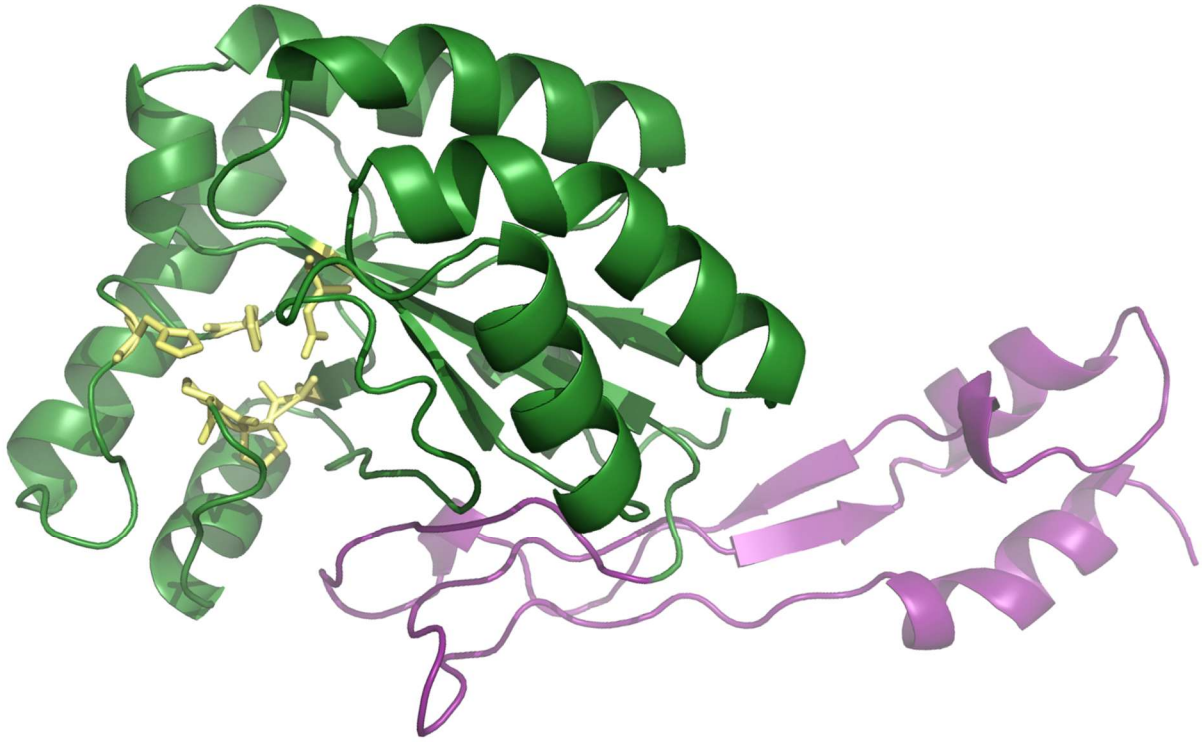
Two deaths and one Resurrection: Tales of Polysaccharide de-acetylases.



Ονοματεπώνυμο	Αλέξιος Μολφέτας-Λυγκιάρης
Επιστημονικός υπεύθυνος	Μιχαήλ Κοκκινίδης
Πτυχίο	Μοριακή Βιολογία- Βιοιατρική
Αριθμός Λέξεων	8185

UNIVERSITY OF CRETE

Abstract

The project aims to elucidate the enzymatic mechanisms of polysaccharide deacetylases(PDA) from pathogenic bacteria, and to investigate the role(s) of their inactive homologues. The genomes of *Bacillus sp.* and especially *B.cereus sensu lato* including *B.anthracis* contain ten putative polysaccharide deacetylase (PDA) genes with high sequence homologies. Crystallographic studies complemented by mass-spectrometry and mutagenesis, reveal a catalytic domain exhibiting a NodB fold, and imply an autocatalytic process which modifies with high specificity a conserved Proline. This targeted hydroxylation of Pro functions as a maturation step for the active site of the de-N-acetylation reaction. One of the PDA's namely *BA3943* lacks a key aspartic acid of the catalytic and metal binding site. Mass spectrometry and enzymatic activity experiments, documented the absence of both Pro-hydroxylation and de-N-acetylation mechanisms. During my master thesis, the X-ray crystallographic structure of *Ba3943* was determined (1.5Å). Three-dimensional superposition with an active homologue revealed significant differences in the catalytic centre. Mutation of three key residues, restored both hydroxylation and de-N-acetylation activities. A computational search aiming to determine the conservation of active site geometry of PDAs, for structures deposited with the Protein Data Bank (PDB), returned ≃500 hits. Thorough data-mining, might reveal common structural basis for the pro-hydroxylation mechanism. Additionally structure solution revealed a conserved intrinsically disordered hinge domain at the N-terminus, rich with lysines, suggesting an important function inside the cell.

Περίληψη

Η παρακάτω εργασία έχει ως στόχο, να χαρακτηρίσει τον ενζυματικό μηχανισμό των απακετυλασών πολυσακχαριτών, από παθογενή βακτήρια, και των ανενεργών ομολόγων της. Στο γονιδίωμα των Bacillus, και ιδιαίτερα στα είδη B.anthracis και B.cereus, υπάρχουν δέκα απακετυλάσες πολυσακχαριτών. Κρυσταλλογραφικές μελέτες από την ομάδας μας, αποκάλυψαν μία καταλυτική (NodB) περιοχή, όπου ένας αυτοκαταλυτικός μηχανισμός υδροξυλίωσης, τροποποιεί, τον α-άνθρακα μίας διατηρημένης προλίνης. Αυτή η ασυνήθιστη μέτα-μεταφραστική τροποποίηση, λειτουργεί ως ενδιάμεσο βήμα, επάγοντας τις απαραίτητες μετατοπίσεις στο ενεργό κέντρο, έτσι ώστε να επιτευχθεί η απακετυλιωτική δράση των ενζύμων. Η νέα αυτή υδροξυλίωση μπορεί να αποτελεί την βάση ενός καθολικού ενζυμικού μηχανισμού που δεν έχει ταυτοποιηθεί ακόμα, καθώς και να συνδέεται με διάφορες μορφές βακτηριακής παθογένειας. Ένα από τα εκφραζόμενα ένζυμα αυτής της οικογένειας, η Ba3943, έχει χάσει κάποια από τα βασικά κατάλοιπα που συμμετέχουν στην κατάλυση και την δέσμευση του μετάλλου στο ενεργό κέντρο. Ενζυμολογικά πειράματα επιβεβαίωσαν την ανενεργότητα του ενζύμου, και ταυτόχρονα αποδείχθηκε και η απουσία του υδροξυλίου στον α-άνθρακα της προλίνης με φασματομετρία μάζας. Κατά τη διάρκεια του μεταπτυχιακού μου, προσδιόρισα την κρυσταλλική δομή της *BA3943* στα 1.5 Å. Τρισ-διάσταση υπέρθεση με ένα ενεργό ένζυμο της ίδιας οικογένειας (*BC1960*), αποκάλυψε ουσιαστικές διαφορές στην γεωμετρία του καταλυτικού κέντρου, που οδήγησαν στον σχεδιασμό στοχευμένων μεταλλάξεων. Μια τριπλή μετάλλαξη του ενζύμου αυτού, τελικά αποκατέστησε την υδροξυλίωση του α-ανθρακα, και ως αποτέλεσμα την απακετυλιωτική του δράση. Η συγκεκριμένη δομή του *BA3943* αποκάλυψε μια αναδίπλωση στο αμινοτελικό άκρο που δεν έχει ξαναχαρακτηριστεί, που περιέχει μεγαλό αριθμό λυσινών που παίζουν συγκεκριμένο ρόλο, στο κύτταρο.

Acknowledgements

This project was completed under the supervision of Professor Kokkinidis, who provided valuable guidance throughout the project. A lot of work has to be attributed to Professor Bouriotis lab, and the work Natasa Tomatsidou has done, on the enzymatic activity of the studied protein. All MASSPEC analysis was performed by Dr. Aivaliotis group, and the script development was done under the supervision of George Eustathiou. Additionally, I have to thank Professor Petratos and Professor Bouriotis, for consulting me, when faced with difficulties.

Furthermore I thank the entire group, and in particular Dina Kotsifaki for withstanding my presence inside the lab, Vicky Fadouloglou for her continual guidance and and Stratos Mylonas for his help with software and hardware issues, that came up during my thesis. In conclusion, I thank the football club “Tiganitis”, and the radio station of the University of Crete.

Contents

INTRODUCTION

Significance of the project

Bacillus anthracis, is a Gram+ (positive) rod-shaped bacterium. It has the ability to sporulate, and survive even in the most adverse environmental conditions. When these dormant spores find the suitable environment, they re-establish vegetative growth, and often this occurs inside a mammalian host. Ingestion or inhalation of *B. anthracis* spores can lead to the popular anthrax disease, which can be a terminal disease, as it spreads very quickly and releases three deadly toxins [Mock and Fouet, 2001]. These toxins termed protective antigen, oedema factor and lethal factor, act in combination to weaken the host and allow a lethal infection to be established rapidly, and lead to death in a matter of days [Mock and Mignot, 2003]. This makes anthrax one of the most lethal pathogens, as it is difficult to diagnose and provide treatment, once it has invaded into a host. Because of its deadly properties it is a prominent bioterrorism agent that can cause massive mortality through the intentional release of aerosolized spores [Inglesby et al., 2002].

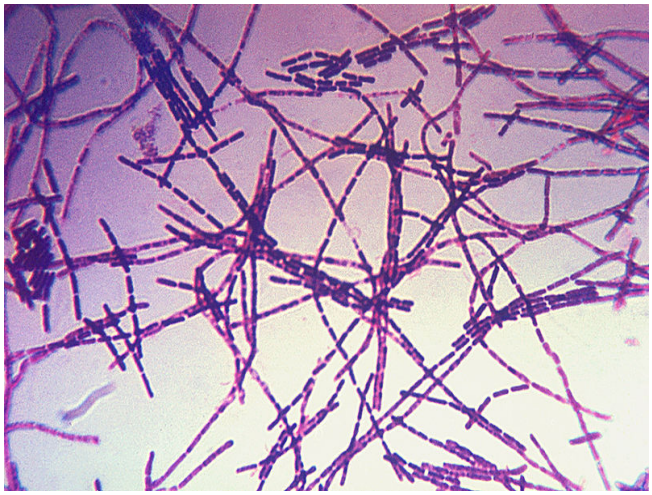


Figure 1. A photomicrograph of Bacillus anthracis bacteria using Gram-stain technique. [CDC, 2009]

Research to battle this deadly disease has been ongoing since the last decades of the nineteenth century, when R.Koch and L. Pasteur identified the causing bacterial pathogen of Anthrax disease. However, most of this research has focused on inhibiting the activity of the three toxins that infiltrate into the host's organism. Since the genome of *B. anthracis* was sequenced, research regarding other aspects of its biology has begun, and alternative ways to battle this disease are endeavoured. One of many ways to inhibit the bacteria from accumulating inside the host and unleashing its devastating toxins, is by blocking its defence mechanisms against the initial immune response. *B. anthracis* modifies the peptidoglycan layer that surrounds it, and this way it becomes invisible to the host's immune system. During the last few years, our group has commenced an attempt, to reveal the complete enzymatic mechanism that achieves such a modification on the peptidoglycan membrane, find ways to block it, and understand the mechanisms *B. anthracis* uses to invade and multiply.

Scientific Background

Gram-negative bacteria possess two phospholipid membranes and a peptidoglycan layer in between the two. Gram-positive bacteria on the other side lack an outer membrane; instead they use a peptidoglycan layer to withstand osmotic pressure and retain their cell shape and physiology. Its composition includes repeating units of *N*-acetyl-glucosamine and *N*-acetyl-muramic acid held together by β(1—4) glycosidic bonds (see Figure 2 below).

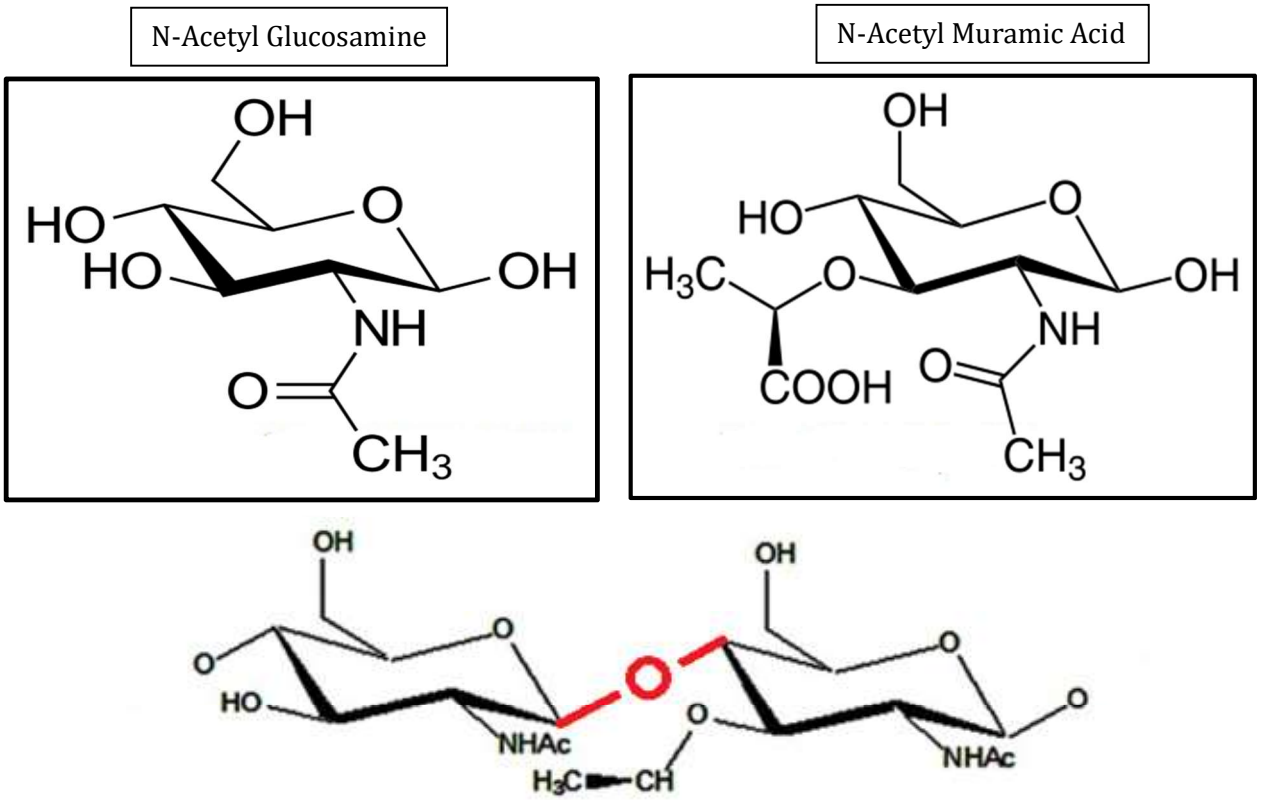


Figure 2. The main components of the peptidoglycan layer, and the glyosidic bond they form.

N-acetyl muramic acid, harbors a chain of 5 amino acids, composed of alanines and glutamic acids, and a variable amino acid at the third position of this chain [Fouet, 2009]. A bridge is usually created between the 3rd and 4th amino acid of the next stem peptide [Fouet, 2009].

In order to elicit an immune response, the host organism must first identify specific recognition patterns in the peptidoglycan or lipoprotein outer layer. Responsible for this recognition, are peptidoglycan recognition proteins, C-type lectins, Toll and Nod like receptors. All these are termed Pattern Recognition Receptors, although each one has specialized in different steps of the recognition process. One family of enzymes, key to the stimulation of the immune response in animals, are lysozymes. Lysozyme is a positively charged macromolecule, which binds to the negatively charged peptidoglycan layer, and

hydrolyzes the glycosidic bond formed between the constituents of this membrane [Ibrahim et al., 2001]. As a result, the glycosidic bond is cleaved, and this results into fragmentation of the peptidoglycan layer, which exposes the components of peptidoglycan to be easily recognized by Pattern Recognition Receptors [Boneca, 2005]. Muramyl di-peptides which are products of this hydrolysis are recognized by the Nucleotide-binding Oligomerization Domain-containing protein 2 (Nod2) protein, which initiates a signal cascade, involving Nf-κB and several kinases that constitute the immune response [Philpott et al., 2014].

Bacteria have developed several strategies to evade recognition by the immune system. Gram- negative bacteria possess an outer membrane to protect the peptidoglycan layer, from lysozyme binding [Callewaert al., 2010]. On the other side, Gram-positive bacteria have evolved different mechanisms of protection against the host's immune system. One of the most sophisticated, involves the modification of the peptidoglycan's layer charge, by adding D-alanine residues in teichoic acids, thus neutralizing the negative charge, and not allowing lysozyme to bind there [Peschel et al., 1999]. However, the predominant mechanism for avoiding detection, employed by all bacteria is molecular camouflage. Slight modifications in the outer layer that render the bacteria practically invisible to lysozymes. Three of these modifications have been identified and characterized. These include 1) *N*-glycolylation of *N*-acetyl-muramic acid, 2) *O*-acetylation of *N*-acetylmuramic acid and 3) *N*-deacetylation of *N*-acetyl-glucosamine.

"Lysozyme binds peptidoglycan by H-bonding between amino acid residues that line the six binding subsites of the enzyme and both the *N*-acetyl and the C-6 hydroxyl groups of three successive GlcNAc and MurNAc residues, respectively" [Blake et al., 1965]. O-acetylation of the 6^{th} carbon, causes steric hindrance, and prevents lysozyme from binding there. De-N-acetylation, which is the most commonly reported modification, happens on β-carbon of *N*-acetyl-glucosamine residue. The cationic nature of these residues impedes lysozyme binding, and thus fragmentation of the glycosidic bond. The de-N-acetylation modification happens post-translationally, by a specific family of enzymes termed polysaccharide de-acetylases. Studies on this family of enzymes, and their enzymatic activity, might reveal possible ways to inhibit the modification, and battle the disease, before it can even establish itself and replicate.

Polysaccharide de-acetylases

Polysaccharide de-acetylases (PDA's) are metalloenzymes, members of the carbohydrate esterase family 4 (CE4), which includes "chitin deacetylases, acetylxylan esterases, xylanases, chitooligosacharide de-acetylases and peptidoglycan deacetylases" [Balomenou et al., 2015]. The genomes of *Bacillus cereus* and *anthracis* contain 10 putative PDA genes [Figure 4]. Proteins of this family share a conserved NodB domain, where the active site lies. This "domain adopts a deformed (beta/alpha) barrel fold, comprising eight parallel beta-strands, with the C-terminal ends of five of these strands forming the solvent-exposed active site region, surrounded by eight alpha- helices"(figure 3) [Interpro]. Except of their prominent role in providing lysozyme resistance, peptidoglycan deacetylases have been found to participate in several cellular processes, such as cell shape maintenance, membrane osmotic ability, cell division/elongation and formation of biofilms. [Balomenou et al., 2015]
The wide range of cellular processes they fulfil make them an interesting object to study, as there are multiple ways to inhibit the accumulation of the bacteria, by targeting one particular family of enzymes. In addition, it is a wonderful example of how an enzyme family has diverged to shape proteins that perform other functions inside the cell, other than catalysis.

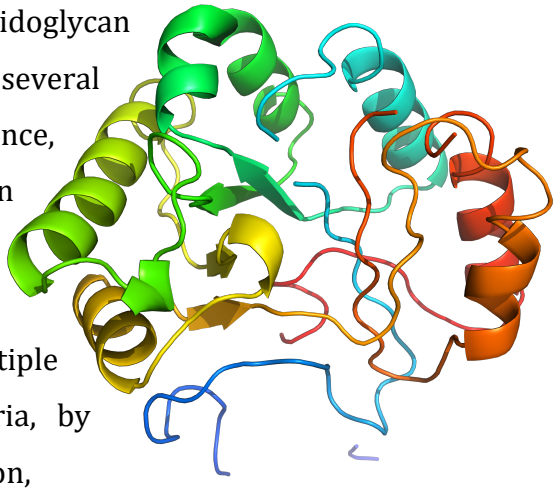


Figure 3. A typical NodB fold.

The main enzymatic function of PDA's is to remove acetyl groups, from polysaccharide substrates [Balomenou et al., 2015]. Products of such a reaction are free amine groups, and the modified polysaccharide. With such a modifications PDA's are able to regulate the localized ionic charge, and serve as a signalling component of the cell [Balomenou et al., 2015]. Such a property makes them, irreplaceable component of the cellular machinery.

B. cereus ATCC 14579	*B. anthracis* st. Ames	Possible function	Identity	Similarity
NP_831730 (275) (BC1960)	NP_844369 (275) (BA1961)	Peptidoglycan GlcNAc deacetylase	94	97
NP_833348 (213) (BC3618)	NP_845942 (213) (BA3679)	Peptidoglycan GlcNAc deacetylase	97	100
NP_832677 (275) (BC2929)	NP_845280 (275) (BA2944)	Peptidoglycan GlcNAc deacetylase	94	97
NP_834868 (245) (BC5204)	NP_847604 (245) (BA5436)	Peptidoglycan GlcNAc deacetylase	93	96
NP_831744 (273) (BC1974)	NP_844383 (273) (BA1977)	Peptidoglycan GlcNAc deacetylase	98	99
NP_830306 (260) (BC0467)	NP_842967 (273) (BA0424)	Peptidoglycan MurNAc deacetylase	98	99
NP_830050 (254) (BC0171)	NP_842717 (254) (BA0150)	Chitooligosaccharide deacetylase	95	99
NP_831543 (234) (BC1768)	NP_844255 (234) (BA1836)	Chitooligosaccharide deacetylase	92	96
NP_833526 (299) (BC3804)	NP_846187 (299) (BA3943)	Chitooligosaccharide deacetylase	95	97
NP_830200 (360) (BC0361)	NP_842877 (360) (BA0330)	PDA	91	94
NP_830200 (360) (BC0361)	NP_842878 (367) (BA0331)	PDA	53	69

Figure 4. Ten putative PDA's in Bacillus cereus and anthracis families.

Consistent with such an assumption are findings suggesting its involvement with cell growth and daughter cell separation [Balomenou et al., 2015]. Differences in the ionic micro-environment influence the function of other protein components of the cell wall. Therefore de-acetylases act as master-switches that control the function of the peptidoglycan layer.

Their involvement with biofilm formation makes them all more important. The formation of biofilm is a key mechanism with which the bacteria can flourish. There have been reports of direct implication of this family of enzymes with the proper formation of biofilms, whose main component are exopolysaccharides (Vlamakis et al., 2013).

Furthermore, *Bacillus Anthracis*, has a unique cell wall neutral polysaccharide that is covalently bonded to peptidoglycan. Two PDA's have been involved with peptidoglycan biogenesis, and other two with the attachment of the neutral polysaccharide to the peptidoglycan [Balomenou et al., 2015]. Finally two more members of this family, BA0330 and BA0331, have been implicated with the ability to grow in the presence of a high salt concentration with the maintenance of a uniform cell shape [Arnaouteli et al., 2015].

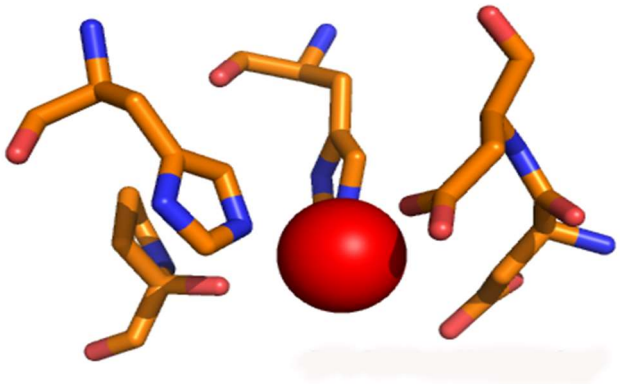


Figure #. Metal Co-ordination by a His-His-Asp triad.

The de-acetylation reaction is metal-dependent, which means that it requires a Cobalt (2+) ion in order for the reaction to go about. Yet, due to the limited bioavailability of this cation in nature other divalent metals have also been found to bind to the catalytic center like Zinc (2+), Iron (II) or Nickel (2+). This metal is co-ordinated by a highly conserved metal-binding triad comprising of 2 Histidines (H) and one Aspartic acid (D) (H-D-D).

The de-acetylation reaction performed by PDA's, follows a general acid-base mechanism. Some well-conserved motifs have been identified and proposed to participate in the catalytic process (See figure 5). The first motif (TFDD) contains two aspartic acids; one which acts as a base to activate H_2O, and the other to coordinate the metal. The second motif (H(S/T)xxH) contributes two histidines required for metal coordination as well. The third motif contains a well conserved arginine, which may act to co-ordinate the catalytic aspartic acid, and a

strictly conserved proline. Two more motifs not shown in the image below, seem to occur, where they contribute a histidine residue which probably acts as the catalytic acid, and it is co-ordinated by an aspartic acid in motif four.

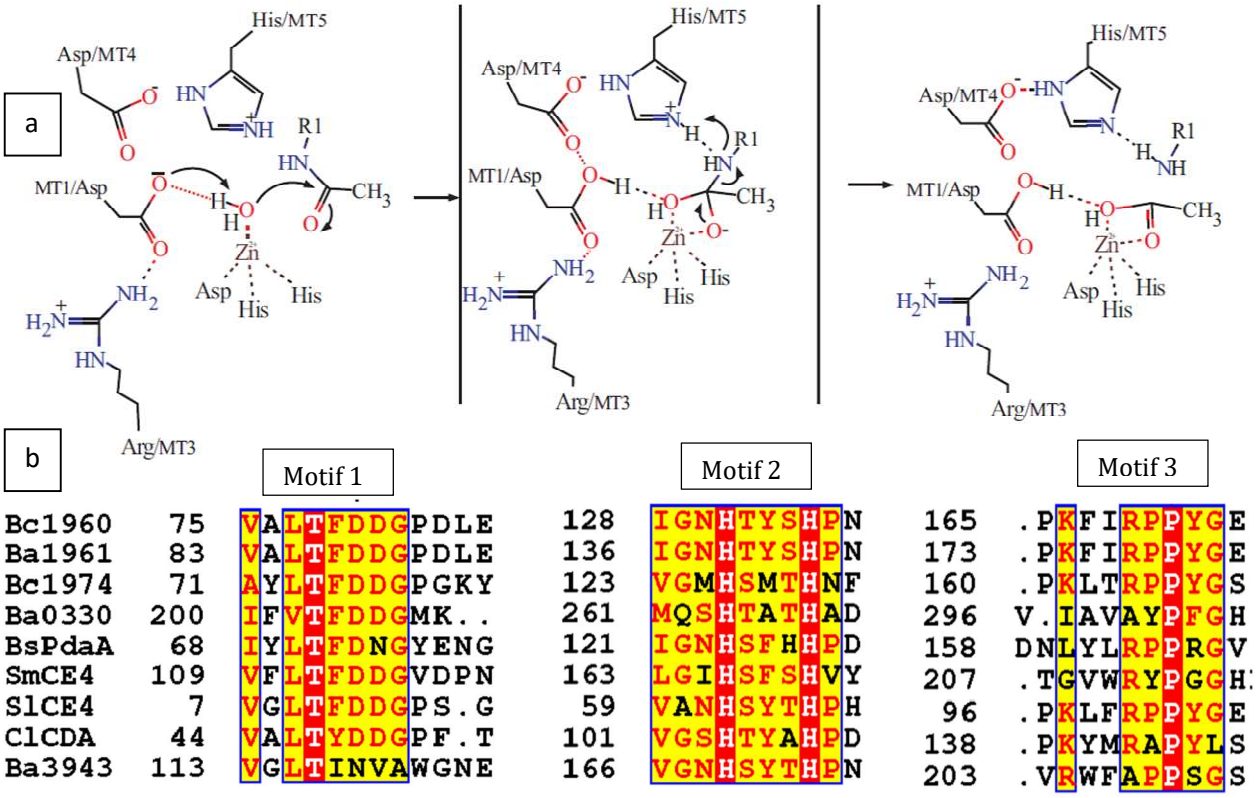


Figure 5. a) Proposed enzymatic mechanism [Balomenou et al., 2015]. Sequence alignment of three well-conserved motifs that form the catalytic centre in PDA's.

2-Hyp (Hydroxyproline)

Proline hydroxylation is one of the most prominent post-translational modifications eukaryotes. It is found in abundance in the body of mammals, as it forms one of the main components of collagen. Collagen is the main constituent of the connective tissue, so it accounts for 30% of the whole protein content in an animal. This post-translational modification takes place in the endoplasmatic reticulum, and it involves the enzyme prolyl-4-hydroxylase (P4H) [Myllyharju et al., 2001]. The identified modification involves the addition of a hydroxyl group (OH^-), in the γ-carbon of a proline residue which is essential for the stability of the collagen triple helix [Shoulders et al., 2011].

Prolyl-4-hydroxylases have also been linked to RNA-induced silencing complexes (RISCs), by associating Argonaute 2 stability, with the hydroxylation of an endogenous proline. [Qi et al., 2008]This hydroxylation is attributed to interactions with P4H.

One of the most important roles of proline hydroxylation can be found in oxygen sensing mechanisms. Oxygen deficiency involves rapid accumulation of non-hydroxylated hypoxia Inducible Factor–α (HIF-a), in order to activate transcription involved with glycolysis, to compensate for the energy loss due to reduced oxidative phosphorylation [Fong et al., 2008]. When HiF-a subunits undergo oxygen-dependent hydroxylation on proline residues, they are recognized by the Von Hippel-Lindau (VHL), which in turn targets them for proteosomal degradation. [Fong et al., 2008]. HIF-proly hydroxylases form an evolutionarily conserved family of dioxygenases that uses oxygen and 2-oxoglutarate (2-OG) as co-substrates, and iron and ascorbate as cofactors [Bruick and Mcknight, 2001]. Hydroxylation occurs at position 4 of proline residues.

In addition, a post-translational hydroxylation of the ribosomal protein Rps23p has been reported recently [Loenarz et al., 2014]. A prolyl-3-hydroxylation regulates the termination of translation, by disturbing stop codon read-through in a sequence specific manner. [Loenarz et al., 2014]. This finding demonstrates that 2-OG-oxygenases contribute to the regulation of gene expression, by hydroxylating ribosomal protein S23 (RPS23).

In pro-karyotes however, the only reports of proline hydroxylation have been on free proline residues. [Gorres and Raines, 2010]. Recently crystallographic studies by our group, complemented by mass-spectrometry have identified a novel proline hydroxylation (Pro→2-Hyp conversion), on the active site of polysaccharide de-acetylases on the Bacillus family of bacteria [Fadouloglou et al., 2013]. This unusual post-translational modification, targets the α-carbon of the proline residue, instead of the usual γ-carbon found on the other instances of hydroxylated proline. The only other known case of C_α hydroxylation is the enzymatic conversion of Gly-extended precursor peptides (pro-hormones) to a peptidyl-α-hydroxyglycine intermediate

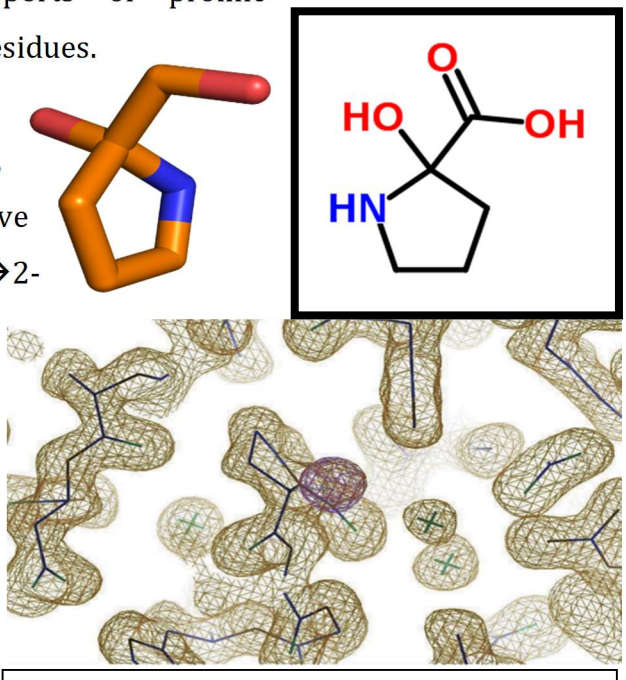


Figure 6. a) Chemical composition of hydroxyproline. b) Sticks representation. c) Coot visualization of electron density (2mFo-DFc) protruding out of the α-carbon of a proline residue, indicating the presence of a hydroxyl group controured at 5σ. [Fadouloglou et al., 2013]

state that is subsequently decomposed. [Takahashi et al., 2002]. Data shown on a later section of this thesis, suggest that this hydroxylation is intertwined with the catalytic function of these enzymes. According to the proposed mechanism, they share at least one common amino acid residue. Although the complete mechanism of autocatalysis has not been unravelled, evidence confirms that this hydroxylation acts as a maturation step, for the active site, and triggers the conformational changes required for de-N-acetylation to go about. Recent unpublished data by our group demonstrate that the mechanism of hydroxylation is dependent on molecular oxygen. This finding compares with the Fe(II)-utilizing mechanisms of prolyl hydroxylases from the superfamily of Fe(II)- and 2-oxoglutarate (2-OG)-dependent dioxygenases [Longbotham et al, 2015]. A proposed model of how this reaction takes place can be seen on figure 7. The 2-Hyp residue (A) gives rise to a hydrogen bond (B) which provides additional stabilization to the transition state of the de-N-acetylation reaction,

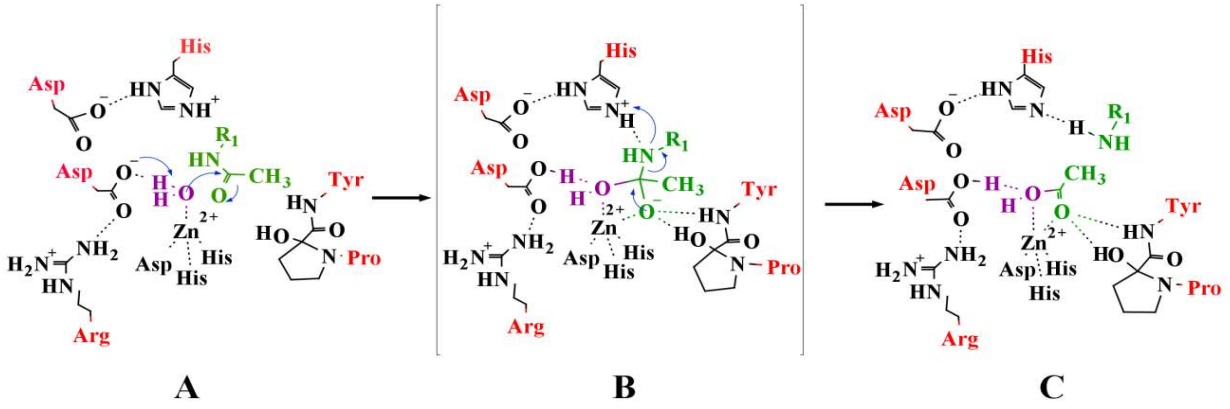


Figure 7. Proposed intertwining of deacetylation with the Pro→2-Hyp conversion in PDAs

complementing its interactions with the divalent metal and with the backbone -NH group (of the residue following 2-Hyp). Zn^{2+} plays the role of Lewis acid, Asp and His residues play the roles of a catalytic base activating the nucleophilic water (magenta), which aid the generation (C) of a free amine and of an acetate product which is hydrogen bonded to 2-Hyp and the -NH group.

It is a frequent phenomenon, to detect novel structures and stereochemical formations, inside the active site of enzymes. In order to perform a catalytic function efficiently, the protein opts for energetically unfavourable conformations inside the environment of the active site.

The dead relative

One member of this family, *Ba3943* (see figure 5), lacks important catalytic and metal binding residues. More specifically, it lacks the proposed catalytic and metal binding aspartic acids, required for the catalytic function of the enzyme. These properties make it a very interesting subject to study, for multiple reasons. Firstly, its possible inactivity, might confirm or deny the proposed theory of de-N-acetylation. The possibility of 2-hyp hydroxylation or its absence, will provide further evidence regarding the association of these two reactions. Mutagenesis studies might reveal the complete amino acid ensemble that orchestrate the autocatalytic, and enzymatic mechanisms. Furthermore, identifying its role in the cell, might reveal an important function, and make it a prominent target for the development of inhibitors. Since the publication of a computational study on kinases, showing the emergence of pseudoenzymes and their importance in cellular processes, it has become clear that evolution does not simply retain inactive relatives [Manning, 2002]. There are still a lot to learn regarding the mechanisms that shape enzymes to adopt novel functions.

Pseudoenzymes

An inactive enzyme is, usually, product of a duplication event, the evolutionary value of which remains unclear. Adaptation relies on the accumulation of point mutations, therefore if mutations in the catalytic machinery are preserved, then it can only be assumed that these amino acids are not required anymore. Although incapable of performing the catalytic activity, the gene is conserved because it serves some different role in the environment of the living cell. Hence, a pseudoenzyme is catalytically inert, nonetheless functional.

The accumulation of new function can materialize via two different genetic mechanisms both involving a duplication event, and divergence. In the neo-functionalisation model, the mutation conferring a new function occurs after the duplication event, where one paralog retains the enzymatic activity and the other shapes into a new protein with a new purpose [Stoltzfus, 1999]. In the sub-functionalisation model, it is presumed that the ancestral protein could perform two different functions and the duplication resulted in a division of labour, where the two distinct gene products, optimized for the different function [Lynch and Force, 2000]. These models though, are just conceptual, and do not correspond to the reality, which can be much more complex.

An example that perfectly illustrates how an enzyme can adopt a novel function can be found on the crystallin protein family. The same product gene can have two different functions, depending on the tissue, and the time of development, it is expressed. For example, the same product gene that results into a chaperone, also produces α-crystallin, and in its crystallized form it is important for maintaining the cytoskeleton and preventing apoptosis [Kim et al., 1987]. Different animals use completely unrelated enzymes to act as crystallins.

The family where the most pseudoenzymes have been reported, is the kinase and phosphatase families [Manning, 2002]. Studies on pseudoenzymes from these two families, reveal a large variety of different roles, mostly regulatory, for these newly found gene products. Often they have been found to participate in the catalytic process, without contributing to the electron transfer process of catalysis. In the EGFR family of kinases, in order for the enzyme to be activated, and be able to complete catalysis, it has to form a dimer. A member of this family, Her3, which has lost essential catalytic residues appears to be able to form dimer, with other members of this family, via its C-lobe, thus contributing to catalysis after all [Zequiraj and Van Aalten, 2010].

Project Goals

The protein studied in this project has all the characteristics that term an enzyme as dead. It retains protein expression, subcellular localization and typical protein folding, but has lost the key amino acids required to complete catalysis.

Throughout the course of this project, the biochemical, enzymatic and structural characterisation of the dead homologue of the PDA family was performed, and the results from this work will be presented. In an attempt to restore the activity of the pseudoenzyme three mutants were synthesized. The conclusions that were drawn from this study will be presented, as well as the future experiments that will attempt to unravel the complete mechanism of autocatalysis via the α-carbon hydroxylation, and its effects on the enzymatic activity of PDA's.

MATERIALS AND METHODS

Mutagenesis Studies

For the purposes of the project, the wild-type ba3943 gene was cloned into a plasmid vector. In an attempt to produce an enzyme that contains the standard residues for catalysis, site directed mutagenesis of *ba3943* wild type gene using PCR (primer extension method) was performed. Three mutants that were produced during the course of the project. The first one was a mutation aiming to restore the catalytic aspartic acid N94D, the second one had an additional mutation aiming at restoring the metal binding aspartic acid V95D, and the triple mutant had also a mutation of an alanine residue in the vicinity of the catalytic centre A183R. The cloning and mutagenesis experiments are being described below, for the wild type gene and the three mutants. All mutagenesis experiments were performed by Professor Bouriotis lab.

Cloning of ba3943 gene of B.anthracis into pET26b expression vector

The *ba3943* gene was amplified from cDNA clones of B.anthracis 7702 strain using PCR. The amplified gene did not contain the signal peptide (amino acid 1-25). Primers were synthesized to incorporate an NdeI at the start and an XhoI site at the end of the gene. The amplified genes were purified, digested with the corresponding enzymes and ligated into pET26b. The resulting product pET26b-ba3943 was an in-frame C-terminal His_6 tag-fused construct in pET-26b, placing the PGNG-dac gene under the transcriptional control of the T_7 lac promoter. C43, pLysS and Star pLysS E.coli strains under a series of conditions were tested for the efficacy to overexpress pET26b-ba3943 construct and C43 (DE3) were selected as the best.

Cloning of *ba3943mut1* gene of *B.anthracis* into pET26b expression vector to produce BA3943 N94D V95D

Ba3943mut1 gene was amplified from purified wild type *ba3943* gene using PCR (primer extension method). *Ba3943mut1* gene contains three point mutations: A on position 279 was converted to G, T on position 283 was converted to A and G on position 284 was converted to C. The amplified gene did not contain the signal peptide (amino acid 1-25). As for the wild type gene, primers were synthesized to incorporate an NdeI at its start and an XhoI site at its end. The amplified genes were purified, digested with the corresponding enzymes and

ligated into pET26b. The resulting product pET26b-ba3943mut1 was an in-frame C-terminal His_6 tag-fused construct in pET-26b, placing the PGNG-dac gene under the transcriptional control of the T_7 *lac* promoter.

As mentioned above, in order to produce *BA3943 N94D V95D*, Asparagine and Valine of MT1 were mutated to Aspartic acids. Asparagine can be substituted by Aspartic acid as they are both polar amino acids. They differ only in that Asparagine contains an amino group in place of a lone oxygen in the aspartic acid. Since *ba3943mut1* gene encodes for a mis-folded protein, it was decided to perform a point mutation. Asparagine was mutated to Aspartic acid (catalytic residue) therefore converting MT1 from TINV to TIDV. The Valine residue remained intact as there are members of the CE4 family which do not retain the conserved Aspartic acid residue which coordinates the metal ion. *B.subtilis* PdaA and BA0424, for example, possess an Asparagine residue in the place of the otherwise conserved Aspartic acid residue (48). Oberbarnscheidt at al., proposed that although an aspartic acid is common at this position, asparagine as well as hydrophobic residues such as Alanine and Valine can also occupy this position. Thus, it is not strictly necessary for a conserved metal ion binding triad.

Cloning of *ba3943mut2* gene of *B.anthracis* into pET26b expression vector to produce BA3943 N94D

Ba3943mut2 gene was amplified from purified wild type *ba3943* gene using PCR (primer extension method). *Ba3943mut2* gene contains a point mutation: A on position 279 was converted to G. The amplified gene did not contain the signal peptide (amino acid 1-25). As for the wild type gene, primers were synthesized to incorporate an NdeI at its start and an XhoI site at its end. The amplified genes were purified, digested with the corresponding enzymes and ligated into pET26b. The resulting product pET26b-ba3943 (279A>G) was an in-frame C-terminal His_6 tag-fused construct in pET-26b, placing the PGNG-dac gene under the transcriptional control of the T_7 *lac* promoter.

After the structure of BA3943 N94D was solved, it was determined that an alanine has to be mutated to back to a well conserved arginine, so a third mutant was designed.

Cloning of *ba3943mut3* gene of *B.anthracis* into pET26b expression vector to produce BA3943 N94D V95D A183R

Ba3943mut3 gene was amplified from purified wild type *ba3943* gene using PCR (primer extension method). *Ba3943mut3* gene contains three point mutations: A on position 279 was converted to G, T on position 283 was converted to A, G on position 284 was converted to C, G on position 554 was converted to A, and C on position 555 was converted to G. The amplified gene did not contain the signal peptide (amino acid 1-25). As for the wild type gene, primers were synthesized to incorporate an NdeI at its start and an XhoI site at its end. The amplified genes were purified, digested with the corresponding enzymes and ligated into pET26b. The resulting product pET26b-ba3943mut1 was an in-frame C-terminal His_6 tag-fused construct in pET-26b, placing the PGNG-dac gene under the transcriptional control of the T_7 *lac* promoter.

Biochemical assays

All biochemical assay experiments were performed by Professor Bouriotis lab. BA3943 was incubated with radiolabeled glycol chitin but no activity was detected. Although glycol chitin is a commonly used substrate for the estimation of deacetylase activity, a variety of substrates, including chitooligomers and peptidoglycan, should be tested before characterizing *BA3943* as inactive.

BA3943 and *BA3943 N94D* were incubated with $GlcNac_6$ and $GlcNAc_7$ at 37°C, overnight. The biochemical assays were performed in a mixture containing the following: 0.2mgr/ml protein, 1mgr/ml substrate, 25mM Tris-HCl pH=7.0 and 1mM $CoCl_2$.

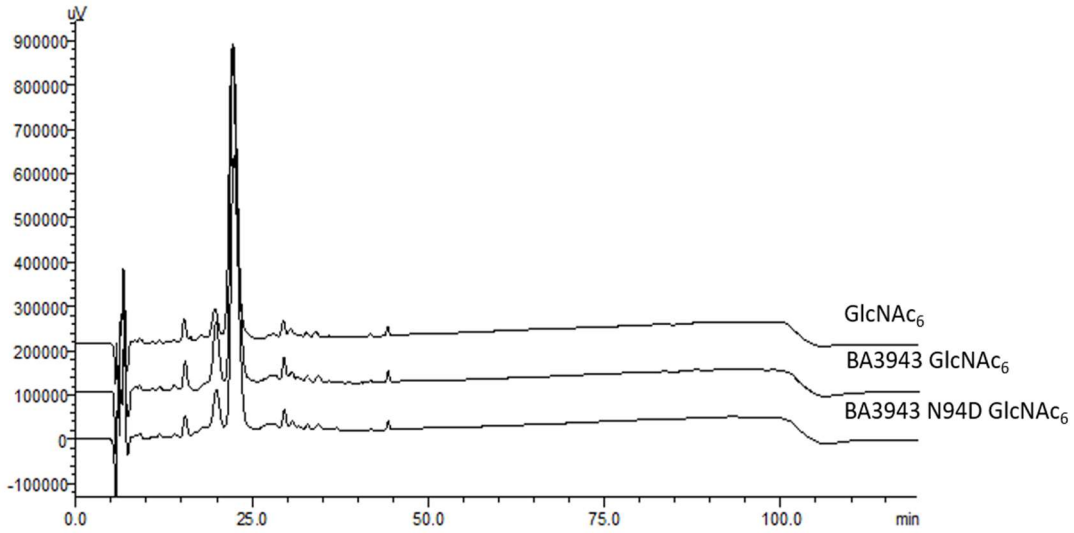


Figure 8. HPLC analysis of BA3943 and BA3943 N94D incubated with $GlcNAc_6$

The reactions were subsequently separated by HPLC using a C18 reverse phase column. Chitooligomers were separated with H_2O, 0.05% TFA (Buffer A) and 50% Acetonitrile, 0.05% TFA (Buffer B) gradients as follows: 85 min 27.78% Buffer B, 90 min 100% Buffer A, 120 min 100% Buffer A at a flow rate of 0.5 ml/min. Neither the wild type nor the mutated protein displayed activity towards the two tested substrates (Figure 8 and 9).

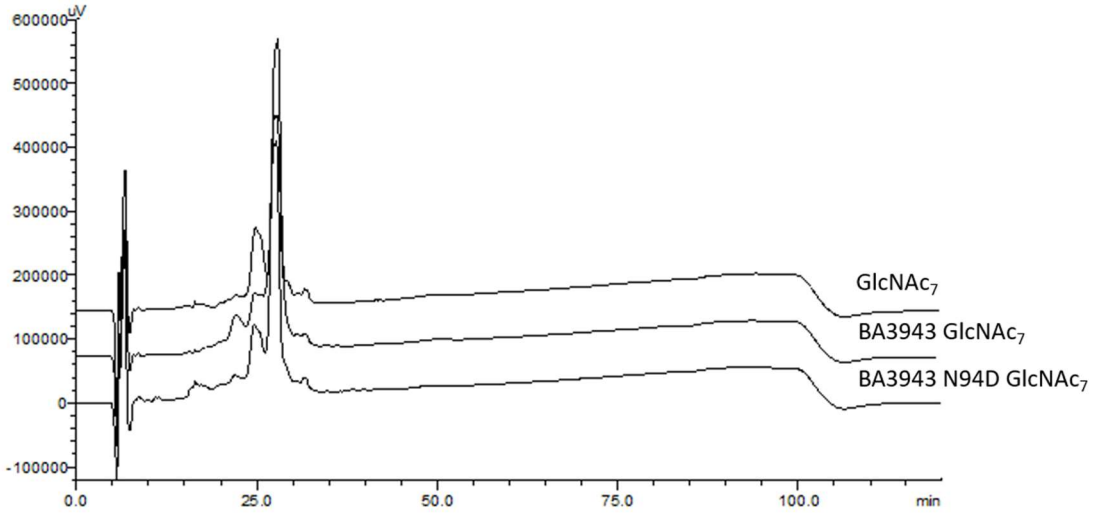


Figure 9. HPLC analysis of BA3943 and BA3943 N94D incubated with $GlcNAc_7$.

Transformation of Ba3943 N94D to C43 *E.coli* competent cells

1.5μλ of the pet26B vector containing the gene of interest, was inserted in 100μλ of the C43 competent cells, and incubate don ice for 25'. Then, it was heat-shocked at 42°C, for 2' and bathed on ice for another 2'. 1ml LB was added to the competent cells and they were incubated for 35' at 37°C. The cells then were centrifuged, the supernatant was removed and 75μλ of the re-suspended pellet were plated in agar plates with Kanamycin. The plates were left to grow O/N at 37°C, at 230rpm.

From the O/N agar plates, 5 colonies far apart from each other were chosen and added to 5ml LB tubes + 50μg/ml Kanamycin. They were left to grow O/N at 37°C, at 230rpm. The next day glycerol stabs from these cultures were taken and kept at -80°C.

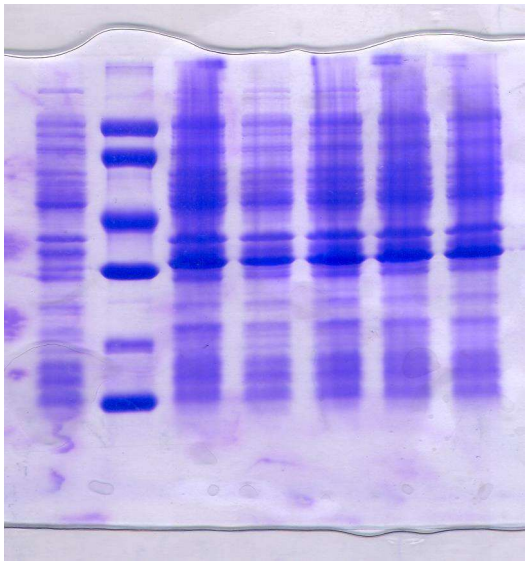


Figure 10 . Gel-electrophoresis of the cell cultures, showing the increased production of a protein with molecular weight 32.164.

Production and Isolation

Using the glycerol stabs, a tip was inserted and discarded in a 200 ml LB with Kanamycin 50mg/ml. The culture was left to grow O/N at 37°C, incubatεδ at 200rpm. The saturated cultures of the transformed deacetylase expression strain were inoculated into 2 Litres of LB medium containing kanamycin as antibiotic and were incubated at 30°C in a shaker incubator to an OD_{600} of 0.6. IPTG at 0.5mM f.c. was added and the cultures were incubated at 30°C O/N. The O/n culture was centrifuged and the resulting pellet, was inoculated in 25mM Tris and 100 mM Nacl. Inhibitors of proteases were added, and then the bacterial cells were lysed and sonicated for 10 30" overlaps, with 30" pauses in between. The resulting mixture was centrifuged at x12500, 6°C, for 60 minutes, and the sample was then loaded to a Nickel chromatography column.

Ni-NTA chromatography column

All the solutions contained 20mM Tris and 300mM NaCl and 10% Glycerol. Imidazole was used to break the NTA-Ni–Histidine bond and elute the protein of interest. The concentration of imidazole increased progressively in the following manner.

Stage	Volume (ml)	Concentration of Imidazole
Wash 1	100	5mM
Wash 2	200	10mM
Wash 3	20	20mM
Elution 1	2x5	100mM
Elution 2	2x5	100mM
Elution 3	2x5	200mM
Elution 4	2x5	200mM
Elution 5-6	2x10	300mM
Elution 7-10	40	300mM

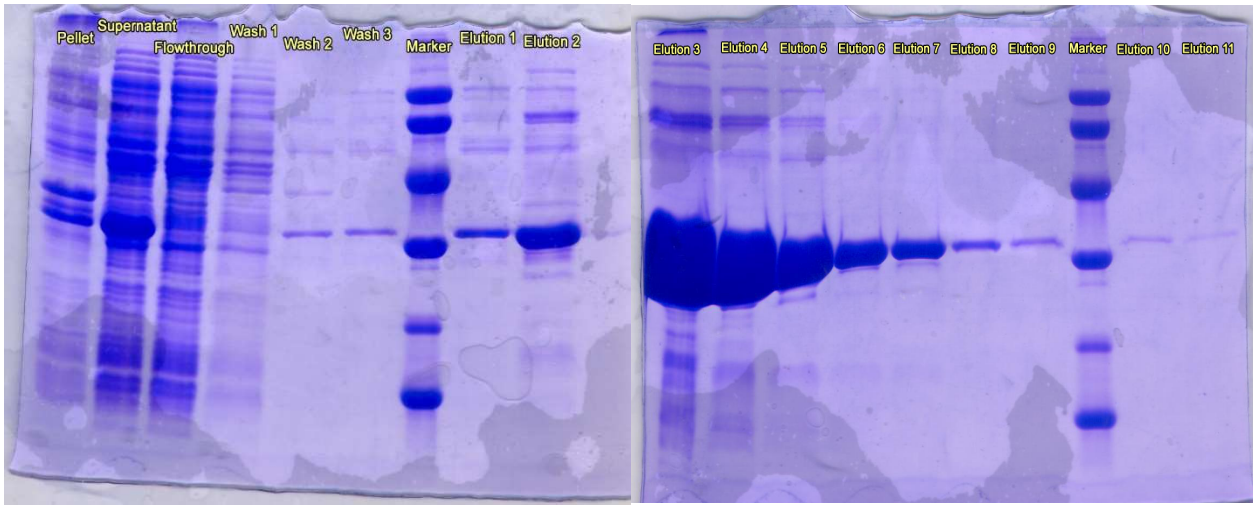


Figure 11. The products of the Ni-chromatography column. Elution 3 was used to perform the crystallographic experiments.

The elutions were separated into volumes of 5ml, in order to achieve maximum protein concentration. The protein had a tendency to aggregate so it was impossible to use gel filtration in order to get a clear sample. That is why immediately after the Ni-chromatography, the most concentrated protein sample (elution 3 = 9.50 mg/ml) was chosen to use for crystallisation experiments. Below there is a table describing the crystallization conditions, which formed the crystal that produced the dataset, with which the structure of BA3943 N94D was solved.

Crystallography

Method	hanging drop vapour diffusion
Plate type	24-well Linbro cell culture plates
Temperature (K)	291
Protein concentration	9,6 mg/ml
Buffer composition of protein solution	20mM Tris 300mM NaCl 100mM Imidazole 10% Glycerol
Composition of reservoir solution	Sodium Acetate CH3COONa (0,3M) 18% Poly-ethene Glycol
Volume and ratio of drop	5μl, 3:2
Volume of reservoir	1000μl

Crystal Measurement and Data Analysis

Diffraction source	X06SA beamline at the Swiss Light Source, Paul Scherrer Institut
Wavelength (Å)	1.000
Temperature (K)	291
Detector	Eiger 16MX (133Hz) detector.
Crystal-detector distance (mm)	166
Rotation range per image (°)	0.2
Total rotation range (°)	360
Exposure time per image (s)	0.1
Space group	$P2_12_12_1$
a, *b*, *c* (Å)	A=58 b=59 c=79
Resolution range (Å)	47.37
Total No. of reflections	77400
No. of unique reflections	41285
Multiplicity	2

Data processing

Data indexing was performed with XDS, and integration with HKL2000. The phase determination step was carried out using the molecular replacement method with the active homologue *BC1960* as a model protein. For this purpose PHASER MR was used. For the purposes of molecular replacement a poly-alanine model was built, and then the sequence was replaced with the actual sequence of *BA3943 N94D*. For this purpose a combination of Pymol, spdbviewer, and Sculptor was used.

Refinement was carried out using PHENIX algorithms. A rigid body refinement was performed, and then a combination of torsion angle and simulated annealing algorithms were implemented. A script that would run, a series of simulated annealing runs, and select the best candidate was then used. This script was designed by Vicky Fadouloglou.

```
set last = 0
set cycles = 500
set count =
while ($count != $cycles)
echo '**********************************************************************'
echo  $count
echo '**********************************************************************'
@ curr = $last + 1
phenix.refine .mtz .pdb \
simulated_annealing=true simulated_annealing.start_temperature=3000
main.random_seed=${curr} output.prefix=ane${curr} << eof
END
eof
if ($status) exit
#@ last++
```

Then the Arp-wArp model building algorithm was implemented to assign the 80 missing residues of the N-terminal domain. Further refinement and addition of waters was followed to finalize the structure.

In Silico methods

The discovery of a hydroxyl (OH group) in the α-carbon of a proline residue in the active site of polysaccharide de-acetylases of *Bacillus cereus*, calls for a careful examination of similar residues, to determine the frequency of this post-translational modification. The characterisation of such a small modification would require either high resolution crystallographic data, or a targeted mass spectrometry analysis (MASSPEC). Otherwise, it would be impossible to identify the addition of a hydroxyl group in the long side chain of a proline residue. Furthermore, this modified proline might be part of an autocatalytic mechanism that promotes active site maturation for the de-N-acetylation reaction to unfold. Similar active site geometries might use the same mechanism of autocatalysis that might as well include the hydroxylated proline. To test for the consistency of the active site geometry, a python script was developed and executed in the entirety of the protein data bank.

The full script is presented is the next page, but the main architectural components, that were used to define the specific geometry are annotated below. To achieve this, the whole protein databank library (till 12/9) was downloaded, and the script performed an execution whereby, it opened each pdb file, and tested it for specific requirements. The cmd python module was used, to perform the structural analysis. The parameters set by the script are described below:

- Presence of a metal
- Metal bonded to 2 Histidines and 1 Aspartic acid
- Maximum metal-Proline distance 8 A

If the pdb file matched the requirements set by the script, its PDB code was returned in a text file, along with information regarding the metal, its distance to proline, and the degrees of angle it forms with the Ca of the proline residue. The whole script used can be found below.

```
from pymol import cmd
import os.path
from pymol import stored
import glob

import glob
a = glob.glob("C:/Python27/Metal/*.pdb")
for f in a:
    cmd.load(f)
    cmd.h_add ("all")
    m = cmd.select ("m", "metals")
    mx = cmd.index ("m")
    for i in mx:
        cmd.select ("mi", "index "+ str(i[1]))
        cmd.select ("btm", "bto. mi")
        cmd.select ("ch", "btm and resn his")
        cmd.select ("ca", "btm and resn asp")
        cmd.select ("pr", "mi around 8 and resn pro")
        cmd.select ("px", "mi around 8 and resn pxu")
        cmd.select ("pr", "br. pr")
        cmd.select ("px", "br. px")
        cmd.select ("pxo", "px and name O")
        cmd.select ("pxc", "px and name Ca")
        cmd.select ("prc", "pr and name Ca")
        pro = cmd.index ("pr")
        pxu = cmd.index ("px")
        pxc = cmd.index ("pxc")
        prc = cmd.index ("prc")
        pxo = cmd.index ("pxo")
        myspace = {'ara': []}
        cmd.iterate("ch","ara.append(resn)", space=myspace)
        cmd.iterate("ca","ara.append(resn)", space=myspace)
        if myspace['ara'].count('HIS') == 2 and myspace['ara'].count('ASP') == 1:
            if len(pro)>0:
                for ii in prc:
                    cmd.select ("pci", "index "+ str(ii[1]))
                    cmd.select ("hg", "neighbor pci")
                    cmd.select ("hg", "hg and elem H")
                    hg = cmd.index ("hg")
                    d = cmd.get_distance('index ' +  str(i[1]), 'index ' +  str(ii[1]), int(0)-1)
                    b = cmd.get_angle('index ' +  str(hg[0][1]), 'index ' +  str(ii[1]), 'index ' +  str(i[1]), state=0)
                    myspace = {'names': []}
                    cmd.iterate('index ' +  str(i[1]), 'names.append(resn)', space=myspace)
                    print "PDBID: " + f[18:22] + "  Metal:", myspace['names'][0]+ "/" + str(i[1]) + "  PRO/" + str(ii[1]) + "  Distance:", d, "
Angle B:", b
            elif len(pxu)>0:
                for iii in pxc:
                    cmd.select ("pxi", "index "+ str(iii[1]))
                    d = cmd.get_distance('index ' +  str(i[1]), 'index ' +  str(iii[1]), int(0)-1)
                    a = cmd.get_angle('index ' +  str(pxo[0][1]), 'index ' +  str(iii[1]), 'index ' +  str(i[1]), state=0)
                    myspace = {'names': []}
                    cmd.iterate('index ' +  str(i[1]), 'names.append(resn)', space=myspace)
                    print "PDBID: " + f[18:22] + "  Metal:", myspace['names'][0]+ "/" + str(i[1]) + "  PXU/" + str(iii[1]) + "  Distance:", d, "
Angle A:", a

    cmd.delete("all")

pymol.cmd.quit()
```

The resulting format of the text file:

PDBID:4hd5 Metal: ZN/5079 PRO/4111 Distance: 4.975 Angle A: 138.30

RESULTS AND DISCUSSION

Biochemical Characterisation of *BA3943*

The protocol developed for the purification and characterisation of this protein was somewhat unusual, and required several months of optimization in order to achieve measurable crystals. Difficulties in concentrating the protein, and its tendency towards aggregation during the first 24 hours after its purification, made it impossible to use gel filtration in order to get a sample clear of other proteins, and very difficult to find a working precipitant for crystallisation. This led to a carefully optimized and executed Nickel chromatography experiment, were the bulk of other proteins was removed in extensive washing steps, and the protein of interest was collected in 5ml tubes, in order to achieve a high concentration directly from the Ni column, and move forward directly to crystallizing the sample, without further delays. Fortunately the bacterial cells achieved a huge production of this protein, as illustrated in figure 12, which resulted in 5ml samples of high concentration.

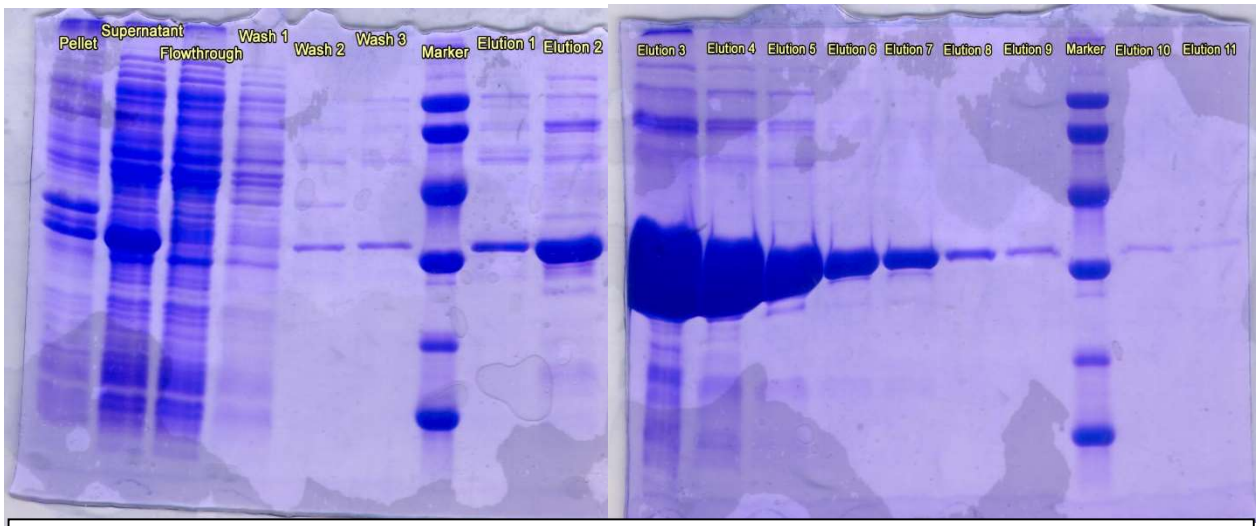


Figure 12 . Gel electrophoresis of the Nickel chromatography experiment, in order to isolate Ba3943 with molecular weight 32,614 KDa.

In order to increase the stability of the protein 10% glycerol was added in the wash and elution solutions used in the purification column. In addition, neither the imidazole used to break the NTA-Ni-Histidine bond nor the glycerol were removed from the solution, as any attempt to do so via dialysis, resulted in fast aggregation of the protein. Early experiments indicated that although glycerol could be removed, and still retain a concentration close to 5-

6 mg/ml, the imidazole was absolutely essential for the stability of the protein. How could the imidazole ring act in order to stabilize the protein is still unknown, though several propositions can be made. One theory could be that the free electron pair of the imidazole ring, could interact with the multiple positively charged lysine residues found on the N-terminus domain of the protein, and stabilize them. The stabilizing nature of imidazole has not been reported in the literature, ever again, however no further experiments to explain this behaviour have been made. Another proposition is that it could bind in the catalytic pocket and stabilize the folding of the active site in such a way, that the whole protein molecule remains stable.

The inability to alter the soluble environment of the protein, meant that the crystallisation conditions would include a sum of different chemical ingredients. After extended optimization, the crystallisation conditions included both glycerol and imidazole, as well as a precipitant, a buffer, a salt and sodium acetate (CH3COONa-) as an additive. Quite remarkable is the fact that despite the variety of chemical constituents in the protein solution, crystallisation was achieved in two different precipitants, and in a range of PH values. The two different precipitants also resulted, in a different crystal shape. When using PEG4000 as a precipitant the resulting crystals were rod-shaped, while when using ammonium sulphate as the precipitant the crystals were cubic shaped (see figure 13). The crystals were then mounted and stored at liquid nitrogen, in order to measure them in a synchrotron facility.

a.

b.

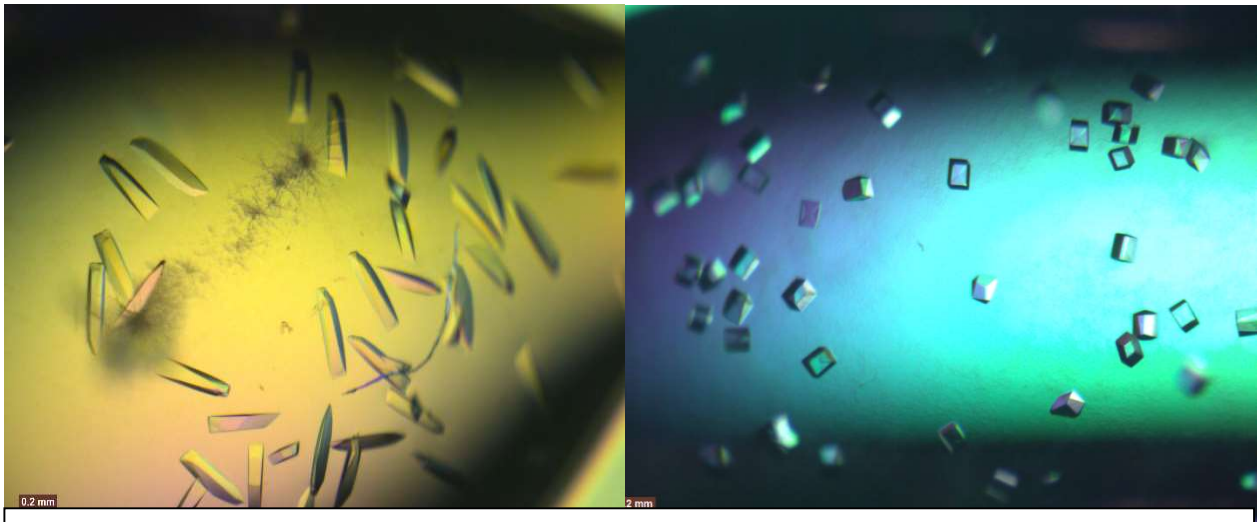


Figure 13. a) Rod-shaped crystals using PEG4000 as a precipitant. b) Cubic shaped crystals, with ammonium sulphate ($NH_2SO_4^-$).

Structure determination by X-ray Crystallography

Data collection was carried out at the PSI synchrotron facility. Several datasets were collected. The highest resolution obtained was at 1.54 Å. Basic data indexing was performed using the XDS software and corresponded to an orthorhombic $P2_12_12_1$ space group, for both types of crystals. Integration and scaling of the experimental data followed. Phase determination was carried out in two steps.

1) Standard molecular replacement algorithms (Phaser) were implemented to determine the catalytic NodB domain, using *BC1960* as the model protein. Their close evolutionary relationship suggest small differences in the basic fold of the catalytic domain. They shared 25% sequence identity.
2) These 80 amino acids, were placed using the Arp-wArp model building algorithm [Langer et al., 2008].

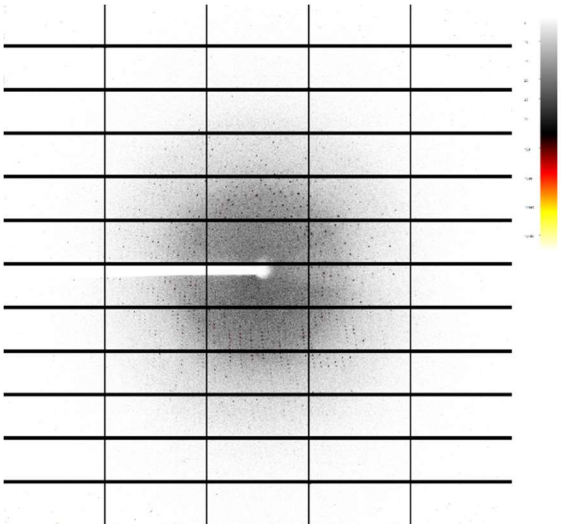


Figure 14. Diffraction pattern.

The combination of these two algorithms, resulted in a model with R-value close to 0.25. An extensive refinement process followed, to unravel the whole structure. The current R-values, are presented on figure 15. Although the refinement process has been finalised for most of the structure, there is still some density left which has not been characterized at the heart of the catalytic centre. The empty density illustrated in figure 16, is still present at 8σ on a standard Fo-Fc map, suggesting the presence of a metal ion (see figure 16). Experiments regarding the exact identity of the metal found in the catalytic centre are still ongoing.

	Final
R-work	0.1891
R-free	0.2031
Bonds	0.006
Angles	0.830

Figure 15. Final R-values

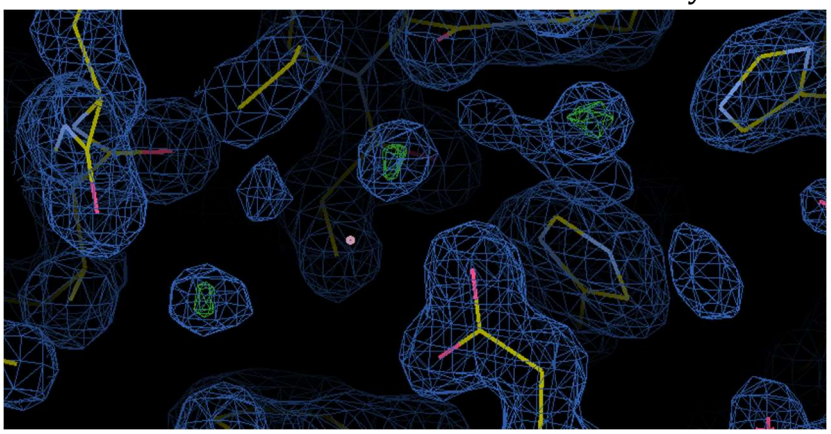
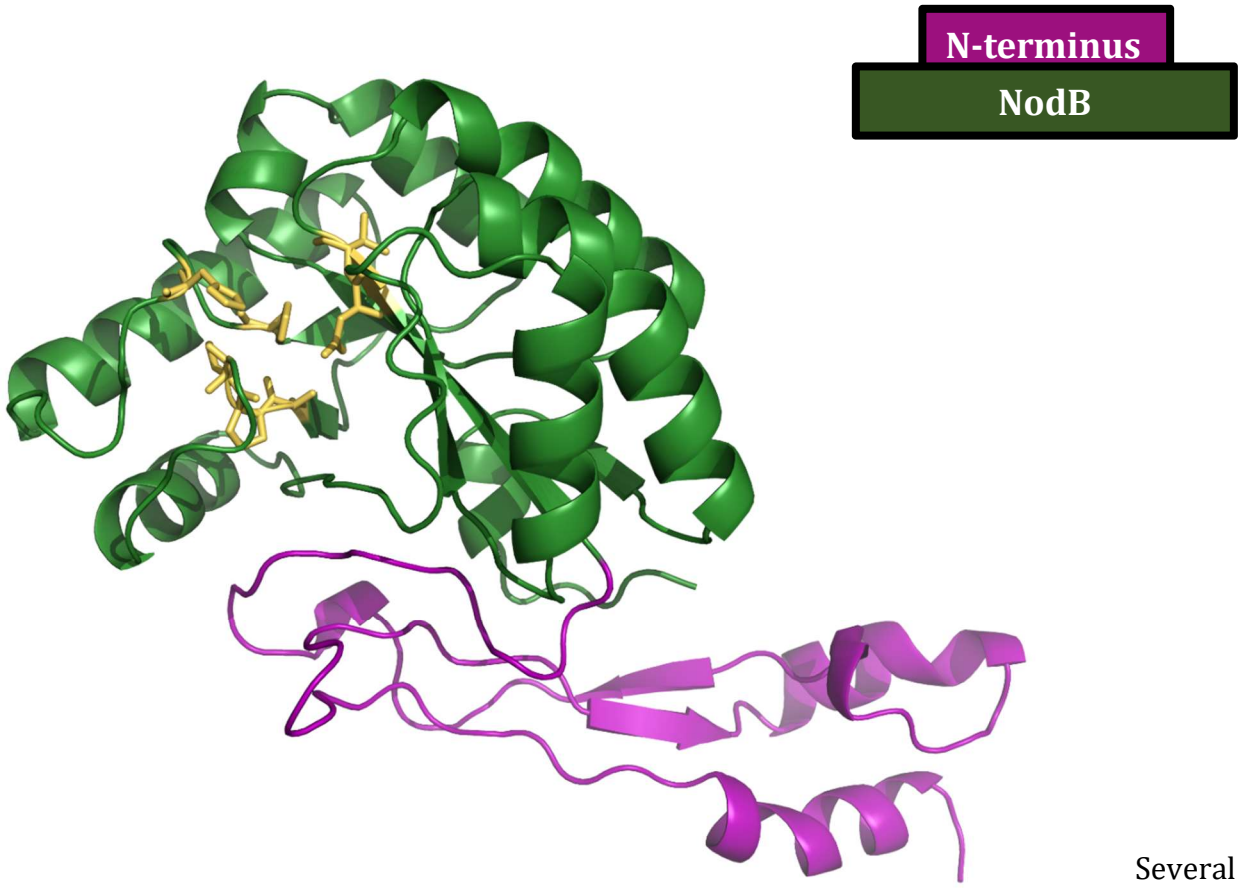


Figure 16. Fo-Fc map at 8σ, and a 2Fo-Fc map at 1.5σ, indicating the presence of a metal ion.

Structural Insights of Ba3943 N94D



Several

Figure 17. The resulting structure of Ba3943 N94D as determined by X-ray crystallography. The catalytic center (pale yellow) is represented in sticks and lies on the catalytic NodB fold (green). The N-terminus (purple) domain, presents an intrinsically disordered hinge domain that has not been characterized before in the literature and its function is completely unknown.

conclusions can be drawn from this structure, regarding 1) the catalytic mechanism, 2) how the 2-hyp formation contributes to catalysis, 3) and the possible function of the protein. According to the proposed theory, wild-type *Ba3943* has lost essential components of the catalytic machinery, which explicate the enzyme's inability to perform catalysis. In the sequence alignment below, it is evident that the catalytic aspartic acid (D) from motif 1 is essential for hydroxylation, and thus de-N-acetylation to occur. When mutated in an active enzyme such as *BA0330 D205A* (see figures 18 & 21), it abolishes activity, and strongly favours the theory that proposes this aspartic acid as the catalytic base for the de-N-acetylation reaction. As presented in figure18, ba3943 has lost the catalytic and metal-binding aspartic acids of motif 1. Mutagenesis experiments were performed to revive its activity, by restoring both the catalytic, and the metal-binding aspartic acid. However, both attempts failed to show any enzymatic activity when tested with glycol-chitin as a substrate.

(see figure 8 & 9).Likewise, the levels of α-carbon hydroxylation of the proline (P)residue of *BA3943* depicted in motif 3 of figure 18, which has been conserved in the dead homologue. Mass spectrometry analysis, performed by Dr. Aivaliotis group, indicates a hydroxylation

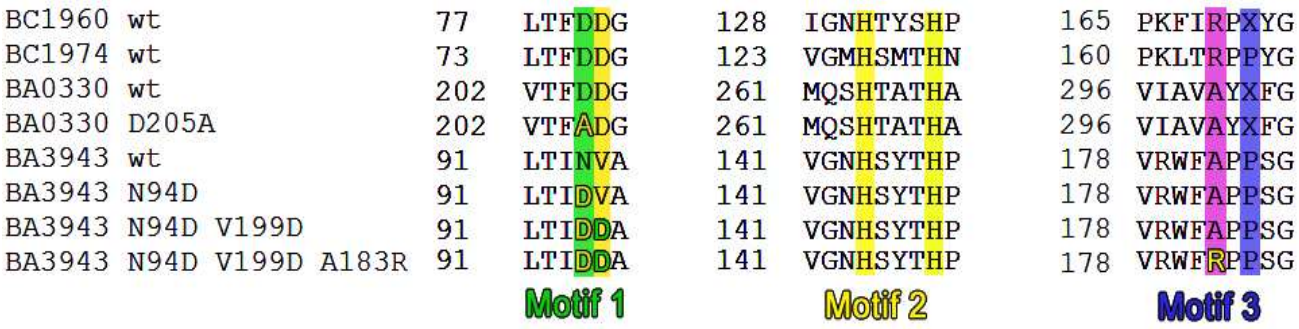


Figure 18. The majority of the deacetylases contain a divalent ion in the active site coordinated by a **His-His-Asp** triad. The first **Asp** residue of **Motif 1** is believed to act as the catalytic base, which activates the catalytic water and the second **Asp** coordinates the metal ion. **Motif 2** contributes the two **His** residues, which coordinate the metal. **Motif 3** contributes a conserved **Arg** residue that coordinates the catalytic base and a strictly conserved **Pro** residue, which may be modified to 2-Hydroxyproline (**Pxu**).

level of approximately 5% (see figure 21), even on the two mutants *Ba3943 N94D*, and *Ba3943 N94D V95D*. In addition, there was no density near the α-carbon of the proline residue, indicating the presence of a carboxyl group.

Structural Superposition of *BA3943 N94D* with *BC1960*.

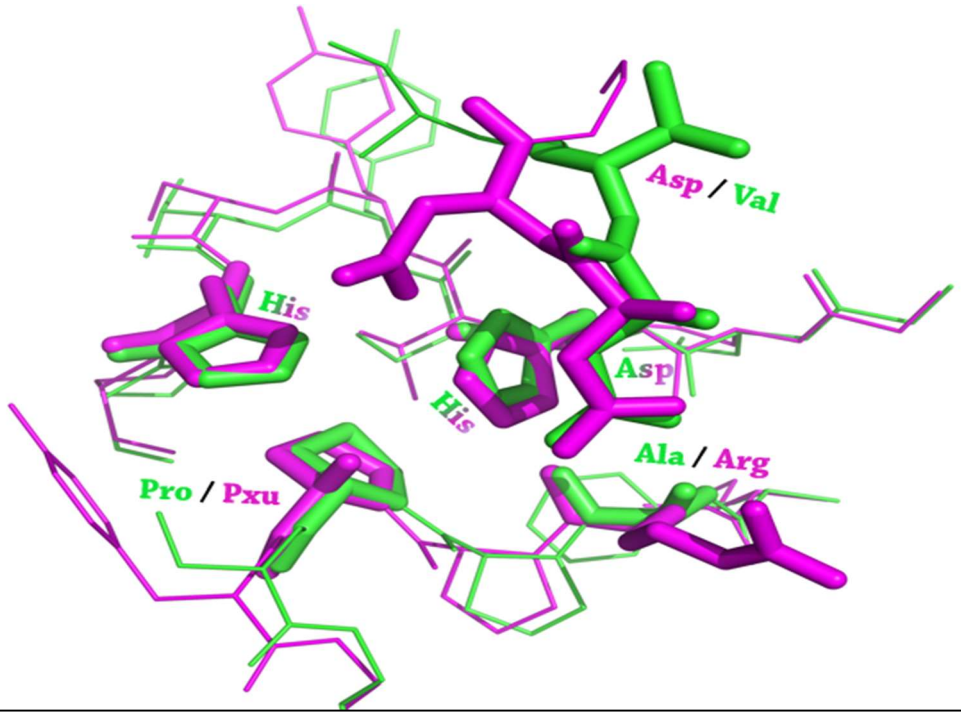


Figure 19. Structural superposition of the catalytic centre of *BA3943 N94D* (green), with an active homologue of the same family, *BC1960* (Magenta).

Structural superposition with an active homologue of the same family, reveals several differences in the catalytic centre that could explain, the absence of enzymatic activity. First and foremost, the metal-binding triad has been disrupted as the valine residue replacing the aspartic acid has a different orientation, protruding on the opposite way in relation to the supposed metal. However, in the double mutant, where this valine residue was mutated back to the metal-binding aspartic acid, the enzymatic activity was absent, suggesting that there are additional amino acid residues that contribute to catalysis, that are absent from this structure. One of the biggest inconsistencies, near the active is the absence of an arginine residue, which is replaced by an alanine in the structure of the dead enzyme, creating potentially a gap and a de-stabilization of the active site. The well-conserved Arg which lies in the vicinity of the catalytic aspartic acid, might contribute to catalysis by co-ordinating the catalytic Aspartic acid. In addition, it may just be a structural component as its presence covers a hole that offers a way for other molecules to enter into the active site (see figure 20).

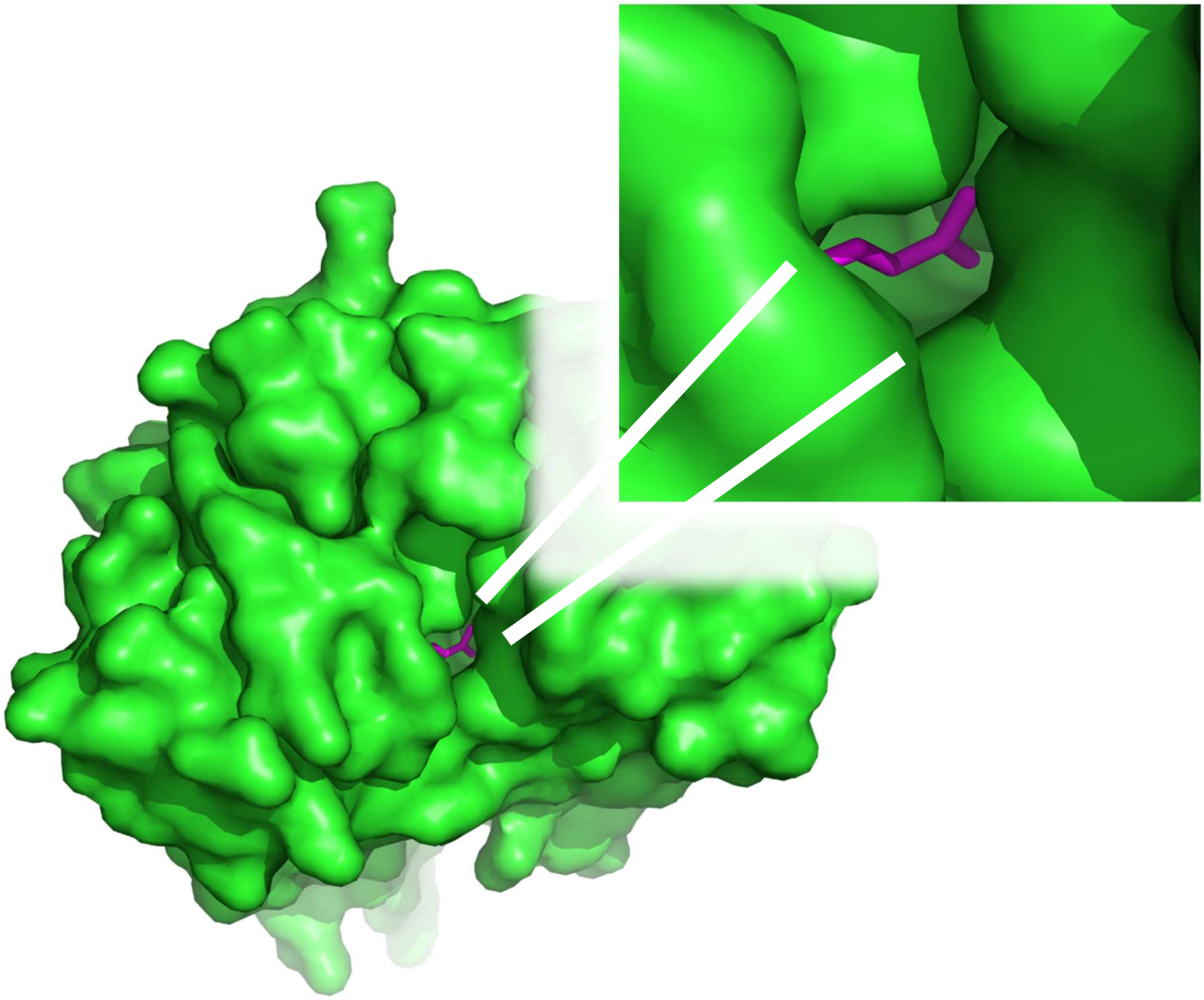


Figure 20. Mesh representation of the NodB domain of Ba3943, showing the cavity created by the absence of the arginine residue.

Source Organism/Uniprot AC	Strain	PDB Code	Function	Conserved Pro (motif3)	%Hyp	De-acetylase activity
B. cereus / BC1960	wt	4L1G	PDA	171	79.88	Active
B. cereus / BC1974	wt	-	PDA	166	31.28	Active
B. anthracis / BA0330	wt	4V33	PDA	302	63.12	Unknown Substrate
B. anthracis / BA0330	D205A	-	PDA	302	6.24	Inactive
B. anthracis / BA3943	wt	-	-	185	8.67	Inactive
B. anthracis / BA3943	N94D	-	-	185	4.05	Inactive
B. anthracis / BA3943	N94D - V95D	-	-	185	6.35	Inactive
B. anthracis / BA3943	N94D-V95D-A183R	-	-	185	25	Inactive

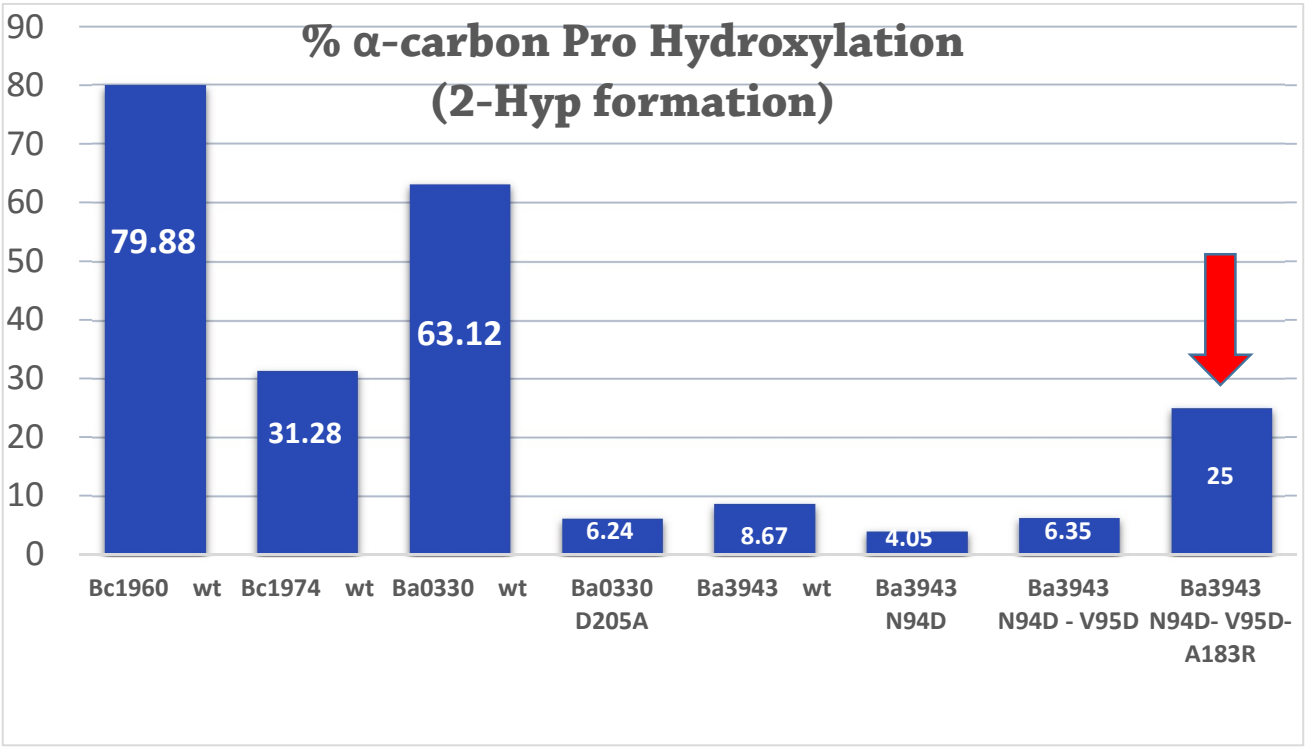


Figure 21. A) The relationship between hydroxylation and de-N-acetylation in the PDA family.

B) Percentage of 2-Hyp in different members of the PDA family, including the mutants of *BA3943*.

Resurrection of the dead

According to the conclusions drawn for the solved structure a new mutant was designed. Using the already synthesized double mutant as a starting template, an additional mutation was performed where the alanine residue was mutated to the well-conserved arginine residue (A183R). Results were astonishing as they reported, an increase in the hydroxylation of the α-carbon, which also coincided with a massive increase in the enzymatic activity of the dead enzyme (see figure 22). Though the enzymatic activity was not absolute, this finding suggests that the arginine residue of the catalytic site contributes indirectly to catalysis, either by filling a gap, which blocks other molecules from inhibiting the reaction, or by co-ordinating the catalytic aspartic acid which acts as a base.

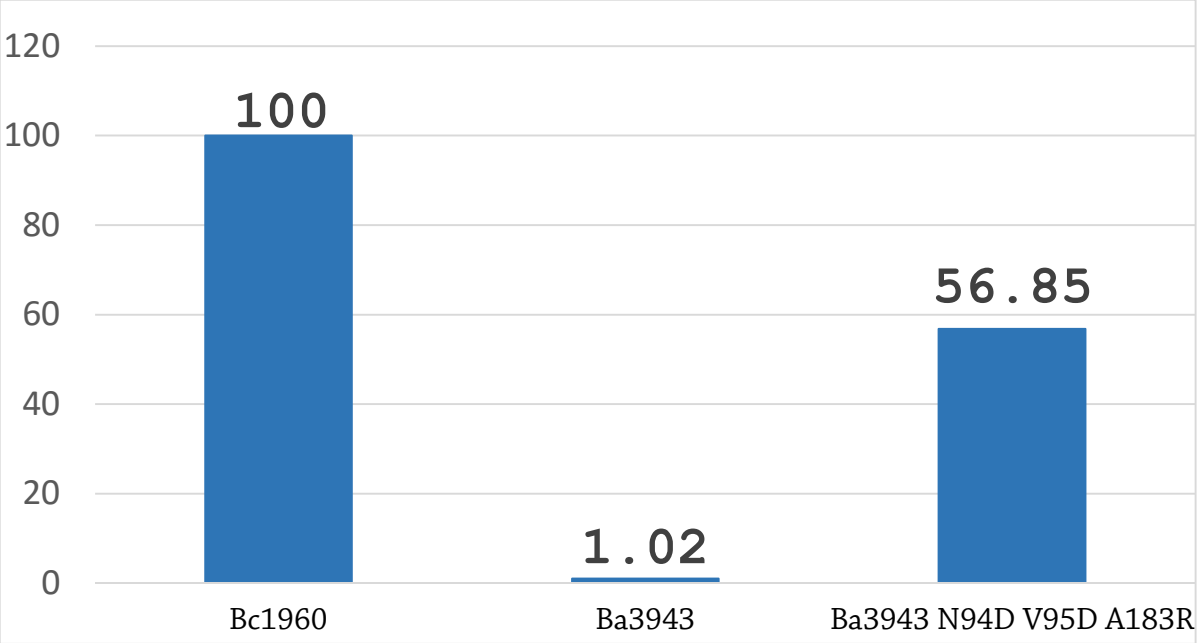


Figure 22. Percentage of enzymatic activity, using glycol chitin as a substrate.

Further optimization by doing site-directed mutagenesis on the *ba3943* gene, can achieve even higher yields of enzymatic efficiency. This process will further clarify, and possibly add even more complexity, on the detailed mechanics of catalysis.

The gap between activity inactivity was five DNA bases apart, when these two proteins share only a quarter of identical amino acid residues. Although it is possible to estimate the time in evolutionary history, when this protein obtained a different function, it has to be noted that, these are dynamic processes, and the estimation would refer to the process of surrendering one function for another.

In silico analysis

The computational analysis as described in the materials and methods section, returned more than 500 deposited structures. This result means that more than 500 of the deposited structures in the PDB database, are consistent with the active site geometry of PDA's, involving a metal and a proline residue with a maximum distance of 8 angstrom apart. These results, hint for a communal mechanism of autocatalysis, involving the α-carbon hydroxylation mechanism. It is also suggestive of similar post-translational modifications, occurring in other amino acid residues other than prolines.

The resulting structures were then plotted, according to the distance between the corresponding metal and the proline residue, and the angles α and β formed between them. A cluster of proteins with a distance of 5Å, (see the circled dots on figure 23), corresponds to closely related enzymes, and enzymes of the same family, that have been known to adopt this active site geometry. Nevertheless, there is a big number of proteins that share extended similarities in their active site, but exhibit variations in the specific geometrical parameters that associate them. Extensive data-mining might provide useful information regarding the hydroxylation mechanism, and the frequency with which it may occur, throughout different proteins from different organisms.

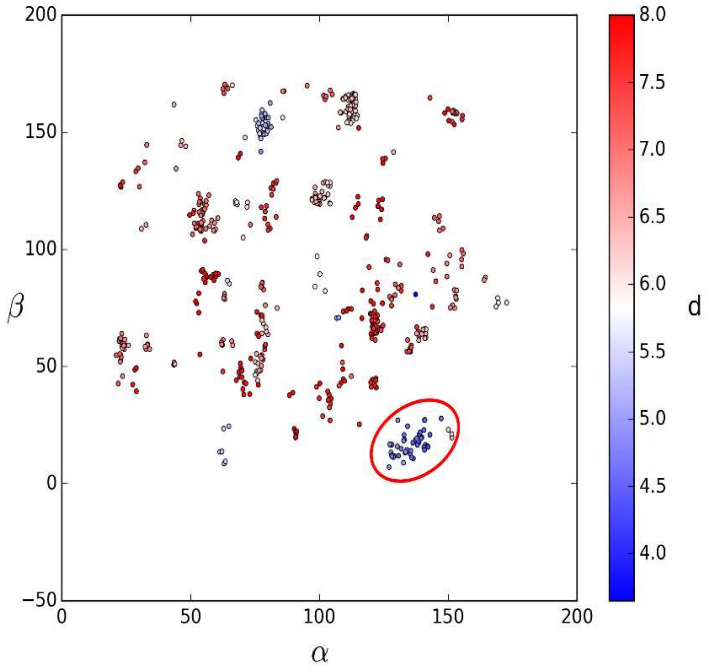


Figure 23. In silico determination of active site geometry. The graph was a product of George's Eustathiou work.

Intrinsically disordered hinge domain

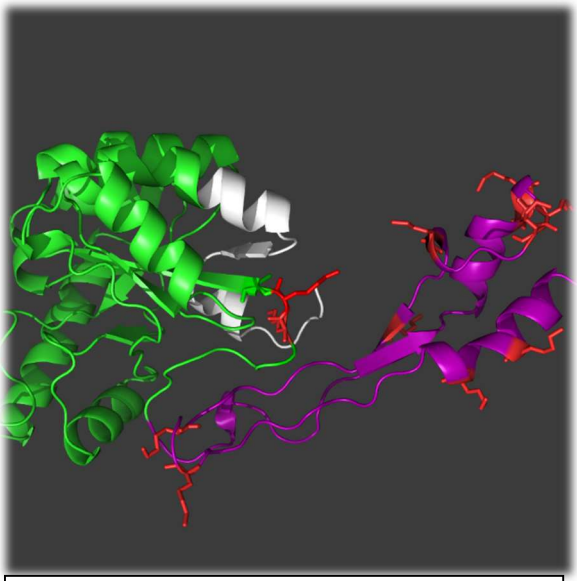


Figure 24. BA3943 Stucture. Nodb(Green), Active Site(White), Lysines (Red), N-terminus (Magenta)

The N-terminus of Ba3943, is composed of 82 amino acids. 11 of those are Lysines (see Figure 24), which translates to a frequency of13,5%. Lysine is a positively charged aliphatic amino acid, were the ε-amino group often engages in hydrogen bonding. Inside the soluble environment of the bacterial membrane, it is inevitable that it will participate in communications, with a variety of macromolecular structures. Unfortunately, experimentally it is trivial to succeed on isolating membrane proteins, thus making it particularly difficult to perform protein-protein interactions, in order to find potential functions of this protein. The temperature factor visualization, demonstrates the high disorder on the lysine rich region. These lysines, may provide the scaffold for the imidazolium to bind, and as a consequence neutralize them. This way the macromolecule retains its solubility and folding. When the amino acid sequence was blasted, all the protein homologues from the entire Bacillus family, were returned (See Figure 26). 15 of them share more than 65% identical residues. This suggests that this novel macromolecular structure, has been retained in the genome, because it performs an essential function. Studying the same protein product of a different organism might give some tips regarding its role.

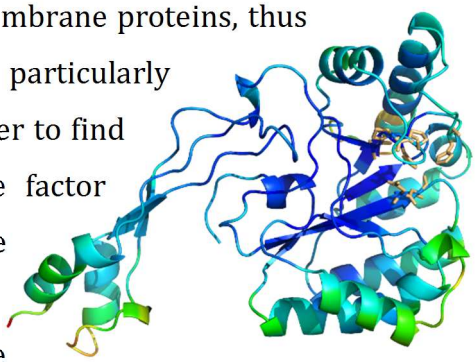


Figure 25. B-factor visualization. (Blue)Order ←→Disorder (Red)

>Bacillus_cereus_26-111_99%

>Bacillus_sp.TD41_26-111_99%

>Streptococcus_pneumoniae_26-111_99%

>Bacillus_thuringiensis_26-111_99%

>Bacillus_wiedmannii_26-111_95%

>Bacillus_weihenstephanensis_26-111_92%

>Bacillus_mycoides_26-111_80%

>Bacillus_gaemokensis_26-111_78%

>Bacillus_bingmayongensis_27-111_80%

>Bacillus_cytotoxicus_27-111_73%

>Geobacillus_thermodenitrificans_37-121_66%

>Geobacillus_stearothermophilus_45-129_66%

>Geobacillus_icigianus_46-130_66%

>Parageobacillus_thermoglucosidans_46-130_67%

>Geobacillus_thermoleovorans_45-129_65%

>Geobacillus_kaustophilus_45-129_65%

Figure 26. Blast search results of the N-terminus domain.

Future Experiments

First and foremost, the structure of the rest of the mutants of *BA3943* will be solved, and careful examination will reveal the conformational changes, that revived the activity of the dead enzyme. This might provide some important clues regarding the autocatalytic hydroxylation mechanism, and how it is associated with the enzymatic activity of PDA's. These experiments are ongoing. Additionally experiments to determine the role of imidazolium as a stabilizer can be designed. Site-directed mutagenesis studies on the lysine residues, of the intrinsically disordered hinge domain, could prove that they are the cause of protein aggregation, and simple dialysis experiments, can showcase the direct effect of imidazolium on the solubility of the protein.

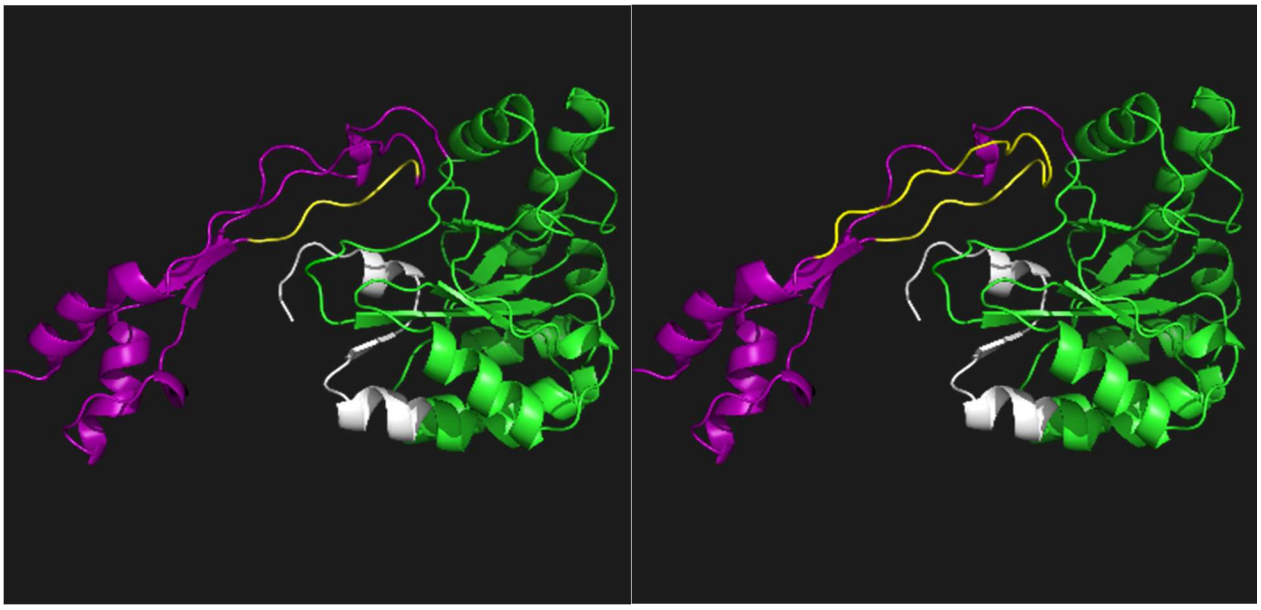


Figure 26. NodB(domain), C-terminus (white), N-terminal domain (purple), Proposed truncation s of the loop of the N-terminal domain(Yellow).

Furthermore, the exact role of this hinge domain can be revealed using truncation experiments. A loop domain that protrudes towards the heart of the NodB fold, and returns to form a second anti-parallel alpha-helix might be important for the stability of the protein. Specific truncations have been designed and are displayed on a cartoon version on figure 26. Enthalpy and entropy values can be measured using a thermidometer facility, to determine the amount of disorder this hinge loop adds, to the overall disorder of the folded protein molecule. Isolation of this domain and protein-protein interactions experiments can also provide evidence regarding the activity of the dead enzyme. The evolutionary conservation of this domain hints for an important regulatory function

Further studies can also be made *in silico* to determine the frequency of such a mechanism. As more details about the autocatalytic mechanism are revealed, more data can be used for the computational searches, to determine the frequency of this unusual post-transaltional modification.

The potential development of inhibitors, to battle the evasive anthrax disease, is of massive importance. Nonetheless, the PDA family of enzymes, have uncovered an even more important mechanism that might be exploited by enzymes of all kinds. Furthermore this family of enzymes includes homologues that have diverged, and have gained a new function. This study provides valuable insight on the mechanisms by which nature shapes proteins to adopt novel functions. The identification of the intrinsically disordered hinge domain is one step forward to the identification of as yet unknown macromolecular assemblies.

BIBLIOGRAPHY

- **Arnaouteli, Sofia, Petros Giastas, Athina Andreou, Mary Tzanodaskalaki, Christine Aldridge, Socrates J. Tzartos, Waldemar Vollmer, Elias Eliopoulos, and Vassilis Bouriotis.** "Two Putative Polysaccharide Deacetylases Are Required for Osmotic Stability and Cell Shape Maintenance InBacillus Anthracis." *Journal of Biological Chemistry* 290.21 (2015): 13465-3478.
- **Balomenou, Stavroula, Sofia Arnaouteli, Dimitris Koutsioulis, Vassiliki E. Fadouloglou, and Vassilis Bouriotis.** "Polysaccharide Deacetylases: New Antibacterial Drug Targets."*Frontiers in Anti-Infective Drug Discovery* 4 (2015): 68-130.
- **Blake CC, Koenig DF, Mair GA, North AC, Phillips DC, Sarma VR.** Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. Nature; 206(4986): (1965) 757-61.
- **Boneca, Ivo G.** "The Role of Peptidoglycan in Pathogenesis." *Current Opinion in Microbiology* 8.1 (2005): 46-53.
- **Bruick RK, McKnight SL**. A conserved family of prolyl-4-hydroxylases that modify HIF. Science 294: (2001); 1337–1340.
- **Callewaert L, Michiels CW**. Lysozymes in the animal kingdom. J Biosci 35(1): (2010); 127-60.
- **CDC, Tm**. "Bacillus Anthracis Gram.jpg." *Wikipedia*. N.p., 30 Mar. 2009. Web. 4 Feb. 2017. <https://commons.wikimedia.org/wiki/File:Bacillus_anthracis_Gram.jpg>.
- **Fadouloglou, Vasiliki, Tsalafouta Aleka, Vassilis Bouriotis, Nicholas Glykos, and Michael Kokkinidis.** "Crystal Structures of Two Polysaccharide Deacetylases FromBacillus Cereus." *Acta Crystallographica Section A Foundations of Crystallography* 69.A1 (2013): n. page. 912.
- **Fadouloglou, Vasiliki E., Maria Kapanidou, Athanasia Agiomirgianaki, Sofia Arnaouteli, Vassilis Bouriotis, Nicholas M. Glykos, and Michael Kokkinidis.** "Structure Determination through Homology Modelling and Torsion-angle Simulated Annealing: Application to a Polysaccharide Deacetylase FromBacillus Cereus." *Acta Crystallographica Section D Biological Crystallography* 69.2 (2013): 276-83.
- **Fong, G-H, and K. Takeda.** "Role and Regulation of Prolyl Hydroxylase Domain Proteins."*Cell Death and Differentiation* 15.4 (2008): 635-41.
- **Fouet A.** "The surface of Bacillus anthracis." Mol Aspects Med ; 30(6) (2009) 374-85.
- **Gorres, K. L., and R. T. Raines.** "Prolyl 4-hydroxylase." *Crit Rev Biochem Mol Biol.* 2.45 (2010): 106-24.

- **Ibrahim HR, Matsuzaki T, Aoki T**. Genetic evidence that antibacterial activity of lysozyme is independent of its catalytic function. FEBS Lett; 506(1): (2001) 27-32
- **Interpro.** N.p., n.d. "NodB Homology Domain (IPR002509)." Web. 01 Mar. 2017. <https://www.ebi.ac.uk/interpro/entry/IPR002509>.
- **Kim, Robert Y., Thomas Lietman, Joram Piatigorsky, and Graeme J. Wistow.** "Structure and Expression of the Duck α-enolase/τ-crystallin-encoding Gene." *Gene* 103.2 (1991): 193-200 Wistow, G., and J. Piatigorsky. "Recruitment of Enzymes as Lens Structural Proteins." *Science* 236.4808 (1987): 1554-556.
- **Langer, G.G., Cohen, S.X., Perrakis, A. & Lamzin, V.S.** Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nature Protocols 3 (2008), 1171-1179.
- **Longbotham, James E., Colin Levy, Linus O. Johannissen, Hanna Tarhonskaya, Shuo Jiang, Christoph Loenarz, Emily Flashman, Sam Hay, Christopher J. Schofield, and Nigel S. Scrutton.** "Structure and Mechanism of a Viral Collagen Prolyl Hydroxylase."*Biochemistry* 54.39 (2015): 6093-105.
- **M. Lynch and a Force**, "The probability of duplicate gene preservation by subfunctionalization.," Genetics, vol. 154, no. 1, ,(2000) pp. 459-73
- **Mock, Michele, and Agnas Fouet.** "Anthrax." *Annual Review of Microbiology* 55.1 (2001): 647-71.
- **Manning, G**. "The Protein Kinase Complement of the Human Genome." *Science* 298.5600 (2002): 1912-934.
- **Mock, Michele, and Tam Mignot.** "Anthrax Toxins and the Host: A Story of Intimacy." *Cellular Microbiology* 5.1 (2003): 15-23.
- **Myllyharju, Johanna, and Kari I. Kivirikko.** "Collagens and Collagen-related Diseases."*Annals of Medicine* 33.1 (2001): 7-21.
- **Loenarz, C., R. Sekirnik, A. Thalhammer, W. Ge, E. Spivakovsky, M. M. Mackeen, M. A. Mcdonough, M. E. Cockman, B. M. Kessler, P. J. Ratcliffe, A. Wolf, and C. J. Schofield**. "Hydroxylation of the Eukaryotic Ribosomal Decoding Center Affects Translational Accuracy." *Proceedings of the National Academy of Sciences* 111.11 (2014): 4019-024.
- **Peschel A, Otto M, Jack RW, Kalbacher H, Jung G, Gotz F. "**Inactivation of the dlt operon in Staphylococcus aureus confers sensitivity to defensins, protegrins, and other antimicrobial peptides." JBiol Chem 274(13): (1999) 8405-10.
- **Philpott DJ, Sorbara MT, Robertson SJ, Croitoru K, Girardin SE**. NOD proteins: regulators of inflammation in health and disease. Nat Rev Immunol; 14(1): (2014) 9-23.
- **Pugliese, Gina, and Martin S. Favero.** "Anthrax as a Biological Weapon: Updated Recommendations for Management." *Infection Control & Hospital Epidemiology* 23.07 (2002): 411.

- **Oberbarnscheidt L., Taylor E. J., Davies G. J., Gloster T. M.** (2007) Structure of a carbohydrate esterase from *Bacillus anthracis*. Proteins 66, 250–252
- **Qi, Hank H., Pat P. Ongusaha, Johanna Myllyharju, Dongmei Cheng, Outi Pakkanen, Yujiang Shi, Sam W. Lee, Junmin Peng, and Yang Shi.** "Prolyl 4-hydroxylation Regulates Argonaute 2 Stability." *Nature* 455.7211 (2008): 421-24.
- **Stoltzfus, Arlin**. "On the possibility of constructive neutral evolution." Journal of Molecular Evolution 49.2 (1999): 169-181.
- **Shoulders, M. D., and R. T. Raines.** "Interstrand Dipole-Dipole Interactions Can Stabilize the Collagen Triple Helix." *Journal of Biological Chemistry* 286.26 (2011): 22905-2
- **Takahashi, Kenichi, Manabu Satani, Yun Gao, and Masato Noguchi**. "The Reaction Mechanism of Peptidylglycine Î±-hydroxylating Monooxygenase." *International Congress Series* 1233 (2002): 235-40.
- **Vlamakis H, Chai Y, Beauregard P, Losick R, Kolter R. Sticking** together: building a biofilm the Bacillus subtilis way. Nat Rev Microbiol 11(3): (2013); 157-68.
- **Zeqiraj, Elton, and Daan Mf Van Aalten.** "Pseudokinases-remnants of Evolution or Key Allosteric Regulators?" *Current Opinion in Structural Biology* 20.6 (2010): 772-81.