

ÉCOLE DOCTORALE des Sciences de la Vie et de la Santé

IGBMC, UMR7104, U964, Illkirch

THÈSE présentée par :

Firas FADEL

Soutenue le : **13 octobre 2014**

pour obtenir le grade de : **Docteur de l'Université de Strasbourg**

Discipline/ Spécialité : Biophysique et Biologie Structurale

**High resolution structural and mechanistic
study of human chitotriosidase (CHIT1)**

THÈSE dirigée par :

Dr. PODJARNY Alberto

DR, CNRS, IGBMC, Illkirch, France

RAPPORTEURS :

Dr. MOUREY Lionel

DR, UMR 5089, Toulouse, France.

Dr. MORERA Solange

DR, LEBS, Gif-sur-Yvette, France

Prof. WESTHOF Eric

Professor, CNRS, IBMC, Strasbourg, France

INVITES :

Dr. LAMOUR Valérie

DR, CNRS, IGBMC, Illkirch, France

Dr. CACHAU Raul

DR, National Cancer Institute, Frederick, USA

Table of Contents

TABLE OF CONTENTS	1
ACKNOWLEDGMENTS	5
RESUME DE THESE	6
INTRODUCTION	6
OBJECTIFS	8
RÉSULTATS	8
CONCLUSIONS	11
PERSPECTIVES	12
INTRODUCTION	17
1.1 CHITIN AND CHITINASES	17
1.1.1 Classification of chitinases	19
1.1.2 Role of GH18 chitinases in lower organisms	24
1.1.3 MAMMALIAN CHITINASES	27
1.1.3.1 Phylogenetic analyses of mammalian GH18 members	27
1.1.3.2 Human and murine GH18 members	29
1.1.3.3 Role of CHIT1 activation in diseases	37
1.2 STRUCTURAL AND MECHANISTIC STUDIES OF GH18 CHITINASES	40
1.2.1 STRUCTURAL ANALYSIS OF THE CATALYTIC DOMAIN OF GH18 CHITINASES	40
1.2.2 ROLE OF THE CONSERVED AROMATIC RESIDUES IN THE SUBSTRATE BINDING FEATURE OF CHIT1 AND ITS COMPARISON WITHIN GH18 CHITINASES	44
1.2.3 ROLE OF CONSERVED RESIDUES WITHIN THE ACTIVE SITE OF GH18 CHITINASES IN THE HYDROLYSIS MECHANISM	48
1.2.3.1 Inverting mechanism vs retaining mechanism	48
1.2.3.2 Substrate-assisted chitinolytic mechanism	50
1.2.3.3 The coordinated role of a conserved catalytic triad in GH18 chitinases	52
1.2.3.4 Transglycosylation	55

1.2.4 PROCESSIVITY IN GH18 CHITINASE.	59
1.3 THE DIFFERENT MODULES IN GH18 CHITINASES	62
1.3.1 CBMs in bacterial chitinases	63
1.3.2 Human ChBDs belong to same family of CBM found in invertebrates	65
1.4 STRUCTURAL-BASED INHIBITION OF GH18 CHITINASES.	67
1.4 Allosamidin, a natural compound inhibitor of GH18 proteins, and its derivatives	67
2. AIMS AND OBJECTIVES OF MY PHD PROJECT	75
3. MATERIALS AND METHODS	77
3.1 PLASMID	77
3.2 CLONING	77
3.2.1 DNA sequencing	79
3.2.2 DNA purification	80
3.3 PROTEIN PRODUCTION	80
3.3.1 HEK293T cell maintenance protocol	80
3.3.2 Transferring HEK293T cells from flasks to roller bottles	80
3.3.3 Small scale transfection and expression test via Western blot	81
3.3.4 Transfection in roller bottles	81
3.4 PURIFICATION PROTOCOL	82
3.5 ENZYME ACTIVITY MEASUREMENTS	82
3.6 IC ₅₀	82
3.7 CRYSTALLOGRAPHIC METHODS	83
3.7.1 Crystallogenesis	83
3.7.2 Crystallization of CHIT1 catalytic domain	86
3.7.3 Co-crystallization of CHIT1 catalytic domain with substrate	87
3.7.4 Crystal soaking of CHIT1 catalytic domain with four inhibitor compounds	88
3.7.5 Crystallization of CHIT1-FL is mentioned in page (127-128)	88
3.8 X-RAY DATA COLLECTION AND PROCESSING	88
3.4 ELECTROSPRAY IONIZATION MASS SPECTROMETRY (ESI-MS) (SEE PAGE 129)	89
3.5 STRUCTURAL CONSERVATION ANALYSIS (SEE PAGE 130)	89

4. RESULTS **90**

NEW INSIGHTS IN THE ENZYMATIC MECHANISM OF HUMAN CHITOTRIOSIDASE (CHIT1) CATALYTIC DOMAIN BY ATOMIC X-RAY AND HYBRID QM/MM **91**

ABSTRACT	91
1-INTRODUCTION	92
2. MATERIALS AND METHODS	95
2.1 Cloning, expression and purification	95
2.2 Enzyme Activity Measurements	95
2.3 Crystallization and data collection	96
2.3 Quantum mechanics/molecular mechanics (QM/MM) calculation method	97
3. RESULTS AND DISCUSSION	98
3.1 Atomic resolution structures of the catalytic domain of CHIT1 reveal double conformation of key catalytic residues	99
3.2 Protonation state of the catalytic triad residues of CHIT1 apo form provides insight into its hydrolysis mechanism	101
3.3 Atomic resolution structures of the catalytic domain of CHIT1 in complex with chitobiose provide insight into catalytic mechanism	104
3.4 Structural analysis of the catalytic triad residues in the CHIT1-chitobiose structure reveals the coexistence of two enzymatic states in the same crystal form	106
3.5 Detailed structural analysis sheds new light into the hydrolytic step	109
3.6 A new scenario for the processive hydrolysis	110
3.7 Detection of product in the catalytic groove reveals insight into the transglycosylation mechanism	111
4. CONCLUSION	112
ACKNOWLEDGEMENTS	113
FIGURE LEGENDS	114

X-RAY CRYSTAL STRUCTURE OF THE FULL-LENGTH HUMAN CHITOTRIOSIDASE (CHIT1) REVEALS FEATURES OF ITS CHITIN-BINDING DOMAIN **123**

ABSTRACT	123
-----------------	------------

1. INTRODUCTION	124
2. MATERIALS AND METHODS	126
2.1 CLONING, EXPRESSION AND PURIFICATION	126
2.2 CRYSTALLIZATION, CROSS-SEEDING AND MICRO-SEEDING	127
2.3. CRYO-COOLING, DATA COLLECTION AND MOLECULAR REPLACEMENT	128
2.4 ELECTROSPRAY IONIZATION MASS SPECTROMETRY (ESI-MS)	129
2.5 STRUCTURAL CONSERVATION ANALYSIS	130
3. RESULTS AND DISCUSSION	130
3.1. CRYSTALLIZATION OF CHIT1-FULL LENGTH (CHIT1-FL)	130
3.2. ANALYSIS OF THE CRYSTAL CONTACTS AND PACKING	131
3.3 OVERALL STRUCTURE OF CHBD _{CHIT1}	134
3.4 EVOLUTIONARY ANALYSIS OF CONSERVED CYSTEINE RESIDUES IN CHBDS	135
3.5 ANALYSIS OF THE EVOLUTIONARY CONSERVED AROMATIC RINGS	136
3.6 CHBD _{CHIT1} PRESENTS A POSITIVELY CHARGED FACE	138
3.7 HIGH SIMILARITY BETWEEN CHBD _{CHIT1} AND CHBD _{AMCASE}	139
4. CONCLUSION	140
ACKNOWLEDGEMENTS	141
FIGURE LEGENDS	142
GENERAL CONCLUSIONS AND PERSPECTIVES	157
4.1 CONCLUSIONS AND PERSPECTIVES REGARDING THE STRUCTURAL AND MECHANISTIC STUDY OF CHIT1 CATALYTIC DOMAIN.	157
4.2 CONCLUSIONS AND FUTURE WORK ON THE CHIT1-FL.	159
APPENDIX1	160
TABLE 1. DATA COLLECTION AND REFINEMENT STATISTICS	161
APPENDIX 2	162
APPENDIX: 3	164
REFERENCES	167

Acknowledgments

I would like to thank Dr. Alberto Podjarny for giving me the confidence to conduct an interesting and challenging project that allowed me to develop a wide range of research skills.

I wish to acknowledge Dr. André Mistchter for teaching me X-ray data processing and for showing high perseverance to reach our aims.

I would like to express my gratitude to Alexandra Cousido-Siah, for her support, her rigorous corrections and especially for showing a high level of responsibility.

I wish to acknowledge Dr. Xavier Ruiz Figueras for giving me moral supports and for being available to make constructive corrections in my manuscript.

I am very grateful to Dr. Yuguang Zhao for his generosity and for teaching me protein expression in human cells.

I would like to acknowledge Dr. Raul Cachau for performing QM/MM calculations.

Many thanks to Dr. Valérie Lamour, Dr. Eduardo Howard and Dr. Claire Batisse for their advices.

I wish to acknowledge Dr. Alexey Rozov for being available for thesis corrections and data processing.

I would like to thank my dearest friends Iskander, Alexey, Heena, Nicolas, Sergey, Anna, Hussein, Irina, Xavier, Ahmad, Katarzyna, Gregory, Justine, Mélanie, Karima, Claire, Vova, Simone et Perrine.

To Irene I would like to express my gratitude for showing love, patience and understanding during hard episodes while I was undertaking this work. Without your support, help, advices and care I could not have achieved it.

To my mother I would like to thank you for your endless love and support throughout my life and for giving me strength to reach goals and chase my dreams.

To my sister, you also deserve my wholehearted. I am remembering when we were kids, our difficult childhood and how we overcome all that. I will still your byebabay.

To my father, I think you would have been happy.

Résumé de thèse

Introduction

La chitine, un polymère de β -1,4-N-acétylglucosamine, est synthétisée pour des fins structurales car elle est un constituant majeur de l'exosquelette des insectes, des carapaces des crustacés, de la paroi cellulaire des champignons et des algues. Dans la nature, il existe un équilibre entre la biosynthèse et la dégradation de la chitine. Cette dernière est assurée par des enzymes chitinolytiques parmi lesquelles se trouvent les chitinases, enzymes appartenant à la famille des glycosyl hydrolases 18 (GH18) et 19 (GH19). Elles hydrolysent la chitine en clivant les ponts β (1,4)-glycosidiques. Les chitinases ont été identifiées dans de nombreux organismes, des bactéries aux êtres humains, où elles jouent des rôles physiologiques différents selon l'espèce qui les produit. En effet, chez les bactéries, les chitinases fournissent les sources de carbone et d'azote en dégradant la chitine. Elles sont impliquées dans la morphogénèse et la croissance des champignons et des insectes. Chez les plantes et les mammifères, ces protéines jouent un rôle protecteur contre des pathogènes contenant de la chitine.

Au cours de la dernière décennie, plusieurs structures cristallines des protéines de la famille GH18 ont été résolues. La comparaison de ces structures montre que les domaines catalytiques de cette famille d'enzymes partagent une architecture tridimensionnelle conservée de type tonneau (α/β)₈. Selon les données structurales, le domaine catalytique est caractérisé par une longue cavité comprenant des résidus aromatiques qui contribuent à la fixation du substrat. Le site actif est situé au fond de cette cavité et est constitué d'une triade catalytique contenue dans le motif consensus conservé DXDXXE. Le modèle proposé de l'hydrolyse de la chitine est basé sur le rôle clé du glutamate de la triade catalytique qui est le donneur du proton conduisant au clivage de la liaison glycosidique du substrat. En plus de la réaction d'hydrolyse, plusieurs chitinases, y compris celles exprimées chez l'homme, sont capables de catalyser la formation de nouvelles liaisons glycosidiques entre des sucres donneurs et accepteurs aboutissant à la re-polymérisation du substrat. Cependant, le mécanisme exact de ce changement réactionnel, appelé transglycosylation, n'a pas encore été clairement élucidé.

Récemment, des efforts considérables ont été fournis afin de comprendre la relation structure-fonction des chitinases de la famille GH18 par combinaison des données de diffraction des rayons X, des expériences biochimiques et des calculs bioinformatiques. Cependant, malgré ces travaux, les détails structuraux du mode d'action et du mécanisme de ces enzymes ne sont pas complètement compris et plusieurs aspects restent controversés.

Au cours de mon projet de thèse, j'ai étudié le domaine catalytique de la chitotriosidase humaine (CHIT1) en tant que modèle du mécanisme catalytique de la famille GH18. Cette enzyme est l'une des deux chitinases actives identifiées chez l'homme. Ayant une masse moléculaire de 50 kDa, elle est composée d'un domaine catalytique (CAT) de 39 kDa relié par une région charnière au domaine de fixation à la chitine. Bien que le rôle physiologique de CHIT1 ne soit toujours pas complètement compris, il a été lié à la réponse immunitaire innée. Par ailleurs, une surexpression de cette protéine a été mise en évidence dans plusieurs maladies, notamment dans la maladie de Gaucher où elle est considérée comme son principal bio-marqueur, ainsi que dans la polypose nasale où elle est surexprimée au niveau de l'épithélium nasale. En conséquence, la résolution de la structure de CHIT1 est essentielle, pour d'une part, permettre de mieux comprendre la relation entre son repliement 3D et son rôle physiopathologique et d'autre part, de contribuer au développement des traitements thérapeutiques contre les maladies dans lesquels elle est impliquée. Dans ce contexte, Fusetti et ses collaborateurs ont résolu en 2002 la première structure cristalline du domaine catalytique de CHIT1 à 2,3 Å. Cependant, la structure complète de cette protéine n'a pas été déterminée. L'analyse de cette structure montre qu'elle adopte un repliement 3D similaire à celui des autres domaines catalytiques des chitinases des autres organismes ayant le motif conservé (DXDXE) situé au niveau du site actif. Bien que les structures de plusieurs complexes CHIT1-inhibiteurs ont été obtenues les années suivantes, la résolution de ces structures y compris de la forme apo n'a pas permis de déterminer l'état d'ionisation et de protonation du site actif.

Objectifs

Au vu des études déjà publiées et afin d'approfondir nos connaissances sur le mécanisme d'action de CHIT1 ainsi que la famille d'enzyme GH18, les objectifs de ma thèse sont les suivants :

- Optimiser la cristallisation de CHIT1 afin d'obtenir des données de diffraction des rayons X à très haute résolution.
- Etudier l'état de protonation du site actif de CHIT1 et comprendre le mécanisme de transfert du proton au cours de la réaction catalytique de CHIT1.
- Cristalliser et résoudre la structure cristalline de CHIT1 complète comportant le domaine catalytique et le domaine de fixation de la chitine (ChBD).

Résultats

Dans un premier temps, j'ai produit la protéine CHIT1 sécrétée dans les cellules HEK293T (Human Embryonic Kidney 293) et l'ai ensuite purifiée par chromatographie d'affinité (IMAC) et chromatographie d'exclusion. Dans le but d'améliorer la résolution des données de diffraction des rayons X, j'ai déterminé de nouvelles conditions de cristallisation pour cette protéine. Suite à l'optimisation de la cristallisation par des expériences de micro-ensemencement, j'ai résolu par remplacement moléculaire une nouvelle structure de la forme apo de CHIT1 à 1.0 Å et une structure pseudo-apo à 0.95 Å (Fig. 1A, B). La structure pseudo-apo correspond à une structure où la protéine a été cristallisée avec la chitine mais où aucune densité électronique de ce polymère n'a été observée. Ces structures montrent que l'aspartate 138 (D138), qui se situe au milieu de la triade catalytique, présente deux conformations, suggérant que ce résidu possède une barrière énergétique similaire pour adopter chaque conformation. Dans la conformation A (D138-confA), le D138 est orienté vers l'aspartate 136 (D136) et dans la conformation B (D138-confB), il est orienté vers le glutamate catalytique E140. Par ailleurs, deux conformations de E140 ont été détectées : la première présente une légère rotation de E140, appelée par convention E140-confA et la deuxième est un rotamère, nommé E140-confB. Les deux conformations de D138 et E140

ont été corrélées de la façon suivante : quand D138 tourne vers D136, E140-confA se forme, tandis que la rotation de D138 vers E140 donne lieu à la formation de E140-confB. L'analyse des différentes conformations dans le site actif de CHIT1 révèle également que les molécules d'eau qui sont en contact avec D138 et E140 subissent également des changements de position dû au mouvement de ces deux résidus. La présence des deux conformations de D138 et de E140 dans la forme apo n'a pas été visualisée dans la structure publiée de CHIT1 à 2.3 Å de résolution. En revanche, les deux conformations de D138 ont été observées dans plusieurs structures de chitinases apo provenant d'autres organismes. Ceci suggère que la triade catalytique de la famille GH18 a conservé son arrangement conformationnel et son profil énergétique qui privilégient une rotation de son aspartate central.

Par la suite, je me suis intéressé à l'état de protonation du site actif de CHIT1. J'ai réalisé des affinements avec le programme SHELXL permettant de déterminer les distances des liaisons carbone-oxygène des carboxylates selon le pic de la densité électronique des résidus de la structure apo et pseudo-apo. Les valeurs de ces distances m'ont permis d'établir une relation avec l'état de protonation du site actif (Fig. 1C). En parallèle, des calculs quantiques ont été réalisés afin de compléter et de valider les données cristallographiques dans le cadre d'une collaboration avec Raul Cachau du Advanced Biomedical Computer Center aux Etats-Unis. Les résultats obtenus grâce à la combinaison des deux approches, cristallographique et quantique, montrent que D138 partage un proton avec D136 quand il est orienté vers ce dernier qui à son tour partage un proton avec la tyrosine Y27. Ces résultats révèlent que D138 partage également un proton quand il tourne vers le résidu catalytique clé E140. En outre, l'analyse du résidu catalytique E140 indique que l'oxygène externe de sa chaîne latérale n'est pas protoné. Cette observation n'est pas cohérente avec le mécanisme proposé : l'oxygène externe de E140 est protoné afin de pouvoir hydrolyser le substrat. Alors, pour approfondir la compréhension de l'effet du substrat sur la triade catalytique, j'ai co-cristallisé la CHIT1 avec trois concentrations différentes de substrat synthétique 4-méthylumbelliféryl β -N,N',N''-triacétylchitotrioside (4-MU-NAG₃) comportant trois sous unités de N-acétyl glycosamine (NAG). Les trois structures obtenues avec les concentrations finales de substrat 0.3 mM, 1 mM et 2.5 mM ont été résolues à 1.10, 1.05 et 1.10 Å de résolution respectivement. Ces structures montrent une

hydrolyse du substrat en deux sous-unités (NAG₂) appelés chitobiose, et une augmentation de l'occupation de celles-ci dans le site catalytique en fonction de la concentration de 4-MU-NAG₃ (Fig. 2B). Cette augmentation de l'occupation est corrélée avec une diminution graduelle de la conformation de E140-confA (Fig. 2A, C). De plus, nos résultats suggèrent que D138-confB transfère un proton à E140 suivi par une rotation qui serait nécessaire pour que le carboxylate du résidu E140 puisse avoir accès à l'oxygène de la liaison glycosidique lors de l'hydrolyse pour transférer le proton. Il semblerait donc qu'une rotation de E140 ait eu lieu pendant la réaction d'hydrolyse du substrat avant que ce résidu prenne une conformation plane après la fin de cette réaction. L'analyse du site actif du complexe CHIT1-chitobiose obtenue avec la concentration de 2.5 mM de substrat montre qu'il reflète au plus près l'état après hydrolyse par rapport aux deux autres structures puisque l'occupation de la chitobiose (80%) est la plus élevée (Fig. 2B, C). De plus, D138 et E140 adoptent quasi-uniquement la conformation B. Les résultats cristallographiques et de calculs quantiques concernant l'état après hydrolyse indiquent que E140-confB devient chargé ce qui est cohérent avec le fait que ce résidu a transféré un proton pour couper la liaison glycosique du substrat. En outre, les données obtenues montrent que D138-confB devient protoné après l'hydrolyse et qu'il forme une liaison ionique forte avec E140. Cela signifie que D138 a reçu un nouveau proton de D136 et Y27 lors de la rotation vers ces derniers qui partageaient eux-mêmes un proton. Ces données ont permis de proposer un nouveau mécanisme d'hydrolyse du substrat. Selon ce mécanisme, dans la forme apo de CHIT1 le proton est stocké entre E140, D138 et D136, maintenu par le changement de conformation de D138 et le changement de position de molécules d'eau (Fig. 3A). Au moment de la fixation du substrat, grâce à un changement de pKa, D138 transfère un proton à E140. Ce dernier subit une rotation pour avoir accès à l'oxygène de la liaison glycosidique et le protoner, ce qui aboutit à la coupure du substrat (Fig. 3B, C, D). Simultanément, D138 tourne vers D136 pour récupérer un nouveau proton, partagé entre D136-Y27 puis il réalise une seconde rotation vers E140 (Fig. 3B, C, D). A ce stade, D138, en conformation B reste protoné et stabilisé par une liaison hydrogène ionique forte avec E140 et une liaison hydrogène avec le groupe N-acetyl du substrat (Fig. 3E). Ceci permet à E140 de recevoir un nouveau proton en présence d'une nouvelle liaison glycosidique à travers D138.

La deuxième partie de mon projet a eu pour but de déterminer la structure cristalline de CHIT1 complète comportant une charnière flexible. En raison de la grande flexibilité de la protéine, les premiers essais de cristallisation ont échoué. J'ai alors opté pour une stratégie consistant à ensemercer les gouttes contenant la CHIT1 complète avec des semences provenant de cristaux du domaine catalytique de CHIT1 obtenus auparavant. Ceci m'a permis d'obtenir les premiers cristaux de la CHIT1 complète. En réalisant de nouveaux ensemcements, des cristaux avec un nouveau groupe d'espace ont été obtenus. Ce groupe d'espace a permis un empilement dans le cristal propre à la protéine entière, ce qui a abouti à stabiliser le domaine ChBD (Fig. 4A). L'amélioration de ces cristaux a permis de résoudre la structure de la forme complète de cette enzyme à 1.9 Å. Cependant, en raison de la flexibilité élevée de la région charnière, la densité électronique qui correspond à cette zone n'a pas pu être visualisée. Cette structure montre que le repliement 3D du domaine ChBD consiste essentiellement en des feuillets β anti-parallèles repartis en triplet et doublet (Fig. 4B, C). La comparaison du domaine ChBD avec celui présent chez les invertébrés montre une conservation de ce repliement. Ce domaine contenant 49 résidus révèle la présence de 3 ponts disulfures formés par 6 cystéines hautement conservées. La position des résidus aromatiques est également hautement conservée ce qui suggère que le tryptophane W465 interagit avec la chitine. La structure du domaine ChBD montre qu'il est caractérisé par une surface positivement chargée.

Conclusions

- Pendant mon travail de thèse, j'ai réussi à optimiser les conditions de cristallisation du domaine catalytique de CHIT1, ce qui a permis d'obtenir des données de diffraction des rayons X à une résolution atomique. Une telle résolution étant atteinte pour la première fois pour cette enzyme et pour un membre de la famille GH18 a permis de découvrir de nouveaux éléments structuraux. Ces éléments ont contribué à l'amélioration de l'analyse de l'état de protonation du site actif et les arrangements de conformation du site actif ainsi que de proposer un nouveau mécanisme de la réaction d'hydrolyse du substrat.

- La stratégie adoptée pour cristalliser la forme complète de CHIT1 a abouti à la détermination d'une nouvelle structure comprenant son domaine ChBD permettant

d'analyser les caractéristiques structurales de ce domaine et de mettre en évidence la nature hautement conservée de son repliement 3D des invertébrés jusqu'à l'homme.

Perspectives

- Déterminer la structure cristalline à haute résolution de CHIT1 avec un analogue de l'état intermédiaire du substrat.

- Obtenir une structure de CHIT1 par cristallographie des neutrons afin de confirmer les données obtenues par diffraction de rayons X concernant l'état de protonation du site actif.

- Réaliser des calculs quantiques à partir des structures de CHIT1 et CHIT1-chitobiose pour modéliser la réaction d'hydrolyse.

- Réaliser des expériences de SAXS avec CHIT1 complète afin de déterminer l'ensemble de ces conformations en solution.

- Modéliser l'interaction du domaine ChBD en complexe avec la chitine.

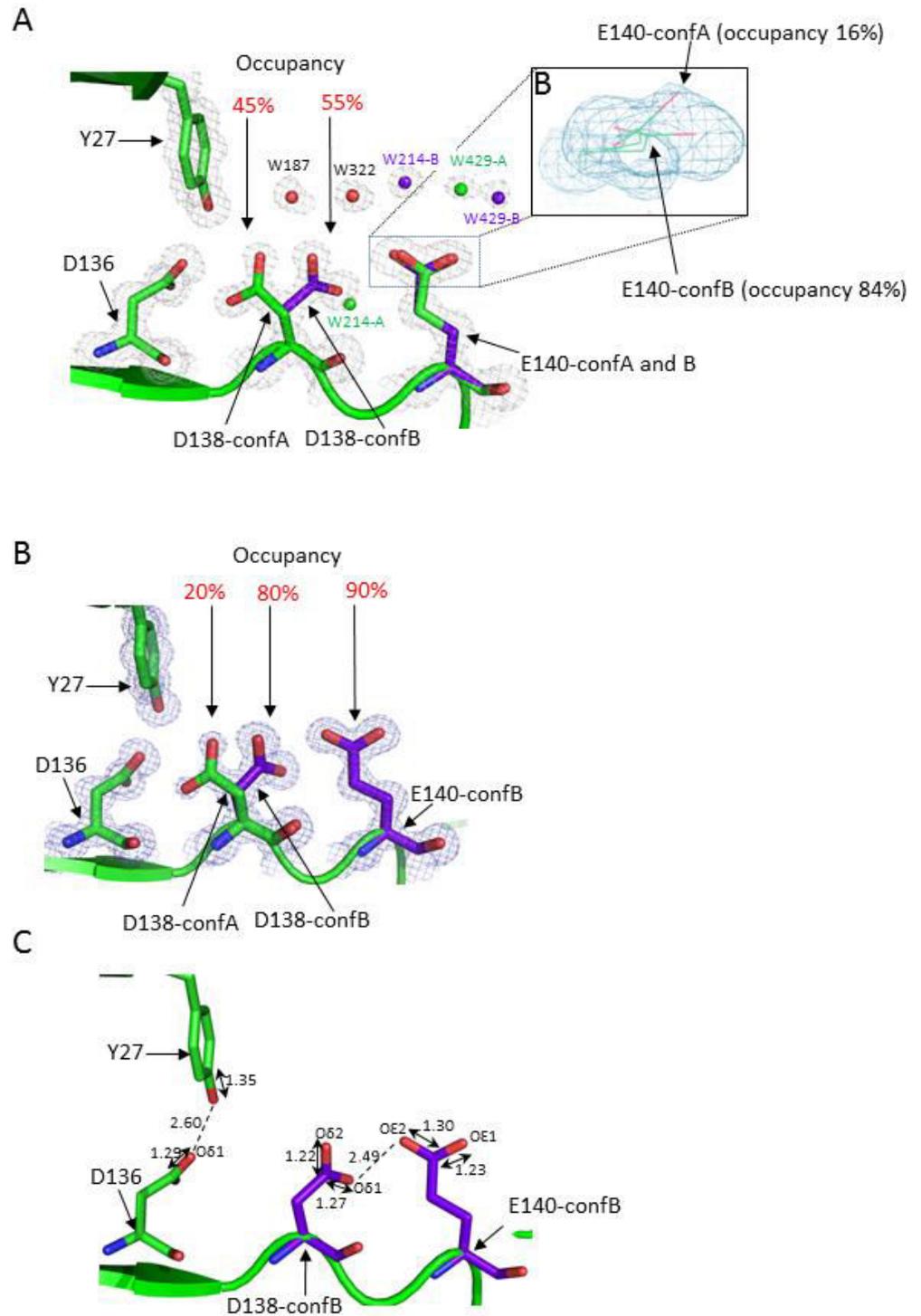


Figure 1. Structures apo et pseudo-apo de CHIT1. A) Site actif de CHIT1 en forme apo. Densité électronique $2F_o - F_c$ des résidus Y27, D136, D138, E140 en gris. Résidus Y27, D136, D138-confA, E140-confA en vert, résidus D138-confB et E140B en violet. Les occupations des deux conformations de D138 et E140 sont mentionnées. B) Site actif de la structure pseudo-apo de CHIT1 montrant la densité électronique et les différentes occupations des mêmes résidus mentionnés en A. C) Site actif de la structure pseudo-apo de CHIT1 montrant les valeurs des longueurs des liaisons C-O des chaînes latérales des résidus Y27, D136, D138, E140 obtenues avec le programme SHELXL.

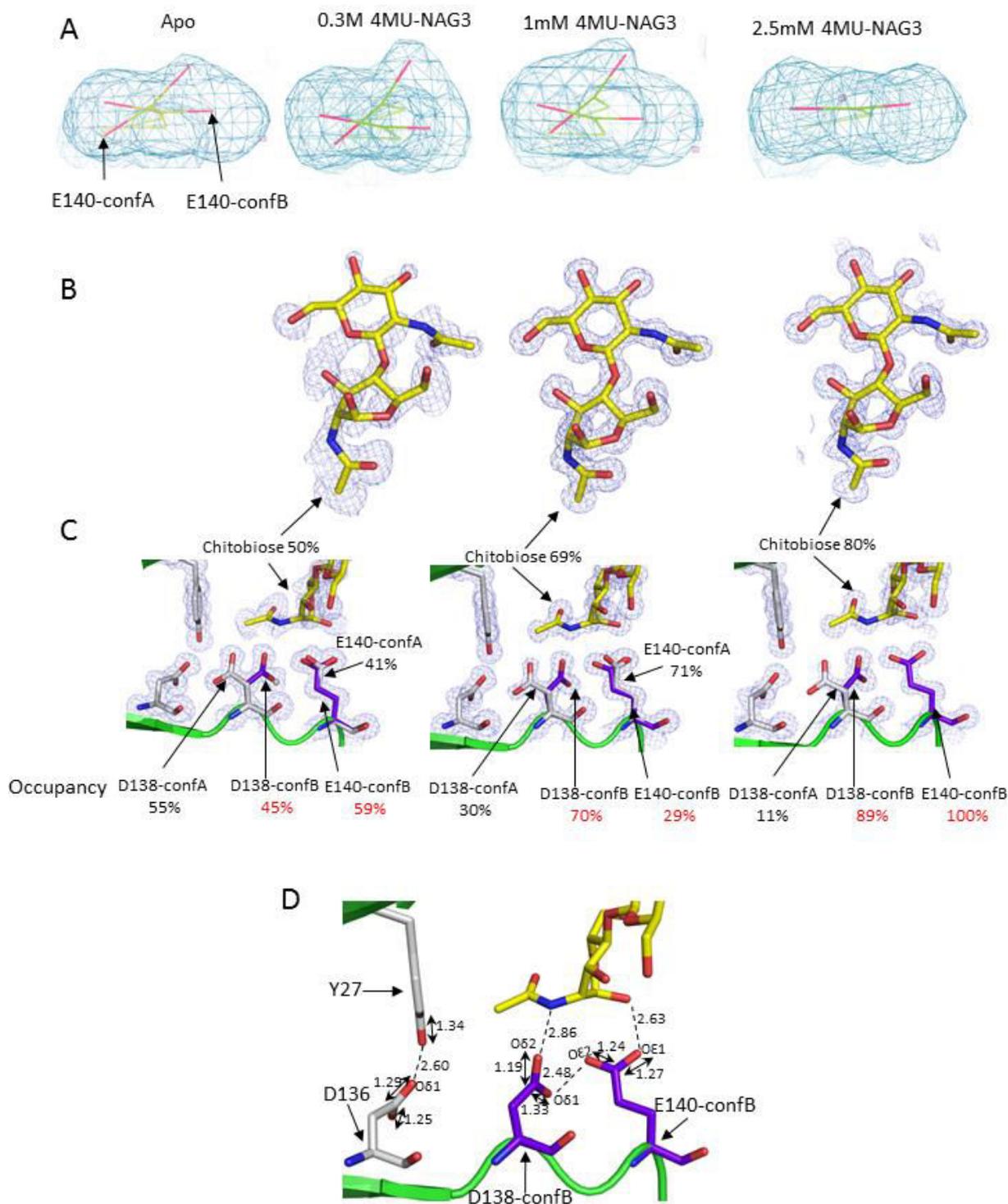


Figure 2. Structures de CHIT1 en complexe avec la chitobiose obtenue par co-cristallisation avec le substrat synthétique 4-MU-NAG₃. A) Densité électronique 2F_o-F_c de E140 montrant les doubles conformations de ce résidu. B) Densité électronique 2F_o-F_c de La chitobiose ou (NAG₂) dans les trois structures cristallines du complexe CHIT1-chitobiose. C) Site actif de CHIT1 avec les occupations des doubles conformations observées pour D138 et E140. D) Structure du site actif de CHIT1 en présence de la chitobiose obtenue avec une concentration de 2.5 mM de 4-MU-NAG₃ montrant les valeurs des longueurs des liaisons C-O obtenues avec le programme SHELXL.

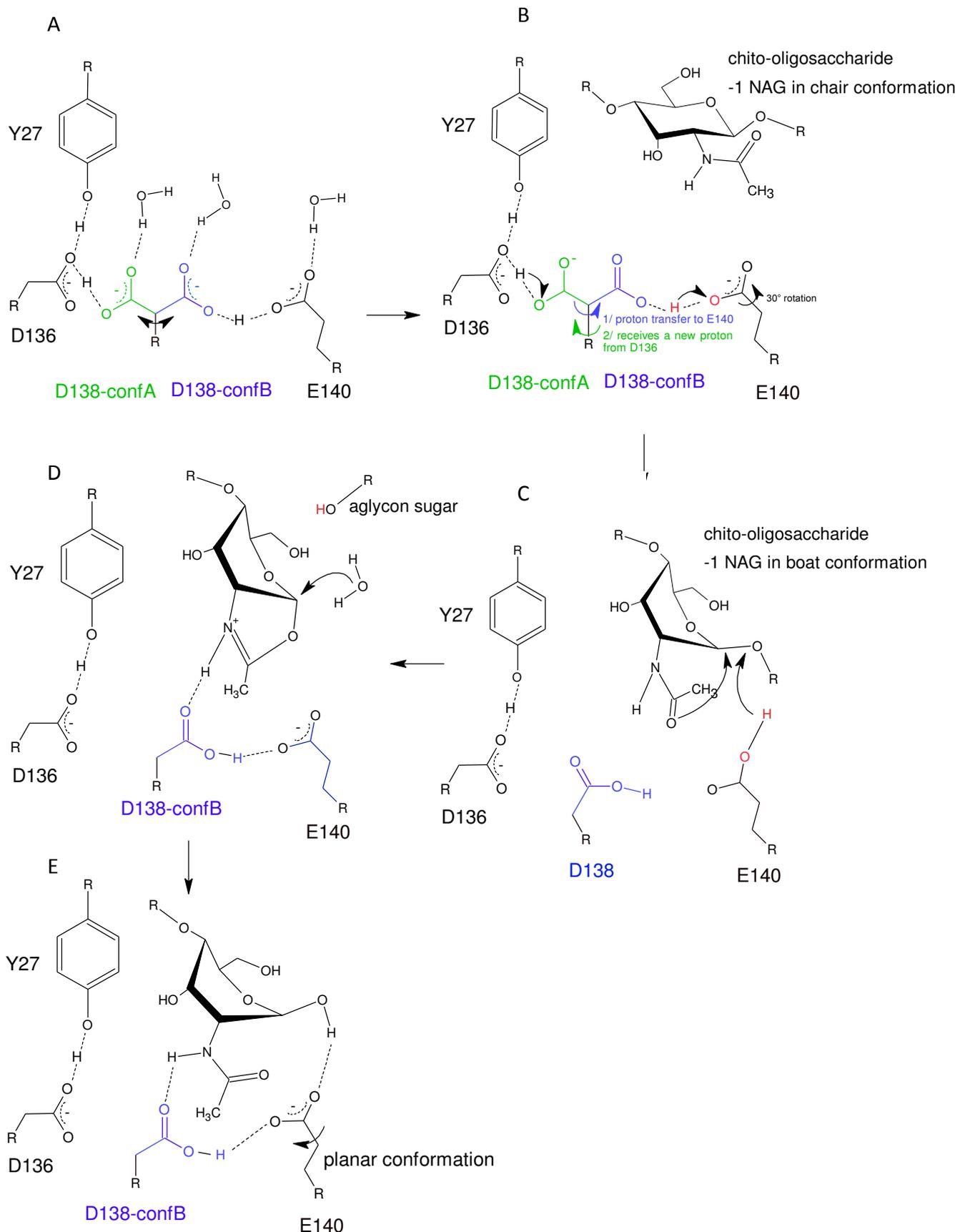


Figure 3. Mécanisme proposé pour l'hydrolyse des chito-oligosaccharides. A) Stockage du proton dans le site actif de CHIT1 en forme apo. B) L'arrivée du substrat et le transfert du proton de D138 à E140 suivi par une rotation de D138 vers D136-Y27 pour récupérer un nouveau proton. C) Protonation du substrat par E140. D) Coupure du substrat, formation de l'intermédiaire et activation de la molécule d'eau qui subit à son tour une attaque nucléophile sur l'intermédiaire. E) Etape après hydrolyse et la reformation de la sous-unité NAG en position -1.

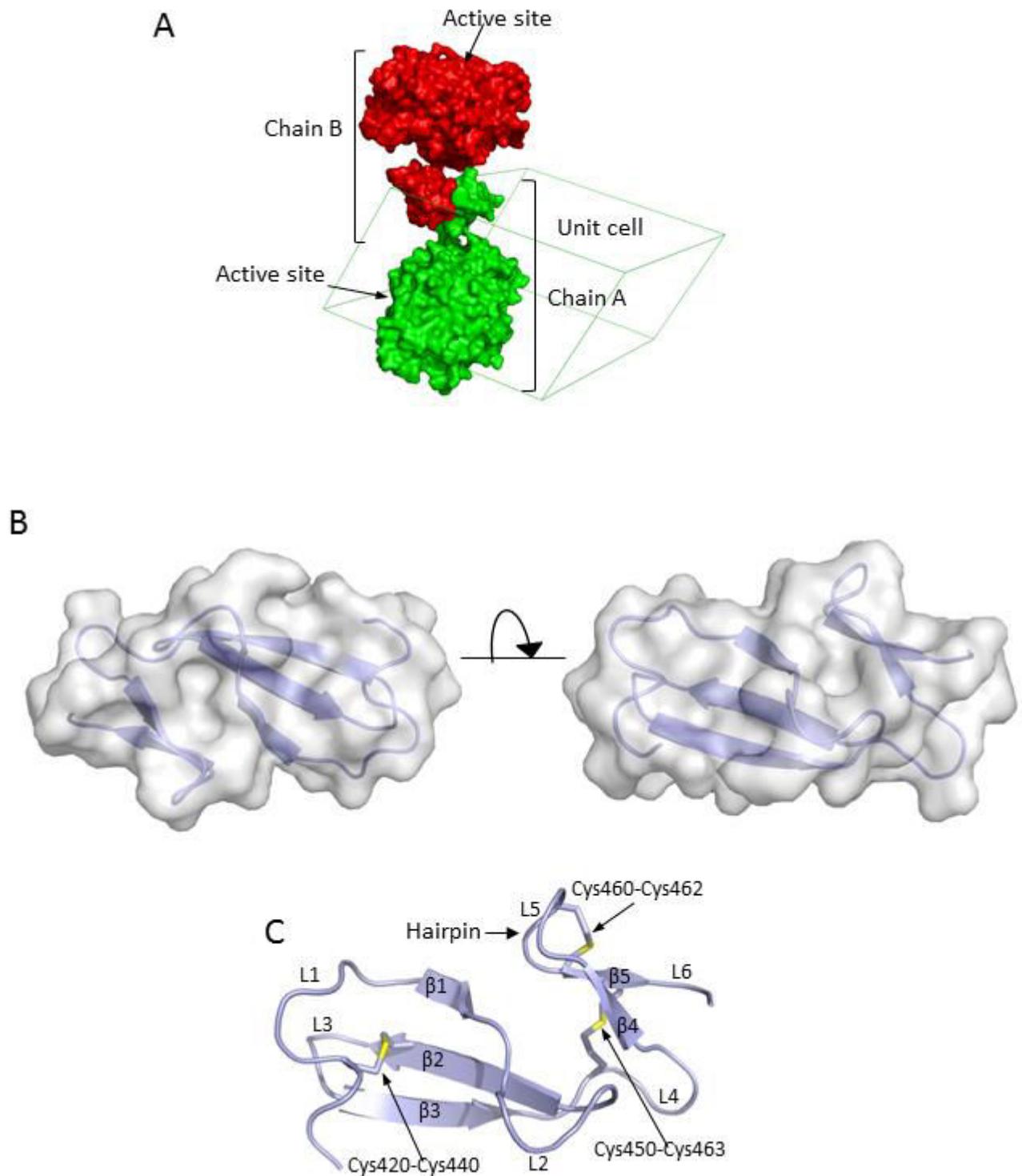


Figure 4. Structure de CHIT1 complète. A) Structure de CHIT1 complète dans l'unité asymétrique. B) Structure du domaine ChBD et sa surface moléculaire. C) Structure du domaine ChBD montrant les trois ponts disulfures.

Introduction

1.1 Chitin and chitinases

The word chitin is derived from the Greek word χιτών (chiton), meaning “envelope” and came to light in 1811 (Ruiz-Herrera, 1978). Chitin, a β -1,4-linked polymer of N-acetylglucosamine (NAG), is the second most abundant natural and insoluble polysaccharide after cellulose (Fig. 1). Each year, approximately 10 gigatons (1×10^{13} kg) of chitin, are produced and degraded in the biosphere. Chitin polymers have the tendency to form microfibrils (also referred to as rods or crystallites) of ~ 3 nm in diameter that are stabilized by hydrogen bonds formed between the amine and carbonyl groups (Fig. 1, 2).

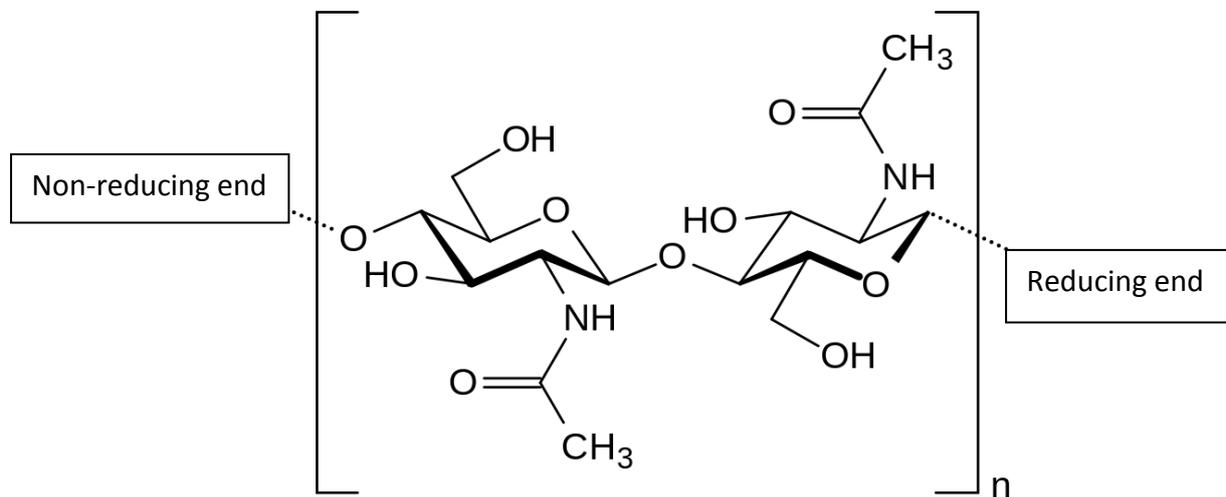


Figure 1. Structure of chitin. Multiple acetyl-glucosamine molecules form long chains via β -1,4 linkages.

The polymorphic forms of chitin vary in packing and polarities affecting its physiological role as well as its physico-chemical properties. Structural analysis suggested the existence of three different crystalline forms of chitin, named α , β and γ (Fig. 2A, B, C). In the α form, all chains are tightly packed and aligned in an anti-parallel orientation; in the β form the chains are arranged in a parallel manner while in the γ form sets of two parallel strands alternate with single antiparallel strands (Fig. 2B). Interestingly, in the β and γ forms the microfibrils are less compact, which results in more contacts with water molecules via

hydrogen bonds leading overall to a relatively more flexible structure (Merzendorfer & Zimoch, 2003).

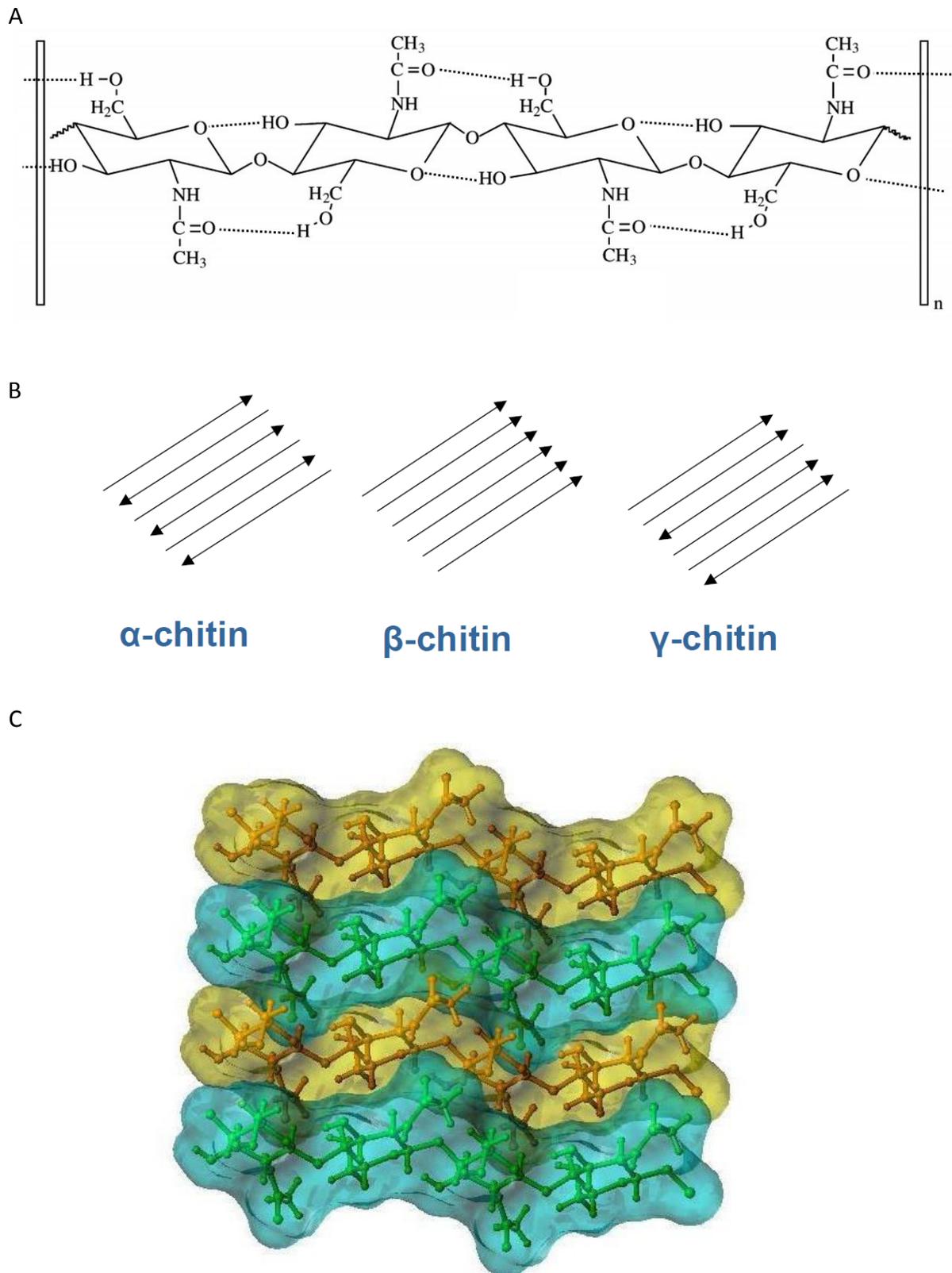


Figure 2. A) Chitin is a homopolymer of N-acetylglucosamine monosaccharide units in $\beta(1\rightarrow4)$ linkages with intra-chain hydrogen bonds. B) Representation of the three highly ordered crystalline structures of chitin α , β and γ . C) 3D representation of the crystal structure of crystalline β -chitin (Kobayashi *et al.*, 2013).

As far as its function is concerned, chitin represents a key structural constituent in a large variety of organisms. Indeed, chitin is widely present in insect exoskeletons, shells of crustaceans, the fungal cell wall and nematode eggs. Moreover, in chitin-containing organisms, the metabolism of NAG chains appears to play an important role in their life cycle, morphogenesis and growth. Interestingly, it was shown that plants and mammals do not synthesize chitin (Bulawa *et al.*, 1995); (Soulie *et al.*, 2006).

In nature, there is a balance between chitin synthesis and degradation. This balance is maintained by two types of enzymes: the chitin synthases, which belong to the family of glycosyl transferases and generate the chitin chain, and the chitinolytic enzymes, which degrade chitin structures. Among the chitinolytic enzymes, chitinases cover a large array of the chitin-degrading enzymes.

1.1.1 Classification of chitinases

Chitinases belong to the superfamily of glycosyl hydrolases (GH) that groups enzymes related to carbohydrate metabolism (Henrissat, 1991). Chitinases have a size range from 20 kDa to about 90 kDa and they can be divided into two major categories according to their catalytic features. These two categories are endochitinases and exochitinases. Endochitinases cleave the glycosidic bonds randomly along the chitin chain, providing a variety of soluble NAG polymers. Exochitinases have been further divided into subcategories: chitobiosidases, which hydrolyze chitin from the non-reducing end generating diacetylchitobiose units, and β -(1,4)-N-acetyl-glucosaminidases (NAGases), which cleave NAG oligomers into NAG monomers (Sahai & Manocha, 1993); (Rai & Bridge, 2009); (Azzouz, 2001).

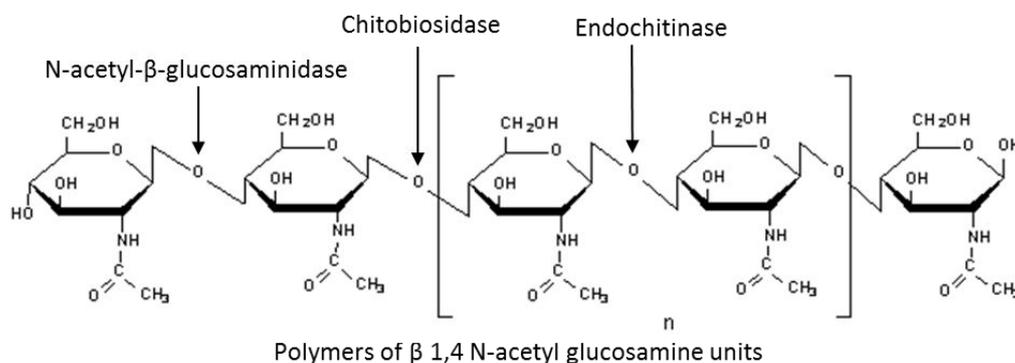


Figure 3. Specificity of endochitinase and exochitinase (chitobiosidase and N-acetyl- β -glucosaminidase) enzymes on the chitin polymer.

Based on the amino acid sequence similarity of chitinases from various organisms, five distinct classes of chitinases have been proposed (Fig. 4).

Class I chitinases have a highly conserved main structure, with an N-terminal cysteine-rich domain, reported to be implicated in chitin binding (Melchers *et al.*, 1994); (Iseli *et al.*, 1993). Most class I chitinases have a C-terminal extension that targets them to the vacuole (Neuhaus *et al.*, 1991).

Class II chitinases have a high similarity with the catalytic domain amino acid sequence of class I chitinases but lack the N-terminal cysteine-rich domain and the C-terminal extension.

Class III chitinases do not show any sequence similarity with class I and II (Collinge *et al.*, 1993).

Class IV chitinases contain a cysteine-rich domain and have a 41 - 47% sequence identity with class I. Moreover, class IV chitinases are different from those of class I as they possess four deletions in the catalytic domain (Collinge *et al.*, 1993).

Class V chitinases have been first isolated from tobacco and reported to act as endochitinases. They lack sequence identity with the previously described classes I-IV but share amino acid sequence identity with bacterial exochitinases (Melchers *et al.*, 1994); (Iseli *et al.*, 1993).

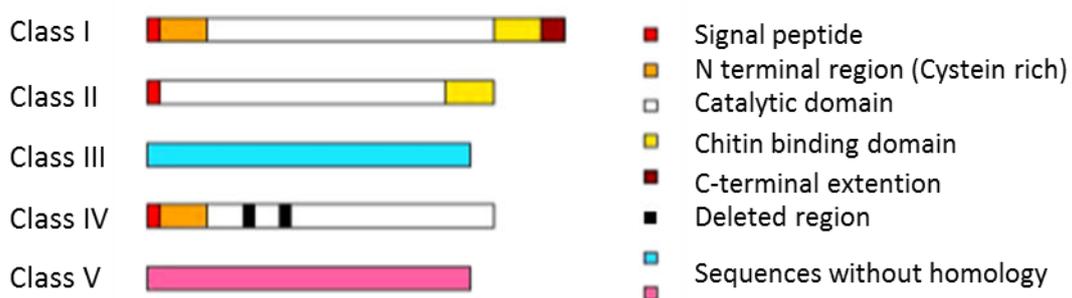


Figure 4. Schematic representation of the different domains in the five chitinase classes

Structural and biochemical studies have led to another classification of chitinases according to differences in three-dimensional (3D) structures and catalytic mechanisms.

These differences have allowed to group chitinases into two families of glycosyl hydrolases, 18 and 19 (GH18 and GH19). Thus, the five classes of chitinases are divided among two families (Henrissat & Bairoch, 1993); (Hamel *et al.*, 1997). Chitinases of these 2 different families do not share amino acid sequence similarity, and have completely different structures and molecular mechanisms of action.

Classes I, II, and IV are essentially expressed in plants and belong to family 19. Classes III and V form part of family 18 (Henrissat & Bairoch, 1993); (Kasprzewska, 2003) and consist of chitinases expressed in plants, bacteria, fungi, insects and mammals (Table 1).

Table 1. Classification of chitinases from various organisms, based on their amino acid sequence similarity.

Family 18 chitinases		Family 19 chitinases		
Plants				
Class III	Class V	Class I	Class II	Class IV
Virus Bacteria Fungi Nematodes Invertebrates Vertebrates		Nematodes		Bacteria (<i>Streptomyces spp.</i>)

Family 18 chitinases are widely distributed in all five kingdoms of nature including species of Archaea [*Thermococcus kodakarensis* (Fukui *et al.*, 2005)], Bacteria [*Serratia marcescens* (*S. marcescens*) (Brurberg *et al.*, 1994)], Fungi [*Coccidioides immitis* (*C. immitis*) (Bortone *et al.*, 2002); (Hollis *et al.*, 2000)], Plantae [tobacco (Melchers *et al.*, 1994)], and Animalia [sandfly (Ramalho-Ortigao & Traub-Cseko, 2003)] and human [chitotriosidase (CHIT1) (Fusetti *et al.*, 2002a)]. Multi-disciplinary approaches have been used to elucidate the origin, the architecture and the function of family 18 proteins. During evolution from invertebrates to vertebrates, glycosyl hydrolases 18 (GH18) proteins have gained new functions such as growth control, innate immune response role, among others that will be discussed later. GH18 family includes chitinases, the majority members, other chitinolytic enzymes and non-enzymatic proteins. Members from this family show a multi-domain

architecture, which consists of a combination of signal peptide, core domain, chitin-binding domain (ChBD) and serine/threonine-rich linkers, with many conserved amino acids repeats (Huang *et al.*, 2012). However, some GH18 members lack a ChBD (Huang *et al.*, 2012). According to structural analyses, GH18 enzymes' core contains the catalytic domain. In a large number of chitinases, the catalytic domain is flanked by a smaller C-terminal ChBD, separated by a short linker region (Li & Greene, 2010); (Brurberg *et al.*, 1994); (Blaiseau *et al.*, 1992); (Jekel *et al.*, 1991). Besides the catalytic domain and ChBD, some bacterial GH18 chitinases contain a fibronectin type III-like domain which plays a role in substrate binding (Horn, Sorbotten, *et al.*, 2006). Various combinations of domains in GH18 chitinases provide these enzymes different functional features regarding the catalytic efficiency on diverse chitin substrates (Huang *et al.*, 2012).

GH18 family also encompasses non enzymatic proteins highly homologous to chitinases called chitinase-like or chi-lectins. These chitinases-like proteins have naturally occurring mutations in their active site leading to the loss of the hydrolytic activity. During evolution, duplication and mutations resulted in the appearance of chi-lectin proteins making them unable to cleave chitin but since the residues involved in the substrate binding are conserved, these proteins preserved the feature of binding NAG polymers. Despite the enzymatic defection of chi-lectins, they show a marked conservation in their 3D folding with the catalytic domain of the active members of GH18 chitinases. Moreover, chi-lectins lack the ChBD, yet they bind to the substrate through their binding groove (Boot *et al.*, 1995); (Renkema *et al.*, 1997); (Boot *et al.*, 2001). Although there is detailed knowledge regarding their structure, insight in the exact physiological functions of the various chi-lectins remains limited (Bussink *et al.*, 2007).

In addition to chitinases and chi-lectins, GH18 family also includes certain enzymes with specificities for other GlcNAc-containing structures such as peptidoglycan (Bokma *et al.*, 1997)

i) Catalytic domain of GH18 chitinases

Structural studies on GH18 family catalytic domains showed that they share a common (α/β)₈ triosephosphate isomerase (TIM) barrel fold. This 3D folding consists of 8 stranded anti-parallel β -sheets laid out in staves tilted at 30° to the central axis, surrounded

by 8 α -helices, forming therefore a ring towards the outside (Terwisscha van Scheltinga *et al.*, 1996); (Fusetti *et al.*, 2002a). The active site of GH18 chitinases contains the conserved consensus sequence: DXXDXDXE that spans strand 4 of $(\alpha/\beta)_8$. The last aspartic and glutamic acid residues were reported to be essential for catalysis as shown by site-directed mutagenesis studies (Watanabe *et al.*, 1993). To complement mutagenesis studies, structural data have suggested that during degradation of chitin the glutamate functions in a general acid/base manner and acts as the proton donor. Additionally, at the C-terminus region of the core of many GH18 chitinases, there is an additional $\alpha + \beta$ fold region which gives a groove character to the active site. This region is composed of six antiparallel β -strands and one α -helix inserted in the loop between strand $\beta 7$ and helix $\alpha 7$ (Fusetti *et al.*, 2002b).

The substrate binds to the catalytic domain of GH18 chitinases in the cleft which appears in the center of the $(\alpha/\beta)_8$ fold. Note that substrate binding sites in GH enzymes use the $-n$ to $+n$ nomenclature where n is a digit representing the monomeric sugar residues in the substrate polymer. Monomers labeled as $-n$ are located towards the non-reducing end of the substrate and $+n$ towards the reducing end relative to the cleavage site located between the -1 and $+1$ monomers (Davies *et al.*, 1997) (Fig. 5). Computational and crystallographic studies have suggested that in general the binding cleft of GH18 chitinases consists of $(-4)(-3)(-2)(-1)(+1)(+2)$ monomers, whereas the binding cleft of GH19 chitinases encompasses monomers $(-3)(-2)(-1)(+1)(+2)(+3)$ (Hashimoto, Honda, *et al.*, 2000). Additionally, the active site is lined with aromatic residues that form stacking interactions with the hydrophobic ring of NAG monomers (Fusetti *et al.*, 2002b); (van Aalten *et al.*, 2001).

As previously mentioned, naturally occurring mutations in the active site, particularly on the catalytic glutamate, have resulted in proteins called chi-lectins which only bind to the substrate without degrading it (Bussink *et al.*, 2007).

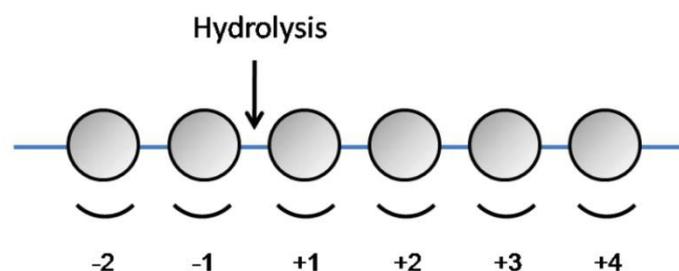


Figure 5. Diagrammatic representation of sugar-binding site in chitinases (based on the scheme proposed by (Davies *et al.*, 1997).

ii) Chitin-binding domain in chitinases

Chitin-binding domains (ChBD) are non-catalytic domains which contain between 45 and 70 residues and belong to the carbohydrate-binding module (CBM) protein domain type, known to group modules that interact with polysaccharides through aromatic residues exposed on its surface (Katouno *et al.*, 2004). The ChBD is a β -strand-rich domain which particularly binds to insoluble chitin (Uni *et al.*, 2012); (Suetake *et al.*, 2002); (Itoh *et al.*, 2002). The role of the ChBD was investigated by using truncated proteins lacking this domain as well as site-directed mutagenesis experiments. Indeed, deletion of ChBD decreased the binding and degradation efficiencies toward insoluble chitin such as colloidal chitin. However, several authors have reported that ChBDs do not bind to soluble chito-oligosaccharides or to soluble chitin derivatives. This suggested that the high affinity of ChBDs to insoluble chitin is essential for enhancing its degradation by the catalytic domain (Uni *et al.*, 2012); (Tjoelker *et al.*, 2000a); (Katouno *et al.*, 2004). Crystallographic and NMR spectroscopy studies have shown that few aromatic residues are involved in chitin binding (Ikegami *et al.*, 2000); (van Aalten *et al.*, 2000). Moreover, it was reported that several cysteine residues that form disulfide bonds are conserved in ChBDs and their mutation have led to complete loss in the binding function (Tjoelker *et al.*, 2000a).

1.1.2 Role of GH18 chitinases in lower organisms

Genome analysis shows that there are two GH18 members in *Saccharomyces cerevisiae*, 18 in filamentous fungus *Aspergillus nidulans*, 35 in *Caenorhabditis elegans*, 17 in *Drosophila melanogaster*, 8 in *Danio rerio*, 9 in *Mus musculus* and 6 in *Homo sapiens*. Interestingly, this wide distribution of GH18 chitinases across species is accompanied by a plethora of physiological functions which have diversified in an evolutionary manner (Huang *et al.*, 2012). Examples of reported roles related to members from this protein family in lower organisms are described in the paragraphs (i-v) and summarized in Table 2.

i) Bacteria

Although bacterial chitinases have been identified in families GH18 and GH19, (Dahiya *et al.*, 2006); (Udaya Prakash *et al.*, 2010); (Ueda *et al.*, 2009), most of them belong to the GH18 family (Larsen *et al.*, 2010). The distribution of GH19 chitinases among bacteria

appears to be restricted to actinobacteria and purple bacteria (Udaya Prakash *et al.*, 2010). Bacteria produce chitinases for different purposes such as nutrition and parasitism (Dahiya *et al.*, 2006); (Faramarzi *et al.*, 2009). In support to their role in nutrition process, it was proposed that the degradation of chitin by chitinases contributes to the supply of nitrogen and carbon (Cohen-Kupiec & Chet, 1998). The process by which chitin degradation occurs in bacteria is complex and involves the combined activity of different chitinases. For example *S. marcescens* produces both an endochitinase (Chitinase A) and an exochitinase (Chitinase B) which act in synergy to degrade chitin (van Aalten *et al.*, 2000). In addition, some bacterial chitinases like e.g. from *S. marcescens* and *Enterobacter agglomerans* protect plants from phytopathogenic fungi by degrading chitin from their cell wall, therefore inhibiting fungal growth (Chernin *et al.*, 1997); (Downing & Thomson, 2000).

ii) Parasites

GH18 chitinases play an important role in the life cycle of many parasites. It has been shown that pathogenic parasites use GH18 chitinases to degrade the chitin-containing barrier in insect vectors in order to invade them, which leads thereby to the transmission of infection to humans by these insect vectors. An interesting example is the case of *P. falciparum*, the causative parasite of human malaria. If the parasitic infective form for mosquito-vector is ingested with the bloodmeal of an infected vertebrate then, during the sporogonic cycle in the vector, *P. falciparum* uses chitinase to degrade the chitin-containing barrier and enter the midgut of the mosquito. Subsequently, the parasites migrate to the salivary glands of the vector and during the following bloodmeal from a new vertebrate the mosquito vector injects the parasites with the saliva completing by that the infectious transmission cycle (Bhatnagar *et al.*, 2003); (Isaacs *et al.*, 2012); (Vernick *et al.*, 2005); (Shahabuddin *et al.*, 1995).

iii) Fungi

All chitinases analyzed so far in the Fungi kingdom belong to the GH18 family (Hartl *et al.*, 2012). In fungi, GH18 members are reported to be involved in processes that include cell-wall degradation and modification, such as spore germination, tip growth and branching of hyphae, spore differentiation, autolysis and mycoparasitism (Adams, 2004). During the

cell division in yeast (particularly in *S. cerevisiae*), a chitin disk in the bud site septum is formed between the mother and the daughter cell by chitin synthase II. This is then followed by degradation via an endochitinase that leads to cell separation. The disruption of the gene encoding this endochitinase leads to defects in cell separation resulting in pseudohyphae formation (Kuranda & Robbins, 1991); (King & Butler, 1998). GH18 chitinases have also been identified in filamentous fungi. For example, in support of their morphogenetic role, disruption of the gene encoding the *Aspergillus nidulans* chitinase, leads to a defect in spore germination and a hyphal growth rate (Takaya *et al.*, 1998). In *Candida albicans* (*C. albicans*), a dimorphic fungus which can grow as a single cell or as a filamentous form, GH18 chitinases have been reported to be required for cell division and for the switch between forms (Dunkler *et al.*, 2005); (Kuranda & Robbins, 1991); (McCreath *et al.*, 1996).

iv) Insects

In insects, chitin is widely present in the exoskeleton. Therefore, chitin regeneration is a key process required for molting, exoskeleton development and thus essential for insect metamorphosis, growth and protection. Chitin synthases and chitinolytic enzymes work hand in hand in remodeling chitinous structures. Apart from the GH48 chitinase identified in leaf beetle *Gastrophysa atrocyanea* (Fujita *et al.*, 2006), all known insect chitinases belong to the family GH18 (Zhang *et al.*, 2011). In insects, chitin-degrading enzymes play a crucial role in postembryonic development, especially during larval molt and pupation. For example, during their life cycle, insects replace their chitinous cuticle. In order to achieve this, insect epidermal secrete and accumulate chitinases in the molting fluid between the epidermis and the old cuticle leading to the digestion of the latter. Interestingly, the expression of chitinolytic enzymes is under hormonal control at the different stages of metamorphosis (Merzendorfer & Zimoch, 2003).

v) Plants

Although plants do not synthesize chitin, they produce chitinases from GH18 and GH19 families. As mentioned before, only two classes of plant chitinases come from the GH18 family, classes III and V. Members from these two families belong to the pathogenesis-related proteins: a group of proteins which include hydrolytic enzymes, enzyme inhibitors,

and cell membrane-permeabilizing peptides that are produced by plants in response to invading pathogens and abiotic factors (Edreva, 2005); (Sels *et al.*, 2008). In fact, chitinases secreted by plants contribute to their defense mechanism by degrading fungal cell walls (Samac & Shah, 1991); (Lawton *et al.*, 1993).

In addition to the expression of active GH18 chitinases, plants also express chi-lectins, the enzymatically inactive GH18 proteins. They are produced by a broad range of plants and, similar to active chitinases, appear to be involved in different physiological processes (Van Damme *et al.*, 2007). For instance, in *Arabidopsis thaliana* and *Oryza sativa*, it has been shown that chi-lectins proteins are involved in cellulose biosynthesis (Sanchez-Rodriguez *et al.*, 2012); (Wu *et al.*, 2012).

Table 2. Summary of the different functions of GH18 family chitinases in lower organisms.

Organism	Function
Bacteria	Antifungal protection/energy source
Parasites	Migration in host/egg hatching
Fungi	Cell wall component/growth
Arthropods	Exoskeleton component/growth
Plants	Antifungal immunity

1.1.3 Mammalian chitinases

1.1.3.1 Phylogenetic analyses of mammalian GH18 members

Mammals lack endogenous chitin and they contain only GH18 family members in their genomes (Bussink *et al.*, 2007); (Collinge *et al.*, 1993); (Li & Greene, 2010); (Funkhouser & Aronson, 2007). From the functional point of view, the lack of chitin and chitin synthases in mammals strongly suggests a different role/s from those in insects and fungi. Functional and phylogenetic studies of the GH18 family in mammals have been focused so far on human and murine genes. These studies have divided the vertebrate GH18 genes into 3 distinct phylogenetic groups: chitinases/chi-lectins, chitinase-domain containing (CHIDs) and chitobiasis (CTBSs) (Table 3). In these three groups, only three active chitinolytic enzymes

have been found, two of which display chitinase activity, chitotriosidase (CHIT1) and acidic mammalian chitinase (AMCase) while the third one (CTBS) is only capable of hydrolyzing monosaccharides from oligosaccharides present in glycoproteins and does not act on chitin (Funkhouser & Aronson, 2007); (Synstad *et al.*, 2004); (Boot *et al.*, 2001); (Bussink *et al.*, 2007).

Evolutionary analyses have proposed that ancient gene duplication first allowed the specialization of the two active chitinases, CHIT1 and AMCase. Indeed, some authors have suggested that the gene duplication allowing evolution of CHIT1 and AMCases occurred very early in tetrapod evolution (Bussink *et al.*, 2007). Moreover, it has been proposed that subsequent gene duplications of active chitinase genes followed by mutation causing loss of enzymatic function have led to the evolution of the chi-lectins in mammals. Table 3 details the different human and murine GH18 genes described.

According to phylogenetic data, the two active mammalian chitinases, CHIT1 and AMCase, are present in all mammals for which their genomic information is available and are extremely conserved in these organisms (Bussink *et al.*, 2007). These enzymes have a molecular weight of 50 kDa which corresponds to a 39 kDa catalytic domain linked by a hinge to a small chitin-binding domain. The X-ray crystal structures of the catalytic domains of the two human chitinases have shown that they share a conserved 3D folding consisting of a $(\alpha/\beta)_8$ TIM barrel, where the catalytic motif (DXDXE) is located (Fusetti *et al.*, 2002b); (Olland *et al.*, 2009a). However, no 3D structures of the human chitin-binding domain have been elucidated yet (Tjoelker *et al.*, 2000b). The detailed features of CHIT1 and AMCase will be developed in section 1.1.3.2 of the introduction.

In contrast to the two mammalian chitinases, the chi-lectin genes are present only in some particular species. For example, the CHI3L2 (YKL-39) can be found in the primate and cow genomes but not in the genomes of rodents (Table 3). The opposite is the case of chitinase 3-like 3 (Chi3l3) or (Ym1), chitinase 3-like 4 (Chi3l4) or (Ym2), and brain chitinase-like protein 2 genes (Bclp2) which exist only in rodents but not in primate genomes (Table 3). Regarding their structure, chi-lectins consist of a 39 kDa domain retaining the TIM-barrel structure like the one found in active chitinases, (Sun *et al.*, 2001); (Wierenga, 2001); (Fusetti *et al.*,

2002a); (Houston *et al.*, 2003) but they lack the chitinbinding domain. Only oviductin (OVGP1) became basic and gained a glycosylate serine/threonine-rich domain (Buhi, 2002).

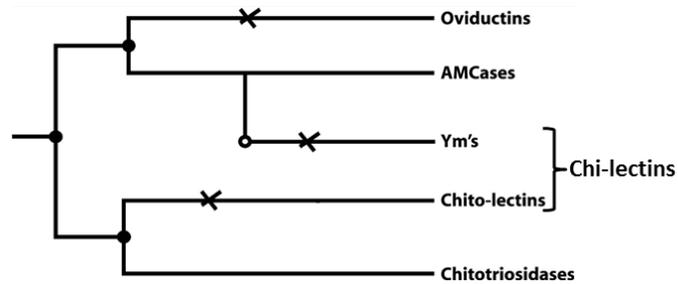


Figure 6. Overview of the evolution of the four subsets chitinase genes, ● the “ancestral” gene duplications, a cross indicates the loss of catalytic activity mutations. “Chito-lectins” are chi-lectins evolved from the chitotriosidase gene (duplication). Yms are rodent chi-lectin. Adapted from (Bussink *et al.*, 2007).

Table 3. Overview of the human and murine GH18 genes, common aliases from the literature and hydrolytic activity. Colors are according to the groups: chitinases/CLPs in blue, CTBS in red and CHID1 in green

gene Name	Common aliases	Chitin hydrolytic activity	Group	Genome: Human(H) Mouse (M)
CHIA	AMCase	Endo	Chitinase/CLPs	H/M
CHIT1	Chitotriosidase/ chitinase1	Endo	Chitinase/CLPs	H/M
OVGP1	Oviductin, MUC9	-	Chitinase/CLPs	H/M
CHI3L1	YKL-40, BRP-39, cartilage glycoprotein 1	-	Chitinase/CLPs	H/M
CHI3L2	Chondrocyte protein 39	-	Chitinase/CLPs	H
Chi3L3	Ym1	-	Chitinase/CLPs	M
Chi3L4	Ym2	-	Chitinase/CLPs	M
Chi3L7	BCLP	-	Chitinase/CLPs	M
BC051070	BYm	-	Chitinase/CLPs	M
CTBS		Exo	CTBS	H/M
CHID1	SI-CLP	-	CHID1	H/M

1.1.3.2 Human and murine GH18 members

The human genome encodes two active chitinases, CHIT1 and AMCase, three chi-lectins OVGP1, CHI3L1 and CHI3L2, one CTBS and one CHID (Table 3). All of the human GH18

family members are located on human chromosome 1 except CHID, located on chromosome 11 (Bussink *et al.*, 2007); (Funkhouser & Aronson, 2007).

The following section will provide a review of the two active mammalian chitinases, CHIT1 and AMCase, as well as CHI3L1, which are the most well-described chitinases and chitinases in humans and mice. So far the acquired knowledge of their functions indicates that they are under complex regulation and their roles are widely interconnected in immune regulation, tissue remodeling and fibrosis, involving both chitinous and non-chitinous substrates.

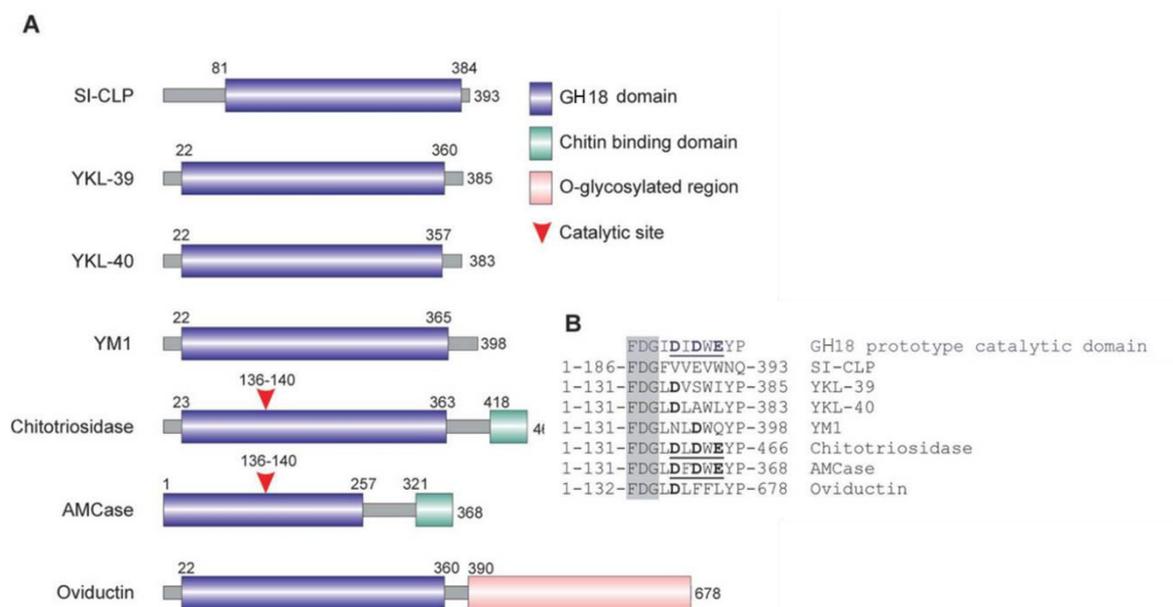


Figure 7. A) Schematic representation of domain organization of mammalian GH18 proteins. B) Critical amino acid in catalytic sites. The conserved FDG sequence preceding catalytic motif is shown in shadowed box. Catalytic residues are shown in bold. Complete active catalytic motifs are underlined. Adapted from (Kzhyshkowska *et al.*, 2007).

i. Chitotriosidase (CHIT1)

The existence of CHIT1 in humans was discovered serendipitously by researchers during biochemical investigations on plasma specimens from patients suffering from Gaucher disease (details on Gaucher disease are described in section 1.1.3.3 paragraph i). CHIT1 was the first active chitinase described in humans and mammals, and is found encoded in all mammalian genomes (Hollak *et al.*, 1994); (Boot *et al.*, 1995).

CHIT1 is an endochitinase which shows a dual activity towards chitin, the classical hydrolysis activity and transglycosylation (detailed in section 1.2.3.3 and 1.2.3.4) (Aguilera *et al.*, 2003); (Fusetti *et al.*, 2002a); (Renkema *et al.*, 1997); (Boot *et al.*, 1995).

The CHIT1 gene is located on chromosome 1, contains 12 exons spanning ~20 kb and encoding multiple splice forms (Eiberg & Den Tandt, 1997) (Fig. 8A). As described in section 1.1.3.1, the 50 kDa isoform contains the catalytic domain linked by a hinge to the ChBD. This isoform is predominantly secreted, but in part is processed by a proteolytic event into a 39 kDa form that accumulates in lysosomes. Moreover, in macrophages, alternative splicing generates a minor level of a distinct CHIT1 mRNA variant, encoding a 40 kDa isoform of CHIT1, which is C-terminally truncated and almost identical to the 39 kDa cleaved form (Fig. 8B). Both isoforms with molecular masses of 50 and 39 kDa were isolated and both were shown to be enzymatically active chitinases. However, the 39 kDa form corresponds only to the active catalytic domain and does not bind to colloidal chitin (Boot *et al.*, 1998); (Renkema *et al.*, 1997). It is worth to note that the secreted 50 kDa form is found in the bloodstream whereas the 39 kDa form is predominantly found in tissues (Renkema *et al.*, 1997). In some populations there is a common polymorphism in the gene CHIT1 resulting in a 24-base pair duplication and leading to the production of an inactive enzyme (Malaguarnera, 2006); (Maver *et al.*, 2010); (Boot *et al.*, 1998). Heterozygotes were reported to have about 50% of CHIT1 activity compared to wild type, whereas homozygotes have no CHIT1 activity in any body tissue (Malaguarnera, 2006).

Recent experiments have shown a wide expression of CHIT1 in healthy human tissues, especially in lung, followed by spleen, fetal liver, thymus and lacrimal gland (Ohno *et al.*, 2013); (Hall *et al.*, 2008). Interestingly, functional studies have demonstrated that the major active chitinase in human lung is CHIT1 (Seibold *et al.*, 2008).

At the cellular level, CHIT1 is mainly expressed, stored, and secreted by cells involved in the innate immune system such as macrophages and neutrophils. In some tissues, it is expressed by epithelial cells. The role of CHIT1 in immune cells is related to their maturation. Indeed, during the differentiation of monocytes to macrophages and phagocytes, CHIT1 gene is gradually upregulated, with a particularly high expression in the later phases of monocyte differentiation (Di Rosa, De Gregorio, *et al.*, 2013); (van Eijk *et al.*, 2007); (Boot *et*

al., 1995). This led to the suggestion that CHIT1 plays an active role in the innate immune response. Effectively, *in vitro* studies with recombinant human CHIT1 have shown the ability to stop the hyphal growth of *C. albicans* and inhibit *Cryptococcus neoformans* proliferation as well as causing hyphal tip lysis in *Mucor rouxii* (Boot *et al.*, 2001); (van Eijk *et al.*, 2005). Furthermore, *in vivo* experiments have demonstrated that administration of recombinant CHIT1 clearly improves survival in mouse models challenged with lethal doses of *C. albicans* or *A. fumigatus*, the main causes of mortality in immuno-compromised individuals. These studies confirmed the antifungal activity of CHIT1 by both *in vitro* experiments and animal models, supporting thereby the proposed role of CHIT1 in innate immune response against chitin-containing invaders (van Eijk *et al.*, 2005). This was further supported by clinical observations where CHIT1 activity was detected to be highly elevated in plasma and urine of fungal infected children or parasitic infection such as *P. falciparum* malaria (Barone *et al.*, 2003). Interestingly, health improvement of these subjects led to a decrease in CHIT1 activity (Labadaridis *et al.*, 2005). However, Hall *et al.* clearly showed that CHIT1 does not have any effect on bacterial growth, consistent with the fact that bacteria do not contain chitin (Hall *et al.*, 2008).

A lot of research has been dedicated to the role of CHIT1 in innate immune response. The upregulation of CHIT1 gene expression was shown to be correlated with stimulation of monocyte-derived macrophages via cytokines (GM-CSF, TNF- α and IFN- γ), prolactin hormone and LPS (a liposaccharidic component of Gram negative bacteria found in the outer membrane, a strong inducer of the immune response) (Malaguarnera *et al.*, 2004); (Malaguarnera *et al.*, 2005). Moreover, several pieces of evidence have revealed that the prolactin-mediated induction of CHIT1 gene is regulated by protein tyrosine kinase (PTK), phosphoinositide-3-kinase (PI3K), and mitogen-activated protein kinases (MAPK), p38 and ERK1/2 signaling transduction components (Di Rosa *et al.*, 2009). On the other hand, toll-like receptor (TLR), which can recognize fungi such as *C. albicans*, *A. fumigatus* and *Cryptococcus neoformans*, was found to trigger the release of CHIT1 from specific granules in neutrophils (van Eijk *et al.*, 2007).

Altogether, these data indicate that CHIT1 is induced to different extents by a variety of cytokines and hormonal cues and in different cells of the immune system supporting the

thought of its biological relevance for the host immune response in the early phases of infections.

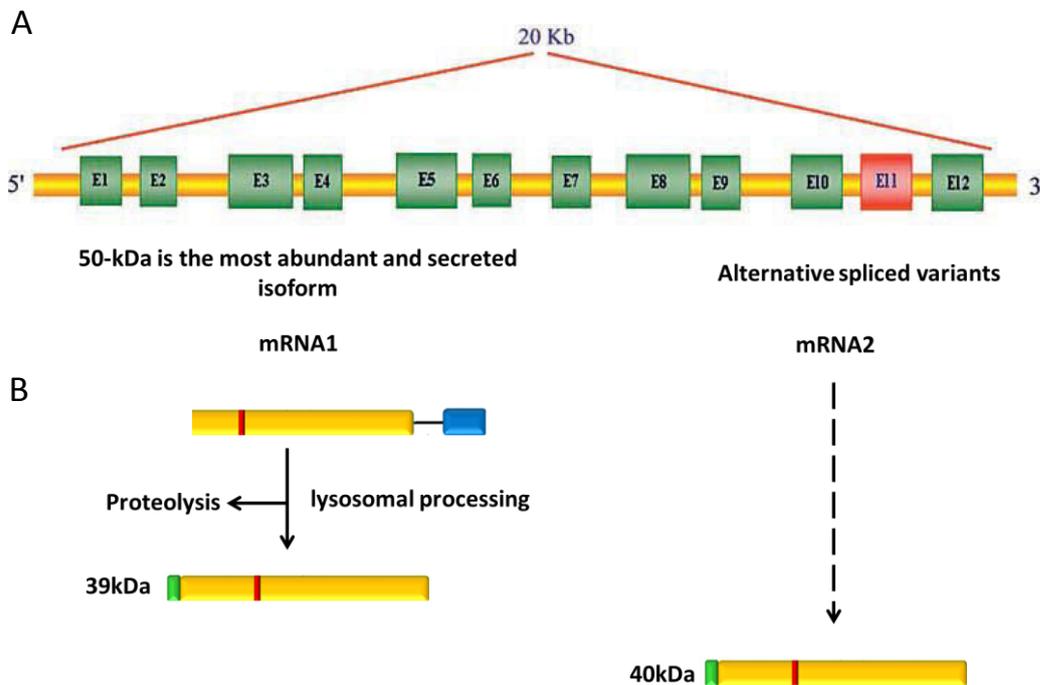


Figure 8. A) Gene structure of CHIT1 composed of 12 exons and spans about 20 kb. B) The predominant mRNA species encoding the 50 kDa protein correspond to the secreted isoform. Alternative splicing generates CHIT1 mRNA that encodes a 40 kDa isoform as the exon 11 introduces a premature stop codon. The 40 kDa isoform is almost identical to the 39 kDa isoform generated by proteolytic processing of the 50 kDa CHIT1.

ii. Acidic mammalian chitinase (AMCase)

Acidic mammalian chitinase (AMCase) was the second chitinase identified in human and is able to degrade chitinous substrates and fungal cell wall chitin (Boot *et al.*, 2001) (Boot *et al.*, 2005). The AMCase gene (*CHIA*) is located on human chromosome 1 and, as mentioned before, it has a high sequence homology with CHIT1, suggesting that these genes arose from a duplication event of an ancestor gene (Boot *et al.*, 2001). Like CHIT1, AMCase is synthesized as a 50 kDa protein that contains a 39 kDa catalytic domain, linked to a C-terminal ChBD by a hinge region. Although, there is an overall high similarity between human chitinases, AMCase shows a low isoelectric point and a distinct acidic pH optimum, between 4 and 6 (Boot *et al.*, 2001).

The main tissues and organs that express AMCase have been identified by qPCR in humans and mice. While the fetal brain, and liver, lung, heart and thyroid gland are the

major tissues of *CHIA* transcription in humans, in mice *CHIA* gene is most highly expressed in the stomach (Ohno *et al.*, 2013). Indeed, in insectivorous mammals it has been suggested that AMCase may actually be involved in chitin digestion (Strobel *et al.*, 2013). This is probably not the case for humans, as AMCase expression level in the human gastrointestinal tract is significantly lower than that observed in mice (Ohno *et al.*, 2013).

AMCase was found to have an anti-parasitic role in a murine model of toxoplasmosis. Indeed, upon infection by *Toxoplasma gondii*, AMCase expression is highly upregulated in macrophages of the mouse brain. The AMCase activity within the central nervous system of mice was reported to be an essential immunological factor contributing to the lysis of chitinous cysts and eradication of parasitic infections (Nance *et al.*, 2012).

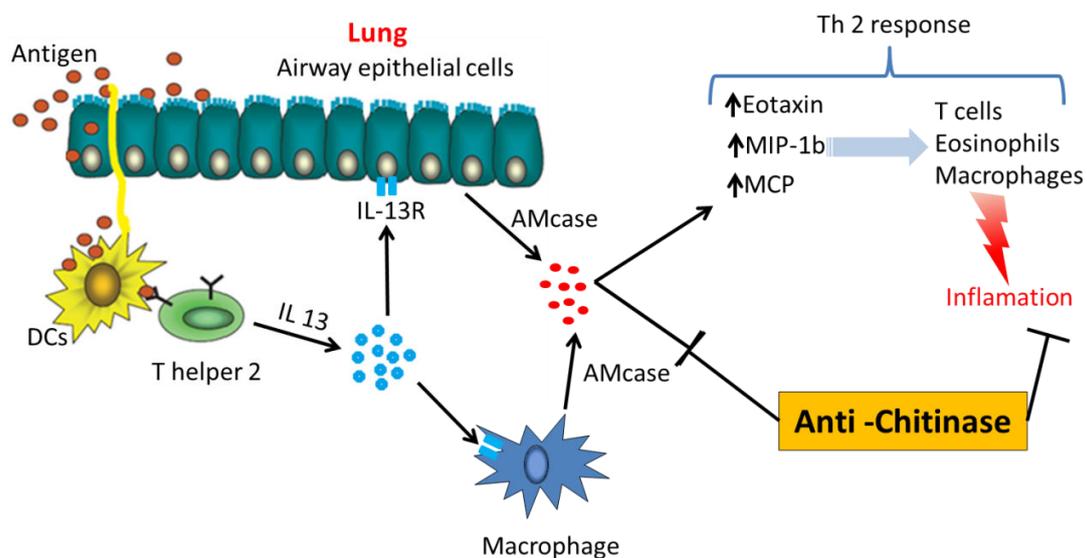


Figure 9 Schematic representation of the mechanism of chitinases mediated airway hyperresponsiveness and airway inflammation. Antigen presenting cells such as dendritic cells (DC) actively uptake antigens (presumably contain chitin) and present these antigens to Th2 cells. These cells produce Th2 cytokines such as IL-13. Of note, IL-13 plays a major role to induce the production of AMCase by airway epithelial cells and macrophages, which express IL-13 receptor (IL-13R) on their surface. AMCase induces the production of cytokines such as MCP and eotaxin, which induce the recruitment of T cells, eosinophils and macrophages in the lung and further exacerbates the inflammation and airway hyperresponsiveness. Anti-AMCase antibody and chitinase inhibitor suppress both airway hyperresponsiveness and inflammation. Adapted from (Kawada *et al.*, 2007).

Moreover, this enzyme has attracted considerable attention since a study conducted in an aeroallergen asthma mouse model on ovalbumin-sensitized mice showed an important role of AMCase in the pathology of asthma and allergic inflammation (Zhu *et al.*, 2004) (Fig. 9). In this study, both mRNA and AMCase protein were found overexpressed by the

epithelial airways and macrophages in the lungs of asthmatic mice. In addition, *in situ* hybridization experiments have demonstrated overexpression of AMCcase mRNA in the epithelial lung cells of patients with asthma. Asthma is a chronic disease characterized by an exaggerated adaptive immune response-mediated by T helper 2 (Th2) cells which enhance airway inflammations (Ray & Cohn, 1999). Indeed, the high expression of AMCcase was shown to occur in response to the stimulation by interleukin-13, a Th2 cytokine recognized to be involved in asthma development. Interestingly, administration of anti-AMCcase antibodies, a siRNA approach to knockdown AMCcase in mice or nonspecific AMCcase inhibition by allosamidin (see section 1.4) have led to the decrease of Th2-inflammation and tissue eosinophilia (Zhu *et al.*, 2004) (Fig. 9). Further, it has been demonstrated that epidermal growth factor receptor (EGFR) participates in the induction of AMCcase secretion which then stimulates inflammatory chemokine production by pulmonary epithelial cells (Hartl *et al.*, 2008). In addition, research for common genetic variants of human AMCcase in pediatric asthma revealed that AMCcase polymorphisms are associated with bronchial asthma in children (Bierbaum *et al.*, 2005). Altogether, these findings have suggested that AMCcase is involved in asthma progression (Sutherland *et al.*, 2009); (Zhu *et al.*, 2004); (Sutherland *et al.*, 2011).

Accordingly, AMCcase has been considered a potential pharmaceutical target for novel anti-asthma treatment (Sutherland *et al.*, 2011); (Sutherland *et al.*, 2009); (Zhu *et al.*, 2004); (Elias *et al.*, 2005); (Cole *et al.*, 2010). However, in seemingly contrast to this proposition, later studies in mouse models of asthma have provided conflicting evidence regarding the beneficial effect on eosinophilic lung inflammation of inhibiting AMCcase (Zhu *et al.*, 2004); (Matsumoto *et al.*, 2009); (Yang *et al.*, 2009); (Van Dyken *et al.*, 2011); (Fitz *et al.*, 2012).. Potential explanations for these contradictory data include differences in allergen challenge protocols used on mice, or with the use of allosamidin, an unspecific chitinase inhibitor which could have targeted other GH18 members. Additionally, one may consider that there might be several pathways which require AMCcase expression and very likely AMCcase mutant mice show a deregulation on all on these pathways (Sutherland *et al.*, 2011).

With regards to the complex implication of chitinases in human and mouse physiological processes, an interesting finding has shown that AMCcase possesses other

biological effects independent from its chitinolytic activity. Indeed, evidence has indicated that this protein displays anti-apoptotic properties on the epithelium via phosphoinositide 3-kinase (PI3K) and AKT signaling which do not require its enzymatic activity. Interestingly, AMCCase point mutations, which keep the enzyme capable of binding chitin but abrogate its enzymatic function, did not alter its anti-apoptotic role. However, this new function was found to be abolished by allosamidin which competes with chitin binding on both active AMCCase and mutated non-active AMCCase. Thus, these results demonstrate that only AMCCase-substrate binding feature is required to protect cells from apoptosis. Importantly, the anti-apoptotic effect of AMCCase was proposed to alter proliferation/apoptosis ratio in epithelial cell favoring epithelial cell accumulation and causing hyperplasia and airway remodeling, such in inflammation (Hartl *et al.*, 2009).

Other studies have reported that AMCCase and/or CHIT1 may also be implicated in other inflammatory related disorders such as conjunctivitis, nasal polyp pathogenesis, adenoid hypertrophy, neuromyelitis, gastritis and most recently eosinophilic esophagitis (Bucolo *et al.*, 2011); (Correale & Fiol, 2011); (Cozzarini *et al.*, 2009); (Heo *et al.*, 2011); (Park *et al.*, 2011); (Ramanathan *et al.*, 2006); (Cho *et al.*, 2014).

Importantly, most of the information regarding targeting AMCCase has been generated from work using allosamidin, a nonspecific inhibitor of GH18 chitinases. Surprisingly, introducing specific AMCCase inhibitors has resulted in an imbalance of the immune response involving an accumulation of neutrophils, which are not associated to Th2 response. One should also bear in mind that the beneficial effects of allosamidin in comparison to other specific inhibitors of AMCCase may be due to actions that are independent of direct chitinase activity like the substrate binding features and/or because it may have targeted simultaneously other members of the chitinase family such as the chitinase-like proteins. Thus, new specific inhibitors could potentially be developed with activity against chi-lectins proteins. An understanding of the actions of the chi-lectin protein family and the chitinolytic-independent actions of AMCCase merits high attention in the future (Sutherland *et al.*, 2011).

1.1.3.3 Role of CHIT1 activation in diseases

i) CHIT1 and Gaucher disease

As pointed out in the previous section, CHIT1 was detected in Gaucher disease. This disease is a rare, recessively inherited lysosomal storage disorder, resulting from a mutation in the glucocerebrosidase (GCase) gene, leading to a defect in the encoded enzyme and to an accumulation of lipid-laden macrophages known as Gaucher cells (Hollak *et al.*, 1994). As a result, Gaucher patients develop gross hepatosplenomegaly, bone lesions, and less frequently, neurological abnormalities. Importantly, in the serum of Gaucher patients, the activity of CHIT1 increases 10 – 1000 fold due to its excessive expression by Gaucher cells. The elevated activity has led to consider CHIT1 as a biomarker of Gaucher disease for the clinical diagnosis and for monitoring the efficacy of treatment. In fact, CHIT1 activity decreases sharply upon Enzyme Replacement Therapy (ERT), coinciding with clinical improvements in Gaucher patients (Hollak *et al.*, 1994); (de Fost *et al.*, 2006); (Pastores *et al.*, 2005); (Pastores & Barnett, 2005).

ii) CHIT1 and Malaria infection

As aforementioned, plasmatic CHIT1 activity is highly increased in *P. falciparum* malaria infected subjects, consistent with the host defense role of CHIT1 (Barone *et al.*, 2003); (Sutherland *et al.*, 2009). Barone *et al.* have demonstrated that the red blood cell destruction in malaria patients triggers CHIT1 overproduction by active macrophages. Therefore, it has been suggested to monitor the macrophage functions through plasma CHIT1 levels in malaria patients (Barone *et al.*, 2003). On the other hand, malaria parasites produce their own chitinase (PfCHT1) whose role is to digest and cross the chitinous barriers in insect-vectors thereby completing the parasitic cycle transmission (Tsai *et al.*, 2001). Interestingly, it has been reported that insects fed with blood from malaria patients, with higher plasma CHIT1 activity, could therefore aid the malaria parasite to complete its life cycle in the vector increasing its transmissibility (see section 1.1.2 paragraph ii) (Shahabuddin & Kaslow, 1993); (Shahabuddin *et al.*, 1995); (Di Luca *et al.*, 2007); (Tsai *et al.*, 2001). Sequence analysis of CHIT1 polymorphisms and PfCHT1 reveals the existence of amino acid sequence similarities (Vinetz *et al.*, 1999). These findings have led to some confusion over the role of CHIT1 activity upon parasitic infection: the high activity of CHIT1

related to the immune response of the host, instead of being protective, could be beneficial for the parasites (Sutherland *et al.*, 2009).

iii) CHIT1 and non-alcoholic liver disease-steatohepatitis

The expression of CHIT1 mRNA is induced in patients with non-alcoholic fatty liver disease steatohepatitis (NASH) (Malaguarnera, Di Rosa, Zambito, dell'Ombra, Nicoletti, *et al.*, 2006). A crucial event in the initiation of NASH involves lipid accumulation and lipid peroxidation in the hepatocytes. This is followed by the activation of the liver macrophages, called Kupffer cells, and the hepatic stellate cells. Overproduction of CHIT1 by the Kupffer cells results in the activation of hepatic stellate cells suggesting that this enzyme is implicated in the progression of hepatic fibrosis (Malaguarnera, Di Rosa, Zambito, dell'Ombra, Nicoletti, *et al.*, 2006); (Malaguarnera, Di Rosa, Zambito, dell'Ombra, Di Marco, *et al.*, 2006).

iv) CHIT1 and atherosclerosis

Atherosclerosis is an inflammatory disease characterized by progressive deposition of lipids and fibrous matrix in the arterial wall. In patients with atherosclerosis, serum levels of CHIT1 are increased up to 55-fold compared to normal individuals demonstrating the presence of activated macrophages in these subjects. Further, a correlation has been established between the increased CHIT1 expression and the accumulation of lipid-laden macrophages inside human atherosclerotic vessel walls (Karadag *et al.*, 2008). Although the underlying reasons are still unknown, CHIT1 activity has been related to the severity of the atherosclerotic lesions, making it a putative biomarker of the atherosclerotic extension (Moreno *et al.*, 1994).

v) CHIT1 and sarcoidosis

Sarcoidosis is a disorder of unknown origin characterized by the accumulation of activated immune cells in affected organs which results in granuloma formation (Baughman *et al.*, 2003). Most patients with active sarcoidosis have highly elevated activity of CHIT1 in serum and bronchoalveolar lavage. In sarcoidosis, CHIT1 is expressed and released by the macrophages and treatment with an anti-inflammatory agent such as corticosteroids causes a decrease in CHIT1 activity in the majority of patients (Boot *et al.*, 2010); (Kanneganti *et al.*,

2012); (van Eijk *et al.*, 2005). It has been suggested that fungi may be the causative factor in sarcoidosis and the higher activity of CHIT1 is reflecting a specific reaction against chitin from these fungi (Boot *et al.*, 2010); (Tercej *et al.*, 2007); (Tercej *et al.*, 2008).

vi) CHIT1 and Alzheimer's disease

The expression of CHIT1 mRNA was found elevated in some neurodegenerative disorders, particularly in Alzheimer's disease (AD) (Di Rosa *et al.*, 2006). Alzheimer's disease is a progressive neurodegenerative disorder resulting in the loss of higher cognitive function caused by accumulation of β -amyloid plaques. A feature in the brain of AD patients is the presence of activated microglial cells, the resident macrophages of the central nervous system. In this respect, the elevated expression of several cytokines correlated well with the significantly elevated expression of CHIT1 in AD patients. However, it is not clear if CHIT1 activity affects directly the central nervous system functions. Sotgiu *et al.* have proposed a dual role of CHIT1: the first one reflecting the strong macrophage-microglia activation due to β -amyloid deposition and testifies the existence of an inflammatory process. The second one suggesting a protective role of CHIT1 in AD brains by degrading the chitin-like substrates decreasing like that the β -amyloid fibers. (Sotgiu *et al.*, 2008); (Castellani *et al.*, 2005); (Di Rosa *et al.*, 2006).

vii) CHIT1 and cancer

In primary prostatic cancer patients, high CHIT1 activity was observed only in patients with high Gleason score, the most widely used diagnostic method of grading prostate cancer tissue. The correlation between Gleason scores and CHIT1 activity shows the importance of macrophage involvement in cancer progression. This indicates that CHIT1 may have a role in the progression of cancer to the malignant state. Indeed, nowadays it has been well established that macrophage-produced factors play a critical role in cancer progression through paracrine signaling pathways and/or through destruction of extracellular matrix, which enhances invasion and metastasis (Kucur *et al.*, 2008); (Kanneganti *et al.*, 2012).

vii) Summary of CHIT1 in diseases

Enzymatically active CHIT1 has been reported to be associated with several diseases, which comprise macrophage activation and in some cases, it has been validated as a disease

biomarker. The overall effects of CHIT1 seem to be dependent on many factors, including the stage of inflammation and the specific cell types and organs involved. It is worth to point out that the mechanisms by which CHIT1 increases and the consequences of its elevated activity remain not fully understood. It also remains under debate whether CHIT1 represents a valid therapeutic target for some of these disorders (Di Rosa *et al.*, 2014). It is then clear that depending on the context, this protein may be beneficial or detrimental for the host (Malaguarnera, 2006); (Sutherland *et al.*, 2009); (Kanneganti *et al.*, 2012).

1.2 Structural and mechanistic studies of GH18 chitinases

The first published structures of the GH18 chitinase family were those of ChiA from *S. marcescens* (Perrakis *et al.*, 1994) and of a plant single domain endochitinase called hevamine (Terwisscha van Scheltinga *et al.*, 1996); (Terwisscha van Scheltinga *et al.*, 1994). Afterwards, crystal structures for several bacterial, fungal, plant and human GH18 chitinases have been determined (Terwisscha van Scheltinga *et al.*, 1994); (Hollis *et al.*, 2000); (Hart *et al.*, 1995); (Perrakis *et al.*, 1994); (Fusetti *et al.*, 2002b).

1.2.1 Structural analysis of the catalytic domain of GH18 chitinases

Although the overall sequence similarity between GH18 chitinases is not particularly high (average pairwise identity 21%; <http://www.sanger.ac.uk/Software/Pfam>), all the catalytic domains of GH18 chitinases have a common TIM-barrel (α/β)₈ fold characterized by several conserved sequence motifs (Fusetti *et al.*, 2002b); (Hollis *et al.*, 2000); (Watanabe *et al.*, 2001); (Perrakis *et al.*, 1994); (Terwisscha van Scheltinga *et al.*, 1994) and (van Aalten *et al.*, 2000) (Fig. 10). The most prominent motif is the DXXDXDXE consensus site located in β 4 strand of the TIM barrel, where the last residue in the motif is the catalytic glutamate (Fig. 11A). Additionally, the second Asp (D2) is supposed to contribute to the stabilization of the substrate during the catalytic reaction (McCarter & Withers, 1994); (Perrakis *et al.*, 1994); (Terwisscha van Scheltinga *et al.*, 1994); (Tews *et al.*, 1997); (Brameld & Goddard, 1998); (van Aalten *et al.*, 2000); (van Aalten *et al.*, 2001); (Fusetti *et al.*, 2002a). In contrast, GH19 chitinases have a different 3D folding, with more α -helices, and share the bilobal $\alpha+\beta$ motif

of lysozyme, which forms a well-defined substrate binding cleft between the lobes (Hart *et al.*, 1995).

The crystal structure of the 39 kDa CHIT1 catalytic domain (CHIT1-CAT), has been determined by Fusetti and collaborators (Fusetti *et al.*, 2002a), presenting the conserved TIM barrel folding observed in all the family 18 chitinases, encompassing the DXXDXDXE motif located at the end of its fourth β strand (Fig. 11A, D).

Furthermore, an insertion of an ($\alpha + \beta$) domain in the (α/β)₈ barrel is observed in human GH18 chitinases (CHIT1 and AMCase). This insertion is conserved in several GH18 chitinases from various species and is thought to form a wall leading to a profound binding groove (cleft) in the core of the GH18 catalytic domain (van Aalten *et al.*, 2000); (Fusetti *et al.*, 2002a); (Olland *et al.*, 2009b); (Zees *et al.*, 2009). The ($\alpha + \beta$) insertion region has a conserved folding composed of six anti-parallel β strands and one α -helix and it inserts between the seventh α -helix and seventh β -strand of the TIM barrel (Fig 11B, C, D). In CHIT1, this insertion consists of 67 residues. In ChiB, this additional domain is much bigger (79-residue $\alpha + \beta$ domain) but less so in ChiA (61-residue $\alpha + \beta$ domain). The catalytic binding groove contains a large number of conserved solvent-exposed aromatic residues as well as polar residues which are thought to contribute to substrate binding (Watanabe *et al.*, 2003); (Katouno *et al.*, 2004), (Fig. 12), (Table 4, 5). The lack of the ($\alpha + \beta$) domain in chitinase C (ChiC) of *S. marcescens* and hevamine, where the catalytic domain is composed only of a TIM barrel structure makes the substrate binding cleft much shallower and leads to losing most of the aromatic residues seen in CHIT1, AMCase, ChiA and ChiB (Terwisscha van Scheltinga *et al.*, 1994); (Horn, Sorbotten, *et al.*, 2006). An important feature of the CHIT1 active site cleft consists of its extended and more open character on one face of the enzyme (Fig. 11D).

Apart from the catalytic domain, some of the GH18 chitinases contain one or several additional domains that participate in substrate binding and are presumed to promote activity on chitin (Horn, Sorbotten, *et al.*, 2006). Section 1.3 of this manuscript is devoted to the description of the structural features of such additional domains.

As far as the mode of action of chitinases is concerned, the position of aromatic residues within the active cleft is thought to influence their mode of the action regulating

features such as the substrate binding, endo- or exo-chitinolytic activity, co-existence of dual hydrolysis and transglycosylation, processivity, etc. Indeed, CHIT1 is suggested to be a processive (see section 1.2.4) enzyme with endochitinase activity (see section 1.1.1) which shows a dual activity: hydrolysis and a high rate of transglycosylation (see section 1.2.3.4).

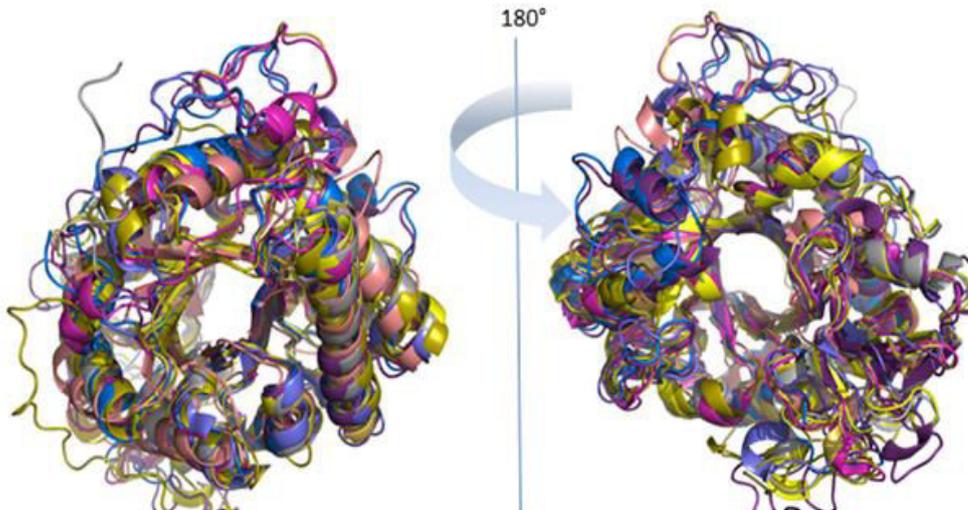


Figure 10. Superimposition of eight GH18 chitinases and chitinase-like structures showing the high conservation of the TIM-barrel (α/β)₈ fold. Protein Data Bank (PDB) code: [CHIT1: 1HKM], [MGP-40: 1LJY], [*B. circulans* chitinase A1: 1ITX], [*C. immitis* chitinase: 1D2K], [*S. marcescens* ChiA: 1FFR], [*S. marcescens* ChiB: 1UR9], [Arthobacter psychrophilic chitinase: 1KFW] and [YKL-40: 1NWT] (Mohanty *et al.*, 2003; Vaaje-Kolstad *et al.*, 2004; Papanikolau *et al.*, 2001); (Fusetti *et al.*, 2003).

As pointed out before, chitinases from both families are further divided into exo- and endochitinases. Exochitinase activity represents a progressive action that starts at the non-reducing end of chitin chain and successively releases chitobiose (NAG₂) units, whereas, endochitinase activity involves random cleavage at internal points within a chitin chain (Robbins *et al.*, 1988). A correlation between the exo-endo chitinolytic feature and architecture of the active site has been established. Indeed, the family-18 endochitinases, such as *S. marcescens* ChiA, *Hevea brasiliensis* chitinase (hevamine) and human CHIT1 are groove-like structures with openings above and at both ends (Brameld & Goddard, 1998); (Hart *et al.*, 1995). In contrast, the active sites of exochitinases, such as *S. marcescens* ChiB, have tunnel-like morphologies (van Aalten *et al.*, 2001).

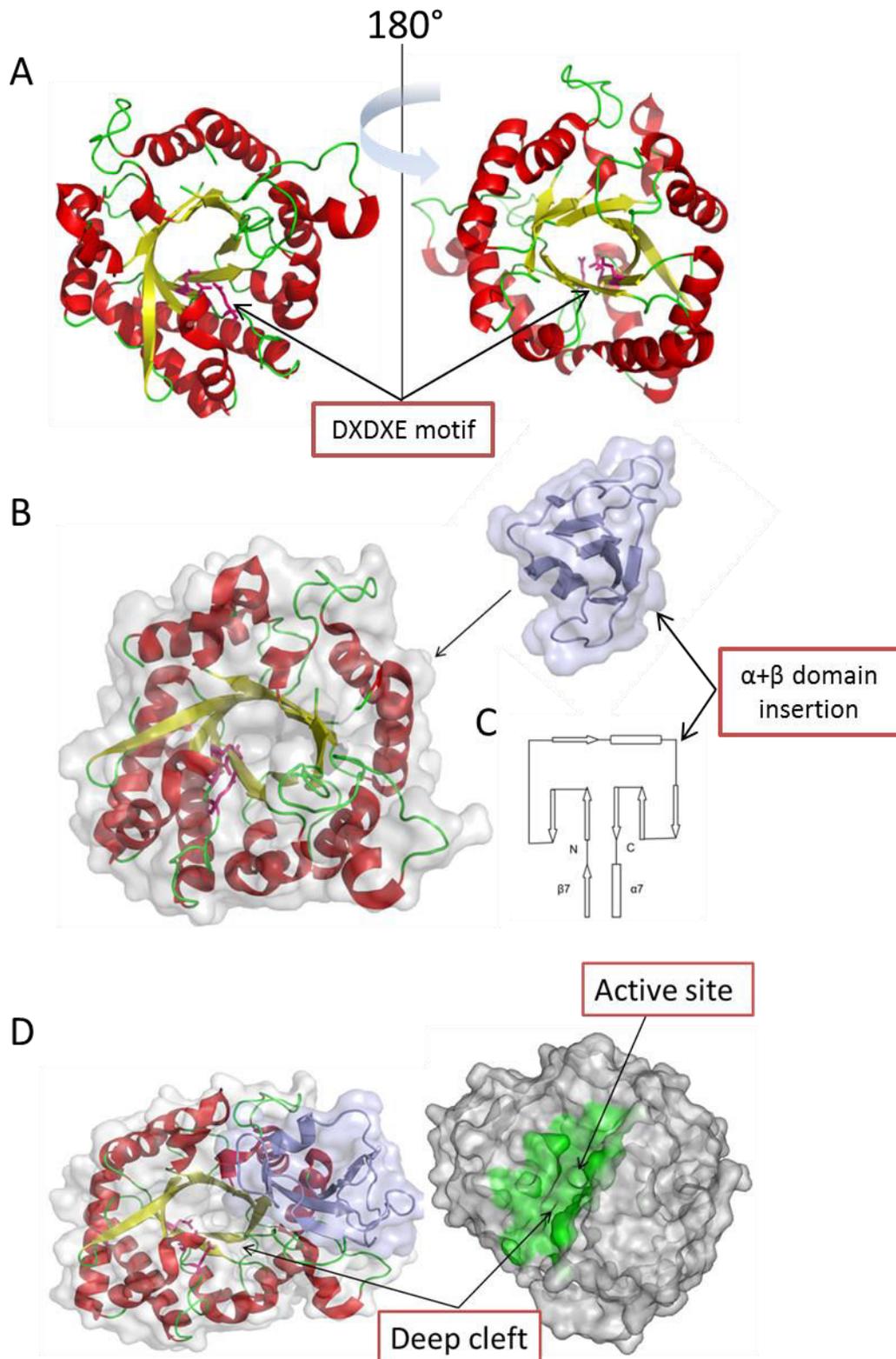


Figure 11. A) Cartoon representation of the core domain of CHIT1 (PDB code: 1GUV) (Fusetti *et al.*, 2002a) consisting of an $(\alpha/\beta)_8$ TIM-barrel fold. The three amino acids of the conserved DXDXE motif are colored in pink. B) Cartoon and schematic representation of the $(\alpha + \beta)$ domain insertion domain between $\beta 7$ and $\alpha 7$ on the TIM barrel, which is composed of two anti-parallel β -strands followed by one β -strand, one short α -helix, and lastly three anti-parallel β -strands. C) The arrows indicate β -strands and the rectangles are α -helices. The lines stand for the loops connecting α -helices or β -strands. D) Overall structure of the CHIT1 catalytic domain structure comprising the TIM barrel with the $(\alpha + \beta)$ domain insertion leading to the formation of a deep groove in the catalytic site.

1.2.2 Role of the conserved aromatic residues in the substrate binding feature of CHIT1 and its comparison within GH18 chitinases

The structure of CHIT1-CAT has revealed the presence of stretches of aromatic residues forming two conserved aromatic motifs. The first motif (Trp-Tyr-Trp-Trp) is lined along the binding cleft (subsites -6 to -1) including Trp71, Tyr34, Trp31, Trp358 in CHIT1 (Fig. 12) (Table 4). The second motif, which consists of two consecutive tryptophans, appears in the +1 and +2 subsites and corresponds to Trp99 and Trp218 in CHIT1 (based on the nomenclature proposed by Davies *et al.* (Davies *et al.*, 1997) (Fig. 12, 13) (Table 4). Interestingly, besides the overall similarity between CHIT1 and AMCase structures, both aromatic motifs are identical and located in the same positions pointing to the high conservation of the substrate-binding cleft in human chitinases (Fig. 12A, B). In addition, these aromatic motifs are highly conserved across species in several well-studied GH18 chitinases (Fig. 12 and Table 4), suggesting that many GH18 enzymes from different species share common substrate binding features. In this respect, the tryptophan in the -1 subsite (Trp358 in CHIT1) is suggested to act as an “anvil” onto which the -1 NAG is stacked. This stacking interaction is coordinated with specific H-bonds with other residues contributing to accommodate the -1 sugar in the boat conformation required for achieving the hydrolysis (Fig. 12, 13). In some GH18 chitinases like in hevamine and ChiC, residues from this aromatic motif are not fully conserved. However, the equivalent Trp358 (in CHIT1) appears in all GH18 chitinases. The mutation of this residue abolishes indeed the chitinase activity (van Aalten *et al.*, 2000); (Papanikolaou *et al.*, 2001); (Fusetti *et al.*, 2002a); (Songsiriritthigul *et al.*, 2008); (Yang *et al.*, 2010) (Table 4).

It is interesting to note that Met300 forms a hydrophobic interaction with the -2 sugar by stacking against the methyl of its *N*-acetyl group (Fusetti *et al.*, 2002b); (Olland *et al.*, 2009b). It has been reported that this amino acid, together with neighboring aromatic residues from the TIM domain, may constitute part of the hydrophobic core of the substrate-binding site (Li & Greene, 2010) (Table 5).

While the structures of the catalytic domains of CHIT1 and AMCase (Fusetti *et al.*, 2002b); (Olland *et al.*, 2009b) are known, there are no structures of complexes with extended ligands providing sufficient comprehensive insight into the subsite arrangement

within the enzyme cleft (Eide *et al.*, 2013). In this regard, the crystallographic data from ChiA (*S. marcescens*) has shown that a chito-octamer binds from -6 to +2 subsites where the exposed aromatic residues lining the binding cleft stack against the hydrophobic face of the sugars (Papanikolau *et al.*, 2001). As the structural arrangement of these exposed aromatic residues is significantly conserved between ChiA and CHIT1, this finding has contributed to understand the binding characteristics of the substrate in the catalytic cleft of CHIT1. Hence, structure comparison based on the structural data provided by the ChiA-NAG₈ complex have led to suggest a binding model of CHIT1 with a NAG polymer underlying the stacking role of the conserved aromatic motifs (Fusetti *et al.*, 2002a) (Fig. 14). Furthermore, thermodynamic and substrate degradation studies of ligand binding between CHIT1 and ChiA have suggested that CHIT1 degrades chitin with the same directionality as ChiA from the non-reducing end towards the reducing end (Hult *et al.*, 2005); (Zakariassen *et al.*, 2009) (Fig. 14). Recently, mass spectrometry studies on CHIT1 substrate affinity have demonstrated a preference affinity on +3 subsite from the reducing end (Eide *et al.*, 2013).

Chitinase	Aromatic residue ID						
	-6	-5	-4	-3	-1	+1	+2
Subsite position							
ChiB	-		-N		Trp403	Trp97	Trp220
ChiC	Tyr78		-	Trp34	Trp300	-	Trp223
ChiA	Phe232		Tyr170	Trp167	Trp539	Trp275	Phe396
AMCase	Trp71		Tyr34	Trp31	Trp360	Trp99	Trp218
CHIT1	Trp71		Tyr34	Trp31	Trp358	Trp99	Trp218
OfChitI-CAD	Phe61		Tyr37	Trp34	Trp372	Trp107	Trp241
ChiA1	Trp122		Tyr56	Trp53	Trp433	Trp164	Trp285
<i>B. circulans</i>							
Chitinase A	Trp231		Tyr171	Trp168	Trp570	Trp275	Tyr397
<i>V. harveyi</i>							
AfChiB	Trp92		Tyr55	Trp52	Trp384	Trp137	Phe251
CiX1	Phe71		Tyr50	Trp47	Trp378	Trp134	
CrChi1	Phe74		Tyr53	Trp50	Trp381	Trp131	Trp218

Table 4. Conserved aromatic residues in the positions of the motif WYWW and WW.

Table 5. Conserved residues in the substrate binding sites of GH18 chitinase PDB codes of 1GUV, 1D2K, and 1ITX and their proposed roles. PDB code 1GUV: CHIT1; 1D2K: *C. immitis* chitinase; 1ITX: *B. circulans* chitinase A1.

Code	Interaction with the substrate				Formation of the hydrophobic core				
1GUV	Tyr267	Arg269	Glu297	Met300	Tyr303	Val306	Ala312	Val332	Phe334
1D2K	Tyr293	Arg295	Glu316	Val319	Tyr322	Met325	Ala330	Ile352	Tyr354
1ITX	Tyr338	Arg340	Glu366	Ser369	Phe372	Leu375	Tyr385	Ile407	Tyr409

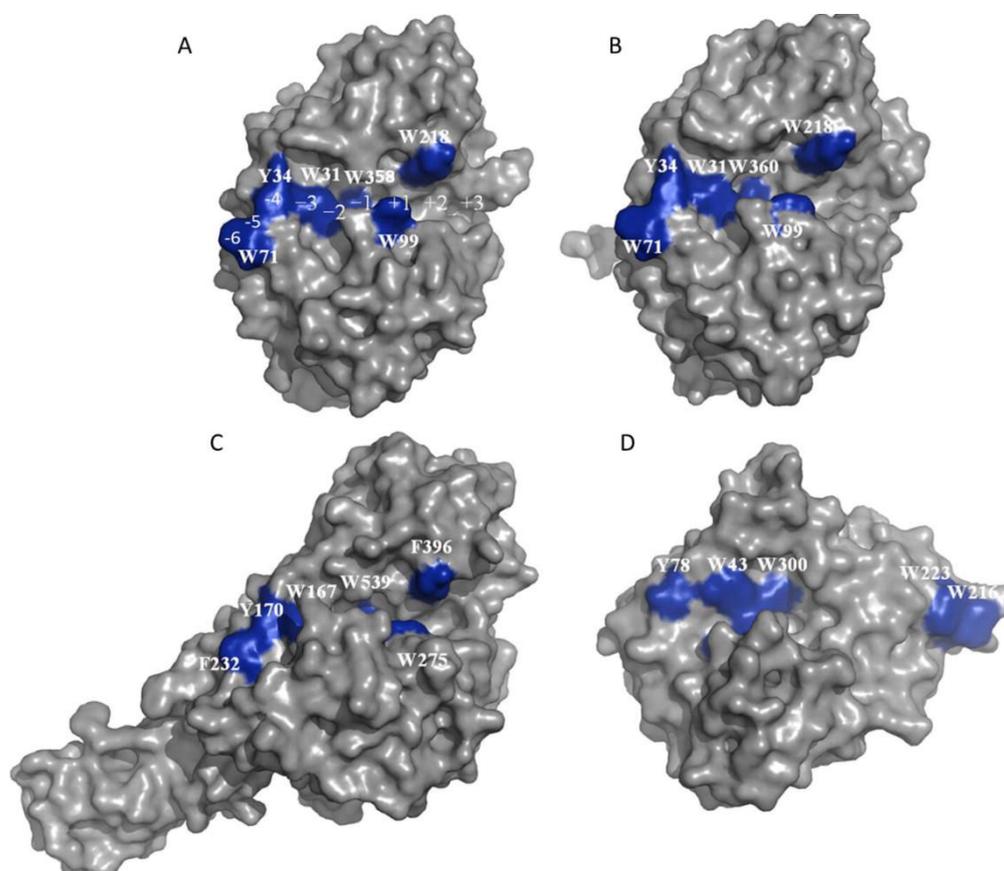


Figure 12. A) Crystal structures of CHIT1 (PDB code: 1GUV); B) AMCase (PDB code: 3FXV); C) ChiA from *S. marcescens* (PDB code: 1EHN) D) ChiC (PDB code: 4AXN). Aromatic amino acids lining the substrate-binding cleft and known to be important determinants of enzyme properties are colored blue. Individual subsites are marked for CHIT1 dividing the catalytic cleft into +n and -n subsites according to the nomenclature proposed by Davies *et al.* (Davies *et al.*, 1997). Adapted from ref. (Eide *et al.*, 2013)

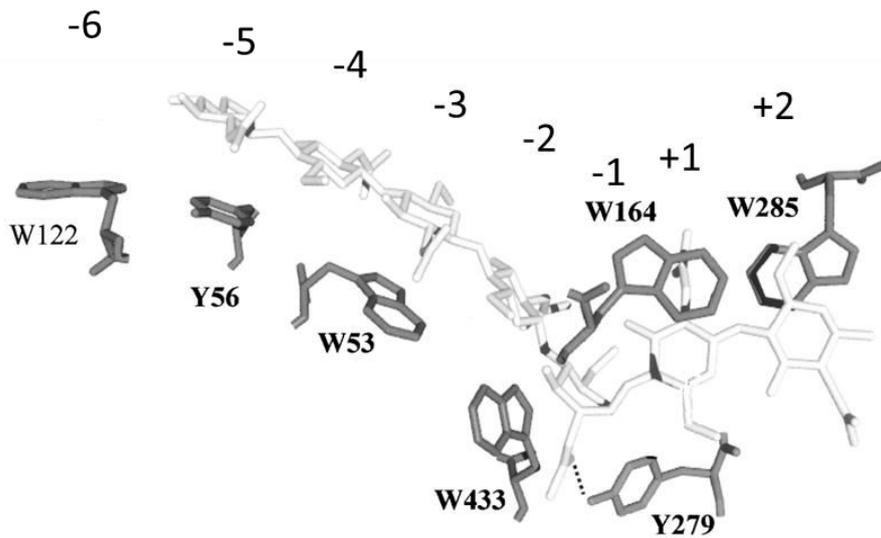


Figure 13. Stick model of enzyme-substrate interactions in GH18 chitinases. These interactions are highly conserved in this enzyme family. The example model comes from *B. circulans* chitinase A1. Adapted from (Watanabe *et al.*, 2003). The first motif, WYWW lined along the binding cleft corresponding to subsites -6 to -1 (Table 4). The second motif, which consists of two tryptophanes, appears in the +1 and +2 subsites and corresponds to W164 and W285 equivalent to W99 and W218 in CHIT1. The hydrolysis occurs between the -1 and +1 NAG. W433 serve as an “anvil” contributing to the distortion of the -1 NAG into a boat configuration. In some GH18 chitinases, tryptophanes from these two motifs (except the W433 equivalent to W358 in CHIT1) could be substituted into other aromatic residues which keep the substrate binding pattern relatively similar. Adapted from ref (Watanabe *et al.*, 2003)

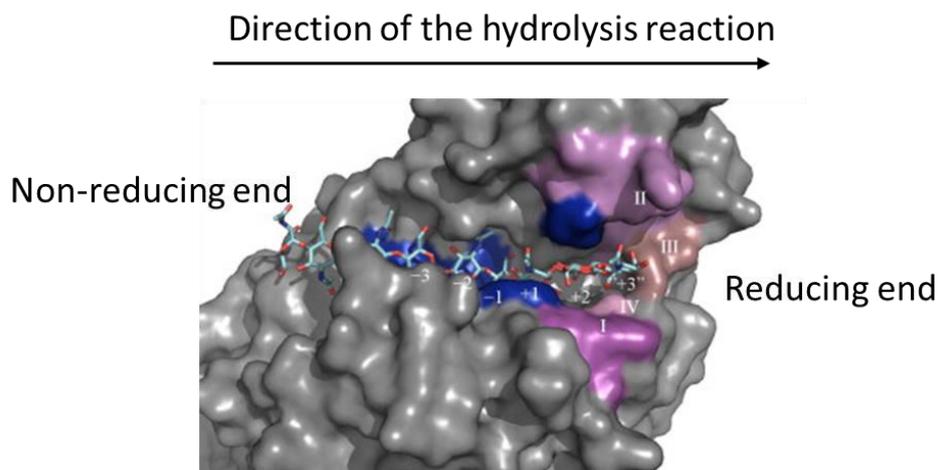


Figure 14. Crystal structure of the active site of ChiA in complex with a (GlcNAc)₈ molecule bound to subsites -6 to +2, with numbered subsites (PDB code: 1EHN) (Papanikolaou *et al.*, 2001). Four aromatic amino acids interacting with the ligand and located near the catalytic center are colored blue. These are: Tyr170 (-4) Trp167 (-3), Trp539 (-1/-2), Trp275 (+1), and Phe396 (+2). From ref. (Norberg *et al.*, 2011)

1.2.3 Role of conserved residues within the active site of GH18 chitinases in the hydrolysis mechanism

Besides the aromatic motifs and the hydrophobic core, active GH18 chitinases have seven residues at the bottom of the catalytic groove which are highly conserved. In CHIT1 these seven residues are Ser181, Tyr27, Asp136, Asp138, Glu140, Tyr212, Asp213 and correspond to Tyr10, Ser93, Asp140, Asp142, Glu144, Tyr214, Asp215 in ChiB where the mechanistic studies were mostly performed. Mutation of subsets of these conserved residues in various GH18 chitinases has shown for the majority that they are important for catalysis (Bokma *et al.*, 2002); (Watanabe *et al.*, 1993); (Watanabe *et al.*, 1994); (Watanabe *et al.*, 1994); (Lin *et al.*, 1999); (Lu *et al.*, 2002); (Kolstad *et al.*, 2002). To illustrate the reported roles of these amino acids, it is important to introduce some notions related to the mechanism of carbohydrate catalysis.

1.2.3.1 Inverting mechanism vs retaining mechanism

Enzymatic hydrolysis of the glycosidic bond occurs via general acid-base catalysis that requires two critical residues: a proton donor and a nucleophile/base (Koshland, 1953); (Sinnott, 1990). Most GH proteins follow one of the two possible general mechanistic pathways during the hydrolysis reaction. These are called the retaining and inverting mechanisms. They depend on the configuration of the anomeric oxygen resulting from the outcome of the hydrolysis. The inverting mechanisms involves an inversion of the stereochemistry from equatorial to axial configuration and vice-versa (Tews *et al.*, 1997) (Fig. 15a), while the retaining mechanism preserves the stereochemistry of the anomeric oxygen at C1 of the resulted product in a configuration similar to the initial one i.e. equatorial substrate to equatorial product, or axial substrate to axial product (Koshland, 1953); (Sinnott, 1990); (Tews *et al.*, 1997) (Fig. 15b).

Both retaining and inverting enzymes use two carboxylate groups provided by either a glutamate or an aspartate amino acid residue. The inverting GH enzymes require a catalytic acid residue and a catalytic base residue while those acting in a retaining manner contain a general acid/base residue and a nucleophilic residue (Vuong & Wilson, 2010) (Fig. 15a, b).

The acid catalytic residue is the proton donor which makes a hydrogen-bond (H-bond) with the glycosidic oxygen of the substrate to be hydrolyzed. The position of the proton donor residue is identical in both retaining and inverting GHs. However, the nucleophilic catalytic residue is located differently. In the retaining enzymes, this residue is in close vicinity to the sugar anomeric carbon, whereas the basic nucleophilic catalyst is more distant in the inverting enzymes, accommodating thereby a water molecule between the base and the sugar moiety (McCarter & Withers, 1994); (Sinnott, 1990). In inverting GHs, the hydrolysis reaction occurs in one step. The catalytic acid residue donates a proton to the glycosidic oxygen atom of the leaving group while the second catalytic base residue receives a proton from a water molecule, increasing its nucleophilicity and assisting its attack on the anomeric center of the sugar. The nucleophilic attack is followed by a subsequent cleavage of the substrate, which yields a product with opposite stereochemistry relative to the substrate (Vuong & Wilson, 2010); (Davies & Henrissat, 1995); (Sinnott, 1990).

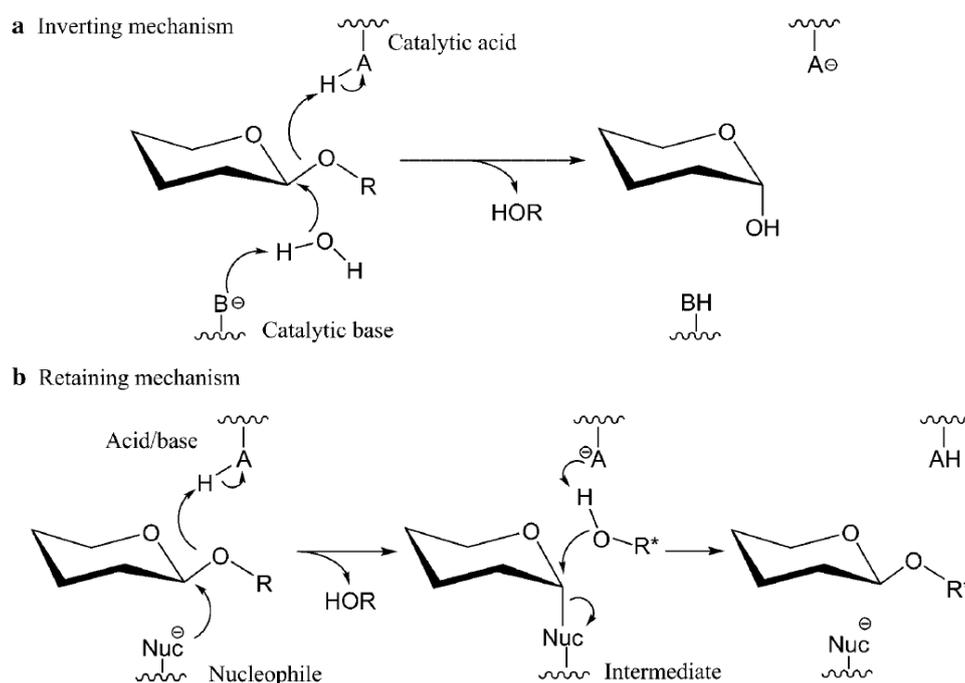


Figure 15. Proposed inverting (a) and retaining (b) mechanisms. AH: a catalytic acid residue, B⁻: a catalytic base residue, Nuc: a nucleophile, and R: a carbohydrate derivative. HOR: an exogenous nucleophile, often a water molecule (Vuong & Wilson, 2010); (Sinnott, 1990).

In retaining GHs, the hydrolysis reaction occurs in two steps. In the first step (called the glycosylation step), the nucleophilic residue attacks the anomeric carbon to displace the leaving group and forms in the transition state a covalent enzyme-substrate intermediate. At the same time, the general acid/base catalytic residue donates a proton to the glycosyl oxygen atom releasing the leaving group. The second step, called deglycosylation, leads to the cleavage of the enzyme-substrate intermediate. The deprotonated acid/base residue, which has donated the proton during the first step, acts in the second step as a base deprotonating a water molecule and assisting its nucleophilic attack on the anomeric C1 of substrate, resulting in the release of the product and the regeneration of the enzyme (Davies & Henrissat, 1995); (Vuong & Wilson, 2010); (Koshland, 1953).

It has been reported that GH19 chitinases hydrolyze the chitin chain through an inverting mechanism, while GH18 chitinases degrade chitin through a retaining mechanism (Iseli *et al.*, 1996); (Tews *et al.*, 1997); (Synstad *et al.*, 2004).

1.2.3.2 Substrate-assisted chitinolytic mechanism

With the development of X-ray crystallography, the resolution of the 3D structures of the GH enzymes has widely contributed to validate the predicted catalytic residues and to understand their mode of action (Henrissat & Bairoch, 1993). Indeed, the crystal structures of many GH18 chitinases, supported by site-directed mutagenesis studies, have revealed the presence of a catalytic glutamate residue acting as an acid/base residue at the end of the aforementioned β 4-strand. However, the structural data have revealed the absence of the second nucleophilic residue to stabilize the intermediate state. This observation was therefore not consistent with the classical view of GHs retaining enzymes. Thus, in order to understand how this enzymatic family performs hydrolysis in a retaining manner with only one acid/base catalytic residue, extensive enzymatic and structural analysis efforts have been pursued. In this regard, the X-ray structures of two chitinases, hevamine and Chitinase A, in the apo form, bound to chitotetraose and to NAG₈ respectively, have revealed a distortion of the -1 sugar of the substrate upon binding to the active site (Tews *et al.*, 1997); (Terwisscha van Scheltinga *et al.*, 1994; Papanikolau *et al.*, 2003); (Papanikolau *et al.*, 2001). This distortion consists of the accommodation of the -1 sugar in boat conformation (Fig. 13).

Afterwards, the comparison of these enzyme-substrate 3D structures with the structure of hevamine in complex with allosamidin, an inhibitor mimicking the intermediate, has contributed to observe special structural features regarding the interaction between the GH18 enzymes and the substrate. Accordingly, the distortion of the -1 NAG allows its N-acetyl group to accommodate its position within the active site in a way that its carbonyl oxygen site points toward the anomeric C1 belonging to the same -1 sugar (Fig. 16). Therefore, some authors have suggested that when the catalytic glutamate acts as acid and protonates the glycosidic oxygen between the -1 and +1 NAG, this leads to its cleavage and displaces the +1 NAG. Simultaneously, an intramolecular nucleophilic attack is performed by the N-acetyl group of the -1 NAG on the anomeric C1, thereby enhancing the formation of the intermediate. As shown in fig. 16 and 17, this occurs during the transition state where a covalent bond is formed between the carbonyl oxygen of the C2 N-acetyl group and the anomeric C1 leading to the appearance of an oxazolinium ion intermediate. Thus, all these findings suggest that the nucleophile 'missing' from the structure is provided by the substrate itself (van Aalten *et al.*, 2001); (Synstad *et al.*, 2004); (Terwisscha van Scheltinga *et al.*, 1995); (Tews *et al.*, 1997); (Bokma *et al.*, 2002); (Brameld & Goddard, 1998).

This unusual mechanism for the hydrolysis reaction in GH18 chitinases was called substrate-assisted chitinolytic mechanism. It satisfies the condition of retained stereochemistry of hydrolysis products. At the same time, it is different from the classical mechanism that requires two catalytic residues in the active site. The results of further studies on hevamine (Bokma *et al.*, 2002) and chitinase B from *S. marcescens* (van Aalten *et al.*, 2001), in addition to modeling studies (Brameld & Goddard, 1998), all support a mechanism that includes the formation of an oxazolinium ion (Fig. 16).

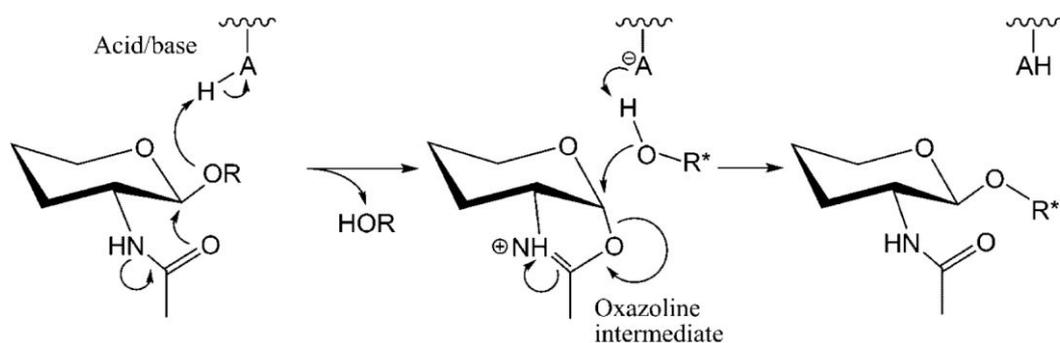


Figure 16. Substrate-assisted mechanism for retaining GHs. A 2-acetamide group in the substrate acts as a nucleophile to form an oxazoline intermediate.

1.2.3.3 The coordinated role of a conserved catalytic triad in GH18 chitinases

In 2001, the crystallographic studies on the GH18 *S. marcescens* Chitinase B (ChiB) gave new insights regarding the catalytic mechanism of chitinases. This prompted Van Aalten *et al.* to develop a mechanistic model to support the original ideas based on the substrate distortion and assistance. Importantly, the novelty in the catalytic model of Van Aalten lies in that it involves three critical catalytic residues, namely Asp140, Asp142 and Glu144, (Asp136, Asp138, Glu140 in CHIT1 respectively) thereby providing a better understanding regarding the active site mode of action (van Aalten *et al.*, 2001).

Notably, these three residues correspond to the highly conserved amino acids in the consensus sequence DXDXE, the signature of active GH18 chitinases. Hence, although the active GH18 chitinases from different organisms share low sequence similarity (<http://www.sanger.ac.uk/Software/Pfam>), their 3D folding is very conserved (Fig. 11), with high similarity in the binding cleft and more importantly they share the same catalytic triad (Fusetti *et al.*, 2002b); (Hollis *et al.*, 2000); (Watanabe *et al.*, 2001); (Perrakis *et al.*, 1994); (Terwisscha van Scheltinga *et al.*, 1994) and (van Aalten *et al.*, 2000). Therefore, it has been suggested that these chitinases share a common catalytic mechanism. Consequently, the mechanistic analysis on this triad in one GH18 chitinase might be applicable to the other active chitinases. As ChiB is one of the best studied GH18 enzymes and shares the same catalytic engine (Asp140, Asp142 and Glu144) observed in this enzyme family, ChiB has been considered as an appropriate reference for discussing the catalytic mechanism of all the other active family 18 chitinases including the human ones, CHIT1 and AMcase.

The catalytic mechanism of active family 18 chitinases proposed by Van Aalten *et al.* follows the nomenclature residues of ChiB. According to their mechanism, the enzyme in the apo form shows the central Asp142 making a H-bond contact with Asp140 and accommodated thereby in a position spatially separated from the protonated catalytic Glu144. The binding of the substrate causes the rotation of the protonated Asp142 towards Glu144 enabling H-bonding between these two residues and lowering Glu144 pK_a. The rotation of the Asp142 is important to stabilize the distortion of the pyranose ring of the -1 NAG to a skewed boat conformation by H-bonding with the acetamido-group. The Glu144 which has now acquired a lower pK_a donates the proton to the glycosidic oxygen of the

scissile bond between -1 and +1 sugar leading to the formation of an oxazolinium ion intermediate stabilized by Tyr214 and Asp142 (Fig. 17). At the same time, the deprotonated Glu144 is stabilized by Asp142. Ultimately, Glu144 receives a proton from the hydrolytic water molecule, increasing its nucleophilicity and assisting its attack on the anomeric carbon of -1 NAG. At the end of the reaction, Asp142 rotates to its original position where it interacts with Asp140. It has been suggested that a 'flip'-like conformational change of Asp142 may also play an important role in adjusting the pK_a of the catalytic Glu during the catalysis (van Aalten *et al.*, 2001); (Oku & Ishikawa, 2006).

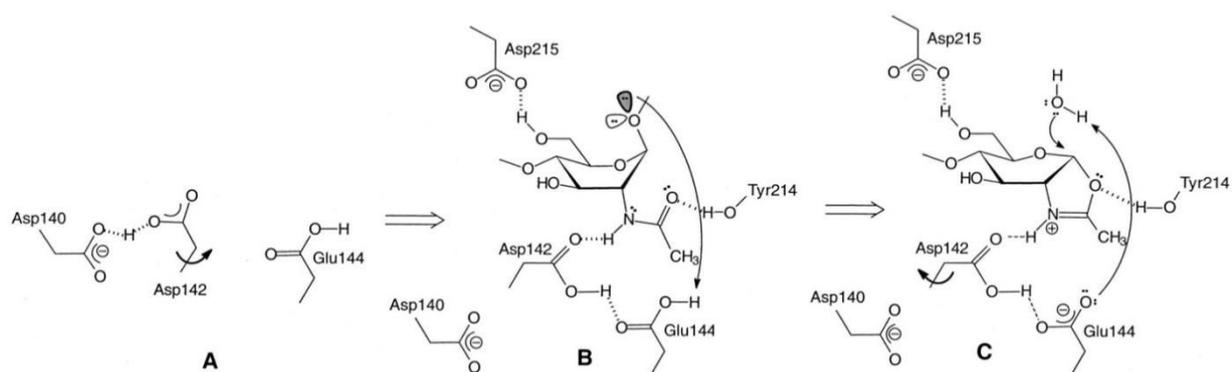


Figure 17. Proposed catalytic mechanism in GH18 family chitinase. Asp140, Asp142, and Glu144, conserved in most GH18 chitinases, are shown during separate stages of the hydrolysis reaction. A) Resting enzyme. Asp142 is too far away to interact with Glu144. B) Binding of substrate (only -1 NAG residue is shown) causes distortion of the pyranose ring to a boat or skewed boat conformation (see also Fig. 2) and rotation of Asp142 toward Glu144, enabling H-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). From ref. (van Aalten *et al.*, 2001)

Most of the recent mechanistic studies on GH18 chitinases from human to bacteria, including X-ray crystallography, site-directed mutagenesis and computational analysis studies, support the catalytic model of Van Aalten *et al.* (van Aalten *et al.*, 2001; Synstad *et al.*, 2004); (Vaaje-Kolstad *et al.*, 2004); (Bokma *et al.*, 2002); (Suginta & Sritho, 2012); (Papanikolaou *et al.*, 2003); (Malecki *et al.*, 2013); (Yang *et al.*, 2010). It is worth to notice that a recent study in ChiA on the catalytic triad, based on mutation of the middle aspartate in ChiA to Asn has extended the attributed roles of this aspartate and concluded that it has multiple functions in the catalytic mechanism of GH18 chitinases. According to this study the middle aspartate (i) affects the positioning and nucleophilicity of the acetamido-group involved in catalysis, (ii) interacts with the catalytic acid, presumably contributing to the

necessary cycling of its pKa during catalysis and thus to the positioning and/or activation of the catalytic water molecule, and (iii) interacts with and stabilizes the oxazolinium ion intermediate (Zakariassen *et al.*, 2011).

The work on the catalytic mechanism in GH18 chitinases has extended to the study of additional highly conserved residues surrounding the catalytic triad. Indeed, these residues were shown to be involved in the interaction with residues from the catalytic triad and in the stabilization of the substrate distortion. According to the nomenclature of ChiB, these amino acids are Ser93, which belongs to the conserved motif SXGG, Tyr10, Tyr214 and Asp215/Asn215 (Ser181 and Tyr27, Tyr212, Asp213 in CHIT1). Ser93 and Tyr10 are located at the bottom of the catalytic core (Fig. 18). The crystallographic observations show that rotation of protonated Asp142 towards Glu144 upon binding of the substrate leads to adjustments in the positions of Tyr10 and Ser93 that have two major consequences. First, the adjusted side chains of Tyr10 and Ser93 partly fill the space left by the rotation of Asp142. Second, and possibly more important, these adjustments allow these residues to protonate Asp140 via strong H-bond, thus compensating in part for the negative charge of the latter (Vaaje-Kolstad *et al.*, 2013); (Papanikolau *et al.*, 2001); (Bokma *et al.*, 2002).

Regarding Tyr214 contribution, as shown in (Fig. 18), the phenolic hydroxyl of the Tyr214 side chain stabilizes the acetamido-group in the distorted conformation, therefore promoting the nucleophilic attack of this group on the anomeric carbon. Several of the mutational studies cited above have shown that mutation of Tyr214 (to Phe or Ala) has a detrimental effect on the catalytic efficiency (Vaaje-Kolstad *et al.*, 2013); (Jitonnom *et al.*, 2014)

Asp215 is involved in H-bonding -1 sugar in its distorted conformation (Fig. 18). It has been suggested that the negative charge on this residue contributes to increase the pKa of the catalytic amino acid Glu144, especially in the enzyme–substrate complex where the glycosidic oxygen is located between the carboxyl groups of Asp215 and Glu144. In this regard, mutation of Asp215 to Asn (which can still form a H-bond with the substrate) has only a small effect on the maximum activity but leads to an acidic shift of the pH optimum for activity. In contrast, mutation of Asp215 to Ala decreases the activity which points the importance of this residue in the catalytic activity (Synstad *et al.*, 2004).

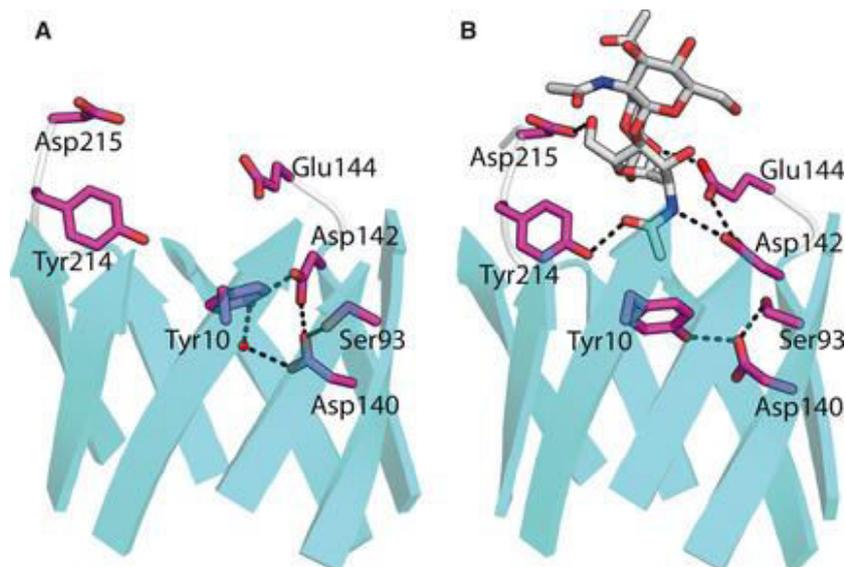


Figure 18. Key elements of the catalytic machinery of family 18 chitinases. The picture shows two situations that have been observed by crystallography for ChiB (van Aalten *et al.*, 2001). Main chains are shown in cartoon representation and illustrate the β -barrel structure of the catalytic domain. Carbon atoms of the side chains and substrate (in B) are colored magenta and grey, respectively. Asp140, Asp142 and Glu144 are part of strand 4 and comprise the diagnostic DXDXE sequence motif. In the apo form structure A), Asp140 and Asp142 share a proton. Upon substrate binding B), Asp142, with its proton, rotates away from Asp140. This rotation is accompanied by changes deeper down in the barrel that stabilize the now 'lonely' (and partly buried) Asp140 (Synstad *et al.*, 2004) (Vaaje-Kolstad *et al.*, 2013). From ref (Vaaje-Kolstad *et al.*, 2013)

1.2.3.4 Transglycosylation

In 2003, Aguilera *et al.* reported that CHIT1 has a marked rate of transglycosylation activity, especially under high concentration of substrate (Aguilera *et al.*, 2003). As described before, CHIT1 is highly elevated in many diseases; therefore monitoring its activity is exploited for clinical diagnostic and evaluation of therapeutic efficacy. To do so, the CHIT1 enzymatic activity is detected by artificial chromogenic or fluorogenic substrates, e.g. 4-methylumbelliferyl di-N-acetyl- β -D-chitobiose (4-MU-chitobiose), used in (Aguilera *et al.*, 2003). Once cleaved by CHIT1, the substrate loses the fluorescent 4-MU moiety (Fig. 19). During the measurements of CHIT1 activity, a strong decrease of the fluorescence intensity was detected upon incubation of the enzyme under saturating substrate concentrations (Fig. 19A). This has caused to the clinicians a considerable limitation regarding the reliability of the enzymatic activity evaluation. This fluorescence emission decrease was explained by the transglycosylation which consists on the repolymerization of a chitobiose unit generated by

the previous hydrolytic cycle to a new 4-MU-chitobiose molecule resulting in the lack of a fluorescent product (Fig. 19B, 220) (Aguilera *et al.*, 2003).

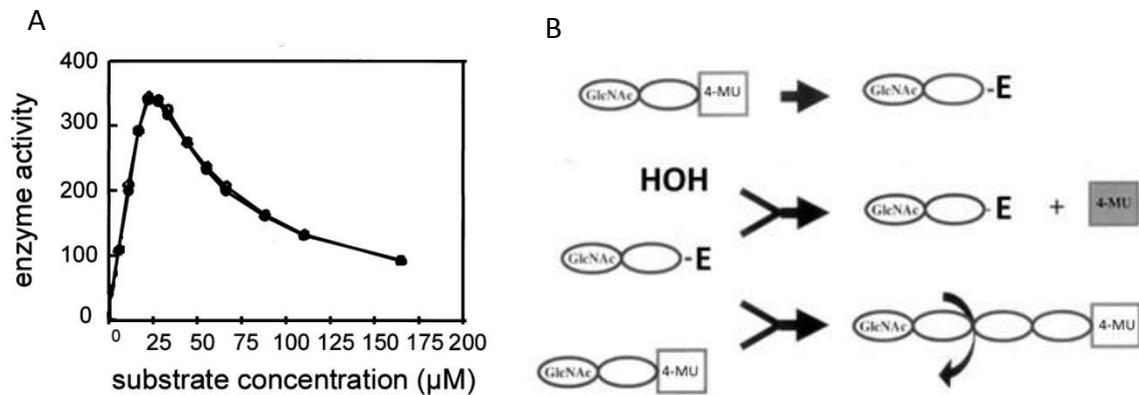


Figure 19. A) Absolute values for CHIT1 activity with 4-MU-chitobiose substrate. B) With increasing concentration of 4-MU-chitobiose there is an increasing formation of 4-MU-chitotetraose by transglycosidase activity. Chitobiose units can be subsequently transferred from 4-MU-chitotetraose to 4-MU-chitobiose molecules, a futile cycle generating no new products and not releasing fluorescent 4-MU. The experimenter observes this process as inhibition of enzyme activity, since that is monitored by formation of fluorescent 4-MU. GlcNAc: N-aceyl-glucosamine; E:CHIT1 enzyme From ref. (Aguilera *et al.*, 2003)

The demonstration of a marked transglycosidase capacity in CHIT1 was not entirely surprising, as it had been previously described for other GH18 chitinases (Sharon & Seifter, 1964); (Kravchenko, 1967); (Chipman *et al.*, 1968); (Aronson *et al.*, 2006); (Lu *et al.*, 2009); (Zakariassen *et al.*, 2009); (Zakariassen *et al.*, 2011). To understand the enzymatic features that control and lead to the occurrence of such activity, a combination of approaches have been applied consisting of site-directed mutagenesis, structural analysis and computational calculations (Fig. 20).

These approaches were performed on many GH18 chitinases as they share closely similar properties. According to the reported conclusions from these studies, it has been suggested that in transglycosylation a special architecture of the catalytic site could disfavor the positioning of the hydrolytic water responsible for the nucleophilic attack in the hydrolysis cycle (Aronson *et al.*, 2006); (Umemoto *et al.*, 2013). This special architecture, which involves the existence of aromatic residues in the product release subsites, favors therefore binding of incoming chito-oligosaccharide through strong stacking interactions (Fig. 21).

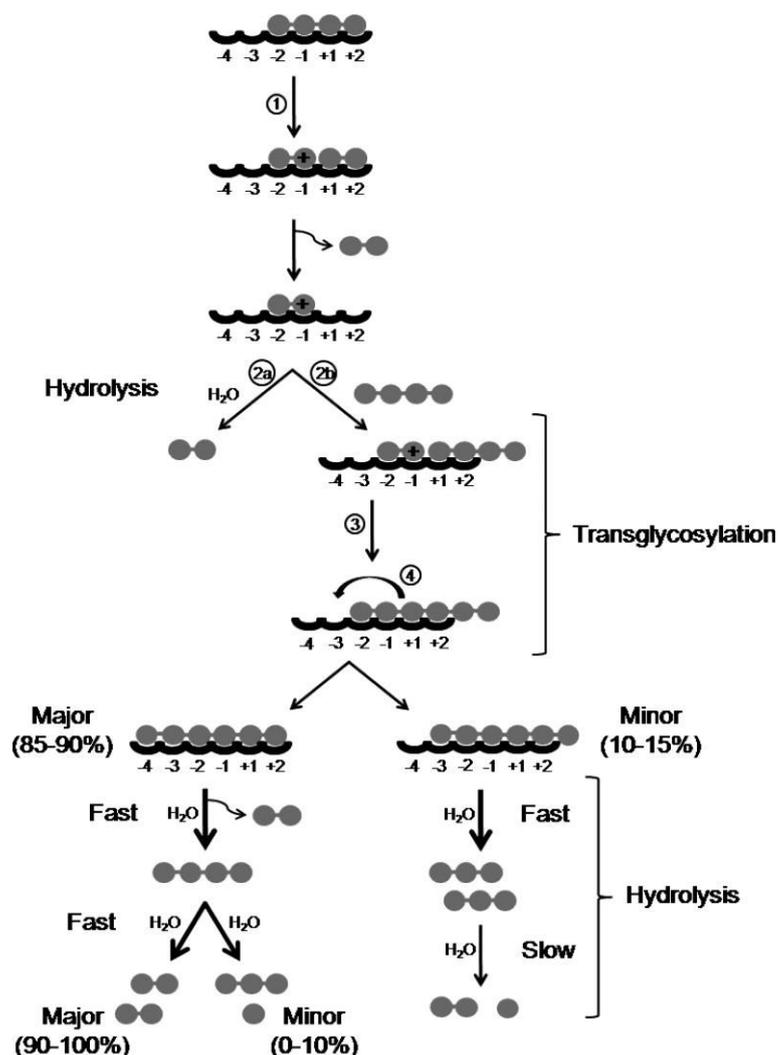


Figure 20. Schematic depiction of a possible transglycosylation reaction using $(\text{GlcNAc})_4$ as substrate. GlcNAc residues are shown as gray circles, and the binding subsites of GH18 chitinases are numbered. The $(\text{GlcNAc})_4$ substrate binds to the -2 to +2 subsites, and the glycosidic bond is cleaved to form a positive charged “ $(\text{GlcNAc})_2$ ” oxazolinium ion intermediate (marked with +) in the -1 and -2 subsites (step 1). The $(\text{GlcNAc})_2$ aglycon product then exits and is replaced by an incoming water (step 2a) or acceptor $(\text{GlcNAc})_4$ (step 2b). In the second case (transglycosylation), the O4 of the nonreducing-end GlcNAc of $(\text{GlcNAc})_4$ completes the reaction by nucleophilic attack on C1 of the glycosyl in the -1 site and forms $(\text{GlcNAc})_6$ (step 3). This transglycosylation product is released and may after rebinding (step 4) be subjected to hydrolysis or further transglycosylation reactions. Further transglycosylation may lead to production of longer oligosaccharides of “any” length. These longer oligomers will be good substrates, and they will eventually be converted to trimers and dimers by hydrolysis but may yield novel tetramers along the way. From ref. (Zakariassen *et al.*, 2011)

By correlating the highlighted roles of the aromatic residues and the multiple interpreted functions of the middle aspartate, the hydrolysis model proposed by Van Aalten *et al.* has been extended by Zakariassen *et al.* with a suggested mechanism for transglycosylation.

According to the mechanism proposed by Zakariassen *et al.*, when the catalytic glutamate acts as a general acid by protonating and cleaving the glycosidic oxygen, an oxazolinium ion intermediate is formed simultaneously. Upon the arrival of an acceptor chito-oligosaccharide molecule, this latter attacks the intermediate ion and the glutamate, which has acted as a general acid in the previous step, functions in this step as a general base leading to the generation of a glycosidic bond between the acceptor molecule and the previously cleaved substrate (Zakariassen *et al.*, 2011) (Fig. 21, 22). Moreover, transglycosylation has attracted significant attention because it has been proposed as a technology for oligosaccharide enzymatic synthesis, considered advantageous over synthetic chemical strategies, especially in terms of time and cost efficiency (Zakariassen *et al.*, 2011). Thus, it is considered that understanding the enzymatic mechanism of transglycosylation is of high interest to be able to engineer mutant hydrolase glycosylase enzymes with lower hydrolysis activity and pronounced hypertransglycosylation to produce longer oligosaccharide of specific length and sequence of acetylated and deacetylated units (Zakariassen *et al.*, 2011).

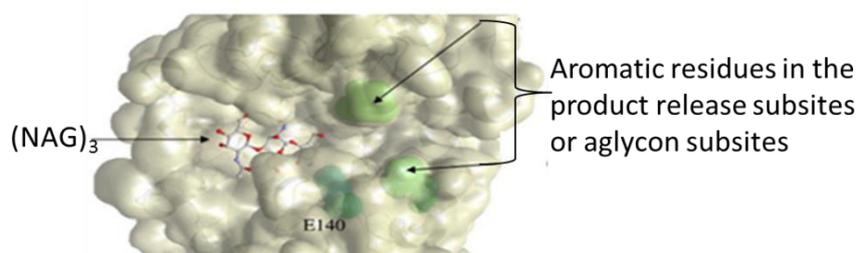


Figure 21. Active site of GH18 chitinases occupied by a polymer of three N-acetyl glycosamine (NAG₃) in the -1 to -3 subsites. The catalytic glutamate corresponding to E140 in CHIT1 is colored in green. Two aromatic residues (colored in green) in the product release subsites (positive subsites) favor the binding of incoming chito-oligosaccharide through strong stacking. Adapted from ref. (Taira *et al.*, 2010)

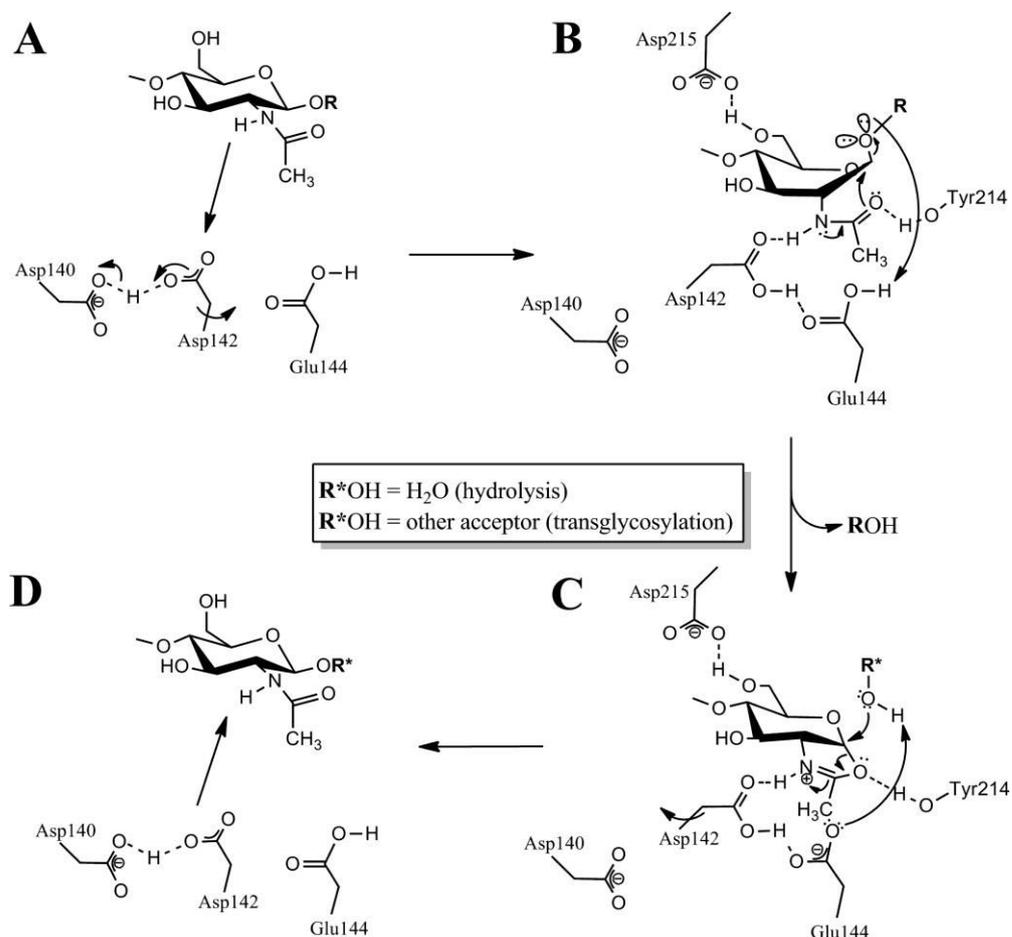


Figure 22. Catalytic mechanism of GH18 chitinases exemplified with ChiB from *S. marcescens*. The figure is modified from ref (van Aalten *et al.*, 2001). A) Substrate binds the resting enzyme and induces rotation of Asp142 toward Glu144. B + C) Substrate binding causes distortion of the -1 sugar to a skew-boat conformation (only the -1 sugar is shown); Glu144 functions as a general acid and facilitates leaving group departure by protonating the glycosidic oxygen. Simultaneously, a nucleophilic attack by the acetamido group of the -1 sugar also promotes leaving group departure, leading to the formation of an oxazolinium ion intermediate C). C + D) Finally, Glu144 acts as a general base and activates an acceptor molecule that attacks the oxazolinium ion. This results in a hydrolytic or a transglycosylation product with retained conformation at the anomeric carbon D).

1.2.4 Processivity in GH18 chitinase.

Due to the crystalline nature of chitin, substrate accessibility is energetically demanding and enzyme-substrate binding is considered the rate limiting step of hydrolysis (Zakariassen *et al.*, 2011); (Beckham & Crowley, 2011). Hence, enzymes have developed special tactics to ensure efficient binding and degradation. The additional carbohydrate-binding modules observed in a large number of chitinases joined to their catalytic domain is presumed to be beneficial for enzyme efficiency as it increases the adherence ability to the

chitin chain (McCartney *et al.*, 2004); (Lehtio *et al.*, 2003); (Carrard *et al.*, 2000); (Watanabe *et al.*, 1994); (Din *et al.*, 1994); (Boraston *et al.*, 2004). As a consequence of the existence of these accessory modules often associated to long and deep substrate-binding, chitinases will not release the substrate after a cleavage. However, they remain attached to their substrates between the subsequent hydrolytic cycles which leads to several cleavages on the substrate chain resulting in a sliding process, called processivity (Davies & Henrissat, 1995); (Rouvinen *et al.*, 1990); (Robyt & French, 1967); (Robyt & French, 1970). Practically, since GH18 chitinase hydrolyze by using a substrate-assisted mechanism which involves the N-acetyl group, hence its position is important to perform this mechanism. As the successive sugar units in chitin are rotated by 180°, therefore during sliding of the substrate only the second sugar that has a correct position of the N-acetyl group in the -1 subsite will have a productive binding meaning that will lead to the cleavage of the polymeric substrate in the active site. As a result, the products of processive hydrolysis are disaccharides (Fig. 23) (Purushotham & Podile, 2012); (Zakariassen *et al.*, 2009). Processive degradation is thought to improve catalytic efficiency on crystalline substrates because single polymer chains are prevented from reassociating with the insoluble material in between catalytic cycles (Teeri, 1997); (Horn, Sikorski, *et al.*, 2006), thus dispensing the energy cost of gaining access to a single chain for every hydrolysis cycle (Rouvinen *et al.*, 1990); (Divne *et al.*, 1994).

Both the endo- and exo- mode of action can occur in combination with processivity. Interestingly, studies on processive enzymes have highlighted the critical role of the surface-exposed aromatic residues starting from the substrate binding-domain and extending to the catalytic cleft. Indeed, it is presumed that the hydrophobic interaction between the aromatic residues and sugar moieties explain how processive chitinases manage to remain attached to the substrate (Divne *et al.*, 1994); (Varrot *et al.*, 2003); (Horn, Sikorski, *et al.*, 2006); (Zakariassen *et al.*, 2009) (Fig. 24).

Many pieces of evidence have shown that CHIT1 is a processive endo-chitinase as it is capable to perform random binding to the substrate, and cleavage within the polymer chain (Eide *et al.*, 2012); (Eide *et al.*, 2013). Site-directed mutagenesis studies targeting the processivity feature of enzymes have been done on ChiA, which is presumed to share an “aromatic signature” similar to CHIT1 in the substrate-binding cleft with the same hydrolysis

direction from the reducing end (Norberg *et al.*, 2011); (Eide *et al.*, 2012). The mutation of aromatic residues (Trp275 and Phe396) located at positions +1 and +2 respectively, have shown only limited effect on processivity. These positions have been considered as product release subsites in ChiA and CHIT1 in the processive hydrolysis of polymeric chitin (Eide *et al.*, 2013). However, Trp167 (Trp31 in CHIT1) in the -3 subsite was demonstrated to crucially affect the processivity in ChiA (Zakariassen *et al.*, 2009). In ChiA and CHIT1, the polymeric part of the substrate is thought to bind to the $-n$ subsites, therefore, a Trp in the -3 subsite seems to be vital for maintaining the processivity action, supporting the thought which underlines the significant implication of conserved exposed aromatic amino acids in the processivity mode of action (Zakariassen *et al.*, 2009); (Eide *et al.*, 2012); (Norberg *et al.*, 2011). Moreover, CHIT1 acts on fungal cell walls, and it is conceivable that its processive mechanism contributes to its fungistatic effect (van Eijk *et al.*, 2005); (Eide *et al.*, 2013). This information is important for further work aimed at understanding the behavior of GH18 chitinases, especially the human chitinases, as well as the development of inhibitors that are specific for certain chitinases (Eide *et al.*, 2012); (Vaaje-Kolstad *et al.*, 2013).

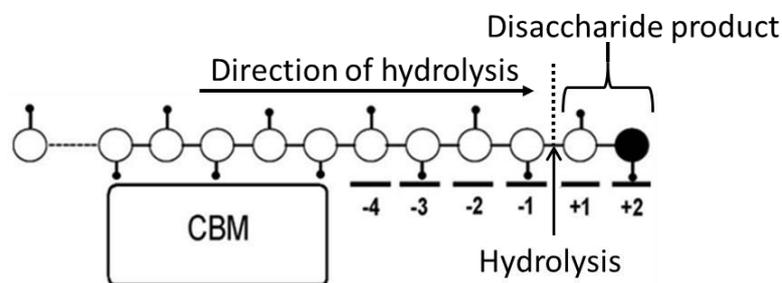


Figure 23. Schematic picture of GH18 chitinase in complex with a single chitin chain. The catalytic domain of the enzyme has six subsites, numbered from -4 to +2. CBM indicates a carbohydrate-binding module involved in chitin-binding (see section 3). Correctly positioned N-acetyl groups (symbolized by small black balls on sticks) show successive sugar units in chitin rotated by 180°. The reducing end sugar is colored *black*. The glycosidic bond between the sugar residues in subsite +1 and -1 is enzymatically cleaved (illustrated with an *arrow*) resulting in disaccharide products. The scheme shows the situation during processive action when only dimers are produced. The arrow indicates the direction of the sliding of the substrate through the active site cleft. Adapted from ref. (Zakariassen *et al.*, 2009)

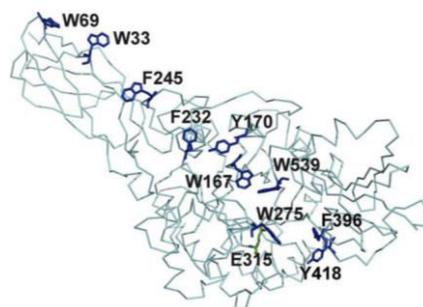


Figure 24. Position of aromatic residues and enzyme-substrate interactions in ChiA and B from *S. marcescens*. The α -carbon chain of the complete ChiA (PDB code: 1EDQ) presented in ribbon. The aromatic side chains that line the substrate-binding cleft of the catalytic domain and the surface of the chitin-binding domain are in stick. From ref. (Zakariassen *et al.*, 2009)

1.3 The different modules in GH18 chitinases

In addition to the catalytic domain, most enzymes involved in cellulose and chitin degradation usually contain one or more domains that are involved in substrate binding (Tews *et al.*, 1997); (Bayer *et al.*, 1998); (Svitil & Kirchman, 1998). Indeed, GH18 chitinases can contain, apart from catalytic domains, extra substrate-binding domains thought to promote activity on insoluble substrates, i.e. they are multimodular (Horn, Sikorski, *et al.*, 2006) (Fig. 25). Some of these extra domains can be ChBDs belonging to different carbohydrate-binding modules (CBM) families, linked by a hinge to the catalytic domain. Besides ChBDs, some chitinases could also contain fibronectin type III-like (FnIII-like) or immunoglobulin-like (Ig-like) domains which are thought to give the protein a long reach over the chitin surface (Malecki *et al.*, 2013) (Fig. 26). These domains are thought to promote binding on insoluble chitin especially to the crystalline form and therefore increase the efficiency of the catalytic domain activity (Jee *et al.*, 2002); (Uchiyama *et al.*, 2001); (Boraston *et al.*, 2004); (Eijsink *et al.*, 2008); (Nimlos *et al.*, 2012b); (Hashimoto, Ikegami, *et al.*, 2000); (Watanabe *et al.*, 1994); (Horn, Sikorski, *et al.*, 2006). CBMs are the most common non-catalytic modules associated to carbohydrate-targeting enzymes e.g. GH). The majority of these domains bind to specific carbohydrate polymers such as cellulose and chitin. Currently, there are 64 families of CBMs based on amino acid sequence similarity in the CAZy database (Cantarel *et al.*, 2009). CBMs can be structurally diverse (Malecki *et al.*, 2013).

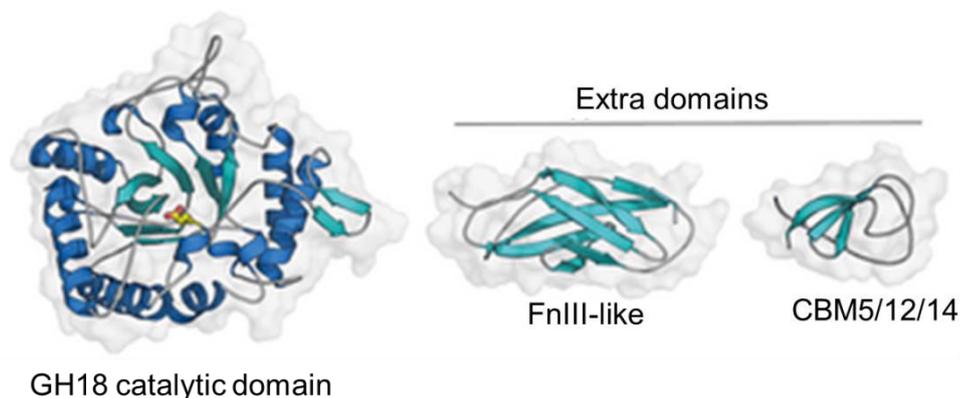


Figure 25. Structural overview of the different modules that exist in GH18 chitinases. The figure shows the structure of the TIM barrel of GH18 catalytic domain, FnIII domain and the ChBD which belong to CBM5-12 or CBM14 families. Helices are colored dark blue, β -strands are colored cyan (Vaaje-Kolstad *et al.*, 2013).

Many putative CBMs have been identified by amino acid sequence similarity but only a few representatives have been shown experimentally to have a carbohydrate-binding function (Roske *et al.*, 2004). The roles of the CBMs and FnIII-like modules have been investigated in several studies (Fig. 26). Many of the substrate binding features of CBMs are in general well documented (Boraston *et al.*, 2004) but it is still not completely clear how CBMs participate to the positioning of the catalytic domain, the processive action, and the local decrystallization of the substrate (Eijsink *et al.*, 2008); (Nimlos *et al.*, 2012a).

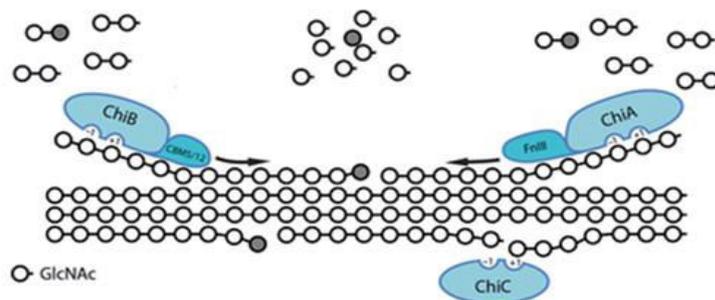


Figure 26. Schematic overview of the *S. marcescens* multimodular GH18 chitinases ChiA, ChiB, ChiC and their synergic action on chitin. Adapted from (Vaaje-Kolstad *et al.*, 2013)

1.3.1 CBMs in bacterial chitinases

The first full length structure of ChiB from *S. marcescens* was the first solved structure of a chitinase catalytic domain with its C-terminal ChBD (Fig. 27). This latter belongs to the CBM family 5 (CBM5) consisting of three antiparallel β -strands connected by long loops and contains two surface-exposed aromatic residues (Trp479 and Tyr481) (van Aalten *et al.*, 2000) (Fig. 28). Interestingly, other known ChBDs belonging to CBM5 have shown conservation of these two aromatic residues. For example, in the structures of the ChBD_{ChiC} from *Streptomyces griseus* (Akagi *et al.*, 2006) (Fig. 28) and ChBD_{MmChi60} from psychrophilic bacterium *Moritella marina* (Malecki *et al.*, 2013), these conserved aromatic rings appear as two tryptophans with coplanar side chains. On the other hand, ChiA1 from *B. circulans* has in addition to the catalytic domain, two FnIII-like domains and a C-terminal ChBD belonging to the CBM12 family (Fig. 28). The structure of the CBM12 isolated from the catalytic domain was solved by NMR, and shows two antiparallel β -sheets with Pro689 and Pro693 exposed on the surface, and only one tryptophan, Trp687, highly conserved among

different bacterial chitinases and demonstrated to be essential for the binding to chitin (Watanabe *et al.*, 2003); (Ferrandon *et al.*, 2003) (Fig. 28). CBM5 and CBM12 modules are distantly related and appear as one family in Pfam (PF02839; www.pfam.org). Both are characterized by the presence of conserved exposed tryptophans that interact with the substrate (Vaaje-Kolstad *et al.*, 2013).

On the basis of sequence analyses, bacterial chitinases ChBDs have been suggested to share sequence homology with carbohydrate-binding domains found in chitinases, cellulases and xylanases (Svitil & Kirchman, 1998); (Brurberg *et al.*, 1996); (Akagi *et al.*, 2006). It is worth to note that many cellulose binding domains (CBDs) have shown to be able to bind to chitin in addition to cellulose (Goldstein *et al.*, 1993); (Linder *et al.*, 1996); (Tomme *et al.*, 1995); (Tomme *et al.*, 1994). On the other hand, ChBD binds specifically to chitin and not to cellulose suggesting that there must be significant differences in the mechanisms of substrate recognition between ChBD and CBDs (Hashimoto, Honda, *et al.*, 2000). Akagi *et al.* have proposed that many ChBDs and CBDs have retained almost the same backbone conformations. However, the substrate specificity (chitin or cellulose) depends not only on the number of aromatic residues within these domains but also on their versatile conformations in the binding interface (Akagi *et al.*, 2006) (Fig. 28).

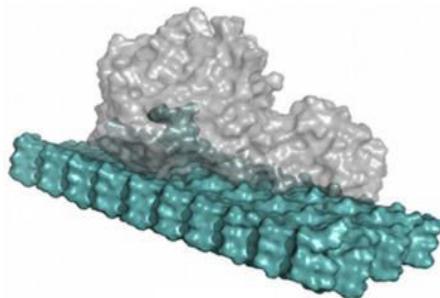


Figure 27. Example of the mode of action of GH18 chitinase CBMs on chitin. Surface representation of the *S. marcescens* ChiB (catalytic domain and ChBD) modeled, bound to crystalline chitin. Adapted from (Payne *et al.*, 2012).

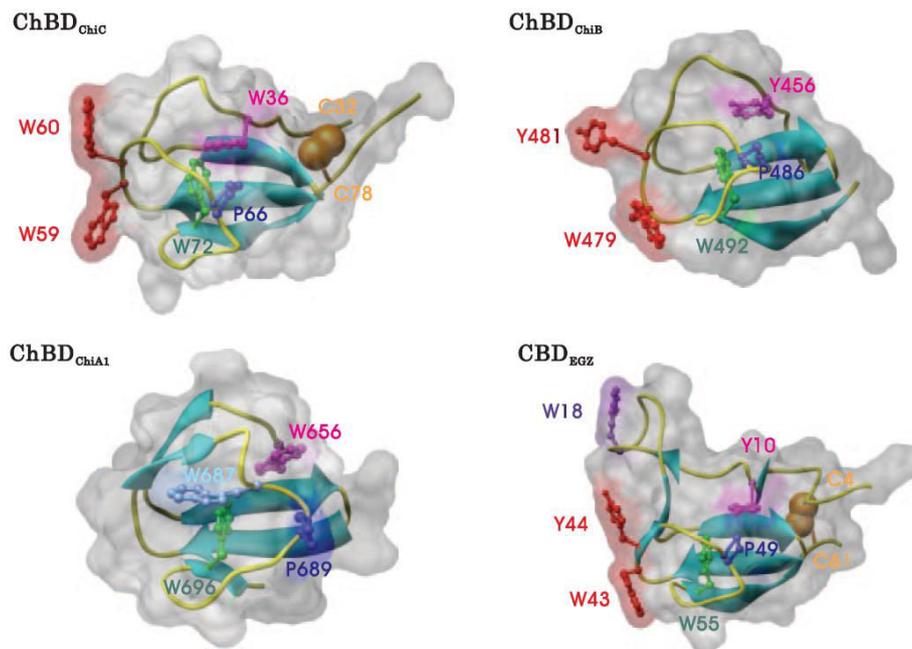


Figure 28. Comparison of the 3D structures of representative chitin- and cellulose-binding domains that belong to CBM family 5 or 12. The figure shows surfaces of the chitin binding domain (ChBD) from the chitinases *Streptomyces griseus* HUT6037 chitinase C (ChBD_{ChiC}), *Bacillus circulans* chitinase A1 (ChBD_{ChiA1}), *S. marcescens* chitinase ChBD_{ChiB}, and the cellulose *Erwinia chrysanthemi* endoglucanase Z precursor (CBD_{EGZ}) with a ribbon display for their backbones. Some important residues for structural comparison are displayed as a ball-and-stick model. The sulfur atoms in the disulfide bridges are indicated by yellow balls. The important residues for comparison are indicated. PDB code: ChBD_{ChiC} (2D49), ChBD_{ChiA1} (1E15), ChBD_{ChiB} (1E15), and CBD_{EGZ} (1AIW). From ref. (Akagi *et al.*, 2006).

1.3.2 Human ChBDs belong to same family of CBM found in invertebrates

The C-terminus ChBDs found in the 50 kDa forms of CHIT1 and AMCCase (associated by a hinge to their N-terminal catalytic domains) have been reported to not have homology to known structures, but show sequence similarities with nematodal and insect ChBDs. This observation has suggested that these human ChBDs belong to the CBM14 family. In fact, both invertebrates and mammals ChBDs have been reported to belong to the CBM14, are cysteine-rich and have several highly conserved aromatic residues (Funkhouser & Aronson, 2007). Sequence comparisons indicate that invertebrate and plant ChBDs do not share significant amino acid sequence similarity suggesting that they are not co-ancestral. Thus, plant ChBDs are reported to belong to family CBM18. However, both the invertebrate and the plant ChBDs are cysteine-rich and have several highly conserved aromatic residues (Shen & Jacobs-Lorena, 1998); (Wright, 1987); (Akagi *et al.*, 2006). These cysteines have as a role to maintain protein folding, while the aromatic amino acids are related to the interaction with saccharides (Wright *et al.*, 1991); (Shen & Jacobs-Lorena, 1999). The only structure available

for an invertebrate CBM14 motif is that for tachycitin, a 73-residue polypeptide with antimicrobial activity from hemocytes of the horseshoe crab (*Tachypleus tridentatus*). The 3D structure of tachycitin was solved by NMR and consists of a single ChBD of three N-terminal β -stranded and two C-terminal β -stranded anti-parallel β -sheets, β -helical turn and 5 disulfide bridges (Suetake *et al.*, 2002). Although no sequence homology is observed between the CBM14 and CBM18, yet the aforementioned 3D structures revealed that they share some structural features (Suetake *et al.*, 2002); (Suetake *et al.*, 2000) (Fig. 29).

The ChBD in CHIT1 consist of 49 residues, eight of them are aromatic amino acids and six are cysteine residues. Mutation of any of these cysteines results in complete loss of chitin binding capacity (Tjoelker *et al.*, 2000b). Tjoelker and co-workers have compared the chitinolytic activity between the full-length and truncated CHIT1 (containing only the catalytic domain without the ChBD). The results show that the absence of ChBD_{CHIT1} does not affect the ability to hydrolyze the soluble substrate, but a major decrease in the hydrolysis of colloidal chitin, suggesting the ChBD in humans is essential for hydrolyzing insoluble chitin. In fact, the recombinant ChBD_{CHIT1} binds to the cell wall of several fungi, e.g. *C. albicans*, *Mucor rouxii*, and *Neurospora crassa* and binds specifically to chitin but not to other polysaccharides, e.g. chitosan cellulose (Tjoelker *et al.*, 2000b). A recent report has tested the affinity of three ChBDs toward fungal cell walls, ChBD_{ChiA1} from *B. circulans* and human ChBD_{AMCase} and ChBD_{CHIT1}. Interestingly, ChBD_{CHIT1} displays the strongest affinity in every tested fungal species (Vandevenne *et al.*, 2011). Further, Eide and collaborators have suggested that the ChBD_{CHIT1} may increase the processivity of CHIT1 and have pointed out that the lack of the 3D structure of the ChBD_{CHIT1} makes the understanding of its mode of action incomplete (Eide *et al.*, 2012); (Eide *et al.*, 2013).

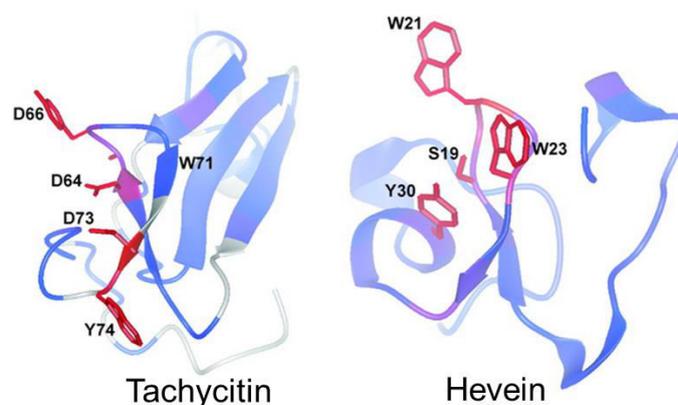


Figure 29. 3D representation of the chitin-binding sites of tachycitin (PDB code: 1DQC) and hevein (PDB code: 1HEV). From ref. (van den Burg *et al.*, 2004).

1.4 Structural-based inhibition of GH18 chitinases.

As chitinases are involved in several vital processes of various species and have similarities in their binding and active sites, many studies have highlighted the interest in chitinase inhibitors as molecules with multiple potentials. Indeed, chitinase inhibitors have been exploited i) to investigate the catalytic mechanism of these enzymes; ii) to inhibit the growth of pathogenic fungi (Takaya *et al.*, 1998) and insects (Arai, Shiomi, Yamaguchi, *et al.*, 2000); (Arai, Shiomi, Iwai, *et al.*, 2000) and to hinder malaria transmission (Vinetz *et al.*, 1999); (Vinetz *et al.*, 2000) and (Tsai *et al.*, 2001); iii) and recently for anti-inflammatory potential against asthma and allergic diseases, including atopic dermatitis and allergic rhinitis (Zhu *et al.*, 2004). Thus, several different classes of chitinase inhibitors have been reported and their effects have been investigated on different chitinases in order to evaluate their specificity against the chitinase targets. So far, many natural compounds and synthetic chitinase inhibitors for GH18 members exist and some of them are effective inhibitors (nM range). Chitinases inhibitors vary widely in structure and include pseudotrisaccharides (Nishimoto *et al.*, 1991); (Sakuda *et al.*, 1986), (Sakuda *et al.*, 1987), cyclic peptides (Izumida *et al.*, 1996); (Omura *et al.*, 2000); (Arai, Shiomi, Yamaguchi, *et al.*, 2000); (Arai, Shiomi, Iwai, *et al.*, 2000), amino acid derived materials (Tabudravu *et al.*, 2002), complex alkaloids (Kato *et al.*, 1995) and simple purine-derived heterocycles (Rao, Houston, *et al.*, 2005).

To understand the mode of action of such inhibitors, biochemical and structural investigations combined with computational studies have been conducted to detail the interactions in chitinase–inhibitor complexes (Rao, Andersen, *et al.*, 2005); (Rao, Houston, *et al.*, 2005); (van Aalten *et al.*, 2001).

1.4 Allosamidin, a natural compound inhibitor of GH18 proteins, and its derivatives

i) The multiple roles of allosamidin

Allosamidin, a pseudotrisaccharide, was first isolated from the mycelium of *Streptomyces* sp. and is the most extensively studied chitinase inhibitor (Sakuda *et al.*, 1986); (Sakuda *et al.*, 1987) (Fig. 30) (Table 6). Various biological properties have been reported regarding its function as a chitinase inhibitor, including inhibition of cell separation in fungi (Kuranda & Robbins, 1991); (Sakuda *et al.*, 1993), mortality induction towards insect larvae

(Sakuda *et al.*, 1986); (Sakuda *et al.*, 1987); (Blattner *et al.*, 1996) and blocking of malaria parasite penetration into the mosquito midgut (Shahabuddin & Kaslow, 1993). Recently, allosamidin has been shown to decrease lung inflammation in a mouse model of asthma (Zhu *et al.*, 2004). Allosamidin is thought to be a competitive chitinase inhibitor and is composed of two N-acetylallosamine molecules and one allosamizoline (or demethylallosamizoline) (Dickinson *et al.*, 1989) (Fig. 30).

Table 6. IC₅₀ values of allosamidin against various family 18 chitinases (nM).

Chitinase	Allosamidin IC ₅₀ (nM)
CHIT1 (Rao <i>et al.</i>, 2003)	40
Mouse AMcase (Boot <i>et al.</i>, 2001)	400
<i>Neurospora crassa</i> (McNab & Glover, 1991)	400
<i>Trichoderma harzianum</i> (Muzzarelli, 1996)	1600
<i>Bombyx mori</i> (Muzzarelli, 1996)	48
<i>Candida albicans</i> (Nishimoto <i>et al.</i>, 1991)	300
<i>Saccharomyces cerevisiae</i> (Nishimoto <i>et al.</i>, 1991)	54000
<i>Trichoderma sp.</i> (Nishimoto <i>et al.</i>, 1991)	1300

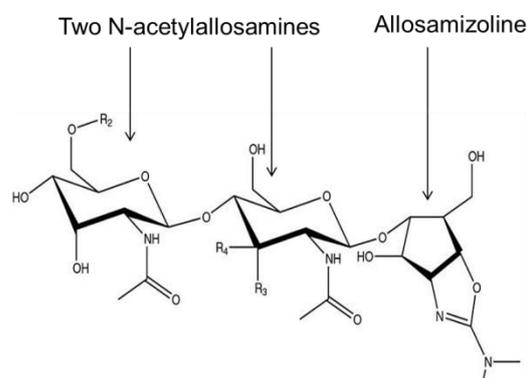


Figure 30. Allosamidin formula

ii) Allosamidin, a tool to understand the mode of action of GH18 members

Binding of allosamidin to chitinases has been investigated through crystal structures using hevamine (Terwisscha van Scheltinga *et al.*, 1996), ChiB and ChiA from *S. marcescens* (Vaaje-Kolstad *et al.*, 2004); (van Aalten *et al.*, 2001); (Papanikolau *et al.*, 2003), CiChi from *C. immitis* (Bortone *et al.*, 2002), human chitotriosidase (CHIT1) (Rao *et al.*, 2003) (Fig. 31A), and AfChiB1 from *A. fumigatus* (Rao, Andersen, *et al.*, 2005) and human AMCase (Olland *et al.*, 2009b) (Fig. 31B). According to the structural data, the allosamizoline unit of allosamidin binds deep in the -1 subsite by stacking with the fully conserved Trp in GH18 chitinases (Trp358 in CHIT1, Trp360 in AMCase) and interacts with the middle aspartate (Asp138 in CHIT1 and AMCase) and catalytic glutamate (Glu140 in CHIT1 and AMCase) of the consensus motif DXDXE found in all active GH18 chitinases (Fig. 31A, B). Strikingly, it has been presumed that this unit mimics the oxazolinium ion intermediate of the substrate-assisted reaction, thus generating favorable interactions and tight binding in the active site. These interactions have been reported to be conserved along the GH18 chitinases (Andersen *et al.*, 2005). Moreover, while the apo form chitinases have shown the middle aspartate residue turning towards to and forming a H-bond with the first aspartate (Asp136 in CHIT1 and AMCase); this middle aspartate was found flipped towards the catalytic glutamate and interacting with its side chains in presence of allosamidin (Fig. 31B). The second orientation of the aspartate has been suggested to occur similarly within the real intermediate during the catalytic reaction (Papanikolau *et al.*, 2003); (Rao *et al.*, 2003); (Hennig *et al.*, 1995); (Vaaje-Kolstad *et al.*, 2004); (van Aalten *et al.*, 2001). In the structure of ChiB from *S. marcescens* in complex with allosamidin, an ordered water molecule has been observed within 3.3 Å of the allosamizoline C1 carbon, and subsequent analysis of the *C. immitis* CTS1-allosamidin and hevamine chitinase-allosamidin complexes have also revealed such a water molecule (van Aalten *et al.*, 2001). This interaction is thought to be reminiscent of the attack of a water molecule, which hydrolyzes the oxazolinium ion reaction intermediate (van Aalten *et al.*, 2001). Furthermore, structure alignments have shown that allosamidin binds to the different chitinases in a similar way, occupying the -3, -2 and -1 subsites, with an identical binding fashion to the corresponding substrate NAG units within the binding site. Indeed, in addition to the interactions highlighted above, allosamidin binding is mediated by several

hydrophobic contacts and forms H-bond with several residues in the chitinases binding groove (Fig. 31C) as well as with water molecules. Thus, taken together structural studies on allosamidin inhibitor have been used as a tool to investigate the substrate binding features and the catalytic mechanism in GH18 chitinases.

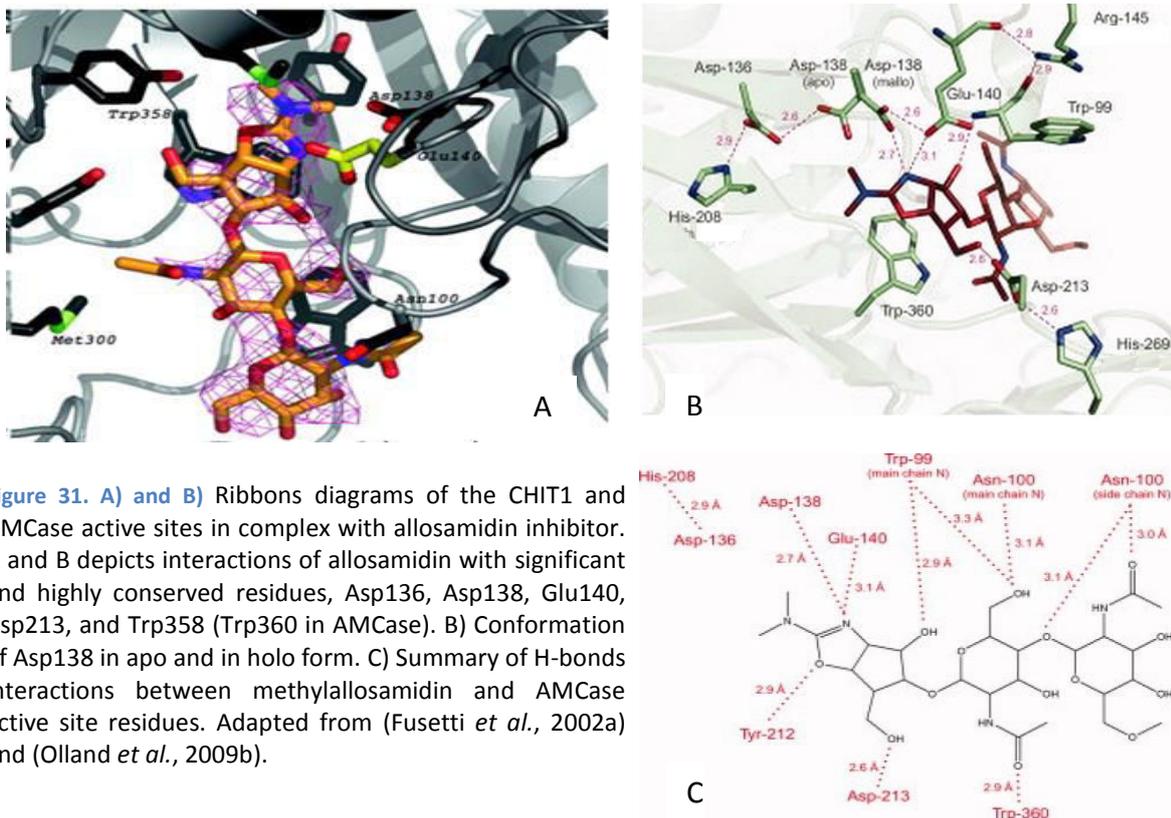


Figure 31. A) and B) Ribbons diagrams of the CHIT1 and AMCase active sites in complex with allosamidin inhibitor. A and B depicts interactions of allosamidin with significant and highly conserved residues, Asp136, Asp138, Glu140, Asp213, and Trp358 (Trp360 in AMCase). B) Conformation of Asp138 in apo and in holo form. C) Summary of H-bonds interactions between methylallosamidin and AMCase active site residues. Adapted from (Fusetti *et al.*, 2002a) and (Olland *et al.*, 2009b).

iii) Investigation of the causes and the results of the differential inhibition values of allosamidin in GH18 chitinases

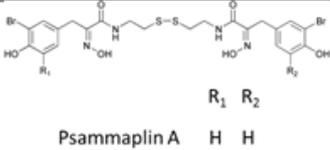
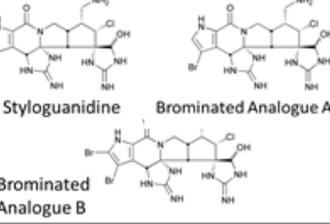
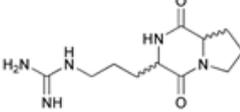
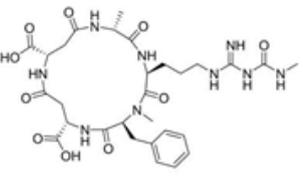
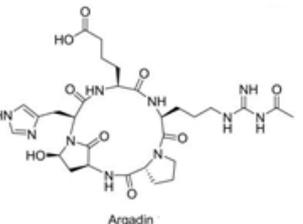
IC₅₀ values have been reported for allosamidin against numerous chitinases (Table 6) ranging from strong inhibition of CHIT1 (Rao *et al.*, 2003) and *Bombyx mori* chitinase (Koga) to relatively weak inhibition of *C. albicans* and *S. cerevisiae* chitinases (Nishimoto *et al.*, 1991). Comparison of the allosamidin-chitinase co-structures, the most studied being the human and fungal complexes, shows that although many contacts are presumed to be conserved, yet differences in interactions are noted between distinct chitinases and this inhibitor. These different interactions may be the basis for differences in inhibition. First, the

S. cerevisiae and *C. albicans* chitinases are similar to the relatively small plant chitinase hevamine, which lacks the extra $\alpha+\beta$ domain that gives the active site a groove character and provides several contacts with the inhibitor (Tyr267, Glu297, and Met300 in CHIT1; Fig. 31A). Moreover, Asn100, which interacts via H-bonds with the oxygen of the *N*-acetyl group on the -3 sugar, is replaced by the shorter threonine in the fungal enzyme. Met300 (in CHIT1), which stacks against the methyl of the *N*-acetyl group on the -2 sugar, is a valine in the fungal enzyme, creating a small void (Fig. 31A). Thus, these structural observations led to the suggestion of developing inhibitors against chitinases with differential specificity based on the changes around the -2 and -3 subsites.

i. Other chitinase inhibitors.

Although several naturally occurring analogues of allosamidin have been identified and isolated (Nishimoto *et al.*, 1991); (Isogai *et al.*, 1989); (Somers *et al.*, 1987); (Zhou *et al.*, 1993), allosamidin and its derivatives have been considered to possess physicochemical properties not compatible with a drug-like compound, such as high molecular weight and poor ligand efficiency (Vaaje-Kolstad *et al.*, 2004). In addition, allosamidin has a broad range of inhibition for GH18 chitinases (Berecibar *et al.*, 1999). Thus, this has prompted investigators to identify, test and design new chitinase inhibitors for specific targets (Table 7). It is worth to note that chitinase inhibitor design should lead to the generation of ligands specific on the target chitinase and should have the minimum effect on chitinases from other species. Thus, when inhibitors against parasite, fungal or other pathogen chitinases are developed, the effects on human chitinases should be negligible (Fusetti *et al.*, 2002a).

Table 7. Chemical structures of inhibitors tested on GH18 chitinases.

Chitinase inhibitors	Chemical Structures	Target	IC ₅₀
<p>Psammaplins</p> <p>Originated from sponge <i>Psammaplysilla Purpurea</i> Quinoa et al., 1987</p>	 <p style="text-align: center;">R₁ R₂ Psammaplins A H H</p>	<p><i>Bacillus</i> sp. Chitinase</p> <p>SmChiB</p>	<p>IC₅₀ = 68 μM ^{Tabudravu et al., 2002}</p> <p>IC₅₀ = 100 μM ^{Tabudravu et al., 2002}</p>
<p>Styloguanidines</p> <p>Originated from Sponge <i>Stylorella aurantium</i> (Kato et al., 1995)</p>	 <p style="text-align: center;">Styloguanidine Brominated Analogue A Brominated Analogue B</p>	<p>Bacterial chitinase <i>Schwanella</i> sp.</p> <p>Inhibitor of cyprid larvae of barnacles</p>	
<p>Cyclic proline-containing dipeptides</p> <p>Originated from marine bacterium <i>Pseudomonas</i> sp. IZ208 (Izumida et al., 1996a)</p>	 <p style="text-align: center;">Arg Pro</p> <p>101 L 102 D } Stereoisomer</p>	<p><i>Bacillus</i> sp. Chitinase</p> <p>SmChiB</p> <p><i>Saccharomyces cerevisiae</i> and <i>Candida albicans</i> Chitinases. ^{Izumida et al., 1996a}</p>	<p>IC₅₀¹⁰¹ = 6.3 mM ^{Houston et al., 2004}</p> <p>IC₅₀¹⁰² = 1.2 mM ^{Houston et al., 2004}</p>
<p>Cyclopentapeptide argifin</p> <p>Originated from fungi <i>Gliocladium</i> sp. FTD-0668 (Omura et al., 2000; Arai et al., 2000; Shiomi et al., 2000)</p>	 <p style="text-align: center;">Argifin</p> <p>Consist of: N-methyl carbamoyl-derivatised L-Arg, an N-methyl L-Phe, two β-L-Asp and a D-Ala residue.</p>	<p><i>Lucilia cuprina</i> (blowfly) chitinase</p> <p><i>Af ChiB1</i></p> <p>SmChiB</p>	<p>IC₅₀ = 3.7 μM ^{Shiomi et al., 2000}</p> <p>IC₅₀ = 1.1 μM ^{Rao et al., 2005a}</p> <p>IC₅₀ = 1.1 μM ^{Rao et al., 2005a}</p>
<p>Cyclopentapeptide argadin</p> <p>Originated from fungi <i>Clonostachys</i> sp. FO-7314 (Arai et al., 2000).</p>	 <p style="text-align: center;">Argadin</p> <p>Consist of: Acetylated L-Arg, a D-Pro, a backbone cyclised L-Asp-bsemialdehyde, an L-His and an L-amino adipic acid residue.</p>	<p><i>Lucilia cuprina</i> (blowfly) chitinase</p> <p><i>Af ChiB1</i></p> <p>SmChiB</p> <p>CHIT1</p>	<p>IC₅₀ = 150 nM ^{Shiomi et al., 2000}</p> <p>IC₅₀ = 0.5 μM ^{Rao et al., 2005a}</p> <p>IC₅₀ = 13 nM ^{Rao et al., 2005a}</p>

propose that GH18 members could actually promote inflammation-associated carcinogenesis (Qureshi *et al.*, 2011); (Kanneganti *et al.*, 2012); (Shao *et al.*, 2014); (Di Rosa *et al.*, 2014); (Di Rosa, De Gregorio, *et al.*, 2013); (Di Rosa, Malaguarnera, *et al.*, 2013). Interestingly, as AMCase has been considered to be a pro-inflammatory mediator in allergy, the anti-inflammatory effect of allosamidin in a mouse model of asthma was thought to be achieved by targeting this chitinase. This prompted investigators to develop specific inhibitors against AMCase as a therapeutic strategy against allergic inflammation (Table 7, two last two lines). Surprisingly and diversely to allosamidin treatment, the AMCase inhibitor *in vivo* has led to an imbalance in the immune response resulting in high neutrophil counts. This is explained by the fact that allosamidin is an unspecific inhibitor for GH18 proteins and since they share similarities in the binding and activity properties, it has been suggested that allosamidin could inhibit many members simultaneously which results in the decreasing of the allergic symptoms in mouse models and which is not the case with monotherapy of AMCase. As the physiological role of all GH18 proteins is not exactly known, it has been recognized that developing compounds that are drug-like inhibitors to target GH18 members with differential specificities against each member or multiple specificities remains of high interest. These compounds could be used, i) to explore and dissect the different roles of GH18 proteins in normal and physiopathological processes, ii) to develop novel therapeutic strategies that combine many approaches or targets to find novel treatments against for example allergies (Sutherland *et al.*, 2011). As the evolutionary pathway driving the emergence of GH18 proteins is likely shared across human and mouse, the studies in mice should allow assessing fundamental functions related to the different chitinases and chi-lectins in humans.

2. Aims and objectives of my PhD project

One of the main research interests of our laboratory consists in studying in fine detail the structural basis of enzyme mechanism involved in important biological processes in order to establish the corresponding structure activity relationships (SAR). For that, our laboratory determines the crystallographic structures of proteins and macromolecular complexes, in the presence and in the absence of natural or synthetic ligands.

Enzymes frequently involve proton translocation steps during their catalytic reactions which mean that hydrogens play important roles in the enzymatic mechanism and function. The investigation of protonation states of catalytic and relevant residues involves the determination of H-atom positions, which provides essential biochemical information that helps in understanding the electrostatic interactions and the catalytic mechanisms. However, the positions of H atoms in macromolecules cannot be routinely determined from crystallographic X-ray data at $<1.2 \text{ \AA}$ (Blakeley *et al.*, 2008); (Afonine *et al.*, 2010). In this regard, our laboratory is strongly interested in developing new methodologies, in order to improve the resolution and quality of protein structures and complexes with ligands as well as to determine the protonation states of the studied enzymes. For this last purpose, we have a particular focus on (very) high resolution X-ray and neutron diffraction. Indeed, previous work done in the lab has allowed the structural determination of the human aldose reductase, at 0.66 \AA , the highest resolution obtained in X-Ray Crystallography for a medium size enzyme (36 kDa) (Howard *et al.*, 2004).

In this context, my PhD project consisted in conducting high resolution structural studies of CHIT1. This enzyme displays several poorly understood features such as the activity shift from hydrolysis to transglycosylation and the processivity action, which exist in a large number of GH18 chitinases. As already mentioned in the introduction, the X-ray crystal structure of the catalytic domain of CHIT1 apo form has been already solved up to 2.3 \AA . This resolution and the obtained resolutions in other homologous GH18 chitinases do not allow observing hydrogen atoms in the structures and therefore are not informative about the mechanistic features in this enzyme family. Thus, improving the resolution of CHIT1 crystal structures is an important step to conduct a mechanistic study on this enzyme, determine

the protonation state of its active site and gain a deeper insight onto the conserved enzymatic features in GH18 family. Also, a very high resolution structural study will facilitate the understanding of the proton pathway translocation during the substrate catalysis and will be a useful basis for drug-design purposes.

Moreover, since the solved crystal structures of CHIT1 consisted only on its catalytic domain, the structure of the full length enzyme which contains the ChBD has so far never been solved. This domain is rarely studied from a structural point of view and at the same time recent data indicate that it is involved in several biological processes.

Based on all the above, my PhD work has had several goals which can be summarized as follows:

1) To improve the resolution of the X-ray structure of CHIT1 catalytic domain to the atomic resolution.

2) To investigate the protonation state of the active site of CHIT1 and the structure activity relationship.

3) To increase our understanding on the enzymatic mechanism in GH18 chitinases.

4) To obtain the X-ray crystal structure of the full length human CHIT1.

The results of these objectives will be developed and discussed in the next sections. Importantly, the results of my PhD work led to the following two papers that will be submitted before my defense:

- New insight in the enzymatic mechanism of human chitotriosidase (CHIT1) catalytic domain by atomic X-ray data and Hybrid QM/MM.

- X-ray crystal structure of the full-length human chitotriosidase (CHIT1) reveals features of its chitin binding domain.

Besides my PhD project, work done in collaboration with an industrial partner has allowed me to solve the crystal structures of CHIT1 in complex with four inhibitors. However, for reasons of intellectual property, only the data collection and refinements statistics as well as the IC_{50} will be included in the annex section but will not be discussed in this thesis manuscript as they fall under a confidentiality clause.

3. Materials and Methods

3.1 Plasmid

The following criteria were taken into account for choosing the plasmid backbone of the expression vector: the plasmid should have a very high copy number in *Escherichia coli* (*E. coli*) (for efficient DNA production), its promoter should be the strongest available for mammalian cells and, finally, it should be small enough to allow the efficient cloning of variable-length constructs. In view of these criteria, the pHL vector was chosen (kindly provided by Yuguang Zhao, The Division of Structural Biology at the University of Oxford). pHL is based on the pLEXm backbone (see appendix 3, fig. 1) and contains the following features:

- the pBR322 origin of replication giving the high copy number in *E. coli* needed to obtain the high amounts of plasmid DNA required for transient transfection;
- ampicillin resistance;
- the cytomegalovirus enhancer;
- the chick β -actin promoter to give high levels of expression;
- the rabbit β -globin intron to increase RNA production;
- a poly-A signal to increase RNA stability.

In addition, several new features were introduced to the pLEXm backbone:

- a Kozak sequence;
- a secretion signal sequence (from the native CHIT1 protein) to allow the secretion of the recombinant protein in the media;
- a C-terminal thrombin site followed by a His-tag to allow protein purification.

3.2 Cloning

The cloning procedure of the CHIT1 full length and CHIT1 catalytic domain was performed by sequence and ligation-independent cloning (SLIC) method (Li & Elledge, 2012). 15 μ g of pHLsec plasmid were digested by restriction enzymes BamHI and XhoI and treated with T4 DNA polymerase (New England Biolabs). The conditions and concentration of DNA and the used enzymes were chosen according to the New England Biolabs technical reference catalogue.

- **pHLsec sequence after *Bam*HI + *Xho*I cleavage and T4 DNA polymerase treatment**

GAA
CTTAAGTTCGAACGGTGGTACC**CCTAG** TCGAGACTAGTATCGCGATAATT
A

The cDNA construct corresponding to CHIT1 catalytic domain (1-386) was used as template to generate the C-terminal thrombin site by using the forward primer 1 and reverse primer 1. The resulting Polymerase Chain Reaction (PCR) product was used to amplify the construct and generate the C-terminal His-tag by performing a second PCR and using the forward primer 1 and reverse primer 2.

For the CHIT1 full length the cDNA construct corresponding to the 50 kDa form was used as a template to add the C-terminal thrombin site by using the forward primer 2 and reverse primer 3. The resulting PCR product was used to amplify the construct and generate the C-terminal His-tag through a second PCR where the forward primer 2 and reverse primer 4 were used.

Primer	Sequence (5' → 3')
Forward primer 1	AATTC AAGCTTGCCACCATGgtgcggtctgtgg
Reverse primer 1	GTGATGGTGATGGTGGTGAGAACCGCGTGGCACCAGactcagttcctgccgtagc
Reverse primer 2	ATTATCGCGATACTAGTCTCGAGTCATTAGTGATGGTGATGGTGGTG
Forward primer 2	AATTC AAGCTTGCCACCATGgtgcggtctgtgg
Reverse primer 3	GTGATGGTGATGGTGGTGAGAACCGCGTGGCACCAGattccaggtgcagcatttg
Reverse primer 4	ATTATCGCGATACTAGTCTCGAGTCATTAGTGATGGTGATGGTGGTG
pHL-F	CTCATCATTTTGGCAAAG
pHL-R	CTCAGTGGTATTTGTGAGC

The enzymes used in all PCRs were Expand High Fidelity enzymes (Roche) and the PCR program consisted of the following steps:

	Temperature	Time	Cycles
Initial	94°C	2 min	1X
Denaturation			
Denaturation	94°C	15 s	10 X
Annealing	63°C	30 s	
Elongation	72°C	60 s	
Denaturation	94°C	15 s	15X
Annealing	63°C	30 s	
Elongation	72°C	60 s	
Final	72°C	7 min	1X
Elongation			
Cooling 4°C	unlimited time		

The PCR products from the second PCR performed for each construct were also treated with T4 DNA polymerase. To stop the T4 DNA polymerase activity, dCTP were added and incubated 10 min at room temperature.

For each construct, the ligation was performed by mixing the digested pHL vector with the PCR product of the second PCR at 37 °C during 1 h. The ligation mixture was next used to transform competent *E. coli* DH5α. The resulting clones were checked by PCR with pHL-F and pHL-R primers.

3.2.1 DNA sequencing

To confirm the successful cloning of the desired insert into the pHL vector, sequence determination of the respective DNA region was performed by the company GATC Biotech using the dye terminator method. The results were provided as ABI chromatogram files. These chromatograms were compared and analyzed using BioEdit Sequence alignment Editor.

3.2.2 DNA purification

Plasmid DNA containing the insert of the full length CHIT1 or CHIT1 catalytic domain were purified by using the Endotoxin-Free Plasmid High speed Giga Kit (Qiagen). As high-quality DNA is essential for successful transfection, only samples with an OD₂₆₀/OD₂₈₀ ratio of 1.8 or higher were used for this purpose. Since the pHL plasmid is very high copy number, between 10 or 12 mg of pure DNA were obtained from a 2 L overnight bacterial culture (*E. coli* DH5 α strain). The precipitated DNA samples were washed with 70 % ethanol before dissolving them in sterile 10 mM Tris pH 8.0.

3.3 Protein Production

3.3.1 HEK293T cell maintenance protocol

Adherent Human Embryonic Kidney 293 (HEK293T) cells were used for production of recombinant full length CHIT1 and recombinant CHIT1 catalytic domain as they have a high capacity for recombinant protein expression and low-cost media requirements. The HEK293T cells are also easy to handle and maintain, they have a good growth rate and a high transfectability. Cells were grown in Dulbecco's Modified Eagle's Medium (DMEM high glucose, SIGMA) containing 1 % L-glutamine, 1 % non-essential amino-acids (Gibco) and 10 % foetal calf serum (FCS), (SIGMA). Cells were manually cultured from stocks stored at -196 °C, using a T25 flask (CORNING) in a humidified 310 K incubator with 5 % CO₂. Upon reaching 90 % confluence, cells were detached using 5 ml trypsin–EDTA (SIGMA) during 5 min. Cells were diluted five times by adding 21 ml DMEM (10 % FCS). 4 ml of cells were split to each flask T175 flask (BD-Falcon) containing 23 ml DMEM (10 % FCS). Cells were normally grown in a humidified 310 K incubator with 5 % CO₂.

3.3.2 Transferring HEK293T cells from flasks to roller bottles

Large-scale cultures for protein production were performed in expanded-surface polystyrene roller bottles (2125 cm², Greiner Bio-One). To transfer cells from T175 flasks to roller bottles, the medium from T175 flask was first removed. Cells were then washed with phosphate buffered saline (PBS) and detached by adding 5 ml trypsin–EDTA during 5 min. 21 ml of DMEM (10 % FCS) were added to the detached cells and then all the cells were pipetted to the roller bottle containing 200 ml DMEM (1 % L-glutamine, 1 % non-essential

amino-acids and 10 % FCS). Since these roller bottles do not have a gas-permeable cap, CO₂ gas has to be flown into each bottle for 20–30 s before tightening the cap.

3.3.3 Small scale transfection and expression test via Western blot

For rapid expression screening of both constructs corresponding to full length CHIT1 and CHIT1 catalytic domain, DNA transfection was performed in 10 ml cell culture incubated in 6-well plates using MiniPrep-purified DNA (QIAprep Spin Kit, QIAGEN). 1 mg plasmid quantity of each construct was used in each expression test. 3 days after transfection, small aliquots of conditioned media (10 µl) were analysed by Western blot using the PentaHis monoclonal primary antibody (1:1000 dilution, QIAGEN) and goat antimouse IgG peroxidase-conjugated secondary antibody (1:2000, SIGMA). Bands corresponding to the molecular weight of each construct were visualized by chemiluminescence using the ECL kit (GE Healthcare).

3.3.4 Transfection in roller bottles

Before the preparation of the transfection cocktail, 80 µl of chloroform were added to 1 ml of the plasmid DNA (1-2 mg/roller bottle) to avoid any contamination of HEK293T cells after transfection. The DNA-chloroform mix solution was centrifuged for 10 min at 13200 rpm. Afterwards, only the plasmid DNA was mixed to 50 ml of serum-free DMEM media. The DNA-media solution was supplemented by 1.5 ml of sterile polyethylenimine (PEI) '25 kDa branched' (SIGMA) at 100 mg/ml stock and pH 7.0. The PEI is the transfection agent which was reported to be effective in transfecting various HEK293T cell lines (Aricescu et al., 2006; Durocher et al., 2002). Thus, the transfection cocktail composed of DNA, PEI and media was supplemented by 375 µl of the N-glycosylation inhibitor, kifunensine (1 mg/ml) (Chang *et al.*, 2007), vortexed and incubated during 10 min at room temperature to allow for DNA-PEI complex formation. During complex formation, media from the roller bottles was changed, lowering the serum concentration (200 ml of DMEM with 2 % FCS were added per roller bottle). Finally, the DNA-PEI complex was added to each bottle, briefly rotated to allow mixing, after which the cells were placed in the incubator. 5 days later, conditioned medium was ready for collection and protein purification.

3.4 Purification protocol

Recombinant full length CHIT1 and CHIT1 catalytic domain were readily purified from media. To do so, conditioned cell media were poured into 500 ml centrifuge tubes and centrifuged at room temperature at 3000 rpm during 10 min. Subsequently, the supernatant was filtered by using 0.22 μm filters through a 0.2 mm membrane (Express filter, Merck Millipore), and 500 ml of the filtered supernatant were poured into 50 cm dialysis tubes. Each dialysis tube was placed in 15 L phosphate buffer (49.63 g Na_2HPO_4 ; 219 g NaCl; 25.5 ml 1 M NaH_2PO_4) at 4 °C stirred and dialyzed during 36 h. The 15 L phosphate buffer was renewed each 5 h. After dialysis, 750 ml aliquots of conditioned media were transferred to 2 L conical flasks and incubated for 1h30 min at 16 °C/120 rpm with 1.5 ml cobalt-coated Talon beads (Clontech) equilibrated with phosphate buffer. The beads were applied into a column and were washed with 75 ml of phosphate buffer, pH 8.0; a second round of washing was applied by adding two column volumes of phosphate buffer (pH 8.0 with 10 mM imidazole). The recombinant full length CHIT1 and CHIT1 catalytic domain were eluted by adding 4 ml of 20 mM Tris-HCl, pH 8.0, 300 mM imidazol phosphate buffer. The eluted proteins were further purified by size exclusion chromatography on a Superdex 200 16/60 column (GE Healthcare) in 10 mM HEPES, 150 mM NaCl pH 7.5. The protein purity was assessed by SDS–PAGE.

3.5 Enzyme activity measurements

The enzymatic activity was determined using commercial synthetic fluorogenic substrates as described before (Hollak *et al.*, 1994). Briefly, 0.25 nM of CHIT1 was incubated at 37 °C with various concentrations (0-200 mM) of the substrate 4-methylumbelliferyl β -N,N',N''-triacetylchitotrioside (SIGMA). The assays were performed in 100 mM citrate phosphate buffer, pH 5.6 containing 1 mg/ml bovine serum albumin. After 30 min, the assay was stopped by addition of 100 μL 0.3 M glycine-NaOH, pH 10.3. The product 4-methylumbelliferone was quantified using a microplate fluorometer (excitation 360 nm/emission 440 nm).

3.6 IC₅₀

CHIT1 at 0.25 nM was incubated with various concentrations of each inhibitor plus 0.022 μM substrate, 100 mM citrate phosphate buffer, pH 5.2 in a final volume of 50 μL . The

IC₅₀ values were derived from the following concentrations for each inhibitor (0, 1 nM, 10 nM, 100 nM, 1 μM, 10 μM, 100 μM). After 30 min incubation at 37 °C, the reaction was stopped with the addition of 500 μl sodium carbonate buffer (0.5 M sodium carbonate and 0.5 M sodium bicarbonate pH 10.6). Liberated 4-methylumbeliferone was quantified using a microplate fluorometer (excitation 360 nm/emission 440 nm).

3.7 Crystallographic methods

3.7.1 Crystallogenesis

Crystallization is a transition from a solute state to an ordered solid state. Although a lot of progress has been achieved in the last decade, especially through the use of robotics, obtaining protein crystals remains a great challenge in X-ray crystallography. Indeed, even though a lot of proteins have been crystallized, no general rule for this process can be applied to all proteins. For this reason, obtaining high quality protein crystals involves screening for suitable crystallization conditions and relies on an empirical procedure. The general crystallization procedure consists of reducing the solubility of the protein of interest. It is generally known that the solubility curve divides the phase diagram of a protein solution into two zones: the undersaturated and the supersaturated zone (Fig. 1). A protein will stay in solution only up to a certain concentration. Thus, under this concentration the protein is in the undersaturated region where neither nucleation nor crystals can emerge. On the other hand, when the concentration of a protein is above its solubility limit, the solution becomes supersaturated (Fig. 1). At this point, nucleation and crystal growth can occur at different regions within the supersaturated area. The nucleation zone or labile zone is where the beginning of crystal formation occurs. This is achieved by the emergence of nuclei able to initiate aggregation in an ordered manner which could lead to crystals formation and growth. In contrast, the metastable zone enables crystal growth but nucleation cannot take place in this region. Once nuclei are formed, the concentration of the protein in solution is lowered and ideally drops back into the metastable zone, where further growth of the nuclei is supported. If very high supersaturation is reached, the nucleation process, which can be described as ordered precipitation, might be overruled by unordered precipitation (Fig. 2). Under these conditions, an amorphous precipitate, a phenomenon often encountered in

protein crystallization experiments, can be formed. Therefore, the ideal crystallization experiment establishes conditions where the protein reaches a level of supersaturation without forming amorphous precipitant but where nucleation could be enhanced.

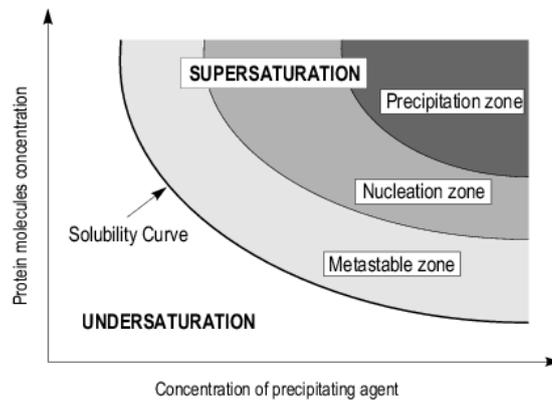


Figure 1. Crystallization phase diagram. Schematic representation of a two-dimensional phase diagram, illustrating the change of protein molecules concentration against precipitating agent concentration. The concentration space is divided by the solubility curve into two areas corresponding to undersaturated and supersaturated state of a protein solution. The supersaturated area comprises of the metastable, nucleation and precipitation zones (Ducruix, 1992).

There are at least six practical methods used for macromolecule crystallization comprising vapor diffusion, bulk crystallization, batch, free-interface diffusion, dialysis, temperature-induced. During my PhD work, vapor diffusion technique was the main method used for crystallization employing both the hanging drop and sitting drop techniques (Fig. 2A, B). In the vapor diffusion method, the initial crystallizing reagent concentration in the droplet is less than that in the reservoir. With time, the reservoir will pull water from the droplet in a vapor phase which leads to an equilibration of water vapor between the sample droplet and the reservoir solution. During the equilibration process, the sample becomes concentrated until the precipitating concentration in the drop becomes almost equivalent to the reservoir concentration, which increases the supersaturation phase of the protein in the drop. As a result, this eventually leads to the nucleation and crystal growth.

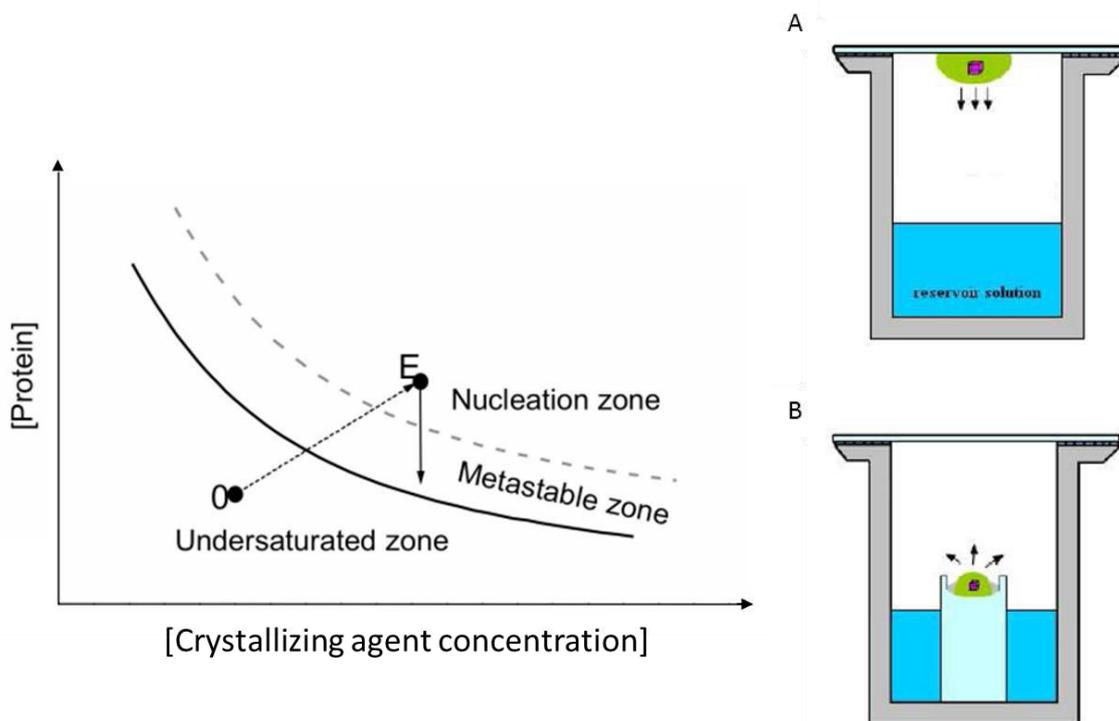


Figure 2. Phase diagram and representation of standard vapor diffusion crystallization methods (A) hanging drop (B) sitting drop. Through loss of water vapor to the larger volume reservoir, the droplet comes into equilibrium with the crystallizing agent concentration in reservoir, thus inducing crystallization.

To change the protein solubility and reach the condition that favors the crystallization of the protein of interest, different approaches can be adopted including altering the buffer pH, temperature, protein concentration, dielectric constant of the solvent or using precipitant agents (McPherson, 1999). These are added to the crystallization solution in addition to solvent and buffer. Precipitant agents change the protein solvation either by interacting directly with the molecule or by changing the solvent properties. They can be divided in three groups: salts, organic solvents and polymers capable of interacting directly with the protein and reducing the dielectric constant of the solvent. The most used is the polyethylene glycol (PEG), which I also used in this work.

Following the screening of crystallization conditions, the optimization of initial crystallization hits is important to adapt the kinetics of crystal growth to generate large single 3D crystals suitable for very high resolution X-ray diffraction.

Furthermore, among the crystallization optimization techniques, seeding, a widely used method to control nucleation and adjust the way in which crystals grow in order to

obtain large, single crystal with high quality diffraction. There are three major types of commonly used seeding: macro-seeding, micro-seeding, and cross-seeding (Bergfors, 2003). During my PhD, I mainly employed micro-seeding and cross-seeding. Micro-seeding involves the introduction of micro-crystals of the protein of interest into a drop in the metastable zone with lower supersaturation phase insufficient for spontaneous nucleation (Rupp, 2009). The seeded crystals will then continue to grow. The crystal growth conditions can be further optimized independently without the need to induce nucleation by the proteins themselves (Luft & DeTitta, 1999). Micro-seeding can be achieved either by employing the streak seeding (with a horse tail hair) method or by adding a small aliquot of a seed solution to a crystallization solution (Stura & Wilson, 1990); (Stura & Wilson, 1991). In practical terms, to perform micro-seeding experiments, pre-formed crystals are crushed and a seed stock solution is prepared and diluted. Each diluted seed solution is used to test the best seed concentration which should avoid the occurrence of many nuclei and should favor the growth of large single crystals. The same protocol can be applied in the case of cross-seeding but in this case, the micro-crystals used for seeding come from a homologue protein of the protein of interest (Bergfors, 2003).

3.7.2 Crystallization of CHIT1 catalytic domain

Screening of crystallization conditions hits were applied for both CHIT1 catalytic domain and full length CHIT1 by setting sitting drops via the Mosquito crystallization robot (TTP LabTech) in 96-well crystallization plates. Each drop consisted of 0.1 μ l protein solution mixed with an equal volume of reservoir solution equilibrated against 40 μ L of the reservoir solution. The advantage of sitting drop is that it requires a small amount of material and is ideal for screening a great number of different conditions by using the different screening kits supplied by different companies. For both construct of CHIT1 the screen kits were Index from Hampton Research, the PEGs and Classics from QIAGEN. In the case of full length CHIT1 an additional screen was performed, the silver bullets from Hampton Research which contains a large number of additives.

For the crystallization of CHIT1 catalytic domain, many crystallization hits were identified. The best crystal quality obtained from these screens was 20 % polyethylene glycol

(PEG) 3350; 0.2 M potassium sodium tartrate (PST) pH 7.2 at 20 °C. This crystallization condition was optimized and the protein concentration tested varied from 7 to 9 mg/ml. The manual crystallization experiments were performed by the hanging-drop vapour-diffusion technique at 17 °C by mixing 1.5 to 3 µL protein with an equal volume of reservoir. The crystals of the apo form of CHIT1 appeared after 4 to 6 days of equilibration against 500 µL of reservoir solution containing 22-26 % (w/v) polyethylene glycol (PEG) 3350, 0.2 M (PST) pH 7.2. To obtain large crystals, rounds of seeding were performed by adding diluted seed solution to the crystallization solution. To prepare the stock solution for seeding pre-obtained crystals were transferred to 1.5 ml eppendorf tube containing Seed Bead (Hampton Research) and 100 µl of reservoir solution. Several dilutions were examined varying from 1 to 14000 dilution fold. The reservoir condition of 25 % PEG 3350, 0.2 M PST pH 7.2 containing a seed dilution of 14000 gave a stick-shaped mono-crystal that grew to a maximum size of 1 x 0.14 x 0.1 mm³.

The pseudo-apo crystal form was obtained after failing to co-crystallize the protein with chitin (chitin from shrimp shells, SIGMA). A saturated stock solution of chitin was prepared in DMSO and dissolved in the reservoir solution (also containing a low concentration of micro-seeds) which was added to the drop. The crystals obtained were collected at 0.95 Å resolution but chitin was not observed, therefore this structure was considered as a pseudo-apo form.

3.7.3 Co-crystallization of CHIT1 catalytic domain with substrate

Crystals containing the artificial substrate 4-MU-NAG₃ (SIGMA) were obtained in the same crystallization condition than the CHIT1 apo form combined with micro-seeding, by mixing 1.5 µL proteins with an equal volume of the reservoir solution containing a final range concentration of 4-MU-NAG₃ between 0.3 mM and 2.5 mM. Hydrolysis of the substrate occurred in the drop, therefore, crystal of CHIT1 complexed with only two NAG monomers, also called chitobiose appeared after 10 to 30 days. All the crystals obtained for CHIT1 apo form, pseudo-apo form or in complex with chitobiose were cryo-cooled in liquid nitrogen, using a solution containing 35 % PEG 3350, 0.2 M PST.

3.7.4 Crystal soaking of CHIT1 catalytic domain with four inhibitor compounds

Soaking of inhibitors has been performed on apo crystals of CHIT1 catalytic domain. These crystals were transferred to 10 μL of a reservoir solution in sitting drop experiment. The four compounds were dissolved in 100% DMSO to a concentration of 200 mM. The soaking solution contained 5 mM of each inhibitor in the reservoir solution (25 % PEG3350, 0.2 M PTS, pH 7.2) and 2.5 % DMSO. Soaking time was 1 hour for each compound.

3.7.5 Crystallization of CHIT1-FL is mentioned in page (127-128)

3.8 X-ray Data collection and processing

When a crystal is placed in an X-ray beam, the radiation is scattered by the electrons clouds of the atoms within the crystal. Upon X-ray scattering, interference occurs due to the regular three-dimensional arrangement within the crystal lattice. Waves can cancel out each other due to negative interference. Only waves which interfere constructively between X-rays fulfill the Bragg condition for diffraction and will give rise to a set of well-defined spots arranged with a characteristic geometry that can be recorded on a detector as a diffraction pattern. The constructive interference between X-rays scattered from successive planes in the crystal will only take place if the path difference ($2d$) between the X-rays is equivalent to an integral number of wavelength (λ) with θ the scattering angle (Fig. 3).

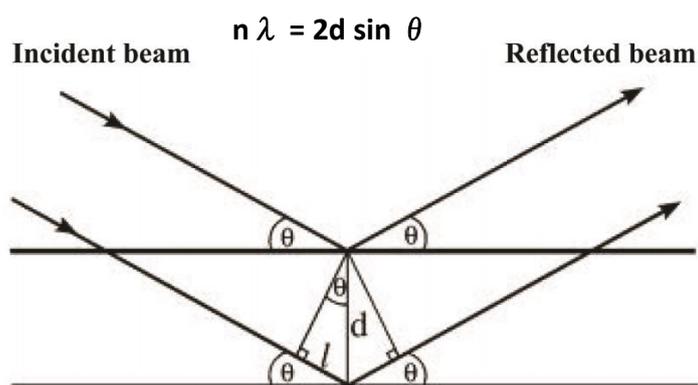


Figure 3. Geometry of diffraction and its relationship to Bragg's Law (Glusker and Trueblood, 1985).

Since the crystal is composed of a number of repeating patterns (unit cells) in a regular and ordered manner, thus each molecule within the crystal diffracts equally, and

strong diffracted X-ray beams can be measured (Rhodes, 2000) Since only a few reflections will fulfill the reflection conditions, the crystal is rotated to cover the entire reciprocal space.

- Details of cryo-cooling and X-ray data collection regarding the different structures obtained with CHIT1 catalytic domain are mentioned in pages (96-97)
- Details of cryo-cooling and X-ray data collection regarding obtained structure of CHIT1-FL are mentioned in page (128-129)

3.4 Electrospray Ionization Mass Spectrometry (ESI-MS) (See page 129)

3.5 Structural Conservation Analysis (See page 130)

4. Results

This section is divided into two main parts and is presented as two papers. The first part deals with the structural and mechanistic study of the catalytic domain CHIT1 through a detailed investigation of the protonation state of the active site which allowed us to propose a new mechanism for the hydrolysis mechanism. The second part relates to the crystallogensis approach and structural characterization of the full length CHIT1 (CHIT1-FL) with a particular focus on the structural features of CHIT1 ChBD.

New insights in the enzymatic mechanism of human chitotriosidase (CHIT1) catalytic domain by atomic X-ray and hybrid QM/MM

Firas Fadel¹, Yuguang Zhao², Raul Cachau³, Alexandra Cousido-Siah¹, Eduardo Howard^{1,4}, Andre Mitschler¹ and Alberto Podjarny¹

1. Department of Integrative Structural Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/UdS, 1, Rue Laurent Fries, 67404 Illkirch CEDEX, France.
2. Division of Structural Biology, Wellcome Trust Centre for Human Genetics, Oxford University, Roosevelt Drive, Headington, Oxford, United Kingdom
3. SAIC Frederick, Inc. Advanced Biomedical Computer Center, Information Systems Program, Frederick National Laboratory for Cancer Research, Frederick, MD, USA.
4. IFLYSIB, Conicet, UNLP, Calle 59 N° 789, La Plata, Argentina

ABSTRACT

Chitotriosidase (CHIT1) is a human chitinase belonging to the highly conserved glycosyl hydrolase family 18 (GH18). GH18 enzymes hydrolyze chitin, a N-acetyl glucosamine polymer synthesized by lower organisms for structural purposes. In the last decade, CHIT1 has attracted attention due to its upregulation in immune system disorders and for being a marker of Gaucher disease. The 39 kDa catalytic domain shows a conserved cluster of three acidic residues, E140, D138 and D136, involved in the hydrolysis reaction. Under excess concentration of substrate, CHIT1 and other homologues within GH18 family are able to perform an additional activity, transglycosylation. To increase our knowledge on the catalytic mechanism of GH18 chitinases and the coexistence of a dual enzymatic activity, we conducted a detailed structural and mechanistic analysis on CHIT1. For this, we improved the resolution of the X-ray crystal structure of the catalytic domain of CHIT1 from the previous

highest resolution of 1.65 Å (PDB code: 1WAW) up to the range between 0.95-1.10 Å, for the apo and pseudo-apo forms and in complex with chitobiose, allowing the determination of the protonation states within the active site. This information was further extended by hybrid quantum mechanics/molecular mechanics (QM/MM) calculations. Our results allow us to suggest a new mechanism involving changes in the conformation and protonation state of the catalytic triad, as well as a new role for Y27, providing new insights into the hydrolysis and transglycosylation activities.

Keywords CHIT1, GH18 chitinase, crystal structures, protonation states, hydrolysis, catalytic mechanism.

1-Introduction

Chitinases belong to the class of glycosyl hydrolases (GH) that degrade chitin, an abundant natural polysaccharide, by cleaving the β -(1,4) linkages of its N-acetylglucosamine (NAG) chain (Gooday, 1990). Chitinases can be subdivided into two families, family 18 (GH18) and family 19 (GH19) that differ in structure and mechanism (Henrissat & Davies, 1997). GH18 chitinases have been identified in many organisms varying from lower organisms to humans. Interestingly, chitinases fulfill diverse functional roles in different species. While they insure carbon and nitrogen sources by degrading chitin in bacteria; they are involved in growth and morphogenesis in fungi and insects. Further, chitinases have been shown to play a protective role in plants and mammals against chitin-containing pathogens.

In the past decade, several crystal structures of GH18 chitinases have been solved. According to these structures, the catalytic domains of this enzyme family consist of a $(\alpha/\beta)_8$ TIM barrel fold with a high degree of conservation in their active site cleft composed of aromatic residues that contribute to the substrate binding (van Aalten *et al.*, 2000); (Perrakis *et al.*, 1994); (Fusetto *et al.*, 2002a); (Olland *et al.*, 2009a); (Yang *et al.*, 2010); (Terwisscha van Scheltinga *et al.*, 1996). The cleavage of the chitin polymer takes place between subsites -1 and +1. The catalytic triad is positioned at the subsite -1, located at the bottom of the substrate binding cleft. This catalytic triad, which consists of two aspartates and a glutamate, belongs to the highly conserved motif (DXDXE) that characterizes the GH18 chitinases (van

Aalten *et al.*, 2001). The glutamate of this motif has been identified as the key catalytic acid/base residue, which is supposed to be protonated on the outer oxygen of its side chain. In general, the majority of GH18 apo form crystal structures have shown the middle aspartate forming a short hydrogen bond (H-bond) with the first aspartate (conformation A). However, it has been reported that in presence of the substrate the middle aspartate turns towards the catalytic glutamate and forms a H-bond with this latter (Conformation B) (van Aalten *et al.*, 2000). In addition to its participation in the catalysis, the middle aspartate has been reported to assist the substrate conformation stabilization subsequent to binding (Fusetti *et al.*, 2002a); (Olland *et al.*, 2009a); (van Aalten *et al.*, 2000); (Songsiriritthigul *et al.*, 2008).

Regarding the enzymatic mechanism, GH18 chitinases are considered as retaining enzymes, which implies the retention of the initial anomeric carbon configuration in the yielding hydrolysis product (White & Rose, 1997); (Davies & Henrissat, 1995). Previous studies have proposed that the binding of the substrate generates a distortion of the -1 NAG subunit to a boat conformation preceding the formation of the non-covalent oxazolinium ion intermediate. The acetamido group of the -1 distorted sugar will perform the nucleophilic attack, enabling the formation of this intermediate ion i.e. the so-called substrate-assisted mechanism (Brameld & Goddard, 1998); (Tews *et al.*, 1997). Simultaneously to the nucleophilic attack, the catalytic glutamate will protonate the glycosidic oxygen located between the -1 +1 NAG, which leads to the cleavage of the chitin (van Aalten *et al.*, 2001); (Suginta & Sritho, 2012).

In addition to the hydrolysis reaction, many chitinases, including the human ones, show a distinctive property consisting in the capacity of shifting the activity from hydrolyzation to transglycosilation (TG) in the presence of excess substrate concentrations. In TG, enzymes catalyze the formation of glycosidic bonds between donor and acceptor sugar units which leads to the re-polymerization of the substrate (Zakariassen *et al.*, 2011). However, the precise mechanism of this reaction is not yet clearly understood.

By combining X-ray data, biochemical experiments and computational calculations, extensive efforts have been devoted to elucidate the structure-function relationship including the features of the substrate binding in GH18 chitinases. Despite these efforts, the

detailed structural mechanistic basis is not fully understood and many aspects remain controversial. Indeed, the protonation pattern of the catalytic site and the proton pathway during the enzymatic reaction has not yet been elucidated. To determine the protonation states of the catalytic site, it is necessary to obtain crystals of better quality to achieve atomic resolution ($\sim 1 \text{ \AA}$).

In this study, we investigated the catalytic domain of the human chitotriosidase (CHIT1), as a model to probe the mode of action of the GH18 chitinases. This enzyme is one of the two active chitinases that have been identified in humans and reported to be involved in the innate immune response as well as being a biomarker of Gaucher disease (Hollak *et al.*, 1994). Further to its hydrolysis activity, it shows a high transglycosylation rate, a widespread phenomenon in GH18 chitinases (Aguilera *et al.*, 2003). In 2002, the first X-ray crystal structure of CHIT1 (PDB code: 1GUV) was determined at 2.35 \AA by Fusetti and collaborators (Fusetti *et al.*, 2002a) and, later on, Rao and collaborators obtained a structure of CHIT1 in complex with the inhibitor argifin at 1.65 \AA (PDB code: 1WAW) (Rao *et al.*, 2003). Those structures showed that this enzyme shares the same TIM barrel 3D folding of GH18 chitinases with the conserved catalytic motif (DXDXE) located in the active site (Fusetti *et al.*, 2002a). However, more detailed information is still required for a full explanation of proton translocation processes. Thus, in order to extend our knowledge regarding the catalytic properties, we obtained new X-ray data up to the range between $0.95\text{-}1.10 \text{ \AA}$ of CHIT1 in apo and pseudo-apo forms, and in complex with chitobiose i.e N- acetylglucosamine (NAG) dimer (from now on, chitobiose). Our crystal structures reveal the detailed internal organization of the active site residues, as well as their interactions with the substrate and allow us to investigate the protonation state of catalytic triad. Quantum mechanics calculations further supplied our crystallographic findings, confirming the observed protonation states of the catalytic triad and providing novel insights into the proton pathway during the hydrolysis reaction.

2. MATERIALS AND METHODS

2.1 Cloning, expression and purification

The full length human chitotriosidase gene CHIT1 was used as a template to generate the construct corresponding to the catalytic domain 1-386 residues (CHIT1) by the polymerase chain reaction (PCR) using the following primers (SIGMA): 5'-AATTCAAGCTTGCCACCATGGTGC GG TCTGTGG-3' (N-terminal derived sense primer) and two antisense primers to generate the 3' end encoding S386 with an additional thrombin site and a His-tag 5'-GTGATGGT GATGGTGGT GAGA ACCGCGTGGCACCAGACTCAGTTCCTGCCGTAG C-3'; 5'-ATTATCGCGATACTAGTCTCGAGTCATTAGTGATGGT GATGGTGGT G-3'. The PCR product was cloned into the pHL expression vector (Aricescu *et al.*, 2006). CHIT1 was transiently expressed in adherent HEK293T cells grown on roller bottles as described (Zhao *et al.*, 2011). After dialysis against 25 mM phosphate buffer saline (PBS), pH 8.0 at 4 °C, the secreted protein was purified from the media using an immobilized metal affinity chromatography (IMAC) batch procedure. CHIT1 was further purified by size exclusion chromatography on a Superdex 200 16/60 (GE Healthcare) column in 10 mM HEPES, 150 mM NaCl pH 7.5. The protein purity was assessed by SDS-PAGE followed by Coomassie Brilliant Blue staining (Laemmli, 1970). The enzyme concentration was determined from the absorption at 280 nm using an UV NanoDrop 1000 Spectrophotometer (Thermo Scientific). The molar extinction coefficient was calculated using the ProtParam tool on the ExPasy server (Gasteiger *et al.*, 2005) to 73590 M⁻¹ cm⁻¹. (See appendix 2)

2.2 Enzyme Activity Measurements

The enzymatic activity was determined using commercial synthetic fluorogenic substrates (Hollak *et al.*, 1994). Briefly, 0.25 nM of CHIT1 was incubated at 37 °C with various concentrations (0-200 μM) of the substrate 4-methylumbelliferyl β-N,N',N''-triacylchitotrioside (4-MU-NAG₃, SIGMA). The assays were performed in 100 mM citrate phosphate buffer, pH 5.6 containing 1 mg/ml bovine serum albumin. After 30 min, the assay was stopped by addition of 100 μL 0.3 M glycine-NaOH, pH 10.3. The product 4-methylumbelliferone was quantified using a microplate fluorometer (excitation 360 nm/emission 440 nm).

2.3 Crystallization and data collection

For crystallization, of CHIT1 apo form, the protein was concentrated to 9 mg/ml in 10 mM HEPES, 150 mM NaCl pH 7.5 buffer and crystals were grown by the hanging-drop vapour-diffusion method at 17 °C by mixing 1.5 µL proteins with an equal volume of reservoir containing a low concentration of micro-seeds. The crystals that could reach a maximum size of 1 x 0.14 x 0.10 mm³, appeared after 4 to 6 days of equilibration against 500 µL of reservoir solution containing 24-26 % (w/v) polyethylene glycol (PEG) 3350, 0.2 M potassium sodium tartrate (PST) at pH 7.2. (See appendix 2)

The pseudo-apo crystal form was obtained after failing to co-crystallize the protein with chitin (chitin from shrimp shells, SIGMA). A saturated stock solution of chitin was prepared in DMSO and dissolved in the reservoir solution (also containing a low concentration of micro-seeds) which was added to the drop. The crystals obtained were collected at 0.95 Å resolution but chitin was not observed, therefore this structure was considered as a pseudo-apo form.

Crystals containing the artificial substrate 4-MU-NAG₃ (SIGMA) were obtained in the same crystallization condition than the CHIT1 apo form combined with micro-seeding, by mixing 1.5 µL protein with an equal volume of the reservoir solution containing a range of final concentration of 4-MU-NAG₃ between 0.3 mM and 2.5 mM. Hydrolysis of the substrate occurred in the drop, therefore, crystal of CHIT1 complexed with only two N-acetylglucosamine monomers i.e. chitobiose appeared after 10 to 30 days. All the crystals obtained for CHIT1 apo form, pseudo-apo form or in complex with chitobiose were cryo-cooled in liquid nitrogen, using a solution containing 35 % PEG 3350, 0.2 M PST.

X-ray diffraction data were collected on the PXIII X06DA beamline of the Swiss light Source (SLS) synchrotron, Villigen, Switzerland. All data sets were integrated, merged and scaled using HKL-2000 (Otwinowski & Minor, 1997). The structures were solved using molecular replacement with Phaser (McCoy *et al.*, 2007) using the coordinates of the native structure of the same protein as an initial search model (PDB code: 1GUV) (Fusetti *et al.*, 2002a). The model was improved by alternating cycles of manual model building using Coot (Emsley *et al.*, 2010), refined using REFMAC5 (Murshudov *et al.*, 2011) and using PHENIX (Adams *et al.*, 2010). The stereochemistry quality of the final model was assessed with

PROCHECK (Laskowski *et al.*, 1993). To determine the protonation states of some polar residues in the active site an additional refinement with removed stereochemical restraints for the C-O bond lengths was performed using SHELXL (Sheldrick, 2008). The protonation states were determined by measuring and analyzing the C-O bond lengths. Generally, neutral carboxyls have unequal lengths of around 1.21 and 1.32 Å for the C=O and C—OH bonds, respectively. In contrast, negatively charged carboxyls are expected to have identical C—O bond lengths of around 1.26 Å (Erskine *et al.*, 2003). Structural figures were prepared using PyMOL (<http://www.pymol.org>). A summary of the data-collection and structure-refinement statistics is given in Table 1.

2.3 Quantum mechanics/molecular mechanics (QM/MM) calculation method

For the combined QM/X-ray refinement we use the program QMRx, based on the general purpose driver DYNGA (Parker *et al.*, 2003). This program is similar to other available codes capable of hybrid QM/X-ray refinement like ChemShell (Metz *et al.*, 2014), ComQum (Ryde *et al.*, 2002) or Phenix/DivCon (Borbulevych *et al.*, 2014) but offers the main advantage that arbitrary slave QM codes can be used with it, resulting on a much larger set of options. Among those options was the possibility to use the PM7 hamiltonian for the QM calculations instead of the usual PM6 parameterization (Borbulevych *et al.*, 2014), (Stewart, 2009). The much higher accuracy of PM7 has been discussed in the literature in great detail (Stewart, 2013). We use the MOZYME (Stewart, 2009) implementation of PM7 as available in MOPAC2012 under QMRx. This implementation offers several advantages. Firstly, we can avoid the use of a divide and conquer or other highly localized approaches during the QM calculation. The use of such methods has been the subject of controversy since it has been demonstrated that the degree of localization of the wave function can greatly affect the charge distribution in the enzyme (Wick *et al.*, 2014). The problem of charge assignments and the ways in which the QM treatment affects them has also been a subject of discussion in the literature (Yu, Hayik, *et al.*, 2006). In our calculations we address this problem in two significant ways: by drastically increasing the QM integral cutoffs, and by introducing a dielectric screening function as recommended in the literature (Andersson *et al.*, 2013). The effect of these additions is a large increase in the total CPU time which limits the total

number of conformations that can be explored in a reasonable amount of time but affords a much more accurate description of the system. For the X-ray section QMRx either uses Xplor-NIH (Wick *et al.*, 2014) or ShelX as a slave program to obtain the first order derivatives of the X-ray restrain function. The mixing of the QM and X-ray restrains was performed using the standard protocols as described in the literature (Metz *et al.*, 2014); (Ryde *et al.*, 2002); (Yu, Li, *et al.*, 2006); (Falklöf *et al.*, 2012). The procedure uses the standard approach for constraint weight evaluation as described in MM/X-ray methods for the evaluation of the restrain (Brunger & Rice, 1997) albeit applied to smaller random regions of the structure due to the much larger computational cost of the QM calculation. The restrain weight factor in our case was confined to the range 0.2/0.3. The convergence during optimization was monitored by following the norm of the total gradient (ng) until we achieved $ng < 1.25$. It should be noted that the speed of convergence of the minimization is much slower when using a QM/X-ray method than what is normally observed in molecular mechanics optimization due to the presence of much softer intramolecular interactions than what is described in an MM method.

3. Results and discussion

In this study we present a detailed structural and mechanistic analysis of the CHIT1 catalytic domain based on several X-ray crystal structures at atomic resolutions. To the best of our knowledge, this is the first report where a GH18 chitinase member is studied at such a resolution. This family, which is structurally characterized by a conserved 3D folding consisting of $(\alpha/\beta)_8$ TIM barrel, does not show a particular high sequence similarity (average pairwise identity 21%; <http://www.sanger.ac.uk/Software/Pfam>) (Synstad *et al.*, 2004). Nonetheless, the catalytic triad with the DXDXE signature is fully conserved in all active GH18 chitinases (Bussink *et al.*, 2007). Accordingly, CHIT1 is a reliable model to understand the detailed structural basis of the catalytic mechanism in GH18 chitinases. Our work has focused on studying the geometry changes within the catalytic triad of CHIT1 during different stages in the enzyme catalysis. Additionally, extensive analyses were applied to combine the observed geometric shifts to the repercussive protonation state modifications within the side chains of the main residues in the catalytic site. Consequently, correlations

between the geometric rearrangements and the proton translocations have been established.

3.1 Atomic resolution structures of the catalytic domain of CHIT1 reveal double conformation of key catalytic residues

The previously reported crystal structures of CHIT1 in the apo form (PDB code: 1GUV) or in complex with a substrate (chitobiose, PDB code: 1LG1), determined by Fusetti and colleagues (Fusetti *et al.*, 2002a), showed the catalytic domain at 2.35 Å and 2.78 Å resolution respectively. Even with the presence of an inhibitor, the highest resolution obtained was 1.65 Å (in complex with the inhibitor argifin, PDB code: 1WAW) (Rao, Andersen, *et al.*, 2005). Since our aim was to perform detailed structural mechanistic studies and protonation states analysis, we needed to improve the resolution of those structures to atomic resolution. We therefore set up new crystallization conditions combined with micro-seeding, which allowed us to obtain crystals for CHIT1 with higher diffraction quality ~1.0 Å (section 2.3 in materials and methods).

As expected, the final model of both CHIT1 apo form and pseudo-apo forms comprised 365 amino acids with 2-3 additional residues observed belonging to the thrombin site. The His-tag did not appear and is presumed to be disordered. The overall tertiary structure, determined at 0.95 Å (pseudo-apo form) and at 1.0 Å (apo form), is a TIM barrel and resembles the one already determined at 2.35 Å (Fusetti *et al.*, 2002a). Increasing the resolution of CHIT1 apo form allowed us to gain more insight into the structural features of the catalytic groove and of the active site, in particular concerning its flexibility. Interestingly, at this resolution, several residues in the active site are detected in multiple conformations and, in particular, the residues of the catalytic triad (D₁XD₂XE) adopt double conformations (Fig. 1A, 2A). In contrast, at 2.35 Å resolution (PDB code: 1GUV) the multiple conformations of many of these residues were unresolved.

In the 1.0 Å apo structure, the catalytic glutamate shows a double conformation: E140-confA, a slightly rotated non-planar rotamer (16% occupancy), and E140-confB, a major planar rotamer (84% occupancy) (Fig. 1A, B). In addition, the middle aspartate (D138) adopts two conformations (Fig. 1A, C, D and E). D138-confA is oriented towards D136 (45% occupancy) and forms a short H-bond with it (2.55 Å, suggesting the presence of a low

barrier hydrogen bond (LBHB) (Fig. 1D, E). D138-confB (55% occupancy) is flipped towards E140 with whom it forms a short H-bond (2.57 Å, suggesting again a LBHB), stabilizing its conformation (Fig. 1E). Thus, D138-confA is linked to D136 while D138-confB to E140-conf B (Fig. 1D, E). Moreover, the outer oxygen of E140-confA forms a H-bond with the conformation A of the water molecule w429 (w429-A, 49% occupancy). w429-A shows a slight displacement of 1.57 Å in relation to its B conformation (w429-B, 51% occupancy) (Fig. 1D, E). This slight shift is a consequence of two different E140 conformations.

During our analysis, we compared the overall occupancy distributions between the residues of the catalytic triad and the surrounding water molecules. Indeed, E140 seemed practically unaffected by D138 conformational changes. This can be explained by the presence of a second water molecule which occupies the same position of D138-confB (w214-A, 40% occupancy, Fig. 1C). Interestingly, w214-B (60% occupancy) is interacting with E140-confB, which will happen only when D138-confB is interacting with E140-confB. On the other hand, when D138 i.e. D138-confA is flipped towards D136, w214-B might move from this frontal position to the location occupied by D138-confB inner oxygen (Fig. 1D, E).

It is worth noting that D138-confB was not detected in the previously published 2.35 Å apo structure and the corresponding lower resolution density was modeled as a water molecule. It was then considered that the catalytic triad adopted only one stable state (corresponding to D138-confA, E140-confA and w214-A in our 1.0 Å apo structure). In contrast, our data suggest the existence of a dynamic equilibrium of the catalytic triad in the apo form. To note also that the conformation of D136 remains unchanged because it is stabilized by H-bonds with Y27 and S181 (Fig. 1D, E). When D138 turns toward D136, w214 moves down and mimics the position of the inner oxygen of D138, thus H-bonding the inner oxygen of E140 side chain. Simultaneously, w429 stabilizes the outer oxygen of E140 (Fig. 1D). Simply put, the position and the reorganization of water molecules occurring in coordination with D138 flipping limit the flexibility of E140 which explains the dominant planar conformation.

In most of the apo structures from GH18 members only a single conformation of the middle aspartate (D₂) is observed. This is mainly due to limitation given by resolutions lower than 2 Å, and to the use of molecules such as glycerol or ethylene glycol in the cryo-solution,

capable of interacting with the catalytic glutamate (Yang *et al.*, 2010); (Fusetti *et al.*, 2002a). Interestingly, D₂ double conformation has been described before in other apo structures of GH18 members, such as ChiA from *Serratia marcescens* (Papanikolau *et al.*, 2001) and chitinase D from *Serratia proteamaculans* (Madhuprakash *et al.*, 2013), with resolutions of 1.55 and 1.49 Å, respectively. Besides, D₂ mutation has been shown to abolish the enzymatic activity (Papanikolau *et al.*, 2001); (Bokma *et al.*, 2002). Indeed, the X-ray crystal structure of the mutant archaeal chitinase complexed to NAG₄ (PDB code: 3A4X), where D₂ was substituted to an alanine, showed an altered conformation of the catalytic glutamate. This resulted in making this latter 5 Å far from the scissile oxygen of glycosidic bond (Bokma *et al.*, 2002).

Altogether, this behavior of D₂ seems to be a conserved feature of this residue in GH18 chitinases and reinforces the idea that these flipping conformational changes could be essential to stabilize the catalytic glutamate in the apo form. This geometric stabilization could in fact maintain a favorable conformation for the substrate cleavage.

3.2 Protonation state of the catalytic triad residues of CHIT1 apo form provides insight into its hydrolysis mechanism

To gain insight into the physicochemical properties of the catalytic triad in CHIT1, we decided to investigate its protonation pattern in the apo form. Usually, ultrahigh-resolution X-ray crystal structures have the potential to reveal exact H atoms-atom positions (Afonine *et al.*, 2010), or an alternative method is the measure and analysis of the bond lengths (Ahmed *et al.*, 2007). The advantage of the last approach is that it avoids the technical difficulty of locating H atoms from weak electron density. C-O bond lengths can be determined by performing a supplementary refinement with no stereochemical restraints using SHELXL (Sheldrick, 2008) This refinement optimizes the carboxyl bond lengths according to the electron density. However, in the case of our apo structure at 1.0 Å, this strategy turned out to be insufficient. Indeed, our data corresponds to a snapshot which reflects two states of the catalytic triad with D138 and E140 adopting partial conformations. Therefore, the measured carboxyl bond lengths of E140 and D136 represent an average of the two states and do not reflect the situation of one of the two conformations. Moreover,

the standard average error for the carboxyl bond length of D138 in each conformation A is relatively high ($> 0.035 \text{ \AA}$).

Unexpectedly, when we attempted to co-crystallize CHIT1 with crystalline chitin, we improved the resolution up to 0.95 \AA , but we did not find any electron density corresponding to such polymer in this structure; therefore, we consider it as a pseudo-apo form. Crystalline chitin is known to not interact with the catalytic domain in absence of the chitin binding domain which is our case. This structure shows D138-confB as dominant conformation (80% occupancy) as well as E140-confB (90% occupancy) (Fig. 2A). Moreover, the organization of the water molecule network in the catalytic groove was closely similar to the 1.0 \AA apo structure, with the exception that water molecules w214 and w429 showed only one conformation (for the sake of clarity regarding the residue conformations, we will call them B), consistent with a predominant D138-confB. We then employed the 0.95 \AA pseudo-apo structure as an improved model for analyzing the protonation pattern of the catalytic triad through the C-O bond lengths measure.

Following the refinement in phenix.refine, the 0.95 \AA pseudo-apo structure was refined using SHELXL (Sheldrick, 2008), with removed stereochemical restraints on the carboxyl moieties of all glutamate and aspartate residues. Our results show that on D136 inner oxygen, the (C—O δ 1) bond was refined to 1.26 \AA while on the outer oxygen O δ 2, the (C—O δ 2) bond was refined to 1.29 \AA , suggesting a partial protonation on this O δ 2. In parallel, the phenol oxygen of Y27 reveals a C—O bond of 1.35 \AA , suggesting also partial to total protonation of the hydroxyl of Y27 (Fig. 2B, C). Taken together and since Y27 and D136 make a short H-bond (2.60 \AA), these data suggest that these two residues share a proton within a LBHB (Fig. 2B, C). To further validate our results, QM/MM was performed with the 0.95 \AA pseudo-apo structure. The aforementioned interpretation was also supported and expanded by the QM/MM calculations, which showed low barrier energy for the proton sharing between these two residues.

In the case of D138-confB, the (C—O δ 2) bond displayed a bond length of 1.22 \AA , indicating that it is deprotonated, while the (C—O δ 1) bond of D138-confB was refined to 1.27 \AA , meaning that it is partially protonated (Fig. 2B, C). In parallel, the (C—O ϵ 1) bond of dominant conformation E140-confB was refined to 1.23 \AA while the (C—O ϵ 2) showed a 1.30

Å bond length. The carboxylate bond lengths between E140 and D138 are consistent with the presence of a short LBHB of 2.49 Å between these two residues and underline the surprising finding that the outer oxygen is not protonated in the apo form when D138 flips towards E140. In fact the values obtained led us to conclude that the inner oxygen of D138-confB (O δ 1) shares a proton with the inner oxygen of E140 (O ϵ 2) where the affinities to the proton between these two residues are closely similar. In this regard, a round electron density signal appears between D138-confB and E140-confB, which may correspond to this shared proton (Fig. 2D).

Furthermore, the QM/MM calculations confirm the X-ray observations and also indicate that a proton is shared between O δ 1 D138-confB and O ϵ 2 E140-confB in the apo form of CHIT1.

To investigate the protonation state of the catalytic triad when D138-confB flips to D138-confA, we based our study on QM/MM calculations data. This is because in the 0.95 Å pseudo-apo structure the D138-confA has an occupancy of 20% which is not sufficient to make an accurate bond length analysis as the standard errors were high (see table 2). Interestingly, QM/MM calculations revealed that, when D138-confA forms a LBHB with D136, there is a proton sharing phenomenon between Y27, D136, and D138. Moreover, in this conformation, the outer oxygen of E140 O ϵ 1 stays deprotonated and the inner oxygen O ϵ 2 forms a H-bond with water molecule w214A as shown in Figure 1D for the apo structure at 1.0 Å. Combining the X-ray data and QM/MM from both structures (apo and pseudo-apo), we propose that CHIT1 possesses an unusual system to “stock” the proton before hydrolysis. This system involves at least four residues (Y27, D136, D138 and E140), where D138, by flipping constantly, swings the proton to each side of the catalytic site from D136 to E140. Remarkably, our finding reveals that in contrast with the previously reported data (van Aalten *et al.*, 2001); (Papanikolaou *et al.*, 2001); (Fusetti *et al.*, 2002a); (Jitonnom *et al.*, 2011); (Jitonnom *et al.*, 2014), CHIT1 maintains the outer oxygen O ϵ 1 of E140, which is supposed to donate the proton to cleave the glycosidic bond, deprotonated. Taken together, our data questions the previous published model of the hydrolysis mechanism (van Aalten *et al.*, 2001) as well as the proposed transglycosylation model (Zakariassen *et al.*, 2011).

3.3 Atomic resolution structures of the catalytic domain of CHIT1 in complex with chitobiose provide insight into catalytic mechanism

Studies of the complex of CHIT1 with a substrate were conducted to determine how the catalytic residue E140 could protonate the oxygen of the scissile glycosidic bond. The limiting step for the comparison with the apo CHIT1 structure was again to obtain a resolution of 1.0 Å or lower. As the soaking experiments were destabilizing the crystals resulting in loss of high resolution diffraction, we then conducted co-crystallization experiments of CHIT1 with different concentrations of the synthetic substrate 4-MU-NAG₃. Fortunately, we succeeded in developing a robust protocol to co-crystallize CHIT1 with 4-MU-NAG₃ by means of micro-seeding. This methodology allowed us to control crystal growth and get high quality crystals diffracting at atomic resolution. Effectively, co-crystals with 4-MU-NAG₃ at 0.3 mM, 1 mM and 2.5 mM concentrations reached X-ray data resolution at 1.10 Å, 1.05 Å, and 1.10 Å respectively (Table 1 and Fig. 3).

However, all the structures (solved, by molecular replacement (MR) using the apo structure as an initial model) have revealed CHIT1 complexed to a dimer of N-acetyl glucosamine (chitobiose) located in the -1 and -2 subsites. This indicates that the hydrolysis occurred in the drop and thus allows us to analyze the post-hydrolysis 3D structure of CHIT1. All our electron density $2F_o-F_c$ maps of the three structures show the -2 NAG in chair conformation, while the -1 NAG adopts a boat conformation (Fig. 3B), which disagrees with the previously published structure of CHIT1-chitobiose at 2.78 Å in which the -1 NAG was modeled in chair conformation (PDB code: 1LG1) (Fusetti *et al.*, 2002b). Most probably, the low resolution of this structure impeded the clear determination of the -1 NAG configuration and could thus explain this disagreement. Moreover, the boat conformation of the -1 NAG seen in our structures is consistent with the substrate distortion event, described in other GH18 chitinases and reported to be required for the substrate-assisted mechanism in this enzyme family (Brameld & Goddard, 1998); (Songsiriritthigul *et al.*, 2008); (van Aalten *et al.*, 2001).

Interestingly, by comparing the three structures, our data indicate a gradual increase of chitobiose occupancies in the catalytic groove, consistent with the augmentation of substrate in the drop. As a result, the occupancy of chitobiose was refined to 50%, 69% and

80% for the structures with 0.3 mM, 1 mM and 2.5 mM substrate concentrations, respectively (Fig. 3B). Remarkably, in the condition where the lowest substrate concentration (0.3 mM) and occupancy (50%), we have noticed that the occupancy of E140-confA, which is minimal (16%) in the apo form, significantly increases to 41%, becoming closer to the occupancy of the E140 planar conformation (E140-confB) (Fig. 3A ,B). In the same condition, D138 also shows a quasi-equal occupancy of conformations A (55%) and B (45%) (Fig. 3C). Notably, by supplementing the substrate amount in the drop, the occupancy of chitobiose in the binding site gradually increases (69% in the structure with 1 mM and 80% in the structure with 2.5 mM of substrate concentration). This augmentation was also accompanied by higher occupancies of E140-confB (71% and 100% in the structures with 1 mM and 2.5 mM of substrate, respectively) and D138-confB (70% and 89% in the structures with 1 mM and 2.5 mM of substrate concentration, respectively), as well as a decrease in the slightly rotated non-planar conformation of E140 (E140-confA) (29% and 0% in the structures with 1 mM and 2.5 mM of substrate, respectively) (Fig. 3C). On one hand, this confirms the previously reported data indicating that the presence of the substrate induces the rotation of D138 towards E140 (van Aalten *et al.*, 2000). On the other hand, our observations reveal for the first time two clear conformations for E140 in presence of a hydrolysable substrate: one planar and one rotated, indicating that such movement occurs (Fig. 3A). This observation prompted us to think that we detected at least two states of the enzyme. In the state where the planar conformation is adopted by E140 (E140-confB), this conformation is stabilized by H-bond contacts provided via the chitobiose on the outer side (OE1) and D138-confB on the inner side (OE2) (Fig. 3E). In contrast, it seems that in the conformation A of our CHIT1-chitobiose structures, in which there is no stable interacting with the catalytic triad, E140 displays a rotated conformation (E140-confA) (Fig. 3A). This idea is reinforced by the fact that when the occupancy of chitobiose increases, the occupancy of the rotated conformation of E140 decreases until it becomes negligible in the co-crystal grown at 2.5 mM of substrate.

Regarding substrate recognition and binding, Songsiriritthigul *et al.* (2008) have reported that the chito-oligosaccharide chain is in a linear form during the initial step of substrate recognition. In the next step, the substrate chain performs a bend step leading to

the distortion of the -1 NAG to a boat conformation (Songsiriritthigul *et al.*, 2008). According to Songsiriritthigul and collaborators, the bend is accompanied by a twist of the glycosidic bond to make it accessible for the cleavage by the catalytic glutamate. Consistent with their evidences, in our structures with 0.3 mM and 1 mM of substrate, the rotated E140confA establishes a distance of 2.63 Å and 2.54 Å respectively with the chitobiose (E140-OE2 — O1B-chitobiose). This distance is increased to 2.73 Å (in the structure at 0.3 mM substrate) or 2.70 Å (in the structure at 1 mM of substrate) when E140 adopts the planar E140confB conformation, indicating that the rotated conformation (E140-confA) favors the cleavage of the substrate (Fig. 3D). In this regard, a similar rotation of E140 has already been detected in CHIT1 and AMCase complexes with allosamidin derivatives mimicking the intermediate (at a lower resolution than the CHIT1-chitobiose complex aforementioned), meaning that the E140 adopts this rotated position during the transition state (Fusetti *et al.*, 2003); (Olland *et al.*, 2009a). Altogether, our results suggest that upon substrate bending and twisting, the catalytic glutamate also rotates to gain access to the glycosidic bond. Such rotation could not be possible if D138 is in conformation B as it stabilizes the inner side oxygen OE2 of E140. Based on this analysis, we propose that the arrival and distortion of the -1 NAG displaces the water molecules which interact with E140 and D138 in apo form. Thus, E140-confA should rotate when D138 adopts conformation A. The displacement of the water network leads to the loss of the dynamic equilibrium described in 3.1 and 3.2 (Fig. 1D, E and 2B) which was limiting the mobility of E140. Hence, E140 is “free” to rotate when D138 turns towards D136.

Overall, one can say that upon substrate arrival, the “apo” dynamic equilibrium is destabilized allowing E140 to rotate for having the scissile oxygen of the glycosidic bond in an accessible position.

3.4 Structural analysis of the catalytic triad residues in the CHIT1-chitobiose structure reveals the coexistence of two enzymatic states in the same crystal form

Having highlighted the importance of the rotation of E140 in the process of hydrolysis, we next wanted to investigate its role in the substrate. We therefore studied the protonation state of the catalytic triad based on the analysis of the carboxyl bond length distances and QM/MM in CHIT1 complexed to chitobiose.

As aforementioned, in the structure with 0.3 mM and with 1 mM substrate, the chitobiose has a lower occupation than the structure with 2.5 mM. This latter has a single conformation for E140 while in the two others structures E140 and D138 display double conformations leading to decrease the electron density peak for each conformation, and thereby increasing the average error for the conformation with lower occupation. As a result, we could not determine the protonation pattern of the E140-confA since this conformation did not reach more than 41% occupancy in all the solved structures with substrate. This occupancy is not enough to obtain a low standard average by using SHELXL refinement. Consequently, to overcome this problem, we performed QM/MM calculations using the structure with 1 mM substrate, because this latter contains two conformations of D138 and E140 in presence of chitobiose. Therefore, this structure allowed determining the charge of the catalytic triad when the conformation of E140 is rotated (E140-confA) and planar (E140-confB). Importantly, our QM/MM calculations have shown that the rotated E140-confB is protonated while D138-confA is deprotonated. This indicates that D138-confA has transferred a proton to E140 before it flips towards D136 when deprotonated. Once E140 gets protonated, it rotates to gain access to the oxygen of the scissile bond. In addition to the QM/MM analysis done with the 1 mM structure concentration of substrate, we performed an unrestrained refinement by SHELXL on the structure with 2.5 mM concentration, as it represents the most accurate model of the post hydrolysis state among our three structures (100 % planar conformation for E140 and D138-confB as a dominant conformation at 89%). We compared the results of the protonation states of SHELXL on the 2.5mM substrate structure with the QM/MM calculations for the 1 mM substrate complex.

Hence, the unrestrained refinement of the 2.5 mM substrate structure shows that the planar conformation (equivalent to E140-confB in the other structures) has a C—O ϵ 1 bond length of 1.27 Å and a C—O ϵ 2 length of 1.24 Å indicating that both oxygen atoms share the charge. On the other hand, D138-confB reveals a C—O δ ₁ bond length of 1.33 Å and a C—O δ ₂ of 1.19 Å indicating that the C—O δ ₁ is protonated (Fig. 3E). In this state, the E140-confB carboxyl side chain O ϵ 1 is stabilized by the scissile oxygen of -1 NAG and forms with it a short H-bond of 2.63 Å while the other carboxyl oxygen of E140-confB, O ϵ 2, forms a strong H-Bond (2.50 Å) with the O δ 1 D138-confB (Fig. 3E). Even though the interatomic distance

between the two oxygens (O ϵ 2-E140-confB and O δ 1-D138-confB) is similar between this and the pseudo-apo structure at 0.95 Å resolution, an important difference is revealed when we measured the carboxylate bond length of E140-confB. Thus, our results show that E140-confB becomes charged in the presence of 2.5 mM substrate concentration. In fact, the presence of such short interatomic distance O ϵ 2—O δ 1, an ionic profile for E140 together with a protonated oxygen in the C—O δ 1 bond of D138-confB prompt us to suggest that the short H-bond between O ϵ 2—O δ 1 is not a LBHB but a strong ionic hydrogen bond (SIHB) (Meot-Ner, 2012). Such information needs to be confirmed by neutron diffraction or NMR. Indeed, this strong unusual nonstandard short H-bonds have been recently revealed by neutron crystallography in elastase and photoactive yellow protein (Yamaguchi *et al.*, 2009); (Tamada *et al.*, 2009). Regarding the 2.63 Å interatomic distance between O ϵ 1-E140 and O1 of chitobiose, it is not possible to determine whether it is LBHB or SIHB by X-ray crystallography. Nonetheless, we can propose that after a hydrolysis cycle the carboxylate of E140 becomes charged and bordered by two short H-bonds on each side of its carboxylate.

To further support our data, we decided to calculate the pKa of all the polar residues within CHIT1 in both the apo form and chitobiose structures. The calculated pKa via the server <http://propka.ki.ku.dk/> for the chitobiose structures indicate 7.6 and 13.4 for D138 and E140, respectively while they are 12.3 and 6.55 in the apo form for D138 and E140, respectively (Table 2). The significant shift in the pKa values indicates that in presence of the substrate an inversion of the acid/base profile of E140 and D138 occurs. In fact, E140 which was acidic in the apo form becomes basic in presence of chitobiose, whereas D138 which was basic gets converted into an acid residue. The decrease in the D138 pKa in presence of chitobiose is most likely due to the formation of a H-bond between the N-acetyl group of the -1 NAG moiety and the outer oxygen of D138.

Altogether, upon arrival of the substrate, this pKa shift supports that D138-confB will transfer a proton to E140. This is then followed by D138 flipping and E140 rotation to gain access and protonate the oxygen of the glycosidic bond.

3.5 Detailed structural analysis sheds new light into the hydrolytic step

All the proposed hydrolysis mechanisms in GH18 chitinases have reported that after the cleavage of the glycosidic bond through its protonation by the catalytic glutamate (E140 in CHIT1), an oxazolinium ion intermediate is generated and a water molecule gets activated by the same catalytic glutamate (van Aalten *et al.*, 2001); (Songsiriritthigul *et al.*, 2008); (Tews *et al.*, 1997); (Papanikolau *et al.*, 2001); (Jitonnom *et al.*, 2014); (Jitonnom *et al.*, 2011). According to these proposed hydrolysis models, during the activation of the hydrolytic water molecule, the catalytic glutamate receives a hydrogen and the -OH group of the water molecule performs a nucleophilic attack on the anomeric carbon leading to the re-formation of the -1 NAG moiety with retention of the initial configuration (van Aalten *et al.*, 2001); (Songsiriritthigul *et al.*, 2008). In contrast to the generally accepted mechanism, our results show that after the hydrolysis, the E140 is ionic and forms a H-bond with the acquired -OH on the anomeric carbon C1. This leads us to think that the activation of the hydrolytic water might have not been carried out by E140 since as a result of the catalysis, E140 is not protonated. We can therefore hypothesize that another residue could participate/be responsible in the activation of the hydrolytic water molecule. One candidate residue for this task is probably D213, which is located opposite to E140 and possesses an outer oxygen forming a short contact with the water molecule (w300) (Fig. 4A, B, C). The contact of D213 with the water molecule (w300) occurs in the apo form and in presence of chitobiose where it forms a H-bond with $-O_6H$ of the -1 NAG moiety. However, when we superimposed a long chain of NAG polymer in the subsites -4 to +2 based on the crystal structure of mutant ChiA, whose active site is highly similar to the one of CHIT1, the w300 overlaps with the +1 NAG. This means that upon substrate sliding to +1 and +2 subsites, it displaces this water molecule which was present in apo, whereas after the cleavage this water molecule regains its +1 position and gets stabilized by D213 after the displacement of the aglycon. It is worth noting that D213 is highly conserved in GH18 chitinases and its stabilization of the water molecule (w300) also appears in the crystal structure of chitinase D from *Serratia proteamaculans* at 1.49 Å (Madhuprakash *et al.*, 2013). In ChiB from *Serratia marcescens* the mutation of D215 (D213 in CHIT1) resulted in a mild activity of this enzyme and to an acidic shift in its pH optimum (Synstad *et al.*, 2004). Strikingly, several studies have

demonstrated that the mutation of the equivalent D213 to an alanine is deleterious for the chitinase activity (Synstad *et al.*, 2004); (Papanikolau *et al.*, 2001). Taken altogether, these data suggest that the activation of the water molecule might not be performed by the catalytic glutamate (as it becomes charged after a full hydrolysis cycle) but by another residue, which could be through D213 in CHIT1.

3.6 A new scenario for the processive hydrolysis

Based on the analysis of the different observed occupancies, H-bond sorts, C—O bond lengths as well as QM/MM calculations, we have join it to the scenario of the hydrolysis reaction linked to the enzymatic processivity of CHIT1. The mechanism we suggest consists of:

i) In the apo form of CHIT1, a dynamic equilibrium is established within the catalytic triad together with Y27, allowing the storage of the catalytic proton by a flipping conformational change of D138 as well as a back and forth movement of the water molecules w214 and w429 (Fig. 5A). ii) Upon the arrival of the substrate, the water molecules in the binding site get displaced and due to the loss of the dynamic equilibrium as well as the shift in the pKa, the D138 transfers a proton to E140 and turns, once deprotonated, towards D136 (Fig. 5B). Simultaneously, E140 rotates towards the twisted glycosidic bond and due to the elevation in its pKa, E140 protonates the scissile oxygen leading to the formation of the oxazolinium ion intermediate (Fig. 5C, D). This occurs accompanied by the displacement of the aglycon sugar which allows the hydrolytic water molecule to access the vicinity of the active site (Fig. 5D). At this point, D138 having received a new proton from D136, turns towards E140 stabilizing its rotated conformation to a planar conformation by a SIHB and establishes a H- bond with the nitrogen of the N-acetyl group in the -1 NAG moiety (Fig 5B, C, D). iii) A nucleophile residue (probably D213) activates the hydrolytic water molecule which in turn performs a nucleophilic attack on the anomeric carbon C1 of the intermediate ion leading the formation of the -1 NAG with retention of its configuration (Fig. 5E). iv) a hydrolysis cycle is completed at this point and the enzyme slides the substrate. During the substrate sliding, the -2 NAG arrives to the -1 subsite resulting in an unproductive binding as the N-acetyl group is placed in the opposite direction to the

catalytic triad. This allows the D138 to transfer the proton which was stocked between O ϵ 2—O δ 2 via a SIHB to E140. Moreover, as the outer oxygen O δ 2 of D138 is not stabilized by the N-acetyl group, this residue can now flip towards D136 and acquire a new proton and turn again towards E140 to continue a new hydrolysis cycle as described in ii).

This process repeats itself during the hydrolysis of the chitin chain. Consequently, the flipping conformational change of D138 in apo form is important for the storage of the proton where it swings the proton. However, in presence of a substrate chain, it functions as a shuttle of the proton from Y27 and D136 to E140. The repeating cycle in the processive hydrolysis involving many amino acids brings to mind that this enzyme processes the chitin chain according to a “fordist model”.

3.7 Detection of product in the catalytic groove reveals insight into the transglycosylation mechanism

The detection of chitobiose in the CHIT1 active site is not surprising as was previously also detected by soaking crystals of CHIT1 itself (Fusetti *et al.*, 2002a) or other native bacterial GH18 chitinases with chito-oligosaccharides (Malecki *et al.*, 2013); (Perrakis *et al.*, 1994). The presence of chitobiose in the catalytic groove could be explained by the stacking interactions made by two tryptophans (W31 which exist in -3 and -2 subsites and W358 in -1 subsites) as well as H-bond contacts with polar residues and water molecules in the binding site (Fig. 4B, C). Consistently, in-depth studies conducted by Eide *et al.* (Eide *et al.*, 2012) have also shown a high binding affinity to the NAG moieties in the subsites -2 and -1 subsites in CHIT1. Moreover, CHIT1 is known to be processive as well as other bacterial GH18 chitinases. Parenthetically, the equivalents of W31 and W358 in other GH18 chitinases are key residues relevant to ensure the processive capacity of these enzymes. For example, the mutation of W137 in chitinase A (W31 in CHIT1) in *Serratia marcescens* was reported to strongly affect processivity (Zakariassen *et al.*, 2009). Therefore, it is believed that presence of such tryptophans in the binding site is important to prevent the chito-oligosaccharide chain from leaving processive GH18 chitinases thereby allowing these enzymes to slide the polymer. On the other hand, it is known that chito-oligomer substrates have a successive alternation of the N-acetyl group position as each NAG unit is rotated 180° (Fig. 4B) in relation to the other. Thus, the sliding of 2 NAG units is sufficient to obtain a N-acetyl group

accommodated in the -1 subsite on the side of the catalytic triad. Once accommodated, this fulfills the condition for the substrate-assisted mechanism to be carried out, thereby leading to the cleavage of the glycosidic bond. As a result, the products of processive hydrolysis are disaccharides. Given that CHIT1 cleaves by dimers, chitobiose is the last cleaving unit which cannot be further cleaved nor slid as it is stabilized by interactions along the dimer. Hence, the fact that we obtained the complex CHIT1-chitobiose not by soaking but by long duration co-crystallization experiments (4 weeks) together with the observation that the chitobiose occupancy increases proportionally to the substrate concentration makes us suggest that the complex CHIT1-chitobiose is a relatively stable one. We propose that the high affinity of -2 and -1 subsites which causes a relatively high stability of the dimeric product (chitobiose) in the catalytic groove blocks the -n subsite and thus represents the basis for a low saturation enzymatic capacity. Furthermore, CHIT1 is known to display a high affinity for NAG moieties at the +n subsite due to the presence of aromatic residues (W99, W218, Y190) (Fig. 4D). This suggests that the combination of both, the obstruction at the -n subsites with the high substrate affinity at the +n subsites turns these subsites into substrate acceptors. Hence, we believe that this new substrate repositioning lies at the foundation of the re-polymerization phenomenon known as transglycosylation (Taira *et al.*, 2010); (Zakariassen *et al.*, 2011). Based on this result, it is possible that the transglycosylation occurs after the end of a hydrolysis cycle and not necessary during hydrolysis meaning that it can occur for example after the generation of -1 NAG moiety in the -1 subsite.

Altogether, these data provide further structural insights on the previously reported high transglycosylation rate in CHIT1 (Aguilera *et al.*, 2003).

4 CONCLUSION

In this study, for the first time, we report atomic resolution structures of CHIT1 apo form and in complex with chitobiose, by means of X-ray crystallography. We have extended our study to the protonation state of the catalytic residues with the combined use of unrestrained refinement with SHELXL and QM/MM calculations, which have revealed new insights regarding the catalytic mechanism of the hydrolysis reaction in CHIT1, whose main features are conserved in the GH18 chitinase family. Indeed, we hereby provide new findings

regarding the role of D138 as a swing in the apo form and a proton shuttle during hydrolysis. Strikingly and in contrast to what was previously assumed, our study on the protonation state of the key catalytic residue E140 reveals that the outer oxygen of E140 is deprotonated in apo form and adopts an ionic state after hydrolysis. Furthermore, our investigation on the geometry of E140 showed for the first time a rotation that liberates E140 from D138 and therefore allows the protonated oxygen to better access to the glycosidic bond and to cleave it. Importantly, we indicate a shift in the type of H-bond established between D138 and E140 from a LBHB in the apo form to a SIHB in complex with chitobiose which could be important to maintain the ability to perform many hydrolysis cycles. Moreover, our results underline the low barrier phenomenon of proton sharing taking place between Y27 together with D136 and D138 in the apo form for the proton storage. In addition, our findings highlight the putative role of Y27 and D136 in “supplying” protons to D138 thanks to a low energy barrier for proton translocation between these three residues during the hydrolysis cycle. Besides providing a deeper understanding of the hydrolytic mechanism, our structures of CHIT1-chitobiose have provided additional insights regarding the structural basis of the high rate transglycosylation in CHIT1. Finally, it is relevant to remark that this work also opens the door to a joint atomic X-ray plus neutron diffraction studies, which we aim to develop to gain further certainty in the modified CHIT1 catalytic mechanism that we have started to unveil. Overall, this advance in a fundamental science question also provides new knowledge useful for the design of more specific and powerful CHIT1 and GH18 chitinases inhibitors.

Acknowledgements

We thank the IGBMC Structural Biology and Genomics Platform staff for technical assistance. We are grateful to Dr. Irene Yujnovsky, Dr. Sergey Melnikov, and Dr. Francesc Xavier Ruiz Figuras for helpful discussions. The crystallographic experiments were performed on the X06DA (PXIII) beamline at the Swiss Light Source synchrotron, Paul Scherrer Institut, Villigen, Switzerland. We thank in particular, Ezequiel Panepucci and Vincent Olieric for their help on data collection. This work has been funded by the CNRS, the INSERM, the Université de Strasbourg, Biostruct-X (FP7, contract 283570) Instruct (European Strategy Forum of

Research Infrastructures; ESFRI) and the French Infrastructure for Integrated Structural Biology (FRISBI).

FIGURE LEGENDS

Figure 1. CHIT1 active site in apo form. A) Cluster of D136, D138, E140, Y27 in sticks and water molecules w187, w322, w214-A and B, w429 A and B in sphere. Y27, D136, D138confA, E140-confA, w214-A and w429-A colored in green. D138-confB, E140-confB, w214-B and w429-B colored in blue purple. $2F_o-F_c$ electron density map (1σ cutoff) of the cluster in mesh and colored in grey. B) Two conformations (rotated E140-confA and planar E140-confB) shown in lines within the $2F_o-F_c$ the electron density map (1σ cutoff). C) $2F_o-F_c$ electron density map (1σ cutoff) of D138 and E140-confB in grey and F_o-F_c in green of D138 and E140. D) and E) Stereoview of the hydrogen bonding network, occupancies percentage and distances in each conformation of the cluster shown in A). S181 is also nearby the catalytic triad, but it is not shown for the sake of clarity.

Figure 2. CHIT1 in pseudo-apo form. A) Cluster of D136, D138, E140, Y27 in sticks with occupancies percentage. $2F_o-F_c$ electron density map (1σ cutoff) of the cluster in mesh and colored in blue. Y27, D136, D138confA, E140-confA, in green. D138-confB and E140-confB, in blue purple. B) The hydrogen bonding network with the distances in conformation B of the cluster shown in A. C) Cluster of Y27, D136, D138-confB and E140-confB in sticks with the obtained bond length C—O by SHELXL refinement. D) $2F_o-F_c$ electron density map (1σ cutoff) of E140confB and D138-confB in blue and the F_o-F_c map (3σ cutoff) in green show a signal that could correspond to shared hydrogen between E140 and D138 in conformation B.

Figure 3. Structures of CHIT1-chitobiose complex obtained by using different concentrations of 4-MU-NAG₃. A) Two conformations of E140 shown in lines within the $2F_o-F_c$ electron density map (1σ cutoff) in blue. B) $2F_o-F_c$ electron density map (1σ cutoff) of chitobiose in the three structures of the complex with CHIT1. C) Cluster of Y27, D136, D138, E140, shown in sticks with the occupancies percentages in the complex CHIT1-chitobiose. Y27, D136, D138-confA, E140-confA colored in grey. D138-confB, E140-confB colored in blue purple. D) Zoom on the interaction between E140 and O1 of chitobiose. The H-bond length

are mentioned. E) The cluster of Y27, D136, D138-confB, E140-confB and chitobiose shown in sticks with the obtained C—O bond lengths by SHELXL refinement.

Figure 4. Stereoviews of the interactions of chitobiose in the binding and active site. A) and B) Different residues of the binding and active site involved in the interaction with chitobiose shown in stick. C) Zoom on the position of D138-confB, E140-confB, D213 and the water molecule w300 in presence of the chitobiose. H-bond distances are indicated. D) CHIT1 and chitobiose in surface and the aromatic residue W99, W218, Y190 in the subsites +1, +2, +3 represented in stick and colored in green.

Figure 5. Proposed hydrolysis mechanism in CHIT1 (See text). A) Dynamic equilibrium of the active site CHIT1 in the apo form. B) Substrate arrival. 1- Proton transfer from D138-confB to E140. 2- D138 Flip towards D136 to receive a new proton from this latter. C) -1 NAG twist to boat conformation and E140 protonate the scissile oxygen of the glycosidic bond. D) Aglycon sugar displacement, formation of oxazolinium ion intermediate and stabilization of D138-confB protonated. E) End of the hydrolysis cycle. Re-generation of -1 NAG with retention of its configuration. SIHB formation between D138-confB and E140 in the planar form.

Table 1. Data collection and refinement statistics

	CHIT1 pseudo-apo form	CHIT1 apo form	CHIT1- chitobiose 0.3 mM	CHIT1-chitobiose 1 mM	CHIT1-chitobiose 2.5 mM
PDB code					
Synchrotron, beamline	SLS, X06DA (PXIII)	SLS, X06DA (PXIII)	SLS, X06DA (PXIII)	SLS,X06DA (PXIII)	SLS,X06DA (PXIII)
Wavelength (Å)	0.8	0.8	0.8	0.8	0.8
Resolution range (Å)	50-0.95 (0.98-0.95)	50-1.0 (1.04-1.0)	50-1.1 (1.14-1.10)	50-1.05 (1.09-1.05)	50-1.101 (1.14-1.10)
Space group	P 21 21 2	P 21 21 2	P 21 21 2	P 21 21 2	P 21 21 2
Unit cell (Å)	85.33 103.70 41.69	85.69 105.75 41.52	85.67 106.18 41.43	85.5 105.52 41.475	85.502 103.434 41.58
Total reflections	2811923 (19720)	1121774 (10116)	1805445	1091078	1041198
Unique reflections	223038 (18926)	188444 (9732)	152249	175071	144158 (13705)
Multiplicity	12.6 (9.9)	6.0 (2.4)	11.9 (8.6)	6.2 (5.0)	7.2 (5.7)
Completeness (%)	95.99 (82.51)	92.51 (48.18)	99.51 (95.97)	99.78 (98.46)	96.32 (92.70)
Mean I/sigma(I)	36.08 (2.57)	27.50 (1.77)	40.0 (3.57)	31.92 (2.24)	20.38 (2.61)
Wilson B-factor	8.82	10.13	9.82	10.59	7.66
R-sym	0.059 (0.766)	0.041 (0.439)	0.057 (0.552)	0.046 (0.681)	0.085 (0.602)
R-factor	0.1142 (0.194)	0.1355 (0.231)	0.1396 (0.167)	0.1434 (0.220)	0.1433 (0.189)
R-free	0.1222 (0.189)	0.1476 (0.247)	0.1545 (0.185)	0.1534 (0.2293)	0.1622 (0.2083)
Number of atoms	7587	4107	3660	3778	6516
macromolecules	3518	3526	3096	3171	3062
ligands	20	40	29	87	58
water molecules	586	526	535	512	367
Protein residues	369	369	370	369	368
RMS (bonds)	0.008	0.005	0.006	0.013	0.011
RMS (angles)	1.38	1.15	1.19	1.35	1.36
Ramachandran favored (%)	98	98	98	99	99
Ramachandran outliers(%)	0	0	0	0	0
Clashscore	5.65	3.87	3.18	3.45	3.06
Average B-factor	12.70	13.80	13.40	14.00	10.20
macromolecules	10.80	12.10	11.50	12.20	9.20
ligands	24.80	29.00	13.90	14.50	7.50
solvent	22.90	23.90	24.50	24.80	18.70

Statistics for the highest-resolution shell are shown in parentheses

Table. 1
Fadel et al.

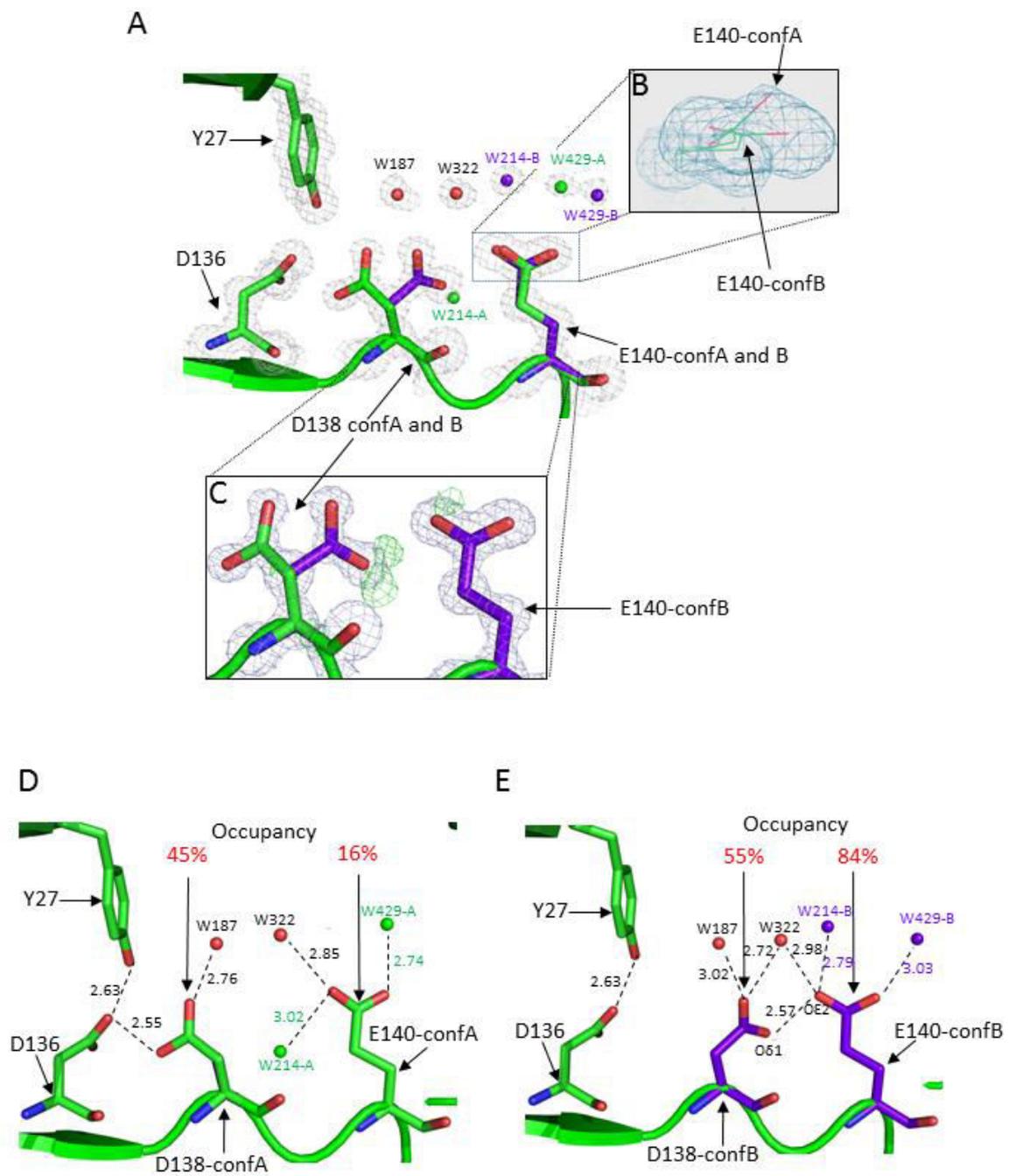


Fig. 1
Fadel et al.

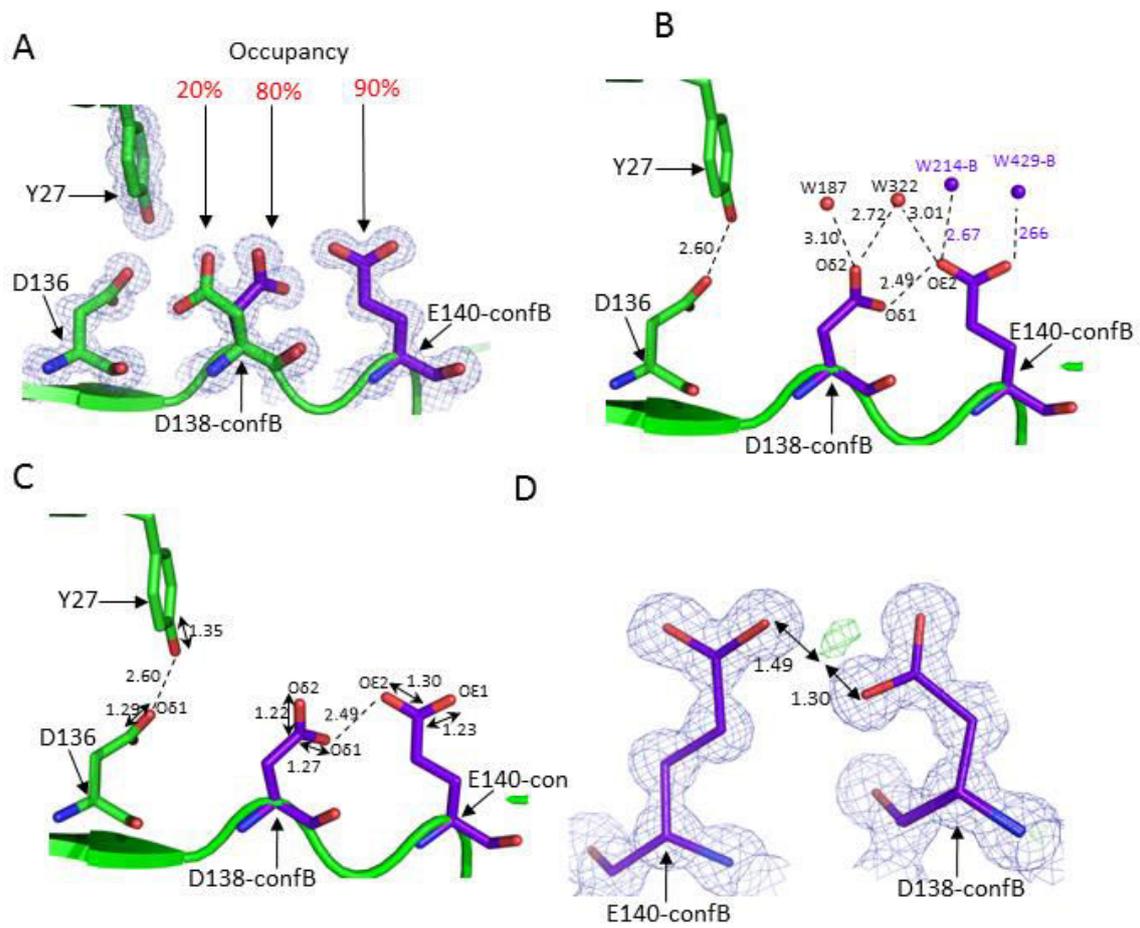
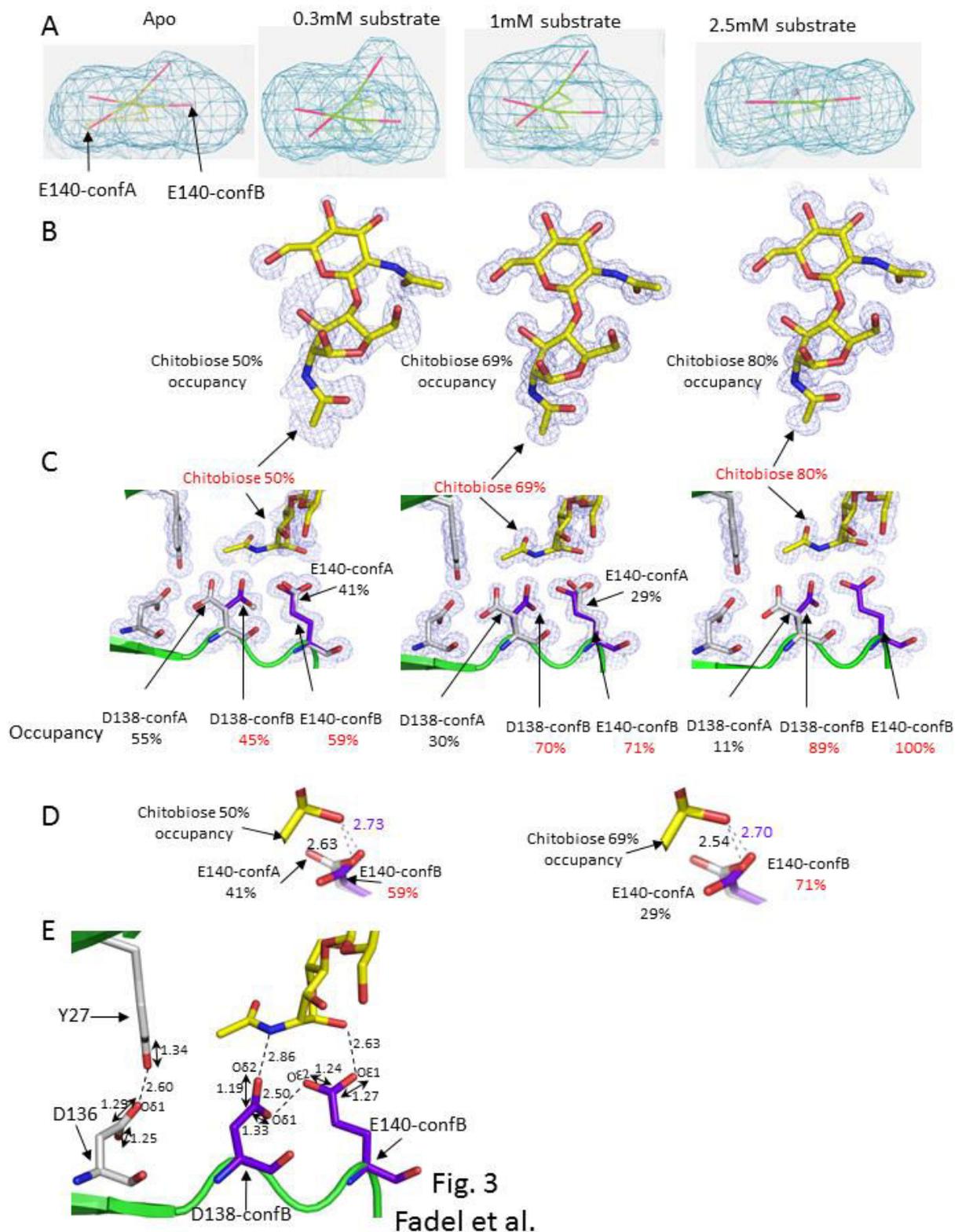


Fig. 2
Fadel et al.

Tableau 2 C—O bond length, standard error and pKa for the structure of CHIT1 in pseudo-apo form and at 2.5 mM substrate.

Residue	Bond length (in Å)	Error (in Å)	Bond length (in Å)	Error (in Å)	pKa	
					ConfA	ConfB
Pseudo-apo structure (0.95 Å)						
E140- confB	C-Oε1 1.2311	0.020	C-Oε2 1.3075	0.0192	8.13	6.55
D138- confB	C-Oδ1 1.218	0.0175	C-Oδ2 1.271	0.0175	13.07	12.34
D136	C-Oδ1 1.266	0.0126	C-Oδ2 1.2877	0.0125	4.64	6.02
Y27	C-OH 1.342	0.0127				
2.5 mM substrate (1.1 Å) CHIT1-chitobiose						
E140- confB	C-Oε1 1.2736	0.0200	C-Oε2 1.2397	0.0207	13.4	
D138- confB	C-Oδ1 1.3261	0.0208	C-Oδ2 1.1907	0.0179	7.6	
D136	C-Oδ1 1.2506	0.0167	C-Oδ2 1.2878	0.0170	6.12	
Y27	C-OH 1.347	0.0158				

Table. 2
Fadel et al.



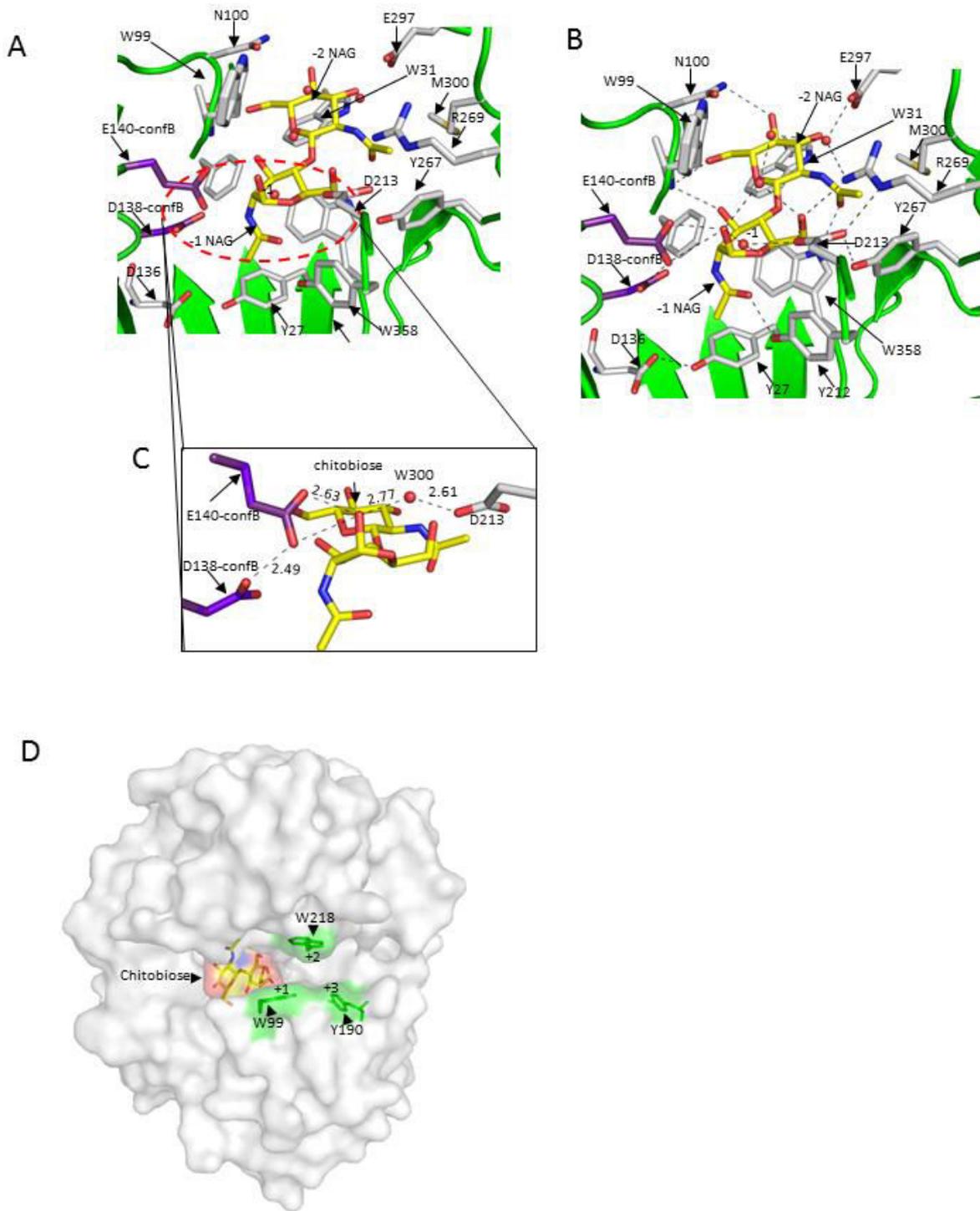


Fig. 4
Fadel et al.

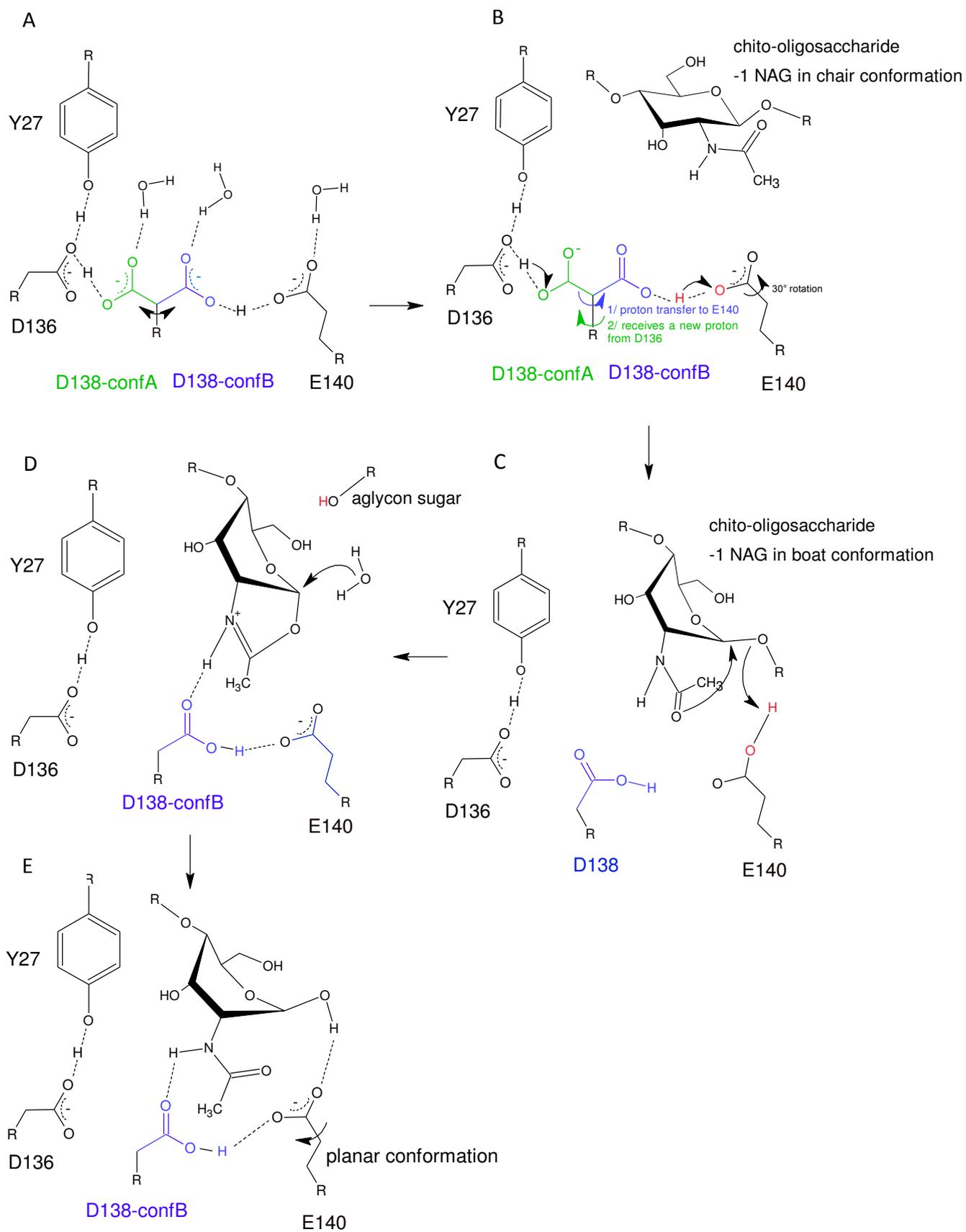


Fig. 5
Fadel et al.

X-ray crystal structure of the full-length human chitotriosidase (CHIT1) reveals features of its chitin-binding domain

Firas Fadel¹, Yuguang Zhao², Alexandra Cousido-Siah¹, André Mitschler¹ and Alberto Podjarny¹.

¹Department of Integrative Biology, IGBMC, CNRS UMR 7104, INSERM U 964, Université de Strasbourg, 1, rue Laurent Fries, Illkirch, France.

²Division of Structural Biology, University of Oxford, The Henry Wellcome Building for Genomic Medicine, Headington, Oxford OX3 7BN, UK.

Keywords: chitotriosidase, chitin binding domain, CBM14, seeding, X-ray crystal structure

Abstract

Chitinases are enzymes which catalyze the hydrolysis of chitin. Human chitotriosidase-1 (CHIT1) is one of the two active human chitinases, involved in innate immune response and highly expressed in a variety of diseases. CHIT1 is composed of a catalytic domain linked by a hinge to its chitin binding domain (ChBD_{CHIT1}). This last domain belongs to the CBM14 family and is thought to mediate the accessibility of the protein towards insoluble chitin. So far, only the crystal structures of the catalytic domains of both human chitinases CHIT1 and AMCase have been solved, resulting in a lack of structural data regarding ChBDs. Here, we report the crystallogenes strategy combining cross-seeding and several micro-seeding cycles for obtaining the first crystal structure of full length CHIT1 at 1.95 Å resolution. The structure of ChBD_{CHIT1} reveals a distorted β-sandwich showing the conservation of this 3D fold within the CBM14 family. All six highly conserved cysteine residues maintain the structural conformation of this domain by forming disulfide bridges.

The aromatic ring pattern of ChBD_{CHIT1} highlights the high conservation of their 3D positions and reveals that the binding interface contains the conserved Trp465 supposed to bind to sugar moieties. Furthermore, ChBD_{CHIT1} presents a positively charged surface. Finally, the high homology between the two human chitinases suggests that the AMC_{Case} ChBD will also adopt a distorted β -sandwich fold.

1. Introduction

Carbohydrate-protein interactions and carbohydrate-related catalysis has attracted significant attention due to their importance in numerous biological processes, such as cell-cell recognition, cell adhesion, and carbohydrate turnover. Recently, protein domains involved in such interactions have been classified into distinct carbohydrate-binding module families (CBMs). Indeed, currently there are 64 families of CBMs classified based on amino acid sequence similarity in the CAZy database (Cantarel *et al.*, 2009), but which can be structurally diverse (Malecki *et al.*, 2013). The CBMs are non-catalytic modules usually associated to carbohydrate-degrading enzymes (Bolam *et al.*, 1998); (Brunner *et al.*, 1998) and are thought to promote binding of insoluble carbohydrate polymers increasing the efficiency of the catalytic domain activity (Jee *et al.*, 2002); (Ikegami *et al.*, 2000); (Uchiyama *et al.*, 2001); (Boraston *et al.*, 2004); (Eijsink *et al.*, 2008); (Nimlos *et al.*, 2012a); (Hashimoto, Ikegami, *et al.*, 2000); (Watanabe *et al.*, 1994); (Horn, Sikorski, *et al.*, 2006); (Bolam *et al.*, 1998); (Brunner *et al.*, 1998).

One of the most widely present carbohydrates in nature is chitin, which consists of a linear β -1,4-linked polymer of N-acetylglucosamine (GlcNAc). Indeed, chitin is the second most abundant, natural and insoluble polysaccharide after cellulose and is a major component of fungal cell walls including those of fungal plants and human pathogens (Gooday, 1990). Chitinases [EC 3.2.1.14] are carbohydrate enzymes that catalyze the hydrolytic degradation of chitin. Chitinases belong to the group of glycosyl hydrolases (GH) and can be subdivided into two families (families GH18 and GH19) that differ in structure and mechanism. GH18 chitinases have been identified in a large number of organisms varying from lower organisms to humans. A large number of GH18 chitinases are multi-modular meaning that, in addition to their catalytic domains, they contain one or many extra

domain/s. These extra domains can be fibronectin type III-like domain (FnIII) or chitin binding domain (ChBD), the last mediating the interaction with insoluble chitin leading to a more efficient substrate hydrolysis (Horn, Sikorski, *et al.*, 2006); (Malecki *et al.*, 2013). ChBDs belong to different CBM families and are most commonly located at the C-terminus of the catalytic domain linked to it by a hinge region. Of these families, CBM5 and CBM12 are the most extensively studied ChBDs associated to a chitinase catalytic domain and are usually found in bacteria. CBM18 almost exclusively includes ChBDs from plants, with the exception of one CBM18 identified in *Streptomyces griseus* (Ohno *et al.*, 1996). On the other hand, CBM14 is commonly present in chitinases from baculoviridae, invertebrates, and mammals up to humans. CBM14 and CBM18 can be expressed solely as an individual module or linked to a chitinase catalytic domain. However, when they form part of a chitinase enzyme, CBM18 is only found fused to GH19 catalytic domain, while CBM14 is associated to GH18 chitinases of both invertebrates and mammals (Funkhouser & Aronson, 2007); (van den Burg *et al.*, 2004).

Chitotriosidase-1 (CHIT1) and acidic mammalian chitinase (AMCase) are the only active human chitinases. They are composed of a GH18 catalytic domain linked by a hinge to a CBM14 ChBD. The crystal structure of the catalytic domain alone, without their ChBDs, of both human chitinases has been solved (Fusetti *et al.*, 2002a); (Olland *et al.*, 2009a). CHIT1 is reported to be involved in the innate immune response against chitin-containing pathogens (Gordon-Thomson *et al.*, 2009) and is produced by macrophages and neutrophils (Lee, 2009); (Kzhyshkowska *et al.*, 2007); (Renkema *et al.*, 1997). CHIT1 can exist in two isoforms, the 39 kDa isoform which is found in lysosomes and the full length 50 kDa secreted isoform (CHIT1-FL). This isoform has been detected in Gaucher disease where the activity of CHIT1 increases between 10 – 1000 fold. Thus, it is a biomarker of Gaucher disease and a tool to monitor the therapy efficacy (Hollak *et al.*, 1994). Additionally, CHIT1 was found upregulated in patients with several disorders, including infections, chronic inflammation and degenerative disorders (Malaguarnera, 2006); (Kanneganti *et al.*, 2012). Although CHIT1 has been well-characterized as a clinical marker, its specific functions and effects under normal and pathological conditions remain not fully understood. Interestingly, a recent study has inferred the interaction of CHIT1 with glycan substrates found in association with

the surface of epithelial cells and macrophages (Larsen *et al.*, 2014). More recently, CHIT1 ChBD (ChBD_{CHIT1}) has been proposed to be involved in tumor metastasis of osteolytic lesions (Di Rosa *et al.*, 2014). Thus, it is of considerable interest to determine the structural properties of the CHIT1-FL in order to get new insights regarding its mode of action and in particular, to reveal the yet unknown structure of the CBM14 ChBD_{CHIT1}. Here, we report the first crystal structure of human CHIT1-FL at 1.95 Å resolution, determined with an adapted crystallogenesi approach combining cross-seeding and several micro-seeding cycles, including the first CBM14 domain solved by X-ray crystallography and being one of the few complete bi-modular chitinase structures available. Our structural and evolutionary analysis on the ChBD_{CHIT1} shows the importance of the conserved residues in maintaining the functionality of ChBD_{CHIT1} and reveals new features of this domain.

2. Materials and methods

2.1 Cloning, expression and purification

The human CHIT1-FL cDNA (GenBank: BC105682) of the 50 kDa CHIT1 isoform was used as a template to generate the C-terminal thrombin site (Thr-site) and His-tag by two polymerase chain reactions (PCR) using the following primers (SIGMA), 5'-AATTCAAGCTTGCCACCATGGTGC GGTCTGTGG-3' (N-terminal derived sense primer) and two antisense primers 5'-GTGATGGT GATGGTGGT GAGAACCGCGTGGCACCAGATTCCAGGTGCAGCA TTTG-3'; 5'-ATTATCGCGATACTAGTCTCGAGTCATTAGTGATGGT GATGGTGGT G-3'. The PCR product was cloned into the pHL expression vector (Aricescu *et al.*, 2006). The CHIT1-FL was transiently expressed in adherent HEK293T cells grown in roller bottles in the presence of the N-glycosylation inhibitor kifunensine (Chang *et al.*, 2007) as previously described (Aricescu *et al.*, 2006). After dialysis against 25 mM phosphate buffer saline (PBS) pH 8.0 at 4 °C, the secreted protein was purified from the media using an immobilized metal affinity chromatography (IMAC) batch procedure. CHIT1-FL was further purified by size exclusion chromatography on a Superdex 200 HR 16/60 (GE Healthcare) in 10 mM HEPES, 150 mM NaCl pH 7.5. The protein purity was assessed by SDS-PAGE (0.1 % SDS, 12.5% polyacrylamide (Laemmli, 1970) followed by Coomassie Brilliant Blue staining. The enzyme

concentration was determined from the absorption at 280 nm using an UV NanoDrop 1000 Spectrophotometer (Thermo Scientific). The molar extinction coefficient was calculated using the ProtParam tool on the ExPasy server (Gasteiger *et al.*, 2005) to be $83935 \text{ M}^{-1} \text{ cm}^{-1}$.

2.2 Crystallization, cross-seeding and micro-seeding

A Tecan Temo 96 head robot (Tecan) has been used to set sparse matrix screen containing commercially available crystallization reagents into the reservoirs of 96-well crystallization plates for sitting drops vapor diffusion method. The initial crystallization trials of CHIT1-FL were performed using Mosquito crystallization robot (TTP LabTech) to set sitting drops composed of $0.1 \mu\text{l}$ protein solution mixed with an equal volume of reservoir solution equilibrated against $40 \mu\text{L}$ of the reservoir solution. Although hundreds of crystallization conditions were tested, none of them have succeeded in inducing the crystallization of the CHIT1-FL. Next, we have tried cross-seeding using micro-crystals of the CHIT1 39 kDa catalytic domain (CHIT1-CAT) previously crystallized (Fusetti *et al.*, 2002a).

The CHIT1-CAT starter crystals were obtained at $17 \text{ }^\circ\text{C}$ by hanging drop technique through mixing equal volume of the well solution containing 25 % PEG, 0.2 M potassium sodium tartrate (PST) with the protein at $9 \text{ mg}\cdot\text{ml}^{-1}$ (in 0.01 M HEPES pH 7.0, 0.15 M NaCl), equilibrated against 0.5 ml of reservoir solution.

High quality CHIT1-CAT crystals were placed in an Eppendorf tube containing $100 \mu\text{l}$ of well solution and one Seed Bead (Hampton Research). Crystals in the eppendorf were then crushed and homogenized by vortexing for 2 min. The obtained micro-crystals of CHIT1-CAT were then used for automated high throughput cross-seeding screens using the Mosquito crystallization robot, as mentioned above. Each sitting drop consisted of $0.1 \mu\text{l}$ of the screening reservoir solution with $0.07 \mu\text{l}$ of the CHIT1-FL solution at $9 \text{ mg} \cdot \text{ml}^{-1}$ and $0.03 \mu\text{l}$ of the seeding stock. The drops were equilibrated against $40 \mu\text{L}$ of reservoir screen solution at $20 \text{ }^\circ\text{C}$. Nine thin and fragile crystals of CHIT1-FL have been obtained after this round of screening, but they were diffracting at a very low resolution. Nevertheless, CHIT1-FL crystals from the crystallization condition A (15 % PEG3350, 0.2 M sodium sulfate) have shown the best X-ray diffraction resolution in comparison with the eight other crystals and have been crushed for micro-seeding in the same way as indicated above. The new stock of

CHIT1-FL-Thr-His micro-crystals was used to set up manual hanging drop vapor diffusion in 24-well greased plates (Hampton Research). In each crystallization drop, 1.5 μL of the reservoir solution A were added to 1 μL of the CHIT1-FL solution ($9 \text{ mg} \cdot \text{ml}^{-1}$) with 0.5 μL of the second micro-seed CHIT1-Thr-His stock. Drops were equilibrated over 500 μL of reservoir solution at 17 °C and 24 °C. Crystals grew after one week, but their diffraction pattern was highly anisotropic. To stabilize the obtained crystals, many techniques have been applied including dehydration, cross-linking with glutaraldehyde, soaking with sodium malonate and annealing (Heras & Martin, 2005). All these methods have failed in decreasing the anisotropy of the crystals. Therefore, crystals grown by hanging drop in the condition A have been used for a third round of high-throughput micro-seed crystallization screen. After this screening round, new CHIT1-FL crystals have been obtained under different crystallization conditions. However, only the F6 condition from the Silver Bullets screen (Hampton Research) has shown good diffraction quality. Hence, a new micro-seeding solution stock has been prepared from crystals present within a drop of 0.4 μL , set up by the Mosquito robot after the third round screen and this new stock was used to optimize manual hanging drop vapor diffusion experiments of the new crystallization hit. The final improved F6 condition consisted of drops composed of 1 μL of the F6 condition of the Silver Bullets additive containing (0.2 % w/v 2-Methyl-2,4-pentanediol; 0.2 % w/v 1,2,3-Heptanetriol; 0.2 % w/v Diethylenetriaminepentakis (methylphosphonic acid); 0.2 % w/v D-Sorbitol; 0.2 % w/v Glycerol; Buffer 0.02 M HEPES sodium pH 6.8) added to 2 μL of the reservoir solution B (15 % PEG 3350, 0.02 M HEPES pH 6.8) with 1 μL of the CHIT1-FL solution ($8 \text{ mg} \cdot \text{ml}^{-1}$ in 0.01 M HEPES pH 7.5, 0.15 M NaCl) and 0.5 μL of the micro-seeding stock prepared from last round screen appeared in condition F6. The drops were equilibrated at 17 °C against 500 μL of the reservoir solution of solution B.

2.3. Cryo-cooling, data collection and molecular replacement

Crystals of CHIT1-FL grown in condition A were cryo-protected by sequential incubation for 30 seconds in two solutions containing increasing concentration of ethylene glycol (15 % and 25 %) in 20 % PEG 3350, 0.2 M of sodium sulfate pH 7.2, prior to flash-freezing in liquid nitrogen at 100 K.

Crystals of CHIT1-FL grown in optimized condition F6 were cryo-protected by sequential incubation for 30 seconds in two solutions containing increasing concentration of glycerol (15 % and 25 %) in 15 % PEG 3350, 0.02 M HEPES pH 6.8, the crystals were picked from the cryo-protectant solution into fiber loops (Hampton Research) and flash-frozen in liquid nitrogen at 100 K.

Crystals grown in different crystallization conditions were tested for diffraction at 100 K on our in-house Rigaku MicroMax 007 HF X-ray source equipped with an OSMIC Confocal optic VariMax HF and the Saturn 944 CCD detector.

Data sets were collected from crystals at the Swiss Light Source (SLS) synchrotron on the X06DA (PXIII) beamline. After the optimization of the procedure, 800 diffraction images were collected using a Pilatus 2M detector, up to a resolution of 1.95 Å, with an oscillation range of 0.25° and an exposure time of 0.3 s per frame, with a none attenuated beam of a 1.0 Å X-ray wavelength. All data sets were integrated, merged and scaled using the program HKL-2000 (Otwinowski & Minor, 1997) and XDS (Kabsch, 2010). The structure was solved by molecular replacement (MR) with Phaser (McCoy *et al.*, 2007) using the coordinates of the native structure of the same protein as an initial search model (Protein Data Bank (PDB) code 1GUV) (Fusetti *et al.*, 2002a). The model was improved by alternating cycles of manual model building using Coot (Emsley *et al.*, 2010); refined using REFMAC5 (Murshudov *et al.*, 2011) and PHENIX (Adams *et al.*, 2010). The stereochemical quality of the final model was checked with PROCHECK (Laskowski *et al.*, 1993). Structural figures were prepared using PyMOL (<http://www.pymol.org>). A summary of the data-collection and structure-refinement statistics is given in Table 1.

2.4 Electrospray Ionization Mass Spectrometry (ESI-MS)

Prior to ESI-MS analysis, CHIT1-FL was desalted on Zeba Spin Desalting Columns (Pierce) in 50 mM ammonium acetate. ESI-MS measurements were performed on an electrospray time-of-flight mass spectrometer (MicroTOF, BrukerDaltonic). Purity of the protein was verified by mass spectrometry in denaturing conditions (samples were diluted at 2 pmol· μL^{-1} in a 1:1 water–acetonitrile mixture (v/v) acidified with 1 % formic acid).

2.5 Structural Conservation Analysis

Homologous sequences to ChBD_{CHIT1} were obtained by BLAST (or PSI-BLAST) with an inclusion threshold of $e = 0.0001$ in UniRef90 (Altschul *et al.*, 2005); (Altschul *et al.*, 1997). Alignments of sequences were performed using MAFFT (Katoh & Standley, 2013). The amino acid sequences used are given in the supplementary data. The rate of evolution at each site is calculated using the empirical Bayesian (Mayrose *et al.*, 2004). Structural conservation analysis was performed using the ConSurf server (Ashkenazy *et al.*, 2010); (Glaser *et al.*, 2003).

3. Results and discussion

3.1. Crystallization of CHIT1-Full Length (CHIT1-FL)

Crystallization of the CHIT1-FL enzyme was a challenging task which might be explained by the high flexibility of the hinge region linking the catalytic domain to the chitin binding domain. Indeed, a large number of crystallization conditions turned out to be ineffective for achieving the crystallization of CHIT1-FL. Nevertheless, we found a strategy that promoted the crystallization of this enzyme by cross-seeding crystals from the CHIT1-CAT construct to induce the crystallization of the full length protein (Fig. 1). Crystals of CHIT1-FL construct obtained in the crystallization condition A were 700×100×70 μm in size. Despite the anisotropy of their diffraction pattern, datasets from these crystals were obtained at 2.6 Å resolution, in space group $P2_12_12_1$, with unit-cell parameters $a = 85.95$, $b = 108.30$, $c = 106.05$ Å, different from the space group of the seeding crystal CHIT1-CAT ($P2_12_12$). Surprisingly, the structure solved by MR showed two protein molecules in the asymmetric unit but displayed electron density corresponding only for the catalytic domain without any clear electron density for the hinge and the ChBD, (due probably to the flexibility of the hinge and the ChBD). This prompted us to launch a new crystallization screen by means of micro-seeding using crystals from the crystallization condition A. As detailed in materials and methods, we improved the obtained new crystals in the condition F6 of the Silver Bullets screen (Hampton Research), which resulted in good diffraction quality in terms of decreasing anisotropy and mosaicity. Data from the best crystal obtained with this condition were processed at 1.95 Å resolution in a new space group $P2_1$ with unit-cell

parameters $a = 51.12$, $b = 106.66$, $c = 85.66$ Å, $\beta = 107.11^\circ$. The new structure was then solved by MR with two protein molecules in the asymmetric unit (Fig. 1 and supplementary data Fig. 3)

Altogether, our results demonstrate that the cross-seeding enhances the crystallization of the full length protein. Indeed, the crystals obtained by cross-seeding were able to induce the growth of crystals with a different space group of lower symmetry, thereby improving the packing and the diffraction pattern. Our strategy consisted in applying several cycles of micro-seeding experiments combined to automated high-throughput crystallization conditions. The application of this approach, where each cycle improved the quality of the crystals, allowed us to generate good quality crystals diffracting at 1.95 Å resolution and thus to solve the so far elusive CHIT1-FL structure.

3.2. Analysis of the crystal contacts and packing

The first X-ray structure of the full length human CHIT1 form was solved by MR to a resolution of 1.95 Å, using the catalytic domain of CHIT1 as a search model. Data collection and processing statistics are presented in Table 1. The final refined model has R_{work} and R_{free} values 20.51 and 24.54 % respectively. There are two protein molecules in the asymmetric unit named chain A and B (Fig. 2A). Consequently, four domains were built, two catalytic domains and two ChBD corresponding to the regions from Ala22-Leu386 and Asn417-Asn466 respectively (Fig. 2B). As the extension consisting of the Thr-site with the His-tag was not cleaved, the residues belonging to the Thr-site were modeled at the C-terminus of ChBD from chains A and B while the electron density corresponding to four histidines from the His-tag was observed only on chain A. These histidine residues are stabilized by the crystal packing as they were interacting with the binding site of a neighboring molecule in the unit cell. Although both ChBD in the asymmetric unit are similar with a rmsd value of 0.24 Å, the conformation of the Thr-site extension is different in each ChBD of the asymmetric units which allows us to distinguish them (Fig. 2B, C). The hinge region comprising 31 amino acids which links the catalytic domain and ChBD was not built due to a lack of interpretable electron density, presumably because it is a highly mobile and disordered region. Our data show an asymmetric unit consisting of two monomeric molecules arranged as four

separated and disconnected domains including two differently oriented catalytic domains and two ChBDs, each displaying a distinct Thr-site conformation (Fig. 2B, C, supplementary data Fig. 3). Due to the lack of electron density for the hinge region, we thus wondered which ChBD should be paired to each catalytic domain to form two CHIT1-FL monomers. In order to obtain such information, we have analyzed the different possibilities in the crystal packing. In one of the two independent CHIT1-FL molecules within the asymmetric unit, the ChBD in chain A is located close to the non-reducing side of the catalytic domain while the ChBD in chain B appears below the catalytic domain (Fig. 2A, B, C). The distance between the termini of both domains (catalytic domain Leu386-Asn416 ChBD) is ~ 30 Å in both chains A and B. In this configuration, the ChBD does not contact the catalytic domain from the same chain. However, the ChBDs from both monomers interact with each other forming a homodimer with a buried surface of 1063.4 Å² (~ 6 % of the monomer surface area) (Fig. 2A). The calculated free energy change (ΔG) of the interdomain contact (ΔG^{int}) calculated by PISA (<http://www.ebi.ac.uk/msd-srv/pisa/cgi-bin/piserver?qa=4lyz>) gave a value of -12.2 kcal mol⁻¹ (see supplementary data fig. 4). The comparison of this ΔG value with the other possible crystallographic interfaces and the different domain arrangements in the asymmetric unit has shown that the configuration described above in figure 2A is energetically favored. Indeed, the second next possible CHIT1-FL domain configuration shows two pseudo anti-parallel monomers with 775.5 Å² of buried surface with a ΔG^{int} value of -1.7 kcal mol⁻¹, rendering it energetically unfavorable in comparison to the previous configuration.

Interestingly, our data seems to indicate that the appearance of the electron density corresponding to the ChBDs has occurred thanks to interactions of these two domains with each other i.e. because of the packing. Since the size-exclusion chromatography and Mass Spectrometry (MS) data confirm that the CHIT1-FL is monomeric in solution (supplementary data), the dimer formation was brought about by the crystal packing of the space group $P2_1$ which resulted in the stabilization of the ChBD_{CHIT1} monomers. This result suggests that a stable dimer may not have been formed in the crystal packing with the space group $P2_12_12_1$ which could explain the reason why the electron density associated to ChBDs was not detected.

As mentioned previously, the position of the ChBD relative to the catalytic domain is not the same in chains A and B. The different positions of the ChBDs in the asymmetric unit could be explained by the high mobility of the hinge region that allows the ChBD to move randomly affecting in turn the orientation of the catalytic domain.

Our crystal packing analysis allows us to suggest that the ChBD_{CHIT1} is characterized by a high mobility and is not situated in a precise direction with regards to the corresponding catalytic domain to which it is linked. This high mobility was also reflected by significant high B-factors for the ChBD (from 26 Å² to ~80 Å²) in comparison to the B-factors of the catalytic domain residues (Fig. 3A). This is a main difference when compared to other crystal structures of full length bacterial chitinases such as ChiA and ChiB (fig. 3B) from *S. marcescens* which display the carbohydrate binding domain situated in a clearly defined direction relative to the catalytic domain. For example ChiB has a C-terminal ChBD similar to CHIT1, however the hinge region in ChBD_{ChiB} is not flexible and show an average B-factor of 24.8 Å² (van Aalten *et al.*, 2000). The low flexibility of the hinge in ChiB results in that its ChBD is located towards the C-terminus (Vaaje-Kolstad *et al.*, 2013). On the other hand, ChiA has an N-terminal carbohydrate binding domain extending to the substrate binding cleft at the non-reducing side (Vaaje-Kolstad *et al.*, 2013). This was suggested to affect how the two enzymes degrade the chitin polymer from different ends: ChiA acts from the reducing end while ChiB does it from the non-reducing end. Although the ChBD_{CHIT1} is C-terminal to the catalytic domain, it hydrolyzes from the reducing end in contrast to ChiB. This could be explained by the high flexibility of the hinge in CHIT1-FL and the open groove architecture of its binding site. Moreover, the high flexibility of the hinge in CHIT1-FL suggests that the ChBD_{CHIT1} is not aligned with the catalytic domain but that it is moving in different directions in such a way that looks like a probe inspecting the environment for the presence of substrate. Based on this, we hypothesized the role played by ChBD_{CHIT1} in the step-wise mechanism of CHIT1-FL action. Once the ChBD_{CHIT1} locates the presence of chitin in the environment, it binds to it and then drives the catalytic domain to the substrate location. Once bound to chitin, the ChBD_{CHIT1} disrupts its crystalline structure making it accessible to be hydrolyzed by the catalytic domain.

3.3 Overall structure of ChBD_{CHIT1}

We hereby describe the crystal structure of ChBD_{CHIT1} comprising 49 residues (417-466). In fact, functional analysis defined the minimal sequence required for chitin binding activity in CHIT1 corresponding to the most C-terminal 49 amino acids of the protein (Tjoelker *et al.*, 2000b), which agrees with the ChBD_{CHIT1} in our CHIT1-FL structure. The structure of the catalytic domain, which adopts the conserved (α/β)₈ TIM barrel fold found in all GH18 family, is essentially the same as the CHIT1-CAT already described by (Fusetti *et al.*, 2002a). The structure of ChBD_{CHIT1} reveals an elongated conformation (dimensions 60 x 17 x 14 Å), which is different from the globular and compact conformation of ChBD from bacteria and plants belonging to CBM5/12 and CBM18 respectively (Ikegami *et al.*, 2000); (Akagi *et al.*, 2006). ChBD_{CHIT1} fold consists of a distorted β -sandwich composed of two β -sheets containing three N-terminal anti-parallel β -strands (β 1, β 2, β 3; residues 427-428, 436-440, 445-449) and two C-terminal anti-parallel β -strands (β 4, β 5; 455-457, 460-464) (Fig. 4A, B). By sequence similarity, ChBD_{CHIT1} has been attributed to the CBM14 family which also exists in invertebrates and in particular, in insects and nematodes (Bussink *et al.*, 2007). So far the only known structure of a CBM14 was solved by NMR and corresponds to the ChBD polypeptide tachycitin (73 residues), found in horseshoe crab (*Tachypleus tridentatus*) (Fig. 4C) (Suetake *et al.*, 2000). Tachycitin is a ChBD, produced as a single domain, not linked to any chitinase enzyme, yet it shares 34 % of the sequence identity with ChBD_{CHIT1} which is fused to a catalytic domain. Importantly, structural alignment between the human ChBD_{CHIT1} and the NMR structure of tachycitin reveals that they share the same distorted β -sandwich fold (Fig. 4C, D). Indeed, after superimposition of both domains, we observe a rmsd deviation in α -carbon positions of 1.27 Å between ChBD_{CHIT1} and tachycitin (Fig. 4D). On the other hand, the ChBD_{CHIT1} lacks the C-terminal α -helical turn (Fig. 4C, D) as it is shorter than tachycitin (73 residues). We can therefore conclude that the most prominent feature that can be deduced from the structural alignment between tachycitin and ChBD_{CHIT1} is that the distorted β -sandwich fold is a structural characteristic of CBM14, which is conserved from invertebrates to vertebrates.

Moreover, structural analysis of the backbone conformations of CBM14 ChBD_{CHIT1} and ChBD from CBM5/12 found in bacterial chitinase structures shows the presence of 5 β -

strands as main secondary structure components (Fig. 5A, B). Nonetheless, the main difference between these two CBM families is that in the case of CBM5-12, the β -sheet containing two β -strands is composed by the first and fifth β -strand, while in CBM14 it is formed by two consecutive β -strands (Fig. 4B and Fig 5A, B). This gives the CBM5/12 a compact and globular conformation, while the separated β -sheets in CBM14 exhibit an extended conformation (Fig. 4B and Fig 5A, B). Our observations indicate that ChBD_{CHIT1} possesses separated β -sheets which in turn make the domain adopt an overall extended conformation. We hypothesize that this structural feature may be functionally relevant as a way for the ChBD_{CHIT1} to scan for the presence of substrate in the protein environment.

3.4 Evolutionary analysis of conserved cysteine residues in ChBDs

As it has been reported before, ChBD_{CHIT1} contains 6 highly conserved cysteine residues (Tjoelker *et al.*, 2000b). Indeed, our electron density map confirms that they form three disulfide bonds. The one between Cys420 and Cys440 connects the β -strand 2 (β 2) to the beginning of the first loop (L1). The second disulfide bond is in the C-terminal region of ChBD_{CHIT1} (Cys450-Cys463) and links the last β 5 with the L4 (Fig. 4B). These two disulfide bonds exist in equivalent locations in tachycitin suggesting that the conservation of these disulfide bridges contributes to the structural conservation of the ChBD global folding within the CBM14 family, both in invertebrates and mammals (Fig. 4B, C, D). The remaining disulfide bond is established between two cysteine residues (Cys460-Cys462) linking the hairpin to the β 5 (Fig. 4B). Although these two latter cysteines do not exist in tachycitin, the evolutionary structural conservation analysis performed by the Consurf server over 150 chitinases and chitinase-like proteins from different species in invertebrates and vertebrates demonstrates that all the six cysteine residues from ChBD_{CHIT1} are fully conserved (Fig. 6A and supplementary data fig. 2). The non-conservation of the equivalent Cys460-Cys462 in tachycitin could be explained by the fact that tachycitin is an individual domain which is not part of a chitinase enzyme. However the full conservation of ChBD_{CHIT1} cysteine residues is observed when we studied the sequence-conservation pattern of ChBDs linked to chitinases. Interestingly, site-directed mutagenesis experiments in CHIT1 have shown that each of these cysteine residues is critical for the binding activity to chitin (Tjoelker *et al.*, 2000b). In fact,

this work underlines the indispensability of each of the six cysteines and point out their implication disulfide bond formation our structure support it. Combining these evidences with our finding regarding the high conservation of the six cysteine residues, we hypothesize that their evolutionary conserved role is to maintain the integrity of the ChBD in a functional folded conformation.

3.5 Analysis of the evolutionary conserved aromatic rings

It is believed that the interaction of the carbohydrate crystalline substrates i.e. chitin and cellulose with their respective binding domains (ChBDs and cellulose binding domains (CBDs)) occurs via exposed aromatic residues (Asensio et al., 2000); (Akagi et al., 2006). Interestingly, although ChBDCHIT1 is a small module, it contains 7 conserved aromatic residues among which, 6 of them are solvent exposed (Fig. 6B). This gives a hydrophobic character to the ChBDCHIT1 surface which could result in an increased affinity for the crystalline chitin within the fungal cell wall (Tjoelker et al., 2000b). In fact, within the core of the domain where the three anti-parallel β -strands are located, there are 4 exposed aromatic residues. Two of them are phenylalanine residues (Phe446 and Phe437) (Fig. 6B, C), located on β 2 and β 3 respectively and pointing to the same face of the domain. The other two are two tyrosine residues (Tyr428 and Tyr438) pointing to the opposite face of ChBDCHIT1 (Fig. 6B, C). Remarkably, in the region where these two tyrosine residues are positioned, there are two proline residues (Pro429 and Pro431) facing each other which makes this side “canal-like” and rich in aromatic rings (Fig. 6C). This observation leads us to propose that such surface may play a role in assisting ChBDCHIT1 to bind chitinous substrate via stacking interactions.

Moreover, the ChBDCHIT1 is surrounded by two additional aromatic residues (Phe419 and Trp465) from two sides of the domain, on the L1 and L5, respectively (Fig. 6B, C). The position of many aromatic residues in ChBDCHIT1 is similar to the only known structure of a ChBD from the CBM14 family, the tachycitin. This means that the aromatic rings in the CBM14 family are distributed on the entire molecular surface of the domain and not concentrated on one side as is the case of plants ChBDs belonging to CBM18 such as hevein (Andersen *et al.*, 1993) (Fig. 5C) and OsChia1b from rice chitinase (Kezuka *et al.*,

2010). Moreover, the sequence alignment of ChBDs shows that the positions of the aromatic rings and their distribution patterns are evolutionary conserved between vertebrates and invertebrates (Fig. 6A) (supplementary data Fig.2). The strong conservation of the position of aromatic residues underlines the importance of their evolutionary conserved biological role in mediating the access towards chitin within the fungal cell wall.

The binding interfaces of several ChBDs belonging to different CBM families (CBM5/12, CBM14, and CBM18) interacting with sugar moieties of crystalline chitin have been determined by site-directed mutagenesis studies (Uni *et al.*, 2012); (Tjoelker *et al.*, 2000a); (Katouno *et al.*, 2004) combined with sequence and structural alignment analysis as well as NMR oligosaccharide titration (Asensio *et al.*, 2000); (Akagi *et al.*, 2006) and computational studies (Kezuka *et al.*, 2010). This interface region was found to be the same not only in ChBDs but also in CBDs, even though the reason why ChBDs display specificity only toward crystalline chitin and not toward cellulose remains a question to answer. It is worth noting that the presumed interface region in the different CBMs contains often two or three aromatic rings which directly stack to sugar moieties in the crystalline carbohydrate (Fig. 5A, B). In fact, by making analogous observations and analyzing the different published structures of CBMs from different species (Suetake *et al.*, 2000); (van den Burg *et al.*, 2004); (Akagi *et al.*, 2006), we can propose that the binding region interface of ChBD_{CHIT1} corresponds to the side domain comprising the C-terminal portion which contains the Trp465 (Fig. 4D, 6B, C). Indeed, Tjoelker *et al.* (Tjoelker *et al.*, 2000a) have reported that the construct of ChBD_{CHIT1} lacking Trp465 and Asn466 completely loses the binding activity toward chitin. In another study on the ChBD from *Bacillus circulans* chitinase A1 (ChiA1), mutagenesis experiments have demonstrated that a single tryptophan present in the interacting interface region can pivot to interact with the chitin (Ferrandon *et al.*, 2003). Thus, taken together, all available evidence indicates that Trp465 is a key aromatic residue for the binding activity to chitin. Accordingly, although the binding region of ChBD_{CHIT1} is characterized by the presence of only one tryptophan, other ChBDs have at least two aromatic residues in this region. Nonetheless, ChBD_{CHIT1} displays a high affinity to crystalline chitin as it has been demonstrated by *in vitro* experiments (Vandevenne *et al.*, 2011). Interestingly, the Trp465 is a highly conserved aromatic residue across vertebrates and

invertebrates (supplementary data Fig. 2), which adopts a planar solvent exposed conformation closely similar to the planar conformation detected in the binding interface of many known structures of CBMs (Akagi *et al.*, 2006); (Suetake *et al.*, 2000); (Malecki *et al.*, 2013); (Andersen *et al.*, 1993); (van den Burg *et al.*, 2004) (Fig. 5 A,B and Fig 6B, C). From an evolutionary point of view, this suggests that the human genome has conserved the minimum number of aromatic residues that allows an efficient binding function to chitin which makes the CHIT1 able to attack fungal invaders by having access to their chitin cell wall. On the other hand, the structure of ChBD from *Bacillus circulans* chitinase A1 (ChBD_{ChiA1}) corresponds to the only structure of a ChBD where the tryptophan in the binding interface does not adopt a solvent exposed planar conformation (Fig. 5B). Additionally, it has been demonstrated that the ChBD_{ChiA1} affinity towards chitin is lower than the affinity of ChBD_{CHIT1} (Vandevenne *et al.*, 2011). Thus, our analysis combined with available evidence suggests that the solvent exposed planar conformation of tryptophans on the binding side of the ChBD of different families is a key structural feature affecting the affinity towards the crystalline carbohydrate sugar.

3.6 ChBD_{CHIT1} presents a positively charged face

A recent study has demonstrated the secretion of CHIT1-FL by osteoclasts to the extracellular matrix. In this study, it has been put forward that CHIT1 is involved in bone matrix digestion leading to bone resorption. Importantly, ChBD_{CHIT1} has been proposed to bind to the calcium phosphate surface, a main component of bone surface, thereby mediating the CHIT1 function during the digestion of calcium phosphate substrate (Di Rosa *et al.*, 2014). To gain insight into such putative interaction we have investigated the electrostatic surface properties of ChBD_{CHIT1}. Remarkably, the ChBD_{CHIT1} domain has essentially two different charged faces. One face is mostly neutral with a negatively charged spot which spans two residues (Ser439, Asp425), while the second face is highly positively charged due mainly to four residues (Arg444, Arg434, Lys423 and Lys461) (Fig. 6D). Since phosphate ions are negatively charged components of bone minerals, it could be hypothesized that these negatively charged phosphate ions have electrostatic affinity to the

positively charged face of ChBD_{CHIT1}, thereby facilitating the access of the enzyme and the digestion of the bone matrix.

3.7 High similarity between ChBD_{CHIT1} and ChBD_{AMCase}

The structures of CHIT1 and AMCase catalytic domain have revealed that both adopt the same $(\alpha/\beta)_8$ TIM barrel fold showing a high similarity in their active sites (Olland *et al.*, 2009a); (Fusetti *et al.*, 2002a). However, the structure of ChBD_{AMCase} has not been solved yet; we therefore wondered if our X-ray crystal structure of ChBD_{CHIT1} could give insight into the 3D folding and structural characteristics of ChBD_{AMCase}. Indeed, sequence alignment of ChBD belonging to CHIT1 and AMCase reveals that they are characterized by a high sequence similarity of 63 % and sequence identity of 49 % (23 out of 49 residues are identical) (Fig. 7). Remarkably, cysteine residues and aromatic rings positions are completely conserved in both domains. Furthermore, in addition to the fact that many residues located on the same β -strand in CHIT1 are conserved in ChBD_{AMCase}, residues of β_4 of ChBD_{CHIT1} are identical in ChBD_{AMCase}. Accordingly, as both human ChBDs show a high sequence similarity and given that they belong to CBM14 characterized by the conserved distorted β -sandwich, we can therefore propose that both human ChBDs share the same 3D folding and the same surface aromatic character.

It is worth pointing out that recent data regarding human chitinases are highlighting increasingly their interaction with non-chitinous components and suggest possible interactions in processes with extracellular receptors and extracellular matrix (ECM) components. For example, Hartl *et al.* have demonstrated a physical interaction between AMCase and the epidermal growth factor receptor (EGFR) which leads to a stimulation of CCL2, CCL17, and CXCL8 chemokines production (Hartl *et al.*, 2008). On the other hand, recent *in vitro* data have proposed that glycans on the epithelial and macrophage cell surface could be hydrolyzed by CHIT1 contributing to the pathogenic response (Larsen *et al.*, 2014). In addition, it has been reported that AMCase has an anti-apoptotic function not mediated by its catalytic activity (Hartl *et al.*, 2009). Although the current experimental data are limited regarding such hypothesis, all these studies underline the importance of full length human chitinases in several biological processes. Indeed, evidence points out to the

fact that chitinase functions do not necessarily rely exclusively on the catalytic activity but rather that other regions such as ChBDs could also contribute to the protein action and be affected under pathological conditions. Up to our study, only the crystal structures of the catalytic domain alone, without their ChBDs, of both human chitinases had been solved. Here we have reported the 3D crystal structure of the full length CHIT1 and performed a detailed structural analysis of ChBD_{CHIT1}. In this context, our structural investigation on ChBD_{CHIT1} reveals that ChBD_{CHIT1} and ChBD_{AMCase} share significant structural properties suggesting a comparative physiological mode of action of these two domains.

4. Conclusion

In this study, we have reported an original crystallogenesi approach for obtaining for the first time the full length structure of a mammalian chitinase, CHIT1. This strategy included many cycles of cross-seeding and micro-seeding which have allowed us to solve the full length structure of CHIT1 and the first structure of a CBM14 ChBD associated to a chitinase. The lack of electron density corresponding to the hinge region linking the catalytic domain to the ChBD prompts us to suggest a high mobility of this region resulting in a high mobility of the entire ChBD. Thus, thanks to the adopted crystallogenesi strategy, we have succeeded in overcoming the flexibility issue of the hinge which impeded the direct observation of the ChBD electron density. The structure of ChBD_{CHIT1} reveals a distorted β -sandwich fold which appears to be conserved within the CBM14 family across invertebrates and humans. In these ChDB modules, the highly conserved cysteine residues seem to have an essential role in maintaining the functional conformation of the domain. The investigation of the aromatic ring pattern of ChBD_{CHIT1} highlights their high conservation and reveals that the binding interface contains a conserved aromatic residue (Trp465) adopting a solvent planar conformation convenient for efficient binding to sugar moieties. Furthermore, the ChBD_{CHIT1} presents a positively charged surface which could be responsible for the binding to negatively charged phosphate ion of the bone matrix, facilitating its digestion during bone resorption. Finally, we have shed light onto the high similarity between ChBD_{CHIT1} and ChBD_{AMCase} suggesting that this latter could also adopt a distorted β -sandwich fold similar to ChBD_{CHIT1} and CBM14. Indeed, we believe that our developed crystallogenesi methodology

could be used for co-crystallization or soaking experiments with different ChBD substrates or for solving the 3D structure of AMCcase-FL. All in all, our results have revealed many structural aspects of human ChBDs which give useful insights onto their properties/features.

Acknowledgements

We thank the IGBMC Structural Biology and Genomics Platform staff for technical assistance, in particular Pierre Poussin-Courmontagne and Alastair Mc Ewen. We also thank the IGBMC proteomics platform for mass spectrometry analyses, in particular Frank Ruffenach and Adeline Page. We are grateful to Dr. Bruno Klaholz, Dr. Alex Rozov and Dr. Francesc Xavier Ruiz Figuras for helpful discussions about X-ray data processing. The crystallographic experiments were performed on the X06DA (PXIII) beamline at the Swiss Light Source synchrotron, Paul Scherrer Institut, Villigen, Switzerland. We thank in particular, Ezequiel Panepucci and Vincent Olieric for their help on data collection. This work has been funded by the CNRS, the INSERM, the Université de Strasbourg, Biostruct-X (FP7, contract 283570) Instruct (European Strategy Forum of Research Infrastructures; ESFRI) and the French Infrastructure for Integrated Structural Biology (FRISBI).

Table 1. Data collection, processing and refinement statistics.

	CHIT1-FL
PDB code	
Synchrotron, beamline	SLS, X06DA (PXIII)
Wavelength (Å)	1.0
Resolution range (Å)	44.69 - 1.95 (2.013 - 1.95)
Space group	P 1 2 ₁ 1
Unit cell (Å)	a= 51.14 b= 106.66 c= 85.67 α=γ= 90 β= 107.13
Total reflections	242490 (22495)
Unique reflections	62392 (5738)
Multiplicity	3.9 (3.9)
Completeness (%)	96.63 (89.40)
Mean I/sigma(I)	15.06 (1.85)
Wilson B-factor	29.66
R-sym	0.057 (0.63)
R-meas	0.077 (85.3)
CC(1/2)	99.8 (68.6)

R-factor	0.2051 (0.3605)
R-free	0.2454 (0.4082)
Number of atoms	6951
macromolecules	6750
ligands	24
water	177
Protein residues	849
RMS (bonds)	0.007
RMS (angles)	1.10
Ramachandran favored (%)	97
Ramachandran outliers (%)	0
Clashscore	7.79
Average B-factor	31.60
macromolecules	31.60
ligands	33.60
solvent	29.60

Statistics for the highest-resolution shell are shown in parentheses.

Figure legends

Figure 1. Crystallogenesi s strategy for crystallization of CHIT1-FL. (A) crystals from the catalytic domain of CHIT1 crushed and used for initial automated cross-seeding. (B) Crystals of CHIT1-FL obtained after the cross-seeding round in the first hit crystallization condition A. (C) Crystals of CHIT1-FL obtained after optimization through manual hanging drop. (D) Crystals of CHIT1-FL from condition A crushed and used for another cycle of automated micro-seeding leading to crystallization condition F6. (E) (F) Crystals of CHIT1-FL obtained after optimizing F6 condition.

Figure 2. 3D structure of CHIT1-FL. (A) Left, in unit cell, surface representation of two CHIT1-FL molecules (chain A and B). The position of the active site of each monomer shows that they are not in the same direction. Right, in crystal packing, surface representation of 4 CHIT1-FL monomers. (B) Surface and ribbon representation of chain A comprising the catalytic domain and ChBD in two views. Thr-site (yellow), His-tag (light violet) and hinge region (dotted line). (C) Surface and ribbon representation of chain B comprising the catalytic domain and ChBD in two views. Thr-site (yellow), hinge region (dotted line).

Figure 3. Representation of the thermal parameter distribution shown as B-factor 'putty' as implemented in PyMOL (<http://www.pymol.org>). A) CHIT1-FL with a zoom on the ChBD_{CHIT1}. B) ChiB from *Serratia marcescens* with a zoom on the hinge and the ChBD_{ChiB}. The

α -atom B factors are depicted on the structure in dark blue (lowest B factor) through to red (highest Bfactor), with the radius of the ribbon increasing from low to high B factor.

Figure 4. 3D structure of ChBD_{CHIT1} and structural comparison with tachycitin (A) Solvent accessible surface of ChBD_{CHIT1} comprising its backbone represented as violet ribbons is shown in two orientations. (B) Ribbon diagram of ChBD_{CHIT1} colored as in A displaying the distorted β -sheet sandwich. The six cysteines forming disulfide bonds are shown as sticks. The sulfur atoms of the 6 cysteines are labelled in yellow. (C) Ribbon representation of the tachycitin 3D structure illustrates a distorted β -sheet sandwich fold and its C-terminal α -helical turn. Trp55 (red sticks). (D) Superimposition of ChBD_{CHIT1} (light violet) and tachycitin (green) backbones. The conserved disulfide bonds in both structures are marked as red sticks. Conserved tryptophans are indicated as sticks, in green for CHIT1 and in red for tachycitin.

Figure 5. 3D structures of bacterial and plant ChBDs. (A) Left, 3D structure of ChBD in *Streptomyces griseus* HUT6037 chitinase C (ChBD_{Chic}) and right, 3D structure of ChBD in chitinase A1 from *Bacillus circulans* WL-12 (ChBD_{ChiA1}). (B) Superimposition of the backbone from ChBD_{ChiA1} and *Streptomyces griseus* HUT6037 chitinase C (ChBD_{Chic}). Relevant tryptophans are represented as lines. Trp687 and Trp696 belong to ChBD_{ChiA1} in black and Trp30 and Trp31 belong to ChBD_{Chic} in blue. C) 3D structure of ChBD in hevein and aromatic residue are represented as lines in blue.

Figure 6. Evolutionary and structural features of ChBD_{CHIT1} (A) Sequence conservation in ChBD_{CHIT1} is represented as surface, ribbon and line in 2 views. Color-codes depend on the residue conservation degree (conserved, magenta to variable, cyan). Relevant conserved residues are indicated with arrows. (B) ChBD_{CHIT1} backbone is presented in light violet ribbon. Aromatic residues are labeled and represented as green sticks. (C) Representation of the hydrophobic potential surfaces of ChBD_{CHIT1} in two orientations. Color-codes depend on the hydrophobic potential (from blue to yellow with increasing hydrophobicity). Hydrophobic residues are indicated with arrows. (D) Representation of the electrostatic potential at the surface of ChBD_{CHIT1} in two orientations. The protein is shown as solvent-accessible surface colored by electrostatic potential at ± 5 kT/e. Color-codes depend on the electrostatic potential (red: negative charge; blue, positive charge; and white: neutral charge)

Figure 7. Amino acid sequence alignment of the ChBD in both CHIT1 and AMCase generated by Clustal Omega. The residues are colored according to their physico-chemical properties using Clustal color code. Stars point out conserved residues. Beta sheets(green) and loops (blue).

Supplementary figures

Supplementary figure 1.

(A) Domain organization of CHIT1-FL. The signal peptide is colored in green. The catalytic domain which belongs to family 18 glycosyl hydrolase is colored in yellow. The

active site is indicated in red. The ChBD_{CHIT1} colored in blue is linked by a hinge region to the catalytic domain. (B) Western-blot analysis by anti-His antibody of four days conditioned media confirms the expression and secretion of CHIT1-FL at the predicted molecular weight. (C) On the left, Chromatogram showing the elution peak during purification of the CHIT1-FL by size-exclusion chromatography. On the right, CHIT1-FL after migration on SDS-PAGE and stained by Coomassie brilliant blue. (D) Negative-ion mode ESI-MS spectrum of the native CHIT1-FL. The negative ion peaks with m/z ratios of 50800 Da correlate with the monomer form of CHIT1-FL which has a molecular weight of 51051.3 Da.

Supplementary figure 2.

Sequence alignment and conservation of ChBD homologues of ChBD_{CHIT1}. Color-codes depend on the residue conservation degree (conserved, magenta to variable, cyan).

Supplementary figure 3.

Model of CHIT1-FL and electronic density maps (1σ cutoff) for the asymmetric unit ($2F_o-F_c$ map – grey, F_o-F_c map – green).

Supplementary figure 4.

Data after submitting the structure coordiante in PDBe server PISA (Protein Interfaces, Surfaces and Assemblies).

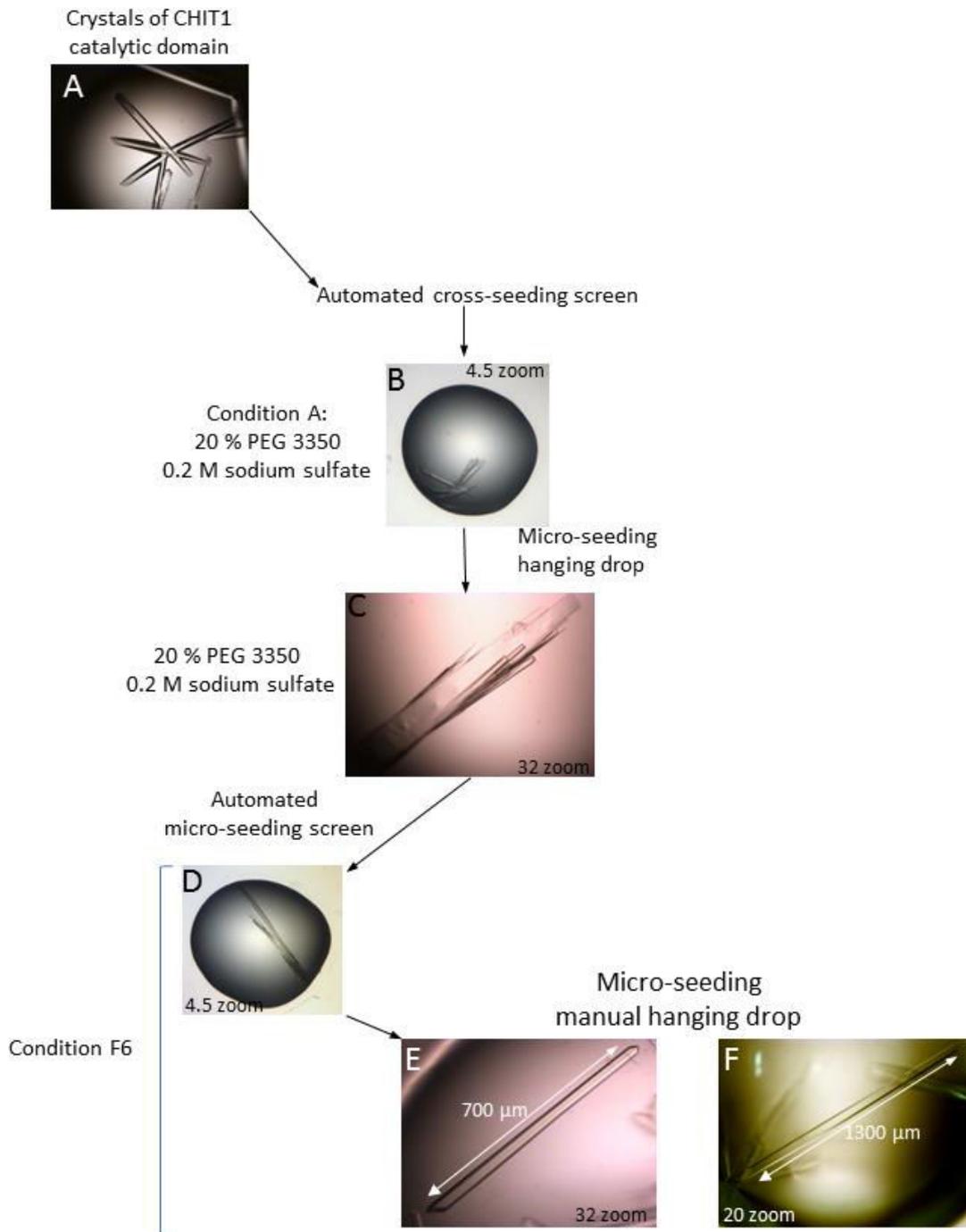


Figure 1
Fadel et al.

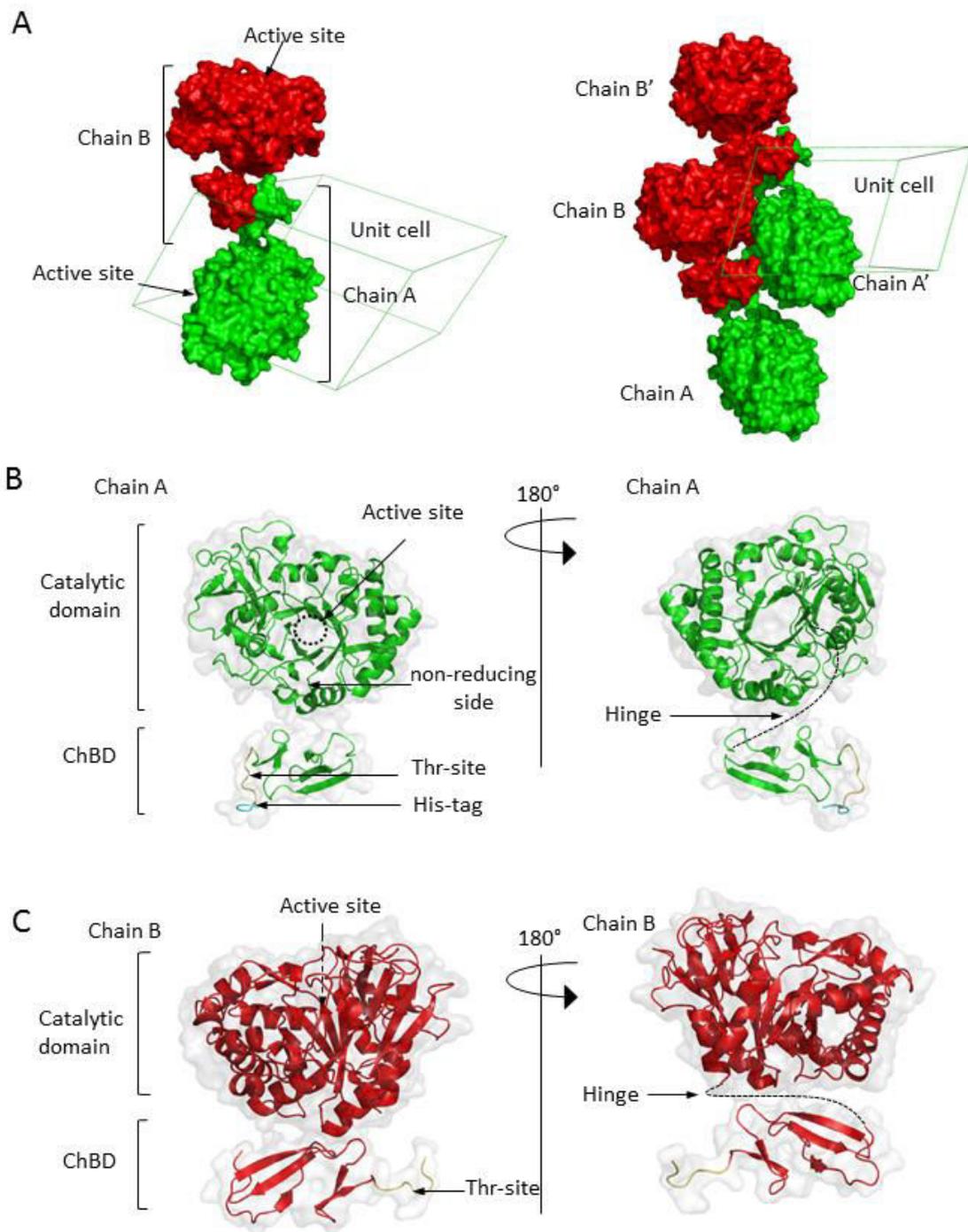


Figure 2
Fadel et al.

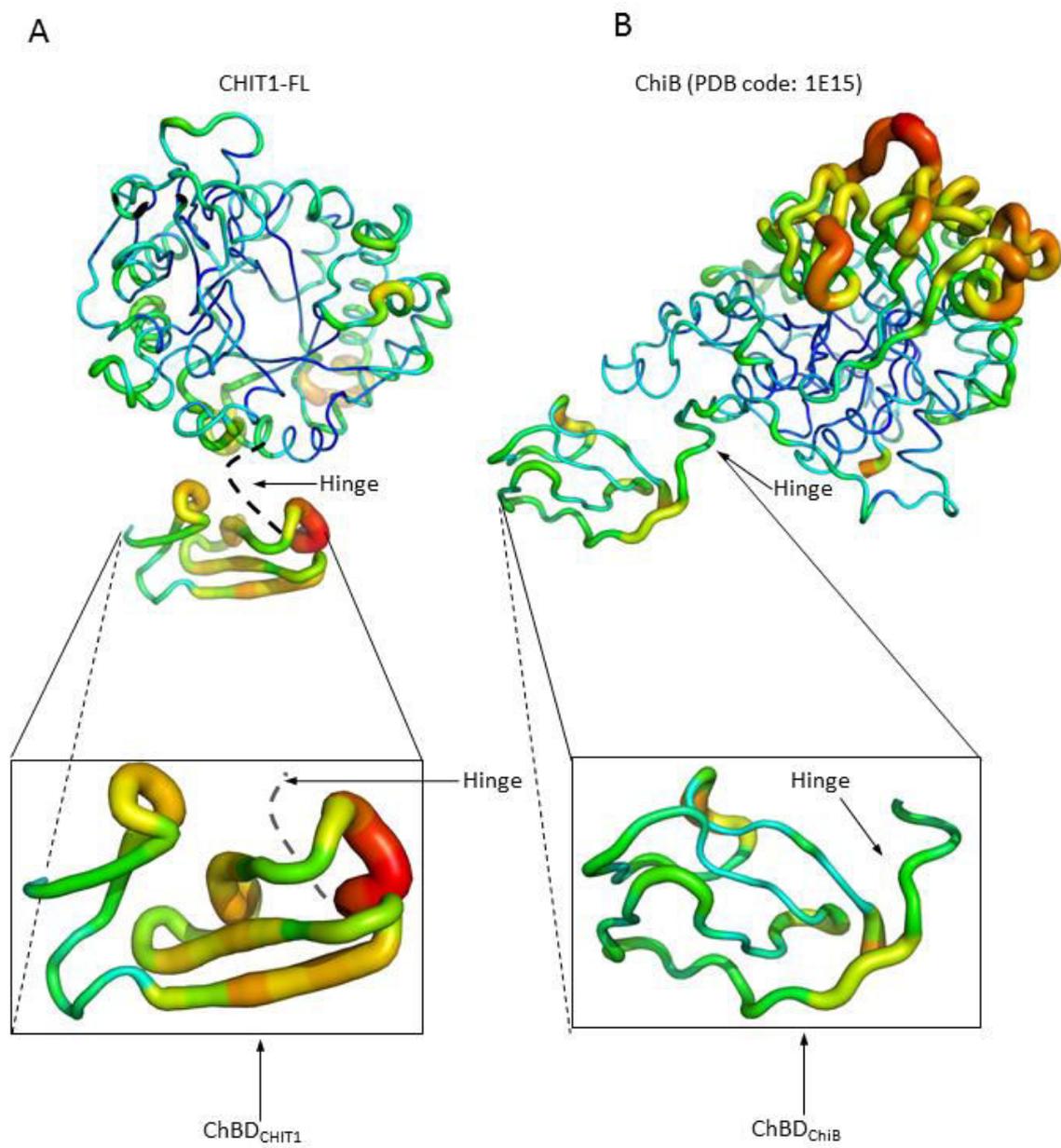


Figure 3
Fadel et al.

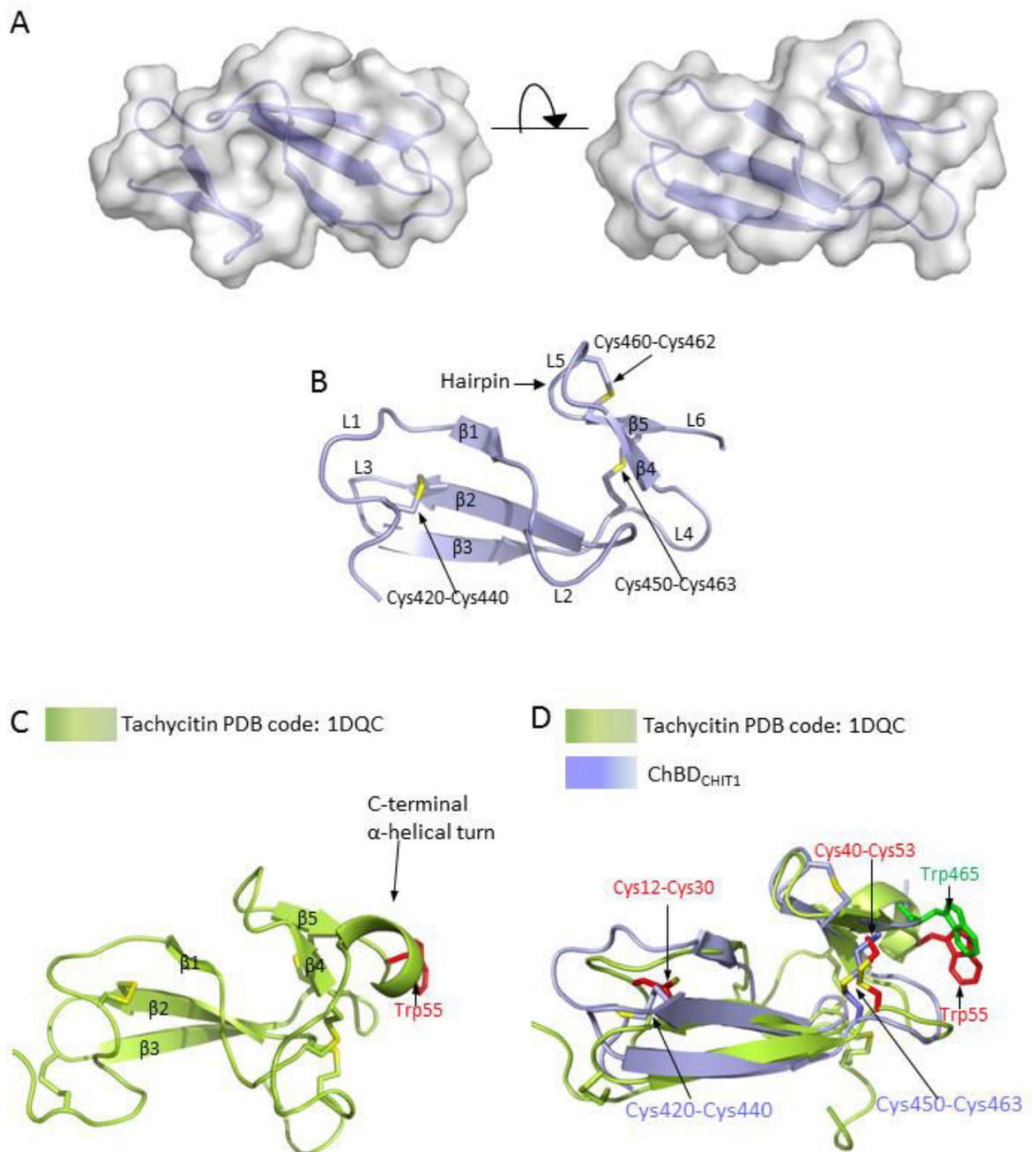


Figure 4
Fadel et al.

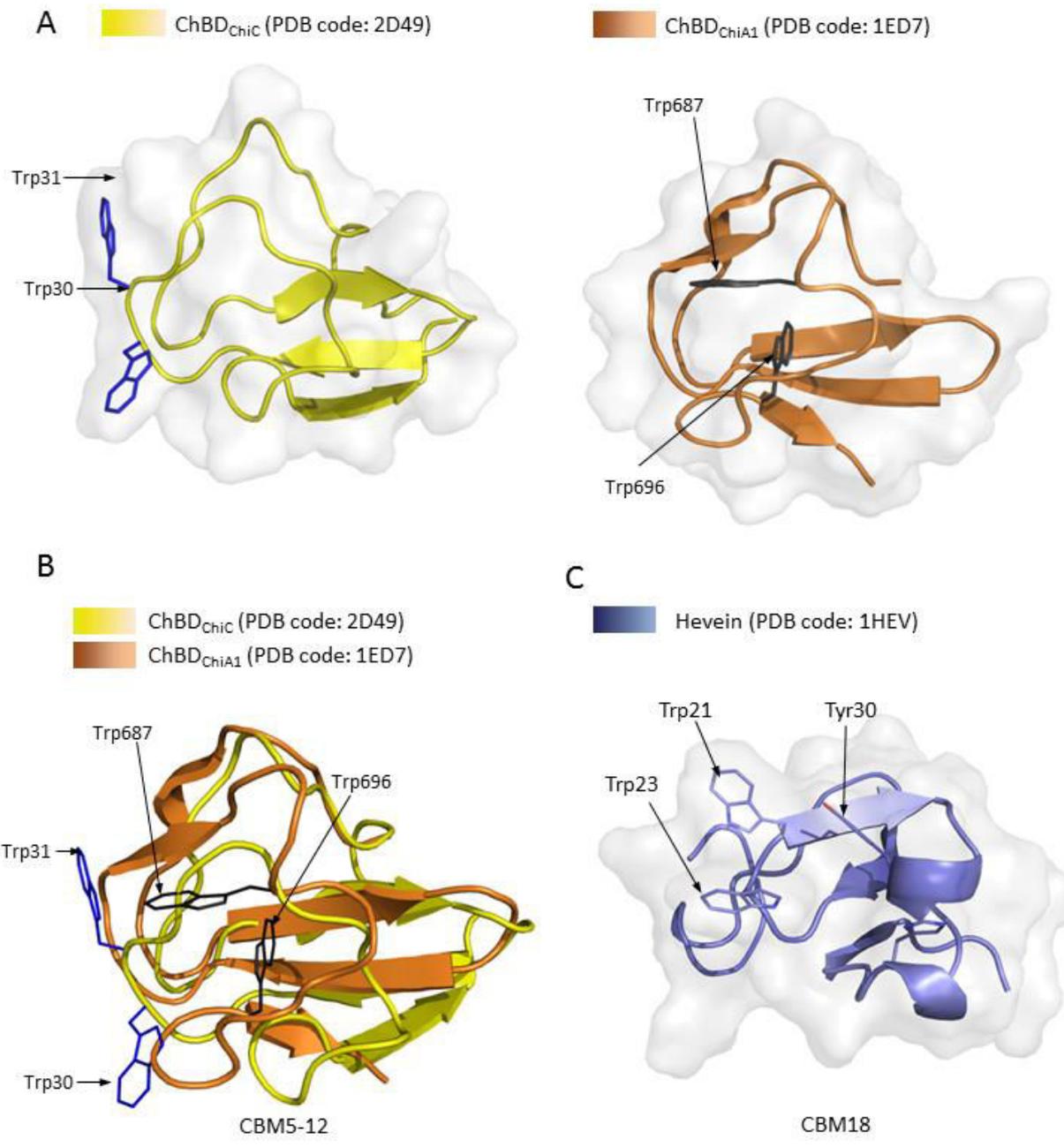
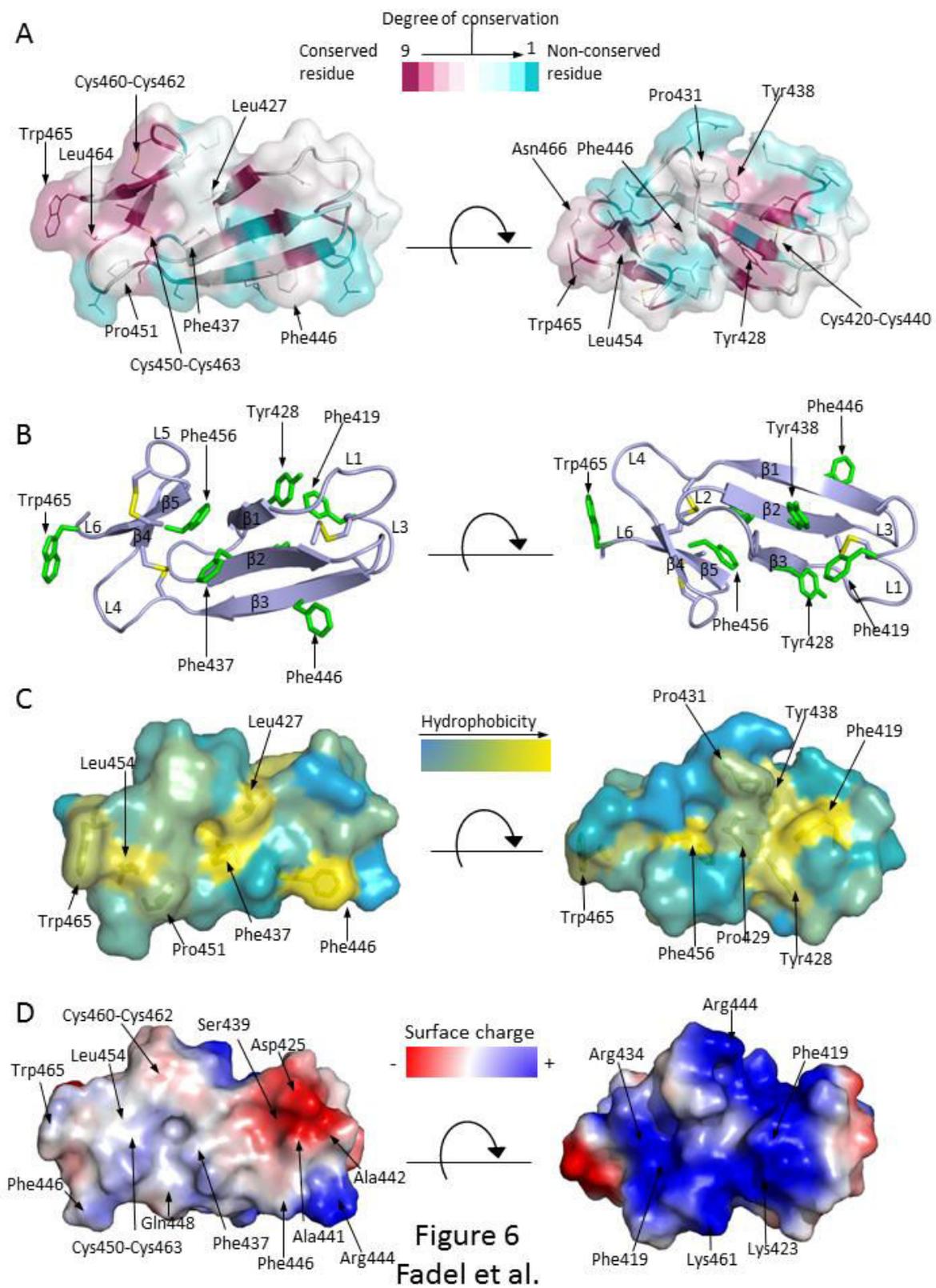


Figure 5
Fadel et al.



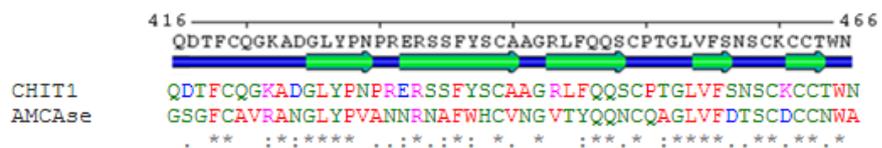
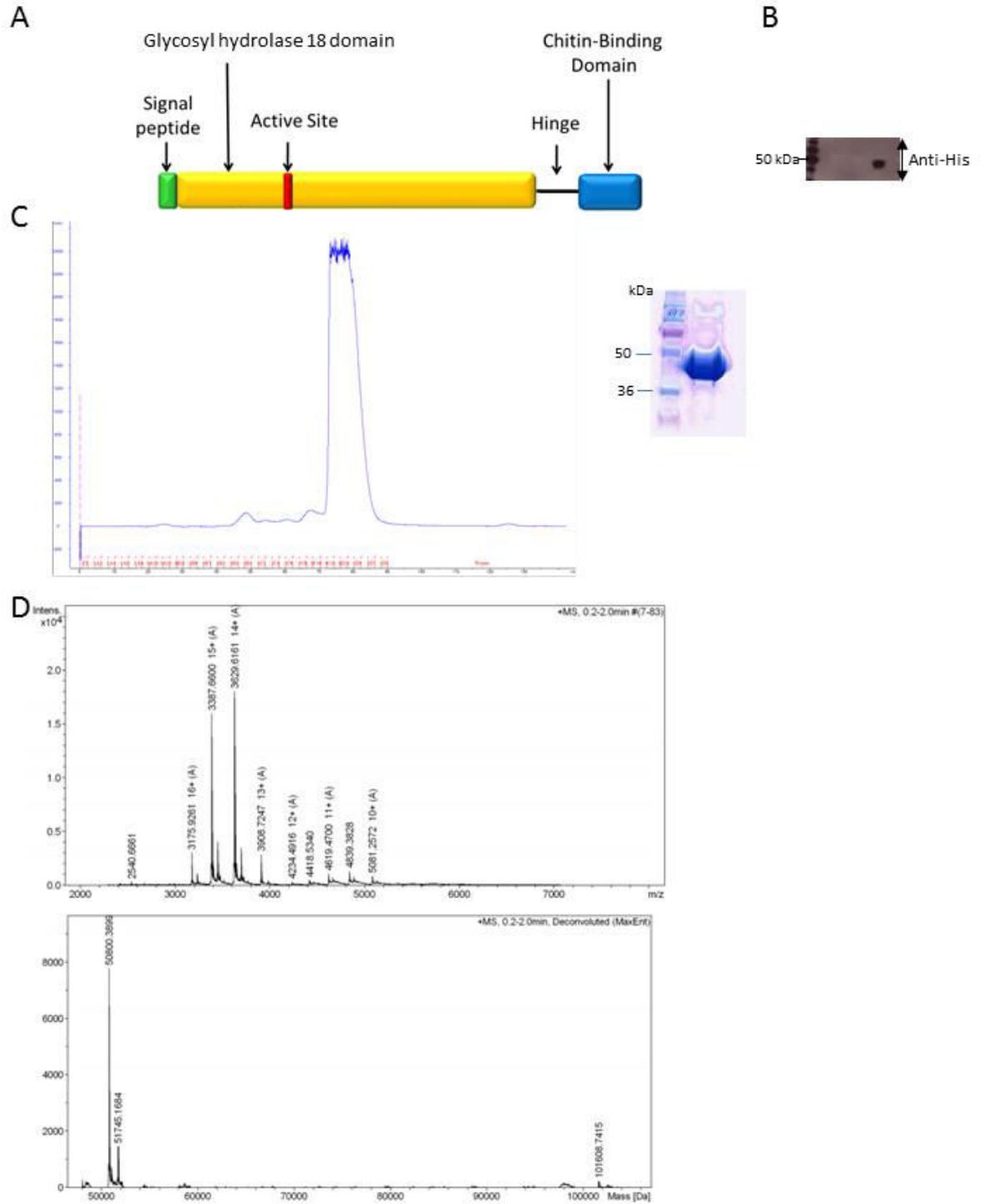


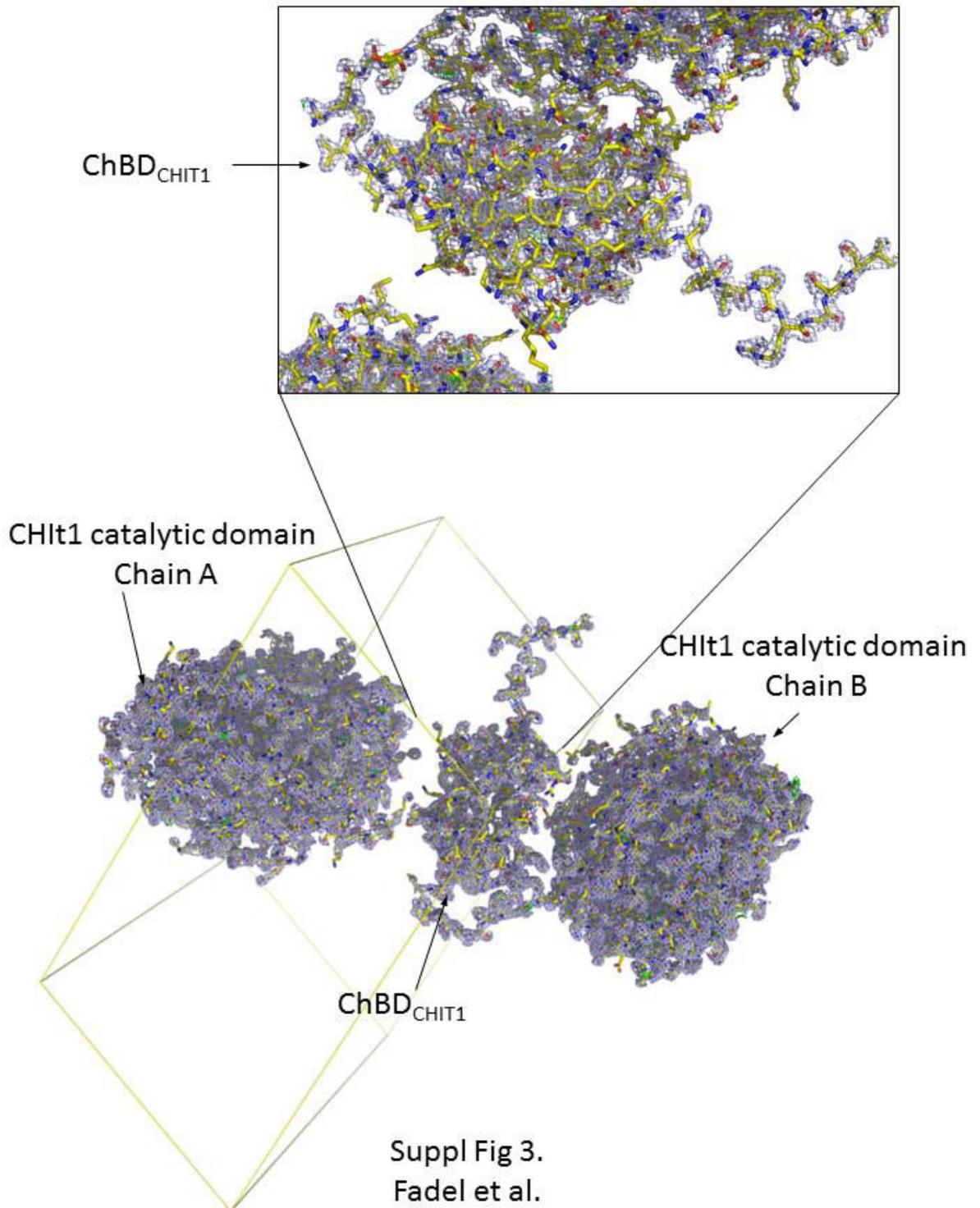
Figure 7
Fadel et al.

Supplementary data



Suppl. Fig 1
Fadel et al.

UniRef90_VP10003959996_474_520	--	F	C	A	G	K	A	N	G	I	Y	A	D	P	N	N	K	K	N	F	Y	N	C	L	N	G	Q	T	F	V	Q	S	C	M	E	G	L	V	F	D	A	C	S	Q	C	C	C	C	N	W	--	
UniRef90_F7AY19_406_455	ONT	F	C	K	D	K	A	N	G	L	Y	V	N	P	Q	D	H	T	A	F	Y	N	C	L	D	K	Q	T	F	T	Y	S	C	P	K	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--	
UniRef90_Q75NB9_437_483	--	F	C	A	G	K	S	N	G	L	Y	P	D	A	T	N	K	N	H	F	Y	E	C	S	Q	Q	K	T	F	T	D	Q	H	C	A	V	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--
UniRef90_VP10004443889_675_723	--	E	G	F	C	S	G	K	A	D	G	T	Y	S	D	P	E	D	M	T	K	F	I	K	C	V	G	K	T	F	T	F	H	C	A	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--	
UniRef90_I3JC30_423_472	GPG	F	C	N	G	K	P	D	G	L	Y	P	N	P	D	D	Q	T	S	F	Y	N	C	A	G	G	V	T	T	Y	V	S	Q	Q	G	A	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--
UniRef90_Q28F51_444_490	--	F	C	V	G	K	S	N	G	L	Y	P	P	D	T	N	K	S	O	F	F	N	C	L	N	G	I	T	Y	Q	Q	A	O	Q	T	S	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--	
UniRef90_W5KIH3_434_480	--	F	C	A	G	K	S	N	G	L	Y	P	N	P	T	N	K	N	Q	F	Y	S	G	D	Q	G	K	T	Y	L	E	H	C	A	T	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--	
UniRef90_VP10003C917D8_447_493	--	F	C	T	G	K	A	D	G	I	Y	S	D	P	E	N	T	K	F	Y	Q	C	V	Q	G	K	T	F	H	L	Q	C	A	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--			
UniRef90_VP10004571960_414_463	DGS	F	C	D	G	K	P	D	G	V	Y	A	H	P	D	A	S	N	K	F	Y	E	C	A	G	G	K	T	F	E	K	T	C	P	N	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--	
UniRef90_H3CKT8_419_468	GSD	F	C	S	G	K	A	D	G	M	Y	A	N	P	A	D	R	N	S	F	Y	V	C	A	G	G	I	T	Y	V	R	F	C	G	A	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--	
UniRef90_VP10003C29EF6_436_482	--	F	C	L	N	K	D	D	G	I	Y	P	V	P	E	N	T	K	F	Y	I	C	A	N	R	A	T	F	F	R	M	S	C	P	D	G	L	V	Y	N	G	T	S	G	C	C	C	C	C	N	W	--
UniRef90_VP100042BF4C6_235_281	--	F	C	A	S	M	A	S	G	L	F	P	D	P	S	N	S	N	S	F	Y	H	C	M	N	G	K	T	Y	L	E	N	C	Q	A	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--	
UniRef90_VP10003EDFD15_408_454	--	F	C	A	G	K	A	N	G	L	Y	P	V	A	N	N	R	N	A	F	W	H	C	L	N	G	V	T	Y	Q	Q	N	C	Q	A	R	H	V	F	D	M	S	G	C	C	C	C	C	N	W	--	
UniRef90_VP1000273CD1D_847_893	--	F	C	T	G	K	S	K	G	L	Y	P	S	P	T	S	K	C	A	F	Y	N	C	V	D	E	H	T	Y	E	A	O	Q	E	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--		
UniRef90_I3KJP1_416_466	GSO	F	C	A	G	K	P	D	G	L	Y	V	N	D	E	D	P	H	T	F	Y	Q	C	V	H	G	H	A	Y	I	Q	F	C	P	N	N	L	V	I	S	D	A	C	K	C	C	C	C	N	W	--	
UniRef90_F1S635_451_497	--	F	C	T	G	K	A	D	G	I	Y	S	D	P	K	S	T	K	Y	Q	C	A	G	G	R	T	F	H	F	Q	C	A	Q	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--			
UniRef90_H2M4Q2_430_479	GSO	F	C	A	G	K	P	D	G	L	Y	P	N	P	D	R	A	T	F	Y	M	C	A	G	G	I	T	H	L	T	C	G	A	G	L	V	I	N	D	S	G	C	C	C	C	C	N	W	--			
UniRef90_VP100046C3CC0_527_576	DDT	F	C	A	G	R	Q	D	G	T	Y	A	D	P	K	D	T	Q	F	Y	V	C	A	G	G	R	T	Y	H	F	T	C	A	A	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--		
UniRef90_G9MD29_439_488	DGS	F	C	A	G	K	A	N	G	M	Y	G	N	P	K	N	K	N	F	Y	N	C	S	A	G	K	T	Y	L	E	H	C	A	S	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--			
UniRef90_I3LL32_431_476	--	F	C	A	G	K	A	D	G	L	Y	P	V	A	N	N	R	N	A	F	W	H	C	Q	N	G	I	T	Y	E	Q	T	O	T	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--		
UniRef90_VP10002BCE0D6_431_480	ESQ	F	C	V	G	K	S	N	G	L	Y	P	S	P	T	S	K	C	T	F	Y	N	C	V	D	G	T	Y	E	A	O	Q	T	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--			
UniRef90_R4GAC5_269_315	--	F	C	A	G	K	A	N	G	L	Y	P	V	A	N	N	R	N	A	F	W	H	C	L	D	G	T	Y	E	Q	T	T	Y	E	Q	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--		
UniRef90_Q95M17_424_471	--	G	C	A	G	K	A	D	G	L	Y	P	V	A	N	N	R	N	A	F	W	H	C	V	N	G	I	T	Y	Q	K	N	O	L	T	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--	
UniRef90_VP10003343825_431_477	--	F	C	A	G	K	S	N	G	L	Y	P	V	V	N	T	H	A	F	W	R	C	L	N	G	I	T	Y	Q	Q	N	C	Q	A	G	L	V	I	N	D	S	G	C	C	C	C	C	N	W	--		
UniRef90_VP1000273B3FD_675_724	ONT	F	C	E	N	Q	E	N	G	L	H	P	N	P	Q	D	F	A	T	F	Y	N	C	V	K	R	T	F	R	L	S	C	P	A	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--		
UniRef90_M3K1N7_413_458	--	C	T	G	R	A	E	G	I	Y	P	D	P	E	N	T	K	F	Y	I	C	A	N	R	A	T	F	F	R	M	S	C	P	R	G	L	V	I	N	D	S	G	C	C	C	C	C	N	W	--		
UniRef90_W5N240_419_468	DPR	F	C	V	G	K	V	D	G	L	Y	A	N	V	M	D	P	N	S	F	Y	Q	C	S	R	G	V	T	Y	I	Q	R	O	V	G	L	V	F	K	D	K	C	C	C	C	C	C	N	W	--		
UniRef90_VP10003C47A20_419_468	DGD	F	C	A	G	K	Q	D	G	I	H	A	D	P	Q	I	T	K	Y	M	C	A	G	G	Q	T	Y	H	F	T	C	A	A	G	L	V	I	D	S	G	C	C	C	C	C	N	W	--				
UniRef90_H2TD25_354_403	GSO	F	C	S	G	K	P	D	G	L	Y	A	N	P	D	N	K	N	S	F	Y	M	C	S	G	G	I	T	Y	V	R	F	Q	T	G	S	V	F	D	M	S	G	C	C	C	C	C	N	W	--		
UniRef90_VP10003W5LHJ3_252_298	--	F	C	L	G	K	S	D	G	I	Y	P	N	P	S	D	F	T	S	F	Y	Q	C	A	N	G	I	K	F	V	Q	T	C	P	A	G	L	V	I	S	D	S	G	C	C	C	C	C	N	W	--	
UniRef90_H2MET2_423_472	DTN	F	C	A	T	K	I	G	G	V	Y	A	K	P	D	A	F	G	S	F	Y	N	C	A	N	G	I	T	W	I	Q	N	C	P	H	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--	
UniRef90_VP10003F7BC08_403_449	--	F	C	E	G	K	A	D	G	I	Y	A	D	P	D	F	S	R	F	Y	N	C	A	G	G	T	Y	V	A	K	R	C	A	G	L	V	F	N	D	S	G	C	C	C	C	C	N	W	--			
UniRef90_VP10003317678_443_489	--	F	C	T	G	K	A	D	G	T	Y	T	D	P	E	D	M	T	K	F	Y	K	G	V	G	K	T	F	H	F	S	C	A	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--			
UniRef90_F1MVE6_418_467	GQO	F	C	T	G	K	A	D	G	I	Y	S	D	P	E	D	M	T	K	F	Y	Q	C	A	G	G	R	T	F	Y	F	Q	C	N	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--		
UniRef90_I3JC26_405_452	--	D	F	C	H	A	R	P	D	G	L	Y	E	N	P	A	D	K	T	F	F	Q	C	F	Q	O	N	T	Y	L	H	R	C	Q	F	G	L	I	Y	W	D	S	G	C	C	C	C	C	N	W	--	
UniRef90_VP10002C42F99_417_467	EDT	F	C	X	Q	G	K	A	D	G	L	Y	P	I	P	Q	K	S	S	Y	I	C	E	W	O	R	L	F	L	R	S	C	P	A	G	L	V	F	N	T	S	G	C	C	C	C	C	N	W	--		
UniRef90_H2N6E4_275_321	--	F	C	A	N	R	A	S	S	L	Y	P	D	P	T	D	K	N	A	F	Y	N	C	M	K	K	T	F	T	Q	R	C	Q	A	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--		
UniRef90_VP10004439B81_451_501	GSO	F	C	S	G	K	S	D	G	I	Y	K	N	E	D	R	A	T	F	Y	Q	C	V	S	G	I	T	Y	L	Q	R	C	P	L	N	L	V	F	S	D	R	C	N	C	C	C	C	N	W	--		
UniRef90_VP10004431BF9_460_506	--	F	C	A	G	K	A	D	G	M	Y	I	D	P	E	N	T	K	F	Y	Q	C	A	G	G	K	T	F	H	F	Q	C	A	N	G	L	V	F	K	E	N	C	M	C	C	C	C	N	W	--		
UniRef90_H1A4C0_188_234	--	F	C	A	G	K	P	N	G	I	Y	A	D	P	S	N	R	N	F	Y	N	C	L	N	G	Q	T	F	V	Q	S	C	P	E	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--		
UniRef90_VP10003C92066_406_452	--	F	C	E	D	K	V	D	G	I	Y	G	D	P	D	D	S	R	F	E	C	V	N	G	T	Y	F	K	R	C	A	E	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--				
UniRef90_W5UCJ6_450_497	--	F	C	V	G	K	T	N	G	L	Y	P	C	A	T	N	S	N	K	F	Y	E	C	S	R	G	Q	T	Y	F	Q	C	A	E	T	G	L	V	F	D	M	S	G	C	C	C	C	C	N	W	--	
UniRef90_VP1000388D21F_434_480	--	F	C	L	H	K	A	N	G	I	Y	P	V	P	E	A	M	A	F	Y	I	C	A	N	R	R	T	F	R	M	S	C	P	E	G	L	J	F	D	Y	L	C	K	C	C	C						



Engaged interfaces

##	Interfacing structures	N _{occ}	Diss.	Sym.ID	Buried area, Å ²	ΔG, kcal/mol	N _{HB}	N _{SB}	N _{DS}	CSS
1	B + A	1	*	1_555	1063.4 (50%)	-12.2 (100%)	11 (100%)	0	0	1.000

##	NN ↔	Range	Structure 1			×	Structure 2			Interface	ΔG	ΔG	N _{HB}	N _{SB}	N _{DS}	CSS		
			N _H	N _{DS}	Surface Å ²	Range	Symmetry op-n	Sym.ID	N _H	N _{DS}	Surface Å ²	area, Å ²	kcal/mol	P-value				
1	*	B	108	20	19364	o A	x,y,z	1_555	104	31	19364	1063.4	-12.2	0.056	11	0	0	1.000
2	o	A	123	31	19364	≠ A	x-1,y,z	1_455	90	21	19364	1010.9	-8.1	0.166	16	3	0	0.000
3	o	B	82	19	19533	≠ B	x-1,y,z	1_455	108	31	19533	873.2	-1.4	0.553	12	6	0	0.000
4	o	A	85	28	19364	o B	x-1,y,z	1_455	88	24	19533	775.5	-1.7	0.535	10	2	0	0.000
5	o	A	64	15	19364	o B	x-2,y,z-1	1_354	63	19	19533	566.7	-0.6	0.575	9	2	0	0.000
6	o	B	37	12	19533	o A	-x,y-1/2,-z	2_545	48	16	19364	373.8	-2.5	0.374	4	0	0	0.000
7	o	B	24	9	19533	≠ B	-x+2,y-1/2,-z+1	2_746	26	8	19533	259.6	2.9	0.858	1	0	0	0.000
8	o	A	26	13	19364	≠ A	-x-1,y-1/2,-z	2_445	23	9	19364	214.8	1.6	0.758	2	0	0	0.000
9	o	A	7	3	19364	o B	x-1,y,z-1	1_454	9	3	19533	82.1	-0.5	0.399	0	0	0	0.000
10	o	B	4	2	19533	o A	-x-1,y-1/2,-z	2_445	5	3	19364	35.1	0.4	0.722	0	0	0	0.000
11	o	A	1	1	19364	o B	-x,y-1/2,-z	2_545	2	2	19533	3.3	0.2	0.783	0	0	0	0.000

Suppl Fig. 4
Fadel et al.

General conclusions and perspectives

Several aspects of the results presented in this thesis manuscript are novel and have generated new knowledge regarding the catalytic mechanism of CHIT1 and the main structural features of its ChBD. In the following next two sections, the main conclusions of my PhD work will be discussed and the perspectives of this study will be presented. The first part concerns the structural and mechanistic study on the catalytic domain of CHIT1 while the second section will deal with the conclusions and perspectives from the crystal structure of the full length CHIT1.

4.1 Conclusions and perspectives regarding the structural and mechanistic study of CHIT1 catalytic domain.

The optimization of the crystallization conditions has opened the door to obtain atomic resolution of CHIT1 and subsequently, the examination of new structural details that characterize this protein. The investigation of the protonation states of the CHIT1 active site revealed unexpected findings which set the basis to propose a new mechanism of chitin chain hydrolysis. Indeed, based on this study, new roles have been attributed to D138 which consists of a “swing” in the apo form and a “proton shuttle” during hydrolysis. Moreover, the rotation of E140 which liberates E140 from D138 provides an explanation to the surprising observation regarding the deprotonated outer oxygen of the catalytic glutamate in the CHIT1 apo form. This rotation appears to be important to protonate the oxygen of the glycosidic bond. My results highlight the importance of, Y27, together with the catalytic triad in the enzymatic mechanism. The obtained results show a shift in the sort of H-bond established between D138 and E140 from a LBHB in the apo form to a SIHB in complex with chitobiose which may be important to perform several cycles of hydrolysis. Since the studied residues in the CHIT1 active site are highly conserved in the GH18 family, therefore the proposed protonation pattern and hydrolysis mechanism could be extended to other active chitinases from the same family.

At the same time, since the obtained results are based on X-ray diffraction data, combining these data to neutron diffraction studies is obviously relevant to gain further

certainty regarding the proposed hydrolysis mechanism. In this respect, optimization of perdeuterated CHIT1 expression in the yeast *Pichia pastoris* has been started to perform a more detailed investigation of the hydrogen network.

Moreover, to gain further insight into the kinetics of the hydrolysis reaction, it would be interesting to perform assays by performing soaking experiments at low temperature (4°C) in order to detect different steps of hydrolysis in the crystal packing. Such assays are under optimization and I could confirm that decreasing the temperature of the 17°C to 4°C does not affect the crystallogenesis process.

Following the positive outcome of the cocrystallization of CHIT1 with 4-MU-NAG, the cocrystals of CHIT1-chitobiose could be used to perform soaking experiments with high concentration of the substrate in order to trigger the transglycosylation reaction in the crystal packing.

Furthermore, my results were not sufficient to predict which residue activates the hydrolytic water, but it is of high importance to determine how this water is activated. Therefore, obtaining the structure of CHIT1 at very high resolution with an analogue of oxazolinium intermediate might be helpful to obtain such information.

Further to crystallographic studies, to validate the proposed hydrolysis mechanism, QM/MM calculations will be pursued on a model of CHIT1 in complex with a polymeric substrate in order to follow the proton transfer in the active site from the apo form to the cleavage of the substrate.

It is worth noting that understanding and predicting protein-ligand binding is essential for drug design. Moreover, an important component of the binding energy is related to electrostatic interactions which depend on the protonation states of protein residues. It is also known that several GH18 chitinases in addition to CHIT1 are considered as pharmaceutical drug targets (Cole *et al.*, 2010), (Yang *et al.*, 2010); (Fusetti *et al.*, 2002a); (Rao, Houston, *et al.*, 2005). The majority of these chitinases have a conserved active site similar to the one in CHIT1. Therefore, our unexpected findings on the protonation states might turn out useful for the design of more specific and potent inhibitor for CHIT1 and GH18 chitinases.

4.2 CONCLUSIONS AND FUTURE WORK ON THE CHIT1-FL.

In this study, an original crystallogenesi approach based on many cycles of cross-seeding and micro-seeding has led to determine for the first time the full length structure of a mammalian chitinase and the first structure of a CBM14 ChBD associated to a chitinase. The crystallization and the structure of CHIT1-FL have revealed a high flexibility of its hinge leading to high mobility of the ChBD_{CHIT1}, which could be an unusual mechanism to get access to chitinous substrate in chitin containing organisms in order to protect from them. Therefore, to get further insight regarding the behavior of the CHIT1-FL in solution, experiments of Small Angle X-Ray Scattering (SAXS) could be conducted. Such experiments will provide the different conformational ensembles that could be adopted by this protein.

Last but not least, since there is no structural data regarding the interaction of ChBD_{CHIT1} with its substrate, it would be interesting to generate the inactive mutant form of CHIT1 in order to cocrystallize it or soak it with chitinous substrate. Another option for the near future is to model the interaction of ChBD_{CHIT1} with NAG chito-polymer.

Appendix 1

In the context of our collaboration with an industrial partner, I solved the crystal structures of CHIT1 in complex with 4 inhibitors. However for reasons of intellectual property, only the data collection and refinements statistics as well as the IC_{50} will be included in this annex section. The four inhibitors are named from 1 to 4.

Figure 1. Comparative IC_{50} curves on for a variety of measurements/well for each the four inhibitors

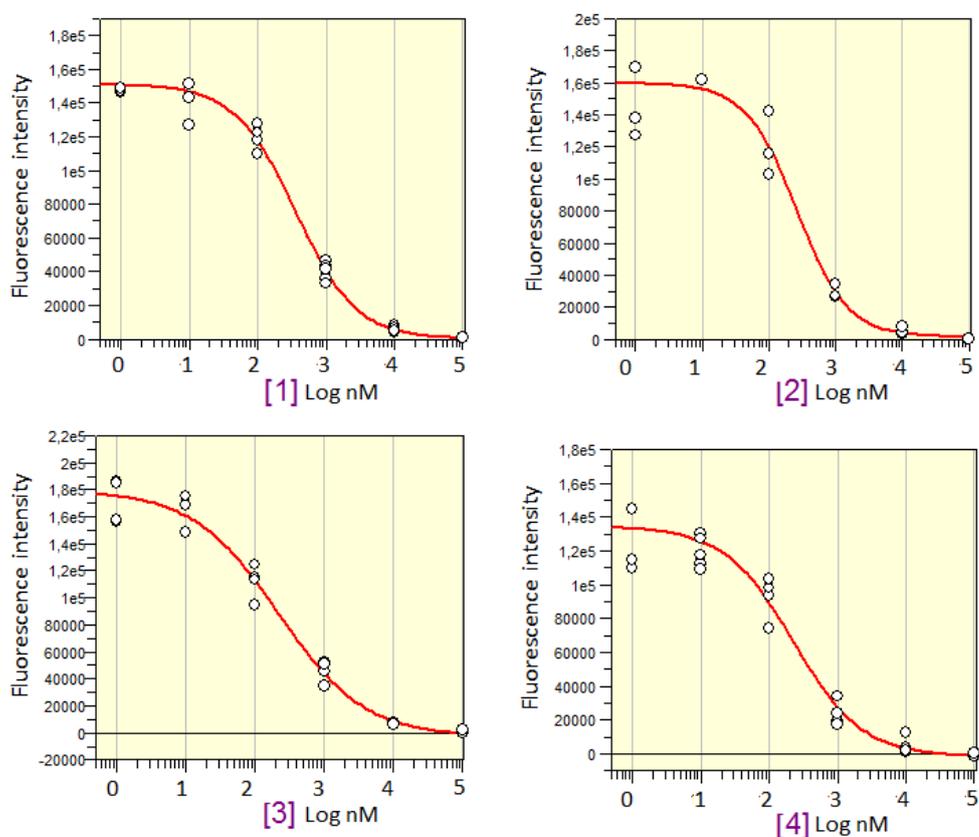


Table 1 IC_{50} of the four inhibitors.

Inhibitor	IC_{50} (nM)	Std Error
1	360.65	52.07
2	262.21	50.65
3	223.56	50.65
4	227.3	52.07

Table 1. Data collection and refinement statistics

	1	2	3	4
Wavelength (Å)	0.8	0.8	0.8	0.8
Resolution range (Å)	30.09 - 1.349 (1.397 - 1.349)	17.59 - 1.267 (1.313 - 1.267)	28.93 - 1.447 (1.499 - 1.447)	29.93 - 1.433 (1.484 - 1.433)
Space group	P 2 ₁ 2 ₁ 2			
Unit cell	85.9 106.0 42.15	85.45 105.6 42.137	85.6 105.99 42.24	85.5 105.7 41.9
Total reflections	531590	590678	381836	447878
Unique reflections	83374 (8047)	100288 (9387)	67782 (6543)	69842 (6647)
Multiplicity	6.4 (6.4)	5.9 (6.1)	5.9 (5.3)	6.4 (6.2)
Completeness (%)	97.44 (95.50)	98.49 (93.41)	97.70 (95.48)	98.75 (95.31)
Mean I/sigma(I)	34.08 (3.0)	25.9 (2.9)	26.32 (3.11)	30.29 (3.16)
Wilson B-factor	14.19	11.74	13.30	14.35
R-sym	2.8 (52.4)	5.8 (53.0)	5.7 (55.3)	5.5 (56.9)
R-factor	0.1321 (0.1744)	0.1369 (0.1866)	0.1440 (0.1599)	0.1373 (0.1593)
R-free	0.1560 (0.2186)	0.1630 (0.2225)	0.1693 (0.2023)	0.1697 (0.2232)
Number of atoms	6550	3472	3396	3433
macromolecules	3146	3061	3155	3048
ligands	65	32	41	55
water	222	379	200	328
Protein residues	377	377	377	372
RMS(bonds)	0.009	0.006	0.008	0.006
RMS(angles)	1.31	1.12	1.16	1.15
Ramachandran favored (%)	98	99	99	99
Ramachandran outliers (%)	0	0	0	0
Clashscore	5.95	4.51	5.40	4.82
Average B-factor	16.90	15.30	16.80	20.30
macromolecules	16.10	13.60	16.20	18.40
ligands	26.30	11.60	16.60	24.00
solvent	25.80	28.70	26.00	36.60

Statistics for the highest-resolution shell are shown in parentheses.

Appendix 2

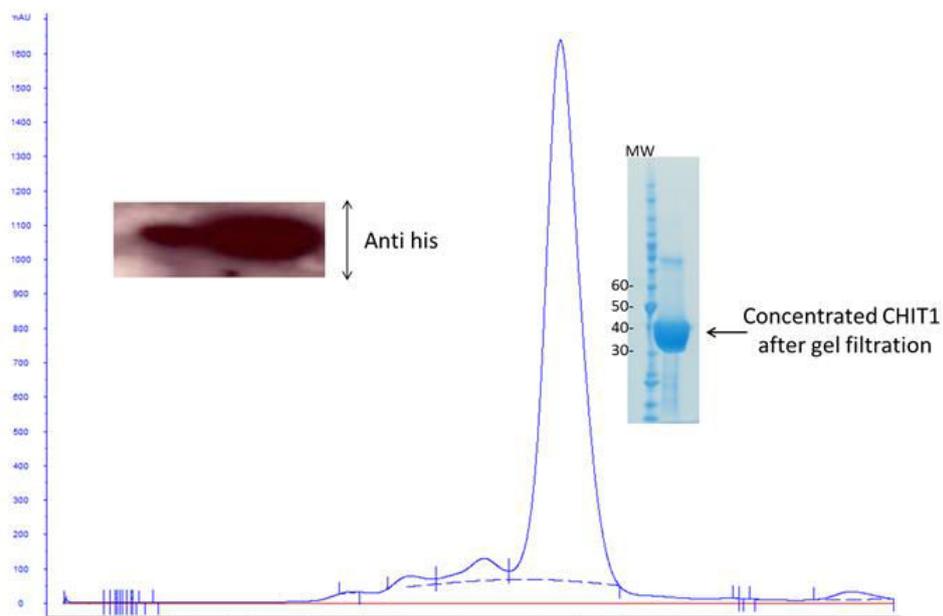


Figure 1 Western-blot analysis by anti-His antibody of four days conditioned media confirms the expression and secretion of CHIT1 catalytic domain at the predicted molecular weight. Chromatogram showing the elution peak during purification of the CHIT1-FL by size-exclusion chromatography. Band of CHIT1 after migration on SDS-PAGE followed by Coomassie Brilliant Blue staining.

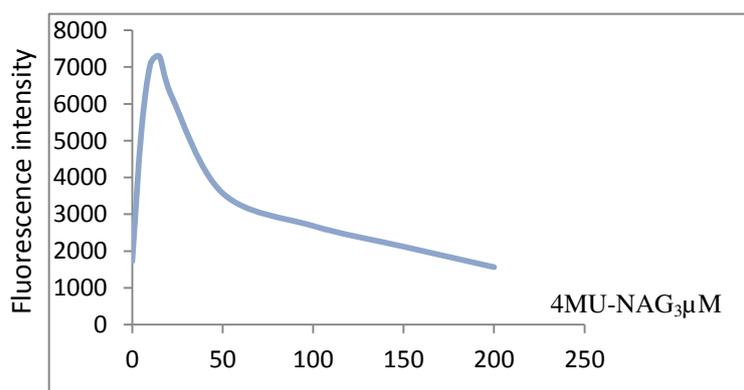


Figure 2. Fluorescence intensity of CHIT1 catalytic domain showing the hydrolysis of the substrate and the transglycosylation activity under excess of substrate concentration as reported by {Aguilera, 2003 #891}

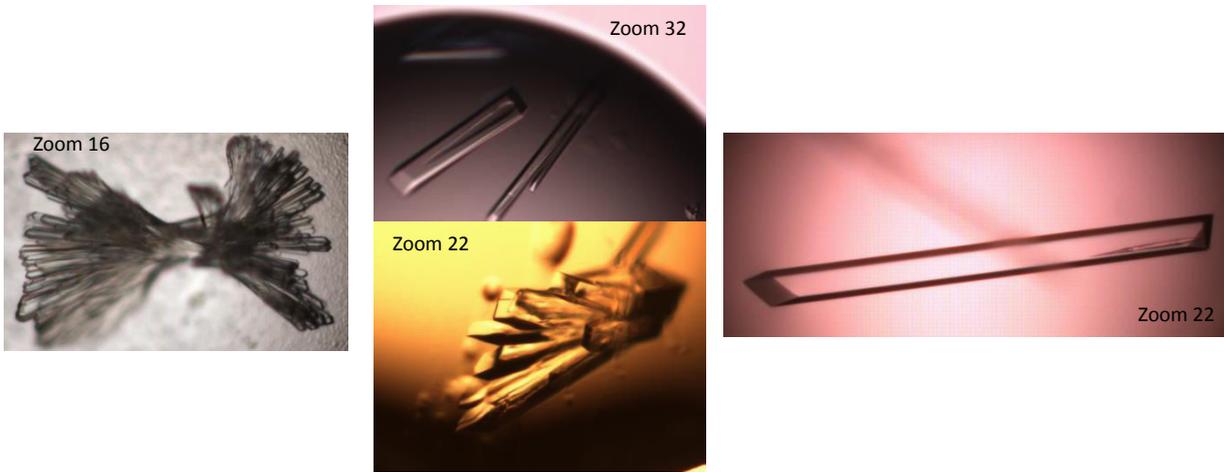


Figure 3. Crystals of CHIT1 catalytic domain. In the Left and middle before optimization of the crystallization condition. In the right a crystal with after optimization by microseeding with dimensions of 1 x 0.14 x 0.10 mm.

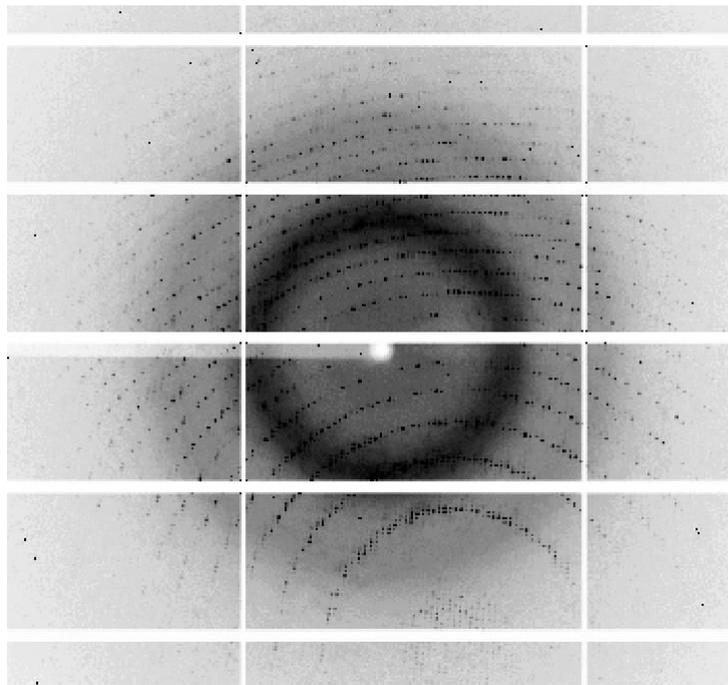
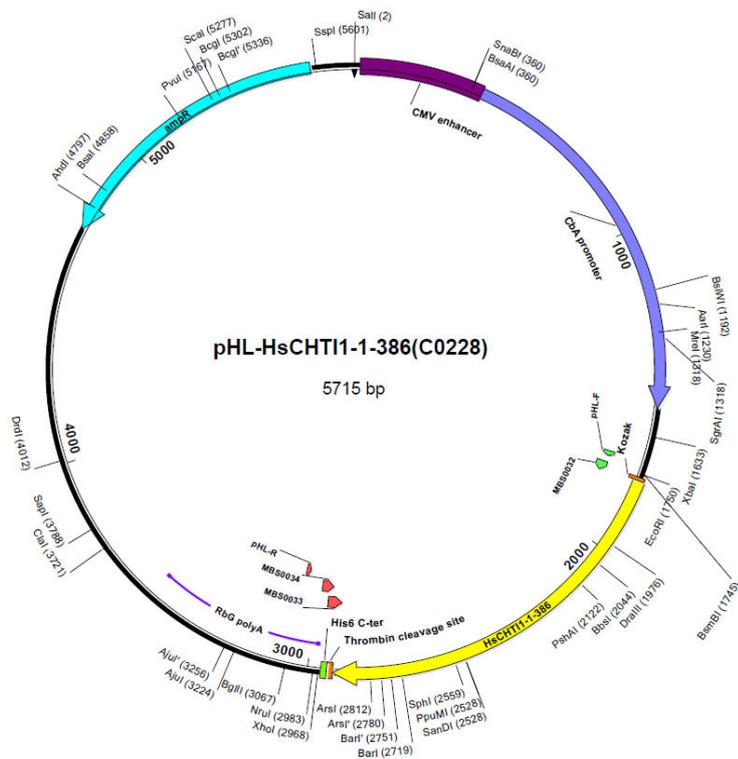
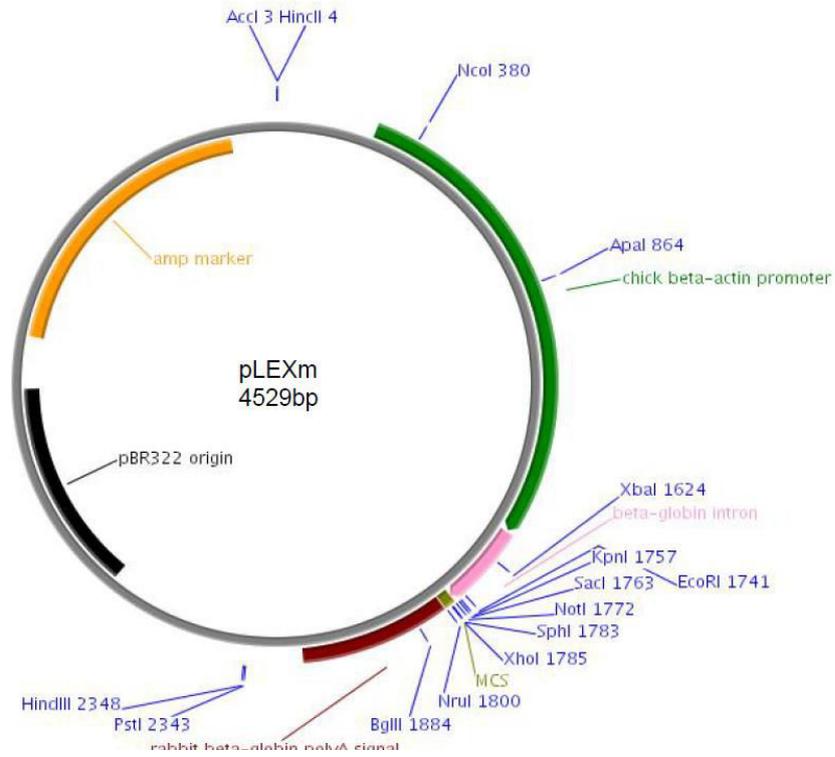


Figure 4. The X-ray diffraction pattern obtained with 1 second exposure and 0.25° oscillation range about the horizontal axis.

Appendix: 3



The final sequence of the PCR product encoding for CHIT1 catalytic domain:

ATGgtgcggtctgtggcctgggcaggtttcatggtcctgctgatgatcccatggggct
ctgctgcaaaactgggtctgctacttcaccaactgggcccagtacagacagggggaggctcgc
ttcctgcccaggacttggaccccagcctttgcaccacctcatctacgccttcgctggcat
gaccaaccaccagctgagcaccactgagtggaatgacgagactctctaccaggagtccaatg
gcctgaagaagatgaatcccaagctgaagaccctgtagccatcggaggctggaatttcggc
actcagaagttcacagatatggtagccacggccaacaaccgtcagacctttgtcaactcggc
catcaggtttctgcgcaaatacagctttgacggccttgaccttgactgggagtaccaggaa
gccaggggagccctgccgtagacaaggagcgttcacaaccctggtagcaggacttggccaat
gccttcagcaggaagcccagacctcaggggaaggaacgccttcttctgagtgcagcggttcc
agctgggcagacctatgtggatgctggatacagaggtggacaaaatcgcccagaacctggatt
ttgtcaaccttatggcctacgacttccatggctcctgggagaagggtcacgggacataacagc
ccctctacaagaggcaagaagagagtgggtgcagcagccagcctcaacgtggatgctgctgt
gcaacagtggtgcagaaggggacccctgccagcaagctgatccttggcatgcctacctacg
gacgctccttcacactggcctcctcatcagacaccagagtggggggcccagccacaggggtct
ggcactccaggccccttcaccaaggaaggaggatgctggcctactatgaagtctgctcctg
gaagggggccaccaaacagagaatccaggatcagaaggtgcctacatcttccgggacaacc
agtgggtgggctttgatgatgtggagagcttcaaaaccaaggtcagctatctgaagcagaag
ggactgggcggggccatgggtctgggcactggacttagatgactttgccggcttctcctgcaa
ccagggccgataccccctcatccagacgctacggcaggaactgagtCTGGTGCCACGCGGTT
CTCACCACCATCACCATCACTAATGA

The final sequence of the PCR product encoding for CHIT1-FL

ATGGTGCGGTCTGTGGCCTGGGCAGGTTTCATGGTCCTGCTGATGATCCCATGGGGC
TCTGCTGCAAACTGGTCTGCTACTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCTCG
CTTCCTGCCCAAGGACTTGGACCCCAGCCTTTGCACCCACCTCATCTACGCCTTCGCTGGCAT
GACCAACCACCAGCTGAGCACCCTGAGTGGAATGACGAGACTCTCTACCAGGAGTTCAATG
GCCTGAAGAAGATGAATCCCAAGCTGAAGACCCTGTTAGCCATCGGAGGCTGGAATTTCCGGC
ACTCAGAAGTTTACAGATATGGTAGCCACGGCCAACAACCGTCAGACCTTTGTCAACTCGGC
CATCAGGTTTCTGCGCAAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGAA
GCCAGGGGAGCCCTGCCGTAGACAAGGAGCGCTTCACAACCCTGGTACAGGACTTGGCCAAT

GCCTTCCAGCAGGAAGCCCAGACCTCAGGGAAGGAACGCCTTCTTCTGAGTGCAGCGGTTCC
AGCTGGGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAAATCGCCCAGAACCTGGATT
TTGTCAACCTTATGGCCTACGACTTCCATGGCTCTTGGGAGAAGGTCACGGGACATAACAGC
CCCCTCTACAAGAGGCAAGAAGAGAGTGGTGCAGCAGCCAGCCTCAACGTGGATGCTGCTGT
GCAACAGTGGCTGCAGAAGGGGACCCCTGCCAGCAAGCTGATCCTTGGCATGCCTACCTACG
GACGCTCCTTCACACTGGCCTCCTCATCAGACACCAGAGTGGGGGCCCCAGCCACAGGGTCT
GGCACTCCAGGCCCTTCACCAAGGAAGGAGGGATGCTGGCCTACTATGAAGTCTGCTCCTG
GAAGGGGGCCACCAAACAGAGAATCCAGGATCAGAAGGTGCCCTACATCTTCCGGGACAACC
AGTGGGTGGGCTTTGATGATGTGGAGAGCTTCAAAACCAAGGTCAGCTATCTGAAGCAGAAG
GGACTGGGCGGGCCATGGTCTGGGCACTGGACTTAGATGACTTTGCCGGCTTCTCCTGCAA
CCAGGGCCGATACCCCTCATCCAGACGCTACGGCAGGAACTGAGTCTTCCATACTTGCCTT
CAGGCACCCCAGAGCTTGAAGTTCCAAAACCAGGTCAGCCCTCTGAACCTGAGCATGGCCCC
AGCCCTGGACAAGACACGTTCTGCCAGGGCAAAGCTGATGGGCTCTATCCCAATCCTCGGGA
ACGGTCCAGCTTCTACAGCTGTGCAGCGGGCGGCTGTTCCAGCAAAGCTGCCCGACAGGCC
TGGTGTTCAGCAACTCCTGCAAATGCTGCACCTGGAATTTAGTTCCTCGTGGATCACTCGAGC
ACCACCACCACCACCACTGA

References

- Adams, D. J. (2004). *Microbiology* **150**, 2029-2035.
- Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 213-221.
- Afonine, P. V., Mustyakimov, M., Grosse-Kunstleve, R. W., Moriarty, N. W., Langan, P. & Adams, P. D. (2010). *Acta Crystallographica Section D* **66**, 1153-1163.
- Aguilera, B., Ghauharali-van der Vlugt, K., Helmond, M. T., Out, J. M., Donker-Koopman, W. E., Groener, J. E., Boot, R. G., Renkema, G. H., van der Marel, G. A., van Boom, J. H., Overkleeft, H. S. & Aerts, J. M. (2003). *J Biol Chem* **278**, 40911-40916.
- Ahmed, H. U., Blakeley, M. P., Cianci, M., Cruickshank, D. W., Hubbard, J. A. & Helliwell, J. R. (2007). *Acta Crystallogr D Biol Crystallogr* **63**, 906-922.
- Akagi, K., Watanabe, J., Hara, M., Kezuka, Y., Chikaishi, E., Yamaguchi, T., Akutsu, H., Nonaka, T., Watanabe, T. & Ikegami, T. (2006). *J Biochem* **139**, 483-493.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res* **25**, 3389-3402.
- Altschul, S. F., Wootton, J. C., Gertz, E. M., Agarwala, R., Morgulis, A., Schaffer, A. A. & Yu, Y. K. (2005). *FEBS J* **272**, 5101-5109.
- Andersen, N. H., Cao, B., Rodriguez-Romero, A. & Arreguin, B. (1993). *Biochemistry* **32**, 1407-1422.
- Andersen, O. A., Dixon, M. J., Eggleston, I. M. & van Aalten, D. M. (2005). *Nat Prod Rep* **22**, 563-579.
- Andersson, M. P., Jensen, J. H. & Stipp, S. L. (2013). *PeerJ* **1**, e198.
- Arai, N., Shiomi, K., Iwai, Y. & Omura, S. (2000). *J Antibiot (Tokyo)* **53**, 609-614.
- Arai, N., Shiomi, K., Yamaguchi, Y., Masuma, R., Iwai, Y., Turberg, A., Kolbl, H. & Omura, S. (2000). *Chem Pharm Bull (Tokyo)* **48**, 1442-1446.
- Aricescu, A. R., Lu, W. & Jones, E. Y. (2006). *Acta Crystallogr D Biol Crystallogr* **62**, 1243-1250.
- Aronson, N. N., Jr., Halloran, B. A., Alexeyev, M. F., Zhou, X. E., Wang, Y., Meehan, E. J. & Chen, L. (2006). *Biosci Biotechnol Biochem* **70**, 243-251.
- Asensio, J. L., Canada, F. J., Siebert, H. C., Laynez, J., Poveda, A., Nieto, P. M., Soedjanaamadja, U. M., Gabius, H. J. & Jimenez-Barbero, J. (2000). *Chem Biol* **7**, 529-543.
- Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. (2010). *Nucleic Acids Res* **38**, W529-533.
- Azzouz, F. (2001). PhD thesis, Neuchâtel.
- Barone, R., Simpoire, J., Malaguarnera, L., Pignatelli, S. & Musumeci, S. (2003). *Clin Chim Acta* **331**, 79-85.
- Baughman, R. P., Lower, E. E. & du Bois, R. M. (2003). *Lancet* **361**, 1111-1118.
- Bayer, E. A., Chanzy, H., Lamed, R. & Shoham, Y. (1998). *Curr Opin Struct Biol* **8**, 548-557.
- Beckham, G. T. & Crowley, M. F. (2011). *J Phys Chem B* **115**, 4516-4522.
- Berecibar, A., Grandjean, C. & Siriwardena, A. (1999). *Chem Rev* **99**, 779-844.
- Bergfors, T. (2003). *Seeds to crystals*, 2003/04/30 ed.

- Bhatnagar, R. K., Arora, N., Sachidanand, S., Shahabuddin, M., Keister, D. & Chauhan, V. S. (2003). *Biochem Biophys Res Commun* **304**, 783-787.
- Bierbaum, S., Nickel, R., Koch, A., Lau, S., Deichmann, K. A., Wahn, U., Superti-Furga, A. & Heinzmann, A. (2005). *Am J Respir Crit Care Med* **172**, 1505-1509.
- Blaiseau, P. L., Kunz, C., Grison, R., Bertheau, Y. & Brygoo, Y. (1992). *Curr Genet* **21**, 61-66.
- Blakeley, M. P., Ruiz, F., Cachau, R., Hazemann, I., Meilleur, F., Mitschler, A., Ginell, S., Afonine, P., Ventura, O. N., Cousido-Siah, A., Haertlein, M., Joachimiak, A., Myles, D. & Podjarny, A. (2008). *Proceedings of the National Academy of Sciences* **105**, 1844-1848.
- Blattner, R., Furneaux, R. H. & Lynch, G. P. (1996). *Carbohydr Res* **294**, 29-39.
- Bokma, E., Rozeboom, H. J., Sibbald, M., Dijkstra, B. W. & Beintema, J. J. (2002). *Eur J Biochem* **269**, 893-901.
- Bokma, E., van Koningsveld, G. A., Jeronimus-Stratingh, M. & Beintema, J. J. (1997). *FEBS Lett* **411**, 161-163.
- Bolam, D. N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M. P., Rixon, J. E., Boraston, A., Hazlewood, G. P. & Gilbert, H. J. (1998). *The Biochemical journal* **331 (Pt 3)**, 775-781.
- Boot, R. G., Blommaart, E. F., Swart, E., Ghauharali-van der Vlugt, K., Bijl, N., Moe, C., Place, A. & Aerts, J. M. (2001). *J Biol Chem* **276**, 6770-6778.
- Boot, R. G., Bussink, A. P. & Aerts, J. M. (2005). *J Immunol* **175**, 2041-2042.
- Boot, R. G., Hollak, C. E., Verhoek, M., Alberts, C., Jonkers, R. E. & Aerts, J. M. (2010). *Clin Chim Acta* **411**, 31-36.
- Boot, R. G., Renkema, G. H., Strijland, A., van Zonneveld, A. J. & Aerts, J. M. (1995). *J Biol Chem* **270**, 26252-26256.
- Boot, R. G., Renkema, G. H., Verhoek, M., Strijland, A., Blik, J., de Meulemeester, T. M., Mannens, M. M. & Aerts, J. M. (1998). *J Biol Chem* **273**, 25680-25685.
- Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. (2004). *The Biochemical journal* **382**, 769-781.
- Borbulevych, O. Y., Plumley, J. A., Martin, R. I., Merz, K. M., Jr. & Westerhoff, L. M. (2014). *Acta Crystallogr D Biol Crystallogr* **70**, 1233-1247.
- Bortone, K., Monzingo, A. F., Ernst, S. & Robertus, J. D. (2002). *J Mol Biol* **320**, 293-302.
- Brameld, K. A. & Goddard, W. A. (1998). *Journal of the American Chemical Society* **120**, 3571-3580.
- Brunger, A. T. & Rice, L. M. (1997). *Methods Enzymol* **277**, 243-269.
- Brunner, F., Stintzi, A., Fritig, B. & Legrand, M. (1998). *Plant J* **14**, 225-234.
- Brurberg, M. B., Eijsink, V. G. & Nes, I. F. (1994). *FEMS Microbiol Lett* **124**, 399-404.
- Brurberg, M. B., Nes, I. F. & Eijsink, V. G. (1996). *Microbiology* **142 (Pt 7)**, 1581-1589.
- Bucolo, C., Musumeci, M., Musumeci, S. & Drago, F. (2011). *Front Pharmacol* **2**, 43.
- Buhi, W. C. (2002). *Reproduction* **123**, 355-362.
- Bulawa, C. E., Miller, D. W., Henry, L. K. & Becker, J. M. (1995). *Proc Natl Acad Sci U S A* **92**, 10570-10574.
- Bussink, A. P., Speijer, D., Aerts, J. M. & Boot, R. G. (2007). *Genetics* **177**, 959-970.
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res* **37**, D233-238.

- Carrard, G., Koivula, A., Soderlund, H. & Beguin, P. (2000). *Proc Natl Acad Sci U S A* **97**, 10342-10347.
- Castellani, R. J., Siedlak, S. L., Fortino, A. E., Perry, G., Ghetti, B. & Smith, M. A. (2005). *Curr Alzheimer Res* **2**, 419-423.
- Chang, V. T., Crispin, M., Aricescu, A. R., Harvey, D. J., Nettleship, J. E., Fennelly, J. A., Yu, C., Boles, K. S., Evans, E. J., Stuart, D. I., Dwek, R. A., Jones, E. Y., Owens, R. J. & Davis, S. J. (2007). *Structure* **15**, 267-273.
- Chernin, L. S., De la Fuente, L., Sobolev, V., Haran, S., Vorgias, C. E., Oppenheim, A. B. & Chet, I. (1997). *Appl Environ Microbiol* **63**, 834-839.
- Chipman, D. M., Pollock, J. J. & Sharon, N. (1968). *J Biol Chem* **243**, 487-496.
- Cho, J. Y., Rosenthal, P., Miller, M., Pham, A., Aceves, S., Sakuda, S. & Broide, D. H. (2014). *Int Immunopharmacol* **18**, 35-42.
- Cohen-Kupiec, R. & Chet, I. (1998). *Curr Opin Biotechnol* **9**, 270-277.
- Cole, D. C., Olland, A. M., Jacob, J., Brooks, J., Bursavich, M. G., Czerwinski, R., DeClercq, C., Johnson, M., Joseph-McCarthy, D., Ellingboe, J. W., Lin, L., Nowak, P., Presman, E., Strand, J., Tam, A., Williams, C. M., Yao, S., Tsao, D. H. & Fitz, L. J. (2010). *J Med Chem* **53**, 6122-6128.
- Collinge, D. B., Kragh, K. M., Mikkelsen, J. D., Nielsen, K. K., Rasmussen, U. & Vad, K. (1993). *Plant J* **3**, 31-40.
- Correale, J. & Fiol, M. (2011). *Mult Scler* **17**, 521-531.
- Cozzarini, E., Bellin, M., Norberto, L., Polese, L., Musumeci, S., Lanfranchi, G. & Paoletti, M. G. (2009). *Eur J Gastroenterol Hepatol* **21**, 1119-1126.
- Dahiya, N., Tewari, R. & Hoondal, G. S. (2006). *Appl Microbiol Biotechnol* **71**, 773-782.
- Davies, G. & Henrissat, B. (1995). *Structure* **3**, 853-859.
- Davies, G. J., Wilson, K. S. & Henrissat, B. (1997). *Biochem J* **321 (Pt 2)**, 557-559.
- de Fost, M., Hollak, C. E., Groener, J. E., Aerts, J. M., Maas, M., Poll, L. W., Wiersma, M. G., Haussinger, D., Brett, S., Brill, N. & vom Dahl, S. (2006). *Blood* **108**, 830-835.
- Di Luca, M., Romi, R., Severini, F., Toma, L., Musumeci, M., Fausto, A. M., Mazzini, M., Gambellini, G. & Musumeci, S. (2007). *Parasitol Res* **100**, 1033-1039.
- Di Rosa, M., De Gregorio, C., Malaguarnera, G., Tuttobene, M., Biazzo, F. & Malaguarnera, L. (2013). *Mol Cell Biochem* **374**, 73-80.
- Di Rosa, M., Dell'Ombra, N., Zambito, A. M., Malaguarnera, M., Nicoletti, F. & Malaguarnera, L. (2006). *Eur J Neurosci* **23**, 2648-2656.
- Di Rosa, M., Malaguarnera, G., De Gregorio, C., Drago, F. & Malaguarnera, L. (2013). *Inflammation* **36**, 482-492.
- Di Rosa, M., Tibullo, D., Vecchio, M., Nunnari, G., Saccone, S., Di Raimondo, F. & Malaguarnera, L. (2014). *Bone* **61**, 55-63.
- Di Rosa, M., Zambito, A. M., Marsullo, A. R., Li Volti, G. & Malaguarnera, L. (2009). *J Cell Biochem* **107**, 881-889.
- Dickinson, K., Keer, V., Hitchcock, C. A. & Adams, D. J. (1989). *J Gen Microbiol* **135**, 1417-1421.
- Din, N., Damude, H. G., Gilkes, N. R., Miller, R. C., Jr., Warren, R. A. & Kilburn, D. G. (1994). *Proc Natl Acad Sci U S A* **91**, 11383-11387.
- Divne, C., Stahlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J. K., Teeri, T. & Jones, T. A. (1994). *Science* **265**, 524-528.

- Downing, K. J. & Thomson, J. A. (2000). *Can J Microbiol* **46**, 363-369.
- Ducruix, A. a. G., R. (1992).
- Dunkler, A., Walther, A., Specht, C. A. & Wendland, J. (2005). *Fungal Genet Biol* **42**, 935-947.
- Edreva, A. (2005). *Gen Appl Plant Physiol* **31**, 105-124.
- Eiberg, H. & Den Tandt, W. R. (1997). *Hum Genet* **101**, 205-207.
- Eide, K. B., Lindbom, A. R., Eijsink, V. G., Norberg, A. L. & Sorlie, M. (2013). *FEBS Lett* **587**, 3508-3513.
- Eide, K. B., Norberg, A. L., Heggset, E. B., Lindbom, A. R., Varum, K. M., Eijsink, V. G. & Sorlie, M. (2012). *Biochemistry* **51**, 487-495.
- Eijsink, V. G., Vaaje-Kolstad, G., Varum, K. M. & Horn, S. J. (2008). *Trends Biotechnol* **26**, 228-235.
- Elias, J. A., Homer, R. J., Hamid, Q. & Lee, C. G. (2005). *J Allergy Clin Immunol* **116**, 497-500.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 486-501.
- Erskine, P. T., Coates, L., Mall, S., Gill, R. S., Wood, S. P., Myles, D. A. & Cooper, J. B. (2003). *Protein Sci* **12**, 1741-1749.
- Falklöf, O., Collyer, C. & Reimers, J. (2012). *Theor Chem Acc* **131**, 1-16.
- Faramarzi, M., Fazeli, M., Yazdi, M. T., Adrangi, S., Al-Ahmadi, K. J., Tasharrofi, N. & Mohseni, F. A. (2009). *Biotechnology* **8**, 93-99.
- Ferrandon, S., Sterzenbach, T., Mersha, F. B. & Xu, M. Q. (2003). *Biochim Biophys Acta* **1621**, 31-40.
- Fitz, L. J., DeClercq, C., Brooks, J., Kuang, W., Bates, B., Demers, D., Winkler, A., Nocka, K., Jiao, A., Greco, R. M., Mason, L. E., Fleming, M., Quazi, A., Wright, J., Goldman, S., Hubeau, C. & Williams, C. M. (2012). *Am J Respir Cell Mol Biol* **46**, 71-79.
- Fujita, K., Shimomura, K., Yamamoto, K., Yamashita, T. & Suzuki, K. (2006). *Biochem Biophys Res Commun* **345**, 502-507.
- Fukui, T., Atomi, H., Kanai, T., Matsumi, R., Fujiwara, S. & Imanaka, T. (2005). *Genome Res* **15**, 352-363.
- Funkhouser, J. D. & Aronson, N. N., Jr. (2007). *BMC Evol Biol* **7**, 96.
- Fusetti, F., Pijning, T., Kalk, K. H., Bos, E. & Dijkstra, B. W. (2003). *J Biol Chem* **278**, 37753-37760.
- Fusetti, F., von Moeller, H., Houston, D., Rozeboom, H. J., Dijkstra, B. W., Boot, R. G., Aerts, J. M. & van Aalten, D. M. (2002a). *J Biol Chem* **277**, 25537-25544.
- Fusetti, F., von Moeller, H., Houston, D., Rozeboom, H. J., Dijkstra, B. W., Boot, R. G., Aerts, J. M. & van Aalten, D. M. (2002b). *The Journal of biological chemistry* **277**, 25537-25544.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. e., Wilkins, M., Appel, R. & Bairoch, A. (2005). *The Proteomics Protocols Handbook*, edited by J. Walker, pp. 571-607. Humana Press.
- Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E. & Ben-Tal, N. (2003). *Bioinformatics* **19**, 163-164.
- Goldstein, M. A., Takagi, M., Hashida, S., Shoseyov, O., Doi, R. H. & Segel, I. H. (1993). *J Bacteriol* **175**, 5762-5768.
- Gooday, G. (1990). *Advances in Microbial Ecology*, Vol. 11, *Advances in Microbial Ecology*, edited by K. C. Marshall, pp. 387-430. Springer US.

- Gordon-Thomson, C., Kumari, A., Tomkins, L., Holford, P., Djordjevic, J. T., Wright, L. C., Sorrell, T. C. & Moore, G. P. (2009). *Cell Mol Life Sci* **66**, 1116-1125.
- Hall, A. J., Morroll, S., Tighe, P., Gotz, F. & Falcone, F. H. (2008). *Microbes Infect* **10**, 69-78.
- Hamel, F., Boivin, R., Tremblay, C. & Bellemare, G. (1997). *J Mol Evol* **44**, 614-624.
- Hart, P. J., Pfluger, H. D., Monzingo, A. F., Hollis, T. & Robertus, J. D. (1995). *J Mol Biol* **248**, 402-413.
- Hartl, D., He, C. H., Koller, B., Da Silva, C. A., Homer, R., Lee, C. G. & Elias, J. A. (2008). *J Biol Chem* **283**, 33472-33482.
- Hartl, D., He, C. H., Koller, B., Da Silva, C. A., Kobayashi, Y., Lee, C. G., Flavell, R. A. & Elias, J. A. (2009). *J Immunol* **182**, 5098-5106.
- Hartl, L., Zach, S. & Seidl-Seiboth, V. (2012). *Appl Microbiol Biotechnol* **93**, 533-543.
- Hashimoto, M., Honda, Y., Nikaidou, N., Fukamizo, T. & Watanabe, T. (2000). *J Biosci Bioeng* **89**, 100-102.
- Hashimoto, M., Ikegami, T., Seino, S., Ohuchi, N., Fukada, H., Sugiyama, J., Shirakawa, M. & Watanabe, T. (2000). *Journal of bacteriology* **182**, 3045-3054.
- Hennig, M., Jansonius, J. N., Terwisscha van Scheltinga, A. C., Dijkstra, B. W. & Schlesier, B. (1995). *J Mol Biol* **254**, 237-246.
- Henrissat, B. (1991). *Biochem J* **280 (Pt 2)**, 309-316.
- Henrissat, B. & Bairoch, A. (1993). *Biochem J* **293 (Pt 3)**, 781-788.
- Henrissat, B. & Davies, G. (1997). *Current Opinion in Structural Biology* **7**, 637-644.
- Heo, K. W., Hur, D. Y., Park, S. K., Yang, Y. I., Kwak, H. H. & Kim, T. Y. (2011). *Otolaryngol Head Neck Surg* **145**, 660-665.
- Heras, B. & Martin, J. L. (2005). *Acta Crystallographica Section D* **61**, 1173-1180.
- Hollak, C. E., van Weely, S., van Oers, M. H. & Aerts, J. M. (1994). *J Clin Invest* **93**, 1288-1292.
- Hollis, T., Monzingo, A. F., Bortone, K., Ernst, S., Cox, R. & Robertus, J. D. (2000). *Protein Sci* **9**, 544-551.
- Horn, S. J., Sikorski, P., Cedervist, J. B., Vaaje-Kolstad, G., Sorlie, M., Synstad, B., Vriend, G., Varum, K. M. & Eijsink, V. G. (2006). *Proc Natl Acad Sci U S A* **103**, 18089-18094.
- Horn, S. J., Sorbotten, A., Synstad, B., Sikorski, P., Sorlie, M., Varum, K. M. & Eijsink, V. G. (2006). *FEBS J* **273**, 491-503.
- Houston, D. R., Recklies, A. D., Krupa, J. C. & van Aalten, D. M. (2003). *J Biol Chem* **278**, 30206-30212.
- Howard, E. I., Sanishvili, R., Cachau, R. E., Mitschler, A., Chevrier, B., Barth, P., Lamour, V., Van Zandt, M., Sibley, E., Bon, C., Moras, D., Schneider, T. R., Joachimiak, A. & Podjarny, A. (2004). *Proteins* **55**, 792-804.
- Huang, Q. S., Xie, X. L., Liang, G., Gong, F., Wang, Y., Wei, X. Q., Wang, Q., Ji, Z. L. & Chen, Q. X. (2012). *Glycobiology* **22**, 23-34.
- Hult, E. L., Katouno, F., Uchiyama, T., Watanabe, T. & Sugiyama, J. (2005). *Biochem J* **388**, 851-856.
- Ikegami, T., Okada, T., Hashimoto, M., Seino, S., Watanabe, T. & Shirakawa, M. (2000). *J Biol Chem* **275**, 13654-13661.
- Isaacs, A. T., Jasinskiene, N., Tretiakov, M., Thiery, I., Zettor, A., Bourgouin, C. & James, A. A. (2012). *Proc Natl Acad Sci U S A* **109**, E1922-1930.
- Iseli, B., Armand, S., Boller, T., Neuhaus, J. M. & Henrissat, B. (1996). *FEBS Lett* **382**, 186-188.
- Iseli, B., Boller, T. & Neuhaus, J. M. (1993). *Plant Physiol* **103**, 221-226.

- Isogai, A., Sato, M., Sakuda, S., Nakayama, J. & Suzuki, A. (1989). *Agricultural and Biological Chemistry* **53**, 2825-2826.
- Itoh, Y., Kawase, T., Nikaidou, N., Fukada, H., Mitsutomi, M. & Watanabe, T. (2002). *Biosci Biotechnol Biochem* **66**, 1084-1092.
- Izumida, H., Nishijima, M., Takadera, T., Nomoto, A. M. & Sano, H. (1996). *J Antibiot (Tokyo)* **49**, 829-831.
- Jee, J. G., Ikegami, T., Hashimoto, M., Kawabata, T., Ikeguchi, M., Watanabe, T. & Shirakawa, M. (2002). *The Journal of biological chemistry* **277**, 1388-1397.
- Jekel, P. A., Hartmann, B. H. & Beintema, J. J. (1991). *Eur J Biochem* **200**, 123-130.
- Jitonnom, J., Lee, V. S., Nimmanpipug, P., Rowlands, H. A. & Mulholland, A. J. (2011). *Biochemistry* **50**, 4697-4711.
- Jitonnom, J., Limb, M. A. & Mulholland, A. J. (2014). *J Phys Chem B* **118**, 4771-4783.
- Kabsch, W. (2010). *Acta Crystallographica Section D* **66**, 125-132.
- Kanneganti, M., Kamba, A. & Mizoguchi, E. (2012). *J Epithel Biol Pharmacol* **5**, 1-9.
- Karadag, B., Kucur, M., Isman, F. K., Hacibekiroglu, M. & Vural, V. A. (2008). *Circ J* **72**, 71-75.
- Kasprzewska, A. (2003). *Cell Mol Biol Lett* **8**, 809-824.
- Kato, M., Kamigaito, M., Sawamoto, M. & Higashimura, T. (1995). *Macromolecules* **28**, 1721-1723.
- Katoh, K. & Standley, D. M. (2013). *Mol Biol Evol* **30**, 772-780.
- Katouno, F., Taguchi, M., Sakurai, K., Uchiyama, T., Nikaidou, N., Nonaka, T., Sugiyama, J. & Watanabe, T. (2004). *J Biochem* **136**, 163-168.
- Kawada, M., Hachiya, Y., Arihiro, A. & Mizoguchi, E. (2007). *Keio J Med* **56**, 21-27.
- Kezuka, Y., Kojima, M., Mizuno, R., Suzuki, K., Watanabe, T. & Nonaka, T. (2010). *Proteins* **78**, 2295-2305.
- King, L. & Butler, G. (1998). *Curr Genet* **34**, 183-191.
- Kobayashi, K., Kimura, S., Heux, L. & Wada, M. (2013). *Carbohydr Polym* **97**, 105-110.
- Koga, D. *Specific inhibition of Bombyx mori chitinase by allosamidin.*
- Kolstad, G., Synstad, B., Eijsink, V. G. & van Aalten, D. M. (2002). *Acta Crystallogr D Biol Crystallogr* **58**, 377-379.
- Koshland, D. E. (1953). *Biological Reviews* **28**, 416-436.
- Kravchenko, N. A. (1967). *Proc R Soc Lond B Biol Sci* **167**, 429-430.
- Kucur, M., Isman, F. K., Balci, C., Onal, B., Hacibekiroglu, M., Ozkan, F. & Ozkan, A. (2008). *Urol Oncol* **26**, 47-52.
- Kuranda, M. J. & Robbins, P. W. (1991). *J Biol Chem* **266**, 19758-19767.
- Kzhyshkowska, J., Gratchev, A. & Goerdts, S. (2007). *Biomark Insights* **2**, 128-146.
- Labadaridis, I., Dimitriou, E., Theodorakis, M., Kafalidis, G., Velegraki, A. & Michelakakis, H. (2005). *Arch Dis Child Fetal Neonatal Ed* **90**, F531-532.
- Laemmli, U. K. (1970). *Nature* **227**, 680-685.
- Larsen, M. H., Leisner, J. J. & Ingmer, H. (2010). *Appl Environ Microbiol* **76**, 6470-6476.
- Larsen, T., Yoshimura, Y., Voldborg, B. G., Cazzamali, G., Bovin, N. V., Westerlind, U., Palcic, M. M. & Leisner, J. J. (2014). *FEBS letters* **588**, 746-751.
- Laskowski, R. A., Moss, D. S. & Thornton, J. M. (1993). *Journal of molecular biology* **231**, 1049-1067.
- Lawton, K., Uknes, S., Friedrich, L., Gaffney, T., Alexander, D., Goodman, R., Metraux, J. P., Kessmann, H., Goy, P. A., Rella, M. G., Ward, E. & Ryals, J. (1993). *Developments in*

- Plant Pathology*, Vol. 2, *Mechanisms of Plant Defense Responses*, edited by B. Fritig & M. Legrand, pp. 422-432. Springer Netherlands.
- Lee, C. G. (2009). *Yonsei Med J* **50**, 22-30.
- Lehtio, J., Sugiyama, J., Gustavsson, M., Fransson, L., Linder, M. & Teeri, T. T. (2003). *Proc Natl Acad Sci U S A* **100**, 484-489.
- Li, H. & Greene, L. H. (2010). *PLoS One* **5**, e8654.
- Li, M. & Elledge, S. (2012). *Methods in Molecular Biology*, Vol. 852, *Gene Synthesis*, edited by J. Peccoud, pp. 51-59. Humana Press.
- Lin, F. P., Chen, H. C. & Lin, C. S. (1999). *IUBMB Life* **48**, 199-204.
- Linder, M., Salovuori, I., Ruohonen, L. & Teeri, T. T. (1996). *J Biol Chem* **271**, 21268-21272.
- Lu, Y., Yang, H., Hu, H., Wang, Y., Rao, Z. & Jin, C. (2009). *Glycoconj J* **26**, 525-534.
- Lu, Y., Zen, K. C., Muthukrishnan, S. & Kramer, K. J. (2002). *Insect Biochem Mol Biol* **32**, 1369-1382.
- Luft, J. R. & DeTitta, G. T. (1999). *Acta Crystallographica Section D* **55**, 988-993.
- Madhuprakash, J., Singh, A., Kumar, S., Sinha, M., Kaur, P., Sharma, S., Podile, A. R. & Singh, T. P. (2013). *Int J Biochem Mol Biol* **4**, 166-178.
- Malaguarnera, L. (2006). *Cellular and molecular life sciences : CMLS* **63**, 3018-3029.
- Malaguarnera, L., Di Rosa, M., Zambito, A. M., dell'Ombra, N., Di Marco, R. & Malaguarnera, M. (2006). *The American journal of gastroenterology* **101**, 2060-2069.
- Malaguarnera, L., Di Rosa, M., Zambito, A. M., dell'Ombra, N., Nicoletti, F. & Malaguarnera, M. (2006). *Gut* **55**, 1313-1320.
- Malaguarnera, L., Musumeci, M., Di Rosa, M., Scuto, A. & Musumeci, S. (2005). *J Clin Lab Anal* **19**, 128-132.
- Malaguarnera, L., Musumeci, M., Licata, F., Di Rosa, M., Messina, A. & Musumeci, S. (2004). *Immunol Lett* **94**, 57-63.
- Malecki, P. H., Raczynska, J. E., Vorgias, C. E. & Rypniewski, W. (2013). *Acta Crystallogr D Biol Crystallogr* **69**, 821-829.
- Matsumoto, T., Inoue, H., Sato, Y., Kita, Y., Nakano, T., Noda, N., Eguchi-Tsuda, M., Moriwaki, A., Kan, O. K., Matsumoto, K., Shimizu, T., Nagasawa, H., Sakuda, S. & Nakanishi, Y. (2009). *Biochem Biophys Res Commun* **390**, 103-108.
- Maver, A., Medica, I., Salobir, B., Tercelj, M. & Peterlin, B. (2010). *Genet Mol Res* **9**, 58-68.
- Mayrose, I., Graur, D., Ben-Tal, N. & Pupko, T. (2004). *Mol Biol Evol* **21**, 1781-1791.
- McCarter, J. D. & Withers, S. G. (1994). *Curr Opin Struct Biol* **4**, 885-892.
- McCartney, L., Gilbert, H. J., Bolam, D. N., Boraston, A. B. & Knox, J. P. (2004). *Anal Biochem* **326**, 49-54.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *Journal of Applied Crystallography* **40**, 658-674.
- McCreath, K. J., Specht, C. A., Liu, Y. & Robbins, P. W. (1996). *Yeast* **12**, 501-504.
- McNab, R. & Glover, L. A. (1991). *FEMS Microbiology Letters* **82**, 79-82.
- McPherson, A. (1999). *Crystallization of Biological Macromolecules*. Cold Spring Harbor Laboratory Press.
- Melchers, L. S., Apotheker-de Groot, M., van der Knaap, J. A., Ponstein, A. S., Sela-Buurlage, M. B., Bol, J. F., Cornelissen, B. J., van den Elzen, P. J. & Linthorst, H. J. (1994). *Plant J* **5**, 469-480.
- Meot-Ner, M. (2012). *Chemical Reviews* **112**, PR22-PR103.

- Merzendorfer, H. & Zimoch, L. (2003). *J Exp Biol* **206**, 4393-4412.
- Metz, S., Kästner, J., Sokol, A. A., Keal, T. W. & Sherwood, P. (2014). *Wiley Interdisciplinary Reviews: Computational Molecular Science* **4**, 101-110.
- Mohanty, A. K., Singh, G., Paramasivam, M., Saravanan, K., Jabeen, T., Sharma, S., Yadav, S., Kaur, P., Kumar, P., Srinivasan, A. & Singh, T. P. (2003). *J Biol Chem* **278**, 14451-14460.
- Moreno, P. R., Falk, E., Palacios, I. F., Newell, J. B., Fuster, V. & Fallon, J. T. (1994). *Circulation* **90**, 775-778.
- Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Crystallogr D Biol Crystallogr* **67**, 355-367.
- Muzzarelli, R. A. A. (1996). *Chitin enzymology: proceedings of the 2nd International Symposium on Chitin Enzymology, Senigallia (Italy), 8-11 May, 1996*. Atec.
- Nance, J. P., Vannella, K. M., Worth, D., David, C., Carter, D., Noor, S., Hubeau, C., Fitz, L., Lane, T. E., Wynn, T. A. & Wilson, E. H. (2012). *PLoS Pathog* **8**, e1002990.
- Neuhaus, J. M., Sticher, L., Meins, F., Jr. & Boller, T. (1991). *Proc Natl Acad Sci U S A* **88**, 10362-10366.
- Nimlos, M. R., Beckham, G. T., Matthews, J. F., Bu, L., Himmel, M. E. & Crowley, M. F. (2012a). *J Biol Chem* **287**, 20603-20612.
- Nimlos, M. R., Beckham, G. T., Matthews, J. F., Bu, L., Himmel, M. E. & Crowley, M. F. (2012b). *The Journal of biological chemistry* **287**, 20603-20612.
- Nishimoto, Y., Sakuda, S., Takayama, S. & Yamada, Y. (1991). *J Antibiot (Tokyo)* **44**, 716-722.
- Norberg, A. L., Dybvik, A. I., Zakariassen, H., Mormann, M., Peter-Katalinic, J., Eijsink, V. G. & Sorlie, M. (2011). *FEBS Lett* **585**, 2339-2344.
- Ober, C. & Chupp, G. L. (2009). *Curr Opin Allergy Clin Immunol* **9**, 401-408.
- Ohno, M., Togashi, Y., Tsuda, K., Okawa, K., Kamaya, M., Sakaguchi, M., Sugahara, Y. & Oyama, F. (2013). *PLoS One* **8**, e67399.
- Ohno, T., Armand, S., Hata, T., Nikaidou, N., Henrissat, B., Mitsutomi, M. & Watanabe, T. (1996). *J Bacteriol* **178**, 5065-5070.
- Oku, T. & Ishikawa, K. (2006). *Biosci Biotechnol Biochem* **70**, 1696-1701.
- Olland, A. M., Strand, J., Presman, E., Czerwinski, R., Joseph-McCarthy, D., Krykbaev, R., Schlingmann, G., Chopra, R., Lin, L., Fleming, M., Kriz, R., Stahl, M., Somers, W., Fitz, L. & Mosyak, L. (2009a). *Protein Sci* **18**, 569-578.
- Olland, A. M., Strand, J., Presman, E., Czerwinski, R., Joseph-McCarthy, D., Krykbaev, R., Schlingmann, G., Chopra, R., Lin, L., Fleming, M., Kriz, R., Stahl, M., Somers, W., Fitz, L. & Mosyak, L. (2009b). *Protein science : a publication of the Protein Society* **18**, 569-578.
- Omura, S., Arai, N., Yamaguchi, Y., Masuma, R., Iwai, Y., Namikoshi, M., Turberg, A., Kolbl, H. & Shiomi, K. (2000). *J Antibiot (Tokyo)* **53**, 603-608.
- Otwinowski, Z. & Minor, W. (1997). Vol. 276, pp. 307-326. Elsevier.
- Papanikolau, Y., Prag, G., Tavlas, G., Vorgias, C. E., Oppenheim, A. B. & Petratos, K. (2001). *Biochemistry* **40**, 11338-11343.
- Papanikolau, Y., Tavlas, G., Vorgias, C. E. & Petratos, K. (2003). *Acta Crystallogr D Biol Crystallogr* **59**, 400-403.
- Park, S. K., Heo, K. W., Hur, D. Y. & Yang, Y. I. (2011). *Am J Rhinol Allergy* **25**, 12-14.

- Parker, C. L., Ventura, O. N., Burt, S. K. & Cachau, R. E. (2003). *Molecular Physics* **101**, 2659-2668.
- Pastores, G. M. & Barnett, N. L. (2005). *Expert Opin Emerg Drugs* **10**, 891-902.
- Pastores, G. M., Barnett, N. L. & Kolodny, E. H. (2005). *Clin Ther* **27**, 1215-1227.
- Payne, C. M., Baban, J., Horn, S. J., Backe, P. H., Arvai, A. S., Dalhus, B., Bjoras, M., Eijsink, V. G., Sorlie, M., Beckham, G. T. & Vaaje-Kolstad, G. (2012). *J Biol Chem* **287**, 36322-36330.
- Perrakis, A., Tews, I., Dauter, Z., Oppenheim, A. B., Chet, I., Wilson, K. S. & Vorgias, C. E. (1994). *Structure* **2**, 1169-1180.
- Purushotham, P. & Podile, A. R. (2012). *J Bacteriol* **194**, 4260-4271.
- Qureshi, A. M., Hannigan, A., Campbell, D., Nixon, C. & Wilson, J. B. (2011). *Genes Cancer* **2**, 74-87.
- Rai, M. & Bridge, P. D. (2009). *Applied Mycology*. CABI.
- Ramalho-Ortigao, J. M. & Traub-Cseko, Y. M. (2003). *Insect Biochem Mol Biol* **33**, 279-287.
- Ramanathan, M., Jr., Lee, W. K. & Lane, A. P. (2006). *Am J Rhinol* **20**, 330-335.
- Rao, F. V., Andersen, O. A., Vora, K. A., Demartino, J. A. & van Aalten, D. M. (2005). *Chem Biol* **12**, 973-980.
- Rao, F. V., Houston, D. R., Boot, R. G., Aerts, J. M., Hodkinson, M., Adams, D. J., Shiomi, K., Omura, S. & van Aalten, D. M. (2005). *Chem Biol* **12**, 65-76.
- Rao, F. V., Houston, D. R., Boot, R. G., Aerts, J. M., Sakuda, S. & van Aalten, D. M. (2003). *J Biol Chem* **278**, 20110-20116.
- Rao, F. V., Houston, D. R., Boot, R. G., Aerts, J. M. F. G., Hodkinson, M., Adams, D. J., Shiomi, K., Omura, S. & van Aalten, D. M. F. *Chemistry & Biology* **12**, 65-76.
- Ray, A. & Cohn, L. (1999). *J Clin Invest* **104**, 985-993.
- Renkema, G. H., Boot, R. G., Strijland, A., Donker-Koopman, W. E., van den Berg, M., Muijsers, A. O. & Aerts, J. M. (1997). *Eur J Biochem* **244**, 279-285.
- Rhodes, G. (2000). *Crystallography made crystal clear: A guide for users of macromolecular models*. San Diego: Academic Press.
- Robbins, P. W., Albright, C. & Benfield, B. (1988). *J Biol Chem* **263**, 443-447.
- Robyt, J. F. & French, D. (1967). *Arch Biochem Biophys* **122**, 8-16.
- Robyt, J. F. & French, D. (1970). *J Biol Chem* **245**, 3917-3927.
- Roske, Y., Sunna, A., Pfeil, W. & Heinemann, U. (2004). *J Mol Biol* **340**, 543-554.
- Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. & Jones, T. A. (1990). *Science* **249**, 380-386.
- Ruiz-Herrera, J. (1978). *Proc. First Internat. Conf. Chitin and Chitosan.*, **78(7)**, 11-21.
- Rupp, B. (2009). *Biomolecular Crystallography: Principles Practice, and Applications to Structural Biology*. New York: Garland Science.
- Ryde, U., Olsen, L. & Nilsson, K. (2002). *J Comput Chem* **23**, 1058-1070.
- Sahai, A. S. & Manocha, M. S. (1993). *FEMS Microbiology Reviews* **11**, 317-338.
- Sakuda, S., Isogai, A., Matsumoto, S. & Suzuki, A. (1987). *J Antibiot (Tokyo)* **40**, 296-300.
- Sakuda, S., Isogai, A., Matsumoto, S., Suzuki, A. & Koseki, K. (1986). *Tetrahedron Letters* **27**, 2475-2478.
- Sakuda, S., Isogai, A., Suzuki, A. & Yamada, Y. (1993). *Actinomycetologica* **7**, 50-57.
- Samac, D. A. & Shah, D. M. (1991). *Plant Cell* **3**, 1063-1072.

- Sanchez-Rodriguez, C., Bauer, S., Hematy, K., Saxe, F., Ibanez, A. B., Vodermaier, V., Konlechner, C., Sampathkumar, A., Ruggeberg, M., Aichinger, E., Neumetzler, L., Burgert, I., Somerville, C., Hauser, M. T. & Persson, S. (2012). *Plant Cell* **24**, 589-607.
- Seibold, M. A., Donnelly, S., Solon, M., Innes, A., Woodruff, P. G., Boot, R. G., Burchard, E. G. & Fahy, J. V. (2008). *J Allergy Clin Immunol* **122**, 944-950 e943.
- Sels, J., Mathys, J., De Coninck, B. M., Cammue, B. P. & De Bolle, M. F. (2008). *Plant Physiol Biochem* **46**, 941-950.
- Shahabuddin, M., Kaidoh, T., Aikawa, M. & Kaslow, D. C. (1995). *Exp Parasitol* **81**, 386-393.
- Shahabuddin, M. & Kaslow, D. C. (1993). *Parasitol Today* **9**, 252-255.
- Shao, R., Francescone, R., Ngernyuang, N., Bentley, B., Taylor, S. L., Moral, L. & Yan, W. (2014). *Carcinogenesis* **35**, 373-382.
- Sharon, N. & Seifter, S. (1964). *J Biol Chem* **239**, PC2398-2399.
- Sheldrick, G. (2008). *Acta Crystallographica Section A* **64**, 112-122.
- Shen, Z. & Jacobs-Lorena, M. (1998). *J Biol Chem* **273**, 17665-17670.
- Shen, Z. & Jacobs-Lorena, M. (1999). *J Mol Evol* **48**, 341-347.
- Sinnott, M. L. (1990). *Chemical Reviews* **90**, 1171-1202.
- Somers, P. J., Yao, R. C., Doolin, L. E., McGowan, M. J., Fukuda, D. S. & Mynderse, J. S. (1987). *J Antibiot (Tokyo)* **40**, 1751-1756.
- Songsiriritthigul, C., Pantoom, S., Aguda, A. H., Robinson, R. C. & Suginta, W. (2008). *J Struct Biol* **162**, 491-499.
- Sotgiu, S., Musumeci, S., Marconi, S., Gini, B. & Bonetti, B. (2008). *J Neuroimmunol* **197**, 70-73.
- Soulie, M. C., Perino, C., Piffeteau, A., Choquer, M., Malfatti, P., Cimerman, A., Kunz, C., Boccara, M. & Vidal-Cros, A. (2006). *Cell Microbiol* **8**, 1310-1321.
- Stewart, J. J. (2009). *J Mol Model* **15**, 765-805.
- Stewart, J. J. (2013). *J Mol Model* **19**, 1-32.
- Strobel, S., Roswag, A., Becker, N. I., Trenczek, T. E. & Encarnacao, J. A. (2013). *PLoS One* **8**, e72770.
- Stura, E. A. & Wilson, I. A. (1990). *Methods* **1**, 38-49.
- Stura, E. A. & Wilson, I. A. (1991). *Journal of Crystal Growth* **110**, 270-282.
- Suetake, T., Aizawa, T., Koganesawa, N., Osaki, T., Kobashigawa, Y., Demura, M., Kawabata, S., Kawano, K., Tsuda, S. & Nitta, K. (2002). *Protein Eng* **15**, 763-769.
- Suetake, T., Tsuda, S., Kawabata, S., Miura, K., Iwanaga, S., Hikichi, K., Nitta, K. & Kawano, K. (2000). *J Biol Chem* **275**, 17929-17932.
- Suginta, W. & Sritho, N. (2012). *Biosci Biotechnol Biochem* **76**, 2275-2281.
- Sun, Y. J., Chang, N. C., Hung, S. I., Chang, A. C., Chou, C. C. & Hsiao, C. D. (2001). *J Biol Chem* **276**, 17507-17514.
- Sutherland, T. E., Andersen, O. A., Betou, M., Eggleston, I. M., Maizels, R. M., van Aalten, D. & Allen, J. E. (2011). *Chem Biol* **18**, 569-579.
- Sutherland, T. E., Maizels, R. M. & Allen, J. E. (2009). *Clin Exp Allergy* **39**, 943-955.
- Svitil, A. L. & Kirchman, D. L. (1998). *Microbiology* **144 (Pt 5)**, 1299-1308.
- Synstad, B., Gaseidnes, S., Van Aalten, D. M., Vriend, G., Nielsen, J. E. & Eijsink, V. G. (2004). *Eur J Biochem* **271**, 253-262.
- Tabudravu, J. N., Eijsink, V. G., Gooday, G. W., Jaspars, M., Komander, D., Legg, M., Synstad, B. & van Aalten, D. M. (2002). *Bioorg Med Chem* **10**, 1123-1128.

- Taira, T., Fujiwara, M., Denhart, N., Hayashi, H., Onaga, S., Ohnuma, T., Letzel, T., Sakuda, S. & Fukamizo, T. (2010). *Biochim Biophys Acta* **1804**, 668-675.
- Takaya, N., Yamazaki, D., Horiuchi, H., Ohta, A. & Takagi, M. (1998). *Biosci Biotechnol Biochem* **62**, 60-65.
- Tamada, T., Kinoshita, T., Kurihara, K., Adachi, M., Ohhara, T., Imai, K., Kuroki, R. & Tada, T. (2009). *J Am Chem Soc* **131**, 11033-11040.
- Teeri, T. T. (1997). *Trends Biotechnol* **15**, 160-167.
- Tercelj, M., Rott, T. & Rylander, R. (2007). *Respir Med* **101**, 774-778.
- Tercelj, M., Salobir, B. & Rylander, R. (2008). *Med Hypotheses* **70**, 831-834.
- Terwisscha van Scheltinga, A. C., Armand, S., Kalk, K. H., Isogai, A., Henrissat, B. & Dijkstra, B. W. (1995). *Biochemistry* **34**, 15619-15623.
- Terwisscha van Scheltinga, A. C., Hennig, M. & Dijkstra, B. W. (1996). *J Mol Biol* **262**, 243-257.
- Terwisscha van Scheltinga, A. C., Kalk, K. H., Beintema, J. J. & Dijkstra, B. W. (1994). *Structure* **2**, 1181-1189.
- Tews, I., Terwisscha van Scheltinga, A. C., Perrakis, A., Wilson, K. S. & Dijkstra, B. W. (1997). *Journal of the American Chemical Society* **119**, 7954-7959.
- Tjoelker, L. W., Gosting, L., Frey, S., Hunter, C. L., Trong, H. L., Steiner, B., Brammer, H. & Gray, P. W. (2000a). *J Biol Chem* **275**, 514-520.
- Tjoelker, L. W., Gosting, L., Frey, S., Hunter, C. L., Trong, H. L., Steiner, B., Brammer, H. & Gray, P. W. (2000b). *The Journal of biological chemistry* **275**, 514-520.
- Tomme, P., Driver, D. P., Amandoron, E. A., Miller, R. C., Jr., Antony, R., Warren, J. & Kilburn, D. G. (1995). *J Bacteriol* **177**, 4356-4363.
- Tomme, P., Gilkes, N. R., Miller, R. C., Jr., Warren, A. J. & Kilburn, D. G. (1994). *Protein Eng* **7**, 117-123.
- Tsai, Y. L., Hayward, R. E., Langer, R. C., Fidock, D. A. & Vinetz, J. M. (2001). *Infect Immun* **69**, 4048-4054.
- Uchiyama, T., Katouno, F., Nikaidou, N., Nonaka, T., Sugiyama, J. & Watanabe, T. (2001). *The Journal of biological chemistry* **276**, 41343-41349.
- Udaya Prakash, N. A., Jayanthi, M., Sabarinathan, R., Kanguane, P., Mathew, L. & Sekar, K. (2010). *J Mol Evol* **70**, 466-478.
- Ueda, M., Ohata, K., Konishi, T., Sutrisno, A., Okada, H., Nakazawa, M. & Miyatake, K. (2009). *Appl Microbiol Biotechnol* **81**, 1077-1085.
- Umemoto, N., Ohnuma, T., Mizuhara, M., Sato, H., Skriver, K. & Fukamizo, T. (2013). *Glycobiology* **23**, 81-90.
- Uni, F., Lee, S., Yatsunami, R., Fukui, T. & Nakamura, S. (2012). *Biosci Biotechnol Biochem* **76**, 530-535.
- Vaaje-Kolstad, G., Horn, S. J., Sorlie, M. & Eijsink, V. G. (2013). *FEBS J* **280**, 3028-3049.
- Vaaje-Kolstad, G., Vasella, A., Peter, M. G., Netter, C., Houston, D. R., Westereng, B., Synstad, B., Eijsink, V. G. H. & van Aalten, D. M. F. (2004). *Journal of Biological Chemistry* **279**, 3612-3619.
- van Aalten, D. M., Komander, D., Synstad, B., Gaseidnes, S., Peter, M. G. & Eijsink, V. G. (2001). *Proc Natl Acad Sci U S A* **98**, 8979-8984.
- van Aalten, D. M., Synstad, B., Brurberg, M. B., Hough, E., Riise, B. W., Eijsink, V. G. & Wierenga, R. K. (2000). *Proc Natl Acad Sci U S A* **97**, 5842-5847.

- Van Damme, E. J., Culerrier, R., Barre, A., Alvarez, R., Rouge, P. & Peumans, W. J. (2007). *Plant Physiol* **144**, 662-672.
- van den Burg, H. A., Spronk, C. A. E. M., Boeren, S., Kennedy, M. A., Vissers, J. P. C., Vuister, G. W., de Wit, P. J. G. M. & Vervoort, J. (2004). *Journal of Biological Chemistry* **279**, 16786-16796.
- Van Dyken, S. J., Garcia, D., Porter, P., Huang, X., Quinlan, P. J., Blanc, P. D., Corry, D. B. & Locksley, R. M. (2011). *J Immunol* **187**, 2261-2267.
- van Eijk, M., Scheij, S. S., van Roomen, C. P., Speijer, D., Boot, R. G. & Aerts, J. M. (2007). *FEBS Lett* **581**, 5389-5395.
- van Eijk, M., van Roomen, C. P., Renkema, G. H., Bussink, A. P., Andrews, L., Blommaart, E. F., Sugar, A., Verhoeven, A. J., Boot, R. G. & Aerts, J. M. (2005). *Int Immunol* **17**, 1505-1512.
- Vandevenne, M., Campisi, V., Freichels, A., Gillard, C., Gaspard, G., Frere, J. M., Galleni, M. & Filee, P. (2011). *Protein science : a publication of the Protein Society* **20**, 1451-1463.
- Varrot, A., Frandsen, T. P., von Ossowski, I., Boyer, V., Cottaz, S., Driguez, H., Schulein, M. & Davies, G. J. (2003). *Structure* **11**, 855-864.
- Vernick, K. D., Oduol, F., Lazzaro, B. P., Glazebrook, J., Xu, J., Riehle, M. & Li, J. (2005). *Curr Top Microbiol Immunol* **295**, 383-415.
- Vinetz, J. M., Dave, S. K., Specht, C. A., Brameld, K. A., Xu, B., Hayward, R. & Fidock, D. A. (1999). *Proc Natl Acad Sci U S A* **96**, 14061-14066.
- Vinetz, J. M., Valenzuela, J. G., Specht, C. A., Aravind, L., Langer, R. C., Ribeiro, J. M. & Kaslow, D. C. (2000). *J Biol Chem* **275**, 10331-10341.
- Vuong, T. V. & Wilson, D. B. (2010). *Biotechnol Bioeng* **107**, 195-205.
- Watanabe, T., Ariga, Y., Sato, U., Toratani, T., Hashimoto, M., Nikaidou, N., Kezuka, Y., Nonaka, T. & Sugiyama, J. (2003). *Biochem J* **376**, 237-244.
- Watanabe, T., Ishibashi, A., Ariga, Y., Hashimoto, M., Nikaidou, N., Sugiyama, J., Matsumoto, T. & Nonaka, T. (2001). *FEBS Lett* **494**, 74-78.
- Watanabe, T., Ito, Y., Yamada, T., Hashimoto, M., Sekine, S. & Tanaka, H. (1994). *Journal of bacteriology* **176**, 4465-4472.
- Watanabe, T., Kobori, K., Miyashita, K., Fujii, T., Sakai, H., Uchida, M. & Tanaka, H. (1993). *J Biol Chem* **268**, 18567-18572.
- White, A. & Rose, D. R. (1997). *Curr Opin Struct Biol* **7**, 645-651.
- Wick, C. R., Hennemann, M., Stewart, J. J. & Clark, T. (2014). *J Mol Model* **20**, 2159.
- Wierenga, R. K. (2001). *FEBS Lett* **492**, 193-198.
- Wright, C. S. (1987). *J Mol Biol* **194**, 501-529.
- Wright, H. T., Sandrasegaram, G. & Wright, C. S. (1991). *J Mol Evol* **33**, 283-294.
- Wu, B., Zhang, B., Dai, Y., Zhang, L., Shang-Guan, K., Peng, Y., Zhou, Y. & Zhu, Z. (2012). *Plant Physiol* **159**, 1440-1452.
- Yamaguchi, S., Kamikubo, H., Kurihara, K., Kuroki, R., Niimura, N., Shimizu, N., Yamazaki, Y. & Kataoka, M. (2009). *Proc Natl Acad Sci U S A* **106**, 440-444.
- Yang, C. J., Liu, Y. K., Liu, C. L., Shen, C. N., Kuo, M. L., Su, C. C., Tseng, C. P., Yen, T. C. & Shen, C. R. (2009). *Hum Gene Ther* **20**, 1597-1606.
- Yang, J., Gan, Z., Lou, Z., Tao, N., Mi, Q., Liang, L., Sun, Y., Guo, Y., Huang, X., Zou, C., Rao, Z., Meng, Z. & Zhang, K. Q. (2010). *Microbiology* **156**, 3566-3574.

- Yu, N., Hayik, S. A., Wang, B., Liao, N., Reynolds, C. H. & Merz, K. M., Jr. (2006). *J Chem Theory Comput* **2**, 1057-1069.
- Yu, N., Li, X., Cui, G., Hayik, S. A. & Merz, K. M., 2nd (2006). *Protein Sci* **15**, 2773-2784.
- Zakariassen, H., Aam, B. B., Horn, S. J., Varum, K. M., Sorlie, M. & Eijsink, V. G. (2009). *J Biol Chem* **284**, 10610-10617.
- Zakariassen, H., Hansen, M. C., Joranli, M., Eijsink, V. G. & Sorlie, M. (2011). *Biochemistry* **50**, 5693-5703.
- Zees, A. C., Pyrpassopoulos, S. & Vorgias, C. E. (2009). *Biochim Biophys Acta* **1794**, 23-31.
- Zhang, J., Zhang, X., Arakane, Y., Muthukrishnan, S., Kramer, K. J., Ma, E. & Zhu, K. Y. (2011). *PLoS One* **6**, e19899.
- Zhao, Y., Bishop, B., Clay, J. E., Lu, W., Jones, M., Daenke, S., Siebold, C., Stuart, D. I., Jones, E. Y. & Aricescu, A. R. (2011). *J Struct Biol* **175**, 209-215.
- Zhou, Z. Y., Sakuda, S., Kinoshita, M. & Yamada, Y. (1993). *J Antibiot (Tokyo)* **46**, 1582-1588.
- Zhu, Z., Zheng, T., Homer, R. J., Kim, Y. K., Chen, N. Y., Cohn, L., Hamid, Q. & Elias, J. A. (2004). *Science* **304**, 1678-1682.

Firas FADEL

Résumé

La chitotriosidase (CHIT1) est l'une des deux chitinases humaines appartenant à la famille glycosyl hydrolase 18 (GH18). Cette famille d'enzymes, hautement conservée, hydrolyse la chitine, un polymère de N-acétyl glucosamine. Dans la dernière décennie, CHIT1 a attiré l'attention en raison de sa surexpression dans les troubles du système immunitaire. La protéine CHIT1 complète est composée d'un domaine catalytique lié par une charnière au domaine de liaison à la chitine (ChBD_{CHIT1}). CHIT1 présente plusieurs caractéristiques enzymatiques conservées dans la famille GH18 qui ne sont pas complètement comprises. Pour renforcer nos connaissances sur le mécanisme catalytique de CHIT1 et de la famille GH18, j'ai amélioré la résolution des structures obtenues par diffraction de rayon-X du domaine catalytique de CHIT1. Ces structures correspondent à la forme apo de CHIT1, pseudo-apo ainsi qu'en complexe avec un substrat synthétique ont été obtenues à des résolutions comprises entre 0.95 Å et 1.10 Å. Mes résultats combinés aux données des calculs quantiques QM/MM m'ont permis de proposer un nouveau mécanisme pour l'hydrolyse basé sur un réarrangement géométrique qui accompagne le transfert des protons dans le site actif. Nos données fournissent ainsi de nouvelles informations sur l'action de processivité et sur l'activité de transglycosylation. En outre, grâce à une nouvelle stratégie de cristallogénèse, la première structure cristalline de CHIT1 complète a pu être obtenue à une résolution de 1.95 Å. Cette nouvelle structure a révélé que le domaine ChBD_{CHIT1} adopte une configuration en β -sandwich distordue montrant une forte conservation de ses caractéristiques structurales avec les domaines ChBD des invertébrés.

Mon étude présente de nombreux aspects structuraux et mécanistiques de CHIT1 qui donnent de nouvelles perspectives sur son mode d'action et une nouvelle compréhension des caractéristiques enzymatiques conservées dans la famille GH18.

Mots-clés: CHIT1, famille GH18, structures cristallines, état de protonation, hydrolyse, mécanisme catalytique, domaine ChBD

Résumé en anglais

Chitotriosidase (CHIT1) is one of the two active human chitinases belonging to the glycosyl hydrolase family 18 (GH18), a highly conserved enzyme family. GH18 enzymes hydrolyze chitin, a N-acetyl glucosamine polymer. In the last decade, CHIT1 has attracted attention due to its upregulation in immune system disorders. Full length CHIT1 is composed of a catalytic domain linked by a hinge to its chitin-binding domain (ChBD_{CHIT1}). CHIT1 is characterized by many enzymatic features that are conserved in GH18 and not completely understood. To increase our knowledge on the catalytic mechanism in CHIT1 and GH18 family, I improved the X-ray resolution crystal structure of CHIT1 catalytic domain in apo and pseudo apo forms as well as in complex with a synthetic substrate to a resolution range between 0.95Å and at 1.10Å. My results supplied by QM/MM calculations allow me to suggest a new mechanism for hydrolysis based on a geometric rearrangement of the active site during protons translocation. Our data provide as well new insight into the processivity action and the transglycosylation activity. Moreover, thanks to a new a crystallogenesis strategy, I obtained the first crystal structure of full length CHIT1 at 1.95 Å resolution. This new structure allowed revealing the distorted β -sandwich fold structure of ChBD_{CHIT1} showing a high conservation of its structural features with ChBDs in invertebrates.

My study presents many structural and mechanistic aspects of CHIT1 which gives new insights onto its mode of action and shed light into the conserved enzymatic features in GH18 chitinase family.

Keywords: CHIT1, GH18 chitinase, crystal structures, protonation states, hydrolysis, catalytic mechanism, ChBD.