

**FUNCTIONAL AND STRUCTURAL STUDIES OF GH20  
 $\beta$ -N-ACETYLGLUCOSAMINIDASE (GlcNAcase) FROM  
*VIBRIO HARVEYI* AND pK<sub>a</sub> CALCULATIONS  
OF GH18 CHITINASES**

**Piyanat Meekrathok**



**A Thesis Submitted in Partial Fulfillment of the Requirements for the  
Degree of Doctor of Philosophy in Biochemistry  
Suranaree University of Technology  
Academic Year 2015**

การศึกษาหน้าที่และโครงสร้างของเอนไซม์บีต้า-เอ็น-อะซิติลกลูโคซามินิเดส  
(กลูคแนคเอส) เฟมิลี 20 จากเชื้อ *Vibrio harveyi* และการคำนวณค่า  $pK_a$   
ของเอนไซม์ไคตินเนสเฟมิลี 18



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต  
สาขาวิชาชีวเคมี  
มหาวิทยาลัยเทคโนโลยีสุรนารี  
ปีการศึกษา 2558

**FUNCTIONAL AND STRUCTURAL STUDIES OF GH20**  
***β*-N-ACETYLGLUCOSAMINIDASE (GlcNAcase) FROM *VIBRIO***  
***HARVEYI* AND pK<sub>a</sub> CALCULATIONS OF GH18 CHITINASES**

Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

Thesis Examining Committee

\_\_\_\_\_  
(Assoc. Prof. Dr. Jatuporn Wittayakun)

Chairperson

\_\_\_\_\_  
(Assoc. Prof. Dr. Wipa Suginta)

Member (Thesis Advisor)

\_\_\_\_\_  
(Dr. David Apps)

Member

\_\_\_\_\_  
(Asst. Prof. Dr. Jeerus Sucharitakul)

Member

\_\_\_\_\_  
(Prof. Dr. James R. Ketudat-Cairns)

Member

\_\_\_\_\_  
(Prof. Dr. Sukit Limpijumnong)

Vice Rector for Academic Affairs  
and Innovation

\_\_\_\_\_  
(Prof. Dr. Santi Maensiri)

Dean of Institute of Science

ปิยฉัฐ หมี่กระโทก : การศึกษาหน้าที่และโครงสร้างของเอนไซม์บีต้า-เอ็น-อะซิติลกลูโคซามินิเดส (กลูคแนคเอส) แฟมิลี 20 จากเชื้อ *Vibrio harveyi* และการคำนวณค่า  $pK_a$  ของเอนไซม์ไคตินเนสแฟมิลี 18 (FUNCTIONAL AND STRUCTURAL STUDIES OF GH20  $\beta$ -N-ACETYLGLUCOSAMINIDASE (GlcNAcase) FROM *VIBRIO HARVEYI* AND  $pK_a$  CALCULATIONS OF GH18 CHITINASES) อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร. วิชา สุจินต์, 253 หน้า

เอนไซม์ไคตินเนสจัดอยู่ในกลุ่มของไกลโคไซดีไฮโดรเลสแฟมิลีที่ 18 และ 19 ซึ่งมีหน้าที่ในการย่อยสลายไคตินให้เป็นน้ำตาลโอลิโกแซคคาไรด์สายสั้นที่ถูกย่อยสลายต่อให้เป็นน้ำตาลกลูคานโมเลกุลเดี่ยวโดยเอนไซม์กลูคแนคเอสแฟมิลีที่ 3 20 และ 84 วิทยานิพนธ์ฉบับนี้ได้แบ่งการศึกษาออกเป็น 3 ส่วน โดยส่วนแรกทำการศึกษาเกี่ยวกับการตรวจสอบกรดอะมิโนที่ทำหน้าที่ในการเร่งปฏิกิริยาของเอนไซม์กลูคแนคเอสโดยวิธีการช่วยเหลือทางเคมี กรดอะมิโน 2 คู่ (Asp303-Asp304 และ Asp437-Glu438) ของเอนไซม์กลูคแนคเอสถูกทำให้กลายพันธุ์โดยเทคนิคการกลายพันธุ์แบบเฉพาะตำแหน่ง การแทนที่กรดอะมิโน Asp303 Asp304 Asp437 และ Glu438 ด้วยกรดอะมิโนอะลานีน แอสพาราจีน หรือกลูตามีนทำให้สูญเสียความสามารถในการเร่งปฏิกิริยาของเอนไซม์กลูคแนคเอส อย่างไรก็ตาม ความสามารถในการเร่งปฏิกิริยาที่ลดลงของเอนไซม์กลายพันธุ์ D437A สามารถถูกกู้คืนกลับมาได้อย่างมากตามปริมาณความเข้มข้นของนิวคลีโอไฟล์ (ฟอร์มेटไอออน) ที่ใส่เพิ่มขึ้น ในขณะที่ความสามารถในการเร่งปฏิกิริยาของเอนไซม์กลายพันธุ์ตัวอื่นยังคงเหมือนเดิม ดังนั้น การช่วยเหลือทางเคมีของเอนไซม์กลูคแนคเอสกลายพันธุ์ D437A โดยการใส่นิวคลีโอไฟล์ช่วยระบุว่า กรดอะมิโน D437 เป็นตัวเร่งแบบนิวคลีโอไฟล์หรือเบส และคู่กรดอะมิโน E438 เป็นตัวเร่งแบบให้โปรตอนหรือรับโปรตอน

ในงานวิจัยส่วนที่สอง หน้าที่และโครงสร้างของเอนไซม์กลูคแนคเอสได้ถูกศึกษาโดยการศึกษาทางด้านผลึกศาสตร์และจลนพลศาสตร์ของโปรตีน การทำโครมาโตกราฟีโดยแยกตามขนาดของโปรตีนและการทำอิเล็กโตรโฟรีซิสของอะคริลาไมด์เจลแบบตั้งต้นเสนอแนะว่าเอนไซม์กลูคแนคเอสถูกผสมเป็นแบบโมโนเมอร์โดยมีขนาดเท่ากับ 75 กิโลดาลตัน การศึกษาโครงสร้างแสดงให้เห็นว่า เอนไซม์กลูคแนคเอสประกอบไปด้วยสามโดเมนที่แตกต่างกันคือโดเมนที่จับคาร์โบไฮเดรตทางด้านเอ็นเทอมินอล โดเมนที่มีโครงสร้างเป็น  $\alpha+\beta$  และโดเมนเร่งปฏิกิริยาที่มีโครงสร้างแบบ TIM-barrel บริเวณช่องที่จับกับซับสเตรทของเอนไซม์กลูคแนคเอสมีโครงสร้างเหมือนกับอูโมงค์ขนาดเล็กซึ่งเหมาะกับการจับน้ำตาลโอลิโกแซคคาไรด์สายสั้น การซ้อนทับกันของโครงสร้างที่มีลิแกนด์กับไม่มีลิแกนด์เสนอแนะว่า การจับกับน้ำตาลชักนำ

ให้เกิดการเปลี่ยนแปลงโครงรูปบริเวณร่องที่จับกับน้ำตาล การวิเคราะห์ทางจลนพลศาสตร์ของเอนไซม์ กลายพันธุ์บริเวณเร่งปฏิกิริยาแสดงให้เห็นว่า คู่กรดอะมิโน D437-E438 มีความสำคัญอย่างมากต่อการเร่งปฏิกิริยาของเอนไซม์กลูคอกเนส

ในงานวิจัยส่วนที่สาม ค่า  $pK_a$  ของกรดอะมิโนในบริเวณเร่งปฏิกิริยาที่ตำแหน่ง DxDxE sequence motif (ใช้สัญลักษณ์แทนว่า  $D_1$   $D_2$  และ E) และ pH profile ของเอนไซม์โคติเนสแฟมิลี 18 ได้ถูกศึกษา การวิเคราะห์พบว่าแขนงข้างกรดอะมิโน  $D_1$  อยู่ในตำแหน่ง “ขึ้น” อย่างสมบูรณ์ ในขณะที่แขนงข้างของกรดกลูตามิก E อยู่ในตำแหน่ง “ลง” เป็นส่วนใหญ่ทั้งในรูปของเอนไซม์อิสระและที่จับกับลิแกนด์ ค่า  $pK_a$  ที่ได้จากการคำนวณของกรดอะมิโนสามตัวถูกพบว่า ค่า  $pK_a$  ของ  $D_1$  มีค่าน้อยกว่า 2 ค่า  $pK_a$  ของ  $D_2$  อยู่ระหว่าง 8.0 ถึง 13.0 และค่า  $pK_a$  ของ E ถูกพบอยู่ในช่วงตั้งแต่ 6.0 ถึง 9.0 ค่า  $pK_a$  ที่คำนวณได้ระบุว่า คู่กรดอะมิโนแอสพาเทต  $D_1$  และ  $D_2$  ในเอนไซม์โคติเนสใช้ประจุลบเพียงหนึ่งตัวร่วมกันในทุกช่วง pH ในขณะที่กรดกลูตามิก E ถูกทำให้มีโปรตอนที่ค่า pH ที่เอนไซม์อยู่ในรูปที่ทำงานได้ โดยสรุป กรดอะมิโนสามตัวนี้ถูกเสนอแนะเพื่อแสดงบทบาทร่วมกันในการเร่งปฏิกิริยาของเอนไซม์โคติเนสแฟมิลี 18 และกรดอะมิโน  $D_1$   $D_2$  และ E ต้องอยู่ในรูปที่มีประจุ ไม่มีประจุ และไม่มีประจุตามลำดับสำหรับโคติเนสที่อยู่ในรูปที่ทำงานได้



PIYANAT MEEKRATHOK : FUNCTIONAL AND STRUCTURAL  
STUDIES OF GH20  $\beta$ -*N*-ACETYLGLUCOSAMINIDASE (GlcNAcase)  
FROM *VIBRIO HARYEYI* AND  $pK_a$  CALCULATIONS OF GH18  
CHITINASES. THESIS ADVISOR : ASSOC. PROF. WIPA SUGINTA,  
Ph.D. 253 PP.

*$\beta$ -N-ACETYLGLUCOSAMINIDASE, CHITINASES, VIBRIO HARVEYI, PROTEIN  
CRYSTALLOGRAPHY, ENZYME KINETICS,  $pK_a$  CALCULATIONS*

Chitinases are a member of family 18 and 19 glycoside hydrolases that are responsible for the successive degradation of insoluble chitin to yield soluble chitooligosaccharides, which are then further hydrolysed to GlcNAc monomers by family 3, 20 and 84 GlcNAcases. This thesis is divided into three parts. The first part is involved with probing the catalytic residues of *Vh*GlcNAcase by a chemical rescue approach. Two invariant acidic pairs (Asp303-Asp304 and Asp437-Glu438) of *Vh*GlcNAcase were mutated using a site-directed mutagenesis strategy. Substitution of Asp303, Asp304, Asp437 or Glu438 with Ala/Asn/Gln produced a dramatic loss of the GlcNAcase activity. However, the activity of the inactive D437A mutant was largely recovered in the presence of an exogenous nucleophile (formate ion) in a concentration-dependent manner, while the activity of other mutants was restored only slightly. In conclusions, chemical rescue of the *Vh*GlcNAcase D437A inactive mutant by an added nucleophile helped to identify Asp437 as the catalytic nucleophile/base, and hence its acidic partner Glu438 as the catalytic proton donor/acceptor.

In the second part, the function and structure of *Vh*GlcNAcase were investigated

by protein crystallography and enzyme kinetics. Size-exclusion chromatography and native-PAGE suggested that the recombinant *VhGlcNAcase* is a monomeric enzyme with molecular weight of 75 kDa. Structural investigation revealed that *VhGlcNAcase* comprises three distinct domains, designated as the *N*-terminal carbohydrate-binding domain, the  $\alpha$ + $\beta$  topology domain and the TIM-barrel catalytic domain. The substrate binding pocket of *VhGlcNAcase* has a small tunnel-like structure, which is suitable to accommodate a short-chain chitooligosaccharide. Superimposition of the crystal structures of ligand-free and ligand-bound *VhGlcNAcase* suggests that binding of the GlcNAc induces local conformational changes around the sugar binding pocket. Kinetic analysis of the active-site mutants revealed that the adjacent D437-E438 pair is significantly important for the enzymic activity of *VhGlcNAcase*.

In the third part, the  $pK_a$  of the active site residues in the DxDxE sequence motif (referred as 'D<sub>1</sub>, D<sub>2</sub> and E') and its pH profiles of family 18 chitinases were investigated. The analysis has found that the side chain of D<sub>1</sub> is mainly in the 'up' position, whereas the side chain of E is mainly in the 'down' position in the apo and holo forms. The  $pK_a$  values calculated for the three residues are as follows:  $pK_a$  (D<sub>1</sub>) < 2.0,  $pK_a$  (D<sub>2</sub>) in the range between 8.0-13.0 and  $pK_a$  (E) in the range from 6.0 to 9.0. The calculated  $pK_a$  values indicate that the D<sub>1</sub>-D<sub>2</sub> pair holds exactly one negative charge over the whole accessible pH range, whereas the catalytic acid E is protonated over the catalytically competent pH range. In summary, these three acidic groups D<sub>1</sub>, D<sub>2</sub> and E are proposed to play a concerted role in the catalysis of family 18 chitinases and must be charged, neutral and neutral, respectively for the chitinases to be active.

School of Biochemistry

Student's signature\_\_\_\_\_

Academic Year 2015

Advisor's signature\_\_\_\_\_

## ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my thesis advisor, Assoc. Prof. Dr. Wipa Suginta for providing me a great opportunity to work on this project and her valuable guidance that has made this thesis successful. I am also deeply grateful to my co-advisors, Dr. Ingrid R. Vetter at the Max Planck Institute (MPI) of Molecular Physiology, Dortmund in Germany, who gave me a useful guidance and valuable advice on the protein crystallography, and more importantly, provided me the access to the X-ray crystallization facilities at MPI Dortmund and Dr. Jens Erik Nielsen at University College Dublin in Ireland, who gave me a valuable technical skills and his precious advice on the protein  $pK_a$  calculation, and also his kind support during my visit in Ireland. Additionally, I would also like to thank Dr. Heino Prinz, at MPI Dortmund, who gave me the valuable suggestions on the human protein expression in bacterial system, and more importantly, provided me the access to the Dortmund Protein Facility (DPF) at MPI Dortmund.

I would also like to thank my special colleagues, Marco Bürger and Arthur T. Porfetye in the department of Mechanistic Cell Biology at MPI Dortmund for their help in the in-house X-ray data collection and initial structural analysis and also thank the beamline staffs at the SLS synchrotron at the Paul Scherrer Institute, Villigen, Switzerland, Max Planck Institute for Medical Research, Heidelberg in Germany, and Falk Syberg at Ruhr-Universität Bochum in Germany for help with data collection.

I would also like to thank my special colleagues, Predrag Kukic and Damien Farrell at the UCD Conway Institute, University College Dublin in Ireland for their computer setup, teaching in linux language and a valuable technical skills on the protein  $pK_a$  calculation and all my friends for providing a nice collaborative working environment.

All lecturers of biochemistry courses at Suranaree University of Technology that I have taken are acknowledged, particularly to Prof. Dr. James R. Ketudat-Cairns, who taught me in the bioinformatics and protein crystallography courses, Asst. Prof. Dr. Rodjana Opasiri and Assoc. Prof. Dr. Wipa Suginta, who introduced me to the enzymology class.

I would like to acknowledge the Biochemistry-Electrochemistry Research Unit Laboratory and the Center for Scientific and Technological Equipment, Suranaree University of Technology for providing all research facilities.

Most importantly, I would like to thank for the financial support by SUT outstanding academic performance scholarship, the Thailand Research Fund (TRF) through the Royal Golden Jubilee Ph.D. Program and Suranaree University of Technology as well as the Deutscher Akademischer Austausch Dienst e.V. (DAAD, German Academic Exchange Service) during my one-year research in Germany.

Finally, my special thanks are given to my beloved family for their support, understanding, infinite love and care throughout my life. Completion of this thesis would not been possible without them.

Piyanat Meekrathok

# CONTENTS

	<b>Page</b>
ABSTRACT IN THAI.....	I
ABSTRACT IN ENGLISH .....	III
ACKNOWLEDGEMENTS.....	V
CONTENTS.....	VII
LIST OF TABLES .....	XIII
LIST OF FIGURES .....	XV
LIST OF ABBREVIATIONS.....	XX
<b>CHAPTER</b>	
<b>I INTRODUCTION.....</b>	<b>1</b>
1.1 Chitin and applications .....	1
1.2 Roles of chitin-degrading enzymes.....	3
1.3 Classification of chitinases and GlcNAcases.....	5
1.4 Catalytic mechanism of chitinases and GlcNAcases .....	11
1.5 Studies of the chemical rescue in retaining enzymes.....	18
1.6 Structural analysis of family 18 chitinases .....	21
1.6.1 Structural analysis of chitinase A and B from <i>S. marcescens</i> .....	21
1.6.1 Structural analysis of chitinase A from <i>V. harveyi</i> .....	23
1.7 Studies of p <i>K</i> <sub>a</sub> calculations .....	25
1.7.1 Theory of protein p <i>K</i> <sub>a</sub> calculation.....	25
1.7.2 Running a p <i>K</i> <sub>a</sub> calculation.....	32

## CONTENTS (Continued)

	<b>Page</b>
1.7.3 Extracting $pK_a$ values from calculated titration curves.....	32
1.7.4 Calculating pH activity profiles .....	33
1.7.5 Accuracy of protein $pK_a$ calculation methods.....	34
1.7 Studies of <i>Vibrio harveyi</i> GlcNAcases .....	35
1.8 Research objectives.....	36
<b>II MATERIALS AND METHODS</b> .....	<b>39</b>
2.1 Chemicals and reagents.....	39
2.1.1 Bacterial strains and expression plasmids.....	39
2.1.2 Site-directed mutagenesis and plasmid purification .....	39
2.1.3 Protein expression and purification .....	40
2.1.4 Enzyme kinetics .....	41
2.1.5 Protein crystallization .....	41
2.2 Analytical programs.....	42
2.3 Instrumentation .....	43
2.4 Sequence analysis and homology modeling of <i>Vh</i> GlcNAcases .....	44
2.5 Mutational design and site-directed mutagenesis of <i>Vh</i> GlcNAcases .....	45
2.6 Expression and purification of <i>Vh</i> GlcNAcase and its mutants.....	48
2.7 Investigation of protein state by size-exclusion chromatography.....	51
2.8 GlcNAcase activity assay .....	52
2.9 Determination of the pH optima of <i>Vh</i> GlcNACase WT and D437A mutant.....	52

## CONTENTS (Continued)

	<b>Page</b>
2.10 Chemical rescue assay of <i>VhGlcNAcase</i> and its mutants.....	53
2.11 Time-course and kinetics of sodium formate effects on the activity of <i>VhGlcNAcase</i> D437A mutant.....	53
2.12 Steady-state kinetics of <i>VhGlcNAcase</i> and its mutants.....	54
2.13 Determination of three-dimensional structures of <i>VhGlcNAcases</i> .....	55
2.13.1 Crystallization of <i>VhGlcNAcase</i> WT and the D437A mutant.....	55
2.13.2 Data collection and processing.....	58
2.13.3 Structural determination and refinement.....	59
2.14 $pK_a$ calculation of GH18 chitinases.....	61
2.14.1 Selection and preparation of X-ray structures.....	61
2.14.2 $pK_a$ calculations.....	62
2.14.3 Prediction of pH profiles.....	64
<b>III RESULTS</b> .....	66
<b>PART I</b> .....	66
3.1 Sequence analysis and homology modelling.....	66
3.2 Mutational design and site-directed mutagenesis.....	69
3.3 Expression and purification of recombinant <i>VhGlcNAcase</i> and its mutants.....	72
3.4 Effects of the active-site mutation on the specific activity of <i>VhGlcNAcase</i> .....	75
3.5 Effects of sodium formate concentration on the rescued activity of the D37A mutant.....	78

## CONTENTS (Continued)

	<b>Page</b>
3.6 Steady state kinetics of activation by sodium formate.....	80
<b>PART II.....</b>	<b>85</b>
3.7 Functional characterization of <i>VhGlcNAcases</i> .....	85
3.7.1 Expression and purification .....	85
3.7.2 Molecular weight determination of <i>VhGlcNAcase</i> .....	87
3.8 Crystallization of the wild-type <i>VhGlcNAcase</i> and mutant D437A .....	89
3.8.1 Initial screening and optimization.....	89
3.9 Data collection and processing of the wild-type <i>VhGlcNAcase</i> and mutant D437A.....	99
3.10 Structural determination of the wild-type <i>VhGlcNAcase</i> and mutant D437A.....	101
3.10.1 Phase determination by molecular replacement .....	101
3.10.2 Refinements and structural determination .....	102
3.10.3 The overall structure of <i>VhGlcNAcase</i> .....	105
3.10.4 Dimer interface of GH20 <i>VhGlcNAcase</i> .....	109
3.10.5 Sugar binding induces conformational changes .....	112
3.10.6 Structural comparison of <i>VhGlcNAcase</i> with other bacterial GH20 members .....	114
3.10.7 Specific interactions of <i>VhGlcNAcase</i> with sugar.....	119
3.11 Specific hydrolyzing activity of <i>VhGlcNAcase</i> and its mutants.....	123
3.12 Kinetic parameters of <i>VhGlcNAcase</i> and its mutants.....	124
3.13 Substrate specificity analysis .....	127

## CONTENTS (Continued)

	<b>Page</b>
<b>PART III</b> .....	130
3.14 Sequence analysis and the catalytic residue of chitinases .....	130
3.15 p <i>K</i> <sub>a</sub> calculations of the active-site residues.....	133
3.16 Predicting the pH-activity profiles.....	141
3.16.1 Computational pH-profiles of the apo structures.....	144
3.16.2 Computational pH-profiles for holo structures.....	150
3.17 p <i>K</i> <sub>a</sub> comparison between bacterial and human chitinases .....	154
<b>IV DISCUSSION</b> .....	158
4.1 Chitin degradation pathway of <i>Vibrio harveyi</i> .....	158
4.2 Probing the catalytic residues of <i>VhGlcNAcases</i> by chemical rescue.....	160
4.3 Expression, purification and molecular weight determination of <i>VhGlcNAcases</i> .....	165
4.4 Crystallization and structure determination of <i>VhGlcNAcases</i> .....	166
4.5 Comparison of <i>VhGlcNAcases</i> with other chitinases and <i>GlcNAcases</i> .....	169
4.6 Kinetics of wild-type <i>VhGlcNAcase</i> and its mutants .....	172
4.7 Common determinants of p <i>K</i> <sub>a</sub> calculations of GH18 chitinases.....	177
4.8 Computational pH activity profiles of GH18 chitinases.....	182
4.9 Comparison of p <i>K</i> <sub>a</sub> values between bacterial and human chitinases .....	185
<b>V CONCLUSION</b> .....	189
<b>REFERENCES</b> .....	192
<b>APPENDICES</b> .....	221

**CONTENTS (Continued)**

	<b>Page</b>
APPENDIX A STANDARD CURVES .....	222
APPENDIX B COMMON $pK_a$ VALUE .....	223
APPENDIX C PUBLICATIONS .....	224
APPENDIX D LIST OF PRESENTATION.....	250
CURRICULUM VITAE.....	253

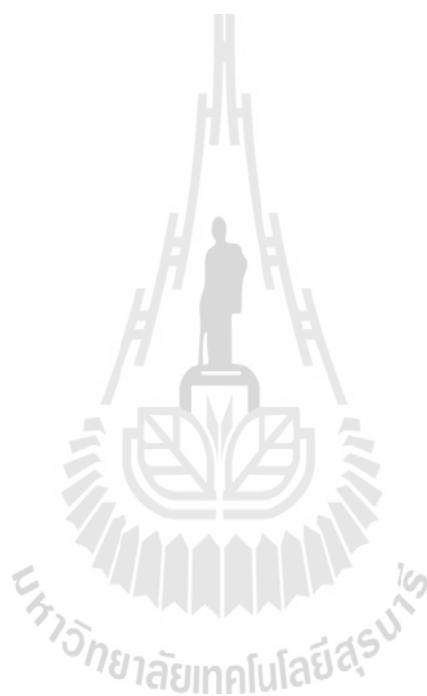


## LIST OF TABLES

Table	Page
1.1 Possible protonation states for a hypothetical protein consisting of three titratable groups. ....	29
2.1 Primers used for site-directed mutagenesis.....	46
2.2 The PCR reaction used for site-directed mutagenesis .....	48
2.3 The PCR running conditions used for site-directed mutagenesis .....	48
3.4 Specific activity of wild-type <i>VhGlcNAcase</i> and its mutants against <i>pNP</i> - GlcNAc in the absence and presence of 1 M sodium salts .....	77
3.5 Kinetic parameters for the hydrolytic activity of <i>VhGlcNAcase</i> mutant D437A in the presence of increasing sodium formate (HCOONa) concentration .....	83
3.6 Data collection and processing statistics.....	100
3.7 Refinement statistics .....	104
3.8 Summary of the interactions of GLCNACs with the binding residues in the substrate binding pocket of GH20 <i>VhGlcNAcase</i> , <i>SpHex</i> and <i>SmCHB</i> .....	122
3.9 Kinetic parameters of wild-type and mutants of <i>VhGlcNAcase</i> .....	126
3.10 Conformation and the $pK_a$ values calculated from the apo and apo- generated structures of three acidic residues in the DxDxE motif of GH18 chitinases .....	137
3.11 Conformation and the $pK_a$ values calculated from the holo structures of three acidic residues in the DxDxE motif of family 18 chitinases and comparison with its apo and apo-generated structures .....	140

**LIST OF TABLES (Continued)**

<b>Table</b>	<b>Page</b>
S1 Model $pK_a$ values used in this study.....	223



## LIST OF FIGURES

Figure	Page
1.1 Chemical structure of chitin.....	2
1.2 Schematic drawing of the predominant cleaved patterns of chitinolytic enzymes.....	6
1.3 Ribbon representation of the main folds of the catalytic domain of the family 18 (PDB code: 3B9A) and family 19 (PDB code: 3CQL) chitinases .....	8
1.4 The domain organization of bacterial chitinases .....	9
1.5 The catalytic mechanism of glycoside hydrolases.....	13
1.6 The proposed catalytic mechanism of family 18 chitinase B from <i>Serratia marcescens</i> ( <i>SmChiB</i> ).....	15
1.7 The refined catalytic cycle of chitin degradation by <i>Vibrio harveyi</i> chitinase A ( <i>VhChiA</i> ).....	16
1.8 The catalytic mechanisms of the two classes of GlcNAcases .....	18
1.9 Proposed mechanism of azide-mediated rescue with the <i>SpHex</i> D313A mutant .....	20
1.10 The crystal structure of <i>S. marcescens</i> chitinase A mutant E315L in complex with GlcNAc <sub>6</sub> (PDB code: 1NH6) .....	22
1.11 The crystal structure of <i>S. marcescens</i> chitinase B mutant E144Q in complex with GlcNAc <sub>5</sub> (PDB code: 1E6N).....	23
1.12 The structure of inactive mutant E315M complexed with GlcNAc <sub>6</sub> of <i>V. harveyi</i> chitinase A (PDB code: 3B9A) .....	25

## LIST OF FIGURES (Continued)

Figure	Page
1.13 The titration curve of an acid with a $pK_a$ value of 4.0 calculated using equation 1.2.....	26
1.14 pH dependence of population of protonation states for three-group system and its titration curves.....	28
1.15 The thermodynamic cycle for the transfer of a titratable group from solvent to a protein environment.....	31
2.1 Optimization of crystallization conditions of wild-type <i>VhGlcNAcase</i> .....	56
2.2 Optimization of crystallization condition of the mutant D437A of <i>VhGlcNAcase</i> .....	58
3.1 Comparison of the modelled structure of <i>VhGlcNAcase</i> with other <i>GlcNAcases</i> .....	68
3.2 Point mutations of active site residues.....	70
3.3 Identification of the amplified PCR product ~5,300 bp of the <i>VhGlcNAcase</i> and pQE-60 vector .....	71
3.4 Size confirmation of mutagenic plasmids of the <i>VhGlcNAcase</i> /pQE-60 vector by <i>BamHI</i> restriction enzyme digestion.....	72
3.5 SDS-PAGE of the expression and purification of <i>VhGlcNAcases</i> .....	73
3.6 The purity of <i>VhGlcNAcase</i> protein in 12% SDS-PAGE by increasing protein concentration from 1-20 $\mu\text{g}$ .....	74
3.7 SDS-PAGE analysis of <i>VhGlcNAcase</i> and its mutants, after IPTG-induced expression overnight .....	74
3.8 The chemical structures of azide and formate ions used in this study.....	75

## LIST OF FIGURES (Continued)

Figure	Page
3.9 Specific hydrolytic activity of wild-type <i>VhGlcNAcase</i> against <i>pNP-GlcNAc</i> in the presence of various concentrations of sodium salts in 100 mM sodium phosphate buffer, pH 7.0 .....	76
3.10 Activity/pH profiles of <i>VhGlcNAcase</i> and its mutant D437A.....	78
3.11 Effect of sodium formate on <i>pNP-GlcNAc</i> hydrolysis by <i>VhGlcNAcase</i> and its mutants .....	79
3.12 Kinetic properties of activation by sodium formate .....	81
3.13 Chemical rescue of the D437A mutant by sodium formate.....	84
3.14 Purification of <i>VhGlcNAcase</i> expressed from <i>E. coli</i> M15 cells using a HiPrep 16/60 Sephacryl S-200 prepacked gel-filtration column.....	86
3.15 A Chromatographic profile and calibration curve of <i>VhGlcNAcase</i> and standard proteins by size-exclusion chromatography .....	88
3.16 Crystals of ligand-free <i>VhGlcNAcase</i> from initial screenings using sitting drop technique.....	89
3.17 Crystals of the inactive mutant D437A of <i>VhGlcNAcase</i> from initial screenings using sitting drop technique .....	91
3.18 Grid screen with variation of PEG 3350, Bis-Tris and sodium/potassium phosphate .....	93
3.19 Grid screen with variation of sodium malonate and sodium acetate .....	94
3.20 Photographs of wild-type <i>VhGlcNAcase</i> crystals obtained from the optimization using a variation of sodium malonate and sodium acetate concentrations .....	95

## LIST OF FIGURES (Continued)

Figure	Page
3.21 Grid screen with variation of PEG 3350, Bis-Tris and sodium acetate .....	97
3.22 Refined crystals after optimization .....	98
3.23 X-ray diffraction images .....	101
3.24 Ramachandran plot for all non-Pro/Gly residues.....	105
3.25 Domain organization and overall structure of <i>VhGlcNAcase</i> .....	108
3.26 The dimer interface of GH20 <i>VhGlcNAcase</i> .....	111
3.27 Conformational changes of <i>VhGlcNAcase</i> structures in the absence and presence of GLCNAC .....	113
3.28 Overview of structural comparison of <i>VhGlcNAcase</i> with other bacterial GH20 structures .....	116
3.29 Superimposition of bacterial GH20 GlcNAcases .....	119
3.30 Specific interactions within the substrate binding pocket of <i>VhGlcNAcase</i> ...	121
3.31 Specific hydrolyzing activity of <i>VhGlcNAcase</i> and its mutants.....	124
3.32 Comparison of substrate specificity based on the 3D structures .....	128
3.33 Overview of family 18 chitinases .....	132
3.34 Contribution of $pK_a$ values of acidic active-site residues, D <sub>1</sub> , D <sub>2</sub> and E plotted in three dimensions .....	134
3.35 The quality of the titration curves from apo structures (A-F) used in determining the $pK_a$ values.....	142
3.36 The computational pH-activity profiles of family 18 chitinases using the ligand-free structures .....	143

## LIST OF FIGURES (Continued)

Figure	Page
3.37 The quality of the titration curves from the holo structures (A-B) used in determining the $pK_a$ values.....	152
3.38 The computational pH-activity profiles of family 18 chitinases using the complex structures with the natural substrates .....	153
3.39 A comparison of bacterial and human chitinases .....	157
4.1 Model of the chitin degradation cascade of the marine bacterium <i>Vibrio harveyi</i> .....	159
4.2 Proposed mechanism of nucleophile-mediated chemical rescue with the inactive mutant of GH20 GlcNAcases.....	164
4.3 Proposed catalytic mechanism of GH20 <i>Vh</i> GlcNAcase .....	177
4.4 Summary of the effect of placing a titratable group in a negative, positive and hydrophobic environment.....	187
4.5 The refined catalytic mechanism of family 18 chitinases during chitin degradation supported by the calculated active-site $pK_a$ values.....	188

## LIST OF ABBREVIATIONS

A	Absorbance
Å	Angstrom
BSA	Bovine serum albumin
Bp	Base pair(s)
CCPS	Catalytically competent protonation state
cDNA	Complementary deoxyribonucleic acid
Chi	Chitinase
DNA	Deoxyribonucleic acid
DNPGlcNAc	Dinitrophenyl- <i>N</i> -acetyl-glucosaminide
dNTP	Deoxynucleotide triphosphate
EDTA	Ethylenediamine tetraacetic acid
GH	Glycoside hydrolase
GH18	Glycoside hydrolase family 18
GH20	Glycoside hydrolase family 20
GlcNAc	<i>N</i> -acetyl-glucosamine
GlcNAc <sub>2</sub>	Di- <i>N</i> -acetyl-chitobiose
GlcNAc <sub>3</sub>	Tri- <i>N</i> -acetylchitotriose
GlcNAc <sub>4</sub>	Tetra- <i>N</i> -acetyl-chitotetraose
GlcNAc <sub>5</sub>	Penta- <i>N</i> -acetyl-chitopentaose
GlcNAc <sub>6</sub>	Hexa- <i>N</i> -acetylchitohexaose

## LIST OF ABBREVIATIONS (Continued)

GlcNAcase	$\beta$ - <i>N</i> -acetylglucosaminidase
h	Hour
HDL	High-density lipoprotein
Hex	$\beta$ -Hexosaminidase
HPLC	High performance liquid chromatography
IPTG	Isopropyl thio- $\beta$ - <i>D</i> -galactoside
LB	Luria-Bertani lysogeny broth
LDL	Low-density lipoprotein
nm	Nanometer(s)
NGA2B	Bisected glycan <i>N</i> -acetyl- <i>D</i> -glucosamine $\beta$ -1,2- <i>D</i> -mannose $\alpha$ -1,3 ( <i>N</i> -acetyl- <i>D</i> -glucosamine $\beta$ -1,2- <i>D</i> -mannose $\alpha$ -1,6( <i>N</i> -acetyl- <i>D</i> -glucosamine $\beta$ -1,4))- <i>D</i> -mannose $\beta$ -1,4- <i>N</i> -acetyl- <i>D</i> -glucosamine
Ni-NTA	Ni-nitrilotriacetic acid
min	Minute
Mr	Relative molecular mass
MW	Molecular weight
°C	Degree Celcius
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEG	Polyethylene glycol

**LIST OF ABBREVIATIONS (Continued)**

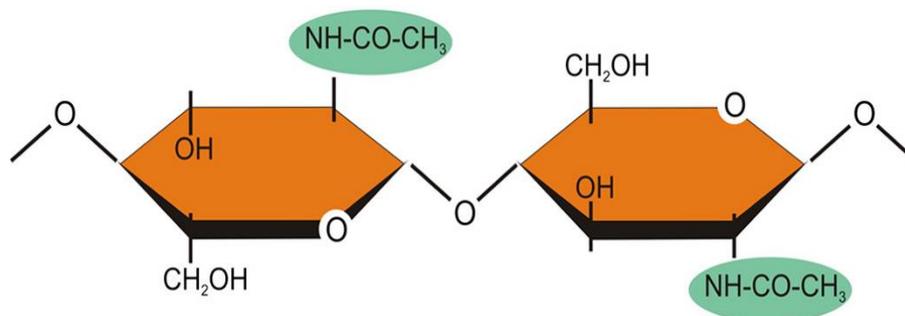
pH	Negative logarithm of hydrogen ion activity
PMSF	Phenylmethylsulfonylfluoride
<i>p</i> NP	<i>para</i> -nitrophenol
<i>p</i> NP-GlcNAc	<i>para</i> -nitrophenyl- <i>N</i> -acetyl-glucosaminide
<i>p</i> NP-GlcNAc <sub>2</sub>	<i>para</i> -nitrophenyl- di- <i>N</i> -acetyl-chitobioside
RMSD	root mean-square deviation
Rpm	Round(s) per minute
s	Second
SDS-PAGE	Polyacrylamide gel electrophoresis
TB	Terrific broth
TEMED	Tetramethylenediamine
TIM	Triose-phosphate isomerase
Tris	Tris-(hydroxymethyl)-aminoethane
TLC	Thin layer chromatography
v/v	Volume/volume
WIpK <sub>a</sub>	WHAT IF pK <sub>a</sub> calculation
WT	Wild-type
w/v	Weight/volume

# CHAPTER I

## INTRODUCTION

### 1.1 Chitin and applications

Chitin is a homopolysaccharide of *N*-acetyl-D-glucosamine (GlcNAc) units linked together with  $\beta$ -(1,4) glycosidic linkages (Figure 1.1). Chitin is the second most abundant polysaccharide in nature, next to cellulose and it serves as a major component in fungal cell wall, exoskeleton of crustaceans and insects, mollusks, protozoa and microfilaria sheaths of parasitic nematodes (Bussink *et al.*, 2007; Lee, 2009; Rinaudo, 2006). Chitin is hard, inelastic and its chains can run either parallel ( $\beta$ -chitin) or anti-parallel ( $\alpha$ -chitin) (Blackwell, 1969; Minke and Blackwell, 1978). This natural polymer and its derivatives are suitable to be used as functional materials because of their biocompatibility, biodegradability, non-toxicity, and well-adsorption properties, etc (Ravi Kumar, 2000). However, biological degradation of chitin is difficult, since its chains contain intra and intermolecular hydrogen bonds, which cause chitin to be a highly insoluble material, which limits its use (Blackwell, 1969; Ravi Kumar, 2000).



**Figure 1.1** Chemical structure of chitin (Seidl, 2008).

Several strategies have been developed to convert chitin into small soluble derivatives by chemical or enzymatic methods (Ilankovan *et al.*, 2006). For a chemical method, a strong acid such as HCl is employed for chitin hydrolysis but several problems are found in GlcNAc<sub>n</sub> production because of its limitation by the acid, including low yields, high cost and acidic waste created by a strong acid. In contrast, the enzymatic method can catalyse the chitin degradation completely with less time consumed under mild conditions, has lower cost and is non-toxic to the environment (Chen *et al.*, 2010; Sashiwa *et al.*, 2003).

Chitin and its derivatives can be used in various fields, including biomedicine, nutrition, material sciences, biotechnology, agriculture and food industries (Harish Prashanth and Tharanathan, 2007; Kato *et al.*, 2003; Ravi Kumar, 2000). In the biomedical field, chitin exhibits a wide variety of applications. For example, chitin sutures resist attack in bile, urine and pancreatic juice (Nakajima *et al.*, 1986), whereas deacetylated chitooligomers can be used for fatty acid absorption, decreasing low-density lipoprotein (LDL) and increasing high-density lipoprotein (HDL) cholesterol levels (Koide, 1998). Chitin derivatives also possess anti-tumor and anti-bacterial properties, and can be used for drug delivery carriers, wound dressing, stimulating connective

tissue formation, activating the immune response, decreasing sugar in the blood, and enhancing absorption of calcium (Kim and Rajapakse, 2005; Koide, 1998; Struszczyk, 2006). In food industries, deacetylated forms of chitin are used as food, drink supplements, animal nutritional supplements, and as a measure of mold contamination of agricultural commodities and food products (Cousin, 1996; Qin *et al.*, 2006; Ravi Kumar, 2000). In agriculture, chitin is used to increase crop yields and protect harvested products by eliminating pests (Chang *et al.*, 2007). In biotechnology, chitin films, sponges and fibres are used to enhance wound healing in tree-bark tissues (Hirano, 1999).

## 1.2 Roles of chitin-degrading enzymes

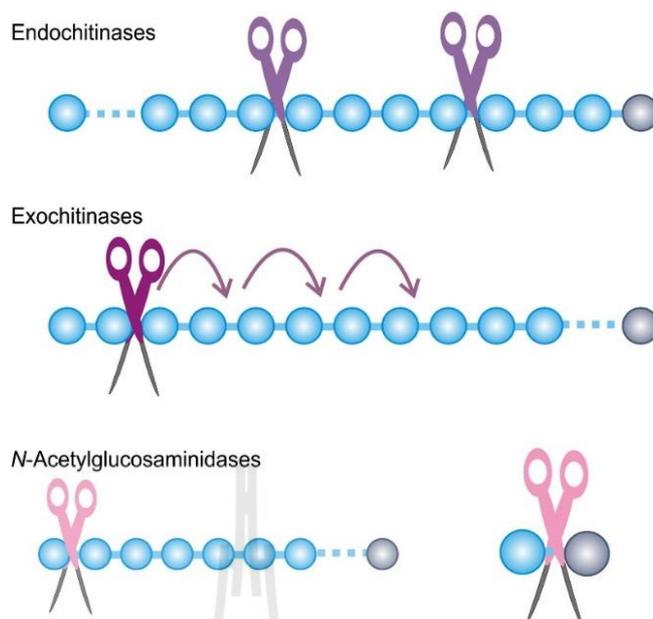
Chitin is degraded by chitinolytic enzymes, which are divided into chitinases (EC 3.2.1.4) and  $\beta$ -*N*-acetylglucosaminidases (GlcNAcases) (3.2.1.52). Chitinases hydrolyse the  $\beta$ -(1,4) glycosidic linkages between the adjacent *N*-acetylglucosamine residues of insoluble chitin to yield soluble chitooligosaccharides and then further hydrolysed to the *N*-acetylglucosamines by GlcNAcases. Bacterial chitinases and GlcNAcases utilize chitin as important carbon and nitrogen sources for the cells and their chitinolytic enzymes are involved in the natural recycling of chitin biomass (Armand *et al.*, 1994; Bassler *et al.*, 1991; Keyhani and Roseman, 1999; Park *et al.*, 2000; Yu *et al.*, 1991). Fungal chitinases and GlcNAcases are involved in fungal development and morphogenesis, including cell wall digestion, germination and hyphal growth (Kim *et al.*, 2002; Kuranda and Robbins, 1991; Sahai and Manocha, 1993). Most chitinases and GlcNAcases from insects play a vital role in the degradation of old cuticle chitin during insect metamorphosis, especially in the moulting process

(Hogenkamp *et al.*, 2008; Liu *et al.*, 2011; Merzendorfer and Zimoch, 2003). Plants have inducible defence mechanisms using chitinases and GlcNAcases which act as the biological control agents by degradation of the fungal cell walls, resulting in death or inhibition of fungal growth and invading insects (Bolar *et al.*, 2001; Dowd *et al.*, 2007; Herrera-Estrella and Chet, 1999; Melchers and Stuiver, 2000). Plant GlcNAcases are also present at high levels during the ripening of many fruits, suggesting a role in *N*-glycan degradation or processing (Hossain *et al.*, 2014; Jagadeesh *et al.*, 2004a, 2004b). In animals, chitinases and GlcNAcases are involved in the digestive system (Rinaudo, 2006). In human, chitinases are found to be highly expressed in macrophages that are involved in inflammatory conditions (Kanneganti *et al.*, 2013; Kzhyshkowska *et al.*, 2007; Rosa *et al.*, 2012). Both chitinases and GlcNAcases from human have been also detected at high levels in patients infected with *Plasmodium falciparum* that may reflect an immunological response to malarial infection induced by the human enzymes (Patil *et al.*, 2000). Specifically, human acidic mammalian chitinase (AMCase) is associated with T helper-2 inflammation such as asthma and allergic diseases (Zhu *et al.*, 2004; Donnelly and Barnes, 2004; Elias *et al.*, 2005; Kawada *et al.*, 2007) and chitotriosidase has been described in an increase in symptomatic Gauchers patients (Hollak *et al.*, 1994). For human GlcNAcases, deficiency of HexA or HexB leads to the accumulation of gangliosides, causing Tay-Sachs and Sandhoff diseases, respectively (Gravel *et al.*, 1995; Myerowitz, 1997). The chitin-degrading enzymes are of great interest as the potential target for the development of specific pesticides, antifungal agents and human drug design and seem to be potential biocatalysts for chitin conversion.

### 1.3 Classification of chitinases and GlcNAcases

Chitinolytic enzymes are divided into chitinases and *N*-acetyl-glucosaminidases (GlcNAcases). Chitinases are a diverse group of enzymes that catalyse the conversion of chitin to soluble chitooligosaccharides. As shown in Figure 1.2, chitinases are further classified into endo- and exochitinases, depending upon their cleavage patterns (Cohen-Kupiec and Chet, 1998; Seidl, 2008). Endochitinases degrade chitin from any point along the polymer chain, yielding soluble low-molecular mass multimers of chitooligo fragments such as (GlcNAc)<sub>2</sub>, (GlcNAc)<sub>3</sub>, and (GlcNAc)<sub>4</sub> (Sahai and Manocha, 1993). On the other hand, exochitinases cleave the glycosidic bonds from the non-reducing end and release (GlcNAc)<sub>2</sub> as the major product. GlcNAcases further catalyse the release of terminal non-reducing GlcNAc residues from chitin with highest affinity for GlcNAc<sub>2</sub> as the major product of chitin hydrolysis from chitinases and then convert it into two GlcNAcs as shown in Figure 1.2 (Horsch *et al.*, 1997). Finally, GlcNAc is then taken up by the bacterial cells and used as carbon and nitrogen sources (Yu *et al.*, 1991).



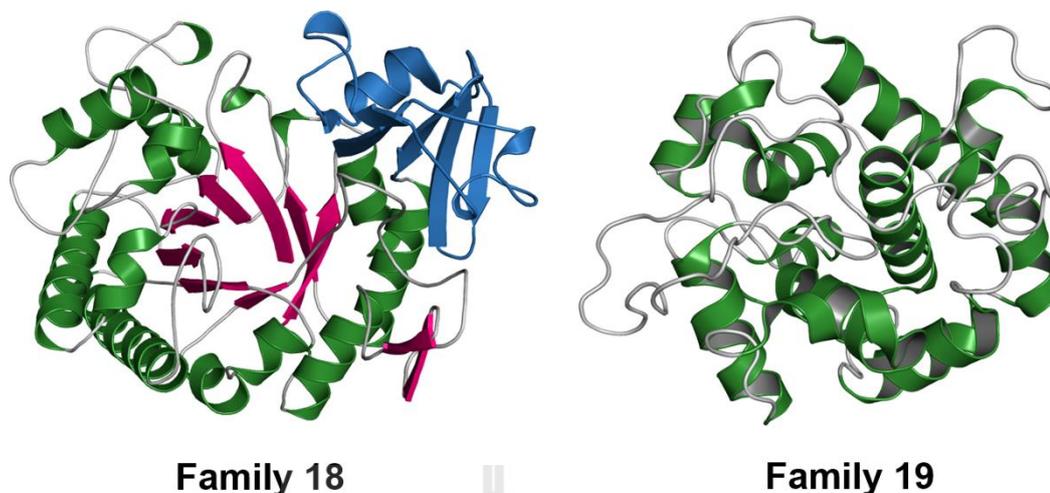


**Figure 1.2** Schematic drawing of the predominant cleaved patterns of chitinolytic enzymes. Subunits of the chitin are shown in light blue and the reducing end sugar in gray (Seidl, 2008).

Chitinases are a member of the glycoside hydrolase superfamily, which has been currently classified into 135 families based on their primary sequence similarity, substrate specificity and occasionally on their molecular mechanism (Davies and Henrissat, 1995; Henrissat, 1991; Henrissat and Bairoch, 1993) and further classified for families lacking significant sequence similarity but having similar three dimensional structure as the same clan (Henrissat and Davies, 1997). According to the Carbohydrate-Active enZYmes database (CAZy database; <http://www.cazy.org>), chitinases belong to families 18 and 19 (GH18 and GH19) and family 18 chitinases are also classified to clan GH-K that characterized by the structure. Enzymes in the two GH families do not share any significant sequence identity and the mechanisms and 3D-structures of representative members are not similar (Davies and Henrissat, 1995).

Family 18 chitinases are mainly found in bacteria, fungi, viruses, insects, mammals, and some plants that produce classes III and V chitinases. In contrast, family 19 chitinases are mainly found in higher plants like classes I, II, and IV plant chitinases and some Gram-positive bacteria, such as *Streptomyces* species (Iseli *et al.*, 1996; Ohno *et al.*, 1996).

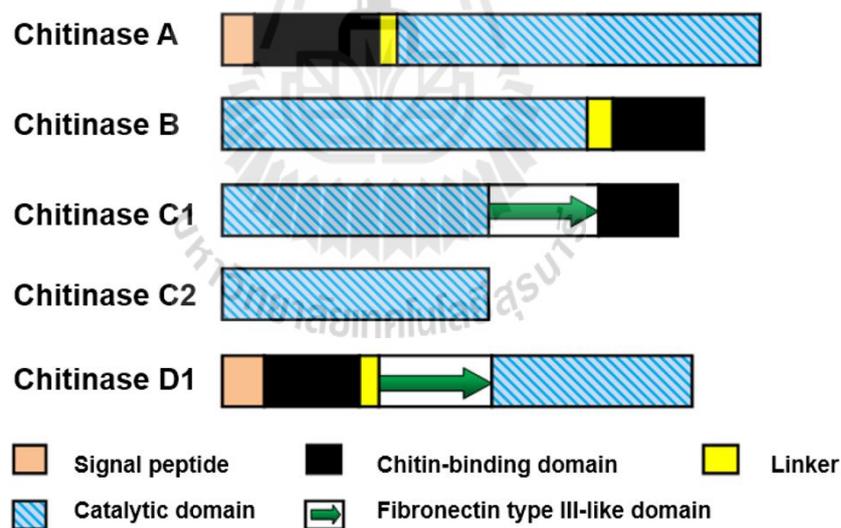
Generally, family 18 chitinases contain multiple functional domains, such as chitin-binding domains (ChBDs) and fibronectin type III-like domains (Fn3 domains), linked to the catalytic domain (Suzuki *et al.*, 1999). The catalytic domain of family 18 chitinases has a  $(\beta/\alpha)_8$  TIM-barrel containing eight  $\alpha$ -helices and eight  $\beta$ -strands, which possess a deep and long substrate-binding cleft formed by loops following the C-termini of the eight parallel  $\beta$ -strand (Figure 2.3) (Aronson *et al.*, 2003; Hollis *et al.*, 2000; Perrakis *et al.*, 1994; Songsiriritthigul *et al.*, 2008; Terwisscha van Scheltinga *et al.*, 1994; van Aalten *et al.*, 2000). On the other hand, the catalytic domain of family 19 chitinases does not possess a TIM-barrel but comprises two lobes that have the  $\alpha$ -helix-rich fold (Figure 1.3) (Hart *et al.*, 1995; Kezuka *et al.*, 2006). In GH19 barley chitinase docked with GlcNAc<sub>6</sub>, the substrate binding cleft is composed of two  $\alpha$ -helices and three stranded  $\beta$ -sheets lying between the two-lobe fold (Figure 1.3) (Davies and Henrissat, 1995; Hart *et al.*, 1995; Henrissat and Davies, 2000). Additionally, these two families use different catalytic mechanisms with two possible stereochemical products, retention or inversion at the anomeric centre (Davies and Henrissat, 1995).



**Figure 1.3** Ribbon representation of the main folds of the catalytic domain of the family 18 (PDB code: 3B9A) and family 19 (PDB code: 3CQL) chitinases (modified from Davies and Henrissat, 1995).

Bacterial family 18 chitinases are further classified into chitinase A, B, C, and D, based on the amino acid sequence identity of their catalytic domains (Suzuki *et al.*, 1999). Chitinase A (ChiA) consists of an *N*-terminal signal peptide preceding a ChBD and a catalytic domain with a small  $\alpha+\beta$  domain inserted between the seventh and eighth  $\beta$ -strands of  $(\beta/\alpha)_8$ -barrel (Brurberg *et al.*, 2001; Suzuki *et al.*, 2002). Chitinase B (ChiB) consists of the catalytic domain that has a fold similar to ChiA and a small putative ChBD at the *C*-terminal part (Brurberg *et al.*, 2001; Suzuki *et al.*, 2002). The difference in their domain topology leads ChiA and ChiB to digest chitin from opposite directions. ChiA degrades a chitin chain from the reducing end, but ChiB from the non-reducing end (Uchiyama *et al.*, 2001; van Aalten *et al.*, 2000). Recently, chitinase C1 (ChiC1) and chitinase C2 (ChiC2) were reported. ChiC1 has a fibronectin type III-like domain (FnIIID) at the *C*-terminal end that is absent in ChiA and ChiB (Suzuki *et al.*, 2002), whilst ChiC2 is a proteolytic derivative of ChiC1 (Alam *et al.*,

1996; Suzuki *et al.*, 2002) generated by cleavage of the *C*-terminal portion. The catalytic domains of ChiB and ChiC1 are located at the *N*-terminus which lacks the signal peptide. Chitinase D (ChiD1) is another chitinase that was identified in *B. circulans* WL-12. ChiD1 contains a signal peptide before the *N*-terminal ChBD, which is similar to ChiA, connected with an FnIII domain (Alam *et al.*, 1996; Brurberg *et al.*, 2001), and then the TIM barrel domain at the *C*-terminus. All GH18 chitinases contain a catalytic domain with the SxGG and DxTxE conserved motifs. These two motifs are the characteristic of a family 18 glycoside hydrolase (Brurberg *et al.*, 2001). The domain organization of chitinases A, B, C1 and C2 from *S. marcescens* 2170 and ChiD1 from *B. circulans* is summarized in Figure 1.4 (Suzuki *et al.*, 1999).



**Figure 1.4** The domain organization of bacterial chitinases (modified from Suzuki *et al.*, 1999).

GlcNAcases are classified into glycoside hydrolase family 3 (GH3), family 20 (GH20) and family 84 (GH84) based on their amino acid sequence similarity, substrate specificity and the mode of enzyme action (Carbohydrate-Active enZymes Database; <http://www.cazy.org>; Davies and Henrissat, 1995; Henrissat and Bairoch, 1993; Henrissat and Davies, 1997). Family 3 glycoside hydrolases include  $\beta$ -glucosidases (EC 3.2.1.21),  $\beta$ -xylosidases (EC 3.2.1.37),  $\beta$ -glucosylceramidases (EC 3.2.1.45),  $\beta$ -*N*-acetylhexosaminidases (EC 3.2.1.52),  $\beta$ -*N*-acetylglucosaminidases (EC 3.2.1.52),  $\alpha$ -L-arabinofuranosidases (EC 3.2.1.55), exo-glucanases (EC 3.2.1.-). Family 20 glycoside hydrolases include  $\beta$ -*N*-acetylhexosaminidases (EC 3.2.1.52),  $\beta$ -*N*-acetylglucosaminidases (EC 3.2.1.52) and lacto-*N*-biosidases (EC 3.2.1.140). Family 84 glycoside hydrolases include  $\beta$ -*N*-acetylglucosaminidases (EC 3.2.1.52) and hyaluronidases (EC 3.2.1.35).  $\beta$ -*N*-acetylglucosaminidases (GlcNAcases) hydrolyse the  $\beta$ -1,4-glycosidic linkage between two *N*-acetyl-D-glucosamine residues (Tews *et al.*, 1996) whereas  $\beta$ -*N*-acetylhexosaminidases ( $\beta$ -Hexs) catalyse the removal of both gluco- and galacto- configurations of *N*-acetyl- $\beta$ -D-hexosaminides from the non-reducing end of oligosaccharide and their glycoconjugates as well as a hydrolysis of G<sub>M2</sub> ganglioside (Robbins *et al.*, 1992; Sandhoff and Kolter, 1998). There are only five bacterial GH3 GlcNAcases have been characterized, including NagZ from *Bacillus subtilis* (Litzinger *et al.*, 2010), NagZ from *Salmonella typhimurium* (Bacik *et al.*, 2012), NagA from *Thermotoga maritima* (Mine *et al.*, 2014), CbsA from *Thermotoga neapolitana* (Choi *et al.*, 2009) and NagZ from *Vibrio cholera* (Balcewich *et al.*, 2009). In contrast, the GH20 GlcNAcases have been characterized from a wide variety of organisms ranging from bacteria to human, including from insect (Liu *et al.*, 2011), human (Lemieux *et al.*, 2006; Mark *et al.*, 2003) and bacterial GlcNAcases such as chitobiase from *Serratia*

*marcescense* (Tews *et al.*, 1996),  $\beta$ -hexosaminidase from *Streptomyces plicatus* (Mark *et al.*, 2001), disperin B ( $\beta$ -1,6-*N*-acetylglucosaminidase) from *Actinobacillus actinomycetemcomitans* (Ramasubbu, *et al.*, 2005), *N*-acetyl- $\beta$ -D-glucosaminidase from *Streptococcus gordonii* (Langley *et al.*, 2008),  $\beta$ -*N*-acetylhexosaminidase from *Paenibacillus* sp. (Sumida *et al.*, 2009),  $\beta$ -*N*-acetylglucosaminidases from *Vibrio harveyi* 650 (Suginta *et al.*, 2010),  $\beta$ -*N*-acetylhexosaminidase from *Streptococcus pneumoniae* R6 (Jiang *et al.*, 2011),  $\beta$ -*N*-acetylglucosaminidase from *Streptococcus pneumoniae* TIGR4 (Pluvinaige *et al.*, 2011) and  $\beta$ -*N*-acetylhexosaminidase from *Streptomyces coelicolor* A3(2) (Thi *et al.*, 2014). For GH84 GlcNAcase, there is only one  $\beta$ -*N*-acetylglucosaminidase (NagJ) from *Clostridium perfringens* has been characterized (Rao *et al.*, 2006).

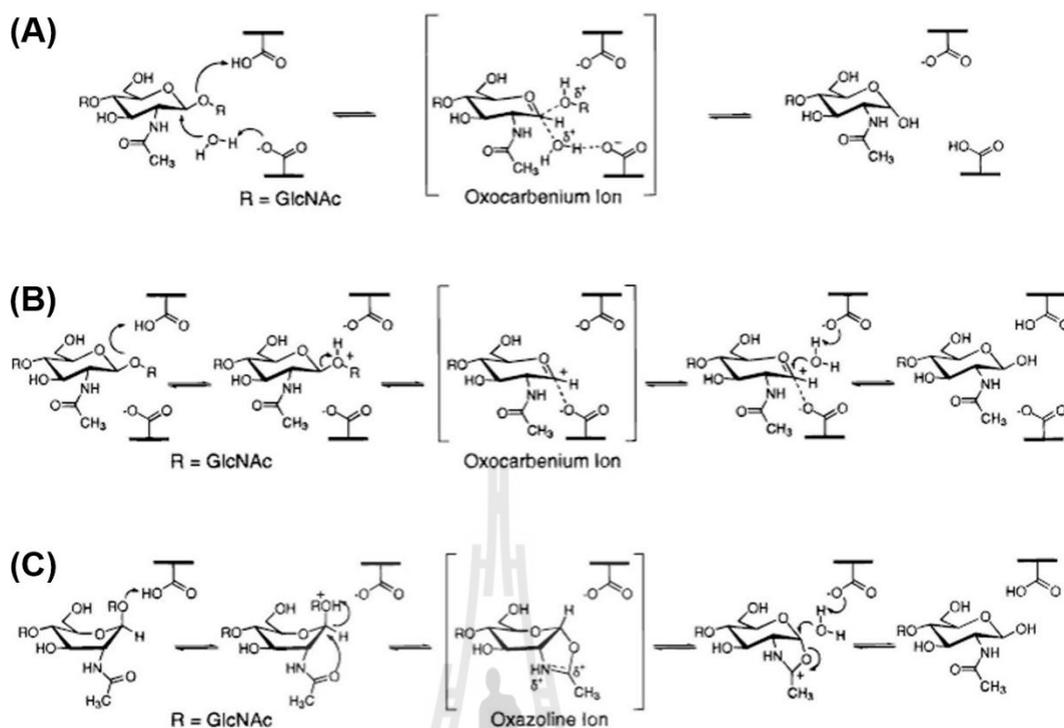
#### 1.4 Catalytic mechanism of chitinases and GlcNAcases

In general, enzymatic hydrolysis of  $\beta$ -(1,4) glycosidic linkages takes place by a general acid-base catalysis that requires two amino acid residues, one acting as a proton donor and the other as a nucleophile or base (Koshland, 1953; Sinnott, 1990). The mechanistic pathway for acid-catalyzed glycosyl hydrolases is shown in Figure 1.5. Figure 1.5A represents inversion of the stereochemistry whereas Figure 1.5B-C represent retention of the stereochemistry of the anomeric oxygen at C1 relative to the initial configuration (Brameld and Goddard, 1998).

With both retaining and inverting mechanisms, the protein donor position is identical or within the hydrogen-bonding distance to the glycosidic oxygen. The model of the inverting mechanism, described as the single displacement mechanism (see Figure 1.5A), is proposed for family 19 chitinases (Brameld and Goddard, 1998, 3rd).

The single displacement catalysis requires two acidic residues within the active site. The first acidic residue acts as a general acid and the second acts as a nucleophile. An example of GH19 chitinase is chitinase from barley (*Hordeum vulgare* L.). This enzyme requires Glu67 (the proton donor) and Glu89 (the nucleophile) for catalysis (Hart *et al.*, 1995).

In retaining enzymes, the nucleophilic catalytic base is in close vicinity of the sugar anomeric carbon whereas the base in inverting enzymes is more distal because they must accommodate a water molecule between the base and the sugar (Davies and Henrissat, 1995). Hen egg white lysozyme (HEWL) was the first glycoside hydrolase to have its three dimensional structure solved (Matthews and Remington, 1974). HEWL is an example of a retaining enzyme that requires two acidic residues, one protonated (Glu35) and the other one deprotonated (Asp52) (Phillips, 1967). Following the retaining mechanism, the  $\beta$ -(1,4)-glycosidic oxygen is first protonated by the carboxylic group of Glu35, which act as a general acid, leading to an positively charged oxocarbenium ion intermediate at the -1 sugar residue. This intermediate is stabilized by a second carboxylate group of Asp52 via covalent or electrostatic interactions. The leaving group (or the reaction product) subsequently diffuses out of the active site and is replaced by a water molecule. A nucleophile attack by water yields the hydrolytic product, which retains the initial  $\beta$ -anomeric configuration. This hydrolytic model is designated as the double displacement mechanism (see Figure 1.5B) (Brameld and Goddard, 1998).

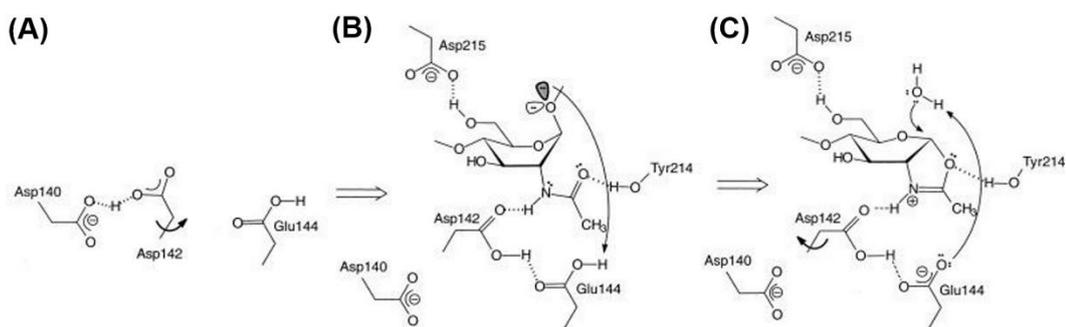


**Figure 1.5** The catalytic mechanism of glycoside hydrolases. (A) Single-displacement mechanism. (B) Double-displacement mechanism. (C) Substrate-assisted mechanism. (modified from Brameld and Goddard, 1998).

The second model of the retaining mechanism is the substrate-assisted mechanism (see Figure 1.5C). The initial pathway of this mechanism is identical to the double displacement one, but an oxazolinium ion intermediate is formed via anchimeric assistance by the neighboring *N*-acetyl group (Brameld *et al.*, 1998).

On the basis of structural and stereochemical studies of chitin hydrolysis, all of the family 18 chitinases catalyse their reactions through the substrate-assisted mechanism or retaining mechanism in which the substrate at subsite -1 is distorted to a boat conformation prior to protonation, leading to a glycosidic bond cleavage and formation of the oxazolinium ion intermediate (Armand *et al.*, 1994; Bortone *et al.*, 2002; Brameld *et al.*, 1998; Papanikolaou *et al.*, 2001; Terwisscha van Scheltinga *et al.*,

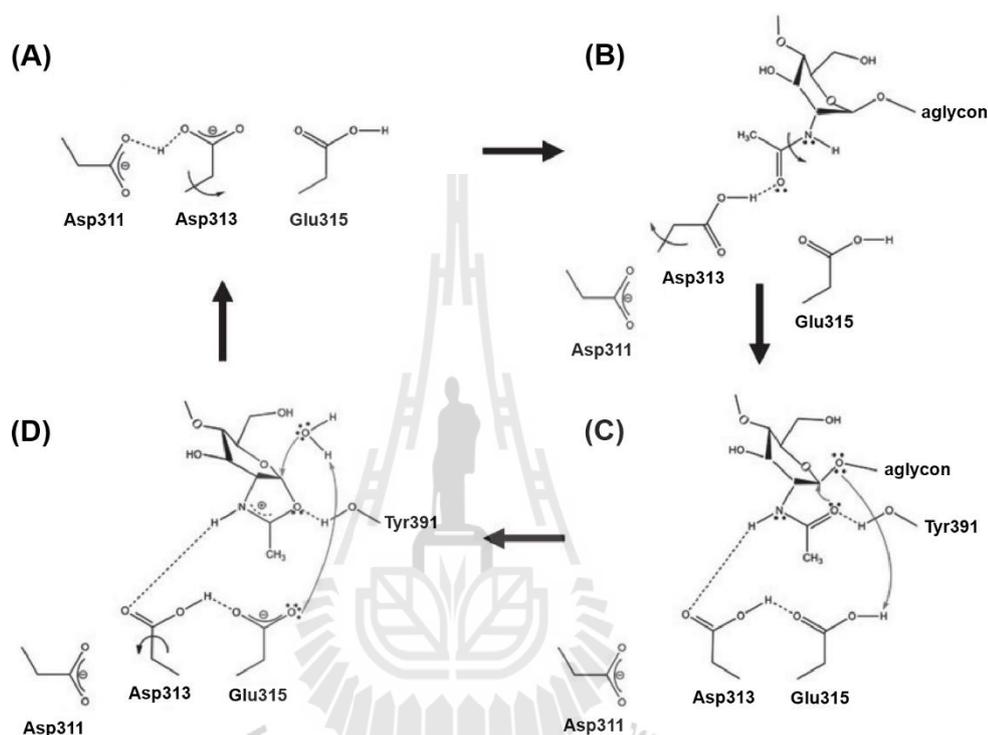
1995; Tews *et al.*, 1997). From the multiple sequence alignment, all family 18 chitinases share a conserved DxDxE motif that is buried in the core of the  $(\beta/\alpha)_8$  fold. These acidic residues include the catalytic acid/base residue Glu315 in *SmChiA* or Glu144 in *SmChiB* that acts as a proton donor due to its position in the vicinity to donate a proton to the oxygen O4 of +1 subsite sugar unit (Kolstad *et al.*, 2002; Kolstad *et al.*, 2004; Papanikolau *et al.*, 2001; Perrakis *et al.*, 1994). Subsequent cleavage of glycosidic bond leads to a formation of an oxazolinium ion intermediate that is stabilized by hydrogen bonding interaction with the protonated Asp142 in *SmChiB* as shown in Figure 1.6A-B. The rotation of Asp142 also causes lowering  $pK_a$  of Glu144 of about 0.8 pH units (Synstad *et al.*, 2004), which promotes a proton transfer to the oxygen in the scissile glycosidic bond. Then, a proton from a water molecule is abstracted by the  $\gamma$ -carboxylate of Glu144 in *SmChiB* and the remaining hydroxide anion is taken up by the C1 carbon of -1 sugar, yielding the hydrolytic product with retained anomeric configuration (Figure 1.6C) (Kolstad *et al.*, 2002; van Aalten *et al.*, 2001).



**Figure 1.6** The proposed catalytic mechanism of family 18 chitinase B from *Serratia marcescens* (*SmChiB*). (A) Resting enzyme. (B) Binding of substrate and rotation of Asp142 toward Glu144, enabling hydrogen bond interactions between the hydrogen of the acetamido group, Asp142 and Glu144. (C) Hydrolysis of the oxazolinium ion intermediate leads to protonation of Glu144 and rotation of Asp142 to its original position where it shares a proton with Asp140 (van Aalten *et al.*, 2001).

Recently, the refined catalytic cycle of chitin degradation by *VhChiA* has been reported and the acidic residues in the DxDxE sequence motif were suggested to play a concerted role in the catalysis (Suginta and Sritho, 2012). A study of the enzyme-substrate interactions of *VhChiA* revealed two conformations of Asp313 and -1 GlcNAc. The first conformation, likely to be the initial conformation, showed that the  $\beta$ -COOH of Asp313 detaches from Asp311 and rotates to interact with the  $\text{-C=O}$  of the *N*-acetamido group of the -1GlcNAc through hydrogen bonding (Figure 1.7A-B). The second conformation, is derived from the first conformation, in which the Asp313 side chain makes hydrogen bonds between the  $\text{-NH}$  of the *N*-acetyl group and the  $\gamma$ -COOH of Glu315 and then facilitates bond cleavage by nucleophilic attack of the  $\text{-C=O}$  of the *N*-acetamido group on the transient C1-GlcNAc cation and protonation of the glycosidic oxygen by Glu315 (Figure 1.7C), leading to formation of oxazolinium ion

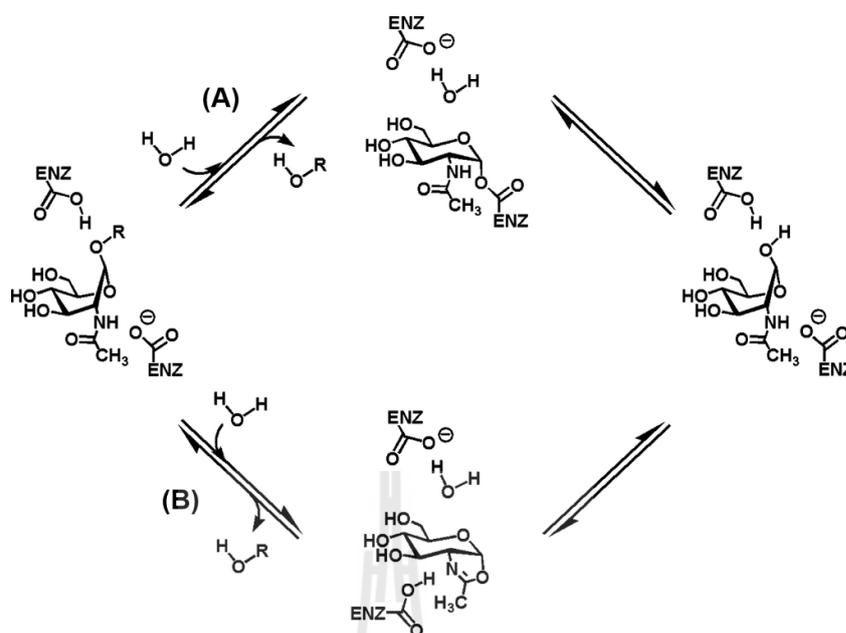
intermediate which is stabilized by Asp313 (Figure 1.7D). The second nucleophilic attack on the reaction intermediate by a water molecule leads to the product in retention configuration (Figure 1.7D) (Suginta and Sritho, 2012).



**Figure 1.7** The refined catalytic cycle of chitin degradation by *Vibrio harveyi* chitinase A (*VhChiA*). (A) Pre-priming. (B) Substrate binding. (C) Bond cleavage. (D) Formation of the reaction intermediate (modified from Suginta and Sritho, 2012).

The chemical mechanism in glycoside hydrolase families 3, 20 and 84 is distinct based on the association of the C2-acetamido group of the substrate. The GH3 enzymes use a typical retaining mechanism involving the formation and breakdown of a covalent  $\alpha$ -glycosyl enzyme intermediate formed on an aspartate residue (Figure 1.8A) (Vocadlo *et al.*, 2000; Vocadlo and Withers, 2005). In the first step of the reaction, the formation of the intermediate, departure of the aglycon leaving group is typically facilitated by a

general acid/base catalytic residue, which donates a hydrogen atom to the glycosidic oxygen whereas the nucleophile forms a covalent glycosyl-enzyme intermediate. In the second step of the reaction, the breakdown of the intermediate, the same catalytic residue acts as a general base by enhancing the nucleophilicity of a water molecule poised near the anomeric center. This water molecule attacks on the anomeric center of substrate being the formation of the product with retained stereochemistry (Vocadlo *et al.*, 2005). In contrast, GH20 and GH84 enzymes are typically employed the substrate-assisted retaining mechanism (Drouillard *et al.*, 1997; Mark *et al.*, 2001; Rao *et al.*, 2006), which is similar to family 18 chitinases (Vocadlo and Withers, 2005). This substrate-assisted mechanism is involved with the carbonyl oxygen of the C2-acetamido group that acts as a nucleophile to displace the aglycon leaving group with the formation of an oxazolinium ion intermediate. Afterwards, a nucleophilic attack from a water molecule at the anomeric center breaks down the oxazolinium ring to generate the product with retained stereochemistry (Figure 1.8B) (Vocadlo *et al.*, 2005). Recent studies of GlcNAcases have also provided reasonable evidence supporting substrate-assisted catalysis by both enzyme kinetics (Drouillard *et al.*, 1997; Williams *et al.*, 2002) and structural studies (Mark *et al.*, 2003; Mark *et al.*, 2001; Tews *et al.*, 1996) show an adjacent Asp-Glu pair in the GH20 GlcNAcases as the key catalytic residues in the catalytic cycle (Hou *et al.*, 2001).



**Figure 1.8** The catalytic mechanisms of the two classes of GlcNAcases. (A) Family 3 GlcNAcases use an anionic enzymic carboxylate group as the nucleophile to form a covalent glycosyl enzyme intermediate. (B) Families 20 and 84 GlcNAcases use the 2-acetamido group of the substrate that acts as a nucleophile to form an oxazolinium ion intermediate (modified from Vocadlo *et al.*, 2005).

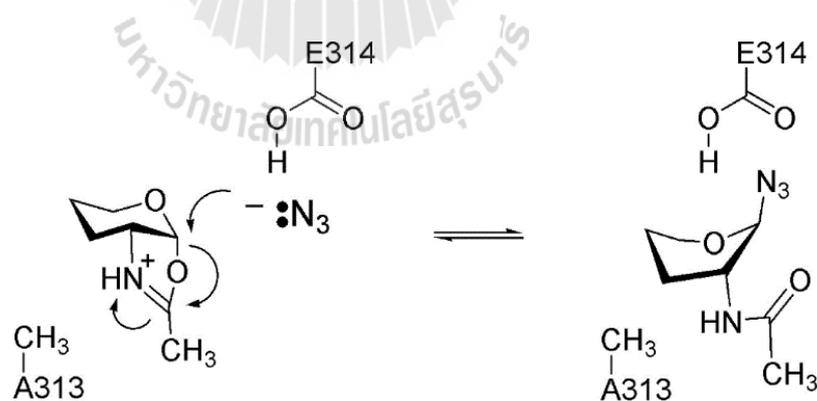
## 1.5 Studies of the chemical rescue in retaining enzymes

In the substrate-assisted mechanism of GH20 enzymes, two adjacent acidic side-chains (Asp-Glu) in the active site form a catalytic pair. The aspartate residue typically acts as the catalytic base/nucleophile, while the glutamic acid acts as the catalytic proton donor/acceptor (Maier *et al.*, 2003; Williams *et al.*, 2002). The effects of exogenous anions, such as azide or formate, can provide direct evidence identifying the catalytic acid/base residues in retaining glycoside hydrolases. Following mutation of the acid-base residue or the nucleophilic residue, hydrolytic activity can be rescued by the addition of an exogenous nucleophile, such as azide or formate (Comfort *et al.*,

2007; Zechel and Withers, 2001), resulting in the formation of products with the  $\alpha$  or  $\beta$  configuration. Examples of enzymes studied by use of this approach include *Streptomyces plicatus* GH20 hexosaminidase (*SpHex*) (Williams *et al.*, 2002), *Arthrobacter protophormiae* GH85 endo- $\beta$ -N-acetyl-glucosaminidase (Endo A) (Fujita *et al.*, 2007), *Streptomyces* sp. GH1  $\beta$ -glucosidase (Vallmitjana *et al.*, 2001), *Paenibacillus* sp. TS12 GH3 glucosylceraminidase (Paal *et al.*, 2004), *Cellulomonas fimi* GH10 exoglucanase/xylanase (MacLeod *et al.*, 1994), *Bacillus licheniformis* GH16 1,3-1,4- $\beta$ -glucanase (Viladot *et al.*, 1998), *Sulfolobus solfataricus* GH29  $\alpha$ -L-fucosidase (Cobucci-Ponzano *et al.*, 2003) and *Geobacillus stearothermophilus* T-6 GH51  $\alpha$ -L-arabinofuranosidase (Shallom *et al.*, 2002). All studies demonstrated that sodium azide produced a significant recovery of the glycoside hydrolase activity of inactive mutants in which one of the catalytic pair of acidic residues was mutated, for example, the key catalytic residues Glu134 and Glu138 of *Bacillus licheniformis* 1,3-1,4- $\beta$ -glucanase were mutated to alanine by site-directed mutagenesis and showed a drastic decrease in hydrolytic activity of the mutants E138A and E134A. Addition of sodium azide, which acts as an exogenous nucleophile, reactivated the hydrolytic activity of inactive mutant E138A, yielding a  $\beta$ -glycosyl azide product, arising from nucleophilic attack of azide anion on the glycosyl-enzyme intermediate in the deglycosylation step whereas azide anions reactivated the inactive mutant E134A through a single inverting displacement, forming the  $\alpha$ -glycosyl azide product in the glycosylation step. Such results suggested that Glu138 acts as a general acid/base residue while Glu134 acts as the catalytic nucleophile (Viladot *et al.*, 1998).

In the most relevant case of the GH20 GlcNAcases, rescue of the activity of *SpHex* from *Streptomyces plicatus* (Williams *et al.*, 2002) has been demonstrated.

*SpHex* catalyses the hydrolysis of *N*-acetyl- $\beta$ -hexosaminides in which an acidic pair Asp313-Glu314 is identified to be the catalytic residues. Point mutation of Asp313 of *SpHex* to Ala or Asn (mutants D313A or D313N) almost abolished the enzyme's hydrolytic activity, but the catalytic activity of the mutant D313A was significantly increased with the inclusion of sodium azide in the assay medium. For such results, the predicted functional roles of Asp313 in *SpHex* were to aid the C2-acetamido group of -1GlcNAc in acting as a powerful nucleophile and provide charge stabilization of the transition state that flanks the oxazolinium ion intermediate during the glycosylation and deglycosylation steps of substrate hydrolysis. Alternatively, azide anion may reactivate the inactive mutant by acting as a strong nucleophile that competes with the hydroxyl group of a water molecule to interact with the oxazolinium ion intermediate, yielding the  $\beta$ -glycosyl azide product in the deglycosylation step (Figure 1.9) (Williams *et al.*, 2002).



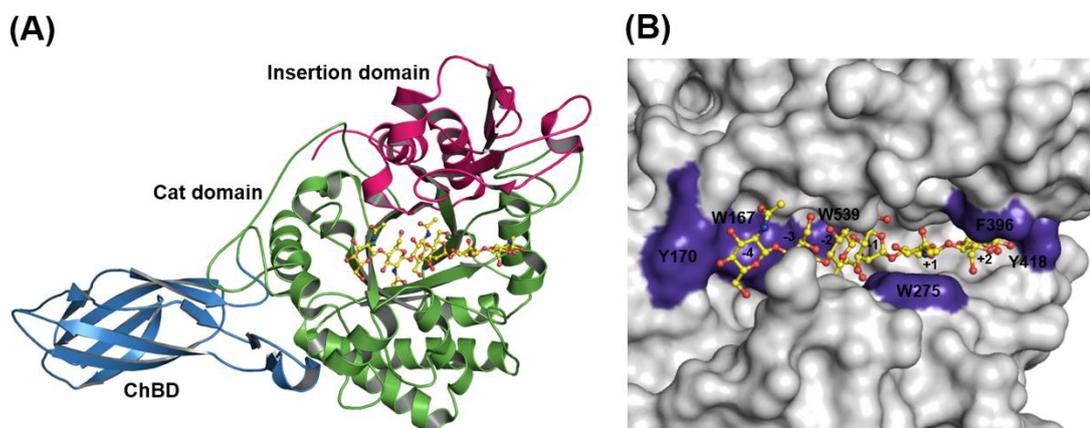
**Figure 1.9** Proposed mechanism of azide-mediated rescue with the *SpHex* D313A mutant. Azide anion as an alternative nucleophile to water acts to open the oxazolinium ion intermediate (Williams *et al.*, 2002).

## 1.6 Structural analysis of family 18 chitinases

### 1.6.1 Structural analysis of chitinase A and B from *S. marcescens*

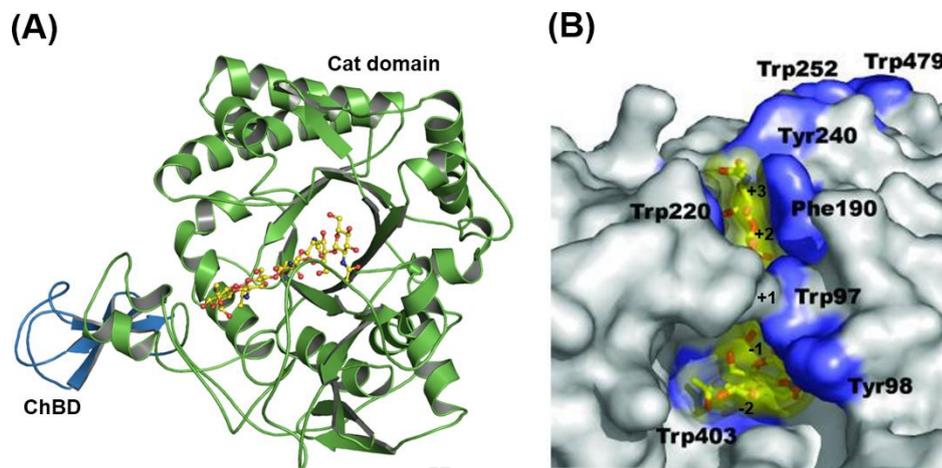
The three-dimensional structure of bacterial chitinase A from *S. marcescens* was the first structure to be solved and refined to 2.3 Å resolution (Perrakis *et al.*, 1994). The overall structure of *S. marcescens* ChiA consists of three domains designated as: i) an *N*-terminal chitin-binding domain (ChBD) comprising mostly  $\beta$ -stands ii) a *C*-terminal catalytic (Cat) domain containing the  $(\beta/\alpha)_8$ -barrel fold and iii) a small insertion domain comprising three  $\alpha$ -helices and five  $\beta$ -strands that are inserted into the TIM barrel. The crystal structure of *S. marcescens* chitinase A mutant E315L complexed with the GlcNAc<sub>6</sub> is shown in Figure 1.10A.

*S. marcescens* chitinase A was found to degrade a chitin chain from the reducing end and the catalytic site of the enzyme contains six substrate binding subsites extending from subsite -4 to +2 (Figure 1.10B). Trp275 and Phe396 are important for substrate binding at the +1 and +2 subsites and stacked against the hydrophobic faces of the corresponding GlcNAcs. Tyr418 seems to mark the end of the binding cleft, however, this residue did not interfere with the extension of the reducing end beyond the subsite +2 (Aronson *et al.*, 2003; Papanikolaou *et al.*, 2001).



**Figure 1.10** The crystal structure of *S. marcescens* chitinase A mutant E315L in complex with GlcNAc<sub>6</sub> (PDB code: 1NH6). (A) A ribbon representation of the domain organization of *S. marcescens* chitinase A. (B) A surface representation of the substrate-binding cleft. The aromatic residues that interact with the substrate are highlighted in blue (modified from Aronson *et al.*, 2003).

Unlike the chitinase A from the same species, a three-dimensional structure of chitinase B from *S. marcescens* refined to 1.9 Å resolutions (van Aalten *et al.*, 2000) reveals that chitinase B consists of an *N*-terminal (β/α)<sub>8</sub>-TIM barrel catalytic domain, a linker and a small *C*-terminal chitin-binding domain. The TIM barrel catalytic domain has a fold similar to that of chitinase A as shown in Figure 1.11A. The catalytic site of the enzyme is defined as subsites -3 to +3 and its substrate-binding cleft has a tunnel-like character (van Aalten *et al.*, 2000) lined with exposed aromatic residues as shown in Figure 1.11B. Binding of chitin oligomers is blocked beyond the subsite -3, which explains why chitinase B degrades chitin chains from the nonreducing end (van Aalten *et al.*, 2000; 2001) and primarily converts chitin to dimers with exoactivity (Brurberg *et al.*, 1996; Suzuki *et al.*, 2002).

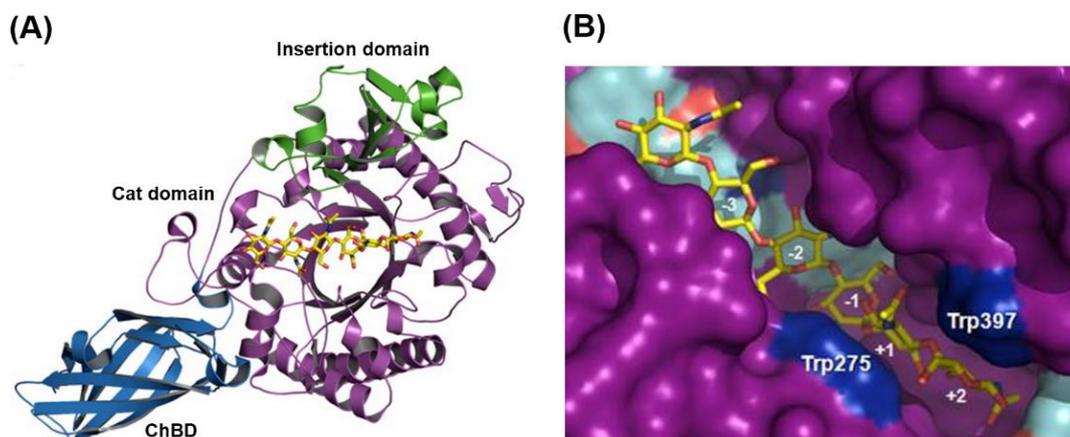


**Figure 1.11** The crystal structure of *S. marcescens* chitinase B mutant E144Q in complex with GlcNAc<sub>5</sub> (PDB code: 1E6N). (A) A ribbon representation of the domain organization of *S. marcescens* chitinase B. (B) A surface representation of the substrate-binding cleft with GlcNAc<sub>5</sub> bound from subsites -2 to +3. The surface exposed aromatic residues that interact with the substrate are highlighted in blue (modified from Horn *et al.*, 2006).

### 1.6.2 Structural analysis of chitinase A from *V. harveyi*

*V. harveyi* (formerly *V. carchariae*) is a Gram-negative marine bacterium that shows high level of chitinase A activity in the presence of chitin. *V. harveyi* chitinase A is classified as a member of family 18 chitinases and active as a monomer with MW of 63 kDa (Suginta *et al.*, 2000). Analysis of chitin hydrolysis using HPLC/ESI-MS revealed that *V. harveyi* chitinase A acts as an endochitinase (Suginta *et al.*, 2004). The enzyme shows a broad range of substrate specificity with various chitin oligomers. At the initial time of reaction, chitinase A yields predominantly  $\beta$ -anomers, which supports the substrate-assisted mechanism as described for other family 18 chitinases (Suginta *et al.*, 2005).

The crystal structures of the wild-type and mutated chitinase A revealed that chitinase A consists of three domains, an *N*-terminal chitin binding domain, a  $(\beta/\alpha)_8$  TIM-barrel catalytic domain and a small  $(\alpha+\beta)$  insertion domain (Figure 1.12A) (Songsiriritthigul *et al.*, 2008). The catalytic residue (Glu315) is positioned in the loop of the strand  $\beta_4$  which is part of a DxDxE sequence motif whereas the Asp313 positioned at the bottom of the binding cleft next to catalytic residue 315 plays several important roles in the catalytic cycle of *VhChiA* by contributing to substrate binding, stabilizing the oxazolinium ion intermediate and lowering the  $pK_a$  of catalytic residue 315 to facilitate a bond cleavage (Suginta and Sritho, 2012). The structure of the mutant E315M complexed with  $(\text{GlcNAc})_6$  displays the substrate-binding cleft as a long deep groove, which contains six-binding subsites (-4)(-3)(-2)(-1)(+1)(+2) (Figure 1.12B). This subsite topology defines subsite -4 at the non-reducing end (NRE), subsite +2 at the reducing end (RE) and the cleavage site between -1 and +1 subsites. Trp231 and Tyr245, and Tyr31 and Trp70 appear to line up in the position for binding the longer chain chitins (Songsiriritthigul *et al.*, 2008). Trp275 and Trp397 stacked against the +1 GlcNAc and +2GlcNAc, respectively (Figure 1.12B) are suggested to be important for the primary interaction with soluble substrates (Suginta *et al.*, 2007) and are also crucial for the progressive degradation of insoluble chitin (Suginta *et al.*, 2009).



**Figure 1.12** The structure of inactive mutant E315M complexed with GlcNAc<sub>6</sub> of *V. harveyi* chitinase A (PDB code: 3B9A). (A) A ribbon representation of the domain organization of *V. harveyi* chitinase A. (B) A surface representation of E315M in complex with GlcNAc<sub>6</sub> that fully occupied subsites -4 to +2 within the substrate binding cleft lining with surface-exposed aromatic residues (Songsiriritthigul *et al.*, 2008; Suginta *et al.*, 2009).

## 1.7 Studies of pK<sub>a</sub> calculations

### 1.7.1 Theory of protein pK<sub>a</sub> calculation

The pH-dependent ionization of a protein titratable group is typically described using the Henderson-Hasselbalch equation:

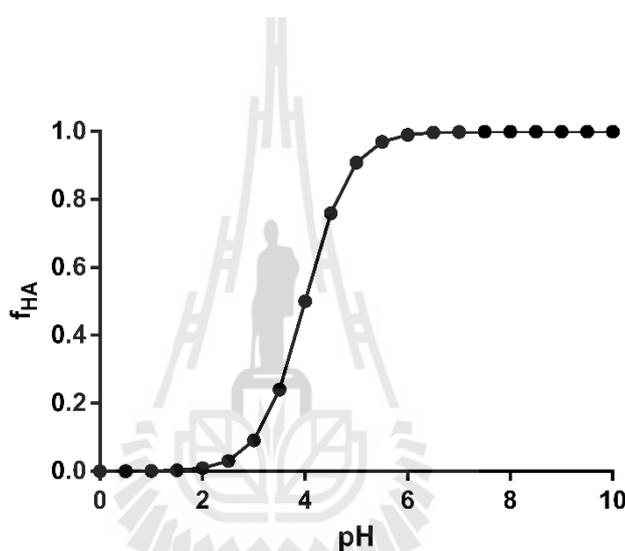
$$\text{pH} = \text{pK}_a + \log \frac{[\text{A}^-]}{[\text{HA}]} \quad (1.1)$$

where,  $\text{pH} = -\log[\text{H}_3\text{O}^+]$ ,  $\text{pK}_a = -\log K_a$ ,  $K_a = [\text{H}_3\text{O}^+][\text{A}^-]/[\text{HA}]$  and  $[\text{HA}]$  and  $[\text{A}^-]$  are the concentrations of the acid and its conjugate base, respectively. The dissociation of a proton from an acid HA to A<sup>-</sup> causes a change of charge on the titratable group. It is seen that the pK<sub>a</sub> values of an acid is the pH value where the concentrations of the protonated and deprotonated forms of an acid are the same. Furthermore by rearranging

the Henderson-Hasselbalch equation, the fractional charge  $f_{\text{HA}}$  of an acid can be obtained as the equation:

$$f_{\text{HA}} = \frac{[\text{HA}]}{[\text{HA}] + [\text{A}^-]} = \frac{1}{K_a + 1} = \frac{1}{10^{\text{pH} - \text{p}K_a} + 1} \quad (1.2)$$

and plotting of  $f_{\text{HA}}$  as a function of pH will give the well-known sigmoid titration curve as shown in Figure 1.13).



**Figure 1.13** The titration curve of an acid with a  $\text{p}K_a$  value of 4.0 calculated using equation 1.2.

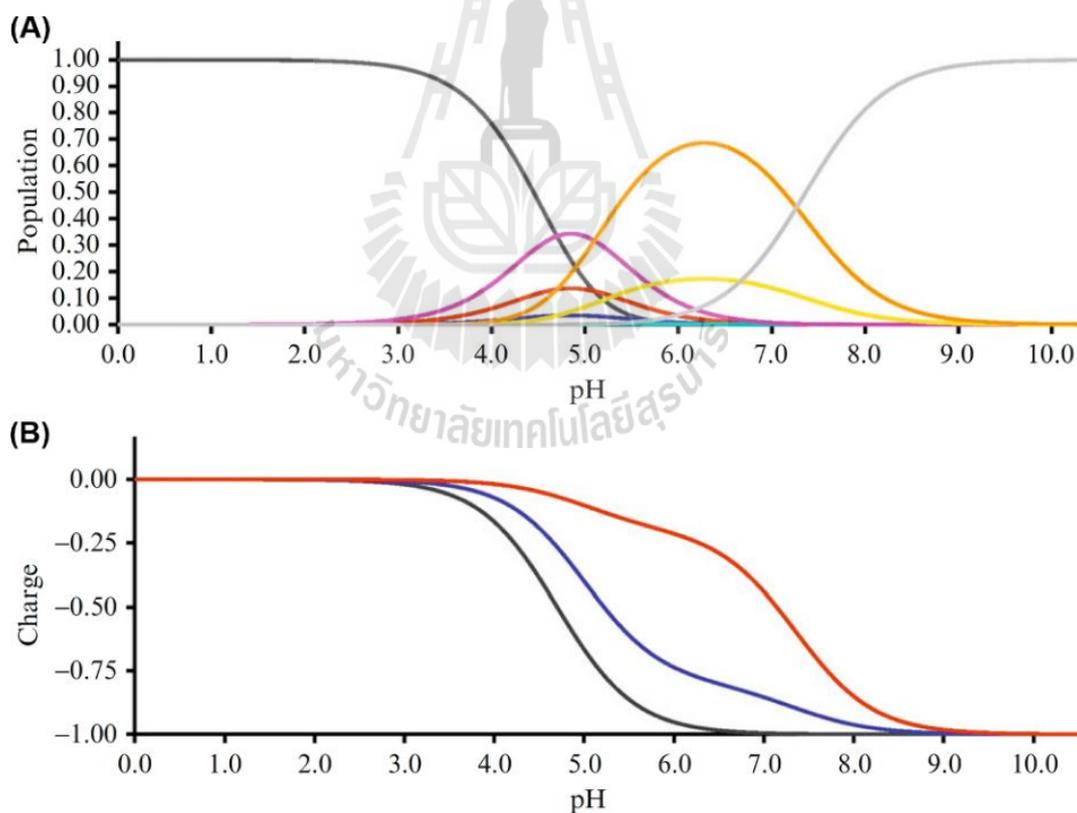
In proteins, the chemical reactions with water is the uptake and release of protons by amino acids in which the parts of amino acids in a protein that can uptake or release protons will be referred as protein titratable groups. The titration of titratable groups that have the  $\text{p}K_a$  values in water within the range from 0-14 will be considered, for example, the titratable groups in the side chains of Asp, Glu, Tyr, Cys, His, Lys and Arg as well as the two terminal groups. These titratable groups are defined as bases when titratable groups convert from positive to neutral and acids as titratable groups

that exist in neutral and negative forms. In the following, Asp, Glu, Cys, Tyr and the C-terminus are defined as acids whereas His, Arg, Lys and the N-terminus are defined as bases.

The  $pK_a$  value of a titratable group is a measure of the free energy difference between the neutral and the charged forms of a titratable group. It is therefore possible to calculate the  $pK_a$  value of the group if we can calculate the energy difference between the neutral and the charged form of a titratable group. This energy difference can be calculated using a sampling or simulation technique, which most  $pK_a$  calculation algorithms in proteins do exactly.

A protein that has  $N$  titratable groups can typically occupy  $2^N$  different protonation states. Each titratable group can exist in at least two protonation states: charged and neutral. Each of all possible protonation states can be associated with a specific free energy, and when the pH dependence of these energies is known, the fractional charges are typically obtained by evaluating the Boltzmann sum. For proteins containing more than 35 titratable groups, it is no longer possible to evaluate the Boltzmann sum. For a large system, a Monte Carlo sampling method (Beroza *et al.*, 1991) and the cluster approach (Gilson, 1993) have been typically used to produce the accurate fractional charges at every pH values. The energy of every possible protonation state of the protein at a particular pH is then converted into fractional charge for each residue at each pH value in order to get the titration curves. A plot of the pH-dependent population of the protonation states for three-group system is obtained by evaluation of the partition function displayed in pKaTool program (Figure 1.14A) (Nielsen, 2007, 2009).

The fractional charge of a particular group is simply the sum of the partition function for all the states where a specific group is charged at a given pH value. For example, if there are eight ( $2^3$ ) protonation states, the charge of a specific group 1 (G1) is the sum of  $p_i$  in states 1, 2, 3, and 4 where G1 is charged (Table 1.1). This calculation is repeated for each pH value of interest, and the charge is plotted as a function of pH to construct titration curves as shown in Figure 1.14B. According to the calculated titration curves, the final  $pK_a$  value for each titratable group is determined as the pH value where the group is half-protonated.



**Figure 1.14** pH dependence of population of protonation states for three-group system and its titration curves. (A) pH dependence of the relative populations of eight possible protonation states. The pink bell-shaped curve represents one deprotonated group,

whereas the orange bell-shaped curve represents the protonation states with two groups deprotonated. The single black curve at low pH represents protonation states with all groups protonated whereas the single gray curve at high pH represents protonation states with all groups deprotonated. (B) Titration curves of three titratable acids: black titration curve with  $pK_a$  value 4.7, blue titration curve with  $pK_a$  value 5.1 and red titration curve with  $pK_a$  value 5.7 (Nielsen, 2009).

**Table 1.1** Possible protonation states for a hypothetical protein consisting of three titratable groups.

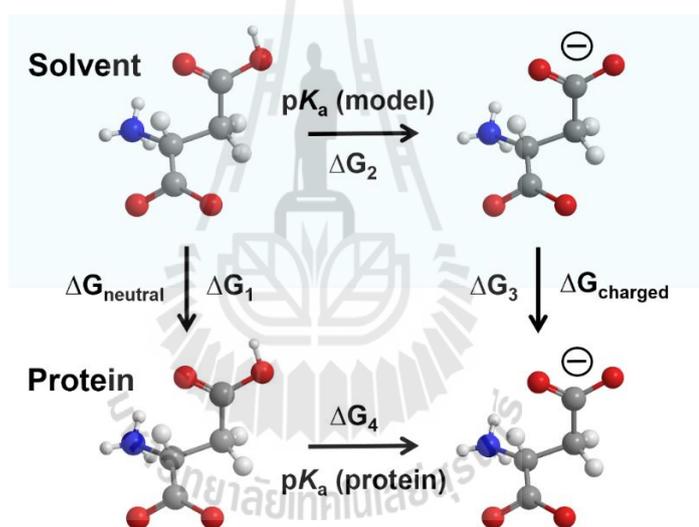
State	G1	G2	G3	Energy
1	+	+	+	$\Delta G_{pH}(1) + \Delta G_{pH}(2) + \Delta G_{pH}(3) + (1=2) + (1=3) + (2=3)$
2	+	+	0	$\Delta G_{pH}(1) + \Delta G_{pH}(2) + (1=2)$
3	+	0	+	$\Delta G_{pH}(1) + \Delta G_{pH}(3) + (1=3)$
4	+	0	0	$\Delta G_{pH}(1)$
5	0	+	+	$\Delta G_{pH}(2) + \Delta G_{pH}(3) + (2=3)$
6	0	+	0	$\Delta G_{pH}(2)$
7	0	0	+	$\Delta G_{pH}(3)$
8	0	0	0	0

Note that: + is charged, 0 is uncharged. Energy is relative to state 8. (X=Y) indicates the interaction energy between the charge form of group X and Y.  $\Delta G_{pH}(X)$  is the free energy difference between the charged and uncharged forms of group X at a fixed pH values.

The Henderson-Hasselbalch equation (Eq. 1.1) can be used to extract a  $pK_a$  value for the group that has only a simple titration curve. The case of more complicated titration curves such as the titration curve containing tight association of more than one titratable group, cannot be described by the Henderson-Hasselbalch equation. The methods for calculating the pH dependence of the free energy for each protonation state have been considered by calculating the intrinsic  $pK_a$  values and pairwise interaction energies (electrostatic interaction energies) and this allows us to extract the  $pK_a$  values from its titration curves. The intrinsic  $pK_a$  values and pairwise interaction energies can be calculated using the information from its three-dimensional structure. Electrostatic interaction energies can be calculated from protein structures by solving the Poisson-Boltzmann equation (PBE) that can be applied to most other  $pK_a$  calculation methods (Baker *et al.*, 2001; Madura *et al.*, 1995; Nicholls and Honig, 1991). In the PBE calculation scheme (Yang *et al.*, 1993) it is possible to calculate three types of energies: desolvation energies, background interaction energies, and site-site interaction energies that are required in  $pK_a$  calculation. Namely, desolvation energy and background interaction energy are pH-independent interactions, whereas the electrostatic interaction between titratable groups (charge-charge interaction) is pH-dependent interaction. The desolvation energy and background interaction energy describe the protein environment that influences the  $pK_a$  values of each titratable residue when disregarding the titration of all other titratable groups in the enzyme. The model  $pK_a$  value (the  $pK_a$  value of the residue in water) (Nielsen and Vriend, 2001) integrated with the  $pK_a$  differences of desolvation energy and background interaction energy is called the intrinsic  $pK_a$  which is the  $pK_a$  values when all other titratable groups in the protein are fixed in their neutral state. This  $pK_a$  can be calculated based on the

thermodynamic cycle as depicted in Figure 1.15. The intrinsic  $pK_a$  values of each residue integrated with all charge-charge interaction energies can be used to calculate the energy of every possible protonation state of the protein at a particular pH.

The input for a PBE solver consisted of a protein structure expressed as charges and a set of calculation parameters. The preparation of a protein structure shows a large effect on the electrostatic energies from the PBE run, and therefore results in a large effect on the calculated  $pK_a$  values.



**Figure 1.15** The thermodynamic cycle for the transfer of a titratable group from solvent to a protein environment.  $pK_a$  (model) is the model  $pK_a$  value of the group in solvent which is typically water.  $pK_a$  (protein) is the  $pK_a$  value of the group in the protein, disregarding the effects from other titratable groups. The  $\Delta G_{\text{charged}}$  and  $\Delta G_{\text{neutral}}$  are the energies associated with transferring the charged and neutral form from water into the protein. The net free energy,  $\Delta G_1 + \Delta G_4 - \Delta G_3 - \Delta G_2$  is equal to zero.

### 1.7.2 Running a $pK_a$ calculation

The  $pK_a$  calculation is run by choosing a set of protein structures, examining the protein structures and adapting the protein structures to the experimental condition, for example, removal of all water molecules and all ions that are crystallization artifacts and modelling any missing atoms in the X-ray structures. The PBE-based methods for calculating the  $pK_a$  values have been developed for modelling a change in protein structure by optimizing the hydrogen bond network for each protonation state (Alexov and Gunner, 1997; Nielsen and Vriend, 2001). These methods make the protein respond to changes in the protonation state by rearranging dipoles to minimize the energy of each protein structure. In general, optimization or relaxation of the hydrogen bonding network gives improved results, especially for buried titratable sites that are typically involved with a movement of hydrogen bond-mediated dipoles. However, the improvement obtained by hydrogen bond optimization has less effect for the surface residues since the surface residues themselves typically change their conformation more than buried titratable sites as a deprotonation reaction occurs. The buried titratable sites are kept fixed in the protein interior and a change in protein environment only comes from reorientation of hydrogen bond dipoles whereas the surface residues can move more freely and change the rotamer states (Alexov and Gunner, 1999; Georgescu *et al.*, 2002).

### 1.7.3 Extracting $pK_a$ values from calculated titration curves

The calculated  $pK_a$  values have been derived from calculated titration curves. A protein structure including charges and radii and a set of calculation parameters is converted into titration curves in a standard  $pK_a$  calculation algorithm.

The extraction of  $pK_a$  values from titration curves can be performed with most methods assuming that the  $pK_a$  value of a titration curve is the  $pK^{1/2}$  value which is the pH value at which the titratable group is half-protonated (Nielsen, 2007). This method has been used in many  $pK_a$  calculation methods because the non-Henderson-Hasselbalch titration curves observed with standard  $pK_a$  calculation methods prevent the extraction of classical  $pK_a$  values from the Henderson-Hasselbalch equation and so use of  $pK^{1/2}$  values is better to define a  $pK_a$  value from non-Henderson-Hasselbalch titration curves. When making a complete set of  $pK_a$  calculations, the titration curves for all residues that one is interested in should be checked manually to identify titration curves that cannot be described by a single  $pK_a$  value, using a freely available graphical tool such as pKaTool (Nielsen, 2007) that allows easy interpretation of calculated  $pK_a$  values at the click of a button or web-based solutions such as the H++ server (Gordon *et al.*, 2005).

#### 1.7.4 Calculating pH activity profiles

The  $pK_a$  values in the active site of an enzyme are calculated in order to predict the pH activity profile. Therefore, it is necessary to identify the catalytically competent protonation state (CCPS) for the enzyme before predicting the pH activity profile. Two or more titratable groups are typically required to be in a specific protonation state to provide the protons or charges for the enzyme to be active in the catalytic mechanism. Hen egg white lysozyme is a good example of this in which the catalytic acid/base Glu35 donates a proton to the glycosidic bond whereas the catalytic nucleophile Asp52 attacks at the C1 of the sugar substrate (Phillips, 1967).

When the CCPS has been identified, the pH activity profile can be then predicted simply by extracting the pH-dependent population of the CCPS. It should be mentioned that only in the case of perfect Henderson-Hasselbalch titration curves of the catalytic groups is it possible to correctly predict the pH activity profile from the calculated  $pK_a$  values. Many enzymes are known to have pH-activity profiles with shoulders or tails, and it is therefore clear that at least two titratable groups in the active site are generally important to be in a catalytic protonation state for enzyme catalysis.

### 1.7.5 Accuracy of protein $pK_a$ calculation methods

A calculated  $pK_a$  value depends on the X-ray structure, the method used in  $pK_a$  calculation and the residues of interest (Nielsen, 2007; Nielsen and McCammon, 2003b). The accuracy of a protein  $pK_a$  calculation result is evaluated by the root mean square deviation (RMSD) between the computational and experimental  $pK_a$  values of all protein titratable groups. The performance of individual  $pK_a$  calculation packages can vary from the average value. Most  $pK_a$  calculation algorithms have been developed to give low RMSD values for a set of calculated  $pK_a$  values, compared with experimentally measured  $pK_a$  values (van Vlijmen *et al.*, 1998; Nielsen and Vriend, 2001; Toseland *et al.*, 2006).

Presently, the best  $pK_a$  calculation algorithms reach an accuracy within  $\pm 0.5$   $pK_a$  units when benchmarked with experimentally measured  $pK_a$  values (Nielsen, 2007). WHAT IF PBE-based  $pK_a$  calculation package (WI $pK_a$ ) is comparable in accuracy to other  $pK_a$  calculation packages and has been benchmarked extensively to assess its sensitivity to structural errors (Nielsen and McCammon, 2003b). WI $pK_a$  has been developed by optimizing the hydrogen bond network for each protonation state

used with the PBE solver in the  $pK_a$  calculations (Alexov and Gunner, 1997; Nielsen and Vriend, 2001). This method allows the protein to respond to changes in the protonation state by rearranging dipoles to minimize the energy of each protein structure. This global optimization gives significantly improved results, especially for buried residues that are typically involved with the movement of hydrogen bond-mediated dipoles (Nielsen and Vriend, 2001).

### 1.7 Studies of *Vibrio harveyi* GlcNAcases

The complete genome sequence of *Vibrio harveyi* encodes two GlcNAcases (*VhNag1* and *VhNag2*) that belong to the new members of family 20 glycoside hydrolases. The full-length DNAs of these GlcNAcases were successfully cloned into the pQE-60 expression vector, which provides high-level expression in *E. coli* M15 host cells as a C-terminally His<sub>6</sub>-tagged polypeptide. *VhNag1* has a molecular mass of 89 kDa and an optimum pH of 7.5, whereas *VhNag2* has a molecular mass of 73 kDa and an optimum pH of 7.0. When specific hydrolytic activity was assayed with various substrates, *pNP*-GlcNAc was found to be the most effective. Both GlcNAcases could hydrolyse all the natural substrates, *VhNag2* being ten-fold more active than *VhNag1*. A time course of chitin oligosaccharide hydrolysis by TLC and quantitative HPLC showed that *VhNag2* acts exolytically for chitin degradation in a sequential manner, yielding GlcNAc as the end product and chitotetraose was found to be the best substrate for *VhNag2* due to its highest activity as compared with the other natural glycosides. Kinetic modeling of the enzymic reaction suggested that binding at subsite (-2) and (+4) had unfavourable binding free energy changes and that the active site of *VhNag2* comprises four GlcNAc binding subsites, designated (-1), (+1), (+2), and (+3). In living

cells, these intracellular GlcNAcases may work after endolytic chitinases to complete chitin degradation and could potentially serve as biocatalysts in the production of chitin derivatives during the recycling of chitin biomass (Suginta *et al.*, 2010).

## 1.8 Research objectives

In marine ecosystems, chitin biomaterials are initially hydrolyzed to short-chain oligosaccharides by chitinases, hydrolytic enzymes that are secreted mainly by marine bacteria such as *V. harveyi*, which has been reported to express high levels of chitinases and GlcNAcases in order to efficiently utilize chitin biomaterials as its sole source of energy. In the chitin-degradation pathway of *V. harveyi*, chitinases initially degrade chitin to small chitooligosaccharide fragments, which are further transported across the bacterial cell wall through the chitooligosaccharide-uptake channel, which is known as chitoporin. In the periplasm, GlcNAcases (EC 3.2.1.52) degrade chitin oligosaccharides to GlcNAc monomers, which are transported across the inner membrane by a specific ABC transporter into the cytoplasm and are then further metabolized to metabolic intermediates that can readily be converted to carbon and nitrogen sources for cells. However, the human homologues of bacterial GlcNAcases, known as HexA and HexB, are clinically important since they are critically involved in the degradation of glycosphingolipids, which are deposited in the form of gangliosides  $G_{M1}$ ,  $G_{M2}$  and  $G_{M3}$  on the plasma membranes of nerve cells. Mutations of these enzymes lead to the accumulation of  $G_{M2}$  gangliosides, which results in fatal diseases such as Tay-Sachs and Sandhoff diseases.

The effects of exogenous anions, such as azide or formate, can provide direct evidence identifying the catalytic acid/base residues in retaining glycoside hydrolases.

Following mutation of the acid-base residue or the nucleophilic residue, hydrolytic activity of inactive mutants can be rescued by the addition of an exogenous nucleophile, such as azide or formate. However, the effects of external nucleophiles on the hydrolytic activity of inactive mutant *VhGlcNAcases* have not been thoroughly investigated yet. The first part of this research aims to investigate the effects of mutations on the enzymatic activity of *VhGlcNAcase* and probe the catalytic residues of *VhGlcNAcases* by chemical rescue to support the proposal of a substrate-assisted mechanism of GH20 GlcNAcases.

Chitinase A from the marine bacterium *V. harveyi* was previously reported for functional and structural characterization. The crystal structures of *V. harveyi* chitinase A and its catalytically inactive mutant (E315M) in the presence and absence of substrates give an insight in catalytic mechanism and provide evidence that the interacting sugars undergo conformational changes prior to hydrolysis. However, a crystal structure of  $\beta$ -*N*-acetylglucosaminidase from any marine bacterium has not been reported yet. Therefore, in the second part of this research, we firstly determine the molecular weight of *VhGlcNAcase* in the native state and then solve the 3D-structure of GlcNAcases with and without the natural substrate as well as determining the kinetic properties of *VhGlcNAcase* and its mutants.

The  $pK_a$  values of the active-site residues in an enzyme are important to the functionality of the catalytic mechanisms, enzyme activity, pH-dependent conformational changes and protein stability. The pH-activity profile of chitinases is typically bell-shaped and that is determined by the  $pK_a$  values of the acidic active-site residues. The computational algorithm in defining the  $pK_a$  values in the active site is of great interest because it can predict the pH activity profile and supports the catalytic

mechanism. Therefore, the third part of this research aims to calculate the active site  $pK_a$  values of the GH18 chitinases and confirm the catalytic mechanism based on the  $pK_a$  calculation as well as predict the pH-activity profiles of chitinases and compare these to experimentally measured pH-activity profiles.

The objectives of this research include:

1. To express and purify the wild-type and *VhGlcNAcase* mutants.
2. To investigate effects of mutations on the enzymatic activity of *VhGlcNAcase*.
3. To probe the catalytic residues in the catalytic cycle of *VhGlcNAcase* using the chemical rescue assay for supporting the proposal of a substrate-assisted mechanism of GH20 GlcNAcases.
4. To determine the molecular weight of *VhGlcNAcase* in the native state.
5. To solve the 3D-structure of GlcNAcase inactive mutants with and without the natural substrate and understand the mechanism of enzyme catalysis.
6. To determine the kinetic properties of *VhGlcNAcase* and its mutants.
7. To calculate the active site  $pK_a$  values of the GH18 chitinases and confirm the catalytic mechanism based on the  $pK_a$  calculation.
8. To predict pH-activity profiles of chitinases and compare these to experimentally measured pH-activity profiles.

## CHAPTER II

### MATERIALS AND METHODS

#### 2.1 Chemicals and reagents

##### 2.1.1 Bacterial strains and expression plasmids

*Escherichia coli* type strain DH5 $\alpha$  was used for cloning, subcloning and plasmid preparation. Supercompetent *E. coli* XL1Blue (Stratagene, La Jolla, CA, USA) was the host strain for the production of mutagenized plasmid. *E. coli* strain M15 (pREP) host cells (Qiagen, Valencia, CA, USA) and the recombinant plasmid of pQE-60 vector containing *GlcNAcase* gene fragments were used for a high-level expression of recombinant enzyme (Suginta *et al.*, 2010).

##### 2.1.2 Site-directed mutagenesis and plasmid purification

Chemicals and reagents used for site-directed mutagenesis were of molecular biology grade. The mutagenic primers used for site-directed mutagenesis were ordered from BioDesign (Bangkok, Thailand) and Bio Basic Canada (Ontario, Canada). *Pfu* DNA polymerase, *Pfu* polymerase 10x buffer, dNTP mix, *DpnI*, *BamHI* and *BglII* restriction enzymes, were purchased from Promega (Madison, WI, USA). High-Speed Plasmid Mini Kit was product of Geneid (Taipei, Taiwan), QIAquick Gel Extraction Kit of QIAGEN (QIAGEN GmbH, Hilden, Germany) and QuickClean II Plasmid Miniprep Kits of GenScript (GenScript, Piscataway, New Jersey, USA). SYBR<sup>®</sup> Green I nucleic acid gel stain was bought from Sigma-Aldrich (St. Louis, MO,

USA) and HyperLadder<sup>TM</sup> I DNA marker from Bioline (Tacerton MA, USA).

### 2.1.3 Protein expression and purification

Chemicals and reagents used for protein expression and purification were of analytical grade. Glacial acetic acid, hydrochloric acid, methanol, n-butanol, phosphoric acid, potassium hydroxide, sodium chloride, sodium hydroxide, sodium dihydrogen phosphate and disodium hydrogen phosphate were obtained from Carlo ERBA (Rodano, Milano, Italy). Acrylamide, tris (hydroxymethyl)amine, glycerol, ammonium persulfate, 2-mercaptoethanol, bromophenol blue, coomassie blue R250, coomassie blue G250, *N, N', N'', N'''*-tetramethylethylenediamine (TEMED), bis-*N, N''*-methylenebisacrylamide, tris(2-carboxyethyl)phosphine (TCEP), ethylenediamine tetra-acetic acid (EDTA), calcium chloride, magnesium chloride, sodium dodecyl sulfate (SDS) were products of Sigma-Aldrich (St. Louis, MO, USA). Glycine was from Vivantis (Oceanside, CA, USA). Ampicillin, kanamycin, phenyl methylsulfonyl fluoride (PMSF), imidazole, hen egg white lysozyme, Triton X-100 were product of USB Corporation (Cleveland, OH, USA). DNase I was from Bio basic (Markham, Ontario, Canada). Isopropyl thio- $\beta$ -D-galactoside (IPTG) was from Merck Millipore (Billerica, MA, USA). MF-Millipore Membrane Filters (0.22  $\mu$ m and 0.45  $\mu$ m pore size) were purchased from Millipore Corporation (Beverly, MA, USA). Vivaspin-20 ultrafiltration membrane concentrators (10 kDa molecular-weight cutoff) were product of Vivascience (Hanover, Germany). Ni-nitrilotriacetic acid (Ni-NTA) agarose resin was a product of QIAGEN. The TALON<sup>®</sup> Superflow<sup>TM</sup> metal affinity resin was the product of Clontech Laboratories, Inc., USA. Disposable columns for Ni/Co agarose chromatography (1.0  $\times$  10 cm) from Biorad, (Bio-Rad Laboratories, Inc., CA, USA).

HisTALON<sup>®</sup> pre-packed columns (1 × 5 ml) were purchased from Clontech Laboratories, Inc., USA. HiPrep 16/60 Sephacryl S-200 HR and HiPrep 26/60 Sephacryl S-300 HR columns were products of GE Healthcare (Munich, Germany) and BCA protein assay kit was a product of EMD Chemicals (San Diego, CA, USA).

#### 2.1.4 Enzyme kinetics

All chemicals and reagents were obtained from the following sources: *p*-nitrophenol (*p*NP) and *p*-nitrophenyl-*N*-acetyl-glucosaminide (*p*NP-GlcNAc) were purchased from Sigma-Aldrich (St. Louis, MO, USA); sodium dihydrogen phosphate, disodium hydrogen phosphate, citric acid, sodium carbonate, sodium nitrate, sodium formate and sodium chloride were purchased from Carlo Erba (Rodano, Milano, Italy); sodium azide was purchased from LabChem Inc. (Zelienople, PA, USA) and a 96-well microtiter plate was purchased from Nunc (Roskilde, Denmark).

#### 2.1.5 Protein crystallization

Commercially crystal screens: the sparse-matrix screens, the JCSG Core Suites I, II, III and IV, the Classics and Classics II Suites and the PACT Suite and the grid screens the PEGs Suite and the Anions Suite were from Qiagen, Hilden, Germany. The 20 mm siliconized glass cover, 0.02-0.5 mm CryoLoops<sup>™</sup>, CrystalCap Magnetic<sup>™</sup> with vial, Magnetic Crystal Wand<sup>™</sup>, CryoCane, CryoSleeve, Vial Clamp<sup>™</sup>-Curved, Long CryoTongs<sup>™</sup>, CrystalWand<sup>™</sup>, Crystal clear sealing tape and Glass Dewar were products of Hampton Research (Hampton Research Corp., CA, USA). The Baysilone-Paste<sup>®</sup> vacuum grease was the product of Carl Roth (Carl Roth GmbH & Co. KG, Karlsruhe, Germany). The 96-well CrystalQuick plates were from

Greiner Bio-One, Frickenhausen, Germany. Linbro Plates (24-well tissue culture plates) with cover were from Linbro<sup>®</sup> (Linbro Division, Flow laboratories Inc., CA, USA).

The natural substrate, *N, N'*-diacetylchitobiose (GlcNAc<sub>2</sub>) were purchased from Dextra (Dextra Laboratories Ltd., West Berkshire, United Kingdom). PEG3350, PEG 6000, MES (2-(*N*-morpholino)ethanesulfonic acid), sodium/potassium phosphate and sodium acetate were a product of USB Corporation. Tris (hydroxymethyl)amine hydrochloride, potassium sodium tartrate, sodium malonate dibasic, 1,3-Bis[tris(hydroxymethyl)methylamino]propane (Bis-Tris propane), 2, 2-Bis(hydroxyl methyl)-2, 2', 2''-nitrilotriethanol (Bis-Tris), were purchased from Sigma-Aldrich.

## 2.2 Analytical programs

An amino acid sequence alignment of GlcNAcases and chitinases were constructed using the CLUSTALW algorithm in a GCG package (Thompson *et al.*, 1994) and displayed in the DNASTAR package (DNASTAR, Inc., Madison, WI, USA) and the CLC Main Workbench (CLC Bio, Aarhus, Denmark). The structure-based alignment was further generated using the program ESPript, v3.0 (Robert and Gouet, 2014). The crystallographic programs were as follows. The iMosflm (Battye *et al.*, 2011) and XDS (Kabsch, 2010) programs were used for data indexing and scaling. The PHASER program (McCoy *et al.*, 2007) was used for molecular replacement calculations, the COOT program (Emsley and Cowtan, 2004) for model building, PRODRG servers (Schüttelkopf and van Aalten, 2004) for topology of ligand, REFMAC5 in the CCP4 suit (Murshudov *et al.*, 1997) for refinement, PROCHECK (Laskowski *et al.*, 1993) and MolProbity (Chen *et al.*, 2010) for validation of the

stereochemistry of the final model, LigPlot+ (Laskowski and Swindells, 2011) for two-dimensional diagram of ligand-protein interactions and PYMOL ([www.pymol.org](http://www.pymol.org)) for 3D visualization of protein. For the  $pK_a$  calculation, WHAT IF (Vriend, 1990) was used for rebuilding missing atoms and correcting nomenclature problems, WHAT IF  $pK_a$  calculation package (WIpKa) (Nielsen and Vriend, 2001) for  $pK_a$  calculation, pKaTool program (Nielsen, 2007) for titration curve evaluation and g3data program for extracting the graphical data points.

### 2.3 Instrumentation

The instruments required for site-directed mutagenesis, protein expression, purification and functional characterization are located at the Biochemistry-Electrochemistry Research Unit Laboratory and the Center for Scientific and Technological Equipment, Suranaree University of Technology, Nakhon Ratchasima, Thailand. These instruments included a Sonopuls Ultrasonic homogenizer with a 6-mm diameter probe, a Mastercycler® personal PCR thermocycler (Eppendorf AG., Hamburg, Germany), a DNA gel apparatus (Pharmacia Biotech, SF, USA), a Jenway UV-VIS spectrophotometer (Bibby Scientific Ltd., Staffordshire, UK), a Gel-Doc 2000 Gel document system (Bio-Rad Laboratories, CA, USA), a Mini-PROTEAN® 3 Cell (Bio-Rad, Hercules, CA, USA), a shaking incubator (MRC, Holon, Israel), a microcentrifuge Denville 26OD (Denville Scientific, Metuchen, NJ, USA), a high-speed microcentrifuge CF16RX II (Hitachi, Tokyo, Japan), an ÄKTA purifier system (Amersham Bioscience, Piscataway, NJ, USA), a Thermomixer comfort (Eppendorf AG, Hamburg, Germany) and a Biochrom Anthos Multiread 400 Microplate Reader (Biochrom, Cambridge, UK).

The instruments used for protein expression, purification, crystallization and structural characterization located at the Max-Planck Institute (MPI) of Molecular Physiology, Dortmund, Germany were a HC-2000 microfluidizer (Microfluidics, Lampertheim, Germany), ÄKTA purifier system (GE Healthcare), a Mosquito Crystallization robot (TTP Labtech, Melbourn, England), a Zeiss stemi 200-C stereo microscope (Carl Zeiss MicroImaging GmbH., Göttingen, Germany), mounted with a color video camera (Sony Corp., Tokyo, Japan), a Rigaku Micromax HF-007 rotating anode generator and two MAR Research image plate detectors on a Nonius FR-591 generator (Rigaku Corp., Berlin, Germany). For the instruments used for data collections at the beamline PX-II of Swiss Light Source located in Villigen, Switzerland was a MAR Research CCD detector.

## 2.4 Sequence analysis and homology modeling of *Vh*GlcNAcases

The amino acid sequence of the matured *Vh*GlcNAcase was submitted to Swiss-Model (<http://swissmodel.expasy.org/>) for the tertiary structure prediction using the crystal structure of *S. marcescens* chitobiase (PDB code: 1QBA) as a structural template. To obtain detailed information about the enzyme's active site, the modelled structure of *Vh*GlcNAcase was superimposed on the 3D structure of *S. marcescens* chitobiase (*Sm*CHB) docked with GlcNAc<sub>2</sub> coordinates. The annotated structures were edited and displayed in PyMOL ([www.pymol.org](http://www.pymol.org)). The structure-based alignment was generated by aligning the amino acid sequence of *Vh*GlcNAcase with five GH20 GlcNAcases with known 3D-structures, including *S. marcescens* chitobiase, *Sm*CHB (PDB code: 1QBA); *Streptomyces plicatus*  $\beta$ -*N*-acetylhexosaminidase, *Sp*Hex (PDB code: 1HP4); *Paenibacillus* sp.  $\beta$ -hexosaminidase, *Ps*Hex1T (PDB code: 3GH4);

human  $\beta$ -hexosaminidase A ( $\alpha$ -chain), *HsHexA* (PDB code: 2GJX) and human  $\beta$ -hexosaminidase B ( $\beta$ -chain), *HsHexB* (PDB code: 1NOU). The amino acid sequence alignment was carried out in ClustalW, and the structure-based alignment was further generated using the program ESPript, v3.0 (Robert and Gouet, 2014).

## 2.5 Mutational design and site-directed mutagenesis of *VhGlcNAcases*

The full-length *GlcNAcase* cDNA (accession No. HM175716) was isolated from the genome of *V. harveyi* by the PCR technique, and the gene encoding *VhGlcNAcase* (amino acids 5-642, lacking the signal peptide) was then cloned into the pQE-60 expression vector (Qiagen, Valencia, California, USA), generating a four-amino-acid (MGGS) cloning artefact at the *N*-terminus as described previously (Suginta *et al.*, 2010). The recombinant protein was shown to be expressed at a high level in *E. coli* M15 (pREP4) cells as a *C*-terminally His<sub>6</sub>-tagged polypeptide (Suginta *et al.*, 2010). The pQE-60 expression vector harboring the full length *VhGlcNAcase* was used as DNA template for point mutations. The mutagenic primers were synthesized by commercial sources (BioDesign Co., Ltd Bangkok, Thailand and Bio Basic Canada Inc., Ontario, Canada) and the oligonucleotide sequences of these primers are listed in Table 2.1. The underlined sequences represent the mutated codon.

**Table 2.1** Primers used for site-directed mutagenesis.

<b>Mutation</b>	<b>Oligonucleotide sequence<sup>a</sup></b>
D303N	forward 5'- CATTGGCATCTCACT <u>AAC</u> GATGAAGGCTGGCGTG -3' reverse 5'- CACGCCAGCCTTCATC <u>GTT</u> AGTGAGATGCCAATG -3'
D303A	forward 5'- CATTGGCATCTCACT <u>GCG</u> GATGAAGGCTGGCGTG -3' reverse 5'- CACGCCAGCCTTCATC <u>CGC</u> AGTGAGATGCCAATG -3'
D304N	forward 5'- GCATCTCACTGACA <u>ACG</u> AAGGCTGGCGTGTC -3' reverse 5'- GACACGCCAGCCTTC <u>GTT</u> GTCAGTGAGATGC -3'
D304A	forward 5'- GCATCTCACTGAC <u>GCG</u> GAAGGCTGGCGTGTC -3' reverse 5'- GACACGCCAGCCTTC <u>CGC</u> GTCAGTGAGATGC -3'
H373A	forward 5'- GAAATTGATGTACCTGGT <u>GCG</u> TGCCGCGCCGCAATTAAG -3' reverse 5'- CTTAATTGCGGCGCGGC <u>ACG</u> CACCAGGTACATCAATTTTC -3'
D437N	forward 5'- GTTCACATTGGCGCGA <u>ACG</u> AAGTGCCTAACGGC -3' reverse 5'- GCCGTTAGGCACTT <u>CGT</u> TCGCGCCAATGTGAAC -3'
D437A	forward 5'- GTTCACATTGGCGCG <u>GCG</u> GAAGTGCCTAACGGC -3' reverse 5'- GCCGTTAGGCACTT <u>CGC</u> CGCGCCAATGTGAAC -3'
E438Q	forward 5'- GTTCACATTGGCGCGG <u>ACC</u> AGTGCCTAACGGCGTGTG -3' reverse 5'- CACACGCCGTTAGGCAC <u>CTG</u> GTCCGCGCCAATGTGAAC -3'
E438D	forward 5'- CATTGGCGCGGAC <u>GAT</u> GTGCCTAACGGCGTG -3' reverse 5'- CACGCCGTTAGGCAC <u>ATC</u> GTCCGCGCCAATG -3'
E438A	forward 5'- CACATTGGCGCGGAC <u>GCG</u> GTGCCTAACGGCGTGTG -3' reverse 5'- CACACGCCGTTAGGCAC <u>CGC</u> GTCCGCGCCAATGTG -3'
D532A	forward 5'- CAAACTACTTATTTG <u>GCG</u> GATGACCCAAGACTACGC -3' reverse 5'- GCGTAGTCTTGGGTCATC <u>GCC</u> AAATAAGTAGTTTG -3'
E584A	forward 5'- CCGCTCTATGGTG <u>GCG</u> GATCATCAACAACCC -3' reverse 5'- GGGTTGTTGATGATC <u>GCG</u> CACCATAGAGCGG -3'

<sup>a</sup> Sequences underlined indicate the mutated codons.

The mutants D303A, D304A, D437N, D437A, E438Q, E438D, E438A, D532A and E584A of the *VhGlcNAcase* were made using the QuickChange Site-Directed Mutagenesis Kit (Stratagene, California, USA), following the Manufacturer's instruction. The PCR reaction mix and PCR conditions used for site-directed mutagenesis are shown in Table 2.2 and Table 2.3, respectively. After gene amplification, the PCR products were verified by agarose gel electrophoresis. The mutated DNA was treated with DpnI for 1 hour at 37°C and then transformed into *E. coli* XL1-Blue competent cells. Afterwards, the transformed cells were spread on LB-agar containing 100 µg/mL ampicillin. The mutated plasmids obtained from positive clones were extracted using QuickClean II Plasmid Miniprep Kits (GenScript, Piscataway, New Jersey, USA) and the correct insertion of the mutated sequences was verified by automated DNA sequencing in both directions from First BASE Laboratories Sdn Bhd, Seri Kembangan, Malaysia. The DNA sequencing chromatograms were displayed using the program Chromas Lite, Technelysium Pty Ltd, South Brisbane, Australia and the nucleotide sequence analyses were obtained from the DNASTAR package (DNASTAR, Inc., Madison, USA).

**Table 2.2** The PCR reaction used for site-directed mutagenesis

Reagent	Volume ( $\mu\text{L}$ )
DNA template (50 ng/ $\mu\text{L}$ )	1.0
<i>Pfu</i> DNA polymerase (2.5 U/ $\mu\text{L}$ )	1.0
10X reaction buffer	2.5
Oligonucleotide forward primer (10 $\mu\text{M}$ )	2.5
Oligonucleotide reverse primer (10 $\mu\text{M}$ )	2.5
dNTP mix (2 mM)	2.5
Milli-Q water	13
Total volume	25

**Table 2.3** The PCR running conditions used for site-directed mutagenesis

PCR step	Cycle	Temperature	Time
Hot start	1	98°C	5 sec
Denaturing start	1	98°C	3 min
Denaturing		98°C	30 sec
Annealing	30	55°C	1 min
Extension		65°C	12 min
Final extension	1	65°C	10 min

## 2.6 Expression and purification of *VhGlcNAcase* and its mutants

In this study, the expression and purification of recombinant wild-type *VhGlcNAcase* and its mutants were optimized in order to improve the yield and the purity of the proteins for crystallization purposes. The recombinant wild-type *VhGlcNAcase* was expressed in *E. coli* M15 (pREP) cells as a 652-amino acid polypeptide, including the C-terminal (His)<sub>6</sub> sequence (Suginta *et al.*, 2010). Expression of all *GlcNAcase* variants was based on the protocol described previously (Meekrathok *et al.*, 2015). Briefly, the transformed cells harbouring the recombinant

pQE-60/GlcNAcase plasmid were grown at 37°C in Terrific Broth (TB) containing 100 µg/mL ampicillin and 25 µg/mL kanamycin until the cell density reached an OD<sub>600</sub> of 0.6. The cell culture was cooled to 20°C, before isopropyl thio-β-D-galactoside (IPTG) was added to a final concentration of 0.4 mM for GlcNAcase expression. Cell growth was continued at 20°C for an additional 18 hours, and cells were harvested by centrifugation at 4,500 rpm for 30 min. The bacterial pellet was resuspended in lysis buffer (20 mM Tris-HCl buffer, pH 8.0, 150 mM NaCl, 1 mM phenylmethylsulphonyl fluoride (PMSF), 5% (v/v) glycerol and 1 mg/mL lysozyme and 1 mg/mL DNase I), and then lysed on ice using a Sonopuls ultrasonic homogenizer with a 6-mm diameter probe (50% duty cycle; amplitude setting, 30%; total time, 30 s, 6-8 times). Unbroken cells and cell debris were removed by centrifugation at 13,000 rpm for 1 hour. The supernatant containing *Vh*GlcNAcase was immediately applied to a polypyrine column packed with 5 mL of TALON<sup>®</sup> Superflow<sup>™</sup> metal affinity resin (Clontech Laboratories, Inc., USA) operated at 4°C with gravity-dependent flow. The column was washed with 8 column volumes (cv) of equilibration buffer (20 mM Tris-HCl buffer, pH 8.0), followed by 7 cv of the equilibration buffer containing 10 mM imidazole. The protein was then eluted with 250 mM imidazole in the same buffer. Eluted fractions of 10 mL were collected and 15 µL of each fraction was further analyzed by 12% SDS-PAGE, according to the method of Laemmli (Laemmli, 1970), to confirm the purity of the protein. Fractions with GlcNAcase activity were pooled, exchanged into 20 mM Tris-HCl buffer pH 8.0 containing 150 mM NaCl buffer, and subjected to several rounds of centrifugation in Vivaspın-20 ultrafiltration membrane concentrators (10 kDa molecular-weight cutoff, Vivascience AG, Hannover, Germany) for complete removal of imidazole. For structural characterization, the GlcNAcases were further purified on

a HiPrep 16/60 Sephacryl S-200 prepacked column (Amersham Bioscience, Piscataway, NJ, USA) connected to an ÄKTAprime purification system (Amersham Bioscience, Piscataway, New Jersey, USA.) The chromatography column was operated at a flow rate of 0.2 mL/min, and 1.4 mL fractions were collected. Note that, the running buffer for the wild-type protein was low-salt equilibration buffer (20 mM Tris-HCl buffer pH 8.0 containing 30 mM NaCl and 1 mM TCEP) to avoid interference by salt in the crystallization trials. However, some of the wild-type protein appeared as an aggregation peak in the gel-filtration profile. *VhGlcNAcase*-containing fractions were pooled and concentrated in the equilibration buffer (20 mM Tris-HCl buffer, pH 8.0, containing 150 mM NaCl) using the same type of Vivaspin membrane concentrator and then the purity was judged by 12% SDS-PAGE. Protein aggregation after purification was evaluated using native PAGE according to a protocol described elsewhere (Arndt *et al.*, 2012). The final concentration of *VhGlcNAcase* was determined by the  $A_{280}$  (Edelhoch, 1967; Gill and von Hippel, 1989; Pace *et al.*, 1995) and BCA (bicinchoninic acid) assay (Smith *et al.*, 1985) (Novagen<sup>®</sup> BCA protein assay kit, EMD Chemicals Inc., San Diego, CA, USA) using a standard calibration curve constructed from BSA (0-250  $\mu\text{g/ml}$ ). The freshly prepared protein was aliquoted and flash-frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until used.

## 2.7 Investigation of protein state by size-exclusion chromatography

The molecular mass (MW) of wild-type *VhGlcNAcase* in solution was investigated using size-exclusion chromatography. The well-defined globular protein standards ranging from 349 to 669,000 Da (GE healthcare, USA and Sigma-Aldrich Pte Ltd., Singapore) were employed for making a calibration curve of the HiPrep 26/60 Sephacryl S-300 prepacked columns (GE Healthcare Bio-Sciences AB, Uppsala, Sweden). The protein standards used in this experiment were  $N_\epsilon$ -DNP-L-lysine hydrochloride (349 Da), ribonuclease A (13,700 Da), ovalbumin (43,000 Da), bovine serum albumin (66,000 Da), aldolase (158,000 Da), ferritin (440,000 Da) and thyroglobulin (669,000 Da). The blue dextran 2000 was used to determine the void fraction in the column. Each standard protein including *VhGlcNAcase* (0.5-3  $\mu$ g) was mixed with  $N_\epsilon$ -DNP-L-lysine hydrochloride that acts as a control in equilibration buffer containing 20 mM Tris-HCl buffer, pH 8.0, containing 150 mM NaCl and then loaded to the HiPrep 26/60 Sephacryl S-300 prepacked columns connected with the ÄKTAprime system using the same equilibration buffer as a running buffer. The process was set up with a flow rate of 2.0 mL/min and 5.1 mL fractions were collected and assayed for GlcNAcase activity. The gel-phase distribution coefficient ( $K_d$ ) was calculated from the equation:  $K_d = (V_e - V_o)/V_s$  where  $V_e$  is elution volume,  $V_o$  is void volume (determined using Blue Dextran), and  $V_s$  is volume of stationary phase. The estimated molecular mass of *VhGlcNAcase* was determined from the calibration curve (plot of  $K_d$  versus log MW) once its  $K_d$  value is calculated from the measured elution volume.

## 2.8 GlcNAcase activity assay

GlcNAcase activity was determined by a colorimetric assay using *p*NP-GlcNAc (Sigma-Aldrich Pte Ltd., St. Louis, MO, USA) as substrate. The reaction mixture in a 96-well microtiter plate contained an optimal amount of *Vh*GlcNAcase (0.1  $\mu$ g for WT and 5-30  $\mu$ g for mutants), 500  $\mu$ M *p*NP-GlcNAc and 100 mM phosphate buffer, pH 7.0 in a total volume of 200  $\mu$ L. The assay was carried out at 37°C with constant agitation in an Eppendorf ThermoMixer<sup>®</sup> comfort (Eppendorf AG, Hamburg, Germany), and was terminated by adding 100  $\mu$ L of 3 M Na<sub>2</sub>CO<sub>3</sub> to each well after 10 min. The concentration of *p*-nitrophenol (*p*NP) released was determined at 405 nm in a Biochrom Anthos MultiRead 400 Microplate Reader (Biochrom, Cambridge, UK). The molar quantity of the liberated *p*NP was calculated from a calibration curve of *p*NP standard varying from 0 to 20 nmol. The hydrolytic activity of the enzyme was expressed as the quantity of *p*NP (nmol) produced in 1 min at 37°C.

## 2.9 Determination of the pH optima of *Vh*GlcNAcase WT and D437A mutant

To determine the activity/pH profile, the specific activity of *Vh*GlcNAcase WT and D437A mutant was determined in a discontinuous assay. The reaction mixture contained 0.05  $\mu$ g *Vh*GlcNAcase or 5  $\mu$ g D437A, 500  $\mu$ M *p*NP-GlcNAc, and McIlvaine's sodium phosphate-citric acid buffer, pH 3.0-9.0 (McIlvaine, 1921) at different pH values ranging from 3.0 to 9.0, in a total volume of 200  $\mu$ L. The reaction was carried out as described for the GlcNAcase activity assay.

## 2.10 Chemical rescue assay of *VhGlcNAcase* and its mutants

Sodium azide and sodium formate were initially tested for their ability to rescue the enzymic activity of *VhGlcNAcase* inactive mutants. A 200- $\mu$ L assay mixture, prepared in a 96-well microtiter plate, contained 500  $\mu$ M *p*NP-GlcNAc, 5  $\mu$ g of enzyme, 1 M sodium azide or formate, and 100 mM sodium phosphate buffer, pH 7.0. The reaction mixture was incubated at 37°C for 10 min with constant agitation, and the reaction was terminated by the addition of 100  $\mu$ L 3 M Na<sub>2</sub>CO<sub>3</sub>. The reaction of wild-type *VhGlcNAcase* was carried out as described for mutants, but with 0.1  $\mu$ g of the enzyme in the assay.

To determine the effect of concentration on the rescue activity, the reaction was incubated with different concentrations of azide or formate for a longer time. A 200- $\mu$ L assay mixture contained 500  $\mu$ M *p*NP-GlcNAc, 0.2  $\mu$ g of mutants and 0.1-2.0 M sodium azide or formate in 100 mM sodium phosphate buffer, pH 7.0. The reaction was allowed to proceed at 37°C for 60 min, and then terminated by the addition of 100  $\mu$ L of 3 M Na<sub>2</sub>CO<sub>3</sub>. The amount of *p*NP released was calculated as described above.

## 2.11 Time-course and kinetics of sodium formate effects on the activity of *VhGlcNAcase* D437A mutant

Chemical rescue of the D437A inactive mutant by sodium formate was further observed at different times of incubation. A 200- $\mu$ L assay mixture contained 500  $\mu$ M *p*NP-GlcNAc, 0.2  $\mu$ g of the mutant D437A and 0.1-2.0 M sodium formate in 100 mM sodium phosphate, pH 7.0. The reaction mixture was incubated at 37°C for times of 0, 2.5, 5, 10, 30, and 60 min. For kinetic experiments, a 200- $\mu$ L reaction mixture,

containing 0-500  $\mu\text{M}$  *p*NP-GlcNAc, 5  $\mu\text{g}$  of the mutant D437A, 0.1-2.0 M sodium formate and 100 mM sodium phosphate buffer, pH 7.0, was incubated for 10 min at 37°C and the reaction terminated with 100  $\mu\text{L}$  of 3 M  $\text{Na}_2\text{CO}_3$ . The amount of *p*NP formed during the reaction was estimated as described previously. The kinetic parameters (apparent  $K_m$ , apparent  $k_{\text{cat}}$  and apparent  $k_{\text{cat}}/K_m$ ) were determined by non-linear regression function available in GraphPad Prism v.5.0 (GraphPad Software Inc., San Diego, CA).

## 2.12 Steady-state kinetics of *Vh*GlcNAcase and its mutants

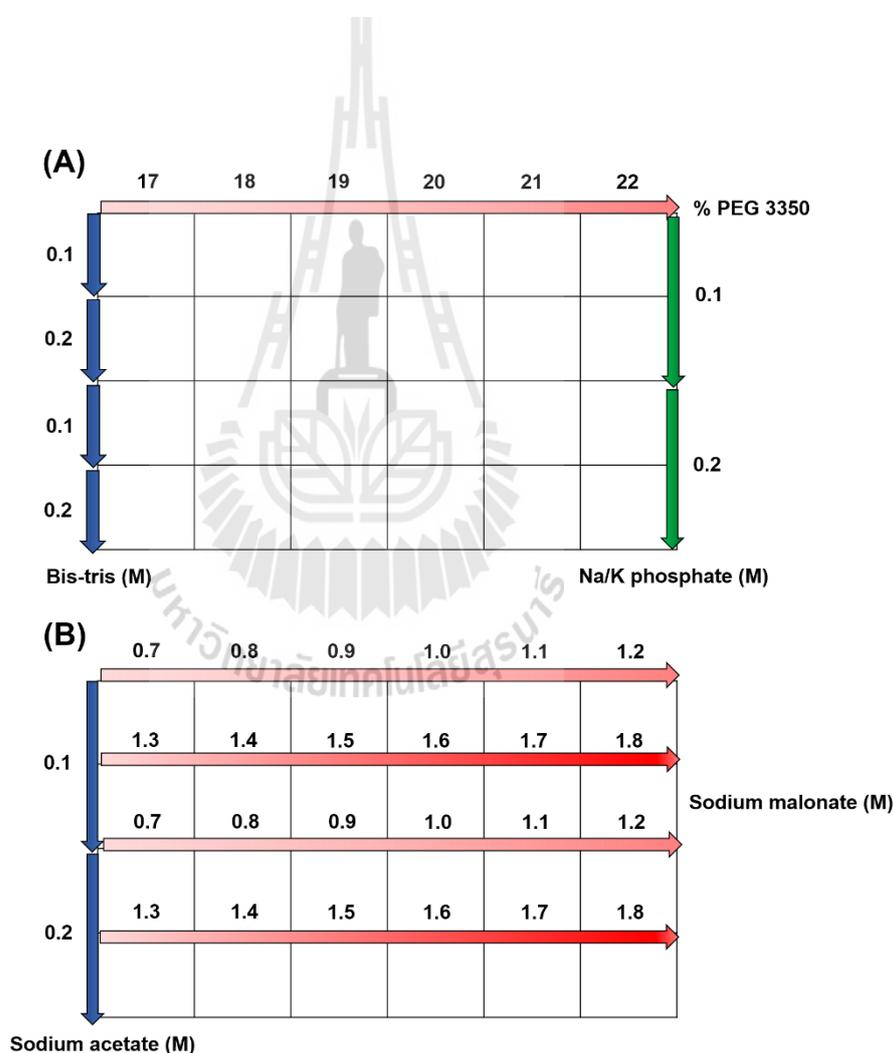
Kinetic studies of wild-type *Vh*GlcNAcase and its mutants were carried out by colorimetric assay in a microtiter plate reader as described above, with substrate concentrations varying from 0-500  $\mu\text{M}$ . Briefly, a 200- $\mu\text{L}$  reaction mixture containing 0-500  $\mu\text{M}$  *p*NP-GlcNAc, prepared in 100 mM sodium phosphate buffer, pH 7.0 and 0.1-30  $\mu\text{g}$  enzyme, was incubated at 37°C with constant shaking for 10 min. The enzymic reactions were then terminated with 100  $\mu\text{L}$  of 3 M  $\text{Na}_2\text{CO}_3$ . The amount of the reaction products was measured at 405 nm, then converted to molar quantities using a standard curve of *p*NP (0-20 nmol) as described previously. The kinetic parameters ( $K_m$ ,  $k_{\text{cat}}$  and  $k_{\text{cat}}/K_m$ ) were determined from triplicate assays of data by the nonlinear regression function of the GraphPad Prism software v.5.0 (GraphPad Software Inc., San Diego, CA).

## 2.13 Determination of three-dimensional structures of *VhGlcNAcases*

### 2.13.1 Crystallization of *VhGlcNACase* WT and the D437A mutant

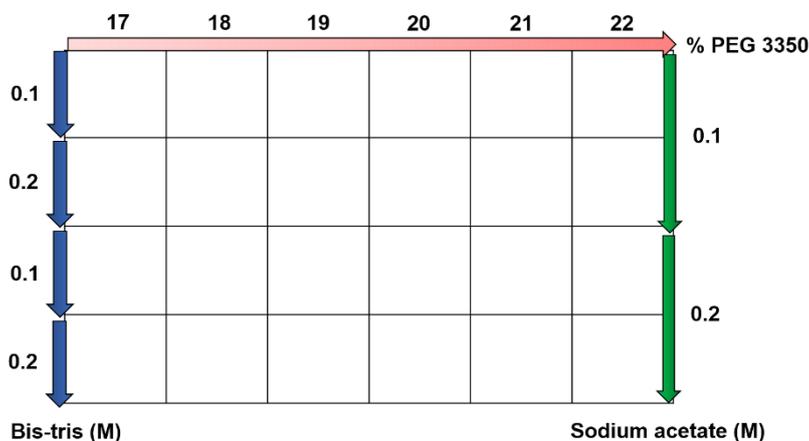
Initial crystallization screening of wild-type *VhGlcNAcase* was performed at 20°C using commercially available screens from Qiagen, Hilden, Germany: the sparse-matrix screens the JCSG Core Suites I, II, III and IV, the Classics and Classics II Suites and the PACT Suite and the grid screens the PEGs Suite and the Anions Suite. The screens were set up in 96-well CrystalQuick plates (Greiner Bio-One, Frickenhausen, Germany) using a Mosquito Crystallization robot (TTP Labtech, Melbourn, England) with the sitting-drop method. The purified enzyme was centrifuged at 13,000 rpm for 5 min and 0.1 µL of the freshly prepared wild-type protein at 10 mg/mL in equilibration buffer consisting of 20 mM Tris-HCl pH 8.0, 30 mM NaCl was then pipetted into 0.1 µL of each precipitant in the crystallization screens. Small plate-shaped crystals of wild-type *GlcNAcase* were observed within 1 day in condition F10 from the PACT Suite (20% (w/v) PEG 3350, 0.1 M Bis-Tris propane pH 6.5, 0.2 M sodium/potassium phosphate) and small crystals also appeared within 3 days in condition C6 of the Anions Suite (0.1 M sodium acetate pH 4.6, 1.2 M sodium malonate). Crystals obtained from both conditions were further optimized by the hanging-drop vapour-diffusion technique using different concentrations of two precipitants: (i) 17-22% (w/v) PEG 3350, 0.1-0.2 M Bis-Tris propane pH 6.5, 0.1-0.2 M sodium/potassium phosphate and (ii) 0.1-0.2 M sodium acetate pH 4.6, 0.7-1.8 M sodium malonate as shown in Figure 2.1. In each drop, 1.5 µL *GlcNAcase* solution was mixed with 1.5 µL of each precipitant and then equilibrated over 1.0 mL of the respective precipitant in a 24-well Linbro tissue-culture plate. The wild-type crystals in 0.1 M sodium acetate pH 4.6, 1.4 M sodium malonate were transferred into drops

consisting of 2  $\mu\text{L}$  cryoprotectant solution (0.1 M sodium acetate pH 4.6, 2.9 M sodium malonate). The substrate complex of GlcNAcase was obtained by soaking the wild-type crystals with the natural substrate (GlcNAc<sub>2</sub>; 10 mM) in the corresponding mother liquor at 20°C for 5, 10, 15 and 30 min. Afterwards, the soaked crystals were immersed for a few seconds in a cryoprotectant consisting of the mother liquor with 2.9 M sodium malonate and 10 mM GlcNAc<sub>2</sub> before flash-cooling in liquid nitrogen for further storage.



**Figure 2.1** Optimization of crystallization conditions of wild-type *VhGlcNAcase*. Grid screen of variable concentration of the conditions F10 from the PACT Suite (A) and C6 from the Anions Suite (B).

For crystallization of the GlcNAcase D437A mutant, 0.1  $\mu\text{L}$  protein solution (18 mg/mL in 20 mM Tris-HCl buffer pH 8.0 containing 150 mM NaCl and 1 mM TCEP) was pipetted into 0.1  $\mu\text{L}$  of each precipitant using the same set of crystal screens as described above. Small plate-shaped crystals (form I) were obtained within 3 days in condition G7 from the PACT Suite 9 (20% (w/v) PEG 3350, 0.1 M Bis-Tris pH 7.5, 0.2 M sodium acetate), condition D9 from the PEGs Suite (25% (w/v) PEG 6000, 0.1 M Tris-HCl pH 8.5) and condition H1 from the PEGs Suite (20% (w/v) PEG 3350, 0.2 M potassium sodium tartrate). Three-dimensional rod-shaped crystals (form II) were also obtained within 14 days of incubation in condition C12 from the Anions Suite (0.1 M MES pH 6.5, 1.2 M sodium malonate) at 20°C. The D437A crystals from condition G7 of the PACT Suite were further optimized by the hanging-drop vapour diffusion method in a 24-well Linbro tissue-culture plate as described for the wild-type enzyme. A 1.5  $\mu\text{L}$  droplet of the D437A mutant was mixed with 1.5  $\mu\text{L}$  of precipitants at various concentrations (17-22% (w/v) PEG 3350, 0.1 M Bis-Tris pH 7.5, 0.1-0.2 M sodium acetate) and then equilibrated over 1.0 mL of the respective precipitant. After the D437A crystals had been immersed in a cryoprotectant solution consisting of mother liquor supplemented with 25% (v/v) glycerol, they were rapidly transferred to liquid nitrogen and stored.



**Figure 2.2** Optimization of crystallization condition of the mutant D437A of VhGlcNAcase. Grid screen of variable concentration of the condition G7 from the PACT Suite.

### 2.13.2 Data collection and processing

For in-house crystal testing and the collection of preliminary data sets, crystals were mounted in nylon loops (Hampton Research, Aliso Viejo, California, USA) on a goniometer and data were collected on a Bruker MICROSTAR or Rigaku MicroMax-007 HF rotating-anode generator with a copper anode as the X-ray source. X-ray diffraction data were collected from high-quality crystals using a Pilatus 6M detector on beamline PX-II at the Swiss Light Source (SLS), Paul Scherrer Institute, Villigen, Switzerland. All X-ray images were recorded using 0.25 Å oscillations at -173°C and a wavelength of 1.0 Å with crystal-to-detector distances of 381, 350, 440 and 500 mm for the wild type, the wild type complexed with GlcNAc, the D437A mutant form I and the D437A mutant form II, respectively. Data indexing was carried out using iMosflm (Battye *et al.*, 2011) from the CCP4 suite. The data were further processed using XDS (Kabsch, 2010). The protein content of the asymmetric unit was estimated by calculating the Matthews coefficient (Matthews, 1968), while the solvent

content was calculated based on two subunits of protein, each with six attached histidine residues. The maximum resolution for each data set was determined by using the average ratio of measured intensity to its standard deviation ( $I/\sigma(I)$ ) that was more than 2-fold in the outer shell. The global quality indicator to measure the spread of individual intensity measurements around the average value for the group of equivalent reflections is the  $R_{\text{merge}}$  (Rhodes, 2000), which is defined as:

$$R_{\text{merge}} = [\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|] / [\sum_{hkl} \sum_i I_i(hkl)] \quad (2.1)$$

where  $I_i$  is the intensity for the  $i$ th measurement of an equivalent reflection with indices  $hkl$ .

### 2.13.3 Structural determination and refinement

Molecular replacement was employed to obtain phase information using the program *PHASER* (McCoy *et al.*, 2007) from the CCP4 suite using the structure of  $\beta$ -hexosaminidase from *Arthrobacter aurescens* (PDB code: 3RCN; 35% identical to GlcNAcase from *V. harveyi*; Midwest Center for Structural Genomics, unpublished work) as a search model. Afterwards, the final model of wild-type *VhGlcNAcase* was used as a template to obtain the phases of all data sets of the *VhGlcNAcase*-substrate complex and the mutant D437A. After obtaining initial phases from the search model, the COOT program (Emsley and Cowtan, 2004) was used to manually rebuild the model structure. The electron density maps ( $F_o - F_c$  and  $2F_o - F_c$  maps) were calculated from the observed structure-factor amplitudes ( $|F_{\text{obs}}|$ ) and the calculated structure factor amplitudes ( $|F_{\text{calc}}|$ ), where each  $|F_{\text{obs}}|$  is derived from the measured reflection intensity and each  $|F_{\text{calc}}|$  is the amplitude of the corresponding structure factor calculated from the current model (Rhodes, 2000). A new model rebuilding was carried out based on

those two maps which were recalculated for each cycle of rebuilding and further rounds of structural refinement with the program REFMAC5 available in the CCP4 suite (Collaborative Computational Project, 1994; Murshudov *et al.*, 1997). The resultant electron density map showed the disagreement regions between the search model and the new structure, which was determined through rebuilding and refinement processes that were continued until the correlation between the model and the diffraction data was maximized. The agreement between the crystallographic model and the experimental X-ray diffraction data was measured by  $R_{\text{factor}}$  which is defined as:

$$R = \frac{\sum \| |F_{\text{obs}}| - |F_{\text{calc}}| \|}{\sum |F_{\text{obs}}|} \quad (2.2)$$

where a similar quality criterion,  $R_{\text{free}}$  was calculated from 5% of the reflections selected randomly and omitted from the refinement process.

During the model rebuilding process, the electron density of the second half of GlcNAc<sub>2</sub> molecule could not be found in the structure. The molecular topology of the GlcNAc was then taken from the protein data bank (PDB code: 3GH5) and modeled into the corresponding  $2F_o - F_c$  and  $F_o - F_c$  maps. The geometry of each final model was verified by PROCHECK (Laskowski *et al.*, 1993) and MolProbity (Chen *et al.*, 2010). The refined structures of the three enzyme forms were compared within the program Superpose and direct contacts determined in the program Contact (CCP4 suite). A two-dimensional diagram of ligand-protein interactions from 3D coordinates was automatically visualized using the program LigPlot+ (Laskowski and Swindells, 2011) and the graphical structures and electron density maps were generated and displayed by PyMOL ([www.pymol.org](http://www.pymol.org)).

## 2.14 pK<sub>a</sub> calculation of GH18 chitinases

### 2.14.1 Selection and preparation of X-ray structures

Structures used in this study were selected using the CAZy database (<http://www.cazy.org>) and the Protein Data Bank (PDB; Berman *et al.* 2000). All specific PDB files used in this study were: one structure from *P. furiosus* chitinase A (PDB code: 2DSK); one structure from *Arthrobacter* sp. chitinase B (PDB code: 1KFW); one structure from *B. circulans* WL-12 chitinase A1 (PDB code: 1ITX); nine structures from *S. marcescens* chitinase B (PDB codes: 1E15, 1E6N, 1E6P, 1E6R, 1E6Z, 1GOI, 1H0G, 1OGG, 1UR9); eleven structures from *S. marcescens* chitinase A (PDB codes: 1CTN, 1EDQ, 1EHN, 1EIB, 1FFQ, 1FFR, 1K9T, 1NH6, 1RD6, 1X6L, 1X6N); one structure from *S. coelicolor* chitinase A (PDB code: 3EBV); four structures from *V. harveyi* chitinase A (PDB codes: 3B8S, 3B9A, 3B9D, 3B9E); fourteen structures from *A. fumigatus* chitinase B1 (PDB codes: 1W9P, 1W9U, 1W9V, 2A3A, 2A3B, 2A3C, 2A3E, 2IUZ, 3CH9, 3CHC, 3CHD, 3CHE, 3CHF, 1WNO); four structures from *C. immitis* chitinase 1 (PDB codes: 1D2K, 1LL4, 1LL6, 1LL7); seven structures from *H. brasiliensis* hevamine A (PDB codes: 1HVQ, 1KQY, 1KQZ, 1KR0, 1KR1, 1LLO, 2HVM); ten structures from human chitotriosidase (PDB codes: 1GUV, 1HKI, 1HKJ, 1HKK, 1HKM, 1LG1, 1LG2, 1LQ0, 1WAW, 1WB0); two structures from human AMCCase (PDB codes: 3FXY, 3FY1); one structure from *P. platycephala* endochitinase (PDB code: 2GSJ) and four structures from *S. cerevisiae* chitinase 1 (PDB codes: 2UY2, 2UY3, 2UY4, 2UY5). All 70 PDB files from the protein database are either apo structures, or holo structures with and without mutations. The PDB files were first regularized using WHAT IF (Vriend, 1990) to rebuild missing atoms and correct nomenclature problems. All crystallographic water molecules were deleted

before carrying out  $pK_a$  calculations, and all ions were removed. ‘apo-generated’ structures were also constructed by removing ligands and ions from the holo forms. Holo structures included the charges and radii for ligands calculated using PRODRG server (Schüttelkopf and van Aalten, 2004).

### 2.14.2 $pK_a$ calculations

Poisson-Boltzmann equation (PBE)-based  $pK_a$  calculation methods compute the effect of transferring a titratable group from water to its position in the protein (Yang *et al.* 1993). The protein titratable groups used in the  $pK_a$  calculation were the *N*-terminus, Asp, Glu, Tyr, Cys, Lys, Arg, His, and *C*-terminus. The estimated  $pK_a$  value of the residue in water is called the model  $pK_a$  and is determined by interpolating the  $pK_a$  values of compounds that resemble amino acid side chains. The model  $pK_a$  values used in this study were described by Nielsen and Vriend, 2001. The  $pK_a$  value of a titratable group in the protein is then computed by adding the desolvation energy with non-titratable and the interaction energy with titratable groups. Once the pH dependence of these energies is known, the fractional charge is calculated by evaluating the Boltzmann sum for every titratable group at each pH value of interest using the equation 2.3 below:

$$p_i = \frac{e^{-\frac{E_i}{kT}}}{\sum_{j=1}^{2^N} e^{-\frac{E_j}{kT}}} \quad (2.3)$$

where  $p_i$  is the fractional charge on group  $i$ ,  $E_i$  is the energy of state  $i$ ,  $E_j$  is the energy of state  $j$ , and the sum is over all protonation states in the protein. A protein consisting of  $N$  titratable groups can exist in  $2^N$  protonation states. The methods used for

calculating the pH dependence of the energy for each protonation state can be described using equation 2.4 below:

$$E_x = \sum_{i=1}^N \gamma_i \ln(10)kT(pH - pK_{a,int,i}) + \sum_{i=1}^N \sum_{j=1}^N \gamma_i \gamma_j E(i, j) \quad (2.4)$$

where  $E_x$  is the energy of protonation state  $i$  at a given pH,  $N$  is the number of active-site residues considered,  $\gamma_i$  is 1 if group  $i$  is charged and 0 if neutral, and  $E(i, j)$  is the electrostatic interaction energy between groups  $i$  and  $j$ . The ionization of a single titratable group in the absence of all other titratable groups is described by the first term. The second term describes interactions between titratable groups and it is assumed that the energy of any protonation state of the protein can be tabulated by the pairwise interaction energies between two titratable groups.

Standard  $pK_a$  calculations ( $pK_a$  calculations where all groups are included) of the whole protein were carried out using the WHAT IF  $pK_a$  calculation package (WIpKa) as described previously (Nielsen and Vriend, 2001). WIpKa employs a global hydrogen-bond network optimization procedure (Hoofst *et al.*, 1996) to arrive at the best structural configuration for every protonation state evaluated during the course of the  $pK_a$  calculation. All parameters were set as described previously (Nielsen and Vriend, 2001), except that a single dielectric constant of 8 was used for the protein. The OPLS force field (Jorgensen and Tirado-Rives, 1988) was used as a source of charges and radii. The WHAT IF  $pK_a$  calculation uses Delphi II (Nicholls and Honig, 1991) to solve the linearized form of the PBE to obtain the electrostatic energies. Desolvation energies, background interaction energies and charged-charged interaction energies were fully calculated for all titratable groups in the protein.

### 2.14.3 Prediction of pH profiles

A plot of the pH-dependent protonation state populations for a hypothetical three acidic active-site residues (D<sub>1</sub>, D<sub>2</sub> and E) was obtained by evaluation of the partition function or fractional charge using the pKaTool program (Nielsen, 2007, <http://enzyme.ucd.ie/Science/pKa/pKaTool>). The titration curves for each titratable group were constructed by adding the population of the protonation states that have a specific group in its charged state. For a case of three active-site residues (D<sub>1</sub>, D<sub>2</sub> and E), there are eight (2<sup>3</sup>) possible protonation states. The fractional charge of a specific residue (D<sub>1</sub>, D<sub>2</sub> or E) is the sum of the populations of the protonation states when the specific residue is charged at a given pH value. The titration curves were produced from the fractional charge plotted as a function of pH and displayed in pKaTool. The calculated pK<sub>a</sub> of three active-site residues (D<sub>1</sub>, D<sub>2</sub> and E) were employed in order to predict the pH activity profile of the chitinases. These D<sub>1</sub>-D<sub>2</sub>-E groups must be in a specific protonation state to provide the proton or charges required in catalytic mechanism. Several possible protonation states were investigated in pKaTool. Here, the catalytically competent protonation state (CCPS) of the D<sub>1</sub>, D<sub>2</sub>, and E in chitinases was supposed to be charged, neutral and neutral, respectively. The experimental pH profiles for family 18 chitinases used in comparison with the computational pH profiles were listed as follows: *S. marcescens* chitinase B (van Aalten *et al.*, 2001); *S. marcescens* chitinase A (Zees *et al.*, 2009); *V. harveyi* chitinase A (Suginta and Sritho, 2012); *S. cerevisiae* chitinase 1 (Hurtado-Guerrero and van Aalten, 2007); human chitotriosidase (Boot *et al.*, 2005); human AMCase (Chou *et al.*, 2006) and *H. brasiliensis* hevamine A (Bokma *et al.*, 2002). In cases where the graphical data were

available, data points at a given pH were digitized using g3data (<http://www.frantz.fi/software/g3data.php>).



## CHAPTER III

### RESULTS

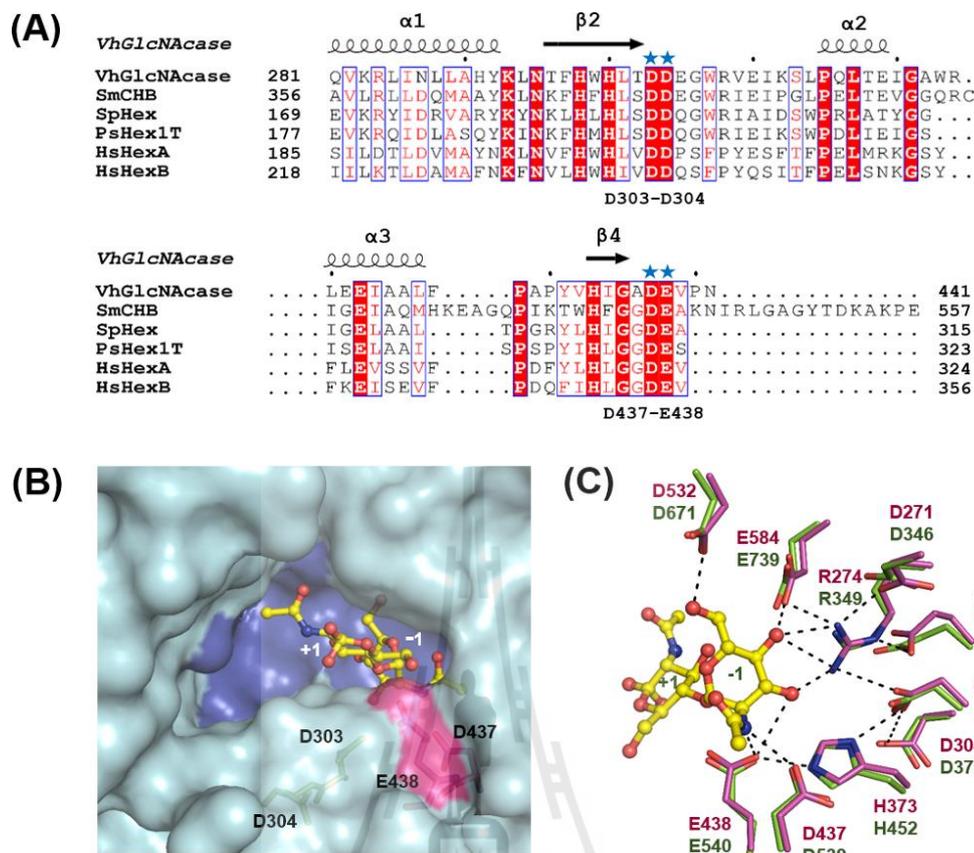
#### PART I

#### PROBING THE CATALYTIC RESIDUES OF *VhGlcNAcase* BY CHEMICAL RESCUE

##### 3.1 Sequence analysis and homology modelling

We previously reported cloning and recombinant expression of the gene encoding GH20  $\beta$ -*N*-acetylglucosaminidase from *V. harveyi* (Suginta *et al.*, 2010). The enzyme, known as *VhGlcNAcase* (formerly *VhNag2*), exhibited exolytic activity, degrading chitin oligosaccharides from the non-reducing end in a sequential manner, with GlcNAc monomer as the final product. Since the crystal structure of *VhGlcNAcase* is undetermined, we first gained preliminary information on structural identity of *VhGlcNAcase* by carrying out sequence alignment of *VhGlcNAcase* with other GH20 GlcNAcases of known structure. The results showed that the highest sequence identity of *VhGlcNAcase* was with *Serratia marcescens* chitobiase (*SmCHB*), with 24% identity (Tews *et al.*, 1996), followed by *Streptomyces plicatus*  $\beta$ -*N*-acetylhexosaminidase (*SpHex*) with 21% identity (Mark *et al.*, 2001), human  $\beta$ -hexosaminidase A (*HsHexA*) (Lemieux *et al.*, 2006) and human  $\beta$ -hexosaminidase B

(*HsHexB*) (Mark *et al.*, 2003) with 17% identity, while the lowest was with *Paenibacillus* sp.  $\beta$ -hexosaminidase (*PsHex1T*) which had 13% identity (Sumida *et al.*, 2009). Structure-based alignment of *VhGlcNAcase* and *SmCHB* (Figure 3.1A) indicated two separate conserved segments on the surface of the  $(\beta/\alpha)_8$  TIM barrel domain of the two enzymes. For *VhGlcNAcase*, the preceding segment comprises the acidic pair Asp303-Asp304, located at the end of loop2 (L2), which links strand  $\beta$ 2 and helix  $\alpha$ 2 (Figure 3.1A, upper sequence portion). The second pair, Asp437-Glu438, is present at the start of loop4 (L4) connecting strand  $\beta$ 4 and helix  $\alpha$ 2 (Figure 3.1A, lower sequence portion). Superimposition of the modelled structure of *VhGlcNAcase* with the crystal structure of *SmCHB* gave an RMSD of 0.651 Å for 390 C $\alpha$  atoms (Figure 3.1B) and showed both conserved acidic pairs to be part of the GlcNAc-binding pocket. The Asp437-Glu438 pair were located close to the scissile bond joining -1GlcNAc and +1GlcNAc, which suggested that these amino acids could play a catalytic role. Structural alignment of the active site residues (Figure 3.1C) showed that the location of the Asp303-Asp304 pair is equivalent to that of Asp378-Asp379 in *SmCHB*, whereas the Asp437-Glu438 pair corresponded with Asp539-Glu540 (Prag *et al.*, 2000). Based on the crystal structure and kinetic data, the Asp539-Glu540 pair had been suggested to have a catalytic function for *SmCHB* (Prag *et al.*, 2000).



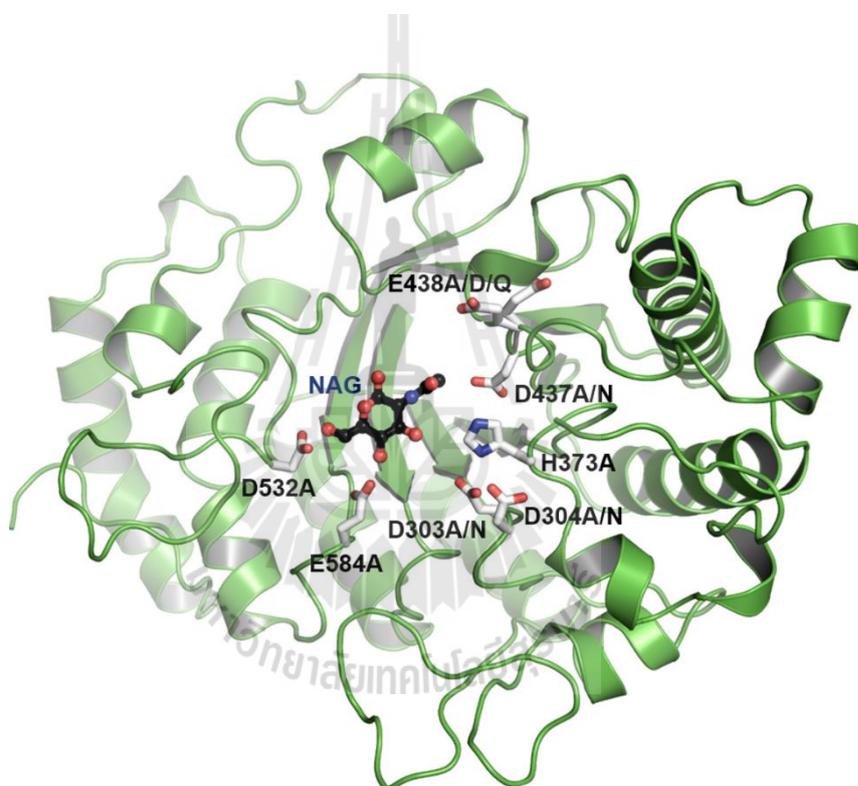
**Figure 3.1** Comparison of the modelled structure of *VhGlcNAcase* with other GlcNAcases. (A) Multiple sequence alignment of GH20 glycoside hydrolases. The amino acid sequence of *Vibrio harveyi*  $\beta$ -*N*-acetylglucosaminidase, *VhGlcNAcase* (SwissProt: D9ISE0) was retrieved from the Uniprot database. This sequence was aligned with those of *Serratia marcescens* chitobiase, *SmCHB* (SwissProt: Q54468), *Streptomyces plicatus*  $\beta$ -*N*-acetylhexosaminidase, *SpHex* (SwissProt: O85361), *Paenibacillus* sp.  $\beta$ -hexosaminidase, *PsHex1T* (SwissProt: D2KW09), human  $\beta$ -hexosaminidase A ( $\alpha$ -chain), *HsHexA* (SwissProt: P06865) and human  $\beta$ -hexosaminidase B ( $\beta$ -chain), *HsHexB* (SwissProt: P07686). The putative amino acid residues that are important for GlcNAcase activity are indicated with blue stars. (B) Surface representation of the active-site pocket of *VhGlcNAcase* (in blue) docked with GlcNAc<sub>2</sub> (in yellow stick) from *SmCHB* (PDB code: 1QBB). The solvent-accessible

surface of D437-E438 is highlighted in pink and the buried surface of D303-D304 is highlighted in green. (C) Superimposition of both conserved acidic pairs (Asp303-Asp304 and Asp437-Glu438) of modelled *VhGlcNAcase* (in magenta stick) with the crystal structure of *SmCHB* (in green stick) in complex with GlcNAc<sub>2</sub>. N atoms are shown in blue and O atoms in red.

### 3.2 Mutational design and site-directed mutagenesis

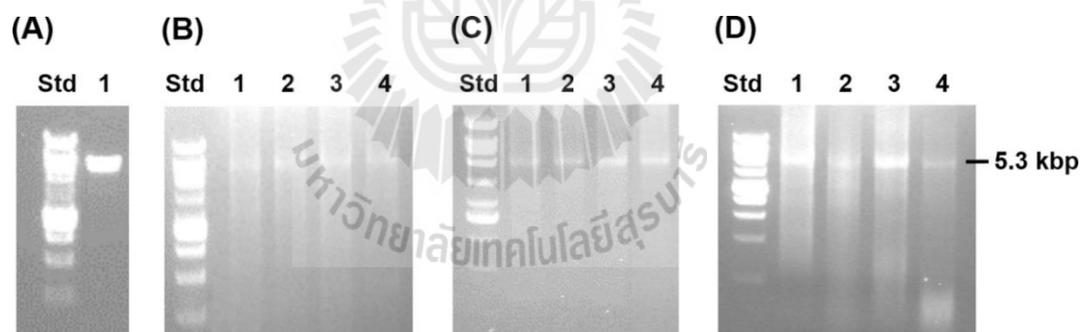
The sequence alignment of *VhGlcNAcase* with family 20 enzymes previously showed that the acidic groups were highly conserved in the TIM-barrel catalytic domain serving for enzyme catalysis. The preliminary structure of *VhGlcNAcase* in complex with sugar revealed that the acidic residues such as Asp303, Asp304, Asp437, Glu438, Asp532 and Glu584 that are completely aligned with Asp191, Asp192, Asp313, Glu314, Asp395 and Glu444, respectively, of the *SpHex*, are located in the active site pocket and interact with the GlcNAc. Interestingly, His373 is located between these acidic groups and also highly conserved with other family 20 glycoside hydrolases. The acidic groups may play an important role in the catalytic mechanism of the enzyme based on the acid-base catalysis. To investigate a role of these conserved residues for catalysis in *VhGlcNAcase*, site-directed mutagenesis was designed for all acidic active-site residues and a histidine residue. As shown in Figure 3.2, its catalytic domain in complex with substrate was represented as a ribbon in green and the hydrolysed GlcNAc was shown as a black ball-and-stick model and all target acidic residues represented in white stick model, including Asp303, Asp304, His373, Asp437, Glu438, Asp532 and Glu584 which are around the substrate binding pocket at the top of the TIM-barrel catalytic domain (Figure 3.2). In *VhGlcNAcase*, the acidic active-site

residues were virtually mutated to alanine or other polar/charged residues as a single mutation according to the primer design shown in Table 2.1. As seen in Figure 3.2, His373, Asp532 and Glu584 were only mutated to alanine whereas Asp303, Asp304, Asp437 was mutated to alanine and asparagine and Glu438 was mutated to alanine, aspartate and glutamine. Amino acid substitution of *VhGlcNAcase* generated twelve mutants changing the size and charge of the suspected catalytic residues.

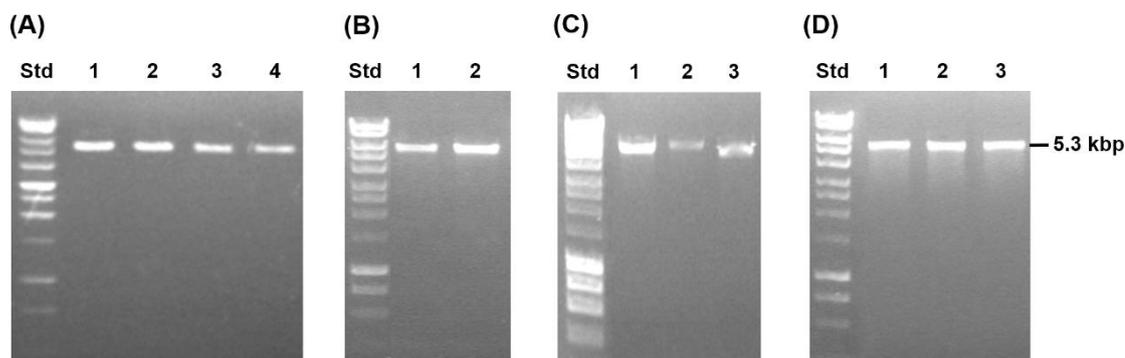


**Figure 3.2** Point mutations of active site residues. (A) Ribbon representation of the 3D structure of *VhGlcNAcase* (in green) was shown in complex with GlcNAc (in black ball-and-stick model with N atoms in blue and O atoms in red). The mutated acidic residues (Asp303, Asp304, Asp437, Glu438, Asp532 and Glu584) and His373 located in the active site of *VhGlcNAcase* are presented in white-stick model. Point mutations were introduced by site-directed mutagenesis as described in the text.

After mutational design, site-directed mutagenesis was carried out using the pQE-60 expression vector harboring the gene encoding full-length *VhGlcNAcase* as DNA template (Suginta *et al.*, 2010). The mutants D303N, D303A, D304N, D304A, H373A, D437N, D437A, E438Q, E438D, E438A, D532A and E584A were successfully amplified by PCR technique using the mutagenic primers listed in Table 2.1. The expected PCR products (~5.3 kbp) were analyzed by agarose gel electrophoresis (Figure 3.3). The PCR of the mutated DNA was treated with DpnI and then transformed into *E. coli* XL1-Blue competent cells. The mutated plasmids obtained from positive clones were extracted and examined the size by single digestion with *Bam*H1 restriction enzymes (Figure 3.4). The DNA sequencing confirmed that all the mutants were generated successfully.



**Figure 3.3** Identification of the amplified PCR product ~5,300 bp of the *VhGlcNAcase* and pQE-60 vector. The PCR products of the *VhGlcNAcase* mutants were analyzed on 1% (w/v) agarose containing sybergreen. (A) Lane Std: 1 kb DNA marker; lane 1: wild-type *VhGlcNAcase*. (B) Lanes 1-4: the mutants D437A, E438A, E438D, E438Q, respectively. (C) Lanes 1-4: the mutants D437N, H373A, D532A, E584A, respectively. (D) Lanes 1-4: the mutants D303N, D303A, D304N, D304A, respectively.

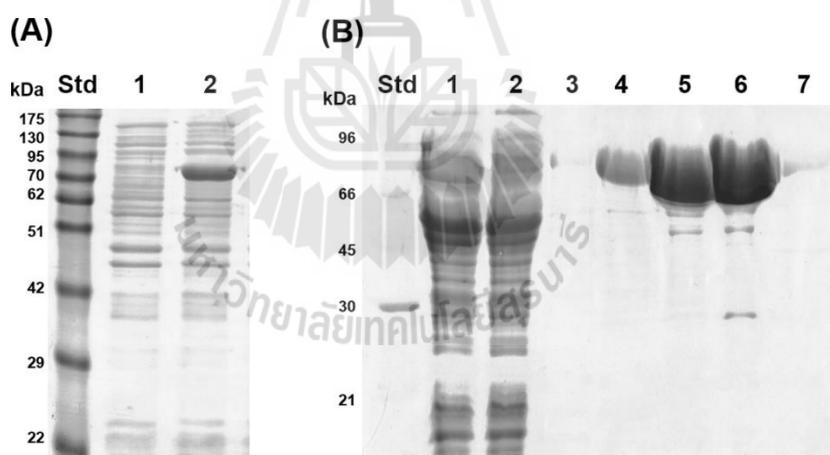


**Figure 3.4** Size confirmation of mutagenic plasmids of the *VhGlcNAcase*/pQE-60 vector by *Bam*HI restriction enzyme digestion. (A) Lane Std: 1 kb DNA marker; lanes 1-4: the mutants D437A, E438A, E438D, E438Q, respectively. (B) Lanes 1-2: the mutants D437N and H373, respectively. (C) Lanes 1-3: the mutants D532A, E584A and D303A, respectively. (D) Lanes 1-3: the mutants D303N, D304N, D304A, respectively.

### 3.3 Expression and purification of recombinant *VhGlcNAcase* and its mutants

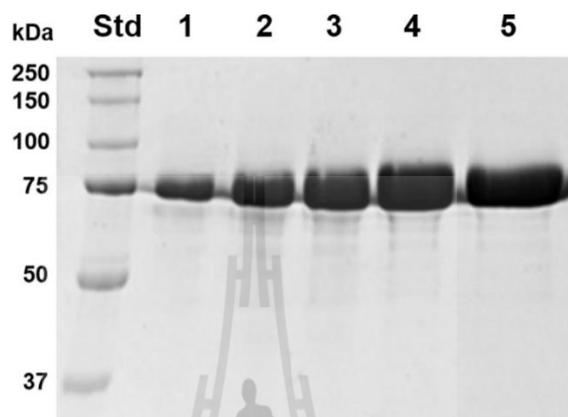
After verifying the nucleotide sequences of the mutagenized *VhGlcNAcase* cDNAs, protein expression of the *E. coli* expressed *VhGlcNAcase* recombinant protein and its mutants were evaluated. The full-length recombinant *VhGlcNAcase* and its mutants were successfully expressed at high levels when induced with IPTG (Figure 3.5A, lane 1), but not in the non-induced control group (Figure 3.5A, lane 2). The final yields obtained from this expression system were approximately 20 mg of purified protein per liter of bacterial culture. SDS-PAGE analysis under denaturing conditions showed a single band of all purified mutants migrating to ~75 kDa showing the molecular weight corresponded to the wild-type *VhGlcNAcase* as reported previously

(Suginta *et al.*, 2010). All GlcNAcase forms expressed as the 652-amino acid fragment with a C-terminal (His)<sub>6</sub>-tagged sequence were readily purified using cobalt affinity chromatography (Figure 3.5B). The 75 kDa protein corresponding to the expected size of recombinant *VhGlcNAcase* was detected as a soluble form (Figure 3.5B; lane 1) and highly expressed in eluted fractions (Figure 3.5B; lanes 5 and 6). *GlcNAcase* protein showed a high purity when examined in SDS-PAGE by increasing protein concentration from 1-20 µg (Figure 3.6). For the other mutants, the final yields obtained from this expression system were approximately 20 mg of purified protein per liter of bacterial culture with the exception of D303A and D304A showing less protein expression of about 2-5 mg. The protein purity of all mutants was shown in Figure 3.7.

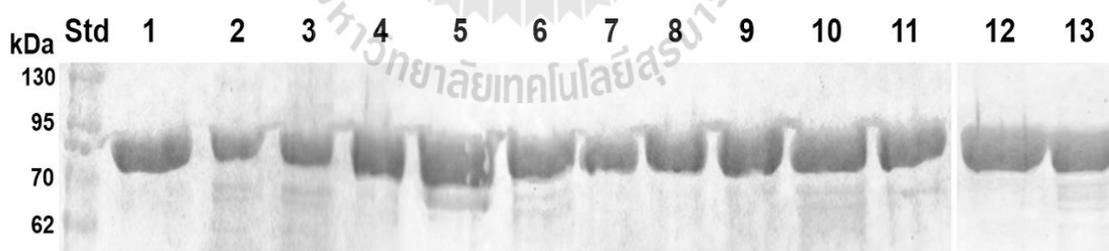


**Figure 3.5** SDS-PAGE of the expression and purification of *VhGlcNAcases*. (A) Protein expression analysis. Lane Std: protein molecular weight marker; lane 1: whole cell lysate of non-induced M15 pREP host cells containing *VhGlcNAcase/pQE-60*; lanes 2: whole cell lysate of induced M15 pREP host cells containing *VhGlcNAcase/pQE-60* with 0.4 mM IPTG. (B) Protein purification by cobalt affinity chromatography. Lane Std: protein molecular weight marker; lane 1: supernatant of cell

lysate after induction; lane 2: flow-through fraction; lane 3: wash fraction without imidazole; lane 4: wash fraction with 10 mM imidazole; lanes 5-6: eluted fractions (10 ml for each) with 250 mM imidazole.



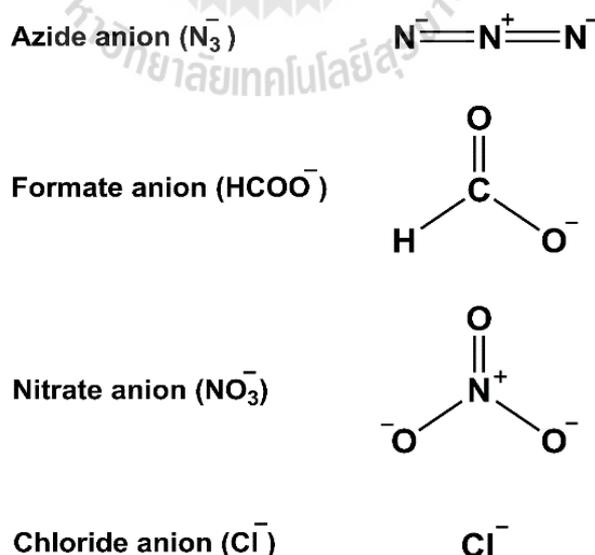
**Figure 3.6** The purity of *VhGlcNAcase* protein in 12% SDS-PAGE by increasing protein concentration from 1-20  $\mu$ g. Lane Std: protein molecular weight marker; lanes 1-5: the purified enzyme 1  $\mu$ g, 2.5  $\mu$ g, 5  $\mu$ g, 10  $\mu$ g and 20  $\mu$ g, respectively.



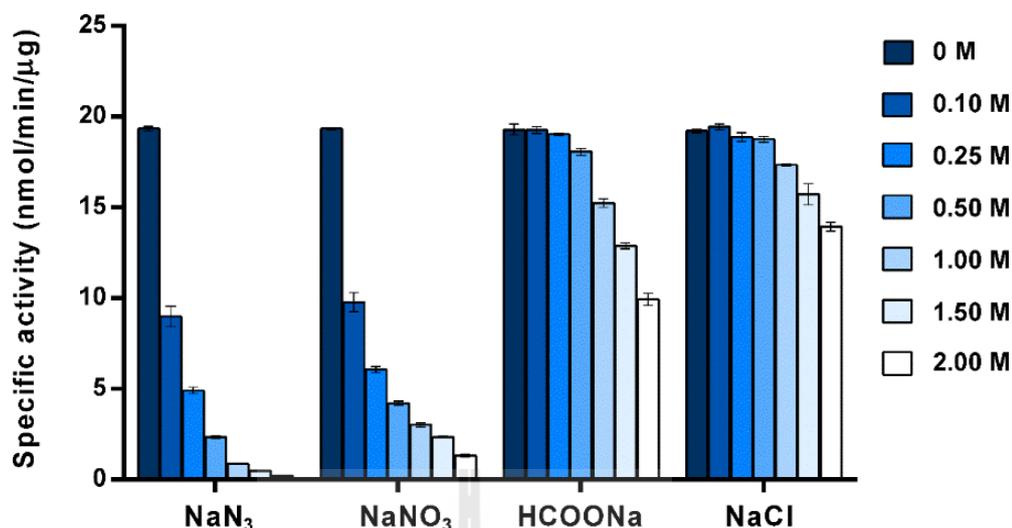
**Figure 3.7** SDS-PAGE analysis of *VhGlcNAcase* and its mutants, after IPTG-induced expression overnight. The purified enzymes (10  $\mu$ g) were electrophoresed through a 12% SDS-PAGE gel, which was then stained with Coomassie blue. Lane Std: protein molecular weight marker; lanes 1-13: wild-type, D303A, D303N, D304A, D404N, H373A, D437A, D437N, E438A, E438D, E438Q, D532A and E584A, respectively.

### 3.4 Effects of the active-site mutation on the specific activity of *VhGlcNAcase*

To study the rescue of activity by external nucleophiles, sodium azide and sodium formate, along with other sodium salts (for the chemical structures of azide and formate ions, see Figure 3.8), were first tested for their physicochemical effects on the activity of *VhGlcNAcase* WT. We recently observed that the enzymic activity of the unmutated (wild-type) *VhGlcNAcase* was inhibited by various sodium salts, including azide, nitrate, formate and chloride (Sirimontree *et al.*, 2015). Here, we confirmed their inhibitory effects on the WT activity. As shown in Figure 3.9, the specific activity of *VhGlcNAcase* WT decreased greatly when sodium azide or sodium nitrate was included in the assay medium. On the other hand, sodium formate and sodium chloride showed only moderate effects. For all the ions tested, the degree of inhibition increased with increasing concentration.



**Figure 3.8** The chemical structures of azide, formate, nitrate and chloride ions used in this study.



**Figure 3.9** Specific hydrolytic activity of wild-type *VhGlcNAcase* against *pNP-GlcNAc* in the presence of various concentrations of sodium salts in 100 mM sodium phosphate buffer, pH 7.0. The *GlcNAcase* assay was carried out as described in the text.

Point mutations of the selected residues (Asp303, Asp304, Asp 437 and Glu438) caused a drastic loss of enzymatic activity (Table 3.4, column 2). The activity of mutants D303A, D304A, D437N was undetectable, while the residual activity of other mutants, including D303N, D304N, D437A and E438A/Q, was less than 5% of that of WT *VhGlcNAcase* (Table 3.4, column 2).

Next we tested two selected compounds, sodium azide and sodium formate, for their ability to rescue the enzymic activity of the inactive mutants. The results clearly showed that sodium formate had much less inhibitory effect on the *VhGlcNAcase* mutants than on the WT enzyme (Table 3.4, column 3). When 1 M sodium azide was included in the assay medium the specific activity of the *VhGlcNAcase* WT was less than 5% of the original activity, while the residual activity of the mutants was 27-65%

of the original. On addition of sodium formate (Table 3.4, column 4), relatively less inhibition, or even enhancement of activity, was also observed with the enzyme variants. Notably, the specific activity of the D437A mutant was enhanced to 182% of the activity in the absence this compound.

**Table 3.4** Specific activity of wild-type *VhGlcNAcase* and its mutants against *pNP-GlcNAc* in the absence and presence of 1 M sodium salts.

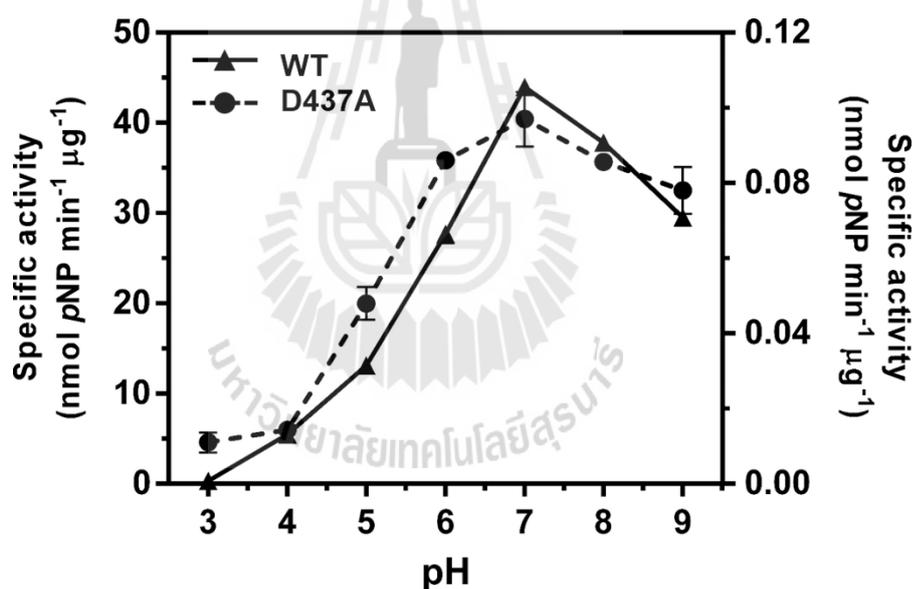
GlcNAcase mutant	Specific activity (nmol/min/ $\mu$ g)		
	No sodium salt	1 M NaN <sub>3</sub>	1 M HCOONa
Wild-type	19.4 $\pm$ 0.22 (100) <sup>a</sup>	0.89 $\pm$ 0.14 (5)	15.4 $\pm$ 0.29 (80)
D303A	n.d.	n.d.	n.d.
D303N	0.81 $\pm$ 0.02 (100)	0.31 $\pm$ 0.02 (39)	0.75 $\pm$ 0.01 (93)
D304A	n.d.	n.d.	n.d.
D304N	0.29 $\pm$ 0.01 (100)	0.10 $\pm$ 0.01 (35)	0.27 $\pm$ 0.02 (92)
D437A	0.048 $\pm$ 0.004 (100)	0.013 $\pm$ 0.003 (27)	0.087 $\pm$ 0.006 (182)
D437N	n.d.	n.d.	n.d.
E438A	0.074 $\pm$ 0.007 (100)	0.034 $\pm$ 0.003 (46)	0.075 $\pm$ 0.004 (102)
E438Q	0.089 $\pm$ 0.005 (100)	0.058 $\pm$ 0.005 (65)	0.096 $\pm$ 0.003 (108)

<sup>a</sup> Numbers in brackets indicate the relative specific activities of *VhGlcNAcase* and its variants with each sodium salt, in comparison with *VhGlcNAcase* without added sodium salt (set to 100).

n.d: undetectable activity.

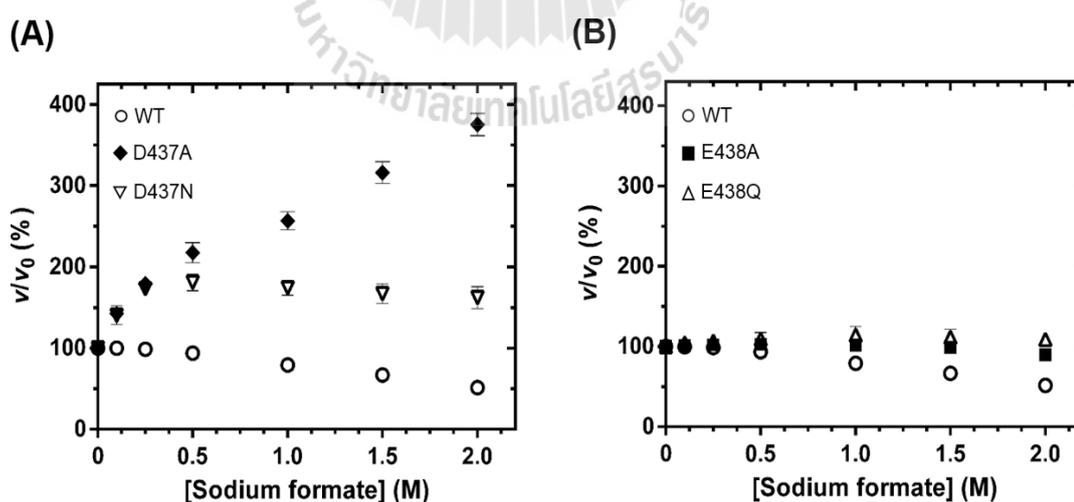
### 3.5 Effects of sodium formate concentration on the rescued activity of the D37A mutant

Since only for mutant D437A was the specific activity significantly enhanced by sodium formate, we examined whether this mutant showed a shift in the activity/pH curve compared to the WT enzyme. Figure 3.10 shows the similar response of the activity of the two *VhGlcNAcase* forms to pH variation. Although mutant D437A had a slightly broader activity/pH curve than the WT enzyme, the two forms had a similar optimal pH of around 7.0.



**Figure 3.10** Activity/pH profiles of *VhGlcNAcase* and its mutant D437A. The specific activity of *VhGlcNAcase* (solid line, left y axis) and the mutant D437A (dashed line, right y axis) was measured at pH = 3.0, 4.0, 5.0, 6.0, 7.0, 8.0 and 9.0 in the McIlvaine's sodium phosphate-citric acid buffer system. *pNP-GlcNAc* was used as substrate and the reaction was carried out for 10 min at 37°C.

Next, we investigated whether the enzyme activity of the *VhGlcNAcase* D437A was modified by sodium formate in a concentration-dependent manner. In this set of experiments, we also included the effect of concentration on the activity of the E438A mutant, for comparison. Figure 3.11 shows plots of the fractional activity ( $v_i/v_0$ ) of the enzyme at discrete concentrations of sodium formate. The relative activity of *VhGlcNAcase* WT was found to decline in response to increasing concentrations of sodium formate from 0.1 to 2.0 M. At the highest concentration of sodium formate, the residual activity of the WT enzyme was reduced to less than half of its original value. In marked contrast, the relative activity of mutant D437A was elevated with increasing sodium formate concentration, and at 2.0 M sodium formate was four times the original activity, while the relative activity of mutant D437N increased slightly (about 1.7-fold) (Figure 3.11A). However an increase in concentration of sodium formate did not restore the enzymatic activity of the E438A or E438Q mutants (Figure 3.11B).

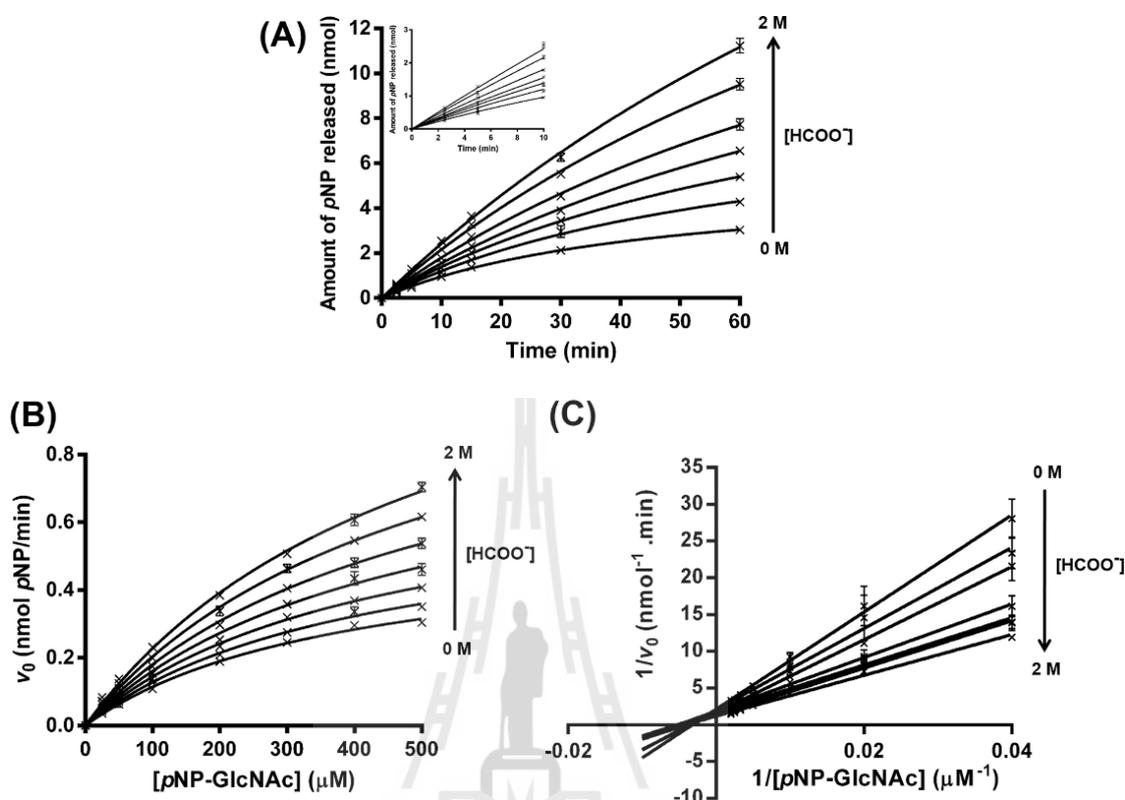


**Figure 3.11** Effect of sodium formate on *pNP-GlcNAc* hydrolysis by *VhGlcNAcase* and its mutants. Various concentrations of sodium formate (0.0-2.0 M) were added to the reaction mixture, which contained 500  $\mu\text{M}$  *pNP-GlcNAc* and 100 mM sodium

phosphate buffer, pH 7.0, at 37°C.  $v/v_0$  is fractional activity of the enzyme, i.e. activity in the presence of sodium formate relative to that in its absence. (A) The D437A (filled diamonds) and D437N (open inverted triangles) mutants. (B) The mutants E438A (filled squares) and E438Q (open triangles). The wild-type  $V_h$ GlcNAcase activity is shown as open circles in both Figure A and B.

### 3.6 Steady state kinetics of activation by sodium formate

Time-course experiments were carried to determine the initial rate of reaction. The product generated in the course of *p*NP-GlcNAc hydrolysis by the mutant D437A was monitored at different time points. Figure 3.12A is a plot of *p*NP release against time, showing that the initial rate of the reaction with and without sodium formate could be determined within 10 min. The amount of product formed up to this time was linearly proportional to the time of incubation (Figure 3.12A, inset). More detailed kinetic experiment was carried out in an attempt to analyse the mutant D437A-catalyzed rates of hydrolysis. *p*NP-GlcNAc hydrolysis without and with sodium formate from 0.1-2.0 M were performed to measure the initial rate ( $v_0$ ) of the enzyme within the incubation period of 10 min. Figure 3.12B shows non-linear increases in the initial reaction rates for the D437A mutant, with discrete increases in sodium formate concentrations from 0.1 to 2.0 M. These non-linear plots exhibit typical Michaelis-Menten kinetics, where the apparent maximum rate of reaction (app  $V_{max}$ ) is approached at a concentration of *p*NP-GlcNAc above 500  $\mu$ M. Inverse transformation of the non-linear plots in Figure 3.12B yields linear Lineweaver-Burk (LB) plots, as shown in Figure 3.12C. Each plot, representing the relation of  $1/v_0$  and  $1/[S]$ , allows estimation of the kinetic parameters of the enzyme in the absence and presence of sodium formate.



**Figure 3.12** Kinetic properties of activation by sodium formate. (A) Time-courses of reactions of the D437A mutant with and without sodium formate. Reaction mixtures (200  $\mu\text{L}$ ), containing 2  $\mu\text{g}$  of D437A mutant and 500  $\mu\text{M}$  of *pNP-GlcNAc* without sodium formate and with 0.1-2.0 M sodium formate and 100 mM sodium phosphate buffer, pH 7.0, were incubated at 37°C for 0-60 min, and the reaction terminated with 100  $\mu\text{L}$  of 3 M  $\text{Na}_2\text{CO}_3$ . Release of *pNP*, monitored at  $A_{405}$ , was converted to molar quantities using a calibration curve of *pNP* (0-20 nmol). The linear part of the reaction progress was shown as an inset. (B) Initial reaction rates for the mutant D437A of *VhGlcNAcase* in the presence of sodium formate were obtained from Michaelis-Menten plots. Reaction rates were measured using *pNP-GlcNAc* (0-500  $\mu\text{M}$ ) as the substrate, 5  $\mu\text{g}$  of the mutant D437A of *VhGlcNAcase* without sodium formate and

with 0.1-2.0 M sodium formate. (C) Activation by formate anion was evaluated by means of Lineweaver-Burk plots of initial reaction rates.

Curve fittings from Michaelis-Menten plots were conducted for individual concentrations of sodium formate (Figure 3.12B), yielding the kinetic parameters apparent  $k_{cat}$  and apparent  $K_m$  as presented in Table 3.5. At low concentration of substrate ( $[S] \ll K_m$ ), the reaction rates are defined as:

$$v_0 = k_{cat} [E][S] / K_m \quad (3.1)$$

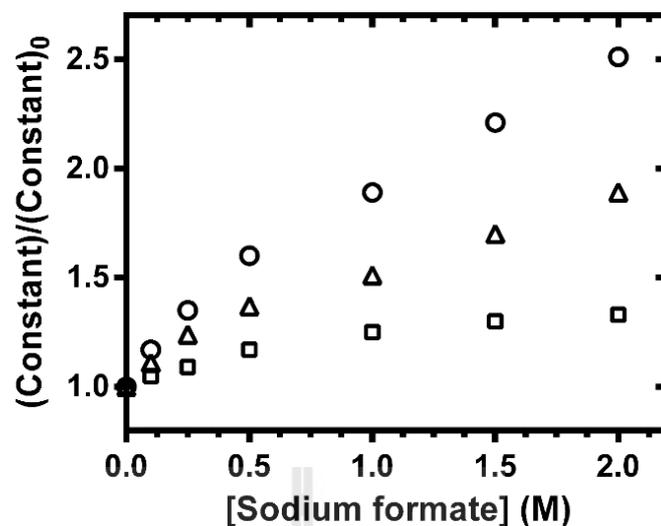
where  $v_0$  is the initial velocity,  $k_{cat}$  is the first-order rate constant,  $K_m$  is the Michaelis constant and  $[S]$  is the concentration of *p*NP-GlcNAc substrate.

The kinetic parameters were extracted from these data by non-linear fitting to the Michaelis-Menten equation using the Prism v5.0 software (GraphPad Software Inc., San Diego, CA). The kinetic data for the mutant D437A in Table 3.5 indicate discrete increases in apparent values of  $K_m$ ,  $k_{cat}$ , and  $k_{cat}K_m$ . The enhancement of the enzymic activity of *Vh*GlcNAcase on addition of sodium formate is shown in Figure 3.13 as plots of  $(app\ K_m)/K_m$ ,  $(app\ k_{cat})/k_{cat}$ , and  $(app\ k_{cat}/K_m)/(k_{cat}/K_m)$ , all relative to the values in the absence of formate ( $constant/constant_0$ ), as a function of formate concentration. The data analysis indicates a small, concentration-dependent increase in the  $K_m$  value, reaching 1.3-times the reference value at 2.0 M sodium formate. In contrast, very significant increases in the constant  $k_{cat}$  were observed, and at 2.0 M sodium formate,  $k_{cat}$  was 2.5 fold greater than at 0 M. Hence, the ratio  $k_{cat}/K_m$  was increased to 1.9-times the reference ratio, in the presence of 2.0 M sodium formate.

**Table 3.5** Kinetic parameters for the hydrolytic activity of *VhGlcNAcase* mutant D437A in the presence of increasing sodium formate (HCOONa) concentration. The presented values are Mean  $\pm$  S.D. obtained from experiments carried out in triplicate.

Sodium formate concentration (M)	$K_m$ ( $\mu\text{M}$ )	$k_{\text{cat}}$ ( $\text{s}^{-1}$ )	$k_{\text{cat}}/K_m$ ( $\text{s}^{-1} \text{mM}^{-1}$ )
0	$390 \pm 32$ (100) <sup>a</sup>	$0.14 \pm 0.006$ (100)	0.36 (100)
0.10	$411 \pm 37$ (105)	$0.16 \pm 0.008$ (117)	0.40 (111)
0.25	$425 \pm 29$ (109)	$0.19 \pm 0.007$ (135)	0.44 (124)
0.50	$456 \pm 56$ (117)	$0.22 \pm 0.02$ (160)	0.49 (137)
1.00	$489 \pm 41$ (125)	$0.26 \pm 0.01$ (189)	0.54 (151)
1.50	$509 \pm 33$ (130)	$0.31 \pm 0.01$ (221)	0.60 (170)
2.00	$519 \pm 38$ (133)	$0.35 \pm 0.02$ (251)	0.67 (189)

<sup>a</sup> Values in brackets represent relative activity compared to that without sodium formate (set as 100).



**Figure 3.13** Chemical rescue of the D437A mutant by sodium formate. Three kinetic constant ratios: the apparent first-order rate constants,  $k_{cat}$  (open circles), apparent Michaelis constants,  $K_m$  (open squares) and apparent second-order rate constants,  $k_{cat}/K_m$  (open triangles), were plotted as a function of sodium formate concentration.



## PART II

### FUNCTIONAL AND STRUCTURAL CHARACTERIZATION OF *VhGlcNAcase*

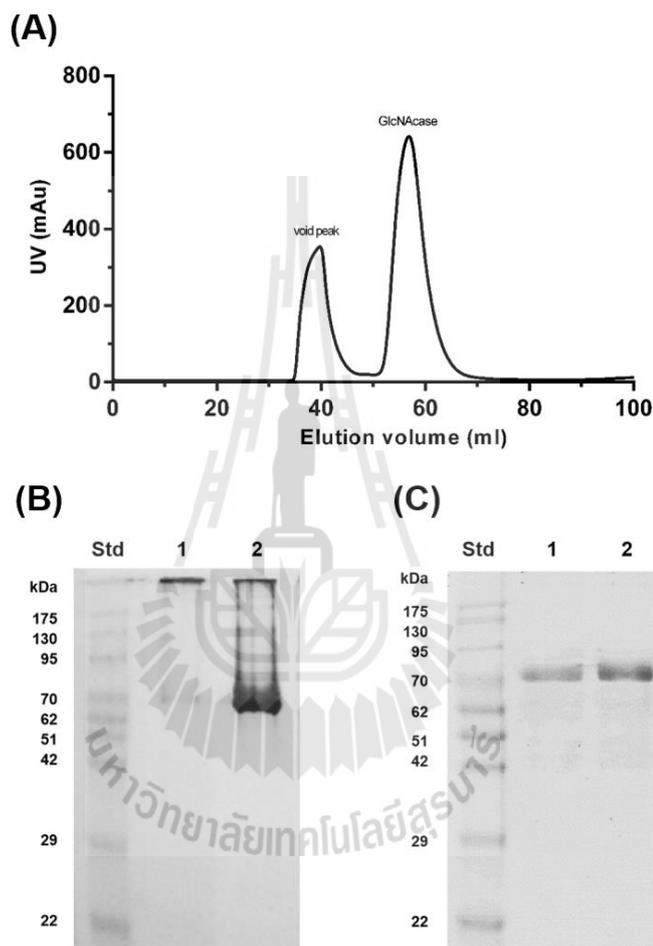
#### 3.7 Functional characterization of *VhGlcNAcases*

##### 3.7.1 Expression and purification

For crystallization purposes, both wild-type and mutated GlcNAcases were highly expressed in *E. coli* M15 (pREP4) cells as C-terminally His<sub>6</sub>-tagged polypeptides that could be purified by affinity chromatography. Both GlcNAcases, the wild type and the D437A mutant, were purified to homogeneity by a two-step protocol involving TALON metal resin affinity chromatography followed by HiPrep (16/60) Sephacryl S-200 HR gel-filtration chromatography.

Figure 3.14A shows a chromatographic profile of the elution of the wild-type protein from a HiPrep 16/60 Sephacryl S-200 prepacked column. The enzyme was found in two peaks. The first peak, which eluted at the void volume, was shown by native PAGE to be aggregated protein with a molecular mass that was too large to enter the gel (Figure 3.14B, lane 1). However the second peak, which eluted at 45-70 ml, was non-aggregated protein that migrated with an apparent molecular mass of 75 kDa, corresponding to the molecular mass of the GlcNAcase monomer (Figure 3.14B, lane 2). The pooled fraction of each protein peak was found to migrate similarly under denaturing conditions on SDS-PAGE (Figure 3.14C), indicating that they were likely

to contain the same protein. When the GlcNAcase activity was assayed with *pNP*-GlcNAc as the substrate, the second peak showed high activity, while no GlcNAcase activity was detected for the void peak.



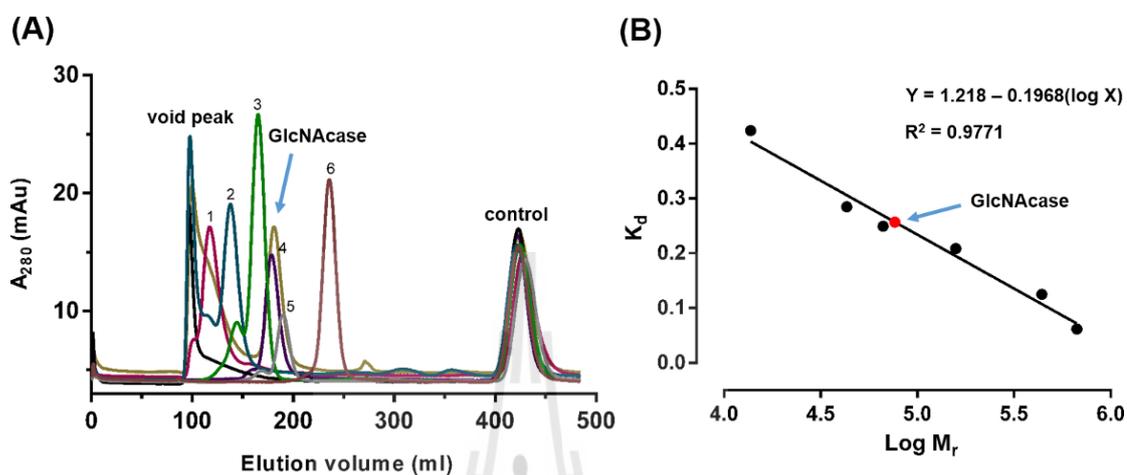
**Figure 3.14** Purification of *VhGlcNAcase* expressed from *E. coli* M15 cells using a HiPrep 16/60 Sephacryl S-200 prepacked gel-filtration column. (A) A chromatographic elution profile of wild-type *VhGlcNAcase* obtained from an ÄKTApurifier system. (B) GlcNAcase-containing fractions obtained from the two peaks were pooled separately and then loaded onto native PAGE followed by Coomassie Blue staining. (C) The same protein fractions were analysed by SDS-PAGE. Lane Std, low-molecular-weight

protein markers (labelled in kDa); lane 1, pooled fractions from the void peak; lane 2, pooled fractions from the second peak.

### 3.7.2 Molecular weight determination of *VhGlcNAcase*

To verify the *VhGlcNAcase* state in solution, molecular weight of *VhGlcNAcase* was determined by size-exclusion chromatography. Under denaturing conditions, the wild-type *VhGlcNAcase* showed a single protein band migrating to ~75 kDa. The *VhGlcNAcase* was previously suggested to be a monomeric enzyme as suggested by native-PAGE analysis (Meekrathok *et al.*, 2015). The well-defined globular protein standards ranging from 349 to 669,000 Da were loaded to the HiPrep 26/60 Sephacryl S-300 prepacked column using the  $N_\epsilon$ -DNP-L-lysine hydrochloride (349 Da) as a control for the retention volume of each standard protein. The chromatographic profile of *VhGlcNAcase* and standard proteins is shown in Figure 3.15A. The chromatographic profile of *VhGlcNAcase* was in the vicinity of the chromatographic profile of bovine serum albumin (BSA) that has the molecular weight about 66 kDa. The retention volume of each standard protein was then converted to the gel-phase distribution coefficient ( $K_d$ ) according to the equation:  $K_d = (V_e - V_o)/V_s$  described in Materials and Methods for making a calibration curve of standard proteins. From a plot of  $K_d$  versus log MW in Figure 3.15B, the estimated molecular mass of *VhGlcNAcase* from size-exclusion chromatography was calculated to be 76.42 kDa, suggesting a monomeric form of *VhGlcNAcase* in solution which is in a good agreement with the theoretical mass from mass spectrometry and SDS-PAGE analysis showing the protein band migrated to ~75 kDa. In addition, PISA analysis (Krissinel and Henrick, 2007) predicted no dimer formation in solution, which is consistent with

the molecular mass of the active enzyme (~75 kDa) obtained from size-exclusion chromatography.

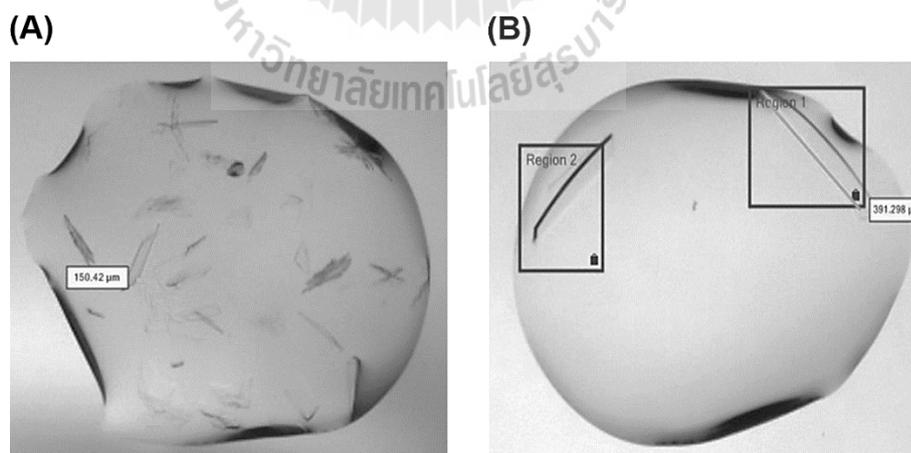


**Figure 3.15** A chromatographic profile and calibration curve of *VhGlcNAcase* and standard proteins by size-exclusion chromatography. (A) The well-defined globular protein standards ranging from 349 to 669,000 Da were employed for making a calibration curve of the HiPrep 26/60 Sephacryl S-300 prepacked columns. The protein standard used as a control is  $N_\epsilon$ -DNP-L-lysine hydrochloride (349 Da). The protein standards: 1, thyroglobulin (669 kDa); 2, ferritin (440 kDa); 3, aldolase (158 kDa); 4, bovine serum albumin (66 kDa); 5, ovalbumin (43 kDa); 6, ribonuclease A (13.7 kDa). The blue dextran 2000 was used to determine the void fraction in the column. (B) The gel-phase distribution coefficient ( $K_d$ ) was calculated from the equation:  $K_d = (V_e - V_o)/V_s$  where  $V_e$  is elution volume,  $V_o$  is void volume (determined using Blue Dextran), and  $V_s$  is volume of stationary phase. The estimated molecular mass of *VhGlcNAcase* was determined from the calibration curve (plot of  $K_d$  versus  $\log MW$ ) once its  $K_d$  value is calculated from the measured elution volume.

## 3.8 Crystallization of the wild-type *VhGlcNAcase* and mutant D437A

### 3.8.1 Initial screening and optimization

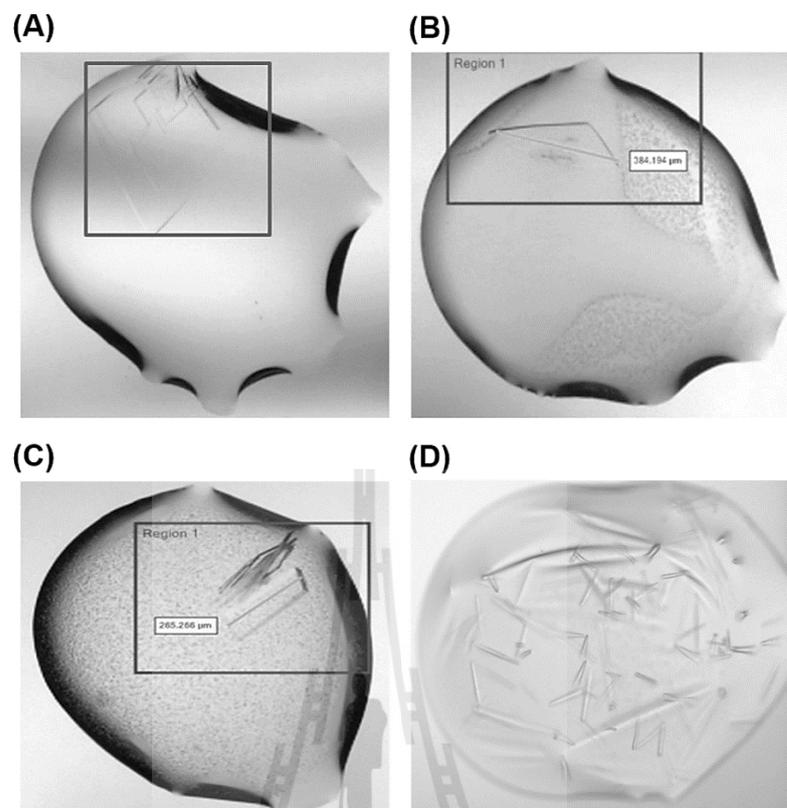
Crystallization trials with the sitting-drop vapour-diffusion method yielded small crystals from various conditions and initial screening of the crystal quality with an in-house X-ray diffractometer showed that most conditions did not give X-ray-quality crystals, with the exception of those obtained from conditions F10 from the PACT Suite (20% (w/v) PEG 3350, 0.1 M Bis-Tris propane pH 6.5, 0.2 M sodium/potassium phosphate) and C6 of the Anions Suite (0.1 M sodium acetate pH 4.6, 1.2 M sodium malonate). From condition F10 from the PACT Suite, the crystals with small-plate clusters (Figure 3.16A), obtained within 1 day at 20°C, whereas only two 3D-plate crystals (Figure 3.16B) were obtained within 3 days at 20°C in the condition C6 of the Anions Suite. These crystals could be diffracted to 2.4 Å resolution using the in-house X-ray diffraction system.



**Figure 3.16** Crystals of ligand-free *VhGlcNAcase* from initial screenings using sitting drop technique. The conditions included (A) 20% (w/v) PEG 3350, 0.1 M Bis-Tris

propane pH 6.5, 0.2 M sodium/potassium phosphate and (B) 0.1 M sodium acetate pH 4.6, 1.2 M sodium malonate.

In a case of inactive mutant D437A, initial screening of the crystal by the sitting-drop vapour-diffusion method yielded small crystals from various conditions with lower quality diffraction when monitored with an in-house X-ray diffractometer. The good-quality crystals were obtained under conditions, including G7 from the PACT Suite (20% (w/v) PEG 3350, 0.1 M Bis-Tris pH 7.5, 0.2 M sodium acetate) at 20°C within 3 days (Figure 3.17A); D9 from the PEGs Suite (25% (w/v) PEG 6000, 0.1 M Tris-HCl pH 8.5) at 20°C within 3 days (Figure 3.17B); H1 from the PEGs Suite (20% (w/v) PEG 3350, 0.2 M potassium sodium tartrate) at 20°C within 3 days (Figure 3.17C); and C12 from the Anions Suite (0.1 M MES pH 6.5, 1.2 M sodium malonate) at 20°C within 14 days (figure 3.17D). Small plate-shaped crystals (form I) were observed under conditions G7 from the PACT Suite, D9 from the PEGs Suite and H1 from the PEGs Suite while three-dimensional rod-shaped crystals (form II) were observed under condition C12 from the Anions Suite. The precipitant PEG 3350 and the salt sodium malonate were found to be appropriate for further optimization of this protein.

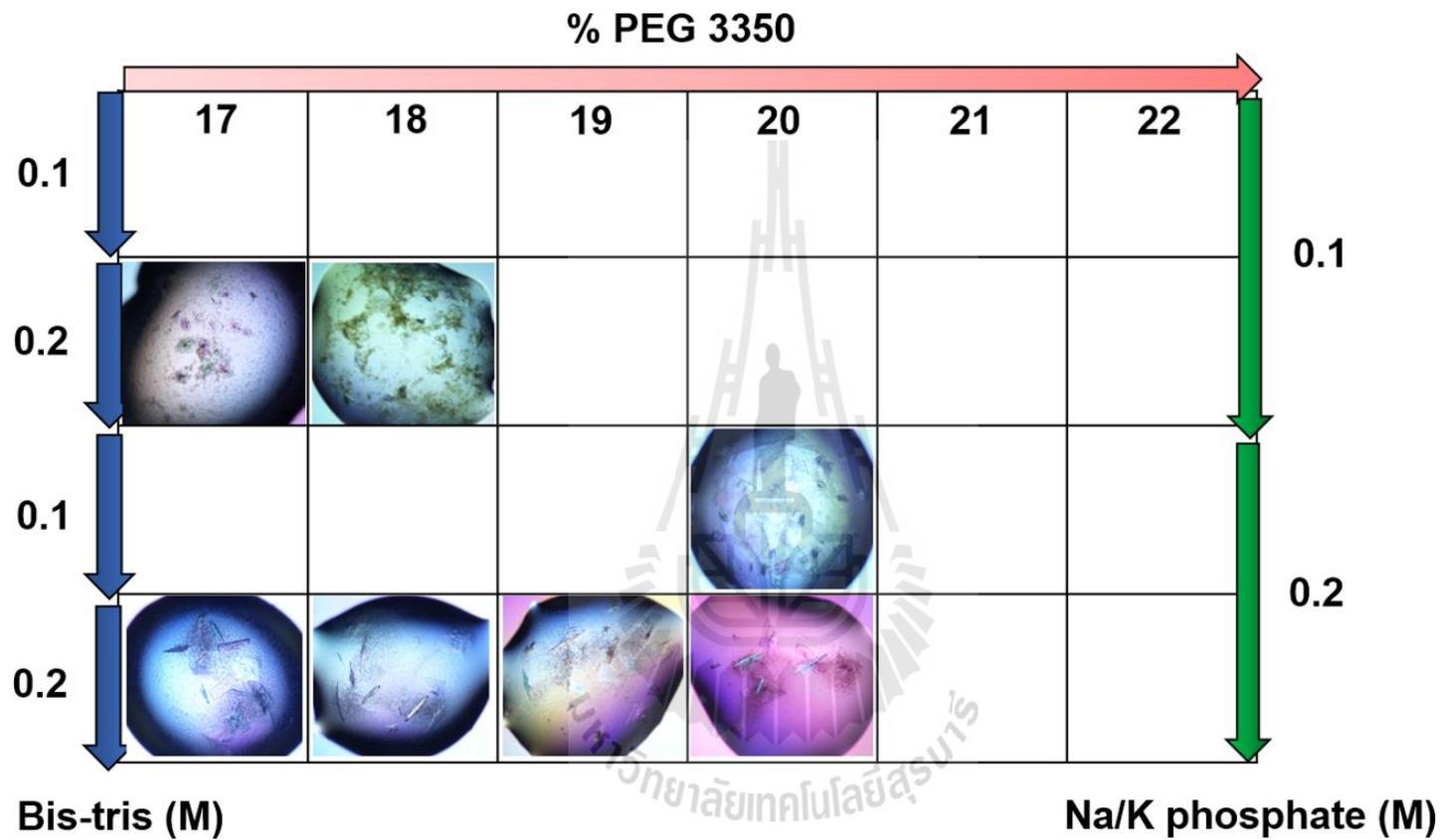


**Figure 3.17** Crystals of the inactive mutant D437A of *VhGlcNAcase* from initial screenings using sitting drop technique. The conditions included (A) 20% (w/v) PEG 3350, 0.1 M Bis-Tris pH 7.5, 0.2 M sodium acetate (B) 25% (w/v) PEG 6000, 0.1 M Tris-HCl pH 8.5 (C) 20% (w/v) PEG 3350, 0.2 M potassium sodium tartrate and (D) 0.1 M MES pH 6.5, 1.2 M sodium malonate.

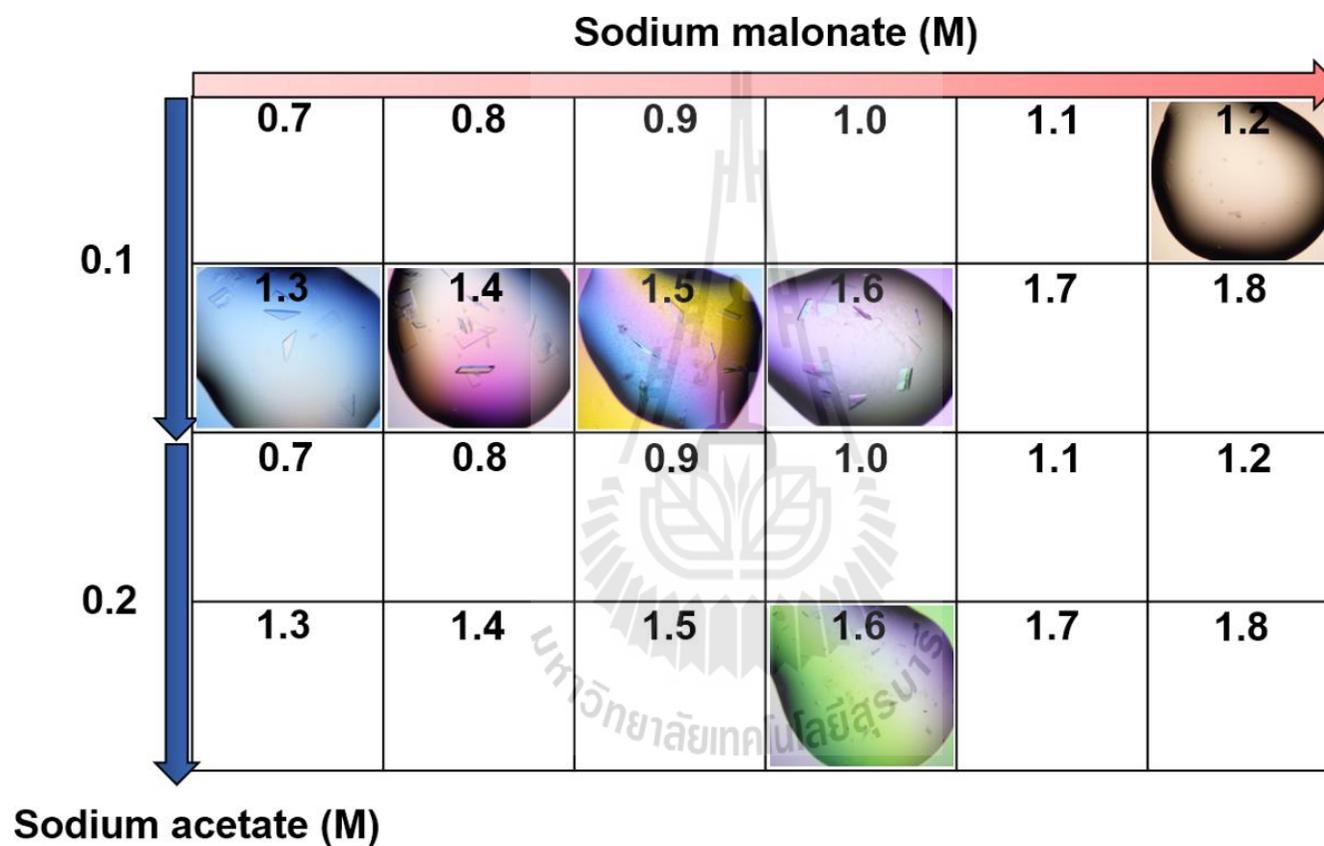
From all the positive screening conditions, the crystals of wild-type *GlcNAcase* grown under the conditions F10 from the PACT Suite and C6 of the Anions Suite were further optimized by the hanging drop method. For the conditions F10 from the PACT Suite, the concentrations of PEG 3350 were varied in range of 17-22% (w/v), Bis-Tris and sodium/potassium phosphate were varied from 0.1-0.2 M in three dimension grid screen as shown in Figure 3.18. More 3D plate crystals were found within 1 day at 20°C

in conditions including 17-18% PEG 3350, 0.2 M Bis-Tris, 0.1 M sodium/potassium phosphate; 20% PEG 3350, 0.1 M Bis-Tris, 0.2 M sodium/potassium phosphate; and 17-20% PEG 3350, 0.2 M Bis-Tris, 0.2 M sodium/potassium phosphate. Although all crystals were diffracted and gave diffraction patterns similar to that of protein diffraction pattern, the diffraction was only to low resolution (more than 3Å), which was not enough for collection and processing.

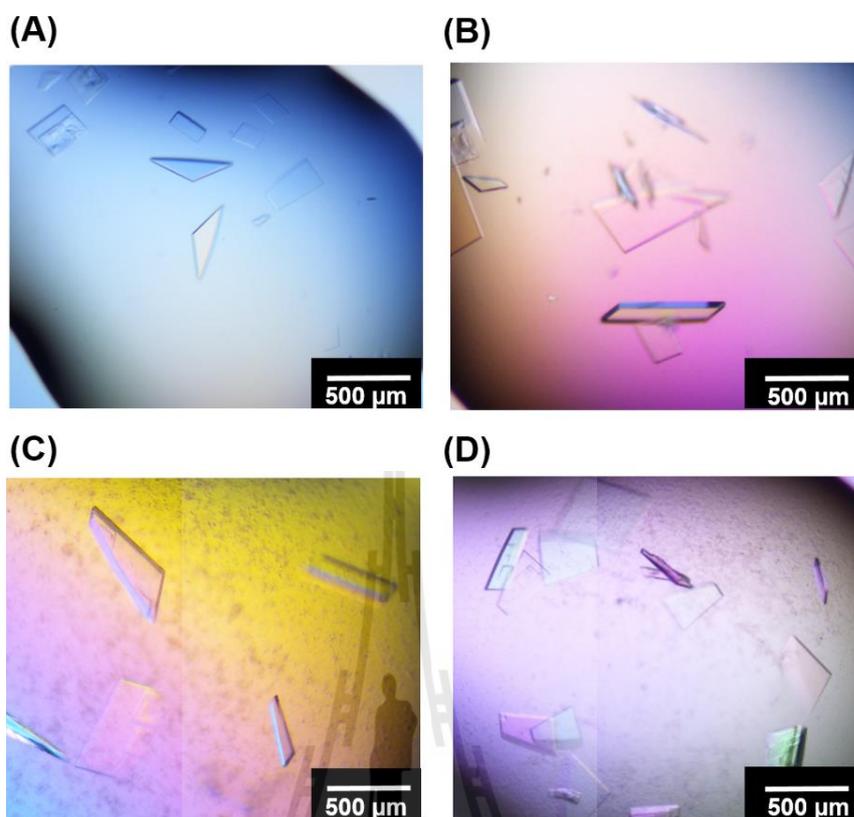
For the conditions C6 of the Anions Suite, the concentrations of sodium malonate were varied in range of 0.7-1.8 M and sodium acetate were varied from 0.1-0.2 M in two dimension grid screen as shown in Figure 3.19. More 3D plate crystals were found within 3 days at 20°C in conditions including 1.2-1.6 M sodium malonate and 0.1 M sodium acetate and 1.6 M sodium malonate and 0.2 M sodium acetate. The high-quality crystals of wild-type *VhGlcNAcase* were the conditions 1.3-1.6 M sodium malonate and 0.1 M sodium acetate (Figure 3.20), giving the diffraction to 2.4 Å resolution.



**Figure 3.18** Grid screen with variation of PEG 3350, Bis-Tris and sodium/potassium phosphate.

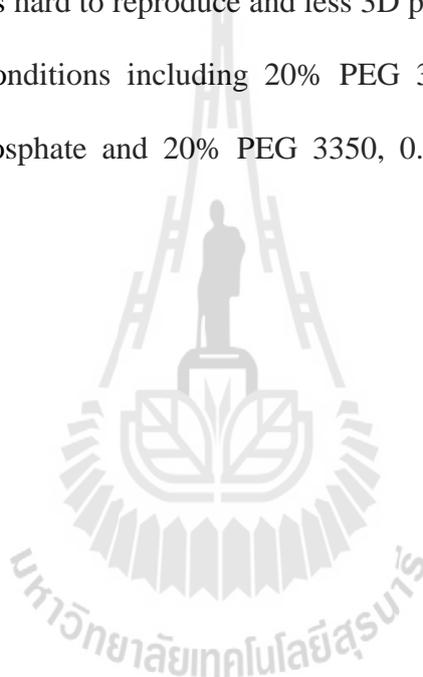


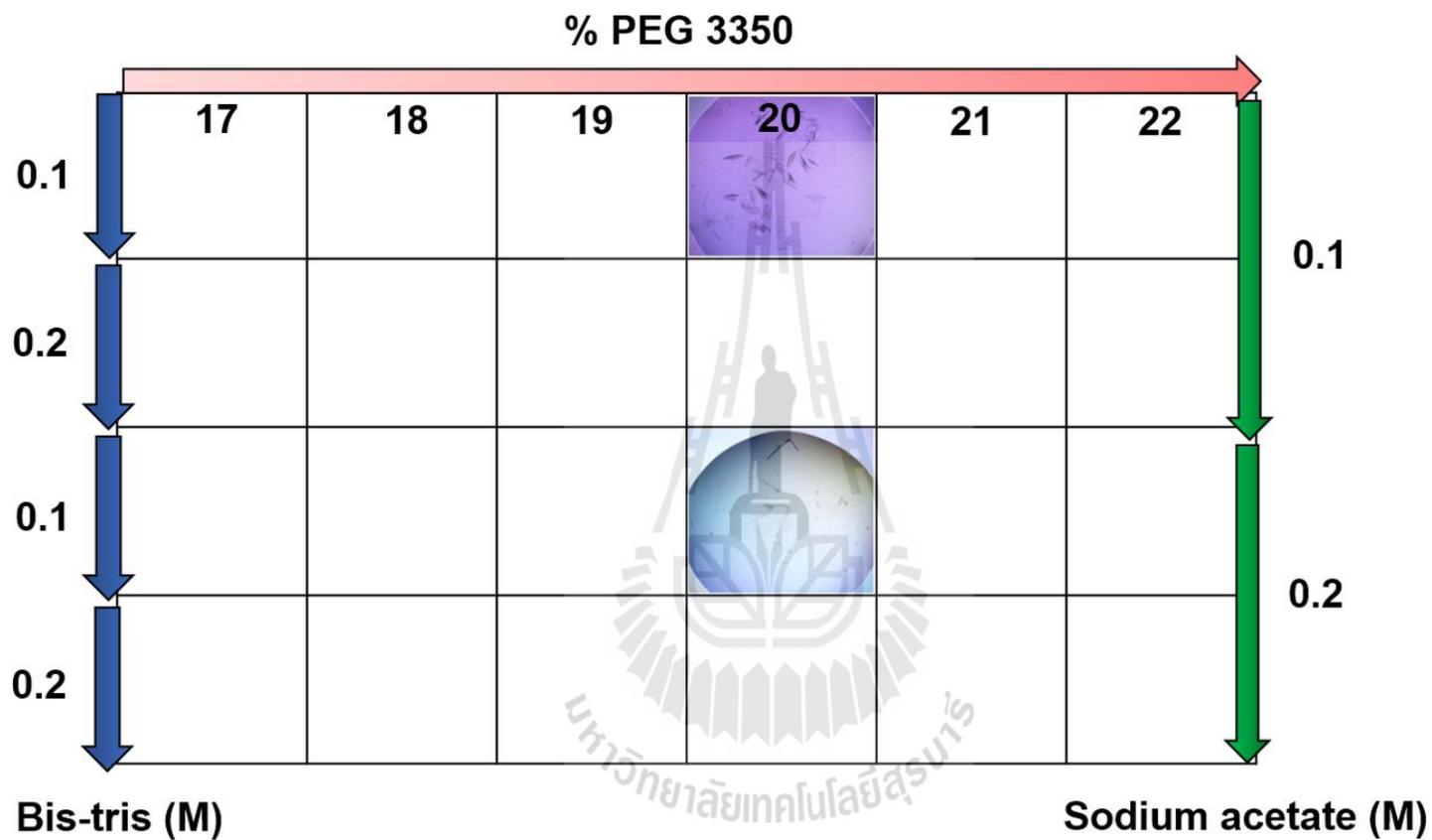
**Figure 3.19** Grid screen with variation of sodium malonate and sodium acetate.



**Figure 3.20** Photographs of wild-type *VhGlcNAcase* crystals obtained from the optimization using a variation of sodium malonate and sodium acetate concentrations. (A) 1.3 M sodium malonate and 0.1 M sodium acetate (B) 1.4 M sodium malonate and 0.1 M sodium acetate (C) 1.5 M sodium malonate and 0.1 M sodium acetate and (D) 1.6 M sodium malonate and 0.1 M sodium acetate.

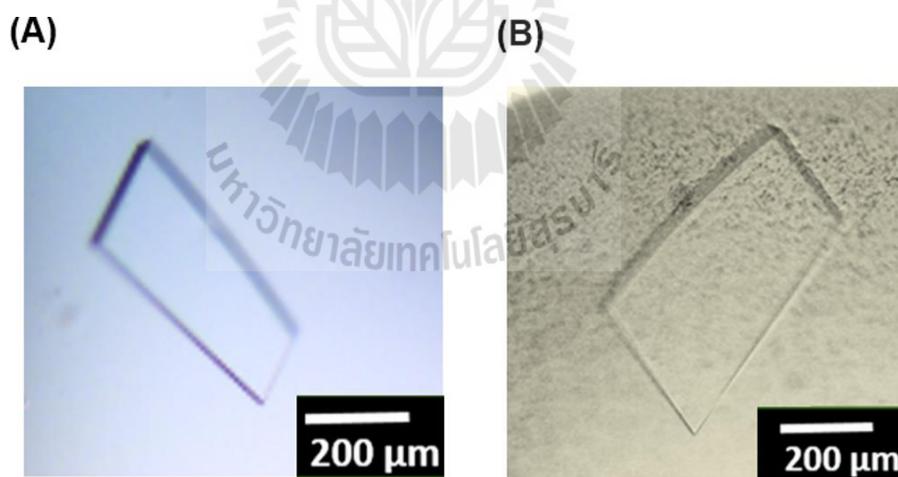
From all the positive screening conditions, the crystals of inactive mutant D437A grown under the conditions G7 from the PACT Suite (20% (w/v) PEG 3350, 0.1 M Bis-Tris pH 7.5, 0.2 M sodium acetate) was further optimized by the hanging drop method. For the conditions G7 from the PACT Suite, the concentrations of PEG 3350 were varied in range of 17-22% (w/v), Bis-Tris and sodium acetate were varied from 0.1-0.2 M in three dimension grid screen as shown in Figure 3.21. The crystal of the mutant D437A was hard to reproduce and less 3D plate crystals were found within 3 days at 20°C in conditions including 20% PEG 3350, 0.1 M Bis-Tris, 0.1 M sodium/potassium phosphate and 20% PEG 3350, 0.1 M Bis-Tris, 0.2 M sodium acetate.





**Figure 3.21** Grid screen with variation of PEG 3350, Bis-Tris and sodium acetate.

In summary, large single crystals of wild-type GlcNAcase appeared as thick three-dimensional plates with approximate dimensions of  $400 \times 200 \times 20 \mu\text{m}$  (Figure 3.22A) after optimization with 0.1 M sodium acetate pH 4.6 and 1.4 M sodium malonate. In the case of the crystal complex of wild-type *Vh*GlcNAcase, soaking the wild-type crystal with the GlcNAc2 substrate for a short time (5 min) was attempted but without success. We eventually obtained X-ray diffraction data for the native enzyme complexed with a single GlcNAc molecule. After further optimization, the inactive D437A mutant crystals appeared as thick three-dimensional plates in a condition consisting of 20% (w/v) PEG 3350, 0.1 M Bis-Tris pH 7.5, 0.1 M sodium acetate. The D437A crystals grew to approximate final dimensions of  $500 \times 300 \times 50 \mu\text{m}$  (Figure 3.22B) in a space group similar to that of the wild-type crystals.



**Figure 3.22** Refined crystals after optimization (A) A crystal of wild-type *Vh*GlcNAcase, with dimensions of  $400 \times 200 \times 20 \mu\text{m}$ , obtained from a hanging-drop vapour-diffusion setup using 0.1 M sodium acetate pH 4.6 containing 1.4 M sodium malonate. (B) A crystal of the D437A mutant (dimensions of  $500 \times 300 \times 50 \mu\text{m}$ ) obtained from a hanging-drop vapour-diffusion setup using 20% (w/v) PEG 3350, 0.1

M Bis-Tris pH 7.5, 0.1 M sodium acetate. Both crystals were obtained within 3 days of incubation at 20°C.

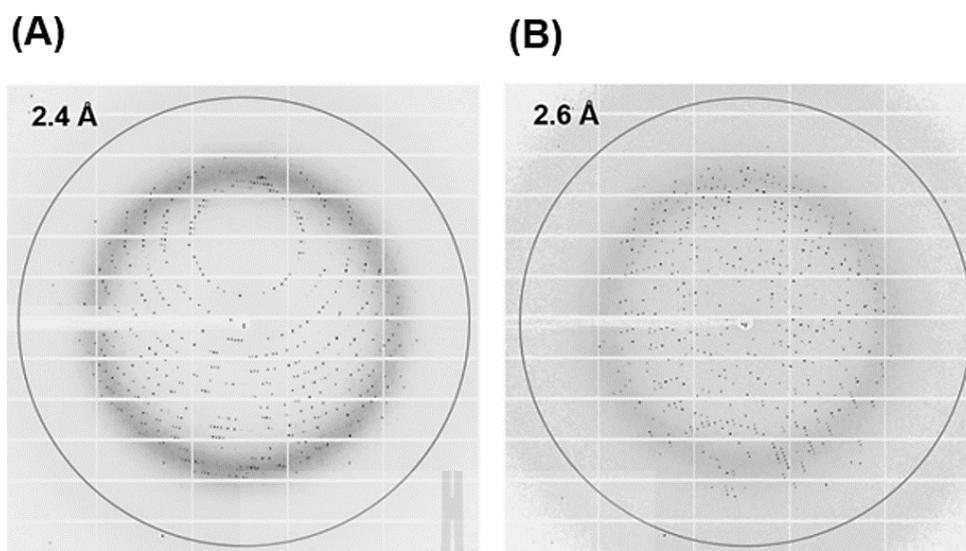
### 3.9 Data collection and processing of the wild-type *VhGlcNAcase* and mutant D437A

The final statistics for X-ray data-collection and processing of the *VhGlcNAcase* crystal variants are shown in Table 3.6. The refined unit-cell parameters of the wild-type crystal are  $a = 90.2$ ,  $b = 130.7$ ,  $c = 98.5$  Å,  $\alpha = 90^\circ$ ,  $\beta = 113^\circ$  and  $\gamma = 90^\circ$ . The crystal contains two molecules per asymmetric unit, with an estimated Matthews coefficient ( $V_M$ ) of  $3.59$  Å<sup>3</sup> Da<sup>-1</sup>. The indexing statistics were compatible with the monoclinic space group  $P2_1$ . For the wild-type crystal in complex with GlcNAc, the unit-cell parameters are  $a = 91.3$ ,  $b = 129.6$ ,  $c = 100.0$  Å,  $\alpha = 90^\circ$ ,  $\beta = 114.4^\circ$  and  $\gamma = 90^\circ$ , which are very similar to those of the unliganded wild-type crystal in the same space group. For the D437A mutant crystal (form I), the refined unit-cell parameters are again very similar:  $a = 89.4$ ,  $b = 129.3$ ,  $c = 98.4$  Å,  $\alpha = 90^\circ$ ,  $\beta = 112.2^\circ$  and  $\gamma = 90^\circ$ . D437A mutant crystals also grew as rod-shaped crystals (form II) belonging to the tetragonal space group  $P4_32_12$  (Figure 3.17D), with unit-cell parameters  $a = 165.1$ ,  $b = 165.1$ ,  $c = 155.2$  Å,  $\alpha = 90^\circ$ ,  $\beta = 90^\circ$  and  $\gamma = 90^\circ$  and again with two molecules per asymmetric unit. Figure 3.23 shows representative diffraction images for both wild-type and monoclinic D437A mutant crystals, which diffracted to 2.4 and 2.6 Å, respectively.

**Table 3.6** Data collection and processing statistics.

Values for the outer shell are given in parentheses.

	WT	WT-GlcNAc	D437A (form I)	D437A (form II)
Space group	$P2_1$	$P2_1$	$P2_1$	$P4_32_12$
Unit-cell parameter (Å)				
$a, b, c$ (Å)	90.2, 130.7, 98.5	91.3, 129.6, 100.0	89.4, 129.3, 98.4	165.1, 165.1, 155.2
$\alpha, \beta, \gamma$ (°)	90.0, 113.0, 90.0	90.0, 114.4, 90.0	90.0, 112.2, 90.0	90.0, 90.0, 90.0
Resolution range (Å)	48.55 - 2.37 (2.43 - 2.37)	46.21 - 2.50 (2.56 - 2.50)	48.43 - 2.60 (2.67 - 2.60)	49.50 - 3.0 (3.08 - 3.0)
$V_M$ (Å <sup>3</sup> Da <sup>-1</sup> )	3.59	3.62	3.53	3.55
Solvent content (%)	65.73	66.03	65.22	65.39
Subunits per AU	2	2	2	2
Total No. reflections	577490	342683	436477	569579
No. of unique reflections	84783	72950	63617	43294
Wavelength (Å)	0.99980	0.97889	0.97889	0.97779
Data completeness (%)	99.4 (99.1)	99.4 (99.9)	99.8 (99.7)	99.5 (99.5)
Redundancy	6.6 (7.1)	4.7 (4.5)	6.9 (7.1)	13.2 (13.9)
Mean $\langle I/\sigma(I) \rangle$	9.93 (2.21)	9.71 (3.17)	9.7 (2.18)	12.8 (2.06)
$R_{\text{merge}}$ (%)	14.6 (108.7)	11.8 (53.3)	12.5 (77.5)	21.0 (167.1)
CC (1/2)	99.4 (84.1)	99.2 (85.1)	99.3 (86.5)	99.6 (79.0)



**Figure 3.23** X-ray diffraction images of (A) wild-type *VhGlcNAcase* in the absence of GlcNAc2 and (B) the inactive D437A mutant with resolutions of 2.4 and 2.6 Å, respectively. The X-ray data were collected on beamline PX-II at the Swiss Light Source, Villigen, Switzerland.

### 3.10 Structural determination of the wild-type *VhGlcNAcase* and mutant D437A

#### 3.10.1 Phase determination by molecular replacement

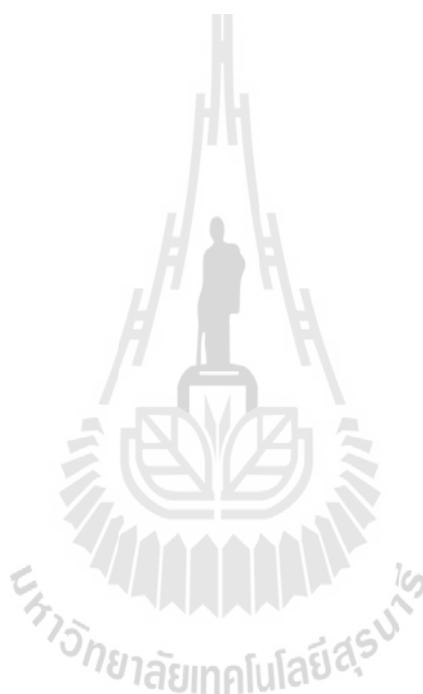
The first data set belonging to wild-type *VhGlcNAcase* that was processed to 2.37 Å was successfully solved by the molecular replacement with Phaser (McCoy *et al.*, 2007) from the CCP4 package using the known structure of  $\beta$ -hexosaminidase from *Arthrobacter aurescens* (PDB code: 3RCN; 35% identity to GlcNAcase from *V. harveyi*; Midwest Center for Structural Genomics, unpublished work) as a model. The preliminary structure of wild-type *VhGlcNAcase* was then

employed to obtain the phases of all the data sets of the mutant D437A and *VhGlcNAcase* in complex with *N*-acetylglucosamine.

### 3.10.2 Refinements and structural determination

Since the preliminary structural model of the D437A mutant from the tetragonal crystal form contained exactly the same dimer as the monoclinic crystal form (RMSD of 0.82 Å), but with much weaker diffraction of the crystals, this crystal form was not further investigated. The crystallographic data and refinement statistics resulting in the finalized model of the *VhGlcNAcase* structures were summarized in Table 3.7. The space group symmetry of all three forms of *VhGlcNAcases* were determined as monoclinic  $P2_1$  with similar unit cell parameters, as summarized in the Table 3.6. The final model of the wild-type *VhGlcNAcase*, which was used as a model of the other data sets, was refined at 2.37 Å resolution with an  $R_{\text{factor}}$  of 23.4%,  $R_{\text{free}}$  of 26.4% with the R.M.S. deviations of bond length of 0.007 Å and bond angle of 1.12°. The electron density map of the 10,333 protein atoms (639 amino acid residues for one molecule in the asymmetric unit) showed a good fit with an average B-factor of 49.0 Å<sup>2</sup>. For a complex structure of *VhGlcNAcase* with GlcNAc, the final  $2F_o - F_c$  map, contoured at 1.0 $\sigma$ , showed a fit of electron density map for GlcNAc (29 ligand atoms) in subsite -1, with an average B-factor of 42.1 Å<sup>2</sup>, whereas an average B-factor of 685 water molecules was 33.2 Å<sup>2</sup>. This complex structure was refined at 2.50 Å resolution with an  $R_{\text{factor}}$  of 20.8%,  $R_{\text{free}}$  of 25.2% with the R.M.S. deviations of bond length of 0.009 Å and bond angle of 1.20°. The ligand-free form of the D437A mutant, was refined at 2.60 Å resolution with an  $R_{\text{factor}}$  of 19.9%,  $R_{\text{free}}$  of 25.0% with the RMSD of bond length of 0.010 Å and bond angle of 1.36°. The electron density map of the D437A mutant containing 10,294 protein atoms (639 amino acid residues for one molecule in

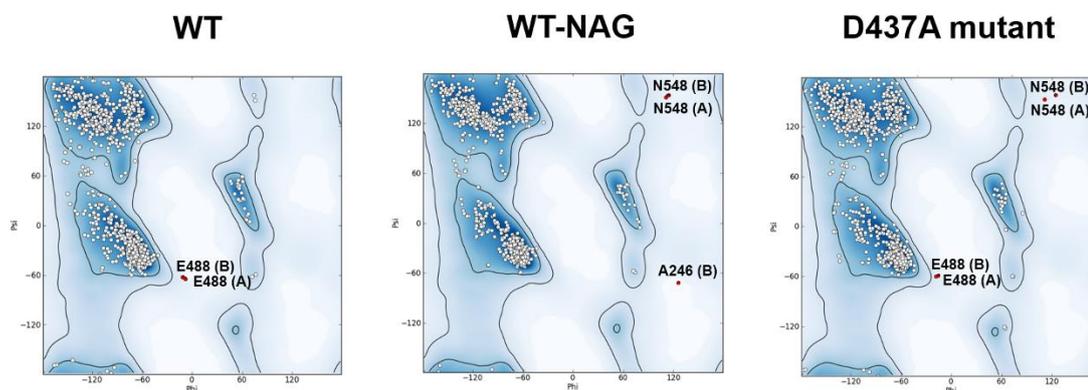
the asymmetric unit) showed a clear electron density map at the position of Asp437 that was mutated to alanine with an average B-factor of  $63.2 \text{ \AA}^2$ . The fitted model was in a good agreement with the calculated model, with a few residue in the outlier regions of the Ramachandran plots such as Glu488 (A,B) in wild-type *VhGlcNAcase*, Ala246 (B) and Asn548 (A,B) in wild-type *VhGlcNAcase* in complex with GlcNAc, Glu488 (A,B) and Asn548 (A,B) in the D437A mutant (Table 3.7 and Figure 3.24)



**Table 3.7** Refinement statistics.

	WT	WT-GlcNAc	D437A mutant
<b>Refinement</b>			
Resolution (Å)	2.37	2.50	2.60
$R_{\text{factor}}$ (%)	23.4 (43.2)	20.8 (29.6)	19.9 (33.2)
$R_{\text{free}}$ (%)	26.4 (42.6)	25.2 (33.7)	25.0 (36.4)
Number of atoms			
Protein	10333	10301	10294
Ligand/ion	-	58	-
Water	579	685	393
RMSD			
Bond lengths (Å)	0.007	0.009	0.010
Bond angles (°)	1.122	1.200	1.364
Average $B$ -factors			
Protein	49.0	36.0	63.2
Ligand	-	42.7	-
Solvent	44.9	33.2	55.5
Ramachandran plot			
Favoured regions (%)	98.05	98.19	97.01
Allowed Region (%)	1.79	1.57	2.68
Outlier regions (%)	0.16	0.24	0.31

Numbers in parentheses are outer shell parameters.



**Figure 3.24** Ramachandran plot for all non-Pro/Gly residues. Plot created using PHENIX.

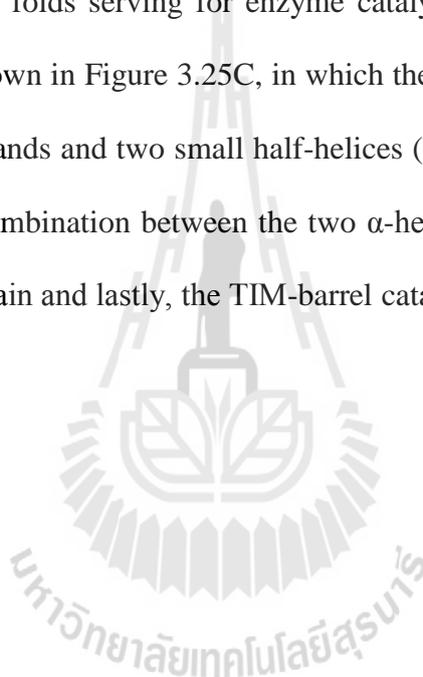
### 3.10.3 The overall structure of *VhGlcNAcase*

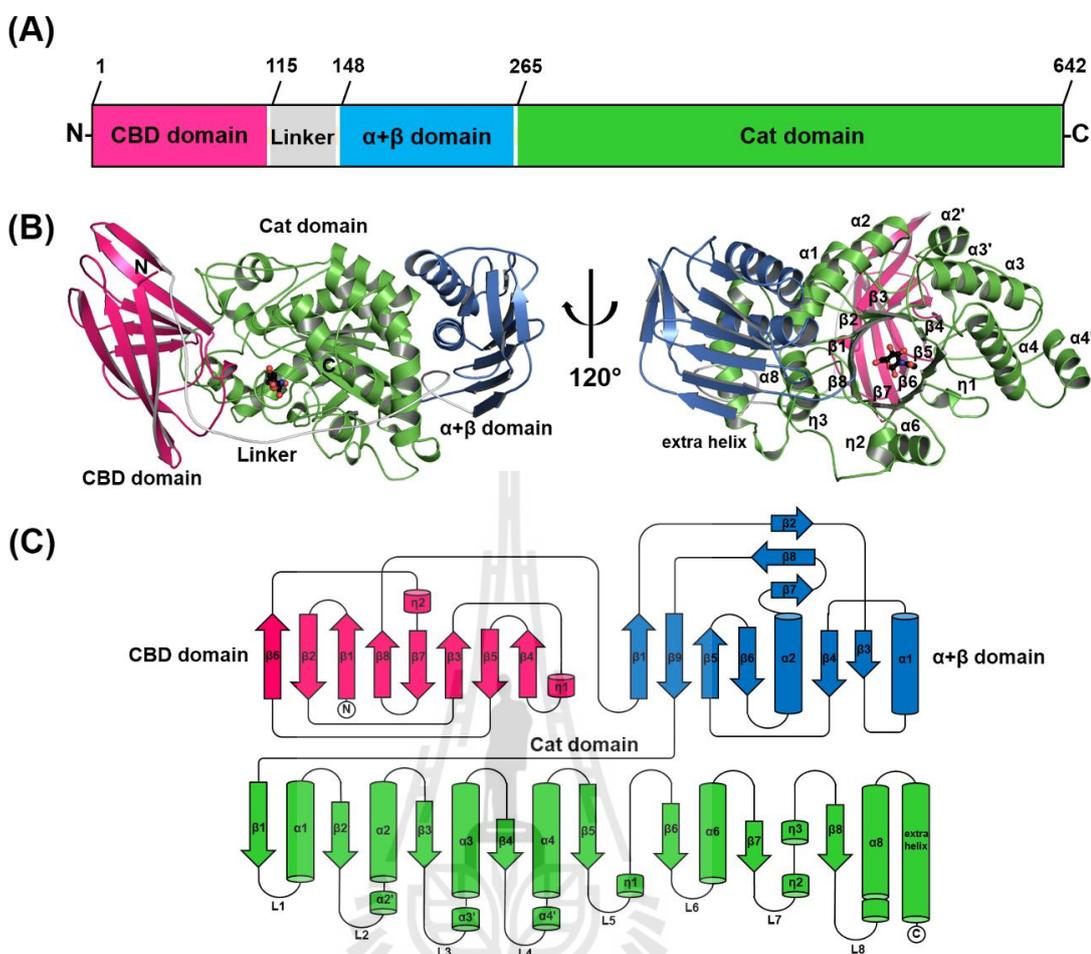
The wild-type and mutated *VhGlcNAcases* were expressed as the C-terminal (His)<sub>6</sub> tagged proteins and there is no disulfide bridge found in *VhGlcNAcase* structure. The three-dimensional structure of 642-residue *VhGlcNAcase* is composed of three distinct domains: a putative N-terminal carbohydrate-binding domain of 114 residues, followed by a relatively long linker (33 residues; residues 115-147) that is tightly bound on the surface of the TIM-barrel catalytic domain and connects the N-terminal carbohydrate-binding domain and the  $\alpha+\beta$  topology domain (residues 148-264), and lastly the C-terminal domain with a TIM-barrel fold (residues 265-642) that contains the highly charged active site serving for the enzyme catalysis. The domain organization of *VhGlcNAcase* was summarized in Figure 3.25A as the N-terminal carbohydrate-binding domain (in pink), the linker (in gray), the  $\alpha+\beta$  topology domain (in skyblue) and the TIM-barrel catalytic domain (in green). The overall structure of *VhGlcNAcase* was shown in Figure 3.25B. As seen in Figure 3.25B,

when rotating the crystal structure of *VhGlcNAcase*, a sugar molecule found in the active site was shown in the middle of the TIM barrel fold of the catalytic domain.

The carbohydrate-binding domain belongs to the carbohydrate-binding module family 2 (CBM2) according to the CAZy database (Lombard *et al.*, 2014). A search with DALI (<http://ekhidna.biocenter.helsinki.fi>) (Holm and Sander, 1993) indeed identifies the CBM2 domains of endoglucanase D of *Clostridium cellulovorans* as closest relative (RMSD 1.6-1.7 Å over 100 residues, 8% sequence identity, PDB codes: 3NDZ, 3NDY and 3ICG; unpublished work), followed by the cellulose-binding CBM2 domain of exo-1,4- $\beta$ -D-glycanase from *Cellulomonas fimi* (RMSD 2.0-2.1 Å over 99-100 residues, 9-10% sequence identity, PDB codes 1EXG and 1EXH, CBM2 domain only) (Xu *et al.*, 1995) and the more distantly related CBM domain of chitobiase which is in a similar location relative to the active site as the CBM2 in our structures (RMSD 2.5-2.6 Å over 113 residues, 10% sequence identity, PDB codes: 1QBB, 1QBA, 1C7T and 1C7S) (Prag *et al.*, 2000; Tews *et al.*, 1996). In the most similar structure of this subdomain, endoglucanase D from *Clostridium cellulovorans* (PDB code 3ICG; unpublished work), the CBM2 domain is attached to the C-terminus of the catalytic TIM barrel domain and is located far away from the active site. The sequence identity is only 13.7%, but the r.m.s.d is quite low with 1.71 Å over 102 residues, indicating a high structural similarity. The second domain,  $\alpha$ + $\beta$  topology domain had the most similar topology with the  $\beta$ -Hex from *Arthrobacter aureescens* (RMSD 1.5 Å over 112 residues, 28% sequence identity, PDB code: 3RCN; unpublished work), followed by the  $\beta$ -Hex from *Bacteroides fragilis* (RMSD 1.6 Å over 108 residues, 25% sequence identity, PDB code 4PYS; unpublished work) and  $\beta$ -Hex from *Paenibacillus* sp. TS12 (RMSD 2.5 Å over 116 residues, 21% sequence

identity, PDB code: 3GH7) (Sumida *et al.*, 2009). In the most similar structure of this subdomain, the  $\alpha+\beta$  topology domain from *Arthrobacter aurescens* (PDB code: 3RCN), is attached to the C-terminus of the catalytic TIM barrel domain and has the structural similarity with the  $\alpha+\beta$  topology domain of GlcNAcase from *Vibrio harveyi* with the very low RMSD of 1.5 Å over 112 residues, indicating a high structural similarity. The third domain of *VhGlcNAcase* is the TIM barrel which is one of the most common protein folds serving for enzyme catalysis. The topology diagram of *VhGlcNAcase* was shown in Figure 3.25C, in which the carbohydrate-binding domain contains of eight  $\beta$ -strands and two small half-helices (here represented as theta). The second domain is a combination between the two  $\alpha$ -helices and nine  $\beta$ -strands called the  $\alpha+\beta$  topology domain and lastly, the TIM-barrel catalytic domain.





**Figure 3.25** Domain organization and overall structure of *VhGlcNAcase*. (A) Three distinct domains of *VhGlcNAcase* were drawn manually and the order of amino-acid sequence was labelled sequentially. (B) Overall structure of GH20 *VhGlcNAcase* consists of three domains. The *N*-terminal carbohydrate-binding domain is presented in dark pink, the  $\alpha$ + $\beta$  domain is presented in skyblue, a linker between the *N*-terminal carbohydrate-binding domain (CBD) and the  $\alpha$ + $\beta$  domain is presented in gray and the TIM-barrel catalytic (Cat) domain is presented in green. The GlcNAc in the active site located in the middle of the TIM-barrel catalytic domain of *VhGlcNAcase* is shown as a black ball-and-stick model with N atoms in blue and O atoms in red. (C) The topology diagram of *VhGlcNAcase* was drawn manually based on its 3D structure analyzed in

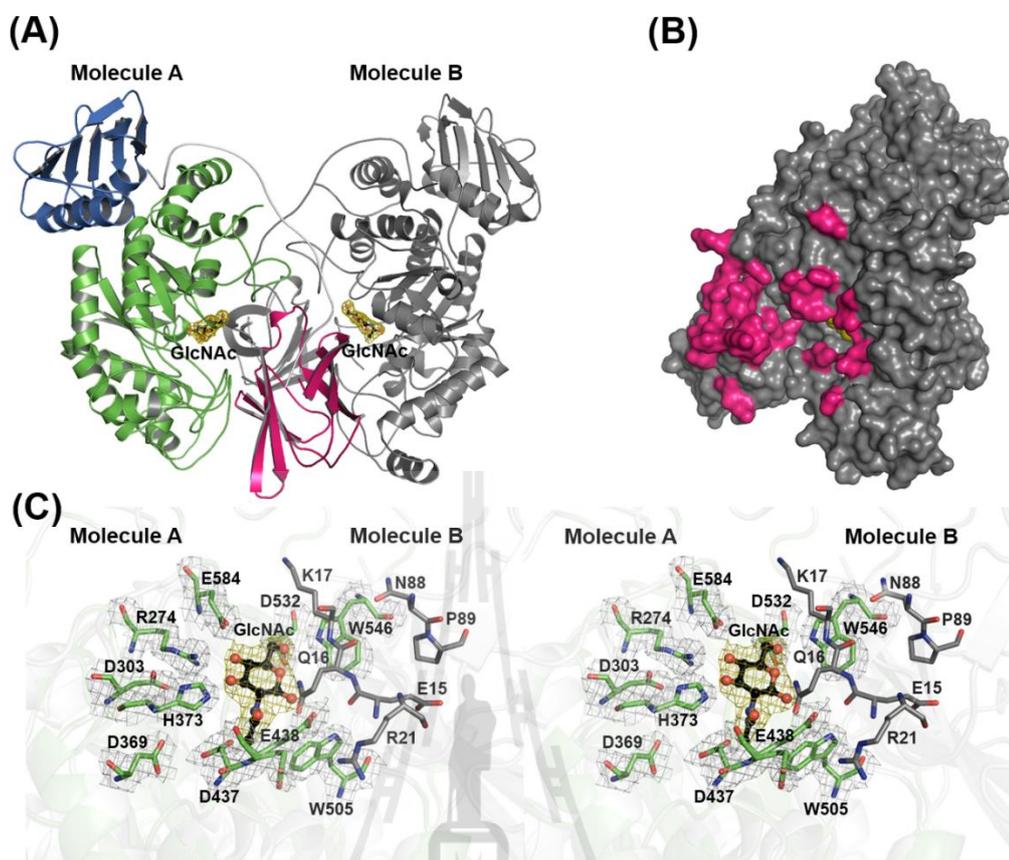
PDBSum server. The  $\beta$ -strand,  $\alpha$ -helix and loop are represented by yellow arrows, green cylinders and black lines, respectively. Very short helices are depicted by green short cylinders labelled ' $\eta$ '. The carbohydrate-binding domain (residues 4-114) contains at most ten  $\beta$ -strands and two short helices. The  $\alpha+\beta$  domain (residues 148-259) is a combination of nine  $\beta$ -strands and two  $\alpha$ -helices. The catalytic domain (residues 292-633) contains eight  $\beta$ -strands as referred to  $\beta 1$  to  $\beta 8$  and the six main  $\alpha$ -helices connecting the  $\beta$ -strands are referred as  $\alpha 1$  to  $\alpha 8$  according to the typical order of a TIM-barrel domain. The segmented helices or short helices are referred as  $\alpha'$ , for instance  $\alpha 2'$  indicating the short helix within the region of the  $\alpha 2$  helix.

#### 3.10.4 Dimer interface of GH20 *VhGlcNAcase*

As seen in Figure 3.26A, the asymmetric unit in the crystal structure of *VhGlcNAcase* contains two identical molecules (RMSD is 0.082 over 566 residues). The stereo view for the experimental electron density map contoured at  $1\sigma$  for a region of the *VhGlcNAcase* in complex with the natural product, GlcNAc shown in Figure 3.10.3A. The GlcNAc<sub>2</sub> was first soaked with wild-type crystals for 30 min (Meekrathok *et al.*, 2015). However, the GlcNAc<sub>2</sub> was then hydrolysed to GlcNAc molecules of which one GlcNAc molecule remained at the active site at the most stable subsite -1, showing a fit to the  $2F_o-F_c$  density map (Figure 3.26A). This affinity is supported by the B-factor of GlcNAc in chain A being low ( $29.4 \text{ \AA}^2$  compared to  $36.1 \text{ \AA}^2$  as overall B factor), indicating high rigidity through strong interactions at the active site. The hydrolyzed GlcNAc in the active site has the sugar ring conformation as  ${}^4C_1$  chair conformation based on the Cremer-Pople parameter calculator (Cremer and Pople, 1975; Jeffrey and Yates, 1979). Surface representation of a single unit of chain B shows

the dimer interface area buried by chain A was determined using the PDBSum server (Figure 3.26B). The hydrophilic and hydrophobic interface area between two molecules, mainly contributed by the *N*-terminal carbohydrate-binding and the catalytic domain, is around 1800 Å<sup>2</sup> and there are 33 residues involved in the dimer interface.

A close-up stereo image of the dimer interface in the active site pocket of *VhGlcNAcase* is shown in Figure 3.26C. The GlcNAc bound to Arg274 is coordinated by the protein charged groups such as Asp303, Asp437, Glu438, Asp532 and Glu584 (Figure 3.26C). Interestingly, the side chain of Glu438 can be fitted into two conformations following to the electron density in which one direction rotates away from the sugar and the other direction rotates to interact with the GlcNAc. This residue, Glu438 is flexible (B factor is 36.6 Å<sup>2</sup>) depending on the presence of sugar. Additionally, the electron density map of the charged and polar residues that interact with GlcNAc is shown in Figure 3.26C. The GlcNAc to Arg274 is stabilized and held in place by charge-charge and polar interactions with Asp271, Asp303, Asp304, Glu305 and water molecules. These polar groups are obviously located at the edge of the binding pocket and accommodate the inner GlcNAc. There are three residues, Glu438, Trp505 and Trp546 from chain A that are affected by the other molecule, chain B. The exit of the negatively charged catalytic pocket is blocked by Ser14-Asn19 loop of the *N*-terminal carbohydrate-binding domain of the neighbor molecule, which has a hydrophilic, but rather uncharged surface. Especially, the side-chains of Gln16 and Lys17 of the carbohydrate-binding domain of monomer B push between GlcNAc and the side chain of Glu438 located at the TIM-barrel catalytic domain of monomer A, resulting in Glu438 of monomer A being located 7.5Å away from the GlcNAc product (Figure 3.26C).



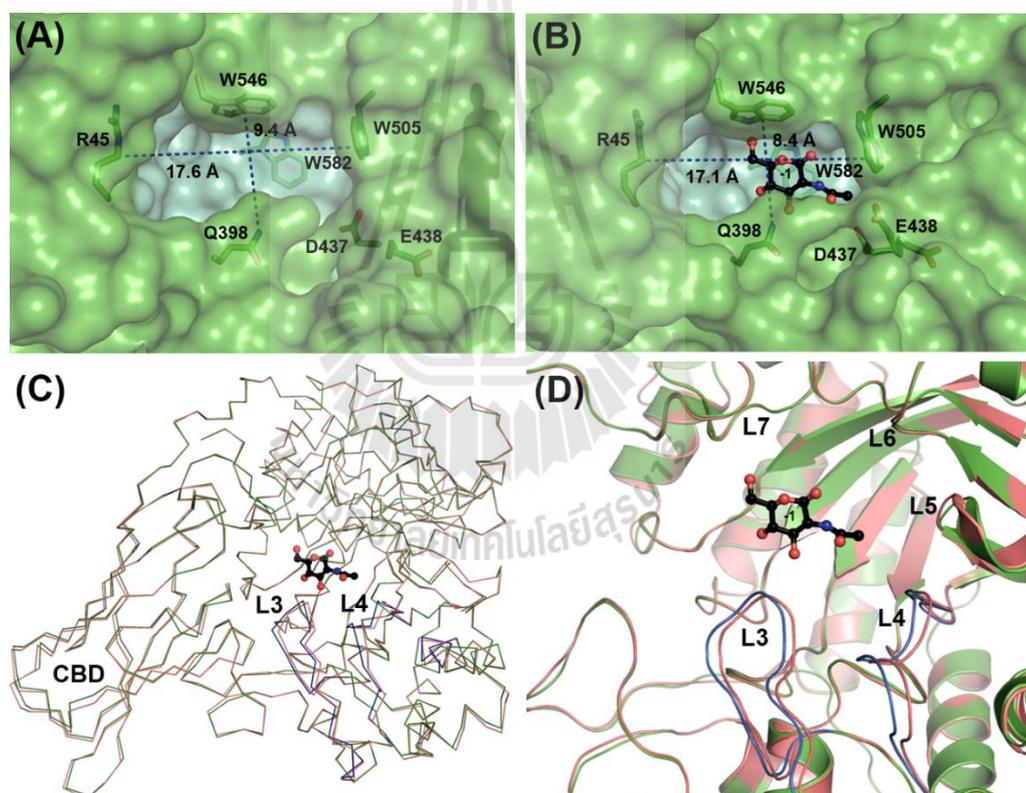
**Figure 3.26** The dimer interface of GH20 *VhGlcNAcase*. (A) The overall structure of GH20 *VhGlcNAcase* presenting two identical molecules per asymmetric unit. The *N*-acetylglucosamine product was found in the active site of each molecule as shown as black ball-and-stick. (B) Surface representation of a single unit of chain B shows the dimer interface between the two molecules in the asymmetric unit. Residues involved in the dimer interface are colored in dark pink, whereas residues not involved in dimerization are colored gray. The GlcNAc in the active site pocket is colored in yellow spheres. (C) Stereo image of the dimer interface showing the interactions from the other molecule B (in gray stick) to molecule A (in green stick). The  $2F_o - F_c$  electron density map for a region of the putative active site is contoured at  $1\sigma$  and complete to 2.5 Å resolution. The refined model of *VhGlcNAcase* in complex with GlcNAc (chain A) is drawn as sticks with the carbon atoms of the active site residues in green, nitrogen in

blue and oxygen in red whereas the GlcNAc is drawn as ball-and-stick with the carbon atoms in black. There are three residues, Glu438, Trp505 and Trp546 from molecule A that are affected by the other molecule, chain B.

### 3.10.5 Sugar binding induces conformational changes

In the ligand-free form, a surface of the substrate binding pocket (in cyan) is shown in Figure 3.27A. The side chain of Glu438, located at loop L4, swings out from the active site, widening the pocket at subsite -1 whereas the side chain of Asp437 moves to the acetamido group of the GlcNAc (Figure 3.27A). Upon GlcNAc binding Asp437 moves toward the acetamido group of GlcNAc, Gln398 and the indole side chain Trp505 rotates from its original position closer to GlcNAc and the side chain of Glu438 also swings into the binding pocket (Figure 3.27B), resulting in the edge of the substrate-binding pocket around subsite -1 narrowing to around 1 Å. As a result of sugar binding, the whole CBD (see Figure 3.27C) is slightly tilted and the volume of the pocket is reduced to the shape of the bound sugar (Figure 3.27B). Superimposition of the apo-*Vh*GlcNAcase structure on the structure in complex with GlcNAc, gives  $C_{\alpha}$  RMSD for 1,278 residues (two chains) of 0.91 Å, respectively (RMSD 0.35 Å<sup>2</sup> over 639 residues/chain A). Figure 3.27C shows the overall structure of *Vh*GlcNAcases in absence and presence of GlcNAc, with the structural changes observed in the carbohydrate-binding domain (CBD) and small conformational changes observed in the loop regions near the surface of the substrate binding pocket. The temperature factors (B factor) of CBD, L3 and L4 are equal to 37.5, 41.2 and 42.5 Å<sup>2</sup>, respectively. A close-up investigation of the GlcNAc-binding residues of the apo enzyme (deep salmon) (Figure 3.27D) in comparison with that of the enzyme in complex with the GlcNAc

(green) (Figure 3.27D) reveals substantial movements of loops L3 (blue) (residues 387-403; RMSD 0.22 Å<sup>2</sup> over 17 residues) and L4 (blue) (residues 434-467; RMSD 0.40 Å<sup>2</sup> over 34 residues) on the surface of subsite -1 (see Figure 3.25C for loop topology). As compared with the native protein (Figure 3.27A-B), four key residues were found to move considerably towards the center of the substrate-binding pocket: Gln398 (part of the middle of loop L3), Asp437 and Glu438 (part of the beginning of loop L4) and Trp505, which is located at L6 (in between β6 and α6).



**Figure 3.27** Conformational changes of *VhGlcNAcase* structures in the absence and presence of GlcNAc. (A) A surface representation of the sugar-binding pocket (cyan) of the unliganded form of *VhGlcNAcase*. (B) A surface representation of the sugar-binding pocket (cyan) of the *VhGlcNAcase*/GlcNAc complex, showing local changes causing the closure of the surface area of the substrate-binding pocket around subsite

-1, relative to the binding pocket in the apo form of *VhGlcNAcase*. (C) A line representation of the structure of ligand-free *VhGlcNAcase* (deep salmon) was superimposed on the structure of *VhGlcNAcase* in complex with GlcNAc (green and blue), showing the movement upon the GlcNAc binding. (D) A closeup of the GlcNAc binding pocket from (C). Q398 located on loop 3 (L3) and D437-E438 located on loop 4 (L4) move towards the GlcNAc.

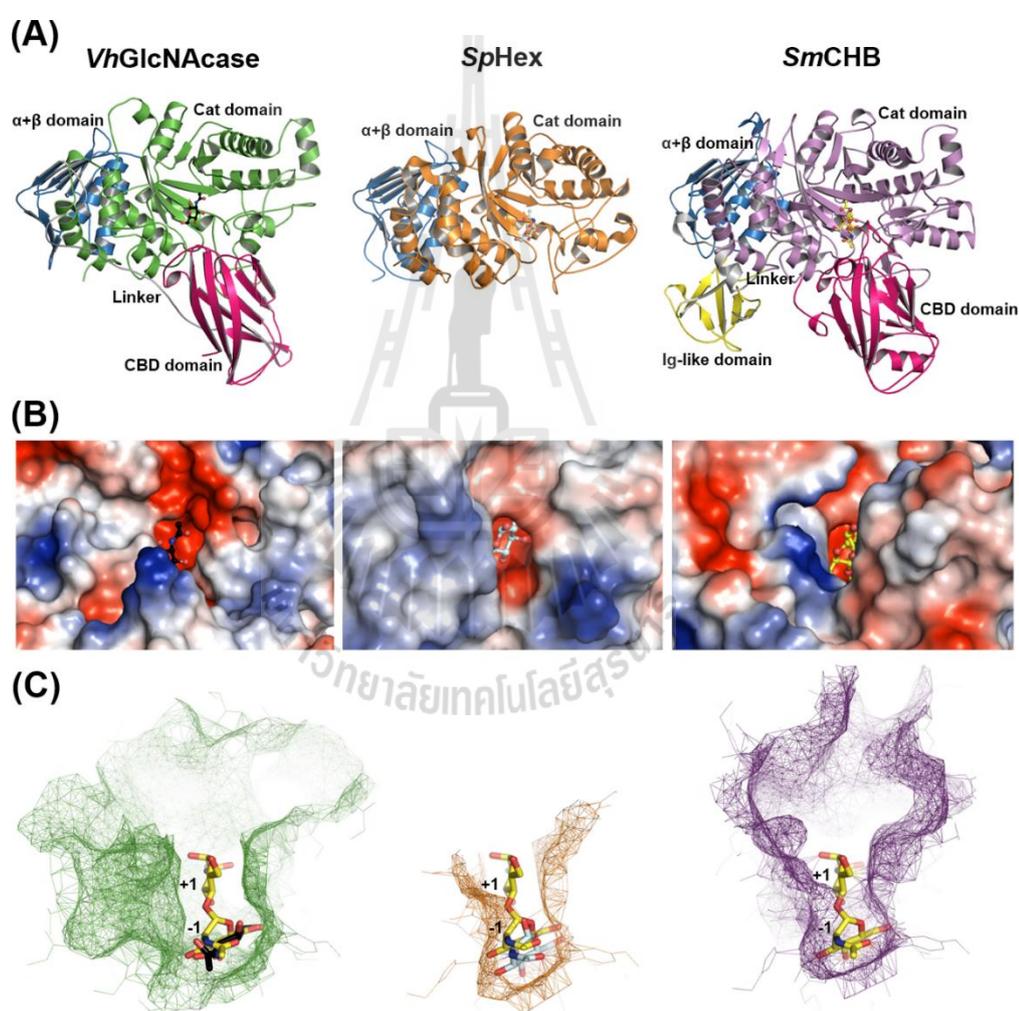
### **3.10.6 Structural comparison of *VhGlcNAcase* with other bacterial GH20 members**

The crystal structure of *VhGlcNAcase* was compared with the other bacterial family 20 glycoside hydrolases, many of which have also been solved in complexed with GlcNAc. A DALI search showed that the *VhGlcNAcase* structure is most similar to the crystal structure of  $\beta$ -*N*-acetylhexosaminidase from *Arthrobacter aurescens* (Z-score = 47.8, RMSD 1.8 Å over 469 residues, 36% sequence identity, PDB code: 3RCN), followed by  $\beta$ -*N*-acetylhexosaminidase from *Streptomyces coelicolor* A3(2) (ScHex) (Z-score = 43.9, RMSD 2.1 Å over 445 residues, 30% sequence identity, PDB code 4C4D) (Thi *et al.*, 2014) and  $\beta$ -*N*-acetylhexosaminidase from *Streptomyces plicatus* (SpHex) (Z-score = 43.6, RMSD 2.1 Å over 446 residues, 29% sequence identity, PDB code: 1M01) (Williams *et al.*, 2002). However, there is no functional and structural analysis of *Arthrobacter aurescens*  $\beta$ -*N*-acetylhexosaminidase yet. Therefore, the well-characterized *Streptomyces plicatus*  $\beta$ -*N*-acetylhexosaminidase was selected for further comparative study with *VhGlcNAcase*. In addition, chitobiase from *Serratia marcescens* (SmCHB) also showed high structural similarity with *VhGlcNAcase* with the Z-score = 37.4, RMSD 2.4 Å

over 611 residues, 28% sequence identity, PDB code: 1QBA) (Tews *et al.*, 1996) and was further refined by complexing with the natural substrate, GlcNAc<sub>2</sub> (PDB code: 1QBB).

The domain organization of GH20 enzymes is different according to their amino acid sequence, length and arrangement. The crystal structures of *VhGlcNAcase*, *SpHex* and *SmCHB* are compared in terms of domain organization in Figure 3.28A. The crystal structure of *VhGlcNAcase* consists of three domains, as mentioned previously. The *SpHex* has two domains comprising of 506 amino acids that fold into the *N*-terminal domain (I) called  $\alpha+\beta$  topology domain and the catalytic domain (II) containing a  $(\beta/\alpha)_8$  barrel at the *C*-terminus (Mark *et al.*, 2001), whereas the *SmCHB* consists of four domains designated as the *N*-terminal carbohydrate-binding domain (I),  $\alpha+\beta$  topology domain (II), TIM-barrel catalytic domain (III) and the *C*-terminal immunoglobulin-like domain (IV) (Tews *et al.*, 1996). In Figure 3.28B, the electrostatic surface potential of *VhGlcNAcase*, *SpHex* and *SmCHB* displayed in the same orientation with Figure 3.28A was calculated in PyMOL. The sugar-binding pocket of GH20 proteins is negatively charged, and these acidic groups play an important role both for proton donation and acceptance in catalysis and substrate coordination through hydrogen bonds. As expected, the sugar-binding pockets of *SpHex* and *SmCHB* are conserved in charge and size through all three proteins. Here, *SpHex* is the most similar to *VhGlcNAcase* in terms of amino acid sequence and structural folding. In Figure 3.28C, a topology of the substrate binding pocket of each bacterial GlcNAcase structure is shown in the same orientation and its substrate (GlcNAc) of *VhGlcNAcase* and *SpHex* were superimposed with the non-reducing GlcNAc residue of GlcNAc<sub>2</sub> from *SmCHB*. The substrate binding pockets of

*VhGlcNAcase*, *SpHex* and *SmCHB* have estimated dimensions (width  $\times$  length  $\times$  depth) of 9.4 Å  $\times$  17.6 Å  $\times$  12.2 Å, 7.7 Å  $\times$  10.3 Å  $\times$  11.0 Å and 12.2 Å  $\times$  14.3 Å  $\times$  19.2 Å, respectively. The active sites of *VhGlcNAcase* and *SmCHB* are the tunnel-like structure, whereas the active site of *SpHex* is a small and narrow pocket for accommodating small chitooligosaccharides.

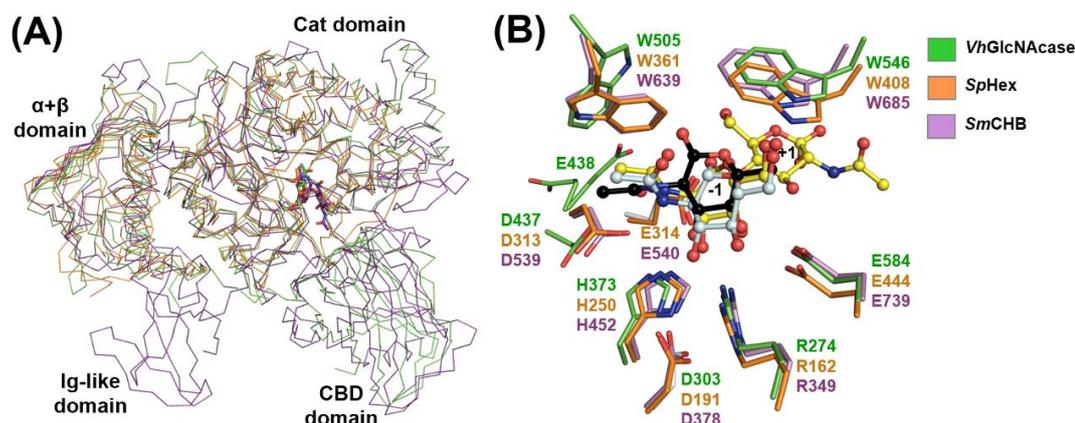


**Figure 3.28** Overview of structural comparison of *VhGlcNAcase* with other bacterial GH20 structures. (A) The domain organization of related GH20 enzymes was shown as cartoon representation. Each figure was kept fixed in the same orientation. (B) The molecular surface calculated from the crystal structures of *VhGlcNAcase*, *SpHex* (PDB

entry: 1M01) and *Sm*CHB (PDB entry: 1QBB), are compared by electrostatic surface potential, calculated in PyMOL. The overall bacterial GH20 structures are neutral (white surface) whereas the active site of all three structures is more acidic (red surface) and narrow for accommodating small oligosaccharides. (C) The topologies of the substrate binding pockets of the GH20 GlcNAcases generated using the PDBSum server and displayed in PyMOL. Each structure was kept in the same orientation and its substrate (GlcNAc) of *Vh*GlcNAcase and *Sp*Hex were superimposed with GlcNAc<sub>2</sub> from *Sm*CHB.

Superimposition of three bacterial GH20 GlcNAcases, *Vh*GlcNAcase, *Sp*Hex (PDB code: 1M01) (Mark *et al.*, 2001) and *Sm*CHB (PDB code 1QBB) (Tews *et al.*, 1996) was displayed as a line representation (Figure 3.29A). Only the  $\alpha+\beta$  topology domain and the TIM-barrel catalytic domain showed conservation for all three bacterial GlcNAcases. A close up of the sugar-binding residues of *Vh*GlcNAcase, *Sp*Hex and *Sm*CHB is shown in Figure 3.29B. As compared with *Sp*Hex (orange stick) in complex with GlcNAc (pale cyan ball-and-stick) and *Sm*CHB (violet stick) in complex with GlcNAc<sub>2</sub> (yellow ball-and-stick) in Figure 3.29B, the active-site residues of *Vh*GlcNAcase (green stick) with GlcNAc (black ball-and-stick) showed that the sugar ring of *Vh*GlcNAcase is well aligned with a chair conformation of -1GlcNAc of *Sp*Hex with the exception of the C2-acetamido group and well aligned with the sugar ring of -1GlcNAc of *Sm*CHB but in a different conformation (Figure 3.29B). At subsite -1, the GlcNAc ring in *Vh*GlcNAcase adopts the chair conformation whereas the GlcNAc ring in *Sm*CHB adopts a boat conformation. However, the outer GlcNAc from GlcNAc<sub>2</sub> of *Sm*CHB showed a chair conformation at subsite +1.

There are conformational differences in residues Arg274, His373, Asp437, Glu438, Trp505, Trp546 and Glu584 as compared to *SpHex* and *SmCHB* (Figure 3.29B). The most significant difference is the position of D437/E438 compared to the respective amino acids in the homologues, as due to the contact between the two protein molecules in the asymmetric unit in the crystal of *VhGlcNAcase*, D437/E438 is pushed away from the sugar binding pocket. Since the sugar rings of GlcNAc were slightly shifted, the side chain of Asp437 (equivalent to Asp313 in *SpHex*) is also slightly displaced towards the acetamido group of GlcNAc to orient the substrate and stabilize the transition states in the substrate-assisted catalytic mechanism. The side chains of Glu438 can swing in and out, depending on the substrate as mentioned previously, while the side chain of Glu314 in *SpHex* and Glu540 in *SmCHB* has only one conformation around 3.6-3.7 Å from the acetamido group of GlcNAc. However, as the Glu438 side chain of *VhGlcNAcase* forms a salt bridge with the guanidyl group of Arg21 of the other molecule in the protein interface, so the large distance to the GlcNAc is most likely a result of the dimer formation in the crystal and does not represent active the conformation of the monomer in solution. The local conformation changes of the indole ring of Trp505 to be oriented perpendicular to Trp361 in *SpHex* and Trp639 in *SmCHB*, which may suggest it can act as a lid residue that controls the entrance of the substrate due to its flexibility upon substrate binding.



**Figure 3.29** Superimposition of bacterial GH20 GlcNAcases. (A) A line representation of the structure of ligand-bound *VhGlcNAcase* (green) complexed with GlcNAc (black) was superimposed on the structures of *SpHex* (orange) complexed with GlcNAc (cyan) and *SmCHB* (violet) complexed with GlcNAc<sub>2</sub> (yellow). (B) A close up from figure (C) of the active site residues of *VhGlcNAcase* (green stick), *SpHex* (orange stick) and *SmCHB* (violet stick) that participate in sugar binding. The GlcNAc substrates of *VhGlcNAcase* and *SpHex* are shown as black and pale cyan ball-and-stick models, respectively whereas the GlcNAc<sub>2</sub> of *SmCHB* is shown as yellow ball-and-stick model.

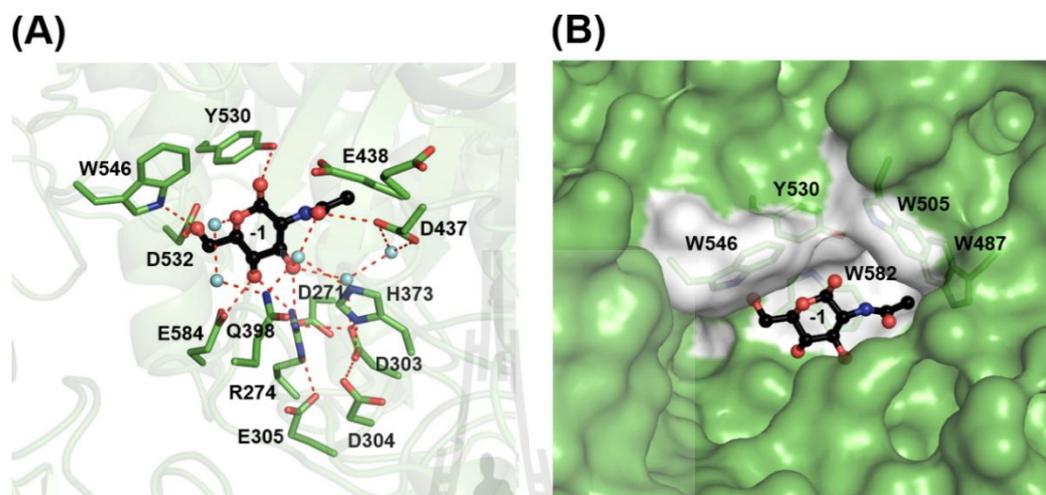
### 3.10.7 Specific interactions of *VhGlcNAcase* with sugar

The substrate-binding pocket of *VhGlcNAcase* was investigated using the complex with GlcNAc. Figure 3.30 shows the stereo image of the amino acid side-chains and water molecules that form the hydrogen bonding network with the GlcNAc in the active site, located at the middle of the TIM-barrel structure of the catalytic domain at subsite -1. The hydrolytic product, GlcNAc, derived from GlcNAc<sub>2</sub> and the DE pair of Asp437-Glu438 are in a position close to the  $\beta$ -1,4 glycosidic oxygen, which is the ester bond to be cleaved. One of the water molecules located above the sugar ring of GlcNAc may be the incoming water that acts as the nucleophile attacking at the

anomeric center of the cyclic intermediate to produce a product with retained configuration. The non-reducing GlcNAc is tightly anchored via hydrogen bonds at the OH3 and OH4 groups of the GlcNAc to Arg274, which is held in place by polar interactions with Asp271, Asp303, Asp304, Glu305 and water molecules. The terminal amines of the guanidino group of Arg274 make hydrogen bonds to Asp271 and Asp303, whereas the imine forms a hydrogen bond with Glu305. The OH4 and OH6 groups of GlcNAc are stabilized by Glu584 and Asp532, respectively. The carboxyl group of Asp303 forms hydrogen bonds with Asp304, Arg274 and a water molecule coordinated with the GlcNAc substrate. The imidazole ring of His373 located between the acidic groups coordinates via a hydrogen bond with Asp303 and indirectly interacts with the GlcNAc through a water molecule. The side chain of Asp437 is within 3.1 Å of the acetamido group of the GlcNAc and may play an important role involved in substrate-assisted mechanism. The side-chain of Glu438 has two conformations, one swinging toward the GlcNAc while the other conformation rotates away from the GlcNAc, similar to the apo form presented in the electron density plot. Considering the conformation rotating toward the GlcNAc, Glu438 faces towards the oxygen atom of the glycosidic linkage, and may be the catalytic acid-base residue (although due to the crystal packing the D437-E438 containing loop is pushed away from the GlcNAc binding pocket).

The hydrophobic interactions between GlcNAc and aromatic/hydrophobic residues in the active site surrounding the subsite -1 were further analyzed as shown in Figure 3.30B. Five aromatic residues, Trp487, Trp505, Trp546, Trp582 and Tyr530, create the active pocket wall in the active site. Trp582 was found to stack directly with the plane of the pyranose ring of the GlcNAc at subsite -1. The interactions

of GlcNAcs with the binding residues in the substrate binding pocket of GH20 *Vh*GlcNAcase, *Sp*Hex and *Sm*CHB are summarized in Table 3.8.



**Figure 3.30** Specific interactions within the substrate binding pocket of *Vh*GlcNAcase.

(A) The image of amino acid side-chains (in green stick) and water molecules that form the hydrogen bonding network with the GlcNAc (in black ball-and-stick) in the active site. Water molecules are shown as cyan spheres and hydrogen bonds are shown as red dashed lines. (B) The image of surface representations of the aromatic side chains involving the hydrophobic interactions with the sugar molecule around the -1GlcNAc. The protein surface is green with hydrophobic patches shown as white surface and green stick models of the corresponding residues.

**Table 3.8** Summary of the interactions of GlcNAcs with the binding residues in the substrate binding pocket of GH20 *VhGlcNAcase*, *SpHex* and *SmCHB*

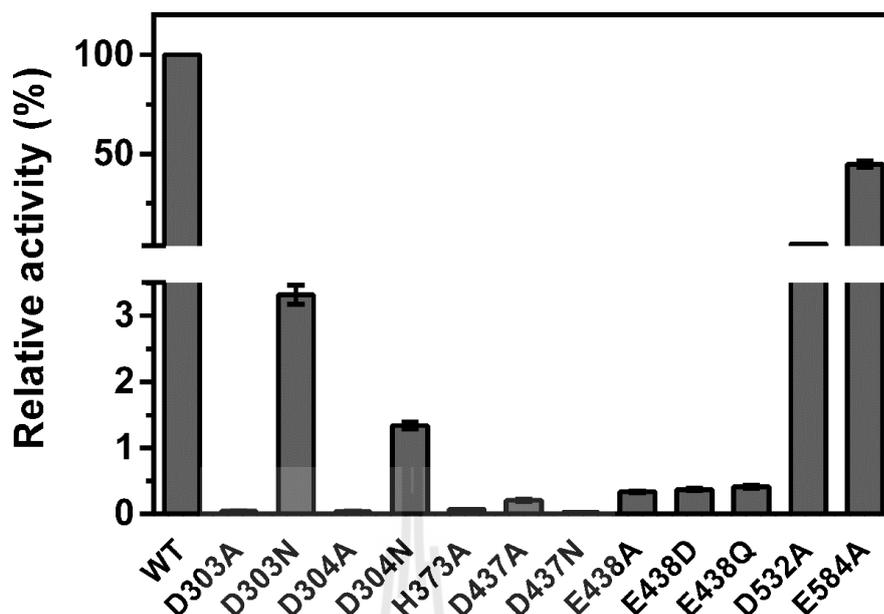
Subsite	GH20 enzyme		
	<i>VhGlcNAcase</i>	<i>SpHex</i>	<i>SmCHB</i>
-1	<u>R274</u> , <u>Q398</u> , <u>D437</u> , <b>W487</b> , <b>W505</b> , <b>Y530</b> , <u>D532</u> , M533, <b>W546</b> , <b>W582</b> , <u>E584</u>	<u>R162</u> , <u>D313</u> , <u>E314</u> , <b>W344</b> , <b>W361</b> , <b>Y393</b> , <u>D395</u> , M396, <b>W408</b> , <b>W442</b> , <u>E444</u>	<u>R349</u> , <u>Q494</u> , <u>D539</u> , <u>E540</u> , <b>W616</b> , <b>W639</b> , <b>Y669</b> , <u>D671</u> , <b>F672</b> , <b>W685</b> , <b>W737</b> , <u>E739</u>
+1		<u>E314</u>	<u>E540</u> , <b>W685</b>

Normal-and-underlined: hydrogen bonding residues

Bold: hydrophobic interaction residues

### 3.11 Specific hydrolyzing activity of *VhGlcNAcase* and its mutants

The specific hydrolyzing activity of *VhGlcNAcase* was measured with *pNP*-GlcNAc as substrate following the previous assay (Suginta *et al.*, 2010). Here, the relative hydrolytic activity of wild-type *VhGlcNAcase* and its mutants was assayed against 125  $\mu$ M *pNP*-GlcNAc for 10 min (Figure 3.31). All mutants showed a decrease in hydrolyzing activity against *pNP*-GlcNAc. Mutations of the expected catalytic DE pair, D437-E438 showed a strongly decreased activity as compared to the wild-type enzyme. The mutant D437N showed the most severe effects on hydrolyzing activity, having essentially no activity against *pNP*-GlcNAc (0.02% remaining activity) whereas the mutant D437A showed a less drastic decrease in hydrolyzing activity (0.21% remaining activity). With mutation of E438, the mutants E438A, E438D and E438Q showed decreased activity of about 0.34%, 0.37% and 0.41% of wild-type, respectively. In contrast, the mutant E584A showed a high remaining activity of 45%. Strikingly, the DD pair, D303-D304 showed a strongly decreased hydrolyzing activity when this DD pair was mutated to alanine (0.04% remaining activity). However, mutations to asparagine revealed that both D303N and D304N still had remaining activity of 3.3% and 1.3%, respectively. Additionally, the mutant H373A showed a drastic decrease in hydrolytic activity as compared with the wild-type enzyme.



**Figure 3.31** Specific hydrolyzing activity of *VhGlcNAcase* and its mutants. The enzymatic activity was determined as described in Materials and Methods using the colorimetric assay against 125  $\mu$ M *pNP-GlcNAc* at 405 nm. All specific hydrolyzing activities were normalized to the hydrolytic activity of wild-type *VhGlcNAcase*, set to 100%. Bar graphs are average values  $\pm$  SD obtained from experiments performed in triplicate.

### 3.12 Kinetic parameters of *VhGlcNAcase* and its mutants

The initial velocities of *VhGlcNAcase* and its mutants were measured for kinetic experiments. The kinetic parameters of the hydrolytic activity of wild-type *VhGlcNAcase* and its mutants were analyzed based on Michaelis-Menten kinetics. As shown in Table 3.9, all mutants except E438 lost the enzymatic activity against *pNP-GlcNAc*, which is reflected by the  $k_{cat}/K_m$ . Mutations of the expected catalytic DE pair, D437-E438, showed a strongly decreased activity as compared with the wild-type enzyme. Mutations of D437 increase the  $K_m$  against *pNP-GlcNAc* about 4.2-fold and

3.7-fold for the mutants D437A and D437N, respectively. The mutant D437A had very low  $k_{cat}$ , yielding the very low  $k_{cat}/K_m$  of  $0.36 \text{ s}^{-1} \text{ mM}^{-1}$  whereas the mutant D437N had high  $K_m$  and the lowest  $k_{cat}$ , yielding the lowest  $k_{cat}/K_m$  of  $0.04 \text{ s}^{-1} \text{ mM}^{-1}$  towards *p*NP-GlcNAc. The reduction in catalytic activity of the mutants D437A and D437N are the result of the changed charge of the amino acid, which results in the destabilization of reaction intermediates in the transition states. Mutations of E438 also showed that the mutants E438A, E438D and E438Q had a slightly increased  $K_m$  (correlating to a decreased binding affinity) but strongly decreased hydrolyzing activity against *p*NP-GlcNAc. The relative  $k_{cat}/K_m$  values of E438A, E438D and E438Q were 0.28%, 0.30% and 0.35%, respectively. Together with its 3D structure and presence next to the cleavage site, this suggests that E438 is important as the catalytic residue. On the other hand, the mutant E584A showed an overall catalytic activity of 32% that of wild-type, suggesting that this residue is not essential for enzyme catalysis. Surprisingly, mutations of the DD pair showed decreased activity to various extents. The mutants D303A and D304A had a drastically decreased hydrolyzing activity (non-detectable activity). However, mutations to asparagine revealed that both D303N and D304N still had remaining activity that is reflected by  $k_{cat}/K_m$  of 1.8% and 0.75%, respectively.

**Table 3.9** Kinetic parameters of wild-type and mutants of *VhGlcNAcase*.

<i>VhGlcNAcase</i> mutants	$K_m^a$ ( $\mu\text{M}$ )	$k_{\text{cat}}$ ( $\text{s}^{-1}$ )	$\frac{k_{\text{cat}}}{K_m}$ ( $\times 10^3 \text{ s}^{-1} \text{ M}^{-1}$ )
Wild-type	$92 \pm 6$	$28.3 \pm 0.6$	306 (100) <sup>b</sup>
D303A	n.d. <sup>c</sup>	n.d.	n.d.
D303N	$325 \pm 30$	$1.8 \pm 0.08$	5.45 (1.78)
D304A	n.d.	n.d.	n.d.
D304N	$236 \pm 11$	$0.54 \pm 0.01$	2.30 (0.75)
H373A	$341 \pm 36$	$0.039 \pm 0.002$	0.11 (0.04)
D437A	$390 \pm 32$	$0.14 \pm 0.006$	0.36 (0.12)
D437N	$339 \pm 47$	$0.015 \pm 0.001$	0.04 (0.01)
E438A	$128 \pm 16$	$0.11 \pm 0.005$	0.87 (0.28)
E438D	$138 \pm 13$	$0.13 \pm 0.004$	0.91 (0.30)
E438Q	$120 \pm 13$	$0.13 \pm 0.005$	1.08 (0.35)
D532A	$153 \pm 9$	$1.6 \pm 0.03$	10.2 (3.32)
E584A	$178 \pm 20$	$17.5 \pm 0.7$	98.1 (32.0)

The hydrolysis of *p*NP-GlcNAc at varying concentrations of 25-500  $\mu\text{M}$  was carried out with 0.1  $\mu\text{g}$  *VhGlcNAcase* in 100 mM sodium phosphate buffer, pH 7.0 at 37°C for 10 min, and the reaction was then terminated with 100  $\mu\text{l}$  of 3 M  $\text{Na}_2\text{CO}_3$  and measured at  $A_{405}$ . Release of *p*NP was converted to molar quantities using a calibration curve of *p*NP (0-20 nmol). The kinetic values ( $K_m$ ,  $k_{\text{cat}}$  and  $k_{\text{cat}}/K_m$ ) were determined by

nonlinear regression using GraphPad Prism software (GraphPad Software Inc., San Diego, CA).

<sup>a</sup> Kinetic assays were performed with *p*NP-GlcNAc as described in Materials and methods. Results are average of three independent experiments.

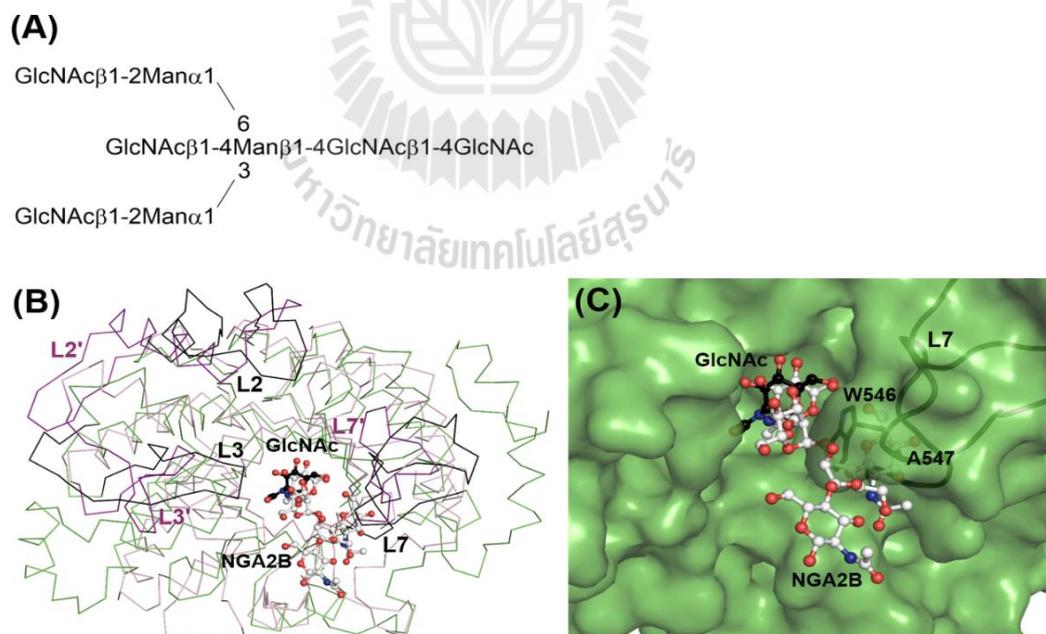
<sup>b</sup> Numbers in brackets reveal relative  $k_{cat}/K_m$  values of the generated mutants by comparing with the wild-type value (set to 100).

<sup>c</sup> Non-detectable activity.

### 3.13 Substrate specificity analysis

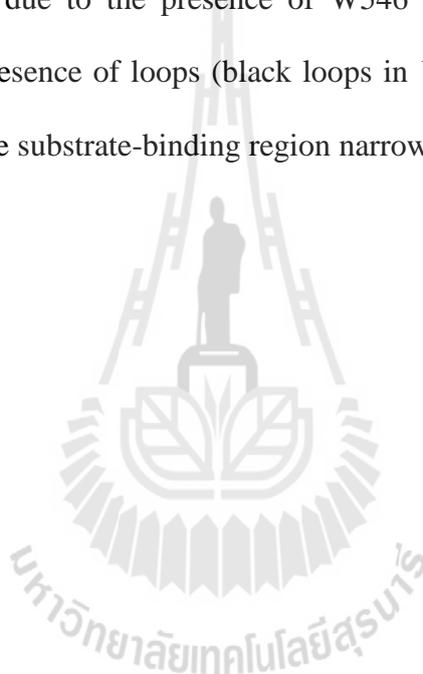
A DALI search of the protein structure showed that the topology of *Vh*GlcNAcase closely matches with other  $\beta$ -*N*-acetylhexosamidases. Interestingly, this includes a homologue of *Vh*GlcNAcase, *Streptococcus pneumoniae* exo- $\beta$ -*N*-acetylglucosaminidase (StrH) (RMSD 2.8 Å over 286 residues, 16% sequence identity, PDB code: 2YLA) (PluviGlcNAce *et al.*, 2011) that is involved in the complete degradation of *N*-linked glycans of its human host (King, 2010). The presence of this enzyme enables degradation of a component of the extracellular matrix and facilitates invasion of the host tissue by the pathogenic bacterium. However, the *Vh*GlcNAcase cannot hydrolyze a broader range of substrates such as *N*-linked glycans or gangliosides (Suginta *et al.*, 2010). Therefore, the crystal structure of *Vh*GlcNAcase was compared with the StrH in complex with the branched glycan *N*-acetyl-D-glucosamine  $\beta$ -1,2-D-mannose  $\alpha$ -1,3 (*N*-acetyl-D-glucosamine  $\beta$ -1,2-D-mannose  $\alpha$ -1,6(*N*-acetyl-D-glucosamine  $\beta$ -1,4))-D-mannose  $\beta$ -1,4-*N*-acetyl-D-glucosamine (NGA2B) (Figure 3.32A) for the inspection of substrate specificity. The TIM-barrel catalytic domains of *Vh*GlcNAcase and StrH (PDB code: 2YLA) were superimposed as shown in Figure

3.32B. The long loops of *VhGlcNAcase* (in green and black) connecting  $\beta 2$  to  $\alpha 2$  (L2; residues 315-343),  $\beta 3$  to  $\alpha 3$  (L3; residues 386-406) and  $\beta 7$  to  $\beta 8$  (L7; residues 525-551) are significantly different with the loops of StrH (light pink and purple) that has the loops connecting residues 681-712 (L2'), residues 754-774 (L3') and residues 898-914 (L7'), respectively (Figure 3.32B). Loop L3 has the highest thermal motion with B factor around  $40.6 \text{ \AA}^2$ , followed by loops L2 and L7 that have B factor approximately  $34.3$  and  $26.9 \text{ \AA}^2$ , respectively. Overhanging loops L2, L3 and L7 of *VhGlcNAcase* (in black wire) protrude into the active site of StrH complexed with NGA2B, causing the substrate-binding region to be narrower (Figure 3.32B). In particular, Trp546 and Ala547 located on loop 7 (L7) in *VhGlcNAcase* clash with the NGA2B substrate of StrH (Figure 3.32C).



**Figure 3.32** Comparison of substrate specificity based on the 3D structures. (A) Chemical structure of NGA2B. (B) A line representation of *VhGlcNAcase* (green and black) was superimposed with the StrH (pale pink and purple) showing the different

orientations of the three overhanging loops as depicted in L2, L3 and L7 for *VhGlcNAcase* and L2', L3' and L7' for StrH, respectively. The GlcNAc in the active site of *VhGlcNAcase* is represented by a black ball-and-stick model whereas the NGA2B in the active site of StrH is represented by a white ball-and-stick model. (C) Superposition of *VhGlcNAcase* (green surface) complexed with GlcNAc (black ball-and-stick) and StrH complexed with NGA2B (white ball-and-stick) showed a steric clash to the NGA2B due to the presence of W546 and A547 on loop 7 (L7) of *VhGlcNAcase*. The presence of loops (black loops in *VhGlcNAcase*) protruding into the active site make the substrate-binding region narrower in *VhGlcNAcase*.



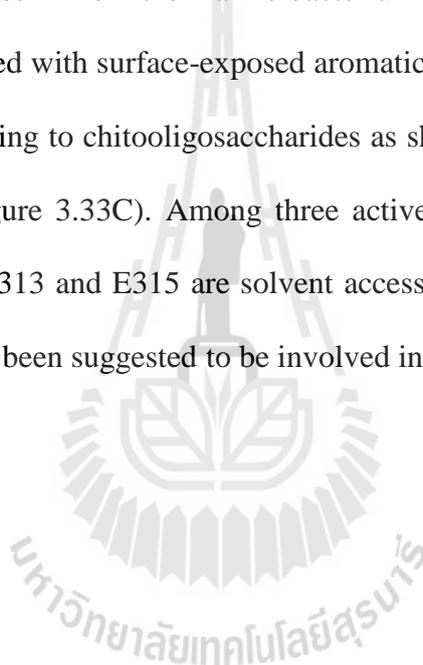
## PART III

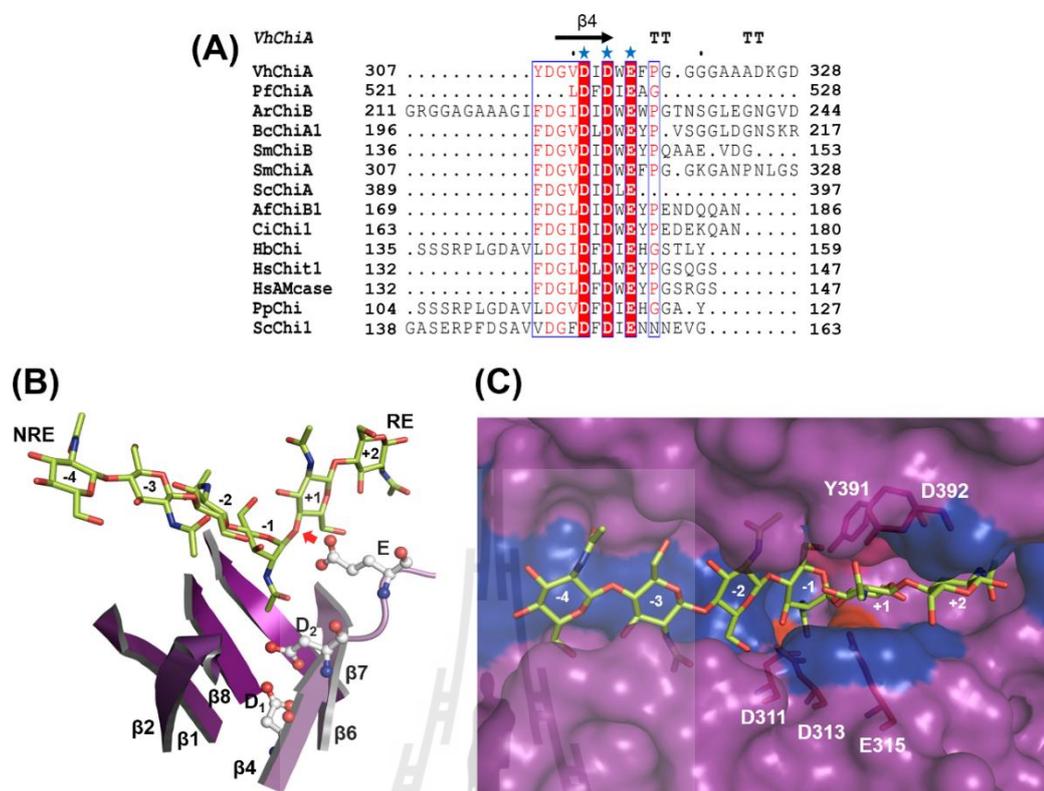
### pK<sub>a</sub> CALCULATIONS OF GH18 CHITINASES

#### 3.14 Sequence analysis and the catalytic residue of chitinases

For the purpose of this study, we have used 14 different chitinases for which we were able to retrieve X-ray structures, namely: *P. furiosus* chitinase (*PfChiA*); *Arthrobacter* sp. chitinase B (*ArChiB*); *B. circulans* WL-12 chitinase A1 (*BcChiA1*); *S. marcescens* chitinase B (*SmChiB*); *S. marcescens* chitinase A, (*SmChiA*); *S. coelicolor* chitinase A (*ScChiA*); *V. harveyi* chitinase A, (*VhChiA*); *A. fumigatus* chitinase B1 (*AfChiB1*); *C. immitis* chitinase (*CiChi1*); *H. brasiliensis* hevamine A (*HbChi*); *H. sapiens* chitotriosidase (*HsChit1*); *H. sapiens* acidic mammalian chitinase (*HsAMCase*); *P. platycephala* endochitinase (*PpChi*) and *S. cerevisiae* chitinase 1 (*ScChi1*). First we carried out multiple sequence alignment of the chosen chitinases (Figure 3.14.1A). The alignment showed that the *V. harveyi* chitinase A (*VhChiA*) sequence (Suginta *et al.*, 2000) is closest to the full sequence of *S. marcescens* chitinase A (*SmChiA*, 36% similarity) (Jones *et al.*, 1986), followed by *B. circulans* WL-12 chitinase A1 (*BcChiA1*, 16% similarity) (Watanabe *et al.*, 1990), *Arthrobacter* sp. chitinase B (*ArChiB*, 15% similarity) (Lonhienne *et al.*, 2001), *A. fumigatus* chitinase B1 (*AfChiB1*, 15% similarity) (Jaques *et al.*, 2003) and *S. marcescens* chitinase B (*SmChiB*, 14% similarity) (Brurberg *et al.*, 1995). As shown in Figure 3.33A, the multiple sequence alignment of family 18 chitinases showed a completely conserved Dx Dx E sequence motif that has been suggested to play a concerted role in the catalysis

of chitinases (Lu *et al.*, 2002; Terwisscha van Scheltinga *et al.*, 1994). For example, the acidic groups of Asp311-Asp313-Glu315 in *VhChiA* are completely aligned with the Asp311-Asp313-Glu315 in *SmChiA*. The DxDxE sequence motif that spans strand  $\beta$ 4 of the TIM barrel will be referred here as D<sub>1</sub>, D<sub>2</sub> and E (Figure 3.33B). The residue ‘E’ located on the flexible loop after strand  $\beta$ 4 is the catalytic proton donor to the glycosidic oxygen between the subsite -1 and +1 of GlcNAcs for a bond cleavage (red arrow, Figure 3.33B). Chitinase A from the marine bacterium *Vibrio harveyi*, has a long and deep catalytic cleft lined with surface-exposed aromatic residues (blue surface), which is responsible for binding to chito oligosaccharides as shown in a complex of *VhChiA* and chitohexaose (Figure 3.33C). Among three active-site acidic residues, D311 is buried, whereas the D313 and E315 are solvent accessible (red). Residues Y391 and D392 (pink) have also been suggested to be involved in catalysis.





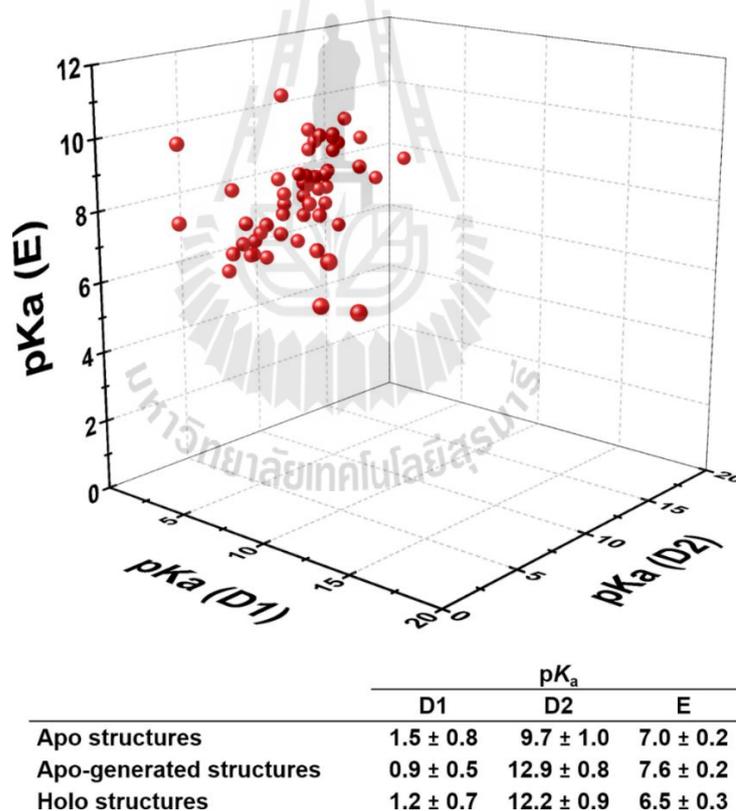
**Figure 3.33** Overview of family 18 chitinases. (A) Multiple sequence alignment of glycoside hydrolases in family 18. The amino acid sequences of fourteen chitinases used in this study were taken from the uniprot database. The secondary structure of *Vibrio harveyi* chitinase A, VhChiA (SwissProt: Q9AMP1) was aligned with the amino acid sequences of *P. furiosus* chitinase A, PfChiA (SwissProt: Q8U1H5); *Arthrobacter* sp. chitinase B, ArChiB (SwissProt: Q9REI6); *B. circulans* WL-12 chitinase A1, BcChiA1 (SwissProt: P20533); *S. marcescens* chitinase B, SmChiB (SwissProt: Q54276); *S. marcescens* chitinase A, SmChiA (SwissProt: P07254); *S. coelicolor* chitinase A, ScChiA (SwissProt: Q9Z9M9); *V. harveyi* chitinase A, VhChiA (SwissProt: Q9AMP1); *A. fumigatus* chitinase B1, AfChiB1 (SwissProt: Q873X9); *C. immitis* chitinase, CiChi1 (SwissProt: P0CB51); *H. brasiliensis* hevamine A, HbChi (SwissProt: P23472); *H. sapiens* chitotriosidase, HsChit1 (SwissProt: Q13231); *H. sapiens* acidic mammalian chitinase, HsAMCase (SwissProt: Q9BZP6);

*P. platycephala* endochitinase, *PpChi* (SwissProt: A6ZJJ1) and *S. cerevisiae* chitinase 1, *ScChi1* (SwissProt: P29029). The completely conserved amino acid residues that are crucial for chitinase activity are indicated with blue stars. (B) Ribbon representation of family 18 chitinase A from *Vibrio harveyi* showing the positions of three active-site acidic residues Asp (D<sub>1</sub>), Asp (D<sub>2</sub>) and Glu (E) (in white ball-and-stick model) in the DxDxE motif of the strand  $\beta$ 4 of the TIM-barrel structure. Only the  $\beta$ -strands (in purple) of the TIM-barrel structure are shown for clarity and here are defined as  $\beta$ 1 (residues 160-166);  $\beta$ 2 (residues 187-193);  $\beta$ 4 (residues 309-313);  $\beta$ 6 (residues 385-389);  $\beta$ 7 (residues 454-459) and  $\beta$ 8 (residues 566-570). The lemon stick of GlcNAc<sub>6</sub> is shown along with the catalytic TIM-barrel domain labeled as non-reducing end (NRE) and reducing end (RE). A red arrow indicates the cleavage site. (C) Surface representation of the catalytic cleft of chitinase A from *Vibrio harveyi*. The long track of surface-exposed hydrophobic residues responsible for binding with GlcNAc<sub>6</sub> is shown as blue surface. The three active-site acidic residues D311, D313 and E315 are shown as red surface whereas the Y391 and D392 that are expected to be involved in catalysis are shown as pink surface.

### 3.15 pK<sub>a</sub> calculations of the active-site residues

All family 18 chitinases contain the DxDxE sequence motif that positions three carboxylates, D<sub>1</sub>, D<sub>2</sub> and E in close proximity in an almost linear arrangement. The pK<sub>a</sub> values of D<sub>1</sub>, D<sub>2</sub> and E were successfully calculated for 75 structures of family 18 chitinases. The final dataset contained calculations for 24 apo structures, 46 apo-generated structures and 5 structures that contained either an inhibitor or a substrate-like molecule. The calculated pK<sub>a</sub> values (in red sphere) of three acidic residues D<sub>1</sub>, D<sub>2</sub>

and E were  $pK_a(D_1) < 2.0$ ,  $pK_a(D_2)$  in between 8.0-13.0 and  $pK_a(E)$  in the range from 6.0 to 9.0 (Figure 3.34). The differences in calculated  $pK_a$  values of the apo and apo-generated structures are statistically insignificant according to the post hoc tests (Duncan) in ANOVA analysis and show the average of the calculated  $pK_a$  values of D<sub>1</sub>, D<sub>2</sub> and E as  $1.5 \pm 0.8$ ,  $9.7 \pm 1.0$ ,  $7.0 \pm 0.2$  for the apo structures and  $0.9 \pm 0.5$ ,  $12.9 \pm 0.8$ ,  $7.6 \pm 0.2$  for the apo-generated structures, respectively (Figure 3.34). For the calculation of chitinases with ligand, the calculated  $pK_a$  values of D<sub>1</sub>, D<sub>2</sub> and E are found to be  $0.8 \pm 0.6$ ,  $12.4 \pm 1.1$ ,  $6.4 \pm 0.3$ , respectively (Figure 3.34).



**Figure 3.34** Contribution of  $pK_a$  values of acidic active-site residues, D<sub>1</sub>, D<sub>2</sub> and E plotted in three dimensions. Except for a few outliers, the  $pK_a$  values of the three active site carboxyls are found to be  $pK_a(D_1) < 2.0$ ,  $pK_a(D_2)$  in between 8.0-13.0 and the  $pK_a$  of E is found to lie in the range from 6.0 to 9.0.

The conformations of the three acidic residues, D<sub>1</sub>, D<sub>2</sub> and E of family 18 chitinases were visualised and details of their pK<sub>a</sub> predictions are given in Table 3.10. As seen in Table 3.10, the conformation of D<sub>1</sub> is completely in the 'up' position, whereas the conformation of E is almost in the 'down' position in both free and bound enzymes. Interestingly, the conformation of D<sub>2</sub> is in the 'up' or 'down' position depending on the presence of ligand. Namely, in the ligand-free enzyme (without mutation), a 59% majority of D<sub>2</sub> is in the 'down' position, interacting with D<sub>1</sub>, which is completely in the 'up' position. In 82% of cases, the calculated pK<sub>a</sub> values of D<sub>1</sub> are less than 0, whereas the calculated pK<sub>a</sub> values of D<sub>2</sub> and E are found to be 10.2 and 6.9, respectively. However, in the enzyme-ligand complex (without mutation), in 68% of cases D<sub>2</sub> moves into the 'up' position and interacts with either a substrate or an inhibitor and the glutamate, E. The D<sub>1</sub> is completely in the 'up' position for all structures in which the calculated pK<sub>a</sub> values of D<sub>1</sub> is found to be less than 0. In six structures, the D<sub>1</sub> could not be determined due to a strong interaction with D<sub>2</sub> (fluctuation titration curve overlapping between D<sub>1</sub> and D<sub>2</sub>). Whatever the position of D<sub>2</sub>, the pK<sub>a</sub> values of D<sub>1</sub> are still less than 0 while the average pK<sub>a</sub> values of D<sub>2</sub> and E are found to be 12.2 and 7.3, respectively.

In addition, mutations of D<sub>1</sub> or D<sub>2</sub> in the free forms of chitinases increase the acidity of the adjacent aspartate residue (Table 3.10), i.e., the mutation D140N (equivalent to D<sub>1</sub>) in *SmChiB* showed a downward shift in the pK<sub>a</sub> value of D142 (equivalent to D<sub>2</sub>) by 3.9 pH units while the mutation D169N (equivalent to D<sub>2</sub>) in *CiChi1* showed a downward shift in the pK<sub>a</sub> value of D167 (equivalent to D<sub>1</sub>) by 5.6 pH units.

The  $pK_a$  values of D<sub>1</sub>, D<sub>2</sub> and E calculated from the structures complexed with the natural substrates or inhibitors (holo structure) were further compared to its ligand-free structures (apo and apo-generated structures) as shown in Table 3.11. The conformations of D<sub>1</sub>, D<sub>2</sub> and E in five chitinases are found to be completely ‘up’, ‘up’ and ‘down’ upon ligand binding. In the absence of ligand, most calculated  $pK_a$  values of D<sub>1</sub> are less than 0, however, the calculated  $pK_a$  values of D<sub>1</sub> are increased in the presence of the natural substrates as shown in hevamine A in complex with GlcNAc<sub>3</sub> (PDB code: 1HVQ) and chitotriosidase in complex with GlcNAc<sub>2</sub> (PDB code: 1LG1). The calculated  $pK_a$  values of D<sub>2</sub> for all five structures are still higher than D<sub>1</sub> and showed average values of  $13.1 \pm 0.8$ ,  $13.7 \pm 1.2$  and  $12.4 \pm 1.1$  for the apo, apo-generated and holo structures, respectively. Interestingly, there are statistically significant differences on post hoc tests (Duncan) in ANOVA analysis in the calculated  $pK_a$  values of E between the apo-generated and holo structures of chitinases. As shown in Table 3.11, bacterial chitinases (*SmChiA* and *SmChiB*) showed a decrease in the  $pK_a$  values of E when they are bound with the ligand and the D<sub>2</sub> changes its original rotation from D<sub>1</sub> to E (see 1E15 vs 1E6R and 1CTN vs 1FFQ in Table 3.15.2). All calculated  $pK_a$  values of E in the holo structures (bound enzyme) showed a decrease by 0.7-2.8 pH units as compared with the  $pK_a$  values in the apo-generated structure (free enzyme) (Table 3.11).

**Table 3.10** Conformation and the  $pK_a$  values calculated from the apo and apo-generated structures of three acidic residues in the DxDxE motif of GH18 chitinases.

Protein	Source	PDBID	Res. (Å)	Mutation(s)	Ligand	Conformation			$pK_a$			Residue number
						D1	D2	E	D1	D2	E	
Chitinase A	<i>P. furiosus</i>	2DSK	1.50	-	-	up	up	down	< 0.0 <sup>a</sup>	9.9	6.0	D522, D524, E526
Chitinase B	<i>Arthrobacter</i> sp.	1KFW	1.74	-	-	up	down	down	< 0.0	15.6	8.2	D188, D190, E192
Chitinase A1	<i>B. circulans</i>	1ITX	1.10	-	-	up	down	down	n.d. <sup>b</sup>	n.d.	8.7	D200, D202, E204
Chitinase B	<i>S. marcescens</i>	1E15	1.90	-	-	up	down	down	< 0.0	15.1	7.1	D140, D142, E144
		1E6N	2.25	E144Q	GlcNAc <sub>5</sub>	up	up	-	< 0.0	6.9	-	
		1E6P	1.70	E144Q	-	up	down	-	n.d.	n.d.	-	
		1E6R	2.50	-	Allosamidin	up	up	down	< 0.0	13.4	7.4	
		1E6Z	1.99	-	Oxazolinium ion intermediate	up	up/down	down	14.0	< 0.0	8.1	
		1GOI	1.45	D140N	-	-	up	down	-	11.2	8.3	
		1H0G	2.00	-	Argadin	up	down	down	< 0.0	12.0	6.9	
		1OGG	1.97	D142N	Allosamidin	up	-	down	< 0.0	-	7.2	
		1UR9	1.80	D142N	HM508	up	-	down	< 0.0	-	7.1	
		Chitinase A	<i>S. marcescens</i>	1CTN	2.30	-	-	up	down	down	< 0.0	14.3
1EDQ	1.55			-	-	up	up/down	down	n.d.	n.d.	7.4	
1EHN	1.90			E315Q	GlcNAc <sub>5</sub>	up	down	-	n.d.	n.d.	-	
1EIB	1.80			D313A	GlcNAc <sub>5</sub>	up	-	down	< 0.0	-	6.1	
1FFQ	1.90			-	Allosamidin	up	up	down	< 0.0	15.6	8.6	
1FFR	1.80			Y390F	GlcNAc <sub>6</sub>	up	up	down	< 0.0	15.6	8.7	
1K9T	1.80			D391A	GlcNAc <sub>4</sub>	up	up	down	< 0.0	15.2	7.6	
1NH6	2.05			E315L	GlcNAc <sub>6</sub>	up	down	-	< 0.0	14.3	-	
1RD6	2.60			W167A	-	up	up	down	< 0.0	4.9	9.4	
1X6L	1.90			W167A	-	up	up	down	< 0.0	8.4	7.7	
1X6N	2.00			W167A	Allosamidin	up	up	down	< 0.0	17.7	8.4	
Chitinase A	<i>S. coelicolor</i>			3EBV	1.50	-	-	up	up	down	< 0.0	4.9
Chitinase A	<i>V. harveyi</i>	3B8S	2.00	-	-	up	down	down	< 0.0	16.0	8.4	D311, D313, E315
		3B9A	1.80	E315M	GlcNAc <sub>6</sub>	up	down	-	< 0.0	16.7	-	
		3B9D	1.72	E315M	GlcNAc <sub>5</sub>	up	down	-	n.d.	n.d.	-	
		3B9E	1.70	E315M	-	up	down	-	n.d.	n.d.	-	

**Table 3.10** Conformation and the  $pK_a$  values calculated from the apo and apo-generated structures of three acidic residues in the Dx Dx E motif of GH18 chitinases (Continued).

Protein	Source	PDBID	Res. (Å)	Mutation(s)	Ligand	Conformation			$pK_a$			Residue number			
						D1	D2	E	D1	D2	E				
Chitinase B1	<i>A. fumigatus</i>	1W9P	1.70	-	-	up	up	down	< 0.0	9.1	6.0	D173, D175, E177			
		1W9U	1.85	-	Argadin	up	down	down	15.6	< 0.0	7.0				
		1W9V	2.00	-	Argifin	up	up	down	< 0.0	13.6	7.6				
		2A3A	2.10	-	Theophylline	up	down	down	< 0.0	15.0	6.6				
		2A3B	1.90	-	Caffeine	up	down	down	n.d.	n.d.	7.9				
		2A3C	2.07	-	Pentoxifylline	up	down	down	n.d.	n.d.	8.4				
		2A3E	1.95	-	Allosamidin	up	up	down	< 0.0	11.9	10.2				
		2IUZ	1.95	-	C2-Dicaffeine	up	down	down	n.d.	n.d.	7.4				
		3CH9	2.20	-	Dimethylguanyurea	up	up/down	down	< 0.0	13.1	7.7				
		3CHC	1.90	-	Monopeptide	up	up	down	< 0.0	13.7	7.3				
		3CHD	2.00	-	Dipeptide	up	up	down	< 0.0	14.5	7.1				
		3CHE	2.05	-	Tripeptide	up	up	down	< 0.0	14.2	7.5				
		3CHF	1.95	-	Tetrapeptide	up	up	down	< 0.0	15.0	7.5				
		Chitinase 1	<i>C. immitis</i>	1WNO	2.10	-	-	up	down	down	12.7		0.9	6.7	D135, D137, E139 D167, D169, E171
				1D2K	2.20	-	-	up	down	down	6.1		7.8	6.6	
1LL4	2.80			-	Allosamidin	up	up	down	0.5	7.9	5.9				
1LL6	2.80			D169N	-	up	-	down	0.5	-	6.5				
Hevamine A	<i>H. brasiliensis</i>	1LL7	2.00	E171Q	-	up	down	-	8.3	< 0.0	-				
		1HVQ	2.20	-	GlcNAc <sub>3</sub>	up	up	down	< 0.0	9.3	6.6	D123, D125, E127			
		1KQY	1.92	D125A/E127A/ Y183F	GlcNAc <sub>5</sub>	up	-	-	< 0.0	-	-				
		1KQZ	1.92	D125A/E127A/ Y183F	GlcNAc <sub>4</sub>	up	-	-	< 0.0	-	-				
		1KR0	1.92	D125A/Y183F	GlcNAc <sub>4</sub>	up	-	down	< 0.0	-	5.2				
		1KR1	2.00	D125A/E127A	GlcNAc <sub>4</sub>	up	-	-	< 0.0	-	-				
		1LLO	1.85	-	Allosamidin	up	up	down	< 0.0	12.0	7.2				
		2HVM	1.80	-	-	up	up	down	< 0.0	11.9	6.6				

**Table 3.10** Conformation and the  $pK_a$  values calculated from the apo and apo-generated structures of three acidic residues in the DxDxE motif of GH18 chitinases (Continued).

Protein	Source	PDBID	Res. (Å)	Mutation(s)	Ligand	Conformation			$pK_a$			Residue number
						D1	D2	E	D1	D2	E	
Chitotriosidase	<i>H. sapiens</i>	1GUV	2.35	-	-	up	down	down	< 0.0	9.6	5.6	D136, D138, E140
		1HKI	2.55	-	Glucosaminidase B	up	up	down	2.6	14.7	8.0	
		1HKJ	2.60	-	Methylallosamidin	up	up	down	1.4	13.1	9.0	
		1HKK	1.85	-	Allosamidin	up	up	down	< 0.0	13.8	8.4	
		1HKM	2.55	-	Demethylallosamidin	up	up	down	< 0.0	13.8	9.0	
		1LG1	2.78	-	GlcNAc <sub>2</sub>	up	up	down	0.5	13.7	8.7	
		1LG2	2.10	-	Ethylene glycol	up	down	down	1.0	> 20.0 <sup>c</sup>	7.6	
		1LQ0	2.20	-	-	up	down	down	0.9	10.8	6.2	
		1WAW	1.75	-	Argadin	up	down	down	< 0.0	18.9	7.0	
		1WB0	1.65	-	Argifin	up	up	down	< 0.0	16.5	9.1	
AMCase	<i>H. sapiens</i>	3FXV	2.00	-	-	up	down	down	< 0.0	13.4	6.4	D136, D138, E140
		3FY1	1.70	-	Methylallosamidin	up	up	down	< 0.0	16.5	9.1	
Endochitinase	<i>P. platycephala</i>	2GSJ	1.73	-	-	up	up	down	< 0.0	9.8	5.6	D123, D125, E127
Chitinase 1	<i>S. cerevisiae</i>	2UY2	1.60	-	-	up	up	down	< 0.0	10.7	5.4	
		2UY3	1.90	-	8-Chlorotheophylline	up	up	down	< 0.0	11.6	7.7	D153, D155, E157
		2UY4	1.75	-	Acetazolamide	up	up	down	1.4	9.3	6.7	
		2UY5	1.60	-	Kinetin	up	up	down	< 0.0	8.1	5.3	

<sup>a</sup> Residue is always deprotonated.

<sup>b</sup> n.d., not determined due to D1 and D2 interacting strongly (interaction energy ~20.0 kT/e)

<sup>c</sup> Residue is always protonated.

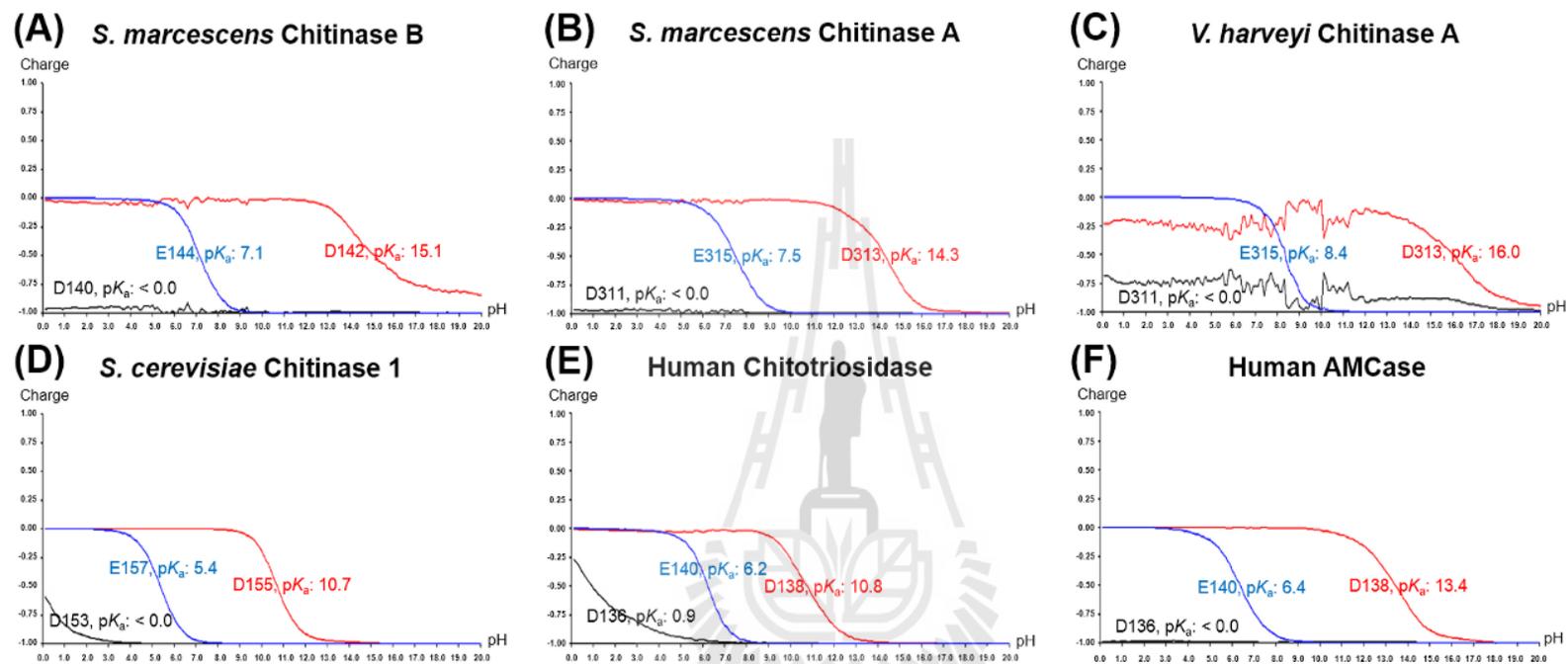
**Table 3.11** Conformation and the  $pK_a$  values calculated from the holo structures of three acidic residues in the DxDxE motif of family 18 chitinases and comparison with its apo and apo-generated structures.

Protein	Source	PDBID	Res. (Å)	Type	Ligand	Conformation			$pK_a$			Remark
						D1	D2	E	D1	D2	E	
Chitinase B	<i>S. marcescens</i>	1E15	1.90	Apo	-	up	down	down	<0.0	15.1	7.1	D140, D142, E144
		1E6R	2.50	Apo-generated	-	up	up	down	<0.0	13.4	7.4	
		1E6R	2.50	Holo	Allosamidin	up	up	down	<0.0	10.3	6.2	
Chitinase A	<i>S. marcescens</i>	1CTN	2.30	Apo	-	up	down	down	<0.0	14.3	7.5	D311, D313, E315
		1FFQ	1.90	Apo-generated	-	up	up	down	<0.0	15.6	8.6	
		1FFQ	1.90	Holo	Allosamidin	up	up	down	<0.0	16.0	5.8	
Hevamine A	<i>H. brasiliensis</i>	2HVM	1.80	Apo	-	up	up	down	<0.0	11.9	6.6	D123, D125, E127
		1HVQ	2.20	Apo-generated	-	up	up	down	<0.0	9.3	6.6	
		1HVQ	2.20	Holo	GlcNAc <sub>3</sub>	up	up	down	0.7	12.2	5.8	
Chitotriosidase	<i>H. sapiens</i>	1LQ0	2.20	Apo	-	up	down	down	0.9	10.8	6.2	D136, D138, E140
		1LG1	2.78	Apo-generated	-	up	up	down	0.5	13.7	8.7	
		1LG1	2.78	Holo	GlcNAc <sub>2</sub>	up	up	down	3.3	10.3	7.4	
AMCase	<i>H. sapiens</i>	3FXY	2.00	Apo	-	up	down	down	<0.0	13.4	6.4	D136, D138, E140
		3FY1	1.70	Apo-generated	-	up	up	down	<0.0	16.5	9.1	
		3FY1	1.70	Holo	Methylallosamidin	up	up	down	<0.0	13.4	7.0	

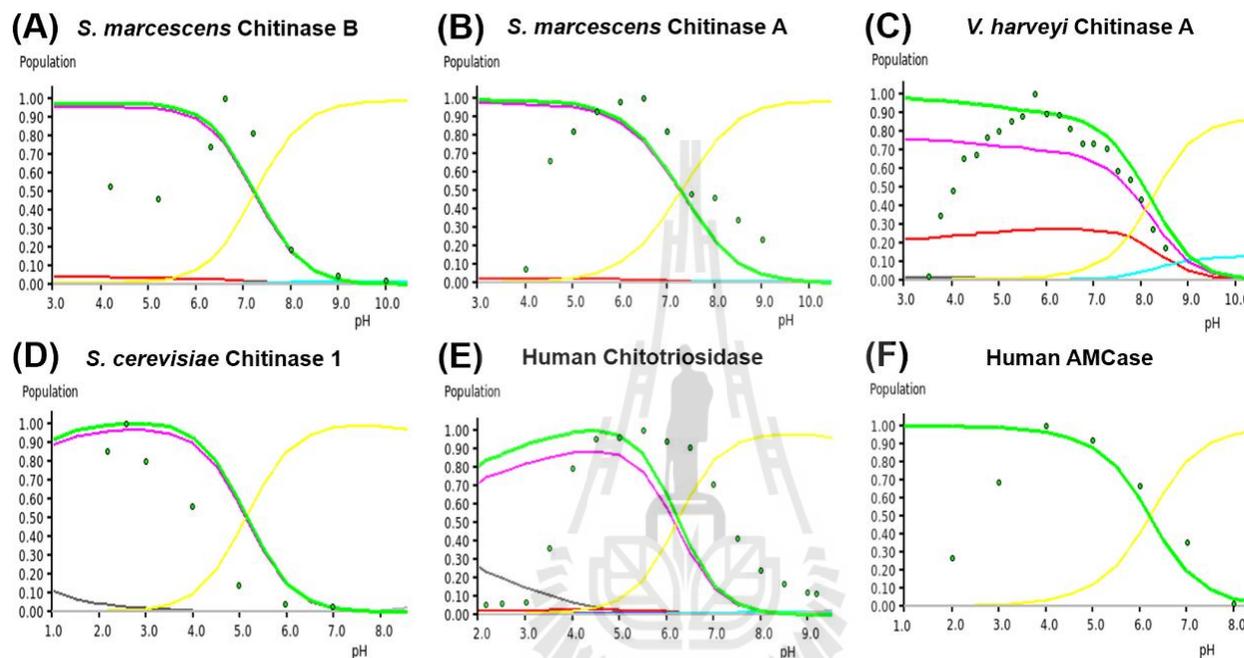
### 3.16 Predicting the pH-activity profiles

Due to a close proximity of the three carboxylates in the DxDxE motif sequence that is completely conserved in family 18 chitinases, the pH-dependence of catalytic activity is hypothesized to be dependent on the protonation state called the catalytically competent protonation state (CCPS) of all three residues, D<sub>1</sub>, D<sub>2</sub> and E. The calculated titration curves used in determining the pK<sub>a</sub> values of D<sub>1</sub>, D<sub>2</sub> and E in ligand-free structures for each chitinase were of good quality with the exception of *VhChiA* showing fluctuations in the curves for D<sub>1</sub> and D<sub>2</sub> (Figure 3.35).

For the chitinases used here we find that the D<sub>1</sub>, D<sub>2</sub> and E cluster behaves autonomously and only small perturbations in the titration curves and protonation state populations are observed when more active site groups are treated explicitly in Equation 2.4. Specifically, it is hypothesized that D<sub>1</sub> is required to be negatively charged whereas D<sub>2</sub> and E are required to be neutral for chitinases to be active. The computational pH profiles of family 18 chitinases were then predicted simply by extracting the pH dependence of the population of the identified CCPS and comparing these to experimentally measured pH-activity profiles of chitinases as follows: chitinase B from *S. marcescens* (*SmChiB*), chitinase A from *S. marcescens* (*SmChiA*), chitinase A from *V. harveyi* (*VhChiA*), chitinase 1 from *S. cerevisiae* (*ScChi1*), human chitotriosidase and human AMCase, respectively (Figure 3.36A-F). As shown in Figure 3.36A-F, the computational pH-activity profiles (green curves) are compared with the experimentally measured pH-activity profiles (green circles) that are scaled to unity. The green curve is the population in the catalytically competent protonation state normalized when the catalytic D<sub>1</sub>-D<sub>2</sub>-E triad was modelled with Asp (D<sub>1</sub>) charged, Asp (D<sub>2</sub>) neutral, and Glu (E) neutral.



**Figure 3.35** The quality of the titration curves from apo structures (A-F) used in determining the  $pK_a$  values.



**Figure 3.36** The computational pH-activity profiles of family 18 chitinases using the ligand-free structures. As seen in Figure 3.36A-F, the computational pH-activity profiles (green curve) of six chitinases are compared with the experimentally measured pH-activity profiles (green circles) that are scaled to the same units. The green curve is the pink curve normalized when the catalytic D<sub>1</sub>-D<sub>2</sub>-E triad was modelled with D<sub>1</sub> charged, D<sub>2</sub> neutral, and E neutral. The populations of the protonation states are also shown as the contributions of (0,0,0) (grey), (0,0,1) (blue), (0,1,0) (red), (0,1,1) (cyan), (1,0,0) (pink), (1,0,1) (yellow), (1,1,0) (orange), and (1,1,1) (light grey), where 0 and 1 defines as neutral and charged forms of D<sub>1</sub>, D<sub>2</sub> and E, respectively.

### 3.16.1 Computational pH-profiles of the apo structures

#### 3.16.1.1 Chitinase B from *Serratia marcescens* (*SmChiB*)

The experimentally measured  $k_{\text{cat}}$ -pH profile of *SmChiB* was determined using an assay with 4-methylumbelliferyl  $\beta$ -D-N, N'-diacetylchitobioside (4MU-GlcNAc<sub>2</sub>) and showed a bell-shaped profile with an optimum pH around 6-7 (van Aalten *et al.*, 2001). The calculated  $\text{p}K_{\text{a}}$  values of D140, D142 and E144 were found to be less than 0, 15.1 and 7.1, respectively (Figure 3.35A). The Asp140 is suggested to be deprotonated whereas the Asp142 and Glu144 are suggested to be protonated. *SmChiB* has a computational pH-activity profile with much more inverse sigmoid shape. For a hypothetical three-group system of D140, D142 and E144, the protonation state populations in the pH range 3-5 are pH-independent, showing a maximum population around 0.98. The protonation state populations are pH dependent in the pH range 6-10 showing a decrease in population to zero at pH 9 (Figure 3.36A). The basic limb of an experimentally measured pH-activity profile (green circles) of *SmChiB* is in a good agreement with a computational pH-activity profile (green line). The basic side of the curve with a  $\text{p}K_{\text{a}}$  of  $7.5 \pm 0.1$  and  $7.0 \pm 0.1$  from experimental and computational profiles can be rationalized by the ionization of Glu144 for which the predicted  $\text{p}K_{\text{a}}$  value was in this range. However, the  $\text{p}K_{\text{a}}$  that represents the ionization of the residue on the acidic side of the curve could not be determined accurately, since the shoulder seen between pH 3 and 5 indicates a residue that is always deprotonated.

### 3.16.1.2 Chitinase A from *Serratia marcescens* (*SmChiA*)

The experimentally measured  $k_{\text{cat}}$ -pH profile of *SmChiA* was determined using an assay with 4-methylumbelliferyl  $\beta$ -D-N, N', N''-triacetylchitotrioside (4MU-GlcNAc<sub>3</sub>) and showed a bell-shaped profile with an optimum pH at 6.5 (Zees *et al.*, 2009). The calculated  $\text{p}K_{\text{a}}$  values of D311, D313 and E315 were found to be less than 0, 14.3 and 7.5, respectively (Figure 3.35B). The Asp311 is suggested to be deprotonated whereas the Asp313 and Glu315 are suggested to be protonated. *SmChiB* has a computational pH-activity profile which has much more inverse sigmoid shape. For a hypothetical three-group system of D311, D313 and E315, the protonation state populations in the pH range 3-5 are pH-independent, showing a maximum population around 0.98 whereas the protonation state populations are pH dependent in the pH range 6-10 showing a decrease in population to zero at pH 9 (Figure 3.36B). The basic limb of an experimentally measured pH-activity profile (green circles) of *SmChiB* showed a shift upward as compared with the computational pH-activity profile (green line). This  $\text{p}K_{\text{a}}$  can be interpreted simply as a result of the ionization of the residue on the basic side of the curve with a  $\text{p}K_{\text{a}}$  of  $7.5 \pm 0.1$  and  $7.3 \pm 0.1$  from experimental and computational profiles that correspond to the calculated  $\text{p}K_{\text{a}}$  of Glu315. However, the  $\text{p}K_{\text{a}}$  which represents the ionization of the residue on the acidic side of the curve could not be determined accurately, since the shoulder seen between pH 3 and 5 indicates a residue that is always deprotonated.

### 3.16.1.3 Chitinase A from *Vibrio harveyi* (*VhChiA*)

The experimentally measured  $k_{\text{cat}}/K_{\text{m}}$ -pH profile of *VhChiA* was determined using an assay with 4-nitrophenyl *N, N'*-diacetyl- $\beta$ -D-chitobioside (*p*NP-GlcNAc<sub>2</sub>) and showed a bell-shaped profile with an optimum pH at around 5-6 (Suginta and Sritho, 2012). The calculated  $\text{p}K_{\text{a}}$  values of D311, D313 and E315 were found to be less than 0, 16.0 and 8.4, respectively (Figure 3.35C). The Asp311 is suggested to be deprotonated whereas the Asp313 and Glu315 are suggested to be protonated. *VhChiA* has a computational pH-activity profile with inverse sigmoid shape. For a hypothetical three-group system of D311, D313 and E315, the protonation state populations at pH range 3-10 are pH dependent with a decrease in population from a maximum population around 0.98 at pH 3 to zero at pH 10 (Figure 3.36C). The basic limb of an experimentally measured pH-activity profile (green circles) of *VhChiA* showed a good agreement with a computational pH-activity profile (green line). This  $\text{p}K_{\text{a}}$  can be interpreted simply as a result of the ionization of a residue on the basic side of the curve with a  $\text{p}K_{\text{a}}$  of  $7.8 \pm 0.1$  and  $8.0 \pm 0.1$  from experimental and computational profiles that correspond to the calculated  $\text{p}K_{\text{a}}$  of Glu315. However, the  $\text{p}K_{\text{a}}$  which represents the ionization of the residue on the acidic side of the curve could not be determined accurately, since the shoulder seen between pH 3 and 6 indicates a residue that is always deprotonated.

#### 3.16.1.4 Chitinase 1 from *Saccharomyces cerevisiae* (ScChi1)

The experimentally measured  $k_{\text{cat}}/K_m$ -pH profile of ScChi1 was determined using an assay with 4-methylumbelliferyl  $\beta$ -D-N, N', N''-triacetylchitotrioside (4MU-GlcNAc<sub>3</sub>) and showed a bell-shaped profile with an optimum pH at around 2.6 (Hurtado-Guerrero and van Aalten, 2007). The calculated  $pK_a$  values of D153, D155 and E157 were found to be less than 0, 10.7 and 5.4, respectively (Figure 3.35D). The Asp153 is suggested to be deprotonated whereas the Asp155 and Glu157 are suggested to be protonated. ScChi1 has a computational pH-activity profile with a much more inverse sigmoid shape. For a hypothetical three-group system of D153, D155 and E157, the protonation state populations at pH range 1-3 are pH dependent with an increase in population from 0.90 to a maximum population around 0.98 at pH 3 and then decreased to zero at pH 7 (Figure 3.36D). The basic limb of an experimentally measured pH-activity profile (green circles) of ScChi1 showed a downward shift as compared with the computational pH-activity profile (green line). This  $pK_a$  can be interpreted simply as a result of the ionization of a residue on the basic side of the curve with a  $pK_a$  of  $4.2 \pm 0.1$  and  $5.0 \pm 0.1$  from experimental and computational profiles that correspond to the calculated  $pK_a$  of Glu157. However, the  $pK_a$  which represents the ionization of the residue on the acidic side of the curve, could not be determined accurately, since the shoulder seen between pH 1 and 3 indicates a residue that is always deprotonated.

### 3.16.1.5 Chitotriosidase from human

The experimentally measured  $k_{\text{cat}}$ -pH profile of human chitotriosidase was determined using an assay with 4-methylumbelliferyl  $\beta$ -D-N, N', N''-triacetylchitotrioside (4MU-GlcNAc<sub>3</sub>) and showed a bell-shaped profile with an optimum pH that was relatively broad and peaks around pH 6 (Boot *et al.*, 2005). The calculated  $\text{p}K_{\text{a}}$  values of D136, D138 and E140 were found to be 0.9, 10.8 and 6.2, respectively (Figure 3.35E). The Asp136 is suggested to be deprotonated whereas the Asp138 and Glu140 are suggested to be protonated. Human chitotriosidase has a computational pH-activity profile which is approximately bell-shaped. For a hypothetical three-group system of D136, D138 and E140, the protonation state populations at pH range 2-5 are pH dependent with an increase in populations from 0.80 to a maximum population around 0.98 at pH 5 and then decreased to zero at pH 8 (Figure 3.36E). The basic limb of an experimentally measured pH-activity profile (green circles) of human chitotriosidase showed an upward shift as compared with the computational pH-activity profile (green line). This  $\text{p}K_{\text{a}}$  can be interpreted simply as a result of the ionization of a residue on the basic side of the curve with a  $\text{p}K_{\text{a}}$  of  $7.2 \pm 0.1$  and  $6.3 \pm 0.1$  from experimental and computational profiles that correspond to the calculated  $\text{p}K_{\text{a}}$  of Glu140. However, the  $\text{p}K_{\text{a}}$  that represents the ionization of the residue on the acidic side of the curve could not be determined accurately, since the shoulder seen between pH 2 and 5 indicates a residue that is always deprotonated.

### 3.16.1.6 AMCase from human

The experimentally measured  $k_{\text{cat}}/K_{\text{m}}$ -pH profile of human AMCase was determined using an assay with 4-methylumbelliferyl  $\beta$ -D-N, N'-diacetylchitobioside (4MU-GlcNAc<sub>2</sub>) and showed a bell-shaped profile with an optimum pH around 4-5 (Chou *et al.*, 2006). The calculated  $\text{p}K_{\text{a}}$  values of D136, D138 and E140 were found to be less than 0, 13.4 and 6.4, respectively (Figure 3.35F). The Asp136 is suggested to be deprotonated whereas the Asp138 and Glu140 are suggested to be protonated. Human AMCase has a computational pH-activity profile with a more inverse sigmoid shape. For a hypothetical three-group system of D136, D138 and E140, the protonation state populations at pH range 1-4 are pH independent, showing a maximum population around 1.0 whereas the protonation state populations are pH dependent at pH range 5-8 showing a decrease in population to zero at pH 8 (Figure 3.36F). The basic limb of an experimentally measured pH-activity profile (green circles) of human AMCase is in a good agreement with the computational pH-activity profile (green line). This  $\text{p}K_{\text{a}}$  can be interpreted simply as a result of the ionization of the residue on the basic side of the curve with a  $\text{p}K_{\text{a}}$  of  $6.3 \pm 0.1$  and  $6.2 \pm 0.1$  from experimental and computational profiles that correspond to the calculated  $\text{p}K_{\text{a}}$  of Glu140. However, the  $\text{p}K_{\text{a}}$  which represents the ionization of the residue on the acidic side of the curve, could not be determined accurately, since the shoulder seen between pH 1 and 4 indicates a residue that is always deprotonated.

### 3.16.2 Computational pH-profiles for holo structures

#### 3.16.2.1 Chitotriosidase from human

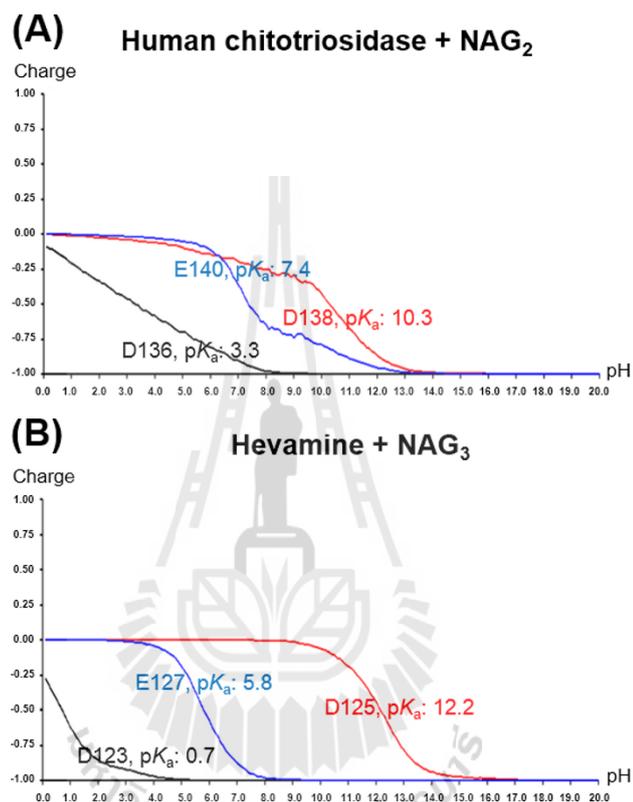
The experimentally measured  $k_{\text{cat}}$ -pH profile of human chitotriosidase was determined using an assay with 4-methylumbelliferyl  $\beta$ -D-N, N', N''-triacetylchitotrioside (4MU-GlcNAc<sub>3</sub>) and showed a bell-shaped profile with an optimum pH that was relatively broad and peaks around pH 6 (Boot *et al.*, 2005). The calculated  $pK_a$  values of D136, D138 and E140 were found to be 3.3, 10.3 and 7.4, respectively (Figure 3.37A). The Asp136 is suggested to be deprotonated whereas the Asp138 and Glu140 are suggested to be protonated. Human chitotriosidase in complex with chitobiose (GlcNAc<sub>2</sub>) has a computational pH-activity profile which is approximately bell-shaped with an optimum pH around 5-6. For a hypothetical three-group system of D136, D138 and E140, the protonation state populations are pH dependent. In the pH range 1-5, populations rapidly increased from 0.46 to a maximum population around 0.99 at pH 5.5 and then rapidly decreased to zero at pH 9 (Figure 3.38A). The acidic limb of an experimentally measured pH-activity profile (green circles) of human chitotriosidase showed an upward shift compared to the computational pH-activity profile (green line). This  $pK_a$  can be interpreted simply as a result of the ionization of a residue on the acidic side of the curve with a  $pK_a$  of  $3.6 \pm 0.1$  and  $1.8 \pm 0.1$  from experimental and computational profiles that are close to the calculated  $pK_a$  of Asp136. The basic limb of an experimentally measured pH-activity profile (green circles) of human chitotriosidase showed an upward shift using the computational pH-activity profile (green line) and this  $pK_a$  can be interpreted simply as a result of the ionization of a residue on the basic side of the curve with a  $pK_a$  of  $7.3 \pm$

0.1 and  $6.9 \pm 0.1$  from experimental and computational profiles that are close to the calculated  $pK_a$  of Glu140.

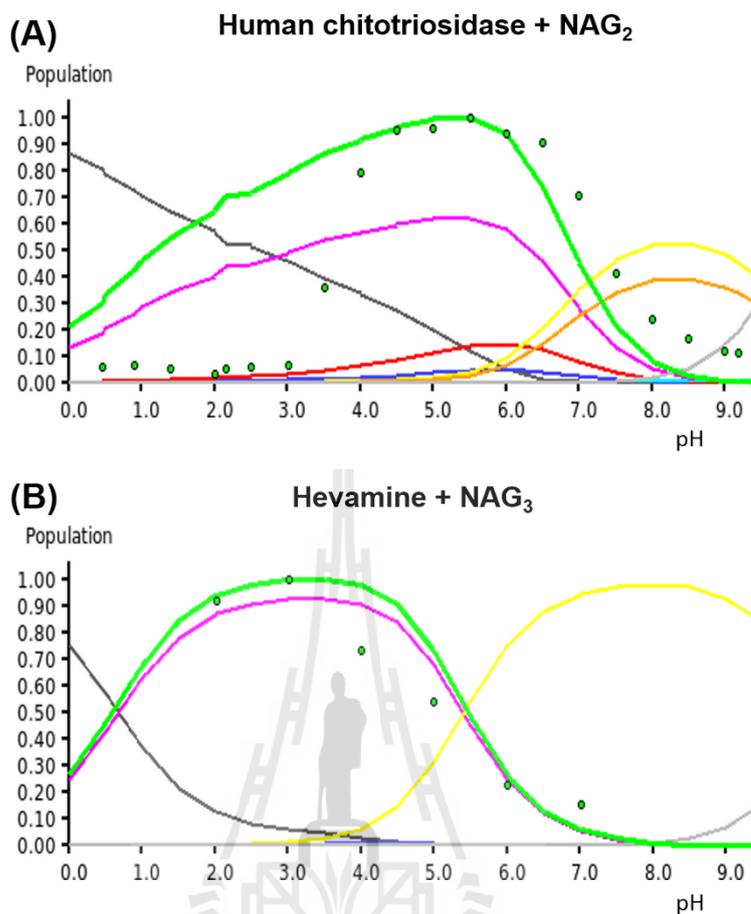
### 3.16.2.2 Hevamine A from *Hevea brasiliensis* (*HbChi*)

The experimentally measured  $k_{\text{cat}}$ -pH profile of *HbChi* was determined using an assay with chitopentaose (GlcNAc<sub>5</sub>) and showed a bell-shaped profile with an optimum pH at pH 2-3 with no activity remaining at pH 8 and above (Bokma *et al.*, 2002). The calculated  $pK_a$  values of D123, D125 and E127 were found to be 0.7, 12.2 and 5.8, respectively (Figure 3.37B). The Asp123 is suggested to be deprotonated whereas the Asp125 and Glu127 are suggested to be protonated. Hevamine A in complex with chitotriose (GlcNAc<sub>3</sub>) has a computational pH-activity profile which is approximately bell-shaped with an optimum pH around 2-4. For a hypothetical three-group system of D123, D125 and E127, the protonation state populations are pH dependent. At pH range 1-3, populations rapidly increased from 0.26 to a maximum population around 0.99 at pH 3 and then decreased to zero at pH 8 (Figure 3.38B). The acidic limb of an experimentally measured pH-activity profile (green circles) of *HbChi* showed a good agreement with the computed pH-activity profile (green line). This  $pK_a$  can be interpreted simply as a result of the ionization of a residue on the acidic side of the curve with a  $pK_a$  of  $1.0 \pm 0.1$  from only the computational profile, that is close to the calculated  $pK_a$  of Asp123 whereas the  $pK_a$  from an experimentally measured  $k_{\text{cat}}$ -pH profile cannot be determined since the enzyme activity at pH below 2 was not measured. The basic limb of an experimentally measured pH-activity profile (green circles) of *HbChi* showed a good agreement with the computational pH-activity profile (green line) and this  $pK_a$  can be interpreted simply

as a result of the ionization of a residue on the basic side of the curve with a  $pK_a$  of  $5.2 \pm 0.1$  and  $5.4 \pm 0.1$  from experimental and computational profiles that are close to the calculated  $pK_a$  of Glu127.



**Figure 3.37** The quality of the titration curves from the holo structures (A-B) used in determining the  $pK_a$  values.



**Figure 3.38** The computational pH-activity profiles of family 18 chitinases using the complex structures with the natural substrates. The computational pH-activity profiles of human chitotriosidase complexed with chitobiose (A) and hevamine A complexed with chitotriose (B) are shown in a pH range from 0-9. The computational pH-activity profiles (green curve) are compared with the experimentally measured pH-activity profiles (green circles) that are scaled to the same unit. The green curve is the pink curve (the population of the CCPS) normalized when the catalytic D<sub>1</sub>-D<sub>2</sub>-E triad was modelled with D<sub>1</sub> charged, D<sub>2</sub> neutral and E neutral. The populations of the various protonation states are also shown as mentioned above.

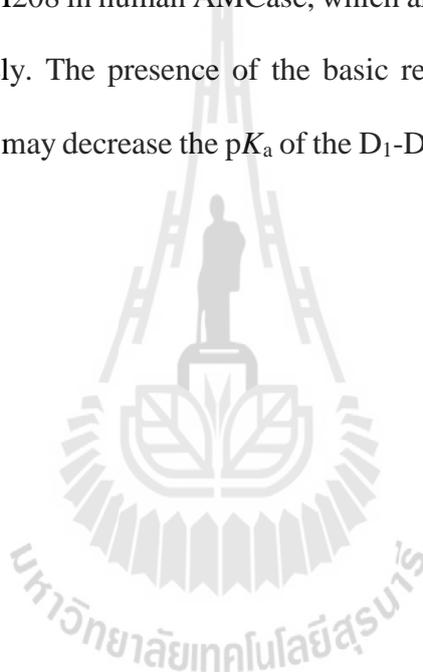
### 3.17 $pK_a$ comparison between bacterial and human chitinases

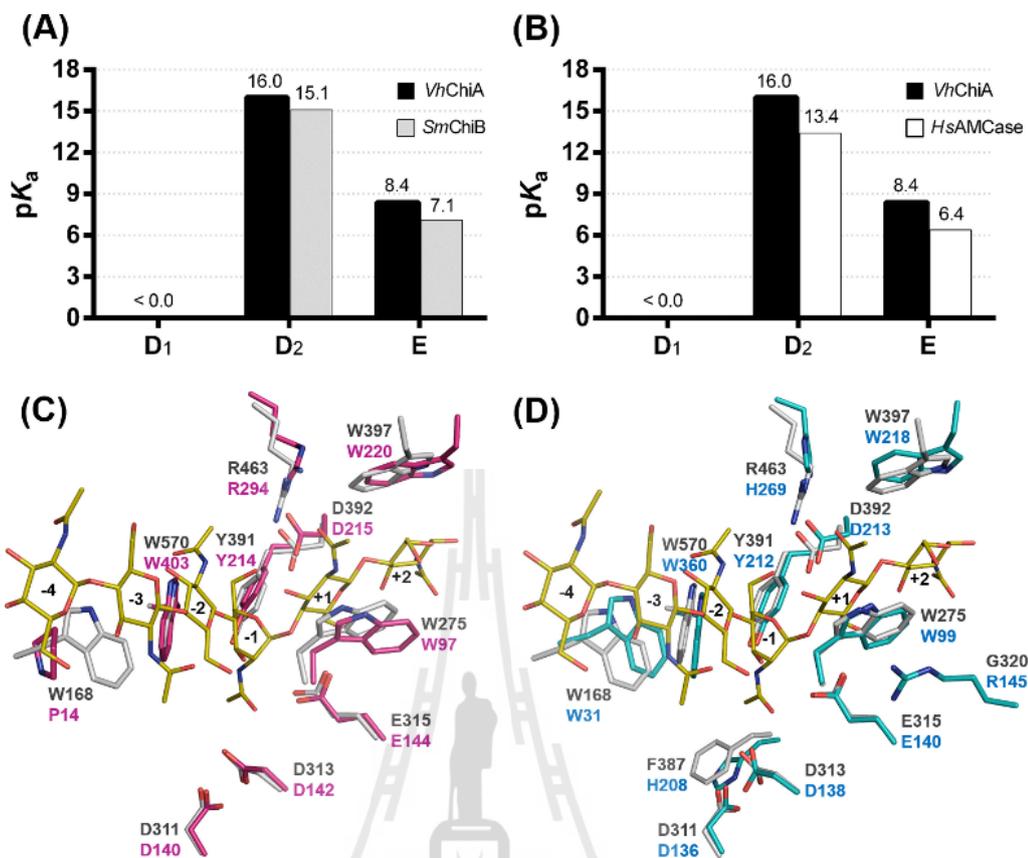
$pK_a$  values of the active-site residues in an enzyme are important for the catalytic mechanisms, enzyme activity, pH-dependent conformational changes and protein stability. Therefore comparison of  $pK_a$  values of chitinases used here is important having in mind that they differ in their structures and functions. The pH-activity profile of family 18 chitinases has a broad pH-range from 2-9 and is determined by the  $pK_a$  values of its acidic active-site residues. In bacterial chitinases, *VhChiA* showed a  $k_{cat}$ -pH profile with a broad optimum pH around 5-9 whereas *SmChiB* showed a  $k_{cat}$ -pH profile with an optimum pH around 6-7. The difference in the optimum pH profiles may stem from the different  $pK_a$  values of active site residues, D<sub>1</sub>, D<sub>2</sub> and E in both bacterial enzymes. Namely, the calculated  $pK_a$  values of D311, D313 and E315 of *VhChiA* are found to be less than 0, 16.0 and 8.4, respectively, whereas the calculated  $pK_a$  values of D140, D142 and E144 of *SmChiB* are found to be less than 0, 15.1 and 7.1, respectively (Figure 3.39A). The  $pK_a$  of D<sub>1</sub> (D311 in *VhChiA* and D140 in *SmChiB*) cannot be distinguished between both bacterial chitinases since its  $pK_a$  is found to be less than 0. The  $pK_a$  of D<sub>2</sub> (D313 in *VhChiA* and D142 in *SmChiB*) is different by only 0.9 pH units whereas the  $pK_a$  of E (E315 in *VhChiA* and E144 in *SmChiB*) is different by about 1.3 pH units. In comparison with human AMCase, *VhChiA* showed a  $k_{cat}/K_m$ -pH profile with an optimum pH around 5-6 whereas human AMCase showed a  $k_{cat}/K_m$ -pH profile with an optimum pH around 4-5. The  $pK_a$  of D<sub>1</sub> (D311 in *VhChiA* and D136 in *HsAMCase*) cannot be distinguished between bacterial and human chitinases since its  $pK_a$  is found to be less than 0. The  $pK_a$  of D<sub>2</sub> (D313 in *VhChiA* and D138 in *HsAMCase*) is different by about 2.6 pH units whereas the  $pK_a$  of E (E315 in *VhChiA* and E140 in *HsAMCase*) is different by about 2.0 pH units (Figure 3.39B).

Interestingly, the active site topologies between chitinases are not the same. The crystal structure of *VhChiA* was compared with the crystal structures of *SmChiB* and *HsAMCase*. A DALI pairwise comparison (Hasegawa and Holm, 2009) showed that the crystal structure of the catalytic domain of *VhChiA* (PDB code: 3B8S) has a fold similar to the crystal structure of the catalytic domain of *SmChiB* (PDB code: 1E15) with Z-score = 37.7, RMSD 2.3 Å over 352 residues and 30% sequence identity, whereas the crystal structure of the catalytic domain of *VhChiA* (PDB code: 3B8S) has a fold similar to the crystal structure of the catalytic domain of *HsChiB* (PDB code: 3FXY) with Z-score = 44.4, RMSD 1.8 Å over 348 residues and 28% sequence identity. A DALI alignment suggests that *VhChiA* has a fold more similar to *HsAMCase* than to *SmChiB*.

A close up of the sugar-binding pocket of *VhChiA*, *SmChiB* and *HsAMCase* is shown in (Figure 3.39C-D). Even though the three chitinases have a conserved sugar-binding pocket for longer chitooligosaccharides, some differences in substrate-binding features between *VhChiA* and other two homologues were observed. As compared with *SmChiB* (pink stick), the active-site residues of *VhChiA* (gray stick) with GlcNAc<sub>6</sub> (yellow stick) showed that the sugar rings of GlcNAc<sub>6</sub> from subsite -4 to +2 are in the chair conformation with the exception of the distortion of -1 GlcNAc into a boat conformation and the twist of the scissile bond at subsite -1 and +1 before the bond cleavage. The active site residues, D311, D313 and E315 of *VhChiA* are well aligned with the active site residues, D140, D142, E144 of *SmChiB* and there are some shifts in residues Y391, D392, W275, W397 and R463 of *VhChiA* as compared to the equivalent residues in *SmChiB* (Figure 3.39C). The most significant difference is in the position of W168 in *VhChiA* (equivalent to P14 in *SmChiB*). The presence of W168

instead of P14 and other aromatic residues in the substrate-binding cleft may increase the  $pK_a$  of the D<sub>1</sub>-D<sub>2</sub>-E system. In comparison with human AMCase, the active site residues D311, D313 and E315 of *VhChiA* are well aligned with the active site residues, D136, D138, E140 of human AMCase and there are some shifts in residues Y391, D392, W168, W275, W397, R463 and W570 of *VhChiA* as compared to the equivalent residues in human AMCase (Figure 3.39D). The most significant difference is the presence of R145 and H208 in human AMCase, which are equivalent to G320 and F387 in *VhChiA*, respectively. The presence of the basic residues R145 and H208 in the substrate-binding cleft may decrease the  $pK_a$  of the D<sub>1</sub>-D<sub>2</sub>-E system, resulting a lowered pH optimum.





**Figure 3.39** A comparison of bacterial and human chitinases. The calculated active-site pKa values D<sub>1</sub>, D<sub>2</sub> and E of *VhChiA* were compared with the calculated active-site pKa values D<sub>1</sub>, D<sub>2</sub> and E of *SmChiB* (A) and *HsAMCase* (B). Superimposition of the active site residues of *VhChiA* (PDB code: 3B8S) in gray stick with *SmChiB* (PDB code: 1E15) in dark pink stick (C) and *HsAMCase* (PDB code: 3FXV) in blue stick (D) that participate in sugar binding is shown. The substrate, GlcNAc<sub>6</sub> was taken from the crystal structure of *VhChiA* in complex with GlcNAc<sub>6</sub> (PDB code: 3B9A) and displayed as yellow stick.

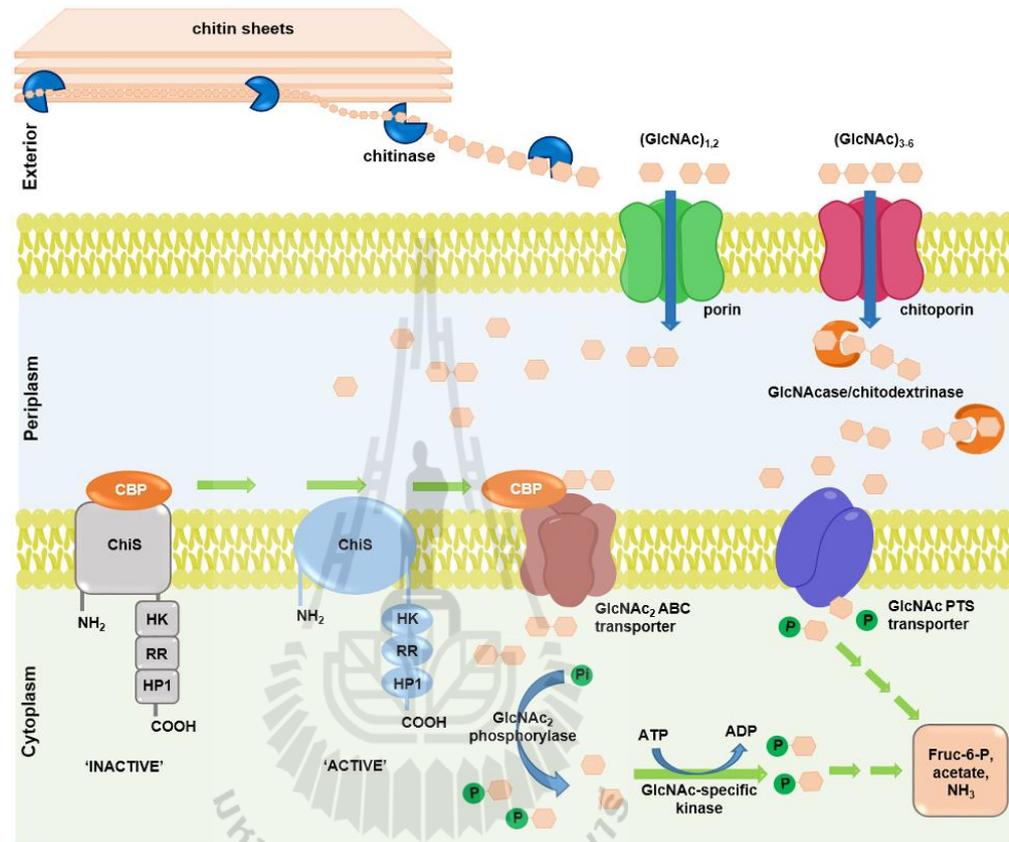
## CHAPTER IV

### DISCUSSION

#### 4.1 Chitin degradation pathway of *Vibrio harveyi*

Chitin turnover in the marine biosphere depends upon the activities of marine *Vibrios* (Jeuniaux and Voss-Foucart, 1991; Zobell and Rittenberg, 1938). The chitin catabolic cascade of the *Vibrios* has been demonstrated to incorporate a large number of genes and enzymes, which are orchestrated in a complex signal transduction pathway (Figure 4.1) (Bassler *et al.*, 1991; Jeuniaux and Voss-Foucart, 1991; Keyhani and Roseman, 1999; Li and Roseman, 2004; Park *et al.*, 2000; Yu *et al.*, 1991; Zobell and Rittenberg, 1938). We previously identified and characterized three biological components of the chitin catabolic pathway that are essential for chitin degradation and chitin uptake by *V. harveyi*. Chitinase A (so-called *VhChiA*) is an endolytic enzyme responsible for the breakdown of insoluble chitin chains into small, soluble chitooligosaccharides (Suginta *et al.*, 2004; Suginta *et al.*, 2005), while chitoporin (so-called *VhChiP*), a sugar-specific porin located in the outer membrane of the bacterium, is responsible for chitooligosaccharide uptake (Suginta, Chumjan, Mahendran, Janning *et al.*, 2013; Suginta, Chumjan, Mahendran, Schulte *et al.*, 2013). The last component is  $\beta$ -*N*-acetylglucosaminidase (known as *VhGlcNAcase*), an exolytic enzyme capable of degrading the transported chitooligosaccharides to GlcNAc monomers, which then act as signalling molecules that regulate the downstream cascade of the chitin catabolic

pathway, through the activation of the chitin sensor (ChiS) (Hunt *et al.*, 2008; Keyhani and Roseman, 1999).



**Figure 4.1** Model of the chitin degradation cascade of the marine bacterium *Vibrio harveyi* (modified from Li and Roseman, 2004; Suginta *et al.*, 2013).

## 4.2 Probing the catalytic residues of *VhGlcNAcases* by chemical rescue

*VhGlcNAcase*, a novel member of the GH20 GlcNAcase family, contains four GlcNAc binding subsites (-1), (+1), (+2) and (+3), and exhibits its greatest activity with chitotetraose (Suginta *et al.*, 2010). Amino acid sequence comparison with other GlcNAcases and our 3D-structure, modelled on the known 3D structure of *SmCHB* (Figure 3.1A-C), suggested that two invariant acidic side-chain pairs, Asp303-Asp304 and Asp437-Glu438, could be important for catalysis. Both acidic pairs lie in close proximity to the cleavage site (-1 subsite) and have equal opportunity to act as the catalytic couple. In this study, we performed site-directed mutagenesis, followed by chemical rescue assay to identify the catalytic couple. In the first set of experiments, we observed that point mutations of four invariant acid residues (Asp303, Asp304, Asp437 and Glu438) caused a drastic loss of the enzymic activity of *VhGlcNAcase* toward a synthetic substrate, *pNP*-GlcNAc. Notably, mutations of Asp437 to Ala (mutant D437A) and Glu438 to Ala (mutant E438A) abolished the activity almost completely, confirming that these acidic residues play important roles in chitin degradation.

In the next experiment, we observed that among various sodium salts, sodium azide greatly inhibited the activity of *VhGlcNAcase* WT, but sodium formate produced only weak inhibition (Figure 3.9). Such observations were consistent with our previous report that sodium azide acted as a potent competitive inhibitor of *VhGlcNAcase* (Sirimontree *et al.*, 2015). Both azide and formate ions, the forms of sodium azide and sodium formate, respectively, that exist in buffered solution, are strong nucleophiles (Comfort *et al.*, 2007; Zechel and Withers, 2001). Therefore, their ability to rescue enzymic activity of inactive mutants through nucleophilic effect has been employed to

elucidate the catalytic mechanism of several retaining glycoside hydrolases (Cobucci-Ponzano *et al.*, 2003; Paal *et al.*, 2004; Shallom *et al.*, 2002; Vallmitjana *et al.*, 2001; Viladot *et al.*, 1998). In our study, their inhibitory effects on the mutant forms of *VhGlcNAcase* were significantly less than on WT, suggesting that the inactivating effects of the active-site mutations were partially eliminated when azide or formate was included in the assay reaction. Formate ion seemed to act as the more potent nucleophile, as we observed its greater chemical rescue effect on the mutants D437A and E438A/Q, as compared to azide ion; the enzymic activity of the mutant D437A was even enhanced by formate, but not by azide. The less effective chemical rescue produced by azide ion may result from its linear geometry, which allows only a poor fit into the catalytic pocket of *VhGlcNAcase*. In contrast formate ion, which has trigonal planar geometry, may accurately mimic the carboxylate side chain of Asp437 (Figure 3.8). Therefore, the activity loss due to the interruption of the catalytic cycle, caused by loss of the natural nucleophile upon replacement of Asp437 with Ala, could be re-established in the presence of this small exogenous nucleophile.

Lineweaver Burk plots of  $1/v_0$  vs.  $1/[S]$  at different formate concentrations (Figure 3.12C) yielded linear lines that intersect above the x-axis, agreeing with the mixed-type mode of binding. The results suggested that formate ion could interact with both unliganded D437A ( $E^{\text{mut}}$ ) and ligand-bound D437A ( $E^{\text{mut}}S$ ), but the enhanced activity would occur only when formate ion bound to the enzyme-substrate complex. Figure 4.2 shows the proposed mechanism, in which formate ion increases the rate of *pNP*-glycoside hydrolysis by replacing the substituted side-chain of Asp437 in the catalytic pocket of the D437A-substrate complex. The restoration of the enzyme activity observed with the mutant D437A suggested a crucial role for Asp437 as the

catalytic nucleophile in both the glycosylation and deglycosylation steps of the substrate-assisted mechanism proposed for GH20 GlcNAcases (Williams *et al.*, 2002).

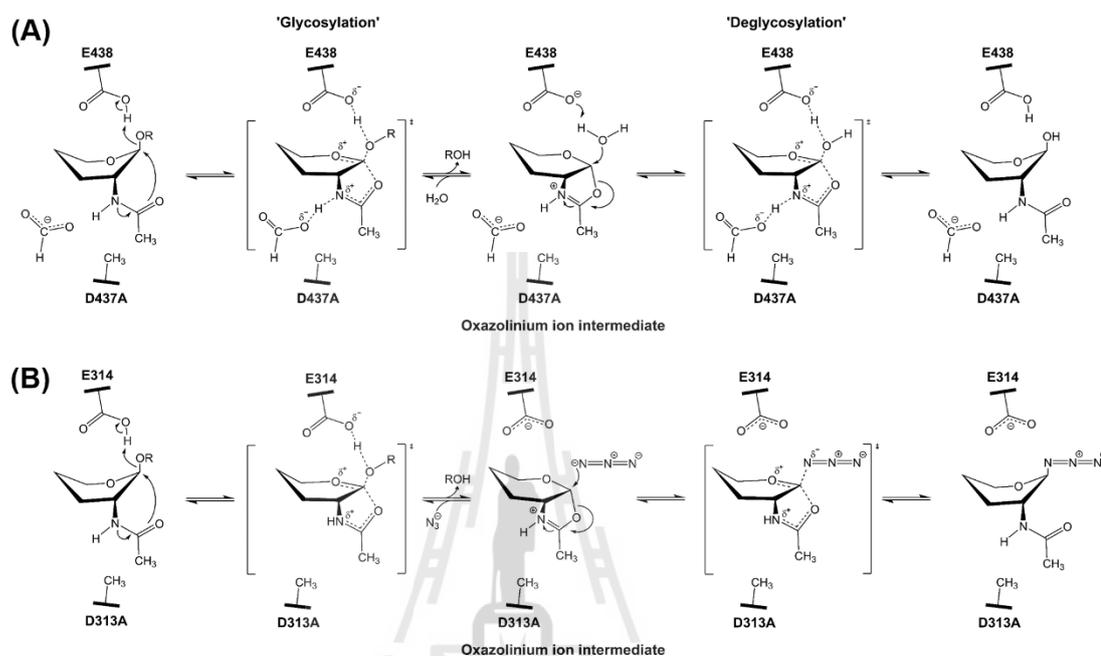
The absence of any shift in the optimal pH in the pH-activity profiles of the *VhGlcNAcase* WT and the D437A mutant suggested that D437 did not facilitate bond cleavage by lowering the  $pK_a$  value of its catalytic partner Glu438. This is a major difference in the catalytic role of Asp437 in *VhGlcNAcase* from Asp313 in *SpHex*. In the case of *SpHex*, the D313A mutant was shown to increase the optimal pH value from 5.0 to 7.5, the pH/activity data suggesting a significant contribution to bond cleavage by Asp313 (Williams *et al.*, 2002).

When compared with *SpHex* (Williams *et al.*, 2002), the effects of formate ion on the *VhGlcNAcase* inactive mutant showed completely different mechanistic details from the effects of azide ion on the *SpHex* inactive mutant. As clearly illustrated in Figure 4.2A, formate ion acts as an alternative nucleophile to Asp437. In our study, the planar geometry of formate ion optimally replaced the missing  $-\beta\text{COO}^-$  side chain of D437A. In our proposed mechanism, formate ion ( $\text{HCOO}^-$ ) accepts a proton from the  $-\text{NH}$  of the C2-acetamido group of the oxazolinium intermediate that is generated in course of the scissile-bond cleavage by the acid catalyst Glu438. Such a covalent bond formation would aid the stabilization of the oxazolinium intermediate in the glycosylation step. In addition, the nucleophilic attack by formate ion would help to orient the positively-charged C1 of the reaction intermediate, so as to react with the neighbouring water molecule in the subsequent deglycosylation step, leading to hydrolysis and retaining the configuration of the GlcNAc product. The diagram shown in Figure 4.2A clearly emphasizes the role of Asp437 as a catalytic nucleophile, through

the rescue function of formate ion, to stabilize the reaction intermediate at glycosylation and deglycosylation steps of the enzyme-substrate complex.

In marked contrast, azide ion rescued the GlcNAcase activity of *SpHex* inactive mutant (D313A) by acting as an alternative nucleophile to water (not to Asp313). As shown in Figure 4.2B, azide ion does not mimic the nucleophilic role of Asp313, but the result of the nucleophilic attack by azide ion was to open the oxazolinium ion intermediate mainly in the deglycosylation step. This proposed mechanism was supported by the kinetic analysis, which showed a much greater increase in the first-order rate constant ratio (apparent  $k_{cat}/k_{cat}$ ) (16-fold) than in the second-order rate constant ratio (apparent  $k_{cat}/K_m)/(k_{cat}/K_m)$  (5-fold). With *VhGlcNAcase*, formate-mediated chemical rescue produced similar increases in (apparent  $k_{cat}/k_{cat}$ ), and (apparent  $k_{cat}/K_m)/(k_{cat}/K_m)$  (2.5- and 1.9-fold respectively). Such analysis suggested that the exogenous nucleophile restored the activity of the D437A mutant by accelerating both rates of deglycosylation (as reflected by the apparent  $k_{cat}$ ) and glycosylation (as reflected by the apparent  $k_{cat}/K_m$ ). It is noteworthy that the rescue effect observed for our inactive *VhGlcNAcase* was not dramatic, and this may reflect the reactivity of the leaving group on the tested substrate. *pNP-GlcNAc* contains a poor leaving group, so is much less susceptible to enzymic hydrolysis than 2,4-DNPGlcNAc and 3,5-DNPGlcNAc, which contain strong leaving groups. Similar results were reported by Vallmitjana *et al.*, 2001. They observed only a 3 fold enhancement of  $k_{cat}$  when *pNP-GlcNAc* was used as the substrate for the  $\beta$ -glucosidase assay of the nucleophilic inactive mutant E178A, while 188 fold  $k_{cat}$  enhancement was observed when 2,4-DNPGlcNAc was the substrate. This would explain the modest 2.5 fold increase in  $k_{cat}$  for *VhGlcNAcase* inactive mutant with *pNP-GlcNAc* as substrate,

compared to *SpHex*, which showed a 16 fold increase in  $k_{cat}$  with 3,5-DNPGlcNAc substrate (Williams *et al.*, 2002).



**Figure 4.2** Proposed mechanism of nucleophile-mediated chemical rescue with the inactive mutant of GH20 GlcNAcases. (A) Proposed mechanism of formate-mediated chemical rescue with the *Vh*GlcNAcase D437A mutant. Formate ion is involved in both glycosylation and deglycosylation steps by providing charge stabilization of transition states that flank the oxazolinium ion. (B) Proposed mechanism of azide-mediated chemical rescue with the *SpHex* D313A mutant. The azide ion acts in the deglycosylation step by opening the oxazolinium ion intermediate (Williams *et al.*, 2002). Hydroxyl groups and C6 have been omitted for clarity.

### 4.3 Expression, purification and molecular weight determination of *VhGlcNAcases*

The gene encoding *VhGlcNAcase* was cloned in the pQE-60 expression vector and all of the mutants were generated using the full-length *VhGlcNAcase* cDNA as a template (Suginta *et al.*, 2010). Both wild-type and mutated *GlcNAcases* were expressed in *E. coli* M15 (pREP) cells as C-terminally His<sub>6</sub>-tagged polypeptides that could be readily purified by cobalt affinity chromatography. For crystallization purposes, the wild type *VhGlcNAcases* was further purified using low-salt equilibration buffer (20 mM Tris-HCl buffer pH 8.0 containing 30 mM NaCl and 1 mM TCEP) by HiPrep (16/60) Sephacryl S-200 HR gel-filtration chromatography for getting a high purity and avoiding salt crystals. A chromatographic profile of the wild-type *VhGlcNAcases* obtained from gel filtration showed that the enzyme was found in two peaks (Figure 3.14A). As confirmed by native PAGE, the first peak was supposed to be aggregated protein with a molecular mass that was too large to enter the gel (Figure 3.14B, lane 1), while the second peak was non-aggregated protein that migrated with an apparent molecular mass of 75 kDa (Figure 3.14B, lane 2), corresponding to the monomeric form of *VhGlcNAcase*. In addition, the molecular mass of *VhGlcNAcase* calculated from a protein standard curve taken from size-exclusion chromatography was found to be 76.42 kDa (Figure 3.15B), confirming a monomeric form of *VhGlcNAcase* in solution. PISA analysis (Krissinel and Henrick, 2007) predicted no dimer formation in solution through this interface, which is consistent with the molecular mass of the active enzyme (~75 kDa) obtained from size-exclusion chromatography. The chromatographic profile of the D437A mutant was similar to that

of the wild-type enzyme, but the void peak was not seen, since the higher salt concentration helped to improve the protein solubility.

#### 4.4 Crystallization and structure determination of *VhGlcNAcases*

Conditions for *VhGlcNAcase* crystallization were initially screened using the sitting-drop vapour-diffusion method and further optimized by the hanging-drop vapour-diffusion method. After optimization in various conditions, good-quality crystals of the wild-type enzyme were obtained from the condition: 0.1 M sodium acetate pH 4.6, 1.4 M sodium malonate (Figure 3.22A), which diffracted to 2.4 Å resolution. The crystals of wild-type *VhGlcNAcase* grown under this condition were soaked with the GlcNAc<sub>2</sub> substrate for a short time (5 min) but without success. We also tried soaking the crystals of the native enzyme with the natural substrate GlcNAc<sub>2</sub> for various times, but we could not obtain X-ray diffraction data for the GlcNAcase-GlcNAc<sub>2</sub> complex because the substrate was degraded by the active enzyme. As the dimers formed in the crystals block some part of the active site, especially the +1 subsite, there would be insufficient space for accommodation of a larger ligand like GlcNAc<sub>2</sub>. Finally, the native enzyme complexed with a single GlcNAc molecule at -1 subsite was obtained under this condition. For the D437A mutant, the good-quality crystals were obtained from the condition: 20% (w/v) PEG 3350, 0.1 M Bis-Tris pH 7.5, 0.2 M sodium acetate (Figure 3.22B). We failed to obtain the inactive mutant D437A complexed with GlcNAc<sub>2</sub> either by both soaking or by co-crystallization, probably because substitution of the Asp437 residue with Ala affects not only the catalytic activity but also the affinity of the enzyme for its substrate. Both wild-type *VhGlcNAcase* and the D437A mutant crystals appeared as thick three-dimensional

plates and had a similar monoclinic space group  $P2_1$  in which the refined unit-cell parameters of the wild-type crystal are  $a = 90.2$ ,  $b = 130.7$ ,  $c = 98.5$  Å,  $\alpha = 90^\circ$ ,  $\beta = 113^\circ$  and  $\gamma = 90^\circ$ . The D437A mutant crystals also grew under the condition: 0.1 M MES pH 6.5 and 1.2 M sodium malonate as rod-shaped crystals belonging to the tetragonal space group  $P4_32_12$  (Figure 3.17D), with unit-cell parameters  $a = 165.1$ ,  $b = 165.1$ ,  $c = 155.2$  Å (Table 3.6). This difference could be explained by the crystal packing from different conditions used for growing crystals. The final statistics for the refinement of three forms of wild-type *VhGlcNAcase*, wild-type *VhGlcNAcase* in complex with GlcNAc and the D437A mutant were completed with  $R_{\text{factor}}$  values of 23.4%, 20.8% and 19.9%, respectively and  $R_{\text{free}}$  values of 26.4%, 25.2% and 25.0%, respectively (Table 3.7), indicating that all the structures were well refined.

The three-dimensional structure of GH20 *VhGlcNAcase* consists of three distinct domains (Figure 3.25). The first *N*-terminal domain of *VhGlcNAcase* belongs to the carbohydrate-binding domain (CBD) in which the topology is very similar to the chitin-binding domain of chitinase A from the same species (Songsiriritthigul *et al.*, 2008). The second domain of *VhGlcNAcase* has the most similar topology with the  $\alpha+\beta$  topology domain of other GH20  $\beta$ -hexosaminidases (Mark *et al.*, 2001; Tews *et al.*, 1996) but the function of this domain is still unknown. Third, the major  $(\beta/\alpha)_8$  TIM-barrel fold serves for the enzyme catalysis. However, the positions of  $\alpha 5$  and  $\alpha 7$  of the  $(\beta/\alpha)_8$ -barrel structure of *VhGlcNAcase* are missing and there is an extra helix at the end of  $\alpha 8$ . A similar structure is also found in human HexB (Mark *et al.*, 2003) and seems the common structural feature with other family 20 glycoside hydrolases.

Although there are two molecules per asymmetric unit in the crystal structures of *VhGlcNAcase*, the estimated molecular mass of *VhGlcNAcase* from size-exclusion

chromatography was calculated to be 76.42 kDa, suggesting a monomeric form of *VhGlcNAcase* in solution. Therefore, the two molecules per asymmetric unit seen in the crystal structures are a structural artifact.

The crystal structure of *VhGlcNAcase* in complex with the natural substrate showed that the  $\text{GlcNAc}_2$  was hydrolyzed by the wild-type crystals to GlcNAc molecules (Meekrathok *et al.*, 2015) of which one GlcNAc molecule remained at the active site (Figure 3.25B). The hydrolytic product, GlcNAc was found at the most stable subsite -1 and its conformation is similar to the other sugar conformations found in  $\beta$ -hexosaminidase from *Paenibacillus* sp. TS12 (Hex1T; PDB code: 3GH5) (Sumida *et al.*, 2009) and from *Streptomyces plicatus* (*SpHex*; PDB code: 1M01) (Williams *et al.*, 2002).

The dimer interface was found in the active site pocket of *VhGlcNAcase* (Figure 3.26B-C). Interestingly, the side chain of Glu438 assumes two conformations (B factor is  $36.6 \text{ \AA}^2$ ) depending on the binding with sugar. In the absence of sugar, the side chain of Glu438 rotated away from the active site. However, in the presence of sugar, the side chain of Glu438 rotated to the active site (Figure 3.26C). A similar conformational change of the catalytic residue was also found in the insect  $\beta$ -*N*-acetylhexosaminidase (*OfHex1*). The crystal structures of wild-type *OfHex1* and *OfHex1* in complex with TMG-chitotriomycin and PUGNAc revealed the obvious conformational changes of the catalytic residue Glu368 (equivalent to Glu438 in *VhGlcNAcase*) (Liu, Zhang, Liu, Chen *et al.*, 2011; Liu, Zhang, Liu, Wu *et al.*, 2011). In the absence of ligand, the side-chain of Glu368 rotated away from the active site. In contrast, the side chain of Glu368 rotated toward the active site in the presence of TMG-chitotriomycin and PUGNAc. In *VhGlcNAcase*, the polar groups are obviously located at the edge of the binding pocket

and accommodate the inner GlcNAc for specific interactions. The exit of the negatively charged catalytic pocket is blocked by Ser14-Asn19 loop of the *N*-terminal carbohydrate-binding domain of the neighbor molecule, which has a hydrophilic, but rather uncharged surface, resulting in Glu438 of monomer A being located 7.4 Å far from the sugar.

In the absence of GlcNAc, the catalytic Glu438 located at loop L4, widens the pocket around the subsite -1 whereas the side chain of Asp437 moves to the acetamido group of the GlcNAc (Figure 3.27A). Upon GlcNAc binding, the side chain of Glu438 also swings into the binding pocket, Asp437 moves toward the acetamido group of GlcNAc, Gln398 and the indole side chain Trp505 rotates from its original position closer to GlcNAc (Figure 3.27B), resulting in the edge of the substrate-binding pocket around subsite -1 is more narrow around 1 Å. This reflects conformational changes that are induced upon sugar binding. The flexibility of Glu438 enables the bond cleavage of chitooligosaccharide bound between subsites -1 and +1 while the movement of Asp437 may help the enzyme to stabilize the transition state via interaction with acetamido group of the substrate through substrate-assisted mechanism efficiency.

#### **4.5 Comparison of *Vh*GlcNAcases with other chitinases and GlcNAcases**

Apparently, the active site topologies of GlcNAcases and chitinases are not the same. Most chitinases have an endolytic mode of action, containing a long and deep binding groove for randomly binding with a very long chitooligosaccharide chain along the substrate binding cleft (Perrakis *et al.*, 1994; Songsiriritthigul *et al.*, 2008). On the other hand, most GlcNAcases are exolytic enzymes that digest the  $\beta$ -1,4-glycosidic

bonds from the non-reducing end in *N*-acetylglucosamine oligomers (mainly dimers) to *N*-acetylglucosamine (monomer). The active site of GlcNAcases is a small pocket in the  $(\beta/\alpha)_8$ -barrel which is suitable for the recognition of smaller oligosaccharides or a saccharide non-reducing extremity (Mark *et al.*, 2001; Tews *et al.*, 1996). This might indicate that this enzyme group prefers smaller oligosaccharides.

GH20 enzymes have the property to hydrolyze a broad range of substrates, including  $\beta(1-4)$ ,  $\beta(1-3)$ ,  $\beta(1-2)$  and  $\beta(1-6)$  glycosidic linkages as well as glycolipids, glycoproteins or sulfated glycoconjugates (Conzelmann *et al.*, 1978; Hechtman, 1977; Intra *et al.*, 2008; Jiang *et al.*, 2011; Manuel *et al.*, 2007; Mark *et al.*, 1998; Sumida *et al.*, 2009). However, the *VhGlcNAcase* cannot hydrolyse substrates such as *N*-linked glycans or gangliosides (Suginta *et al.*, 2010). The substrate specificity of *VhGlcNAcase* was compared with that of a homologue of *VhGlcNAcase*, *Streptococcus pneumoniae* *exo-β-N*-acetylglucosaminidase (StrH) (Pluvinage *et al.*, 2011), which is involved in the complete degradation of *N*-linked glycans of its human host (King, 2010), facilitating this pathogenic bacterium to invade host tissue. A comparison of the crystal structure of *VhGlcNAcase* with the StrH in complex with NGA2B reveals that overhanging loops L2, L3 and L7 of *VhGlcNAcase* protrude to the active site of StrH complexed with NGA2B, causing the substrate-binding region to be more narrow, particularly, loop L7 in *VhGlcNAcase* clashing the NGA2B of StrH (Figure 3.32B-C). This study confirmed that *VhGlcNAcase* cannot hydrolyse branched substrates like glycans, unlike in the case of StrH.

The crystal structure of *VhGlcNAcase* was further compared with other GH20 GlcNAcases, many of which were also complexed with GlcNAc. The domain organization of GH20 is different according to their amino acid sequence, length and

arrangement. The crystal structures of *VhGlcNAcase*, *SpHex* and *SmCHB* are compared in term of domain organization (Figure 3.28A). *VhGlcNAcase* is composed of three domains as mentioned previously whereas of the other GH20 enzymes, the *SpHex* has two domains comprising of 506 amino acids that are folded into the *N*-terminal domain (I), called  $\alpha+\beta$  topology domain, and the catalytic domain (II) containing a  $(\beta/\alpha)_8$  barrel at the *C*-terminus (Mark *et al.*, 2001). In marked contrast, *SmCHB* is composed of four domains designated as the *N*-terminal carbohydrate-binding domain (I),  $\alpha+\beta$  topology domain (II), TIM-barrel catalytic domain (III) and the *C*-terminal immunoglobulin-like domain (IV) (Tews *et al.*, 1996). The absence of *N*-terminal carbohydrate-binding domain (I) and the *C*-terminal immunoglobulin-like domain (IV) opens up a large groove that leads directly to a solvent-exposed active site at the *C*-terminal end of the TIM barrel catalytic domain in *SpHex*. In GH20 GlcNAcases, the acidic groups found in the sugar-binding pocket play an important role in proton donor/acceptor in catalysis and substrate coordination through hydrogen bonds. The active sites of *VhGlcNAcase* and the other two homologues are different. The sugar rings of *VhGlcNAcase* are well aligned with the chair conformation of -1GlcNAc of *SpHex* and with the sugar ring of -1GlcNAc of *SmCHB* but in different conformations (Figure 3.29). Namely, at subsite -1, the GlcNAc ring in *VhGlcNAcase* adopts the chair conformation whereas the GlcNAc ring of *SmCHB* adopts the boat conformation, with the torsion angles of the  $C_\alpha$  backbone that makes it susceptible to cleavage in active GlcNAcases, as found in the GH20 enzymes (Liu, Zhang, Liu, Wu *et al.*, 2011; Tews *et al.*, 1996). However, the outer GlcNAc from GlcNAc<sub>2</sub> showed a chair conformation at subsite +1. There are conformational changes in residues Arg274, His373, Asp437, Glu438, Trp505, Trp546 and Glu584 as compared to *SpHex* and *SmCHB* (Figure 3.29). The most significant

difference is the position of D437/E438 compared to the respective amino acids in the homologues, as due to the dimer contact in the crystal of *VhGlcNAcase* D437/E438 is pushed away from the sugar-binding pocket. However, as the Glu438 side chain of *VhGlcNAcase* forms a salt bridge with the guanidyl group of Arg21 of the other molecule in the dimer interface, the large distance to the GlcNAc is most likely a result of the dimer formed in the crystal and does not represent active the conformation of the monomer in solution. The dimer formation most likely causes the structural rearrangement of the loop next to D437/E438 causing their positions to be further from GlcNAc than in *SmCHB* (1QBB) and *SpHex* (1M01). The presence of the charged residues of H373, D303, R275 and E584 causes the active site to be very tight and seems to block the entrance of the sugar beyond to subsite -2. The active site pocket was found to be more open at +1 subsite and the presence of GlcNAc at subsite +1 in *SmCHB* suggests that Trp546 in *VhGlcNAcase* may play a role in stacking interactions with the incoming sugar at subsite +1.

#### **4.6 Kinetics of wild-type *VhGlcNAcase* and its mutants**

The main binding features in the active site of GH20 GlcNAcases are polar and charged residues, which form the hydrogen bonding network with the sugar as well as aromatic residues that bind with the GlcNAc ring through hydrophobic interactions (Figure 3.30). Within the active site, the amino acid side-chains and water molecules form the hydrogen bonding network with the GlcNAc, located in the middle of the TIM-barrel structure of catalytic domain at subsite -1. Considering the conformation rotating to the GlcNAc, Glu438 faces the oxygen atom of the glycosidic linkage, and may be the catalytic acid-base residue (although due to the crystal packing the D437-

E438 containing loop is pressed away from the GlcNAc binding pocket). The side chain of Asp437 is within 3.1 Å of the acetamido group of the GlcNAc and may facilitate the substrate-assisted retaining mechanism. The hydrolytic product, GlcNAc, derived from GlcNAc<sub>2</sub> and the position of the DE pair of Asp437-Glu438 are in a position around the β-1,4 glycosidic oxygen which is the ester bond being cleaved. The hydrophobic interactions between GlcNAc and aromatic/ hydrophobic residues in the active site surrounding the subsite -1 create the hydrophobic pocket wall in the active site. Trp582 was found to stack directly with the plane of the pyranose ring of the GlcNAc at subsite -1, and the interactions between the aromatic π electrons and polarized hydrogens of the sugar ring are expected to contribute to the binding with sugar. All aromatic residues are completely conserved with other GH20 enzymes and play an important role in substrate binding (Mark *et al.*, 2003; Mark *et al.*, 2001).

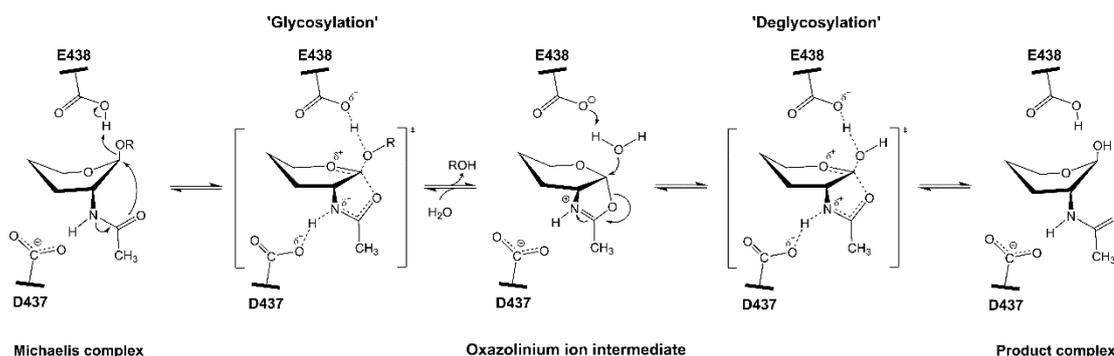
Based on the sequence alignment of *Vh*GlcNAcase with other GH20 GlcNAcases, the acidic groups were highly conserved in the TIM-barrel catalytic domain serving for enzyme catalysis. The crystal structure of *Vh*GlcNAcase in complex with GlcNAc revealed that the acidic residues such as Asp303, Asp304, Asp437, Glu438, Asp532 and Glu584 that are completely aligned with Asp191, Asp192, Asp313, Glu314, Asp395 and Glu444, respectively, of the *Sp*Hex, are located in the active site pocket and interact with the GlcNAc. Interestingly, His373 is located between these acidic groups and also highly conserved with other GH20 GlcNAcases. The specific hydrolyzing activity and kinetics of *Vh*GlcNAcase and its mutants D303A, D304A, D437N, D437A, E438Q, E438D, E438A, D532A and E584A were measured with *p*NP-GlcNAc as substrate using the previously described assay (Suginta *et al.*, 2010). All mutants showed a decrease in hydrolyzing activity against *p*NP-GlcNAc

(Figure 3.31). Mutations of the expected catalytic DE pair, D437-E438 showed a strongly decreased hydrolyzing activity as compared to the wild-type enzyme. Kinetics of wild-type *VhGlcNAcase* and its mutants showed that all mutants except E438 lost the binding affinity and enzymatic activity against *pNP-GlcNAc* which are reflected by the  $K_m$  and  $k_{cat}/K_m$ , respectively (Table 3.9). Mutations of the expected catalytic DE pair, D437-E438 showed a strongly decreased activity as compared with the wild-type enzyme. Mutants of D437 increase  $K_m$  against *pNP-GlcNAc* about 4.2-fold and 3.7-fold for the mutants D437A and D437N, respectively. The mutant D437A had very low  $k_{cat}$ , yielding the very low  $k_{cat}/K_m$  of  $0.36 \text{ s}^{-1} \text{ mM}^{-1}$  whereas the mutant D437N had high  $K_m$  and the lowest  $k_{cat}$ , yielding the lowest  $k_{cat}/K_m$  of  $0.04 \text{ s}^{-1} \text{ mM}^{-1}$  towards *pNP-GlcNAc*. The reductions in the  $k_{cat}/K_m$  values for these mutants are mainly reflected by the decrease in the  $k_{cat}$  values. The replacement with either Ala or Asn in the side chain of Asp437 may have effects on the transition state stabilization. The reduction in catalytic activity of the mutants D437A and D437N are a result of the changed charge of the amino acid, which results in the destabilization of reaction intermediates in the transition states. Mutations of E438 also showed that the mutants E438A, E438D and E438Q showed a slightly increased  $K_m$  (correlating to a decreased binding affinity) but strongly decreased hydrolyzing activity against *pNP-GlcNAc*. The relative  $k_{cat}/K_m$  values of E438A, E438D, E438Q were 0.28%, 0.30% and 0.35%, respectively. This finding provided supporting evidence that the presence of E438 around the cleavage site is important as the catalytic residue. Additionally, it was previously reported that changing a pair of Asp313-Glu314 in *SpHex* (Mark *et al.*, 1998; Williams *et al.*, 2002) and Asp539-Glu540 in *SmCHB* (Prag *et al.*, 2000), which are the catalytic pair for both *GlcNAcases* (equivalent to Asp437-Glu438 in *VhGlcNAcase*), showed a severely

decreased enzymatic activity as compared with the wild-type enzyme. For example, in wild-type *Sm*CHB, the  $k_{\text{cat}}/K_{\text{m}}$  value was reported to be  $1.3 \times 10^7 \text{ s}^{-1} \text{ mM}^{-1}$ . The mutant D539A in *Sm*CHB had very low  $k_{\text{cat}}$ , yielding the very low  $k_{\text{cat}}/K_{\text{m}}$  of  $8.0 \times 10^3 \text{ s}^{-1} \text{ mM}^{-1}$  (0.06% remaining activity) and the mutant E540A also had very low  $k_{\text{cat}}$ , yielding the low  $k_{\text{cat}}/K_{\text{m}}$  of  $4.2 \times 10^5 \text{ s}^{-1} \text{ mM}^{-1}$  (3.23% remaining activity) towards *p*NP-GlcNAc (Prag *et al.*, 2000). In human, the catalytic pair of Asp354-Glu355 in HexB also showed a strongly decreased activity after mutation (Hou *et al.*, 2001). On the other hand, the mutant E584A of *Vh*GlcNAcase showed an overall catalytic activity of 32%, suggesting that this residue is not crucial for enzyme catalysis. The most striking observations were made with mutations of the DD pair, showing decreased activity to various extents. The mutants D303A and D304A had a drastically decreased hydrolyzing activity (non-detectable activity). However, mutations to asparagine revealed that both D303N and D304N still had remaining activity that is reflected by  $k_{\text{cat}}/K_{\text{m}}$  of 1.8% and 0.75%, respectively. The difference in residual activity is a consequence of the amino-acid substitutions with different properties. As seen in Figure 3.30A, the GlcNAc is tightly anchored via hydrogen bonds at the OH3 and OH4 groups of the GlcNAc to Arg274 which is held in place by polar interactions with Asp271, Asp303, Asp304, Glu305 and water molecules in which the terminal amino groups of Arg274 make hydrogen bonds to Asp271 and Asp303. The carboxyl group of Asp303 also makes hydrogen bonds with Asp304, Arg274 and a water molecule coordinated with the GlcNAc substrate and the salt bridge most often arises from the anionic carboxylate of Asp or Glu (D271, D303, D304, E305 and E584) and the cationic guanidinium of Arg274. Mutations to Ala completely abolished the hydrogen bonding network, polar and electrostatic interactions whereas mutations to Asn may maintain the polar interactions to various

extents. The acidic pair, Asp303-Asp304 is suggested to play a concerted role with the other charged/polar residues in the active-site pocket for holding the Arg274 in place by polar interaction and hydrogen bonding. Additionally, mutation of H373 showed a drastic decrease in overall catalytic efficiency ( $k_{cat}/K_m$ ) as compared with the wild-type enzyme. H373 is located between the acidic groups of D303, D304 and D437, and deprotonates D437, thereby increasing its potential to stabilize the substrate-mediated transition state. Removing the imidazole ring by H373A mutation prevents stabilization of D437 and proton subtraction as well as an indirect interaction with the GlcNAc through a water molecule. The mutation studies indicate that E438-D437 are the catalytic residues, while the D303-D304 and E584 are involved in substrate coordination, and reduce binding upon charge inversion or removal.

Taken together with data obtained from the chemical rescue assay, kinetics and complex structures, *VhGlcNAcases* supports the proposal of a substrate-assisted mechanism of GH20 GlcNAcases, requiring the catalytic pair Asp437-Glu438 for catalysis. The  $\beta$ -(1,4)-glycosidic oxygen of the GlcNAc is protonated by the catalytic E438 in *VhGlcNAcase* which acts as a general acid to facilitate departure of the leaving group, leading to a bond cleavage (Figure 4.3). Subsequently, the primary nucleophile which is the C2-acetamido group of the GlcNAc moiety forms a positively charged oxazolinium ion intermediate during the transition states. The deprotonated D437 in *VhGlcNAcase* plays a critical role in orienting the C2-acetamido group and stabilizing the transition states, while the deprotonated E438 in *VhGlcNAcase*, which acts as a general base, aids attack of a water molecule at the anomeric carbon to hydrolyse the reaction intermediate, yielding the product complex with the retained  $\beta$ -configuration.



**Figure 4.3** Proposed catalytic mechanism of GH20 VhGlcNAcase.

#### 4.7 Common determinants of pK<sub>a</sub> calculations of GH18 chitinases

Family 18 chitinases contain a completely conserved DxDxE sequence motif. The first Asp (D<sub>1</sub>) is located in the TIM-barrel core at strand  $\beta$ 4 and rather far from the catalytic proton donor, glutamic acid (E) with the average distance between D<sub>1</sub> and E around  $8.3 \pm 0.1$  Å, which is too far for making a hydrogen bond between the D<sub>1</sub> and E whereas the distances between the catalytic proton donor, glutamic acid (E) and the oxygen in the scissile glycosidic bond range from 2.6-3.6 Å. Mutational studies of the DxDxE sequence motif were previously reported and suggested that the active-site acidic residues are crucial for the catalytic function of chitinases. Similar findings were also generated by mutations of the residues D<sub>1</sub>, D<sub>2</sub> or E, forming inactive enzymes, for example, D140A/N, D142A, E144A/Q in *SmChiB* (equivalent to D<sub>1</sub>, D<sub>2</sub> and E, respectively) (Synstad *et al.*, 2004), D313A/N and E315M/Q in *VhChiA* (equivalent to D<sub>2</sub> and E, respectively) (Suginta *et al.*, 2005; Suginta and Sritho, 2012) and residues D313A and E315Q in *SmChiA* (equivalent to D<sub>2</sub> and E, respectively) (Papanikolaou *et al.*, 2001). Additionally, the adjacent residues, tyrosine and aspartate (equivalent to Y214 and D215 in *SmChiB*) are highly conserved in chitinases and proposed to help in

catalysis. Mutation of Y214F in *SmChiB* showed a reduced activity by two orders of magnitude whereas a mutation of D215 showed a decrease in activity to different extents, namely, a greatly reduced activity of  $1 \times 10^4$ -fold was displayed in D215A but 4.2% remaining activity was found in D215N (Synstad *et al.*, 2004).

The  $pK_a$  values of three acidic residues D<sub>1</sub>, D<sub>2</sub> and E for 76 structures of 14 various chitinases were successfully calculated (Table 3.10 and 3.11) using the WIpK<sub>a</sub> used for  $pK_a$  calculation of active-site residues in other glycoside hydrolases (Joshi *et al.*, 2001; Nielsen and McCammon, 2003a) and can give accurate and useful information (Nielsen and Vriend, 2001; Nielsen and McCammon, 2003a; Nielsen and McCammon, 2003b).

An investigation of the conformation changes of three acidic residues, D<sub>1</sub>, D<sub>2</sub> and E in family 18 chitinases revealed that the conformation of D<sub>1</sub> is completely in the 'up' position whereas the conformation of E is almost in the 'down' position in both free and bound enzymes (Table 3.10). The conformation of D<sub>2</sub> is in the 'up' or 'down' position depending on the ligand binding (Table 3.10). In the absence of ligand, the distance between D<sub>1</sub> ('up' position) and D<sub>2</sub> ('down' position) for chitinases is found to be in a range 2.5-2.6 Å, and these findings suggest that the D<sub>1</sub> and D<sub>2</sub> make a hydrogen bond with each other. This aspect is found in family 18 chitinases such as ChiB from *Arthrobacter* sp., ChiA1 from *B. circulans* WL-12, ChiB and ChiA from *S. marcescens*, ChiA from *V. harveyi*, Chi1 from *C. immitis*, chitotriosidase and AMCase from human. In most chitinases, the calculated  $pK_a$  values of D<sub>1</sub> are less than 0 whereas the  $pK_a$  values of D<sub>2</sub> are on average around 11.3. The  $pK_a$  calculations of D<sub>1</sub> and D<sub>2</sub> strongly support the proposal that the D<sub>1</sub>-D<sub>2</sub> pair in chitinases holds exactly one negative charge over the whole accessible pH range, contributing to the destabilization of the buried negative

charge on the D<sub>1</sub>. The calculated pK<sub>a</sub> (E) in all chitinase forms varied from 6.0 to 9.0, indicating that this glutamate residue is protonated at pH values where the enzyme is in the active form. A contribution of the calculated pK<sub>a</sub> values of three acidic residues D<sub>1</sub>, D<sub>2</sub> and E (Figure 3.34) for every chitinase form with some outliers, suggests that the pK<sub>a</sub> values in the active sites of chitinases are quite consistent and the data presented here provide a general picture of the active-site pK<sub>a</sub> in chitinases.

However, some D<sub>2</sub> residues do not move to the 'up' position in the presence of ligand (Table 3.10). Ligands that are not fully occupied in the binding cleft, especially at subsite -1, may not induce the conformation change of D<sub>2</sub>. This aspect is supported by the crystal structures of *AfChiB1* in complex with the inhibitors, i.e., the D<sub>2</sub> is in the 'down' when it bound with theophylline (PDB code: 2A3A), caffeine (PDB code: 2A3B) and pentoxifylline (PDB code: 2A3C). In contrast, the D<sub>2</sub> is in the 'up' conformation for making a hydrogen bond with the reaction intermediate analog, allosamidin (Rao *et al.*, 2005).

In six structures, D<sub>1</sub> and D<sub>2</sub> showed calculated non-HH titration curves with fluctuations, resulting in an inaccurate pK<sub>a</sub> value. This aspect is not uncommon in detecting functional sites in proteins since the inter-atomic distance between D<sub>1</sub> and D<sub>2</sub> may be too short and have strong electrostatic interaction energies, which are typically found in enzyme active sites. In a related study, the key residues with such perturbed curves occur in the active site (Ondrechen *et al.*, 2001) and many of these key residues are charged residues that are located in unfavorable electrostatic interaction energy environments (Elcock, 2001). Interestingly, the conformation of D<sub>2</sub> is also occupied either in the 'down' or 'up' position (Table 3.10). In a high-resolution crystal structure of *SmChiA* without ligand (PDB code: 1EDQ), the side chain of Asp142 (equivalent to

D<sub>2</sub>) has two conformations. Due to the very short distance (2.3 Å) between D<sub>1</sub> and D<sub>2</sub>, the pK<sub>a</sub> values of D<sub>1</sub> and D<sub>2</sub> could not be calculated due to a strong electrostatic interaction between D<sub>1</sub> and D<sub>2</sub> as mentioned previously, but the pK<sub>a</sub> of E is still around 7.4 as usual.

The pK<sub>a</sub> values of D<sub>1</sub>, D<sub>2</sub> and E calculated from the structures complexed with the natural substrates or inhibitors (holo structure) were further investigated in details and compared to the ligand-free structures (apo and apo-generated structures) as shown in Table 3.11. The conformations of D<sub>1</sub>, D<sub>2</sub> and E in five chitinases are found completely to be 'up', 'up' and 'down' upon ligand binding. Most bound chitinases show that a rotation of Asp (D<sub>2</sub>) towards Glu (E) is more favourable as this brings the change closer to the solvent, allowing charge dispersion through sharing a proton with glutamate. The conformation of the Asp142 from *SmChiB* and the Asp313 from *SmChiA* (equivalent to D<sub>2</sub>) is found to be in the 'down' position in the apo enzyme as mentioned previously (Table 3.10) but the conformation of D<sub>2</sub> is changed to the 'up' position in the complex with allosamidin (van Aalten *et al.*, 2001; Papanikolau *et al.*, 2003) which is similar to what was found in chitotriosidase in complex with the substrate, GlcNAc<sub>2</sub> (Fusetti *et al.*, 2002). Crystal structures of *SmChiA*-Y390F in complex with GlcNAc<sub>6</sub> (Papanikolau *et al.*, 2001) and *SmChiB*-E144Q in complex with GlcNAc<sub>5</sub> (van Aalten *et al.*, 2001) showed that there is no space for D<sub>2</sub> to rotate back and forth when the substrate is bound. The results obtained from this investigation indicated that the rotation of D<sub>2</sub> towards the substrate may happen concomitantly with ligand-binding and substrate-distortion (Synstad *et al.*, 2004). In the absence of ligand, most calculated pK<sub>a</sub> values of D<sub>1</sub> are less than 0, however, the calculated pK<sub>a</sub> values of D<sub>1</sub> are increased in the presence of the natural substrates as shown in hevine A in

complex with GlcNAc<sub>3</sub> (PDB code: 1HVQ) (Terwisscha van Scheltinga *et al.*, 1994) and chitotriosidase in complex with GlcNAc<sub>2</sub> (PDB code: 1LG1) (Fusetti *et al.*, 2002), suggesting that the charges and radii of ligands affect the p*K*<sub>a</sub> values of D<sub>1</sub>. The calculated p*K*<sub>a</sub> values of D<sub>2</sub> for all five structures are still higher than D<sub>1</sub> and showed average values around 13.1 ± 0.8, 13.7 ± 1.2 and 12.4 ± 1.1 for the apo, apo-generated and holo structures, respectively. D<sub>2</sub> is located between D<sub>1</sub> and E in which the mutual electrostatic repulsion between the carboxylate groups may elevate the p*K*<sub>a</sub> values. In bacterial chitinases (*SmChiA* and *SmChiB*) bound with the ligand, the D<sub>2</sub> side chain detaches from D<sub>1</sub> and rotates to form a hydrogen bond with the side chain of Glu (E), resulting in a decrease of the calculated p*K*<sub>a</sub>(E) (see 1E15 vs 1E6R and 1CTN vs 1FFQ in Table 3.10). It is plausible that the protonated D<sub>2</sub> close to the protonated E (2.5 Å in *SmChiA* and 2.7 Å in *SmChiB* for D<sub>2</sub> and E) increases the acidity of the proton or lowers the p*K*<sub>a</sub> value on the protonated E that acts as the catalytic proton donor, leading to powerful assistance of leaving group departure and pulls electrons from the protonated E towards the developing positive charge that will become the oxazolinium ion intermediate during catalysis. In addition, a large decrease in all calculated p*K*<sub>a</sub> values of E between the holo structures (bound enzyme) and the apo-generated structure (free enzyme) (Table 3.11), suggests that the ligand also helps to lower the p*K*<sub>a</sub> values of the catalytic acid E, facilitating bond cleavage.

p*K*<sub>a</sub> calculations by the theoretical program (WIp*K*<sub>a</sub>) do not apply to the real situation during catalysis such as when the partial charges are formed, and with covalent bonds being formed and broken in the transition states. Structural details such as resolution and protein structure check scores should be considered (Table 3.10 and 3.11) and very low resolution structures should be avoided since different structures

can give different results in the quality of the calculated  $pK_a$  values, especially in the active site (Nielsen and McCammon, 2003b).

#### 4.8 Computational pH activity profiles of GH18 chitinases

The computational pH-activity profiles of family 18 chitinases were predicted based on three acidic active-site residues (D<sub>1</sub>, D<sub>2</sub> and E) that are in close proximity and completely conserved in family 18 chitinases (Figure 3.33A-B). These D<sub>1</sub>-D<sub>2</sub>-E groups must be in a specific protonation state to provide the proton or charges required in catalytic mechanism and in this study, the catalytically competent protonation state (CCPS) of the D<sub>1</sub>-D<sub>2</sub>-E of chitinases was supposed to be charged, neutral and neutral, respectively. This hypothesis has been supported by inspecting the crystal structures in complex with their natural substrates and the reaction intermediate analogues and kinetics by site-directed mutagenesis also give an insight into the catalytic mechanism (Terwisscha van Scheltinga *et al.*, 1994; Papanikolau *et al.*, 2001; van Aalten *et al.*, 2001 Bokma *et al.*, 2002; Bortone *et al.*, 2002; Suginta and Sritho, 2012).

The computational pH-profiles predicted based on the calculations of D<sub>1</sub>, D<sub>2</sub> and E in apo enzymes (Figure 3.36) are equivalent to the  $k_{cat}/K_m$ -pH profiles. However, the experimentally measured pH profiles available in previous work exist as both  $k_{cat}/K_m$ -pH profiles (Chou *et al.*, 2006; Hurtado-Guerrero and van Aalten, 2007; Suginta and Sritho, 2012) and  $k_{cat}$ -pH profiles (van Aalten *et al.*, 2001; Boot *et al.*, 2005; Zees *et al.*, 2009). The basic limb of computational pH profiles (equivalent to  $k_{cat}/K_m$ -pH profiles) of *VhChiA* and human AMCase showed a good agreement with their experimentally measured  $k_{cat}/K_m$ -pH profiles and the  $pK_a$  from the experimental and computational profiles are most likely to be the calculated  $pK_a$  of Glu (E). The acidic

limb of the computational pH profiles from *VhChiA*, *ScChi* and human AMCase could not be determined accurately (Figure 3.36) since the residue on the acidic side of the curve is always deprotonated and most likely to be the charged form of D<sub>1</sub>, showing its p*K*<sub>a</sub> less than zero. Unlike with *VhChiA* and human AMCase, the computational pH profiles of *ScChi1* show an upward shift around 1 pH unit as compared with its experimentally measured *k*<sub>cat</sub>/*K*<sub>m</sub>-pH profile. In the other case, the computational pH-profiles (equivalent to *k*<sub>cat</sub>/*K*<sub>m</sub>-pH profiles) of *SmChiB*, *SmChiA* and human chitotriosidase showed a downward shift as compared with the experimentally measured *k*<sub>cat</sub>-pH profiles. This is not uncommon for a comparison of *k*<sub>cat</sub>/*K*<sub>m</sub>- and *k*<sub>cat</sub>-pH profiles in which the p*K*<sub>a</sub> calculation with charge and radii of neutral substrates are most likely to increase the p*K*<sub>a</sub> of Glu (E).

Unlike the computational pH-profiles predicted using apo enzymes, the computational pH-profiles predicted based on the calculations of D<sub>1</sub>, D<sub>2</sub> and E in holo enzymes (Figure 3.38) are equivalent to the *k*<sub>cat</sub>-pH profiles in which the apparent p*K*<sub>a</sub> values reflect to the p*K*<sub>a</sub> values in the enzyme-substrate complex (Kyte, 1995). The basic limb of computational pH profiles of human chitotriosidase complexed with chitobiose and hevamine A complexed with chitotriose shows a good agreement with their experimentally measured *k*<sub>cat</sub>-pH profiles and its p*K*<sub>a</sub> from the experimental and computational profiles are most likely to be the calculated p*K*<sub>a</sub> of Glu (E) whereas the acidic limb and the p*K*<sub>a</sub> shows the ionization of the residue that are close to the calculated p*K*<sub>a</sub> of D<sub>1</sub>.

The mutational effects on the active site residues, D<sub>1</sub>, D<sub>2</sub> or E to the pH activity profile were previously reported. In *SmChiB*, the D140N mutant (equivalent to D<sub>1</sub>) displayed an acidic shift in the pH profile, indicating that the D140N mutation lowers

the  $pK_a$  of key catalytic residues (Synstad *et al.*, 2004). Mutations of D313A/N in *VhChiA* showed that the pH optimum for the D313 mutants (equivalent to D<sub>2</sub>) was shifted upward by 0.8 pH units, indicating that the ionization of D313 influences the protein environment around the cleavage site (Suginta and Sritho, 2012). These strongly support the  $pK_a$  calculations (Table 3.11) that D<sub>2</sub> helps to lower the  $pK_a$  of the catalytic residue Glu (E). The D<sub>2</sub> is likely to play a role in pulling electrons from E towards the developing positive charge that will become the oxazolinium ion intermediate during catalysis, as discussed previously. Replacement by a less polarizable asparagine of D<sub>2</sub> would reduce the ability for electron withdrawal.

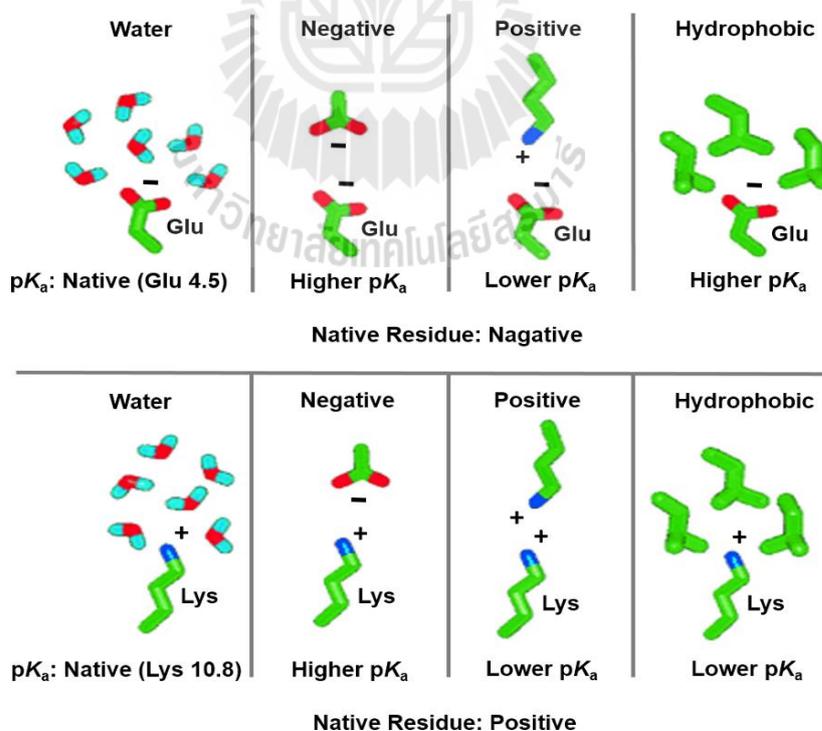
The pH profiles from the calculations with substrate show a great improvement in producing a more bell-shaped profile as compared with the profiles calculated without the substrate. The difference in pH profiles can be caused by strong interactions and mobility in the D<sub>1</sub>-D<sub>2</sub>-E system involved with several interacting groups. The  $pK_a$  calculations with the natural substrate are close to the real situation for the chitinase activity assay against the chromogenic or fluorescent substrates. However, the differences can be listed as follows: chitinase assays were carried out with different substrates, buffers and pH that are not similar to the free energy-based  $pK_a$  calculation. In general, any  $pK_a$  calculation will be more accurate if the pH used for solving the crystal is closest to the  $pK_a$  value that one is interested in. Such a complex of the human chitotriosidase with GlcNAc<sub>2</sub> was obtained by soaking the crystals in GlcNAc<sub>5</sub> with CAPS buffer pH 10.6 to prevent acid-catalyzed hydrolysis of the substrate (Fusetti *et al.*, 2002) even though the pH optimum of the enzyme is around pH 6 (Boot *et al.*, 2005). However, the accuracy of the predictions and analyses using a  $pK_a$  calculation will always depend on the intrinsic  $pK_a$  values and the site-site interaction energy

matrix, which are highly dependent on the protein structure, force field,  $pK_a$  calculation parameter set up and procedures used in the  $pK_a$  calculation.

#### 4.9 Comparison of $pK_a$ values between bacterial and human chitinases

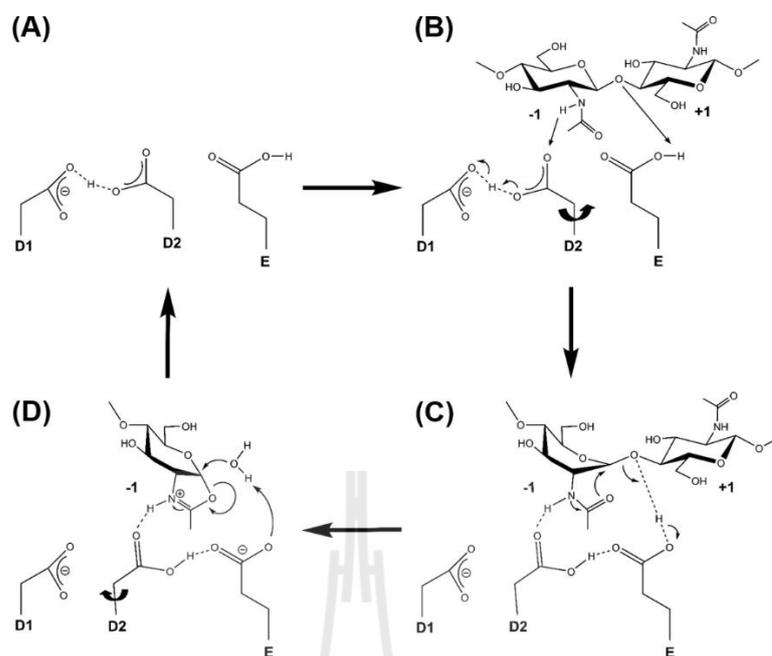
Chitinases from different sources have a broad range of pH optima, which are determined by the  $pK_a$  values of the titratable groups in the active site. In comparison within bacterial chitinases, the  $k_{cat}$ -pH profiles of *VhChiA* and *SmChiB* are different, namely *VhChiA* has a broad optimum pH around 5-9 while *SmChiB* has an optimum pH around 6-7. The difference in the optimum pH profiles is from the calculated  $pK_a$  values of D<sub>2</sub> and E in *VhChiA* showing an increase as compared with *SmChiB* (Figure 3.39A). *VhChiA* is an endochitinase which cleaves the chitin chain randomly at internal sites (Suginta *et al.*, 2004) and has a groove-like structure (Songsiriritthigul *et al.*, 2008) whereas *SmChiB* is an exochitinase which hydrolyses the chitin chain from the non-reducing end and successively releases diacetylchitobiose (GlcNAc<sub>2</sub>) units and has a tunnel-like structure (van Aalten *et al.*, 2001). Although *VhChiA* and *SmChiB* have different modes of chitin degradation and domain organization, the  $pK_a$  calculation of the active site residues D<sub>1</sub>, D<sub>2</sub> and E do not show statistically different  $pK_a$  values, suggesting that the mode of enzyme action and 3D structure do not apply in the  $pK_a$  system. As compared with the human AMCase with its optimum  $k_{cat}/K_m$ -pH profile around pH 4-5 (Chou *et al.*, 2006), the  $pK_a$  of D<sub>1</sub> (D311 in *VhChiA* and D136 in *HsAMCase*) cannot be distinguished but the  $pK_a$  values of D<sub>2</sub> and E of *HsAMCase* showed a decrease of about 2.6 and 2.0 pH units, respectively as compared with *VhChiA* (Figure 3.39B). A decrease in the  $pK_a$  values of active site residues results in the downward shift in its pH optimum to 4-5.

In most chitinases, the active site clefts are lined with aromatic amino acids that play an important role for substrate binding (van Aalten *et al.*, 2000; Uchiyama *et al.*, 2001; Suginta *et al.*, 2007; Songsiriritthigul *et al.*, 2008). In *VhChiA*, the presence of W168 instead of P14 in *SmChiB* and other aromatic residues line in the substrate-binding cleft (Figure 3.39C) may increase the  $pK_a$  of the D<sub>1</sub>-D<sub>2</sub>-E system. In human AMCase, the presence of the basic residues R145 and H208 (equivalent to G320 and F387 in *VhChiA*) in the substrate-binding cleft of human AMCase (Figure 3.39D) may decrease the  $pK_a$  of the D<sub>1</sub>-D<sub>2</sub>-E system, causing the pH optimum of human AMCase to be lower than *VhChiA*. The surrounding environment of a titratable group plays an important role in determining the  $pK_a$  value of that residue. Environmental effects on the  $pK_a$  of titratable residues are summarized in Figure 4.4.



**Figure 4.4** Summary of the effect of placing a titratable group in a negative, positive and hydrophobic environment (modified from Nielsen *et al.*, 2001).

Taken together, the  $pK_a$  calculations of the active site residues and pH-profile prediction lead to a proposed catalytic mechanism of family18 chitinases in general (Figure 4.5). First: Pre-priming (Figure 4.5A), the enzyme is in the free form and the acidic pair, D<sub>1</sub> and D<sub>2</sub> share a proton together with a distance around 2.5-2.6 Å since most  $pK_a$  values of D<sub>1</sub> are less than 0 whereas the  $pK_a$  values of D<sub>2</sub> are on average around 11.3. Second: Binding of substrate (Figure 4.5B), the side chain of D<sub>2</sub> abstracts a proton from D<sub>1</sub> and rotates to form a hydrogen bond with the C2-acetamido group of the GlcNAc at subsite -1. This can be seen in Table 3.10 and 3.11 showing the rotation of D<sub>2</sub> upon the ligand binding. Third: On formation of oxazolinium ion intermediate (Figure 4.5C), the pyranose ring of -1GlcNAc is distorted to a boat or skewed boat conformation. The side chain of D<sub>2</sub> rotates toward the side chain of E and stabilizes the reaction intermediate by forming two hydrogen bonds with the NH- of the C2-acetamido group and with the carboxyl group of E. The glycosidic bond is then cleaved by the nucleophilic attack of the carbonyl oxygen of the C2-acetamido group on C1, with protonation of the glycosidic oxygen by E. Fourth: During hydrolysis of the oxazolinium ion intermediate (Figure 4.5D), the nucleophilic attack from the activated water molecule leads to the protonation of E and rotation back of the side chain of D<sub>2</sub> to its original conformation where it shares a proton with D<sub>1</sub>.



**Figure 4.5** The refined catalytic mechanism of family 18 chitinases during chitin degradation supported by the calculated active-site  $pK_a$  values.

## CHAPTER V

### CONCLUSION

This research describes the functional and structural characterization of GH20  $\beta$ -*N*-acetylglucosaminidase (GlcNAcase) from *Vibrio harveyi* and  $pK_a$  calculations of GH18 chitinases. The studies are divided into three parts. The first part is focused on the investigation of the catalytic residues of *Vh*GlcNAcase by chemical rescue. All mutants D303A, D303N, D304A, D304N, H373A, D437A, D437N, E438A, E438D, E438Q, D532A and E584A of the *Vh*GlcNAcase were successfully cloned with correct sequence using the site-directed mutagenesis method. Each mutant was successfully expressed in *E. coli* M15 (pREP) cells as a 652-amino acid polypeptide, including the C-terminal (His)<sub>6</sub> sequence. In this study, we have demonstrated that an exogenous nucleophile (formate ion) selectively enhances the enzymatic activity of an inactive mutant *Vh*GlcNAcase, D437A, in a concentration-dependent manner. However, the activity of other active-site mutants (D303A/N, D304A/N, and E438A/Q) was not significantly affected by the addition of this strong nucleophile. The rescued activity of the D437A mutant suggests that Asp437 is the catalytic nucleophile, while its invariant acidic partner Glu438 likely acts as a catalytic proton-donating residue. This experimental evidence confirms that the residues Asp437 and Glu438, located in the middle of the substrate-binding cleft in the modelled structure of GH20 *Vh*GlcNAcase, act as the catalytic pair in the catalytic cycle of chitooligosaccharide hydrolysis by this enzyme.

The second part provides the enzyme kinetics and structural insights into the catalytic mechanism and binding of substrate to *VhGlcNAcase*, which plays an important role in the chitin degradation process by hydrolyzing chitin fragments to *N*-acetylglucosamine (GlcNAc) monomers. Three crystal structures of *VhGlcNAcase*, a catalytically inactive mutant D437A and the wild type, in the absence and presence of *N*-acetylglucosamine (GlcNAc), were solved. The *VhGlcNAcase* structure contains the carbohydrate-binding domain, the  $\alpha+\beta$  topology domain and the major  $(\beta/\alpha)_8$  TIM barrel catalytic domain which is similar to other GH20 GlcNAcases. Size-exclusion chromatography showed that *VhGlcNAcase* is a monomeric enzyme with a MW of 75 kDa in solution. Enzyme kinetic analysis of the mutants in the active site suggests that the catalytic DE pair, Asp437 and Glu438 may play an important role in the enzyme catalysis. Superimposition of the crystal structures of ligand-free and ligand-bound *VhGlcNAcase* suggests that binding of the GlcNAc induces local conformational changes around the sugar-binding pocket that promotes the sugar-enzyme interactions for substrate hydrolysis. The crystal structure of *VhGlcNAcase* in complex with GlcNAc shows that the interaction with the sugar is mainly through a hydrogen bonding network as well as hydrophobic interactions. The presence of loops L7 (residues 525-551) seems to obstruct the binding pocket to larger substrates that contain branches. As a result, the active site of *VhGlcNAcase* is a small pocket or crater that is suitable for the recognition of smaller oligosaccharides at the non-reducing extremity.

The third part involved a calculation of the active site  $pK_a$  values of the GH18 chitinases and confirmation of the catalytic mechanism based on the  $pK_a$  calculation as well as prediction the pH-activity profiles of chitinases. Family 18 chitinases contain a completely conserved DxDxE sequence motif that plays a crucial role in catalysis. The

conformation changes and  $pK_a$  values of the active-site residues, D<sub>1</sub>, D<sub>2</sub> and E studied here are able to give us a general picture of the catalytic mechanism of GH18 chitinases. We found that D<sub>1</sub> is completely in the 'up' position whereas the catalytic proton donor E is mainly in the 'down' position. D<sub>2</sub> is found in two conformations depending on the bound ligand. The calculated  $pK_a$  values of D<sub>1</sub> are less than 0 whereas the  $pK_a$  values of D<sub>2</sub> have an average value around 11.3. The calculations of D<sub>1</sub> and D<sub>2</sub> strongly suggest that the D<sub>1</sub>-D<sub>2</sub> pair in chitinases holds exactly one negative charge over the whole accessible pH range contributing to the destabilization of the buried negative charge on the D<sub>1</sub>. The calculated  $pK_a$  values of E in all the chitinases varied from 6.0 to 9.0, indicating that this glutamate residue is protonated at pH values where the enzyme is in the active form. The D<sub>2</sub> orients the C2-acetamido group of the -1GlcNAc and helps to lower the  $pK_a$  of the catalytic acid E, facilitating a bond cleavage. The computational pH-profiles suggest that D<sub>1</sub>-D<sub>2</sub>-E must be charged, neutral and neutral, respectively, for most chitinases to be active. The acidic and basic limbs of the computational pH profile are most likely to be determined by the D<sub>1</sub> and E, respectively. Taken together, the calculated active-site  $pK_a$  values and computational pH-profiles give us a general picture for the catalytic mechanism of GH18 chitinases.



## **REFERENCES**

## REFERENCES

- Alam, M. M., Mizutani, T., Isono, M., Nikaidou, N., and Watanabe, T. (1996). Three chitinase genes (*chiA*, *chiC* and *chiD*) comprise the chitinase system of *Bacillus circulans* WL-12. **J. Ferment. Bioeng.** 82: 28-36.
- Alexov, E. G. and Gunner, M. R. (1997). Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. **Biophys. J.** 72: 2075-2093.
- Alexov, E. G. and Gunner, M. R. (1999). Calculated protein and proton motions coupled to electron transfer: electron transfer from QA<sup>-</sup> to QB in bacterial photosynthetic reaction centers. **Biochemistry.** 38: 8253-8270.
- Armand, S., Tomita, H., Heyraud, A., Gey, C., Watanabe, T., and Henrissat, B. (1994). Stereochemical course of the hydrolysis reaction catalyzed by chitinases A1 and D from *Bacillus circulans* WL-12. **FEBS. Lett.** 343: 177-180.
- Arndt, C., Koristka, S., Bartsch, H., and Bachmann, M. (2012). Native polyacrylamide gels. **Methods Mol. Biol.** 869: 49-53.
- Aronson, N. N. Jr., Halloran, B. A., Alexyev, M. F., Amable, L., Madura, J. D., Pasupulati, L., Worth, C., and Van Roey, P. (2003). Family 18 chitinaseoligosaccharide substrate interaction: subsite preference and anomer selectivity of *Serratia marcescens* chitinase A. **Biochem. J.** 376: 87-95.
- Bacik, J. P., Whitworth, G. E., Stubbs, K. A., Vocadlo, D. J., and Mark, B. L. (2012). Active site plasticity within the glycoside hydrolase NagZ underlies a dynamic

- mechanism of substrate distortion. **Chem. Biol.** 19: 1471-1482.
- Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. **Proc. Natl. Acad. Sci. USA.** 98: 10037-10041.
- Balcewich, M. D., Stubbs, K. A., He, Y., James, T. W., Davies, G. J., Vocadlo, D. J., and Mark, B. L. (2009). Insight into a strategy for attenuating AmpC-mediated beta-lactam resistance: structural basis for selective inhibition of the glycoside hydrolase NagZ. **Protein Sci.** 18: 1541-1551.
- Bassler, B. L., Gibbons, P. J., Yu, C., and Roseman, S. (1991). Chitin utilization by marine bacteria. Chemotaxis to chitin oligosaccharides by *Vibrio furnissii*. **J. Biol. Chem.** 266: 24268-24275.
- Bassler, B. L., Yu, C., Lee, Y. C., and Roseman, S. (1991). Chitin utilization by marine bacteria. Degradation and catabolism of chitin oligosaccharides by *Vibrio furnissii*. **J. Biol. Chem.** 266: 24276-24286.
- Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. **Acta Crystallogr. Sect. D-Biol. Crystallogr.** 67: 271-281.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The Protein Data Bank. **Nucleic Acids Res.** 28: 235-242.
- Beroza, P., Fredkin, D. R., Okamura, M. Y., and Feher, G. (1991). Protonation of interacting residues in a protein by a Monte Carlo method: application to lysozyme and the photosynthetic reaction center of *Rhodobacter sphaeroides*. **Proc. Natl. Acad. Sci. USA.** 88: 5804-5808.

- Blackwell, J. (1969). Structure of  $\beta$ -chitin or parallel chain systems of poly- $\beta$ -(1-4)-*N*-acetyl-D-glucosamine. **Biopolymers**. 7: 281-298.
- Bokma, E., Rozeboom, H. J., Sibbald, M., Dijkstra, B. W., and Beintema, J. J. (2002). Expression and characterization of active site mutants of hevamine, a chitinase from the rubber tree *Hevea brasiliensis*. **Eur. J. Biochem**. 269: 893-901.
- Bolar, J. P., Norelli, J. L., Harman, G. E., Brown, S. K., and Aldwinckle, H. S. (2001). Synergistic activity of endochitinase and exochitinase from *Trichoderma atroviride* (*T. harzianum*) against the pathogenic fungus (*Venturia inaequalis*) in transgenic apple plants. **Transgenic Res**. 10: 533-543.
- Boot, R. G., Bussink, A. P., Verhoek, M., de Boer, P. A., Moorman, A. F., and Aerts, J. M. (2005). Marked differences in tissue-specific expression of chitinases in mouse and man. **J. Histochem. Cytochem**. 53: 1283-1292.
- Bortone, K., Monzingo, A. F., Ernst, S., and 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.
- Brameld, K. A. and Goddard, W. A., 3rd. (1998). The role of enzyme distortion in the single displacement mechanism of family 19 chitinases. **Proc. Natl. Acad. Sci. USA**. 95: 4276-4281.
- Brameld, K. A., Shrader, W. D., Imperiali, B., and Goddard, W. A., 3rd. (1998). Substrate assistance in the mechanism of family 18 chitinases: theoretical studies of potential intermediates and inhibitors. **J. Mol. Biol**. 280: 913-923.

- Brameld, K. A. and Goddard, W. A. (1998). Substrate distortion to a boat conformation at subsite -1 is critical in the mechanism of family 18 chitinases. **J. Am. Chem. Soc.** 120: 3571-3580.
- Brurberg, M. B., Eijsink, V. G., Haandrikman, A. J., Venema, G., and Nes, I. F. (1995). Chitinase B from *Serratia marcescens* B JL200 is exported to the periplasm without processing. **Microbiology.** 141: 123-131.
- Brurberg, M. B., Nes, I. F., and Eijsink, V. G. (1996). Comparative studies of chitinases A and B from *Serratia marcescens*. **Microbiology.** 142: 1581-1589.
- Brurberg, M. B., Synstad, B., Klemsdal, S. S., van Aalten, D. M. F., Sundheim, L., and Eijsink, V. G. H. (2001). Chitinases from *Serratia marcescens*. **Recent Res. Dev. Microbiol.** 5: 187-204.
- Bussink, A. P., Speijer, D., Aerts, J. M., and Boot, R. G. (2007). Evolution of mammalian chitinase(-like) members of family 18 glycosyl hydrolases. **Genetics.** 177: 959-970.
- Chang, W. T., Chen, Y. C., and Jao, C. L. (2007). Antifungal activity and enhancement of plant growth by *Bacillus cereus* grown on shellfish chitin wastes. **Bioresour. Technol.** 98: 1224-1230.
- Chen, J. K., Shen, C. R., and Liu, C. L. (2010). *N*-acetylglucosamine: production and applications. **Mar. Drugs.** 8: 2493-2516.
- Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. **Acta Crystallogr. Sect. D-Biol. Crystallogr.** 66: 12-21.

- Choi, K-H., Seo, J. Y., Park, K-M., Park, C-S., and Cha, J. (2009). Characterization of glycosyl hydrolase family 3  $\beta$ -N-acetylglucosaminidases from *Thermotoga maritima* and *Thermotoga neapolitana*. **J. Biosci. Bioeng.** 108: 455-459.
- Chou, Y. T., Yao, S., Czerwinski, R., Fleming, M., Krykbaev, R., Xuan, D., Zhou, H., Brooks, J., Fitz, L., Strand, J., Presman, E., Lin, L., Aulabaugh, A., and Huang, X. (2006). Kinetic characterization of recombinant human acidic mammalian chitinase. **Biochemistry.** 45: 4444-4454.
- Cobucci-Ponzano, B., Trincone, A., Giordano, A., Rossi, M., and Moracci, M. (2003). Identification of the catalytic nucleophile of the family 29  $\alpha$ -L-fucosidase from *Sulfolobus solfataricus* via chemical rescue of an inactive mutant. **Biochemistry.** 42: 9525-9531.
- Cohen-Kupiec, R. and Chet, I. (1998). The molecular biology of chitin digestion. **Curr.Opin. Biotechnol.** 9: 270-277.
- Collaborative Computational Project Number 4. (1994). The CCP4 suite: programs for protein crystallography. **Acta Crystallogr. Sect. D-Biol. Crystallogr.** 50: 760-763.
- Comfort, D. A., Bobrov, K. S., Ivanen, D. R., Shabalin, K. A., Harris, J. M., Kulminskaya, A. A., Brumer, H., and Kelly, R. M. (2007). Biochemical analysis of *Thermotoga maritima* GH36  $\alpha$ -Galactosidase (*TmGalA*) confirms the mechanistic commonality of clan GH-D glycoside hydrolases. **Biochemistry.** 46: 3319-3330.
- Conzelmann, E., Sandhoff, K., Nehr Korn, H., Geiger, B., and Arnon, R. (1978). Purification, biochemical and immunological characterisation of

- hexosaminidase A from variant AB of infantile G<sub>M2</sub> gangliosidosis. **Eur. J. Biochem.** 84: 27-33.
- Cousin, M. A. (1996). Chitin as a measure of mold contamination of agricultural commodities and foods. **J. Food Prot.** 59: 73-81.
- Cremer, D. and Pople, J. A. (1975). General definition of ring puckering coordinates. **J. Am. Chem. Soc.** 97: 1354-1358.
- Davies, G. and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. **Structure.** 3: 853-859.
- Di Rosa, M., Malaguarnera, G., De Gregorio, C., Drago, F., and Malaguarnera, L. (2013). Evaluation of CHI3L-1 and CHIT-1 expression in differentiated and polarized macrophages. **Inflammation.** 36: 482-492.
- Donnelly, L. E. and Barnes, P. J. (2004). Acidic mammalian chitinase - a potential target for asthma therapy. **Trends Pharmacol. Sci.** 25: 509-511.
- Dowd, P. F., Johnson, E. T., and Pinkerton, T. S. (2007). Oral toxicity of  $\beta$ -N-acetyl hexosaminidase to insects. **J. Agric. Food Chem.** 55: 3421-3428.
- Drouillard, S., Armand, S., Davies, G. J., Vorgias, C. E., and Henrissat, B. (1997). *Serratia marcescens* chitobiase is a retaining glycosidase utilizing substrate acetamido group participation. **Biochem. J.** 328: 945-949.
- Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. **Biochemistry.** 6: 1948-1954.
- Elcock, A. H. (2001). Prediction of functionally important residues based solely on the computed energetics of protein structure. **J. Mol. Biol.** 312: 885-896.

- Elias, J. A., Homer, R. J., Hamid, Q., and Lee, C. G. (2005). Chitinases and chitinase-like proteins in  $T_H2$  inflammation and asthma. **J. Allergy Clin. Immunol.** 116: 497-500.
- Emsley, P. and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. **Acta Crystallogr. Sect. D-Biol. Crystallogr.** 60: 2126-2132.
- Fujita, K., Sato, R., Toma, K., Kitahara, K., Suganuma, T., Yamamoto, K., and Takegawa, K. (2007). Identification of the catalytic acid base residue of arthrobacter endo-beta-*N*-acetylglucosaminidase by chemical rescue of an inactive mutant, **J. Biochem.** 142: 301-306.
- Fusetti, F., von Moeller, H., Houston, D., Rozeboom, H. J., Dijkstra, B. W., Boot, R. G., Aerts, J. M., and 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: 25537-25544.
- Georgescu, R. E., Alexov, E. G., and Gunner, M. R. (2002). Combining conformational flexibility and continuum electrostatics for calculating  $pK_{as}$  in proteins. **Biophys. J.** 83: 1731-1748.
- Gill, S. C. and von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. **Anal. Biochem.** 182: 319-326.
- Gilson, M. K. (1993). Multiple-site titration and molecular modeling: two rapid methods for computing energies and forces for ionizable groups in proteins. **Proteins.** 15: 266-282.
- Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S., and Onufriev, A. (2005). H++: a server for estimating  $pK_{as}$  and adding missing hydrogens to macromolecules. **Nucleic Acids Res.** 33: W368-371.

- Gravel, R. A., Clarke, J. T. R., Kaback, M. M., Mahuran, D., Sandhoff, K., and Suzuki, K. (1995). The GM<sub>2</sub> gangliosidoses. In *The Metabolic Basis of Inherited Disease* (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds) 7th Ed., pp. 1807-1839, McGraw-Hill, New York.
- Hart, P. J., Pfluger, H. D., Monzingo, A. F., Hollis, T., and Robertus, J. D. (1995). The refined crystal structure of an endochitinase from *Hordeum vulgare* L. seeds at 1.8 Å resolution. **J. Mol. Biol.** 248: 402-413.
- Harish Prashanth, K. V. and Tharanathan, R. N. (2007). Chitin/chitosan: modifications and their unlimited application potential - an overview. **Trends Food Sci. Technol.** 18: 117-131.
- Hasegawa, H. and Holm, L. (2009). Advances and pitfalls of protein structural alignment. **Curr. Opin. Struct. Biol.** 19: 341-348.
- Hechtman, P. (1977). Characterization of an activating factor required for hydrolysis of Gm<sub>2</sub> ganglioside catalyzed by hexosaminidase A. **Can. J. Biochem.** 55: 315-324.
- Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. **Biochem. J.** 280: 309-316.
- Henrissat, B. and Bairoch, A. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. **Biochem. J.** 293: 781-788.
- Henrissat, B. and Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. **Curr. Opin. Struct. Biol.** 7: 637-644.

- Henrissat, B. and Davies, G. J. (2000). Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics. **Plant Physiol.** 124: 1515-1519.
- Herrera-Estrella, A. and Chet, I. (1999). Chitinases in biological control. **EXS.** 87: 171-184.
- Hirano, S. (1999). Chitin and chitosan as novel biotechnological materials. **Polym. Int.** 48: 732-734.
- Hogenkamp, D. G., Arakane, Y., Kramer, K. J., Muthukrishnan, S., and Beeman, R. W. (2008). Characterization and expression of the  $\beta$ -*N*-acetylhexosaminidase gene family of *Tribolium castaneum*. **Insect Biochem. Mol. Biol.** 38: 478-489.
- Hollak, C. E., van Weely, S., van Oers, M. H., and Aerts, J. M. (1994). Marked elevation of plasma chitotriosidase activity. A novel hallmark of Gaucher disease. **J. Clin. Invest.** 93: 1288-1292.
- Hollis, T., Monzingo, A. F., Bortone, K., Ernst, S., Cox, R., and Robertus, J. D. (2000). The X-ray structure of a chitinase from the pathogenic fungus *Coccidioides immitis*. **Protein Sci.** 9: 544-551.
- Holm, L. and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. **J. Mol. Biol.** 233: 123-138.
- Hoof, R. W., Sander, C., and Vriend, G. (1996). Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. **Proteins.** 26: 363-376.
- Horn, S. J., Sikorski, P., Cederkvist, J. B., Vaaje-Kolstad, G., Sorlie, M., Synstad, B., Vriend, G., Varum, K. M., and Eijsink, V. G. (2006). Costs and benefits of

- processivity in enzymatic degradation of recalcitrant polysaccharides. **Proc. Natl. Acad. Sci. USA.** 103: 18089-18094.
- Horsch, M., Mayer, C., Sennhauser, U., and Rast, D. M. (1997).  $\beta$ -*N*-acetylhexosaminidase: a target for the design of antifungal agents. **Pharmacol. Ther.** 76: 187-218.
- Hossain, M. A., Rana, M. M., Kimura, Y., and Roslan, H. A. (2014). Changes in biochemical characteristics and activities of ripening associated enzymes in mango fruit during the storage at different temperatures. **Biomed Res. Int.** 2014, 232969.
- Hou, Yongmin, Vocadlo, David J., Leung, Amy, Withers, Stephen G., and Mahuran, Don. (2001). Characterization of the Glu and Asp residues in the active site of human  $\beta$ -Hexosaminidase B. **Biochemistry.** 40: 2201-2209.
- Hunt, D. E., Gevers, D., Vahora, N. M., and Polz, M. F. (2008). Conservation of the chitin utilization pathway in the *Vibrionaceae*. **Appl. Environ. Microbiol.** 74: 44-51.
- Hurtado-Guerrero, R. and van Aalten, D. M. (2007). Structure of *Saccharomyces cerevisiae* chitinase 1 and screening-based discovery of potent inhibitors. **Chem. Biol.** 14: 589-599.
- Ilnkovan, P., Hein, S., Ng, C-H., Trung, T. S., and Stevens, W. F. (2006). Production of *N*-acetyl chitobiose from various chitin substrates using commercial enzymes. **Carbohydr. Polym.** 63: 245-250.
- Intra, J., Pavesi, G., and Horner, D. S. (2008). Phylogenetic analyses suggest multiple changes of substrate specificity within the glycosyl hydrolase 20 family. **BMC Evol. Biol.** 8: 214.

- Iseli, B., Armand, S., Boller, T., Neuhaus, Jean-Marc, and Henrissat, B. (1996). Plant chitinases use two different hydrolytic mechanisms. **FEBS. Lett.** 382: 186-188.
- Jagadeesh, B. H., Prabha, T. N., and Srinivasan, K. (2004a). Activities of glycosidases during fruit development and ripening of tomato (*Lycopersicon esculantum* L.): implication in fruit ripening. **Plant Sci.** 166: 1451-1459.
- Jagadeesh, B. H., Prabha, T. N., and Srinivasan, K. (2004b). Activities of  $\beta$ -hexosaminidase and  $\alpha$ -mannosidase during development and ripening of bell capsicum (*Capsicum annuum* var. *variata*). **Plant Sci.** 167: 1263-1271.
- Jaques, A. K., Fukamizo, T., Hall, D., Barton, R. C., Escott, G. M., Parkinson, T., Hitchcock, C. A., and Adams, D. J. (2003). Disruption of the gene encoding the ChiB1 chitinase of *Aspergillus fumigatus* and characterization of a recombinant gene product. **Microbiology.** 149: 2931-2939.
- Jeffrey, G. A. and Yates, J. H. (1979). Stereographic representation of the cremer-pople ring-puckering parameters for pyranoid rings. **Carbohydr. Res.** 74: 319-322.
- Jeuniaux, C., and Voss-Foucart, M. F. (1991). Chitin biomass and production in the marine environment. **Biochem. Syst. Ecol.** 19: 347-356.
- Jiang, Y. L., Yu, W. L., Zhang, J. W., Frolet, C., Di Guilmi, A. M., Zhou, C. Z., Vernet, T., and Chen, Y. (2011). Structural basis for the substrate specificity of a novel  $\beta$ -N-acetylhexosaminidase StrH protein from *Streptococcus pneumoniae* R6. **J. Biol. Chem.** 286: 43004-43012.
- Jones, J. D., Grady, K. L., Suslow, T. V., and Bedbrook, J. R. (1986). Isolation and characterization of genes encoding two chitinase enzymes from *Serratia marcescens*. **EMBO J.** 5: 467-473.

- Jorgensen, W. L. and Tirado-Rives, J. (1988). The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. **J. Am. Chem. Soc.** 110: 1657-1666.
- Joshi, M. D., Sidhu, G., Nielsen, J. E., Brayer, G. D., Withers, S. G., and McIntosh, L. P. (2001). Dissecting the electrostatic interactions and pH-dependent activity of a family 11 glycosidase. **Biochemistry.** 40: 10115-10139.
- Kabsch, W. (2010). XDS. **Acta Crystallogr. Sect. D-Biol. Crystallogr.** 66: 125-132.
- Kanneganti, M., Kamba, A., and Mizoguchi, E. (2012). Role of chitotriosidase (chitinase 1) under normal and disease conditions. **J. Epithel. Biol. Pharmacol.** 5: 1-9.
- Kato, Y., Onishi, H., and Machida, Y. (2003). Application of chitin and chitosan derivatives in the pharmaceutical field. **Curr. Pharm. Biotechnol.** 4: 303-309.
- Kawada, M., Hachiya, Y., Arihiro, A., and Mizoguchi, E. (2007). Role of mammalian chitinases in inflammatory conditions. **Keio J. Med.** 56: 21-27.
- Keyhani, N. O. and Roseman, S. (1999). Physiological aspects of chitin catabolism in marine bacteria. **Biochim. Biophys. Acta.** 1473: 108-122.
- Kezuka, Y., Ohishi, M., Itoh, Y., Watanabe, J., Mitsutomi, M., Watanabe, T., and Nonaka, T. (2006). Structural studies of a two-domain chitinase from *Streptomyces griseus* HUT6037. **J. Mol. Biol.** 358: 472-484.
- Kim, S. and Rajapakse, N. (2005). Enzymatic production and biological activities of chitosan oligosaccharides (COS): A review. **Carbohydr. Polym.** 62: 357-368.
- Kim, S., Matsuo, I., Ajisaka, K., Nakajima, H., and Kitamoto, K. (2002). Cloning and characterization of the *nagA* gene that encodes  $\beta$ -N-acetylglucosaminidases

- from *Aspergillus nidulans* and its expression in *Aspergillus oryzae*. **Biosci. Biotechnol. Biochem.** 10: 2168-2175.
- King, S. J. (2010). Pneumococcal modification of host sugars: a major contributor to colonization of the human airway? **Mol. Oral. Microbiol.** 25: 15-24.
- Koide, S. S. (1998). Chitin-chitosan: Properties, benefits and risks. **Nutr. Res.** 18: 1091-1101.
- Kolstad, G., Synstad, B., Eijsink, V. G., and van Aalten, D. M. (2002). Structure of the D140N mutant of chitinase B from *Serratia marcescens* at 1.45 Å resolution. **Acta Crystallogr. Sect. D-Biol. Crystallogr.** 58: 377-379.
- Koshland, D. E., Jr. and Clarke, E. (1953). Mechanism of hydrolysis of adenosinetriphosphate catalyzed by lobster muscle. **J. Biol. Chem.** 205: 917-924.
- Krissinel, E. and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. **J. Mol. Biol.** 372: 774-797.
- Kuranda, M. J. and Robbins, P. W. (1991). Chitinase is required for cell separation during growth of *Saccharomyces cerevisiae*. **J. Biol. Chem.** 266: 19758-19767.
- Kyte, J. (1995). *Mechanism in protein chemistry*, pp. 258-283, Garland Publishing, New York.
- Kzhyshkowska, J., Gratchev, A., and Goerdts, S. (2007). Human chitinases and chitinase-like proteins as indicators for inflammation and cancer. **Biomark. Insights.** 2: 128-146.
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. **Nature.** 227: 680-685.

- Langley, D. B., Harty, D. W., Jacques, N. A., Hunter, N., Guss, J. M., and Collyer, C. A. (2008). Structure of *N*-acetyl- $\beta$ -D-glucosaminidase (GcnA) from the endocarditis pathogen *Streptococcus gordonii* and its complex with the mechanism-based inhibitor NAG-thiazoline. **J. Mol. Biol.** 377: 104-116.
- Laskowski, R. A., Moss, D. S., and Thornton, J. M. (1993). Main-chain bond lengths and bond angles in protein structures. **J. Mol. Biol.** 231: 1049-1067.
- Laskowski, R. A. and Swindells, M. B. (2011). LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. **J. Chem. Inf. Model.** 51: 2778-2786.
- Lee, C. G. (2009). Chitin, chitinases and chitinase-like proteins in allergic inflammation and tissue remodeling. **Yonsei Med. J.** 50: 22-30.
- Lemieux, M. J., Mark, B. L., Cherney, M. M., Withers, S. G., Mahuran, D. J., and James, M. N. (2006). Crystallographic structure of human  $\beta$ -hexosaminidase A: interpretation of Tay-Sachs mutations and loss of GM<sub>2</sub> ganglioside hydrolysis. **J. Mol. Biol.** 359: 913-929.
- Li, X. and Roseman, S. (2004). The chitinolytic cascade in *Vibrios* is regulated by chitin oligosaccharides and a two-component chitin catabolic sensor/kinase. **Proc. Natl. Acad. Sci. USA.** 101: 627-631.
- Litzinger, S., Fischer, S., Polzer, P., Diederichs, K., Welte, W., and Mayer, C. (2010). Structural and kinetic analysis of *Bacillus subtilis* *N*-acetylglucosaminidase reveals a unique Asp-His dyad mechanism. **J. Biol. Chem.** 285: 35675-35684.
- Liu, T., Zhang, H., Liu, F., Chen, L., Shen, X., and Yang, Q. (2011). Active-pocket size differentiating insectile from bacterial chitinolytic  $\beta$ -*N*-acetyl-D-hexosaminidases. **Biochem. J.** 438: 467-474.

- Liu, T., Zhang, H., Liu, F., Wu, Q., Shen, X., and Yang, Q. (2011). Structural determinants of an insect  $\beta$ -N-acetyl-D-hexosaminidase specialized as a chitinolytic enzyme. **J. Biol. Chem.** 286: 4049-4058.
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. **Nucleic Acids Res.** 42: D490-495.
- Lonhienne, T., Mavromatis, K., Vorgias, C. E., Buchon, L., Gerday, C., and Bouriotis, V. (2001). Cloning, sequences, and characterization of two chitinase genes from the Antarctic *Arthrobacter* sp. strain TAD20: isolation and partial characterization of the enzymes. **J. Bacteriol.** 183: 1773-1779.
- Lu, Y., Zen, K. C., Muthukrishnan, S., and Kramer, K. J. (2002). Site-directed mutagenesis and functional analysis of active site acidic amino acid residues D142, D144 and E146 in *Manduca sexta* (tobacco hornworm) chitinase. **Insect Biochem. Mol. Biol.** 32: 1369-1382.
- MacLeod, A. M., Lindhorst, T., Withers, S. G., and Warren, R. A. (1994). The acid/base catalyst in the exoglucanase/xylanase from *Cellulomonas fimi* is glutamic acid 127: evidence from detailed kinetic studies of mutants. **Biochemistry.** 33: 6371-6376.
- Madura, J. D., Briggs, J. M., Wade, R. C., Davis, M. E., Luty, B. A., Ilin, A., Antosiewicz, J., Gilson, M. K., Bagheri, B., and Scott, L R. (1995). Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. **Comput. Phys. Commun.** 91: 57-95.

- Maier, T., Strater, N., Schuette, C. G., Klingenstein, R., Sandhoff, K., and Saenger, W. (2003). The X-ray crystal structure of human  $\beta$ -Hexosaminidase B provides new insights into Sandhoff disease. **J. Mol. Biol.** 328: 669-681.
- Manuel, S. G., Ragunath, C., Sait, H. B., Izano, E. A., Kaplan, J. B., and Ramasubbu, N. (2007). Role of active-site residues of dispersin B, a biofilm-releasing  $\beta$ -hexosaminidase from a periodontal pathogen, in substrate hydrolysis. **FEBS J.** 274: 5987-5999.
- Mark, B. L., Mahuran, D. J., Cherney, M. M., Zhao, D., Knapp, S., and James, M. N. (2003). Crystal structure of human  $\beta$ -hexosaminidase B: understanding the molecular basis of Sandhoff and Tay-Sachs disease. **J. Mol. Biol.** 327: 1093-1109.
- Mark, B. L., Vocadlo, D. J., Knapp, S., Triggs-Raine, B. L., Withers, S. G., and James, M. N. (2001). Crystallographic evidence for substrate-assisted catalysis in a bacterial  $\beta$ -hexosaminidase. **J. Biol. Chem.** 276: 10330-10337.
- Mark, B. L., Wasney, G. A., Salo, T. J., Khan, A. R., Cao, Z., Robbins, P. W., James, M. N., and Triggs-Raine, B. L. (1998). Structural and functional characterization of *Streptomyces plicatus*  $\beta$ -N-acetylhexosaminidase by comparative molecular modeling and site-directed mutagenesis. **J. Biol. Chem.** 273: 19618-19624.
- Matthews, B. W. (1968). Solvent content of protein crystals. **J. Mol. Biol.** 33: 491-497.
- Matthews, B. W. and Remington, S. J. (1974). The three dimensional structure of the lysozyme from bacteriophage T4. **Proc. Natl. Acad. Sci. USA.** 71: 4178-4182.

- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007). Phaser crystallographic software. **J. Appl. Crystallogr.** 40: 658-674.
- Meekrathok, P., Burger, M., Porfetye, A. T., Vetter, I. R., and Suginta, W. (2015). Expression, purification, crystallization and preliminary crystallographic analysis of a GH20  $\beta$ -N-acetylglucosaminidase from the marine bacterium *Vibrio harveyi*. **Acta Crystallogr. Sect. F-Struct. Biol. Commun.** 71: 427-433.
- Melchers, L. S. and Stuiver, M. H. (2000). Novel genes for disease-resistance breeding. **Curr. Opin. Plant. Biol.** 3: 147-152.
- Merzendorfer, H. and Zimoch, L. (2003). Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. **J. Exp. Biol.** 206: 4393-4412.
- Mine, S., Kado, Y., Watanabe, M., Fukuda, Y., Abe, Y., Ueda, T., Kawarabayasi, Y., Inoue, T., and Ishikawa, K. (2014). The structure of hyperthermophilic  $\beta$ -N-acetylglucosaminidase reveals a novel dimer architecture associated with the active site. **FEBS J.** 281: 5092-5103.
- Minke, R. and Blackwell, J. (1978). The structure of  $\alpha$ -chitin. **J. Mol. Biol.** 120: 167-181.
- Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. **Acta Crystallogr. Sect. D-Biol. Crystallogr.** 53: 240-255.
- Myerowitz, R. (1997). Tay-Sachs disease-causing mutations and neutral polymorphisms in the Hex A gene. **Hum. Mutat.** 9: 195-208.

- Nakajima, M., Atsumi, K., Kifune, K., Miura, K., and Kanamaru, H. (1986). Chitin is an effective material for sutures. **Jpn. J. Surg.** 16: 418-424.
- Nicholls, A. and Honig, B. (1991). A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. **J. Comput. Chem.** 12: 435-445.
- Nielsen, J. E. (2007). Analysing the pH-dependent properties of proteins using  $pK_a$  calculations. **J. Mol. Graph. Model.** 25: 691-699.
- Nielsen, J. E. (2009). Analyzing enzymatic pH activity profiles and protein titration curves using structure-based  $pK_a$  calculations and titration curve fitting. **Methods Enzymol.** 454: 233-258.
- Nielsen, J. E., Borchert, T. V., and Vriend, G. (2001). The determinants of  $\alpha$ -amylase pH-activity profiles. **Protein Eng.** 14: 505-512.
- Nielsen, J. E. and McCammon, J. A. (2003a). Calculating  $pK_a$  values in enzyme active sites. **Protein Sci.** 12: 1894-1901.
- Nielsen, J. E. and McCammon, J. A. (2003b). On the evaluation and optimization of protein X-ray structures for  $pK_a$  calculations. **Protein Sci.** 12: 313-326.
- Nielsen, J. E. and Vriend, G. (2001). Optimizing the hydrogen-bond network in Poisson-Boltzmann equation-based  $pK_a$  calculations. **Proteins.** 43: 403-412.
- Ohno, T., Armand, S., Hata, T., Nikaidou, N., Henrissat, B., Mitsutomi, M., and Watanabe, T. (1996). A modular family 19 chitinase found in the prokaryotic organism *Streptomyces griseus* HUT 6037. **J. Bacteriol.** 178: 5065-5070.
- Ondrechen, M. J., Clifton, J. G., and Ringe, D. (2001). THEMATIC: a simple computational predictor of enzyme function from structure. **Proc. Natl. Acad. Sci. USA.** 98: 12473-12478.

- Paal, K., Ito, M., and Withers, S. G. (2004). *Paenibacillus* sp. TS12 glucosylceramidase: kinetic studies of a novel sub-family of family 3 glycosidases and identification of the catalytic residues. **Biochem. J.** 378: 141-149.
- Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. **Protein Sci.** 4: 2411-2423.
- Papanikolau, Y., Prag, G., Tavlas, G., Vorgias, C. E., Oppenheim, A. B., and Petratos, K. (2001). High resolution structural analyses of mutant chitinase A complexes with substrates provide new insight into the mechanism of catalysis. **Biochemistry.** 40: 11338-11343.
- Papanikolau, Y., Tavlas, G., Vorgias, C. E., and Petratos, K. (2003). De novo purification scheme and crystallization conditions yield high-resolution structures of chitinase A and its complex with the inhibitor allosamidin. **Acta Crystallogr. Sect. D-Biol. Crystallogr.** 59: 400-403.
- Park, J. K., Keyhani, N. O., and Roseman, S. (2000). Chitin catabolism in the marine bacterium *Vibrio furnissii*. Identification, molecular cloning, and characterization of A *N*, *N'*-diacetylchitobiose phosphorylase. **J. Biol. Chem.** 275: 33077-33083.
- Patil, R. S., Ghormade, V. V., and Deshpande, M. V. (2000). Chitinolytic enzymes: an exploration. **Enzyme Microb. Technol.** 26: 473-483.
- Perrakis, A., Tews, I., Dauter, Z., Oppenheim, A. B., Chet, I., Wilson, K. S., and Vorgias, C. E. (1994). Crystal structure of a bacterial chitinase at 2.3 Å resolution. **Structure.** 2: 1169-1180.

- Phillips, David C. (1967). The hen egg-white lysozyme molecule. **Proc. Natl. Acad. Sci. USA.** 57: 483-495.
- Pluvinaige, B., Higgins, M. A., Abbott, D. W., Robb, C., Dalia, A. B., Deng, L., Weiser, J. N., Parsons, T. B., Fairbanks, A. J., Vocadlo, D. J., and Boraston, A. B. (2011). Inhibition of the pneumococcal virulence factor StrH and molecular insights into N-glycan recognition and hydrolysis. **Structure.** 19: 1603-1614.
- Prag, G., Papanikolaou, Y., Tavlas, G., Vorgias, C. E., Petratos, K., and Oppenheim, A. B. (2000). Structures of chitinase mutants complexed with the substrate Di-N-acetyl-D-glucosamine: the catalytic role of the conserved acidic pair, aspartate 539 and glutamate 540. **J. Mol. Biol.** 300: 611-617.
- Qin, C., Li, H., Xiao, Q., Liu, Y., Zhu, J., and Du, Y. (2006). Water-solubility of chitosan and its antimicrobial activity. **Carbohydr. Polym.** 63: 367-374.
- Ramasubbu, N., Thomas, L. M., Raganath, C., and Kaplan, J. B. (2005). Structural analysis of dispersin B, a biofilm-releasing glycoside hydrolase from the periodontopathogen *Actinobacillus actinomycetemcomitans*. **J. Mol. Biol.** 349: 475-486.
- Rao, F. V., Dorfmüller, H. C., Villa, F., Allwood, M., Eggleston, I. M., and van Aalten, D. M. F. (2006). Structural insights into the mechanism and inhibition of eukaryotic O-GlcNAc hydrolysis. **EMBO J.** 25: 1569-1578.
- Ravi Kumar, M. N. V. (2000). A review of chitin and chitosan applications. **React. Funct. Polym.** 46: 1-27.
- Rhodes, G. (2000). Crystallography made crystal clear. A guide for users of macromolecular models (2nd ed.), pp. 133-158, Academic Press, New York.

- Rinaudo, M. (2006). Chitin and chitosan: properties and applications. **Prog. Polym. Sci.** 31: 603-632.
- Robbins, P. W., Overbye, K., Albright, C., Benfield, B., and Pero, J. (1992). Cloning and high-level expression of chitinase-encoding gene of *Streptomyces plicatus*. **Gene.** 111: 69-76.
- Robert, X. and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. **Nucleic Acids Res.** 42: W320-324.
- Sahai, A. S. and Manocha, M. S. (1993). Chitinases of fungi and plants: their involvement in morphogenesis and host-parasite interaction. **FEMS Microbiol. Rev.** 11: 317-338.
- Sandhoff, K. and Kolter, T. (1998). Processing of sphingolipid activator proteins and the topology of lysosomal digestion. **Acta Biochim. Pol.** 45: 373-384.
- Sashiwa, H., Fujishima, S., Yamano, N., Kawasaki, N., Nakayama, A., Muraki, E., Sukwattanasinitt, M., Pichyangkurac, R., and Aiba, S. (2003). Enzymatic production of *N*-acetyl-D-glucosamine from chitin. Degradation study of *N*-acetyl chitooligosaccharide and the effect of mixing of crude enzymes. **Carbohydr. Polym.** 51: 391-395.
- Schüttelkopf, A. W. and van Aalten, D. M. F. (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. **Acta Crystallogr. Sect. D-Biol. Crystallogr.** 60: 1355-1363.
- Seidl, V. (2008). Chitinases of filamentous fungi: a large group of diverse proteins with multiple physiological functions. **Fungal Biol. Rev.** 22: 36-42.
- Shallom, D., Belakhov, V., Solomon, D., Shoham, G., Baasov, T., and Shoham, Y. (2002). Detailed kinetic analysis and identification of the nucleophile in  $\alpha$ -L-

- arabinofuranosidase from *Geobacillus stearothermophilus* T-6, a Family 51 glycoside hydrolase. **J. Biol. Chem.** 277: 43667-43673.
- Sinnott, M. L. (1990). Catalytic mechanism of enzymic glycosyl transfer. **Chem. Rev.** 90: 1171-1202.
- Sirimontree, P., Fukamizo, T., and Suginta, W. (2015). Azide anions inhibit GH-18 endochitinase and GH-20 exo  $\beta$ -*N*-acetylglucosaminidase from the marine bacterium *Vibrio harveyi*. **J. Biochem.** pii: mvv087. [Epub ahead of print].
- Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Klenk, D. C. (1985). Measurement of protein using bicinchoninic acid. **Anal. Biochem.** 150: 76-85.
- Songsiriritthigul, C., Pantoom, S., Aguda, A. H., Robinson, R. C., and Suginta, W. (2008). Crystal structures of *Vibrio harveyi* chitinase A complexed with chitooligosaccharides: implications for the catalytic mechanism. **J. Struct. Biol.** 162: 491-499.
- Struszczyk, M. H. (2006). Global requirements for medical applications of chitin and its derivatives. **Pol. Chitin Soc. Monograph XI**: 95-102.
- Suginta, W., Chuenark, D., Mizuhara, M., and Fukamizo, T. (2010). Novel  $\beta$ -*N*-acetylglucosaminidases from *Vibrio harveyi* 650: cloning, expression, enzymatic properties, and subsite identification. **BMC Biochem.** 11: 40.
- Suginta, W., Chumjan, W., Mahendran, K. R., Janning, P., Schulte, A., and Winterhalter, M. (2013). Molecular uptake of chitooligosaccharides through chitoporin from the marine bacterium *Vibrio harveyi*. **PLoS One.** 8: e55126.

- Suginta, W., Chumjan, W., Mahendran, K. R., Schulte, A., and Winterhalter, M. (2013). Chitoporin from *Vibrio harveyi*, a channel with exceptional sugar specificity. **J. Biol. Chem.** 288: 11038-11046.
- Suginta, W., Pantoom, S., and Prinz, H. (2009). Substrate binding modes and anomer selectivity of chitinase A from *Vibrio harveyi*. **J. Chem. Biol.** 2: 191-202.
- Suginta, W., Robertson, P. A., Austin, B., Fry, S. C., and Fothergill-Gilmore, L. A. (2000). Chitinases from *Vibrio*: activity screening and purification of chiA from *Vibrio carchariae*. **J. Appl. Microbiol.** 89: 76-84.
- Suginta, W., Songsiriritthigul, C., Kobdaj, A., Opassiri, R., and Svasti, J. (2007). Mutations of Trp275 and Trp397 altered the binding selectivity of *Vibrio carchariae* chitinase A. **Biochim. Biophys. Acta.** 1770: 1151-1160.
- Suginta, W. and Sritho, N. (2012). Multiple roles of Asp313 in the refined catalytic cycle of chitin degradation by *Vibrio harveyi* chitinase A. **Biosci. Biotechnol. Biochem.** 76: 2275-2281.
- Suginta, W., Vongsuwan, A., Songsiriritthigul, C., Prinz, H., Estibeiro, P., Duncan, R. R., Svasti, J., and Fothergill-Gilmore, L. A. (2004). An endochitinase A from *Vibrio carchariae*: cloning, expression, mass and sequence analyses, and chitin hydrolysis. **Arch. Biochem. Biophys.** 424: 171-180.
- Suginta, W., Vongsuwan, A., Songsiriritthigul, C., Svasti, J., and Prinz, H. (2005). Enzymatic properties of wild-type and active site mutants of chitinase A from *Vibrio carchariae*, as revealed by HPLC-MS. **FEBS J.** 272: 3376-3386.
- Sumida, T., Ishii, R., Yanagisawa, T., Yokoyama, S., and Ito, M. (2009). Molecular cloning and crystal structural analysis of a novel  $\beta$ -N-acetylhexosaminidase

from *Paenibacillus* sp. TS12 capable of degrading glycosphingolipids. **J. Mol. Biol.** 392: 87-99.

Suzuki, K., Sugawara, N., Suzuki, M., Uchiyama, T., Katouno, F., Nikaidou, N., and Watanabe, T. (2002). Chitinases A, B, and C1 of *Serratia marcescens* 2170 produced by recombinant *Escherichia coli*: enzymatic properties and synergism on chitin degradation. **Biosci. Biotechnol. Biochem.** 66: 1075-1083.

Suzuki, K., Taiyoji, M., Sugawara, N., Nikaidou, N., Henrissat, B., and Watanabe, T. (1999). The third chitinase gene (*chiC*) of *Serratia marcescens* 2170 and the relationship of its product to other bacterial chitinases. **Biochem. J.** 343: 587-596.

Synstad, B., Gaseidnes, S., Van Aalten, D. M., Vriend, G., Nielsen, J. E., and Eijsink, V. G. (2004). Mutational and computational analysis of the role of conserved residues in the active site of a family 18 chitinase. **Eur. J. Biochem.** 271: 253-262.

Terwisscha van Scheltinga, A. C., Armand, S., Kalk, K. H., Isogai, A., Henrissat, B., and Dijkstra, B. W. (1995). Stereochemistry of chitin hydrolysis by a plant chitinase/lysozyme and X-ray structure of a complex with allosamidin: evidence for substrate assisted catalysis. **Biochemistry.** 34: 15619-15623.

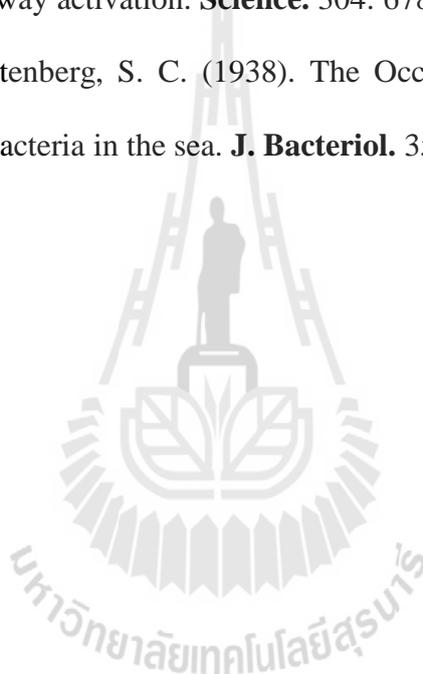
Terwisscha van Scheltinga, A. C., Kalk, K. H., Beintema, J. J., and Dijkstra, B. W. (1994). Crystal structures of hevamine, a plant defence protein with chitinase and lysozyme activity, and its complex with an inhibitor. **Structure.** 2: 1181-1189.

- Tews, I., Perrakis, A., Oppenheim, A., Dauter, Z., Wilson, K. S., and Vorgias, C. E. (1996). Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease. **Nature Struct. Biol.** 3: 638-648.
- Tews, I., Terwisscha van Scheltinga, A. C., Perrakis, A., Wilson, K. S., and Dijkstra, B. W. (1997). Substrate-assisted catalysis unifies two families of chitinolytic enzymes. **J. Am. Chem. Soc.** 119: 7954-7959.
- Thi, N. N., Offen, W. A., Shareck, F., Davies, G. J., and Doucet, N. (2014). Structure and activity of the *Streptomyces coelicolor* A3(2)  $\beta$ -N-acetylhexosaminidase provides further insight into GH20 family catalysis and inhibition. **Biochemistry.** 53: 1789-1800.
- Thompson, J. D., Higgins, D. G., and 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: 4673-4680.
- Toseland, C. P., McSparron, H., Davies, M. N., and Flower, D. R. (2006). PPD v1.0 - an integrated, web-accessible database of experimentally determined protein  $pK_a$  values. **Nucleic Acids Res.** 34: D199-203.
- Uchiyama, T., Katouno, F., Nikaidou, N., Nonaka, T., Sugiyama, J., and Watanabe, T. (2001). Roles of the exposed aromatic residues in crystalline chitin hydrolysis by chitinase A from *Serratia marcescens* 2170. **J. Biol. Chem.** 276: 41343-41349.
- Vaaje-Kolstad, G., Houston, D. R., Rao, F. V., Peter, M. G., Synstad, B., van Aalten, D. M., and Eijsink, V. G. (2004). Structure of the D142N mutant of the family

- 18 chitinase ChiB from *Serratia marcescens* and its complex with allosamidin. **Biochim. Biophys. Acta.** 1696: 103-111.
- Vallmitjana, M., Ferrer-Navarro, M., Planell, R., Abel, M., Ausín, C., Querol, E., Planas, A., and Pérez-Pons, J. A. (2001). Mechanism of the family 1 beta-glucosidase from *Streptomyces* sp: catalytic residues and kinetic studies. **Biochemistry.** 40: 5975-5982.
- van Aalten, D. M., Komander, D., Synstad, B., Gaseidnes, S., Peter, M. G., and Eijnsink, V. G. (2001). Structural insights into the catalytic mechanism of a family 18 exo-chitinase. **Proc. Natl. Acad. Sci. USA.** 98: 8979-8984.
- van Aalten, D. M., Synstad, B., Brurberg, M. B., Hough, E., Riise, B. W., Eijnsink, V. G., and Wierenga, R. K. (2000). Structure of a two-domain chitotriosidase from *Serratia marcescens* at 1.9-Å resolution. **Proc. Natl. Acad. Sci. USA.** 97: 5842-5847.
- van Vlijmen, H. W., Schaefer, M., and Karplus, M. (1998). Improving the accuracy of protein pK<sub>a</sub> calculations: conformational averaging versus the average structure. **Proteins.** 33: 145-158.
- Viladot, J. L., de Ramon, E., Durany, O., and Planas, A. (1998). Probing the mechanism of *Bacillus* 1,3-1,4-β-D-glucan 4-glucanohydrolases by chemical rescue of inactive mutants at catalytically essential residues. **Biochemistry.** 37: 11332-11342.
- Vocadlo, D. J., Mayer, C., He, S., and Withers, S. G. (2000). Mechanism of action and identification of Asp242 as the catalytic nucleophile of *Vibrio furnisii* N-acetyl-β-D-glucosaminidase using 2-acetamido-2-deoxy-5-fluoro-α-L-idopyranosyl fluoride. **Biochemistry.** 39: 117-126.

- Vocadlo, D. J. and Withers, S. G. (2005). Detailed comparative analysis of the catalytic mechanisms of  $\beta$ -*N*-acetylglucosaminidases from families 3 and 20 of glycoside hydrolases. **Biochemistry**. 44: 12809-12818.
- Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. **J. Mol. Graph.** 8: 52-56, 29.
- Watanabe, T., Suzuki, K., Oyanagi, W., Ohnishi, K., and Tanaka, H. (1990). Gene cloning of chitinase A1 from *Bacillus circulans* WL-12 revealed its evolutionary relationship to Serratia chitinase and to the type III homology units of fibronectin. **J. Biol. Chem.** 265: 15659-15665.
- Williams, S. J., Mark, B. L., Vocadlo, D. J., James, M. N. G., and Withers, S. G. (2002). Aspartate 313 in the *Streptomyces plicatus* hexosaminidase plays a critical role in substrate-assisted catalysis by orienting the 2-acetamido group and stabilizing the transition state. **J. Biol. Chem.** 277: 40055-40065.
- Xu, G. Y., Ong, E., Gilkes, N. R., Kilburn, D. G., Muhandiram, D. R., Harris-Brandts, M., Carver, J. P., Kay, L. E., and Harvey, T. S. (1995). Solution structure of a cellulose-binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy. **Biochemistry**. 34: 6993-7009.
- Yang, A. S., Gunner, M. R., Sampogna, R., Sharp, K., and Honig, B. (1993). On the calculation of  $pK_{as}$  in proteins. **Proteins**. 15: 252-265.
- Yu, C., Lee, A. M., Bassler, B. L., and Roseman, S. (1991). Chitin utilization by marine bacteria. A physiological function for bacterial adhesion to immobilized carbohydrates. **J. Biol. Chem.** 266: 24260-24267.
- Zechel, D. L. and Withers, S. G. (2001). Dissection of nucleophilic and acid-base catalysis in glycosidases. **Curr. Opin. Chem. Biol.** 5: 643-649.

- Zees, A. C., Pyrpassopoulos, S., and Vorgias, C. E. (2009). Insights into the role of the ( $\alpha+\beta$ ) insertion in the TIM-barrel catalytic domain, regarding the stability and the enzymatic activity of chitinase A from *Serratia marcescens*. **Biochim. Biophys. Acta.** 1794: 23-31.
- Zhu, Z., Zheng, T., Homer, R. J., Kim, Y. K., Chen, N. Y., Cohn, L., Hamid, Q., and Elias, J. A. (2004). Acidic mammalian chitinase in asthmatic Th2 inflammation and IL-13 pathway activation. **Science.** 304: 678-1682.
- Zobell, C. E. and Rittenberg, S. C. (1938). The Occurrence and characteristics of chitinoclastic bacteria in the sea. **J. Bacteriol.** 35: 275-287.



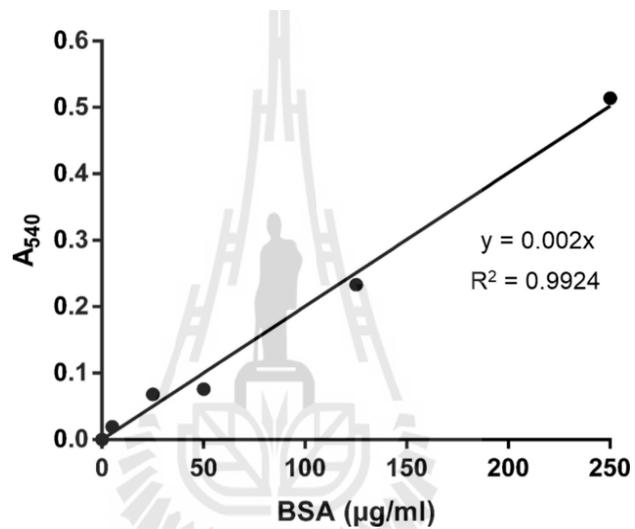


**APPENDICES**

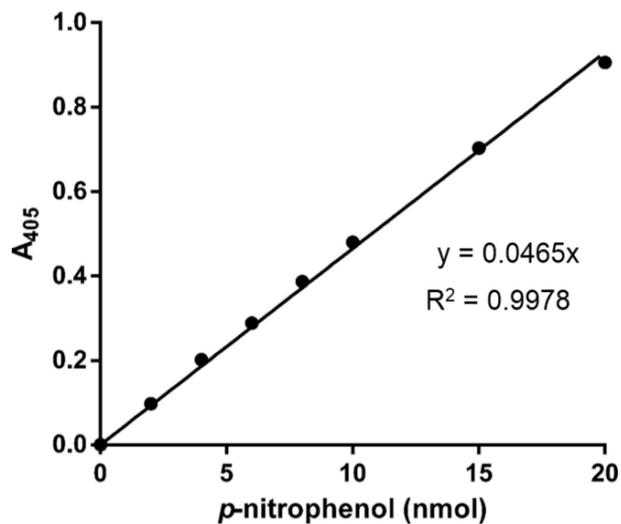
# APPENDIX A

## STANDARD CURVES

### 1. Stand curve of BSA by BCA method



### 2. Standard curve of *p*-nitrophenol



**APPENDIX B**  
**COMMON p*K*<sub>a</sub> VALUE**

**Table S1 Model p*K*<sub>a</sub> values used in this study**

Residue type	Model p <i>K</i> <sub>a</sub> value
Arginine	13.0
Aspartic acid	4.0
Cysteine	8.7
C-terminus	3.8
Glutamic acid	4.4
Histidine	6.3
Lysine	10.4
N-terminus	8.0
Tyrosine	9.6

## APPENDIX C

### PUBLICATIONS

#### Publication outputs:

**Meekrathok**, P., Bürger, M., Porfetye, A. T., Vetter, I. R. and Suginta, W. (2015).

Expression, purification, crystallization and preliminary crystallographic analysis of a GH20  $\beta$ -N-acetylglucosaminidase from the marine bacterium *Vibrio harveyi*. **Acta Crystallogr. Sect. F-Struct. Biol. Commun.** 71: 427-433.

**Meekrathok**, P., and Suginta, W. (2016) Probing the catalytic mechanism of *Vibrio harveyi* GH20  $\beta$ -N-acetylglucosaminidase by chemical rescue. **PLoS One.** 11(2): e0149228.

#### In preparations:

**Meekrathok**, P., Porfetye, A. T., Bürger, M., Vetter, I. R. and Suginta, W. (2016)

Structural and functional analysis of a novel GH20  $\beta$ -N-acetylglucosaminidase (GlcNAcase) from the marine bacterium *Vibrio harveyi* (In preparations).

**Meekrathok**, P., Kukic, P., Nielsen, J. E., and Suginta, W. (2016) Understanding the concerted roles of three acidic residues in the DxDxE motif of GH18 chitinases through pK<sub>a</sub> calculations (In preparations).

Lowhalidanon, K., **Meekrathok**, P., Thongsom, S., Suginta, W., and Khunkaewla, P. (2016). Human acidic mammalian chitinase (AMCase): molecular cloning, protein expression, and production of monoclonal antibody (In preparations).





## Expression, purification, crystallization and preliminary crystallographic analysis of a GH20 $\beta$ -*N*-acetylglucosaminidase from the marine bacterium *Vibrio harveyi*

Piyanat Meekrathok,<sup>a</sup> Marco Bürger,<sup>b</sup> Arthur T. Porfetye,<sup>b</sup> Ingrid R. Vetter<sup>b\*</sup> and Wipa Suginta<sup>a\*</sup>

Received 11 January 2015  
Accepted 27 February 2015

Edited by T. Bergfors, Uppsala University, Sweden

**Keywords:**  $\beta$ -*N*-acetylglucosaminidase; *Vibrio harveyi*; GH20 glycoside hydrolase family.

**Supporting information:** this article has supporting information at journals.iucr.org/f

<sup>a</sup>Biochemistry–Electrochemistry Research Unit, School of Biochemistry, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand, and <sup>b</sup>Max Planck Institute of Molecular Physiology, 44227 Dortmund, Germany. \*Correspondence e-mail: ingrid.vetter@mpi-dortmund.mpg.de, wipa@sut.ac.th

*Vibrio harveyi*  $\beta$ -*N*-acetylglucosaminidase (*Vh*GlcNAcase) is a new member of the GH20 glycoside hydrolase family responsible for the complete degradation of chitin fragments, with *N*-acetylglucosamine (GlcNAc) monomers as the final products. In this study, the crystallization and preliminary crystallographic data of wild-type *Vh*GlcNAcase and its catalytically inactive mutant D437A in the absence and the presence of substrate are reported. Crystals of wild-type *Vh*GlcNAcase were grown in 0.1 *M* sodium acetate pH 4.6, 1.4 *M* sodium malonate, while crystals of the D437A mutant were obtained in 0.1 *M* bis-tris pH 7.5, 0.1 *M* sodium acetate, 20% PEG 3350. X-ray data from the wild-type and the mutant crystals were collected at a synchrotron-radiation light source and were complete to a resolution of 2.5 Å. All crystals were composed of the same type of dimer, with the substrate *N,N'*-diacetylglucosamine (GlcNAc<sub>2</sub> or diNAG) used for soaking was cleaved by the active enzyme, leaving only a single GlcNAc molecule bound to the protein.

### 1. Introduction

Chitin is a  $\beta$ -1,4-linked *N*-acetylglucosamine (GlcNAc) polymer and is one of the most abundant biopolymers in nature. Chitin serves as a major component of fungal cell walls and also of the exoskeletons of crustaceans and insects and the microfilaria sheaths of parasitic nematodes, molluscs, protozoa and other marine organisms (Armand *et al.*, 1994; Yuli *et al.*, 2004). In marine ecosystems, chitin biomaterials are initially hydrolyzed to short-chain oligosaccharides by chitinases, hydrolytic enzymes that are secreted mainly by marine bacteria such as *Vibrio* sp. (Ohishi *et al.*, 1996; Suginta *et al.*, 2000; Kadokura *et al.*, 2007). *V. harveyi* is a Gram-negative marine bacterium that has been reported to express high levels of chitinases and GlcNAcases in order to efficiently utilize chitin biomaterials as its sole source of energy (Suginta *et al.*, 2000). The bacterium is both a primary and an opportunistic pathogen of marine animals, triggering a lethal disease called luminous vibriosis (Austin & Zhang, 2006), which affects marine fish and prawn farming operations worldwide (Liuxy *et al.*, 1996; Actis *et al.*, 2011). In the chitin-degradation pathway of *V. harveyi*, chitinases initially degrade chitin to small chitooligosaccharide fragments, which are further transported across the bacterial cell wall through the chitooligosaccharide-uptake channel, which is known as chitoporin (Keyhani *et al.*, 2000; Li & Roseman, 2004; Suginta, Chumjan,



© 2015 International Union of Crystallography

## research communications

Mahendran, Janning *et al.*, 2013; Suginta, Chumjan, Mahendran, Schulte *et al.*, 2013). In the periplasm, GlcNAcases (EC 3.2.1.52) degrade chitin oligosaccharides to GlcNAc monomers, which are transported across the inner membrane by a specific ABC transporter into the cytoplasm and are then further metabolized to metabolic intermediates that can readily be converted to carbon and nitrogen sources for cells. In bacteria, GlcNAcases act as biocatalysts during the natural recycling of chitin biomaterials. However, the human homologues of bacterial GlcNAcases, known as HexA and HexB, are clinically important since they are critically involved in the degradation of glycosphingolipids, which are deposited in the form of gangliosides GM1, GM2 and GM3 on the plasma membranes of nerve cells. Mutations of these enzymes lead to the accumulation of GM2 gangliosides, which results in fatal diseases such as Tay–Sachs disease and Sandhoff disease (Myerowitz, 1997).

In the Carbohydrate-Active Enzymes (CAZy) database (<http://www.cazy.org/>), GlcNAcases are classified as glycosidic hydrolases of family 3 (GH3) or family 20 (GH20), which differ from each other in their sequences and their mechanisms of enzyme action (Henrissat & Bairoch, 1993; Henrissat & Davies, 1997; Lombard *et al.*, 2014). GH3 GlcNAcases catalyze the release of terminal 2-acetamido-2-deoxy- $\beta$ -D-glycopyranoside from various glycoconjugates (Drouillard *et al.*, 1997) using the covalent  $\alpha$ -glycosidic enzyme intermediate mechanism (Vocadlo & Withers, 2005). In contrast, GH20 GlcNAcases typically hydrolyze chitin fragments employing a substrate-assisted mechanism, in which the carbonyl O atom of the C<sub>2</sub> N-acetyl moiety takes part in cleavage of the scissile bond, yielding an oxazolinium ion as the reaction intermediate (Mark *et al.*, 2001, 2003; Lemieux *et al.*, 2006; Langley *et al.*, 2008).

In terms of structural studies, the crystal structure of *Serratia marcescens* chitobiase (*SmChiB*) was the first GH20 GlcNAcase structure to be solved (in a complex with diNAG; Tews *et al.*, 1996; Prag *et al.*, 2000). To date, a few other GH20 structures have been reported, including those of *Streptomyces plicatus*  $\beta$ -Hex (Williams *et al.*, 2002), human HexB (Mark *et al.*, 2003; Maier *et al.*, 2003), human HexA (Lemieux *et al.*, 2006), *Streptococcus gordonii* GenA (Langley *et al.*, 2008), *Paenibacillus* sp. TS12  $\beta$ -Hex (Sumida *et al.*, 2009) and *S. pneumoniae* R6  $\beta$ -Hex (Jiang *et al.*, 2011). All GH20 hexosaminidases have a structural feature in common: a central catalytic ( $\beta/\alpha$ )<sub>8</sub> TIM-barrel domain with a substrate-binding pocket located within the ( $\beta/\alpha$ )<sub>8</sub> TIM barrel.

In a previous study, we isolated two GH20 GlcNAcase genes encoding two GlcNAcase isoforms, *VhNag1* and *VhNag2*, from the genome of *V. harveyi* strain 650 (Suginta *et al.*, 2010). Both genes were cloned in the pQE60 expression vector and were expressed at high levels in *Escherichia coli* M15 (pREP) host cells. *VhNag1* was inactive and had a molecular mass of about 93 kDa. *VhNag2* had a molecular mass of 75 kDa and was catalytically active, with a pH optimum of 7.0. *VhNag2* is an exolytic enzyme that sequentially hydrolyzes chitin oligosaccharides, releasing GlcNAc as the end product. Kinetic studies of the hydrolysis of the

natural glycosides showed that chitotetraose was the best substrate of *VhNag2* (Suginta *et al.*, 2010). In the present study, we report the expression, purification and crystallization trials of *VhNag2* (hereafter referred to as *VhGlcNAcase*) in the absence and presence of the substrate *N,N'*-diacetylglucosamine (GlcNAc<sub>2</sub> or diNAG). Based on a literature review and multiple sequence alignment, the Asp437–Glu438 residues are equivalent to the catalytic pair Asp539–Glu540 in *SmChiB* (Tews *et al.*, 1996) and Asp313–Glu314 in *S. plicatus*  $\beta$ -Hex (Williams *et al.*, 2002). Therefore, it is presumed that both residues are important in substrate-assisted catalysis and that Asp437 may play the same role as Asp539 in *SmChiB* and Asp313 in  $\beta$ -Hex. Accordingly, we mutated this residue and the corresponding mutant D437A was expressed, purified and crystallized in an attempt to obtain an inactive protein that might still interact with the substrate diNAG, in order to obtain detailed information on the protein–sugar interactions.

## 2. Materials and methods

### 2.1. Cloning, expression and purification

The full-length GlcNAcase cDNA (accession No. HM175716) was isolated from the genome of *V. harveyi* by the PCR technique, and the gene encoding *VhGlcNAcase* (amino acids 5–642, lacking the signal peptide) was then cloned into the pQE60 expression vector (Qiagen, Valencia, California, USA), generating a four-amino-acid (MGGS) cloning artefact at the N-terminus as described previously (Suginta *et al.*, 2010). The recombinant protein was shown to be expressed at a high level in *E. coli* M15 (pREP4) cells as a C-terminally His<sub>6</sub>-tagged polypeptide (Suginta *et al.*, 2010). The D437A mutant of the GlcNAcase was generated in this study using the QuikChange Site-Directed Mutagenesis Kit (Stratagene, California, USA) according to the manufacturer's protocols, using the full-length GlcNAcase cDNA as template. The mutagenic forward and reverse primers were 5'-GTTCA-CATTGGCGCGCGGAAGTGCCTGTGCCTAACGGC-3' and 5'-GCCGTTAGGCACTTCCGCGCGCCAATGTGAC-3', respectively. The underlined sequences represent the mutated codon. After gene amplification, the mutated DNA was treated with DpnI and then transformed into *E. coli* XL1-Blue competent cells. The mutated plasmids were extracted using QuickClean II Plasmid Miniprep Kits (GenScript, Piscataway, New Jersey, USA) and the mutated sequence was verified by automated sequencing (First Base Laboratories Sdn Bhd, Selangor Darul Ehsan, Malaysia).

In this study, the expression and purification of recombinant wild-type *VhGlcNAcase* and the D437A mutant were optimized in order to improve the yield and the purity of the proteins for crystallization purposes. Cells harbouring the recombinant pQE60/*VhGlcNAcase* plasmid were grown at 310 K in Terrific Broth (TB) medium containing 100  $\mu$ g ml<sup>-1</sup> ampicillin and 25  $\mu$ g ml<sup>-1</sup> kanamycin until the OD<sub>600</sub> of the cell culture reached 0.6. GlcNAcase expression was induced by the addition of isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.4 mM and cell growth continued at

293 K for 18 h. The cell pellet was then collected by centrifugation and resuspended in lysis buffer [20 mM Tris-HCl pH 8.0 containing 150 mM NaCl, 1 mM phenylmethylsulfonyl fluoride (PMSF), 5% (v/v) glycerol and 1 mg ml<sup>-1</sup> lysozyme]. The resuspended cells were then lysed on ice using a Sonopuls ultrasonic homogenizer with a 6 mm diameter probe (50% duty cycle; amplitude setting 30%; total time 30 s; 6–8 pulses).

For purification, freshly prepared crude lysate obtained after homogenization was centrifuged at 12 000g for 1 h at 277 K and then applied onto a TALON metal-affinity resin (1.0 × 10 cm; Clontech, USA). Affinity chromatography was carried out at 277 K under gravity-dependent flow following the manufacturer's instructions (<http://www.clontech.com>). The column was washed thoroughly with equilibration buffer (20 mM Tris-HCl buffer pH 8.0 containing 150 mM NaCl) followed by 10 mM imidazole in equilibration buffer; GlcNAcase was eluted with 250 mM imidazole in equilibration buffer and three eluted fractions (10 ml each) were collected. After the purity of the proteins had been analyzed by SDS-PAGE, the GlcNAcase-containing fractions were pooled, concentrated using a Vivaspin 20 ultrafiltration membrane concentrator (10 kDa molecular-weight cutoff; Vivascience AG, Hannover, Germany) and then further purified on a HiPrep 16/60 Sephacryl S-200 prepacked column connected to an ÄKTAprime purification system (Amersham Bioscience, Piscataway, New Jersey, USA). For the wild-type protein, the running buffer was low-salt equilibration buffer (20 mM Tris-HCl buffer pH 8.0 containing 30 mM NaCl and 1 mM TCEP). The wild-type protein was prepared in low-salt equilibration buffer to avoid interference by salt in the crystallization trials. However, some of the wild-type protein appeared as an aggregation peak in the gel-filtration profile. Purification of the D437A mutant at a low salt concentration showed complete aggregation of the protein. Therefore, the solubility of the mutated protein was improved by increasing the salt concentration to 150 mM in the same equilibration buffer as used previously (Suginta *et al.*, 2010). The chromatography column was operated at a flow rate of 0.2 ml min<sup>-1</sup>, and 1.4 ml fractions were collected and assayed for GlcNAcase activity using a colorimetric assay with pNP-GlcNAc as the substrate. The D437A mutant had only 1–2% of the activity of the wild type (data not shown). GlcNAcase-containing fractions were pooled and concentrated using the same type of Vivaspin membrane concentrator. Protein concentrations were initially determined by measuring the absorption at 280 nm (Edelhoch, 1967; Gill & von Hippel, 1989; Pace *et al.*, 1995) and then by the method of Bradford (1976) using a standard calibration curve constructed with BSA. The purity and the molecular mass of the enzyme monomer were determined by SDS-PAGE (Laemmli, 1970). Protein aggregation after purification was evaluated using native PAGE according to a protocol described elsewhere (Arndt *et al.*, 2012).

## 2.2. Protein crystallization

Initial crystallization screening of wild-type *Vh*GlcNAcase was performed at 293 K using commercially available screens

from Qiagen, Hilden, Germany: the sparse-matrix screens the JCSG Core Suites I, II, III and IV, the Classics and Classics II Suites and the PACT Suite and the grid screens the PEGs Suite and the Anions Suite. The screens were set up in 96-well CrystalQuick plates (Greiner Bio-One, Frickenhausen, Germany) using a Mosquito Crystallization robot (TTP Labtech, Melbourn, England) with the sitting-drop method. The purified enzyme was centrifuged at 13 000g for 5 min and 0.1 µl of the freshly prepared wild-type protein at 10 mg ml<sup>-1</sup> in equilibration buffer consisting of 20 mM Tris-HCl pH 8.0, 30 mM NaCl was then pipetted into 0.1 µl of each precipitant in the crystallization screens. Small plate-shaped crystals of wild-type GlcNAcase were observed within 1 d in condition F10 from the PACT Suite [20% (w/v) PEG 3350, 0.1 M bis-tris propane pH 6.5, 0.2 M sodium/potassium phosphate] and small crystals also appeared within 3 d in condition C6 of the Anions Suite (0.1 M sodium acetate pH 4.6, 1.2 M sodium malonate). Crystals obtained from both conditions were further optimized by the hanging-drop vapour-diffusion technique using different concentrations of two precipitants: (i) 17–22% (w/v) PEG 3350, 0.1–0.2 M bis-tris propane pH 6.5, 0.1–0.2 M sodium/potassium phosphate and (ii) 0.1–0.2 M sodium acetate pH 4.6, 0.7–1.8 M sodium malonate. In each drop, 1.5 µl GlcNAcase solution was mixed with 1.5 µl of each precipitant and then equilibrated over 1.0 ml of the respective precipitant in a 24-well Linbro tissue-culture plate. The wild-type crystals in 0.1 M sodium acetate pH 4.6, 1.4 M sodium malonate were transferred into drops consisting of 2 µl cryoprotectant solution (0.1 M sodium acetate pH 4.6, 2.9 M sodium malonate). The substrate complex of GlcNAcase was obtained by soaking the wild-type crystals with the natural substrate (diNAG; 10 mM) in the corresponding mother liquor at 293 K for 5, 10, 15 and 30 min. Afterwards, the soaked crystals were immersed for a few seconds into a cryoprotectant consisting of the mother liquor with 2.9 M sodium malonate and 10 mM diNAG before flash-cooling in liquid nitrogen for further storage.

For crystallization of the GlcNAcase D437A mutant, 0.1 µl protein solution (18 mg ml<sup>-1</sup> in 20 mM Tris-HCl buffer pH 8.0 containing 150 mM NaCl and 1 mM TCEP) was pipetted into 0.1 µl of each precipitant using the same set of crystal screens as described above. Small plate-shaped crystals (form I) were obtained within 3 d in condition G7 from the PACT Suite [20% (w/v) PEG 3350, 0.1 M bis-tris pH 7.5, 0.2 M sodium acetate], condition D9 from the PEGs Suite [25% (w/v) PEG 6000, 0.1 M Tris-HCl pH 8.5] and condition H1 from the PEGs Suite [20% (w/v) PEG 3350, 0.2 M potassium sodium tartrate]. Three-dimensional rod-shaped crystals (form II) were also obtained within 14 d of incubation in condition C12 from the Anions Suite (0.1 M MES pH 6.5, 1.2 M sodium malonate) at 293 K. The D437A crystals from condition G7 of the PACT Suite were further optimized by the hanging-drop vapour-diffusion method in a 24-well Linbro tissue-culture plate as described for the wild-type enzyme. A 1.5 µl droplet of the D437A mutant was mixed with 1.5 µl of precipitants at various concentrations [17–22% (w/v) PEG 3350, 0.1 M bis-tris pH 7.5, 0.1–0.2 M sodium acetate] and then equilibrated over 1.0 ml of

## research communications

Table 1  
Data-collection and processing statistics.  
Values in parentheses are for the outer shell.

	Wild type	Wild type, GlcNAc complex	D437A mutant (form I)	D437A mutant (form II)
Space group	$P2_1$	$P2_1$	$P2_1$	$P4_22_2$
Unit-cell parameters				
$a, b, c$ (Å)	90.2, 130.7, 98.5	91.3, 129.6, 100.0	89.4, 129.3, 98.4	165.1, 165.1, 155.2
$\alpha, \beta, \gamma$ (°)	90.0, 113.0, 90.0	90.0, 114.4, 90.0	90.0, 112.2, 90.0	90.0, 90.0, 90.0
Resolution range (Å)	48.55–2.43 (2.43–2.37)	46.21–2.50 (2.56–2.50)	48.43–2.60 (2.67–2.60)	49.50–3.00 (3.08–3.00)
$V_{SI}$ (Å <sup>3</sup> Da <sup>-1</sup> )	3.59	3.62	3.53	3.55
Solvent content (%)	65.73	66.03	65.22	65.39
Subunits per asymmetric unit	2	2	2	2
Total No. of reflections	577490	342683	436477	569579
No. of unique reflections	84783	72950	63617	43294
Wavelength (Å)	0.99980	0.97889	0.97889	0.97779
Data completeness (%)	99.4 (99.1)	99.4 (99.9)	99.8 (99.7)	99.5 (99.5)
Multiplicity	6.6 (7.1)	4.7 (4.5)	6.9 (7.1)	13.2 (13.9)
Mean $I(\sigma(I))$	9.93 (2.21)	9.71 (3.17)	9.7 (2.18)	12.8 (2.06)
$R_{\text{merge}}^{\dagger}$ (%)	14.6 (108.7)	11.8 (53.3)	12.5 (77.5)	21.0 (167.1)
$CC_{1/2}$	99.4 (84.1)	99.2 (85.1)	99.3 (86.5)	99.6 (79.0)

$\dagger R_{\text{merge}} = \sum_{\text{obs}} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{\text{obs}} \sum_i I_i(hkl)$ , where  $I_i(hkl)$  is the intensity for the  $i$ th measurement of an equivalent reflection with indices  $hkl$ .

the respective precipitant. After the D437A crystals had been immersed in a cryoprotectant solution consisting of mother liquor supplemented with 25%(v/v) glycerol, they were rapidly transferred to liquid nitrogen and stored.

### 2.3. Data collection and processing

For in-house crystal testing and the collection of preliminary data sets, crystals were mounted in nylon loops (Hampton Research, Aliso Viejo, California, USA) on a goniometer and data were collected on a Bruker MICROSTAR or Rigaku MicroMax-007 HF rotating-anode generator with a copper anode as the X-ray source. X-ray diffraction data were collected from high-quality crystals using a Pilatus 6M detector on beamline PX-II at the Swiss Light Source (SLS), Paul Scherrer Institute, Villigen, Switzerland. All X-ray images were recorded using 0.25 Å oscillations at 100 K and a wavelength of 1.0 Å with crystal-to-detector distances of 381, 350, 440 and 500 mm for the wild type, the wild type complexed with GlcNAc, the D437A mutant form I and the D437A mutant form II, respectively. Data indexing was carried out using *iMosflm* (Battye *et al.*, 2011) from the *CCP4* suite (Winn *et al.*, 2011). The data were further processed using *XDS* (Kabsch, 2010). The protein content of the asymmetric unit was estimated by calculating the Matthews coefficient (Matthews, 1968), while the solvent content was calculated based on two subunits of protein, each with six attached histidine residues.

### 3. Results and discussion

The gene encoding *VhGlcNAcase* was cloned in the pQE60 expression vector and the D437A mutant was generated using the full-length *VhGlcNAcase* cDNA as a template as described above. Both wild-type and mutated *GlcNAcases* were expressed in *E. coli* M15 (pREP4) cells as C-terminally

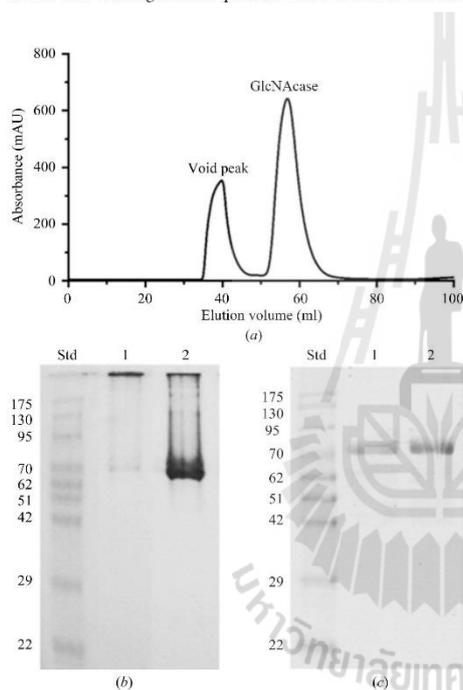
His<sub>6</sub>-tagged polypeptides that could be purified by affinity chromatography. Both *GlcNAcases*, the wild type and the D437A mutant, were purified to homogeneity by a two-step protocol involving TALON metal resin affinity chromatography followed by HiPrep (16/60) Sephacryl S-200 HR gel-filtration chromatography. The final yield of the purified enzymes was approximately 20 mg protein per litre of bacterial culture. Fig. 1(a) shows a chromatographic profile of the elution of the wild-type protein from a HiPrep 16/60 Sephacryl S-200 prepacked column. The enzyme was found in two peaks. The first peak, which eluted at the void volume, was shown by native PAGE to be aggregated protein with a molecular mass that was too large to enter the gel (Fig. 1b, lane 1). However the second peak, which eluted at 45–70 ml, was non-aggregated protein that migrated with an apparent molecular mass of 75 kDa, corresponding to the molecular mass of the *GlcNAcase* monomer (Fig. 1b, lane 2). The pooled fraction of each protein peak was found to migrate similarly under denaturing conditions on SDS-PAGE (Fig. 1c), indicating that they were likely to contain the same protein. When the *GlcNAcase* activity was assayed with *pNP-GlcNAc* as the substrate, the second peak showed high activity, while no *GlcNAcase* activity was detected for the void peak. The chromatographic profile of the D437A mutant was similar to that of the wild-type enzyme, but the void peak was not seen, since the higher salt concentration helped to improve the protein solubility (data not shown).

Crystallization trials with the sitting-drop vapour-diffusion method yielded small crystals from various conditions and initial screening of the crystal quality with an in-house X-ray diffractometer showed that most conditions did not give X-ray-quality crystals, with the exception of those obtained from condition C6 of The Anions Suite (0.1 M sodium acetate pH 4.6, 1.2 M sodium malonate), which diffracted to 2.4 Å resolution. Therefore, crystals of wild-type *GlcNAcase* grown under this condition were further optimized. Small single

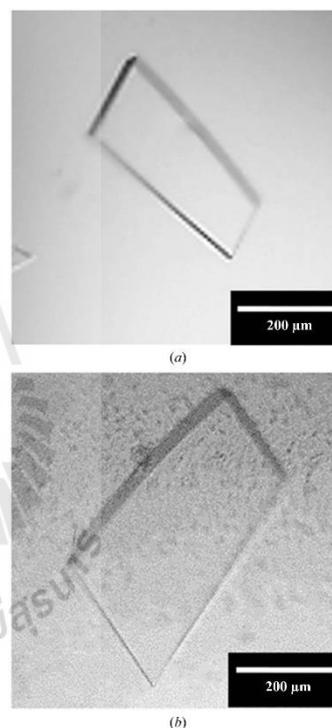
crystals of wild-type GlcNAcase were finally observed from 0.1 M sodium acetate pH 4.6, 1.3–1.6 M sodium malonate. After optimization, large single crystals of wild-type GlcNAcase appeared as thick three-dimensional plates with approximate dimensions of  $400 \times 200 \times 20 \mu\text{m}$  (Fig. 2a). In the case of the crystal complex of wild-type *Vh*GlcNAcase, soaking the wild-type crystal with the diNAG substrate for a short time (5 min) was attempted but without success. We also tried soaking the crystals of the native enzyme with the natural substrate diNAG for various times, but we could not obtain X-ray diffraction data for the GlcNAcase–diNAG complex because the substrate was degraded by the active enzyme. We eventually obtained X-ray diffraction data for the native enzyme complexed with a single GlcNAc molecule. Fig. 3 shows that the degradation product GlcNAc fitted with full

occupancy into the well defined electron density found in the active site of *Vh*GlcNAcase. In the crystallization of the D437A mutant, small plate-shaped crystals (form I) were obtained during optimization of condition G7 from the PACT screen with a reservoir solution consisting of 20% (w/v) PEG 3350, 0.1 M bis-tris pH 7.5, 0.2 M sodium acetate. After further optimization, the D437A mutant crystals appeared as thick three-dimensional plates in a condition consisting of 20% (w/v) PEG 3350, 0.1 M bis-tris pH 7.5, 0.1 M sodium acetate. The D437A crystals grew to approximate final dimensions of  $500 \times 300 \times 50 \mu\text{m}$  (Fig. 2b) in a space group similar to that of the wild-type crystals.

The final statistics for X-ray data-collection and processing of the GlcNAcase crystal variants are shown in Table 1. The refined unit-cell parameters of the wild-type crystal are



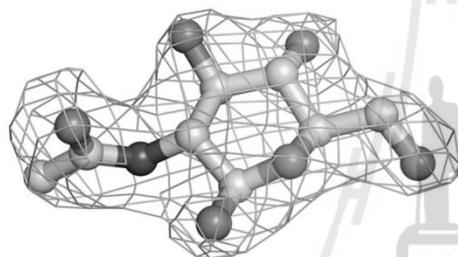
**Figure 1**  
Purification of *Vh*GlcNAcase expressed from *E. coli* M15 cells using a HiPrep 16/60 Sephacryl S-200 prepacked gel-filtration column. Chromatography was performed under a low salt concentration, as described in the text. (a) A chromatographic elution profile of wild-type *Vh*GlcNAcase obtained from an AKTApurifier system. (b) GlcNAcase-containing fractions obtained from the two peaks were pooled separately and then loaded onto native PAGE followed by Coomassie Blue staining. (c) The same protein fractions were analysed by SDS-PAGE. Lane Std, low-molecular-weight protein markers (labelled in kDa); lane 1, pooled fractions from the void peak; lane 2, pooled fractions from the second peak.



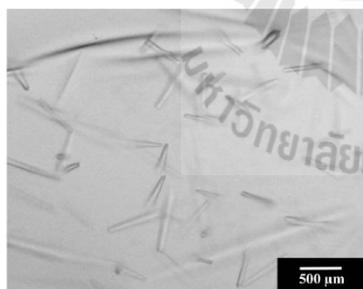
**Figure 2**  
(a) A crystal of wild-type *Vh*GlcNAcase, with dimensions of  $400 \times 200 \times 20 \mu\text{m}$ , obtained from a hanging-drop vapour-diffusion setup using 0.1 M sodium acetate pH 4.6 containing 1.4 M sodium malonate. (b) A crystal of the D437A mutant (dimensions of  $500 \times 300 \times 50 \mu\text{m}$ ) obtained from a hanging-drop vapour-diffusion setup using 20% (w/v) PEG 3350, 0.1 M bis-tris pH 7.5, 0.1 M sodium acetate. Both crystals were obtained within 3 d of incubation at 293 K.

## research communications

$a = 90.2$ ,  $b = 130.7$ ,  $c = 98.5$  Å,  $\beta = 113^\circ$ . The crystal contains two molecules per asymmetric unit, with an estimated Matthews coefficient ( $V_M$ ) of  $3.59$  Å<sup>3</sup> Da<sup>-1</sup>. The indexing statistics were compatible with the monoclinic space groups  $P2$  or  $P2_1$ . Molecular replacement yielded an unambiguous solution in space group  $P2_1$ . For the wild-type crystal in complex with GlcNAc, the unit-cell parameters are  $a = 91.3$ ,  $b = 129.6$ ,  $c = 100.00$  Å,  $\beta = 114.4^\circ$ , which are very similar to those of the unliganded wild-type crystal in the same space group. For the D437A mutant crystal, the refined unit-cell parameters are again very similar:  $a = 89.4$ ,  $b = 129.3$ ,  $c = 98.4$  Å,  $\beta = 112.2^\circ$ . D437A mutant crystals also grew as rod-shaped crystals (form II) belonging to the tetragonal space group  $P4_32_12$  (Fig. 4), with unit-cell parameters  $a = 165.1$ ,  $b = 165.1$ ,  $c = 155.2$  Å and again with two molecules per asymmetric unit. Since the preliminary structural model of the D437A mutant from the tetragonal crystal form contained exactly the same dimer as the monoclinic crystal form (r.m.s.d. of 0.82 Å), but with much weaker diffraction of the crystals, this crystal form was not further investigated. Fig. 5 shows



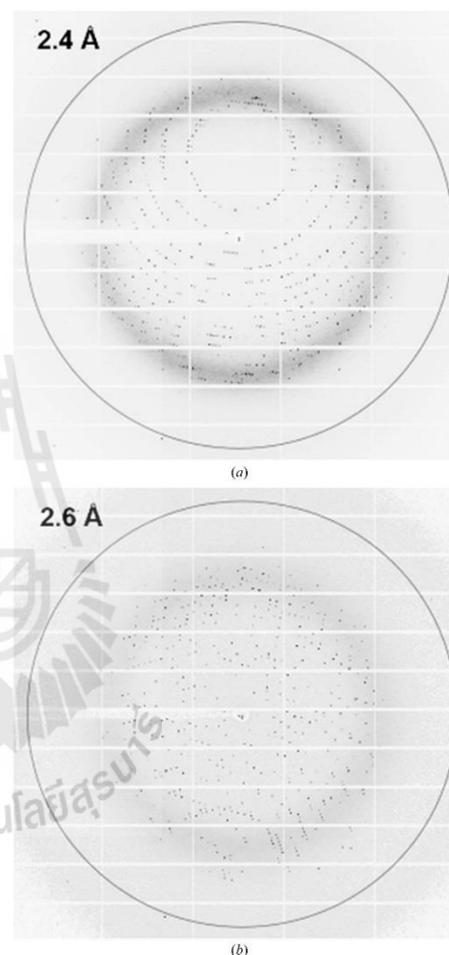
**Figure 3**  
The electron-density map fitted with a GlcNAc molecule after the wild-type *VhGlcNAcase* crystals had been soaked with the substrate diNAG for 30 min. The structure is presented at 2.5 Å resolution and is contoured at the  $1\sigma$  level.



**Figure 4**  
Crystals of the D437A mutant belonging to the tetragonal space group  $P4_32_12$ . Crystals grew as thin rod-shaped crystals within 14 d of incubation in condition C12 from The Anions Suite (0.1 M MES pH 6.5, 1.2 M sodium malonate) at 293 K.

representative diffraction images for both wild-type and monoclinic D437A mutant crystals, which diffracted to 2.4 and 2.6 Å, respectively.

For molecular-replacement (MR) calculations, *Phaser* (McCoy *et al.*, 2007) from the *CCP4* package was employed to obtain an initial solution of the structure of *VhGlcNAcase*. The crystal structure of  $\beta$ -hexosaminidase from *Arthrobacter*



**Figure 5**  
X-ray diffraction images of (a) wild-type *VhGlcNAcase* in the absence of diNAG and (b) the inactive D437A mutant with resolutions of 2.4 and 2.6 Å, respectively. The X-ray data were collected on beamline PX-II at the Swiss Light Source, Villigen, Switzerland.

*aurescens* (PDB entry 3rcn; 35% identity to GlcNAcase from *V. harveyi*; Midwest Center for Structural Genomics, unpublished work), which lacks the N-terminal carbohydrate-binding domain, was used as a search model. After ten cycles of refinement, the  $R_{\text{free}}$  values for the wild type, the wild type in complex with GlcNAc, the D437A mutant (form I) and the D437A mutant (form II) were 25.8, 24.2, 25.0 and 23.4%, respectively.

The preliminary crystallographic data suggested the presence of two protein molecules in the asymmetric unit, with a hydrophilic interface area between two molecules contributed mainly by the N-terminal carbohydrate-binding and catalytic domains, with a surface area of about 1800 Å<sup>2</sup>. PISA analysis (Krissinel & Henrick, 2007) predicted no dimer formation in solution through this interface, which is consistent with the molecular mass of the active enzyme (~75 kDa) obtained from size-exclusion chromatography (Fig. 1). Initial structural refinements showed a well defined electron-density map corresponding to a single GlcNAc molecule in the active site of each wild-type enzyme molecule, indicating that the substrate diNAG was hydrolyzed during the soaking process. We failed to obtain the inactive mutant D437A complexed with diNAG either by both soaking or by co-crystallization, probably because substitution of the Asp437 residue with Ala affects not only the catalytic activity but also the affinity of the enzyme for its substrate. Further refinement to obtain final models of the three GlcNAcase structures is in progress.

#### Acknowledgements

We thank the beamline staff of beamline PXII at the SLS synchrotron at the Paul Scherrer Institute, Villigen, Switzerland and our colleagues Toni Meinhart, Juliane Benz and Ilme Schlichting (Max Planck Institute for Medical Research, Heidelberg) and Eckhard Hofmann and Falk Syberg (Ruhr-Universität Bochum) for help with data collection. This research was financially supported by the Thailand Research Fund (TRF) through the Royal Golden Jubilee PhD Program (Grant No. PHD/0084/2550) and the German Academic Exchange Service (DAAD; Code No. A/10/98020) to P.M. WS was also funded by TRF Basic Research Grant (Grant No. BRG578001).

#### References

- Actis, L. A., Tolmasky, M. E. & Crosa, J. H. (2011). *Fish Diseases and Disorders*, edited by P. T. K. Woo & D. W. Bruno, pp. 570–605. Wallingford: CABI.
- Armand, S., Tomita, H., Heyraud, A., Gey, C., Watanabe, T. & Henrissat, B. (1994). *FEBS Lett.* **343**, 177–180.
- Arndt, C., Koristka, S., Bartsch, H. & Bachmann, M. (2012). *Methods Mol. Biol.* **869**, 49–53.
- Austin, B. & Zhang, X.-H. (2006). *Lett. Appl. Microbiol.* **43**, 119–124.
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
- Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
- Drouillard, S., Armand, S., Davies, G. J., Vorgias, C. E. & Henrissat, B. (1997). *Biochem. J.* **328**, 945–949.
- Edelhoch, H. (1967). *Biochemistry*, **6**, 1948–1954.
- Gill, S. C. & von Hippel, P. H. (1989). *Anal. Biochem.* **182**, 319–326.
- Henrissat, B. & Bairoch, A. (1993). *Biochem. J.* **293**, 781–788.
- Henrissat, B. & Davies, G. (1997). *Curr. Opin. Struct. Biol.* **7**, 637–644.
- Jiang, Y.-L., Yu, W.-L., Zhang, J.-W., Frolet, C., Di Guilmi, A. M., Zhou, C.-Z., Vernet, T. & Chen, Y. (2011). *J. Biol. Chem.* **286**, 43004–43012.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kadokura, K., Rokutani, A., Yamamoto, M., Ikegami, T., Sugita, H., Itoi, S., Hakamata, W., Oku, T. & Nishio, T. (2007). *Appl. Microbiol. Biotechnol.* **75**, 357–365.
- Keyhani, N. O., Li, X.-B. & Roseman, S. (2000). *J. Biol. Chem.* **275**, 33068–33076.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.
- Langley, D. B., Harty, D. W., Jacques, N. A., Hunter, N., Guss, J. M. & Collyer, C. A. (2008). *J. Mol. Biol.* **377**, 104–116.
- Lemieux, M. J., Mark, B. L., Cherney, M. M., Withers, S. G., Mahuran, D. J. & James, M. N. G. (2006). *J. Mol. Biol.* **359**, 913–929.
- Li, X. & Roseman, S. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 627–631.
- Luxy, P.-C., Lee, K.-K. & Chen, S.-N. (1996). *Lett. Appl. Microbiol.* **22**, 413–416.
- Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M. & Henrissat, B. (2014). *Nucleic Acids Res.* **42**, D490–D495.
- Maier, T., Strater, N., Schuette, C., Klingenstein, R., Sandhoff, K. & Saenger, W. (2003). *J. Mol. Biol.* **328**, 669–681.
- Mark, B. L., Mahuran, D. J., Cherney, M. M., Zhao, D., Knapp, S. & James, M. N. G. (2003). *J. Mol. Biol.* **327**, 1093–1109.
- Mark, B. L., Vocadlo, D. J., Knapp, S., Triggs-Raine, B. L., Withers, S. G. & James, M. N. G. (2001). *J. Biol. Chem.* **276**, 10330–10337.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Myerowitz, R. (1997). *Hum. Mutat.* **9**, 195–208.
- Ohishi, K., Yamagishi, M., Ohta, T., Suzuki, M., Izumida, H., Sano, H., Nishijima, M. & Miwa, T. (1996). *J. Ferment. Bioeng.* **82**, 598–600.
- Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. (1995). *Protein Sci.* **4**, 2411–2423.
- Prag, G., Papanikolaou, Y., Tavlas, G., Vorgias, C. E., Petratos, K. & Oppenheim, A. B. (2000). *J. Mol. Biol.* **300**, 611–617.
- Suginta, W., Chuenark, D., Mizuhara, M. & Fukamizo, T. (2010). *BMC Biochem.* **11**, 40.
- Suginta, W., Chumjan, W., Mahendran, K. R., Janning, P., Schulte, A. & Winterhalter, M. (2013). *PLoS One*, **8**, e55126.
- Suginta, W., Chumjan, W., Mahendran, K. R., Schulte, A. & Winterhalter, M. (2013). *J. Biol. Chem.* **288**, 11038–11046.
- Suginta, W., Robertson, P. A., Austin, B., Fry, S. C. & Fothergill-Gilmore, L. A. (2000). *J. Appl. Microbiol.* **89**, 76–84.
- Sumida, T., Ishii, R., Yanagisawa, T., Yokoyama, S. & Ito, M. (2009). *J. Mol. Biol.* **392**, 87–99.
- Tews, I., Perrakis, A., Oppenheim, A., Dauter, Z., Wilson, K. S. & Vorgias, C. E. (1996). *Nature Struct. Mol. Biol.* **3**, 638–648.
- Vocadlo, D. J. & Withers, S. G. (2005). *Biochemistry*, **44**, 12809–12818.
- Williams, S. J., Mark, B. L., Vocadlo, D. J., James, M. N. & Withers, S. G. (2002). *J. Biol. Chem.* **277**, 40055–40065.
- Winn, M. D. et al. (2011). *Acta Cryst.* **D67**, 235–242.
- Yuli, P. E., Suhartono, M. T., Rukayadi, Y., Hwang, J. K. & Pyun, Y. R. (2004). *Enzyme Microb. Technol.* **35**, 147–153.

## RESEARCH ARTICLE

# Probing the Catalytic Mechanism of *Vibrio harveyi* GH20 $\beta$ -N-Acetylglucosaminidase by Chemical Rescue

Piyanat Meekrathok<sup>1</sup>, Wipa Suginta<sup>1,2\*</sup>

**1** Biochemistry-Electrochemistry Research Unit and School of Biochemistry, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima, 30000, Thailand, **2** Center of Excellence in Advanced Functional Materials, Suranaree University of Technology, Nakhon Ratchasima, 30000, Thailand

\* [wipa@sut.ac.th](mailto:wipa@sut.ac.th)

click for updates

## Abstract

### Background

*Vibrio harveyi* GH20  $\beta$ -N-acetylglucosaminidase (*VhGlcNAcase*) is a chitinolytic enzyme responsible for the successive degradation of chitin fragments to GlcNAc monomers, activating the onset of the chitin catabolic cascade in marine *Vibrios*.

### Methods

Two invariant acidic pairs (Asp303-Asp304 and Asp437-Glu438) of *VhGlcNAcase* were mutated using a site-directed mutagenesis strategy. The effects of these mutations were examined and the catalytic roles of these active-site residues were elucidated using a chemical rescue approach. Enhancement of the enzymic activity of the *VhGlcNAcase* mutants was evaluated by a colorimetric assay using pNP-GlcNAc as substrate.

### Results

Substitution of Asp303, Asp304, Asp437 or Glu438 with Ala/Asn/Gln produced a dramatic loss of the *GlcNAcase* activity. However, the activity of the inactive D437A mutant was recovered in the presence of sodium formate. Our kinetic data suggest that formate ion plays a nucleophilic role by mimicking the  $\beta$ -COO<sup>-</sup> side chain of Asp437, thereby stabilizing the reaction intermediate during both the glycosylation and the deglycosylation steps.

### Conclusions

Chemical rescue of the inactive D437A mutant of *VhGlcNAcase* by an added nucleophile helped to identify Asp437 as the catalytic nucleophile/base, and hence its acidic partner Glu438 as the catalytic proton donor/acceptor.

### General Significance

Identification of the catalytic nucleophile of *VhGlcNAcases* supports the proposal of a substrate-assisted mechanism of GH20 *GlcNAcases*, requiring the catalytic pair Asp437-

## OPEN ACCESS

**Citation:** Meekrathok P, Suginta W (2016) Probing the Catalytic Mechanism of *Vibrio harveyi* GH20  $\beta$ -N-Acetylglucosaminidase by Chemical Rescue. PLoS ONE 11(2): e0149228. doi:10.1371/journal.pone.0149228

**Editor:** Eugene A. Permyakov, Russian Academy of Sciences, Institute for Biological Instrumentation, RUSSIAN FEDERATION

**Received:** November 21, 2015

**Accepted:** January 28, 2016

**Published:** February 12, 2016

**Copyright:** © 2016 Meekrathok, Suginta. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This work was financially supported by Suranaree University of Technology (SUT) and the Office of the Higher Education Commission under the NRU project of Thailand. PM received a Ph.D. Scholarship through the Royal Golden Jubilee Ph.D. Program from the Thailand Research Fund (TRF) and SUT (Grant no. PHD/0084/2550). WS was funded by TRF and SUT through the Basic Research Grant (Grant no. BRG578001). The funders had no

role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

**Abbreviations:** *Vh*GlcNAcase, *Vibrio harveyi*  $\beta$ -N-acetylglucosaminidase; GH, glycoside hydrolase; GlcNAc, N-acetylglucosamine; pNP-GlcNAc, para-nitrophenyl-N-acetylglucosaminide; IPTG, isopropyl  $\beta$ -D-1-thiogalactopyranoside; PMSF, phenylmethylsulfonyl fluoride.

Glu438 for catalysis. The results suggest the mechanistic basis of the participation of  $\beta$ -N-acetylglucosaminidase in the chitin catabolic pathway of marine *Vibrios*.

## Introduction

*Vibrio harveyi* is a bioluminescent marine bacterium that utilizes chitin biomaterials, which are abundantly available in the aquatic environment, as its sole source of energy. The initial step of chitin breakdown by *Vibrios* involves the synergistic action of various chitin-related proteins [1–3]. Lytic polysaccharide monooxygenase, a copper-dependent enzyme, attacks recalcitrant chitin polysaccharides [4,5], while endochitinases hydrolyse long chitin chains to chitin oligosaccharides, which are then transported through the bacterial cell wall by chitoporin or ChiP [6–8]. In the periplasm, these chitin fragments are degraded by *exo*  $\beta$ -N-acetylglucosaminidases (GlcNAcases) and the resultant GlcNAc monomers are transported through the inner membrane by the GlcNAc-PTS transporter and finally metabolized in the cytoplasm, finally acting as sources of carbon and nitrogen [9]. Based on CAZy (Carbohydrate-Active enZymes Database; <http://www.cazy.org>), bacterial GlcNAcases belong to either glycoside hydrolase family 3 (GH3), family 20 (GH20) or family 84 (GH84). In general, GlcNAcases from all families break the  $\beta$ -1,4-glycosidic linkage next to the non-reducing end of GlcNAc-containing oligosaccharides, generating GlcNAc units as the end product [10,11]. However, GH3 GlcNAcases differ in their amino acid sequence identity and mode of action from those in the GH20 and GH84 families [12,13]. GH3 GlcNAcases catalyse the hydrolytic reaction through a standard 'retaining' mechanism involving a covalent glycosyl-enzyme intermediate [14,15], while GH20 and GH84 GlcNAcases hydrolyse chitooligosaccharides through a 'substrate-assisted' mechanism involving the transient formation of an oxazolinium ion intermediate [16–18].

A 'chemical rescue' approach has been employed to identify the catalytic mechanism of several glycoside hydrolases. The effects of exogenous anions, such as azide or formate, can provide direct evidence identifying the catalytic acid/base residues in retaining glycoside hydrolases. Following mutation of the acid-base residue or the nucleophilic residue, hydrolytic activity of the mutants can be rescued by the addition of an exogenous nucleophile, such as azide ion, resulting in the formation of products with the  $\alpha$  or  $\beta$  configuration. An example is a study on *Bacillus* 1,3-1,4- $\beta$ -D-glucan 4-glucanohydrolases [19]. Sodium azide was shown to rescue the glucanase activity, but with a different mechanism, when either the nucleophilic (Glu134) or the catalytic acid/base (Glu138) residues were mutated to Ala. E138A yielded a  $\beta$ -glycosyl azide product, arising from nucleophilic attack of azide on the glycosyl-enzyme intermediate, thus proving the role Glu138 as the catalytic acid-base residue. In contrast, azide reactivated the E134A mutant through a single inverting displacement to give the  $\alpha$ -glycosyl azide product, consistent with Glu134 being the catalytic nucleophile.

In the substrate-assisted mechanism of GH20 enzymes, chemical rescue helps to directly identify the catalytic nucleophile in the catalytic pair (typically the invariant Asp-Glu couple) in the enzyme's active site. The Asp residue normally acts as the catalytic base/nucleophile, while the glutamic acid acts as the catalytic proton donor/acceptor [20,21]. Examples of enzymes studied by use of this approach include *Streptomyces plicatus* GH20 hexosaminidase (SpHex) [21], *Arthrobacter protophormiae* GH85 endo- $\beta$ -N-acetylglucosaminidase (Endo A) [22], *Streptomyces* sp. GH1  $\beta$ -glucosidase [23], *Paenibacillus* sp. TS12 GH3 glucosylceramidase [24], *Cellulomonas fimi* GH10 exoglucanase/xylanase [25], *Bacillus licheniformis* GH16

1,3-1,4- $\beta$ -glucanase [19], *Sulfolobus solfataricus* GH29  $\alpha$ -L-fucosidase [26] and *Geobacillus stearothermophilus* T-6 GH51  $\alpha$ -L-arabinofuranosidase [27]. In the case of GH20 GlcNAcases, rescue of the activity of SpHex from *Streptomyces plicatus* [21] has been demonstrated. SpHex catalyses the hydrolysis of *N*-acetyl- $\beta$ -hexosaminides. Point mutation of Asp313 of SpHex to Ala or Asn (mutants D313A or D313N) almost abolished the enzyme's hydrolytic activity, but the catalytic activity of the mutant D313A was significantly increased with the inclusion of sodium azide in the assay medium.

We previously cloned, expressed and characterized a novel member of the GH20 GlcNAcase family, from the marine bacterium *V. harveyi* (so-called VhGlcNAcase) [9]. Based on amino acid sequence alignment with other GlcNAcases, the catalytic pair of VhGlcNAcase was predicted to be Asp437-Glu438. We have now employed the chemical rescue approach to identify the functional roles of Asp437 as the catalytic nucleophile and Glu438 as the catalytic acidic residue of VhGlcNAcase.

## Materials and Methods

### Bacterial strains and chemicals

*Escherichia coli* type strain DH5 $\alpha$  was used for cloning, subcloning and plasmid preparation. Supercompetent *E. coli* XL1Blue (Stratagene, La Jolla, CA, USA) was the host strain for the production of mutagenized plasmid. *E. coli* strain M15 (pREP) host cells (Qiagen, Valencia, CA, USA) and the recombinant plasmid of pQE 60 vector containing *GlcNAcase* gene fragments were used for high-level expression of recombinant enzyme. Chemicals and reagents used for protein expression, purification and characterization of VhGlcNAcase were of analytical grade unless otherwise stated. A QuickChange Site-Directed Mutagenesis Kit including *Pfu* Turbo DNA polymerase was purchased from Stratagene. Restriction enzymes and DNA modifying enzymes were the products of New England Biolabs, Inc. (Beverly, MA, USA). All other chemicals and reagents were obtained from the following sources: reagents for bacterial media (Scharlau Chemie S.A., Barcelona, Spain); *p*-nitrophenol (*p*NP) and *p*-nitrophenyl-*N*-acetylglucosaminide (*p*NP-GlcNAc) were purchased from Sigma-Aldrich (St. Louis, MO, USA); sodium azide was purchased from LabChem Inc. (Zelienople, PA, USA); sodium nitrate, sodium formate and sodium chloride were purchased from Carlo Erba (Rodano, Milano, Italy).

### Amino acid sequence analysis and homology modeling

The amino acid sequence of the matured VhGlcNAcase was submitted to Swiss-Model (<http://swissmodel.expasy.org/>) for tertiary structure prediction using the crystal structure of *S. marcescens* chitobiase (PDB entry: 1QBA) as a structural template. To obtain detailed information about the enzyme's active site, the modelled structure of VhGlcNAcase was superimposed on the 3D structure of *S. marcescens* chitobiase (SmCHB) docked with diNAG coordinates. The annotated structures were edited and displayed in PyMOL ([www.pymol.org/](http://www.pymol.org/)). The structure-based alignment was generated by aligning the amino acid sequence of VhGlcNAcase with five GH20 GlcNAcases with known 3D-structures, including *S. marcescens* chitobiase, SmCHB (PDB code: 1QBA); *Streptomyces plicatus*  $\beta$ -*N*-acetylhexosaminidase, SpHex (PDB code: 1HP4); *Paenibacillus* sp.  $\beta$ -hexosaminidase, PsHex1T (PDB code: 3GH4); human  $\beta$ -hexosaminidase A ( $\alpha$ -chain), HsHexA (PDB code: 2GJX) and human  $\beta$ -hexosaminidase B ( $\beta$ -chain), HsHexB (PDB code: 1NOU). The amino acid sequence alignment was carried out in ClustalW, and the structure-based alignment was further generated using the program ESPript, v3.0 [28].

### Site-directed mutagenesis

The pQE 60 expression vector harboring the full length *VhGlcNAcase* cDNA [9] was used as DNA template. Site-directed mutagenesis was carried out using the QuickChange Site-Directed Mutagenesis Kit (Stratagene), following the Manufacturer's instruction. The mutagenic primers were synthesized by commercial sources (BioDesign Co., Ltd Bangkok, Thailand and Bio Basic Canada Inc., Ontario, Canada) and the oligonucleotide sequences of these primers are listed in Table 1. Eight single mutants, namely D303A, D303N, D304A, D304N, D437A, D437N, E438A and E438Q, were generated and the success of the designed mutations was verified by automated DNA sequencing (First BASE Laboratories Sdn Bhd, Selangor Darul Ehsan, Malaysia).

### Protein expression and purification

The recombinant wild-type *VhGlcNAcase* was expressed in *E. coli* M15 (pREP) cells as a 652-amino acid polypeptide, including the C-terminal (His)<sub>6</sub> sequence [9]. Expression of all GlcNAcase variants was based on the protocol described previously by Suginta *et al.* [9]. Briefly, the transformed cells were grown at 37°C in Terrific Broth (TB) containing 100 µg mL<sup>-1</sup> ampicillin and 25 µg mL<sup>-1</sup> kanamycin until the cell density reached an OD<sub>600</sub> of 0.6. The cell culture was cooled to 20°C, before isopropyl thio-β-D-galactoside (IPTG) was added to a final concentration of 0.4 mM for GlcNAcase expression. Cell growth was continued at 20°C for an additional 18 h, and cells were harvested by centrifugation at 4,500 ×g for 30 min. The bacterial pellet was re-suspended in lysis buffer (20 mM Tris-HCl buffer, pH 8.0, 150 mM NaCl, 1 mM phenylmethylsulphonyl fluoride (PMSF), 5% (v/v) glycerol and 1 mg mL<sup>-1</sup> lysozyme, and then lysed on ice using a Sonopuls ultrasonic homogenizer with a 6-mm diameter probe (50% duty cycle; amplitude setting, 30%; total time, 30 s, 6–8 times). Unbroken cells and cell debris were removed by centrifugation at 12,000 ×g for 1 h. The supernatant containing *VhGlcNAcase* was immediately applied to a polypyrone column packed with 5 mL of TALON<sup>®</sup> Superflow<sup>™</sup> metal affinity resin (Clontech Laboratories, Inc., USA) operated at 4°C with gravity-dependent flow. The column was washed with 8 column volumes (cv) of equilibration buffer (20 mM Tris-HCl buffer, pH 8.0 containing 150 mM NaCl), followed by 7 cv of the equilibration buffer containing 10 mM imidazole. The protein was then eluted with 250 mM imidazole in the same buffer. Eluted fractions of 10 mL were collected and 15 µL of each fraction was analyzed by 12% SDS-PAGE, according to the method of Laemmli [29], to confirm the purity of the protein. Fractions with GlcNAcase activity were pooled and subjected to several rounds of centrifugation in Vivaspin-20 ultrafiltration membrane concentrators (10 kDa molecular-weight cut-off, Vivascience AG, Hannover, Germany) for complete removal of imidazole. The final concentration of the protein was determined by the BCA method [30]. The freshly prepared protein was either immediately used for functional characterization or stored at -80°C until used.

### GlcNAcase activity assay

GlcNAcase activity was determined by a colorimetric assay using *p*NP-GlcNAc as substrate. The reaction mixture in a 96-well microtiter plate contained an optimal amount of *VhGlcNAcase* (0.1 µg for WT and 5 µg for mutants), 500 µM *p*NP-GlcNAc and 100 mM phosphate buffer, pH 7.0 in a total volume of 200 µL. When sodium formate was added, the final pH in the reaction mixture was found to be 7.0 ± 0.3. The assay was carried out at 37°C with constant agitation in an Eppendorf ThermoMixer<sup>®</sup> Comfort (Eppendorf AG, Hamburg, Germany), and was terminated by adding 100 µL of 3 M Na<sub>2</sub>CO<sub>3</sub> to each well after 10 min. The concentration of *p*-nitrophenol (*p*NP) released was determined at 405 nm in a Biochrom Anthos Multi-Read 400 Microplate Reader (Biochrom, Cambridge, UK). The molar quantity of the liberated

Table 1. Primers used for site-directed mutagenesis.

Mutation	Oligonucleotide sequence <sup>a</sup>
D303A	forward 5' - CATTGGCATCTCACTGCGGATGAAGGCTGGCGTG - 3'
	reverse 5' - CAGCCAGCCTTCATCCGCAGTGAGATGCCAATG - 3'
D303N	forward 5' - CATTGGCATCTCACTAAGCATCAAGGCTGGCGTG - 3'
	reverse 5' - CAGCCAGCCTTCATCGTTAGTGAGATGCCAATG - 3'
D304A	forward 5' - GCATCTCACTGACGGGAAGCTGGCGTGTC - 3'
	reverse 5' - GACACGCCAGCCTTCGCGTCAGTGAGATGC - 3'
D304N	forward 5' - GCATCTCACTGACAAAGAAGCTGGCGTGTC - 3'
	reverse 5' - GACACGCCAGCCTTCGTTGTCAAGTGAGATGC - 3'
D437A	forward 5' - GTTACATTGGCGCGGGAAGTGCCTAACGGC - 3'
	reverse 5' - GCCGTTAGGCATTCGCGCGGCAATGTGAAC - 3'
D437N	forward 5' - GTTACATTGGCGGGAACGAAGTGCCTAACGGC - 3'
	reverse 5' - GCCGTTAGGCATTCGTTGCGGCCAATGTGAAC - 3'
E438A	forward 5' - CACATTGGCGCGGACCGGTCCTAACGGCGTCTG - 3'
	reverse 5' - CACACGCCGTTAGGCACCGGTCGCGCCAATGTG - 3'
E438Q	forward 5' - GTTACATTGGCGCGGACCGGTCCTAACGGCGTCTG - 3'
	reverse 5' - CACACGCCGTTAGGCACCTGGTCCGCGCCAATGTGAAC - 3'

<sup>a</sup> Sequences underlined indicate mutated codons.

doi:10.1371/journal.pone.0149228.t001

*p*NP was calculated from a calibration curve with *p*NP concentration varied from 0 to 20 nmol. The hydrolytic activity of the enzyme was expressed as the quantity of *p*NP (nmol) produced in 1 min at 37°C.

#### Determination of the pH optima of *Vh*GlcNAcase WT and D437A mutant

To obtain the activity/pH profiles, the specific activity of *Vh*GlcNAcase WT and D437A mutant was determined in a discontinuous assay. The reaction mixture contained 0.05 µg *Vh*GlcNAcase or 5 µg D437A, 500 µM *p*NP-GlcNAc, and McIlvaine's sodium phosphate-citric acid buffer, pH 3.0–9.0 [31] at different pH values ranging from 3.0 to 9.0, in a total volume of 200 µL. The reaction was carried out as described for the GlcNAcase activity assay.

#### Chemical rescue assay

Sodium azide and sodium formate were initially tested for their ability to rescue the enzymic activity of inactive *Vh*GlcNAcase mutants. A 200-µL assay mixture, prepared in a 96-well microtiter plate, contained 500 µM *p*NP-GlcNAc, 5 µg of enzyme, 1 M sodium azide or formate and 100 mM sodium phosphate buffer, pH 7.0. The reaction mixture was incubated at 37°C for 10 min with constant agitation, and the reaction was terminated by the addition of 100 µL of 3 M Na<sub>2</sub>CO<sub>3</sub>. The reaction of wild-type *Vh*GlcNAcase was carried out as described for mutants, but with 0.1 µg of the enzyme in the assay.

To determine the effect of concentrations on the rescued activity, the reaction mixture was incubated with different concentrations of azide or formate for a longer time. A 200-µL assay mixture contained 500 µM *p*NP-GlcNAc, 0.2 µg of mutants and 0.1–2.0 M sodium azide or formate in 100 mM sodium phosphate buffer, pH 7.0. The reaction was allowed to proceed at 37°C for 60 min, and then terminated by the addition of 100 µL of 3 M Na<sub>2</sub>CO<sub>3</sub>. The amount of *p*NP released was calculated as described above.

### Time-course and kinetics of sodium formate effects on the activity of *VhGlcNAcase* D437A mutant

Chemical rescue of the inactive D437A mutant by sodium formate was further observed at different times of incubation. A 200- $\mu$ L assay mixture contained 500  $\mu$ M *p*NP-GlcNAc, 0.2  $\mu$ g of the mutant D437A and 0.1–2.0 M sodium formate in 100 mM sodium phosphate, pH 7.0. The reaction mixture was incubated at 37°C for times of 0, 2.5, 5, 10, 30, and 60 min. For kinetic experiments, a 200- $\mu$ L reaction mixture, containing 0–500  $\mu$ M *p*NP-GlcNAc, 5  $\mu$ g of the mutant D437A, 0.1–2.0 M sodium formate and 100 mM sodium phosphate buffer, pH 7.0, was incubated for 10 min at 37°C and the reaction terminated with 100  $\mu$ L of 3 M Na<sub>2</sub>CO<sub>3</sub>. The amount of the *p*NP formed during the reaction was estimated as described previously. The kinetic parameters (apparent  $K_m$ , apparent  $k_{cat}$  and apparent  $k_{cat}/K_m$ ) were determined with a non-linear regression function available in GraphPad Prism v.5.0 (GraphPad Software Inc., San Diego, CA).

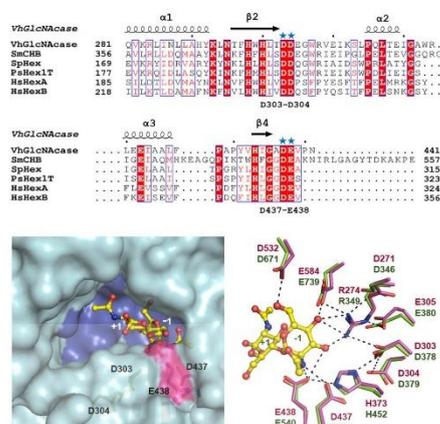
## Results

### Sequence analysis and homology modeling

We previously reported cloning and recombinant expression of the gene encoding GH20  $\beta$ -*N*-acetylglucosaminidase from the marine bacterium *V. harveyi* [9]. The enzyme, known as *VhGlcNAcase* (formerly *VhNag2*), exhibited exolytic activity, degrading chitin oligosaccharides from the non-reducing end in a sequential manner, with GlcNAc monomer as the final product. Since the crystal structure of *VhGlcNAcase* is undetermined, we first gained preliminary information on the structural identity of *VhGlcNAcase* by aligning its sequence with those of other GH20 GlcNAcases of known structure. The results showed that the highest sequence identity of *VhGlcNAcase* was with *Serratia marcescens* chitobiase (*SmCHB*), with 24% identity [32], followed by *Streptomyces plicatus*  $\beta$ -*N*-acetylhexosaminidase (*SpHex*) with 21% identity [17], human  $\beta$ -hexosaminidase A (*HsHexA*) [33] and human  $\beta$ -hexosaminidase B (*HsHexB*) [10] with 17% identity, while the lowest was with *Paenibacillus* sp.  $\beta$ -hexosaminidase (*PsHex1T*) which had 13% identity [34]. Structure-based alignment of *VhGlcNAcase* and *SmCHB* (Fig 1A) indicated two separate conserved segments on the surface of the ( $\beta/\alpha$ )<sub>8</sub> TIM barrel domain of the two enzymes. For *VhGlcNAcase*, the preceding segment comprises the acidic pair Asp303-Asp304, located at the end of loop2 (L2), which links strand  $\beta$ 2 and helix  $\alpha$ 2 (Fig 1A, upper sequence portion). The second pair, Asp437-Glu438, is present at the start of loop4 (L4) connecting strand  $\beta$ 4 and helix  $\alpha$ 4 (Fig 1A, lower sequence portion). Superimposition of the modelled structure of *VhGlcNAcase* with the crystal structure of *SmCHB* gave an R.M.S.D. of 0.651 Å for 390 C $\alpha$  atoms (Fig 1B) and showed both conserved acidic pairs to be part of the GlcNAc-binding pocket. The Asp437-Glu438 pair was located close to the scissile bond joining -1GlcNAc and +1GlcNAc, which suggested that these amino acids could play a catalytic role. Structural alignment of the active site residues (Fig 1C) showed that the location of the Asp303-Asp304 pair is equivalent to that of Asp378-Asp379 in *SmCHB*, whereas the Asp437-Glu438 pair corresponded with Asp539-Glu540 [35]. Based on the crystal structure and kinetic data, the Asp539-Glu540 pair had been suggested to have a catalytic function for *SmCHB* [35].

### Effects of the active-site mutation on the specific activity of *VhGlcNAcase*

Point mutations of the selected residues (Asp303, Asp304, Asp 437 and Glu438) caused a drastic loss of enzymatic activity (Table 2, column 2). The activity of mutants D303A, D304A and



**Fig 1. Comparison of the modelled structure of *VhGlcNAcase* with other *GlcNAcases*.** (A) Multiple sequence alignment of GH20 glycoside hydrolases. The amino acid sequence of *Vibrio harveyi*  $\beta$ -N-acetylglucosaminidase, *VhGlcNAcase* (SwissProt: D9ISE0) was retrieved from the Uniprot database. This sequence was aligned with those of *Serratia marcescens* chitinase, *SmChb* (SwissProt: Q54468), *Streptomyces plicatus*  $\beta$ -N-acetylhexosaminidase, *SpHex* (SwissProt: O85361), *Paenibacillus* sp.  $\beta$ -hexosaminidase, *Pshex1T* (SwissProt: D2KW09), human  $\beta$ -hexosaminidase A ( $\alpha$ -chain), *HsHexA* (SwissProt: P06865) and human  $\beta$ -hexosaminidase B ( $\beta$ -chain), *HsHexB* (SwissProt: P07686). The putative amino acid residues that are important for *GlcNAcase* activity are indicated with blue stars. (B) Surface representation of the active-site pocket of *VhGlcNAcase* (in blue) docked with *GlcNAc*<sub>2</sub> (in yellow stick) from *SmChb* (PDB entry: 1QBB). The solvent-accessible surface of D437-E438 is highlighted in pink and the buried surface of D303-D304 is highlighted in green. (C) Superimposition of both conserved acidic pairs (Asp303-Asp304 and Asp437-Glu438) of modelled *VhGlcNAcase* (in magenta stick) with the crystal structure of *SmChb* (in green stick) in complex with *GlcNAc*<sub>2</sub>. N atoms are shown in blue and O atoms in red.

doi:10.1371/journal.pone.0149228.g001

**Table 2. Specific activity of wild-type *VhGlcNAcase* and its mutants with pNP-GlcNAc as substrate in the absence and presence of 1 M sodium salts.** The presented values are Mean  $\pm$  S.D. obtained from experiments carried out in triplicate.

GlcNAcase mutant	Specific activity (nmol/min/ $\mu$ g)		
	No sodium salt	1 M NaH <sub>2</sub> PO <sub>4</sub>	1 M HCOONa
Wild-type	19.4 $\pm$ 0.22 (100) <sup>a</sup>	0.89 $\pm$ 0.14 (5)	15.4 $\pm$ 0.29 (80)
D303A	n.d.	n.d.	n.d.
D303N	0.81 $\pm$ 0.02 (100)	0.31 $\pm$ 0.02 (39)	0.75 $\pm$ 0.01 (93)
D304A	n.d.	n.d.	n.d.
D304N	0.29 $\pm$ 0.01 (100)	0.10 $\pm$ 0.01 (35)	0.27 $\pm$ 0.02 (92)
D437A	0.05 $\pm$ 0.01 (100)	0.01 $\pm$ 0.01 (27)	0.09 $\pm$ 0.01 (182)
D437N	n.d.	n.d.	n.d.
E438A	0.07 $\pm$ 0.01 (100)	0.03 $\pm$ 0.01 (46)	0.08 $\pm$ 0.01 (102)
E438Q	0.09 $\pm$ 0.01 (100)	0.06 $\pm$ 0.01 (65)	0.10 $\pm$ 0.01 (108)

<sup>a</sup> Numbers in brackets indicate the relative specific activities of *VhGlcNAcase* and its variants with each sodium salt, in comparison with *VhGlcNAcase* without added sodium salt (set to 100).

n.d: undetectable activity.

doi:10.1371/journal.pone.0149228.t002

D437N was undetectable, while the residual activity of other mutants, including D303N, D304N, D437A and E438A/Q, was less than 5% of that of WT *VhGlcNAcase* (Table 2, column 2). To study the rescue of activity by external nucleophiles, sodium azide and sodium formate, along with other sodium salts (for the chemical structures of azide and formate ions, see Fig A in S1 File), were first tested for their physicochemical effects on the activity of *VhGlcNAcase* WT. We recently observed that the enzymic activity of the unmutated (wild-type) *VhGlcNAcase* was inhibited by various sodium salts, including azide, nitrate, formate and chloride [36]. Here, we confirmed their inhibitory effects on the WT activity. As shown in Fig B in S1 File, the specific activity of *VhGlcNAcase* WT decreased greatly when sodium azide or sodium nitrate was included in the assay medium. On the other hand, sodium formate and sodium chloride showed only moderate effects. For all the ions tested, the degree of inhibition increased with increasing concentration.

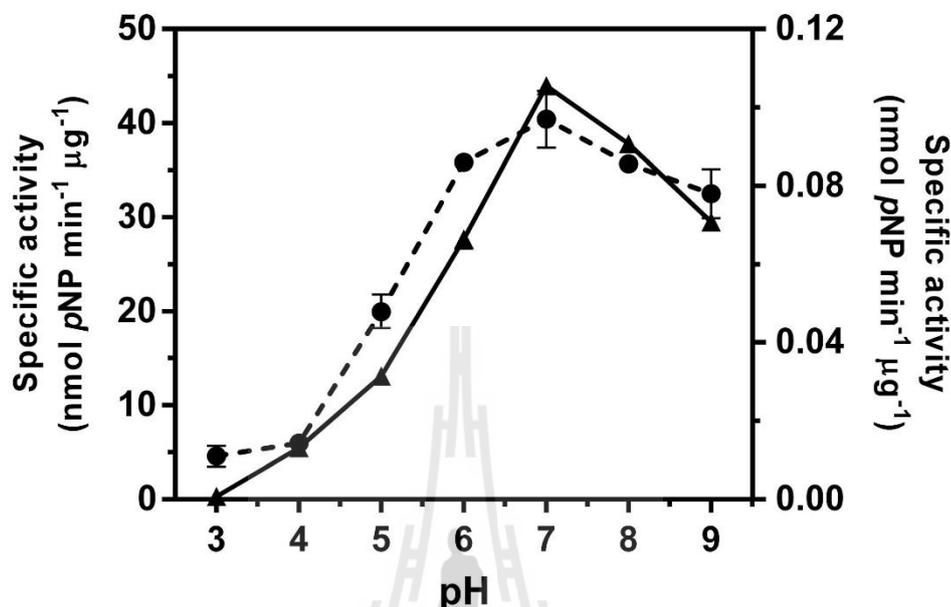
Next we tested two selected compounds, sodium azide and sodium formate, for their ability to rescue the enzymic activity of the inactive mutants. The results clearly showed that sodium formate had much less inhibitory effect on the *VhGlcNAcase* mutants than on the WT enzyme (Table 2, column 3). When 1 M sodium azide was included in the assay medium the specific activity of the *VhGlcNAcase* WT was less than 5% of the original activity, while the residual activity of the mutants was 27–65% of the original. On addition of sodium formate (Table 2, column 4), relatively less inhibition, or even enhancement of activity, was also observed with the enzyme variants. Notably, the specific activity of the D437A mutant was enhanced to 182% of the basal activity in the presence this compound.

#### Effects of sodium formate concentration on the rescued activity of the D437A mutant

Since only for mutant D437A was the specific activity significantly enhanced by sodium formate, we examined whether this mutant showed a shift in the activity/pH curve compared to the WT enzyme. Fig 2 shows the similar response of the activity of the two *VhGlcNAcase* forms to pH variation. Although mutant D437A had a slightly broader activity/pH curve than the WT enzyme, the two forms had a similar optimal pH of around 7.0. Next, we investigated whether the enzyme activity of the *VhGlcNAcase* D437A was modified by sodium formate in a concentration-dependent manner. In this set of experiments, we also included the effect of concentration on the activity of the E438A mutant, for comparison. Fig 3 shows plots of the fractional activity ( $v_i/v_0$ ) of the enzyme at discrete concentrations of sodium formate. The relative activity of *VhGlcNAcase* WT was found to decline in response to increasing concentrations of sodium formate from 0.1 to 2.0 M; at the highest concentration, the residual activity of the WT enzyme was reduced to less than half of its original value. In marked contrast, the relative activity of mutant D437A was elevated with increasing sodium formate concentration, and at 2.0 M sodium formate was four times the original activity, while the relative activity of mutant D437N increased slightly (about 1.7-fold) (Fig 3A). However an increase in concentration of sodium formate did not restore the enzymatic activity of the E438A or E438Q mutants (Fig 3B). When the same set of enzyme variants was tested with sodium azide, decreases in the fractional activity were observed (data not shown).

#### Steady state kinetics of activation by sodium formate

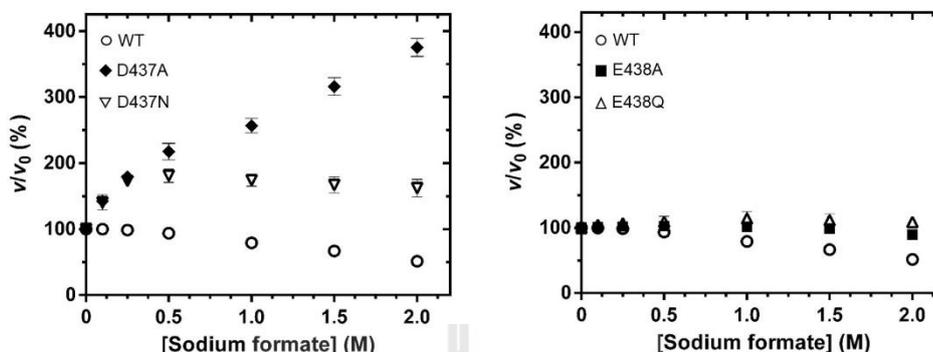
In order to determine the initial rate of reaction, the product generated in the course of pNP-GlcNAc hydrolysis by the mutant D437A was monitored at different time points. Fig 4A is a plot of pNP release against time, showing that the initial rate of the reaction with and without sodium formate could be determined within 10 min, the amount of product formed up to



**Fig 2. Activity/pH profiles of VhGlcNAcase and its mutant D437A.** The specific activity of VhGlcNAcase (solid line, left y axis) and the mutant D437A (dashed line, right y axis) was measured at pH = 3.0, 4.0, 5.0, 6.0, 7.0, 8.0 and 9.0 in the McIlvaine's sodium phosphate-citric acid buffer system. pNP-GlcNAc was used as substrate and the reaction was carried out for 10 min at 37°C.

doi:10.1371/journal.pone.0149228.g002

this time being directly proportional to the time of incubation (Fig 4A, inset). Fig 4B shows non-linear increases in the initial reaction rates for the D437A mutant with increasing substrate concentration, and discrete increases in sodium formate concentrations from 0.1 to 2.0 M. These plots exhibit typical Michaelis-Menten kinetics, where the apparent maximum rate of reaction ( $\text{app } V_{\text{max}}$ ) is approached at concentrations of pNP-GlcNAc above 500  $\mu\text{M}$ . Inverse transformation of the non-linear plots in Fig 4B yields linear Lineweaver-Burk (LB) plots, as shown in Fig 4C. Each plot, representing the relation of  $1/v_0$  and  $1/[S]$ , allows estimation of the kinetic parameters of the enzyme in the absence and presence of sodium formate. The kinetic data for the mutant D437A in Table 3 indicate discrete increases in apparent values of  $K_m$ ,  $k_{\text{cat}}$ , and  $k_{\text{cat}}/K_m$ . The enhancement of the enzymic activity of VhGlcNAcase on addition of sodium formate is shown in Fig 5 as plots of  $(\text{app } K_m)/K_m$ ,  $(\text{app } k_{\text{cat}})/k_{\text{cat}}$ , and  $(\text{app } k_{\text{cat}}/K_m)/(k_{\text{cat}}/K_m)$ , all relative to the values in the absence of formate (constant/constant<sub>0</sub>), as a function of formate concentration. The data analysis indicates a small, concentration-dependent increase in the  $K_m$  value, reaching 1.3-times the reference value at 2.0 M sodium formate. In contrast, very significant increases in the constant  $k_{\text{cat}}$  were observed, and at 2.0 M sodium formate,  $k_{\text{cat}}$  was 2.5 fold greater than at 0 M. Hence, the ratio  $k_{\text{cat}}/K_m$  ration was increased to 1.9-times the reference ratio, in the presence of 2.0 M sodium formate.

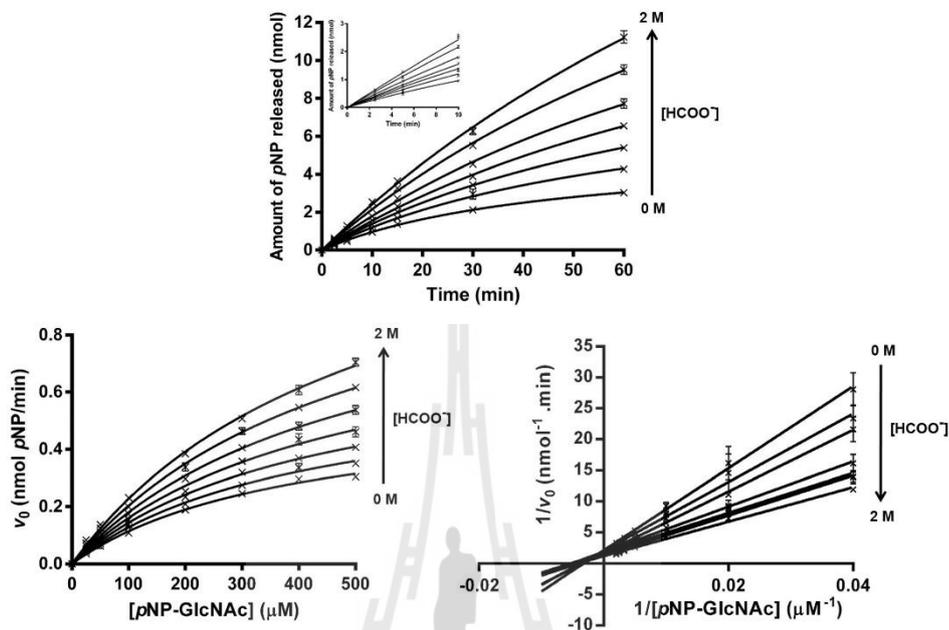


**Fig 3. Effect of sodium formate on pNP-GlcNAc hydrolysis by  $VhGlcNAcase$  and its mutants.** Various concentrations of sodium formate (0.0–2.0 M) were added to the reaction mixture, which contained 500  $\mu$ M pNP-GlcNAc and 100 mM sodium phosphate buffer, pH 7.0, at 37°C.  $v/v_0$  is the fractional activity of the enzyme, i.e. activity in the presence of sodium formate relative to that in its absence. (A) The D437A (filled diamonds) and D437N (open inverted triangles) mutants. (B) The mutants E438A (filled squares) and E438Q (open triangles). The wild-type  $VhGlcNAcase$  (open circles) are shown in both A and B.

doi:10.1371/journal.pone.0149228.g003

## Discussion

Chitin turnover in the marine biosphere depends upon the activities of marine *Vibrios* [37,38]. The chitin catabolic cascade of the *Vibrios* has been demonstrated to involve a large number of genes and enzymes, which are orchestrated in a complex signal transduction pathway [2,3,37–41]. We previously identified and characterized three biological components of the chitin catabolic pathway that are essential for chitin degradation and chitin uptake by *V. harveyi*. Chitinase A (so-called  $VhChiA$ ) is an endolytic enzyme responsible for the breakdown of insoluble chitin chains into small, soluble chitooligosaccharides [8,42], while chitoporin (so-called  $VhChiP$ ), a sugar-specific porin located in the outer membrane of the bacterium, is responsible for chitooligosaccharide uptake [6,7]. The last component is  $\beta$ -N-acetylglucosaminidase (known as  $VhGlcNAcase$  or formerly  $VhNag2$ ), an exolytic enzyme capable of degrading the transported chitooligosaccharides to GlcNAc monomers, which then act as signalling molecules that regulate the downstream cascade of the chitin catabolic pathway, through the activation of the chitin sensor ( $ChiS$ ) [40,43].  $VhGlcNAcase$ , a member of the GH20 GlcNAcase family, contains four GlcNAc binding subsites (-1), (+1), (+2) and (+3), and exhibits its greatest activity with chitotetraose [9]. Amino acid sequence comparison with other GlcNAcases and our 3D-structure, modelled on the known 3D structure of *SmCHB* (Fig 1A, 1B and 1C), suggested that two invariant acidic side-chain pairs, Asp303-Asp304 and Asp437-Glu438, could be important for catalysis. Both acidic pairs lie in close proximity to the cleavage site (-1 subsite) and have equal opportunity to act as the catalytic couple. In this study, we performed site-directed mutagenesis, followed by a chemical rescue assay, to identify the catalytic couple. In the first set of experiments, we observed that point mutations of four invariant acidic residues (Asp303, Asp304, Asp437 and Glu438) caused a drastic loss of the enzymic activity of  $VhGlcNAcase$  toward a synthetic substrate, pNP-GlcNAc. Notably, mutations of Asp437 to Ala (mutant D437A) and of Glu438 to Ala (mutant E438A) abolished the activity almost completely, confirming that these acidic residues play important roles in chitin degradation.



**Fig 4.** (A) Time-courses of reactions of the D437A mutant with and without sodium formate. Reaction mixtures (200  $\mu$ L), containing 2  $\mu$ g of D437A mutant and 500  $\mu$ M of pNP-GlcNAc and varied concentrations of sodium formate (0, 0.1, 0.25, 0.5, 1.0, and 2.0 M) and 100 mM sodium phosphate buffer, pH 7.0, were incubated at 37°C for 0–60 min, and the reaction terminated with 100  $\mu$ L of 3 M  $\text{Na}_2\text{CO}_3$ . Release of pNP, monitored at  $A_{405}$ , was converted to molar quantities using a calibration curve of pNP (0–20 nmol). The linear part of the reaction progress was shown as an inset. (B) Initial reaction rates for the mutant D437A of VhGlcNAcase in the presence of sodium formate were obtained from Michaelis-Menten plots. Reaction rates were measured using pNP-GlcNAc (0–500  $\mu$ M) as the substrate, 5  $\mu$ g of the mutant D437A of VhGlcNAcase and sodium formate at the same range of concentrations as described above. (C) Activation by formate anion was evaluated by means of Lineweaver-Burk plots of initial reaction rates.

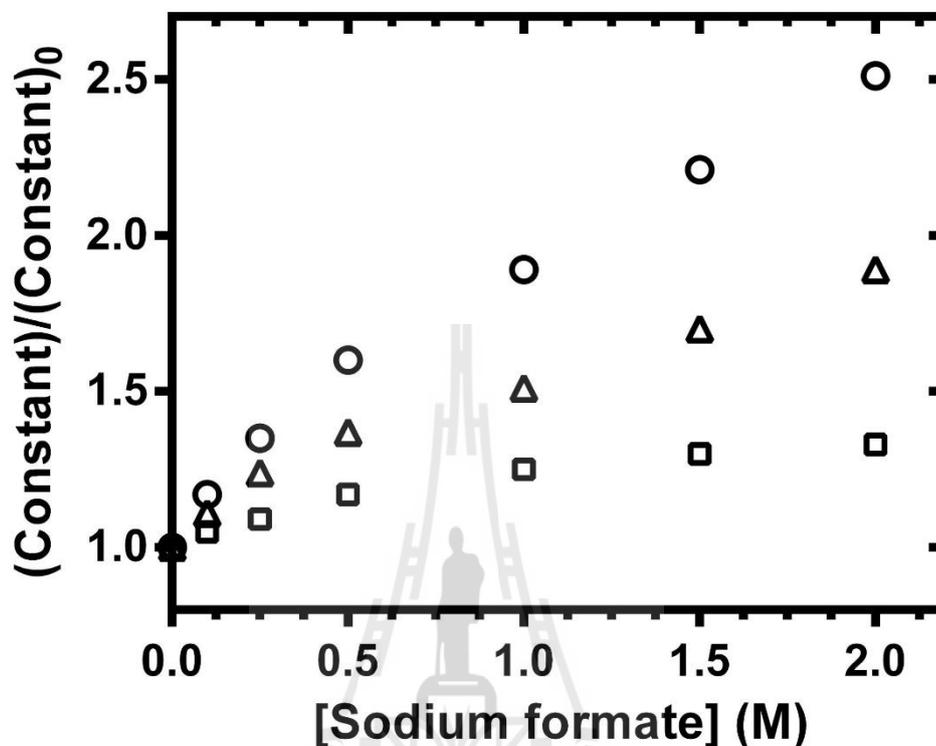
doi:10.1371/journal.pone.0149228.g004

**Table 3.** Kinetic parameters for the hydrolytic activity of VhGlcNAcase mutant D437A, in the presence of increasing sodium formate ( $\text{HCOONa}$ ) concentration. The presented values are Mean  $\pm$  S.D. obtained from experiments carried out in triplicate.

Sodium formate concentration (M)	$K_m$ ( $\mu$ M)	$k_{cat}$ ( $\text{s}^{-1}$ )	$k_{cat}/K_m$ ( $\text{s}^{-1} \text{mM}^{-1}$ )
0	390 $\pm$ 32 (100) <sup>a</sup>	0.14 $\pm$ 0.01 (100)	0.36 (100)
0.10	411 $\pm$ 37 (105)	0.16 $\pm$ 0.01 (117)	0.40 (111)
0.25	425 $\pm$ 29 (109)	0.19 $\pm$ 0.007 (135)	0.44 (124)
0.50	456 $\pm$ 56 (117)	0.22 $\pm$ 0.02 (160)	0.49 (137)
1.00	489 $\pm$ 41 (125)	0.26 $\pm$ 0.01 (189)	0.54 (151)
1.50	509 $\pm$ 33 (130)	0.31 $\pm$ 0.01 (221)	0.60 (170)
2.00	519 $\pm$ 38 (133)	0.35 $\pm$ 0.02 (251)	0.67 (189)

<sup>a</sup> Values in brackets represent relative activity compared to that without sodium formate (set as 100)

doi:10.1371/journal.pone.0149228.t003



**Fig 5. Chemical rescue of the D437A mutant by sodium formate.** Three kinetic constant ratios; the apparent first-order rate constants,  $k_{cat}$  (open circles), apparent Michaelis constants,  $K_m$  (open squares) and apparent second-order rate constants,  $k_{cat}/K_m$  (open triangles), were plotted as a function of sodium formate concentration.

doi:10.1371/journal.pone.0149228.g005

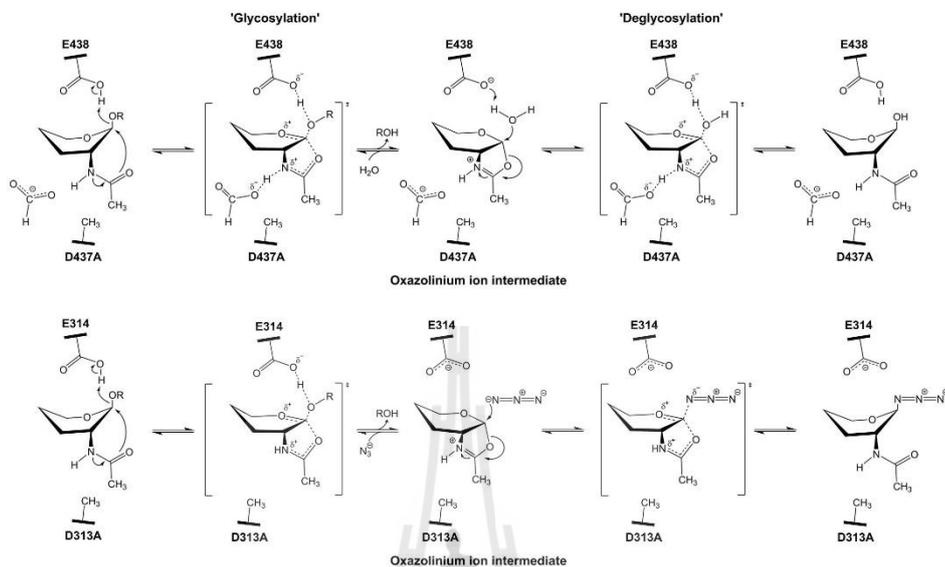
In the next experiment, we observed that among various sodium salts, sodium azide greatly inhibited the activity of *VhGlcNAcase* WT, but sodium formate produced only weak inhibition (Fig B in [S1 File](#)). Such observations were consistent with our previous report that sodium azide acted as a potent competitive inhibitor of *VhGlcNAcase* [36]. Both azide and formate ions, the forms of sodium azide and sodium formate, respectively, that exist in buffered solution, are strong nucleophiles [44,45]. Therefore, their ability to rescue enzymic activity of inactive mutants through nucleophilic effect has been employed to elucidate the catalytic mechanism of several retaining glycoside hydrolases [19,23–27]. In our study, their inhibitory effects on the mutant forms of *VhGlcNAcase* were significantly less than on WT, suggesting that the inactivating effects of the active-site mutations were partially eliminated when azide or formate was included in the assayed reaction. Formate ion appeared to act as the more potent nucleophile, as we observed its greater chemical rescue effect on the mutants D437A and

E438A/Q, as compared to that of azide ion; the enzymic activity of the mutant D437A was even enhanced by formate, but not by azide. The less effective chemical rescue produced by azide ion may result from its linear geometry, which allows only a poor fit into the catalytic pocket of *VhGlcNAcase*. In contrast formate ion, which has trigonal planar geometry, may accurately mimic the carboxylate side chain of Asp437 (Fig A in S1 File). Therefore, the activity loss due to the interruption of the catalytic cycle, caused by loss of the natural nucleophile upon replacement of Asp437 with Ala, could be re-established in the presence of this small exogenous nucleophile.

Lineweaver Burk plots of  $1/v_o$  vs.  $1/[s]$  at different formate concentrations (Fig 4C) yielded lines that intersect above the x-axis, agreeing with the mix-type mode of binding. The results suggested that formate ion could interact with both unliganded D437A ( $E^{mut}$ ) and ligand-bound D437A ( $E^{mut}S$ ), but the enhanced activity would occur only when formate ion bound to the enzyme-substrate complex. Fig 6 shows the proposed mechanism, in which formate ion increases the rate of *pNP*-glycoside hydrolysis by replacing the substituted side-chain of Asp437 in the catalytic pocket of the D437A-substrate complex. The restoration of the enzyme activity observed with the mutant D437A suggested a crucial role for Asp437 as the catalytic nucleophile in both the glycosylation and deglycosylation steps of the substrate-assisted mechanism proposed for GH20 GlcNAcases [21].

The absence of any shift in the optimal pH in the pH-activity profiles of the *VhGlcNAcase* WT and the D437A mutant suggested that D437 did not facilitate bond cleavage by lowering the  $pK_a$  value of its catalytic partner Glu438. This is a major difference in the catalytic role of Asp437 in *VhGlcNAcase* from Asp313 in *SpHEX*. In the case of *SpHEX*, the D313A mutant was shown to have its optimal pH value increased from 5.0 to 7.5, the pH/activity data suggesting a significant contribution to bond cleavage by Asp313 [21].

When compared with *SpHEX* [21], we propose that the effects of formate ion on the *VhGlcNAcase* inactive mutant are mechanistically different from the effects of azide ion on the *SpHEX* inactive mutant. As shown in Fig 6A, formate ion acts as a nucleophilic substitute for Asp437 in the *VhGlcNAcase* D437A mutant, the planar formate ion ( $HCOO^-$ ) optimally filling the volume occupied in WT *VhGlcNAcase* by the  $\beta-COO^-$  side chain. Formate then accepts a proton from the -NH of the C2-acetamido group of the oxazolinium intermediate that is generated in course of the scissile-bond cleavage by the acid catalyst Glu438. Such covalent bond formation aids the stabilization of the oxazolinium intermediate in the glycosylation step, and also helps to orient the positively charged C1 of the reaction intermediate, so as to react with the neighbouring water molecule in the subsequent deglycosylation step. Our data show that formate-mediated chemical rescue produced similar increases in (apparent  $k_{cat}$ )/ $k_{cat}$ , and (apparent  $k_{cat}/K_m$ )/( $k_{cat}/K_m$ ) (2.5- and 1.9-fold respectively), supporting the catalytic role of Asp437 in both glycosylation and deglycosylation steps in the enzyme-substrate complex. In marked contrast, azide ion rescued the GlcNAcase activity of *SpHEX* inactive mutant (D313A) by acting as an alternative nucleophile to water (not to Asp313), generation a  $\beta$ -glycosyl azide product. As shown in Fig 6B, azide ion does not mimic the nucleophilic role of Asp313, but the result of the nucleophilic attack by azide ion is to open the oxazolinium ion intermediate during the deglycosylation step. This proposed mechanism was revealed by the kinetic analysis, which showed a much greater increase in the first-order rate constant ratio (apparent  $k_{cat}$ )/ $k_{cat}$  (16-fold) than in the second-order rate constant ratio (apparent  $k_{cat}/K_m$ )/( $k_{cat}/K_m$ ) (5-fold). In *VhGlcNAcase* such analysis suggested that the exogenous nucleophile restored the activity of the D437A mutant by accelerating both the rate of deglycosylation (as reflected by the apparent  $k_{cat}$ ) and of glycosylation (as reflected by the apparent  $k_{cat}/K_m$ ). It is noteworthy that the rescue effect observed for our inactive *VhGlcNAcase* was not dramatic, and this may reflect the reactivity of the leaving group on the tested substrate. *pNPGlcNAc* contains a poor leaving group,



**Fig 6.** (A) Proposed mechanism of formate-mediated chemical rescue of the activity of the *VhGlcNAcase* D437A mutant. Formate ion is involved in both the glycosylation and deglycosylation steps by providing charge stabilization of transition states that flank the oxazolinium ion. (B) Proposed mechanism of azide-mediated chemical rescue with the Sp-Hex D313A mutant. Azide ion is involved only in the deglycosylation step, acting to open the cyclic oxazolinium ion intermediate [21]. Hydroxyl groups and  $C_6$  have been omitted for clarity.

doi:10.1371/journal.pone.0149228.g006

so is much less susceptible to enzymic hydrolysis than 2,4DNPGlcNAc and 3,5DNPGlcNAc, which contain strong leaving groups. Similar results were reported by Vallmitjana et al. [23]. They observed only a 3 fold enhancement of  $k_{cat}$  when *p*NPG was used as the substrate for the  $\beta$ -glucosidase assay of the nucleophilic inactive mutant E178A, while 188 fold  $k_{cat}$  enhancement was observed when 2,4DNPG was the substrate. This would explain the modest 2.5 fold increase in  $k_{cat}$  for *VhGlcNAcase* inactive mutant with *p*NPGlcNAc as substrate, compared to SpHEX, which showed a 16 fold increase in  $k_{cat}$  with 3,5DNPGlcNAc substrate [21].

## Conclusions

In this study, we have demonstrated that an exogenous nucleophile (formate ion) selectively enhances the enzymatic activity of an inactive mutant *VhGlcNAcase*, D437A, in a concentration-dependent manner. However, the activity of other active-site mutants (D303A/N, D304A/N, and E438A/Q) was not significantly affected by the addition of this strong nucleophile. The rescued activity of the D437A mutant suggests that Asp437 is the catalytic nucleophile, while its invariant acidic partner Glu438 likely acts as a catalytic proton-donating residue. This experimental evidence confirms that the residues Asp437 and Glu438, located in the middle of the substrate-binding cleft in the modelled structure of GH20 *VhGlcNAcase*, act as the catalytic pair in the catalytic cycle of chitooligosaccharide hydrolysis by this enzyme.

## Supporting Information

**S1 File. The chemical structures of azide and formate ions used in this study (Fig A) and specific hydrolytic activity of wild-type *VhGlcNAcase* against pNP-GlcNAc, in the presence of various concentrations of sodium salts (Fig B).**  
(PDF)

## Acknowledgments

We would like to acknowledge Biochemistry-Electrochemistry Research Unit and Biochemistry Laboratory, and the Centre for Scientific and Technological Equipment, Suranaree University of Technology for providing the facilities for carrying out this research. We greatly appreciate a critical proofreading of this manuscript by Dr. David Apps, Centre for Integrative Physiology, School of Biomedical Sciences, University of Edinburgh, United Kingdom.

## Author Contributions

Conceived and designed the experiments: PM WS. Performed the experiments: PM. Analyzed the data: PM WS. Contributed reagents/materials/analysis tools: WS. Wrote the paper: PM WS.

## References

1. Bassler BL, Yu C, Lee YC, Roseman S. Chitin utilization by marine bacteria. Degradation and catabolism of chitin oligosaccharides by *Vibrio furnissii*. *J Biol Chem* 1991; 266:24276–24286. PMID: [1761533](#)
2. Li X, Roseman S. The chitinolytic cascade in *Vibrios* is regulated by chitin oligosaccharides and a two-component chitin catabolic sensor/kinase. *Proc Natl Acad Sci USA* 2004; 101:627–631. PMID: [14699052](#)
3. Yu C, Lee AM, Bassler BL, Roseman S. Chitin utilization by marine bacteria. A physiological function for bacterial adhesion to immobilized carbohydrates. *J Biol Chem* 1991; 266:24260–24267. PMID: [1761531](#)
4. Hamre AG, Eide KB, Wold HH, Sorlie M. Activation of enzymatic chitin degradation by a lytic polysaccharide monooxygenase. *Carbohydr Res* 2015; 407:166–169. doi: [10.1016/j.carres.2015.02.010](#) PMID: [25812992](#)
5. Lo Leggio L, Simmons TJ, Poulsen JC, Frandsen KE, Hemsworth GR, Stringer MA, et al. Structure and boosting activity of a starch-degrading lytic polysaccharide monooxygenase. *Nat Commun* 2015; 6:5961. doi: [10.1038/ncomms6961](#) PMID: [25608804](#)
6. Suginta W, Chumjan W, Mahendran KR, Janning P, Schulte A, Winterhalter M. Molecular uptake of chitooligosaccharides through chitoporin from the marine bacterium *Vibrio harveyi*. *PLoS One* 2013; 8: e55126. doi: [10.1371/journal.pone.0055126](#) PMID: [23383078](#)
7. Suginta W, Chumjan W, Mahendran KR, Schulte A, Winterhalter M. Chitoporin from *Vibrio harveyi*, a channel with exceptional sugar specificity. *J Biol Chem* 2013; 288:11038–11046. doi: [10.1074/jbc.M113.454108](#) PMID: [23447539](#)
8. Suginta W, Vongsuwan A, Songsiririthigul C, Prinz H, Estibeiro P, Duncan RR, et al. An endochitinase A from *Vibrio carchariae*: cloning, expression, mass and sequence analyses, and chitin hydrolysis. *Arch Biochem Biophys* 2004; 424:171–180. PMID: [15047189](#)
9. Suginta W, Chuenark D, Mizuhara M, Fukamizo T. Novel  $\beta$ -N-acetylglucosaminidases from *Vibrio harveyi* 650: cloning, expression, enzymatic properties, and subsite identification. *BMC Biochem* 2010; 11:40. doi: [10.1186/1471-2091-11-40](#) PMID: [20920218](#)
10. Mark BL, Mahuran DJ, Chemey MM, Zhao D, Knapp S, James MN. Crystal structure of human  $\beta$ -hexosaminidase B: understanding the molecular basis of Sandhoff and Tay-Sachs disease. *J Mol Biol* 2003; 327:1093–1109. PMID: [12662933](#)
11. Mark BL, Wasney GA, Salo TJ, Khan AR, Cao Z, Robbins PW, et al. Structural and functional characterization of *Streptomyces plicatus*  $\beta$ -N-acetylhexosaminidase by comparative molecular modeling and site-directed mutagenesis. *J Biol Chem* 1998; 273:19618–19624. PMID: [9677388](#)

12. Henrissat B, Bairoch A. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* 1993; 293:781–788. PMID: [8352747](#)
13. Henrissat B, Davies G. Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol* 1997; 7:637–644. PMID: [9345621](#)
14. Vocadlo DJ, Mayer C, He S, Withers SG. Mechanism of action and identification of Asp242 as the catalytic nucleophile of *Vibrio furnisii* N-acetyl-β-D-glucosaminidase using 2-acetamido-2-deoxy-5-fluoro-α-L-idopyranosyl fluoride. *Biochemistry* 2000; 39:117–126. PMID: [10625486](#)
15. Vocadlo DJ, Withers SG. Detailed comparative analysis of the catalytic mechanisms of β-N-acetylglucosaminidases from families 3 and 20 of glycoside hydrolases. *Biochemistry* 2005; 44:12809–12818. PMID: [16171396](#)
16. Drouillard S, Armand S, Davies GJ, Vorgias CE, Henrissat B. *Serratia marcescens* chitinase is a retaining glycosidase utilizing substrate acetamido group participation. *Biochem J* 1997; 328:945–949. PMID: [9396742](#)
17. Mark BL, Vocadlo DJ, Knapp S, Triggs-Raine BL, Withers SG, James MN. Crystallographic evidence for substrate-assisted catalysis in a bacterial β-hexosaminidase. *J Biol Chem* 2001; 276:10330–10337. PMID: [11124970](#)
18. Rao FV, Dorfmueller HC, Villa F, Allwood M, Eggleston IM, van Aalten DMF. Structural insights into the mechanism and inhibition of eukaryotic O-GlcNAc hydrolysis. *EMBO J* 2006; 25:1569–1578. PMID: [16541109](#)
19. Viladot JL, de Ramon E, Durany O, Planas A. Probing the mechanism of *Bacillus* 1,3–1,4-β-D-glucan 4-glucanohydrolases by chemical rescue of inactive mutants at catalytically essential residues. *Biochemistry* 1998; 37:11332–11342. PMID: [9698381](#)
20. Maier T, Sträter N, Schuette CG, Klungenstein R, Sandhoff K, Saenger W. The X-ray crystal structure of human β-hexosaminidase B provides new insights into Sandhoff disease. *J Mol Biol* 2003; 328:669–681. PMID: [12706724](#)
21. Williams SJ, Mark BL, Vocadlo DJ, James MN, Withers SG. Aspartate 313 in the *Streptomyces plicatus* hexosaminidase plays a critical role in substrate-assisted catalysis by orienting the 2-acetamido group and stabilizing the transition state. *J Biol Chem* 2002; 277:40055–40065. PMID: [12171933](#)
22. Fujita K, Sato R, Toma K, Kitahara K, Suganuma T, Yamamoto K, et al. Identification of the catalytic acid base residue of *Arthrobacter endo*-β-N-acetylglucosaminidase by chemical rescue of an inactive mutant. *J Biochem* 2007; 142:301–306. PMID: [17567654](#)
23. Vallmitjana M, Ferrer-Navarro M, Planell R, Abel M, Ausin C, Querol E, et al. Mechanism of the family 1 β-glucosidase from *Streptomyces* sp: catalytic residues and kinetic studies. *Biochemistry* 2001; 40:5975–5982. PMID: [11352732](#)
24. Paal K, Ito M, Withers SG. *Paenibacillus* sp. TS12 glucosylceramidase: kinetic studies of a novel sub-family of family 3 glycosidases and identification of the catalytic residues. *Biochem J* 2004; 378:141–149. PMID: [14561218](#)
25. MacLeod AM, Lindhorst T, Withers SG, Warren RAJ. The acid/base catalyst in the exoglucanase/xylanase from *Cellulomonas fimi* is glutamic acid 127: evidence from detailed kinetic studies of mutants. *Biochemistry* 1994; 33:6371–6376. PMID: [7910761](#)
26. Cobucci-Ponzano B, Trincone A, Giordano A, Rossi M, Moracci M. Identification of the catalytic nucleophile of the family 29 α-L-fucosidase from *Sulfolobus solfataricus* via chemical rescue of an inactive mutant. *Biochemistry* 2003; 42:9525–9531. PMID: [12911294](#)
27. Shallom D, Belakhov V, Solomon D, Shoham G, Baasov T, Shoham Y. Detailed kinetic analysis and identification of the nucleophile in α-L-arabinofuranosidase from *Geobacillus stearothermophilus* T-6, a family 51 glycoside hydrolase. *J Biol Chem* 2002; 277:43667–43673. PMID: [12221104](#)
28. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 2014; 42:W320–324. doi: [10.1093/nar/gku316](#) PMID: [24753421](#)
29. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 1970; 227:680–685. PMID: [5432063](#)
30. Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, Provenzano MD, et al. Measurement of protein using bicinchoninic acid. *Anal Biochem* 1985; 150:76–85. PMID: [3843705](#)
31. McLivaine TC. A buffer solution for colorimetric comparison. *J Biol Chem* 1921; 49:183–186.
32. Tews I, Perrakis A, Oppenheim A, Dauter Z, Wilson KS, Vorgias CE. Bacterial chitinase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease. *Nature Struct Biol* 1996; 3:638–648. PMID: [8673609](#)
33. Lemieux MJ, Mark BL, Cherney MM, Withers SG, Mahuran DJ, James MN. Crystallographic structure of human β-hexosaminidase A: interpretation of Tay-Sachs mutations and loss of G<sub>M2</sub> ganglioside hydrolysis. *J Mol Biol* 2006; 359:913–929. PMID: [16698036](#)

34. Sumida T, Ishii R, Yanagisawa T, Yokoyama S, Ito M. Molecular cloning and crystal structural analysis of a novel  $\beta$ -N-acetylhexosaminidase from *Paenibacillus* sp. TS12 capable of degrading glycosphingolipids. *J Mol Biol* 2009; 392:87–99. PMID: [19524595](#)
35. Prag G, Papanikolaou Y, Tavlas G, Vorgias CE, Petratos K, Oppenheim AB. Structures of chitinase mutants complexed with the substrate di-N-acetyl-D-glucosamine: the catalytic role of the conserved acidic pair, aspartate 539 and glutamate 540. *J Mol Biol* 2000; 300:611–617. PMID: [10884356](#)
36. Sirmontree P, Fukamizo T, Suginta W. Azide anions inhibit GH-18 endochitinase and GH-20 Exo  $\beta$ -N-acetylglucosaminidase from the marine bacterium *Vibrio harveyi*. *J Biochem* 2016; 159:191–200. doi: [10.1093/jb/mw087](#) PMID: [26330565](#)
37. Jeuniaux C, Voss-Foucart MF. Chitin biomass and production in the marine environment. *Biochem Syst Ecol* 1991; 19:347–356.
38. Zobell CE, Rittenberg SC. The occurrence and characteristics of chitinoclastic bacteria in the sea. *J Bacteriol* 1938; 35:275–287. PMID: [16560102](#)
39. Bassler BL, Gibbons PJ, Yu C, Roseman S. Chitin utilization by marine bacteria. Chemotaxis to chitin oligosaccharides by *Vibrio furnissii*. *J Biol Chem* 1991; 266:24268–24275. PMID: [1761532](#)
40. Keyhani NO, Roseman S. Physiological aspects of chitin catabolism in marine bacteria. *Biochim Biophys Acta* 1999; 1473:108–122. PMID: [10580132](#)
41. Park JK, Keyhani NO, Roseman S. Chitin catabolism in the marine bacterium *Vibrio furnissii*. Identification, molecular cloning, and characterization of a *N*, *N*-diacetylchitobiose phosphorylase. *J Biol Chem* 2000; 275:33077–33083. PMID: [10913116](#)
42. Suginta W, Vongsuwan A, Songsiririthigul C, Svasti J, Prinz H. Enzymatic properties of wild-type and active site mutants of chitinase A from *Vibrio carchariae*, as revealed by HPLC-MS. *FEBS J* 2005; 272:3376–3386. PMID: [15978043](#)
43. Hunt DE, Gevers D, Vahora NM, Polz MF. Conservation of the chitin utilization pathway in the *Vibrionaceae*. *Appl Environ Microbiol* 2008; 74:44–51. PMID: [17933912](#)
44. Comfort DA, Bobrov KS, Ivanen DR, Shabalina KA, Harris JM, Kulminkaya AA, et al. Biochemical analysis of *Thermotoga maritima* GH36  $\alpha$ -galactosidase (TmGalA) confirms the mechanistic commonality of clan GH-D glycoside hydrolases. *Biochemistry* 2007; 46:3319–3330. PMID: [17323919](#)
45. Zechel DL, Withers SG. Dissection of nucleophilic and acid-base catalysis in glycosidases. *Curr Opin Chem Biol* 2001; 5:643–649. PMID: [11738173](#)



## APPENDIX D

### LIST OF PRESENTATION

**Piyanat Meekrathok**, Arthur T. Porfetye, Marco Bürger, Ingrid R. Vetter and Wipa Suginta. Functional and structural analysis of a GH20  $\beta$ -N-acetylglucosaminidase from the marine bacterium *Vibrio harveyi*. The 29<sup>th</sup> Annual Symposium of the Protein Society, Fira De Barcelona-Montjuïc, Barcelona, Spain, July 22<sup>nd</sup>-25<sup>th</sup>, 2015. P22, *Poster presentation*.

**Piyanat Meekrathok** and Wipa Suginta. Identification of the catalytic base/nucleophile of a GH20  $\beta$ -N-acetylglucosaminidase from *Vibrio harveyi* through chemical rescue of inactive mutants. The 10<sup>th</sup> International Symposium of the Protein Society of Thailand, Chulabhorn Research Institute Convention Center, Bangkok, Thailand, July 15<sup>th</sup>-17<sup>th</sup>, 2015. P36, *Oral presentation*.

**Piyanat Meekrathok**, Arthur T. Porfetye, Marco Bürger, Ingrid R. Vetter and Wipa Suginta. Structural and functional characterization of a novel GH20  $\beta$ -N-acetylglucosaminidase (GlcNAcase) from marine bacterium *Vibrio harveyi*. The 10<sup>th</sup> International Symposium of the Protein Society of Thailand, Chulabhorn Research Institute Convention Center, Bangkok, Thailand, July 15<sup>th</sup>-17<sup>th</sup>, 2015. P37, *Poster presentation*.

**Piyanat Meekrathok** and Wipa Suginta. Identification of the catalytic base/nucleophile of a GH20  $\beta$ -N-acetylglucosaminidase from *Vibrio harveyi* through chemical rescue of inactive mutants. The 10<sup>th</sup> International Symposium of the Protein Society of Thailand, Chulabhorn Research Institute Convention Center, Bangkok, Thailand, July 15<sup>th</sup>-17<sup>th</sup>, 2015. P36, *Poster presentation*.

**Piyanat Meekrathok**, Marco Bürger, Arthur T. Porfetye, Ingrid R. Vetter and Wipa Suginta. Structure and function of  $\beta$ -N-acetylglucosaminidase (GlcNAcase) from *Vibrio harveyi*. 7<sup>th</sup> Asia Oceania Human Proteome Organization (AOHUPO) Congress and 9<sup>th</sup> International Symposium of the Protein Society of Thailand “Frontiers in Protein and Proteomic Research”, Miracle Grand Convention Hotel, Bangkok, Thailand, August 6<sup>th</sup>-8<sup>th</sup>, 2014. P106, *Invited oral presentation*.

**Piyanat Meekrathok**, Marco Bürger, Arthur T. Porfetye, Ingrid R. Vetter and Wipa Suginta. Structure and function of  $\beta$ -N-acetylglucosaminidase (GlcNAcase) from *Vibrio harveyi*. 7<sup>th</sup> Asia Oceania Human Proteome Organization (AOHUPO) Congress and 9<sup>th</sup> International Symposium of the Protein Society of Thailand “Frontiers in Protein and Proteomic Research”, Miracle Grand Convention Hotel, Bangkok, Thailand, August 6<sup>th</sup>-8<sup>th</sup>, 2014. P106, *Poster presentation*.

**Piyanat Meekrathok**, Marco Bürger, Arthur T. Porfetye, Ingrid R. Vetter and Wipa Suginta. Structures and functions of  $\beta$ -N-acetylglucosaminidase (GlcNAcase) from *Vibrio harveyi*. RGJ-Ph.D. Congress XV “Towards Translational Research: Publications to Products, P2P”, Jomtien Palm Beach Resort Pattaya, Chonburi, Thailand, May 28<sup>th</sup>-30<sup>th</sup>, 2014. P185, *Oral presentation*.

Kanokwan Lowhalidanon, **Piyanat Meekrathok**, Sunisa Thongsom, Wipa Suginta and Panida Khunkaewla. Human acidic mammalian chitinase (AMCase): molecular cloning, protein expression, and production of monoclonal antibody. The 8<sup>th</sup> International Symposium of the Protein Society of Thailand, Chulabhorn Research Institute Convention Center, Bangkok, Thailand, August 5<sup>th</sup>-7<sup>th</sup>, 2013. P84, *Poster presentation*.

**Piyanat Meekrathok**, Marco Bürger, Arthur T. Porfetye, Ingrid R. Vetter and Wipa Suginta. Three-dimensional structures of  $\beta$ -N-acetylglucosaminidase (GlcNAcase) from *Vibrio harveyi*. The 8<sup>th</sup> International Symposium of the Protein Society of Thailand, Chulabhorn Research Institute Convention Center, Bangkok, Thailand, August 5<sup>th</sup>-7<sup>th</sup>, 2013. P54, *Poster presentation*.

**Piyanat Meekrathok**, Wipa Suginta, Predrag Kukic and Jens Erik Nielsen. Predicting the active site pK<sub>a</sub> values and pH-activity profiles of the GH-18 chitinases. The 5<sup>th</sup> Annual Symposium of Protein Society of Thailand “From Basic Approachs to Modern Technologies”, Chulabhorn Research Institute Conference Center, Bangkok, Thailand, June 23<sup>rd</sup>-25<sup>th</sup>, 2010. P49, *Poster presentation*.

# CURRICULUM VITAE

**NAME:** Mr. Piyanat Meekrathok

**DATE OF BIRTH:** 31 October 1985

**PLACE OF BIRTH:** Nakhon Ratchasima, Thailand

## EDUCATION:

- Academic year 2007, Bachelor of Science (Crop Production Technology)  
First Class Honors (Gold Medal), Suranaree University of Technology,  
Nakhon Ratchasima, Thailand
- Academic year 2007 Start as Ph.D. student, Doctor of Philosophy  
(Biochemistry), Suranaree University of Technology, Nakhon Ratchasima,  
Thailand

## SCHOLARSHIPS AND AWARDS:

- The Royal Golden Jubilee Ph.D. Program scholarship, from the Thailand  
Research Fund
- Deutscher Akademischer Austausch Dienst e.V. German Academic Exchange  
Service (DAAD scholarship), Germany
- Honorable mention poster award at 7<sup>th</sup> Asia Oceania Human Proteome  
Organization (AOHUPO) Congress and 9<sup>th</sup> International Symposium of the  
Protein Society of Thailand, Bangkok, Thailand