unit.

The Schrödinger equation is an eigenvalue equation. We obtain two crucial values from it: E, the energy of the system, and ψ, the wavefunction. The wavefunction has no significance by itself. However, the square of the wavefunction can be used to find the probable location of a particle at time t and at position x. $|\psi(x,t)|^2$ = probability density. To find this probability for three dimensional space (*x,y* and *z* coordinate system), you use Schrödinger's equation with a specified radius, the squared wavefunction, and multiply the wavefunction by 4(pi)r^2 the volume of the sphere.

The molecular Hamiltonian (*4*);

$$H = -\frac{\hbar^2}{2m}\nabla^2 + V(x)$$

Eq.4

Where:

m= mass of the electron.

V(x)= potential energy of the particle at point *x.*

∇^2= Laplace operator of the particle.

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

Eq.5

Under the representation of positions and wavefunction of a molecular system this Hamiltonian has the following form:

$$H = T_{nuc} + T_{elec} + V_{nuc-nuc} + V_{elec-elec} + V_{elec-nuc}$$

Eq.6

T= kinetic energy.

V= Coulomb attraction or repulsion.

Approximations need to be made to solve Schrödinger.

First approximation: Born-Oppenheimer Approximation

The approximation assumes that we can study the behaviour of electrons in a field of frozen nuclei, correct Hamiltonian (*5*, *6*),

$$H_{exact} = T_{elec} + V_{elec-elec} + T_{nuc} + V_{nuc-nuc} + V_{elec-nuc}$$

Eq.7

$V_{elec\text{-}nuc}$= electron-nucleus cross term, it is not separable ie. It has a fixed nuclear position. Therefore:

$$H_{exact} = T_{elec} + V_{elec-elec} + V_{elec-nuc} = H_{elec}$$

Eq.8

Note that the term describing nuclear kinetic energy is left out as it is zero, and the nuclear-nuclear Coulomb term is a constant.

Now, this would be acceptable in ground electron state, as it assumes that electronic and nuclear motions are independent, which is not entirely true, this becomes problematic at excited states, where different surfaces may cross and give rise to vibronic interactions, which leads to the second approximation.

The second approximation: the orbital approximation, the Hartree Fock

For the electronic Hamiltonian

$$\mathrm{T}_{elec} + \mathrm{V}_{elec-elec} + \mathrm{V}_{elec-nuc} = \mathrm{H}_{elec}$$

Eq.9

The term$\mathrm{V}_{\text{elec-elec}}$ is not separable and therefore requires an orbital approximation.

The independent particle model assumes each electron moves in its own orbital and ignores the correlation of behaviour of an electron with other electrons. This can lead to serious problems.

The one-electron Hamiltonian *h* for electron *i* in the molecular system is given by its kinetic energy and its potential energy of interaction with all nuclei, with *Z* number of protons at the nucleus α.

$$h_i = -\frac{1}{2}\nabla_i^2 - \sum_{\alpha=1}^{nuc}\frac{Z_\alpha}{x_{i\alpha}}$$

Eq.10

Each electron is contained within one spin molecular orbital, x_i (where there are 2 spin molecular orbitals to a special molecular orbital).

The total wave function is written in the form of the Slater determinant of one-particle wave functions (or orbitals) to asymmetrize the Hartree product (*4*). The generalized expression, written as a determinant, for an N-electron system, is defined as:

$$\psi^{slater} = \frac{1}{\sqrt{\mathrm{N!}}}\begin{vmatrix} x_1(1)x_2(1) & \cdots & x_n(1) \\ \vdots & \ddots & \vdots \\ x_1(N)x_2(N) & \cdots & x_n(N) \end{vmatrix}$$

Eq.11

The set of molecular orbitals that lead to the lowest energy are obtained by a process referred to as "self-consistent field" or SCF procedure. The typical SCF procedure is the HF procedure, and they all lead to an equation in the form

$$f(i)X(x_i) = {}^{\varepsilon}X(x_i)$$

Eq.12

x_i= spin and spacial coordinates of electrons *i.*

f (i) is the Fock operator and can be written as:

$$f(i) = -\frac{1}{2}\nabla_i^2 + V^{eff}(i)$$

Eq.13

Where,

V^{eff}= effective potential seen by electron *i*, which depends on spin orbitals of the other electrons.

The Hartree-Fock approximation leads to the HF equations which are a set of coupled differentials equations, each involving the coordinates of a single electron. These may be solved numerically however; further approximations are introduced to transform the HF equation into a set of algebraic equations.

LCAO

In practice the spacial molecular orbitals ψ_i are formulated as a linear combination of atomic orbitals (LCAO). (*7*):

$$\psi_i = \sum_{j=1}^{T} C_j \phi_j$$

Eq.14

Where; *T* atomic orbitals ϕ contribute to forming *T* molecular orbitals ψ

C_j= the magnitude of contribution that atomic orbital ϕ_j makes to the eventual orbital ψ_j.

Roothaan-Hall Equation

The HF and LCAO approximations, taken jointly and applied to the Schrödinger equation lead to the Roothaan-Hall equation:

$$F_c = \varepsilon S_c$$

Eq.15

Where,

ε= orbital energies

S= overlap matrix

F= Fock matrix which is analogous to the Hamiltonian in the Schrödinger equation.

Correlated Methods

Within the HF method of quantum chemistry, the antisymmetric wavefunction is approximated by a single Slater determination. However a single determinant cannot be used to express exact wavefunctions, as it does not take Coulomb correlation into account. This leads to an overestimation of the total electronic energy. Therefore Hartree Fock limit is always above the exact energy. This difference is termed the Correlation energy (Figure 11):

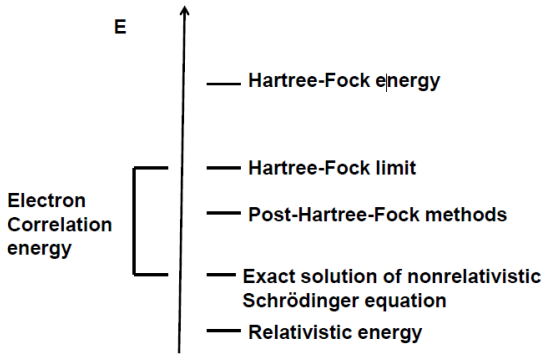

Figure 11: Electron correlation energy in terms of various levels of theory of solutions for the schrödinger equation (*8*).

Many different correlation methods have been introduced; some of them are Kohn-Sham, Møller-Plesset, and configuration interaction.

3.3.1.3. AM1 Semi empirical approximation

In enzymatic systems, the SCF process explained above is too expensive. The size and nature of our systems and the high number of energy evaluation have necessitated the development of cheaper methods.

The Semi-empirical approach is formulated within the *ab initio* method frame work, but many smaller integrals are neglected to speed up calculations. Empirical parameters are introduced to the remaining integrals and calibrated against reliable experimental and theoretical data to compensate for errors caused by these approximations.

AM1 (Austin Model 1) is based on the Neglect of Diatomic Differential Overlap (NDDO) (*9-11*). It approximates 2-electron integrals and uses a modified expression for the nuclear-nuclear core repulsion. This results in a non-physical attractive force that mimics Van der Waals interactions. Reparametization is carried out on dipole moments, ionization potentials and the geometry of molecules (*11*).

If, 2 electrons are described by 2 basis functions, then the 2-electron integral is:(*8*)

$$\iint \phi_a(1)\phi_b(1)\left(\frac{1}{r_{12}}\right)\phi_c(2)\phi_d(2)d\tau_1 d\tau_2$$

Eq.16

a and b= basis function used to describe electron1

c and d= basis function used to describe electron 2

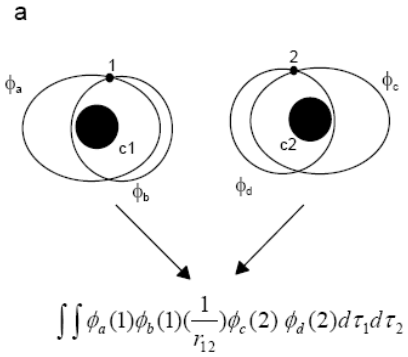

Figure 12: The AM1 Evaluation for two electron interaction integrals where electron repulsion integrals are calculated(*8*).

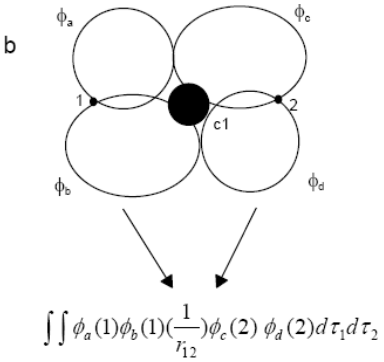

Figure 13: The AM1 evaluation for one centre two electron interaction integrals(*8*).

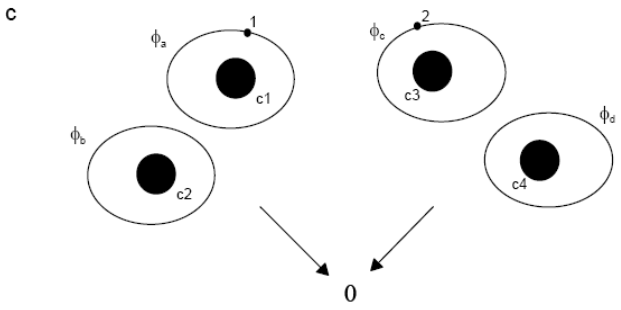

Figure 14: A case where AM1 method is not applicable because it omits two electron repulsions over three and four centres (*8*).

If a and b are on different centers (Figure14) this integral is set to zero under AM1 technique, same is the case with c and d. This is in contrast to the *ab initio* levels of theory that do calculate these 3- and 4-center repulsion integrals. AM1 thus evaluates the 2-center 2-electron integrals (Figure 12) the 1-center 2-electron integrals (Figure 13), the core-electron attraction integrals and the core-core repulsion integrals.

In addition to calculating a smaller number of integrals, the AM1 method incorporates the 1s electrons and nucleus into a single core. The rationale for this approximation is that important chemical changes involve only the valence electrons.

Semiempiric QM methods are currently too computationally demanding to model systems with more than 300 atoms. Enzymes and explicit solvation models that consist of many thousands of atoms are unable to be described at that level of theory. Molecular mechanics may be used to study small molecules as well as large biological systems.

3.3.2. Molecular Mechanics

Molecular mechanics describes molecules in terms of "bonded atoms" which have been disturbed due to non-bonded van der Waals and Coulombic interactions (*10*). Quantum mechanics on the other hand, makes no reference to chemical bonding (*7*).

Molecular mechanics is an expression of the total energy of a system as a sum of individual energy terms for every pair of bonded atoms; it also adds additional potential energy terms coming from angle, dihedral, vdW energy and electrostatics (*6, 8, 12, 13*).

$$E = E_{cov} + E_{non-cov}$$

Eq.17

The components of covalent and non-covalent interactions are:

$$\mathrm{E}_{cov} = \mathrm{E}_{bond} + \mathrm{E}_{\mathrm{angle}} + \mathrm{E}_{\mathrm{dihed}}$$

Eq.18

$$\mathrm{E}_{non-cov} = \mathrm{E}_{electrostatic} + \mathrm{E}_{VDW}$$

Eq.19

This energy level is called an empiric force field and the strategy is Molecular Mechanics.

3.3.2.1. Structural terms

$$\mathrm{E}_{struc} = \sum_{bond} \frac{k_b}{2}(l - l_0)^2 + \sum_{angle} \frac{k_a}{2}(\theta - \theta_0)^2 + \sum_{tortion} \frac{k_t}{2}[1 - \cos n\omega] + \sum_{impropers} \frac{k_i}{2}(\omega - \omega_0)^2$$

Eq.20

Where,

θ, ω, and l =values of bond angle, torsion and length respectively

The last term accounts for out-of-plane torsion and is included to maintain planarity about planar centers (eg. Carbonyl system)

θ_0, ω_0 and l_0= equilibrium geometric parameters in the absence of any other term in molecular mechanics energy.

Additionally, the non-bonded terms that include electrostatic and VdW

$$\mathrm{E}_{non-bond} = \sum_{i=1} \sum_{j=1} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} + \sum_{i=1} \sum_{j=1} \left[\frac{\mathrm{A}}{r_{ij}^{12}} - \frac{\mathrm{B}}{r_{ij}^{6}}\right]$$

Eq.21

The first term of the equation is the Coulombic interaction between two atom-centered partial charges. The second term is the VdW interaction between spicies i and j.

The total energy of the system would thus be the sum of equations E_{struc} and $E_{non\text{-}bond}$ this is E_{train} (*8*).

3.3.2.2. Force fields

The AMBER, CHARMM, OPLS and GROMOS force fields are the most widely used sets of parameters applied to the simulation of biomolecules (*8, 14, 15*).

Each forcefield defines its own sets of atoms, therefore, parameters of an atom in one forcefield would not be transferable to another force field. This is of little concern as the chemical interpretation of parameters is not of great importance, rather the set of variable are adjusted to give a global good fitting. This means, that after a molecular dynamic simulation some properties such as the conformation distribution are similar in any force field (*8*).

3.3.2.3. Limitations

The primary advantage of molecular mechanics models over quantum mechanical methods is that it may be applied to molecular systems consisting of thousands of atoms due to the simplicity of the method. Conformational analysis is possibly one of the most significant applications of MM.

Another advantage is that they are parameterized and may be used to reproduce known experimental data and even anticipate potential outcomes of closely related systems.

There are however, some limitations to the use of MM (*16*). 'Strain energy' is specific to a given molecule this limits the description of geometries and conformations to equilibrium states and so it cannot be used in thermodynamic calculations.

MM cannot be used to explore reactivity or binding selectivity as it does not provide any information about bonding or electronic distribution in molecules.

Another limitation is that the currently available force fields are parameterized using results from experiments and could poorly handle non-equilibrium transition states.

And finally, MM does not depend only on the choice of parameters, but also on systematic among related molecules. This would mean that its success would be limited in describing the structure and conformation of new molecules outside the range of parameterization.

3.3.3. Hybrid methods

Methods for modelling large systems, such as solute/solvent or core/environment interactions can be divided into 2 main groups.

3.3.3.1. The continuum Model

The model considers the solvent as a dielectric continuum and the solute as a molecule embedded in a cavity in the continuum. The most commonly used continuum salvation method used with *ab initio* levels of theory is the isodensity polarized continuum model (IPCM) of Fish & co-workers.

3.3.3.2. Quantum Mechanics/Molecular Mechanics

A method for incorporating quantum mechanics into enzyme kinetics is the QM/MM model which models the environment explicitly (*17*).

In QM/MM a small reactive part of the chemical system is described quantum mechanically while the remaining large non-reactive part is described by molecular mechanics see Figure 15.

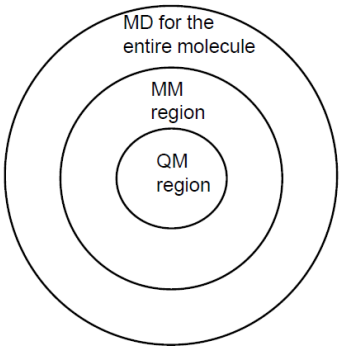

Figure 15: An example of a hybrid QM/MM model.

QM/MM yields good chemical results and is easy to implement in computational codes. It however, produces qualitative results and is not able to provide quantitative calculations (*8, 16, 18*). This is due to the following:

- Good *ab initio* description of the QM region is required however; only semi empirical calculations are possible due to the size of the QM part.
- It is computationally too expensive to access free energy numbers through extensive sampling.

- The great difficulty experienced in calibration of the interaction between the QM and MM regions in biochemical systems.

Spiltting the system

The main idea in QM/MM is to split the chemical system into 2 parts: a *quantum part*, which is described by quantum mechanics and a *classical part*, which constitutes the rest of the system described by molecular mechanics.

The full Hamiltonian for the system is expressed as

$$\mathrm{H}_{eff} = \mathrm{H}_{\mathrm{QM}} + \mathrm{H}_{\mathrm{QM/MM}} + \mathrm{H}_{\mathrm{MM}}$$

Eq.22

Owing to the strong QM-MM interactions, the total energy of the system cannot be written simply as the sum of the energy of the subsystems. Coupling term need to be considered and special precaution needs to be taken at the boundry between the subsystems, especially when it cuts across covalent bonds.

Since van der Waal terms on MM atoms provide the only difference in interaction of one atom type versus another we attribute VdW parameters to all QM atoms and the Hamiltonian describing the interaction between the QM and MM part of the system is given by

$$\mathrm{H}_{\mathrm{QM/MM}} = -\sum_{i=1}^{elec}\sum_{m=1}^{\mathrm{MM}}\frac{\mathrm{Qm}}{r_{im}} + \sum_{\alpha=1}^{nuc}\sum_{m=1}^{\mathrm{MM}}\frac{\mathrm{Z}_{\alpha}\mathrm{Q}_{m}}{r_{\alpha m}} + \sum_{\alpha=1}^{nuc}\sum_{m=1}^{\mathrm{MM}}\left[\frac{\mathrm{A}_{\propto m}}{r_{\propto m}^{12}} - \frac{\mathrm{B}_{\propto m}}{r_{\propto m}^{6}}\right]$$

Eq.23

Where,

Qm= point charge in QM/MM system

Z=charge of QM nuclei

Q_m=charge of MM atoms

r =distance of separation

the Van der Waal term also models electronic repulsion and dispersion interactions, which do not exist between QM and MM atoms because MM atoms do not possess explicit electrons (*19*).

Dividing covalent bonds across QM and MM regions

Link atom approximation

The enzyme residues to be treated quantum mechanically are first identified. However, since the remaining parts of the enzyme will be treated molecular mechanically the link atom method is applied to satisfy the free valency created by the QM-MM division, by adding an atom, usually a hydrogen atom, which is covalently bonded to the QM system. The link atom is included in $H_{QM/MM}$ calculation as represented in Figure 16. The bond Q-M is described at MM level.

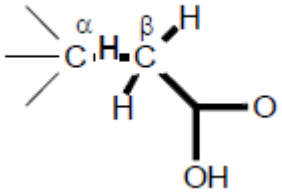

Figure 16: QM link atom (H) addition to an enzyme glutamic acid residue between C_α and C_β. This yields 'acetic acid' QM system.(*8*)

3.4. Molecular Dynamics

Biomolecular dynamics occur over a wide variety of scales in terms of time and space. Two of the most common ways to produce statistically reliable equilibrium ensembles are the Monte Carlo method and the Molecular Dynamic method (*20*).

In Molecular Dynamics the movement of atoms and molecules are simulated. Atoms and molecules are allowed to interact for a period of time so that they can provide a view of their motion. By numerically solving Newton's law of motion for the interacting system of particles results in the production of trajectories of these molecules and atoms (*12, 21*). Potential energy and interacting forces between particles are defined by molecular mechanic force fields.

Newton's equation of motion

$$-\frac{dV}{dx} = m\frac{d^2x}{dt^2}$$

Eq.24

Where;

V= potential energy of the system.

m= mass of the particle.

x= direction of the coordinate

t= time

This is solved numerically propagating a trajectory at small time-steps. MD equations are typically propagated until 10^9-10^{12} steps to observe a chemically interesting reaction.

This method is most useful when the aim is to predict the structure and function of proteins. Results may also be used to determine macroscopic thermodynamic properties of the system.

3.4.1. Temperature and pressure control in MD simulations

In the equation of motion above, the total energy of the system is constant. Time averages obtained for this MD simulation would be equivalent to the average macrocanonical ensemble (NVE), in which the number of particles, the volume and energy of the system are constant. This is not the most appropriate method however, if we are trying to mimic experimental conditions. It is common to do experiments under conditions in which temperature and/or pressure are constant. Thermodynamic ensembles that correspond to these are canonicle (NVT) ensembles, isothermal-isobaric (NPT) ensembles and isobaric isoenthalpic (NPH) ensembles, respectively.

The idea behind the NVT method is that the system being simulated is not isolated but interacts, or is coupled, with an external bath. In microcanonical ensembles, the energy of the system remains constant and the temperature fluctuates. Coupling to an external system means that the energy can be transferred in and out of the system that is, the energy fluctuates. It is this transfer, properly formulated, that allows the algorithm to control the temperature.

The temperature of a particle system is related to the time average of the velocity of the particles.

$$\langle \sum_i^n \frac{1}{2} m_i v_i^2 = \frac{3}{2} n k_B T \rangle$$

Eq.25

Where;

k_B= Boltzman's constant.

T= temperature.

m_i= mass of paticle *i.*

v_i= velocity of particle *i.*

The initial velocities can be given from a Maxwell-Boltzman distribution at the desired temperature. The most intuitive strategy to maintain constant temperature is to multiply the velocity of the particles by a scaling factor $\lambda = {\mathrm{T}_{required}}/{\mathrm{T}_{current}}$ at every step. One such method is the Berendsen method in which the system is coupled to a pretend external bath kept at fixed temperature T. The exchange of thermal energy between the two is gradual. So instead of a sudden resetting of the velocity of a particle to a new value, the velocity is gradually scaled by multiplying it by a factor k given by factor λ.

$$\lambda = \left[1 + \frac{\Delta \mathrm{T}}{\tau_{\mathrm{T}}} \left(\frac{\mathrm{T}}{\mathrm{T}_{ins}} - 1 \right) \right]^{1/2}$$

Eq.26

Where,

ΔT=time step.

τ_T = time constant of the coupling.

T_{ins}=instantaneous kinetic temperature.

In this way the velocity is adjusted such that the instantaneous kinetic temperature approaches the desired temperature T. The strength of the coupling is controlled through the use of an appropriate coupling time constant τ_T.

NPT methods maintain constant temperature, pressure and a constant number of particles. The partition function can be written as the weighted sum of the partition function of canonical ensembles.

$$\Delta NPT = \int (NVT) exp(-\beta PV)\, CdV$$

Eq.27

Where,

B=$^{1}/_{k_B T}$ ($k_{B=}$ Boltzman's constant)

V=volume of the system.

C=$^{N}/_{V}$

3.4.2. Perodic boundry conditions

In simulations in condensed phase the solution is usually surrounded by a bulk system, a solvent. If the solvent is reproduced by a single box then the edge of the box will have an artificial surface tension. To circumvent this difficulty of limited box size, periodic boundary conditions are used. In this approach we surround the original box containing solute and solvent molecules by images of itself. By using this approach we obtain an infinite sized system. Since MD simulations are performed as NVE (constant number of particles, volume and energy) or NVT (constant number of particles, volume and pressure) the volume of the boxes do not vary during simulations and the number of particles are also kept constant by periodicity of the lattice, for example., a particle leaving the box on the left enters it on the right.

3.5. Free binding energy

Bimolecular systems are controlled and governed by changes in the free binding energies in the active site. This is a fundamental law of thermodynamics, and is the property of the entire configuration space consisting of enthalpic and entropic contributions.

Enthalpy is the internal energy, averaged over a Boltzmann distribution over the entire configuration space. Entropy is the negative expected log probability over the same distribution. The internal energy of the system should not be confused with binding free energy. Internal energy is the property of a specific configuration of a system and accounts for electrostatic and hydrophobic interactions among others.

In order to study the mechanics and dynamics of a protein-ligand system, Molecular Dynamic simulations has proven to be a useful tool. Thermodynamic integration (TI) (*22*), linear response (LR) (*23*), free energy perturbation (FEP)(*24*, *25*) and molecular mechanic Gerneralized Born Surface Area (MMGBSA) are but some of the computational methods that exist that estimate ligand binding affinities and selectivities (*26*-*28*).

MMGBSA is fast and versatile and can incorporate the effects of thermal averaging with force field/continuum solvent models to post process series of representative snapshots from MD trajectories.

3.5.1. MMGBSA

The MMGBSA (*29*) method can be summerized as:

$$\Delta G_{bind=} G_{complex-} \left(G_{protein} + G_{ligand} \right)$$

Eq.28

Where,

G_{compex}, $G_{protein}$ and G_{ligand} are the free energies of the complex the protein and the ligand respectively.

$$\Delta G_{bind} = \Delta H - T\Delta S$$

Eq.29

We see that ΔG_{bind} contains both an enthalpic ΔH and entropic ($-T\Delta S$) contribution. The enthalpic ΔH component is composed of the gas-phase molecular mechanics energy (ΔG_{MM}) the salvation free energy (ΔG_{solv}).

ΔG_{MM} is further divided into non-covalent van der Waal (ΔE_{VDW}) and electrostatic energy components (ΔE_{elec})

$$\Delta G_{MM} = \Delta E_{VDW} + \Delta E_{elec}$$

Eq.30

The solvation free energy ΔG_{solv}, is further divided into a polar component (ΔG_{pol}) and a non-polar component ($\Delta G_{non\text{-}pol}$)

$$\Delta G_{sol} = \Delta G_{pol} + \Delta G_{non-pol}$$

Eq.31

The polar component is calculated on the GB module on the Amber suite. The non-polar component ($\Delta G_{non\text{-}pol}$) is determined by:

$$\Delta G_{non-pol} = \gamma SASA + \beta$$

Eq.32

Where,

SASA= solvent accessible surface area that is determined with the MSMS program, using a probe radius of 1.4 Å.

γ and β =empirical constants.

3.5.2. Ligand residue interaction decomposition

The interaction between each of the active site residues and the ligand is analyzed and the binding interaction consists of:

$$\Delta G_{inhib-res} = \Delta E_{VDW} + \Delta E_{elec}\Delta E_{pol} + \Delta E_{non-pol}$$

Eq.33

Reference

1. Nye, M. J. (1994) *Chemical Philosophy to Theoretical Chemistry: Dynamics of Matter and Dynamics of Disciplines 1800-1950*, University of California press.
2. http://www.thephysicsmill.com. (2012) Unreal Truths the Bohr Model of the Atom.
3. Atkins, P., and de Paula, J. (2006) Atkin's Physical Chemistry
4. Schwabl, F. (2007) *Quantum Mechanics*, 4th ed., Springer-verlag Berlin Heidelberg New York.
5. Leach, A. R. (2001) *Molecular Modeling: Principles and Aplications*, 2nd ed.
6. Hinchliffe A (2003) *Molecular Modeling for Beginners* John Wiley & sons Ltd.
7. Hehre, W. J. (2003) *A Guide to Molecular Mechanics and Quantum Chemical Calculations*, Wavefunction, Inc.
8. Soliman, M. E. S. (2009) Computational Modelling of Glycoside Hydrolase Mechanism, In *Department of chemistry*, University of Bath.
9. Cramer, C. J. (1961) *Essentials of Computational Chemistry: Theories and Models* 2nd ed.
10. Rogers, D. W. (2003) *Computational Chemistry Using Pc*, 3rd ed., John Wiley & Sons, Inc.
11. Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., and Stewart, J. J. P. (1985) Am1: A New General Purpose Quantum Mechanical Molecular Model, *J. Am. Chem. Soc. 107*, 3902-3909.
12. Jensen, F. (1999) *Introduction to Computational Chemistry*, John Wiley & Son.
13. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., and Weiner, P. (1984) A New Force Field for Molecular Mechanic Simulation for Nucleic Acid and Proteins, *J. Am. Chem. Soc. 106*, 765-784.
14. Sansom, C. E., and Smith, C. A. (1998) Computer Applications in the Biomolecular Sciences. Part 1: Molecular Modelling *Biochem. Educ.*, 103-111.
15. Keserü, G., and Kolossváry, I. (1999) *Molecular Mechanics and Conformational Analysis in Drug Design. In.1*, Blackwell Science Ltd.
16. Bultinck, P., de Winter, H., langenaeker, W., and Tolleneare, J. P. (2004) *Computational Medicinal Chemistry for Drug Discovery*, CRC press.
17. Senn, H. M., and Thiel, W. (2009) QMMM Methods for Biomolecular Systems, *Angew. Chem. 48*, 1198-1229.
18. Amara, P., Field, M. J., Alhambra, C., and Gao, J. L. (2000) The Generalized Hybrid Orbital Method for Combined Quantum Mechanical/Molecular Mechanical Calculations: Formulation and Tests of the Analytical Derivatives, *Theor. Chem. Acc. 104*, 336-343.
19. Amara, P., and Field, M. J. (1990) Combined Quantum Mechanical and Molecular Mechanical Potentials, *CCA26*.
20. Wang, W., Donini, O., Reyes, C. M., and Kollman, P. A. (2001) Biomolecular Simulations: Recent Developments in Force Fields, Simulations of Enzyme Catalysis, Protein-Ligand, Protein-Protein, and Protein-Nucleic Acid Noncovalent Interactions, *Annu. Rev. Biophys. Biomol. Struct. 30*, 211-243.
21. Taylor, R. D., Jewsbury, P. J., and Essex, J. W. (2002) A Review of Protein-Small Molecule Docking Methods, *J. Comput. Aided Mol. Des. 16*, 151-166.
22. Steinbrecher, T., Mobley, D. L., and Case, D. A. (2007) Nonlinear Scaling Schemes for Lennard-Jones Interactions in Free Energy Calculations, *J. Chem. Phys. 127*, 2141081-21410813.

23. Yokogawa, D. (2013) Linear Response Approximation to Reference Interaction Site Model Self-Consistent Field Explicitly Including Spatial Electron Density Distribution. Free Energy, *J. Chem. Phys. 138*, 1641091-1641095.
24. Roux, B., Nina, M., Pomes, R., and Smith, J. C. (1996) Thermodynamic Stability of Water Molecules in the Bacteriorhodopsin Proton Channel: A Molecular Dynamics Free Energy Perturbation Study, *Biophys. J. 71*, 670-681.
25. Liu, P., Dehez, F., Cai, W., and Chipot, C. (2012) A Toolkit for the Analysis of Free-Energy Perturbation Calculations, *J. Chem. Theory. Comput. 8*, 2606-2616.
26. Hu, G., Wang, D., Liu, X., and Zhang, Q. (2010) A Computational Analysis of the Binding Model of Mdm2 with Inhibitors, *J. Comput. Aided Mol. Des.*, 687-697.
27. Brandsdal, B. O., Österberg, F., Almlöf, M., Feierberg, I., Luzhkov, V. B., and Åqvist, J. (2003) Free Energy Calculations and Ligand Binding, In *Adv. Protein Chem.*, pp 123-158, Uppsala University.
28. Hou, T., Wang, J., Li, Y., and Wang, W. (2011) Assessing the Performance of the Mmpbsa and Mmgbsa Methods: I. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamic Simulations, *J. chem. inf. model. 24*, 69-82.
29. Hu, G., Zhang, Q., and Chen, L. Y. (2011) Insights into Scfv:Drug Binding Using the Molecular Dynamics Simulation and Free Energy Calculation, In *J. Mol. Model.*, pp 1919-1926.

Chapter 4

Submitted article

Insight into the conformational changes of glycosidases during the catalytic pathway: molecular dynamics simulations on GH- 11 xylanase

Mahasin Ebrahim[a], Adam A. Skelton[a], Ndumiso N. Mhlongo[a], Hendrik G. Kruger[b], Ian H. Williams[c] and Mahmoud E. S. Soliman[a *]

[a]School of Health Sciences, University of KwaZulu-Natal, Westville, Durban 4001, South Africa

[b]Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban 4001, South Africa

[c]Department of Chemistry, University of Bath, Bath, United Kingdom

** Corresponding author: email: Soliman@ukzn.ac.za*

Abstract

The structure and dynamics of β-1,4-xylanase from *B. circulans* (BCX) are studied by all-atom molecular dynamics. Three separate 10 ns simulations at different stages of the glycosylation process are performed to gain insight about the structural changes involved. These are, (a) the free enzyme, (b) the enzyme with the substrate, non-covalently associated in the active site and (c) the substrate covalently bound to the active site, representing the known intermediate. In this report, we implement several metrics to accurately estimate the thumb-finger dynamics of the enzyme during the catalytic course. It is shown that the covalently bound substrate causes a closing of the active site and the non-covalently bound substrate causes an opening compared to the free enzyme. These results are rationalized in terms of steric effects for the non-covalent case and a closing of the active site to facilitate the

transition occurring in the covalently bonded case. Furthermore, it is shown that there is a visible closing of the thumb-finger around the covalently bound substrate. Covalent attachment combined with a hydrogen bond between one of the hydroxyl hydrogen atoms of the substrate and the amide oxygen atom of Pro116 is most likely responsible for the tighter conformation.

Keywords: Glycosyl hydrolases, molecular dynamics, xylanases

1. Introduction

Glycoside hydrolases (GH) are a complex group of enzymes that are responsible for the hydrolysis of the glycosidic bond in sugars (*1, 2*). These enzymes possess a diverse range of functions in living organisms, both pathological and physiological, thus making them important therapeutic targets (*3*). The inhibition of some of these enzymes have be utilized for functional drugs for diseases such as influenza (*4*) and diabetes, viral infections (*3*), as well as for potential anticancer drugs (cite). For example, tamiflu inhibits the Neuraminidase enzyme, a type of glycoside hydrolase responsible for breaking down the glycoproteins that are required for influenza virus replication. In-depth knowledge of the structure and function of these enzymes is essential if we are to design better enzyme inhibitors for a greater variety of enzyme targets. In this study, the β-1,4-xylanase from *B. circulans* (BCX) will be used to better understand the mechanism and function of glycoside hydrolase enzymes in general, as it is experimentally (*5-7*) and computationally (*8*, *9*) well-characterised.

This enzyme was, therefore, used as a model for the family of GHs, as a whole, despite it being a bacterial enzyme responsible for breaking down plant matter and, therefore, its inhibition serving no medicinal purpose.

Previous studies have concluded that intermolecular motions of enzymes, in-terms of substrate binding and succeeding chemical and dissociation steps, are important in their function (*10-12*). Thus, the conformational change the enzyme undergoes during catalysis is one aspect that warrants deeper investigation (*13, 14*).

The β-1,4-xylanase from *B. circulans* (BCX) is a small enzyme with two conserved carboxylic residues Glu78 and Glu172 that act as the neucleophile and acid/base residues

respectively, leading to the hydrolysis of the glycosidic link with the net retention of the anomeric carbon configuration. This is a two-step mechanism that occurs via a covalent glycosyl-enzyme intermediate (*7, 15, 16*) (Figure 1). The X-ray crystal structure of the β-1,4-xylanase from *B. circulans* covalently bonded with 2-deoxyfluoroxylose (2DFX) has been resolved (PDB code: 1BVV) (*6*); however, the fluoro group was removed to better represent the natural substrate.

We have previously reported the effect of active site mutant Tyr69Phe (*8*) as well as the ring distortion on the enzyme structural and mechanistic features(*9*). Herein, we expand on our previous reports by providing more insight into the enzyme dynamics associated with the catalytic reaction pathway.

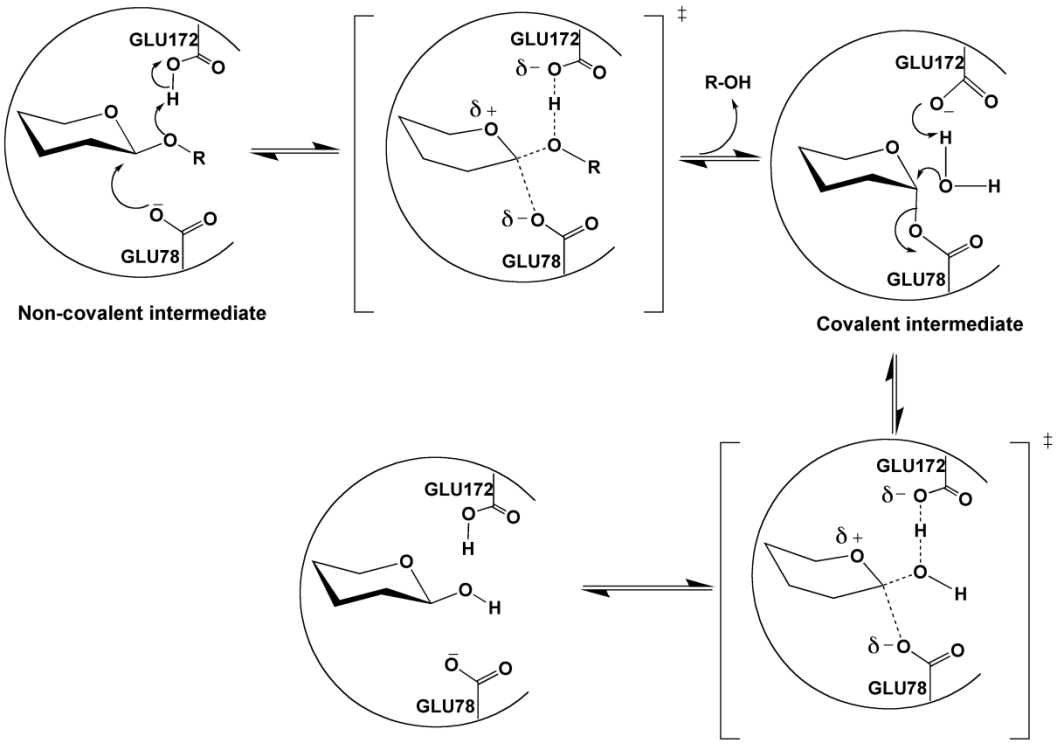

Figure 1. Mechanism of retaining *endo*-1,4-b-xylanase: catalytic residues areGlu78 and Glu172. (Sugar-ring distortion not shown.). R = xylopyranose.

The structure of the GH11 xylanase can be described as a partially closed 'hand' consisting of one domain folding into two sheets, which is twisted to form a cleft on one side of the

protein where the active site is found (*15*). The cleft is covered by a long loop called the thumb region, shown in Figure 2, which involves 11 residues (residue Tyr113-Thr124) connecting β-strands B7 and B8. The thumb is structured as a classical hairpin, containing a Type1 β-turn (residue Ser117-Gly120) and six internal hydrogen bonds. The finger region is formed from A and B β-sheets (residue 10-15) and the palm is made up of an α-helix and a twisted β-sheet.

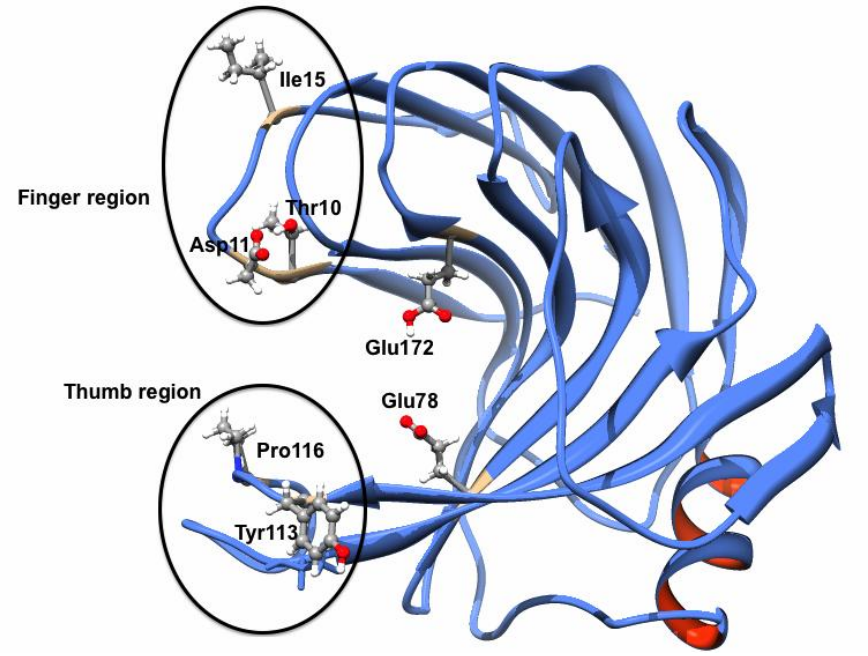

Figure 2. Structure of the BCX thumb (res 113-124) and finger (res 10-15) regions as well as the carboxylic residues Glu78 and Glu172

Crystallographic structures give a rigid view of the location of the atoms that are often partly disturbed by crystal packing. A crystal structure represents only one particular snapshot of many possible conformations. A study of several X-ray structures of the enzyme provides some information about the dynamic movement that should occur going from one conformation to another. The crystallographic B-factors reveal some clues to the general mobility of residues, still the precise dynamic features of enzyme structures are difficult to infer. Therefore, molecular dynamics could provide a robust tool to explore the conformational landscape of biological systems.

Previous experimental and computational studies into the function of the thumb-like structure have resulted in many inconsistent conclusions (*17-19*). Most of these studies have not considered the different enzyme states in the simulation or they may not have used a long

enough MD time scale (maximum 1 ns were used) (*18*). In this study, in order to accurately probe the dynamics of the thumb-like region, different stages of the reaction pathway will be investigated. To this end, 10 ns MD simulations have been performed for the free-enzyme state, the non-covalent complex as well as the covalently bonded glycosyl-enzyme intermediate (see Figure 1). It is worth mentioning that, in the previous reports, estimating the thumb-finger dynamics was mainly based on measuring one parameter which is the distance between the two central residues of the thumb and the finger regions (*18*). In the current study, we will use various metrics to describe the motion of the thumb-finger region. Results obtained from the current study should therefore provide a much better understanding of the role of the thumb-finger regions in GH catalytic process compared to previously reported studies (*17, 18*).

2. Material and Methods

Enzyme models – three enzyme states

The full computational scenario for modelling the three enzyme systems, free enzyme, non-covalent and covalent intermediates has been described in our previous publication (*8*). However, in our previous work a phenolate moiety was used as a leaving group to reproduce the experimental kinetic data (*8*), in this work we have used the xylopyranose moiety to retain the natural substrate structure.

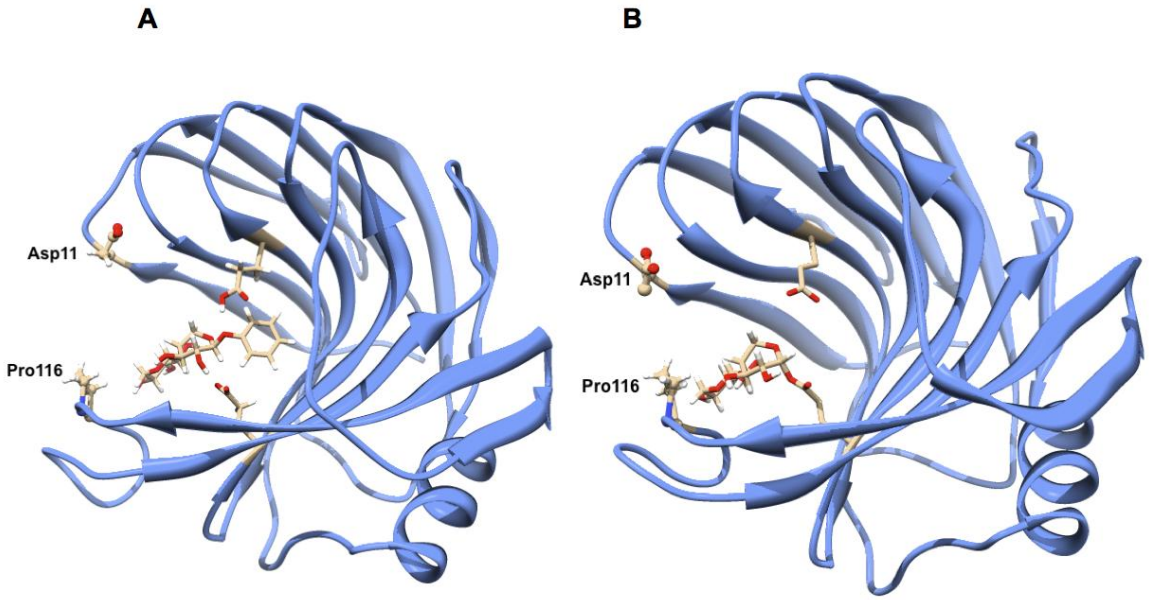

Figure 3. Representative structures for the non-covalent and covalent bound intermediates, A and B, respectively.

Systems set-up and molecular dynamics simulations

The X-ray crystallographic structure of the covalent enzyme-inhibitor complex of wild type BCX was obtained from the Protein Data Bank (PDB accession code 1BVV)(cite). Due to the importance of protonation, the acid/base residue Glu172 in the non-covalent system was protonated and was visually inspected, using Chimera.

Hydrogen atoms were added to the protein by using the LEap module of AMBER10 (*20*). The motivation for the protonation states of the enzyme was addressed in our previous reports (*8*). In this study, optimization of the substrate was first carried out at the HF/6-31G* level with the Gaussian 03 package (*21*). The restrained electrostatic potential (RESP) procedure (*22*), which is also part of the AMBER package, is used to calculate the partial atomic charges, utilizing the Gaussian output. GAFF (*23*) force field parameters and RESP partial charges were assigned using the ANTECHAMBER module in the AMBER 10 package. The standard AMBER force field for bioorganic systems (FF03) (*24*) was used to describe the protein parameters. To neutralize the charge of the systems, an appropriate number of chloride counter ions are added. Each system is then enveloped by a truncated, cubic periodic box of TIP3P (*25*) water molecules at a distance of 10 Å from the solute atoms.

The three enzyme systems were minimized with the SANDER module in a constant volume by 1000 cycles of steepest descent minimization followed by 1000 cycles of conjugated gradient minimization. These procedures ensure that the initial structures were maintained while the solvent is allowed to relax. After energy minimization, the systems were gradually heated from 0 K to 300 K while applying harmonic restraints with force constants of 2 kcal/(mol·Å^2) to restrict the movement of solute atoms for 70 ps using canonical ensemble (NVT)-MD. Amber supports the Berendsen barostat using default parameters. A cutoff of 12Å was used for Lennard-Jones interactions, and the long-range electrostatic interactions were calculated with the particle mesh Ewald (PME) method. The subsequent isothermal isobaric ensemble (NPT)-MD was used for 500 ps to equilibrate the systems. Finally, 10 ns of isothermal isobaric ensemble (NPT)-MD simulation was applied without any restraints where the temperature was regulated at 300 K using the Langevin thermostat and the pressure kept at 1.0 atm. .Trajectories were analysed at every 1 ps using the PTRAJ module.

3. Results and discussion

Estimation of thumb-finger motion: parameters and metrics

Root of Mean Square Fluctuation (RMSF)

To provide information on the local structure, flexibility, and heterogeneity of the enzyme, the root mean square fluctuation (RMSF) (*26, 27*) was analysed. RMSF is a measure of the deviation between the position of a particle, *i* and its reference the mean position, taken over time.

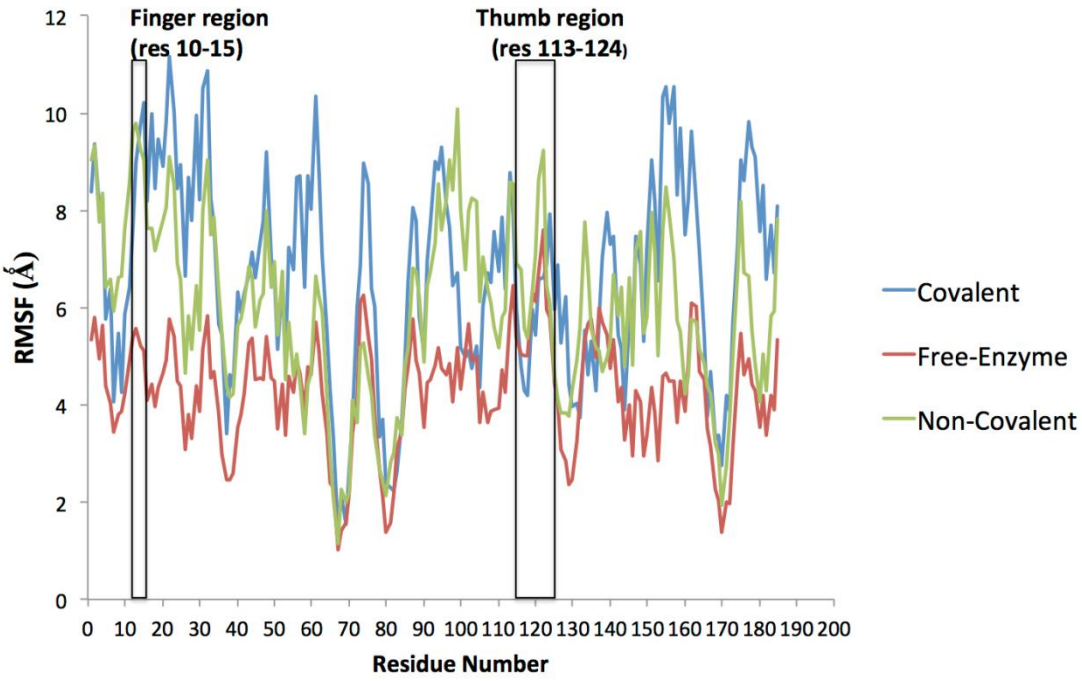

Figure 4. RMSF of backbone atoms per residue of the BCX free enzyme (red), enzyme covalently bonded to the substrate (blue) and non-covalently bonded (green).

For the first step of our investigation, a 10 ns trajectory was analysed to calculate the fluctuation per residue. The RMSF per residue plot for each of the three systems is shown in Figure 4. It can be noticed that the free enzyme RMSF is consistently smaller than those of the covalent and non-covalent systems. This is surprising since it was assumed that the substrate would stabilize the enzyme; however, perhaps the substrate induces changes in the enzyme, making it more flexible.

Distance Metrics

The opening and closing of the active site were determined by the thumb and finger movement and position. Five different distances were carefully chosen to provide information about the dynamics of the thumb-finger regions. These distances were between α-carbon atoms of four different residues, two representing the thumb-finger region, Asp11 and Pro116, and two representing the active site, Glu172 and Glu78. These distances are (a) Asp11-Pro116 (will be hereafter named the 'finger-thumb distance'), (b) Asp11-Glu172, (c) Asp11-Glu78, (d) Pro116-Glu172 and (e) Pro116-Glu78 (see figure 2).

The finger-thumb distance, Pro116-Asp11, provides the extent to which the enzyme opens and closes. Asp11-Glu172 distance gives a measure of whether the finger region becomes compressed in response to the substrate. The Pro116-Glu78 distance is a similar measure but for the thumb. The cross distances, Pro116-Glu172 and Asp11-Glu78, were also considered to roughly estimate the vector of the motion. Collective monitoring of all these distances along the MD trajectory will provide a comprehensive picture on the entire dynamics of the thumb-finger regions, hence the conformational changes during the catalytic cycle. Moreover, in all cases, the distributions of the distances are reported; this gives further information about the population and range of each distance.

As is clearly shown in Figure 5, the thumb-finger distance plot of the covalent system (blue curve) exhibits on average a small distance, indicating a more closed conformation. Conversely, the free-enzyme on average shows a larger distance indicating a more open conformation. This is to be expected as in the covalent phase of the mechanism cycle, the substrate and enzyme are complexed together; this causes the thumb to close and prohibit the complex from being dislodged.

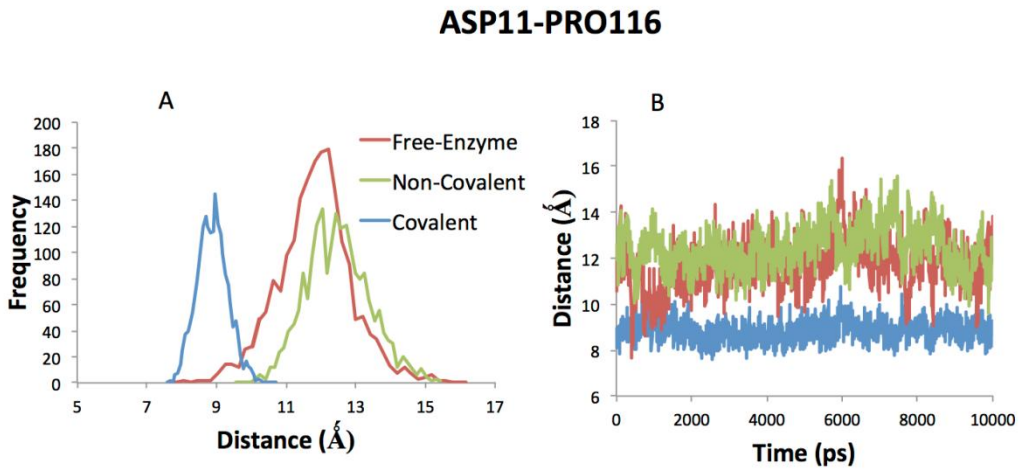

Figure 5. The extent of opening and closing of the enzyme. A) Finger-thumb distance as a function of time and B) a distribution of distances for the free enzyme (red), covalently (blue) and non-covalently (green) bonded systems.

There is a hydrogen bond between one of the hydroxyl hydrogen atoms of the substrate and the amide oxygen atom of Pro116 (Figure 6). This holds the thumb fixed in position since the other end of the substrate is covalently bound to Glu78. For the free and non-covalently bound substrate systems, however, the absence of this interactions cause the thumb to be much more flexible, producing a larger range of distributions, at larger distances.

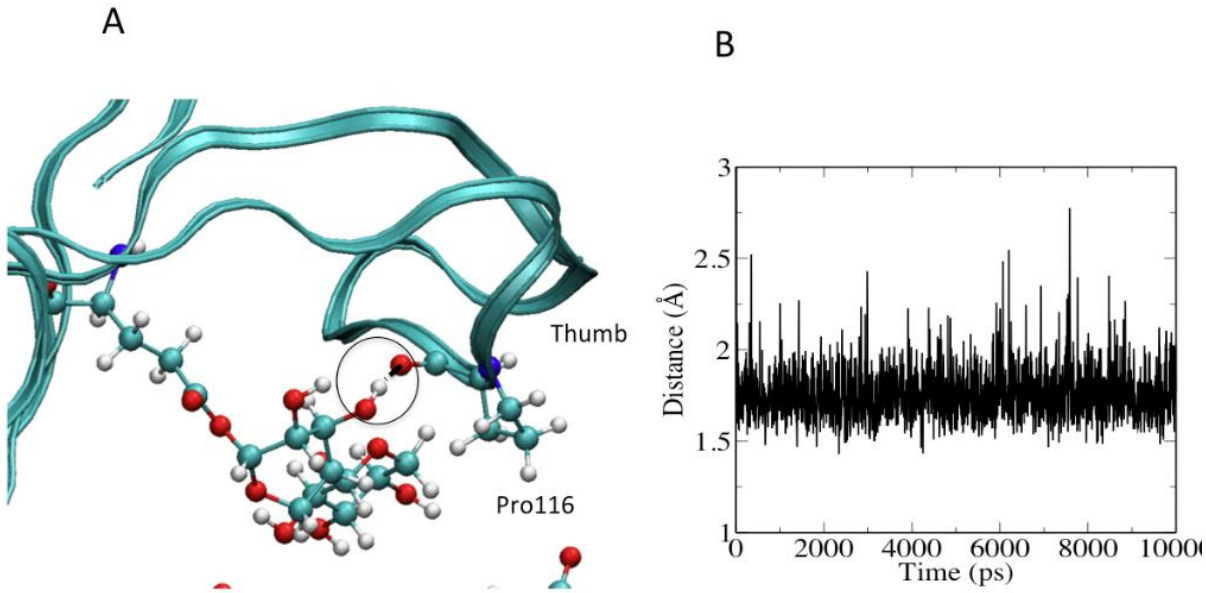

Figure 6. (A) Representative structure for the hydrogen bond between the substrate and Pro116 in the covalently bonded system, B) The hydrogen bond distance as a function of time graph.

The Asp11-Glu172 distance plot, Figure 7, shows that the covalent system also has, on average, the smallest distance followed by the non covalent, with the free enzyme having a similar but narrower distribution on the higher end of the distances.

The covalently bonded intermediate displays the largest average range of distances and the free enzyme the smallest, which is roughly the opposite than for the case of the Asp11-Pro116 distance. This indicates that the substrate induces some flexibility to the fingers and this is especially the case for the covalent system where the distribution is the widest. Perhaps this extra flexibility provides a way of inducing the substrate to leave the enzyme. There is no simple molecular explanation for the difference in distances for the different

systems since the substrate (both covalent and non-covalent) is attached or hydrogen bonded to the thumb rather than the fingers. It is likely, however, that changes in the other distances (due to hydrogen bonding interactions between the substrate and the thumb residues) induce changes in this distance in a more subtle way. It should be noted that the difference between the different systems is quite small (the difference in the peaks of the non-covalent and free enzyme is ~ 1 Å compared to ~4 Å for the Asp11-Pro116 distance.

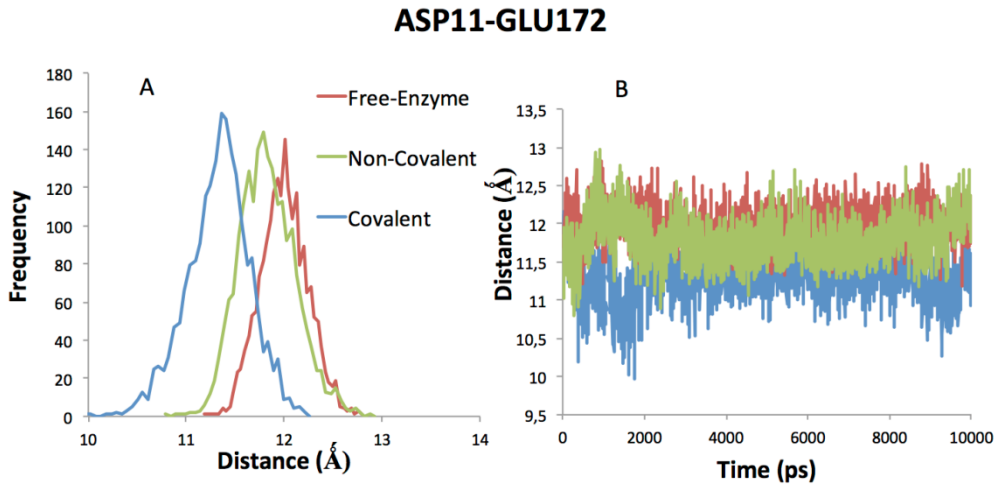

Figure 7. The extent of compression of the fingers region of the enzyme. A) Asp11-Glu172 distance as a function of time and B) a distribution of distances for the free enzyme (red), covalently (blue) and non-covalently bonded (green) systems.

The Pro116-Glu78 distance plots, as shown in Figure 8, indicate that the free enzyme and covalently bonded systems almost exhibit similar distributions (except the covalently bonded system has a narrower distribution at smaller distances). The non-covalently bonded system, however, shows a more complicated distribution; at the start of the simulation the distance varies from 11 to 12 Å, similar to that of the other systems, but switches to larger distances after 2 ns. This is because at the start of the simulation there are hydrogen bonds between hydroxyl hydrogen atoms of the substrate and the amide oxygen atom of Pro116 and the carbonyl oxygen atom of Glu78, constraining this distance and holding the two residues closer to each other (Figure 9).

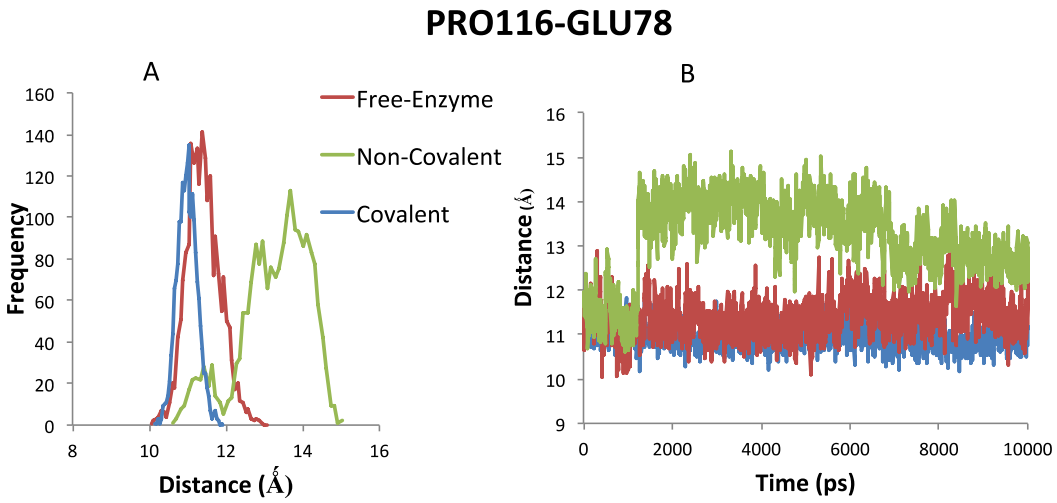

Figure 8. The extent of compression on the thumb region of the enzyme. A) Pro116-Glu78 distance as a function of time and B) a distribution of distances for the free enzyme (red), covalently (blue) and non-covalently bonded (green) systems.

When these hydrogen bonds break, at approximately the same time (~1250 ps), the distance is able to increase (Figure 9B). The increase in the distance at that point, compared with the free system, can probably be explained by steric considerations; that is, the presence of the substrate may push this distance farther apart. We notice later hydrogen bond formation again between the substrate and the thumb residue Pro116 at around 9ns (Figure 9B). For the covalent system, the hydrogen bond between the substrate hydroxyl hydrogen and the amide oxygen of Pro116 (see Figure 6) explains why the distance and range of distances is small; that is, since the other end of the substrate is actually connected to the enzyme, this hydrogen bonding interaction restricts the Pro116-Glu78 distance.

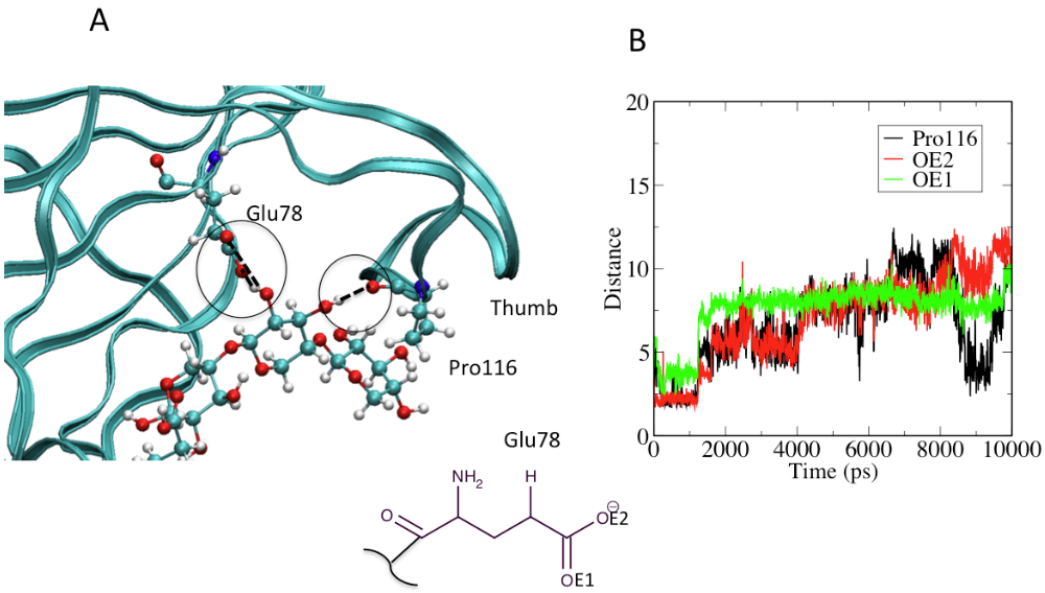

Figure 9. A) Representative structure for the hydrogen bonds between the substrate and Pro116 and the substrate and OE1 and OE2 in the non-covalently bonded system, B) The hydrogen bond distance as a function of time graph.

Analysis of the Pro116-Glu172 cross distances show that the non-covalently bonded system exhibits the largest values followed by the free enzyme, then the covalent system. This order is not surprising considering that the cross distances are related to the other distances; since the Pro116-Glu172 distance is a combination of the Asp11-Pro116 and the pro116-Glu78 cases. The free enzyme dimensions is intermediate between the other two systems since it's Asp11-Pro116 distance is similar to the non-covalent value and it's pro116-Glu78 distance is similar to the covalent value.

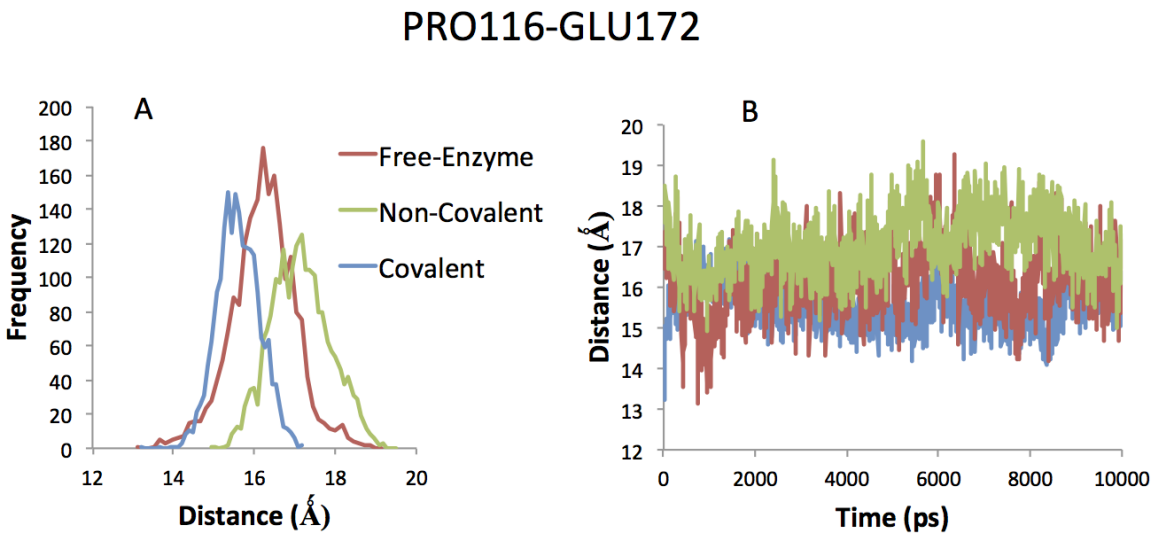

Figure 10. The extent of compression of the thumb region of the enzyme A) Pro116-Glu172 cross-distance as a function of time and B) a distribution of distances for the free enzyme (red), covalently (blue) and non-covalently bonded (green) systems.

Figure 11 shows the Asp11-Glu78 cross-distance. In this case, the covalent distance is the smallest with the free-enzyme at an intermediate position and the non-covalent distance the greatest; this corresponds to the Pro116-Glu78 cross-distances but the distances are more spread out in this case. This distance is a combination of the Asp11-Pro116 and the Pro116-Glu78 distances; both these distances show a considerable separation between non-covalent and covalent systems.

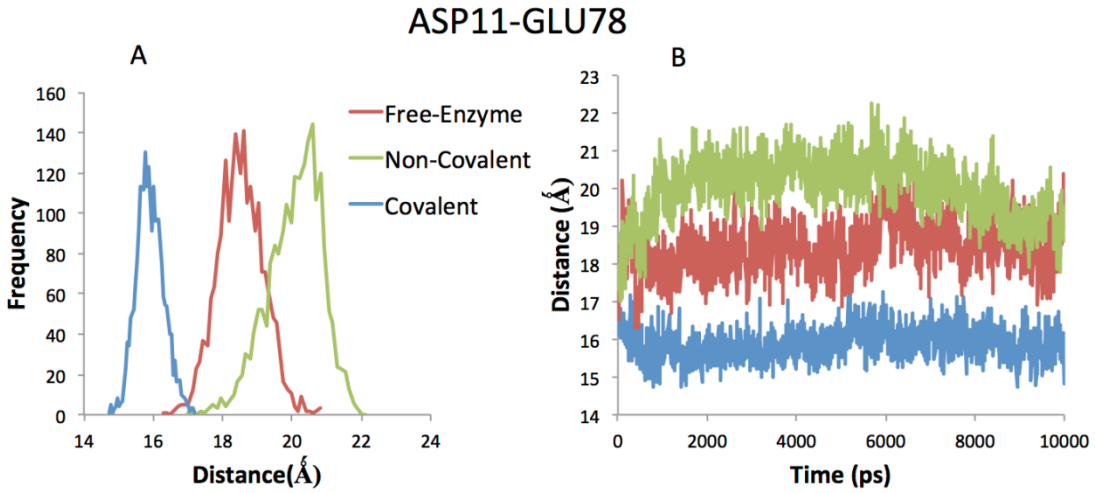

Figure 11. A) Asp11-Glu78 cross distance as a function of time and B) a distribution of distances for the free enzyme (red), covalently (blue) and non-covalently bonded (green) systems.

It should be noticed that in all of the distances reported here, the covalent system is the smallest and the non-covalent the largest. The free enzyme is either intermediate or more similar to the covalent or non-covalent systems. The main reason behind these general orderings are constriction of the thumb by the covalent bond and hydrogen bonds causing a shortening of the distances for the covalent case and steric effects of the non-covalent substrate lengthening the distances.

The work by Paes *et al* (*17*) suggests that the thumb holds the substrate in place and then opens up to facilitate with the expulsion of the product. The results here show that the thumb-finger distance Asp11-Glu172 (Figure 7) and Pro116-Glu78 (Figure 8) cross-distances are all decreased for the covalently bound complex. For the non-covalently bound case, however, the Pro116-Glu78 cross-distance is increased compared to the free enzyme and Asp11-Glu172 shows a similar distribution to the free enzyme. This allows the substrate to enter and the product to exit the active site cleft.

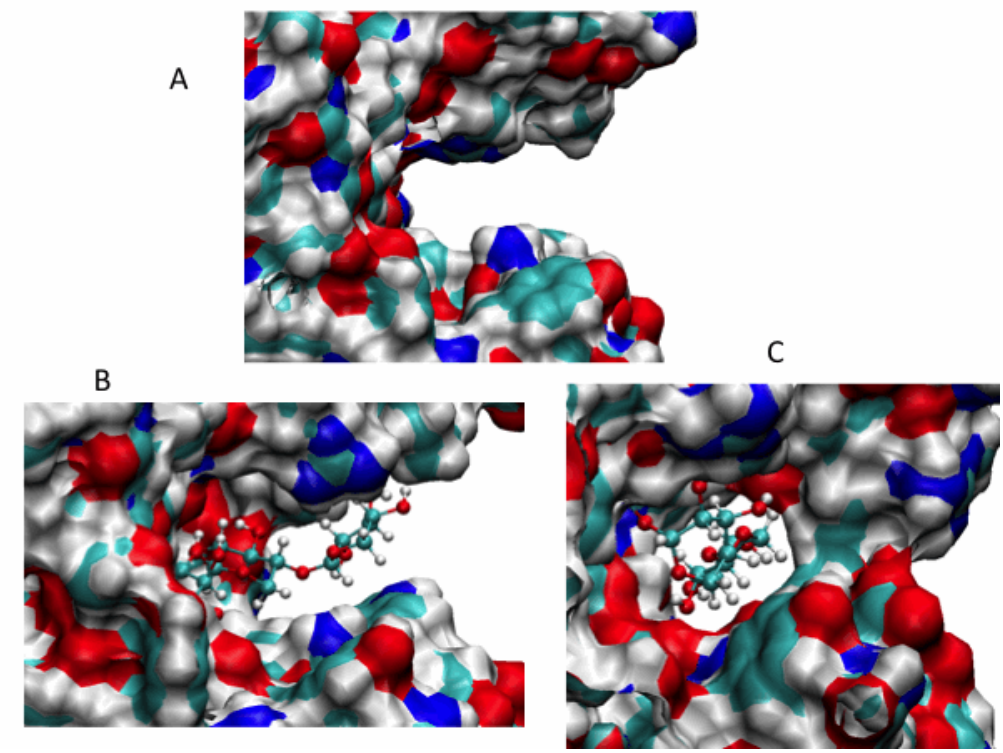

Figure 12. Surface representation of the enzyme in the three systems A) Free enzyme, B) Non-covalent system and C) Covalent system. The probe radius for the surface is 1 Å.

Figure 12 shows a surface representation of the enzyme in the three systems. This Figure provides further insight into the effect of the substrate on the conformation of the enzyme. It can be noticed that the free and non-covalent systems are visually the same; however the covalent system is visually much more closed, where the finger and thumb surfaces make contact. This agrees with the previous distance measures, where the thumb-finger distance is much smaller for the covalent system than the free and non-covalent systems. Furthermore, Figure 12C shows that for the covalent system there is effectively only a small cavity for the substrate to fit.

4. Conclusions

The structure and dynamics of β-1,4-xylanase from *B. circulans* interacting with the xylose disaccharide substrate was studied by all-atom molecular dynamics using the Amber, *ff99SB* force field. Three separate 10 ns simulations at different stages of the glycosylation process

were performed. The free enzyme was modelled to provide information about the dynamics before interaction with the substrate. The enzyme with the substrate, non-covalently, associated between the fingers and thumb of the active site cleft was modelled to give information about the initial approach of the substrate and the substrate. Then, the substrate covalently bound to the active site, was modelled to represent the known intermediate in the glycosolation process. The root mean squared fluctuation (RMSF) of each residue showed that in the important 'thumb' region of the enzyme, the dynamics were significantly reduced due to the presence of the substrate both covalently and non-covalently bound to the enzyme. Furthermore, the covalently bound substrate caused tightening of the active site and the non-covalently bound substrate caused relaxation compared with the free enzyme. It is postulated that this closing of the thumb-finger distance holds the covalently bound substrate within the cleft. This is coupled with the results of the Asp11-Glu172 and the Pro116-Glu78 distances which demonstrate tightening of the covalently bound system. This can be visualized as the fingers and thumb curling around the substrate further holding it close to the active site. This behaviour is expected to speed up the hydrolysis of the covalently bonded sugar moiety for ultimate expulsion as the hydrolysed sugar product.

Calculations have also shown a similarity in behaviour between the covalent system and the non-covalent system for the first 2ns. The hydrogen bond in the non-covalent system acts similar to the covalent bond in restricting the compression, in other words causing the thumb to 'curl.' It does not, however, affect the opening and closing movement, the finger-thumb distance. Thereafter the covalent bond is maintained whereas the hydrogen bond of the non-covalent system breaks after 2ns allowing less compression in the non-covalent system.

This work is the first investigation into the detailed dynamics and structure of the enzyme in response to the substrate at different stages of the glycosylation process. This information, combined with the existing experimental information, assisted us to shine light on the factors affecting the GH catalytic process. Moreover, it can potentially assist with the design of new compounds that can interfere with the natural dynamics of the thumb-finger regions leading to potential GH inhibitors.

Acknowledgment

This research was financially supported by School of Health Sciences, UKZN. Authors also acknowledge the computational facility provided by the Centre of high Performance Computing, CHPC (www.chpc.ac.za).

References

1. Henrissat, B. (1991) A Classification of Glycosyl Hydrollases Based on Amino-Acid-Sequence Similarities, *Biochem J 280*, 309-316.
2. Davies, G., and Henrissat, B. (1995) Structures and Mechanisms of Glycosyl Hydrolases, *Structure 3*, 853-859.
3. Asano, N. (2003) Glycosidase Inhibitors: Update and Perspectives on Practical Use, *Glycobiology 13*, 93-104.
4. Moscona, A. (2005) Drug Therapy- Neraminidase Inhibitors for Influenza, *N Engl J Med*, 11.
5. Joshi, M. D., Sidhu, G., Pot, I., Brayer, G. D., Withers, S. G., and McIntosh, L. P. (2000) Hydrogen Bonding and Catalysis: A Novel Explanation for How a Single Amino Acid Substitution Can Change the Ph Optimum of a Glycosidase, *J mol biol 299*, 255-280.
6. Sidhu, G., Withers, S. G., Nguyen, N. T., McIntosh, L. P., Ziser, L., and Brayer, G. D. (1999) Sugar Ring Distortion in the Glycosyl-Enzyme Intermediate of a Family G/11 Xylanase, *Biochemistry 38*, 5346-5355.
7. Wakarchuk, W. W., Campbell, R. L., Sung, W. L., Davoodi, J., and YAGUCHI, M. (1994) Mutational and Crystallographic Analyses of the Active Site Residues of the Bacillus Circulans Xylanase, *Protein sci 3*, 467-475.
8. Soliman, M. E. S., Ruggiero, G. D., Pernia, J. J. R., Greig, I. R., and Williams, I. H. (2009) Computational Mutagenesis Reveals the Role of Active-Site Tyrosine in Stabilising a Boat Conformation for the Substrate: Qm/Mm Molecular Dynamics Studies of Wild-Type and Mutant Xylanases, *Org Biomol Chem 7*, 460-468.
9. Soliman, M. E. S., Pernia, J. J. R., Greig, I. R., and Williams, I. H. (2009) Mechanism of Glycoside Hydrolysis: A Comparative Qm/Mm Molecular Dynamics Analysis for Wild Type and Y69f Mutant Retaining Xylanases, *Org. Biomol. Chem.*, 5236-5244.
10. Akke, M., Skelton, N. J., Kördel, J., Palmer, A. G., and Chazin, W. J. (1993) Effects of Ion Binding on the Backbone Dynamics of Calbindin D9k Determined by N Nmr Relaxation, *Biochemistry 32*, 9832-9844.
11. Stivers, J. T., Abeygunawardana, C., Mildvan, A. S., and Whitman, C. P. (1996) N-15 Nmr Relaxation Studies of Free and Inhibitor-Bound 4-Oxalocrotonate Tautomerase: Backbone Dynamics and Entropy Changes of an Enzyme Upon Inhibitor Binding, *Biochemistry 35*, 16036-16047.
12. Mandel, A. M., Akke, M., and Palmer, A. G. (1995) Backbone Dynamics of Escherichia-Coli Ribonuclease-H1 - Correlations with Structure and Function in an Active Enzyme, *J Cell Biochem*, 29-29.
13. Gerstein, M., Lesk, A. M., and Chothia, C. (1994) Structural Mechanisms for Domain Movements in Proteins, *Biochemistry 33*, 6739-6749.
14. Faber, H. R., and Matthews, B. M. (1990) A Mutant T4 Lysozyme Displays Five Different Crystal Conformations, *Nature Struct Biol 348*, 263-266.
15. Törrönen A, Anu Harkki, and Juha Rouvinen. (1994) Three-Dimensional Structure of Endo-1,4- B-Xylanase 11 from Trichoderma Reesei: Two Conformational States in the Active Site, *EMBO J 13*, 2493-2501.
16. Miao, S. C., Ziser, L., Aebersold, R., and Withers, S. G. (1994) Identification of Glutamic-Acid-78 as the Active-Site Nucleophile in Bacillus-Subtilis Xylanase Using Electrospray Tandem Mass-Spectrometry, *Biochemistry 33*, 7027-7032.

17. Paës, G., Tran, V., Takahashi, M., Boukari, I., and O'Donohue, M. J. (2007) New Insights into the Role of the Thumb-Like Loop in Gh-11 Xylanases, *Protein Eng Des Sel 20*, 15-23.
18. Muilu, J., Törrönen, A., Peräkylä, M., and Rouvinen, J. (1998) Functional Conformational Changes of Endo-1,4-Xylanase Ii from Trichoderma Reesei: A Molecular Dynamics Study, *Proteins:Struc Funct Gen 31*, 434-444.
19. Kempf, D. J., Norbeck, D. W., Codacovi, L. M., Wang, X. C., Kohlbrenner, W. E., Wideburg, N. E., Paul, D. A., Knigge, M. F., Vasavanonda, S., Craigkennard, A., Saldivar, A., Rosenbrook, W., Clement, J. J., Plattner, J. J., and Erickson, J. (1990) Structure-Based, C2 Symmetrical Inhibitors of Hiv Protease, *J Med Chem 33*, 2687-2689.
20. Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M. J., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber Biomolecular Simulation Programs, *J Comput Chem*, 1668-1688.
21. Frisch MJ, T. G., Schlegel HB, Scuseria GE, Robb MA,, Cheeseman JR, M. J., Vreven T, Kudin KN, Burant, JC, M. J., Iyengar SS, Tomasi J, Barone V, Mennucci B,Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada, M, E. M., Toyota K, Fukuda R, Hasegawa J, Ishida M,, Nakajima T, H. Y., Kitao O, Nakai H, Klene M, Li X, Knox, JE, H. H., Cross JB, Bakken V, Adamo C, Jaramillo J,, Gomperts R, S. R., Yazyev O, Austin AJ, Cammi R,, Pomelli C, O. J., Ayala PY, Morokuma K, Voth GA,, Salvador P, D. J., Zakrzewski VG, Dapprich S, Daniels, AD, S. M., Farkas O, Malick DK, Rabuck AD, Raghavachari, K, F. J., Ortiz JV, Cui Q, Baboul AG, Clifford S,, Cioslowski J, S. B., Liu G, Liashenko A, Piskorz P,, Komaromi I, M. R., Fox DJ, Keith T, Al-Laham MA, Peng, CY, N. A., Challacombe M, Gill PMW, Johnson B,, and Chen W, W. M., Gonzalez C, Pople JA. (2004) *Gaussian Inc, Wallingford, CT*.
22. Cieplak, P., Cornell, W. D., Bayly, C., and Kollman, P. A. (1995) Application of the Multimolecule and Multiconformational Resp Methodology to Biopolymers: Charge Derivation for DNA, Rna, and Proteins, *J Comput Chem 16*, 1357-1377.
23. Wang, J., Wolf, R. F., Caldwell, J. W., Kollman, P. F., and Case, D. A. (2004) Development and Testing of a General Amber Force Field, *J Comput Chem 25*, 1157-1174.
24. Wang, W., and Kollman, P. A. (2000) Free Energy Calculations on Dimer Stability of the Hiv Protease Using Molecular Dynamics and a Continuum Solvent Model, *J. Mol. Biol. 303*, 567-582.
25. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Wate, *J Chem Phys 79*, 926–935.
26. Król, M., Roterman, I., Piekarska, B., Konieczny, L., Rybarska, J., Stopa, B., and P., S. (2005) Analysis of Correlated Domain Motions in Igg Light Chain Reveals Possible Mechanisms of Immunological Signal Transduction, *Proteins: Struct Funct Bioinf 59*, 545-554.
27. Yin, J., Bowen, D., and Southerland, W. M. (2006) Barnase Thermal Titration Via Molecular Dynamics Simulations:Detection of Early Denaturation Sites, *J. Mol. Graphics. Modell. 24*, 233-243.

Chapter 5

5. Computational binding free energy insight into 1,4-β-xylanase

5.1. Introduction

The binding of enzymes to their substrate is done in a very specific manner, this is biologically important feature. Binding energies provides a useful and significant link between the structure and function of biomolecular systems. Being able to predict the strength of non-covalent interactions, binding free energies and the structure of molecular complexes, provides a better understanding of biomolecular interactions and the ability to optimize them.

Free binding energy studies have been used for many areas in biology, such as protein structure determination, drug design and to examine protein-ligand binding (*1-5*). These calculations use MD or MC simulations (*6*) to compute the changes in free energy usually associated with binding (*7*). MMPBSA and MMGBSA have proven to be versatile tools for calculating binding energies (*8-10*). The absolute free energies of all bound and unbound reference states are estimated in these methods.

Mutational studies combined with X-ray crystallographic studies of enzyme-substrate complexes have provided detail description of the active site of this enzyme. Active site residues include two glutamic acid residues Glu78 and Glu172 that are mainly involved in catalysis with Glu78 acting as a neucleophille and Glu172 playing a dual role of an acid/base catalyst(*11*). Arg112 also plays a role in catalysis, Tyr69 and Tyr80 are involved in substrate binding.

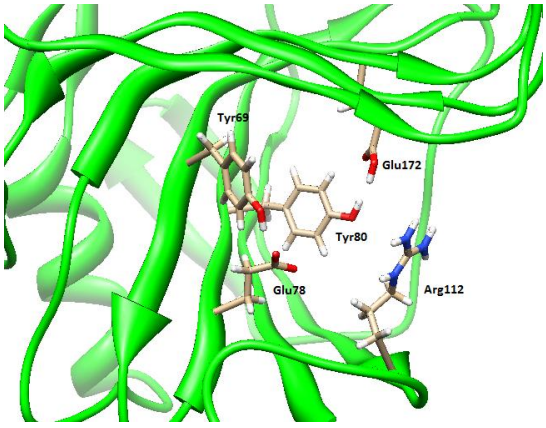

Figure 1: The active site of the *β Circulans* xylanase.

Studies(*11*) probing the effect of altering the distance between the two catalytic residues has revealed that shortening the neuceophilic side chain (Glu78Asp) decreased k_{cat}/K_m values at least 1600 fold, whereas extending length side chain (Carboxymethylated Glu78Cys) decreased k_{cat}/K_m values by 16-100 fold.

Also mutation of the conserved (Tyr80Phe) resulted in a dramatic loss of activity leaving only 0.03%residual activity and mutation (Tyr69Phe) caused a total loss of enzyme activity.

Mutation of (Arg112Asn) results in 68% loss in specific activity.

Recent studies have investigated the role of Tyr69, and showed that it plays a key role in sugar ring distortion, stabilizing the boat as compared to the chair conformation with a 20 $kJ.mol^{-1}$ difference in relative free energy (*12*). Additionally 2D potential mean force calculations of the BCX wild type was compared to Tyr69Phe mutant, which indicated that it also lowers the free energy barrier for the glycosylation step by 42 $kJ.mol^{-1}$ (*13*).

This study aims to provide further insight on the role of residues Glu78, Glu172, Tyr69, Tyr80, and Arg112 in the BCX enzyme non-covalently bound to the substrate, on substrate binding and catalytic activity using binding free energy calculations.

5.2.Materials and methods

MD simulations were performed using Amber10 software at an all-atom MM level on the wild-type as well as Glu78Asp, Glu172Aps, Tyr69Phe, Tyr80Phe and Arg112Asn mutants. Mutations were performed to each of the residues of interest using Swiss-pdbView. For us to quantitatively describe the influence of the mutations, MM-PBSA and MM-GBSA methods (*8-10*), available on the Amber10 package was used to calculate binding free energies.

5.2.1. System set up

The X-ray crystallographic structure of the covalent enzyme-inhibitor complex of wild-type BCX was obtained from the Protein Data Bank (PDB accession code 1BVV)(*14*).

The enzyme was modified and prepared as described in our previous publication (*12*).

Due to the importance of protonation of the Glu172 residue a proton was added using LEaP module on Amber10 (*15*). Charges for the substrate xylose were calculated using AM1-BCC procedure (*12, 16, 17*) after minimizing the molecule at the AM1 semi-empirical level. GAFF

force field (*18*) parameters and AM1-BCC partial charges were assigned using the *Antechamber* module in the Amber10 package (*19*). Missing hydrogen atoms were added using the LEaP module. The *ff99SB* force field (*20*) with the TIP3P water model was used for all simulations. The system was solvated with cubic boxes of TIP3P water with side-length 8 Å. An appropriate number of Cl^- counter ions were added to neutralize the net positive charge on the system. Periodic boundry conditions were used to avoid edge effects. The system was equilibrated using the canonical (NVT) ensemble and, the SHAKE algorithm was used to constrain all bonds involving hydrogen.

5.2.2. Molecular dynamic simulations

Structures were optimized using Sybyl prior to minimization to remove any bad contacts in the structure. The system is minimized in four phases. First, all water molecules are relaxed with a gradient minimizer, while keeping the protein structure frozen; next the system was subjected to a short MD pre-equilibration using mild constraining force to maintain the desired interactions between the substrate and the catalytic residues. Then, the system was heated to 300 K at intervals of 50 K over 10 ps simulations runs with a 1 fs time step. In the last phase the MD simulation is run for 10 ns with a time step of 2 fs.

5.2.3. Parameters and matrics

A recent evaluation of the performance of MMGBSA and MMPBSA for predicting binding energies based on MD simulations has shown that in predicting absolute binding free energy MMPBSA proved to be a better method, however this wasn't true for the relative binding energies (*9*). Interestingly, GB methods demonstrated greater efficiency and accuracy in continuum solvation models for drug-like molecule (*21*).

Free binding energy was therefore calculated using Poisson-Boltzmann-MMPBSA as well as Generalized-Born- MMGBSA (*9, 22, 23*) methods supplied with Amber10 package.

In MMPBSA or MMGBSA (ΔG_{bind}) between a ligand (L) and a receptor (R) to form a complex (RL) is calculated as:

$$\Delta G_{bind} = G_{cpx} - \left(G_{rec} + G_{lig}\right)$$

Eq. 1

In which, the binding free energy ΔG_{bind} ,is evaluated as follows:

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{sol} - T\Delta S$$

Eq. 2

ΔG_{bind} was considered as a sum of changes in the molecular mechanic (MM) gas-phase binding energy $\Delta E_{MM,}$ solvation energy $\Delta G_{sol}$and entropy term ($T\Delta S$). The ΔE_{MM} is calculated by:

$$\Delta E_{MM} = \Delta E_{int} + \Delta E_{vdw} + \Delta E_{ele}$$

Eq. 3

ΔE_{MM} is further divided into the amber force field internal energy terms (bind, angle and torsion), ΔE_{int}: the non-covalent Van der Waals, ΔE_{vdw}; and electrostatic energies component, $\Delta E_{ele.}$

In MMGBSA method, the solvation free energy can be calculated as follows:

$$G_{sol} = G_{GB} + G_{nonpolar}$$

Eq. 4

$$G_{nonpolar} + \gamma SASA + b$$

Eq. 5

G_{GB} is the polar solvation contribution calculated by solving the GB equation (*24, 25*), $G_{nonpolar}$ is the non-polar solvation contribution and is estimated by the solvent accessible surface area (SASA), which was determined using water probe radius of 1.4Å. The surface tension constant γ was set to 0.0072 kcal.mol^{-1} and *b* to 0 kcal.mol^{-1} (*26*).

For the calculations we have chosen a number of 200 snapshots evenly spaced over the 2 ns trajectories with an interval of 10 ps.

5.3. Results and discussion

The following table shows the preliminary results obtained from our calculations:

	Activity (% of wild type)*	MM (kcal/mol)	QM/MM (AM1)
Wild			
GBSA	100	-39.31	-45.35
PBSA		-23.14	-29.70
Tyr69Phe			
GBSA	<0.01	-31.88	-39.40
PBSA		-19.93	-18.94
Arg112Asn			
GBSA	35	-32.17	-44.89
PBSA		-19.18	-27.18
Tyr80Phe			
GBSA	0.045	-34.59	-42.49
PBSA		-18.38	-26.01
Glu78Asp			
GBSA	0.043	-36.40	Results not obtained
PBSA		-20.51	
Glu172Asp			
GBSA	0.9	-42.98	-47.46
PBSA		-22.02	-32,1

*k_{cat}/K_m

Table 1: Binding free energy results for BCXwild type and mutants using GBSA and PBSA.

Kcat is the rate constant for the reaction, while Km is the concentration of substrate required for efficient catalysis. k_{cat}/K_m represents the ratio measure of enzyme efficiency related to free energy barrier.

It would be scientifically incorrect to relate binding energy to the k_{cat}/K_m ratio as the full free energy profile, the PMF, and importantly the transition state has not been considered. However, the binding energies calculated and shown in Table 1, enable us to understand the binding forces in the enzyme-substrate complex. From the above table as well as previous publications (*12, 13*) we know that Tyr69 via a hydrogen bond, stabilizes the boat conformation required for transition state to occur. Mutation therefore has a direct effect on binding. A similar effect is seen with Tyr80 ans Arg112. On the other hand we notice that

catalytic residues Glu78 and Glu172 do not contribute to the binding of the substrate, however, they are essential for the catalytic mechanism.

Our results may not be consistent with experimental findings since these binding figures are taken at the reactant state and may differ from those taken at transition state and during the product phase of the reaction.

References

1. Homeyer, N., and Gohlke, H. (2012) Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method, *Mol. Inform. 31*, 114-122.
2. Brandsdal, B. O., Österberg, F., Almlöf, M., Feierberg, I., Luzhkov, V. B., and Åqvist, J. (2003) Free Energy Calculations and Ligand Binding, *Adv. Protein Chem.*, 123-158.
3. de Ruiter, A., and Oostenbrink, C. (2011) Free Energy Calculations of Protein-Ligand Interactions, *Curr. Opin. Chem. Biol. 15*, 547-552.
4. Mobley, D. L., Graves, A. P., Chodera, J. D., McReynolds, A. C., Shoichet, B. K., and Dill, K. A. (2007) Predicting Absolute Ligand Binding Free Energies to a Simple Model Site, *J. Mol. Biol. 371*, 1118-1134.
5. Boyce, S. E., Mobley, D. L., Rocklin, G. J., Graves, A. P., Dill, K. A., and Shoichet, B. K. (2009) Predicting Ligand Binding Affinity with Alchemical Free Energy Methods in a Polar Model Binding Site, *J. Mol. Biol. 394*, 747-763.
6. Wang, J., Hou, T., and Xu, X. (2006) Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models, *Curr. Compu.t Aided Drug Des. 2*, 287-306.
7. Gilson, M. K., Given, J. A., Bush, B. L., and McCammon, J. A. (1997) The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review, *Biophys. J. 72*, 1047-1069.
8. Miller, B. R., III, McGee, J. T. D., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012) Mmpbsa.Py: An Efficient Program for End-State Free Energy Calculations, *J. Chem. Theory Comput. 8*, 3314-3321.
9. Hou, T., Wang, J., Li, Y., and Wang, W. (2011) Assessing the Performance of the Mmpbsa and Mmgbsa Methods: I. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamic Simulations, *J. chem. inf. model 24*, 69-82.
10. Stoica, I., Sadiq, S. K., and Coveney, P. V. (2008) Rapid and Accurate Prediction of Binding Free Energies for Saquinavir-Bound Hiv-1 Proteases, *J. Am. Chem. Soc. 130*, 2639-2648.
11. Wakarchuk, W. W., Campbell, R., Sung, W., Davoodi, J., and Yaguchi, M. (1994) Mutational and Crystallographic Analyses of the Active Site Residues of the Bacillus Circulans Xylanase, *Protein Sci. 3*, 467-475.
12. Soliman, M. E. S., Ruggiero, G. D., Pernia, J. J. R., Greig, I. R., and Williams, I. H. (2009) Computational Mutagenesis Reveals the Role of Active-Site Tyrosine in Stabilising a Boat Conformation for the Substrate: Qm/Mm Molecular Dynamics Studies of Wild-Type and Mutant Xylanases, *Org. Biomol. Chem. 7*, 460-468.
13. Soliman, M. E. S., Pernia, J. J. R., Greig, I. R., and Williams, I. H. (2009) Mechanism of Glycoside Hydrolysis: A Comparative Qm/Mm Molecular Dynamics Analysis for Wild Type and Y69f Mutant Retaining Xylanases, *Org. Biomol. Chem.*, 5236-5244.
14. Sidhu, G., Withers, S. G., Nguyen, N. T., McIntosh, L. P., Ziser, L., and Brayer, G. D. (1999) Sugar Ring Distortion in the Glycosyl-Enzyme Intermediate of a Family G/11 Xylanase, *Biochemistry 38*, 5346-5355.
15. Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber Biomolecular Simulation Programs, *J. Comput. Chem. 26*, 1668-1688.
16. Jakalian, A., Bush, B. L., Jack, D. B., and Bayly, C. I. (2000) Fast, Efficient Generation of High-Quality Atomic Charges. Am1-Bcc Model: I. Method, *J. Comput. Chem. 21*, 132-146.

17. Jakalian, A., Bush, B. L., Jack, D. B., and Bayly, C. I. (2000) Fast, Efficient Generation of High-Quality Atomic Charges. Am1-Bcc Model: I Method, *J. Comput. Chem. 21*, 132-146.
18. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and Testing of a General Amber Force Field, *J. Comput. Chem. 25*, 1157-1174.
19. Crowley, M., Walker, R. C., Zhang, W., Merz, K. M., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., and Kollman, P. A. (2008) Amber10 Manual.
20. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters, *Proteins: Struct. Funct. Bioinform. 65*, 712-725.
21. Kongsted, J., Söderhjelm, P., and Ryde, U. (2009) How Accurate Are Continuum Solvation Models for Drug-Like Molecules., *J. Comput. Aided Mol. Des. 23*, 395-409.
22. Shen, M., Zhou, S., Li, Y., Pan, P., L., Z., and Hou, T. (2013) Discovery and Optimization of Triazine Derivatives as Rock1 Inhibitors: Molecular Docking, Molecular Dynamics Simulations and Free Energy Calculations, *Mol. Biosyst. 9*, 361-373.
23. Hayes, J. M., and Archontis, G. (2012) Mm-Gb(Pb)Sa Calculations of Protein-Ligand Binding Free Energies *http://www.intechopen.com*, 21.
24. Feig, M., and Brooks, C. L. (2002) Evaluating Casp4 Predictions with Physical Energy Functions, *Proteins-Struct. Funct. Bioinf. 49*, 232-245.
25. Gourmala, C., Luo, Y., Barbault, F., Zhang, Y., Ghalem, S., Maurel, F., and Fan, B. (2007) Elucidation of the Lewisx-Lewisx Carbohydrate Interaction with Molecular Dynamics Simulations: A Glycosynapse Model, *J. Mol. Struct. (Theochem) 821*, 22-29.
26. Sitkoff, D., Sharp, K. A., and Honig, B. (1994) Accurate Calculation of Hydration Free-Energies Using Macriscopic Solvent Models, *J. Phys. Chem. 98*, 1978-1988.

Chapter 6

6. General conclusions and recommendations for future studies

6.1. General conclusions

The aim of this work was to study the dynamics of the β-1,4-xylanase enzyme and investigate the effect of the substrate in the active site on changes in its conformation in terms of the widening and narrowing of the distance between the thumb and finger region. This was done by performing three separate simulations at the different stages of the glycosylation process. To a large extent, our objectives have been accomplished. Results have confirmed the following:

1. The covalently bound substrate caused a tightening of the active site, while the non-covalently bound substrate caused relaxation when compared to the free enzyme.
2. Cross distances between the thumb and finger with the catalytic residue at the opposite end, demonstrate a tightening in the covalently bound system, this is probably to hold the substrate in the active site for the purpose of speeding up hydrolysis.

6.2. Recommendation and future studies

One of the major challenges in this work is the lack of experimental data against which we can compare our computational results. However, our results clearly contribute towards the molecular understanding of the dynamic and structure of the enzyme in response to the substrate. Recently developed computational approaches such as Principle Component Analysis (PCA)(*1*), Residue Interaction Network (RIN)(*2*) Coarse grained Molecular Dynamics (*3*), and Substrate Envelope Analysis (SEA) (*4*) would certainly provide more detail on the dynamics and conformational preferences as well as the drug-enzyme networks of interaction.

References

1. Haider, S., Parkinson, G., and Neidle, S. (2008) Molecular Dynamics and Principle Components Analysis of Human Telomeric Quadruplex Multimers, *J Biophys 95*, 296-311.
2. Doncheva, N. T., Klein, K., Domingues, F. S., and Albrecht, M. (2011) Analysing and Visualizing Residue Networks of Protein Structures, *Trends Biochem. Sci 36*, 179-182.
3. Chu, J. W., and Voth, G. (2007) Course-Grained Free Energy Functions for Studying Protein Conformational Changes: A Double-Well Network Model, *J Biophys 93*, 3860-3871.
4. Altman, M. D., Ali, A., Kumar Reddy, G. S. K., Nalam, M. N. L., Anjum, S. G., Cao, H., Chellappan, S., Kairys, V., Fernandes, M. X., Gilson, M. K., Schiffer, C. A., Rana, T. M., and Tidor, B. (2008) Hiv-1 Protease Inhibitors from Inverse Design in the Substrate Enveloipe Exhibit Subnanomolar Binding to Drug-Resistant Variants *J. Am. Chem. Soc. 130*, 6099-6113.