EXPERIMENTAL AND COMPUTATIONAL STUDIES OF A FUNGAL CHITINASE

By

Faez Iqbal Khan

(Reg. No: 21243044)

Submitted in fulfillment of the requirements of the degree of Doctor of Philosophy: Chemistry in the Faculty of Applied Sciences at the Durban University of Technology

March 2015

DECLARATION

I **Faez Iqbal Khan** declare that the thesis submitted for the degree of Doctor of Philosophy (Ph.D.): Chemistry at the Durban University of Technology is the result of my own investigation and has not already been accepted in substance for any degree, and is not being concurrently submitted for any other degree. All the work was done by the Author.

_____ _____

Signature of the student Date

_____ _____

Signature of the promoter Date

_____ _____

Signature of the promoter Date

_____ _____

Signature of the promoter Date

ACKNOWLEDGEMENTS

I thank Almighty Allah, The most beneficent, the most merciful, for giving me the patience and perseverance throughout my life to move in the right direction.

I owe profound gratitude and sincere regards to my supervisors Prof K Bisetty, Head of department of Chemistry, for his benevolent support and direction throughout my research. His careful corrections during regular meetings and discussions always ensured a more thoughtful approach towards problem solving.

It is my privilege to express my gratitude and indebtedness to my supervisor in the Department of Biotechnology and Food Technology Prof Suren Singh, Executive Dean of Faculty of Applied Sciences, for his guidance, continuous encouragement, constructive suggestions and sagacious advice throughout the course of my research work. His keen sense of scientific temperament was a source of inspiration to me.

I express my special thanks to Prof Kugen Permaul, Head of Department of Biotechnology and Food Technology for his excellent guidance and valuable suggestions through my research periods.

I am also thankful to Dr. Imtaiyaz Hassan, Assistant Professor, Center for Interdisciplinary research in Basic Sciences, Jamia Millia Islamia, New Delhi, India, for his support and time.

I would also like to extend my acknowledgements to Dr. Algasan Govender who has been very compassionate and informative from the preliminary studies of this project.

I would like to express my gratitude to the Centre for high performance computing, an initiative support by the Department of Science and Technology of South Africa.

I would like to thank all my Computational Modeling and Bio-Analytical Chemistry (CMBAC) group members Myalo, Parvesh, Ayyappa, Kanchi, Deepali and Shahbaaz for unforgettable co-operation.

I would like to thank my Lab mates Ashira, Evashni and Khadija from the Department of Biotechnology and Food technology at DUT, with whom I share numerous memorable moments which made my research really enjoyable.

My heartfelt thanks to Danish, Zeya, Ehsan, Alamgir, Moneer, Haider, Rafat, Aamir, Waquar, Rashid, Abhishek, Munazzah, Zubair, Shafi, Anand, Meraj, Razique, Nadir, Saad, Bilal, Poonam, Faiz, Tanushree, Parul and Abilar, who unconditionally supported me throughout my research duration.

I also take this opportunity to thank all the staff members in the Department of chemistry who have helped me in various ways during my research work.

My parents, brothers and sisters deserve special mention for their inseparable support, love and care. I cannot thank them enough for being the source of strength and inspiration.

I dedicate this thesis to my beautiful nieces Aroush Khan, Aleena Khan, Areeba Khan, Ayesha Khan, Zainab Khan, Fatima Khan and handsome nephews Aayan Khan, Ammar Khan and Hadeed Khan. They brought joy and happiness to my life, I love them so much and I thank Allah for them, they are a great blessing.

I wish I could mention each individually. I am thankful to all those who directly or indirectly contributed in this work.

Faez Iqbal Khan

ABSTRACT

Chitin, the second most abundant natural biopolymer, is composed of repeating units of *N*-acetyl-β-D-glucosamine and primarily forms the structural component of protective biological matrices such as fungal cell walls and exoskeletons of insects. Chitinases are a ubiquitous class of extracellular enzymes that have gained attention in the past few years due to their wide range of biotechnological applications, especially in the field of agriculture for bio-control of fungal phytopathogens. They play an important role in the defense of organisms against chitin-containing parasites by hydrolyzing the β-1,4-linkages in chitin and hence act as anti-fungal as well as anti-biofouling agents. Moreover, the effectiveness of conventional insecticides is increasingly compromised by the occurrence of resistance and thus, chitinases offer a potential alternative to the use of chemical fungicides. In recent years, thermostable enzymes isolated from thermophilic microorganisms have gained widespread attention in industrial, medical, environmental and biotechnological applications due to their inherent stability at high temperatures and a wide range of pH optima. Determination of the three-dimensional structure of a protein can provide important details about its biological functions and its mode of action. However, despite their significance, the precise three-dimensional structures of most of the chitinases, including those isolated from *Thermomyces lanuginosus* is not fully characterized so far. Hence, the main focus of the present study was to gain a better understanding of the structural features of chitinases obtained from this thermostable fungus using both experimental and computational techniques, and their relationship with their activity profiles. The genes encoding thermostable chitinase II from *T. lanuginosus* were isolated and cloned in both *E. coli* as well as the *Pichia pastoris* expression system. Analysis of the nucleotide sequences revealed that the chitinase II gene (1196 bp) encodes a 343 amino acid

protein of molecular weight 36.65 kDa whereas the chitinase I gene (1538 bp) encodes a 400 amino acid protein of molecular weight 44.14 kDa. *In silico* protein modeling was helpful in predicting the 3D models of the novel chitinase II enzyme, followed by the prediction of its active sites. The presence of Glu176 was found to be essential for the activity of chitinase II. Similarly, analysis of chitinase I revealed several active sites in its structural framework. A 10 ns Molecular dynamics (MD) simulations was implemented to assess the conformational preferences of chitinases. The MD trajectories at different temperatures clearly revealed that the stability of the enzymes were maintained at higher temperatures. Additionally, a constant pH molecular dynamics simulations at a pH range 2-6 was performed to establish the optimum activity and stability profiles of chitinase I and chitinase II. For this purpose, the Molecular Dynamics simulations were carried out at fixed protonation states in an explicit water environment to evaluate the effect of the physiological pH on chitinase I and II enzymes obtained from *T. lanuginosus*. The results suggest a strong conformational pH dependence of chitinases. These enzymes retained their characteristic TIM Barrel fold at the respective protonated conditions, thus validated the experimental outcomes. Further, the different stability and flexibility predictions were used to assess the relation of point mutations and enzyme stabilities. Our results pave the way to engineer new and better thermostable enzymes.

LIST OF CONTENTS

DECLARATION..... I

ACKNOWLEDGEMENTS II

ABSTRACT..... IV

LIST OF CONTENT VI

LIST OF TABLES XIV

LIST OF FIGURES XV

LIST OF ACRONYMS AND SYMBOLS XXI

AIM XXV

OBJECTIVES XXV

RESEARCH OUTPUTS XXVI

CHAPTER 1

INTRODUCTION 1

CHAPTER 2

LITERATURE REVIEW 5

2.1 Chitin..... 5

2.1.1 Chemical structure 5

2.1.2 Chitin in nature 6

2.1.3 Industrial production of chitin 8

2.1.4 Production from fungal sources 9

2.1.5 Applications of chitin 10

2.1.5.1 Enzyme immobilization 11

2.1.5.2 Antioxidant property 11

2.1.5.3 Wound healing property 11

2.1.5.4 Membranes 12

2.1.5.5 Controlled drug-release 12

2.1.5.6 Antimicrobial property 13

2.1.5.7 Anticoagulant activity 13

2.1.5.8 Gene therapy 13

2.1.5.9 Miscellaneous applications 14

2.2 Chitinase 14

2.2.1 Chitinase sources 15

2.2.1.1 Plant chitinases 15

2.2.1.2 Insect chitinases 15

2.2.1.3 Mammalian chitinases 16

2.2.1.4 Microbial chitinases 16

2.2.1.5 Fungal chitinases 17

2.2.2 Classification of chitinase 21

2.2.3 Industrial usage of chitinases 23

2.2.3.1 Fungal protoplast generation.....23

2.2.3.2 Single cell protein production 23

2.2.3.3 Production of chitooligosaccharides 23

2.2.3.4 Biocontrol 24

2.3 *Thermomyces lanuginosus* 24

2.3 Scope of the study.....25

CHAPTER 3

EXPERIMENTAL 28

3.1 Material and Methods 29

3.1.1 Strains, plasmids and growth conditions 29

3.1.2 Total RNA isolation from *Thermomyces lanuginosus* SSBP and first strand cDNA synthesis.....30

3.1.3 Primer design 30

3.1.4 Polymerase chain reaction (PCR) 31

3.1.5 Agarose gel electrophoresis 32

3.1.6 DNA quantification.....33

3.1.7 Restriction digestion and ligation 33

3.1.8 Preparation of competent BL21 cells.....34

3.1.9 Preparation of electro-competent *Pichia pastoris* cells 35

3.1.10 Transformation and selection.....35

3.1.11 Transformation of Pichia pastoris GS115.....36

3.1.12 Protein expression.....36

3.1.13 Plate assay for chitinolytic enzyme activity.....37

3.2 Results and Discussions.....38

3.2.1 Amplification and Cloning of chitinase. 38

3.2.2 Expression of chitinase 40

3.3 Conclusion 42

CHAPTER 4

COMPUTATIONAL 43

4.1 Structure prediction and analysis.....43

4.1.1 Material and Methods 46

4.1.1.1 Template Selection and Fold Assignment 47

4.1.1.2 Template-Target Alignment.....49

4.1.1.3 Three Dimensional (3D) Model Building.....50

4.1.1.4 Loop Modelling 51

4.1.1.5 Side chain Modelling.....52

4.1.1.6 Evaluation and refinement of models 52

4.1.2 Result and Discussion.....53

4.1.2.1 Template search 53

4.1.2.2 Structure prediction and analysis.....54

4.1.3 Conclusions.....60

4.2 Molecular Docking 61

4.2.1 Material and methods.....63

4.2.1.1 Active pocket prediction.....63

4.2.1.2 Ligand docking with CDOCKER 64

4.2.1.3 Preparing Inputs.....64

4.2.1.4 Invoking CDOCKER.....64

4.2.1.5 Analyzing CDOCKER Results.....65

4.2.2 Results and Discussion 65

4.2.3 Conclusions.....68

4.3 Molecular Dynamics simulations.....69

4.3.1 Periodic Boundary Conditions (PBC).....71

4.3.2 Ewald Summation Techniques 72

4.3.3 Particle Mesh Ewald (PME) 73

4.3.4 Thermostats in MD 74

4.3.5 Solvent models.....75

4.3.6 Energy-Minimization Procedures in Simulations.....76

4.3.7 Steepest Descent Method.....76

4.3.8 Conjugate Gradient Method.....77

4.3.9 Molecular Dynamics analysis 79

4.3.10 Material and Methods 80

4.3.11 Results and Discussion 86

4.3.11.1 Root mean square deviation.....86

4.3.11.2 Root mean square fluctuation 87

4.3.11.3 Radius of gyration.....88

4.3.11.4 Secondary Structure 89

4.3.11.5 Total energy of system.....92

4.3.12 Conclusions.....92

4.4 Constant pH MD simulations 93

4.4.1 Materials and Methods.....94

4.4.1.1 Protein Ionization and Residue pK calculations 94

4.4.1.2 Constant pH MD simulations 96

4.4.2 Results and Discussion 97

4.4.3 Conclusions.....105

4.5 *In Silico* stable mutants prediction of chitinase II 106

4.5.1 Materials and Methods.....108

4.5.1.1 Tolerant/Intolerant mutations.....108

4.5.1.2 Reliability index/Stability prediction.....109

4.5.1.3 Flexibility in the protein and its mutants 110

4.5.1.4 Molecular dynamics simulations 110

4.5.2 Results and Discussion 111

4.5.2.1 Sorting intolerant from tolerant 111

4.5.2.2 Prediction and selection of stable mutants.....112

4.5.2.3 Flexibility in the protein and its mutants 116

4.5.2.4 Molecular dynamics simulations 117

4.5.3 Conclusions.....123

CHAPTER 5

SUMMARY 124

CHAPTER 6

CONCLUSIONS 128

CHAPTER 7

REFERENCES.....131

APPENDICES

APPENDIX I: Commands of Gromacs 154

APPENDIX II: Ions.mdp 155

APPENDIX III: Minim.mdp.....156

APPENDIX IV: Nvt.mdp.....157

APPENDIX V: Npt.mdp.....158

APPENDIX VI: Md.mdp 159

APPENDIX VII: Lists of predicted point mutations on the basis of their RI and stability score 160

APPENDIX VIII: RMSF, B-factor and confidence index for (A) normal chitinase II (B) S45E, (C) S45I, (D) S50E, (E) S50I, (F) S50V, (G) normal chitinase II, (H) Q135C, (I) S136I, (J) S136K, (K) S136L and (L) S136V 166

LIST OF TABLES

Table 2.1: Enzymes from *Thermomyces lanuginosus* with their optimum pH and temperature 25

Table 4.1: The obtained CDOCKER energy and CDOCKER interaction energies of 10 docked poses of chitinase II with allosamidin derivatives 66

Table 4.2: Typical vibration frequencies in molecules (wave numbers) and hydrogen bonded liquids [Source: Gromacs user manual 6.5.4]. 71

Table 4.3: Calculation of protein ionization and pK_a value for chitinase I and II 98

Table 4.4: The list of all intolerant positions in the amino acid chain that are less likely to mutate. There exist certain positions where mutation would be lethal thus causing a increase/decrease in protein stability. 111

Table 4.5: The classification of selected highly stable mutants (RI score > 6, (+) indicates the positions that are more prone to mutation 112

Table 4.6: Analysis of selected mutants using I-Mutant 2.0, (pH=7, T=25 ^{o}C; T=37 ^{o}C), MuStab (pH=7, T=25 ^{o}C; T=37 ^{o}C) and SDM server. 115

Table 4.7: Average Total Energy (kJ/mol) of chitinase and its mutants during the time scale of 2000 ps simulations.....118

LIST OF FIGURES

Figure: 1.1: The overall work flow of the research methodology 4

Figure: 2.1 Primary structures of chitin and cellulose. A- Chitin, and B- Cellulose. 6

Figure: 2.2 Chemical structure of chitin depicting its intramolecular hydrogen bonds (dotted lines) between the neigbouring sugar rings in chitin (Smidsrod and Moe, 1995) 7

Figure: 2.3 Schematic representation of the three polymorphic forms of chitin a) α-chitin, b) β-chitin and c) γ-chitin. 8

Figure: 2.4 schematic representations of the wide range of applications of chitin and its derivates. 10

Figure: 2.5 Domain organization of fungal chitinases. SP, signal peptide. GH 18, glycoside hydrolase family 18, BD, binding domain. 19

Figure: 2.6. Schematic representation of action of chitinases on chitin (www.sigmaaldrich.com) 22

Figure: 3.1: Chitinase gene sequence showing primer binding sites (highlighted). 31

Figure: 3.2: chitinase II protein PCR product on agarose gel. Lane 1: 1 kb DNA size marker, Lane 3, 4: Single expected band of size 1032 bp. 38

Figure: 3.3 Restriction digestion of (A) pTz construct, Lane 1: 1 kb DNA size marker, lane 3, 4: expected excised DNA band of size 1032 bp (B) pBGP1construct, Lane 1: 1 kb DNA size marker, lane 3: expected excised DNA band of size 1032 bp (C) pET21c construct, Lane 1: DNA marker, lane 3: expected excised DNA band of size 1032 bp 39

Figure: 3.4 Colony PCR of positive clone of citinase II into pBGP1, Lane 1: 1 kb DNA size marker, lane 2: expected DNA band of size 1032 bp. 40

Figure: 3.5 SDS-PAGE analysis of chitinase II protein stained with Coomassie. Lane 1: Protein

marker, lane 5 -pellet, lane 7 -36KDa chitinase II enzyme. 41

Figure: 3.6 Chitinolytic enzyme plate assay (1, 2, 3 , 4 and 5) chitinase II production medium as negative control (0). 41

Figure: 4.1: Computational workflow adopted for the reliable model generation using homology modelling 46

Figure: 4.2: The predicted three dimensional (3D) structure of (A) chitinases II and (B) chitinase I highlighting the characteristic TIM barrel. 55

Figure: 4.3: The superimposition of predicted model (RED) with template (GREEN) and their calculated DOPE profiles using MODELLER. (A) The prediction for chitinase II. (B) Analysis results for chitinase I. 57

Figure: 4.4 The Ramachandran plot for predicted (A) chitinase II (B) chitinase I enzymes respectively. 58

Figure: 4.5 The topology of (A) chitinase II and (B) chitinase I generated using PDBsum. 59

Figure: 4.6 Structure analysis with template (PDB id: 2UY2) shows active site pocket in chitinase II comprises residues 36, 70, 174, 176, 208, 234, 232 and 307.....65

Figure: 4.7 (A, B and C): The top three docked poses confirming that Glu176 is critical for the activity of chitinase II. 67

Figure: 4.8 Splitting of charges into discrete and smeared distributions in the real and reciprocal space.....72

Figure: 4.9: A 2D schematic representation of particle-mesh techniques used in most Fourier-based methods **(a)** A system of charged particles. **(b)** The charges are interpolated on a 2D grid. **(c)** Using FFT, the potential and forces are calculated at grid points. **(d)** Interpolate forces back to particles and update coordinates. 74

Figure: 4.10: The computational workflow for the molecular dynamics simulations using GROMACS package.....82

Figure: 4.11: The chitinase II molecule immersed in water is subjected to NVT equilibration at (A) 300 K, (B) 325 K and (C) 350 K.....84

Figure: 4.12: The chitinase II molecule immersed in water is subjected to NPT equilibration at (A) 300 K, (B) 325 K, and (C) 350 K 85

Figure: 4.13: RMS deviation values for (A) chitinase II and (B) chitinase I at different temperatures as a function of time. Black, Red, Green and Blue colour represent RMS deviations obtained from 300 K, 325 K, 350 K and 375 K respectively. The duration of each simulation is 10 ns.....87

Figure: 4.14: Average RMS deviation values as a function of amino acid sequence numbers for (A) chitinase II and (B) chitinase I. Values were calculated with the use of Cα atoms. Black, Red, Green and Blue colour represent RMS fluctuation obtained from 300 K, 325 K, 350 K and 375 K respectively. 88

Figure: 4.15: The radius of gyration values as a function of time for (A) chitinase II and (B) chitinase I. Values were calculated with the use of Cα atoms. Black, Red, Green and Blue colour represent radius of gyration obtained from 300 K, 325 K, 350 K and 375 K respectively. 89

Figure: 4.16: The graphical representation showing the stability of structural elements of chitinase II at (A)300 K, (B)325 K, (C)350 K, and (D)375 K. The secondary structure elements were comparable at 300 K-350 K.....90

Figure: 4.17: The graphical representation showing the stability of structural elements of chitinase I at (A)300 K, (B)325 K, (C)350 K, and (D)375 K. The secondary structure elements were comparable at 300 K-350 K.....91

Figure: 4.18: The pH dependence of total charge at different protonated conditions of (A) chitinase I, A1 and A2 represents CHARMm and CHARMm Polar H force fields respectively (B) chitinase II, B1, B2 represents chitinase II protein with CHARMm and CHARMm Polar H force fields respectively, while B3, B4 represent chitinase II without signal peptide typed with CHARMm and CHARMm Polar H force fields respectively. 97

Figure: 4.19: Relative Folding energy at different protonated conditions. (A) chitinase I, A1 and A2 represents CHARMm and CHARMm Polar H force fields respectively (B) chitinase II, B1, B2 represents chitinase II protein with CHARMm and CHARMm Polar H force fields respectively, while B3, B4 represent chitinase II without signal peptide typed with CHARMm and CHARMm Polar H force fields respectively. 98

Figure: 4.20: Electrostatic Energy at different protonated conditions. (A) chitinase I, A1 and A2 represents CHARMm and CHARMm Polar H force fields respectively (B) chitinase II, B1, B2 represents chitinase II protein with CHARMm and CHARMm Polar H force fields respectively, while B3, B4 represent chitinase II without signal peptide typed with CHARMm and CHARMm Polar H force fields respectively. 99

Figure: 4.21: The Average total energy values as a function of pH for (A) chitinase I and (B) chitinase II. The values obtained from pH 2-6 respectively. 100

Figure: 4.22: The radius of gyration values as a function of time for (A) chitinase I and (B) chitinase II. Values were calculated with the use of Cα atoms. Black, Red, Green, Blue and Yellow colour represent radius of gyration obtained from pH 2-6 respectively. 101

Figure: 4.23: RMS deviation values for (A) chitinase I and (B) chitinase II at different pH as a function of time. Black, Red, Green, Blue and Yellow colour represent RMS deviations obtained from pH 2-6 respectively. The duration of each simulation is 20 ns. 102

Figure: 4.24: Average RMS deviation values as a function of amino acid sequence numbers for (A) chitinase I and (B) chitinase II. Values were calculated with the use of Cα atoms. Black, Red, Green, Blue and Yellow colour represent RMS fluctuation obtained from pH 2-6 respectively. 103

Figure: 4.25: The superimposition of native and simulated structures of chitinase I and II at pH 2-6, 300 K. Green and Blue represents native and the simulated chitinases structures respectively. Thickness of the coil represents the fluctuations of heavy atoms around the average structure.....104

Figure: 4.26: Frequency of amino acids substitutions observed in chitinase II on the basis of RI. 111

Figure: 4.27: Graphical representations of selected stable mutants with changes in their polar contacts (A) S45E and S45I; S(Red), E (green) and I(blue) (B) S50E, S50I and S50V; S(red), E(green), I(blue) and V(yellow) (C) Q135C; Q(red) and C(green) (D) S136I, S136L and S136V; S(red), I(green), L(blue) and V(yellow). 117

Figure: 4.28: The radius of gyration values as a function of time for (A) wild chitinase II (black), S45E (red) and S45I (green) (B) S50E (red), S50I (green) and S50V (blue) (C) Q135C (red) and Q135K (green) (D) S136I (red), S136L (green) and S136V (blue). Values were calculated with the use of Cα atoms. 119

Figure: 4.29: RMS deviation values as a function of time for (A) wild chitinase II (black), S45E (red) and S45I (green) (B) S50E (red), S50I (green) and S50V (blue) (C) Q135C (red) and Q135K (green) (D) S136I (red), S136L (green) and S136V (blue). Values were calculated with the use of Cα atoms.....121

Figure: 4.30: Average RMS deviation values as a function of amino acid sequence numbers for (A) wild chitinase II (black), S45E (red) and S45I (green) (B) S50E (red), S50I (green) and S50V (blue) (C) Q135C (red) and Q135K (green) (D) S136I (red), S136L (green) and S136V (blue). Values were calculated with the use of Cα atoms. 123

List of Acronyms and Symbols

HCl	Hydro Chloric Acid
NaOH	Sodium Hydroxide
DNA	Deoxyribonucleic acid
AMCase	Acidic mammalian chitinase
CBM	Carbohydrate-binding module
COSs	Chito-oligosaccharides
SCPs	Single cell proteins
AOX1	Methanol-induced alcohol oxidase
zeoR	Zeocin resistance gene
ampR	Ampicillin resistance gene
MCS	Multiple cloning site
RNA	Ribonucleic acid
RT-PCR	Reverse transcription polymerase chain reaction
3-D	Three Dimensional
NMR	Nuclear magnetic resonance
EC	Enzyme Commission

PDB	Protein Data Bank
DS	Discovery Studio
SCOP	Structural Classification of Proteins
BLAST	Basic Local Alignment Search Tool
HMM	Hidden Markov Model
SMART	Simple Modular Architecture Research Tool
SYSTERS	SYSTEmatic Re-Searching
PRALINE	PRofile ALIgNEment
RMSD	Root Mean Square Deviation
DOPE	Discrete optimized protein energy
Mol PDF	Molecular Probability Density Function
TM-score	Template modeling score
CPU	Central processing unit
MD	Molecular dynamics
MC	Monte Carlo
GA	Genetic Algorithm
DG	Distance Geometry

GOLD	Genetic Optimisation for Ligand *Docking*
AMBER	Assisted Model Building with Energy Refinement
GROMOS	GROningen MOlecular Simulation
GROMACS	GROningen MAchine for Chemical Simulations
PBC	Periodic Boundary Conditions
PME	Particle-Mesh Ewald
GB	Generalized Born model
PE	Poisson Equation
OPLS	Optimized potential for liquid simulation
GRACE	GRaphing Advanced Computation and Exploration of data
VMD	Visual Molecular Dynamics
OPLS-AA/L	Optimized Potential for Liquid Simulations/ all atoms
NVT	Constant number of particles, volume and temperature
NPT	Constant number of particles, pressure and temperature
CHPC	Center for High performance Computing
RMSF	Root Mean Square Fluctuations
SIFT	Sorting Intolerant from Tolerant

PSSM	Positon specific scoring matrix
PredyFlexy	Flexibility and Local Structure Prediction
RI	Reliability Index

AIM

The main focus of this work was to study the conformational profile of fungal chitinases using experimental and computational methods.

OBJECTIVES

- To clone, express and screen the biological activity of chitinase.
- To determine the 3D structure of chitinases using computational techniques.
- To predict the active site using evolutionary, structural information and molecular docking study.
- To employ molecular dynamics (MD) simulations at various temperatures to assess the conformational preferences and thermostable nature of chitinases.
- To employ constant pH molecular dynamics simulation to establish the optimum activity and stability profiles of chitinases.
- To predict and compare the stability of normal and mutated chitinase in order to generate a better thermostable enzymes.

RESEARCH OUTPUTS

ORAL PRESENTATIONS

- Molecular Dynamics (MD) simulations study of a chitinase, National Conference on Recent Trends in Protein Structural Biology, December 16-18, 2013, Jamia Millia Islamia, New Delhi, India-25.
- Computational approaches of thermostable xylanase" International Interdisciplinary Science conference on Protein Folding and Diseases, 2012, Jamia Millia Islamia, New Delhi, India-25.

PUBLICATIONS

- Creation of thermostable and alkaline stable xylanase variants by DNA shuffling, 2014, Dawn Elizabeth Stephens, **Faez Iqbal Khan**, Parvesh Singh, Krishna Bisetty, Suren Singh, Kugen Permaul. Journal of Biotech, DOI: 10.1016/j.jbiotec.2014.07.446
- Role of N-Terminal Residues on Folding and Stability of C-Phycoerythrin: Simulation and Urea-induced Denaturation Studies, Khalid Anwer, Ravi Sonani, DattaMadamwar, Parvesh Singh, **Faez Khan**, Krishna Bisetty, Faizan Ahmad and Md. Imtaiyaz Hassan, Journal of Biomolecular Structure & Dynamics, 2013, DOI:10.1080/07391102.2013.855144
- Thermostable chitinase II from *Thermomyces lanuginosus* SSBP: Cloning, Structure Prediction and Molecular Dynamics Simulations, **Faez Iqbal Khan**, Algasan Govender, Kugen Permaul, Suren Singh, Krishna Bisetty, Journal of Theroretical Biology (under review).

- A constant pH Molecular Dynamics study of chitinase I and II isolated from *Thermomyces lanuginosus*. **Faez Iqbal Khan**, Krishna Bisetty, Kugen Permaul, Suren Singh, Journal of computational Biology and Chemistry (under review).
- *In Silico* stable mutantion prediction of thermostable chitinase obtained from *Thermomyces lanuginosus*. **Faez Iqbal Khan**, Krishna Bisetty, Kugen Permaul, Suren Singh (in preparation)
- A Molecular dynamics study of a chitinase I isolated from *Thermomyces lanuginosus*, **Faez Iqbal Khan**, Krishna Bisetty, Kugen Permaul, Suren Singh (in preparation)
- *In Silico* Mutational Analysis of Human β-glucuronidase: Enzyme responsible for Mucopolysaccharidosis VII (Sly Syndrome), **Faez Iqbal Khan**, Mohd. Shahbaaz, Md. Imtaiyaz Hassan, Krishna Bisetty and William S. Sly (in preparation).

POSTER PRESENTATION

- Experimental and computational studies of thermostable chitinases isolated from *Thermomyces lanuginosus*, February 14-17, 2015,. National Symposium on Biophysics & Golden Jubilee Meeting of the Indian Biophysical Society, Faez Iqbal Khan, Mohd. Shahbaaz, Suren Singh, Kugen Permaul, Md. Imtaiyaz Hassan, Krishna Bisetty. Jamia Millia Islamia, New Delhi, India.
- Molecular Dynamics simulations study of a chitinase, National Conference on Recent Trends in Protein Structural Biology, Faez Iqbal Khan, Kugen Permaul, Suren Singh, Md. Imtaiyaz Hassan, Krishna Bisetty December 18, 2013, Jamia Millia Islamia, New Delhi, India -25.

- Functional Annotation of Conserved Hypothetical Proteins from *Mycoplasma pneumoniae* 309 using computational methods. Mohd. Shahbaaz, Faez Iqbal Khan, Md. Imtaiyaz Hassan, Faizan Ahmad and Krishna Bisetty. Center for High performance computing (CHPC) conference, National Meeting 1 - 5 December 2014, Kruger National Park, South Africa.

CHAPTER 1

INTRODUCTION

Chitinases are a class of enzymes that hydrolyze the naturally occurring polysaccharide chitins, a linear polymer of β (1,4) N-acetyl β-D-glucosamine (Henrissat, 1999). Chitinases are ubiquitous and are present in all forms of life including bacteria, fungi, plants, animals and also in viruses. They form a class of highly conserved enzymes and exhibit diverse functions in these organisms. Chitinases possess an extraordinary ability to hydrolyze the highly insoluble chitin polymer directly to the lower molecular weight chitooligomers, which are widely used in agricultural, biotechnological, industrial and medical fields. In recent years chitinase has been the key focus of research owing to its important biophysiological functions and applications. They play a crucial role in the defense of plants against chitin-containing pathogens and also serves as an effective biocontrol agent against phytopathogens. Fungal chitinases primarily belong to the family of 18 glycosyl hydrolases (Henrissat B., 1999) and displays a high amino acid homology with class III plant chitinases (Hayes et al., 1994). The family of 18 fungal chitinases comprises of 5 domains viz., catalytic domain, N-terminal signal peptide region, chitin-binding domain, serine/threonine rich-region, and C-terminal extension region. The signal peptide, predicted at the N-terminus is an indicator of the protein secretion that targets the protein outside the cells through secretory pathways. The catalytic domain is the most important domain of this enzyme which is responsible for the chitin substrate hydrolysis. A comparison of the amino acid sequences of chitinases revealed two highly conserved motifs in the family of GH18 chitinases: $D_{XX}D_XD_XE$ and S_XGG, corresponding to the catalytic domains and substrate-binding sites respectively (Henrissat and Bairoch, 1996). The residues Glu (E) and Asp (D) are highly conserved within the catalytic domains of chitinases, indicating their direct involvement in the

hydrolysis of the glycosidic bond.

In fungi, chitinases have a diverse role in nutrition, morphogenesis, defense and development processes. Other significant applications of fungal chitinases include the possibility for improvement of plant resistance using genetic manipulation techniques. Fungal chitinases are also employed in insect control. Chitinases have several industrial and agricultural applications. The ability of chitinases to degrade makes them valuable for pollution treatment, pest and mosquito control and for biocontrol applications. In addition, chitinases play an important role in fungal protoplast generation, single cell protein production, preparation of bioactive chitooligosaccharides and in the degradation of shellfish waste.

The present study focuses mainly on a fungal chitinase obtained from *Thermomyces lanuginosus*, a thermophilic Deuteromycete that grows between 293.15 K– 333.15 K, with an optimum growth temperature of 323.15 K (Singh et al., 2003), and optimum growth pH of 6.5. Some of them are thermostable and very suitable for the industrial applications.

In this study, the chitinase gene from *Thermomyces lanuginosus* was isolated, successfully amplified by PCR using cDNA as template and purified from agarose. The gene was successfully cloned into the *E. coli* BL21 expression vector, pET21c and the *P. pastoris* GS115 expression vector, pBGP1 which was confirmed by colony PCR and enzymatic digestion. Expression was confirmed by clear zone in the chitinolytic plate assay. Structure predictions of the chitinases were performed using various algorithms of "Homology modelling" and "*ab initio*" methods. The models generated were further refined using the molecular dynamics (MD) simulations implemented in GROMACS (Van Der Spoel et al., 2005). Active sites in the protein were predicted by an assay of bioinformatic tools as well as by comparing the amino acids

sequence with other similar proteins. Analysis of the active sites was further confirmed by molecular docking studies using CDOCKER of Discovery Studio. The MD simulations of chitinase showed that chitinase was able to maintain its stability at temperatures ranging from 300-350K, comparable to experimental results. A constant pH molecular dynamics simulation at a pH range 2-6 was performed. The purpose of this study was to establish the optimum activity and stability profiles of chitinase I and chitinase II. For this purpose, the Molecular Dynamics simulations were carried out at fixed protonation states in an explicit water environment to evaluate the effect of the physiological pH on chitinase I and II enzymes obtained from *T. lanuginosus*. The results suggest a strong conformational pH dependence of chitinases. These enzymes retained their characteristic TIM Barrel fold at the respective protonated conditions, thus validated the experimental outcomes. The effect of amino acid mutations at a particular location can be utilized to observe stability change in the protein. In this study, different stability and flexibility predictions were used to assess the relation of point mutation and protein stability. The generation of mutation induced models by using various *in silico* techniques was used to predict the stability in the absence of experimental 3D structure. The results showed coherent patterns in all predictions. These results can be further validated using experimental techniques. The overall work flow of current research are represented in Figure 1.1.

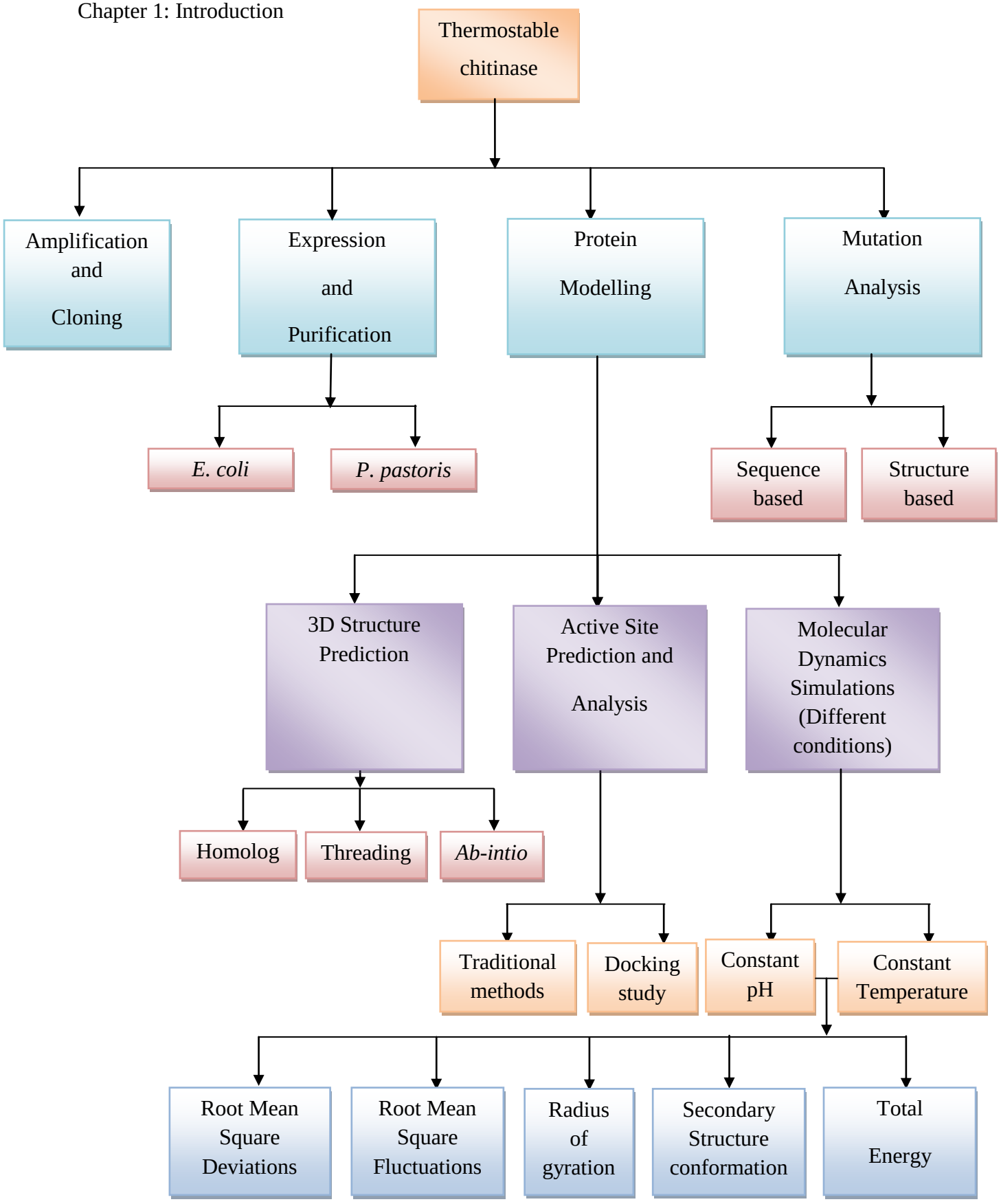


Figure 1.1: The overall work flow of the research methodology

CHAPTER 2

LITERATURE REVIEW

2.1 Chitin

The term ‘Chitin’ originated from a Greek word meaning tunic, a kind of clothing worn in ancient Greece, was first discovered by Henri Braconnot in mushroom in 1811 (Braconnot 1811). Chitin is one of the most abundant amino-polysaccharides present in animals, primarily in the shells of molluscs, crustaceans and insects where it forms a key constituent of their exoskeleton. It is also found in plants such as algae and in the structural membranes and cell walls of fungal mycelia (Mathur & Narang, 1990).

Chitin has traditionally been considered a structural material and has been less important than other functional natural polymers such as proteins and nucleic acids. However, in recent years, with a better understanding of its biological and physiological properties, chitin has emerged as an essential polymer and is being used in a vast array of different fields such as pharmaceutical, cosmetic, biomedical and food industries.

2.1.1 Chemical structure

Chitin is an unbranched polymer of 2-acetamido-2-deoxy-D-glucopyranose. It has a structural and functional resemblance to cellulose and may be considered as a derivative of cellulose with an acetamido group at carbon 2 (**Figure 2.1**).

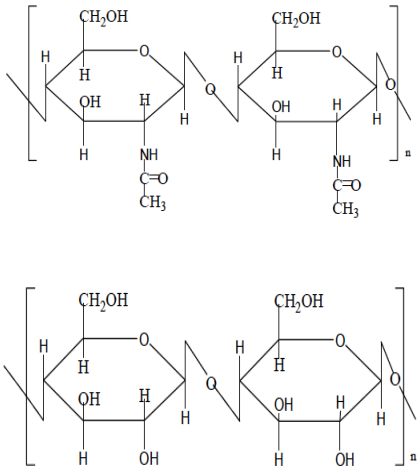


Figure 2.1 Primary structures of chitin and cellulose. Top- Chitin and Bottom- Cellulose.

Both polymers primarily aid in supporting cell and body surfaces, thus functioning as structural components, while the cellulose provides strength to the cell wall of plants, chitin accounts for the mechanical strength of arthropods exoskeletons and fungal cell walls (Gooday, 1990). The idealized chemical structure of chitin is regarded as a homopolymer but in nature it rarely exists in this form. Naturally occurring chitin possesses limited de-**N**-acetylation, owing to the activity of chitin deacetylases in chitin-containing organisms, such that the fraction of acetylated units (FA) ranges between 0.9 and 1.0.

2.1.2 Chitin in nature

Chitin is the second most abundant biopolymer and the most abundant amino-polysaccharide in nature. In animals, chitin mainly exists in the shells of mollusks and crustaceans, in the backbone of squids and in the cuticle of insects. Long chain chitin molecules are associated with proteins by covalent bonds and together they form a complex structural network. Calcium carbonate are deposited into the network contributing to strength of the shells and protection of the organisms

(Felse, A. P.; Panda, T., 1999). In fungi, chitin exists in the cell wall of spores and hyphae. It is associated with glucan molecules in the form of microfibrils, which are embedded in an amorphous matrix thus providing the framework in cell wall morphology (Kumar, N. V. R. M.). The chitin crystal structure, comprises of individual chains associated by C=O and H-N-groups forming hydrogen-bonded sheets. In addition, each chain involves intramolecular hydrogen bonding between the carbonyl group and the hydroxyl group on C-6 of neighbouring sugar rings. Also another hydrogen bond exists between the ring oxygen and the OH-group on C-3, similar to that observed in cellulose (Minke and Blackwell, 1978) and this extensive hydrogen bonding accounts for the stiffness of the chitin chain (**Figure 2.2**).

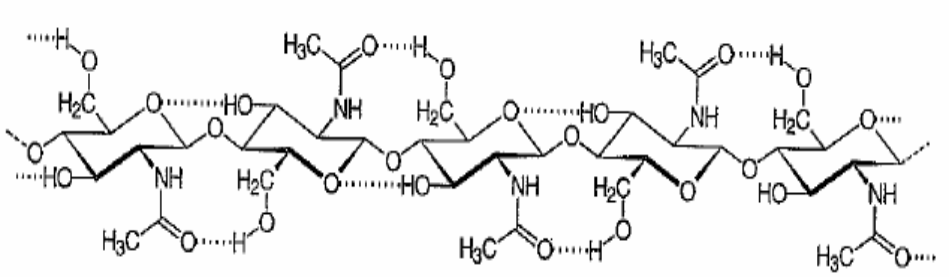


Figure 2.2 Chemical structure of chitin depicting its intramolecular hydrogen bonds (dotted lines) between the neigbouring sugar rings in chitin (Smidsrod and Moe, 1995)

Chitin possesses a highly ordered, crystalline structure and exists in three polymorphic forms namely α-, β- and γ-chitin (Hackman, 1954) (Hackman and Goldberg, 1965). These variants differ in packing, size of the unit cell and in the number of chitin chains per unit cell, their degree of hydration and polarities of adjacent chitin chains in successive sheets (Aam et al., 2010, Chen et al., 2010) as shown in Figure 2.3.

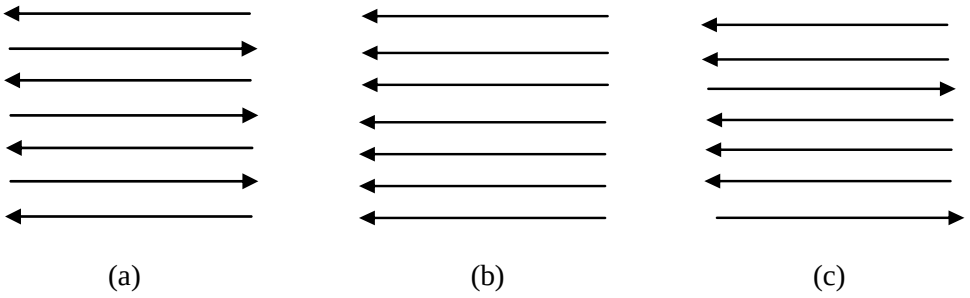


Figure 2.3 Schematic representation of the three polymorphic forms of chitin (a) α-chitin, (b) β-chitin and (c) γ-chitin.

α-chitin is the most predominant form and is isolated from the exoskeletons of crustaceans, particularly shrimps and crabs. It is also found in fungal and yeast cell walls as well as in insect cuticle. It is the most crystalline and compact of all forms comprising of individual chains arranged in an antiparallel fashion (Carlstrom, 1957). β-chitin comprises of parallel chains and is a more reactive form possessing a higher affinity for solvents (Sannan T et al., 1976). In γ-chitin two out of three chitin chains are arranged in parallel fashion while the third chain is oriented in the opposite direction. The γ-chitin has been considered to be a combination of α and β structures, rather than a different crystalline form. α-Chitin is mostly associated with extreme hardness (Rudall & Kenching, 1973) while β- and γ-chitin account for toughness, flexibility and motility and may have physiological functions other than support (Muzzarelli, 1977). The extensive intermolecular hydrogen bonding explains the inability of α-chitin to swell upon soaking in water (Minke and Blackwell, 1978) whereas β-chitin readily swells in water as it lacks these interchain hydrogen bonds (Blackwell, 1969).

2.1.3 Industrial production of chitin

The annual biosynthesis of chitin is estimated to reach 100 billion tons (Synowiecki and Al-

Khateeb, 2003). The cuticles of various crustaceans, majorly crabs and shrimps obtained as a by product in the seafood industry, serve as the major source of raw material for industrial scale production of chitin. Crustacean chitin is naturally closely associated with proteins, lipids, minerals and pigments and the extraction process involves three basic steps: demineralization for removal of calcium carbonate; deproteinization for removal of protein; and decolouration for removal of pigments (Teng et al., 2001). Demineralization is generally carried out with 0.275 – 2 M HCl at a temperature of 273 K - 373.15 K for 1 - 48 h (Khor, E., 2001). Deproteinization is performed using 1 M NaOH for 1 - 72 h at 338.15 K – 373.15 K (Ashford, N. A.; Hattis, D, and Murray, A. E., 1977) and decolouration is attained by ethanol, acetone, or hydrogen peroxide (Khor, E., 2001).

2.1.4 Production from fungal sources

In the recent years, chitin production from fungal mycelium has received increased attention due to significant advantages over the current source namely seafood waste. Seafood waste supplies are limited by seasons. The sites of fishing industry whereas fungal mycelium can be obtained easily by fermentation that possesses no geographic or seasonal limitations (Cano-Canchola et al., 1992). The crustacean chitin may vary in the physico-chemical properties, while fungal chitin have relatively consistent properties because of the controlled fermentation conditions (Novaes-Ledieu et al., 1987). The fungal mycelia contain lesser inorganic materials compared to crustacean wastes thus requiring no demineralization treatment during the processing (Calonje et al., 2000). Apparently fungal chitin is more effective in inducing the plant immune response and is potentially a better candidate for agricultural applications (Hammond, B. W. J 1979).

2.1.5 Applications of chitin

Chitin possesses low toxicity and is an inert polymer. It is biorenewable, biodegradable, owing to the ubiquitous occurence of chitinases in a broad range of taxons. Chitin, both in its native as well as modified forms, exhibits wide range of potential application in food industry, biotechnology, agriculture, drugs and pharmaceuticals, environmental science and in gene therapy as illustrated in Figure 2.4 below.

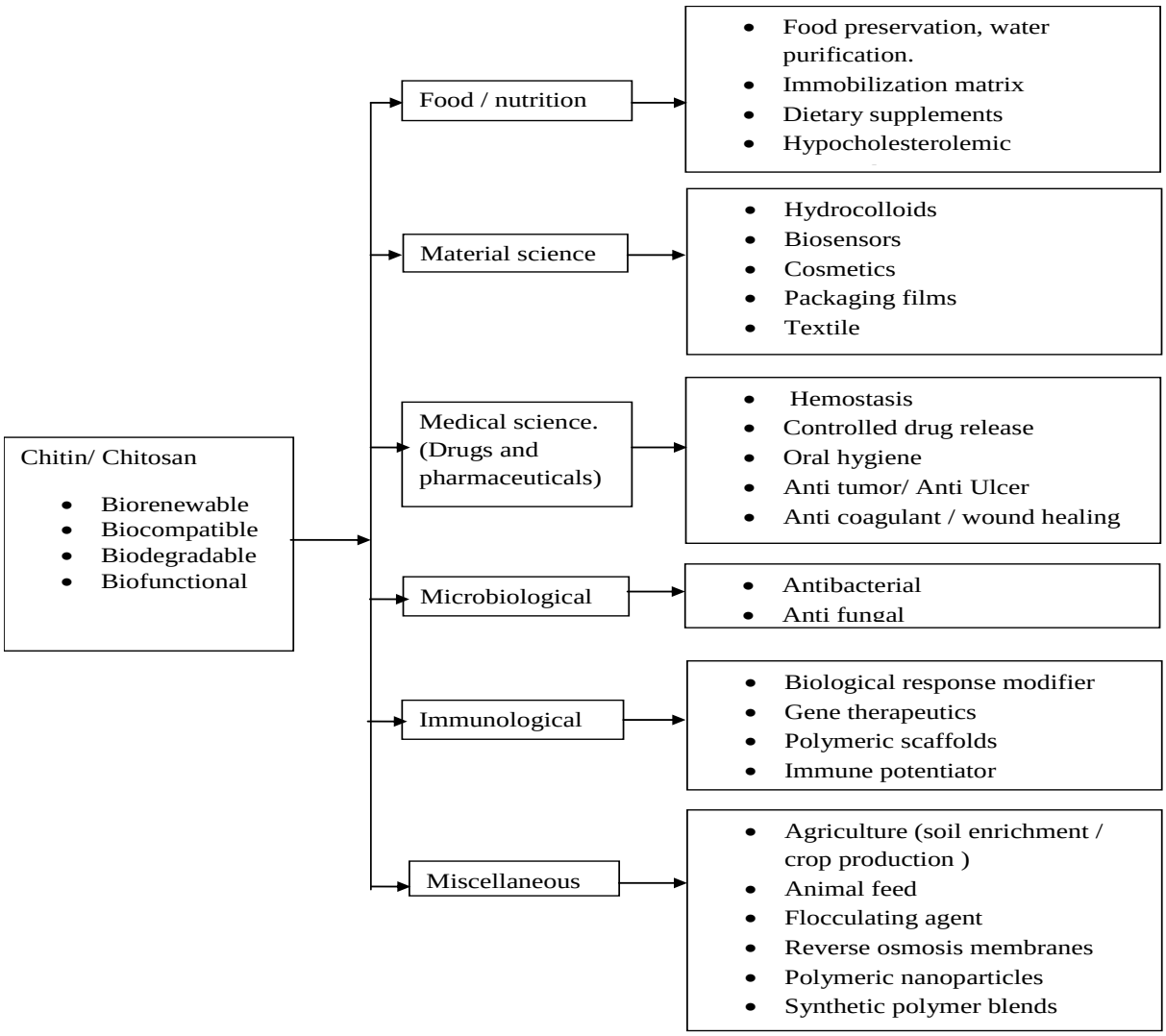


Figure 2.4 schematic representations of the wide range of applications of chitin and its derivates.

2.1.5.1 Enzyme immobilization

Enzyme immobilization has enormous practical utility owing to enhanced catalytic potential, enhanced pH and temperature resistance, and prolonged reusability. Chitin and its derivatives provide an excellent base material for immobilization of numerous carbohydrate degrading enzymes, due to an increased thermostability as compared to the free enzyme. Enzyme immobilization finds its applications in the food industry for fruit juice clarification and milk processing when α - and β -amylases are grafted on chitin (Krajewska B 2004).

2.1.5.2 Antioxidant property

The demand for potent, naturally derived antioxidant molecules has seen an exponential increase over those of synthetic origin in the past years. Chitin/chitosan and derivatives, being safe and non-toxic provide protection against highly reactive free radicals, thus inhibiting the advancement of numerous chronic diseases (Tiwari, 2004). The molecular weight and viscosity of chitosan effects its antioxidant property (Jeon et al., 2002)(Xie et al., 2001), where the low viscosity chitosan exhibited the greatest antioxidative effect (Chao et al., 2004).

2.1.5.3 Wound healing property

Wound healing is a process associated with rapid dermal regeneration. For this purpose, a novel chitosan membrane has been designed comprising of macroporous sponge-like sublayer with a top layer of skin surface. This membrane exhibited controlled evaporative water loss, excellent oxygen permeability, and promoted fluid drainage ability, effectively inhibiting invasion of exogenous microorganisms at the same time (Mi et al., 2001). Chitosan-based wound dressing led to quick wound healing by enhanced epithelialization rate and well organized dermal

deposition of collagen. Also it resulted in reduced scar tissue (fibroplasias) by preventing fibrin formation in wounds and formed a protective film coating (Kennedy, Lloyd, Paterson, & Knill, Methacanon, 1998). Chitosan degradation products were internally absorbed being a lysozyme substrate, affecting the macrophage activity.

2.1.5.4 Membranes

Of late, the paradigm shift from synthetic packaging materials to biopolymer-derived biodegradable packaging films has open up numerous opportunities for chitin/chitosan-derived packaging films. These have an advantage over the synthetic ones owing to their eco-friendly and user-friendly features, and also the raw materials are generally obtained from renewable natural resources.

2.1.5.5 Controlled drug-release

Chitosan acts as a versatile carrier for drugs and other biologically active molecules owing to its low toxicity and the abundance of free amino groups. High viscosity chitosan films displayed better sustainable release due to drug-polymer interactions. The drug release profiles of chitosan-alginate combination based drug exhibited a long period of induction followed by a rapid drug release phase in the artificial intestinal fluid. CM-chitin was modified selectively for obtaining antitumor drug conjugates (Ouchi T et al., 1992). 5-fluorouracil (choice drug for colon carcinomas) and the D-glucose analog of muramyl- L-alanyl-isoglutamine, possessing immune-adjuvant activity were grafted on CM-chitin utilizing a specific spacer and an ester bond. Sustained release of antibiotic oxytetracycline from chitosan microspheres for both oral administration and injection has been reported (Mi et al., 1997).

2.1.5.6 Antimicrobial property

In recent years, the antimicrobial property of natural biopolymers like chitin and its derivatives, has received significant attention due to the many advantages over the synthetic chemical agents used currently. Chitosan displays a broad-spectrum antimicrobial activity against bacteria, including both gram-positive and gram-negative, as well as fungi (Vishu Kumar et al., 2005). The possible mechanisms suggested for this antimicrobial function are, the interaction of the polycationic molecule with the predominantly negatively charged cell wall components (proteins and lipopolysaccharides) of the microorganism. It may result in the leaky intracellular components owing to the modifications in permeability barrier; prevention of nutrient uptake by cell; binding to DNA (upon entry into the cell), and thus inhibiting transcription and translation etc.

2.1.5.7 Anticoagulant activity

Blood coagulation is a phenomenon involving the sequential activation of a number of serine proteinases generating thrombin and subsequent thrombin-catalysed conversion of fibrinogen into insoluble fibrin. Several sulfated glycosaminoglycans find their use as anticoagulants owing to their ability to interfere with the process of blood coagulation (Vongchan et al., 2002). Sulfated chitosan, for example has been claimed to possess a high potency as an anticoagulant as compared to the well-known anti-thrombic agent, heparin.

2.1.5.8 Gene therapy

Gene therapy involves delivering a gene into a cell exogenously with the aim of combating genetic disease, impeding tumor progression and fighting viral infections. Chitin, a naturally

occurring cationic biopolymer, has emerged as an alternative nonviral gene delivery system in recent years. Chitin proves to be an excellent vector for gene delivery due to its biocompatible and non-toxic nature, thus causes no side effects even at very high doses, unlike the other nonviral vectors such as liposomes and cationic polyelectrolytes.

2.1.5.9 Miscellaneous applications

Chitin has been used in the preparation of affinity chromatography columns for isolation of lectins and determinination of their structure. Chitin and 6-O-carboxymethylchitin have been found to activate peritoneal macrophages *in vivo* and elicit nonspecific host resistance against *E. coli* infection. Chitin and chitosan derivatives are capable of adsorbing silver thiosulfate complexes, thus are also used for the treatment of industrial pollutants (Nakbanpote W, Songkroah C, and Thiravetyan P., 2004) and actinides (Kosyakov VN, Yakovlev NG and Veleshko IE., 2002). Regenerated chitin derivative fibers find their use in the paper making process as binders. Another interesting application is in treatment of periodontal bony defects as a hydroxyapatite– chitin–chitosan composite bone-filling material, which forms a self-hardening paste for guided tissue regeneration (Ito M, Matahira Y and Sakai K., 1998).

2.2. Chitinase

Chitinases are hydrolytic enzymes that cleave the β (1,4) linkage of chitin polymer down to N-acetyl β-D-glucosamine monomers (Muzzarelli, 1999). This accounts for an extraordinary performance since the substrate chitin is a highly insoluble and crystalline solid, which occurs in innumerable forms depending on the role of the living tissue. Chitinases are ubiquitously present in a wide range of taxons including bacteria, fungi, insects, plants, animals and viruses. They

form a class of highly conserved enzymes and exhibit diverse function in these organisms. Chitinase has been the key focus of research in the recent times owing to its important biophysiological functions and applications.

2.2.1 Chitinase sources

2.2.1.1 Plant chitinases

Chitinases are present in plant seeds, stems, tubers and flowers and are induced by the attack of phytopathogens as pathogenesis-related proteins in plant self-defence or by contact with elicitors such as chitooligosaccharides or growth regulators such as ethylene (Koga et al., 1997) (Gooday G. W. 1996). Their role is not only defense, but also in growth and development; this is supported by the fact that plant chitinases have been detected in the early stages of seed development. In plants, chitinases generally belong to endochitinase category and have smaller molecular weight than the insect chitinases. They found to inhibit fungal growth along with other enzymes like the β-1,3-glucanases that are present in citrus fruits, corn, potatoes, tobacco, , beans, tomatoes, yam and peas (Flach et al., 1992)(Koga D et al., 1996)(Mayer et al., 1996).

2.2.1.2 Insect chitinases

Chitinases of insects are degradative enzymes that fuction during ecdysis. These are primarily endochitinases that randomly cleave chitin present in the cuticle to chitooligosaccharides that are further hydrolized by exoenzymes to N-acetyl-glucosamine monomer, which is reused to synthesize a new cuticle. Chitinases from *Bombyx mori*, *Manduca sexta* have been extensively studied (Flach. J. et al., 1992). In addition, insect chitinases have a defensive role against their own parasites. Chitinases are largely found in crustaceans such as prawns, shrimps, and krills,

where they are induced before molting process in the integument (Flach. J. et al., 1992).

2.2.1.3 Mammalian chitinases

Until 1994, it was believed that humans do not produce chitinases due of the lack of endogenous chitin in human tissues suggesting no requirement for enzymatic turnover and restructuring of chitin molecules. However, recent studies have reported about nine mammalian chitinases or chitinase-like genes that exist in human and other mammals, all belonging to the family 18 glycosyl hydrolase (Boot et al., 2001) (Fusetti et al., 2002) (Bierbaum et al., 2005) (Lee et al., 2007a) (Seibold et al., 2009). The first chitinase in man was described by Hollak, van Weely,van Oers, and Aerts (1994) when they observed high chitinolytic activity in serum of Gaucher disease-affected patients. It was named chitotriosidase because of its ability to hydrolyze chitotriose. In the progressive years, Bleau, Massicotte, Merlen, and Boisvert (1999) and Chang et al. (2001) identified and characterized four other mammalian chitinase-like enzymes in search of other chitinases that could possibly compensate for chitotriosidase deficiency. Boot et al. (2001) discovered yet another chitinase, acidic mammalian chitinase (AMCase); which is expressed in lung macrophages, in gastric epithelia, and in the pulmonary epithelia during asthmatic inflammation and possesses high catalytic activity at pH 2 (Zhu et al., 2004).

2.2.1.4 Microbial chitinases

The bacterial producers of chitinases include *Serratia, Chromobacterium, Klebsiella, Bacillus, Streptomyces*, while the fungal producers are *Trichoderma, Oenicillium, Lecanicillium, Neurospora, Mucor, Metarhizium, Beauveria, Lycoperdon and Aspergillus*. These microorganisms produce large amounts of chitinases than animals and plants, which are of two

types namely, endochitinases and exochitinases. Chitin is not degraded intracellularly because of its size, molecular complexity, insolubility, and heterogeneous composition. The microbes extracellularly secrete enzymes with different specificity, to transform or hydrolyse chitin (Cottrell et al., 1999). Yeast and fungal chitinases participate in morphogenesis where hydrolytic cleavage of the chitin polymer is crucial for hyphal growth and branching, septum formation and spore germination.

2.2.1.5 Fungal chitinases

Fungal chitinases primarily belong to the family 18 of the glycosyl hydrolase superfamily (Henrissat, 1999) and display a high amino acid homology with class III plant chitinases (Hayes et al., 1994). The basic structure of family 18 fungal chitinases comprises of 5 domains: (1) catalytic domain, (2) N-terminal signal peptide region, (3) chitin-binding domain, (4) serine/threonine rich-region, and (5) C-terminal extension region.

The signal peptide, predicted at the N-terminus, is an indicator of protein secretion that targets the protein outside the cell via secretory pathway. During protein secretion across the cell membrane, the signal peptide is cleaved by signal peptidase releasing the mature protein outside of the cell. Chitinases lacking this domain indicate that they are intracellular proteins, and they may function during morphogenesis (Seidl et al., 2005, Takaya et al., 1998).

The catalytic domain is the most important domain of any enzyme and it is responsible for chitin substrate hydrolysis. Comparison of amino acid sequence of chitinases reveals two highly conserved motifs in family GH18 chitinase: $D_{XX}D_XD_XE$ and S_XGG, correspond to catalytic domain and substrate-binding site, respectively (Henrissat and Bairoch, 1996). Glu (E) and Asp

(D) residues are highly conserved within the catalytic domain of chitinases indicating their direct involvement in hydrolysis of the glycosidic bond.

Degradation of chitin begins when insoluble chitin binds to the chitin-binding domain. Lack of this domain has no effect on enzyme activity to soluble chitin, but activity is lost to insoluble substrate (Tjoelker et al., 2000). Insoluble chitin binds to chitin-binding domain by highly conserved aromatic residues like tryptophan, phenylalanine and/or tyrosine.

Serine/threonine-rich region acts as a linker between the catalytic domain and the chitin-binding domain and is necessary for maintenance of fungal chitinase stability. These regions of fungal chitinase are generally glycosylated with sugar chains, which help to resist hydrolysis by proteases (Arakane et al., 2003). However, most of the fungal chitinases lack the serine/threonine rich-region, chitin-binding domain, and C-terminal extension region suggesting that these domains are unnecessary for chitinase activity because naturally-occurring chitinases that lack these regions are still enzymatically active.

In general, the molecular masses of fungal chitinases range between 30 kDa-200 kDa and pH optimal range from 4 to 7. The optimal temperature for fungal chitinase activity varies between 293.15 K - 313.15 K. The enzyme activity of fungal chitinases is significantly affected by metal ions. Generally, Mg^{2+}, Ca^{2+}, Na^{+} and K^{+} act as activators of fungal chitinases , while heavy metal ions like Hg^{2+}, Ag^{+}, Fe^{2+} and Cu^{2+} have significant inhibitory effects (Duo-Chuan, 2006, Hartl et al., 2012).

In fungi, chitinases have diverse roles in nutrition, morphogenesis, defense and development processes. For example, mutation of the chitinase gene (CTS1) in *Saccharomyces cerevisiae* prevents the separation of cells after division and results in cell clumping, while exogenous expression of chitosanase and chitinase influence morphogenesis in the *Schizosaccharomyces*

pombe (Shimono et al., 2002).

Reports suggest a role of fungal cell wall chitinases in filamentous fungal sporulation since the chitinase inhibitors demethylallosamidin or allosamidin resulted in inhibition of fragmentation of hyphae into arthroconidia (Sandor et al., 1998) (Yamanaka S et al. 1994). *Trichoderma spp*. are considered as powerful biocontrol agents against soil borne fungal pathogens amongst other chitinolytic fungi and bacteria (Lorito et al., 1993). Other reports emphasize on the mycoparasitic effect of *Trichoderma* on fungi, which results in induced resistance and increased development in plants (Bolar et al., 2000) (Yedidia I et al., 2000). Other significant applications of fungal chitinases include the possibility for improvement of plant resistance using genetic manipulation techniques. For example, the chi42 gene of *T. harzianum* encodes an endochitinase possessing a much stronger anti-fungal activity against a number of phytopathogenic fungi, and is constitutively expressed in apple, tobacco, and potato (Bolar et al., 2000, Lorito et al., 1998). Fungal chitinases are also employed in insect control. A through phylogenetic study of chitinases from sequenced fungal genomes classified them into three different subgroups- A, B and C as shown in figure 2.5 (Seidl et al., 2005).

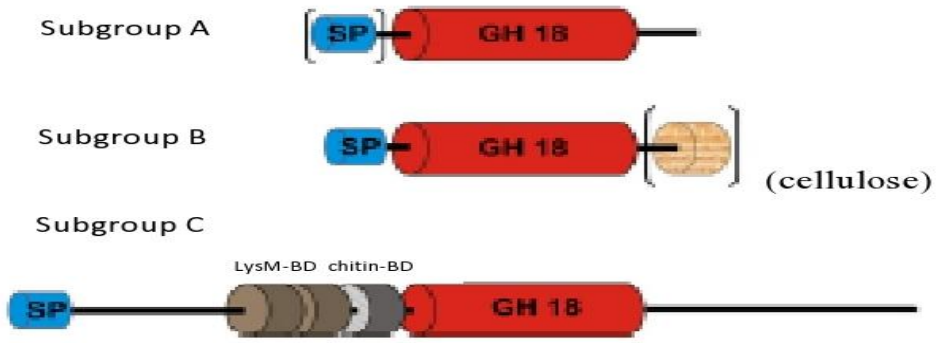


Figure 2.5 Domain organization of fungal chitinases. SP, signal peptide. GH 18, glycoside hydrolase family 18, BD, binding domain

Subgroups A and B included all previously recognised fungal chitinases, while subgroup C comprised of a novel group of high molecular weight chitinases. Subgroup A chitinases has an average molecular mass of 40–50 kDa and possesses a catalytic domain, but lack CBMs. Also they include highly conserved and profusely expressed chitinases of the fungal kingdom. Several subgroup A chitinases are extracellular proteins expressing an N-terminal signal peptide, which guides them to the secretory pathway while others either posses an ER-targeting sequence or are predicted to be intracellular proteins.

Subgroup B chitinases are highly variable in their domain structure as well as size and their molecular masses range between 30–90 kDa. A feature that distinguishes subgroup B chitinases from other members is that frequently a CBM is present.

The number of subgroup B members hugely varies among different fungal species. Most ascomycetes (Mg. grisea,. N. crassa, A. niger, A. nidulans, Mycosphaerella fijiensis, Myc. graminicola and Coccidoides immitis) and basidiomycetes (Postia placenta, Phanerochaete chrysosporium, Laccaria bicolour) are characterized by the presence of only 2–3 ORFs encoding subgroup B chitinases. Subgroup C comprises of a novel category of fungal chitinases, their molecular masses ranging between 140–170 kDa. They possess a characteristic N-terminal signal peptide that targets them to the secretory pathway. Presence of a chitin-binding domain (CBM 18) and several LysM-motifs clearly distinguishes them from other fungal chitinases. Interestingly, the unique domain architecture of subgroup C chitinase members distinguishes it from subgroups A and B chitinases indicating towards a specialized and distinctive role of these enzymes in chitin hydrolysis.

2.2.2 Classification of chitinases

On the basis of the amino acid sequence similarity, chitinases are grouped into glycosyl hydrolase families 18 and 19 (Henrissat and Bairoch, 1993). These two families differ in their sequence similarity and display different three dimensional structures suggesting their evolution from different ancestors. Chitinases form a unique group of enzymes that can efficiently hydrolyze polymeric chitin by degrading it into chitooligosaccharides with a minimum chain length of n=2. The family 18 chitinases possess a characteristic (α/β) 8-barrel domain, comprising of eight α-helices and eight β-strands. They are widely distributed in organisms ranging from bacteria, fungi, viruses, insects, plants and animals. The family 19 chitinases are similar to lysozyme and chitosanase and are found in plants and some *Streptomyces* strains. Their catalytic domains possess high α-helical content. Substrate assisted catalysis is the preferred catalytic mechanism of family 18 chitinases, while a general acid and base mechanism has been reported in the case of family 19 chitinases (Sasaki, C. et al., 2002). The family 18 chitinases are sensitive to a potent chitinase inhibitor, allosamidin while family 19 chitinases are not (Koga, D. et al., 1987). Family 18 chitinases are capable of cleaving GlcNAc-GlcNAc and GlcNAc-GlcN linkages, whereas members of family 19 chitinases cleave GlcNAc-GlcNAc and GlcN-GlcNAc (Watanabe et al., 1999).

The chitinases are also classified according to N-terminal sequence, isoelectric pH, localization of the enzyme, the inducers and presence or absence of signal peptide into different classes: the Class I chitinases are reported in plants, whereas Class II enzymes are found in bacteria, fungi and plants. Class III includes endochitinases that have a shallow and open active site architecture (Terwisscha van Scheltinga et al., 1996) and do not display any sequence similarity to Class I or II enzymes. Class IV chitinases exhibit similar characteristics to that of Class I chitinases but

they are significantly smaller than Class I chitinases (Patil R. et al., 2004) (Patil et al., 2000). Class V chitinases are exochitinases comprising a deep and tunnelshaped active site (van Aalten et al., 2000, Bortone et al., 2002) and have been detected in studies on symbiotic plant–microbe interactions (Salzer et al., 2004).

The enzymatic hydrolysis of chitin is carried out by a chitinolytic system classified into: endo-chitinases (EC 3.2.1.1.4), exo-chitinases or chitobiosidases (EC 3.2.1.29), and N-acetyl-β-glucosaminidases (EC 3.2.1.52) (**Figure 2.6**).

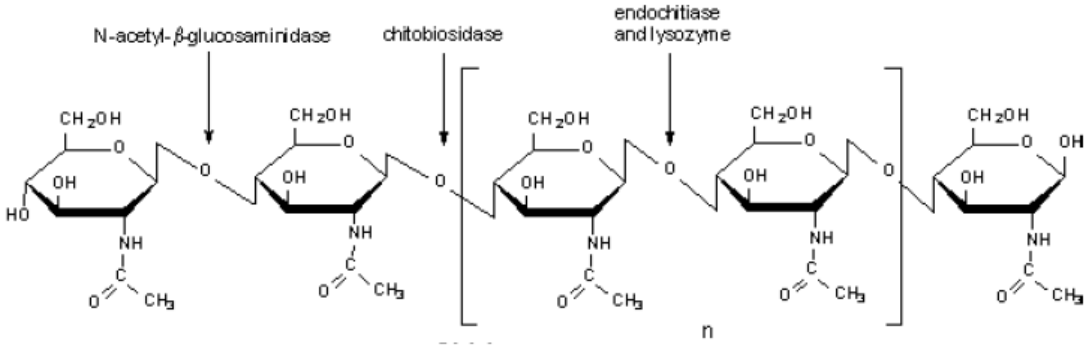


Figure 2.6. Schematic representation of action of chitinases on chitin (www.sigmaaldrich.com)

Endochitinases cleave along the internal chain of chitin randomly, generating low molecular oligomers of N-acetyl glucosamine and eventually giving diacetylchitobiose as the predominant product. On the contrary, exochitinases hydrolyze terminal non-reducing sugars releasing diacetylchitobiose without production of N-acetyl glucosamine or oligomers. N-acetyl-β-glucosaminidases remove sugar residues from the non-reducing end of the chitin chain and catalyse the breakdown of diacetylchitobiose as well as chitotriose and chitotetraose into N-acetyl glucosamine monomers.

2.2.3 Industrial usage of chitinases

Chitinases have several industrial and agricultural applications. The ability of chitinases to degrade chitin makes them valuable for pollution treatment, pest and mosquito control and for biocontrol applications. In addition, chitinases play an important role in fungal protoplast generation, single cell protein production, preparation of bioactive chitooligosaccharides and in the degradation of shellfish waste. Besides, chitinases can be used in human health care too, in fungal disease therapy and as additives in antifungal creams (Dahiya et al., 2006).

2.2.3.1 Fungal protoplast generation

Fungal protoplasts are used to study cell wall synthesis, enzyme synthesis and secretion or strain improvement (Dahiya et al, 2006). Since fungi contain chitin in their cell walls, chitinases can be used in fungal protoplast formation. Dahiya et al. (2005) produced chitinase in submerged and solid-state fermentation with *Enterobacter sp.* NRG4, which was effective in release of protoplasts from *Pleurotus florida, Trichoderma reesei, Agaricus bisporus* and *Aspergillus niger*.

2.2.3.2 Single cell protein production

Single cell proteins (SCPs) are proteins produced from the culture of a single celled microorganism, and can be used as food or feed supplement or substitute. Using chitinous waste as a substrate, chitinolytic enzymes can be applied in degradation of chitin and production of SCPs (Dahiya et al, 2006).

2.2.3.3 Production of chitooligosaccharides

Chitooligosaccharides, N-acetyl glucosamines and glucosamines have a tremendous

pharmaceutical potential and can be used in human medicines. Chitohexose and chitoheptose, for example, have antitumor activity (Dahiya et al, 2006); glucosamine is commonly used for osteoarthritis treatment, while N-acetyl glucosamine can be used as a nutritional supplement, where it is important for intestinal function. Chitooligosaccharides and N-acetyl glucosamine are either produced as a result of acid hydrolysis of chitin or chemical acetylation of glucosamine and its oligomers. Recently, production of N-acetyl glucosamine has been reported by enzymes derived from *Serratia marcescens, Bacillus huringiensis subsp. pakistani, Aeromonas hydrophila H2330, Trichoderma viride, Acremonium cellulolyticus and Aeromonas sp.* (Jung et al, 2007).

2.2.3.4 Biocontrol

Biocontrol is an alternative method in modern agriculture for reducing the dependence on chemical fungicides that may cause environmental pollution and produce resistant strains. Since chitin is the major cell wall component of plant pathogens, like *Botrytis cinerea, Rhizoctonia solani, Sclerotinia sclerotiorum or Fusarium sp*., chitinases can be used as biocontrol agents as they are considered to be crucial enzymes in mycoparasitism.

2.3 *Thermomyces lanuginosus*

Thermomyces lanuginosus was first isolated in 1899 by Tsiklinskaya. it is a thermophilic Deuteromycete(unicellular or septate and reproduces asexually by forming aleurioconidia) that grows between 293.15 K – 333.15 K with an optimum growth temperature of 323.15 K (Singh et al., 2003). The optimum growth pH of most strains of *T. lanuginosus* is 6.5. It is reported to be non-cellulytic and probably grows commensally with celluloytic fungi by utilizing the sugar generated by these fungi and possibly also by using their mycelial breakdown products (Singh et

al., 2000c). *T. lanuginosus* secretes a variety of enzymes. Some of them are thermostable and very suitable for the industrial applications (**Table 2.1**)

Table 2.1 Enzymes from *Thermomyces lanuginosus* with their optimum pH and temperature.

Enzyme	Optimal pH	Optimal temperature	Reference
Lipase	8.0	333.15 K	Arima et al., (1972)
β-galactosidase	6.7 to 7.2,	-	(Fischer et al., 1995)
Invertase	6.5	333.15 K	Chaudhuri and Maheshwari (1996)
Protease	5·0 and 9·0	343.15 K	Li et al. (1997)
Trehalase	5.0	323.15 K	Bharadwaj and Maheshwari (1999)
Hemicellulases	6.0	323.15 K	Singh et al. (2000b)
Glucoamylase	6.0	343.15 K	Odibo and Ulbrich-Hofmann (2001)
A-amylase	5.0	333.15 K	Odibo and Ulbrich-Hofmann (2001)
Phytase	5.0	343.15 K	Gulati et al. (2007)
α-galactosidase	5-5.5	333.15 K	Rezessy-Szabo et al. (2007)
Xylanases	6.0-7.0	333.15 K -348.15 K	Khucharoenphaisan and Sinma (2010)

2.4 Scope of the study

Thermophilic microorganisms like *T. lanuginosus* produce thermostable enzymes that are able to survive high temperatures as well as extreme pH condition (Haki and Rakshit, 2003). These thermostable enzymes include xylanases, lipases, amylases and chitinases that exhibit very high

specificity and efficiency even at extreme conditions and therefore have substantial potential regarding many industrial purposes (Haki and Rakshit, 2003). The utility of these enzymes in the paper, pharmacy and food industries are extensively studied (Haki and Rakshit, 2003). Xylanase was produced by the cloning of rapid-amplifying mesophiles using recombinant DNA technology (Stephens et al., 2014). In industry, xylanases have been used in the process of pulp-bleaching for reducing the chlorine chemicals (Gruber et al., 1998). Despite their superior stability at high temperatures, various efforts such as site directed mutagenesis and the introduction of extra disulfide bonds have been made to increase the thermostability (Gruber et al., 1998). Various alkali stable and thermostable mutants of the xylanases were created using random mutagenesis and DNA shuffling techniques (Stephens et al., 2014). Among different thermostable enzymes, chitinase from *T. lanuginosus* possesses a vast array of applications in diverse fields like food industry, medicine, biotechnology and agriculture. However, higher expenses and lower yield of industrial production, restricts the large scale applications of chitinase (Kadokura et al., 2007). Recently, there has been a growing interest in increasing the production of chitinase using recombination DNA technology as it provides a better option for higher yields, cost efficient, and easy enzyme production (Adrio and Demain, 2010). The 3D structure of chitinase from *T. lanuginosus* has not yet been fully characterized. In this study, MD simulations were performed to provide valuable insights into the structure and stability of the enzymes at different temperatures and pH range by analysing various thermodynamic parameters. The effects of the physiological pH on chitinases enzymes were determined. The mechanisms of the mutational effects are useful to study the nature and stability of enzymes. For this purpose, various stability and flexibility predictors were used to assess the relationship between point mutation and enzyme stability. The present work mainly focused on the isolation,

characterization, cloning and expression of chitinase. Furthermore, the experimental studies were complimented with computational modelling and MD simulations aimed at assessing the conformational profile of the thermostable chitinases. Future computational and experimental results will be key to successfully predict optimal assay conditions of the enzymes. This will guide the performing of enzyme assays in lab, thereby speeding up commercial application of enzymes. The detailed analysis of amino acid sequence in the 3D structure of enzymes will help to identify the critical residues involved in functions and stability. Furthermore, thermally fluctuating residues can be modified to generate enzymes with advanced properties.

CHAPTER 3

EXPERIMENTAL

Due to wide range of application in several fields, chitinase has received global attention in both industrial and scientific areas. However, higher expenses and lower yield of industrial chitinase production, restricts the large scale applications of chitinase (Kadokura et al., 2007). Recently, there has been growing interest in increasing the production of chitinase using recombination DNA technology. Recombinant DNA technology provides a better option for high yield, cost efficient, and easy enzyme production (Adrio and Demain, 2010). Recombinant DNA is created by molecular cloning (Carroll, 1993). Recombinant DNA formation comprises a cloning vector joined to a DNA molecule that replicates within a cell. Vectors are generally plasmids that represent relatively small DNA segments (Lodish, 2000). Vectors contain essential genetic signals for replication, additional regions for foreign DNA insertion, recognition of cells with recombinant DNA, and expression of the foreign DNA. The vector selection depends on the host organism, size of the DNA to be cloned and expression of genes (Griffiths, 1999). Restriction enzymes and DNA ligase are the main tools in making recombinant DNA molecules (Berg, 2002). Precise changes to the host organism may enhance the expression of the gene (Holmes, 1996). The basic steps of molecular cloning include the following steps. First, the DNA fragments to be cloned are generated using PCR and restriction endonucleases. Second, the fragments produced are ligated to desired vectors. Third, the recombinant DNA molecule is transported to a host cell where it produce clones by replication. The recombinant DNA in the host cells is spread to all progeny cells, thus creating a population that all carry the identical sequence. Finally, the cloned DNA segments can be recovered by plasmid isolation and purified for further analysis (Lessard, 2013). Recombinant DNA is commonly used in medicine,

biotechnology and research. Eukaryotic genes are different from prokaryotic genes. In eukaryotes, the genes coding for protein are interrupted by introns that must be spliced out of gene before the RNA is translated. This chapter deals with the cloning of chitinase gene into *E. coli* and *P. pastoris* expression vectors followed by the expression and purification of chitinase enzymes.

3.1 Material and Methods

3.1.1 Strains, plasmids and growth conditions

T. lanuginosus SSBP (Singh et al., 2000a) was grown at 323.15 K on potato dextrose medium (Merck), sub-cultured every two weeks and stored at 277.15 K. *E. coli* BL21 (Stratagene) was grown and maintained on Luria Bertani (LB) medium (10 g/l peptone, 5 g/l yeast extract powder, 10 g/l NaCl and 15 g/l technical agar). Transformed *E. coli* was maintained on LB medium supplemented with 100 µg/ml ampicillin and grown at 310.15 K. *Pichia pastoris* GS115 (Invitrogen) was maintained on YPD agar (20 g/l peptone, 20 g/l dextrose, 10 g/l yeast extract and 20 g/l agar). Transformed *P. pastoris* GS115 was maintained on YPD supplemented with 100 µg/ml zeocin (Sigma-Aldrich). For short term storage, sub-culturing was done every two weeks, whilst for long term storage cultures were supplemented with 15% glycerol and stored at 203.15 K. Plasmids used in this study were pET21c (5.4 kb) (Novagen) and pBGP1 (4.6 kb) (Lee et al., 2005), which possess an ampicillin resistance gene for selection in *E. coli* and a zeocin resistance gene for selection in *P. pastoris* GS115.

3.1.2 Total RNA isolation from *Thermomyces lanuginosus* SSBP and first-strand cDNA synthesis

One ml of an aqueous spore suspension (1×10^6 spores/ml) was inoculated into chitinase production medium [10 g/l; K_2HPO_4, 0.87 g L^{-1}; KH_2PO_4, 0.68 g L_{-1}; KCl, 0.2 g L^{-1}; NH_4NO_3, 1 g L^{-1}; $MgSO_4$, 0.2 g L^{-1}; yeast extract, 4 g L^{-1}; pH 6.5] (Guo et al., 2005) and incubated at 323.15 K on a rotary shaker at 150 rpm. Mycelia were harvested by centrifugation following 48 h incubation and stored at 203.15 K. Total RNA was isolated using the TRIzol® method and the integrity was confirmed by electrophoresis (0.8% agarose gel) and NanoDrop Spectrophotometer (Thermo Scientific) (OD260/OD280 > 1.8) analysis. The first-strand cDNA was synthesized from 5 µg total RNA using the iScriptTM cDNA synthesis kit (Biorad).

3.1.3 Primer design

Primers were designed using Oligo version 6.0 software (Rychlik, 2007), based on chitinase gene identified in the currently available annotated genome sequence data of *Thermomyces lanuginosus* SSBP (GenBank Accession no. KJ740647), (**Figure 3.1**). The first step involved using this chitinase gene sequence as an input and searching for appropriate enzymes whose restriction sites could be used for cloning. *EcoRI, XbaI, NheI and XhoI* sites were selected as these enzymes were found not to cut within the coding region of chitinase gene. Additionally, these enzymes also had unique site in the multiple cloning site (MCS) of the cloning vector pET21c. The primers were synthesized and supplied in lyophilized form by Inqaba Biotech. These were dissolved in appropriate volumes of sterilized water as indicated by the company to obtain a final concentration of 100 µM and stored at 253.15 K until further use.

ATGCCTTCCTTCAAAAGCGTTGTCTCTTCTCTGCCGGTTATCCTCACGGCTCTTCCGAGTGTGCAGGCAGGTCTGGACCTGTCCT
CGACATCAAATGTTGTTGTGTATTGGGGCCAGAACTCCGCCGCGGCGTCTGGCGGTGGCCCAAGTCAGCAGCCACTTGCCACC
TATTGTGAAGATCCGAACATTGATACTCTTGTCATGGCATTTATGACCCGGATTAACGGGGCTGGGGGCGTTCCGGAGATCAA
CCTTGCCAACATTGGTGATTCTTGCGGCACATTTGACGGTACAAACCTGAAGGACTGCCCTCAGGTCGGCGAGGATATCAAGA
AGTGCCAGAGCTTGGGTAAAACCATCCTCCTCTCTATCGGCGGTGCCACCTATACCGAAGGTGGCTTCCAGAGTGCCGAGGCC
GCCGAAGCTGGTGCCAGAATGGTTTGGGAAACTTTTGGACCCGTGACGAACGGAGATGCCCTCCGTCCGTTCGGTGACGCCG
TTGTTGATGGTTTCGACCTCGACTTTGAGGCCACTGTTTCCAACATGGTTCCGTTTGCGAACACCCTTCGCTCGCTCATGGACAG
CGATTCCAGCAAGCAGTACTTCTTGACTGCGGCTCCTCAATGCCCTTTCCCTGACGCGGCGAACAAGGAAATGCTCGATGGTGC
TGTTTCCTTTGACGCTATTTGGGTGCAGTTCTACAACAACTACTGCGGCGTCAACTCCTACCCGGACAACTTCAACTTCAATACC
TGGGACGATTGGGCACAGAACACTTCGAAGAACAAGAACGTCAAAGTTCTCGTCGGCGTGCCGGCTAACACCGGGGCTGCTG
GATCTGGATACTTGCCTGTGGATCAATTGGCTCCAGTCATTGAGCATGCCAGGACGTTCCCTAGCTTCGGAGGTGTCATGATGT
GGGATGCAAGCCAGGCTTACGCCAATGATGGATTCCTTTCTGGCATCAAGTCTATTCTGGGCAGTGTTATTTCGCGTGTTAAGC
GCATGT**TCTTCCGCCGCGACTTCTGGTGA**

Figure 3.1: Chitinase gene sequence showing primer binding sites (highlighted).

Sequences of primers used in this study, with relevant restriction sites, are as follows

CHIT_F1: 5′ CCGGAATTCATGCCTTCCTTCAAAAGCGTTGTC 3′

CHIT_R1: 5′ CCGTCTAGAACACCAGAAGTCGCGGCGGAAGA 3′

CHIT_F2: 5′ CCGGCTAGCATGCCTTCCTTCAAAAGCGTTGTC 3′

CHIT_R2: 5′ CCG CTCGAGATGCCTTCCTTCAAAAGCGTTGTC 3′

3.1.4 Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) is one of the most fundamental techniques in molecular biology, which is employed to amplify a single or a few copies of DNA across several orders of magnitude, producing thousands to millions of copies of the desired sequence. In 1984 when Kary Mullis discovered the technique, where he added native *E. coli* DNA polymerase after each cycle due to its destruction at high temperatures used for DNA denaturation, till the present era, PCR has been at the forefront of technological advancement. This has made amplification much more easier and time saving. The present method involves the use of a thermostable enzyme, the

Taq DNA polymerase, which is added at the start of the reaction. There are three major steps involved in PCR: denaturation, annealing and extension. As the name suggests, it commences with a high temperature induced denaturation of DNA into two strands, followed by annealing/ hybridization of the primers to the complementary sequences within the target gene and finally the polymerase adds nucleotides (dNTPs) complementary to the parent DNA strand sequence in 5′ to 3′ direction to extend the new daughter strand.

The optimized conditions used for amplification of chitinase II gene using the primers in a thermocycler;

Initial denaturation at 371.15 K for 30 s

Cyclic denaturation at 371.15 K for 10 s
Primer annealing at 328.15 K for 20 s } 30 cycles
Extension at 345.15 K for 30 s

Final extension at 343.15 K for 7 min.

3.1.5 Agarose gel electrophoresis

Presence of the PCR product was confirmed using 1% agarose gel electrophoresis. The desired amount of agarose was placed in an Erlenmeyer flask, together with the required amount of 1 X TAE buffer, which was diluted from a 50 X TAE stock (242 g Tris, 57.1 ml acetic acid, 100 ml 0.5 M EDTA). The agarose was then heated in a microwave for 1-2 min to form a clear solution, which was then poured into a casting tray with well combs and allowed to set. PCR products were mixed with 6X gel loading buffer in a ratio of 1:5, which was then loaded into the gel and run at 100 V for approximately 1 h. The size of the PCR product was confirmed with a DNA

molecular weight marker. Gels were then subjected to staining in ethidium bromide (0.05 mg/ml) for 20 min and destained in distilled water for a further 5 - 10 min. Stained gels were viewed on a UV transilluminator. Ethidium bromide intercalates between the bases in DNA double helix and results in a strong, UV-excitable orange fluorescence that makes the position of DNA bands visible on an agarose gel. The desired bands were excised from the gel and purified using the DNA Clean and Concentrator Kit (Zymo Research), according to manufacturer's protocol.

3.1.6 DNA quantification

DNA concentration was determined using the NanoDrop Spectrophotometer ND-1000 (Thermo Scientific). A one microlitre aliquot of the sample was loaded onto the stage, and its absorbance was measured at 260 nm. This was done in triplicate for both vector and insert and the average value was used as the DNA concentration. TE buffer was used as a blank to "zero" the spectrophotometer before samples were measured.

3.1.7 Restriction digestion and ligation

Restriction endonucleases were used for specific digestion of both the pET21c vector and the insert DNA. These cuts produce compatible sticky ends for ligation with T4 DNA ligase. Standard protocols were followed for restriction analysis (Sambrook *et al*., 1989; 2001). For digestion, 0.1 volumes of the corresponding restriction buffer (10X) was added to the DNA solution. Mixtures were incubated with the desire restriction enzymes (Fermentas) at 310.15 K overnight. Restricted DNA samples were analyzed on 1% agarose gels, as previously described. The desired bands were excised and purified using DNA Clean and Concentration Kit (Zymo Research), according to manufacturer's protocol.

For ligation of the restricted insert DNA into the vector pET21c, a molar vector: insert ratio of 1:3 was used for high ligation efficiency. The ligation reaction containing the vector, insert, ligation buffer (has ATP) and T4 DNA ligase was incubated overnight at $4^{o}C$. The ligation mixtures were then transformed into competent host cells.

3.1.8 Preparation of competent BL21 cells

Host cells were made "competent" or capable of taking up DNA from their surrounding environment, by exposing them to Ca^{2+}, which interacts with their cell envelopes. This procedure made their cell envelopes more permeable for DNA uptake. *E. coli* was cultured on LB medium. A single colony was used to inoculate 5 ml sterile LB medium and incubated at 310.15 K, 180 rpm overnight. One millilitre of this culture was used to inoculate 29 ml of fresh LB medium and shaken at 310.15 K, 180 rpm until it reached an OD of 0.375 at 590 nm. The culture was immediately placed on ice and kept cold for the duration of the procedure. The cell pellet was recovered by centrifugation at 3000 x g for 10 min and the supernatant was discarded. The cells were resuspended in 10 ml cold 100 mM $CaCl_2$ and centrifuged at the same speed and resuspended in 10 ml cold 100 mM $CaCl_2$. The entire mixture was incubated on ice for 20 min and then centrifuged. The competent cells were then resuspended in 2 ml 100 mM $CaCl_2$ containing 10% glycerol. One hundred microlitres of the prepared competent cells were dispensed into eppendorfs, incubated at 277.15 K overnight, and then stored at 203.15 K. According to the protocol followed (Ausubel *et al*., 1989), it is suggested that competent cells are most efficient when prepared 24 h prior to transformation. The preparations were therefore incubated at 277.15 K overnight to enhance the effectiveness of the subsequent transformation procedure.

3.1.9 Preparation of electro-competent *Pichia pastoris* cells

Pichia pastoris cells from glycerol stock were streaked onto the YPD plate to grow single colonies. One single colony were used to inoculate 10 ml of YPD and grown for 12-14 h with shaking at 303.15 K to an OD_{600} of 3.0. Then, 250 ml of YPD was inoculated with an aliquot of the overnight culture to reach an OD_{600} of 0.005 and grow 6 h to an OD_{600} of 1-1.3. The cells were harvested by centrifugation at 1500 x g for 5 minutes at 4^oC. Cells were gently resuspended in 250 ml ice cold milli-Q H_2O using a glass pipette and centrifuge as above. The washing steps were repeated with 125 ml ice cold milli-Q H_2O. The cells were resuspended in 20 ml of 1 M ice cold sorbitol and centrifuged as above. Finally cells were resuspended in 0.5 ml of 1 M sorbitol and kept on ice.

3.1.10 Transformation and selection

The ligated DNA cocktail was added to one aliquot of the *E. coli* BL21 competent cells and incubated on ice for 30 min. The cells were thereafter subjected to heat shock for 30 s at 315.15 K and then transferred to ice. 900 µl of fresh LB medium was added to the tube and cells were then given an outgrowth for 1 h at 310.15 K. The transformation mixtures were plated on LB agar plates containing ampicillin (100 µg/ml) and kept at 310.15 K overnight in an incubator. This served as a selection method since only those cells that had taken up the ligated plasmid were able to grow on the selective medium supplemented with ampicillin. Colonies that grew on LB-ampicillin plates were tested by colony PCR to determine the presence of chitinase gene. A positive clone was then sent for DNA sequencing.

3.1.11 Transformation of *Pichia pastoris* GS115

Once pBGP1 clones were confirmed by DNA band size through restriction and gel electrophoresis, they were cloned into electrocompetent yeast cells. One colony was inoculated into 20 ml YPD and incubated at 303.15 K, 200 rpm for 12-16 h. The culture was then transferred to 180 ml YPD and allowed a further incubation period on the shaker until an OD_{600} of 1.4 was reached. The cell pellet was then collected by centrifugation at 4000 rpm for 10 min and resuspended in sterilized distilled water. The cells were washed once again with cold distilled water followed by a final wash with 1 M sorbitol (Sigma-Aldrich).The cells were resuspended in a final volume of 2 ml 1 M sorbitol and kept on ice until the transformation. Two microliters of purified plasmid was mixed with 80 µl electrocompetent yeast cells in an eppendorf tube. The mixture was then transferred to a sterile, cold 0.2 cm electroporation cuvette (Bio-Rad) which was incubated on ice. The following electroporation conditions were used: 2000 V, 25 µF, 200 Ω. The mixture was pulsed using the Gene Pulser II Xcell (Bio-rad) using the pre-set protocol for *P. pastoris*. One millilitre of ice-cold 1M sorbitol was added to the mixture which was left on ice for 15 min. This mixture was followed by 1 ml of YPD. The mixture was incubated at 303.15 K for 1 h. One hundred microlitre aliquots of cells were plated on YPD supplemented with 100 µg/ml zeocin and incubated at 303.15 K for 3-4 days. The positive clone was used in subsequent expression and characterization studies.

3.1.12 Protein expression

A positive clone that was confirmed by sequencing was inoculated in 5 ml LB medium and incubated at 310.15 K overnight in a shaking incubator. This served as a starter culture for large

scale cultures that were used for protein expression. 5 ml of this starter culture was added to 500 ml of fresh LB medium and cells were allowed to grow till log phase. At regular intervals, cell growth was monitored by measuring optical density at 600 nm. At optical density of 0.6 (log phase), cells were induced with 0.5 mM IPTG and incubated at 310.15 K for further 3 h. The cells were then harvested by centrifugation at 7000 rpm for 5 min. Supernatant was discarded and the pellet was dissolved in lysis buffer and kept on ice for 15 min. The cell lysate was further subjected to freeze-thaw cycles and sonication to achieve complete lysis (Structural Genomics et al., 2008). The lysed cells were then centrifuged at 12000 rpm for 30 min to separate the insoluble fraction and cell debris that were pelleted. Supernatant was separated and stored at 277.15 K until further purification process.

3.1.13 Plate assay for chitinolytic enzyme activity

Glycol chitin was prepared by Trudel and Asselin (1989) and Lee et al. (2007b) methods. 1 g of glycol chitosan was dissolved in 20 ml of 10% acetic acid by vigorous stirring. The viscous solution was allowed to stand overnight at room temperature with generally stirring. Methanol (90 ml) was added slowly using Whatman no. 4 filter paper. The filtrate was transferred into a beaker and 1.25 ml of acetic anhydride was added with magnetic stirring. The gel was allowed to set for 45 min at room temperature, then transferred to a blender, cover with methanol, and homogenized for 4 min at highest speed. The resulting suspension was centrifuged for 15 min at 15,000 x g at 277.15 K. The gelatinous pellet was resuspended in about 1 volume of methanol, homogenizes, and centrifuged as in the preceding step. The pellet was resuspended in distilled water (100 ml) containing 0.02% (w/v) sodium azide and homogenized for 4 min. This was the 1% (w/v) stock solution of glycol chitin. Plate assay was accorded to Anil et al. (2007) with few

modifications. Glycol chitin was added to 0.2% of melted agarose and transfered into a petri plate. Holes were punched on agarose medium after solidification. 200 - 400 microliters of enzyme suspension was added in separate wells and incubated at 323.15 K for 2 h. Chitinase production medium served as negative control. The plate were stained with 2 % Calcofluor White M2R for 1.5 h and washed with distilled water for 2 h and observed under UV light to detect the zone of clearance around the wells.

3.2 Results and Discussions

3.2.1 Amplification and Cloning of chitinase

The chitinase gene was successfully amplified by PCR using cDNA as template which was further purified from agarose. A 1000 bp DNA ladder was used to estimate the product size. After amplification, a single band of approx 1032 bp was observed under UV light on an ethidium bromide-stained agarose gel (**Figure 3.2**).

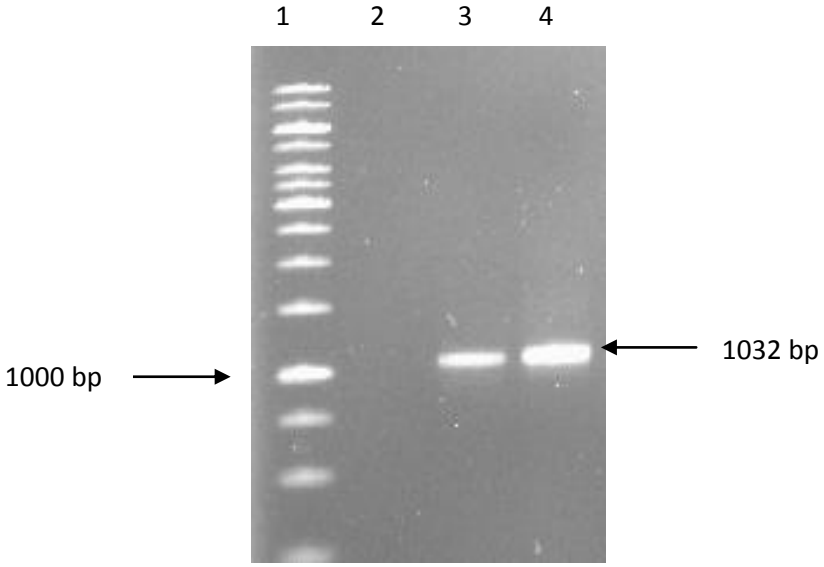


Figure 3.2: Chitinase II protein PCR product on agarose gel. Lane 1: 1 kb DNA size marker, Lane 3, 4: Single expected band of size 1032 bp.

The gene was successfully cloned into the pET21c and pBGP1 and expressed in *E. coli* BL21 and *P. pastoris* GS115, which was confirmed by enzymatic digestion, ensuring that the DNA fragments contain same size. Restriction digestions of chitinase II insert TA cloning vectors pTz (**Figure 3.3 A**), pBGP1 (**Figure 3.3 B**) and pET21c were performed (**Figure 3.3 C**)

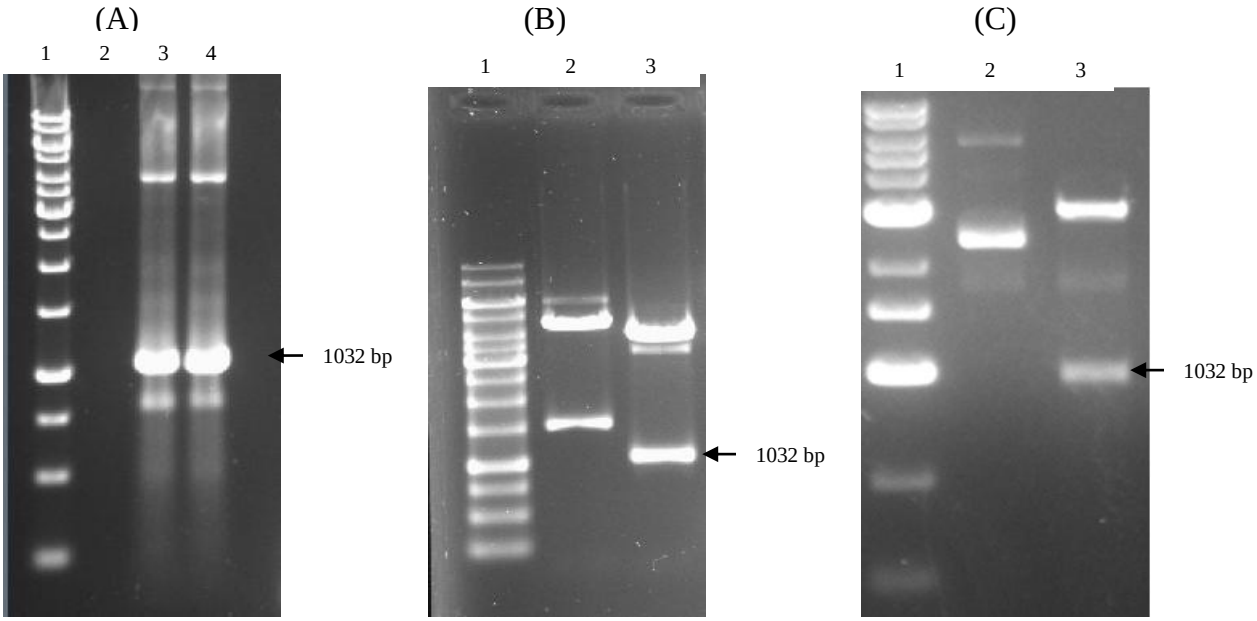


Figure 3.3 Restriction digestion of (A) pTz construct, Lane 1: 1 kb DNA size marker, lane 3, 4: expected excised DNA band of size 1032 bp (B) pBGP1construct, Lane 1: 1 kb DNA size marker, lane 3: expected excised DNA band of size 1032 bp (C) pET21c construct, Lane 1: DNA marker, lane 3: expected excised DNA band of size 1032 bp

Amplification of the plasmid insert colonies obtained from positive clone was achieved by colony PCR using the desired primers. Colony PCR was performed for chitinase gene insert into pBGP1 (**Figure 3.4**)

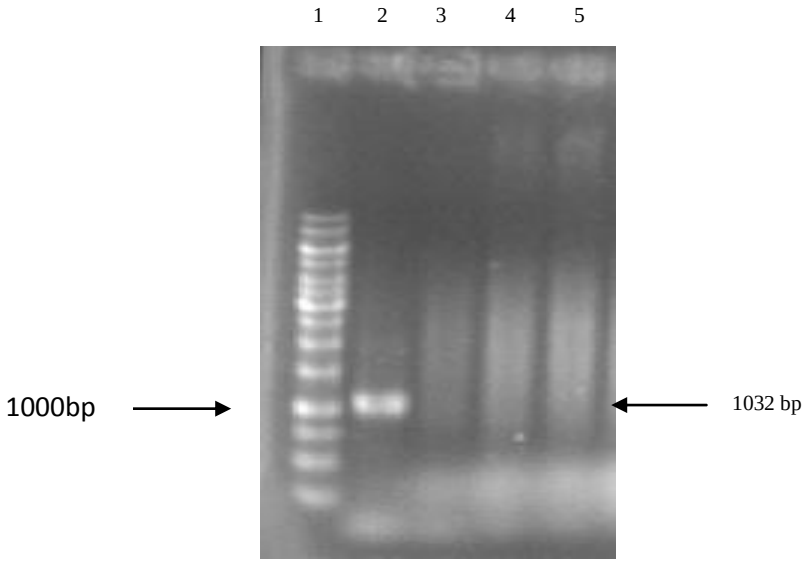


Figure 3.4 Colony PCR of positive clone of citinase II into pBGP1, Lane 1: 1 kb DNA size marker, lane 2: expected DNA band of size 1032 bp.

3.2.2 Expression of chitinase

SDS-PAGE was performed for the expression of chitinase II into *E. Coli* BL21 as shown in Figure 3.5. The predicted size of chitinase II protein was 36.6 KDa. There is only one potential N-linked glycosylation site at position 258 occupied by asparagine as predicted by NetNGlyc 1.0 server.

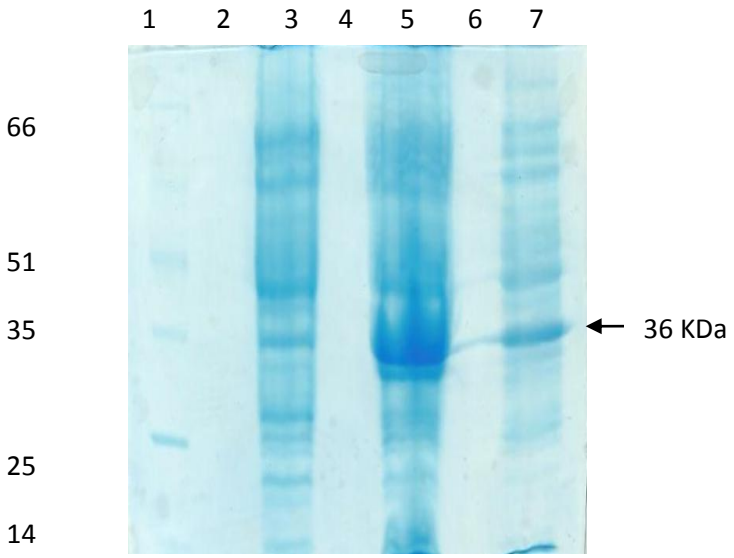


Figure 3.5 SDS-PAGE analysis of chitinase II protein stained with Coomassie. Lane 1: Protein marker, lane 5 - pellet, lane 7 -36KDa chitinase II enzyme.

The expression of chitinase II into *P. pastoris* GS115 was confirmed by clear zone in the chitinolytic plate assay as shown in Figure 3.6.

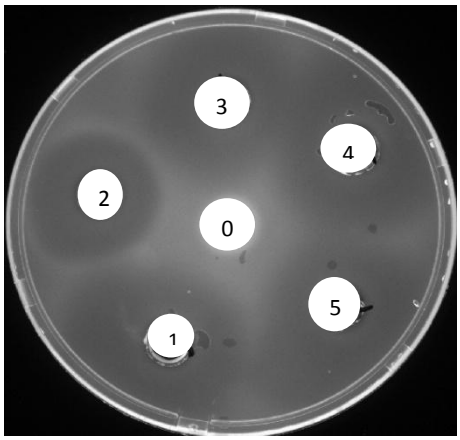


Figure 3.6 Chitinolytic enzyme plate assay (1, 2, 3 , 4 and 5) and chitinase II production medium as negative control (0).

3.3 Conclusion

The chitinase II gene was successfully amplified by PCR using cDNA as a template purified from agarose. The size of the product was 1032 bp and the gene was successfully cloned into the *E. coli* BL21 expression vector, pET21c and the *P. pastoris* GS115 expression vector, pBGP1 which was confirmed by colony PCR and enzymatic digestion. The expected size of the chitinase II was 36.6 KDa. There is only one potential N-linked glycosylation site at position 258 occupied by asparagine as predicted by the NetNGlyc 1.0 server. Expressions were confirmed by the clear zone in the chitinolytic plate assay. The 3D structure of chitinase obtained from *T. lanuginosus* has not been characterized to date. However, to better understand the structural features of chitinase II, we obtained a 3D model with a higher precision by performing homology modelling and *ab initio* methods as discussed in next chapter.

CHAPTER 4

COMPUTATIONAL

This chapter deals with the 3D structure prediction of chitinases using Homology modelling followed by an active site analysis using sequence based and structure based approaches. Docking studies were performed with one of its inhibitors to confirm the binding sites. Molecular Dynamics simulations provided valuable insights into the structure and stability of the enzymes. A 10 ns MD simulation of chitinases were performed at different temperatures ranging from 300 K-375 K to check the thermostable nature of the enzymes and by analysing various thermodynamic parameters. A constant pH molecular dynamics simulation at a pH range 2-6 was performed. The purpose of this study was to establish the optimum activity and stability profiles of chitinase I and II. The MD simulations were carried out at fixed protonation states in an explicit water environment to evaluate the effect of the physiological pH on chitinases enzymes. The mechanisms of mutational effects are useful to study the nature of enzymes and to generate the stable mutants. For this purpose, various stability and flexibility predictors were used to assess the relationship between point mutation and enzyme stability.

4.1 Structure prediction and analysis

Structure predictions of chitinase I and II were performed using various algorithms of "Homology modelling" and "*ab initio*" methods. The aim of homology or comparative protein modelling is to construct a probable three-dimensional (3D) replica for the protein with an unknown experimental structure (i.e. the target) by means of sequence similarities to that of a

known structure (i.e. the templates) present in databases (John and Sali, 2003). However, here two criteria must be achieved for reliable model building. First, there must be a detectable likeness between the structure of the template and the target of the protein sequence. Secondly, a significant correct alignment between the structure of the template and the target sequence must be computed. The homology modelling is achieved by minor changes in the sequence of a given protein with slight changes in the 3D structure (Chothia and Lesk, 1986). The comparative or homology modelling of the protein 3D structure continues to be the most precise method in contrast to other available methods such as *ab initio* methods (Koehl and Levitt, 1999). The overall accuracy range is achieved for the comparative models, from the minimal resolution protein models along with an accurate fold in the structure to a more precise model obtained from nuclear magnetic resonance (NMR) spectroscopy or X-ray crystallography (Sanchez and Sali, 1997). Sometimes models with lower resolutions are helpful in function predictions since certain attributes can only be inferred from the crude approximation of the structural elements of the model. Higher conservations are assumed in the 3D structures of proteins in comparison to their corresponding primary structures (Lesk and Chothia, 1980). Therefore, structural similarity can typically be presumed, if the likenesses amongst the proteins are noticeable at their sequence level. Furthermore, protein sequences with non-measurable similarities show the presence of analogous structures. There is an assumption that a fraction of all sequences submitted in the databases are noticeably related to no less than one identified protein structure (Flockner et al., 1995). Thus homology modelling could potentially be applied to over 150 000 protein sequences amongst approximately 500 000 submitted sequences (Bairoch and Apweiler, 1999) which matches up to about 10 000 experimentally determined protein 3D structures (Berman et al., 2000). However, the importance of homology model prediction is gradually growing as the

quantity of distinctive structural folds that proteins assumed are limited due to less number of experimentally determined structures (Holm and Sander, 1996).

If the resulting model of the query is not satisfactory because of the low sequence identity (< 25%) with the template structure, then *ab initio* structure prediction protocols such as I-TASSER (Roy et al., 2010) and ROBETTA servers (Kim et al., 2004) are used. The I-TASSER server (Roy et al., 2010) based on *ab initio* algorithms, originally produces 3D models from a manifold threading based alignment and repeated structural compiling simulations. It also predicts the function of the query proteins by means of the structural comparisons of the 3D predicted models with other proteins of known structures which generates the outputs that contain secondary structures as well as full-length tertiary predictions, including Enzyme Commission (EC) numbers and ligand-binding sites etc. (Roy et al., 2010). Similarly, the ROBETTA server (Kim et al., 2004) uses *de novo* or *ab initio* methods to calculate the structural features of proteins which do not have any structural analogs in the PDB (Berman et al., 2000). First, the K*Sync alignment module present in the ROBETTA server is utilized to perform the alignment of the given protein sequences upon the structure of the parent proteins. Therefore, it calculates the flexible regions via permitting them to sample the conformational window through sections in a manner analogous to the protocols based on the *de novo* scheme in the context of the template used. However, if no homologues structures are present in the database, then the server predicts the domain using *de novo* protocols of Rosetta (Misura et al., 2006) that permits the full length of the domain to discover conformational spaces through fragment inclusion, generating a sizeable collection and the concluding models are chosen from this group.

4.1.1 Materials and Methods

The protocols for homology modelling used in this study, consist of four chronological steps outlined in **Figure 4.1** (Sanchez and Sali, 1997).

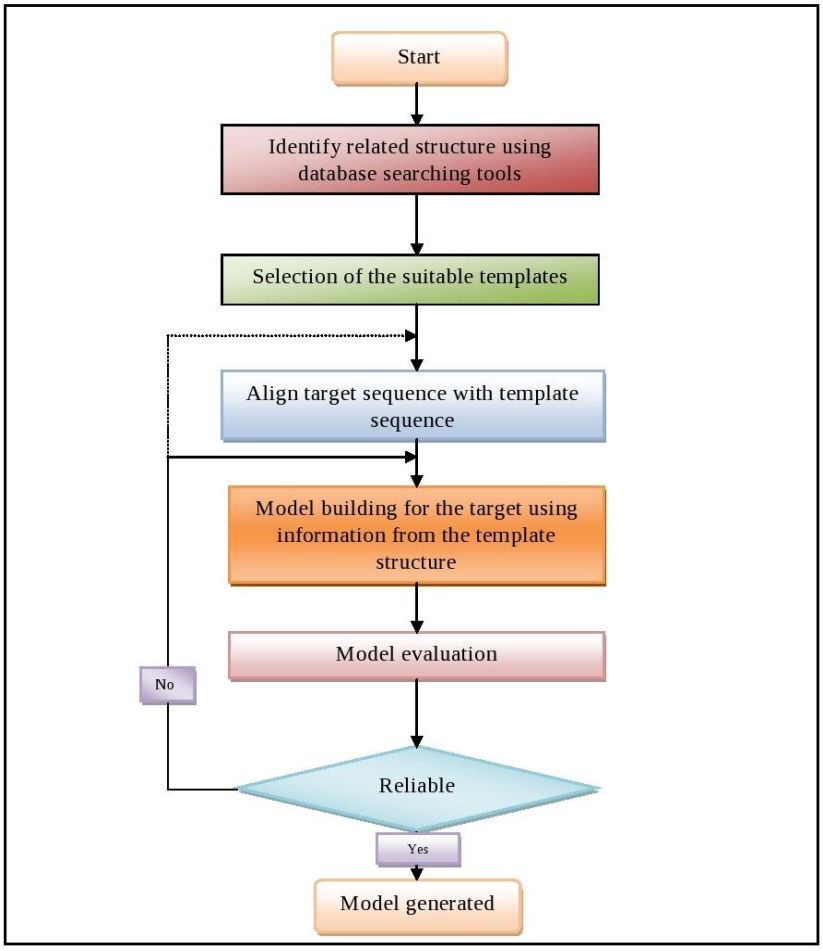


Figure 4.1: Computational workflow adopted for the reliable model generation using homology modelling

In this work, each step of the structure prediction method using homology modelling was performed using various modules present in Discovery Studio (DS) 4.0 (Accelrys Software Inc,

2013). The techniques used for all the steps in the homology modelling are explained below:

4.1.1.1 Template Selection and Fold Assignment

The initial step in homology modelling is to recognize the related protein structures to the target sequence in the structural databases like Protein Data Bank (PDB) (Berman et al., 2000), followed by the selection of template among those proteins. The numerous available protein structures and sequence databases facilitates this phase. The sequence of the target proteins are used for searching the template in the PDB (Berman et al., 2000), DALI (Holm and Rosenstrom, 2010), SCOP (Murzin et al., 1995) and CATH (Orengo et al., 1997). Depending upon the complexity of a genome, there is a 20% to 50% likelihood of calculating an associated protein structure for a respective sequence arbitrarily selected from the given genome (Fischer and Eisenberg, 1997). The protein comparison methods are divided into three major subdivisions which are helpful in similar fold identifications. The primary class are based on pairwise comparisons of sequences that includes the independent comparison of the query sequence with every sequence of the database (Apostolico and Giancarlo, 1998). Frequently used programs for pairwise comparison are FASTA (Pearson, 1998) and BLAST (Altschul et al., 1990). The secondary class of approaches are based on comparisons of multiple protein sequences to increase the accuracy of the template exploration (Altschul et al., 1997). PSI-BLAST (Altschul et al., 1997) are widely utilized for this purpose, which iteratively searches an array of homologs of the query sequence. Regarding a particular protein sequence, a preliminary array of protein homologs from a structural database is gathered, then a weighted multiple sequence alignments is constructed, thereafter a PSSM (position-specific scoring matrix) is created using this

alignment, and the created matrix is utilized to explore the database for the homologs. The database searching steps are reiterated until no novel homologs are discovered. PSI-BLAST locate proteins of similar structures for about twofold as much sequences in comparison to BLAST (Park et al., 1998). An interrelated methodology (Rychlewski et al., 1998) also starts by discovering all sequences evidently linked to the query sequence to find the profile of the query sequence. Furthermore, comparable profiles are created for all identified homologs. Then templates are identified by evaluating the profile of the query sequences with every constructed profile for the established template structures. An additional alternative utilizes the multiple alignments of the protein sequence united with information of the structural elements calculated from the given protein sequences (Fischer and Eisenberg, 1997), which are helpful in the identification of significant relationships between the structures when the sequence identity falls under 25 % i.e. for identifying remote sequence–structure associations. The HMMER (Finn et al., 2011) is an HMM based software used for the profile-HMM comparison of the query protein with various protein databases. The "threading" or fold recognition methods are the listed third class algorithms for structure prediction (Bowie et al., 1991) that utilize the pairwise likeness of a protein of known structure and a query sequence. The given methods are particularly helpful when there are no related sequences for the modelling query. The HHpred server (Soding et al., 2005) used for implementing fold recognition utilizes the pairwise comparison of profile hidden Markov models (HMMs) for remote protein homology detection and structure prediction. Various other domain analysis tools were also used for the reliable template prediction. This analysis was performed using SMART (Schultz et al., 2000), InterProScan 5 (Jones et al., 2014), ProtoNet (Rappoport et al., 2012) and SYSTERS (Schultz et al., 2000) in order to identify the template more precisely. InterProScan 5 searches the similar signature in the interpro consortium

databases. Similarly, Simple Modular Architecture Research Tool (SMART) annotates the genetically distant domains in the sequence of the given proteins and SYSTERS which uses the clustering-based algorithms for the function prediction of proteins. The Classify Your Protein module of ProtoNet was used to analyze the protein family of the respective proteins. The templates suitable for modelling the given proteins were selected on the basis of the following rules.

- The template showing the identity (>30 %) is considered as the reliable homolog. The template belonging to a similar subfamily to that of target sequence was selected on the basis of multiple sequence alignments and a phylogenetic analysis.
- The template descriptor factors like ligands, pH, solvent and quaternary interactions etc. should be evaluated to the required parameters for the model.
- The excellence of the experimental accuracy of the template structure in terms of NMR structure's number of restraints per residue and the R-factor and resolution of a crystallographic structure is considered as important factors for template selection.

The templates fulfilling the respective criteria were selected for the construction of the comparative model. Multiple templates can be used for the homology modelling as the utilization of various templates central from the sequence of the target usually amplifies the predicted model precision (Fiser and Sali, 2003).

4.1.1.2 Template-Target Alignment

The predictions of homology modelling methods are based on folding assignments which makes an alignment among the template structures and sequence of the target. However, it is not considered as the reliable target–template alignment for homology modelling. The database

probing methods are frequently adjusted to find the remote identity. Consequently, when particular templates are selected, a particular means should be utilized to perform the alignment of template structures and the target sequence (Baxevanis, 1998). The alignment is approximately correct with identity over 40 % for closely related protein sequences, but if it is lower than 40 %, then the regions of minimal local sequence similarity turns out to be frequent (Saqi et al., 1998) (Rost, 1999). The alignments contain progressively more alignment errors and a number of gaps as the sequence identity decreases. In order to obtain the most accurate and highly reliable alignment the methods like CLUSTAL W (Thompson et al., 1994) and PRALINE (PRofile ALIgNEment) (Simossis and Heringa, 2005) were used. The latter method utilizes the secondary structure elements for the comparative-extended multiple sequence alignment (Simossis and Heringa, 2005).

4.1.1.3 Three Dimensional (3D) Model Building

After a preliminary alignment of the target–template, an array of techniques can be utilized to predict a 3D model of the query protein. There are three methods widely used for the model building i.e. modelling using rigid body assembly (Blundell et al., 1987), segment matching (Claessens et al., 1989) and satisfaction of spatial restraints (Aszodi and Taylor, 1996). Modelling through assembly of rigid bodies form the 3-D model of the protein from a few of the rigid bodies generated from the alignment of the protein structures (Blundell et al., 1987). This method is centred on the inherent dissection into variable loops, conserved core regions and side chains of the various protein folds (Blundell et al., 1987). 3D model constructions by Coordinate Reconstruction or Segment Matching forms the basis on the assumption that the majority of the hexapeptide segments in the structure of proteins can be grouped into roughly 100 structural

classes (Unger et al., 1989). Therefore, the atomic positions from the template structures can be utilized to build homology models by recognizing and collecting brief segments which mount these positions (Unger et al., 1989). The Cα atoms are assumed to be the guiding positions of the segments conserved in the aligned sequence of target and the structure of the template (Unger et al., 1989). The prediction of the protein structure using satisfaction of the spatial restraints produces many restraints or constraints obtained from the alignment with the template structure on the framework of the query sequence. The assumption that the equivalent angles and distances among aligned residues in the query and the template structures are alike, leads to the generation of the restraints. The stereochemical restraints on dihedral angles, bond angles, bond lengths and non-bonded atoms acquired from an empirical force field, supplemented these comparatively derived restraints. The 3D model was then constructed by reducing the infringements of all the restraints. In this study, the MODELLER (Fiser and Sali, 2003) module present in DS was used to construct the three dimensional structure of the query protein by satisfying the spatial restraint.

4.1.1.4 Loop Modelling

Variations in the structure resulted from insertions, deletions and substitutions among the members of a given fold family. These respective changes often occur in the loop regions that link the fundamental secondary structural elements present in the protein 3D characteristic fold. Therefore, specificity in the functional behaviour of a protein is often determined by the conformational behaviour of the loops. The binding and active sites are usually present in the loop region. As a result, the precision adopted in modelling the loop regions are key aspects in shaping the efficiency of homology predicted models by examining receptor-ligand interactions.

The loop regions predicted can also be assumed as a mini–protein folding challenge. The accurate conformational preferences of a given protein chain fragment have to be predicted principally from the primary structure of the protein itself. The loop regions were modelled using MODELLER (Fiser and Sali, 2003).

4.1.1.5 Side chain Modelling

In contrast to loop modelling, the conformations of the side chains are calculated from analogous structures and from energetic or steric considerations (Sali, 1995). The predictions of disulphide bridges can be assumed as a unique task in side chain modelling of the proteins. The information present in the structure of proteins are used in the side chain modelling (Jung et al., 1994) and from the corresponding disulfide bridges in analogous protein structures (Sali and Overington, 1994). The two effects that are considered as significant in side chain modelling includes the coupling amongst the side chains and main chains and the incessant nature of the allocations of dihedral angles in side-chain (Vasquez, 1996). While comparing stability and packing energies both of these effects were found to be significant (Lee, 1996). Thus, the SCWRL4.0 (Wang et al., 2008) used is a command line based software, designed particularly for the calculation of side-chain conformations assuming that a fixed backbone typically resulted from the structures obtained through experimental techniques like NMR or X-ray crystallography.

4.1.1.6 Evaluation and refinement of models

An estimation of the precision of the predicted 3D models of the proteins is necessary for the understanding of derived structural information. The model evaluation was performed using whole structure as well as in distinct regions. The fold analysis was the first step using the energy

based Z-score, assuming that for a score less than 2.5, then the template and the query protein belong to the same fold (Sanchez and Sali, 1998). Various available methods were used for the evaluation of the model such as TM score and DOPE Profile. Moreover, the 3D models were further verified using the modules of the SAVES server (http://nihserver.mbi.ucla.edu/SAVES/) such as PROCHECK, VERIFY3D, WHAT_CHECK, PROVE and ERRAT modules. The PROCHECK assesses the stereo-chemical quality of the structure residue by residue in comparison with a well refined structure of the proteins with similar resolutions. Similarly, WHAT_CHECK algorithm performs the extensive calculation of the stereo-chemical parameters. ERRAT calculates the statistical parameters of the non-bonded interactions between different atom types, while PROVE computes the Z-score deviation by utilizing the comparisons with highly refined PDB structures and VERIFY_3D performs the compatibility of three dimensional protein models with its own 1D structure i.e. amino acid sequences. If the evaluation score was not satisfactory, then the selected models were refined using GROMOS energy minimization algorithm implemented in DeepView (Kaplan and Littlejohn, 2001), CHARMm (Vanommeslaeghe et al., 2010) energy minimization using ChiRotor algorithm of DS and SCWRL 4.0 (Wang et al., 2008).

4.1.2 Results and Discussion

4.1.2.1 Template searches

The sequence similarity searches indicated the presence of chitinase like activities in their framework. The top 10 blast hits revealed that one of the given chitinase proteins (designated as chitinase II) was similar to the fungal class III chitinase, with the highest identity of 65 %

belonging to *Byssochlamys spectabilis* chitinase III having an e value of 3e-157 in the NCBI non-redundant database, while in the PDB search identified *Saccharomyces cerevisiae* chitinase **(**PDB Id: 2UY2) as the suitable template for the structure prediction of chitinase II. Similarly, another protein (designated as chitinase I) query sequence was similar to the chitinase of *Thermomyces lanuginosus* showed 98 % identity with an e-value of 0.00. Chitinase I is similar to chain A of the chitinase present in *Aspergillus fumigatus* (PDB ID: 1WNO) with identity of 62% selected to be a suitable template for chitinase I in Protein Data Bank. Moreover, the InterProScan revealed that both sequences contain domains were similar to the Glycoside hydrolase family 18, a group of enzymes that hydrolyzes the glycosidic linkages between the carbohydrate units. The given proteins were also found to be a member of 4164978 and 136700 clusters of ProtoNet and SYSTERS respectively, which are the groups of chitinase like enzymes.

4.1.2.2 Structure prediction and analysis

MODELLER (version - 9.14) was used to align the template structures and to query sequences. The predicted structures of chitinase II and chitinase I showed low violations of the restraints which were considered to be more precise. Consequently, 20 models were generated and evaluated on the basis of RMSD, TM-score, DOPE profile and Molecular Probability Density Function (Mol PDF). The generated structure of the both chitinase II and chitinase I are shown in **Figure 4.2.**

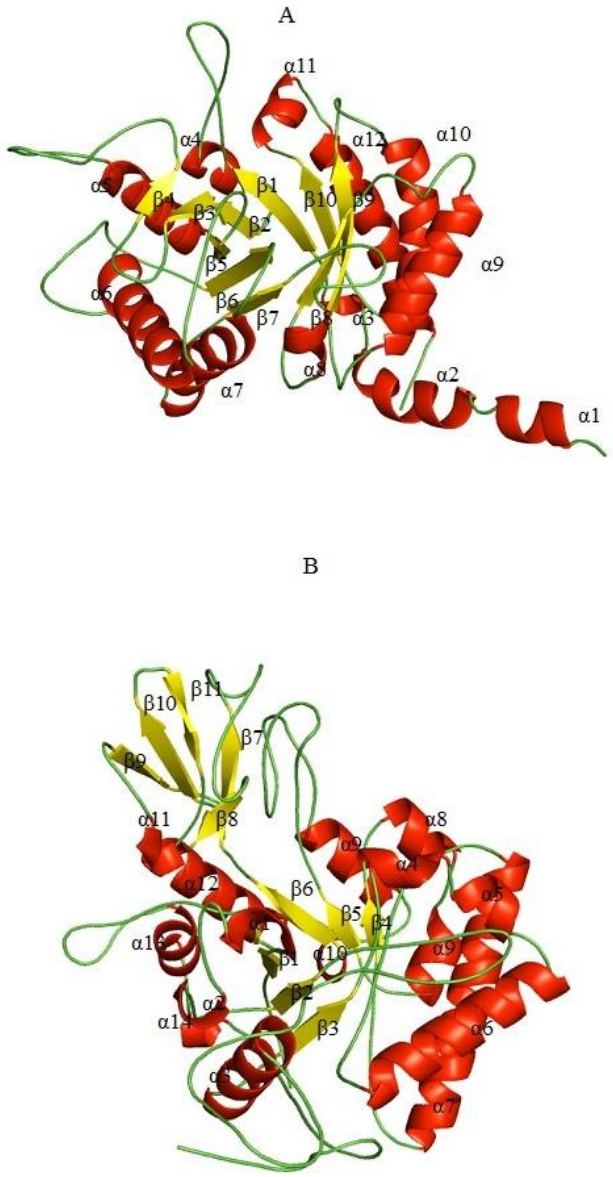


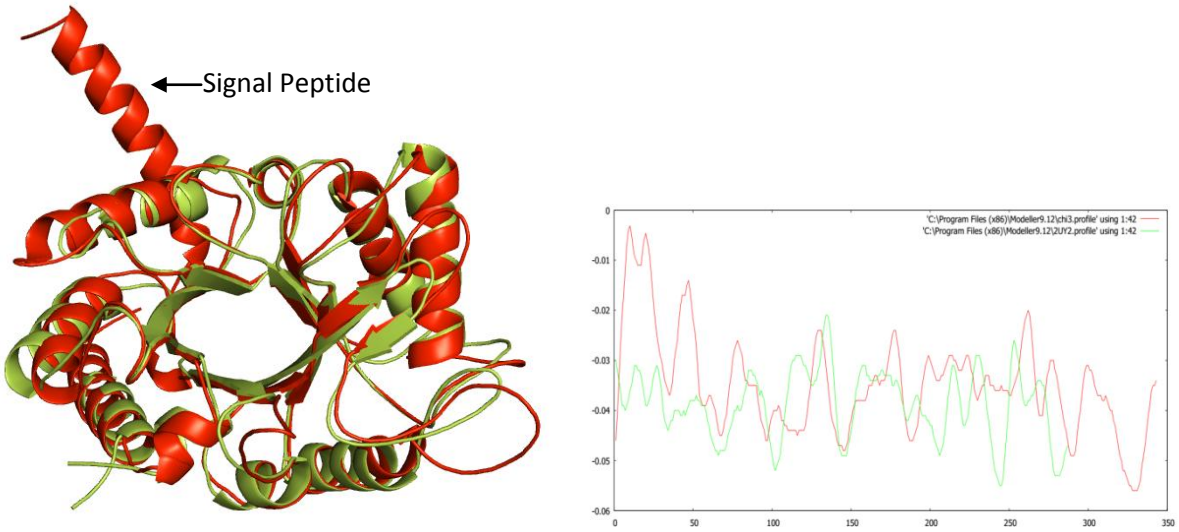
Figure 4.2: The predicted three dimensional (3D) structure of (A) chitinases II and (B) chitinase I highlighting the characteristic TIM barrel.

The predicted models showed overhanging sequences which were further modelled using *ab intio* protocols implemented in the I-TASSER and ROBETTA servers. Dope energies of the

models and the templates structures were computed, which revealed loop regions from residues 24 to 105 of the model with higher DOPE energies than the template (**Figure: 4.3A**). The DOPE scores for the chitinase II and its template (2UY2) were -37797.9609 and -37746.6328 respectively. While for chitinase I the DOPE profile was comparable with template shown in Figure 4.3 B. The DOPE score for chitinase I and its template (1WNO) were -41595.5585 and -49313.0078 respectively. This indicates the reliability of the predicted models, which were subjected to further refinement. The generated models were optimized using the CHARMM 22 forcefield present in the DS. Thereafter the energy minimizations of the 3-D structures were performed in order to avoid bad molecular contacts by using the Deepview program. The relative energies for chitinase I and chitinase II models were -9060.353 kJ/mol and -5523.179 kJ/mol respectively.

The RMSD and TM scores between the chitinase II and its template were 0.945 and 0.1719 respectively, indicating that both belong to the same fold. In contrast, the RMSD value for chitinase I was calculated to be 0.712 after superimposing it with 1WNO and with a TMscore of 0.7148 between them indicating a high similarity of the structures. The models were further evaluated using SAVES. The PROCHECK showed 98.3 % for chitinase II of the total residues in the Ramachandran Plot and for chitinase I it was 98.5 % (**Figure 4.4)**. Verify_3D showed 88.66 % and 96.26 % of the total residues had an averaged 3D-1D score (>0.2) for chitinase II and chitinase I respectively, indicating the reliability of the model prediction. The Overall quality factor score predicted by Errat was 30.102 and 63.582 for the chitinase I and II respectively.

A



B

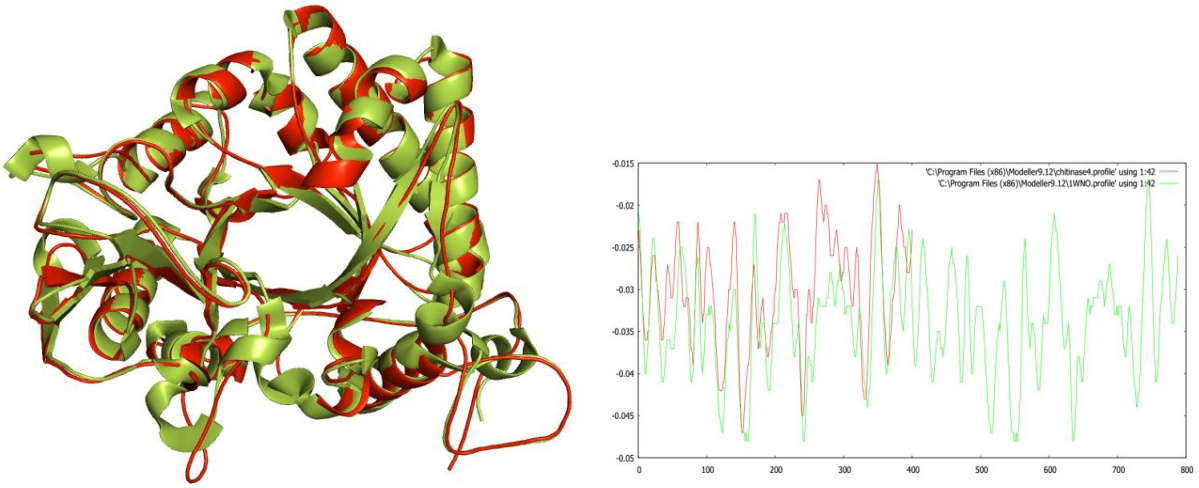


Figure 4.3: The superimposition of predicted model (RED) with template (GREEN) and their calculated DOPE profiles using MODELLER. (A) The prediction for chitinase II (B) Analysis results for chitinase I**.**

(A) (B)

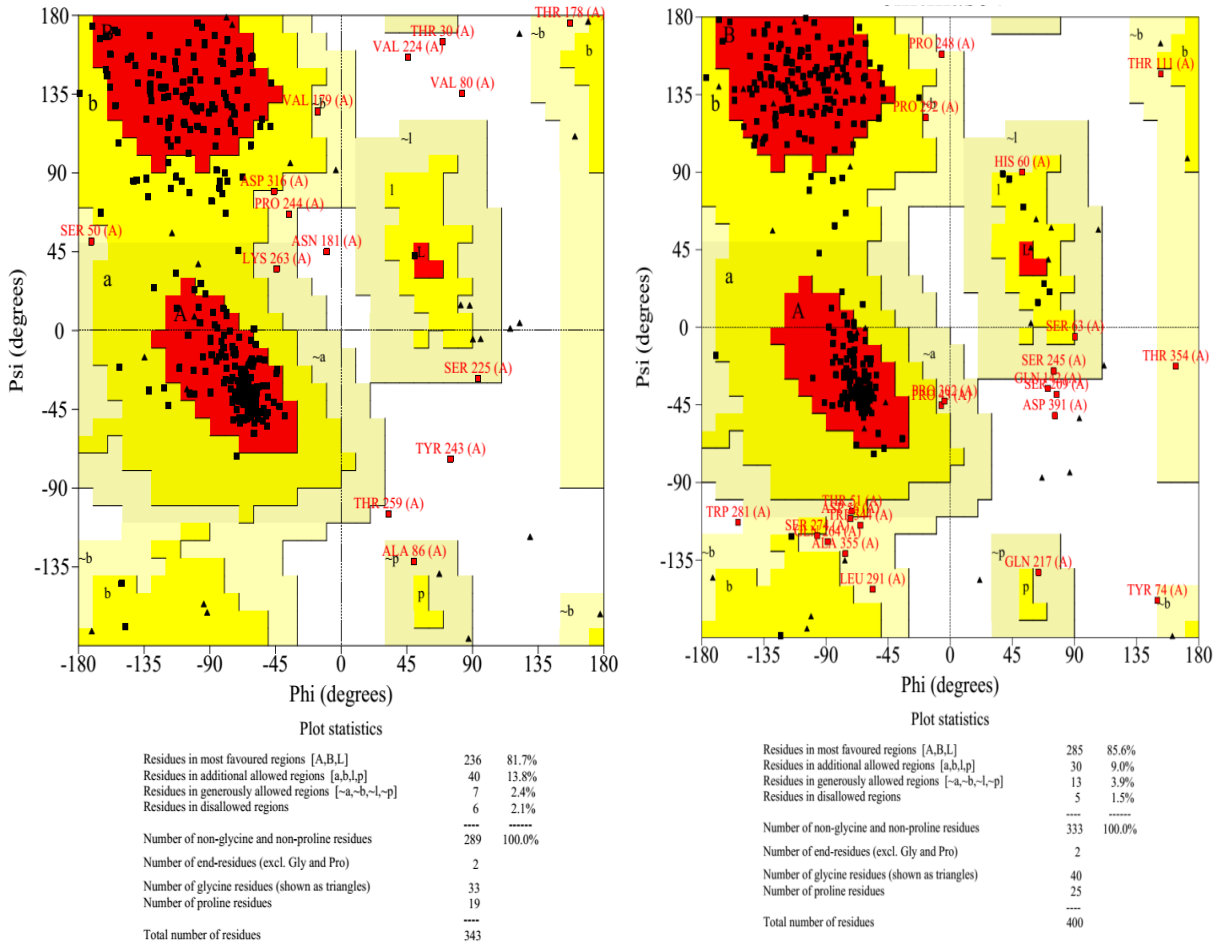


Figure 4.4 The Ramachandran plot for predicted (A) chitinase II (B) chitinase I enzymes respectively.

The topology for the given structures of both chitinases were generated using PDBsum (de Beer et al., 2014) in order to understand the detailed structural features. The structurally conserved $(\beta/\alpha)_8$ TIM-barrel were observed in the framework of chitinase II (**Figure 4.2 A and 4.5 A**). The chitinase II contains 8.7 % beta strands (30 AA), 34.7 % alpha helix (119 AA) and 0.9 % 3-10 helix (3 AA) (**Figure 4.5 A**). The structure showed the presence of 1 beta sheet comprising of 8 parallel beta strands, 12 helices and 11 helix-helix interactions and 6 characteristics beta-alpha-

beta motifs that may be responsible for the activity of the enzymes. Moreover, there is a presence of a disulphide cysteine bridge between 92Cys-103Cys residues (**Figure 4.5 A**). The chitinase II contains well defined TIM-barrel with a large β/α domain inserted into several loops (**Figure 4.2 A**). The structure contains a shallow substrate binding groove as indicated after a structural comparison with the template. Chitinase II showed a presence of one conserved solvent exposed amino acid Trp307.

(A) (B)

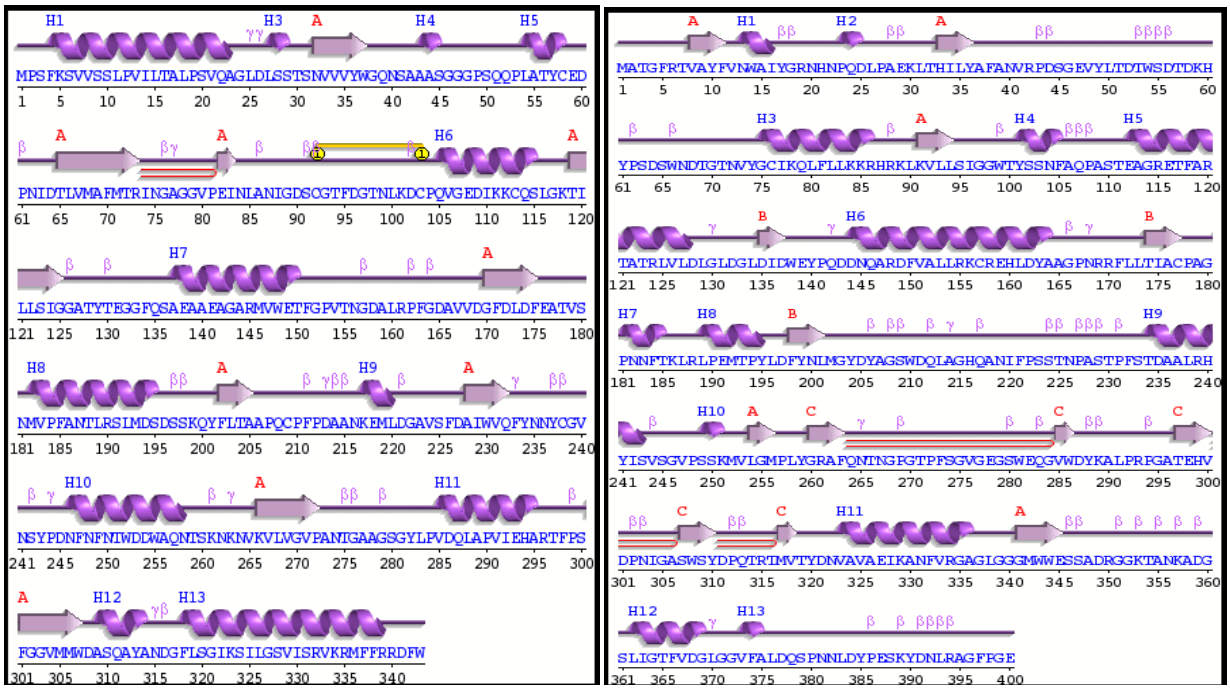


Figure 4.5 The topology of (A) chitinase II and (B) chitinase I generated using PDBsum.

The structure of chitinase I showed (Trans) glycosidases like β/α TIM-barrel as the characteristic domain of its structure with 47 (11.8 %), 94 (23.5 %) and 18 (4.5 %) of amino acids are Strand, Alpha helix and 3-10 helix, respectively (**Figure 4.2 B and 4.5 B**). The structure of chitinase I

revealed three beta sheets which are composed of eight parallel beta strands and five anti parallel beta strands (**Figure 4.5 B**). There are 13 helices with 10 helix-helix interactions as well as 3 particular beta-alpha-beta motifs in chitinase I.

4.1.3 Conclusions

The models were generated with high precision refinement using various available prediction algorithms. These models showed high structure similarities to the chitinase-like proteins with the characteristic $(\beta/\alpha)_8$ TIM barrel found in both proteins. The active site analysis is predicted in section 4.2 followed by further structure refinement using the molecular dynamics protocols implemented in GROMACS (Van Der Spoel et al., 2005) as described in the section 4.3.

4.2 Molecular Docking

The characteristics of an enzyme derived from its amino acid sequence determine the shape of the enzymes. A closer fit between an active site of an enzyme and its substrate molecule increases the efficiency of a reaction (Sullivan and Holyoak, 2008). The active site is typically found in a 3D groove or pocket of the enzyme molecule aligned with amino acid residues. The molecular utility of a protein is generally limited to its active site residues, which may comprise of an interaction with small size ligands, nucleic acids or other proteins (Powers et al., 2006). Integrity of the essential structural constituent of the active site is vital for preserving the functional activity of the protein.

The interaction studies between biomolecules are central to all biological processes. The interaction between biologically relevant molecules such as nucleic acids, proteins, carbohydrates and fatty acids play an important role in the process of signal transduction. Complex regulatory and metabolic interaction networks within the living systems are governed by these interactions. Experimental observations and computer based theoretical analysis are the main scientific tools for understanding this phenomenon.

Molecular docking is a computational method which tries to predict the preferred orientation of one molecule to a second when bound to each other to form a stable complex (Kitchen et al., 2004). The docking of small molecules i.e. ligands against larger receptor molecules is a complex task. Docking between these respective molecules is often considered to be a *Lock and Key* mechanism, where the receptor and ligand do not change the conformation during binding. The ligands are believed to be more flexible and assume multiple conformations in solution space. Moreover, even though the receptors show well defined conformations, they can also

show the alteration if the binding of ligand through *Induced Fit* mechanism. Knowledge of the preferred alignment in turn may be used to predict the strength of association or binding affinity between two molecules using scoring functions. Docking simulations are utilised for rational drug design and virtual screening of the library of small compounds (Kitchen et al., 2004). Most of the docking algorithms generate a large number of likely conformations, some of which can be rejected immediately due to high energy clashes with the protein. We require a means to score each pose to identify those of most biological relevance. The free binding energy can be written as an additive equation of various intermolecular energies to reflect their relative contributions in binding.

$$\Delta G_{bind} = \Delta G_{solvent} + \Delta G_{conformation} + \Delta G_{interactions} + \Delta G_{rotations} + \Delta G_{vibrations}$$

Generally, the docking practices between ligands and the receptors can be executed in a rigid mode, where a single conformation is used. This methodology is of little importance in contrast to docking procedures. The universal assumption for docking studies is to grasp the protein in a strict rigid conformation and to dock a chain of ligand conformations against the active site of the receptor protein. Most of the docking algorithms are based on these assumptions. On the other hand flexible docking allows the optimization of a particular arrangement of side-chains during the process of molecular docking. Flexible docking also facilitates the specification of pre-generated conformations of the receptor that comprises backbone and side-chain flexibility. Flexible docking usually takes additional CPU time in comparison with the rigid docking practices. In this work, we predicted the active site of chitinase using the information present in the literature of similar proteins, various bioinformatics tools and docking studies.

4.2.1 Material and methods

4.2.1.1 Active pocket prediction

Active or binding sites on the protein surfaces play a central role in the protein functions. The identification of those binding sites on the protein molecule is often the first step to study protein functions and structure-based drug design. In this study, the active site was predicted using various bioinformatics tools as well as by comparing the chitinase II amino acids sequence with the template and other similar proteins. The active site pockets were predicted by metaPocket 2.0 (Zhang et al., 2011), COFACTOR (Roy et al., 2012) and COACH (Yang et al., 2013). The metaPocket 2.0 algorithm uses LIGSITEcs, PASS, Q-SiteFinder, SURFNET, Fpocket, GHECOM, ConCavity and POCASA predictors to identify the pocket sites. COFACTOR identifies template proteins of similar folds and functional sites by threading the target structure through three representative template libraries that have known as protein–ligand binding interactions, Enzyme Commission number or Gene Ontology terms. COACH was based on two methods, (i) based on binding-specific substructure comparison (TM-SITE) and (ii) based on sequence profile alignment (S-SITE), for complementary binding site predictions. The docking studies were further done with allosamidin derivative allosamizoline as ligand, a reported chitinase inhibitor (Suzuki et al., 2006) to visualize the interaction with predicted active pocket site. The receptor and ligand were optimized using various protocols available in Discovery Studio (DS) (Koska et al., 2008). Docking was performed using CDOCKER (Wu et al., 2003), which is a CHARMm-based docking method (Wu et al., 2003) that generates highly accurate docked poses.

4.2.1.2 Ligand docking with CDOCKER

CDOCKER is a powerful docking method based on CHARMm algorithm that has been presented to generate highly accurate docked poses. CDOCKER is a molecular dynamics (MD) simulated-annealing-based algorithm. An important advance to the CDOCKER protocols has been the introduction of the soft-core potentials. Soft-core potentials are found to be effective in exploring the conformational space of small organics and macromolecules and are being used in various applications, including docking The steps includes Preparing Inputs, Launching CDOCKER and Analyzing CDOCKER Results.

4.2.1.3 Preparing Inputs

The generated 3D chitinase model was loaded and the "Prepare Protein" module was used, to remove the water molecules, to add a suitable forcefield and hydrogens to the suitable sites. The "Preparing a Receptor Binding site module" and information present in the literature was used to predict the probable active site present in the protein. Similarly, the "Prepare Ligand" module was used for the preparation of the ligands to make them suitable for the docking procedures.

4.2.1.4 Invoking CDOCKER

The chitinase and allosamizoline were docked against each other by using CDOCKER module. To ensure the docked poses are diverse, the RMSD threshold was set to 0.5 Å. Then the Top Hits parameter was expanded and the Pose Cluster Radius was set to 0.5.

4.2.1.5 Analyzing CDOCKER Results

The “Receptor-Ligand Interactions” tools were used to visualize non-bond interactions. The 10 docked poses were obtained and ranked according to the CDOCKER energy with best pose has highest negative CDOCKER_ENERGY.

4.2.2 Result and Discussion

Using the information present in the literature and the comparison of sequence with template, the active pocket of the chitinase II may possess Tyr36, Phe70, Asp174, Glu176, Gln208, Gln232, Tyr234 and Trp307. The Glu176 was predicted to be essential for the activity of the chitinase (**Figure 4.6**).

```
                                    10        20        30        40
gi|702076139|gb|AIW06012.1|      MPSFKSVVSSLPVILTALPSVQAGLDLSSTSNVVVYWGQNSAAASGGGPS 50
2UY2_A|PDBID|CHAIN|SEQUENCE      -------------------------DRSANTNIAVYWGQNSAG------T 19
                                                          * *:.:*:.********.      :
                                    60        70        80        90
gi|702076139|gb|AIW06012.1|      QQPLATYCEDPNIDTLVMAFMTRINGAGGVPEINLANIGDSCGTFDGTNL 100
2UY2_A|PDBID|CHAIN|SEQUENCE      QESLATYCESSDADIFLLSFLNQFPTLG----LNFANA--CSDTFS-DGL 62
                                 *:.******..: * ::::*:.::   *    :*:**   ...**.  .*
                                    110       120       130       140
gi|702076139|gb|AIW06012.1|      KDCPQVGEDIKKCQSLGKTILLSIGGATYTEGGFQSAEAAEAGARMVWET 150
2UY2_A|PDBID|CHAIN|SEQUENCE      LHCTQIAEDIETCQSLGKKVLLSLGGASGSY-LFSDDSQAETFAQTLWDT 111
                                  .*.*:.***:.******.:***:***: :   *.. . **: *: :*:*
                                    160       170       180        190
gi|702076139|gb|AIW06012.1|      FGPVTNGDALRPFGDAVVDGFDLDFEATVSN-MVPFANTLRSLMDSDSSK 199
2UY2_A|PDBID|CHAIN|SEQUENCE      FGEGT-GASERPFDSAVVDGFDFDIENNNEVGYSALATKLRTLF-AEGTK 159
                                 **  * * : ***..*******:*:* . .    .:*..**:*: ::.:*
                                      210       220       230       240
gi|702076139|gb|AIW06012.1|      QYFLTAAPQCPFPDAANKEMLDGAVSFDAIWVQFYNNYCGVNSYPDNFNF 249
2UY2_A|PDBID|CHAIN|SEQUENCE      QYYLSAAPQCPYPDASVGDLLENAD-IDFAFIQFYNNYCSVS---GQFNW 205
                                 **:*:******:***:  ::*:.*  :*  ::********.*.   .:**:
                                      250       260       270         280
gi|702076139|gb|AIW06012.1|      NTWDDWAQNTSKNKNVKVLVGVPANTGAAGSGYLP-VDQLAPVIEHARTF 298
2UY2_A|PDBID|CHAIN|SEQUENCE      DTWLTYAQTVSPNKNIKLFLGLPGSASAAGSGYISDTSLLESTIADIASS 255
                                 :**  :**..* ***:*:::*:*..:.******:. .. * ..* .  :
                                       310       320       330       340
gi|702076139|gb|AIW06012.1|      PSFGGVMMWDASQAYANDGFLSGIKSILGSVISRVKRMFFRRDFW 343
2UY2_A|PDBID|CHAIN|SEQUENCE      SSFGGIALWDASQAFSNELNGEPYVEILKNLLTSASQTA------ 294
                                 .****: :******::*:   .   .** .::: ..:
```

Figure 4.6: Structure analysis with template (PDB id: 2UY2) shows active site pocket in chitinase II comprises residues 36, 70, 174, 176, 208, 232, 234 and 307.

On the basis of the outputs obtained from the metaPocket 2.0, COFACTOR and COACH servers, the above active site positions were predicted to be potential ligand binding site in chitinase. In order to validate the above predicted results the docking analysis were performed using allosamizoline as ligand and chitinase as receptor. The allosamizoline was docked at the active site pocket of the chitinase using CDOCKER protocol of the Discovery Studio. The top ten conformations of the docked allosamizoline-chitinase II complex were selected for the evaluation as tabulated in **Table 4.1**

Table 4.1 The obtained cdocker energy and cdocker interaction energies of 10 docked poses of chitinase II with allosamidin derivatives.

DOCKED POSITIONS	CDOCKER_ENERGY (kcal/mol)	CDOCKER INTERACTION ENERGY (kcal/mol)
1	-8.29987	-29.1129
2	-8.39712	-29.3752
3	-10.128	-28.7152
4	-10.4287	-27.5459
5	-10.561	-26.854
6	-10.6359	-27.4068
7	-11.0569	-27.2111
8	-11.0672	-26.5673
9	-11.2785	-26.4762
10	-11.4503	-28.836

Three stable docked pose conformations shown in **Figure 4.7 (A), (B) and (C)**. The docked pocket contains Ala127, Asp174, Glu176, Ala206 and Gln232. The CDOCKER energy for these predicted conformations was shown in Table 4.1.

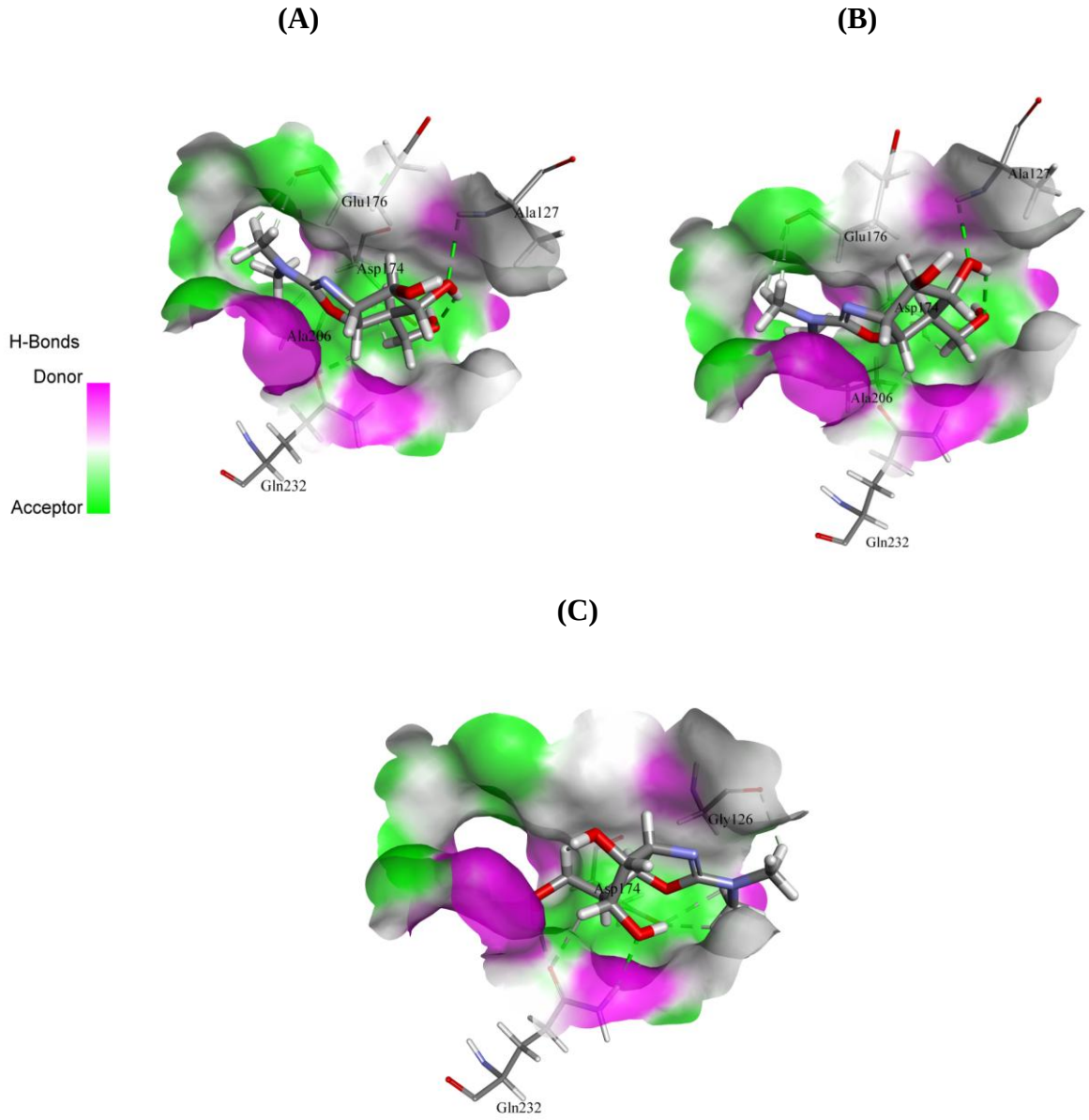


Figure 4.7 (A, B and C): The top three docked poses confirming that Glu176 is critical for the activity of chitinase II.

4.2.3 Conclusions

A closer inspection of the predicted 3D structures of chitinase II showed that it contains Tyr36, Phe70, Asp174, Glu176, Gln208, Gln232, Tyr234 and Trp307 in the active site pocket. Specifically Glu176 was observed to be essential for the activity of chitinase II. These analyses of the active site was further confirmed by molecular docking studies using CDOCKER of Discovery Studio. These analysis were helpful in understanding the active sites as this enzyme is thermostable. Accordingly MD simulations were performed to understand the structural conformation of chitinase at different temperatures as discussed in the next section.

4.3 Molecular Dynamics simulations

Molecular dynamics (MD) simulations are one of the most important tools for the computational study of biomolecules (Lindahl, 2008). It computes the time dependent behavior of a molecular system and therefore provides comprehensive information of changes in protein fluctuations and conformation. They are also used in structure determination from experimental NMR and X-ray crystallography. Molecular dynamics (MD) is described as a type of computer simulation in which the atoms and the molecules are allowed to interact for a time period by approximations of known physical attributes, resulting in the simulation of the motion (Lindahl, 2008) of a system of particles (Lindahl, 2008). The analytical study of the properties of such complex systems is a challenging task and therefore, an MD simulation provides a solution by means of numerical methods. It represents an interface between laboratory experiments and the theory and can be understood as a effective experiment thereby probing the link between molecular structure, function and its movement. The laws and the theory of MD simulations are derived from interdisciplinary science like physics, chemistry and mathematics. It utilizes algorithms from the computer science and information theory. It is applied today generally in materials science to study complex dynamic processes that occur in biological systems, including protein folding, protein stability, molecular recognition, conformational changes, drug design, determination of 3D structures, enzyme reactions and ion transport in biological systems. (Lindahl, 2008, Karplus and McCammon, 2002). It is believed that the atoms in the molecule interact with each other according to the rules of the employed force field. The configurations of the systems are generated by Newton's Laws of motion, which specifies the variations of the atomic positions and particle velocities with respect to time. This can be explained by determining the force on each particle (F_i) as a function of time, i.e. negative gradient of the potential energy:

$$F_i = -\frac{\partial U}{\partial r_i}$$

Where U is the potential energy function and r is the position of a particle.

MD simulations solve Newton's equations of motion for a system of N interacting atoms:

$$F_i = m_i \frac{\partial^2 r_i}{\partial t^2}, i = 1 \ldots N.$$

The acceleration (a_i) of each particle can then be determined by dividing the force acting on it by the mass of the particle:

$$a_i = \frac{F_i}{m_i}$$

The kinetic energy can be defined in terms of both the velocities and momentum of the particles:

$$K(v) = \frac{1}{2} \sum_{i=1}^{N} m_i {v_i}^2$$

$$K(p) = \frac{1}{2} \sum_{i=1}^{N} \frac{{p_i}^2}{m_i}$$

The total energy of the system i.e. Hamiltonian, is the total of the kinetic and potential energies:

$$H(q, p) = K(p) + U(q)$$

Where q is the set of Cartesian coordinates, p is the momenta of the particles and U(q) represents the potential energy function.

The velocities, $v_i(t)$ are the first derivative of the positions with respect to time:

$$V_i(t) = \frac{d}{dt} q_i(t)$$

where $q_i(t)$ represents the atomic positions at a particular time, t.

New velocities and positions of the atoms are calculated at time t based on initial atom coordinates of the system. The atoms will be moved to these new positions. As a result, a new conformation is formed. The cycle is then repeated for a number of steps described that gives a collection of energetically accessible conformations called an ensemble. The vibration frequencies in molecules are shown in Table 4.2.

Table 4.2: Typical vibration frequencies in molecules (wave numbers) and hydrogen bonded liquids. [Source: Gromacs user manual 6.5.4]

S.No.	Type of bonds	Type of vibrations	Wave number (cm^{-1})
1.	C-H, O-H, N-H	Stretch	3000 – 3500
2.	C=C. C=O	Stretch	1700 – 2000
3.	HOH	Bending	1600
4.	C-C	Stretch	1400 – 1600
5.	H_2CX	Sciss, Rock	1000 – 1500
6.	CCC	Bending	800 – 1000
7.	O-H…O	Liberation	400 – 700
8.	O-H…O	Stretch	50 – 200

4.3.1 Periodic Boundary Conditions (PBC)

A more accurate approach in simulations is to employ the solvent explicitly, which is prepared by soaking the molecule in a box of solvent which require extra computational effort. Periodic Boundary Conditions (PBC) are usually employed to model in the bulk solvent. In the case of an infinite PBC, the simulation box is infinitely replicated in every path to form a lattice. Practically, most of the MD simulations determine the potentials with some cutoff scheme for

the computational efficiency. In these cutoff schemes, every particle interacts with the adjoining images of the other n-1 particles (minimum image convention). The cutoff methods has been shown to present simulated behavior and significant errors in simulations (Holden et al., 2013).

4.3.2 Ewald Summation Techniques

The long range coulombic interactions are time consuming in most of the MD simulation methods. Ewald summation (ES) was presented in 1921 (Fennell and Gezelter, 2006) as a method to figure out the long range interactions among infinite particles and their countless periodic images. Long-range interactions are evaluated as sums that converge very slowly. The principle of obtaining the Ewald sum is by the conversion of the summation of the two series potential energy as shown in Figure 4.8 (Fennell and Gezelter, 2006). A Gaussian charge distribution is normally used. The sum over the point charges is changed to a sum of the interactions between the charges and the neutralizing distributions according to the equation below:

$$\mathbf{U_{Ewald} = U^{r} + U^{m} + U^{0}}$$

This part is the real space sum $\mathbf{U^{r}}$. A second charge distribution is included to the system that exactly neutralizes the first distribution. This summation is achieved in the reciprocal space and is named as $\mathbf{U^{m}}$. The dipole ($\mathbf{U^{0}}$) includes the effect of the entire dipole moment of the unit cell, the macroscopic lattice shape, and the dielectric constant of the surrounding medium (Fennell and Gezelter, 2006).

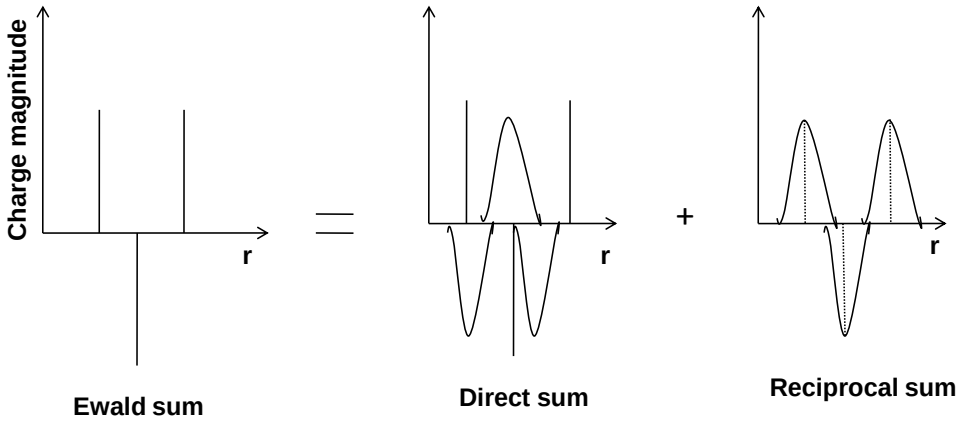


Figure 4.8 Splitting of charges into discrete and smeared distributions in the real and reciprocal space.

4.3.3 Particle Mesh Ewald (PME)

The Particle Mesh Ewald method (PME) divides the potential energy into Ewald's standard direct and reciprocal sums. It utilizes the conventional Gaussian charge distributions (Norberto de Souza and Ornstein, 1999). The direct sum is calculated explicitly by means of cutoffs whereas the reciprocal sum is estimated via Fast Fourier Transform (FFT) with convolutions on a grid where charges are interpolated in the grid points (**Figure 4.9**). Furthermore, it does not interpolate but somewhat calculates the forces by analytically differentiating of energies, thus reducing memory requirements substantially (Norberto de Souza and Ornstein, 1999).

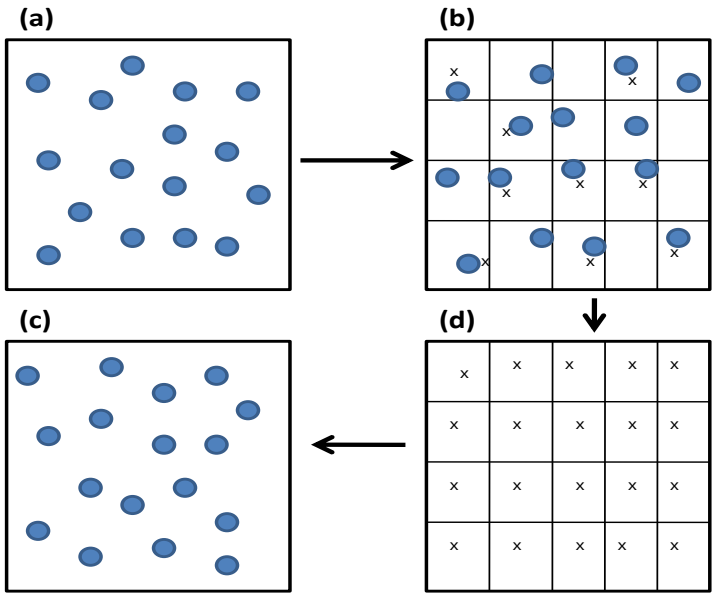


Figure 4.9 A 2D **s**chematic representation of the particle-mesh techniques used in most Fourier-based methods **(a)** A system of charged particles. **(b)** The charges are interpolated on a 2D grid. **(c)** Using FFT, the potential and forces are calculated at grid points. **(d)** Interpolate forces back to particles and update coordinates.

4.3.4 Thermostats in MD

Various thermostat methods are available to add and eliminate energy from the boundaries of an MD system in a comparatively realistic way, approximating the canonical ensemble (Fuzo and Degreve, 2012). In the canonical ensemble, the particles number (N), temperature (T) and the volume (V) are conserved. There is a exchange of energy between endothermic and exothermic processes with a thermostat (Fuzo and Degreve, 2012). However, since the temperature is described by the ensemble average kinetic energies of all particles as in Equation 1.

$$\langle \frac{1}{2} m v^2 \rangle = \frac{k_B T}{2} \quad \ldots\ldots\ldots (1)$$

Thus, different types of thermostats, such as Nosé-Hoover, Berendsen and Langevin thermostats have been suggested to manage the particle motions. A Berendsen thermostat, a proportional type of thermostat, that corrects the deviations (T) from set point (T_0) by multiplying the velocities by a factor to manage the value of (T) (Fuzo and Degreve, 2012). Nosé-Hoover thermostat, an integral type of thermostat, that introduces extra degrees of freedom (momentum) into the Hamiltonian of a system (Fuzo and Degreve, 2012). The Langevin thermostats follow the Langevin equation of motion instead of Newton's equation of motion (Fuzo and Degreve, 2012). In the Langevin equation, a frictional force added to the conservative force is proportional to the velocity. It regulates the kinetic energy of the particle as a result the temperature matches the set temperature as shown in Equation 2.

$$ma = -\xi v + f(r) + f' \quad \ldots\ldots\ldots\ldots (2)$$

Where m is mass, a is acceleration, $f(r)$ is conservative force, v is the velocity, ξ is a frictional constant, and f' is a random force. The frictional force $-\xi v$ decreases the temperature as ξ is a fixed positive value. The force is determined from a Gaussian distribution to add kinetic energy to the particle, and its variance is the function of set temperatures and time steps. The balance of random force with frictional force therefore, maintains the system temperature at a set value.

4.3.5 Solvent models

The MD simulations of proteins were carried out in explicit solvent, where water is represented by all atom force field models. Due to its significance and ubiquity, water is the most explored liquid by molecular simulations (Nguyen et al., 2014). Currently there are several water models

used in biomolecular simulations such as SPC/E, TIP3P and TIP5P (Jorgensen and Tirado-Rives, 2005). These models are optimized to one or more physical properties of water, such as the radial distribution function, diffusivity, density anomaly etc. In this study we selected SPC/E 216 models for all MD simulations. It is realistic to include water molecules explicitly in simulations of protein.

4.3.6 Energy-Minimization Procedures in Simulations

Energy minimization methods are divided into several classes based on the order of derivative employed for locating a minimum on the potential energy surface. Zero order methods are those which use the energy function to identify low energy regions through a grid search method. The first-derivative techniques include the steepest descent method or the conjugate gradient method, using the gradient of energy function. Whereas, the second-derivative methods, like the Newton-Raphson algorithm make use of the Hessian function to locate the minima (Kini and Evans, 1991).

4.3.7 Steepest Descent Method

The steepest descent method is a useful for removing bad contacts and correcting bad geometries (Kini and Evans, 1991). It moves directly down the steepest slope on the potential energy surface, with limited changes to the molecular structure, and is most effective when the molecular system is far from a minimum. Since, the calculation does not readily converge and

can oscillate, it is often recommended that a small step-size value be chosen. The steepest descent method computes the gradient at its current location, and then travels in the opposite direction of the gradient until it reaches a minimum. The energy is computed for the initial geometry and after the movements of one of the atoms in a small increment in one of the directions of the coordinate system. This process is repeated for all the atoms which are finally moved to a new position downhill on the energy surface, due to the fact that every new step is at right angles to the one before it. This makes numerous smaller steps to proceed down along a narrow valley, and stops when a predetermined threshold condition is fulfilled. The optimization process is slow near the minimum, and consequently, the steepest descent method is often used for the structures far from the minimum as a first rough and introductory run, followed by a subsequent minimization. A more advanced algorithm i.e. conjugate gradient is employed (Kini and Evans, 1991).

4.3.8 Conjugate Gradient Method

The conjugate gradient algorithm is a primary order minimizer and varied from the steepest descent protocol, since it uses mutually the current gradient and the preceding search command to perform the process of minimization (Kini and Evans, 1991). The advantage of the conjugate gradient minimizer is that it uses the history of minimization to compute the search direction, and congregate faster than the technique of steepest descent. This is attributed to the fact that it is the first derivative rate of change of the total energy regarding atomic positions with units of the gradient are kcal mol^{-1} $Å^{-1}$. The conjugate gradient method produces an array of directions that prevail over the oscillatory behavior of the steepest descents in narrow valleys. Successive

directions are not at right angles to each other (Kini and Evans, 1991)**,** and the algorithm of conjugate gradient accumulates the information about the generated function from primary iteration to the consecutive one. In every minimization step, there is calculation of gradient and utilized as additional information for computing the vectors for new direction of the minimization procedure. Therefore, each consecutive step refines the direction until the minimum. The calculation effort as well as storage space necessity are greater as compared those required for the protocol of steepest descent, regarding the large systems, the conjugate gradient method is preferred. The expense of total computational and the long time period per iteration is compensated by the additional efficient convergence to the minimum, achieved in the case of conjugate gradients (Kini and Evans, 1991). There are numerous means in minimization of molecular structures to describe the criteria of convergence. In the case of non-gradient minimizers, simply the augmentations in the energy as well as the coordinates can be considered in order to conclude the excellence of the real geometry of the particular molecular system. However, atomic gradients are applied for this purpose among all gradient minimizers. The top method in this regard is to compute the RMS (root-mean-square) gradients of the forces on every atom of a molecule (Kini and Evans, 1991). Depending upon the objective of the minimization an appropriate value is chosen as a maximum derivative. If there is preference for the minimal relaxation of a stressed molecule, a coarse criterion of convergence like a utmost derivative of 0.1 kcal $mol^{-1}Å^{-1}$ is adequate, whereas for further cases 0.001 kcal $mol^{-1}Å^{-1}$ for a absolute minimum is satisfactory (Kini and Evans, 1991).

4.3.9 Molecular Dynamics analysis

Root mean square deviation (RMSD) is one of the most important fundamental properties for analysis whether the protein is stable and close to the experimental structure (Kuzmanic and Zagrovic, 2010). The RMSD of certain atoms in a protein with respect to a reference native structure, r_{ref} , is calculated as

$$RMSD(t) = \left[\frac{1}{M}\sum_{i=1}^{N} m_i \left| r_i(t) - r_i^{ref}\right|^2\right]^{1/2}$$

$$RMSD(t) = \left[\frac{1}{M}\sum_{i=1}^{N} m_i \left| r_i(t) - r_i^{ref}\right|^2\right]^{1/2}$$

where $M=\Sigma_i m_i$ and $r_i(t)$ is the position of atom i at time t after least square fitting the structure to the reference structure.

To calculate the average fluctuation of all residues during the simulation, the root mean square fluctuation (RMSF) of the Cα atoms of protein from the primary structure were plotted as a function of residue number (Kuzmanic and Zagrovic, 2010). The RMSF is a measure of the deviation between the position of particle i with reference position as

$$\mathrm{RMSF}_i = \left[\frac{1}{T}\sum_{t_j=1}^{T} \left|r_i(t_j) - r_i^{ref}\right|^2\right]^{1/2}$$

where T is the time and r_i^{ref} is the reference position of particle i.

The radius of gyration of a protein is a measure of its compactness (Lobanov et al., 2008). If a protein is folded in stable configuration, it will likely maintain a relatively steady value of R_g. The compactness of a structure is calculated as

$$R_g = \left(\frac{\sum_i |r_i|^2 m_i}{\sum_i m_i}\right)^2$$

Where m_i is the mass of atom i and r_i the position of atom i with respect to the center of mass of the molecule.

The total mean energy of the system was calculated as;

$$\langle E \rangle = \frac{1}{N} \sum_{i=1}^{N} E_i$$

Where, E_i is the energy of atom i.

4.3.10 Materials and Methods

The MD simulations were performed at the molecular mechanics level using GROMACS 4.6.5 simulations program (Van Der Spoel et al., 2005) and all atom functions by OPLS (optimized potential for liquid simulation). GROMACS is user-friendly, high performance research tool with topologies and parameter files written in a clear text format designed for the protein dynamics study using classical molecular dynamics theory (Pronk et al., 2013). MD simulation methods were performed on chitinase II and chitinase I at four different temperatures 300 K, 325 K, 350 K and 375 K in order to investigate its stability profile. The proteins were soaked in a

cubic box of water molecules with a dimension of 10 Å using editconf module for creating boundary conditions and genbox for solvation. The spc216 template was used to solvate the protein. The charges on the protein were neutralized by the addition of Na^+ and Cl^- ions to maintain neutrality. The system was then minimized using 1500 steps of steepest descent. The temperature of all the systems was subsequently raised from 0 to 300 K, 325 K, 350K and 375 K during their equilibration period (100 ps) at a constant volume under periodic boundary conditions. The equilibration was achieved in two phases NVT ensemble (constant number of particles, volume, and temperature at 100 ps) and NPT ensemble (constant number of particles, pressure, and temperature at 100 ps). After the equilibration phase, the particle-mesh Ewald method (Horng et al., 2005) was applied, and the production phase consisting of 10 ns was performed at 300 K, 325 K, 350 K and 375 K. All the resulting trajectories were analyzed using g_rms, g_rmsf, g_energy and do_dssp utilities of GROMACS. The output was plotted using GRACE (GRaphing Advanced Computation and Exploration of data) plotting tool present in Linux operating system. All the graphic presentations of the 3D model were prepared using Discovery studio 4.0 and VMD (Visual Molecular Dynamics) (Humphrey et al., 1996). The step by step protocols are given below (**Figure 4.10**).

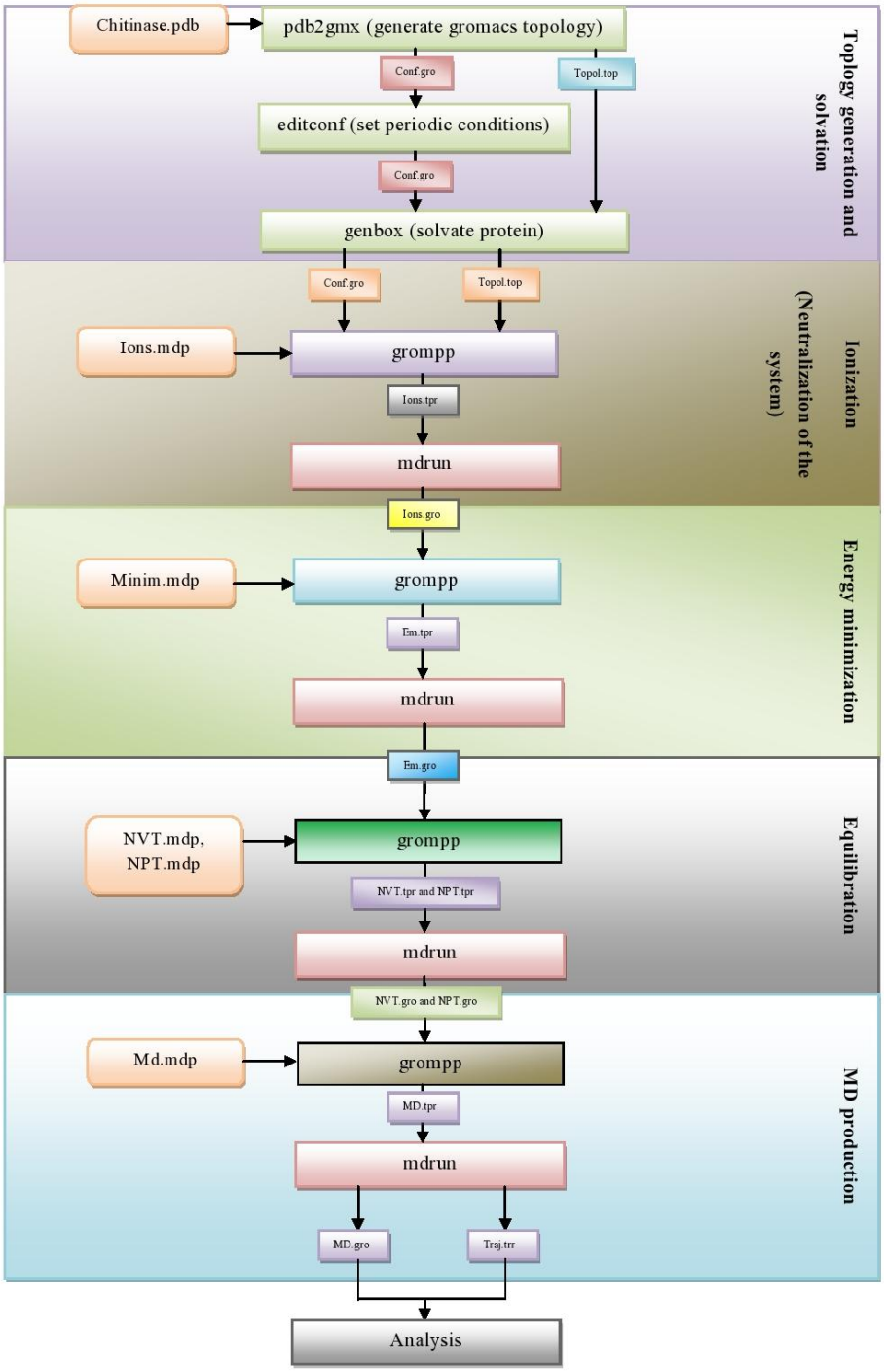


Figure 4.10: The computational workflow for the molecular dynamics simaution using GROMACS package

The structure must be of high resolution. There should not be any ligands and missing residue. Visualize the structure using VMD (Humphrey et al., 1996, Hsin et al., 2008) and Pymol (Seeliger and de Groot, 2010) to remove water and ligand if present. The residue HOH in PDB file was removed. The PDB file must have only protein atoms, for the input into the first GROMACS module, pdb2gmx. Incomplete sequences or some amino acid residues that have missing atoms will cause pdb2gmx to fail. PDB structure of chitinase was refined using scwrl4. The purpose of pdb2gmx is to create the topology of molecule, a post-processed structure file and a position restraint file using "OPLS-AA/L" (Optimized Potential for Liquid Simulations/ all atoms) force field. Once the GROMACS pdb2gmx module generates the topology file, the protein was solvated in aqueous system. There are two steps to define the box and filling it with solvent:

1. Define the box dimensions using the editconf module.
2. Fill the box with water using the solvate module (formerly called genbox).

The cubic boundary conditions were defined with a diameter of 1.0 nm. After defining the dimensions of the solvation box, the whole system was solvated using SPC/E216 water model. The solvated system contains a charged protein. In order to neutralize the system ions are added using the information available through pdb2gmx. By utilizing the molecular dynamics parameter file (.mdp), the genion module neutralizes the system. The neutralization was carried out using steepest descent algorithm and Verlet cut off scheme. After establishing neutralized solvated system, the energy minimization was performed in order to combat inappropriate geometry and structure distortion. The maximum number of 50000 steepest descent steps was used to perform energy minimization. This process was executed until a maximum force (<1000

kJ/mol/nm) is obtained. Since, the minimized structure of the chitinase was obtained it was then equilibrated using various ensembles. Equilibration of solvent and ions around the protein is required because the solvent is mostly optimum within itself and not with the protein. During equilibration, the system is brought to the temperature designated for MD simulation to establish proper orientation. The first equilibration was performed at a constant number of particles, volume and temperature (NVT) which is also known as canonical or isothermal-isochoric conducted at 100 ps at different temperatures (**Figure 4.11 A, B** and **C**).

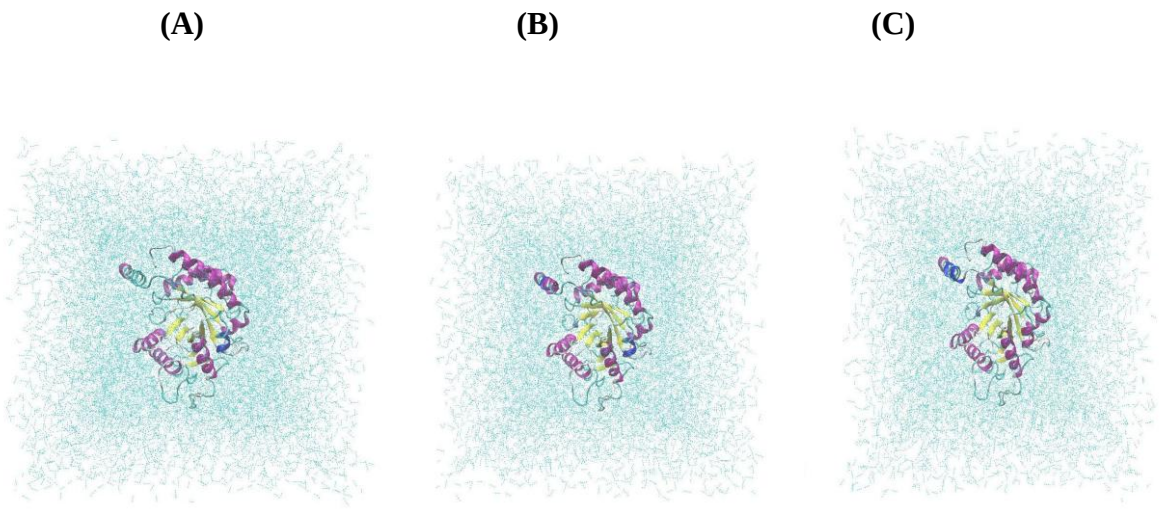


Figure 4.11: The chitinase II molecules immersed in water is subjected to NVT equilibration at (A) 300 K, (B) 325 K and (C) 350 K.

Similarly, in order to stabilize the density (pressure) of the system, equilibration was performed at a constant number of particles, pressure and temperature (NPT) conducted at 100 ps which is also known as isothermal-isobaric ensemble (**Figure 4.12 A, B** and **C**).

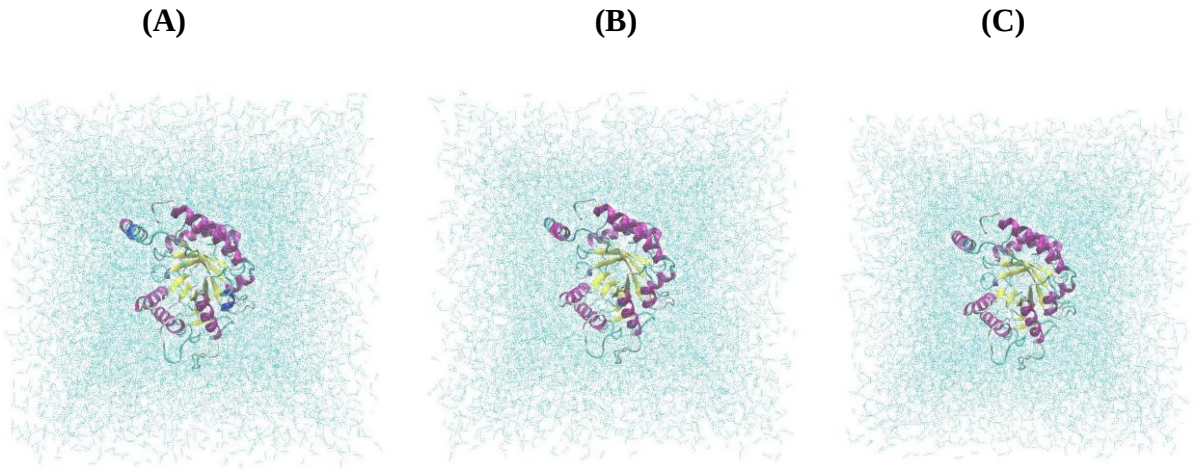


Figure 4.12: The chitinase II molecule immersed in water is subjected to NPT equilibration at (A) 300 K, (B) 325 K, and (C) 350 K.

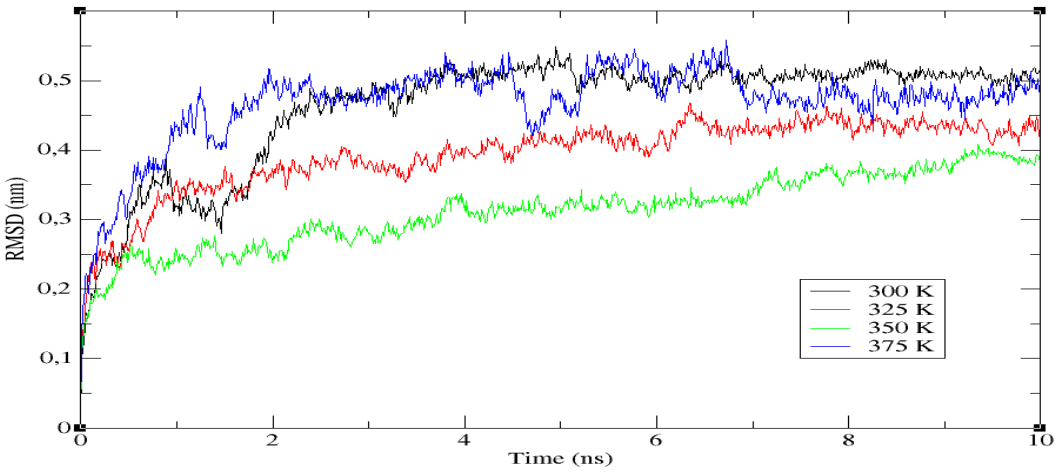
This produced an equilibrated system with respect to temperature and pressure. Finally MD simulations were performed at 10 ns on GPU nodes using mdrun module available at the Center for High performance Computing (CHPC) in Cape Town (Appendices I-VI).

4.3.11 Results and Discussion

4.3.11.1 Root mean square deviation

The MD simulation of chitinase II displayed an acceptable stability profile at a temperature of 300 K, 325 K and 350 K respectively, while fluctuations were observed at 375 K. The relatively stable nature of the protein was observed for all temperatures by average total energy values. Similarly, the RMSD values were found to be around 0.5 nm, 0.4 nm and 0.3 nm respectively, for temperatures ranging from 300-350 K (**Figure 4.13 (A)**). The average RMSD values of the chitinase II structure was found to be low at 350 K as compared to 325 K and 300 K. Initial fluctuations were observed at 300 K and thereafter a constant equilibrium was achieved. The slight fluctuations were observed for 325 K and 350 K suggests the stable nature of the protein at higher temperatures. Similarly, the MD simulation of chitinase I displayed a constant RMSD values (0.25 nm) at a temperature of 300 K, 325 K and 350 K respectively (**Figure 4.13 (B)**).

(A)



(B)

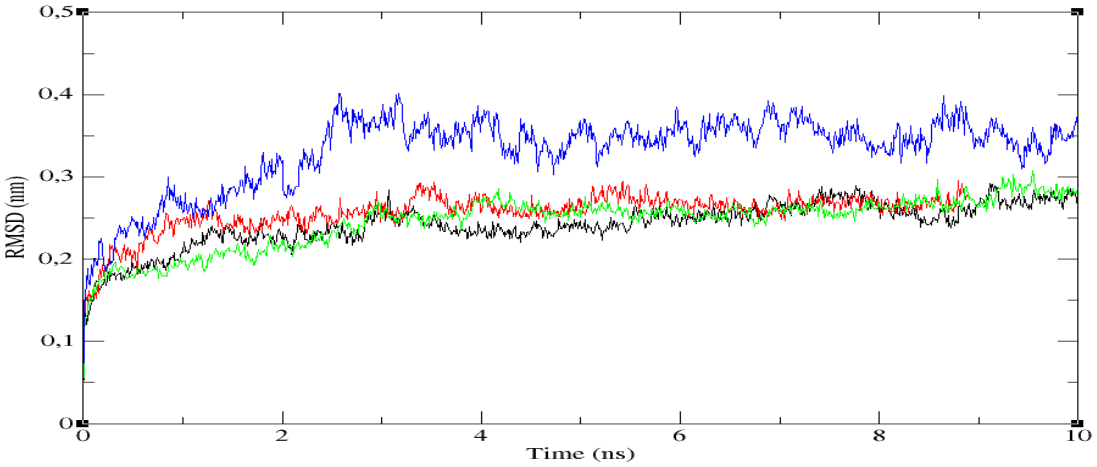
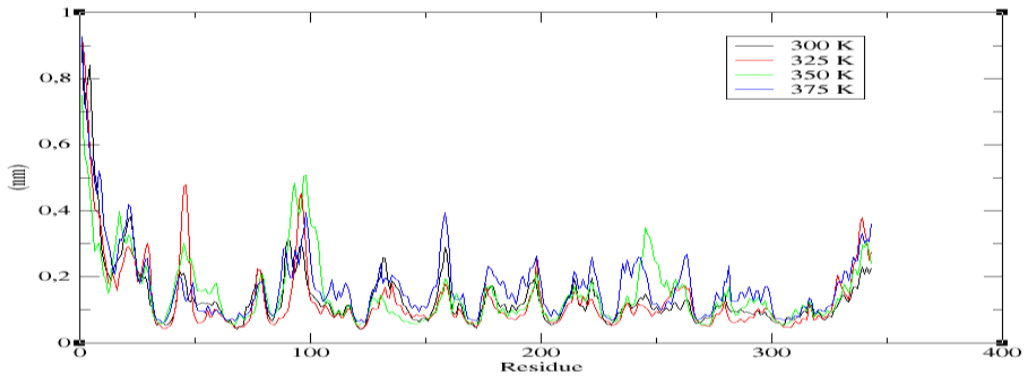


Figure 4.13: RMS deviation values for (A) chitinase II and (B) chitinase I at different temperatures as a function of time. Black, Red, Green and Blue colour represent RMS deviations obtained from 300 K, 325 K, 350 K and 375 K respectively. The duration of each simulation is 10 ns.

4.3.11.2 Root mean square fluctuation

The average fluctuations of each residue were evaluated during the course of the simulation using the g_rmsf module. The RMSF of the Cα atoms of chitinase II from the initial structure were plotted as a function of residue number (**Figure 4.14 (A)**). The RMSF of the Cα value of the chitinase II structure was comparable for all temperature except 375 K. At 375 K, the residues present in the chitnase II showed high fluctuations. At 325 K, the loop forming residues ranging from 40-50 showed higher fluctuations. Similarly, the region spanning residues 80 to 100 showed higher values of RMSF at 325 K and 350 K, while the coiled region of 240–250 residues displayed a higher value at 350 K. However, the enhanced localized flexibility in the loop regions suggested a greater stability of the chitinase II enzymes at higher temperatures. Similarly, chitinase I showed higher values of RMSF at 375 K (**Figure 4.14 (A)**).

(A)



(B)

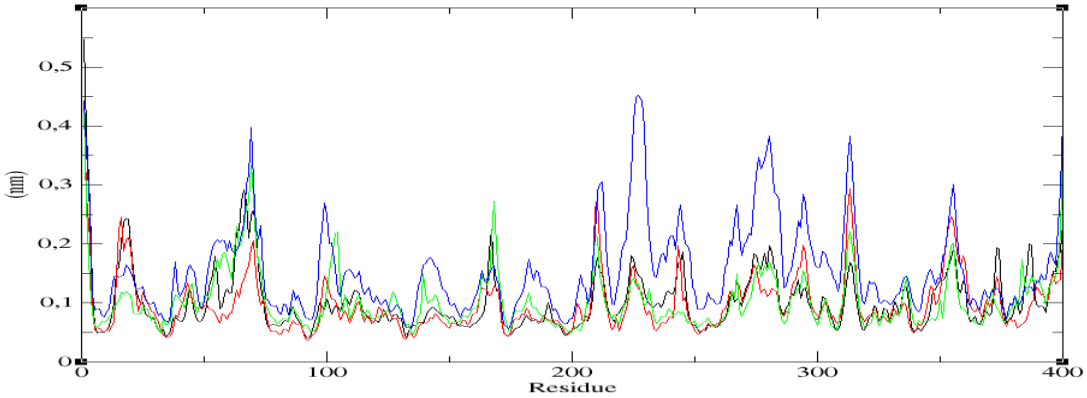
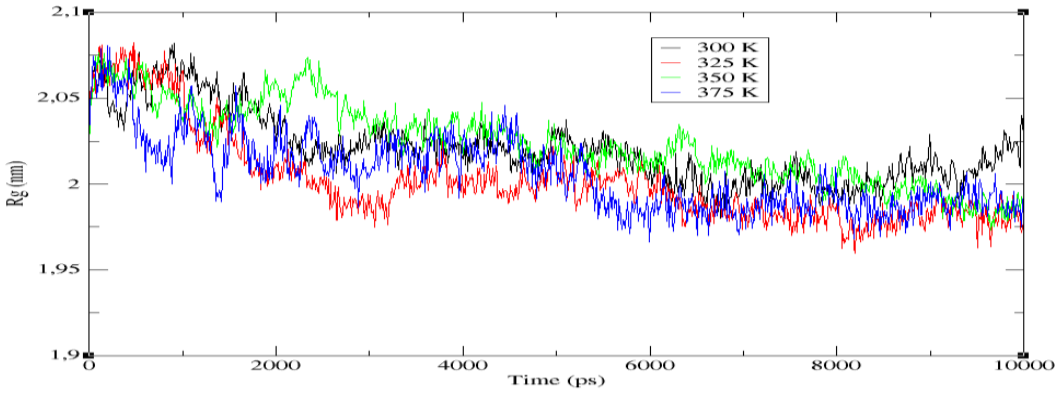


Figure 4.14: Average RMS deviation values as a function of amino acid sequence numbers for (A) chitinase II and (B) chitinase I. Values were calculated with the use of Cα atoms. Black, Red, Green and Blue colour represent RMS fluctuation obtained from 300 K, 325 K, 350 K and 375 K respectively.

4.3.11.3 Radius of gyration

The R_g values suggests that the chitinases remain stable in its compact (folded) form over the course of simulations, with little fluctuations at 300 K, 325 K and 350 K (**Figure: 4.15**). The Rg values for temperatures 300 K-350 K were comparable for both the proteins. However, the chitinase I showed higher fluctuations values of Rg than chitinase II at 375 K.

(A)



(B)

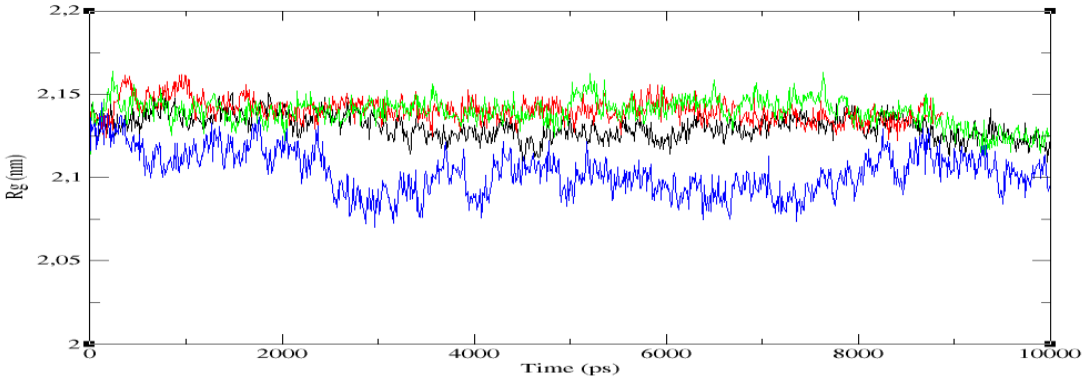


Figure 4.15: The radius of gyration values as a function of time for (A) chitinase II and (B) chitinase I. Values were calculated with the use of Cα atoms. Black, Red, Green and Blue colour represent radius of gyration obtained from 300 K, 325 K, 350 K and 375 K respectively.

4.3.11.4 Secondary Structure

The secondary structures obtained during the MD simulations are highlighted by the do_dssp module of GROMACS at different temperatures in (**Figure 4.16 and 4.17**). This clearly shows the global conformation of chitinases, which is conserved with no significant changes in its

secondary structure elements at higher temperatures (300 K – 350 K). However, large fluctuations were observed in the structural elements of chitnases at 375 K.

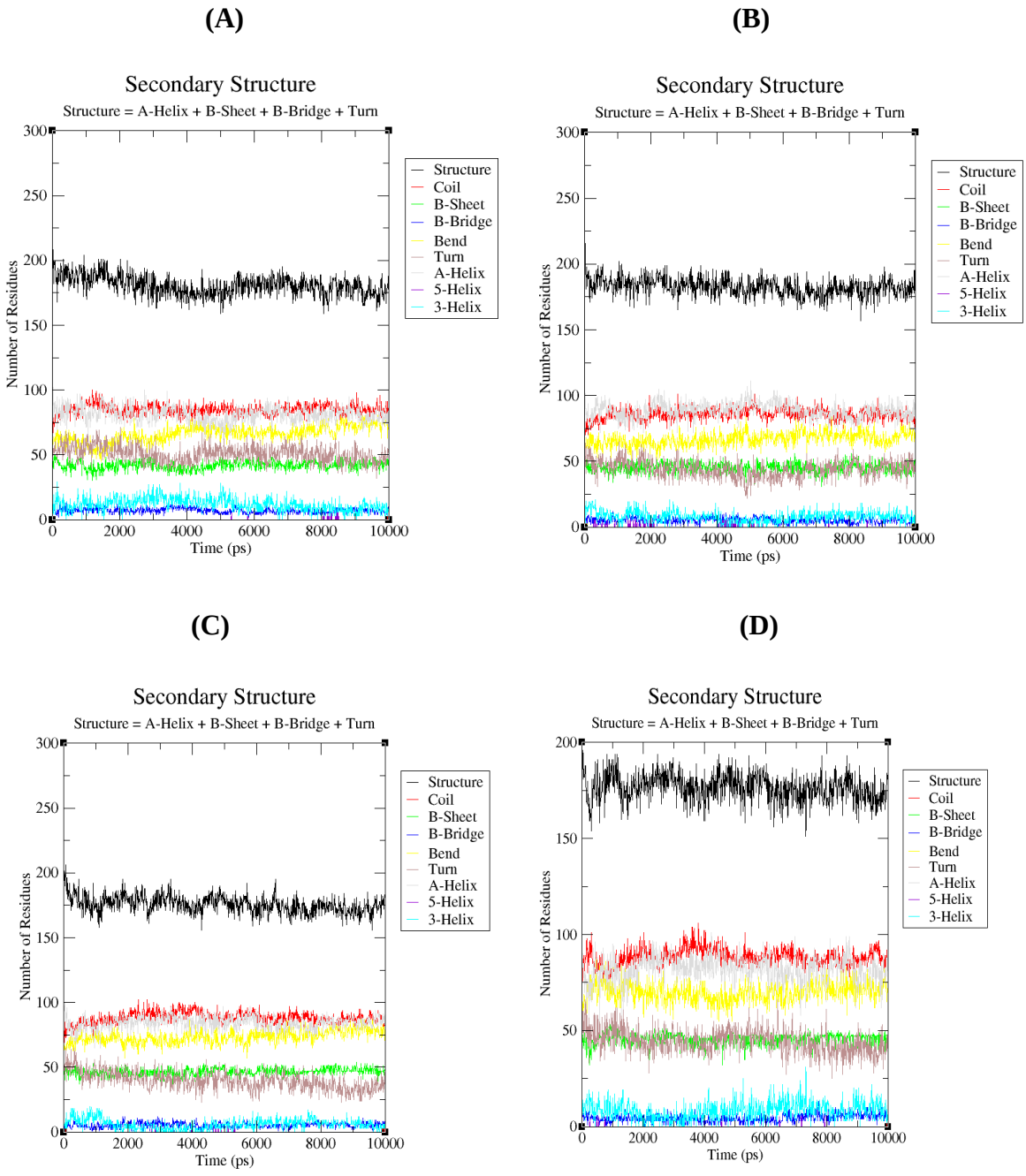


Figure 4.16: The graphical representation showing the stability of structural elements of chitinase II at (A)300 K, (B)325 K, (C)350 K, and (D)375 K. The secondary structure elements were comparable at 300 K-350 K.

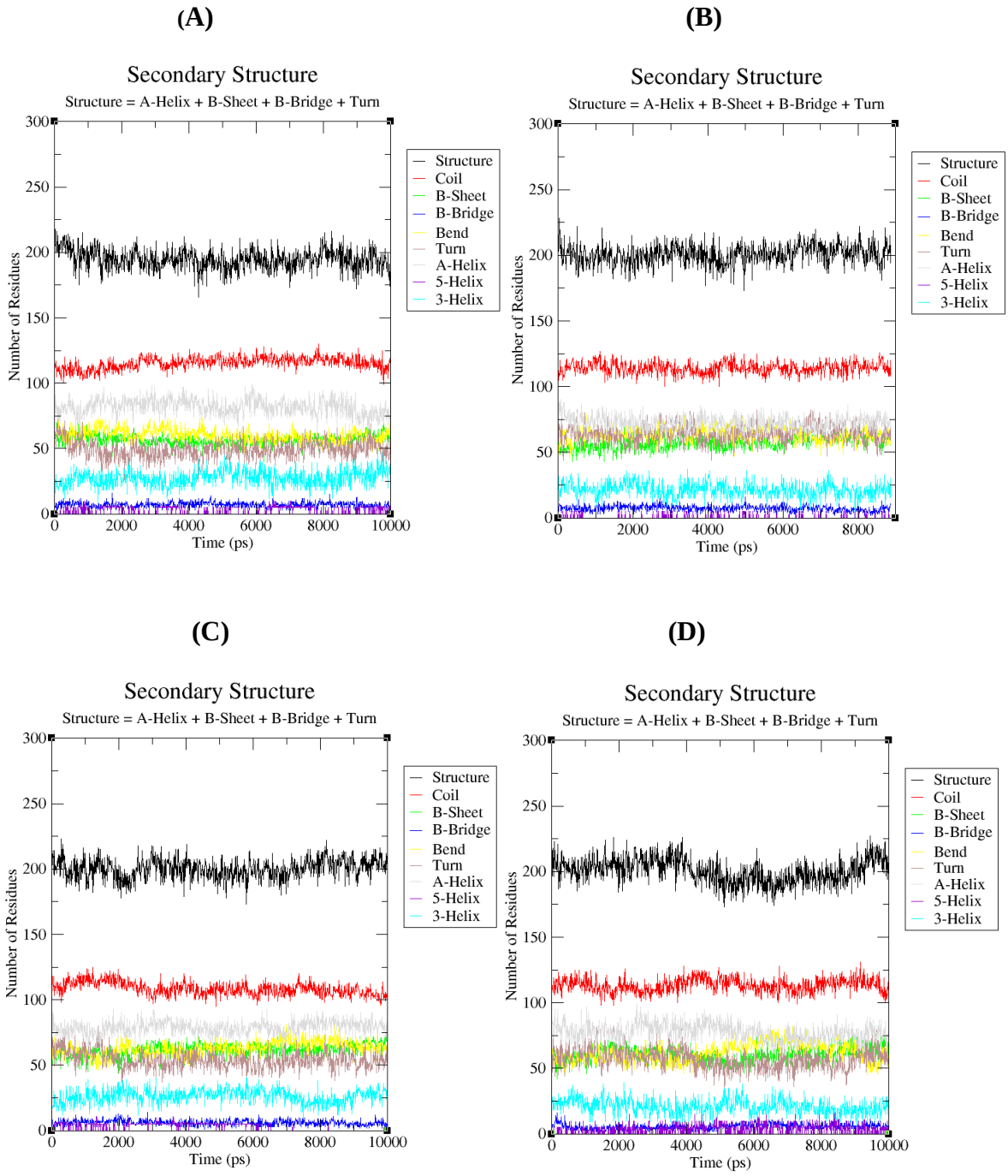


Figure 4.17: The graphical representation showing the stability of structural elements of chitinase I at (A)300 K, (B)325 K, (C)350 K, and (D)375 K. The secondary structure elements were comparable at 300 K-350 K.

4.3.11.5 Total energy of system

The relative energy for of chitinase II was found to be around -1002760 kJ/mol, -951579 kJ/mol and -1003070 KJ/mol at 300 K, 325 K and 350 K respectively. While the relative energy of chitinase I was found to be around -1002550 kJ/mol, -1003210 kJ/mol and -1003130 kJ/mol showing relatively stable nature of protein at 300 K, 325 K and 350 K respectively. The relative energy of chitinase II and chitinase I were found to less at 350 K and 325 K respectively.

4.3.12 Conclusions

The MD simulation of chitinases showed an increase in the stability over a temperature ranging from 300 K-350 K, while at 375 K it showed a decrease in the stability. The calculated RMSD value between the processed and the starting structures of chitinase II at 300 K, 325 K and 350 K were 1.950 Å, 1.315 Å and 1.718 Å, respectively. While for 375 K the RMSD value showed an increase in the magnitude of 2.815 Å, in contrast to the above calculated values. Likewise, the chitinase I also showed an increase in the stability over a temperature ranging from 300 K-350 K. This indicated that the chitinases were able to maintain its stability at temperatures ranging from 300-350 K, but decreased as the temperature changed to 375 K. The previous study on thermostable chitinases suggested its maximum stability at low pH (Meng Zhang, 2014). In order to validate these results, constant pH MD simulations were performed to establish the relationship between the physiological pH with the structural and functional aspect of the chitinases enzymes as discussed in section 4.4.

4.4 Constant pH simulations

The pH determines the charges of the ionizable residues, exhibiting a direct influence on the intra-molecular electrostatic interactions thereby affecting the molecular structure (Flanagan et al., 1981). The pH-dependent protein stability is a critical factor for protein folding and thus it is a key determinant in various molecular mechanisms. When a protein folds, the titratable amino acids in the protein molecule are modified, depending upon the nature of buried amino acids in the protein with no exposure to the solvent. Specifically, the pKa values of the amino acid side chains play a vital role in defining the pH-dependent characteristics of a protein (Song et al., 2009). Literature studies revealed that the pH has a dramatic impact on biomolecules (Matthew et al., 1985, Yang and Honig, 1993), therefore pH-dependent molecular dynamics simulations would provide a deeper insight into the structural and functional aspects of the enzyme (Kim et al., 2014). The recently developed constant pH MD methods directly capture the protonation of residues, thereby increasing the realism of MD simulations. Accordingly in this study the MD simulations were performed with a fixed protonation state of the enzymes (Machuqueiro and Baptista, 2006). The prime objective of this study was to evaluate the effect of the physiological pH on chitinase I and II enzymes obtained from *Thermomyces lanuginosus*. Accordingly, the pKa values of the titratable groups were obtained by modulating the protonation state of the various titratable functional groups present in the system (Burgi et al., 2002). The generated change in charge distribution showed a profound impact on the structure and function of the enzymes. The MD simulations were performed in explicit water solvent at different pH conditions. This study was performed to establish the relationship between the physiological pH with the structural and functional aspect of the chitinase I and II enzymes. The results demonstrate the influence of pH on protein structure and dynamics. MD simulations were useful

to study the deviation of secondary structure elements present in chitinases at each pH. However the results detected small structural changes at low pH. To the best of my knowledge, constant pH MD simulations were performed first time on thermostable chitinases.

4.4.1 Materials and Methods

4.4.1.1 Protein Ionization and Residue pK calculations

pK_a values of the amino acid side chain aid in defining the pH-dependent characteristics of a protein. The pH-dependent activity and protein stability displayed by the enzymes are the properties that are determined by the pKa values of amino acid side chains. The process of acid dissociation is given by:

$$AH \rightarrow A^- + H^+$$

Where AH is an acid, A^- is the conjugated base and H^+ is the proton. In this case the equilibrium constant (Ka) is defined as:

$$K_a = [A^-]\,[H^+]/\,[AH]$$

The equilibrium constant have orders of magnitude for different acids, the negative decimal logarithm (pK_a) is often used:

$$pKa = pH - \log [A^-]/[AH]$$

This is known as the Henderson-Hasselbalch equation and is used to calculate the ratio of protonated and deprotonated species for a given pH. This equation is correct only if there is one type of acid in solution. However in protein, ionizable amino acid residues interact with each other and often their behaviour is described more accurately by Hill equation:

$$pKa = pH - n \log [A^-]/[AH]$$

where n is the Hill coefficient and it is measure of the cooperativity of the system: $n < 1$ indicating negative cooperativity and $n > 1$ positive cooperativity. The equilibrium constant, of a chemical reaction, depends on its free energy. The relationship between the standard free energy of deprotonation and the pKa of the acid, is given by:

$$\Delta G_o = -2.3RT\ pK_a$$

The free energy difference of the system being protonated or deprotonated at a given pH is given by:

$$\Delta G = -2.3RT\ pKa + 2.3RT\ pH$$

and the pKa obtained:

$$pKa = -\Delta G/\ 2.3RT + pH$$

In order to obtain the pKa, one must calculate the free energy of deprotonation. These free energy can be calculated using methods developed to work with MD simulations. The combination of the Generalized Born (GB) and Iterative Mobile Clustering (IMC) approaches provided new computational protocols for pK calculations effect of protein ionization. The pK_a and titration curves for each one of the titratable amino acid residues present in chitinase I and chitinase II were calculated using modules present in Discovery studio 4.0 (Spassov and Yan, 2008). The chitinases structures were prepared and typed by the CHARMm and CHARMm Polar hydrogen force field and the protein structures were cleaned. The total charge of the protein as a function of pH and electrostatic contribution to the free energy as a function of pH were calculated. The isoelectric points (pI) of the chitinases were predicted at a particular pH.

4.4.1.1 Constant pH MD simulations

A constant pH MD simulation method using GROMACS package developed by Baptista (Machuqueiro and Baptista, 2007) was used. The change in the protonation states were performed using the pdb2gmx module of GROMACS 4.6.5 packages (Van Der Spoel et al., 2005) by utilizing the information obtained from DS. These generated protonation states were used as inputs for molecular dynamics simulations (Van Der Spoel et al., 2005, Witham et al., 2011). The MD simulations for the protonated states of chitinase I and chitinase II were set at a pH range of 2-6. The MD simulations were performed at a molecular mechanics level using GROMACS (Van Der Spoel et al., 2005) and the all atom functions by OPLS (optimized potential for liquid simulations). The proteins were soaked in a cubic box of water molecules with a dimension of 10 Å using the editconf module for creating boundary conditions and the genbox module for solvation. The spc216 template was used to solvate the protein. The charges on the protein were neutralized using the verlet cut-off scheme by the addition of Na^+ and Cl^- ions to maintain neutrality. The system was then minimized in order to combat inappropriate geometry and structure distortion using 1500 steps of steepest descent. This process was executed until a maximum force (<1000 kJ/mol/nm) was reached. The temperatures of all the systems were subsequently raised from 0 to 300 K during their equilibration period (100 ps) at a constant volume under periodic boundary conditions. The equilibration was achieved in two phases using the NVT and NPT ensembles (constant number of particles, volume, and temperature at 100 ps). After the equilibration phase, the Particle-Mesh Ewald summation method (Horng et al., 2005) was applied, and the production phase consisting of 20 ns were performed. The resulting trajectories were analyzed using g_gyrate, g_rms, g_rmsf, g_energy, g_hbond and do_dssp utilities of GROMACS. The outputs were analysed using X-MGRACE

(GRaphing Advanced Computation and Exploration of data) plotting tool. All the graphic presentations of the 3D model were prepared using Discovery studio 4.0 and Visual Molecular Dynamics (Humphrey et al., 1996).

4.4.2 Results and Discussions

The total titration curves of chitinase I and chitinase II were obtained by calculating the average net charge at each pH value (**Figure 4.18**). The net charge and isoelectric pH (pI) of a protein depend on the content of ionizable groups and their pKa values.

(A) (B)

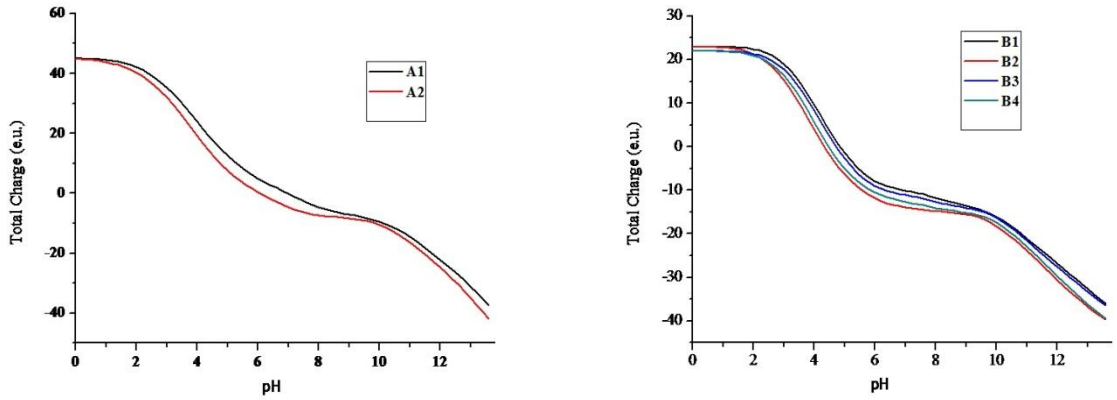


Figure 4.18: The pH dependence of total charge at different protonated conditions of (A) chitinase I, A1 and A2 represents CHARMm and CHARMm Polar H force fields respectively (B) chitinase II, B1, B2 represents chitinase II protein with CHARMm and CHARMm Polar H force fields respectively, while B3, B4 represent chitinase II without signal peptide typed with CHARMm and CHARMm Polar H force fields respectively.

The maximum stability of chitinase I and chitinase II using CHARMm and CHARMm Polar H force fields were obtained at pH ranging from 3.6 to 4.0 and 2.2 to 3.8 respectively (**Figure 4.19**) and the isoelectric point of 6.08 to 6.98 and 4.35 to 4.87 respectively (**Table 4.3**).

(A) (B)

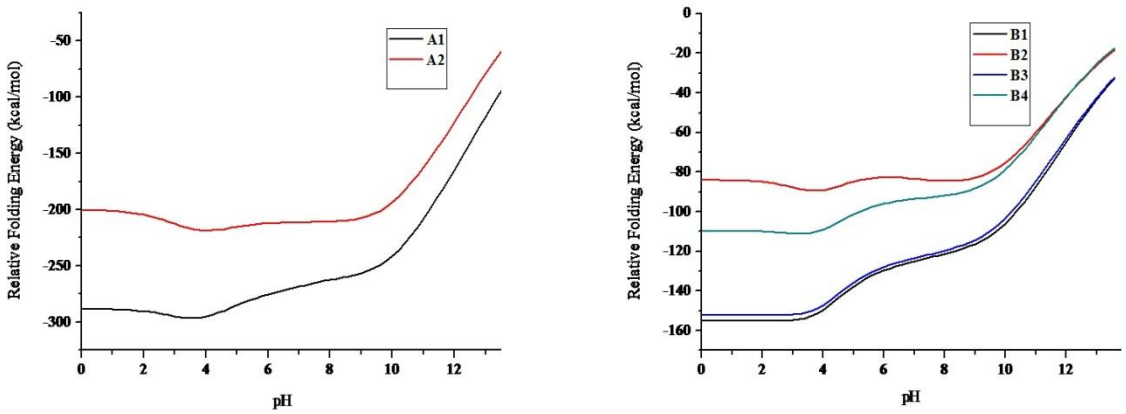


Figure 4.19: Relative Folding energy at different protonated conditions. (A) chitinase I, A1 and A2 represents CHARMm and CHARMm Polar H force fields respectively (B) chitinase II, B1, B2 represents chitinase II protein with CHARMm and CHARMm Polar H force fields respectively, while B3, B4 represent chitinase II without signal peptide typed with CHARMm and CHARMm Polar H force fields respectively

Table 4.3: Calculation of protein ionization and pKa value for chitinase I and II.

	Chitinase I		**Chitinase II**		**Chitinase II (without signal peptide)**	
Forcefield	**CHARMm**	**CHARMm Polar H**	**CHARMm**	**CHARMm Polar H**	**CHARMm**	**CHARMm Polar H**
Maximum Stability at pH	3.6	4	2.2	3.8	2.6	3.2
Isoelectric Point	6.98	6.08	4.8700	4.35	4.74	4.49
Electrostatic_Energy (kcal/mol)	-7962.0	-7728.8	-6895.7	-6666.3	-6490.0	-6245.5

The total charge at a particular pH was found to be less when typed with CHARMm Polar H force fields while the relative folding energy was found to be less when typed with CHARMm force fields. No significant changes were observed for both the models of chitinase II in the plot

of total charge vs pH. This suggests that the amino acids present in the signal peptide region of chitinase II have little or no influence on pH. Electrostatic interactions are widely believed to be primary factors upon which the pH-dependent phenomena are dependent. Since the charge carried by protein is pH-dependent, the electrostatic interaction between residues of the protein would be modulated by pH. The Electrostatic energy of chitinase I and II were found to be less at low pH indicating that the chitinases are stable at low pH (**Figure 4.20**) (Meng Zhang, 2014).

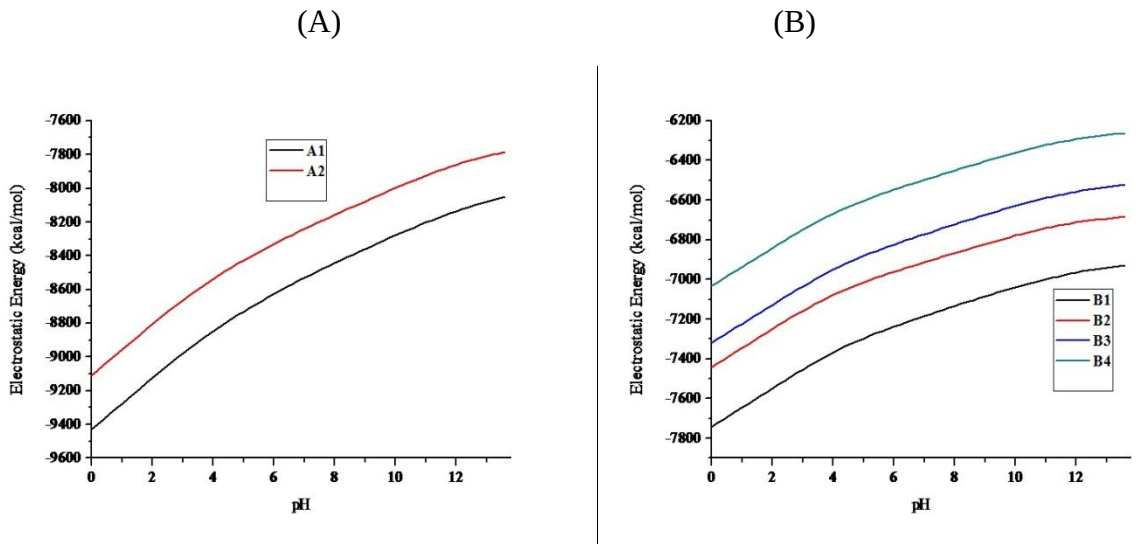
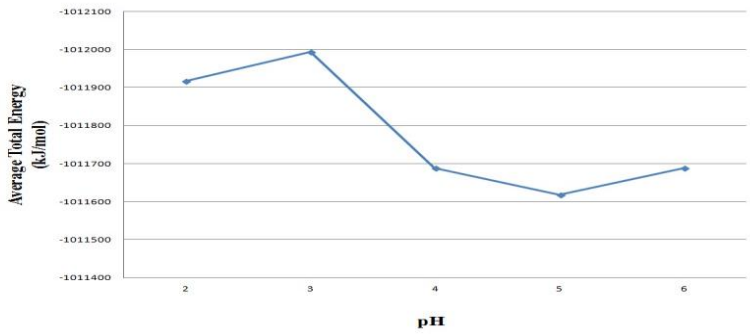


Figure 4.20: Electrostatic Energy at different protonated conditions. (A) chitinase I, A1 and A2 represents CHARMm and CHARMm Polar H force fields respectively (B) chitinase II, B1, B2 represents chitinase II protein with CHARMm and CHARMm Polar H force fields respectively, while B3, B4 represent chitinase II without signal peptide typed with CHARMm and CHARMm Polar H force fields respectively.

The pdb2gmx module of GROMACS generated the desirable protonation states by altering the configuration of amino acids. Within the protein the respective titratable residues were selected on the basis of information obtained from titration analysis. It was observed that chitinases

attained more stable conformation at low pH. Ranges of pH 2-6 were selected for 20 ns simulations. The purpose of the energy value calculation was to ensure the stability of the protein during the MD simulations at a constant pH and temperature. The constant average fluctuation of temperature around 300 K suggested stable and accurate nature of the MD simulations performed. The results suggest that chitinase I and chitinase II showed a low value of total relative energy of -1011618 kJ/mol and -1452083 kJ/mol respectively at pH 5 (**Figure 4.21**).

(A)



(B)

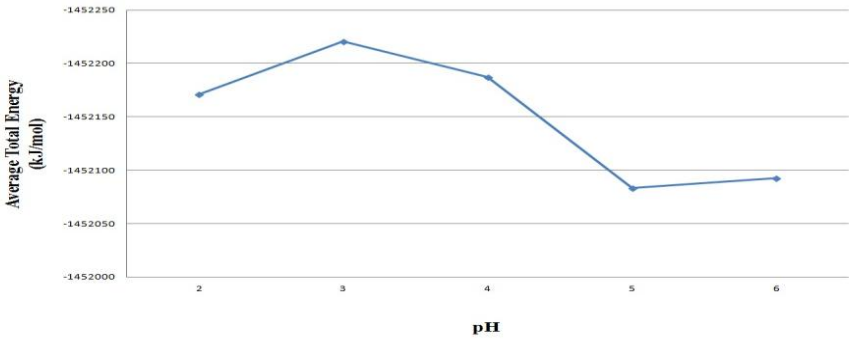


Figure 4.21: The Average total energy values as a function of pH for (A) chitinase I and (B) chitinase II. The values obtained from pH 2-6 respectively.

The Radius of gyration (Rg) analysis provided insight into the compactness of the protein during the simulations (Lobanov et al., 2008). Rg is defined as the root mean square distance of the

collection of atoms from their common centre of gravity. The calculated average Rg values ranged from 2.10 - 2.13 nm for chitinases I. The Rg plot suggested that chitinase I remained in its compact state at pH 2, pH 3, pH 5 and pH 6, while at pH 4 protein loses its compactness after 10 ns simulations (**Figure 4.22 A**). The calculated Rg values for chitinase II were found to be 2.03-2.12 nm during the simulations. pH vs time scale plot for chitinase II suggested that the protein showed fluctuations throughout the simulations. The chitinase II acquire a stable Rg value after 10 ns at pH 3 (**Figure 4.22 B**).

(A) (B)

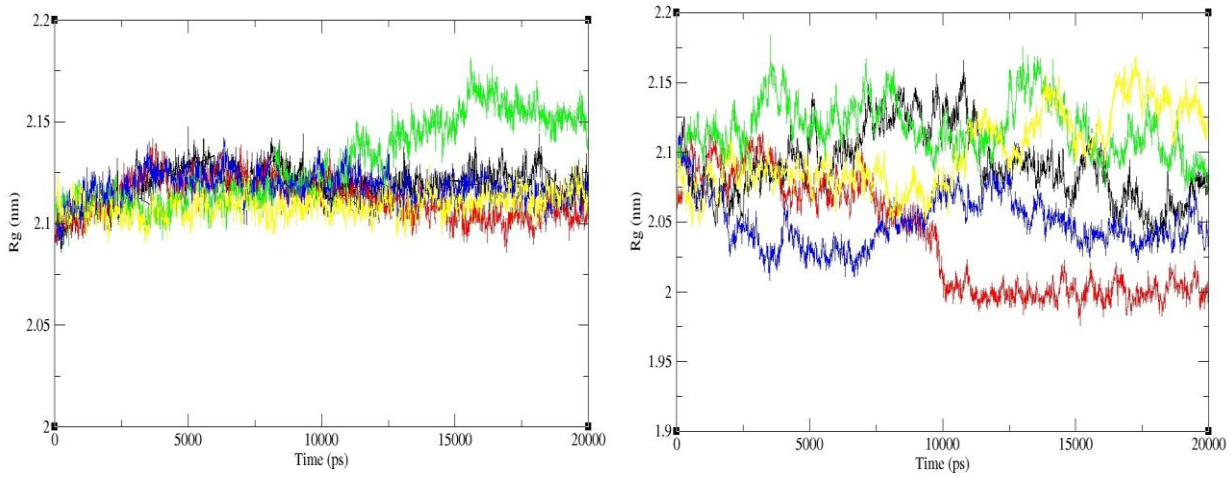


Figure 4.22: The radius of gyration values as a function of time for (A) chitinase I and (B) chitinase II. Values were calculated with the use of Cα atoms. Black, Red, Green, Blue and Yellow colour represent radius of gyration obtained from pH 2-6 respectively.

The Root mean square deviation (RMSD) is a measure of the deviation of the conformational stability of the proteins from backbone structure of the initial starting structure (Kuzmanic and

Zagrovic, 2010). The observed RMSD plots were obtained from molecular dynamics simulations. The constant RMSD values between 0.16-0.19 nm indicate the stable nature of chitinase I. The 20 ns MD simulations showed that the chitinase I has a less average RMSD value at pH 3 (0.16 nm), pH 5 (0.17 nm) and pH 6 (0.16 nm). From the plot, the RMSD values were found to be constant and stable throughout the time scale at pH 5 than pH 3 and pH 6. This indicates that chitinase I acquires high stability at pH 5 while, high fluctuations were observed at pH 2. The average RMSD values for chitinase II were found to be in the range of 0.35 – 0.66 indicating that pH changes exhibit more effect on the structure. Chitinase II showed a lesser RMSD value at pH 2 followed by pH 5 (**Figure 4.23**).

(A) (B)

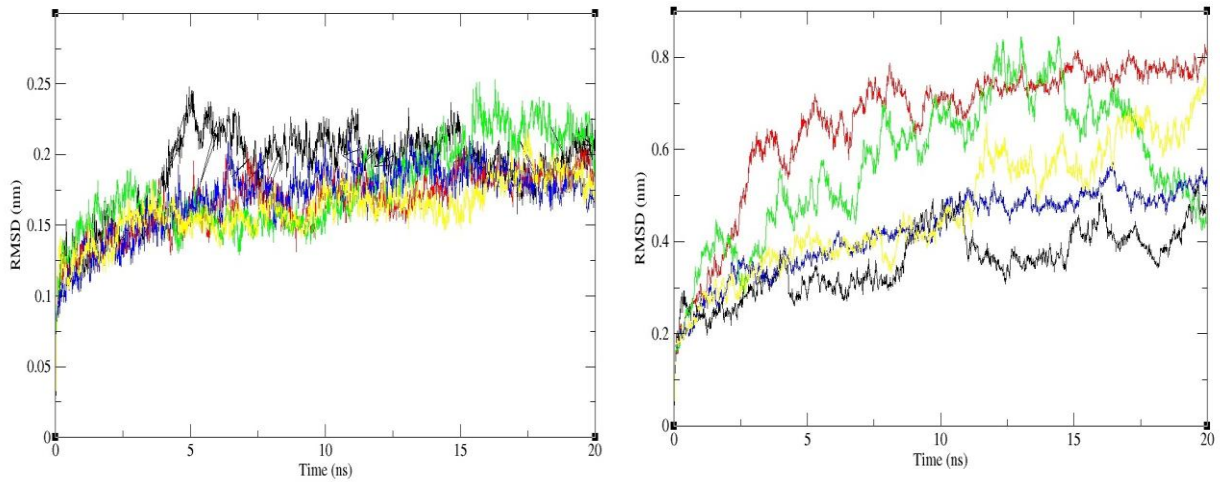


Figure 4.23: RMS deviation values for (A) chitinase I and (B) chitinase II at different pH as a function of time. Black, Red, Green, Blue and Yellow colour represent RMS deviations obtained from pH 2-6 respectively. The duration of each simulation is 20 ns.

The root mean square fluctuations (RMSF) with reference to the residues were calculated for the protonated enzymes during simulations. Our results show that chitinase I at pH 3, pH 5 and pH 6

have a less and comparable average RMSF value, but a slightly high deviation in the α-helical region of residue 50-60 at pH 3 were found (**Figure 4.24**). However, high and unstable fluctuations were observed at pH 4 in the regions spanning residues 40-70. The RMSF values for most of the residues lie below 0.1 nm at pH 5 indicating the stable nature of chitinase I. The RMS fluctuations for chitinase II were comparable at all pH except pH 4 indicating that the chitinase II is active at wide range of pH.

(A) (B)

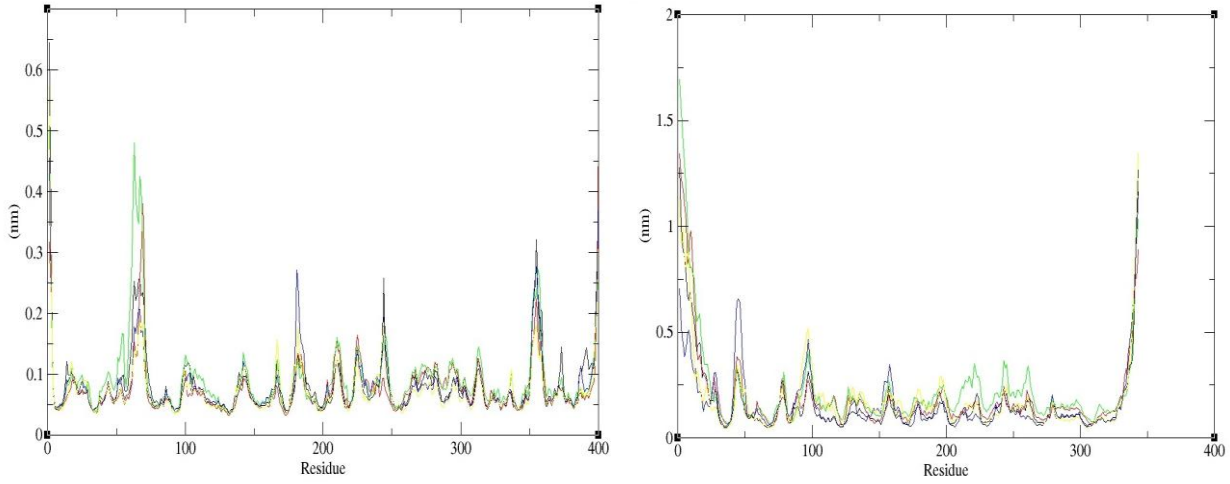
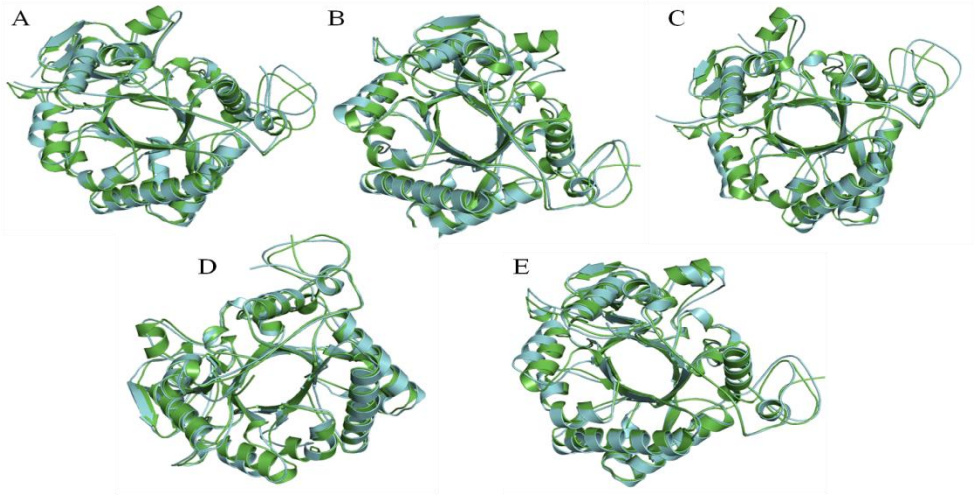


Figure 4.24: Average RMS deviation values as a function of amino acid sequence numbers for (A) chitinase I and (B) chitinase II. Values were calculated with the use of Cα atoms. Black, Red, Green, Blue and Yellow colour represent RMS fluctuation obtained from pH 2-6 respectively.

The TIM-Barrel conformation of chitinases were maintained during the 20 ns time scales of simulations. The RMS deviations of native and simulated structures of chitinase I at pH 2-6 were found to be 1.174 Å, 1.134 Å, 1.359 Å, 1.065 Å and 1.277 Å respectively (**Figure 4.25 A**), while chitinase II have 2.029 Å, 1.924 Å, 2.086 Å, 1.708 Å and 2.108 Å respectively at pH 2-6. The

chitinase I showed lowest RMS deviation at pH 5 while chitinase II showed lowest deviations at pH 3 and pH 5 (**Figure 4.25 B**).

(A)



(B)

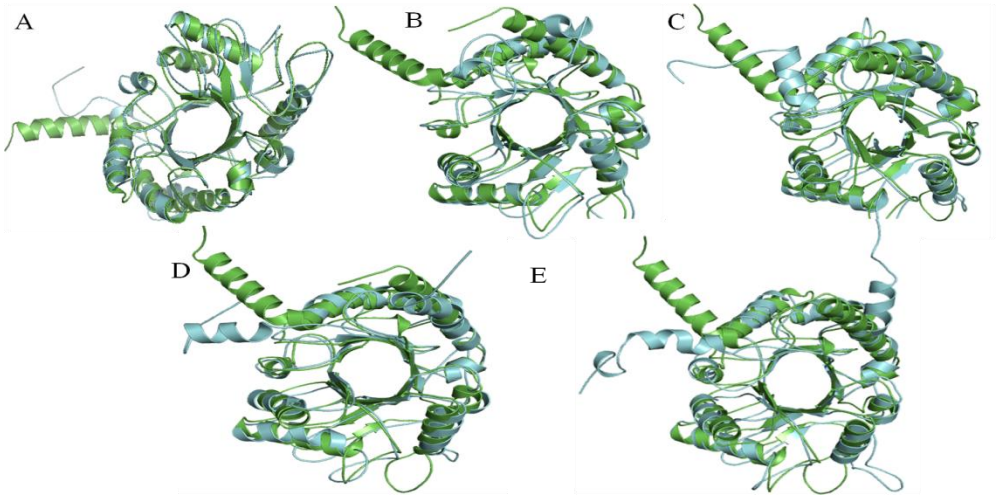


Figure 4.25: The superimposition of native and simulated structures of chitinase I and II at pH 2-6, 300 K. Green and Blue represents native and the simulated chitinases structures respectively. Thickness of the coil represents the fluctuations of heavy atoms around the average structure.

The hydrogen bond is a significant factor in stabilizing protein conformations (Fleming and Rose, 2005). The average number of hydrogen bonds between the main chain and side chains within the protein during each steps of the simulations were counted. The numbers of internal hydrogen bonds vary as pH changes. The average number of hydrogen bonds in the main chain-side chain were found to be comparable at pH 3 (45) and pH 5 (44), while pH 2, pH 4 and pH 6 have less, indicating that chitinase I exhibits stable folding at pH 3 and pH 5. The average number of hydrogen bonds in chitinase II varies between 27 to 29 at pH range 2-6.

4.4.3 Conclusions

The structural effects arising from pH induced changes on chitinase I and chitinase II enzymes have been performed using molecular dynamics simulations. This is considered to be a major factor in a solvent for the proper functioning of enzymes. We presented a titration curve and pK calculations of the titratable sites at a particular pH. Our results suggested that the maximum stability and optimal activity of chitinase I and II lies at an acidic pH in accordance with experimental results. The chitinases remained in their close TIM barrel conformations during the simulations. The superimposition of simulated structures and native structures of chitinase I and II showed less RMS deviation at pH 5. We also describe the results from simulations focusing on how the conformational change is linked to the pKa values. We observed strong conformational pH dependence. Our results were comparable of reproducing what was observed experimentally, showing that constant-pH MD simulations are a reliable tools to study pH induced conformational changes.

4.5 *In Silico* stable mutants prediction of chitinase II

The protein primary structure is a linear polypeptide chain comprising of a series of amino acids. It transforms from the unfolded conformation to a unique folded active conformation after synthesis on ribosome. The folding pathways defined from a primary to tertiary structure has been a challenging task, despite the simple concept of protein folding (Fitzkee et al., 2005, Kang and Kini, 2009). However, very little is known about the forces that differentiate unfolded and folded forms. The protein's thermodynamic stability is described by the free-energy difference between the unfolded and the folded state (Pace et al., 1996) as represented by the expression below:

$$\Delta\Delta G=\Delta G_{unfold} - \Delta G_{fold}$$

The difference in energy between the unfolded state and folded state is due to differences in the intramolecular interactions, entropic factors and interactions with the surrounding medium. The protein is more stable (lower ΔG) in its folded state than the unfolded state, which influences its activity and regulation. Stability changes are often encountered for mutated proteins. The knowledge of the mechanisms underlying the mutations affecting the stability is a important area of research. Accurate prediction of changes in protein stability that arise upon mutagenesis is essential for the understanding of structure function relationship and designing of new protein. The function of protein depends on its conformation (Hsu et al., 2008, Mohamed et al., 2009, Muller et al., 1996) therefore; conformational rigidity and flexibility must be balanced. The over packing, reduction in hydrophobic area, backbone strain, and loss of electrostatic interactions caused due to single mutations lead to changes in protein stability. The free energy difference (ΔG) between the folded and unfolded states of proteins is affected by changes in atom-atom interactions. Changes in the interaction among residues within a protein and its surroundings

affect the entropy of the system, thus affecting flexibility/rigidity of the structure (Yue et al., 2005). Proteins are stabilized by covalent disulphide bonds and the non-covalent hydrophobic, electrostatic, van der Waals, and hydrogen bonds. The polarity/charge of functional residues in the active site is located in hydrophobic clefts (Dessailly et al., 2007) and the residues in the catalytic site often have unfavorable backbone angles. For these reasons, the substitution at active site positions may significantly increase the stability, but with the loss of activity. The overall structure, stability of a protein, and the regions that undergo conformational rearrangements are defined by the intra-molecular interactions. Several computational techniques have been developed to predict the free energy differences of unfolding ($\Delta\Delta G$) between a wild-type protein and its mutant based on energy function and machine-learning approaches. Computational and statistical methods may be implemented to predict effects caused by mutations and its biological mechanisms (Thusberg and Vihinen, 2009). The stability of enzymes and proteins, particularly with regards to thermal stability remains a critical issue in industrial biotechnology. Thermostable enzymes are often preferred over their mesophilic counterparts since most industrial processes are often carried at high process temperatures. Some of these advantages includes, reduced risk of mesophiles contamination, improved solubility and bioavailability (Becker et al., 1997), higher reaction rates due to a decreased viscosity (Krahe et al., 1996), and increased diffusion coefficient of substrates and higher process yields due to improved solubility of substrates. The changes in protein stability due to mutations can be predicted computationally by stability predictors. We performed a systematic analysis of various stability predictors using bioinformatics tools. These predictors included Sorting Intolerant from Tolerant (SIFT) (Sim et al., 2012), I-Mutant2.0 (Capriotti et al., 2005), FoldX (Schymkowitz et

al., 2005), SDM server (Worth et al., 2011), MuStab (Teng et al., 2010) and Molecular Dynamics (MD) simulations (Lindahl, 2008).

4.5.1 Materials and Methods

4.5.1.1 Tolerant/Intolerant mutations

The Sorting Intolerant from Tolerant (SIFT) algorithm predicts the effect of the potential impact of amino acid substitutions and coding variants on protein function. It employs sequence homology to compute the adverse effect on protein function upon amino acid substitution. The basic assumption is that the regions that are evolutionarily conserved tends to be less tolerant of mutations, and hence amino acid substitutions/insertions/deletions in these regions are more likely to influence the function. SIFT is based on the degree of conservation of amino acid residues in the sequences obtained from closely related sequences derived through PSI-BLAST. The workflow of SIFT begins with a protein query that is searched against a protein database to get homologous protein sequences. Sequences with suitable diversity are selected and aligned at a specific position to look at the amino acid composition and compute the scores. SIFT scores the mutation at a specific position based on a position-specific scoring matrix (PSSM). It also takes into consideration the physiochemical properties of each amino acid thereby predicting if the amino acid change is tolerant or intolerant, i.e whether the change severely affects the protein function. The output from SIFT includes a score from 0-1, where <=0.05 is the threshold for tolerance. The amino acid sequence of chitinase was submitted to SIFT server and the output obtained was carefully examined. The positions that showed intolerance were grouped on the basis of the magnitude of scoring function.

4.5.1.2 Reliability index/Stability prediction

Several bioinformatics software tools were used to predict protein stability like sequence based I-Mutant 2.0, MuStab, structure based SDM server and FoldX. I-Mutant2.0 is a tool based on support vector machine (SVM) for the automatic prediction of protein stability changes due to single point mutations under different conditions. I-Mutant2.0 correctly predicts the extent to which the stability is affected by a mutation in a new protein with respect to the parent type by utilizing protein sequence or structure. It reaches an 80% accuracy and 0.71 Pearson correlation value, when correlated with experimental ΔΔGs starting from the protein structure. ΔΔG values for stabilizing mutation is greater than zero and for destabilizing mutation is less than zero. Predicting Mutant Protein Stability Change or MuStab is another SVM based web server that takes into consideration various biochemical, structural and biological features. MuStab predicts the effect of the amino acid substitution on the protein stability and function. It utilizes amino acid composition, conformational parameters, polarity and average buried area present in folded state. It has 84.59 % accuracy, 70.29 % sensitivity and 90.98 % specificity. Site Directed Mutator (SDM) based on a statistical potential energy function, predicts the effect of SNPs on the proteins stability. The selected mutants generated from I-Mutant2.0 were further analyzed and compared with different servers such as MuStab and SDM server at different temperature and pH. FoldX, a protein design algorithm uses an empirical force field that computes the energetic effect of point mutations as well as the interaction energy of protein utilizing a full atomic description of the protein structure. FoldX mutates protein using a probability based rotamer library, while search alternative conformations of the side chain present in surrounding. It is used for several reasons, including the effect of point mutations, protein design based on an improved stability, modify affinity/specificity and homology modeling.

4.5.1.3 Flexibility in the protein and its mutants

Protein structures are not rigid macromolecules. The flexibility in the mutated regions were predicted by Flexibility and Local Structure prediction (PredyFlexy) server (de Brevern et al., 2012). The analysis was done on the basis of local structure flexibility features in proteins by relying on the B-factors from X-ray experiments and backbone fluctuations in solution observed in molecular dynamics simulations.

4.5.1.4 Molecular dynamics simulations

The MD simulations were performed at molecular mechanics level by GROMACS 4.6.5 packages (Van Der Spoel et al., 2005) and all atom functions by OPLS (optimized potential for liquid simulation). MD simulation methods were performed on chitinase II mutants (S45E, S45I, S50E, S50I, S50V, Q135C, Q135K, S136I, S136L and S136V) in order to investigate its stability profile. The chitinase II mutants were soaked in a cubic box of water molecules with a dimension of 10 Å. The spc216 template was used to solvate the protein and the charges were neutralized by the addition of Na^+ and Cl^- ions. The system was then minimized using 1500 steps of steepest descent and the temperature was subsequently raised from 0 to 300 K during the equilibration period (100 ps). The equilibration was achieved in two phases NVT ensemble and NPT ensemble. After the equilibration phase, the particle-mesh Ewald method (Horng et al., 2005) was applied and the production phase consisting of 2000 ps was performed at 300 K.

4.5.2 Results and Discussion

4.5.2.1 Sorting intolerant from tolerant

The amino acid residues in the active site possess maximum extent of intolerance and hence, mutation at these positions would lead to a decrease in the stability, loss of activity and denaturation of proteins. The active site residues Tyr36, Glu176, Ala206, Gln208, Gln232, Tyr234 have the maximum number of intolerance 19 while Ala127, Asp174 have 18 and Phe70, Trp307 have 17 (**Table 4.4**).

Table 4.4: The list of all intolerant positions in the amino acid chain that are less likely to mutate. There exist certain positions where mutation would be lethal thus causing a increase/decrease in protein stability.

Number of intolerant mutations	Amino acid positions
19	1M, 36Y, 37W, 92C, 109D, 101I, 113C, 114Q, 117G, 123S, 125G, 126G, 131E, 133G, 134F, 140A, 144A, 148W, 151F, 152G, 153P, 161R, 162P, 163F, 166A, 169D, 170G, 171F, 172D, 175F, 176E, 181N, 189L, 190R, 206A, 207P, 208Q, 209C, 212P, 213D, 220L, 227D, 232Q, 233F, 234Y, 235N, 236N, 238C, 239G, 248N, 252W, 256A, 260S, 262N, 270G, 278A, 281G, 300S, 301F, 302G, 303G, 305M, 307W, 308D, 315N, 343W.
18	32N, 54L, 58C, 69A, 85L, 103C, 106V, 118K, 119T, 120I, 121L, 122L, 127A, 128T, 168V, 173L, 174D, 185F, 186A, 192L, 201Y, 204T, 211F, 228A, 231V, 243Y, 247F, 255W, 265V, 266K, 267V, 272P, 282Y, 298F, 310S, 311Q, 318F, 326L.
17	23A, 25L, 33V, 38G, 68M, 70F, 74I, 84N, 95F, 97G, 98T, 100L, 104P, 124I, 129Y, 198S, 199K, 202F, 203L, 205A, 224V, 229I, 249F, 253D, 273A, 274N, 177A, 288L, 304V, 322I, 331S, 333V.
16	8V, 29S, 35V, 65T, 76G, 79G, 81P, 86A, 105Q, 132G, 136S, 143G, 147V, 179V, 219M, 222G, 240V, 269V, 271V, 279G, 292I, 294H, 317G, 319L.
15	5K, 7V, 21V, 26D, 34V, 71M, 78G, 93G, 108E, 177A, 182M, 195S, 196D, 216N, 226F, 230W, 257Q, 280S, 283L, 306M, 314A.
14	11L, 16T, 18L, 158D, 165D, 167V, 218E, 237Y, 242S, 259T, 268L, 309A, 325I, 332R, 342F.
13	15L, 31S, 57Y, 66L, 83I, 89G, 142A, 150T, 156N, 164G, 258N, 312A.
12	64D, 82E, 184P, 241N.
11	39Q, 72T, 187N, 188T, 210P, 313Y.
10	13V, 87N, 91S, 246N, 264N, 285V.
9	28S, 107G, 130T, 155T.

8	40N, 41S, 75N, 276G, 284P, 327G, 341D.
7	149E, 191S, 215A, 286D, 296R, 299P, 320S.
6	20S, 47G, 50S, 51Q, 90D, 99N, 244P, 324S.
5	94T, 96D, 115S, 139A, 183V, 200Q, 250N, 295A, 321G, 323K.
4	27L, 48G, 52Q, 59E, 62N, 102D, 111K, 112K, 135Q, 137A, 138E, 145R, 157G, 159A, 178T, 194D, 217K, 221D, 245D, 261K, 289A, 290P, 293E, 316D, 328S, 329V, 335R.
3	3S, 46G, 67V, 254D, 263K.
2	6S, 19P, 44A, 60D, 63I, 80V, 146M, 223A, 251T, 287Q, 291V, 297T.
1	17A, 53P, 61P, 73R, 77A, 116L, 141E, 154V, 180S, 197S, 275T.
0	2P, 4F, 9S, 10S, 12P, 14I, 22Q, 24G, 30T, 42A, 43A, 45S, 49P, 55A, 56T, 88I, 101K, 160L, 193M, 214A, 225S, 330I, 334K, 336M, 337F, 338F, 339R, 340R.

4.5.2.2 Prediction and selection of stable mutants

The stability score (0-10) and reliability index (RI) were generated using I-Mutant 2.0 for each amino acid substitution at a pH of 7 and a temperature of 298.15 K. The positions with highest reliability index were classified in the ascending order of magnitude (Appendix VII). The mutations with high RI scores ranging from 6 to 9 were selected for further evaluation. The lists of generated mutants with increased stability are expressed in terms of ΔΔG (**Table 4.5**). The maximum number of substitutions were possible for V, L, E, I, M, P and Y whereas least substitutions were observed for S, H, A, Q and substitution for G was unlikely (**Figure 4.26**)

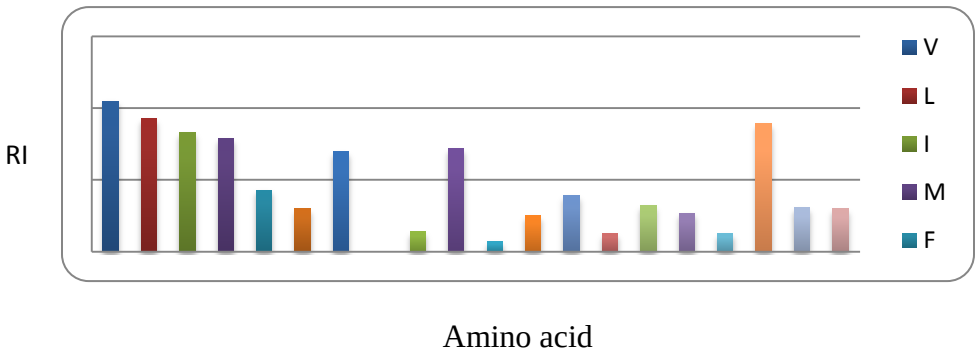


Figure 4.26: Frequency of amino acids substitutions observed in chitinase II on the basis of RI.

The maximum RI value was observed at the position 136 such that Serine can be replaced by Valine, Leucine, Isoleucine in order to gain maximum stability. Combinations of these mutations can be employed experimentally to achieve maximum stability of chitinase II protein. S136M, S136F, S136Y, S136K, S136E, S136N, S45I, S50I and Q135L (Appendix VII) and are the best predicted highly stable mutants. The predicted ΔΔG values of mutants at 298.15 K and pH 7 are for Q135L, S136L, S136I, S45I, S136V, S45E, S50I, S50V, Q135C, S50E and Q135K are 2.08, 2.5, 1.71, 1.64, 1.63, 1.6, 1.45, 1.43, 1.35, 1.07 and 0.55 kcal/mol respectively.

Table 4.5: The classification of selected highly stable mutants (RI score > 6, (+) indicates the positions that are more prone to mutation).

Reliability Index	Amino Acid positions	V	L	I	M	F	W	Y	G	A	P	S	T	C	H	R	K	Q	E	N	D
9	136S	+	+	+																	
8	45S			+																	
	50S			+																	
	135Q		+																		
	136S				+	+		+									+		+	+	
7	45S																		+		
	50S	+																	+		
	108E				+																
	112K	+									+										
	118K	+			+						+										
	123S																		+		
	136S						+				+			+		+		+			
	138E		+																		
	141E		+																		
	180S	+																			
	195S																		+		
	197S			+															+		
	198S			+															+		
	199K	+									+										
	260S					+													+		
	261K	+			+	+															
	263K	+				+					+										
	264N	+																			
	266K	+																			
	274N	+																			

	280S	+																			
	334K	+																			
6	39Q		+																		
	45S	+			+		+	+													
	46G									+											
	50S							+			+						+				
	111K										+										
	112K				+			+													
	123S										+										+
	125G																		+		
	126G																		+		
	131E		+																		
	133G		+																		
	135Q													+			+				
	136S												+		+						+
	143G		+																		
	180S																		+		
	191S																		+		
	195S			+							+										
	197S	+									+										
	198S	+									+										
	200Q			+															+		
	213D																		+		
	217K	+																			
	218E		+																		
	257Q		+																		
	258N		+		+								+								
	260S															+					
	261K							+			+										
	263K							+													
	264N				+	+							+								
	266T					+					+										
	280S			+															+		
	286D																		+		
	310S		+	+	+											+					
	334										+										

These mutated positions were further analyzed by MuStab (pH=7,T= 298.15 K & 310.15 K). According to MuStab, Q135K, Q135L, Q135C, S45E, S50E, S50I, S50V have the good confidence score of 36.79 %, 30.89 %, and 30.89 %, 28.04 %, 25.18 %, 25 % and 22.86 % respectively. According to combine results of I-Mutant 2.0, (pH=7,T=298.15 K & 310.15 K) and MuStab (pH=7, T=298.15 K & 310.15 K), Q135L, S45E, S50I, S50V, Q135C, S50E and Q135K are the stable mutants. Further WHAT IF server suggests stabilizing proline mutations at 35, 53,

66, 121, 136, 207, 216, 229, 239, 262, 306 and 321 amino acid residue position. Swiss-PdbViewer (DeepView) was used to create *in-silico* mutations on the three dimensional structure of chitinase II and difference between wild and mutant residue was visualized **(Table 4.6)**. According to combine results of I-Mutant 2.0, (pH=7, T= 298.15 K & 310.15 K) and MuStab (pH=7, T= 298.15 K & 310.15 K), Q135L, S45E, S50I, S50V, Q135C, S50E and Q135K are the stable mutants.

Table 4.6: Analysis of selected mutants using I-Mutant 2.0, (pH=7, T= 298.15 K & 310.15 K), MuStab (pH=7, T= 298.15 K & 310.15 K) and SDM server.

SEQUENCE BASED STABILITY ANALYSIS									
Amino Acid Substitutions	**I-Mutant 2.0, pH=7,T= 298.15 K**		**I-Mutant 2.0, pH=7,T= 310.15 K**		**MuStab pH=7,T= 298.15 K prediction confidence**		**MuStab pH=7,T= 310.15 K**		**Structure based stability analysis (SDM server)**
	Stability	**ΔΔG* (Kcal/ mol)**	**Stability**	**ΔΔG* (Kcal/ mol)**	**Stability**	**Confid -ence score**	**Stability**	**Confid -ence score**	
Q135L	INCREASED	2.08	INCREASED	2.17	INCREASED	30.36%	INCREASED	30.89%	-
S136L	INCREASED	2.5	INCREASED	2.60	-	-	-	-	Stabilizing (0.70)
S136I	INCREASED	1.71	INCREASED	1.83	-	-	-	-	Stabilizing (0.96)
S45I	INCREASED	1.64	INCREASED	1.77	-	-	-	-	highly stabilizing (2.72)
S136V	INCREASED	1.63	INCREASED	1.72	-	-	-	-	-
S45E	INCREASED	1.6	INCREASED	1.65	INCREASED	27.5%	INCREASED	28.04%	Stabilizing (1.35)
S50I	INCREASED	1.45	INCREASED	1.58	INCREASED	24.11%	INCREASED	25%	-

S50V	INCREASED	1.43	INCREASED	1.53	-	-	INCREASED	22.86%	-
Q135C	INCREASED	1.35	INCREASED	2.17	INCREASED	30.36%	INCREASED	30.89%	highly stabilizing (2.51)
S50E	INCREASED	1.07	INCREASED	1.10	INCREASED	24.46%	INCREASED	25.18%	-
Q135K	INCREASED	0.55	INCREASED	0.57	INCREASED	35.18%	INCREASED	36.79%	-

*ΔΔG = ΔG(NewProtein)-ΔG(WildType) in Kcal/mol, ΔΔG<0: Decrease Stability, ΔΔG>0: Increase Stability

4.5.2.3 Flexibility in the protein and its mutants

At position 45 of chitinase II, the RMS fluctuations were found to decrease for S45E and S45I with predicted flexibility of 0.6 and 0.4, respectively as compared to the normal value of 0.9. This decrease in RMS fluctuation values indicates the stable nature of S45E and S45I. In the case of mutants S50E, S50I and S50V the RMS fluctuations remained comparable with no change in predicted flexibility. Similarly, at positions 135 and 136, the mutants showed decrease in RMSF values. Therefore, these mutations at the respective positions may enhance the stability and flexibility of this enzyme (Appendix VIII)**.** Moreover, the introduction of Glu and Ile at position 45 of chitinase II, decreases the polar contacts between neighboring side chains, whereas at position 50 the interactions increased with the addition of Ile. At position 135, there were no changes in the polar interactions upon mutation as no polar side chains were present, while S136I showed an increase in polar contacts (**Figure 4.27**).

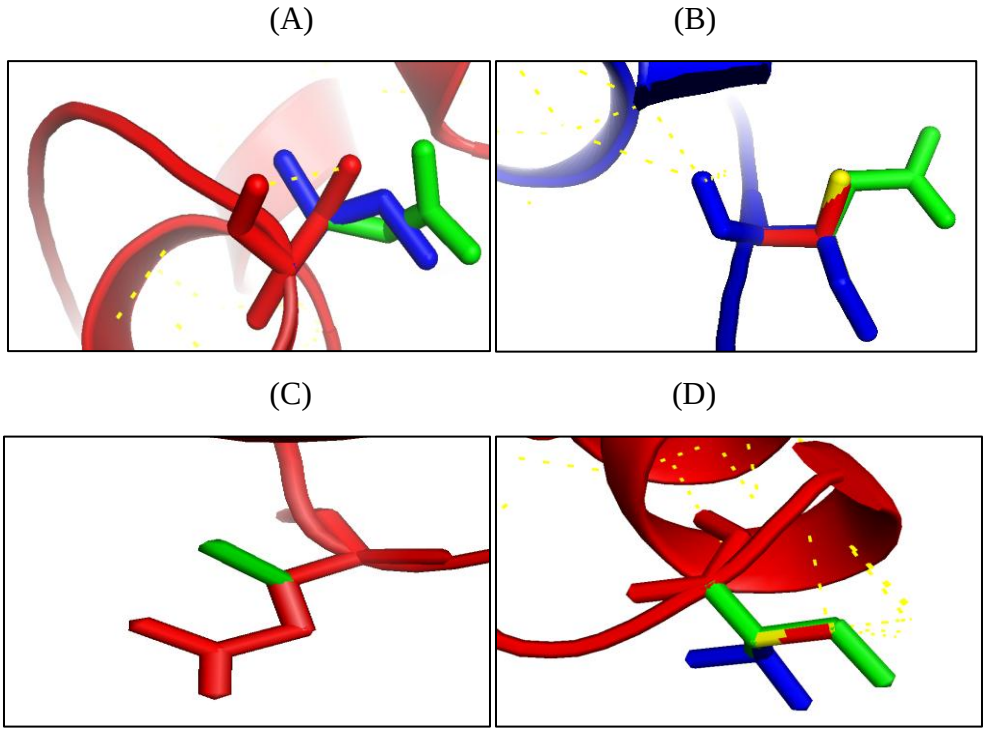


Figure 4.27: Graphical representations of selected stable mutants with changes in their polar contacts (A) S45E and S45I; S(Red), E (green) and I(blue) (B) S50E, S50I and S50V; S(red), E(green), I(blue) and V(yellow) (C) Q135C; Q(red) and C(green) (D) S136I, S136L and S136V; S(red), I(green), L(blue) and V(yellow).

4.5.2.4 Molecular dynamics simulations

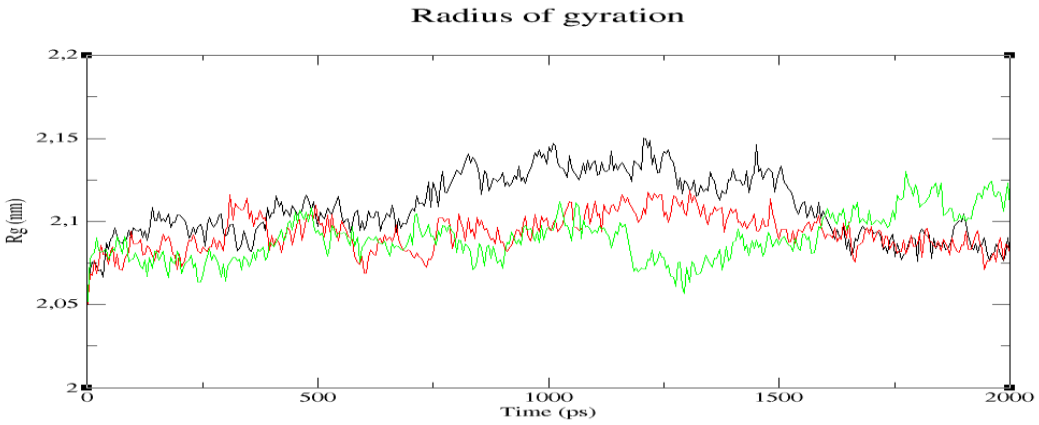
The constant average fluctuation of temperature around 300 K suggested stable and accurate nature of the MD simulations performed. The average total energy during the simulations suggested that the designed mutants have less energy than the wild type chitinase (**Table 4.7**).

Table 4.7: Relative Energy (kJ/mol) of chitinase and its mutants during the time scale of 2000 ps simulations.

S. No.	Protein type	Average Total Energy (kJ/mol)
1	Wild	-1458740
2	S45E	-1451940
3	S45I	-1450760
4	S50E	-1451970
5	S50I	-1451050
6	S50V	-1451030
7	Q135C	-1451310
8	Q135K	-1450590
9	S136I	-1450620
10	S136L	-1450770
11	S136V	-1458510

The Radius of gyration (Rg) analysis provided insight into the compactness of the protein during the simulations (Lobanov et al., 2008). The Rg values suggested that all the predicted stable mutants of chitinase II remained in its compact (folded) form over the course of simulations as compared with wild type chitinase II. The overall Rg (nm) values was found to be lower thus suggesting to be more stable than wild type chitinase II. The Rg values for the mutants were found to be lower than wild type chitinase II and lies below 2.1 nm except initial fluctuations for Q135C. The Rg plot suggested that the mutants S45I, S50I, Q135K and S136L represent a more stable isoform of protein (**Figure 4.28**).

(A)



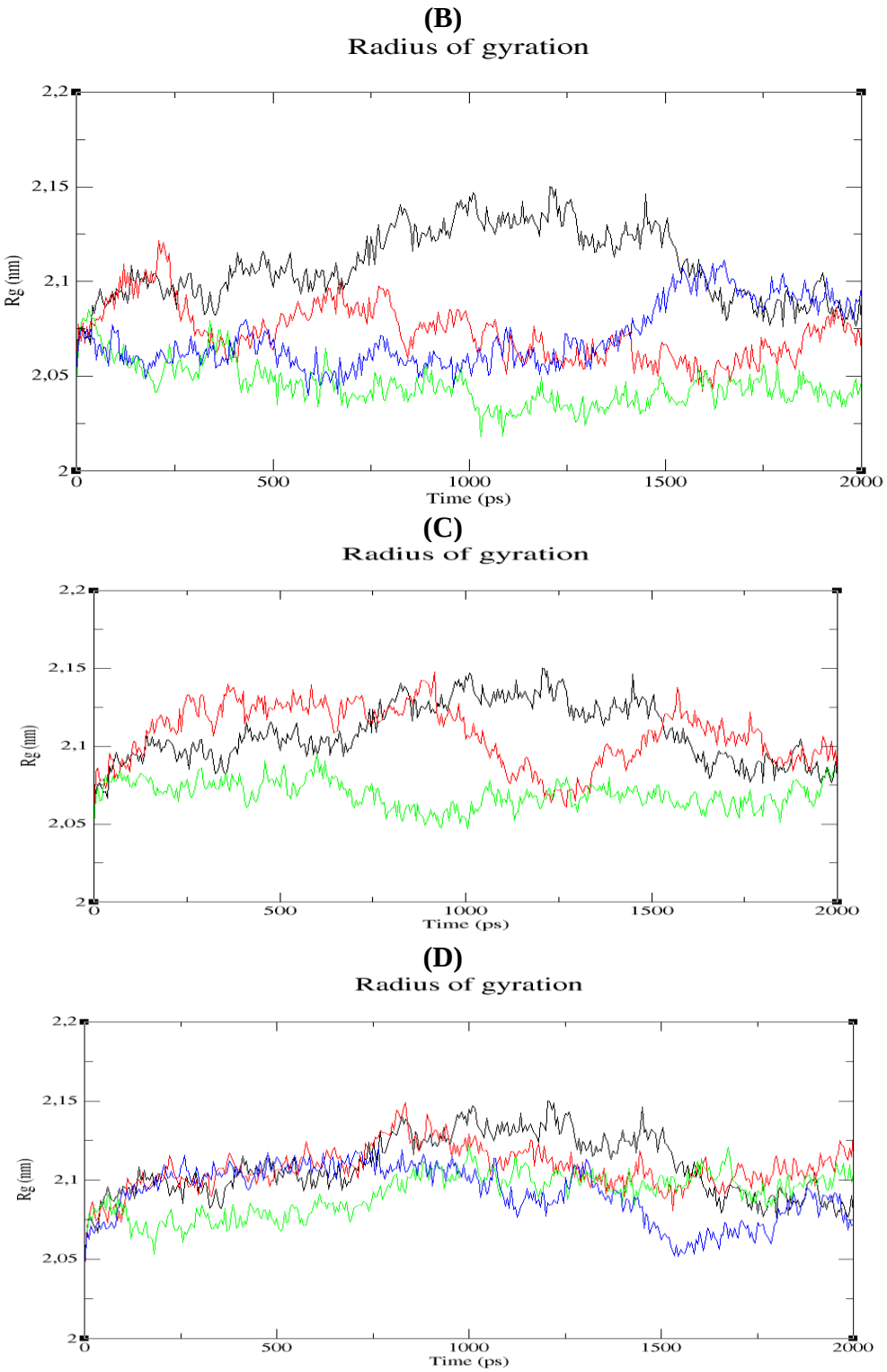
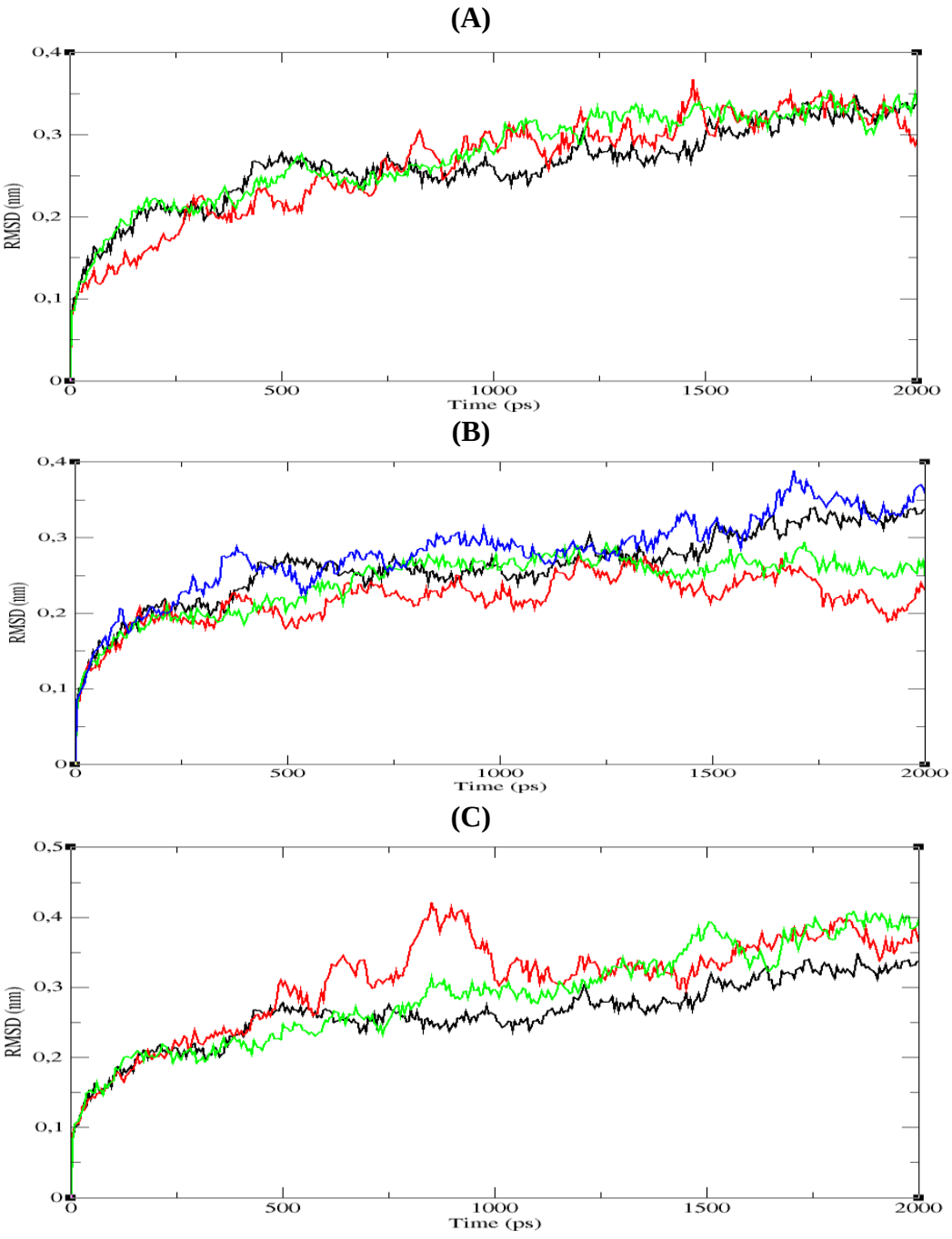


Figure 4.28: The radius of gyration values as a function of time for wild chitinase II (black), (A) S45E (red) and S45I (green) (B) S50E (red), S50I (green) and S50V (blue) (C) Q135C (red) and Q135K (green) (D) S136I (red), S136L (green) and S136V (blue). Values were calculated with the use of Cα atoms.

The MD simulation of chitinase II mutants displayed an acceptable stability profile. The RMSD values for all the proteins were found to be around 0.2 nm - 0.4 nm. All the selected stable mutants show almost constant and comparable values of RMSD deviations while Q135C at 500-1000 ps and S136I at 15-2000 ps shows a little higher fluctuations (**Figure 4.29**).



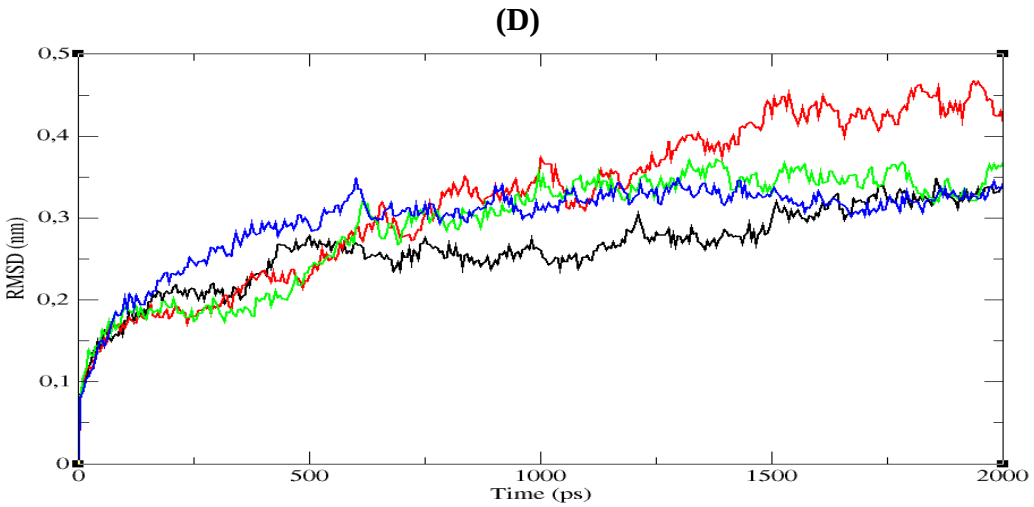
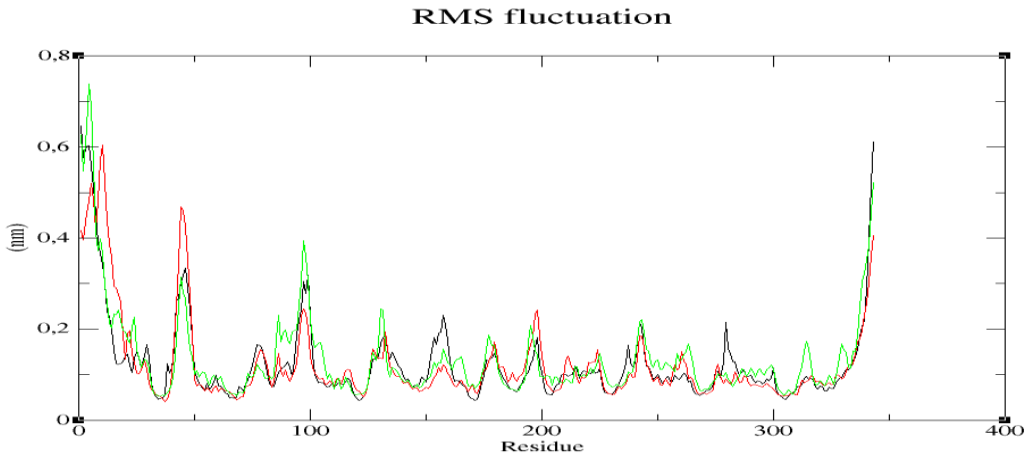


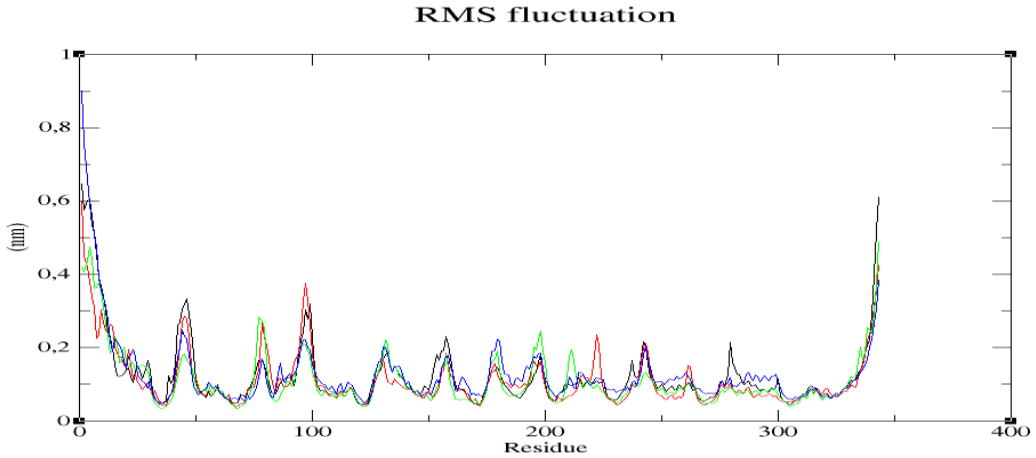
Figure 4.29: RMS deviation values as a function of time for wild chitinase II (black), (A) S45E (red) and S45I (green) (B) S50E (red), S50I (green) and S50V (blue) (C) Q135C (red) and Q135K (green) (D) S136I (red), S136L (green) and S136V (blue). Values were calculated with the use of Cα atoms.

The root mean square fluctuations (RMSF) with reference to the residues were evaluated during the course of the simulation using the g_rmsf module. The RMSF of the Cα atoms of chitinase II mutants from the initial structure were plotted as a function of residue number. The mutants S50E, S50I, S50V, Q135C and Q135K showed smaller RMS fluctuations at the mutated positions, while S45I, S136I, S136L and S136V showed a comparable RMSF values with wild type chitinase II. S45E showed a little higher fluctuation at the point mutation site. The RMS fluctuations around the mutated positions suggested that the new protein designed have less RMS fluctuations from the initial structure thus suggesting the stable nature of these mutants. This information can be utilized to design new and stable protein molecules (**Figure 4.30**).

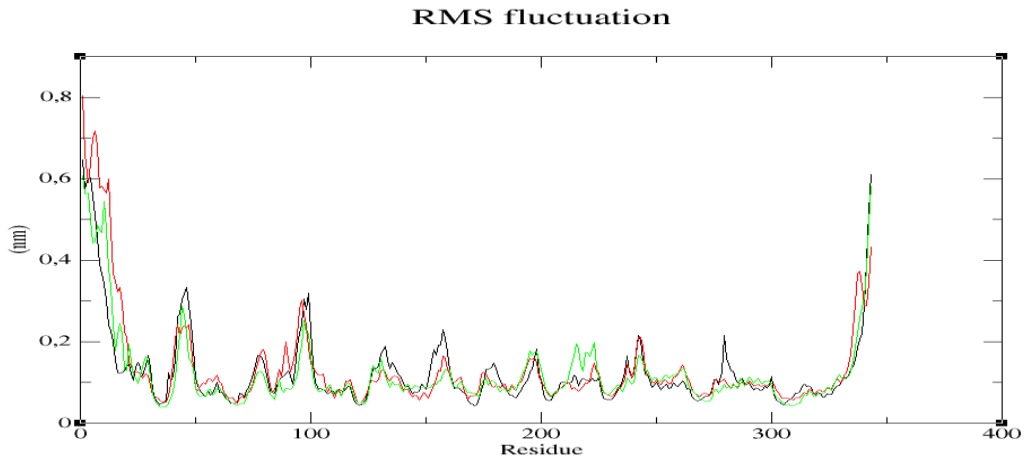
(A)


RMS fluctuation
0,8
0,6
0,4
0,2
0
(nm)
0
100
200
300
400
Residue

(B)


RMS fluctuation
1
0,8
0,6
0,4
0,2
0
(nm)
0
100
200
300
400
Residue

(C)


RMS fluctuation
0,8
0,6
0,4
0,2
0
(nm)
0
100
200
300
400
Residue

(D)

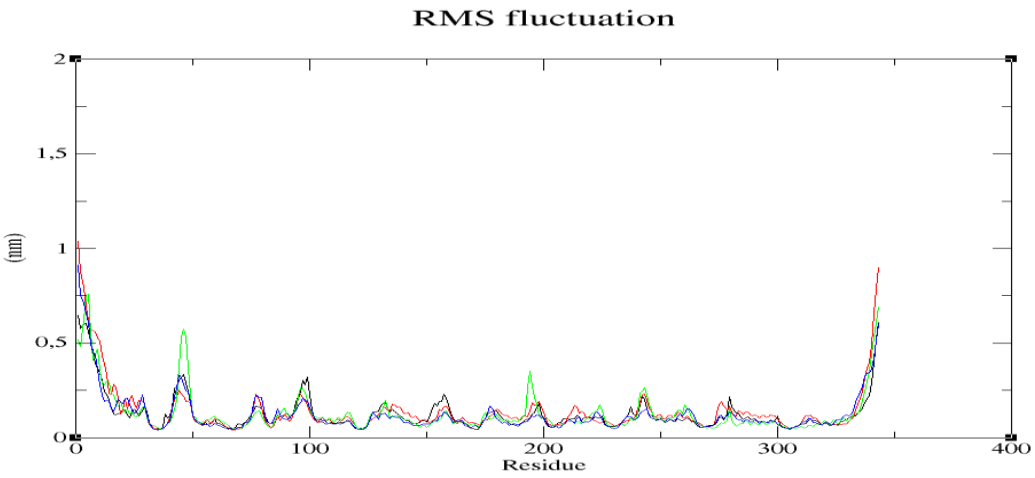


Figure 4.30: Average RMS deviation values as a function of amino acid sequence numbers for wild chitinase II (black), (A) S45E (red) and S45I (green) (B) S50E (red), S50I (green) and S50V (blue) (C) Q135C (red) and Q135K (green) (D) S136I (red), S136L (green) and S136V (blue). Values were calculated with the use of Cα atoms.

4.5.3 Conclusions

Proteins are dynamic molecules with marginal stability. The effect of the amino acid mutations at a particular location can be utilized to observe stability changes in the protein. In this study, different stability and flexibility predictions were used to assess the relation of point mutation and protein stability. The generation of mutation induced models by using various *in silico* techniques can be used to predict the stability in the absence of experimental 3D structures. The results showed a coherent pattern in all predictions. Using MD simulations, the common fluctuating residues which were considered as thermally weak-point residues, were investigated. These stable mutants sustain a stable conformation during the course of MD simulation. The results obtained from this study can be further validated to design a better thermostable enzyme.

CHAPTER 5

SUMMARY

The current study mainly focuses on experimental characterization and computational analysis of thermostable chitinase. The experimental methodology deals with the cloning of chitinase gene obtained from *Thermomyces lanuginosus* into *E. coli* and *P. pastoris* expression vectors followed by the expression and purification of chitinase enzymes. *Thermomyces lanuginosus* SSBP was grown at 323.15 K on potato dextrose medium. *Pichia pastoris* GS115 (Invitrogen) was maintained on YPD agar supplemented with 100 µg/ml zeocin. Total RNA was isolated from *T. lanuginosus* SSBP and the first-strand cDNA was synthesized. Primer were designed using Oligo version 6.0 software, based on chitinase gene identified in the currently available annotated genome sequence data of *T. lanuginosus* SSBP with EcoRI, XbaI, NheI and XhoI restriction sites. Polymerase chain reaction (PCR) was carried out with initial denaturation at 371.15 K for 30 s, cyclic denaturation at 371.15 K for 10s, primer annealing at 328.15 K for 20 s, extension at 345.15 K for 30 s and the final extension at 343.15 K for 7 min. The PCR product was confirmed using 1 % agarose gel electrophoresis. Size of the PCR product was confirmed by running a DNA ladder alongside the samples. Gels were then subjected to staining in ethidium bromide (0.05 mg/ml) for 20 min and destained in distilled water for a further 5 - 10 min. Stained gels were viewed on a UV transilluminator. The size of the product was 1032 bp. The desired bands were excised from the gel and purified using the DNA Clean and Concentrator Kit (Zymo Research). Restriction endonucleases were used for specific digestion of both the vector and the insert DNA. For ligation, a molar vector: insert ratio of 1:3 was used to guarantee high ligation efficiency. The ligation reaction containing the vector, insert, ligation buffer (has ATP) and T4 DNA ligase was incubated overnight at 4 ^{o}C. The ligation mixtures were then further transformed

into competent host cells. DNA concentration was determined using the NanoDrop Spectrophotometer ND-1000 (Thermo Scientific). The competent BL21 and electro-competent *Pichia pastoris* cells were prepared. The ligated DNA was transformed into competent cells. The transformation mixtures were plated. Colonies that grew were tested by colony PCR to determine the presence of *chitinase* gene. DNA sequencing of a positive clone was carried out. A positive clone was cultured and processed for protein expression. Glycol chitin was prepared and the enzyme activity was confirmed by plate assay. The gene was successfully cloned into the *E. coli* BL21 expression vector, pET21c and the *P. pastoris* GS115 expression vector, pBGP1. The size of chitinase II protein was 36.6 KDa. The 3D structure of chitinase obtained from *T. lanuginosus* has not been characterized thus far. In order to study the detail structure and dynamics of the enzymes, a 3D model was obtained with a higher precision using homology modelling and *ab initio* approaches.

The computational approaches involves 3D structure prediction of chitinases using the MODELLER (version - 9.14) followed by the active site analysis using sequence-based and structure-based approaches. Here, twenty models were generated and evaluated on the basis of RMSD, TM-score, DOPE profile and Molecular Probability Density Function (Mol PDF). The identification of the catalytic residues is key in the understanding of the functions of enzymes. For this purpose, a variety of computational methods have been implemented for this task. The best existing method exploits information from the sequences and the 3D structures of similar proteins. The active site pockets were predicted by metaPocket 2.0, COFACTOR and COACH softwares. Docking studies were performed with allosamizoline, one of its inhibitors to confirm the binding sites. Structural analysis of the predicted 3D structures of chitinase II revealed that it contains Tyr36, Phe70, Asp174, Glu176, Gln208, Gln232, Tyr234 and Trp307 in the active site

pocket. The Glu176 was observed to be essential for the activity of chitinase II. MD simulations provided valuable insights into the structure and stability of the enzymes. 10 ns MD simulations of chitinase II and chitinase I were performed at different temperatures ranging from 300 K-375 K to check the thermostable nature of this enzyme and by analyzing various parameters. The MD simulations of chitinases displayed an acceptable stability profile at a temperature of 300 K, 325 K and 350 K respectively, while instability was observed at 375 K, in agreement with the experimental data. The average RMSD values of the chitinases structures were found to be lower at 350 K, as compared to 325 K and 300 K. The RMSF of the Cα value of the chitinases structures were comparable at all temperatures except 375 K. However, the loop forming residues ranging from 40-50 showed higher fluctuations at 325 K. Similarly, the regions spanning residues from 80 to 100 showed higher values of RMSF at 325 K and 350 K, while the coiled regions of 240–250 residues displayed an elevated value at 350 K. However, the enhanced localized flexibility in the loop regions suggested a greater stability in the case of chitinase II enzymes at higher temperatures. The R_g values explain that the protein remained stable in its folded form over the MD trajectory with little fluctuations observed at 300 K, 325 K and 350 K. The secondary structures during the MD simulations clearly show the global conformation of chitinases conserved with no significant changes in its secondary structure elements, even at the higher temperatures. We also performed a constant pH molecular dynamics simulation at a pH range 2-6. The purpose of this study was to establish the optimum activity and stability profiles of chitinase I and chitinase II respectively. The Molecular Dynamics simulations were carried out at fixed protonation states in an explicit water environment to evaluate the effect of the physiological pH on chitinase I and II enzymes. The results suggest a strong conformational pH dependence of chitinases. These enzymes retained their characteristic TIM Barrel fold at the

respective protonated conditions, thus validating the experimental outcomes. The mechanisms of effects of mutation are useful to study the nature of this enzyme and to generate the stable mutants. Various stability and flexibility predictors were used to assess the relation of point mutation and enzyme stability. According to our results, Q135L, S45E, S50I, S50V, Q135C, S50E and Q135K are the stable mutants. All possible stable protein mutants can be validated experimentally for future study.

CHAPTER 6

CONCLUSIONS

This work involved experimental studies supported with computational methods to better understand the conformational profile of chitinase. The experimental techniques employed in this study included cloning, expression and enzyme activity assay revealed several novel attributes of the thermostable chitinase obtained from *T. lanuginosus*. Additionally, computational methods were also employed to complement the experimental work aimed at analysing the thermostable nature of this enzyme. The sequence based function prediction, structure analysis and active site predictions revealed that it has a domain similar to Glycoside hydrolase family 18, which hydrolyzes the glycosidic bond between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety. The predicted functions were further confirmed by various structure based prediction methods. The 3D models were predicted with higher precisions using various protein structure prediction methods. These models showed that the characteristic chitinase $(\beta/\alpha)_8$ TIM barrel is in their structural topology with Tyr36, Phe70, Asp174, Glu176, Gln208, Gln232, Tyr234 and Trp307 as the active site residues. However, the presence of the exposed surface of Glu176 was essential for the activity and is anticipated to have a profound effect on the substrate binding. These active sites were confirmed by sequence and structure based analysis. Docking studies with one of the inhibitor allosamidin showed various patterns of binding with their active site residues.

Moreover, molecular dynamics simulations provided valuable insights into the 3D structure and stability of chitinases at different temperatures ranging from 300K-375K. The RMSD values obtained at all temperatures clearly revealed that the stability of this enzyme was maintained

until 350 K, in accordance with experimentally observed results. On the other hand, an increase in the RMSD value at 375 K, revealed drastic conformational changes in the secondary structure that indicate the enzyme is in the process of denaturation.

The structural effects arising from pH induced effects on chitinase I and chitinase II enzymes have been performed by using molecular dynamics simulation method. This is considered to be a major factor in solvent medium in order for the enzymes to keep its proper function. We presented a titration curve and pK calculations of the titrable sites at a particular pH. Our results suggest that the pH calculations supported the experimental data, that chitinase I and chitinase II are stable at lower pH. The results from simulations focusing on how the conformational change is linked to the pKa values. The constant pH MD simulations of both the enzymes showed a strong conformational pH dependence and the chitinases remain stable and in closed conformation at lower pH.

Furthermore, the effects of the amino acid mutations at particular locations were utilized to observe the stability changes in the protein, due to their dynamic and marginal stabilities. For this purpose, the mechanisms of effects of mutation were useful to study the nature of the proteins and to generate the stable mutants. Accordingly in this study, different stability and flexibility predictions were employed to assess the relation of the point mutations and protein stabilities. The generation of mutation-induced models using various *in silico* techniques were used to predict the stability of chitinase in the absence of experimental data. These results showed a coherent pattern in all predictions. Enhancing intramolecular interactions and lowering the overall RMSD of highly fluctuating residues can improve enzyme rigidity and thermostability. All possible stable protein mutants can be validated experimentally for future

studies. These findings may pave the way to engineer a new and better thermostable protein for future purposes.

REFERENCES

FENNELL, C. J. & GEZELTER, J. D. 2006. Is the Ewald summation still necessary? Pairwise alternatives to the accepted standard for long-range electrostatics. *J Chem Phys,* 124**,** 234104.

FUZO, C. A. & DEGREVE, L. 2012. Effect of the thermostat in the molecular dynamics simulation on the folding of the model protein chignolin. *J Mol Model,* 18**,** 2785-94.

HOLDEN, Z. C., RICHARD, R. M. & HERBERT, J. M. 2013. Periodic boundary conditions for QM/MM calculations: Ewald summation for extended Gaussian basis sets. *J Chem Phys,* 139**,** 244108.

HUMPHREY, W., DALKE, A. & SCHULTEN, K. 1996. VMD: visual molecular dynamics. *J Mol Graph,* 14**,** 33-8, 27-8.

JORGENSEN, W. L. & TIRADO-RIVES, J. 2005. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc Natl Acad Sci U S A,* 102**,** 6665-70.

KARPLUS, M. & MCCAMMON, J. A. 2002. Molecular dynamics simulations of biomolecules. *Nat Struct Biol,* 9**,** 646-52.

KINI, R. M. & EVANS, H. J. 1991. Molecular modeling of proteins: a strategy for energy minimization by molecular mechanics in the AMBER force field. *J Biomol Struct Dyn,* 9**,** 475-88.

LINDAHL, E. R. 2008. Molecular dynamics simulations. *Methods Mol Biol,* 443**,** 3-23.

NGUYEN, T. T., VIET, M. H. & LI, M. S. 2014. Effects of water models on binding affinity: evidence from all-atom simulation of binding of tamiflu to A/H5N1 neuraminidase. *ScientificWorldJournal,* 2014**,** 536084.

NORBERTO DE SOUZA, O. & ORNSTEIN, R. L. 1999. Molecular dynamics simulations of a protein-protein dimer: particle-mesh Ewald electrostatic model yields far superior results to standard cutoff model. *J Biomol Struct Dyn,* 16**,** 1205-18.

References

VAN DER SPOEL, D., LINDAHL, E., HESS, B., GROENHOF, G., MARK, A. E. & BERENDSEN, H. J. 2005. GROMACS: fast, flexible, and free. *J Comput Chem,* 26**,** 1701-18.

KITCHEN, D. B., DECORNEZ, H., FURR, J. R. & BAJORATH, J. 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov,* 3**,** 935-49.

POWERS, R., COPELAND, J. C., GERMER, K., MERCIER, K. A., RAMANATHAN, V. & REVESZ, P. 2006. Comparison of protein active site structures for functional annotation of proteins and drug design. *Proteins,* 65**,** 124-35.

ROY, A., YANG, J. & ZHANG, Y. 2012. COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Res,* 40**,** W471-7.

SULLIVAN, S. M. & HOLYOAK, T. 2008. Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection. *Proc Natl Acad Sci U S A,* 105**,** 13829-34.

SUZUKI, S., NAKANISHI, E., OHIRA, T., KAWACHI, R., NAGASAWA, H. & SAKUDA, S. 2006. Chitinase inhibitor allosamidin is a signal molecule for chitinase production in its producing Streptomyces I. Analysis of the chitinase whose production is promoted by allosamidin and growth accelerating activity of allosamidin. *J Antibiot (Tokyo),* 59**,** 402-9.

WU, G., ROBERTSON, D. H., BROOKS, C. L., 3RD & VIETH, M. 2003. Detailed analysis of grid-based molecular docking: A case study of CDOCKER-A CHARMm-based MD docking algorithm. *J Comput Chem,* 24**,** 1549-62.

YANG, J., ROY, A. & ZHANG, Y. 2013. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics,* 29**,** 2588-95.

ZHANG, Z., LI, Y., LIN, B., SCHROEDER, M. & HUANG, B. 2011. Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. *Bioinformatics,* 27**,** 2083-8.

References

AAM, B. B., HEGGSET, E. B., NORBERG, A. L., SORLIE, M., VARUM, K. M. & EIJSINK, V. G. 2010. Production of chitooligosaccharides and their potential applications in medicine. *Mar Drugs,* 8**,** 1482-517.

ADRIO, J. L. & DEMAIN, A. L. 2010. Recombinant organisms for production of industrial products. *Bioeng Bugs,* 1**,** 116-31.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-10.

ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res,* 25**,** 3389-402.

ANIL, K., SESHAGIRIRAO, K. & PODILE, A. R. 2007. A simple, rapid and yet less expensive method to detect chitinase in agarose plates. *J Biochem Biophys Methods,* 70**,** 683-4.

APOSTOLICO, A. & GIANCARLO, R. 1998. Sequence alignment in molecular biology. *J Comput Biol,* 5**,** 173-96.

ARAKANE, Y., ZHU, Q., MATSUMIYA, M., MUTHUKRISHNAN, S. & KRAMER, K. J. 2003. Properties of catalytic, linker and chitin-binding domains of insect chitinase. *Insect Biochem Mol Biol,* 33**,** 631-48.

ASZODI, A. & TAYLOR, W. R. 1996. Homology modelling by distance geometry. *Fold Des,* 1**,** 325-34.

BAIROCH, A. & APWEILER, R. 1999. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Res,* 27**,** 49-54.

BAXEVANIS, A. D. 1998. Practical aspects of multiple sequence alignment. *Methods Biochem Anal,* 39**,** 172-88.

References

BECKER, P., ABU-REESH, I., MARKOSSIAN, S., ANTRANIKIAN, G. & MARKL, H. 1997. Determination of the kinetic parameters during continuous cultivation of the lipase-producing thermophile Bacillus sp. IHI-91 on olive oil. *Appl Microbiol Biotechnol,* 48**,** 184-90.

BERMAN, H. M., WESTBROOK, J., FENG, Z., GILLILAND, G., BHAT, T. N., WEISSIG, H., SHINDYALOV, I. N. & BOURNE, P. E. 2000. The Protein Data Bank. *Nucleic Acids Res,* 28**,** 235-42.

BHARADWAJ, G. & MAHESHWARI, R. 1999. A comparison of thermal characteristics and kinetic parameters of trehalases from a thermophilic and a mesophilic fungus. *FEMS Microbiol Lett,* 181**,** 187-93.

BIERBAUM, S., NICKEL, R., KOCH, A., LAU, S., DEICHMANN, K. A., WAHN, U., SUPERTI-FURGA, A. & HEINZMANN, A. 2005. Polymorphisms and haplotypes of acid mammalian chitinase are associated with bronchial asthma. *Am J Respir Crit Care Med,* 172**,** 1505-9.

BLACKWELL, J. 1969. Structure of beta-chitin or parallel chain systems of poly-beta-(1-4)-N-acetyl-D-glucosamine. *Biopolymers,* 7**,** 281-98.

BLUNDELL, T. L., SIBANDA, B. L., STERNBERG, M. J. & THORNTON, J. M. 1987. Knowledge-based prediction of protein structures and the design of novel molecules. *Nature,* 326**,** 347-52.

BOLAR, J. P., NORELLI, J. L., WONG, K. W., HAYES, C. K., HARMAN, G. E. & ALDWINCKLE, H. S. 2000. Expression of Endochitinase from Trichoderma harzianum in Transgenic Apple Increases Resistance to Apple Scab and Reduces Vigor. *Phytopathology,* 90**,** 72-7.

BOOT, R. G., BLOMMAART, E. F., SWART, E., GHAUHARALI-VAN DER VLUGT, K., BIJL, N., MOE, C., PLACE, A. & AERTS, J. M. 2001. Identification of a novel acidic mammalian chitinase distinct from chitotriosidase. *J Biol Chem,* 276**,** 6770-8.

BORTONE, K., MONZINGO, A. F., ERNST, S. & ROBERTUS, J. D. 2002. The structure of an allosamidin complex with the Coccidioides immitis chitinase defines a role for a second acid residue in substrate-assisted mechanism. *J Mol Biol,* 320**,** 293-302.

References

BOWIE, J. U., LUTHY, R. & EISENBERG, D. 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science,* 253**,** 164-70.

BURGI, R., KOLLMAN, P. A. & VAN GUNSTEREN, W. F. 2002. Simulating proteins at constant pH: An approach combining molecular dynamics and Monte Carlo simulation. *Proteins,* 47**,** 469-80.

CALONJE, M., NOVAES-LEDIEU, M., BERNARDO, D., AHRAZEM, O. & GARCIA MENDOZA, C. 2000. Chemical components and their locations in the Verticillium fungicola cell wall. *Can J Microbiol,* 46**,** 101-9.

CANO-CANCHOLA, C., SOSA, L., FONZI, W., SYPHERD, P. & RUIZ-HERRERA, J. 1992. Developmental regulation of CUP gene expression through DNA methylation in Mucor spp. *J Bacteriol,* 174**,** 362-6.

CAPRIOTTI, E., FARISELLI, P. & CASADIO, R. 2005. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res,* 33**,** W306-10.

CARLSTROM, D. 1957. The crystal structure of alpha-chitin (poly-N-acetyl-D-glucosamine). *J Biophys Biochem Cytol,* 3**,** 669-83.

CARROLL, W. L. 1993. Introduction to recombinant-DNA technology. *Am J Clin Nutr,* 58**,** 249S-258S.

CHAO, A. C., SHYU, S. S., LIN, Y. C. & MI, F. L. 2004. Enzymatic grafting of carboxyl groups on to chitosan--to confer on chitosan the property of a cationic dye adsorbent. *Bioresour Technol,* 91**,** 157-62.

CHAUDHURI, A. & MAHESHWARI, R. 1996. A Novel Invertase from a Thermophilic FungusThermomyces lanuginosus:Its Requirement of Thiol and Protein for Activation. *Archives of Biochemistry and Biophysics,* 327**,** 98-106.

CHEN, J. K., SHEN, C. R. & LIU, C. L. 2010. N-acetylglucosamine: production and applications. *Mar Drugs,* 8**,** 2493-516.

References

CHOTHIA, C. & LESK, A. M. 1986. The relation between the divergence of sequence and structure in proteins. *EMBO J,* 5**,** 823-6.

CLAESSENS, M., VAN CUTSEM, E., LASTERS, I. & WODAK, S. 1989. Modelling the polypeptide backbone with 'spare parts' from known protein structures. *Protein Eng,* 2**,** 335-45.

COTTRELL, M. T., MOORE, J. A. & KIRCHMAN, D. L. 1999. Chitinases from uncultured marine microorganisms. *Appl Environ Microbiol,* 65**,** 2553-7.

DAHIYA, N., TEWARI, R. & HOONDAL, G. S. 2006. Biotechnological aspects of chitinolytic enzymes: a review. *Appl Microbiol Biotechnol,* 71**,** 773-82.

DE BEER, T. A., BERKA, K., THORNTON, J. M. & LASKOWSKI, R. A. 2014. PDBsum additions. *Nucleic Acids Res,* 42**,** D292-6.

DE BREVERN, A. G., BORNOT, A., CRAVEUR, P., ETCHEBEST, C. & GELLY, J. C. 2012. PredyFlexy: flexibility and local structure prediction from sequence. *Nucleic Acids Res,* 40**,** W317-22.

DESSAILLY, B. H., LENSINK, M. F. & WODAK, S. J. 2007. Relating destabilizing regions to known functional sites in proteins. *BMC Bioinformatics,* 8**,** 141.

DUO-CHUAN, L. 2006. Review of fungal chitinases. *Mycopathologia,* 161**,** 345-60.

FENNELL, C. J. & GEZELTER, J. D. 2006. Is the Ewald summation still necessary? Pairwise alternatives to the accepted standard for long-range electrostatics. *J Chem Phys,* 124**,** 234104.

FINN, R. D., CLEMENTS, J. & EDDY, S. R. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res,* 39**,** W29-37.

FISCHER, D. & EISENBERG, D. 1997. Assigning folds to the proteins encoded by the genome of Mycoplasma genitalium. *Proc Natl Acad Sci U S A,* 94**,** 11929-34.

FISCHER, L., SCHECKERMANN, C. & WAGNER, F. 1995. Purification and characterization of a thermotolerant beta-galactosidase from Thermomyces lanuginosus. *Appl Environ Microbiol,* 61**,** 1497-501.

FISER, A. & SALI, A. 2003. Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol,* 374**,** 461-91.

FITZKEE, N. C., FLEMING, P. J., GONG, H., PANASIK, N., JR., STREET, T. O. & ROSE, G. D. 2005. Are proteins made from a limited parts list? *Trends Biochem Sci,* 30**,** 73-80.

FLACH, J., PILET, P. E. & JOLLES, P. 1992. What's new in chitinase research? *Experientia,* 48**,** 701-16.

FLANAGAN, M. A., ACKERS, G. K., MATTHEW, J. B., HANANIA, G. I. & GURD, F. R. 1981. Electrostatic contributions to the energetics of dimer-tetramer assembly in human hemoglobin: pH dependence and effect of specifically bound chloride ions. *Biochemistry,* 20**,** 7439-49.

FLEMING, P. J. & ROSE, G. D. 2005. Do all backbone polar groups in proteins form hydrogen bonds? *Protein Sci,* 14**,** 1911-7.

FLOCKNER, H., BRAXENTHALER, M., LACKNER, P., JARITZ, M., ORTNER, M. & SIPPL, M. J. 1995. Progress in fold recognition. *Proteins,* 23**,** 376-86.

FUSETTI, F., VON MOELLER, H., HOUSTON, D., ROZEBOOM, H. J., DIJKSTRA, B. W., BOOT, R. G., AERTS, J. M. & VAN AALTEN, D. M. 2002. Structure of human chitotriosidase. Implications for specific inhibitor design and function of mammalian chitinase-like lectins. *J Biol Chem,* 277**,** 25537-44.

FUZO, C. A. & DEGREVE, L. 2012. Effect of the thermostat in the molecular dynamics simulation on the folding of the model protein chignolin. *J Mol Model,* 18**,** 2785-94.

GRUBER, K., KLINTSCHAR, G., HAYN, M., SCHLACHER, A., STEINER, W. & KRATKY, C. 1998. Thermophilic xylanase from Thermomyces lanuginosus: high-resolution X-ray structure and modeling studies. *Biochemistry,* 37**,** 13475-85.

References

GULATI, H., CHADHA, B. & SAINI, H. 2007. Production, purification and characterization of thermostable phytase from thermophilic fungus <i>Thermomyces lanuginosus TL-7. *Acta Microbiologica et Immunologica Hungarica,* 54**,** 121-138.

HACKMAN, R. H. 1954. Studies on chitin. I. Enzymic degradation of chitin and chitin esters. *Aust J Biol Sci,* 7**,** 168-78.

HACKMAN, R. H. & GOLDBERG, M. 1965. Studies on chitin. VI. The nature of alpha- and beta-chitins. *Aust J Biol Sci,* 18**,** 935-46.

HAKI, G. D. & RAKSHIT, S. K. 2003. Developments in industrially important thermostable enzymes: a review. *Bioresour Technol,* 89**,** 17-34.

HARTL, L., ZACH, S. & SEIDL-SEIBOTH, V. 2012. Fungal chitinases: diversity, mechanistic properties and biotechnological potential. *Appl Microbiol Biotechnol,* 93**,** 533-43.

HAYES, C. K., KLEMSDAL, S., LORITO, M., DI PIETRO, A., PETERBAUER, C., NAKAS, J. P., TRONSMO, A. & HARMAN, G. E. 1994. Isolation and sequence of an endochitinase-encoding gene from a cDNA library of Trichoderma harzianum. *Gene,* 138**,** 143-8.

HENRISSAT, B. 1999. Classification of chitinases modules. *EXS,* 87**,** 137-56.

HENRISSAT, B. & BAIROCH, A. 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J,* 293 (Pt 3)**,** 781-8.

HENRISSAT, B. & BAIROCH, A. 1996. Updating the sequence-based classification of glycosyl hydrolases. *Biochem J,* 316 (Pt 2)**,** 695-6.

HOLDEN, Z. C., RICHARD, R. M. & HERBERT, J. M. 2013. Periodic boundary conditions for QM/MM calculations: Ewald summation for extended Gaussian basis sets. *J Chem Phys,* 139**,** 244108.

HOLM, L. & ROSENSTROM, P. 2010. Dali server: conservation mapping in 3D. *Nucleic Acids Res,* 38**,** W545-9.

References

HOLM, L. & SANDER, C. 1996. Mapping the protein universe. *Science,* 273**,** 595-603.

HORNG, J. C., TRACZ, S. M., LUMB, K. J. & RALEIGH, D. P. 2005. Slow folding of a three-helix protein via a compact intermediate. *Biochemistry,* 44**,** 627-34.

HSIN, J., ARKHIPOV, A., YIN, Y., STONE, J. E. & SCHULTEN, K. 2008. Using VMD: an introductory tutorial. *Curr Protoc Bioinformatics,* Chapter 5**,** Unit 5 7.

HSU, Y. H., JOHNSON, D. A. & TRAUGH, J. A. 2008. Analysis of conformational changes during activation of protein kinase Pak2 by amide hydrogen/deuterium exchange. *J Biol Chem,* 283**,** 36397-405.

HUMPHREY, W., DALKE, A. & SCHULTEN, K. 1996. VMD: visual molecular dynamics. *J Mol Graph,* 14**,** 33-8, 27-8.

INC, A. S. 2013. Ref. Accelrys Software Inc., Discovery Studio Modeling Environment, Release 4.0, San Diego: Accelrys Software Inc.

JEON, Y. J., KAMIL, J. Y. & SHAHIDI, F. 2002. Chitosan as an edible invisible film for quality preservation of herring and atlantic cod. *J Agric Food Chem,* 50**,** 5167-78.

JOHN, B. & SALI, A. 2003. Comparative protein structure modeling by iterative alignment, model building and model assessment. *Nucleic Acids Res,* 31**,** 3982-92.

JONES, P., BINNS, D., CHANG, H. Y., FRASER, M., LI, W., MCANULLA, C., MCWILLIAM, H., MASLEN, J., MITCHELL, A., NUKA, G., PESSEAT, S., QUINN, A. F., SANGRADOR-VEGAS, A., SCHEREMETJEW, M., YONG, S. Y., LOPEZ, R. & HUNTER, S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics,* 30**,** 1236-40.

JORGENSEN, W. L. & TIRADO-RIVES, J. 2005. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc Natl Acad Sci U S A,* 102**,** 6665-70.

JUNG, S. H., PASTAN, I. & LEE, B. 1994. Design of interchain disulfide bonds in the framework region of the Fv fragment of the monoclonal antibody B3. *Proteins,* 19**,** 35-47.

KADOKURA, K., ROKUTANI, A., YAMAMOTO, M., IKEGAMI, T., SUGITA, H., ITOI, S., HAKAMATA, W., OKU, T. & NISHIO, T. 2007. Purification and characterization of Vibrio parahaemolyticus extracellular chitinase and chitin oligosaccharide deacetylase involved in the production of heterodisaccharide from chitin. *Appl Microbiol Biotechnol,* 75**,** 357-65.

KANG, T. S. & KINI, R. M. 2009. Structural determinants of protein folding. *Cell Mol Life Sci,* 66**,** 2341-61.

KAPLAN, W. & LITTLEJOHN, T. G. 2001. Swiss-PDB Viewer (Deep View). *Brief Bioinform,* 2**,** 195-7.

KARPLUS, M. & MCCAMMON, J. A. 2002. Molecular dynamics simulations of biomolecules. *Nat Struct Biol,* 9**,** 646-52.

KHUCHAROENPHAISAN, K. & SINMA, K. 2010. Beta-xylanase from Thermomyces lanuginosus and its biobleaching application. *Pak J Biol Sci,* 13**,** 513-26.

KIM, D. E., CHIVIAN, D. & BAKER, D. 2004. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res,* 32**,** W526-31.

KIM, M. O., BLACHLY, P. G., KAUS, J. W. & MCCAMMON, J. A. 2014. Protocols Utilizing Constant pH Molecular Dynamics to Compute pH-Dependent Binding Free Energies. *J Phys Chem B.*

KINI, R. M. & EVANS, H. J. 1991. Molecular modeling of proteins: a strategy for energy minimization by molecular mechanics in the AMBER force field. *J Biomol Struct Dyn,* 9**,** 475-88.

KITCHEN, D. B., DECORNEZ, H., FURR, J. R. & BAJORATH, J. 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov,* 3**,** 935-49.

KOEHL, P. & LEVITT, M. 1999. A brighter future for protein structure prediction. *Nat Struct Biol,* 6**,** 108-11.

References

KOGA, D., SASAKI, Y., UCHIUMI, Y., HIRAI, N., ARAKANE, Y. & NAGAMATSU, Y. 1997. Purification and characterization of Bombyx mori chitinases. *Insect Biochem Mol Biol,* 27**,** 757-67.

KOSKA, J., SPASSOV, V. Z., MAYNARD, A. J., YAN, L., AUSTIN, N., FLOOK, P. K. & VENKATACHALAM, C. M. 2008. Fully automated molecular mechanics based induced fit protein-ligand docking method. *J Chem Inf Model,* 48**,** 1965-73.

KUZMANIC, A. & ZAGROVIC, B. 2010. Determination of ensemble-average pairwise root mean-square deviation from experimental B-factors. *Biophys J,* 98**,** 861-71.

LEE, C. 1996. Testing homology modeling on mutant proteins: predicting structural and thermodynamic effects in the Ala98-->Val mutants of T4 lysozyme. *Fold Des,* 1**,** 1-12.

LEE, C. C., WILLIAMS, T. G., WONG, D. W. & ROBERTSON, G. H. 2005. An episomal expression vector for screening mutant gene libraries in Pichia pastoris. *Plasmid,* 54**,** 80-5.

LEE, P., WAALEN, J., CRAIN, K., SMARGON, A. & BEUTLER, E. 2007a. Human chitotriosidase polymorphisms G354R and A442V associated with reduced enzyme activity. *Blood Cells Mol Dis,* 39**,** 353-60.

LEE, Y. S., PARK, I. H., YOO, J. S., CHUNG, S. Y., LEE, Y. C., CHO, Y. S., AHN, S. C., KIM, C. M. & CHOI, Y. L. 2007b. Cloning, purification, and characterization of chitinase from Bacillus sp. DAU101. *Bioresour Technol,* 98**,** 2734-41.

LESK, A. M. & CHOTHIA, C. 1980. How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J Mol Biol,* 136**,** 225-70.

LESSARD, J. C. 2013. Molecular cloning. *Methods Enzymol,* 529**,** 85-98.

LI, D.-C., YANG, Y.-J. & SHEN, C.-Y. 1997. Protease production by the thermophilic fungus Thermomyces lanuginosus. *Mycological Research,* 101**,** 18-22.

LINDAHL, E. R. 2008. Molecular dynamics simulations. *Methods Mol Biol,* 443**,** 3-23.

References

LOBANOV, M., BOGATYREVA, N. S. & GALZITSKAIA, O. V. 2008. [Radius of gyration is indicator of compactness of protein structure]. *Mol Biol (Mosk),* 42**,** 701-6.

LORITO, M., HAYES, C. K., DI PIETRO, A. & HARMAN, G. E. 1993. Biolistic transformation of Trichoderma harzianum and Gliocladium virens using plasmid and genomic DNA. *Curr Genet,* 24**,** 349-56.

LORITO, M., WOO, S. L., GARCIA, I., COLUCCI, G., HARMAN, G. E., PINTOR-TORO, J. A., FILIPPONE, E., MUCCIFORA, S., LAWRENCE, C. B., ZOINA, A., TUZUN, S. & SCALA, F. 1998. Genes from mycoparasitic fungi as a source for improving plant resistance to fungal pathogens. *Proc Natl Acad Sci U S A,* 95**,** 7860-5.

MACHUQUEIRO, M. & BAPTISTA, A. M. 2006. Constant-pH molecular dynamics with ionic strength effects: protonation-conformation coupling in decalysine. *J Phys Chem B,* 110**,** 2927-33.

MACHUQUEIRO, M. & BAPTISTA, A. M. 2007. The pH-dependent conformational states of kyotorphin: a constant-pH molecular dynamics study. *Biophys J,* 92**,** 1836-45.

MATTHEW, J. B., GURD, F. R., GARCIA-MORENO, B., FLANAGAN, M. A., MARCH, K. L. & SHIRE, S. J. 1985. pH-dependent processes in proteins. *CRC Crit Rev Biochem,* 18**,** 91-197.

MAYER, R. T., MCCOLLUM, T. G., NIEDZ, R. P., HEARN, C. J., MCDONALD, R. E., BERDIS, E. & DOOSTDAR, H. 1996. Characterization of seven basic endochitinases isolated from cell cultures of Citrus sinensis (L.). *Planta,* 200**,** 289-95.

MENG ZHANG, A. K. P., ALGASAN GOVENDER, ZHENGXIANG WANG,SUREN SINGH, KUGENTHIREN PERMAUL 2014. The multi-chitinolytic enzyme system of the compost-dwellingthermophilic fungus Thermomyces lanuginosus. *Process Biochemistry*.

MI, F. L., TAN, Y. C., LIANG, H. C., HUANG, R. N. & SUNG, H. W. 2001. In vitro evaluation of a chitosan membrane cross-linked with genipin. *J Biomater Sci Polym Ed,* 12**,** 835-50.

References

MI, F. L., WONG, T. B. & SHYU, S. S. 1997. Sustained-release of oxytetracycline from chitosan microspheres prepared by interfacial acylation and spray hardening methods. *J Microencapsul,* 14**,** 577-91.

MINKE, R. & BLACKWELL, J. 1978. The structure of alpha-chitin. *J Mol Biol,* 120**,** 167-81.

MISURA, K. M., CHIVIAN, D., ROHL, C. A., KIM, D. E. & BAKER, D. 2006. Physically realistic homology models built with ROSETTA can be more accurate than their templates. *Proc Natl Acad Sci U S A,* 103**,** 5361-6.

MOHAMED, A. J., YU, L., BACKESJO, C. M., VARGAS, L., FARYAL, R., AINTS, A., CHRISTENSSON, B., BERGLOF, A., VIHINEN, M., NORE, B. F. & SMITH, C. I. 2009. Bruton's tyrosine kinase (Btk): function, regulation, and transformation with special emphasis on the PH domain. *Immunol Rev,* 228**,** 58-73.

MULLER, C. W., SCHLAUDERER, G. J., REINSTEIN, J. & SCHULZ, G. E. 1996. Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. *Structure,* 4**,** 147-56.

MURZIN, A. G., BRENNER, S. E., HUBBARD, T. & CHOTHIA, C. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol,* 247**,** 536-40.

MUZZARELLI, R. A. 1999. Native, industrial and fossil chitins. *EXS,* 87**,** 1-6.

NGUYEN, T. T., VIET, M. H. & LI, M. S. 2014. Effects of water models on binding affinity: evidence from all-atom simulation of binding of tamiflu to A/H5N1 neuraminidase. *ScientificWorldJournal,* 2014**,** 536084.

NORBERTO DE SOUZA, O. & ORNSTEIN, R. L. 1999. Molecular dynamics simulations of a protein-protein dimer: particle-mesh Ewald electrostatic model yields far superior results to standard cutoff model. *J Biomol Struct Dyn,* 16**,** 1205-18.

References

NOVAES-LEDIEU, M., MARTINEZ COBO, J. A. & GARCIA MENDOZA, C. 1987. The structure of the mycelial wall of Agaricus bisporus. *Microbiologia,* 3**,** 13-23.

ODIBO, F. J. C. & ULBRICH-HOFMANN, R. 2001. Thermostable α-Amylase and Glucoamylase from Thermomyces lanuginosus F1. *Acta Biotechnologica,* 21**,** 141-153.

ORENGO, C. A., MICHIE, A. D., JONES, S., JONES, D. T., SWINDELLS, M. B. & THORNTON, J. M. 1997. CATH--a hierarchic classification of protein domain structures. *Structure,* 5**,** 1093-108.

PACE, C. N., SHIRLEY, B. A., MCNUTT, M. & GAJIWALA, K. 1996. Forces contributing to the conformational stability of proteins. *FASEB J,* 10**,** 75-83.

PARK, J., KARPLUS, K., BARRETT, C., HUGHEY, R., HAUSSLER, D., HUBBARD, T. & CHOTHIA, C. 1998. Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods. *J Mol Biol,* 284**,** 1201-10.

PATIL, R. S., GHORMADE, V. V. & DESHPANDE, M. V. 2000. Chitinolytic enzymes: an exploration. *Enzyme Microb Technol,* 26**,** 473-483.

PEARSON, W. R. 1998. Empirical statistical estimates for sequence similarity searches. *J Mol Biol,* 276**,** 71-84.

POWERS, R., COPELAND, J. C., GERMER, K., MERCIER, K. A., RAMANATHAN, V. & REVESZ, P. 2006. Comparison of protein active site structures for functional annotation of proteins and drug design. *Proteins,* 65**,** 124-35.

PRONK, S., PALL, S., SCHULZ, R., LARSSON, P., BJELKMAR, P., APOSTOLOV, R., SHIRTS, M. R., SMITH, J. C., KASSON, P. M., VAN DER SPOEL, D., HESS, B. & LINDAHL, E. 2013. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics,* 29**,** 845-54.

RAPPOPORT, N., KARSENTY, S., STERN, A., LINIAL, N. & LINIAL, M. 2012. ProtoNet 6.0: organizing 10 million protein sequences in a compact hierarchical family tree. *Nucleic Acids Res,* 40**,** D313-20.

REZESSY-SZABO, J. M., NGUYEN, Q. D., HOSCHKE, A., BRAET, C., HAJOS, G. & CLAEYSSENS, M. 2007. A novel thermostable alpha-galactosidase from the thermophilic fungus Thermomyces lanuginosus CBS 395.62/b: purification and characterization. *Biochim Biophys Acta,* 1770**,** 55-62.

ROST, B. 1999. Twilight zone of protein sequence alignments. *Protein Eng,* 12**,** 85-94.

ROY, A., KUCUKURAL, A. & ZHANG, Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc,* 5**,** 725-38.

ROY, A., YANG, J. & ZHANG, Y. 2012. COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Res,* 40**,** W471-7.

RYCHLEWSKI, L., ZHANG, B. & GODZIK, A. 1998. Fold and function predictions for Mycoplasma genitalium proteins. *Fold Des,* 3**,** 229-38.

RYCHLIK, W. 2007. OLIGO 7 primer analysis software. *Methods Mol Biol,* 402**,** 35-60.

SALI, A. 1995. Modeling mutations and homologous proteins. *Curr Opin Biotechnol,* 6**,** 437-51.

SALI, A. & OVERINGTON, J. P. 1994. Derivation of rules for comparative protein modeling from a database of protein structure alignments. *Protein Sci,* 3**,** 1582-96.

SALZER, P., FEDDERMANN, N., WIEMKEN, A., BOLLER, T. & STAEHELIN, C. 2004. Sinorhizobium meliloti-induced chitinase gene expression in Medicago truncatula ecotype R108-1: a comparison between symbiosis-specific class V and defence-related class IV chitinases. *Planta,* 219**,** 626-38.

SANCHEZ, R. & SALI, A. 1997. Advances in comparative protein-structure modelling. *Curr Opin Struct Biol,* 7**,** 206-14.

SANCHEZ, R. & SALI, A. 1998. Large-scale protein structure modeling of the Saccharomyces cerevisiae genome. *Proc Natl Acad Sci U S A,* 95**,** 13597-602.

SANDOR, E., PUSZTAHELYI, T., KARAFFA, L., KARANYI, Z., POCSI, I., BIRO, S., SZENTIRMAI, A. & POCSI, I. 1998. Allosamidin inhibits the fragmentation of Acremonium chrysogenum but does not influence the cephalosporin-C production of the fungus. *FEMS Microbiol Lett,* 164**,** 231-6.

SAQI, M. A., RUSSELL, R. B. & STERNBERG, M. J. 1998. Misleading local sequence alignments: implications for comparative protein modelling. *Protein Eng,* 11**,** 627-30.

SCHULTZ, J., COPLEY, R. R., DOERKS, T., PONTING, C. P. & BORK, P. 2000. SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res,* 28**,** 231-4.

SCHYMKOWITZ, J., BORG, J., STRICHER, F., NYS, R., ROUSSEAU, F. & SERRANO, L. 2005. The FoldX web server: an online force field. *Nucleic Acids Res,* 33**,** W382-8.

SEELIGER, D. & DE GROOT, B. L. 2010. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J Comput Aided Mol Des,* 24**,** 417-22.

SEIDL, V., HUEMER, B., SEIBOTH, B. & KUBICEK, C. P. 2005. A complete survey of Trichoderma chitinases reveals three distinct subgroups of family 18 chitinases. *FEBS J,* 272**,** 5923-39.

SHIMONO, K., MATSUDA, H. & KAWAMUKAI, M. 2002. Functional expression of chitinase and chitosanase, and their effects on morphologies in the yeast Schizosaccharomyces pombe. *Biosci Biotechnol Biochem,* 66**,** 1143-7.

SIM, N. L., KUMAR, P., HU, J., HENIKOFF, S., SCHNEIDER, G. & NG, P. C. 2012. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res,* 40**,** W452-7.

SIMOSSIS, V. A. & HERINGA, J. 2005. PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res,* 33**,** W289-94.

SINGH, S., DU PREEZ, J. C., PILLAY, B. & PRIOR, B. A. 2000a. The production of hemicellulases by Thermomyces lanuginosus strain SSBP: influence of agitation and dissolved oxygen tension. *Appl Microbiol Biotechnol,* 54**,** 698-704.

SINGH, S., MADLALA, A. M. & PRIOR, B. A. 2003. Thermomyces lanuginosus: properties of strains and their hemicellulases. *FEMS Microbiol Rev,* 27**,** 3-16.

SINGH, S., PILLAY, B., DILSOOK, V. & PRIOR, B. A. 2000b. Production and properties of hemicellulases by a Thermomyces lanuginosus strain. *J Appl Microbiol,* 88**,** 975-82.

SINGH, S., REDDY, P., HAARHOFF, J., BIELY, P., JANSE, B., PILLAY, B., PILLAY, D. & PRIOR, B. A. 2000c. Relatedness of Thermomyces lanuginosus strains producing a thermostable xylanase. *J Biotechnol,* 81**,** 119-28.

SODING, J., BIEGERT, A. & LUPAS, A. N. 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res,* 33**,** W244-8.

SONG, Y., MAO, J. & GUNNER, M. R. 2009. MCCE2: improving protein pKa calculations with extensive side chain rotamer sampling. *J Comput Chem,* 30**,** 2231-47.

SPASSOV, V. Z. & YAN, L. 2008. A fast and accurate computational approach to protein ionization. *Protein Sci,* 17**,** 1955-70.

STEPHENS, D. E., KHAN, F. I., SINGH, P., BISETTY, K., SINGH, S. & PERMAUL, K. 2014. Creation of thermostable and alkaline stable xylanase variants by DNA shuffling. *J Biotechnol,* 187**,** 139-46.

STRUCTURAL GENOMICS, C., CHINA STRUCTURAL GENOMICS, C., NORTHEAST STRUCTURAL GENOMICS, C., GRASLUND, S., NORDLUND, P., WEIGELT, J., HALLBERG, B. M., BRAY, J., GILEADI, O., KNAPP, S., OPPERMANN, U., ARROWSMITH, C., HUI, R., MING, J., DHE-PAGANON, S., PARK, H. W., SAVCHENKO, A., YEE, A., EDWARDS, A., VINCENTELLI, R., CAMBILLAU, C., KIM, R., KIM, S. H., RAO, Z., SHI, Y., TERWILLIGER, T. C., KIM, C. Y., HUNG, L. W., WALDO, G. S., PELEG, Y., ALBECK, S., UNGER, T., DYM, O., PRILUSKY, J., SUSSMAN, J. L., STEVENS, R. C., LESLEY, S. A., WILSON, I. A., JOACHIMIAK, A., COLLART, F., DEMENTIEVA, I., DONNELLY, M. I., ESCHENFELDT, W. H., KIM, Y., STOLS, L., WU, R., ZHOU, M., BURLEY, S. K., EMTAGE, J. S., SAUDER, J. M., THOMPSON, D., BAIN, K., LUZ, J., GHEYI, T., ZHANG, F., ATWELL, S., ALMO, S. C., BONANNO, J. B., FISER, A.,

SWAMINATHAN, S., STUDIER, F. W., CHANCE, M. R., SALI, A., ACTON, T. B., XIAO, R., ZHAO, L., MA, L. C., HUNT, J. F., TONG, L., CUNNINGHAM, K., INOUYE, M., ANDERSON, S., JANJUA, H., SHASTRY, R., HO, C. K., WANG, D., WANG, H., JIANG, M., MONTELIONE, G. T., STUART, D. I., OWENS, R. J., DAENKE, S., SCHUTZ, A., HEINEMANN, U., YOKOYAMA, S., BUSSOW, K. & GUNSALUS, K. C. 2008. Protein production and purification. *Nat Methods,* 5**,** 135-46.

SULLIVAN, S. M. & HOLYOAK, T. 2008. Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection. *Proc Natl Acad Sci U S A,* 105**,** 13829-34.

SUZUKI, S., NAKANISHI, E., OHIRA, T., KAWACHI, R., NAGASAWA, H. & SAKUDA, S. 2006. Chitinase inhibitor allosamidin is a signal molecule for chitinase production in its producing Streptomyces I. Analysis of the chitinase whose production is promoted by allosamidin and growth accelerating activity of allosamidin. *J Antibiot (Tokyo),* 59**,** 402-9.

SYNOWIECKI, J. & AL-KHATEEB, N. A. 2003. Production, properties, and some new applications of chitin and its derivatives. *Crit Rev Food Sci Nutr,* 43**,** 145-71.

TAKAYA, N., YAMAZAKI, D., HORIUCHI, H., OHTA, A. & TAKAGI, M. 1998. Intracellular chitinase gene from Rhizopus oligosporus: molecular cloning and characterization. *Microbiology,* 144 (Pt 9)**,** 2647-54.

TENG, S., SRIVASTAVA, A. K. & WANG, L. 2010. Sequence feature-based prediction of protein stability changes upon amino acid substitutions. *BMC Genomics,* 11 Suppl 2**,** S5.

TENG, W. L., KHOR, E., TAN, T. K., LIM, L. Y. & TAN, S. C. 2001. Concurrent production of chitin from shrimp shells and fungi. *Carbohydr Res,* 332**,** 305-16.

TERWISSCHA VAN SCHELTINGA, A. C., HENNIG, M. & DIJKSTRA, B. W. 1996. The 1.8 A resolution structure of hevamine, a plant chitinase/lysozyme, and analysis of the conserved sequence and structure motifs of glycosyl hydrolase family 18. *J Mol Biol,* 262**,** 243-57.

THOMPSON, J. D., HIGGINS, D. G. & GIBSON, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res,* 22**,** 4673-80.

THUSBERG, J. & VIHINEN, M. 2009. Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods. *Hum Mutat,* 30**,** 703-14.

TJOELKER, L. W., GOSTING, L., FREY, S., HUNTER, C. L., TRONG, H. L., STEINER, B., BRAMMER, H. & GRAY, P. W. 2000. Structural and functional definition of the human chitinase chitin-binding domain. *J Biol Chem,* 275**,** 514-20.

TRUDEL, J. & ASSELIN, A. 1989. Detection of chitinase activity after polyacrylamide gel electrophoresis. *Anal Biochem,* 178**,** 362-6.

UNGER, R., HAREL, D., WHERLAND, S. & SUSSMAN, J. L. 1989. A 3D building blocks approach to analyzing and predicting structure of proteins. *Proteins,* 5**,** 355-73.

VAN AALTEN, D. M., SYNSTAD, B., BRURBERG, M. B., HOUGH, E., RIISE, B. W., EIJSINK, V. G. & WIERENGA, R. K. 2000. Structure of a two-domain chitotriosidase from Serratia marcescens at 1.9-A resolution. *Proc Natl Acad Sci U S A,* 97**,** 5842-7.

VAN DER SPOEL, D., LINDAHL, E., HESS, B., GROENHOF, G., MARK, A. E. & BERENDSEN, H. J. 2005. GROMACS: fast, flexible, and free. *J Comput Chem,* 26**,** 1701-18.

VANOMMESLAEGHE, K., HATCHER, E., ACHARYA, C., KUNDU, S., ZHONG, S., SHIM, J., DARIAN, E., GUVENCH, O., LOPES, P., VOROBYOV, I. & MACKERELL, A. D., JR. 2010. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem,* 31**,** 671-90.

VASQUEZ, M. 1996. Modeling side-chain conformation. *Curr Opin Struct Biol,* 6**,** 217-21.

VISHU KUMAR, A. B., VARADARAJ, M. C., GOWDA, L. R. & THARANATHAN, R. N. 2005. Characterization of chito-oligosaccharides prepared by chitosanolysis with the aid of papain and

Pronase, and their bactericidal action against Bacillus cereus and Escherichia coli. *Biochem J,* 391**,** 167-75.

VONGCHAN, P., SAJOMSANG, W., SUBYEN, D. & KONGTAWELERT, P. 2002. Anticoagulant activity of a sulfated chitosan. *Carbohydr Res,* 337**,** 1239-42.

WANG, Q., CANUTESCU, A. A. & DUNBRACK, R. L., JR. 2008. SCWRL and MolIDE: computer programs for side-chain conformation prediction and homology modeling. *Nat Protoc,* 3**,** 1832-47.

WATANABE, T., KANAI, R., KAWASE, T., TANABE, T., MITSUTOMI, M., SAKUDA, S. & MIYASHITA, K. 1999. Family 19 chitinases of Streptomyces species: characterization and distribution. *Microbiology,* 145 (Pt 12)**,** 3353-63.

WITHAM, S., TALLEY, K., WANG, L., ZHANG, Z., SARKAR, S., GAO, D., YANG, W. & ALEXOV, E. 2011. Developing hybrid approaches to predict pKa values of ionizable groups. *Proteins,* 79**,** 3389-99.

WORTH, C. L., PREISSNER, R. & BLUNDELL, T. L. 2011. SDM--a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res,* 39**,** W215-22.

WU, G., ROBERTSON, D. H., BROOKS, C. L., 3RD & VIETH, M. 2003. Detailed analysis of grid-based molecular docking: A case study of CDOCKER-A CHARMm-based MD docking algorithm. *J Comput Chem,* 24**,** 1549-62.

XIE, W., XU, P. & LIU, Q. 2001. Antioxidant activity of water-soluble chitosan derivatives. *Bioorg Med Chem Lett,* 11**,** 1699-701.

YANG, A. S. & HONIG, B. 1993. On the pH dependence of protein stability. *J Mol Biol,* 231**,** 459-74.

YANG, J., ROY, A. & ZHANG, Y. 2013. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics,* 29**,** 2588-95.

References

YUE, P., LI, Z. & MOULT, J. 2005. Loss of protein structure stability as a major causative factor in monogenic disease. *J Mol Biol,* 353**,** 459-73.

ZHANG, Z., LI, Y., LIN, B., SCHROEDER, M. & HUANG, B. 2011. Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. *Bioinformatics,* 27**,** 2083-8.

BRACONNOT H. Sur la nature des champignons. Ann Chim Phys. 1811;79:265–304.

GOODAY GW. The ecology of chitin degradation. Adv Microb Ecol. 1990;11:387–430.

FELSE PA, PANDA T. 1999. Studies on applications of chitin and its derivatives. Bioprocess Eng. 20:505–512

KUMAR, N. V. R. M. A review of chitin and chitosan applications. React. Funct. Polym. 2000, 46, 1-27

SMIDSRØD O. AND MOE S.T. (1995). Biopolymerkjemi. Tapir Forlag, Trondheim.

SANNAN T, KURITA K, IWAKURA Y. Studies on Chitin, 2. Effect of deacetylation on solubility. Macromol Chem. 1976;177:3589–3600.

RUDALL KM, KENCHINGTON W. The chitin system. Biol Rev. 1973;48:597–633.

MUZZARELLI RAA. Chitin. Pergamon Press; Oxford, UK: 1977.

ASHFORD NA, HATTIS DB, MURRAY AE. MIT Sea Grant Program. Massachusetts Institute of Technology; 1977

KRAJEWSKA B. 2004. Application of chitin- and chitosan-based materials for enzyme immobilizations: a review. Enzyme Microb. Technol. 35:126–139

References

DAHIYA N, TEWARI R, TIWARI RP, HOONDAL GS. Production of an antifungal chitinase from Enterobacter sp. NRG4 and its application in protoplast production. World J Microbiol Biotechnol. 2005;21(8–9):1611–1616.

LIN HY, CHOU CC. Antioxidant activities of water-soluble disaccharide chitosan derivatives. Food Res Int. 2004;37:883–889. doi: 10.1016/j.foodres.2004.04.007

LLOYD LL, KENNEDY JF, METHACANON P, PATERSON M, KNILL CJ. Carbohydrate polymers as wound management aids. Carbohydr Polym. 1998;37:315–322. doi: 10.1016/S0144-8617(98)00077-0

OHYA Y, SHIRATANI M, KOBAYASHI H, OUCHI T. Release behavior of 5-fluorouracil from chitosan-gel nanospheres immobilizing 5-fluorouracil coated with polysaccharides and their cell specific cytotoxicity. J Macromol Sci Pure Appl Chem. 1994;31:629.

GOODAY GW. Aggresive and Defensive Roles for Chitinases. In: Muzzarelli RA, editor. Chitin Enzymology. Italia: Atec Edizioni; 1996. pp. 125–133.

YAMANAKA S, TSUYOSHI N, KIKUCHI R, TAKAYAMA S, SAKUDA S, YAMADA Y. Effect of demethylallosamidin, a chitinase inhibitor, on morphology of fungus Geotrichum candidum. J Gen Appl Microbiol. 1994;40:171–4.

ARIMA K, LIU W-H, BEPPU T. Studies on the lipase of thermophilic fungus Humicola lanuginosa. Agric Biol Chem. 1972;36:893–895

LODISH H, BERK A, ZIPURSKY SL, et al. Molecular Cell Biology. 4th edition. New York: W. H. Freeman; 2000. Section 7.1, DNA Cloning with Plasmid Vectors.

GRIFFITHS AJF, GELBART WM, MILLER JH, et al. Modern Genetic Analysis. New York: W. H. Freeman; 1999. Cloning a Specific Gene.

References

BERG JM, TYMOCZKO JL, STRYER L. Biochemistry. 5th edition. New York: W H Freeman; 2002. Section 6.2, Recombinant DNA Technology Has Revolutionized All Aspects of Biology.

HOLMES RK, JOBLING MG. Genetics. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996.

Schrodinger The PyMOL Molecular Graphics System, Version 1.2r3pre. LLC.

KRAHE, M., ANTRANIKIAN, G., MARKEL, H., 1996. Fermentation of extremophilic microorganisms. FEMS Microbiol. Rev. 18, 271–285

Appendix I

Commands of Gromacs

(1) Creating a gromacs topology file from PDB file

pdb2gmx -f chitinase.pdb -o chi3_processed.gro -water spce

(2) Defining the Unit Cell using editconf & Adding Solvent is done by genbox

editconf -f chi3_processed.gro -o chi3_newbox.gro -c -d 1.0 -bt cubic

The command centers the protein in the box (-c), and put it at least 1.0 nm from the box edge (-d 1.0). The type of box was defined as a cube (-bt cubic). A solute box distance of 1.0 nm means that there is at least 2.0 nm between any two protein periodic images.

genbox -cp chi3_newbox.gro -cs spc216 -o chi3_solvate.gro -p topol.top

The protein configuration (-cp) is present in the output of previous step and the SPC/E 216 solvent configuration (-cs) was selected.

(3) Neutralization by adding ions

The neutralization was carried out in two steps:

(i) Preparation of ion.tpr using grompp module:

grompp -maxwarn 1 -f ions.mdp -c chi3_solvate.gro -p topol.top -o ions.tpr

The ions.tpr has the input that has all the parameters of atoms present in the system.

(ii) Neutralization of system using genion:

genion -s ions.tpr -o chi3_ions.gro -p topol.top -pname NA -nname CL -conc 0.150 -neutral

(4) Running energy minimization

(i) grompp -maxwarn 1 -f minim.mdp -c chi3_ions.gro -p topol.top -o em.tpr

The 50000 of steepest descent was used to minimize the system.

(ii) mdrun -v -deffnm em

(5) NVT Equilibration

(i) grompp -maxwarn 1 -f nvt.mdp -c em.gro -p topol.top -o nvt.tpr

Creation of nvt.tpr take place in this step which is input for mdrun module that perform equilibration.

(ii) mdrun -v -deffnm nvt

(6) NPT Equlilibration

(i) grompp -maxwarn 1 -f npt.mdp -c nvt.gro -t nvt.cpt -p topol.top -o npt.tpr
(ii) mdrun -v -deffnm npt

(7) MD production at CHPC cluster

The MD production is take place by executing the following PBS script at CHPC cluster

Appendix II

Ions.mdp

```
integrator  = steep                 ; Algorithm (steep = steepest descent minimization)
emtol               = 1000.0   ; Stop minimization when the maximum force < 1000.0 kJ/mol/nm
emstep      = 0.01     ; Energy step size
nsteps              = 50000             ; Maximum number of (minimization) steps to perform

; Parameters describing how to find the neighbors of each atom and how to calculate the interactions
nstlist             = 1                 ; Frequency to update the neighbor list and long range forces
cutoff-scheme   = Verlet
ns_type             = grid              ; Method to determine neighbor list (simple, grid)
coulombtype         = PME               ; Treatment of long range electrostatic interactions
rcoulomb      = 1.0                     ; Short-range electrostatic cut-off
rvdw                = 1.0               ; Short-range Van der Waals cut-off
pbc                   = xyz             ; Periodic Boundary Conditions (yes/no)
```

Appendix III

Minim.mdp

```
integrator  = steep                 ; Algorithm (steep = steepest descent minimization)
emtol                = 1000.0   ; Stop minimization when the maximum force < 1000.0 kJ/mol/nm
emstep     = 0.01     ; Energy step size
nsteps               = 50000              ; Maximum number of (minimization) steps to perform

; Parameters describing how to find the neighbors of each atom and how to calculate the interactions
nstlist              = 1                  ; Frequency to update the neighbor list and long range forces
cutoff-scheme   = Verlet
ns_type              = grid               ; Method to determine neighbor list (simple, grid)
coulombtype          = PME                ; Treatment of long range electrostatic interactions
rcoulomb     = 1.0                ; Short-range electrostatic cut-off
rvdw                 = 1.0                ; Short-range Van der Waals cut-off
pbc                  = xyz                ; Periodic Boundary Conditions (yes/no)
```

Appendix IV

Nvt.mdp

```
title                    = OPLS chitinases NVT equilibration
define                   = -DPOSRES          ; position restrain the protein
; Run parameters
integrator  = md                    ; leap-frog integrator
nsteps                   = 50000             ; 2 * 50000 = 100 ps
dt                        = 0.002            ; 2 fs
; Output control
nstxout                  = 500               ; save coordinates every 1.0 ps
nstvout                  = 500               ; save velocities every 1.0 ps
nstenergy   = 500                   ; save energies every 1.0 ps
nstlog                   = 500               ; update log file every 1.0 ps
; Bond parameters
continuation                = no            ; first dynamics run
constraint_algorithm    = lincs         ; holonomic constraints
constraints          = all-bonds     ; all bonds (even heavy atom-H bonds) constrained
lincs_iter           = 1                 ; accuracy of LINCS
lincs_order          = 4                 ; also related to accuracy
; Neighborsearching
cutoff-scheme   = Verlet
ns_type                  = grid              ; search neighboring grid cells
nstlist                  = 10                ; 20 fs, largely irrelevant with Verlet
rcoulomb      = 1.0                 ; short-range electrostatic cutoff (in nm)
rvdw                     = 1.0               ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype              = PME  ; Particle Mesh Ewald for long-range electrostatics
pme_order    = 4                    ; cubic interpolation
fourierspacing           = 0.16      ; grid spacing for FFT
; Temperature coupling is on
tcoupl                   = V-rescale                     ; modified Berendsen thermostat
tc-grps                  = Protein Non-Protein; two coupling groups - more accurate
tau_t                    = 0.1          0.1          ; time constant, in ps
ref_t                    = T*           T*           ; reference temperature, one for each group, in K
; Pressure coupling is off
pcoupl                   = no                ; no pressure coupling in NVT
; Periodic boundary conditions
pbc                      = xyz                   ; 3-D PBC
; Dispersion correction
DispCorr   = EnerPres; account for cut-off vdW scheme
; Velocity generation
gen_vel                  = yes               ; assign velocities from Maxwell distribution
gen_temp   = T*                     ; temperature for Maxwell distribution
gen_seed   = -1                     ; generate a random seed                        *T= 300, 325, 350, 375 K
```

Appendix V

Npt.mdp

```
title                    = OPLS chitinases NPT equilibration
define                   = -DPOSRES              ; position restrain the protein
; Run parameters
integrator     = md                  ; leap-frog integrator
nsteps                   = 50000                 ; 2 * 50000 = 100 ps
dt                         = 0.002               ; 2 fs
; Output control
nstxout                  = 500                   ; save coordinates every 1.0 ps
nstvout                  = 500                   ; save velocities every 1.0 ps
nstenergy     = 500                  ; save energies every 1.0 ps
nstlog                   = 500                   ; update log file every 1.0 ps
; Bond parameters
continuation         = yes               ; Restarting after NVT
constraint_algorithm    = lincs              ; holonomic constraints
constraints              = all-bonds         ; all bonds (even heavy atom-H bonds) constrained
lincs_iter               = 1                 ; accuracy of LINCS
lincs_order              = 4                 ; also related to accuracy
; Neighborsearching
cutoff-scheme   = Verlet
ns_type                  = grid              ; search neighboring grid cells
nstlist                  = 10          ; 20 fs, largely irrelevant with Verlet scheme
rcoulomb        = 1.0                ; short-range electrostatic cutoff (in nm)
rvdw                     = 1.0               ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype              = PME               ; Particle Mesh Ewald for long-range electrostatics
pme_order       = 4                  ; cubic interpolation
fourierspacing           = 0.16              ; grid spacing for FFT
; Temperature coupling is on
tcoupl                   = V-rescale            ; modified Berendsen thermostat
tc-grps                  = Protein Non-Protein     ; two coupling groups - more accurate
tau_t                    = 0.1          0.1             ; time constant, in ps
ref_t                    = T*           T*              ; reference temperature, one for each group, in K
; Pressure coupling is on
pcoupl                       = Parrinello-Rahman             ; Pressure coupling on in NPT
pcoupltype          = isotropic                 ; uniform scaling of box vectors
tau_p                        = 2.0                       ; time constant, in ps
ref_p                        = 1.0                       ; reference pressure, in bar
compressibility     = 4.5e-5                ; isothermal compressibility of water, bar^-1
refcoord_scaling    = com
; Periodic boundary conditions
pbc                      = xyz               ; 3-D PBC
; Dispersion correction
DispCorr     = EnerPres   ; account for cut-off vdW scheme
; Velocity generation
gen_vel                  = no                ; Velocity generation is off
  *T= 300, 325, 350, 370 K
```

Appendix VI

Md.mdp

```
title                    = chitinases MD simulation
; Run parameters
integrator     = md                    ; leap-frog integrator
nsteps                   = 5000000   ; 2 * 5000000 = 10000 ps (10 ns)
dt                       = 0.002                 ; 2 fs
; Output control
nstxout                  = 5000              ; save coordinates every 10.0 ps
nstvout                  = 5000              ; save velocities every 10.0 ps
nstenergy          = 5000              ; save energies every 10.0 ps
nstlog                   = 5000              ; update log file every 10.0 ps
nstxout-compressed  = 5000      ; save compressed coordinates every 10.0 ps
                    ; nstxout-compressed replaces nstxtcout
compressed-x-grps   = System    ; replaces xtc-grps
; Bond parameters
continuation        = yes               ; Restarting after NPT
constraint_algorithm    = lincs             ; holonomic constraints
constraints              = all-bonds        ; all bonds (even heavy atom-H bonds) constrained
lincs_iter               = 1                ; accuracy of LINCS
lincs_order              = 4                ; also related to accuracy
; Neighborsearching
cutoff-scheme   = Verlet
ns_type                  = grid                  ; search neighboring grid cells
nstlist                  = 10          ; 20 fs, largely irrelevant with Verlet scheme
rcoulomb        = 1.0                   ; short-range electrostatic cutoff (in nm)
rvdw                     = 1.0                   ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype              = PME                   ; Particle Mesh Ewald for long-range electrostatics
pme_order       = 4                     ; cubic interpolation
fourierspacing           = 0.16                  ; grid spacing for FFT
; Temperature coupling is on
tcoupl                   = V-rescale             ; modified Berendsen thermostat
tc-grps                  = Protein Non-Protein     ; two coupling groups - more accurate
tau_t                    = 0.1           0.1             ; time constant, in ps
ref_t                    = T*            T*              ; reference temperature, one for each group, in K
; Pressure coupling is on
pcoupl                   = Parrinello-Rahman             ; Pressure coupling on in NPT
pcoupltype         = isotropic                 ; uniform scaling of box vectors
tau_p                    = 2.0                       ; time constant, in ps
ref_p                    = 1.0                       ; reference pressure, in bar
compressibility     = 4.5e-5                ; isothermal compressibility of water, bar^-1
; Periodic boundary conditions
pbc                      = xyz                   ; 3-D PBC
; Dispersion correction
DispCorr     = EnerPres   ; account for cut-off vdW scheme
; Velocity generation
gen_vel                  = no                    ; Velocity generation is off
*T= 300, 325, 350, 375 K
```

Appendix VII

Lists of predicted point mutations on the basis of their RI and stability score. (WT= wild type)

	Amino Acids Substitution Stability(+) and RI(reliability index 0-10), pH=7, Temp=25℃																			
	V	**L**	**I**	**M**	**F**	**W**	**Y**	**G**	**A**	**P**	**S**	**T**	**C**	**H**	**R**	**K**	**Q**	**E**	**N**	**D**
24	1							WT		1								4		
26																		2		WT
28	1		1		0					2	WT							4		0
29	1		1							2	WT							4		1
31	2		3	1	2	2	1			2	WT							4		2
32	1		1	1														0	WT	
36							WT											1		
38								WT										0		
39	0	6	3		0		1		1				3			4	WT	4		
40	5	3	4	4						1								2	WT	
41	5	2	5	4	4	3	3			4	WT				2	3		5	2	3
45	6	4	8	6	5	6	6		3	5	WT	0	1	2	4	5	5	7	5	5
46	4	3	5	2			1	WT	6	1			1					4		
47	5	4	4	1			2	WT	5	2			2					5		
48	4	2	3				0	WT	3	0			0					4		
	V	L	I	M	F	W	Y	G	A	P	S	T	C	H	R	K	Q	E	N	D
50	7	5	8	5	5	5	6		0	6	WT	0	2	1	3	6	5	7	5	4
51		4	5				1						2			3	WT	3		
52		2	4				1						2			3	WT	2		
55									WT	2										
56			0				1					WT								
59	2	2	2	3			1			3								WT		
60							0											5		WT
62	5	3	3	2						1								1	WT	
63		0	WT																	
64		1																4		WT
74		1	WT																	
75	4	5	1	3						0		0							WT	
76		1						WT												
82	2	4	1	5			1			4				2				WT		
84	3	2	3	5						1		2							WT	

87	2	0	1	2															WT	
90																		2		WT
91	1			2			1			1	WT							1		
	V	L	I	M	F	W	Y	G	A	P	S	T	C	H	R	K	Q	E	N	D
96																		2		WT
99	2		3	2														1	WT	
101	4			3			4			5						WT				
102							0											3		WT
103							0					4	WT							
	V	L	I	M	F	W	Y	G	A	P	S	T	C	H	R	K	Q	E	N	D
105			0														WT			
108	3	2	5	7	5	0	4			3	1			1	2			WT		
109					0		0				2	0						2		WT
111	5			5	4		5			6	1	0				WT				
112	7	1		6	5		6			7	3	1	0			WT	1	1	1	
113			1	1	2	1	2			2	2	4	WT					3		
114		1	3				1										WT	4		
115	0		3	2	2	1				1	WT	1						4		2
117								WT										2		
118	7	3	1	7	3		5			7	5	3	1			WT	3	2	0	1
123	5	2	5	4	3	4	5			6	WT	3		0		4	4	7	3	6
124			WT															0	1	
125	4	1		1			1	WT	1	3								6		
126	3	2	0	0			1	WT	1	2								6		
127	1								WT	4										
128				0			1					WT								
130	0		2	2			2					WT							1	
131	5	6	4	4			4			4			0	3				WT		
132	4	5	1					WT					0					4		
133	5	6	2	0			1	WT	0				2					4		
134	0	4	1	0	WT	2	4			1			2			1		2	0	
135	4	8	5	1	1	1	4		1				6	1	2	6	WT	4	1	
136	9	9	9	8	8	7	8		4	7	WT	6	7	6	7	8	7	8	8	6
137		0							WT											

138	5	7	3	3			3			3			1	1				WT		
	V	L	I	M	F	W	Y	G	A	P	S	T	C	H	R	K	Q	E	N	D
141	4	7	4	3			3			2			1	1				WT		
143	3	6	2					WT					2					3		
145	0	1											2		WT					
146		1		WT	0								2					3		
147	WT	1																		
149	2	3																WT		
150	0						0					WT			0				1	
	V	L	I	M	F	W	Y	G	A	P	S	T	C	H	R	K	Q	E	N	D
156	5	1	2	1														1	WT	
158	1						0											3		WT
159	3								WT											
166	1								WT	0										
169																		1		WT
172																			1	WT
174																		2		WT
176	4	2		2						1								WT		
177	1								WT											
178	0			1			0					WT			3				1	
180	7	4	4	3	5	2	4			4	WT		1		4	4	1	6	3	3
181	4	2	0	0															WT	
182				WT														0		
184		0								WT										
	V	L	I	M	F	W	Y	G	A	P	S	T	C	H	R	K	Q	E	N	D
187	3	2	1	1						1								1	WT	
191	4	4	5	3	5	4	2			4	WT				4	2		6	3	4
194																			0	WT
195	5	3	6	5	5	5	4			6	WT	0			3	2	2	7	4	5
196																		1		WT
197	6	5	7	5	5	5	5			6	WT	1	1		4	4	3	7	5	5
198	6	4	7	4	5	4	4			6	WT				4	3	2	7	4	4

199	7	4	1	5	3		5			7			2			WT	0		2	
200		5	6				3						4			2	WT	6		
201	1	0	3				WT						3					1		
204			4	0			2					WT			2					
205			1						WT	1										
206	0		1						WT	0										
	V	L	I	M	F	W	Y	G	A	P	S	T	C	H	R	K	Q	E	N	D
208		3	3										1				WT	1		
210		1								WT										
212		0								WT										
213	4	5	4	4	3		5		1		2	0	5	2	5	2	3	6	1	WT
216	5	2	4	2															WT	
217	6	2		2	1		3			3						WT				
218	5	6	4	5			2			2				2	2			WT		
221	0	4	2	3			2						2		4	0		4		WT
225	4	3	5	4	5	2	2			1	WT				4	0		4	2	2
227																		1		WT
232		4	1										1				WT	1		
	V	L	I	M	F	W	Y	G	A	P	S	T	C	H	R	K	Q	E	N	D
235	2		1	1															WT	
236	2		1	2															WT	
241	2			1															WT	
242	3			1	3	1	2			1	WT				0			0		0
246	2		2	2															WT	
248	1	1	2	2															WT	
250	2	3	2	3															WT	
257		6	3	1	4	1	3					2	4		3		WT	5		
258	5	6	5	6	5	2	2			0		6	2		3			4	WT	0
260	3	3	5	5	7	5	4			2	WT	5			6	1		7	1	4
	V	L	I	M	F	W	Y	G	A	P	S	T	C	H	R	K	Q	E	N	D
261	7	3		7	7	0	6			6		3	2		3	WT			2	1
262	5	3	4	5	5	1	0			0		5			1			3	WT	
263	7	2		5	7		6			7		2	1		1	WT			1	1
264	7	5	5	6	6	2	2			2		6	1		2			5	WT	0

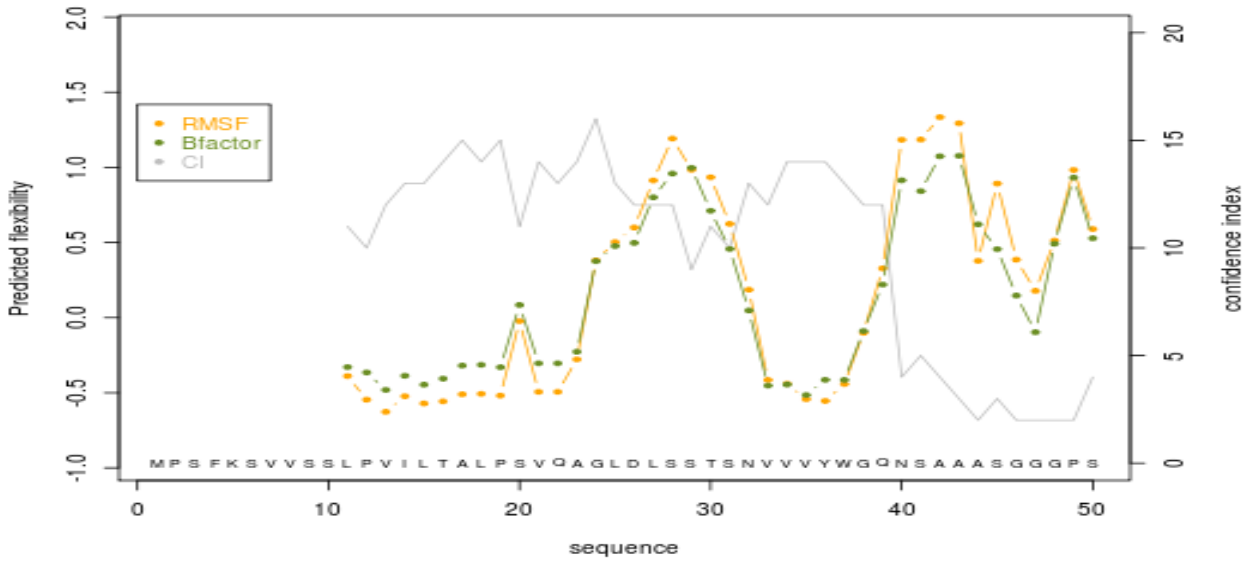
266	7	2		4	6		5			6		1	1			WT				0
273									WT	2										
274	7	2	4	5	2		1					2						4	WT	
275	0											WT								
276	4							WT	1	1								5		
277	2								WT	5										
278									WT	2										
279	3							WT	0									3		
280	7	2	6	4	4	3	5			5	WT				4	5	3	6	3	3
281	3							WT										2		
282	2						WT									0		0		
	V	L	I	M	F	W	Y	G	A	P	S	T	C	H	R	K	Q	E	N	D
286	3		1	1			2								1	1		6		WT
287		1	2													2	WT	1		
289	2								WT	4										
293	5	3	2	1						2								W T		
294	3	5	1				2			4			3	WT				5		
295	2								WT	2										
297	2	0	1									WT			1				2	
299		1								WT										
300	5	5	4	2	4	2	2			2	WT				3	2		4	3	4
302	0	3						WT										2		
303	0	4	0					WT										2		
304	W T	1																		
308	1	4	1	3	0		2						1	0	2			4		WT
	V	L	I	M	F	W	Y	G	A	P	S	T	C	H	R	K	Q	E	N	D
310	5	6	6	6	5	3	4			3	WT		0		6	2	2	4	5	3
311		6	3										2			1	WT	0		
313		2					WT						1							
315	3	4	5	4															WT	
316																		0		WT
320	2		3	3	0	1	0			1	WT							3		1

323	2			2						4						WT				
327								WT										2		
328						0				1	WT							2		2
	V	L	I	M	F	W	Y	G	A	P	S	T	C	H	R	K	Q	E	N	D
331	1	1			1	1				1	WT							1		3
334	7	5		4	3		3			6			0			WT			3	1
341		0																		WT

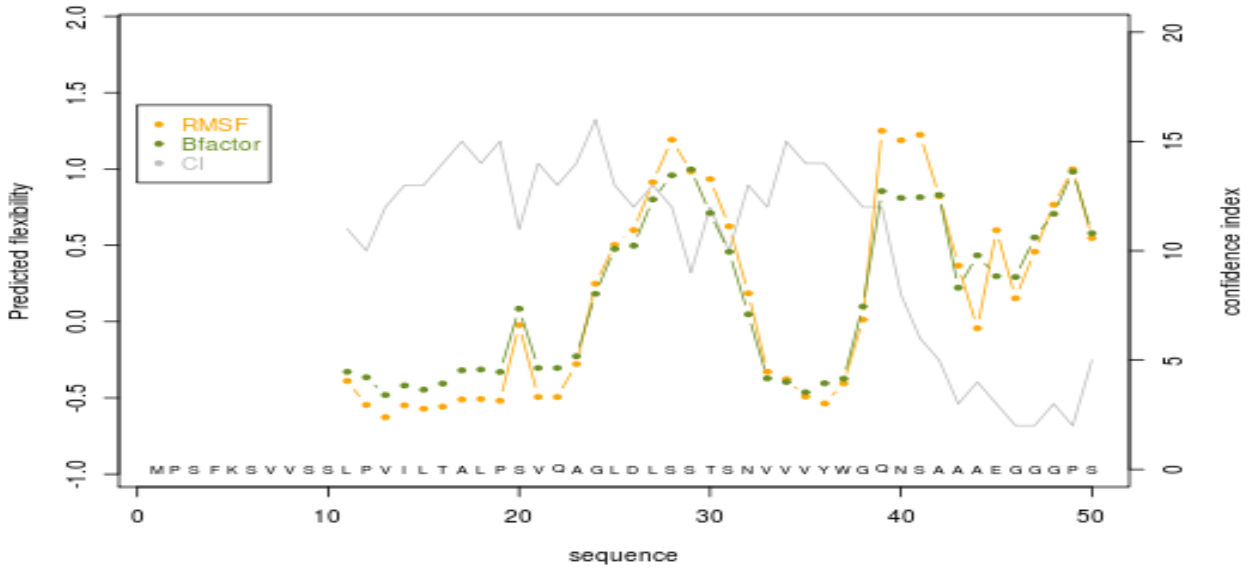
Appendix VIII

RMSF, B-factor and confidence index for (A) normal chitinase II (B) S45E, (C) S45I, (D) S50E, (E) S50I, (F) S50V, (G) normal chitinase II, (H) Q135C, (I) S136I, (J) S136K, (K) S136L and (L) S136V

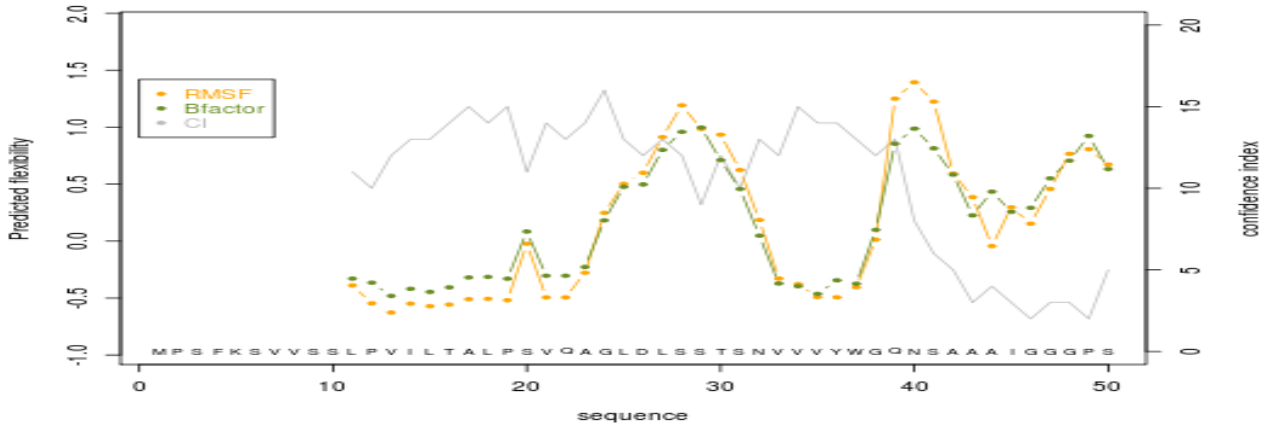
(A)



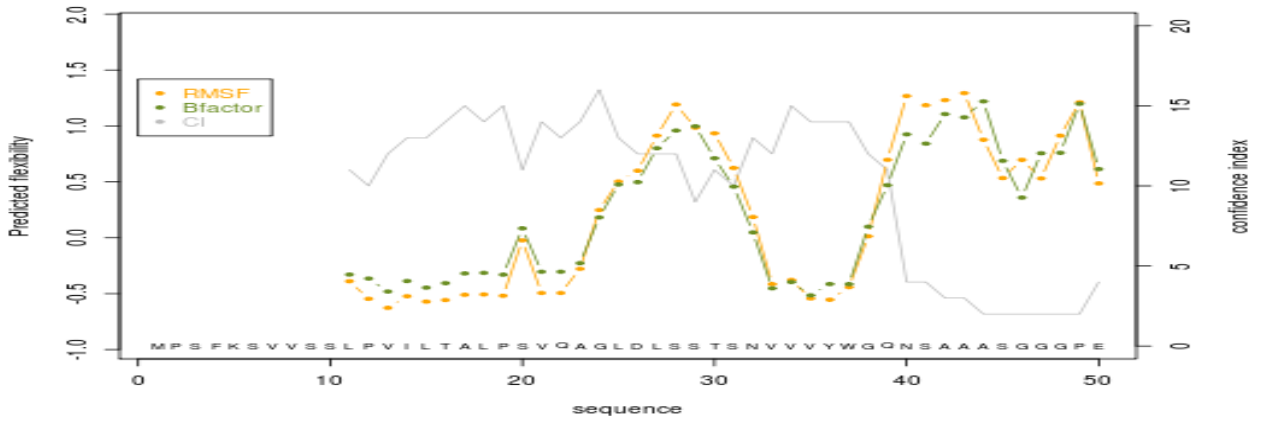
(B)



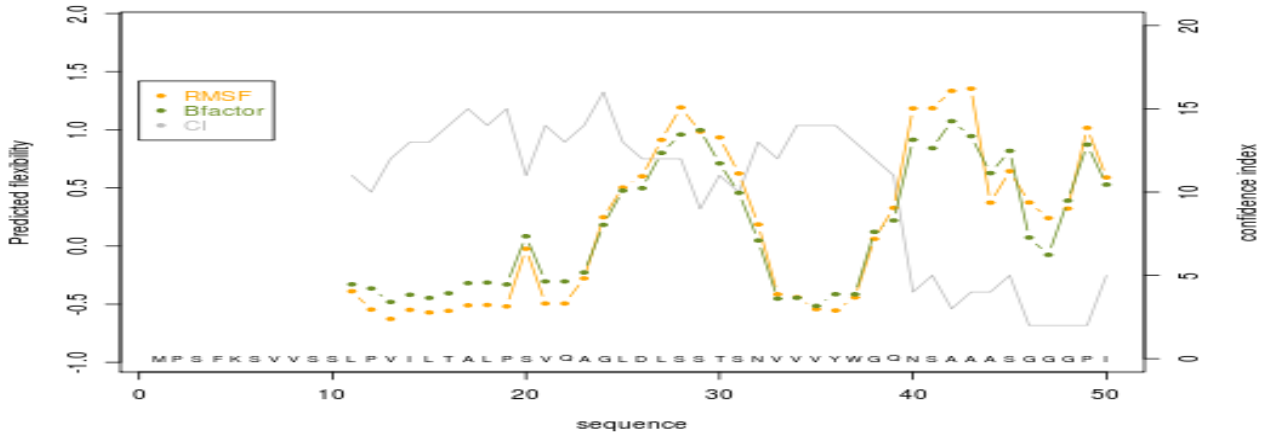
(C)


RMSF
Bfactor
CI
Predicted flexibility
confidence index
sequence
MPSFKSVVSSLPVILTALPSVQAGLDLSSTSNVVVYWGQNSAAAIGGGPS

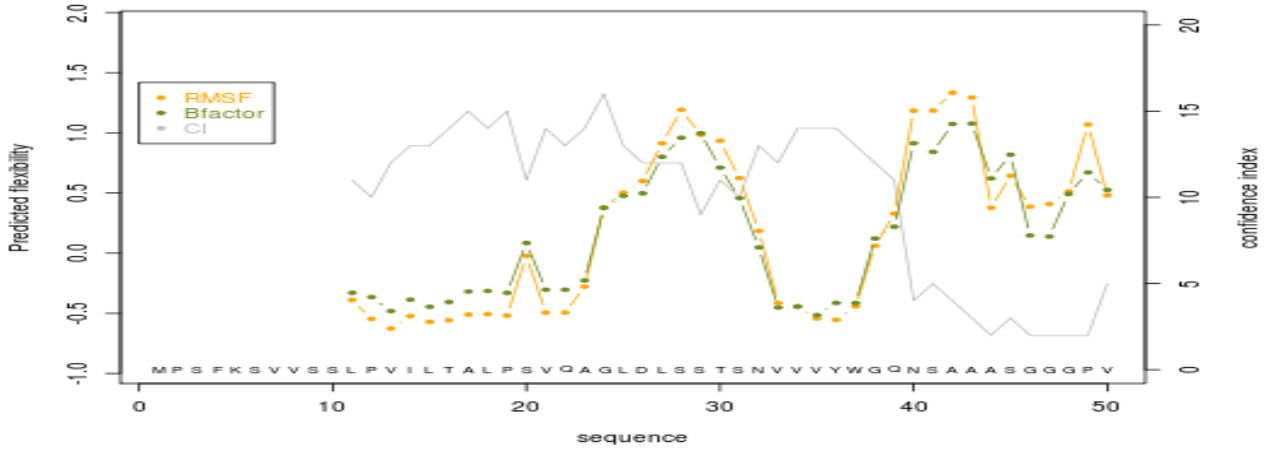
(D)


RMSF
Bfactor
CI
Predicted flexibility
confidence index
sequence
MPSFKSVVSSLPVILTALPSVQAGLDLSSTSNVVVYWGQNSAAASGGGPE

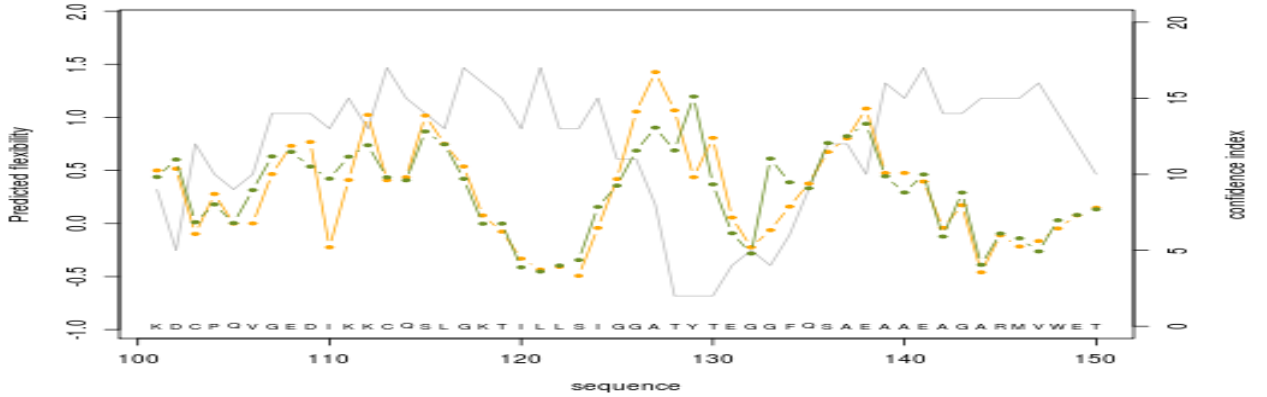
(E)


RMSF
Bfactor
CI
Predicted flexibility
confidence index
sequence
MPSFKSVVSSLPVILTALPSVQAGLDLSSTSNVVVYWGQNSAAASGGGPI

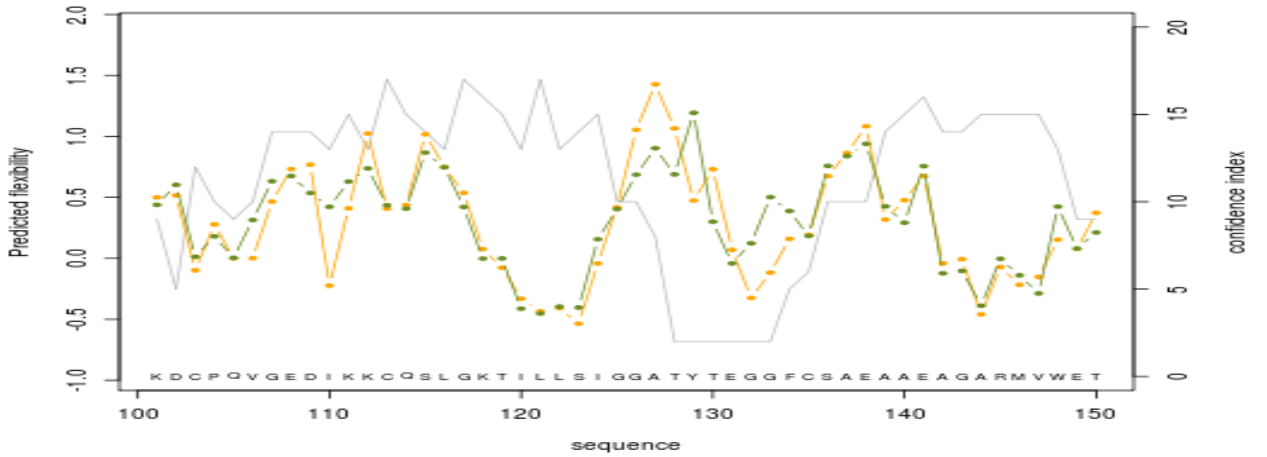
(F)


RMSF
Bfactor
CI
Predicted flexibility
confidence index
sequence
M P S F K S V V S S L P V I L T A L P S V Q A G L D L S S T S N V V V Y W G Q N S A A A S G G G P V

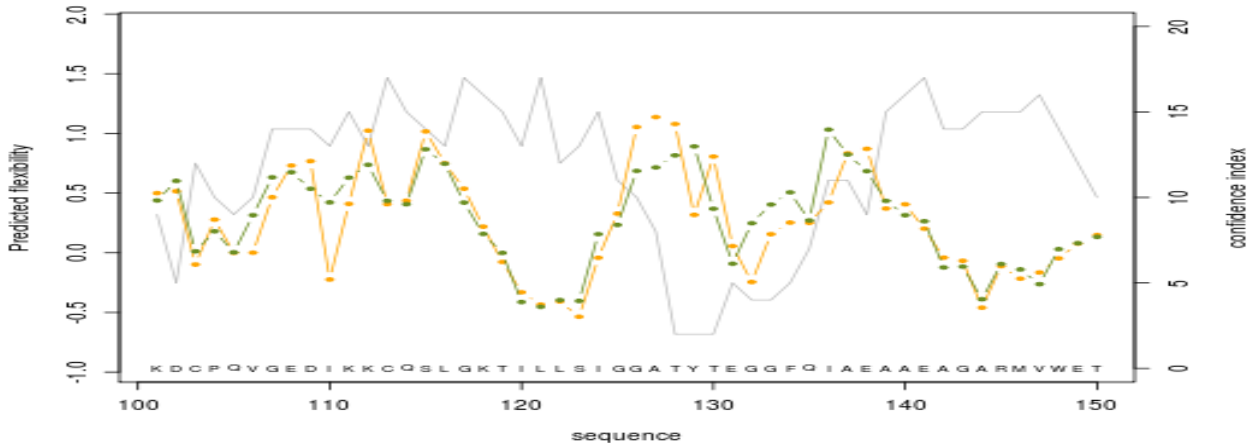
(G)


Predicted flexibility
confidence index
sequence
K D C P Q V G E D I K K C Q S L G K T I L L S I G G A T Y T E G G F Q S A E A A E A G A R M V W E T

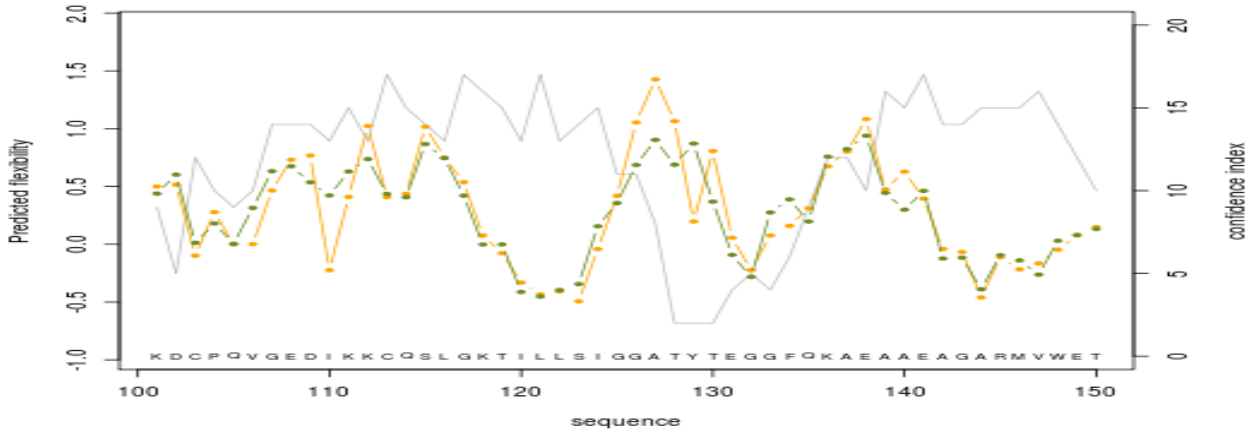
(H)


Predicted flexibility
confidence index
sequence
K D C P Q V G E D I K K C Q S L G K T I L L S I G G A T Y T E G G F C S A E A A E A G A R M V W E T

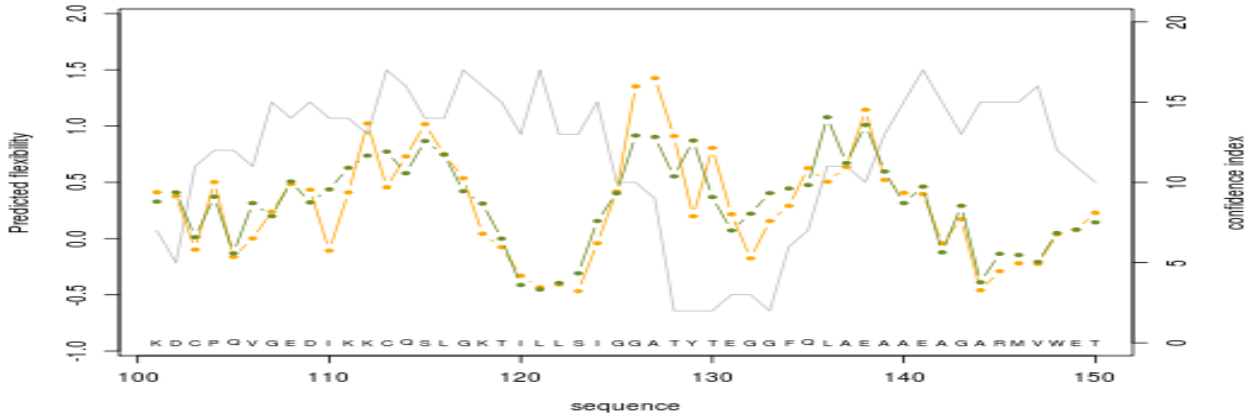
(I)


Predicted flexibility
confidence index
sequence
K D C P Q V G E D I K K C Q S L G K T I L L S I G G A T Y T E G G F Q I A E A A E A G A R M V W E T

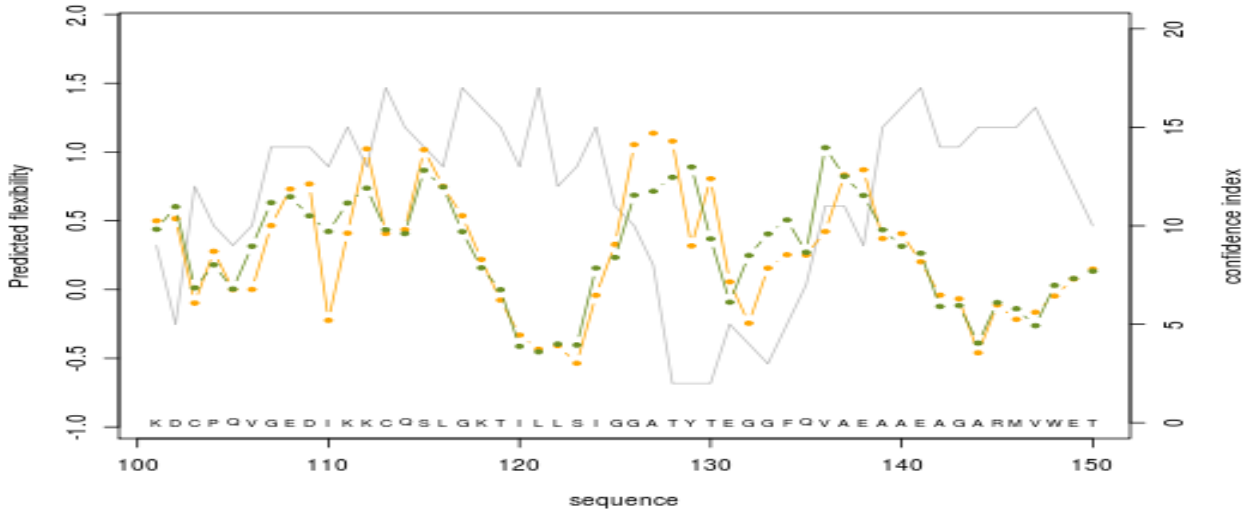
(J)


Predicted flexibility
confidence index
sequence
K D C P Q V G E D I K K C Q S L G K T I L L S I G G A T Y T E G G F Q K A E A A E A G A R M V W E T

(K)


Predicted flexibility
confidence index
sequence
K D C P Q V G E D I K K C Q S L G K T I L L S I G G A T Y T E G G F Q L A E A A E A G A R M V W E T

(L)


Predicted flexibility
confidence index
sequence
K D C P Q V G E D I K K C Q S L G K T I L L S I G G A T Y T E G G F Q V A E A A E A G A R M V W E T
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
0
5
10
15
20
100
110
120
130
140
150