
Structural studies
of chitinolytic enzymes
from *Pyrococcus chitonophagus*

Katarzyna Biniak-Antosiak

2022

The research described in this thesis has been carried out
at the Institute of Bioorganic Chemistry, Polish Academy of Sciences in
Poznan
in the Department of Structure and Function of Biomolecules
Under the supervision of Professor Wojciech Rypniewski.

Auxiliary Supervisor:
Dr. Magdalena Bejger

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

The research conducted in this dissertation was supported by grants:

FNP, MASTER 2014, Engineering a “chitosome” – a toolbox for chitin degradation, (8./2014).

NSC, OPUS 14, The study of chitinolytic enzymes, aiming at constructing an artificial chitosome, (2017/27/B/NZ1/02201).



INSTITUTE OF BIOORGANIC CHEMISTRY
Polish Academy of Sciences



The studies described in this PhD dissertation have been included in the following article, which is currently under review:

Structural, thermodynamic and enzymatic characterization of N,N-diacetylchitobiose deacetylase from *Pyrococcus chitonophagus*

Katarzyna Biniek-Antosiak, Magdalena Bejger, Joanna Śliwiak, Daniel Baranowski, Ahmed Mohammed, Dmitri I. Svergun , Wojciech Rypniewski

*I would like to thank:
My advisor, Professor Wojciech Rypniewski
For support, patience and understanding.*

*I also want to thank
Dr. Magdalena Bejger
for her willingness to share her knowledge
and her time.*

*And also for a lot of valuable advices
Dr. habil. Agnieszka Kiliszek
Dr. Joanna Śliwiak
Dr. habil. Miłosz Ruszkowski
Dr Jakub Barciszewski
and Stanisław Wosicki, M.Sc.*

*I dedicate this work to my husband Patryk
and my Dad.*

Table Of Contents

1. Abbreviations	9
2. The aim of the study	11
3.1 Extremophiles	12
3.2 Thermophiles	13
3.3 <i>Pyrococcus chitonophagus</i>	14
3.4 Chitin	15
3.5 Extremophilic enzymes	16
3.6 Chitinolytic enzymes	17
3.7 Chitinases	19
3.8 Diacetylchitobiose deacetylase	21
3.9 <i>Exo</i> - β -D-glucosaminidase	22
4. Materials and methods	23
4.1 Protein overexpression system used	23
4.1.1 Amplification	23
4.1.2 Expression vector	24
4.1.3 TOPO cloning method	24
4.1.4 Protein expression	26
4.2 Protein purification	26
4.3 Verification of the quality of the protein preparation	27
4.4 Protein crystallization	28
4.5 Diffraction measurements	29
4.6 Definition and methods of solving the phase problem	30
4.7 Refinement of structures	32
4.8 Assessing the quality of structures	33
4.9 Small Angle X-ray Scattering (SAXS)	35
4.10 Differential Scanning Calorimetry (DSC)	35
4.11 Other programs and servers used for analysis and presentation of results	37
5. Experimental procedures	38

5.1 Plasmid expression vectors	38
5.2 Host cell transformation	39
5.3 Scaling-up the cell culture	40
5.4 Protein overexpression	40
5.5 Cell lysis	42
5.6 Protein purification	42
5.6.1 Chitinases	43
CH16_B1	43
CH19_K3	45
5.6.2 Diacetylchitobiose deacetylase - Dac-74	48
5.6.3 <i>Exo</i> - β -D-glucosaminidase - GlmA-01	50
5.7 Protein crystallization	52
5.9 Solving and refinement of the structures	56
6. Results and discussion	58
6.1 Diacetylchitobiose deacetylase - Dac-74	58
6.1.1 Packing in the unit cell	58
6.1.2 Dac-74 oligomerization in solution	60
6.1.3 Identifying the prosthetic group by anomalous X-ray scattering. .	61
6.1.4 Differential Scanning Calorimetry (DSC)	61
6.1.5 The structure of Dac-74	62
6.1.6 Substrate binding and proposed reaction mechanisms for Dac-74	64
6.1.7 Enzymatic activity	70
6.1.8 Conclusions of Dac-74	72
6.2 <i>Exo</i> - β -D-glucosaminidase - GlmA-01	74
6.2.1 Overall structure	74
6.2.2 Active site of GlmA-01	77
6.2.3 Enzymatic activity	78

6.2.4 Differential Scanning Calorimetry (DSC).....	79
6.2.5 Conclusion for GlmA-01	80
7. Summary.....	81
8. Streszczenie.....	83
9. References.....	85

1. Abbreviations

ADPs Atomic Displacement Parameters

ATP Adenosine Triphosphate

CAZy Database of Carbohydrate-Active enZymes

CH16_B1 Chitinase from *Pyrococcus chitonophagus*

CH19_K3 Chitinase from *Pyrococcus chitonophagus*

Dac-74 Diacetylchitobiose deacetylase from *Pyrococcus chitonophagus*

DSC Differential Scanning Calorimetry

DNA Deoxyribonucleic Acid

GlmA-01 *Exo*- β -D-glucosaminidase from *Pyrococcus chitonophagus*

IPTG Isopropyl- β -D-Thiogalactoside

IMAC Ion Metal Affinity Chromatography

MAD Multi-Wavelength Anomalous Diffraction

MCSG Midwest Center for Structural Genomics

MES 2-(N-morpholino)ethanesulfonic acid

MIR Multiple Isomorphous Replacement

MPD 2-Methyl-2,4-pentanediol

MR Molecular replacement

GlcN Glucosamine

(GlcN)₂ Chitobiose

GlcN-GlcNAc 1,4- β -D-Glucosaminyl-D-N-Acetylglucosamine

GlcNAc-GlcN 1,4- β -D-N-Acetylglucosaminyl-D-Glucosamine

GlcNAc N-acetyl-D-glucosamine

(GlcNAc)₂ N,N'-Diacetylchitobiose

(GlcNAc)₃ N,N',N''-Triacetylchitotriose

NCS Non-Crystallographic Symmetry

NMR Nuclear Magnetic Resonance

PCR Polymerase Chain Reaction

PDB Protein Data Bank

PEG Polyethylene glycol

SAD Single-Wavelength Anomalous Diffraction

SAXS Small-Angle X-ray Scattering

SEC Size-Exclusion Chromatography

SIR Single Isomorphous Replacement.

TCEP 3,3',3''-Phosphanetriyltripropanoic acid, reducing agent

TIM barrel β / α -barrel, named after triosephosphate isomerase

TPEN (N,N,N',N'-tetrakis(2-pyridinylmethyl)-1,2-ethanediamine

2. The aim of the study

The objects of my research were the following chitinolytic enzymes: chitinases (CH16_B1, CH19_K3), diacetylchitobiose deacetylase (Dac-74) and *exo*- β -D-glucosaminidase (GlmA-01).

The aim of the dissertation was to conduct structural and mechanistic studies of proteins from the chitin degradation pathway in the hyperthermophilic archaeon *Pyrococcus chitonophagus*. The results were to be interpreted with an emphasis on understanding the properties and function of the proteins in terms of their structure. Structural investigations of enzymes are facilitated if they can be observed in interactions with relevant ligands. Therefore, among my aims was a crystallographic study of protein-ligand complexes. Working with hyperthermophilic enzymes offers an opportunity to observe their special properties and investigate the bases of their adaptation to extreme environmental conditions. With this in mind, I also aimed to observe any characteristic properties of these thermostable proteins. The enzymes studied here are part of a broader investigation of chitin degradation. Therefore, I was also interested to see how the functions of the studied enzymes complement each other.

3. Introduction

3.1 Extremophiles

Extremophile is an organism that is adapted to or requires extreme ranges of environmental factors to live. In most cases, extremophiles are single-celled organisms that are mainly archaea or bacteria. Multicellular organisms living in extreme environmental conditions are also known.

Types of extremophiles:

- ✓ acidophile: an organism for which the optimum environmental pH is <3 ,
- ✓ alkaliphile: an organism for which the optimum environmental pH is >9 ,
- ✓ anaerobe: an organism that does not require oxygen,
- ✓ halophile: an organism that tolerates high salt concentrations, especially NaCl (above 10%),
- ✓ mesophile: an organism that lives in the temperature range of 15-60° C,
- ✓ metallotolerant: an organism capable of living in an environment containing a high concentration of heavy metals (copper, cadmium, arsenic, zinc),
- ✓ microaerophile: an organism that requires less oxygen than is present in the Earth's atmosphere,
- ✓ piezophile: an organism with optimal growth in hydrostatic pressures above 10 MPa (= 99 atm = 1,450 psi). Also referred to as barophile.
- ✓ psychrophile: an organism with an optimum in the temperature range $<15^{\circ}\text{C}$ (also living around 0°C),
- ✓ thermophile: a thermophilic organism that requires a temperature above 45°C

3.2 Thermophiles

Thermophiles are microorganisms that live in a hot environment, for example in hot springs, deep-sea thermal vents or decaying wood. They may be generally classified, depending on the optimal growth temperature, into three groups:

- ✓ moderate thermophiles (50-60 °C),
- ✓ extreme thermophiles (60-80 °C) and
- ✓ hyperthermophiles (80-110 °C).

At high temperatures, the fluidity of cell membranes increases. The body's strategy to overcome this is to adjust the lipid composition of the membrane, usually the ratio of saturated to unsaturated lipids.

An example of a hyperthermophilic organism is *Pyrococcus chitonophagus*.

3.3 *Pyrococcus chitonophagus*

Pyrococcus (formerly known as *Thermococcus*) *chitonophagus* is a chitin-degrading, hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent off the Mexican west coast at a depth of 2600 m (Huber et al., 1995). Figure 1 shows the location where *P. chitonophagus* was found.



Figure 1 Known occurrences, collected specimens and observations of *Pyrococcus chitonophagus* (Encyclopedia of Life. Available from <http://eol.org>. Accessed 15 July 2018).

The archaeon was previously suggested to belong to the *Thermococcus* rather than the *Pyrococcus* genus. Whole genome phylogeny as well as whole proteome comparisons using all available complete genomes in *Thermococcales* clearly showed that the species belongs to the *Pyrococcus chitonophagus* genus (Papadimitriou et al., 2016). It is anaerobic, round to slightly irregular coccus-shaped, 1.2–2.5 μm in diameter, and motile by means of a tuft of flagella. The appearance of *P. chitonophagus* is shown on Figure 2

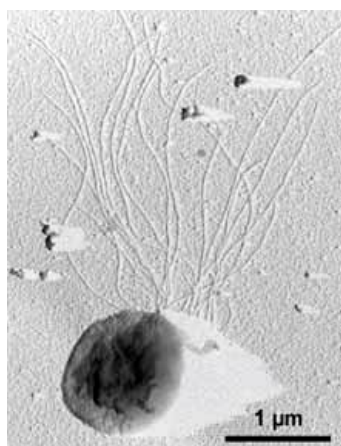


Figure 2 *Pyrococcus chitonophagus* image by K.O. Stetter & R. Rachel, Univ. Regensburg

3.4 Chitin

Chitin is the second most common biopolymer after cellulose. It is estimated that chitin production reaches 10^{10} - 10^{11} tons a year and for its decomposition are responsible archaea, bacteria and, to a lesser extent, some species of fungi (Gooday, n.d. 1990)

For organisms that recycle chitin it is a rich source of energy and macroelements necessary for life. As distinct from cellulose, chitin, in addition to carbon and oxygen, also contains nitrogen, which is harder to obtain. Despite its universality, chitin has been used to a small extent by humans. Chitin is a very stable substance and a recalcitrant substrate.

Chitin is found in three forms (Kurita, n.d. 2001) : α -chitin - chains of GlnNAc residues are arranged antiparallel allowing chitin to create strong hydrogen bonds; this form is the most abundant, most stable and highly insoluble, β -chitin - GlnNAc chains are arranged parallel and spaced further apart which makes the H-bonds weaker, γ -chitin - mixed composition of α - and β -chitin (Figure 3).

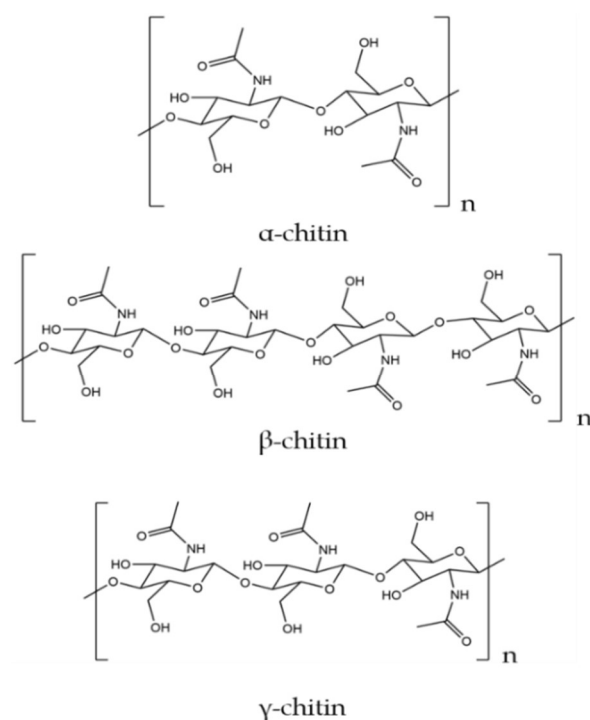


Figure 3 Molecular structures of: α -chitin, β -chitin and γ -chitin

3.5 Extremophilic enzymes

Organisms living at high temperature have developed mechanisms for their protection and produce macromolecules that resist thermal denaturation and preserve their activity. While most proteins are irreversibly denatured and lose their activity above 40°C, some thermophilic enzymes remain stable at temperatures close to 115°C (Pyrpassopoulos et al., 2006).

There are no simple determinants of thermostability, but some characteristics of thermostable proteins have been observed. Often they have shorter loops of polypeptide chains on their surface than mesophilic analogues. It has been also observed that thermophilic proteins compared with their mesophilic counterparts may have an increased number of hydrogen bonds, in particular salt bridges, and a relatively higher number of disulphide bridges (Yip et al., 1995), but they can also have a more hydrophobic character (Haney et al., 1999). However, not all studies confirm the general trends. Increased thermostability can also result from the mutation of single amino acid residues, especially on the surface of the protein (Kumwenda et al., 2013). A protein has also recently been described in which thermostability is due to the presence of a special domain at the N-terminal end of the protein. It has been proposed that thermostability results from the synergy of various factors rather than from one specific cause (Vogt et al., 1997).

Some of the thermophilic enzymes are used in molecular biology, for example the Taq polymerase used in PCR.

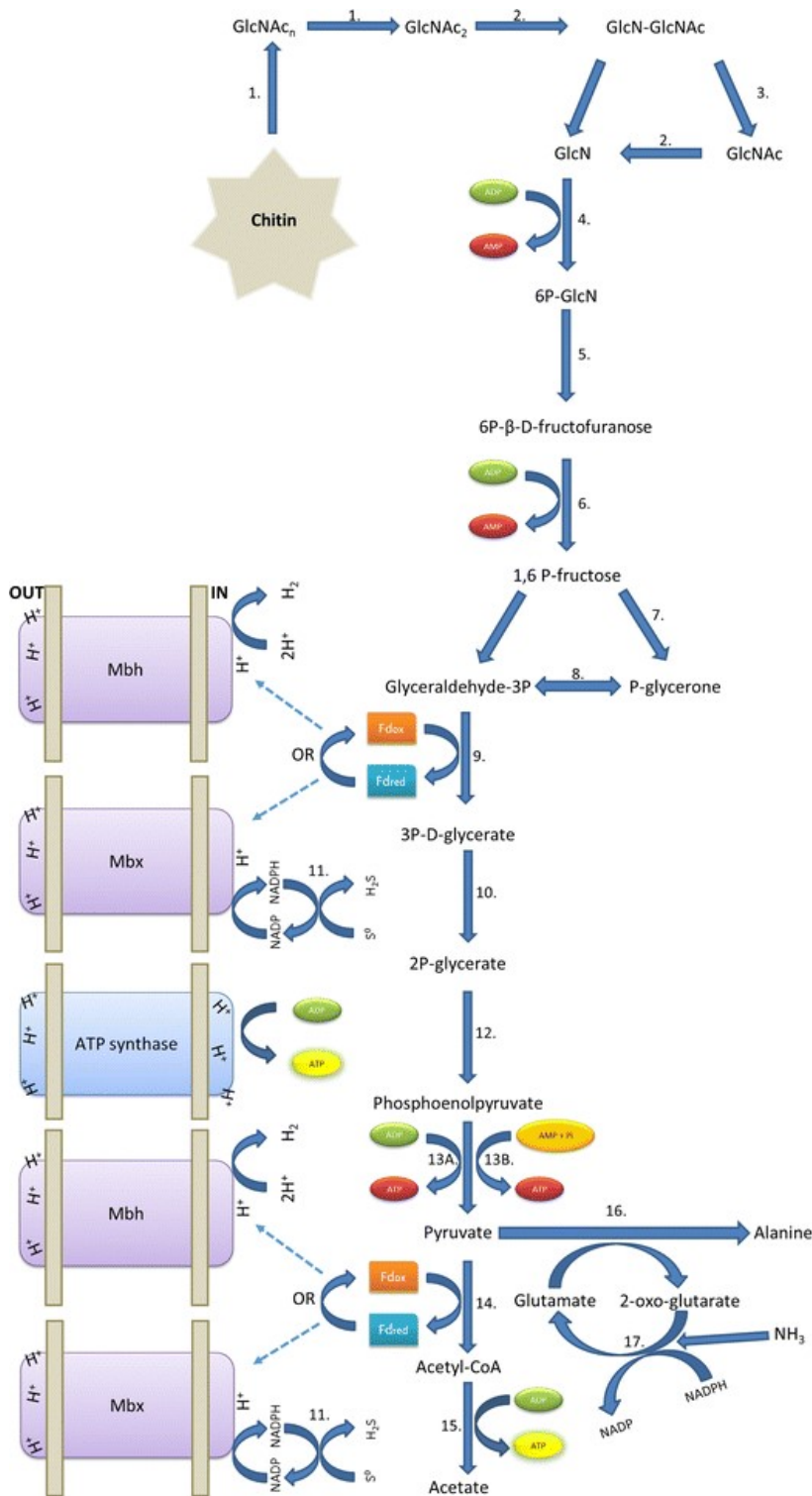
Enzymes from psychrophilic organisms are relatively heat labile but show an extraordinarily high activity at low temperatures, near 0°C, which gives them a competitive edge under those conditions. Considering the area of the oceans, Arctic and Antarctic regions, high mountains and permafrost, in most of the biosphere the temperature is below 5°C. Therefore, low temperatures are norm rather than exception. Cold-adapted organisms have colonised even the regions of Antarctic where the temperature drops to -60°C

(Casanueva et al., 2010). The main determining factor for life under those conditions is maintaining functionality of the enzymes responsible for metabolism and the cell cycle. Although the speed of chemical reactions decreases approximately exponentially with temperature, the psychrophilic enzymes can maintain activity at a comparable rate to their mesophilic counterparts.

3.6 Chitinolytic enzymes

Chitinolytic enzymes include a variety of chitinases (processive, non-processive, acting exo- and endo-) chitosanases, deacetylases, deaminases and even cellulases. Yet, it is possible that only three enzymes could degrade chitin all the way to glucosamine, which is readily usable and easily ingestible. This minimum set of enzymes consists of a chitinase, deacetylase and glucosaminidase. The main product of chitin is $(\text{GlcNAc})_2$, which can be deacetylated at one end by diacetylchitobiose deacetylase (Dac). The product of Dac could then be cleaved by glucosaminidase (GlmA) into a mixture of glucosamine and N-acetylgucosamine. The latter could be fed back to Dac to be deacetylated, yielding glucosamine as the final product. The enzymes described in this work are those first three enzymes involved in the degradation of chitin.

Their action is illustrated on the top part of the scheme of the chitin degradation pathway (Figure 4).



Chitinase (1),
 diacetylchitobiose deacetylase (2),
 exo -β-D-glucosaminidase (3),
 ADP-dependent D-glucosamine
 kinase (4),
 glucosamine 6-phosphate
 deaminase (5),
 ADP-dependent
 phosphofructokinase (6)
 fructose-biphosphate aldolase (7),
 triose-phosphate isomerase (8),
 glyceraldehyde-3-phosphate-
 ferredoxin oxidoreductase (9),
 phosphoglycerate mutase (10),
 CoASH-dependent NADPH S0
 reductase (11),
 enolase (12),
 pyruvate kinase (13A),
 phosphoenolpyruvate synthase
 (13B),
 pyruvate: ferredoxin
 oxidoreductase (14),
 acetyl-CoA synthetase (15),
 glutamate: pyruvate transaminase
 (16),
 NADP-dependent glutamate
 dehydrogenase (17).

Figure 4 Pathways for chitin degradation and central energy production of *P. chitonophagus*. (Papadimitriou et al., 2016).

3.7 Chitinases

Chitinases are common enzymes belonging to the glycoside hydrolase (GH) group. This group includes enzymes that hydrolyze glycosidic linkages in polysaccharides, oligosaccharides and glycosides. This very numerous group and has been divided into more than 115 families based on the amino acid sequences of their catalytic domains. Chitinases have been assigned to families 18, 19, 20 and 48 (Kasprzewska, 2003).

Many archaea and bacteria rely on the ability to completely break down chitin, and synthesize a range of chitinolytic enzymes. Bacteria produce chitinases belonging mainly to the GH 18 family, but also those belonging to the GH 19 family (Larsen et al., 2011). All fungal chitinases are also included in the GH 18 family, as are insect-produced chitinases (Hartl et al., 2012) .

Chitinases from the GH 19 family are mainly found in higher plants and, as I mentioned, in some bacteria. The origins of these two families of hydrolases are believed to be different because they differ in their amino acid sequences, three-dimensional structures and the mechanism of the catalytic reactions (Ohnuma et al., 2011).

Chitinases catalyze the hydrolysis of the β -glycosidic bonds between the C1 and C4 carbon of two adjacent N-acetylglucosamines in the chitin chain (Felse & Panda, 2000). Two types of chitinases have been distinguished: endochitinases (EC 3.2.1.14) and exochitinases (EC 3.2.1.52). This division results from different reaction mechanisms (Sahai & Manocha, 1993). Endochitinases randomly cleave bonds inside the chitin polymer to form soluble low molecular weight N-acetylglucosamine (GlcNAc) oligomers, such as, chitotetrose, chitotriose and N, N'-diacetylchitobiose dimers, those chitinases are non-processive enzymes (Thakur et al. 2022). On the other hand, exochitinases are processive enzymes i.e. they move along the poly GlcNAc chain as they hydrolyse it , in one direction or another depending on the type of chitinases i.e. catalyze the gradual release of (GlcNAc)₁₋₂ from the reducing or non-reducing end of the chitin microfiber.

The structure of bacterial chitinases usually consist of a removable signal sequence, a catalytic domain and, possibly, additional domains, such as chitin binding domains, fibronectin III domain or region rich in serine or threonine. Catalytic domains play the main role in the chitin hydrolysis process, however, depending on the hydrolase family, they differ in their spatial structure. The GH18 family chitinases contain the characteristic DxDxE amino acid sequence motif in which glutamate is the catalytic residue necessary for activity. In many chitinases, the catalytic domains contain inserted subdomains, however, only in a few cases are they assigned a function. Bacterial GH18 chitinases have a catalytic domain (CatD) in the form of a β/α -barrel (TIM-barrel) fold with catalytic residues at the carboxyl ends of the β -strands. In turn, the catalytic domains of chitinases belonging to GH19 are rich in α -helical structures (Kezuka et al., 2006). The presence of a chitin-binding domain increases the substrate affinity and the efficiency of chitin hydrolysis, indicating the key role of these domains (Morimoto et al., 1997). FnIII domains, or similar to them Ig-like domains, on the other hand, usually do not appear to be directly involved in chitin binding, but may be important for the enzyme activity, possibly influencing its properties and structure, facilitating the access of the catalytic domain to the substrate (Yan & Fong, 2015), or, as it was shown in Nakamura et al., FnIII domain can be involved in chitin decrystallization and chain sliding (Nakamura et al. 2018).

In terms of size, chitinases are very diverse. They range from about 20 kDa to about 120 kDa, with most bacterial chitinases ranging from 20 to 60 kDa. For comparison, the size of plant chitinases ranges from about 25 to about 40 kDa, and insect chitinases from 40 to 85 kDa. The chitinases discussed further in this work come from the archaea *P. chitonophagus* and their sizes are 90 kDa for CH19_K3 and 62 kDa for CH16_B1.

3.8 Diacetylchitobiose deacetylase

Chitobiose is an acetylation disaccharide formed during the enzymatic reaction of chitinase on chitin.

N,N' -diacetylchitobiose deacetylases (Dac) catalyze hydrolysis of the N-acetyl group of the reducing or non-reducing end of N,N' -diacetylchitobiose (GlcNAc)₂ and are part of the chitinolytic pathway in archaeons. In the CAZy database Dacs are defined as carbohydrate esterase and are dependent on zinc.

Different chitin deacetylases act either on chitin to turn it into chitosan, the de-N-acetylated product, (EC 3.5.1.41) (Blair et al., 2006) , or they act on shorter GlcNAc oligomers, most commonly the (GlcNAc)₂ product of chitin degradation by a chitinase (Figure 5). Many such enzymes are grouped into either the carbohydrate esterase (CE) family 4 (CE-4), when they act on the reducing end of the carbohydrate oligomer (EC 3.5.1.105) or CE-14 when they act on the non-reducing end (EC 3.5.1.136).

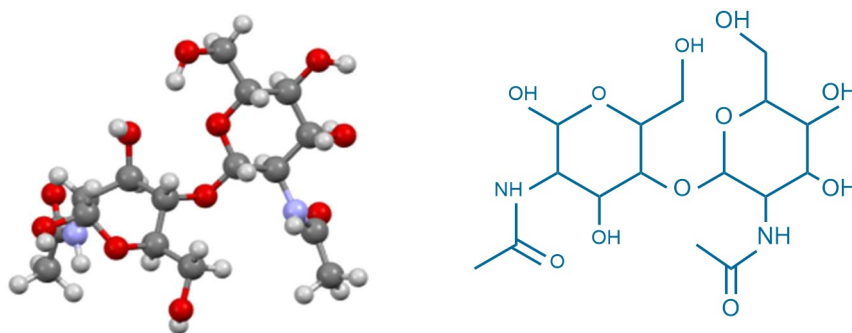


Figure 5 The substrate (GlcNAc)₂ of the enzymatic reaction of diacetylchitobiose deacetylase (Dac).

3.9 *Exo*- β -D-glucosaminidase

Exo- β -D-glucosaminidase is the last enzyme in this short chitin degradation path (Figure 6).

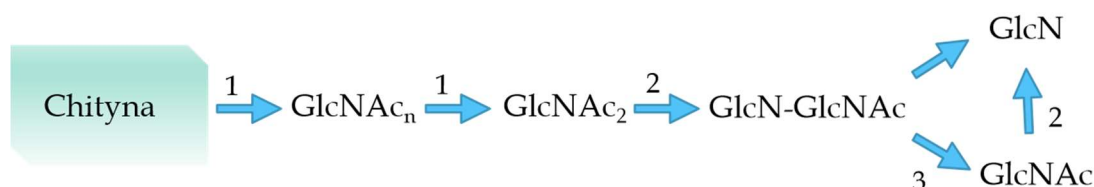


Figure 6 Fragment of the chitin metabolic pathway for *Pyrococcus chitonophagus*. (1) chitinase, (2) diacetylchitobiose deacetylase (3) *exo*- β -D-glucosaminidase

The archaeal *exo*- β -D-glucosaminidase (GlmA) is a thermostable enzyme belonging to the glycosidase hydrolase (GH) 35 family, hydrolyses chitosan oligosaccharides into monomer glucosamines. GlmA is a relatively novel enzyme in terms of its primary structure, as it is homologous to both GH35 and GH42 β -galactosidases (Mine & Watanabe, 2019). The resulting product of diacetylchitobiose deacetylase (GlcN-GlcNAc), shown on Figure 7, is subsequently hydrolyzed into GlcN and GlcNAc by an *exo*- β -D-glucosaminidase (GlmA) (EC 3.2.1.165).

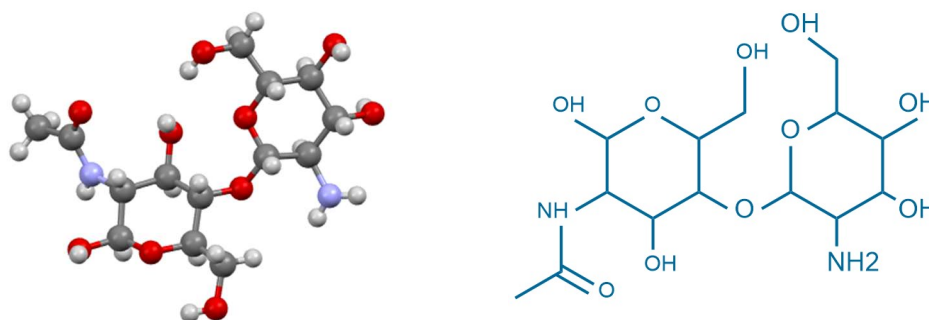


Figure 7 The product (GlcN-GlcNAc) of the enzymatic reaction of GH14 diacetylchitobiose deacetylase (Dac) and the substrate of *exo*- β -D-glucosaminidase (GlmA) GlmA

4. Materials and methods

4.1 Protein overexpression system used

In the process of determining the structure of proteins by X-ray crystallography it is necessary to obtain a protein crystal, which in turn requires significant amounts of high-purity protein. Hence, in biocrystallography, expression systems are necessary to allow relatively fast, cheap and efficient production of recombinant proteins. The preparation of an expression system for a given protein requires the use of a variety of methods and materials. The expression systems used in this study consisted of plasmid expression vectors containing genetic information about the protein under study - the coding sequence in the form of a DNA insert and the bacterial host, in this case, *Escherichia coli*, in which the protein is overexpressed.

4.1.1 Amplification

Any DNA fragment can be amplified *in vitro* by performing PCR (Polymerase Chain Reaction) (Mullis et al., 1986). A standard PCR reaction consists of three cyclically repeated steps:

- (I) denaturation of template DNA at 95 ° C,
 - (II) hybridization of primers to the template at their melting point,
 - (III) primer elongation complement with the sequence of the template DNA strand carried out by DNA polymerase at a temperature of approx. 72 ° C
- and
- (IV) final step: additional elongation of unfinished single-stranded fragments of the amplified DNA molecules.

In my experimental work I used this reaction to pick out the desired gene from the genomic DNA (insert) so that it could be connected to the expression vector.

4.1.2 Expression vector

Plasmids are circular double-stranded DNA molecules, naturally occurring in the cytoplasm of bacterial cells, which, due to the presence of the *ori* sequence, are capable of self-replication independently of the bacterial chromosome. On the basis of plasmids, commercial vectors were created, which were then used as carriers of heterologous genetic information, introduced into bacterial cells. The appropriate design of the vectors allows transcription and translation of this information using the enzymatic apparatus of the host cell. All vectors also have selective markers, allowing the growth of only those bacteria that have the vector. Currently, there are many types of vectors that carry a variety of additional features in addition to the transported gene. In my research, I used a commercial pET151/D-topo vector (Invitrogen). The vector I used has a gene encoding β -lactamase (*bla*), which determines resistance to the antibiotic ampicillin (which I added to the culture medium).

4.1.3 TOPO cloning method

The process of inserting the protein coding sequence into the pET151/D-topo vector is carried out in two steps. The first step is the preparation of the DNA insert obtained as a result of the PCR reaction (during which an appropriately designed primer attaches an additional CACC sequence to the 5'-end of the gene). The second step is ligation of the insert and vector. The commercial pET151/D-topo vector is a DNA strand rather than a circular closed molecule with a blunt end and a second sticky end, containing a GTGG sequence complementary to the CACC sequence of the insert. This allows it to bind in the correct orientation. The pET151/D-topo heterologous gene insertion is under the control of the T7 promoter, containing a lactose operator sequence (*lac O*) specific to the RNA polymerase from the T7 phage. The pET151/D-topo vector also has a gene encoding a

repressor protein of the lactose operon (*lacI*). The Lac repressor in pET-type vectors provides dual control of expression. First, by blocking transcription of the encapsulated foreign gene from the T7 promoter. Second, it prevents the expression of T7 polymerase, whose gene is under the control of the *lacUV5* promoter in the bacterial genome of cells expressing cells carrying the λ DE3 fragment derived from the λ phage genome. Expression of T7 polymerase and the target gene is induced by IPTG. The pET151/D-topo vector introduces at the N-terminus of the target protein a fragment of a polypeptide chain (4 kDa) that contains a histidine tag (a sequence of six histidines) and a TEV protease recognition site. (Figure 8).

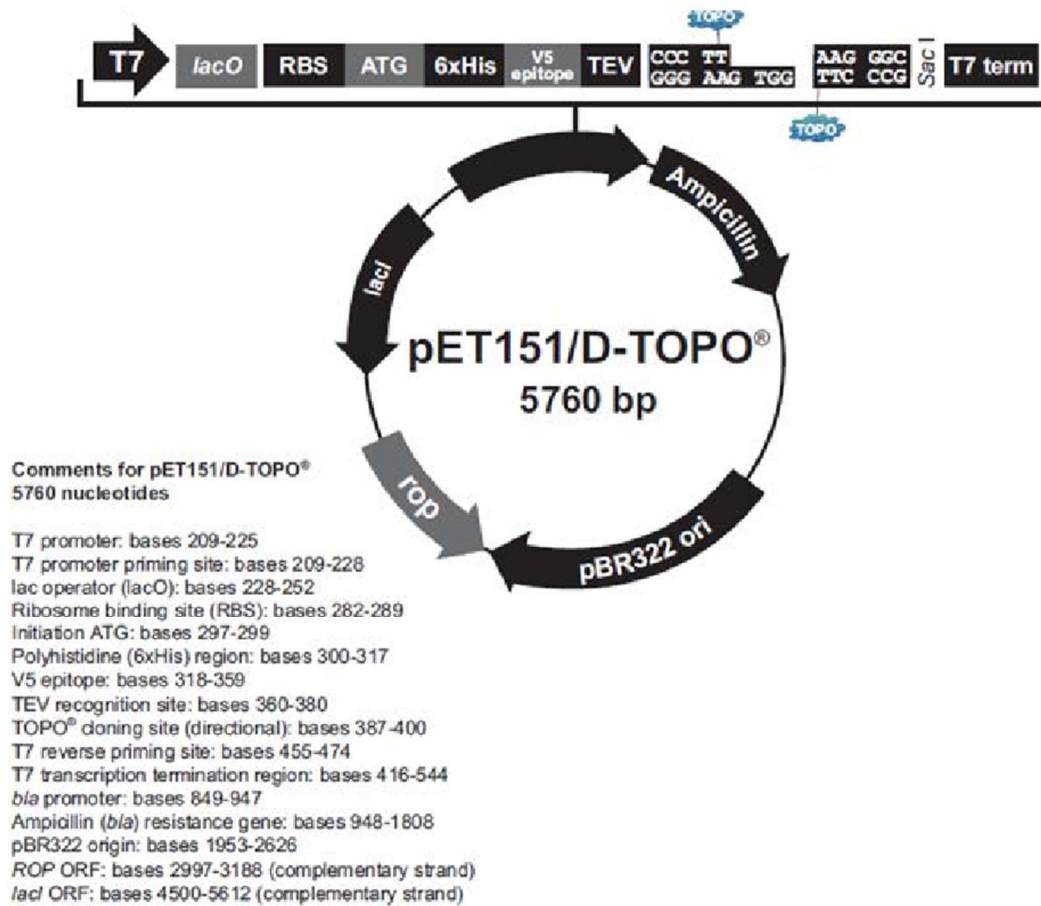


Figure 8 Schematic of pET151/D-topo. Diagram borrowed from the manual: Champion™ pET151 Directional TOPO™ Expression Kit (Invitrogen)

4.1.4 Protein expression

In order to express the recombinant proteins from the prepared expression vectors, the competent expression cells are chemically transformed while subjecting them to heat shock. The cells are then grown in media with antibiotic until they reach the log phase of growth ($OD_{600} = 0.8$) and the expression is induced by adding IPTG. For expression, I used BL21 "magic" cells and two different media (LB broth and TB broth). I prepared LB and TB media according to standard recipes (Sambrook & Russel, 2001) BL21 "magic" competent cells (MCSG) are characterized by having:

- the present λ DE3 fragment allowing the expression of T7 RNA polymerase
- "magic" plasmid encoding arginine (AGG and AGA codon recognition), isoleucine (ATA) tRNAs rarely used by bacteria,
- kanamycin resistance gene.

4.2 Protein purification

To isolate the protein obtained by overexpression, bacterial cells are lysed. In this process the cell membrane is destroyed and the protein is released into the solution where, hopefully, they maintain properties, e.g. enzymatic activity, structure, etc. The lysis of the bacteria can be performed with ultrasound (sonication), pressure (French press), appropriate chemical factors (organic solvents detergents,) or biochemical (enzymatic digestion). For my preparations, I used lysis with sonication.

Soluble protein is subjected to a purification process that allows to get rid of all unwanted cell elements (such as cell membranes, nucleic acids, carbohydrates, natural *E. coli* proteins) and obtain a homogeneous, suitable for crystallization target protein.

Purification is a multi step process in which many different methods are used, among them, the most important chromatographic methods, such as:

- Ion-exchange chromatography where differences in the charges of protein molecules are used
- Gel filtration where differences in the hydrodynamic radii of the separated proteins are used
- Affinity chromatography which is based on the specific interaction between different kinds of resin and short peptide tags or fusion proteins, which are fused to recombinant target protein as a result of a modification introduced into the protein sequence at the cloning stage. An example is the Histag fused to N-terminal end of my target proteins which is easily bound by resins with immobilized nickel ion. Short peptide tags and fusion proteins can usually be removed with proteases (e.g. TEV or factor Xa).

Chromatographic separations can be performed by the traditional gravity method (sometimes vacuum accelerated) or in the FPLC (Fast Performance Liquid Chromatography) system using overpressure.

During the purification of proteins, in addition to chromatographic methods, I also used techniques relying on the phenomenon of osmosis-dialysis. The solutions used, apart from the basic ingredients (buffer and salt), often contained additional ingredients, such as reducing agents or glycerol, with which I tried to influence the physicochemical properties of the purified objects (Bondos & Bicknell, 2003).

4.3 Verification of the quality of the protein preparation

In assessing the quality of the preparations, the basic method was denaturing polyacrylamide gel electrophoresis (SDS-PAGE). In this method, SDS unfolds the polypeptide chain and gives it a uniform negative charge proportional to the size of the molecule. The separation of denatured charged particles takes place in an electric field and the rate of migration through the pores of the gel depends on their mass. Thanks to the use of an appropriate mass marker the separation of proteins by size, and more precisely the separation of protein subunits (consisting of single chains), enables the

estimation of the size of the proteins, subunits present in the tested sample, and on this basis, their initial identification, while determining the purity of the preparation.

To estimate the protein concentration in the obtained samples, I used the spectrophotometric method and the Nano Drop spectrophotometer. The method consists of measuring the absorption of light with a wavelength $\lambda = 280$ nm by the aromatic residues of tyrosine and tryptophan for a protein with a known sequence. The Beer-Lambert law is used to calculate the protein concentration.

Another method, the results of which I used, is mass spectrometry. I commissioned the Mass Spectrometry Laboratory of the IBCh PAS to perform the experiments. The obtained results were necessary to confirm that the purified proteins are those which I expected.

4.4 Protein crystallization

Isolation of protein from solution in the form of single crystals proceeds similarly to crystallization of small organic or inorganic compounds. Each solvent (buffer) chemical compound (protein) has a certain concentration limit called the saturation concentration, at which precipitation or crystallization begins.

The first stage of crystallization is the formation of microscopic nuclei of nascent single crystals (nucleation). In the next phase, the single crystals grow (propagation crystallization). An undesirable effect of any crystallization is the agglomeration of single crystals into larger structures, known as adhesions. With the right choice of crystallization method and conditions, this can be inhibited, thus obtaining single crystals suitable for diffraction studies.

To obtain the crystals, I used the vapor diffusion method, the basis of which is that a closed system strives to reach equilibrium. In this method, a drop of a solution containing protein is placed over a container filled with a

solution of a precipitating agent . The drop consists of a protein solution of a certain concentration mixed with a precipitating agent solution in a certain ratio. In the drop containing the protein, the initial concentration of the precipitating agent is lower than in the reservoir solution. The diffusion of volatile components within the system leads to an equalization of the concentrations of precipitants in the two solutions, thus bringing the protein to a supersaturated state and eventually to its crystallization or precipitation.

I sought the initial conditions for crystallization in a seated drop system using a Gryphon crystallization robot (Art Robbins Instruments) with plates having 96 wells (Art Robbins Instruments) sealed with Crystal Clear (Hampton Research) tape. I obtained crystallization solutions from commercial crystallization reagent kits (Hampton Research, Molecular Dimensions Ltd).

I optimized the initial conditions of crystallization at the indicated temperatures, changing the volumes of the drop and container, changing solutions concentration and sometimes the arrangement of crystallization from a sitting drop to a hanging drop. The goal in optimizing the crystallization was to obtain more single crystals of sufficient size for X-ray diffraction measurements.

4.5 Diffraction measurements

The source of X-rays that I used was synchrotron radiation, which is now commonly used in biocrystallography. It is characterized by high intensity which is necessary in studies of protein crystals as they reach small sizes and have relatively large unit cell. Moreover, a great advantage of synchrotron radiation is the possibility of selecting an appropriate wavelength, which is of great importance when measuring the anomalous scattering signal (MAD or SAD).

4.6 Definition and methods of solving the phase problem

The purpose of the study of the crystal structure by X-ray diffraction methods is to determine the electron density distribution in a symmetrically independent part of the unit cell, because the maxima of the electron density electron will indicate the places where the atoms are located. Information about the electron density distribution is hidden in the diffraction pattern from the examined crystal.

The function that gives a mathematical description of the diffraction pattern is called the structure factor, indicated by the symbol $F(hkl)$. The electron density distribution, based on the diffraction pattern which we analyse, can be mathematically expressed as the inverse Fourier transform of the structure factor.

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

where:

V - unit cell volume

$|F(hkl)|$ - structure factor

$\rho(xyz)$ - coordinates of the atoms

Phase angle values cannot be measured with a detector, while the amplitude can be calculated from the intensity of the reflection $I(hkl)$ based on the relationship (2) :

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

The phase problem is therefore the fundamental problem to be solved in structural analysis. There are three main methods to solve the phase problem.

The first is isomorphic substitution (SIR or MIR), where the crystals are impregnated with heavy metal ions (e.g. mercury or platinum) or halides

(e.g. bromides or iodides). The effect of soaking is the appearance (in the crystal) of regions with a large number of electrons which are control points for determining the phases. The use of one isomorphous derivative (SIR) results in an ambiguity, in the sense that two solutions are obtained to the phase problem, with the same probability. An unequivocal solution can be obtained with the use of several different derivatives (MIR). A weak point of this method is a possible lack of isomorphism between the derivatives, i.e. the presence of differences in unit cell parameters or changes in the composition of the solvent.

The second method is tuned anomalous diffraction, multiple (MAD) or single (SAD), which consists of inducing resonance in electron clouds of atoms of specific elements with the help of radiation with an appropriately selected wavelength. The method is based on the use of synchrotron radiation having a wide range of possible wavelengths. The advantage of this method over isomorphous substitution is the possibility of using one crystal for measurements, the diffraction of which is measured at several wavelengths. λ_1 corresponds to the maximum absorption (f''), λ_2 corresponds to the inflection point of the absorption curve and maximum dispersion (f'), and at a wavelength λ_3 above the absorption threshold. A theoretical X-ray absorption spectrum is shown on Figure 9. Selenium (recombinant proteins with selenomethionine), heavy atoms or sulfur naturally occurring in proteins are most often used as anomalous atoms.

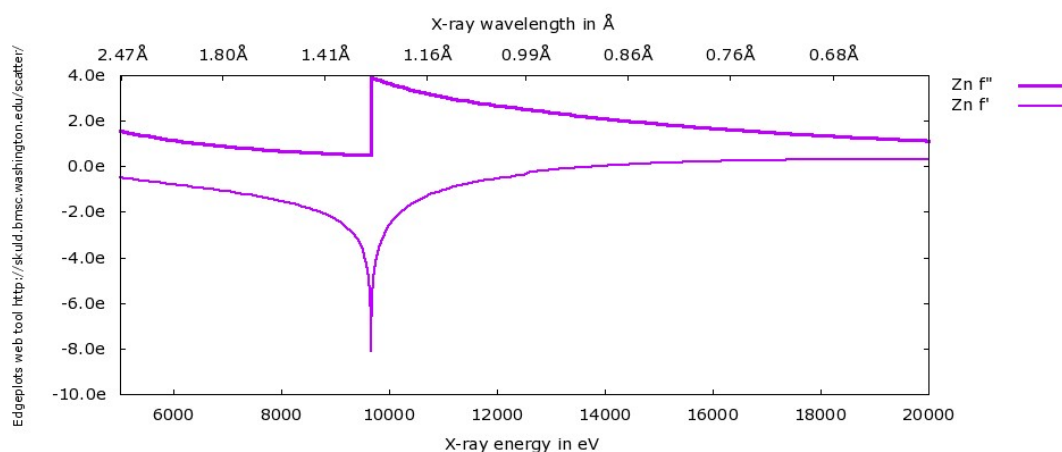


Figure 9 The absorption curve for a zinc atom generated by Edgeplots web tool <http://skuld.bmsc.washington.edu/scatter/>

The last method is molecular substitution/replacement (MR). It can be used when the investigated protein is structurally similar to an already existing model. The determination of the phases is based on the superposition of the Patterson functions, one derived from the experimental structure factor amplitudes and the second from amplitudes calculated from the 'search' model, in which the greatest correlation between the experimental vectors and those derived from a known structure is sought. The model for molecular substitution should meet two criteria. Its identity with the tested protein at the sequence level should be greater than $\sim 25\%$ and the mean square of the deviation of the positions of the C_{α} atoms of the model and the new structure should not exceed approx. 1.6 \AA . The source of the models is usually the PDB data bank.

Solving the phase problem with described above methods allows the calculation of the electron density map. The next stage of the work is the proper construction of the model based on the electron density. All the structures described in this paper were solved by the molecular substitution method using the PHASER program (McCoy et al., 2007) included in the CCP4 program package.

4.7 Refinement of structures

The structures solution is only the first stage that allows to define the coordinates of individual atoms - we only get an approximate model. In the case of the MR method, only the atoms of the initial model are defined in the structure. Therefore, the coordinates of all atoms must be refined, and in some cases it is also necessary to rebuild parts of the model. Refinement takes place in successive cycles. Each step includes a manual model correction and a round of calculations to minimize the differences between the experimental and calculated values of the structure factors. For the construction of the model, I used a specialized graphics program Coot. Based on electron density maps ($2Fo-Fc$ and $Fo-Fc$), it is possible to correct molecular models, for

example add and remove amino acid residues, ligands or water molecules. The rebuilt model is refined using programs usually based on the maximum likelihood method such as Refmac5 (Murshudov et al., 2011) or phenix.refine (Afonine et al., 2012) .

For the refined model, the electron density map is recalculated for the next rebuild cycle. To best refine the model parameters, it is necessary to include additional parameters such as soft constraints into the process. They contain information about the geometry of proteins developed on the basis of the geometry of small molecules deposited in the Cambridge (CSD) database (Allen & Motherwell, 2002) The resolution and quality of the dataset determine how strong constraints should be applied. In addition to stereochemical constraints, we also use the TLS (translation, libration, screw) parameters to refine the model. It makes it possible to describe anisotropic particle vibrations with only a small number of parameters. The TLS parameters are based on the assumption that groups of atoms can move like a rigid body. I used the Coot program to visualize and rebuild the model structures described in this paper. The structures were refined using the Refmac5 program included in the CCP4-packag

4.8 Assessing the quality of structures

Each step of refinement is monitored by checking the R and R_{free} discrepancy indicators. These are the values of the discrepancy between the experimental structure factor amplitudes and those calculated on the basis of the model. The R index, calculated for the entire set of reflections, has the disadvantage that it can improve (decrease in value) due to overfitting the model and the data, with each extension of the model with additional parameters (even if they are not confirmed by the experimental data). To prevent this, a test R_{free} indicator was introduced (Kleywegt & Jones, 1997) based on randomly selected ~ 1000 reflections, which are excluded from the refinement process from the very beginning (Brünger, 1992). For well-refined

structures, the R index should usually not be greater than 20%, while the difference between R_{free} and R should not be greater than approx. 3-7 percentage points (Wlodawer et al., 2008)

An important criterion for assessing the quality of the structure are the values of the torsion angles ϕ and ψ of the main chain, which are not constrained during refinement. The use of the Ramachandran diagram (Ramakrishnan & Ramachandran, 1965) allows the identification of amino acid residues with unusual or misconformation that may result from specific interactions of this residue or incorrect model structure. In order to verify the conformation of the residues in the model, the electron density 'omit' map should be analysed. The 'omit' map is an electron density map calculated after omitting the part of the molecule that is to be verified.

4.9 Small Angle X-ray Scattering (SAXS)

The technique of small angle X-ray scattering (SAXS) is popular in the study of biological systems and more specifically protein solutions. SAXS allows the determination of the macromolecular shape. The SAXS method can be combined with size-exclusion chromatography (SEC-SAXS) to determine the dominant oligomer species of the protein in solution.

In the SAXS instrument, a monochromatic X-ray beam is introduced into the sample, from which some of the X-rays scatter, while most pass through the sample without interacting with it. The scattered X-rays create a scattering pattern. The scattering pattern contains information about the structure of the sample. The technical problem that needs to be solved in SAXS is the separation of the weak scattering intensity from the strong main beam. The smaller the desired angle, the more difficult it is (usually 0.1-10°, hence the "small angle" in the name). The radially averaged scattering profile is interpreted in terms of the overall shape of the molecule in solution that caused the scattering. Depending on the angular range in which a clear scattering signal can be recorded, SAXS is able to provide structural information with dimensions from 1 to 100 nm and with repetitive distances in partially ordered systems up to 150 nm.

4.10 Differential Scanning Calorimetry (DSC)

In this work, the DSC method - differential scanning calorimetry - was used to measure thermal stability of the proteins. At the basis of the DSC method is the principle that if during heating, physical transformations take place in the material under study, the rate of heat delivery will be different from that of the reference material (here, no transformations will take place in the set temperature range). If endothermic processes occur in the sample, the rate of heat delivery will be higher, otherwise we will have an exothermic reaction.

The heat denaturation was studied by MICROCAL PEAQ-DSC system (Malvern Instruments Ltd.). Standard DSC experiment consisted of two measurements, both at the same instrument conditions: a) five reference scans with buffer-filled cells (both) to establish instrument thermal history and to achieve good baseline repeatability; b) one sample-buffer scan to prepare melting temperature data for analysis. Before every DSC experiment protein sample was dialyzed against specified buffer (which was also later used in DSC scans) and protein concentration was measured using NanoDrop.

4.11 Other programs and servers used for analysis and presentation of results

I also used the following program and servers to solve, refine and analyse the structures:

- ✓ SSM - program for comparison of structures (in the COOT program) (Krissinel & Henrick, 2004)
- ✓ BLAST - server to search amino acid sequence search and comparison (Altschul et al., 1997)
- ✓ ALIGN - server to least squares superimposition (Satow et al., 1986)
- ✓ DSSP - server to secondary structure (Kabsch & Sander, 1983)
- ✓ PDBsum - server to analysis of the quaternary structure (Laskowski, 2009)
- ✓ UCSF Chimera - program to preparation illustrations (Pettersen et al., 2004)
- ✓ ARP / wARP - program for automatic model building and refinement (Lamzin & Wilson, 1993)
- ✓ The STARANISO Server - the server aims to perform an anisotropic cut-off of the combined intensity data, to perform a Bayesian estimate of structure amplitudes, and to apply an anisotropic correction to the data (Tickle et al., 2018).
- ✓ Alpha fold 2 - a server for predicting the structure of a protein based on the sequence of amino acids (Mirdita et al., 2021)
- ✓ Dali (Holm, 2020)

5. Experimental procedures

5.1 Plasmid expression vectors

I received the material for research in the form of a genomic DNA from *Pyrococcus chitonophagus* DSM 10152 (formerly *Thermococcus chitonophagus*) from Prof. Constantinos E. Vorgias (The University of Athens). After analyzing the genome (GenBank: LN999010.1) (Papadimitriou et al., 2016), I identified 37 protein coding genes that may be involved in chitin degradation and then designated four of them for laboratory studies. The selected genes (shown in Table 1) were cloned into pET151D-topo vectors which although expression with N-terminal tag containing codons for six histidines and TEV cleavage site.

Table 1 Information contained in GenBank for the enzymes described in the thesis.

Locus tag	GenBank code	Encoded enzyme	The short name of protein used in this thesis
Chiton_1716	CUX78495.1	chitinase	CH16_B1
Chiton_1119	CUX77898.1	chitinase	CH19_K3
Chiton_0574	CUX77353.1	diacetylchitobiose deacetylase	Dac-74
Chiton_1401	CUX78180.1	exo- β -D- glucosaminidase	GlmA-01

When creating pET vectors, firstly I prepared DNA inserts for every selected gene. Using PCR from the genetic material of *Pyrococcus chitonophagus* I extracted a DNA fragment constituting the genes of the proteins I was interested in. During these reactions I used PWO polymerase (Roche) and primers (F-forward and R-reverse) designed to match the beginning and end of the gene sequence being extracted, with the F primer introducing an additional four CACC nucleotides at the 5' end of the coding strand of the inserts. Samples of each PCR reaction products (inserts) were subjected to electrophoresis on a 1% agarose gel in order to confirm the length of the amplified DNA fragments. In those cases where in one PCR-reaction I obtained several products, I separated the entire post-reaction mixture electrophoretically, and then I isolated from the gel only those DNA fragments of the appropriate length. The inserts were isolated using a commercial QIAquick Gel Extraction Kit from Quiagen. Ligation reactions, which I carried out according to the protocol provided with the pET151/D-topo vector by Invitrogen, essentially consisted of mixing the insert with a vector. Ligation of the DNA insert and the pET151/D-topo plasmid vector was caused by the presence of enzyme topoisomerase. Introduced into the insert, additional nucleotides (5' CACC-) ensured the correct direction of the ligation.

5.2 Host cell transformation

When introducing an insert containing vector into *E. coli* host cells I used heat shock. The eppendorf tube containing the plasmid and specific competent cells was incubated firstly on ice for 30 minutes, then in a heating block at 42°C for approximately 30 seconds, and again on ice for 3 minutes. I used BL21 “magic” expression cells for transformation.

5.3 Scaling-up the cell culture

The cell culture procedure for all four proteins followed the same protocol, all were performed at 37°C and all fresh cultures of bacteria (BL21-“magic” *E. coli*), were inoculated with cells from the previous culture. The procedure contain the following steps after the transformation:

- ✓ Growing the culture for an hour, with shaking (250 rpm), in SOC medium (200 µl);
- ✓ Growing the culture overnight (approx. 16 hours), on LB or TB medium (3 ml) with antibiotics (kanamycin, ampicillin);
- ✓ Growing the culture for 2 - 4 hours, with shaking (250 rpm), in a medium with antibiotics (kanamycin, ampicillin) (40 ml);
- ✓ The last liquid culture was carried out on a scale of typically about 4 liters to an OD₆₀₀ value in the range of 0.5 - 0.8, and on LB or TB media.

5.4 Protein overexpression

During the pilot expressions, I determined the appropriate IPTG concentration, temperature and expression time for the individual proteins. I also checked the types of competent bacteria in which expression was most effective. For example I used BL21(DE3) Star (Invitrogen), OverExpress C41(DE3) and OverExpress C43 (Lucigen Corporation).

The results of the pilot expressions in terms of the concentration of the inducer were similar, therefore the final concentration for all proteins was 0.5mM IPTG. I conducted my expressions at two temperatures: 18°C (duration: 16 hours) and 37°C (duration: 4 hours). The temperature of expression can influence the solubility of the produced proteins (Schein & Noteborn, 1988). The bacterial cells, from 4L of culture, were collected by centrifugation and after decantation of the culture media cells pellets were suspended in 60ml (or 120ml) of water and frozen at -80°C for future use. The expression conditions for all proteins are summarized in Table 2

Table 2 Conditions of gene expression of the chitinolytic enzymes used in this work.

Protein	CH19_K3	CH16_B1	Dac-74	GlmA-01
Competent cells	BL21-“magic” (<i>E. coli</i>)	BL21-“magic” (<i>E. coli</i>)	BL21-“magic” (<i>E. coli</i>)	BL21-“magic” (<i>E. coli</i>)
Medium	TB	TB	LB	LB
Antibiotics	kanamycin, ampicillin	kanamycin, ampicillin	kanamycin, ampicillin	kanamycin, ampicillin
Temperature expressions	37°C	18°C	37°C	18°C
Time expressions	4h	16h	4h	16h

5.5 Cell lysis

The first step in protein purification is to subject the preparation to cell lysis by sonication. The lysis buffers used are summarized in Table 3.

Table 3 The composition of the buffers used for lysis.

Protein	CH19_K3	CH16_B1	Dac-74	GlmA-01
Lysis buffer	50mM TRIS pH 7.5	50mM TRIS pH 8.5	50mM TRIS pH 7.5	50mM TRIS pH 8.5
	200mM NaCl	200mM NaCl	200mM NaCl	300mM NaCl
	50mM imidazole	25mM imidazole	50mM imidazole	30mM imidazole
	10% glycerol	10% glycerol	10% glycerol	10% glycerol
	0.5% Triton X100	0.5% Triton X100	0.5% Triton X100	0.5% Triton X100
	1mM PMSF	1mM PMSF	1mM PMSF	1mM PMSF

Sonication was performed at intervals of 2 seconds to 20 seconds for 4.5 minutes total sonication time with amplitude 70%. After approximately 1 hour, the sample was ready for centrifugation at 16000 rpm for 30 minutes at 4°C. After centrifugation, I placed the supernatant with soluble proteins on a nickel column in order to bind the protein to the resin. Cell pellet contained insoluble protein was not used.

5.6 Protein purification

I purified proteins using affinity chromatography, what was possible because of the modification (His-tag) introduced into the target proteins by genetic engineering at the cloning stage. More GE Healthcare Ni Sepharose High Performance resins with immobilized Nickel ion. The basic properties of the analyzed proteins are summarized in Table 4 .

Table 4 Properties of the studied proteins

Protein	CH19_K3	CH16_B1	Dac-74	GlmA-01
Number of amino acids	805	546	267	781
Molecular weight [Da]	90564.15	62345.86	30931.62	91032..08
Theoretical pI	4.92	7.78	5.35	5.57

5.6.1 Chitinases

CH16_B1

After lysis, I placed the supernatant on a nickel column and followed the protocol described below in Table 5. I used Bradford's reagent to check the presence of protein in 5 ml portions of the liquids eluted in each step.

Table 5 Preparation of the protein on Ni²⁺column

5ml resins preparation	Binding step	Wash step	Elution step	Trash step
3x 15ml H ₂ O 2x 15ml basic buffer + 50mM imidazole <u>Basic buffer</u> <u>50mM TRIS</u> <u>pH 8.5</u> <u>200mM NaCl</u>	Pour the supernatant into the column, mix and leave it for 40 minutes in the cold room. The rest of the steps are carried out at room temperature.	Basic buffer + 50mM imidazole (about 3-4 times) + 100mM imidazole (about 2-3 times) until all not bounded proteins are washed away.	Elution of bounded proteins with basic buffer +300mM imidazole (about 30 ml).	Resins cleaning with basic buffer +600 mM imidazole and regeneration of the resins.

Then I placed the eluted protein solution in a dialysis bag with TEV protease added in order to cut off the His-tag (at a ratio of 1mg Histag fused TEV protease per 10mg protein) and performed a two-step dialysis. The first stage lasted approximately 2 hours, the second stage approximately 18 hours in the cold room. I dialyzed into a buffer: 50mM TRIS pH 8.5, 200mM NaCl, 5mM 2-Mercaptoethanol (which was added because TEV protease needs it to

be active). Table 6 summarizes the protein purification on the second nickel column which was preformed after dialysis.

Table 6 II Ni²⁺column conditions.

5ml resins preparation	Binding step	Flow through step	Trash step
3x 15ml H2O 2x 15ml basic buffer + 50mM imidazole	Pour the dialyzed proteins into the columns, mix and leave for 10-15 minutes in the cold room.	TEV protease and His-tag bind to the column. Target protein do not bind but flow through column.	Resins cleaning with basic buffer +600 mM imidazole and regeneration of the resins.

CH16_B1 was applied to a HiPrep 16/60 Superdex S-200 GE Healthcare column pre-equilibrated with 25 mM TRIS pH 8.5 and 200 mM NaCl in the AKTA FPLC system. The fractions (12-23) containing CH16_B1 were concentrated on Amicon Ultra 15 ml centrifugal filter units with a 30000MW cut-off to 10 mg/ml, ready for crystallization. Figure 10 shows the chromatogram for the CH16_B1 protein. The purity of the samples was checked with the SDS-PAGE as shown in Figure 11.

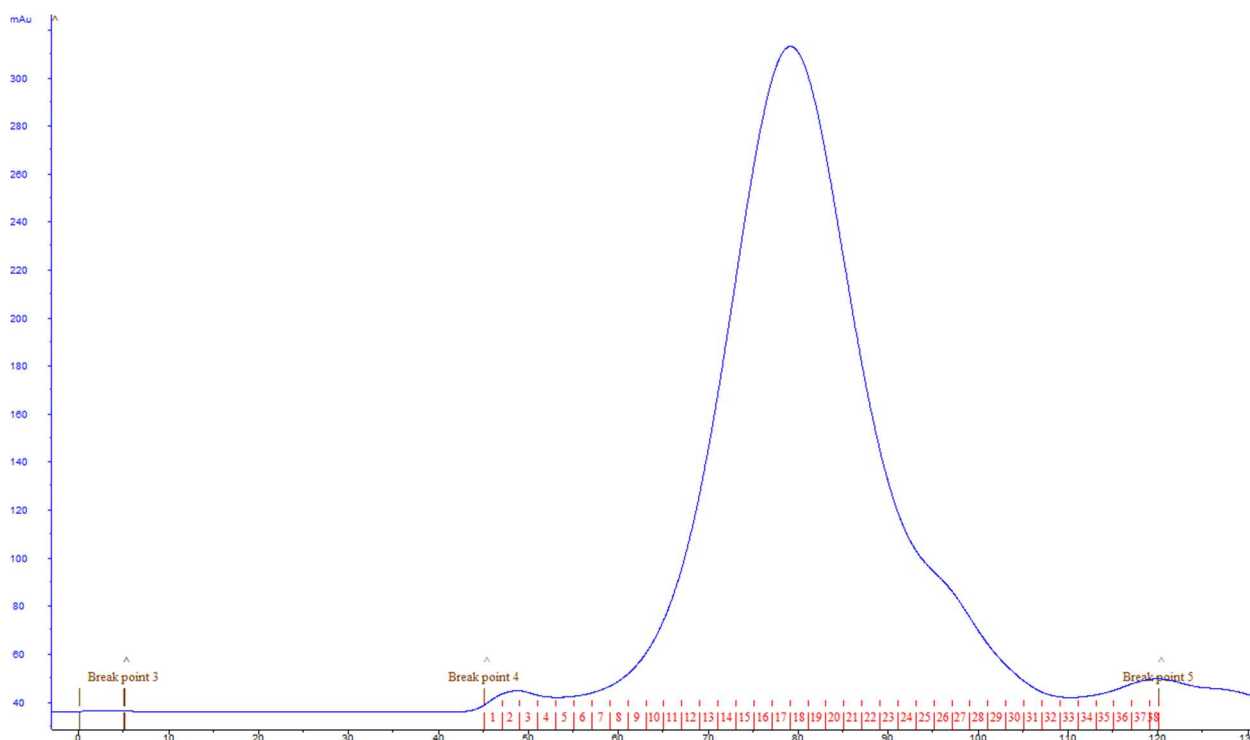


Figure 10 SEC chromatogram on AKTA system of CH16_B1 sample.

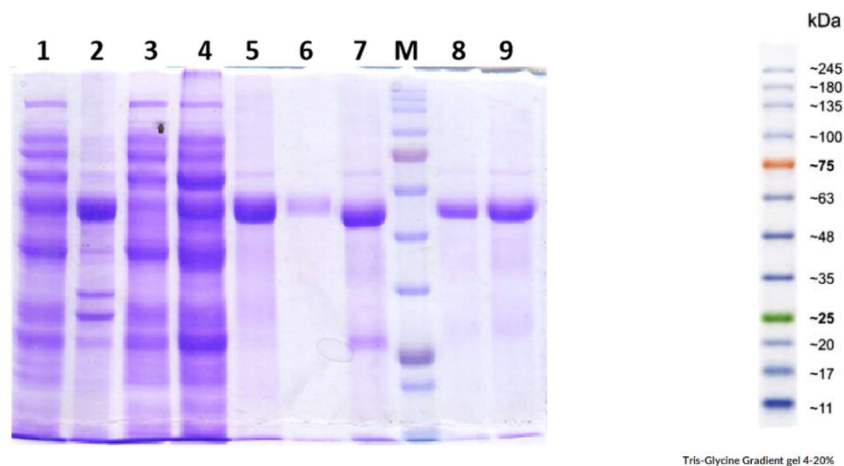


Figure 11 SDS polyacrylamide gel showing the purification steps of CH16_B1., 1. Supernatant 2. Insoluble proteins fraction 3. First nickel column - fraction of not bound proteins, 4. Wash step, 5. Elution step, 6. Trash step, 7. Target protein after dialysis with TEV-shorter by 4kDa (His-tag), M- Protein marker (3-colour prestained DNA-Gdańsk), 8. Second nickel column flow through step, 9. Concentrated sample for gel filtration .

CH19_K3

In this case, after lysis, I heated the supernatant in a water bath at 80°C / 15 minutes (bath temperature 79-83°C). After the sample had cooled down, I centrifuged get it again at 16,000 rpm / 30 minutes / 4 ° C. I applied the prepared supernatant to the nickel column. The sediment consisted of denaturated proteins are treated as impurities in the sample. I continued following the protocol below (Table 7). I used Bradford's reagent to check the presence of protein in each step.

Table 7 Preparation of the Ni²⁺ column:

5ml resins preparation	Binding step	Wash step	Elution step	Trash step
3x 15ml H ₂ O 2x 15ml basic buffer + 50mM imidazole <u>Basic buffer</u> 50mM TRIS pH 7.5 200mM NaCl	Pour the supernatant into the column, mix and leave it for 40 minutes in the cold room. The rest of the steps are carried out at room temperature.	Basic buffer + 50mM imidazole (about 3-4 times) + 100mM imidazole (about 2-3 times) until all not bounded proteins are washed away.	Elution of bounded proteins with basic buffer +300mM imidazole (about 30 ml).	Resins cleaning with basic buffer +600 mM imidazole and regeneration of the resins.

Then I placed the protein solution in a dialysis bag with TEV protease added and performed a two-step dialysis. The first stage lasted approximately 2 hours, the second stage approximately 18 hours in the cold room. I dialyzed into a buffer: 50mM TRIS pH 7.5, 200mM NaCl, 5mM 2-Mercaptoethanol. Table 8 summarizes the purification on the second nickel column which was preformed after dialysis.

Table 8 II Ni²⁺column

5ml resins preparation	Binding step	Flow through step	Trash step
3x 15ml H ₂ O 2x 15ml basic buffer + 50mM imidazole	Pour the dialyzed proteins into the columns, mix and leave for 10-15 minutes in the cold room.	TEV protease and His-tag bind to the column. Target protein do not bind but flow through column.	Resins cleaning with basic buffer +600 mM imidazole and regeneration of the resins.

CH19_K3 was applied to a HiPrep 16/60 Sephacryl S-100 column pre-equilibrated with 25 mM TRIS pH 7.5 and 200 mM NaCl in the AKTA chromatographic system. The fractions (6-11) containing CH19_K3 were concentrated on Amicon Ultra 15 ml centrifugal filter units with a 50000MW cut-off to 7 mg/ml, ready for crystallization. Figure 12 shows the chromatogram for the CH19_K3 protein. The purity of the samples was checked with the SDS-PAGE (Figure 13).

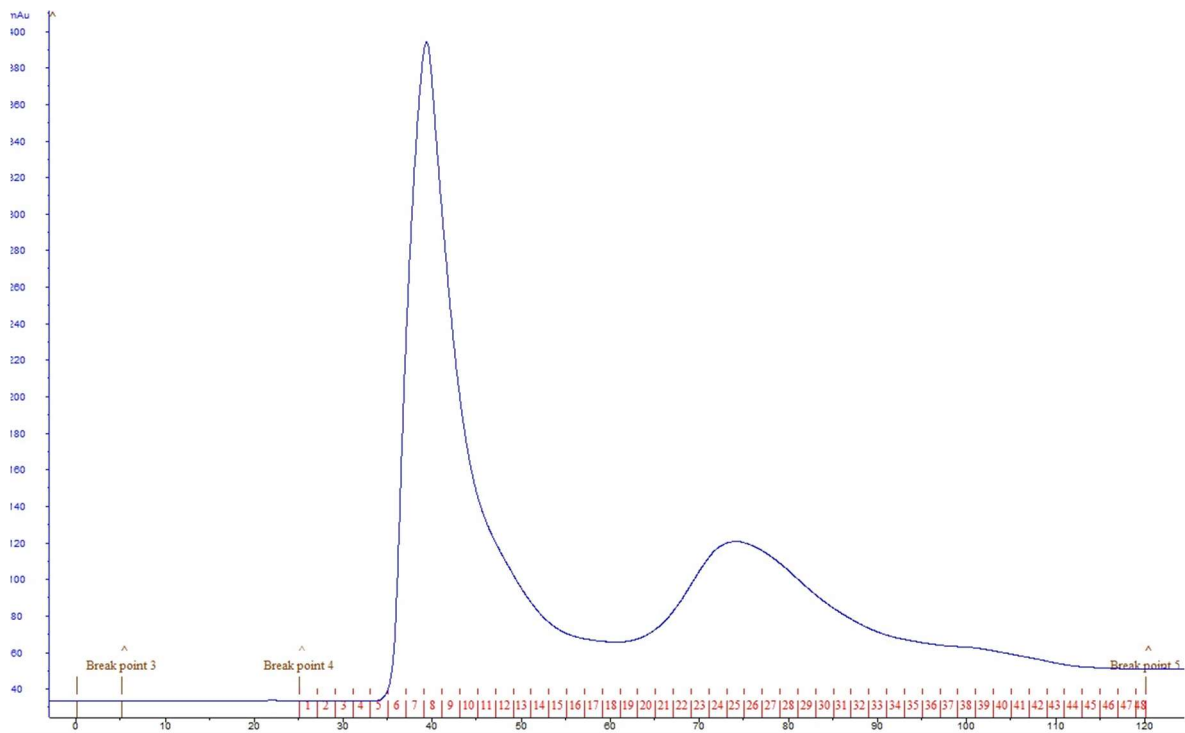


Figure 12 SEC chromatogram on AKTA system of CH19_K3 sample.

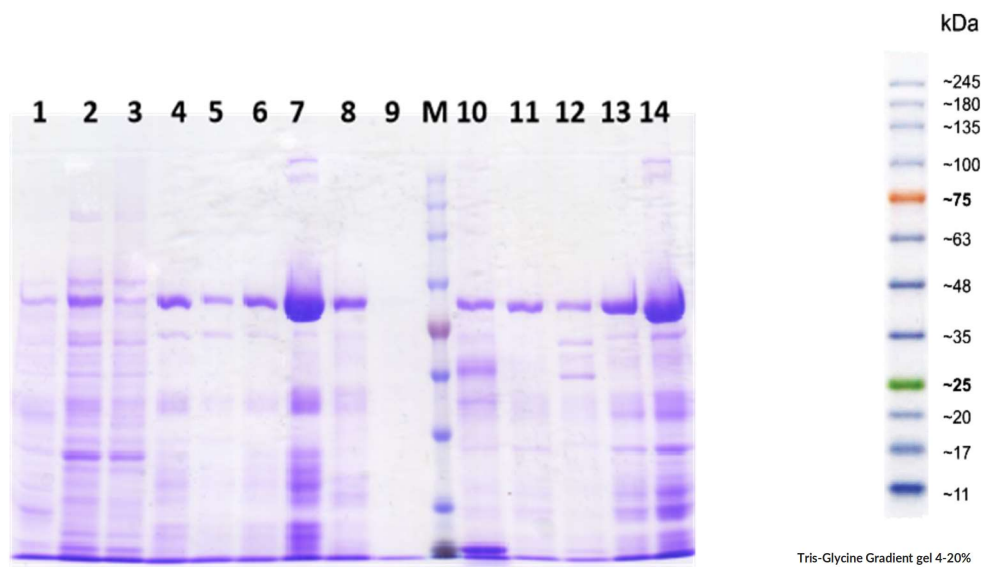


Figure 13 SDS-PAGE showing the purification steps of CH19_K3. 1. Insoluble proteins fraction, 2. Supernatant, 3. Insoluble proteins fraction after heating at 80°C, 4. Supernatant after heating at 80°C 5. First nickel column fraction of not bound proteins, 6. Wash step with 50mM imidazole, 7. Wash step with 100mM imidazole, 8. Elution step, 9. Trash step, M-Protein marker (3-colour prestained DNA-Gdańsk), 10. Protein after dialysis with TEV shorter by 4kDa (His-tag) 11. Second nickel column-flow through step, 12. Second nickel column trash step, 13. Protein prepared SEC, 14. Concentrated protein prepared for SEC.

5.6.2 Diacetylchitobiose deacetylase – Dac-74

After the lysis, I placed the supernatant on a nickel column and followed the protocol described below (Table 9). I used Bradford's reagent to check the presence of protein in each steps.

Table 9 Preparation of the Ni²⁺column

5ml resins preparation	Binding step	Wash step	Elution step	Trash step
3x 15ml H ₂ O 2x 15ml basic buffer + 50mM imidazole <u>Basic buffer</u> <u>50mM TRIS</u> <u>pH 7.5</u> <u>200mM NaCl</u>	Pour the supernatant into the column, mix and leave it for 40 minutes in the cold room. The rest of the steps are carried out at room temperature.	Basic buffer + 50mM imidazole (about 3-4 times) + 100mM imidazole (about 2-3 times) until all not bounded proteins are washed away.	Elution of bounded proteins with basic buffer +300mM imidazole (about 30 ml).	Resins cleaning with basic buffer +600 mM imidazole and regeneration of the resins.

Then I placed the protein solution in a dialysis bag with TEV protease added and performed a two-step dialysis. The first stage lasted approximately 2 hours, the second stage approximately 18 hours in the cold room. I dialyzed into a buffer: 50mM TRIS pH 7.5, 200mM NaCl, 5mM 2-Mercaptoethanol. Table 10 summarizes the purification on the second nickel column.

Table 10 II Ni²⁺column

5ml resins preparation	Binding step	Flow through step	Trash step
3x 15ml H ₂ O 2x 15ml basic buffer + 50mM imidazole	Pour the dialyzed proteins into the columns, mix and leave for 10-15 minutes in the cold room.	TEV protease and His-tag bind to the column. Target protein do not bind but flow through column.	Resins cleaning with basic buffer +600 mM imidazole and regeneration of the resins.

I put the protein solutions in a dialysis bag. I conducted a two-stage dialysis in order to get rid of imidazole and 2-mercaptoethanol. The first stage takes about an hour, the second stage overnight in the cold room. I dialyzed into a buffer: 25mM TRIS pH 7.5, 200mM NaCl, 200 μ M ZnCl₂ (or 10 μ M if the protein solution was prepared for ITC measurements). The protein containing Dac-74 were concentrated on Amicon Ultra 15 ml centrifugal filter units with a 10000MW cut-off to 4-5 mg/ml, ready for crystallization. The purity of the samples was checked with the SDS gel (Figure 14).

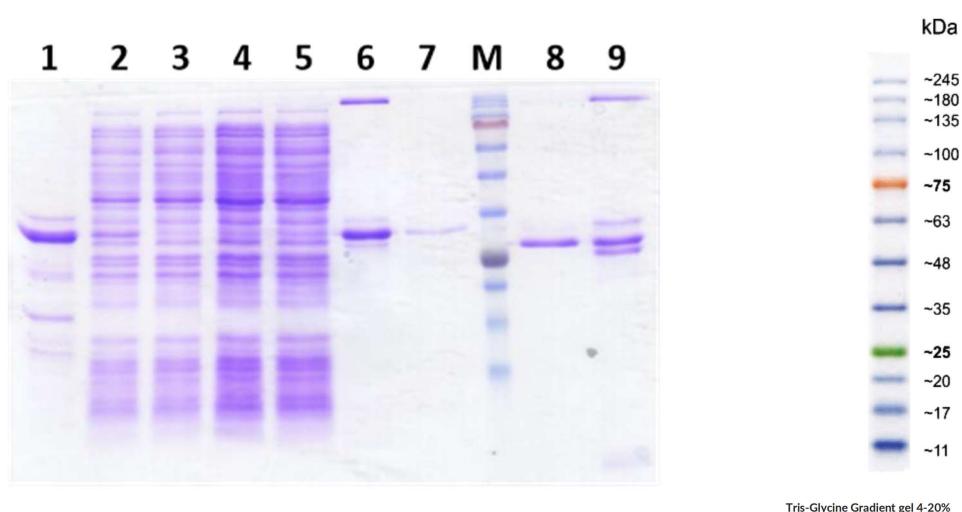


Figure 14 SDS-PAGE showing the purification steps of Dac_74. 1 Insoluble proteins fraction, 2. supernatant, 3. First nickel column fraction of not bound proteins, 4. Wash step with 50mM imidazole, 5. Wash step with 100mM imidazole, 6. Elution step, 7. Trash step, M- Protein marker (3-colour prestained DNA-Gdańsk), 8. Second nickel column flow through step, 9. Trash step.

5.6.3 *Exo*- β -D-glucosaminidase - GImA-01

After the lysis step the supernatant was placed on a nickel column and followed the protocol described below (Table 11). I used Bradford's reagent to check the presence of protein in each steps.

Table 11 Preparation of the Ni²⁺column

5ml resins preparation	Binding step	Wash step	Elution step	Trash step
3x 15ml H ₂ O 2x 15ml basic buffer + 30mM imidazole <u>Basic buffer</u> 50mM TRIS pH 8.5 300mM NaCl	Pour the supernatant into the column, mix and leave it for 40 minutes in the cold room. The rest of the steps are carried out at room temperature.	Basic buffer + 30mM imidazole (about 3-4 times) + 50mM imidazole (about 2-3 times) until all not bounded proteins are washed away.	Elution of bounded proteins with basic buffer +200mM imidazole (about 30 ml).	Resins cleaning with basic buffer +600 mM imidazole and regeneration of the resins.

Then I placed the protein solution in a dialysis bag with TEV protease added and performed a two-step dialysis. The first stage lasted approximately 2 hours, the second stage approximately 18 hours in the cold room. I dialyzed into a buffer 50mM TRIS pH 8.5, 300mM NaCl, 5mM 2-Mercaptoethanol. Table 12 summarizes the purification on the second nickel column.

Table 12 II Ni²⁺column

5ml resins preparation	Binding step	Flow through step	Trash step
3x 15ml H ₂ O 2x 15ml basic buffer + 30mM imidazole	Pour the dialyzed proteins into the columns, mix and leave for 10-15 minutes in the cold room.	TEV protease and His-tag bind to the column. Target protein do not bind but flow through column.	Resins cleaning with basic buffer +600 mM imidazole and regeneration of the resins.

GlmA-01 was applied to a HiPrep 16/60 Sephacryl S-100 column pre-equilibrated with 25 mM TRIS pH 8.5 and 50 mM NaCl in the AKTA chromatographic system. Desalting carried out on the AKTA system, was necessary element for crystallization, because with high salt the protein did not crystallize. The fractions (5-10) containing GlmA-01 were concentrated to 12 mg/ml, ready for crystallization. Figure 15 shows the chromatogram for the GlmA-01 protein. The purity of the samples was checked with the SDS-PAGE (Figure 16).

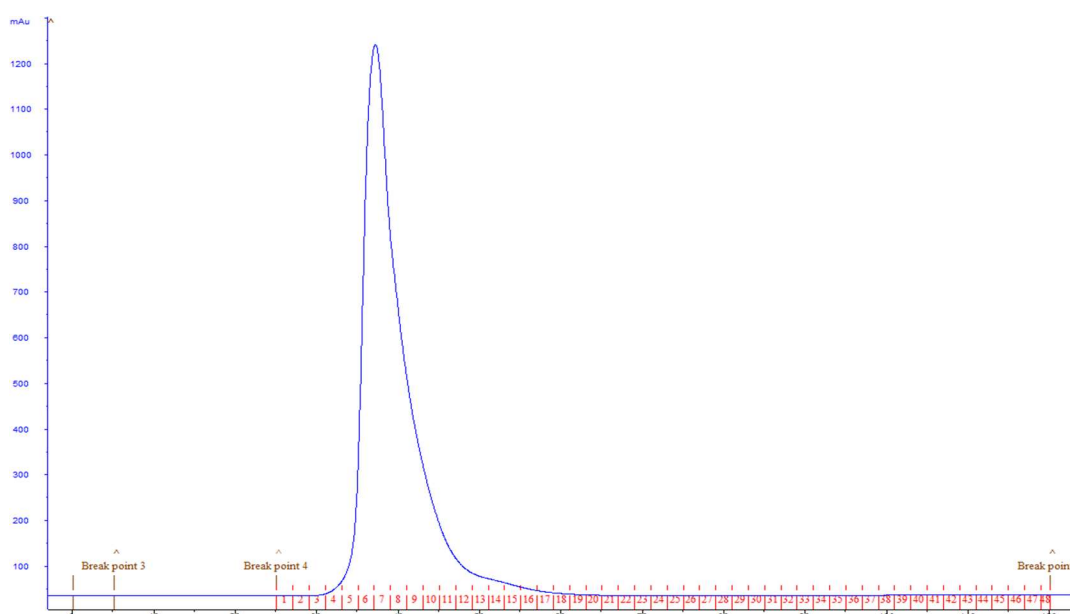


Figure 15 SEC chromatogram on AKTA system of GlmA-01 sample.

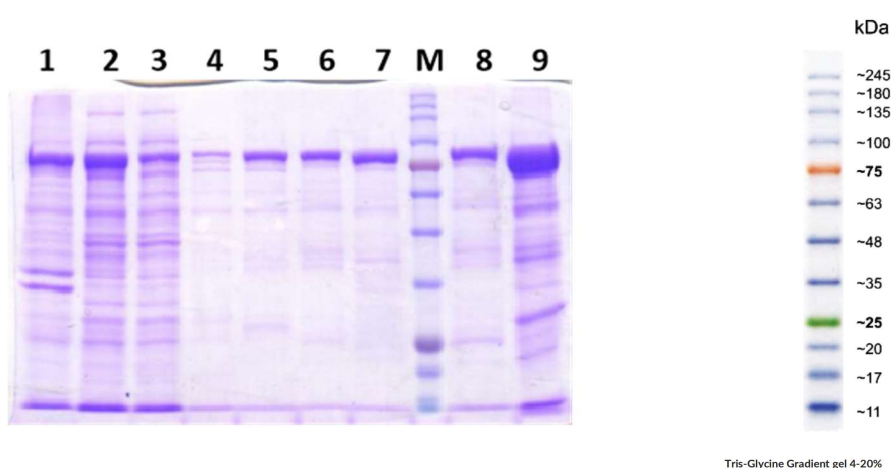


Figure 16 SDS-PAGE showing the purification steps of GlmA-01. 1. Insoluble proteins fraction 2. Supernatant, 3. First nickel column fraction of not bound proteins, 4. Wash step with 30mM imidazole, 5. Wash step with 50mM imidazole, 6. Elution step GlmA-01, 7. Second nickel column flow through step, M- Protein marker (3-colour prestained DNA-Gdańsk), 8.=6. Elution step, 9. Protein prepared for gel filtration.

5.7 Protein crystallization

I established the initial crystallization conditions for Dac-74 & GlmA-01 using the Gryphon crystallization robot (Art Robbins Instruments) and commercial crystallization kits from Hampton Research and Molecular Dimensions Ltd.. I used for this 96-well plates, where I put 60 μ l of the appropriate buffer in each reservoir. The sitting drops had a volume of 0.9 μ l, set in three volumetric ratios: 2: 1, 1:1 and 1: 2.

The optimization I carried out in two temperatures, 18°C and 35°C, in two types of plates for both proteins:

- ✓ the sitting drops: 2.1 μ l, ratios: 1:2, 2:1, 80 μ l in reservoir
- ✓ the hanging drops: 4.5 μ l, ratios: 1:2, 1:1, 2:1 500 μ l in reservoir

To crystallize Dac-74, GlcNAc, (GlcNAc)₂, (GlcNAc)₃ and zinc were used. I was able to obtain only the (GlcNAc)₂ complex. Liganded Dac (Dac-74-lig) crystallized in space group *P*2₁ in the presence of (GlcNAc)₂ added in the 5:1 molar excess of the ligand with respect to the protein, and in 0.2 M ammonium chloride, 0.1 M MES pH 6.0, 20% v/v PEG 6000, at 18°C. Dac-74 (unliganded) crystallized in space group *P*3₂2₁ in 2.0 M ammonium sulfate, 0.1 M HEPES, pH 7.5, 2% v/v PEG 400, at 35°C. Crystals for the anomalous dispersion measurements (Dac-74-anom) were obtained from purified Dac-74 treated with a 1000:1 molar excess of TPEN, for 1.5h, followed by a 3-step dialysis (1h, 4h and overnight). The protein was then incubated in ZnCl₂ in 10:1 molar excess of the metal cation. The protein crystallized in space group *P*2₁2₁2₁ in 2.0 M ammonium sulfate, 0.1 M HEPES, pH 7.5, 2% v/v PEG 400, at 35°C, similarly to the trigonal Dac-74 crystals (Figure 17).

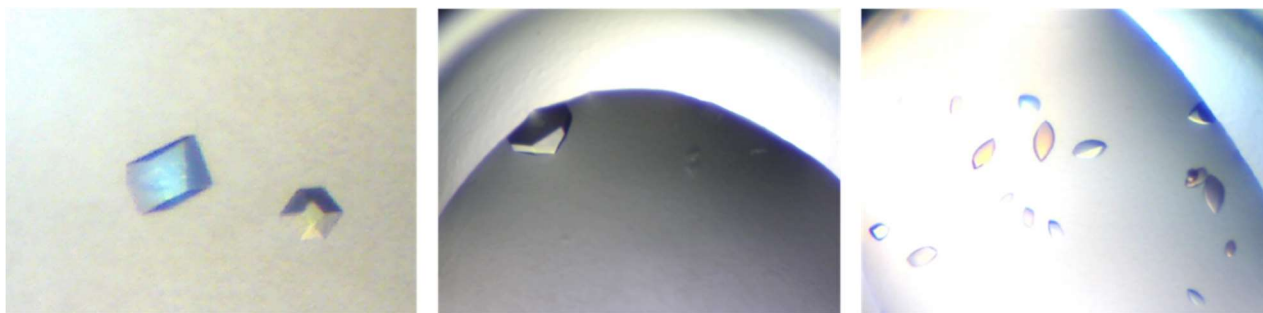


Figure 17 Dac-74 crystals obtained under various crystallization conditions.

For crystallization of GlmA-01 I added GlcNAc and (GlcNAc)₂, however in these cases I did not obtain any crystals with ligand. Unfortunately, due to the limited amount of substrate (GlcN-GlcNAc) crystallization setting was impossible. GlmA-01 crystallized in space group C222₁ in 0.1 M SPG buffer (Molecular Dimensions Ltd., cat. no. MD2-59), pH 8.0, 25% v/v PEG 1500 at 18°C.

For purified chitinases CH16_B1 and CH19_K3 screening crystallizations were carried out by using a crystallization robot Gryphon (Art Robbins Instruments). In this method, seven sets of crystallization solutions were tested, each consisted of 96 conditions, which gives a total of 672 tested conditions. Moreover, the crystallizations were set at three different temperatures: 4, 18 and 35°C. However, none of these attempts were successful.

Therefore, I decided to test the capabilities of the Alpha Fold 2 server (Mirdita et al., 2021) for predicting structure models. The results showed the limitations of this method and how dependent it is on the deposited in the PDB database structures, especially those obtained experimentally. It is easy to see in Figure 18 that the structure for CH19_K3 looks better than that of CH16_B1.

Working with CH19_K3 protein I was searching constantly for information about this protein in different data bases. In 2016 I found that Horiuchi et al. published the structure of chitinase from *Pyrococcus chitonophagus* (sequence identical to CH19_K3) but reduced to catalytic domain only (Horiuchi et al. 2016). Alpha Fold 2 easily found that domain but

didn't managed with the rest of the protein (predicted two chitin binding domains, present in the CH19_K3 sequence), that is why only loops are seen on prediction structure.

For CH16_B1 chitinase it is expected that protein will have one catalytic and one chitin binding domain. However sequence similarity searching (e.g. BLAST) allowed me to find similar structures of catalytic domains only (PDB codes: 1ITX , 5GZT, 5GZK) at about 35% level. Hence, the proposed model is less accurate, because the written algorithms did not have a good base to use.

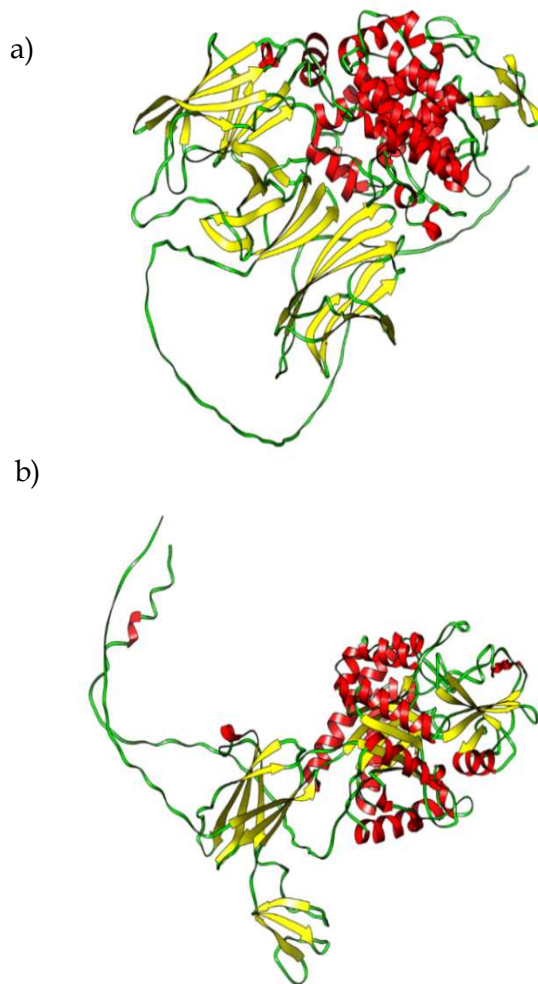


Figure 18 The predicted structures of chitinases from *P. chitonophagus* (a) CH19_K3 , (b) CH16_B1 created with the help of Alpha Fold 2 server (Mirdita et al., 2021).

5.8 Diffraction measurements

The synchrotron X-ray diffraction data were collected at the temperature of 100 K, at beamline P13 operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany) (Cianci et al., 2017). Prior to freezing, the crystals were immersed in cryoprotecting solution obtained by mixing the reservoir solution with ethylene glycol, to the final concentration of the latter of 25% v/v. (GlcNAc)₂ or Zn²⁺ were added to the cryoprotection solutions, as appropriate for Dac-74-lig and Dac-74-anom, to prevent their disassociation from the protein.

Dac-74-lig and Dac-74 X-ray data were collected at a convenient fixed X-ray energy, whereas the Dac-74-anom data collection experiment was preceded by an energy scan to determine the X-ray fluorescence peak as the appropriate energy for data collection. Examples of recorded diffraction images are shown in Figure 19.

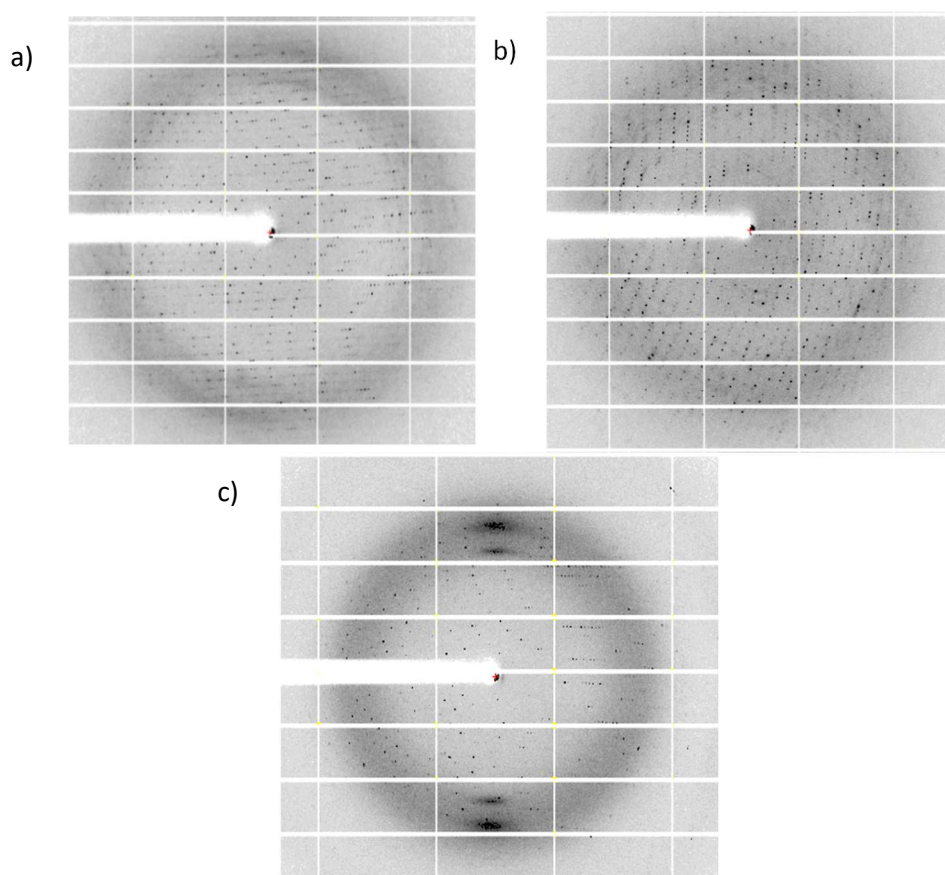


Figure 19 Examples of diffraction frames from three data sets collected from (a) Dac-74-lig (b) Dac74 (c) GlmA-01.

5.9 Solving and refinement of the structures

The X-ray diffraction data were processed using the XDS program (Kabsch & IUCr, 2010). The X-ray data are summarized in the table 13. The structures were solved using Phaser from the CCP4 program suite. The structure of Dac-74 was solved by molecular replacement using the PDB entry 4XM0 as the starting model (Nakamura et al., 2015). The structures of Dac-74-lig and Dac-74-anom were solved using the Dac-74 as the starting model. The anomalous density maps for Dac-74-anom were calculated using as Fourier coefficients the anomalous amplitudes of the reflections, $|F^+-F^-|$, and phases calculated from the refined atomic coordinates and retarded by 90° . The structure of GlnA-01 was solved likewise, using the PDB model 5GSL (Mine et al., 2017). The molecular models were built using Coot (Emsley et al., 2010) and refined with Refmac5 (Murshudov et al., 2011).

Table 13 Summary of the X-ray data and the final models.

Protein	Dac-74-lig	Dac-74	Dac-74-anom[‡]	GlmA-01
PDB code	8BGO	8BGN	8BGP	
Space group	<i>P</i> 2 ₁	<i>P</i> 3 ₂ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>C</i> 222 ₁
Unit cell parameters:				
<i>a</i> (Å)	79.2	150.6	78.6	89.4
<i>b</i> (Å)	151.7	150.6	158.7	120.5
<i>c</i> (Å)	167.1	72.2	158.6	145.6
α (°)	90	90	90	90
β (°)	93.3	90	90	90
γ (°)	90	120	90	90
Wavelength (Å)	0.9797	0.9797	1.2815	0.9763
Resolution (Å)	3.08	2.76	2.51	2.1
R _{merge} ^{*#}	0.077 (0.657)	0.114 (0.966)	0.199 (1.090)	0.111 (1.66)
R _{pim} [†]			0.057 (0.303)	
Completeness (%)	99.3	99.08	94.2	99.5
Observed reflections	488298	470509	565577	612215
Unique reflections	6943	24234	43577	44987
<I/ σ (I)>	14.6 (2.0)	18.4 (2.4)	11.1 (3.03)	19.4 (1.4)
Protein subunits/asymmetric unit	12	3	6	1
R [§] /R _{free} [§]	0.1687/0.2147	0.1763/0.2271	0.2183/0.2614	0.165/0.219

[‡]Diffraction data were processed and statistics calculated by STARANISO (Tickle et al., 2018)

*Values in brackets are for the highest resolution shell.

#R_{merge} = $\frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ is the integrated intensity of a given reflection and $\langle I(hkl) \rangle$ is the mean intensity of multiple corresponding symmetry-related reflections.

†R_{pim} = $\frac{\sum_{hkl} (1/n-1)^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$.

§R = $\frac{\sum_{hkl} ||F_{obs}| - |F_{calc}||}{\sum_{hkl} |F_{obs}|}$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.

§R_{free} is R calculated using a randomly chosen subset of reflections excluded from the refinement.

6. Results and discussion

6.1 Diacetylchitobiose deacetylase - Dac-74

I obtained three structures for diacetylchitobiose deacetylase. In the first (Dac-74), there was no ligand added. I observed a metal ion binding site, but no anomalous data have been collected for it. The second (Dac-74-anom) is the structure for which anomalous data were collected, confirming the presence of zinc in the structure. The third (Dac-74-lig) has the lowest resolution but ligand molecules are present in the structure.

6.1.1 Packing in the unit cell

The three deacetylase structures belong to three different crystalline forms. The unliganded Dac-74 crystallized in space group $P3_221$, with half a protein hexamer in the asymmetric unit. The other three subunits are related by the crystallographic two-fold axis, resulting in a hexameric assembly. Dac-74-anom gave $P2_12_12_1$ crystals containing one hexamer in the asymmetric unit. Dac-74-lig crystallized in the $P2_1$ space group and contains a complex of Dac-74 with its substrate, (GlcNAc)₂. There are two protein hexamers in the asymmetric unit. Figures 20 to 22 show the packing of molecules in the unit cell.

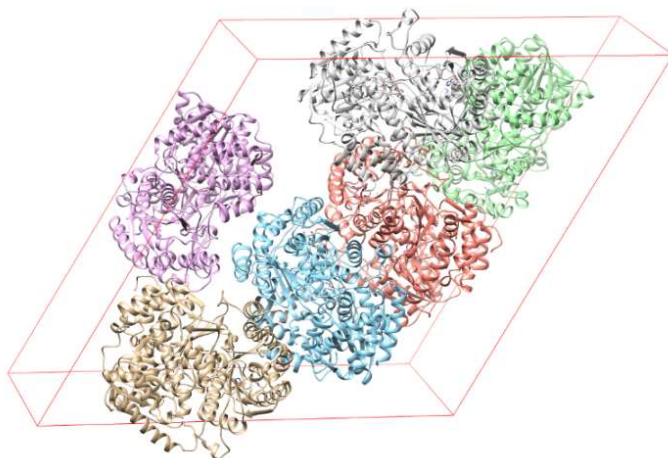


Figure 20 The packing of Dac-74 in the unit cell.

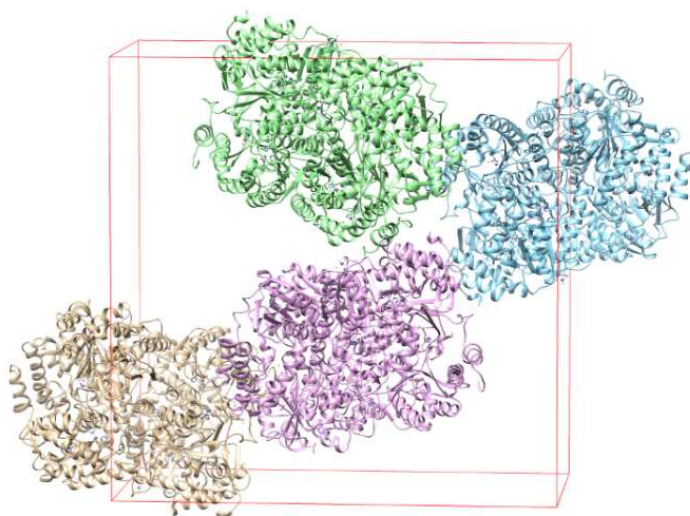


Figure 21 The packing of Dac-74-anom in the unit cell.

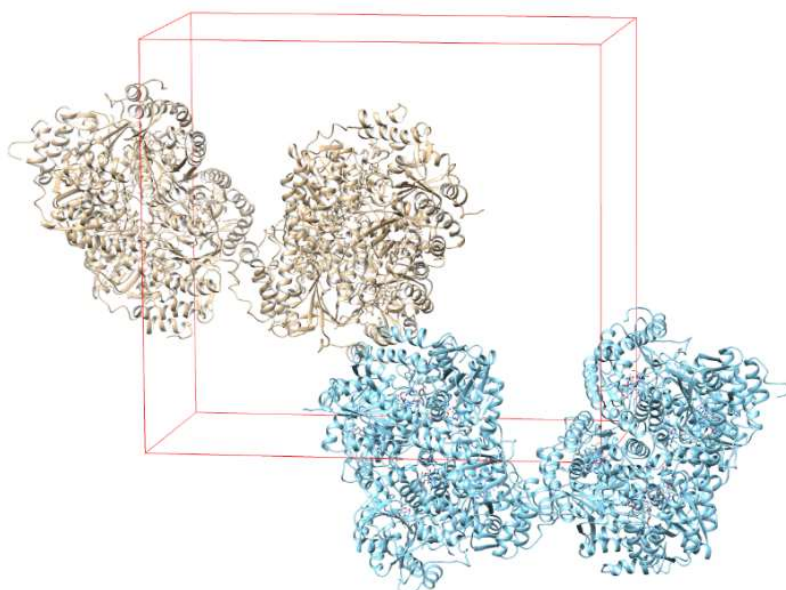


Figure 22 The way of packing Dac-74-lig in the unit cell.

Comparison of the different crystal forms shows a consistent hexameric association of the subunits, indicating that hexamer is the physiological form of the protein.

6.1.2 Dac-74 oligomerization in solution

The oligomerization of the protein was verified in solution by means of small-angle X-ray scattering (SAXS) at PETRA III synchrotron by Dr. Ahmed Mohammed. The SAXS studies indicated that Dac-74 in solution occurs mainly as a hexamer, with some indication of higher oligoassociation, accounting for approximately 15% of the signal (Figure 23). SAXS combined with size-exclusion chromatography (SEC-SAXS) confirmed hexamer as the dominant form of oligomerization in solution (Figure 24).

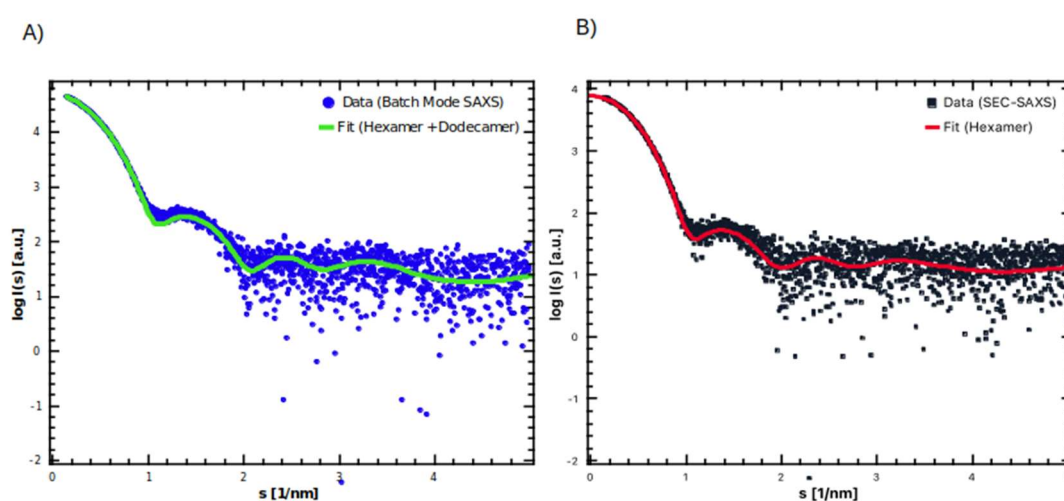


Figure 23 (A) Small-angle X-ray Scattering data from Dac-74 deacetylase at 4.7 mg/ml collected in batch mode (blue) and the scattering from a mixture of model hexamers and dodecamers computed by OLIGOMER (green); (B) SEC-SAXS data from the main elution peak of deacetylase (black) and the intensity from the hexamer model computed by CRY SOL (red).

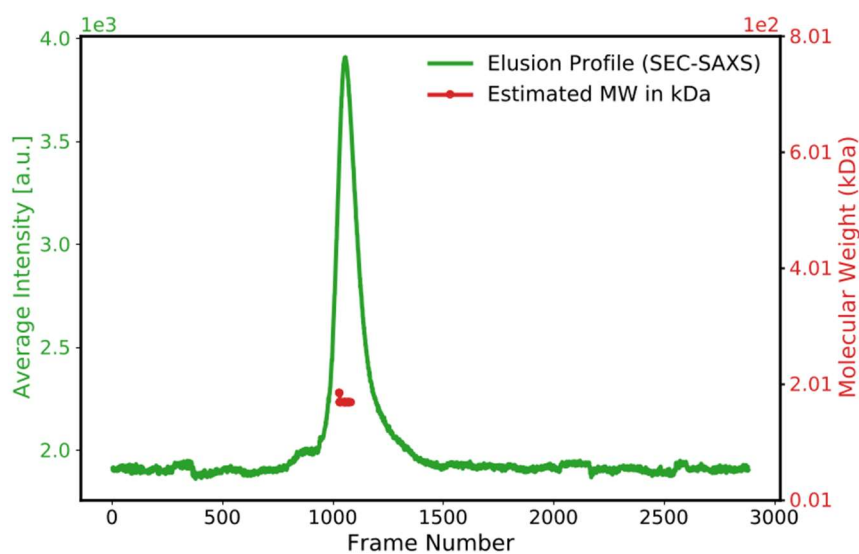


Figure 24 SEC-SAXS elution profile of deacetylase computed by CHROMICS and displaying one predominant elution peak (the MW across the peak is displayed in red).

6.1.3 Identifying the prosthetic group by anomalous X-ray scattering.

The presence of Zn^{2+} in the crystal structure of Dac-74 was confirmed by X-ray anomalous dispersion. Energy scan performed on the Dac-74-anom crystals revealed X-ray absorption profile characteristic of zinc, while anomalous density maps clearly identified the Zn^{2+} sites as peaks at the level of 18-25 r.m.s.d.

Additional measurements, by means of microcalorimetry (ITC) were performed by Dr. Joanna Śliwiak to determine Zn^{2+} affinity and stoichiometry. The ITC results confirmed that Zn^{2+} is the expected metal cation to bind Dac-74, although other metal cations, such as Cd^{2+} or Ni^{2+} show comparable affinities and they can substitute Zn^{2+} , if they are present.

6.1.4 Differential Scanning Calorimetry (DSC)

Differential Scanning Calorimetry experiments for Dac-74, which I performed, indicated extremely high protein melting temperature 99 °C (Figure 25). Asymmetric shape of the melting profile indicates that unfolding of the protein is not a simple one-step mechanism. It probably involves de-oligomerisation before unfolding of the individual subunits.

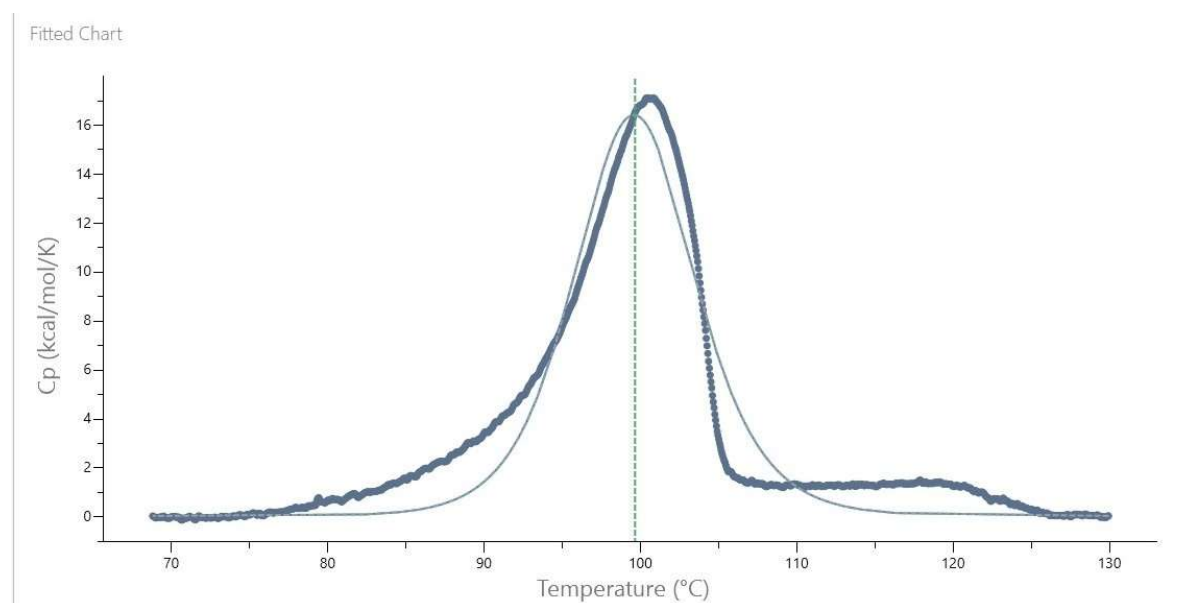


Figure 25 DSC profiles for heating scans of Dac-74. The thick line is the instrumental data, the thin line marks the fitted profile.

6.1.5 The structure of Dac-74

Dac-74 can be described as ‘a dimer of trimers’, with two doughnut-shaped trimers stacked back-to-back along a molecular three-fold axis, with a 8-10 Å channel through the middle of this assembly (Figure 26).

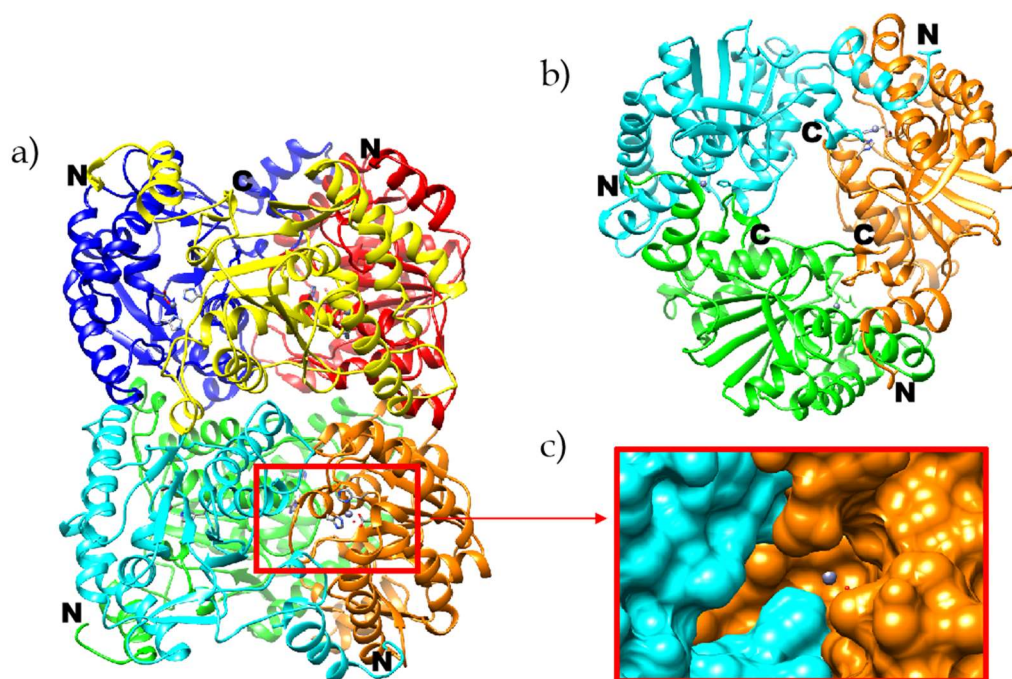


Figure 26 Ribbon diagram of a Dac-74 hexamer (a), viewed along the long molecular axis (b), only three polypeptide chains are shown). The Zn^{2+} cations (blue spheres) and their chelating protein residues are drawn in detail, indicating the active sites. The N- and C-termini are marked. Surface representation in the magnified inset illustrates the shape of the substrate-binding pocket (c).

Each Dac-74 subunit is an α/β domain comprising seven β -strands (of which one is anti-parallel to the others) sandwiched between α -helices (Figure 27). Approximately 20 N-terminal residues and 13 C-terminal residues form an extension reaching to the neighboring subunit and forming part of its substrate-binding site. Each Dac-74 subunit contains a metal cation bound by a conserved His-Asp-His metal-ion-binding triad (His40-Asp43-His151), characteristic of the Carbohydrate Esterase Family 14 (CE-14) (Figure 28) of which the most similar known structures are diacetylchitobiose deacetylases from *Pyrococcus horikoshii* and *Pyrococcus furiosus* (Mine et al. 2012).

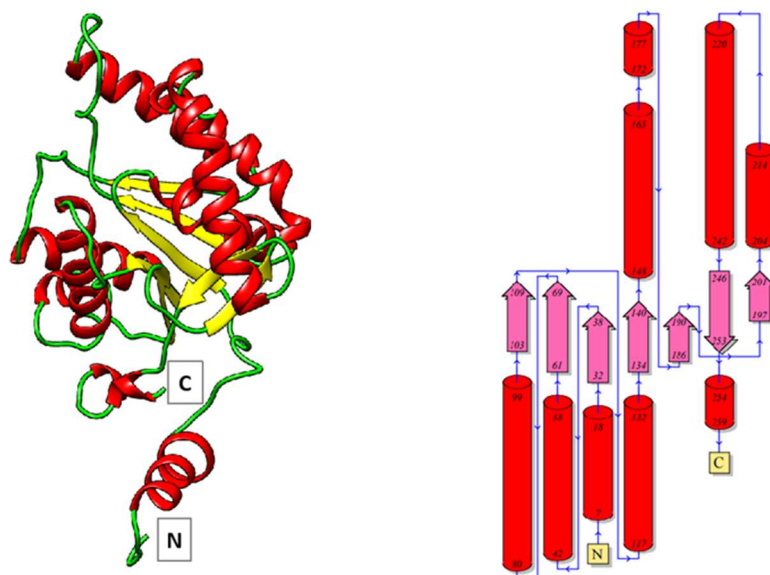


Figure 27 The ribbon diagram of the Dac-74 subunit and its topology diagram.

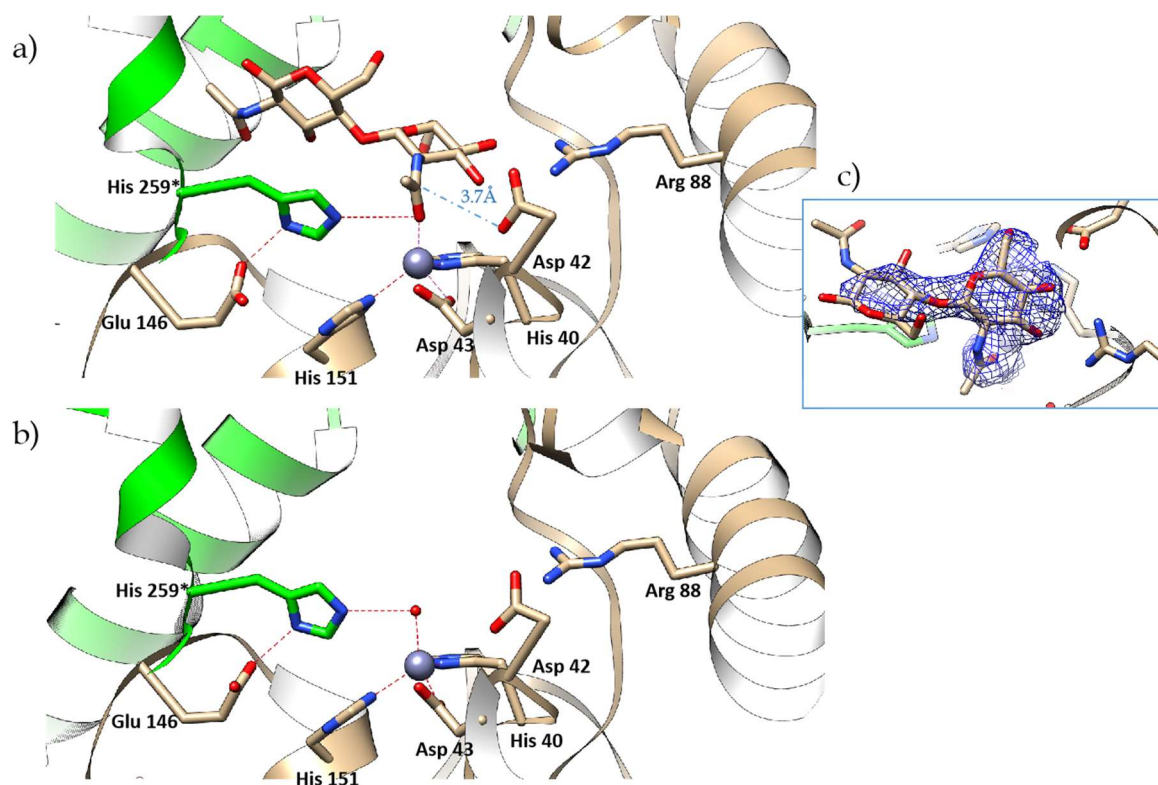


Figure 28 Details of (a) the substrate-binding site of Dac-74-lig structure with bound (GlcNAc)₂ and (b) Dac-74 structure with a single water molecule bound to Zn²⁺. The Zn²⁺ cation is drawn as a blue sphere, water is a small red sphere. Hydrogen bonds are marked with red dotted lines. The distance from Asp42 to the C atom of the scissile bond of the substrate is indicated by a blue dashed line. Asterisk (*) marks His259 that forms part of the substrate-binding site although it comes from a neighboring protein subunit. A comparison of the substrate-bound and unliganded structures shows that the presence of the substrate does not seem to make any significant change to the structure of the protein. Part c shows electron density, 2Fo-Fc, at 1σ level, of the ligand.

6.1.6 Substrate binding and proposed reaction mechanisms for Dac-74

A (GlcNAc)₂ substrate molecule was found in each substrate-binding cavity of the 12 crystallographically independent subunits in the Dac-74-lig structure. The electron density clearly indicated two connected sugar rings (Figure 29).

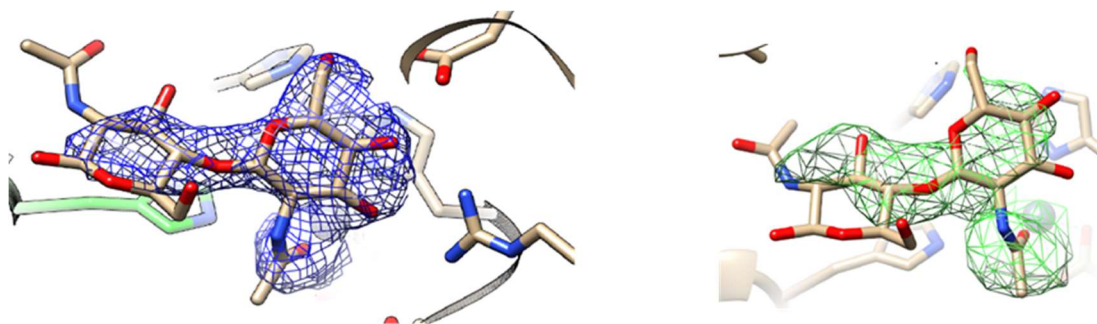


Figure 29 Density map for ligand -(GlcNAc)₂ 2Fo-Fc, at 1σ level (left), and the omit map at 3σ level (right).

The bound carbohydrate is oriented with its non-reducing end toward the active site, consistently with other enzymes of the CE-14 family and differently from CE4 enzymes which act on the reducing end of their sugar substrates. The GlcNAc moiety that is closer to the active site is held in place by a network of hydrogen bonds to the surrounding protein residues (Figure 30).

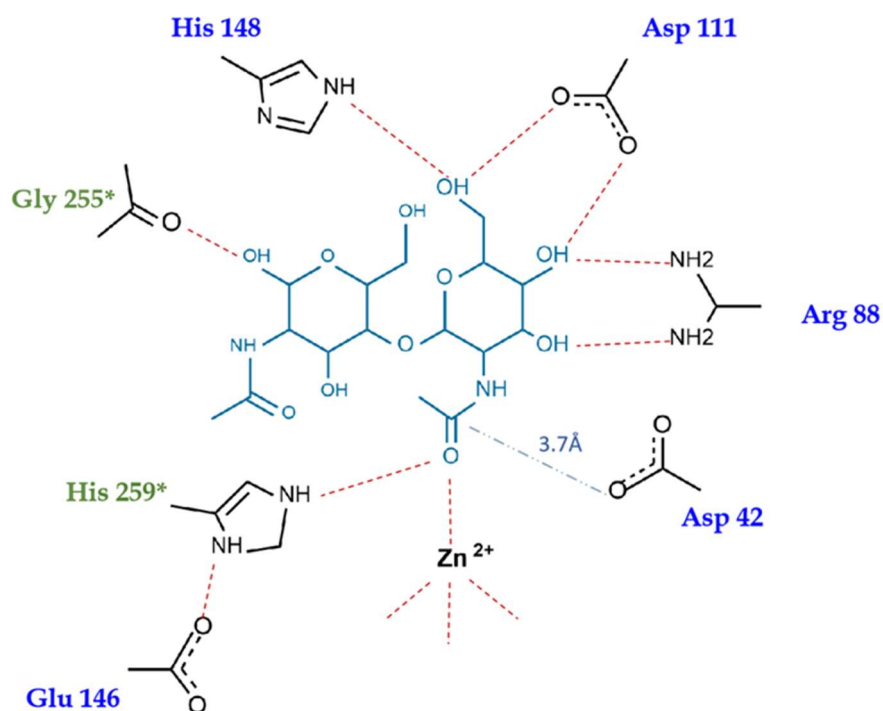


Figure 30 Scheme of (GlcNAc)₂ substrate binding to Dac-74. Hydrogen bonds are indicated by red dotted lines, the distance from Asp42 to the C atom at the scissile bond is marked by a blue dashed line. Residues that form part of the substrate binding site, that come from the neighboring subunit, are marked with.

The N-acetyl group is directed toward Zn²⁺ and its carbonyl oxygen atom completes the tetrahedral coordination sphere of the Zn²⁺ ion (the Zn²⁺ to O distance is 2.1 ± 0.03 Å). The substrate's N-acetyl group is flanked by Asp42 and His259* (* denotes that the residue comes from the neighboring subunit). Its position is equidistant from the two flanking residues (3.5 ± 0.2 Å from the carbonyl oxygen to Nε2 of His259 and to Oδ1 of Asp42). Interestingly, Asp42 is consistently the most outlying residue on the Ramachandran plot (Figure 31).

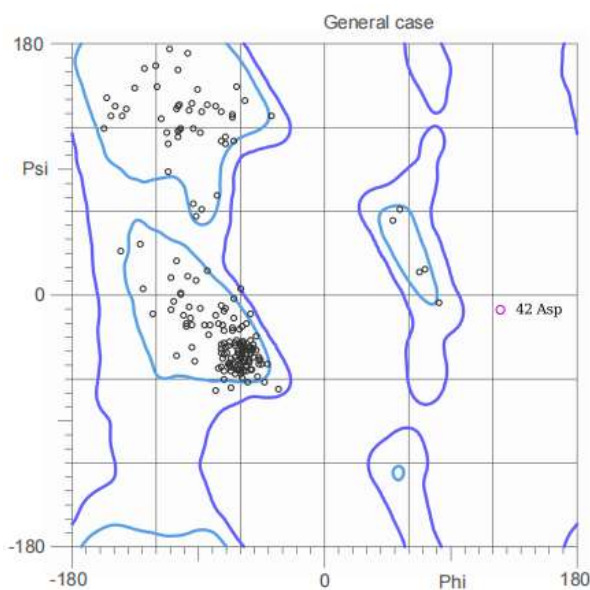


Figure 31 Ramachandran plot of Dac-74, showing Asp42

Asp42 lies at the beginning of a short helix preceded by a Pro residue. The methyl part of the N-acetyl group of the substrate is lodged in a hydrophobic pocket formed by Ile46, Phe219, Trp227 and Ile260*.

Position of the other sugar moiety is less well defined but can be fitted in the electron density. It makes one H-bonding interaction with the protein: between the O1 hydroxyl group and the carbonyl oxygen atom of Gly255 of the neighboring subunit. It is unusual in crystallography to find a substrate molecule in an enzyme's active site. One would rather expect the substrate to be processed by the enzyme before the diffraction measurements can be carried out. Possible factors that could explain the observed stability of the substrate-enzyme complex are: (1) sub-optimal temperature (ca. 20°C) and pH (6.0) of the solution containing the crystal. The optimal temperature for the reaction is above 75°C, and optimal pH is 7.5 (see ITC results below). (2) A large excess of fresh substrate present in the cryosolution in which the crystal was immersed just before being frozen. (3) The diffraction measurements were performed in cryogenic conditions (100 K). In Dac-74 and Dac-74-anom, there was no ligand in the substrate-binding sites and the coordination sphere of Zn^{2+} consists of the three conserved His-Asp-His residues and a single water molecule that occupies the same position as the

carbonyl oxygen of the substrate's N-acetyl group in the Dac-74-lig complex. Thus with or without the substrate, the Zn^{2+} cation is coordinated tetrahedrally (Figure 30).

This calls for some discussion with regard to the reaction mechanism. Two different models have been proposed in the literature concerning the broad family of hydrolytic enzymes containing Zn^{2+} , that include proteinases/peptidases and deacetylases (Feliciano & da Silva, 2015). In one model the reaction involves a hydrolytic water molecule as the first step, leading to a tetrahedral intermediate product. An alternative model involves a nucleophilic attack by a conserved Glu or Asp, with the formation of an anhydride intermediate, followed by hydrolysis as the second step. Different studies, by various techniques (NMR, crystallography, QM) of the related enzymes support one model or the other (Feliciano & da Silva, 2015). Figure 32 below illustrates both the proposed reaction mechanisms.

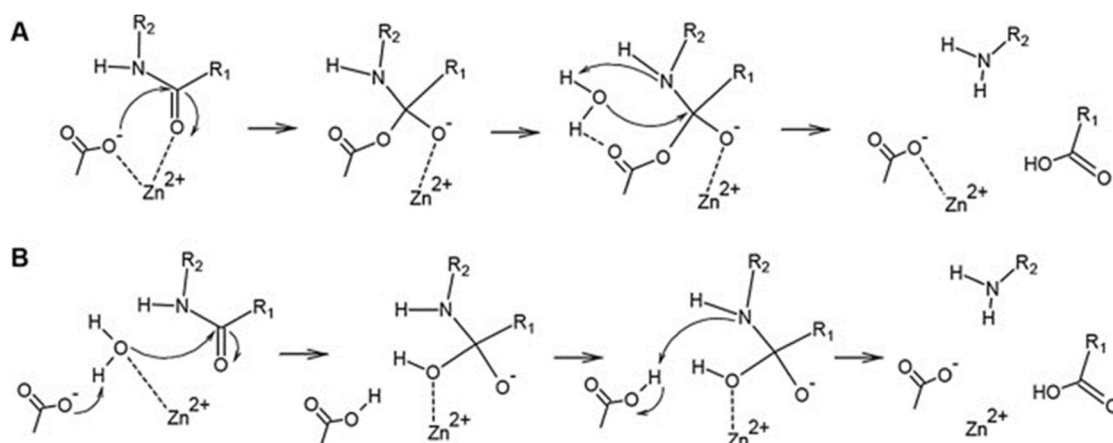


Figure 32 Alternative reaction mechanisms (A) glutamate nucleophilic attack and (B) water nucleophilic attack (Feliciano & da Silva, 2015).

Related crystal structures described in the literature show the following:

- ✓ The unliganded lipoglycopeptide antibiotic deacetylase (Orf2*) has two clear water molecules bound to the Zn^{2+} cation, thus making the Zn^{2+} pentacoordinated (PDB ID 3DFF) (Zou et al., 2008). This is consistent with the model according to which one water molecules is displaced

by the substrate while the second is the 'hydrolytic water' acting as the nucleophile. This enzyme, however, has a different architecture of the binding site from Dac-74. The gap between the Asp and His residues, flanking the active site, is wider in Orf2* by approximately 2 Å (ca 9 Å), compared to Dac-74 (7 Å), probably due to the different nature of its substrate. The wider substrate binding site in Orf2* can more easily accommodate the two water molecules. An amino acid sequence comparison of the two enzymes shows that whereas the Asp residues come from corresponding places, the His residues are unrelated. The monomeric Orf2* is self-contained, whereas in the oligomeric Dac-74 the catalytic His259* comes from the neighboring subunit.

- ✓ The two deacetylases from *Pyrococcus horikoshii* and *Pyrococcus furiosus* that are closely related to Dac-74 have been described with Zn²⁺ or Cd²⁺ in the metal-binding site and different ligands (Mine et al. 2014a) (Nakamura et al. 2015a) (Nakamura et al. 2016). In the complex with the reaction intermediate analog (MPG), two oxygen atoms of the ligand approach the active site but they interact asymmetrically: one makes a direct interaction with Zn²⁺ (the average distance is 1.9 Å) while the other is more distant (2.8 Å). In these complexes, Zn²⁺ appears tetrahedrally coordinated (Nakamura et al. 2016). Asymmetric interactions with Zn²⁺ are also observed in complexes with the phosphate: one of the oxygen atoms is closer to the cation (1.9 Å) than the other (3.1 Å) (PDB ID 3W13) and in a complex with an acetate ion (1.9 and 2.4 Å) (PDB ID 3WE7) (Mine et al. 2014a). In the absence of ligands, the six crystallographically independent subunits contain water molecules that can be interpreted as either partially disordered or they can be modeled as a single site near Zn²⁺ or as two water molecules but also at different distances to the Zn²⁺ cation (PDB ID 4XM0, 4XM2) (Nakamura et al. 2015a). Overall, Zn²⁺ shows a clear tendency to be tetrahedrally coordinated, taking one ligand in addition to the conserved His-Asp-His triad. This is different from the Cd²⁺-

substituted proteins, in which the Cd^{2+} cations show a tendency to be octahedrally coordinated, with some distortions to this geometry or one ligand missing in the geometrically octahedral six-ligand coordination shell (PDB ID 3W14, 4XM1, 4XLZ) (Mine et al. 2014b) (Nakamura et al. 2015)

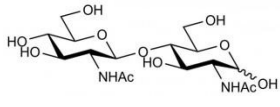
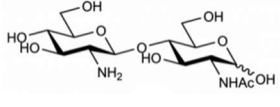
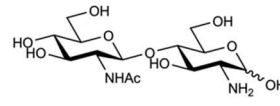
In summary, the Zn^{2+} cations in Dac-74 and in closely related enzymes show a propensity to be tetraordinated and external ligands approach it in a monodentate rather than bidentate manner. This disfavors the model according to which the enzyme in its resting state has two water molecules at the Zn^{2+} ion, one of which is displaced by the substrate's carbonyl oxygen, while the other acts as the hydrolytic water. In Dac-74, the Zn^{2+} is clearly tetraordinated and the single water molecule is displaced by the substrate. This does not exclude the possibility of hydrolysis as the first step of the reaction, as the environment is of course fully hydrated. There is, however, no defined site for a water molecule, from which it could carry out the hydrolytic attack, and it is not clear how this water could approach the substrate that is lodged equidistantly between Asp42 and His259* (ca. 3.5 Å from each). It seems necessary for the structure to change before the hydrolysis could take place. One could also consider the possibility of the reaction proceeding through an anhydride intermediate.

6.1.7 Enzymatic activity

The enzymatic activity for Dac-74 deacetylases was verified by ITC studies performed in the Laboratory of Protein Engineering by Dr. Joanna Śliwiak, and NMR tests performed in the Department of Biomolecular NMR by Dr. Daniel Baranowski

Table 14 summarizes the study of Dac-74 activity by ITC and NMR on substrates (GlcNAc)₂ and on compounds singly acetylated on the reducing end or the non-reducing end. When I was preparing the enzyme for the activity tests, I observe to activity it was necessary to anneal the protein by incubating it at 60°C, for at least 5 min. The result of these tests showed that the enzyme's performance is best for the doubly acetylated substrate, while the singly acetylated substrate on the non-reducing end works but with lower activity, and for the singly acetylated substrate on the reducing end, the reaction does not take place.

Table 13 Activity tests for Dac-74.

Substrate	ITC	NMR
	+	+
	-	-
	+	+

The obtained test results indicate that the enzyme works in accordance with what is described in the literature for deacetylases from family 14, for which substrate deacetylation takes place at the non-reducing end.

Activity tests by means of ITC were also performed for GlcNAc and (GlcNAc)₃, however in these cases the enzyme efficiency was lower and its working conditions were different. Which gave me a confirmation that the

target substrate was (GlcNAc)₂. The shape and size of the substrate-binding pocket also seems to be well suited to accommodate (GlcNAc)₂ (Figure 33), whereas the smaller compound would not fit in the binding cavity as tightly and its binding specificity is expected to be additionally reduced at elevated temperatures due to diffusion. On the other hand, ligands larger than (GlcNAc)₂ would protrude from the binding cavity and the hydrogen bond with Gly255* could not form due to the glycosidic bond between the second and third sugar moiety.

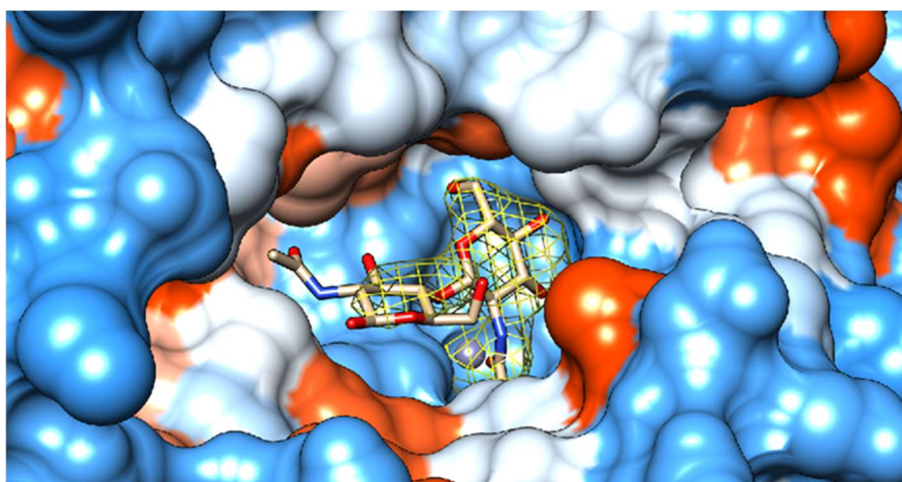


Figure 33 Surface view of the substrate-binding pocket of *Pch*-Dac with bound (GlcNAc)₂. The color of the protein surface indicates the electrostatic charge: negative (*red*), positive (*blue*) and neutral (*white*). The Zn²⁺ ion is shown as a blue sphere. The bound substrate is drawn as sticks. The contours show the 2F_o-F_c electron density of the ligand, at 1σ level.

6.1.8 Conclusions of Dac-74

The extended study of Dac-74 revealed its molecular structure in the crystal. This was supplemented by collaborative structural studies in solution and studies of the protein's thermodynamic stability, enzymatic activity under varying conditions, and substrate specificity. Dac-74 was consistently hexameric in three different crystal forms and in solution, which implied that this was the biologically relevant assembly. The oligomeric structure and individual subunits remained stable up to temperatures of approximately 100°C. The identity of the bound metal cation was confirmed to be Zn^{2+} , although other metal cations were found to also bind with comparable affinities. The enzyme showed a relatively high activity toward $(GlcNAc)_2$, the main product of chitin degradation by chitinases, but $GlcNAc$ and $(GlcNAc)_3$ were also processed, albeit with lower efficiency. Only the acetyl group at the non-reducing end of the sugar was cleaved, with the sugar ring being held in the enzyme's binding cavity by a network of hydrogen bonds.

The crystallographic study afforded an analysis of the enzyme in complex with its substrate. This was possible probably due to an excess of the substrate while the enzyme was far from its optimal working conditions (the efficiency of this hyperthermophilic enzyme drops fast with temperature). A comparison of the enzyme-substrate complex with the unliganded enzyme indicated that the Zn^{2+} cation is tetrahedrally coordinated. Three of its ligands are provided by the protein's conserved His-Asp-His triad and the fourth, in an unliganded structure, is a single water molecule wedged between two catalytic residues, His259* and Asp42. The substrate approaches the Zn^{2+} ion in a monodentate manner, displacing the water molecule with the oxygen atom of its acetyl group. The structure of the protein does not change significantly upon the substrate binding. The N-acetyl group is positioned between the two active residues, approximately 3.5 Å from each. There is no evidence in the electron density of a water molecule that could act as the hydrolytic water and there is no place where a water molecule could be fitted

in the immediate vicinity of the active site. This does not mean that water could not penetrate the complex, but this crystal structure does not show how this could happen. One needs to allow for the active enzyme showing dynamics not seen in the crystal structure and hydrolysis as the first step cannot be ruled out, but this structure is consistent with the reaction mechanism that involves an anhydride intermediate.

Comparing structures and/or amino acid sequences of related proteins should make it possible to identify features responsible for thermostability. In the case of Dac-74 this was difficult due to the fact that the closely related proteins were also from hyperthermophiles, while the closest mesophilic deacetylases belonged to a different family of proteins (GH4, rather than GH14).

6.2 *Exo*- β -D-glucosaminidase - GlmA-01

6.2.1 Overall structure

Glucosaminidase crystals belong to the space group $C222_1$. There is one GlmA-01 subunit in the asymmetric unit and four dimeric assemblies in the unit cell, formed by the operation of crystallographic two-fold axes and centering. The dimers thus formed are analogous to the most similar known crystal structures of *exo*- β -D-glucosaminidases (Mine & Watanabe, 2019) (Mine et al. 2017). The GlmA-01 dimer has a central cavity that contains the ligand-binding and active sites that were identified in an analogous structure of GlmA from *T. kodakarensis* (PDB code 5GSM) (Mine & Watanabe, 2019). The cavity is accessible from the solvent by channels approximately 20 Å wide (Figure 34).

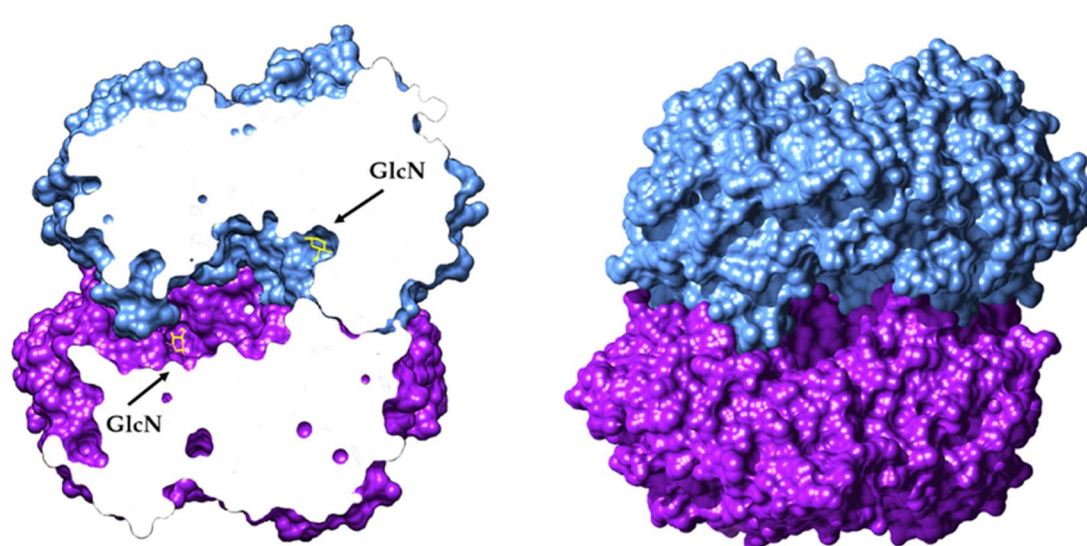


Figure 34 GlmA-01 dimer with the central cavity containing the ligand-binding and active sites. There is no ligand bound in the structure of GlmA-01 and the molecules of glucosamine (GlcN) that are visible on the figure have been modelled based on a superposition of GlmA-01 with the related liganded structure of GlmA from *T. kodakarensis* (Mine et al. 2017).

GlmA-01 subunit consists of 781 amino acid residues. Three domains can be distinguished in the subunit structure: TIM-barrel domain is made up of residues 1 to 434, α/β -domain consists of residues 435 to 647 and β -domain is made of residues 648 to 781 (Figures 35 and 36).

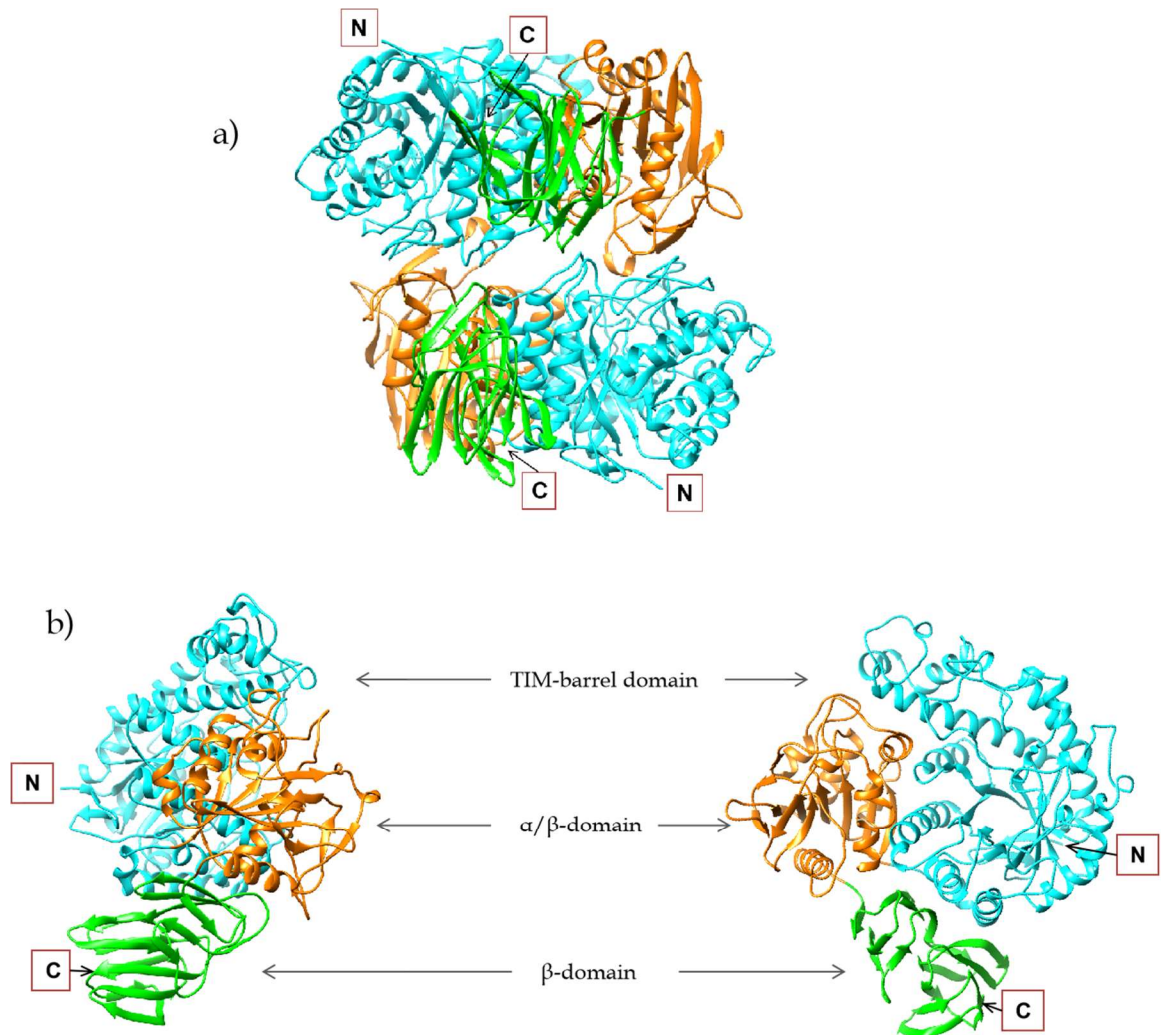


Figure 35 The biological dimer (a) and the three domains of GlmA-01 subunit: TIM-barrel domain (light blue), α/β -domain (orange) and β -domain (green)(b).

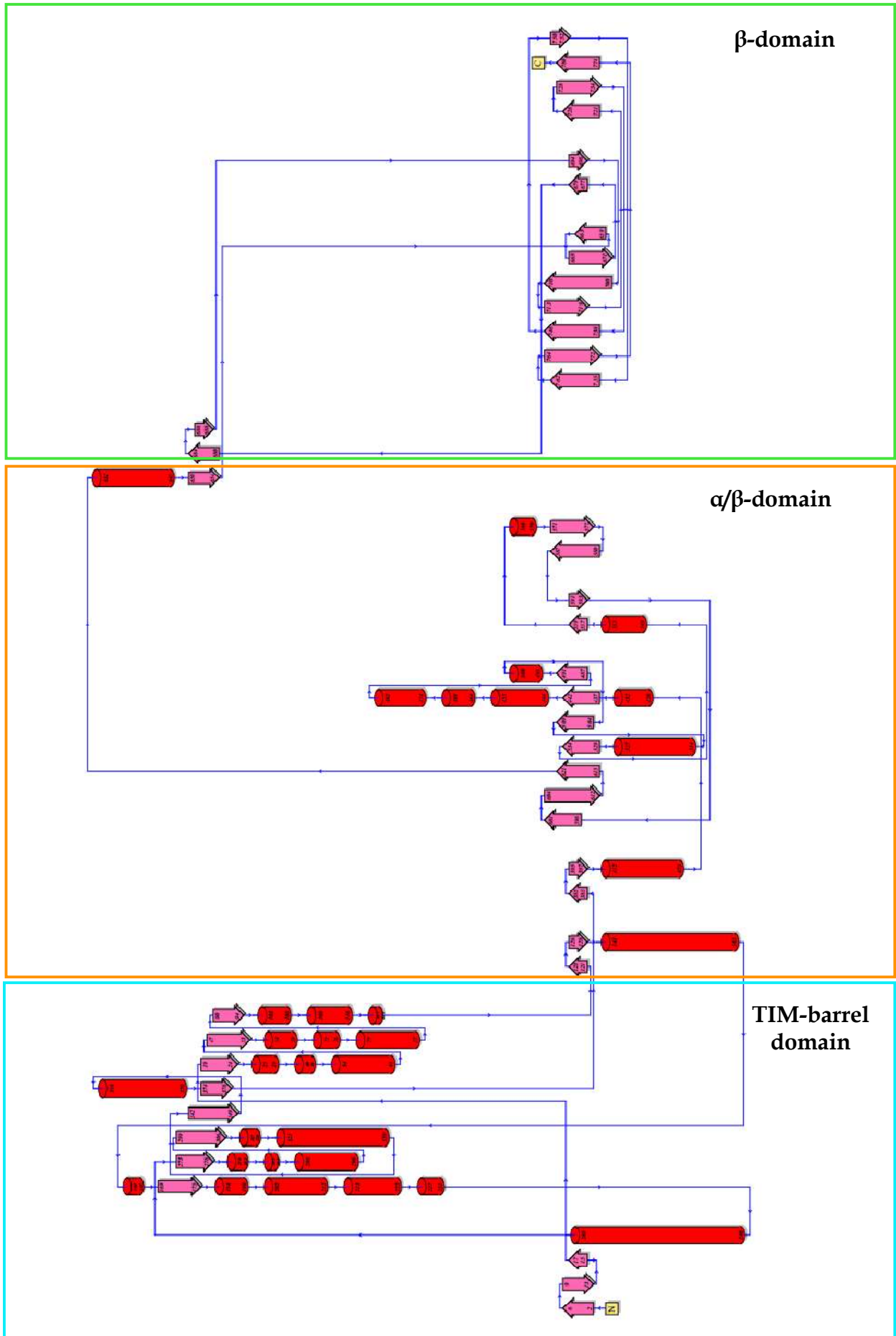


Figure 36 Topology diagram of the GlmA-01 subunit

A bioinformatics analysis was performed using the DALI server. Structural comparisons with homologous GlnA proteins revealed that the structure of GlnA-01 resembles three other known structures of the glycoside hydrolases (GH) 35 family of proteins, also derived from thermophilic organisms and exhibiting glucosaminidase activity. They are glucosaminidases from *T. kodakarensis* (PDB code 5GSM), *P. horikoshii* (5GSL) and *P. furiosus* (6JOW) (Mine and Watanabe 2019). Relatively low sequence similarity (15–29% identity) is also shared with the GH42 family of enzymes. They have a similar subunit topology to GlnA but form trimers and have a different enzymatic specificity (β -galactosidases).

6.2.2 Active site of GlnA-01

The active site of the enzyme has been defined in the structure analogous to GlnA-01, from *T. kodakaraensis*, in which the authors observed the reaction product - glucosamine (Mine & Watanabe, 2019) (PDB code 5GSM). The amino acid residues that interact with the product in the related structure correspond closely to the analogous residues in GlnA-01 (Figure 37). They are mainly aspartate and glutamate residues that create a dense network of hydrogen bonds with all the hydroxyl groups and the amine group of the ligand. The positions of the amino group and the O4 hydroxyl group are occupied in my structure of GlnA-01 by water molecules.

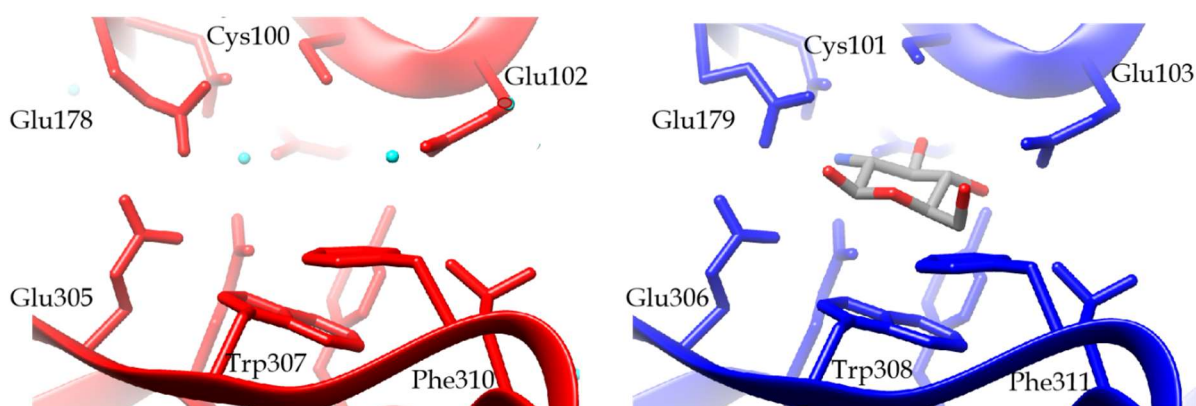


Figure 37 The residues responsible for binding the product. Red is GlnA-01, blue is GlnA from *T. kodakarensis* (PDB code 5GSM).

6.2.3 Enzymatic activity

When I was preparing the enzyme for the activity tests, I found that to observe any activity it was necessary to anneal the protein by incubating it at 60°C, for at least 5 min.

Preliminary ITC tests were performed on GImA-01 in the Laboratory of Protein Engineering by Dr. Joanna Śliwiak. Four potential substrates were tested: (GlcNAc)₂, (GlcN)₂, GlcN-GlcNAc and GlcNAc-GlcN. Some activity was detected only for GlcN-GlcNAc. Which is consistent with the postulated mechanism of the reaction described in the literature (Tanaka et al. 2004). No actual tests were performed previously on singly acetylated potential substrates (GlcN-GlcNAc or GlcNAc-GlcN). We were able to perform only preliminary tests before the compounds were withdrawn from the market. They are currently unavailable. I plan to produce them in the lab to finish the activity tests and find the optimal working conditions for the enzyme.

6.2.4 Differential Scanning Calorimetry (DSC)

I performed differential scanning calorimetry experiments with GImA-01. They indicated an extremely high melting point of the protein, one peak appears at 94°C, the other at 99°C (Figure 38). The first peak could correspond to deoligomerization of the protein and the second to unfolding of the subunits.

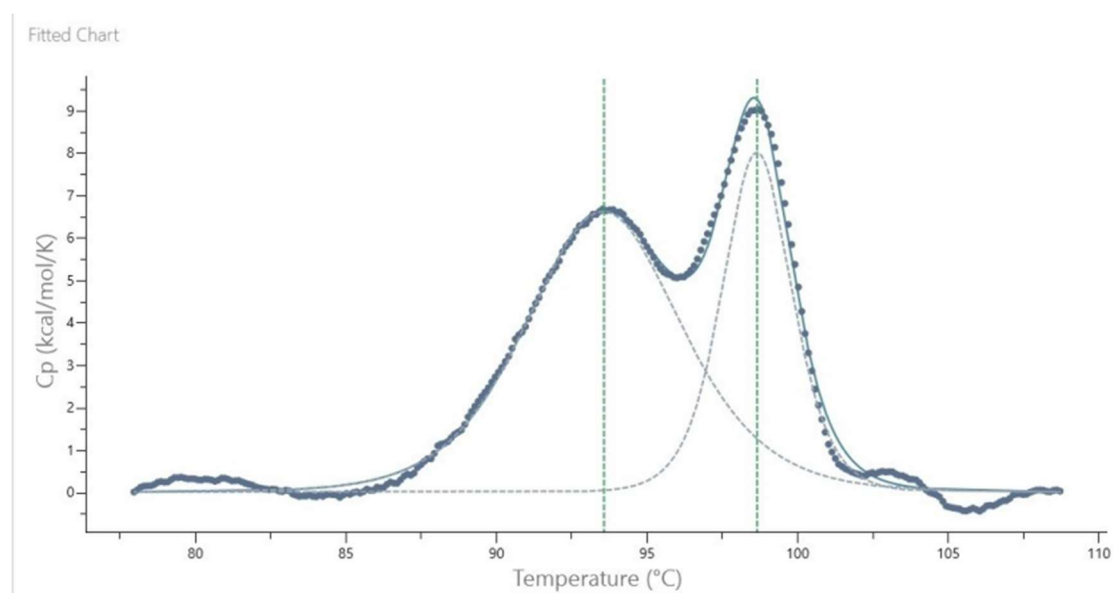


Figure 38 DSC profile for heating scans of GImA-01. The dotted lines represent an interpretation of the profile as two distinct peaks.

6.2.5 Conclusion for GlnA-01

The structure and the preliminary enzymatic tests place GlnA-01 in the GH35 family, rather than the related GH42 family (Mine et al., 2017). It is the fourth described structure of *exo*- β -D-glucosaminidase. GH35 enzymes were relatively recently classified as being distinct from GH42, and recognized as participating in chitin degradation. Also, all the known GlnA proteins, including GlnA-01 come from hyperthermophilic archaea.

The enzymatic tests performed to date indicate that GlnA-01 is specific to GlcN-GlcNAc as the substrate, which is the product of Dac-74. This means that GlnA-01 could work in tandem with Dac-74, as proposed for related enzymes by (Tanaka et al. 2004), in the sense that the product of Dac-74 is cleaved by GlnA-01 into a mixture of GlcN and GlcNAc, the latter of which can then be deacetylated by Dac-74, giving only GlcN as the final product.

7. Summary

This thesis presents structural and functional studies of enzymes involved in the chitin degradation pathway. Four proteins belonging to three different enzyme classes (chitinase, diacetylchitobiose deacetylase and *exo*- β -D-glucosaminidase) were selected for the study and purification protocols was established for them.

I obtained crystals for two of the enzymes (under several crystallization conditions) for which diffraction data were recorded. For diacetylchitobiose deacetylase, three data sets were recorded: for the unliganded protein, anomalous data at the zinc absorption wavelength, and for the enzyme-substrate complex. Thanks to additional research, I was able to confirm that the deacetylase belongs to the carbohydrate esterase (CE) family 14, as suggested by the literature, which includes zinc-dependent deacetylases. For *exo*- β -D-glucosaminidase, crystallographic data were obtained and the structure was solved of the unliganded enzyme.

In addition to structural research, which was the focus of my work, the scope of activities was extended through collaboration to complementary research by means of techniques such as SAXS, NMR, ITC, DSC.

The three enzymes that I investigated – chitinase, deacetylase and glucosaminidase – form a minimum set that should be able to degrade chitin all the way to glucosamine, provided that the latter two enzymes are not “too specific”. Otherwise intermediate products will accumulate. The deacetylase in this study removes the acetyl group only from the non-reducing end of (GlcNAc)₂ and (GlcNAc)₃, but it can also deacetylate the monomeric GlcNAc. This means that it is sufficiently unspecific to be part of the chitinolytic trio, defined above. The specificity of glucosaminidase still remains to be determined and will be subject to further studies, but its structure has been solved and early enzymatic studies indicate that it complements the activity of diacetylchitobiose deacetylase.

It is worth noting that both the hyperthermophilic enzymes, Dac-74 and GlmA-01, could be prepared at room temperature but required annealing before any enzymatic activity could be detected. The term “annealing” was defined in metallurgy and materials science, to describe heat treatment that alters physical or chemical properties of materials. It apparently applies also to at least some hyperthermophilic enzymes. The details of what actually takes place during the annealing process remain to be determined.

8. Streszczenie

W pracy przedstawiono badania strukturalne i funkcjonalne enzymów uczestniczących w szlaku degradacji chityny. Do badań wybrano cztery białka należące do trzech różnych klas enzymów (chitynaz, deacetylaz diacetylochitobiozy i *egzo*- β -D-glukozaminidaz) i opracowałam dla nich protokoły oczyszczania.

Dla dwóch enzymów uzyskałam kryształy (w różnych warunkach krystalizacyjnych), dla których zarejestrowano dane dyfrakcyjne. Dla deacetylazy diacetylochitobiozowej zarejestrowano trzy zestawy danych: dla białka bez ligandu, dane anomalne przy długości fali absorpcji cynku oraz dla kompleksu enzym-substrat. Dzięki dodatkowym badaniom udało mi się potwierdzić, że deacetylaza należy do rodziny 14 esteraz węglowodanowych (CE) obejmującej deacetylazy zależne od cynku, co sugeruje literatura. Dla *egzo*- β -D-glukozaminidazy uzyskano dane krystalograficzne i rozwiązano strukturę enzymu.

Oprócz badań strukturalnych, na których skupiała się moja praca, dzięki współpracy zakres działań został rozszerzony o badania komplementarne z wykorzystaniem takich technik jak SAXS, NMR, ITC, DSC.

Trzy badane przeze mnie enzymy - chitynaza, deacetylaza i glukozaminidaza - tworzą minimalny zestaw, wystarczający do tego żeby zdegradować chitynę do glukozaminy, pod warunkiem, że dwa ostatnie enzymy nie będą "zbyt specyficzne". W przeciwnym razie produkty pośrednie będą się gromadzić. Deacetylaza w tym badaniu usuwa grupę acetylową tylko z nieredukującego końca (GlcNAc)₂ i (GlcNAc)₃, ale może również deacetylować monomeryczny GlcNAc. Oznacza to, że jest wystarczająco niespecyficzna, aby wchodzić w skład triady chitynolitycznej, zdefiniowanej powyżej. Specyficzność glukozaminidazy wciąż nie została określona i będzie przedmiotem dalszych badań, ale struktura dla niej została

rozwiązana, a wczesne badania enzymatyczne wskazują, że uzupełnia ona aktywność deacetylazy diacetylochitobiozowej.

Warto zauważyć, że oba enzymy są hipertermofilne, Dac-74 i GImA-01, mogły być przygotowywane w temperaturze pokojowej, ale wymagały wyżarzania, zanim można było wykryć jakąkolwiek aktywność enzymatyczną. Termin "wyżarzanie" został zdefiniowany w metalurgii i materiałoznawstwie, aby opisać obróbkę cieplną, która zmienia fizyczne lub chemiczne właściwości materiałów. Najwyraźniej odnosi się on również do niektórych enzymów hipertermofilnych. Szczegóły tego, co faktycznie ma miejsce podczas procesu wyżarzania, pozostają do ustalenia.

9. References

- Afonine, Pavel v., Ralf W. Grosse-Kunstleve, Nathaniel Echols, Jeffrey J. Headd, Nigel W. Moriarty, Marat Mustyakimov, Thomas C. Terwilliger, Alexandre Urzhumtsev, Peter H. Zwart, and Paul D. Adams. 2012. "Towards Automated Crystallographic Structure Refinement with Phenix.Refine." *Urn:Issn:0907-4449* 68 (4): 352-67. <https://doi.org/10.1107/S0907444912001308>.
- Allen, Frank H., and W. D. Samuel Motherwell. 2002. "Applications of the Cambridge Structural Database in Organic Chemistry and Crystal Chemistry." *Urn:Issn:0108-7681* 58 (3): 407-22. <https://doi.org/10.1107/S0108768102004895>.
- Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman. 1997. "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs." *Nucleic Acids Research* 25 (17): 3389-3402. <https://doi.org/10.1093/NAR/25.17.3389>.
- Blair, David E., Omid Hekmat, Alexander W. Schüttelkopf, Binesh Shrestha, Ken Tokuyasu, Stephen G. Withers, and Daan M.F. van Aalten. 2006. "Structure and Mechanism of Chitin Deacetylase from the Fungal Pathogen *Colletotrichum Lindemuthianum*." *Biochemistry* 45 (31): 9416-26. <https://doi.org/10.1021/BI0606694/ASSET/IMAGES/MEDIUM/BI0606694N00001.GIF>.
- Bondos, Sarah E., and Alicia Bicknell. 2003. "Detection and Prevention of Protein Aggregation before, during, and after Purification." *Analytical Biochemistry* 316 (2): 223-31. [https://doi.org/10.1016/S0003-2697\(03\)00059-9](https://doi.org/10.1016/S0003-2697(03)00059-9).
- Brünger, Axel T. 1992. "Free R Value: A Novel Statistical Quantity for Assessing the Accuracy of Crystal Structures." *Nature* 1992 355:6359 355 (6359): 472-75. <https://doi.org/10.1038/355472a0>.
- Casanueva, Ana, Marla Tuffin, Craig Cary, and Don A. Cowan. 2010. "Molecular Adaptations to Psychrophily: The Impact of 'Omic' Technologies." *Trends in Microbiology* 18 (8): 374-81. <https://doi.org/10.1016/J.TIM.2010.05.002>.

- Cianci, Michele, Gleb Bourenkov, Guillaume Pompidor, Ivars Karpics, Johanna Kallio, Isabel Bento, Manfred Roessle, Florent Cipriani, Stefan Fiedler, and Thomas R. Schneider. 2017. "P13, the EMBL Macromolecular Crystallography Beamline at the Low-Emittance PETRA III Ring for High- and Low-Energy Phasing with Variable Beam Focusing." *Journal of Synchrotron Radiation* 24 (1): 323–32. <https://doi.org/10.1107/S1600577516016465/IG5037SUP2.AVI>.
- Emsley, P., B. Lohkamp, W. G. Scott, and K. Cowtan. 2010. "Features and Development of Coot." *Urn:Issn:0907-4449* 66 (4): 486–501. <https://doi.org/10.1107/S0907444910007493>.
- Feliciano, Gustavo Troiano, and Antônio José Roque da Silva. 2015. "Unravelling the Reaction Mechanism of Matrix Metalloproteinase 3 Using QM/MM Calculations." *Journal of Molecular Structure* 1091 (July): 125–32. <https://doi.org/10.1016/j.molstruc.2015.02.079>.
- Felse, P. A., and T. Panda. 2000. "Production of Microbial Chitinases – A Revisit." *Bioprocess Engineering* 23:2 23 (2): 127–34. <https://doi.org/10.1007/PL00009117>.
- Gooday, Graham W. n.d. "The Ecology of Chitin Degradation." Accessed September 26, 2022.
- Haney, P J, J H Badger, G L Buldak, C I Reich, C R Woese, and G J Olsen. 1999. "Thermal Adaptation Analyzed by Comparison of Protein Sequences from Mesophilic and Extremely Thermophilic Methanococcus Species." *Proceedings of the National Academy of Sciences* 96 (7): 3578–83. <https://doi.org/10.1073/pnas.96.7.3578>.
- Hartl, Lukas, Simone Zach, and Verena Seidl-Seiboth. 2012. "Fungal Chitinases: Diversity, Mechanistic Properties and Biotechnological Potential." *Applied Microbiology and Biotechnology* 93 (2): 533–43. <https://doi.org/10.1007/S00253-011-3723-3/FIGURES/3>.
- Holm, Liisa. 2020. "Using Dali for Protein Structure Comparison." *Methods in Molecular Biology* 2112: 29–42. https://doi.org/10.1007/978-1-0716-0270-6_3/COVER.

- Huber, R., Josef Stöhr, Sabine Hohenhaus, Reinhard Rachel, Siegfried Burggraf, Holger W. Jannasch, and Karl O. Stetter. 1995. "Thermococcus Chitonophagus Sp. Nov., a Novel, Chitin-Degrading, Hyperthermophilic Archaeum from a Deep-Sea Hydrothermal Vent Environment." *Archives of Microbiology*.
<https://doi.org/10.1007/s002030050262>.
- Kabsch, W., and IUCr. 2010. "XDS." *Urn:Issn:0907-4449* 66 (2): 125–32.
<https://doi.org/10.1107/S0907444909047337>.
- Kabsch, Wolfgang, and Christian Sander. 1983. "Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features." *Biopolymers* 22 (12): 2577–2637.
<https://doi.org/10.1002/BIP.360221211>.
- Kasprzewska, A. 2003. "Plant Chitinases - Regulation and Function." *Cellular and Molecular Biology Letters* 08 (3).
- Kezuka, Yuichiro, Manabu Ohishi, Yoshikane Itoh, Jun Watanabe, Masaru Mitsutomi, Takeshi Watanabe, and Takamasa Nonaka. 2006. "Structural Studies of a Two-Domain Chitinase from *Streptomyces Griseus* HUT6037." *Journal of Molecular Biology* 358 (2): 472–84.
<https://doi.org/10.1016/J.JMB.2006.02.013>.
- Kleywegt, Gerard J., and T. Alwyn Jones. 1997. "[11] Model Building and Refinement Practice." *Methods in Enzymology* 277 (January): 208–30.
[https://doi.org/10.1016/S0076-6879\(97\)77013-7](https://doi.org/10.1016/S0076-6879(97)77013-7).
- Krissinel, E., and K. Henrick. 2004. "Secondary-Structure Matching (SSM), a New Tool for Fast Protein Structure Alignment in Three Dimensions." *Acta Crystallographica Section D-Biological Crystallography* 60 (12): 2256–68.
<https://doi.org/10.1107/S0907444904026460>.
- Kumwenda, Benjamin, Derek Litthauer, Özlem Tastan Bishop, and Oleg Reva. 2013. "Analysis of Protein Thermostability Enhancing Factors in Industrially Important *Thermus* Bacteria Species Submit a Paper." *Evolutionary Bioinformatics* 2013 (9): 327–42. <https://doi.org/10.4137/EBO.S12539>.
- Kurita, Keisuke. n.d. "Controlled Functionalization of the Polysaccharide Chitin." www.elsevier.com/locate/ppolysci.

- Lamzin, V. S., and K. S. Wilson. 1993. "Automated Refinement of Protein Models." *Acta Crystallographica. Section D, Biological Crystallography* 49 (Pt 1): 129–47. <https://doi.org/10.1107/S0907444992008886>.
- Larsen, Tanja, Bent O. Petersen, Birgit G. Storgaard, Jens O. Duus, Monica M. Palcic, and Jørgen J. Leisner. 2011. "Characterization of a Novel Salmonella Typhimurium Chitinase Which Hydrolyzes Chitin, Chitooligosaccharides and an N-Acetylactosamine Conjugate." *Glycobiology* 21 (4): 426–36. <https://doi.org/10.1093/GLYCOB/CWQ174>.
- Laskowski, Roman A. 2009. "PDBsum New Things." *Nucleic Acids Research* 37 (Database issue). <https://doi.org/10.1093/NAR/GKN860>.
- McCoy, Airlie J., Ralf W. Grosse-Kunstleve, Paul D. Adams, Martyn D. Winn, Laurent C. Storoni, and Randy J. Read. 2007. "Phaser Crystallographic Software." *Urn:Issn:0021-8898* 40 (4): 658–74. <https://doi.org/10.1107/S0021889807021206>.
- Mine, Shouhei, Takahisa Ikegami, Kazunori Kawasaki, Tsutomu Nakamura, and Koichi Uegaki. 2012. "Expression, Refolding, and Purification of Active Diacetylchitobiose Deacetylase from *Pyrococcus Horikoshii*." *Protein Expression and Purification* 84 (2): 265–69. <https://doi.org/10.1016/J.PEP.2012.06.002>.
- Mine, Shouhei, and Masahiro Watanabe. 2019a. "Structural Insights into the Molecular Evolution of the Archaeal Exo- β -d-Glucosaminidase." *International Journal of Molecular Sciences*. MDPI AG. <https://doi.org/10.3390/ijms20102460>.
- Mine, Shouhei, Masahiro Watanabe, Saori Kamachi, Yoshito Abe, and Tadashi Ueda. 2017. "The Structure of an Archaeal β -Glucosaminidase Provides Insight into Glycoside Hydrolase Evolution." *The Journal of Biological Chemistry* 292 (12): 4996–5006. <https://doi.org/10.1074/JBC.M116.766535>.
- Mirdita, Milot, Sergey Ovchinnikov, and Martin Steinegger. 2021. "ColabFold - Making Protein Folding Accessible to All." *BioRxiv*, August, 2021.08.15.456425. <https://doi.org/10.1101/2021.08.15.456425>.

- Morimoto, Kenji, Shuichi Karita, Tetsuya Kimura, Kazuo Sakka, and Kunio Ohmiya. 1997. "Cloning, Sequencing, and Expression of the Gene Encoding *Clostridium Paraputrificum* Chitinase ChiB and Analysis of the Functions of Novel Cadherin-like Domains and a Chitin-Binding Domain." *Journal of Bacteriology* 179 (23): 7306–14. <https://doi.org/10.1128/JB.179.23.7306-7314.1997>.
- Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn, and H. Erlich. 1986. "Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction." *Cold Spring Harbor Symposia on Quantitative Biology* 51 (1): 263–73. <https://doi.org/10.1101/SQB.1986.051.01.032>.
- Murshudov, Garib N., Pavol Skubák, Andrey A. Lebedev, Navraj S. Pannu, Roberto A. Steiner, Robert A. Nicholls, Martyn D. Winn, Fei Long, and Alexei A. Vagin. 2011. "REFMAC5 for the Refinement of Macromolecular Crystal Structures." *Urn:Issn:0907-4449* 67 (4): 355–67. <https://doi.org/10.1107/S0907444911001314>.
- Nakamura, Akihiko, Kei ichi Okazaki, Tadaomi Furuta, Minoru Sakurai, and Ryota Iino. 2018. "Processive Chitinase Is Brownian Monorail Operated by Fast Catalysis after Peeling Rail from Crystalline Chitin." *Nature Communications* 9 (1). <https://doi.org/10.1038/s41467-018-06362-3>.
- Nakamura, Tsutomu, Mayumi Niiyama, Wakana Hashimoto, Kurumi Ida, Manabu Abe, Junji Morita, and Koichi Uegaki. 2015. "Multiple Crystal Forms of N,N'-Diacetylchitobiose Deacetylase from *Pyrococcus Furiosus*." *Acta Crystallographica Section F: Structural Biology Communications* 71 (June): 657–62. <https://doi.org/10.1107/S2053230X15005695>.
- Papadimitriou, Konstantinos, Panagiotis K. Baharidis, Anastasios Georgoulis, Marion Engel, Maria Louka, Georgia Karamolegkou, Aggeliki Tsoka, et al. 2016. "Analysis of the Complete Genome Sequence of the Archaeon *Pyrococcus Chitonophagus* DSM 10152 (Formerly *Thermococcus Chitonophagus*)." *Extremophiles* 20 (3): 351–61. <https://doi.org/10.1007/s00792-016-0826-x>.

- Pettersen, Eric F., Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. 2004. "UCSF Chimera – A Visualization System for Exploratory Research and Analysis." *Journal of Computational Chemistry* 25 (13): 1605–12. <https://doi.org/10.1002/JCC.20084>.
- Pyrpassopoulos, Serapion, Metaxia Vlassi, Achilleas Tsortos, Yannis Papanikolaou, Kyriacos Petratos, Constantinos E Vorgias, and George Nounesis. 2006. "Equilibrium Heat-Induced Denaturation of Chitinase 40 from *Streptomyces Thermoviolaceus*." <https://doi.org/10.1002/prot.21003>.
- Ramakrishnan, C., and G. N. Ramachandran. 1965. "Stereochemical Criteria for Polypeptide and Protein Chain Conformations: II. Allowed Conformations for a Pair of Peptide Units." *Biophysical Journal* 5 (6): 909–33. [https://doi.org/10.1016/S0006-3495\(65\)86759-5](https://doi.org/10.1016/S0006-3495(65)86759-5).
- Sahai, A. S., and M. S. Manocha. 1993. "Chitinases of Fungi and Plants: Their Involvement in Morphogenesis and Host – Parasite Interaction." *FEMS Microbiology Reviews* 11 (4): 317–38. <https://doi.org/10.1111/J.1574-6976.1993.TB00004.X>.
- Sambrook, Joseph, and David W. Russel. 2001. "Screening Expression Libraries." *Molecular Cloning: A Laboratory Manual*, 14.1-14.51. https://books.google.pl/books?id=BzTgvgEACAAJ&hl=pl&source=gbs_book_other_versions.
- Satow, Yoshinori, Gerson H. Cohen, Eduardo A. Padlan, and David R. Davies. 1986. "Phosphocholine Binding Immunoglobulin Fab McPC603: An X-Ray Diffraction Study at 2.7 Å." *Journal of Molecular Biology* 190 (4): 593–604. [https://doi.org/10.1016/0022-2836\(86\)90245-7](https://doi.org/10.1016/0022-2836(86)90245-7).
- Schein, Catherine H., and Mathieu H.M. Noteborn. 1988. "Formation of Soluble Recombinant Proteins in *Escherichia Coli* Is Favored by Lower Growth Temperature." *Bio/Technology* 1988 6:3 6 (3): 291–94. <https://doi.org/10.1038/nbt0388-291>.
- Shuman, Stewart. 1994. "THE JOURNAL OF BIOLOGICAL CHEMISTRY Novel Approach to Molecular Cloning and Polynucleotide Synthesis Using *Vaccinia*

- DNA Topoisomerase*." *Journal of Biological Chemistry* 269 (51): 32678–84.
[https://doi.org/10.1016/S0021-9258\(18\)31688-0](https://doi.org/10.1016/S0021-9258(18)31688-0).
- Thakur, Deepali, Anjali Chauhan, Prakriti Jhila, Rajesh Kaushal, and Bhawna Dipta. 2022. "Microbial Chitinases and Their Relevance in Various Industries." *Folia Microbiologica*. Springer Science and Business Media B.V.
<https://doi.org/10.1007/s12223-022-00999-w>.
- Tickle, I. J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C. & Bricogne, G. (2018). STARANISO. Global Phasing Ltd, Cambridge, UK. STARANISO Anisotropy & Bayesian Estimation Server." n.d. Accessed September 28, 2022. <https://staraniso.globalphasing.org/cgi-bin/staraniso.cgi>.
- Vogt, Gerhard, Stefanie Woell, and Patrick Argos. 1997. "Protein Thermal Stability, Hydrogen Bonds, and Ion Pairs." *Journal of Molecular Biology* 269 (4): 631–43. <https://doi.org/10.1006/jmbi.1997.1042>.
- Wlodawer, Alexander, Wladek Minor, Zbigniew Dauter, and Mariusz Jaskolski. 2008. "Protein Crystallography for Non-Crystallographers, or How to Get the Best (but Not More) from Published Macromolecular Structures." *The FEBS Journal* 275 (1): 1. <https://doi.org/10.1111/J.1742-4658.2007.06178.X>.
- Yan, Qiang, and Stephen S. Fong. 2015. "Bacterial Chitinase: Nature and Perspectives for Sustainable Bioproduction." *Bioresources and Bioprocessing* 2 (1): 1–9. <https://doi.org/10.1186/S40643-015-0057-5/FIGURES/3>.
- Yip, KSP S.P., TJ J. Stillman, KL L. Britton, PJ J. Artymiuk, PJ J. Baker, SE E. Sedelnikova, PC C. Engel, et al. 1995. "The Structure of *Pyrococcus Furiosus* Glutamate Dehydrogenase Reveals a Key Role for Ion-Pair Networks in Maintaining Enzyme Stability at Extreme Temperatures." *Structure* 3 (11): 1147–58. [https://doi.org/10.1016/S0969-2126\(01\)00251-9](https://doi.org/10.1016/S0969-2126(01)00251-9).