STRUCTURE DYNAMICS GUIDED ENZYME IMPROVEMENT OF ENDO-BETA-1, 4-XYLANASE I

A Dissertation

by

UGUR UZUNER

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Joshua S. Yuan
Co-Chair of Committee,	James L. Starr
Committee Members,	Dennis C. Gross
	Mike V. Kolomiets
	Susie Y. Dai
Head of Department,	Leland S. Pierson III

August 2013

Major Subject: Plant Pathology

Copyright 2013 Ugur Uzuner

OZET

ENDO-BETA-1, 4-KSILANAZ I ENZIMININ YAPI DINAMIKLERI REHBERLIGINDE GELISTIRILMESI

Enzim yapı dinamiklerinin, yapı-fonksiyon ilişkisi bakımından önem derecesi son yıllarda gerçekleştirilen çalışmalarla ortaya konmuştur. Degişik yapı dinamiklerine özgü analiz platformları arasında, Hidrojen Deuterium Değişimi Kütle Spektrometrisi; (HDX-MS) ligand bağlanması, protein bağlanması ve enzim katalizi üzerinden protein dinamiklerini açıklamak için etkili ve yüksek işlem hacmine sahip metod olarak göze çarpmaktadır. Hidrojen deuterium kütle spektrometrisi, HDX deneylerindeki ayrışmış peptidler için deuterium verilerine ait yüzde degerleri veya m/z değerine dayalı olarak proteinlerin bölgesel dinamiklerini belirlemek ve çalışmak icin kullanılabilmektedir.

Şimdiye kadar HDX kütle spektrometrisi verilerini analiz etmek için farklı yazılım paketleri geliştirildi. Buna rağmen kütle spektrometrisi verilerinin doğru, ileri düzeyde gelişmiş ve belirgin istatistiki analizi pratikte mevcut değildi. Bu eksiklikleri gidermek adına HDXanalyzer programı mevcut çalışma dahilinde geliştirilmiştir. HDXanalyzer yazılımının HDX-MS analizi açısından yeterliliği, enzim yapı dinamiklerini tanımlamak adına, *Trichoderma longibrachiatum*'dan model kataliz endoksilanaz A (XYN I) kullanılarak gerçekleştirildi. XYN I enziminin iki farklı ksilanaz substrati ile etkileşimi HDXanalyzer programıyla değerlendirilmiş ve HDX-MS analiz sonuçları oldukça yüksek derecede farklılaşmış dinamik değişikliklerin varlığını ortaya koymuştur.

XYN I için elde edilmiş ve protein mühendisliği çabaları için hedeflenebilir bölgeleri belirgin şekilde sınırlandıran dinamik veriler, XYN I enziminin katalizi için önemli yapısal modifikasyonlar hakkında kullanışlı bilgiler sağlamıştır. Elde edilmiş bu kullanışlı bilgiler, XYN I enziminin belirli karakteristiklerini geliştirmek için enzim mühendisliği çalışmalarında kullanılmıştır. HDX-MS verilerine dayalı olarak XYN I enziminin yüksek düzeyde stabilizasyonu gerçekleştirildi ve oldukça aktif ve kabul edilebilir derecede ısıya dayanıklı iki XYN I rekombinantı başarıyla geliştirildi.

HDX-MS yoluyla yapısal açıdan ileri derecede benzerlik gösteren iki katalizin karşılaştırmalı dinamik analizleri de mevcut çalışma kapsamında gerçekleştirilmiştir. Fonksiyon ve amino asit sekansları farklılık göstermesine rağmen yapısal açıdan oldukça benzer XYN I ve endoglukanaz (Eg1A) enzimleri, birbirinden oldukça farklı yapı değişiklikleri göstermiştir. XYN I ile karşılaştırıldığında, *Aspergillus niger* mantarından Eg1A enzimi oldukça sınırlı yapısal pozisyon değişimleri göstermiştir. Elde edilen bulgular açıkça göstermiştir ki enzimatik kataliz esnasında meydana gelen içsel hareketlenmeler, tepkimeleri katalizlemek için her durumda tek ve yeterli etken değildir.

Sonuç olarak, enzimatik reaksiyonların mevcut biyokimyasal ve biyofiziksel bilgilerine, protein yapı dinamiklerinin bilgisinin entegre edilmesi enzim mühendisliği ve enzim karakteristiklerinin geliştirilmesi çalışmaları için kullanışlı öngörüler sağlayabilmektedir.

ABSTRACT

Enzyme structure dynamics has recently been revealed to be essential for structure-function relationship. Among various structure dynamics analysis platforms, hydrogen deuterium exchange mass spectrometry stands as an efficient and high-throughput way to analyze protein dynamics upon ligand binding, protein folding, and enzyme catalysis. HDX-MS can be used to study the regional dynamics of proteins based on the m/z value or percentage of deuterium incorporation for the digested peptides in the HDX experiments.

Various software packages have been developed to analyze HDX-MS data. However, for the accurate, enhanced, and explicit statistical analysis of HDX-MS data HDXanalyzer software was developed.

The capability of HDX-MS analysis for the identification of enzyme structure dynamics was tested by using model catalysis endoxylanase A (XYN I) from *Trichoderma longibrachiatum*. The HDX data of XYN I revealed a highly dynamic personality of XYN I through the interaction with two substrates. The dynamic data which certainly restricts the targeted regions for the protein engineering efforts provided useful knowledge about the essential structural modifications for the catalysis of XYN I. The obtained knowledge was then employed for the engineering studies in order to improve the certain characteristics of XYN I protein.

The high level stabilization of XYN I protein was gathered and the two highly active and moderately thermostable XYN I recombinants were developed based on the

HDX-MS data which further confirmed the efficiency of the current strategy for the rational designs of catalytic proteins.

A differential dynamics analysis of the two structurally similar catalysts was also performed through HDX-MS. The functionally and sequentially different but structurally highly similar XYN I and endoglucanase (Eg1A) enzymes revealed distinct structure dynamic characteristics. Compared to XYN I, Eg1A from *Aspergillus niger* indicated quite restricted structural motions. The data clearly postulated that the intrinsic dynamic modifications of during the enzymatic catalysis may not be the only driving force in all cases.

In summary, the integration of the structure dynamics knowledge to the current biochemical and biophysical data of catalysts may provide novel insights to further enzyme improvement applications.

DEDICATION

This dissertation is devoted to my family and my parents Sami and Ayse Uzuner, who taught me that life presents many obstacles, and the strongest are not those who never fall, but those who do, and learn how to get back up.

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to my advisor Dr. Joshua S. Yuan, for his invaluable guidance and suggestions, his constant encouragement and financial support, and his understanding, patience, and friendship in my pursuit of a PhD degree. I would like to thank my committee members, Dr. Starr, Dr. Gross, Dr. Kolomiets, and Dr. Dai, for their guidance, support, and valuable inputs throughout the course of this research. I also thank to Dr. Ziyu Dai and Pacific Northwest National Laboratory at Washington State, for their guidance, cooperation, and endless help.

Thanks also go to my friends, especially Dr. Gray W. Shi and Dr. Xing Zhuo, Su Sun, all of my lab members, colleagues, and the department faculty and staff for making my time at Texas A&M University a great experience. I also want to extend my gratitude to the Ministry of National Education of The Republic of Turkey, which provided financial support during my PhD research.

Finally, thanks to all of my family members of, my mother and father for their encouragement, endless support and love. I also would like to thank to my friends, Hakan Sahin, Emre Aksoy, and Derya Yanmis. Every achievement of my graduate study comes with their deep love and encouragement.

NOMENCLATURE

CMC	Carboxymethyl Cellulose
DNS	3,5-Dinitrosalicylic Acid
LB	Luria-Bertani
LC-LTQ	Liquid Chromatography-Linear Quadrupole Ion Trap
LCL	Lower Bound of 100 % Confidence Interval for Individual Prediction
MES	2-(N-Morpholino) Ethanesulfonic Acid
MM	Minimal Medium
Ni-NTA	Nickel-Nitrilotriacetic Acid
OD	Optical Density
PDA	Potato Dextrose Agar
PEG	Polyethylene Glycol
SDS	Sodium Dodecyl Sulfate
UCL	Upper Bound of 100 % Confidence Interval for Individual Prediction

TABLE OF CONTENTS

Page

ABSTRACT iii

DEDICATION vii

ACKNOWLEDGEMENTS viii

NOMENCLATURE ix

TABLE OF CONTENTS x

LIST OF FIGURES xiii

LIST OF TABLES xv

CHAPTER I Introduction and Literature Review 1

1.1. Introduction 1
1.2. Structure and Dynamics Relevance of Enzymes 3
1.3. Engineering Strategies for Biocatalysts 5
1.4. Engineering Strategies to Improve Xylanase Production by *Trichoderma reesei* 7
1.5. The Importance and Phylogeny of *Trichoderma reesei* 9
1.6. Importance of Biocatalyst Improvement Enterprises for Biofuel Applications 11
1.7. Xylanase Production by *T. reesei* 13
1.8. Hydrogen Deuterium Exchange Mass Spectrometry as an Alternative Approach to Guide Enzyme Improvement 16

Chapter II HDX-ANALYZER: A NOVEL PACKAGE FOR STATISTICAL ANALYSIS OF PROTEIN STRUCTURE DYNAMICS* 18

2.1. Introduction 18
2.2. Implementation 23
2.2.1. Data Processing as Implemented by Python 23
2.2.2. Statistical Models and Implementation 25
2.2.3. RPY for Integrating the Different Components 27
2.2.4. User Interface as Implemented by Wxpython 28

Page

2.3. Results and Discussion 28
2.3.1. The Input Data Format and the Usage of the Package 29
2.3.2. The Output of HDXanalyzer 29
2.3.3. The Interpretation and Comparison of Different Statistical Models 33
2.3.4. The Overlay of 3D Structure for Differential Structure Dynamics of Xylanase 35

Chapter III ENZYME STRUCTURE DYNAMICS OF XYLANASE I FROM *TRICHODERMA LONGIBRACHIATUM** 38

3.1. Introduction 38
3.2. Material and Methods 41
3.2.1. Protein and Reagents 41
3.2.2. HDX Experiment 42
3.2.3. Peptide Identification and HDX Data Processing 43
3.2.4. Data Processing for HDX Mass Spectrometry 44
3.2.5. Structure Dynamics of Apo XYN I Revealed by HDX-MS Analysis 46
3.2.6. Differential HDX Analysis of XYN I 47
3.2.7. Statistical Analysis for Differential HDX Data 49
3.2.8. Correlation of HDX Data with X-Ray Structure 50
3.2.9. Comparative Analysis of XYN I Structure Dynamics When Binding with Two Different Substrates 53
3.2.10. HDX Profile and Enzyme Improvement for the Future 54
3.3. Results and Discussion 62

Chapter IV STRUCTURE DYNAMICS GUIDED-ENZYME IMPROVEMENT STUDIES FOR XYN I FROM *TRICHODERMA REESEI* 64

4.1. Introduction 64
4.2. Material and Methods 66
4.2.1. Materials 66
4.2.2. The Introduction of Amino Acid Substitutions into XYN I Coding Sequence 66
4.2.3. The Yeast-Based Construction of Fungal Expression Cassettes 70
4.2.4. *E. coli* Transformation 73
4.2.5. Fungal Transformation 74
4.2.6. Enzyme Expression and Purification 76
4.2.7. Xylanase Activity Assays 79
4.3. Results and Discussion 80

Chapter V DIFFERENTIAL DYNAMICS OF THE TWO STRUCTURALLY SIMILAR BIOCATALYSTS 88

Page

5.1. Introduction 88
5.2. Materials and Methods 91
5.2.1. Protein and Reagents 91
5.2.2. HDX Experiments 91
5.2.3. Peptide Identification and HDX Data Processing 93
5.2.4. Statistical Analysis for Differential HDX Data 94
5.3. Results and Discussion 96
5.3.1. Structure Dynamics of Apo Eg1A 96
5.3.2. Active Site upon Ligand Binding Revealed by HDX 100
5.3.3. Comparison the Dynamic Motions of XYN I and Eg1A 103

Chapter VI CONCLUSIONS AND PERSEPCTIVES 110

References 112

Appendix 128

LIST OF FIGURES

Page

Figure 1. The three-dimensional structure of right hand-shaped XYN I enzyme with its catalytic amino acid residues marked as yellow. 9

Figure 2. The needs for biocatalyst improvement to optimize biorefinery processes.....12

Figure 3. The implementation flow of HDXanalyzer. 23

Figure 4. The overview of the HDXanalyzer user interface. 30

Figure 5. The deuteration-rate related figure of a representative peptide from apo and ligand-bound protein. 31

Figure 6. The output for the statistical analysis window.....32

Figure 7. Verification of data analysis using previously published estrogen receptor data. 35

Figure 8. The overlay of *p* values from the HDX analysis onto the 3D structure of XYN I. 36

Figure 9. The peptides analyzed in the HDX experiment for XYN I (PDB code: 1XYN). 44

Figure 10. The HDX profile of apo XYN I through representative peptides.....45

Figure 11.The mass spectra of XY N I peptide "YTIWENTRVNEPSIQGTAT" cover the residues; 102-129.....46

Figure 12. The deuterium incorporation curves of four XYN I peptides.....48

Figure 13. The HDX profiles overlaid onto the 3D structure of XYN I.. 52

Figure 14. Multiple sequence alignment of homologs of *T. longibrachiatum* xylanase.. 61

Figure 15. Consecutive PCR-based introduction of amino acid substitutions into XYN I. 68

Page

Figure 16. The schematic visualization of the *T. reesei* expression cassettes constructed within yeast cells. 71

Figure 17. High-level expression of XYN I recombinants in *T. reesei* 78

Figure 18. The Ni-NTA superflow-aided purification of XYN I recombinants from the fungal cultures of *T. reesei*.. 79

Figure 19. The enzymatic activity results of XYN I recombinants through different temperatures 82

Figure 20.The pH optima of the two recombinants of XYN I protein.. 83

Figure 21. The residual activity and related half-life results of the two XYN I recombinants as compared to wild type (X12). 84

Figure 22.The specific positions of the E7 recombinant-specific substitutions on the 3D structure of XYN I. 86

Figure 23. The specific positions of the Q4 recombinant-specific substitutions on the 3D structure of XYN I.. 87

Figure 24. Peptides analyzed in the HDX experiment for Eg1A. 99

Figure 25. The differential dynamic characteristics of Eg1A upon binding with various ligands.. 101

Figure 26. Regional deuteration rates of both XYN I and Eg1A through HDX-MS analysis. 104

Figure 27. Globular dynamics of both XYN I and Eg1A revealed by MALDI-TOF analysis. 105

Figure 28. The pairwise sequence alignment of *A. niger* Eg1A and *T. reesei* XYN I together with secondary structure assignment. 106

Figure 29. RMSD-based structure dynamics comparison of both XYN I and Eg1A. .. 107

Figure 30. Multiple sequence alignment for Eg1A indicating the highly stabilized (red boxes) and destabilized regions (blue boxes) in the protein sequence through HDX analysis. 109

LIST OF TABLES

Page

Table 1. Properties of xylanases produced by *T. reesei.* 14

Table 2. The comparison of statistical analysis results from the Paired-students test and multiple regression. 34

Table 3. The HDX analysis of the sequence-specific peptides from XYN I and related statistical data. 55

Table 4. The HDX rates of the specific peptides through XYN I-xylan interaction and related statistical data.....57

Table 5. List of the primers used to engineer XYN I enzyme.....67

Table 6. The list of the XYN I recombinants created. 69

Table 7. Primers operated to build the fungal expression cassettes specific to filamentous fungus *T. reesei.* 71

Table 8. Differential deuteration rates of six notable peptides from Eg1A. 95

Table 9. The deuteration levels and statistical data of the specific peptides from Eg1A in response to various ligands 97

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

1.1. Introduction

Proteins are macromolecules that exist in all biological systems, from lower prokaryotes to higher eukaryotes. They consist of the prominent amounts of living organisms, both quantitatively and qualitatively. The majority of the key biochemical processes in cells and organisms are controlled through the coordinated action of protein molecules [1]. In these multi protein complexes, proteins are involved in a variety of different interactions and form a coordinated network for different cellular signals and most metabolic pathways. The term protein was first pronounced by Jons Jakob Berzelius in 1838 and their prominence for life was attributed to the Greek word "protos", which means first rank of importance [2].

Similar to polysaccharide and nucleic acid molecules, proteins are a major group of catalysts in the cells, highly essential for metabolism, and participate in almost every process within the organelles and cells. In addition, they provide structural and mechanical support to the organisms. Some protein types also are involved in cell signaling, cell cycle, adhesion, and immune responses [1]. For the last three decades, it was firmly believed that the proteins were the only catalytic macromolecules; however,

since then a variety of RNA molecules have also been identified with catalytic activities [3, 4].

Enzymes are the molecules in protein forms and tailored protein catalysts responsible for thousands of chemical reactions that take place naturally. As key players in most biological processes, they catalyze nearly all important chemical reactions in the body as well as being involved in the control of the transcription of genetic information, signal transduction, and cell regulation [5]. Enzymes are well suited to their functions by having highly exquisite catalytic specificities, enormous catalytic powers, and tunable catalytic reactions. They accelerate the rates of chemical reactions by decreasing the catalytic energy barrier through binding and stabilizing the transition states and this acceleration may be as much as 10^{20} fold [6]. They are not consumed or altered during the reaction and do not change the equilibrium, but lessen the reaction time.

The majority of the dynamic metabolic activities in living organisms are mostly regulated by specific proteins and enzymes, hence understanding their roles on these biochemical events is hidden in their structures [7]. Enzymes and proteins are unique macromolecules consisting of a sequence-specific composition of twenty different amino acids. Their complexity and uniqueness are defined by the sequence of the gene from which the protein is produced. The linear amino acid sequences of proteins, which is known as the primary structure, promote all sorts of biological and chemical characteristics such as their catalytic activities, substrate specificities, and stabilities. The primary structure of proteins is produced by ribosomal machinery in the cells based on genetic code and coding of gene transcripts. Through hydrogen (H) bond interactions

among amino acid residues, α- helix, β-sheet, and turns are derived which are specified as the secondary structure of proteins. Overall folded polypeptide chains are called as the tertiary structure of proteins, which harbors the variations of different chemical interactions such as disulfide bridges, ionic and/or hydrophobic interactions, hydrogen bonds, and stemming from side chain R groups located on amino acid molecules. Tertiary structures are the functional units of proteins with the properties more consistent and stable compared to those primary and secondary formations. Certain metabolic reactions in living organisms, however, requires the cooperative action of two or more polypeptide chains or proteins, which may have identical or different amino acid sequences. In this case, the complex generated by several proteins are called quaternary structures of protein molecules [8]. In brief, the ultimate functionality of enzymes and proteins are derived from the cooperative actions of entire chemical interactions occurring during protein folding stages from primary to tertiary structure.

Catalytically reactive atoms embedded within folded enzyme structures are the major key components responsible for the unique feature of chemical catalysis. Unlike their unquestionable importance for biochemical reactions, there are still many unknown factors contributing to enzyme catalysis. To some extent, current technological advancements allow disclosure of the mechanisms of their function through chemical reactions and static three-dimensional structures (3D); however, the dynamic nature of enzymes has recently been proposed to have notable functions in catalysis [9-11]. It is currently well accepted that the intrinsic dynamics of enzymes implement substantial roles in enzymatic catalysis. The term enzyme dynamics is typically attributed to the

conformational flexibility and alterations in the atomic coordinates of enzyme molecules. The atomic displacements of enzyme-structure fluctuations typically occurs in a wide range of timescales changing from femtoseconds to seconds which result in an enzyme molecule having an assortment of different conformational states [12]. Although it is still under debate, the potential enzyme backbone fluctuations such as the movement of flexible loops, rotations of side-chains, and motions of larger domains highlighted up to date may occur at times spanning from femtosecond to millisecond tiers. Some fast timescale (femtoseconds–picoseconds) fluctuations on enzyme molecular machinery may also occur as local collective protein motions, in particular bond stretching, side-chain rotamers, and angle bending within specific atomic groups [12]. Nevertheless, the mechanisms underlying the great majority of the functions of biological macromolecules are most likely driven at least partially by large-scale structural conformational changes.

1.2. Structure and Dynamics Relevance to Enzymes

Understanding the biochemical and biophysical properties and unraveling the functional roles of proteins within the context of their natural environments is still one of the greatest challenges for many disciplines such as biochemistry, biophysics, molecular biology, enzymology, and structural biology. Recent technological advancements towards this goal have been made with the development of both chemical and biochemical tools; however, the knowledge of enzyme-structure and enzyme-function relationship has not yet been well documented.

Catalytic properties of enzymes are deeply associated with both their biochemical and structural characteristics and physically interaction partners. Today, the structure–function relationship of proteins constitutes the main dogma of structural biology. Over hundred years ago, fundamental mechanisms of enzyme catalysis was the subject of intensive research over time and described by Fischer (1894) with a “lock and key” model [13]. As another hypothesis overlapping to some extent with recent dynamic considerations, the induced-fit model introduced by Koshland in 1958 [14], extended the lock and key model by incorporating flexibility of protein structures which introduced a new perspective into substrate recognition and enzyme catalysis. The induced-fit scenario was typically postulated as a conformational change in the enzyme structure through intimate and specific interactions between enzyme and its substrate. In other words, free and ligand-bound protein conformations are altered and induced as a result of binding interaction that triggers the structure of a protein to a new conformation that is more complementary to its binding partner. Unlike the induced-fit model, the conformational selection model as a fresh theory of enzyme dynamics, points out that enzymes and proteins are naturally available in their environments with ensembles of conformational states [15]. Thermally accessible conformational substates of proteins are more favorable for ligand binding rather than its native (the lowest energy) conformation [16]. The conformational selection model posits that weakly populated, higher energy conformations are capable of binding to ligands, indicating the presence of a shifting event for the distribution of conformers towards ligand-bound states. In this scenario, intrinsic dynamics of the protein initiates a spontaneous transition between a

stable unbound and a less stable ligand-bound conformation. The conformational selection model mainly stem from the free energy landscape theory of protein structure and dynamics [16]. A protein free energy landscape defines the existence of different conformations and substates in the dynamic equilibrium which was well addressed by Henzler et al [11]. It is therefore postulated that the small changes in the free energy are responsible for the conformational fluctuations and intrinsic dynamic changes.

All aforementioned models mainly serve the same goal and intend to shed light on the puzzle about the different protein motions, unique protein-ligand binding patterns, and exquisite protein-protein interactions.

1.3. Engineering Strategies for Biocatalysts

The engineering approaches of protein biocatalysts can be primarily performed through several different ways such as directed evolution, computational *de novo* design, and rational design [17-19]. Directed evolution usually employs a batch of molecular biology techniques enabling the imitation of natural evolutionary processes in laboratory environments in order to optimize the specific properties of functional parental proteins. Directed evolution requires the formation of a remarkably large library of randomly distributed variants which later needs to be screened to obtain the desired recombinants of parental enzyme in a labor-intensive and time consuming manner. Directed evolution applications, however, are predominantly the best choice when the knowledge of the protein of interest is highly limited or the protein of interest is typically novel and there

is no further information beyond its amino acid sequence. Although the outcomes from directed evolution and derivatives-based applications [DNA shuffling, site directed saturation mutagenesis (SDSM), error-prone polymerase chain reaction (epPCR), look-through mutagenesis (LTM), iterative saturation mutagenesis (ISM), etc.] are mostly favored and predominantly overcome the results from rational design-based engineering efforts, it often lacks of insights towards the mechanistic knowledge and the functional characteristics of recombinant biocatalysts and does not provide further information to the enzyme improvement efforts [18-21].

To engineer enzymes, the rational design technique typically uses different strategies such as site directed mutagenesis (SDM), domain exchange, secondary element swapping, and protein fusion to obtain recombinants with desirable features such as higher activity, better selectivity, increased stability, and resistance to inhibitors and tolerance to extreme conditions [22, 23]. Rational design is a very information-intensive process since it utilizes the harmony of different knowledge from amino acid sequence, 3D structure, structure-function relationship, and the mechanisms of catalysis or inhibition [22-24]. Extensive X-ray and NMR studies have been previously carried out to elucidate the structure-function relationship of some cellulases [25, 26]. Exquisite and fine-tuned analytical techniques such as X-ray scattering, nuclear magnetic resonance (NMR), molecular dynamics simulations, hydrogen deuterium exchange mass spectrometry (HDX-MS), recently developed X-ray spectroscopy, Kinetic Terahertz absorption (kinetic THz), single-molecule fluorescence resonance energy transfer

(FRET), quantum mechanical/molecular mechanical methods (QM/MM) empower the hand of rational design for better engineering efforts [12, 27].

1.4. Engineering Strategies to Improve Xylanase Production by *Trichoderma reesei*

As an industrially valuable and model strain for the large-scale production of cellulosic and hemicellulosic enzymes, *T. reesei* currently contributes to the biomass degradation by expressing four different xylanases [28]. Many approaches and efforts intended to increase the enzymatic production of xylanases from *T. reesei* by optimizing the culturing conditions, and incorporating a variety of different biomass feedstocks have provided remarkable improvements over time, in particular for the last three decades [29-33]. In order to further improve the degradation of industrially valuable feedstock by *T. reesei*, and increase and refine the amount of end-products obtained after hydrolysis, the expression of xylanase genes under high expression promoters of *T. reesei* and the heterologous expression of thermophilic enzymes in different *T. reesei* strains were also examined through various strategies [34-39].

As another approach beyond the optimizing culturing conditions, molecular genome engineering, and heterologous expression endeavors, the sequence and structure-based engineering of biocatalysts from a variety of different organisms for both improved catalysis and enhanced hydrolytic activities have also been performed over time.

Rational design of cell wall degrading enzymes based on structure information has produced fungal and bacterial enzymes with higher catalytic activity, altered stability and substrate specificity [40-44]. Identification of the 3D structures has led to the prediction of active sites, coordination spheres, and mechanisms for catalysis and substrate binding [26, 45-49]. In the early 1990s, the identification of the 3D structures of both XYN I and XYN II *T. reesei* endoxylanases by X-ray crystallography improved the knowledge of their biochemical mechanisms (Figure 1) [50]. However, considering the tremendous amount of input to date, the success in cell wall degrading enzyme improvement through rational design has been rather limited [24].

Rational design of cell wall degrading enzymes based on the structure-driven knowledge has led to the production of fungal and bacterial enzymatic recombinants with higher catalytic activity, altered stability, and substrate specificity [51-55]. All these efforts greatly manifest the importance of structure-dynamics and structure-function relationships of catalytic machineries.

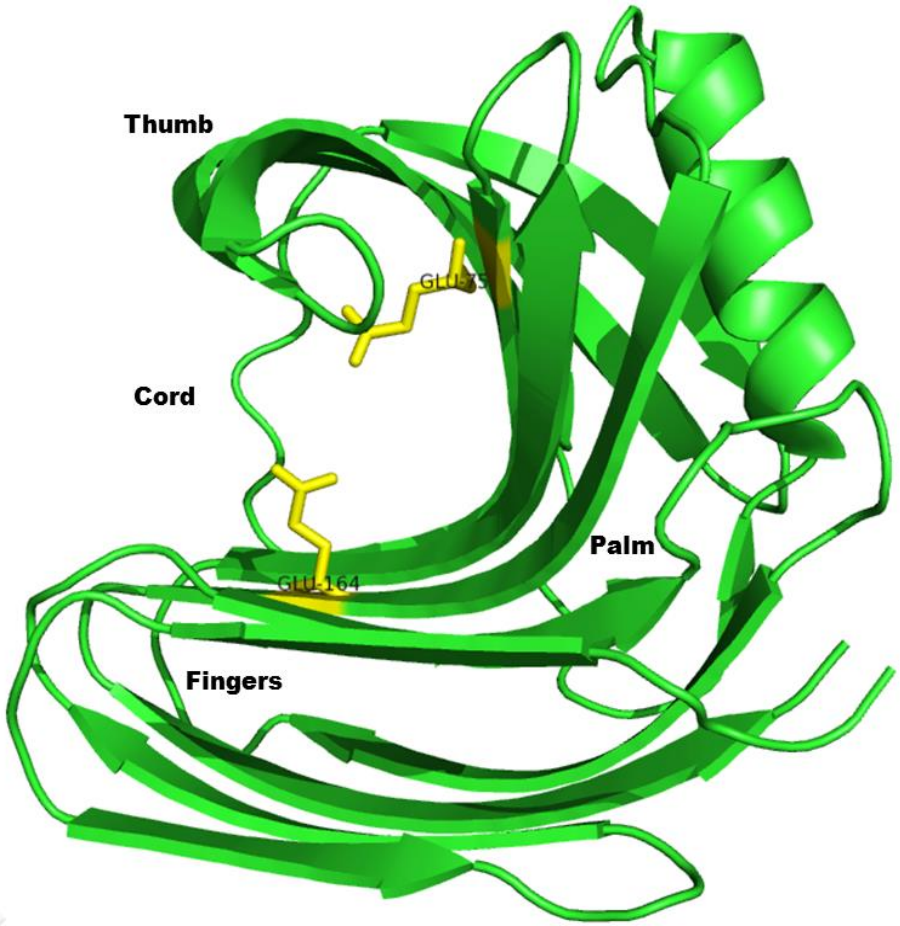

Figure 1. The three-dimensional structure of right hand-shaped XYN I enzyme with its catalytic amino acid residues marked as yellow.

1.5. The Importance and Phylogeny of *Trichoderma reesei*

The filamentous Ascomycota fungus *Trichoderma reesei* (teleomorph *Hypocrea jecorina*) produces a great number of cellulolytic and hemicellulolytic enzymes essential for xylan and lignocellulosic biomass degradation which make it a highly preferable species for industrial and biotechnology-based applications. The genus

Hypocrea/*Trichoderma* containing the different members of soil-borne or wood-decaying fungi are essential to humankind as producers of industrial enzymes and biocontrol agents against plant pathogens, but also as opportunistic pathogens of immunocompromised hosts. In terms of heterologous expression compatibility, efficient secretory mechanisms, and large scale production of eukaryotic proteins, *T. reesei* fungus stands as an excellent and ideal model system. It is of great interest especially for the degradation of a variety of different lignocellulosic biomasses thanks to its ability to naturally produce and secrete a range of enzymes important for biofuel production and plant cell wall hydrolysis.

Throughout the last two decades, there have been numerous efforts to elucidate the geographic distribution and phylogeny of *Trichoderma* members. The industrially most preferred species and well-known cellulase producer *T. reesei* is phylogenetically positioned within *Trichoderma* – section *Longibrachiatum* which is the most diverse group of *Trichoderma*. Besides *T. reesei,* the monophyletic group section *Longibrachiatum* contains some other intensively studied species, such as the facultative opportunistic human pathogens *T. longibrachiatum, T. citrinoviride,* and *H. orientalis* [56]. The most studied mycoparasitic biocontrol agents of the genus *Trichoderma* are most likely *T. atroviride* (teleomorph *Hypocrea atroviridis*) and *T. virens* (teleomorph *Hypocrea virens)* both of which are positioned within the paraphyletic clade section *Pachybasium.* This situation undoubtedly points out the phylogenetic distinction of hydrolytic enzyme producers *T. reesei* and *T. longibrachiatum* from these two biocontrol agents [57].

1.6. Importance of Biocatalyst Improvement Enterprises for Biofuel Applications

Increasing environmental concerns, mounting energy needs with the ascending population of earth, and diminutions on the potential of fossil fuels have shifted the focus of energy production towards sustainable and renewable energy sources such as lignocellulosics most abundantly available feedstock on earth. Although lignocellulosic biomass feedstock stands as one of the truly sustainable promising and renewable energy resources, the major challenge in producing biofuels from lignocellulosics is to improve the biological and chemical conversion efficiency [58-61]. Improving the effectiveness of biomass conversion is indispensable for the viability of advanced biofuels as alternative to crude oil [58-64]. Throughout the past two decades, extensive efforts were dedicated to improve the biomass conversion efficiency by focusing on the conditions of pretreatment, saccharification, and fermentation [58, 60, 62, 65]. The biomass conversion improvement endeavors depend heavily on the availability of biocatalysts with suitable features. Figure 2 simply exemplifies the potential strategies to improve different biorefinery processes in the presence of efficient biocatalysts.

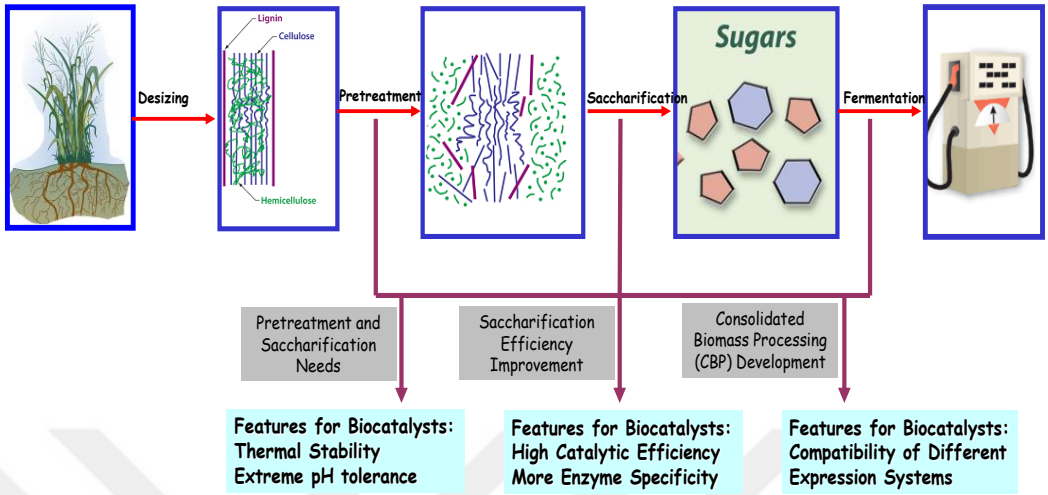

Figure 2. The needs for biocatalyst improvement to optimize biorefinery processes.

Biocatalysts tolerant to the severe environmental conditions such as high temperature and extreme pH may allow the integration of a better pretreatment with the other steps of lignocellulosic conversion.

The engineering efforts of biocatalysts are of great interest and indispensable need for industrial applications in order to simplify biomass processing, reduce the current processing costs, and improve efficiency of hydrolysis. Basically, exogenous enzymes are expressed in *Escherichia coli* (*E. coli*) or other biofuel fermenting organisms such as yeast to reduce the enzyme load and processing complexity [60, 66-68]. However, the expression of some commonly used fungal biocatalysts with sufficient activity in fermentation organisms like yeast is highly challenging and mostly results in insufficient biocatalyst activity [66, 68-71]. Enzymes secreted by microorganisms for cell wall digestion can be generally classified into two categories; free cell wall degrading enzymes and the multi-enzyme cellulosome systems. The free cell wall

degrading enzymes secreted by the filamentous fungus *T. reesei* and its close relatives are still the most common commercial enzymes used in biomass conversion due to their high efficiency and robustness [72]. Xylan is the major hemicellulose component of the plant cell wall and the second most abundant natural polysaccharide, and its degradation requires the concerted and synergistic function of several enzymes including endo-β-1,4-xylanases (EC 3.2.1.8) [73]. Due to the broad applications in biopulping and biobleaching in the paper industry, xylanase has been one of the major research focuses for bioconversion [74]. Xylanases with high substrate binding specificity, enhanced enzymatic activity, and increased thermostability are needed for various biorefinery applications. Tremendous efforts have been devoted to improve xylanase and cellulase enzyme performances by manipulating the protein amino sequences in the past [75]. It is experimentally infeasible to test all possible mutants of a protein, and it is time consuming since the majority of the manipulated sequences do not fold properly into functional proteins [76]. Suitable techniques are thus needed to guide the enzyme improvement with structure-function relationship for better enzyme rational design and engineering.

1.7. Xylanase Production by *T. reesei*

Besides the production of a variety of different cellulases, four major xylanases in *T. reesei* have also been identified to date [28]. Although they are grouped under various glycosyl hydrolase (GH) families, three of them were reported to have specific

endo-β-1,4-xylanase activity. Both xylanase I (XYN I) and xylanase II (XYN II), the two main xylanases with molecular masses of 19 and 20 kDa and isoelectric points of 5.5 and 9.0, respectively, belong to the GH family 11 hydrolases (Table 1). A GH family 10 member endo xylanase III (XYN III) from *T. reesei* has also been identified. The XYN III production in *T. reesei* can be induced with cellulosic substrate and L-sorbose rather than the induction by xylan [77]. Another unique xylanase, a GH family 30 member, was also identified from *T. reesei* as xylanase IV (XYN IV) with a molecular mass about 43 kDa (Table 1), displaying hydrolytic activity on hardwood glucuronoxylan, wheat arabinoxylan, and rhodymenan [28].

The typical 3D structures of xylanases contain the two β-sheet and one α- helix units which together form a β-sandwich structure [78]. XYN I and XYN II in general exhibit a right hand-shape structure where thumb, cord, palm, and finger segments are specified suggested by Torronen et al [79]. The two catalytic residues (Glu75 and Glu164) are positioned within the binding cleft (palm region) of the XYN I (Figure 1).

Table 1. Properties of xylanases produced by *T. reesei.*

Enzymes	amino acid sequence	pH optimum	pI	PDB codes	molecular weight (kDa)
XYN I	178	3.5-4.5	5.5	1XYN	19
XYN II	190	4.5-5.5	9.0	1ENX	20
XYN III	347	6.0-6.5	9.1	ND	32
XYN IV	465	3.5-4.0	7.0	ND	43

The proposed research therefore focuses on endo-β-1,4-xylanase 1 (XYN I) from *T. longibrachiatum* as a model cell wall degrading enzyme due to the extensive previous studies and its commercial value [72, 80, 81]. XYN I is an important enzyme with broad applications in particular for the biofuel industry [82]. As a GH family 11 enzyme, XYN I xylanase serves as a good model for other cellulolytic and hemicellulolytic enzymes. The XYN I of *T. longibrachiatum* displays 100% sequence identity to that of *Trichoderma reesei*. The two closely related filamentous fungi have the ability to abundantly secrete native hydrolytic enzymes. Considering all the identical characteristics of XYN I from both organisms, the 3D structure of *T. reesei* XYN I can be used to overlay the HDX results of XYN I from *T. longibrachiatum* in order to visualize and unravel its dynamic characteristics.

By considering the needs for better biocatalysts in the bioprocessing applications, it is hypothesized that the targeted modification of enzyme dynamics can be used to improve heterologous expression compatibility, thermostability, and other characteristics of biocatalysts. The knowledge of the structure dynamics during enzymatic catalysis can be derived from molecular dynamics simulations, NMR, and mass spectrometry-based methods.

1.8. Hydrogen Deuterium Exchange Mass Spectrometry as an Alternative Approach to Guide Enzyme Improvement

Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS) has been broadly applied to study protein dynamics and structure, particularly for protein binding with ligands, substrates, DNA and other molecules [83-88]. Such analysis has enabled the illustration of enzyme-substrate interaction mechanisms and the protein binding molecular determinants [89, 90]. The same approach can be used to study the conformational changes of xylanase when binding with substrates or inhibitors.

The fundamental concept of HDX-MS analysis is based on the mass increase of a protein when the protein protons (H) exchange with the solvent deuterium (D) [91]. The knowledge of the rates of the H/D exchange in the protein structure can be utilized to elucidate the structure dynamics of the protein as mass to charge ratio (m/z). Moreover, HDX-MS can also be used to study the global and regional protein conformational changes with different platforms [92, 93]. Coupled with protein digestion and chromatographic separation, HDX-MS is able to profile different regions of protein for H/D exchange based on the peptide H/D exchange rate and percentage. The information allows understanding which region of the protein is more stabilized or destabilized based on the H/D exchange information [88, 93-102]. If more H/D exchange is observed, it is postulated that the protein region is more exposed to the solution phase, and the region is thus considered to be more flexible. The structure dynamics information provided by novel HDX-MS platforms can be employed for enzyme engineering applications [97,

103]. Recent research efforts also support this idea by unveiling the relation of the structure dynamics to enzyme function [8, 97, 104]. Recently, uncovering the information of structure dynamics has become an important consideration for enzyme engineering efforts [8, 9, 104-108].

The structure dynamics-guided approach has been successfully used for several enzyme activity improvement applications [8]. With this study, it is therefore proposed as a novel strategy to unravel the intrinsic dynamic patterns of XYN I during enzymatic catalysis and provide useful knowledge to the downstream engineering applications. The related research is innovative, transformative, and translational. It is innovative because no previous work has been carried out to improve cell wall degrading enzymes based on HDX-MS based structure dynamics analyses. A combination of the structure dynamics data from the HDX-MS analyses and the static structure information from the X-ray crystallography provides novel insights for further enzyme improvement such as rational design and domain swapping. Furthermore, the related research introduces a new approach to enzyme engineering efforts to improve enzyme thermostability, specificity, and efficiency through rational design. Overall, this research study aims to address several challenges in biocatalyst improvement using structure dynamics-guided enzyme improvement as a novel strategy.

CHAPTER II

HDX-ANALYZER: A NOVEL PACKAGE FOR STATISTICAL ANALYSIS OF PROTEIN STRUCTURE DYNAMICS

2.1. Introduction

Protein intrinsic dynamics has been increasingly recognized as an important consideration for protein function [11]. Several recent studies revealed that protein dynamics play essential roles in catalysis and other functions [11, 81]. Among the different techniques, HDX mass spectrometry stands out as a relatively high throughput platform to probe the backbone dynamics of the proteins [88, 91, 96, 109]. HDX mass spectrometry has been broadly applied to study protein dynamics and structure, in particular for protein binding with ligands, substrates, DNA, and other molecules [84-88].

Such analysis has enabled the illustration of enzyme substrate interaction mechanism and the protein binding molecular determinants [89, 90]. The fundamental concept of HDX mass spectrometry analysis is based on the mass increase of a protein when the protein protons are exchanged with solvent deuterium [91]. The rate and

---

*Reprinted with permission from:" HDX-Analyzer: A Novel Package For Statistical Analysis of Protein Structure Dynamics" Sanmin Liu, Lantao Liu, Ugur Uzuner, Xin Zhou, Manxi Gu, Weibing Shi, Yixiang Zhang, Susie Y Dai, Joshua S. Yuan *BMC Bioinformatics*, Volume 12(Suppl 1):S43. Copyright (2011) BioMed Central

percentage of the H/D exchange can be measured by mass to charge ratio (m/z) of the protein. The HDX mass spectrometry can be used to study the global and regional protein conformational changes with different platforms [92, 93]. Coupled with protein digestion and chromatography separation, the HDX mass spectrometry is able to profile different regions of protein for H/D exchange based on the peptide H/D exchange rate or the m/z of the peptides. In a differential HDX experiment, usually two protein forms are compared. The *apo* protein and the ligand bound protein are subjected to HDX experiment in a parallel mode. The information allows understanding which region of the protein is more stabilized or destabilized upon ligand binding in the solvent exchange reaction [88, 93-102, 109]. If more H/D exchange is observed for a particular region, the protein region is more dynamic in the solvent exchange reaction, and the region is thus considered to be more flexible or less stable in HDX. In a typical differential HDX experiments, the protein of interest was subject to different exchange times in its *apo* form and protein ligand complex. The data processing for HDX mass spectrometry thus requires comparison to a large set of the m/z values or percentages of deuterium incorporation for the same peptides derived from *apo* protein and ligand bound protein at different exchange times with some technical replicates.

Various software platforms have been developed to analyze the HDX data. Among them include HX-Express, Deuterator, HD Desktop, DEX, Hydra, TOF2H and others. Most of the HDX data analysis software packages developed to date focus on calculating the m/z value from the MS raw data for the deuterated peptide, then evaluate the m/z value increase according to time. One of the aims for those software packages is

targeting processing a large set of HDX data, which includes calculating the centroid m/z value for varied HDX exchange times in a batch mode. For example, HX-Express is a semi-automated software package which exports deuterium uptake curve and peak width plots based on Microsoft excel application [110]. Compared with HX-Express, Deuterator is more automated and can deconvolute overlapping mass peaks. At the same time, Deuterator implements a server for processing multiple HDX data sets comparing the protein *apo* and ligand binding mode based on the web application [111]. Furthermore, HD Desktop is built on top of deuterator, which integrates more tools for data extraction with visualization components [112]. DEX is a Fourier deconvolution method that has been developed for high-resolution mass spectrometry data [113]. Hydra executes through a user-defined workflow, by which deuterium incorporation values are extracted and can be visualized in tabular and graphical formats. Hydra also automates the extraction and visualization of deuterium distribution values for large data sets [114]. TOF2H focuses on interpreting MALDI based HDX data and also builds up a pipeline for automated data processing [115]. Despite significant progress in the field, most software uses absolute differences between HD exchange rates as an evaluation of the differential structure dynamics changes. Some software did not integrate the differential HD exchange evaluation at all. It is noted that CalcDeut [116] evaluates the statistical distribution of deuterium incorporated into protease digested peptide fragments to compensate data truncation due to instrument signal to noise ratio. Nevertheless, most of the software programs have not integrated strong statistical evaluation of HD exchange rates for peptides. The statistical analysis of the differential H/D exchange rates of a

specific peptide is crucial in terms of the eventual decision whether the significant structural dynamic changes occur on that region or not. In many studies, a large difference in exchange rate may not reflect the differential structure dynamics changes if the standard deviation for the exchange rate is high. For this reason, new platforms for the statistical evaluation of HDX-MS data have also been integrated into HDXanalyzer.

The integration of statistical analysis with data processing is challenging. In terms of statistical analysis, several software environment including SAS, SPSS, and R can be employed. Among these packages, R is the open source choice with a strong visualization component and has significant advantages for developing the open source software. Despite the various advantages of R, the software environment does not have strong user-interface supports and thus requires the users to have basic knowledge of the statistical language R. In order to develop user-friendly statistical software for HDX mass spectrometry analysis, the latest RPY2 package to connect the statistical module of R with a data processing module implemented by Python has been employed and the user-interface by wxpython has been implemented. Many programming languages including C, java, Perl, Python can be used for data processing and UI development, and each programming language has its pros and cons. Among these programming languages, Python stands out as a unique choice for two reasons. First, the RPY package allowed the seamless and effective implementation between the data processing module and statistical module in R. Second, Python is a script language suitable for handling strings and biological data, and various BioPython packages have been developed for the

analysis of biological data. For these reasons, the HDX-MS analysis software HDXanalyzer has been developed using Python, R, and RPY2 packages.

A novel software package HDXanalyzer for the statistical analysis of HDX-MS data to evaluate the protein structure dynamics motions has been developed with this study. The software package includes three major components, the data intake and processing module, the user interface, and the statistical analysis module. The data processing module is developed in Python to process excel input files containing the m/z value of the peptides from different experiments. The pre-formatted m/z value of the peptides is processed to derive the centroid of the peptide peaks or the percentage of deuterium incorporation for the statistical analysis. The data is then processed by the Figure Generator to create graphs to visualize the differential HD exchange rates in *apo* and ligand bound protein. Further statistical analysis is carried out by R, where two statistical methods are used. The Paired Student's t-test is used to compare either the centroid values of the m/z value or the deuterium incorporation rate to derive point estimation, confidence intervals, and *p* value to indicate if significant differences in structure dynamics exist or not. In addition, the multiple regression (or ANCOVA) model is also involved to carry out the similar analysis through linear combination of the intercepts. HDXanalyzer thus provides novel solutions toward ultimate quantification and statistical evaluation of the structure dynamics changes when studied by HDX mass spectrometry. The software package addresses the imminent need of statistical evaluation for the HDX mass spectrometry analysis and can be expanded to other applications for HD exchange studies by other techniques.

2.2. Implementation

2.2.1. Data Processing as Implemented by Python

As shown in Figure 3, HDXanalyzer includes several modules, data processor, statistical analysis module, and user interface. HDX mass spectrometry often involves two types of data, the MS/MS analysis for peptide sequence identification and the MS analysis for m/z value of the peptide peaks in different protein forma and status (i.e. *apo* form, ligand bound form, proteins that has been subjected to different HDX exchange time points).

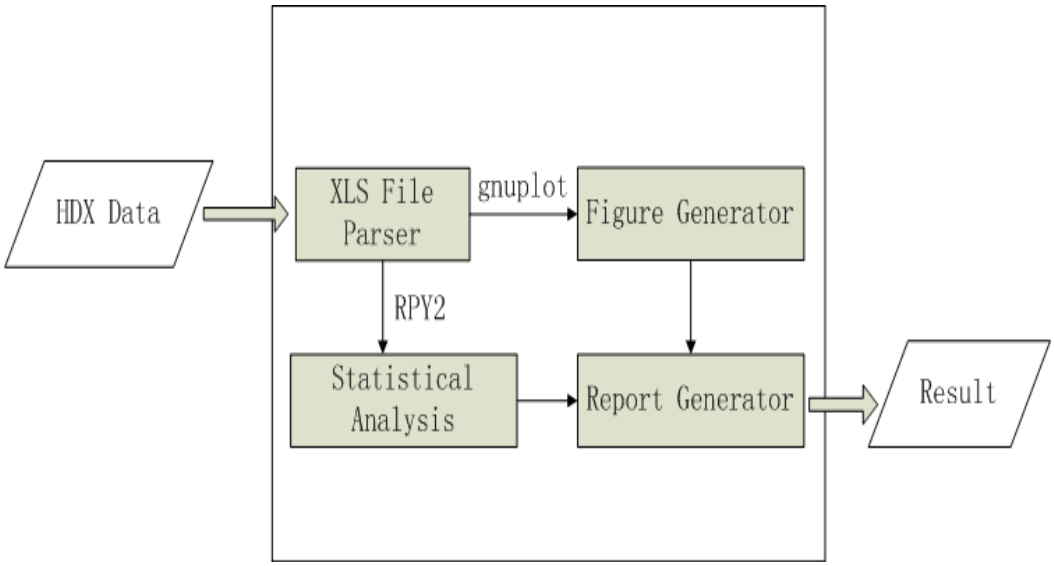

Figure 3. The implementation flow of HDXanalyzer.

The MS/MS analysis often allows us to correlate the peptide ID with a sequence, which is beyond the scope of HDXanalyzer. The software is mainly dealing with the MS analysis data. The raw MS data was exported for intensity and m/z values in the excel format for each peptide that has been identified in the HDX experiment. The HDXanalyzer then takes the pre-formatted Excel file of peptide ID and m/z values of the peptides at different HDX exchange time points as the input. A python package for reading and formatting information from excel files named as xlrd was used to develop a parser to extract and process Excel spreedsheets. The parser allows us to extract all m/z values and intensity of peptide at different retention time. The peptide ID and the treatment (*apo* or ligand bound) are also extracted.

The data processing can derive two types of variables for the statistical analysis, centroid of the peptide in form of m/z value or deuterium incorporation rate. Centroid values of each peptide are derived based on the m/z value of the peaks generated by MS analysis. For deuterium incorporation rate, the weighted average m/z values of each peptide ion isotopic cluster are calculated. Basically, the deuterium incorporation rate is calculated based on the centroid of the peptide m/z value and is in form of percentage. The deuteration level of each peptide is calculated based on the equation 1, and corrections for back-exchange are made based on 70% deuterium recovery and accounting for 80% deuterium content in the ion-exchange buffer. These corrections can be defined by users in the data processing procedure. The deuteration levels are calculated based on the equation given below;

$$Deuteration\ level(\%) = \left(\frac{m/z\ (P)\ -\ m/z\ (N)}{m/z\ (F)\ -\ m/z\ (N)} \right) \times 100 \qquad \textit{Equation 1}$$

where m/z (P), m/z (N), and m/z (F) are the centroid value of partially deuterated peptide, nondeuterated peptide, and fully deuterated peptide, respectively [111].

The resulted data are then processed into a table format and loaded to the *Figure Generator* to create visualization of the dynamic status of peptide at different time points. Specifically, the figures are graphs displaying the m/z value or deuterium incorporation rate of a peptide at different exchange time. Gnuplot was employed to implement the Figure Generator. Gnuplot is a cross-platform graphing utility compatible with Linux, Windows, and many other systems. The advantages of gnuplot lies in two aspects including the high quality 2D plots for scattered data and automate gnuplot using shell scripts. Besides the graphic display of the HDX data, statistical analysis is carried out to generate the point estimation of differential m/z value or incorporation rate, the confidence intervals and the *p* value.

2.2.2. Statistical Models and Implementation

Statistical analysis was employed to evaluate if a peptide or a region of the protein has significant changes in structure dynamics as revealed by HDX mass spectrometry or not. Such changes are reflected in the differences of either m/z value

during the HDX experiments or the deuterium incorporation rate. The m/z value or deuterium incorporation rate from different peptides can be compared with different statistical models to derive parameter estimation and *p* value. The parameter estimations allow us to evaluate the levels and variations of the differences in structure dynamics of a protein region, and the *p* value allows us to determine if the differences are significant or not.

Two types of statistical models were used. First, a Paired Student's t-test (Pairwise t-test) was used to compare the m/z value or percentage of deuterium incorporation for peptides from *apo* or ligand binding proteins. Paired Student's T-test is used instead of the regular T-test because of the time effects in the HDX experiments. The m/z values or the percentages of deuterium incorporation for a peptide increase as the hydrogen deuterium exchange time gets longer. As these two values are time dependent, they reach a plateau when the exchange time is long enough. For this reason, the Paired t-test is used to avoid the time point effects. Besides the pairwise t-test, the multiple regression model or ANCOVA model were also used to compare the m/z or incorporation rate differences between peptides from the two types of proteins (*apo* and ligand bound). The linear combination of the group effects allows comparing the differences between *apo* and ligand-bound proteins. For either model, the point estimation of mean differences, confidence intervals, and *p* values are rendered. The multiple regression model is given in equation 2;

$$Y = \beta_T X_{Time} + \beta_G X_{Group} + \beta_{TG} X_{Time} \times X_{Group} \qquad \textit{Equation 2}$$

where Y is the dependent variable that can be either the m/z value or the deuterium incorporation rates of different peptides. Y is dependent on the effects of time points and different groups from either *apo* or ligand bound proteins. The combination of the two effects may also influence the dependent variable.

2.2.3. RPY for Integrating the Different Components

The integration of statistical analysis, data processing, and visualization is always challenging, and the recent developed RPY allows us to integrate the statistical capacity of R and the user interface and data processing capacity of Python. As an open-source language, R has the unique advantages over other statistical languages for software development. RPY enables the use of R for statistical analysis of HDX mass spectrometry data. RPY2 provides a low-level interface to R. The Python-based system thus can directly call R function through RPY and the software efficiency and effectiveness are greatly improved.

2.2.4. User Interface as Implemented by Wxpython

The user interface of HDXanalyzer was developed with Wxpython. Wxpython is a Python extension model, which works as a wrapper for the cross-platform GUI API wxWidgets for the Python programming language. A user friendly interface including a menu bar, a tool bar, and four windows have been developed. The four windows are data manager window, figure browser, enlarged figure, and statistical analysis windows. Data manager window shows the spreadsheets (or the peptides) for the data analysis. The figure browser window shows a list of the graphs comparing the m/z values or deuterium incorporation rates of all peptides listed in the data manager window. All of the graphs are clickable and can be viewed in the enlarged figure window. The statistical analysis window displays the statistical analysis results.

2.3. Results and Discussion

HDXanalyzer is implemented as a software package to enable the statistical analysis of HDX mass spectrometry data and to allow the evaluation of protein structure dynamics changes. In order to test the confidence of HDXanalyzer, a previously published dataset for the HDX-MS analysis from XYN I enzyme has been employed.

2.3.1. The Input Data Format and the Usage of the Package

The HDXanalyzer aims to integrate statistical analysis for comparing structure dynamics of protein upon ligand or substrate binding. As discussed in the implementation part, the software takes a pre-formatted excel files containing m/z values for multiple peptides of different treatment and time points. The data pre-formatting allows to the software to process a uniform input of HDX mass spec data from different instruments. The sample input file is derived from a xylanase structure dynamics study and the m/z values of the peak area for the peptides are included. Each input Excel file contains several sheets for the data from different peptides and treatments. The spreadsheet contains peptide ID, m/z value, charge state, time points for deuterium treatment, and the ligand name to distinguish different experimental sets like apo set and ligand set. The peptide ID can be corresponding to a certain peptide sequence. The upload function is available from the user interface, where input file can be read and processed to generate centroids and deuterium incorporation rates of the peptides as aforementioned. The data are therefore further analyzed for visualization and statistical analysis.

2.3.2. The Output of HDXanalyzer

The data output includes three parts as shown in Figure 4. The upper left panel is the data manager window and contains the peptide list. The upper middle panel is the

figure browser window and contains the graphs to compare the trends of HD exchange for all the peptides from the input file.

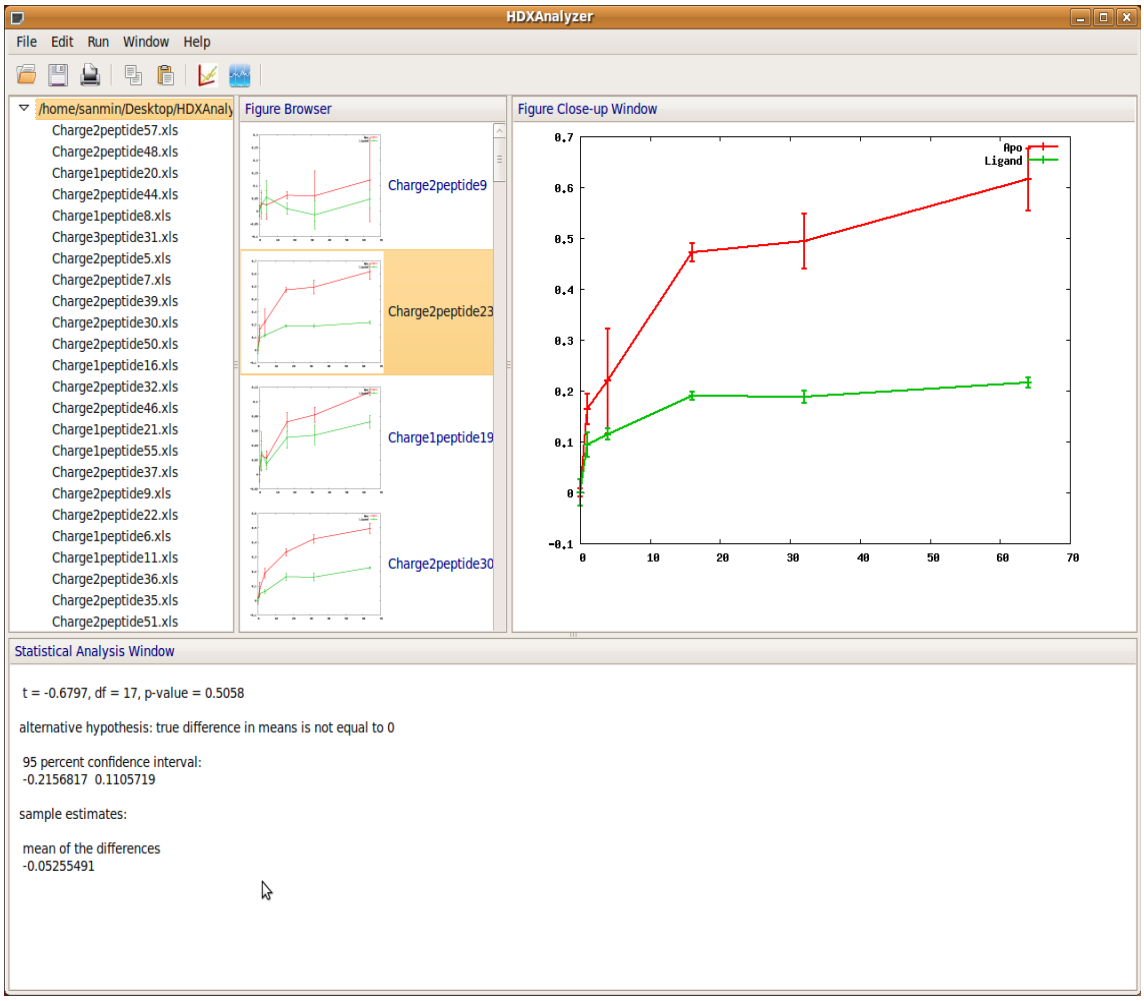

Figure 4. The overview of the HDXanalyzer user interface.

In each graph, the X axis is the time after the deuterium incubation, and the Y axis is either the m/z value of the centroid or deuterium incorporation rate. The user can choose

the output in either format. The same peptide from the *apo* protein and the ligand-bound protein are marked with different colors in the graph (Figure 5).

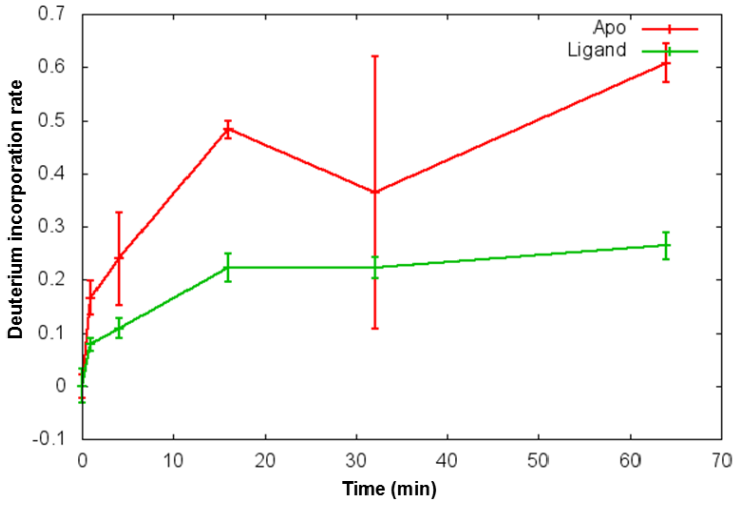

Figure 5. The deuteration-rate related figure of a representative peptide from apo and ligand-bound protein.

Each graph in the middle panel is clickable, and the enlarged graph for each peptide is presented in the expanded figure window in the upper right panel. When the graph is clicked, the statistical analysis can be carried out for the particular peptide and the output is presented in the statistical analysis window at the bottom (Figure 6).

```
Statistical Analysis Window

Start processing file: Charge1peptide1.xls

        Paired t-test

data:  c(925.456900473987, 925.682936642653, 926.037709898089, 926.26022276934,
925.86020382893,  and 925.849438731747, 926.128295019922, 925.874373466418, 925
926.846911272986,  925.532116978845, 925.980601890983, 926.314560378207, 926.51
926.538252530693)
t = -2.4096, df = 17, p-value = 0.02759
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.81149651 -0.05382831
sample estimates:
mean of the differences
       -0.4326624

lm model parameter contrast

   Contrast     S.E.     Lower     Upper     t df Pr(>|t|)
1 -0.4326624 0.1682729 -0.7799605 -0.0853643 -2.57 24   0.0168
```

Figure 6. The output for the statistical analysis window.

The statistical analysis output is typical R output with the point and confidence interval estimation as well as the *p* value. The parameter and *p* value estimation can be generated for either centroid of the peptide or the deuterium incorporation percentage. Statistical analysis can be carried out for all the peptides and the output can be presented together in the statistical analysis window.

2.3.3. The Interpretation and Comparison of Different Statistical Models

As aforementioned, two types of statistical analyses were implemented for HDXanalyzer. Both Paired Student's T-test and the linear combination of intercept for group (*apo* vs. ligand) effects in multiple regression models were used to derive parameter and *p* value estimation. A very important decision for data processing is the choice of time point. The early time points after deuterium exchange, especially for exchange less than 1 minute, may lead to very limited exchange even in *apo* protein. In such case, the statistical analysis of data from these time points cannot represent the real deuterium incorporation level. Therefore, HDXanalyzer offers users the option to choose the time points for the analysis.

The data interpretation is another important consideration. Several key values are presented in the statistical analysis readouts. The *p* value presents the estimation of whether a region has significant differential structure dynamics or not. A small *p* value on either deuterium incorporation rate or centroid of m/z indicated that the peptide has significant differential structure dynamics changes between the *apo* and ligand interaction. The point estimation of the mean differences and the confidence intervals estimates the levels of the changes. Both parameter estimation and the *p* value are important in interpreting the HDX-MS data to render reliable conclusions. In addition, the statistical analysis results also provide both the paired student's t-test

results and the outputs from multiple regression model. As shown in Table 2, the results from both types of analysis correlate with one another, highlighting the reliability of the results.

Table 2. The comparison of statistical analysis results from the Paired-students test and multiple regression.

Analysis Type	**Mean**	**LCL**	**UCL**	***p* values**
Multiple Regression	0.4522	0.3389	0.5654	0
Paired t-test	0.4452	0.3393	0.5649	0
Multiple Regression	0.2362	0.1157	0.3567	0.0004
Paired t-test	0.2362	0.1237	0.3487	0

To further validate the analysis results, the HDX data from another previously published dataset were also tested [117]. Figure 7 reveals the results of statistical analysis for the two peptides presented in the previous publication by Dai et al. [30]. The peptide in panel A did not show significant differential HDX, whilst the peptide in panel B shows significant differential HDX with $P < 0.001$ (Figure 7). The results from HDXanalyzer correlates well with the previous statistical analyses carried out by regular statistical analysis software PRISM by Dai et al. [3]. The comparison highlighted that HDXanalyzer is reliable.

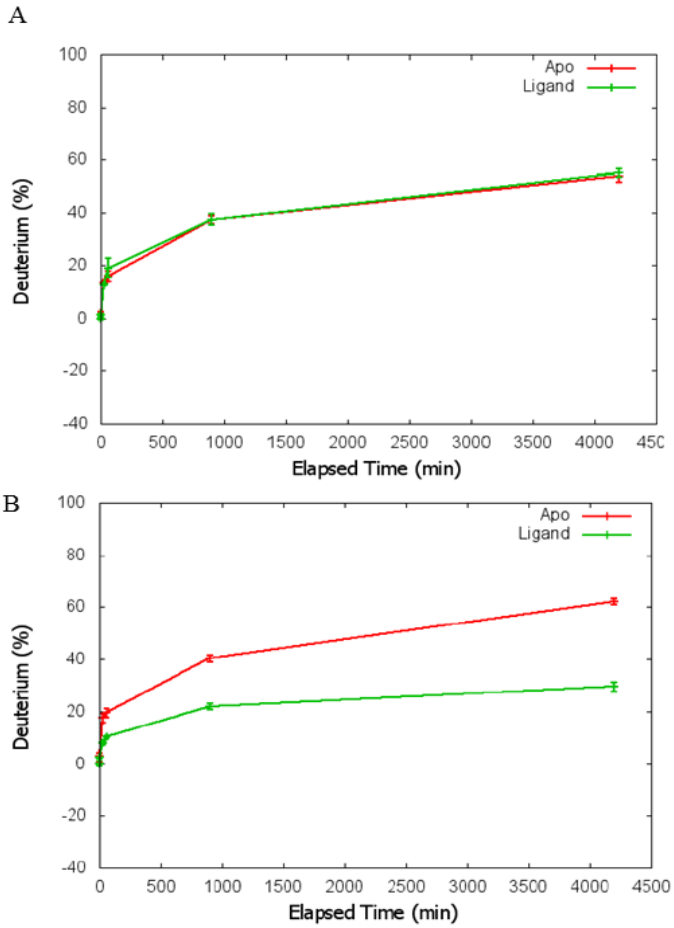

Figure 7. Verification of data analysis using previously published estrogen receptor data. A panel is a peptide showing no significant changes in the HDX; and B panel illustrates a peptide with significant changes in the HDX.

2.3.4. The Overlay of 3D Structure for Differential Structure Dynamics of Xylanase

The statistical analysis allows identification of those regions of the protein that exhibit significant structure dynamics changes upon ligand binding. As shown in Figure 8, the substrate binding of XYN I trigger substantial structural dynamic changes in

various regions of the protein. The overlay of the HDX-MS data onto the 3D structure facilitates the evaluation and interpretation of dynamic patterns of the protein.

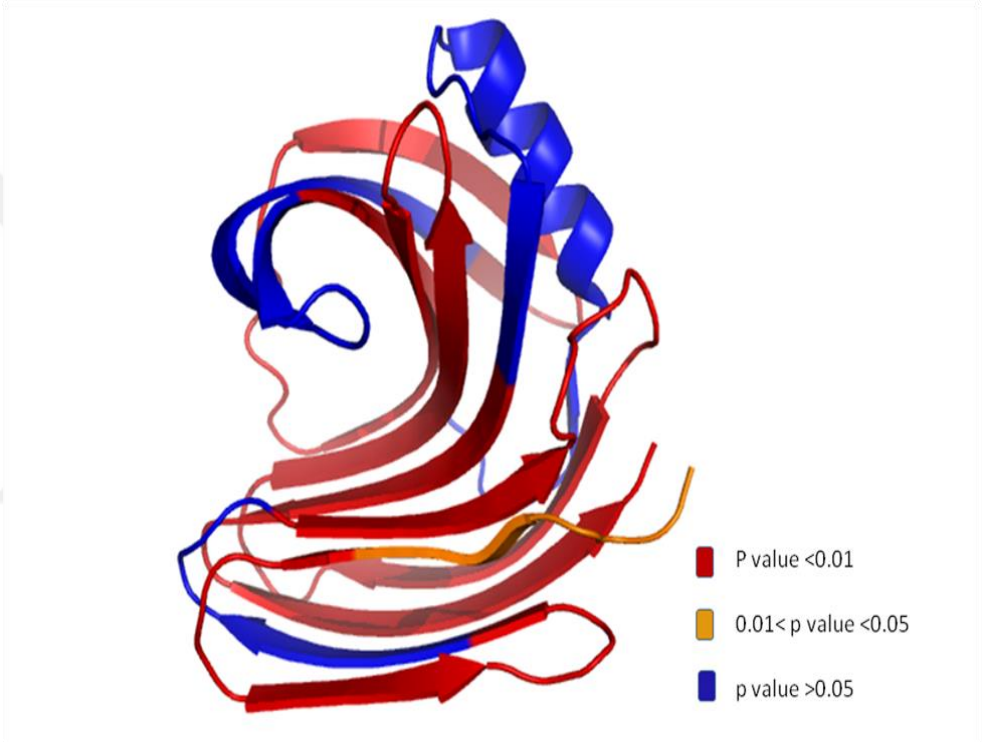

Figure 8. The overlay of *p* values from the HDX analysis onto the 3D structure of XYN I. The color legend indicates the level of confidence.

For example, the significant *p* values of the different peptides from the protein of interest, the regions with significant changes are highlighted on 3D structures and this information assists to understand the enzyme catalysis mechanisms to some extent. HDXanalyzer thus provides a powerful tool for statistical analysis of structure dynamics

data, which has not been achieved with the previous HDX mass spectrometry analysis packages.

HDXanalyzer as a statistical analysis software has enabled the accurate evaluation of the changes of protein structure dynamics. The software thereby enables the quick, timely-efficient and practical evaluation of the HDX-MS based differential structure dynamics analyses through the integration of graphical visualizations and statistical data.

Statistical analysis delivers crucial evaluation of whether a protein region is significantly protected or unprotected during the HDX mass spectrometry studies. In addition, HDXanalyzer stands as a powerful tool for the statistical evaluation of the fluctuations in protein structure dynamics based on HDX-MS data. The open-free developed software can be reached from the link given below;

http://people.tamu.edu/~syuan/hdxanalyzer

CHAPTER III

ENZYME STRUCTURE DYNAMICS OF XYLANASE I FROM *Trichoderma longibrachiatum*

3.1. Introduction

Xylan is the major hemicellulose component of the plant cell wall and the second most abundant natural polysaccharide. Most of xylan is a heteropolysaccharide consisting of β-1,4-linked D-xylose monomers in connection with side branches of arabinosyl, glucuronosyl, acetyl, uronyl, and mannosyl residues [74]. Complete degradation of xylan structures requires the concerted and synergistic function of several enzymes including endo- β-1, 4-xylanases (EC 3.2.1.8) [73]. Due to the broad applications in biopulping and biobleaching in paper industry, xylanase has been one of the major research focuses for bioconversion [74]. In particular, endoxylanases have been thoroughly studied as the major lignocellulosic biomass degradation enzymes. Xylanases with high substrate binding specificity, enhanced enzymatic activity, and increased thermostability are needed for various biorefinery applications. Tremendous efforts have been devoted to improve xylanase and cellulase enzyme performance by manipulating the protein amino sequence in the past [75]. However, sequence-based

---

*Reprinted with permission from: "Enzyme Structure Dynamics of Xylanase I from *Trichoderma longibrachiatum"* Ugur Uzuner, Weibing Shi, Lantao Liu, Sanmin Liu, Susie Y. Dai, Joshua S. Yuan. *BMC Bioinformatics*, volume 11(Suppl 6):S12. Copyright (2010) BioMed Central

protein modification has its limitations [118, 119]. It is experimentally infeasible to test all possible mutants of a protein, and it is time consuming since the majority of the manipulated sequences do not fold properly into functional proteins [76]. Suitable techniques thus are needed to guide enzyme improvement with structure-function relationship for better enzyme rational design and engineering.

Structure dynamics has become an important consideration for enzyme engineering [8, 9, 11, 104-108]. The structure dynamics during the enzyme catalysis can be derived from molecular dynamics simulations, NMR, and mass spectrometry-based methods. In particular, novel HDX mass spectrometry platforms provide the structure dynamics information for enzyme engineering [97, 103]. Recent research unveiled how structure dynamics is related to enzyme function [9, 97, 104]. The structure dynamics-guided approach has been successfully used for enzyme activity improvement [8, 11]. Hydrogen/deuterium exchange mass spectrometry (HDX-MS) represents one of the most widely used platforms for exploring protein conformational dynamics, folding, and binding [87, 120-124]. HDX mass spectrometry has been broadly applied to study protein dynamics and structure, in particularly for the protein binding with ligands, substrates, DNA and other molecules [84-87, 125]. Such analysis has enabled the illustration of the enzyme substrate interaction mechanism and the protein binding molecular determinants [88, 126].

The fundamental concept of HDX mass spectrometry analysis is based on the mass increase of a protein when the protein protons exchange with the solvent deuterium [90]. The rate and percentage of the H/D exchange can be measured by mass to charge

ratio (m/z) of the protein. The HDX mass spectrometry can be used to study the global and regional protein conformational changes with different platforms [91, 92]. Coupled with protein digestion and chromatography separation, the HDX mass spectrometry is able to profile different regions of protein for H/D exchange based on the peptide H/D exchange rate and percentage. The underlying cause of HDX structure dynamics may involve changes in hydrogen bonds and other forces [87]. For instance, if the protein binding with ligand leads to more H/D exchange in a region, the ligand binding is expected to induce conformational changes to destabilize the region. HDX mass spectrometry thus allows probing the protein structure dynamics changes during enzymatic reactions [9, 88, 93].

The HDX platform comes with the advantages of mass spectrometry analysis: fast, straightforward, and environmentally friendly [87]. The HDX mass spectrometry technologies thus provide user-friendly alternatives to study the structure and dynamics of xylanase in a way that is not possible with other technologies. The advantages of HDX mass spectrometry over X-Ray and NMR are higher throughput, less protein purity requirements, and the dynamics and stability information rendered [127, 128]. Compared with X-Ray or NMR techniques, HDX mass spectrometry is difficult to resolve structures at a single amino acid residue resolution. The resolution of the techniques relies on the protease digestion, which produces peptides of varied length in a protein-dependent manner. In addition, the conformational changes observed are the backbone changes, and the side chain information is limited. However, it is of great interest to resolve peptide regions that span several amino acid residues to localize the stabilized or

destabilized region during the catalysis or inhibition. The backbone changes contribute to the conformational changes involved in enzyme reactions [84, 88]. The combination of HDX mass spectrometry with X-ray data and computational modeling is a potent way to provide more detailed information regarding the structure, stability, and dynamics of enzyme/substrate or enzyme/inhibitor interactions.

In the present study, the structure dynamics changes of the *Trichoderma longibrachiatum* (*T. longibrachiatum*) xylanase enzyme upon binding with xylohexaose and xylan ligands was evaluated with HDX mass spectrometry analysis. The analyses revealed important regional dynamics of xylanase upon ligand binding. Combination of the structure dynamics data from the HDX analyses and the static structure information from the X-ray crystallography provided novel insights for further enzyme improvement.

3.2. Material and Methods

3.2.1. Protein and Reagents

The endo-1,4-β-Xylanase M2 (EC 3.2.1.8) from *T. longibrachiatum* was purchased from *Megazyme* (Megazyme International Ireland Ltd., Wicklow, Ireland) and used throughout the all HDX experimental processes without further purification. The protein solution was provided as a mixture of ammonium sulfate 45%, sodium azide 0.02%, water 45%, and Xylanase M2 (*T. longibrachiatum*) 2.5%. The substrate, xylohexaose (molecular weight, MW: 810.70 g^{-1}, Cat No: O-XHE) with >95% purity, was also provided from *Megazyme*. The second substrate of xylanase enzyme for this

study was 2% oats spelts xylan obtained from *TCI AMERICA* (Portland, OR, Cat No: X0011).

3.2.2. HDX Experiment

HDX experiments were similar to those previously described except without using the LEAP Technologies Twin HTS PAL liquid handling robot [39-40]. Briefly, the xylanase protein solution was used without further purification at the concentration of 13.6 mg/mL in solution. The xylohexaose was dissolved in a D_2O buffer (20mM Tris-HCL, 100mM KCL, and 1mM DTT in D_2O, pD 7.9) to reach a final concentration of 25mM. Xylan was dissolved in D_2O buffer to make up a 1% solution. Four µL of the xylanase solution was mixed with 16 µL of the ligand D_2O buffer for HDX experiments at room temperature for 0, 60, 240, 960, 1920, and 3840 seconds, respectively. After the incubation in D_2O at an aforementioned hydrogen deuterium exchange time, the exchange reaction was quenched with 30 µL ice-cold solution containing 2M urea and 1% Trifluoroacetic acid (TFA), injected into an injection valve with 50 µL sample loop, and then passed through a pepsin column (Applied Biosystems, Foster City, CA) by a solvent pump (0.1% TFA in water) with flow rate at 200 µL/min. The pepsin column was kept on ice. The digested xylanase peptides were then eluted through a micro peptide cartridge (Michrom Bioresources, Inc., Auburn, CA) and desalted. The digestion and desalting takes a total of 5 min. Peptides were then eluted across a 2.1mm x 5cm C18 column (Thermo Scientific, Waltham, MA) with a linear gradient of 2%-50%

Solution B over Solution A for 10 min (Solvent A, 0.1% formic acid in water; solvent B, 0.1 formic acid 80% acetonitrile, 20% water; flow rate 200 µL/min). Mass spectrometric analyses were carried out with the capillary temperature at 280 ºC using LC-LTQ mass spectrometer (Thermo Scientific, Waltham, MA). The *apo* xylanase HDX experiment was performed with the same protocol except that the D_2O solution contained no ligand.

3.2.3. Peptide Identification and HDX Data Processing

Product ion spectra were acquired in a data-dependent MS/MS mode. The precursor ion survey scan was performed and the five most abundant ions were selected for product ion analysis. MS/MS *.raw data was first converted into *.MS2 file and then searched against the database containing xylanase using SEQUEST algorithm (Bioworks, Thermo Finnigan, CA). All peptide ion assignments were inspected manually.

The weighted average m/z values of each peptide ion isotopic cluster were calculated with the in-house developed software named as HDXanalyzer (Chapter I). The deuteration level was calculated based on the equation given in Chapter I, and corrections for back-exchange were made based on 70% deuterium recovery and accounting for 80% deuterium content in the ion-exchange buffer.

3.2.4. Data Processing for HDX Mass Spectrometry

The xylanase enzyme of *T. longibrachiatum* shared 100% sequence identity to that of *T. reesei*. A total of 57 digested peptides were identified in the MSMS data acquisition with the sequence coverage of 91% (Figure 9) for XYN I chain A.

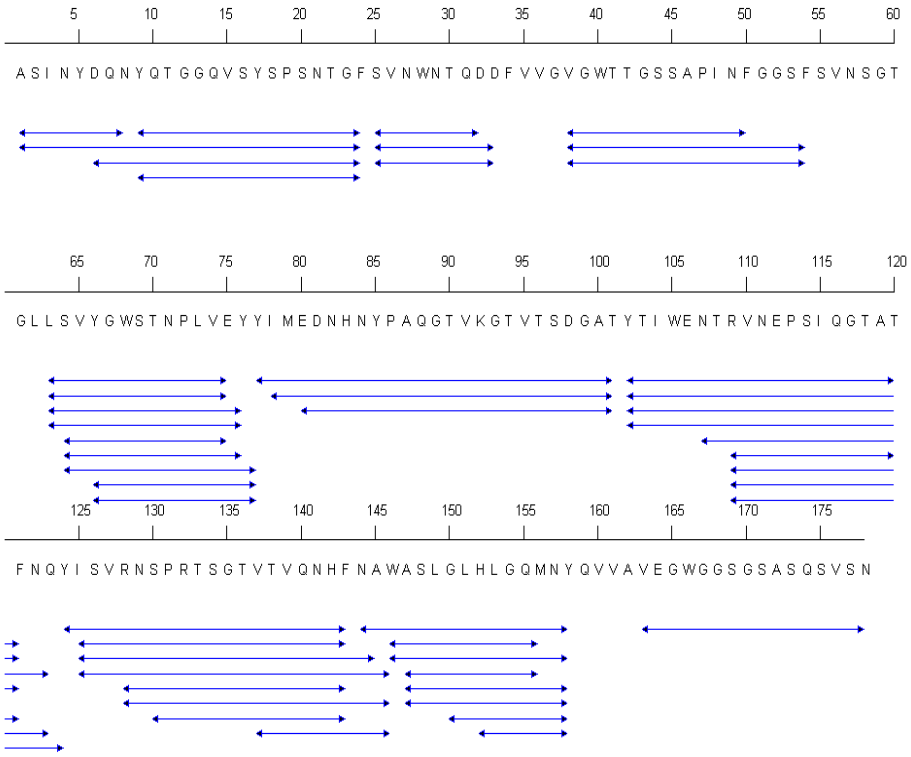

Figure 9. The peptides analyzed in the HDX experiment for XYN I (PDB code: 1XYN). The blue arrowed lines represent the sequence of the peptides.

In addition, the HDX analyses rendered 45 peptides with significant signal to noise ratio (S/N>3) and also gave 71% sequence coverage for XYN I enzyme. The covered region is given in Figure 10. Twelve peptides identified in the MSMS acquisition could not be measured accurately in the HDX experiments most likely due to the co-elution problems or weak signals after long exchange times (for precise description, see Table 3). The HDX profile of apo XYN I through the representative peptides is revealed in Figure 10. In order to illustrate the molecular mass increases of the peptides over time, a representative deuterium incorporation spectrum for the peptide "YTIWENTRVNEPSIQGTAT" (residues; 102-120) was displayed in Figure 11.

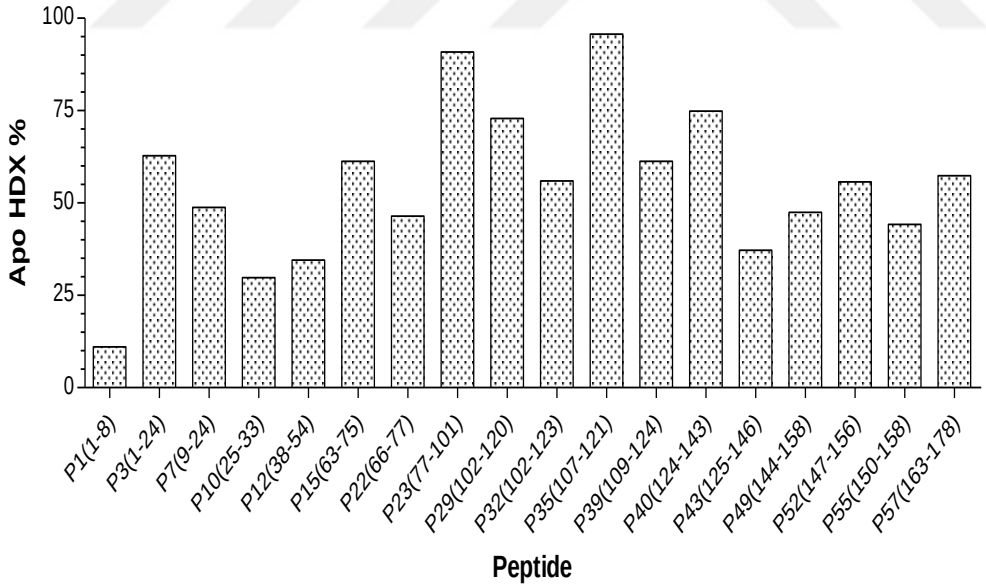

Figure 10. The HDX profile of apo XYN I through representative peptides. The selected peptides have a total of 91 % sequence coverage. The percentage is the averaged values of five HDX experiments with different exchange times (60, 240. 960, 1920, and 3840 seconds).

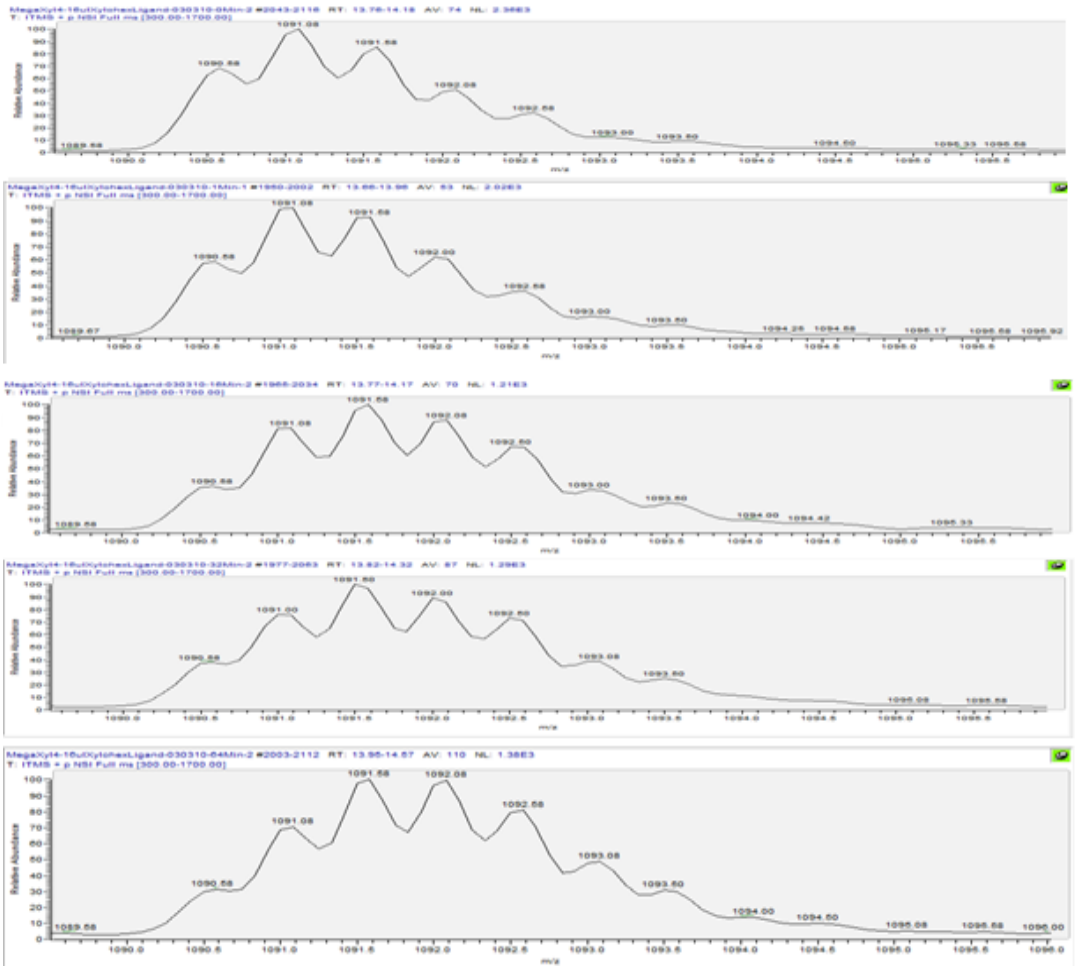

Figure 11.The mass spectra of XY N I peptide "YTIWENTRVNEPSIQGTAT" cover the residues; 102-129. The exchange times from top to bottom are 60, 240, 960, 1920, and 3840 seconds, respectively. The shift of the peak centroid toward to the left indicates the incorporation of deuterium into the backbone of the region.

3.2.5. Structure Dynamics of Apo XYN I Revealed by HDX-MS Analysis

The HDX analysis of apo xylanase showed that the enzyme dynamics is regional specific. Regions that include residues 77 to 101 and 107 to 121 showed the greatest hydrogen deuterium exchange in the *apo* enzyme, and the N-terminal of the enzyme is

less dynamic according to the overall hydrogen deuterium exchange. Overall, the enzyme is dynamic based on the HDX experiment considering that the average exchange percentage is 50 % for all five exchange times (i.e. 60, 240, 960, 1920, and 3840 seconds).

3.2.6. Differential HDX Analysis of XYN I

Differential HDX analysis of the *apo* protein versus the protein ligand complex revealed the structure dynamics change of the enzyme upon ligand binding. Figure 12 illustrates the two examples of the regions highly protected in the HDX experiment. The protected regions however displayed less deuterium incorporation for the same exchange times. The peptide “VGWTTGSSAPINF” in Figure 12A (residues; 38-50) and the peptide “YTIWENTRVNEPSIQGTAT” in Figure 12B (residues; 102-120) showed significant protection in the HDX measurements. Contrary to that, the peptide “SVYGWSTNPLVEY” (residues; 64-76), and the peptide “SVYGWSTNPLVEYY” (residues; 64-77) revealed similar exchange pattern in the *apo* protein and holo protein, which suggested those regions had little structure dynamics change upon xylohexaose binding (Figure 12C and Figure 12D). More importantly, the two peptides had only one amino acid differences and had an essentially similar HDX profile, which highlighted the reproducibility of the current data. The differences in the structure dynamics probed by hydrogen deuterium exchange potentially reflected the differential local structure protections. The common speculation of mechanism for HDX involves H bonding and

possible solvent accessibility, which lead to different hydrogen/deuterium exchange rates in different protein regions [36]. Nevertheless, the HDX analysis revealed the differential structure dynamics of various xylanase regions when binding with substrate xylohexaose.

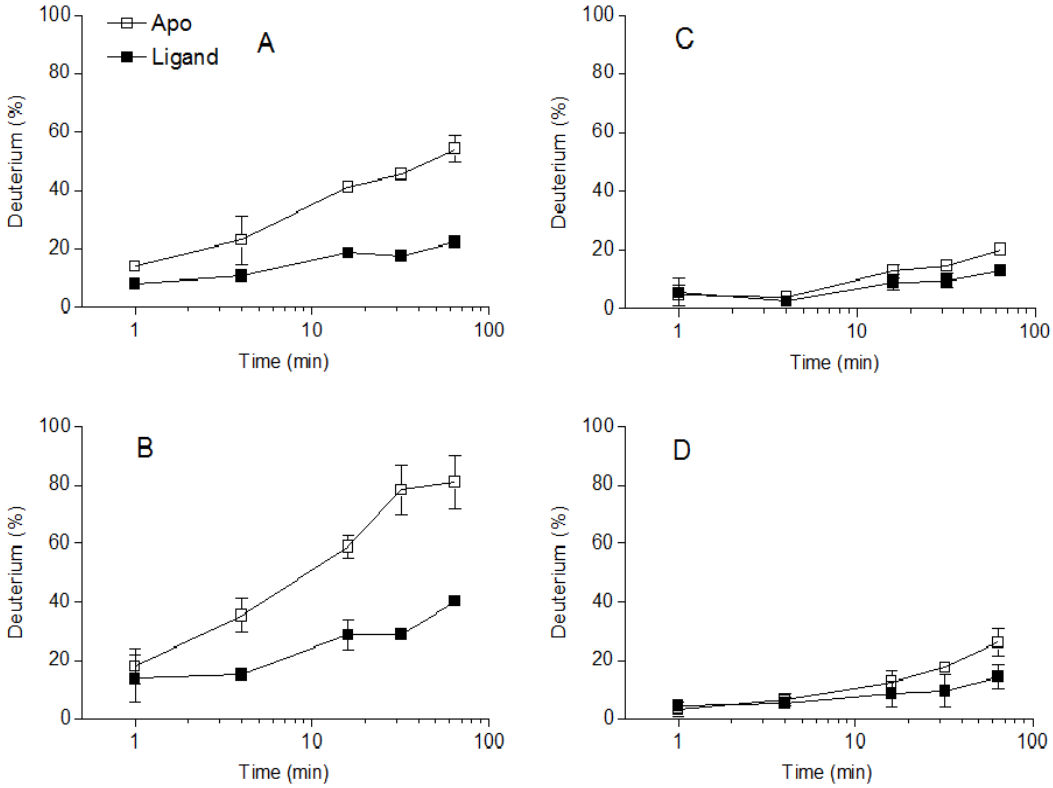

Figure 12. The deuterium incorporation curves of four XYN I peptides. The open squares illustrate the HDX data for apo XYN I and the closed squares display the HDX data for the XYN I-xylohexaose bound complex. A) The HDX data for the peptide "VGWTTGSSAPINF" containing residues; 38-50. B) The HDX data for the peptide "YTIWENTRVNEPSIQGTAT" containing residues; 102-120. C) The HDX data for the peptide "SVYGWSTNPLVEY" containing residues; 64-76, D) The HDX data for the peptide "SVYGWSTNPLVEYY" containing residues; 64-77.

3.2.7. Statistical Analysis for Differential HDX Data

Despite the visualization, the differential HDX profile needs to be determined statistically. We chose the last three time points for the paired student *t* test analysis to compare the *apo* and ligand binding mass spectra centroid value as shown in Table 3. For the *apo* enzyme, the maximum D_2O exchange percentage (HDX percentage) for each peptide ranged from 10% to 95.7%. In parallel, HDX percentage was also defined for the binding of XYN I with its substrate. The level of protection for a particular peptide can be determined by two methods. First, the differences for the HDX percentage between the *apo* and *holo* protein can be calculated. Second, the differences of the centroid of the mass spectra for the peptides can also be evaluated. The advantage for the percentage calculation is that it gives a numeric value that is normalized against the size and the charge state of the protein. The advantage of centroid value is that it is the direct measurement of the HDX mass spectra and thus can be more readily subject to various statistical analyses. It has been therefore combined the strength of the two types of measurement. For the visualization in, the deuterium rates as percentage differences were overlaid onto the3D structure of the protein. For the statistical analysis in Table 3 and Table 4, the mass spectra centroid values of the peptides to compare the *apo* and *holo* proteins were also evaluated by employing paired student's t-test.

The results revealed that the substrate xylohexaose played an important role in the stabilization of the xylanase chain A, because the HDX rate for most of the peptides were smaller than *apo*, and the centroid values were significantly lower than those of

apo protein for most peptides (P <0.01). Only a few peptides such as the peptide "SVNWNTQDD" and "NTRVNEPSIQGTATF" showed no significant differences in the peptide centroid values between the *apo* and *holo* protein (t =0.9553 and 0.9681, P =0.3627 and 0.3614).

In contrast, most of the peptides were not significantly stabilized by xylan binding Table 4. Upon the xylan binding, only a few peptides including "SVNWNTQDD" (t=4.76, P<0.01), "LSVYGWSTNPLVEY" (t=3.08, P=0.02), "YGWSTNPLVEYY" (t =4.52, P <0.01), "ISVRNSPRTSGTVTVQNHF" (t =3.23, P =0.01), and "NAWASLGLHLGQMNY" (t =2.81, P =0.02) were significantly stabilized through xylan binding Table 4. Even though some peptides of enzyme-xylan binding complex exhibited large HDX percentage differences, these peptides did not display significant variations in the centroid values. Overall, the results highlighted the importance of statistical analysis applications with respect to interpretation of structure dynamics of XYN I.

3.2.8. Correlation of HDX Data with X-Ray Structure

Correlation of the HDX data with the X-ray three dimensional structure data can also provide useful information to understand the enzyme function. The 3D structure of xylanase from *T. reesei* was used as the template to overlay the structure dynamics with the static structure since the enzyme shares 100% amino acid sequence identity with the *T. longibrachiatum* Xylanase. Excellent consistency between the X-ray structure and the

HDX analyses were also observed. The active sites of XYN I, Glu75 and Glu164 are located in B6 and B4 β strands [37]. The protein was previously described to have the analogy of the right hand. As shown in Figure 13A, the HDX profile revealed that the regions close to the reaction center had the highest protection when the enzyme bound with xylohexaose ligand, suggesting that the enzymatic activity also involves chemical interactions between substrate and several specific amino acid residues closely located to active sites. Correspondingly, the results highlighted the potential regions essential for the ligand binding process and substrate specificity (Figure 13A). In addition, the thumb region of the protein as suggested in the previous publication [37] was also highly stabilized during the ligand binding. The stabilization indicated that the thumb region of xylanase might also play an important role in binding and processing of the xylohexaose substrates. HDX analysis thus had the potential to reveal important regions of substrate binding and enzyme reaction beyond the reaction sites. Besides the thumb regions, the existence of some other stabilized regions outside of the substrate binding groove was also determined. For instance, the peptide containing the residues between from 77 to 101 located only one residue away from the active site Glu75 exhibited the greatest protection in the HDX experiment. The dynamic characteristics of XYNI-xylan complex however, was less dynamic compared to that of xylohexaose (Figure 13B). The protection was also observed with the peptide containing the residues from 6 to 24 which represents one of the fingers of the right hand structure of the protein.

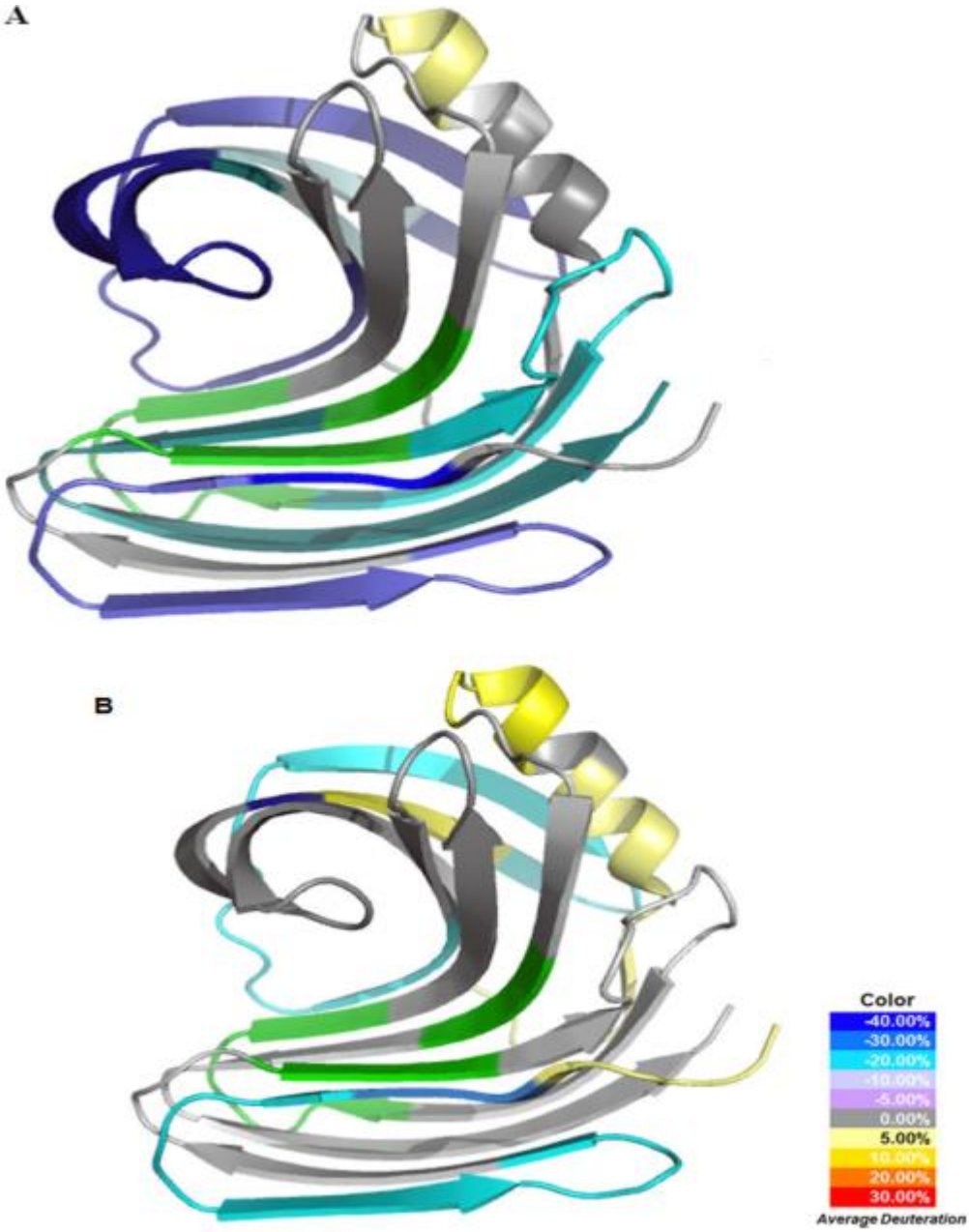

Figure 13. The HDX profiles overlaid onto the 3D structure of XYN I. A) HDX profile of XYN I through the binding to xylohexaose soluble substrate. B) HDX profile of XYN I through its interaction with xylan (XYN I/xylan complex). The color legend reveals the rates of deuterium incorporation difference specifically obtained by subtracting the deuteration percentage of apo xylanase from holo xylanase. The green colored sections represent the regions that are not detected after pepsin digestion or cannot be measured accurately in the HDX experiments due to co-elution problems.

3.2.9. Comparative Analysis of XYN I Structure Dynamics When Binding with Two Different Substrates

As aforementioned, the HDX profile of the two substrates, xylan and xylohexaose, was analyzed. The results showed quite altered structure dynamics induced by different substrate bindings as shown in Figure 13, Table 3, and Table 4 . Overall, the xylan binding induced the less enzyme structure dynamics changes on the conformational structure of XYN I compared to that of xylohexaose binding. Even though the xylan binding displayed some regional HDX deuterium changes as shown in Figure 13B, most peptides did not exhibit statistically significant changes in terms of HDX percentage. The results through the HDX profiles of the two substrates uncovered the existence of the differential binding ability of XYN I over various substrates. Such variation between the substrates may be explained with the solubility problem of xylan molecules. Xylan is not entirely soluble and the solution usually forms an emulsion during the experiment which might have led to the significant variation of the HDX analysis during the xylan binding experiments. Overall results indicated that xylohexaose binding might have rendered much more reliable data for mechanism studies.

3.2.10. HDX Profile and Enzyme Improvement for the Future

Engineering protein flexibility (plasticity) can be used to enhance substrate/ligand specificity of proteins by increasing the rigidity of flexible residues [38]. For this reason, the HDX analysis of XYN I was carried out to identify the critical sequences that could be stabilized to improve the local interactions of the groove region and neighboring residues of XYN I. The current data provided significant information regarding the global and local dynamics of xylanase enzymes through binding with different substrates. The obtained knowledge related to the structure dynamics of XYN I can be engaged to the future enzyme engineering applications. Those highly stabilized regions through ligand bindings are excellent candidates for similar downstream protein engineering treatments. In addition to the thumb region, the cord region of XYN I was also highly deuterated during the experiment, indicating that the cord region could also be another target region towards to enzyme specificity and stabilization-based engineering applications on XYN I.

Table 3. The HDX analysis of the sequence-specific peptides from XYN I and related statistical data.

			Apo	Xylohexaose		Centroid difference				
ID	**Peptide**	**Residue&Charge**	**D2O%**	**D2O%**	**Change rate%**	**Mean**	**SD**	***T*-test**	**df**	***p*-value**
1	ASINYDQN	1-8 (1)	11	9	-2	-0.492	1.03	1.432	8	< 0.01
2	ASINYDQNYQTGGQ	1-14 (2)								
3	ASINYDQNYQTGGQVSYSPSNTGF	1-24 (2)	62.8	27	-35.8	0.9661	0.34	8.528	8	< 0.01
4	SINYDQN	2-8 (1)								
5	DQNYQTGGQVSYSPSNTGF	6-24 (2)	61.6	21.7	-39.9	0.7114	0.182	11.72	8	< 0.01
6	YQTGGQVSYSPSNTGF	9-24 (1)	45.6	18	-27.6	1.1621	0.263	13.27	8	< 0.01
7	YQTGGQVSYSPSNTGF	9-24 (2)	48.78	20.2	-28.58	0.6794	0.182	11.18	8	< 0.01
8	SVNWNTQD	25-32 (1)	12.45	7.57	-4.88	0.6672	0.268	7.472	8	< 0.01
9	SVNWNTQDD	25-33 (2)	14.5	4.26	-10.24	0.1268	0.394	0.965	8	0.3627
10	SVNWNTQDD	25-33 (1)	29.8	12.3	-17.5	0.6465	0.121	16	8	< 0.01
11	VGWTTGSSAPINF	38-50 (1)	47	19.61	-27.39	1.0629	0.191	16.74	8	< 0.01
12	VGWTTGSSAPINFGGSF	38-54 (1)	34.52	15.82	-18.7	0.9381	0.354	7.943	8	< 0.01
13	VGWTTGSSAPINFGGSF	38-54 (2)	32.28	7.68	-24.6	0.4687	0.184	7.652	8	< 0.01
14	LSVYGWSTNPLVE	63-75 (1)	16.75	10.55	-6.2	0.1773	0.151	3.518	8	< 0.01
15	LSVYGWSTNPLVE	63-75 (2)	61.26	24.68	-36.58	0.6243	0.144	12.98	8	< 0.01
16	LSVYGWSTNPLVEY	63-76 (1)	17.21	7.85	-9.36	0.3248	0.171	5.687	8	< 0.01
17	LSVYGWSTNPLVEY	63-76 (2)	20.8	14.43	-6.37	-0.114	0.228	1.505	8	< 0.01
18	SVYGWSTNPLVE	64-75 (1)	9.97	5.63	-4.34	0.2211	0.143	4.63	8	< 0.01
19	SVYGWSTNPLVEY	64-76 (1)	15.88	10.52	-5.36	0.2603	0.095	8.242	8	< 0.01
20	SVYGWSTNPLVEYY	64-77 (1)	18.76	10.92	-7.84	0.3695	0.237	4.677	8	< 0.01
21	YGWSTNPLVEYY	66-77 (1)	14.4	8.36	-6.04	0.314	0.157	6.011	8	< 0.01
22	YGWSTNPLVEYY	66-77 (2)	46.42	26.5	-19.92	0.4475	0.135	9.933	8	< 0.01
23	YIMEDNHNYPAQGTVKGTVTSDGAT	77-101 (2)	90.88	35.44	-55.44	1.0861	0.235	13.87	8	< 0.01
24	IMEDNHNYPAQGTVKGTVTSDGAT	78-101 (2)	83.74	42.21	-41.53	0.8823	0.527	5.018	8	< 0.01
25	EDNHNYPAQGTVKGTVTSDGAT	80-101 (2)	82.38	33	-49.38	0.8533	0.326	7.858	8	< 0.01

Table 3. Continued

			Apo	Xylohexaose		Centroid difference				
ID	**Peptide**	**Residue&Charge**	**D2O%**	**D2O%**	**Change rate%**	**Mean**	**SD**	***T*-test**	**df**	***p*-value**
26	DNHNYPAQGTVKGTVTSDGAT	81-101 (2)								
27	YTIWENTRVNE	102- 112 (1)								
28	YTIWENTRVNEPSIQG	102-117 (2)								
29	YTIWENTRVNEPSIQGTAT	102-120 (2)	72.88	36.68	-36.2	0.8337	0.193	12.98	8	< 0.01
30	YTIWENTRVNEPSIQGTATF	102-121 (2)	74.39	32.65	-41.74	0.7598	0.199	11.57	8	< 0.01
31	YTIWENTRVNEPSIQGTATF	102-121 (3)	69	34.8	-34.2	0.4264	0.172	7.419	8	< 0.01
32	YTIWENTRVNEPSIQGTATFNQ	102-123 (2)	55.96	32.91	-23.05	0.506	0.585	2.592	8	0.0319
33	IWENTRVNEPSIQGTAT	104-120 (2)								
34	NTRVNEPSIQGTAT	107-120 (2)								
35	NTRVNEPSIQGTATF	109-120 (2)	95.7	33.6	-62.1	0.2885	0.894	0.968	8	0.3614
36	RVNEPSIQGTAT	109-120 (2)								
37	RVNEPSIQGTATF	109-121 (2)	44.54	21.4	-23.14	0.546	0.243	6.741	8	< 0.01
38	RVNEPSIQGTATFNQ	109-123 (2)	39.31	16.7	-22.61	0.273	0.112	7.344	8	< 0.01
39	RVNEPSIQGTATFNQY	109-124 (2)	61.25	32.23	-29.02	0.6612	0.102	19.47	8	< 0.01
40	YISVRNSPRTSGTVTVQNHF	124-143 (2)	74.86	38	-36.86	0.8307	0.323	7.727	8	< 0.01
41	ISVRNSPRTSGTVTVQNHF	125-143 (3)	29.37	15.76	-13.61	0.3136	0.105	9.007	8	< 0.01
42	ISVRNSPRTSGTVTVQNHFNA	125-145 (2)	36.33	12.97	-23.36	0.3182	0.169	5.662	8	< 0.01
43	ISVRNSPRTSGTVTVQNHFNAW	125-146 (2)	37.19	13.31	-23.88	0.3188	0.122	7.847	8	< 0.01
44	RNSPRTSGTVTVQNHF	128-143 (2)	23.13	14.5	-8.63	0.2763	0.169	4.896	8	< 0.01
45	RNSPRTSGTVTVQNHFNAW	128-146 (2)	24.23	7.74	-16.49	0.2565	0.124	6.205	8	< 0.01
46	SPRTSGTVTVQNHF	130-143 (2)	10	6.17	-3.83	0.4537	0.258	5.279	8	< 0.01
47	SPRTSGTVTVQNHF	130-143 (2)								
48	VTVQNHFNAW	137-146 (2)								
49	NAWASLGLHLGQMNY	144-158 (2)	47.44	29.65	-17.79	0.8253	0.182	13.57	8	< 0.01
50	WASLGLHLGQM	146-156 (2)	10.59	14.31	3.72	0.2383	0.104	6.847	8	< 0.01
51	WASLGLHLGQMNY	146-157 (2)	45.47	25	-20.47	0.5188	0.109	14.33	8	< 0.01
52	ASLGLHLGQM	147-156 (2)	55.68	32.27	-23.41	0.3143	0.171	5.509	8	< 0.01

Table 3. Continued

			Apo	Xylohexaose		Centroid difference				
ID	**Peptide**	**Residue&Charge**	**D2O%**	**D2O%**	**Change rate%**	**Mean**	**SD**	***T*-test**	**df**	***p*-value**
53	ASLGLHLGQMNY	147-158 (1)	32.86	2.45	-30.41	1.0308	0.406	7.624	8	< 0.01
54	ASLGLHLGQMNY	147-158 (2)								
55	GLHLGQMNY	150-158 (1)	44.18	28.38	-15.8	0.465	0.32	4.356	8	< 0.01
56	HLGQMNY	152-158 (1)	19.95	7.65	-12.3	0.4594	0.274	5.028	8	< 0.01
57	VEGWGGSGSASQSVSN	163-178 (2)	57.35	29.37	-27.98	0.6036	0.162	11.21	8	< 0.01

Table 4. The HDX rates of the specific peptides through XYN I-xylan interaction and related statistical data.

			Apo	Xylan		Centroid difference				
ID	**Peptide**	**Residue&Charge**	**D2O%**	**D2O%**	**Change rate%**	**Mean**	**SD**	***T*-test**	**df**	***p*-value**
1	ASINYDQN	1-8 (1)	11	17.47	6.47	-0.568	1.216	1.4012	8	0.1987
2	ASINYDQNYQTGGQ	1-14 (2)								
3	ASINYDQNYQTGGQVSYSPSNTGF	1-24 (2)	62.8	31.6	-31.2	0.2922	0.2969	2.9521	8	0.0184
4	SINYDQN	2-8 (1)								
5	DQNYQTGGQVSYSPSNTGF	6-24 (2)	61.6	26.45	-35.15	0.0919	0.2317	1.1893	8	0.2684
6	YQTGGQVSYSPSNTGF	9-24 (1)	45.6	24.4	-21.2	0.2032	0.1721	3.5408	8	0.0076
7	YQTGGQVSYSPSNTGF	9-24 (2)	48.78	21.3	-27.48	0.1322	0.1717	2.3084	8	0.0498
8	SVNWNTQD	25-32 (1)	12.45	9.29	-3.16	-0.0074	0.2488	0.0896	8	0.9308
9	SVNWNTQDD	25-33 (2)	14.5	9.06	-5.44	-0.4882	0.3086	4.7457	8	0.0015
10	SVNWNTQDD	25-33 (1)	29.8	20.42	-9.38	-0.0559	0.1813	0.0388	8	0.3824
11	VGWTTGSSAPINF	38-50 (1)	47	33.16	-13.84	0.0681	0.2478	0.8244	8	0.4336
12	VGWTTGSSAPINFGGSF	38-54 (1)	34.52	20.61	-13.91	0.0495	0.3242	0.4578	8	0.6592
13	VGWTTGSSAPINFGGSF	38-54 (2)	32.28	29.77	-2.51	-0.0746	0.1771	1.2635	8	0.242

Table 4. Continued

			Apo	Xylan		Centroid difference				
ID	**Peptide**	**Residue&Charge**	**D2O%**	**D2O%**	**Change rate%**	**Mean**	**SD**	**T-test**	**df**	**p-value**
16	LSVYGWSTNPLVEY	63-76 (1)	17.21	16.54	-0.67	-0.1646	0.1606	3.0765	8	0.0152
17	LSVYGWSTNPLVEY	63-76 (2)	20.8	21.65	0.85	-0.0678	0.1864	1.092	8	0.3066
18	SVYGWSTNPLVE	64-75 (1)	9.97	15.61	5.64	-0.1651	0.191	2.5937	8	0.0319
19	SVYGWSTNPLVEY	64-76 (1)	15.88	13.58	-2.3	-0.1467	0.2064	2.1333	8	0.0655
20	SVYGWSTNPLVEYY	64-77 (1)	18.76	16.48	-2.28	-0.1772	0.3409	1.5597	8	0.1574
21	YGWSTNPLVEYY	66-77 (1)	14.4	17.3	2.9	-0.1605	0.1946	2.4742	8	0.0385
22	YGWSTNPLVEYY	66-77 (2)	46.42	25.77	-20.65	0.1567	0.104	4.5203	8	0.0019
23	YIMEDNHNYPAQGTVKGTVTSDGAT	77-101 (2)	90.88	60.26	-30.62	0.0787	0.2673	0.883	8	0.403
24	IMEDNHNYPAQGTVKGTVTSDGAT	78-101 (2)	83.74	63.1	-20.64	-0.0628	0.5804	0.0345	8	0.0739
25	EDNHNYPAQGTVKGTVTSDGAT	80-101 (2)	82.38	62.95	-19.43	-0.0682	0.2842	0.72	8	0.492
26	DNHNYPAQGTVKGTVTSDGAT	81-101 (2)								
27	YTIWENTRVNE	102- 112 (1)								
28	YTIWENTRVNEPSIQG	102-117 (2)								
29	YTIWENTRVNEPSIQGTAT	102-120 (2)	72.88	60.03	-12.85	0.0164	0.1753	0.2799	8	0.7867
30	YTIWENTRVNEPSIQGTATF	102-121 (2)	74.39	60.2	-14.19				8	
31	YTIWENTRVNEPSIQGTATF	102-121 (3)	69	59.07	-9.93	-0.0347	0.2216	0.4698	8	0.651
32	YTIWENTRVNEPSIQGTATFNQ	102-123 (2)	55.96	65.1	9.14	-0.3886	0.5815	2.0046	8	0.0799
33	IWENTRVNEPSIQGTAT	104-120 (2)								
34	NTRVNEPSIQGTAT	107-120 (2)								
35	NTRVNEPSIQGTATF	109-120 (2)	95.7	44.12	-51.58	-0.4629	0.9855	1.409	8	0.1965
36	RVNEPSIQGTAT	109-120 (2)			0					
37	RVNEPSIQGTATF	109-121 (2)	44.54	45.25	0.71	-0.1663	0.2423	2.0597	8	0.0734
38	RVNEPSIQGTATFNQ	109-123 (2)	39.31	34.85	-4.46	0.0313	0.1623	0.0589	8	0.5786
39	RVNEPSIQGTATFNQY	109-124 (2)	61.25	56.38	-4.87	-0.1092	0.2907	1.1267	8	0.2925

Table 4. Continued

			Apo	Xylan		Centroid difference				
ID	**Peptide**	**Residue&Charge**	**D2O%**	**D2O%**	**Change rate%**	**Mean**	**SD**	***T*-test**	**df**	***p*-value**
40	YISVRNSPRTSGTVTVQNHF	124-143 (2)	74.86	57.56	-17.3	-0.1694	0.3407	1.4918	8	0.1741
41	ISVRNSPRTSGTVTVQNHF	125-143 (3)	29.37	20.45	-8.92	-0.1261	0.1172	3.2288	8	0.0121
42	ISVRNSPRTSGTVTVQNHFNA	125-145 (2)	36.33	27.1	-9.23	-0.0848	0.19	1.3397	8	0.2172
43	ISVRNSPRTSGTVTVQNHFNAW	125-146 (2)	37.19	26.7	-10.49	-0.1107	0.1559	2.1303	8	0.0658
44	RNSPRTSGTVTVQNHF	128-143 (2)	23.13	21.46	-1.67	-0.1006	0.2016	1.4967	8	0.1728
45	RNSPRTSGTVTVQNHFNAW	128-146 (2)	24.23	21.05	-3.18	-0.1174	0.1684	2.091	8	0.0699
46	SPRTSGTVTVQNHF	130-143 (2)	10	11.53	1.53	0.0291	0.172	0.5083	8	0.625
47	SPRTSGTVTVQNHF	130-143 (2)								
48	VTVQNHFNAW	137-146 (2)								
49	NAWASLGLHLGQMNY	144-158 (2)	47.44	44.19	-3.25	0.1296	0.1384	2.8082	8	0.0229
50	WASLGLHLGQM	146-156 (2)	10.59	19.87	9.28	-0.0516	0.0991	1.5623	8	0.1568
51	WASLGLHLGQMNY	146-157 (2)	45.47	25.88	-19.59	0.1403	0.138	3.0484	8	0.0159
52	ASLGLHLGQM	147-156 (2)	55.68	20.98	-34.7	0.0291	0.1149	0.7586	8	0.4698
53	ASLGLHLGQMNY	147-158 (1)	32.86	22.55	-10.31	0.3286	0.3459	2.8497	8	0.0215
54	ASLGLHLGQMNY	147-158 (2)								
55	GLHLGQMNY	150-158 (1)	44.18	36.3	-7.88	-0.1752	0.4382	1.995	8	0.2646
56	HLGQMNY	152-158 (1)	19.95	10.76	-9.19	-0.0472	0.2264	0.6252	8	0.5492
57	VEGWGGSGSASQSVSN	163-178 (2)	57.35	42.86	-14.49	-0.0602	0.1874	0.0929	8	0.3638

In order to relay the HDX analysis results to enzyme evolution and improvement, the multiple sequence alignment of xylanase enzymes from different species and families were also carried out as shown in Figure 14. The 10 fungal and 2 bacterial xylanase 1 sequences were aligned together with the *T. longibrachiatum* xylanase. The multiple sequence alignment revealed that some highly conserved regions like peptide "RVNEPSIQGTATFNQY" (residues; 109-124) might be crucial for enzymatic activity since it was significantly stabilized during the substrate binding. This peptide normally refers to the thumb region of the xylanase and locates very close to the active site (Glu75) of the protein (Figure 13A and B). The significant protection indicated its crucial role in substrate recruitment or holding. The evolutionary evidence and structure dynamics information correlated with one another indicate that the region might be one of the structure dynamics determinants for enzyme function. Besides the peptide (covers the residues between 109-124), several other regions were also determined with significantly protected dynamic patterns following the substrate binding, and some of these regions however were less conserved.

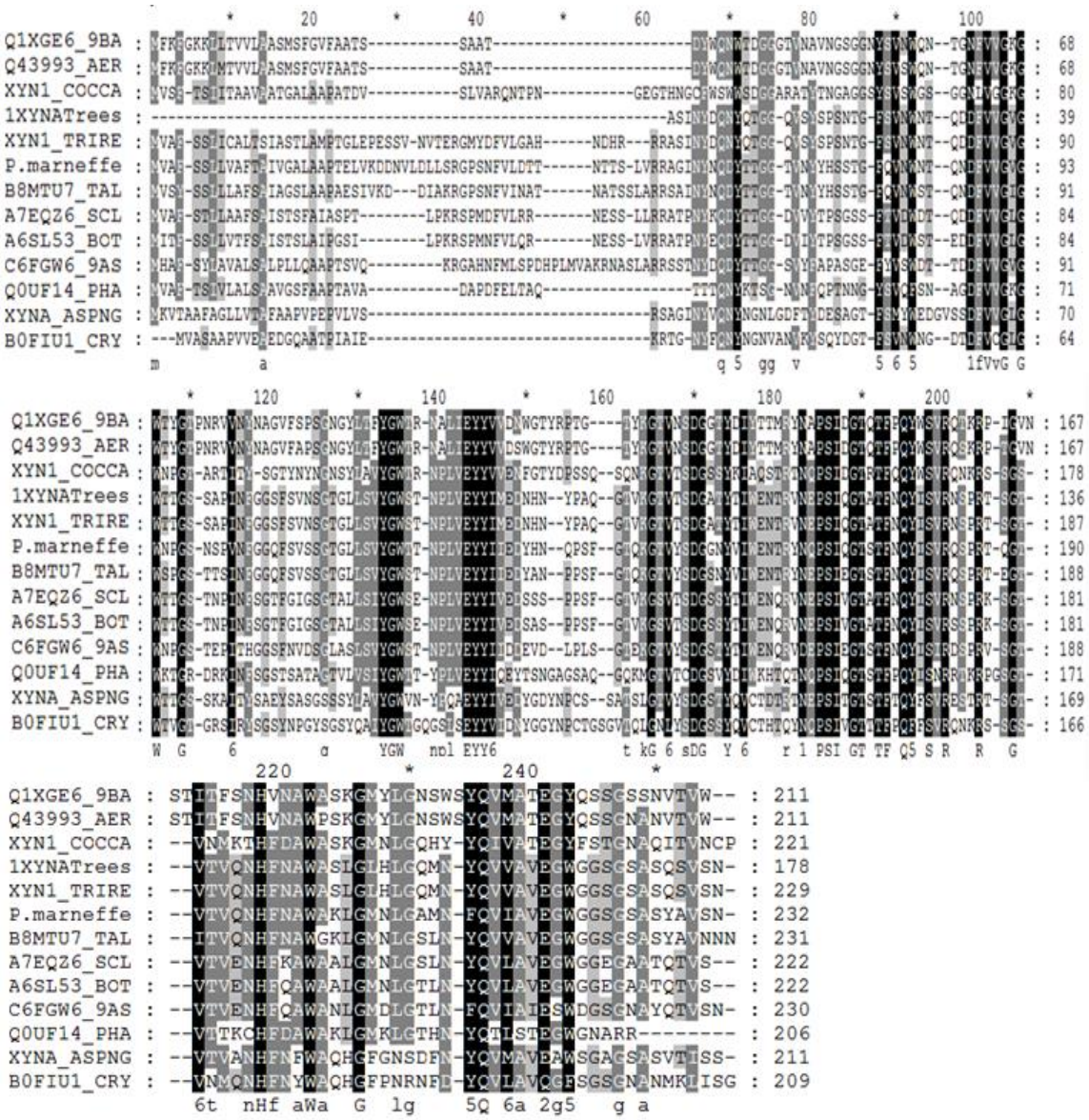

Figure 14. Multiple sequence alignment of homologs of *T. longibrachiatum* xylanase. The alignment includes only XYN I sequences from different organisms. The sequences are endo-1,4-beta-xylanase A from *Paenibacillus sp*. (Q1XGE6_9BA), *Aeromonas punctata* (Q43993_AER), *Cochliobolus carbonum* (XYN I_COCCA), *T. reesei* (XYN I_TRIRE), *P. marneffei* (P. Marneffei), *Talaromyces stipitatus* (B8MTU7_TAL), *Sclerotinia sclerotiorum* (A7EQZ6_SCL), and *Botryotinia fuckeliana* (A6SL53_BOT), *Bispora sp.* (C6FGW6_9AS), *Phaeosphaeria nodorum* (Q0UF14_PHA), *A. niger* (XYNA_ASPNG), and *Cryptococcus flavus* (B0FIU1_CRY).

The current findings therefore point out that HDX-MS analysis can be used to explore a broader range of cell wall degrading enzymes for both mechanistic study and enzyme engineering.

3.3. Results and Discussion

Overall, HDX analysis has revealed significant intrinsic dynamics for xylanase enzyme and such dynamics might be important for the enzyme function. Specifically, different regions of the *apo* xylanase showed a differential HDX rate. The substrate binding leads to significant protection or stabilization effects, where some regions near the reaction sites were significantly stabilized by the enzyme substrate xylohexaose. The xylohexaose binding also induced protection on regions beyond the reaction center. In contrary to xylohexaose, xylan binding induced fewer changes in enzyme structure dynamics, assumingly due to the insufficient binding resulted from the insolubility of xylan. The structure dynamics information for substrate binding indicated that many regions in the protein coordinately changed conformation to fulfill the function of substrate docking and catalysis.

The structure dynamics information also correlated with the enzyme evolution to a certain degree, where the evolutionarily conserved regions were very well protected by the substrate binding. These regions were expected to be essential for the enzyme function. Overall, HDX mass spec analysis allows identifying the novel structure determinants for the enzyme function that could not be found with traditional X-ray or

NMR techniques. HDX mass spectrometry thus provided a novel platform to guide the rational design of enzymes.

HDX-MS findings unveiled the existence of strong dynamic-function relevance for XYN I and such relevance can be explored for the future enzyme improvement. The typical ligand-binding interactions can lead to significant dynamic motions and subsequent stabilizations at both regional and global levels in the structure of catalysts like XYN I. These types of conformational fluctuations occur during ligand bindings and are most likely related to the molecular mechanisms of the enzymatic function that can be identified and explored by using powerful and high-throughput HDX-MS platforms. By considering the current data, it was postulated that the HDX-MS analysis of cell wall degrading enzymes provides a novel platform to guide the rational design of biocatalysts.

CHAPTER IV

STRUCTURE DYNAMICS GUIDED-ENZYME IMPROVEMENT STUDIES FOR XYN I FROM *Trichoderma reesei*

4.1. Introduction

To produce ideal and desired biocatalysts for industrial applications, natural enzymes often require the optimization of their catalytic efficiencies. The engineering of proteins for improved function and new characteristics is a fast-growing field in biotechnology. A number of enzyme engineering strategies are currently employed to modify biocatalysts in terms of improving their suitability for large-scale industrial applications [129]. In general, these attempts include various directed evolution practices, semi rational design techniques, and more recently, the de novo design of novel enzymes.

The directed evolution of catalysts primarily requires the natural evolution of gene sequences by random mutations which result in the creation of large mutant libraries (10^3-10^6 mutants) displaying a high level of sequence diversity. The created libraries are then investigated through high-throughput screening to identify recombinants with desired characteristics. Directed evolution enabled some success in producing cellulases and xylanases with improved activity [130-132]. Nevertheless, it is not quite sufficient to provide mechanistic insights into the protein structure-function relationship.

De novo design of enzymes is another engineering strategy and commonly requires the modeling of the enzyme transition states specific to functionality. It theoretically offers the modification of atomic positions in a particular configuration and aims to obtain the enzymes with desired activities through the computational rational design.

The rational design of catalysts however, is an information-concerted process requiring the understanding of the mechanisms such as prior knowledge of the sequence, 3D crystal structure, structure-function relationship, and the mechanisms of catalysis or inhibition. The rational design strategies are capable of integrating a variety of several strategies to generate engineered catalysts with enhanced activity, efficient specificity, and increased stability [22, 23].

HDX-MS stands as a high-throughput tool enabling the rational design of catalysts through the identification of protein structure dynamics during ligand binding. By taking advantage of HDX-MS, the previously identified structure dynamics knowledge (Chapter III) of the model enzyme XYN I from *T. longibrachiatum* was employed in order to engineer and obtain better xylanase biocatalysts for various industrial applications. The highly dynamic regions identified in the 3D structure of XYN I protein were targeted for site specific modifications. The thumb and cord regions of the protein revealed a high degree of dynamic changes through binding with different substrates compared to the other parts of the protein [81]. The related findings call attention to these regions in terms of their roles during substrate binding. It is therefore hypothesized that the stabilization of the thumb and/or cord regions can help to improve

the various catalytic characteristics of XYN I. Due to its highest dynamic motion profile, the thumb region of the XYN I protein was the primarily focus in terms of the engineering applications targeted to obtain better XYN I recombinants.

4.2. Material and Methods

4.2.1. Materials

Oligonucleotide primers were synthesized by Integrated DNA Technologies Inc (IDT) and the sequences are presented in Table 5. Beechwood xylan was purchased from Sigma Aldrich. *T. reesei pyrG* mutant (*T. reesei* Qm9414, *pyrG* auxotroph), *Saccharomyces cerevisiae* YM10 strain (*S. cerevisiae*), and pRS426 phagemid plasmid for the fungal expression cassette construction were kindly provided by Dr. Ziyu Dai, at the Pacific Northwest National Laboratory (PNNL) in Washington state, USA. *E. coli*-based expression vector pET32a was purchased from *Novagen Inc.* and used for the cloning and sequence-based confirmations of the XYN I recombinants.

4.2.2. The Introduction of Amino Acid Substitutions into XYN I Coding Sequence

The engineering of XYN I protein was performed via the traditional PCR-based introduction of amino acid substitutions. When the substitutions designed, codon preference of *T. reesei* was always considered. Three consecutive PCR reactions were followed for each substitution within the coding sequence.

Table 5. List of the primers used to engineer XYN I enzyme.

Primer	*Sequence*
N111D F	GACGAGCCTTCCATCCAGGGC
N111D R	CTGCTCGTTGACACGGGTATT
N111D Ext	GAGAATACCCGTGTCAACGACGAGCCTTCCATCCAGGGC
E112C F	TGTCCTTCCATCCAGGGCACAGC
E112C R	ACAGTTGACACGGGTATTCTCCCA
E112C Ext	GAGAATACCCGTGTCAACTGTCCTTCCATCCAGGGCA
E112Q F	CAGCCTTCCATCCAGGGCACA
E112Q R	CTGGTTGACACGGGTATTCTCCCA
E112Q Ext	GAGAATACCCGTGTCAACCAGCCTTCCATCCAGGGCA
E112I F	ATCCCTTCCATCCAGGGCACA
E112I F	GATGTTGACACGGGTATTCTCCCA
E112I Ext	GAGAATACCCGTGTCAACATCCCTTCCATCCAGGGCA
S97C F	TGCGACGGAGCCACTTACACCATC
S97C R	GCAGGTGACGGTTCCCTTGAC
S97C Ext	GTCAAGGGAACCGTCACCTGCGACGGAGCCACTTACACC
S64C F	TGTGTCTATGGCTGGAGCACCAAC
S64C R	ACAAAGCAGGCCAGTTCCGCT
S64C Ext	AGCGGAACTGGCCTGCTTTGTGTCTATGGCTGG
N141C F	TGCCACTTCAATGCTTGGGCC
N141C R	GCACTGCACAGTAACAGTTCCGCT
N141C Ext	GGAACTGTTACTGTGCAGTGCCACTTCAATGCTTGGGCC
T70R F	CGCAACCCACTGGTTGAGTAC
T70R R	GCGGCTCCAGCCATAGA
T70R Ext	TCTATGGCTGGAGCCGCAACCCACTGGTTGAGTAC
Q116E F	GAGGGCACAGCGACCTTCAAC
Q116E R	CTCGATGGAAGGCTCGTTGACACG
Q116E Ext	GTCAACGAGCCTTCCATCGAGGGCACAGCGACCTTC

Internal primers were designed and used to introduce the specific mutations into the coding sequence of XYN I (Figure 15). A specific extension primer was also synthesized for each mutation and used to extend the fragment (usually C terminus fragment) length in order to increase the overlapping sequence between the two pieces of the gene during a second PCR. Those two pieces previously amplified were recombined back into one piece of DNA sequence as a result of the third PCR.

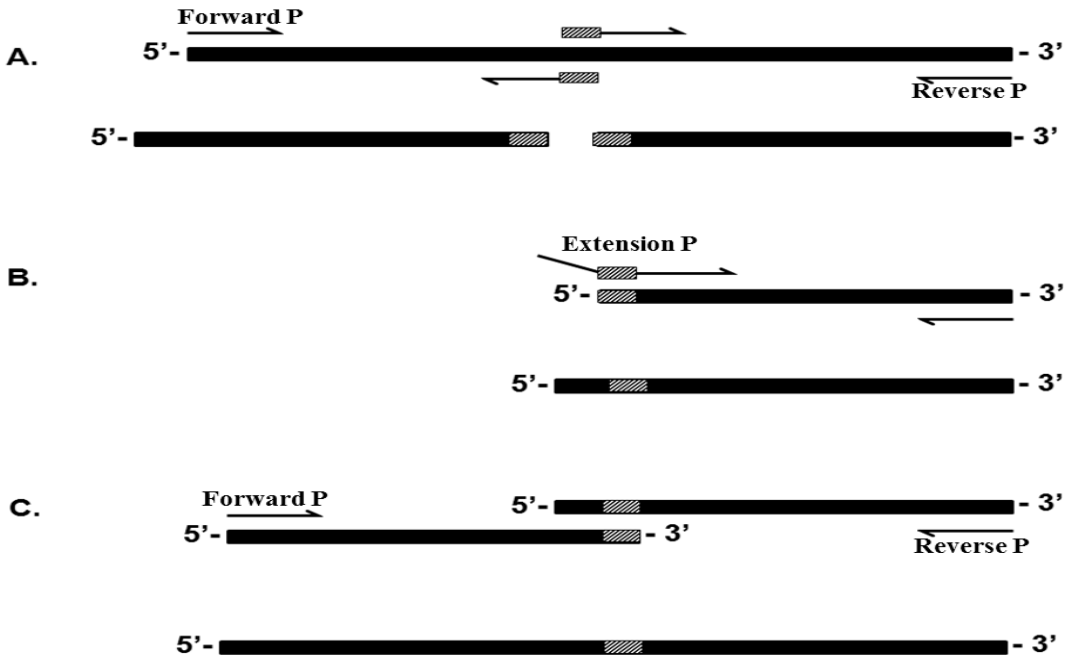

Figure 15. Consecutive PCR-based introduction of amino acid substitutions into XYN I gene fragment. A) Internal and external primers were employed to introduce single amino acid substitutions into XYN I during first PCR reaction. B) During the second PCR, an extension primer was designed and used together with reverse primer in order to increase the length of overlapping sequences between the two fragments from 1st PCR. C) By using parental primers of XYN I, a consecutive third PCR was performed to recombine the two fragments of XYN I gene before cloning and sequence confirmation.

This strategy was followed for the each amino acid substitution related to the engineering and creating HDX data-driven specific recombinants of XYN I. The created recombinants of XYN I together with wild type XYN I are given in Table 6. In order to easily purify the XYN I proteins from fungal cultures, a specific His tag sequence was added to the 3' ends of each recombinants.

4.2.1. The construction of fungal expression cassettes for each XYN I recombinant

In light of the HDX-MS data for the protein, XYN I recombinants were created by employing traditional PCR strategy. In order to test the possible characteristics of each recombinant, the expression cassettes targeting the production of

Table 6. The list of the XYN I recombinants created.

XWT
N111D
E112Q
E112I
E112C
S64C
E7:T70R,Q116E
Q4: T70R, Q116E, S97C, N141C

proteins in fungal host were constructed. The required components for the construction of the expression cassettes were PCR amplified from the specific templates mentioned above. The specific primers carrying about 30 bp overlapping sequence in total were used to construct the expression cassettes in yeast cells. The primers used for the expression cassette construction are given in Table 7. A specific *S. cerevisiae*-based construction strategy was employed to combine all components. The italic sequences

refer to the *Pme I* restriction sites which were later utilized in order to exclude the expression cassettes from pRS426 phagemid backbone vector.

4.2.3. The Yeast-Based Construction of Fungal Expression Cassettes

In order to express each XYNI recombinants in *T. reesei*, a yeast-based construct building method, also known as non-homologous end joining (NHEJ) strategy, was followed [133]. This in vivo gap-repair cloning method in yeasts has been recently recognized as one of the most efficient applications for error-free and high fidelity construction of expression cassettes [134].

A representative scheme of the construct built for the *T. reesei* based expression of XYN I recombinants are shown in Figure 16. Yeast-based gap repair enabled the successful integration of each incorporated DNA fragments together with specific selection marker, and mitigated any PCR-based mismatching and related amplification errors during the expression cassette constructions for fungal transformation and homologous recombination.

Table 7. Primers operated to build the fungal expression cassettes specific to filamentous fungus *T. reesei.*

Primers	Sequence
A1	gtaacgccagggttttcccagtcacgacg*GTTTAAAC*GTGGCAACAAGAGGCCAGAG
B1	ATTGGACTGAGTGAAGAAGCCGTTGGCAAATTAC
A2	ATTTGCCAACGGCTTCTTCACTCAGTCCAATCTCA
B2	GAGGCCTGTGGGCATAGCACGAGCTGTGGCCAAGAA
A3	GCCACAGCTCGTGCTATGCCCACAGGCCTCGAGCCT
B3	AGGAAATCATGACTCGTGGTGATGGTG ATGATGAGAACCACGGTTGCTGACACTCTGTGA
A4	CATCATCACCATCACCACgagtcatgatttcctcttgg
B4	TTTCGCCACGGAGCTtcatgacttgccgcatactc
A5	**TGCGGCAAGTCATGAAGCTCCGTGGCGAAA**
B5	gcggataacaatttcacacaggaaacagc*GTTTAAAC*CCGCGCCGGGAAATTCTTT

A uracil and tryptophan auxotroph YM10 strain of *S. cerevisiae* was used during the construction of expression cassettes of the XYN I recombinants.

The orotidine-5′-phosphate decarboxylase (*pyrG*) gene of *T. reesei* was used as selective marker during the fungal expression of recombinants. The co-transformation of competent yeast cells by using 100 ng of the each construct component amplified by

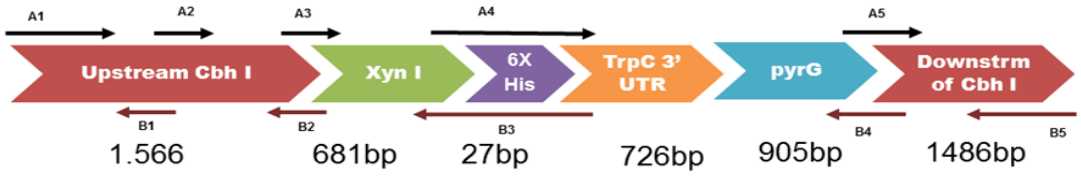

Figure 16. The schematic visualization of the *T. reesei* expression cassettes constructed within yeast cells.

PCR was performed together with 100 ng of linearized pRS426 phagemid backbone vector. YM10 yeast strain was first inoculated into 5-10 ml of YPD (Bacteriological peptone 20 g/L, yeast extract 10 g/L, and glucose 20 g/L) culture, incubated overnight at 30 °C with continuous shaking. The following day, 50 ml YPD in 250 ml flask was inoculated with 1 ml of the overnight culture and incubated at 30 °C until the $O.D_{600}$ value reaches 1. The cells were collected by centrifugation at 5000 rpm for 2-5 min, washed with sterile H_2O, and then, resuspended in 1 ml 100 mM lithium acetate, centrifuged at top speed for 15 second; the supernatant was discarded. The cell pellet was then resuspended in 400 µl 100 mM lithium acetate (if original O.D= 1) and delivered into the eppendorf tubes as 50 µl aliquots. Before transformation, the aliquots were spun down and the supernatants were removed.

The co-transformation of each expression cassette component into the yeast cells was performed by using a 360-X µl transformation mixture without the fragments (where X is the total volume of all PCR fragments, vector etc.). The transformation mix per sample was as followed; 240 µl 50 % PEG 3350, 36 µl 1 M lithium acetate, 50 µl sheared salmon sperm DNA (2 mg/ml), 34-X µl sterile H_2O. After that, each component of the construct, linearized vector DNA and specific recombinant of XYN I were added into the tubes, homogenized by inverting and then, heat shock was applied in a 42 °C water bath for 40 min. The cells were then spun down for 15 second at 6000 rpm; the cell pellets were washed with 1 ml sterile H_2O by gentle pipetting. 100-200 µl of YPD medium was used to resuspend cells and spread onto SC-Ura plates and the plates were incubated for 2-3 days in a 30 °C incubator. Positive colonies, the representatives of

successful transformation, were scraped off in order to perform the genomic DNA isolations as described earlier [135].

4.2.4. *E. coli* Transformation

Although total genomic DNAs can be directly used to enrich the constructs via PCR, the size of the constructs and PCR conditions chosen, however, may adversely affect the sequence quality and render sequence based errors in the coding sequences. In order to ensure the sequence fidelity, an *E. coli* based enrichment strategy was successfully performed for the yeast-based expression constructs. The genomic DNAs isolated from yeast cells were used for the general *E. coli* transformation. The positive transformants were selected on LB plates including ampicillin antibiotic. In light of current experiences, nevertheless, a common *E. coli* transformation strategy based on heat shock at 42 °C may not work all the time. In this case, it is the best choice to use an electroporation method to transfer the recombinant plasmid DNAs into TOP10 or DH5α (*Invitrogen Inc.*) *E. coli* competent cells. Positive colonies were picked and inoculated into LB broth medium. Plasmid isolation from liquid cultures was performed by following a general plasmid isolation protocol in order to isolate pRS426 plasmids specific to each XYN I recombinant.

In order to confirm the true size of the constructs built in yeast cells, size-specific PCR amplifications and restriction reactions were examined. Specific *Pme I* restriction sites located at both ends of the constructs were used to confirm the correct size and

assembly of the constructs. The XYN I recombinants-enclosed expression cassettes excluded from the pRS426 backbone vector were next used as the templates for the fungal transformation of *T. reesei.*

4.2.5. Fungal Transformation

The pRS426-excluded linear expression cassettes of each XYN I recombinants were used to transform a *pyrG* auxotroph of the *T. reesei* fungus by following the protoplast-based transformation strategy. The upstream and downstream untranslated (UTR) regions of the cellobiohydrolase I (CBH I) gene from *T. reesei* were amplified and used to perform the true integration and homologous recombination of expression cassettes into the *T. reesei* genome. The CBH I promoter together with its secretion signal was selected to ensure the large scale expression, production, and secretion of the recombinant XYN I enzymes into the liquid growth medium. For this reason, the native signal peptide sequence of XYN I protein was replaced with that of CBH I.

The conidia were harvested from 7-12 day old PDA plates of *T. reesei pyrG*-mutant strain and 10^6 -10^7 spores/ml inoculated into 100 ml PDB growth medium. The suspension was incubated at 28 °C for 14-16 hours by shaking at 150 rpm on orbital shaker. The fungal material was collected using 100 micron nylon filter (HC3-100, sefar American; Tetko) or two layers of sterile miracloth, washed three times with ice cold ddH2O. The hydrolyzing enzyme mixture was prepared within protoplasting buffer (50 mM MES, 50 mM CaCl2, and 0.5 M Mannitol, pH 5.5), filter sterilized. The enzyme

complex to create protoplasts from the fungal mycelia was prepared within protoplasting buffer as a combination of 2 mg chitinase (approx. 300 units/g solid; Sigma C-6137), 7 mg lyticase (approx. 275 units/mg solids; Sigma L-4025), and 44 mg cellulase (Onozuka R-10). Fresh fungal mycelia (0.6-0.7 gram) were transferred into 15 ml centrifuge tubes and dissolved with 2.4 ml of protoplasting buffer, and incubated at 28 °C for 30 - 45 min by shaking at 200 rpm. The hydrolyzing enzyme mixture was prepared within the protoplasting buffer (50 mM MES, 50 mM $CaCl_2$, and 0.5 M Mannitol, pH 5.5).

In order to separate the undigested mycelia from the protoplasts produced, the suspension was filtered through a 10 µm nylon filter. The spores were collected by centrifuging the supernatants at 4000 rpm for 5 min, eluted within 1 ml of protoplasting buffer to determine the concentration. The 1-2 µg from the linear expression DNAs of XYN I recombinants were added to a 100 µl of protoplast suspension, incubated on ice for 20 min. A 60 µl amount of 50% PEG 6000 (sterile) prepared in protoplasting buffer was added to the tubes at RT. The suspensions were mixed by inverting the eppendorf tubes several times and the second addition of 60 µl of 50% PEG 6000 was performed and mixed, incubated at room temperature for 30 minutes. For the regeneration of the protoplasts, the osmotic density of MM was stabilized with 1 M sorbitol. The mixture was then homogenized with 0.8 % agar, poured onto MM selection plates, and incubated at 28 °C for 3 days.

4.2.6. Enzyme Expression and Purification

The positive transformants were first transferred onto the minimal medium (MM) selective cultures. The transformants were incubated on MM medium for four generations in order to obtain homologous recombinants. Since the *T. reesei pyrG* strain is not able to grow on MM, it can only grow in MM when a specific concentration (0.5%) of uridine is added. Hence it was used as negative control during the all transformation experiments performed. The recombinant *T. reesei* strains were inoculated onto PDA agar plates and incubated for about 5-7 days at 28 °C until the sporulation occurs. In order to initiate the expression of XYN I recombinants under the CBH I promoter, the spores were washed from PDA plates by using 10 ml of MM. After that, the spores were filtered via miracloth before inoculating into MM to exclude the mycelia from the inocula.

Together with *T. reesei pyrG* mutant strain as a negative control, a 10^6 spores/ml of each recombinant line were inoculated into 100 ml of MM containing 1 % glucose as sole carbon source. Spore germination and mycelia formation were induced as in 250 ml flasks on an orbital shaker with 48 hours incubation at 28 °C (200 rpm). Following formation of mycelia after two days, the biomass produced was filtered through 1 layer of miracloth to replace the media with fresh MM including 2 % wheat bran instead of glucose. Two days incubation was performed for the induction of the recombinant protein expression. The supernatants from each culture were then harvested by filtering through miracloth. The extracts were further centrifuged at 10000 rpm for 15 min in

order to remove possible mycelium artifacts and cell debris. The culture filtrate was then precipitated by adding gradual concentrations of ammonium sulfate (up to 80 %). The suspension was centrifuged at 10000 rpm for 30 min. The protein pellets were eluted within “His Buffer A” (100 mM NaCl, 40 mM imidazole, 1 % glycerol, 20 mM Tris-HCL pH 8.0).

The protein suspensions were treated with Qiagen Ni-NTA superflow resins in order to purify His-tagged XYN I recombinants out of the secretome of *T. reesei* cultures and incubated on a rotary shaker at 4 °C for four hours. After 6XHis resin-based incubation, the His resins were washed three times with washing buffer (50 mM NaH_2PO_4, 300 mM NaCl, and 20 mM imidazole; pH 8.0) in order to remove the non-specific protein contaminants. The bound proteins were then eluted by using elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, and 250 mM imidazole; pH 8.0). To collect the all bound proteins completely, the His resins were washed two times with elution buffer. The dialysis of the eluted proteins within sodium phosphate citrate buffer (pH 4.5) was performed overnight. The fungal expression levels and the purity of the XYN I recombinant proteins were examined by SDS page analysis (Figure 17 and Figure 18). The concentrations of the eluted proteins were identified by using a Pierce Commassie Protein Assay kit (*Thermo Scientific Inc*) with spectrophotometric measurements at 595nm.

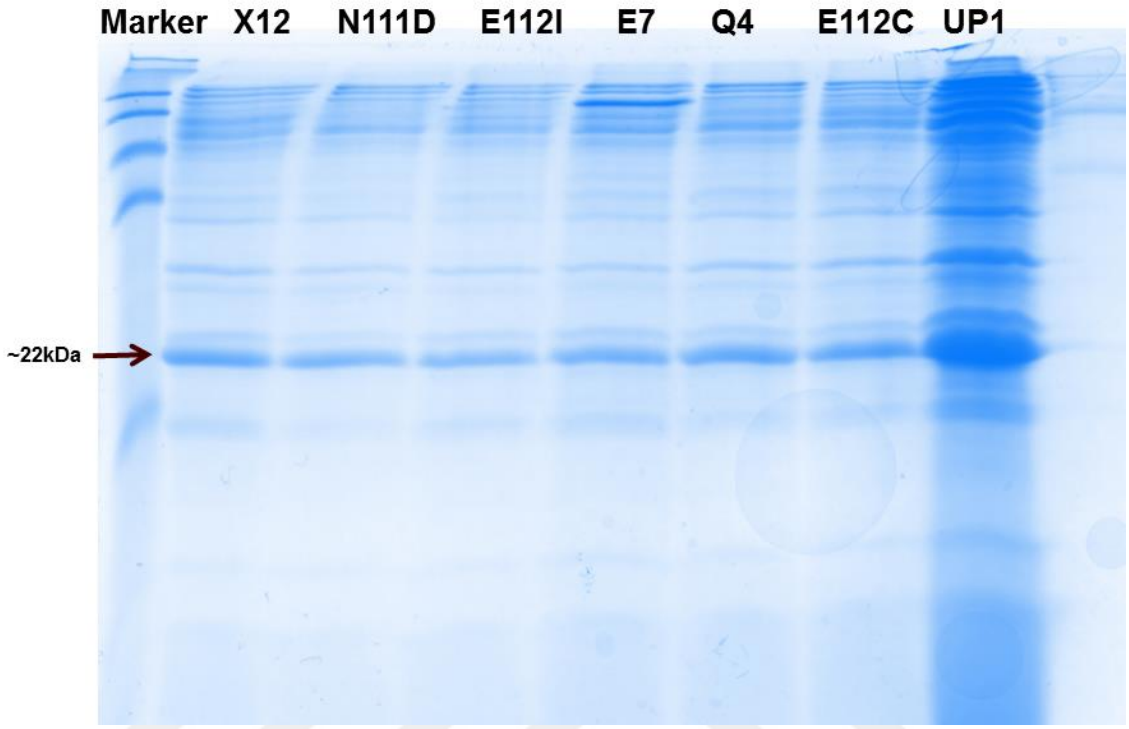

Figure 17. High-level expression of XYN I recombinants in *T. reesei*. The recombinants were expressed under the CBH I promoter and secreted into the culture medium. The figure illustrates the expression levels of the proteins before purification. X12- wild type XYN I; UP1: unpurified total protein before any 6XHis resin treatment.

The protein concentrations of the recombinants were normalized to the same enzyme concentration by using sodium citrate buffer before performing any activity-based quantitative assays. This normalization action enabled the identification of the first hand activities of the engineered recombinants compared to the wild type XYN I.

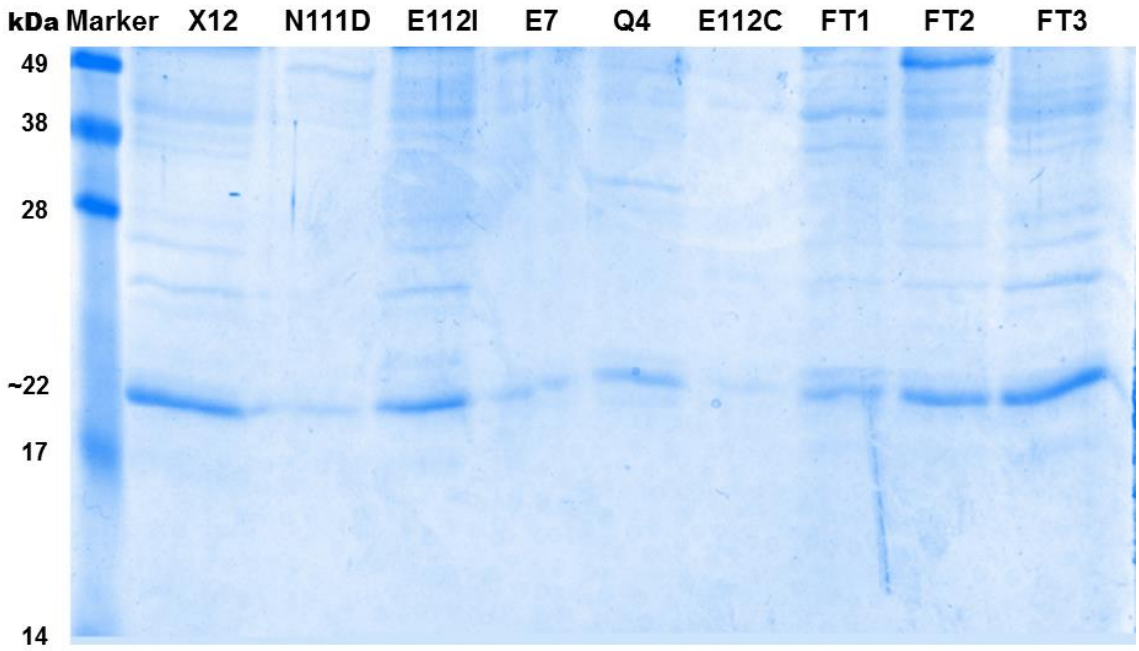

Figure 18. The Ni-NTA superflow-aided purification of XYN I recombinants from the fungal cultures of *T. reesei*. FT refers to flow through from the first washed his resins. FT1: flow through from X12; FT2: flow through from N111D (Asn111Asp); FT3: flow through from E7.

4.2.7. Xylanase Activity Assays

Xylanase activity of the XYN I recombinants was determined by measuring the amount of reducing sugars from hydrolyzed beechwood xylan using the DNS (3,5-dinitrosalisylic acid) method previously described by Bailey [136]. The enzymatic activity evaluations of the XYN I recombinants were carried out at pH 4.0 at 50°C, with the reaction time of 30 min. These conditions were also followed for the analysis of enzyme kinetics. The main substrate beechwood xylan was prepared within 0.05 M sodium citrate buffer (pH 4.0), boiled and then cooled down to room temperature while

stirring. 250 µl of 1 % beechwood xylan substrate was mixed with 125 µl of enzyme suspension following the dilution of recombinant enzyme concentrations to the measurable levels. After incubation at 50°C for 30 min, the reaction was terminated by adding 500 µl of DNS reagent [8]. The mixture was then boiled for 10 min and cooled to room temperature. The absorbance values of released sugar molecules were identified at 540 nm by using Beckman Coulter DU 640 spectrophotometer. The same experimental design was also followed for the identification of the kinetic parameters of each XYN I recombinant. Each reaction and the related controls were run in triplicate.

4.3. Results and Discussion

T. reesei (teleomorph of *Hypocrea jecorina*) is an industrially important filamentous fungus due to harboring both excellent secretion capacity of numerous enzymes essential for lignocellulosic biomass degradation and accessible fermentation conditions. As a eukaryotic model system, it has been well developed for the production of various hydrolyzing enzymes (e.g., 40 g/L of CBH I) [137]. In addition, having the capacity for both extracellular and large amount of production of cell wall degradation enzymes enables the successful expression of heterologous eukaryotic proteins in this model system. Thus, the homologous expression of XYN I recombinants was performed under the highly inducible CBH I promoter in *T. reesei.* The purified proteins from *T. reesei* cultures were subject to quantitative enzyme analysis to identify the effects of each specific amino acid substitution.

The enzyme assay results for the XYN I recombinants created are shown in Figure 19. The engineering studies of XYN I was mostly focused on the thumb region of the structure regarding the high regional dynamics. Therefore, the hydrophilic amino acid residues in the thumb region were first determined, through the comprehensive structural and residue-based analysis, few amino acid residues were identified and substituted with hydrophobic amino acids in order to increase the hydrophobicity index of the thumb region. E7 and Q4 are the two recombinants of XYN I with higher enzymatic activity. Both recombinants contain two different types of stabilization within the structure of XYN I. The design of the E7 mutant aims to stabilize the thumb region slightly by creating an ionic interaction between the thumb region and the neighboring β-sheet wall.

It is hypothesized that the slight stabilization of the thumb region may not disrupt its binding function and would still allow functionality specific to ligand binding. The XYN I recombinants were also tested in terms of their pH optima. The best two XYN I recombinants were characterized in terms of their kinetic parameters, pH, and residual activities. The sodium phosphate-citrate buffer was utilized in order to perform the pH-based activity assays of the mutants.

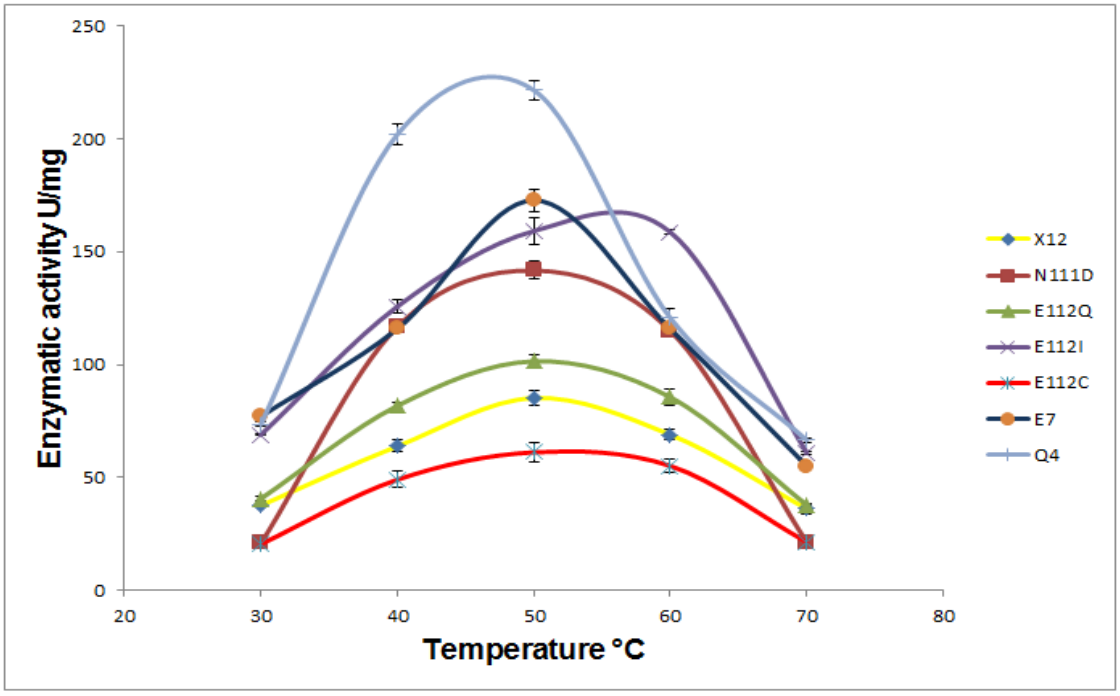

Figure 19. The enzymatic activity results of XYN I recombinants through different temperatures. E7 includes the substitutions; Thr70Arg and Gln116Glu. Q4 contains four amino acid substitutions: Thr70Arg, Gln116Glu, Ser97Cys, and Asn141Cys.

As XYN I is an acidic enzyme compared to XYN II, the activities were tested with the pH interval from 2 to 8 over different pH values shown in Figure 20.

This data suggest that the enzyme activity of the Q4 mutant showed greater pH dependence than E7 and X12. On the other hand, the E7 mutant displayed a better pH interval compared to wild type X12 and Q4, indicating that the amino acid residues (Asn and Ser) substituted with cysteine for the disulfide bridge introduction might have some roles to accept or release protons at different pH values due to their OH groups.

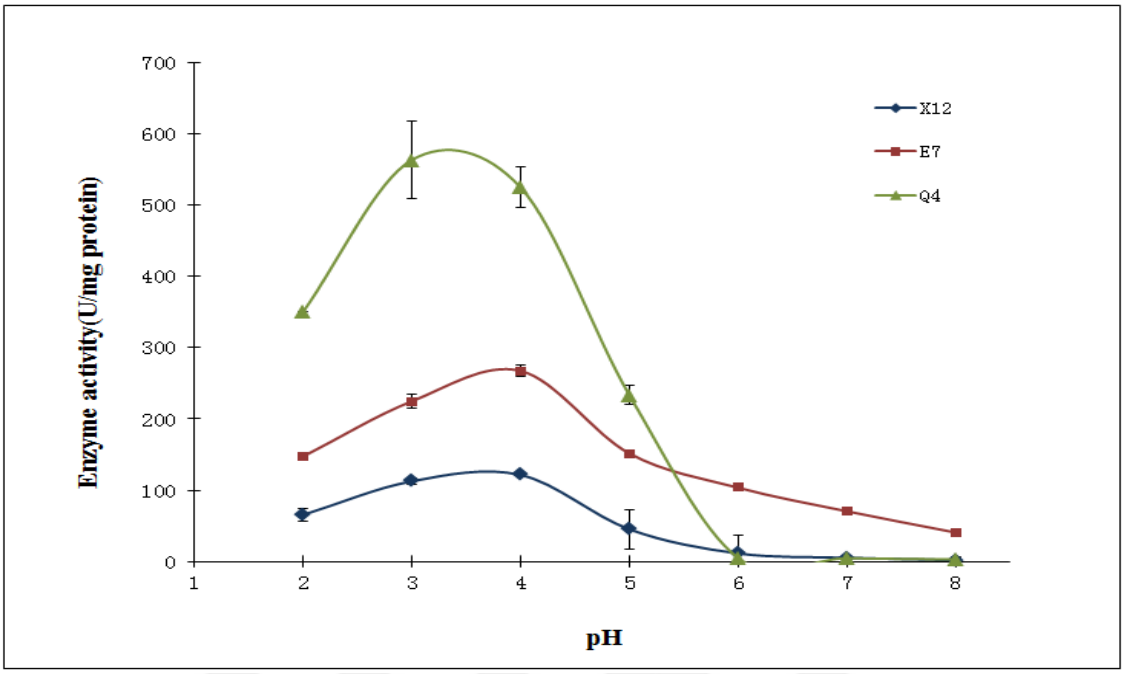

Figure 20. The pH optima of the two recombinants of XYN I protein. X12 is wild type XYN I. E7 and Q4 were tested through different pH values for their best activity results. The reaction conditions are identical to those previously described.

The residual activity based analysis data further proved that the Q4 and E7 recombinants are better enzymes than the wild type. The half-life of the wild type XYN I (X12) is around 20 min in the current study, the half times of both E7 and Q4 mutants were much higher. The half-life of E7 is around 150 min, indicating the improvement on the half-life of the wild type XYN I over seven times. On the other hand, the Q4 mutant shows a half-life of 250 minutes, which is more than a 10 fold increase compared to that of the wild type (Figure 21).

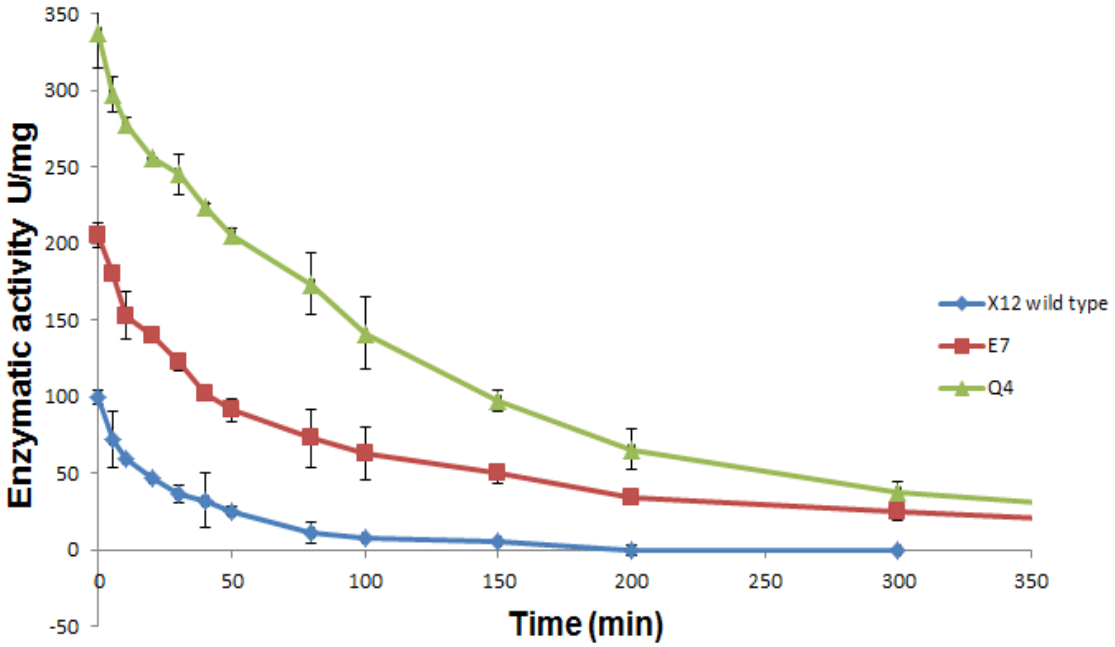

Figure 21. The residual activity and related half-life results of the two XYN I recombinants as compared to wild type (X12). The enzymes were treated at 50 °C without any substrate over time and then the substrate was added to the enzyme aliquots, incubation was performed for 30 min for the same conditions. The hydrolysis reaction was terminated by adding DNS solution.

In order to better understand the mechanism of the stabilization on the 3D structures of the both E7 and Q4 mutants, a deep look into the structure of XYN I is essential. Considering the prominent role of the thumb region for the open and closed conformations of XYN I enzyme for ligand binding, it is hypothesized that optimized stabilization of thumb region may result in desired characteristics. Thus, an ionic interaction was introduced to reduce the flexibility of the thumb region. Negatively (Glu) and positively charged (Arg) amino acid residues were chosen to replace glutamine

(Gln116) and threonine (Thr70) in the original sequence and produced an improved enzyme of E7 (Figure 22).

Furthermore, the introduction of two cysteine amino acids into the other two highly dynamic regions of E7 mutant resulted in an improved enzyme Q4. The hypothesis is that introduction of a disulfide bond might have strongly stabilized the mutated region. In turn, it is possible that the introduction of a disulfide bridge onto a place beyond the reaction center of the protein may increase the stability but not affect the enzymatic activity. Meanwhile, the previous data suggested that the α-helix of XYN I and the downstream part of the thumb region on the back site is quite dynamic (see the peptide 23 and the peptide 40 from the XYN I HDX data for better understanding of their deuteration levels; Table 3).

For instance, the peptide 23 "YIMEDNHNYPAQGTVKGTVTSDGAT" from XYNI showed high deuteration level suggested that the peptide region is quite dynamic in the hydrogen deuterium exchange experiment.

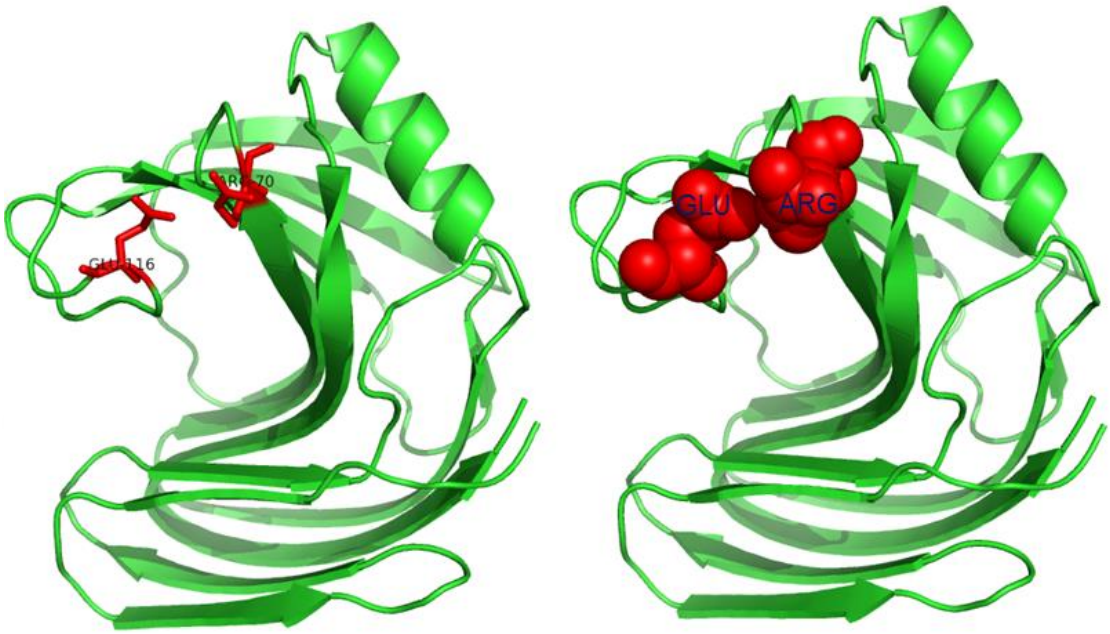

Figure 22.The specific positions of the E7 recombinant-specific substitutions on the 3D structure of XYN I. The related substitutions are Gln116Glu and Thr70Arg.

In addition, the neighboring region peptide; the peptide 40 "YISVRNSPRTSGTVTVQNHF" also revealed a high level of deuteration (74.86%) for the apo enzyme. The close neutral amino acid residues (Ser97 and Asn141) on these regions were identified and replaced with cysteine residues in order to introduce a disulfide bridge to improve thermostability of the E7 recombinant of XYN I (Figure 23).

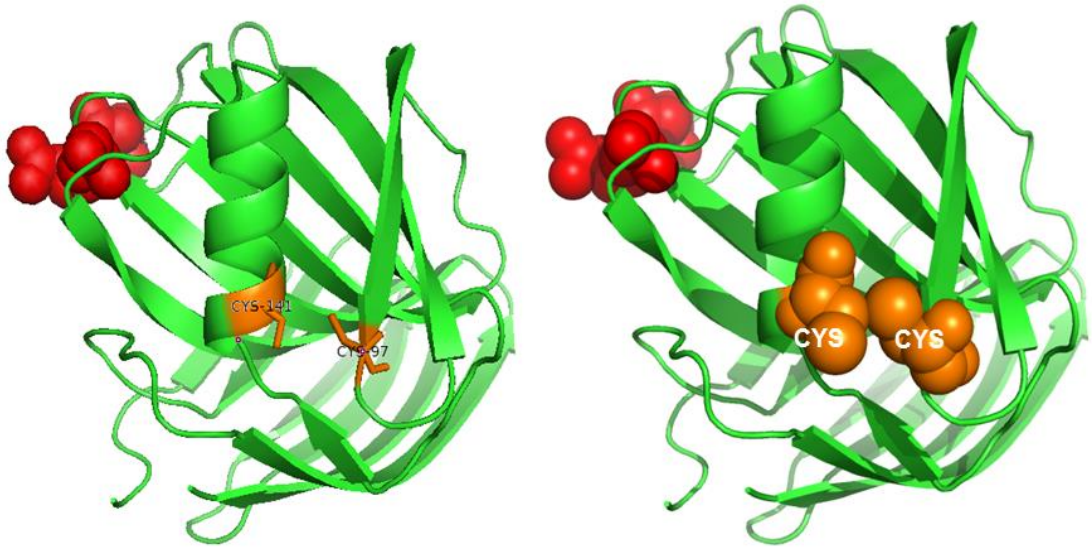

Figure 23. The specific positions of the Q4recombinant-specific substitutions on the 3D structure of XYN I. The related substitutions are Gln116Glu, Thr70Arg, Ser97Cys, and Asn141Cys.

The Q4 recombinant therefore contains the synergistic impact of four substituted amino acid residues. This attempt resulted in a better recombinant compared to both wild type XYN I and E7 proteins.

In conclusion, the stabilization efforts on XYN I protein based on the HDX-MS data resulted in promising findings and provide insights into the enzyme engineering applications through the knowledge of structure dynamic and structure function relationships.

CHAPTER V

DIFFERENTIAL DYNAMICS OF THE TWO STRUCTURALLY SIMILAR BIOCATALYSTS

5.1. Introduction

Utilization of cellulolytic and hemicellulolytic enzymes in industry, especially for the saccharification of lignocellulosic biomass in biorefinery are increasing at a rapid rate [58, 60, 61, 80, 81]. Due to their potential for industrial applications, these enzymes have been subject to vast engineering studies, in particular through rational design or direct evolution [22-24]. The fundamental knowledge obtained from the extensive previous studies make these enzymes perfect models to study protein structure-function relationship. Despite the tremendous amount of work, limited progress has been made as a result of the enzyme improvement efforts through rational design based solely on 3D structure [24]. Therefore, additional information in enzyme structure-dynamics and structure-function relationship is needed.

Recent studies have shown that structural motions might have significant roles during enzymatic catalysis because enzyme functions can depend heavily on the intrinsic dynamics of proteins [8, 11, 22, 104-107, 138-141]. In particular, Dorothy Kern's group recently discovered how the intrinsic dynamics of a protein can be harnessed for enzyme

catalysis [9, 106]. Moreover, because of its possible importance in functionality, structure dynamics has been suggested as driving force for enzyme evolution [142].

In spite of the current progress, only few studies have reported the comparison of the intrinsic dynamic patterns of enzymes with similar 3D structures, yet such a comparison can help to elucidate the role of structure dynamics in evolvability and functionality of enzymes. The structure dynamics analysis of the two similar enzymes from different organisms presented below provides insights into to their intrinsic dynamics and functionality patterns.

Several platforms can enable the study of enzyme structure dynamics analyses like NMR, X-ray, molecular dynamics, fluorescence resonance, and others [11]. Each technology has its pros and cons. Among these techniques, HDX-MS platforms emerge as relatively high-throughput solution enabling multiple enzyme comparisons [97, 103]. The fundamental concept of HDX-MS analysis is based on the mass increase of a protein when the protein protons exchanged with solvent deuterium [91]. The rate and percentage of the H/D exchange can be measured by mass to charge ratio (m/z) of the protein. HDX-MS can be also used to study the global and regional protein conformational changes with different platforms [92, 93]. Coupled with protein digestion and chromatographic separation, HDX mass spectrometry is able to profile different regions of protein for H/D exchange based on the peptide H/D exchange rate and percentage. The gathered information allows understanding of which region of the protein is more stabilized or destabilized based on solvent exchange information and this information is translated as the knowledge of the protein structure modifications towards

ligand binding [8, 88, 89, 93-102, 127].HDX –MS platform comes with the advantages of mass spectrometry analysis: fast, straightforward, radioactivity free, environmental friendly, and less protein input and purity required [88].

Even though the recently developed HDX-MS techniques are recognized by the scientific community and have been used to study protein-ligand interactions, HDX-MS coupled with enzyme digestion for regional protein dynamics analysis has not been widely applied to cellulolytic and other enzyme studies. The recent studies indicate that HDX-MS can potentially provide useful structure dynamics information to enzyme engineering [81]. Through the HDX-MS based analysis, it was identified that the XYN I enzyme exhibits a highly dynamic pattern during ligand binding and pointed out the possible importance of several peptide regions essential for the functionality of XYN I with highly remarkable motions.

In the present study, cellulolytic endo-β-1,4-glucanase (Eg1A) A from *Aspergillus niger* was chosen as a model catalyst and its structure based motions were defined using HDX-MS platform through binding with different ligands. One of the most remarkable reasons of this study is that *A. niger* Eg1A reveals a high structural similarity to that of *T. longibrachiatum* XYN I. The identification of the intrinsic dynamics of the two structurally similar, but functionally different biocatalysts therefore can provide insights into their ligand binding patterns and structure-function specificities.

5.2. Materials and Methods

5.2.1. Protein and Reagents

Purified cellulase (endo-1,4-β-D-glucanase) of *A. niger* was purchased from Megazyme (Megazyme International Ireland Ltd., Wicklow, Ireland) and was used throughout all HDX experiments. Reference protein myoglobin was obtained from Sigma-Aldrich (St. Louis, MO). The protein solution was provided in 3.2 M ammonium sulphate with a specific activity of 86 U / mg against carboxymethyl cellulose (CMC) at 40°C and pH 4.5. The substrate of endoglucanase for this study was 0.5% CMC (average molecular weight, 250 000 Da, TCI AMERICA, Portland, OR). The inhibitor, D-(+)-cellobiose (molecular weight, MW: 342.30, Cat No: 22150) with >99.0% purity, was provided from sigma-Aldrich (St. Louis, MO). The second inhibitor of cellulase enzyme for this study was palladium chloride ($PdCl_2$, MW: 177.33, Cat No. 520659) was also purchased from Sigma-Aldrich.

5.2.2. HDX Experiments

For the Eg1A HDX-MS analysis, followed experimental design was similar to that of the structure dynamic analysis of XYN I from *T. longibrachitaum* [81]. Briefly, the *A. niger* Eg1A protein solution was directly used for HDX experiments, and the final concentration of Eg1A was 12.9 mg/ml. The substrate, CMC, was dissolved in a D_2O buffer (20mM Tris-HCL, 100mM KCL, and 1mM DTT in D_2O, pH 7.9) to make up a 0.5% solution. In previous studies, it has been revealed that addition of 100 μM $PdCl_2$ to

1μg/ml Eg1A inhibits more than 94% of enzyme activity against 1% CMC [143], and cellobiose has also defined as a potential inhibitor of cellulase [144]. Thus, the two inhibitors, cellobiose (competitive inhibitor) and $PdCl_2$ (uncompetitive inhibitor), were selected and also dissolved in the same D_2O buffer to obtain final concentrations of 146 and 100 mM, respectively. For the *apo* enzyme HDX reaction, 5 μl of the Eg1A solution was mixed with 15 μl of D_2O buffer and the same experimental design was also followed for the enzyme-ligand interactions (D_2O + ligand) at room temperature for 0, 1, 15, and 60 minutes, respectively. Next, the exchange reaction was terminated by adding 30 μl ice-cold quenching buffer (2 M urea and 1% trifluroacetic acid, TFA). The solution was then injected into a loading valve with 50 μl sample loop, passed through a pepsin column (Applied Biosystems, Foster City, CA) for 5 minutes. The digested Eg1A peptides were then eluted through a micro peptide cartridge (Michrom Bioresources, Inc., Auburn, CA) and desalted. The digested peptides were then eluted across a 2.1 mm × 5 cm C18 column (Thermo Scientific, Waltham, MA) with a linear gradient of 2%-50% Solution B (0.1% formic acid in 80% acetonitrile) over Solution A (0.1% formic acid in water) with a flow rate of 200 μL/min in 10 minutes. The eluted peptides were analyzed using LC-LTQ mass spectrometer (Thermo Scientific, Waltham, MA).

The global HDX experiment was carried out by monitoring the HDX of the intact enzyme with matrix assisted laser desorption ionization time of flight (MALDI- TOF) mass spectrometer (Shimadzu Scientific Instruments, Columbia, MD). Hydrogen exchange was initiated by adding 10-fold excess deuterated exchange buffer (20 mM

Tris-HCl, 100 mM KCL, and 1 mM DTT in D_2O, pH 7.9) to the original Eg1A stock solution (12.9 mg/mL). At a given exchange time, 1 μl of the exchange reaction was added into 9 μl of matrix solution, which has been kept on ice before addition of the protein solution. The matrix solution was prepared as the saturated solution of sinapinic acid in 45% acetonitrile/ water with 0.1 % of trifluroacetic acid. The reference protein myoglobin was added into the matrix solution in order to get an accurate measurement of the molecular weight of the intact protein. After the protein solution was introduced into the ice cold matrix solution, 2 μl of the solution was immediately placed onto a MALDI plate and dried under the gentle air flow. For a given time point, at least ten MALDI mass spectra were collected and the results were averaged.

5.2.3. Peptide Identification and HDX Data Processing

The identification of each peptide region was achieved by acquiring the product ion in a data-dependent MS/MS mode, and the precursor ion survey scan was performed and the five most abundant ions were selected for product ion analysis [81, 109]. The MS/MS raw data was converted into MS2 file and searched against the database with *A. niger* Eg1A using SEQUEST algorithm (Bioworks, Thermo Finnigan, CA). All peptide ion assignments were inspected and determined manually.

The weighted average m/z value were analyzed and calculated by the recently developed software, HDXanalyzer [80]. The deuteration level was calculated based on

the equation which was given in Chapter II and corrections for back-exchange were made based on 70% deuterium recovery and accounting for 80% deuterium content in the ion-exchange buffer.

5.2.4. Statistical Analysis for Differential HDX Data

As the Eg1A deuteration rates over time were slightly low, the HDX profiles during the longest exchange time (60 minutes) were chosen for the student *t*-test analysis to compare the *apo* and ligand-bound protein deuteration rates. Furthermore, the cumulative deuterium rates of six peptides from Eg1A-substrate/inhibitor were analyzed using one-factor ANOVA and Tukey's HSD procedures (

ID	residue	Peptide sequence	Treatment	$D_2O\% \pm SD^a$
2	24-37	DSASSPPYSVNQNL	Apo	13.8 ± 0.7 a
			CMC	6.7 ± 0.6 b
			Cellobiose	2.1 ± 0.9 c
			$PdCl_2$	3.1 ± 2.8 bc
			F-test	$^bF_{3,8} = 35.5$, $P < 0.01$
15	101-116	WKQDNTNVNADVAYDL	Apo	7.3 ± 1.8 a
			CMC	8.8 ± 2.3 a
			Cellobiose	1.7 ± 1.0 b
			$PdCl_2$	2.0 ± 1.5 b
			F-test	$F_{3,8} = 13.4$, $P < 0.01$
16	117-129	FTAANVDHATSSG	Apo	5.7 ± 1.4 b
			CMC	12.7 ± 2.4 a
			Cellobiose	2.6 ± 2.5 b
			$PdCl_2$	4.2 ± 1.9 b
			F-test	$F_{3,8} = 13.5$, $P < 0.01$
22	135-159	IWLARYGNIQPIGKQIATATVGGKS	Apo	16.1 ± 1.2 a
			CMC	12.6 ± 0.7 ab
			Cellobiose	10.9 ± 2.2 b

			$PdCl_2$	11.4 ± 2.2 b
			F-test	$F_{3,8} = 5.6$, $P = 0.02$
34	180-195	VSESPINSYSGDINAF	Apo	13.1 ± 0.9 a
			CMC	3.2 ± 0.9 b
			Cellobiose	7.9 ± 3.4 b
			$PdCl_2$	6.3 ± 1.3 b
			F-test	$F_{3,8} = 13.8$, $P < 0.01$
40	230-239	TVDNWTASVN	Apo	15.9 ± 3.3 a
			CMC	16.0 ± 1.3 a
			Cellobiose	4.6 ± 3.5 b
			$PdCl_2$	5.1 ± 2.1 b
			F-test	$F_{3,8} = 16.9$, $P < 0.01$

Table 8). All analyses were performed using DPS software [145].

ID	residue	Peptide sequence	Treatment	D_2O% ± SD[a]
2	24-37	DSASSPPYSVNQNL	Apo	13.8 ± 0.7 a
			CMC	6.7 ± 0.6 b
			Cellobiose	2.1 ± 0.9 c
			$PdCl_2$	3.1 ± 2.8 bc
			F-test	[b]$F_{3,8} = 35.5$, $P < 0.01$
15	101-116	WKQDNTNVNADVAYDL	Apo	7.3 ± 1.8 a
			CMC	8.8 ± 2.3 a
			Cellobiose	1.7 ± 1.0 b
			$PdCl_2$	2.0 ± 1.5 b
			F-test	$F_{3,8} = 13.4$, $P < 0.01$

16	117-129	FTAANVDHATSSG	Apo	5.7 ± 1.4 b
			CMC	12.7 ± 2.4 a
			Cellobiose	2.6 ± 2.5 b
			$PdCl_2$	4.2 ± 1.9 b
			F-test	$F_{3,8} = 13.5$, $P < 0.01$
22	135-159	IWLARYGNIQPIGKQIATATVGGKS	Apo	16.1 ± 1.2 a
			CMC	12.6 ± 0.7 ab
			Cellobiose	10.9 ± 2.2 b
			$PdCl_2$	11.4 ± 2.2 b
			F-test	$F_{3,8} = 5.6$, $P = 0.02$
34	180-195	VSESPINSYSGDINAF	Apo	13.1 ± 0.9 a
			CMC	3.2 ± 0.9 b
			Cellobiose	7.9 ± 3.4 b
			$PdCl_2$	6.3 ± 1.3 b
			F-test	$F_{3,8} = 13.8$, $P < 0.01$
40	230-239	TVDNWTASVN	Apo	15.9 ± 3.3 a
			CMC	16.0 ± 1.3 a
			Cellobiose	4.6 ± 3.5 b
			$PdCl_2$	5.1 ± 2.1 b
			F-test	$F_{3,8} = 16.9$, $P < 0.01$

Table 8. Differential deuteration rates of six notable peptides from Eg1A.

5.3. Results and Discussion

5.3.1. Structure Dynamics of Apo Eg1A

The structure dynamics of Eg1A from *A. niger* was probed both at the global and regional levels. The whole sequence of Eg1A protein is consisted of 239 amino acids, 159 of which represent the GH family 12 conserved domain and 223 of which are involved in the functional structure of Eg1A [143].From the MALDI-TOF measurement, it was observed that the Eg1A protein has a truncated version which starts Differential deuteration rates of six notable peptides from Eg1A from the residue 17th, from the N terminus, with a molecular weight of 24272 Da. In this study, a total of 40 digested

peptides were identified in the MSMS data acquisition (Table 9).

Table 9. The deuteration levels and statistical data of the specific peptides from Eg1A in response to various ligands

ID	Peptide	Residue & Charge	Apo				CMC				Cellobiose				PdCl2		
			D2O %	D2O %	T-test	p-value	D2O%	T-test	p-value	D2O %	T-test	p-value	D2O %	T-test	p-value
1	DSASSPPYSVNQNL	(8-21)1	8.93	5.49	1.4452	0.2219	2.1	3.0829	0.0368	1.85	3.4088	0.0271
2	DSASSPPYSVNQNL	(8-21)2	13.84	6.68	12.594	0.0002	2.09	17.7513	0.0001	3.15	6.4143	0.003
3	YVDKLSSSGASW	(34-45)1	6.55	4.18	1.1714	0.3065	6.94	0.152	0.8865	11.04	1.1717	0.3063
4	YVDKLSSSGASW	(34-45)2	8.23	4.5	1.2494	0.2796	5.48	0.4474	0.6777	5.48	0.4474	0.6777
5	HTEWTWSGGEGTVKS	(46-60)2	9.09	7.21	0.3625	0.7375	2.71	1.1932	0.2987	3.92	1.001	0.4222
6	HTEWTWSGGEGTVKSYSNSGVTF	(46-68)2	5.16	6.9	1.481	0.2127	4.07	0.7449	0.4977	4.03	1.682	0.1679
7	HTEWTWSGGEGTVKSYSNSGVTFNKKLVSDVSSIPTSVE	(46-84)3	6.17	7.26	0.5959	0.5833	4.8	0.667	0.4129	7.49	0.8257	0.4554
8	NKKLVSDVSSIPTSVE	(69-84)2	3.87	5.02	0.7101	0.5169	2.62	0.913	0.4129	3.95	0.0477	0.9643
9	NKKLVSDVSSIPTSVEWKQDNTNVNA	(69-94)3	3.18	4.83	2.2879	0.0841	1.89	0.7914	0.473	3.97	0.6453	0.5539
10	NKKLVSDVSSIPTSVEWKQDNTNVNAD	(69-95)3	4.96				5.19	0.1913	0.8576	4.84	0.2015	0.8501
11	VSDVSSIPTSVEWKQDNTNVNA	(76-94)2	7.47	8.55	0.7717	0.4833	4.41	1.9478	0.1233	7.34	0.0964	0.932
12	EWKQDNTNVNA	(86-94)2	6.44	6.27	0.057	0.9573	3.07	1.455	0.2194	5.25	0.6203	0.5687
13	WKQDNTNVNA	(87-94)2	8.26	9.2	0.6239	0.5665	3.43	2.5427	0.0638	7.77	0.4775	0.6579
14	WKQDNTNVNAD	(87-95)2	8.37	4.82	1.485	0.2117	4.05	1.6698	0.1703	5.84	1.1286	0.3222
15	WKQDNTNVNADVAYDL	(85-100)2	7.29	8.79	0.89	0.4238	1.71	4.5626	0.0103	2.21	4.0205	0.0159
16	FTAANVDHATSSG	(101-113)2	5.7	12.71	4.3403	0.0122	2.56	1.8792	0.1334	4.19	1.1049	0.3312
17	FTAANVDHATSSGD	(101-114)1	3.84	5.96	0.3822	0.7218	3.89	0.008	0.9944	5.44	0.0752	0.9437
18	FTAANVDHATSSGD	(101-114)2	11.71	10.61	0.3674	0.732	4.96	2.0205	0.1366	7.1	1.5508	0.1959
19	FTAANVDHATSSGDYE	(101-116)2	7.52	10.08	1.0703	0.3448	2.34	2.2526	0.0874	10.46	1.1029	0.332
20	FTAANVDHATSSGDYEL	(101-117)2	11.66	7.25	0.504	0.6643	6.28	0.6122	0.6027	8.99	0.2995	0.7795
21	MIWLARYGNIQPIGKQIATATVGGKS	(118-143)3	12.43	8.89	2.1345	0.0997	8.58	2.2369	0.0899	12.15	0.2618	0.8064
22	IWLARYGNIQPIGKQIATATVGGKS	(119-143)2	16.11	12.6	4.3166	0.0125	10.9	3.604	0.0227	12.84	3.1316	0.0351
23	IWLARYGNIQPIGKQIATATVGGKS	(119-143)3	14.9	12.55	2.2011	0.0925	10.6	2.2807	0.0847	11.06	2.2837	0.0844

Table 9. Continued

ID	Peptide	Residue & Charge	Apo	CMC			Cellobiose			PdCl2		
			D2O %	D2O %	T-test	p-value	D2O%	T-test	p-value	D2O %	T-test	p-value
24	LARYGNIQPIGKQIATATVGGKS	(121-143)2	12.22	14.79	1.3992	0.2343	11.9	0.1423	0.8937	13.4	1.4459	0.2217
25	LARYGNIQPIGKQIATATVGGKS	(121-143)3	17.32	16.45	0.7564	0.4915	13.4	2.3255	0.0807	15.32	1.4625	0.2174
26	ARYGNIQPIGKQIATATVGGKS	(122-143)2	17.89	17.85	0.0141	0.9894	14.3	1.6389	0.1766	15.28	1.2451	0.2811
27	ARYGNIQPIGKQIATATVGGKS	(122-143)3	18.22	17.01	1.2607	0.276	13.1	2.1558	0.0973	14.24	2.9297	0.0428
28	WEVWYGSTTQAGAEQRT	(144-160)2	7.32	6.89	1.1814	0.3029	7.14	0.0786	0.9412	7.44	0.8146	0.461
29	WEVWYGSTTQAGAEQRTYSF	(144-163)2	6.37	8.72	0.9555	0.3934	5.76	0.3013	0.7782	5.55	0.6507	0.5507
30	YGSTTQAGAEQRTYSF	(148-163)2	6.9	6.66	0.1033	0.9271	5.23	0.6188	0.5695	7.26	0.1398	0.8956
31	VSESPINSYSGDINA	(164-194)1	10.58				7.95	1.583	0.1886	6.93	2.1989	0.0928
32	VSESPINSYSGDINA	(164-178)2	10.87	8	10.803	0.0004	9.31	1.1724	0.3618	5.54	3.7231	0.0652
33	VSESPINSYSGDINAF	(164-179)1										
34	VSESPINSYSGDINAF	(164-179)2	13.1	7.02	2.8396	0.0469	7.94	2.581	0.0613	7.3	10.9996	0.0004
35	FSYLTQNQGFPASSQY	(180-195)2	5.78	6.57	0.4158	0.6989	3.31	1.8706	0.1347	6.1	0.2364	0.8248
36	LTQNQGFPASSQY	(183-195)2	7.24	5.03	1.6898	0.2331	9.72	1.5528	0.2607	6.96	0.1598	0.8877
37	LINLQFGTE	(196-204)1	4.97	6.74	0.6434	0.555	0.58	1.3947	0.2356	6.54	0.2965	0.7816
38	QFGTEAFTGGPATF	(200-213)1	3.62	6.2	1.6127	0.1821	1.77	1.5228	0.2025	3.23	0.2875	0.788
39	QFGTEAFTGGPATF	(200-213)2										
40	TVDNWTASVN	(214-223)2	15.92	15.95	0.0162	0.9879	4.63	4.1043	0.0148	5.08	4.8254	0.0085

A sequence coverage of 85.4% for the whole enzyme and 100% coverage for conserved GH family 12 domain have been identified (Figure 24). After 60 minutes exchange with D_2O, the HDX analysis revealed a minimum exchange rate of the enzyme when it was incubated with deuterium solvent. Most of the peptide regions exhibited less than 10% exchange as compared to the full exchange capacity of the corresponding peptide. Only six peptide regions were detected to have more than 10% but less than 20% of deuteration exchange rate in the HDX experiment.

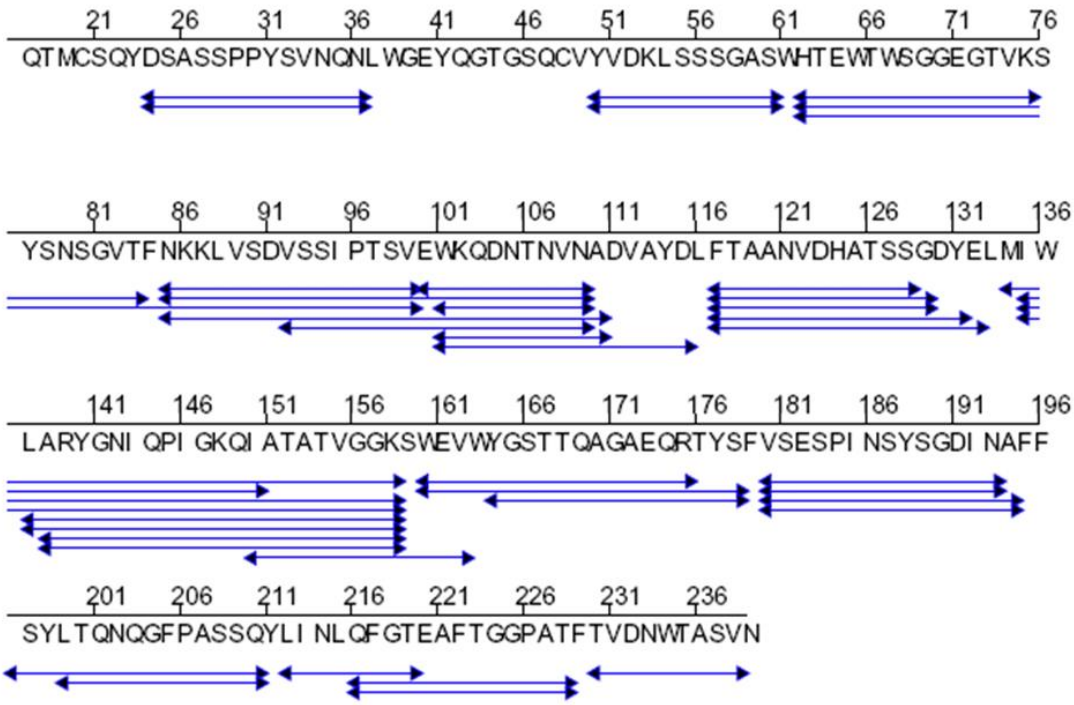

Figure 24. Peptides analyzed in the HDX experiment for Eg1A. The blue arrows donate the sequence of the peptides analyzed successfully.

Careful examination of the Eg1A crystal structure revealed that the enzyme has a "jelly-roll" folding with two antiparallel β-sheets, which presents as a common folding motif in endoglucanase (Figure 25). The approximate dimensions of the Eg1A catalytic core is about 40 X 40 X 35 A and consists of a single-domain polypeptide chain (GH12) [143]. The hydrophilic surface of sheet B forms a long open cleft which is the binding site for substrate. The β-sheet A contains six strands (A1- A6) and β-sheet B contains nine strands (B1-B9). B1, B2, the loop region between B5 and B6 sheets, the loop between B6 and B9, B8, part of the helix region, and the A4 region in the C terminus are the regions identified with HDX as slightly dynamic regions of Eg1A. It was suggested that the hydrophilic face of the β-sheet B forms along the open cleft, which harbors the binding site of the substrate. The HDX-MS analysis findings for Eg1A are consistent with the specific protein-ligand interaction sites, where the most dynamic regions identified are located on the β-sheet regions.

5.3.2. Active Site upon Ligand Binding Revealed by HDX

When the *apo* enzyme was treated with solvent deuterium under specific conditions, the HDX-MS data of Eg1A indicated that the several regions of the protein reveal higher level of structural motions than the other regions (Figure 25). In order to examine the dynamic behavior of the enzyme to the different substrate or inhibitors, carboxymethyl cellulose (CMC), cellobiose, and $PdCl_2$ were interacted with *apo* protein

through in a defined time scale. CMC is one of the main substrates of cellulases. Cellobiose and $PdCl_2$ are the two inhibitors of Eg1A. The major regions whose dynamics have been changed upon ligand binding were shown in Figure 25 by overlaying

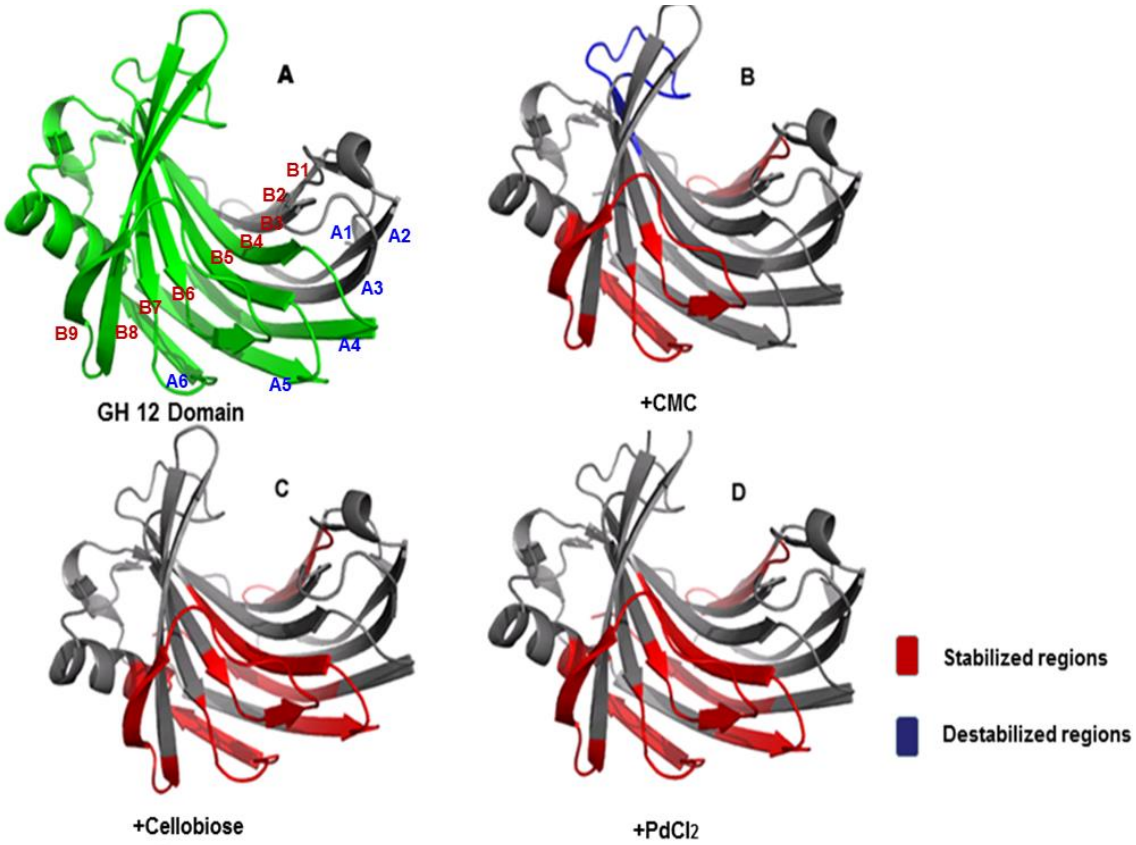

Figure 25. The differential dynamic characteristics of Eg1A upon binding with various ligands. A) Apo *A. niger* Eg1Acrystal structure (PDB: 1KS4) to show the GH12 domain region (green); B) The HDX profile of Eg1A/CMC interaction overlaid onto the *apo* Eg1A 3D structure; C) The HDX delineation of Eg1A/cellobiose binding overlaid onto *apo* Eg1A 3D structure; D) The HDX report of Eg1A/cellobiose interaction overlaid onto *apo* Eg1A crystal structure. The stabilized (red) and destabilized (blue) regions identified through HDX analysis are indicated.

the HDX change data onto the protein X-ray structure. The two B1 and B2 regions, the loop regions before the helix, and part of the helix region have been shown to be stabilized binding with both CMC and two inhibitors. Part of A5 and B5, the loop region between A5 and B5, the A4 region in β-sheet A, however, showed stabilization effects when binding with the inhibitors, but displayed no change through interacting with CMC. Interestingly, the loop region between B5 and B6 was destabilized through interacting with CMC as revealed by HDX. Conversely, the same region showed no dynamic difference when binding with both cellobiose and $PdCl_2$ inhibitor (Figure 25). The amino acid residues located in this loop are highly conserved (Phe117, Thr118, Ala119, Ala120, His124, Thr126, Ser128, and Gly129). This loop is an important region within the active site cleft, and the destabilization effect observed in HDX might be explained as a different binding mode of this region upon interaction with substrate versus inhibitors. Another important region referred to as the "cord" region in the both XYN I and Eg1A [79, 81], is the surface loop region between strands B6 and B9. Amino acids in this region are also highly conserved. In the HDX experiment, this region is slightly protected when binding with both the substrate and the inhibitors. The peptide "IWLARYGNIQPIGKQIATATVGGKS" (residue 135-159) showed 16% of exchange rate after 60 minutes incubation with solvent deuterium, which belongs to one of the regions that are most dynamic in the HDX experiment. The same peptide showed 11% of exchange rate when incubated with inhibitors. The X-ray structure showed that when Pd^{2+} ions are binding with Eg1A, one ion is residing in the binding cleft and close to the cord loop [143]. That Pd^{2+} forms coordinate covalent bonds with Met134, and Glu132. It

has been suggested that the loop may have conformational changes upon substrate binding and facilitate the interaction of the two conserved tryptophan residues (Phe136 and Phe163) with the substrate. The HDX analysis results are consistent with the X-ray data, indicating the differential dynamics behavior of the enzyme when binding with various ligands.

5.3.3. Comparison the Dynamic Motions of XYN I and Eg1A

Even though the sequence similarity between Eg1A and XYN I is extremely low, structure analyses suggest that they share overall the same folding [143, 146, 147]. Hemicellulolytic XYN I protein from *T. longibrachiatum* exhibits only 16% sequence identity with Eg1A analyzed in this work. Both XYN I and Eg1A, however, share well known "right hand" structure including two β-strands and one α-helix to form the palm. However, in the Eg1A structure, the strands are more expanded with one more strand in β-sheet A. One notable structure feature about Eg1A is that cysteine 20 and cysteine 48 forms a disulfide bond by connecting strand A1 and A2, which might contribute to the less dynamic nature of Eg1A. However, the presence of one disulfide bond connecting two β-sheet strands might impact the dynamics locally and the contribution of such single disulfide bond cannot be inclusive. Since no denaturant was used during the HDX experiments, the measurement of the backbone hydrogens only probes the protein under native conditions. The exchange rates of the backbone

hydrogens are highly dependent on the three dimensions of the protein structure, which involves H bonding, protein distortional motion, and solvent penetration etc. The similarity of the 3D structures provides a good model to investigate the comparative structure dynamics.

Interestingly, a big dynamic difference was observed for the two structurally similar protein folds. In the previous study, XYN I showed much higher exchange dynamics when incubated with solvent deuterium. Through a 64 minutes exchange with D_2O, most peptides representing various regions of the XYN I enzyme experienced more than 50% of total exchange; some regions further displayed over 90% exchange rate with deuterium as compared to Eg1A protein (Figure 26) [81].

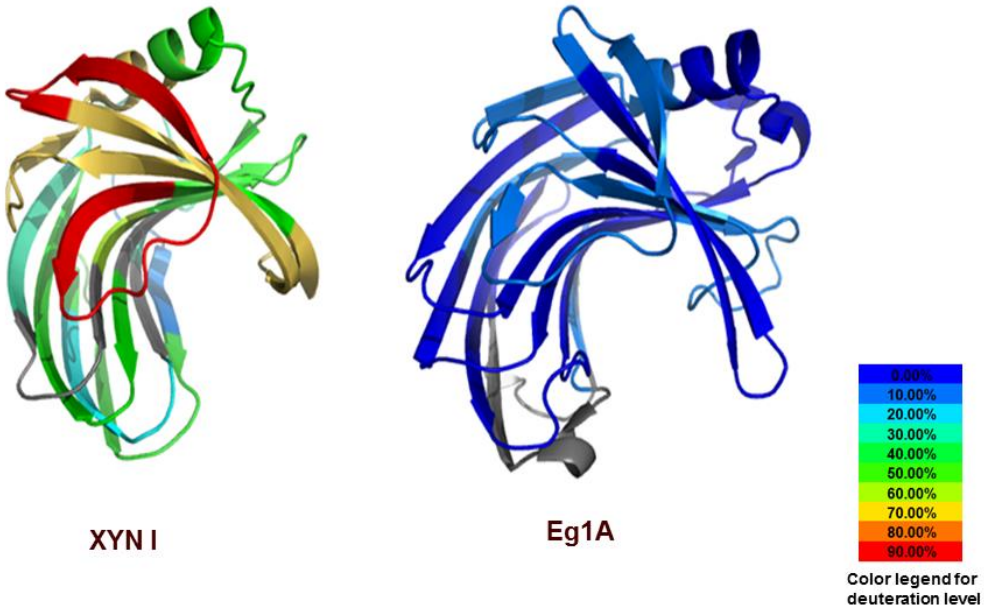

Figure 26. Regional deuteration rates of both XYN I and Eg1A through HDX-MS analysis.

In order to directly compare the two enzymes, the global hydrogen deuterium exchange rates were probed by measuring the mass increases of the intact proteins when they treated with D_2O (Figure 27).

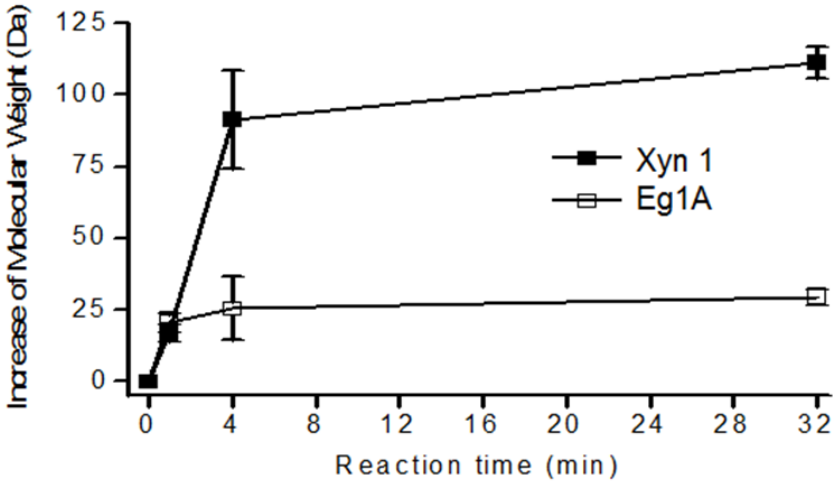

Figure 27. Globular dynamics of both XYN I and Eg1A revealed by MALDI-TOF analysis.

The MALDI-TOF measurements revealed that the molecular weight of the XYN I is 19047 Da, given that the enzyme is composed of 178 amino acids. This is also a truncated version of the enzyme, which starts from the amino acid fifty second (Ala52). The exchange rates of both Eg1A and XYN I intact proteins through the 32 minutes incubation with D_2O resulted in a mass increase of 39 Da for Eg1A and 60 Da for XYN I.

Given the whole sequence of each enzyme used in the experiments, Eg1A is composed of 223 amino acids and XYN I only contains 178 amino acids (Figure 28). Furthermore, the MALDI-TOF analysis revealed that the molecular weight of XYN I protein starts increasing dramatically after the first four minutes incubation with D_2O as compared to Eg1A, suggesting the existence of a very rigid tertiary structure for Eg1A protein.

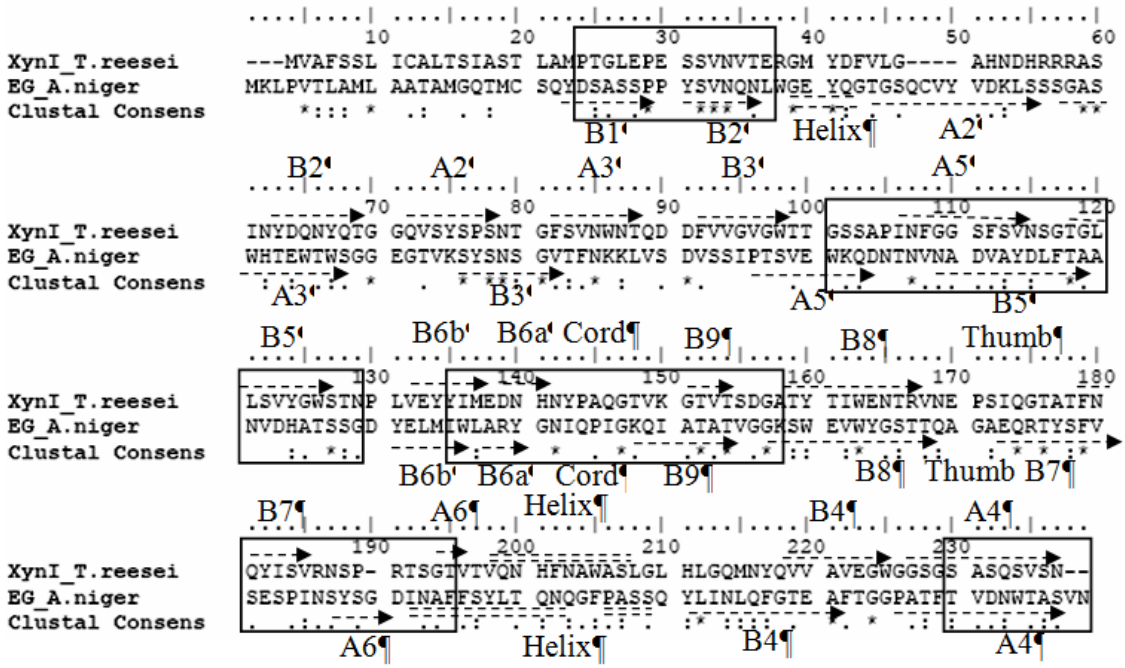

Figure 28. The pairwise sequence alignment of *A. niger* Eg1A and *T. reesei* XYN I together with secondary structure assignment. β-strands are marked with arrows; and α-helices with double dashed lines. The "cord" and "thumb" regions are also indicated.

Eg1A has more backbone hydrogen available for exchange if global unfolding can be obtained. It was suggested that the HDX data of the *apo* protein with solvent might involve segmental unfolding reactions, even though the global unfolding is not reached [143].

In addition, the root mean square deviation (RMSD) analysis confirmed the highly dynamic characteristic of XYN I compared to Eg1A (Figure 29). The RMSD is regarded as an important indicator of the structure flexibilities and dynamics of proteins. It is frequently used for the comparison of similar structures through the calculations of the average distances between the molecular atoms.

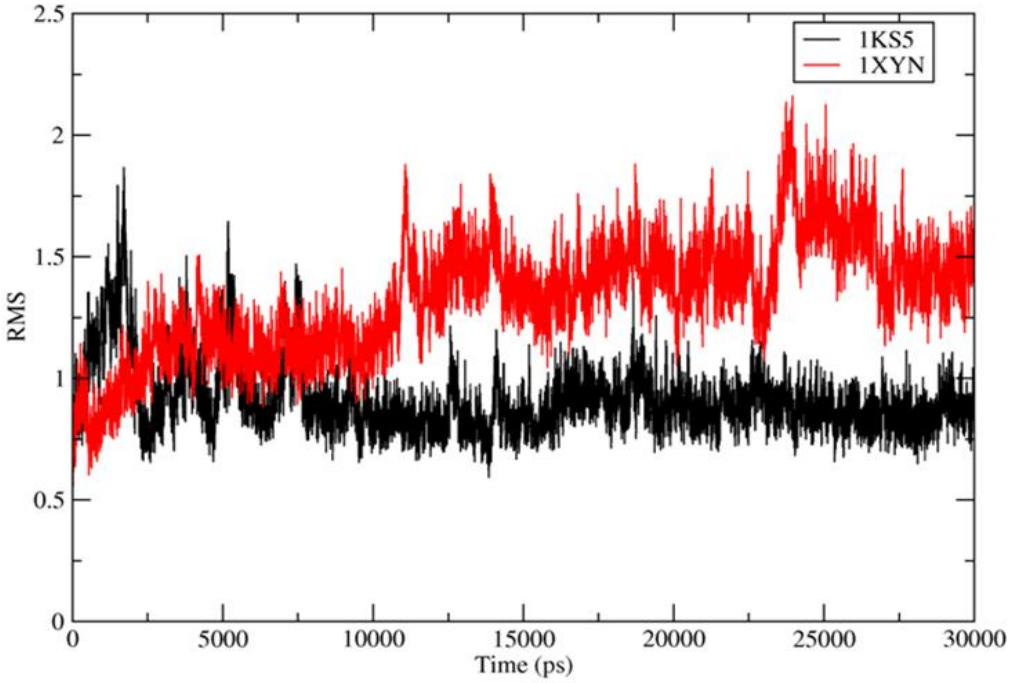

Figure 29. RMSD-based structure dynamics comparison of both XYN I and Eg1A. The XYN I protein starts changing in 10 ns in terms of the structural flexibility, the

Eg1A protein however exhibits highly rigid structural patterns throughout the analysis

The study unveiled potential new structure determinants for Eg1A enzyme function. The structure dynamics of Eg1A was compared to the previously studied XYN I and highlighted dramatic differences in enzyme structure dynamics, even though the 3D structure of the two enzymes are similar. The HDX rates were compared both at global and regional levels for the two enzymes, and the global dynamics data confirmed the regional dynamics observations. Molecular dynamics of the two enzymes were also compared at the nanosecond scale. The comparison of nano second scale dynamics also pointed to the same conclusion that XYN1 is more dynamic than Eg1A.The structural flexibilities revealed by RMSD further confirmed the more dynamic personality of XYN I protein compared to Eg1A. XYN I was shown to be highly dynamic in the *apo* form when in the solution phase exchanging with solvent deuterium. Binding with substrate protected XYN I for a variety of peptide regions. However, the *apo* Eg1A showed very limited exchange with solvent deuterium. Binding with substrate introduced protection into Eg1A, but such protection is significantly less as compared to XYN I.

Further overlay of the current structure dynamics data of Eg1A onto a multiple sequence alignment of the Eg1A proteins from different fungal organisms pointed out that the highly conserved surface loop region (the cord region; the residues between 138 and 160) may be crucial for differential dynamics (Figure 30). Overall, the current results highlight that enzyme dynamics can be a driving force for the differential functionality and biocatalyst engineering applications.

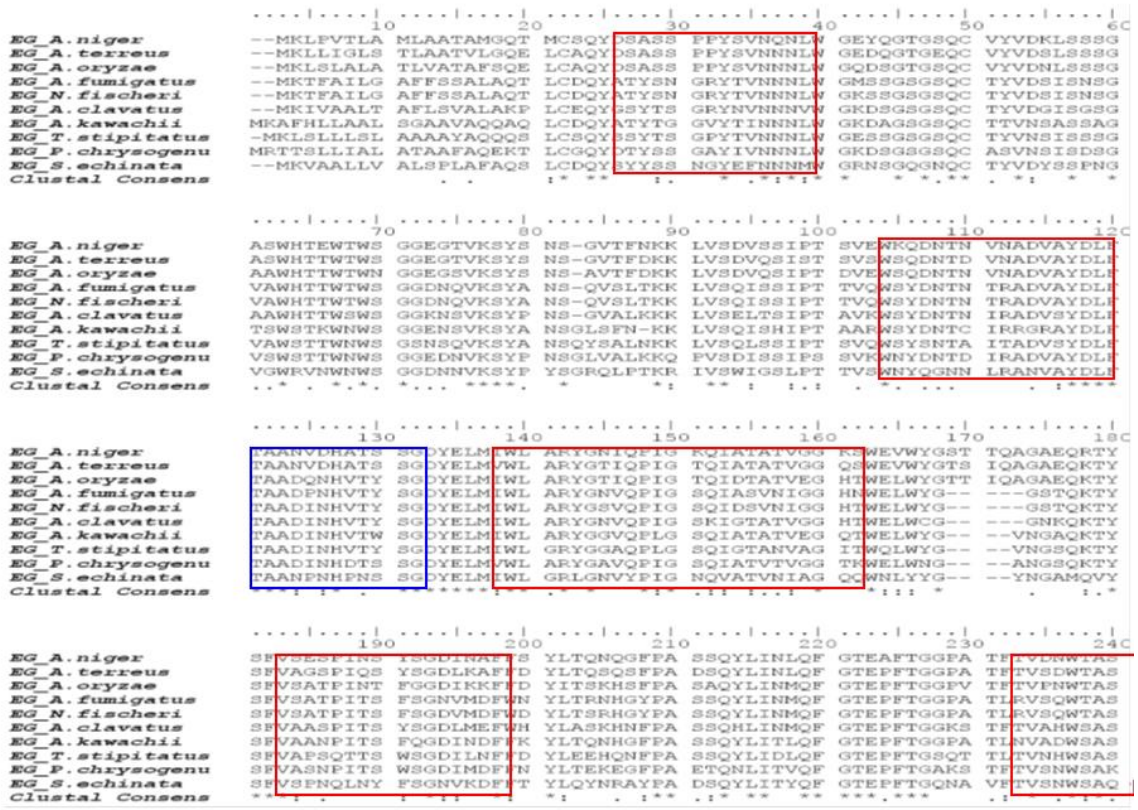

Figure 30. Multiple sequence alignment for Eg1A indicating the highly stabilized (red boxes) and destabilized regions (blue boxes) in the protein sequence through HDX analysis.

CHAPTER VI

CONCLUSIONS AND PERSEPCTIVES

Studying the structure dynamics of proteins is of great interest regarding its potential role in structure-function relationships. Among various structure analysis platforms, HDX-MS stands out as an efficient way to analyze protein dynamics upon ligand binding. Demonstrated here is the systematic use of HDX-MS with model biocatalysts to reveal the application of HDX-MS toward structure-function relationships for bioenergy research.

The HDXanalyzer software, developed to quantify the HDX-MS analyses of related data, enables rapid and statistical analyses of protein structure dynamics, which can further assist more accurate interpretation of HDX-MS data.

For the XYN1 model system, high intrinsic dynamics has been identified. Interaction with different substrates renders the differential stabilization of the enzyme structure. The thumb, cord regions, and the binding groove were detected as the key dynamic regions of XYN I which exhibit differential dynamic patterns through binding with specific substrates. Further modification of the enzyme based on the HDX-MS data leads to successful enzyme engineering with improved enzyme activity. It would be expected that the findings through HDX-MS analysis can translate into more protein engineering work for better enzymes.

The structure dynamics comparison between the two similar protein folds revealed different dynamics in the HDX-MS analysis. In the model system here, even

though the two enzymes, *T. reesei* XYN I and *A. niger* Eg1A, are efficient catalysts for their substrates, the enzymes showed distinct enzyme dynamics, indicating that enzyme dynamics do not always underlie catalysis. Molecular dynamics modeling also conforms to the HDX analysis results, further questioning the current understanding of the correlation between catalysis and protein structure dynamics. Future applications of the enzyme dynamics analysis with more model systems will help better elucidate structure-function relationships with a broader perspective.

REFERENCES

1. Houseweart MK, Cleveland DW: **Intermediate filaments and their associated proteins: multiple dynamic personalities**. *Current opinion in cell biology* 1998, **10**(1):93-101.

2. McGowan LC, Hamelberg D: **Conformational plasticity of an enzyme during catalysis: Intricate coupling between cyclophilin A dynamics and substrate turnover**. *Biophys J* 2013, **104**(1):216-226.

3. Fieulaine S, Boularot A, Artaud I, Desmadril M, Dardel F, Meinnel T, Giglione C: **Trapping conformational states along ligand-binding dynamics of peptide deformylase: the impact of induced fit on enzyme catalysis**. *PLoS biology* 2011, **9**(5):e1001066.

4. Ma B, Nussinov R: **Enzyme dynamics point to stepwise conformational selection in catalysis**. *Curr Opin Chem Biol* 2010, **14**(5):652-659.

5. Warshel A: **Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites**. *Journal of Biological Chemistry* 1998, **273**(42):27035-27038.

6. Wolfenden R, Snider MJ: **The depth of chemical time and the power of enzymes as catalysts**. *Accounts of Chemical Research* 2001, **34**(12):938-945.

7. Min W, Xie XS, Bagchi B: **Two-dimensional reaction free energy surfaces of catalytic reaction: effects of protein conformational dynamics on enzyme catalysis**. *J Phys Chem B* 2008, **112**(2):454-466.

8. Liu YH, Konermann L: **Enzyme conformational dynamics during catalysis and in the 'resting state' monitored by hydrogen/deuterium exchange mass spectrometry**. *Febs Lett* 2006, **580**(22):5137-5142.

9. Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev DM, Wolf-Watz M, Bosco DA, Skalicky JJ, Kay LE, Kern D: **Intrinsic dynamics of an enzyme underlies catalysis**. *Nature* 2005, **438**(7064):117-121.

10. Bernado PB, Martin: **Structural biology: Proteins in dynamic equilibrium**. *Nature* 2010, **468**(7327):1046-1048.

11. Henzler-Wildman KK, Dorothee: **Dynamic personalities of proteins**. *Nature* 2007, **450**(7172):964-972.

12. Grossman M, Sela-Passwell N, Sagi I: **Achieving broad molecular insights into dynamic protein interactions by integrated structural-kinetic approaches**. *Current Opinion in Structural Biology* 2011, **21**(5):678-685.

13. Fischer E: **Einfluss der configuration auf die wirkung der enzyme**. *Berichte der deutschen chemischen Gesellschaft* 1894, **27**(3):2985-2993.

14. Koshland DE: **Application of a theory of enzyme specificity to protein synthesis**. *Proc Nat Acad Sci USA* 1958, **44**(2):98-104.

15. Frauenfelder H SS, and Wolynes PG: **The energy landscapes and motions of proteins**. *Science* 1991, **254**(5038):1598-1603.

16. Boehr DD, Nussinov, Ruth, Wright, Peter E: **The role of dynamic conformational ensembles in biomolecular recognition**. *Nature Chemical Biology* 2009, **5**(12):789-796.

17. Damborsky J, Brezovsky J: **Computational tools for designing and engineering biocatalysts**. *Current Opinion in Chemical Biology* 2009, **13**(1):26-34.

18. Illanes A, Cauerhff A, Wilson L, Castro GR: **Recent trends in biocatalysis engineering**. *Bioresource technology* 2012, **115**(0):48-57.

19. Kaul P AY: **Strategies for discovery and improvement of enzyme function: state of art and opportunities**. *Microbial Biotechnology* 2012, **5**(1):18-33.

20. Dalby PA: **Strategy and success for the directed evolution of enzymes**. *Current Opinion in Structural Biology* 2011, **21**(4):473-480.

21. Wang M, Si T, Zhao H: **Biocatalyst development by directed evolution**. *Bioresource technology* 2012, **115**(0):117-125.

22. Eijsink VGH, Bjork A, Gaseidnes S, Sirevag R, Synstad B, van den Burg B, Vriend G: **Rational engineering of enzyme stability**. *Journal of biotechnology* 2004, **113**(1-3):105-120.

23. Schulein M: **Protein engineering of cellulases**. *Bba-Protein Struct M* 2000, **1543**(2):239-252.

24. Zhang YHP, Himmel ME, Mielenz JR: **Outlook for cellulase improvement: screening and selection strategies**. *Biotechnology Advances* 2006, **24**(5):452-481.

25. Kraulis PJ, Clore GM, Nilges M, Jones TA, Pettersson G, Knowles J, Gronenborn AM: **Determination of the 3-dimensional solution structure of the C-terminal domain of cellobiohydrolase-I from *Trichoderma reesei* : A study using nuclear magnetic-resonance and hybrid distance geometry dynamical simulated annealing**. *Biochemistry* 1989, **28**(18):7241-7257.

26. Rouvinen J, Bergfors T, Teeri T, Knowles JKC, Jones TA: **3-Dimensional structure of cellobiohydrolase-II from *Trichoderma reesei***. *Science* 1990, **249**(4967):380-386.

27. Estácio SG: **Chapter 9 - In silico strategies toward enzyme function and dynamics**. In: *Advances in Protein Chemistry and Structural Biology.* Edited by Christo C, Tatyana K-C, vol. Volume 87: Academic Press; 2012: 249-292.

28. Tenkanen M, Vršanská M, Siika-aho M, Wong DW, Puchart V, Penttilä M, Saloheimo M, Biely P: **Xylanase XYN IV from *Trichoderma reesei* showing exo- and endo-xylanase activity**. *Febs J* 2013, **280**(1):285-301.

29. Yu X-B, Nam J-H, Yun H, Koo Y-M: **Optimization of cellulase production in batch fermentation by *Trichoderma reesei***. *Biotechnol Bioprocess Eng* 1998, **3**(1):44-47.

30. Rossi-Rodrigues BC, Brochetto-Braga MR, Tauk-Tornisielo SM, Carmona EC, Arruda VM, Chaud Netto J: **Comparative growth of Trichoderma strains in different nutritional sources, using bioscreen C-automated system**. *Brazilian Journal of Microbiology* 2009, **40**:404-410.

31. Kar S, Mandal A, Mohapatra PK, Samanta S, Pati BR, Mondal KC: **Production of xylanase by immobilized *Trichoderma reesei* SAF3 in Ca-alginate beads**. *Journal of industrial microbiology & biotechnology* 2008, **35**(4):245-249.

32. Kar S, Sona Gauri S, Das A, Jana A, Maity C, Mandal A, Das Mohapatra P, Pati B, Mondal K: **Process optimization of xylanase production using cheap solid substrate by *Trichoderma reesei* SAF3 and study on the alteration of behavioral properties of enzyme obtained from SSF and SmF**. *Bioproc Biosyst Eng* 2013, **36**(1):57-68.

33. Xiong H, Turunen O, Pastinen O, Leisola M, von Weymarn N: **Improved xylanase production by *Trichoderma reesei* grown on L-arabinose and lactose or D-glucose mixtures**. *Appl Microbiol Biotech* 2004, **64**(3):353-358.

34. Paloheimo M, Mäntylä A, Kallio J, Puranen T, Suominen P: **Increased production of xylanase by expression of a truncated version of the xyn11A gene from *Nonomuraea flexuosa* in *Trichoderma reesei***. *Applied and environmental microbiology* 2007, **73**(10):3215-3224.

35. Zou G SS, Jiang Y, Brink JVD, Vries RP, Chen L, Zhang J, Ma L, Wang C, and Zhou Z: **Construction of a cellulase hyper-expression system in *Trichoderma reesei* by promoter and enzyme engineering**. *Microb Cell Fact* 2012, **11**(21):1-21.

36. Te'o VSJ, Cziferszky AE, Bergquist PL, Nevalainen KMH: **Codon optimization of xylanase gene xynB from the thermophilic bacterium *Dictyoglomus thermophilum* for expression in the filamentous fungus *Trichoderma reesei***. *FEMS microbiology letters* 2000, **190**(1):13-19.

37. Benko Z, Drahos E, Szengyel Z, Puranen T, Vehmaanperä J, Réczey K: ***Thermoascus aurantiacus* CBHI/Cel7A production in *Trichoderma reesei* on alternative carbon sources**. *Appl Biochem Biotech* 2007, **137-140**(1-12):195-204.

38. de Faria FP, Te OV, Bergquist PL, Azevedo MO, Nevalainen KM: **Expression and processing of a major xylanase (XYN2) from the thermophilic fungus *Humicola grisea* var. *thermoidea* in *Trichoderma reesei***. *Letters in applied microbiology* 2002, **34**(2):119-123.

39. Geng A, Zou G, Yan X, Wang Q, Zhang J, Liu F, Zhu B, Zhou Z: **Expression and characterization of a novel metagenome-derived cellulase Exo2b and its application to improve cellulase activity in *Trichoderma reesei***. *Applied microbiology and biotechnology* 2012, **96**(4):951-962.

40. Vikramathithan J, Ravikumar S, Muthuraman P, Nirmalkumar G, Shayamala S, Srikumar K: **Purification and biochemical characterization of two major thermophilic xylanase isoforms (T-70 and T-90) from xerophytic *Opuntia vulgaris* plant spp.** *Cellulose* 2012, **19**(4):1373-1383.

41. Wang ZY, Jin Y, Wu HJ, Tian ZF, Wu YY, Xie XM: **A novel, alkali-tolerant thermostable xylanase from *Saccharomonospora viridis*: direct gene cloning, expression and enzyme characterization**. *World J Microb Biot* 2012, **28**(8):2741-2748.

42. El-Bondkly AMA: **Molecular identification using ITS sequences and genome shuffling to improve 2-deoxyglucose tolerance and xylanase activity of marine-derived fungus, *Aspergillus sp.* NRCF5**. *Appl Biochem Biotech* 2012, **167**(8):2160-2173.

43. Cui FJ, Zhao LM: **Optimization of xylanase production from *Penicillium sp*.WX-Z1 by a two-step statistical strategy: Plackett-Burman and Box-Behnken experimental design**. *Int J Mol Sci* 2012, **13**(8):10630-10646.

44. Masui DC, Zimbardi ALRL, Souza FHM, Guimaraes LHS, Furriel RPM, Jorge JA: **Production of a xylose-stimulated beta-glucosidase and a cellulase-free thermostable xylanase by the thermophilic fungus *Humicola brevis* var. *thermoidea* under solid state fermentation**. *World J Microb Biot* 2012, **28**(8):2689-2701.

45. Zhang F, Chen JJ, Ren WZ, Lin LB, Zhou Y, Zhi XY, Tang SK, Li WJ: **Cloning, expression, and characterization of an alkaline thermostable GH11 xylanase from *Thermobifida halotolerans* YIM 90462(T)**. *Journal of industrial microbiology & biotechnology* 2012, **39**(8):1109-1116.

46. Davies G, Henrissat B: **Structures and mechanisms of glycosyl hydrolases**. *Structure* 1995, **3**(9):853-859.

47. Kleywegt GJ, Zou JY, Divne C, Davies GJ, Sinning I, Stahlberg J, Reinikainen T, Srisodsuk M, Teeri TT, Jones TA: **The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 angstrom resolution, and a comparison with related enzymes**. *J Mol Biol* 1997, **272**(3):383-397.

48. Divne C, Stahlberg J, Reinikainen T, Ruohonen L, Pettersson G, Knowles JKC, Teeri TT, Jones TA: **The 3-dimensional crystal structure of the catalytic core of cellobiohydrolase-I from *Trichoderma reesei***. *Science* 1994, **265**(5171):524-528.

49. Dominguez R, Souchon H, Lascombe MB, Alzari PM: **The crystal structure of a family 5 endoglucanase mutant in complexed and uncomplexed forms reveals an induced fit activation mechanism**. *J Mol Biol* 1996, **257**(5):1042-1051.

50. Torronen A, Rouvinen J: **Structural comparison of two major endo-1,4-xylanases from *Trichoderma reesei***. *Biochemistry* 1995, **34**(3):847-856.

51. Zhang S, Wilson DB: **Surface residue mutations which change the substrate specificity of *Thermomonospora fusca* endoglucanase E2**. *Journal of biotechnology* 1997, **57**(1-3):101-113.

52. Zhang S, Barr BK, Wilson DB: **Effects of noncatalytic residue mutations on substrate specificity and ligand binding of *Thermobifida fusca* endocellulase Cel6A**. *Eur J Biochem* 2000, **267**(1):244-252.

53. Zhang S, Irwin DC, Wilson DB: **Site-directed mutation of noncatalytic residues of *Thermobifida fusca* exocellulase Cel6B**. *Eur J Biochem* 2000, **267**(11):3101-3115.

54. Baker JO, McCarley JR, Lovettt R, Yu CH, Adney WS, Rignall TR, Vinzant TB, Decker SR, Sakon J, Himmel ME: **Catalytically enhanced endocellulase Cel5A from *Acidothermus cellulolyticus***. *Applied biochemistry and biotechnology* 2005, **121**:129-148.

55. Escovar-Kousen JM, Wilson D, Irwin D: **Integration of computer modeling and initial studies of site-directed mutagenesis to improve cellulase activity on Cel9A from *Thermobifida fusca***. *Applied biochemistry and biotechnology* 2004, **113-16**:287-297.

56. Druzhinina IS, Komoń-Zelazowska M, Ismaiel A, Jaklitsch W, Mullaw T, Samuels GJ, Kubicek CP: **Molecular phylogeny and species delimitation in the section *Longibrachiatum* of Trichoderma**. *Fungal Genetics and Biology* 2012, **49**(5):358-368.

57. Chaverri P SG: **Hypocrea/Trichoderma (Ascomycota, Hypocreales, *Hypocreaceae*): Species with green ascospores**. *Studies in Mycology* 2003, **48**:116.

58. Yuan JS, Tiller KH, Al-Ahmad H, Stewart NR, Stewart CN, Jr.: **Plants to power: bioenergy to fuel the future**. *Trends Plant Sci* 2008, **13**(8):421-429.

59. Schmer MR, Vogel KP, Mitchell RB, Perrin RK: **Net energy of cellulosic ethanol from switchgrass**. *Proc Nat Acad Sci USA* 2008, **105**:464-469.

60. Lynd LR, Laser MS, Bransby D, Dale BE, Davison B, Hamilton R, Himmel M, Keller M, McMillan JD, Sheehan J *et al*: **How biotech can transform biofuels**. *Nat Biotech* 2008, **26**(2):169-172.

61. Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJ, Hallett JP, Leak DJ, Liotta CL *et al*: **The path forward for biofuels and biomaterials**. *Science* 2006, **311**(5760):484-489.

62. Somerville C: **Biofuels**. *Current Biology* 2007, **17**(4):R115-R119.

63. Somerville C: **The billion-ton biofuels vision**. *Science* 2006, **312**(5778):1277-1277.

64. Himmel ME: **Biomass recalcitrance: engineering plants and enzymes for biofuels production (vol 315, pg 804, 2007)**. *Science* 2007, **316**(5827):982-982.

65. Murnen HK, Balan V, Chundawat SPS, Bals B, Sousa LD, Dale BE: **Optimization of ammonia fiber expansion (AFEX) pretreatment and enzymatic hydrolysis of *Miscanthus x giganteus* to fermentable sugars**. *Biotechnol Progr* 2007, **23**(4):846-850.

66. Den Haan R, Rose SH, Lynd LR, van Zyl WH: **Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae***. *Metabolic Engineering* 2007, **9**(1):87 - 94.

67. Lu YP, Zhang YHP, Lynd LR: **Enzyme-microbe synergy during cellulose hydrolysis by *Clostridium thermocellum***. *Proc Nat Acad Sci USA* 2006, **103**(44):16165-16169.

68. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS: **Microbial cellulose utilization: Fundamentals and biotechnology**. *Microbiol Mol Biol Rev* 2002, **66**(3):506-577.

69. Eneyskaya EV, Kulminskaya AA, Savel'ev AN, Shabalin KA, Golubev AM, Neustroev KN: **alpha-mannosidase from *Trichoderma reesei* participates in the postsecretory deglycosylation of glycoproteins**. *Biochem Biophys Res Commun* 1998, **245**(1):43 - 49.

70. Imperiali B, O'Connor SE: **Effect of N-linked glycosylation on glycopeptide and glycoprotein structure**. *Current Opinion in Chemical Biology* 1999, **3**(6):643 - 649.

71. Jeoh T, Michener W, Himmel M, Decker S, Adney W: **Implications of cellobiohydrolase glycosylation for use in biomass conversion**. *Biotechnol Biofuels* 2008, **1**(1):10.

72. Merino ST, Cherry J: **Progress and challenges in enzyme development for biomass utilization**.*Biofuels.* vol. 108; 2007: 95-120.

73. Jun H, Bing Y, Keying Z, Xuemei D, Daiwen C: **Thermostable carbohydrate binding module increases the thermostability and substrate-binding capacity of *Trichoderma reesei* xylanase 2**. *N Biotechnol* 2009.

74. Tsujibo H, Ohtsuki T, Iio T, Yamazaki I, Miyamoto K, Sugiyama M, Inamori Y: **Cloning and sequence analysis of genes encoding xylanases and acetyl xylan esterase from *Streptomyces thermoviolaceus* OPC-520**. *Appl Environ Microbiol* 1997, **63**(2):661-664.

75. Barak Y, Nov Y, Ackerley DF, Matin A: **Enzyme improvement in the absence of structural knowledge: a novel statistical approach**. *ISME Journal: Multidisciplinary Journal of Microbial Ecology* 2008, **2**(2):171-179.

76. Arnold FH: **Fancy footwork in the sequence space shuffle**. *Nat Biotech* 2006, **24**(3):328-330.

77. Xu J, Takakuwa N, Nogawa M, Okada H, Morikawa Y: **A third xylanase from *Trichoderma reesei* PC-3-7**. *Applied microbiology and biotechnology* 1998, **49**(6):718-724.

78. Purmonen M, Valjakka J, Takkinen K, Laitinen T, Rouvinen J: **Molecular dynamics studies on the thermostability of family 11 xylanases**. *Protein Engineering Design and Selection* 2007, **20**(11):551-559.

79. Torronen A, Harkki A, Rouvinen J: **Three-dimensional structure of endo-1,4-beta-xylanase II from *Trichoderma reesei*: two conformational states in the active site**. *The EMBO journal* 1994, **13**(11):2493-2501.

80. Liu SM, Liu LT, Uzuner U, Zhou X, Gu MX, Shi WB, Zhang YX, Dai SY, Yuan JS: **HDX-Analyzer: a novel package for statistical analysis of protein structure dynamics**. *Bmc Bioinformatics* 2011, **12**.

81. Uzuner U, Shi WB, Liu LT, Liu SM, Dai SY, Yuan JS: **Enzyme structure dynamics of xylanase I from *Trichoderma longibrachiatum***. *Bmc Bioinformatics* 2010, **11**.

82. Collins T, Gerday C, Feller G: **Xylanases, xylanase families and extremophilic xylanases**. *FEMS Microbiology Reviews* 2005, **29**(1):3-23.

83. Ben-Nissan G, Sharon M: **Capturing protein structural kinetics by mass spectrometry**. *Chemical Society Reviews* 2011, **40**(7):3627-3637.

84. Li J, Lim MS, Li S, Brock M, Pique ME, Woods VL, Craig L: ***Vibrio cholerae* toxin-coregulated pilus structure analyzed by hydrogen/deuterium exchange mass spectrometry**. *Structure* 2008, **16**(1):137-148.

85. Brock M, Fan F, Mei FC, Li S, Gessner C, Woods VL, Cheng X: **Conformational analysis of epac activation using amide hydrogen/deuterium exchange mass spectrometry**. *Journal of Biological Chemistry* 2007, **282**(44):32256-32263.

86. Lambris JD, Sfyroera G, Schuster M, Chen H, Tzekou A, Papp K, Winters M, Woods VL: **Studies on the solvent accessibility of native C3 and its fragments, as analyzed by HDX-MS**. *Molecular Immunology* 2007, **44**(1-3):202-202.

87. Eyles SJ, Kaltashov IA: **Methods to study protein dynamics and folding by mass spectrometry**. *Methods* 2004, **34**(1):88-99.

88. Englander SW: **Hydrogen exchange and mass spectrometry: A historical perspective**. *J Am Soc Mass Spectrom* 2006.

89. Begley MJ, Taylor GS, Brock MA, Ghosh P, Woods VL, Dixon JE: **Molecular basis for substrate recognition by MTMR2, a myotubularin family phosphoinositide phosphatase**. *Proc Nat Acad Sci USA* 2006, **103**(4):927-932.

90. Derunes C, Burgess R, Iraheta E, Kellerer R, Becherer K, Gessner CR, Li S, Hewitt K, Vuori K, Pasquale EB *et al*: **Molecular determinants for interaction of SHEP1 with Cas localize to a highly solvent-protected region in the complex**. *FEBS letters* 2006, **580**(1):175-178.

91. Roder H, Elove GA, Englander SW: **Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR**. *Nature* 1988, **335**(6192):700-704.

92. Black BE, Foltz DR, Chakravarthy S, Luger K, Woods VL, Cleveland DW: **Structural determinants for generating centromeric chromatin**. *Nature* 2004, **430**(6999):578-582.

93. Powell KD, Fitzgerald MC: **High-throughput screening assay for the tunable selection of protein ligands**. *Journal of combinatorial chemistry* 2004, **6**(2):262-269.

94. Spraggon G, Pantazatos D, Klock HE, Wilson IA, Woods VL, Lesley SA: **On the use of DXMS to produce more crystallizable proteins: Structures of the *Thermotoga maritima* proteins TM0160 and TM1171**. *Protein Science* 2005, **14**(6):1688-1688.

95. Wong L, Miyashita O, Woods VL, Onuchic JN, Adams JA, Jennings PA: **Protein-protein interactions of Csk probed by H-D exchange and computational analysis**. *Protein Sci* 2004, **13**:201-201.

96. Krishna MM, Hoang L, Lin Y, Englander SW: **Hydrogen exchange methods to study protein folding**. *Methods* 2004, **34**:51-64.

97. Chalmers MJ, Busby SA, Pascal BD, He Y, Hendrickson CL, Marshall AG, Griffin PR: **Probing protein ligand interactions by automated hydrogen/deuterium exchange mass spectrometry**. *Anal Chem* 2006, **78**(4):1005-1014.

98. Tong Y, Wuebbens MM, Rajagopalan KV, Fitzgerald MC: **Thermodynamic analysis of subunit interactions in *Escherichia coli* molybdopterin synthase**. *Biochemistry* 2005, **44**(7):2595-2601.

99. Powell K, Wang M, Silinski P, Ma L, Wales T, Dai S, Warner A, Yang X, Fitzgerald M: **The accuracy and precision of a new H/D exchange- and mass spectrometry-based technique for measuring the thermodynamic stability of proteins**. *Analytica Chimica Acta,* 2003, **496**(1-2):225-232.

100. Ma L, Fitzgerald MC: **A new H/D exchange- and mass spectrometry-based method for thermodynamic analysis of protein-DNA interactions**. *Chem Biol* 2003, **10**(12):1205-1213.

101. Powell KD, Fitzgerald MC: **Measurements of protein stability by H/D exchange and matrix-assisted laser desorption/ionization mass spectrometry using picomoles of material**. *Anal Chem* 2001, **73**(14):3300-3304.

102. Chalmers MJ, Busby SA, Pascal BD, Southern MR, Griffin PR: **A two-stage differential hydrogen deuterium exchange method for the rapid characterization of protein/ligand interactions**. *J Biomol Tech* 2007, **18**(4):194-204.

103. Maity H, Lim WK, Rumbley JN, Englander SW: **Protein hydrogen exchange mechanism: Local fluctuations**. *Protein Sci* 2003, **12**(1):153-160.

104. Agarwal PK: **Role of protein dynamics in reaction rate enhancement by enzymes**. *Journal Of The American Chemical Society* 2005, **127**(43):15248-15256.

105. Daniel RM, Dunn RV, Finney JL, Smith JC: **The role of dynamics in enzyme activity**. *Annual Review of Biophysics and Biomolecular Structure* 2003, **32**(1):69-92.

106. Eisenmesser E, Bosco D, Akke M, Kern D: **Enzyme dynamics during catalysis**. *Science* 2002, **295**:1520 - 1523.

107. Hammes-Schiffer S, Benkovic SJ: **Relating protein motion to catalysis**. *Annual Review of Biochemistry* 2006, **75**(1):519-541.

108. Agarwal PK: **Enzymes: An integrated view of structure, dynamics, and function**. *Microbial Cell Factories* 2006, **5**:2-12.

109. Dai SY, Chalmers MJ, Bruning J, Bramlett KS, Osborne HE, Montrose-Rafizadeh C, Barr RJ, Wang Y, Wang MM, Burris TP *et al*: **Prediction of the tissue-specificity of selective estrogen receptor modulators by using a single biochemical method**. *Proc Nat Acad Sci USA* 2008, **105**(20):7171-7176.

110. Weis DD, Engen JR, Kass IJ: **Semi-automated data processing of hydrogen exchange mass spectra using HX-Express**. *J Am Soc Mass Spectrom* 2006, **17**(12):1700-1703.

111. Pascal BD, Chalmers MJ, Busby SA, Mader CC, Southern MR, Tsinoremas NF, Griffin PR: **The deuterator: software for the determination of backbone**

amide deuterium levels from H/D exchange MS data. *Bmc Bioinformatics* 2007, **8**:156.

112. Pascal BD, Chalmers MJ, Busby SA, Griffin PR: **HD desktop: an integrated platform for the analysis and visualization of H/D exchange data**. *J Am Soc Mass Spectrom* 2009, **20**(4):601-610.

113. Hotchko M, Anand GS, Komives EA, Ten Eyck LF: **Automated extraction of backbone deuteration levels from amide H/2H mass spectrometry experiments**. *Protein Science* 2006, **15**(3):583-601.

114. Slysz GW, Baker CA, Bozsa BM, Dang A, Percy AJ, Bennett M, Schriemer DC: **Hydra: software for tailored processing of H/D exchange data from MS or tandem MS analyses**. *Bmc Bioinformatics* 2009, **10**:162.

115. Nikamanon P, Pun E, Chou W, Koter MD, Gershon PD: **"TOF2H": a precision toolbox for rapid, high density/high coverage hydrogen-deuterium exchange mass spectrometry via an LC-MALDI approach, covering the data pipeline from spectral acquisition to HDX rate analysis**. *Bmc Bioinformatics* 2008, **9**:387.

116. Chik JK, Vande Graaf JL, Schriemer DC: **Quantitating the statistical distribution of deuterium incorporation to extend the utility of H/D exchange MS data**. *Anal Chem* 2006, **78**(1):207-214.

117. Dai SY, Burris TP, Dodge JA, Montrose-Rafizadeh C, Wang Y, Pascal BD, Chalmers MJ, Griffin PR: **Unique ligand binding patterns between estrogen receptor α and β revealed by hydrogen−deuterium exchange**. *Biochemistry* 2009, **48**(40):9668-9676.

118. DC. D: **Darwin's dangerous idea: Evolution and the meanings of life**. New York, NY: Simon & Schuster Inc.; 1995.

119. Chatterjee R, Yuan L: **Directed evolution of metabolic pathways**. *Trends in Biotechnology* 2006, **24**(1):28-38.

120. Bae E, Phillips GN: **Roles of static and dynamic domains in stability and catalysis of adenylate kinase**. *Proc Nat Acad Sci USA* 2006, **103**(7):2132-2137.

121. Konermann L, Tong X, Pan Y: **Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches**. *Journal of Mass Spectrometry* 2008, **43**(8):1021-1036.

122. Igor A, Kaltashov SJE: **Mass spectrometry in biophysics**: John Wiley & Sons; 2005.

123. Pan J, Rintala-Dempsey AC, Li Y, Shaw GS, Konermann L: **Folding kinetics of the S100A11 protein dimer studied by time-resolved electrospray mass spectrometry and pulsed hydrogen−deuterium exchange**. *Biochemistry* 2006, **45**(9):3005-3013.

124. Miranker A, Robinson C, Radford S, Aplin R, Dobson C: **Detection of transient protein folding populations by mass spectrometry**. *Science* 1993, **262**(5135):896-900.

125. Powell KD, Ghaemmaghami S, Wang MZ, Ma L, Oas TG, Fitzgerald MC: **A General mass spectrometry-based assay for the quantitation of protein−ligand binding interactions in solution**. *J Am Chem Soc* 2002, **124**(35):10256-10257.

126. Begley MJ, Taylor GS, Brock MA, Ghosh P, Woods VL, Dixon JE: **Molecular basis for substrate recognition by MTMR2, a myotubularin family phosphoinositide phosphatase**. *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(4):927-932.

127. Iyer GH, Garrod S, Woods Jr VL, Taylor SS: **Catalytic independent functions of a protein kinase as revealed by a kinase-dead mutant: Study of the Lys72His mutant of cAMP-dependent kinase**. *J Mol Biol* 2005, **351**(5):1110-1122.

128. Hoofnagle AN, Resing KA, Ahn NG: **Protein analysis by hydrogen exchange**. *Annual Review of Biophysics and Biomolecular Structure* 2003, **32**(1):1-25.

129. Quin MB, Schmidt-Dannert C: **Engineering of biocatalysts: from evolution to creation**. *ACS Catalysis* 2011, **1**(9):1017-1021.

130. Kim Y-S, Jung H-C, Pan J-G: **Bacterial cell surface display of an enzyme library for selective screening of improved cellulase variants**. *Appl Environ Microbiol* 2000, **66**(2):788-793.

131. Murashima K, Kosugi A, Doi RH: **Thermostabilization of cellulosomal endoglucanase EngB from *Clostridium cellulovorans* by in vitro DNA recombination with non-cellulosomal endoglucanase EngD**. *Molecular Microbiology* 2002, **45**(3):617-626.

132. Wang T, Liu X, Yu Q, Zhang X, Qu Y, Gao P, Wang T: **Directed evolution for engineering pH profile of endoglucanase III from *Trichoderma reesei***. *Biomolecular Engineering* 2005, **22**(1-3):89-94.

133. Colot HV, Park G, Turner GE, Ringelberg C, Crew CM, Litvinkova L, Weiss RL, Borkovich KA, Dunlap JC: **A high-throughput gene knockout procedure for Neurospora reveals functions for multiple transcription factors**. *Proc Nat Acad Sci USA* 2006, **103**(27):10352-10357.

134. Kitazono AA: **Improved gap-repair cloning method that uses oligonucleotides to target cognate sequences**. *Yeast* 2009, **26**(9):497-505.

135. Harju S, Fedosyuk H, Peterson KR: **Rapid isolation of yeast genomic DNA: Bust n' Grab**. *BMC Biotechnol* 2004, **4**.

136. Bailey MJ, Biely P, Poutanen K: **Interlaboratory testing of methods for assay of xylanase activity**. *Journal of biotechnology* 1992, **23**(3):257-270.

137. Liu T, Wang T, Li X, Liu X: **Improved heterologous gene expression in *Trichoderma reesei* by cellobiohydrolase I gene (cbh1) promoter optimization**. *Acta Biochimica et Biophysica Sinica* 2008, **40**(2):158-165.

138. Schueler-Furman O, Wang C, Bradley P, Misura K, Baker D: **Progress on modeling of protein structures and interactions**. *Science* 2005, **310**(5748):638-642.

139. Mendes J, Guerois R, Serrano L: **Energy estimation in protein design**. *Current Opinion In Structural Biology* 2002, **12**(4):441-446.

140. Butterfoss GL, Kuhlman B: **Computer-based design of novel protein structures**. *Annual Review Of Biophysics And Biomolecular Structure* 2006, **35**:49-65.

141. Polizzi KM, Bommarius AS, Broering JM, Chaparro-Riggers JF: **Stability of biocatalysts**. *Current Opinion in Chemical Biology* 2007, **11**(2):220-225.

142. Tokuriki N, Tawfik DS: **Protein dynamism and evolvability**. *Science* 2009, **324**(5924):203-207.

143. Khademi S, Zhang DC, Swanson SM, Wartenberg A, Witte K, Meyer EF: **Determination of the structure of an endoglucanase from *Aspergillus niger* and its mode of inhibition by palladium chloride**. *Acta Crystallogr D* 2002, **58**:660-667.

144. Bezerra RMF, Dias AA: **Enzymatic kinetic of cellulose hydrolysis - Inhibition by ethanol and cellobiose**. *Appl Biochem Biotech* 2005, **126**(1):49-59.

145. Tang QY, Feng MG: **DPS data processing system for practical analysis**. Beijing: Science Press; 2002.

146. Torronen A, Kubicek CP, Henrissat B: **Amino acid sequence similarities between low molecular weight endo-1,4-beta-xylanases and family H cellulases revealed by clustering analysis**. *Febs Lett* 1993, **321**(2-3):135-139.

147. Torronen A, Rouvinen J, Ahlgren M, Harkki A, Visuri K: **Crystallization and preliminary X-ray analysis of two major xylanases from *Trichoderma reesei***. *J Mol Biol* 1993, **233**(2):313-316.

APPENDIX

VI.1. of

VI.2.

```
                        10        20        30        40        50        60
                         |         |         |         |         |         |
1XYN          ---------ASINYDQNYQTG-GQVSYSP-SNTGFSVNWN--TQDDFVVGVGWTTGSSAP
1BK1          --------SAGINYVQNYNGNLGDFTYDE-SAGTFSMYWEDGVSSDFVVGLGWTTGSSNA
1UKR          --------SAGINYVQNYNGNLGDFTYDE-SAGTFSMYWEDGVSSDFVVGLGWTTGSSNA
1ENX          ETIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGN--FVGGKGWQPGTKNK
1YNA          ETTPNSEGWHDGYYYSWWSDGGAQATYTNLEGGTYEISWGDGGN--LVGGKGWNPGLNAR
1F5J          ALTSNASGTFDGYYYELWKDT-GNTTMTVYTQGRFSCQWSNINNALFRTGKKYNQNWQSL
1XNB          ---------ASTDYWQNWTDGGGIVNAVNGSGGNYSVNWSNTGN--FVVGKGWTTGSPFR
2DCY          ---------ASTDYWQNWTDGGGIVNAVNGSGGNYSVNWSNTGN--FVVGKGWTTGSPFR
Consensus                a   % qnw dg g vty   s g %svnW    n  fvvGkgwttgs
Prim.cons.    ETT3N33GSAS22YYQNWNDGGG3VTYTNGSGGTFSVNWSNGGNSDFVVGKGWTTGSSNR

                        70        80        90       100       110       120
                         |         |         |         |         |         |
1XYN          ----INFGGSFSVNSGTGLLSVYGWSTNPLVEYYIMED--NHNYPAQGTVKGTVTSDGAT
1BK1          ----ITYSAEYSASGSSSYLAVYGWVNYPQAEYYIVEDYGDYNPCSSATSLGTVYSDGST
1UKR          ----ITYSAEYSASGSASYLAVYGWVNYPQAEYYIVEDYGDYNPCSSATSLGTVYSDGST
1ENX          V---INFS-GSYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSV
1YNA          A---IHFE-GVYQPNGNSYLAVYGWTRNPLVEYYIVENFGTYDPSSGATDLGTVECDGSI
1F5J          GTIRITYS-ATYNPNGNSYLCIYGWSTNPLVEFYIVESWGNWRPP-GATSLGQVTIDGGT
1XNB          T---INYNAGVWAPNGNGYLTLYGWTRSPLIEYYVVDSWGTYRPT--GTYKGTVKSDGGT
2DCY          T---INYNAGVWAPNGNGYLTLYGWTRSPLIEYYVVDSWGTYRPT--GTYKGTVKSDGGT
Consensus        In%      pngnsyL vYGW  nPl E%Y!v#  g ynp   aT lGtV sDG t
Prim.cons.    TTIRINYSAGV2APNGNSYLAVYGW2RNPL2EYYIVE2WGTYNP4SGATSLGTVTSDGST

                       130       140       150       160       170       180
                         |         |         |         |         |         |
1XYN          YTIWENTRVNEPSIQGT-ATFNQYISVRNSPRTSG---TVTVQNHFNAWASLGLHL-GQM
1BK1          YQVCTDTRTNEPSITGT-STFTQYFSVRESTRTSG---TVTVANHFNFWAQHGFGN-SDF
1UKR          YQVCTDTRTNEPSITGT-STFTQYFSVRESTRTSG---TVTVANHFNFWAHHGFGN-SDF
1ENX          YDIYRTQRVNQPSIIGT-ATFYQYWSVRRNHRSSG---SVNTANHFNAWAQQGLTL-GTM
1YNA          YRLGKTTRVNAPSIDGT-QTFDQYWSVRQDKRTSG---TVQTGCHFDAWARAGLNVNGDH
1F5J          YDIYRTTRVNQPSIVGT-ATFDQYWSVRTSKRTSG---TVTVTDHFRAWANRGLNL-GTI
1XNB          YDIYTTTRYNAPSIDGDRTTFTQYWSVRQSKRPTGSNATITFTNHVNAWKSHGMNLGSNW
2DCY          YDIYTTTRYNAPSIDGDRTTFTQYWSVRQSKRPTGSNATITFSNHVNAWKSHGMNLGSNW
Consensus     Y i  ttRvN PSI Gt  TF QYwSVR s RtsG   t!tv nHfnaWa  Gl l g
Prim.cons.    YDIYTTTRVN2PSIDGTRATFTQYWSVRQSKRTSGSNATVTVANHFNAWASHGLNLG2D3

                       190       200
                         |         |
1XYN          NYQVVAVEGWGGSGSASQSVSN-----
1BK1          NYQVMAVEAWSGAGSASVTISS-----
1UKR          NYQVVAVEAWSGAGSASVTISS-----
1ENX          DYQIVAVEGYFSSGSASITVS------
1YNA          YYQIVATEGYFSSGYARITVADVG---
1F5J          DQITLCVEGYQSSGSANITQNTFSQSS
1XNB          AYQVMATEGYQSSGSSNVTVW------
2DCY          AYQVMATEGYQSSGSSNVTVW------
Consensus      yqvvavEgy ssGsas tvs
Prim.cons.    NYQVVAVEGYQSSGSASVTVSS22QSS
```

Figure 1. Multiple sequence alignment for XYN I enzyme from different organisms. A multiple sequence alignment of XYNI enzymes from various fungal and bacterial sources is given in order to elucidate the sequence similarity of family 11 (GH 11) xylanases. 1XYN: XYNI from *T. reesei*; 1ENX: XYNII from *T. reesei*; 1BK1: xylanase 1 from *Aspergillus kawachii*; 1UKR: xylanase 1 from *Aspergillus niger*; 1YNA: xylanase 1 from *Thermomyces lanuginosus*; 1F5J: xylanase 1 from *Dictyoglomus thermophilum*; 1XNB: xylanase 1 from *Bacillus circulans*; 2DCY: xylanase 1 from *Bacillus subtilis*.