PIOTR MAŁECKI

STRUCTURAL AND FUNCTIONAL STUDIES OF CHITINASES FROM EXTREMOPHILES

Thesis presented to the Scientific Council of the Institute of Bioorganic Chemistry Polish Academy of Sciences in Poznan as a Ph.D. dissertation.

POZNAN 2014

THE RESEARCH DESCRIBED IN THIS THESIS HAS BEEN CARRIED OUT
AT THE INSTITUTE OF BIOORGANIC CHEMISTRY, POLISH ACADEMY OF SCIENCES IN POZNAN
IN THE STRUCTURE-FUNCTION RELATIONSHIP GROUP
AFFILIATED TO THE CENTER FOR BIOCRYSTALLOGRAPHIC RESEARCH
UNDER THE SUPERVISION OF PROFESSOR WOJCIECH RYPNIEWSKI.

FINANCIAL SUPPORT FOR THIS WORK WAS PROVIDED
BY THE EUROPEAN UNION WITHIN THE EUROPEAN REGIONAL DEVELOPMENTAL FUND.
INTERNATIONAL PHD PROGRAMME WAS SUPERVISED BY THE FOUNDATION FOR POLISH SCIENCE.

I WOULD LIKE TO THANK:

MY ADVISOR, PROFESSOR WOJCIECH RYPNIEWSKI
FOR SUPPORT, PATIENCE AND PRICELESS ADVICE

PROFESSOR CONSTANTINOS VORGIAS
FOR INTRODUCING ME TO ADVANCED PROTEIN BIOCHEMISTRY

PROFESSOR MARIUSZ JASKÓLSKI AND PROFESSOR MICHAŁ SIKORSKI
FOR MANY VALUABLE DISCUSSIONS

MY BELOVED WIFE AGATA AND SON LEON
FOR SUPPORT, UNDERSTANDING AND LOVE

MY WHOLE FAMILY
MY LAB MATES
FRIENDS

'It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.'

Sherlock Holmes Quote

-A Scandal in Bohemia

‘Data! Data! Data!’ he cried impatiently. ‘I can't make bricks without clay.’

Sherlock Holmes Quote

-The Adventure of the Copper Beeches

Table of contents

ABBREVIATIONS:

ADPs	Atomic Displacement Parameters
ATP	Adenosine Triphosphate
β/α-barrel	Conserved protein fold consisting of eight α-helices and eight parallel β-strands
CAZy	Database of Carbohydrate-Active enZymes
CBMs	Carbohydrate Binding Modules
ChBD	Chitin-Binding Domain
ChiA	Chitinase A from *Serratia marcescens*
ChiB	Chitinase B from *Serratia marcescens*
Chi40	Chitinase 40 from *Streptomyces thermoviolaceus*
dCTP	Deoxycytidine Triphosphate
dGTP	Deoxyguanosine Triphosphate
DSC	Differential Scanning Calorimetry
DLS	Dynamic Light Scattering
DNA	Deoxyribonucleic Acid
FnIII	Fibronectin Type III Domain
GH	Glycoside Hydrolases
GH18	Glycoside Hydrolase Family 18 (CAZy)
G+C content	Guanine-Cytosine Content
HIC	Hydrophobic Interaction Chromatography
Ig-like	Immunoglobin-like Domain
IPTG	Isopropyl-β-D-Thiogalactoside
IMAC	Ion Metal Affinity Chromatography
LIC	Ligation-Independent Cloning
MAD	Multi-Wavelength Anomalous Diffraction
MES	2-(N-morpholino)ethanesulfonic acid
*Mm*Chi60	Chitinase 60 from *Moritella marina*

MPD	2-Methyl-2,4-pentanediol
MR	Molecular replacement
MR-SAD	Combination of MR with SAD methods
NAG	N-acetyl-D-glucosamine
NAG_2	N,N'-Diacetylchitobiose
NAG_3	N,N',N''-Triacetylchitotriose
NAG_4	N,N',N'',N'''-Tetraacetyl chitotetraose
NAG_5	N,N',N'',N''',N''''-Pentaacetyl chitopentaose
NAG_6	N,N',N'',N''',N'''',N'''''-Hexaacetylchitohexaose
NCS	Non-Crystallography Symmetry
NMR	Nuclear Magnetic Resonance
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PIPE	Polymerase Incomplete Primer Extension
SAD	Single-Wavelength Anomalous Diffraction
SAXS	Small-Angle X-ray Scattering
SDM	Site-Directed Mutagenesis
SEC	Size Exclusion Chromatography
SPRINP	Single-Primer Reactions in Parallel
TCEP	3,3′,3″-Phosphanetriyltripropanoic acid, reducing agent
TIM barrel	β/α-barrel, named after triosephosphate isomerase
TLS	Translation-Liberation-Screw approximation

THE RESEARCH DESCRIBED IN THE THESIS HAS BEEN PRESENTED IN THE FOLLOWING PUBLICATIONS:

Małecki P.H., Raczynska J.E., Vorgias C.E., Rypniewski W.

Structure of a complete four-domain chitinase from Moritella marina, a marine psychrophilic bacterium, Acta Crystallographica Section D, 2013 69: 821-9

The article was distinguished as the cover story and cover illustration of the issue.

The article presents unliganded *Mm*Chi60 and its complexes with reaction intermediate and reaction product.

Małecki P.H., Vorgias C.E., Petoukhow M.V., Svergun D.I., Rypniewski W.

Crystal structures of substrate-bound chitinase from the psychrophilic bacterium Moritella marina and its structure in solution, Acta Crystallographica Section D, 2014 70: 676-84

The article presents E153Q *Mm*Chi60 mutant structure, E153Q in complex with ligand and SAXS structures.

Małecki P.H., Rypniewski W.

Survival in extremes, BioTechnologia, 2013 vol. 94(1): 27-30

Other publications by Piotr Małecki

Małecki P.H., Rypniewski W., Szymański M., Barciszewski J., Meyer A.

Binding of the plant hormone kinetin in the active site of mistletoe lectin I from Viscum album, BBA - Proteins and Proteomics, 2012, 1824:334-338

This paper describes a project that was Piotr Małecki's initiation into protein crystallography at the beginning of his post-graduate studies.

STRUCTURAL AND FUNCTIONAL STUDIES OF CHITINASES FROM EXTREMOPHILES.

THE AIM OF THE STUDY.

The aim of this study was to solve structures of chitinases from extremophilic organisms. The three dimensional structure of the protein defines its function. This relationship holds especially for enzymes because their structure must be retained for enzymatic activity. Three-dimensional structures of biological molecules are better conserved in evolution than their genetic or amino acid sequences. Therefore structural data are needed to answer many questions which cannot be answered by analysing the primary sequences of proteins.

The question addressed in my research project is: *What structural features in extremophile's proteins are responsible for their unusual properties?* Enzymes from psychrophiles are structurally adapted to work efficiently in a cold environment, while enzymes from thermophiles have structural features that keep them stable at high temperatures. Insights into proteins' ability to withstand and function optimally under extreme conditions can be gained by analysing and comparing proteins from psychrophiles and thermophiles. To answer the question about the structural bases of the adaptation I chose two proteins with unknown structures: chitinase 60 from a marine psychrophilic bacterium *Moritella marina* and chitinase 40 from a thermophilic *Streptomyces thermoviolaceus*. In the Protein Data Bank, there were only two incomplete structures of chitinases from extremophiles: one from a psychrophile and another from a hyperthermophile, whereas many structures were available from mezophilic organisms.

The two selected structures were to be compared with other known chitinases. Both of the chitinases that I chose belong to the same Glycoside Hydrolase Family 18 and share the similar β/α-barrel fold of the catalytic domain, also called a TIM-barrel. The β/α-barrel

enzymes share features that are comparable: the active site is always located at the carboxyl ends of the eight-stranded β-sheet that constitutes the core of the barrel, whereas the other side of the barrel is reported to affect stability.

Relating the structure of temperature extremophilic proteins to their physical and the chemical properties and explaining the principles of the enzymes' efficiency and stability is interesting on the level of basic science. It is also important for designing enzymes for the industrial usage.

INTRODUCTION

EXTREMOPHILES AND THEIR OCCURRENCE.

The ancient Romans created the world ‘extremus’, which is superlative of ‘exter’, meaning ‘being on the outside’. In the fifteenth century the word ‘extreme’ was adopted to English via Middle French (Rothschild & Mancinelli, 2001). At the beginning of the twenty-first century we know that life is spread across extremes of the environment, in which from anthropocentric point of view life is hard to imagine. Those organisms who 'love' (‘philos’ in ancient Greek texts) to live and thrive in extreme environments have been named by Macelroy ‘extremophiles’(MacElroy, 1974). An organism that lives in an environment that is extreme in more than one way is called a polyextremophile.

Extreme conditions can be classified according to the nature of the extremum. Physical extrema include temperature, radiation or pressure while geochemical extrema are: salinity, pH, desiccation, oxygen species or redox potential. Extremophiles can be found in all three taxonomic domains. Most hyperthermophiles are members of Archea and Bacteria. Eukaryotes are common among psychrophiles, acidophiles, alkalophiles, piezophiles, xerophiles and halophiles (respectively pertaining to low temperature, low pH, high pH, extremes of pressure, desiccation and salinity).

In this classification the definition of extreme conditions is formulated from the anthropocentric point of view. Otherwise, even conditions of the environment in which mankind lives could be viewed as extreme. For instance, free oxygen, whose reactive forms cause damage to nucleic acids, lipids and proteins can result in mutations or even death. Thus aerobes would be regarded as extremophilic as opposed to anaerobes. The aerobic environment, however, allows the production of ATP 18 times more efficiently than in anaerobic conditions.

Temperature is among the crucial factors that determine the diversity of life and this work is focused on temperature extremophiles. At one end of the temperature scale, ice destroys cells, whereas on the another end, heat causes denaturation of biomolecules.

Normally, temperature approaching 100ºC denatures proteins and nucleic acids and increases fluidity of cell membranes, disrupting cells and its compartments. Chlorophyll is unstable above 75ºC and photosynthesis stops. Gas solubility in aqueous environment is another factor that depends on temperature. It is difficult for organisms requiring O_2 or CO_2 to assimilate it at high temperature. Thus, temperature extremophiles living either in the heat or in the cold must adjust cell parameters or guard against external environment to maintain stable intracellular conditions.

PSYCHROPHILES

Cold environments are the most abundant on Earth and temperatures permanently below 5ºC extend over an estimated 75% of the biosphere, including the deep sea, the polar and alpine regions, and subterranean caves. This area is successfully colonised by organisms known as psychrophiles (optimal growth < 15ºC), including bacteria (Deming, 2002), archaea (Cavicchioli, 2006), eukarya (Peck, 2002) and also viruses (Wells & Deming, 2006). It is worth mentioning that psychrophilic organisms can catalyse reactions at 4ºC at rates comparable to their mesophilic counterparts at 37ºC (Pomeroy & Wiebe, 2001). In addition, some of them possess a doubling time comparable to *E. coli* at 37 ºC (Feller & Gerday, 2003).

The major obstacle to living in the cold is the freezing of water, which is generally lethal. There are only a few known exceptional organisms that take advantage of this, such as a nematode *Panagrolaimus davidi* which can withstand freezing of all the water in its body (Wharton & Ferns, 1995) and some frogs, turtles and snake whose winter surviving strategies is based on external water freezing (Storey & Storey, 1996). In addition to many general strategies developed for survival in the cold, such as alterations in membrane fluidity (Russell, 1997), expression of anti-freezing proteins (Muryoi *et al.*, 2004), production of compatible solutes and exopolysaccharides (Nichols *et al.*, 2005), proteins from cold-adapted organisms must possess structural alterations, especially in their enzymes, compared with their mesophilic counterparts. It seems that in terms of evolution there are many paths for the adaptation of psychrophilic enzymes and it has been the aim of many scientists to elucidate the features responsible for their unusual properties. The high

catalytic activity of psychrophilic enzymes in the cold environment, where the level of free energy is low, is generally explained by an increased flexibility of the molecules (Cummings & Black, 1999, D'Amico *et al.*, 2002, Gianese *et al.*, 2002).

THERMOPHILES

In contrast to cold-adapted microorganisms, thermophiles thrive in the hot environment and can be found in hot springs, deep see thermal vents or decaying wood. They can be generally classified, depending on the optimum growth temperature, into 3 groups: moderate thermophiles (50-60ºC), extreme thermophiles (60-80ºC) and hyperthermophiles (80-110ºC). At high temperature the fluidity of the cell membranes is increased. The organism's strategy to overcome this is to adjust the composition of the membrane's lipids, usually the ratio of the saturated to unsaturated lipids. In temperatures higher than 70ºC, DNA is also prone to denaturation and chemical modification. However, DNA of hyperthermophiles, such as *Pyrococcus furiosus* has been reported to be more stable *in vivo* than that of a mesophile such as *Esherichia coli* (Peak *et al.*, 1995). The strategy to increase the stability of DNA, protecting it from depurination and hydrolysis involves monovalent and divalent salts binding to the nucleic acids (Marguet & Forterre, 1998).

GENERAL ADAPTATION 'STRATEGIES' IN EXTREMOPHILIC ENZYMES

Most enzymes are active within a short range of a defined set of conditions. However, taking into consideration temperature extremophiles, enzymatic machinery of the host must be adapted to the conditions of the environment it lives in. Their enzymes are called extremozymes and can perform the same reactions as their mesophilic counterparts at a temperature that would inhibit or denature the less extreme forms.

PSYCHROPHILIC ENZYMES

The basic challenge for psychrophilic enzymes is low temperature. They must confront mainly two physical challenges: low thermal energy, because any decrease in temperature exponentially affects the rate of biochemical reactions, and high viscosity of water which increases by a factor higher than two when the temperature decreases from 37ºC to 0ºC. The higher viscosity negatively affects the rates of diffusion of such molecules as water, salts or enzyme's substrates (D'Amico *et al.*, 2006, Demchenko *et al.*, 1989, Siddiqui *et al.*, 2004). According to Arrhenius' equation the rate of the reaction falls two to four fold for every 10ºC decrease of temperature (Georlette *et al.*, 2004). It seems that the psychrophilic enzymes must have some structural adaptation features to maintain a high level of enzymatic activity in the cold. Thus, enzymes from psychrophiles hold the key to the answer of the question: *What makes an enzyme efficient*?

There are relatively few crystal structures of cold-adapted enzymes compared to mesophilic or thermophilic ones. At low temperature proteins become more rigid. It has been suggested, that an increase in structural flexibility contributes to maintaining activity in the cold environment (Aurilia *et al.*, 2009, D'Amico *et al.*, 2003, Feller & Gerday, 1997, Fields & Somero, 1998, Karan *et al.*, 2012, Margesin & Feller, 2010, Mavromatis *et al.*, 2003, Merlino *et al.*, 2010, Rasmussen *et al.*, 1992, Siddiqui & Cavicchioli, 2006, Somero, 1975). This increased flexibility might apply to the entire protein or just the part of the structure that is taking part in catalysis, and is probably responsible for the decreased stability of cold-adapted enzymes (Collins *et al.*, 2003, D'Amico *et al.*, 2003). Comparison studies of cold–active superoxide dismutases from *Pseudoalteromonas haloplanktis* and *Allivibrio salmonicida* and mesophilic counterparts revealed an increased flexibility in residues of the active site in the psychrophilic enzymes (Merlino *et al.*, 2010). The same situation was observed in adenylate kinases by Bae and Philips, who concluded that flexibility was essential for enzymes to be active at low temperature (Bae & Phillips, 2004). This additional flexibility could be achieved perhaps by amino acid substitution, by decreasing the number of hydrogen bonds within the molecule (Aghajari *et al.*, 1998, Michaux *et al.*, 2008), by an increased flexibility in the surroundings of the active site, reduced ion pair interactions and enhancing the surface charge distribution for substrate or cofactor binding (Kim *et al.*, 1999).

More information on low-temperature studies comes from bioinformatics. It has been reported that protein surface-solvent interactions play an important role in cold adaptation (Saelensminde *et al.*, 2009). Among other significant factors are amino acid composition and distribution that affect the overall protein charge, the hydrophobic core, the number of hydrogen bonds and salt bridges (Feller & Gerday, 2003, Georlette *et al.*, 2004, Russell, 2000, Zartler *et al.*, 2001).

It is worth noting that while cold-adapted enzymes show tendencies to higher flexibility and thus increased activity in a low temperature, there is an example of a psychrophilic enzyme that exhibits temperature-independent character - triosephosphatate isomerase from *Moritella marina* (Cavicchioli *et al.*, 2011). All this shows that there are many ways to achieve high catalytic rate by psychrophilic enzymes.

THERMOPHILIC ENZYMES

Living in high temperature, requires from an organism a means to protect its macromolecules from the heat and an adaptation of the molecules themselves to prevent heat denaturation and preserve their activities. Some thermophilic enzymes remain stable in temperatures as high as 113ºC (Pyrpassopoulos *et al.*, 2006), whereas most proteins denature irreversibly and lose activity above 40ºC.

Increasing temperature causes weakening of internal interactions in the molecule. At first, long range interactions between subunits are affected causing a disruption of the quaternary structure, if present. Once those interactions are loosened, the protein becomes more flexible and many hydrophobic groups are exposed to the surrounding solvent. The next step in the process of heat denaturation is breaking of hydrogen bonds in helices, making them easily accessible by water molecules which then interact with the newly exposed amino groups and carbonyl oxygen atoms of the peptide bonds. It further leads to an interruption of the hydrophobic core, such that the hydrophobic groups are no longer buried inside the protein but are exposed to the solvent. When water molecules flood the protein's internal hydrogen bonds, it unfolds. This can be detected as an increase of the radius of the molecule. The thermodynamically driven tendency for the protein to bury its

hydrophobic groups and expose its polar groups still operates at high temperature but at this stage it is likely to result in a structure different from native. Refolding of the protein is very unlikely and energetically unfavourable, although possible. Chitinase 40 from *Streptomyces thermoviolaceus* can be refolded with an almost 100% efficiency, which makes it very interesting for looking into high-temperature adaptation features (Pyrpassopoulos *et al.*, 2006).

From the wide range of studies and diverse results it is difficult to identify a general factor contributing to the increase of thermal stability of proteins from thermophiles. The first crystal structure of a thermophilic protein was thermolysin, reported in 1974 (Matthews *et al.*, 1974). Since that work, several investigators have focused on the structural bases of thermostable proteins. Among factors most often reported are:

- increased hydrophobicity of the protein's core (Haney *et al.*, 1997)
- increased compactness and shortening of surface loops (Russell *et al.*, 1997)
- decreased number of surface cavities and oligomerisation to bury the surface area (Salminen *et al.*, 1996)
- substitution of moderately hydrophobic amino acids in mesophilic compared to highly hydrophobic in thermophilic counterparts, with Met to Leu and Val to Ile being most common (Haney *et al.*, 1997, Russell *et al.*, 1998, Zuber, 1988)
- increased rigidity, due to an increased number of proline residues (Bogin *et al.*, 1998, Haney *et al.*, 1997, Watanabe *et al.*, 1997)
- decreased number of thermolabile residues in a molecule, such as Asn, Gln and Cys (Russell *et al.*, 1997)
- increased number of α-helices and polar residues on the surface (Haney *et al.*, 1997, Vogt & Argos, 1997, Vogt *et al.*, 1997)
- increased intramolecular interactions – hydrogen bonds and salt bridges (Elcock, 1998, Haney *et al.*, 1997, Kumar *et al.*, 2000, Russell *et al.*, 1997, Russell *et al.*, 1998, Vogt & Argos, 1997, Vogt *et al.*, 1997, Xiao & Honig, 1999, Yip *et al.*, 1998, Yip *et al.*, 1995).

However, proteins' stabilization 'strategies' might be different in different protein families. Some observed changes can be due to the phylogenetic distance between thermophilic and mesophilic organisms, unrelated to thermal adaptation of the protein. What is more,

describing a selected adaptation feature might not be straightforward when more than one factor could be at play and thus thermostability could be achieved by a variety of means.

GLYCOSIDE HYDROLASES

Organisms take advantage of a wide occurrence and variety of carbohydrates in nature, using oligosaccharides and polysaccharides in many biological processes as a store of easily accessible energy or as structural elements, such as cellulose and chitin, or for signalling (*sakchar*, meaning sugar or sweetness, derived from Greek). Thus, specific enzymatic machinery must be employed, by the cell to hydrolyse selectively glycosidic bonds, necessary in the processes of energy uptake, cell wall expansion and degradation, and turnover of signalling glycoproteins (Davies & Henrissat, 1995). Enzymes, that catalyse the hydrolysis of O-, N- and S-linked glycosides, are classified to glycoside hydrolases family (GH, EC 3.2.1.-). Enzyme nomenclature is based on substrate specificity and occasionally on their mechanism. Classification of GH in families is based on the amino acids sequence similarities (CAZY, http://www.cazy.org/ (Cantarel *et al.*, 2009). This classification system allows identifying the catalytic mechanism and key catalytic residues, as they are conserved in most GH families (Henrissat, 1991, Henrissat & Bairoch, 1996). However, GH might be classified within or outside the families in many different ways depending on the particular criterion.

Polysaccharide can be cleaved by GH at the end of the polymer (most frequently the non-reducing end) which distinguishes the enzyme as exo-acting, or it can be cleaved in the middle of the chain, which makes them endo-acting (Fig. 1).

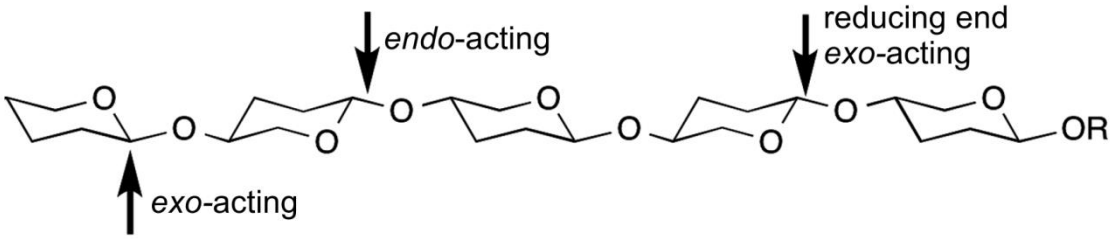


Figure 1. Different working modes of glycoside hydrolases. Cleavage sited of saccharide polymer chain are marked by arrows.

Hydrolysis of the glycosidic bond takes place *via* general catalysis that requires two critical residues: a proton donor and a nucleophile/base (Koshland, 1953, Sinnott, 1990). However several interesting variations on this mechanism have been reported.

The hydrolysis of glycoside with inversion of the anomeric configuration is generally achieved in one step, a single displacement mechanism involving an oxocarbenium ion-like transition state. The reaction typically occurs with general acid and general base assistance from two amino acids side chains, normally glutamic or aspartic acids located 6-11 Å apart (McCarter & Withers, 1994). One interacts with a water molecule, leading to the production of a hydroxyl ion, which launches a nucleophilic attack on the anomeric carbon of the sugar ring. The second residue acts as an acid–base catalyst and protonates the leaving group.

The hydrolysis with net retention of the anomeric configuration is most common and is achieved *via* a two-step double-displacement mechanism typically involving a covalent glycosyl-enzyme intermediate. Each step passes through the oxocarbenium ion-like transition state. Typically it requires two amino acids, an acid and a base separated by approximately 5.5 Å. At first, one residue, the nucleophile, attacks the anomeric centre displacing aglycon and forms the glycosyl-enzyme intermediate. The other residue functioning as an acid catalyst protonates the glycosidic oxygen as the bond cleaves. At the second stage the second residue acts as a base catalyst deprotonating a water molecule and then the hydrolysis of the intermediate takes place. After the resultant hydrolysis, the anomeric carbon is retained in its original configuration (Beguin & Aubert, 1994, Davies, 1998).

The situation is different in enzymes belonging to families 18, 20, 25, 56, 84 and 85 that hydrolyse substrates with N-acetyl (acetamido) or N-glycolyl group. There is no special amino acid designed to work as a nucleophile. Instead, the substrate's acetamido group itself is engaged in the nucleophilic attack. The formation of an oxazolinium ion intermediate, which is the result of this substrate-assisted mechanism, was deduced from crystal structures of chitinases with their substrates and natural inhibitors (Terwisscha van Scheltinga *et al.*, 1995).

GLYCOSIDE HYDROLASES FAMILY 18

Glycoside hydrolase family 18 (GH18) includes both catalytically active chitinases (EV3.2.1.14) and endo- β-N-acetylglucosaminidases (EC 3.2.1.96), and also non-enzymatic proteins with a function of carbohydrate binding modules, 'lectin'or xylanase inhibitors. Representatives of this family are found among eukaryotes – animals, plants and fungi, prokaryotes and viruses. In chitin-containing organisms, chitinases play a role during morphogenesis and cell division. Plant and human chitinases serve as defence factors against fungal pathogens. Prokaryotes and fungi contain chitinolytic enzymes to convert chitin into an energy source and as a supply of carbon and nitrogen.

Although similarity in primary sequence between chitinases in GH18 family could be relatively low (<15% identity) all those enzymatically active chitinases possess the canonical DxDxE catalytic motif and share a common structural fold of the catalytic domain which consists of a β/α-barrel framework made of eight parallel β-strands in the core surrounded by eight α-helices. It is worth noting that chitinases found in higher plants and some from the Gram-positive *Streptomyces* bacteria are classified as GH19. Their catalytic domains have a lysozyme-like fold with a shallow substrate-binding groove which, contrary to GH18, is not rich in aromatic residues.

Family 18 chitinases hydrolyse all forms of chitin and chitosan, whose degrees of acetylation might be as low as 13% (Sorbotten *et al.*, 2005), in different working manners. GH18 chitinases bind the substrate in an extended substrate recognition site. The sugar moieties at the non-reducing end, counting from the scissile glycosidic bond, are given

negative numbers (glycon side) and those on the reducing side are given positive numbers (aglycon side). Some chitinases work in a processive manner, being attached to the substrate and after hydrolysis sliding along the carbohydrate chain, or non-processive endo-acting way. Most of the chitinases produce chitobiose (dimer of β-1, 4-linked glucosamine units, NAG_2) and do not cleave short oligomers, such as NAG_3 or NAG_4.

GH18 enzymatic catalysis retains the configuration on the anomeric carbon atom in a double displacement reaction mechanism. As mention before, acetamido group of the substrate is a nucleophile (substrate-assisted reaction or anchimeric assistance) whereas the enzyme provides the general acid/base which protonates the glycosidic oxygen to facilitate the departure of the aglycon and deprotonates a water molecule for the hydrolysis of the oxazolinium ion. The active residue is Glu from the conserved DxDxE motif (Fig. 2). It has been reported that in addition to the active Glu, the neighbouring Asp residues are also important for catalytic activity. Acetamido group generates an oxazolinium ion intermediate which subsequently is hydrolysed or in some cases transglycosylated.

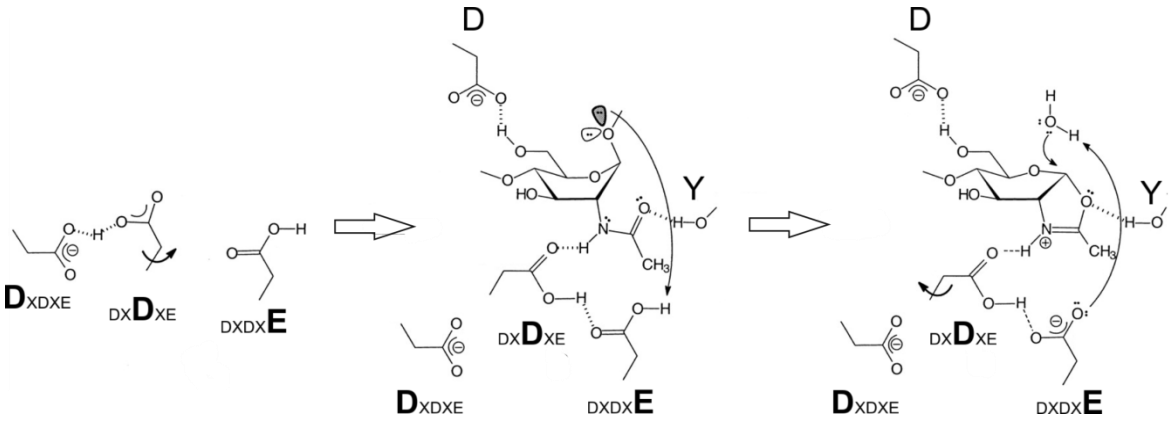


Figure 2. Catalytic mechanism for glycoside hydrolases family 18. DxDxE, conserved in most family 18 chitinases, are shown in successive stages of the catalysis, as proposed by Tews et al. (Tews *et al.*, 1997) and based on structural studies of chitinase B (van Aalten *et al.*, 2001). From left: resting enzyme. Second D is too far away to interact with E. In the middle: binding of NAG substrate in the −1 sub-site causes distortion of the sugar ring to a boat or skewed boat conformation and second D interacts with E via a hydrogen bond. Right: hydrolysis of the oxazolinium ion intermediate leads to protonation of E and rotation of second D back to its original position.

The substrate-assisted mechanism has consequences for the inhibition process of the enzyme. The naturally occurring pseudotrisaccharide consisting of two N-cetylglucosamines linked to allosamizoline moiety (allosamidin, Fig. 3) is a potent inhibitor of family 18 chitinases (Sakuda *et al.*, 1987).

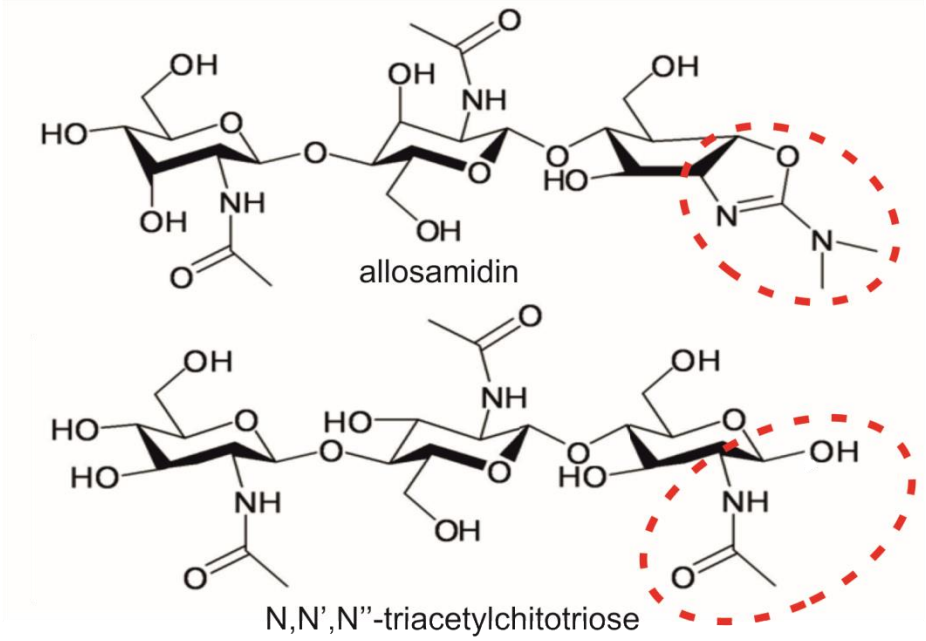


Figure 3. Structures of allosamidin and N,N',N''-triacetylchitotriose. Atoms essential for the inhibition in the allosamidin and for the substrate-assisted creation of the reaction intermediate in NAG_3 are marked in dashed ovals (red).

CHITIN

Chitin is an highly insoluble polymer of β-1,4-linked N-acetyl glucosamine and the second most abundant organic compounds in nature, after cellulose. Each NAG moiety is rotated 180 degrees relative to its neighbours along the chain. The structure of chitin may be compared to cellulose, but with additional hydrogen bonds, due to the acetamido groups, linking adjacent chains. Chitin is found in three forms (Kurita, 2001):

- α-chitin – NAG chains are arranged antiparallel allowing chitin to create strong hydrogen bonds; this form is the most abundant, most stable and highly insoluble

- β-chitin – NAG chains are arranged parallel and spaced further apart which makes the H-bonds weaker
- γ-chitin – mixed composition of α- and β-chitin

Chitin is found throughout the taxonomic kingdoms as the main component of the cell walls of fungi and major green algae, exoskeletons of arthropods such as crustaceans (crabs, lobsters, shrimps) and insects, the radulae of molluscs, and the beaks and internal shells of cephalopods (squid, octopi). In the aquatic biosphere alone, more than 10^{11} tons are produced annually (Keyhani & Roseman, 1999). All this chitin is recycled as a source of carbon, nitrogen and energy mainly by microorganisms exhibiting chitinolytic activity. Both marine and land microorganisms possess the ability to: find chitin, adhere to the chitinaceous substrate, select a single polymer and cleave it to oligosaccharides, transport it to the cell and finally catabolise it to the final product of fructose-6-P, acetate and ammonia.

STRUCTURES OF CHITINASES

Many members of GH18 display multi-domain architecture with various arrangements of the domains (Carbohydrate Active Enzymes database, CAZy)(Cantarel *et al.*, 2009). The common feature is a catalytic domain with the fold of a β/α-barrel (also known as TIM barrel), whose function is to hydrolyse the β-1,4 linkage between N-acetyl-D-glucosamine (NAG) residues in chitin. Additionally, many catalytic domains contain an α+β domain or extended loops inserted in the TIM barrel, which participate in the formation of the substrate-binding groove, making it deeper or even tunnel-like. Characteristic for chitinolytic enzymes are carbohydrate binding modules (CBMs) designed to bind chitin (ChBD, chitin-binding domains) (Akagi *et al.*, 2006). They are found in plant, fungal and bacterial proteins and can be structurally diverse. In addition to those, many chitinases also contain fibronectin type III (FnIII) or immunoglobulin like (Ig-like) domains (Toratani *et al.*, 2006). One or more of them can be found between the catalytic domain and ChBD. They are postulated to serve as spacers to adjust the position of ChBD

relative to the catalytic domain (Toratani *et al.*, 2006). However, the exact role of linking domains is still unclear.

Although many multi-domain chitinases have been identified in gene sequences, crystallographic studies are few, presumably because the molecules are flexible and difficult to crystallise. *Serratia marcescens* has been described as the most effective chitinolytic bacterium with its enzymatic machinery containing at least four chitinases and chitin-binding protein (Monreal & Reese, 1969). The chitinolytic machinery of *S. marcescens* is of interest because to date it is the most studied and best characterised. Crystal structures of many complexes and mutants have been solved for chitinase A (Perrakis *et al.*, 1994), chitinase B (van Aalten *et al.*, 2000) and the catalytic domain for chitinase C (Payne *et al.*, 2012) as well as a chitobiase. All structures provide insight into the overall structures, architecture of the active site and the enzymatic mechanism.

CHITINASE A

The first solved crystal structure of a chitinase is chitinase A (ChiA) from *S. marcescens* (Perrakis *et al.*, 1994). The 563-residue chitinase precursor is secreted from the cell. This is accompanied by the cleavage of the signal peptide. The resulting enzyme has 540 amino acid residues and the molecular mass of 58.5kDa. The structure is modular, which is very common for enzymes working on insoluble substrates, such as cellulose or chitin (Fig. 4).

The structure of ChiA was a landmark in research on family 18 chitinases. It demonstrated that the catalytic domain is a β/α-barrel and revealed spatial arrangement of residues in the active site. The enzyme has a deep substrate-binding groove formed in part by the α+β domain which extends to the N-terminal domain *via* a patch of aromatic residues on the surface. This N-terminal domain is classified as a FNIII and is believed to be engaged in chitin binding. Up to now, 15 models of ChiA have been deposited in the Protein Data Bank (PDB) as miscellaneous variants, mutants and complexes with substrates and inhibitors. Based on the structural data and kinetics it was reported that ChiA hydrolyses chitin from the reducing end in an exo-processive manner. After substrate hydrolysis, the dimeric product is required to dissociate from sub-sites +1 and +2 (aglycon

side). Three aromatic residues have been identified to be engaged in processivity: Trp167, Trp275 and Phe396 (respectively -3, +1 and +2 sub-sites).

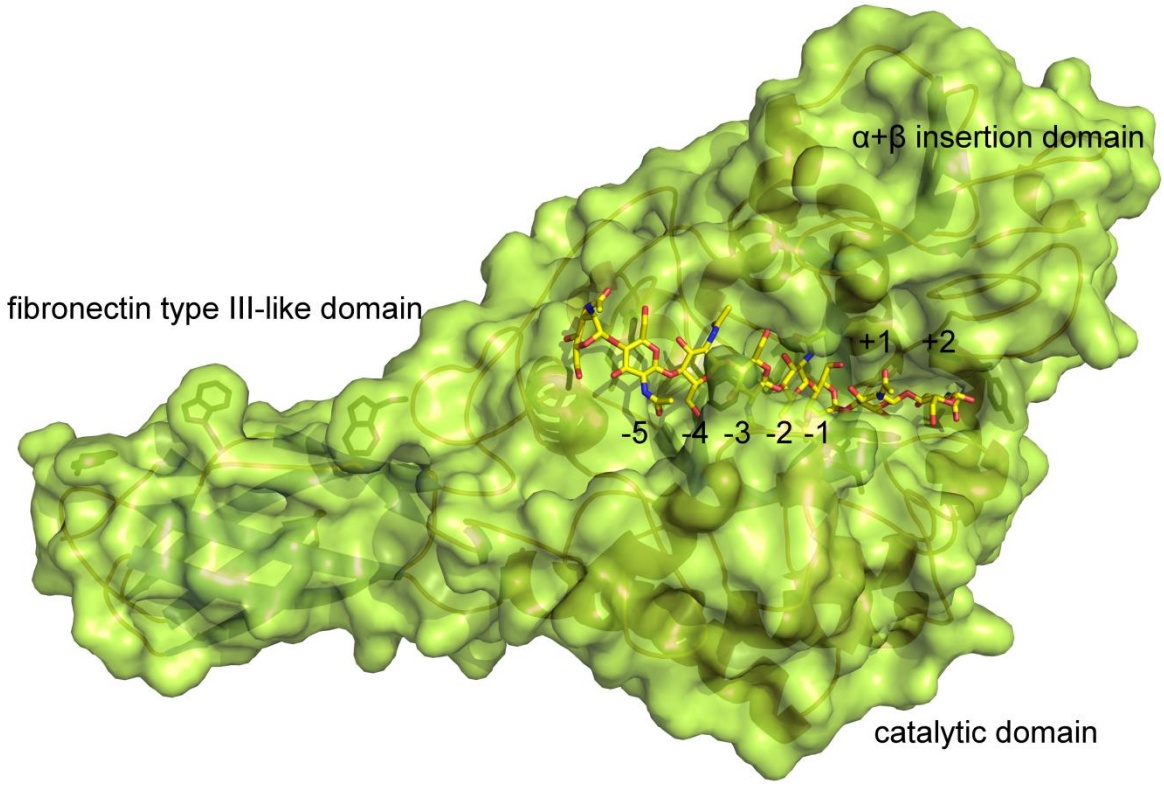


Figure 4. Structure of chitinase A molecule (PDB code 1ehn)(Papanikolau et al., 2001). Semi-transparent surface representation with the scheme of the secondary structure. Aromatic residues responsible for chitin binding are marked with sticks. Eight NAG moieties span the substrate-binding groove. The substrate binding sub-sites are numbered from +2 to -5. This and the other figures of structures have been prepared with the program PyMOL (DeLano, 2002).

CHITINASE B

The chitinase B (ChiB) from *S. marcescens* contains 498 residues with a calculated molecular weight of 55.4kDa. Similar to ChiA, ChiB also has a modular structure.

There are more than 23 crystal structures of ChiB available in PDB, including mutants, complexes with inhibitors and an inactive form in complex with substrates. The main difference from ChiA is the presence of an additional, to the catalytic β/α-barrel, chitin-binding domain, on the opposite side of the catalytic centre. The substrate binding site is 'tunnel-like' permitting tight interactions with the polymeric substrate (Fig. 5). The binding groove is extended, from the aglycon (+) site to the chitin binding domain with its

patch of aromatic residues that is up to 55Å from the active residues. It corresponds to the length of ten NAG moieties. The architecture of the active site and the substrate binding site shows that the chitinase cleaves the substrate from the non-reducing end of the sugar chain by a combined chitobiosidase and chitotriosidase activity and the product reaction dissociates from the glycon (-) site.

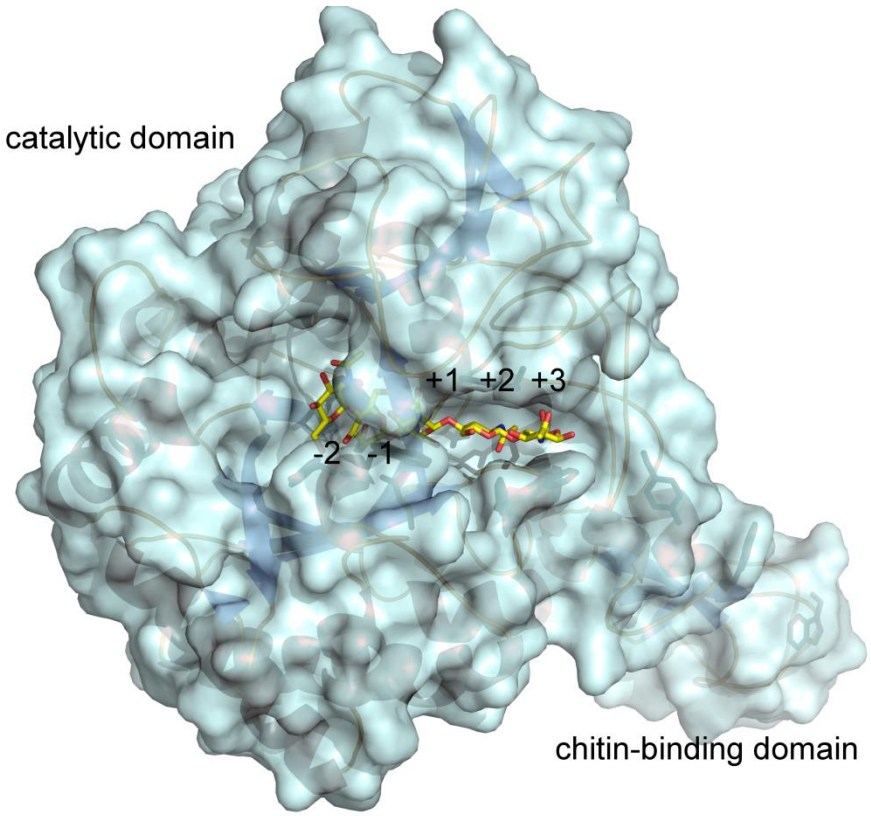


Figure 5. Structure of the monomer of chitinase B (PDB code 1e6n)(van Aalten *et al.*, 2001).Semi-transparent surface representation with secoundary structure indicated. Aromatic residues responsible for the chitin binding are marked with sticks. Five NAG moieties span the substrate binding groove. Binding sub-sites are numbert from +3 to -2.

OTHER CHITINASES

Apart from the ChiA and ChiB from *S.marcescens*, chitinase C (ChiC) from *S.marcescens* has been solved (PDB code 4axn, Fig. 6A)(Payne *et al.*, 2012). Compared to ChiA and ChiB this chitinase is composed only of a catalytic domain lacking several loops and the α+β domain, which results in a very shallow substrate-binding cleft.

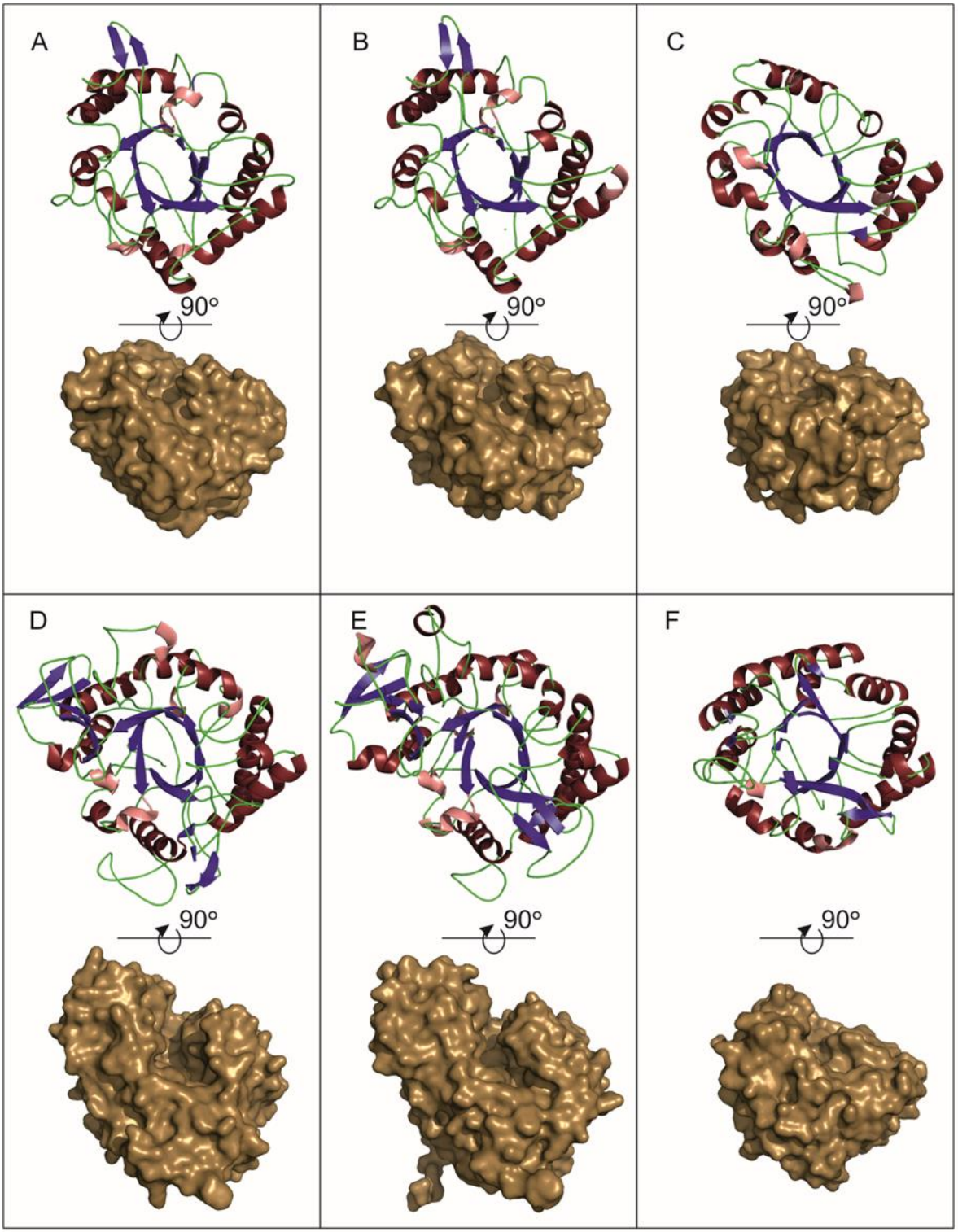


Figure 6 Ribbon representation of the secondary structures of different chitinases and their surface representation with a visible substrate binding groove (in the middle of the top surface of each structure). PDB codes: A. 4axn B. 3ian C. 1llo D. 1kfw E. 1itx F. 2dsk

Remarkable is the extended aglycon substrate-binding site by means of a small β-hairpin subdomain with two solvent-exposed tryptophan residues. Similar features are found in the chitinase from *Lactococcus lactis* (PDB code 3ian, Fig. 6B)(Bonnano *et al*., unpublished). Shallow substrate binding site characterises also the chitinase ChiNCTU2 from *Bacillus cereus* (Hsieh *et al.*, 2010)(PDB code 3n11), *Streptomyces coelicolor* (Vigdorovich et al., unpublished, PDB code 3ebv), chitinase A1 from *Aspergillus fumigatus* (Rush *et al.*, 2010)(PDB code 2xvn) and a plant chitinase/lysozyme (hevamine) from *Hevea brasiliensis* (Terwisscha van Scheltinga *et al.*, 1996)(PDB code 1llo, Fig. 6C).

Most chitinases whose crystallographic structures have been solved are derived from mesophilic organisms. However, the structure of the catalytic domain of *Arthrobacter* TAD20 psychrophilic chitinase is known (Ayati et al., unpublished; PDB code 1kfw, Fig. 6D). From the other end of the temperature scale, one structure of a catalytic domain is available from hyperthermophilic archaeon *Pyrococcus furiosus* (Nakamura *et al.*, 2007)(PDB code 2dsk, Fig. 6F).

It is worth noting that a multi-domain structure of chitinase from *Bacillus circulans* WL-12 (Toratani *et al.*, 2006) has been deduced by a combination of techniques. It revealed a very elongated arrangement of domains (by SAXS): catalytic β/α-barrel at one end (X-ray; PDB code 1itx, Fig. 6E) and the chitin-binding domain at the other end (NMR), and a tandem of two FNIII domains in between (Toratani *et al.*, 2006).

MORITELLA MARINA AND CHITINASE 60

Moritella marina (*Vibrio marinus*) is an obligatory psychrophilic marine bacterium. It was isolated from a sample raised from the depth of 1200 meters in the northern Pacific Ocean, where the temperature was measured to be 3.24 C° (Morita & Haight, 1964), and further taxonomically characterized (Colwell & Morita, 1964).

Moritella marina is a Gram-negative facultative anaerobe in the form of curved or straight rods and is motile owing to polar flagella. In addition to low-temperature

adaptation it is also resistant to high salinity and pressure. The bacterium in the laboratory conditions has an optimum temperature growth between 15 and 16 C°.

Recently the draft genome of the sequence of the *Moritella marina* was publishes and partially annotated with a G+C content of 40.5% (Kautharapu & Jarboe, 2012)

The bacteria were reported to secrete chitinase in response to chitin induction. The chitinase gene consists of 1650 nucleotides and encodes a protein of 550 amino acids with a calculated molecular weight of ~60kDa, named *Mm*Chi6. The gene was inserted into the plasmid pET-11a and *E. coli* cells were transformed (Stefanidi & Vorgias, 2008). The expressed and purified protein was characterized biochemically and showed a relatively high level of catalysis at low temperatures. DSC analysis has shown melting temperature (T_m) – 56.4°C ±0.5°C which is close to T_m of another psychrophilic chitinase from *Arthrobacter* sp. – 54.3°C.

CHITINASE 40 FROM *STREPTOMYCES THERMOVIOLACEUS*

Streptomycetes are Gram-positive bacteria found in soil, characterised by high GC content in the genome and are able to grow in a wide range of environmental conditions. By secreting a variety of enzymes to degrade naturally occurring macromolecules they possess an ability to digest chitin, cellulose, xylan and starch (McCarthy & Williams, 1992).

Streptomyces thermoviolaceus is a thermophilic bacterium isolated from decayed wood. It is reported to grow on xylan but also grows actively on chitin as a source of carbon and nitrogen, but does not have a cellulase activity. The chitinolytic system of *S. thermoviolaceus* is composed of four chitinases: Chi40, Chi30 belonging to GH18 and Chi30, Chi25 belonging to GH19, and two N-acetyl-β-glucosamidases: NagA and NagB.

The gene for Chi40 was cloned into many vectors in order to find a suitable vector/host strain combination for expression and subsequent purification of correctly folded protein. The G+C content of the gene encoding the protein is about 70%. The

optimal temperature for the activity of the enzyme was determined to be in the range 60-65 °C and the optimum pH for the activity was 6.0. Chi 40 was reported to refold in almost 100% after heat denaturation as the first chitinase containing the TIM barrel catalytic domain (Pyrpassopoulos *et al.*, 2006).

MATERIALS AND METHODS

BASIC MATERIALS

Table 1 Materials used in the experiments

Category	Material	Producer	Cat. number
SDS-polyacrylamide gel electrophoresis	Acrylamide	Molekula	35672700
	N,N'-Methylene-bis-acrylamide	Molekula	22797959
	Glycine	Sigma-Adrich	22609174
	Tris-HCl	BioShop	TRS001.1
	SDS	Sigma-Adrich	L-4509
	TEMED	Sigma-Adrich	T-7024
	APS	Sigma-Adrich	A-3679
	β-mercaptoethanol	BioShop	MER002
	Glycerol	BioShop	GLY001
	Brilliant Blue R	Sigma-Adrich	B-0149
	PageRuler Plus	Thermo Scientific	26691
	Protein Molecular Weight Markers	Promega	V8491
	Methanol	Chempur	
	Glacial acetic acid	Chempur	
	Ready Gel Tris-HCl 4-20%	Bio-Rad	161-1123
Bacteria culturing media	Agar	BioShop	AGR001.1
	Ampicillin	BioShop	AMP222
	Kanamycin	BioShop	KAN201
	Bio-Tryptone	BioShop	TRP402.205
	Sodium chloride	Chempur	117941206
	Yeast extract	BioShop	YEX401.205
Protein chromatography	HiPrep 16/60 Sephacryl 100	GE Healthcare	17-1165-01
	HiLoad 16/60 Superdex 200	GE Healthcare	17-1069-01
	Phenyl-Sepharose 6 Fast Flow	GE Healthcare	17-0973-03

Table 1 Continuation

Category	Material	Producer	Cat. number
	Q-sepharose Fast Flow	Sigma-Aldrich	Q1126
Crystallisation screens	PACT premier HT-96	Molecular Dimensions	MD1-36
	JCSG-plus HT-96	Molecular Dimensions	MD1-40
	Morpheus HT-96	Molecular Dimensions	MD1-46
	PGA Screen HT-96	Molecular Dimensions	MD1-51
	Midas HT-96	Molecular Dimensions	MD1-59
	Grid Screen Ammonium Sulphate	Hampton Research	HR2-211
	Grid Screen Sodium Malonate	Hampton Research	HR2-247
	Grid Screen PEG 6000	Hampton Research	HR2-213
	Grid Screen MPD	Hampton Research	HR2-215
Ligands	N,N'-Diacetylchitobiose	Carbosynth	OD00769
	N,N',N''-Triacetylchitotriose	Carbosynth	OT06497
	N,N',N'',N'''-Tetraacetylchitotetraose	Carbosynth	OT04211
	N,N',N'',N''',N''''-Pentaacetylchitopentaose	Carbosynth	OP04834
	N,N',N'',N''',N'''',N'''''-Hexaacetylchitohexaose	Carbosynth	OH07433
	PNP-GlcNAc	Carbosynth	EN04911
Buffers ingredients	Sodium phosphate Monobasic	BioShop	SPM400
	Sodium phosphate Dibasic	BioShop	SPD600
	Sodium hydroxide	Chempur	118109252
	Hydrochloric acid	Chempur	115752837
	EDTA	BioShop	EDT001
	Imidazole	Molekula	M10820769
	Sucrose	POCH	772090110
	Triton X-100	Sigma-Aldrich	T-9284

Table 1 Continuation

Category	Material	Producer	Cat. number
Protein handling	Amicon Ultra Centrifugal Filters Ultracel -30K	Millipore	UFC903024 UFC803024
	Amicon Ultra Centrifugal Filters Ultracel -10K	Millipore	UFC901024 UFC801024
	Millex-GV Syringe-driven Filter Unit	Millipore	SLGV013SL
	PureLink® PCR Purification Kit	Invitrogen	K3100-01
	Ultrafree-MC Centrifugal Filter Device	Millipore	UFC30VV25
DNA electrophoresis	Agarose	Sigma-Aldrich	A9539
	Midori green	Nippon Genetics	MG02
	Ethydium bromide	BioShop	ETB444
	GeneRuler 1 kb DNA Ladder	Thermo Scientific	SM0311
	GeneRuler 100 bp Plus DNA Ladder	Thermo Scientific	SM0321
	Bromophenol blue	Sigma-Aldrich	B-8026
	Xylene cyanol	Sigma-Aldrich	X4126
Other	Oligonucleotides	Genomed	
	Z-Competent E. coli Transformation Kit	Zymo Research	T3001
	GeneMATRIX Plasmid Miniprep DNA Purification Kit	EURx	E3500

Expression plasmids:

- pET-TOPO-151D
- pET11a
- pMCSG7 (Midwest Center for Structural Genomics, Argonne, IL, USA)

Bacteria strains:

- BL21(DE3)
- DH5α
- BL21Magic (Midwest Center for Structural Genomics, Argonne, IL, USA)

Molecular biology enzymes:

- Hot start Kod polymerase
- T4 polymerase (exonuclease)
- DpnI restriction enzyme
- TEV protease

CULTURE MEDIA

Table 2 Culture media used for growing bacteria

Liquid LB Medium (Luria-Bertani Medium)		Solid LB Agar	
Tryptone	10 g	**Tryptone**	10 g
Yeast extract	5 g	**Yeast**	5 g
NaCl	10 g	**NaCl**	10 g
		Agar	15 g

Deionised water was added and shaken until solutes have dissolved; the volume was adjusted to 1000 ml. The LB-agar was warmed up in the microwave oven until the agar dissolved. In the case of immediate use, there was no need for autoclaving. LB-agar medium was allowed to cool to 50-60ºC before adding the thermolabile substances. Antibiotics were added, depending on the cell resistance: 1ml ampicillin to the final concentration of 100 μg/ml or/and 1ml kanamycin to 50 μg/ml. Liquid agar medium was poured on Ø 90mm Petri plates directly from the flask under the laminar flow cabinet and left to cool and harden.

Table 3 Super optimal broth with catabolite repression (SOC)

SOC media	
Tryptone	20 g
Yeast extract	5 g
NaCl	0.5 g
KCl	2.5 mM
Glucose	20 mM

Deionised water was added and shaken until solutes have dissolved and the volume was adjusted to 1000 ml. SOK medium was kept in 1ml aliquots at -18 ºC until used.

Antibiotics stock solutions:

Ampicillin – 100 mg/ml

Kanamycin – 50 mg/ml

Solid state antibiotics were dissolved in the MiliQ water. Then, solutions were kept at -18ºC to avoid thermal degradation.

SOLUTIONS FOR CELL BREAKING AND PROTEIN PURIFICATION.

All buffers were prepared using chemicals in the solid state or from previously prepared stock solutions. The given values are for preparing 1000 ml of the buffers. After shaking until solutes have dissolved the pH was adjusted using HCl or NaOH and water was added to the final volume of 1000 ml in a measuring cylinder. In the phosphate buffers used during protein chromatography the ratio of Na_2HPO_4 and NaH_2PO_4 was calculated from the Henderson–Hasselbalch equation to get a buffer at pH 8.0 (Equation 1).

$$pH = pK_\alpha + \log_{10}\left(\frac{[A^-]}{[HA]}\right)$$

Equation 1. Henderson–Hasselbalch equation. [HA] is the molar concentration of the undissociated weak acid; [A⁻] is the molar concentration of this acid's conjugate base. $pK_\alpha = 6.86$ **at 25ºC**

Stock solutions for preparing buffers:

- Sodium Phosphate Monobasic: 0.2 M
- Sodium Phosphate Dibasic: 0.2 M
- Tris HCl pH 8.0: 1 M
- Imidazole: 1 M
- TCEP: 0.5 M
- EDTA: 1 M
- NaCl: 5 M

Buffers

Table 4 Cell lysis buffer.

Lysis buffer $P_{20}EDTA_1TX\text{-}100_{0.5\%}$ pH 8.0		Final concentration
Sodium Phosphate Monobasic: 0.2M	6.8 ml	20 m M
Sodium Phosphate Dibasic: 0.2M	93.2 ml	
EDTA	1 ml	1 mM
Triton X-100	5 ml	0.5%

The buffer was stored in the cold room (4 ºC), and used cold.

Table 5 Cell lysis buffer for His-tagged proteins.

HisTrap lysis and binding pH 8.0		Final concentration
Tris-HCl	50 ml	50 mM
NaCl	100 ml	500 mM
Imidazole	20 ml	20 mM
TCEP	2 ml	1mM

The buffer was stored in the cold room (4 ºC), and used cold.

Table 6 Elution buffer for His-tagged proteins.

HisTrap elution buffer pH 8.0		Final concentration
Tris-HCl	50 ml	50 mM
NaCl	100 ml	500 mM
Imidazole	300 ml	300 mM
TCEP	2 ml	1mM

The buffer was stored in the cold room (4 ºC), and used cold.

Table 7 Osmotic shock buffer for periplasmic proteins

Osmotic shock buffer $Tris_{30}EDTA_3S_{20\%}$pH 8.0		Final concentration
Tris-HCl	30 ml	30 mM
EDTA	3 ml	3 mM
Sucrose	200 g	20 % w/v

The buffer was always prepared fresh and used at room temperature.

Table 8 Binding buffer for hydrophobic interaction using phenyl-sepharose

$P_{20}E_1AS_{1M}$pH 8.0		Final concentration
Sodium Phosphate Monobasic: 0.2M	6.8 ml	20 mM
Sodium Phosphate Dibasic: 0.2M	93.2 ml	
EDTA	1 ml	1 mM
Ammonium sulphate	132.14 g	1 M

The buffer was stored in the cold room (4 ºC), and used cold.

Table 9 No salt buffer for elution from phenyl sepharose and binding with ion exchange - Q-sepharose

Phenyl-Sepharose elution buffer $P_{20}E_1$ pH 8.0 Q-Sepharose binding buffer $P_{20}E_1$ pH 8.0		Final concentration
Sodium Phosphate Monobasic: 0.2M	6.8 ml	20 mM
Sodium Phosphate Dibasic: 0.2M	93.2 ml	
EDTA	1 ml	1 mM

The buffer was stored in the cold room (4 ºC), and used cold.

Table 10 Q-Sepharose elution buffer

$P_{20}E_1NaCl_{1M}$ pH 8.0		Final concentration
Sodium Phosphate Monobasic: 0.2M	6.8 ml	20 mM
Sodium Phosphate Dibasic: 0.2M	93.2 ml	
EDTA	1 ml	1 mM
NaCl	200 ml	1 M

The buffer was stored in the cold room (4 ºC), and used cold.

Table 11 Size exclusion buffer (SEC I) without reducing agent

$Tris_{20}E_1NaCl_{200}$ pH 8.0		Final concentration
Tris-HCl	20 ml	20 mM
EDTA	1 ml	1 mM
NaCl	40 ml	200 mM

The buffer was filtered through 0.2 micrometre filter and stored in the cold room (4 ºC)

Table 12 Size exclusion buffer (SEC II) with reducing agent - TCEP

$Tris_{50}NaCl_{200}TCEP_1$ pH 8.0		Final concentration
Tris-HCl	50 ml	50 mM
NaCl	40 ml	200 mM
TCEP	2 ml	1 mM

The buffer was filtered through 0.2 micrometre filter and stored in the cold room (4 ºC)

Table 13 Dialysis buffer

$Tris_{50}NaCl_{200}TCEP_1$ pH 8.0		Final concentration
Tris-HCl	50 ml	50 mM
NaCl	200 ml	500 mM
TCEP	2 ml	1 mM

ELECTROPHORESIS

Table 14 DNA electrophoresis buffer 10x TBE

TBE10x pH ~8.3		Working solution TBE
Tris base	108 g	45 mM tris-borate
Boric acid	55 g	
EDTA	20 ml of 1 M	1 mM EDTA

The stock solution was passed through a 0.45 micrometre filter to prevent or delay the formation of precipitates and then stored at room temperature and diluted before use: 100 ml of 10X to one litre with deionised water.

Table 15 DNA electrophoresis gel-loading buffer 4X

Agarose Gel-loading buffer 4X	
Bromophenol blue	0.17%
Xylen cyanol FF	0.17%
Glycerol	20%

Buffer was stored at -18ºC, used in the 1:3 ratio to DNA sample.

SDS-PAGE

- Stock solutions:
- SDS: 10% (w/v)
- Tris-HCl: 1M pH 6.8
- Tris-HCl: 1.5M pH 8.8
- APS: 10% (w/v)

Table 16 Acrylamide mix stock sollution 30%

Acrylamide mix stock sollution 30%	
Acrylamide	29% (w/v)
N,N'-methylene-bis acrylamide	1% (w/v)

The acrylamide mixture was stored in a dark bottle in the cold room.

Table 17 Denaturing protein electrophoresis running buffer

Tris-glycine electrophoresis buffer	
Tris base	25 mM
Glycine	250 mM
SDS	0.1%

Table 18 Denaturing protein electrophoresis sample loading buffer

SDS gel-loading buffer 4X	
Tris –HCl (pH 6.8)	200 mM
β-mercaptoethanol	400 mM
SDS	8% (w/v)
Bromophenol blue	0.4 %
Glycerol	40% (v/v)

Buffer was stored at -18ºC and used in the 1:3 ratio to protein sample

Table 19 Denaturing protein electrophoresis separating gels: 12% and 15%

Solutions for preparing resolving SDS-PAGE gel (10 ml)		
Gel percentage	12 % gel	15 % gel
H_2O	3.3 ml	2.3 ml
30% acrylamide mix	4 ml	4 ml
1.5M Tris pH 8.8	2.5 ml	2.5 ml
10 % SDS	0.1 ml	0.1 ml
10 % APS	0.1 ml	0.1 ml
TEMED	0.004 ml	0.004 ml

Table 20 Denaturing protein electrophoresis stacking gel

5% stacking SDS-PAGE gel (3 ml)	
H_2O	2.1 ml
30% acrylamide mix	0.5 ml
1.0M Tris pH 6.8	0.38 ml
10 % SDS	0.03 ml
10 % APS	0.03 ml
TEMED	0.003 ml

Table 21 Staining and distaining buffers for SDS- polyacrylamide gels.

	Staining buffer	Distaining buffer
Methanol	500 ml	300 ml
Glacial acetic acid	100 ml	100 ml
Coomassie Brilliant Blue R-250	0.1 % (w/v)	

Table 22 Standard cloning PCR master mix

PCR master mix for cloning – 50 µl		Final concentration
10X PCR buffer	5 µl	1X
$MgSO_4$ 25 mM	3 µl	1.5 mM
Betain 5 M	10 µl	1 M
dNTPs 2 mM each	5 µl	0.2 mM
F primer 10 µM	1.5 µl	0.3 µM
R primer10 µM	1.5 µl	0.3 µM
KOD polymerase 1U/µl	1 µl	
Template	1 µl	
H_2O	22 µl	

Table 23 T4 exonuclease reaction mix

T4 reaction - 15 µl		Final concentration
NEB2 buffer	1.5 µl	1x
dGTP/dCTP	0.75 µl	5 mM
DTT 40mM	1.5 µl	4 mM
T4 polimerase	0.15 µl	1U
BSA	0.15 µl	1x
Template	1 µl	
H_2O	9.95 µl	

BIOCHEMICAL METHODS:

DNA SEQUENCING, GENE MANIPULATION AND CLONING

In order to know the exact gene sequence and thus the primary structure of the encoded protein of interest, or to create a genetically manipulated protein, the plasmid from the previously transformed *E.coli* cells had to be isolated.

Cultures for mini-preps

Bacteria containing the desired plasmids were cultured in 12 ml tubes in 2 ml of liquid LB-media overnight at 37 ºC, in a shaker at 220 rpm. Antibiotics were added to the media depending on the resistance of the host cells: *Mm*Chi60 with ampicillin, while other mutants of *Mm*Chi60 (E153Q, D2, D12) and Chi40 with both ampicillin and kanamycin. The LB-medium was inoculated from the -80ºC stock by scraping the frozen surface of the culture with a sterile tip, or from agar plates by touching the bacterial colonies with a sterile tip and transferring immediately to a test tube. Mini-prep cultures were centrifuged in 2 ml Eppendorf tubes at 14.000 rpm for 2 minutes. The supernatant was discarded, while the cell pellet was used immediately for the plasmid isolation or it was frozen at -18ºC for future handling.

Plasmid isolation

Plasmids were isolated using the Plasmid Miniprep DNA Purification Kit according to the protocol delivered with the kit. Kits for plasmid preparation are based on the alkaline lysis method invented by Birboim and Doly (Birnboim & Doly, 1979). The Miniprep method generally relies on a column alterable affinity for DNA which depends on the buffer that is washing through. The first step consists of suspending the cell pellet in EDTA- and RNaseA-containing buffer that chelates divalent metal cations which are cofactors of DNAses and disrupt the cells membranes and digest RNA. Then strong alkaline lysis buffer (containing SDS and NaOH) is added to break the cells and denature fully the proteins and the chromosomal DNA while the plasmid DNA is only partially denatured. After adding acetate-containing buffer, all cell debris along with the chromosomal DNA is precipitated, while small plasmid DNA can renature and stay in solution. After centrifugation, both

phases can be separated and the supernatant can be loaded on the affinity column and washed from contaminations. Changing to a weak buffer or water causes the plasmid DNA to be eluted from the column. The pure plasmid was used for sequencing, transforming the expression cells or as a template in mutational PCR.

Measuring plasmid concentration

Concentrations of the plasmid DNA and the purified product of PCR were measured using a NanoDrop, placing 1-2µl of the mini-prepped DNA onto the pedestal. Previously, spectrophotometer was calibrated. The base line of the instrument was set with a DNA elution buffer and 1-2 µl of DNA sample were applied onto the pedestal. The measurement of the concentration of the DNA was based on absorbance at 260 nm while the purity of the sample by the 260/280 nm absorbance ratio (nucleic acids *versus* proteins).

Mutagenesis and cloning

Competent cells

For the preparation of chemically competent *E.coli,* cells were treated with a buffer that contained $CaCl_2$ to disrupt the cell membranes. Calcium chloride ions neutralise the repulsion between the negatively charged phospholipid heads and lipopolysaccharides of the cell membrane and the negatively charged phosphate groups on the DNA. Subsequent heat shock creates a thermal gradient which, in turn, creates a tunnel leading into the cell, allowing extra-chromosomal DNA (such as plasmids) to enter the cell.

Competent cells were prepared using the Z-Competent *E.coli* Transformation Kit according to the protocol supplied with the kit. Briefly, 0.5 ml of fresh overnight *E. coli* in LB culture was inoculated in 50 ml of SOB in 500 ml flask and shaken vigorously at 18ºC for about 30 hours until OD_{600} reached 0.4 to 0.6. The culture was placed on ice for 10 minutes and pelleted at 3000rpm for 6 minutes at 4ºC. The whole procedure included a sequence of mixing with a wash buffer, centrifuging, mixing with a competent buffer, and was performed in the cold. Aliquots of 0.2 ml cells in 1.5 ml Eppendorf tubes were immediately frozen in liquid nitrogen and stored at -80ºC.

Competent cells yielding a high number of plasmid copies used in cloning and for sequencing purpose are called cloning strains (DH5α). Expression strains (BL21Magic and BL21(DE3)) were used to express the protein efficiently from the derived construct.

Mutagenesis using the PIPE method

*Mm*Chi60 deletion mutants D2 (E423_A504del) and D12 (M348_A504del) were prepared using the PIPE method which takes advantage on the observation that, contrary to the common assumption, normal PCR amplification results is a mixture of products, which are not fully double stranded (Fig. 7). The 5' ends of the reaction product (which are exactly the same fragments as delivered by the synthetic primers) can be left unpaired in the final round of PCR. Therefore, simply by changing the primer sequence of those ends, mutation can be promoted in a simple PCR reaction without additional steps (Klock & Lesley, 2009).

The first 12-18 bases on the 5' ends of the primers were designed to be directionally complementary to the second primer's annealing site to the template, so that the PCR product could anneal as desired and become functional upon transformation.

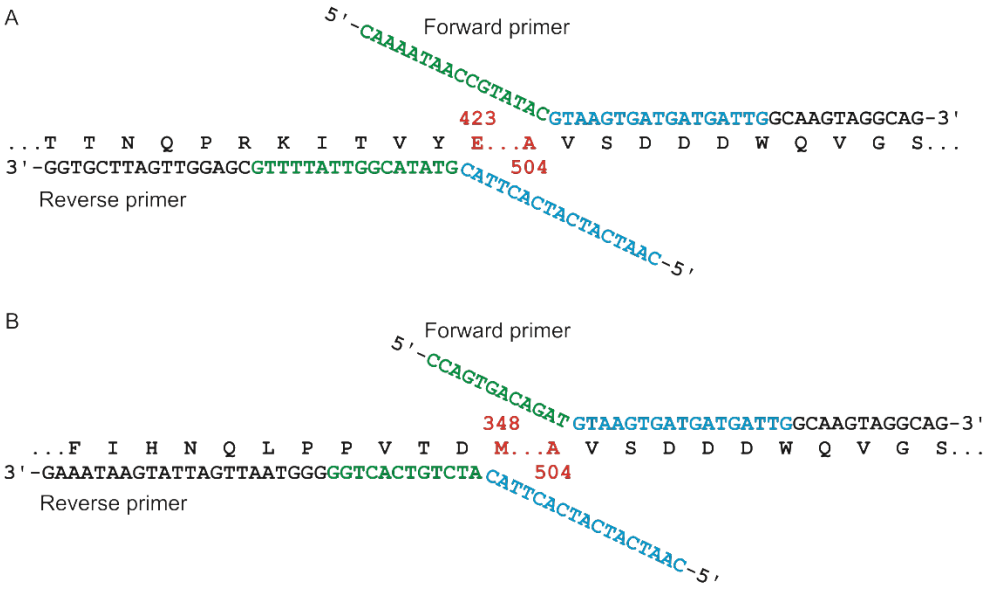


Figure 7 PIPE mutagenesis. Primers designed for delitions E423_A504del (A) and M348_A504del (B) presented together with the corresponding fragment of *Mm*Chi60. Complementary fragments of primers are coloured green and blue. The deletion fragment is in red showing the range of the deletion, with the numbering of amino acids consistent with the native protein.

Both primers were designed to flank the deletion part of the enzyme to amplify the plasmid except for the part to be deleted. The PCR reactions were performed using the primers:

*Mm*Chi60_E423_A504del

forward: 5'CAAAATAACCGTATACGTAAGTGATGATGATTGGCAAGTAGGCAG3',

reverse:

5'CAATCATCATCACTTACGTATACGGTTATTTTGCGAGGTTGATTCGTGG3'

*Mm*Chi60_M348_A504del

forward:5'CCAGTGACAGATGTAAGTGATGATGATTGGCAAGTAGGCAG3',

reverse:

5'CAATCATCATCACTTACATCTGTCACTGGGGGTAATTGATTATGAATAAAG3'

in which the underlined fragment is complementary to the second primer's binding fragment.

The PCR reaction components were the same as described in the PCR master mix for cloning and the reaction parameters are specified below:

Table 24 PCR reaction parameters for the PIPE cloning of deletion mutants of *Mm*Chi60

	Temperature [°C]	Time [s]
1. Polymerase activation	95	120
2. Denaturing	95	20
3. Annealing	61	10
4. Extension	70	270
Steps 2 to 4 were repeated 26 times.		

To eliminate the transformation background, parental DNA was digested by DpnI according to the protocol.

Side directed mutagenesis using two single-primer reactions in parallel

For site-directed mutagenesis (SDM) reactions Single-Primer Reactions in Parallel (SPRINP) was used to prevent the problem with primers pairing and in consequence tandem repeats of primer in the cloned DNA (Edelheit *et al.*, 2009). The active site mutant

of *Mm*Chi60 was produced by introducing E153Q mutation by site directed mutagenesis, using the forward primer 5'CGATATTGATTTACAACAAGCAGCGATCACAGC3' and the reverse 5'GCTGTGATCGCTGCTTGTTGTAAATCAATATCG3' in which the mutated codon GAA to CAA is marked by underline. The protocol includes steps:

1. Amplifying the plasmid DNA in two separate primer extension reactions with the reaction mixture and thermocycler parameters specified below:

Table 25 Master Mix for side directed mutagenesis using two single-primer reactions in parallel

Reaction components		
	Reaction I	**Reaction II**
10X PCR buffer	5 µl	5 µl
$MgSO_4$ 25 mM	22 µl	22 µl
Betain 5 M	10 µl	10 µl
dNTPs 2 mM each	5 µl	5 µl
F primer 10 µM	1.5 µl	-
R primer10 µM	-	1.5 µl
KOD polymerase 1U/µl	1 µl	1 µl
Template	~200 ng	~200 ng
H_2O	Up to 50 µl	Up to 50 µl

Table 26 PCR reaction parameters for the site-directed mutagenesis of Chi60

Thermocycler reaction parameters		
	Temperature [ºC]	Time [s]
1. Polymerase activation	95	120
2. Denaturing	95	20
3. Annealing	60	10
4. Extension	70	270
Steps 2 to 4 were repeated 35 times.		

2. Combining both products in one tube to denature and reanneal the complementary strands by slow cooling.

Table 27 Reannealing conditions for combined two complementary strands amplified in separate PCR reactions.

Denaturation and slow cooling conditions	
Temperature [ºC]	Time [min]
95	5
90	1
80	1
70	0.5
60	0.5
50	0.5
40	0.5
37	Holding

3. DpnI digestion of the parental template DNA according to the protocol described further.

Mutagenesis using His-tag vector

Chi40 N-terminal deletions were performed and several truncations of the gene were cloned into pMCSG7 (Fig. 8) and pMCSG48 vectors designed for ligase-independent cloning (LIC). The vectors were kindly provided by Midwest Center for Structural Genomics, Argonne, IL, USA. Four mutants were created with fusion with the His6-tag that has affinity for nickel resins.

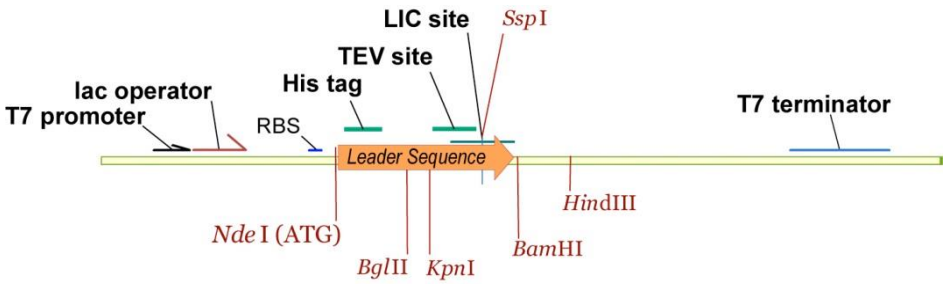


Figure 8. pMCSG7 vector map

For the vector amplification, PCR was performed using the primers: forward 5'ATTGGAAGTGGATAACGGATCCGAATT3' and reverse 5'ATTGGATTGGAAGTACAGGTTCTCGGT in a standard master mix for cloning and the reaction parameters specified below:

	Temperature	Time [s]
1. Polymerase activation	95	120
2. Denaturing	95	20
3. Annealing	60	10
4. Extension	70	270
Steps 2 to 4 were repeated 35 times.		

Linear products of PCR reactions were purified with PureLink PCR Purification Kit and further used in a ligase-independent cloning.

For the insert amplification, PCR was performed using primers designed for ligase-independent cloning into the pMCSG7 and pMCSG48 vectors using the web site server: http://bioinformatics.anl.gov/cgi-bin/tools/primer_design.pl
reverse
Chi40M_allR5'TTATCCACTTCCAATGTTATCACGCCAGGCCGTTGCTCA3',
forward
Chi40M29F5'TACTTCCAATCCAATGCCACCGTCGAGACCCGCGC3',
Chi40M32F5'TACTTCCAATCCAATGCCACCCGCGCCGCCGCC3',
Chi40M36F5'TACTTCCAATCCAATGCCGCCGACAACGGCACGGTCAA3',
Chi40M39F5'TACTTCCAATCCAATGCCGGCACGGTCAAGCTCGGCTA3'
in a standard master mix for cloning and the reaction parameters specified below:

	Temperature	Time [s]
1. Polymerase activation	95	120
2. Denaturing	95	20
3. Annealing	60	10
4. Extension	70	270
Steps 2 to 4 were repeated 35 times.		

All four PCR products were purified with PureLink PCR Purification Kit and further used in a ligase-independent cloning.

PCR product purification

For further DNA manipulation and performing LIC, the PCR product must be purified. For the purpose of removing primers, dNTPs, enzymes, salts, and short-failed PCR products PureLink PCR Purification Kit was used according to the protocol supplied with the kit.

The used method relies on selective binding of dsDNA to silica-based membrane in the presence of chaotropic salts. All impurities were washed away while the DNA was bound to the membrane. For elution, a low salt buffer or water were used.

Ligase-Independent Cloning (LIC)

The LIC method is a fast and relatively cheap way to produce expression constructs. It is based on homology regions present in the primers used for amplification of the insert and the ends of linearized cloning vector. Advantage is taken of the specific exonuclease activity of the T4 DNA polymerase. It digests 3' ends of DNA until it reaches the nucleotide that is present in the reaction mixture. Therefore, the digestion reaction of the vector is performed in the presence of a nucleotide complementary to one that is in the insert digestion reaction. In this way, complementary single-stranded DNA overhangs are created and are able to anneal with each other (Fig. 9). Thus the annealed insert with the vector can be directly transferred to *E. coli* cells with no additional treatment with ligases.

The purified products of the PCR: the linear form of the pMCSG7 and pMCSG48 vectors and different mutants of Chi40 were separately treated with the T4 reaction buffer. The vector reaction was supplemented with dGTP while dCTP was added to the insert reaction. Both were incubated in PCR tubes for 40 minutes at 22ºC and then for 20 minutes at 75 ºC (for the polymerase thermal inactivation).

LIC reaction was performed by mixing the previously T4 treated: 5µl of the vector with 15 µl of the insert. After 5 min. incubation at room temperature 25mM EDTA was added and incubated for 5 min. at room temperature. The product was ready to be used for transforming the *E. coli* cells.

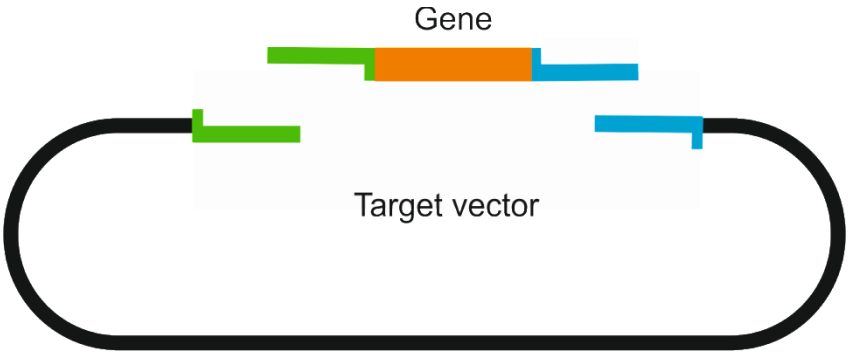


Figure 9. The picture presents an insert and a vector with complementary overhangs after treatment with T4 polymerase

DpnI digestion

DpnI is a restriction enzyme that recognises methylated, and hemi-methylated sites of the DNA and thus cleaves cell-derived DNA (Fig. 10).

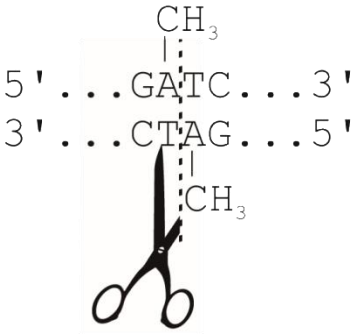


Figure 10. DpnI mode of action

Amplification reactions for creating mutants of *Mm*Chi60 were performed on the cell derived plasmids that still can effectively transform bacterial cells after PCR. Therefore in order to get rid of any transformation background, before the transformation, DNA was incubated overnight at 37 ºC with the DpnI restriction enzyme (0.5-1 µl per each 20 µl of DNA). Inactivation of the DpnI was performed before the transformation by 20 min. incubation at 80ºC.

Agarose gel

For visualisation and assessment of the size of the PCR products, agarose gel electrophoresis was performed. The direction of migration of the DNA is from the negative to the positive electrode and is due to the naturally occurring negative charge carried by their sugar-phosphate backbone.

A 1% gel was used for inserts while longer products and plasmids were analysed using a 0.6% gel. The gels were prepared by mixing an appropriate amount of molecular grade agarose (1g per 100ml for 1%, 0.6g per 100ml for 0.6%) with 1X TBE buffer and heating in the microwave oven until the agarose dissolved. Upon cooling to 50ºC, ethydium bromide or Midori Green was added for UV illumination. Gel was poured and samples, previously mixed with a loading buffer, were transferred to the gel pockets. Molecular markers were also used to assess the size of migrating DNA.

Transformation

Prepared plasmids were used to transform competent *E. coli* cells (Fig. 11). In the beginning the cells were taken out of -80°C and put on ice to thaw slowly. Meanwhile the SOC medium was placed in a thermo block to reach 37 °C. The DNA after cloning was put on ice together with the thawing cells to reach the same temperature. 5 to 10 µl of DNA were transferred to tubes containing the cells and mixed by flicking. The tubes were put again on ice for 30 minutes. After this time a heat shock was performed at 42°C for 90 seconds and the cells were transferred again on ice where they stayed for 5 minutes. Warm SOC medium was added to bacteria and the tubes were transferred to a shaking thermo block at 37°C for 45 minutes. The agar plates were inoculated and cultured overnight at 37°C.

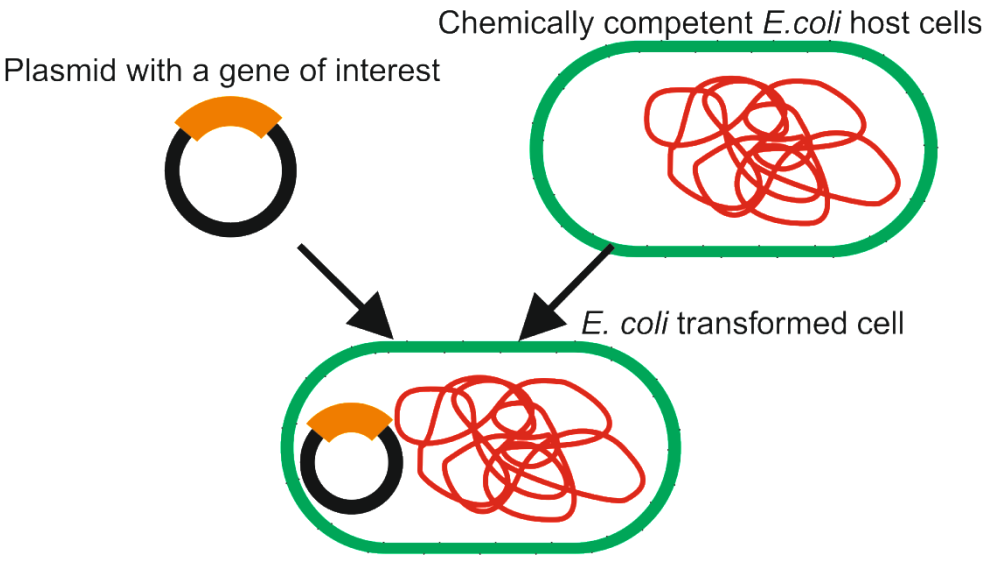


Figure 11. *E.coli* transformation scheme.

Storage of bacterial cultures – glycerol stocks

Long-term storage of bacteria was achieved using glycerol stocks. Bacteria were grown in the same way as cultures for plasmid isolation. The LB medium was inoculated from an overnight culture or from a Petri dish by touching the bacterial colony with a sterile tip and transferring it to the tube. Subsequently, 0.5 ml of the overnight culture was transferred to the Eppendorf tube and 0.5 ml of sterile 50% glycerol was added. The glycerol was dispersed evenly by vortexing, frozen immediately in liquid nitrogen and transferred to -80ºC for long-term storage. To recover, the frozen surface of the culture was scratched by a sterile tip and transferred either to liquid media or streaked onto the surface of an agar plate.

LARGE SCALE PROTEIN PRODUCTION

Proteins of interest were overexpressed using *E.coli* transformed with a plasmid carrying isopropyl-β-D-thiogalactoside (IPTG) inducible promoters. This promoter is derived from the *E.coli* cells in which it is a part of the *lac* operon. Natively, under this promoter there are three genes that are engaged in lactose digestion: β-galactosidase, lactose permease and thiogalactosidase transacetylase which are triggered by the availability of the lactose substrate. The lactose promoter in vectors is located upstream the cloned gene so the expression of the protein of interest depends on the presence of an inducer. The inducer, which naturally is allolactose binds to the repressor allowing the transcription of genes located downstream the promoter. Unlike allolactose, the IPTG is not hydrolysable, therefore after inducing the genes, they are expressed constantly.

Overnight pre-cultures were grown in 150 ml flasks with 50 ml of liquid LB medium with appropriate antibiotics at 37ºC in a shaker. 10-20 ml of overnight culture was transferred to 1 l of freshly prepared LB medium in a 2 l flask or in a 2 l coca-cola plastic bottle. New cultures were grown in a shaker at 37ºC and 220 rpm, until the cells reached mid-log growth (OD_{600}~1). From this step two different protocols were used:

1. In the case of *Mm*Chi60 and its mutants, as well as Chi40 truncation mutants, the temperature was decreased to 18 ºC and after cooling, IPTG was added to a 1mM

final concentration. The bacteria were grown overnight at 18 ºC in a shaker. The bacterial culture media were centrifuged at 8000 rpm for 10 min. and supernatant was discarded while the cells were kept for further processing.

2. In the case of secreted protein (Chi40) IPTG (1mM final concentration) was added directly at 37 ºC to the bacterial cultures and shaken for further 3-5 hours. The bacterial culture media were centrifuged at 10000 rpm for 20 min. The supernatant with the secreted protein was used in a further purification procedure, so were the pelleted cells.

Cell lysis

Pelleted cells were suspended in a lysis buffer in 50 ml falcon tubes, placed in an ice bath and disrupted by sonication. Bursts of a total duration of 4 min. with appropriate intervals for cooling were used. Cell debris was removed by centrifugation at 15 000 rpm for 30 min. at 4°C. The supernatant was further exposed to ammonium sulphate for precipitation in the case of the wild type *Mm*Chi60. Chi40 truncation mutants were applied on a column packed with HisTrap HP resin and equilibrated with a binding buffer.

Osmotic shock

In gram-negative bacteria, expressed proteins, which have a special secretion signal peptide, are transported to the periplasmic space, which is located between the biological membranes. Secretory production of recombinant proteins has several advantages compared with cytosolic proteins. The N-terminal end of the protein is cleaved during the secretion which makes the protein similar to the natural gene product. The activity of proteases is much lower in the periplasm than in the cytoplasm and there are fewer proteins located between the membranes which makes them easier to purify. A more oxidative environment and chaperone proteins facilitating the formation of disulphide bonds give the protein a better chance for correct folding.

The osmotic shock procedure was implemented to disrupt only the outer membrane of the cells without breaking the inner membrane and thus release only the content of the periplasm. The pelleted cells were treated by osmotic shock according to the published procedure with some alternations (Poole & Hancock, 1984). *E. coli* cells were suspended

by pipetting in and out using 10 ml tips and mixed for 10 min. in osmotic buffer and spun down at 6,000 rpm. Cells were re-suspended in ice-cold water by in and out pipetting using 10 ml tips and the periplasmic proteins were released to the solution by mixing the cold solution for 10 minutes. Cells were centrifuged at 10,000 rpm. The supernatant was adjusted to 1M ammonium sulphate and 20mM Na-phosphate buffer (pH 8.0) and directly applied on a 10 ml Phenyl-Sepharose 6 Fast Flow column (previously equilibrated in the Phenyl-sepharose washing and binding buffer).

Supernatant treatment

The supernatant with the secreted Chi40 was kept in the cold room and was adjusted to a 20mM phosphate buffer and 1mM EDTA using stock solutions. Solid ammonium sulphate was added in small portions, with constant mixing with a magnetic stirrer to a 1 M final concentration. Then the pH of the supernatant was adjusted to 8.0 using diluted NaOH and the supernatant was applied on 10 ml Phenyl-Sepharose 6 Fast Flow column (previously equilibrated in the Phenyl-sepharose washing and binding buffer).

Ammonium sulphate precipitation

Solubility of proteins varies according to the salt concentration in the buffer solution and the presence of hydrophobic regions on the surface of the protein. In the low concentration of salt we can observe an increased solubility of the proteins and this process is called salting-in. However, by increasing the salt concentration and thus the ionic strength the protein solubility decreases up to the point of precipitation. The theoretical basis of salting-out is complex, but probably a competition between the protein and salt ions for available water molecules is at play. As mentioned before, different proteins have different hydrophobic regions on the surface, what means that it is possible to separate proteins from the mixture by changing the concentration of ammonium sulphate.

Solid ammonium sulphate was added in small portions to the supernatant containing a soluble mixture of proteins after bacterial cell lysis, up to the final saturation of 40% (22.6 g/100ml) and stirred at 4ºC for 20 min. It was centrifuged for 20 min. at14.000 rpm and the volume of the supernatant was measured. Ammonium sulphate was added to the supernatant to the final saturation of 60% (12.0 g/100ml) and stirred the same way as before. After centrifuging for 20 min. (14.000 rpm) the pellet with *Mm*Chi60 was kept and

solubilised using the Phenyl-Sepharose binding buffer and applied on a 10 ml Phenyl-Sepharose 6 Fast Flow column (previously equilibrated with the same buffer).

Q-sepharose ion exchange chromatography

Ion exchange chromatography is the most common method of purification of charged molecules. Two types of ion exchange chromatography are based on the same principle: in cation exchange positively charged molecules binds to negatively charged media; in anion exchange the media are positively charged, so negatively charged molecules can bind. The binding of the molecule to the media is based on ionic interaction and the strength is determined by the number and location of charges on the molecule. By increasing the salt concentration, molecules with a lower affinity start to elute from the column, whereas stronger bound molecules need more salt to be released from the column.

The column was packed with 8 ml of anion exchange Q-sepharose medium and equilibrated with a binding buffer. Then a diluted sample of *Mm*Chi60 (in the same buffer) was applied on the column. Subsequently, the column was washed with 50 ml of the binding buffer and 100 ml of a gradient of linearly increasing salt concentration was applied while 5 ml fractions were collected (50 ml of binding buffer/50 ml of elution buffer). Fractions were tested on SDS-PAGE and the most pure and concentrated were pooled for further use.

Phenyl-Sepharose hydrophobic interaction chromatography (HIC)

Hydrophobic amino-acids are usually buried in the protein core. However, many biomolecules have sufficiently exposed hydrophobic groups to be captured by hydrophobic moieties on the chromatography media. The interaction is enhanced by buffers with high ionic strength. It is an excellent technique to be used after ammonium sulphate precipitation and for capturing proteins from large volumes of solvent.

The column was packed with 8 ml of phenyl-sepharose medium and equilibrated with the binding buffer. Protein samples in the same buffers were applied on the column. Subsequently the column was washed with 50 ml of the binding buffer. In the case of the *Mm*Chi60 wild-type enzyme, a 100 ml gradient of linearly decreasing ammonium sulphate concentration was applied while 5 ml fractions were collected (50 ml of binding buffer/50

ml of elution buffer). Fractions were tested on SDS-PAGE chromatography and the most pure and concentrated were pooled.

In other cases: for Chi40, both the supernatant and the periplasmic protein, and the *Mm*Chi60 deletion and low-activity mutants, phenyl-sepharose was used for binding the protein which was then eluted in one step by 40 ml of elution buffer. The eluted solution was concentrated and applied on the SEC column before crystallisation.

His-tag proteins purification

His Trap media are designed for a simple, one-step purification of histidine-tagged proteins. The resins have immobilised nickel ions which have a high affinity to the histidine tail attached to the protein of interest (IMAC – ion metal affinity chromatography).

The supernatant of the lysate was applied on the column previously packed with 6 ml of the HisTrap HP resin, connected to the VacMan (Vacuum Manifold, Promega), and the chromatography process was accelerated with a vacuum pump. After protein binding the column was washed 4 times by 30 ml of the binding buffer. The bound protein was eluted using 15 ml of elution buffer and the eluted liquid was supplemented with TEV protease and loaded into dialysis bags. During the dialysis conducted overnight in a dialysis buffer at 4 ºC the His6-tag and MBP were cleaved and the excess of imidazole was removed. Subsequently, the dialysed solution was mixed with HisTrap HP resin to bind cleaved His_6-tag, His_6-tag with MBP and His_6-tagged TEV protease. The flow-through containing the purified protein was collected and concentrated.

Size exclusion chromatography (SEC)

Proteins can be separated by their size and shape using size exclusion chromatography. The column is made of porous gel beads. Small molecules move slowly entering many pores and are eluted from the column slowly, whereas the largest molecules enter few pores and move very fast, and elute first.

The media used include dextran (HiLoad 200Superdex 16/60 column) and dextran-polyacrylamide (HiPrep 100 Sephacryl 16/60). Size exclusion chromatography was performed on the AKTA Prime FPLC system. Columns were equilibrated using the size exclusion buffer and samples before applying were concentrated and filtered on 0.4

micrometer filters. In the case of Sephacryl a 1.5 ml sample was injected and the flow rate 0.5 ml/min was used, while in the case of Sephadex 1-3 ml samples were injected and 1 ml/min flow rate was applied. 2ml fractions were collected and the fractions of the highest purity and concentration were pooled and concentrated.

Protein concentration

Protein samples were concentrated before being applied on the size exclusion column and before crystallisation trials. Amicon UItra 4 ml and 15 ml centrifugal filter units with a 10000 and 30000MW cut-off were used. Prior to use, the cellulose membrane of the Amicon filters was moistened with the buffer in which the protein was suspended. The concentrator was spun in the fixed angle rotor at 4ºC and at force not exceeding 5000 g. The concentrated protein was collected from the filter reservoir by pipetting and the ultra-filtrate was discarded.

SDS-POLYACRYLAMIDE GEL ELECTROPHORESIS OF PROTEINS

SDS PAGE electrophoresis was carried out in a discontinuous buffer system (Laemmli, 1970) using 12 %, 15% and 4-20 % gels depending on the size of the protein of interest. Before loading samples onto the gel, the protein was mixed with an SDS gel loading buffer in a 3:1 ratio. In this process a strong ionic detergent (SDS), reducing agent (β-mercaptoethanol) and incubation of the sample at 95ºC for 5 min. made the protein dissociate into individual denatured amino acid chains. Negatively charged SDS binds to the polypeptide proportionally to its length. Therefore the separation of proteins on the gels depends on the molecular weight because the speed of migration through the gel is proportional to the charge. The concentration of the polyacrylamide used to cast the gel also impacts on the speed of protein migration. 10µl of samples were transferred to gel pockets, and 5 µl of molecular weight marker. By using PageRuler Plus molecular weight marker it was possible to assess the size of the separated proteins. Until the samples reached the resolving gel, the electrophoresis was conducted using 80 V and subsequently 160 V, until the bromophenol blue reached the bottom of the gel. Following that, the gel was washed in water and transferred to the staining solution, where it was mixed gently for

30 min. and then it was transferred to distaining solution until stained bands of the protein were visible.

ABSORBANCE ASSAY (280NM)

The concentration of proteins was measured in the NanoDrop spectrophotometer based on the absorption of near UV by proteins' Tyr, Trp, Phe residues and disulphide bonds. The absorbance at 280nm and an estimated extinction coefficient was used to calculate the concentration. The extinction coefficient of the protein was estimated on the basis of the amino acid sequence using the Expasy's ProtParam tool (Gasteiger *et al.*, 2005).

PRINCIPLES OF PROTEIN X-RAY CRYSTALLOGRAPHY

PROTEIN PREPARATION

Crystallisation requires a protein of a high quality and in large quantity to ensure a reasonable success rate. The two features that define the quality of the protein sample crucial for crystallisation to be successful are: purity and homogeneity. Different proteins behave in different ways, but generally more than 90% purity is required for starting crystallisation trials. The purity and homogeneity can be achieved by appropriate methods of protein purification and handling. The state of the protein sample can be evaluated by biochemical and physical methods like SDS-PAGE (purity), size exclusion chromatography (to observe different folding forms or aggregation) or DLS (polydispersity indicates inhomogeneity in solution).

CRYSTALLISATION

Screening for initial protein crystallisation conditions

In order to obtain protein crystals the molecules must assemble into a periodic three-dimensional lattice. The process of crystal forming is controlled by the laws of physical chemistry and thermodynamics. In terms of thermodynamics crystallisation of biological macromolecules is not very different from crystallisation of simple salts. One difference is that protein crystals are held together by weak intermolecular interactions, such as hydrogen bonds, salt bridges and hydrophobic interactions, and contain large volumes filled with solvent. Matthews (Matthews, 1968) observed that the solvent content can range from 27% to 65%. These factors explain the fragility of protein crystals.

Successful crystallisation usually requires a large number of experiments exploring various conditions, called initial screening. It usually starts with the protein solution at the concentration of approximately 10 mg/ml mixed with compounds which reduce the solubility of the protein. This mixture can be brought to supersaturation by the vapour diffusion method in which the concentration of the precipitant and the protein is gradually

increased. The slow concentration results from the fact that the initial concentration of precipitant in the protein/precipitant mixture is lower than in the reservoir solution, therefore water evaporates from the drop into the reservoir increasing the concentration of both the protein and the precipitant. At supersaturation the protein could precipitate or crystallise, if under the right conditions nucleation sites are formed (Fig. 12). The general problem in protein crystallisation is its empirical character. It is practically impossible to predict crystallisation conditions in advance. The important variables include: type of precipitant, pH, ionic strength, protein to screen solution ratio, temperature, protein concentration, the presence of various salts, ligands and additives and the crystallisation method.

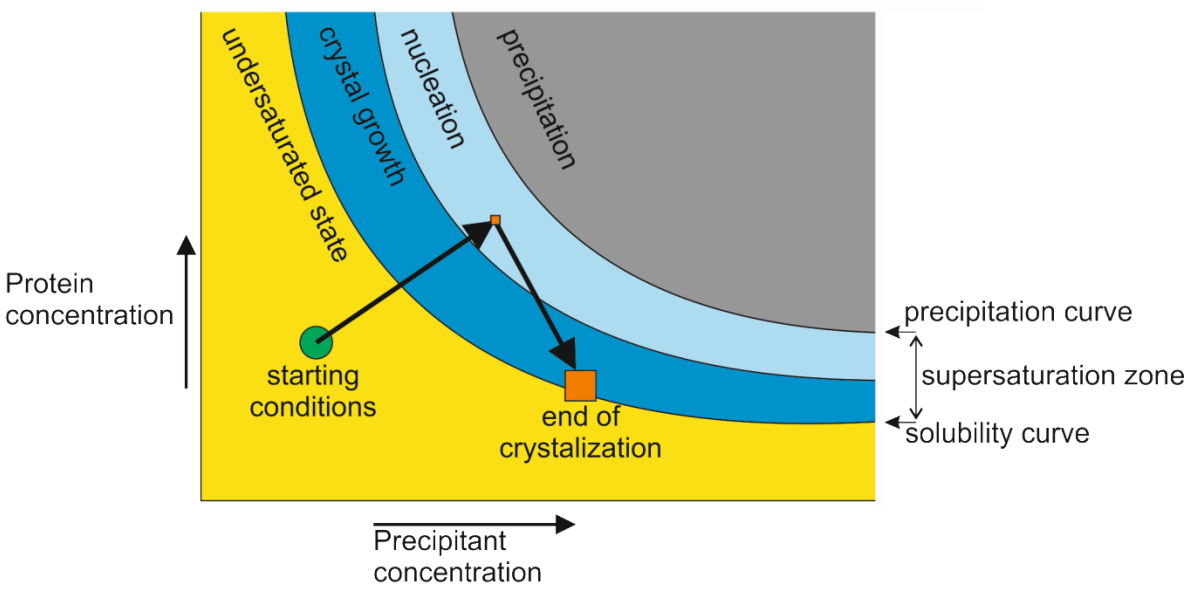


Figure 12. The diagram illustrates the principles of vapour diffusion. The arrows show the path of successful crystallisation. The protein and precipitant mixture concentrates and reaches the nucleation region. Once it is reached crystallisation nuclei form. Crystal growth causes a decrease of the protein concentration in the solution and the growth eventually stops.

The initial screening in this work was facilitated by the Crystal Gryphon (Art Robbins Instruments) crystallisation robot and a number of ready-to-use crystallisation solutions. The standard crystallisation plate used in the experiments (Intelli-Plate 96) consists of 96 reservoir wells located next to three wells for the protein .The robotic method

is designed for sitting drop vapour diffusion and placing in a drop as little as 100 nl of protein mixed with the appropriate volume of screen solution, automatically. The plate with three crystallisation wells per reservoir well is ideal for optimizing protein concentration, additive screens, drop ratio and combinatorial experiments.

The optimization by hanging drop method

Crystals can grow within hours, days, weeks or even, in extreme situations, months. In order to improve the crystal shape, size and diffraction properties, initial conditions form the robot plates could be optimized using standard Hampton Research 24-well plate. Precipitant concentration, pH, protein concentration and the presence of additives are systematically varied in order to improve crystallisation. Drops are prepared on a siliconized glass cover slip by mixing 0.5-3 microliters of protein solution with the same volume of precipitant solution. The slip is placed upside down over a precipitant solution in the well (0.5 ml). The chamber is sealed with a silicone oil or silicone grease applied on the circumference.

The drop is slowly equilibrated against a larger reservoir of solution containing the precipitant or another dehydrating agent.

Post-crystallisation treatment

Once the crystal has grown, before mounting it for data collection, in many cases it needs a post-crystallisation treatment. This is performed to achieve two goals: changing the 'mother liquor' to a cryo-protectant and, if required, to obtain a protein-ligand complex.

Most X-ray crystallographic data collection is done at a low temperature, typically 100 K, to minimize crystal damage caused by free radicals generated by the X-ray beam. A soaking solution which, when frozen prevents ice formation is known as a 'cryo-protectant'. In practice it is an agent such as glycerol, ethylene glycol, low molecular weight PEG, sugars or MPD. The crystal is soaked in this solution for as short as a few seconds up to 24 hours prior to freezing in liquid nitrogen. Some crystallisation solutions already contain cryo-protectants and crystals grown under those conditions can be frozen directly. This is an advantage in the case of fragile crystals.

In order to determine a protein structure with a bound ligand or heavy metals, the method of choice could be crystal soaking or co-crystallisation with a ligand. In the case of crystal soaking, the crystal is placed in artificial mother liquor containing an excess of ligand, which can be done at the same time as cryo-protection. The soaking time and the concentration of the substrate need optimization and depend on the crystal fragility, size of water channels and the size of the ligand. In this work the time of crystal soaking varied between few seconds to 16 hours.

X-RAY SOURCE - SYNCHROTRON RADIATION

Nowadays, the elucidation of protein structures is carried out primarily using the synchrotron radiation. Apart from the exceptional X-ray brilliance and highly collimated beam, synchrotron radiation offers a possibility to tune (adjust) the wavelength which is desirable in particular experiments, for instance with single wavelength anomalous dispersion (SAD) or multi-wavelength anomalous dispersion (MAD). Data were collected using the synchrotron light source at:

- EMBL Hamburg, Petra III, beamline X12, X13, X14
- BESSY Berlin, beamline BL14.1, BL14.2
- MAXLAB II Lund, beamline I911-2, I911-3

Crystals were transported to the beamlines on crystallisation plates or frozen in a dry-shipper at the temperature below 100K.

X-RAY DATA COLLECTION STRATEGY

The data collection it the last experimental step of the crystal structure determination and should be carried out with great care. Its superior aim is to collect every unique reflection, at least once, to the highest resolution available, making the data set complete. Protein crystals differ a lot and should be treated individually. The choice of strategy of data collection could be aided by computer software such as Mosflm (Leslie & Powell, 2007), as there are several factors to be taken into account, but the overall strategy must be

decided by the experimenter because it depends on individual crystal properties. The crystal symmetry, the unit cell parameters, the crystal orientation in relation to position of the beam as well as the resolution limits need to be considered. A good strategy maximizes the quantity and the quality of the data, making the data complete to the highest possible resolution with minimal radiation damage. The diffraction measurements are usually carried out at 100K in order to protect the crystal from free radicals diffusion and to decrease the thermal motion of the atoms in the crystal.

Once the crystal is cantered on the goniometer, preliminary rotational images can be collected to assess the crystal quality and select parameters for further steps.

- Exposure time

  Theoretically, the longer crystal exposure time, the higher the intensities measured and thus the higher the signal to noise ratio. Doubling the intensity increases the signal-to-noise ratio by $2^{1/2}$ (Dauter, 1999). Longer exposure is even more important for weak high resolution reflections. However too long exposure time could result in reflections overload - higher intensities that the limit of the detectors, even though the new pixel detectors (e.g. Pilatus) provide a greater dynamic range than CCD detectors. On the other hand, the higher the exposure time the longer data collection and the higher radiation damage of the crystal which results in lower data quality.

- Detector distance

  The closer the detector to the diffracting crystal the higher the Bragg angle intercepted by the detector. Of course, the distance depends on the resolution limits of the diffracting crystal and should be adjusted accordingly. In the case of setting the crystal to detector distance to collect maximum diffracting resolution it could lead to an overlap of high resolution reflection profiles and overloads at low resolution. The first applies especially in crystals with one large unit cell dimension when this dimension is oriented perpendicular to the rotation axis (Dauter, 1999) and increasing detector distance would increase the inter-spot distances. When the exposure is adjusted to measure weak high resolution reflections the contrast between reflection intensities results in the overloads of strong reflections at low

resolution. It this case both low resolution and high resolutions reflections should be measured in separate rotation pass, with different exposure time and different crystal-to-detector distance.

- Oscillation angle

  The oscillation angle (phi) describes the amount of rotation of the crystal during one exposure and is related to the number of reflections on a single image. Decreasing the oscillation angle to less than 0.5°, known as fine slicing, could help avoid the overlap of reflections in the case of large unit cell dimensions. The other advantage of fine slicing is the reduction of background noise and thus increasing the signal-to-noise ratio. Nowadays, with the new Pilatus detectors, fine slicing is the norm. On the other hand it is not recommended for poor diffracting crystals when short oscillation angle and short time of exposure is insufficient to record clear diffraction.

- The number of degrees covered in data collection

  The answer to the question '*how many degrees should be collected*?' can be obtained using the strategy software, but in general the minimum range is dictated by the space group symmetry and the crystal orientation. The aim of the native data collection is to obtain a complete set of unique reflections and to have reasonable redundancy in order to be able to assess the errors in the intensities. In the case of anomalous scattering, one generally needs to cover a wider range of crystal rotation, because of Friedel pairs are no longer equivalent; one also needs high accuracy and redundancy of the data in order to detect the weak anomalous signal.

COMPUTATIONAL TECHNIQUES – CONVERTING SPOTS TO NUMBERS

For data processing two suites of programs have been used: HKL (Otwinowski & Minor, 1997) containing xdisp (data visualization), denzo (data indexing and integration) and scalepack (data averaging and scaling) & XDS (Kabsch, 2010)**.**

There are two kinds of information to be extracted from diffraction images. The first comes from the arrangement of reflections and gives information about the crystal lattice and symmetry. The second is hidden in diffraction intensities and gives information

about the crystal content. This information is incomplete because the phases are not measured. In order to read the information from diffraction images the reflections are indexed and the unit cell is defined by three distances (*a, b, c*) and angels (α, β, γ), and Miller indices (h, k, l) are assigned to each reflection. Based on the information about the position of strong reflections, the position of all reflections can be predicted.

After the data have been indexed, corrections for instrumental factors (e.g. crystal to detector distance, position of the direct beam), polarization effects, X-ray absorption and crystal decomposition must be applied to the entire data set. To calculate intensities of the reflections the detector background must be estimated and subtracted from the reflection profile. The integration step reduces the raw diffraction images to a set of integrated intensities of particular reflections. Those procedures are carried out by the software.

Subsequently, the integrated intensities (I) must be put on the same scale. Intensities measured more than once, due to the symmetry, e.g. Friedel pairs, must be merged unless the Bijvoet differences are needed for anomalous phasing. During the scaling process, the various parameters, like the unit cell dimensions, and crystal orientation, are refined with a greater accuracy than in the initial assessment of individual images. An important final step is to confirm the space group assignment by an inspection of the systematic absences in the case of screw axes.

DIFFRACTION DATA QUALITY

At the end of the data processing the quality of the data must be assessed by statistics. Generally, the first factor that one could look at is the resolution. In general, the higher the better. There are many ways of assessing the high resolution limit of diffraction. The first is the signal-to-noise ratio, i.e. the ratio of the intensity to the error associated with the intensity, $I/\sigma(I)$ which generally should be significantly higher than one in the highest resolution shell. Another parameter that should be considered is the completeness of the data in the highest resolution shell.

The most frequently reported descriptor of the data quality is the merging *R* factor R_{merge} (Stout & Jensen, 1968)(equation 2) which tells us about differences between

symmetry related reflections, which ideally should equal. The lower the R_{merge} the more consistent the data, but one should remember that this parameter is redundancy-dependent. In practise, if reflections are measured only once the R_{merge} is 0. However, the higher the redundancy, the more reliable the data.

$$R_{merge}(I) = \frac{\sum_{hkl} \sum_I \left| I_i(hkl) - \overline{I(hkl)} \right|}{\sum_{hkl} \sum_I I_i(hkl)}$$

Equation 2. Merging R factor. $I_i(hkl)$ is the integrated intensity of a given reflection, $\overline{I(hkl)}$ is the mean intensity of multiple corresponding symmetry-related reflections

Another useful global quality indicator is the redundancy-independent *R* merge factor $R_{r.i.m.}$ (Weiss & Hilgenfeld, 1997) which describes the accuracy of the averaged measurement. A useful indicator of crystallographic information in each resolution shell is a correlation factor CC(1/2) between random half-datasets (Karplus & Diederichs, 2012).

STRUCTURE SOLUTION

Electron density

Solving a crystal structure requires evaluating a mathematical function that describes the electron density at every point in the unit cell (equation 3).

$$\rho(xyz) = \frac{1}{V} \sum_{hkl} |F_{hkl}| e^{-2\pi i(hx+hy+hz)-\varphi(hkl)}$$

Equation 3. Function defining the electron density at a given point of the unit cell. This function, represented by the Greek letter ρ, at each point in the unit cell given by the coordinates (x, y, z). $|F_{hkl}|$, known as structure factor amplitudes, represent the diffracted beams from all atoms contained in the crystal lattice; h, k, l are the Miller indices of the diffracted beams and φ(hkl) represents the phases associated with the structure factors.

The equation above represents the Fourier transform between direct space and the reciprocal (diffraction) space, represented by the structure factor amplitudes and their phases. It shows the holistic nature of the diffraction because to calculate the electron density in one point in the unit cell it is necessary to use the contribution of all structure factors.

The phase problem

X-ray detectors can only read intensities but not phases of the electromagnetic waves. The amplitude can be easily calculated taking the square root of the intensity (equation 3), however to evaluate the electron density function (equation 4) one must solve the so-called phase problem.

$$I(hkl) = |F(hkl)|^2$$

Equation 4. Relationship between intensity and amplitude.

The structure factor can be represented as a vector F_{hkl} on a complex plane, having the amplitude $|F_{hkl}|$ and the phase φ_{hkl}. This representation is called the Argand diagram (Fig. 13).

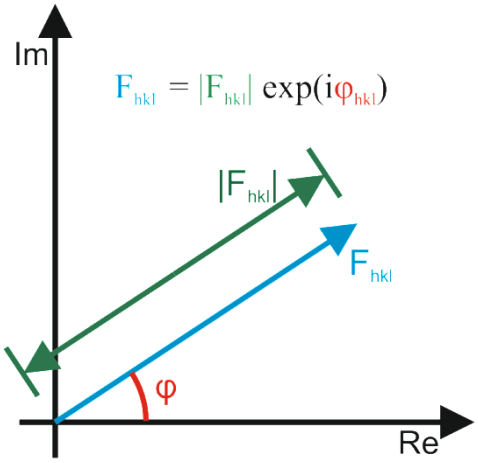


Figure 13. F_{hkl} is complex and can be represented with an Argand diagram. We determine $|F_{hkl}|$ experimentally but we still need φ_{hkl}

Patterson map

The solution to the phase problem was proposed by Arthur Lindo Patterson. Based on the electron density function he introduced a new $P_{(uvw)}$ function with a simplified formula that does not involve phases and amplitudes of the structure factors are replaced by their squares (intensities) (equation 5).

$$P(uvw) = \frac{1}{V}\sum_{hkl}|F_{hkl}|^2 \cos 2\pi(hu + hv + hw)$$

Equation 5. Patterson function

Patterson map is a vector map with the peaks corresponding to the vectors between atom pairs in the crystal. The intensity of each peak is proportional to the product of the atomic numbers of the two atoms to which the vector refers. This makes the Patterson function particularly effective for the identification of the position of heavy atoms (and subsequently the other atoms) in case when the remaining atoms have lower atomic numbers. Once the map is calculated it allows obtaining the absolute position (x, y, z) of the heavy atoms within the unit cell. The Patterson maps can be interpreted for small molecules. For a protein structure, the Patterson map is far too complex to allow the resolving of the individual inter atomic vectors *de novo*. Nevertheless, various phasing methods for solving large molecules structures, such as proteins and nucleic acids, are based on the Patterson function.

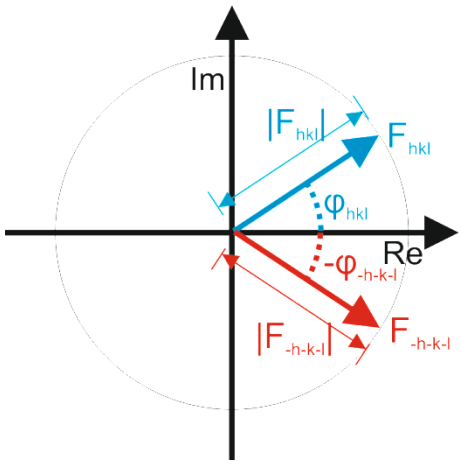
Isomorphous replacement

This method involves introducing heavy atoms in the crystal structure without changing its properties in comparison to the native form, hence it should be isomorphic. Heavy atoms characterized by higher scattering power in comparison to proteins normal atoms, change significantly the intensities of the diffraction spots in comparison with the native data. This is used to calculate a map of interatomic vectors involving the heavy atoms, finding their positions and using the partial structure to determine the phases of the entire structure.

Anomalous scattering

Since the beginning of protein crystallography anomalous diffraction signal was found to be useful for phase estimation. In the beginning of the 90s multiwavelength anomalous diffraction method (MAD) was widely used (Hendrickson, 1991) especially after the development of selenomethionine incorporation to proteins (Hendrickson *et al.*, 1990). Nowadays the method of choice is the simpler and often successful single wavelength anomalous dispersion/diffraction (SAD)(Blow, 2003).

Anomalous diffraction phasing is based on differences between intensities in Friedel pairs. Friedel pairs are Bragg reflections related by inversion through the origin and Friedel's Law states that members of a Friedel pair have equal amplitude $|F_{hkl}|$ and opposite phase φ_{hkl} (Fig. 14).



$$|F_{hkl}| = |F_{\overline{hkl}}|$$

$$\varphi_{hkl} = -\varphi_{\overline{hkl}}$$

Figure 14. On an Argan diagram Friedel's pairs appear to be reflected across the real axis. It is true for crystals in the absence of anomalous scattering. In data processing, Friedel pairs are averaged together.

This law is broken whenever an anomalous scatterer is introduced to the crystal. In case all atoms scatter equally, the amplitudes remain the same but the phase relationship no longer holds. In case some atoms scatter anomalously and some don't then both the amplitude and phase relationship is broken. The latter is used in SAD and MAD phasing.

The largest changes in anomalous scattering factors of atoms occur around the characteristic absorption edges since their values are directly related to the atomic absorption coefficient. Wavelength-tunable synchrotron radiation allows changing the energy so the incident X-ray beam has a frequency close to natural vibration frequency of the electrons in the incorporated heavy atoms. The total atomic scattering factor which in addition to normal scattering term f^0 that depend on the scattering angle is modified by two components depending on the wavelength: real f' and imaginary f" (equation 6). For SAD experiment f" is maximized.

$$f = f^0 + f' + f''$$

Equation 6. Total atomic scattering factor in case of anomalous scattering

If the wavelength is optimized to give the maximum effect, the incident beam makes atoms scatter anomalously: they absorb and re-emit the photons with the phase retarded by 90° relative to the normally scattered photons. This can be visualised on the Argand diagram as a vector at the right angle and always in the 'anticlockwise' direction added to each vector that represents the classical scattering (Fig. 15).

It breaks down the Fridel's law giving rise to anomalous differences that can be used to locate the anomalous scatterers. By increasing the quality of the data measurement and the redundancy of the diffraction data set, even weak anomalous signals can be used to estimate the phases of the entire data set.

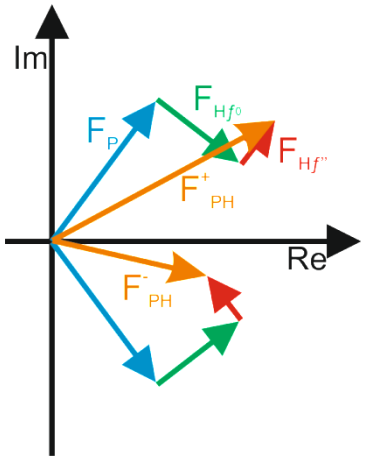


Figure 15. The structure factors of the light atom (FP) and the contribution of the heavy atoms (FHf°, FH*f*'') add up to give the structure factor of the whole structure (FPH). The anomalous component (FH*f*'') is perpendicular to the previous FHf° and always positive. This breaks down the Friedel's law and pars F^{+}_{PH} and F^{-}_{PH} are not equal.

Molecular replacement & Patterson map

A common way to solve the phase problem is the molecular replacement (MR) which requires a model of a homologous known structure with a sufficient sequence similarity. It is based on the observation that proteins with similar primary structure have similar folds. The idea is to place the initial (search) model in the new unit cell by discovering its correct orientation and position. Both are calculated using the rotation and the translation functions where they show the highest correlation between the experimental measurements and those calculated from the initial model (Rossmann, 1972, Rossmann & Blow, 1962).

The rotation function relies on the Patterson function, which can be computed without the phase information. The conventional interpretation of the Patterson function is that it consists of the interatomic vectors integrated over the entire unit cell. Patterson map is calculated from the experimental intensities and compared to a theoretical Patterson based on the coordinates of the homologous structure. When the model is oriented and placed in the correct position for the new structure, the two Patterson functions should be similar and many intramolecular vectors will overlap. Thus the correct orientation is detected. Determining the correct position of the molecule usually also relies on comparing interatomic vectors, but the method focuses on intermolecular vectors instead. Once the

known protein structure is properly oriented and translated within the new unit cell, an electron density map is calculated using phases calculated from these atomic positions and structure factors amplitudes based on the experimental intensities.

STEPS IN MODEL BUILDING AND REFINEMENT

Model building

Depending on the method of phasing, the initial models are usually biased and incomplete. In the case of isomorphous replacement or anomalous scattering methods, we get a map that present the experimenter with a 3-D jigsaw puzzle in which he has to anchor the amino acid residues. After MR the initial coordinates are inaccurate, the local structure could be wrong or absent and details of the structure, such as local disorder, solvent molecules or ligands, could not be observed. The initial map suffers from the so-called phase bias, coming from the initial model, and can be misleading to a significant extent. However, the initial model is usually sufficient for the proper refinement to begin. Phases can be gradually improved following adjustments to the model. The new map, calculated using these improved phases, gives an indication for further adjustments. This iterative process leads to improvements and extension of the model until it is judged to be refined and final. A well refined model fits the electron density and accounts for all its significant features. In addition to the protein it usually includes ordered solvent and any other ligands present in the crystal lattice. Electron density maps used in the refinement include:

- $2F_o$ - Fc computed using phases calculated from the atomic model. With the well refined model this map should show the realistic electron density.
- F_o-F_c – shows the difference between the observed and calculated structure factors (from the model). Positive peaks show density unaccounted for in the model, while negative peaks generally indicate misplaced atoms in the model.

The model could be manually corrected by displaying and manipulating in a program like COOT (Crystallographic Object-Oriented Toolkit)(Emsley *et al.*, 2010). By examination of the F_o-F_c and $2F_o$-F_c maps, new atoms can be introduced, atoms' positions can be changed or 'bad' ones can be removed.

The amount of detail that can be interpreted in the electron density depends on the resolution and the quality of the phases. Some regions can remain invisible in the electron density due to disorder: either static, i.e. caused by variations between one molecule to the next within the crystal, or dynamic disorder in case when the region of the molecule is mobile. The latter is usually decreased by cryogenic conditions during data collection.

Each round of manual fitting is followed by automated refinement using structure refinement software, such as phenix.refine (Afonine *et al.*, 2012) or Refmac (Murshudov *et al.*, 2011). Both are based on the maximum likelihood function. The ultimate goal is to optimize the simultaneous agreement of an atomic model with the observed data and with a priori chemical information (Hendrickson, 1985). The target function has two components: how the model fits the experimental data and how well it agrees with the known chemical – targets. Restraints are proper values with a standard deviation of the bond length, angles, chirality, planarity, non-crystallography symmetry (NCS) if available, torsion angels or reference structures. Restraints can be loosened with higher resolution data or with reliable maps.

Atoms in the crystal are subject to thermal motions and their displacements around their central positions must be modelled. These atomic displacement parameters (ADPs) could be refined either isotropically or anisotropically. During the isotropic refinement, where atoms are assumed to vibrate equally in all directions, ADP requires one parameter in addition to the three parameters needed for describing the position (x, y, z). In higher resolution structures, where more reflections are available, atoms can be assumed to vibrate differently in different directions, and their vibrational profile is represented as an ellipsoid. This requires six parameters in addition to the three positional coordinates. In both cases, a large number of adjustable parameters give rise to complicated target function.

The observation that the displacement of atoms are strongly coupled led to the introduction the Translation-Liberation-Screw approximation (TLS)(Winn *et al.*, 2001). The translations and liberations are defined with respect to a set of orthogonal axes, with translations along the three axes and liberations around them. Those two simultaneous movements results in a screw component. The origin of the axes is in the centre of the rigid group and directions can be refined. The key part of this method is the choice of the TLS groups performed either manually, by dividing the structure into domains or sub-domains,

or automatically, based on the TLS Motion Determination software (Painter & Merritt, 2006).

To assess the correctness of the model the R_{work} statistic, describing the agreement of the model and the data, and the cross-validation R_{free} (Brunger, 1997) statistic are commonly used (equation 7). For this purpose the diffraction data are divided into two sets: a large working set which is used in the refinement and a test set (not less than 1000 reflections). The R_{free} value computed with the test set data significantly reduced the danger of overfitting the diffraction data and is a more objective guide during the model building and refinement.

$$R_{work} = \frac{\sum_{hkl(work)} |F_o - F_c|}{\sum_{hkl(work)} |F_o|}$$

$$R_{free} = \frac{\sum_{hkl(free)} |F_o - F_c|}{\sum_{hkl(free)} |F_o|}$$

Equation 7. R_{work} and R_{free}

At the end of the refinement the model should be validated and be reasonable in terms of physics – contacts and packing, chemistry – bonds, angles, planarity, chirality, charge interactions, crystallography – R-factors, B-factors, real space correlation and statistically- no under-modelling and no over-fitting.

CRYSTAL STRUCTURES OF THE PSYCHROPHILIC *MM*CHI60

PROTEIN PURIFICATION

The chitinase gene named *Mm*Chi60 from *M. marina* was cloned as described in (Stefanidi & Vorgias, 2008) and overexpressed and purified as described in the Materials and Methods section. Briefly, cell extract was exposed to ammonium sulphate precipitation and further purified on Phenyl-Sepharose and Q-sepharose column (Fig. 16).

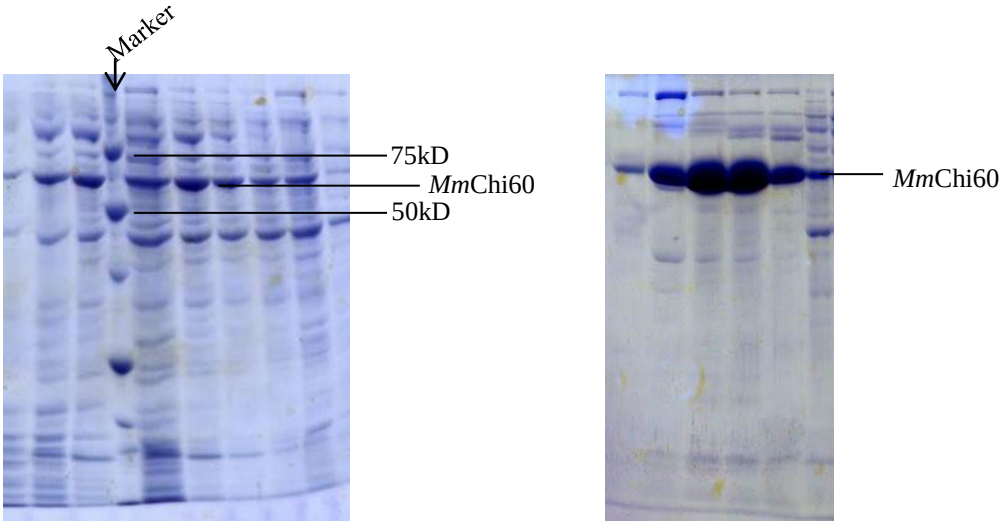


Figure 16. SDS-PAGE electrophoresis of *Mm*Chi60 elution fractions. Left Phenyl-Sepharose and right Q-Sepharose.

The size exclusion chromatography performed at end of the purification separated minor impurities and yielded a homogenous and a pure protein. Fractions collected from the chromatogram peak, with a maximum UV absorbance around 75 ml. (Fig. 17) were pooled together, concentrated to 3 mg/ml and stored at 4°C. *Mm*Chi60 was examined using Dynamic Light Scattering (DLS) and showed a high homogeneity of the sample, indicating that the protein is suitable for crystallisation trials.

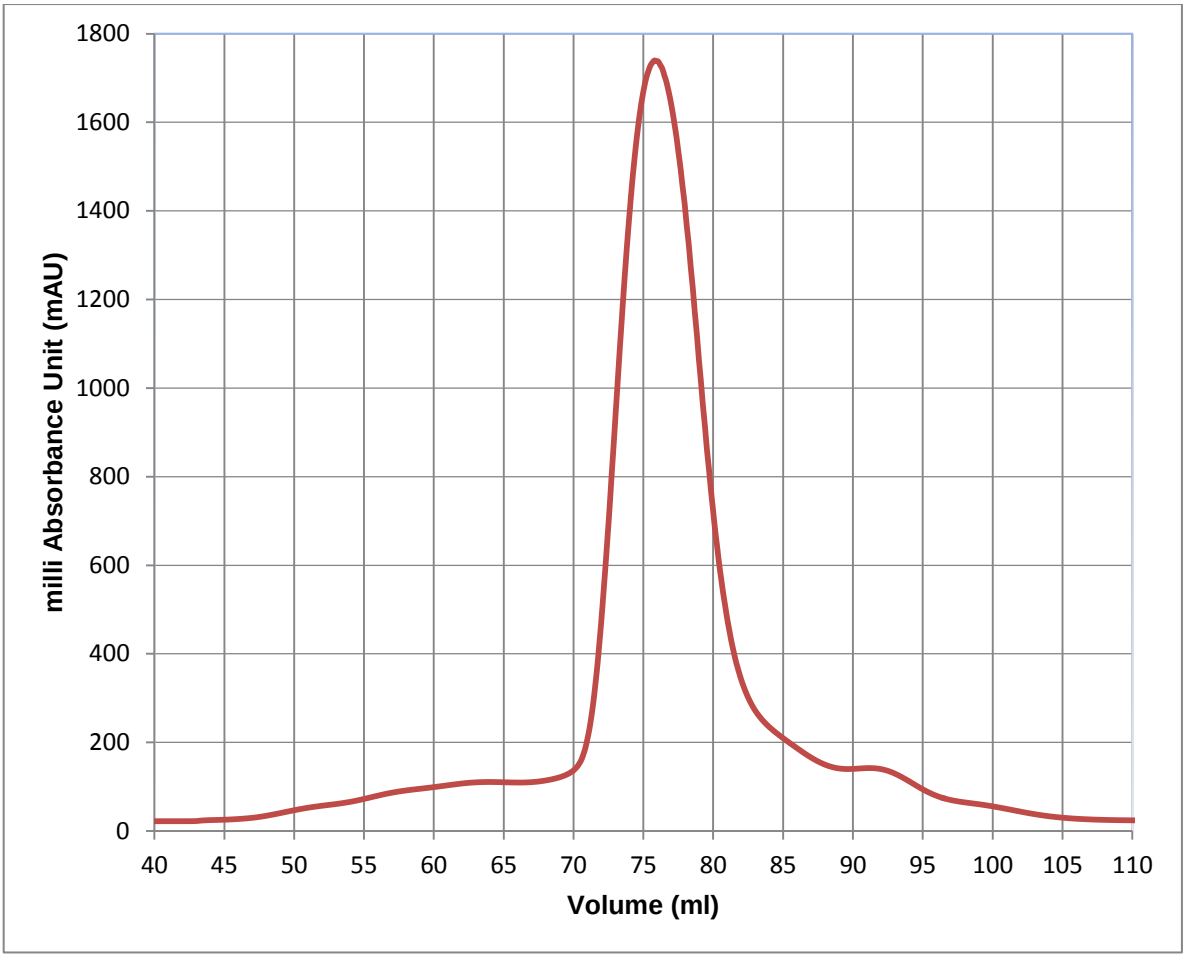


Figure 17 Chromatogram of *Mm*Chi60 sample. Fractions from the peak with a maximum at 75 ml. were collected and pooled for further crysatallisation trials.

CRYSTALLISATION CONDITIONS

First *Mm*Chi60 crystals were grown using the method of hanging drop vapour diffusion at 4°C. The well solution contained 23% PEG 4000 w/v, 0.16 M ammonium sulphate and 0.1 M citrate buffer at pH 5.5. The protein solution contained 3 mg/ml of *Mm*Chi60, 0.02 M Tris buffer at pH 7.0 and 0.5 M NaCl. Crystallisation drops were prepared by mixing the protein solution with the well solution in the ratio 2:1 or 3:1. Protein crystals appeared after approximately two weeks (Fig. 18).

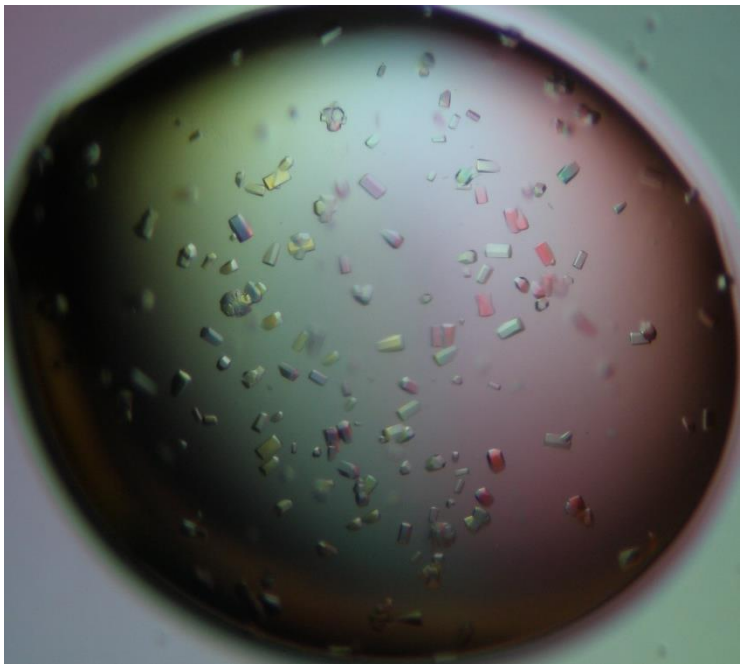


Figure 18 *Mm*Chi60 crystals in crystallisation drop of ~2µl

The crystal used in the SAD measurement was soaked for 20 minutes in the cryo-solution to which KBr had been added to the concentration of 0.5 M. The cryo-solution contained 25% glycerol (v/v) in a solution similar to the well solution.

In the case of NAG ligands soaking experiment, crystals were soaked in by adding the ligand to the cryo-solution at the concentration exceeding the protein concentration by approximately 5:1. Two experiments were performed: with N,N',N''-triacetyl chitotriose (NAG_3) and with N,N',N'',N'''-tetraacetyl chitotetraose (NAG_4). The crystals were immersed in the cryo-solution for approximately 30 seconds prior to freezing at 100 K in the cryostream. All the soaking experiments were performed at 4°C, close to *M. marina's* living temperature, and care was generally taken to maintain the *Mm*Chi60 crystals at the cold-room temperature up to the moment of freezing.

DATA COLLECTION AND PROCESSING

X-ray diffraction data were collected on the beam line 14.2 at the BESSY synchrotron in Berlin, using the marresearch MX-225 detector for. The data were integrated and scaled using the HKL2000 software (Otwinowski & Minor, 1997) and XDS (Kabsch, 2010). Summary of the X-ray data collection and processing is given in Table 28.

Table 28 Summary of the X-ray data collection and processing

	*Mm*Chi60/Br	*Mm*Chi60	*Mm*Chi60/NAG_3	*Mm*Chi60/NAG_4
Space group	$P3_112$	$P3_112$	$P3_112$	$P3_112$
Unit cell parameters				
a = *b* (Å)	67.81	67.62	66.00	67.09
c (Å)	259.68	257.19	257.54	258.03
Wavelength (Å)	0.92001	0.91841	0.91841	0.91841
Oscillation (°)	0.3	0.5	0.5	0.5
Crystal-to-detector distance (mm)	280	210	210	220
Resolution (Å)	50.0-2.70 (2.75-2.70)	50.0-2.1 (2.14-2.1)	50.0-2.07 (2.11-2.07)	50.0-2.26 (2.32-2.26)
R_{merge}	0.134 (0.869)	0.071 (0.451)	0.114 (0.94)	0.098 (0.979)
Completeness (%)	100.0 (100.0)	95.5 (70)	99.5 (99.1)	99.3 (98.0)
Observed reflections	154117	323034	359867	292099
Unique reflections	19182	38154	39522	31416
$\langle I/\sigma(I) \rangle$	14.4 (1.8)	28.1 (2.3)	20 (1.9)	21.1 (2.4)
Multiplicity	8.0	8.5	9.1	7.5

Values in parentheses correspond to the highest resolution shell

SOLVING STRUCTURE – MR-SAD METHOD

The phase problem for *Mm*Chi60 was solved using the *Mm*Chi60/Br data (Table 28) by the MR-SAD method implemented in the AutoRickShaw pipeline (Panjikar *et al.*, 2005). Briefly, a partial molecular replacement (MR) solution was obtained with MOLREP (Vagin & Teplyakov, 2010), using as the search model the crystal structure of the chitinase from *Lactococcus lactis*, stripped of solvent molecules (Bonanno *et al.*, unpublished; PDB code 3ian). The MR model at this stage, consisting of a TIM β/α-barrel, included 309 amino acid residues and corresponded to approximately 60% of the *Mm*Chi60 structure. The MR model was pre-refined with Refmac5 (Murshudov *et al.*, 2011) and used in a maximum-likelihood SAD phasing calculation with PHASER (McCoy *et al.*, 2007) to improve phases and identify sites of Br anomalous scatterers. The positions and occupancy factors of the heavy atoms were refined with MLPHARE (Winn *et al.*, 2011), SHARP (Bricogne *et al.*,

2003) and BP3 (Pannu & Read, 2004). The resulting phases were improved using PIRATE by applying real space constraints based on known features of a protein electron density map (Cowtan, 2000). Next, a polyalanine model of yet uninterpreted regions of electron density was constructed using SHELXE (Sheldrick, 2008). Following that, a round of automatic model building was performed using RESOLVE (Terwilliger, 2003) and BUCCANEER (Cowtan, 2008).

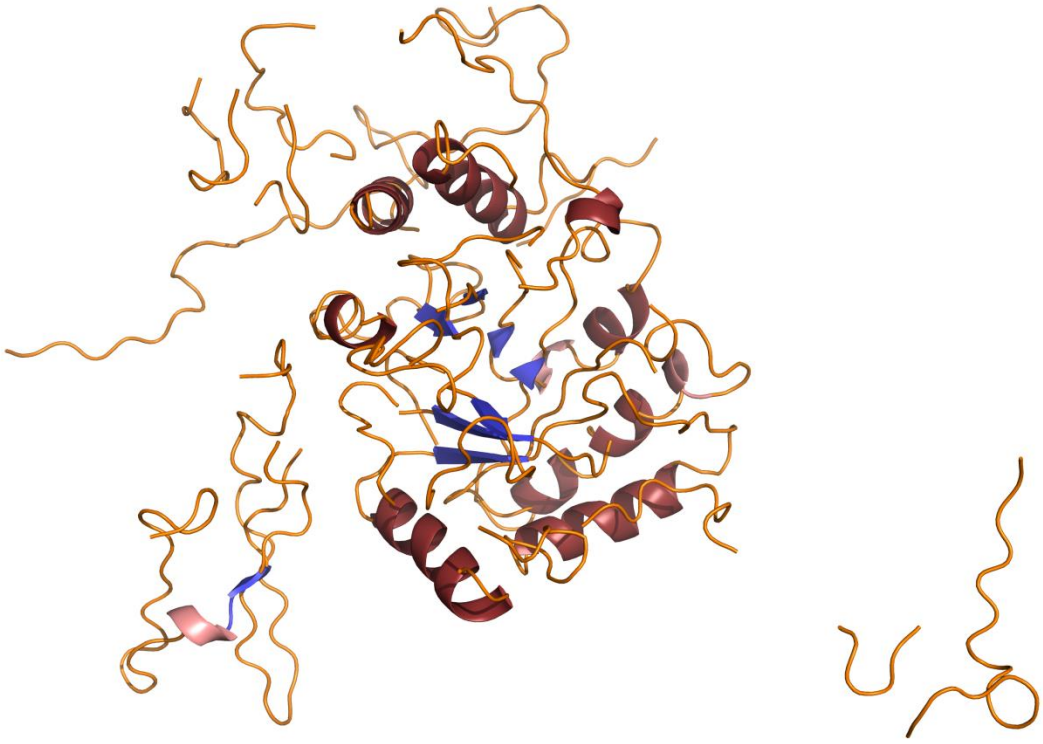


Figure 19. The model obtained from the Auto Rickshaw pipeline by the MR-SAD method. In the middle of the figure, the outline of β/α-barrel can be identified. Parts of the model are spread across symmetry related molecules.

The updated model was refined with Refmac5 (Murshudov *et al.*, 2011). A total of six cycles were performed of phasing, heavy atom search and model building, as described above. In the end, twelve bromide sites were identified in the anomalous map above the 3σ level. Three of the peaks were greater than 10 σ (Fig. 20). The R-factor and R_{free} for the atomic model at this stage were 0.31 and 0.38(Fig. 19).

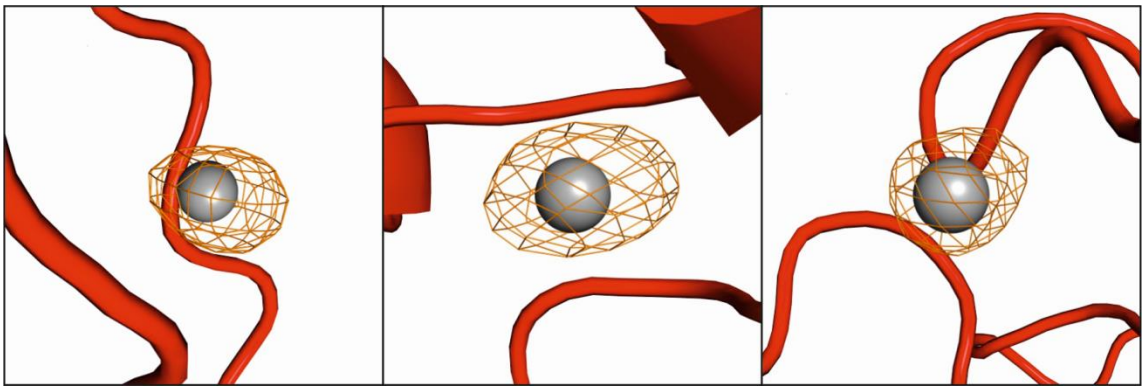


Figure 20 Three bromides modeled in phased anomalous difference map contour at 10σ.

MODEL REFINEMENT

The final model was obtained after cycles of manual adjustments using COOT (Emsley *et al.*, 2010) and Refmac5 (Murshudov *et al.*, 2011). ARP/wARP (Langer *et al.*, 2008) was also applied for automatic model building, however only few residues have been anchored in a sequence. The other *Mm*Chi60 structures were solved by MR using the model obtained as described above. In addition to Refmac5, phenix.refine was used for model refinement (Afonine *et al.*, 2012).

Table 29 The final refinement statistics

	***Mm*Chi60**	***Mm*Chi60/NAG_3**	***Mm*Chi60/NAG_4**
PDB code	4hmc	4hme	4hmd
TLS groups	3	3	3
R_{work}/ R_{free}	0.19/0.24	0.18/0.21	0.19/0.25
Protein atoms	4137	4137	4137
Ligand atoms	7	54	35
Water molecules	261	292	167
Average B factor ($Å^2$)	44	51	47
R.m.s.d. from ideal			
Bond lengths (Å)	0.018	0.020	0.016
Bond angles (°)	1.87	1.85	1.76
Ramachandran plot (%)			
Favored	96.2	97.9	96.0
Allowed	3.6	1.9	3.8
Outliers	0.2	0.2	0.2

ADPs in all models were refined isotropically and TLS parameters (Winn *et al.*, 2001) were refined for three groups, as calculated by TLSMD server (Painter & Merritt, 2006). For the cross-validation more than 1000 reflections were randomly selected. The final refinement statistics are listed in Table 29.

THE OVERALL STRUCTURE

In the crystal lattice, the *Mm*Chi60 molecules are nested pairwise against each other, around the two-fold axes (Fig. 21). The *Mm*Chi60 molecule consists of 528 amino acid residues (23-550) arranged in four domains: the N-terminal catalytic β/α-barrel domain (residues 23-342, classed as glycoside hydrolase family 18 – GH18), two bacterial immunoglobulin-like (Ig-like) domains, known as Big 3 domains (Pfam database ID PF07523, residues 348-421 and 422-501) and the C-terminal chitin-binding domain (residues 507-550). The fold of the chitin-binding domain is recognised as the carbohydrate-binding module of type A (Boraston *et al.*, 2004), Family 5 (Akagi *et al.*, 2006), abbreviated as CBM5, with some specific chitin-binding features. The domains are arranged linearly, giving the molecule an elongated, sea-horse shape (Fig. 21, 22). Electron density of the main chain is continuous except at Val505 between the chitin-binding domain and the adjacent Ig-like domain. The two cysteine residues present in the sequence (Cys41 and Cys53) form a disulphide bridge at the base of the hairpin formed by β-strands 43-44 and 50-51. A sodium-binding site was identified in the catalytic domain. The cation is coordinated by the side chains of Asn105, Asp146, the carbonyl oxygen atom of Gly144, the

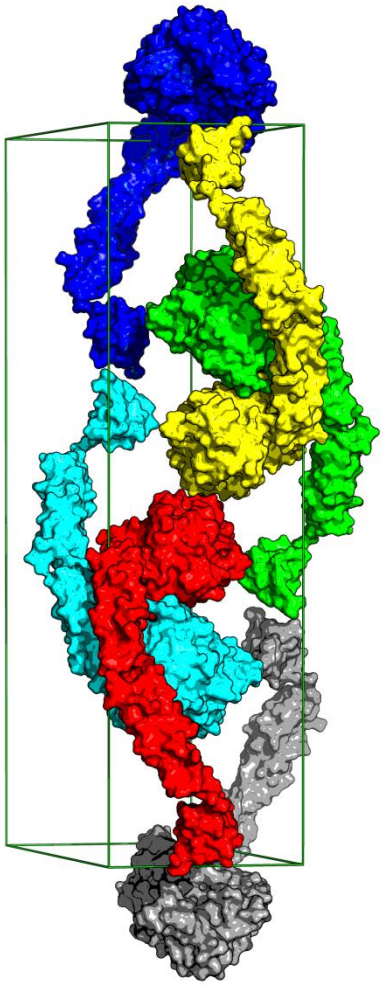


Figure 21. Crystal packing. The molecules pack pairwise around the crystal two-fold axes. There is one protein molecule in the asymmetric unit and six protein molecules in the unit cell.

carbonyl oxygen atom of Thr24 (2.6-2.7 Å) and a water molecule (distances 2.2-2.4 Å). A sodium site was found in a corresponding place also in the chitinase from *L. lactis* (unpublished, PDB code 3ian). The linker between the catalytic domain and the first Ig-like domain begins with a pair of proline residues (Pro343, Pro344). The linker is well ordered and it contacts the β/α-barrel *via* buried water molecules and polar side chains on the catalytic domain's surface. Each Ig-like domain is a bundle of five β-strands (the third strand being interrupted by a 3_{10}-turn). The β-strands are antiparallel except for the first with the last, which are parallel (a consequence of an odd number of strands in the bundle). The two Ig-like Big 3 domains show 40% identity (75% similarity) between their core regions of 77 amino acids. There is no clear pattern to the conservation. According to the Pfam server (Finn *et al.*, 2010) the Big 3 domains are found in cell wall adhesion proteins and in various glycoside hydrolyses.

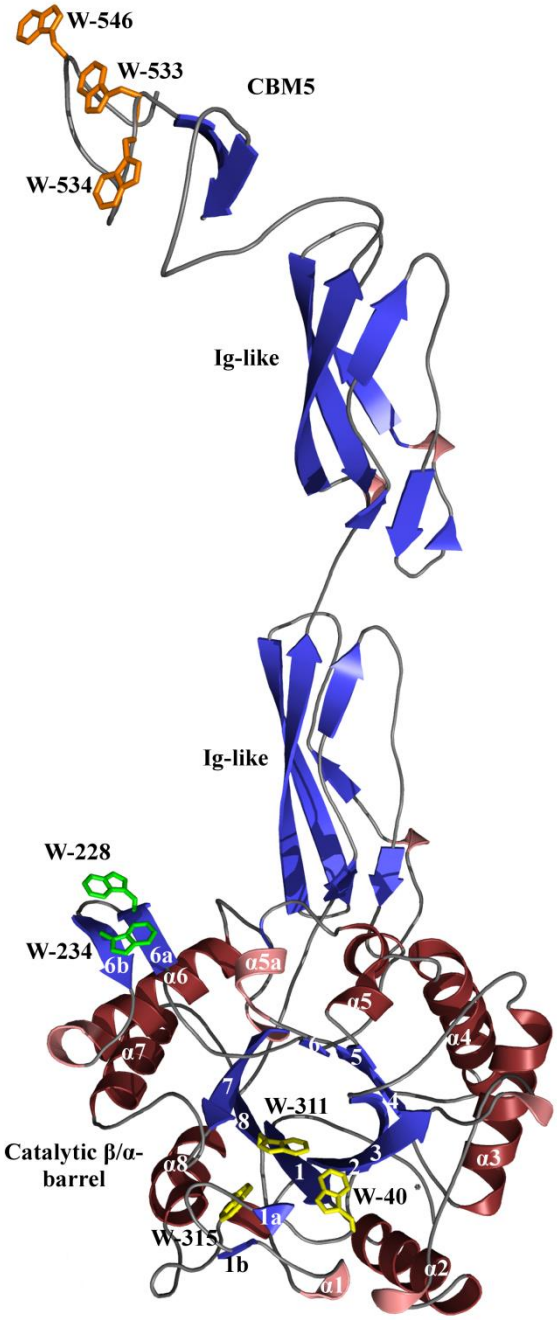


Figure 22. Cartoon plot of *Mm*Chi60. The four protein domains are labeled. The secondary structure elements – α-helices (red), β-sheets (blue), 310-helices (pink) – of the catalytic domain are labeled according to the convention for the β/α-barrels. Tryptophan residues believed to be significant in chitin binding are shown and labeled: those of the chitin-binding domain (orange), the Trp patch on the catalytic domain (green) and Trp residues lining the active site (yellow).

CHITIN BINDING

The small 44-residue chitin-binding (CMB5) domain contains six Trp residues, of which three contribute to the hydrophobic core (Trp510, Trp529, Trp549) and the other three (Trp533, Trp534 and Trp546) are exposed on the surface and their side chains are co-planar (Fig. 22, 23). Two of the exposed Trp residues are conserved in other known chitin-binding domains of the same CBM5 class: in ChiC from *S. griceus* (Akagi *et al.*, 2006) and in ChiB from *S. marcescens* (van Aalten *et al.*, 2000). The third, Trp546, has not been observed in other crystal structures. It resides on a three-residue insertion that forms a small loop. The distance between the exposed Trp side chains in *Mm*Chi60 is 10-11 Å, which corresponds with the length of two NAG units, i.e. the period of the poly-NAG chain in chitin.

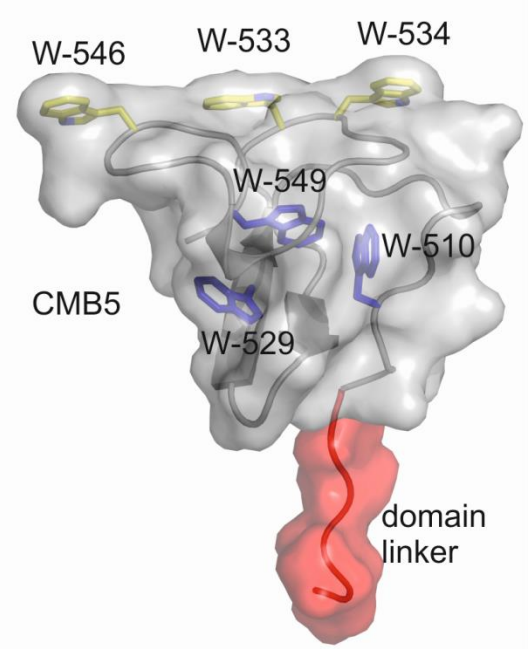


Figure 23. Semi-transparent surface representation of the chitin-binding domain with flexible linking region (red) with the second Ig-like domain. Trp residues exposed to the solvent (yellow sticks) and hydrophobic core Trp (blue sticks) are indicated.

Recently, the solution structure of chitin-binding domain 1 (ChBD1) from hyperthermophilic chitinase from *Pyrococcus furiosus* has been reported (Mine *et al.*, 2014). It has three aromatic residues (two Tyr and Trp residues) in the corresponding place to solvent exposes Trp residues of *Mm*Chi60. In ChBD1 each one was replaced with Ala what resulted in a decrease of the binding affinity to the chitin powder to 30-40% of the affinity of the wild type.

A similar surface feature to *Mm*Chi60 ChBD, consisting of two nearly co-planar Trp residues (Trp228 and Trp234) is found on the catalytic domain (Fig. 21). The distance between the two exposed side chains is 7.3 Å. A similar feature is found in the chitinase from *L. lactis* (PDB code 3ian) and in ChiC from *S. marcescens* (Payne *et al.*, 2012). No NAG ligands have been found in the crystal lattice, associated with the two surface Trp clusters.

SUBSTRATE-BINDING GROOVE

The substrate-binding groove is located at the carboxyl ends of the β-strands of the β/α-barrel, similar to other enzymes with the TIM barrel architecture. In the *Mm*Chi60 unliganded structure, the groove is occupied only by solvent water molecules. In the structure soaked in NAG_3, a NAG_2 ligand was observed in the substrate-binding sub-sites -1 and -2 (Davies *et al.*, 1997) (Fig. 24). In the structure soaked in NAG_4, a ligand of similar size to NAG_2 was observed but with distinct features in the -1 site. There was no electron density for the O1 atom and the density of the N-acetyl chain indicated formation of an oxazolinium ring. Subsequently, the ligand density was interpreted as an oxazolinium ion reaction intermediate (Fig. 24). The differences in the ligand structure are accompanied by local differences in the protein structure and the solvent (see below).

The bottom of the substrate-binding groove is lined with Trp40, Trp311 and Trp315 (Fig. 22, 24a). The first two Trp residues are in stacking interactions with the two NAG moieties in the substrate-binding site. In addition to the stacking, Trp311 H-bonds the carbonyl oxygen atom of the N-acetyl group of the NAG unit in the -2 site. The side chain of Trp315 provides another H-bond with the same oxygen atom (Fig. 24a).

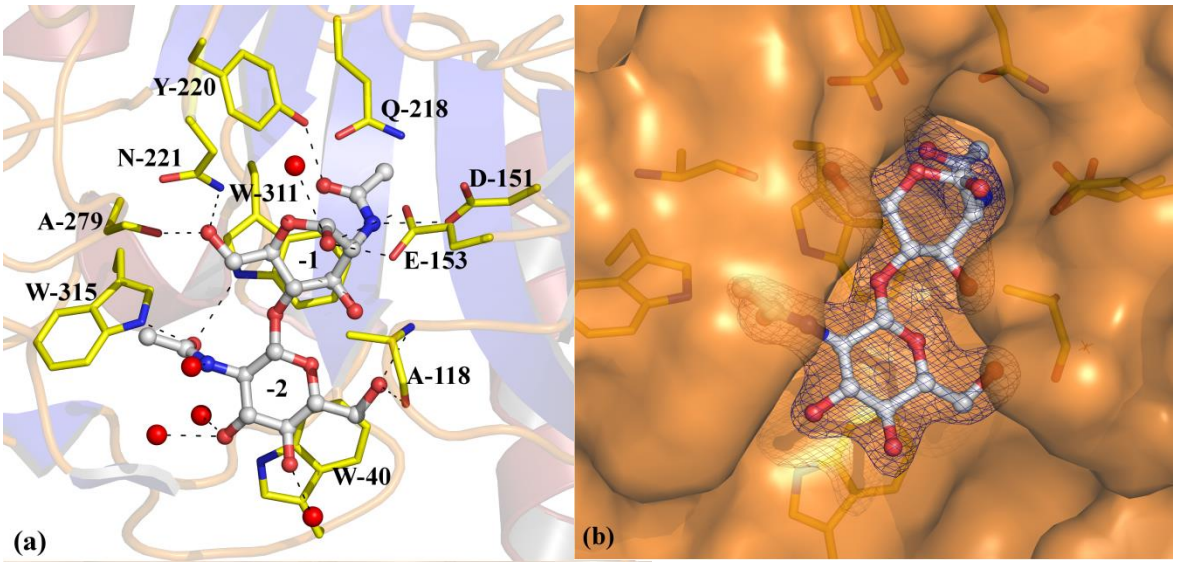


Figure 24. (a) Details of the interaction between the reaction product NAG2 and *Mm*Chi60. Hydrogen bonds between the ligand and the protein's binding site and water molecules (red spheres) are indicated as dashed lines. (b) Surface representation showing the ligand binding cavity and a weighted 2Fo-Fc map, contoured around the ligand at the 2 σ level.

The substrate-binding groove in *Mm*Chi60 is relatively shallow (Fig. 24b), compared with the other well-known chitinase structures, e.g. ChiA (Perrakis *et al.*, 1994) (PDB code 1ctn) or ChiB (van Aalten *et al.*, 2000) (PDB code 1e15) from *S. marcescens*, which both have additional loops on both sides of the substrate-binding groove and an α+β domain inserted in the TIM barrel on the '+' side of the substrate-binding site. Among the known structures, the substrate-binding groove in *Mm*Chi60 is the most similar to the chitinase from *L. lactis* (PDB code 3ian), ChiC from *S. marcescens* (Payne *et al.*, 2012) and the chitinase from *B. cereus* (Hsieh *et al.*, 2010).

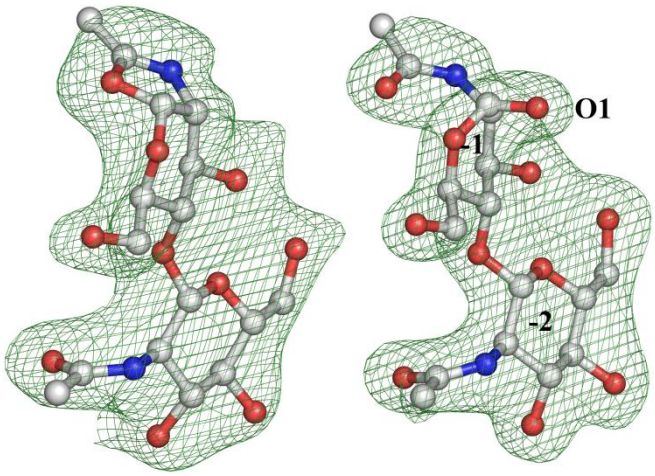


Figure 25. Comparison of the omit Fo-Fc maps, contoured at the 3 σ level, of ligands found in the binding groove of *Mm*Chi60. The reaction intermediate – an oxazolinium ion – in the -1 sub-site (left) and the reaction product – a NAG moiety – featuring the electron density of the O2 atom (right). The -2 sub-site in both complexes is occupied by a NAG residue.

THE ACTIVE SITE

The active site lies at the bottom of the substrate-binding cleft, at the junction between the -1 and +1 sub-sites, where the cleft bends sharply. The curvature of the binding groove is similar to other chitinases and has been shown to induce bending of the poly-NAG chain at the scissile β-1,4-glycosidic bond, forcing the sugar ring in the -1 site into the *boat* conformation (Papanikolau *et al.*, 2001). In the complex of *Mm*Chi60 with the NAG_2 reaction product the sugar ring is indeed found in the *boat* conformation. In the complex with the oxazolinium reaction intermediate the sugar ring is in the *chair* conformation. A cavity at the bottom of the -1/+1 junction is directed into the body of the protein, down the

axis of the β/α-barrel, and accommodates the N-acetyl group of the '-1'NAG unit, or it contains the oxazolinium ring of the bound reaction intermediate.

The active site contains two conserved acidic side chains: Asp151 and Glu153, part of the DxDxE signature of glycosyl-hydrolases belonging to family 18 (Henrissat & Davies, 1997). The conformation of Asp151 is unchanged in the three *Mm*Chi60 structures, but Glu153 changes its position relative to Asp151. In the *Mm*Chi60 complex with NAG_2 the distance between the neighbouring carboxyl O atoms of Asp151 and Glu153 is 3.1 Å, indicating a hydrogen atom between them. In the ligand-free structure and in the complex with the oxazolinium reaction intermediate the corresponding distance between the two residues is approximately 10 Å. This is due to a shift in the main chain and a flip of the Glu153 side chain. The flipping of the side chain results in a distortion from planarity of the peptide bond between residues 152 and 153, which is reflected in the value of the ω torsion angle equal to 160°. The space vacated by the displaced side chain is occupied by electron density which can accommodate a water molecule and has been modelled so. The flipping of the Glu153 side chain is accompanied by a shift of residues Gly117 to Ile122 and a rotation of the main chain Ala120-His121. In effect, the carbonyl atom of Ala120 moves by approximately 5 Å, and its place is taken by the carboxyl group of Glu153, whose oxygen atoms H-bond with N of 119Asp (2.8 Å) and N of 120Ala (2.9 Å) (Fig. 26).

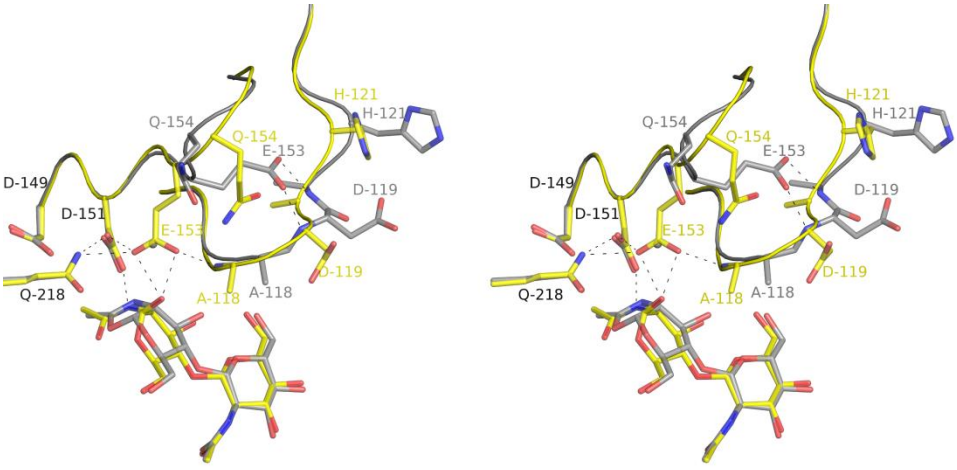


Figure 26. Stereo pair showing a comparison of *Mm*Chi60 structures with the bound oxazolinium reaction intermediate (grey sticks, grey labels) and with the NAG_2 reaction product (yellow sticks and labels). Black labels are used for structurally unchanged residues. The different stages of ligand processing are accompanied by significant rearrangements of residues in the binding site, especially the active residues Glu153 and the flexible loop 117-122. Black dashes indicate hydrogen bonds of Asp151, Glu153 and other residues.

A close contact exists between the C1 atom of the oxazolinium intermediate and the carbonyl oxygen atom of the Gln218 side chain (3.1 Å), indicating a C-H…O hydrogen bond. The distance from the carbonyl oxygen and oxygen of the oxazolinium ring is 2.8 Å – a close Van der Waals contact. Some residual difference electron density in the structure at and around the oxazolinium ion indicates a low-occupancy species similar to the structure with NAG_2 reaction product. However, the occupancy was too low for meaningful modelling.

The O6 hydroxyl of NAG in the -1 sub-site is bound to the Nδ2 amine group of Asn221 and the carbonyl oxygen of Ala279 (Fig. 24*a*). This is a different coordination of the O6 hydroxyl from that found in the chitinases ChiA and ChiB from *S. marcescens* which have an Asp in the place of Asn221 (H-bond acceptor instead of donor) and in which the O6 hydroxyl makes no H-bond with the main chain. The arrangement found in the present structures is similar to that found in the chitinase from *B. cereus* (Hsieh *et al.*, 2010) (PDB code 3n11) and a plant chitinase/lysozyme (hevamine) from *Hevea brasiliensis* (Terwisscha van Scheltinga *et al.*, 1996) (PDB code 1llo).

PROTEIN-SORTING SIGNAL

In *Moritella marina* the protein product of the *Mm*Chi60 gene is secreted out of the bacterial cell (Stefanidi & Vorgias, 2008). The N-terminal end in the protein is recognized as a signal peptide with a cleavage site between position 22 and 23 by a SignalP 4.1 Server (Petersen *et al.*, 2011) which discriminates sorting-peptides from transmembrane regions. The electron density for all the crystal structures is clearly visible only from the 23rd residue which indicates either disorder in the N-terminal region or signal peptide cleavage by the *E.coli* machinery. The mass spectrometry (MS) determined the mass of the *Mm*Chi60 to be 58623 Da (Fig. 27), which is consistent with the theoretical MW of the N-terminal truncated protein calculated by ExPASy server: 58577 Da. The MS experiment was conducted by Paweł Rodziewicz from the Laboratory of Proteomics and Metabolomics of IBCh PAN.

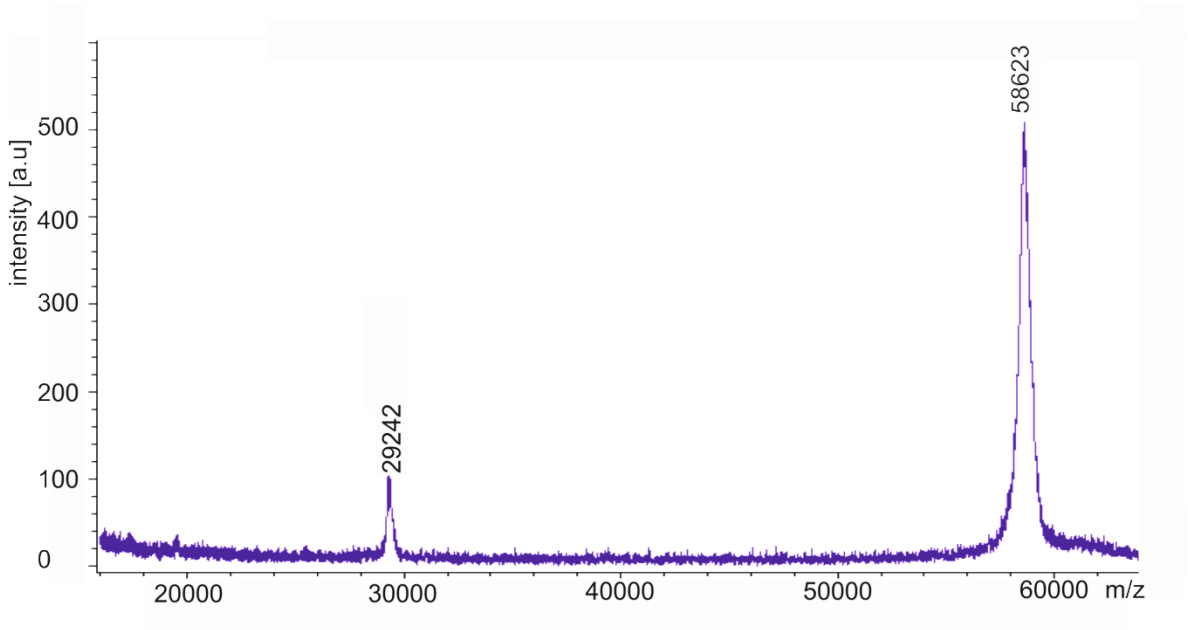


Figure 27 MALDI-TOF mass spectrum of the *Mm*Chi60 sample. Single charged particles, indicating a mass of *Mm*Chi60 is identified as a 58kDa peak. The double charged peak is around 29kDa.

CRYSTAL STRUCTURES OF THE MUTATED E153Q *MM*CHI60

PLASMID AND DNA MANIPULATION

Mutations were introduced into a plasmid isolated from *E. coli* cells previously used for *Mm*Chi60 expression. The active site mutant was produced by introducing a E153Q mutation by site directed mutagenesis using the primers in which the codon for glutamic acid GAA was mutated to glutamine CAA. The PCR product showed tandem repeats of the primer in the cloned DNA. To avoid this, two single primer extension reactions were performed. The resulting recombinant plasmids were verified by DNA sequencing.

PROTEIN PURIFICATION

Overproduction and purification steps are different from the native form of *Mm*Chi60. Advantage was taken of the assumption that the protein of interest is localized in the periplasm because the native sorting-peptide is recognized by *E. coli*. Cells were broken by osmotic shock and further purified according to the Materials and Methods section.

CRYSTALLISATION CONDITIONS

All *Mm*Chi60 E153Q crystals were grown using the hanging-drop vapour-diffusion method at 291 K. Initial crystallisation conditions were found in the Morpheus protein crystallisation screen in solution H4 using the Art Robbins Gryphon crystallisation robot: mixing 0.4µl of 10mg/ml protein solution with 0.4µl well solution. The well solution consisted of a mixture of precipitants: 12.5% w/v PEG 1000, 12.5% w/v PEG 3350, 12.5% v/v MPD, a mix of amino acids, 0.02M of each: Na-L-glutamate; alanine (racemic); glycine; lysine HCl (racemic); serine (racemic) and 0.1M MES/imidazole pH 6.5. The NAG ligands were soaked in by adding the ligand to the cryo-solution at the concentration exceeding the protein concentration by approximately 5:1. Three soaking experiments were performed with N,N',N'',N'''-tetraacetyl chitotetraose (NAG_4), N,N',N'',N''',N''''-pentaacetyl chitopentose (NAG_5) and .N,N',N'',N''',N''''N''''''-hexaacetyl chitohexose

(NAG_6). The time of the crystal immersion in the cryo-solution containing 25% glycerol and the ligand depended on the length of the NAG polymer. In the case of NAG_4, the crystal was immersed for 10 min, whereas crystals soaked in NAG_5 and NAG_6 were incubated overnight. The crystal without ligand did not require additional cryo-solution.

DATA COLLECTION AND PROCESSING

X-ray diffraction data were collected on the beam line 14.2 at the BESSY synchrotron in Berlin, using the marresearch MX-225 detector. The data were integrated and scaled using the XDS software (Kabsch, 2010). Summary of the X-ray data collection and processing is given in Table 30.

Table 30 Summary of the X-ray data collection and processing

	E153Q	E153Q Nag4	E153Q Nag5
Space group	$P3_112$	$P3_112$	$P3_112$
Unit cell parameters			
a (Å)	67.66	67.32	67.06
b (Å)	67.66	67.32	67.06
c (Å)	255.64	255.76	256.63
Wavelength (Å)	0.91841	0.91841	0.91841
Resolution (Å)	31.4-1.55 (1.64-1.55)	33.37-1.48 (1.57-1.48)	34.44-1.64 (1.74-1.64)
Oscillation (°)	0.2	0.2	0.3
Number of frames	500	550	350
Crystal-to-detector distance (mm)	150	150	150
R_{merge} [#]	0.065 (0.88)	0.054 (0.85)	0.072 (0.9)
Completeness (%)	99.8 (98.7)	99.8 (99.4)	99.9 (99.3)
Observed reflections	609238	758729	534266
Unique reflections	97965	110794	81642
$\langle I/\sigma(I) \rangle$	16.4 (1.8)	20.5 (2)	17.7 (1.9)
Multiplicity	6.2	6.9	6.5

Values in parentheses correspond to the highest resolution shell

SOLVING STRUCTURE & MODEL REFINEMENT

The phase problem was solved by molecular replacement using the Phaser program (McCoy *et al.*, 2007) and the crystal structure of *Mm*Chi60 as the search model(PDB code 4hmc). The model was refined with the phenix.refine software (Afonine *et al.*, 2012). In all cases hydrogen atoms were added at riding positions. ADPs were refined anisotropically for all protein atoms (non-H) in the unliganded and complexed with NAG_4 structures which resulted in a significant decrease of the R factors. In the structure E153Q with NAG_5 the TLS parameters (Winn *et al.*, 2001) were refined for three groups. The NAG ligands were modelled only if clear omit electron density was observed for the ligand. The final refinement statistics are listed in Table 31.

Table 31 The final refinement statistics

	E153Q	E153Q NAG_4	E153Q NAG_5
PDB code	4MB3	4MB4	4MB5
R_{work}/ R_{free}	0.14/0.18	0.14/0.17	0.15/0.18
Protein atoms	4205	4217	4200
Ligand atoms	23	102	119
Water molecules	909	711	719
Average B factor (Å^2)	26	25	24
R.m.s. deviation from ideal			
Bond lengths (Å)	0.02	0.02	0.02
Bond angles (°)	1.63	1.76	1.79
Ramachandran plot (%)			
Favoured	97.6	98.16	97.77
Allowed	2.4	1.84	2.23
Outliers	0	0	0

LOW ACTIVITY MUTATED STRUCTURES

The enzymatic activity of the mutated *Mm*Chi60 E153Q is 10^5 lower than the wild type enzyme. The mutant was prepared to study the interactions of NAG substrates with the chitinase. Although the crystallisation conditions were different, the mutant protein

crystallised in the same space group as the wild type, with very similar unit cell parameters (Table 30). Consequently, the overall structure of the *MmChi60* E153Q mutant with and without ligands resembles the wild type structure.

Two complexes have been obtained: with NAG_4 and NAG_5 bound in the enzyme's substrate binding groove. To examine the effect of the mutation on the architecture of the active site in its resting state the structure of the E153Q mutant was also solved without a ligand. In addition, a structure of the E153Q mutant soaked with NAG_6 was examined, but unfortunately electron density was only observed for five NAG moieties.

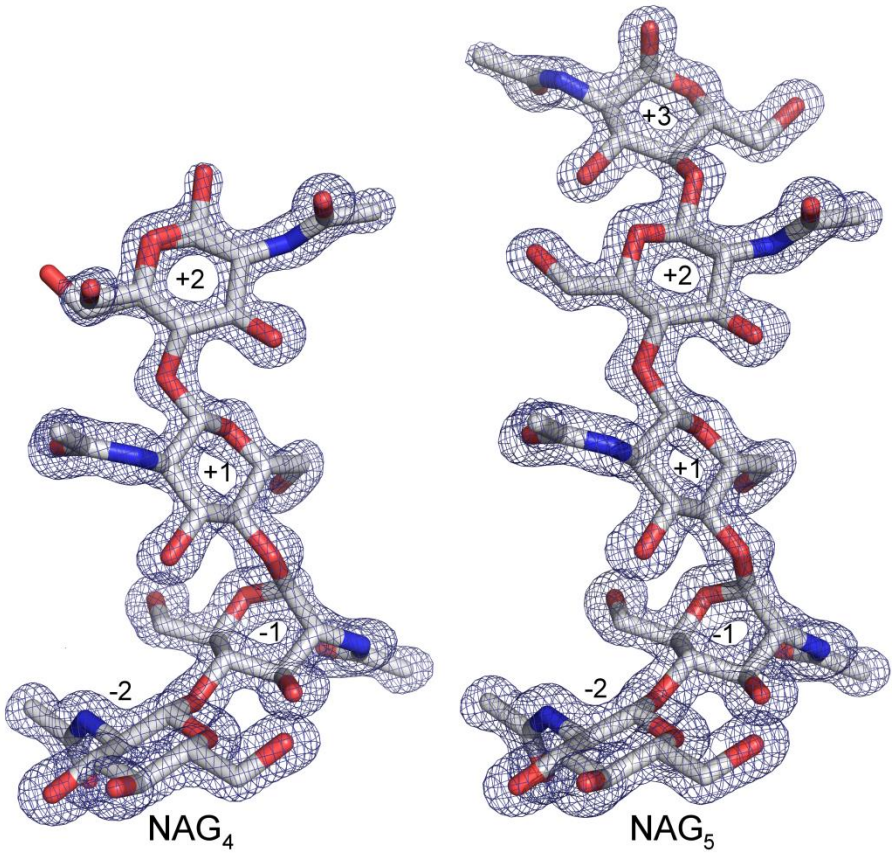


Figure 28. Weighted 2Fo-Fc maps contoured at the 2σ level for ligands bound in the substrate binding groove. The ligands occupy sub-sites from -2 to +2 for crystal soaked with NAG_4 (left) and -2 to +3 after soaking with NAG_5 (right).

In the unliganded E153Q mutant the substrate binding groove is occupied by solvent water molecules and a sodium ion in a position equivalent to the O7 atom of the

acetyl group of NAG in the -1 sub-site. In the crystals soaked with NAG_4 and NAG_5, electron density for the ligands was clearly identified in the substrate binding site. In each structure, a partially occupied (0.6) sulphate ion was found coordinating water molecules bound to the NAG moiety in the -2 site.

Glu153, the active residue in the wild type *Mm*Chi60 was observed previously with the O atom of the side chain stationed against the N atom of the main chain of Asp119 and Ala120 in the nearby flexible loop. In the E153Q mutant the glutamine residue does not occupy the 'away'position seen in the native protein but points toward the substrate-binding groove and is hydrogen-bonded to the second Asp of the DxDxE motif. If a ligand is present it is also H-bonded to the NAG moiety occupying the -1 site. The same position of residue 153 was observed in the active form of the enzyme in complex with the reaction product. The conformations of Asp149 and Asp151 of the DxDxE motif are unchanged in all of the presented structures. In the E153Q mutant, Asp151 is in close proximity (2.7-3Å) to Gln153. The interaction between the two residues is *via* oxygen atoms. The correctness of the modelling was checked by swapping Oε1 and Nε2: the B-factors changed from no difference to a difference of 5 $Å^2$.

SUBSTRATE ACCOMMODATION

The location of the substrate binding groove is similar to other chitinases with the *β/α*-barrel fold but the groove is shallow and clearly defined only in the immediate vicinity of the active site. NAG_4 was observed in the -2 -1 +1 +2 sub-sites (Fig 28). In the crystal soaked in NAG_5, the +3 site was also occupied by a NAG moiety (Fig 28). Similar positioning of the NAG_5 substrate had been observed in chitinase B from *Serratia marcescens* (PDB code 1e6n) in which two NAG subunits were modelled in the glycon, '-' sub-site and three in the aglycon '+' sub-site (van Aalten *et al.*, 2001). This is a different substrate-binding mode from chitinase A (PDB code 1ehn) where only two NAG units were observed in the aglycon site, whereas on the glycon sub-site five and more were modelled (Papanikolau *et al.*, 2001).

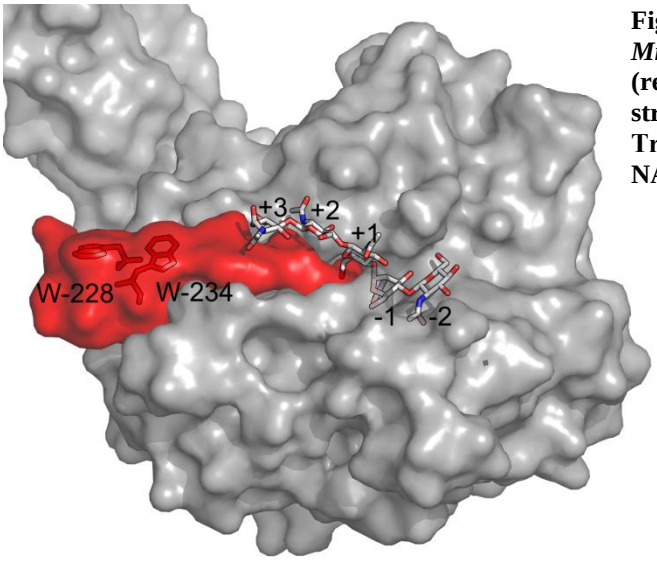


Figure 29. Surface representation of the E153Q *Mm*Chi60 molecule. The insertion domain (residues 218–235) forms a platform-like structure on the aglycon side (red). Trp228 and Trp234 forming the 'Trp patch' are shown. NAG_5 is represented as sticks.

All available chitinase structures in complex with ligands have a tunnel-like aglycon site. In *Mm*Chi60 E153Q there is no tunnel or a clear groove on the aglycon side and the 2nd and 3rd NAG moieties are bound on the flat surface of the protein, hydrogen bonded with the enzyme only on one side, having the other face accessible to the solvent (Fig. 29). All interactions between the NAG units and the enzyme are summarized in Table 32. NAG_4 is closely superimposable on the NAG_5 ligand, with an almost identical conformation of the superposed parts. One difference can be noted in the +2 sub-site, where the terminal sugar ring of the NAG_4 ligand is statically disordered, showing both anomeric states. In addition to the β anomer common to NAG polymers, the alternative configuration of the +2 sugar is α and its O1 atom is hydrogen bonded to the O atom of Asn221. The terminal moiety of NAG_5 is also in the α anomeric state.

The conformations of all the sugar rings can be clearly identified based on electron density maps. In the –2, +1, +2 and +3 sub-sites the sugars adopt the standard chair conformation, whereas the sugar ring in the -1 sub-site is distorted into the boat conformation. Similar distortions were observed previously in other inactive chitinase structures with a bound ligand, or in an active enzyme in complex with the product (Papanikolau *et al.*, 2001, Songsiriritthigul *et al.*, 2008, Tsuji *et al.*, 2010, van Aalten *et al.*, 2001, Winn *et al.*, 2001). The boat conformation makes the sugar chain bent and twisted and helps position the acetyl group of the NAG in the -1 sub-site at the bottom of the active site groove, which is essential for substrate-assisted formation of the reaction intermediate (oxazolinium ion). The angle at which the substrate is bent (measured between C4, C1 of NAG in -1 sub-site and C4 of +1 NAG) is 98°. The average temperature factor of the NAG

units shows that the moieties in the -1 and +1 sites have the lowest values, indicating that they are the best ordered parts of the NAG_5 ligand: 13 $Å^2$ for the -2 site, 11 $Å^2$ for -1, 12 $Å^2$ for +1, 14 $Å^2$ for +2 and 20 $Å^2$ for +3. It is interesting to note that in the substrate, the best ordered NAG moiety is at the active centre, i.e. in the -1 sub-site, whereas in the product (*Mm*Chi60 structure with reaction product), the NAG unit in this position has a higher B-factor than the moiety in the -2 sub-site.

Table 32 NAG_5 ligand interaction with the protein molecule.

NAG moiety	Ligand atom	Bond length (Å)	Residue atom	Residue
-2	O6	3.1	N	Ala118
	O6	2.8	O	Ala118
	O7	3	Nε	Trp311
	O7	2.9	Nε	Trp315
-1	O7	2.6	OH	Tyr 220
	N2	3.1	Oδ	Asp 151
	N2	3.2	Oε	Gln153
	O3	3.1	N	Ala 118
	O6	2.9	Nδ	Asn 221
	O6	2.7	O	Ala 279
1	O6	2.7	Oε	Gln218
	O6	2.9	Nδ	Asn221
	O7	2.7	Oε	Tyr46
2	N2	2.9	Oε	Gln222
	N2	3.2	Oε1	Glu191
	O3	2.6	Oε1	Glu191
3	O7	2.9	N	Gly224
	O6	2.5	Oδ	Asp225
	O1	3	Nε	Gln154 (symm.mol.)

FLEXIBILITY IN THE LIGAND BINDING REGION.

In the *Mm*Chi60 structures the catalytic Glu153 and the flexible loop Gly116-Ile122 were observed to shift toward the ligand upon its binding. In the E153Q mutant, even though the

mutated residue in all cases points toward the binding groove, as observed before in the product-bound state, the loop adopts different conformations depending on whether or not ligand is present (Fig. 30A). In the complex of E153Q *Mm*Chi60 with NAG_5, the O3 atom of NAG in the -1 sub-site is stabilized by the main-chain N atom of Ala118, shifted by 1.8 Å in comparison to the unliganded structure. The same N atom is H-bonded to the O6 atom of the NAG moiety in the -2 sub-site. The carbonyl oxygen of Ala118 also moves by 2.2 Å to partake in bonding the O6 atom of the NAG moiety in the -2 sub-site. The flexible loop in the unliganded E153Q structure is located in an intermediate position between the unliganded native enzyme and liganded mutant (Fig. 30A).

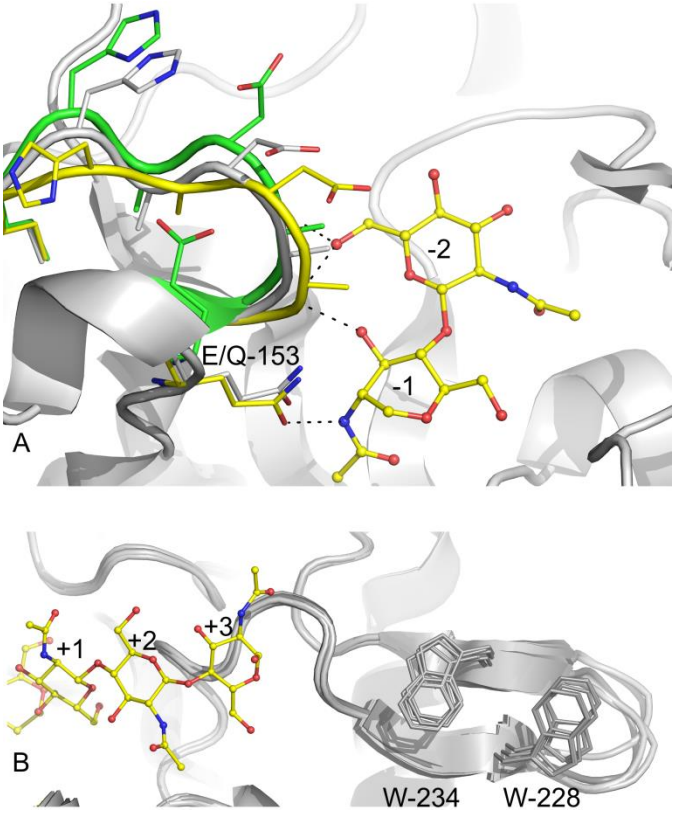


Figure 30. (a) Superposition showing the flexible loop on the glycon side of the substrate-binding site: the wild-type enzyme with the catalytic Glu153 'parked' against the loop in *Mm*Chi60, is shown in green, the unliganded mutant E153Q is shown in grey and the liganded mutant is shown in yellow, showing part of the NAG_5 substrate bound in the -2 and -1 sub-sites. (b) Superposition showing the insertion subdomain forming a tongue-like structure with the Trp patch on its surface. The three E153Q structures are superposed as well as the three *Mm*Chi60 structures. The subdomain exhibits considerable mobility. Part of the NAG_5 substrate is shown on the aglycon side of the substrate-binding site (coloured sticks).

The amino acid segment 218-235 participates in ligand binding on the aglycon sub-side of the substrate binding site. With respect to the overall topology this is an insertion domain on the β/α-barrel, consisting mainly of two anti-parallel β-strands connected by a loop, classified as a β-hairpin. The insertion domain forms a platform-like structure on the surface of the catalytic domain (Fig. 29). The β-hairpin displays mobility of up to 2 Å with

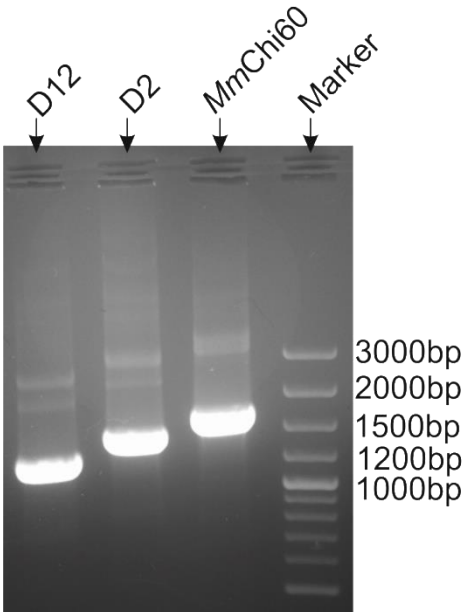
respect to the catalytic domain, seen upon superposition of all mutated and native *Mm*Chi60 structures (Fig. 30B). Residues Gln218, Asn221, Gln222, Gly224 and Asp 225 are hydrogen bonded to the NAG moieties of NAG_5 occupying the aglycon sub-sites. It appears that longer ligands would be blocked from interactions on this side by close contacts with a symmetry-related protein molecule.

The bound substrate on the + site points in the direction of the β-hairpin (residues 226-235). The substrate-binding area extends toward two solvent exposed tryptophan residues Trp228 and Trp234, which can be considered a secondary chitin binding element (the main being the chitin-binding domain).

PLASMID AND DNA MANIPULATION

Mutations were introduced into a plasmid isolated from *E. coli* cells previously used for *Mm*Chi60 expression. Mutation procedures are described in the Materials and Method section. Three constructs were prepared:

- deletion of second Ig-like domain - D2 (E423_A504del)
- deletion of both Ig-like domains - D12 (M348_A504del)
- deletion of both Ig-like domains and chitin-binding domain – D123 (T346_R550del)



The resulting recombined plasmids were verified by agarose electrophoresis using amplified mutated inserts (Fig. 31). The size of all bands corresponds to the length of manipulated gene. DNA sequencing confirmed the gene length after deletion.

Figure 31. UV visualization of an agarose gel of amplified inserts of deletion mutants D12 and D2. As a control, *Mm*Chi60 gene was run together with GeneRuler 100bp Plus DNA Ladder (Thermo Scientific).

PROTEIN PURIFICATION

Recombined proteins were secreted to the periplasm so, similarly to E153Q *Mm*Chi60, cells were broken by osmotic shock and further purified according to Materials and Methods section. The final step before crystallisation was gel filtration. Results for both D12 and D2 mutants are presented on chromatograms (Fig. 32). The purified proteins were checked on SDS-PAGE and DLS which confirmed its purity and homogeneity.

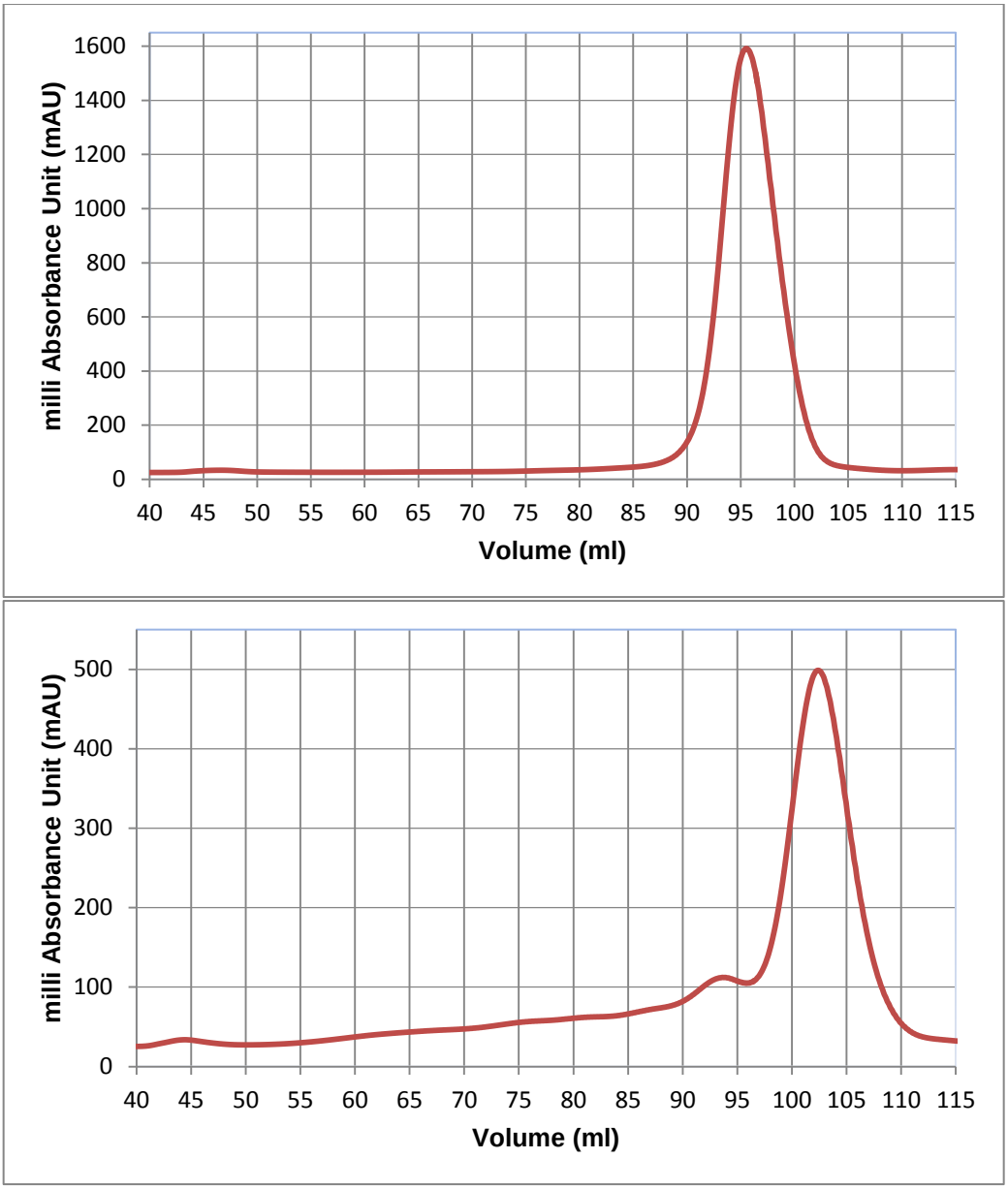


Figure 32 Chromatogram of D2 (top) and D12 (bottom) sample. D2 fractions from the peak with a maximum at 95 ml. were collected and pooled for further experiments. D12 fractions from the peak with a maximum at 102 ml. were collected and pooled for further experiments.

CRYSTALLISATION CONDITIONS

All crystals of *Mm*Chi60 deletion mutant were grown using the hanging-drop vapour-diffusion method at 291 K.

For the D123 the initial crystallisation condition was found in Grid Salt HT crystallisation screen by mixing 0.4µl of 12mg/ml protein solution with 0.4µl well solution.

The well solution consisted of 3.4M sodium malonate at pH 6.0.

For the D2 mutant crystallisation condition was found in Morpheus crystallisation screen by mixing 0.4µl of 10mg/ml protein solution with 0.4µl well solution. The well solution consisted of a mixture of precipitants: 10% w/v PEG 20000, 20% v/v PEG MME 550, a mix of monosaccharides, 0.02M of each: D-glucose; D-mannose; D-galactose; L-fucose; D-xylose; N-acetyl-D-glucosamine and 0.1M MES/imidazole pH 6.5.

For the D12 mutant crystallisation condition was found in JCSG crystallisation screen by mixing 0.4µl of 12mg/ml protein solution with 0.4µl well solution. The well solution consisted of a mixture of precipitants: 45% v/v MPD, 0.2M ammonium acetate and 0.1M Bis Tris pH 5.5.

All of those crystallisation conditions include cryo-protection.

DATA COLLECTION

X-ray diffraction data for the D123 crystal were collected on the beam line 14.2 at the BESSY synchrotron in Berlin, using the marresearch MX-225 detector. For both crystals D12 and D2 data were collected on the beam lines P13 EMBL at the PETRA III synchrotron in Hamburg. The data were integrated and scaled using the XDS software (Kabsch, 2010). Summary of the X-ray data collection and processing is given in Table 33.

SOLVING STRUCTURE & MODEL REFINEMENT

The phase problem was solved by molecular replacement using the Phaser program. The individual domains of the crystal structure of E153Q *Mm*Chi60 (PDB code 4mb3), the catalytic domain, the first Ig-like domain and the ChBD, were used as the search models in case of D2 structure. As the Matthews coefficient for one molecule in the asymmetric unit was calculated to be 6.41 (80% of the solvent crystal content), two and three copies of each domain were searched. After sequential Phaser runs, two complete copies of the D2 mutant were found in the ASU. The final model was obtained after cycles of manual adjustments using COOT and refined with the phenix.refine software. TLS parameters were refined for

fourteen groups. Due to poor electron density in some parts of one catalytic domain, in those regions only the main chain was retained during refinement.

Table 33 Summary of the X-ray data collection and processing

	D2	D12	D234
Space group	R3	$C222_1$	$C222_1$
Unit cell parameters			
a (Å)	225.74	50.7	51.14
b (Å)	225.74	72.69	72.63
c (Å)	65.31	171.19	175
Wavelength (Å)	0.976265	0.976265	0.191841
Oscillation (°)	0.5	0.5	0.3
Number of frames	500	300	600
Crystal-to-detector distance (mm)	503	180	155
Resolution (Å)	112.87-2.68 (2.84-2.68)	85.5-1.32 (1.4-1.32)	19.67-1.48 (1.57-1.48)
R_{merge}	0.09 (0.86)	0.05 (0.95)	0.08 (0.81)
Completeness (%)	98.8 (93.2)	99.1 (95.8)	99.6 (98)
Observed reflections	242568	405015	393315
Unique reflections	34425	74675	54649
$\langle I/\sigma(I) \rangle$	14.06 (1.45)	16.43 (1.72)	14.1 (2.1)
Multiplicity	7.05	5.42	7.2

The D123 mutant was solved by molecular replacement using the catalytic domain as the search model. ADPs were refined isotropically using the Phenix structure refinement software.

For the D12 structure both catalytic domain and ChBD were used by Phaser as the search model. However, only the catalytic domain was found in the asymmetric unit. This is the highest resolution structure available for the catalytic domain of *Mm*Chi60 and ADPs for all atoms accept hydrogen were refined anisotropically. The final refinement statistics are listed in Table 34.

Table 34 The final refinement statistics

	D12	D2	D234
R_{work}/ R_{free}	0.12/0.15	0.2/0.24	0.15/0.20
Protein atoms	2660	6707	2581
Ligand atoms	25	17	11
Water molecules	396	20	342
Average B factor (Å^2)	20	127	21
R.m.s. deviation from ideal			
Bond lengths (Å)	0.022	0.01	0.019
Bond angles (°)	1.37	1.82	1.74
Ramachandran plot (%)			
Favored	97.92	95.8	97.55
Allowed	2.08	3.29	2.45
Outliers	0	0.91	0

D2 STRUCTURE

In the crystal lattice there are two protein molecules in the asymmetric unit, labelled A and B (Fig. 33). The D2 A molecule consists of 446 amino acids while in molecule B 444 residues have been modelled. As designed, protein molecules are arranged in three domains: the N-terminal catalytic β/α-barrel domain, the Ig-like domain and the C-terminal Chitin-binding domain. In the substrate-binding groove additional electron density has been observed and a 2-(N-morpholino)ethanosulfonic acid (MES) molecule and an imidazole have been modelled. MES interacts with Trp residues lining the bottom of the binding groove in the -1 and -2 sub-sites (Trp40, Trp311, Trp315). Only twenty water molecules have been identified in the model.

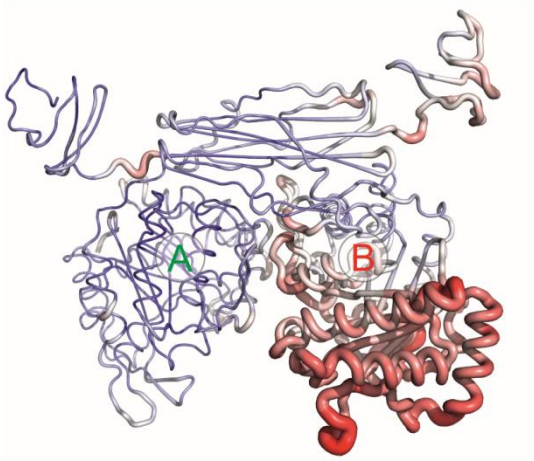


Figure 33. B-factor distribution over molecules A and B in ASU presented in color spectrum blue-white-red, from minimum to maximum (50-220 Å2).

Weak electron density in some parts of the β/α-barrel in molecule B did not allow to model residues' side chains. Neither interactions with symmetry related molecules nor with the second molecule in ASU were observed in this part of the model (Fig. 34) and the B-factors of the many main chain atoms exceed 200Å^2 (Fig. 33).

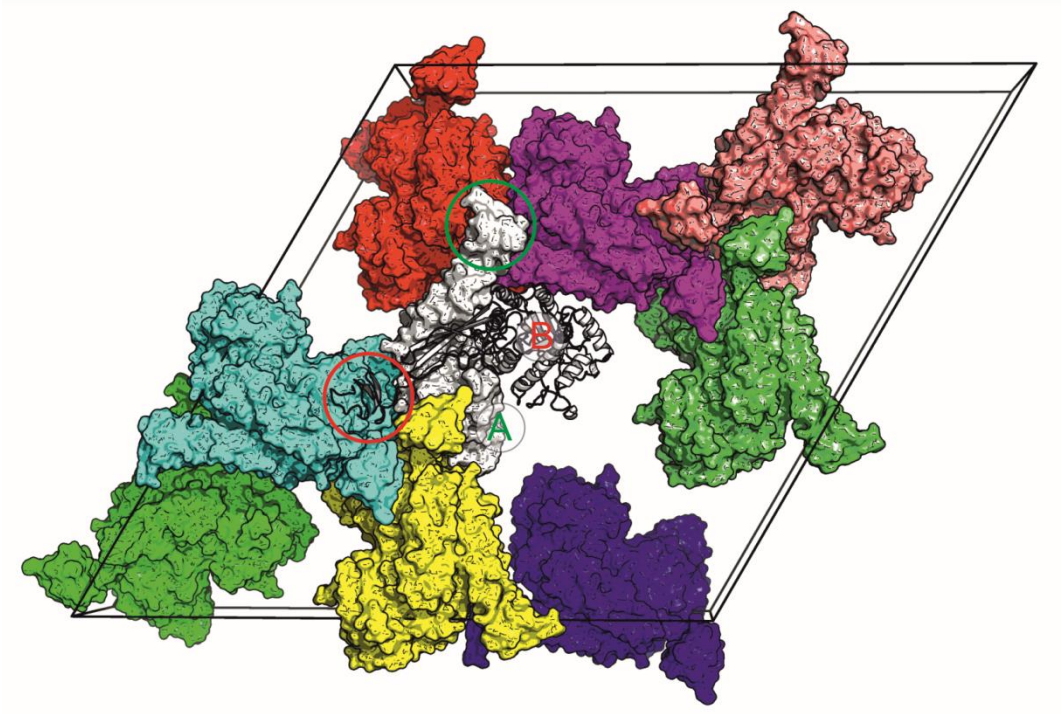


Figure 34. Top. Crystal packing. There are two protein molecules in an asymmetric unit. One crystallographic dimer is marked by letter A (green font) presented as a white surface and B (red fount) presented as a cartoon. Chitin-binding domain of molecule A is marked by a green circle and ChBD of molecule B is marked by a red circle.

DOMAIN POSITION

The individual domains of both molecules in the ASU and in the native *Mm*Chi60 do not differ significantly. The RMS deviations for main chain atoms are listed in Table 35. What is more, the β/α-barrels from the two molecules in the ASU are arranged similarly to the native *Mm*Chi60 catalytic domains related by the crystal two-fold axis.

Table 35 RMS deviations of main chain atoms for corresponding domains.

Compared domains		RMSD Å
D2 β/α-barrel, molecule A	D2 β/α-barrel, molecule B	0.68
D2 Ig-like, molecule A	D2 Ig-like, molecule A	0.37
D2 ChBD, molecule A	D2 ChBD, molecule B	0.34
D2 β/α-barrel, molecule A	*Mm*Chi60 β/α-barrel	0.42
D2 β/α-barrel, molecule B	*Mm*Chi60 β/α-barrel	0.73
D2 ChBD, molecule A	*Mm*Chi60 ChBD	0.34
D2 β/α-barrels, molecule A and B	*Mm*Chi60 β/α-barrel and symmetry related β/α-barrel	2.06

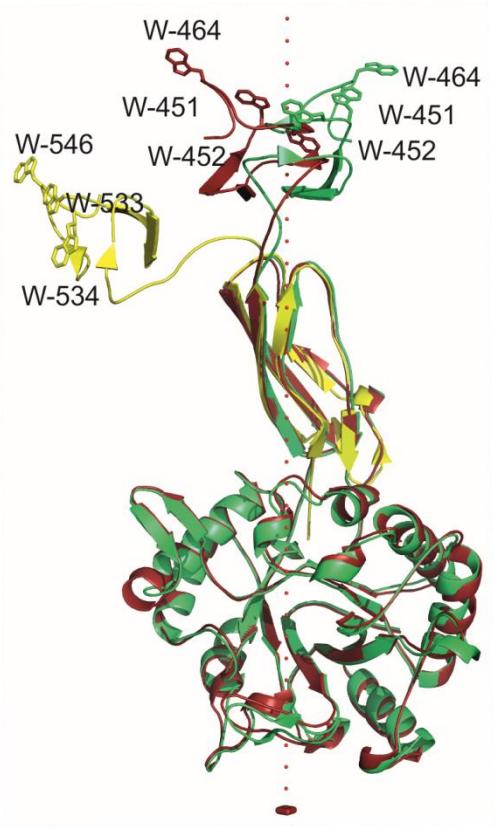


Figure 35 Superposition of the Ig-like domains of the D2 molecule A(green) and B (red) and the Ig-like domain of two C-terminal domains of *Mm*Chi60. The rotation axis of D2 ChBD is presented as red dots.

The position of ChBD in relation with the rest of the mutated protein differs in both molecules. The protein domain motion analysis – DynDom (Hayward & Berendsen, 1998) has shown that the movement of ChBD around the twist axis (the line joining the centre of the mass of a pair of domains) is 138°, the closure is 6% (percentage measure of the degree of closure motion as the square of the projection on the closure axis, perpendicular to the twist axis) and translation -0.1Å. Residues between the Ig-like domain and ChBD, 422-428, were identified as a bending region and are additionally characterised by increased B-factors (Fig 33). The superposition of the second Ig-like domain from *Mm*Chi60 and the D2 molecule has shown that while Ig-like domains overlap, the ChBD of *Mm*Chi60 and molecule B differ in position by in 35Å, counting from the

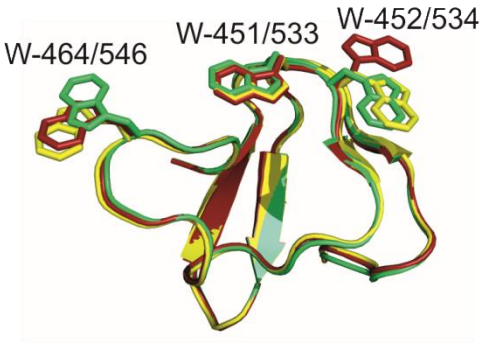


Figure 36 Superposition of ChBD of the deletion mutant D2 molecule A (green), molecule B (red) and MmChi60 (yellow).

middle solvent exposed Trp residues (Fig. 35).

The superposition of ChBD of *Mm*Chi60 and both ChBD of D2 molecules (Fig. 36) shows different arrangements of the solvent exposed Trp residues. In molecule A, Trp464 rotates 180° around the Cβ-Cγ bond while in molecule B, Trp452 rotates 100° around Cα-Cβ bond. However in all cases the co-planarity of all solvent exposed residues is retained.

When comparing the two N-terminal domains of *Mm*Chi60 – the catalytic domain and the Ig-like domain with equivalent domains of the D2 molecule, DynDom server localized the hinge region between the catalytic domain and the first Ig-like domain. The rotation angle of 10°, translation of 0.2Å and 35% closure were identified and residues located between domains 346-349 were selected as a hinge region.

D12 AND D234 STRUCTURES

Each structure consists only of a single β/α-barrel even though the D12 protein used for crystallisation was longer and very likely consisted of two domains – catalytic and ChBD. The RMS deviation of the main chain between D2 and D12 is 0.24Å thus both structures are described together in this paragraph. Each structure consists of 322 residues (23-345) which are very similar to the catalytic domain of *Mm*Chi60 with an RMS deviation of 0.48-0.5Å between them. In the crystal lattice (Fig. 37), the crystal contacts of the catalytic domains of D12 and D123 are different from those observed in *Mm*Chi60 and the D2 mutant, where characteristic crystallographic dimers were identified.

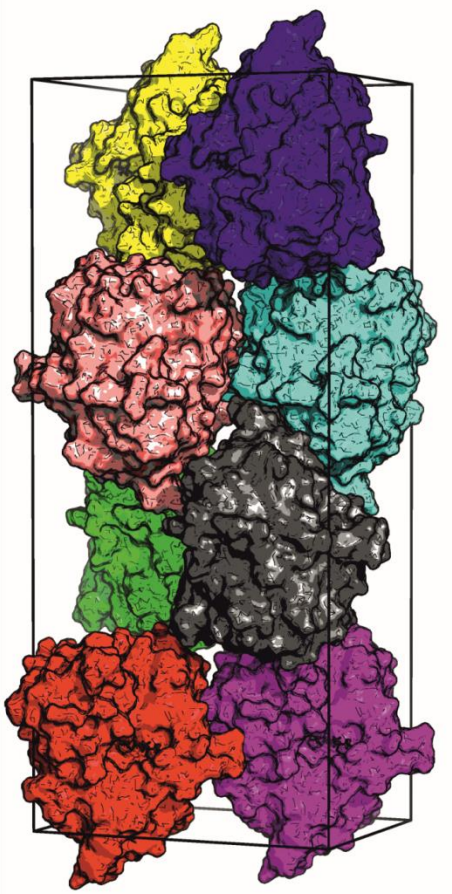


Figure 37. Crystal packing. There is one protein molecule in a asymmetric unit and in harmony with the space group, eight protein molecules in a unit cell.

The D12 structure is the highest resolution (1.32Å) structure of the catalytic domain of *Mm*Chi60. The two cysteine residues present in the sequence (Cys41 and Cys53) form a disulphide bridge as it was observed in other, lower resolution structures. The sodium-binding site identified in *Mm*Chi60 is also occupied in both deletion structures and the Na ion is coordinated by the side chains of Asn105, Asp146, the carbonyl oxygen atom of Gly144, the carbonyl oxygen atom of Thr24. In D12, a 2-methyl-2,4-pentanediol (MPD) molecule has been modelled in the substrate-binding groove. The O4 atom is H-bonded with Nε1 atoms of Trp311 and Trp314 (3Å), the same residues that are engaged in binding the acetyl group of NAG residue in the -2 sub-side in the E153Q structure complexed with NAG_5. The other part of the MPD molecule, the O2 atom, H-bonds Tyr46, which was reported to bind the acetyl group of NAG in the +1 sub-site. Additionally, an acetate molecule interacts with the Asp151 oxygen atom and the hydroxyl group of Tyr220 at similar way as the acetyl group of NAG residue positioned in the -1 sub-side. Additionally seven more sodium cations, two chloride anions and one more acetate have been modelled in various parts of the D12 molecule.

The active site residue (Glu153) is 'parking'against the nitrogen atoms on a flexible loop's nearby, as reported before for an unliganded *Mm*Chi60 structure. The main difference between catalytic domain of unliganded *Mm*Chi60 and the D12 and D123 mutants is on the platform-like structure on the aglycon sub-side of the substrate binding site. The β-hairpin displays mobility up to 2 Å with respect to the catalytic domain, as seen upon superposition. The same situation was observed before upon superposition of all *Mm*Chi60 structures and E153Q mutants.

SMALL ANGLE X-RAY SCATTERING (SAXS) MEASUREMENTS AND ANALYSIS

Synchrotron X-ray solution scattering data were collected at the EMBL P12 beam line (DESY, Hamburg) using a PILATUS detector (Round *et al.*, 2008). The scattering profiles were measured at several solute concentrations ranging from 0.6 to 8.8 mg/ml. Data acquisition was performed at a sample-detector distance of 2.7 m, covering the range of momentum transfer $0.07 < s < 4.6$ nm^{-1} ($s = 4\pi \sin(\theta)/\lambda$ where 2θ is the scattering angle and $\lambda = 0.12$ nm is the X-ray wavelength) in 20 frames of 0.05 seconds each to check for possible radiation damage. SAXS experimental parameters are summarized in Table 36.

Table 36 SAXS experimental parameters

Data-collection parameters	
Beam line	P12
Beam geometry	0.2 x 0.12 mm^2
Wavelength (nm)	0.124
s range (nm^{-1})	0.07-4.6
Exposure time (s)	1
Concentration range (mg/mL)	0.6-8.8
Temperature (K)	283
Structural parameters	
$I(0)$ (relative) [from $P(r)$]	3000 ± 200
R_g (nm) [from $P(r)$]	3.2 ± 0.1
$I(0)$ (relative) [from Guinier]	3000 ± 200
R_g (nm) [from Guinier]	3.2 ± 0.1
D_{max} (nm)	11 ± 0.5
Porod volume estimate (nm^3)	75 ± 10
Dry volume calculated from sequence (nm^3)	70
Molecular-mass determination	
$I(0)$ (relative) for BSA	4500 ± 200
Molecular weight *MW* (from $I(0)$) (kDa)	47 ± 10
Calculated monomeric *MW* from sequence (kDa)	58.6

The data collection was performed by the author of the thesis however it was processed and analysed by Maxim V. Petoukhov and Dmitri I. Svergun from European Molecular Biology Laboratory in Hamburg. Primary data reduction was performed by automated pipeline (Franke *et al.*, 2012) and comprehensive analysis of the scattering profiles was accomplished using the ATSAS software package (Petoukhov *et al.*, 2012). The forward scattering *I(0)* and the radii of gyration R_g were evaluated using the Guinier approximation, assuming that at very small angles ($s < 1.3/R_g$), the intensity is represented as $I(s) = I(0) \exp(-(sR_g)2/3)$. The maximum dimension D_{max} was computed using the indirect transform package GNOM (Svergun, 1992). Molecular weight (*MW*) estimate was made by comparison of the forward scattering with *I(0)* of a bovine serum albumin (BSA) standard. The scattering from the crystallographic monomer of *Mm*Chi60 $I_{calc}(s)$ was calculated using the program CRYSOL (Svergun *et al.*, 1995) which also fitted the experimental data $I_{exp}(s)$ by adjusting the excluded volume and the contrast of the hydration layer to minimize the discrepancy:

$$\chi^2 = \frac{1}{N-1}\sum_{j=1}^{N}\left[\frac{I_{\exp}(s_j) - cI_{calc}(s_j)}{\sigma(s_j)}\right]^2,$$

where *c* is a scaling factor, *N* is the number of points and σ denotes the experimental errors.

An ensemble optimization method (EOM) (Bernado *et al.*, 2007) has been applied to assess the flexibility of *Mm*Chi60 in solution. In the EOM approach, a large pool (of 10,000 *Mm*Chi60 chains) with random orientations of individual domains in respect to each other has been generated and the optimized ensemble has been selected by genetic algorithm which averaged-intensity fits the experimental SAXS data. The distribution of the structural descriptors (*i.e.* R_g, D_{max}) in the optimized ensemble with respect to the original random pool provides an idea of the conformation variability of the macromolecule in solution.

CONFORMATIONAL FREEDOM IN SOLUTION

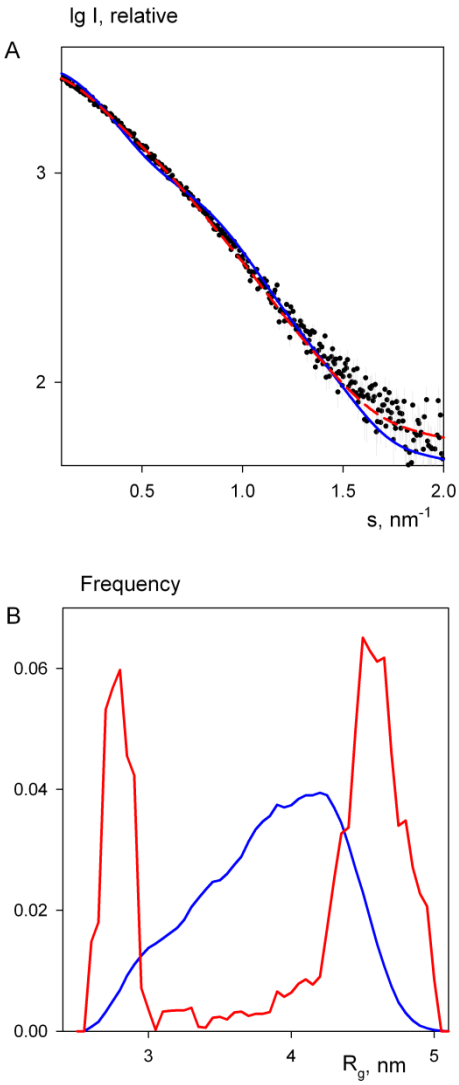


Figure 38. (a) SAXS profiles of *Mm*Chi60. Experimental data are denoted by black dots with the fit from the crystallographic monomer shown as a blue solid line and the fit from the EOM ensemble as a red dashed line. (b) R_g distribution of *Mm*Chi60 in solution obtained by EOM. Random pool, blue; distribution in the optimized ensemble, red.

The experimental SAXS profile of *Mm*Chi60 is presented in Figure 38(a) and the corresponding structural parameters are summarized in Table 36. The SAXS data point to a monomeric state of *Mm*Chi60 in solution, even though the crystal packing of *Mm*Chi60 and its deletion mutant D2 suggests a dimeric form. The overall size and radius of gyration (R_g) computed from the rather elongated crystallographic monomer are noticeably larger than the experimentally deduced values in solution. The scattering profile computed by CRYSOL from the atomic model of *Mm*Chi60 displays several shoulders (Fig. 38A), whereas the experimental curve demonstrates a rather smooth behaviour. These deviations between the experimental and computed profiles yield a discrepancy of χ = 1.7. The crystal structures show that *Mm*Chi60 could possess considerable conformational freedom. The discrepancy between the computed and experimental SAXS profiles, in which the former curve oscillates around the latter curve, is a typical result of flexibility in solution. Indeed, it has been shown in a theoretical study (Bernado, 2010) that domain motion often results in 'smearing' of the SAXS curve.

To check the hypothesis of inter-domain flexibility, the ensemble-optimization method (EOM) was applied. The full-length protein was divided into four individual domains and these were allowed to adopt random orientations with respect to each other to sample the conformational space. The distribution of the R_g values in the original random pool is

presented in Fig. 38B (blue line). The optimized ensemble consisting of compact, completely extended and intermediate conformations yielded a good fit to the experimental data for *Mm*Chi60 with χ = 1.2 (Fig. 38a). The corresponding R_g distribution presented in Fig. 38B (red line) displays two major populations: completely extended (R_g around 4.7 nm) and rather compact (R_g around 2.7 nm). Overall, the SAXS results suggest a high degree of flexibility in the hinge regions of *Mm*Chi60 in solution (Fig. 39). Interestingly, the intermediate conformation displays an overall shape rather similar to the crystal structure. The intermediate conformation appears to be present in solution in smaller amounts and manifests itself as a shoulder of the peak corresponding to the completely open structure.

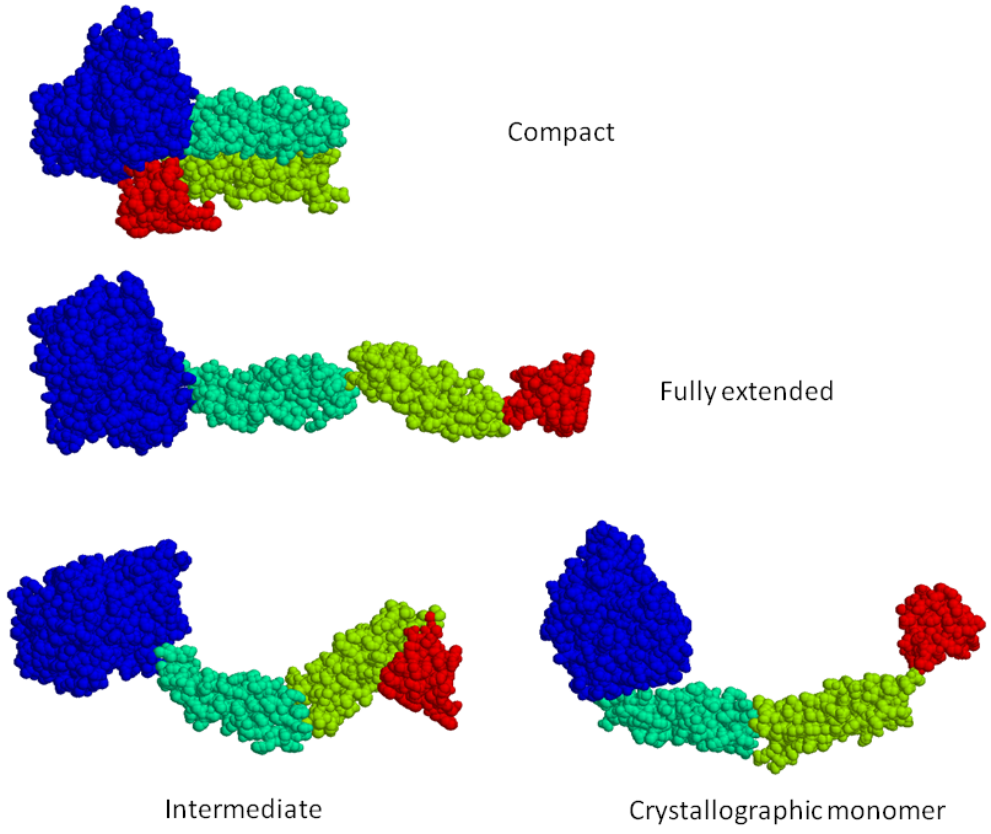


Figure 39. Three conformations of *Mm*Chi60 from the EOM set consistent with the SAXS data and the crystallographic monomer.

PROTEIN PURIFICATION

The gene for a thermostable Chi40 from *Streptomyces thermoviolaceus* was cloned together with the *pel*B secretion peptide as describes by Christodoulou *et al*. (Christodoulou *et al.*, 2001). *Pel*B refers to a 22 amino-acid N-terminal leader sequence of pectatelyase B of *Erwinia carotovora (Lei et al., 1987)* and when attached to a protein, directs the protein to the periplasm and out of the cell, and the sequence is removed by a signal peptidase. The protein produced from this construct was purified both from the periplasm, by an osmotic shock, and from the culture medium, where its occurrence was predominant.

Purification procedure was performed as described in the Material and Methods. Size exclusion chromatography separated minor impurities and aggregates, and yielded a homogenous and a pure protein. Samples were collected from fractions with the maximum UV absorbance around 57th ml. at the chromatogram peak (Fig. 40).

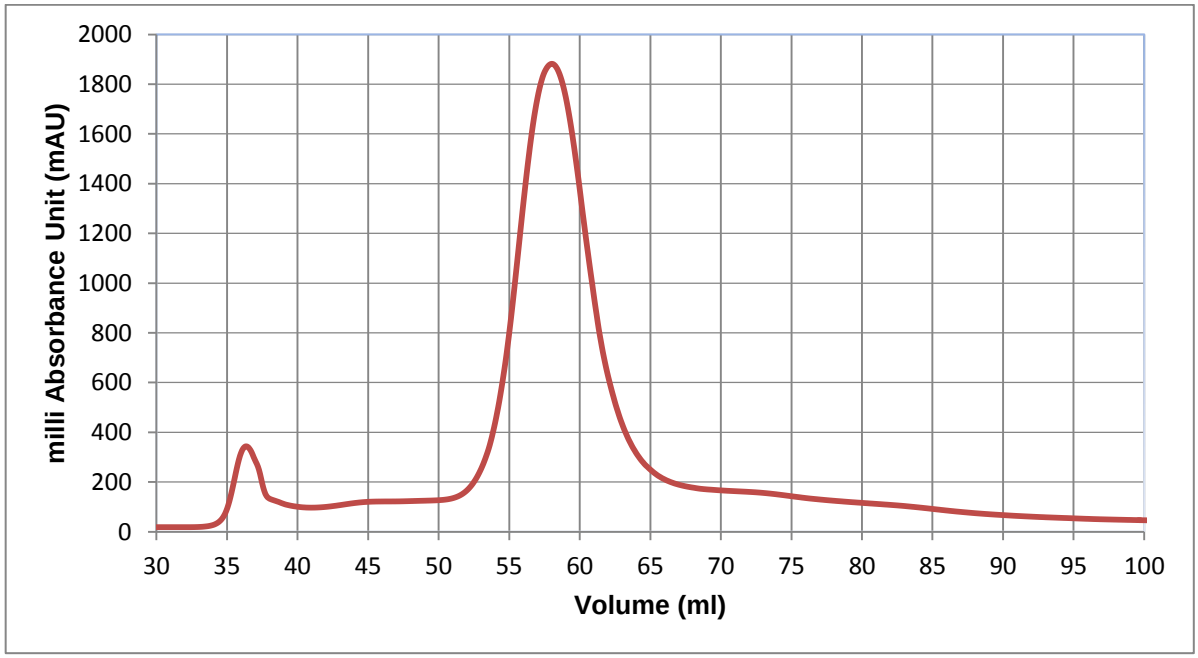


Figure 40. Chromatogram of Chi40 sample. Chi40 fractions from the peak with a maximum at 57 ml. were collected, pooled together for further crystallisation experiments. Minor aggregates were separated around 36 ml.

CHI40 MUTANTS AND THEIR PURIFICATION.

Four additional N-terminally truncated genes coding proteins starting at different residues were prepared:

- 29^{th} – M1_28Sdel
- 32^{nd} – M1_31Vdel
- 36^{th} – M1_35Adel
- 39^{th} – M1_38Ddel

LIC method was incorporated as described in the Material and Methods. All inserts were cloned into both pMCSG7 (His-tag) and pMCSG48 (NusA and His-tag) plasmids. Three colonies of each preparation were tested for protein overproduction and subsequently one of each *E.coli* strain holding a plasmid with the desired mutation was selected: 29^{th} with in pMCSG48 vector and 32^{nd}, $36^{th,}$ 39^{th} in pMCSG7 (Fig. 41). Purification was scaled up and finally, soluble and homogenous Chi40 truncated mutants were obtained and concentrated to 20mg/ml.

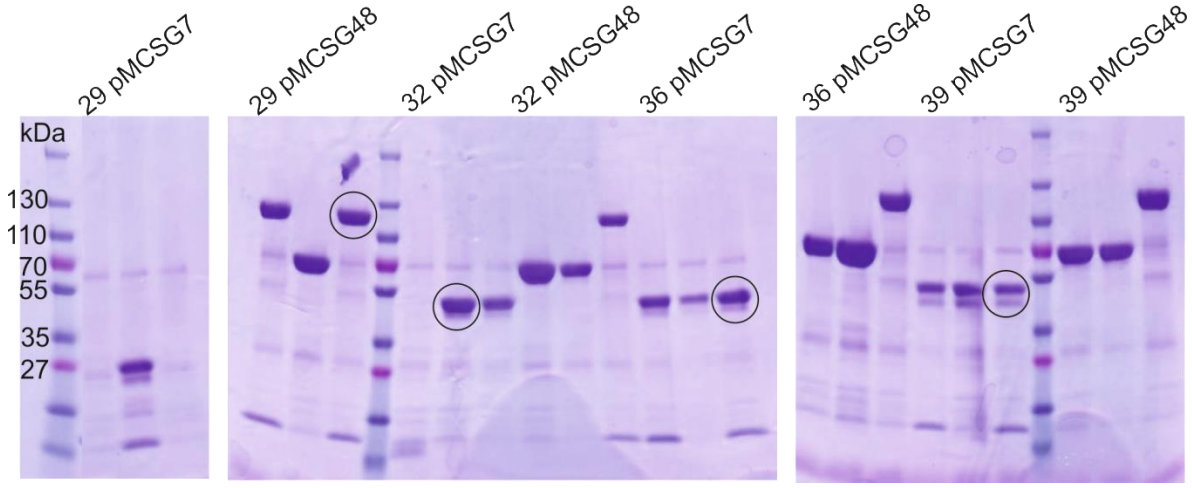


Figure 41. SDS-PAGE electrophoresis of screening for Chi40 deletion mutants. Circles indicate desired protein products. The 40 kDa band correspond to Chi40 mutants with a His-tag (pMCSG7). At 120kDa is Chi40 fused with NusA (pMCSG48). The 70kDa band is NusA without the Chi40 insert. The minor 27kDa band is lizosyme used for cell lysis.

CRYSTALLISATION CONDITIONS

Many crystallisation trials of Chi40 were conducted using the Art Robbins Gryphon crystallisation robot. A large number of crystallisation plates were set up using Chi40 isolated form the supernatant but as the result, only one protein crystal grew after approximately 18 months in condition where at the beginning well solution contained 1.9 M sodium malonate at pH 5.0. Crystallisation drops were prepared by mixing the 30mg/ml protein solution with the well solution in an equal ratio.

Reproduction of crystallisation conditions has not yielded crystals yet. Testing various Chi40 N-terminal mutants in various concentrations, mixing with the reaction products and substrates, changing crystallisation temperatures (4°C, 15°C, 19°C, 30°C) and using a large range of crystallisation screens gave negative crystallisation results.

DATA COLLECTION AND PROCESSING

X-ray diffraction data were collected on the beam line I911-3 at the MAX-lab synchrotron in Lund, using the marresearch MX-225 detector. Due to only one crystal available but thanks to the size of the crystal, three separate data sets were collected from distinct parts of the crystal (Fig. 42). Close inspection of the frames and processing with various parameters allowed the best data set to be selected. The data consisted of 320 images; however a beam injection took place during the data collection after frame 70. The data were integrated and scaled using the XDS (Kabsch, 2010) using frames from 71 to 240. Out of the 53657 strong reflections selected for indexing, only 27986 have been indexed. An attempt to separately index the remaining spots succeeded only partially, because almost half of them belonged to a different crystal lattice. This explains the diffraction pattern which is the result of twinned of the Chi40 crystal. Summary of the X-ray data collection and processing of the best data set is given in Table 37.

SOLVING STRUCTURE & MODEL REFINEMENT

The phase problem was solved by molecular replacement using the Phaser program

(McCoy *et al.*, 2007) and a model structure as the search model. The model structure was generated by the GeneSilico Metaserver, using various methods for protein prediction (Kurowski & Bujnicki, 2003) mainly based on homology modelling. Matthews's coefficients were calculated: 8.02 (84.67%), 4.01 (69.33%), 2.67 (54%) and 2 (38.67%) for respectively 1 to 4 39.5kDa molecules. Finally, Phaser found two protein molecules in the ASU, indicating a solvent content of almost 70%.

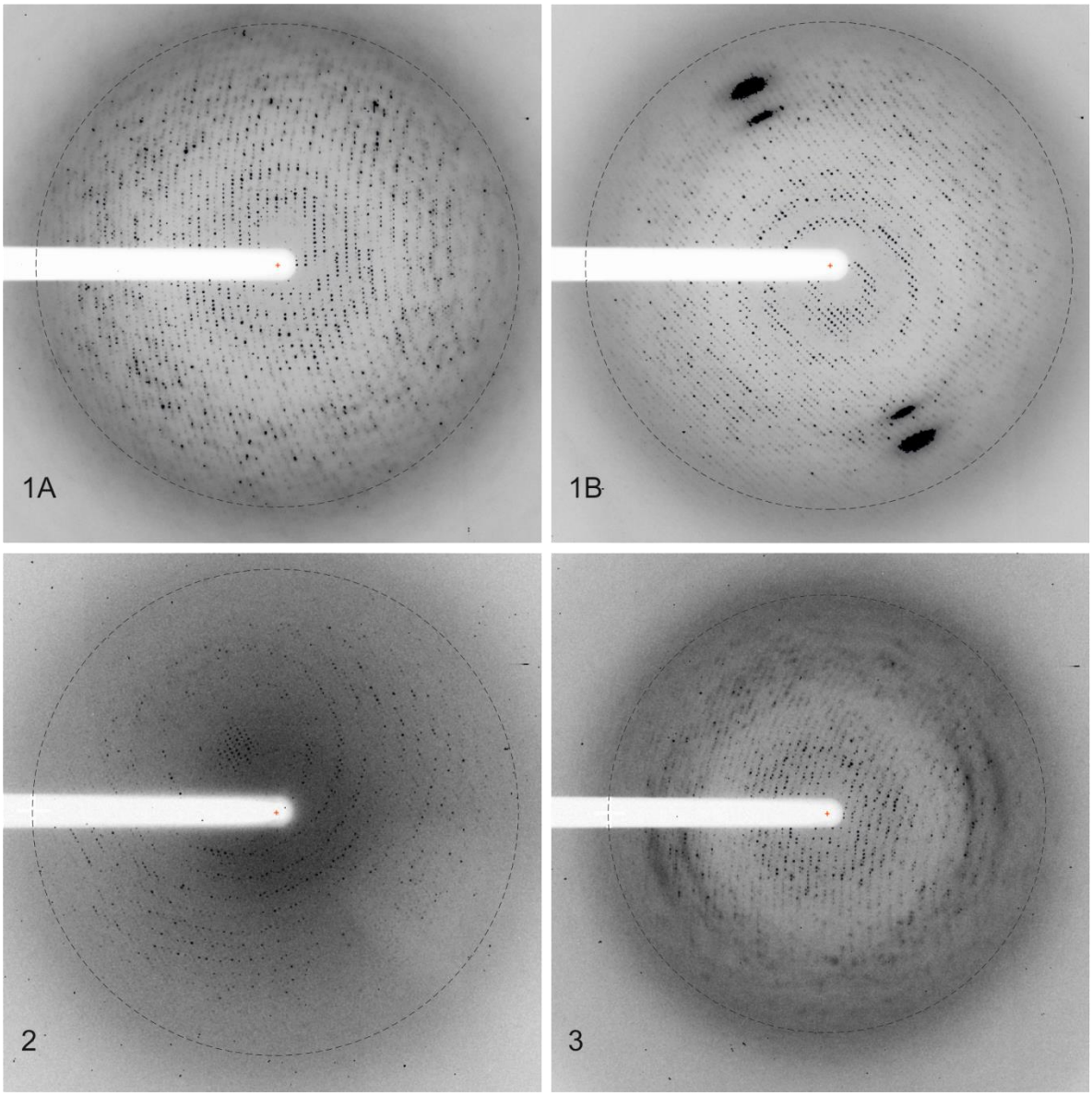


Figure 42. Examples of diffraction frames from three data sets collected from the Chi40 crystal. Dashed circle is a 3Å resolution ring. 1A is a first image (71st frame) in the working data set.1B frame after 40° crystal rotation (150th frame). 2 and 3 are images from the second and third data sets.

The model was refined using the phenix.refine software (Afonine *et al.*, 2012). TLS parameters (Winn *et al.*, 2001) were refined for both protein molecules: four and five groups accordingly. NCS related atoms were identified automatically and restrained in refinement. The final statistics are listed in Table 37.

Table 37 Summary of Chi40 X-ray data collection and final refinement statistics

Summary of the X-ray data collection	
Space group	$P6_122$
Unit cell parameters	
a =b (Å)	183.3
c (Å)	130.9
Wavelength (Å)	1.000
Oscillation (°)	0.5
Number of frames	170
Crystal-to-detector distance (mm)	277
Resolution (Å)	49.05-2.77 (2.94-2.77)
R_{merge}	0.15 (1.01)
Completeness (%)	99.4 (98.1)
Observed reflections	333347
Unique reflections	33240
<I/σ(I)>	13.53 (1.91)
Multiplicity	10.03

The final refinement statistics	
R_{work}/ R_{free}	0.15/0.20
Protein atoms	5584
Ligand atoms	21
Water molecules	116
Average B factor (Å^2)	62
R.m.s. deviation from ideal	
Bond lengths (Å)	0.017
Bond angles (°)	1.63
Ramachandran plot (%)	
Favoured	95.93
Allowed	3.93
Outliers	0.14

OVERALL STRUCTURE

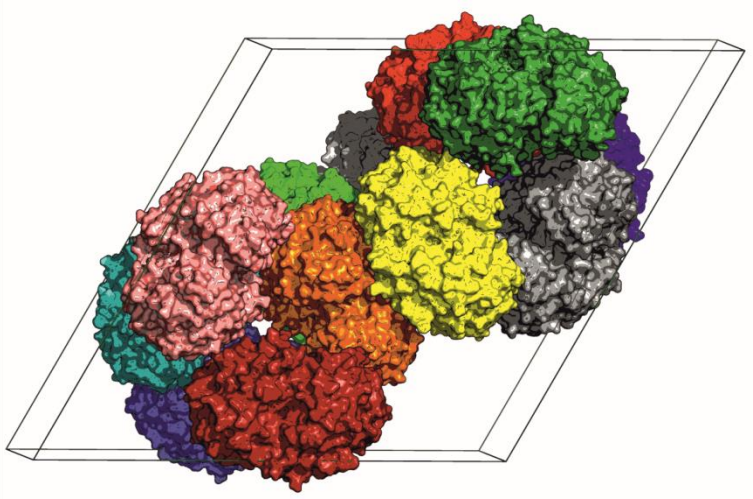


Figure 43. Crystal packing. There are two protein molecules in an asymmetric unit showed at the same colour.

There are two molecules of Chi40 in the asymmetric unit facing each other with their catalytic sides, related by non-crystallographic two-fold symmetry. In the crystal lattice in cell unit, the dimers of Chi40 molecules are packed around the 6_1 axis (Fig. 43). The model of molecule A consists of 369 amino acid residues (40-408) and molecule B consists of 370 amino acids residues (39-408). Both molecules are very similar with the RMS deviation of 0.23Å for the main chain. The protein consists of the N-terminal catalytic β/α-barrel domain (classed as GH18) and an α+β insertion that is mainly composed of five antiparallel strands and an α helix. The insertion domain is located between strand β7 and helix α7 and additionally consists of a loop fragment between strand β6 and helix α6 (Fig. 44, 45).

The most similar structures to Chi40, with identity from 34% to 29% within the Chi40 domains are: the catalytic domain of the psychrophilic chitinase B from *Arthrobacter* TAD20 (PDB code 1kfw), the catalytic domain of chitinase A1 from *Bacillus circulans* WL-12 (PDB code 1itx), chitinase CrChi1 from nematophagous fungus *Clonostachys rosea* (Yang *et al.*, 2010) (PDB code 3g6l), chitinase 1 from *Coccidioides immitis* (Hollis *et al.*, 2000)(PDB code 1d2k), chitinase from *Vibrio harveyi* (Songsiriritthigul *et al.*, 2008) (PDB code 3B8S) and ChiA from *Serratia marcescens* (Perrakis *et al.*, 1994) (PDB code 1ctn). The two last have an N-terminal FN-III domain that is not present in Chi40.

The electron density of the main chain is continuous. Despite the low resolution data, the map of the side chains is clear which allowed observing mismatched residues as the result of wrong primary sequence. Those residues were replaced with correct ones based on the new data from the gene sequencing.

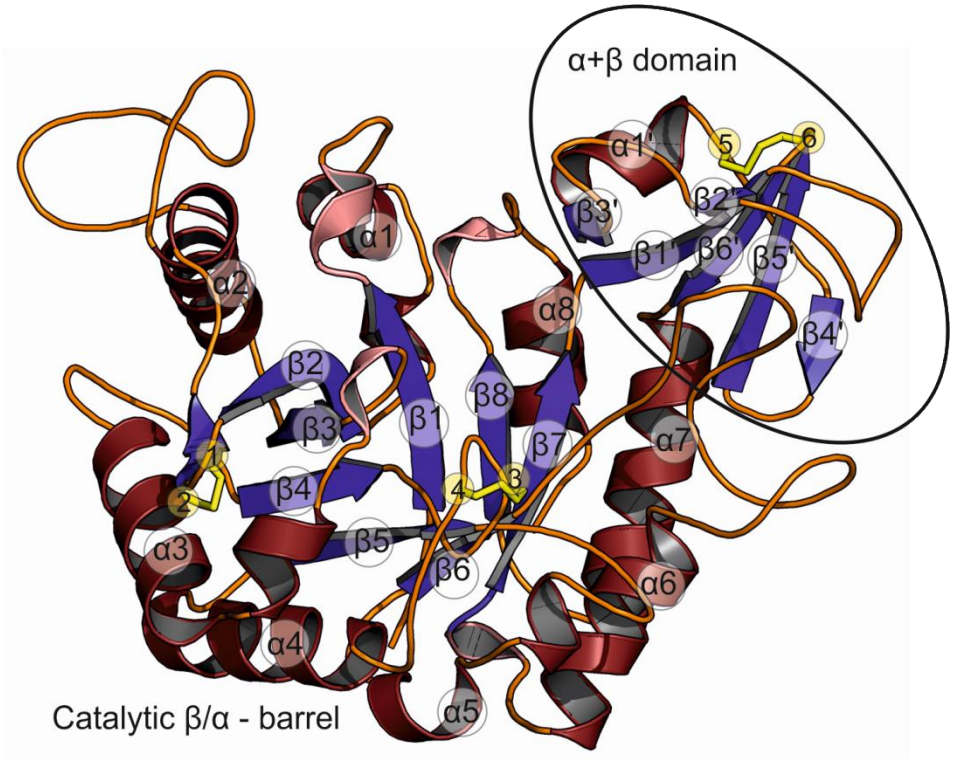


Figure 44. Cartoon plot of Chi40. The two protein domains are labelled. The secondary structure elements – α-helices (red), β-sheets (blue), 3_{10}-helices (pink) – of the catalytic domain are labelled according to the convention for the β/α-barrels. The α + β insertion domain is marked by an oval. Cysteine residues forming disulphide bridges are labelled (yellow sticks) and numbered: 1 – Cys86, 2 – Cys162, 3 – Cys188, 4 – Cys192, 5 – Cys343, 6 – Cys357.

Based on all the previously listed structures many truncations can be identified in Chi40. A short deletion, 2-3 amino acids, is located on the loop proceeding helix α3. In Chi40 the junction between the helix α4 and strand β5 is 7-9 residues shorter than all others.

However, the largest differences are between Chi40 and sequentially most similar chitinase from *Arthrobacter*. Many deletions are visible after sequence alignment; all in

Chi40 (see alignment, Fig. 45). The location of those additional parts not present in Chi40 are visible as green tubes on a Figure 46.

```
Chi40   MKYLLPTAAAGLLLLAAQPAMAMATDHSPTVETRAAADNGTVKLGYFTEWGTYDRNFNVK  60
1KFW    ------------------------------PLTSTVNGYRNVGYFAQWGVYGRAFQAK  37

Chi40   NLDTSGTAAKITHINYAFGNVT-----------------GGKCAIGDSYADYDKAFTADQ 103
1KFW    QLDVSGTAKNLTHINYSFGNINNQTLTCFMANKAQGTGPNGSDGAGDAWADFGMGYAADK  97

Chi40   SVSGQADTWDQPLRGNFNQLRQLKAKYPHIKVLWSFGGWTWSGGFADAAK---DPQGFAQ 160
1KFW    SVSGKADTWDQPLAGSFNQLKQLKAKNPKLKVMISLGGWTWSKNFSKAAATEASRQKLVS 157

Chi40   SCYNLVHD---PRWDG---------VFDGIDIDWEYP-NACGLTCD----SSGPDAFRNL 203
1KFW    SCIDLYIKGNLPNFEGRGGAGAAAGIFDGIDIDWEWPGTNSGLAGNGVDTVNDRANFKAL 217

Chi40   MAALR---STFGDELVTAAVTADGTPG--GKIEATDYAGAAQY--VDWYNVMTYDFFGAW 256
1KFW    LAEFRKQLDAYGSTNNKKYVLSAFLPANPADIDAGGWDDPANFKSLDFGSIQGYDLHGAW 277

Chi40   D--AQGPTAPHSPLTSYDGIPKQGFTSADAIAAFKAQGVPADKLLLGIGFYGRGWTGVTQ 314
1KFW    NPTLTGHQANLYDDPADPRAPSKKFSADKAVKKYLAAGIDPKQLGLGLAAYGRGWTGAKN 337

Chi40   DAPGGTATGPAAGTWEQGIEDYKVLKNT-CPVTGTVAGTAYAHCGSNLWSYDTPDTIASK 373
1KFW    VSPWGPATDGAPGTYETANEDYDKLKTLGTDHYDAATGSAWRYDGTQWWSYDNIATTKQK 397

Chi40   MAWANDQGLRGAFAWDFSGDTADGELIAALSNGLA------------- 408
1KFW    TDYIVSKGLGGGMWWELSGD-RNGELVGAMSDKFRAAAPGPVTEAAPP 444
```

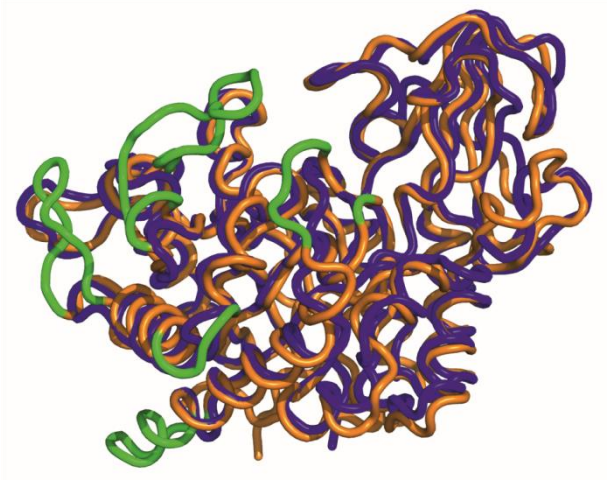


Figure 45. Alignment of Chi40 and chitinase from *Arthrobacter*. Similar (light blue) and identical amino acids (dark blue) are coloured. Gaps are marked ash dashes.

Figure 46. Cartoon presentation of superimposed Chi40 (blue) and chitinase from *Arthrobacter* (orange and green). Fragments absent in Chi40 are colored green.

The helix following the strand β2 is worth mentioning because it is composed of two helices separated by a large loop (96-117), however it resembles a continuous one (Fig. 45). The oxygen atoms: O93 and O94 from the top of the first part of helix form a helix like interactions with the bottom of the second part of the helix receiving hydrogen atoms from

N119 and N120 respectively. This long loop between parts of the helix extend the substrate-binding patch from the + sub-site supplying the solvent exposed Trp112 residue at a fringe of the protein.

DISULPHIDE BRIDGES

All cysteine residues present in the sequence form intra-chain disulphide bridges (Fig 44, 47). The first pair (Cys86 and Cys162) links the small β-strand which is right after strand β2 and a sequentially distant helix α3. This β-strand is a part of the β- hairpin together with the C-terminal end of the proceeding strand β2 which sticks out from the barrel. Both sulphur atoms in cystine, S86 and S162 interact with aromatic rings at a distance 3.6-3.9Å to the centroid of Phe139 and 3.4-3.5Å to the centroid of Phe158 respectively. The sulphur atoms lies S atoms lie close to the normal to the aromatic ring planes. In an equivalent position in the chitinase from psychrophilic *Arthrobacter* disulphide a bridge has been observed, but with no aromatic interactions. Additionally Phe139 is unique in Chi40 among the most similar structures listed above.

The second pair of cysteine residues (Cys188, Cys192) makes a disulphide bond at the base of the β turn (Fig. 44, 47). The cystine is fully exposed to the solvent. This β turn at the top interacts with the conserved substrate binding Trp142 by a hydrophobic interaction. This disulphide bond is unique among the comparable chitinases.

The third disulphide bond (Cys343 and Cys357) is localized on the insertion domain (Fig. 44, 47). This covalent bond connects the end of helix α1' with end of strand β5'. The sulphur atoms from residue 343 lies close to the plane of Trp362 aromatic ring within 3.3 – 3.5Å distance from Cδ1 atom. This disulphide bond is also unique among the most similar chitinases.

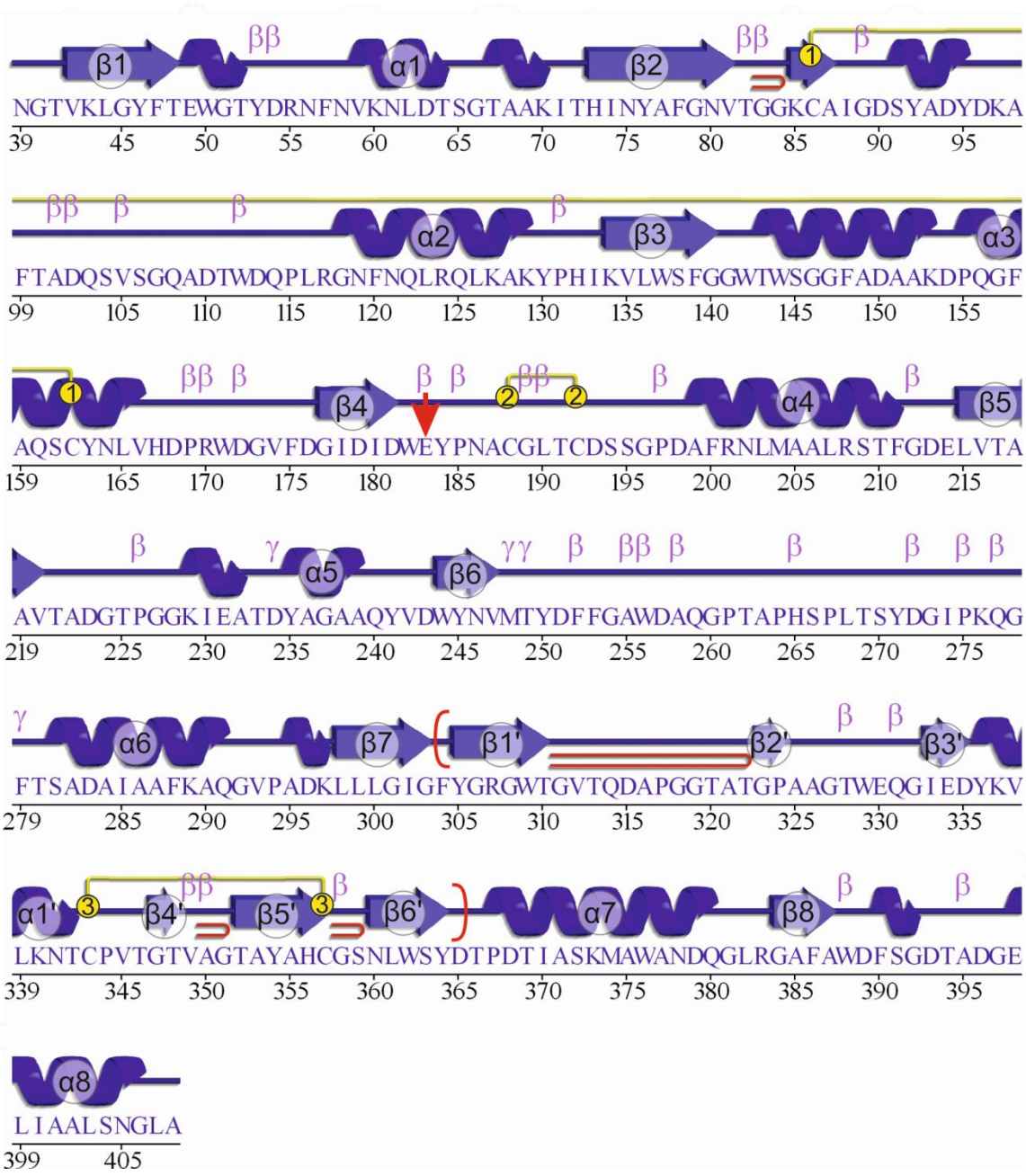


Figure 47. Secondary structure elements of Chi40. β and γ turns (purple letters) are indicated as well as β-hairpins (red hairpin). Helices, β-strands and loops are numbered accordingly to the topology of the β/α-barrel. Half cystines are marked by yellow circles and paired by yellow lines. The active site residue, E183, is marked by a red arrow and the α+β insertion domain is in red brackets.

SUBSTRATE-BINDING GROOVE

The substrate-binding groove is located at the carboxyl ends of the β-strands of the β/α-barrel, similar to *Mm*Chi60 and other enzymes with the TIM barrel architecture. Residues lying along the loops connecting the β-strand with helixes create the substrate binding cleft. The groove is deep and clearly outlined however Chi40 lacks the residues forming the other side of the substrate binding cleft located below -4 sub-side, which is observed in ChiA from *S. marcescens* and chitinase A1 from *B. circulans*.

In Chi40, the groove is occupied only by a few water molecules. In the space between two protein molecules in the ASU, ligand molecules have been modelled. Malonate molecules have been identified in the vicinity of Trp142 and Trp256 in both molecules. A positive peak of electron density was observed at the junction between Trp50 and Trp144 from both protein molecules, on the 2-fold non-crystallographic symmetry axis where the third malonate and water molecules have been modelled; however the electron density is not well defined and it seems that the ligand is in multiple conformations.

Even though Chi40 structure with ligand was not solved, the residues lying along the substrate binding groove together with the active side residue are easily identified based on a structure comparison with an inactive ChiA in complex with NAG8 (Papanikolau *et al.*, 2001)(PDB code 1ehn). The residues forming H-bonds and stacking interactions with the ligand are highly conserved among the most similar chitinases listed above. In Chi40 Glu183 is a catalytic residue, a part of the DxDxE signature, and its side-chain is directed ti the ligand binding groove near the scissile bond. The second D, Asp181 is conserved and is supposed to bind positively charged nitrogen atom from the oxazolinium ring in the -1 sub-site, as observed in the *Mm*Chi60 complex with a product intermediate. According to the ChiA structure in complex with a poly-NAG, two Trp residues, 256 and 142, form a sandwich-like cleft for a ligand. Tyr184, which substitutes Phe316 from ChiA, has an additional hydroxyl group to bind a ligand in the +1 sub-site.

The '–' side of the substrate-binding groove is lined with aromatic residues, conserved in all the similar chitinases: Trp 388 and Trp 50 continuing with Tyr53 and Tyr92. The last aromatic residue on this patch is Trp112. There are several other residues in positions equivalent to those interacting with ligand in ChiA: Glu330, Trp142, Thr143,

Trp329, Arg55. Chi40's Asp54 that points towards the substrate binding groove is substituted by Gly in other structures.

SEQUENCE CHANGES AND THEIR STRUCTURAL CONSEQUENCES

Potential factors stabilising the structure have been examined. First, all salt bridges have been listed (Costantini *et al.*, 2008) and analysed it terms of conservation. Among the most similar structures, comparing only superimposable fragments, the number of interactions between positively charged residues (His, Lys, Arg) with negative (Glu, Asp) in Chi40 is lower: eleven compared to twelve and more in others. Among the salt bridges observed in Chi40, five are present in similar structures (Table 38).

Table 38 Salt bridges in Chi40 chain B conserved among other chitinases

Residue 1	Residue 2	Distance
NH1 ARG B 55	OD1 ASP B 335	3.75
NH1 ARG B 55	OD2 ASP B 335	3.23
NZ LYS B 127	OD1 ASP B 176	2.72
NZ LYS B 127	OD2 ASP B 176	3.67
NH1 ARG B 208	OD1 ASP B 243	3.96
NH1 ARG B 208	OD2 ASP B 243	2.84
NH2 ARG B 208	OD2 ASP B 243	3.92
NH2 ARG B 307	OD1 ASP B 251	2.89
NH2 ARG B 307	OD2 ASP B 251	3.88
NZ LYS B 373	OD1 ASP B 365	3.76
NZ LYS B 373	OD2 ASP B 365	2.86

In some positions, residues conserved in other proteins are substituted in Chi40. For example, instead of Arg or Lys commonly observed on position 167, in Chi40 a histidine residue is present and its nitrogen atoms interact with Oδ1 of Asp172 and Oε1 of Glu214. It is worth mentioning that Glu214 taking part in this interaction is also unique because an aromatic residue is usually located in this position. Pro155 and Pro264 are unique and they substitute the commonly observed on this position Arg and His respectively.

DISCUSSION

WHAT IS *MM*CHI60'S PLACE AMONG OTHER CHITINASES?

The common feature of family 18 chitinases is the β/α-barrel fold of the catalytic domain and the conserved sequence motif DxDxE. However, *Mm*Chi60 shows a clear amino acid sequence similarity with only one other chitinase whose structure has been determined experimentally: the chitinase from *L. lactis* (57% identity within the catalytic domain). The next most similar are the chitinase from *B. cereus* (29%) and the plant chitinase hevamine (23%). The catalytic domains of the three enzymes appear similar to the catalytic domain of *Mm*Chi60 in that they all form shallow substrate binding grooves, as opposed to ChiA or ChiB from *S. marcescens*, which have insertions in the basic topology of the β/α-barrel, making the binding sites deeper or 'tunnel-like'

Some chitinases have been found to process the substrate sequentially. The deep binding groove or even tunnel-shaped of those enzymes were attributed to their processivity (Zakariassen *et al.*, 2009). In processive enzymes, a single carbohydrate chain slides along the substrate binding cleft while disaccharides are cleaved off successively, as observed in ChiA and ChiB from *S. marcescens* (Davies & Henrissat, 1995). On the other hand, the non-processive enzymes are characterised by shallow substrate-binding grooves, as in the three-domain exochitinase ChiC from *S. marcescens* (Horn *et al.*, 2006, Suzuki *et al.*, 1999, Vaaje-Kolstad *et al.*, 2005). Contrary to ChiA and ChiB, in *Mm*Chi60, the shallow substrate-binding groove has fewer aromatic residues, especially on the '+' (aglycan) side. One feature that *Mm*Chi60 shares with ChiA, which is absent in ChiB, is Trp40 in the -3 sub-site. The corresponding residue in ChiA (Trp167) was shown to sustain its processivity toward chitin, while its substitution to Ala increased the enzyme's hydrolytic activity towards chitosan (partially deacetylated, water-soluble chitin derivative). The issue of processivity of *Mm*Chi60 has not been investigated. It can just be noted that according to recently published work, a shallow binding groove and flexible elements in the enzymatic machinery suggest a non-processive working mode (Payne *et al.*, 2012).

A remarkable similarity exists between *Mm*Chi60 and chitinase A1 from *B. circulans* (Toratani *et al.*, 2006). Studies of the latter by a combination of techniques revealed a similar arrangement of domains to *Mm*Chi60. Chitinase A1 has the catalytic β/α-barrel at one end and the chitin-binding domain at the other end, and a tandem of two fibronectin type III (FnIII) domains corresponding to the tandem of bacterial Ig-like (Big 3) domains in *Mm*Chi60. The two types of domain are superficially similar except that the Big 3 domains have only five β-strands each, lacking the two 'meandering' strands of the Greek key motif. The two Big 3 domains of *Mm*Chi60 are the first X-ray structures of this type of domain. One structure of such a domain had been determined by NMR (Aramini *et al.*, unpublished; PDB code 2kpn).

*MM*CHI60 REACTION MECHANISM BASED ON THE CRYSTAL STRUCTURES

The enzymatic reaction mechanism is probably the same among chitinases belonging to the GH18 family (van Aalten *et al.*, 2001), but different enzymes use various strategies to maximise their turnover rate. Chitin is a 'difficult' substrate: in the crystalline form the NAG chains are highly interconnected by H-bonds and form a rather featureless surface.

Analysis of the *Mm*Chi60 complexes shows that the enzyme is active in the crystal from. Enzymatic activity of crystallised enzymes has been noted many times before. In a previously reported experiment with another bacterial chitinase, the protein crystal was soaked in NAG_4 for 24h and only a single sugar ring was found in the enzyme's binding site (Perrakis *et al.*, 1994). In the present study NAG_3 has been reduced to NAG_2, which appears to be a stable product of this enzyme. In the soaking experiment with NAG_4 the oxazolinium reaction intermediate has been captured in the crystal. This is fortunate but not completely unexpected. The soaking times were deliberately short – only as long as it took to cryo-protect the crystal – in the hope of capturing interesting protein-ligand complexes. Trapping reaction intermediates is usually difficult and requires special measures or conditions to slow the reaction, as demonstrated for *O*-GlcNAc hydrolase (He *et al.*, 2010). This was not the case here and the pH 5.5 was close to optimal for the enzymatic reaction (Stefanidi & Vorgias, 2008). Very likely, the observation of the reaction intermediate is

strongly dependent in this case on the time of soaking and the initial concentration of the substrate.

Comparison of the three *Mm*Chi60 structures with other, previously reported chitinases can be used to identify structural features relevant to the catalytic mechanism. First, the DxDxE signature of family 18 chitinases is conserved (it corresponds to Asp149-Ile150-Asp151-Leu152-Glu-153), and the *Mm*Chi60 structures provide new information about their position and conformational changes at different points along the course of the enzymatic reaction. The most mobile residue in this study is Glu153, considered to be the acid/base of the reaction mechanism. In *Mm*Chi60, the residue flips between two positions more than 7 Å apart. In the unliganded structure it points away from the active site and its carboxyl group is 'parked'apparently stably against two main chain NH groups of a flexible loop that positions itself accordingly, while a water molecule occupies its place in the active site. The structure obtained in the absence of ligands can be considered the resting state of the enzyme. A similar arrangement is seen in the complex with the reaction intermediate. In this structure we see one of the elements necessary for hydrolysis of the oxazolinium intermediate – a water molecule – but the base is away. In the complex with NAG_2 the hydrolysis has already taken place. The active Glu has moved in, the water molecule is no longer there, but a hydroxyl group has been added to the sugar ring. A similar remote position of an equivalent Glu has been discovered in the chitinase from *L. lactis*, but not discussed. Recently, it was described in ChiC from *S. marcescens* (Payne *et al.*, 2012). In the chitinase from *B. cereus*, a corresponding flexible loop was described, but without flipping of the catalytic Glu (Hsieh *et al.*, 2010). The flipping of the catalytic Glu153 and shifting of adjacent residues in the presence of the reaction product can be compared to the ligand-induced conformational change described for other types of glycoside hydrolases (Varrot *et al.*, 2000)(Fig. 30).

Asp151 is the middle one of the three conserved residues of the chitinase DxDxE signature. In earlier studies this residue was proposed to rotate between the other two acidic side chains, shuttling a proton to the active site. It was reported to be in the 'down' position in the unliganded (resting) state of ChiB (van Aalten *et al.*, 2001) and in two alternative, 'up' or 'down'orientations in unliganded ChiA (Papanikolau *et al.*, 2003). The switching observed in ChiA and ChiB from *S. marcescens* was accompanied by adjusting movements

of the neighbouring Ser and Tyr residues. This Ser was examined and discovered to be important for the catalytic mechanism; its mutation to Ala reduced the specific activity 20-fold (Synstad *et al.,* 2004). A comparison of *Mm*Chi60 with other chitinases reveals different conformation of the key residues and amino acid substitutions around them. First, Asp151 points away from Asp149 in all three *Mm*Chi60 structures, including the unliganded enzyme. No contact is observed between the two residues. Secondly, the aforementioned Ser, whose mutation to Ala was debilitating to ChiB, occurs naturally in *Mm*Chi60 (Ala114). The 'up' position of Asp151 is stabilised by three H-bonds: (1) with the NH group of the N-acetyl moiety of the ligand in the active site, (2) with the catalytic Glu153, when it is directed toward the active site (3.1 Å in the complex with NAG_2), indicating a proton in between, (3) with the $-NH_2$ group of Gln218 (3.3 Å) (Fig. 24). Interactions (1) and (2) have been described in the previously proposed reaction mechanism (van Aalten *et al.,* 2001) but the third H-bond is not observed in most chitinase structures because they have Met instead of Gln. The residue in this position is interesting for its interaction with the ligand in the active site. In *Mm*Chi60, the carbonyl oxygen of Gln218 is in close contact with the O atom of the oxazolinium ring of the reaction intermediate (3.1 Å). It is also close to C1 (3.1 Å), the target of the nucleophilic attack by water. Such a close contact with the carbonyl O atom is expected to draw the H1 atom away from its riding position on C1, making the carbon atom accessible to the nucleophile. In those chitinases that have Met in this position, analogous interaction takes place between the C1-H1 group and the S atom. Thus the role of Met or Gln is to prime the reaction intermediate for the nucleophilic attack, but Gln218 in *Mm*Chi60 has an additional role – to maintain Asp151 in the 'up' position and to maintain Glu153 in the 'in' position relative to the active site (2.8 Å between Glu153 Oε2 and Gln218 Nε2). This is the active conformation of the Asp151-Glu153 pair, with the closely interacting carboxyl groups likely to hold a hydrogen atom between them.

The first of the conserved triad of acidic residues, Asp149, remains constant in the centre of the β/α-barrel, underneath the active site. Its position in the three *Mm*Chi60 structures is similar to the other known chitinases. The apparent role of this residue is to provide a negative potential to increase the pK_a of Asp151 nearby and indirectly of the catalytic Glu153.

Another interesting residue in the ligand-binding site is Asn221. Most studied chitinases have Asp in this position with the proposed role to increase the pK_a of the catalytic residue. Its mutation to Asn decreased the activity 20-fold in ChiB from *S. marcescens* (Synstad *et al.*, 2004). In *Mm*Chi60 Asn occurs naturally, which shows clearly that this enzyme has a different strategy to maintain an appropriately high pK_a of the catalytic Glu153. The presence of the pair Asn221 and the abovementioned Gln218 is observed not only in *Mm*Chi60. A corresponding pair of residues is found in chitinases from *L. lactis* (Bonnano *et al.*, unpublished, PDB code 3ian) *B. cereus* (Hsieh *et al.*, 2010), *S. coelicolor* (Vigdorovich *et al.*, unpublished, PDB code 3ebv), *A. fumigatus* (Rush *et al.*, 2010) and in hevamine (Terwisscha van Scheltinga *et al.*, 1996). Interestingly, these chitinases from diverse organisms – bacteria, a fungus and a plant – have a common characteristic: shallow substrate-binding groves (no insertions in the β/α-barrel). In contrast, other known chitinase structures have the above pair of residues replaced by Asp accompanied by Met, and most of those enzymes have 'tunnel-like' binding sites.

Structural comparisons allow us to discriminate the enzymes' invariant residues from those which have been evolving. The constant features are probably the most essential to the catalytic activity or substrate recognition. The other elements probably indicate adaptation of different enzymes to their specific environments or point to alternative means to achieve some auxiliary functionality, such as fine-tuning the pK_a of the catalytic residue to be close to the ambient pH.

EFFECT OF THE E153Q MUTATION

The reaction mechanism of *Mm*Chi60 and the role of the flexible active Glu153 were discussed based on three crystal structures: unliganded wild-type enzyme, its complex with an oxazolinium ion – the reaction intermediate, and a complex with the NAG_2 reaction product. The crystal structures of E153Q mutants add to this the initial stage of the enzymatic cycle – substrates bound to a mutant of reduced activity.

Three crystal structures of E153Q mutant have been solved and examined in order to evaluate and understand the process of ligand binding to *Mm*Chi60. The point mutation of active glutamate to glutamine served to debilitate the enzyme's activity in order to

observe a stable enzyme-substrate complex. The mutated active site residue is a part of the DxDxE signature (Henrissat & Davies, 1997). The structures have been compared with the previously solved wild type *Mm*Chi60 unliganded and in complex with shorter NAG ligands.

Kinetic and structural studies have shown that the E153Q mutation not only significantly debilitates the enzymatic activity but also affects the position of the mutated residue 153. The replacement of the carbonyl group by a neutral amide group entailed a considerable structural rearrangement. The unusual flipping of the active site residue upon ligand binding, observed in previously reported structures, is abolished in the mutant. The absence of one carbonyl group prevents it from 'parking' against the main chain amine groups of the flexible loop nearby. This indicates that both O atoms are essential for the 'away' conformation in the resting state. The mutated residue in the unliganded structure is directed toward the active site and is hydrogen bonded to the neighbouring Asp151. In the unliganded and liganded structures Gln153 interacts with Asp151 *via* the O atoms (3 Å), indicating a hydrogen atom in between. This is somewhat surprising because in the interaction between Gln and Asp side chains, one would expect the amine group to participate in the H-bond. The evident preference for the oxygen atom from Gln153 suggests that the oxygen from Asp151 is already protonated.

SUBSTRATE BINDING

Shallowness of the substrate-binding site makes it easily accessible to the substrate. Upon substrate binding, conformational changes can be observed in the glycon and the aglycon binding sites of the protein.

Induced fit of Gly116-Ile122 loop

Three stable positions can be distinguished of the Gly116-Ile122 loop (Fig. 30A). As mentioned before, in the resting state of the wild type enzyme this flexible loop is in the distal position with respect to the substrate glycon binding site. The active Glu153 is stationed against the loop's main chain N atoms. This conformation makes the glycon substrate binding site wide and accessible to the substrate. In the unliganded state of the E153Q mutant, with Gln153 detached from the flexible loop, the loop is shifted towards the

substrate binding groove, to a position which can be considered intermediate. Upon binding NAG oligomers by the mutated E153Q protein, the loop is in the proximal position with respect to the substrate, the same as in the complex of the active enzyme with the reaction product NAG_2. It seems that the binding of the substrate induces conformational changes favourable for the interaction with the protein. Due to the approach of the flexible loop three hydrogen bonds are created, positioning the substrate in the -1 and -2 sub-sides. The closing movement of the loop could be considered an induced fit, because it is observed only in the presence of substrate.

Extent of the binding site on the aglycon (insertion domain) and glycon sides

The ligand binding study reveals the role of the subdomain composed of a β-hairpin that forms a platform-like structure on the protein's surface on the aglycon side of the substrate binding site (Fig. 29). This can be contrasted with ChiA or ChiB from *S. marcescens* which have a deep substrate binding groove in this area. Similar insertion domains to that in *Mm*Chi60 were observed in ChiC from *S. marcescens* (Payne *et al.*, 2012) and a chitinase from *Lactococcus lactis* (PDB entry 3ian), but those structures are unliganded. The platform in *Mm*Chi60 is made up of amino acids 218-235, some of which are engaged in substrate binding *via* hydrogen bonds. The platform-like structure on the aglycon side extends from the +1 sub-site towards to the two nearly co-planar solvent-exposed Trp228 and Trp234. The last observed NAG moiety, in the +3 sub-site is approximately two NAG units short of the Trp patch (Fig. 29 & Fig. 30B).

Unfortunately, observation of longer ligands binding to this platform was not possible, although they were also used in soaking experiments. Short soaking with NAG_6 resulted in diffraction, but only a NAG pentamer was clearly visible in the electron density map, spanning sites +3 to -2. Overnight incubation with NAG_6 resulted in cracking of the crystals and no diffraction. It seems that crystal contacts on the '+' (aglycon) side obstruct longer NAG oligomers from binding. Crystal damage after longer soaking time suggests that the ligand exhibits affinity for an area extending beyond the +3 sub-site.

It is interesting why NAG_6 did not simply occupy sub-sites from +3 to -3, as there seems to be unobstructed space for at least one more NAG residue on the '-' side. Comparison of the liganded Chitinase A from *S. marcescens* with *Mm*Chi60 revealed that

residues involved in ligand binding in the -3 sub-site are not conserved in *Mm*Chi60 thus it is possible that the substrate-binding in *Mm*Chi60 does not extend beyond -2.

DOMAIN MOTION STUDIES

SAXS measurements indicated that *Mm*Chi60 in an aqueous environment adopts a wide range of conformations ranging from compact, having all the domains close to each other, through the elongated sea-horse shape, similar to that found in full length protein crystal, to fully extended (Fig. 39). What is more, the SAXS experiment helped resolve the issue of possible dimerization suggested by crystal contacts observed in the full-length protein and in its D2 mutant, demonstrating that *Mm*Chi60 is a monomer in solution.

The ensemble of conformations generated in modelling the SAXS profile gives a rather rough idea of the dynamic behaviour of the molecule, but allowing inter-domain flexibility is reasonable and there is also some evidence for it in the crystallographic model. In particular, the linker between the chitin binding domain and the rest of the molecule has a clearly elevated temperature factor and poorly defined electron density in the full-length protein structures. Conformation freedom of the ChBD domain was not observed in the crystal structure of *Mm*Chi60, where it is restrained by the crystal contacts. In the mutant structure, with one Ig-like domain (D2) deleted, additionally to elevated temperature factor in the linking region between Ig-like and ChBD, different spatial positions of ChBD of both molecules were observed. Based on the structural comparison of D2 and *Mm*Chi60, a hinge region was also localised between the catalytic domain and the first Ig-like domain.

The multi domain architecture is found in many bacterial chitinases but besides *Mm*Chi60 only the aforementioned chitinase A1 from *Bacillus circulans* has been investigated experimentally. There is no similarity in the primary structure of the two enzymes; however the arrangement of domains is similar: an N-terminal catalytic domain, two linking domains and a C-terminal chitin binding domain. It suggests an advantage of such architecture.

The overall structure of *Mm*Chi60 is suggestive as to the enzyme's mode of action as it skims the surface of chitin:

- the role of the chitin-binding domain would be to provide a general affinity of the protein to chitin. In the similar domain (ChBD1) of *Pyrococcus furiosus* the mutation of the each solvent-exposed aromatic residue, equivalent to those observed in *Mm*Chi60, resulted in a 30-40% decrease of affinity to insoluble chitin. It is worth noting that crystal structures of other chitin-binding domain have only two exposed aromatic residues, thus it seems that *Mm*Chi60 has higher affinity to the chitin. What is more, it was shown that solved exposed residues could be in various conformations but their planarity is always retained.
- the long, flexibly hinged Ig-like domains would afford the catalytic domain the freedom to probe the surface of chitin. It could control the distance between the catalytic domain and ChBD to bring the substrate closer to the binding groove.
- the flexible tongue-like insertion domain with the Trp patch on the surface of the catalytic domain would help orient it toward the substrate and dislodge NAG chains from the chitin's surface prior to hydrolysis

The general flexibility of the molecule and distribution of chitin-binding elements could enable it to disrupt the surface of insoluble chitin.

WHAT IS THE BASIS OF LOW-TEMPERATURE OPTIMISATION OF THE ENZYMATIC ACTIVITY?

It has been proposed that psychrophilic enzymes possess additional 'flexibility' which allows them to retain catalytic efficiency at low temperatures. It remains to be determined what one means by flexibility and to which parts of the protein it pertains. An examination of the *Mm*Chi60 structure indicates that a combination of factors are at play, including features of the active site that improve the enzyme's ability to recognise and admit the substrate under unfavourable thermodynamic conditions, and other elements of the protein involved in substrate-binding, and even the overall extended architectures of *MmChi60*.

In the active site we observe flipping of the catalytic Glu153 coupled with the rearrangement of the nearby loop 117-122. In the unliganded structure the loop is shifted 3 Å [non-breaking space is: ctrl-shift-space] away from the active site, in comparison to the

protein/NAG_2 complex. The open conformation and the shallowness of the substrate-binding groove indicate the enzyme's readiness to admit the substrate. On the other hand, the stable conformation of Asp151, pointing in all the *Mm*Chi60 structures toward the active site rather than into the interior of the protein, is conducive to catalysis. The substrate affinity of the binding groove can be supplemented by additional chitin-binding elements elsewhere on the protein: the chitin-binding domain and the pair of Trp residues on the β/α-barrel. The distribution of substrate affinity over the 100 Å length of the protein would reveal fully its effectiveness toward long substrates. The enzyme's long reach and the ability to cling to the surface of chitin could be a significant feature of its adaptation to the cold watery environment.

WHAT IS CHI40'S PLACE AMONG OTHER CHITINASES

The crystal structure of Chi40 was solved thanks to a single available crystal. The poor crystallisability of Chi40 is surprising and was investigated. The N-terminal end was truncated to remove unstructured or unstable motifs that are likely to interfere with crystallisation. Although cell secretion of Chi40 allows formation of disulphide bonds, all the purification steps and crystallisation were performed in the presence of reducing agents to avoid a mismatch in disulphide bridges which could prevent crystallisation. Even the idea for increasing the temperature of crystallisation, which was previously reported as an enhancing factor in the crystallisation of thermophilic proteins, did not yield crystals in the case of Chi40. Only one crystal did not allow an investigation of ligand binding as in the case of *Mm*Chi60, however this one crystal structure was very rewarding, considering that several groups have tried to crystallise this protein for almost 20 years.

Chi40 shares features characteristic of GH18, having the β/α-barrel fold of the catalytic domain and the conserved sequence motif DxDxE. Contrary to *Mm*Chi60, Chi40 has a deep substrate binding groove, similar to that found in ChiA from *S. marcescens.*

The thermophilic Chi40 shares 34% of its amino acids sequence with the catalytic domain of a psychrophilic chitinase from Arthrobacter TAD20, its structural counterpart adapted to the cold environment. A comparison of the two structures and analysis of amino

acid substitutions that allowed them to adapt to distinct temperatures, could give an answer to the question of adaptation. It is known that there are many pathways by which extremophilic enzymes can adjust to their environments and thus many different features have been described as significant for temperature extremophiles. Since thermal stability is achieved by the cumulative effect of small contributions, it could be difficult to formulate general rules for such enzymes. Alignment of the structures clearly shows that Chi40 has a reduced number of amino acids in loops and other secondary structure elements on its surface. This surface minimization could be the reason for the thermostability and almost 100% refolding of Chi40 after thermal denaturation. Shorter loops could decrease entropy of the unfolded state and thus make it easier for the protein to refold in the native form. Loops are regions where mobility is highest, thus they are likely initiation points for thermal denaturation. Minimization the ratio of the surface area to volume enhances the stability of the protein by reducing the unfavourable surface energy and increasing the interior packing interactions (Hwang *et al.*, 1997).

UNIQUE FEATURES OF CHI40

Disulphide bridges are common in extracellular proteins, stabilising their structures. Most studies point to other forces and devices to stabilize thermostable proteins, such as hydrophobic interaction, ionic interactions, hydrogen bonds rather than disulphide bridges. However, the strength of covalent bonding in disulphide bridges has to be reckoned with. The role of paired Cys residues emerged when the structure of the adenylosuccinate lyase from the hyperthermophilic *Pyrobaculum aerophilum* was solved and showed that six cysteine residues pair up to form three disulphids (Toth *et al.*, 2000).

In Chi40 all six Cys residues are paired up in disulphide bridges. What is more, the S atoms in one cystine made by Cys86 and Cys162 are additionally stabilized by S - π interactions with Phe residues. This kind of interaction has been observed in hydrophobic cores of proteins and is believed to contribute to protein stability (Reid, 1985). It is strongest when the S atoms lie close to the normal to the aromatic ring (Ringer *et al.*, 2007), which is just what is observed here in both the described sulphur atoms. A similar disulphide bridge is also observed in the psychrophilic chitinase A but the aromatic

interactions are absent there. Other similar chitinases from mezophilic organisms do not share this feature either. The other two more disulphide bridges observed in Chi40 are unique among all the other similar chitinases.

The high number of disulphide bridges and the exclusive occurrence of some of them in Chi40 suggest a functional character. Their presence should not affect the reaction mechanism of chitin degradation as all residues taking part in catalysis are conserved among comparable chitinases. Their occurrence could affect the thermodynamics of Chi40, in particular its thermostability and renaturation properties. Cystine residues can decrease the entropy of the unfolded state of proteins because once they are formed, the breaking of the covalent bond is energetically demanding and would requires temperatures higher than ambient even for hyperthermophiles

It has been noted that in the course of evolution enzymes adjusted the number, kind and strength of their internal interactions to optimize the balance between the rigidity – to be stable – and flexibility, to be active at their environmental temperatures (Davail *et al.*, 1994). The outcome of the analysis of salt bridges in Chi40 and its most related chitinase structures showed that Chi40 in this respect does not diverge significantly from other chitinases. Among eleven salt bridges five are highly conserved. Two sites, occupied by positively charged residues in all other chitinases, are substituted by proline residues. Those substitutions perhaps serve to stiffen the local structure. Proline stands out among amino acid by its restricted conformational possibilities, which should lower the entropy of the protein's unfolding state. The observed substitutions could be another device to achieve higher thermostability than mesophilic and psychrophilic counterparts.

SUMMARY

The thesis presents structural and functional studies of proteins from extremophiles. Chitinase 60 (*Mm*Chi60) from a psychrophilic bacteria *Moritella marina* and chitinase 40 (Chi40) from a thermophilic *Streptomyces thermoviolaceus* have been investigated with the aim to answer the question: *What structural features in extremophile's proteins are responsible for their unusual properties?*

X-ray crystallography reveals *Mm*Chi60 to be an elongated molecule which in addition to the catalytic β/α-barrel domain contains two Ig-like domains and a chitin-binding domain (ChBD), all linked in a chain. The four-domain structure of *Mm*Chi60 is outstanding in its complexity. Many glycoside hydrolases, such as chitinases or cellulases, have multi-domain structures, but only a few have been solved. Flexibility of hinge regions between domains apparently makes such proteins difficult to crystallise.

A ligand-binding study, using NAG oligomers, showed the enzyme to be active in the crystal lattice and resulted in complexes of the protein with oxazolinium ion – the reaction intermediate – and with NAG_2, a reaction product. The characteristic motif, DxDxE, containing three acidic amino acid residues, a signature of type 18 chitinases, is conserved in the enzyme. Further analysis of the unliganded enzyme and the two protein-ligand complexes, and a comparison with other known chitinases elucidated the role of other conserved residues near the active site.

Further analysis of the active site mutant of *Mm*Chi60 unliganded and in complex with substrates NAG_4 and NAG_5 revealed significant differences in the substrate-binding site compared to the already known complexes of most studied chitinases. The regions that move upon substrate binding were identified in both the glycon and aglycon substrate-binding sites.

To investigate hinge regions in the multi-domain *Mm*Chi60, deletion mutants were constructed: of one Ig-like domain removed (D2), both Ig-like domains removed (D12), and without both Ig-like domains and the chitin-binding domain (D123). In the first case,

the crystal structure revealed two molecules in the asymmetric unit with different positioning of ChBD and Ig-like domain in comparison with the full length protein. In the other cases the unit cell contained only the N-terminal domain which was used as a high-resolution structure of the *Mm*Chi60 catalytic domain. To supplement the structural studies of flexible regions, a SAXS experiment was performed and it was demonstrated that the protein, in addition to the elongated state found in the crystal, can adapt other conformations in solution, ranging from fully extended to compact.

Several features have been identified that are probably important in the reaction mechanism, substrate binding and the enzyme's efficiency at low temperatures. The chitin-binding domain and the tryptophan patch on the catalytic domain provide general affinity for chitin, in addition to the affinity of the binding site; the two Ig-like domains give the protein a long reach over the chitin surface; the flexible region between the chitin-binding domain and the adjacent Ig-like domain suggests the enzyme's ability to probe the surface of the substrate, while the open shallow substrate-binding groove allows easy access to the active site.

In addition to the psychrophilic chitinase, the crystal structure of a thermophilic chitinase 40 was solved, based on a single available crystal. Crystallographic analysis revealed a compact molecule which in addition to the catalytic β/α-barrel domain contained an insertion α+β domain. Generally, the structure of Chi40 is not unique but when compared with other most related chitinases and going into details, it differs. Chi40 shares 34% of its amino acids sequence with the catalytic domain of a psychrophilic chitinase from *Arthrobacter* TAD20, however many truncations have been identified in Chi40 making the enzyme more compact that its counterpart.

Structural features of Chi40 have been identified that are probably important for the thermophilic properties of the enzyme: its heat stability and almost 100% renaturation. The overall fold is stabilised by three disulphide bridges, two unique and one previously observed in the chitinase from *Arthrobacter*. However, the latter disulphide in Chi40 is stabilised by an interaction with aromatic rings of two phenylalanine residues. Additionally, some residues conserved in other structures, taking part in salt bridges, have been

substituted by prolines. This kind of substitution stabilizes a protein, due to a decrease in conformational entropy of its unfolded state.

The research of chitinases from extremophiles described here shows features that are common among the chitinases and those that are unique, indicating what could be important in chitinases' adaptation to extreme temperatures. A sstructure must embody a compromise between rigidity and flexibility, allowing not only conformational stability but also conformational freedom for its biological function. The balance point depends on the temperature in which the particular enzyme is working. It is not only interesting on the level of basic science, but is also important for designing enzymes for industrial use.

STRESZCZENIE

Rozprawa doktorska przedstawia badania strukturalne i funkcjonalne białek organizmów extremofilnych. Chitynaza 60 (*Mm*Chi60) z psychrofilnej bakterii *Moritella marina* i chitynaza 40 (Chi40) z termofilnej bakterii *Streptomyces thermoviolaceus* zostały poddane szczegółowej analizie celem odpowiedzi na pytanie: *Jakie cechy strukturalne białek z organizmów ekstremofilnych wpływają na ich właściwości adaptacyjne?*

Rozwiązana struktura krystaliczna *Mm*Chi60 pokazała, że cząsteczka białka ma kształt wydłużony i poza domeną katalityczną składa się dodatkowo z dwóch liniowo ułożonych domen Ig-like i domeny wiążącej chitynę (ChBD). Jest to pierwsza tak skomplikowana struktura chitynazy rozwiązana metodami krystalograficznymi, mimo że wiele chitynaz lub celulaz posiada wielodomenową złożoną budowę. Swoboda konformacyjna poszczególnych domen w wielu przypadkach jest przyczyną nieudanych prób uzyskania kryształów.

Badania kompleksów białko-ligand pokazały, że enzym w krysztale zachowuje aktywność. Wynikiem tych badań są kompleksy z produktem pośrednim – jonem oksazolinowym, i z produktem reakcji NAG_2. Charakterystyczny dla chitynaz z rodziny hydrolaz glikozydowych motyw trzech kwasowych reszt DxDxE jest zachowany w

*Mm*Chi60. Analiza porównawcza struktur enzymu niezwiązanego z ligandem i tymi, gdzie jest on obecny w miejscu wiązania substratu pozwoliła wyjaśnić rolę innych zachowanych reszt aminokwasów w pobliżu miejsca aktywnego.

Kolejne struktury mutanta reszty aktywnej E153Q nie związanego z ligandem i w kompleksie z substratem NAG_4 i NAG_5 ukazały znaczną różnicę w sposobie wiązania substratu między *Mm*Chi60 a innymi chitynazami, których struktury w kompleksie z substratem są znane. Miejsce wiązania substratu po stronie glikonowej i aglikonowej charakteryzują się dużą swobodą konformacyjną zmieniając swoje położenie pod wpływem wiązania ligandu.

Mutanty delecyjne jednej domeny Ig-like (D2), dwóch domen Ig-like (D12) i domen Ig-like łącznie z ChBD zostały zaprojektowane i poddane krystalizacji celem zaobserwowania miejsc elastycznych pomiędzy poszczególnymi domenami. W przypadku mutanta D2 w dwóch cząsteczkach w części asymetrycznej zaobserwowano różne ułożenia ChBD względem reszty białka. Także domena Ig-like jest przesunięta w porównaniu do białka o pełnej długości. Struktura krystaliczna mutantów D12 i D123 okazały się być bardzo podobne i składają się tylko z domeny katalitycznej, dlatego też, mając wysoką rozdzielczość, posłużyły jako źródło szczegółów o domenie katalitycznej *Mm*Chi60. Zwieńczeniem badań nad swobodą konformacyjną okazały się badania SAXS, które ukazały, że *Mm*Chi60 w roztworze, oprócz wydłużonego kształtu podobnego do tego obserwowanego w krysztale, może przyjmować inne konformacje, od całkowicie rozwiniętej do bardzo zwartej.

W cząsteczce *Mm*Chi60 zostało zidentyfikowanych wiele charakterystycznych cech, które mogą wpływać na mechanizm reakcji, wiązanie substratu i wydajność enzymatyczną obserwowaną w niskich temperaturach. Domena wiążąca chitynę łącznie z resztami tryptofanu liniowo ułożonymi i wyeksponowanymi w stronę rozpuszczalnika, odpowiedzialna jest za powinowactwo do nierozpuszczalnego substratu; dwie, elastycznie połączone domeny Ig-like pozwalają ustawić ChBD w odpowiedniej pozycji względem domeny katalitycznej, której płytkie miejsce wiązania substratu jest przygotowane na przyłączenie substratu.

Mając tylko jeden kryształ udało się rozwiązać strukturę krystaliczną termofilnej Chi40. Ukazała ona zwartą strukturę, która poza domeną katalityczną składa się z insercyjnej domeny α+β. Rozpatrując ogólnie, struktura ta jest podobna do innych dostępnych w bazie PDB jednak głębsze analizy pozwoliły na identyfikowanie cech nietypowych. Sekwencja aminokwasowa Chi40 wykazuje 34% identyczności z psychrofilną chitynazą z *Arthrobacter* TAD20, jednak Chi40 w wielu miejscach jest skrócona, co czyni ją jeszcze bardziej zwartą niż jej psychrofilna odpowiednik.

Wiele cechy strukturalnych białka Chi40, które mogą wpływać na termofilne właściwości tego białka: termostabilność i prawie 100% renaturację, zostało zidentyfikowanych. Struktura białka jest stabilizowana przez trzy wiązania disiarczkowe, z czego dwa nie są obserwowane w żadnych sekwencyjnie najbardziej podobnych strukturach. Trzecie, w Chi40 jest stabilizowane dodatkowe oddziaływaniami z pierścieniami aromatycznymi reszt fenyloalanin. Dodatkowo zaobserwowano substytucję przez proliny dodatnio naładowanych aminokwasów biorących udział w oddziaływaniu w postaci mostków solnych. Taka zamiana stabilizuje strukturę poprzez obniżenie entropi rozfałdowanego białka.

Przedstawione badania chitynaz organizmów ekstremofilnych ukazują cechy tych białek spotykane we wszystkich chitynazach jak również te, które są charakterystyczne dla enzymów ekstremofilnych, wskazując na ich rolę w adaptacji do określonych warunków temperaturowych. Pokazują, że struktury muszą zachować równowagę pomiędzy byciem sztywnym i elastycznym. Zapewnia to nie tylko stabilizację struktury, ale także swobodę konformacyjną potrzebną do pełnieni biologicznych funkcji. Szala tego kompromisu przechyla się w zależności od warunków temperaturowych, w których dany enzym występuje. Jest to nie tylko interesujące z punktu widzenia nauk podstawowych, ale również ważne z punktu widzenia projektowania wyspecjalizowanych enzymów wykorzystywanych w przemyśle.

REFERENCES

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Crystallogr D Biol Crystallogr* **68**, 352-367.

Aghajari, N., Feller, G., Gerday, C. & Haser, R. (1998). *Structure* **6**, 1503-1516.

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.

Aurilia, V., Rioux-Dube, J. F., Marabotti, A., Pezolet, M. & D'Auria, S. (2009). *The journal of physical chemistry. B* **113**, 7753-7761.

Bae, E. & Phillips, G. N., Jr. (2004). *The Journal of biological chemistry* **279**, 28202-28208.

Beguin, P. & Aubert, J. P. (1994). *FEMS microbiology reviews* **13**, 25-58.

Bernado, P. (2010). *European biophysics journal : EBJ* **39**, 769-780.

Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M. & Svergun, D. I. (2007). *J Am Chem Soc* **129**, 5656-5664.

Birnboim, H. C. & Doly, J. (1979). *Nucleic acids research* **7**, 1513-1523.

Blow, D. M. (2003). *Methods in enzymology* **374**, 3-22.

Bogin, O., Peretz, M., Hacham, Y., Korkhin, Y., Frolow, F., Kalb, A. J. & Burstein, Y. (1998). *Protein science : a publication of the Protein Society* **7**, 1156-1163.

Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. (2004). *The Biochemical journal* **382**, 769-781.

Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. (2003). *Acta Crystallogr D Biol Crystallogr* **59**, 2023-2030.

Brunger, A. T. (1997). *Methods in enzymology* **277**, 366-396.

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic acids research* **37**, D233-238.

Cavicchioli, R. (2006). *Nature reviews. Microbiology* **4**, 331-343.

Cavicchioli, R., Charlton, T., Ertan, H., Mohd Omar, S., Siddiqui, K. S. & Williams, T. J. (2011). *Microbial biotechnology* **4**, 449-460.

Christodoulou, E., Duffner, F. & Vorgias, C. E. (2001). *Protein expression and purification* **23**, 97-105.

Collins, T., Meuwis, M. A., Gerday, C. & Feller, G. (2003). *Journal of molecular biology* **328**, 419-428.

Colwell, R. R. & Morita, R. Y. (1964). *Journal of bacteriology* **88**, 831-837.

Costantini, S., Colonna, G. & Facchiano, A. M. (2008). *Bioinformation* **3**, 137-138.

Cowtan, K. (2000). *Acta Crystallogr D Biol Crystallogr* **56**, 1612-1621.

Cowtan, K. (2008). *Acta Crystallogr D Biol Crystallogr* **64**, 83-89.

Cummings, S. P. & Black, G. W. (1999). *Extremophiles : life under extreme conditions* **3**, 81-87.

D'Amico, S., Claverie, P., Collins, T., Georlette, D., Gratia, E., Hoyoux, A., Meuwis, M. A., Feller, G. & Gerday, C. (2002). *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **357**, 917-925.

D'Amico, S., Collins, T., Marx, J. C., Feller, G. & Gerday, C. (2006). *EMBO reports* **7**, 385-389.

D'Amico, S., Marx, J. C., Gerday, C. & Feller, G. (2003). *The Journal of biological chemistry* **278**, 7891-7896.

Dauter, Z. (1999). *Acta Crystallogr D Biol Crystallogr* **55**, 1703-1717.

Davail, S., Feller, G., Narinx, E. & Gerday, C. (1994). *The Journal of biological chemistry* **269**, 17448-17453.
Davies, G. & Henrissat, B. (1995). *Structure* **3**, 853-859.
Davies, G. J. (1998). *Biochemical Society transactions* **26**, 167-173.
Davies, G. J., Wilson, K. S. & Henrissat, B. (1997). *The Biochemical journal* **321 (Pt 2)**, 557-559.
DeLano, W. L. (2002). *DeLano Scientific, San Carlos, CA, USA*.
Demchenko, A. P., Ruskyn, O. I. & Saburova, E. A. (1989). *Biochimica et biophysica acta* **998**, 196-203.
Deming, J. W. (2002). *Current opinion in microbiology* **5**, 301-309.
Edelheit, O., Hanukoglu, A. & Hanukoglu, I. (2009). *BMC biotechnology* **9**, 61.
Elcock, A. H. (1998). *Journal of molecular biology* **284**, 489-502.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 486-501.
Feller, G. & Gerday, C. (1997). *Cellular and molecular life sciences : CMLS* **53**, 830-841.
Feller, G. & Gerday, C. (2003). *Nature reviews. Microbiology* **1**, 200-208.
Fields, P. A. & Somero, G. N. (1998). *Proceedings of the National Academy of Sciences of the United States of America* **95**, 11476-11481.
Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L., Eddy, S. R. & Bateman, A. (2010). *Nucleic acids research* **38**, D211-222.
Franke, D., Kikhney, A. G. & Svergun, D. I. (2012). *Nuclear Instruments and Methods in Physics Research A* **689**, 52–59.
Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). *The Proteomics Protocols Handbook*, 571-607
Georlette, D., Blaise, V., Collins, T., D'Amico, S., Gratia, E., Hoyoux, A., Marx, J. C., Sonan, G., Feller, G. & Gerday, C. (2004). *FEMS microbiology reviews* **28**, 25-42.
Gianese, G., Bossa, F. & Pascarella, S. (2002). *Proteins* **47**, 236-249.
Haney, P., Konisky, J., Koretke, K. K., Luthey-Schulten, Z. & Wolynes, P. G. (1997). *Proteins* **28**, 117-130.
Hayward, S. & Berendsen, H. J. (1998). *Proteins* **30**, 144-154.
He, Y., Macauley, M. S., Stubbs, K. A., Vocadlo, D. J. & Davies, G. J. (2010). *J Am Chem Soc* **132**, 1807-1809.
Hendrickson, W. A. (1985). *Methods in enzymology* **115**, 252-270.
Hendrickson, W. A. (1991). *Science* **254**, 51-58.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *The EMBO journal* **9**, 1665-1672.
Henrissat, B. (1991). *The Biochemical journal* **280 (Pt 2)**, 309-316.
Henrissat, B. & Bairoch, A. (1996). *The Biochemical journal* **316 (Pt 2)**, 695-696.
Henrissat, B. & Davies, G. (1997). *Current opinion in structural biology* **7**, 637-644.
Hollis, T., Monzingo, A. F., Bortone, K., Ernst, S., Cox, R. & Robertus, J. D. (2000). *Protein science : a publication of the Protein Society* **9**, 544-551.
Horn, S. J., Sikorski, P., Cederkvist, J. B., Vaaje-Kolstad, G., Sorlie, M., Synstad, B., Vriend, G., Varum, K. M. & Eijsink, V. G. (2006). *Proceedings of the National Academy of Sciences of the United States of America* **103**, 18089-18094.
Hsieh, Y. C., Wu, Y. J., Chiang, T. Y., Kuo, C. Y., Shrestha, K. L., Chao, C. F., Huang, Y. C., Chuankhayan, P., Wu, W. G., Li, Y. K. & Chen, C. J. (2010). *The Journal of biological chemistry* **285**, 31603-31615.
Hwang, K. Y., Song, H. K., Chang, C., Lee, J., Lee, S. Y., Kim, K. K., Choe, S., Sweet, R. M. & Suh, S. W. (1997). *Molecules and cells* **7**, 251-258.

Kabsch, W. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 125-132.
Karan, R., Capes, M. D. & Dassarma, S. (2012). *Aquatic biosystems* **8**, 4.
Karplus, P. A. & Diederichs, K. (2012). *Science* **336**, 1030-1033.
Kautharapu, K. B. & Jarboe, L. R. (2012). *Journal of bacteriology* **194**, 6296-6297.
Keyhani, N. O. & Roseman, S. (1999). *Biochimica et biophysica acta* **1473**, 108-122.
Kim, S. Y., Hwang, K. Y., Kim, S. H., Sung, H. C., Han, Y. S. & Cho, Y. (1999). *The Journal of biological chemistry* **274**, 11761-11767.
Klock, H. E. & Lesley, S. A. (2009). *Methods in molecular biology* **498**, 91-103.
Koshland, D. E. (1953). *Biol. Rev. Camb. Philos. Soc.* **28**, 416-436.
Kumar, S., Ma, B., Tsai, C. J. & Nussinov, R. (2000). *Proteins* **38**, 368-383.
Kurita, K. (2001). *Progress in Polymer Science* **26**, 1921-1971.
Kurowski, M. A. & Bujnicki, J. M. (2003). *Nucleic acids research* **31**, 3305-3307.
Laemmli, U. K. (1970). *Nature* **227**, 680-685.
Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature protocols* **3**, 1171-1179.
Lei, S. P., Lin, H. C., Wang, S. S., Callaway, J. & Wilcox, G. (1987). *Journal of bacteriology* **169**, 4379-4383.
Leslie, A. G. W. & Powell, H. R. (2007). *Evolving Methods for Macromolecular Crystallography* **245**, 41-51.
MacElroy, R. D. (1974). *Biosystems*, 74-75.
Margesin, R. & Feller, G. (2010). *Environmental technology* **31**, 835-844.
Marguet, E. & Forterre, P. (1998). *Extremophiles : life under extreme conditions* **2**, 115-122.
Matthews, B. W. (1968). *Journal of molecular biology* **33**, 491-497.
Matthews, B. W., Weaver, L. H. & Kester, W. R. (1974). *The Journal of biological chemistry* **249**, 8030-8044.
Mavromatis, K., Feller, G., Kokkinidis, M. & Bouriotis, V. (2003). *Protein engineering* **16**, 497-503.
McCarter, J. D. & Withers, S. G. (1994). *Current opinion in structural biology* **4**, 885-892.
McCarthy, A. J. & Williams, S. T. (1992). *Gene* **115**, 189-192.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J Appl Crystallogr* **40**, 658-674.
Merlino, A., Russo Krauss, I., Castellano, I., De Vendittis, E., Rossi, B., Conte, M., Vergara, A. & Sica, F. (2010). *Journal of structural biology* **172**, 343-352.
Michaux, C., Massant, J., Kerff, F., Frere, J. M., Docquier, J. D., Vandenberghe, I., Samyn, B., Pierrard, A., Feller, G., Charlier, P., Van Beeumen, J. & Wouters, J. (2008). *The FEBS journal* **275**, 1687-1697.
Mine, S., Nakamura, T., Sato, T., Ikegami, T. & Uegaki, K. (2014). *J Biochem* **155**, 115-122.
Monreal, J. & Reese, E. T. (1969). *Canadian journal of microbiology* **15**, 689-696.
Morita, R. Y. & Haight, R. D. (1964). *Limnol. Oceanogr.* **9**, 103-106.
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.
Muryoi, N., Sato, M., Kaneko, S., Kawahara, H., Obata, H., Yaish, M. W., Griffith, M. & Glick, B. R. (2004). *Journal of bacteriology* **186**, 5661-5671.
Nakamura, T., Mine, S., Hagihara, Y., Ishikawa, K. & Uegaki, K. (2007). *Acta crystallographica. Section F, Structural biology and crystallization communications* **63**, 7-11.
Nichols, C. M., Lardiere, S. G., Bowman, J. P., Nichols, P. D., J, A. E. G. & Guezennec, J. (2005). *Microbial ecology* **49**, 578-589.
Otwinowski, Z. & Minor, W. (1997). *Methods in enzymology*, edited by C. W. J. Carter & R. M. Sweet, pp. 307-326: Academic Press (New York).
Painter, J. & Merritt, E. A. (2006). *Acta Crystallogr D Biol Crystallogr* **62**, 439-450.

Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2005). *Acta Crystallogr D Biol Crystallogr* **61**, 449-457.
Pannu, N. S. & Read, R. J. (2004). *Acta Crystallogr D Biol Crystallogr* **60**, 22-27.
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.
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). *The Journal of biological chemistry* **287**, 36322-36330.
Peak, M. J., Robb, F. T. & Peak, J. G. (1995). *Journal of bacteriology* **177**, 6316-6318.
Peck, L. S. (2002). *Polar Biology*, 31-40.
Perrakis, A., Tews, I., Dauter, Z., Oppenheim, A. B., Chet, I., Wilson, K. S. & Vorgias, C. E. (1994). *Structure* **2**, 1169-1180.
Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. (2011). *Nature methods* **8**, 785-786.
Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V. & Svergun, D. I. (2012). *Journal of applied crystallography* **45**, 342-350.
Pomeroy, L. R. & Wiebe, W. J. (2001). *Aquat. Microb. Ecol* **23**, 187-204.
Poole, K. & Hancock, R. E. (1984). *European journal of biochemistry / FEBS* **144**, 607-612.
Pyrpassopoulos, S., Vlassi, M., Tsortos, A., Papanikolau, Y., Petratos, K., Vorgias, C. E. & Nounesis, G. (2006). *Proteins* **64**, 513-523.
Rasmussen, B. F., Stock, A. M., Ringe, D. & Petsko, G. A. (1992). *Nature* **357**, 423-424.
Reid, K. S. C., Lindley, P.F., Thornton, J.M. (1985). *The FEBS journal* **190**, 209-213.
Ringer, A. L., Senenko, A. & Sherrill, C. D. (2007). *Protein science : a publication of the Protein Society* **16**, 2216-2223.
Rossmann, M. G. & Blow, D. M. (1962). *Acta crystallographica* **15**, 24-31.
Rossmann, M. G., E.d. (1972).
Rothschild, L. J. & Mancinelli, R. L. (2001). *Nature* **409**, 1092-1101.
Round, A. R., Franke, D., Moritz, S., Huchler, R., Fritsche, M., Malthan, D., Klaering, R., Svergun, D. I. & Roessle, M. (2008). *Journal of applied crystallography* **41**, 913-917.
Rush, C. L., Schuttelkopf, A. W., Hurtado-Guerrero, R., Blair, D. E., Ibrahim, A. F., Desvergnes, S., Eggleston, I. M. & van Aalten, D. M. (2010). *Chem Biol* **17**, 1275-1281.
Russell, N. J. (1997). *Comparative biochemistry and physiology. Part A, Physiology* **118**, 489-493.
Russell, N. J. (2000). *Extremophiles : life under extreme conditions* **4**, 83-90.
Russell, R. J., Ferguson, J. M., Hough, D. W., Danson, M. J. & Taylor, G. L. (1997). *Biochemistry* **36**, 9983-9994.
Russell, R. J., Gerike, U., Danson, M. J., Hough, D. W. & Taylor, G. L. (1998). *Structure* **6**, 351-361.
Saelensminde, G., Halskau, O., Jr. & Jonassen, I. (2009). *Extremophiles : life under extreme conditions* **13**, 11-20.
Sakuda, S., Isogai, A., Matsumoto, S. & Suzuki, A. (1987). *The Journal of antibiotics* **40**, 296-300.
Salminen, T., Teplyakov, A., Kankare, J., Cooperman, B. S., Lahti, R. & Goldman, A. (1996). *Protein science : a publication of the Protein Society* **5**, 1014-1025.
Sheldrick, G. M. (2008). *Acta Crystallogr A* **64**, 112-122.
Siddiqui, K. S., Bokhari, S. A., Afzal, A. J. & Singh, S. (2004). *IUBMB life* **56**, 403-407.
Siddiqui, K. S. & Cavicchioli, R. (2006). *Annual review of biochemistry* **75**, 403-433.
Sinnott, M. L. (1990). *Chem. Rev.*, 1171-1202.
Somero, G. N. (1975). *The Journal of experimental zoology* **194**, 175-188.

Songsiriritthigul, C., Pantoom, S., Aguda, A. H., Robinson, R. C. & Suginta, W. (2008). *Journal of structural biology* **162**, 491-499.
Sorbotten, A., Horn, S. J., Eijsink, V. G. & Varum, K. M. (2005). *The FEBS journal* **272**, 538-549.
Stefanidi, E. & Vorgias, C. E. (2008). *Extremophiles : life under extreme conditions* **12**, 541-552.
Storey, K. B. & Storey, J. M. (1996). *Annu Rev Ecol Syst*, 365-386.
Stout, G. H. & Jensen, L. H. (1968).
Suzuki, K., Taiyoji, M., Sugawara, N., Nikaidou, N., Henrissat, B. & Watanabe, T. (1999). *The Biochemical journal* **343 Pt 3**, 587-596.
Svergun, D. I. (1992). *J. Appl. Crystallogr.* **25**, 495-503.
Svergun, D. I., Barberato, C. & Koch, M. H. J. (1995). *J. Appl. Crystallogr.* **28**, 768-773.
Synstad, B., Gaseidnes, S., Van Aalten, D. M., Vriend, G., Nielsen, J. E. & Eijsink, V. G. (2004). *European journal of biochemistry / FEBS* **271**, 253-262.
Terwilliger, T. C. (2003). *Acta Crystallogr D Biol Crystallogr* **59**, 38-44.
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). *Journal of molecular biology* **262**, 243-257.
Tews, I., Terwisscha van Scheltinga, A. C., Perrakis, A., Wilson, K. S. & Dijkstra, B. W. (1997). *Journal of the American Chemical Society*, 7954-7959.
Toratani, T., Kezuka, Y., Nonaka, T., Hiragi, Y. & Watanabe, T. (2006). *Biochem Biophys Res Commun* **348**, 814-818.
Toth, E. A., Worby, C., Dixon, J. E., Goedken, E. R., Marqusee, S. & Yeates, T. O. (2000). *Journal of molecular biology* **301**, 433-450.
Tsuji, H., Nishimura, S., Inui, T., Kado, Y., Ishikawa, K., Nakamura, T. & Uegaki, K. (2010). *The FEBS journal* **277**, 2683-2695.
Vaaje-Kolstad, G., Horn, S. J., van Aalten, D. M., Synstad, B. & Eijsink, V. G. (2005). *The Journal of biological chemistry* **280**, 28492-28497.
Vagin, A. & Teplyakov, A. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 22-25.
van Aalten, D. M., Komander, D., Synstad, B., Gaseidnes, S., Peter, M. G. & Eijsink, V. G. (2001). *Proceedings of the National Academy of Sciences of the United States of America* **98**, 8979-8984.
van Aalten, D. M., Synstad, B., Brurberg, M. B., Hough, E., Riise, B. W., Eijsink, V. G. & Wierenga, R. K. (2000). *Proceedings of the National Academy of Sciences of the United States of America* **97**, 5842-5847.
Varrot, A., Schulein, M. & Davies, G. J. (2000). *Journal of molecular biology* **297**, 819-828.
Vogt, G. & Argos, P. (1997). *Folding & design* **2**, S40-46.
Vogt, G., Woell, S. & Argos, P. (1997). *Journal of molecular biology* **269**, 631-643.
Watanabe, K., Hata, Y., Kizaki, H., Katsube, Y. & Suzuki, Y. (1997). *Journal of molecular biology* **269**, 142-153.
Weiss, M. S. & Hilgenfeld, R. (1997). *J. Appl. Cryst.* **30**, 203-205.
Wells, L. E. & Deming, J. W. (2006). *Environmental microbiology* **8**, 1115-1121.
Wharton, D. & Ferns, D. (1995). *The Journal of experimental biology* **198**, 1381-1387.
Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Crystallogr D Biol Crystallogr* **67**, 235-242.
Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Crystallogr D Biol Crystallogr* **57**, 122-133.

Xiao, L. & Honig, B. (1999). *Journal of molecular biology* **289**, 1435-1444.
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.
Yip, K. S., Britton, K. L., Stillman, T. J., Lebbink, J., de Vos, W. M., Robb, F. T., Vetriani, C., Maeder, D. & Rice, D. W. (1998). *European journal of biochemistry / FEBS* **255**, 336-346.
Yip, K. S., Stillman, T. J., Britton, K. L., Artymiuk, P. J., Baker, P. J., Sedelnikova, S. E., Engel, P. C., Pasquo, A., Chiaraluce, R., Consalvi, V. & et al. (1995). *Structure* **3**, 1147-1158.
Zakariassen, H., Aam, B. B., Horn, S. J., Varum, K. M., Sorlie, M. & Eijsink, V. G. (2009). *The Journal of biological chemistry* **284**, 10610-10617.
Zartler, E. R., Jenney, F. E., Jr., Terrell, M., Eidsness, M. K., Adams, M. W. & Prestegard, J. H. (2001). *Biochemistry* **40**, 7279-7290.
Zuber, H. (1988). *Biophysical chemistry* **29**, 171-179.

(166 references in total)