Chitin and peptidoglycan deacetylases: discovery, characterization and engineering

Laia Grifoll Romero

DOCTORAL THESIS

Title	Chitin and peptidoglycan deacetylases: discovery, characterization and engineering
Presented by	Laia Grifoll Romero
Centre	IQS school of engineering
Department	Bioengineering
Directed by	Dr. Antoni Planas Sauter

Supported by the European project “European Union’s Seventh Framework Programme for research, technological development and demonstration” under grant agreement number 613931 (NANO3BIO) and the grant BFU2016-77427-C2-1-R from *Ministerio de Economía y Empresa* from Spain. Predoctoral contracts from both grants are acknowledged.

Pel padrí

“Research is formalized curiosity.
It is poking and prying with a purpose”
Zola Neale Hurston

ACKNOWLEDGEMENTS

ACKNOWLEDGEMENTS

Escriure els agraïments és un moment especial.
Els escrius al final, però serà el primer (o l'únic) que la majoria llegirà.
Els deixes per l'últim, però et porten a l'inici, a recordar el camí. I quin camí!

És el moment d'agrair:

Al Dr. Antoni Planas per dur a terme la direcció d'aquesta tesi doctoral. Moltes gràcies per deixar-me formar part del grup de Bioquímica i per confiar en mi per realitzar aquest projecte. Gràcies per tot el coneixement compartit, per portar cada discussió més enllà i fer-me pensar sempre més (i millor).

Al Dr. Xevi Biarnés, per haver estat imprescindible en la realització d'aquest treball, en el que espero vegi reflectides les idees brillantment compartides i explicades en les llargues reunions. Per mostrar-me, a cada reunió, amb cada conversa i a cada trobada, que l'excel·lència pot venir acompanyada d'humanitat, de predisposició i d'entusiasme.

A la Dra. Magda Faijes, per despertar el meu interès aquell primer any al laboratori del màster i per seguir-ho fent, amb cada pregunta i aportació, des d'aquell moment. Moltes gràcies per l'alegria amb què sempre celebres cadascuna de les nostres petites fites.

Al Dr. Marcelo Guerin, por darme la oportunidad de realizar parte de mi trabajo en su laboratorio.

A su gran equipo, excepcionales del primero al último, que me acogió desde el primer día. Gracias por mostrarme otra manera de trabajar y por dejarme participar en las interesantes, entusiastas (¡y largas!) discusiones.

Como no, a "la cuadrilla" por compartir conmigo los mejores pinchos de Bilbao y por hacerme sentir como en casa. Eskerrik asko!

To all members of the Thesis Committee, for agreeing to be a part of it and for the time spent reading and evaluating this thesis. I hope you enjoy it.

A tots els membres del Laboratori de Bioquímica amb els que he coincidit durant aquests anys: professors, col·laboradors, investigadors i estudiants. Quina sort treballar en un entorn que et

permet conèixer tantes persones interessants, i aprendre de cadascuna d'elles. Gràcies a tots pel que m'emporto de cadascú.

Especialment, als estudiants de grau i de màster que han col·laborat en aquest projecte: Pau Martínez, Carlota Miranda, Alicia Rodríguez i Carolin Vaquier. Moltes gràcies per fer-me aprendre al mateix temps que intentava ensenyar.

Als meus companys de màster, sobretot als que van realitzar el TFM al Laboratori de Bioquímica, on va començar aquesta aventura. Gràcies per aquell primer any.

A la Nuria Abajo, pel tàndem que vam fer, per portar "la seva química" a "la meva biologia" i per la seva amistat tots aquests anys.

A Siwen, por el tiempo compartido en "Puig-reig", donde me esperaba con las mejores cenas y la mejor sonrisa cuando sabía que volvería tarde del laboratorio. Gracias por ser una fuente de paz.

Als que m'han acompanyat durant incomptables hores i dies, fent el seu camí al costat del meu. Amb els que tot va començar i que hi seran després que acabi.

A Hugo, por el tiempo, la paciencia y el esfuerzo, por las explicaciones, las enseñanzas y las lecciones. Por ser un referente, científico y personal. Pero, sobretodo, gracias por dejarme ver más de ti, por las conversaciones acompañadas de cañas "graciencas", y por las que quedan por venir. Grazas.

A la Mireia, per ser l'alegria personificada i il·luminar el laboratori amb cada "bon dia". Moltes gràcies per saber-me fer treure l'optimisme de cada situació. Per ser un exemple de resiliència i positivisme.

Al Sergi, per ser "una" més i per aportar el perfecte equilibri entre ciència i entreteniment, entre rigor i llibertat. Pels dissabtes que començàvem arreglant l'HPLC i acabàvem arreglant el món amb una copa a la mà.

A l'Almudena, per demostrar-me que 3 mesos són suficients perquè una persona et sorprengui i t'inspiri per sempre. Perquè del laboratori vas arribar i en vas marxar, però tinc la sort que a la meva vida has decidit quedar-t'hi.

A la Cris, per tot el que hem compartit, per saber-me entendre, per cuidar-me i preocupar-se per mi. Per les llargues converses que, espero, no acabin. Moltes gràcies per ser un exemple de superació.

A la Núria, per haver compartit amb mi aquest camí, literalment, des del primer dia fins a l'últim. Perquè no puc imaginar-me una millor companya de viatge. Moltes gràcies per la teva amistat i

suport en tot moment, per ser una constant en aquests anys d'alts i baixos i per ser un referent (d'aquells que saps que mai podràs igualar).

A l'Estela, per ser, sens dubte, el millor que m'emporto d'aquesta etapa. Moltes gràcies per ser la bondat, l'empatia i l'altruisme personificats. Per la llum que desprèn el teu somriure i pel calor que donen les teves abraçades que, espero, puguin ser més freqüents ben aviat. Moltes gràcies per fer que hagi valgut la pena.

Al Marc, el Bernat i l'Álvaro, que van començar quan jo ja acabava, per arribar al laboratori i portar entusiasme renovat, i per ser el millor relleu possible.

Als biòlegs, amb els que vaig descobrir les meravelles del món de la ciència (i del món universitari). Moltes gràcies per compartir amb mi alguns dels millors anys de la meva vida. Per barrejar microscopis amb festes, apunts amb nits a la vila, classes amb cançons a al gespa, hores de pràctiques amb escapades rurals, plaques de petri amb viatges a l'Ampolla. Especialment:

A la Natalia, la primera doctora, la que ens serveix d'exemple a tots. Moltes gràcies per ser la meva companya al món de les lletres.

A Virginio, porqué, desde otro laboratorio, me ha hecho sentir que hemos recorrido un poco de este camino juntos. Gracias por hacerme sentir orgullosa cuando hablas de "tu (¿mejor?) amiga".

A Alexis, por ser Alexis.

A la Laura, la Marta i la Rosario, per donar-me suport sempre i per acompanyar-me fins a tants cims, a aquest també heu arribat amb mi.

A l'Anna i la Núria, per fer que un estiu compartit a l'eixample es convertís en anys d'aventures i viatges. Perquè, seguint amb el símil, si amb algú he fet el camí, és amb elles.

Ultreia et suseia.

A les de sempre, les del poble, les que no entenen d'experiments i proteïnes, però si de fer-me riure i animar-me. Moltes gràcies pels anys viscuts i pels incomptables moments compartits, sempre juntes, amb els que podria omplir una altra tesi. Gràcies per deixar-me créixer al vostre costat.

A la Marina, per arribar quan ho va fer i per quedar-se, sens dubte, per sempre. Gràcies per ser font d'inspiració i exemple de fortalesa.

A la Laura i Judit, les incondicionals, les persones per les que existeix el concepte de "millor amic", les que han rigut i plorat amb mi en cada passa del camí. Moltes gràcies per fer-me sentir la seguretat de que mai estaré sola i per haver-me donat l'oportunitat de veure com, any a any i pas

a pas, us convertíeu en el que ara sou: bones, fortes, independents, treballadores, sensibles i, per mi, les millors.

A la resta dels meus amics. Gràcies per donar-me l'oportunitat de poder dir que sou molts i molt bons.

A tota la meva família, la catalana i la gallega, perquè no me'n podria haver tocat una de millor.
A toda a miña familia, a catalana e a galega, porque non podería terme tocado unha mellor.

Al Marc (*aka* c, *aka* CM), per aparèixer, per canviar-ho tot, per fer-se imprescindible i per quedar-se. Moltes gràcies per ser font de felicitat, d'il·lusions, d'aventures i de somriures. Moltes gràcies perquè, tot i haver-ne viscut només l'últim tram, que m'acompanyessis ha fet que arribi al final del camí. Gràcies pel que vas escriure en aquella llibreta vora la costa gallega.
I per estimar-me, sempre.

Als meus pares, Jaume i Montse, perquè sé que no em tindran en compte que no trobi paraules suficients per agrair el seu suport constant i el seu amor incondicional. Al llarg d'aquests agraïments he parlat de molts referents, però els primers i els últims sou i sereu sempre vosaltres. Referents d'esforç, d'empatia, de valor i de constància. Gràcies per fer-me qui sóc.
I per estimar-me, sempre.

I gràcies a tu, lector, que tens aquesta tesi a les teves mans (o a la teva pantalla). Gràcies per arribar fins aquí, t'animo a continuar, però no t'ho tindré en compte si baixa el nivell d'atenció.

INDEX

INDEX

SUMMARY

1. SUMMARY

Chitin deacetylases (CDAs) and peptidoglycan deacetylases are members of family 4 carbohydrate esterases (CE4-enzymes), which also includes poly-β1,6-N-acetylglucosamine deacetylases and some acetylxylan esterases. Peptidoglycan is a major component of the bacterial cell wall. Pathogenic bacteria utilize acetylation (6-*O*-acetylation of MurNAc) and deacetylation (2-*N*-deacetylation of GlcNAc and/or MurNAc residues) of its own cell wall peptidoglycan to evade detection by the host's innate immune system, with peptidoglycan deacetylases being proposed as novel antibacterial targets. Deacetylation of MurNAc residues of peptidoglycan is also involved in endospore formation and germination. Chitosans and chitooligosacharides (COS) are bioactive molecules with many current and potential new applications in several fields, and their physicochemical and biological properties are determined by their specific structures. There is a growing interest in developing selective enzymatic approaches to produce sequence-defined chitooligosaccharides to evaluate their biological functions and develop new applications, with CE4 enzymes showing great potential for their use as biocatalysts.

Aiming to increase the available knowledge regarding the specificity of chitin and peptidoglycan deacetylases from CE4 family on different substrates and the sequential and structural features determining these specificities, this work reports the comprehensive characterization of the peptidoglycan deacetylase PdaC from *Bacillus subtilis* (*Bs*PdaC), as well as the selection and characterization of novel CE4 enzymes.

COS deacetylation kinetics has revealed that *Bs*PdaC operates by a multiple-chain mechanism, with $GlcNAc_3$ being the minimal substrate, and that the activity increases with the degree of glycan polymerization. The x-ray structure of the *Bs*PdaC CE4 domain has been determined and the structural comparison with canonical MurNAc and GlcNAc deacetylases has led to propose that PdaC is the first member of a new subclass of peptidoglycan MurNAc deacetylases, with yet unknown biological functions, based on the differential functional and structural characteristics. Furthermore, the importance of key enzyme-substrate interactions has been highlighted by mutational analysis and the specificity and pattern of deacetylation has been altered by mutating a key substrate-interacting residue.

The search for new CE4 enzymes based on a phylogenetic and bioinformatic analysis of the family is presented. Several new sequences have been selected and the recombinant proteins initially characterized, from which *Tilletia controversa* CDA have been identified as a novel CDA and *Bacillus cereus* Pda has been proposed as a novel MurNAc deacetylase for further characterization.

1. SUMARI

Les desacetilases de quitina (CDAs) i les desacetilases de peptidoglicà són membres de la família 4 de les esterases de carbohidrats (CE4), que també inclou desacetilases de poli-β1,6-N-acetilglucosamina i esterases d'acetilxilà. El peptidoglicà és un dels components principals de la paret cel·lular bacteriana. Els bacteris patògens utilitzen l'acetilació (6-*O*-acetilació de MurNAc) i desacetilació (2-*N*-desacetilació de residus de GlcNAc i/o MurNac) del seu propi peptidoglicà per evadir la detecció pel sistema immunitari innat de l'hoste, havent sigut les desacetilases de peptidoglicà proposades com a noves dianes antibacterianes. La desacetilació de residus de MurNAc del peptidoglicà també està involucrada amb la formació i germinació d'endòspores. Els quitosans i els oligosacàrids de quitosà (COS) són molècules bioactives amb diverses aplicacions, conegudes i potencials, en varis camps, i les seves propietats fisicoquímiques i biològiques estan determinades per les seves estructures específiques. Hi ha un interès creixent pel desenvolupament d'aproximacions enzimàtiques per la producció d'oligòmers de quitosà definits per tal d'avaluar-ne les funcions biològiques i desenvolupar noves aplicacions, presentant els enzims de la família CE4 un gran potencial pel seu ús com a biocatalitzadors.

Amb l'objectiu d'incrementar el coneixement disponible sobre l'especificitat de les desacetilases de quitina i peptidoglicà en diferents substrats i les característiques seqüencials i estructurals que determinen aquestes especificitats, aquest treball reporta la caracterització exhaustiva de la desacetilasa de peptidoglicà PdaC del microorganisme *Bacillus subtilis* (*Bs*PdaC*)*, així com la selecció i caracterització de nous enzims de la família CE4.

La cinètica de desacetilació amb COS ha revelat que *Bs*PdaC actua seguint un mecanisme de cadena múltiple, essent el $GlcNAc_3$ el seu substrat mínim, i que la seva activitat augmenta amb el grau de polimerització del glicà. S'ha determinat l'estructura de rajos x del domini CE4 de *Bs*PdaC i la comparació estructural amb desacetilases canòniques de MurNAc i GlcNAc ha dut a proposar que, en base a les seves característiques funcionals i estructurals diferencials, PdaC és el primer membre d'una nova subclasse de desacetilases de MurNAc amb funcions biològiques encara desconegudes. A més, s'ha destacat, mitjançant un anàlisi mutacional, la importància d'una sèrie d'interaccions enzim-substrat, i s'han alterat l'especificitat i el patró de desacetilació mutant un residu que forma interaccions clau amb el substrat.

Es presenta la cerca de nous enzims CE4 basat en un anàlisi filogenètic i bioinformàtic de la família. S'han seleccionat diverses noves seqüències i s'ha realitzat la caracterització inicial de les proteïnes recombinants, d'entre les quals s'ha identificat la CDA de *Tilletia controversa* com una nova desacetilasa de quitina i s'ha proposat la proteïna Pda de *Bacillus cereus* com una nova desacetilasa de MurNac per la seva caracterització addicional.

1. RESUMEN

Las desacetilasas de quitina (CDAs) y las desacetilasas de peptidoglicano son miembros de la familia 4 de las esterasas de carbohidratos (CE4), que también incluye desacetilasas de poli-β1,6-N-acetilglucosamina y esterasas de acetilxilano. El peptidoglicano es uno de los principales componentes de la pared celular bacteriana. Las bacterias patógenas utilizan la acetilación (6-*O*-acetilación de MurNAc) y desacetilación (2-*N*-desacetilación de residuos de GlcNAc y/o MurNac) de su propio peptidoglicano para evadir la detección por el sistema inmunitario innato del huésped, habiendo sido las desacetilasas de peptidoglicano propuestas como nuevas dianas antibacterianas. La desacetilación de residuos de MurNAc del peptidoglicano también está relacionada con la formación y germinación de endosporas. Los quitosanos y los oligosacáridos de quitosano (COS) son moléculas bioactivas con diversas aplicaciones, conocidas y potenciales, en varios campos, y sus propiedades fisicoquímicas y biológicas están determinadas por sus estructuras específicas. Existe un interés creciente en el desarrollo de aproximaciones enzimáticas para la producción de oligómeros de quitosano definidos para evaluar sus funciones biológicas y desarrollar nuevas aplicaciones, presentando las enzimas de la familia CE4 un gran potencial para su uso como biocatalizadores.

Con el objetivo de incrementar el conocimiento disponible en relación con la especificidad de las desacetilasas de quitina y peptidoglicano en diferentes sustratos y con las características secuenciales y estructurales que determinan estas especificidades, este trabajo reporta la caracterización exhaustiva de la desacetilasa de peptidoglicano PdaC de *Bacillus subtilis* (*Bs*PdaC), así como la selección y caracterización de nuevas enzimas de la familia CE4.

La cinética de desacetilación con COS ha revelado que *Bs*PaC actúa siguiendo un mecanismo de cadena múltiple, siendo $GlcNAc_3$ su sustrato mínimo, y que su actividad aumenta con el grado de polimerización del glicano. Se ha determinado la estructura de rayos x del dominio CE4 de *Bs*PdaC y la comparación estructural con desacetilasas canónicas de MurNAc y GlcNAc ha llevado a proponer que, en base a sus características funcionales y estructurales diferenciales, PdaC es el primer miembro de una nueva subclase de desacetilasas de MurNAc con funciones biológicas todavía desconocidas. Además, se ha destacado la importancia de una serie de interacciones enzima-sustrato, y se han alterado la especificidad y el patrón de desacetilación mutando un residuo que forma interacciones clave con el sustrato.

Se presenta una búsqueda de nuevas enzimas CE4 basada en el análisis filogenético y bioinformático de la familia. Se han seleccionado diversas nuevas secuencias y se ha realizado la caracterización inicial de las proteínas recombinantes, de entre las cuales se ha identificado la CDA de *Tilletia controversa* como una nueva desacetilasa de quitina y se ha propuesto la proteína Pda de *Bacillus cereus* como una nueva desacetilasa de MurNAc para su caracterización adicional.

LIST OF PUBLICATIONS AND COMMUNICATIONS

LIST OF PUBLICATIONS AND COMMUNICATIONS

Publications

1. Expression and specificity of a chitin deacetylase from the nematophagous fungus *Pochonia chlamydosporia* potentially involved in pathogenicity.

A. Aranda-Martínez, L. Grifoll-Romero, H. Aragunde, E. Sancho-Vaello, X. Biarnés, L. Vicente López-Llorca, A. Planas.

Scientifics Reports 8, 2170 (2018). DOI: 10.1038/s41598-018-19902-0.

2. Chitin deacetylases: structures, specificities, and biotech applications.

L. Grifoll-Romero, H. Aragunde, X. Biarnés, A. Planas.

Polymers 10, 352 (2018). DOI: https://doi.org/10.3390/polym10040352

3. Structure-function relationships underlying the dual N-acetylmuramic and N-acetylglucosamine specificities of the peptidoglycan deacetylase PdaC from *Bacillus subtilis.*

L. Grifoll-Romero, M.A. Sainz-Polo, D. Albesa-Jové, M.E. Guerin, X. Biarnés, A. Planas.

Journal of Biological Chemistry 294, 19066 (2019). DOI: 10.1074/jbc.RA119.009510

Communications at scientific conferences and meetings

1. Carbohydrate esterase family CE-4 enzymes active on chitooligosaccharides. Expanding the repertoire of deacetylation patterns.

L. Grifoll-Romero, H. Aragunde, X. Biarnés, A. Planas.

Poster communication, presented at "XII carhohydrate symposium" (Madrid, March 15/16, 2016).

2. Peptidoglycan deacetylase acting on chitin oligosaccharides.

L. Grifoll-Romero, H. Aragunde, X. Biarnés, A. Planas.

Poster communication, presented at "CBM12 – 12th Carbohydrate Bioengineering Meeting" (Viena, April 14, 2017).

3. Peptidoglycan deacetylase acting on chitin oligosaccharides.

L. Grifoll-Romero

Poster communication, presented at "19th European Carbohydrate Symposium EUROCARB" (Barcelona, July 3, 2017).

4. Expression and specificity of a chitin deacetylase catalytic domain from the nematophagous fungus *Pochonia chlamydosporia* potentially involved in pathogenicity

A. Aranda-Martínez, L. Grifoll-Romero, H. Aragunde, E. Sancho-Vaello, X. Biarnés, L. Vicente López-Llorca, A. Planas.

Poster communication, presented at "19th European Carbohydrate Symposium EUROCARB" (Barcelona, July 3, 2017).

5. Enzyme engineering for the production of chitosans with defined deacetylation pattern.

A. Planas, X. Biarnés, H. Aragunde, C. Alsina, L. Grifoll-Romero, S. Pascual, A. Aranda-Martínez.

Oral communication, presented at "NANO3BIO Project Final Dissemination Event" (Hyderabad (India), September 20, 2017).

6. Peptidoglycan deacetylase acting on chitin oligosaccharides.

L. Grifoll-Romero, H. Aragunde, X. Biarnés, A. Planas.

Poster communication, presented at "NANO3BIO Project Final Dissemination Event" (Hyderabad (India), September 21, 2017).

7. Peptidoglycan deacetylase acting on chitin oligosaccharides. Structure and characterization.

L. Grifoll-Romero, X. Biarnés, A. Planas.

Poster communication, presented at "29th International Carbohydrate Symposium" (Lisbon, July 18, 2018).

8. Peptidoglycan deacetylase acting on chitin oligosaccharides. Structure and characterization.

L. Grifoll-Romero, X. Biarnés, A. Planas.

Poster communication, presented at "XII carhohydrate symposium" (Madrid, March 15/16, 2016).

9. Structure-function relationship underlying the dual specificity of the peptidoglycan deacetylase *Bs*PdaC.

L. Grifoll-Romero, M.A. Sainz-Polo, D. Albesa-Jové, M.E. Guerin, X. Biarnés, A. Planas.

Poster communication, presented at "20th European Carbohydrate Symposium EUROCARB**" (Leiden, July 2, 2019).**

LIST OF FIGURES

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES

ABBREVIATIONS

ABBREVIATIONS

[X]	*Concentration of compound X*
3D	*Three dimensional*
A	*GlcNAc, in oligosaccharide sequences of chitin and chitosan*
AA	*Auxiliary Activity*
AcOMU	*4-methylumbelliferyl acetate*
ADP	*Adenosine diphosphate*
AK	*Acetate Kinase*
ATP	*Adenosine triphosphate*
DNA	*Deoxyribonucleic acid*
BCA	*Bicinchoninic acid*
BM	*Binding mode*
BSA	*Bovine serum albumin*
CAZy	*Carbohydrate Active EnZymes*
CBM	*Carbohydrate Binding Module*
CCME	*Chitin and Chitosan Modifying Enzyme*
CD	*Catalytic domain*
cd	*Circular dichroism*
CDA	*Chitin deacetylase*
CE	*Carbohydrate esterases*
COD	*Chitooligosaccharide deacetylase*
COS	*Chitosan oligosaccharides*
D	*GlcN, in oligosaccharide sequences of chitin and chitosan*
DA	*Degree of acetylation*
DMF	*Dimethylformamide*
DP	*Degree of Polymerization*
DSF	*Differential Scanning Fluorimetry*
EDTA	*Ethylenediaminetetraacetic acid*
ESI	*Electro Spray Ionization*
FL	*Full-length*
FPLC	*Fast Protein Liquid Chromatography*
GH	*Glycosil Hydrolase*
GlcN, $GlcNH_2$	*Glucosamine (*D-GlcN*p*NH_2*)*

GlcNAc	*N-acetyl glucosamine (*D-GlcN*p*NAc*)*
GT	*Glycosyltransferase*
HABA	*2-Hydroxy-4-aminobenzoic acid*
HILIC	*Hydrophylic Interaction Liquid Chromatography*
HMM	*Hidden Markov Model*
HPLC	*High Performance Liquid Chromatography*
HPSEC	*High Performance Size Exclusion Chromatography*
ICP-MS	*Inductively Coupled Plasma Mass Spectrometry*
iGlu	*iso-glutamate*
IMAC	*Ionic Metal Affinity Chromatogtraphy*
IPTG	*Isopropyl β-D-1 thiogalactopyranoside*
LB	*Lubia-Bertani culture medium*
m-DAP	*meso-diamopimelic acid.*
MAFFT	*Multiple alignment using fast Fourier transform*
MALDI-TOF	*Matrix Assisted Laser Desorption/Ionization – Time of Flight*
MCS	*Multiple Cloning Site*
MDP	*N-Acetylmuramyl-L-alanyl-D-isoglutamine*
MS	*Mass Spectrometry*
MS^2	*MS/MS*
MSD	*Mass Spectrum Detector*
MT	*Motif*
MU	*4-methylumbelliferone*
MurNAc	*N-acetyl muramic acid*
Mw	*Molecular weight*
MWCO	*Molecular weight cutoff*
NAD^+	*Oxidized nicotinamide adenine dinucleotide*
NADH	*Reduced nicotinamide adenine dinucleotide*
NF	*Nod Factor*
NJ	*Neighbor-Joining*
NLR	*Nod-Like Receptor*
OD	*Optical density*
OD_{600}	*Optical density at 600 nm*
o/n	*Over night*
PA	*Pattern of Acetylation*

PAMP	*Pathogen Associated Molecular Pattern*
PB	*Phosphate Buffer*
PBS	*Phosphate Buffer Saline*
PCR	*Polymerase Chain Reaction*
PDF	*Protein Data Bank*
PGN	*Peptidoglycan*
PGRP	*Peptidoglycan recognition protein*
PL	*Polysaccharide Lyase*
PMSF	*Phenylmethylsulfonyl fluoride*
PRR	*Pattern Recognition Receptor*
PSSM	*Position Specific Scoring Matrix*
PTA	*Phosphotransacetylase*
QC-PCR	*Quick Change Polymerase Chain Reaction*
RMSD	*Root Mean Square Deviation*
rpm	*Revolutions per minute*
RT	*Room temperature*
RT-PCR	*Real-time PCR*
SDS	*Sodium Dodecyl Sulfate*
SDS-PAGE	*Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis*
SIM	*Single Ion Monitoring*
SEC	*Size Exclusion Chromatography*
SD	*Standard Deviation*
SDM	*Site directed mutagenesis*
T_m	*Melting temperature*
TLR	*Toll-like receptor*
TSA	*Thermal Shift Assay*
UHPLC	*Ultra High-Performance Liquid Chromatography*
UPLC	*Ultra Performance Liquid Chromatography*
UV	*Ultraviolet*
v_0	*Initial velocity*
V_0	*Dead volume*
V_e	*Elution volume*
VMD	*Visual Molecular Dynamics*
VWD	*Variable Wavelength Detector*
wt	*wild type*

INTRODUCTION

2. INTRODUCTION

2.1. Enzymes as biocatalysts

Biotechnology includes any technological application consisting in the utilization or manipulation of biological systems, living organisms or their derivatives, to obtain or modify products or implement processes, generally by means of natural and engineering sciences [1].

Biocatalysis is the process used for the transformation or generation of certain products using a **biocatalyst**, which can either be an enzyme, an enzyme complex, a cell organelle or a whole cell from different sources (microbial, plant or animal origin). A biocatalyst can be free or immobilized, a characteristic that has consequences with respect to substrate supply and mass transfer, and in the case of viable cells, physiology [2]. Scientific and technological advances have positioned biocatalysis as an alternative to synthetic chemistry and traditional organic catalysis. Throughout the years, there have been three main waves of innovations that have let the biocatalysis field to its current position. The first one, more than a century ago, started with the recognition from scientists that components of living cells could be used to perform chemical transformations. The second wave, in the 1980s and 1990s, consisted in the expansion of the enzyme's substrates range via initial protein engineering technologies that allowed the expansion of biocatalysis to the manufacture of pharmaceutical intermediates. Finally, from late 1990s to present, the use of directed evolution methods, with advanced DNA technologies and in combination with new bioinformatics tools, has allowed to engineer remarkable new capabilities into enzymes [3].

In cells, biochemical reactions are catalyzed by **enzymes**, proteins that have evolved to act in an efficient and specific way under biologically viable conditions (mild conditions required for the conservation of functionality and integrity of biological systems) [4]. As any other catalyst, enzymes do not alter the chemical equilibrium of the reaction (which only depends on the free energy difference between reactants and products), but they accelerate the rate at which the equilibrium state is reached. In the absence of enzymes, which are responsible of reducing the activation energy of the reaction ($\Delta G^{\neq}$), most metabolic reactions would be too slow [5]. Enzymes present a higher promoter ability than many inorganic catalysts [6], accelerating reactions by factors of a million or even more [7].

As shown in Figure 2.1, the catalyst is able to reduce the magnitude of the activation energy barrier. One of the main characteristics of the catalytic activity of enzymes is that they provide a certain environment for reactions to take place: the active center. In this area of the enzyme, the side chains of the amino acids allow the temporary immobilization of the substrate and, at the same time, the completion of the reaction, releasing the final transformed product. The ability of enzymes to place reactants (substrates and enzymatic cofactors) in favorable orientations inside this catalytic center leads to the formation of the enzyme-substrate complex *ES*, which favors the transition state.

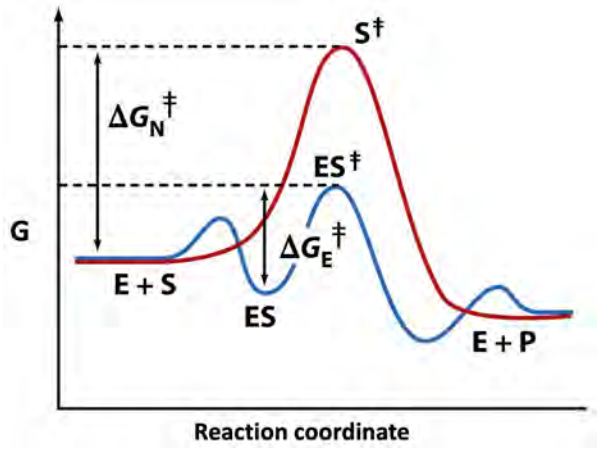


Figure 2.1. Enzyme-catalyzed reaction. *Scheme representing enzyme's role in the reduction of activation energy (ΔG‡) in the conversion reactions in which they participate* [8].

The current use of enzymes as biocatalysts has expanded to a large number of industries such as the pharmaceutical industry (active principle intermediates), energy industry (fermentable sugars for biofuels production), food industry (food processing and supplements), chemical industry (synthesis of organic compounds) and so on [9]. The most used groups of enzymes are the hydrolases (in up to 60% of biotransformations) and the oxidoreductases (in a 20% of the cases), being proteases, lipases, amylases and amidases the enzymes most widely used in industry [1].

2.2. CAZy classification and CE4 family

Carbohydrates are widely distributed in nature and play major structural and metabolic roles in almost every biological system. Their biological functions can be divided into four main categories: (1) *structural properties*, participating in the assembly of complex multicellular organs and organisms as structural component of cell walls; (2) *energy storage*, acting as energy sources for cells and living organisms; (3) *mediation of intra- or intercellular recognition*, modulating events in cell-cell, cell-matrix and cell-multicellular organism interactions - within the same organism or between different organisms (host-parasite, host-pathogen or host-symbiont); and (4) *specific*

recognition by other molecules such as glycan-binding proteins (GBPs), being central elements in several biological recognition processes, such as cell adhesion and signaling, cancer progression or immune response [10]–[15].

Glycobiology is an interdisciplinary field that studies the structure, biosynthesis, biology, evolution and function of carbohydrates and of the proteins that recognize them [16], [17]. There is a need in the field for synthetic approaches that render structurally defined oligosaccharides and glycoconjugates and, considering enzymes' selectivity and specificity, enzymatic synthesis can offer many advantages over chemical approaches [18].

The set of enzymes involved in the assembly, modification and breakdown of carbohydrates and glycoconjugates has been designated as **Carbohydrate-Active enZymes** (**CAZymes**) [17], [19]. The variety of monosaccharide structures and intersugar linkages is abundant and, considering all types of molecules that can be glycosylated (nucleic acids, proteins, lipids, sugar themselves, metabolites, etc.), the diversity of enzymes acting on these glycoconjugates is equally rich. The **CAZy database** (http://www.cazy.org) is a curated database that classifies and describes the families of structurally related catalytic and carbohydrate-binding modules of CAZymes. The families are defined based on significant amino acid sequence similarity and common structural folds and the classification is created module by module, including only protein sequences released in GenBank (http://www.ncbi.nlm.nih.gov/genbank).

Nowadays, there are five major classes in the CAZy classification:

1. **Glycoside Hydrolases** (GHs), that hydrolyze and/or rearrange glycosidic bonds.
2. **Glycosyltransferases** (GTs), that form glycosidic bonds.
3. **Polysaccharide Lyases** (PLs), that catalyze the non-hydrolytic cleavage of glycosidic bonds.
4. **Carbohydrate Esterases** (CEs), that hydrolyze carbohydrate esters.
5. **Auxiliary Activities** (AAs), that include redox enzymes that act in conjunction with other CAZymes.

Carbohydrate-active enzymes often display a modular architecture with non-catalytic modules appended to the catalytic domain of enzyme type classified above and CAZy also includes information on these associated modules:

- **Carbohydrate-Binding Modules** (**CBMs**), that participate in the adhesion to carbohydrates.

The enzymes of interest for this project are included in the CE class which, up to date (June 2020), includes 16 [1] families of enzymes that catalyze the de-*O* or de-*N*-acylation of substituted saccharides. Specifically, the CE4 family (Carbohydrate Esterase Family 4), includes different activities depending on the reaction catalyzed by the enzymes and the substrates on which they act:

- Acetyl xylan esterase (EC 3.1.1.72)
- Chitin deacetylase (EC 3.5.1.41)
- Chitooligosaccharide deacetylase (EC 3.5.1.-)
- Peptidoglycan GlcNAc deacetylase (EC 3.5.1.-)
- Peptidoglycan N-acetylmuramic acid deacetyase (EC 3.5.1.-)

Almost all the sequences in this family are from bacteria (27995), but there is also an important number of sequences that can be found in Eukaryota (600, mainly fungi and insects), with the remaining few members divided between Archea (17) and some open reading frames from viruses (2) (http://www.cazy.org).

The activity of these enzymes is mainly on lineal polysaccharides or oligosaccharides and different members of the family follow distinct modes of action, thus showing diverse deacetylation patterns inside the substrate's chain. These enzymatic action patterns may be divided into three main types, designated multiple-attack, multiple-chain, and single-chain/single site mechanisms (Figure 2.2) [20]:

(A) **Multiple-attack mechanism**, in which, after binding to the polysaccharide chain, the processive enzyme performs sequential deacetylations before dissociating and binding to another region of the chain.
(B) **Multiple-chain mechanism,** in which the enzyme can form more than one active complex with the substrate and the deacetylation of only one residue of the polymeric chain takes place in each binding event.
(C) In the case of CDAs and CODs, there is a third mode of action, the **single-site mechanism**, that covers the few enzymes with specificity for deacetylating a single position in their substrates.

[1]Family CE-10 has been withdrawn because the majority of its members were esterases active on non-carbohydrate substrates. Thus, there are 15 effective CE families [27].

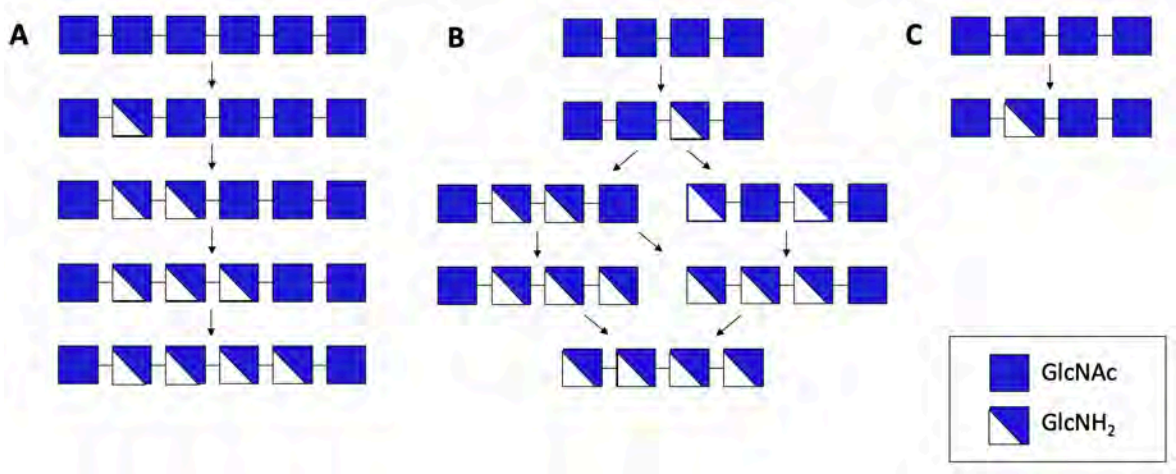


Figure 2.2. Modes of enzymatic action patterns for CE4 enzymes. *A) Multiple-attack mechanism; B) multiple-chain mechanism and C) single-site mechanism.*

Provided that these enzymes act on a broad variety of substrates following different modes of action, their functions in the different organisms that produce them and, consequently, their range of biological and biotechnological applications are also very diverse.

This family of enzymes share a highly conserved region designated as the "NodB homology domain" or "polysaccharide deacetylase domain" of approximately 150 amino acids [20]. NodB is a rhizobial nodulation protein from the symbiotic bacteria of the genus *Rhizobium*. *Rhizobium* species synthetize Nod factors (NFs), lipo-chitooligosaccharide molecules that act as morphogenic signals in legume hosts. The protein is part of the nodABC operon that encodes for three enzymes that participate in the synthesis of the NFs core [21]. In particular, NodB is a chitooligosaccharide deacetylase that deacetylates the residue at the non-reducing end of chitooligosaccharides during NFs synthesis and it was one of the first CE4 enzymes to be characterized [22].

Although some members of the family are monodomain proteins, with their sequence consisting of a single NodB homology domain, some others have a multidomain organization with the presence of other domains. They are mainly domains with unknown function and carbohydrate binding modules (CBMs), with the function of facilitating the solubilization and recognition of substrates or locating the enzyme in the proper cellular compartment [23].

CE4 esterases contain a single divalent metal ion coordinated by conserved residues in the active site and they were proposed to follow a metal-assisted general acid/base catalysis. The publication of the first crystal structure in complex with substrates in a productive manner (CDA from *Vibrio*

cholera in complex with $GlcNAc_2$) [24] allowed to describe the action mechanism of this family of enzymes. First, the general base residue activates a water-bound molecule to form a tetrahedral oxyanion intermediate, which is stabilized by metal coordination and by other residues from the catalytic cleft. Then, the nitrogen group of the intermediate is protonated by the general acid residue to assist C-N bond breaking, generating a free amine in the de-*N*-acetylated product and releasing acetate as a sub-product [24] (Figure 2.3).

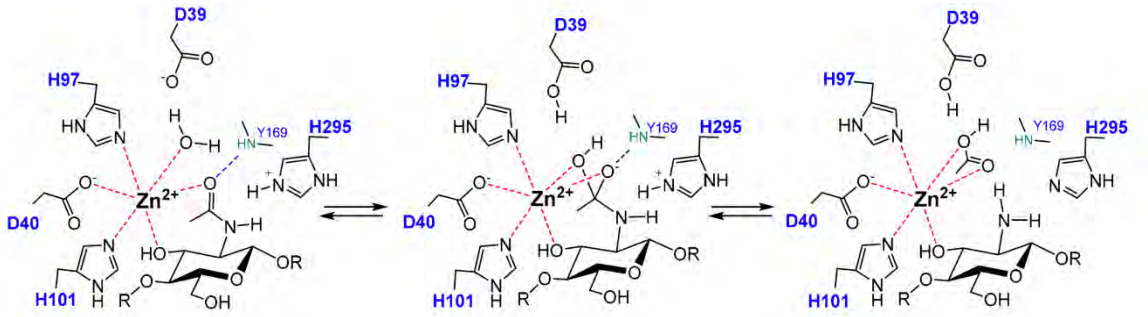


Figure 2.3. Metal-assisted general acid/base mechanism proposed for CE4 deacetylases. *Scheme based on the 3D structure of the enzyme·substrate complex VcCDA-D39S·DP2. D39 is the general base and H295 is the general acid* [24].

2.2.1. Sequence and structural features

The first crystal structure of a CE4 family member was from *Bacillus subtilis* peptidoglycan deacetylase (*Bs*PdaA) and it showed a $(\beta/\alpha)_8$ barrel fold, conserved in most CE4 enzymes' structures that have been later solved (see list of CE4 enzymes with known structure in Table 2.1). This fold is frequently distorted in this family of enzymes, where it can even appear as an $(\beta/\alpha)_7$ barrel, lacking one of the repeats of a typical TIM barrel. A series of loops decorate the ß-barrel and form the majority of the carbohydrate binding pocket [25]–[27]. Another particular feature shared by CE4 members, as mentioned previously, is the presence of a single divalent metal ion (typically Zn^{2+} or Co^{2+}) in their active site, coordinated by a His-His-Asp triad, and the metal ion-dependency of their activity.

Table 2.1. CE4 enzymes with known structure. *Retrieved from CAZy database (www.cazy.org) and from bibliographic research. REF.) 3D structure publication, if available. Table updated on June 2020, without including the enzymes related to the work of this thesis.*

ACTIVITY	ORGANISM	ENZYME/GENE	PDB (YEAR)	REF.
Peptidoglycan deacetylase	*Bacillus anthracis* (B)	BA0330 (a)	4V33 (2015)	[28]
	Bacillus anthracis (B)	BA0150 (a)	4M1B (2014)	[29]
	Bacillus anthracis (B)	BA3943* (a)	6HM9 (2019)	-
	Bacillus cereus (B)	BC0361	4HD5 (2012)	[30]
	Bacillus cereus (B)	*Bc*Pgd	4L1G (2014)	[31]
	Bacillus cereus (B)	BC1974	5NCD (2018)	[32]
	Eubacterium rectale (B)	*Er*Pgd	5JMU (2016)	-
	Streptococcus pneumoniae (B)	*Sp*PgdA	2C1G (2005)	[25]
	Streptococcus mutants (B)	*Sm*PgdA	2W3Z (2008)	[33]
	Bacillus subtilis (B)	*Bs*PdaA	1NY1 (2002)	[34]
	Bacillus anthracis (B)	*Ba*Pda	2J13 (2006)	[35]
Poly-ß-1,6-N-acetyl-D-glucosamine deacetylase	*Aggregatibacter actinomycetemcomitans* HK1651 (B)	*Aa*PgaB (a)	4U10 (2015)	-
	Ammonifex degensii (B)	IcaB	4WCJ (2014)	[36]
	Bordetella bronchiseptica (B)	Bbpb	5BU6 (2015)	[37]
	Escherichia coli (B)	*Ec*PgaB	3VUS (2012)	[38], [39]
Acetyl-xylan esterase	*Clostridium thermocellum* (B)	*Ct*CE4	2C71 (2006)	[40]
	Streptomyces lividans (B)	*Sl*AxeA	2CC0 (2006)	[40]
Chitin deacetylase	*Aspergillus nidulans* (E)	*An*CDA	2Y8U (2012)	[41]
	Colletotrichum lindemuthianum (E)	*Cl*CDA	2IW0 (2006)	[42]
	Arthrobacter sp. (B)	*Ar*CE4	5LFZ (2017)	[43]
	Bombyx mori (E)	*Bm*CDA1	5ZNT (2019)	[44]
	Bombyx mori (E)	*Bm*CDA8	5Z34 (2019)	-
Chitooligosaccharide deacetylase	*Vibrio cholerae* (B)	*Vc*COD	4NY2 (2014)	[24]
	Vibrio parahaemolyticus (B)	*Vp*COD	3WX7 (2014)	[45]
Putative polysaccharide deacetylase	*Burkholderia pseudomallei* (B)	BPSL2118	3S6O (2011)	-
	Pseudomonas aeruginosa PAO1 (B)	PA1517	1Z7A (2005)	-
	Encephalitozoon cuniculi (E)	ECU11_0510	2VYO (2008)	[46]
	Streptococcus pyrogenes (B)	BAM30204.1	6DQ3 (2019)	-

When the 3D structure of *Streptococcus pneumoniae* peptidoglycan deacetylase (*Sp*PgdA) was solved, Blair et *al*. first defined five highly conserved motifs (MT 1 to 5 according to their order of appearance in the proteins' sequences) which were described to assemble the active site [25]. The latest publication of several crystal structures has allowed the redefinition of the description of these consensus motifs [47]. Each one is defined by a short sequence that includes certain amino acids of special relevance that are generally related to catalytic activity (Figure 2.4):

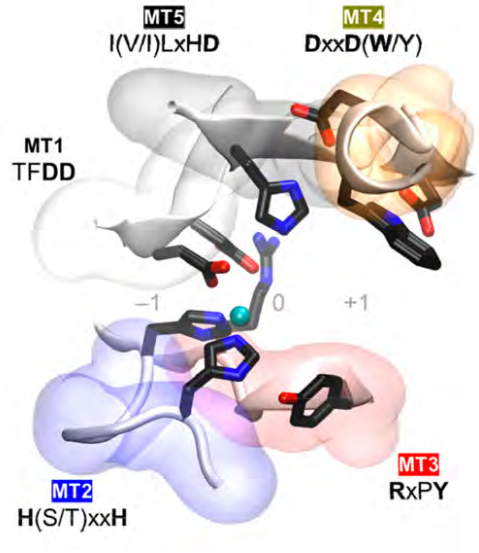


Figure 2.4. Conserved catalytic motifs MT1 to MT5 of CE4 family. *Spatial disposition in the 3D active site structure of CE4 family* [47].

- **Motif 1** (MT1): *TFDDG*, includes the general base aspartate (first D) and the aspartate that is part of the complex with the catalytic metal (second D).
- **Motif 2** (MT2): *H(S/T)XXHP*, in which the two histidine residues are part of the metal-binding system and the serine or threonine residue is an hydrogen bond acceptor that stabilize the structure of the loop-shaped motif in this region. These two histidine residues and the second aspartate of motif 1 form the characteristic His-His-Asp metal-binding triad of CE4 proteins.
- **Motif 3** (MT3): *RXPY*, that is shaping one of the sides of the active site groove, forming stabilizing interactions in this area of the structure.
- **Motif 4** (MT4): *DXXD(W/Y)*, that is shaping the other side of the active site groove and includes a hydrophobic residue exposed to the solvent and an aspartate buried in the structure.
- **Motif 5** (MT5): *I(V/I)LXHD*, includes the catalytic general acid histidine and a leucine that is in the hydrophobic patch that accommodates the acetate methyl group.

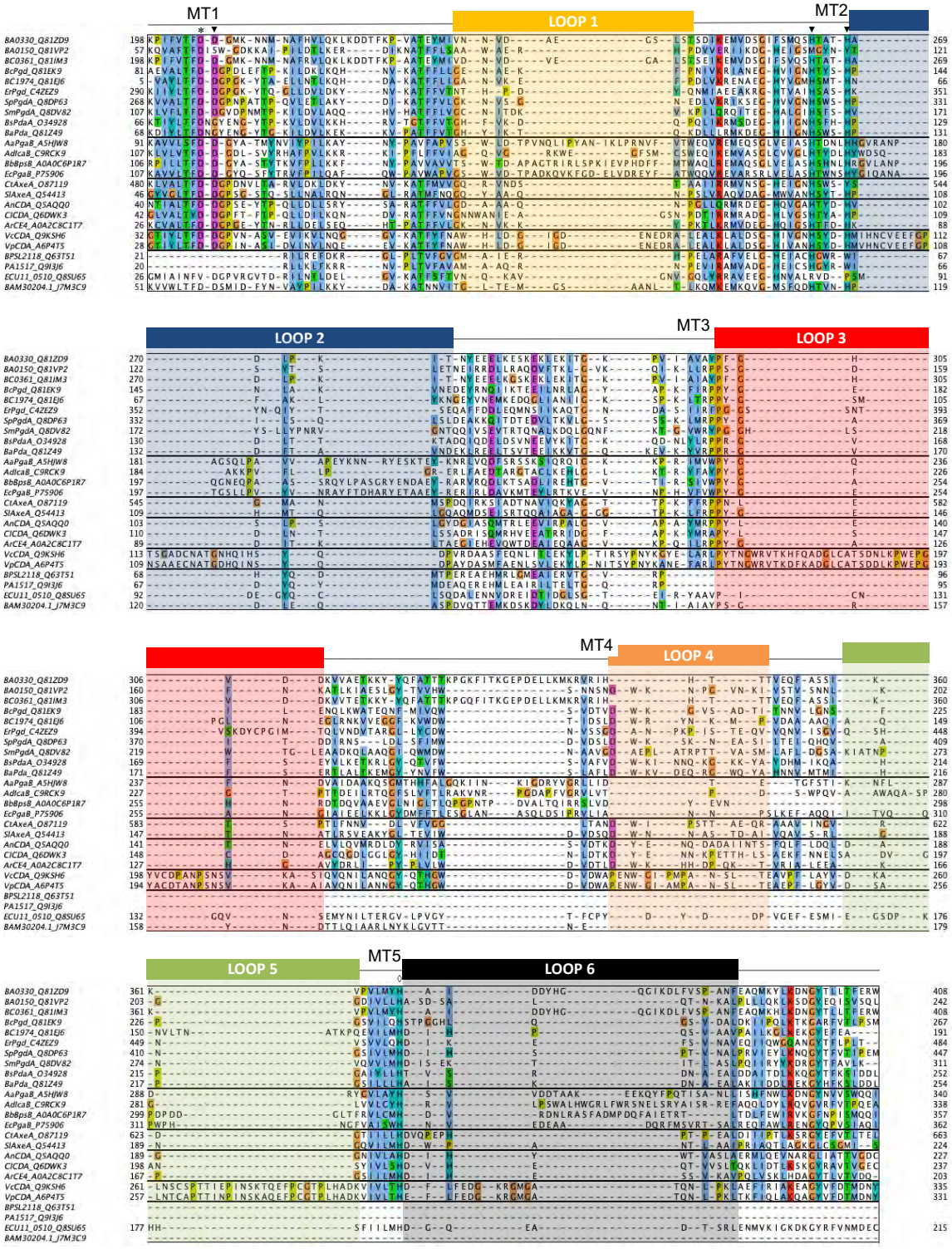


***Figure 2.5*. Multiple sequence alignment of CE4 enzymes.** *Loops are highlighted with colored boxes according to* [24]. *Conserved catalytic motifs are labelled MT1-5. "His-His-Asp" metal binding triad (▼), catalytic base (✶), and catalytic acid (◊) are labelled. The alignment includes sequences of the enzymes listed on Table 2.1, except for BmCDA1 and BmCDA8 from the insect Bombyx more, which were added later to the list and show some differential structural features to other CE4 members* [44].

A multiple sequence alignment of CE4 enzymes with known crystal structure obtained by our group is presented in Figure 2.5. Using standard multiple sequence alignment algorithms based only on sequence composition, does not achieve a proper alignment of the conserved domain of this family of enzymes. Considering that several proteins are multimeric and that there are important differences in sequence length between them, it is necessary to use the structural information in order to reproduce the structural conservation of CE4 family. Therefore, the alignment was guided by the structural superposition of the available 3D structures of these enzymes and manually curated to represent only the catalytic NodB domain. Sequences of enzymes without x-ray structure were added by means of Hidden Markov Model (HMM) comparisons [48].

The structural alignment shows the conserved motifs described above (MT1 to MT5), located at the center of the active site structure, and several non-conserved insertions distributed along the catalytic NodB domain. Both sequence content and length of these non-conserved regions are highly variable between the different enzymes and, when analyzed in the structures, correspond to several loops with different geometry that surround the active site and that are numbered from 1 to 6 (according to their appearance in the sequence) in the multiple alignment. Considering that the rest of the proteins' catalytic domain structure is highly conserved, but they render products with different patterns of acetylation, we proposed these variable loops to be key elements in determining substrate specificity and deacetylation mode of different CE4 enzymes [24], [47].

2.3. Peptidoglycan deacetylases

Peptidoglycan deacetylases can be classified depending on the sugar unit of peptidoglycan that they deacetylate. **Peptidoglycan GlcNAc deacetylases** act on the N-acetylglucosamine units of peptidoglycan and **peptidoglycan MurNAc deacetylases** catalyze the hydrolysis of the acetamido group of N-acetylmuramic acid residues (Figure 2.6).

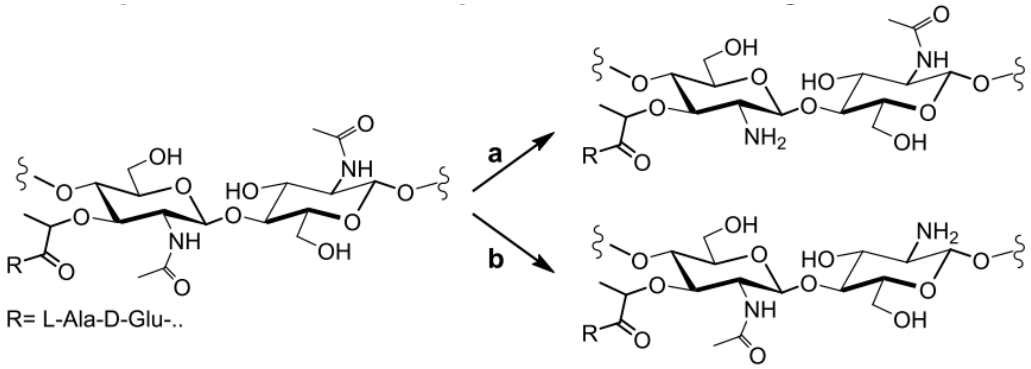


***Figure 2.6**. **Deacetylation reaction by peptidoglycan deacetylases.** Deacetylation of a peptidoglycan fragment by the action of (a) peptidoglycan MurNAc deacetylases or (b) peptidoglycan GlcNAc deacetylases.*

The crystal structures of several peptidoglycan GlcNAc deacetylases have been solved and some of them have been characterized. In general, this subfamily of enzymes is specific for GlcNAc residues on peptidoglycan, but it also shows activity on COS with lower efficiency. In the case of peptidoglycan MurNAc deacetylation, the activity has been annotated to several putative polysaccharide deacetylases based on genetic knock-out studies. However, there are only two available 3D structures of peptidoglycan MurNAc deacetylases and the only biochemically characterized enzyme is PdaA from *B. subtilis*. This enzyme deacetylates a modified peptidoglycan substrate in which MurNAc residues lack the peptidyl substitutions and, contrary to known peptidoglycan GlcNAc deacetylases, it is not active on natural peptidoglycan [34], [49]. Another remarkable difference is that the enzyme does not show activity on COS, being specific for MurNAc residues. Regarding structural features, peptidoglycan MurNAc deacetylases are the only subgroup within the CE4 family that lack the otherwise conserved Asp residue of the metal binding catalytic triad. In the same position, they show an Asn residue which points away from the active metal ion into the core of the protein [34], [35].

Non-acetylated glucosamine residues (GlcN) were found in high proportion in a lysozyme-resistant *Bacillus cereus* strain in 1971, when *N*-deacetylation was first identified [50]. Non-acetylated muramic acid residues (MurN) were later identified in *Bacillus anthracis'* peptidoglycan [51]. Pathogenic bacteria use this cell wall modifications, as well as acetylation (6-*O*-acetylation of MurNAc), as a strategy to evade detection by the immune system of the host. Lysozyme, a muramidase that hydrolises the MurNAc-β-1,4-GlcNAc glycosidic bonds of the glycan strands in peptidoglycan, is an important factor of the innate immune system in humans [52], [53]. Intact peptidoglycan and peptidoglycan fragments generated by lysozyme and other lytic enzymes are recognized by the innate immune system via peptidoglycan recognition proteins, Toll-like receptors, and Nod-like receptors [52]. The degradation of peptidoglycan drives immune response to infection and promotes an increased duration and intensity of this response. Modifications of peptidoglycan's structure by bacterial enzymes like these deacetylases makes it more resistant to lysozyme, consequently evading the antibacterial activity of lysozyme and delaying or suppressing pro-inflammatory immune responses [52].

MurNAc deacetylation on bacterial peptidoglycan has been related with endospore formation and germination. *Bacillus* species present the ability to form spores to survive harsh environmental conditions [54] and these spores have a thick peptidoglycan cortex with a high content of spore-specific muramic acid δ-lactam. Peptidoglycan MurNAc deacetylases are involved in autolysis

during spore germination, carrying out both deacetylation of MurNAc residues and lactam ring formation. First, peptide from the 3-*O*-lactoyl group is removed by the amidase CwlD (muramoyl-L-alanine amidase) and then, residues of the glycan chain are deacetylated by the MurNAc deacetylase [53], [55]. Expression of ClwD in *E. coli* does not result in muramic δ-lactam formation, and expression of the MurNAc deacetylase PdaA alone in the same host does not show any effect *on E. coli* peptidoglycan structure, whereas the introduction of both enzymes leads to muramic δ-lactam formation. Although it has been shown possible to artificially produce this structure in *E. coli*, in vitro activity of PdaA is limited to the deacetylation of MurNAc residues [53].

2.3.1. Substrate: peptidoglycan

Peptidoglycan, also known as murein, is a heteropolymer of long glycan strands crosslinked by short peptides [56]. The glycan chains consist on alternating N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) residues linked by β-1,4 bonds and are formed by oligomerization of monomeric disaccharide peptides units (lipid II) by transglycosylation reactions [57].

One source of diversity in peptidoglycan structure is the variation of the peptide moiety, being its amino acid composition the basis for the peptidoglycan bacterial classification proposed by Schleifer and Kandler [58]. These crosslinking peptides (stem peptides) are attached to the lactoyl groups of MurNAc by an amide peptide (Figure 2.7). They are formed by five amino acids and show some interesting characteristics, such as the presence of alternating D- and L-isomers and non-proteogenic amino acids. In Gram-negative bacteria, the sequence of the stem peptide is L-Ala–D-*iGlu*–*m*-DAP–D-Ala–D-Ala pentapeptide, where *iGlu* is *iso*-glutamate and *m*-DAP corresponds to *meso*-diamopimelic acid. In contrast, in the stem peptide of Gram-positive bacteria the D-iGlu residue is often amidated resulting in D-*i*Gln, L-Lys replaces *m*-DAP in the third position [59], and it has been described the formation of cyclic imides [59], [60]. Crosslinking of the strands is generally between the carboxyl group of D-Ala in position 4 and the amino group of the diamino acid in position 3, although there are some variations between species of bacteria, being these interpeptide bridges another source of diversity [57]. Besides these inter-species differences, there can be variations in the fine structure in the same species, according to the growth conditions of the microorganisms (growth phase, medium composition, intra- or extra-cellular growth, presence of antibiotics) [57].

Figure 2.7. Examples of stem peptides and modes of crosslinking [61]**.**

Peptidoglycan is a major component of the bacterial cell wall and forms a flexible, net-like structure that determines cell shape and serves as scaffold for anchoring other cell wall components [62], [63] and provides mechanical strength and osmotic stability to the cell, preserving its integrity. It is intimately involved in cell growth and cell division processes and any inhibition of its biosynthesis or its degradation during cell growth will result in cell lysis.

Post-synthetic variations of glycan strands such as O-acetylation, O-glycosylation and de-*N*-acetylation of the GlcNAc and/or MurNAc residues represent a strategy for pathogenic bacteria to evade the innate immune system of their host and to control autolysins [61], [64].

2.3.1.1. Peptidoglycan recognition by the host immune system

Cell wall peptidoglycan is constantly being remodeled to allow growth and cell division and its complexity results from the action of different enzymes including glycosyltransferases, transpeptidases and hydrolases (glycosidases, amylases, carboxypeptidases, endopeptidases), among others [65]. Since it is exclusively present in almost all bacterial cell walls, peptidoglycan acts as an excellent pathogen associated molecular pattern (PAMP) to detect bacteria via eukaryotic pattern recognition receptors (PRR) in pathogenic infections [52], [66].

Lysozyme is an important factor of the innate immune system in humans and it is present in many tissues and body fluids [52], [53]. This enzyme is secreted in large amounts by cells of the immune system at the site of infection and it originates (together with other lytic enzymes) muropeptides, small peptidoglycan fragments which are elicitor molecules of the immune response. The main recognition mechanisms of intact peptidoglycan and muropeptides include peptidoglycan recognition proteins, Toll-like receptors, and Nod-like receptors [52].

- **Peptidoglycan recognition proteins (PGRPs):** these soluble proteins are produced by mammals and they bind peptidoglycan in bacterial cell walls by recognition of the stem peptide. PGRPs present antibacterial activity and induce cell lysis [67], [68].
- **Toll-like receptors (TLRs):** the TRL family is a relatively homogenous family of transmembrane proteins composed of 10 – 12 members in most mammals. They are located in the plasma or the endosomal membranes of many types of cells, where they detect exposed ligands. It has been reported that degradation of peptidoglycan in phagosomes releases ligands that activate TLR2 and TLR9 [52].
- **NOD-like receptors (NLRs):** NLRs represent the larger group of intracellular PRR and they share a conserved structure consisting of a N-terminal protein interaction domain, a central NOD domain and a C-terminal leucine-rich repeat (LRR) domain with varying length [52]. Several NLRs can sense small molecules derived from peptidoglycan and induce the secretion of pro-inflammatory cytokines as well as increasing antimicrobial responses, such as the production of reactive oxygen species, nitrogen species and antimicrobial peptides. *N*-acetylmuramyl dipeptide (MDP) is the minimal peptidoglycan fragment to directly interact with NOD2 receptors, ultimately resulting in activation of inflammatory response via the NF-κB and MAP kinase signaling pathways [69].

Bacteria that are able to modify the structure of peptidoglycan to make it more resistant to lysozyme, both evade the antibacterial activity of lysozyme and delay or suppress pro-inflammatory immune responses [52]. Besides protecting peptidoglycan from lysis from exogenous host-defense agents, its modifications also protect it from lysis by the endogenous autolysins involved in its metabolism [64].

These post-synthetic modifications are generally limited to the C2 amine or the C6 hydroxyl moieties of the amino sugars (Figure 2.8), being *N*-deacetylation and *O*-acetylation two of the main ones: (i) *N*-deacetylation of either MurNAc or GlcNAc residues can be mostly found in gram-positive bacteria and its biological role has been related mainly to lysozyme resistance. Recent studies on *Bacillus* species have reported possible physiological functions of this modification, related to PGN

metabolism. However, further research is needed to establish these functions. (ii) All gram-positive and most gram-negative bacteria *O*-acetylate their peptidoglycan, modification that has been related to pathogenicity and it is determinant of lysozyme resistance. In gram negative bacteria, *O*-acetylation is involved on peptidoglycan chain length regulation and on the control of autolysins [64].

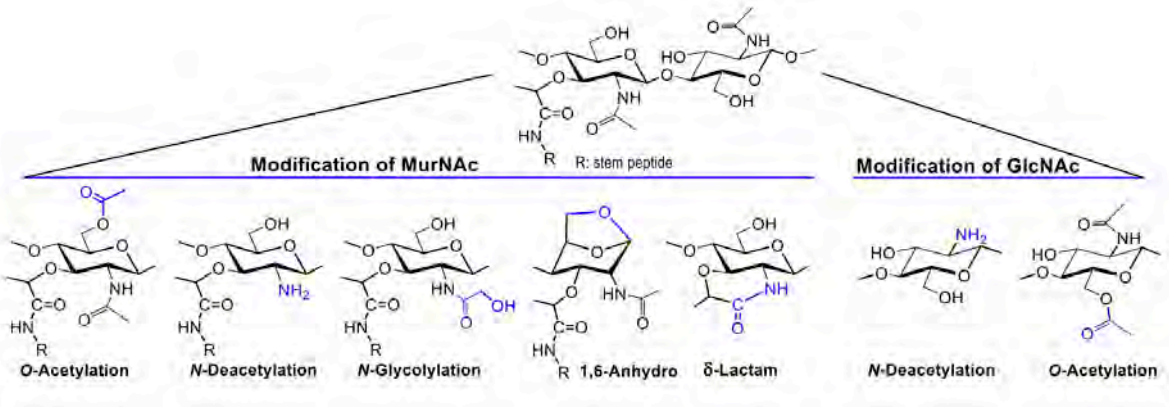


Figure 2.8. Post-synthetic modifications of MurNAc and GlcNAc on glycan strands of peptidoglycan*. Modifications are colored in blue. Adapted from* [64].

2.4. Chitin deacetylases

Chitin deacetylases (**CDAs**) are enzymes that catalyze the hydrolysis of the acetamido group of *N*-acetyl-D-glucosamine monomers of chitin oligomers and polymers, generating the derivatives chitosan and chitooligosaccharides (Figure 2.9).

As a consequence of chitin being one of the most abundant natural polymers, CDAs are present in very diverse organisms performing multiple functions [70]. They have been found mainly in fungi and bacteria, but have also been reported in archaea and insects [71]. Fungal CDAs are involved in fungal nutrition, morphogenesis and development [20], [72], participating in cell wall formation and integrity [73], in spore formation [74], germling adhesion [75], fungal autolysin [76] and in defense mechanisms for host infection [77]. In bacteria, the predominant deacetylases are chitin oligosaccharide deacetylases (also known as CODs), active on low molecular mass COS and essentially inactive on polymeric chitin and chitosan and they are involved in the biosynthesis of Nod factors (in the case of *Sm*NodB) [78] or in the chitin degradation cascades occurring in sea water (in the case of CODs of marine bacteria such as *Vibrio* or *Shewanella* species) [79]–[81].

CDAs seem to be widely present also in insects. Although there is little knowledge on their function and properties, interest in these deacetylases is currently increasing and two structures of CDAs from the organism *Bombyz mori* have been recently reported [44]. These structures show a strong

conservation of the family active site and catalytic residues, suggesting that they follow the same catalytic mechanism. However, they present a much longer, wider and more open substrate binding cleft than other CDAs, as well as two unique loops regions that are a distinct architecture feature not present in other CE4 family members. The lack of steric constraints in the catalytic center may translate into a reduced effectiveness of trapping substrates after binding and can be related to the lower activity of these CDAs in comparison to bacterial ones [44].

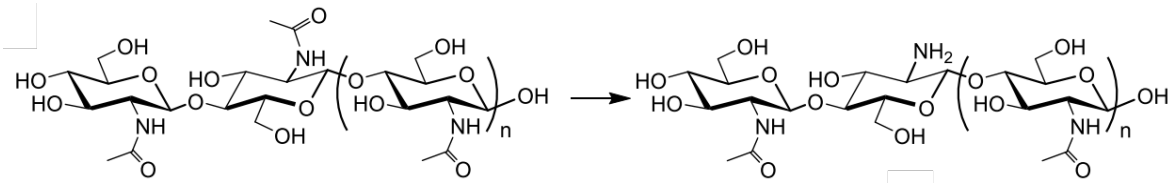


Figure 2.9. Deacetylation reaction by CDAs. *Deacetylation of chitin by the action of a chitin deacetylase resulting in chitosan production with the release of acetate and the formation of a free amino group.*

Just like their sources and functions, these enzymes are very diverse in their characteristics and working conditions, they cover a very extensive range of optimal conditions for performing their activities. Their isoelectric points (pIs) vary from 2.6 to 4.8, their molecular masses vary from 12 to 150 kDa and the optimum pH ranges from 4.5 to 12. One remarkable feature is that some CDAs are significantly thermostable, with optimum temperatures for activity between 30 and 60°C, some of them retaining a substantial percentage of activity after being at 90°C. Regarding their cellular location, most CDAs are secreted to the extracellular medium (secreted to alter the physicochemical properties of the cell wall) or the periplasm (in fungi, tightly coupled to a chitin synthase to rapidly deacetylate newly synthesized chitins before their maturation and crystallization). In bacteria, they can be intracellular (when involved in Nod factors biosynthesis) or extracellular (when involved in the catabolism of chitin).

Most CDAs are inactive on crystalline chitin and have a preference for soluble chitins, such as glycon-chitin or chitin oligomers or for chitosans (partially deacetylated chitin). In general, they present high specificity for polymeric or oligomeric substrates although the preferred DP can vary depending on each enzyme. The inactivity on insoluble chitin is due to the inaccessibility of the acetyl groups in the tightly packed structure. There are some CDAs that present carbohydrate binding modules (CBMs), in addition to their catalytic domain, that seem to increase accessibility of the substrate and enhance the deacetylase activity by the catalytic domain [82]. Moreover, it has been recently reported that oxidative cleavage of the surface polymer chains of crystalline chitin by lytic polysaccharide monooxygenases (LPMOs) greatly enhance CDA activity on this substrate [41].

2.4.1. Substrates: chitin, chitosan and chitooligosacharides

Chitin was first isolated from fungi extracts by H. Braconnot in 1811, when it was named “fungine” [83]. In 1823, when A. Odier obtained the pure polymer from the demineralization of crab shells, it was officially named “chitin” [72], although its chemical structure was not described until 1929 by A. Hoffman [84].

Chitosan was first discovered by C. Rouget in 1859 by partially deacetylating chitin when boiling it in presence of concentrated potassium hydroxide [85], but the polymer was not formally given its name until years later, in 1894, by Hoppe-Seyler [86]. Natural chitosan was first extracted from the cellular wall of the fungi *Phycomyced blakesleanus* by Kreger in 1954 [87] and, nowadays, it is known that several fungi species are chitosan producers [88].

Chitin is a linear, unbranched polysaccharide that consists of β-1,4 linked 2-acetamido-2-deoxy-D-glucose (N-acetyl-D-glucosamine, GlcNAc, A) units and it is one of the most abundant polymers in nature, being present in many taxa as the main structural component supporting the individual cell or the whole organism. It can be found in a broad variety of structures and tissues, such as fungal and diatom cell walls, exoskeleton of arthropods (insects and crustacean) and shell and nacre of mollusks. Moreover, chitin has recently been annotated *in silico* in deuterostomes and vertebrates as well [89]–[91].

This cellulose like polymer is a highly water insoluble molecule due to the intra- and intermolecular hydrogen bonds that the oxygen of the acetamido group forms with adjacent -NH or -OH groups. Chitin fibers form a pseudo crystalline structure and, depending on the source or organism used for its isolation, adopts three different conformations. In all forms, CO-NH hydrogen bonds are formed between neighboring chitin chains, but the production mechanism affects the orientation of the chitin fibrils towards each other:

- **α-chitin**: composed of anti-parallel orientated chains with inter- and intra-sheet hydrogen bonds, it is present mainly in crustaceans.
- **β-chitin**: composed of parallel orientated chains, only intra-sheet hydrogen bonds are found in its structure and it is present mainly in squids’ skeleton.

Water, alcohol and amines can be embedded between β-chitin’s sheets allowing intra-crystalline swelling without breaking hydrogen bonds. On the contrary, α-chitin sheets cannot be penetrated

by water or alcohol, being a more stable form in natural conditions and, hence, the most abundant polymorphic form found in nature [92].

- **γ-chitin**: composed of a mixture of parallel and anti-parallel sheets and present in some fungi.

Chitosan is the most important natural occurring chitin derivative and it is obtained from the deacetylation of the C2 bound N-acetyl group, resulting in 2-amino-2-deoxy-D-glucose (D-glucosamine, GlcN, D) units. Full or partial chitin deacetylation renders lineal copolymers consisting of GlcNAc and GlcN units. Consequently, the term "chitosans" is not used to refer to a unique chemical entity but to a family of compounds that differ in their composition depending on the chitin deacetylation process.

One of the several definitions used to discriminate chitin from chitosan polymers is that from G.A.F. Roberts, who classified them by their solubility and insolubility in 0.1 M acetic acid. Insoluble material in these conditions is named chitin, soluble material is defined as chitosan [93]. The positive charges of chitosan amino groups make it water soluble in slightly acidic conditions.

Chitin, with a high degree of acetylation (DA), is usually characterized based only on its degree of polymerization (DP) but chitosans are also classified by their degree of acetylation and pattern of acetylation (PA). Chitosan functions and corresponding applications can be very diverse depending on these three parameters:

- **DP:** is the number of monomeric units that define a macromolecule, polymer or oligomer. In the case of chitosans, their length can range from small oligomers to large polymers with up to thousands of sugar units.
- **DA:** is the relative number of GlcNAc units in the carbohydrate polymer or oligomer and ranges from 0% in fully deacetylated chitosans to 60% in chitosans that are barely soluble in aqueous solutions. Variations of this parameter influence both their chemical (solubility, charge, reactivity, flexibility, polymer conformation, viscosity, crystallinity, high surface area, porosity, tensile strength, conductivity and photoluminescence) and biological (biodegradability, biocompatibility, muco-adhesion, hemostatic, analgesic, adsorption enhancer, antimicrobial, anticholesterolemic and antioxidant) properties, that can vary with process conditions.

- **PA:** is the distribution of GlcNAc and GlcN units in the polymer or oligomer chain. This distribution of the acetylated units can be block wise, random or alternating. Recently, PA has gotten more attention regarding its influence on chitosan characteristics since it was shown to affect their biological activities [94].

The presence of free amino groups generated from the deacetylation process provides chitosans with positive charge, making them the only known natural polycationic polymers with the ability to interact with several negatively charged derivatives, such as synthetic anionic polymers, polysaccharides, proteins, dyes, phospholipids and cell membranes, enzymes, tumor cells, bacteria cell walls, proteins, and nucleic acids [95], [96]. They can also chelate a wide range of metal ions and the presence of not only these amino groups, but also hydroxyl groups, offers the opportunity to chemically and enzymatically modify their structures. In terms of their biological properties, chitosans are biodegradable and biocompatible as well as safe and non-toxic polymers [97].

The ubiquity of chitin and chitosan in nature in combination with this set of characteristics and properties, which can be further controlled by the DP and DA, make these polymers highly interesting for several industries and suitable for an extensive number of applications.

Besides de-*N*-acetylation, chitin processing by depolymerization renders a series of derivates that include **chitin and chitosan oligomers** (chitooligosaccharides, COS). The lower degree of polymerization of these biomolecules, with low viscosity and small molecular sizes, makes them water soluble and readily absorbed *in vivo* and confers them with different and additional biological functions and applications compared to the polymers.

2.4.1.1. Applications

Due to its insolubility in water, its poor reactivity and the difficulty to work with it (limiting factors for its utilization in living systems), **chitin** is not a suitable material for many applications and the main interest of this polymer is as a starting material for obtaining its derivatives. Nevertheless, chitin has been used to immobilize enzymes, whole cells and antibodies and for the treatment of industrial pollutants. Chitin can be processed as fibers, which have shown to be useful as absorbable sutures, wound-dressing materials and systems for controlled drug release. Furthermore, its physical and biological properties make it a useful supporting material for tissue engineering applications and as a component in hard tissue replacement materials in orthopedics and periodontal applications [98].

Being the main natural occurring chitin derivative and considering their properties, described above, **chitosans**, both polymers and **COS**, have recently gained attention for their potential applications in several fields. Their biological activities and, thus, their applications in these areas, depend on their DP and DA. Some of the most important applications are summarized in Table 2.2, specially highlighting the ones in the biomedical and pharmaceutical fields, in which chitosans and COS are of great interest due to their biocompatibility, biodegradability and non-toxicity.

Table 2.2. Main applications of chitosans and chitosan oligomers in several fields.

INDUSTRY	APPLICATIONS
Agriculture	- **Antifungal and antimicrobial activities**. Chitosans can be used as an alternative treatment to bactericides due to their antifungal and antimicrobial activities. It has been reported that the last is increased with higher DP and lower DA [99]. - **Eliciting activities: plant's immune system promotion**. Chitosans with high DP and also COS with DP $\geq$ 6 show resistance related activities in plant cells. They are recognized as PAMP (pathogen-associated molecular pattern) by PRRs (pattern recognition receptors) inducing signaling cascades, stimulating the innate immunity in plants and, consequently, increasing their resistance against pathogens [99]. - **Vegetal growth promotion** [92]**.**
Water treatment and residues	- As chelating agents, chitosans can be used to **remove** metal ions, dyes and other contaminations [100].
Food industry [101]	- **Dietary fibers**. - Chitosans can **lower cholesterol levels** by blocking the absorption of dietary fat, which also gives them an application for weight and body fat lost [102]. - **Additives.**
Cosmetics industry [92]	- **Skin humectants**. - **Treatment against acne**. - **Make up foundations**. - **Oral treatments** (dentifrice, gums).
Biomedical and pharmaceutical industries	
Biomaterials and tissue engineering	- **Hydrogels, films, fibers and sponges**. [92].

	- **Nanodevices** such as nanoparticles, nanocapsules and composites [101]. These nanoparticles can be loaded for drug delivery, bacterial quorum sensing inhibition or genetic material delivery [102]. Their positive charge allows these materials to form a strong electrostatic interaction with negatively molecules such as DNA or RNA. In order to design an effective gene delivery system based on chitosan, it is necessary to find the optimal parameters, including DA and DP of the polymers and also serum stability, pH and concentration of genetic material.
Wound healing and dressing	- Chitosan scaffolds with macro- and micro-sized pores [102] are used as treatment for wounds.
Immunological potential	- **COS** have shown promising results in promoting functions of the innate and adaptive immune system [103], [104] and can be used as adjuvants in vaccines [102].
Antioxidant potential	- It has been reported that **COS** have an inhibitory effect toward protein oxidation and DNA, they can be used to scavenge free radicals or inhibit the oxidative destruction caused by free radicals, being a possible antioxidative treatment for various oxidative disorders [102].
Antimicrobial potential [102]	
Antiallergy potential	- **COS** with a molecular weight ranging between 1 and 10 kDa significantly inhibited allergic reactions, being possible novel inhibitors against allergic reactions [102].
Antitumoral potential	- **COS** are studied as potential candidates for cancer treatment drugs due to their anti-angiogenic potential and suppression of cancer cell proliferation and metastasis generation [102].
Anti-inflamatory potential	- **COS** have shown to be capable of inhibiting the immune response [105] and are potential candidates capable of preventing or treating diverse chronic inflammations [102].

2.4.1.2. Production

On a commercial scale, the main source of chitin and chitosan are crustacean shells, principally of crabs and shrimps. Among many polysaccharides obtained from marine sources, **chitin** stands out for its availability and richness. Every year, approximately 100 billion tons of chitin are produced by

crustaceans, mollusks, insects, fungi and related organisms [106]. In crustaceans, chitin is part of a complex network with proteins, in which calcium carbonate deposits to form the rigid shell. The removal of these two major components is required for chitin isolation from shellfish: proteins by deproteinization and inorganic calcium carbonate by demineralization, being possible to carry out both methods using chemical or enzymatic treatments. An additional step of decolorization of the obtained product may be needed.

Currently, the most usual method to extract chitin is the chemical extraction. The first chemical deproteinization step is performed heterogeneously using chemicals which also depolymerize the polymer. NaOH is the preferential reagent and it is applied between 0.125 and 0.5M at varying temperatures, up to 160°C and duration, from minutes to days. The use of this alkaline solution results in partial deacetylation and hydrolysis of the chitin polymer, lowering its molecular weight. The chemical demineralization is generally performed with an acidic solution, being the preferential reagent dilute hydrochloric acid (HCl) (Figure 2.10). Demineralization treatments vary with the mineralization degree of each shell, extraction time, temperature, particle size, acid concentration and solute/solvent ratio. Conventionally, it can be accomplished using dilute HCl at different concentrations (up to 10% w/v) at room temperature using different incubation times (from 15 minutes to 48 hours). It was reported that high temperatures can accelerate the demineralization reaction and some reactions are carried out at higher temperatures [107]. For the discoloration treatment, 0.02% potassium permanganate at 60°C is used, which eliminates melanin and carotenoids. In some cases, sodium hypochlorite or hydrogen peroxide are used as alternative or complementary treatments for obtaining the final pure chitin.

Chitin can be converted into **chitosan** through enzymatic or chemical processes, being again the chemical conversion the most common due to its lower cost and suitability for mass production. Either acids or alkalis can be used to deacetylate chitin and obtain chitosan but, since glycosidic bonds are susceptible to acid, alkali deacetylation is more frequently used (Figure 2.10). This process can be performed heterogeneously using a concentrated solution of NaOH at high temperatures (130-150°C) during few hours, or homogeneously, dispersing chitin in concentrated NaOH at 25°C for at least 3 hours and in crushed ice around 0°C. With the heterogenous method, chitosan is produced as an insoluble product deacetylated up to 85-99% with an irregular distribution of GlcNAc and $GlcNH_2$ residues along the polymeric chain, and the homogeneous one results in a soluble chitosan with an average DA of 48-55%. Variations in chitosan preparation may result in changes of DA, distribution of acetyl groups, molecular weight and viscosity in solution among other characteristics. There are many parameters in the deacetylation reaction than can

impact the characteristics of the final product, therefore, a balance between the deacetylation process and the final properties of chitosan should be found.

For the production of **chitosan oligomers,** the most common chemical method is the hydrolysis of polymers that can be carried out with chemical or enzymatic agents, or with high energy impacts [108]. At an industrial level, acids such as HCl are usually used (Figure 2.10). The most frequent conditions are 35% HCl at 80°C for 1-2 hours, being the duration of the treatment, the temperature and the acid concentration parameters that may affect the size of the resulting product [109].

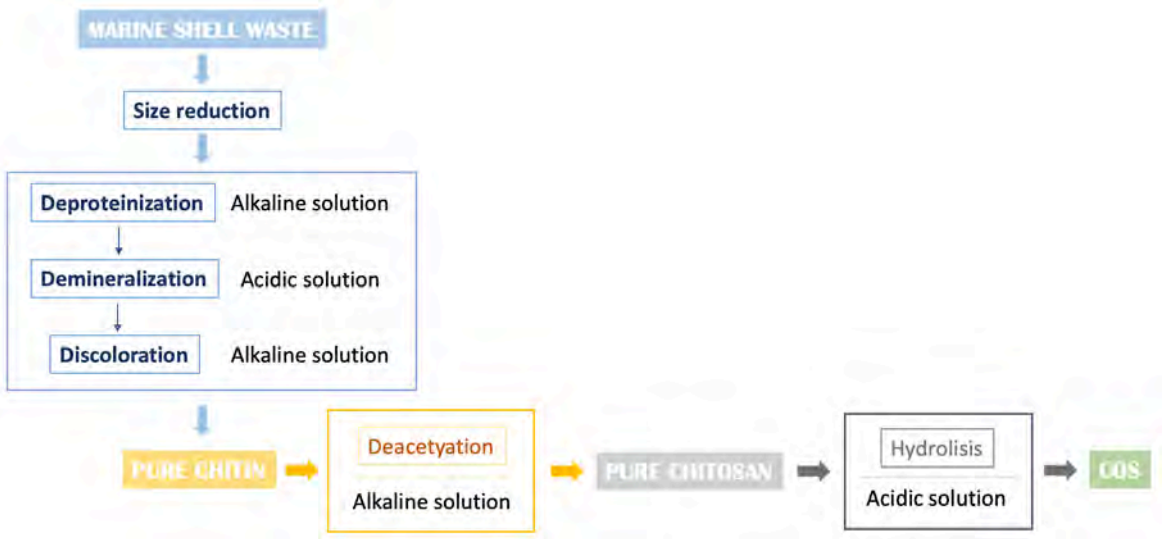


Figure 2.10. General scheme of industrial production of chitin, chitosan and COS from marine shell waste. *Adapted from* [90], [106].

As exposed, physicochemical characteristics of chitin and chitosan can be variable between batches due to seasonality of raw materials, quality of shells, climate, species, and process control. The traditional chemical methods have shown to be expensive, inconsistent in molecular weight and DA and environmentally unfriendly. Considering that DP and DA are defining parameters of the biological activities of chitosans and that PA is shifting more towards the focus of structure-function relationship, chitosans with a highly defined structure are in demand. In order to obtain this control, the alternative biotechnological production of chitosans is gaining importance.

2.4.2. Determinants of substrate specificity: The Subsite Capping Model

2.4.2.1. Defining the Subsite Capping Model

The Subsite Capping Model was first proposed in our group as a hypothesis to explain why CDAs, as a subfamily of CE4 enzymes, present differences in their substrate preference and generate diverse patterns of deacetylation and specificities.

In 2014, our group reported a series of crystal structures of the CDA of *Vibrio cholerae* (*Vc*CDA), which were the first 3D structures of a CE4 enzyme in productive complex with $(GlcNAc)_2$ and $(GlcNAc)_3$ substrates. This CDA is active on chitooligosaccharides with a DP from 2 to 4, being most active on DP2 and showing a reduction of the activity with the increase of the substrate's length [24].

These data provided first insight into structure-function relationship for this family of enzymes and highlighted the role of the 6 loops (already identified at a sequence level in the multiple alignment of Section 2.2) that shape the binding site cleft in substrate binding.

It was possible to start defining the subsites at the catalytic center of the enzyme, being subsite 0 the one in which the residue to be deacetylated is located, negative subsites the ones that are situated towards the non-reducing end and positive subsites the ones located to the reducing end.

It was observed that there is a movement of loop 4 upon substrate binding that changes the structure of the enzyme's active site from an open to a semi-closed or closed conformation depending on the length of the bound substrate (Figure 2.11). In the complex with DP2, the closed conformation leaves space for just two GlcNAc units that are located in the -1 and 0 subsites. In the complex with DP3, the semi-closed conformation of loop 4 allows the positioning of the substrate in the subsites -1, 0 and +1 (Figure 2.11). Negative subsites -2 and -3 are physically blocked by loops 1, 2 and 6 and, on the other site of the catalytic cavity, subsite +2 is also blocked by loop 5. However, previous experimental work showed that the enzyme presents activity on $(GlcNAc)_4$. Analyzing the available structures, it was observed that subsite +2 is necessary for the accommodation of this substrate, indicating a possible movement of loop 5 that would create new positive subsites needed for the catalysis of longer substrates. This dynamic behavior of loop 5 was confirmed initially by docking experiments performed with longer substrates and a bioinformatical model in which the

movement of the loop was forced. Furthermore, extensive engineering of this loop showed the importance of its dynamics for the catalysis of longer chitooligosaccharides [110].

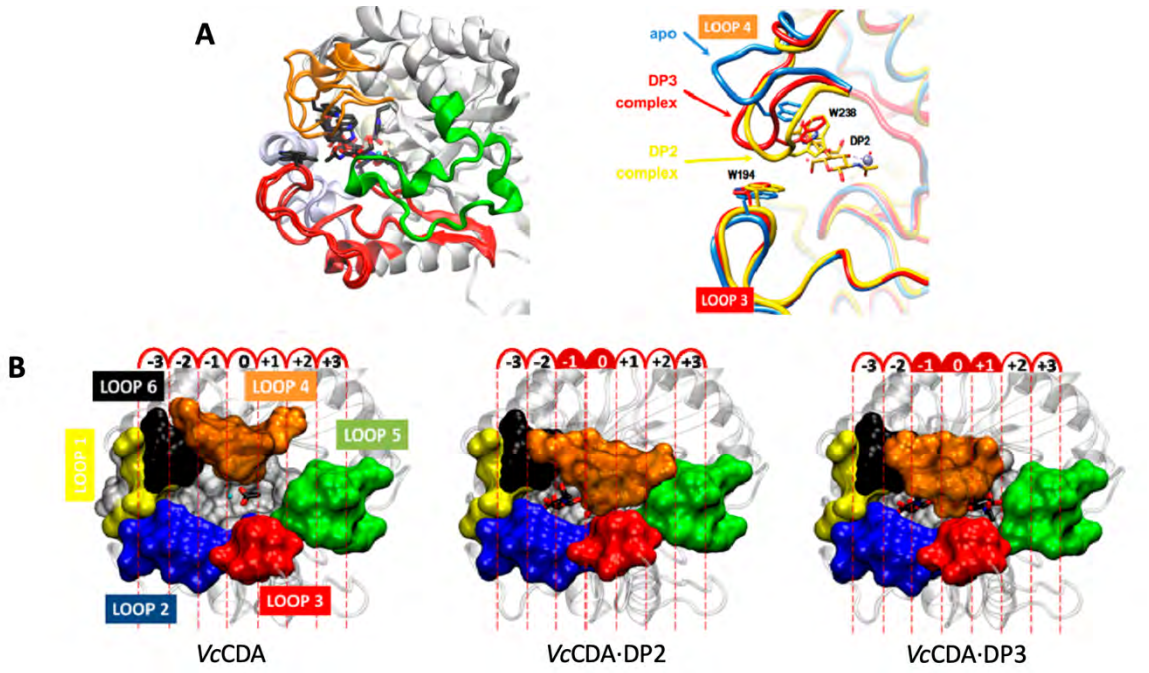


Figure 2.11 Crystallographic structures of VcCDA in the unligated form, in complex with $(GlcNAc)_2$ and in complex with $(GlcNAc)_3$ [24], [47]. A) *Superimposition of the three structures. Left: loop 4 (orange) has different conformations; right: magnification of the active site with loop 4 in the unligated form (blue), and in enzyme-substrate complexes with $(GlcNAc)_2$ (yellow) and $(GlcNAc)_3$ (red) ligands. Only DP2 ligand is shown. B) Overall structures. Loops 1 to 6 are colored as in Figure 2.5.*

Having identified the importance of the loops structure, shape and movement on the substrate specificity showed by *Vc*CDA, the possible extrapolation of this characteristic to other CE4 enzymes was studied. By comparing the different available sequences and structures at the moment, it was observed that all enzymes shared a very conserved core, but the loops showed high variability in sequence, length and shape. Adding this to the fact that, in spite of the highly conserved structure that they share, different CE4 enzymes render products with diverse patterns of acetylation, it is proposed that these variable loops are the possible determinants of their substrate preference and deacetylation mode.

Taking all the previous information into consideration, the definition of the Subsite Capping Model was based on the following key aspects:

- The presence of potential subsites blocked by the structure of the loops that surround the active site.
- The loops are not rigid, but dynamic structures.
- Loops displacement allows the exposition of the needed subsites for the accommodation of certain substrates.

- Different deacetylases show important differences in loops sequence, size and structure, decorating differently the active site of each enzyme.

These four points translate into the Subsite Capping Model statement: the pattern of deacetylation and substrate preference of CE4 enzymes are governed by specific loops and their dynamics.

2.4.2.2. Expanding the Subsite Capping Model

In recent years, some new CE4 enzymes have been characterized and some new 3D structures have been solved, allowing a deeper study of the structure-function relationship of substrate specificity in this proteins' family.

The structural superposition of all available structures of CE4 enzymes active on COS confirmed the significance of loop topology as a determinant of substrate binding specificity. The characteristic α/β barrel is highly conserved, variability being provided by a series of loops surrounding the active site that connect the α/β elements of the barrel. The difference in length, structure and sequence composition of these six loops is what modulates the shape of the binding cleft exposed to the substrate in each particular enzyme. In general, loops 1, 2 and 6 define the negative subsites of the binding cleft and loops 3, 4 and 5 define the positive ones, being the catalytic site located at the center of these loops (subsite 0) that differentially block the accessibility to additional substrate binding subsites.

There are, currently, only two available structures of CE4 enzymes in a productive complex with substrates. Aside from the described *Vc*CDA complexes, recently (in 2017) [43], the 3D structure of *Ar*CEA, a CDA from an *Arthrobacter* sp., was reported. In spite of both being chitin deacetylases and sharing the same molecular function, the binding site topology of these two enzymes is quite different and they represent two models of substrate specificity within the family.

(i) *Vc*CDA. Long Loops and High Specificity

As already described with the definition of the Subsite Capping Model, *Vc*CDA presents long loops that shape a rather closed and small binding cleft where the COS substrates are confined (Figure 2.12, A). These loops are capping both the reducing and non-reducing ends of the substrate, and this constricting topology has two main consequences. The enzyme is most active on the disaccharide and its catalytic efficiency drops substantially on longer oligomers, and the enzyme deacetylates exclusively the penultimate residue from the non-reducing end of the substrates.

There is no space for the substrate to slide along the active site, thus only one GlcNAc can be accommodated at the catalytic site (subsite 0) where deacetylation takes place.

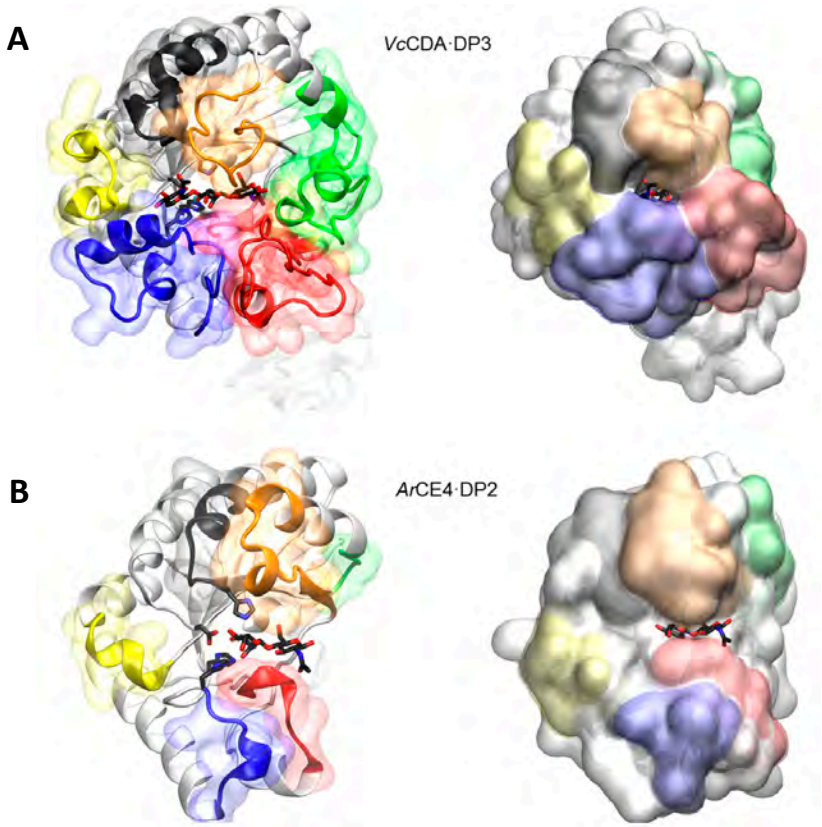


Figure 2.12. 3D structures of enzyme·substrate complex. *A) VcCDA with $(GlcNAc)_3$ substrate. B) ArCE4 with $(GlcNAc)_2$ substrate. Loops 1 to 6 are colored as in Figures 2.5 and 2.11.*

(ii) *Ar*CE4. Short Loops and Broad Specificity

Distinctively, *Ar*CE4's structure in complex with $(GlcNAc)_2$ shows the substrate bound to a more open binding cleft. Although the enzyme was co-crystalized with $(GlcNAc)_4$, only two units are observed in the structure indicating a weak interaction of part of the oligomeric chain with the flat protein surface topology. The main difference between the two model enzymes is the size and shape of the loops surrounding the active site, having both the catalytic center in the same position with respect to the protein core.

In *Ar*CE4's case, the short loops result in a much more open binding cleft that can accommodate longer substrates (Figure 2.12, B). Furthermore, the lack of protein caps at either ends of the substrates allows different binding modes, exposing different GlcNAc units of the oligomeric chain at the catalytic site. These observations are in agreement with the enzyme's properties, showing that the enzymatic activity of *Ar*CE4 increases with the length of the substrates and that it follows

a multiple-chain mechanism fully deacetylating the COS substrates except for the reducing-end units.

As shown in Figure 2.13, patterns of deacetylation differ among enzymes of the same group (with similar length and shape in loops), suggesting that surface charge distribution along the binding cleft and other structural features, yet to be disclosed, may also participate in defining the mode of action and deacetylation pattern of each particular enzyme.

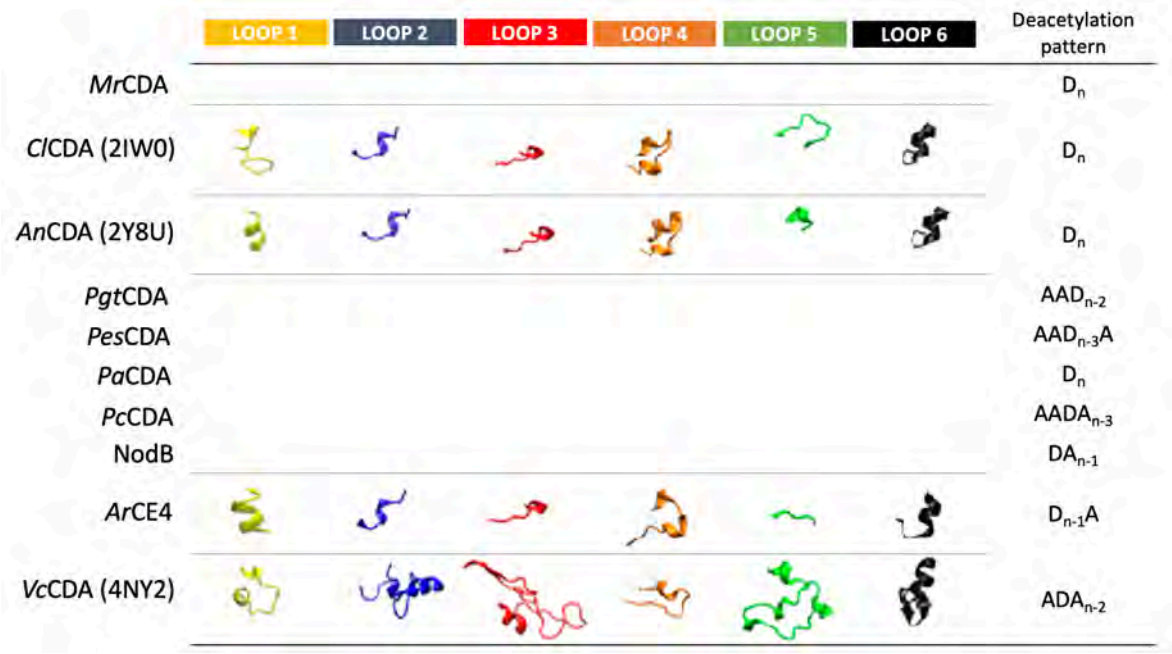


***Figure 2.13**. **CDAs and CODs with known pattern of deacetylation on chitooligoshacarides**. Adapted from* [47].

There is the need of gaining further structural information of enzyme-substrate complexes of CE4 enzymes for better understanding their functionality. This knowledge will be crucial to the design or discovery of novel CDAs with controlled specificities in a rational way. It could be possible to modulate the final result of the catalysis by, for example, exchanging pre-defined loops, but it is necessary to demonstrate the universality and applicability of this theory.

2.5. Nano3Bio Project and application of CE4 enzymes active on COS

As already stated, due to their physico-chemical properties and derivative biological activities, chitosans and COS are some of the most promising functional biomolecules with a wide range of applications in such distant fields as agriculture, cosmetics, water treatment, medicine and the food industry [111]–[115]. These properties have been shown to be strongly dependent on chitosan's DP and DA [116] and, lately, also PA has been proposed to significantly influence the biological

activities of chitosans [117]. However, current available chitosans are obtained with chemical methods that render invariably mixtures of molecules differing in these parameters. Although average DP and DA as well as DP diversity (polydispersity index, Ip) of chitosan samples can be determined, there is no method to measure the polydispersity in terms of DA. In this context, there is a need to find alternative methods to produce partially acetylated COS with known and defined patterns of acetylation that will help to understand the structure-function relationships.

The European Project **NanoBioEngineering of BioInspired Biopolymers** (**Nano3Bio**) was created with the main goal of developing an alternative to the current industrial production of chitosans. The aim of the consortium was to obtain completely defined partially deacetylated chitosan oligomers and polymers through the use of new biotechnological methods, such as genetic, metabolic and enzymatic engineering.

The project included the use of both *in vitro* (based in chemoenzymatic technologies) and *in vivo* (based on metabolic engineering) strategies. The tasks of the different partners that participated in the project went from the discovery of genes and enzymes related to the synthesis and transformation of chitinous substrates, going through the re-design of biocatalysts by enzymatic engineering in order to improve their activities or obtain new ones, to the implementation of metabolic engineering strategies, with which bacterial cell factories were created introducing genes codifying for these new activities into the genetic material of microorganisms. Additionally, it was intended to find new applications for the biotechnologically produced chitosans and to study the ecological impact of their production.

Within chitin and chitosan modifying enzymes (CCMEs), the principal enzymes studied were chitin synthases, chitin deacetylases, chitinases and chitosanases. In one hand, the designed strategy consisted in the use of chitin deacetylases (or related enzymes active on COS) to generate chitosans with regular patterns of acetylation and the use of chitinases and chitosanases to hydrolyze the chitosan polymers' glyosidic bonds in specific sequences in order to generate defined chitosan oligomers. On the other hand, it was intended to also use chitinases with transglycosidase activity for the in vitro polymerization of these defined chitosan oligomers. To complete the strategy, bacterial chitin synthases were first used to produce chitin oligomers in vivo, which can be partially deacetylated by CDAs both *in vivo* or *in vitro,* or polymerized using transglycosidases or glycosynthases.

The in vitro strategy allows obtaining defined chitosans and gives a solution to the lack of reproducibility on the biologic assays performed with chemically produced chitosans. In contrast, with the in vivo strategy, it will be possible to overcome the disadvantage that supposes the use of products with animal origin, mainly on the medical, pharmaceutical and cosmetic areas.

In the recent years, several advances have been made within the consortium. Regarding the in vitro approach and COS deacetylation, the discovery and characterization of several CDAs and related CE4 enzymes contributed to enlarge the knowledge on their specificity.

Enzymes with different specificities were combined to obtain new products that would enlarge the portfolio of partially acetylated COS with defined structures. Initially, it was demonstrated that the two specific CODs, *Rm*NodB (deacetylates specifically the non-reducing end of COS) and *Vc*CDA (deacetylates specifically the penultimate residue from the non-reducing end of COS), each accepted the mono-deacetylated product from the other leading to a specific di-deacetylated product obtained in a one-pot reaction [118]. More recently, all fourteen possible partially acetylated chitosan tetramers have been obtained combining different recombinant bacterial and fungal CDAs, performing both enzymatic deacetylations and N-acetylations [119] (Figure 2.14).

The use of these enzymes as a novel methodology for the production of defined partially acetylated COS is promising, but further structural information on protein-ligand complexes is needed in order to have a deeper knowledge on substrate specificity in this family of enzymes. Determining the structural and sequential determinants of this specificity would allow the rational design or discovery of new deacetylates with controlled deacetylation patterns.

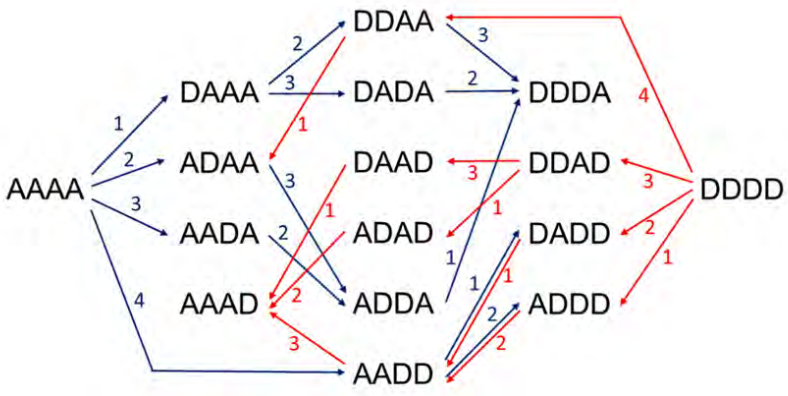


Figure 2.14. Production of all possible chitosan tetramers. *Scheme of the production routes of all possible chitin and chitosan tetramers using 4 different CDAs to specifically deacetylate or N-acetylate COS. A: GlcNAc; D: GlcNH*$_2$*. Blue arrows: deacetylation reactions; red arrows: N-acetylation reactions in the presence of excess acetate* [120].

2.6. Review on chitin deacetylases

Polymers, 10, 352. 2018

Chitin Deacetylases: Structures, Specificities, and Biotech Applications.

Laia Grifoll-Romero, Sergi Pascual, Hugo Aragunde, Xevi Biarnés and Antoni Planas

Laboratory of Biochemistry, Institut Químic de Sarrià, Universitat Ramon Llull, 08017 Barcelona, Spain.

1. Introduction
2. Chitin Deacetylases and the Carbohydrate Esterase Family (CE4)
3. Function and Specificity of CE4 Chitin Deacetylases
4. Structural Determinants of Activity and Specificity
5. Applications of Chitin Deacetylases

During the course of this work, a review was published on the structure, specificity and applications of chitin deacetylases. The complete article can be consulted on Annex 9.6.

OBJECTIVES

3. OBJECTIVES

The work carried out in this doctoral thesis revolves around the study of the Carbohydrate Esterase Family 4 (CE4) of enzymes. The main objective of the project is to discover, characterize and engineer chitin and peptidoglycan deacetylases from this family and to increase the available knowledge regarding the sequence and structural features that determine their substrate specificities. This knowledge will contribute to exploit chitin and peptidoglycan deacetylases both as biocatalysts and as therapeutic targets against fungal and bacterial infections, respectively.

In this context, the specific goals are:

Objectives Chapter 1: PdaC from *Bacillus Subtilis*

1. Phylogenetic analysis of *Bs*PdaC within the CE4 family of enzymes.
2. Expression, purification and characterization of *Bs*PdaC.
3. Structure determination of *Bs*PdaC and identification of the structural determinants of its dual specificity.
4. Rational design of *Bs*PdaC mutants to identify key active site residues and to engineer the enzyme's pattern of deacetylation on chitooligosaccharides.

Objectives Chapter 2: discovery of new CE4 enzymes

1. Selection of new gene sequences encoding for putative chitin deacetylases, using the Subsite Capping Model framework in the search for novel patterns of deacetylation.
2. Expression, purification and characterization of the selected new chitin deacetylases.
3. Identification, selection and characterization of a novel peptidoglycan MurNAc deacetylase.

CHAPTER 1: PdaC from *Bacillus subtilis*. Structure, characterization and engineering

4. CHAPTER 1: PdaC from *Bacillus subtilis*. Structure, characterization and engineering

The peptidoglycan deacetylase PdaC from *Bacillus subtilis* was first identified and initially characterized by Kobayashi and coworkers in 2012 [121]. They reported the enzyme to be a specific MurNAc deacetylase on peptidoglycan that also showed activity on GlcNAc residues of chitin oligomers. This was the first example of a peptidoglycan MurNAc deacetylase with such a dual activity. Furthermore, the pattern of acetylation of the obtained product when using $(GlcNAc)_4$ as substrate was preliminary described to be specific, with only the third residue from the non-reducing end deacetylated (AADA, where A is a GlcNAc residue and D is a $GlcNH_2$ residue).

Being the use of enzymes as biocatalysts a promising strategy for the production of structurally defined chitosan oligomers and polymers, and considering the still scarce available information on the structure-function relationship of CE4 enzymes specificity, further characterization of this new enzyme could bring new insights into the field. The main objective of this work is to solve the crystal structure of PdaC and characterize the enzyme to uncover the molecular basis of its dual activity.

4.1. *Bs*PdaC sequence and phylogeny of CE4 family

*Bs*PdaC is a multimeric protein that consists of 467 amino acids with an initial N-terminal transmembrane domain (from residues 7 to 26), two domains with no assigned function (domains of unknown function by Pfam classification) (from residues 54 to 138 and 156 to 235) and the C-terminus catalytic domain (polysaccharide deacetylase 1 by Pfam classification) (from residues 272 to 392), corresponding to the conserved domain of CE4 enzymes (Figure 4.1). An initial bioinformatic analysis of the protein sequence was performed in order to compare it with other deacetylases from the same family.

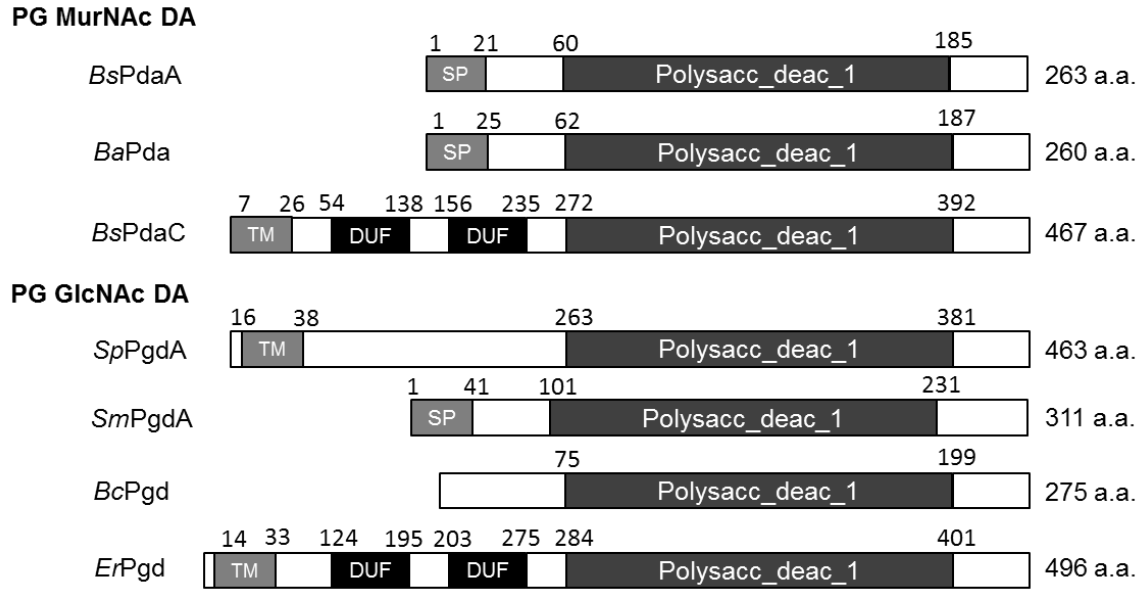


Figure 4.1. Domain organization of peptidoglycan deacetylases*. Numbering according to Pfam database. SP: Signal Peptide; TM: transmembrane region; DUF: Domain of Unknown Function.*

4.1.1. Structural Alignment of CE4 enzymes

A multiple sequence alignment including all CE4 enzymes with known crystallographic structure or with a reported biochemical characterization was built. Several of the enzymes are multimeric and the range of sequence lengths is wide, reasons why using standard multiple sequence alignment algorithms based only on sequence composition does not achieve a proper alignment of the conserved domain of CE4 enzymes, being necessary to consider the available structural information. The Multiseq tool from VMD (Visual Molecular Dynamics) software was used in order to perform a multiple alignment associated to the structural superposition of the available 3D structures of CE4 family members, which was restricted to represent only the catalytic NodB domain. A Hidden Markov Model (HMM) profile was built from the obtained alignment using the hmmbuild program from hmmer software. By means of HMM comparisons, it was possible to include sequences of enzymes without x-ray structure. The final alignment included those members of the CE4 family with known structure or with reported biochemical or kinetic characterization (Table 4.1) and it is shown in Figure 4.2.

The addition of PdaC's sequence to the final alignment (Figure 4.2) allowed the location of the catalytic residues within the conserved motifs of this family of enzymes (MT 1 to 5, Table 4.2) [25]. Motif 1 (TFDDG) contains the catalytic base (first aspartate, Asp285 in *Bs*PdaC) and one residue of the metal coordination triad (second aspartate, Asp286 in *Bs*PdaC). Motif 2 (H(S/T)xxH) contains

the two histidines (His336 and His440 in *Bs*PdaC) that complete the metal coordination triad along with the aspartate in MT1. Motif 5 (I(V/I)LxHD) contains the catalytic acid (His427 in *Bs*PdaC).

Table 4.1. CE4 enzymes with solved x-ray structure or characterized activity on COS (known pattern of deacetylation on chitooligosaccharides or reported in vitro characterization on these substrates). *(B) bacterial, (E) eukaryotic, (*) sequences not included in the multiple sequence alignment presented on Figure 4.2. Table updated on June 2020.*

ACTIVITY	ORGANISM	ENZYME/GENE	Uniprot	PDB
Peptidoglycan deacetylase	*Bacillus anthracis* (B)	BA0330 (a)	Q81ZD9	4V33
	Bacillus anthracis (B)	BA0150 (a)	Q81VP2	4M1B
	Bacillus anthracis (B)	BA3943* (a)	A0A3P1U7W4	6HM9
	Bacillus cereus (B)	BC0361	Q81IM3	4HD5
	Bacillus cereus (B)	*Bc*Pgd	Q81EK9	4L1G
	Bacillus cereus (B)	BC1974	Q81EJ6	5NCD
	Eubacterium rectale (B)	*Er*Pgd	C4ZEZ9	5JMU
	Streptococcus pneumoniae (B)	*Sp*PgdA	Q8DP63	2C1G
	Streptococcus mutants (B)	*Sm*PgdA	Q8DV82	2W3Z
	Bacillus subtilis (B)	*Bs*PdaA	O34298	1NY1
	Bacillus anthracis (B)	*Ba*Pda	Q81Z49	2J13
	Bacillus subtilis (B)	*Bs*PdaC	O34798	This work
Poly-ß-1,6-N-acetyl-D-glucosamine deacetylase	*Aggregatibacter actinomycetemcomitans* HK1651 (B)	*Aa*PgaB (a)	A5HJW8	4U10
	Ammonifex degensii (B)	IcaB	C9RCK9	4WCJ
	Bordetella bronchiseptica (B)	Bbpb	A0A0C6P1R7	5BU6
	Escherichia coli (B)	*Ec*PgaB	P75906	3VUS
Acetyl-xylan esterase	*Clostridium thermocellum* (B)	*Ct*CE4	O87119	2C71
	Streptomyces lividans (B)	*Sl*AxeA	QS4413	2CC0
Chitin deacetylase	*Aspergillus nidulans* (E)	*An*CDA	Q5AQQ0	2Y8U
	Colletotrichum lindemuthianum (E)	*Cl*CDA	Q6DWK3	2IW0
	Arthrobacter sp. (B)	*Ar*CE4	A0A2C8C1T7	5LFZ
	Bombyx mori (E)	*Bm*CDA1	H9J9M0	5ZNT
	Bombyx mori (E)	*Bm*CDA8	-	5Z34
	Mucor rouxii (E)	*Mr*CDA	P50325	-
	Puccinia graminis (E)	*Pgt*CDA	E3K3D7	-

	Pestolotiopsis sp. (E)	*Pes*CDA	A0A1L3THR9	-
	Podospora anserina (E)	*Pa*CDA	B2AAQ0	-
	Pochonia chlamydosporia (E)	*Pc*CDA	-	-
	Saccharomyces cerevisiae (E)	*Sc*CDA 1	Q06702	-
	Saccharomyces cerevisiae (E)	*Sc*CDA 2	Q06703	-
	Rhizopus circinans (E)	*Rc*CDA	Q32XH4	-
	Gongronella butleri (E)	*Gb*CDA	Q8J2N6	-
	Phycomyces blakesleanus (E)	*Pb*CDA	Q9P4U2	-
	Schizophyllum commune (E)	*Sch*CDA	Q9P453	-
	Cryptococcus neoformans (E)	*Cn*CDA 1	Q5KFG8	-
	Cryptococcus neoformans (E)	*Cn*CDA 2	Q5KIC2	-
	Cryptococcus neoformans (E)	*Cn*CDA 3	P0CP76	-
	Flammulina velutipes (E)	*Fv*CDA	-	-
Chitooligosaccharide deacetylase	*Sinorhizobium meliloti* (B)	NodB	P02963	-
	Vibrio cholerae (B)	*Vc*COD	Q9K5H6	4NY2
	Vibrio parahaemolyticus (B)	*Vp*COD	A6P4T5	3WX7
	Shewanella woodyi (B)	*Sw*COD	-	-
	Shewanella báltica (B)	*Sb*COD	-	-
Putative polysaccharide deacetylase	*Burkholderia pseudomallei* (B)	BPSL2118*	Q63T51	3S60
	Pseudomonas aeruginosa PAO1 (B)	PA1517*	Q913J6	1Z7A
	Encephalitozoon cuniculi (E)	ECU11_0510	Q8SU65	2VYO
	Streptococcus p yrogenes (B)	BAM30204.1	J7M3C9	6DQ3

Table 4.2. Conserved motifs of CE4 family. *Comparison between the consensus sequence and the specific residues in the BsPdaC sequence.*

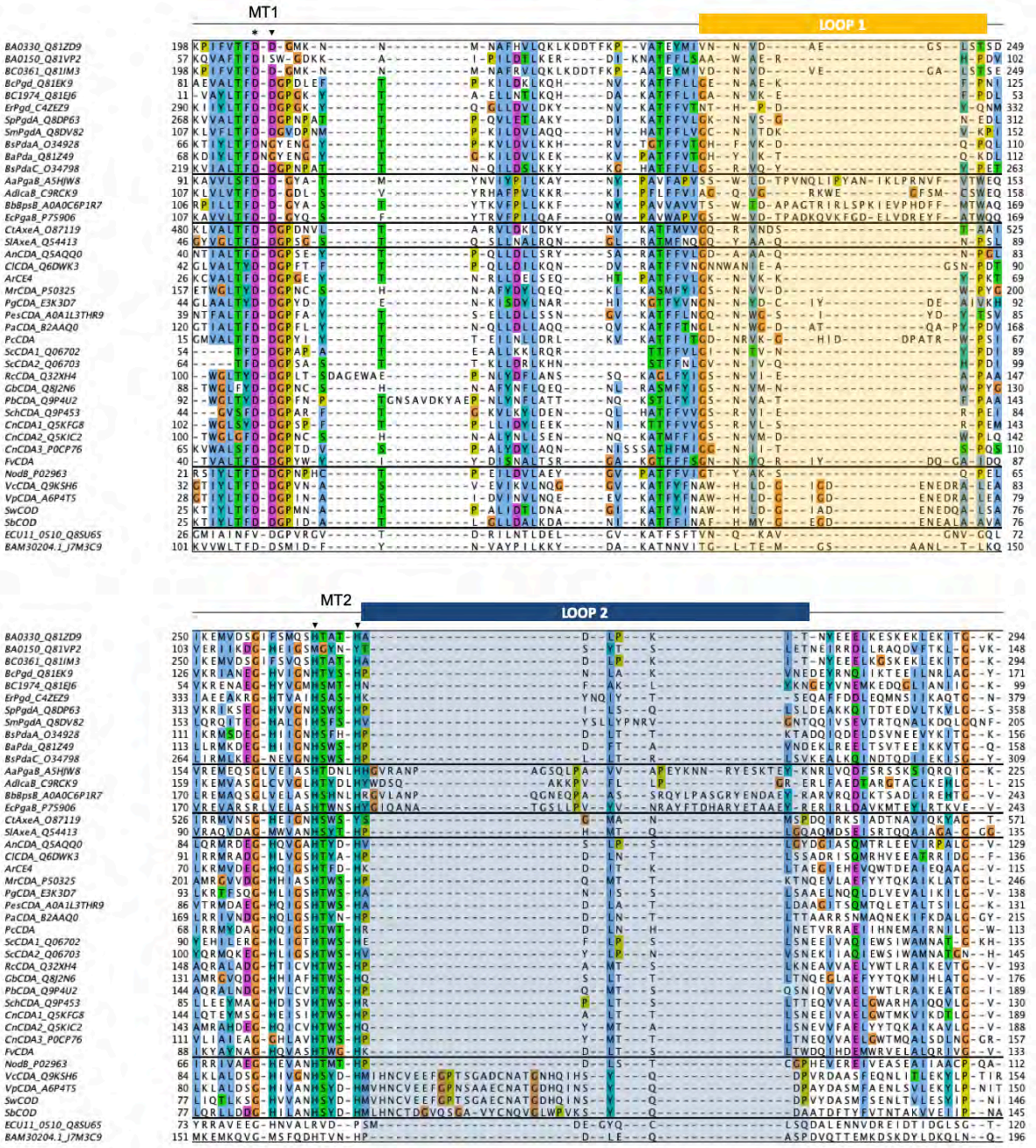
MOTIF	CONSENSUS SEQUENCE	POSITION ON *Bs*PdaC	SEQUENCE ON *Bs*PdaC
MT 1	TFDDG	283 – 287	TFDDG
MT 2	H(S/T)XXHP	336 – 341	HSWSHP
MT 3	RXPY	374 – 377	RPPY
MT 4	DXXD(W/Y)	398 – 402	DPEDWK
MT 5	I(V/I)LXHD	423 – 428	TILIHD

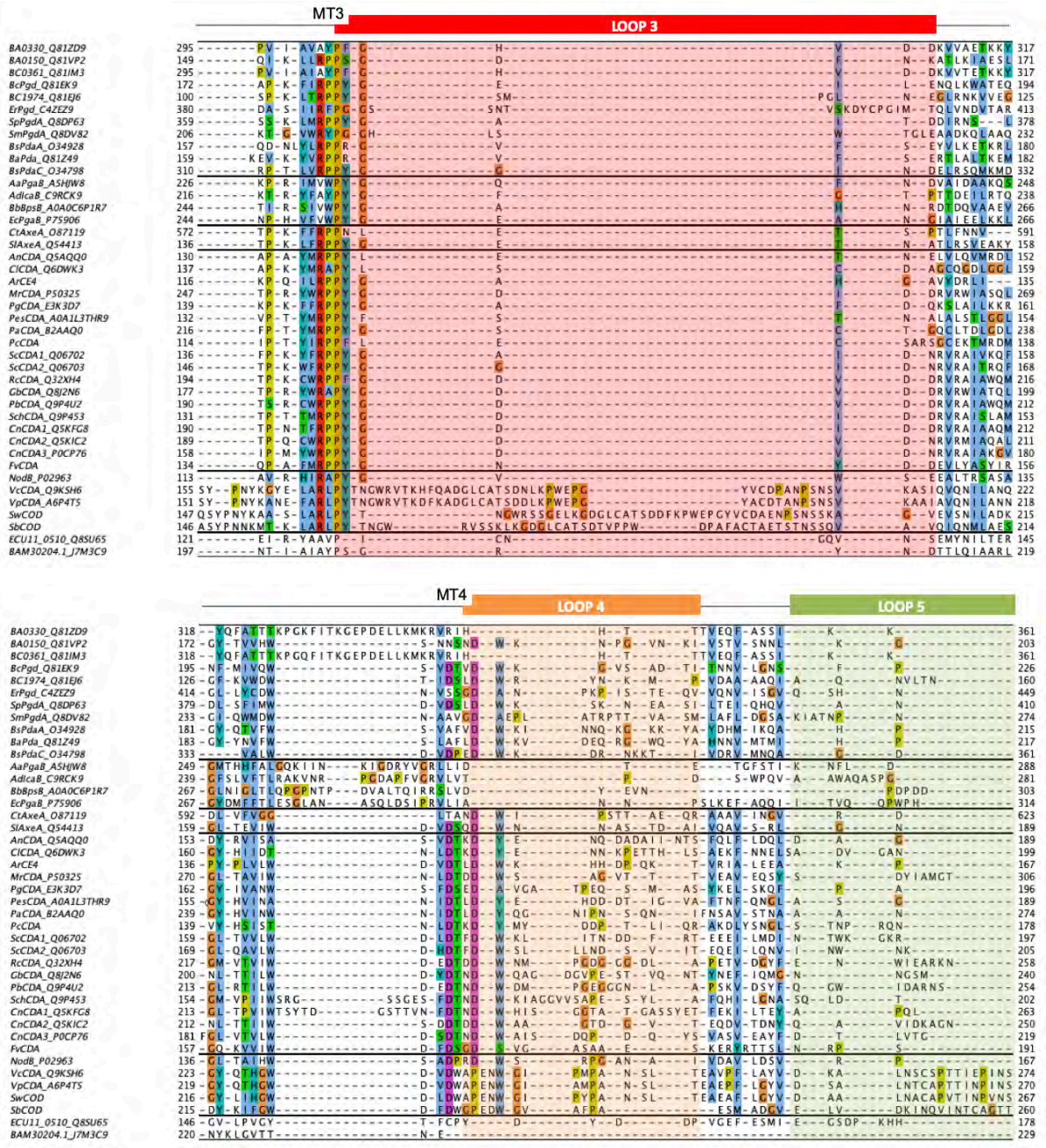
The alignment also identified the position and length of the 6 characteristic loops that shape the active center of the enzyme and that have been related to substrate specificity according to the

Subsite Capping Model [24]. These loops are short in *Bs*PdaC compared to other family members, being loop 4 the longest with 10 amino acids and loop 5 the shortest with 3 amino acids (Table 4.3).

Table 4.3. Loops surrounding active site of CE4 enzymes. *Location and sequence of the 6 loops on BsPdaC enzyme.*

LOOP	POSITION ON *Bs*PdaC (length)	SEQUENCE ON *Bs*PdaC
Loop 1	313-320 (8)	GSRVQYYP
Loop 2	341 – 348 (8)	PLLTRLSV
Loop 3	376 – 382 (7)	PYGGIND
Loop 4	401 – 410 (10)	DWKDENKKTI
Loop 5	419 – 421 (3)	GDG
Loop 6	428 – 437 (10)	DIYRTSADAA


MT1
LOOP 1
MT2
LOOP 2
BA0330_Q81ZD9
BA0150_Q81VP2
BC0361_Q81IM3
BcPgd_Q81EK9
BC1974_Q81EJ6
ErPgd_C4ZEZ9
SpPgdA_Q8DP63
SmPgdA_Q8DV82
BsPdaA_O34928
BaPda_Q81Z49
BsPdaC_O34798
AaPga8_A5HJW8
AdlcaB_C9RCK9
BbBpsB_A0A0C6P1R7
EcPga8_P75906
CtAxeA_O87119
SlAxeA_Q54413
AnCDA_Q5AQQ0
ClCDA_Q6DWK3
ArCE4
MrCDA_P50325
PgCDA_E3K3D7
PesCDA_A0A1L3THR9
PaCDA_B2AAQ0
PcCDA
ScCDA1_Q06702
ScCDA2_Q06703
RcCDA_Q32XH4
GbCDA_Q8J2N6
PbCDA_Q9P4U2
SchCDA_Q9P453
CnCDA1_Q5KFC8
CnCDA2_Q5KIC2
CnCDA3_P0CP76
FvCDA
NodB_P02963
VcCDA_Q9KSH6
VpCDA_A6P4T5
SwCOD
SbCOD
ECU11_0510_Q8SU65
BAM30204.1_J7M3C9


MT3
LOOP 3
MT4
LOOP 4
LOOP 5

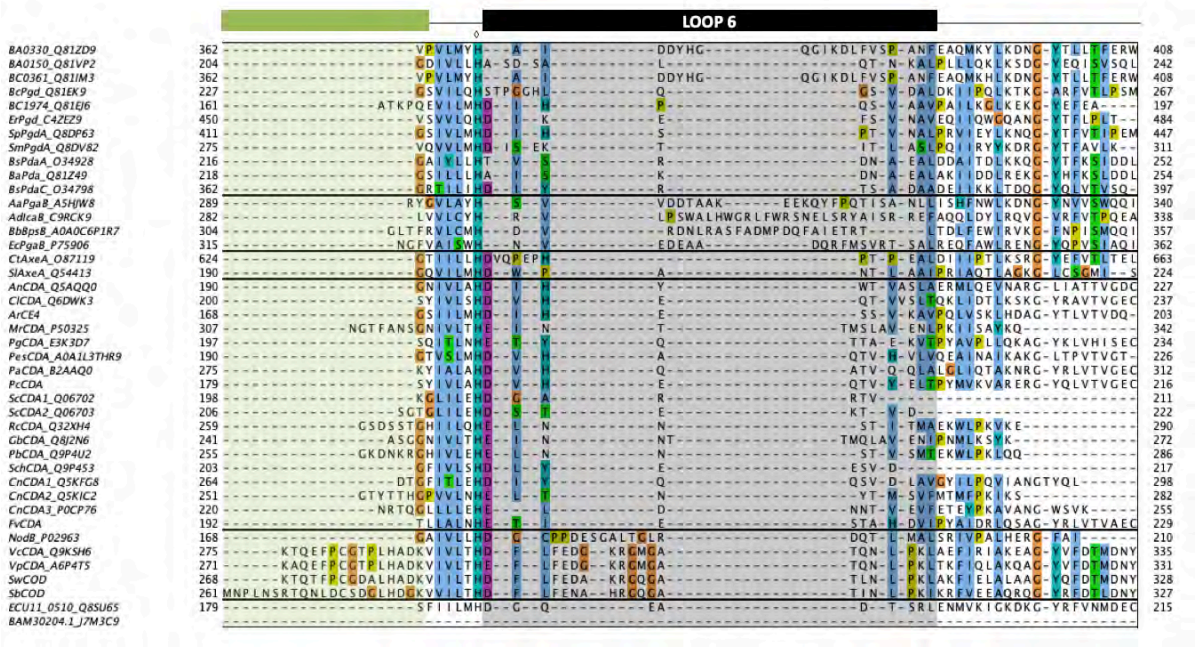


Figure 4.2. Multiple sequence alignment of CE4 enzymes with solved x-ray structure or characterized activity on COS (with sequences from Table 4.1)*. Loops are highlighted with colored boxes according to* [24]*. Conserved catalytic motifs are labelled MT1-5. "His-His-Asp" metal binding triad (▼), catalytic base (✶), and catalytic acid (◊) are labelled.*

4.1.2. Phylogenetic analysis

In order to identify the closest CE4 enzymes to PdaC, a phylogenetic analysis based on the multiple sequence alignment was performed with MEGA software using the Neighbor-joining algorithm (Figure 4.3). This analysis showed that the peptidoglycan deacetylase PgdA from *Streptococcus pneumoniae* (*Sp*PgdA) is the closest CE4 enzyme to *Bs*PdaC, with 30% overal sequence identity and 44% sequence identity for the CE4 catalytic domain. Although they are both peptidoglycan deacetylases, *Sp*PgdA acts on GlcNAc residues of PGN, whereas *Bs*PdaC was reported to deacetylate MurNAc residues. The other two known peptidoglycan MurNAc deacetylases with x-ray structure (*Bs*PdaA and *Ba*CE4) are clustered together in a more distant position from PdaC in the phylogenetic tree. Apart from this, it can be observed from the sequence alignment that these two last enzymes lack the conserved second Asp of MT1, which is replaced by an Asn residue, feature that was so far considered as an indicative of MurNAc deacetylation. However, this trait is not present in PdaC, which seems to be a unique enzyme in the family due to its double activity, both on MurNAc residues (on peptidoglycan) and GlcNAc residues (on chitin oligosaccharides).

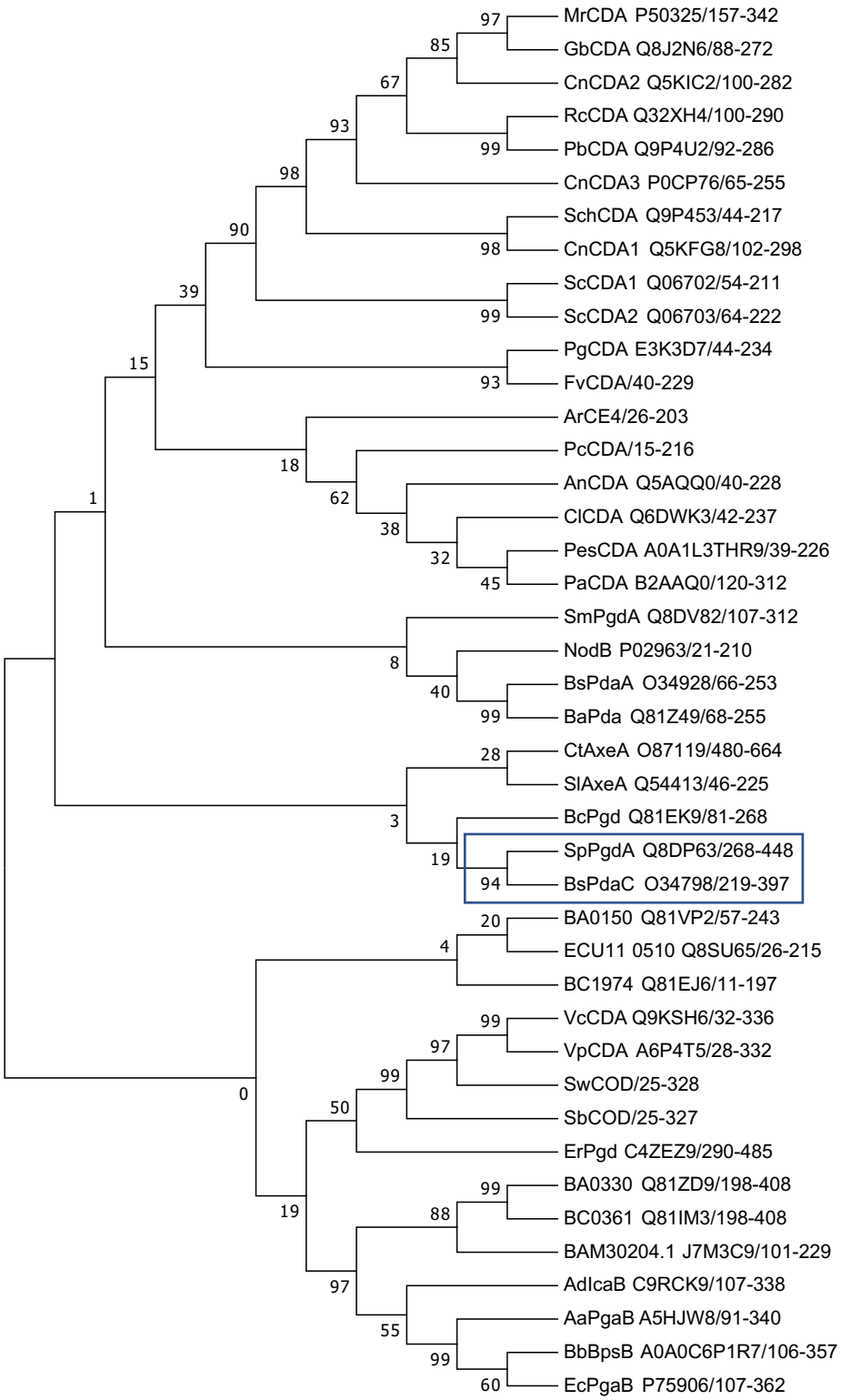


Figure 4.3. Phylogenetic analysis of CDAs from the multiple sequence alignment presented in Figure 4.2*. A bootstrap analysis with 500 replicates was carried out on the trees inferred from the neighbor joining method. The consensus tree is shown with bootstrap values at each node of the tree.*

4.2. Biochemical characterization and enzyme kinetics of *Bs*PdaC catalytic domain

The full-length *Bs*PdaC, *Bs*PdaC-FL (residues 27 to 467, lacking the transmembrane region), and the isolated catalytic domain, *Bs*PdaC-CD (residues 270 to 467), were both recombinantly expressed and purified to homogeneity. Both enzymes displayed similar enzymatic activity on initial characterization studies, and it was decided to proceed with the isolated catalytic domain for further characterization. The results presented in this work include only those of the truncated version of the enzyme but were comparable to those of the full-length enzyme in terms of expression and purification yields (see Annex 9.2).

4.2.1. Expression vector: pET22b with strep tag

For the expression of the enzyme in *E. coli*, it is necessary to clone its encoding gene into a suitable expression vector.

The high copy number plasmid pET22b(+), which allows the fast production of a high quantity of protein once induced, was chosen. The original plasmid includes a His-tag for the purification of the expressed protein by Immobilized Metal Affinity Chromatography (IMAC). Since CE4 enzymes contain metal ions in their catalytic center, the possibility that this tag could negatively influence the enzyme's activity was considered. For this reason, the peptide Strep-Tag II was included in the C-terminus end of the enzyme, followed by a stop codon (Figure 4.6). This additional peptide allows protein purification by affinity chromatography without the presence of metal ions. Furthermore, the amino acid sequence of Strep-Tag II (WSHPQFEK) is rare in nature and allows obtaining higher purity levels in comparison to the original His-tag.

In order to decide the exact sequence that would conform the isolated catalytic domain, a study of the location of this domain on the enzyme's sequence was performed. The assigned location of the enzyme's domains, according to different databases, was determined with the online tool Interpro (Figure 4.4). In addition, the secondary structure prediction of PdaC was obtained with the online tool Psired (Figure 4.5) to ensure that no secondary structure would be disrupted.

When comparing the predicted secondary structure and the domains location on the PdaC sequence, it was observed that the start of the deacetylase domain was always located in an area without secondary structures. It was decided to start the truncated form with Glu270, which is a

negatively charged amino acid, because the C-terminus sequence would finish with a positive Lys residue from the Strep-tag II peptide. Regarding the final residue of the construct, being the catalytic domain at the end of the sequence, it was decided to fully conserve the C-terminus of the enzyme.

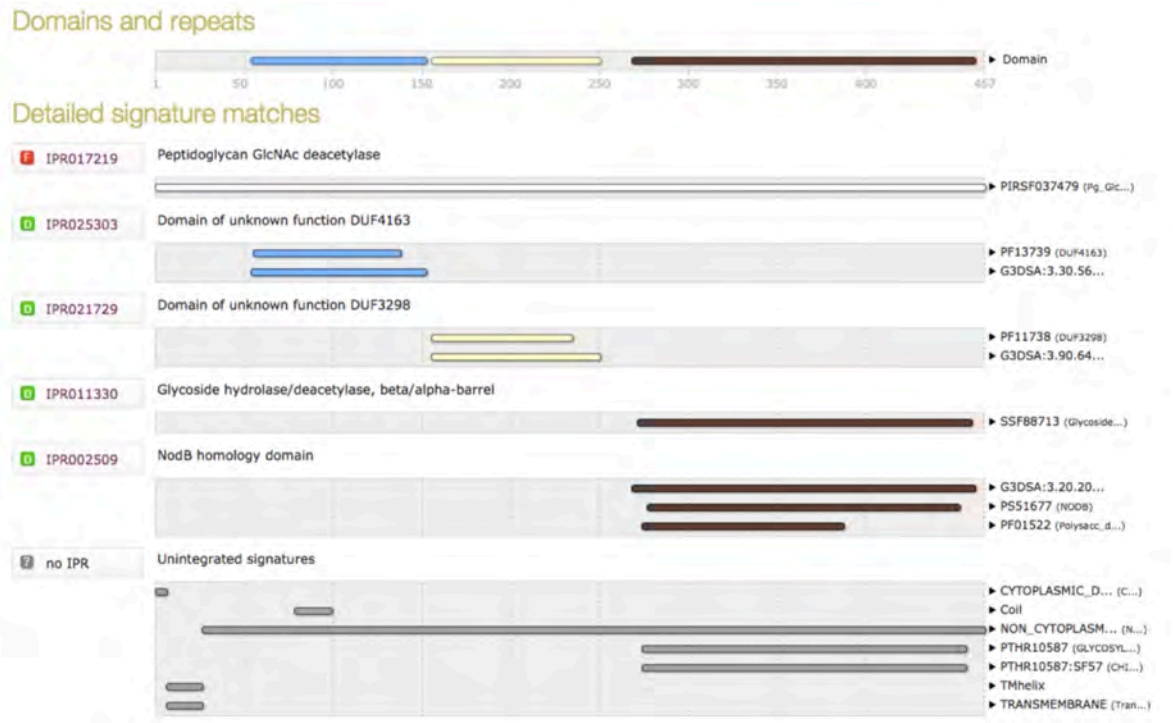


Figure 4.4. Prediction of domains and repeats of BsPdaC. *Domains assignment on the enzyme sequence in different databases. Information retrieved from Interpro, online tool from the European Bioinformation Institute (EBI).*

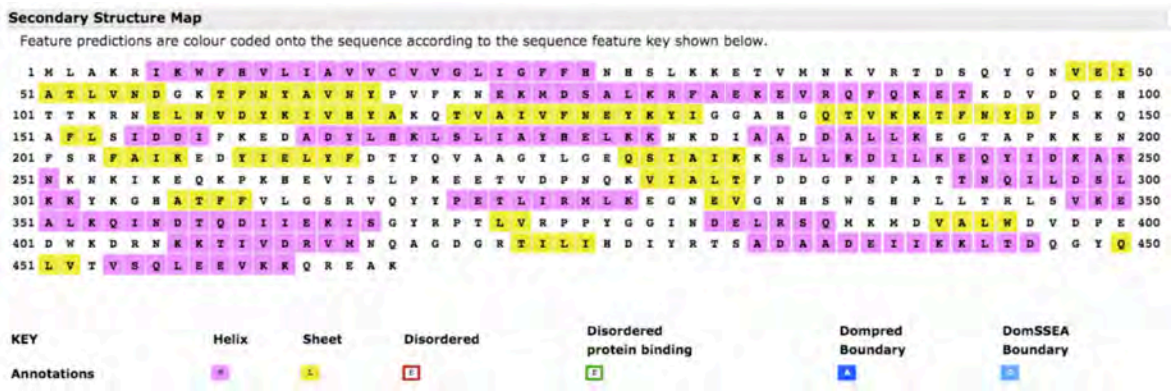


Figure 4.5. Secondary structure prediction of BsPdaC. *Obtained with PSI-blast based secondary structure PREDiction (PSIPRED), method used to investigate protein structure.*

The designed synthetic gene encoding for the catalytic domain of *Bs*PdaC (from residue 270 to 467) with the Strep-tag II sequence at the C-terminus of the protein was produced codon-optimized for its expression in *E. coli*. It was supplied in a pMK-RQ plasmid and it was subcloned into a pET22b(+) plasmid between NdeI and SacI restriction sites to generate the C-terminal strep-tagged protein

(Figure 4.6). After sequencing, competent *E. coli* DH5α and *E. coli* BL21(DE3) STAR cells were transformed with the obtained plasmid.

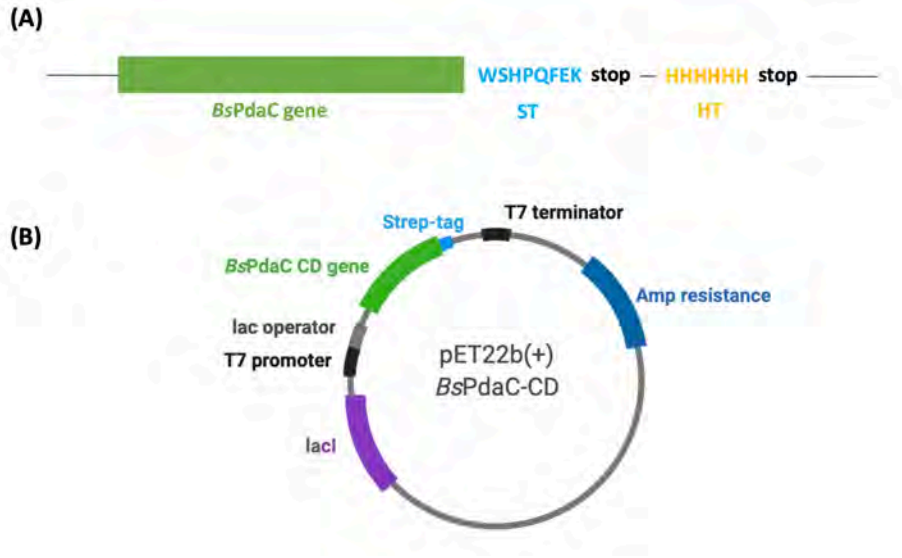


Figure 4.6. pET22b(+) BsPdaC catalytic domain (BsPdaC-CD). *Sequence of the expression vector pET22b(+) with the subcloned sequence of PdaC's catalytic domain and the Strep-tag II sequence in the C-terminus end of the protein.* ST: strep-tag; HT: his-tag.

4.2.2. Expression and purification of *Bs*PdaC catalytic domain

For the expression in *E. coli,* both IPTG induction and the use of an optimized autoinduction medium were considered. This second type of method allows the automatic expression of the protein without the need of monitoring cell growth, changing the medium or adding inducers. Moreover, high optical density is reached, increasing the yield compared with other strategies.

Both methods were performed testing different conditions, such as growth temperature and induction time (for IPTG) or incubation time (for autoinduction). Both showed to be adequate for obtaining high levels of overexpressed protein in the soluble fraction. Bearing in mind the mentioned advantages of working with the autoinduction system, it was chosen as the expression protocol of *Bs*PdaC in *E. coli*.

Final expression conditions involved growing a certain volume of recombinant cells (*E. coli* BL21(DE3) star transformed with the plasmid containing *Bs*PdaC catalytic domain gene) at 25°C for 48 hours. After cell lysis by sonication, the soluble fraction after centrifugation was used for the purification steps.

For the isolation of the enzyme from the cellular lysate, two chromatographic steps were performed: an affinity chromatography with a StrepTrap column followed by a Size Exclusion Chromarography (SEC) with a Superdex 200 (16/600) column.

(i) Affinity chromatography: StrepTrap

Affinity chromatography is one of the most diverse and powerful chromatographic methods for the purification of a specific molecules or group of molecules from complex mixtures, and it is based on highly specific biologic interactions between two molecules. The StrepTrap principle is based on the interaction between d-biotin and streptavidin, one of the strongest non-covalent interactions known.

For *Bs*PdaC purification, between two and four cycles (which included loading, washing, elution and regeneration steps) were performed in the same column, depending on the volume of the initial expression culture (Figure 4.7).

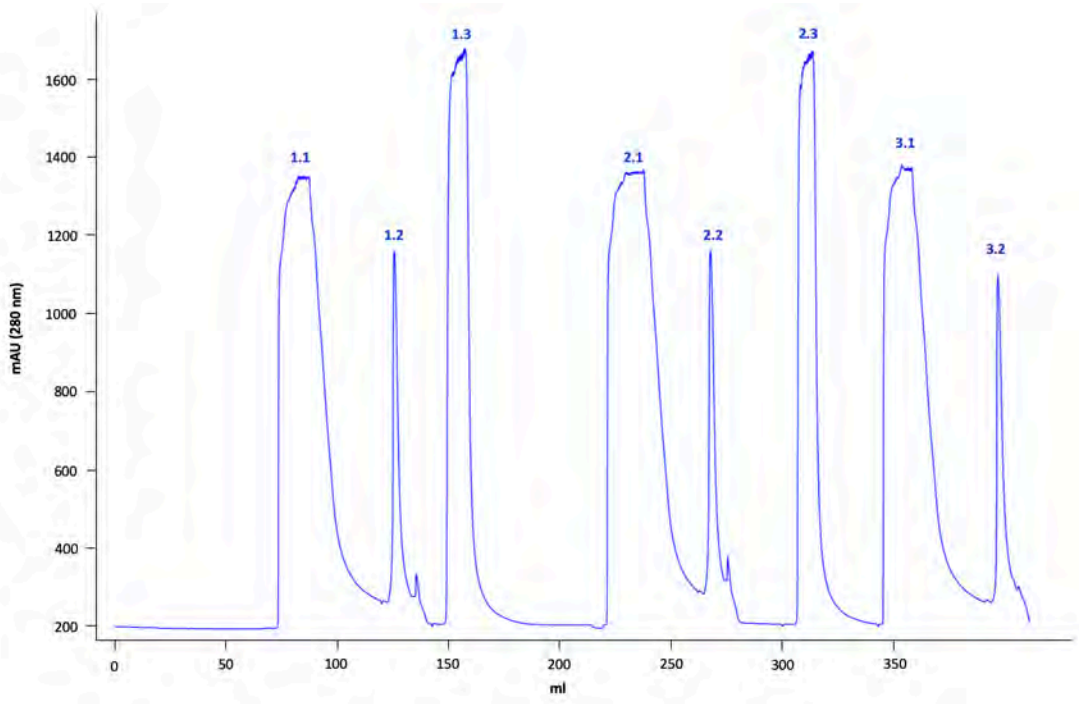


Figure 4.7. Chromatogram of Strep-trap purification of BsPdaC-CD. *Resulting chromatogram from affinity chromatography in which three cycles of loading and elution were performed. Numbers indicate the steps of each loading/elution cycle: 1.1, 2.1 and 3.1: loading and washing; 1.2, 2.2 and 3.2: elution; 1.3 and 2.3: regeneration with HABA and washing.*

(ii) Size Exclusion Chromatograpy: Superdex 200

In order to identify and separate possible oligomeric forms of the expressed protein, a second purification step was performed. Size exclusion chromatography (SEC) or gel filtration is a chromatographic technique used to separate molecules by their molecular weight or size while they pass through the column-packed gel filtration stationary phase.

All elution samples from the first affinity chromatography were pooled together, concentrated up to a 2 ml volume and loaded into a Superdex 200 (16/600) column for this second purification step. The elution was performed at a rate of 1 mL/min using PBS (50 mM phosphate, 300 mM NaCl, pH 7.0) as buffer, with the same composition used in the previous steps.

Three different peaks were observed in the chromatogram (Figure 4.8). Calibration of the column with known standards (see Section 7.2.5) allowed the identification of each fraction composition. Fraction 1 (F1, eluted approximately 45 minutes after injection, at the void volume of the column) probably corresponds to soluble oligomers, fraction 2 (F2, eluted approximately 78 minutes after injection) corresponds to the dimeric form of the enzyme, and fraction 3 (F3, eluted approximately 90 minutes after injection) is the monomeric form of *Bs*PdaC. Several purifications of the enzyme were carried out during the performance of this work and it was observed that, although the chromatographic profile always presented the same three characteristic fractions, their relative amount varied between different batches.

SDS-PAGE analysis was performed with samples corresponding to the different chromatographic steps and it revealed that the recombinant enzyme migrated with an approximate molecular mass of 25.000 kDa, which agrees with the theoretical mass (24478.74 kDa) (Figure 4.9). The three fractions from the SEC show the same band under denaturing SDS-PAGE, confirming the non-covalent oligomeric nature of fractions F2 and F3.

Once the protein's purity was checked, all the volume corresponding to the monomeric fraction was pooled together to measure its concentration and calculate the expression yield. Final protein concentration was determined spectrophotometrically by BCA method measuring Abs at 595 nm using BSA as standard. Average yields of 18 mg of pure monomeric protein per liter of culture were obtained. MALDI-TOF MS analysis of pure fraction 3 showed the main peak corresponding with an almost identical mass to the theoretical molecular mass of the recombinant protein (24478.74 Da)

and a second peak corresponding to the double protonated molecular ion (around 12200 kDa) (Figure 4.10).

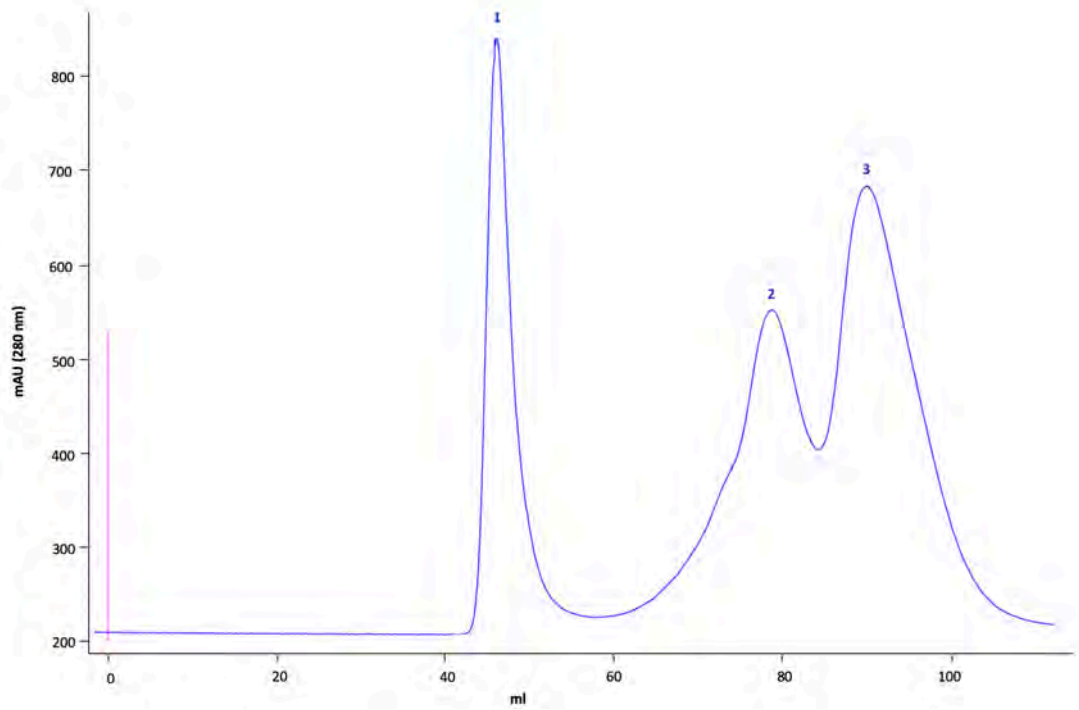


Figure 4.8. Chromatogram of SEC of BsPdaC-CD*. Chromatogram obtained in the gel filtration chromatography using a Superdex 200 (16/600). Three multimeric fractions are shown, from which fraction 2 (F2) corresponds to the dimer and fraction 3 (F3) corresponds to the monomer of the protein.*

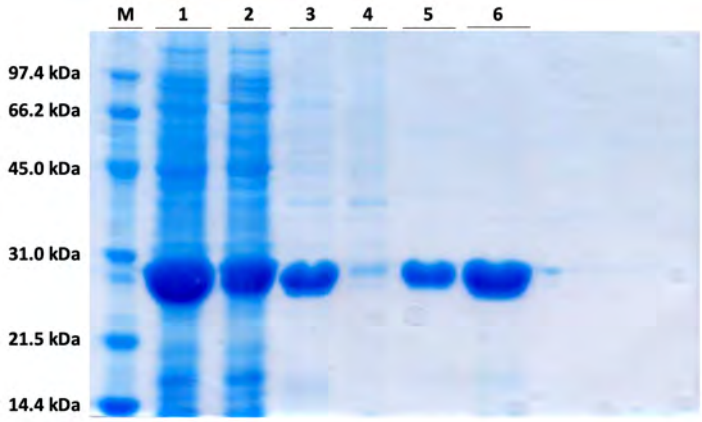


Figure 4.9. SDS-PAGE 14% acrylamide of BsPdaC-CD purification*. M: Marker Low Range (Biorad); 1: cell lysate, soluble fraction; 2: flow through during column loading and washing; 3: elution of protein purified by Strep-trap column; 4: fraction 1 from SEC; 5: fraction 2 from SEC; 6: fraction 3 from SEC.*

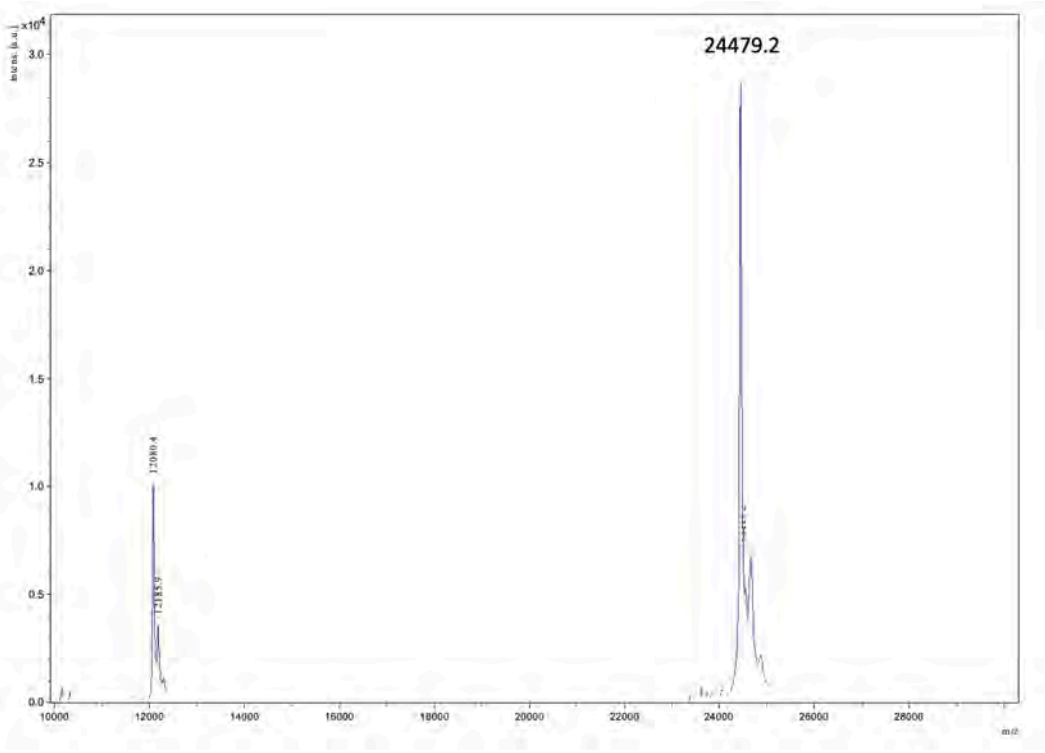


Figure 4.10. MALDI-TOF spectra of pure BsPdaC-CD.

4.2.3. Thermal stability of *Bs*PdaC catalytic domain

The removal of the transmembrane domain and the domains with unknown function from the protein implies a significative disruption of the native *Bs*PdaC protein, reducing in approximately 58% the number of residues of the complete protein sequence. Although they seem to be independent domains connected to the catalytic domain by a linker, it is possible that they have some effect in protein stability and solubility.

One of the most commonly used parameters to estimate proteins stability is their T_m (Melting temperature), at which half of the protein population is in its denatured form and the other half in its native form. Thermal Shift Assays (TSA) are methods in which the changes in the denaturing temperature of a certain protein, under different conditions, are measured. If these conditions remain constant, these assays are a fast method for the T_m comparison between different proteins.

Although it has been shown that the isolated catalytic domain of *Bs*PdaC can be expressed in the soluble fraction with good yields and that it retained its deacetylase activity, it was decided to evaluate the possible effect of the removed domains in protein stability. To determine the melting temperature of both *Bs*PdaC full length and the truncated form consisting of only the catalytic domain, a Differential Scanning Fluorimetry (DSF) assay with the fluorescent dye Spyro Orange

[122] was used. The proteins were subjected to a thermal gradient from 25 to 95°C and the T_m values were determined by fitting the measured fluorescence data to a Boltzman Sigmoidal Equation (see Section 7.3.1). The obtained T_m values were similar for both enzymes, being 60.58 ± 1.4 °C for *Bs*PdaC FL and 59.75 ± 0.6 °C for *Bs*PdaC CD (Figure 4.11). The results indicate that the removal of the initial domains of *Bs*PdaC shows no effect in terms of thermal stability, since both constructs follow practically the same denaturing profile and present the same T_m, and support a well-folded structure for the truncated protein.

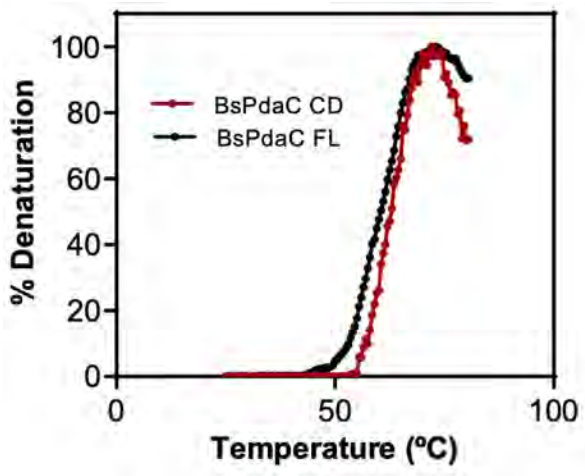


Figure 4.11. Thermal stability of BsPdaC full length (BsPdaC-FL) and catalytic domain (BsPdaC-CD)*. Denaturation curve obtained by the thermal shift assay with Sypro Orange dye and fluorescence monitoring (λex 483 nm, λem 560 nm).*

4.2.4. Enzyme kinetics and deacetyation pattern of *Bs*PdaC catalytic domain

4.2.4.1. Activity profile of oligomeric fractions

Before further characterization of the enzyme, an initial study of the general deacetylation activity of the different fractions obtained from the size exclusion chromatography was performed using the unspecific substrate AcOMU (4-methylumbelliferyl acetate). If hydrolyzed by esterase enzymes, the fluorogenic product 4-methylumbelliferone (MU) would be released and could be detected by measuring fluorescence increase at 340/460 nm along with reaction time.

A reaction with all the 1 ml fractions obtained from SEC was performed with the aim of determining which of the three oligomeric fractions presented activity and, consequently, which ones would later be used for kinetic characterization of the enzyme. The results (Figure 4.12) showed that the fraction with higher activity was the monomeric one, while F1 did not show any activity against AcOMU and the dimeric form showed some activity but at a much lower level than the monomer.

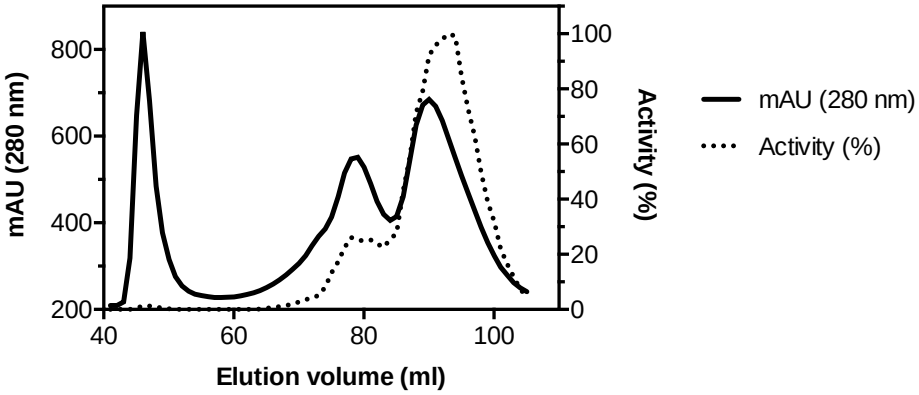


Figure 4.12. Activity profile of SEC fractions from BsPdaC-CD purification. *Gel filtration chromatography of BsPdaC-CD and enzyme activity with AcOMU substrate.*

Having identified the monomer as the most active oligomeric fraction, it was decided to proceed with only this fraction for the subsequent characterization of the enzyme, making sure to always use the monomeric form of the purified protein.

4.2.4.2. Enzyme kinetics on peptidoglycan and chitooligosaccharides

On one hand, the specific activity of the enzyme on *Bacillus subtilis* PGN was measured monitoring acetate release with a coupled enzymatic assay (see Section 7.4.4). The principle of this method is based on the action of an acetate kinase (AK) and a phosphotransacetylase (PTA) and the sequential enzymatic reactions detailed in Figure 4.13. The amount of NAD^+ formed in the reaction pathway is stoichiometric with the amount of acetic acid. NADH consumption is measured by decrease in absorbance at 340 nm.

Reactions were performed with 1 mg/ml of PGN substrate at 37°C. Due to the insolubility of the peptidoglycan polymer, the reaction mixture was a non-homogenous suspension and reactions were always performed with agitation. After enzyme inactivation, the substrate was removed by centrifugation and the supernatant was used to determine acetate release.

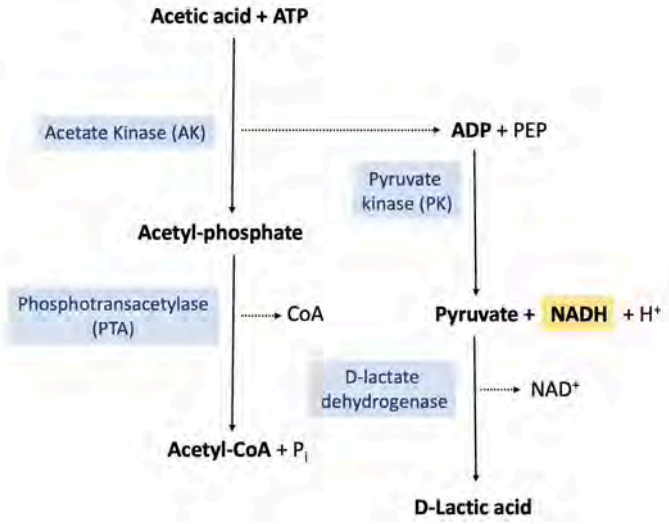


Figure 4.13. Enzymatic coupled reactions for acetate determination.

On the other hand, enzyme activity on COS substrates (DP2 to DP5) was determined by two methods: (i) by quantification of the free amino groups generated from deacetylation of $(GlcNAc)_n$ substrates using fluorescamine labelling, and (ii) by HPLC-MS analysis of the deacetylation reaction in which the sequence of GlcNAc deacetylation events could be monitored with the previously validated method by Aragunde [110]. With both methods, initial rates at short reaction times, corresponding to the first deacetylation event (mono-deacetylated products), were determined.

The first fluorometric assay (see Section 7.4.3) is based on the reaction between the reagent fluorescamine and the primary amines (Figure 4.14) that are formed in the deacetylation reaction by the enzyme. The extent of this reaction as well as the fluorescence intensities of the resulting fluorophores depend on pH, solvent composition and reagent concentration, and optimal values for these variables further depend on the amine under study.

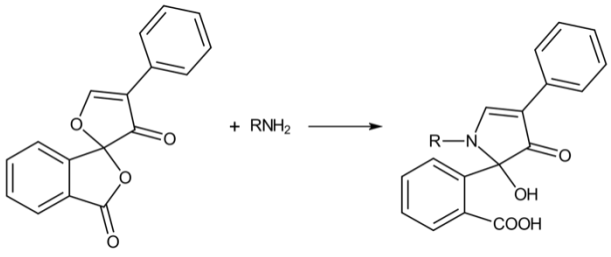


Figure 4.14. Reaction between fluorescamine and primary amine.

The HPLC-MS method used for kinetic characterization of deacetylases (see Section 7.4.2) is based on the direct measurement of chitin oligosaccharides deacetylated by these enzymes. Pure mono-deacetylated standards were obtained from the action of *Vc*CDA on chitin oligomers and a protocol

to stop the reaction at different time points was developed, allowing the separation of the sample preparation and the sample analysis steps simultaneously and, consequently, increasing the efficiency to analyze a high number of samples. The method was validated and showed several advantages with respect to other methods (such as radiometric, colorimetric or fluorometric methods). If sufficiently pure standards are available, the method can be applied to a wide range of products and enzymes [110].

As presented in Table 4.4, the enzyme showed its highest activity on *Bacillus subtillis'* peptidoglycan, its natural substrate. Regarding PdaC's deacetylation activity on chitooligosaccharides, the enzyme showed to be inactive on the disaccharide, being DP3 the minimum substrate for detecting any deacetylation activity. From DP3 to DP5, the specific activity of the enzyme increased with the degree of polymerization. Although the sensitivity of the fluorescamine's method proved insufficient to measure activity with the trisaccharide, in the case of DP4 and DP5, the obtained values are in agreement between the fluorescamine and HPLC methods, being the activity on the longest substrate approximately twice the one on the tetrasaccharide and 10 times higher than the activity on DP3.

Table 4.4. Deacetylation activity of BsPdaC on PGN and chitooligosaccharides. *[1]Peptidoglycan from B. subtilis, specific activity determined by monitoring acetate release. [2]Specific activity for BsPdaC-CD (catalytic domain). [3]Specific activity for BsPdaC-FL (full-length).*

	SPECIFIC ACTIVITY (V_0/[E], min^{-1})	
SUBSTRATE	**FLUORESCAMINE**	**HPLC-MS**
$(GlcNAc)_2$	no activity	no activity
$(GlcNAc)_3$	n.d.	3.99 ± 0.27
$(GlcNAc)_4$	18.9 ± 2.4	18.18 ± 0.22
$(GlcNAc)_5$	38.0 ± 2.8	36.80 ± 3.46
*Bs*PGN [1]	61.2 ± 0.6 [2]	
	81.8 ± 1.7 [3]	

Michaelis-Menten kinetic parameters, which describe the enzymatic behavior, were determined for the $(GlcNAc)_4$ substrate. To do so, 12 enzymatic reactions with different substrate concentration (from 0.5 to 17.5 mM) were performed by duplicate, taking 8 time points for each reaction. All kinetic reactions were performed at 37°C and HPLC-MS monitoring was used as analytical method. The resulting k_{cat} and K_M values were of 0.93±0.03 s^{-1} and 6.25±0.47 mM, respectively, and k_{cat}/K_M

of 149 $M^{-1}s^{-1}$, which are of the same order as those previously reported by Kobayashi and coauthors [121] (Figure 4.15).

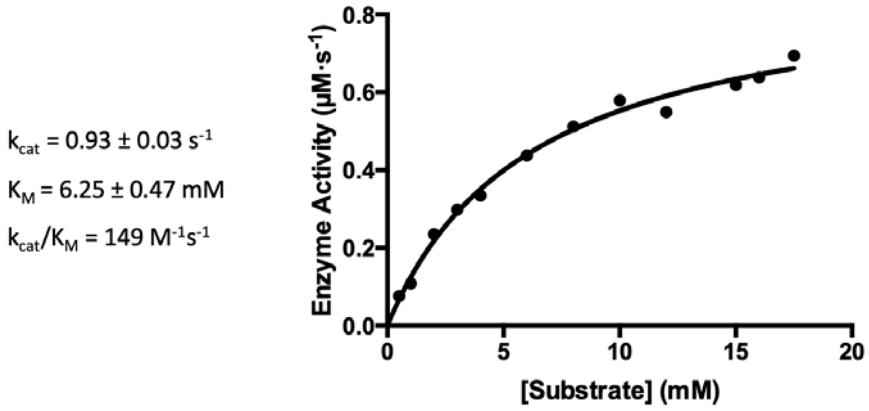


Figure 4.15. Michaelis-Menten kinetic of BsPdaC-CD with $(GlcNAc)_4$ substrate. *Reactions were performed in 50 mM phosphate, 300 mM NaCl, pH 7.0, 37°C. Initial rates of mono-deacetylation were determined by the HPLC-MS method.*

4.2.4.3. Specificity on peptidoglycan

To verify that the truncated *Bs*PdaC-CD maintains the same MurNAc deacetylase activity as the one reported for the full-length enzyme, the specificity was assessed by combining the action of different enzymes on PGN substrate.

First, the specific activity of *Bs*PdaC-FL on *Bacillus subtilis* PGN was determined (as previously described for the truncated form) and showed to be of the same order of magnitude (81.8 min^{-1}, in comparison with 63.1 min^{-1} for the catalytic domain).

PGN was incubated with *Bs*PdaC-FL for an extended period, and the released acetate was quantified. After dialysis to remove the free acetate, the product was incubated with *Bs*PdaC-CD, and it was observed that no further deacetylation took place upon long incubation time. A parallel experiment in which the order of enzymes was inverted (first incubation with *Bs*PdaC-CD, then with *Bs*PdaC-FL) gave the same result. This indicated that both enzymes deacetylate the same residues (MurNAc) of PGN.

In order to have a reference to compare, the final deacetylated PGN products from both experiments were treated with the GlcNAc peptidoglycan deacetylase *Sp*PgdA. The enzyme showed deacetylase activity when incubated with the products, maintaining between 70 and 80% of its maximum activity on intact peptidoglycan (Table 4.5).

Table 4.5. Deacetylation activity of BsPdaC-FL and BsPdaC-CD on Bacillus subtilis PGN*. All reactions were performed at 37°C for 48 hours.*

REACTION	ENZYME/SUBSTRATE	ACETATE RELEASED (μM AcO-/μM E) AT 48 H
1	*Bs*PdaC-CD + PGN	3252
2	*Bs*PdaC-FL + PGN	4010
3	Reaction 1 + *Bs*PdaC-FL	53.6
4	Reaction 2 + *Bs*PdaC-CD	18.4
5	*Sp*PgdA + PGN	3358
6	Reaction 3 + *Sp*PgdA	2680 (80%)[1]
7	Reaction 4 + *Sp*PgdA	2434 (72%)[1]

[1]Percentatge of *Sp*PgdA's activity maintained in comparison with its activity on intact PGN.

The differential behavior exhibited by *Bs*PdaC (both FL and CD) and the GlcNAc peptidoglycan deacetylase *Sp*PgdA indicates that they present a different specificity in the peptidoglycan substrate. After total conversion reactions with both constructs of *Bs*PdaC (which deacetylated the MurNAc residues of peptidoglycan), only *Sp*PgdA (enzyme that deacetylates the GlcNAc residues of this substrate) could further deacetylate the substrate samples.

These results led to the conclusion that the isolated catalytic domain has the same specific activity and retains the same MurNAc specificity as the full-length enzyme.

4.2.4.4. Pattern of deacetylation on COS

HPLC-MS time course monitoring of the deacetylation reaction of *Bs*PdaC catalytic domain on $(GlcNAc)_{3-5}$ revealed that, contrary to the previously reported specificity on the tetrasaccharide [121], the enzyme sequentially deacetylated different GlcNAc units leading to a final product in which all except one GlcNAc unit were deacetylated (Figure 4.16, A).

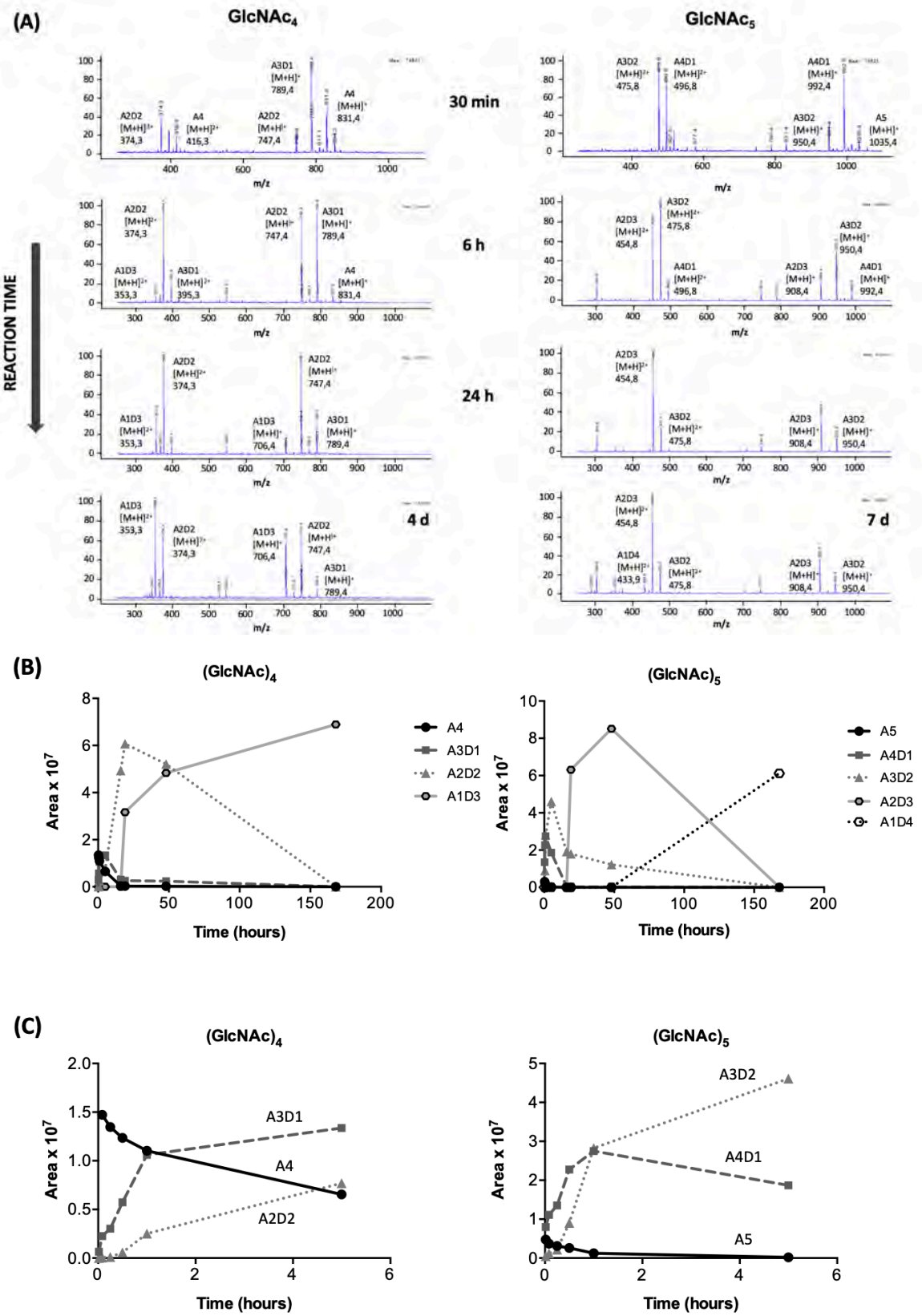


Figure 4.16. MS time monitoring BsPdaC-CD reactions with $(GlcNAc)_4$ and $(GlcNAc)_5$ substrates. *(A) MS spectra at increasing reaction time. Total ion monitoring from SCAN mode (250-1100 m/z scan range) in the 0.9 to 4 minutes time interval, representing the totality of samples; (B) Evolution of peak areas for $[M+H]^+$ ions of the deacetylated products from $(GlcNAc)_4$ and $(GlcNAc)_5$ from SIM (Single Ion Monitoring) mode. Area of the single ions that are monitored for each product; (C) Magnification at short reaction time.*

For the chitotetraose substrate, mono-, di-, and trideacetylated products were sequentially formed (A3D1, A2D2, A1D3, with A being GlcNAc and D being $GlcNH_2$) (Figure 4.16, B). Likewise, with the chitopentaose substrate, A4D1, A3D2, A2D3, and A1D4 were generated along the reaction (Figure 4.16, C). Whereas the final A1D3 product from tetraacetylchitotetraose was detected at rather short reaction times, the final A1D4 product from pentaacetylchitopentaose was only detected after a long reaction time.

In order to know the pattern of acetylation of the different products, several preparative reactions at different reaction times were set up with $(GlcNAc)_4$ and $(GlcNAc)_5$. The structure of the products present on the samples was determined in Dr. Bruno M. Moerschbacher's laboratory in Münster by the MALDI-TOF-MS/MS method designed by Cord-Landwerh and coworkers [123]. The method is based on the detection and identification of COS fragments by UHPLC-MS after their isotopic labelling.

When mixtures of chitin and chitosan oligomers were separated by HILIC UHPLC and detected using ESI-MS, it was revealed that there was a significant difference in the response factors of those oligomers depending on both their DP and DA. In consequence, quantification would require the availability of individual standards for each oligomer.

To overcome the influence of DA on the MS response factor, the oligomers are N-acetylated using $[^2H_6]$ acetic anhydride: the free amino moieties are labelled with this deuterated agent causing glucosamine to convert into $[^2H_3]$N-acetylglucosamine. This reacetylation serves a double objective, in one hand it allows for all molecules with the same DP to elute at the same elution time. On the other hand, it allows the identification of the initially deacetylated residues by the mass increment caused by the re-acetylation with $[^2H_6]$ acetic anhydride.

The second labelling step consists on labelling the reducing end of the oligomers with ^{18}O, which allows for unequivocal discrimination of fragment ions comprising either the reducing end or the non-reducing end (considering two fragments of equivalent acetylation, the fragment coming from the reducing end will show a mass 2 units higher than the one coming from the non-reducing end). This step allows the reconstruction of the molecule's sequence by identifying all the possible fragments from the MS^2 analysis. The PA of the complete molecule can be determined using mainly those fragments with the reducing end labelled.

Double isotopically labelled internal standards are used for quantification and prepared by *N*-acetylation of GlcN oligomers with [$^{13}C_4$, $^{2}H_6$] acetic anhydride, solving the problem of DP influence on the response factor.

As a summary, the method consists of the following four steps, which will allow the identification and quantification of the different products in the reaction mixtures:

(i) A first re-acetylation of free amino groups in chitooligosaccharides using a deuterated agent ([$^{2}H_6$] acetic anhydride).
(ii) Labelling the reducing end with an ^{18}O-tag.
(iii) Quantification using [$^{13}C_2$, $^{2}H_3$]-labelled isotopologues as internal standards.
(iv) Sequencing by tandem MS.

The different patterns of acetylation identified in each sample (corresponding to different reaction times) as well as their relative amount compared to each other are listed on Table 4.6 and the deacetylation pathway followed by the enzyme is represented on Figure 4.17.

Table 4.6. MALDI-TOF-MS/MS analysis of the deacetylation products from $(GlcNAc)_4$ and $(GlcNAc)_5$ by BsPdaC-CD. *PA: pattern of acetylation; A: GlcNAc; R: reacetylated GlcN with [$^{2}H_6$]-acetic anhydre. Procedure according to Cord-Landwehr, 2017.*

$(GlcNAc)_4$					
SAMPLE	**REACTION TIME**	**PA**	**m/z $[M+H]^+$**	**ABUNDANCE (%)**	**PRODUCTS (%) PER TYPE [a]**
DP4.1	30 min	AAAA	831.335	61.0	
		AARA	834.354	19.6	50.3
		ARAA		19.4	49.7
DP4.2	5 h	AAAA	831.335	12.3	
		AARA	834.354	36.0	56.9
		ARAA		27.3	43.1
		ARRA	837.373	24.4	
DP4.3	24 h	AARA	834.354	7.2	27.9
		ARAA		18.7	72.1
		ARRA	837.373	58.8	88.8
		RARA		7.4	11.2
		RRRA	840.392	7.9	
DP4.4	96 h	ARAA	834.354	9.0	
		ARRA	837.373	55.3	89.6
		RRAA		6.4	10.4
		RRRA	840.392	29.3	

[a] % of isobaric products in each sample.

$(GlcNAc)_5$					
SAMPLE	**REACTION TIME**	**PA**	**m/z $[M+H]^+$**	**ABUNDANCE (%)**	**PRODUCTS (%) PER TYPE [a]**
DP5.1	30 min	AAAAA	1034.415	22.9	
		AAARA	1037.434	32.9	50.7
		AARAA		23.0	35.5
		ARAAA		9.0	13.8
		AARRA	1040.452	12.1	
DP5.2	5 h	AARRA	1040.452	54.3	77.4
		ARARA		15.8	22.6
		ARRRA	1043.471	29.9	
DP5.3	24 h	AARRA	1040.452	14.2	45.6
		ARARA, ARRAA[b]		9.2	29.5
		ARRAA, RARAA[b]		7.7	24.9
		ARRRA	1043.471	46.7	81.0
		RARRA		11.0	19.0
		RRRRA	1046.490	11.2	
DP5.4	96 h	ARRAA, RARAA[b]	1040.452	10.8	50.4
		ARARA, ARRAA[b]		10.7	49.6
		ARRRA	1043.471	43.2	
		RRRRA	1046.490	35.2	

[a] % of isobaric products in each sample. [b] Any of both structures, or both.

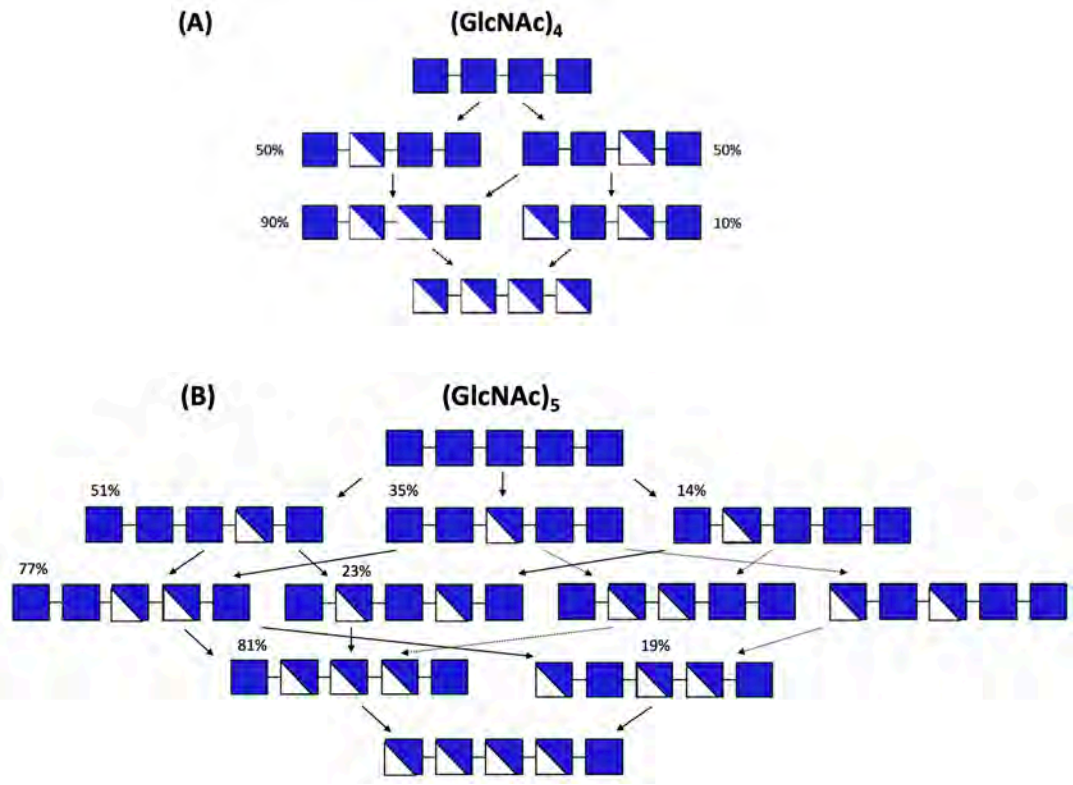


Figure 4.17. BsPdaC-CD pattern of deacetylation. *(A) Deacetylation sequence of $(GlcNAc)_4$. (B) Deacetylation sequence of $(GlcNAc)_5$. Dashed arrow, slower deacetylations at longer reaction times (see Table 4.6).*

PdaC follows a multiple-chain mechanism, deacetylating all residues but the reducing end of COS substrates. With the tetrasaccharide substrate, the first deacetylation occurs at the two central units with equal efficiencies, followed mainly by the deacetylation of the other residue and, finally, of the non-reducing end, being the final reaction product DDDA. With the pentasaccharide substrate, any of the three internal GlcNAc units are first deacetylated but with preference for the second residue from the reducing end (51% as compared with 14 and 35% for the other two internal residues). Deacetylation proceeds in time leading to a final product in which all but the reducing end GlcNAc unit are deacetylated (DDDDA).

The different deacetylation products observed for the first deacetylation event on the pentasaccharide substrate suggest that the binding site cleft has at least 6 subsites, from -3 to +2. Three subsites are enough to explain the observed monodeacetylation pattern on chitotetraose (-1, 0, +1), with two initial binding modes (BM1 and BM2) (Figure 4.18), but additional subsites -2 and +2 may exist, with similar affinities to agree with the fact that both binding modes are alike. With chitopentaose substrate, if -2 and +2 subsites have similar affinities, a new subsite -3 might be required to explain the preferred binding mode BM1' (51%, AAADA) over BM2' (35% AADAA) (Figure 4.18). On the contrary, a potential +3 subsite is not required, since BM3' (14%, ADAAA) is the least preferred binding mode and that may be accounted by the loss of interactions with subsite -2. Binding to subsite +1 is required in any case, since it is always occupied with all substrates. Not only in the first deacetylation events, but also in the subsequent deacetylations, leading to the final products in which the reducing end GlcNAc is not deacetylated.

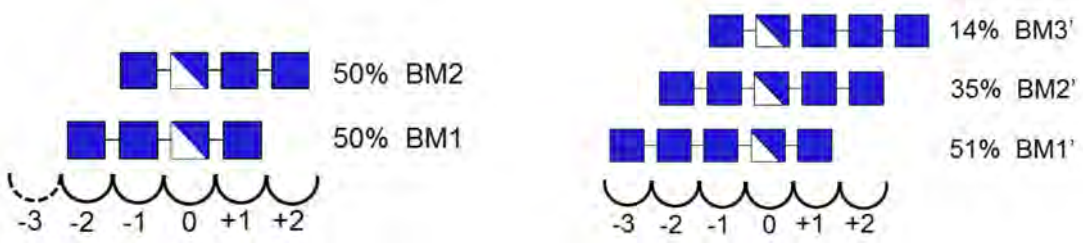


Figure 4.18. Binding of COS ligands to BsPdaC. *Binding modes for the first deacetylation event for substrates $(GlcNAc)_4$ and $(GlcNAc)_5$. Negative subsites on the non-reducing end, positive subsites on the reducing end, subsite 0 is the catalytic site.*

4.2.5. pH and temperature dependence of *Bs*PdaC catalytic domain

The pH dependence of *Bs*PdaC was analyzed using $(GlcNAc)_5$ as substrate in a pH range of 6.0 – 10.6 performing the reactions at 37°C, and the deacetylation activity was measured by the fluorescamine assay to determine initial rates. Maximum activity was registered at pH 8.5, although the enzyme remained active in the 7 – 10 pH range (Figure 4.19, A).

Temperature dependence of the enzyme was analyzed following the same procedure in a temperature range of 25 – 75°C at pH 7.0. Although the enzyme remained active in a wide range of temperature, an environment with a temperature lower than 37°C or higher than 60°C results in a drastic loss of activity. In this case, maximum activity was registered at 55°C (Figure 4.19, B).

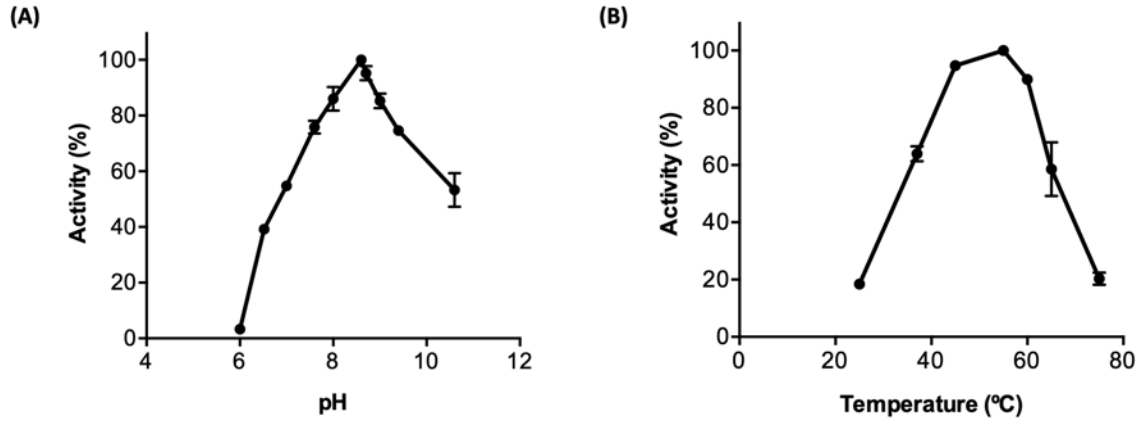


Figure 4.19. pH and temperature profiles of BsPdaC-CD. *(A) pH. (B) Temperature. Activity determined with $(GlcNAc)_5$ in 50 mM phosphate, 300 mM NaCl, at 37°C (for the pH profile), pH 7.0 (for the temperature profile).*

4.2.6. Metal dependence of *Bs*Pdac catalytic domain

One of the key elements of the deacetylation mechanism of CE4 family is the metal cation that is part of the active center. It has been extensively described that this family of enzymes follow a metal dependent mechanism regardless of the substrate. The characterization of different deacetylases has revealed that the divalent cation present on the catalytic site can vary from one enzyme to another, and several studies on preference for different metal ions have been performed. A list of the described metals is presented in Table 4.7, Zn^{2+} and Co^{2+} being the most prevalent cations. It has been observed that, in some cases, the metal that enhances one enzyme's activity acts as an inhibitor of the catalytic capacity of another, a fact that shows the importance of the precise determination of the optimal metal that allows to develop a greater efficiency to each enzyme.

When the initial characterization of the enzyme was reported, Kobayashi and coworkers suggested that *Bs*PdaC might not be a metallo-enzyme since it retained 50% of its activity on PGN and 70% on $(GlcNAc)_4$ after treatment with a chelating agent [121]. However, as shown in the first section of this chapter, PdaC shares the conserved metal-binding triad residues with the rest of the family and it would be expected for *Bs*PdaC to act following the same metal-dependent mechanism.

Table 4.7. Reported metal cations in CE4 enzymes active sites. *In bold: enzymes with more than one reported metal cation. [1]No evidence for native metal but indicated the metal from purification/crystallization experiments. Adapted from* [47].

ACTIVITY	ENZYME	METAL
Chitin deacetylases	*Mr*CDA	Zn^{2+}
	*Cl*CDA	Zn^{2+}
	*P*aCDA	Zn^{2+}
	*Vc*CDA	Zn^{2+}
	*Vp*CDA	Zn^{2+}
	*Ar*CE4	Ni^{2+}
	NodB	Mn^{2+}, Mg^{2+}
Peptidoglycan deacetylases	*Sp*PgdA	Zn^{2+}
	*Sm*PgdA	Zn^{2+}
	*Er*PgdA	Zn^{2+}
	*Ba*Pda	Zn^{2+}
	*Bc*Pgd	Co^{2+}
	BsPdaA[1]	Cd^{2+}
	BC0361	Zn^{2+}
	BA0330	Zn^{2+}
Poly-ß-1,6-N-acetyl-D-glucosamine deacetylase	***Er*PgdAB**	Zn^{2+}, Co^{2+}, Ni^{2+}
	***Ad*IcaB**	Zn^{2+}, Co^{2+}, Ni^{2+}
	*Aa*PgaB	Zn^{2+}

In order to further study the dependence of *Bs*PdaC activity on metal cations, various treatments were tested to obtain the apo enzyme (protein without the metal ion in the active site) for subsequent reconstitution experiments. Initial treatment with 20 mM of the chelating agent EDTA (ethylenediaminetetraacetic acid) only reduced the activity by 10% relative to untreated and freshly purified enzyme, indicating that the metal cation is strongly bound to the active site. Hence, more extreme conditions, which included increasing concentrations of urea in the presence of EDTA, were used. Urea is a protein denaturing agent [124] and it was added to unfold the protein and increase exposure to EDTA. After treatment with 200 mM EDTA in 2 M urea, the recovered protein after dialysis still retained 65% activity and full activity could be restored upon addition of Zn^{2+} cation. The enzyme treated with 7 M urea and 200 mM EDTA was almost completely inactivated, showing that these conditions were required to fully dissociate the metal cation. Protein refolding was quantitative, resulting in an inactive protein that recovered 100% activity upon addition of Zn^{2+} cation (Figure 4.20).

In order to asses that Zn^{2+} was fully removed with the protocol followed to obtain the apo enzyme, the metal content of the sample was checked by ICP-MS, obtaining that $[Zn^{2+}]/[E]$ was 0.1, meaning that approximately 90% of the bound metal was removed in the chosen conditions.

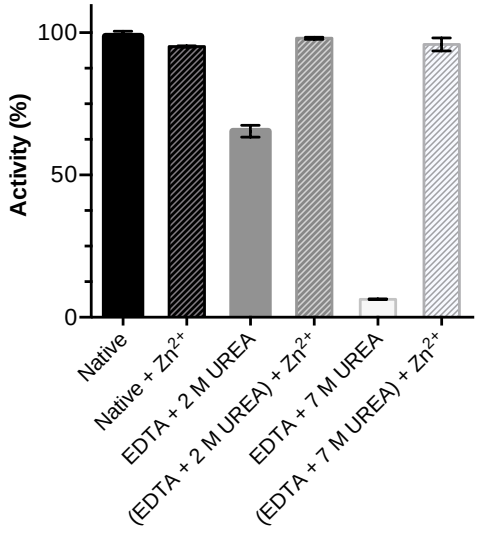


Figure 4.20. Metal dependence of BsPdaC *activity. Residual activity upon treatment with EDTA (200 mM) and urea (2 or 7 M), and reconstitution with Zn^{2+} cation.*

4.2.7. Enzymatic *N*-acetylation of COS by *Bs*PdaC catalytic domain

In 1999, it was reported that the fungal chitin deacetylase from *Colletotrichum lindemuthianum* (*Cl*CDA) was capable of also performing a reverse reaction and acetylate substrates in the presence of high concentrations of sodium acetate [125]. Recently, the ability of several bacterial and fungal CDAs to *N*-acetylate chitosan oligomers (performing the reverse of their natural function of deacetylation) was studied [119]. All tested CDAs showed to be able to *N*-acetylate chitosans and, moreover, to do so while keeping their regio-selectivity. Namely, CDAs that showed a certain specificity for one or more residues of the chitin oligomers substrates when performing its natural deacetylation reaction retained this specificity in the acetylation reactions on chitosan oligomers, generating their reverse product.

As already described, *Bs*PdaC follows a multiple chain mechanism and deacetylates all residues but the reducing end of chitin oligomers, generating the final products DDDA and DDDDA (where D is $GlcNH_2$ and A is GlcNAc). None of the CE4 enzymes active on chitin oligomers characterized to date naturally produces a partially acetylated chitosan oligomer with only the reducing-end unit deacetylated. Considering *Bs*PdaC specificity and the described behavior of other deacetylases of the family, it could be possible for the enzyme to generate the AAAD or AAAAD products from totally deacetylated substrates by performing the reverse *N*-acetylation reaction.

To study this possibility, the time course of long-term reaction with *Bs*PdaC catalytic domain and a totally deacetylated pentasaccharide in the presence of sodium acetate was monitored by HPLC-MS. DP5 substrate (2 mM) was incubated at 37°C with *Bs*PdaC (5 µM) in PBS buffer at pH 7 and 2.5

mM NaAc. Several acetylation products were identified, and their formation and depletion could be followed along the reaction progress (Figures 4.21 and 4.22). The final product of the *N*-acetylation reaction was D1A4 (with a single deacetylated unit), without the presence of further acetylation product. These results showed that PdaC acts in all of the substrate's residues except for one, in the same way as the enzyme does when performing its natural deacetylation reaction. If the enzyme's specificity is maintained in the reverse reaction in the same way that has been reported for other enzymes of the family, the unit in which the enzyme is not acting would be the reducing-end one.

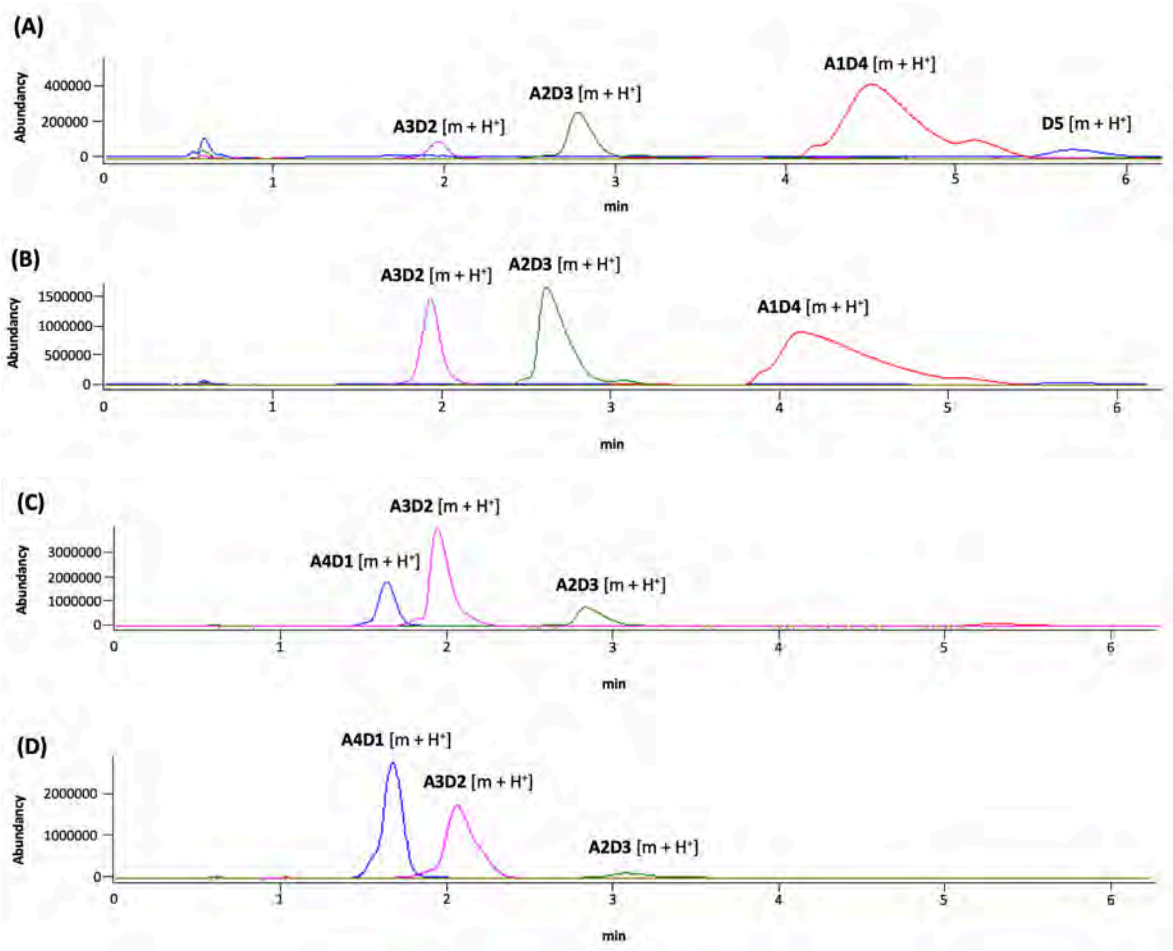


Figure 4.21. SIM (Single Ion Monitoring) HPLC-MS analysis of N-acetylation of D5 with BsPdaC-CD. *Products evolution through reaction time: (A) 1 hour, (B) 3 hours, (C) 24 hours, (D) 138 hours.*

However, with the performed analysis, it is not possible to know the specific sequence of the obtained products and sequencing of their pattern of acetylation would be needed to confirm that the last product is the expected AAAAD.

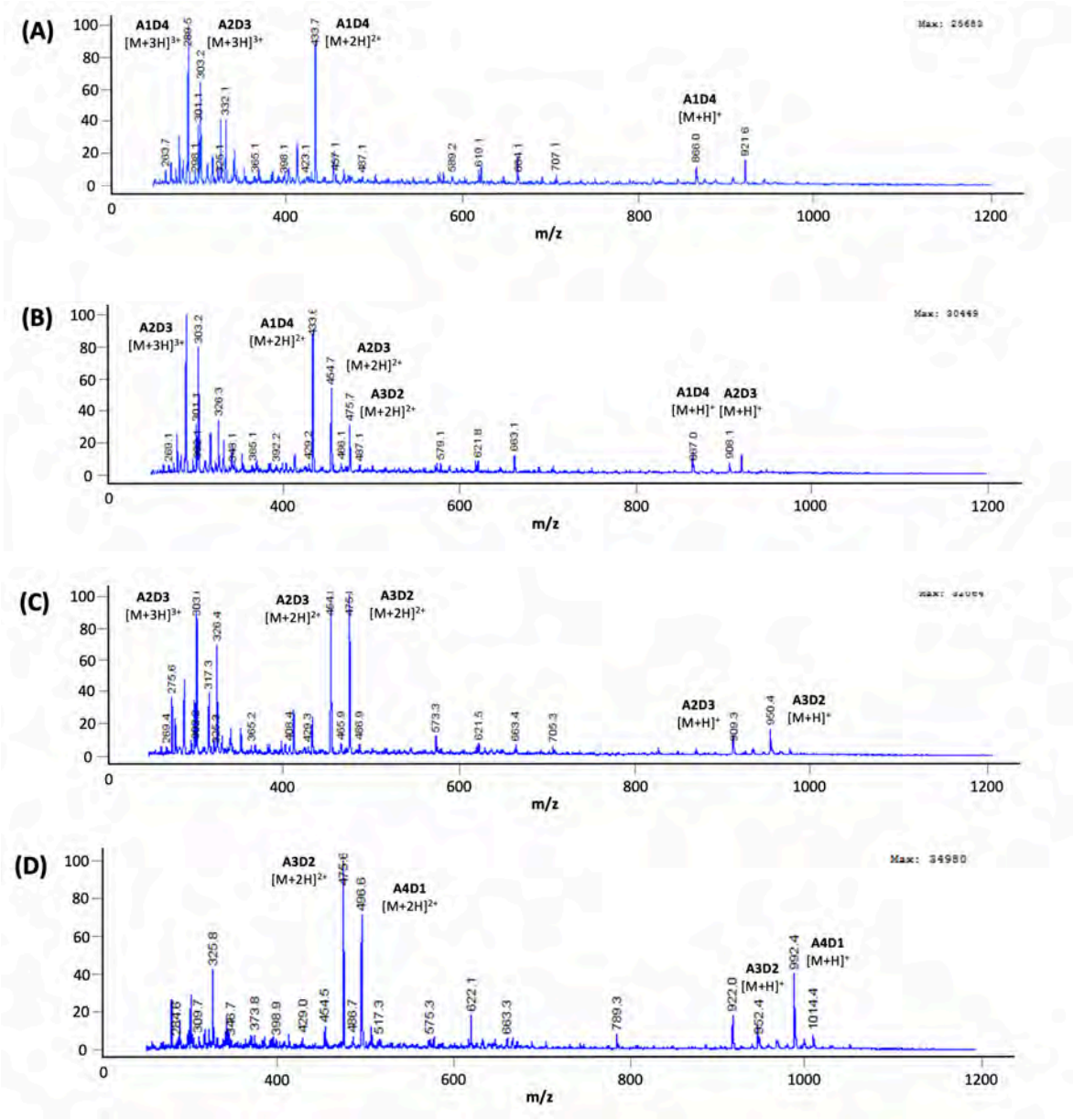


Figure 4.22. SCAN HPLC-MS analysis of N-acetylation of D5 with BsPdaC-CD. *MS spectra at increasing reaction time. Total ion monitoring from SCAN mode (250-1100 m/z scan range) in the 0.9 to 4 minutes time interval, representing the totality of samples. (A) 1 hour, (B) 3 hours, (C) 24 hours, (D) 138 hours.*

Although the final D1A4 product was generated by enzymatic *N*-acetylation with PdaC, intermediate products were always present in the samples corresponding to the endpoint of the reactions. With the aim of obtaining the pure final product, and not a mixture with varying degrees of acetylation, different conditions were tested to optimize the reaction (Table 4.8).

(i) Addition of fresh *Bs*PdaC enzyme at several time points of the reaction

- In order to obtain the final mono-deacetylated product, new enzyme was added after 120 and 192 hours of reaction (Figure 4.23).

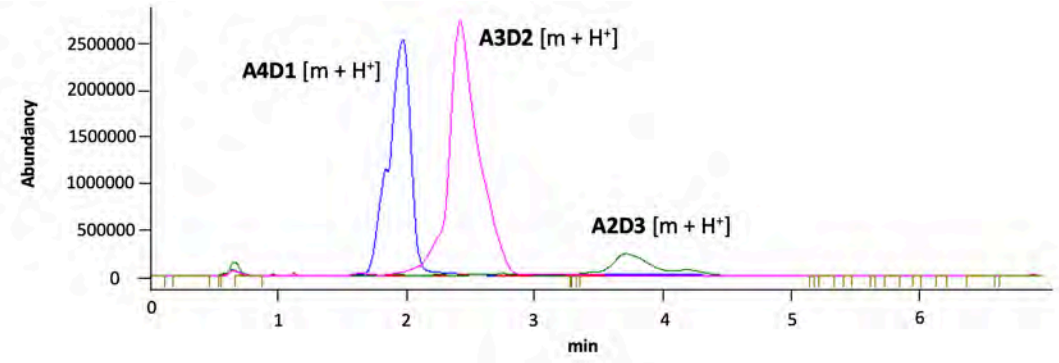


Figure 4.23. SIM (Single Ion Monitoring) HPLC-MS analysis of N-acetylation of D5 with BsPdaC-CD after addition of fresh enzyme.

(ii) Performing the reaction at pH 8.

- The pH of the reaction was increased to be above the pKa of the free amino groups of the substrate/products to avoid products to be protonated and favor the advance of the acetylation reaction (Figure 4.24). The pH at which the reaction was performed (pH 8) is within the action range of *Bs*PdaC, that shows even higher activity in these conditions than at pH 7.

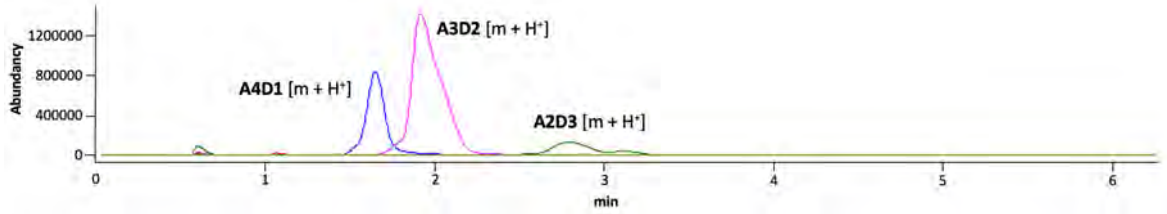


Figure 4.24. SIM (Single Ion Monitoring) HPLC-MS analysis of N-acetylation of D5 with BsPdaC-CD performing the reaction at pH 8.

(iii) Increasing the concentration of sodium acetate from 2.5 mM to 3.5 mM (Figure 4.25).

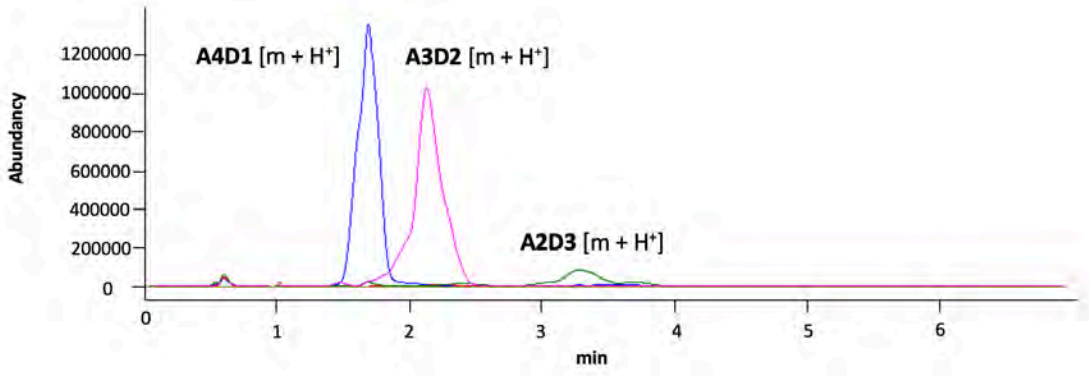


Figure 4.25. SIM (Single Ion Monitoring) HPLC-MS analysis of N-acetylation of D5 with BsPdaC-CD performing the reaction with 3.5 mM NaOAc.

(iv) Performing the reaction with *Bs*PdaC catalytic domain R315S mutant.

- It is possible that the total conversion to the final product is not achieved because the enzyme deacetylates the acetylated products that have been generated during

the reaction. The R315S mutant (described in a later section, see Section 4.4) shows a different preference for deacetylation than the wild type enzyme and it could have less tendency to deacetylate those acetylated products (Figure 4.26).

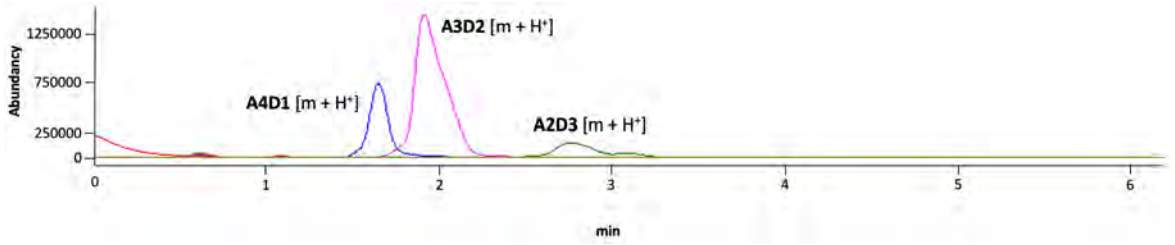


Figure 4.26. SIM (Single Ion Monitoring) HPLC-MS analysis of N-acetylation of D5 with BsPdaC-CD-R315S.

(v) Performing a second reaction with the obtained products.

- The obtained acetylated products were separated from the enzyme by using an ultra-centrifugal filter unit and were used as substrate for a new reaction with fresh enzyme and new sodium acetate buffer (Figure 4.27).

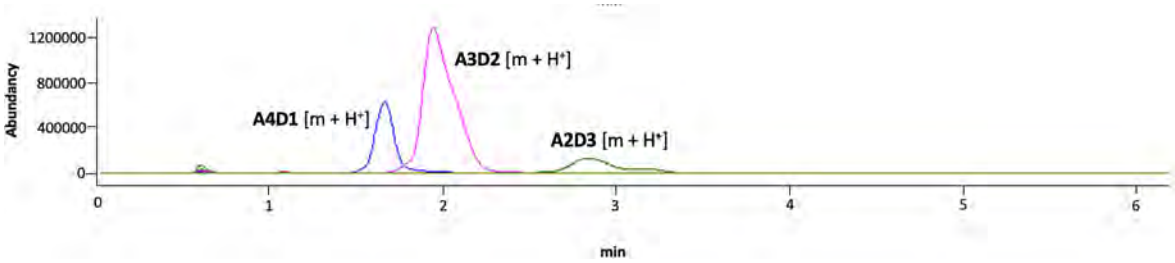


Figure 4.27. SIM (Single Ion Monitoring) HPLC-MS analysis of N-acetylation of the products of a first N-acetylation reaction with BsPdaC-CD.

Table 4.8. Optimization conditions for N-acetylation reaction with BsPdaC.

EXPERIENCE	CONDITIONS
(i)	*Initial*: 2 mM DP5, 5 µM *Bs*PdaC-CD, PBS pH 7 2.5 mM NaAc, 37°C
	After 120 hours: addition of **5 µM *Bs*PdaC-CD** in PBS pH 7 2.5 mM NaAc
	After 192 hours: addition of **5 µM *Bs*PdaC-CD** in PBS pH 7 2.5 mM NaAc
(ii)	2 mM DP5, 5 µM *Bs*PdaC-CD, PBS **pH 8** 2.5 mM NaAc, 37°C, up to 168h
(iii)	2 mM DP5, 5 µM *Bs*PdaC-CD, PBS pH 7 **3.5 mM NaAc**, 37°C, up to 168h
(iv)	2 mM DP5, 5 µM ***Bs*PdaC-R315A**, PBS pH 7 2.5 mM NaAc, 37°C, up to 168h
(v)	*1st reaction:* 2 mM DP5, 5 µM *Bs*PdaC-CD, PBS pH 7 2.5 mM NaAc, 37°C, up to 168h
	2nd reaction: 1st reaction products, 5 µM *Bs*PdaC-CD, PBS pH 7 2.5 mM NaAc, 37°C, up to 72h

The results were the same for all the experiments. Although the initial deacetylated substrate was not present in the final samples, total conversion to the final mono-deacetylated product could not

be achieved by any means, the di-deacetylated product was also present in all the final samples (Figures 4.23 to 4.27). It is possible that the re-acetylated products are deacetylated again and that there is an equilibrium of the products due to the double action of *Bs*PdaC. Due to the lack of standards for all deacetylation products, it is not possible to calculate the final concentration of the major product. However, the reaction with increased acetate concentration was the one that showed the better yield, with approximately 57% of mono-deacetylated product from the total ions' abundancy in the mass spectrum of the sample.

Several methods based on liquid chromatography have been described for the separation of chitooligosaccharides with varying degree of polymerization (DP). Separation between totally acetylated oligomers and products with some degree of acetylation is also possible due to the positive charge of the free amino groups on deacetylated COS. However, the isolation of products with the same DP but different degree of acetylation (DA) in a preparative way is very difficult.

4.3. Structure of *Bs*PdaC

Within the context of searching for new biocatalysts for the production of novel and defined chitosan products, and of studying the structure-function relationship of specificity on CE4 family of enzymes, one of the main goals of this work was to solve the crystal structure of the peptidoglycan deacetylase PdaC from *Bacillus subtilis*. Furthermore, having this new structural information could shed some light in the structural determinants of the enzyme's unique dual activity on GlcNAc and MurNAc residues.

4.3.1. Inactive mutants for co-crystallization

In order to perform co-crystallization assays with substrates, the inactive mutants of both the full-length and the catalytic domain enzymes were prepared by mutating the catalytic aspartate to a serine (D285S). The mutants were generated using the Quick-Change PCR methodology (see Section 7.1.4) with *Bs*PdaC full-length and *Bs*PdaC catalytic domain constructs as templates, and their sequences were confirmed by Sanger sequencing.

Expression in *E. coli* and purification of both recombinant mutants were performed following the same procedure previously described for the wild type enzymes. Similar expression yields were obtained as well as the same characteristic chromatographic profile with three oligomeric fractions and comparable protein purity (Figures 4.28 and 4.29).

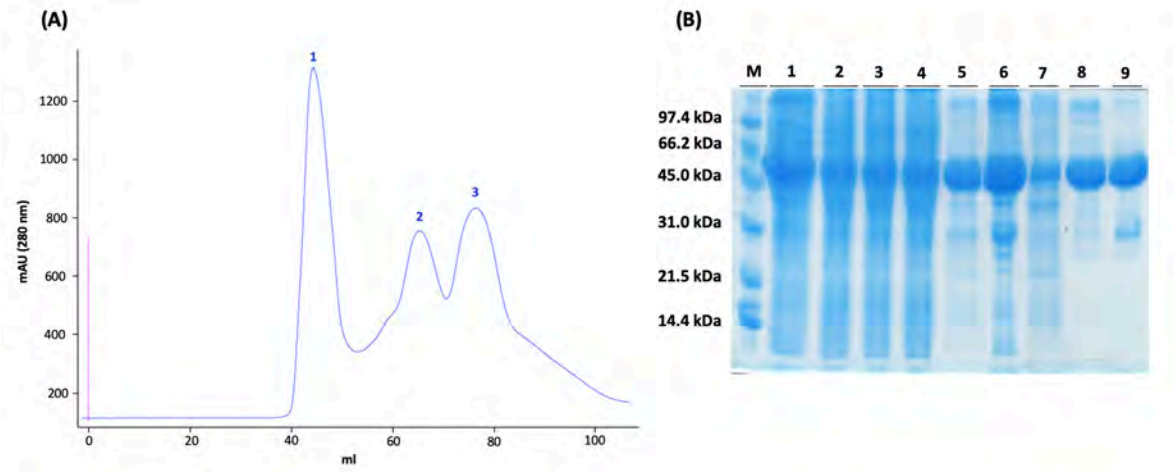


Figure 4.28. Expression and purification of BsPdaC-CD-D285S*. (A) Size Exclusion Chromatography profile (Superdex 200 (16/600)). (B) SDS-PAGE analysis. M: Marker Low Range (Biorad); 1: Lysate; 2-4: Flow through from Strep-trap column loading and washing; 5: Strep-trap elution; 6: Strep-trap elution concentrated; 7: SEC fraction 1; 8: SEC fraction 2; and 9: SEC fraction 3.*

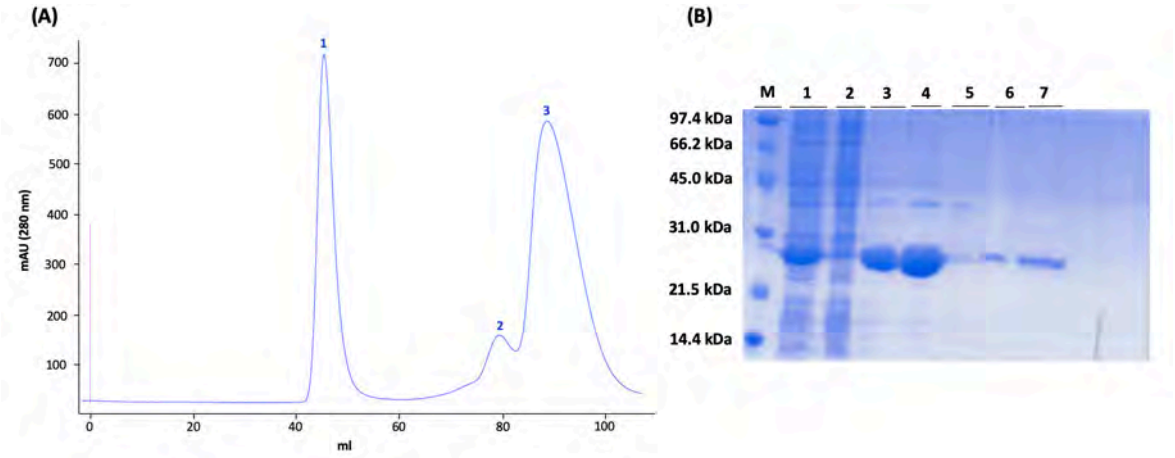


Figure 4.29. Expression and purification of BsPdaC-FL-D285S*. (A) Size Exclusion Chromatography profile (Superdex 200 (16/600)). (B) SDS-PAGE analysis. M: Marker Low Range (Biorad); 1: lysate; 2: flow through from Strep-trap column loading and washing; 3: Strep-trap elution; 4: Strep-trap elution concentrated; 5: SEC fraction 1; 6: SEC fraction; and 7: SEC fraction 3.*

In order to rapidly evaluate the effect of the amino acid change into the mutated proteins, their activity against the unspecific substrate AcOMU (4-methylumbelliferyl acetate) was measured. All the obtained fractions from SEC chromatography of both enzymes were tested for deacetylase activity as previously described (see Section 4.2.4.2 and Section 7.4.1). They showed no deacetylation activity, confirming that the mutation of the catalytic base translated into the inactivation of the proteins (Figure 4.30).

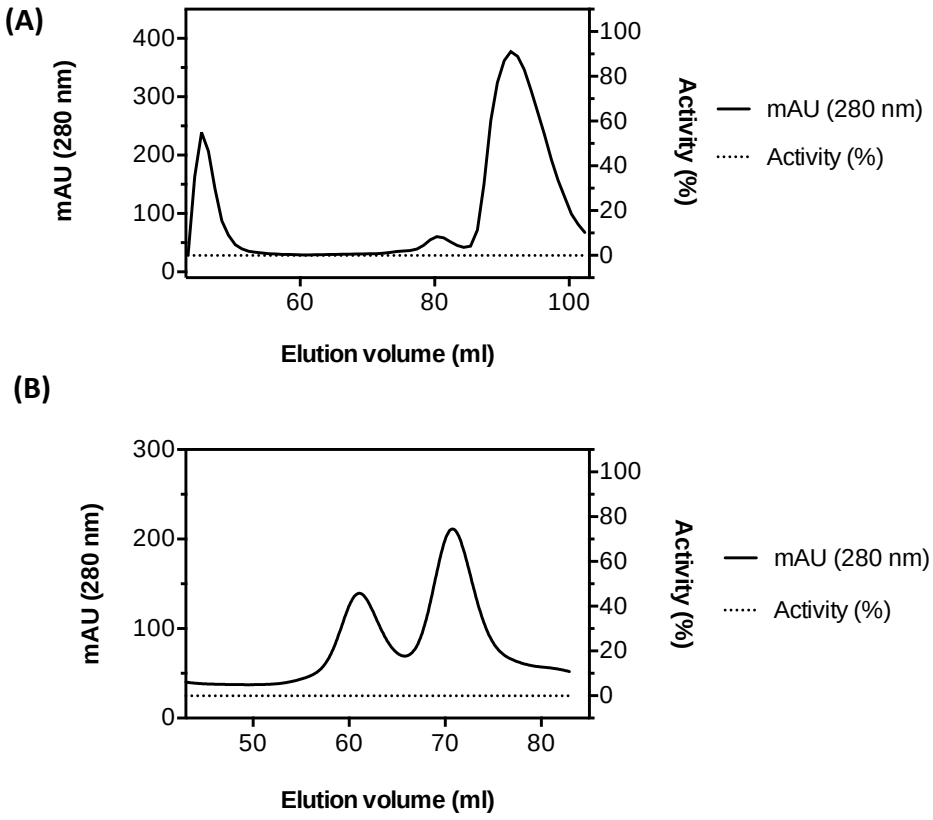


Figure 4.30. Activity profile of SEC fractions against the substrate AcOMU. *(A) BsPdaC-FL-D285S. (B) BsPdaC-CD-D285S.*

Although several crystallogenesis assays were set up with the full-length protein, both with the wild type enzyme and with the inactive mutant in presence of $(GlcNAc)_4$ and $(GlcNAc)_5$ substrates, all attempts were unsuccessful, and no crystals were obtained for this construct.

4.3.2. Crystal structure of *Bs*PdaC catalytic domain

In the case of the truncated protein, the first 3D structure of *Bs*PdaC catalytic domain, which crystallized as a homodimer, was obtained by X-ray crystallography at 1.54 Å resolution[2] (Figure 4.31). The structure of the inactive mutant *Bs*PdaC-CD D285S was also obtained and it superimposes with that of the wild-type enzyme with a RMSD of 0.27 Å. X-ray data collection and refinement statistics are given in Annex 9.3. Both structures were deposited to the Protein Data Bank, with accession codes 6H8L and 6H8N, respectively.

[2] Crystallogenesis and resolution of the X-ray structure was done by Maria Ángela Sanz Polo in Marcelo Guerin's group (Structural Biology lab) at CIC bioGUNE.

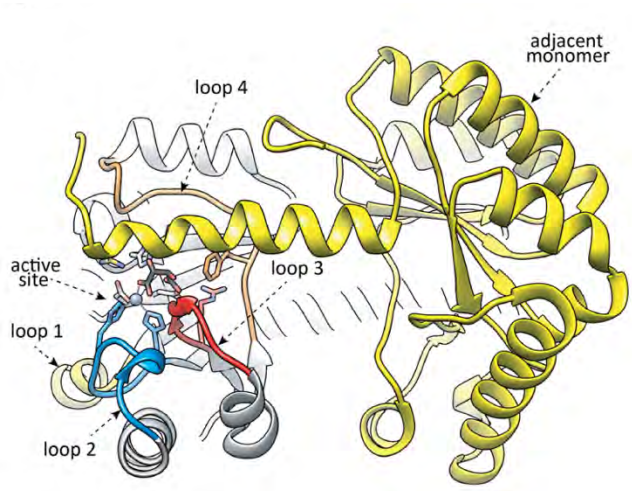


Figure 4.31. Overal structure of BsPdaC-CD. *Crystallographic dimer, one monomer in grey and the other one in yellow.*

The protein adopts the characteristic $(\beta/\alpha)_7$ topology of CE4 enzymes [24], [25], [47]. The central core comprises seven parallel β-strands that form a distorted β-barrel surrounded by α-helixes, with the putative ligand binding site located in the central region of the β-barrel, flanked by a series of connecting loops (Figure 4.32).

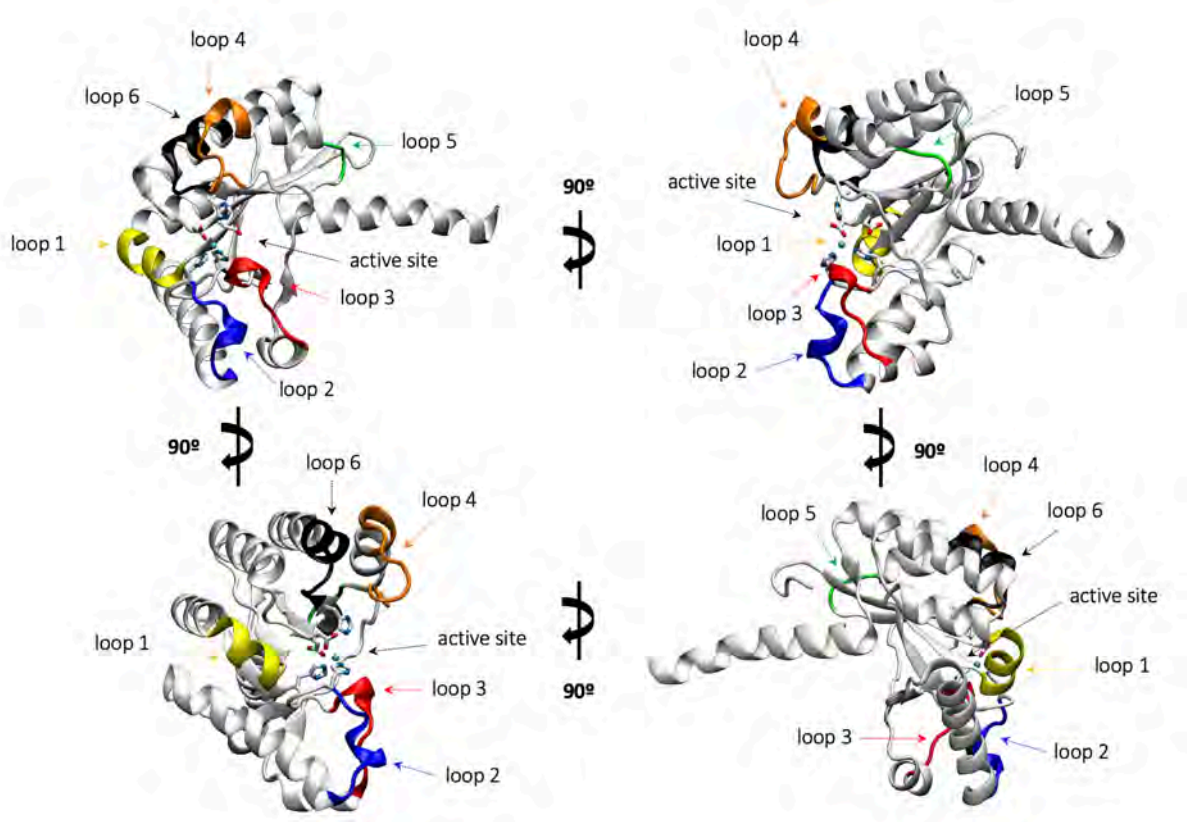


Figure 4.32. BsPdaC-CD 3D structure. *Representation of the overall structure of one monomer, showing the loops (L1-L6) surrounding the active site.*

Another conserved feature of the family that can be observed in the structure is the presence of a Zn^{2+} ion which is coordinated with Asp286, His336, and His340 (conserved metal-binding triad), and with two oxygens of a tartaric acid molecule from the crystallization buffer (Figure 4.33). The conserved catalytic residues, general base (Asp285) and general acid (His427), previously identified in the multiple sequence analysis, are properly oriented in the active site as in other CE4 crystal structures.

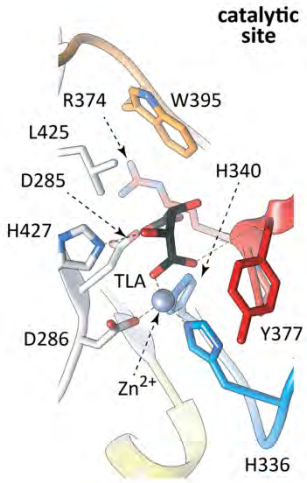


Figure 4.33. BsPdaC-CD 3D structure*. Active site of BsPdaC-CD showing the catalytic residues (D286-H336-H340), a Zn^{2+} ion and carboxylate oxygens of a tartaric acid molecule (TLA) from the crystallization buffer.*

As previously described, CE4 enzymes share a highly conserved core but differ in a series of surface loops that shape their binding cleft. These loops are proposed to be structural elements involved in substrate specificity (Subsite Capping Model) [24], [120].

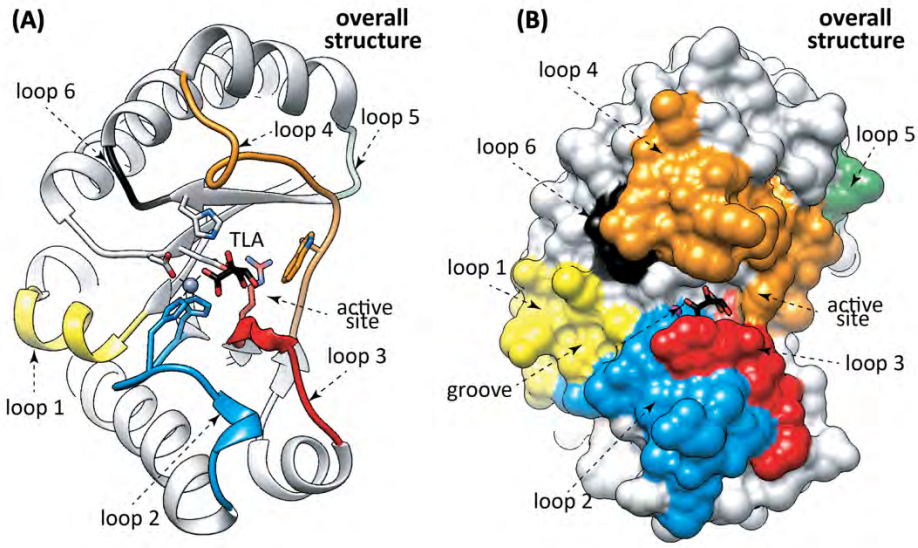


Figure 4.34. BsPdaC-CD 3D structure. *(A) Cartoon representation of the overall structure of BsPdaC-CD showing the loops (L1-L6) surrounding the active site of the CE4 domain. (B) Surface representation of BsPdaC-CD showing the COS binding site. The C-terminus α-helix is not represented.*

These loops were previously assigned in the PdaC sequence according to sequence comparison with CE4 members with known structure. Those locations were confirmed with the crystal structure, in which these structural elements are conformed by almost the same amino acids initially detailed. Loops 1 to 6 (residues 313-320 for loop 1, 341-348 for loop 2, 376-382 for loop 3, 401-410 for loop 4, 419-421 for loop 5 and 428-437 for loop 6) are colored in the 3D structure (Figure 4.34).

The C-terminus contains an α-helix that protrudes from the core of each monomer and interacts with the adjacent monomer, partially covering the active site (Figure 4.35). The helix (from Gly448 to the end of the solved structure) is composed by the final 20 amino acid residues of the protein sequence, plus 5 residues coming from the restriction site sequence used for subcloning and 2 amino acids from the Strep-tag II peptide used for purification (the last 6 C-terminal residues of the Strep-tag are disordered and not seen in the electron density map). The helix of one monomer runs parallel to the substrate binding cleft of the other and the side chain of one of its residues, Arg471, protrudes perpendicularly from the helix axis into the active site of the adjacent monomer, where its guanidyl group interacts with the general acid (His427), the metal coordination aspartate (Asp286) and a tartaric acid molecule from the crystallization liquor. It is thought that the orientation of this terminal helix is a consequence of the crystal packing and it does not represent the protein folding in solution, where the enzyme is fully active. This dimerization contact in the x-ray structure would block access of the substrate to the active site.

Attempts to co-crystallize the inactive mutant with $(GlcNAc)_4$ and $(GlcNAc)_5$ substrates were unsuccessful. The fact that no enzyme-substrate complex was obtained was probably due to the presence of the extended C-terminus α-helix.

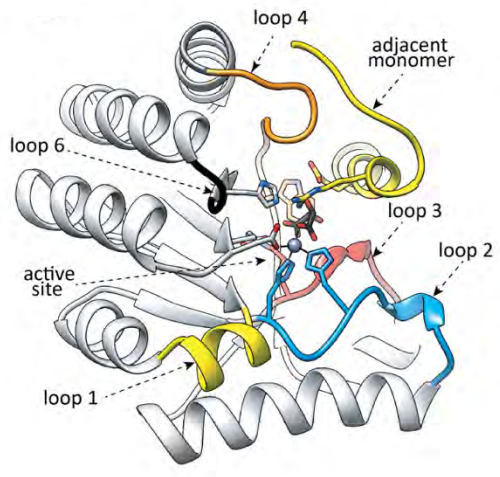


Figure 4.35. BsPdaC-CD 3D structure. *Detail of the dimer interface, in which the side chain of R471 in the C-terminus α-helix from one monomer is inserted into the active site of the adjacent monomer.*

4.3.3. Enzyme·ligand complexes by molecular docking simulations

4.3.3.1. Computational dockings with $(GlcNAc)_4$

Computational dockings of $(GlcNAc)_4$, the most characterized COS substrate for PdaC, on the x-ray structure of the catalytic domain were performed with Autodock Vina (see Section 7.5.3), and binding modes with the lowest energy were considered for analysis.

The results showed two preferred binding modes, with lowest energies (Figure 4.36), in which either of the two internal GlcNAc residues are located in subsite 0 (catalytic site) and would render the AADA and ADAA products. Both binding modes are of similar energy, indicating 50% population each, which is in agreement with the observed enzyme specificity for the first deacetylation event in this substrate as previously showed (see Figure 4.17 on Section 4.2.4.4).

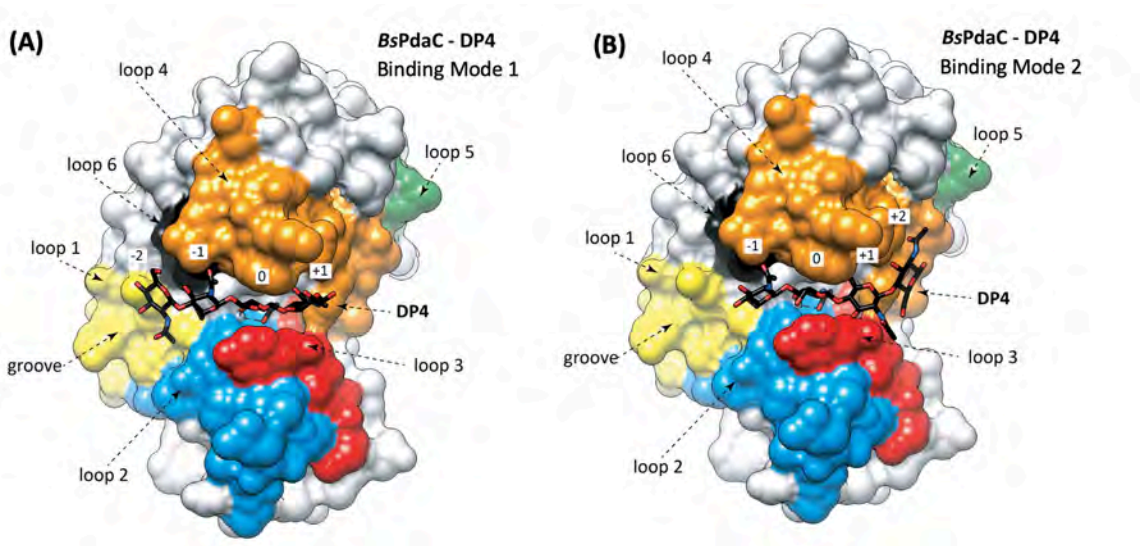


Figure 4.36. Docking simulation of BsPdaC-CD-$(GlcNAc)_4$ complexes. *(A) Binding mode 1 (AAAA). (B) Binding mode 2 (AAAA), where A is the GlcNAc residue in subsite 0 (catalytic site). Structures of minimum energy from the ensemble of docked strutures in each minimum. The C-terminus α-helix is not represented.*

A study of protein-ligand interactions from the obtained docked complexes was performed for each binding mode. The tetrasaccharide substrate expands from subsite -2 to +2, and all GlcNAc units interact with protein residues through hydrogen bonding and stacking interactions (Figure 4.37). A subsite -3 is not seen in these models, but additional interactions may occur with longer and more reactive substrates.

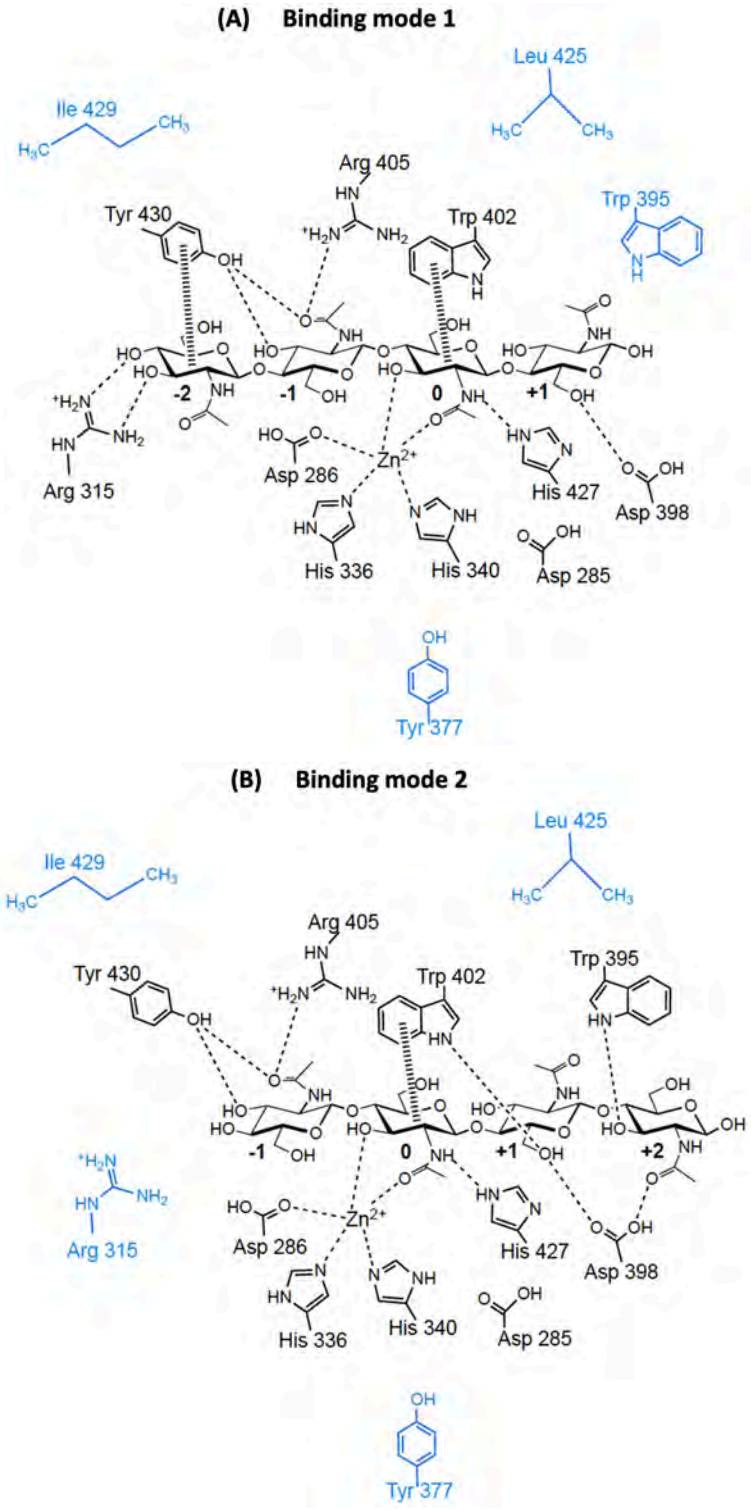


Figure 4.37. Residues from binding cleft that are conserved within CE4 family or that interact with susbtrate in BsPdaC. *Residues in black: protein-ligand interactions in BsPdaC from docking simulations with $(GlcNAc)_4$. Residues in blue: residues that do not establish interactions with the substrate: (A) for binding mode 1 (AAAA); (B) for binding mode 2 (AAAA), where A is the GlcNAc residue in subsite 0 (catalytic site).*

Some of the interacting residues, besides the evident catalytic residues (Asp285 and His427) and the metal-binding triad (Asp286, His336 and His340), are conserved within CE4 family: Asp398, shared by all CDAs, some peptidoglycan deacetylases and the acetylxylan esterase *Sl*AxeA; Trp 395,

conserved in all peptidoglycan deacetylases and most CDAs as well as in *Sl*AxeA; and Trp402, highly conserved in the family with very few exceptions.

If a more exhaustive analysis of the sequences within the whole family is performed, these amino acids (along with other examples) can be described as residues that mediate specific enzyme-substrate interactions related to the different substrate specificities (Table 4.9). These examples show the importance of not only the size and shape of the loops that form the active site, but also of their topology, dynamics and sequence.

Table 4.9. Conserved residues that mediate specific enzyme-substrate interactions in CE4 family. *Adapted from* [47]**.** *In blue: residues that are conserved in BsPdaC sequence but that do not establish interactions with substrate according to docking simulations with* $(GlcNAc)_4$**.**

RESIDUES	
Tyr (MT3 RXP**Y**) (Tyr377 in *Bs*PdaC)	A tyrosine located in **Loop 3** establishes hydrophobic interactions with the substrate residue located at subsite 0. This tyrosine or another aromatic residue can always be found in all the reported CE4 enzymes active on COS except for the MurNAc peptidoglycan deacetylases, which present an arginine at this position.
Trp + Leu (MT5 **L**xHD) (Trp395 and Leu425 in *Bs*PdaC)	A tryptophan and a leucine residue located at **subsite +1** define a hydrophobic pocket that accommodates the methyl group of the GlcNAc residue at subsite 0 and the methylene unit of the hydroxymethyl group of the GlcNAc residue at subsite +1. These two residues are conserved in the family except for some fungal enzymes and poly-β-1,6-GlcNAc deacetylases.
Asp (MT4 **D**xxD) (Asp398 in *Bs*PdaC)	The first aspartate residue of **MT 4** establishes a hydrogen bond with the hydroxyl of the hydroxymethil of the GlcNAc residue at subsite +1 and a water mediated hydrogen bond with the catalytic histidine. The polarity of this residue is conserved except for the case of MurNAc peptidoglycan deacetylases, which present a leucine in that position.
Trp (MT4 DxxD(**W**/Y)) (Trp402 in *Bs*PdaC)	A tryptophan located in **loop 4** establishes a CH-Π interaction with the substrate residue located at subsite 0. The presence of an aromatic residue in this position is highly conserved in the family, with very few exceptions.
Phe (Ile429 in *Bs*PdaC)	A phenylalanine residue located in **loop 6** establishes interactions with the sugar moiety at -1. This is a well conserved hydrophobic position in the family with the exception of poly-β-1,6-GlcNAc deacetylases, which present a polar substituent.

In general, the reducing end of the substrate is not deacetylated or is the least reactive unit for most of CE4 enzymes active of COS. In the structure of *Ar*CE4 in complex with DP2, there is a stacking interaction between the tryptophan in loop 4 (MT4, DxxD(**W**/Y) and the sugar unit in the +1 subsite. As stated before, this aromatic residue is highly conserved and the enzymes that contain it seem to prefer a sugar bound in the +1 subsite. These enzymes either do not deacetylate the reducing end (*Ar*CE4, *Pes*CDA, *Pc*CDA) or it is the slowest position to be deacetylated (*Cl*CDA, AnCDA). *Vc*CDA, and related *Vibrio* and *Shewanella* CODs have the equivalent aromatic residue in a slightly different position after a two-aminoacid insertion in MT4. The residue is located further in Loop 4 and when this loop moves from the open to the closed conformation upon substrate binding, the Trp establishes a stacking interaction, in this case, with the GlcNAc in subsite 0. The preferred substrate of this group of enzymes is DP2 and it is deacetylated at the reducing end. Similarly, the enzyme *Pgt*CDA, which lacks the +1 aromatic residue, deacetylates the reducing end of chitooligosaccharides with a DP ranging from DP4 to DP6.

Although it is possible to make this kind of observations and to relate some key residues or features to differences in substrate preference or specificity, there is still the need of gaining further structural information of enzyme-substrate complexes of CE4 enzymes for better understanding their functionality.

4.3.3.2. Fragment-based docking experiments

*Bs*PdaC presents a dual activity, acting on MurNAc residues of its natural substrate peptidoglycan and also on GlcNAc residues of chitin oligomers, not previously reported for any other peptidoglycan MurNAc deacetylase (note, however, that only PdaA, one MurNAc peptidoglycan deacetylase from the same microorganism, has been biochemically characterized). *Bs*PdaA is not active on chitinous substrates [49]. To get further insights into the substrate specificity of the enzyme, a series of fragment-based docking calculations were carried out on *Bs*PdaC catalytic domain's structure as well as on the 3D structure of *Bs*PdaA and *Sp*PgdA, as representative peptidoglycan deacetylases for MurNAc and GlcNAc units, respectively. *Bs*PdaA is the only characterized peptidoglycan MurNAc deacetylase and *Sp*Pgda, although being a peptidoglycan GlcNAc deacetylase, is the closest CE4 enzyme to *Bs*PdaC. The docking experiments were conducted with monosaccharide (MurNAc and GlcNAc) and disaccharide (GlcNAc-MurNAc and MurNAc-GlcNAc) probes. The 3D structures of enzyme-ligand complexes of the three proteins were modelled by AUTODOCK 4.2 and AUTODOCK VINA (see Section 7.5.3). The obtained complexes

showed significant differences between GlcNAc and MurNAc specific enzymes and *Bs*PdaC, with dual activity depending on the substrate.

(i) *Bs*PdaC

The monosaccharide GlcNAc probe is able to fit into the active site of *Bs*PdaC with a proper orientation for catalysis (with the C=O of the acetamido group coordinated with the Zn^{2+} cation at a distance of 2.1±0.1 Å and the NH at hydrogen bonding distance of the general acid His427 (3.4±0.1 Å) (Figure 4.38, A). No productive binding was observed for the MurNAc probe, although the single monosaccharide was located in subsite 0, all sampled structures had non-productive orientations for catalysis (Figure 4.38, B). However, with the GlcNAc-MurNAc disaccharide probe, the MurNAc unit was placed in subsite 0 and the GlcNAc unit in subsite -1 in a productive manner, indicating a preference for this sugar moiety over the N-acetylglucosamine one in peptidoglycan-based substrates (Figure 4.3-38, C).

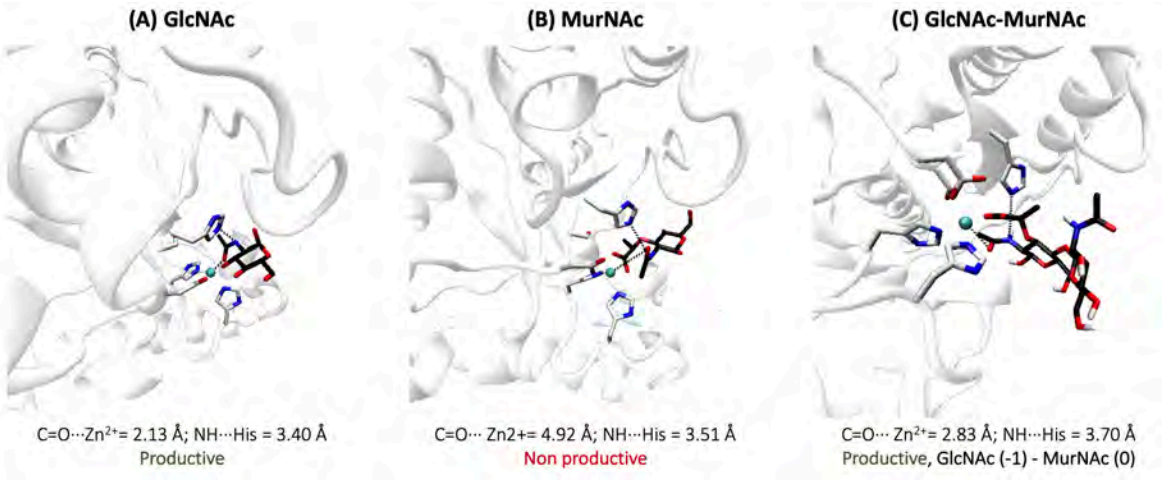


Figure 4.38. Ligand docking simulations on the X-ray structure of BsPdaC-CD*. (A) Monosaccharide GlcNAc. (B) Monosaccharide MurNAc. (C) Disaccharide GlcNAcβ-1,4-MurNAc.*

*Bs*PdaC active site is, therefore, able to accommodate both residues in a productive binding mode for deacetylation. Considering the peptidoglycan structure, which is composed of alternating GlcNAc and MurNAc units and the enzyme activity, it seems that subsites -1 and/or +1 are unable to accommodate a MurNAc residue, even in the absence of peptidyl substitutions, since PGN digested with L-alanine amidase is not deacetylated [121].

Focusing the attention again on the docked complexes obtained with $(GlcNAc)_4$, in which 4 subsites of the binding cleft are occupied, it could be observed that the 3-OH of the GlcNAc unit in subsite 0 does not establish interactions with the protein other than coordination with the metal ion, leaving room for the 3-lactoyl substitution (also shown with the PGN disaccharide just described). In the

other hand, the 3-OH of the GlcNAc unit in subsite -1 establishes hydrogen bond interactions with Tyr430 in both BM1 and BM2, leaving no room for 3-OH substitutions. Lastly, subsite +1 shows a weaker binding of the GlcNAc unit, where only the 6-OH establishes hydrogen bond interactions with protein residues (Asp398 in both binding modes and Trp402 in BM2). Considering all simulation results, they suggest that subsite -1 is specific for GlcNAc residues and dictates the specificity of the enzyme.

(ii) *Bs*PdaA

*Bs*PdaA was chosen as a model of a canonical peptidoglycan MurNAc deacetylase to perform the same fragment-based docking simulations. This enzyme is not active on GlcNAc residues and, in fact, the GlcNAc monosaccharide probe did not get into the active site, but mainly went to subsite +2 (Figure 4.39, A). On the contrary, with the other monosaccharide probe, MurNAc, an ensemble of docked structures compatible with productive binding in the active site (subsite 0) were obtained (Figure 4.39, B). Furthermore, with both disaccharide ligands, MurNAc was the only unit that was placed in subsite 0 in a productive manner, consistent with the MurNAc deacetylase specificity of the enzyme, inactive on COS substrates (Figure 4.39, C and D).

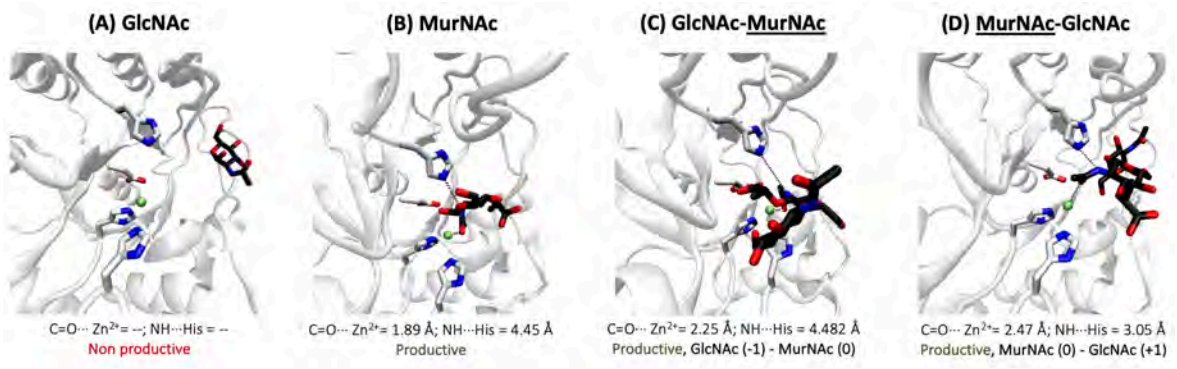


Figure 4.39. Ligand docking simulations on the X-ray structure of BsPdaA MurNAc deacetylase. *(A) Monosaccharide GlcNAc. (B) Monosaccharide MurNAc. (C) Disaccharide GlcNAcβ-1,4-MurNAc. (D) Disaccharide MurNAcβ-1,4-GlcNAc. Protein structure was directly taken from Protein Data Bank with PDB accession code 1W1B.*

(iii) *Sp*PgdA

To complete the comparison, *Sp*PgdA was used as a model of a canonical peptidoglycan GlcNAc deacetylase. The opposite behavior was observed in this case, only the GlcNAc monosaccharide probe was able to productively bind in the catalytic site (Figure 4.40, A), and the disaccharide ligands always placed the GlcNAc sugar in subsite 0, again in agreement with the deacetylation specificity on GlcNAc residues of peptidoglycan and COS substrates (Figure 4.40, C and D).

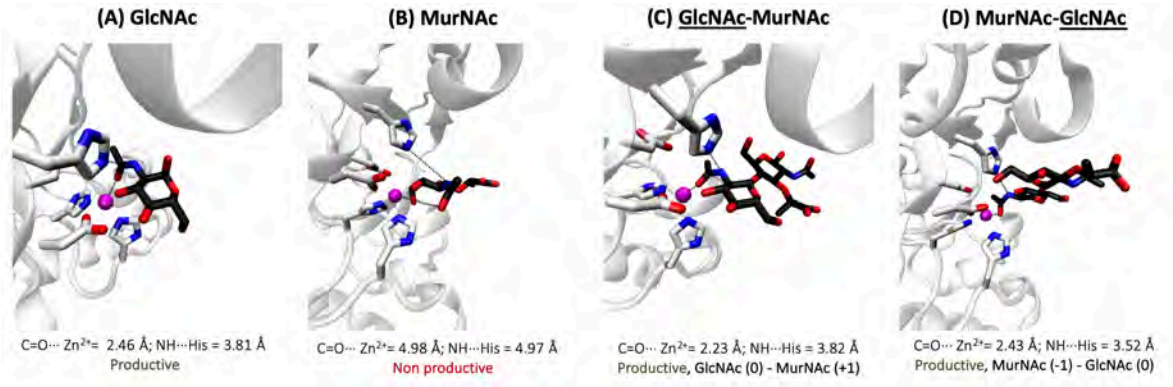


Figure 4.40. Ligand docking simulations on the X-ray structure of SpPgdA GlcNAc deacetylase. *(A) Monosaccharide GlcNAc. (B) Monosaccharide MurNAc. (C) Disaccharide GlcNAcβ-1,4-MurNAc. (D) Disaccharide MurNAcβ-1,4-GlcNAc. Protein structure was directly taken from Protein Data Bank with PDB accession code 2C1G.*

Besides the differences in terms of binding specificity showed by the docking experiments, comparison of 3D structures of the three enzymes unraveled significant differences at their active sites. First, the metal coordination differs between *Bs*PdaC and the other MurNAc deacetylase from *Bacillus subtilis* (PdaA). As already described, PdaC mantains the conserved metal-binding triad found in all characterized CDAs and peptidoglycan GlcNAc deacetylases (such as *Sp*PgdA). In contrast, in MurNAc deacetylases, the Asp residue is replaced by an Asn with a different side-chain orientation, pointing away from the metal ion (structural information currently limited to *Bs*PdaA and *Ba*CE4). Consequently, metal coordination in PdaA's catalytic center involves only two Histidine residues [34], [35]. A structural comparison of the metal coordination in the x-ray structures is represented in Figure 4.41.

Secondly, the surface electrostatic potential at the active site is different between the enzymes (Figure 4.42). *Bs*PdaA has a positively charged binding cleft, as opposed to the GlcNAc deacetylase *Sp*PgdA, in which it is negatively charged. *Bs*PdaC presents positive and negative patches in its binding cleft, laying in between of the two types of peptidoglycan deacetylases.

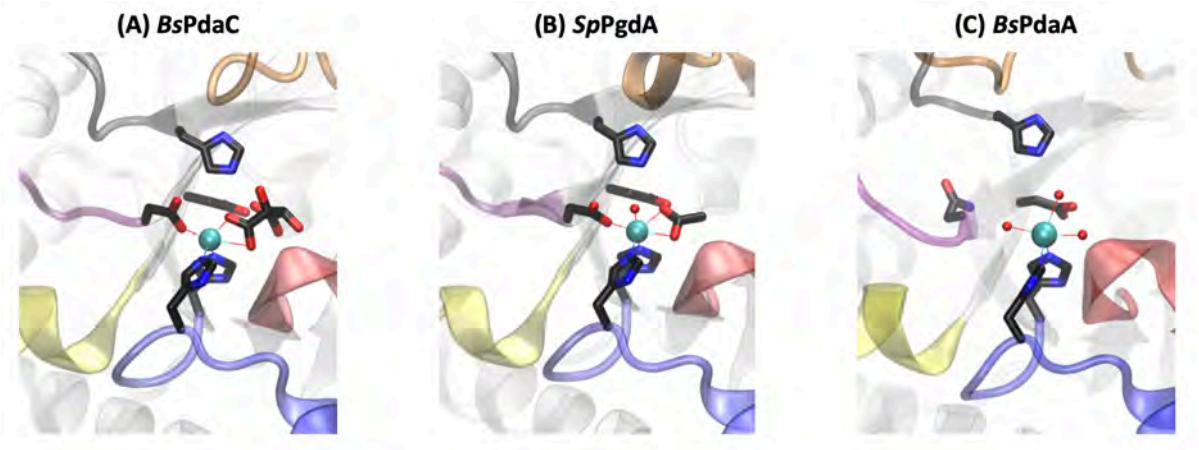


Figure 4.41. Metal coordination in the crystallographic 3D structures of peptidoglycan deacetylases*. (A) BsPdaC. (B) SpPgdA GlcNAc deacetylae. (C) BsPdaA MurNAc deacetylase. Protein structures were directly taken from the Protein Data Bank with PDB accession codes 1W1B (BsPdaA), 2C1G (SpPgd), and 6H8L (BsPdaC, this work).*

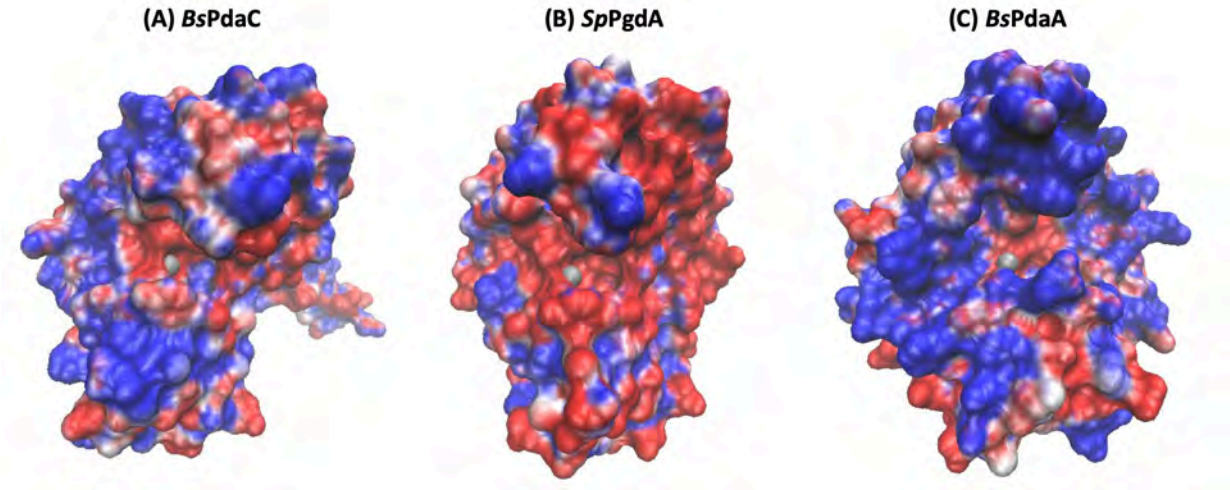
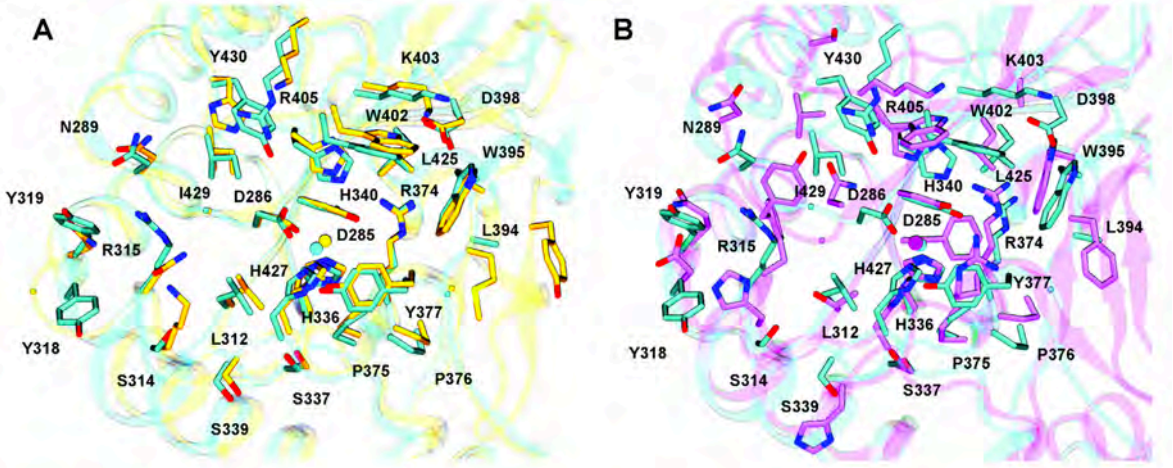


Figure 4.42. Surface electrostatic potential of peptidoglycan deacetylases*. (A) BsPdaC. (B) SpPgdA GlcNAc deacetylase. (C) BsPdaA MurNAc deacetylase. Protein structures were directly taken from the Protein Data Bank with PDB accession codes 1W1B (BsPdaA), 2C1G (SpPgd), and 6H8L (BsPdaC, this work).*

An in-depth comparison of the catalytic center residues of the three structures was conducted and superpositions of their substrate binding cavities are presented in Figure 4.43. The conservation of *Bs*PdaC with *Sp*PgdA is notably higher than with *Bs*PdaA, with very similar 3D structures (RMSD between catalytic domain CA atoms of 1.15 A) and identical structure and sequence composition of subsites 0 and +1. The main differences between PdaC and PgdA are (Figure 4.43, A) in subsite -1: *Bs*PdaC Arg405 and Tyr430 are replaced by Lys395 and His420 in *Sp*PgdA. In terms of specificity, both enzymes have the capacity to accommodate GlcNAc residues at subsite 0 (catalytic site for deacetylation), confirmed by ligand-docking simulation and deacetylation activity on COS. However, *Bs*PdaC does not deacetylate GlcNAc residues from peptidoglycan, behavior than can be explained by the mentioned differences in subsite -1: bulkier side chains of *Bs*PdaC hamper the

binding of the 3-lactoyl substitution of a MurNAc residue in subsite -1, in the same location where it is accommodated in *Sp*PgdA (Figure 4.44).

In contrast, major structural and sequence differences are observed between *Bs*PdaC and *Bs*PdaA at both sides of the active site (Figure 4.43, B). Remarkably, a negative charge is introduced in negative subsites (Tyr318 to Asp106) and a positive charge is removed (Arg315 to Phe103) next to it. *Bs*PdaC and *Bs*PdaA, although both deacetylate MurNAc residues of PGN, differ in the C3 substitution of their substrates, a negatively charged lactoyl group in the PGN substrate of *Bs*PdaA and a peptide-linked lactoyl group in the PGN substrate of *Bs*PdaC. In subsite 0, *Bs*PdaC has Tyr377 as part of the conserved MT3 motif of CE4 enzymes (RxPY). This Tyr is strictly conserved in chitin deacetylases, and highly conserved in PGN GlcNAc deacetylases (with few exceptions where it is replaced by Gly), but differs in *Bs*PdaA, which has an Arg residue instead (Figure 4.43, B). The side chain of Arg166 in *Bs*PdaA is oriented towards the active site, and the positive charge of the guanidinyl group may contribute to the stabilization of the free 3-lactoyl group of MurNAc of its PGN substrate (Figures 4.43, B and 4.44), as opposed to *Bs*PdaC in which the substrate is uncharged.



subsite	-3	-3	-2	-2	-2	-2	-1	-1	-1	-1	-1	0	0	0
*Bs*PdaC	Y318	Y319	N289	-	R315	S314	Y430	I429	L312	S339	R405	D286	H427	D285
*Sp*PgdA	G308	N309	N279	-	N305	K304	H420	I419	L302	S329	K395	D276	H417	D275
***Bs*PdaA**	**D106**	**Q107**	N78	**Y76**	**F103**	**H102**	**S225**	V224	**T100**	**H127**	**K195**	**N74**	H222	D73

subsite	0	0	0	0	0	0	+1	+1	+1	+1	+1	+1	**+2**	**+3**
*Bs*PdaC	H336	H340	S337	W402	Y377	L425	P376	K403	R374	P375	D398	W395	**L394**	-
*Sp*PgdA	H326	H330	S327	W392	Y367	L415	P366	K393	R364	P365	D388	W385	**M384**	**Y459**
***Bs*PdaA**	H124	H128	S125	W194	**R166**	L220	P165	-	R163	P164	**A190**	W187	**F186**	**F58**

Figure 4.43. Structural superposition and sequence alignment of solvent-exposed amino acids surrounding the active site, A) BsPdaC (cyan) and SpPgdA (yellow); B) BsPdaC (cyan) and BsPdaA (purple). *Catalytic center represented by the metal coordination site is in the middle. Amino acid labels correspond to BsPdaC. Subsites are defined by analogy to chitin deacetylases crystal structures in complex with ligands. Amino acid side chains of BsPdaC are labeled in the structures. Protein structures were directly taken from the Protein Data Bank with PDB accession codes 1W1B (BsPdaA), 2C1G (SpPgd), and 6H8L (BsPdaC, this work).*

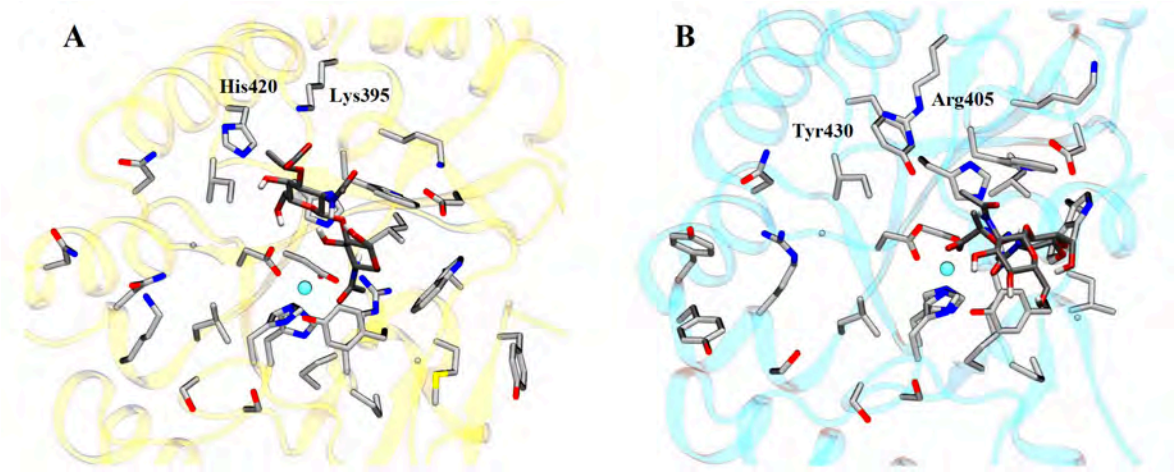


Figure 4.44. Ligand docking of A) MurNAcβ-1,4-GlcNAc on the X-ray structure of* Sp*PgdA GlcNAc deacetylase (in subsites -1, 0) and B) GlcNAcβ-1,4-MurNAc on the X-ray structure of* Bs*PdaC MurNAc deacetylase (in subsites 0, +1). *Solvent-exposed amino acids surrounding the active site are represented. Catalytic center represented by the metal coordination site is in the middle. Protein structures were directly taken from the Protein Data Bank with PDB accession codes 1W1B (BsPdaA), 2C1G (SpPgd), and 6H8L (BsPdaC, this work).*

4.3.4. Second generation of constructs for crystallogenesis

With the aim of obtaining the structure of enzyme·substrate complexes, a set of new constructs of the enzyme was designed. The dimerization packing observed in the solved 2D structure showed the presence of a C-terminal α-helix that protrudes from one monomer and interacts with the active site of the other monomer (the α-helix sequence is composed by the 20 last amino acids from *Bs*PdaC sequence, 5 residues from the subcloning restriction site sequence and 2 amino acids from the Strep-tag II peptide). Considering this structure, it was decided to change the Strep-tag II location to the N-terminus of the proteins.

The chosen expression vector for the subcloning, pPR-IBA2, has the transcriptional control of the strong bacteriophage T7 promoter (in the same way of the previously used pET22b(+) vector) and it presents the Strep-Tactin affinity tag fused with the N-terminus of the recombinant protein. Furthermore, the restriction site of TEV protease was added between this tag and the first codon of the proteins. TEV protease is a highly specific cysteine protease that recognizes the amino acid sequence ENLYFQ(G/S) and it cleaves between the Glc and Gly/Ser residues. Placing this sequence after the Strep-tag II peptide allows the removal of the tag after the purification process. With this concept, different constructs were prepared (Figure 4.45).

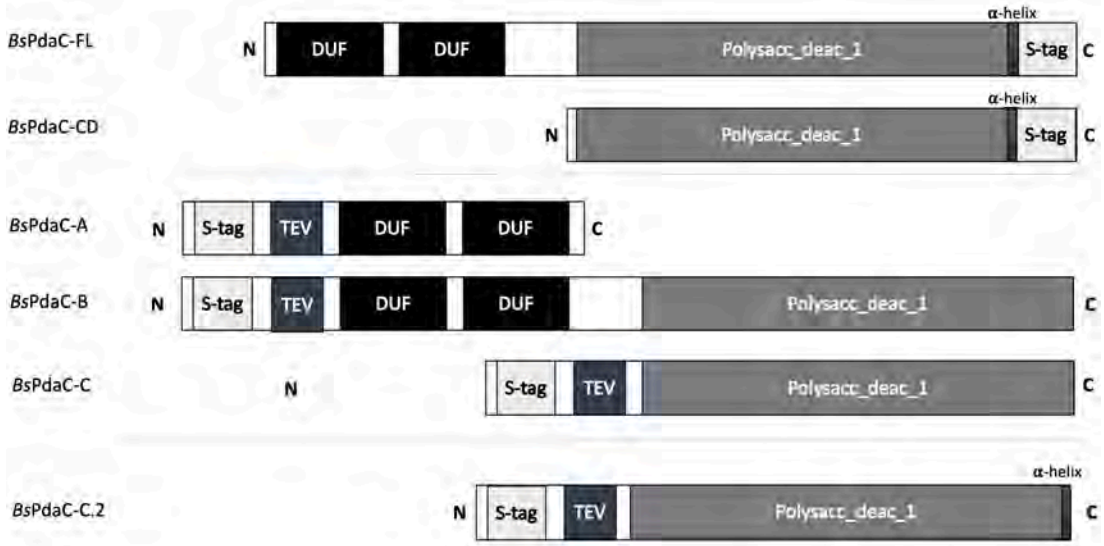


Figure 4.45. Domain organization of all BsPdaC constructs. *α-helix in the C-terminal end: sequence corresponding to the 20 BsPdaC residues that are part of the C-terminus helix that participates in the dimerization on the X-ray structure.*

4.3.4.1. *Bs*PdaC: constructs A, B and C

(i) Construct A

With the aim of gaining further information about their function, the first designed construct consists in the two N-terminus auxiliary domains of *Bs*PdaC (Figure 4.45). As already explained, *Sp*PgdA is the closest CE4 enzyme to *Bs*PdaC and it presents a similar domain architecture, with an initial N-terminal domain that it is linked to the catalytic C-terminal domain by an α-helix. To decide the exact amino acid sequence of each of the new constructs, the sequence alignment of the two enzymes, as well as the domains assignment in several databases and the secondary structure prediction of the complete sequence of *Bs*PdaC, were taken into consideration.

The designed protein sequence for construct A starts at Ser29 (after the signal peptide) and it ends at Lys234, which is, according to the secondary structure prediction, the last amino acid before the α-helix that is thought to act as a linker with the catalytic domain.

(ii) Construct B

With the purpose of trying to obtain the 3D structure of the full-length enzyme, construct B was designed. The main difference between this new construct and the initial full-length protein (besides the change of expression vector) would be the removal of the final 20 residues of the enzyme that were forming the C-terminus α-helix that was occluding the catalytic center on the x-ray structure (Figure 4.45). It was decided to start this construct with Ser29 like the previous one and to end it with Thr453, that was the last amino acid before the start of the α-helix.

(iii) Construct C

Lastly, a new construct with just the catalytic domain of the enzyme and without the α-helix residues was also proposed (Figure 4-45). Its sequence would be the one enclosed between Gln277 and Thr453.

The corresponding sequences were obtained by PCR reaction from the initial *Bs*PdaC full-length gene using primers that also included the KasI (forward) and PstI (reverse) restrictions sites for the later subcloning as well as the codifying sequence for the TEV restriction site (see Section 7.1.3). The sequences were subcloned into a pPR-IBA2 expression vector, which includes the Strep-tag II peptide upstream from the protein sequence (Figure 4.46). After confirming the sequences of the resulting plasmids by Sanger Sequencing, *E. coli* DH5α and BL21(DE3) STAR competent cells were transformed with the final constructs for protein expression and purification.

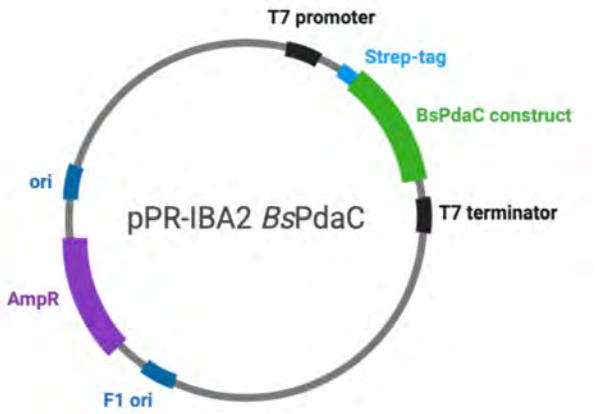


Figure 4.46. New constructs map. *Sequence map of the expression vector pPR-IBA2 with the subcloned sequence of the new constructs (BsPdaC construct on the scheme) with the Strep-tag II sequence and the TEV restriction site in the N-terminus of the protein.*

4.3.4.2. Expression and purification of new constructs

Several expression conditions based on the two methodologies previously described (growing in autoinduction medium and induction by IPTG) were tested with all constructs (Figure 4.47). However, only construct B was successfully produced in the soluble fraction. Expression of both constructs A and C were always observed on the insoluble fraction, indicating low stability and improper folding of the proteins.

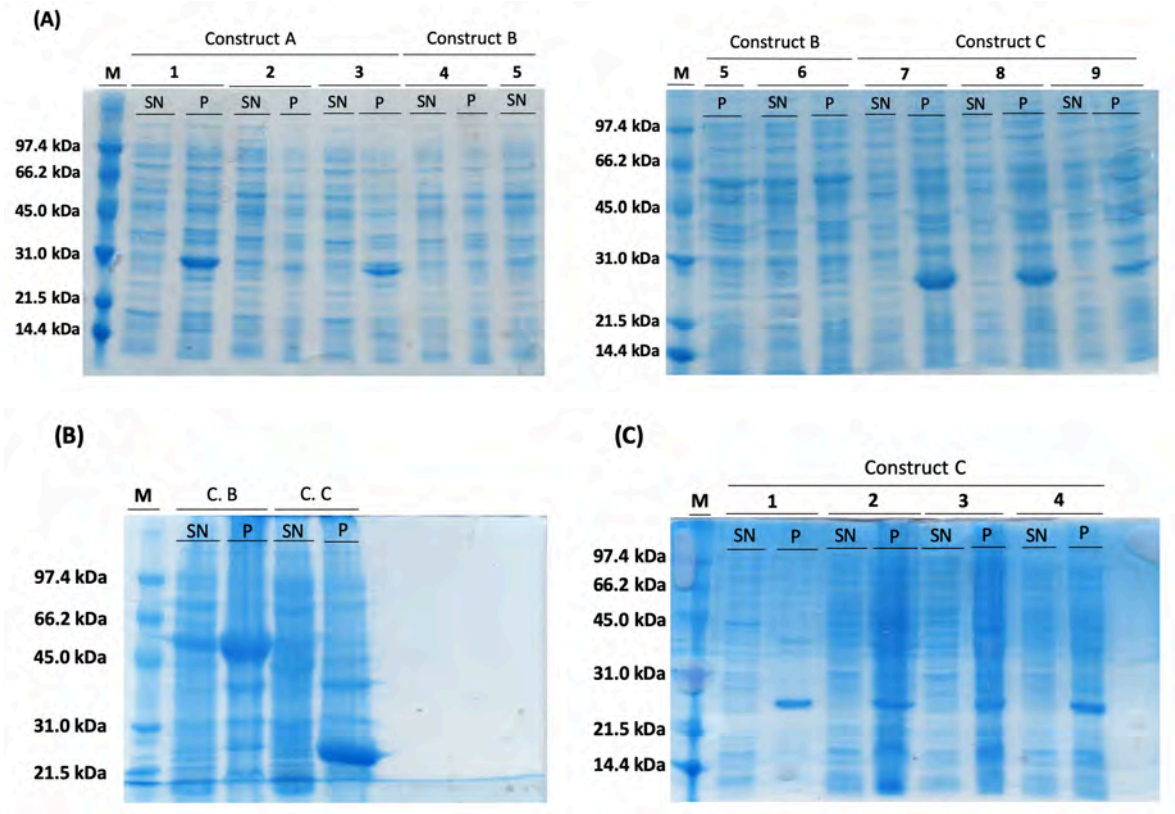


Figure 4.47. SDS-PAGE 14% acrylamide of expression trials of new constructs BsPdaC. (*A) Expression of constructs A, B and C by autoinduction protocol with different growth times. 1, 4 and 7: overnight; 2, 5 and 8: 24 hours; 3, 6 and 9: 58 hours. (B) Expression of constructs B and C by autoinduction protocol with 60h of growth time. C) Expression of construct C by IPTG induction after growing at 37°C until OD = 0.6. 1: induction with 1 mM IPTG for 1 hour at 37°C; 2: induction with 1 mM IPTG for18 hours at 37°C; 3: induction with 1 mM IPTG for 18 hours at 25°C; 4: induction with 0.02 mM IPTG (+ 2% EtOH) for 18 hours at 16°C. SN: supernatant after cellular disruption; P: pellet after cellular disruption.*

When proceeding with *Bs*PdaC construct B expression and purification following the same two chromatographic steps procedure previously explained, important differences were observed in the chromatographic profile obtained with SEC. As shown in Figure 4.48, most of the protein sample that was loaded into the gel filtration column eluted at an early retention time (F1, 40 minutes, corresponding to the void volume of the Superdex 200 (16/600) column), indicating a higher level of protein aggregation. The rise in absorbance for fractions 2 and 3, corresponding to the dimeric and monomeric forms of the protein, was almost unappreciable and much lower than expected. Activity of the three fractions was tested against the unspecific substrate AcOMU and deacetylase activity was not detected in any case.

The changes in expression of constructs B and C in comparison with the previous full-length and catalytic domain proteins, and also the lack of activity of the recombinant construct B, suggested that the removal of the final part of the protein was a change too drastic and that those final amino acids could be of importance for protein folding and activity.

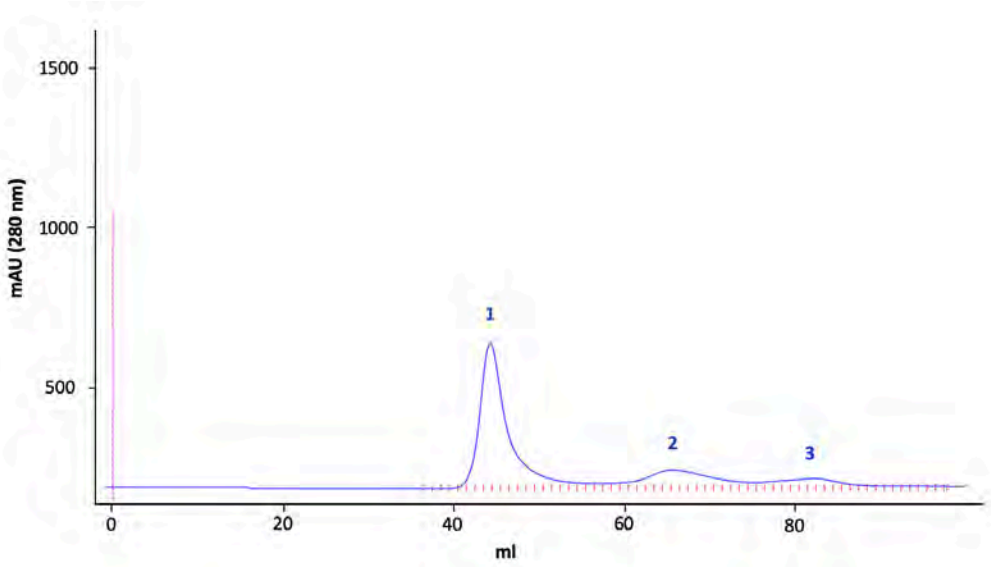


Figure 4.48. Size Exclusion Chromatography profile of BsPdaC-construct B. *Gel filtration performed with a Superdex 200 (16/600) column.*

Studying the structure of those residues on *Sp*PgdA (closest homolog to PdaC), it was observed that they adopt a beta-sheet folding that is covering a hydrophobic patch of the protein (Figure 4.49). Both enzymes share high identity in their catalytic domains and their structures have a very similar topology. If the same folding were taken place in PdaC, the change could have exposed that hydrophobic patch of the protein, which could induce hydrophobic interactions between different monomers and lead to the protein aggregation observed on the SEC profile and, thus, the activity loss.

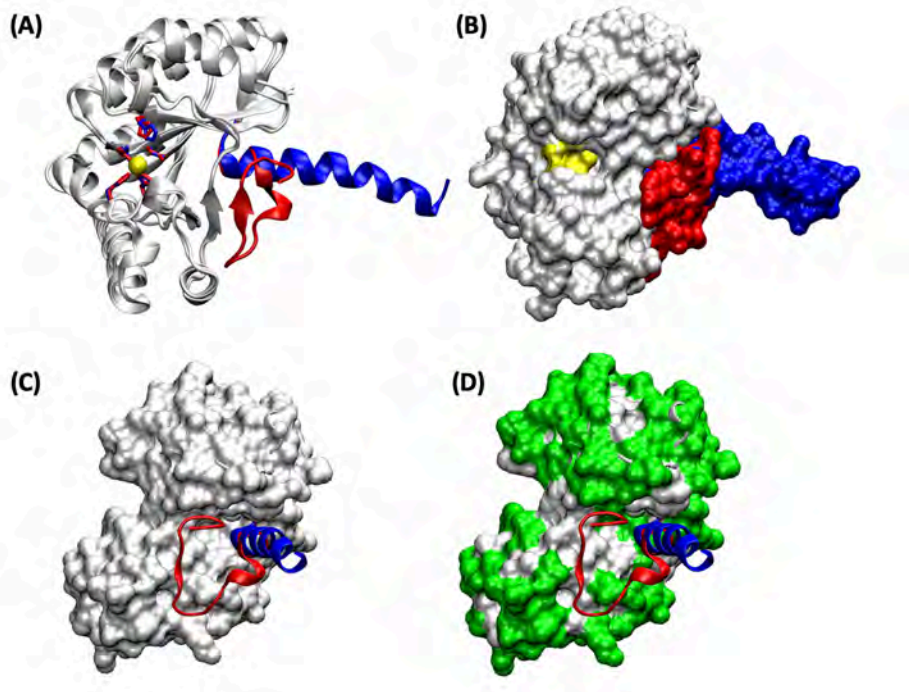


Figure 4.49. Superposition of BsPdaC-CD and SpPgdA 3D structures. *(A to D) Blue: PdaC-CD; Red: SpPgdA; Yellow: Catalytic metal. (D) Green: polar amino acids; white: non-polar/hydrophobic amino acids.*

4.3.4.3. *Bs*PdaC: construct C-2

As a final strategy, construct C-2, consisting on the catalytic domain of the enzyme with the whole protein sequence in the pPR-IBA2 vector, was designed and generated (Figure 4.45). In order to avoid the activity and stability loss observed with the previous attempts, construct C-2 design maintains the 20 last amino acid of *Bs*PdaC sequence (which were the initial part of the C-terminus helix in the x-ray structure). It was chosen to proceed only with the catalytic domain because its initial version crystallized.

Although with low expression yields (between 0.3 and 0.7 mg/L of culture), it was possible to obtain pure soluble recombinant protein following the same expression and purification procedures described for the initial enzymes. The enzyme *Bs*PdaC C-2 showed to be active and displayed similar enzymatic activity to *Bs*PdaC-CD on chitooligosaccharides (Table 4.10).

Table 4.10. Specific activity of construct C-2 in comparison to the original BsPdaC catalytic domain.

	SPECIFIC ACTIVITY (V_0/[E], MIN^{-1})	
SUBSTRATE	***Bs*PdaC-CD**	***Bs*PdaC-C-2**
$(GlcNAc)_2$	no activity	no activity
$(GlcNAc)_3$	3.99 ± 0.27	4.06 ± 0.21
$(GlcNAc)_4$	18.18 ± 0.22	18.45 ± 0.1
$(GlcNAc)_5$	36.80 ± 3.46	37.44 ± 4.24

The inactive mutant (D285S) was generated, produced and purified with similar expression yields to the wild type C-2 construct. After confirming the loss of activity of the mutant, several attempts to co-crystalize with $(GlcNAc)_4$ and $(GlcNAc)_5$ substrates failed again. Although some little crystals were obtained (Figure 4.50) and several optimization conditions were tested (see Section 7.6.3), it was not possible to obtain crystals with the needed characteristics to diffract.

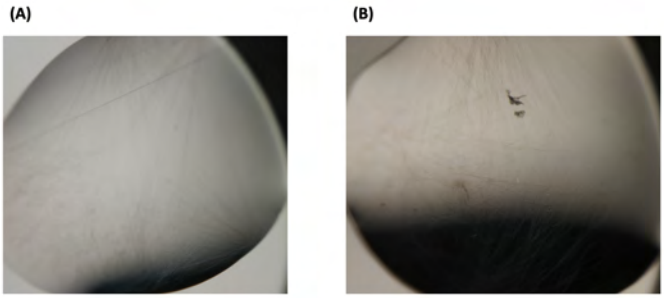


Figure 4.50. Crystals from co-crystallization assays with BsPdaC-C.2-D285S and $(GlcNAc)_4$. *Small crystals were obtained in two different conditions of the crystallogenesis mixtures (see Section 7.6.3).*

Having failed to obtain crystal complexes of *Bs*PdaC with substrates, ligand binding was assessed by circular dichroism (CD) spectroscopy. Both thermal unfolding monitoring at 222 nm and determination of CD spectra between 250 - 300 nm were performed in absence and presence of COS substrates (see Section 7.7.3). The melting temperature (T_m) of the free enzyme and upon addition of COS substrates (DP2, DP4 and DP5) was determined (Figure 4.51 and Table 4.11). An increase between 3.3 and 3.7°C upon binding of ligands was observed, suggesting that binding of these ligands takes place and it results in an increase of the enzyme's thermal stability.

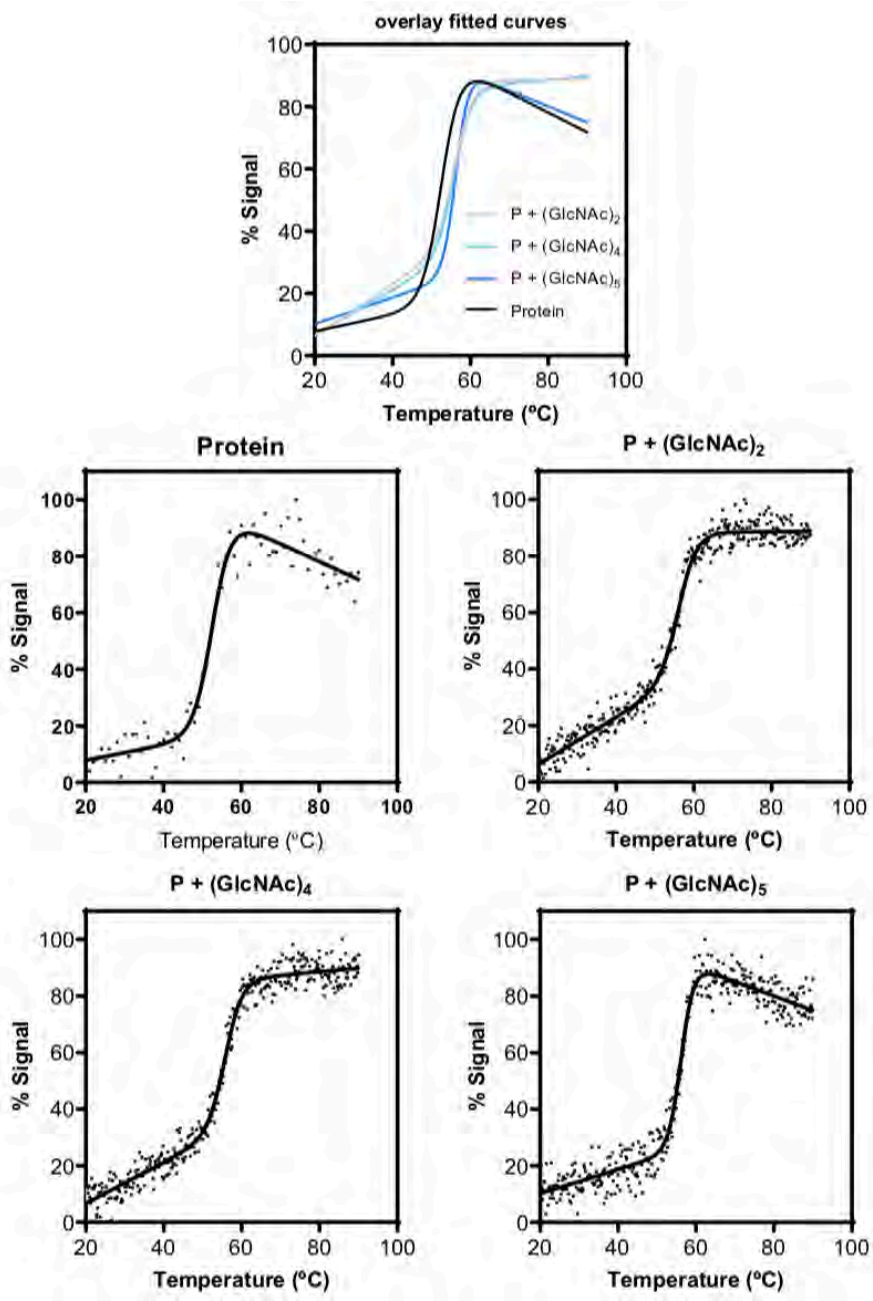


Figure 4.51. Thermal unfolding of BsPdaC-C.2-D285S and effect of ligand binding on its thermal stability. *CD signal data were fitted to a 2-state unfolding model [126]: %Signal = ((αF+βF·T) + (αU+βU·T)·exp (m·(T-Tm)) / (1 + exp (m·(T-Tm))) where αF and βF are the intercept and slope of the base line for the folded protein and αU and βU for the unfolded protein, m the cooperativity of unfolding and Tm the melting temperature.*

Table 4.11. Melting temperature (T_m) values by circular dichroism.

	T_m (°C)
Protein (*Bs*PdaC-C2-D285S)	52.6 ± 0.4
Protein + $(GlcNAc)_2$	56.3 ± 0.2
Protein + $(GlcNAc)_4$	55.9 ± 0.2
Protein + $(GlcNAc)_5$	56.3 ± 0.2

CD spectra can be used empirically as "fingerprints" of the native structure of a particular protein. Aromatic residues (tryptophan, tyrosine and phenylalanine) are the main chromophores in the near-UV region (240 to 320 nm). Thus, changes in the CD spectra of that region are due to changes of the aromatic residues and their environments and can be useful for comparison of tertiary structures between related proteins or for monitoring conformational changes in a protein under different conditions. Addition of specific ligands such as substrates, cofactors or regulatory molecules can lead to structure changes that are essential for their function, and these small conformational changes are more likely to be detected in the near-UV region. *Bs*PdaC contains four aromatic residues in the binding cleft that are at an interaction distance from the substrate (Figure 4.52).

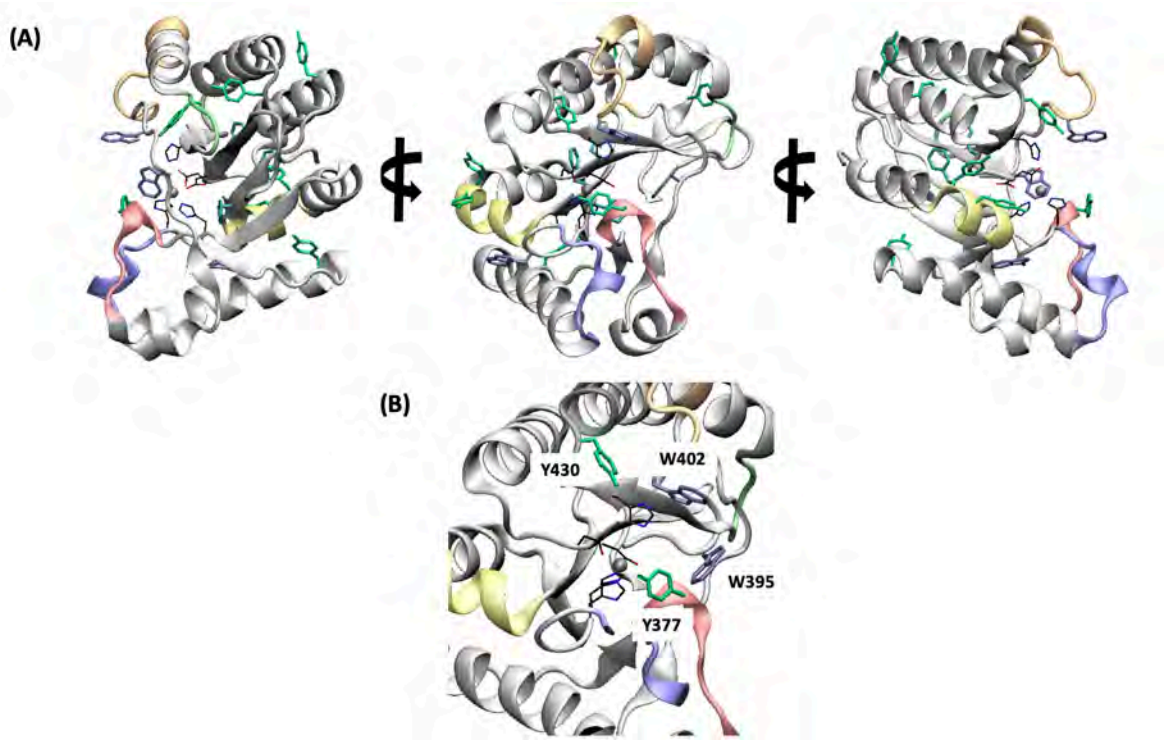


Figure 4.52. Aromatic residues*. (A) From the whole BsPdaC-CD structure. (B) From the catalytic center, within an interacting distance with substrate. Green: tyrosine residues; Cyan: phenylalanine residues; Ice blue: tryptophan residues.*

Small differences can be observed in the CD spectra, specifically in a peak around 255 nm, region corresponding to Phenylalanine residues, and in the region between 285 and 290 nm, corresponding to Tyrosine and Tryptophan residues (Figure 4.53). Although it is not possible to elucidate the change of which specific residues is responsible for the differences in the spectra when ligands are added, the ones of the higher region could be due to the four aromatic residues of the catalytic center.

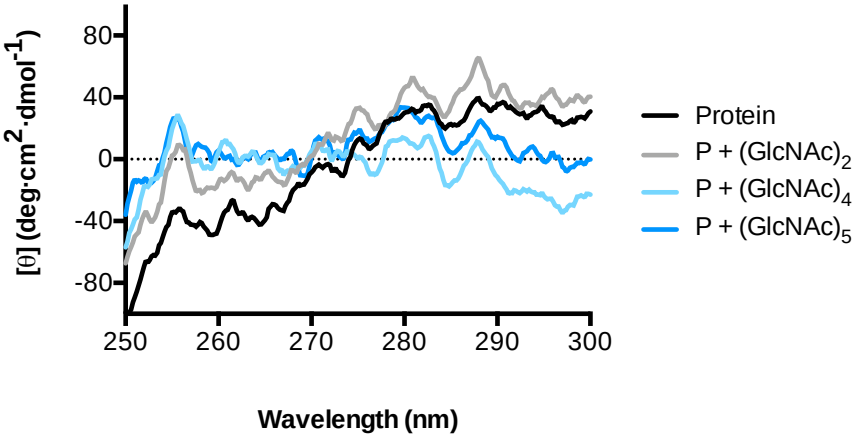


Figure 4.53. Near-UV CD spectra of BsPdaC-C.2-D285S in absence and presence of COS ligands.

4.4. Mutational study on *Bs*PdaC

Due to the short length of the loops that surround the active site, the *Bs*PdaC enzyme shows an open binding cavity and the number of residues that are close enough to interact with the substrate upon binding is limited. Given that it was not possible to obtain a 3D structure in complex with substrates, these residues were previously identified in section 4.3. by docking simulation with $(GlcNAc)_4$ substrate. In this section, firstly, a mutational study of a set of these interacting amino acids is presented in order to experimentally asses their effect on the enzyme's activity. Secondly, it is shown how the mutation of a key interacting residue can modify the specificity of the enzyme.

4.4.1. Probing the active site

Some of the residues that conform the interaction network with the substrate are conserved within the CE4 family, while others seem to be specific of *Bs*PdaC. The selection of seven of these residues (highlighted in Figure 4.54 and represented in Figure 4.55) and the designed mutations for each one of them are described below.

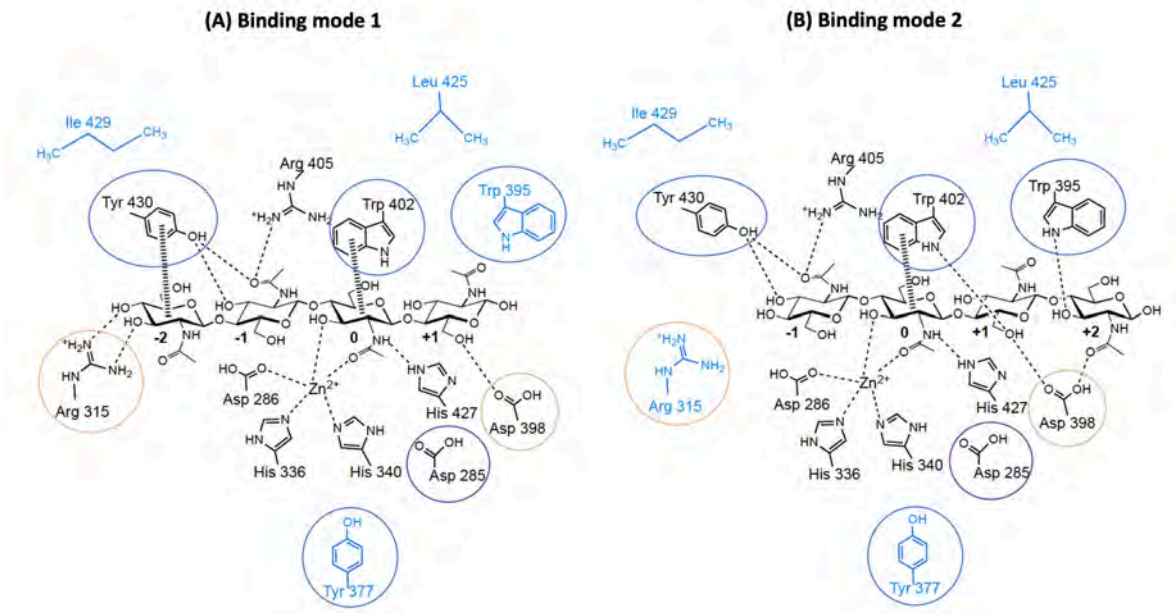


Figure 4.54. Residues from binding cleft that are conserved within CE4 family or that interact with substrate in BsPdaC. *(A) For binding mode 1 (AAAA). (B) For binding mode 2 (AAAA), where A is the GlcNAc residue in subsite 0 (catalytic site). Residues in black: protein-ligand interactions in BsPdaC from docking simulations with $(GlcNAc)_4$. Residues in blue: residues that do not establish interactions with the substrate. Chosen residues for mutational study are circled.*

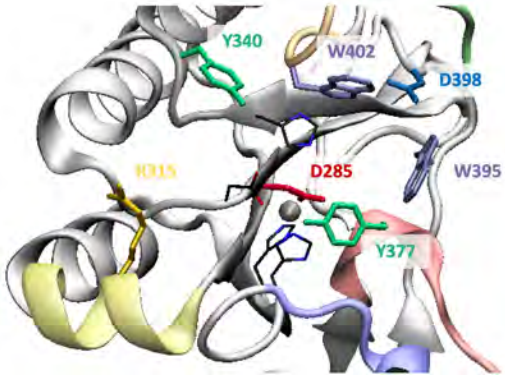


Figure 4.55. BsPdaC 3D structure*. Selected residues for the mutational study of key active site amino acids.*

- **D285**: this aspartate from Motif 1 was previously chosen as the residue to mutate in order to generate the inactive enzyme variant due to its possible role as the general base.

Many carbohydrate-binding proteins and carbohydrate active enzymes contain aromatic amino acid residues in their binding and catalytic sites. These residues interact with carbohydrates with a stacking geometry via CH-π interactions [127]. Four aromatic residues from the active site were selected and mutated to alanine in order to evaluate the importance of the stacking interactions that they may establish with the sugar moieties of the substrate.

- **W402**: this tryptophan is located in loop 4 and establishes a CH-π interaction with the monomer at subsite 0. It is highly conserved within the family, with very few exceptions.

- **Y430**: located in loop 6, this tyrosine establishes both CH-π (with the monomer in subsite -2) and hydrogen bond interactions (with the monomer in subsite -1) in binding mode 1. In binding mode 2, the stacking interaction is lost but the residue maintains its ability to establish two hydrogen bonds with the -1 GlcNAc unit. This is not a conserved residue within the family, being PdaC one of the few members with an aromatic residue in this position (along with *Pg*CDA, *Sch*CDA and *Cn*CDA). Hydrophobic residues like valine, leucine and isoleucine are the most represented in this position (for example, almost all CODs show a leucine), followed by histidine.
- **W395**: all the peptidoglycan deacetylases included in the phylogenetic study of the family share this tryptophan in the same position as well as most CDAs. The residue is located between loops 3 and 4 and, in the case of *Bs*PdaC, interacts with the reducing end moiety in binding mode 2 (subsite -2).
- **Y377**: although this aromatic residue is not strictly located within an interacting distance with the substrate in the docked structures of minimal energy, it was also selected because of its location in loop 3 and the fact that it is highly conserved in the chitin deacetylases subgroup and shared by some other members of the CE4 family.

  In *Vc*CDA, the equivalent residue Tyr169 interacts with the substrate and participates in the stabilization of the tetrahedral oxyanion intermediate during catalysis. Although this interaction is lost in *Bs*PdaC structure, Y377 could interact with the metal coordination residue His340.
- **D398**: this aspartate interacts with the reducing end of the substrate in both binding modes and it is shared by all CDAs, some peptidoglycan deacetylases and the acetylxylan esterase *Sl*AxeA. Different mutations were designed in order to evaluate the effect of those changes:
    - Glutamate: maintaining the negative charge but increasing the size of the lateral chain.
    - Asparagine: with a neutral charge, this residue is able to establish hydrogen bond interaction with the substrate.
    - Leucine: by changing to this hydrophobic residue, possible interactions are eliminated but the shape and size of the side chain are maintained.
    - Alanine: this change to a small, non-polar amino acid eliminates the side chain maintaining the backbone's orientation.
    - Lysine: changing the charge to a positive one and maintaining the shape and size of the side chain.

- **R315**: this arginine residue located in subsite -2 interacts with the non-reducing end of the substrate via two hydrogen bonds in binding mode 1. In binding mode 2, in which the substrate is displaced to positive subsites, the interactions are lost due to the highest distance between the arginine and the oligosaccharide. This is not a conserved residue within the family and the differential role between the two main binding modes made it interesting for the study.

The generation of the proposed mutants was performed by a Quick-Change PCR protocol (see Section 7.1.4) and it was grouped in two sets of experiments. A first set of mutants included residues from positive subsites, to evaluate possible changes in activity when destabilizing interactions between the enzyme and the sugar moiety that binds to +1 or +2 subsites. *Bs*PdaC-FL was used as template to generate the designed mutants at residues W395, D398 and W402. A second series of residues were chosen in order to evaluate the role of aromatic residues in enzyme-substrate binding. In this second case, *Bs*PdaC-CD was used as template for mutants at residues Y370, Y377, R315 and W402. Moreover, the mutant at residue D285 was also generated from both templates. Sequences were confirmed by Sanger sequencing before continuing with their expression, purification and characterization.

4.4.1.1. Expression and purification of mutants

Both the expression and purification protocols were the same than those previously used for the wild type enzymes. Expression of the recombinant proteins in *E. coli* BL21(DE3) star used an autoinduction medium and purification followed two chromatographic steps (strep trap affinity chromatography and size exclusion chromatography with a Superdex 200 column). In all cases, the characteristic profile with three oligomeric fractions from the size exclusion chromatography was observed. The monomeric form of the enzymes was the one used for the subsequent kinetic study in order to compare their deacetylase activity with that of the parental wild type enzyme. In terms of protein expression, most mutants were obtained with similar yields than the wild type enzymes (between 15 and 20 mg/L of culture). The only exception was the aspartate mutant D398N, which showed much lower yields (0.4 mg/L of culture).

4.4.1.2. Specific activity of mutants

In order to contribute with experimental support to the simulation results, from which the interactions network between enzyme and substrate was draught, the specific activity of the designed mutants was determined and compared with the wild type one.

- Mutants from *Bs*PdaC-FL construct

Specific activity toward 2 mM $(GlcNAc)_4$ substrate at 37°C was determined by monitoring the release of acetic acid with a commercial acetate determination kit based on a coupled enzymatic assay (See Section 7.4.4).

- Mutants from *Bs*PdaC-CD construct

In the case of the mutants generated from the catalytic domain, specific activity on 2 mM chitotetraose at 37°C was determined by the HPLC-MS method previously described (see Section 4.2 and Section 7.4.2), monitoring the production of the mono-deacetylated product A3D1 along reaction time.

Results are presented in Figure 4.56 and Tables 4.12 and 4.13, in which the specific activity of each mutant is represented in percentage relative to the parental wild type enzyme, either *Bs*PdaC-CD or *Bs*PdaC-FL.

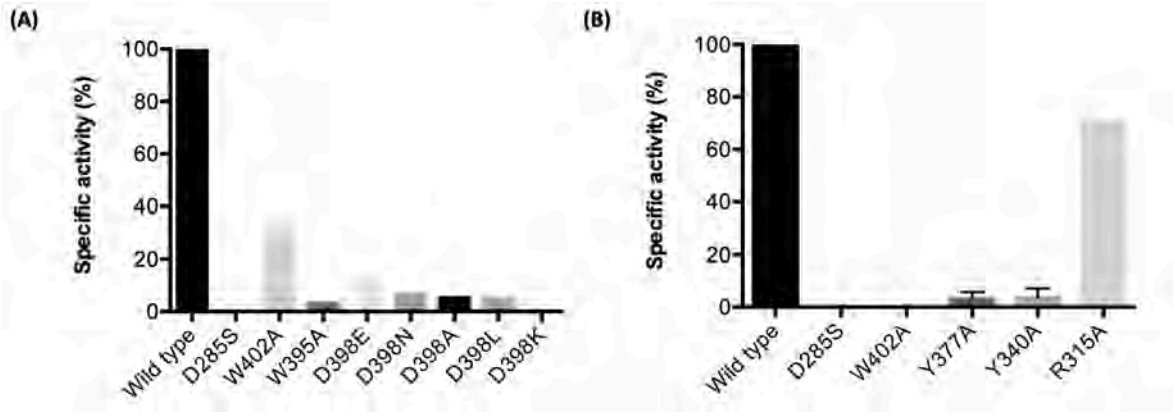


Figure 4.56. Specific activity of active site mutants of BsPdaC. *The specific activity is presented in percentage relative to the wild type enzyme. (A) Mutants from BsPdaC-FL, (B) Mutants from BsPdaC-CD.*

Table 4.12. Specific activity of active site mutants of BsPdaC-FL. *The specific activity is presented in percentage relative to the wild type enzyme.*

From *Bs*PdaC-FL		
ENZYME/MUTANT	**ACTIVITY (%)**	
***Bs*PdaC-FL wild type**	**100**	
*Bs*PdaC D285S	0 (inactive)	
*Bs*PdaC W402A	31	
*Bs*PdaC W395A	3.1	
*Bs*PdaC D398X	D398E	12.5
	D398N	6.6
	D398A	5.3
	D398L	5.1
	D398K	0.24

Table 4.13. Specific activity of BsPdaC-CD active site mutants. *The specific activity is presented in percentage relative to the wild type enzyme.*

From *Bs*PdaC-CD	
ENZYME/MUTANT	**ACTIVITY (%)**
***Bs*PdaC wild type**	**100**
*Bs*PdaC D285S	0 (inactive)
*Bs*PdaC W402A	0.02 – 0.09 (inactive)
*Bs*PdaC Y377A	1 - 5
*Bs*PdaC Y430A	1 - 6
*Bs*PdaC R315A	71

- **D285S**

The position of this residue in both the initial multiple sequence alignment and the 3D structure when $(GlcNAc)_4$ was added in the docking experiments, suggested that it is the one acting as the general base in the catalytic mechanism of the enzyme. The change to serine, a small hydrophilic amino acid, caused a complete loss of activity, confirming the key role of the aspartate in the catalytic mechanism, as already demonstrated in Section 4.3. The inactivation of a CDA by the mutation of the general base to a serine was already reported for *Vc*CDA [24].

- **Aromatic residues**

The set of mutants at the aromatic residues of the active site showed an important loss of activity, retaining only between 1 and 30% of the wild type activity. All mutations were to alanine, a much

smaller hydrophobic amino acid unable to form interactions with the substrate. Thus, the enzymes inactivation was caused by the disruption of the stabilizing interactions (both stacking and hydrogen bonds) established by the tryptophan and tyrosine residues and indicates the important role of these amino acids.

As opposed to the equivalent residue in other CE4 enzymes (like Y168 on *Vc*CDA), Tyr377 does not interact with the substrate during catalysis. Due to the short *Bs*PdaC's loop 3, this residue is not located within an interacting distance with the substrate unit bound to subsite 0. However, the 3D structure shows a possible interaction between Tyr377 and the metal coordination residue His340. Mutation to Ala reduces the activity to 1-5% of the wild type enzyme, probably due to partial disruption of the metal coordination geometry.

Trp402 was the only residue that was mutated to alanine in both constructs, full-length and catalytic domain, with a larger loss of activity in the last. Although both wild type enzymes show similar stability and activity, the alanine mutant from the full-length enzyme retains 30% of the wild type enzyme activity while the alanine mutant from the truncated form showed to be almost inactive, retaining less than 1% of its parental wild type enzyme. Although these results should be verified, they suggest that the lack of the auxiliary N-terminus domains is a possible cause of the lower residual activity of the catalytic domain mutant.

- **D398X**

Although all changes caused a loss of specific activity, the decrease in the values is different depending on the mutation.
Both D398A and D398L mutants showed the same specific activity, suggesting that is not the size of the amino acid on that position that is influencing substrate binding, but the established interactions. The asparagine mutant, with the ability to interact with the substrate via hydrogen bonds, presented a slightly higher specific activity. The mutant that showed the lower residual activity is the lysine one, which implied the loss of one hydrogen bond and the most drastic change, to a positive side chain. The enzyme is almost inactivated with this mutation, suggesting that the negative charge of the original aspartate residue plays an important role in enzymatic activity. Finally, the glutamate mutant (which maintained the original negative charge in another position, due to the larger side chain) was the one that showed the minor change in activity. It is possible that the presence of the negative charge counteracts the change in size, reducing the effect of the

mutation and reinforcing the importance of charge and polarity in the interaction between this residue and the substrate.

- **R315A**

Due to its location in subsite -2, this arginine residue only participates in substrate binding in one of the two binding modes of the substrate (BM 1), being too far from the substrate in the other (BM 2). The alanine mutant conserves a 70% of the wild type activity, indicating that it is not essential for the deacetylation reaction. Nevertheless, the residue could have a role related to the enzyme's specificity, influencing the pattern of deacetylation of PdaC, possibility that will be further explored in the following Section.

From all the studied residues, Y430, W402, W395 and D398 establish direct interactions with residues of the substrate in the two predicted binding modes with $(GlcNAc)_4$. Deletion of any of those interactions, by mutating to Alanine, causes a 2 or 3 orders of magnitude decrease of specific activity, indicating that are important for substrate binding.

4.4.2. Engineering the pattern of deacetylation

In previous sections of this thesis, the specificity and the structure of the enzyme *Bs*PdaC have been described. According to the "Subsite Capping Model", the binding clefts of CE4 enzymes are shaped by a series of structural elements (loops) that are involved in their substrate preference and specificity, although this effect is most important in chitin deacetylases. In the case of PdaC, that reveals short loops and an open binding cleft, this model is probably not relevant for its specificity. The enzyme is able to accommodate substrates in different binding modes and it follows a multiple-chain mechanism on COS substrates with DP≥3. It starts at the internal GlcNAc units and leads to the deacetylation of all residues but the reducing end one. With $(GlcNAc)_4$, the first deacetylation takes place with the same efficiency at the second or third GlcNAc unit.

Considering these characteristics, in order to engineer the pattern of deacetylation of the enzyme, the approach should focus on identifying key residues or features that could play a role on specificity, more than on size or shape of the loops. Docking experiments with $(GlcNAc)_4$ allowed the study of the substrate-enzyme interactions, which led to a rational design mutational strategy to engineer PdaC's specificity and helped to better understand the catalytic mechanism of the enzyme.

4.4.2.1. *Bs*PdaC-CD R315 mutants

Determination of *Bs*PdaC pattern of deacetylation showed that the first deacetylation event on $(GlcNAc)_4$ takes place at either one of the two central residues of the substrate, rendering a 50% population of each product, ADAA and AADA. Computational docking experiments with the solved crystal structure and $(GlcNAc)_4$ were in agreement with the experimental results. Two binding modes (BM1, that would render the product AADA and BM2, that would render the product ADAA) showed the lowest energies. Both were of similar energy (-7.6 and -7.9 kcal/mol respectively), indicating the same probability for each one.

To evaluate the possibility of modifying the pattern of deacetylation exhibited by the enzyme, differential interactions between the two binding modes were explored. Although most of the residues interacting with the substrate are the same in both cases, some of these residues establish interactions with different GlcNAc units of the substrate, as in the following cases:

- Aspartate 398, that interacts only with the sugar located at +1 in BM1 but establishes hydrogen bond interactions with both the GlcNAc units at +1 and +2 in BM2.
- Tyrosine 430, that establishes a hydrogen bond interaction with the GlcNAc unit at subsite -1 in both binding modes but shows an additional staking interaction with the sugar at -2 in BM1.
- Tryptophan 402 (MT 4), that establishes a CH-Π interaction with the substrate residue located at subsite 0 in both binding modes. In addition, it forms a hydrogen bond with the GlcNAc unit at subsite +1 in BM2 that seems to be lost in the BM1 complex.

It should be stressed that these differences are subtle and that the loss of interactions between one binding mode and the other may be attributed to the computational docking protocol. Some of these interactions may be recovered with a proper equilibration of the complex structures.

The only clear differential interacting residues between the two complexes are the following:

- Tryptophan 395, that establishes a hydrogen bond interaction with the GlcNAc unit located at the +2 subsite in BM2.
- Arginine 315, that interacts with the substrate by the formation of two hydrogen bonds between the nitrogens of the guanidinium group and the sugar unit at -2 subsite in BM1.

In these cases, the two residues are clearly too distant from the substrate in the binding mode in which the interaction is lost. W395 is located in subsite +2 which is occupied by the reducing end

unit only in BM2. Similarly, but at the other side of the catalytic groove, R315 is located in subsite -2, where the non-reducing end of the sugar is placed in BM1 (Figure 4.57).

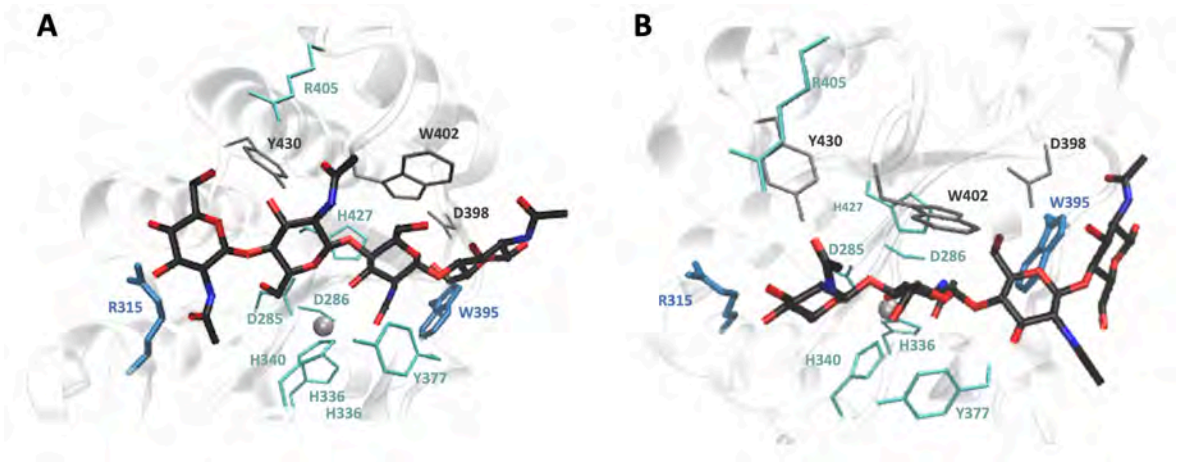


***Figure 4.57**. **Docking simulation of** Bs**PdaC-CD·(GlcNAc)$_4$ complexes.** Catalytic site residues that interact with (GlcNAc)$_4$ in (A) BM1 and (B) BM2. Green) residues that interact in the same way with the substrate in both binding modes (D285, D286, H405, H427, H336 and Y377); Grey) residues that interact differently with the substrate in both binding modes (D398, W402, Y430); Blue) R315, residue that only interacts with the substrate in BM1, and W395, residue that only interacts with the substrate in BM2.*

The change of arginine 315 residue to different amino acids was designed to break the interactions with the non-reducing end and, consequently, destabilize BM1. Four different mutations were designed in order to evaluate the role of this residue and to modify the specificity relative to the wild type enzyme, in which the second and third residue of the substrate are initially deacetylated indistinctly. In all cases, the chosen amino acids are smaller than arginine and occupy less space, possibly weakening the interaction with the substrate being at a higher distance.

- To Lysine: this would be the least drastic change, in which the positive charge of the amino acid is maintained and the formation of at least one hydrogen bond with the substrate is still possible as lysine presents a -NH_3 group in its side chain.
- To Serine: the positive charge is lost as serine presents a polar uncharged side chain. Moreover, just one hydrogen bond can be established between the hydroxyl group of the amino acid and the substrate.
- To Isoleucine: the mutation to this amino acid leads to the loss of both the positive charge and the possibility to form hydrogen bond interactions with the GlcNAc units of the substrate, but the new residue still maintains a bulky side chain.
- To Alanine: this non-charged amino acid is very small and, in the same way as isoleucine, its hydrophobic side chain cannot establish hydrogen bond interactions. This is the most significant change from the original residue, both in size and possible interactions.

The generation of the proposed mutants was performed by a Quick-Change PCR protocol (see Section 7.1.4) using the *Bs*PdaC-CD construct as template and their sequences were confirmed by Sanger sequencing before continuing with their expression, purification and characterization.

4.4.2.1.1. Expression and purification of R315 mutants

The expression protocol of the mutants in *E. coli* BL21(DE3) star was the same as the one used for the wild type enzyme, consisting of expression with LB supplemented medium for the autoinduction production of the protein. Regarding the recombinant proteins' purification, either a single chromatographic step with a Strep-trap column or both chromatographic steps (a first affinity chromatography and a second size exclusion chromatography with a Superdex 200 column) were used depending on the subsequent assays. When purified for the purpose of performing the coupled assay (see 4.4.2.2 section bellow) and specific activity determination (see 4.4.4 section bellow), all enzymes were purified by one chromatographic step. As example, Figure 4.58 shows the chromatographic profile for the R315A mutant and Figure 4.59 shows the SDS-PAGE analysis of the isolated proteins from the Strep-trap affinity chromatography.

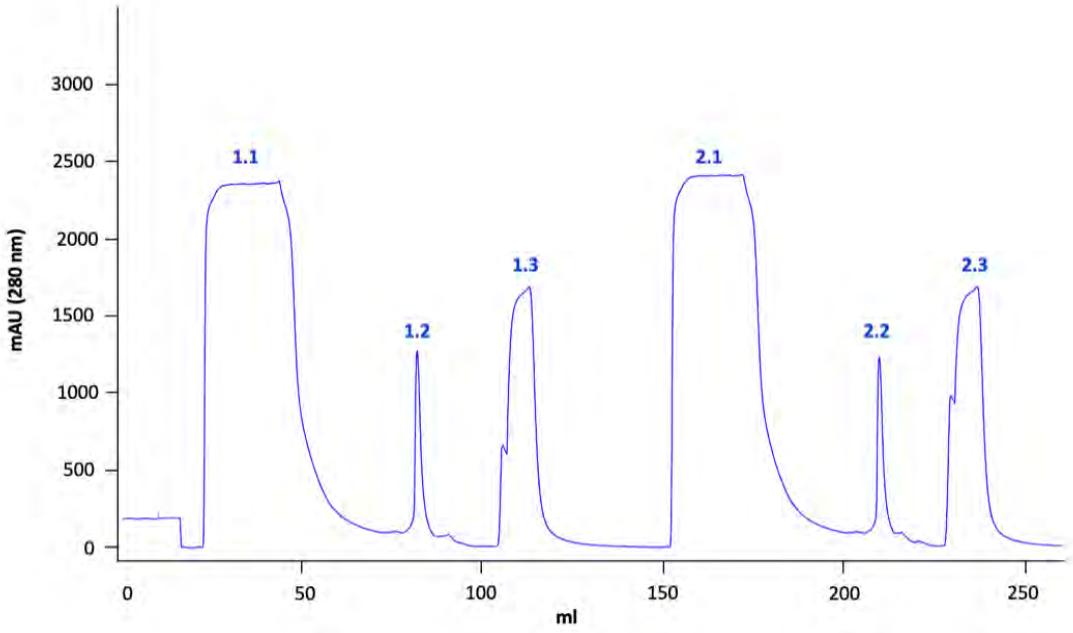


***Figure 4.58. Strep trap chromatography of BsPdaC CD R315A**. Strep-trap column (5 ml). Two loadings (1 and 2) with different steps: 1.1 and 2.1) lysate loading and washing with PBS; 1.2 and 2.2) elution with d-desthiobiotin; 1.3 and 2.3) column regeneration with HABA. The obtained profiles were the same for all the purified mutants.*

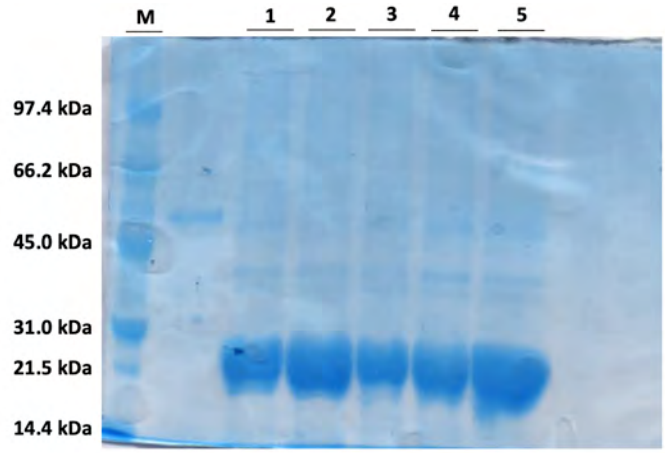


Figure 4.59. SDS-PAGE 14% acrylamide of R315A mutants isolated by affinity chromatography. *M: Marker Low Range (Biorad); 1: BsPdaC-CD-wt; 2: BsPdaC-CD-R315A; 3: BsPdaC-CD-R315I; 4: BsPdaC-CD-R315S; 5: BsPdaC-CD-R315K.*

By contrast, when isolated for the Michaelis-Menten assay (see 4.4.4. section bellow), the R315A mutant was purified with two steps and only the monomeric form obtained from size exclusion chromatography was used for the assay (fraction 3 on Figure 4.60).

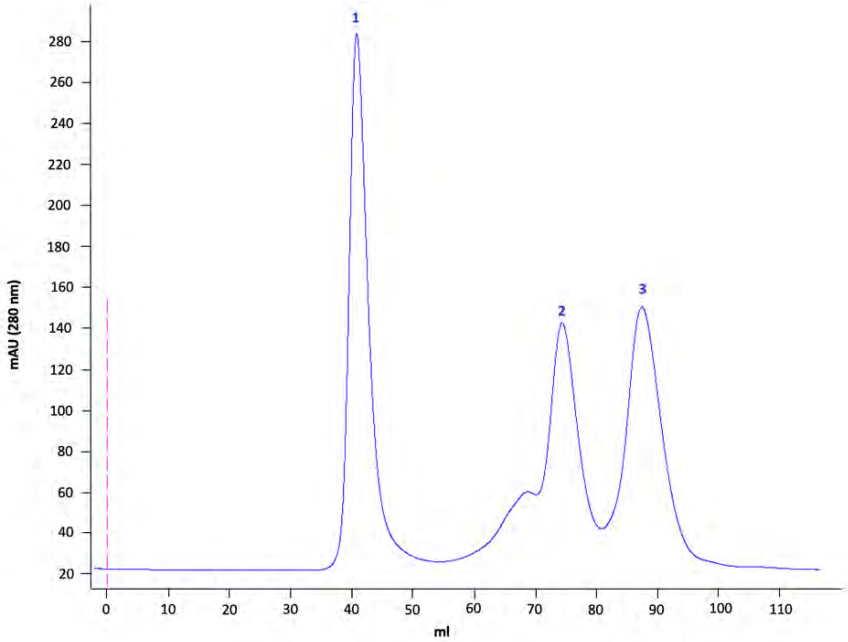


Figure 4.60. Size exclusion chromatography of BsPdaC CD R315A. *Superdex 200. The three characteristic oligomeric fractions of BsPdaC were obtained: 1) soluble oligomers; 2) dimeric fraction; 3) monomeric fraction.*

The expression yields obtained with the mutants were similar to those previously reported for the wild type enzyme, obtaining between 10 and 20 mg of total protein per liter of culture from the strep trap chromatography in all cases. Regarding the size exclusion's profile of the alanine mutant, it was again the same as the one showed by the wild type enzyme, with three characteristic oligomeric fractions of the enzyme (Figure 4.60).

4.4.2.2. Determining the change in specificity: coupled assay

In order to rapidly evaluate if Arg315 mutations induced a change in the enzyme's specificity, a coupled assay based in the combined action of *Bs*PdaC and *Vc*CDA was designed.

As already described, when PdaC acts on $(GlcNAc)_4$, it first renders the two mono-deacetylated products, ADAA and AADA, in the same proportion (approximately 50% each). On the other hand, *Vc*CDA is highly specific and performs a unique catalytic event deacetylating the second position of the substrate from the non-reducing end, generating the ADAA product. Therefore, if this second enzyme is incubated with the mixture of mono-deacetylated products from *Bs*PdaC, it will only be able to act on the AADA product, generating the di-deacetylated product ADDA. By comparing the formation of this di-deacetylated product from the initial mixture generated by *Bs*PdaC, the differences between the wild type enzyme and Arg315 mutants can be assessed and translated into specificity changes (Figure 4.61).

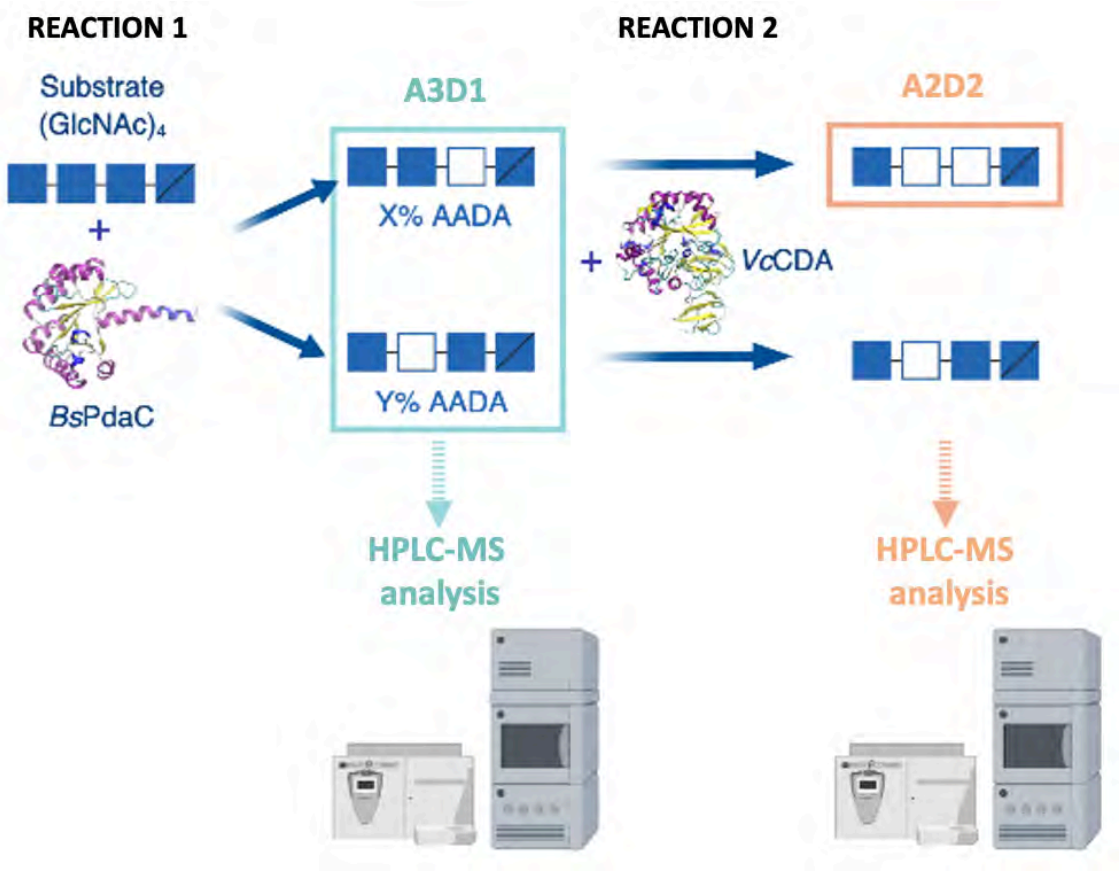


Figure 4.61. Set up of the coupled enzymatic assay. *A first reaction with BsPdaC CD (wild type or mutants) was set up at short time to obtain only mono-deacetylated product. After separating the enzyme from the sample, this was analyzed by HPLC-MS. The enzyme VcCDA was added for the second long-term reaction and the final sample was also analyzed by HPLC-MS. The comparison between the ratio of A2D2 product obtained on the second reaction and the A3D1 product obtained on the first reaction was used to determine possible changes in the mutants' specificity.*

Several considerations had to be taken into account when designing the experimental setup of the assay:

- Reaction 1 conditions (with *Bs*PdaC, wild type or mutants): knowing that PdaC produces several deacetylation products, it was necessary to determine the optimal conditions (enzyme concentration and reaction time) for the incubation of the enzyme with the initial $(GlcNAc)_4$ substrate that would render mainly the mixture of mono-deacetylation products and almost no products of further deacetylations.
- Stopping reaction 1: this first reaction should be stopped, the enzyme should be inactivated or separated from the COS mixture that will be used as substrate for *Vc*CDA in the following reaction.
- Degree of conversion of reaction 2: to be able to compare the ratios between products of the different enzymes is critical that the conversion of the AADA product from the first reaction is totally converted to ADDA by *Vc*CDA in the second reaction. Thus, a suitable amount of *Vc*CDA needed to be added for a sufficient time.

After testing different enzyme concentrations and reaction times with all the enzymes purified by strep-trap affinity chromatography, it was decided to perform the first reaction with an enzyme (*Bs*PdaC wild type or mutants) concentration between 1 and 5 µM during 15 to 30 minutes at 37°C (Figure 4.62). In order to rapidly stop the reaction after the decided time and to obtain the products' mixture sample that will be used for the second reaction with *Vc*CDA, the enzyme *Bs*PdaC was separated from the rest of the reaction components by using an ultra-centrifugal filter unit with a cutoff of 3 kDa. In order to confirm that the deacetylation reaction did not continue after this step, the flow through sample was incubated for 3 hours at 37°C and analyzed (Figure 4.63). For the consecutive enzymatic reaction with *Vc*CDA, between 2,5 and 5 µM of enzyme were added and incubated for at least 48 hours at 37°C before analyzing the final sample with the HPLC-MS (Figure 4.64).

The procedure was followed twice for the wild type enzyme and for the four R315 mutants to determine a possible change in the proportions of the two main mono-deacetylation products generated by the different mutants in comparison to the original enzyme.

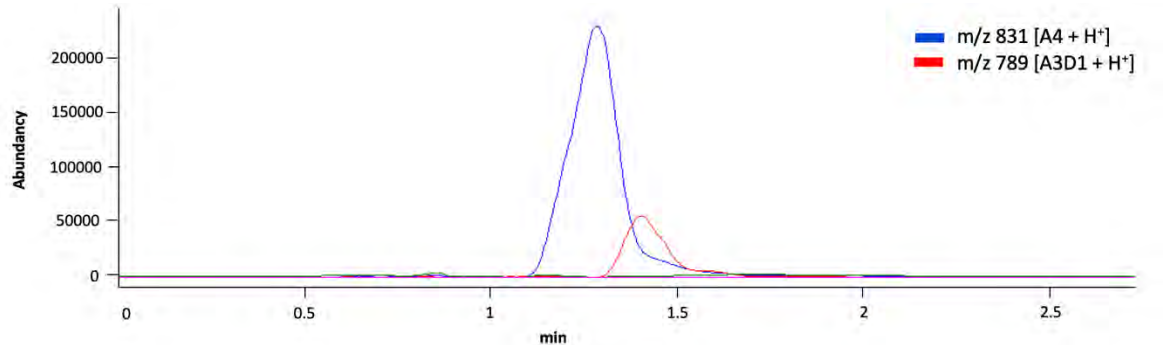


Figure 4.62. HPLC-MS analysis of reaction 1 products. *Example of the chromatographic profile obtained for the first reaction (with BsPdaC CD wild type or the different arginine mutants). The chosen conditions (in this example, 1 µM of BsPdaC, 2 mM $(GlcNAc)_4$, 15 minutes, 37°C) render a sample with the substrate and the product of the first deacetylation only.*

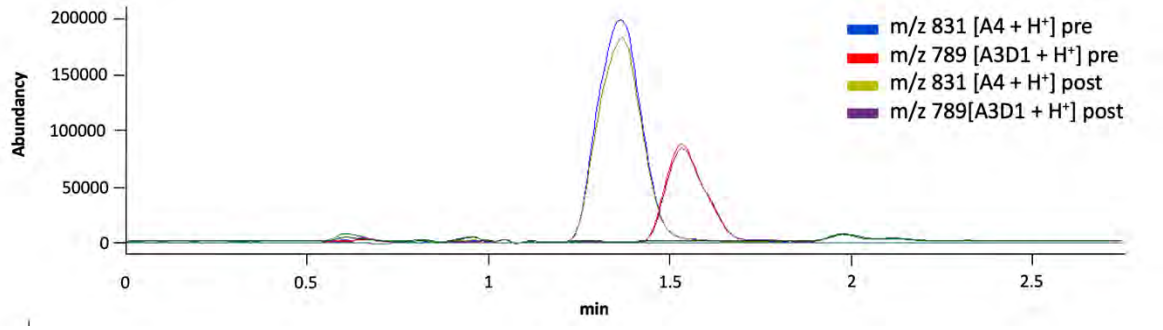


Figure 4.63. HPLC-MS analysis of reaction 1 products before and after enzyme separation*. Verification that the deacetylation reaction doesn't continue after separating the enzyme from the chitooligosaccharides presents in the sample. Pre) sample before passing through the ultra-centrifugal filter unit; post) flow through of the ultra-centrifugal filter unit after being incubated for 3 hours at 37°C.*

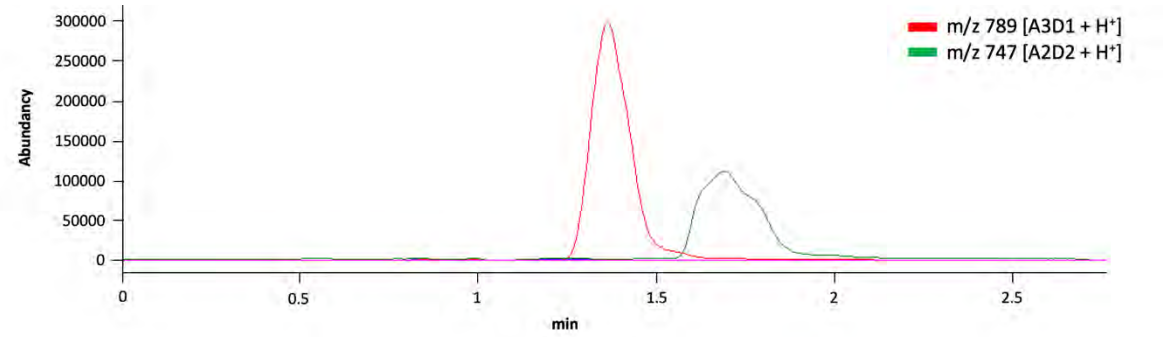


Figure 4.64. HPLC-MS analysis of reaction 2. *Example of the chromatographic profile obtained for the second reaction (with VcCDA). All reactions were set up at final point conditions in order to obtain a total conversion of the enzymes' substrates.*

The monitoring of the deacetylated products from the two consecutive enzymatic reactions was performed by HPLC-MS analysis as described before (see Section 7.4.2). The possible effect that the different mutations may had on the enzyme's specificity was evaluated by the comparison between the amount of di-deacetylated product generated by *Vc*CDA in the second reaction and the amount of the mono-deacetylated ones produced by *Bs*PdaC (wild type or mutants) in the first

reaction. Since there were not available standards for all the products, it was not possible to determine their concentration and the area units obtained with the HPLC-MS were used for calculating the specificity ratio (Equation 4.7).

The response factor for both mono-deacetylated products was considered the same, and different from the response factor of the di-deacetylated product, and the following variables were assigned to each product concentration:

- X = amount of ADAA (product from binding mode 2, BM2)
- Y = amount of AADA (product from binding mode 1, BM1)
- Z = amount of ADDA

$$x + y = A_M \; x \; F_M \qquad \textit{(Equation 4.1)}$$

$$z = A_D \; x \; F_D \qquad \textit{(Equation 4.2)}$$

(where A = area units; F = response factor; M = mono-deaceylated product; and D = di-decatylated product)

At the end of the second reaction, in which there is a complete conversion from AADA to ADDA, the amount of both products is the same (z = y):

$$\frac{BM2}{BM1} = \frac{x}{y} = \frac{(A_M \; x \; F_M) - (A_D \; x \; F_D)}{A_D \; x \; F_D} = \left(\frac{A_M}{A_D}\right) \; x \; \left(\frac{F_M}{F_D}\right) - 1 \qquad \textit{(Equation 4.3)}$$

Knowing that the wild type form of *Bs*PdaC renders the same amount of both mono-deacetylated products, the relation between them is 1 ($x/y = 1$), and it is possible to calculate the relation between response factors of the mono- and di-deacetylated products ($F_{M/D}$):

$$F_{M/D} = \frac{F_M}{F_D} \qquad \textit{(Equation 4.4)}$$

$$\frac{A_M}{A_D} \; x \; F_{M/D} - 1 = 1 \qquad \textit{(Equation 4.5)}$$

$$F_{M/D} = 2 \; x \; \frac{A_D}{A_M} \qquad \textit{(Equation 4.6)}$$

Having determined the relation between response factors, it is possible to calculate the specificity ratio (BM2/BM1) for each mutant (Equation 4.7) and compare the values with the wild type enzyme and between them (Table 4.14).

$$\boldsymbol{\frac{BM2}{BM1} = \frac{A_M}{A_D} \; x \; F_{M/D} - 1} \qquad \textit{(Equation 4.7)}$$

Table 4.14. Specificity ratio. *Ratio between the amount of product from BM2 and the amount of product from BM1 for each Arg315 mutant. Values are compared to the wild type enzyme (BM2/BM1 = 1).*

ENZYME	BM2/BM1
R315K	$1.43 \pm 0{,}06$
R315S	$2.89 \pm 0{,}52$
R315I	$3.99 \pm 0{,}18$
R315A	$5.15 \pm 0{,}05$

Specificity ratios for all mutants are higher than 1, meaning that more of ADAA product was produced than AADA product in all cases. These results suggest that, by mutating the residue arginine 315, which interacts with the non-reducing end of the substrate in BM1, the pattern of deacetylation of the wild type enzyme was altered (destabilizing BM1 and favoring the formation of the BM2 complex).

Differences can be observed in the values for the ratios of the four R315 mutants, being R315K the closest to the wild type. This could possibly be due to the fact that the positive charge of the original amino acid is conserved with this mutation, indicating a role of the charge in substrate binding stabilization. Not just the charge remains unchanged with this mutation but also lysine's ability to form, at least, one of the two hydrogen bond interactions, despite having a smaller side chain.

As for the rest of the mutants, in which the positive charge is lost, the values are notably higher than the wild type's one, and the ratio increases as the number of possible hydrogen bonds formation and the size of the substitute residue are reduced. The serine mutant, that would be able to establish one hydrogen bond, shows a lower ratio than R315I and R315A, both unable to interact with the substrate via hydrogen bond interactions. The alanine mutant, being the smallest substitute residue and the most drastic change from the original arginine, is the one that presents a greater difference with respect to the wild type.

4.4.2.3. Sequencing the product's pattern of acetylation

The described coupled assay showed that the mutations had some effect on *Bs*PdaC's specificity, but it did not provide the final information on the specific pattern of acetylation of the mono-deacetylated products generated by the different mutants.

In order to relate the increasing ratios with the real change in the exhibited deacetylation pattern, preparative reactions with R315 mutants were set up with the previously described conditions. The

PA of the products obtained with these reactions was determined in Dr. Bruno M. Moerschbacher laboratory in Münster (Germany) with the same method that was previously used with the wild type enzyme [123]. With this methodology, it was possible to perform a quantitative analysis and determine the percentage of each mono-deacetylated product inside the A3D1 mixture present in the samples. Values presented in Table 4.15 show that, for all cases, the mixture of mono-deacetylated products generated by the different R315 mutants are composed mainly of the ADAA product (from 65 to 81% depending on the sample). The remaining percentages of mono-deacetylated products of the mixtures are composed mostly the AADA product, appearing also with very low representation the DAAA product.

Table 4.15. MALDI-TOF-MS/MS analysis of the mono-deacetylated products from $(GlcNAc)_4$. *Pattern of acetylation and relative intensity of each product in the samples of deacetylation reactions with the different R315 mutants. A) Product with of relative intensity values < 0.05, below the limit of quantification, are considered "non-formed products".*

	SEQUENCE	RELATIVE INTENSITY
	R315K	
Sample 1	DAAA	0,09
	AADA	0,18
	ADAA	**0,73**
Sample 2	DAAA	0,06
	AADA	0,28
	ADAA	**0,66**
	R315S	
Sample 1	DAAA	0,07
	AADA	0,28
	ADAA	**0,65**
Sample 2	DAAA	0,04
	AADA	0,25
	ADAA	**0,71**
	R315I	
Sample 1	DAAA	0,06
	AADA	0,13
	ADAA	**0,81**
Sample 2	DAAA	0,05
	AADA	0,17
	ADAA	**0,78**
	R315A	
Sample 1	DAAA	0,08
	AADA	0,2
	ADAA	**0,71**
Sample 2	AAAD	0,01[A]
	DAAA	0,06
	AADA	0,11
	ADAA	**0,81**

The lysine and serine mutants are the ones with lower concentrations of ADAA product and the isoleucine and alanine ones have shown to produce a higher percentage of this product. The obtained sequences are in agreement with the changes in ratios observed with the coupled

enzymatic assay, supporting the statement made that the interactions that arginine 315 establishes with the substrate are of high importance for the stabilization of BM1.

Furthermore, the congruity of both sets of data reinforces the usefulness of the assay to determine possible effects on CE4 enzyme's specificities. Having information on the deacetylation pattern exhibited by the characterized deacetylases, it could be possible to adapt it by using different enzymes depending on the initial pattern and the expected changes.

4.4.2.4. Enzyme kinetics of R315 mutants

In the previous points it has been shown that the mutation of the residue arginine 315 affected the specificity of the enzyme, altering the proportions of the mono-deacetylated products generated. In this section, not the specific pattern of acetylation of the products, but the overall deacetylation activity of the different enzymes will be evaluated.

4.4.2.4.1. Specific activity of *Bs*PdaC wt and R315 mutants

It is necessary to stress that the protocol followed for the purification of the enzymes consisted on just an affinity chromatography step with a Strep-trap column, without separating the different oligomeric fractions of the proteins. The previous characterization of the wild type enzyme showed that the monomeric form was the one presenting the highest activity and it was always the fraction used for the kinetic characterization of the enzyme. Taking this into account, it is expected the enzyme's activity of these assays to be lower.

The specific activity for the wild type enzyme and all the mutants was determined by monitoring the total mono-deacetylated product formation by HPLC-MS. Reactions were set up with 2mM $(GlcNAc)_4$ substrate at 37°C and short reaction time (when less than 5% of di-deacetyated products were produced) (see Section 4.4.2).

As a general trend, the specific activity of all mutants is very similar to the one determined for the wild type enzyme (Figure 4.65 and Table 4.16). Although it has been demonstrated that the modification of the arginine residue has an effect on the pattern of the enzyme's deacetylation and the proportion of the mono-deacetylated products has been altered, these changes do not seem to affect the overall mono-deacetylation activity.

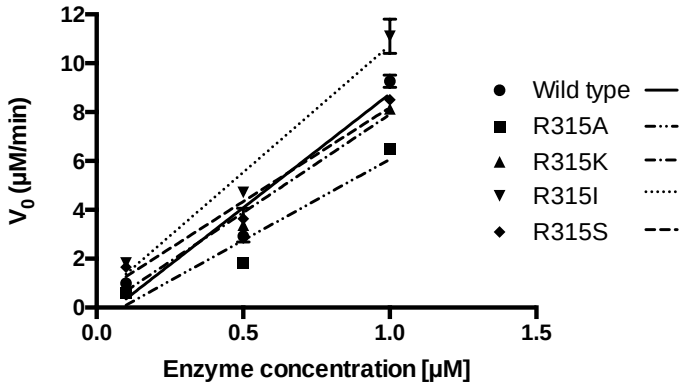


Figure 4.65. Specific activity (µM·min^{-1}·µM) of the wild type enzyme and the different arginine 315 mutants.

Table 4.16. Specific activity (min^{-1}) of the wild type enzyme and the different arginine 315 mutants.

ENZYME	SPECIFIC ACTIVITY (MIN^{-1})	% OF ACTIVITY COMPARED TO WT
Wild type	9.33 ± 0.31	
R315K	7.99 ± 0.06	86%
R315S	8.04 ± 0.6	86%
R315I	10.4 ± 0.8	111%
R315A	6.6 ± 0.02	70%

4.4.2.4.2. Michaelis-Menten kinetics of R315A mutant

The change in the formation of the first deacetylation products represents differences in the way the substrate binds to the catalytic center. The R315A mutant showed a higher preference to form the complex with $(GlcNAc)_4$ that would render the ADAA mono-deacetylated product in comparison to the wild type enzyme. In order to evaluate if the general affinity of the mutant for this substrate had also been altered, the Michaelis constant (K_M) was determined. K_M is the substrate concentration at which the reaction rate is at half-maximum and it is an inverse measure of the substrate's affinity for the enzyme.

It was only feasible to determine the K_M value for the overall generation of mono-deacetylated products, with no possibility to have this information for the ADAA and AADA products separately. In this case, to be able to compare the results with the wild type kinetic parameters previously described, the mutant R315A was isolated following the two steps purification and only the monomeric fraction was used for the enzymatic assay.

To define the K_M parameter that describes the enzymatic behavior, 12 enzymatic reactions with different substrate concentration (from 0,5 to 17,5 mM) were performed by duplicate, taking 8 temporal points for each reaction. The obtained results showed no changes in the value of the

Michaelis constant K_M between the wild type enzyme and the alanine mutant (Figure 4.66 and Table 4.17). It is possible that both binding modes, although being represented in different percentages, present the same kinetic parameters, that being the reason why there is no change in these values when analyzing the mono-deacetylation activity of the mutant. In the same way as what it was observed with the specific activity determination, the conservation of these parameter indicates a similar performance of both enzymes when considering their overall action.

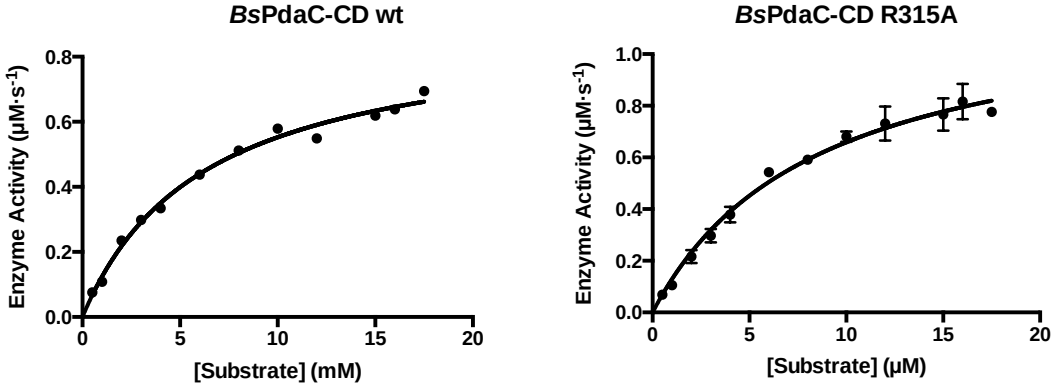


Figure 4.66. Kinetic characterization of wild type enzyme and R315A mutant. *Michaelis-Menten curves.*

Table 4.17. Kinetic parameters K_M of wild type enzyme and R315A mutant.

Enzyme	K_M (mM)	Kcat (s^{-1})
Wild type	6,25 ± 0,47	0.93 ± 0.3
R315A	8,44 ± 0,9	0.28 ± 0.01

4.5. General discussion on *Bs*PdaC

Chapter 1 presents the characterization of the enzyme *Bs*PdaC within the work developed in the area of the deacetylases from family CE4. Previous achievements of the group included solving the first crystallographic structure of a CE4 enzyme in complex with substrates, in collaboration with Dr. Marcelo E. Guerin's group [24], and postulating the Subsite Capping Model, which describes the general structural determinants that dictate the specificity of this family of enzymes. This previous expertise set the basis and perspectives for the work with *Bs*PdaC.

The catalytic domain of the enzyme was isolated (*Bs*PdaC-CD), modification that did not significantly affect protein expression or stability. The availability of a smaller enzymatic unit can be an advantage in terms of production when using it as a biocatalyst or if it has to be incorporated into a production strain.

Enzyme's characterization showed that PdaC is a metallo-enzyme active on both PGN and COS substrates, in which it follows a multiple-chain mechanism. The metal dependence of the enzyme was proved by the fact that the removal of metal cation leads to an inactive enzyme whose activity is restored by reconstitution with Zn^{2+} cation. This is in contrast with the previous report by Kobayashi and coworkers [121] that suggested that *Bs*PdaC might not be a metallo-enzyme since it retained activity after treatment with EDTA. However, as shown here, protein unfolding and refolding was necessary to fully dissociate the metal cation that is strongly bound to the active site. The enzyme showed high activity on intact peptidoglycan and, interestingly, also on COS with DP_{3-5}, deacetylating all substrate residues but the reducing-end ones.

The crystal structure of *Bs*PdaC-CD was solved, confirming the short length of its loops that result in an open binding cleft, characteristics that could be interfered from the initial bioinformatic analysis of the protein sequence. The deacetylation pathway followed by the enzyme on COS substrates fits the Subsite Capping Model, reinforcing the role of the loops in determining the specificity of each CE4 enzyme. The fact that the characterized CE4 enzymes with known structure also fit the model shows that it is possible to generate hypothesis with the Subsite Capping Model as a base that would allow the modification of the enzymes' behavior by engineering their loops. It could be possible to generate tailored deacetylases from a library of loops, defining the substrates in which they would be active as well as the pattern of acetylation of their products. This engineering strategy could allow obtaining new patterns that may not exist in nature or even incorporating non-natural substrates to be deacetylated.

It has also been shown that not only size and shape of loops are guiding substrate specificity of CE4 enzymes, but also specific interacting residues make a contribution. By mutating Arg315 and, thus, disrupting its interaction with the substrate in one of the main binding modes showed by PdaC, the specificity of the enzyme was shifted to produce a higher amount of the other initial product. Furthermore, surface charge distribution along the binding cleft and other structural features may also participate in defining the mode of action and deacetylation pattern exhibited by each particular enzyme [47]. Consequently, fine tuning of PA engineering in this family of enzymes should combine the knowledge based on the Subsite Capping Model and the specific features of sequence and structure of the different enzymes.

The double specificity of the enzyme, active on MurNAc residues of intact peptidoglycan and GlcNAc residues of COS, is a unique behavior among characterized MurNAc deacetylases. *Bs*PdaA

does not deacetylate COS and acts on modified peptidoglycan in which the peptidyl substitutions have been previously removed [49]. Several differential features were identified when comparing the sequence and the obtained crystal structure of PdaC with those of *Bs*PdaA and *Sp*PgdA. Although being a peptidoglycan MurNAc deacetylase, in terms of sequence and metal coordination *Bs*PdaC shows higher similarity to the peptidoglycan GlcNAc deacetylase *Sp*PgdA than to *Bs*PdaA. Considering the surface electrostatic potential at the active site of these enzymes, *Bs*PdaA has a positively charge binding cleft, as opposed to *Sp*PgdA, in which it is negatively charged. *Bs*PdaC (which can productively bind both GlcNac and MurNAc residues) lays in between, with positive and negative patches in its binding cleft.
Comparison of the substrate binding cavities of the three enzymes allowed the identification of key sequence differences that could explain the differential behavior of *Bs*PdaC.

In terms of the biological function of peptidoglycan deacetylation, GlcNAc deacetylases are best characterized for their function in conferring lysozyme resistance to the bacterial cell wall PGN in order to evade the host immune system [52], [53] and have been proposed as novel anti-bacterial targets. In contrast, MurNAc deacetylation of PGN has been involved in sporulation and germination in different *Bacillus* species participating in muramic δ-lactam synthesis [128]. However, *Bs*PdaC is not associated with sporulation, since a pdaC-deficient strain was not affected in sporulation and germination and the only reported phenotype was lysozyme sensitivity [121]. PGN hydrolysis by host lysozymes is an initial event for triggering the innate immune responses. Besides deacetylation of polymeric PGN to confer lysozyme resistance, other modifications of PGN fragments (muropeptides) may impair the detection by host immune receptors. N-acetylmuramyl dipeptide (MDP) was shown to be the minimal PGN fragment to directly interact with Nod2 receptors, ultimately resulting in activation of inflammatory response via the NF-κB and MAP kinase signaling pathways [69]. Recently, it has been reported that N-deacetylated MDP did not activate NF-κB, thus, suggesting that the MurNAc acetyl group is an important feature for recognition by the Nod2 signaling cascade [129]. Whether MurNAc deacetylases act only on the polymeric PGN or also on PGN fragments after the action of lytic enzymes, including lysozyme, is unknown. *Bs*PdaC is currently the only characterized MurNAc deacetylase active on peptide substituted PGN chains, but the minimal size of its PGN substrates is yet unknown. Further characterization of this new subclass of MurNAc deacetylases will certainly contribute to unravel the complex degradation-modification-signaling mechanisms in bacteria-host interactions.

One of the most urgent needs in the field of peptidoglycan enzymology is the availability of pure, defined and biologically relevant substrates that will allow the complete exploration of enzymes' substrate specificity.

Based on the described differential functional and structural characteristics of *Bs*PdaC, it is proposed to be the first member of a new subclass of peptidoglycan MurNAc deacetylases. It is possible that enzymes with PdaC-type MurNAc deacetylase activity are present in pathogenic bacteria, but not yet identified because of the lack of a clear sequence signature to distinguish GlcNAc and MurNAc deacetylases. Further characterization of this new subclass of MurNAc deacetylases will certainly contribute to unravel the complex degradation-modification-signaling mechanisms in bacterial-host interactions.

The discover and characterization of new peptidoglycan deacetylases would contribute to better understand the sequential and structural features that are determining the specificity of these enzymes on peptidoglycan.

4.6. Publication of *Bs*PdaC structure and characterization

J. Biol. Chem, 294(50), 19066. 2019

Structure-function relationships underlying the dual N-acetylmuramic and N-acetylglucosamine specificities of the bacterial peptidoglycan deacetylase PdaC.

Laia Grifoll-Romero[‡], María Angela Sainz-Polo[§], David Albesa-Jové[§], Marcelo E. Guerin[§], Xevi Biarnés[‡], and Antoni Planas[‡].

[‡]Laboratory of Biochemistry, Institut Químic de Sarrià, Universitat Ramon Llull, 08017 Barcelona, Spain.

[§]Structural Biology Unit, Center for Cooperative Research in Biosciences (CIC bioGUNE), Bizkaia Technology Park, Ed. 801A, 48160 Derio, Spain, and the Basque Foundation for Science (IKERBASQUE), 48011 Bilbao, Spain.

1. Introduction
2. Results
3. Discussion
4. Conclusion
5. Materials and methods

During the course of this work, an original article was published on the structure-function relationship of substrate specificity of *Bs*PdaC. The complete article can be consulted on Annex 9.7.

CHAPTER 2: Discovery of new CE4 enzymes

5. CHAPTER 2: Discovery of new CE4 enzymes

The use of chitin deacetylases and related enzymes as biocatalysts for the conversion of chitin to chitosan (or chitin oligomers to chitosan oligomers) in a controlled process offers the possibility of producing partially acetylated COS with tailored patterns of acetylation. Although the number of characterized enzymes has increased in recent years, not all patterns are yet available and enzyme discovery and protein engineering could offer new opportunities for the biotechnological production of these products. Gaining further structural information to decipher both the structural and sequential determinants of substrate specificity in CE4 family will be important to allow the rational design or discovery of novel deacetylases with controlled specificities.

In the case of peptidoglycan deacetylases, which act on either the MurNAc or GlcNAc residues of peptidoglycan, there is even less available information, with few characterized enzymes. The study that was carried out on *Bs*PdaC in the previous chapter showed the enzyme to be a MurNAc deacetylase with activity also on GlcNAc residues. To date, the other characterized peptidoglycan deacetylases are pure GlcNAc or MurNAc deacetylases and their activity on PGN depends on the presence or absence of the peptidic part of the substrate. Discovery and characterization of novel peptidoglycan deacetylases is needed in order to better understand the difference between subgroups and the substrate specificity of these enzymes.

5.1. Selection of novel enzymes

Two different strategies were followed for the searching of novel chitin deacetylases and peptidoglycan deacetylases separately.

5.1.1. New CDAs: in the search of new deacetylation patterns

The search for novel chitin deacetylases that could present new deacetylation patterns was based on a previous phylogenetic and bioinformatic study of the family based on the “Subsite Capping Model” and loop sizes.

5.1.1.1. Phylogenetic and bioinformatics study of CE4 family

In previous work at the laboratory (Romero, Biarnés and Planas, unpublished), a phylogenetic study of CDAs and related proteins was performed. Twelve characterized members of CE4 family were initially chosen (Table 5.1) for the search of related new sequences[3].

Table 5.1. Reference enzymes for phylogenetic analysis. *Enzymes from CE4 family chosen as reference for the search of new sequences. The pattern of deacetylation presented on the third column refers to products of COS deacetylation.*

ACTIVITY	ORGANISM (ENZYME)	PA (COS)	PDB	UNIPROT
Chitin-de-*N*-acetylase (CDA)	*Rhizobiym mellioti* (*Rm*NodB)	DA_{n-1}	-	P02963
	Vibrio cholera (*Vc*CDA)	ADA_{n-2}	4NY2	C3LLX6
	Colletotrichum lindemuthiamun (*Cl*CDA)	D_n	21W0	Q6DWK3
	Aspergilus nidulans (*An*CDA)	D_n	2Y8U	B3VDB86
	Puccinia graminis f. sp Tritici (*Pgt*CDA)	AAD_{n-2}	-	E3K3D7
	Podospora anserina (*Pa*CDA)	D_n	-	B2AAQ0
	Bacillus licheniformis (*Bl*CDA5)	-	-	Q65HY6
Acetylxylan esterase	*Streptomyces lividans* (*Sl*AxeA)	DDD/A1D2	1CC0	D6EHA8
Poly-β-1,6-GlcNAc deacetylase	*Escherichia coli* (*Ec*PgaB)	-	3VUS	P75906
Peptidoglyca deacetylases	*Streptococcus pneumoniae* (*Sp*PgdA)	ADA	2C1G	Q8DP63
	Bacillus subtilis (*Bs*PdaA)	n.a.	1NY1	O34928
	Bacillus subtilis (*Bs*PdaC)	$D_{n-1}A$	6H8L	O34798

From the initial multiple sequence alignment with the 12 reference sequences (alignment-1), a Hidden Markov Model profile (HMM-1) was built and used for the search against the Uniprot database (SwissProt and Trembl) with no specific conditions. A total of 35969 sequences over the threshold were retrieved. After removing sequences with ≥ 90% identity for redundancy reduction, a final set of 17492 sequences was obtained. An alignment with all the sequences guided by this initial HMM profile was performed and manually curated to include only the aligned region corresponding to the conserved catalytic domain of the family (alignment-2). Finally, a phylogenetic

[3] The Multiple sequence alignments and Hidden Markov Model profiles from this Section have been numbered and a detailed description of them can be found in Section 7.5.2.

tree including all sequences, both the initial reference enzymes and the new sequences found in the database search, was built with the *FastTree* method using the *Neighbor-joining* algorithm (Figure 5.1).

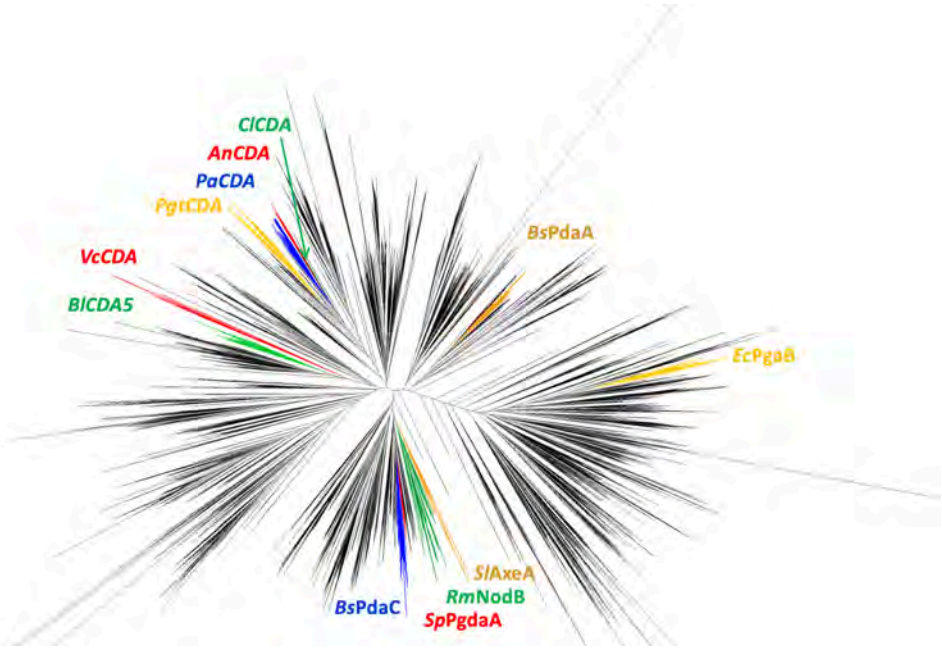


Figure 5.1. Phylogenetic tree with reference enzymes and new sequences retrieved from Uniprot database. *Branches which include each reference enzyme are colored.*

The resulting tree showed that, although there are big clusters of only eukaryotes or bacteria, all clusters share organisms of different kingdoms (Figure 5.2).

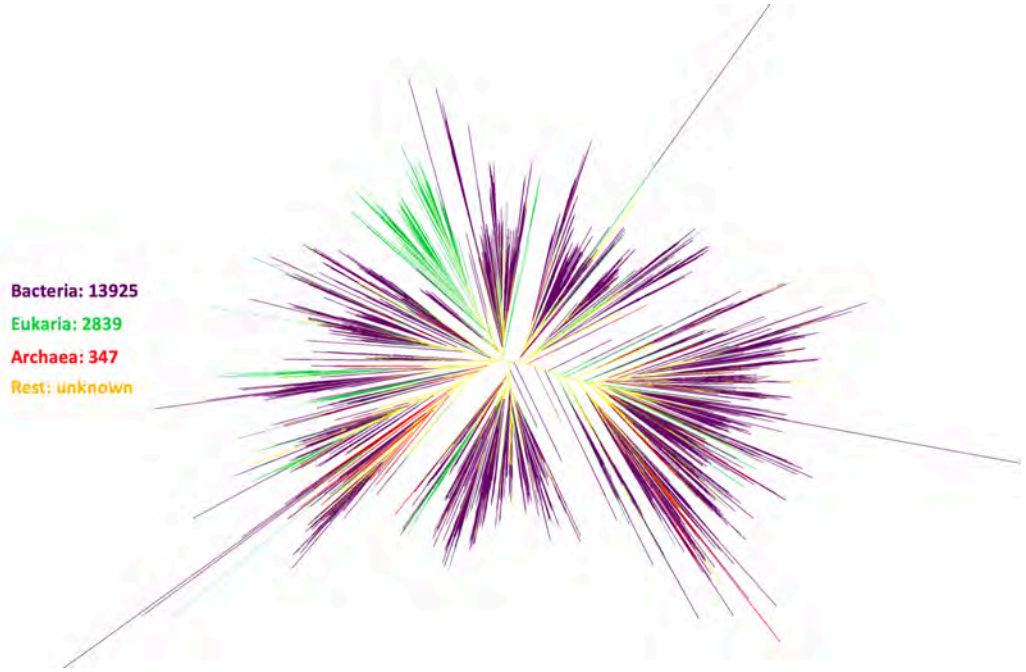


Figure 5.2. Phylogenetic tree with reference enzymes and new sequences retrieved from Uniprot database. *Classification by Kingdom.*

One branch of the tree was selected for each of the 12 reference proteins and rooted by that one. Branch selection was guided by the obtained bootstrap values (trying not to use any lower that 0.7) and by considering the shorter distances (trying not to take clusters considerably far) and the smaller differences in loop sizes (trying not to take branches considerably different). Once selected, each branch was rooted by the corresponding reference protein.

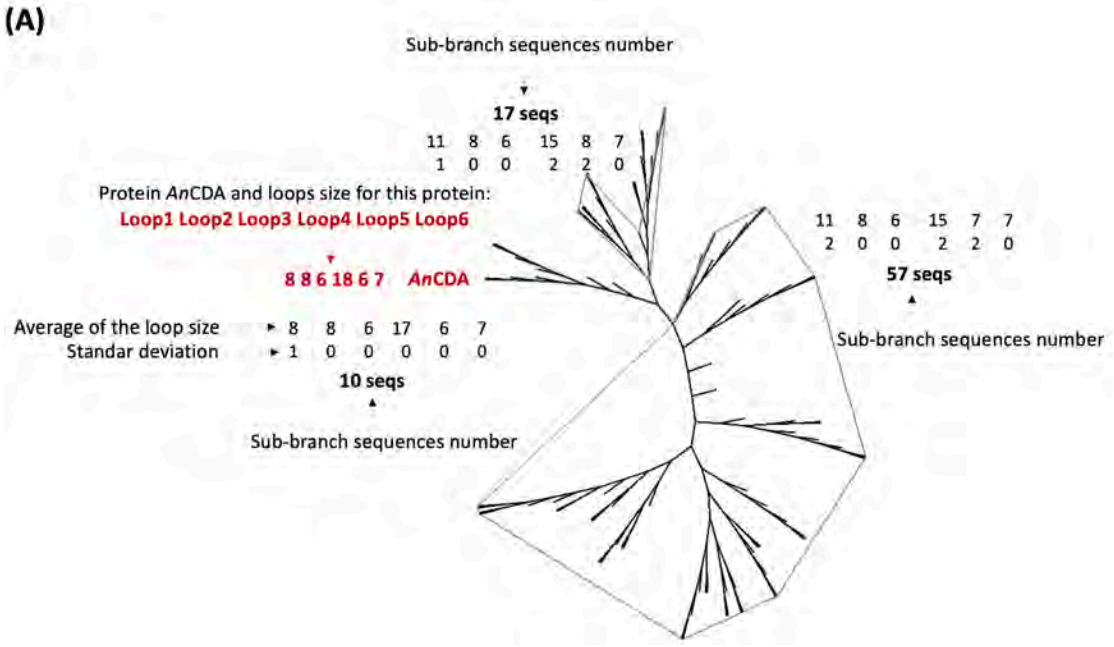
Inside each branch, sub-branches were clustered by loop size according to alignment-2, and the average and standard deviation of those sizes were calculated for all groups. SD values were used as a measure of average homogeneity of the loop size for each branch.

The infoalign tool from the EMBOSS package was used to obtain basic information about the sequences in alignment-2. Displayed information (for each sequence of the alignment) includes two measures of length, counts of gaps, and numbers of identical, similar, and different residues in the sequence when compared to the reference sequence, together with a simple statistic of the percentage of change between the reference sequence and this sequence (https://www.bioinformatics.nl/cgi-bin/emboss/infoalign). The percentage of change (Equation 5.1) was used as a numeric index of protein homology and, in general, proteins with a % change higher than 50 were not considered.

$$\%\ change = \frac{(Alignment\ length - Sequence\ length)\ x\ 100}{Alignment\ length} \qquad \textbf{(Equation 5.1)}$$

Alignment-1, which included the 12 initial enzymes (with the 6 loops from the Subsite Capping Model defined), was used to identify and extract the loops of each sub-branch. Each loop was re-aligned with Jalview and the Mafft algorithm (alignment-3) to generate a new HMM and logo with the web tool Skylign (HMM-2).

An example of the available information for each protein is represented in Figure 5.3., including the selected branch rooted for each reference sequence and the sub-branches clustered according to loop sizes, the information of the multiple sequence alignment, and the HMM logo for each loop and branch (which graphically represents the conservation of each position in the multiple alignment).



(B)

	LOOP SIZE						SeqLen	AlignLen	Gaps	GapLen	Ident	Similar	Differ	% Change
	L1	L2	L3	L4	L5	L6								
AnCDA	8	8	6	18	6	7	188	224	14	36	188	0	0	16,1
tr|?|Eukaryota|Emericella_nidulans|B3VD86	8	8	6	18	6	7	188	224	15	36	184	1	3	17,9
tr|?|Eukaryota|Aspergillus_calidoustus|A0A0U5CII6	8	8	6	18	6	7	188	224	15	36	152	17	19	32,1
tr|?|Eukaryota|Aspergillus_calidoustus|A0A0U5GNN1	8	8	6	17	6	7	187	224	14	37	111	31	45	50,4
tr|?|Eukaryota|Aspergillus_luchuensis|A0A146FX02	8	8	6	17	6	7	187	224	15	37	111	26	50	50,4
tr|?|Eukaryota|Aspergillus_parasiticus|A0A0F0IE87	8	8	6	17	6	7	187	224	15	37	106	35	46	52,7
tr|?|Eukaryota|Aspergillus_flavus|A0A0D9MXP5	8	8	6	17	6	7	187	224	14	37	107	37	43	52,2
tr|?|Eukaryota|Aspergillus_nomius|A0A0L1IM65	8	8	6	17	6	7	187	224	15	37	108	34	45	51,8
tr|?|Eukaryota|Aspergillus_terreus|Q0CZY5	10	8	6	17	6	6	174	209	15	35	86	23	65	58,9
tr|?|Eukaryota|Aspergillus_kawachii|G7XXM2	8	8	6	17	6	7	187	224	15	37	96	27	64	57,1

(C)

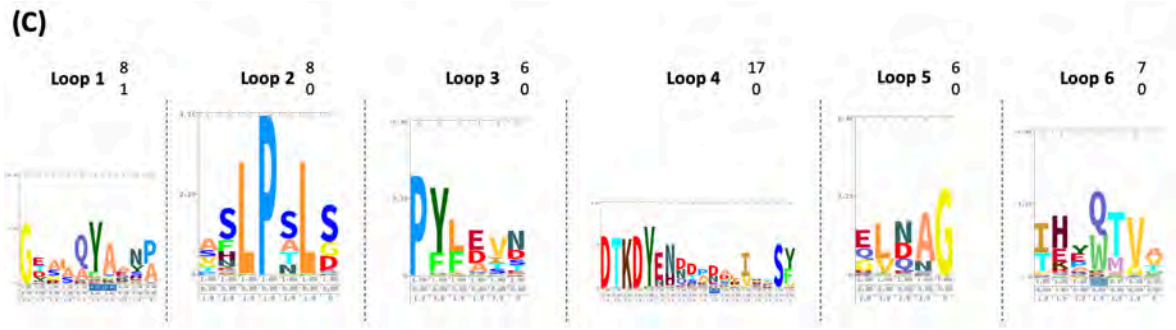


Figure 5.3. Example of the branch selected for the reference enzyme AnCDA. *(A) Branch selected for the protein AnCDA and rooted for this one. Three sub-branches with 10, 17 and 57 sequences, respectively, are shown; (B) Example of one of the sub-branches from AnCDA group (with 10 sequences). Information on loop size and the multiple sequence alignment for each sequence of the cluster; (C) Example of one of the sub-branches from AnCDA group (with 10 sequences). HMM logo for each loop of the cluster.*

5.1.1.2. Selection of new chitin deacetylases

This previous search and the analysis of the sequences retrieved from the bioinformatic study were the basis for the selection of novel chitin deacetylases with possible new patterns of deacetylation. Twelve initial groups with a variable number of sequences, corresponding to the branches selected for each reference enzyme, were considered and analyzed in detail.

Being the aim finding new patterns of deacetylation, the selection criteria was the maximum difference in loop size relative to the reference enzyme. In a first phase, which included the 12 initial groups, each of the loops was considered individually and the number of sequences was reduced by selecting just those sequences with $\pm$ 35% in loop size in comparison with their reference enzyme.

The second selection step was restricted to the seven groups with a CDA as a reference enzyme. Combined divergences in the size of the 6 loops of the sequences in comparison with the reference enzyme were considered, and the ones with the bigger changes were selected for further analysis. Finally, 29 sequences from the seven groups were chosen (Table 5.2). Since the selection was restricted to groups with CDAs as reference enzyme, the rest of the phylogenetic tree was not considered. Those unexplored areas are a potential source for the search of new enzymes.

A Hidden Markov Model guided alignment of the 29 sequences with known CDAs and CODs (list from [120], HMM-3) was performed (alignment-4, see Annex 9.4) and loop sizes were re-assigned according to the alignment.

As it can be observed in Table 5.3, in some of the sequences that showed the biggest differences in loops sizes in comparison with their reference, the value of that length was wrongly assigned. For example, the sequence from *Verticillium longisporum* on *An*CDA's group should have the longest loop 6, but when manually corrected with the new alignment, it was observed that the real length was much shorter and that the sequence did not show any special feature in terms of loops' size. Another case worth mentioning is the sequence from *Aphanomyces astaci* on *Vc*CDA group, initially selected for having a very short loop 1 in comparison with *Vc*CDA. The alignment showed that the initial part of the sequence does not properly align with the other enzymes and that motif 1, which contains the catalytic aspartate and one of the metal-binding residues, is missing.

Table 5.2. Selected sequences according to loop size. *29 sequences from the seven CDA groups (reference enzyme in bold). L1-L6) size of loops; PA) pattern of acetylation of the reference enzymes' products.*

	L1	L2	L3	L4	L5	L6	PA
***An*CDA**	**8**	**8**	**6**	**18**	**6**	**7**	**D_n**
tr|?|Eukaryota|Aspergillus_clavatus|A1CHT9	11	8	6	10	13	7	
sp|?|Eukaryota|Arthroderma_benhamiae|D4B5F9	11	8	6	12	11	7	
tr|?|Eukaryota|Hypocrella_siamensis|A0A173G939	16	8	6	13	10	7	
tr|?|Eukaryota|Microdochium_bolleyi|A0A136IZL6	17	8	6	13	10	7	
tr|?|Eukaryota|Verticillium_longisporum|A0A0G4KZQ2	11	8	6	15	7	42	
***Bl*CDA5**	**8**	**9**	**9**	**15**	**10**	**5**	**?**
sp|CE4|bacteria|Bacillus_subtilis|O07596	8	9	9	13	13	5	
tr|?|Bacteria|Turicibacter_sanguinis|A0A174U1X5	11	12	9	14	10	5	
tr|?|Bacteria|Turicibacter_sanguinis|D4W4M0	11	12	10	13	11	5	
tr|?|Bacteria|Turicibacter_sanguinis|A0A173UFV9	11	12	9	15	10	5	
***Cl*CDA**	**13**	**8**	**6**	**17**	**10**	**7**	**D_n**
tr|?|Eukaryota|Verticillium_longisporum|A0A0G4M7P9	14	8	6	12	15	7	
***Pa*CDA**	**13**	**8**	**6**	**12**	**11**	**7**	**D_n**
tr|?|Eukaryota|Ophiocordyceps_sinensis|T5AHU3	17	40	6	17	10	7	
tr|?|Eukaryota|Colletotrichum_incanum|A0A161WA77	17	41	6	16	12	7	
tr|?|Eukaryota|Blumeria_graminis|N1JI69	31	8	7	20	8	7	
tr|?|Eukaryota|Metarhizium_majus|A0A0B4H3A4	41	8	6	16	12	7	
***Pgt*CDA**	**13**	**8**	**6**	**17**	**5**	**3**	**AAD_{n-2}**
tr|?|Eukaryota|Puccinia_sorghi|A0A0L6U7D9	13	8	14	13	9	3	
tr|?|Eukaryota|Melampsora_larici-populina|F4R898	13	8	6	10	12	3	
tr|?|Eukaryota|Tilletia_controversa|A0A177UF72	13	8	6	10	10	3	
tr|?|Eukaryota|Puccinia_sorghi|A0A0L6UEW6	33	8	12	12	8	3	
***Rm*NodB**	**8**	**8**	**6**	**13**	**5**	**18**	**DA_{n-1}**
sp|CE4|bacteria|Rhizobium_leguminosarum|P04339	8	8	6	10	9	20	
sp|CE4|bacteria|Rhizobium_leguminosarum|P04676	8	8	5	10	9	20	
sp|CE4|bacteria|Rhizobium_galegae|P50354	8	8	6	10	9	20	
sp|CE4|bacteria|Azorhizobium_caulinodans|Q07740	8	8	6	10	9	20	
***Vc*CDA**	**16**	**31**	**45**	**19**	**34**	**16**	**ADA_{n-2}**
tr|?|Eukaryota|Saprolegnia_parasitica|A0A067CB02	27	51	41	13	17	9	
tr|?|Eukaryota|Saprolegnia_parasitica|A0A067CA64	27	51	41	18	14	9	
tr|?|Eukaryota|Saprolegnia_diclina|T0QIU3	26	44	41	16	18	9	
tr|?|Eukaryota|Aphanomyces_astaci|W4GU91	5	39	41	21	14	9	
tr|CE4|bacteria|Geobacter_daltonii|B9M4W6	8	9	6	17	9	8	
tr|?|Bacteria|Weeksella_virosa|F0NZ07	10	5	20	14	16	9	
tr|?|Bacteria|Nitratifractor_salsuginis|E6WXT0	8	12	18	15	16	9	

Table 5.3. Re-assigned loop sizes according to multiple sequence alignment. *For each sequence: upper row) initial loop size; lower row) new loop size, according to alignment-4. Only 4 of the 7 groups are represented because loops lengths from all sequences on BlCDA5, ClCDA and RmNodB groups were correct.*

	L1	L2	L3	L4	L5	L6
***Vc*CDA**	**16**	**31**	**45**	**19**	**34**	**16**
tr|?|Bacteria|Nitratifractor_salsuginis|E6WXT0	8	12	18	15	16	9
	8	12	19	12	14	13
tr|?|Eukaryota|Saprolegnia_parasitica|A0A067CB02	27	51	41	13	17	9
	27	48	42	13	11	14
tr|?|Eukaryota|Saprolegnia_parasitica|A0A067CA64 (W)	27	51	41	18	14	9
	27	54	42	15	12	13
tr|?|Eukaryota|Saprolegnia_diclina|T0QIU3	26	44	41	16	18	9
	23	44	44	14	12	13
tr|?|Eukaryota|Aphanomyces_astaci|W4GU91	5	39	41	21	14	9
	NO	47	39	18	12	13
***Pa*CDA**	**13**	**8**	**6**	**12**	**11**	**7**
tr|?|Eukaryota|Blumeria_graminis|N1JI69	31	8	7	20	8	7
	31	8	8	12	10	11
tr|?|Eukaryota|Metarhizium_majus|A0A0B4H3A4	41	8	6	16	12	7
	17	8	7	12	10	11
tr|?|Eukaryota|Ophiocordyceps_sinensis|T5AHU3	17	40	6	17	10	7
	15	40	7	13	8	7
tr|?|Eukaryota|Colletotrichum_incanum|A0A161WA77	17	41	6	16	12	7
	17	40	7	13	9	12
***Pgt*CDA**	**13**	**8**	**6**	**17**	**5**	**3**
tr|?|Eukaryota|Puccinia_sorghi|A0A0L6U7D9	13	8	14	13	9	3
	13	8	15	14	3	11
tr|?|Eukaryota|Melampsora_larici-populina|F4R898	13	8	6	10	12	3
	13	8	7	7	10	11
tr|?|Eukaryota|Tilletia_controversa|A0A177UF72	13	8	6	10	10	3
	13	8	7	8	6	11
tr|?|Eukaryota|Puccinia_sorghi|A0A0L6UEW6	33	8	12	12	8	3
	11	8	13	9	7	9
***An*CDA**	**8**	**8**	**6**	**18**	**6**	**7**
tr|?|Eukaryota|Verticillium_longisporum|A0A0G4KZQ2	11	8	6	15	7	42
	14	5	7	15	5	8

A phylogenetic tree was built to analyze how the new sequences clustered within the already characterized or annotated CDAs (Figure 5.4). All new sequences were grouped with their reference enzyme (highlighted in Figure 5.4) and, in general, were distributed throughout the tree.

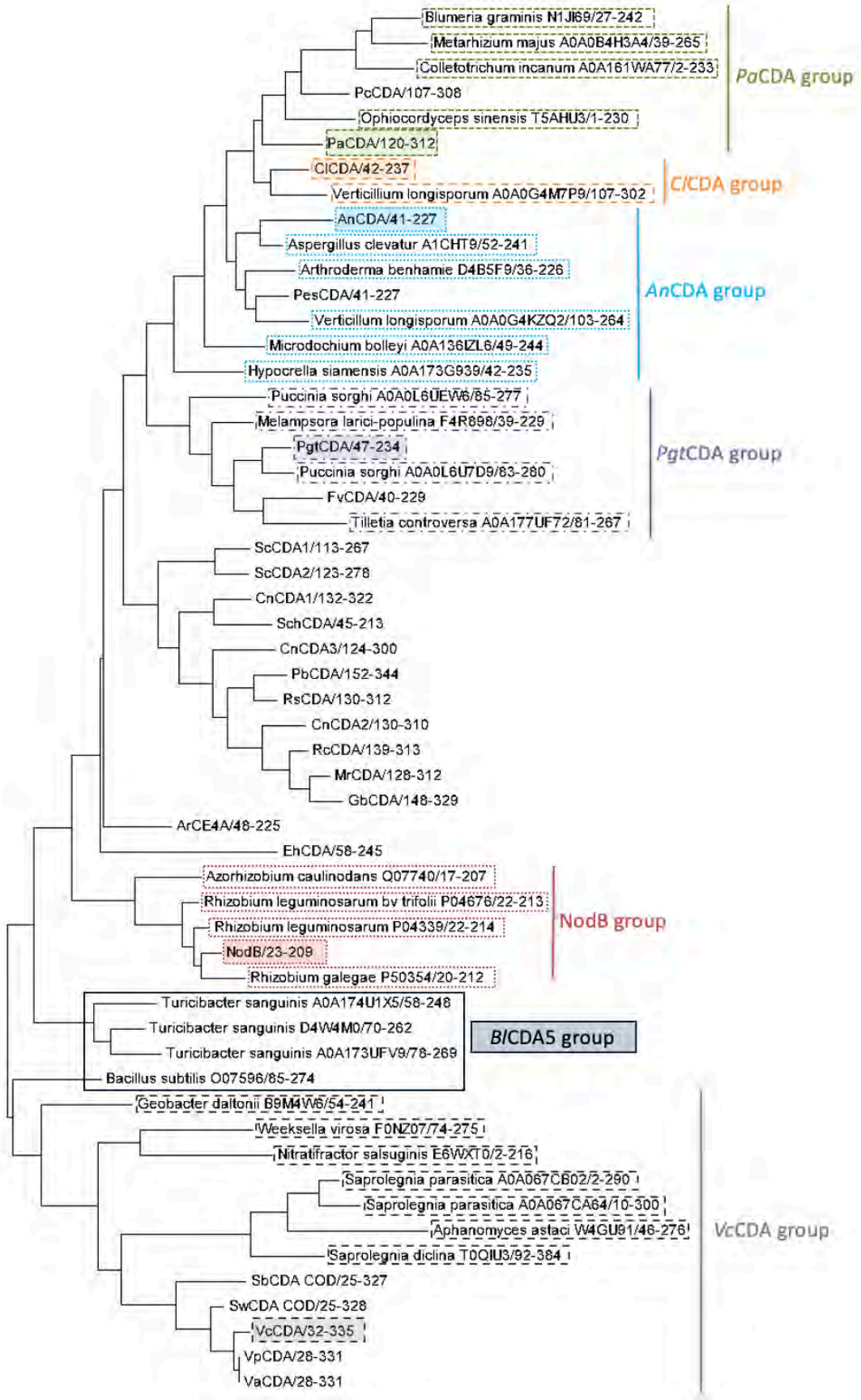


Figure 5.4. Phylogenetic tree with characterized CDAs and 29 selected sequences. *Colored filled boxes) reference enzymes from which groups new sequences were selected; Boxes without color) new sequences from the same group.*

*Pc*CDA, *Pa*CDA, *Cl*CDA and the sequences from their groups form a clade in the phylogenetic tree, showing to be highly related. The three characterized enzymes are extracellular fungal CDAs from the *Ascomycota* phylum and are involved in host infection.

Several sets of sequences with identical loop sizes and origin were detected. In order to select a specific sequence from each set, sequences from *Swiss-prot* database were prioritized. *Swiss-prot* is manually annotated and retrieved, providing a high level of annotation, a minimal level of redundancy and a high level of integration with other databases. Sequences included in this database, which are characterized or annotated, present higher possibilities to be actual CE4 deacetylases. In addition, the sequences were analyzed with the *Interpro* tool to identify their domains and analyze their modular architecture, prioritizing those sequences that contained, at least, one CE4 deacetylase domain and had a simple domain organization. Those in which the domains classification was clear and coincident between different databases and that presented less domains were selected.

Considering those parameters, the final selection included 5 sequences from three of the seven remaining groups (Table 5.4), each of them chosen for the following reasons:

Table 5.4. Final selected sequences for further characterization.

	SOURCE ORGANISM	L1	L2	L3	L4	L5	L6
	From *Vc*CDA group						
1	*Nitrifactor salsuginis* (E6WXT0)	8	12	19	12	14	13
2	*Saprolegnia parasitica* (A0A067CB02)	27	48	42	13	11	14
	From *Pa*CDA group						
3	*Blumeria graminis* (N1JI69)	31	8	8	12	10	11
4	*Ophiocordyceps sinensis* (T5AHU3)	15	40	7	13	8	7
	From *Pgt*CDA group						
5	*Tilletia controversa* (A0A177UF72)	13	8	7	8	6	11

From *Vc*CDA group

*Vc*CDA is a chitin oligosaccharide deacetylase from *Vibrio cholera* which accepts substrates from DP2 to DP6, presenting its highest activity on the shortest substrate [24], [81]. The enzyme specifically deacetylates the penultimate residue from the non-reducing end, generating mono-deacetylated products with the pattern [ADA_{n-2}] [24], [81], [118].

1. *Nitrifactor salsuginis* (E6WXT0)

None of the natural characterized CDAs renders products with just the reducing end deacetylated. According to the described Subsite Capping Model, loops 1, 2 and 6 are capping the binding cleft in negative subsites and loops 3, 4 and 5 in positive subsites. As an example, it has been previously shown the importance of length and flexibility of loop 5 in *Vc*CDA for substrate specificity [110]. The sequence from the bacterium *Nitrifactor salsuginis* was selected because it shows a short loop 1 and, although loop 5 is shorter than the one from *Vc*CDA, it is still longer than the ones from the other characterized enzymes. Having more space in the negative subsites could lead to the displacement of the oligosaccharide towards the binding mode needed for the deacetylation of the reducing end.

2. *Saprolegnia parasitica* (A0A067CB02)

This novel protein combines long loops 1 and 2, the sequences of which are longer than the ones of the reference enzyme. Both loops are located in the non-reducing end of the binding cleft, shaping that side of the active site of the enzyme.

From *Pa*CDA group

The enzyme from the filamentous ascomycete *Podospora anserina* is active on soluble glycol-chitin, chitosans with a high DA, and chitooligosaccharides. It fully deacetylates COS with a DP ≥ 2 and follows a multiple-chain mechanism, with faster deacetylation on longer substrates.

3. *Blumeria graminis* (N1JI69)

Blumeria graminis' sequence was selected for having the longest loop 1 from all the selected new enzymes. This loop is composed by 31 amino acid residues, almost the double of loop 1 from *Vc*CDA (the longest sequence from all the original reference enzymes).

4. *Ophiocordyceps sinensis* (T5AHU3)

In a similar way as the previous candidate, *Ophiocordyceps sinensis'* sequence was chosen because it shows one of the longest loops 2 among all sequences.

From *Pgt*CDA group

*Pgt*CDA from the fungal pathogen *Puccinia graminis* efficiently deacetylates colloidal chitin, glycol-chitin and chitosans, on which its activity increases with the degree of acetylation. With COS substrates, the minimal substrate is DP4. The enzyme acts following a multiple-chain mechanism

on DP4 to DP6 substrates and it specifically deacetylates all but the last two GlcNAc units on the non-reducing end [AA(D)$_{n-2}$] [130].

5. *Tilletia controversa* (A0A177UF72)

Contrary to the other cases, the criteria to select the set of novel sequences from the *Pgt*CDA group was not loop length. While the other sequences were chosen with the objective of finding new patterns of deacetylation, this last selection was guided by the aim of identifying other structural or sequence determinants of specificity.

Most CE4 enzymes share the conserved motif 4 (DXXD(W/Y) that shapes one side of the active site groove and includes a hydrophobic residue in loop 4 (generally, W or Y) that it is exposed to the solvent and establishes a stacking interaction with the sugar unit in the +1 subsite. As already explained, all the characterized enzymes with this residue either do not deacetylate the reducing end of COS or perform the deacetylation of this position at a much slower rate. On the contrary, the enzyme *Pgt*CDA lacks the +1 aromatic residue and deacetylates the reducing end of these substrates (Figure 5.5, A). Sequences from the *Puccinia gramini*'s group, including the selected one, do not have the aromatic amino acid in this position (Figure 5.5, C). The protein from *Tilletia controversa* was chosen with the objective of evaluating the possible function of the residue on the deacetylation pattern.

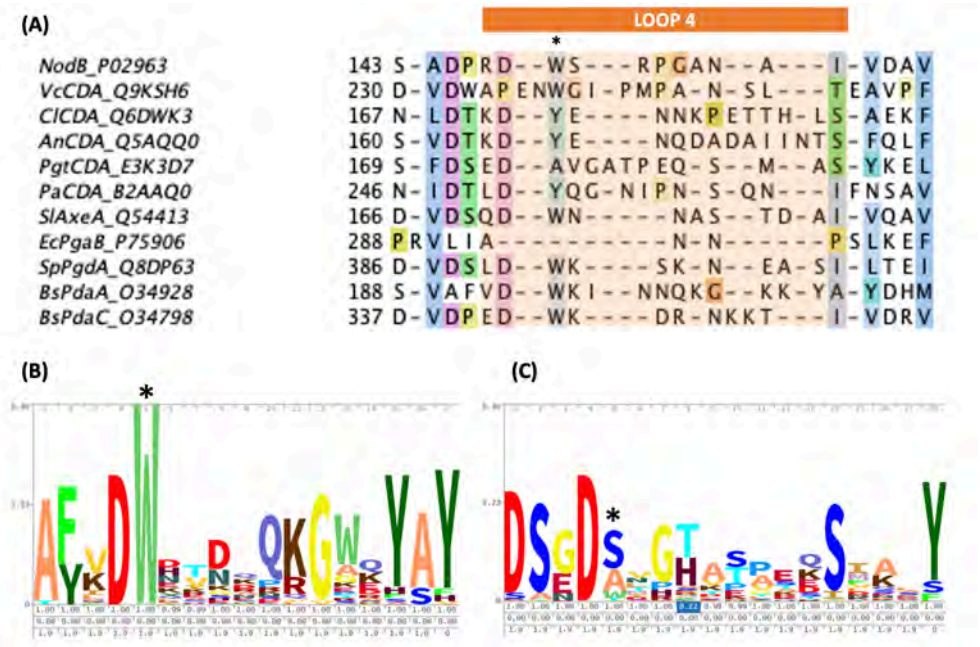


Figure 5.5. Differences in motif 4. *(A) Multiple sequence alignment with the initial reference sequences, manually curated to include loop 4/motif 4 region; (B) As example, HMM logo from Bacillus subtilis' group; (C) HMM from Puccinia gramini's group; *position with the conserved aromatic residue (W/Y) in most of CE4 enzymes.*

5.1.2. New peptidoglycan deacetylase from *Bacillus cereus*

The number of peptidoglycan deacetylases with classified activity and some characterization is very low (Table 5.5). Among them, there are only two annotated enzymes with MurNAc deacetylase activity (PdaA from *Bacillus subtilis,* which has been characterized, and Pda from *Bacillus anthracis,* with solved x-ray structure but not biochemical characterization), being the others peptidoglycan GlcNAc deacetylases.

Table 5.5. Peptidoglycan deacetylases with reported 3D structure or reported activity on peptidoglycan.

SUBTYPE	ENZYME	ORGANISM	ACTIVITY ON PGN
GlcNAc deacetylase	*Sp*PgdA	*Streptococcus pneumoniae*	GlcNAc DA on PGN
	*Sm*PgdA	*Streptococcus mutants*	GlcNAc DA on PGN
	*Bc*Pgd (BC1960)	*Bacillus cereus*	GlcNAc DA on PGN
	*Er*Pgd	*Eubacterium rectale*	GlcNAc DA (annotated)
MurNAc deacetylase	*Bs*PdaA	*Bacillus subtilis*	MurNAc DA on PGN
	*Ba*Pda (BA0424)	*Bacillus anthracis*	MurNAc DA (annotated)

As already described, there are 5 characteristic motifs that are generally conserved in CE4 family, but there are certain differences that could represent subfamilies of enzymes. Motif 1 (TFDDG), which contains the catalytic base and the aspartate from the metal coordination triad, is highly conserved in all CE4 enzymes, except for the two mentioned MurNAc deacetylases. Those enzymes present the insertion of an asparagine residue after the catalytic aspartate and they are missing the metal-binding aspartate (TFDN). This characteristic is not shared by the known GlcNAc deacetylases and it was thought to be a distinctive feature to identify one enzyme's activity from its sequence and distinguish MurNAc deacetylases. However, the enzyme PdaC from *Bacillus subtilis*, the characterization of which has been described on the previous chapter of this work, is a peptidoglycan deacetylase with reported activity on MurNAc residues but with the conserved metal-binding triad. The specificity of this enzyme indicates that the asparagine insertion is not a requirement for an enzyme to be a MurNAc deacetylase. Furthermore, it has been determined that its closest homolog is the GlcNAc deacetylase PgdA from *Streptococcus pneumoniae*, which only deacetylates *N*-acetyl-glucosamine residues. Another differential characteristic of *Bs*PdaC with respect canonical MurNAc deacetylases is that the enzymes is active on intact peptidoglycan, while *Bs*PdaA is only active in peptidoglycan without the peptidic substitutions (glycan chain without the stem peptide). It could be possible that the asparagine insertion is related to this distinction

between different MurNAc deacetylases rather than being a general feature of peptidoglycan MurNAc deacetylases.

In this context, it is clear the need of discovering and characterizing new peptidoglycan deacetylases in order to increase the available information on this subfamily of enzymes and identify the possible determinants of their different specificities: pure GlcNAc deacetylases (such as *Sp*PgdA), pure MurNAc deacetylases (such as *Bs*PdaA) and MurNAc deacetylases with activity also on GlcNAc residues (such as *Bs*PdaC, with dual activity, MurNAc on PGN and GlcNAc on COS).

5.1.2.1. Data base search for new sequences

With the objective of identifying possible characteristic features of each sub-activity of peptidoglycan deacetylases (on MurNAC or GlcNAc residues), new searches in databases were performed to generate a larger set of sequences corresponding to possible peptidoglycan deacetylases.

Different searches in databases were performed using the BLAST (Basic Local Alignment Search Tool) algorithm. BLAST is a sequence similarity searching tool that enables the identification of sequences that resemble a query above a certain threshold from large sequence databanks. It is often used as the first step in analysis of new proteins, performing and scoring a sequence alignment to find similar protein sequences from a large database for a given query sequence. The search maximizes a score that quantifies similarity between sequences given a particular scoring matrix and gap penalty.
Psi-BLAST (Position-Specific Iterated BLAST) aims to find more distant relatives (or a wider range of similar sequences) by performing the BLAST search iteratively. After the first BLAST-P (protein-protein blast) iteration, it produces a PSSM (position specific scoring matrix) from the resulting alignment, generating a larger and more diverse group of similar proteins. In a second step, the alignment is run again and the PSSM is recomputed. These two steps can be repeated multiple times, being this method more sensitive because the matrix is based on the multiple alignment that is specific for the query protein.

Four initial BLAST-P searches were performed, all of them starting with *Sp*PgdA and *Bs*PdaA as reference enzymes for GlcNAc deacetylases and MurNAc deacetylases, respectively. Each one with changes in some parameters with respect the previous one (Table 5.6).

- The first search was performed against the NR (non-redundant protein sequence) database. Selecting this database, the search includes all non-redundant GenBank CDS translations, sequences from PDB, Swiss-prot, PIR and PRE, excluding environmental samples from WGS (whole genome shotgun) projects. In both cases (with *Sp*PgdA and *Bs*PdaA as a query), the resulting list of 100 sequences included the query sequence, identic proteins to it or sequences with negligible variations.
- In the second search, the "RefSeq" database was chosen, which is a well-annotated and manually curated database with less redundancy than the previous one. In addition, the number of output sequences was extended to 500 with the aim of increasing the variability of the resulting sequences. Regardless those changes, lists corresponding to the query sequences were obtained again, mainly because all sequences with the "WP_" accession prefix were included. Since they come from genome sequencing reports, several accession numbers can exist for the same sequence or protein.
- In order to avoid this redundancy, only proteins with the "NP_" accession prefix (from which there is only one for protein) were included in the third search. In this case, it was possible to obtain a set of different related proteins that aligned on the catalytic domain region of the query sequences. Although sequences different from the queries were obtained, all types of CE4 enzymes were included, not just peptidoglycan deacetylases but also chitin deacetylases and other esterases. Furthermore, within the group of peptidoglycan deacetylases, there was no separation between the ones acting on GlcNAc residues and the ones acting on MurNAc residues (specific MurNAc deacetylases appeared on *Sp*PgdA output list and specific GlcNAc deacetylases were included on the one from *Bs*PdaA).
- Trying to obtain a division between the subfamilies, the most restrictive matrices offered by the BLAST-P tool were selected for the last approximation and searches with a maximum of 100 and 500 sequences were performed. Again, different members of the family appeared in both lists and, between them, MurNAc deacetylases and GlcNAc deacetylases without distinction. Furthermore, it was not possible to discard sequences from the obtained lists by means of e-value because no significative change in that parameter was observed in any of the cases.

Table 5.6. Parameters used on the BLAST-P searches performed.

BLAST-P SEARCH	QUERY SEQUENCE	DATABASE	NUM. SEQ.	MATRIX
1	*Sp*PgdA/*Bs*PdaA	NR	100	BLOSUM 62
2	*Sp*PgdA/*Bs*PdaA	RefSeq	500	BLOSUM 62
3	*Sp*PgdA/*Bs*PdaA	RefSeq (just NP_)	100	BLOSUM 62
4	*Sp*PgdA/*Bs*PdaA	RefSeq (just NP_)	100/500	BLOSUM 90/ PAM 30

These first BLAST-P searches results indicate that the presence of the highly conserved motifs, that are shared between all CE4 family members, does not allow to establish a clear separation between the different subtypes of enzymes.

Knowing this limitation, the toll psi-blast (which allow for the search of more distant related sequences) was used for the second round of searches. With this approximation, after a first standard BLAST-P result, several iterations are performed. In each one, it is possible to select which sequences from the previous result are included in the specific matrix. These new searches were performed again with *Sp*PgdA and *Bs*PdaC as query and with the initial conditions presented in Table 5.7.

Table 5.7. Parameters used on the initial BLAST-P search with the psiblast search.

QUERY SEQUENCE	DATABASE	NUM. SEQ.	MATRIX
*Sp*PgdA/*Bs*PdaA	RefSeq	500	BLOSUM 90

From the result of the initial BLAST-P, three different selection criteria were followed to continue with the different iterations:

1. **Psi-blast with low restriction**: choosing all sequences obtained in the initial list except the ones that did not contain the TFDDG/TFND motif, the catalytic histidine or the metal-coordinating histidine in their sequence and the ones with another activity (different to deacetylation) annotated.
2. **Psi-blast restricted to CE4 enzymes**: choosing only the sequences corresponding to characterized CE4 enzymes (with known structure or described activity), also checking that all sequences contained the mentioned residues.
3. **Psi-blast restricted to MurNAc/GlcNAc deacetylases**: choosing only the sequences of characterized enzymes that share activity with the query (MurNAc peptidoglycan deacetylases or GlcNAc peptidoglycan deacetylases).

There was not a change in e-value that allowed to establish a threshold in any of the cases. In consequence, new iterations were performed until the number of output sequences did not increase. Following the described procedure, several sets of sequences to be analyzed were obtained, all of them very similar and sharing most of the sequences. For this reason, the group with a larger number of sequences (210 in total), from the psi-blast search with low restriction with *Sp*PgdA as query, was the one chosen to continue with the analysis.

5.1.2.2. Identification of distinctive motifs

With the aim of identifying possible differences between enzymes with the two subtypes of activities, both the alignment of the set of 210 sequences obtained from the databases searches and the structures of *Sp*PgdA and *Bs*PdaA were studied.

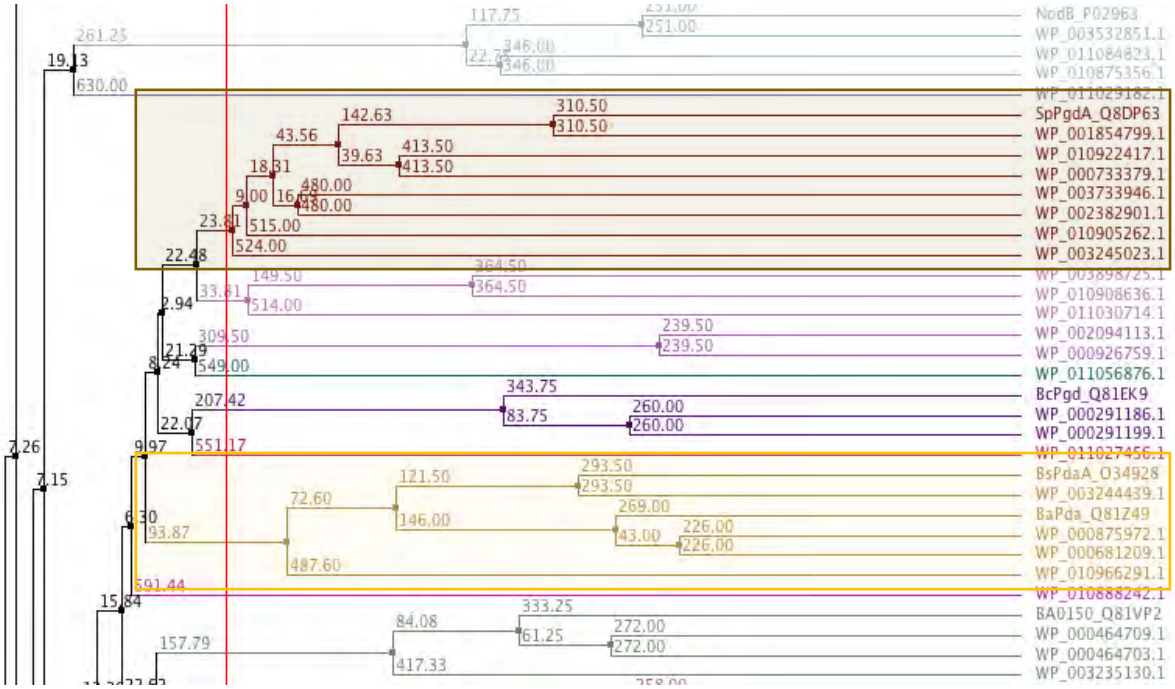


Figure 5.6. Phylogenetic tree with 210 sequences of possible peptidoglycan deacetylases retrieved from databases by BLAST-P and psiblast searches. *Brown) cluster of sequences grouped with the reference GlcNAc peptidoglycan deacetylase SpPgdA; Yellow) cluster of sequences grouped with the reference MurNAc peptidoglycan deacetylase BsPdaA. The complete phylogenetic tree can be found on Annex 9.5.*

A multiple sequence alignment was performed against HMM-4, profile of CE4 enzymes with solved x-ray structure or characterized activity on COS (following the same procedure than in Section 4.1, including the sequences from the alignment on Figure 4.2). From the obtained alignment, a phylogenetic tree was built in order to study how the different sequences clustered and to generate two groups that would represent GlcNAc and MurNAc deacetylases separately. The phylogenetic study showed again the high similarity that all sequences share, being the two reference enzymes

(*Sp*PgdA and *Bs*PdaA) in not so distant positions in the tree (Figure 5.6). It is worth noticing that both classified MurNAc peptidoglycan deacetylases (*Bs*PdaA and *Ba*Pda) appeared in the same cluster but, on the contrary, GlcNAc peptidoglycan deacetylases were distributed throughout the tree. The number of sequences that could be grouped with each reference enzyme is limited (Table 5.8).

The information on annotation and origin of the different sequences was retrieved from their accession codes and it was observed that *Bs*PdaC, the studied enzyme in the previous chapter of this work, appeared inside the GlcNAc deacetylases group together with *Sp*PgdA, although the rest of characterized GlcNAc deacetylases were scattered on the tree. Only two new sequences could be included in *Bs*PdaA group, from which only the one from *Bacillus cereus* shared the asparagine insertion on motif 1 with the two known MurNAc deacetylases.

Table 5.8. Sequences grouped with SpPgdA and BsPdaA on the phylogenetic analysis.

ACCESSION CODE	SOURCE ORGANISM	ANNOTATED ACTIVITY
	***Sp*PgdA**	
*Sp*PgdA	*Streptococcus pneumoniae*	GlcNAc peptidoglycan deacetylase
WP_010922417.1	*Streptococcus pyrogenes*	Deacetylase
WP_000733379.1	*Streptococcus*	Polysaccharide deacetylase
WP_003733946.1	*Listeria monocytogenes*	Endo-1,4-beta-xylanase
WP_002382901.1	*Enterococcus faecalis*	Polysaccharide deacetylase
WP_010905262.1	*Lactococcus lactis*	Endo-1,4-beta-xylanase
WP_003245023.1	*Bacillus subtilis*	*Bs*PdaC (MurNA PGN DA)
	***Bs*PdaA**	
*Bs*PdaA	*Bacillus subtilis*	MurNAc peptidoglycan deacetylase
WP_000875972.1	*Bacillus cereus*	Delta-lactam-biosynthetic de-*N*-acetylase
WP_010966291.1	*Clostridium acetobutylicum*	Delta-lactam-biosynthetic de-*N*-acetylase
*Ba*PdA	*Bacillus anthracis*	MurNAc peptidoglycan deacetylase

These two groups and the crystal structure of the reference enzymes were compared in order to determine possible distinctive patterns, features or residues that could be characteristic for each subtype besides the asparagine insertion previously identified in MurNAc deacetylases. Four distinctive motifs between the two groups were identified and are described below (Figure 5.7 and Tables 5.9-5.12)

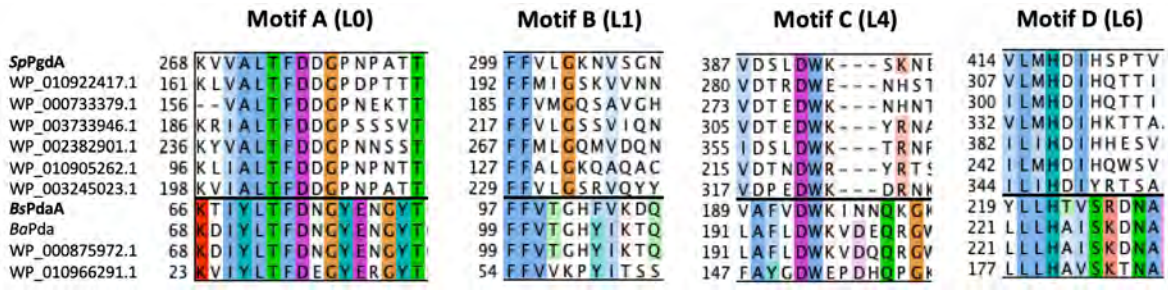


Figure 5.7. Distinctive motives in sequences from SpPgdA and BsPdaA groups. *Upper group) sequences grouped with the reference GlcNAc peptidoglycan deacetylase SpPgdA on the phylogenetic tree; lower group) sequences grouped with the reference MurNAc peptidoglycan deacetylase BsPdaA on the phylogenetic tree.*

- In "loop 0"

The clearest distinctive motif was identified in a sequence location previous to loop 1. On MurNAc deacetylases, it starts with the asparagine next to the catalytic aspartate, it includes a tyrosine residue and it forms a loop (here named loop 0) that is located between loop 1 and loop 6 in the structures. A clear structural difference can be observed between *Bs*PdaA and *Sp*PgdA: on the MurNAc deacetylase, loop 0 is located closer to the catalytic center than on the GlcNAc deacetylase, in which it adopts a more hidden location between L1 and L6.

Table 5.9. Difference in sequences between GlcNAc and MurNAc peptidoglycan deacetylases groups in loop 0.

MOTIF A	
On *Bs*PdaA group	**N**GYENG**Y**
On *Sp*PgdA group	DGPNPXT

- In loop 1

The presence of a phenylalanine/tyrosine close to the previously described loop 0 and to the catalytic center was identified on the group of MurNAc deacetylases. On the contrary, this residue is missing on the GlcNAc deacetylases group (Figure 5.7).

Table 5.10. Difference in sequences between GlcNAc and MurNAc peptidoglycan deacetylases groups in loop 1.

MOTIF B	
On *Bs*PdaA group	GH(**F/Y**)I
On *Sp*PgdA group	GSSV

- In loop 4

Immediately after motif 4, a difference in loop 4, which is longer and includes a glutamine residue close to the active site on the MurNAc deacetylases group, can be observed (Figure 5.7). However, it is known from *Vc*CDA structure that L4 is a very flexible and mobile loop and no clear interaction between the distinctive residues and the catalytic center or the substrate can be identified.

Table 5.11. Difference in sequences between GlcNAc and MurNAc peptidoglycan deacetylases groups in loop 4.

	MOTIF C
On *Bs*PdaA group	WK(I/V)DX**Q**
On *Sp*PgdA group	WK-XX

- In loop 6

Considering its location between L0 and L4, it was though possible for loop 6 to be involved in the interaction between amino acids of these two loops. When analyzing its sequence in the two groups, some differences could be observed (Figure 5.7), but the position of these residues on the crystal structure did not show a possible involvement in establishing interactions between the mentioned loops.

Table 5.12. Difference in sequences between GlcNAc and MurNAc peptidoglycan deacetylases groups in loop 6.

	MOTIF D
On *Bs*PdaA group	H(T/A)VS(R/K)
On *Sp*PgdA group	HDIH(S/Q)

The residues from these distinctive motifs that are located within a short distance to the active site are highlighted in Figure 5.8, being the inserted asparagine previously identified and the tyrosine from L0 the main changes in the active center that could participate differently in the catalytic mechanism. However, as already demonstrated by the initial searches and the phylogenetic study, the high similarity shared by the catalytic domain of all CE4 enzymes makes it very difficult to establish a sequence signature for the different subtype of activities.

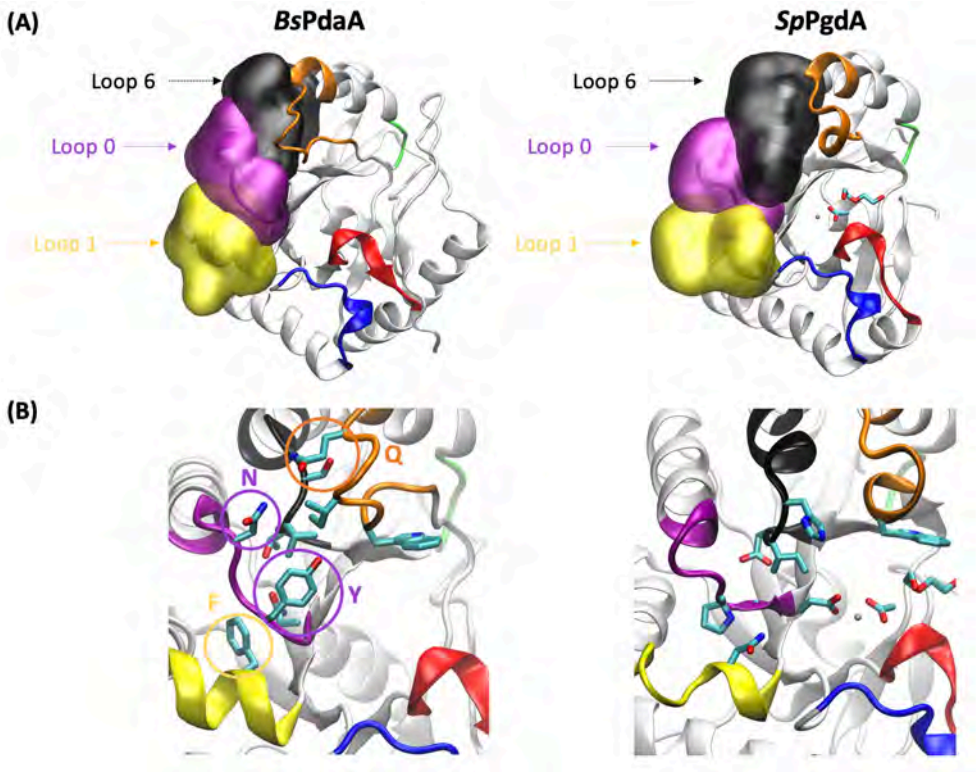


Figure 5.8. Differences in 3D structures of the reference MurNAc peptioglycan deacetylase (BsPdaA) and GlcNAc peptidoglycan deacetylase (SpPgdA). *(A) Surface representation of the space occupied by loops "0" (purple), 1 (yellow) and 6 (black); (B) main distinctive residues in loops "0" (purple, asparagine and tyrosine from motif A), 1 (yellow, phenylalanine from motif B) and 4 (orange, glutamine from motif C).*

5.1.2.3. Peptidoglycan deacetylases on *Bacillus* species and final selection

Bacillus are Gram-positive bacteria and there are three species that present several genes predicted as peptidoglycan deacetylases in their genomes. *Bacillus subtilis* is found in soil and in the gastrointestinal tract of some animals and it has the ability to form spores to survive harsh conditions, *Bacillus cereus* is found on soil and food and it can cause diarrhea in humans and *Bacillus anthracis* is also a spore forming bacteria and is the causative agent of anthrax.

Bacillus subtilis has six identified or predicted polysaccharide deacetylases, three of which have been characterized: PdaA [131] and PdaB (ybaN) [132] are MurNAc deacetylases involved in sporulation. The disruption of PdaA gene lead to a complete lack of spore germination and mutant spores deficient in muramic-δ-lactam and pdaB-deficient spores showed anomalies during the maturation stage and were almost empty or cortex-less, indicating that PdaB is essential to maintain spores after the late stage of sporulation. *Bs*PdaC is also a MurNAc deacetylase but not associated with sporulation and germination, being lysozyme sensitivity the only phenotypic difference showed by knockout PdaC strains [121].

The other three genes (ylxY, yheN and yxKH) do not have an assigned function yet since disruption of these genes did not show any phenotypic difference (including vegetative cell morphology, sporulation and germination, or lysozyme sensitivity) and no further characterization has been performed [22], [72].

Table 5.13. Putative polysaccharide deacetylases from Bacillus cereus and Bacillus anthracis. *From* [133]**.**

B. cereus **ATCC14579**	***B. anthracis*** **str. 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_830200 (360) (BC0361)	NP_842877 (360) (BA0330)	Polysaccharide deacetylase	87	91
NP_830050 (254) (BC0171)	NP_842717 (254) (BA0150)	Chitooligosaccharide deacetylase	95	99
NP_830306 (260) (BC0467)	NP_842967 (260) (BA0424)	Peptidoglycan MurNAc deacetylase	98	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

Both *Bacillus cereus* and *Bacillus anthracis* contain ten ORFs for putative polysaccharide deacetylases, being orthologues with more than 90% identity (Table 5.13). Five were classified as putative peptidoglycan deacetylases and three have been confirmed as peptidoglycan deacetylases [133]. Five enzymes from *B. cereus* were recombinantly expressed and biochemically characterized [133], [134], all of them showing to be active on a wide range of different substrates such as PGN from Gram-positive (*B. cereus*) and Gram-negative (*H. pylori*) bacteria, chitin oligomers with DP between 4 and 6, the synthetic muropeptide GMDP (*N*-acetyl-*D*-glucosaminyl-β-(1,4)-N-acetylmuramyl-*L*-alanyl-*D*-isoglutamine), and glycol chitin. BC2929 and BC5204 were active on lipid II (pentapeptide-substituted MurNAc-GlcNAc with a bactoprenol-pyrophosphate anchor) and the later showed to be also active on UDP-GlcNAc. The biological functions of their homologs from *B. anthracis* was assessed by genetic analysis, generating knockouts. BA1977, BA1961 and BA3679 are

involved in peptidoglycan metabolism while BA5436 and BA2944 are involved in the synthesis of the neutral polysaccharide of *B. anthracis* [133]. A summary of the evidence and characterization level of each enzyme is presented in Table 5.14. Although all of them have been annotated as possible deacetylases from CE4 family, only a few have been characterized.

Table 5.14. Characterization of putative polysaccharide deacetylases from Bacillus cereus and Bacillus anthracis.

ENZYME	KNOCK OUT	EXPRESSION AND PURIFICATION	ENZYMATIC CHARACTERIZATION	X-RAY STRUCTURE
BC1960		Yes	Yes	-
BA1961	Yes			-
BC3618		Yes	Yes	-
BA3679	Yes			-
BC2929		Yes	Yes	-
BA2944	Yes			-
BC5204		Yes	Yes	-
BA5436	Yes			-
BC1974		Yes	Yes	5NCD
BA1977	Yes			-
BC0361	Possible function annotated by sequence similarity			4HD5
BA0330	Possible function annotated by sequence similarity			4V33
BC0171	Possible function annotated by sequence similarity			-
BA0150	Possible function annotated by sequence similarity			4M1B
BC0467/BA0424	Possible function annotated by sequence similarity			-
BC1768/BA1836	Possible function annotated by sequence similarity			-
BC3804/BA3943	Possible function annotated by sequence similarity			-

When studying the phylogenetic analysis of all these possible peptidoglycan deacetylases on *Bacillus*, the only new sequence showing the asparagine insertion in motif 1 was BC0467 from *Bacillus cereus*, which corresponds with the sequence retrieved from the database searches performed: WP_000875972.1. Considering the available information, this was the chosen sequence for its expression as a recombinant enzyme and further characterization.

In comparison with the known characterized MurNAc deacetylases, the protein sequence shares a 53.7% of identity with *Bs*PdaA and a 98% of identity with *Ba*Pda, presenting the same residues as the last one in all the motifs described. On the contrary, the identity percentage of BC0467 with

*Bs*PdaC is only 33,5%. The characteristic of this new sequence suggest that it is a canonical PGN MurNAc deacetylase.

5.2. Chitin deacetylases

5.2.1. Subcloning into pET22b(+) vector

Once all the new genes were selected, a strategy for their subcloning into a suited expression vector for their later expression as recombinant enzymes was designed. The chosen vector was the previously used pET22b(+), in which the introduced gene would be under T7 promotor control and that contains the codifying sequence for a histidine-tag located at the C-terminus of the protein for its later purification. In addition, it was decided to introduce the sequence corresponding to the restriction site of TeV protease between the His-tag sequence and the enzyme sequence to have the possibility to remove the tag after purification if desired.

The modular structure of the selected new proteins was studied (Table 5.15) and most of them presented a signal peptide for cellular location that was removed from the sequence in the final design. The final synthetic genes encoding for the new enzymes that included the TeV restriction sequence were produced codon-optimized for its expression in *E. coli*. They were supplied in pMK-QR vector and they were subcloned into the pET22b(+) plasmid between NdeI and XhoI restriction sites to generate the C-terminal his-tagged proteins. After Sanger sequencing, competent *E. coli* DH5α and *E. coli* BL21(DE3) STAR cells were transformed with the obtained plasmids.

Table 5.15. Modular organization of the selected sequences. *SP: signal peptide; CB: carbohydrate binding domain; CE4: CE4 catalytic domain*

ENZYME	SOURCE ORGANISM	DOMAINS
*Ns*CDA	*Nitratifactor salsuginis*	SP + CE4
*Sp*CDA	*Saprolegnia parasitica*	-
*Bg*CDA	*Blumeria graminis*	SP + CB + CE4 + CB
*Os*CDA	*Ophiocordyceps sinensis*	CE4
*Tc*CDA	*Tilletia controversa*	SP + CE4

5.2.2. Expression and purification of recombinant enzymes

Most of the new enzymes have a fungal origin, which could difficult their expression in a simpler organism like *E. coli*. As an initial approximation, several expression conditions based on the two methodologies previously described (growing in autoinduction medium and induction by IPTG) were tested for all of them (Figure 5.9). The expression of all proteins with the expected molecular weight was obtained except for *Os*CDA, which was not expressed under any condition. However, the only enzymes that showed some expression on the soluble fraction were *Ns*CDA and *Tc*CDA, being *Sp*CDA and *Bg*CDA only expressed on the insoluble fraction after disruption.

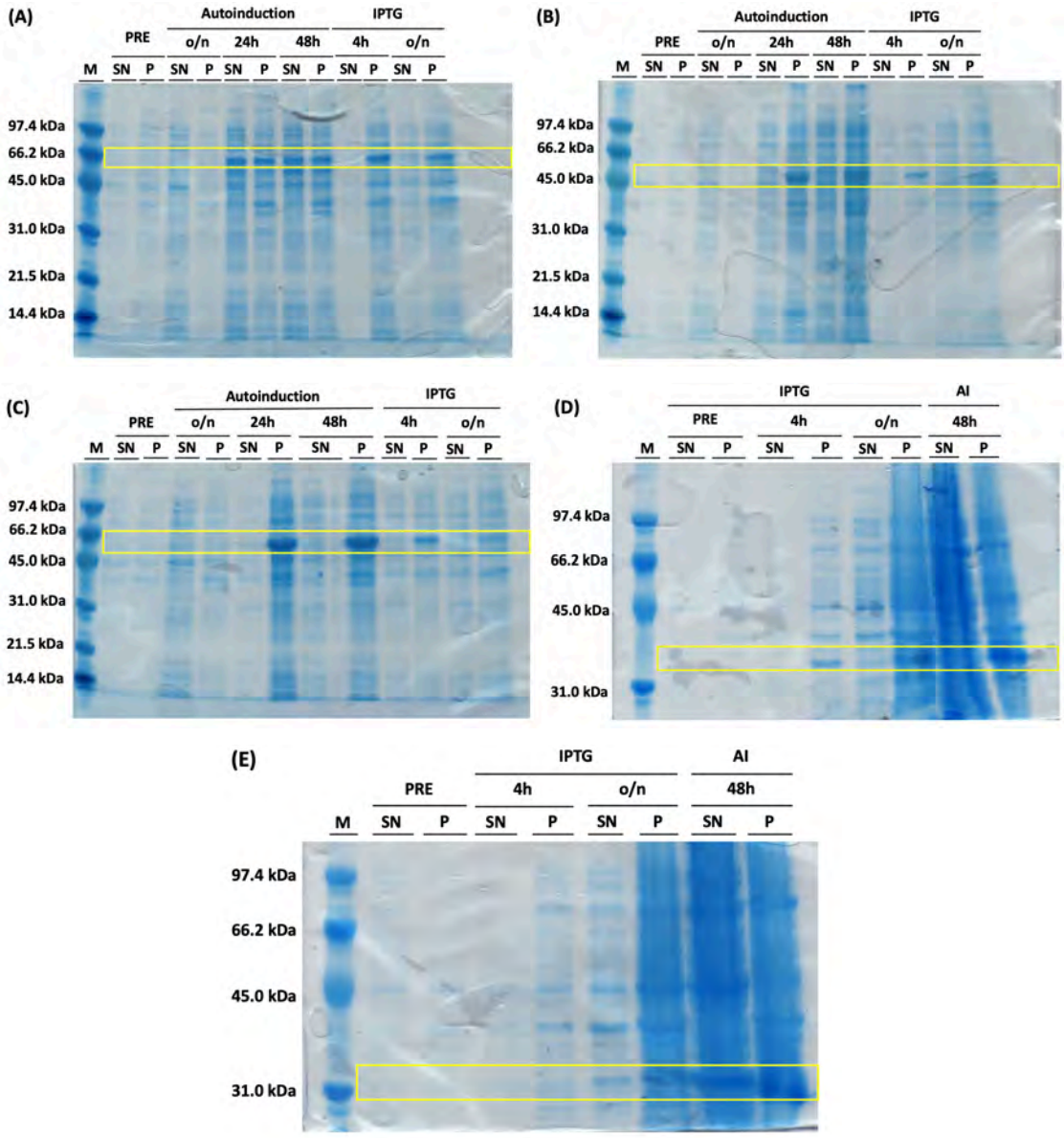


Figure 5.9. SDS-PAGE 14% acrylamide of expression trials of new CDAs. *(A) NsCDA, expected Mw: 58.4 kDa; (B) SpCDA, expected Mw: 48 kDa; (C) BgCDA, expected Mw: 48.3 kDa; (D) OsCDA, expected Mw: 28.4 kDa; (E) TcCDA, expected Mw: 32.5 kDa. PRE) before induction; AI) Autoinduction; M) Molecular weight marker; SN: supernatant; P: pellet.*

It was decided to proceed with the two enzymes from which some soluble protein was recovered. However, the amount of these soluble fractions, when the expression volumes were scaled up, was very low and optimization of the expression conditions was necessary.

5.2.2.1. Expression and purification of *Ns*CDA

Several variations in the expression conditions of the already used methods were tested in order to improve the soluble expression of *Ns*CDA.

- Induction with IPTG

Although different growth conditions (cell density, temperature, incubation time, agitation) and inductor concentration were studied (Table 5.16), the expression of the protein of interest was always observed on the insoluble fraction after cell lysis (Figure 5.10).

Table 5.16. Conditions used for improving NsCDA expression with IPTG induction method. *For experiences 1 to 4, samples at two incubation times after induction were analyzed (4 hours and overnight).*

EXPERIENCE	GROWTH (OD, 600 nm)	[IPTG] (mM)	SOLVENT	INDUCTION Tª (°C)	AGITATION (RPM)	INCUBATION TIME
1	0.845	0.5		37	250	4 h / overnight
2	0.815	0.5		25	170	4 h / overnight
3	0.963	0.2		37	250	4 h / overnight
4	Overnight (saturated)	0.5		37	250	4 h / overnight
5	0.957	0.02	2% EtOH	16	170	20 h

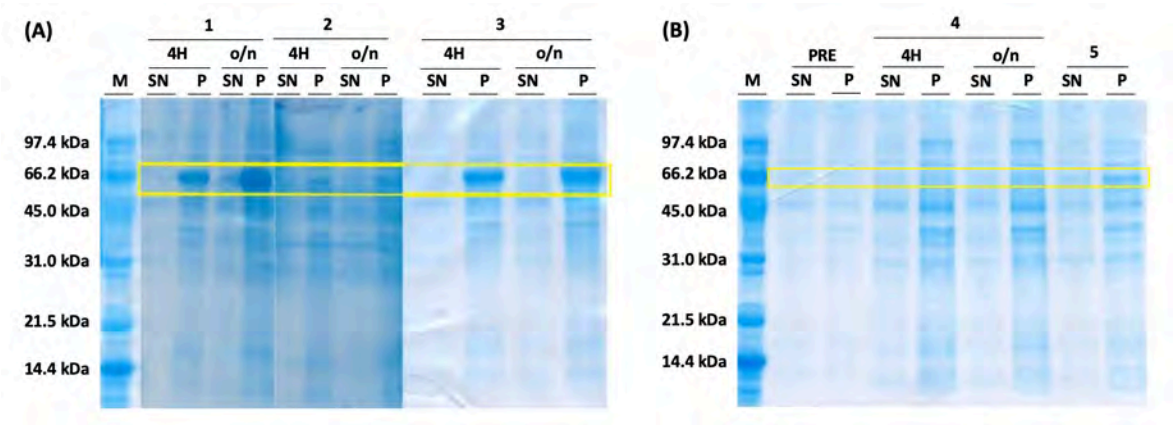


Figure 5.10. SDS-PAGE 14% acrylamide of expression trials of NsCDA. *Expected Mw: 58.4 kDa (yellow square). (A) Experiences 1 to 3 (corresponding to Table 5.16 conditions); (B) Experiences 4 and 5 (corresponding to Table 5.16 conditions). M) Molecular weight marker; PRE) before induction; SN) supernatant; P) pellet; o/n) overnight.*

- Growth in autoinduction medium

Some soluble protein was recovered after cell lysis when expressing *Ns*CDA by growing transformed cells with autoinduction medium at 25°C and 180 rpm for 48 hours.

Purification of the enzyme from a larger expression culture was performed by affinity chromatography using a His-trap column (Figure 5.11), with a final overall yield of 0.2 mg of pure protein per liter of culture. However, an important amount of the expressed protein was eluted in the flow through fraction (not retained in the column) and the imidazole-eluted protein contained additional bands of lower molecular weight (Figure 5.12).

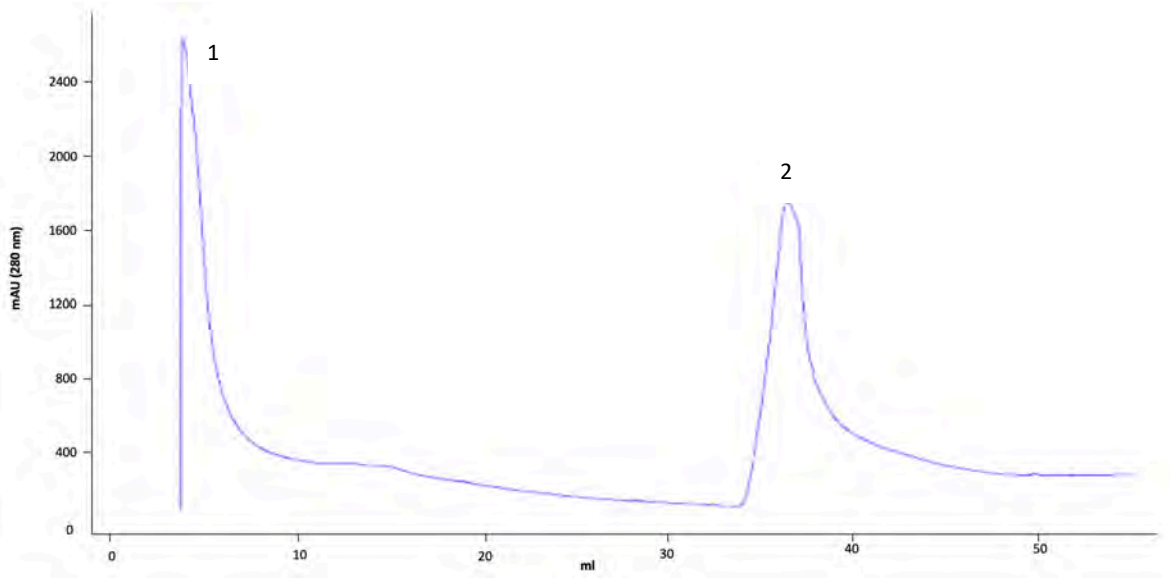


Figure 5.11. Chromatogram of His-trap purification of NsCDA*. Resulting chromatogram from affinity chromatography. 1. Washing with PBS, 2. Elution with PBS + Imidazole.*

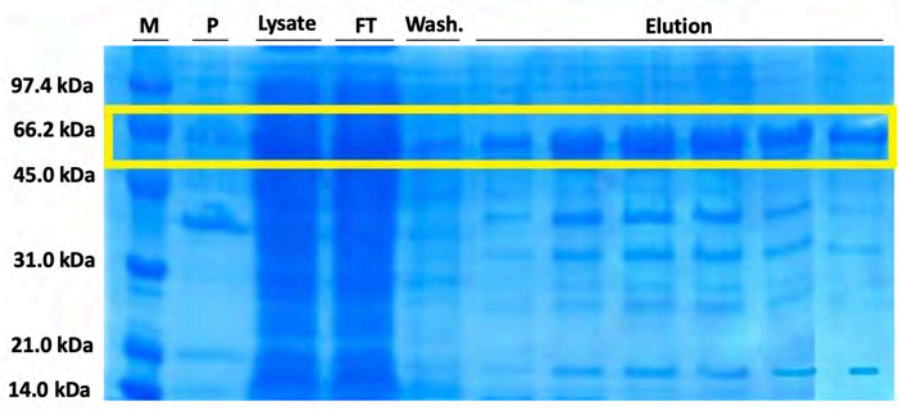


Figure 5.12. SDS-PAGE 14% acrylamide of purification of NsCDA. *M) Molecular weight marker; P) pellet; FT) Flow Through; Wash.) Washing with PBS.*

Since the protein is not pure enough for characterization, further work should investigate improved protocols to both increase the expression yield of the protein and optimize the purification procedure in order to bind all the expressed protein in the affinity chromatography column and include an additional purification step to improve the final purity.

5.2.2.2. Expression and purification of *Tc*CDA

Variations on conditions of the already used methods (IPTG induction and growth on autoinduction medium) as well as new strategies (co-expression with chaperones and use of detergent) were attempted to improve expression of *Tc*CDA.

- Induction with IPTG and growth in autoinduction medium

The same conditions previously tested for *Ns*CDA expression using IPTG as inductor were tested for *Tc*CDA (Table 5.17) obtaining the same results, with either no expression of the target protein or expression only in the insoluble fraction (Figure 5.13). In the same way, when expressed by growing cells in an autoinduction medium as previously described (25°C, 180 rpm), *Tc*CDA was only observed on the insoluble fraction after cell lysis (Figure 5.14).

Table 5.17. Conditions used for improving TcCDA expression with IPTG induction and autoinduction methods. *For experiences 1 to 4, samples at two incubation times after induction were analyzed (4 hours and overnight).*

EXPERIENCE	GROWTH (OD, 600 nm)	[IPTG] (mM)	SOLVENT	INDUCTION Tª (°C)	AGITATION (rpm)	INCUBATION TIME
1	0.625	0.5		37	250	4 h / overnight
2	0.617	0.5		25	170	4 h / overnight
3	0.627	0.2		37	250	4 h / overnight
4	Overnight (saturated)	0.5		37	210	4 h / overnight
5	0.48	0.02	2% EtOH	16	170	20 h

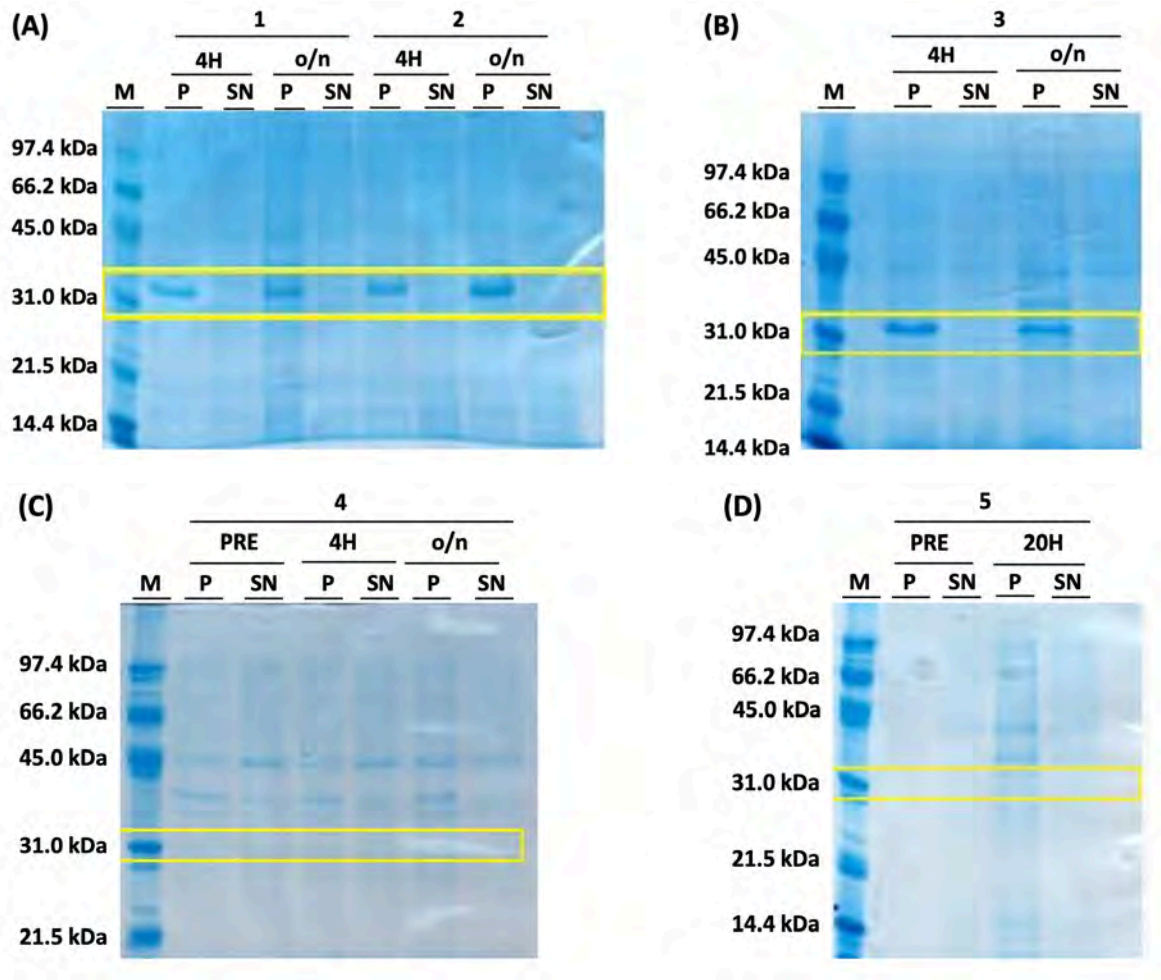


Figure 5.13. SDS-PAGE 14% acrylamide of expression trials of TcCDA. *Expected Mw: 32.5 kDa (yellow square). (A) Experiences 1 and 2 (corresponding to Table 5.17 conditions); (B) Experience 3 (corresponding to Table 5.17 conditions); (C) Experience 4 (corresponding to Table 5.17 conditions); (D) Experience 5 (corresponding to Table 5.17 conditions). M) Molecular weight marker; PRE) before induction; SN) supernatant; P) pellet; o/n) overnight.*

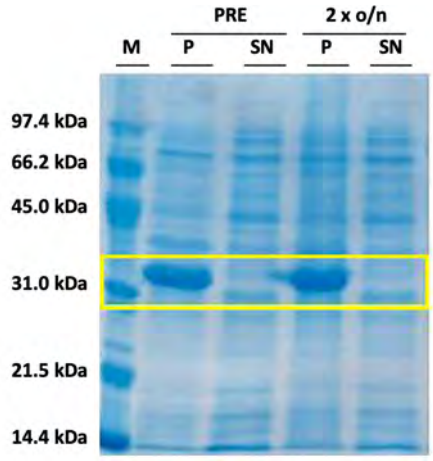


Figure 5.14. SDS-PAGE 14% acrylamide of expression trials of TcCDA. *Expected Mw: 32.5 kDa (yellow square). Expression with autoinduction medium at 25°C and 170 rpm. M) Molecular weight marker; PRE) before induction; SN) supernatant; P) pellet; o/n) overnight.*

- Co-expression with chaperones

Chaperones play a major role in the quality control systems regarding protein conformation and folding. Protein misfolding due to insufficient availability of molecular chaperones is a usual obstacle in recombinant protein production in bacterial cell factories. Co-production of selected sets of cell chaperones along with the target proteins is a common approach to improve the yield of the properly folded form of those proteins [135].

Co-expression of *Tc*CDA with the chaperon system groES-groEL was tested. *E. coli* BL21(DE3) star competent cells were transformed with both the pET22b(+) plasmid containing the *Tc*CDA gene and the pGro7 commercial plasmid, which encodes for the *E. coli* chaperones groES-groEL complex and includes a chloramphenicol resistance gene. The complex is under the control of the araB promoter and, thus, L-arabinose was used for induction (Table 5.18). A protein band with the expected molecular weight corresponding to the chaperone complex (band at Mw 66 kDa) was observed by SDS-PAGE in the soluble fraction after cell lysis. However, even with chaperones expression, *Tc*CDA was still detected only in the insoluble fraction (Figure 5.15).

Table 5.18. Conditions used for improving TcCDA expression by co-expressing it with chaperones.

Experience	[L-Arabinose] (%)	[IPTG] (mM)	OD_{600} of L-Arabinose addition	OD_{600} of IPTG induction	Induction Tª (°C)	Agitation (rpm)
1	0.05	1	0.371	0.567	24	170
2	0.2	1	0.411	0.628	24	170

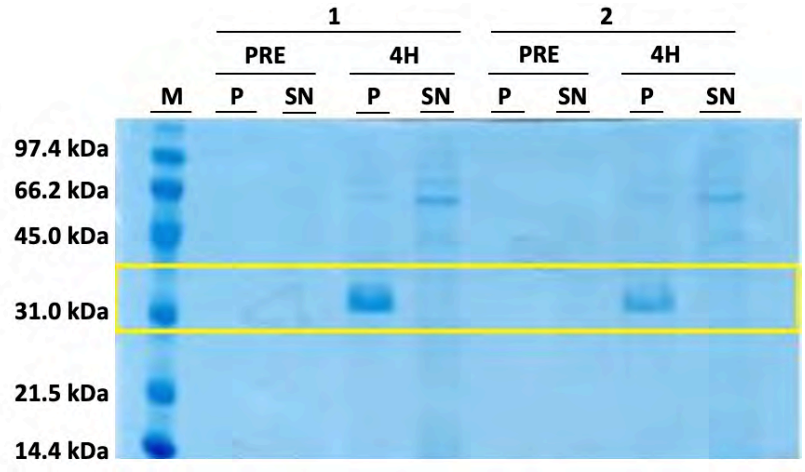


Figure 5.15. SDS-PAGE 14% acrylamide of expression trials of TcCDA. *Expected Mw: 32.5 kDa (yellow square). Experiences 1 and 2 (corresponding to Table 5.18 conditions). M) Molecular weight marker; PRE) before induction; 4H) after 4 hours induction; SN) supernatant (soluble fraction); P) pellet.*

- Use of detergent

Sarkosyl is a mild detergent that has a negatively charged carboxylate group similar to SDS (sodium dodecyl sulfate) but it does not have a similar denaturing effect and that it is used in protein purification protocols due to its demonstrated ability to solubilize folded recombinant proteins from inclusion bodies [136]. Massiah and co-workers developed a protocol in which soaking the insoluble pellet with a higher concentration of sarkosyl (10% w/v) and for longer periods of time (overnight) than usual protocols yielded soluble folded protein. This method was tested with the insoluble fraction (see Section 7.2.2) obtained after cell lysis obtaining successful results. Before the addition of the detergent, the recombinant protein was only detected in the pellet. However, after sarkosyl treatment, an important amount of folded protein was recovered in the soluble fraction (Figure 5.16).

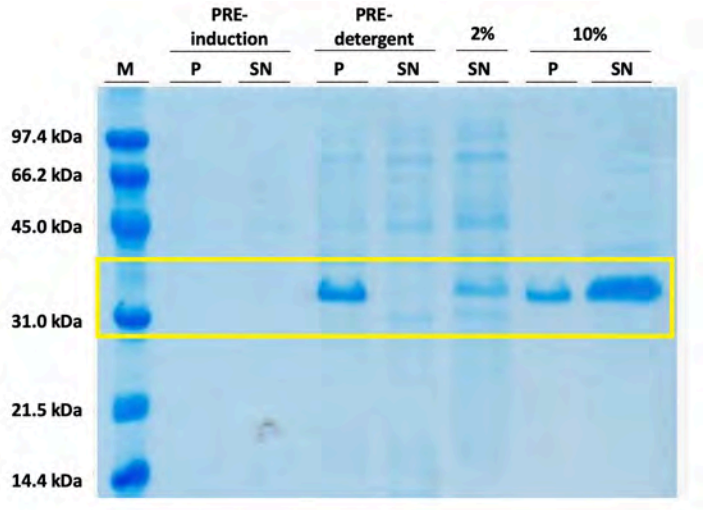


Figure 5.16. SDS-PAGE 14% acrylamide of expression trials of TcCDA. *Expected Mw: 32.5 kDa (yellow square). Overnight induction with IPTG 0.5 mM at 24°C and 170 rpm. M) Molecular weight marker; PRE-induction) before induction; PRE-detergent) before treatment with sarkosyl; 2%) treatment with 2% sarkosyl; 10%) treatment with 10% sarkosyl; SN) supernatant; P) pellet.*

As a summary, different expression conditions were tested, and several methods designed to enhance the proper folding of recombinant proteins in bacterial hosts were used. Treatment of the insoluble fraction (which contained the expressed protein in inclusion bodies) with sarkosyl detergent allowed protein solubilization and increased remarkably the amount of available soluble protein. In consequence, treatment with sarkosyl detergent was the chosen method for the expression of *Tc*CDA on *E. coli.*

The isolation of the recombinant protein was achieved by His-tag affinity chromatography of the sarkosyl-solubilized inclusion bodies (Figure 5.17), obtaining the final pure protein with an approximate yield of 3 – 4 mg of pure protein per liter of culture, enough for enzyme characterization.

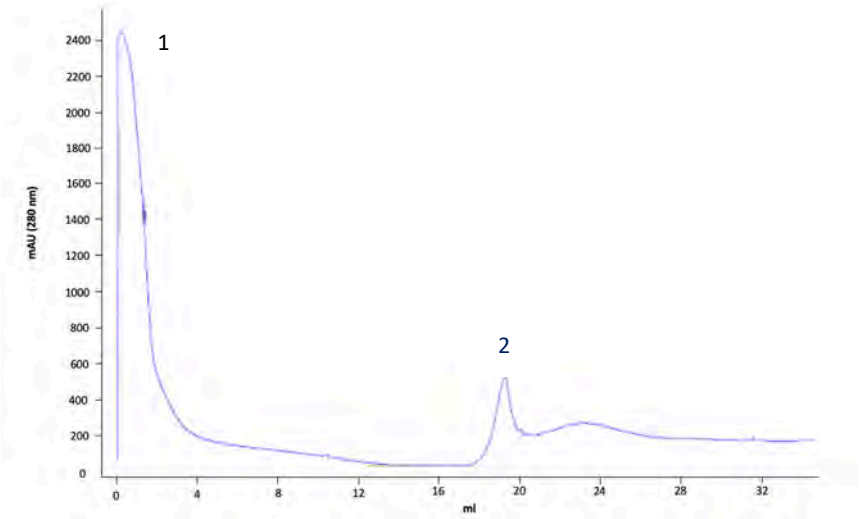


Figure 5.17. Chromatogram of His-trap purification of TcCDA*. Resulting chromatogram from affinity chromatography with TcCDA lysate with 10% sarkosyl. 1. Washing with PBS, 2. Elution with PBS + Imidazole.*

5.2.3. Activity of *Tc*CDA on COS

Once the bacterial recombinant expression of the enzyme was optimized and the pure protein was obtained, the activity of the possible new CDA from *Tilletia controversa* on chitin oligosaccharides was tested. DP2 to DP5 oligomers were used as substrates and deacetylase activity was analyzed by monitoring of those oligomers and their deacetylated products by HPLC-MS as previously described (see Section 7.4.2). The enzyme did not show any activity on $(GlcNAc)_2$ and $(GlcNAc)_3$, but its deacetylase activity on longer substrates was confirmed, being tetraacetylchitotetraose its minimal substrate. After analyzing samples at different reaction times, it was possible to identify up to di-deacetylated product (A2D2) with DP4 as substrate (Figure 5.18) and to tri-deacetylated one with DP5 as substrate (A2D3) (Figure 5.19).

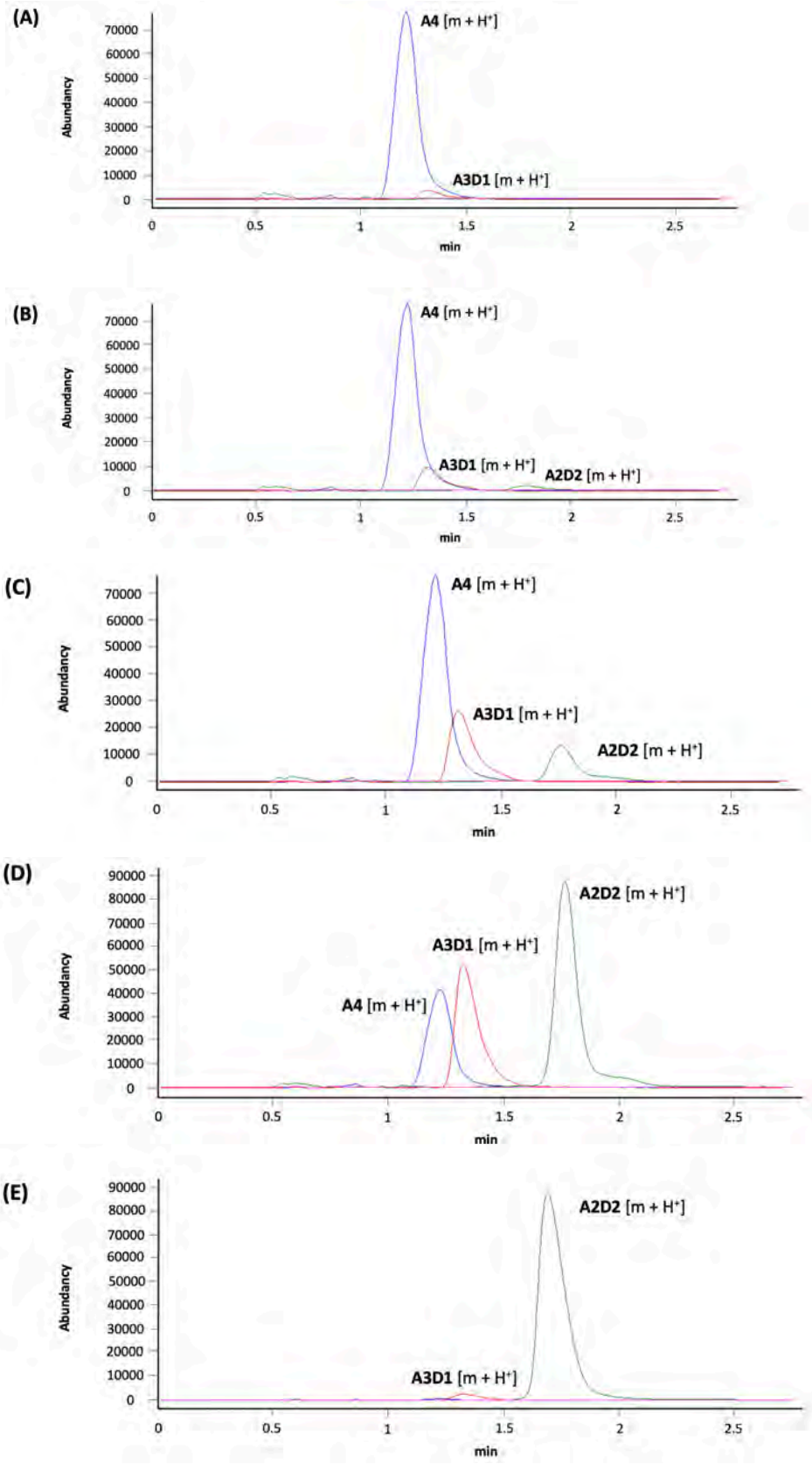


Figure 5.18. SIM (Single Ion Monitoring) HPLC-MS analysis of deacetylation of DP4 with TcCDA. *Products evolution through reaction time: (A) 30minutes, (B) 1 hour, (C) 2 hours, (D) 5 hours, (E) 24 hours.*

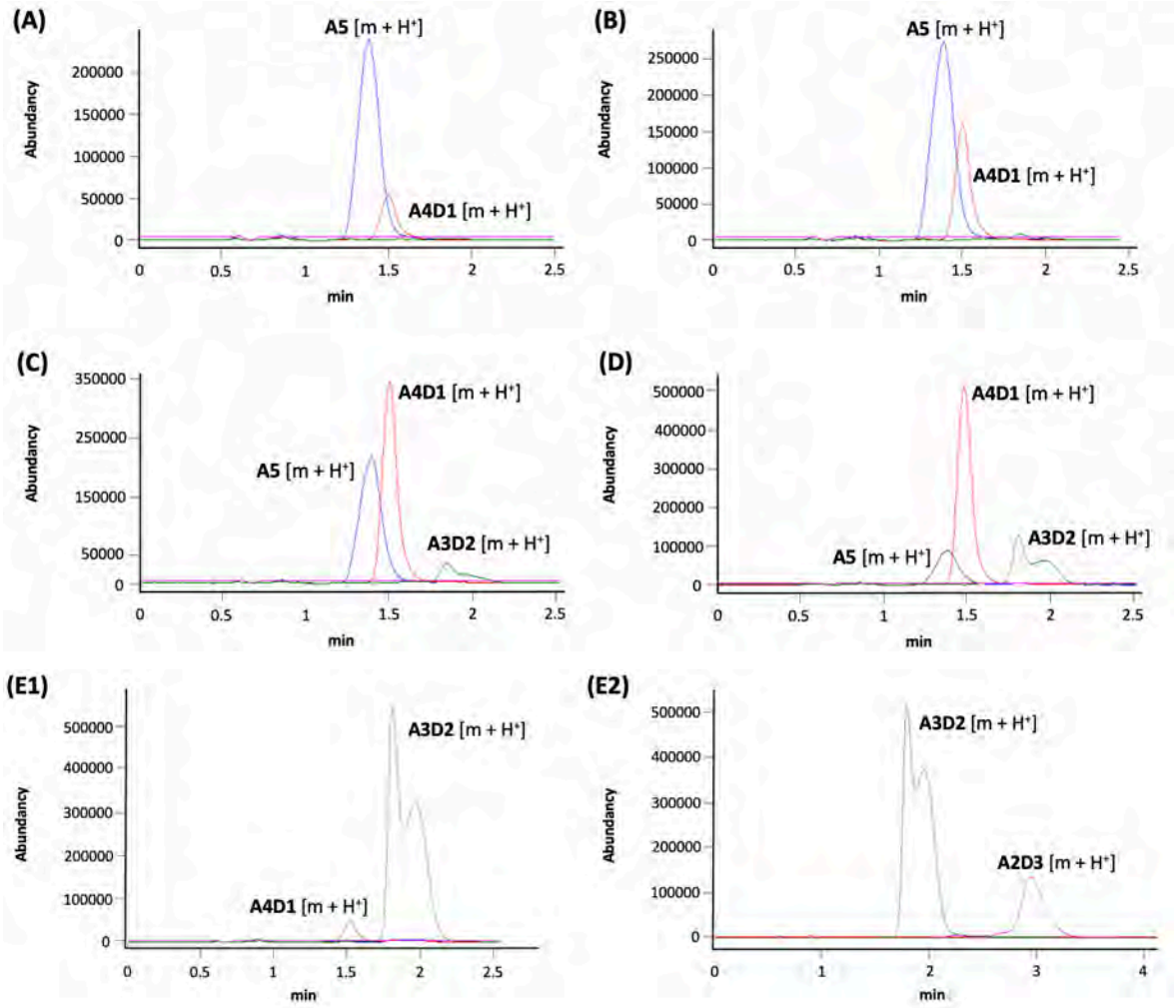


Figure 5.19. SIM (Single Ion Monitoring) HPLC-MS analysis of deacetylation of DP5 with TcCDA. *Products evolution through reaction time: (A) 30minutes, (B) 1 hour, (C) 2 hours, (D) 5 hours, (E) 24 hours. (1) Monitoring of m/z values corresponding to [m + H⁺] of A5, A4D1 and A3D2; (2) Monitoring of m/z values corresponding to [m + H⁺] of A3D2, A2D3, A1D4, and D5.*

These results allowed the identification of the new enzyme as a chitin deacetylase. Considering that its sequence was initially retrieved from *Pgt*CDA's group and that it was selected because it lacks the aromatic residue from motif 4, it is possible that one of the deacetylated GlcNAc units from the substrate oligomers is the reducing end. Further characterization should include the determination of the products' PA in order to get further insights into the possible role of this aromatic residue in preventing the deacetylation of the substrates' reducing end by establishing a strong interaction with it in the +1 subsite.

5.2.4. Next: expression and characterization of all new proteins

The previously conducted phylogenetic and bioinformatics study of family CE4 allowed the selection of several new sequences retrieved from databases that could correspond to new enzymes with novel specificities.

On the one hand, optimization of the recombinant expression of the new proteins is needed to be able to proceed with their further characterization as possible CDAs. Most genes are of eukaryotic origin and those recombinant proteins can have a tendency to aggregate or form inclusion bodies (IBs) in *E. coli*, hampering their proper expression on this simpler bacterial host. Several strategies have been developed and used to improve either protein expression or protein solubility. The possible reason for low or nor expression is protein toxicity (either before or after induction), being the control of the basal expression or of the level of induction the most assayed solutions [137]. Inclusion body formation can be mainly explained by an incorrect disulfide bond formation, the incorrect folding of the protein or its low solubility. The most used methods to overcome misfolding are the co-expression of molecular chaperones or the supplementation of the culture medium with chemical chaperones and co-factors. Fusion of the target protein to solubility enhancers (such as SUMO or Maltose Binding Protein) and treatment with detergents (such as sarkosyl or CHAPS) are other possible solutions to improve protein solubility. However, there are cases in which protein insolubility has to be addressed by changing the host microorganism [135], [136].

On the other hand, once expressed and purified, the final goal is the characterization of the new enzymes as chitin deacetylases by means of studying their activity and specificity on chitin oligosaccharides. Sequence selection was made considering the size of the loops that shape the binding cleft of this family of enzymes with the aim of finding new patterns of deacetylation in nature.

In order to provide further insights into the structure-function relationship of CDAs and decipher the structural and sequential determinants of their substrate specificity, further structural information of protein-ligand complexes is needed. Complete characterization of this set of chosen enzymes (in terms of both their activity and their structure) could be a step towards the discovery of novel CDAs with controlled specificities on the deacetylation of chitin for biotechnological applications.

5.3. Peptidoglycan deacetylase from *Bacillus cereus*

5.3.1. Subcloning, expression and purification

The subcloning strategy designed and followed for the gene of the selected peptidoglycan deacetylase from *B. cereus* was the same described above for all the new chitin deacetylases. In the same way as most of the putative chitin deacetylase genes, the selected sequence presented a N-terminal signal peptide that was not included in the final synthetic gene, which was produced codon optimized for *E. coli* and provided in a pMK-QR plasmid. After subcloning into the pET22b(+) by restriction and ligation, the sequence of the obtained plasmid was confirmed by Sanger sequencing and it was transformed into competent *E. coli* DH5α and *E. coli* BL21(DE3) star cells.

The autoinduction protocol previously used for all *Bs*PdaC constructs was used for the expression of this new recombinant enzyme on *E. coli* BL21(DE3) star cells. The purification protocol consisted in an affinity chromatography with a His-trap column and high yields of more than 100 mg of pure protein per liter of culture were obtained (Figures 5.20 and 5.21).

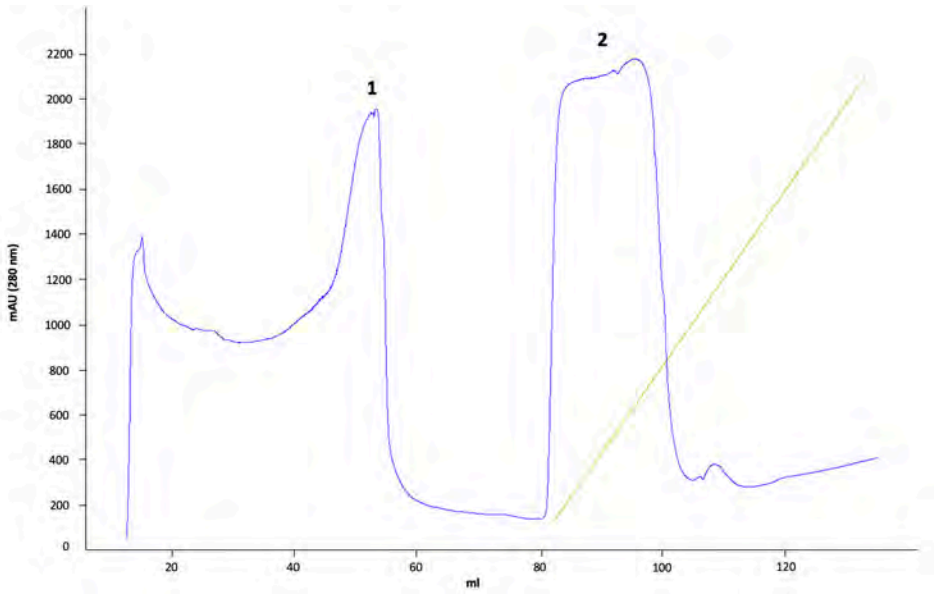


Figure 5.20. Chromatogram of His-trap purification of BcPda*. Resulting chromatogram from affinity chromatography with His-trap column. 1. Washing with PBS, 2. Elution with PBS + Imidazole. Green line) Imidazole %.*

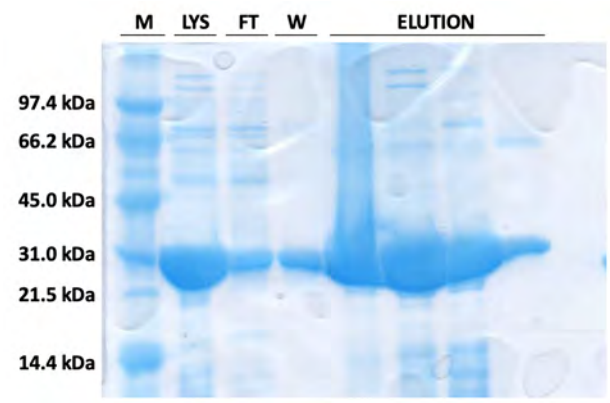


Figure 5.21. SDS-PAGE 14% acrylamide of affinity purification of BcPda. *Expected Mw: 29.1 kDa. M) Molecular weight marker; LYS) lysate; FT) flow through; W) Washing with PBS; ELUTION) different samples of peak 2 in Figure 5.3-1.*

In order to identify possible oligomeric fractions of the enzyme, a second chromatographic step with a size exclusion chromatography column (Superdex 200 16/600) was performed, but the chromatographic profile showed one single peak whith retention time corresponding to the molecular weight of *Bc*Pda's monomer (29.1 kDa) (Figure 5.22).

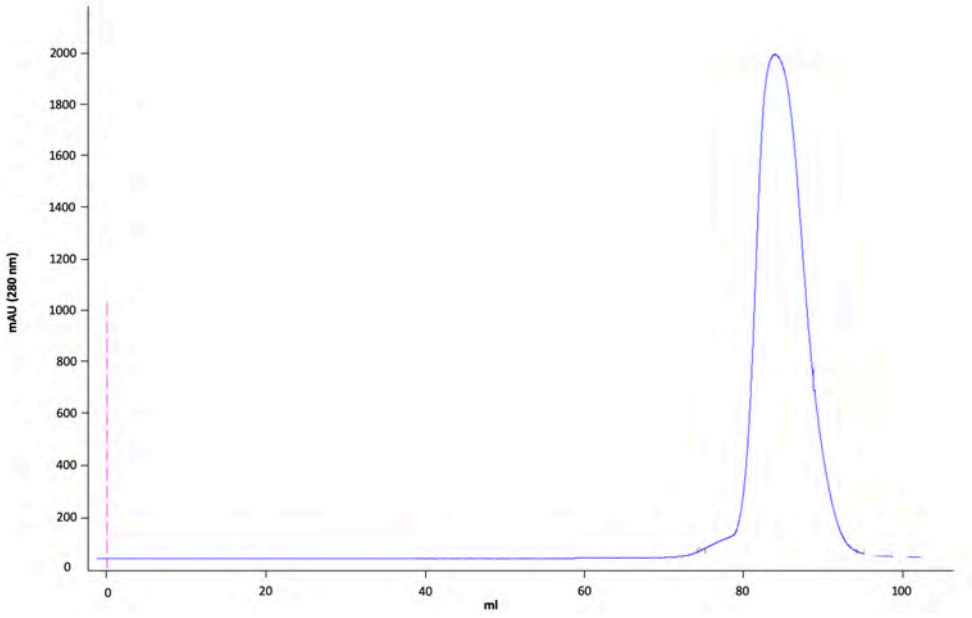


Figure 5.22. Size Exclusion Chromatography of BcPda*. Resulting chromatogram from SEC with a superdex 200 16/60 column.*

5.3.2. Activity characterization of *Bc*Pda

Once the new protein was recombinantly expressed and purified with success, its activity against available substrates was tested. *Bc*Pda was chosen for being phylogenetically close to the two only enzymes classified as MurNAc deacetylases. PdaA from *Bacillus subtilis* is the only one kinetically characterized and it has been reported to only deacetylate *N*-acetylmuramic acid residues without

peptide side chains. It does not recognize native peptidoglycan, soluble chitin, chitin oligomers or GlcNAc as substrates [49].

5.3.2.1. Activity on chitin oligomers

Activity of *Bc*Pda on chitin oligomers, with a degree of polymerization from 2 to 5, was studied following the same procedures previously described, analyzing the possible deacetylase activity by HPLC-MS monitoring (see Section 7.4.2). Several reactions with different enzyme concentration and incubation times were set up but it was not possible to identify deacetylated products in none of the analyzed samples. Furthermore, the signals obtained for the initial substrates were maintained along the reaction time, indicating that the reactions between enzyme and substrates did not occur (Figure 5.23).

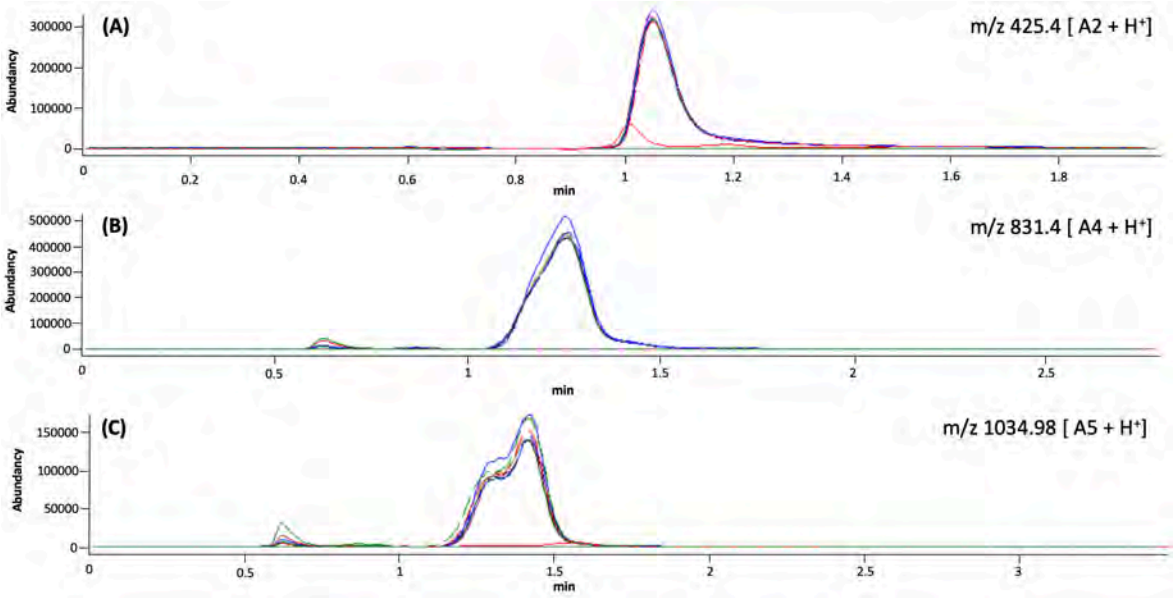


Figure 5.23. SIM (Single Ion Monitoring) HPLC-MS analysis of COS de-N-acetylation with BcPda. *Superposition of samples from different reaction times with different substrates. (A) $GlcNAc_2$ as substrate; (B) $GlcNAc_4$ as substrate; (C) $GlcNAC_5$ as substrate.*

5.3.2.2. Activity on MDP

Considering sequence similarity, this new enzyme should act as a MurNAc peptidoglycan deacetylase. MDP (N-Acetylmuramyl-L-alanyl-D-isoglutamine), consisting in a MurNAc monomer with a dipeptide, was tested as substrate. Again, several reactions with different enzyme concentration and incubation times were set up and the corresponding samples were analyzed by fluorescamine labelling. In the same way as what was previously observed with COS, no deacetylase activity could be detected. Although the monosaccharide is a *N*-acetylmuramic unit, it is probable

that a longer substrate is needed for the catalytic mechanism to take place. Furthermore, knowing that its close homolog PdaA only acts on PGN without peptide side chains, the presence of the di-peptide substitution could also prevent enzyme's activity.

5.3.2.3. Activity on peptidoglycan from *B. subtilis*

The polymeric peptidoglycan from Bacillus subtilis' cell wall was used as a possible substrate for *Bc*Pda. In this case, due to the high insolubility of PGN in aqueous solutions, the reactions were performed as homogeneous suspensions in agitation and the possible deacetylation of the polymer was monitored by acetate release determination using an enzymatic coupled assay (see Section 7.4.4). Although several conditions were assayed, no deacetylase activity on intact peptidoglycan could be detected. As already explained, if this new enzyme had the same specificity than PdaA, it would be necessary to digest the peptidoglycan substrate to remove the stem peptide in order to make it a suitable substrate.

5.3.2.4. Esterase activity on AcOMU

Since the new recombinant protein did not show to be active on any of the tested substrates, its activity against the unspecific substrate AcOMU (4-methylumbelliferyl acetate) was also studied to assess its general esterase activity. The same protocol previously described, with all the obtained fractions from size exclusion chromatography, was performed and the release of the fluorogenic product from the hydrolysis reaction was monitored (Figure 5.24)

In spite of being the only activity that could be detected, the fact that the enzyme acted on this substrate indicated a proper folding of the recombinantly expressed protein.

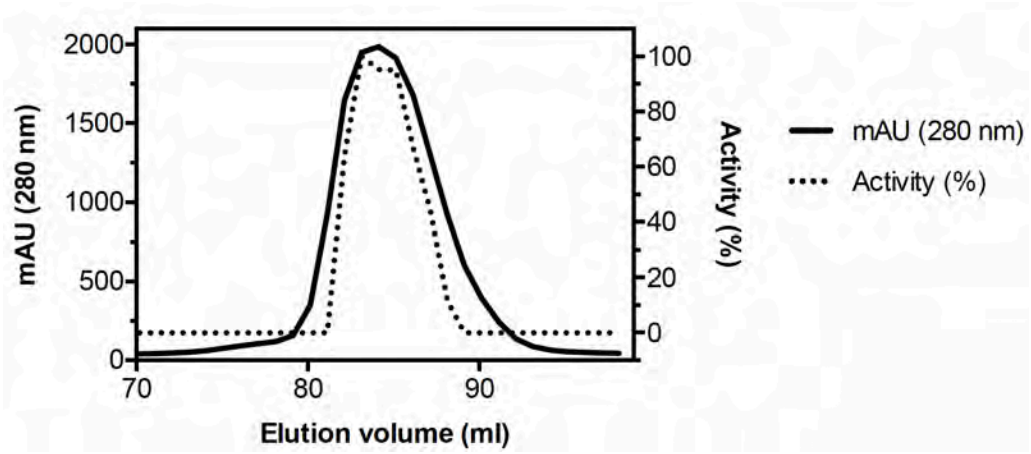


Figure 5.24. Activity profile of SEC fractions from BcPda purification. *Gel filtration chromatography of BcPda and enzyme activity with AcOMU substrate.*

5.3.3. Next: further characterization of *Bc*Pda

Several sequences from different sources have been annotated as possible peptidoglycan deacetylases, both with MurNAc or GlcNAc activity, due to their high similarity to the characterized ones. Yet, few enzymes have been recombinantly expressed and biochemically characterized. The selection of this new enzyme from *Bacillus cereus* for its characterization aimed to increase the available information on peptidoglycan deacetylases and to gain knowledge about the determinants of their specificity. Considering the phylogenetic analysis performed, it is expected for this enzyme to be a specific peptidoglycan MurNAc deacetylase as an ortholog to *Bs*PdaA. In order to confirm its specificity in the same way that was previously done for *Bs*PdaA [49] and *Bs*PdaC [121], the preparation of suitable peptidoglycan substrates following Vollmer and Tomasz procedures [138] is needed.

As previously described, these enzymes cannot only distinguish between MurNAc and GlcNAc residues, but also show different preferences regarding the presence or absence of the MurNAc peptidyl substitution. In order to generate peptidoglycan glycan strands with shorter or none stem peptide, peptidoglycan has to be initially treated with specific cell wall hydrolases from *Bacillus subtilis* (CwlH, a L-alanine amidase that cleaves the peptide leaving peptidoglycan with no peptidyl substitution and CwlE, a DL-endopeptidase that cleaves the peptide between Glu and A_2pm, leaving just the di-peptide L-Ala-D-Glu bound to the MurNAc units of peptidoglycan) that are currently being subcloned and recombinantly expressed.

Further steps on this project should include the production of the needed defined substrates to characterize the new enzyme in terms of kinetics and substrate specificity, as well as solving its crystal structure.

The performed bioinformatics analysis, including the search in the main databases and the comparative study of a large number of sequences, has confirmed that the high degree of sequence similarity between all peptidoglycan deacetylases does not allow the identification of a clear signature to distinguish between the different sub-specificities. It is necessary to enlarge the available knowledge on this family of enzymes, characterizing new proteins from the three subgroups (GlcNAc deacetylases, canonical MurNAc deacetylases and PdaC-like MurNAc deacetylases) in order to elucidate which are the sequence and structural determinants that guide their differential specificity

CONCLUSIONS

6. CONCLUSIONS

The main conclusions drawn from the work carried out in the two chapters of this doctoral thesis are detailed below.

Conclusions Chapter 1: PdaC from *Bacillus Subtilis*

1. The conserved catalytic residues, corresponding to Asp285 and His427, and the metal-coordination triad Asp286, His336 and His340, were identified in the multiple sequence alignment. Moreover, the 6 loops described by the Subsite Capping Model were located in the *Bs*PdaC sequence, showing to be short in comparison with other family members.
2. The phylogenetic analysis based on the multiple sequence alignment showed that the GlcNAc deacetylase *Sp*PgdA (with 30% overall sequence identify with *Bs*PdaC) is the closest CE4 enzyme to *Bs*PdaC.
3. The truncated construct of *Bs*PdaC, including only its CE4 catalytic domain (*Bs*PdaC-CD) showed similar levels of expression, thermal stability and activity than the full-length enzyme (*Bs*PdaC-FL).
4. *Bs*PdaC-CD showed to be a metalloenzyme highly active on intact PGN, with the same specific activity and retaining the same MurNAc specificity that the full-length enzyme. The enzyme is also active on COS with a minimum DP of 3. This behavior is unique among characterized MurNac deacdetylases, which do not deacetylate COS and act on modified PGN without the peptidyl substitution.
5. Determination of the mode of action and the pattern deacetylation of the enzyme on COS substrates showed that *Bs*PdaC follows a multiple-attack mechanism starting with deacetylation of the internal GlcNAc residues and proceeding to deacetylate all but the reducing-end unit of the substrate.
6. The x-ray 3D structure of the CE4 catalytic domain of *Bs*PdaC was solved at 1.54 Å resolution.
   6.1. The catalytic site retains the conserved Asp-His-His metal-binding triad characteristic of CE4 enzymes acting on GlcNAc residues, differing from other MurNAc deacetylases that lack the metal-coordinating Asp.
   6.2. *Bs*PdaC revealed short loops, similar in shape and length to other PGN deacetylases, that result in an open binding cleft able to accommodate polymeric substrates and exhibit the described multiple-chain mode of action.

6.3. Comparison of the 3D structure of *Bs*PdaC with other PGN deacetylases unrevealed significant differences at their active sites that can be related with the distinctive specificity of the enzyme on PGN.

6.4. Docking simulation experiments with $(GlcNAc)_4$ showed two preferred binding modes for the first deacetylation event (BM1 and BM2), which were in agreement with the experimental results.

7. The deletion of interactions between selected residues from the active site (Y430, W402, W395 and D398) and the substrate units caused a decrease of 2 or 3 orders of magnitude of their specific activity, indicating their importance for substrate binding.
8. Mutation of Arg315, which interacts with the non-reducing end of substrate in BM1, altered the pattern of deacetylation of the wild type enzyme by destabilizing BM1 and favoring the formation of the BM2 complex.

Conclusions Chapter 2: discovery of new CE4 enzymes

1. In the search for new specificities, five new sequences corresponding to putative chitin deacetylases were identified and selected from a previous extensive bioinformatics and phylogenetic study of CE4 family.
2. All sequences were subcloned for their expression on a bacterial host, but only *Ns*CDA and *Tc*CDA showed some soluble protein expression and were purified as recombinant enzymes. Expression of *Tc*CDA on *E. coli* was optimized by treatment with sarkosyl detergent, allowing protein solubilization from inclusion bodies.
3. *Tc*CDA showed activity on $(GlcNAc)_4$ and $(GlcNAc)_5$, rendering up to di-deacetylated products with DP4 as substrate and to tri-deacetylated oligomers with DP5 as substrate.
4. From the retrieved sequences from different psi-blast searches against several databases, BC0467 (*Bc*Pda) from *Bacillus cereus* was selected as a potential MurNAc deacetylase, according to sequence similarity to *Bs*PdaA and *Ba*Pda.
5. *Bc*Pda was subcloned, expressed and purified as a recombinant enzyme and did not show activity against any of the tested substrates (chitin oligomers, MDP, and peptidoglycan from *B. subtilis*), although its general esterase activity was confirmed by incubating the enzyme with the unspecific substrate AcOMU. The results are in agreement with the hypothesis of *Bc*Pda being a canonical MurNAc deacetylase, in which case it would be necessary to digest the peptidoglycan substrate to remove the stem peptide in order to make it a suitable substrate.

MATERIALS AND METHODS

7. MATERIALS AND METHODS

In this chapter, the experimental part of the project is described. In some sections, the theoretical foundation of the strategies or techniques is briefly explained. In those cases, this theoretical context is framed in gray boxes and its reading can be omitted if it is not considered necessary.

7.1. DNA manipulation

7.1.2. *Bs*PdaC-FL and *Bs*PdaC-CD subcloning

EXPRESSION VECTORS: PET EXPRESSION SYSTEM

Expression systems are designed to produce multiple copies of a protein of interest in the host cell by introducing an expression vector into the cell. This vector must contain all the necessary genetic codification for the production of the protein, including its own origin of replication, an appropriate promotor, the sequence that terminates transcription and the codifying sequence for ribosomal binding. In addition, expression vectors generally include a selectable genetic marker that allows the isolation of the cells containing the plasmid, such as the codifying sequence for an antibiotic resistance.

The pET expression system (Figure 7.1) allows the fast production of a high quantity of protein once induced and its regulation is based on the use of the promoter of the T7 phage, which will only bind the virus RNA polymerase and not the host one. When the plasmid is transferred to a strain that contains a chromosomal copy of the T7 polymerase under lac operator control and in presence of an inducer, the polymerase is produced and recognizes its T7 promoter in the plasmid, starting the expression of the cloned protein gene. The most used inducer is the lactose analog IPTG (isopropyl β-D-1-thiogalactopyranoside), however, the protein expression can be achieved by other types of inducers, such as the formulation of autoinduction media [139]. The pET22b(+) plasmid includes a His-tag for the later purification of the expressed enzyme by Immobilized Metal Affinity Chromatography (IMAC).

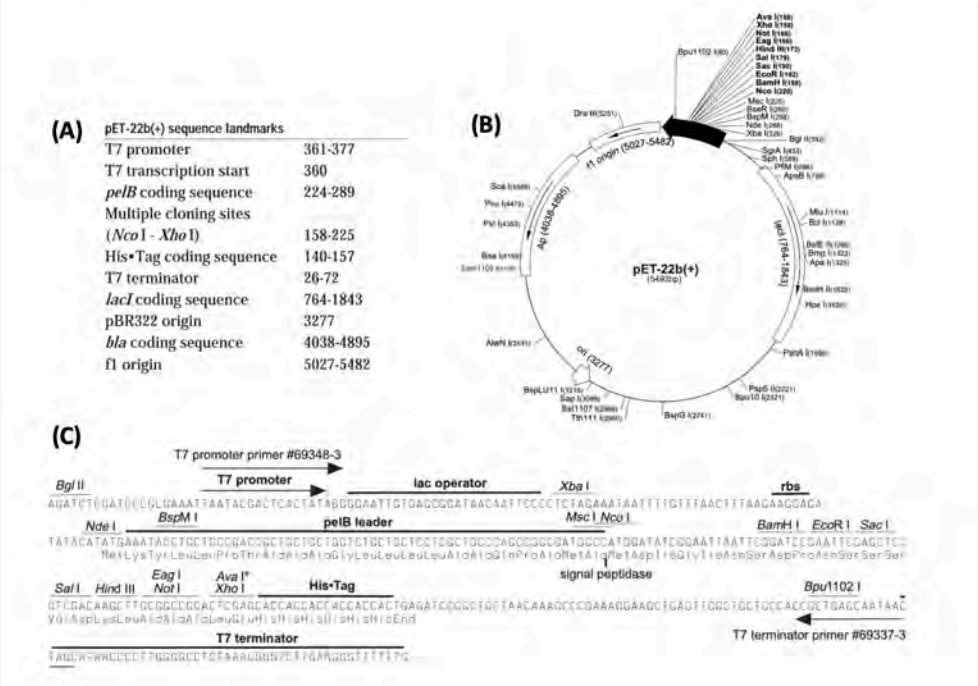


Figure 7.1. pET22b(+) expression system. *(A) Sequence landmarks present on the vector. (B) Schematic representation of the vector and location of restriction sites on the sequence. (C) Sequence of the cloning/expression region.*

A synthetic and codon-optimized gene coding for *Bs*PdaC (amino acid residues 27-467, without signal peptide and Nt-transmembrane region) (GeneArt gene synthesis service, Thermo Fisher Scientific) was cloned into a pET22b(+) vector containing a C-terminal Strep-tag sequence (WSHPQFEK). Both the synthetic gene and the vector were digested with the restriction enzymes NdeI and SacI (New England Biolabs) and ligated with T4 DNA ligase (Bio-Rad). From the generated construct (pET22b-*Bs*PdaC-FL), the catalytic domain (amino acid residues 270-467) was amplified by PCR with primers CD_fwd and CD_rvs (Table 7.1) and cloned into the pET22b(+) following the same strategy.

Table 7.1. Cloning primers for BsPdaC-CD generation.

CLONING PRIMER	DNA SEQUENCE
CD_fwd	5'-TAGTGGCATATGGAAGAAACAGTTGATCCGAATC-3'
CD_rvs	5'-GACCATTGTCGACGGAG-3'

E. coli DH5α competent cells were transformed, positive cells were selected from LB-Ampicillin plates and DNA sequence were confirmed by Sanger sequencing. *E. coli* BL21(DE3) star cells were transformed with the resulting plasmids for later protein expression.

7.1.3. Generation of new *Bs*PdaC constructs

The second generation of constructs were designed with the Strep-tag at the N-terminus end of the proteins. To generate these new constructs, the desired sequences were amplified from pET22b-*Bs*PdaC-FL by PCR and cloned into a pPR-IBA2 vector, which regulations is also based on the T7 phage promoter, but contains an N-terminal Strep-tag sequence. Both the amplified genes and the vector were digested with the restriction enzymes KasI and PstI (New England Biolabs) and ligated with T4 DNA ligase (Bio-Rad).

The cloning primers used to amplify the desired fragments are listed in Table 7.2. The forward primers contain a KasI restriction site (underlined) and the tobacco etch virus proteolysis site (in bold), and the reserve primers contain a PstI restriction site (underlined).

Table 7.2. Cloning primers for BsPdaC's new constructs generation. *Sequence of primers used to amplify the desired fragments from BsPdaC to generate the designed new constructs. Underlined: KasI restriction site in fwd primers and PstI restriction site in rvs primers; Bold: TEV protease restriction site.*

CONSTRUCT (AA)	CLONING PRIMER	DNA SEQUENCE
*Bs*PdaC-A (29 – 234)	A_fwd	5'-TAGTGGCGCC**GAAAACCTGTATTTTCAGGGC**AGCCTGAAAAAA GAAACCGTGATGAATAAAG-3'
	A_rvs	5'-TATCCTGCAGTTTTTTGATTGCAATGCTCTGTTCACCCAG-3'
*Bs*PdaC-B (29 – 453)	B_fwd	5'-TAGTGGCGCC**GAAAACCTGTATTTTCAGGGC**AACCACAGCCTGAA AAAAGAAACCGTG-3'
	B_rvs	5'- TATCCTGCAGGGTAACCAGCTGATAACCCTGATCG-3'
*Bs*PdaC-C (277 – 453)	C_fwd	5'-TAGTGGCGCC**GAAAACCTGTATTTTCAGGGC**CAGAAAGTTATTGCC CTGACCTTTGATGATGG-3'
	C_rvs	5'- TATCCTGCAGGGTAACCAGCTGATAACCCTGATCG-3'
*Bs*PdaC-C-2 (29 – 467)	C-2_fwd	5'- TAGTGGCGCC**GAAAACCTGTATTTTCAGGGC**CAGAAAGTTATTG CCCTGACCTTTGATGATGG-3'
	C-2_rvs	5'- TATCCTGCAGCTATTTTGCTTCACGCTGTTTTTTAAC-3'

E. coli DH5α competent cells were transformed, positive cells were selected from LB-Ampicillin plates and DNA sequence was confirmed by Sanger sequencing. *E. coli* BL21(DE3) star cells were transformed with the resulting plasmids for later protein expression.

7.1.4. Generation of mutants

All mutants were obtained by site-directed mutagenesis, with a modified Quick-Change SMD protocol (Agilent). For each single mutant, a sense/antisense mutagenic primer pair was designed. Each pair contained non-overlapping sequences at their 3' end, primer-primer overlapping sequences at the 5' end and the mismatching nucleotides for the mutation located at the center of the overlapping region. The minimum number of mismatched nucleotides was introduced for each desired mutated amino acid and codon usage for *E. coli* was considered for the design of primers. The mutagenic primers are listed in Table 7.3.

Table 7.3. Mutagenic primers. *Sequence of mutagenic primers used to generate the designed single mutants of BsPdaC. Underlined: changed nucleotides.*

MUTANT	CLONING PRIMER	DNA SEQUENCE
*Bs*PdaC-D285S (of any *Bs*PdaC construct)	Fwd	5'-CTGACTTTTCTGATGGTCCGAATCCGCAACC -3'
	Rvs	5'- GGATTCGGACATCAGAAAAGGTCAGGGCAATAACTTTC -3'
*Bs*PdaC-Y377A	Fwd	5'- GTCCGCCTGCGGGTGGTATTAATGATGAACTGC -3'
	Rvs	5'- ATACCACCCGCAGGCGGACGAACCAGG -3'
*Bs*PdaC-W402A	Fwd	5'- GGATCCGGAAGATGCGAAAGATCGTAACAAAAAAACC
	Rvs	5'- CGATCTTTCGCATCTTCCGGATCCACATCCC -3'
*Bs*PdaC-Y430A	Fwd	5'- TGATATTGCGCGTACCAGCGCAGATG -3'
	Rvs	5'- CTGGTACGATAACGCTCATGAATCAGAATGGTAC -3'
*Bs*PdaC-W395A	Fwd	5'- GAAAATGGATGTTGCACTGGCGGATGTGGATCCGG -3'
	Rvs	5'- CCACATCCGCCAGTGCAACATCCATTTTCATCTGGC -3'
*Bs*PdaC-D398E	Fwd	5'- GGATGTGGAACCGGAAGATTGGAAAGATCG -3'
	Rvs	5'- CTTCCGGTTCCACATCCCACAGTGC -3'
*Bs*PdaC-D398N	Fwd	5'- GGATGTGAATCCGGAAGATTGGAAAGATCG -3'
	Rvs	5'- CTTCCGGATTCACATCCCACAGTG -3'
*Bs*PdaC-D398A	Fwd	5' GGATGTGGCCCCGGAAGATTGGAAAGATCG -3'
	Rvs	5'- CTTCCGGGGCCACATCCCACAGTGC -3'
*Bs*PdaC-D398L	Fwd	5'- GGATGTGCTGCCGGAAGATTGGAAAGATCG -3'
	Rvs	5'- CTTCCGGCAGCACATCCCACAGTGC -3'
*Bs*PdaC-D398K	Fwd	5'- GGATGTGAAACCGGAAGATTGGAAAGATCG -3'
	Rvs	5'- CTTCCGGTTTCACATCCCACAGTG -3'
*Bs*PdaC-R315A	Fwd	5'-TGGGTAGCGCGGTTCAGTATTATCCGGAAACC -3'
	Rvs	5'- ATACTGAACCGCGCTACCCAGAACAAAAAAGG -3'
*Bs*PdaC-R315I	Fwd	5'- TTCTGGGTAGCATTGTTCAGTATTATCCG -3'
	Rvs	5'- ACTGAACAATGCTACCCAGAACAAAAAAG -3'
*Bs*PdaC-R315S	Fwd	5'- GTTCTGGGTAGCAGCGTTCAGTATTATCCG -3'
	Rvs	5'- ACTGAACGCTGCTACCCAGAACAAAAAAG -3'
*Bs*PdaC-R315K	Fwd	5'- TGGGTAGCAAAGTTCAGTATTATCCGGAAAC -3'
	Rvs	5'- ACTGAACTTTGCTACCCAGAACAAAAAAGG -3'

Plasmid pET22b(+)-*Bs*PdaC-FL was used as template for mutagenesis. PCR were carried out using IProof polymerase (0.04 U/µL, Bio-Rad) and a thermocycler Gene Pro Thermal Cycler (Bioer). Amplifications were carried out in a 40 µL final volume of the reaction mixture detailed in Table 7.4 and the PCR reaction detailed in Table 7.5.

Table 7.4. PCR reaction mixture.

	VOLUME (ML)
Template DNA (pET22b-*Bs*PdaC-FL)	X (3-5 ng)
Forward mutagenic primer (5 µM)	3
Reverse mutagenic primer (5 µM)	3
Iproof Master Mix 2x	10
Autoclaved MilliQ water	X (to total volume)
Total	40

Table 7.5. PCR reaction conditions.

STEP	**TEMPERATURE (°C)**	**TIME**	**CYCLES**
Initial denaturation	98	3 min	-
Denaturation	95	30 sec	30x
Annealing	Primers Tm - 5	30 sec	
Extension	72	2 min 30 sec	
Final extension	72	3 min	-

After PCR reaction, samples were digested with Dpn1 for 2-3 hours at 37°C and, then, transformed into *E. coli* DH5α competent cells. Positive transformants were confirmed by Sanger DNA sequencing and the final constructs were used to transform *E. coli* DH5α and *E. coli* BL21(DE3) star competent cells.

7.1.5. New enzymes subcloning

Synthetic and codon-optimized genes coding for *Ns*CDA, *Sp*CDA, *Bg*CDA, *Os*CDA, *Tc*CDA and *Bc*Pda were cloned into a pET22b(+) vector with a C-terminal His-tag sequence. For all cases, both the synthetic gene and the vector were digested with the restriction enzymes NdeI and XhoI (New England Biolabs) and ligated with T4 DNA ligase (Bio-Rad).

E. coli DH5α competent cells were transformed, positive cells were selected from LB-Ampicillin plates and DNA sequence were confirmed by Sanger sequencing. *E. coli* BL21(DE3) star cells were transformed with the resulting plasmids for later protein expression.

7.2. Expression and purification of proteins

7.2.1. Expression with autoinduction medium

AUTOINDUCTION MEDIUM

With this method (contrary to the IPTG induction strategy) is not necessary to add an inducer to the culture at a certain point since the growth of the expression strain is performed in a medium that promotes the induction of the T7 promoter, leading to the expression of the protein of interest.

The medium design is based on the knowledge that lactose induces protein production, which is initially inhibited by other media components that will run out during growth (in this case, another carbon source such as glucose). Initially, the use of glucose as the main carbon source allows the inhibition of protein expression and all cell metabolism is working on division, reaching high cell density. Once this sugar has run out, the use of the lactose (an inducer of the pET system), present on the medium as a new carbon source, will induce protein expression generating a high amount of recombinant enzyme [139]. Besides containing glucose, glycerol and lactose as carbon sources, the medium is composed by LB supplemented with several other additives (Table 7.6) and 100 µg/ml of ampicillin.

Table 7.6. Composition of solutions of autoinduction medium. *The optimized medium is composed of LB (10 g/L tryptone, 5 g/L NaCl, 5 g/L yeast extract) supplemented with the listed solutions and 100 µg/mL ampicillin.*

SOLUTION	COMPONENTS	MEDIUM CONCENTRATION/(%)
Solution M	Na_2HPO_4	0.025
	KH_2PO_4	0.025
	NH_4Cl	0.05
	Na_2SO_4	0.005
Solution 5052	Glycerol	(0.5)
	Glucose	(0.05)
	Lactose	(0.2)
$MgSO_4$	$MgSO_4$	0.002

This type of methods allows the automatic expression of the protein without the need of monitoring cell growth, changing the medium or adding inducers. Moreover, high optical density is reached, increasing the yield compared with other methods.

E. coli BL21(DE3) star cells harboring the plasmids pET22b(+) or pPR-IBA2 with the different *Bs*PdaC constructs and the plasmid pET22b(+) with *Ns*CDA were grown in an autoinduction medium (Table 7.6) at 25°C and 170 rpm for 48 hours. Cells were harvested by centrifugation, suspended in PBS buffer (50 mM phosphate, 300 mM NaCl, pH 7) with 1 mM serine protease inhibitor phenylmethylsulfonyl fluoride (PMSF), and disrupted by sonication with a 6 µM probe (7 minutes, 10 seconds on and 25 seconds off, 50% amplitude) with a Vibra Cell VCX 130 sonicator (SONICS). Cell free extracts were centrifuged (21,000xg, 1 hour at 4°C). Supernatant was used for purification of the recombinant protein.

7.2.2. Expression with IPTG

E. coli BL21(DE3) star cells harboring the plasmids pET22b(+) with *Tc*CDA were grown on LB medium with 100 µg/ml Ampicillin at 250 rpm and 37°C until reaching an OD_{600} of 0.6. Protein expression was induced by adding IPTG at a final concentration of 0.5 mM and incubating overnight at 250 rpm and 25°C. Cells were harvested by centrifugation (4,000xg, 15 min at 4°C), resuspended in 40 ml of PBS with 2% sarcosyl and incubated for 2 hours in agitation and room temperature. Cells were disrupted by sonication as described and centrifuged (12,000xg, 30 min at 4°C). Supernatant was discarded, and the pellet was resuspended in 10 ml PBS with 10% sarcosyl and incubated overnight in agitation at room temperature before using it for purification of the solubilized recombinant protein.

7.2.3. Strep-trap affinity chromatography

A strep-tag II can be found in all plasmids containing *Bs*PdaC constructs, either on the C-terminus (pET22b(+) or the N-terminus (pPR-IBA2) end of proteins. This tag is used for the initial purification by affinity chromatography.

AFFINITY CHROMATOGRAPHY: STREP-TAG

Affinity chromatography is one of the most diverse and powerful chromatographic methods for the purification of a specific molecules or group of molecules from complex mixtures. It is based on highly specific biologic interactions between two molecules such as enzyme-substrate, receptor-ligand or antibody-antigen bindings. These interactions, generally reversible, are used for purifications immobilizing one of the molecules (ligand) in a solid matrix to create a stationary phase while the molecule of interest is on the mobile phase. It is the only technique that allows the isolation of a biomolecule by its biological function or individual chemical structure.

The StrepTrap principle is based on the interaction between d-biotin and streptavidin, one of the strongest non-covalent interactions known. On one hand, a peptide with the capacity to bind to the binding pocket of biotin from streptavidin was designed (strep-tag II, WSHPQFEK). On the other hand, in order to improve the binding properties, the molecule Strep-tactin was developed, an optimized version of streptavidin with 100-fold higher affinity for the peptide [140]–[142]. Due to the chemical composition of its amino acids, the Strep-tag II peptide does not interfere with protein folding or enzyme activity, it does not react with possible buffer impurities, it does not present ionic exchange properties and it does not promote protein aggregation, reasons why, in general, it is not necessary to remove it after purification.

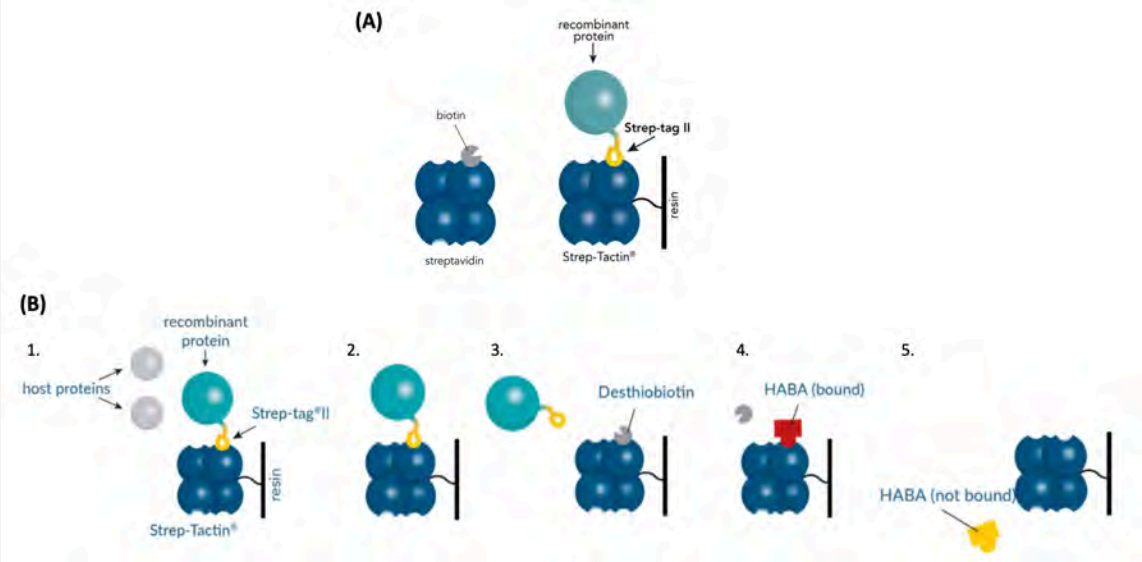


Figure 7.2. Step-tag principle. *(A) Natural binding of biotin to streptavidin versus binding of Strep-tag II peptide with Strep-Tactin on the purification resin. (B) Purification cycle with the Strep-trap system. From IBA-lifesciences.*

The purification is performed with a StrepTrap column into which the clarified cellular lysate can be directly added to pass through. In this step, the recombinant protein with the Strep-Tag II

peptide specifically binds to the Strep-Tactin in the column, being retained. The rest of unspecific not-bound proteins that may remain in the column (*host proteins* in Figure 7.2, B) are washed with a physiologic buffer (the same in which the lysate is dissolved). A buffer solution with 2.5 mM of d-desthibiotin is used to elute the protein of interest. This molecule presents higher affinity for strep-tactin than the recombinant protein and will displace it from the binding sites of the column. Subsequently, in order to regenerate the column, a solution with HABA (2-[4-hydroxy-benzeneazo]benzoic acid) in a great excess is added to displace d-destibiotin. Finally, a last washing step with the initial buffer is performed to remove HABA from the column.

Supernatant from cell extract centrifugation was loaded to a pre-conditioned Strep-Trap 5 ml column in several cycles. The conditioning procedure for a previously used column included several steps, and it was performed at a 2-3 mL/min flow rate:

- 5 column volumes of H_2O.
- 5 column volumes of 0.5M NaOH.
- 5 column volumes of H2O (to completely remove NaOH, it is recommended to check pH).
- 5 column volumes of the buffer in which the protein to purify is resuspended (usually PBS pH 7).
- 2 column volumes of HABA 1 mM in PBS (or until column color has completely changed to red).
- PBS to remove HABA, until column color has completely changed back to white.

When using a new column, it can be directly washed with H_2O to remove the conservation solution and, then, conditioned with the corresponding buffer. All solutions that were loaded to the column were previously filtrated with a 0.45 µm filter.

The procedure for each loading-eluting cycle was the following, the total lyase volume was divided to load 10-12 ml each time. First, the strep-trap column was connected to the Äkta fraction collector (Äkta FPLC, Abersham Biosciences). 10 to 12 mL of clarified lysate were loaded at 2-3 mL/min and monitored with Äkta's UV absorbance detector. The column was washed with PBS until absorbance values reached the baseline and elution was performed by adding d-desthibiotin 2.5 mM in PBS. The fraction corresponding to the appearance of a peak in the absorbance measure was kept. The column was then washed with 2 volumes of PBS, regenerated with 5 volumes of HABA 1 mM in PBS and, finally, washed again with PBS to eliminate HABA.

A standard purification could include 3 to 5 cycles depending on the initial volume and on the quantity of protein observed in the elution steps.

After purification, the column was washed with H_2O and conserved in a 20% EtOH solution.

7.2.4. His-trap affinity chromatography

A his-tag can be found in the pET22b(+) plasmid containing *Ns*CDA, *Tc*CDA and *Bc*Pda genes, on the C-terminus end of proteins. This tag is used for their purification by affinity chromatography.

AFFINITY CHROMATOGRAPHY: HIS-TAG

Immobilized metal-affinity chromatography (IMAC) is a widely employed technic to purify recombinant proteins containing a short affinity tag consisting of polyhistidine residues. This method is based on the interactions between a transition metal ion (such as Co^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}, Co^{2+}) immobilized on a matrix and the side chains of certain amino acids, being histidine the one with the strongest interaction. Electron donor groups on the histidine imidazole ring form coordination bonds with the immobilized transition metal. According to this principle, peptides containing sequences of consecutive histidine residues are efficiently retained on IMAC column matrices and, after washing of the matrix material, can be easily eluted by either adding free imidazole to the column buffer or adjusting the pH its pH [143].

Supernatant from cell extract centrifugation was loaded to a 5 mL Ni^{2+} charged His-Trap chelating column connected to the Äkta fraction collector (Äkta FPLC, Abersham Biosciences). From this loading, all purification steps were monitored with Äkta's UV absorbance detector. The column was washed with PBS buffer (20 mM phosphate, 150 mM NaCl, pH 7) until absorbance values reached the baseline and elution was performed with an elution buffer (20 mM phosphate, 150 mM NaCl, 500 mM imidazole pH 7) gradient (from 0 to 100% in 60 minutes). The fractions corresponding to the appearance of a peak in the absorbance (corresponding to elution buffer concentration between 10 and 40%) was collected and kept. The column was then washed with PBS and H_2O and conserved in a 20% EtOH solution.

7.2.5. Size Exclusion Chromatography

SIZE EXCLUSION CHROMATOGRAPHY

Size exclusion chromatography (SEC) or gel filtration is a chromatographic technique used to separate molecules by their molecular weight or size while they pass through the column packed gel filtration medium. In contrast to affinity chromatography, molecules do not bind to this medium and the buffer composition does not directly affect to the resolution (separation between peaks). Consequently, an important advantage of the technique is that conditions can be changed to adjust to each sample or to the requirements of possible later purification or characterization steps.

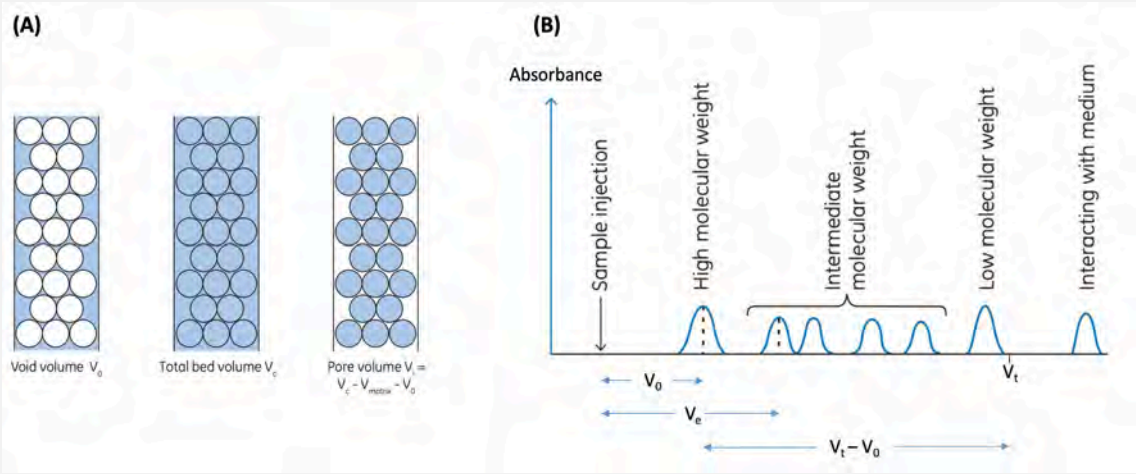


Figure 7.3. Size Exclusion Chromatography principle. *(A) Definition of column volumes. (B) Bigger molecules coming out of the column first, followed by smaller molecules by size order. The separation process is completed when one total column volume has passed through the gel filtration medium.*

Gel filtration medium is a porous matrix in the form of spherical particles that are chosen for their physical and chemical stability and because they are inert (without adsorption properties or reactivity). The medium is packed inside a column and it is equilibrated with a buffer that fills the matrix pores as well as the space between the particles. The liquid inside the pores is referred as stationary phase and it is in equilibrium with the liquid outside the particles (mobile phase). In this environment, the larger molecules will not be retained in the matrix particle's pores and their retention time in the column will be shorter. Contrary, the smaller the molecules are, the more it will take for them to leave the porous matrix in which they are retained (Figure 7.3).

7.2.5.1. Preparative chromatography

For preparative separation of different oligomeric fractions of the studied proteins in this work, a Superdex 200 16/600 column was used (GE Healthcare).
Before using it for purification and connecting it to the monitored fraction collector Äkta, the Superdex column was conditioned with a flow rate of ≤ 1 mL/min, adding 100 mL of H2O and 100 mL of the desired buffer (usually, PBS pH 7). All solutions that were loaded to the column were previously filtrated with a 0.45 μm filter.

All elution fractions from the previous affinity chromatography were pooled together and concentrated to a maximum volume of 2 ml with a 10 kDa MWCO Amicon Ultra-centrifugal filter (Merck Millipore). The concentrated protein was loaded to the column through a previously conditioned 2 ml loop (washed with H_2O and PBS) and elution was performed using the same buffer at 1 mL/min flow rate. The elution of the oligomeric fractions was monitored by online measure of 280 nm absorbance and 1 mL fractions were collected in deep-well microplates.

Once the last fraction was eluted, the column was washed with 100 mL of H_2O and 100 mL of 20% EtOH solution.

All fractions were conserved for their analysis by SDS-PAGE electrophoresis and, once purity and identity (by molecular mass determination) of the interest protein were determined, all aliquots of the desired oligomeric fraction were pooled together. In the case of all *Bs*PdaC variants, this fraction was the monomer, used for further protein characterization.

7.2.5.2. Column calibration

Calibration of the gel filtration column allows to precisely identify the apparent molecular weight of a protein (or any other compound loaded to the column) depending on the needed volume to move it through the column (elution volume).

In order to calibrate the preparative column, the dead volume (V_0) of the column was estimated and a calibration curve was obtained using standard proteins of known weight. The procedure, in which a set of standards from Sigma-Aldrich was used (Gel Filtration Markers Kit for Protein Molecular Weights 6,500-66,000 Da), include conditioning the column as previously described.

Then, a series of solution with the standard proteins suspended in PBS were prepared (Table 7.7) and filtered through a 0.45 µm filters.

***Table 7.7. Superdex 200 calibration**. Compounds and proteins used for calibration of preparative Superdex200, with their respective concentrations and number of injections.*

COMPOUND/PROTEIN	MW (KDA)	CONCENTRATION (mg/ml)	INJECTION
Dextran blue	2000	1.5	#1
Albumin	66	1	#2
Cytochrome C	12.4	1	
Carbonic Anhydrase	29	1	#3
Aprotinin	6.5	1	

The Superdex 200 column was connected to Äkta FPLC and 1 ml of each of the three solutions was loaded to the column through the injection loop in the order that is indicated in Table 7.7. The elution of proteins was performed at the same flow rate used for preparative separation, 1 mL/min, and it was monitored by online measuring of 280 nm absorbance.

Chromatograms data were extracted, and the maximum signal of each protein/compound was considered its elution volume (V_e).

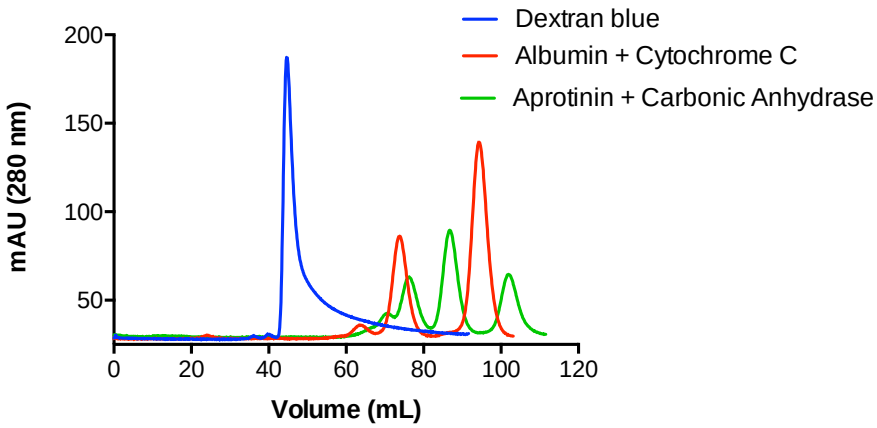


***Figure 7.4. Calibration chromatogram of the preparative Superdex 200 column.** Chromatograms resulting of elutions of the three injections with the reference proteins to calibrate the gel filtration column. The performed injections were: 1. Dextran blue; 2. Albumin and cytochrome C; 3. Aprotinin and carbonic anhydrase.*

A linear regression fit was performed, using Log(Mw) and V_e/V_0 values (Table 7.8 and Figure 7.5).

Table 7.8. Elution volumes for calibration of Superdex 200 column. *Elution times of the proteins and compounds used for the calibration of the gel filtration column and data transformation to obtain the standard curve.*

COMPOUND/PROTEIN	Mw (kDa)	Log(kDa)	V_e (mL)	V_e/V_0
Aprotinin	6.5	0.81	102.24	2.29
Cytochrome C	12.4	1.09	94.24	2.11
Carbonic Anhydrase	29	1.46	86.76	1.94
Albumin	66	1.82	73.83	1.65
Dextran blue (V_0)	2000		44.74	1

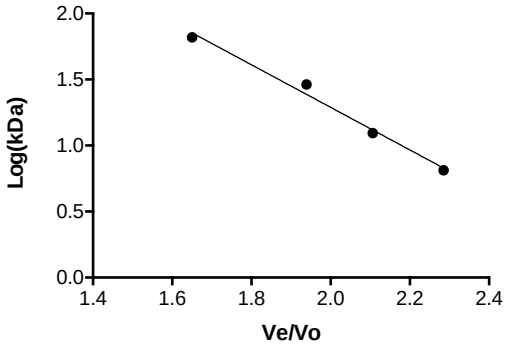


Figure 7.5. Standard curve of the Superdex 200 column calibration. *y = -1.6127x + 4.5148; R^2 = 0.987.*

The obtained equation was used to calculate the apparent Mw of the proteins or fractions that were separated using this methodology (Equation 7.1).

$$Mw = 10^{-1.6217\left(\frac{V_e}{V_0}\right)+4.5148} \quad \textbf{(}\textit{\textbf{Equation 7.1}}\textbf{)}$$

7.3. Biochemical characterization of *Bs*PdaC

7.3.1. Thermal stability determination

THERMAL STABILITY AND DIFFERENTIAL SCANNING FLUORIMETRY

T_m is defined as the temperature at which half of the protein population is in its denatured form and the other half in its native form and it is also used in the study of many other aspects such as ligands binding, protein-protein interactions or effect of denaturing agents. In terms of stability, the more thermally stable a protein is, the more energy is needed to destabilize its structure and, consequently, the higher is its T_m. When comparing melting temperatures, a positive shift in ΔT_m can be related to an increase in structural order and a reduced conformational flexibility, whereas a negative ΔT_m may indicate that induced structural changes lead to a more disordered and unstable protein conformation [144].

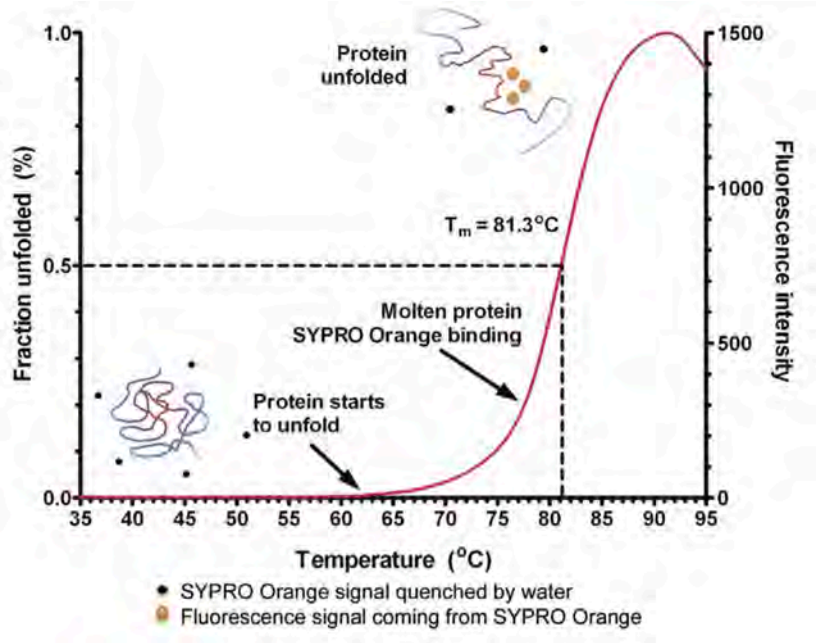


Figure 7.6. Thermal denaturation assay using Thermofluor. *Melting curve of example protein in which the protein solution is heated in presence of the hydrophobic dye Sypro Orange. From Boivin, 2013.*

Thermal Shift Assays (TSA) are methods in which the changes in the denaturing temperature of a certain protein, under different conditions, are measured. If these conditions remain constant, these assays are a fast method for the T_m comparison between different proteins. In 2004, the use of the fluorescent dye Sypro Orange in this methodology was described [122]. This compound specifically binds to hydrophobic surfaces and it can work as a denaturing probe in such a way that, when temperature increases and the protein reaches its denatured state, the hydrophobic areas of the hydrophobic core of the protein that were previously buried are exposed and become accessible to the probe (Figure 7.6). The methodology takes advantage of the increasing fluorescence of the dye when the hydrophobicity of its environment increases. It is quenched in aqueous solutions, but it regains its fluorescence when its aromatic moieties intercalate into the protein hydrophobic pocket. One of the main advantages of using this dye, besides its great sensitivity, is that wavelengths of both its excitation (483 nm) and its radiation emission (560 nm) are found in the majority of RT-PCR equipment. This fact facilitates the fluorescence monitoring of a large number of samples in a system that can efficiently control the temperature ramp of denaturalization. Once the method was established as a reference procedure for the study of denaturalization in proteins, it was named Differential Scanning Fluorimetry (DSF) [145].

Purified protein (*Bs*PdaC-FL or *Bs*PdaC-CD) at 4 mM concentration in PBS (50 mM phosphate, 300 mM NaCL, pH 7) was mixed with a dilution of the commercial dye Sypro Orange (Thermo Fisher Scientific; 1:500 dilution according to manufacturer's protocol) in a final volume of 25 µl.
The samples were subjected to a thermal gradient in a thermocycler (Rotogene 3000, Corbett Research), consisting of a 1 minute at 25°C followed by 1°C increments (30 seconds at each temperature) up to 95°C.
Fluorescence was measured (excitation wavelength = 483 nm, emission wavelength = 560 nm), and Tm was determined by fitting the data to a Boltzmann sigmoidal equation (Equation 7.2) using Prism (GraphPad Software, La Jolla, CA):

$$Fluor. = Fluor._{baseline} + \frac{Fluor._{superior} - Fluor._{baseline}}{1 + e^{(\frac{Tm - T)}{dT}}} \quad \textbf{(Equation 7.2)}$$

7.3.2. Metal dependency of *Bs*PdaC

In order to determine the effect of metal ions and EDTA on the enzyme, an unfolding-refolding procedure was set up.

Purified protein (*Bs*PdaC-CD) in PBS (50 mM phosphate, 300 mM NaCl, pH 7) was dialyzed against PBS with EDTA (200 mM) and urea (2 or 7 M). During the refolding step, EDTA and urea were removed by dialysis: first, with PBS (3 times, for 3 hours) and then with water (3 times, for 3 hours). Activity of the treated proteins and the native enzyme with and without the addition of Zn^{2+} was determined by HPLC-MS (see Section 7.4.2). Reactions were performed with 2 µM *Bs*PdaC-CD (treated or native), 2 mM $(GlcNAc)_4$, PBS (50 mM phosphate, 300 mM NaCl, pH 7) and with or without 1 mM Zn^{2+}.

The metal content of the apoenzyme (after the refolding step from 7 M urea) was determined by ICP-MS (NexION 300XX, PerkinElmer Life Sciences): results showed that $[Zn^{2+}]$ (µM)/[E] (µM) = 0.1, meaning that approximately 90% of the bound metal was removed following the described procedure.

7.3.3. Circular dichroism

In order to test the effect of ligands on *Bs*PdaC-CD thermal stability, ligand binding was assessed by monitoring the thermal unfolding by circular dichroism (CD) spectroscopy in the absence and presence of COS substrates.

A J-810 CD spectropolarimeter (Jasco Corp., Tokyo, Japan) equipped with a Peltier thermal device was used. Measurements were carried out in a range from 20 to 90 °C at 222 nm by using quartz cuvettes with a 1-mm optical path length. Temperature was increased stepwise at 1 °C/min. Samples were 10 µM pPR-IBA2-*Bs*PdaC-CD in 50 mM Tris buffer, 150 mM NaF, pH 7.5. The effect of substrate on the thermal stability was tested under the same conditions, using 1 mM $(GlcNAc)_2$, $(GlcNAc)_4$, and $(GlcNAc)_5$. Transitions were fitted according to a Boltzmann sigmoidal equation using the GraphPad software.

7.4. Kinetic characterization of enzymes

7.4.1. Activity estimation of SEC fractions with AcOMU

This methodology was used in this work mainly to quickly study the relative unspecific activity of each of the fractions from the gel filtration chromatography.

- A stock solution of AcOMU (4-methylumbelliferyl acetate) 50 mM (11 mg/ml) in Acetonitrile was prepared and kept in the dark.
- 50 mL of the buffer in which the protein is dissolved (usually PBS: 50 mM phosphate, 300 mM NaCl, pH 7) were preincubated at 37°C.
- Several dilutions of MU (between 0 and 0.5 mM) were prepared using the same buffer (standard curve).
- The AcOMU stock solution was diluted with the preincubated buffer to a final concentration of 0.5 mM.
- For each gel filtration fraction, 10 µl were mixed with 200 µL of the buffer with substrate solution and fluorescence was measured continuously (excitation wavelength = 340 nm, emission wavelength = 460 nm) both for each of these samples and for the standard curve dilutions.
- Product concentrations were calculated using the standard curve and reaction rates were determined considering time.
- Absorbance (280 nm) values from protein elution and relative rates values were superposed in order to identify which oligomeric fractions showed a higher activity.

7.4.2. Deacetylase activity determination by HPLC-MS

The formation of products with different degrees of acetylation from chitooligosaccharides substrates, $(GlcNAc)_{2\text{-}5}$, was monitored by HPLC-MS (Agilent 1260 HPLC-MS, electrospray ionization

(ESI +), single quadrupole detector) using a XBridge BEH Amide 2.5 µM, 3.0 x 100 mm XP column (Waters) in combination with an XBridge BEH Amide Guard Cartridge (2PK) precolumn (2.5 µM x 20 mm) (Waters).

Reaction were performed by incubating the 2 mM $GlcNAc_n$ substrate (n = 2-5) and 5µM purified enzyme (*Bs*PdaC-CD or *Tc*CDA) in PBS (phosphate 50 mM, NaCl 300 mM, pH 7) at 37°C in a final volume of 100-200 µL in microtiter plates.

At different time intervals, aliquots of 10 µl were withdrawn and added to 90 µl of water/1-propanol (1:1) to stop the reaction. 5 µL samples were injected into the system and eluted at 60 °C with acetonitrile/water (65:35, v/v), 1% formic acid at a flow rate of 0.4 µL/min. MS detection was performed on both SIM mode (for monitoring $[M + H]^+$ of substrate and deacetylation products, Tables 7.9 - 7.12) and SCAN mode (for total ion monitoring, 250–1100 m/z scan range).

***Table 7.9. m/z values monitored by HPLC-MS with $(GlcNAc)_2$ substrate**. m/z values corresponding to $[M + H]^+$ of substrate and deacetylation products. (A: GlcNAc; D: $GlcNH_2$).*

A2	A1D1	D2
425.4	383.4	341.4

***Table 7.10. m/z values monitored by HPLC-MS with $(GlcNAc)_3$ substrate**. m/z values corresponding to $[M + H]^+$ of substrate and deacetylation products. (A: GlcNAc; D: $GlcNH_2$).*

A3	A2D1	A1D2	D3
628.6	586.6	544.6	502.6

***Table 7.11. m/z values monitored by HPLC-MS with $(GlcNAc)_4$ substrate**. m/z values corresponding to $[M + H]^+$ of substrate and deacetylation products. (A: GlcNAc; D: $GlcNH_2$).*

A4	A3D1	A2D2	A1D3	D4
831.4	789.4	747.4	705.4	663.4

***Table 7.12. m/z values monitored by HPLC-MS with $(GlcNAc)_5$ substrate**. m/z values corresponding to $[M + H]^+$ of substrate and deacetylation products. (A: GlcNAc; D: $GlcNH_2$).*

A5	A4D1	A3D2	A2D3	A1D4	D5
1034.98	992.98	950.98	908.98	866.98	824.98

7.4.2.1. Specific activity

Specific activities for the initial monodeacetylation reactions with the same substrates were determined with the same procedure for *Bs*PdaC-CD, *Bs*PdaC-C.2 and all *Bs*PdaC-CD single mutants. Reactions were performed with different enzyme concentration (between 0.25 and 5 µM) and using the corresponding monodeacetylated product standard for quantification. The standards were prepared by reaction of the corresponding $(GlcNAc)_n$ substrate with *Vc*CDA, as previously reported [24], and several dilutions were prepared for each set of standards.

7.4.2.2. Michaelis- Menten kinetics

Michaelis-Menten parameters for both *Bs*PdaC-CD and *Bs*PdaC-CD-R315A were determined for the initial monodeacetylation reaction with $(GlcNAc)_4$ varying the substrate concentration from 0.5 to 17 mM with 3.7 µM *Bs*PdaC-CD or 4.3 µM *Bs*PdaC-CD-R315A in PBS (50 mM phosphate, 300 mM NaCl, pH 7) at 37°C. Kinetic parameters k_{cat}, K_M, and k_{cat}/K_M were calculated by nonlinear regression to the Michaelis–Menten equation.

7.4.3. Deacetylase activity determination by fluorescamine labeling

Enzyme activity of *Bs*PdaC-CD on COS substrates was also determined by fluorescamine labeling of the amino groups generated in the deacetylation reaction. All reactions were performed by incubating 2 mM $(GlcNAc)_{3-5}$ substrate and between 0.15 and 1.3 µM of purified enzyme in PBS (50 mM phosphate, 30 mM NaCl, pH 7) at 37°C in a final volume of 200 µL in microtiter plates.
At different time intervals, aliquots of 20 µL were withdrawn and added to 90 µL of water/1-propanol (1:1) to stop the reaction. Then 20 µL of 2 mg/ml fluorescamine in DMF were added, and the reaction was incubated for 10 minutes at room temperature. The labeling reaction was terminated by adding 150 µL of DMF/water (1:1) and fluorescence was quantified (excitation wavelength = 340 nm, emission wavelength = 460 nm).

A set of glucosamine standards between 0 and 0.5 mM in PBS (50 mM phosphate, 300 mM, pH 7) were prepared, treated following the same protocol and used to determine product concentration.

Determination of the pH and temperature profiles were performed following the same procedure. For determination of the pH optimum, the pH was varied from 6.0 to 10.6 at 37°C. For optimum temperature determination, the temperature was varied from 25 to 75°C at pH 7.

7.4.4. Deacetylase activity determination by acetate release determination

7.4.4.1. Specific activity on PGN

Enzyme activity toward *Bacillus subtilis* peptidoglycan (Sigma-Aldrich) was determined by monitoring the release of acetic acid with a commercial acetate determination kit (Acetic Acid Assay Kit, Megazyme). Reactions were performed by incubating the peptidoglycan substrate (1 mg/ml) and different concentrations of purified *Bs*PdaC-FL or *Bs*PdaC-CD (0.25 – 5 µM) in PBS substrate (50 mM phosphate, 300 mM NaCl, pH 7) at 37°C in shaking conditions. At different time intervals, the reaction samples were heated at 100°C for 5 minutes to inactivate the enzyme and then centrifuged for 2 minutes at 12.000xg. The supernatants were collected and used for measurement of the acetate released following the manufacturer's protocol.
A set of NADH standards between 0 and 0.1 mg/ml in PBS (50 mM phosphate, 300 mM, pH 7) were prepared, treated following the same protocol and used to determine product concentration. Specific activity was expressed as µM acetate/min/µM enzyme (specific activity in min^{-1}).

7.4.4.2. *Bs*PdaC-CD specificity on PGN

To asses that the isolated catalytic domain maintains the same MurNAc deacetylase activity as the full-length enzyme, end-point reactions with both *Bs*PdaC-CD and *Bs*PdaC-FL were set up, and the release of acetic acid was determined as described above. Reaction mixtures with 0.5 µM enzyme, 1 mg/ml PGN in PBS (50 mM phosphate, 300 mM NaCl, pH 7) were incubated at 37°C with agitation for 48 hours. Then samples were dialyzed against PBS buffer to remove the free acetate. *Bs*Pdac-CD was added to the sample of the initial *Bs*PdaC-FL reaction, and *Bs*PdaC-FL was added to the sample of the initial *Bs*PdaC-CD reaction. After 48 hours incubation at 37°C, the released acetate was determined. Finally, the GlcNAc peptidoglycan deacetylase *Sp*PgdA was added to the final reactions at 0.8 µM concentration and acetate release was monitored as described.
As reference, enzymatic activity of *Sp*PgdA on *Bacillus subtilis* peptidoglycan was determined following the same procedure than with *Bs*PdaC.

7.4.4.3. *Bs*PdaC-FL single mutants: activity on $(GlcNAc)_4$

Enzymatic activity of *Bs*PdaC-FL and the single mutants prepared from this template (D285, W395, D398 and W402) was evaluated monitoring the release of acetic acid with the same commercial acetate determination kit and using the same set of NADH standards for product quantification. Reactions mixtures were set up in microtiter plates with several concentrations of enzymes

(between 0.25 and 7 µM), 2 mM $(GlcNAc)_4$ substrate, PBS (50 mM phosphate, 300 mM NaCl, pH 7). Microplates were incubated at 37°C in an absorbance microplate reader (BioTek) and absorbance at 340 nm was measured continuously each 30 seconds.

7.4.5. Determination of the deacetylation pattern

To determine the pattern of acetylation of the different products formed during the time course monitoring, preparative reactions with 2 mM $(GlcNAc)_4$ and $(GlcNAc)_5$ substrates and 0.1 mg of purified *Bs*PdaC-CD in PBS, pH 7.0, at 37 °C, in a final volume of 1 ml, were incubated for different reaction times. The freeze-dried samples were analyzed for pattern of acetylation by MS(MALDI-TOF-MS/MS analysis) after previous ^{18}O-labeling and re-*N*-acetylation with [$^2H_6$]acetic anhydride according to the protocol by Cord-Landwehr *et al*. [123], and the relative amount of each sequence in the mixtures was calculated based on the peak intensities compared for each ion type.

7.5. Bioinformatics

7.5.1. Initial sequence alignment and phylogenetic analysis of CE4 enzymes

A bioinformatic study to compare all sequences from Table 4.1 was performed. First, a structural guided alignment was built with the Multiseq tool from VMD software [146], guided by the structural superposition of all available x-ray structures, retrieved from PDB database. After manually curating the alignment to include only the CE4 domain of the enzymes, a Hidden Markov Model profile was built using the hmmbuild tool from HMMER [48]. Sequences from enzymes without known 3D structure were incorporated into the multiple-sequence alignment by Hidden Markov Model comparison, obtaining the final alignment that included all sequences from Table 4.1.

The phylogenetic relationships were inferred by using the Neighbor-Joining (NJ) algorithm based on the JTT matrix-based model [147] computed from the multiple sequence alignment. The evolutionary analyses were conducted with MEGA7 software [148], with a bootstrap analysis that consisted of 500 replicates.

7.5.2. Phylogenetic study of chitin deacetylases and related proteins

The HMM profile from the multiple sequence alignment with the reference enzymes listed in Table 5.1 (alignment-1) was built using the hmmbuild tool from HMMER [48] (HMM-1) and used to search against the Uniprot database (SwissProt and Trembl) with no specific conditions. After redundancy

reduction, an alignment with 17942 sequences were aligned against HMM-1 with HMMER and a multiple sequence alignment with all new sequences from the database search and the sequences of the reference proteins was generated and manually curated to include only the aligned region corresponding to the conserved catalytic domain of the family (alignment-2). A phylogenetic tree from alignment-2 was built with the FastTree software using the Neighbor-joining algorithm.

One branch of the tree was selected for each of the 12 reference proteins and rooted by that one. Inside each branch, sub-branches were clustered by loop size according to alignment-2. Loops of each sub-branch were identified considering their assigned location on alignment-1, then each loop was extracted and realigned with with Jalview [149] and the MAFFT algorithm (alignment-3) to generate a new HMM and logo with the web tool Skylign (HMM-2).

After performing the analysis of all sequences and choosing 29 of them for the final selection, those sequences were aligned against a HMM including known CDAs and CODs (HMM-3). The resulting alignment (alignment-4, Annex 9.4) was used to re-assign loop size of the 29 sequences. The phylogenetic relationships from alignment-4 were inferred by using the Neighbor-Joining (NJ) algorithm based on the JTT matrix-based model [147], using MEGA7 software [148], with a bootstrap analysis that consisted of 500 replicates.

The multiple sequence alignments and Hidden Markov Model profiles from this Section have been numbered and a detailed description of them can be found bellow (Table 7.13).

Table 7.13. Description of the sequences included in HMM profiles and multiple sequence alignments generated during the bioinformatics and phylogenetic study of CDAs and related proteins.

HIDDEN MARKOV MODEL PROFILES	
HMM-1	Profile built from the 12 characterized members of CE4 family initially selected as reference sequences (*hmm build, HMMER)*
HMM-2	For each selected sub-branch of the phylogenetic tree, an HMM logo was generated for each of the 6 loops. Profiles were built from the corresponding alignment-3 (*Skyling*)
HMM-3	Profile built from the list of known CDAs and CODs extracted from Grifoll-Romero, 2018 (*hmm build, HMMER*).
HMM-4	Profile built from the CE4 enzymes with solved x-ray structure or characterized activity on COS (known pattern of deacetylation on chitooligosaccharides or reported in vitro characterization on these substrates) (*hmm build, HMMER*)

MULTIPLE SEQUENCE ALIGNMENTS	
Alignment-1	Multiple sequence alignment with the 12 characterized members of CE4 family initially selected as reference sequences (*Jalview, Mafft G-INS-I, accuracy oriented*).
Alignment-2	Structural multiple sequence alignment with the set of 17492 sequences retrieved from the initial search against the Uniprot database, guided by HMM-1. The alignment also includes the 12 initial reference sequences (*hmm align*).
Alignment-3	For each selected sub-branch of the phylogenetic tree, an alignment for each of the 6 loops was generated. The sequences of each loop were selected from alignment-1, extracted and re-aligned (*Jalview, Mafft G-INS-I, accuracy oriented*).
Alignment-4	Multiple sequence alignment with the 29 selected sequences and known CDAs and CODs (included in HMM-3) (*Jalview, Mafft G-INS-I, accuracy oriented*).

7.5.3. Modeling of *Bs*PdaC, *Bs*PdaA and *Sp*PgdA structures in complex with substrates

The 3D structures of the enzyme-ligand complexes of *Sp*PgdA, *Bs*PdaA, and *Bs*PdaC were modeled with AUTODOCK 4.2 and AUTODOCK VINA [150], [151]. Protein structures were directly taken from the PDB database. Ligand structures covering the different range of accepted substrates were extracted from other protein structures with PDB accession codes 6AVE (GlcNAc), 1TWQ (MurNAc), 9LYZ (GlcNAcß(1,4)MurNAc and MurNAcß(1,4)GlcNAc), and 1LZC (tetraacetylchitotetraose). Both the protein and ligand structures were first parametrized with AutoDockTools [150]: polar hydrogens were added, AutoDock4.2 atom typing was used, and Gaisteger partial charges were computed. All rotable bonds of the ligands were considered free during the docking calculations, except amide and ring bonds, whereas the whole protein structure was kept fixed.

Tetraacetylchitotetraose, $(GlcNAc)_4$, was docked to the *Bs*PdaC structure with AUTODOCK VINA [151]. The docking search space was confined in a 19.5 x 27.75 x 36.75 $Å^3$ box centered in the active site. Exhaustiveness level was set to 24, and 20 binding modes of the ligands were generated. Only low-energy binding poses were considered for analysis. Distances and surfaces were measured, and pictures of the complexes were generated with VMD [146].

Docking of mono- and disaccharides with AUTODOCK 4.2 was used as a probe for the discovery of putative binding subsites along the enzymatic cavity of *Sp*PgdA, *Bs*PdaA, and *Bs*PdaC structures. A grid box of 31.5 x 25.5 x 28.5 $Å^3$ with a grid point spacing of 0.375 Å, centered at the active site, was used as the search space for docking. For each ligand docking, 1000 rounds of the genetic algorithm implemented in AUTODOCK 4.2 were performed. For each round, an initial population of 150 members was considered, with randomized initial position and orientation coordinates and

randomized conformations of the substrate flexible bonds. The genetic algorithm was extended up to 27,000 off-spring generations, with a maximum of 2,500,000 energy evaluations. Only low-energy and repetitive binding poses were considered for analysis.

7.6. CRYSTALLOGENESIS

Crystallogenesis and resolution of the X-ray structure was done by Maria Ángela Sanz Polo in Marcelo Guerin's group (Structural Biology lab) at CIC bioGUNE.

7.6.1. *Bs*PdaC-CD crystallization and data collection

Crystals of *Bs*PdaC-CD (WT and D285S mutant) were obtained by mixing 0.25 µl of the protein in 50 mM Tris, pH 7, and 150 mM NaCl at 5.6 or 10 mg/ml, respectively, with 0.25 µl of a mother liquor containing 0.2 M ammonium tartrate dibasic, pH6.6, 20% (w/v) PEG 3.350 or 0.4 M ammonium phosphate monobasic by sitting-drop vapor-diffusion crystallization. Crystals appeared after 8–10 days and grew as prisms, reaching 0.1-0.8 x 0.5-0.4 x 0.05-0.07 mm. Single crystals of *Bs*PdaC-CD and *Bs*PdaC-CD-D285S were cryocooled in liquid nitrogen by using a cryoprotectant solution of the mother liquor supplemented with 15% (v/v) ethylene glycerol or 30% (v/v) glycerol, respectively.

X-ray diffraction data were collected on a PILATUS 6MF pixel array detector at the microfocus I02 beamline (λ = 0.9796 Å, Diamond Light Source) and integrated with XDS [152] following standard procedures.

7.6.2. Structure determination and refinement

Structures of *Bs*PdaC-CD and *Bs*PdaC-CD-D285S were solved using as a template the previously reported *S. pneumoniae* peptidoglycan deacetylase *Sp*PgdA (PDB code 2C1G) and molecular replacement methods implemented in Phaser [153] and the PHENIX suite [154]. Model rebuilding was carried out with Buccaneer [155] and the CCP4 suite [156]. The final manual building was performed with Coot [157] and refinement with phenix.refine [154]. The structures were validated by MolProbity [158]. Data collection and refinement statistics are presented in Annex 9.3. Atomic coordinates and structure factors have been deposited with the Protein Data Bank, accession codes 6H8L and 6H8N corresponding to *Bs*PdaC-CD and *Bs*PdaC-CD-D285S crystal structures, respectively. Molecular graphics and structural analyses were performed with the UCSF Chimera package [159].

7.6.3. *Bs*PdaC-C2 crystallogenesis conditions

After expressing and purifying recombinant *Bs*PdaC-C2 as described, buffer exchange and protein were performed with the monomeric fraction of the enzyme. Two screening plates were set up, testing conditions from commercial crystallization reagent kits (PEG/Ion Screen™ and PEG/Ion 2Screen™ from Hampton Research, and Morpheous® HT-96 MD1-47 from Molecular Dimensions). Screening was performed by mixing 0.25 µl of the protein at 5 mg/ml with 1 mM DP4 in 20 mM TRIS, pH 7, and 150 mM NaCl, with 0.25 µl of the mother liquor by the sitting-drop vapor-diffusion technique. Little crystals were obtained in two of the screened conditions:

1. 0.3 M sodium nitrate, 0.3 M sodium phosphate dibasic, 0.3 M ammonium sulfate, 0.1 M **Buffer System 1 (imidazole, MES monohydrate)**, **pH 6.5**, 50% (v/v) Precipitant Mix 1 (40% v/v PEG 5MM MME, 20% v/v PEG 20000).
2. 0.3 M sodium nitrate, 0.3 M sodium phosphate dibasic, 0.3 M ammonium sulfate, 0.1 M **Buffer System 2 (sodium HEPES, MOP)**, **pH 7.5**, 50% (v/v) Precipitant Mix 1 (40% v/v PEG 5MM MME, 20% v/v PEG 20000)

Taking those conditions as a basis, an optimization plate was set up, exploring a pH range between 6 and 8 and a Precipitant Mix 1 concentration between 30 and 70% (v/v). 0.25 µl of *Bs*PdaC-C2 at 5.7 mg/ml (with and without 1 mM DP4) in 20 mM TRIS, pH 7, and 150 mM NaCl were mixed with 0.25 µl of the mother liquor by the sitting-drop vapor-diffusion technique.

7.7. Basic molecular biology and biochemistry protocols

7.7.1. Competent cells preparation and transformation

This protocol was used to obtain chemical competent cells for both DNA subcloning (*E. coli* DH5α, Invitrogen) and protein expression (*E. coli* BL21(DE3) star, Invitrogen).

50 mL of LB were inoculated with target cells and incubated at 37°C and 250 rpm until $0D_{600}$ reached a 0.6 value. Then, cells were kept on ice for 20 minutes and centrifuged for 5 minutes at 5000xg and 4°C. Supernatant was removed by decanting and cells were resuspended in 25 mL of chilled sterile 50 mM $NaCl_2$. A second centrifugation step in the same conditions was performed and cells were resuspended in 5 mL of chilled sterile 50 mM $NaCl_2$ after removing supernatant. 750 µL of sterile glycerol were added to cell suspension (final glycerol concentration of 15%) and, finally, aliquots of 200 µL of these cells were transferred to sterile Eppendorf tubes and stored at -80°C until needed.

For transformation, between 2 and 10 µl of DNA were added to competent cells after thawing them on ice. Several controls were always set up together with the targeted DNA:

- As a negative control of contamination, sterile water was added to competent cells, which at the final step were plated in medium with antibiotic.
- As a positive control of viability of cells, competent cells were plated in medium without antibiotic.
- As positive control of transformation, a plasmid with known transformation efficacy was transformed.

Cells with DNA/water were kept for 30 minutes on ice, then incubated for 2 minutes at 42°C and transferred to ice for 5 minutes. 500 µL of LB medium were added and cells were incubated for 45 minutes on a shaker at 250 rpm and 37°C. After that time, cells were plated to plates with the desired medium and antibiotic.

7.7.2. DNA obtention and purification

For plasmidic DNA extraction, different commercial miniprep kits were used (Sigma Aldrich; Clinisciences). Protocol provided by each manufacturer was followed after growing an overnight culture of the transformed strain with the interest plasmid in 5 mL of LB medium with the required antibiotic.

For purification of DNA fragments or specific bands from an agarose gel, the GenElute™ Gel Extraction Kit (Sigma Aldrich) was used. The protocol supplied by the manufacturer was followed after excising the desired band or bands from 1% agarose gel.

For purification of PCR products, the QIAquick spin – PCR purification Kit (Qiagen) was used, following the protocol provided by the manufacturer.

7.7.3. DNA quantification

The methodology used in this work to quantify DNA was the Qubit® system (Thermo Fisher), based on target-selective dyes that emit fluorescence when bound to DNA. Two different kits were used, depending on the expected concentration of DNA of the sample (high sensitivity, to quantify samples between 10 and 100 ng/µL; and broad range, to quantify samples between 100 and 1000 ng/µL).

The protocol provided by the manufacturer was followed, and DNA concentration was directly provided by the specific fluorometer.

7.7.4. DNA digestion

7.7.4.1. DpnI digestion

DNA was digested with DpnI in order to eliminate the methylated DNA used a template in mutagenic PCRs. Restriction reactions were performed in a final volume of 20 µL, with 1 µL of DpnI-HF (New England Biolas), 2 µL of 10x Cutsmart Buffer and up to 16 µL of DNA sample. Reactions were incubated for 2-3 hours at 37°C.

7.7.4.2. Double digestion

Double restrictions were performed for subcloning of desired protein genes or fragments in expression vectors. All restriction enzymes used with this purpose were from New England Biolabs and presented 100% activity on the provided CutSmart buffer, thus, sample mixtures were prepared with 1 µL of each restriction enzyme, 2 µL of buffer and up to 16 µL of DNA in a final volume of 20 µL. Reactions were incubated for 2-3 hours at 37°C.

7.7.5. DNA ligation

Ligation was performed with T4 DNA ligase (New England Biolabs). Several reactions were set up, ensuring different stoichiometric vector:insert ratios (usually, 1:3, 1:5 and 1:7) and adding 100 ng of DNA, 1 µL of T4 ligase and 2 µL of T4 ligase buffer in a final volume of 20 µL.

Several negative controls were always set up, in addition to the ligation reactions:

- As a control to discard re-ligation: vector with ligase, without insert.
- As a control to ensure that the previous restriction of the vector was performed correctly: just vector, without insert and without ligase.
- As control of possible insert contamination: insert with ligase, without vector.
- As control of possible contamination with another plasmid: just ligase, without any DNA.

Reaction mixtures were incubated at room temperature for 15 minutes. *E. coli* DH5α competent cells were transformed with the ligation products. DNA was extracted with a miniprep kit and its sequence confirmed by Sanger Sequencing.

7.7.2. Protein quantification by BCA

Protein concentration was determined by the BCA method (Thermo Scientific) and Bovine Serum Albumin (BSA) was used as protein reference. Several dilutions of BSA (from 0 to 2 mg/mL in the same buffer in which the interest protein was dissolved) were prepared as standard curve. The protocol indicated by the manufacturer was followed for different dilutions of protein samples and for the standard curve.

20 µL of each sample (different dilutions of the interest protein and standard curve samples) was mixed with 160 µL of the working reagent (mixture described by the manufacturer's protocol) and incubated for 30 minutes at 37°C in a microtiter plate. Absorbance at 595 nm was measured with a microplate reader.

REFERENCES

8. REFERENCES

[1] A. Stryjewska, K. Kiepura, T. Librowski, and S. Lochyński, "Biotechnology and genetic engineering in the new drug development. Part III. Biocatalysis, metabolic engineering and molecular modelling.," *Pharmacol. Rep.*, vol. 65, no. 5, pp. 1102–11, 2013.

[2] L. Hilterhaus, A. Liese, U. Kettling, and G. Antranikian, *Applied biocatalysis: From fundamental science to industrial applications*. Wiley-VCH verlag GmbH & Co. KGaA., 2016.

[3] R. K. Bornscheuer UT1, G.W. Huisman, R.J. Kazlauskas, S. Lutz, and J.C. Moore, "Engineering the third wave of biocatalysis.," *Nature*, vol. 485, no. 7397, pp. 185–194, 2012.

[4] A. Illanes, *Enzyme Biocatalysis. Principles and Applications*. Springer, 2008.

[5] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*. Garland Science, 2002.

[6] D. Nelson, A. Lehninger, and M. Cox, *Lehninger principles of biochemistry*. Macmillan, 2008.

[7] J.M. Berg, J.L. Tymoczko, and L. Stryer, *Biochemistry*, *5th Edition.* W. H. Freeman, 2002.

[8] D. Voet, J.G. Voet, and C.W. Pratt, *Fundamentals of Biochemistry: Life at the Molecular Level, 5th Edition*. Wiley, 2016.

[9] O. Kirk, T.V. Borchert, and C.C. Fuglsang, "Industrial enzyme applications," *Curr. Opin. Biotechnol.*, vol. 13, no. 4, pp. 345–351, 2002.

[10] R.A. Dwek, "Glycobiology: Toward Understanding the Function of Sugars," *Chem. Rev.*, vol. 96, pp. 683-720, 1996.

[11] A. Helenius and M. Aebi, "Intracellular Functions of N-Linked Glycans," *Science*, vol. 291, no. 5512, pp. 2364–2369, 2001.

[12] D.H. Dube and C.R. Bertozzi, "Glycans in cancer and inflammation - Potential for therapeutics and diagnostics," *Nat. Rev. Drug Discov.*, vol. 4, no. 6, pp. 477–488, 2005.

[13] F. Schwarz and M. Aebi, "Mechanisms and principles of N-linked protein glycosylation," *Curr. Opin. Struct. Biol.*, vol. 21, no. 5, pp. 576–582, 2011.

[14] B. Adamczyk, T. Tharmalingam, and P. M. Rudd, "Glycans as cancer biomarkers," *Biochim. Biophys. Acta - Gen. Subj.*, vol. 1820, no. 9, pp. 1347–1353, 2012.

[15] A. Varki, "Biological roles of glycans," *Glycobiology*, vol. 27, no. 1, pp. 3-49, 2017.

[16] D.B. Werz, "Handbook of Chemical Glycosylation. Advances in Stereoselectivity and Therapeutic Relevance," *Angew. Chemie - Int. Ed.*, vol. 47, no. 29, pp. 5281–5282, 2008.

[17] B.I. Cantarel, P.M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, and B. Henrissat, "The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics," *Nucleic Acids Res.*, vol. 37, Database issue, pp. D233-D238, 2009.

[18] A. Planas, M. Faijes, and V. Codera, “When Enzymes Do It Better: Enzymatic Glycosylation Methods,” in *Carbohydrate Chemistry: State of the Art and Challenges for Drug Development*. Imperial College Press, pp. 215-245, 2015.

[19] V. Lombard, H. Golaconda Ramulu, E. Drula, P.M. Coutinho, and B. Henrissat, “The carbohydrate-active enzymes database (CAZy) in 2013,” *Nucleic Acids Res.*, vol. 42, Database issue, pp. D490-D49, 2014.

[20] Y. Zhao, R. D. Park, and R. a Muzzarelli, “Chitin deacetylases: Properties and applications,” *Mar. Drugs*, vol. 8, no. 1, pp. 24–46, 2010.

[21] H.P. Spaink, A.H. M. Wijfjes, K.M.G.M. Drift, J. Haverkamp, J.E. Thomas-Oates, and B.J.J. Lugtenberg, “Structural identification of metabolites produced by the NodB and NodC proteins of Rhizobium leguminosarum,” *Mol. Microbiol.*, vol. 13, no. 5, pp. 821–831, 1994.

[22] M. John, H. Rohrig, J. Schmidt, U. Wieneke, and J. Schell, “Rhizobium NodB protein involved in nodulation signal synthesis is a chitooligosaccharide deacetylase.,” *Proc. Natl. Acad. Sci.*, vol. 90, no. 2, pp. 625–629, 1993.

[23] A. B. Boraston, D. N. Bolam, H. J. Gilbert, and G. J. Davies, “Carbohydrate-binding modules: fine-tuning polysaccharide recognition,” *Biochem. J.*, vol. 382, pp. 769-781, 2004.

[24] E. Andrés, D. Albesa-Jové, X. Biarnés, B.M. Moerschbacher, M.E. Guerin, and A. Planas, “Structural basis of chitin oligosaccharide deacetylation,” *Angew. Chemie - Int. Ed.*, vol. 53, no. 27, pp. 6882–6887, 2014.

[25] D.E. Blair, A.W. Schüttelkopf, J.I. MacRae, and D.M.F. van Aalten, “Structure and metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal virulence factor,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 43, pp. 15429–15434, 2005.

[26] D.J. Little, J. Poloczek, J.C. Whitney, H. Robinson, M. Nitz, and P.L. Howell, “The structure- and metal-dependent activity of *Escherichia coli* PgaB provides insight into the partial de-N-acetylation of poly-β-1,6-N-acetyl-D-glucosamine,” *J. Biol. Chem.*, vol. 287, no. 37, pp. 31126–31137, 2012.

[27] A.M. Nakamura, A.S. Nascimento, and I. Polikarpov, “Structural diversity of carbohydrate esterases,” *Biotechnol. Res. Innov.*, vol. 1, no. 1, pp. 35–51, 2017.

[28] S. Arnaouteli *et al.*, “Two putative polysaccharide deacetylases are required for osmotic stability and cell shape maintenance in *Bacillus anthracis*,” *J. Biol. Chem.*, vol. 290, no. 21, pp. 13465–13478, 2015.

[29] R.J. Strunk *et al.*, “Structure determination of BA0150, a putative polysaccharide deacetylase from *Bacillus anthracis*,” *Acta Crystallogr. Sect. FStructural Biol. Commun.*, vol. 70, no. 2, pp. 156-159, 2014.

[30] V.E. Fadouloglou *et al.*, "Structure determination through homology modelling and torsion-angle simulated annealing: Application to a polysaccharide deacetylase from *Bacillus cereus*," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 69, no. 2, pp. 276–283, 2013.

[31] V.E. Fadouloglou *et al.*, "Unusual α-Carbon Hydroxylation of Proline Promotes Active-Site Maturation," *J. Am. Chem. Soc.*, vol. 139, no. 15, pp. 5330–5337, 2017.

[32] P. Giastas *et al.*, "Structures of the Peptidoglycan *N* -Acetylglucosamine Deacetylase Bc1974 and Its Complexes with Zinc Metalloenzyme Inhibitors," *Biochemistry*, vol. 57, no. 5, pp. 753–763, 2018.

[33] M.D. Dong, J.E. Urch, J.M. Ten Cate, V.A. Rao, D.M.F. Van Aalten, and W. Crielaard, "*Streptococcus mutans* SMU.623c codes for a functional, metal-dependent polysaccharide deacetylase that modulates interactions with salivary agglutinin," *J. Bacteriol.*, vol. 91, no. 1, pp. 394–402, 2009.

[34] D.E. Blair and D.M.F. Van Aalten, "Structures of *Bacillus subtilis* PdaA, a family 4 carbohydrate esterase, and a complex with N-acetyl-glucosamine," *FEBS Lett.*, vol. 570, no. 1–3, pp. 13–19, 2004.

[35] L. Oberbarnscheidt, E.J. Taylor, G.J. Davies, and T.M. Gloster, "Structure of a carbohydrate esterase from *Bacillus anthracis*," *Proteins Struct. Funct. Bioinforma.*, vol. 66, no. 1, pp. 250–252, 2006.

[36] D.J. Little, N.C. Bamford, V. Pokrovskaya, H. Robinson, M. Nitz, and P.L. Howell, "Structural basis for the De-N-acetylation of poly-β-1,6-N-acetyl-D-glucosamine in gram-positive bacteria," *J. Biol. Chem.*, vol. 289, no. 52, pp. 35907–35917, 2014.

[37] D.J. Little *et al.*, "The protein BpsB is a poly-β-1,6-N-acetyl-D-glucosamine deacetylase required for biofilm formation in *Bordetella bronchiseptica*," *J. Biol. Chem.*, vol. 290, no. 37, pp. 22827–22840, 2015.

[38] A. Chibba, J. Poloczek, D.J. Little, P.L. Howell, and M. Nitz, "Synthesis and evaluation of inhibitors of *E. coli* PgaB, a polysaccharide de-N-acetylase involved in biofilm formation," *Org. Biomol. Chem.*, vol. 10, no. 35, p. 7103, 2012.

[39] T. Nishiyama, H. Noguchi, H. Yoshida, S.Y. Park, and J.R.H. Tame, "The structure of the deacetylase domain of *Escherichia coli* PgaB, an enzyme required for biofilm formation: A circularly permuted member of the carbohydrate esterase 4 family," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 69, no. 1, pp. 44–51, 2013.

[40] E.J. Taylor *et al.*, "Structure and activity of two metal ion-dependent acetylxylan esterases involved in plant cell wall degradation reveals a close similarity to peptidoglycan deacetylases," *J. Biol. Chem.*, vol. 281, no. 16, pp. 10968–10975, 2006.

[41] Z. Liu *et al.*, "Structure and function of a broad-specificity chitin deacetylase from *Aspergillus nidulans* FGSC A4," *Sci. Rep.*, vol. 7, no. 1, art. 1746, 2017.

[42] D.E. Blair *et al.*, "Structure and Mechanism of Chitin Deacetylase from the Fungal Pathogen *Colletotrichium lindemuthanium*," *Biochemistry*, vol. 45, pp. 9416–9426, 2006.

[43] T.R. Tuveng, U. Rothweiler, G. Udatha, G. Vaaje-Kolstad, A. Smalås, and V.G.H. Eijsink, "Structure and function of a CE4 deacetylase isolated from a marine environment," *PLoS One*, vol. 12, no. 11, p. e0187544, 2017.

[44] L. Liu *et al.*, "Biochemical characterization of three midgut chitin deacetylases of the Lepidopteran insect *Bombyx mori*," *J. Insect Physiol.*, vol. 113, pp. 42-48, 2019.

[45] T. Hirano, K. Sugiyama, Y. Sakaki, W. Hakamata, S.Y. Park, and T. Nishio, "Structure-based analysis of domain function of chitin oligosaccharide deacetylase from *Vibrio parahaemolyticus*," *FEBS Lett.*, vol. 589, no. 1, pp. 145-151, 2015.

[46] J.E. Urch *et al.*, "Structural and functional characterization of a putative polysaccharide deacetylase of the human parasite *Encephalitozoon cuniculi*," *Protein Sci.*, vol. 18, no. 6, pp. 1197–1209, 2009.

[47] H. Aragunde, X. Biarnés, and A. Planas, "Substrate recognition and specificity of chitin deacetylases and related family 4 carbohydrate esterases," *Int. J. Mol. Sci.*, vol. 19, no. 2, 412, 2018.

[48] S.R. Eddy, "Profile hidden Markov models," *Bioinformatics*, vol. 14, no. 9, pp. 755–763, 1998.

[49] T. Fukushima, T. Kitajima, and J. Sekiguchi, "A polysaccharide deacetylase homologue, PdaA, in *Bacillus subtilis* acts as an N-acetylmuramic acid deacetylase in vitro," *J. Bacteriol.*, vol. 187, no. 4, pp. 1287–1292, 2005.

[50] Y. Araki, T. Nakatani, H. Hayashi, and E. Ito, "Occurrence of non-N-substituted glucosamine residues in lysozyme-resistant peptidoglycan from *Bacillus cereus* cell walls," *Biochem. Biophys. Res. Commun.*, vol. 42, no. 4, pp. 691–697, 1971.

[51] G.F. Zipperle Jr., J.W. Ezzell Jr., and R.J. Doyle, "Glucosamine substitution and muramidase susceptibility in *Bacillus anthracis*," *Can. J. Microbiol.*, vol. 30, no. 5, pp. 553–559, 1984.

[52] A.J. Wolf and D.M. Underhill, "Peptidoglycan recognition by the innate immune system," *Nat. Rev. Immunol.*, vol. 18, no. 4, pp. 243–254, 2018.

[53] W. Vollmer, "Structural variation in the glycan strands of bacterial peptidoglycan," *FEMS Microbiol. Rev.*, vol. 32, no. 2, pp. 287–306, 2008.

[54] P.T. Mckenney, A. Driks, and P. Eichenberger, "The *Bacillus subtilis* endospore: Assembly and functions of the multilayered coat," *Nat. Rev. Microbiol.*, vol. 11, no. 1, pp. 33–44, 2013.

[55] M.E. Gilmore, D. Bandyopadhyay, A. M. Dean, S. D. Linnstaedt, and D. L. Popham,

"Production of Muramic δ-Lactam in *Bacillus subtilis* Spore Peptidoglycan," *J. Bacteriol.*, vol. 186, no. 1, pp. 80–89, 2004.

[56] H.J. Rogers, H. R. Perkins, J. B. Ward, and B. Ward, *Microbial Cell Walls and Membranes*. London: Chapman and Hall, 1980.

[57] W. Vollmer, D. Blanot, and M.A. De Pedro, "Peptidoglycan structure and architecture," *FEMS Microbiol. Rev.*, vol. 32, no. 2, pp. 149–167, 2008.

[58] K.H. Schleifer and O. Kandler, "Peptidoglycan types of bacterial cell walls and their taxonomic implications.," *Bacteriol. Rev.*, vol. 36, no. 4, pp. 407–77, 1972.

[59] A.J.F. Egan, R.M. Cleverley, K. Peters, R.J. Lewis, and W. Vollmer, "Regulation of bacterial cell wall growth," *FEBS J.*, vol. 284, no. 6, pp. 851–867, 2017.

[60] G. Patti, J. Chen, J. Schaefer, and M. Gross, "Characterization of Structural Variations in the Peptidoglycan of Vancomycin-Susceptible *Enterococcus faecium*: Understanding Glycopeptide–Antibiotic Binding Sites Using Mass Spectrometry," *J. Am. Soc. Mass Spectrom.*, vol. 19, no. 10, pp. 1467–1475, 2008.

[61] D. Sychantha, A.S. Brott, C.S. Jones, and A.J. Clarke, "Mechanistic Pathways for Peptidoglycan O-Acetylation and De-O-Acetylation," *Front. Microbiol.*, vol. 9, art. 2332, 2018.

[62] S. Dramsi, S. Magnet, S. Davison, and M. Arthur, "Covalent attachment of proteins to peptidoglycan," *FEMS Microbiol. Rev.*, vol. 32, no. 2, pp. 307–320, 2008.

[63] F.C. Neuhaus and J. Baddiley, "A Continuum of Anionic Charge: Structures and Functions of D-Alanyl-Teichoic Acids in Gram-Positive Bacteria," *Microbiol. Mol. Biol. Rev.*, vol. 67, no. 4, pp. 686–723, 2003.

[64] P.J. Moynihan, D. Sychantha, and A. J. Clarke, "Chemical biology of peptidoglycan acetylation and deacetylation," *Bioorg. Chem.*, vol. 54, pp. 44–50, 2014.

[65] J. Humann and L.L. Lenz, "Bacterial peptidoglycan-degrading enzymes and their impact on host muropeptide detection," *J. Innate Immun.*, vol. 1, no. 2, pp. 88–97, 2009.

[66] C. Chaput and I.G. Boneca, "Peptidoglycan detection by mammals and flies," *Microbes Infect.*, vol. 9, no. 5, pp. 637–647, 2007.

[67] R. Dziarski, D.R. Kashyap, and D. Gupta, "Mammalian peptidoglycan recognition proteins kill bacteria by activating two-component systems and modulate microbiome and inflammation," *Microb. Drug Resist.*, vol. 18, no. 3, pp. 280–285, 2012.

[68] J. Royet, D. Gupta, and R. Dziarski, "Peptidoglycan recognition proteins: Modulators of the microbiome and inflammation," *Nat. Rev. Immunol.*, vol. 11, no. 12, pp. 837–851, 2011.

[69] L.O. Moreira, D.S. Zamboni, J.G. Filep, and J.C. Salazar, "NOD1 and NOD2 signaling in

infection and inflammation," *Front. Immunol*., vol. 3, art. 382, 2012.

[70] F. Caufrier, A. Martinou, C. Dupont, and V. Bouriotis, "Carbohydrate esterase family 4 enzymes: Substrate specificity," *Carbohydr. Res.*, vol. 338, no. 7, pp. 687–692, 2003.

[71] Y. Zhao, W.T. Ju, G.H. Jo, W.J. Jung, and R.-D. Park, "Perspectives of Chitin Deacetylase Research," Chapter 7 from *Biotechnology of Biopolymers*, IntechOpen, 2011.

[72] V. Ghormade, S. Kulkarni, N. Doiphode, P.R. Rajamohanan, and M.V. Deshpande, "Chitin deacetylase: A comprehensive account on its role in nature and its biotechnological applications," *Curr. Res. Technol. Educ. Top. Appl. Microbiol. Microb. Biotechnol.*, vol. 1, pp. 1054–1066, 2010.

[73] L.G. Baker, C.A. Specht, M.J. Donlin, and J.K. Lodge, "Chitosan, the deacetylated form of chitin, is necessary for cell wall integrity in *Cryptococcus neoformans*," *Eukaryot. Cell*, vol. 5, no. 5, pp. 855-867, 2007.

[74] A. Christodoulidou, V. Bouriotis, and G. Thireos, "Two sporulation-specific chitin deacetylase-encoding genes are required for the ascospore wall rigidity of *Saccharomyces cerevisiae*," *J. Biol. Chem.*, vol. 271, no. 49, pp. 31420–31425, 1996.

[75] I.A. Geoghegan and S.J. Gurr, "Chitosan Mediates Germling Adhesion in *Magnaporthe oryzae* and Is Required for Surface Sensing and Germling Morphogenesis," *PLoS Pathog.*, vol. 12, no. 6, pp. 1–34, 2016.

[76] S. White, M. McIntyre, D.R. Berry, and B. McNeil, "The Autolysis of Industrial Filamentous Fungi," *Crit. Rev. Biotechnol.*, vol. 22, no. 1, pp. 1–14, 2002.

[77] A. Sánchez-Vallet, J.R. Mesters, and B.P.H.J. Thomma, "The battle for chitin recognition in plant-microbe interactions," *FEMS Microbiol. Rev.*, vol. 39, no. 2, pp. 171–183, 2015.

[78] P. Roche *et al.*, "The common nodABC genes of *Rhizobium meliloti* are host-range determinants," *Proc. Natl. Acad. Sci.*, vol. 93, no. 26, pp. 15305-15310, 1996.

[79] N. Keyhani, "Physiological aspects of chitin catabolism in marine bacteria," *Biochim. Biophys. Acta - Gen. Subj.*, vol. 1473, no. 1, pp. 108–122, 1999.

[80] K.L. Meibom, X.B. Li, A.T. Nielsen, C.Y. Wu, S. Roseman, and G.K. Schoolnik, "The *Vibrio cholerae* chitin utilization program," *PNAS*, vol. 101, no. 8, pp. 2524-2529, 2004.

[81] X. Li, L. X. Wang, X. Wang, and S. Roseman, "The chitin catabolic cascade in the marine bacterium *Vibrio cholerae*: Characterization of a unique chitin oligosaccharide deacetylase," *Glycobiology*, vol. 17, no. 12, pp. 1377–1387, 2007.

[82] J. Hoßbach, F. Bußwinkel, A. Kranz, J. Wattjes, S. Cord-Landwehr, and B. M. Moerschbacher, "A chitin deacetylase of *Podospora anserina* has two functional chitin binding domains and a unique mode of action," *Carbohydr. Polym.*, vol. 183, pp. 1-10, 2018.

[83] C. . Peniche Covas, W. Argüelles-Monal, and F. Goycoolea, “Chitin and chitosan: Major sources, properties and applications,” in *Monomers, Polymers and Composites from Renewable Resources*, Elsevier, 2008, pp. 517–542.

[84] P. Karrer and A. Hofmann, “Über den enzymatischen Abbau von Chitin und Chitosan,” *Int. Helv. Chim. Acta*, vol. 12, no. 616–637, 1929.

[85] C. Rouget, “Des substances amylacées dans les tissus des animaux, spécialement des Articulés (chitine),” *Comptes Rendus*, vol. 48, p. 792, 1859.

[86] F. Hoppe-Seyler, “Ueber chitin und cellulose,” *Eur. J. Inorg. Chem.*, vol. 27, no. 3, pp. 3329–3331, 1984.

[87] D.R. Kreger, “Observations on cell walls of yeasts and some other fungi by X-ray diffraction and solubility tests,” *Biochim. Biophys. Acta*, vol. 13, pp. 1–9, 1954.

[88] J.P. Zikakis, *Chitin, chitosan, and related enzymes*. Elsevier, 1984.

[89] K. Kurita, “Controlled functionalization of the polysaccharide chitin,” *Prog. Polym. Sci.*, vol. 26, no. 9, pp. 1921–1971, Nov. 2001.

[90] P.K. Dutta, J. Duta, and V.S. Tripathi, “Chitin and Chitosan: Chemistry, properties and applications,” *J. Sci. Ind. Res. (India).*, vol. 63, no. 1, pp. 20–31, 2004.

[91] A. Varki *et al.*, *Essentials of Glycobiology*, *3rd Edition*, Cold Spring Harbor Laboratory Press, 2015.

[92] M. Rinaudo, “Chitin and chitosan: Properties and applications,” *Prog. Polym. Sci.*, vol. 31, no. 7, pp. 603–632, 2006.

[93] G. Roberts, “Chitosan production routes and their role in determining the structure and properties of the product,” in *Advances in Chitin Science*, Eds. Jaques Andre, p. 22, 1997.

[94] M. Hayafune *et al.*, “Chitin-induced activation of immune signaling by the rice receptor CEBiP relies on a unique sandwich-type dimerization,” *PNAS*, vol. 111, no. 3, pp. E404-E413, 2014.

[95] K. Kurita, “Chitin and Chitosan: Functional Biopolymers from Marine Crustaceans,” *Mar. Biotechnol.*, vol. 8, no. 3, pp. 203–226, 2006.

[96] B. Krajewska, “Application of chitin- and chitosan-based materials for enzyme immobilizations: a review,” *Enzyme Microb. Technol.*, vol. 35, no. 2–3, pp. 126–139, 2004.

[97] I. Tsigos, A. Martinou, D. Kafetzopoulos, and V. Bouriotis, “Chitin deacetylases: New, versatile tools in biotechnology,” *Trends Biotechnol.*, vol. 18, no. 7, pp. 305–312, 2000.

[98] R. Singh, K. Shitiz, and A. Singh, “Chitin and chitosan: biopolymers for wound management,” *Int. Wound J.*, vol. 14, no. 6, pp. 1276–1289, 2017.

[99] S.N. Das *et al.*, “Biotechnological approaches for field applications of chitooligosaccharides

(COS) to induce innate immunity in plants," *Critical Reviews in Biotechnology*, vol. 35, no. 1. pp. 29-43, 2015.

[100] P. Nechita, "Applications of Chitosan in Wastewater Treatment," in *Biological Activities and Application of Marine Polysaccharides*, pp. 209-228, 2017.

[101] T. Philibert, B.H. Lee, and N. Fabien, "Current Status and New Perspectives on Chitin and Chitosan as Functional Biopolymers," *Appl. Biochem. Biotechnol.*, vol. 181, no. 4, pp. 1314–1337, 2017.

[102] M.S. Riaz Rajoka, L. Zhao, H.M. Mehwish, Y. Wu, and S. Mahmood, "Chitosan and its derivatives: synthesis, biotechnological applications, and future challenges," *Appl. Microbiol. Biotechnol.*, vol. 103, no. 4, pp. 1557–1571, 2019.

[103] J. Feng, L. Zhao, and Q. Yu, "Receptor-mediated stimulatory effect of oligochitosan in macrophages," *Biochem. Biophys. Res. Commun.*, vol. 317, no. 2, pp. 414–420, Apr. 2004.

[104] P. Zhang, W. Liu, Y. Peng, B. Han, and Y. Yang, "Toll like receptor 4 (TLR4) mediates the stimulating activities of chitosan oligosaccharide on macrophages," *Int. Immunopharmacol.*, vol. 23, no. 1, pp. 254–261, 2014.

[105] M. Yousef, R. Pichyangkura, S. Soodvilai, V. Chatsudthipong, and C. Muanprasat, "Chitosan oligosaccharide as potential therapy of inflammatory bowel disease: Therapeutic efficacy and possible mechanisms of action," *Pharmacol. Res.*, vol. 66, no. 1, pp. 66–79, 2012.

[106] A. Muxika, A. Etxabide, J. Uranga, P. Guerrero, and K. de la Caba, "Chitosan as a bioactive polymer: Processing, properties and applications," *Int. J. Biol. Macromol.*, vol. 105, pp. 1358-1368, 2017.

[107] Y.M. Lv *et al.*, "Highly efficient and selective biocatalytic production of glucosamine from chitin," *Green Chem.*, vol. 19, no. 2, pp. 527–535, 2017.

[108] G. Lodhi *et al.*, "Chitooligosaccharide and its derivatives: Preparation and biological applications," *Biomed Res. Int.*, vol. 2014, art. 654913, 2014.

[109] C. Muanprasat and V. Chatsudthipong, "Chitosan oligosaccharide: Biological activities and potential therapeutic applications," *Pharmacol. Ther.*, vol. 170, pp. 80–97, 2017.

[110] H. Aragunde, "Ingeniería enzimática de quitina desacetilasas y glicosintasas como biocatalizadores: diseño racional de la especificidad y evolución dirigida," Doctoral thesis, IQS School of Engineering (Universitat Ramon Llull), 2017.

[111] I. Younes and M. Rinaudo, "Chitin and chitosan preparation from marine sources. Structure, properties and applications," *Mar. Drugs*, vol. 13, pp. 1133-1174, 2015.

[112] R.C.F. Cheung, T.B. Ng, J.H. Wong, and W.Y. Chan, "Chitosan: An update on potential biomedical and pharmaceutical applications," *Mar. Drugs*, vol. 13, pp. 5156-5186, 2015.

[113] A. Anitha *et al.*, "Chitin and chitosan in selected biomedical applications," *Progress in Polymer Science*, vol. 39, no. 9, pp. 1644-1667, 2014.

[114] A. Khalil HPS *et al.*, "A review on chitosan-cellulose blends and nanocellulose reinforced chitosan biocomposites: Properties and their applications," *Carbohydr. Polym.*, vol. 150, pp. 216–226, 2016.

[115] A. Pestov and S. Bratskaya, "Chitosan and Its Derivatives as Highly Efficient Polymer Ligands," *Molecules*, vol. 21, no. 3, p. 330, 2016.

[116] F.M. Goycoolea *et al.*, "Effect of Molecular Weight and Degree of Acetylation on the Physicochemical Characteristics of Chitosan Nanoparticles," *Adv. Chitin Sci.*, vol. 10, pp. 542–547, 2007.

[117] N. El Gueddari, S. Kolkenbrock, and A. Schaaf, "Chitin and chitosan modifying enzymes: versatile novel tools for the analysis of structure–function relationships of partially acetylated chitosans," *Adv Chitin Sci*, vol. 14, pp. 40–47, 2014.

[118] S.N. Hamer *et al.*, "Enzymatic production of defined chitosan oligomers with a specific pattern of acetylation using a combination of chitin oligosaccharide deacetylases," *Sci. Rep.*, vol. 5, art. 8716, 2015.

[119] L. Hembach, S. Cord-Landwehr, and B.M. Moerschbacher, "Enzymatic production of all fourteen partially acetylated chitosan tetramers using different chitin deacetylases acting in forward or reverse mode," *Sci. Rep.*, vol. 7, art. 17692 , 2017.

[120] L. Grifoll-Romero, S. Pascual, H. Aragunde, X. Biarnés, and A. Planas, "Chitin deacetylases: Structures, specificities, and biotech applications," *Polymers*, vol. 10, art. 352, 2018.

[121] K. Kobayashi *et al.*, "Identification and characterization of a novel polysaccharide deacetylase C (PdaC) from Bacillus subtilis," *J. Biol. Chem.*, vol. 287, no. 13, pp. 9765–9776, 2012.

[122] M.C. Lo *et al.*, "Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery," *Anal. Biochem.*, vol. 332, no. 1, pp. 153–159, 2004.

[123] S. Cord-Landwehr, P. Ihmor, A. Niehues, H. Luftmann, B.M. Moerschbacher, and M. Mormann, "Quantitative Mass-Spectrometric Sequencing of Chitosan Oligomers Revealing Cleavage Sites of Chitosan Hydrolases," *Anal. Chem.*, vol. 85, no. 5, pp. 2893-2900, 2017.

[124] B. J. Bennion and V. Daggett, "The molecular basis for the chemical denaturation of proteins by urea," *PNAS*, vol. 29, no. 9, pp. 5142-5147, 2003.

[125] K. Tokuyasu, H. Ono, K. Hayashi, and Y. Mori, "Reverse hydrolysis reaction of chitin deacetylase and enzymatic synthesis of ß-D-GlcNAc-(1,4)-GlcN from chitobiose," vol. 322, pp. 26–31, 1999.

[126] C.N. Pace, “Determination and Analysis of Urea and Guanidine Hydrochloride Denaturation Curves,” *Methods Enzymol.*, vol. 131, no. 539, pp. 266–280, 1986.

[127] V. Spiwok, “CH/π Interactions in Carbohydrate Recognition,” *Molecules*, vol. 22, art. 1038, 2017.

[128] D.L. Popham, J. Helin, C.E. Costello, and P. Setlow, “Muramic lactam in peptidoglycan of *Bacillus subtilis* spores is required for spore outgrowth but not for spore dehydration or heat resistance,” *Proc. Natl. Acad. Sci.*, vol. 93, pp. 15405-15410, 1996.

[129] J.E. Melnyk, V. Mohanan, A.K. Schaefer, C.W. Hou, and C.L. Grimes, “Peptidoglycan Modifications Tune the Stability and Function of the Innate Immune Receptor Nod2,” *J. Am. Chem. Soc.*, vol. 137, pp. 6987-6990, 2015.

[130] S. Naqvi *et al.*, “A recombinant fungal chitin deacetylase produces fully defined chitosan oligomers with novel patterns of acetylation,” *Appl. Environ. Microbiol.*, vol. 82, no. 22, pp. 6645–6655, 2016.

[131] T. Fukushima, H. Yamamoto, A. Atrih, S.J. Foster, and J. Sekiguchi, “A polysaccharide deacetylase gene (pdaA) is required for germination and for production of muramic ð-lactam residues in the spore cortex of *Bacillus subtilis*,” *J. Bacteriol.*, vol. 184, no. 21, pp. 6007–6015, 2002.

[132] T. Fukushima, “Characterization of a Polysaccharide Deacetylase Gene Homologue (pdaB) on Sporulation of *Bacillus subtilis*,” *J. Biochem.*, vol. 136, no. 3, pp. 283–291, 2004.

[133] S. Balomenou, A. Fouet, M. Tzanodaskalaki, E. Couture-Tosi, V. Bouriotis, and I.G. Boneca, “Distinct functions of polysaccharide deacetylases in cell shape, neutral polysaccharide synthesis and virulence of *Bacillus anthracis*,” *Mol. Microbiol.*, vol. 87, no. 4, pp. 867–883, 2013.

[134] E. Psylinakis *et al.*, “Peptidoglycan N-Acetylglucosamine Deacetylases from *Bacillus cereus*, Highly Conserved Proteins in *Bacillus anthracis*,” *J. Biol. Chem.*, vol. 280, no. 35, pp. 30856-30863, 2005.

[135] M. Martínez-Alonso, E. García-Fruitós, N. Ferrer-Miralles, U. Rinas, and A. Villaverde, “Side effects of chaperone gene co-expression in recombinant protein production,” *Microb. Cell Fact.*, vol. 9, pp. 1–6, 2010.

[136] M. A. Massiah, K. M. Wright, and H. Du, “Obtaining soluble folded proteins from inclusion bodies using sarkosyl, triton X-100, and CHAPS: Application to LB and M9 minimal media,” *Curr. Protoc. Protein Sci.*, vol. 2016, pp. 6.13.1-6.13.24, 2016.

[137] G. L. Rosano and E. A. Ceccarelli, “Recombinant protein expression in *Escherichia coli*: Advances and challenges,” *Front. Microbiol.*, vol. 5, pp. 1–17, 2014.

[138] W. Vollmer and A. Tomasz, “The pgdA gene encodes for a peptidoglycan N-acetylglucosamine deacetylase in *Streptococcus pneumoniae*,” *J. Biol. Chem.*, vol. 275, no. 27, pp. 20496–20501, 2000.

[139] F.W. Studier, “Protein production by auto-induction in high density shaking cultures.,” *Protein Expr. Purif.*, vol. 41, no. 1, pp. 207–234, 2005.

[140] C. Ostermeier, A. Harrenga, U. Ermler, and H. Michel, “Structure at 2.7 Å resolution of the *Paracoccus denitrificans* two-subunit cytochrome c oxidase complexed with an antibody F_V fragment,” *Proc. Natl. Acad. Sci.*, vol. 94, pp. 10547-10533, 1997.

[141] A. Skerra and T.G.M. Schmidt, “Use of the Strep- tag and streptavidin for detection and purification of recombinant proteins,” *Methods Enzymol.*, vol. 326, pp. 271–304, 2000.

[142] T.G. Schmidt and A. Skerra, “The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins,” *Nat. Protoc.*, vol. 2, no. 6, pp. 1528–1535, 2007.

[143] J.A. Bornhorst, J.J. Falke, “Purification of Proteins Using Polyhistidine Affinity Tags,” *Methods Enzym.*, vol. 2000, no. 326, pp. 245–254, 2010.

[144] S. Boivin, S. Kozak, and R. Meijers, “Optimization of protein purification and characterization using Thermofluor screens,” *Protein Expr. Purif.*, vol. 91, no. 2, pp. 192–206, 2013.

[145] F.H. Niesen, H. Berglund, and M. Vedadi, “The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability,” *Nat. Protoc.*, vol. 2, no. 9, pp. 2212–2221, 2007.

[146] W. Humphrey, A. Dalke, and K. Schulten, “VMD: visual molecular dynamics,” *J. Mol. Graph. Model.*, vol. 14, pp. 33–38, 27–28, 1996.

[147] D.T. Jones, W.R. Taylor, and J.M. Thornton, “The rapid generation of mutation data matrices,” *Comput Appl Biosci.*, vol. 8, no. 3, pp. 275–282, 1992.

[148] S. Kumar, G. Stecher, and K. Tamura, “MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets,” *Mol. Biol. Evol.*, vol. 33, no. 7, pp. 1870–1874, 2016.

[149] A. M. Waterhouse, J.B. Procter, D.M.A. Martin, M. Clamp, and G.J. Barton, “Jalview Version 2-A multiple sequence alignment editor and analysis workbench,” *Bioinformatics*, vol. 25, no. 9, pp. 1189–1191, 2009.

[150] G.M. Morris *et al.*, “AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility,” *J. Comput. Chem.*, vol. 30, no. 16, pp. 2785–2791, 2009.

[151] O. Trott, and A.J. Olson, “Autodock vina: improving the speed and accuracy of docking,” *J. Comput. Chem.*, vol. 31, no. 2, pp. 455–461, 2019.

[152] W. Kabsch, “XDS,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. 2, pp. 125–132, 2010.

[153] A.J. McCoy *et al.*, “Phaser crystallographic software,” *J. Appl. Crystallogr.*, vol. 40, no. 4, pp. 658–674, 2007.

[154] P.D. Adams *et al.*, “PHENIX: A comprehensive Python-based system for macromolecular structure solution,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. 2, pp. 213–221, 2010.

[155] K. Cowtan, “The Buccaneer software for automated model building. 1. Tracing protein chains,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 62, no. 9, pp. 1002–1011, 2006.

[156] M.D. Winn *et al.*, “Overview of the CCP4 suite and current developments.,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 67, no. Pt 4, pp. 235–242, 2011.

[157] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan, “Features and development of Coot.,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. Pt 4, pp. 486–501, 2010.

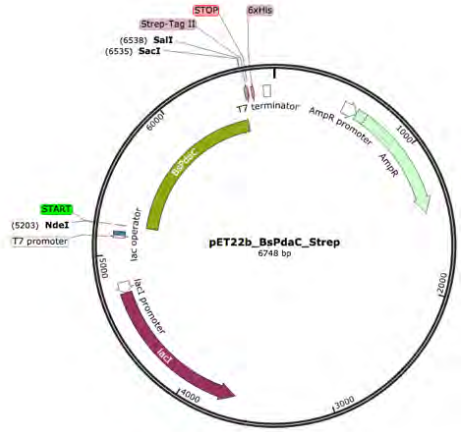
[158] V.B. Chen *et al.*, “MolProbity: all-atom structure validation for macromolecular crystallography.,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. Pt 1, pp. 12–21, 2010.

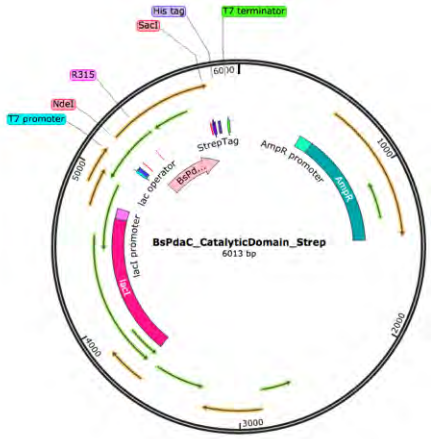
[159] E.F. Pettersen *et al.*, “UCSF Chimera: a visualization system for exploratory research and analysis.,” *J. Comput. Chem.*, vol. 25, no. 13, pp. 1605–1612, 2004.

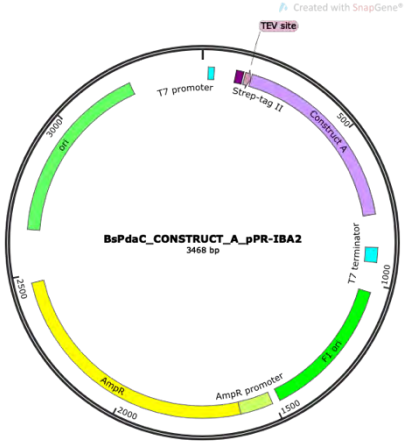
ANNEXES

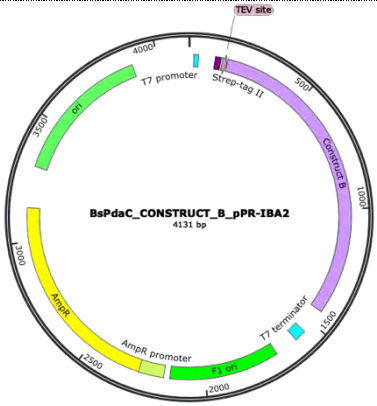
9. ANNEXES

9.1. Nucleotide and amino acid sequences from all enzymes and constructs

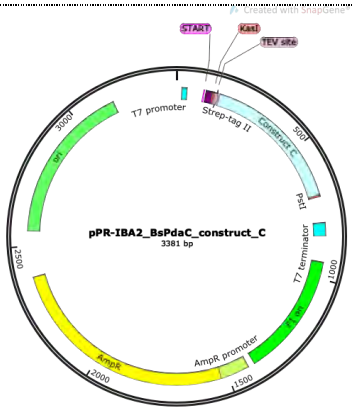
PROTEIN	*Bs*PdaC-FL
VECTOR	pET22b(+)
PLASMID MAP	
GENE SEQUENCE	taatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaaggagatat acatatgaaccacagcctgaaaaaagaaaccgtgatgaataaagttcgtaccgatagccagtatggcaatgttgaaattgc aaccctggtgaatgatggcaaaaccttaactatgccgttaactatccggtgtttaaaaacgagaaaatggacagcgcactg aaacgttttgcagaaaaagaagttcgccagttccagaaagaaaccaaagatgttgatcaagagcataccaccaaacgcaat gaactgaatgtggattacaaaatcgtgcactatgcaaaacagaccgttgccattgtgttcaacgagtataaatacattggtgg tgcacatggtcagaccgtgaaaaaaaccttcaattacgattttagcaaacaggcctttctgagcatcgatgatatcttcaaag aagatgccgattacctgcataaactgagcctgattgcatatcacgagctgaaaaaaaacaaagatatcgcagcagatgatg ccctgctgaaagaaggcaccgcaccgaaaaaagagaattttagccgttttgcgatcaaagaggattacatcgaactgtattt cgatacctatcaggttgcagcaggttatctgggtgaacagagcattgcaatcaaaaaaagtctgctgaaagatatcctgaaa gaacagtatatcgataaagccaaaaacaaaaacaaaatcaaagaacagaaaccgaaacacgaagttattagcctgccta aagaagaaacagttgatccgaatcagaaagttattgccctgacctttgatgatggtccgaatccggcaaccaccaatcagat tctggatagtctgaaaaaatacaaaggccacgccacctttttttgttctgggtagccgtgttcagtattatccggaaaccctgat tcgtatgctgaaagagggtaacgaagttggtaatcatagctggtcacatccgctgctgacccgtctgagcgttaaagaagcc ctgaaacaaattaatgatacccaggacatcatcgagaaaattagcggttatcgtccgaccctggttcgtccgccttatggtgg tattaatgatgaactgcgtagccagatgaaaatggatgttgcactgtgggatgtggatccggaagattggaaagatcgtaac aaaaaaaccattgtggatcgcgttatgaatcaagccggtgatggtcgtaccattctgattcatgatatttatcgtaccagcgca gatgcagccgatgaaattatcaaaaaactgaccgatcagggttatcagctggttaccgttagccagctggaagaggttaaaa aacagcgtgaagcaaaagagctccgtcgacaatggtcacatcctcaatttgaaaaatagggcttgcggccgcactcgagcac caccaccaccaccactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataa ctagcataaccccttggggcctctaaacgggtcttgaggggttttttg
PROTEIN SEQUENCE	MNHSLKKETVMNKVRTDSQYGNVEIATLVNDGKTFNYAVNYPVFKNEKMDSALKRFAEKEVRQF QKETKDVDQEHTTKRNELNVDYKIVHYAKQTVAIVFNEYKYIGGAHGQTVKKTFNYDFSKQAFLSID DIFKEDADYLHKLSLIAYHELKKNKDIAADDALLKEGTAPKKENFSRFAIKEDYIELYFDTYQVAAGYL GEQSIAIKKSLLKDILKEQYIDKAKNKNKIKEQKPKHEVISLPKEETVDPNQKVIALTFDDGPNPATTN QILDSLKKYKGHATFFVLGSRVQYYPETLIRMLKEGNEVGNHSWSHPLLTRLSVKEALKQINDTQDII EKISGYRPTLVRPPYGGINDELRSQMKMDVALWDVDPEDWKDRNKKTIVDRVMNQAGDGRTILI HDIYRTSADAADEIIKKLTDQGYQLVTVSQLEEVKKQREAKELRRQWSHPQFEK*
LEGEND	Underlined: sequence of the protein; start and stop codons; Strep-II-tag; His-tag; T7 promoter and T7 terminator
MUTANTS	D285S, W402A, W395A, D398E, D398N, D398A, D398L, D398K

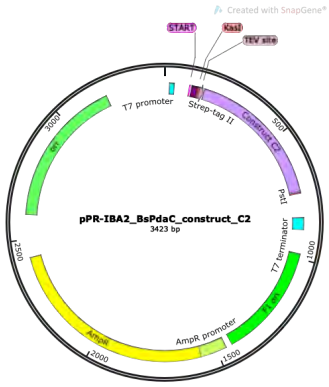
PROTEIN	*Bs*PdaC-CD
VECTOR	pET22b(+)
PLASMID MAP	
GENE SEQUENCE	taatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaaggagatat acatatggaagaaacagttgatccgaatcagaaagttattgccctgacctttgatgatggtccgaatccggcaaccaccaat cagattctggatagtctgaaaaaatacaaaggccacgccacctttttttgttctgggtagccgtgttcagtattatccggaaacc ctgattcgtatgctgaaagagggtaacgaagttggtaatcatagctggtcacatccgctgctgacccgtctgagcgttaaaga agccctgaaacaaattaatgatacccaggacatcatcgagaaaattagcggttatcgtccgaccctggttcgtccgccttatg gtggtattaatgatgaactgcgtagccagatgaaaatggatgttgcactgtgggatgtggatccggaagattggaaagatcg taacaaaaaaaccattgtggatcgcgttatgaatcaagccggtgatggtcgtaccattctgattcatgatatttatcgtaccag cgcagatgcagccgatgaaattatcaaaaaactgaccgatcagggttatcagctggttaccgttagccagctggaagaggtt aaaaaacagcgtgaagcaaaagagctccgtcgacaatggtcacatcctcaatttgaaaaatagcttgcggccgcactcga gcaccaccaccaccaccactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagca ataactagcataaccccttggggcctctaaacgggtcttgaggggttttttg
PROTEIN SEQUENCE	MEETVDPNQKVIALTFDDGPNPATTNQILDSLKKYKGHATFFVLGSRVQYYPETLIRMLKEGNEVG NHSWSHPLLTRLSVKEALKQINDTQDIIEKISGYRPTLVRPPYGGINDELRSQMKMDVALWDVDPE DWKDRNKKTIVDRVMNQAGDGRTILIHDIYRTSADAADEIIKKLTDQGYQLVTVSQLEEVKKQREA KELRRQWSHPQFEK*
LEGEND	Underlined: sequence of the protein; start and stop codons; Strep-II-tag; His-tag; T7 promoter and T7 terminator
MUTANTS	D285S, W402A, Y377A, Y430A, R315A, R315I, R315S, R315K

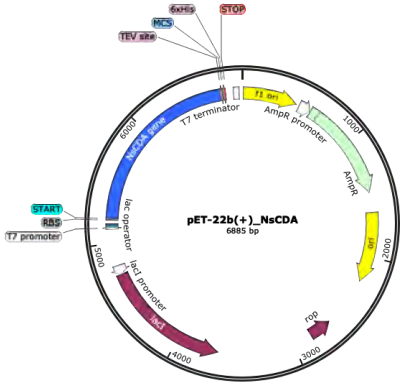
PROTEIN	*Bs*PdaC-A
VECTOR	pPR-IBA2
PLASMID MAP	
GENE SEQUENCE	taatacgactcactatagggaggccacaacggtttccctctagaaataatttgtttaactttaagaaggagatatacatatg gctagctggagccacccgcagttcgaaaaaggcgccgaaaacctgtattttcagggcagcctgaaaaaagaaaccgtgatg aataaagttcgtaccgatagccagtatggcaatgttgaaattgcaaccctggtgaatgatggcaaaacctttaactatgccgt taactatccggtgtttaaaaacgagaaaatggacagcgcactgaaacgttttgcagaaaaagaagttcgccagttccagaa agaaaccaaagatgttgatcaagagcataccaccaaacgcaatgaactgaatgtggattacaaaatcgtgcactatgcaaa acagaccgttgccattgtgttcaacgagtataaatacattggtggtgcacatggtcagaccgtgaaaaaaaccttcaattacg attttagcaaacaggcctttctgagcatcgatgatatcttcaaagaagatgccgattacctgcataaactgagcctgattgcat atcacgagctgaaaaaaaacaaagatatcgcagcagatgatgccctgctgaaagaaggcaccgcaccgaaaaaagagaa ttttagccgttttgcgatcaaagaggattacatcgaactgtatttcgatacctatcaggttgcagcaggttatctgggtgaaca gagcattgcaatcaaaaaatagctgcaggggaccatggtctctgatatctaactaagcttgatccggctgctaacaaagcc cgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacgggtcttgaggg gttttttg
PROTEIN SEQUENCE	MASWSHPQFEKGAENLYFQGSLKKETVMNKVRTDSQYGNVEIATLVNDGKTFNYAVNYPVFKNE KMDSALKRFAEKEVRQFQKETKDVDQEHTTKRNELNVDYKIVHYAKQTVAIVFNEYKYIGGAHGQ TVKKTFNYDFSKQAFLSIDDIFKEDADYLHKLSLIAYHELKKNKDIAADDALLKEGTAPKKENFSRFAIK EDYIELYFDTYQVAAGYLGEQSIAIKK*
LEGEND	Underlined: sequence of the protein; start and stop codons; Strep-II-tag; TEV site; T7 promoter and T7 terminator
MUTANTS	None

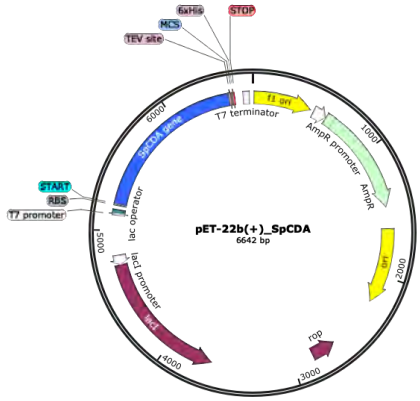
PROTEIN	*Bs*PdaC-B
VECTOR	pPR-IBA2
PLASMID MAP	

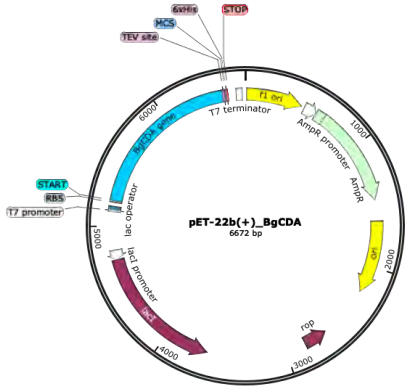
GENE SEQUENCE	taatacgactcactatagggaggccacaacggtttccctctagaaataattttgtttaactttaagaaggagatatacatatg gctagctggagccacccgcagttcgaaaaaggcgccgaaaacctgtattttcagggcaccacagcctgaaaaaagaaacc gtgatgaataaagttcgtaccgatagccagtatggcaatgttgaaattgcaaccctggtgaatgatggcaaaaccttaacta tgccgttaactatccggtgtttaaaaacgagaaaatggacagcgcactgaaacgttttgcagaaaaagaagttcgccagttc cagaaagaaaccaaagatgttgatcaagagcataccaccaaacgcaatgaactgaatgtggattacaaaatcgtgcactat gcaaaacagaccgttgccattgtgttcaacgagtataaatacattggtggtgcacatggtcagaccgtgaaaaaaaccttca attacgattttagcaaacaggcctttctgagcatcgatgatatcttcaaagaagatgccgattacctgcataaactgagcctg attgcatatcacgagctgaaaaaaaacaaagatatcgcagcagatgatgccctgctgaaagaaggcaccgcaccgaaaaa agagaattttagccgttttgcgatcaaagaggattacatcgaactgtatttcgatacctatcaggttgcagcaggttatctggg tgaacagagcattgcaatcaaaaaaagtctgctgaaagatatcctgaaagaacagtatatcgataaagccaaaaacaaaa acaaaatcaaagaacagaaaccgaaacacgaagttattagcctgcctaaagaagaaacagttgatccgaatcagaaagtt attgccctgacctttgatgatggtccgaatccggcaaccaccaatcagattctggatagtctgaaaaaatacaaaggccacg ccacctttttttgttctgggtagccgtgttcagtattatccggaaaccctgattcgtatgctgaaagagggtaacgaagttggta atcatagctggtcacatccgctgctgacccgtctgagcgttaaagaagccctgaaacaaattaatgatacccaggacatcat cgagaaaattagcggttatcgtccgaccctggttcgtccgccttatggtggtattaatgatgaactgcgtagccagatgaaaa tggatgttgcactgtgggatgtggatccggaagattggaaagatcgtaacaaaaaaaccattgtggatcgcgttatgaatca agccggtgatggtcgtaccattctgattcatgatatttatcgtaccagcgcagatgcagccgatgaaattatcaaaaaactga ccgatcagggttatcagctggttacctagctgcaggggaccatggtctctgatatctaactaagcttgatccggctgctaaca aagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacgggtctt gaggggttttttg
PROTEIN SEQUENCE	MASWSHPQFEKGAENLYFQGNHSLKKETVMNKVRTDSQYGNVEIATLVNDGKTFNYAVNYPVFK NEKMDSALKRFAEKEVRQFQKETKDVDQEHTTKRNELNVDYKIVHYAKQTVAIVFNEYKYIGGAHG QTVKKTFNYDFSKQAFLSIDDIFKEDADYLHKLSLIAYHELKKNKDIAADDALLKEGTAPKKENFSRFA IKEDYIELYFDTYQVAAGYLGEQSIAIKKSLLKDILKEQYIDKAKNKNKIKEQKPKHEVISLPKEETVDPN QKVIALTFDDGPNPATTNQILDSLKKYKGHATFFVLGSRVQYYPETLIRMLKEGNEVGNHSWSHPLL TRLSVKEALKQINDTQDIIEKISGYRPTLVRPPYGGINDELRSQMKMDVALWDVDPEDWKDRNKK TIVDRVMNQAGDGRTILIHDIYRTSADAADEIIKKLTDQGYQLVT*
LEGEND	Underlined: sequence of the protein; start and stop codons; Strep-II-tag; TEV site; T7 promoter and T7 terminator
MUTANTS	D285S

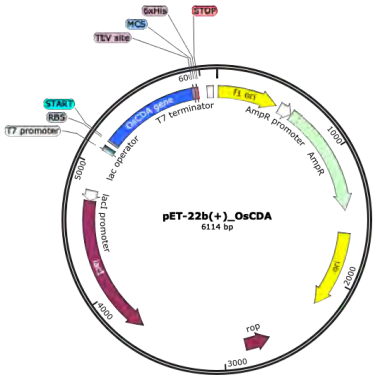
PROTEIN	*Bs*PdaC-C
VECTOR	pPR-IBA2
PLASMID MAP	
GENE SEQUENCE	taatacgactcactatagggaggccacaacggtttccctctagaaataattttgtttaactttaagaaggagatatacatatg gctagctggagccacccgcagttcgaaaaaggcgccgaaaacctgtattttcagggccagaaagttattgccctgacctttg atgatggtccgaatccggcaaccaccaatcagattctggatagtctgaaaaaatacaaaggccacgccacctttttttgttctg ggtagccgtgttcagtattatccggaaaccctgattcgtatgctgaaagagggtaacgaagttggtaatcatagctggtcaca tccgctgctgacccgtctgagcgttaaagaagccctgaaacaaattaatgatacccaggacatcatcgagaaaattagcggt tatcgtccgaccctggttcgtccgccttatggtggtattaatgatgaactgcgtagccagatgaaaatggatgttgcactgtgg gatgtggatccggaagattggaaagatcgtaacaaaaaaaccattgtggatcgcgttatgaatcaagccggtgatggtcgta ccattctgattcatgatatttatcgtaccagcgcagatgcagccgatgaaattatcaaaaaactgaccgatcagggttatcag ctggttacctagctgcaggggaccatggtctctgatatctaactaagcttgatccggctgctaacaaagcccgaaaggaag ctgagttggctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttg
PROTEIN SEQUENCE	MASWSHPQFEKGAENLYFQGQKVIALTFDDGPNPATTNQILDSLKKYKGHATFFVLGSRVQYYPET LIRMLKEGNEVGNHSWSHPLLTRLSVKEALKQINDTQDIIEKISGYRPTLVRPPYGGINDELRSQMK MDVALWDVDPEDWKDRNKKTIVDRVMNQAGDGRTILIHDIYRTSADAADEIIKKLTDQGYQLVT*
LEGEND	Underlined: sequence of the protein; start and stop codons; Strep-II-tag; TEV site; T7 promoter and T7 terminator
MUTANTS	None

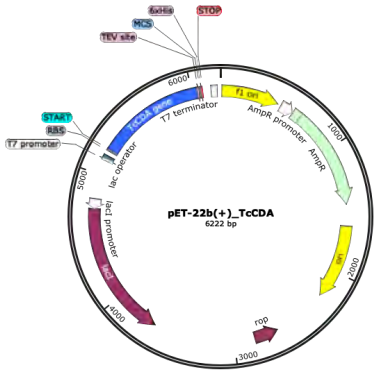
PROTEIN	*Bs*PdaC-C2
VECTOR	pPR-IBA2
PLASMID MAP	
GENE SEQUENCE	taatacgactcactatagggaggccacaacggtttccctctagaaataattttgtttaactttaagaaggagatatacatatg gctagctggagccacccgcagttcgaaaaaggcgccgaaaacctgtattttcagggccagaaagttattgccctgacctttg atgatggtccgaatccggcaaccaccaatcagattctggatagtctgaaaaaatacaaaggccacgccacctttttttgttctg ggtagccgtgttcagtattatccggaaaccctgattcgtatgctgaaagagggtaacgaagttggtaatcatagctggtcaca tccgctgctgacccgtctgagcgttaaagaagccctgaaacaaattaatgatacccaggacatcatcgagaaaattagcggt tatcgtccgaccctggttcgtccgccttatggtggtattaatgatgaactgcgtagccagatgaaaatggatgttgcactgtgg gatgtggatccggaagattggaaagatcgtaacaaaaaaaccattgtggatcgcgttatgaatcaagccggtgatggtcgta ccattctgattcatgatatttatcgtaccagcgcagatgcagccgatgaaattatcaaaaaactgaccgatcagggttatcag ctggttaccgttagccagctggaagaggttaaaaaacagcgtgaagcaaaatagctgcaggggaccatggtctctgatat ctaactaagcttgatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcat aacccccttggggcctctaaacgggtcttgaggggttttttg
PROTEIN SEQUENCE	MASWSHPQFEKGAENLYFQGQKVIALTFDDGPNPATTNQILDSLKKYKGHATFFVLGSRVQYYPET LIRMLKEGNEVGNHSWSHPLLTRLSVKEALKQINDTQDIIEKISGYRPTLVRPPYGGINDELRSQMK MDVALWDVDPEDWKDRNKKTIVDRVMNQAGDGRTILIHDIYRTSADAADEIIKKLTDQGYQLVTV SQLEEVKKQREAK*
LEGEND	Underlined: sequence of the protein; start and stop codons; Strep-II-tag; TEV site; T7 promoter and T7 terminator
MUTANTS	D285S

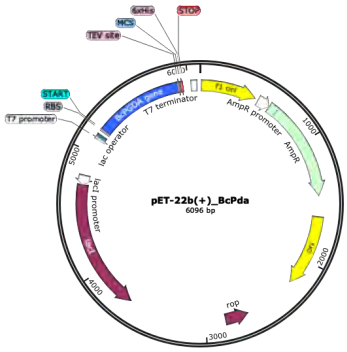
PROTEIN	*Ns*CDA
VECTOR	pET22b(+)
PLASMID MAP	
GENE SEQUENCE	taatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaaggagatat acatatgacaccggttcatccgcctgcaaccgcacagagcggtagtctgaccctggttcgtccgggtatgaccatgcatatta gcagtccgcgtcataatgaaagcacaccgctggatctgctgcgtagtgatgttaatagcgaagcattaggcaccggtagcgt taaaagtccggaaagtcagacccgtaccgatacacatcgtcaggcaattaccccgaaaaccaatccgaatacacagaatat tccgggtaataactttaatcgtagcggcaccaaagaactgtatctgacctttgatgatggtccgctgcgtggtacaggtaatgt tctgaaaattctgaaagaagaaggcgttccggcaaccatgttttgtgttggtcgtcatgcacagcagcatccgggtctgctgc aagaggaactgagcatgccgaatctgctgattgcaaatcacacctatagccatgcaaatggtcattatcgtcgcttttatagc aataccttggtctgctgagcgatattgaacatgcccagctgattttaggtggtcgtaaatatctgcgtctggcaggtcgtaat gtttggcgtacaccggaaattaaacgtgatgatctggcaattgttgcactgcgtggtcgtgttgaagttccggaatatgatagc attagcaaagatggcttctatatctatggttgggataccgaatggcattataatcatgcaaccggtaaaccgattgagtcacc ggaaaaagttgcagccaaaatcgatcatctgtatcgtcgtcatcgtagcgcaaaaccgggtaaagtgattctgctggcacat gattttatgtttcgtacccaggcaggcgcaagccgtctgcgtcgttttattcgtctgatgaaacagcgtggttggagctttcata ccattcgtcattatagccgttataaaccggaaccgctgtatgttgcaaaatactatggtcatagtccgaaagaactttacgca gcaaataaacagcgtagctttcgtccggttcaagaaaatcgtggtgttgaaagcattaatcgttcaagcagctcaccgcgtcc gacagcacagaatggtaaaggtctgatgcagaatcgtctgattgaagcaattcgtcgctataacgcaaaagaagtggaaca tctgattgatcagggtgcaagcgttaatcagccggatagctatggtcacattgcactgaatagcgcagttaaagccaatagc atttacctggttaaaaaactgctggcccatggtgcagatccgatgattaaagatgcacgtggtgaaaatgcactgtataccgc aaaacgttataatcgcggtggtattgaacgctacctgaaaaactattatagcgcacgtagccagaccagcaccattgttaatc aggtgcaagaaaaacaaattgccgaaacaccgcatctggcagatagcagcagcaccgcacataccaatccgctgaaactg ctgcgcatgcgtgaaaatctgtattttcagggcctcgagcaccaccaccaccaccactgagatccggctgctaacaaagccc gaaaggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacgggtcttgagggg ttttttgctgaaaggaggaactatatccggat
PROTEIN SEQUENCE	MTPVHPPATAQSGSLTLVRPGMTMHISSPRHNESTPLDLLRSDVNSEALGTGSVKSPESQTRTDTH RQAITPKTNPNTQNIPGNNFNRSGTKELYLTFDDGPLRGTGNVLKILKEEGVPATMFCVGRHAQQ HPGLLQEELSMPNLLIANHTYSHANGHYRRFYSNTFGLLSDIEHAQLILGGRKYLRLAGRNVWRTPE IKRDDLAIVALRGRVEVPEYDSISKDGFYIYGWDTEWHYNHATGKPIESPEKVAAKIDHLYRRHRSA KPGKVILLAHDFMFRTQAGASRLRRFIRLMKQRGWSFHTIRHYSRYKPEPLYVAKYYGHSPKELYAA NKQRSFRPVQENRGVESINRSSSSPRPTAQNGKGLMQNRLIEAIRRYNAKEVEHLIDQGASVNQPD SYGHIALNSAVKANSIYLVKKLLAHGADPMIKDARGENALYTAKRYNRGGIERYLKNYYSARSQTSTI VNQVQEKQIAETPHLADSSSTAHTNPLKLLRMRENLYFQGLEHHHHHH*
LEGEND	Underlined: sequence of the protein; start and stop codons; TEV site; His-tag; T7 promoter and T7 terminator
MUTANTS	None

PROTEIN	*Sp*CDA
VECTOR	pET22b(+)
PLASMID MAP	
GENE SEQUENCE	taatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaaggagatat acatatgcctccggttgaaacccctccggcaattatggaaagtccggttgatcgtcaggttcaaggtggtaaacgtaaaaaa ctgctgttaggtgttggtgccctggttattgttgcagcagcagcaaccgccattgcagttgcaaccaccgcagaaagcagcac cgcaagcaatgataattcagataatgcagttgttacaccgcagcctccgagcattgttattccgatgccgagcgcaacaggtg caccggcaagcgttctgaaagttaaaacaccgcataacattaagatgccaccggttccgcctgcaaccgatattccgaccaa accgacctgtccgagcagccgtaaagcagaaaaaggtgataccgtttttatgaccctggatgatggtccgagtattaaaggt cgcaaaaatctgctgaccgcactggcacagattaatcagaccattagctttttgagagcagctataacttttgtggtccggaa acctttatgaacaagaactgcgttgtgaaagcccgagtccgtatagcgaagttaccgatttatttgcctataccattaaggca ggtcattttctggcagcacatagcaatacccactattacaataatcagagcgcactgtgtgaatatgccaacatggcaaaatt caccaaaatcgatgcacagtatgaaagctgtggtaatacaccggttgcagatatggttcgtggtgcactgcgtattcagaca gcactgaataatgaatcactgtgggataacgataaagaattcgccatgtatcagaaagccatgagcagcatttggacctatg cacgtctgccgtgtaccagcgcatggcgtctgcctggttatcagaaaaccacactgctgggccataaagatgcaatgcagcc ggaagcgggtgcacgtaatgaagttgcagatgccatgtttcgtggtagcctgccgtgtcgtaatgaaacctttcagagccagc cgtggtatagcattggttgggatgttgaatggcgttgggaaagtctgagcgatatgcaggcagcaaaatgtgcaatgtttaaa accatcgaaggcaattttgcatggggtggtcgtgatgcacagcgtaaacaagaagttgttgttctgggtcacgattatcattat gatacaccggcaaaagccaaaatgtttcgcgatctgctggttgaactgaaactgcatggttatgcactggataccgttgatca tctgaaaatctataccgaaaatctgtattttcagggcctcgagcaccaccaccaccaccactgagatccggctgctaacaaa gcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacgggtcttga ggggttttttgctgaaaggaggaactatatccggat
PROTEIN SEQUENCE	MTPVHPPATAQSGSLTLVRPGMTMHISSPRHNESTPLDLLRSDVNSEALGTGSVKSPESQTRTDTH RQAITPKTNPNTQNIPGNNFNRSGTKELYLTFDDGPLRGTGNVLKILKEEGVPATMFCVGRHAQQ HPGLLQEELSMPNLLIANHTYSHANGHYRRFYSNTFGLLSDIEHAQLILGGRKYLRLAGRNVWRTPE IKRDDLAIVALRGRVEVPEYDSISKDGFYIYGWDTEWHYNHATGKPIESPEKVAAKIDHLYRRHRSA KPGKVILLAHDFMFRTQAGASRLRRFIRLMKQRGWSFHTIRHYSRYKPEPLYVAKYYGHSPKELYAA NKQRSFRPVQENRGVESINRSSSSPRPTAQNGKGLMQNRLIEAIRRYNAKEVEHLIDQGASVNQPD SYGHIALNSAVKANSIYLVKKLLAHGADPMIKDARGENALYTAKRYNRGGIERYLKNYYSARSQTSTI VNQVQEKQIAETPHLADSSSTAHTNPLKLLRMRENLYFQGLEHHHHHH*
LEGEND	Underlined: sequence of the protein; start and stop codons; TEV site; His-tag; T7 promoter and T7 terminator
MUTANTS	None

PROTEIN	*Bg*CDA
VECTOR	pET22b(+)
PLASMID MAP	
GENE SEQUENCE	taatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaaggagatat acatatggaagatgatcatggcgtgccgaaaattgttggtgcacgtcagtttctgagcgaaatcaaaagccgtagcaatatt ctgaccagtccggataccagcaccgttcatcgtagcagcgttaaacgtcagaccagtgatggtcgttgtggtcctggtctggg tagctgtaccaattgttgtagtccggatggttattgtggtagcggtccggaatattgtacctcaccggattgtcagtataaatac ggtccgcagtgtgatgcaaatgttaaaccggcaggtagcagcaccgcaaatgttccgcgtccgcatattggtagcgttccgta tggtggtgcaggtatttatgattgtgttacaccgggtacagttgccgttacctttgatgatggtccgtatctgtataccgcagat gttctggataaattcaaacgctataatgccaaagccacctttttttatcaccggcaacaatatgaataaaggtccgattgatgtt actccggcatggtcaaatgttattaaagtgctgagcatcatcatcaactgctatctgatcaacaaaaacaaacgcatggttag cgaaggtcatcaggttgcaagccatacctggacacatcagagcctggataaagttgatagcgaaacctttcagaaccagatc atctataatgaaatggcctttcgcaacatcctgggttattttccgacctatatgcgtccgccttatagccagtgtgccgaagata attgcggtagcaaactgagccgtctgggttatcatgttgtttattttgatgttgataccgcaggctatctgaacgatgatgcaag ccagattcagaaaagcaaaaacatctgggataaaaccattgccggtgtgaaaccgagcaccgatagtctgctggaaattga acatgatatccattatcagaccgcctataacctgaccgattatattctgcagagcatgtatagcaaaggctttaaaagcgtta ccattggtgaatgtctgggtgatccggcagcaaattggtatcgtagcggtaataatcctcattgtgatgcgaccagcaatcag tgtcagacaagtccggcaccgaatgaaccggcagcaccggttaataacaataaaagcgtgagccaggatggtagctgttca accaccgttacctgtcagggtagcacctttggtaattgttgcagcagcagcggttggtgtggtagcagtaccgattattgcggt gcaggttgtgatagcagctcaggtagctgtgatcgtaaagaaaatctgtattttcagggcctcgagcaccaccaccaccacc actgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataacccc ttggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccggat
PROTEIN SEQUENCE	MEDDHGVPKIVGARQFLSEIKSRSNILTSPDTSTVHRSSVKRQTSDGRCGPGLGSCTNCCSPDGYCG SGPEYCTSPDCQYKYGPQCDANVKPAGSSTANVPRPHIGSVPYGGAGIYDCVTPGTVAVTFDDGP YLYTADVLDKFKRYNAKATFFITGNNMNKGPIDVTPAWSNVIKVLSIIINCYLINKNKRMVSEGHQV ASHTWTHQSLDKVDSETFQNQIIYNEMAFRNILGYFPTYMRPPYSQCAEDNCGSKLSRLGYHVVYF DVDTAGYLNDDASQIQKSKNIWDKTIAGVKPSTDSLLEIEHDIHYQTAYNLTDYILQSMYSKGFKSVT IGECLGDPAANWYRSGNNPHCDATSNQCQTSPAPNEPAAPVNNNKSVSQDGSCSTTVTCQGSTF GNCCSSSGWCGSSTDYCGAGCDSSSGSCDRKENLYFQGLEHHHHHH*
LEGEND	Underlined: sequence of the protein; start and stop codons; TEV site; His-tag; T7 promoter and T7 terminator
MUTANTS	None

PROTEIN	*Os*CDA
VECTOR	pET22b(+)
PLASMID MAP	
GENE SEQUENCE	taatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaaggagatat acatatggccctgacctttgatgatggtccgggtgaatataccagccgtctgctggatattctggatggtctgggtgttcatgc aacctttttatcaccggtggtggtaataacaaaggccatattgatgatgaaagcaccggttatccggaagttctgcgtcgtat gcatgcagcaggtcatcagctggcaagccatacctggtcacatcagagcctggatgcactgagcgcaggcggtggtggaat gcacgctgctggtcatcagttagcatcacacacttggagccatcagtcactggatgccctgtcagcagttgcaagcggttggg caggtcgtaaagccgaagttacctttaatgaaatggcctttcgtaacgtgtttggctttttccgacctattttcgtgcaccgtat ctggaatgtggtagcgcatgtcgtacctttctgggcgaatatggttatcatattatcagcaccaacagcgacaccaaagatta tgaaaatgatgatccggcactgattgcaaatagccgtgcacgttttagcgatagcgttgcacgtcaggcaggcgcacagggt tatattgttctggcacatgatgttcatctgcagagcgttgttaatctggcagaatatatggttcgtgaagcacgtagccgtggtt atgaactggttaccgttggtgaatgtctggatgatgcacgtgaaaattggtatcgcgaagcagaaaatctgtattttcagggc ctcgagcaccaccaccaccaccactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgct gagcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccggat
PROTEIN SEQUENCE	MALTFDDGPGEYTSRLLDILDGLGVHATFFITGGGNNKGHIDDESTGYPEVLRRMHAAGHQLASH TWSHQSLDALSAGGGGMHAAGHQLASHTWSHQSLDALSAVASGWAGRKAEVTFNEMAFRNV FGFFPTYFRAPYLECGSACRTFLGEYGYHIISTNSDTKDYENDDPALIANSRARFSDSVARQAGAQGY IVLAHDVHLQSVVNLAEYMVREARSRGYELVTVGECLDDARENWYREAENLYFQGLEHHHHHH*
LEGEND	Underlined: sequence of the protein; start and stop codons; TEV site; His-tag; T7 promoter and T7 terminator
MUTANTS	None

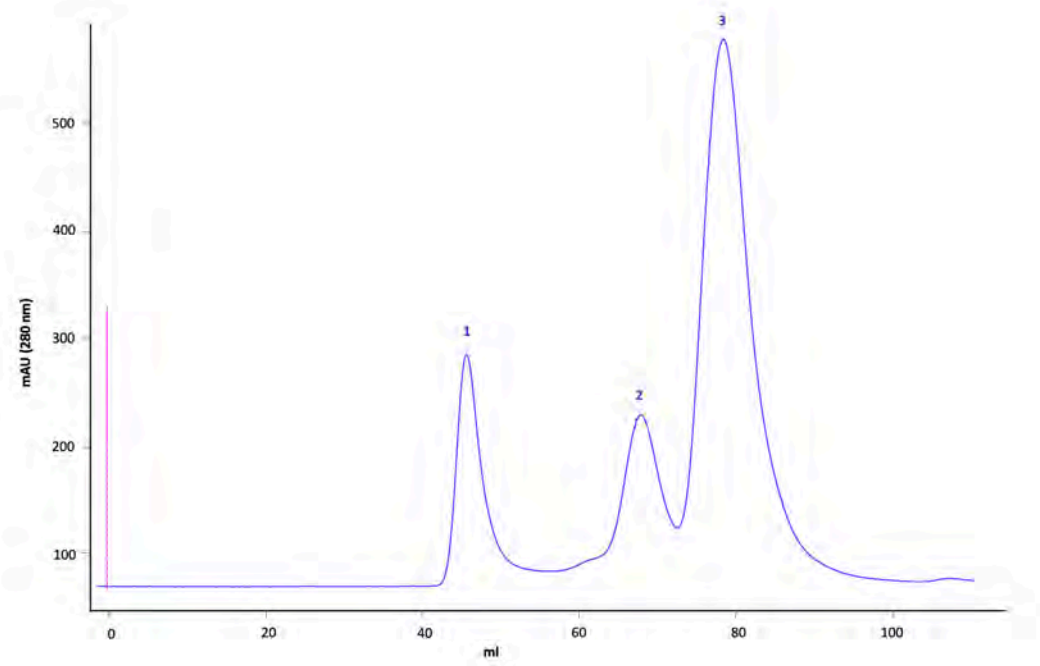
PROTEIN	*Tc*CDA
VECTOR	pET22b(+)
PLASMID MAP	
GENE SEQUENCE	taatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaaggagatat acatatggcaagcgcaccgggtcctgattttccgcctctgagtccgctggttcgtgcagatctggaacgtcgtgatccgagcg aatatgaactgtttcataatgaacagcgtagcggtaatgttaccacacgtgcagttattccaggtggtattacccgttgtacca ccaccaaatgttttagctttacctttgatgatggtccgtataactaccatcagaaaattgcagatgaaattgcaagcattggca cccgtgcaacctttttttgttaatggcaataactatcagtgcatctatgatgatgcagcagttgcagcactgaaatatagttata atgccggtcatcagatttgcagccatacctggtcacatccggatattgaaaccattagccgtgcacgtctggataaagaagtt cagtttattgaagatgccctgtacaaaattctgggtgttgttccggcatgtattcgtccgccttatggtagcgcaagtgatgcaa ccattcagtatctgaataacaaatggggctataatgtggttctgtggaatgcagataccaaagatgcagatggtgcagccgtt agctttagcctgaatcagtatcgtgcactgaaagcaccgaaacatgtgattattctgaatcatgaaaccgttccgaccaccag tagcaaagttattccgcaggcactgcagattgttcaggataatggttatgagccgaaatatatgggcaccgttgcacagacc attggttttgcaccgtataaagttattggcaaaccgggtaaacgtgatagcacctggacctgtgcaggtaaaccgctgcctgg tagcgaaaatctgtattttcagggcctcgagcaccaccaccaccaccactgagatccggctgctaacaaagcccgaaagga agctgagttggctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttgct gaaaggaggaactatatccggat
PROTEIN SEQUENCE	MASAPGPDFPPLSPLVRADLERRDPSEYELFHNEQRSGNVTTRAVIPGGITRCTTTKCFSFTFDDGP YNYHQKIADEIASIGTRATFFVNGNNYQCIYDDAAVAALKYSYNAGHQICSHTWSHPDIETISRARL DKEVQFIEDALYKILGVVPACIRPPYGSASDATIQYLNNKWGYNVVLWNADTKDADGAAVSFSLN QYRALKAPKHVIILNHETVPTTSSKVIPQALQIVQDNGYEPKYMGTVAQTIGFAPYKVIGKPGKRDS TWTCAGKPLPGSENLYFQGLEHHHHHH*
LEGEND	Underlined: sequence of the protein; start and stop codons; TEV site; His-tag; T7 promoter and T7 terminator
MUTANTS	None

PROTEIN	*Bc*Pda
VECTOR	pET22b(+)
PLASMID MAP	
GENE SEQUENCE	taatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaaggagatat acatatgtacaccaacacaccgcataattggggtattccgcgtccgaaaaatgaaaccgttccggatgcaggtaaactgtat accgatctgctgcagaaaaatggtggtttttatctgggtgacaccaagaaaaaagatatctatctgaccttcgataacggcta tgaaaatggttataccggcaaaattctggacgtgctgaaagaaaaaaaagtgcctgcaacctttttgtgacgggtcattata tcaaaacccagaaagacctgctgctgcgtatgaaagatgaaggtcatattattggcaaccacagttggagccatccggatttt accgcagttaatgatgaaaaactgcgtgaagaactgaccagcgttaccgaagaaatcaaaaaagttaccggtcagaaaga agtgaaatacgttcgtccgcctcgtggtgtttttagcgaacgtaccctggcactgaccaaagaaatgggctattataacgtttt ttggagcctggcatttctggattggaaagttgatcagcagcgtggttggcagtatgcacataataatgttatgaccatgattca ccctggtagcattctgctgttacatgcaattagcaaagataatgcagaagccctggccaaaattatcgatgatctgcgtgaaa aaggctaccactttaaaagcctggatgatctggttaaaggtaatcagccggaaaatctgtattttcagggcctcgagcaccac caccaccaccactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataacta gcataaccccttggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccggat
PROTEIN SEQUENCE	MYTNTPHNWGIPRPKNETVPDAGKLYTDLLQKNGGFYLGDTKKKDIYLTFDNGYENGYTGKILDVL KEKKVPATFFVTGHYIKTQKDLLLRMKDEGHIIGNHSWSHPDFTAVNDEKLREELTSVTEEIKKVTG QKEVKYVRPPRGVFSERTLALTKEMGYYNVFWSLAFLDWKVDQQRGWQYAHNNVMTMIHPGSI LLLHAISKDNAEALAKIIDDLREKGYHFKSLDDLVKGNQPENLYFQGLEHHHHHH*
LEGEND	Underlined: sequence of the protein; start and stop codons; TEV site; His-tag; T7 promoter and T7 terminator
MUTANTS	None

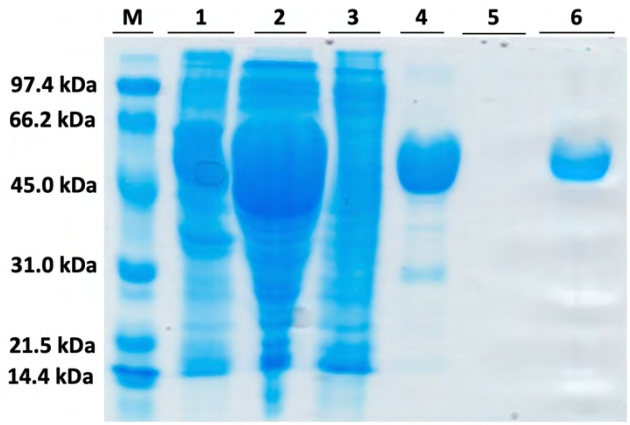
9.2. Expression and purification of *Bs*PdaC-FL

SEC profile (Superdex 200 (16/600)) of *Bs*PdaC-FL

Fraction 2 corresponds to the dimer and fraction 3 corresponds to the monomer of the protein.



SDS-PAGE analysis of *Bs*PdaC-FL purification



M: molecular marker

1: Insoluble fraction after sonication

2: Cell lysate (strep-trap loading)

3: Flow through during loading and washing

4: Strep-trap elution

6: monomeric protein (fraction 3)

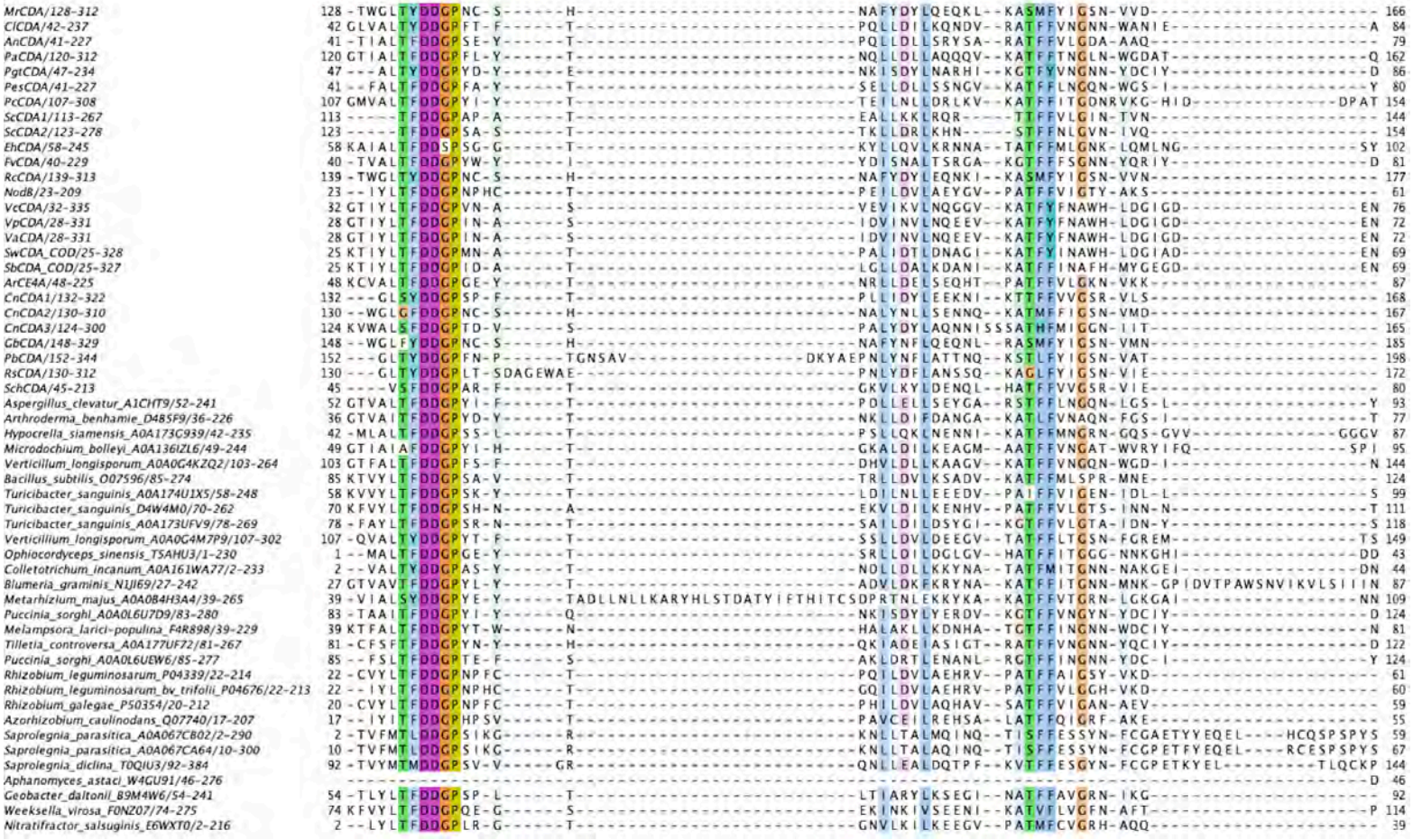
After expression and purification following the same procedures described for the catalytic domain construct, similar yields of 15-20 mg of pure protein per liter of culture were obtained with the full-length enzyme (52.7 kD).

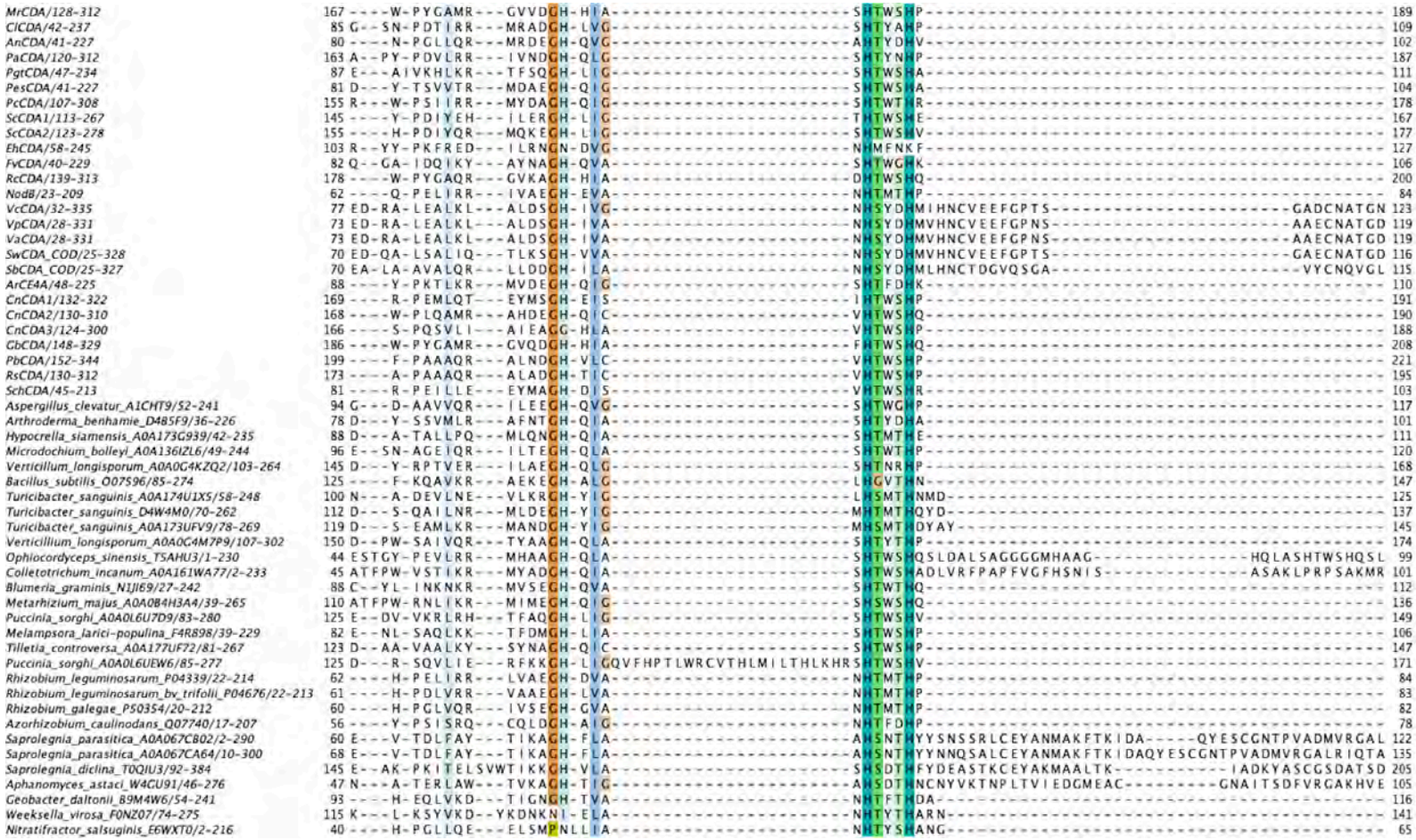
9.3. *Bs*PdaC-CD (wt and D285S mutant) x-ray data collection and refinement statistics

	***Bs*PdaC-CD**	***Bs*PdaC-CD-D285S**
Wavelength	0.9796 Å	0.9796 Å
Resolution range	56.07 - 1.541 (1.596 - 1.541)	21.88 - 1.26 (1.305 - 1.26)
Space group	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1
Unit cell	55.346 61.916 132.226 90 90 90	55.42 61.8 130.99 90 90 90
Total reflections	440385 (44864)	690848 (29571)
Unique reflections	67783 (6689)	120462 (10846)
Multiplicity	6.5 (6.7)	5.7 (2.8)
Completeness (%)	1.00 (1.00)	0.99 (0.89)
Mean I/sigma(I)	11.39 (2.04)	11.01 (1.74)
Wilson B-factor	17.77	12.29
R-merge	0.0877 (0.8135)	0.06804 (0.4831)
R-meas	0.0957 (0.8819)	0.07441 (0.595)
CC1/2	0.998 (0.713)	0.999 (0.707)
CC*	0.999 (0.912)	1 (0.91)
Reflections used in refinement	67782 (6689)	120628 (10846)
Reflections used for R-free	3327 (333)	5936 (546)
R-work	0.1749 (0.2463)	0.1600 (0.3166)
R-free	0.1944 (0.2646)	0.1745 (0.3127)
CC(work)	0.951 (0.841)	0.967 (0.816)
CC(free)	0.958 (0.760)	0.959 (0.830)
Number of non-hydrogen atoms	3697	3830
macromolecules	3277	3334
ligands	22	29
Protein residues	408	408
RMS(bonds)	0.005	0.007
RMS(angles)	0.79	0.98
Ramachandran favored (%)	98	97
Ramachandran allowed (%)	1.9	1.7
Ramachandran outliers (%)	0.49	0.96
Rotamer outliers (%)	0.56	1.1
Clashscore	1.07	2.08
Average B-factor	21.24	17.16
macromolecules	19.78	15.32
ligands	16.44	22.63
solvent	33.59	29.89

9.4. Selection of new chitin deacetylases: alignment-4

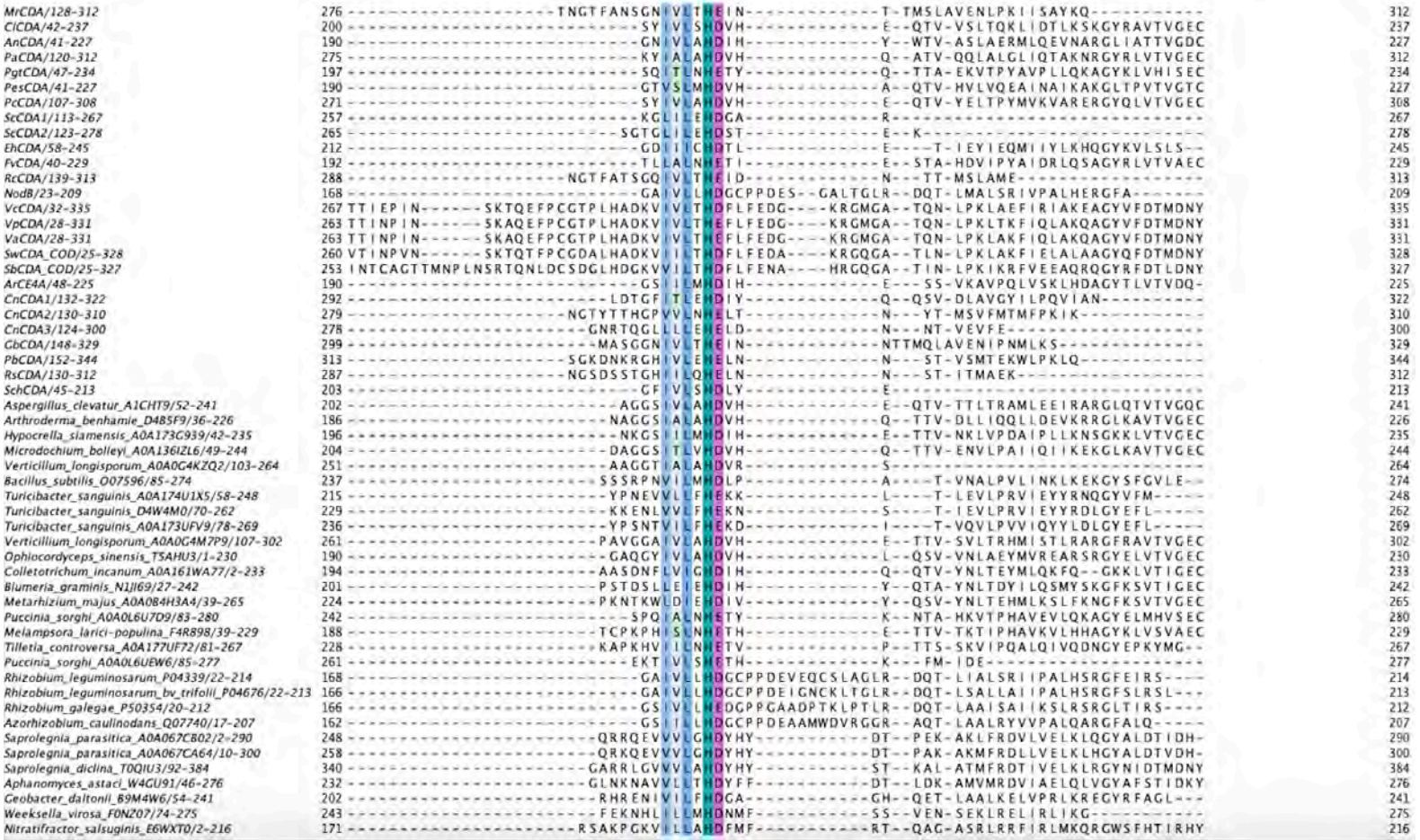
Multiple sequence alignment with the 29 selected sequences in Section 5.1.1.2 and known CDAs and CODs (included in HMM-3) (*Jalview, Mafft G-INS-I, accuracy oriented*).



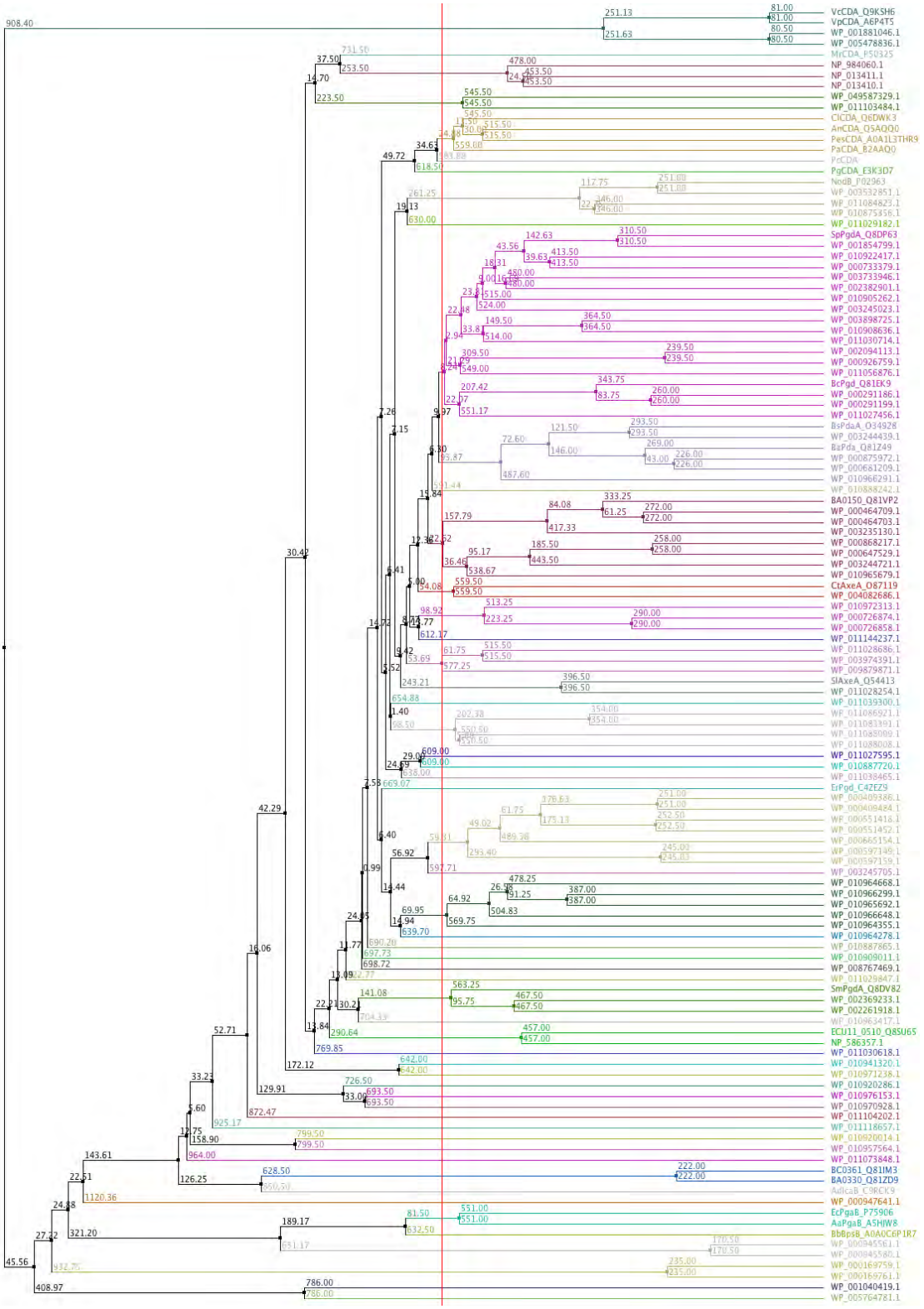



MrCDA/128-312
ClCDA/42-237
AnCDA/41-227
PaCDA/120-312
PgtCDA/47-234
PesCDA/41-227
PcCDA/107-308
ScCDA1/113-267
ScCDA2/123-278
EhCDA/58-245
FvCDA/40-229
RcCDA/139-313
Nod8/23-209
VcCDA/32-335
VpCDA/28-331
VaCDA/28-331
SwCDA_COD/25-328
SbCDA_COD/25-327
ArCE4A/48-225
CnCDA1/132-322
CnCDA2/130-310
CnCDA3/124-300
GbCDA/148-329
PbCDA/152-344
RsCDA/130-312
SchCDA/45-213
Aspergillus_clevatur_A1CHT9/52-241
Arthroderma_benhamie_D4B5F9/36-226
Hypocrella_siamensis_A0A173G939/42-235
Microdochium_bolleyi_A0A136IZL6/49-244
Verticillium_longisporum_A0A0G4KZQ2/103-264
Bacillus_subtilis_O07596/85-274
Turicibacter_sanguinis_A0A174U1X5/58-248
Turicibacter_sanguinis_D4W4M0/70-262
Turicibacter_sanguinis_A0A173UFV9/78-269
Verticillium_longisporum_A0A0G4M7P9/107-302
Ophiocordyceps_sinensis_T5AHU3/1-230
Colletotrichum_incanum_A0A161WA77/2-233
Blumeria_graminis_N1JI69/27-242
Metarhizium_majus_A0A084H3A4/39-265
Puccinia_sorghi_A0A0L6U7D9/83-280
Melampsora_larici-populina_F4R898/39-229
Tilletia_controversa_A0A177UF72/81-267
Puccinia_sorghi_A0A0L6UEW6/85-277
Rhizobium_leguminosarum_P04339/22-214
Rhizobium_leguminosarum_bv_trifolii_P04676/22-213
Rhizobium_galegae_P50354/20-212
Azorhizobium_caulinodans_Q07740/17-207
Saprolegnia_parasitica_A0A067C802/2-290
Saprolegnia_parasitica_A0A067CA64/10-300
Saprolegnia_diclina_T0QIU3/92-384
Aphanomyces_astaci_W4GU91/46-276
Geobacter_daltonii_B9M4W6/54-241
Weeksella_virosa_F0NZ07/74-275
Nitratifractor_salsuginis_E6WXT0/2-216


MrCDA/128-312
ClCDA/42-237
AnCDA/41-227
PaCDA/120-312
PgtCDA/47-234
PesCDA/41-227
PcCDA/107-308
ScCDA1/113-267
ScCDA2/123-278
EhCDA/58-245
FvCDA/40-229
RcCDA/139-313
NodB/23-209
VcCDA/32-335
VpCDA/28-331
VaCDA/28-331
SwCDA_COD/25-328
SbCDA_COD/25-327
ArCE4A/48-225
CnCDA1/132-322
CnCDA2/130-310
CnCDA3/124-300
GbCDA/148-329
PbCDA/152-344
RsCDA/130-312
SchCDA/45-213
Aspergillus_clevatur_A1CHT9/52-241
Arthroderma_benhamie_D4B5F9/36-226
Hypocrella_siamensis_A0A173G939/42-235
Microdochium_bolleyi_A0A136IZL6/49-244
Verticillum_longisporum_A0A0G4KZQ2/103-264
Bacillus_subtilis_O07596/85-274
Turicibacter_sanguinis_A0A174U1X5/58-248
Turicibacter_sanguinis_D4W4M0/70-262
Turicibacter_sanguinis_A0A173UFV9/78-269
Verticillium_longisporum_A0A0G4M7P9/107-302
Ophiocordyceps_sinensis_T5AHU3/1-230
Colletotrichum_incanum_A0A161WA77/2-233
Blumeria_graminis_N1JI69/27-242
Metarhizium_majus_A0A0B4H3A4/39-265
Puccinia_sorghi_A0A0L6U7D9/83-280
Melampsora_larici-populina_F4R898/39-229
Tilletia_controversa_A0A177UF72/81-267
Puccinia_sorghi_A0A0L6UEW6/85-277
Rhizobium_leguminosarum_P04339/22-214
Rhizobium_leguminosarum_bv_trifolii_P04676/22-213
Rhizobium_galegae_P50354/20-212
Azorhizobium_caulinodans_Q07740/17-207
Saprolegnia_parasitica_A0A067CB02/2-290
Saprolegnia_parasitica_A0A067CA64/10-300
Saprolegnia_diclina_T0QIU3/92-384
Aphanomyces_astaci_W4GU91/46-276
Geobacter_daltonii_B9M4W6/54-241
Weeksella_virosa_F0NZ07/74-275
Nitratifractor_salsuginis_E6WXT0/2-216


MrCDA/128-312 276 TNGTFANSGNIVLTHEIN T-TMSLAVENLPKIISAYKQ 312
ClCDA/42-237 200 SYIVLSHDVH E--QTV-VSLTQKLIDTLKSKGYRAVTVGEC 237
AnCDA/41-227 190 GNIVLAHDIH Y--WTV-ASLAERMLQEVNARGLIATTVGDC 227
PaCDA/120-312 275 KYIALAHDVH Q--ATV-QQLALGLIQTAKNRGYRLVTVGEC 312
PgtCDA/47-234 197 SQITLNHETY Q--TTA-EKVTPYAVPLLQKAGYKLVHISEC 234
PesCDA/41-227 190 GTVSLMHDVH A--QTV-HVLVQEAINAIKAKGLTPVTVGTC 227
PcCDA/107-308 271 SYIVLAHDVH E--QTV-YELTPYMVKVARERGYQLVTVGEC 308
ScCDA1/113-267 257 KGLILEHDGA R 267
ScCDA2/123-278 265 SGTGLILEHDST E--K 278
EhCDA/58-245 212 GDIIICHDTL E----T-IEYIEQMIIYLKHQGYKVLSLS-- 245
FvCDA/40-229 192 TLLALNHETI E--STA-HDVIPYAIDRLQSAGYRLVTVAEC 229
RcCDA/139-313 288 NGTFATSGQIVLTHEID N---TT-MSLAME 313
NodB/23-209 168 GAIVLLHDGCPPDES--GALTGLR--DQT-LMALSRIVPALHERGFA 209
VcCDA/32-335 267 TTIEPIN------SKTQEFPCGTPLHADKVIVLTHDFLFEDG----KRGMGA--TQN-LPKLAEFIRIAKEAGYVFDTMDNY 335
VpCDA/28-331 263 TTINPIN------SKAQEFPCGTPLHADKVIVLTHEFLFEDG----KRGMGA--TQN-LPKLTKFIQLAKQAGYVFDTMDNY 331
VaCDA/28-331 263 TTINPIN------SKAQEFPCGTPLHADKVIVLTHEFLFEDG----KRGMGA--TQN-LPKLAKFIQLAKQAGYVFDTMDNY 331
SwCDA_COD/25-328 260 VTINPVN------SKTQTFPCGDALHADKVIILTHDFLFEDA----KRGQGA--TLN-LPKLAKFIELALAAGYQFDTMDNY 328
SbCDA_COD/25-327 253 INTCAGTTMNPLNSRTQNLDCSDGLHDGKVVILTHDFLFENA----HRGQGA--TIN-LPKIKRFVEEAQRQGYRFDTLDNY 327
ArCE4A/48-225 190 GSIILMHDIH E---SS-VKAVPQLVSKLHDAGYTLVTVDQ- 225
CnCDA1/132-322 292 LDTGFITLEHDIY Q--QSV-DLAVGYILPQVIAN 322
CnCDA2/130-310 279 NGTYTTHGPVVLNHELT N---YT-MSVFMTMFPKIK 310
CnCDA3/124-300 278 GNRTQGLLLLEHELD N---NT-VEVFE 300
GbCDA/148-329 299 MASGGNIVLTHEIN NTTMQLAVENIPNMLKS 329
PbCDA/152-344 313 SGKDNKRGHIVLEHELN N---ST-VSMTEKWLPKLQ 344
RsCDA/130-312 287 NGSDSSTGHIILQHELN N---ST-ITMAEK 312
SchCDA/45-213 203 GFIVLSHDLY E 213
Aspergillus_clevatur_A1CHT9/52-241 202 AGGSIVLAHDVH E--QTV-TTLTRAMLEEIRARGLQTVTVGQC 241
Arthroderma_benhamie_D485F9/36-226 186 NAGGSIALAHDVH Q--TTV-DLLIQQLLDEVKRRGLKAVTVGEC 226
Hypocrella_siamensis_A0A173G939/42-235 196 NKGSIILMHDIH E--TTV-NKLVPDAIPLLKNSGKKLVTVGEC 235
Microdochium_bolleyi_A0A136IZL6/49-244 204 DAGGSITLVHDVH Q--TTV-ENVLPAIIQIIKEKGLKAVTVGEC 244
Verticillium_longisporum_A0A0C4KZQ2/103-264 251 AAGGTIALAHDVR S 264
Bacillus_subtilis_O07596/85-274 237 SSSRPNVILMHDLP A----T-VNALPVLINKLKEKGYSFGVLE-- 274
Turicibacter_sanguinis_A0A174U1X5/58-248 215 YPNEVVLLFHEKK L----T-LEVLPRVIEYYRNQGYVFM----- 248
Turicibacter_sanguinis_D4W4M0/70-262 229 KKENLVVLFHEKN S----T-IEVLPRVIEYYRDLGYEFL----- 262
Turicibacter_sanguinis_A0A173UFV9/78-269 236 YPSNTVILFHEKD I----T-VQVLPVVIQYYLDLGYEFL----- 269
Verticillium_longisporum_A0A0G4M7P9/107-302 261 PAVGGAIVLAHDVH E--TTV-SVLTRHMISTLRARGFRAVTVGEC 302
Ophiocordyceps_sinensis_T5AHU3/1-230 190 GAQGYIVLAHDVH L--QSV-VNLAEYMVREARSRGYELVTVGEC 230
Colletotrichum_incanum_A0A161WA77/2-233 194 AASDNFLVIGHDIH Q--QTV-YNLTEYMLQKFQ--GKKLVTIGEC 233
Blumeria_graminis_N1JI69/27-242 201 PSTDSLLEIEHDIH Y--QTA-YNLTDYILQSMYSKGFKSVTIGEC 242
Metarhizium_majus_A0A0B4H3A4/39-265 224 PKNTKWLDIEHDIV Y--QSV-YNLTEHMLKSLFKNGFKSVTVGEC 265
Puccinia_sorghi_A0A0L6U7D9/83-280 242 SPQIALNHETY K--NTA-HKVTPHAVEVLQKAGYELMHVSEC 280
Melampsora_larici-populina_F4R898/39-229 188 TCPKPHISLNHETH E--TTV-TKTIPHAVKVLHHAGYKLVSVAEC 229
Tilletia_controversa_A0A177UF72/81-267 228 KAPKHVILLNHETV P--TTS-SKVIPQALQIVQDNGYEPKYMG-- 267
Puccinia_sorghi_A0A0L6UEW6/85-277 261 EKTIVLSHETH K---FM-IDE 277
Rhizobium_leguminosarum_P04339/22-214 168 GAIVLLHDGCPPDEVEQCSLAGLR--DQT-LIALSRIIPALHSRGFEIRS---- 214
Rhizobium_leguminosarum_bv_trifolii_P04676/22-213 166 GAIVLLHDGCPPDEIGNCKLTGLR--DQT-LSALLAIIPALHSRGFSLRSL--- 213
Rhizobium_galegae_P50354/20-212 166 GSIVLLHEDGPPGAADPTKLPTLR--DQT-LAAISAIIKSLRSRGLTIRS---- 212
Azorhizobium_caulinodans_Q07740/17-207 162 GSIILLHDGCPPDEAAMWDVRGGR--AQT-LAALRYVVPALQARGFALQ----- 207
Saprolegnia_parasitica_A0A067CB02/2-290 248 QRRQEVVVLGHDYHY DT--PEK-AKLFRDVLVELKLQGYALDTIDH- 290
Saprolegnia_parasitica_A0A067CA64/10-300 258 QRKQEVVVLGHDYHY DT--PAK-AKMFRDLLVELKLHGYALDTVDH- 300
Saprolegnia_diclina_T0QIU3/92-384 340 GARRLGVVVLAHDYHY ST--KAL-ATMFRDTIVELKLRGYNIDTMDNY 384
Aphanomyces_astaci_W4GU91/46-276 232 GLNKNAVVLLTHDYFF DT--LDK-AMVMRDVIAELQLVGYAFSTIDKY 276
Geobacter_daltonii_B9M4W6/54-241 202 RHRENIVILFHDGA GH--QET-LAALKELVPRLKREGYRFAGL--- 241
Weeksella_virosa_F0NZ07/74-275 243 FEKNHLILLMHDNMF SS--VEN-SEKLRELIRLIKG 275
Nitratifractor_salsuginis_E6WXT0/2-216 171 RSAKPGKVILLAHDFMF RT--QAG-ASRLRRFIRLMKQRGWSFHTIRHY 216

9.5. Phylogenetic tree with 210 sequences of possible peptidoglycan deacetylases retrieved from databases by BLAST-P and psi-blast searches



9.6. Review on chitin deacetylases

Polymers, 10, 352. 2018

Chitin Deacetylases: Structures, Specificities, and Biotech Applications.

Laia Grifoll-Romero, Sergi Pascual, Hugo Aragunde, Xevi Biarnés and Antoni Planas

Laboratory of Biochemistry, Institut Químic de Sarrià, Universitat Ramon Llull, 08017 Barcelona, Spain.

1. Introduction
2. Chitin Deacetylases and the Carbohydrate Esterase Family (CE4)
3. Function and Specificity of CE4 Chitin Deacetylases
4. Structural Determinants of Activity and Specificity
5. Applications of Chitin Deacetylases

polymers

MDPI

Review

Chitin Deacetylases: Structures, Specificities, and Biotech Applications

Laia Grifoll-Romero, Sergi Pascual, Hugo Aragunde, Xevi Biarnés and Antoni Planas *

Laboratory of Biochemistry, Institut Químic de Sarrià, Universitat Ramon Llull, 08017 Barcelona, Spain; laiagrifollr@iqs.edu (L.G.-R.); sergipascualt@iqs.edu (S.P.); hugoaragunde@gmail.com (H.A.); xavier.biarnes@iqs.edu (X.B.)

* Correspondence: antoni.planas@iqs.edu

Received: 19 February 2018; Accepted: 19 March 2018; Published: 22 March 2018

Abstract: Depolymerization and de-*N*-acetylation of chitin by chitinases and deacetylases generates a series of derivatives including chitosans and chitooligosaccharides (COS), which are involved in molecular recognition events such as modulation of cell signaling and morphogenesis, immune responses, and host-pathogen interactions. Chitosans and COS are also attractive scaffolds for the development of bionanomaterials for drug/gene delivery and tissue engineering applications. Most of the biological activities associated with COS seem to be largely dependent not only on the degree of polymerization but also on the acetylation pattern, which defines the charge density and distribution of GlcNAc and $GlcNH_2$ moieties in chitosans and COS. Chitin de-*N*-acetylases (CDAs) catalyze the hydrolysis of the acetamido group in GlcNAc residues of chitin, chitosan, and COS. The deacetylation patterns are diverse, some CDAs being specific for single positions, others showing multiple attack, processivity or random actions. This review summarizes the current knowledge on substrate specificity of bacterial and fungal CDAs, focusing on the structural and molecular aspects of their modes of action. Understanding the structural determinants of specificity will not only contribute to unravelling structure-function relationships, but also to use and engineer CDAs as biocatalysts for the production of tailor-made chitosans and COS for a growing number of applications.

Keywords: chitin deacetylases; chitosan; chitooligosaccharides; carbohydrate esterases; structure; substrate specificity; deacetylation pattern

1. Introduction

Chitin is a linear polysaccharide of β(1→4)-linked *N*-acetylglucosamine monomers. It was first isolated from fungi in 1811 [1] and its structure was determined in 1929 [2]. Chitin is a major structural component of the exoskeletons of arthropods (insects and crustaceans), of the endoskeletons of mollusks, and it is also found in the cell walls of fungi and diatoms [1,3,4]. It is one of the most abundant organic molecules after cellulose, and the most abundant natural amino polysaccharide. Chitin is present as ordered macrofibrils, mainly in two allomorphs: α-chitin, with antiparallel chains [5], is the most abundant and it is isolated from the exoskeleton of crustaceans, particularly from shrimps and crabs; and β-chitin, with parallel chains [6], is present in the cell walls of diatoms and in the skeletal structures of cephalopods, and commonly extracted from squid pens. β-Chitin is easily converted to α-chitin by alkaline treatment followed by flushing in water [5]. Chitin is also found as γ-chitin in fungi and yeast, which is a combination of the α and β allomorphs [7,8].

Depolymerization (hydrolysis) of chitin by chitinases results in chitooligosaccharides (COS), and de-*N*-acetylation of chitin and COS yields chitosan and partially acetylated COS (paCOS) or fully deacetylated glucosamine oligomers (Figure 1). In nature, the deacetylation of chitin is almost never

complete, and chitosan refers to a family of heteropolysaccharides composed of *N*-acetylglucosamine and glucosamine, characterized by their degree of polymerization (DP), degree of acetylation (DA), and pattern of acetylation (PA). Only some fungi of the *Zygomycota*, *Basidiomycota* and *Ascomycota* phyla have been reported to be capable of naturally producing chitosans [9]. The free amino groups of the deacetylated units in the polymer are protonated at slightly acidic pH, thus making chitosans the only known natural polycationic polysaccharides [1,3,9]. They interact with polyanionic biomolecules such as polyanionic phospholipidic membranes, glycosaminoglycans at cell surfaces, proteins, and nucleic acids. Depolymerization by hydrolysis of the β-1,4-linkages of chitin and chitosan yields their respective oligosaccharides (COS, paCOS) [10]. Whereas chitin and chitosans act as structural polymers, their oligomers are involved in molecular recognition events, such as cell signaling and morphogenesis, and act as immune response elicitors and host-pathogen mediators [11–15]. The catabolism of chitin and chitosan is summarized in Figure 1. Chitin and chitosan oligosaccharides are further degraded following different organism-dependent pathways to end up in the central energy metabolism.

The de-*N*-acetylation of chitin and chitooligosaccharides is catalyzed by chitin deacetylases (CDAs), which exhibit different substrate specificities leading to fully or partially deacetylated products with diverse degrees of acetylation (DA) and patterns of acetylation (PA). In addition to the role CDAs play in the biology of their natural organisms, there is a growing interest in the biochemical characterization of CDAs in order to use them as biocatalysts for the production of partially deacetylated chitooligosaccharides (paCOS) as bioactive molecules in different application fields, or to inhibit them since they are potential targets against pathogenic microorganisms. The aim of this review is to analyze the current knowledge on the biochemistry of chitin deacetylases with regard to their substrate specificity. Although a large number of enzymes have been experimentally identified as chitin deacetylases able to deacetylate chitin either in vivo or in vitro, few of them have been analyzed at the protein level. Here we will focus on those CDAs with characterized activity on chitooligosaccharides (COS) to address the issue of substrate specificity and the deacetylation pattern.

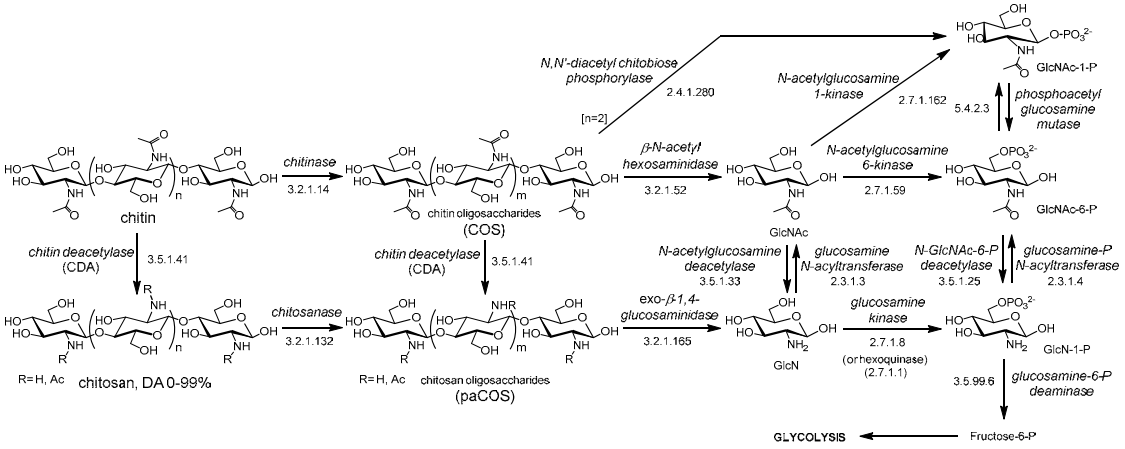


Figure 1. Chitin catabolism. GlcNAc: *N*-acetylglucosamine; GlcN: glucosamine; DA: degree of acetylation. COS: chitooligosaccharides (or chitin oligosaccharides), paCOS: partially acetylated chitooligosaccharides (or chitosan oligosaccharides).

2. Chitin Deacetylases and the Carbohydrate Esterase Family 4 (CE4)

Chitin deacetylases (CDAs, EC 3.5.1.41) and chitooligosaccharides deacetylases (CODs, EC 3.5.1.105) are classified in the carbohydrate esterase family 4 (CE4) in the CAZY database (Carbohydrate Active Enzymes, www.cazy.org) [16]. The CE4 family also contains peptidoglycan *N*-acetylglucosamine deacetylases (EC 3.5.1.104), peptidoglycan *N*-acetylmuramic acid deacetylases (EC 3.5.1.-), poly-β-1,6-*N*-acetylglucosamine deacetylase (EC 3.5.1.-), and some acetyl xylan esterases (EC 3.1.1.72) [17]. These enzymes share a conserved region known as the NodB homology domain

due to its similarity to the NodB oligosaccharide deacetylase, one of the first CE4 enzymes to be characterized [18]. Most currently reported and characterized CDAs and CODs are CE4 enzymes, with the exception of diacetylchitobiose deacetylases (Dacs) from archaea and a COD from *Bacillus cereus* (*Bc*ZBP) that belong to the CE14 family [19,20]. Few other enzymes, such as insect CDAs and a COD from *E. coli* (ChbG) [21] are in the group of "non-classified" in the CAZY database since they do not share sequence similarities to the other CDA families.

The deacetylase activity from extracts of the fungus *Mucor rouxii* was the first active CDA identified and partially purified in the mid 1970s [22,23]. Later on, the NodB from a rhizobium species was the first biochemically characterized COD in 1993 [18]. Many other CDAs and CODs were later identified and purified from very diverse organisms, including archaea, marine bacteria, fungi, and insects [24]. These enzymes are diverse in their biochemical properties: molecular masses in the range from 12 to 150 kDa, acidic isoelectric points (pI from 2.7 to 4.8), optimum pH for activity from 4.5 to 12, and significant thermal stability, with optimum temperatures for activity in the range from 30 to 60 °C. Most CDAs are highly inactive on crystalline chitin due to the inaccessibility of the acetyl groups in the tightly packed chitin structure, and have a preference for soluble forms of chitins such as glycol-chitin or chitin oligomers, as well as partially deacetylated chitin (chitosans). It has recently been shown that CDA activity on crystalline chitin is greatly enhanced by oxidative cleavage of the surface polymer chains by lytic polysaccharide monooxygenases (LPMO) [25]. Some CDAs contain carbohydrate binding modules (CBM) fused to the catalytic domain that seem to enhance the deacetylase activity by increasing the accessibility of the substrate to the catalytic domain [26].

CDAs are localized in different cellular compartments, in the periplasm, in the cytosol, or secreted as extracellular enzymes. Periplasmic fungal CDAs are generally tightly coupled to a chitin synthase to rapidly deacetylate newly synthesized chitin before their maturation and crystallization. Extracellular fungal CDAs are secreted to alter the physicochemical properties of the cell wall, which results in protection against exogenous chitinases, or initiates sporulation or autolysis. In bacteria, CDAs are either intracellular, as those involved in Nod factors biosynthesis in *Rhizobium* species, or extracellular, as those involved in the catabolism of chitin in marine bacteria [24,27,28].

Some CE4 enzymes classified in a specific subfamily also show activity on typical substrates from other subfamilies. Peptidoglycan GlcNAc deacetylases, involved in the de-*N*-acetylation of the bacterial cell wall peptidoglycan with critical functions in the maturation and turnover of peptidoglycan and in bacterial pathogenicity, are also active on COS. Some CDAs have activity on acetylxylan, as well as some acetylxylan esterases are active on COS, which makes their classification doubtful in some cases.

3. Function and Specificity of CE4 Chitin Deacetylases

3.1. Deacetylation Patterns

Chitin deacetylases exhibit diverse deacetylation patterns, reflecting different substrate specificities and pattern recognition on their linear substrates. The mechanisms of action of enzymes that modify in-chain units on a linear polysaccharide are commonly classified as multiple-attack, multiple-chain, and single-chain mechanisms [29]. In the multiple-attack mechanism, binding of the enzyme to the polysaccharide chain is followed by a number of sequential deacetylations, after which the enzyme binds to another region of the polymeric chain. (i.e., *M. rouxii* [30,31]). On a polymeric substrate, this mechanism will result in a block-copolymer structure with blocks of $GlcNH_2$ units within the GlcNAc chain. On COS, it will usually result in full deacetylation of the oligomer. In the multiple-chain mechanism, the enzyme forms an active enzyme-polymer complex and catalyzes the hydrolysis of only one acetyl group before it dissociates and forms a new active complex (i.e., *C. lindemuthianum* CDA [32,33]). It will result in a random distribution of the $GlcNH_2$ and GlcNAc units along the polymeric chain or, in the case of COS substrates, it will render a number of partially deacetylated oligosaccharide intermediates ending in a specific deacetylation pattern or in full deacetylation, depending on the enzyme and the substrate. Finally, a single-chain mechanism includes

processive enzymes in which a number of catalytic events occur on a single substrate molecule leading to sequential deacetylation. Some bacterial chitooligosaccharides deacetylases (CODs), which are specific for a single position leading to mono-deacetylated products (i.e., *Rhizobium* NodB or *Vibrio* CDA or COD) are also included in the last group.

A fundamental question is how these enzymes define their action pattern. This is relevant not only to understand their biological functions but also to use CDAs (native and engineered variants) as biocatalysts for the production of chitosans with non-random deacetylation patterns, and partially deacetylated COS with tailored patterns of acetylation (see Section 6). A structural model on the determinants of substrate specificity is currently emerging from studies on substrate specificity, determination of 3D structures of enzyme-substrate complexes, and multiple sequence alignments. Many CDAs, particularly from fungal origin, have been identified as involved in chitin deacetylation in vivo, but only few of them have been characterized with regard to substrate specificity and mode of action. To the best of our knowledge, Table 1 compiles the CDA enzymes in family CE4 that have been biochemically characterized and have reported activity on COS substrates, some of them with solved three-dimensional structure by X-ray crystallography. Relevant information on their biological function and substrate specificity is summarized below.

CDAs participate in diverse biological processes, which include cell wall morphogenesis and host-pathogen interaction in fungi, generation of signaling molecules in bacteria, and participation in the catabolism of chitin as carbon, nitrogen, and energy sources in marine bacteria and fungi. CDAs were thought to be restricted to fungi and bacteria until a first report in 1986 on their presence in arthropods [34]. CDAs seem to be widely present in insects, in cuticles and the perithrophic midgut matrix, but little is known on the function and properties of insect CDAs [35,36], and they are not included here because scarce information on substrate specificities has been reported.

3.2. Fungal CDAs

Fungal CDAs are involved in fungal nutrition, morphogenesis and development [27,29], participating in cell wall formation and integrity [37], in spore formation [38], germling adhesion [39], fungal autolysis [40], and in defense mechanisms for host infection [41].

Fungi that have chitosan (in addition to chitin) as a structural component of the cell wall, secrete CDAs to the periplasmic space that contribute to chitosan biosynthesis from nascent chitin synthesized by chitin synthases. It occurs during exclusive periods corresponding to their particular biological role in the cell cycle of the fungal species: during cell wall formation (i.e., *M. rouxi* [42] and *A. coerulea* [43]), during sporulation (i.e., *S. cerevisiae* [38,44]), or during vegetative growth (i.e., *C. neoformas* [37]).

Pathogenic fungi secrete CDAs during fungal hyphae penetration to evade plant defense mechanisms and gain access to host tissues. Plants secrete chitinases to break the fungal cell-wall chitin down to chitooligosaccharides (COS), and the released COS are recognized by plant chitin-specific receptors, triggering resistance responses [41]. COS elicitation of resistance mechanisms involve activation of host defense genes [45,46]. There is cumulative evidence that fungi evade plant defense mechanisms by partially deacetylating either their exposed cell wall chitin or the chitooligosaccharides produced by the action of plant chitinases. In both cases, the resulting partially deacetylated oligomers are not well recognized by the specific plant receptors reducing or preventing the elicitation of the defense responses [41,47–49].

Filamentous fungi undergo autolysis by self-digestion of aged hyphal cultures due to carbon starvation [40,50]. During this event there is an increased presence of hydrolytic enzymes, especially those involved in cell wall degradation, and CDAs are secreted to the extracellular medium to deacetylate the chitin oligomers produced by chitinases (i.e., *A. nidulans* [51,52]).

Table 1. CDAs and CODs with characterized activity on COS.

Enzyme [1]	Organism	ID [2]	PDB [3] [Ref.]	COS Substrates [4]	Ref. [5] COS	Metal [6]	PA [7] (on A_n)
*Mr*CDA	*Mucor rouxii*	P50325		≥DP3	[31]	Zn^{2+}	D_n
*Cl*CDA	*Colletotrichum lindemuthianum*	Q6DWK3	2IW0 [53]	DP6 > DP5 > DP4 > DP3 > DP2	[32]	Co^{2+}, Zn^{2+}	D_n
*An*CDA	*Aspergillus nidulans*	Q5AQQ0	2Y8U [25]	DP2 > DP3 > DP4 > DP5	[25]	Co^{2+}	D_n
*Pa*CDA	*Podospora anserine*	XP_001912680.1		≥DP2	[26]	Zn^{2+}	D_n
*Pgt*CDA	*Puccinia graminis*	XP_003323413.1		DP6 > DP5 > DP4	[54]	n.r. [7]	AAD_{n-2}
*Pes*CDA	*Pestolotiopsis* sp.	APH81274.1		DP6-DP5-DP4	[49]	n.r.	$AAD_{n-3}A$
*Pc*CDA	*Pochonia chlamydosporia*			DP5 > DP4	[55]	n.r.	$ADDA_{n-3}$
*Sc*CDA1	*Saccharomyces cerevisiae*	Q06702		DP4, DP6	[56]	n.r.	n.r.
*Sc*CDA2	*Saccharomyces cerevisiae*	Q06703		DP6 > DP5 > DP4 > DP3 > DP2	[57]	Co^{2+}	n.r.
*Mo*CDA *	*Mortierella* sp.			DP7 > DP6 > DP5 > DP4 > DP3 > DP2	[58]	(Co^{2+})	n.r.
*Acoe*CDA *	*Absidia coerulea*			DP5 > DP4 > DP3	[43]	n.r.	n.r.
*Acor*CDA *	*Absidia corymbifera*			DP7 > DP6 > DP5 > DP4 > DP3 > DP2	[59]	(Co^{2+}, Ca^{2+}, Mg^{2+})	n.r.
*Fv*CDA	*Flammulina velutipes*	BAE92728.1		DP5 > DP4 > DP3 > DP2	[60]	(Co^{2+}, Ca^{2+}, Zn^{2+})	n.r.
*Po*CDA *	*Penicillium oxilicum*			DP5 >> DP3 > DP2	[61]	(Co^{2+} Cu^{2+})	n.r.
*Af*CDA *	*Aspergillus flavus*			DP4	[62]	(Zn^{2+}, Mn^{2+})	n.r.
*Sb*CDA *	*Scopulariopsis brevicaulis*			DP6 > DP5 > DP4 > DP3 > DP2	[63]	n.r.	n.r.
*Rc*CDA	*Rhizopus circinans*	A7UMZ0		DP6	[64]	(Mn^{2+}, Mg^{2+})	n.r.
*Rs*CDA	*Rhizopus stolonifer (nigricans)*	Q32XH4		n.r.	[64]		
*Gb*CDA	*Gongronella butleri*	Q8J2N6		n.r.	[65]		
*Pb*CDA	*Phycomyces blakesleeanus*	Q9P4U2		n.r.	[66]		
*Sch*CDA	*Schizophyllum commune*	Q9P453		n.r.	[67]		
*Cn*CDA1, 2, 3	*Cryptococcus neoformans*	Q5KFG8, Q5KIC2, P0CP76		n.r.	[37]		
*Eh*CDA	*Entamoeba histolytica*	XP_656356.1		DP5, DP6	[68]	n.r.	n.r.
NodB	*Sinorhizobium meliloti*	P02963		DP5 > DP2 (DP4, DP3)	[18]	Mn^{2+} Mg^{2+}	DA_{n-1}
*Vc*COD (*Vc*CDA)	*Vibrio cholera*	Q9KSH6	4NY2 [69]	DP2 > DP3 > DP4 > DP5 > DP6	[70]	Zn^{2+}	ADA_{n-2}
*Vp*COD	*Vibrio parahaemolyticus*	Q9KSH6	3WX7 [71]	DP2 > DP3	[72]	Zn^{2+}	n.r.
*Va*COD	*Vibrio alginolyticus*	Q9KSH6		DP2	[73]	Zn^{2+}	AD
*Sw*COD	*Shewanella woodyi*	ACA84860.1		DP2 > DP3 > DP4	[74]	n.r.	AD; [ADA_{n-2}]
*Sb*COD	*Shewanella baltica*	ABN60929.1		DP2 > DP4 > DP3	[75]	n.r.	AD; [ADA_{n-2}]
*Ar*CE4A	*Arthrobacter* sp.	A0A2C8C1T7	5LFZ [76]	DP5 > DP6 ≈ DP4 > DP3 >> DP2	[76]	Ni^{2+} [8]	$D_{n-1}A$

[1] Characterized recombinant enzymes, except those with an asterisk (*) that have been characterized from the native organism and are not included in the sequence alignment, Figure 4. [2] Uniprot or GenBank accession code, [3] PDB accession code and publication of the 3D structure. [4] Activity on chitooligosaccharides (COS) as a function of the degree of polymerization (DP). [5] Selected publication on substrate specificity. [5] Native metal cation or, in parenthesis, metals that enhanced the enzyme activity when added in the reaction buffer. [6] Pattern of acetylation (PA): structure of the main final deacetylated product (A: GlcNAc; D: $GlcNH_2$). Other patterns of acetylation with specific substrates are given in the text. [7] n.r.: not reported. [8] Native metal unknown, Ni^{2+} probably from purification/crystallization.

In general, fungal CDAs deacetylate soluble forms of chitin such as glycol-chitin and chitosans of variable DA, but they are inactive, or show low activity, towards insoluble chitins such as crystalline α- and β-chitin and colloidal chitin. Pretreatment of chitin to make surface fibrils more accessible may result in increased deacetylase activity. It has been recently shown that the activity of an *A. nidulans* CDA on crystalline chitin was enhanced by a lytic polysaccharide monooxygenase (LPMO) that increases substrate accessibility by oxidative cleavage of the chitin chains [25]. Some CDAs also appear to be active on acetylxylan (i.e., *An*CDA), but any of them act on peptidoglycans, typical substrates of other CE4 family members. In addition to polymeric substrate, not many CDA have been analyzed with COS as substrates (Table 1). The analysis of the products from enzymatic deacetylation with regards to the extent and pattern of deacetylated provides information about the specificity and mode of action of these enzymes. The first seven entries in Table 1 correspond to CDAs for which the deacetylation pattern on COS has been reported, whereas the rest of the entries are CDAs active on COS but, to the best of our knowledge, the structure of the deacetylated products has not been analyzed.

Mucor rouxii (*Amylomyces rouxii*) (*Mr*CDA). The dimorphic fungus *M. rouxii* has a cell wall mainly composed of chitin, chitosan, and mucoric acid. While chitin accounts for 10% of the total dry weight of the cell wall, chitosan reaches 30% [77]. The *M. rouxii* CDA enzyme was initially found in the cytosol [23], but it is also secreted into the periplasm where it participates in a tandem synthetic mechanism that involves a chitin synthase and a chitin deacetylase working consecutively and synchronously for synthesis and deposition of chitosan polymers at the outer membrane [78]. Decoupling this mechanism prevents the formation of chitosan [42,79]. *M. rouxii*, like other fungi, has been identified as a suitable microorganism for chitosan production by means of biofermentation processes [80,81]. *Mr*CDA is a monomeric high-mannose-type glycosylated protein with an apparent molecular mass of 75–80 kDa [82]. Kinetic studies indicate that the preferred catalytic metal is Zn^{2+} like many other CDAs. In terms of activity, its optimal pH and temperature values are between 4.5–5.5 and 50 °C, respectively [23,33]. *Mr*CDA deacetylates chitinous polymers such as glycol-chitin, colloidal chitin, chitosan, and chitin, but also deacetylates acetylxylan [17,23]. On chitooligosaccharides, triacetylchitotriose is the smallest substrate and the activity increases with the degree of polymerization (DP) [23,31,42,78]. The enzyme follows a multiple-attack mechanism [30] but the resulting pattern of acetylation (PA) depends on the DP of the substrate: whereas DP3, DP6 and DP7 substrates are not fully deacetylated, leaving the reducing GlcNAc unmodified [$D_{n\_1}A$], DP4 and DP5 substrates are fully deacetylated [D_n]. In all cases, deacetylation starts at the non-reducing end residue and then proceeds to the neighboring monomer towards the reducing end [31].

Colletotrichum lindemuthianum (*Cl*CDA). The deuteromycete *C. lindemuthianum* is a plant pathogen that causes anthracnose, a disease which affects economically important crop species [83]. *Cl*CDA is a heavily glycosylated secreted enzyme allegedly playing a role in the host-pathogen interaction, deacetylating the chitin oligomers resulting from the activity of plant chitinases on the fungal cell wall [83,84]. Less likely is its function in deacetylating the fungal cell wall chitin to evade degradation by plant chitinases, since no chitosan has been observed in the cell wall ultra-structure [85]. Since its discovery in the 1980s it has been purified from its natural host [84,86] as well as expressed in several eukaryotic and bacterial hosts such as *Pichia pastoris* [87,88] and *E. coli* [89,90]. The enzyme has a preference for Co^{2+} and Zn^{2+} as the catalytic metal cation and its activity is substantially inhibited by Cu^{2+} or Ni^{2+}, but not inhibited by EDTA or acetate [53,86]. It is a quite thermostable enzyme with an optimum temperature of 60 °C, and a pH optimum of 8.0. *Cl*CDA is active on both chitin polymers (glycol-chitin) and COS. It fully deacetylates COS with a DP equal to or greater than 3, while it only deacetylates the non-reducing GlcNAc of diacetylchitobiose [32,91]. *Cl*CDA acts by a multiple-chain mechanism following a pathway in which the first residue to be deacetylated is the second from the reducing end [32,33]. The initial mono-deacetylation reaction shows no dependency of k_{cat} on DP and a decrease of K_M with increasing DP [33,53]. However, kinetics of fully deacetylated products formation show an increase in k_{cat} and reduction in K_M that correlate with the increase of DP [86]. It has been reported that this enzyme is reversible, as it is also able to catalyze the acetylation of chitosan oligomers [92–94].

Aspergillus nidulans (*An*CDA). During cell autolysis, *An*CDA is secreted into the extracellular medium to deacetylate the chitin oligomers produced by chitinases [40,50,52,95]. The enzyme has been purified from A. nidulans cultures as a glycosylated enzyme [51]. The recombinant protein has been expressed in E. coli and purified by refolding from inclusion bodies [96] and recently it has been obtained in soluble form [25]. Like ClCDA, *An*CDA is a thermostable protein with an optimal temperature of 50 °C and retaining 68% activity after 1 h at 80 °C. Its optimum pH is 7–8 [51,96]. The enzyme is active on soluble chitins (CM-chitin, glycol-chitin), colloidal chitin, chitosan, acetylxylan, and acetylated glucuronoxylan, but not on peptidoglycan [25,51]. *An*CDA is active on COS with a DP from 2 to 6 [25]. The enzyme catalyzes mono-deacetylation of $(GlcNAc)_2$ and it is inactive on GlcNAc monosaccharide. Longer substrates than DP2 are fully deacetylated. However, the deacetylation rate exhibits a counter-intuitive relationship with the DP of the substrate: odd-numbered COS (DP5, DP3) have higher apparent rate constants than even-numbered oligomers (DP4, DP2). For the DP6 substrate, time-course monitoring of products formation reveals that the first deacetylation event occurs at random positions except for the reducing end, which reacts much slower to yield the fully deacetylated end product [D_n].

Podospora anserina (*Pa*CDA). The filamentous ascomycete *Podospora anserina* lives as a saprophyte on herbivore dung [97]. It has a limited lifespan and it is a model organism in cell aging studies [98]. *Pa*CDA was identified in a search for CDAs containing chitin binding domains. The enzyme has been recombinantly expressed in *Hansenula polymorpha* as a full length protein composed of the CE4 domain flanked by two CBM18 domains [26]. The low activity of the enzyme on colloidal chitin is significantly reduced by deletion of the CBM domains, which supports the hypothesis that the presence of the CBMs helps the enzyme to act on insoluble substrates. *Pa*CDA is active on soluble glycol-chitin, chitosans with a high DA, and COS, with optimum pH and temperature values of 8.0 and 55 °C, respectively. It fully deacetylates COS with a DP $\geq$ 2 and follows a multiple-chain mechanism. With the DP3 substrate, the first deacetylation event has a clear preference for the reducing end, but all possible isomers are found for both mono- and di-deacetylated intermediate products. With DP4 and DP5 substrates, the residue next to the reducing end is preferentially deacetylated first, with the second deacetylation occurring mainly next to the existing $GlcNH_2$ unit on either side. Deacetylation is faster for longer substrates, with deacetylation of the reducing end occurring as a late event [26].

Puccinia graminis f. sp. *Tritici* (*Pgt*CDA). The biotrophic basidiomycete *Puccinia graminis f.* sp. *Tritici* is the causative agent of the stem rust [99]. The appearance of resistant races of *P. graminis* affecting wheat cultivars has been recognized as a serious threat to food security [100,101], boosting the interest in understanding the virulence and defense mechanism of this fungal pathogen. Rust fungi promote the formation of complex structures in order to invade the plant cells but at the same time they must prevent the triggering of immune responses [102]. A main transition during infection is from the ectophytically growing appressorium to the endophytically growing substomatal vesicle; while the former exposes chitin on its surface, the latter exposes chitosan [47,103]. *Pgt*CDA may not only participate in the chitin to chitosan transition, making the cell wall less susceptible to host chitinases [104], but also could deacetylate the chitooligosaccharide products, reducing its elicitor properties [105]. *Pgt*CDA has been recombinantly expressed in *E. coli* as a fusion protein with the maltose binding protein (MBP) [54]. Its optimal pH for activity is between 8 and 9 and its optimal temperature is 50 °C. It is not active on insoluble polymers such as α- or β-chitin, but efficiently deacetylates colloidal chitin, glycol-chitin and chitosans, on which activity increases with the degree of acetylation. With COS substrates, the minimal substrate is tetraacetylchitotetraose (DP4). The structure of the products from enzymatic deacetylation of DP4 to DP6 substrates reveals that the enzyme acts by a multiple-chain mechanism and specifically deacetylates all but the last two GlcNAc units on the non-reducing end [$AA(D)_{n\_2}$] [54].

Pestalotiopsis sp. (*Pes*CDA). The endophytic fungus *Pestalotiopsis* sp. lives inside the tissues of its plant hosts in tropical areas [106]. To successfully survive in their hosts, endophytes also need to avoid being detected by the plant immune system. A secreted *Pestalotiopsis* CDA has been identified when chitosan was present in the culture medium [49]. The recombinantly expressed *Pes*CDA is active on colloidal chitin as substrate, chitosans with a DA of 10–60% (higher activity with a higher DA), and COS,

but inactive on crystalline α- or β-chitin. When analyzing the activity on COS, tetraacetylchitotetraose is the minimal substrate. With a DP5 substrate, the optimum pH and temperature values are 8.0 and 55 °C, respectively. Through a multiple-chain mechanism, the enzyme deacetylates all residues of the substrates except the reducing end and the last two GlcNAc residues from the non-reducing end, with a pattern of deacetylation [AA(D)$_{n_3}$A] [49]. The chitosan oligomers obtained from deacetylation of a DP6 substrate by *Pes*CDA have shown that, as opposed to the fully acetylated oligomer, they are no longer elicitors of the plant immune system in rice cells [49].

Pochonia chlamydosporia (*Pc*CDA). The ascomycete *Pochonia chlamydosporia* infects females and eggs of cyst or root-knot nematodes. It is used as a biocontrol agent against a number of plant parasitic nematodes in food-security crops [107–109]. *P. chlamydosporia* expresses chitosanases and chitin deacetylases during egg infection. Since chitosan is associated with the sites of fungal penetration, it has been suggested that secreted CDAs are involved in nematode infection [110]. A *Pc*CDA has been recently characterized [55]. The full-length protein contains the CE4 catalytic domain flanked by two CBM18 chitin binding domains. The recombinantly expressed *Pc*CDA catalytic domain deacetylates COS with a DP $\geq$ 4, with preference for longer substrates. It starts deacetylating the penultimate residue from the non-reducing end and continues deacetylating the next residue towards the reducing end, with a pattern of acetylation [ADDA$_{n_3}$] [55].

The above described CDAs are well characterized in terms of their deacetylation mode of action on COS and the structure of their deacetylated products. A number of other fungal CDAs (Table 1, and below) have also been assayed on COS substrates but, to the best of our knowledge, the deacetylation pattern of the products has not been reported.

Saccharomyces cerevisiae (*Sc*CDA1 and 2). The *S. cerevisiae* ascospore walls are well ordered structures with two outer layers that confer spore resistance, one made of 95% chitosan and the outermost proteinaceous layer rich in dityrosine [44]. Two CDAs are expressed exclusively during sporulation and are required for spore wall rigidity [38]. Both CDAs have been cloned and expressed in yeast as glycosylated proteins active on glycol-chitin [38,56], and in *E. coli* [57] as soluble proteins with deacetylase activity on glycol-chitin, chitosan (DA 50%) and COS. More detailed characterization of *Sc*CDA2 expressed in *E. coli* revealed that at least two GlcNAc residues are required for activity on COS, with maximum activity on DP6 [57]. When glycol chitin is used as substrate the optimum temperature for enzyme activity is 50 °C and the pH optimum is 8.0. It has also been shown that the *Sc*CDAs may act on nascent chitin chains in an in vitro assay system with chitin synthase [56].

Mortierella sp. (*Mo*CDA). Some *Mortierella* species live as saprotrophs in soil and other organic materials such as decaying plant leafs, fecal pellets or on the exoskeleton of arthropods, whereas other species are endophytes [111]. An extracellular CDA was identified [58] and purified from a *Mortierella* sp. as a highly glycosylated protein with maximum activity at pH 5.5–6 and 60 °C [112]. *Mo*CDA is active on soluble substrates as chitosans and glycol-chitin but with no detectable activity on β-chitin, colloidal chitin, and CM-chitin. It is active on COS with a DP $\geq$ 2, with higher activity with increasing DP of the substrate. With diacetylchitobiose, only monodeacetylation was observed. The structure of the deacetylated products from larger oligomers has not been reported.

Absidia sp. (*Acoe*CDA, *Acory*CDA). *Absidia* strains of *Zygomycetes* produce chitosan in their cell wall through the tandem action of chitin synthases and deacetylases. In *A. coerulea*, chitosan accounts for 10% of the vegetative cells and the DA reaches 95%. *AcoeCDA* was purified and proven to be active on glycol-chitin with a pH optimum of 5 at 50 °C. When the purified enzyme was incubated with a chitin synthase, it converted 90% of the nascent chitin from UDP-GlcNAc into chitosan. It deacetylates COS with more than two GlcNAc units, with increasing activity with longer substrates [43]. Similarly, *Absidia corymbifera* secretes a CDA active on glycol-chitin and chitosans with optimum pH and temperature of 6.5 and 55 °C, respectively, and active on COS with DP $\geq$ 2 [59].

Flammulina velutipes (*Fv*CDA). The basidiomycete *Flammulina velutipes* (called Enokitake in Japan) is commercially cultivated and fruited to produce foods with high nutritional value. A CDA that is expressed at the early stages of fruity body development was recombinantly expressed in *Pichia pastoris* [60]. *Fv*CDA,

active on glycol-chitin and colloidal chitin, deacetylates COS from dimer to pentamer, with activity increasing with the DP of the substrate. The enzyme exhibits the maximum activity at 60 °C and pH 7.

Penicilium oxilicum (*Po*CDA). An extracellular CDA from *Penicilium oxilicum*, purified from culture supernatants, exhibits deacetylase activity on glycol-chitin at pH 9, a common value for extracellularly secreted CDAs as opposed to intracellular CDAs with typical pH optima in the range of 5 to 7. *Po*CDA is active on COS with activity increasing from DP2 to DP5 [61].

Aspergillus flavus (*Af*CDA). In the search for extracellularly secreted CDAs for industrial applications, optimization of solid substrate fermentation and submerged fermentation of *Aspergillus flavus* has been reported [62,113]. The *Af*CDA enzyme purified from the extracellular medium has optimal activity on glycol-chitin and colloidal chitin at pH 8 and 50 °C. When assayed with COS as substrates, *Af*CDA is active on DP4 but has no activity on shorter substrates [62].

Scopulariopsis brevicaulis (*Sb*CDA). *Scopulariopsis* spp. are common soil saprophytes. Few species have been associated with human diseases, including *S. brevicaulis*. They are dermatomycotic molds and mainly have been associated with onychomycosis [114,115]. *Sb*CDA is an extracellular enzyme that is active on chitin and chitosans. The purified native enzyme is also active on COS with at least two GlcNAc units, and the activity increases with the DP of the substrate. With DP6, optimum conditions for deacetylation are pH 7.5 and 55 °C [63].

Rhizopus sp. (*Rc*CDA, *Rs*CDA). *Rhizopus* species have been screened as CDA producers. A *R. circicans* CDA has been cloned and recombinantly expressed in *Pichia pastoris* [64]. *Rc*CDA has maximum activity on glycol-chitin at pH 5–6 and 37 °C. On COS, only activity on a DP6 substrate has been reported. A CDA from *Rhizopus stolonifer* (or *nigricans*) has also been isolated as an active enzyme on glycol-chitin but no activity on COS has been reported [64,116]. Fermentation conditions of other *Rhizopus* species as CDA producers are being studied for the bioconversion of chitin to chitosan [117].

Other chitosan producers have been identified and studied as a source of chitosans, with many reports on screening and fermentation optimization, but the corresponding chitin deacetylases have not been characterized yet. Some examples include *Gongronella butleri* [65,118], *Phycomyces blakesleeanus* [66,119], and *Schizophyllum commune* [67].

Cryptococcus neoformans (*Cn*CDA). *Cryptococcus neoformans* is a dimorphic basidiomycetous human fungal pathogen that causes cryptococcal meningoencephalitis, particularly in immunocompromised patients [37]. *C. neoformans* has substantial chitosan in its cell wall during vegetative growth that is necessary for virulence and persistence in the mammalian host [120,121]. Three CDAs are predicted to be GPI-anchored to the cell wall, suggesting that they transverse the plasma membrane or attach to the cell wall to deacetylate the chitin generated by a chitin synthase as it is extruded through the plasma membrane [37]. The GPI-anchor of *Cn*CDA2 has proven to be required for membrane association but dispensable for cell wall association [122]. Activity of *C. neoformans* CDAs on COS substrates has not been reported. Interestingly, screening studies to identify cryptococcal antigens that stimulate an immune response on murine T cell hybridomas reactive with cryptococcal proteins, have shown that two of the CDAs are immunogenic [123,124].

3.3. Protozoan CDAs

Entamoeba histolytica (*Eh*CDA). *Entamoeba histolytica* is an anaerobic parasitic amoebozoan that predominantly infects humans and other primates causing amoebiasis [125]. The genome contains two putative CDAs, one of which has been cloned and recombinantly expressed in *Saccharomyces cerevisiae* [68]. *Eh*CDA deacetylates COS, being active on DP5 and DP6, but with no detected activity on DP4 [68].

3.4. Bacterial CDAs

The predominant CE4 deacetylases in bacteria are chitin oligosaccharide deacetylases (CODs), active on low molecular mass COS and essentially inactive on polymeric chitin and chitosans. These include rhizobial NodB deacetylases and CODs from marine bacteria. But bacterial CDAs

other than CODs are being discovered from screening programs and data mining of sequenced genomes and metagenomes, as in the recent case of an *Arthobacter* CDA.

Sinorhizobium meliloti (NodB). Rhizobial NodB is part of the Nod operon involved in the biosynthesis of Nod factors, the morphogenic signal molecules produced by rhizobia, which initiate the development of root nodules in leguminous plants [126]. NodB is active on chitooligosaccharides from DP2 to DP5 with no differences in k_{cat}, but K_M decreases with increasing DP [18,127–129]. Specifically, k_{cat}/K_M is 5-fold higher for DP5 than for DP2 substrates. DP4 or DP5 substrates are the natural substrates depending on the Rhizobial strain. *Sm*NodB optimum activity between pH 7 and 8 at 30 °C [18]. NodB is highly specific deacetylating only the non-reducing end residue [DA_{n_1}] although traces of a second deacetylation event have been observed upon long incubations [18,130,131].

Vibrio species (*Vc*CDA, *Vp*CDA, *Va*CDA). Chitin oligosaccharide deacetylases (COD) from the *Vibrionaceae* family are involved in the chitin degradation cascades occurring in sea water [132–135]. They have been identified in many *Vibrio* species, such as *V. algynolyticus* [73,136], *V. parahaemolyticus* [72,137], *V. cholera* [70], *V. harveyi* [138] and others. The *V. parahaemolyticus* and *Vibrio* sp. SN184 CDAs only deacetylate DP2 and DP3 substrates, whereas the *Vibrio cholera* chitin deacetylase (*Vc*CDA) has a broader specificity, accepting substrate from DP2 to DP6 [69,70]. *VcCDA* has a 10-fold higher activity on DP2 than on DP4 [69], and specifically deacetylates the penultimate residue from the non-reducing end, generating monodeacetylated products with the pattern [ADA_{n_2}] [69,70,130].

Shewanella species (*Sw*COD, *Sb*COD). In addition to the *Vibrio* genus, CODs have been recently identified and characterized from the *Shewanella* genus, marine bacteria found in extreme aquatic habitats (low temperature and high pressure). *Shewanella* sp. CODs share high sequence identity (50–60%) with *Vibrio* CODs, and have essentially the same biochemical properties. The *S. woodyi* enzyme (*Sw*COD) contains two CBM12 chitin binding domains at the C-terminus, deacetylates the reducing end on diacetylchitobiose [AD], and the activity drastically decreases from DP2 to DP4 substrates, with no activity detected on a DP5 substrate [74]. The *S. baltica* enzyme (*Sb*COD) contains a single CBM12 at the C-terminus, it is active of diacetylchitobiose with the same deacetylation pattern [AD] but it is less active on a DP3 than on a DP4 substrate [75].

Arthrobacter sp. (*Ar*CE4). A bioinformatics search for monodomain and extracellular CDAs in annotated genomes and metagenomes identified *Ar*CE4 as a CDA from an *Arthrobacter* species [76], a Gram-positive bacteria known to grow on chitin and secrete chitinases [139–141]. *Ar*CE4 is active on chitosan (DA 64%), acetylxylan, and insoluble chitin. It also deacetylates COS substrates with DP $\geq$ 2. The activity increases with increasing DP, with higher activity against DP5 compared to DP6. As shown with the DP5 substrate, the enzyme follows a multiple-chain mechanism where different mono- and di-deacetylated products are obtained. Whereas the first deacetylation occurs at all three internal positions, di-deacetylation mainly takes place at the GlcNAc unit next to the reducing end and at either of the two other internal units (ADDAA and ADADA). The final products have a pattern of acetylation [$D_{n\_1}A$], where the reducing end unit is not deacetylated [76].

4. Structural Determinants of Activity and Specificity

Structural analysis of CE4 enzymes with solved 3D structure have been recently reviewed [142], comparing and highlighting the differences between the different subfamilies based on substrate preferences. Here we focus and summarize the current knowledge on the structure and specificity of CDAs as a subfamily of CE4 enzymes. The closer similarity and activity on the same substrates provides a framework to analyze the structural determinants responsible for the different modes of action that lead to different patterns of deacetylation in their products.

4.1. 3D Structures

Some CDAs are mono-domain proteins and some others have a multi-domain architecture composed of the CE4 catalytic domain (or NodB homology domain), and several other domains, such as

carbohydrate binding modules (CBMs [143]) and domains with unknown function. The function of the CBMs is not clear and might be diverse depending on the biological role of each enzyme in its organism. In extracellular CDAs acting on the cell wall chitin, they may facilitate solubilization and access to the substrate (i.e., *Pa*CDA with two CBMs, where deletion of one or both confirmed their proposed function in supporting the enzymatic conversion of insoluble chitin [26]). In CDAs acting on low molecular weight COS, the CBMs may be involved in enzyme localization. This is the case of COD enzymes from marine bacteria (*Vibrio* and *Shewanella* species), where the small substrate does not span out of the active site, and the CBMs might bind to chitinous material in order to keep the COD activity close to the site where COS are generated by the action of chitinases.

The first CE4 enzymes with 3D structure determined by X-ray crystallography were the peptidoglycan deacetylases *Bs*PdaA [144] and *Sp*PgdA [145], and the first CDA was that from *Colletotrichum lindemuthianum* (*Cl*CDA) [53]. Currently, only five CDAs in the CE4 family have known 3D structure (Figure 2). The CE4 catalytic domain is characterized by a distorted $(\beta/\alpha)_8$ barrel fold. The distorted barrel, which often lacks one of the αβ repeats of regular TIM barrels, creates a groove into which the extended polymer substrate binds [144,146,147]. Seven or eight parallel β-strands form the β-barrel surrounded by α-helices. In addition, a series of loops decorate the β-barrel and make up the majority of the carbohydrate binding pocket as discussed below.

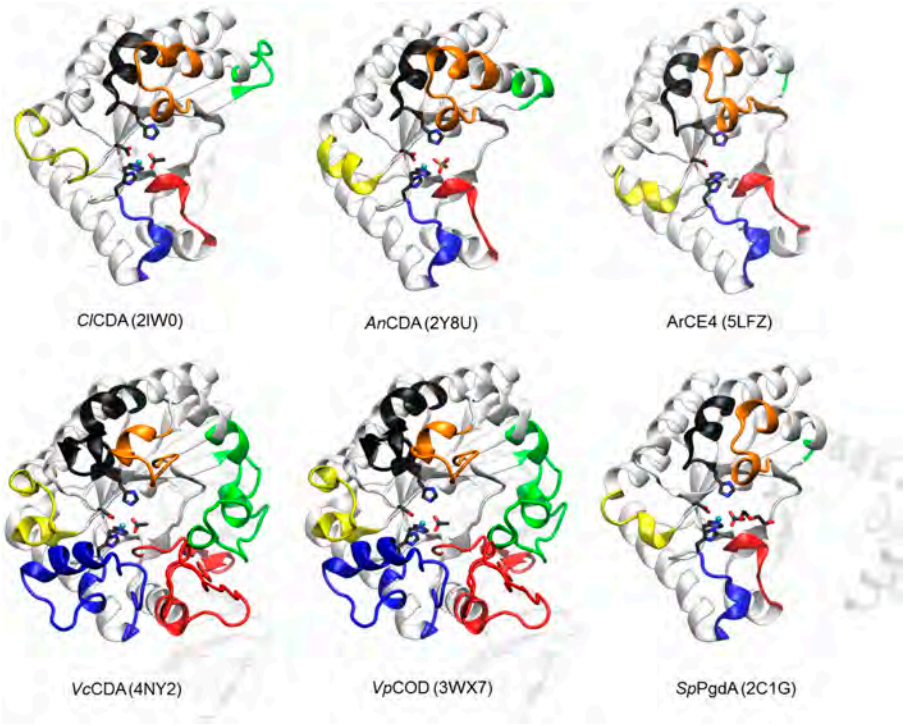


Figure 2. 3D structures of CDAs determined by X-ray crystallography. Loops 1 to 6 colored as in Figure 3. The peptidoglycan deacetylase *Sp*PgdA is also included for comparison (see text). In parenthesis, PDB accession codes.

4.2. The NodB Homology Domain and Conserved Active Site Motifs

The multiple sequence alignment of the CE4 domain for the CDAs listed in Table 1 was guided by the structural superimposition of the available X-ray structures (Figure 2) and is presented in Figure 3. Compared to most of the CDA members, the *Vc*CDA enzyme has substantially longer insertions, and it was key to defining the loops that differentiate CDAs and shape the binding site cleft of these enzymes. Sequences of enzymes without structural date were incorporated into the alignment by means of Hidden Markov Model comparisons. As seen in Figure 3, the conserved motifs and non-conserved insertions are evenly distributed along the sequences of CDAs.

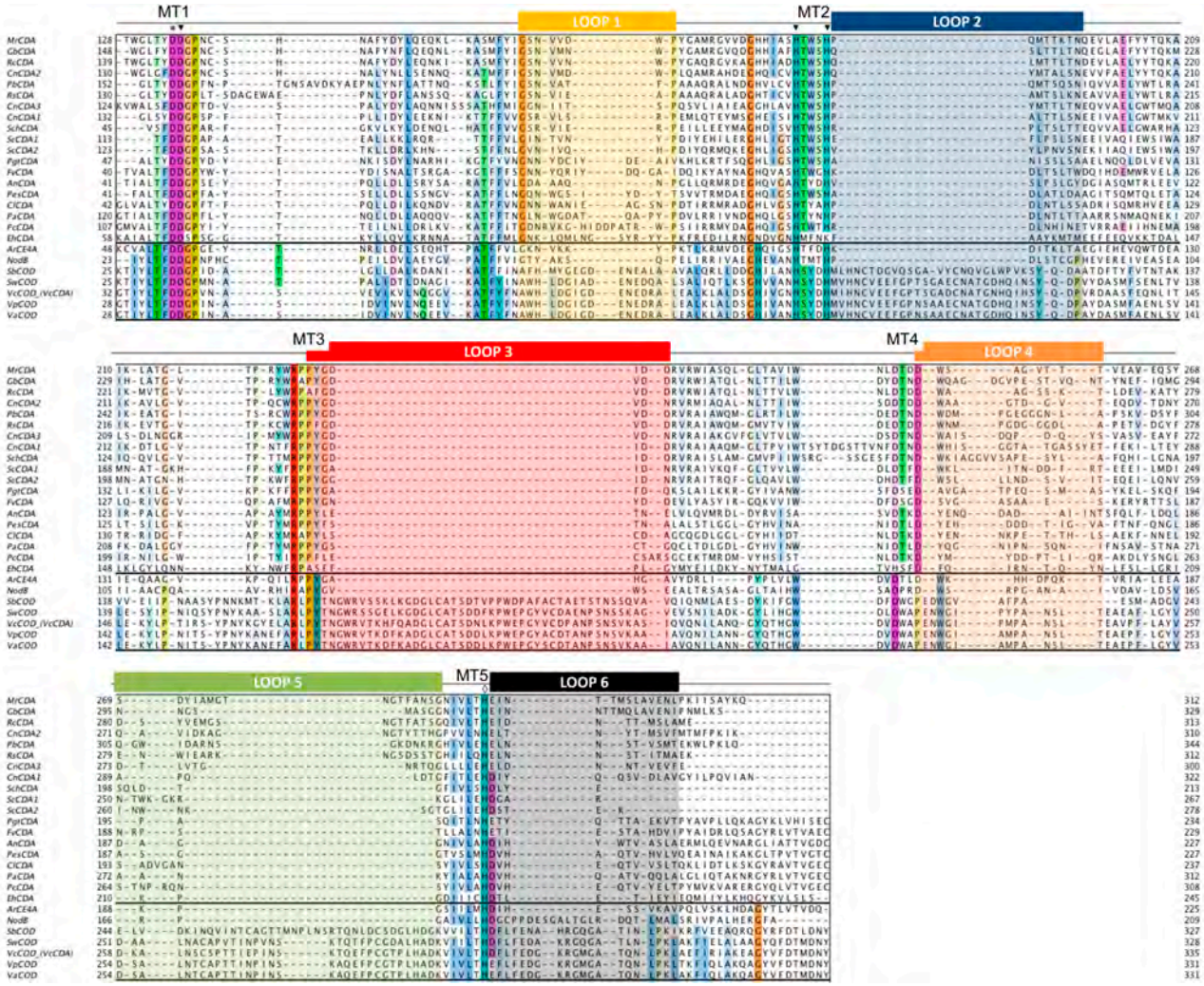


Figure 3. Multiple sequence alignment of the CDA enzymes listed in Table 1. Loops are highlighted with colored boxes according to [69]. Conserved catalytic motifs are labelled MT1-5. The 'His-His-Asp' metal binding triad (▼), catalytic base (*), and catalytic acid (◊) are labelled.

The conserved motifs are related to enzymatic activity (Motifs 1 to 5) and are typically located at the center of the active site structure. The non-conserved insertions correspond to both un-structured and structured loops of variable length, sequence, and geometry that surround the active site. These loops are numbered from Loop 1 to Loop 6 in the sequence alignment (Figure 3). As discussed in Section 4.5, they are key elements in determining the substrate specificity of different CDAs.

As members of the CE4 family, CDAs share the ≈150 aa-long NodB homology domain (CE4 domain). This region is defined by five conserved motifs that, according to the order they appear in the sequence, are named Motif 1 to Motif 5. These consensus motifs were first proposed in 2005 by sequence alignment of representative enzyme members of the CE4 family when the 3D structure of the peptidoglycan deacetylase *Sp*PgdA was solved [144]. Motif 1 (TFDD) is highly conserved in CDAs and contains the general base aspartate (first D) and the metal-binding aspartate (second D). Motif 2 (H(S/T)xxH) is a zinc-binding motif, where the two His residues bind the metal cation and the Ser or Thr residue forms a hydrogen bond with the second His, stabilizing the local conformation of the loop-shaped motif. These two His from Motif 2 plus the metal-binding Asp from Motif 1 are often designated the His-His-Asp metal-binding triad of CE4 enzymes. Motif 3 (RxPY) forms one of the sides of the active site groove and establishes stabilizing interactions with other active site residues. Motif 4 (DxxD(W/Y)) forms the other side of the active site groove, including a hydrophobic residue exposed to the solvent and a buried Asp. Motif 5 (I(V/I)LxHD) contains the catalytic general acid His residues and a Leu, which is part of a hydrophobic pocket that accommodates the acetate methyl group of the substrate.

4.3. Phylogeny of CE4 Chitin Deacetylases

Based on the above multiple sequence alignment, a clustering of the CE4 domain sequences of characterized CDAs based on phylogenetic analysis is presented in Figure 4. This is a reduced phylogenic analysis limited to the CDAs with reported activity on COS as listed in Table 1. Fungal and bacterial CDAs are clearly segregated in two clades, with a protist CDA (*Eh*CDA) located between both groups.

Fungal enzymes from organisms belonging to different phyla (*Zygomycota*, *Basidiomycota*, and *Ascomycota*) are distributed throughout the fungal clade. Within the clade, CDAs appear grouped in two clusters related with their biological function. The first cluster contains orthologous CDAs of different phyla known to have a role in cell wall chitosan biosynthesis at different stages of the fungal cell cycle, such as *Mr*CDA during cell wall formation [42], *Cn*CDAs during vegetative growth [37], or *Sc*CDAs during sporulation [38,44]. Although there is no experimental proof of their biological function, the CDAs from *Gonglonella*, *Phycomyces*, as well as those from *Rhizopus* are likely to be also involved in cell wall formation due to their location in the same cluster of the phylogenetic tree and their taxonomic classification (mucorales inside the *Zygomycota* phylum). Regarding the cellular location of the enzymes in this cluster, most of them are secreted to the periplasm or are GPI-anchored to the cell wall, where they are coupled with chitin synthases for chitosan biosynthesis. The second cluster is mainly composed of extracellular CDAs that participate in host infection, either as a defense mechanism to prevent the elicitation of host defense mechanisms (*Pgt*CDA, *Pes*CDA), or involved in the interaction with the host or as a virulent factor (*Cl*CDA, *Pc*CDA). Extracellular CDAs involved in cell autolysis (*An*CDA) also fall in this group.

Most of the bacterial enzymes included in the alignment are chitin oligosaccharide deacetylases (COD) and are more distantly related to the fungal CDAs. These form a different clade in the phylogenetic tree (Figure 4). The enzymes from marine bacteria (*Vibrio* and *Shewanella* species) are clustered together with high sequence similarity and have similar biological functions and biochemical properties. NodB has a more distant relationship with the other CODs.

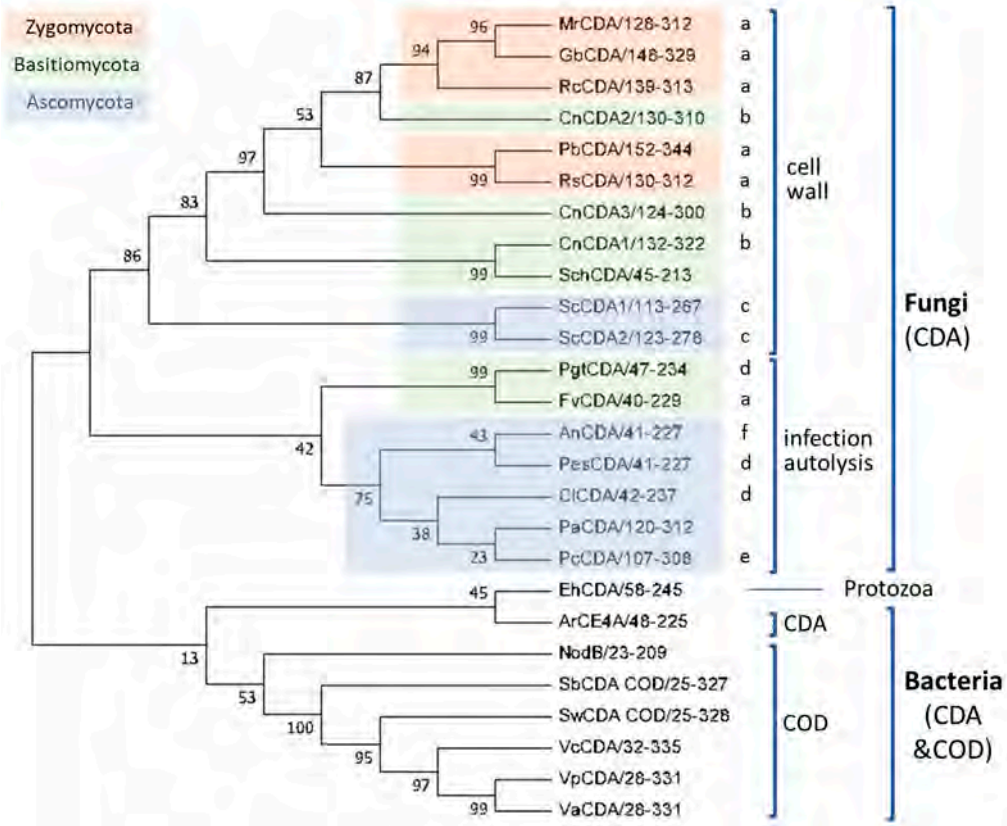


Figure 4. Phylogenetic analysis of CDAs from the multiple sequence alignment presented in Figure 3. A bootstrap analysis with 500 replicates was carried out on the trees inferred from the neighbor joining method. The consensus tree is shown with bootstrap values at each node of the tree. Biological functions: cell wall biosynthesis: (a) cell wall, (b) vegetative growth, (c) sporulation; host infection, (d) defense, (e) interaction/infection, (f) autolysis (see text).

4.4. Catalytic Mechanism

CDA enzymes operate by metal-assisted acid/base catalysis. The general mechanism was first proposed for the peptidoglycan GlcNAC deacetylase *Sp*PgdA when solving its X-ray structure [144] and short after supported by the 3D structures of the acetylxylan esterases *Sl*AxeA and *Ct*AxeA [148]. The catalytic machinery involves the conserved active site motifs containing the metal-binding triad and the general acid and base residues. Only the structures of four different CDA have been solved up to date (*Colletotrichum*, *Aspergillus*, *Vibrio*, and *Arthrobacter* CDAs, Table 1), all consistent with the proposed metal-assisted mechanism. *Vc*CDA was the first CE4 enzyme for which the 3D structure of enzyme-substrate complexes were solved by X-ray crystallography [69]. The structure of complexes of an inactive mutant (at the general base Asp residue) with diacetylchitobiose (DP2) and triacetylchitotriose (DP3) in productive binding for catalysis showed that a sugar hydroxyl group of the substrate also participates in metal coordination. Specifically (Figure 5), the Zn^{2+} cation is coordinated by the imidazole nitrogens of His97 and His101, the carboxylate group of Asp40, and the O7 atom of the *N*-acetyl group and O3 hydroxyl of the GlcNAc ring. The distorted octahedral coordination is completed by a water molecule. Upon activation, this water molecule is proposed to be the nucleophile responsible for removal of the *N*-acetyl group. Just recently, a second structure of an enzyme-substrate complex has been reported for the *Arthrobacter* sp. CDA (*Ar*CE4) [76]. The diacetylchitobiose ligand bound into the active site also shows the same type of interactions with the conserved active site residues.

The proposed mechanism of CDAs and related CE4 enzymes is shown in Figure 5. In the first step, metal coordination polarizes the carbonyl amide of the substrate which reacts with the nucleophilic water molecule activated by the general base (Asp), leading to a tetrahedral oxyanion

intermediate. Next, protonation of the nitrogen group of the intermediate by the general acid (His) facilitates C-N bond breaking with release of acetate and the generation of a free amine in product. Kinetic evidence for an oxyanion tetrahedral intermediate and significant charge development at the first transition state was provided by Hammett linear free energy correlations using the *Cl*CDA enzyme with α-haloacetamido substrate analogues [53]. In most of the enzymes, the catalytic acid and base residues are part of two conserved "charge relay" side chain pairs that may contribute to modulate the pK_a of the catalytic residues [53,144]: the catalytic base (Asp) is tethered by a conserved Arg from MT3 (RxxPY) and the catalytic acid (His) is tethered by a conserved Asp from MT4, DxxD(W/Y).

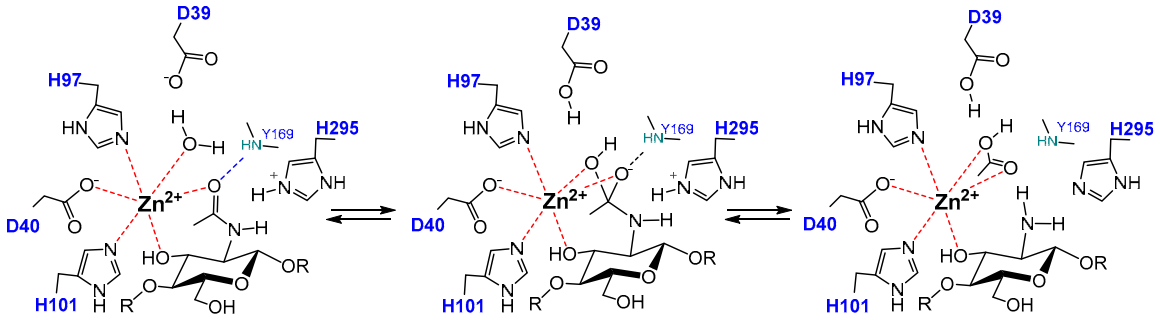


Figure 5. Metal-assisted general acid/base mechanism proposed for CE4 deacetylases. Scheme based on the 3D structure of the enzyme·substrate complex VcCDA$_{D39S}$·DP2 [61]. D39 is the general base and His295 is the general acid.

4.5. Determinants of Substrate Specificity

The series of crystal structures of the *Vibrio cholerae* chitin oligosaccharide deacetylase (*Vc*CDA or *Vc*COD) reported in 2014 [69] were the first 3D structures of a CE4 enzyme in complex with substrates. These data provided a first insight into structure-function relationships for this family of enzymes and highlighted the role of the loops that shape the binding site cleft in substrate binding. Recently by the end of 2017, the 3D structure of an *Arthrobacter* sp. CDA (ArCE4) in complex with substrate has also been reported [76]. Albeit sharing the same molecular function, both enzymes represent two different scenarios regarding the binding site topology and, hence, substrate specificity. *Vc*CDA has a rather closed binding cleft and is highly specific for monodeacetylation of COS, whereas *Ar*CE4 has a more open binding cleft and is able to fully deacetylate their COS substrates (Figure 6). A comparative structural analysis of both enzyme structures has been recently reviewed [142].

4.5.1. *Vc*CDA. Long Loops and High Specificity

Currently available structures of *Vc*CDA include the unliganded form of the enzyme and the binary complexes with *N*-acetylglucosamine (DP1), diacetylchitobiose (DP2), and triacetylchitotriose (DP3). These structures revealed two significant observations: a series of non-conserved loops (labeled Loop 1 to 6 in Figures 3 and 6) that shape the binding cleft, and the dynamics of the loops that assemble the active site for catalysis [69].

In all structures, the substrate is confined in a small binding cleft that is shaped by a series of long loops surrounding the active site (see Figure 6A for the *Vc*CDA·DP3 complex). Given the topology of *Vc*CDA protein surface, the binding of longer COS is prevented because these loops cap both the reducing and non-reducing ends of the substrate. Indeed, the catalytic efficiency of *Vc*CDA drops substantially on oligomers longer than DP2 [69] and attempts to solve the structure of *Vc*CDA bound to substrates longer than DP3 in a catalytically competent mode have been unsuccessful. Another consequence of this constricting topology is the high specificity of *Vc*CDA to exclusively deacetylate the penultimate residue from the non-reducing end of the substrates. There is no room for the ligand to slide along the binding cleft, thus it can only accommodate one GlcNAc unit of the

oligomeric chain at the catalytic site where deacetylation takes place. The binding of substrates induces a conformational change of Loop 4 from an open conformation in the unliganded enzyme to a closed conformation in the enzyme·DP2 complex or a semi-closed conformation in the enzyme·DP3 complex. It is triggered by a staking interaction between a Trp residue located in apical site of Loop 4 and the GlcNAc unit at the catalytic center, locking the substrate in the active site in the proper orientation for catalysis.

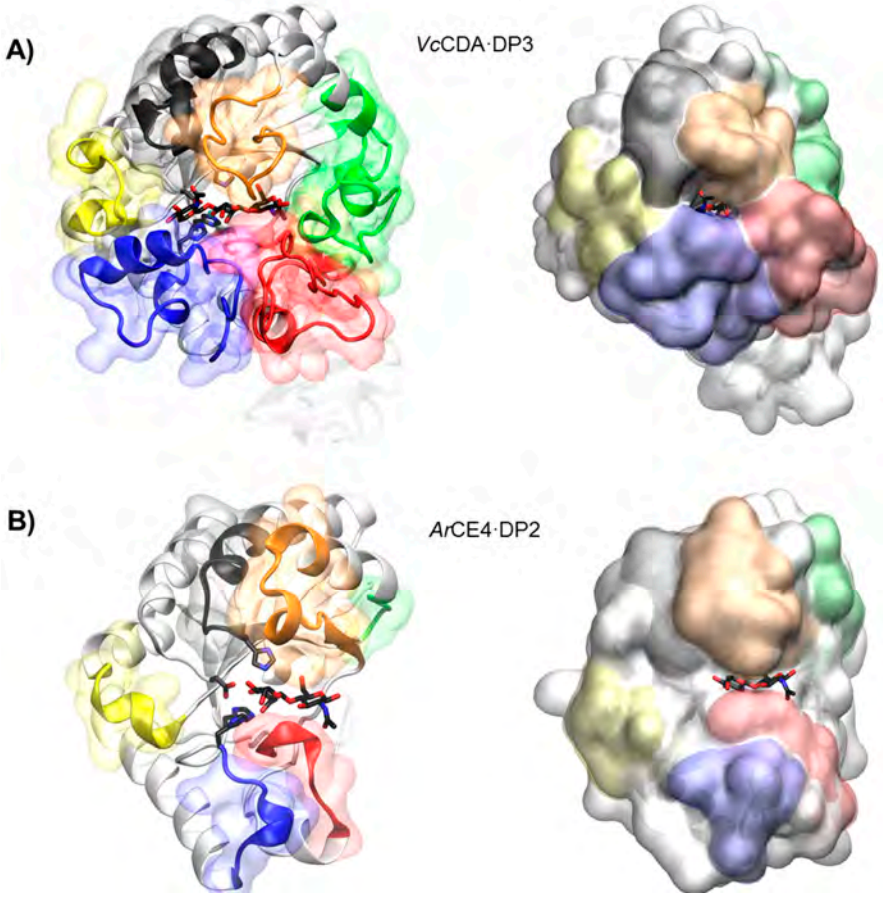


Figure 6. 3D structures of enzyme·substrate complex. (**A**) *Vc*CDA with DP3 substrate and (**B**) *Ar*CE with DP2 substrate. Loops 1 to 6 are colored as in Figure 3.

4.5.2. *Ar*CE4. Short Loops and Broad Specificity

In contrast to *Vc*CDA, the crystal structure of *Ar*CE4 in complex with diacetylchitobiose [76] reveals a flatter protein surface with the substrate bound to a more open binding cleft (Figure 6B). Even though the enzyme was co-crystalized with tetraacetylchitotetraose (DP4), only two GlcNAc units are observed in the structure. This indicates a weak binding of part of the COS substrate on this flat topology of the protein surface. The catalytic center in both enzymes is in the same position with respect to the protein core, being the main difference, the size and shape of the loops surrounding the active site. Since the binding cleft is more open, the enzyme can accommodate longer COS. Indeed, the enzymatic activity of *Ar*CE4 increases as the length of the chitin oligomer chain increases. The lack of protein caps at either the reducing and non-reducing ends of the substrate can also explain the multiple-chain mechanism proposed for this enzyme. Deacetylation takes place at all GlcNAc units of the substrate (except the reducing end) because it can freely bind to *Ar*CE4 in different binding modes exposing different GlcNAc units of the oligomeric chain at the catalytic site.

4.5.3. The Subsite Capping Model

The diversity of deacetylation patterns exhibited by chitin deacetylases and related CE4 enzymes can be attributed to the differential accessibility of the linear chitin oligosaccharide chain to the separate subsites along the substrate binding cleft of their structures. Considering all CE4 enzymes with reported activity on polymeric chitin or COS, these can be classified into two groups. One group is represented by general chitin deacetylases (CDA), and a second group is formed by chitin oligosaccharide deacetylases (COD). The two structures of the enzymes-substrate complexes described above are reference models for the protein surface topologies and substrate binding mechanisms of these two groups: CDAs (*Ar*CE4) and CODs (*Vc*CDA). These two structures provide a unified view of the determinants of substrate specificity in chitin deacetylases in terms of the "subsite capping model" proposed in [69]. According to this model, substrate accessibility is affected by the length, shape, and dynamics of a series of loops surrounding the active site of CE4 enzymes. These loops are numbered from 1 to 6 and their location in the sequences and structures of CDAs and CODs is highlighted in Figures 2 and 3.

The group of CDAs bears short loops, and their structures exhibit a flat and open binding cleft. The substrate binding mechanism in this group of enzymes may be similar to that described for the reference structure of *Ar*CE4. According to the model, the substrate may be able to slide along the binding cleft or to bind in different modes resulting in processive or multiple-chain attack mechanisms of deacetylation. This can already be anticipated for CDA enzymes of known structure (Figure 2) because the flat protein surface is already evident, but also for CDA enzymes of unknown structure given the similar sequence lengths of the loops evidenced in the alignment (Figure 3). This could be an explanation of why CDA enzymes in general are not specific for the deacetylation at a single *N*-acetylglucosamine unit. However, the patterns of deacetylation differ among the different CDAs. The surface charge distribution along the binding cleft and other structural features yet to be disclosed may also participate in defining the mode of action and deacetylation pattern by each particular enzyme.

On the contrary, the group of COD enzymes bears longer loops and their structures have narrower binding pockets and buried active sites. According to the subsite capping model, the substrate is constrained to bind in very specific binding modes resulting in single-site deacetylations. This is the case for the reference structure of *Vc*CDA in complex with substrates, but it can also be anticipated for other COD enzymes for which the 3D structure is still unknown. For instance, Loop 6 in *Rm*NodB is longer than in other CDAs. This loop is located on the non-reducing end site of the binding cleft and may cap the accessibility of the substrate after subsite 0 (the catalytic site) thus defining the deacetylation specificity for the non-reducing end of the substrate. Likewise, the *Shewanella* CODs have a Loop 6 with the same length than the *Vibrio* CODs, but shorter than NodB, and both exhibit the same mono-deacetylation specificity for the penultimate GlcNAc residue from the reducing end of the substrate.

For most CDAs, the reducing end of the substrate is not deacetylated, or it is the least reactive GlcNAc unit. As seen in the *Ar*CE4·DP2 complex 3D structure [64], binding to the +1 subsite seems to be dominated by the stacking interaction of the GlcNAc unit of the substrate with a Trp in Motif 4 at the beginning of Loop 4. This aromatic residue is highly conserved (MT4, DxxD(W/Y), Figure 3). CDA enzymes having this aromatic residue prefer a sugar bound in the +1 subsite; they do not deacetylate the reducing end of their substrates, as it is the case for *Ar*CE4, *Pes*CDA [89] and *Pc*CDA [95], or the reducing end is the slowest position to be deacetylated, as shown for *Cl*CDA [49] and *An*CDA [12]. On the contrary, *Pgt*CDA, which deacetylates the reducing end GlcNAc unit of all substrates from DP4 to DP6, lacks the +1 aromatic residue [89]. Different is the case of *Vibrio* and *Sewanella* CODs that have the equivalent aromatic residue in a slightly different position after a two-amino acid insertion in the MT4 motif, and it is located farther in Loop 4 (Figure 3). In the *Vc*CDA enzyme this loop moves from an open to a closed conformation upon substrate binding, and the same is expected for the other closely related CODs that have the same Loop 4 size. As a consequence of the induced fit, the Trp

residue now establishes a stacking interaction with the GlcNAc unit in subsite 0. DP2 is the preferred substrate for this group of COD enzymes, and it is deacetylated at the reducing end [62].

Gaining further structural information of protein-ligand complexes of CDA and COD enzymes, and other CE4 in general, will contribute further to decipher the structural and sequential determinants of substrate specificity in this family of enzymes. This will pave the way to the rational design or discovery of novel CDA with controlled specificities on the deacetylation of oligomeric and polymeric chitin for the biotechnological production of chitosans and paCOS with defined patterns of acetylation.

5. Application of Chitin Deacetylases

5.1. Targets for Antifungals

Fungal infections have an enormous impact on human health. Fungi are generally opportunistic pathogens affecting immunocompromised individuals including those with AIDS, receiving immunosuppressive drugs or undergoing cancer treatments. The cell wall is a meaningful target for antifungal therapies. Current major classes of antifungal drugs target cell membrane ergosterol biosynthesis (azoles), ergosterol function by disrupting membrane integrity (polyenes), or 1,3-β-glucan synthase preventing the formation of the cell-wall structural polysaccharide 1,3-β-glucan (echinocandins) [149]. New targets to overcome the emerging drug resistance by pathogenic fungi are becoming critical to treat life-threatening fungal infections. Other promising targets are the so called cell wall proteins (CWP) which mediate important cellular processes, including adhesion, invasion, biofilm formation and flocculation [122]. In fungal chitosan producers, chitin deacetylases are a class of CWP and potential targets for drug design. *Cryptococcus neoformans*, one of the most deadly pathogens, requires chitosan for virulence. Lack of chitosan in the cell wall has detrimental consequences in fungal growth and results in the complete loss of sporulation [120,121]. Thus, CDAs represent a promising target for anticryptococcal therapeutics [37,120], but no CDA inhibitors have been reported yet.

In pathogenic plants, major strategies to prevent fungal pathogenesis are related to the inhibition of fungal chitinases, which are required for chitin remodeling in the cell wall [150–152]. Different types of chitinase inhibitors have been reported, including potent natural inhibitors such as allosamidin [153] and the cyclic pentapeptides argifin and argadin [154]. However, inhibition must be selective so as not to interfere with the plant chitinases involved in triggering the plant defense mechanisms. Another potential and promising strategy is the inhibition of extracellularly secreted fungal CDAs since they constitute a defense mechanism to evade the plant immune system, as discussed in Section 3.2. As in the case of human fungal pathogens, no inhibitors have been yet reported against CDAs from plant pathogenic fungi.

5.2. Biocatalysts for the Enzymatic Production of Chitosans and paCOS

Chitosans can be found in a large number of applications in such distant areas as agriculture, cosmetics, water treatment, medicine and the food industry [155–159]. In addition to chitosan polymers, their oligomers (paCOS) have also proven to have relevant potential applications in agriculture and pharmaceutical industries [160]. The physicochemical and biological properties of chitosans and paCOS have been shown to be strongly dependent on their degree of polymerization and their degree of acetylation [161,162]. Many of the identified CDAs arose from screening programs addressed to find efficient biocatalysts to overcome the current industrial chitosan production by highly concentrated alkali treatment of chitin. Some examples include CDAs from *Mortiriella* sp. [58], *Rhizopus* sp. [117], or *Gongronella* sp. [118].

Chemical methods for the production of COS and paCOS are based on chemical depolymerization of chitosan [163,164], total synthesis of chitosan oligomers [165,166], partial chemical deacetylation of fully acetylated COS, or chemical re-*N*-acetylation of glucosamine oligomers based in two-step procedures or one-pot synthesis [167–169]. The drawbacks of chemical strategies are the unwanted

side reactions and the randomness of the chemical reactions. Current efforts are addressed to develop enzymatic routes for COS and paCOS production with defined DP, DA, and PA.

Enzymatic approaches include depolymerization of chitin or chitosan polymers using hydrolytic enzymes, chitinases and chitosanases, enzymatic polymerization by transglycosidation using transglycosylating hydrolases, and enzymatic de-acetylation and re-acetylation of chitin oligomers using chitin deacetylases, strategies recently reviewed in [170]. Based on the current knowledge on the specificity of a number of fungal and bacterial CDAs, recent reports have combined enzymes with different specificities to have access to a large family of paCOS with defined structures. The first proof of concept was to show that two specific CODs, NodB and VcCDA, each accept the monodeacetylated product from the other, leading to specific di-deacetylation, and that both enzymes can work in a one-pot process [130]. Recently, the use of different recombinant CDAs from bacterial and fungal origin to produce all of fourteen possible partially acetylated chitosan tetramers combining different enzymatic deacetylations and enzymatic *N*-acetylations has been reported [171] (Figure 7).

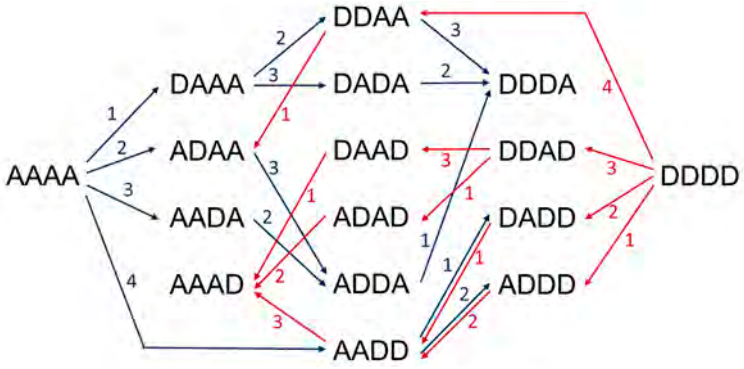


Figure 7. Production routes of all possible chitin and chitosan tetramers using 4 different CDAs to specifically deacetylate or *N*-acetylate paCOS. A: GlcNAc, D: $GlcNH_2$. Blue arrows, deacetylation reactions, red arrows, *N*-acetylation reactions in the presence of excess acetate.

The production of paCOS using in vivo strategies is an alternative to increase the scalability of the process. The first example towards a more general cell factory approach for the in vivo synthesis of paCOS was based on NodB deacetylase. By in vivo studies with *Escherichia coli* expressing different combinations of the nodABCS genes of *Azorhizobium caulinodans*, Nod factor intermediates were identified, as well as the sequence of the biosynthetic steps [172]. The nod gene cluster encodes a series of enzymes, which include the NodC chitin oligosaccharide synthase that produces fully acetylated chitin oligomers, the NodB chitin oligosaccharide deacetylase that deacetylates the non-reducing end unit, the NodA *N*-acyl transferase that transfer a fatty acid chain to the free amine group, and the NodS *N*-methyl transferase. Further transformations by other nod proteins elaborate the final Nod signaling factors [173]. In a first cell factory approach, high density cells of *E. coli* expressing nodC or nodBC genes produced in high yield (up to 2.5 g/L) penta-*N*-acetyl-chitopentaose and its deacetylated derivative tetra-*N*-acetyl-chitopentaose, which were easily purified by charcoal adsorption and ion-exchange chromatography [174]. The strategy was further extended to the production of sulfated and *O*-acetylated derivatives of these two compounds by coexpressing nodC or nodBC with nodH and/or nodL that encode chitooligosaccharide sulfotransferase and chitooligosaccharide *O*-acetyltransferase, respectively [175]. Other Nod analogues have also been generated with further modifications [176–178]. The cell factory approach, currently limited to one deacelylation pattern based on the use of NodB, is a promising technology to be developed by incorporating the diversity of CDAs with different deacetylation patterns in order to access a large family of paCOS and derivatives.

6. Conclusions

We have here summarized the current knowledge on substrate specificity of fungal and bacterial chitin deacetylases, their modes of action, and their use as biocatalysts for the production of chitosans and chitosan oligosaccharides with defined pattern of acetylation. By combining multiple sequence alignments and 3D structures of enzyme·substrate complexes of representative enzymes, a unified view of the determinants of substrate specificity is proposed in terms of the "subsite capping model." According to this model, substrate accessibility is affected by the length, shape, and dynamics of a series of loops surrounding the active site of CE4 enzymes. The group of CDAs active on polymeric substrates and COS bear short loops, and their structures exhibit a flat and open binding cleft. The substrate may be able to slide along the binding cleft or to bind in different modes, resulting in processive or multiple-chain attack mechanisms of deacetylation. Other structural features not yet disclosed, such as the charge distribution along the binding cleft may also participate in defining the mode of action and deacetylation pattern by each particular enzyme. The group of COD enzymes active on low molecular mass COS bear longer loops and their structures have narrower binding pockets and buried active sites. The substrate is constrained to bind in very specific binding modes resulting in single-site deacetylations.

But a deeper knowledge on substrate specificity requires further structural information of protein-ligand complexes of CDA and COD enzymes in order to decipher the structural and sequential determinants of substrate specificity in this family of enzymes aimed at the rational design or discovery of novel CDAs with controlled specificities on the deacetylation of oligomeric and polymeric chitin for biotechnological applications.

Although CDAs have been proposed as targets for antifungal drugs, no specific inhibitors have been yet reported. This is an open field that deserves attention not only for drug design but also to probe the signaling function of CDAs and CODs through their specific deacetylation of COS substrates.

Applications of CDAs and CODs as biocatalyst are currently being developed as a novel methodology to produce partially acetylated COS with tailored patterns of acetylation. Since not all patterns for COS of different sizes are yet available, enzyme discovery and protein engineering offer new opportunities for the biotechnological production of chitosans and paCOS with defined patterns of acetylation.

Acknowledgments: Work supported by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n°613931, and grant BFU2016-77427-C2-1-R from MINECO, Spain. Laia Grifoll and Sergi Pascual acknowledge a predoctoral contract from the NANO3BIO project. Hugo Aragunde acknowledges a predoctoral fellowship from Generalitat de Catalunya.

Author Contributions: Antoni Planas conceived and designed the review; Laia Grifoll, Sergi Pascual, Hugo Aragunde, Xevi Biarnés, and Antoni Planas compiled the references and prepared the figures. Xevi Biarnés and Antoni Planas wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

AXE	Acetylxylan esterase
A_n	$(GlcNAc)_n$
CDA	Chitin deacetylase
COS	Chitooligosaccharides
D_n	$(GlcNH_2)_n$
DA	Degree of acetylation
DP	Degree of polymerization
GlcNAc	*N*-acetylglucosamine
PA	Pattern of acetylation
paCOS	Partially acetylated chiton oligosaccharides
PDB	Protein data bank

References

1. Peniche Covas, C.A.; Argüelles-Monal, W.; Goycoolea, F.M. Chitin and chitosan: Major sources, properties and applications. In *Monomers, Polymers and Composites from Renewable Resources*; Belgacem, M.N., Gandini, A., Eds.; Elsevier: Amsterdam, the Netherlands, 2008; Volume 1, pp. 517–542. ISBN 9780080453163.
2. Karrer, P.; Hofmann, A. Über den enzymatischen Abbau von Chitin und Chitosan I. *Helv. Chim. Acta* **1929**, *12*, 616–637. [CrossRef]
3. Rinaudo, M. Chitin and chitosan: Properties and applications. *Prog. Polym. Sci.* **2006**, *31*, 603–632. [CrossRef]
4. Dutta, P.K.; Duta, J.; Tripathi, V.S. Chitin and Chitosan: Chemistry, properties and applications. *J. Sci. Ind. Res. (India)* **2004**, *63*, 20–31. [CrossRef]
5. Noishiki, Y.; Takami, H.; Nishiyama, Y.; Wada, M.; Okada, S.; Kuga, S. Alkali-induced conversion of β-chitin to α-chitin. *Biomacromolecules* **2003**, *4*, 896–899. [CrossRef] [PubMed]
6. Jang, M.K.; Kong, B.G.; Jeong, Y.I.; Lee, C.H.; Nah, J.W. Physicochemical characterization of α-chitin, β-chitin, and γ-chitin separated from natural resources. *J. Polym. Sci. Part A Polym. Chem.* **2004**, *42*, 3423–3432. [CrossRef]
7. Kumirska, J.; Czerwicka, M.; Kaczyński, Z.; Bychowska, A.; Brzozowski, K.; Thöming, J.; Stepnowski, P. Application of spectroscopic methods for structural analysis of chitin and chitosan. *Mar. Drugs* **2010**, *8*, 1567–1636. [CrossRef] [PubMed]
8. Kaya, M.; Mujtaba, M.; Ehrlich, H.; Salaberria, A.M.; Baran, T.; Amemiya, C.T.; Galli, R.; Akyuz, L.; Sargin, I.; Labidi, J. On chemistry of γ-chitin. *Carbohydr. Polym.* **2017**, *176*, 177–186. [CrossRef] [PubMed]
9. Dhillon, G.S.; Kaur, S.; Brar, S.K.; Verma, M. Green synthesis approach: Extraction of chitosan from fungus mycelia. *Crit. Rev. Biotechnol.* **2013**, *33*, 379–403. [CrossRef] [PubMed]
10. Hoell, I.A.; Vaaje-Kolstad, G.; Eijsink, V.G.H. Structure and function of enzymes acting on chitin and chitosan. *Biotechnol. Genet. Eng. Rev.* **2010**, *27*, 331–366. [CrossRef]
11. Xia, W.; Liu, P.; Zhang, J.; Chen, J. Biological activities of chitosan and chitooligosaccharides. *Food Hydrocoll.* **2011**, *25*, 170–179. [CrossRef]
12. Yu, R.; Liu, W.; Li, D.; Zhao, X.; Ding, G.; Zhang, M.; Ma, E.; Zhu, K.Y.; Li, S.; Moussian, B.; et al. Helicoidal organization of chitin in the cuticle of the migratory locust requires the function of the chitin deacetylase2 enzyme (LmCDA2). *J. Biol. Chem.* **2016**, *291*, 24352–24363. [CrossRef] [PubMed]
13. Winkler, A.J.; Dominguez-Nuñez, J.A.; Aranaz, I.; Poza-Carrión, C.; Ramonell, K.; Somerville, S.; Berrocal-Lobo, M. Short-chain chitin oligomers: Promoters of plant growth. *Mar. Drugs* **2017**, *15*, 40. [CrossRef] [PubMed]
14. Li, X.; Min, M.; Du, N.; Gu, Y.; Hode, T.; Naylor, M.; Chen, D.; Nordquist, R.E.; Chen, W.R. Chitin, chitosan, and glycated chitosan regulate immune responses: The novel adjuvants for cancer vaccine. *Clin. Dev. Immunol.* **2013**, *2013*. [CrossRef] [PubMed]
15. Varki, A.; Sharon, N. *Essentials of Glycobiology*, 2nd ed.; Varki, A., Cummings, R., Esko, J., Freeze, H., Stanley, P., Bertozzi, C.R., Hart, G., Etzler, M.E., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2009.
16. Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P.M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2014**, *42*, 490–495. [CrossRef] [PubMed]
17. Caufrier, F.; Martinou, A.; Dupont, C.; Bouriotis, V. Carbohydrate esterase family 4 enzymes: Substrate specificity. *Carbohydr. Res.* **2003**, *338*, 687–692. [CrossRef]
18. John, M.; Rohrig, H.; Schmidt, J.; Wieneke, U.; Schell, J. Rhizobium NodB protein involved in nodulation signal synthesis is a chitooligosaccharide deacetylase. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 625–629. [CrossRef] [PubMed]
19. Mine, S.; Niiyama, M.; Hashimoto, W.; Ikegami, T.; Koma, D.; Ohmoto, T.; Fukuda, Y.; Inoue, T.; Abe, Y.; Ueda, T.; et al. Expression from engineered *Escherichia coli* chromosome and crystallographic study of archaeal *N*,*N*-diacetylchitobiose deacetylase. *FEBS J.* **2014**, *281*, 2584–2596. [CrossRef] [PubMed]
20. Fadouloglou, V.E.; Deli, A.; Glykos, N.M.; Psylinakis, E.; Bouriotis, V.; Kokkinidis, M. Crystal structure of the BcZBP, a zinc-binding protein from *Bacillus cereus*. *FEBS J.* **2007**, *274*, 3044–3054. [CrossRef] [PubMed]
21. Verma, S.C.; Mahadevan, S. The ChbG gene of the chitobiose (chb) operon of *Escherichia coli* encodes a chitooligosaccharide deacetylase. *J. Bacteriol.* **2012**, *194*, 4959–4971. [CrossRef] [PubMed]
22. Araki, Y.; Ito, E. A pathway of chitosan formation in *Mucor rouxii*: Enzymatic deacetlation of chitin. *Biochem. Biophys. Res. Commun.* **1974**, *56*, 669–675. [CrossRef]

23. Araki, Y.; Ito, E. A Pathway of Chitosan Formation in *Mucor rouxii* Enzymatic Deacetylation of Chitin. *Eur. J. Biochem.* **1975**, *55*, 71–78. [CrossRef] [PubMed]
24. Zhao, Y.; Ju, W.; Jo, G.; Jung, W.; Park, R. Perspectives of Chitin Deacetylase Research. In *Biotechnology of Biopolymers*; InTech: London, UK, 2011; pp. 131–145. [CrossRef]
25. Liu, Z.; Gay, L.M.; Tuveng, T.R.; Agger, J.W.; Westereng, B.; Mathiesen, G.; Horn, S.J.; Vaaje-Kolstad, G.; van Aalten, D.M.F.; Eijsink, V.G.H. Structure and function of a broad-specificity chitin deacetylase from *Aspergillus nidulans* FGSC A4. *Sci. Rep.* **2017**, *7*, 1746. [CrossRef] [PubMed]
26. Hoßbach, J.; Bußwinkel, F.; Kranz, A.; Wattjes, J.; Cord-Landwehr, S.; Moerschbacher, B.M. A chitin deacetylase of *Podospora anserina* has two functional chitin binding domains and a unique mode of action. *Carbohydr. Polym.* **2018**, *183*, 1–10. [CrossRef] [PubMed]
27. Ghormade, V.; Kulkarni, S.; Doiphode, N.; Rajamohanan, P.R.; Deshpande, M.V. Chitin deacetylase: A comprehensive account on its role in nature and its biotechnological applications. In *Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology*; Méndez-Vilas, A., Ed.; Formatex Research Center: Badajoz, Spain, 2010; pp. 1054–1066.
28. Tsigos, I.; Martinou, A.; Kafetzopoulos, D.; Bouriotis, V. Chitin deacetylases: New, versatile tools in biotechnology. *Trends Biotechnol.* **2000**, *18*, 305–312. [CrossRef]
29. Zhao, Y.; Park, R.D.; Muzzarelli, R.A.A. Chitin deacetylases: Properties and applications. *Mar. Drugs* **2010**, *8*, 24–46. [CrossRef] [PubMed]
30. Martinou, A.; Bouriotis, V.; Stokke, B.T.; Vårum, K.M. Mode of action of chitin deacetylase from *Mucor rouxii* on partially *N*-acetylated chitosans. *Carbohydr. Res.* **1998**, *311*, 71–78. [CrossRef]
31. Tsigos, I.; Zydowicz, N.; Martinou, A.; Domard, A.; Bouriotis, V. Mode of action of chitin deacetylase from *Mucor rouxii* on *N*-acetylchitooligosaccharides. *Eur. J. Biochem.* **1999**, *261*, 698–705. [CrossRef] [PubMed]
32. Tokuyasu, K.; Mitsutomi, M.; Yamaguchi, I.; Hayashi, K.; Mori, Y. Recognition of chitooligosaccharides and their *N*-acetyl groups by putative subsites of chitin deacetylase from a Deuteromycete, *Colletotrichum lindemuthianum*. *Biochemistry* **2000**, *39*, 8837–8843. [CrossRef] [PubMed]
33. Hekmat, O.; Tokuyasu, K.; Withers, S.G. Subsite structure of the endo-type chitin deacetylase from a deuteromycete, *Colletotrichum lindemuthianum*: An investigation using steady-state kinetic analysis and MS. *Biochem. J.* **2003**, *374*, 369–380. [CrossRef] [PubMed]
34. Gooday, G.W. Chitin deacetylases in invertebrates. In *Chitin in Nature and Technology*; Springer: Boston, MA, USA, 1986; pp. 263–267.
35. Muthukrishnan, S.; Merzendorfer, H.; Arakane, Y.; Yang, Q. Chitin Metabolic Pathways in Insects and Their Regulation. In *Extracellular Composite Matrices in Arthropods*; Cohen, E., Moussian, B., Eds.; Springer: Berlin, Germany, 2016; pp. 31–65. ISBN 9783319407401.
36. Dixit, R.; Arakane, Y.; Specht, C.A.; Richard, C.; Kramer, K.J.; Beeman, R.W.; Muthukrishnan, S. Domain organization and phylogenetic analysis of proteins from the chitin deacetylase gene family of Tribolium castaneum and three other species of insects. *Insect Biochem. Mol. Biol.* **2008**, *38*, 440–451. [CrossRef] [PubMed]
37. Baker, L.G.; Specht, C.A.; Donlin, M.J.; Lodge, J.K. Chitosan, the deacetylated form of chitin, is necessary for cell wall integrity in *Cryptococcus neoformans*. *Eukaryot. Cell* **2007**, *6*, 855–867. [CrossRef] [PubMed]
38. Christodoulidou, A.; Bouriotis, V.; Thireos, G. Two sporulation-specific chitin deacetylase-encoding genes are required for the ascospore wall rigidity of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **1996**, *271*, 31420–31425. [CrossRef] [PubMed]
39. Geoghegan, I.A.; Gurr, S.J. Chitosan Mediates Germling Adhesion in Magnaporthe oryzae and Is Required for Surface Sensing and Germling Morphogenesis. *PLoS Pathog.* **2016**, *12*, 1–34. [CrossRef] [PubMed]
40. White, S.; McIntyre, M.; Berry, D.R.; McNeil, B. The autolysis of industrial filamentous fungi. *Crit. Rev. Biotechnol.* **2002**, *22*, 1–14. [CrossRef] [PubMed]
41. Sánchez-Vallet, A.; Mesters, J.R.; Thomma, B.P. The battle for chitin recognition in plant-microbe interactions. *FEMS Microbiol. Rev.* **2015**, *39*, 171–183. [CrossRef] [PubMed]
42. Davis, L.L.; Bartnicki-Garcia, S. Chitosan Synthesis by the Tandem Action of Chitin Synthetase and Chitin Deacetylase from *Mucor rouxii*. *Biochemistry* **1984**, *23*, 1065–1073. [CrossRef]
43. Gao, X.-D.; Katsumoto, T.; Onodera, K. Purification and characterization of chitin deacetylase from *Absidia coerulea*. *J. Biochem.* **1995**, *117*, 257–263. [CrossRef] [PubMed]

44. Christodoulidou, A.; Briza, P.; Ellinger, A.; Bouriotis, V. Yeast ascospore wall assembly requires two chitin deacetylase isozymes. *FEBS Lett.* **1999**, *460*, 275–279. [CrossRef]
45. Hadwiger, L.A. Anatomy of a nonhost disease resistance response of pea to Fusarium solani: PR gene elicitation via DNase, chitosan and chromatin alterations. *Front. Plant Sci.* **2015**, *6*, 373. [CrossRef] [PubMed]
46. Hadwiger, L.A. Pea-Fusarium solani interactions contributions of a system toward understanding disease resistance. *Phytopathology* **2008**, *98*, 372–379. [CrossRef] [PubMed]
47. El Gueddari, N.E.; Rauchhaus, U.; Moerschbacher, B.M.; Deising, H.B. Developmentally regulated conversion of surface-exposed chitin to chi-tosan in cell walls of plant pathogenic fungi. *New Phytol.* **2002**, *156*, 103–112. [CrossRef]
48. Liu, T.; Liu, Z.; Song, C.; Hu, Y.; Han, Z.; She, J.; Fan, F.; Wang, J.; Jin, C.; Chang, J.; et al. Chitin-Induced Dimerization Activates a Plant Immune Receptor. *Science* **2012**, *336*, 1160–1164. [CrossRef] [PubMed]
49. Cord-Landwehr, S.; Melcher, R.L.J.; Kolkenbrock, S.; Moerschbacher, B.M. A chitin deacetylase from the endophytic fungus *Pestalotiopsis* sp. efficiently inactivates the elicitor activity of chitin oligomers in rice cells. *Sci. Rep.* **2016**, *6*, 38018. [CrossRef] [PubMed]
50. Emri, T.; Molnár, Z.; Szilágyi, M.; Pócsi, I. Regulation of autolysis in *Aspergillus nidulans*. *Appl. Biochem. Biotechnol.* **2008**, *151*, 211–220. [CrossRef] [PubMed]
51. Alfonso, C.; Nuero, O.M.; Santamaría, F.; Reyes, F. Purification of a heat-stable chitin deacetylase from *Aspergillus nidulans* and its role in cell wall degradation. *Curr. Microbiol.* **1995**, *30*, 49–54. [CrossRef] [PubMed]
52. Reyes, F.; Calatayud, J.; Martinez, M.J. Endochitinase from *Aspergillus nidulans* implicated in the autolysis of its cell wall. *FEMS Microbiol. Lett.* **1989**, *51*, 119–124. [CrossRef] [PubMed]
53. Blair, D.E.; Hekmat, O.; Schüttelkopf, A.W.; Shrestha, B.; Tokuyasu, K.; Withers, S.G.; van Aalten, D.M. Structure and Mechanism of Chitin Deacetylase from the Fungal Pathogen Colletotrichium lindemuthianum. *Biochemistry* **2006**, *45*, 9416–9426. [CrossRef] [PubMed]
54. Naqvi, S.; Cord-Landwehr, S.; Singh, R.; Bernard, F.; Kolkenbrock, S.; El Gueddari, N.E.; Moerschbacher, B.M. A recombinant fungal chitin deacetylase produces fully defined chitosan oligomers with novel patterns of acetylation. *Appl. Environ. Microbiol.* **2016**, *82*, 6645–6655. [CrossRef] [PubMed]
55. Aranda-Martinez, A.; Grifoll-Romero, L.; Aragunde Pazos, H.; Enea Sancho-Vaello; Biarnés, X.; Lopez-Llorca, L.V.; Planas, A. Expression and specificity of a chitin deacetylase catalytic doain from the nematophagous fungus *Pochonia chlamydosporia* potentially involved in pathogenicity. *Sci. Rep.* **2018**, *8*, 2170. [CrossRef] [PubMed]
56. Mishra, C.; Mishra, C.; Semino, C.; Semino, C.; Mccreath, K.J.; Mccreath, K.J.; de la Vega, H.; de la Vega, H.; Jones, B.J.; Jones, B.J.; et al. Cloning and expression of two chitin deacetylase gens of *Saccharomyces cerevisiae*. *Yeast* **1997**, *13*, 327–336. [CrossRef]
57. Martinou, A.; Koutsioulis, D.; Bouriotis, V. Cloning and expression of a chitin deacetylase gene (CDA2) from *Saccharomyces cerevisiae* in *Escherichia coli*: Purification and characterization of the cobalt-dependent recombinant enzyme. *Enzyme Microb. Technol.* **2003**, *32*, 757–763. [CrossRef]
58. Kim, Y.J.; Zhao, Y.; Oh, K.T.; Nguyen, V.N.; Park, R.D. Enzymatic deacetylation of chitin by extracellular chitin deacetylase from a newly screened *Mortierella* sp. DY-52. *J. Microbiol. Biotechnol.* **2008**, *18*, 759–766. [PubMed]
59. Zhao, Y.; Kim, Y.J.; Oh, K.T.; Nguyen, V.N.; Park, R.D. Production and characterization of extracellular chitin deacetylase from *Absidia corymbifera* DY-9. *J. Appl. Biol. Chem.* **2010**, *53*, 119–126. [CrossRef]
60. Yamada, M.; Kurano, M.; Inatomi, S.; Taguchi, G.; Okazaki, M.; Shimosaka, M. Isolation and characterization of a gene coding for chitin deacetylase specifically expressed during fruiting body development in the basidiomycete *Flammulina velutipes* and its expression in the yeast *Pichia pastoris*. *FEMS Microbiol. Lett.* **2008**, *289*, 130–137. [CrossRef] [PubMed]
61. Pareek, N.; Vivekanand, V.; Saroj, S.; Sharma, A.K.; Singh, R.P. Purification and characterization of chitin deacetylase from *Penicillium oxalicum* SAEM-51. *Carbohydr. Polym.* **2012**, *87*, 1091–1097. [CrossRef]
62. Karthik, N.; Binod, P.; Pandey, A. SSF production, purification and characterization of chitin deacetylase from *Aspergillus flavus*. *Biocatal. Biotransform.* **2017**. [CrossRef]
63. Cai, J.; Yang, J.; Du, Y.; Fan, L.; Qiu, Y.; Li, J.; Kennedy, J.F. Purification and characterization of chitin deacetylase from *Scopulariopsis brevicaulis*. *Carbohydr. Polym.* **2006**, *65*, 211–217. [CrossRef]

64. Gauthier, C.; Clerisse, F.; Dommes, J.; Jaspar-Versali, M.F. Characterization and cloning of chitin deacetylases from *Rhizopus circinans*. *Protein Expr. Purif.* **2008**, *59*, 127–137. [CrossRef] [PubMed]
65. Maw, T.; Tan, T.K.; Khor, E.; Wong, S.M. Complete cDNA sequence of chitin deacetylase from *Gongronella butleri* and its phylogenetic analysis revealed clusters corresponding to taxonomic classification of fungi. *J. Biosci. Bioeng*. **2002**, *93*, 376–381. [CrossRef]
66. Mélida, H.; Sain, D.; Stajich, J.E.; Bulone, V. Deciphering the uniqueness of Mucoromycotina cell walls by combining biochemical and phylogenomic approaches. *Environ. Microbiol.* **2015**, *17*, 1649–1662. [CrossRef] [PubMed]
67. Smirnou, D.; Krcmar, M.; Prochazkova, E.V.A. Chitin-Glucan complex production by *Schizophyllum commune* submerged cultivation. *Pol. J. Microbiol.* **2011**, *60*, 223–228. [PubMed]
68. Das, S.; Van Dellen, K.; Bulik, D.; Magnelli, P.; Cui, J.; Head, J.; Robbins, P.W.; Samuelson, J. The cyst wall of *Entamoeba invadens* contains chitosan (deacetylated chitin). *Mol. Biochem. Parasitol.* **2006**, *148*, 86–92. [CrossRef] [PubMed]
69. Andrés, E.; Albesa-Jové, D.; Biarnés, X.; Moerschbacher, B.M.; Guerin, M.E.; Planas, A. Structural basis of chitin oligosaccharide deacetylation. *Angew. Chem. Int. Ed.* **2014**, *53*, 6882–6887. [CrossRef] [PubMed]
70. Li, X.; Wang, L.X.; Wang, X.; Roseman, S. The chitin catabolic cascade in the marine bacterium *Vibrio cholerae*: Characterization of a unique chitin oligosaccharide deacetylase. *Glycobiology* **2007**, *17*, 1377–1387. [CrossRef] [PubMed]
71. Hirano, T.; Sugiyama, K.; Sakaki, Y.; Hakamata, W.; Park, S.Y.; Nishio, T. Structure-based analysis of domain function of chitin oligosaccharide deacetylase from *Vibrio parahaemolyticus*. *FEBS Lett.* **2015**, *589*, 145–151. [CrossRef] [PubMed]
72. Kadokura, K.; Rokutani, A.; Yamamoto, M.; Ikegami, T.; Sugita, H.; Itoi, S.; Hakamata, W.; Oku, T.; Nishio, T. Purification and characterization of *Vibrio parahaemolyticus* extracellular chitinase and chitin oligosaccharide deacetylase involved in the production of heterodisaccharide from chitin. *Appl. Microbiol. Biotechnol.* **2007**, *75*, 357–365. [CrossRef] [PubMed]
73. Ohishi, K.; Yamagishi, M.; Ohta, T.; Motosugi, M.; Izumida, H.; Sano, H.; Adachi, K.; Miwa, T. Purification and Properties of Two Deacetylases Produced by *Vibrio alginolyticus* H-8. *Biosci. Biotechnol. Biochem.* **1997**, *61*, 1113–1117. [CrossRef]
74. Hirano, T.; Uehara, R.; Shiraishi, H.; Hakamata, W.; Nishio1, T. Chitin Oligosaccharide Deacetylase from *Shewanella woodyi* ATCC51908. *J. Appl. Glycosci.* **2015**, *62*, 153–157. [CrossRef]
75. Hirano, T.; Shiraishi, H.; Ikejima, M.; Uehara, R.; Hakamata, W.; Nishio, T. Chitin oligosaccharide deacetylase from *Shewanella baltica* ATCC BAA-1091. *Biosci. Biotechnol. Biochem.* **2017**, *81*, 547–550. [CrossRef] [PubMed]
76. Tuveng, T.R.; Rothweiler, U.; Udatha, G.; Vaaje-Kolstad, G.; Smalås, A.; Eijsink, V.G.H. Structure and function of a CE4 deacetylase isolated from a marine environment. *PLoS ONE* **2017**, *12*, e0187544. [CrossRef] [PubMed]
77. Bartnicki-Garcia, S.; Nickerson, W.J. Isolation, composition, and structure of cell walls of filamentous and yeast-like forms of *Mucor rouxii*. *Biochim. Biophys. Acta* **1962**, *58*, 102–119. [CrossRef]
78. Kafetzopoulos, D.; Martinou, A.; Bouriotis, V. Bioconversion of chitin to chitosan: Purification and characterization of chitin deacetylase from *Mucor rouxii*. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 2564–2568. [CrossRef] [PubMed]
79. Davis, L.L.; Bartnicki-Garcia, S. The co-ordination of chitosan and chitin synthesis in *Mucor rouxii*. *J. Gen. Microbiol.* **1984**, *130*, 2095–2102. [CrossRef] [PubMed]
80. Chatterjee, S.; Adhya, M.; Guha, A.K.; Chatterjee, B.P. Chitosan from *Mucor rouxii*: Production and physico-chemical characterization. *Process Biochem.* **2005**, *40*, 395–400. [CrossRef]
81. Synowiecki, J.; Al-Khateeb, N.A.A.Q. Mycelia of *Mucor rouxii* as a source of chitin and chitosan. *Food Chem.* **1997**, *60*, 605–610. [CrossRef]
82. Martinou, A.; Kafetzopoulos, D.; Bouriotis, V. Isolation of chitin deacetylase from *Mucor rouxii* by immunoaffinity chromatography. *J. Chromatogr. A* **1993**, *644*, 35–41. [CrossRef]
83. Kauss, H.; Bauch, B. Chitin deacetylase from *Colletotrichum lindemuthianum*. *Methods Enzymol.* **1988**, *161*, 518–523. [CrossRef]
84. Tsigos, I.; Bouriotis, V. Purification and characterization of chitin deacetylase from *Colletotrichum lindemuthianum*. *J. Biol. Chem.* **1995**, *270*, 26286–26291. [CrossRef] [PubMed]

85. O'Connell, R.J.; Ride, J.P. Chemical detection and ultrastructural localization of chitin in cell walls of *Colletotrichum lindemuthianum*. *Physiol. Mol. Plant Pathol.* **1990**, *37*, 39–53. [CrossRef]
86. Tokuyasu, K.; Ohnishi-Kameyama, M.; Hayashi, K. Purification and characterization of extracellular chitin deacetylase from *Colletotrichum lindemuthianum*. *Biosci. Biotechnol. Biochem.* **1996**, *60*, 1598–1603. [CrossRef] [PubMed]
87. Shrestha, B.; Blondeau, K.; Stevens, W.F.; Hegarat, F.L. Expression of chitin deacetylase from *Colletotrichum lindemuthianum* in *Pichia pastoris*: Purification and characterization. *Protein Expr. Purif.* **2004**, *38*, 196–204. [CrossRef] [PubMed]
88. Kang, L.; Chen, X.; Zhai, C.; Ma, L. Synthesis and high expression of chitin deacetylase from *Colletotrichum lindemuthianum* in *Pichia pastoris* GS115. *J. Microbiol. Biotechnol.* **2012**, *22*, 1202–1207. [CrossRef] [PubMed]
89. Tokuyasu, K.; Kaneko, S.; Hayashi, K.; Mori, Y. Production of a recombinant chitin deacetylase in the culture medium of *Escherichia coli* cells. *FEBS Lett.* **1999**, *458*, 23–26. [CrossRef]
90. Tokuyasu, K.; Ohnishi-Kameyama, M.; Hayashi, K.; Mori, Y. Cloning and expression of chitin deacetylase gene from a deuteromycete, *Colletotrichum lindemuthianum*. *J. Biosci. Bioeng.* **1999**, *87*, 418–423. [CrossRef]
91. Tokuyasu, K.; Ono, H.; Ohnishi-Kameyama, M.; Hayashi, K.; Mori, Y. Deacetylation of chitin oligosaccharides of dp 2-4 by chitin deacetylase from *Colletotrichum lindemuthianum*. *Carbohydr. Res.* **1997**, *303*, 353–358. [CrossRef]
92. Tokuyasu, K.; Ono, H.; Hayashi, K.; Mori, Y. Reverse hydrolysis reaction of chitin deacetylase and enzymatic synthesis of β-D-GlcNAc-(1→4)-GlcN from chitobiose. *Carbohydr. Res.* **1999**, *322*, 26–31. [CrossRef]
93. Tokuyasu, K.; Ono, H.; Mitsutomi, M.; Hayashi, K.; Mori, Y. Synthesis of a chitosan tetramer derivative, β-D-GlcNAc-(1→4)-β-D-GlcNAc-(1→4)-β-D-GlcNAc-(1→4)-D-GlcN through a partial *N*-acetylation reaction by chitin deacetylase. *Carbohydr. Res.* **2000**, *325*, 211–215. [CrossRef]
94. Kang, L.X.; Liang, Y.X.; Ma, L.X. Novel characteristics of chitin deacetylase from *Colletotrichum lindemuthianum*: Production of fully acetylated chitooligomers, and hydrolysis of deacetylated chitooligomers. *Process Biochem.* **2014**, *49*, 1936–1940. [CrossRef]
95. Reyes, F.; Calatayud, J.; Vazquez, C.; Martínez, M.J. β-*N*-Acetylglucosaminidase from *Aspergillus nidulans* which degrades chitin oligomers during autolysis. *FEMS Microbiol. Lett.* **1990**, *65*, 83–87. [CrossRef]
96. Wang, Y.; Song, J.Z.; Yang, Q.; Liu, Z.H.; Huang, X.M.; Chen, Y. Cloning of a heat-stable chitin deacetylase gene from *Aspergillus nidulans* and its functional expression in *Escherichia coli*. *Appl. Biochem. Biotechnol.* **2010**, *162*, 843–854. [CrossRef] [PubMed]
97. Espagne, E.; Lespinet, O.; Malagnac, F.; Da Silva, C.; Jaillon, O.; Porcel, B.M.; Couloux, A.; Aury, J.-M.; Ségurens, B.; Poulain, J.; Anthouard, V.; et al. The genome sequence of the model ascomycete fungus *Podospora anserina*. *Genome Biol.* **2008**, *9*, R77. [CrossRef] [PubMed]
98. Lorin, S.; Dufour, E.; Sainsard-Chanet, A. Mitochondrial metabolism and aging in the filamentous fungus *Podospora anserina*. *Biochim. Biophys. Acta Bioenerg.* **2006**, *1757*, 604–610. [CrossRef] [PubMed]
99. Figueroa, M.; Upadhyaya, N.M.; Sperschneider, J.; Park, R.F.; Szabo, L.J.; Steffenson, B.; Ellis, J.G.; Dodds, P.N. Changing the Game: Using Integrative Genomics to Probe Virulence Mechanisms of the Stem Rust Pathogen *Puccinia graminis* f. sp. tritici. *Front. Plant Sci.* **2016**, *7*, 1–10. [CrossRef] [PubMed]
100. Singh, R.P.; Hodson, D.P.; Jin, Y.; Lagudah, E.S.; Ayliffe, M.A.; Bhavani, S.; Rouse, M.N.; Pretorius, Z.A.; Szabo, L.J.; Huerta-Espino, J.; et al. Emergence and Spread of New Races of Wheat Stem Rust Fungus: Continued Threat to Food Security and Prospects of Genetic Control. *Phytopathology* **2015**, *105*, 872–884. [CrossRef] [PubMed]
101. Jin, Y.; Singh, R.P. Resistance in U.S. Wheat to Recent Eastern African Isolates of *Puccinia graminis* f. sp. tritici with Virulence to Resistance Gene Sr31. *Plant Dis.* **2006**, *90*, 476–480. [CrossRef]
102. Mendgen, K.; Hahn, M. Plant infection and the establishment of fungal biotrophy. *Trends Plant Sci.* **2002**, *7*, 352–356. [CrossRef]
103. Broeker, K.; Fehser, S.; Tenberge, K.B.; Moerschbacher, B.M. Two class III chitin synthases specifically localized in appressoria and haustoria of *Puccinia graminis* f. sp. tritici. *Physiol. Mol. Plant Pathol.* **2011**, *76*, 27–33. [CrossRef]
104. Ride, J.P.; Barber, M.S. Purification and characterization of multiple forms of endochitinase from wheat leaves. *Plant Sci.* **1990**, *71*, 185–197. [CrossRef]

105. Vander, P.; Vårum, K.M.; Domard, A.; El Gueddari, N.E.; Moerschbacher, B.M. Comparison of the Ability of Partially *N*-Acetylated Chitosans and Chitooligosaccharides to Elicit Resistance Reactions in Wheat Leaves. *Plant Physiol.* **1998**, *118*, 1353–1359. [CrossRef] [PubMed]
106. Maharachchikumbura, S.S.N.; Hyde, K.D.; Groenewald, J.Z.; Xu, J.; Crous, P.W. *Pestalotiopsis* revisited. *Stud. Mycol.* **2014**, *79*, 121–186. [CrossRef] [PubMed]
107. Lopez-Llorca, L.V.; Olivares-Bernabeu, C.; Salinas, J.; Jansson, H.-B.; Kolattukudy, P.E. Pre-penetration events in fungal parasitism of nematode eggs. *Mycol. Res.* **2002**, *106*, 499–506. [CrossRef]
108. Manzanilla-Lopez, R.H.; Esteves, I.; Finetti-Sialer, M.M.; Hirsch, P.R.; Ward, E.; Devonshire, J.; Hidalgo-Diaz, L. *Pochonia chlamydosporia*: Advances and Challenges to Improve Its Performance as a Biological Control Agent of Sedentary Endo-parasitic Nematodes. *J. Nematol.* **2013**, *45*, 1–7. [PubMed]
109. Larriba, E.; Jaime, M.D.L.A.; Carbonell-Caballero, J.; Conesa, A.; Dopazo, J.; Nislow, C.; Martín-Nieto, J.; Lopez-Llorca, L.V. Sequencing and functional analysis of the genome of a nematode egg-parasitic fungus, *Pochonia chlamydosporia*. *Fungal Genet. Biol.* **2014**, *65*, 69–80. [CrossRef] [PubMed]
110. Aranda-Martinez, A.; Lenfant, N.; Escudero, N.; Zavala-Gonzalez, E.A.; Henrissat, B.; Lopez-Llorca, L.V. CAZyme content of *Pochonia chlamydosporia* reflects that chitin and chitosan modification are involved in nematode parasitism. *Environ. Microbiol.* **2016**, *18*, 4200–4215. [CrossRef] [PubMed]
111. Wani, Z.A.; Kumar, A.; Sultan, P.; Bindu, K.; Riyaz-Ul-Hassan, S.; Ashraf, N. *Mortierella* alpina CS10E4, an oleaginous fungal endophyte of *Crocus sativus* L. enhances apocarotenoid biosynthesis and stress tolerance in the host plant. *Sci. Rep.* **2017**, *7*, 8598. [CrossRef] [PubMed]
112. Zhao, Y.; Jo, G.-H.; Ju, W.-T.; Jung, W.-J.; Park, R.-D. A Highly *N*-Glycosylated Chitin Deacetylase Derived from a Novel Strain of *Mortierella* sp. DY-52. *Biosci. Biotechnol. Biochem.* **2011**, *75*, 960–965. [CrossRef] [PubMed]
113. Narayanan, K.; Parameswaran, B.; Pandey, A. Production of chitin deacetylase by *Aspergillus flavus* in submerged conditions. *Prep. Biochem. Biotechnol.* **2016**, *46*, 501–508. [CrossRef] [PubMed]
114. Cuenca-Estrella, M.; Gomez-Lopez, A.; Mellado, E.; Buitrago, M.J.; Monzón, A.; Rodriguez-Tudela, J.L. *Scopulariopsis brevicaulis*, a fungal pathogen resistant to broad-spectrum antifungal agents. *Antimicrob. Agents Chemother.* **2003**, *47*, 2339–2341. [CrossRef] [PubMed]
115. Tosti, A.; PiraccinI, B.M.; Stinchi, C.; Lorenzi, S. Onychomycosis due to *Scopulariopsis brevicaulis*: Clinical features and response to systemic antifungals. *Br. J. Dermatol.* **1996**, *135*, 799–802. [CrossRef] [PubMed]
116. El Ghaouth, A.; Arul, J.; Grenier, J.; Asselin, A. Effect of chitosan and other polyions on chitin deacetylase in *Rhizopus stolonifer*. *Exp. Mycol.* **1992**, *16*, 173–177. [CrossRef]
117. Zhang, H.; Yang, S.; Fang, J.; Deng, Y.; Wang, D.; Zhao, Y. Optimization of the fermentation conditions of *Rhizopus japonicus* M193 for the production of chitin deacetylase and chitosan. *Carbohydr. Polym.* **2014**, *101*, 57–67. [CrossRef] [PubMed]
118. Maw, T.; Tan, T.K.; Khor, E.; Wong, S.M. Selection of *Gongronella butleri* strains for enhanced chitosan yield with UV mutagenesis. *J. Biotechnol.* **2002**, *95*, 189–193. [CrossRef]
119. Yonemura, A.; Nagashima, T.; Murayama, T. *Expression of Chitin Deacetylase Gene from Phycomyces blakesleeanus in Aspergillus oryzae and Neurospora crassa*; The Society for Bioscience and Bioengineering: Osaka, Japan, 2007; p. 129. Available online: http://dl.ndl.go.jp/view/download/digidepo_10529404_po_ART0009175183.pdf?contentNo=1&alternativeNo= (accessed on 19 February 2018).
120. Baker, L.G.; Specht, C.A.; Lodge, J.K. Cell wall chitosan is necessary for virulence in the opportunistic pathogen *Cryptococcus neoformans*. *Eukaryot. Cell* **2011**, *10*, 1264–1268. [CrossRef] [PubMed]
121. Doering, T.L. How Sweet it is! Cell Wall Biogenesis and Polysaccharide Capsule Formation in *Cryptococcus neoformans*. *Annu. Rev. Microbiol.* **2009**, *63*, 223–247. [CrossRef] [PubMed]
122. Gilbert, N.M.; Baker, L.G.; Specht, C.A.; Lodge, J.K. A glycosylphosphatidylinositol anchor is required for membrane localization but dispensable for cell wall association of chitin deacetylase 2 in *Cryptococcus neoformans*. *mBio* **2012**, *3*, e00007–e00012. [CrossRef] [PubMed]
123. Levitz, S.M.; Nong, S.-H.; Mansour, M.K.; Huang, C.; Specht, C.A. Molecular characterization of a mannoprotein with homology to chitin deacetylases that stimulates T cell responses to *Cryptococcus neoformans*. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10422–10427. [CrossRef] [PubMed]
124. Biondo, C.; Beninati, C.; Delfino, D.; Oggioni, M.; Mancuso, G.; Midiri, A.; Tomaselli, G.; Teti, G.; Biondo, C.; Beninati, C.; et al. Identification and Cloning of a Cryptococcal Deacetylase That Produces Protective Immune Responses. *Infect. Immun.* **2002**, *70*, 2383–2391. [CrossRef] [PubMed]

125. Loftus, B.; Anderson, I.; Davies, R.; Alsmark, U.C.M.; Samuelson, J.; Amedeo, P.; Roncaglia, P.; Berriman, M.; Hirt, R.P.; Mann, B.J.; et al. The genome of the protist parasite *Entamoeba histolytica*. *Nature* **2005**, *433*, 865–868. [CrossRef] [PubMed]
126. Roche, P.; Maillet, F.; Plazanet, C.; Debellé, F.; Ferro, M.; Truchet, G.; Promé, J.C.; Dénarié, J. The common nodABC genes of *Rhizobium meliloti* are host-range determinants. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 15305–15310. [CrossRef] [PubMed]
127. Egelhoff, T.T.; Long, S.R. *Rhizobium meliloti* nodulation genes: Identification of nodDABC gene products, purification of nodA protein, and expression of nodA in *Rhizobium meliloti*. *J. Bacteriol.* **1985**, *164*, 591–599. [PubMed]
128. Spaink, H.P.; Wijfjes, A.H.M.; der van Drift, K.M.G.M.; Haverkamp, J.; Thomas-Oates, J.E.; Lugtenberg, B.J.J. Structural identification of metabolites produced by the NodB and NodC proteins of *Rhizobium* leguminosarum. *Mol. Microbiol.* **1994**, *13*, 821–831. [CrossRef] [PubMed]
129. Chambon, R.; Pradeau, S.; Fort, S.; Cottaz, S.; Armand, S. High yield production of *Rhizobium* NodB chitin deacetylase and its use for in vitro synthesis of lipo-chitinoligosaccharide precursors. *Carbohydr. Res.* **2017**, *442*, 25–30. [CrossRef] [PubMed]
130. Hamer, S.N.; Cord-Landwehr, S.; Biarnés, X.; Planas, A.; Waegeman, H.; Moerschbacher, B.M.; Kolkenbrock, S. Enzymatic production of defined chitosan oligomers with a specific pattern of acetylation using a combination of chitin oligosaccharide deacetylases. *Sci. Rep.* **2015**, *5*, 8716. [CrossRef] [PubMed]
131. Röhrig, H.; Schmidt, J.; Wieneke, U.; Kondorosi, E.; Barlier, I.; Schell, J.; John, M. Biosynthesis of lipooligosaccharide nodulation factors: *Rhizobium* NodA protein is involved in *N*-acylation of the chitooligosaccharide backbone. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 3122–3126. [CrossRef] [PubMed]
132. Keyhani, N.O.; Roseman, S. Physiological aspects of chitin catabolism in marine bacteria. *Biochim. Biophys. Acta Gen. Subj.* **1999**, *1473*, 108–122. [CrossRef]
133. Zobell, C.; Rittenberg, S. The occurrence and characteristics of chitinoclastic bacteria in the sea. *J. Bacteriol.* **1937**, *35*, 275–287.
134. Meibom, K.L.; Li, X.B.; Nielsen, A.T.; Wu, C.-Y.; Roseman, S.; Schoolnik, G.K. The *Vibrio cholerae* chitin utilization program. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 2524–2529. [CrossRef] [PubMed]
135. Li, X.; Roseman, S. The chitinolytic cascade in Vibrios is regulated by chitin oligosaccharides and a two-component chitin catabolic sensor/kinase. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 627–631. [CrossRef] [PubMed]
136. Ohishi, K.; Murase, K.; Ohta, T.; Etoh, H. Cloning and sequencing of the deacetylase gene from *Vibrio alginolyticus* H-8. *J. Biosci. Bioeng.* **2000**, *90*, 561–563. [CrossRef]
137. Kadokura, K.; Sakamoto, Y.; Saito, K.; Ikegami, T.; Hirano, T.; Hakamata, W.; Oku, T.; Nishio, T. Production of a recombinant chitin oligosaccharide deacetylase from *Vibrio parahaemolyticus* in the culture medium of *Escherichia coli* cells. *Biotechnol. Lett.* **2007**, *29*, 1209–1215. [CrossRef] [PubMed]
138. Hirano, T.; Maebara, Y.; Uehara, R.; Sakaki, Y.; Shiraishi, H.; Ichimura, S.; Hakamata, W.; Nishio, T. Chitin oligosaccharide deacetylase from Vibrio harveyi ATCC BAA-1116: Gene cloning, overexpression, purification, and characterization. *Chitin Chitosan Res.* **2012**, *19*, 321–324.
139. Jacquiod, S.; Franqueville, L.; Cécillon, S.; Vogel, T.M.; Simonet, P. Soil bacterial community shifts after Chitin enrichment: An integrative metagenomic approach. *PLoS ONE* **2013**, *8*. [CrossRef] [PubMed]
140. Dsouza, M.; Taylor, M.W.; Turner, S.J.; Aislabie, J. Genomic and phenotypic insights into the ecology of *Arthrobacter* from Antarctic soils. *BMC Genom.* **2015**, *16*, 36. [CrossRef] [PubMed]
141. Lonhienne, T.; Mavromatis, K.; Vorgias, C.E.; Buchon, L.; Gerday, C.; Bouriotis, V. Cloning, sequences, and characterization of two chitinase genes from the Antarctic *Arthrobacter* sp. strain TAD20: Isolation and partial characterization of the enzymes. *J. Bacteriol.* **2001**, *183*, 1773–1779. [CrossRef] [PubMed]
142. Aragunde-pazos, H.; Biarnés, X.; Planas, A. Substrate recognition and specificity of chitin deacetylases and related family 4 carbohydrate esterases. *Int. J. Mol. Sci.* **2018**, *19*, 412. [CrossRef] [PubMed]
143. Boraston, A.B.; Bolam, D.N.; Gilbert, H.J.; Davies, G.J. Carbohydrate-Binding modules: Fine-Tuning polysaccharide recognition. *Biochem. J.* **2004**, *382*, 769–781. [CrossRef] [PubMed]
144. Blair, D.E.; Schuttelkopf, A.W.; MacRae, J.I.; van Aalten, D.M.F. Structure and metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal virulence factor. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15429–15434. [CrossRef] [PubMed]

145. Blair, D.E.; Van Aalten, D.M.F. Structures of *Bacillus subtilis* PdaA, a family 4 carbohydrate esterase, and a complex with *N*-acetyl-glucosamine. *FEBS Lett.* **2004**, *570*, 13–19. [CrossRef] [PubMed]
146. Nakamura, A.M.; Nascimento, A.S.; Polikarpov, I. Structural diversity of carbohydrate esterases. *Biotechnol. Res. Innov.* **2017**, *1*, 35–51. [CrossRef]
147. Nishiyama, T.; Noguchi, H.; Yoshida, H.; Park, S.Y.; Tame, J.R.H. The structure of the deacetylase domain of *Escherichia coli* PgaB, an enzyme required for biofilm formation: A circularly permuted member of the carbohydrate esterase 4 family. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69*, 44–51. [CrossRef] [PubMed]
148. Hernick, M.; Fierke, C.A. Zinc hydrolases: The mechanisms of zinc-dependent deacetylases. *Arch. Biochem. Biophys.* **2005**, *433*, 71–84. [CrossRef] [PubMed]
149. Xie, J.L.; Polvi, E.J.; Shekhar-Guturja, T.; Cowen, L.E. Elucidating drug resistance in human fungal pathogens. *Future Microbiol.* **2014**, *9*, 523–542. [CrossRef] [PubMed]
150. Takaya, N.; Yamazaki, D.; Horiuchi, H.; Ohta, A.; Takagi, M. Cloning and characterization of a chitinase-encoding gene (chiA) from *Aspergillus nidulans*, disruption of which decreases germination frequency and hyphal growth. *Biosci. Biotechnol. Biochem.* **1998**, *62*, 60–65. [CrossRef] [PubMed]
151. Hartl, L.; Zach, S.; Seidl-Seiboth, V. Fungal chitinases: Diversity, mechanistic properties and biotechnological potential. *Appl. Microbiol. Biotechnol.* **2012**, *93*, 533–543. [CrossRef] [PubMed]
152. Aoun, M. Host defense mechanisms during fungal pathogenesis and how these are overcome in susceptible plants: A review. *Int. J. Bot.* **2017**, *13*, 82–102. [CrossRef]
153. Huang, G. Chitinase Inhibitor Allosamidin and Its Analogues: An Update. *Curr. Org. Chem.* **2012**, *16*, 115–120. [CrossRef]
154. Rao, F.V.; Houston, D.R.; Boot, R.G.; Aerts, J.M.F.G.; Hodkinson, M.; Adams, D.J.; Shiomi, K.; Omura, S.; Van Aalten, D.M.F. Specificity and affinity of natural product cyclopentapeptide inhibitors against *A. fumigatus*, human, and bacterial chitinases. *Chem. Biol.* **2005**, *12*, 65–76. [CrossRef] [PubMed]
155. Younes, I.; Rinaudo, M. Chitin and chitosan preparation from marine sources. Structure, properties and applications. *Mar. Drugs* **2015**, *13*, 1133–1174. [CrossRef] [PubMed]
156. Cheung, R.C.F.; Ng, T.B.; Wong, J.H.; Chan, W.Y. Chitosan: An update on potential biomedical and pharmaceutical applications. *Mar. Drugs* **2015**, *13*, 5156–5186. [CrossRef] [PubMed]
157. Anitha, A.; Sowmya, S.; Kumar, P.T.T.S.; Deepthi, S.; Chennazhi, K.P.; Ehrlich, H.; Tsurkan, M.; Jayakumar, R. Chitin and chitosan in selected biomedical applications. *Prog. Polym. Sci.* **2014**, *39*, 1644–1667. [CrossRef]
158. Abdul Khalil, H.P.S.; Saurabh, C.K.; Adnan, A.S.; Nurul Fazita, M.R.; Syakir, M.I.; Davoudpour, Y.; Rafatullah, M.; Abdullah, C.K.; Haafiz, M.K.M.; Dungani, R. A review on chitosan-cellulose blends and nanocellulose reinforced chitosan biocomposites: Properties and their applications. *Carbohydr. Polym.* **2016**, *150*, 216–226.
159. Pestov, A.; Bratskaya, S. Chitosan and Its Derivatives as Highly Efficient Polymer Ligands. *Molecules* **2016**, *21*, 330. [CrossRef] [PubMed]
160. Das, S.N.; Madhuprakash, J.; Sarma, P.V.S.R.N.; Purushotham, P.; Suma, K.; Manjeet, K.; Rambabu, S.; Gueddari, N.E.E.; Moerschbacher, B.M.; Podile, A.R. Biotechnological approaches for field applications of chitooligosaccharides (COS) to induce innate immunity in plants. *Crit. Rev. Biotechnol.* **2015**, *35*, 29–43. [CrossRef] [PubMed]
161. Sorlier, P.; Denuzière, A.; Viton, C.; Domard, A. Relation between the degree of acetylation and the electrostatic properties of chitin and chitosan. *Biomacromolecules* **2001**, *2*, 765–772. [CrossRef] [PubMed]
162. Omura, Y.; Shigemoto, M.; Akiyama, T.; Saimoto, H.; Shigemasa, Y.; Nakamura, I.; Tsuchido, T. Antimicrobial Activity of Chitosan with Different Degrees of Acetylation and Molecular Weights. *Biocontrol Sci.* **2003**, *8*, 25–30. [CrossRef]
163. Domard, A.; Cartier, N. Glucosamine oligomers: 1. Preparation and characterization. *Int. J. Biol. Macromol.* **1989**, *11*, 297–302. [CrossRef]
164. Einbu, A.; Varum, K.M. Depolymerization and de-*N*-acetylation of chitin oligomers in hydrochloric acid. *Biomacromolecules* **2007**, *8*, 309–314. [CrossRef] [PubMed]
165. Kuyama, H.; Nakahara, Y.; Nukada, T.; Ito, Y.; Nakahara, Y.; Ogawa, T. Stereocontrolled synthesis of chitosan dodecamer. *Carbohydr. Res.* **1993**, *243*, C1–C7. [CrossRef]
166. Barroca-Aubry, N.; Pernet-Poil-Chevrier, A.; Domard, A.; Trombotto, S. Towards a modular synthesis of well-defined chitooligosaccharides: Synthesis of the four chitodisaccharides. *Carbohydr. Res.* **2010**, *345*, 1685–1697. [CrossRef] [PubMed]

167. Weinhold, M.X.; Sauvageau, J.C.M.; Kumirska, J.; Thöming, J. Studies on acetylation patterns of different chitosan preparations. *Carbohydr. Polym.* **2009**, *78*, 678–684. [CrossRef]
168. Abla, M.; Marmuse, L.; Delolme, F.; Vors, J.P.; Ladavière, C.; Trombotto, S. Access to tetra-*N*-acetyl-chitopentaose by chemical *N*-acetylation of glucosamine pentamer. *Carbohydr. Polym.* **2013**, *98*, 770–777. [CrossRef] [PubMed]
169. Trombotto, S.; Ladavière, C.; Delolme, F.; Domard, A. Chemical preparation and structural characterization of a homogeneous series of chitin/chitosan oligomers. *Biomacromolecules* **2008**, *9*, 1731–1738. [CrossRef] [PubMed]
170. Naqvi, S.; Moerschbacher, B.M. The cell factory approach toward biotechnological production of high-value chitosan oligomers and their derivatives: An update. *Crit. Rev. Biotechnol.* **2017**, *37*, 11–25. [CrossRef] [PubMed]
171. Hembach, L.; Cord-Landwehr, S.; Moerschbacher, B.M. Enzymatic production of all fourteen partially acetylated chitosan tetramers using different chitin deacetylases acting in forward or reverse mode. *Sci. Rep.* **2017**, *7*, 17692. [CrossRef] [PubMed]
172. Mergaert, P.; D'Haeze, W.; Geelen, D.; Promé, D.; Van Montagu, M.; Geremia, R.; Promé, J.C.; Holsters, M. Biosynthesis of *Azorhizobium caulinodans* Nod factors: Study of the activity of the nodABCS proteins by expression of the genes in *Escherichia coli*. *J. Biol. Chem.* **1995**, *270*, 29217–29223. [CrossRef] [PubMed]
173. Poinsot, V.; Crook, M.B.; Erdn, S.; Maillet, F.; Bascaules, A.; Ané, J.M. New insights into Nod factor biosynthesis: Analyses of chitooligomers and lipo-chitooligomers of *Rhizobium* sp. IRBG74 mutants. *Carbohydr. Res.* **2016**, *434*, 83–93. [CrossRef] [PubMed]
174. Samain, E.; Drouillard, S.; Heyraud, A.; Driguez, H.; Geremia, R.A. Gram-scale synthesis of recombiant chitooligosaccharides in *Escherichia coli*. *Carbohydr. Res.* **1997**, *302*, 35–42. [CrossRef]
175. Samain, E.; Chazalet, V.; Geremia, R.A. Production of O-acetylated and sulfated chitooligosaccharides by recombinant *Escherichia coli* strains harboring different combinations of nod genes. *J. Biotechnol.* **1999**, *72*, 33–47. [CrossRef]
176. Cottaz, S.; Samain, E. Genetic engineering of *Escherichia coli* for the production of N^{I}, N^{II}-diacetylchitobiose (chitinbiose) and its utilization as a primer for the synthesis of complex carbohydrates. *Metab. Eng.* **2005**, *7*, 311–317. [CrossRef] [PubMed]
177. Bettler, E.; Samain, E.; Chazalet, V.; Bosso, C.; Heyraud, A.; Joziasse, D.H.; Wakarchuk, W.W.; Imberty, A.; Geremia, R.A. The living factory: In Vivo production of *N*-acetyllactosamine containing carbohydrates in *E. coli*. *Glycoconj. J.* **1999**, *16*, 205–212. [CrossRef] [PubMed]
178. Southwick, A.M.; Wang, L.X.; Long, S.R.; Lee, Y.C. Activity of *Sinorhizobium meliloti* NodAB and nodH enzymes on thiochitooligosaccharides. *J. Bacteriol.* **2002**, *184*, 4039–4043. [CrossRef] [PubMed]

9.7. Publication of *Bs*PdaC structure and characterization

J. Biol. Chem, 294(50), 19066. 2019

Structure-function relationships underlying the dual N-acetylmuramic and N-acetylglucosamine specificities of the bacterial peptidoglycan deacetylase PdaC.

Laia Grifoll-Romero[‡], María Angela Sainz-Polo[§], David Albesa-Jové[§], Marcelo E. Guerin[§], Xevi Biarnés[‡], and Antoni Planas[‡].

[‡]Laboratory of Biochemistry, Institut Químic de Sarrià, Universitat Ramon Llull, 08017 Barcelona, Spain.
[§]Structural Biology Unit, Center for Cooperative Research in Biosciences (CIC bioGUNE), Bizkaia Technology Park, Ed. 801A, 48160 Derio, Spain, and the Basque Foundation for Science (IKERBASQUE), 48011 Bilbao, Spain.

1. Introduction
2. Results
3. Discussion
4. Conclusion
5. Materials and methods

Structure-function relationships underlying the dual *N*-acetylmuramic and *N*-acetylglucosamine specificities of the bacterial peptidoglycan deacetylase PdaC

Received for publication, May 24, 2019, and in revised form, November 1, 2019 Published, Papers in Press, November 5, 2019, DOI 10.1074/jbc.RA119.009510

Laia Grifoll-Romero[‡1], **María Angela Sainz-Polo**[§2], **David Albesa-Jové**[§¶], **Marcelo E. Guerin**[§¶], **Xevi Biarnés**[‡], **and Antoni Planas**[‡3]

From the [‡]*Laboratory of Biochemistry, Institut Químic de Sarrià, University Ramon Llull, 08017 Barcelona, Spain, the* [§]*Structural Biology Unit, Center for Cooperative Research in Biosciences (CIC bioGUNE), Bizkaia Technology Park, Ed. 801A, 48160 Derio, Spain, and the* [¶]*Basque Foundation for Science (IKERBASQUE), 48011 Bilbao, Spain*

Edited by Gerald W. Hart

Bacillus subtilis **PdaC (*Bs*PdaC) is a membrane-bound, multidomain peptidoglycan *N*-deacetylase acting on *N*-acetylmuramic acid (MurNAc) residues and conferring lysozyme resistance to modified cell wall peptidoglycans. *Bs*PdaC contains a C-terminal family 4 carbohydrate esterase (CE4) catalytic domain, but unlike other MurNAc deacetylases, *Bs*PdaC also has GlcNAc deacetylase activity on chitooligosaccharides (COSs), characteristic of chitin deacetylases. To uncover the molecular basis of this dual activity, here we determined the X-ray structure of the *Bs*PdaC CE4 domain at 1.54 Å resolution and analyzed its mode of action on COS substrates. We found that the minimal substrate is $GlcNAc_3$ and that activity increases with the degree of glycan polymerization. COS deacetylation kinetics revealed that *Bs*PdaC operates by a multiple-chain mechanism starting at the internal GlcNAc units and leading to deacetylation of all but the reducing-end GlcNAc residues. Interestingly, *Bs*PdaC shares higher sequence similarity with the peptidoglycan GlcNAc deacetylase *Sp*PgdaA than with other MurNAc deacetylases. Therefore, we used ligand-docking simulations to analyze the dual GlcNAc- and MurNAc-binding specificities of *Bs*PdaC and compared them with those of *Sp*PgdA and *Bs*PdaA, representing peptidoglycan deacetylases highly specific for GlcNAc or MurNAc residues, respectively. *Bs*PdaC retains the conserved Asp-His-His metal-binding triad characteristic of CE4 enzymes acting on GlcNAc residues, differing from MurNAc deacetylases that lack the metal-coordinating Asp residue. *Bs*PdaC contains short loops similar to those in *Sp*PgdA, resulting in an open binding cleft that can accommodate polymeric substrates. We propose that PdaC is the first member of a new subclass of peptidoglycan MurNAc deacetylases.**

The bacterial cell wall peptidoglycan (PGN)[4] is an elaborate polymeric mesh composed of a glycan chain of alternating β1,4-linked *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) units cross-linked via peptidyl bridges attached to the 3-*O*-lactoyl group of MurNAc residues. Pathogenic bacteria utilize acetylation (6-*O*-acetylation of MurNAc) and deacetylation (2-*N*-deacetylation of GlcNAc and/or MurNAc residues) of their cell wall PGN to evade detection by the innate immune system. *N*-Deacetylation was first identified in 1971 in lysozyme-resistant *Bacillus cereus* strains, which contain high proportions of nonacetylated glucosamine (GlcN) residues in the cell wall PGN (1). Later, nonacetylated muramic acid residues (MurN) were identified in the PGN of *Bacillus anthracis* (2). The presence of deacetylated sugars in PGN strongly reduces the activity of the muramidase lysozyme, which hydrolyzes the MurNAcβ(1→4)GlcNAc glycosidic bonds of the glycan strands. Lysozyme is an important factor of the innate immune system in humans, it is present in many tissues and body fluids, and it is secreted in large amounts by cells of the immune system at the site of infection (3, 4). The innate immune system senses intact PGN and PGN fragments originated by lysozyme and other lytic enzymes using diverse recognition mechanisms, which include peptidoglycan recognition proteins, Toll-like receptors, and Nod-like receptors (3). Efficient PGN degradation promotes increased duration and intensity of inflammatory signals and instructs immune responses to infection. Bacteria that are able to modify the structure of PGN to make it more resistant to lysozyme both evade the antibacterial activity of lysozyme and delay or suppress pro-inflammatory immune responses (3).

This work was supported by Ministerio de Economía, Industria y Competitividad, Gobierno de España (MINECO), Spain, Grants BFU2016–77427-C2-1-R (to A. P.), BFU2016-77427-C2-2-R (to M. E. G.), and BFU2017-92223-EXP (to M. E. G.) and Severo Ochoa Excellence Accreditation SEV-2016-0644 (to M. E. G.) and Government of Catalonia Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) Grant 2017SGR-727 (to A. P.) from the Generalitat de Catalunya. The authors declare that they have no conflicts of interest with the contents of this article.

This article contains Tables S1–S5 and Figs. S1–S13.

The atomic coordinates and structure factors (codes 6H8L and 6H8N) have been deposited in the Protein Data Bank (http://wwpdb.org/).

[1] Recipient of a predoctoral contract from Institut Químic de Sarrià.

[2] Recipient of a postdoctoral contract from MINECO under the "Juan de la Cierva Postdoctoral program" (FJCI-2015-25725).

[3] To whom correspondence should be addressed: Institut Químic de Sarrià, Via Augusta 390, 08017 Barcelona, Spain. Tel.: 34-932672000; E-mail: antoni.planas@iqs.edu.

[4] The abbreviations used are: PGN, peptidoglycan; MurNAc, *N*-acetylmuramic acid; GlcN, nonacetylated glucosamine; 3D, three-dimensional; aa, amino acids; RMSD, root mean square deviation; MDP, *N*-acetylmuramyl dipeptide; FL, full-length; CD, catalytic domain; AcOMU, 4-methylumbelliferyl acetate; MU, 4-methylumbelliferone; DMF, *N*,*N*-dimethylformamide; PDB, Protein Data Bank; COS, chitooligosaccharide; CE4, carbohydrate esterase family 4.

Deacetylation of MurNAc residues of PGN is also involved in endospore formation and germination. Spores of *Bacillus* species have a thick peptidoglycan (spore cortex) that contains high abundance of spore-specific muramic acid δ-lactam. Two enzymes are required for its synthesis, the amidase CwlD (muramoyl-L-alanine amidase) that removes the peptide side chain from the 3-*O*-lactoyl group of MurNAc residues and a MurNAc deacetylase to generate the free amino group for lactam formation (4, 5). *Bacillus subtilis* has six identified or predicted polysaccharide deacetylases: PdaA and PdaB are MurNAc deacetylases shown to be involved in sporulation (6, 7). PdaC was reported to be also a MurNAc deacetylase but not associated with sporulation and germination (8). The other gene products, YlxY, YxkH, and YheN, have not been characterized (it is not known whether they deacetylate GlcNAc or MurNAc residues), but they are not associated with sporulation (7, 8).

Peptidoglycan GlcNAc deacetylases (EC 3.5.1.104) and MurNAc deacetylases (EC 3.5.1.-) are classified in carbohydrate esterase family 4 (CE4) in the CAZy (Carbohydrate Active Enzyme) database (www.cazy.org)[5] (9) together with chitin deacetylases (EC 3.5.1.41), poly-β1,6-GlcNAc deacetylases (EC 3.5.1.-), and some acetylxylan esterases (EC 3.1.1.72). CE4 enzymes share a conserved region known as the NodB homologous domain due to its similarity to the NodB oligosaccharide deacetylase, one of the first deacetylases of this family to be characterized (10). They operate by metal-assisted general acid/base catalysis, as first proposed for the *Streptococcus pneumoniae* peptidoglycan GlcNAc deacetylase *Sp*PgdA when its X-ray structure was solved (11). The family is characterized by the conservation of five sequence motifs (named MT1 to -5), which include the Asp-His-His metal-binding triad (second Asp (D) of the MT1 sequence FTDDG and the two histidines (H) of the MT2 sequence H(S/T)*XX*H), the general base Asp (first D in MT1), and the general acid His in MT5 (I(V/I)L*X*HD) (11, 12).

A number of peptidoglycan GlcNAc deacetylases have been characterized, and their 3D structures determined by X-ray crystallography (Table S1) (12). They are specific for GlcNAc residues in the peptidoglycan chain, but they have also been shown to deacetylate COS, homooligomers of GlcNAc residues, but with lower efficiency. In contrast, few peptidoglycan MurNAc deaceylases have been biochemically characterized, although this activity has been annotated to a number of putative polysaccharide deacetylases based on genetic knockout studies (13). The *B. subtilis* PdaA deacetylates MurNAc residues of peptidoglycan devoid of the peptide linked to the muramic acid 3-*O*-lactoyl group of MurNAc residues, which is consistent with its function during sporulation to form muramic acid δ-lactam residues in the spore cortex peptidoglycan (6, 14). It is specific for MurNAc residues, and it is not active on COS. The crystal structure of *Bs*PdaA revealed a modified $(\beta/\alpha)_8$ fold (15) characteristic of CE4 enzymes (some members lacking one α-helix and/or one β-strand of the β/α barrel) (11, 12, 16). A second structure for a putative MurNAc deacetylase, that of the *Ba*CE4 from *B. anthracis*, has been determined (17). Although not biochemically characterized, it is likely to be a MurNAc deacetylase due to its close similarity to *Bs*PdaA. Both enzymes lack the otherwise conserved Asp residue of the metal-binding catalytic triad, having an Asn residue that points away from the active metal ion into the core of the protein (17).

More recently, Kobayashi *et al.* (8) identified a novel peptidoglycan MurNAc deacetylase, *Bs*PdaC, whose expression is regulated by the essential YycFG two-component system that controls cell wall metabolism in *B. subtilis* (18, 19). Interestingly, *Bs*PdaC deacetylates MurNAc residues of intact peptidoglycan, or at least with the L-Ala-D-Glu dipeptide attached to the muramic 3-*O*-lactoyl group, being inactive on the glycan backbone devoid of peptidyl substitutions. More striking is that, as opposed to *Bs*PdaA, *Bs*PdaC also deacetylates COS, an activity that was thought to be restricted to GlcNAc deacetylases.

To understand the molecular bases of such dual activity, we report the X-ray 3D structure of the CE4 catalytic domain of *Bs*PdaC and the mode of action on COS substrates. The dual GlcNAc- and MurNAc-binding specificity at the catalytic site is analyzed by using a combination of X-ray crystallography, enzyme kinetics, ligand-docking simulations, and protein biochemistry/biophysics. We propose that *Bs*PdaC represents the first member of a novel subclass of MurNAc deacetylases.

Table 1
Deacetylation activity of *Bs*PdaC-CD on PGN and chitooligosaccharides
Conditions were as follows: 2 mM $(GlcNAc)_n$ substrates or 1 mg/ml PGN substrate, 0.15–1.3 μM enzyme, 50 mM phosphate buffer, 300 mM NaCl, pH 7.0, 37 °C. ND, not determined.

Substrate	Specific activity (v_0/[*E*])	
	Fluorescamine	HPLC-MS
	min^{-1}	
$(GlcNAc)_2$	No activity	No activity
$(GlcNAc)_3$	ND	3.99 ± 0.27
$(GlcNAc)_4$	18.9 ± 2.4	18.18 ± 0.22
$(GlcNAc)_5$	38.0 ± 2.8	36.80 ± 3.46
*Bs*PGN[a]	63.1 ± 0.6[b]	
	81.8 ± 1.7[c]	

[a] Peptidoglycan from *B. subtilis*, specific activity determined by monitoring acetate release.
[b] Specific activity for *Bs*PdaC-CD (catalytic domain).
[c] Specific activity for *Bs*PdaC-FL (full-length).

Results

Enzyme kinetics and deacetylation pattern

*Bs*PdaC is a multidomain protein composed of a transmembrane region (aa 7–26), two domains with unknown function (aa 54–138 and 156–235), and a CE4 catalytic domain (Fig. S1). The full-length *Bs*PdaC (residues 27–467, lacking the transmembrane region, here referred to as *Bs*PdaC-FL) and the isolated catalytic domain (*Bs*PdaC-CD, aa 270–467) were recombinantly expressed and purified to apparent homogeneity (Fig. S2).

*Bs*PdaC-CD and *Bs*PdaC-FL displayed similar specific activity with PGN (63.1 and 81.8 min^{-1}, respectively) (Table 1) as well as similar thermal stability (Fig. S3). The isolated catalytic domain is expected to have the same MurNAc specificity as the full-length enzyme reported previously (8). To assess it, PGN was incubated with *Bs*PdaC-FL for an extended period, and the

[5] Please note that the JBC is not responsible for the long-term archiving and maintenance of this site or any other third party hosted site.

released acetate was quantified. After dialysis to remove the free acetate, the product was incubated with *Bs*PdaC-CD, and it was observed that essentially no further deacetylation took place upon a long incubation time. A parallel experiment where the order of the enzymes was inverted (first incubation with *Bc*PdaC-CD, then with *Bc*PdaC-FL) gave the same result (Table S2). This indicates that both enzymes deacetylate the same residues (MurNAc) of PGN. Moreover, when the final deacetylated PGN product from both experiments was treated with *Sp*PgdA (a deacetylase specific for GlcNAc residues), an additional release of acetate was observed, with a specific activity of 70–80% that of *Sp*PgdA on intact PGN (Table S2). It is then concluded that the isolated catalytic domain has the same specific activity and retains the same MurNAc specificity as the full-length enzyme. The isolated catalytic domain was used for further studies.

Enzyme activity on COS substrates was determined by two methods: by quantification of the free amino groups generated from deacetylation of $(GlcNAc)_n$ substrates using the fluorescamine reagent (11), and by HPLC-MS analysis of the deacetylation reaction in which the sequence of GlcNAc deacetylation events could be monitored (16). The pH and temperature profile of the enzyme activity with $(GlcNAc)_5$ substrate showed a pH optimum of 8.5 at 37 °C and an optimum temperature of 55 °C at pH 7 (Fig. S3).

*Bs*PdaC is a metal-dependent deacetylase as most of the CE4 enzymes (12). Treatment with 20 mM EDTA only reduced the activity by 10% relative to untreated and freshly purified enzyme, indicating that the metal cation is strongly bound to the active site. Conditions to generate the apo-form of the enzyme for reconstitution experiments included increasing concentrations of urea in the presence of EDTA. After treatment with 200 mM EDTA in 2 M urea, the recovered protein after dialysis still retained 65% activity, and full activity could be restored upon the addition of Zn^{2+} cation. Treatment with 7 M urea and 200 mM EDTA was required to dissociate the metal cation. Protein refolding was quantitative, resulting in an inactive protein that recovered 100% activity upon the addition of Zn^{2+} cation (Fig. S3*C*).

Enzyme activity was evaluated with COS substrates $(GlcNAc)_n$, $n = 2–5$. The enzyme was inactive with the disaccharide substrate, the trisaccharide was the minimal COS substrate, and the activity increased with the degree of polymerization (Table 1). Although Kobayashi *et al.* (8) reported single monodeacetylation activity of *Bs*PdaC on $(GlcNAc)_4$, we have found a more complex mode of action. HPLC-MS time course monitoring of the deacetylation reaction revealed that the enzyme sequentially deacetylates different GlcNAc units, leading to final products in which all but one GlcNAc unit were deacetylated (Fig. 1*A*). For the chitotetraose substrate, mono-, di-, and trideacetylated products were sequentially formed (A3D1, A2D2, and A1D1, where A represents GlcNAc and D is GlcN). Likewise, with the chitopentaose substrate, A4D1, A3D2, A2D3, and A1D4 were generated along the reaction. Fig. 1 (*B* and *C*) summarizes the evolution of the different products. Whereas the final A1D3 from tetraacetylchitotetraose was detected at rather short reaction times, the final A1D4 product from pentaacetylchitopentaose was only detected after a long reaction time. The structure of the different products with regard to their pattern of acetylation was determined by the MALDI-TOF-MS/MS method according to Cord-Landwerh *et al.* (20), consisting in ^{18}O labeling of the reducing end and re-*N*-acetylation with $[^2H_6]$acetic anhydride, which allows the identification and quantification of the different products in the reaction mixtures (Table S3). *Bs*PdaC follows a multiple-chain mechanism (or distributive mechanism) as presented in Fig. 2*A*. Interestingly, with the tetrasaccharide substrate, the first deacetylation occurs at the two central GlcNAc units with equal efficiencies, whereas with the pentasaccharide substrate, any of the three internal GlcNAc units are also deacetylated but with preference for the second residue from the reducing end (51% as compared with 14 and 35% for the other two internal residues). Deacetylation proceeds in time, leading to final products in which all but the reducing-end GlcNAc units are deacetylated (DDDA and DDDDA, respectively).

Initial rates corresponding to the first deacetylation event leading to monodeacetylated products were determined by HPLC-MS. Likewise, the same values were obtained by the fluorescamine method at short reaction times, corresponding to monodeacetylation, as verified by HPLC-MS. The specific activity on $(GlcNAc)_5$ was twice than that on $(GlcNAc)_4$ and 10-fold higher than the activity with the shortest substrate $(GlcNAc)_3$ (Table 1). Michaelis–Menten kinetic parameters were determined for the tetrasaccharide substrate (Fig. S4), with apparent k_{cat} and K_m values of 0.93 ± 0.03 s^{-1} and 6.25 ± 0.47 mM, respectively, and apparent k_{cat}/K_m of 149 M^{-1} s^{-1}, which are of the same order as those reported for a peptidoglycan substrate (8).

The increase in activity from a chitotriose to a chitopentaose substrate suggests that the binding site cleft has at least six subsites, from −3 to +2, to account for the different activities when considering the first deacetylation event (Fig. 2*B*). For the chitotetraose substrate, three subsites (−1, 0, +1) explain the observed monodeacetylation pattern, but additional subsites −2 and +2 may exist, with similar affinities to agree with the fact that both binding modes are alike (50% deacetylation at each central GlcNAc unit). With the chitopentaose substrate, if subsites −2 and +2 have similar affinities, a new subsite −3 might be required to account for the preferred binding mode 1 (BM1′, Fig. 2*B*) (51%) over BM2′ (35%), because BM1′ does not occupy subsite +2. A potential subsite +3 is not required, because BM3′ is the least preferred binding mode (14%) that may be accounted for by the loss of interactions with subsite −2. Binding to subsite +1 is required in any case, because it is always occupied not only in the different binding modes for the first monodeacetylation reactions, but also in the subsequent deacetylations leading to the final products in which the reducing-end GlcNAc unit is not deacetylated.

The crystal structure of BsPdaC-CD

The 3D structure of *Bs*PdaC-CD was obtained by X-ray crystallography at 1.54 Å resolution (Table S4 and Fig. 3). *Bs*PdaC-CD crystallized as a homodimer. The protein monomer adopts a $(\beta/\alpha)_7$ barrel topology. The central core comprises seven parallel β-strands that form a distorted β-barrel

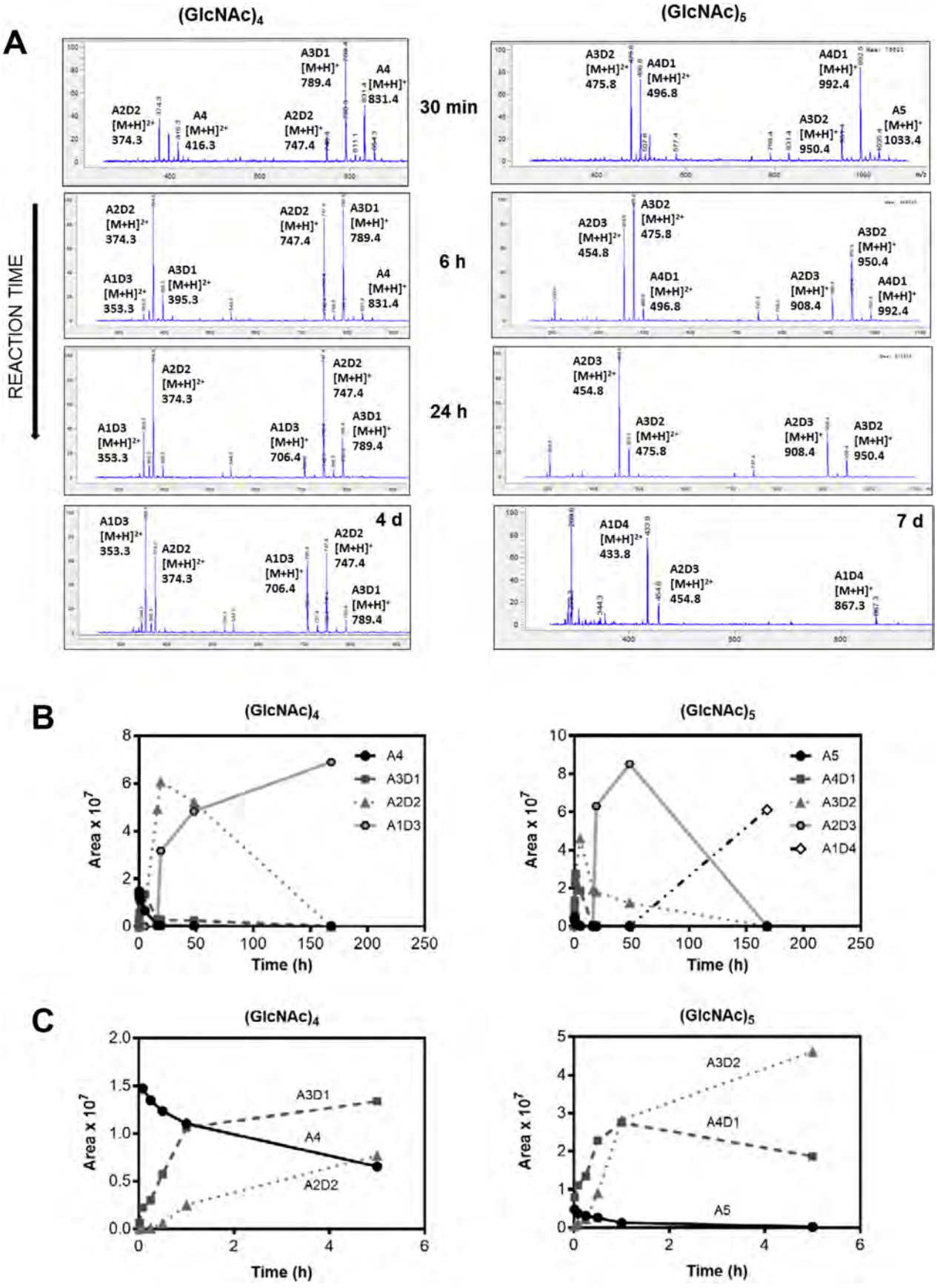


Figure 1. MS time monitoring of the *Bs*PdaC-CD deacetylase reactions with substrates $(GlcNAc)_4$ and $(GlcNAc)_5$. *A*, MS spectra at increasing reaction time. *B*, evolution of peak areas for $[M + H]^+$ ions of the deacetylated products. *C*, magnification of *B* at short reaction time. Relative areas from the MS spectra are given.

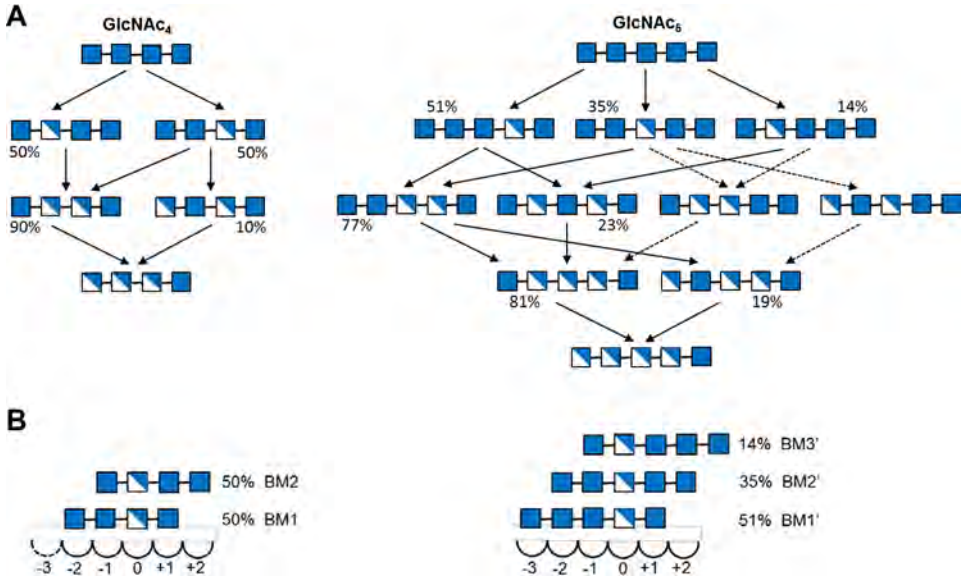


Figure 2. BsPdaC pattern of deacetylation. *A*, deacetylation sequence of COS $(GlcNAc)_4$ and $(GlcNAc)_5$. Symbolic representation of glycans is according to Ref. 47, with the reducing end on the *right*. *Dashed arrow*, slower deacetylations at longer reaction times (see Table S2). *B*, binding of COS ligands to *Bs*PdaC. Shown are binding modes for the first deacetylation event for substrates $(GlcNAc)_4$ and $(GlcNAc)_5$. Negative subsites are on the nonreducing end, and positive subsites are on the reducing end. Subsite 0 is the catalytic site.

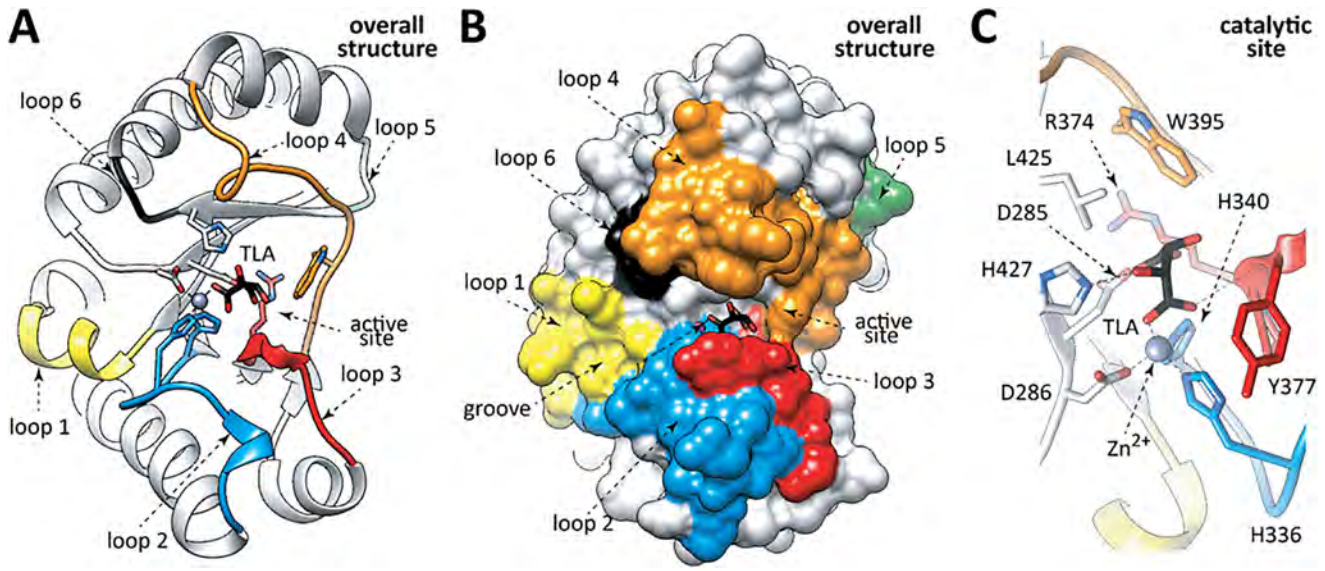


Figure 3. The crystal structure of *Bs*PdaC-CD. *A*, *cartoon representation* of the overall structure of *Bs*PdaC-CD showing the loops (*L1–L6*) surrounding the active site of the CE4 domain. *B*, *surface representation* of *Bs*PdaC-CD showing the COS-binding site. *C*, active site of *Bs*PdaC-CD showing the catalytic residues (Asp^{285}, general base; His^{427}, general acid) and metal-binding triad (Asp^{286}-His^{336}-His^{340}), a Zn^{2+} ion, and carboxylate oxygens of a tartaric acid molecule (*TLA*) from the crystallization buffer. Loops according to the subsite capping model are as in Fig. S11 and flanking secondary structures are as follows: loop 1 (residues 313–320, $\beta2$-$\alpha2$), loop 2 (341–348, $\beta3$-$\alpha3$), loop 3 (376–382, $\beta4$-$\alpha4$), loop 4 (401–410, $\beta5$-$\alpha5$), loop 5 (419–421, $\alpha5$-$\beta6$), and loop 6 (428–437, $\beta6$-$\alpha6$).

surrounded by α-helices. The putative ligand-binding site is located in the central region of the β-barrel and is flanked by a series of connecting loops.

The structure contains a Zn^{2+} ion coordinated with Asp^{286}, His^{336}, and His^{340}, which form the conserved metal-binding triad of CE4 enzymes (with few exceptions; see below), and two oxygens of a tartaric acid molecule from the crystallization buffer (Fig. 3*C*). The conserved catalytic residues, general base Asp^{285} and general acid His^{427}, have the right orientation in the active site for catalysis as in other CE4 crystal structures. CE4 enzymes differ in a series of surface loops that shape the binding site cleft and are structural elements involved in substrate specificity, as proposed in the "subsite-capping model" (16, 21). Loops 1–6 (loop 1, residues 313–320; loop 2, 341–348; loop 3, 376–382; loop 4, 401–410; loop 5, 419–421; loop 6, 428–437) are *colored* in the 3D structure (Fig. 3, *A* and *B*).

The C terminus contains an α-helix that protrudes from the core of each monomer and interacts with the adjacent monomer partially covering the active site (Fig. S6). This helix (from Gly^{448} to the end of the solved structure) is composed of the final 20 amino acid residues of the protein sequence, plus 5 aa coming from the restriction site sequence used for subcloning, one of which (Arg^{471}) inserts the side chain into the active site of the adjacent monomer, and 2 aa from the Strep-tag (the last

6 C-terminal residues of the Strep-tag are disordered and not seen in the electron density map). The C-terminal α-helix of one monomer runs parallel to the substrate binding cleft of the other monomer. The side chain of Arg^{471} protrudes perpendicularly from the helix axis, and the guanidyl group interacts with His^{427} (general acid), Asp^{286} (metal coordination), and a tartaric acid molecule (from the crystallization mother liquor) in the active site of the other monomer. The orientation of this terminal helix is probably a consequence of the crystal packing because the enzyme in solution is fully active, and this dimerization contact in the X-ray structure would block access of the substrate to the active site.

Despite much effort, attempts to co-crystallize *Bs*PdaC-CD-D285S (inactive mutant at the general base residue) with the tetrasaccharide ligand $(GlcNAc)_4$ were unsuccessful, probably due to the presence of the extended C-terminal α-helix. The structure of the mutant superimposes with that of the WT enzyme with an RMSD of 0.27 Å. A new construct moving the Strep-tag to the N terminus and leaving an intact C terminus with the original native sequence was prepared (*Bs*Pdac-CD-NtStrep and purified following the same protocol. The WT enzyme had the same specific activity on COS substrates as the initial *Bs*PdaC-CD enzyme. Upon mutation of the general base residue (Asp^{285}), this new construct again failed to co-crystalize with the $(GlcNAc)_4$ substrate. It is worth noting that ligand binding was assessed by monitoring the thermal unfolding by CD in the absence and presence of COS substrates (Fig. S7), observing a T_m increase of 3.3 and 3.7 °C upon binding of $(GlcNAc)_4$ and $(GlcNAc)_5$, respectively.

Enzyme·ligand complexes by ligand-docking simulations

Computational docking of $(GlcNAc)_4$ on the X-ray structure of *Bs*PdaC-CD showed two preferred binding modes with low energy (Fig. 4) in which either of the two internal GlcNAc residues are located in the catalytic site (subsite 0) and would render the AADA and ADAA products. Both binding modes are of similar energy, indicating 50% population each, which is in agreement with the observed enzyme specificity for the first deacetylation event on this substrate as shown in Fig. 2. The substrate expands subsites −2 to +2, and all GlcNAc units interact with protein residues through hydrogen bonding and stacking interactions (Fig. 4*C*). A subsite −3 is not seen in these models with a tetrasaccharide ligand, but additional surface interactions may occur with a longer and more reactive pentasaccharide substrate.

The enzyme·ligand models (Fig. 4*C*) predict some key interactions that were probed by site-directed mutagenesis. Tyr^{430}, located in loop 6, interacts with the substrate through a stacking interaction with the GlcNAc unit at subsite −2 and hydrogen-bonds with the GlcNAc unit at subsite −1. Mutation to Ala (Y430A) reduced the specific activity to 6% of the WT activity. Likewise, Trp^{402} is located in loop 4 and establishes a tight hydrophobic (stacking) interaction with the GlcNAc unit in subsite 0, on the opposite site (relative to the sugar plane) to the metal binding triad (Aps^{286}-His^{336}-His^{340}). Mutation to Ala (W402A) has a drastic effect, with a residual activity <0.1% than that of the WT enzyme.

To obtain further insights into the substrate specificity of *Bs*PdaC, a series of fragment-based ligand-docking simulations were conducted with monosaccharide (MurNAc and GlcNAc) and disaccharide (GlcNAcβ(1→4)MurNAc and MurNAcβ(1→4)GlcNAc) probes. Similar docking experiments on the 3D structure of *Bs*PdaA and *Sp*PgdA, as representative peptidoglycan deacetylases specific for MurNAc and GlcNAc units, respectively, were also analyzed for comparison (Figs. S8–S10). These docking experiments illuminate significant differences between GlcNAc- and MurNAc-specific enzymes and the here-reported *Bs*PdaC with dual specificity, depending on the substrate.

The monosaccharide GlcNAc probe is able to fit into the active site of *Bs*PdaC with a proper orientation for catalysis (Fig. S8*A*); the C=O of the acetamido group is coordinated with the Zn^{2+} cation (distance 2.1 Å), and the NH is at hydrogen-bonding distance to the general acid His-427 (3.4 Å). No productive binding was observed for the MurNAc probe (Fig. S8*B*). However, with the disaccharide GlcNAcβ(1→4)MurNAc probe, productive binding placed the MurNAc unit in subsite 0 and the GlcNAc unit in subsite −1 (Fig. S8*C*). In addition, as shown before, docking of the $(GlcNAc)_4$ substrate also gave productive complexes for catalysis that were in agreement with the experimentally observed specificity of the enzyme. The active site (subsite 0) is thus able to accommodate both GlcNAc and MurNAc residues in a productive binding mode for catalysis. Because the PGN glycan chain is composed of alternating GlcNAc and MurNAc units, it seems that subsites −1 and/or +1 are unable to accommodate a MurNAc residue, even in the absence of peptidyl substitutions, because PGN digested with L-alanine amidase is not deacetylated (8).

Docking of $(GlcNAc)_4$ showed that the 3-OH of the GlcNAc unit in subsite 0 does not establish any interaction with the protein other than coordination with the Zn^{2+} cation (Fig. 4*C*), thus leaving room for the 3-lactoyl substitution (as also seen in the ligand-docking simulations with the PGN disaccharide (Fig. S8*C*)), but the 3-OH of the GlcNAc unit in subsite −1 hydrogen-bonds with Tyr^{430} in both binding modes, leaving no room for 3-OH substitutions. Subsite +1, on the other side, shows a weaker binding of the GlcNAc unit, where only the 6-OH hydrogen-bonds with protein residues (Asp^{398} in both binding modes and Trp^{402} in binding mode 2). Therefore, the simulations suggest that subsite −1 is specific for GlcNAc residues and dictates the specificity shown by the enzyme. *Bs*PdaC requires at least a dipeptidyl substitution at the 3-*O*-lactoyl group of MurNAc residues, not included in the docking simulations with PGN disaccharide, indicating that additional interactions with the peptide substituent are important for activity on PGN substrates.

Similar fragment-based ligand-docking simulations were performed on *Bs*PdaA (canonical PGN MurNAc deacetylase) with oligosaccharide probes (Fig. S9). The GlcNAc monosaccharide probe did not get into the active site, but mainly went to subsite +2, whereas the MurNAc probe gave an ensemble of docked structures compatible with productive binding in subsite 0. With the disaccharide ligands, only the MurNAc unit was found in subsite 0 in a productive orientation, consistent with the MurNAc deacetylase specificity of this enzyme that is inac-

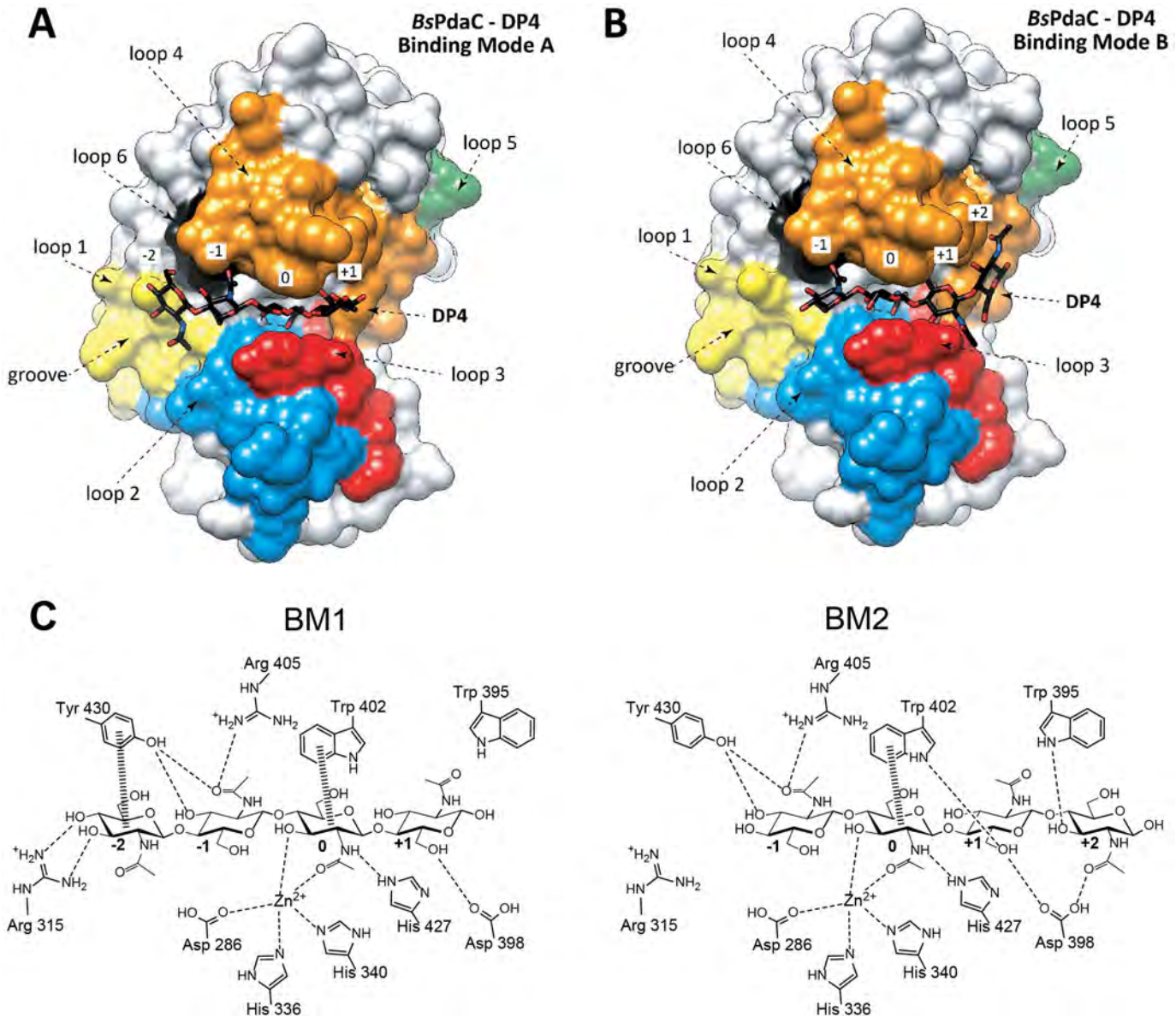


Figure 4. Docking simulation of *Bs*PdaC-CD·(GlcNAc)$_4$ complexes. *A*, binding mode 1 (AAAA). *B*, binding mode 2 (AAAA), where A is the GlcNAc residue in subsite 0 (catalytic site). Shown are structures of minimum energy from the ensemble of docked structures in each minimum. The C terminal α-helix is not represented. Loops are *colored* as in Fig. 3. *C*, protein-ligand interaction from docking simulations. *BM*, binding mode.

tive on COS substrates. The opposite behavior was observed for *Sp*PgdA, a specific GlcNAc deacetylase (Fig. S10). Only the GlcNAc monosaccharide probe was able to productively bind in subsite 0, and the disaccharide ligands always placed the GlcNAc unit into subsite 0, with disaccharide binding to subsites 0 and +1 for GlcNAcβ(1→4)MurNAc and to subsites −1 and 0 for MurNAcβ(1→4)GlcNAc, again consistent with the deacetylation specificity on GlcNAc residues of peptidoglycan and COS substrates.

BsPdaC sequence and phylogeny of family CE4 enzymes

The sequence of the catalytic domain was added to the multiple sequence alignment of CE4 family members (16), which includes chitin deacetylases, peptidoglycan deacetylases, some acetylxylan esterases, and poly-β1,6-GlcNAc deacetylases (Table S1 and Fig. S11), allowing the location of the catalytic residues within the conserved motifs of this family of enzymes (MT1 to -5) (11, 21). Motif 1 (TFDDG) contains the catalytic base (first aspartate, Asp^{285} in *Bs*PdaC) and one residue of the metal coordination triad (second aspartate, Asp^{286} in *Bs*PdaC), motif 2 (H(S/T)*XX*H) contains two histidines (His^{336} and His^{340} in *Bs*PdaC) involved in metal coordination along with the aspartate in MT1, and motif 5 (I(V/I)L*X*HD) contains the catalytic acid (His^{427} in *Bs*PdaC). The sequence alignment also identified the loops that are characteristic of CE4 enzymes and have been related to substrate specificity (16), which are short in *Bs*PdaC as compared with other family members (loops 1–6, Fig. S11).

A phylogenetic analysis based on the multiple-sequence alignment showed that the peptidoglycan deacetylase PgdA from *S. pneumoniae* (*Sp*PgdA) is the closest CE4 enzyme to *Bs*PdaC (Fig. 5). Both proteins share 30% overall sequence iden-

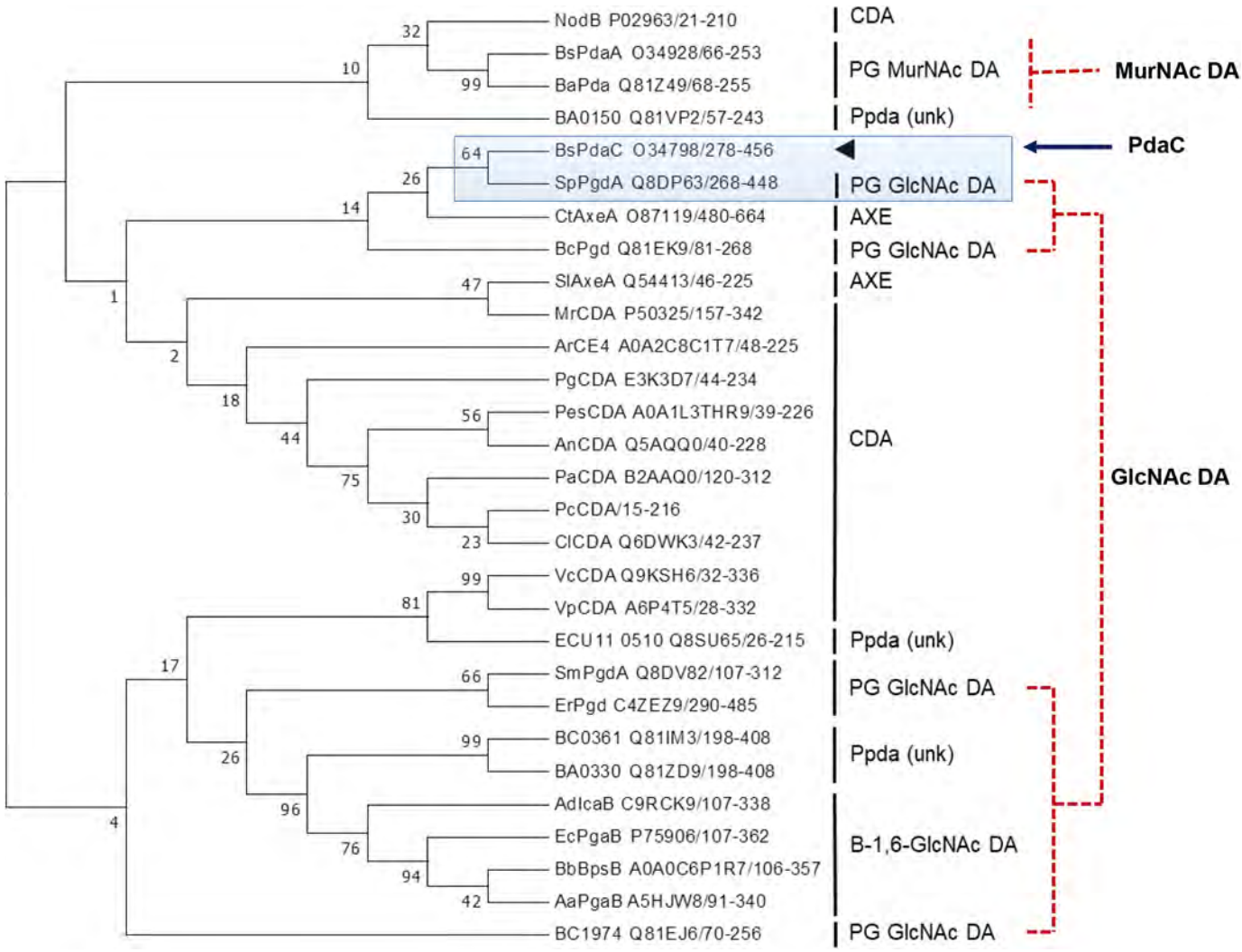


Figure 5. Phylogenetic analysis of CE4 enzymes from the multiple sequence alignment presented in Fig. 1. A bootstrap analysis with 500 replicates was carried out on the trees inferred from the maximum likelihood method.

tity with the same domain organization and 44% sequence identity for the CE4 catalytic domain. Although they are both peptidoglycan deacetylases, *Sp*PgdA acts on GlcNAc residues (11, 22), whereas *Bs*PdaC deacetylates MurNAc residues of peptidoglycan substrates (8). Interestingly, the other two known peptidoglycan MurNAc deacetylases with solved 3D structure, *Bs*PdaA (biochemically characterized) (15) and *Ba*CE4 (annotated) (17), are more distantly clustered in the phylogenetic tree.

Discussion

PdaC is a metalloenzyme that follows a multiple-chain mechanism on COS substrates

*Bs*PdaC is a metalloenzyme as expected for a CE4 enzyme and here proven by the fact that removal of the metal cation leads to an inactive enzyme whose activity is restored by reconstitution with Zn^{2+} cation. This is in contrast with a previous report (8), which suggested that *Bs*PdaC might not be a metalloenzyme because it retained 50% activity when assayed with PGN and 70% when assayed with tetraacetylchitotetraose after treatment with 20 mM EDTA. However, as shown here, these conditions were not sufficient to dissociate the metal cation because protein unfolding and refolding was necessary to dissociate the metal cation that is strongly bound to the active site.

The enzyme is highly active on intact PGN and active on COS with degree of polymerization ≥3. It follows a multiple-attack mechanism starting with deacetylation of the internal GlcNAc residues and proceeding to deacetylate all but the reducing-end unit of the substrate. This behavior is unique among characterized MurNAc deacetylases, which do not deacetylate COS and prefer a modified PGN substrate in which the peptidyl substitutions have been previously removed (14).

PdaC is a new subclass of MurNAc deacetylases with different sequence and structural features

*Bs*PdaC has a different substrate specificity as compared with canonical MurNAc deacetylases, represented by *Bs*PdaA (14, 15). Comparison of the 3D structures unravels significant differences at their active sites.

First, the metal coordination differs between these two types of MurNAc deacetylases. *Bs*PdaC maintains the conserved metal-binding triad (Asp-His-His) found in all characterized

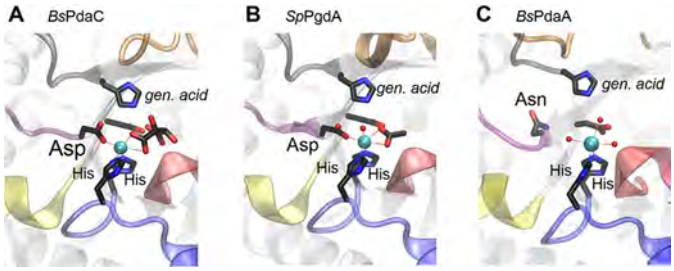


Figure 6. Metal coordination in the crystallographic 3D structures of *Bs*PdaC (*A*), *Sp*Pgda GlcNAc deacetylase (*B*), and *Bs*PdaA MurNAc deacetylase (*C*).

chitin deacetylases and peptidoglycan GlcNAc deacetylases (Fig. S11). In contrast, MurNAc deacetylases (structural information currently limited to *Bs*PdaA and *Ba*CE4(BA0424), Table S1) lack the conserved Asp, which is replaced by an Asn that has a different side-chain orientation pointing away from the metal ion. *Bs*PdaA was initially crystallized as the apoenzyme, but the metal coordination could be solved by soaking experiments with Cd^{2+}, showing that it was only coordinated with two active-site residues, His^{124} and His^{128}, and the remaining ligands were water molecules (15). *Ba*CE4 was solved with Zn^{2+} cation, where only the two conserved His coordinated the metal cation, and coordination was completed with an acetate and cacodylate anions from the crystallization mother liquor (17). A structural comparison of the metal coordination in the X-ray structures of *Bs*PdaC, *Bs*PdaA, and *Sp*Pgda (as representative of a GlcNAc deacetylase) is shown in Fig. 6.

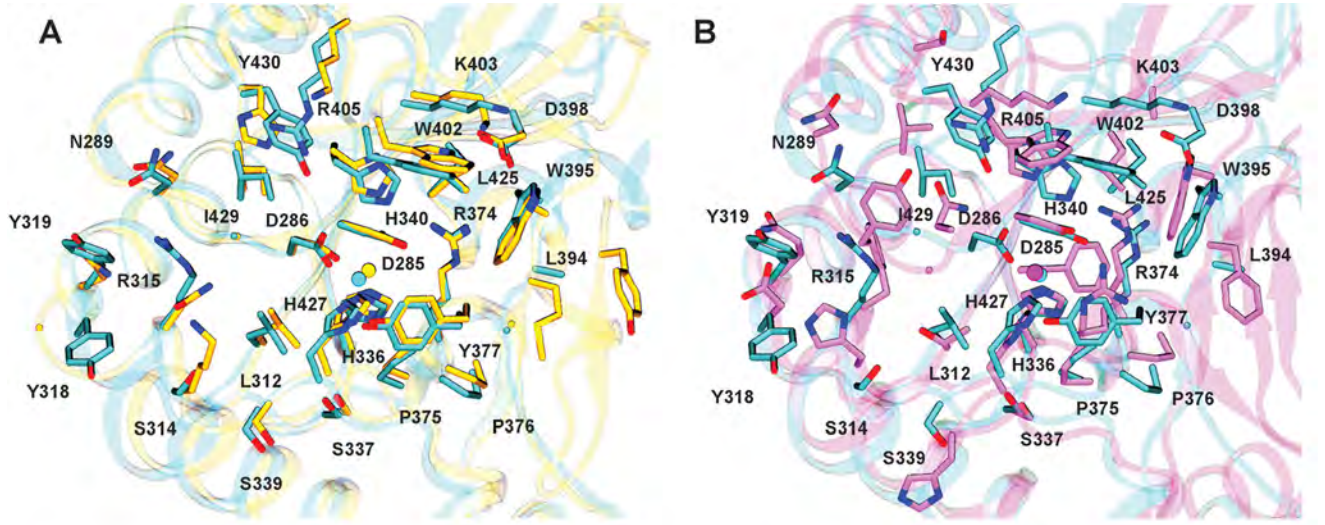
Second, the structural and sequence alignment of *Bs*PdaC with CE4 family members identifies the loops that shape the binding site cleft and that have been proposed to define the substrate specificity of CDAs (16). *Bs*PdaC reveals short loops, similar in shape and length to other PGN deacetylases, that result in an open binding cleft able to accommodate polymeric substrates and exhibit a multiple-chain mode of action. As a consequence of the nonconserved nature and structural arrangement of these loops surrounding the active site, both MurNAc deacetylase subclasses display different surface electrostatic potential at the active site. *Bs*PdaA has a positively charged binding cleft, as opposed to GlcNAc deacetylases, such as *Sp*PgdA, in which it is negatively charged. *Bs*PdaC lies in between, with positive and negative patches in the binding cleft (Fig. S12).

Third, substrate-binding cavities of *Bs*PdaC, *Sp*PgdA, and *Bs*PdaA are compared in Fig. 7. The conservation of *Bs*PdaC with *Sp*PgdA is notably higher than with *Bs*PdaA, with very similar 3D structures (RMSD between catalytic domain CA atoms of 1.15 Å). The structure and sequence composition of subsite 0 is identical in *Bs*PdaC and *Sp*PgdA. This reveals how *Bs*PdaC, similarly to *Sp*PgdA, has the capacity to accommodate GlcNAc residues at the catalytic site for deacetylation (subsite 0), confirmed by ligand-docking simulations and deacetylase activity on COS. However, *Bs*PdaC does not deacetylate GlcNAc residues from PGN. Such a pattern of deacetylation would require the binding of MurNAc units at subsite −1, similarly to *Sp*PgdA. This has not been detected by docking. This different behavior can be explained by the sequence differences at subsite −1 of both structures (Fig. 7*A*): *Bs*PdaC Arg^{405} and Tyr^{430} are replaced by Lys^{395} and His^{420} in *Sp*PgdA. Bulkier side chains of *Bs*PdaC hamper the binding of the 3-lactoyl substitution of a MurNAc residue in subsite −1, in the same location, where it is accommodated in *Sp*PgdA (Fig. S13). Subsite +1 is also identical in *Bs*PdaC and *Sp*Pgda. On the other hand, major structural and sequence differences are observed between *Bs*PdaC and *Bs*PdaA at both sides of the active site (Fig. 7*B*). Remarkably, a negative charge is introduced in negative subsites (Tyr^{318} to Asp^{106}, respectively) and a positive charge is removed (Arg^{315} to Phe^{103}, respectively) next to it. *Bs*PdaC and *Bs*PdaA, although both deacetylate MurNAc residues of PGN, differ in the C3 substitution of their substrates, a negatively charged lactoyl group in the PGN substrate of *Bs*PdaA and a peptide-linked lactoyl group in the PGN substrate of *Bs*PdaC. In subsite 0, *Bs*PdaC has Tyr^{377} as part of the conserved MT3 motif of CE4 enzymes (R*X*PY). This Tyr is strictly conserved in chitin deacetylases, highly conserved in PGN GlcNAc deacetylases (with few exceptions where it is replaced by Gly), but differs in *Bs*PdaA, which has an Arg residue instead (Fig. 5). The side chain of Arg^{166} in *Bs*PdaA is oriented toward the active site, and the positive charge of the guanidinyl group may contribute to the stabilization of the free 3-lactoyl group of MurNAc of its PGN substrate (Fig. 7*A* and Fig. S13), as opposed to *Bs*PdaC, in which the substrate is uncharged.

Biological functions of MurNAc deacetylation

Genes encoding for PGN deacetylases are predominantly expressed in Gram-positive bacteria, especially in *Bacillus* species, *Listeria monocytogenes*, and *Streptococcus pneumoniae* (3, 4, 23). Pathogenic *B. cereus* and *B. anthracis* contain 10 polysaccharide deacetylase homologues, from which at least six are PGN deacetylases, involved in virulence and PGN biogenesis during elongation and cell division (13, 24, 25). PGN GlcNAc deacetylases are best characterized for their function in conferring lysozyme resistance to the bacterial cell wall PGN to evade the host innate immune system (3, 4). These deacetylases have been proposed as novel antibacterial targets, but screening studies in the search of selective inhibitors are still scarce (26–28).

MurNAc deacetylation of PGN has been involved in sporulation and germination in different *Bacillus* species participating in muramic δ-lactam synthesis (29). PdaA-deficient spores have no muramic δ-lactam structure in the spore cortex PGN and cannot germinate (6). These MurNAc deacetylases require previous release of the peptide substitution at the 3-*O*-lactoyl group of MurNAc units by the action of an L-alanine amidase (CwlD). Further evidence was provided by the fact that introduction of the *pdaA* and *cwlD* genes into *Escherichia coli* cells led to lactam formation in its cell wall peptidoglycan (5). However, *Bs*PdaC is not associated with sporulation, because a *pdaC*-deficient strain was not affected in sporulation and germination, and the only reported phenotype was lysozyme sensitivity (8). The function of PdaC in its native organism is unknown, other than conferring lysozyme resistance to the cell wall that may not be relevant for a soil bacteria such as *B. subtilis*. *Bs*PdaC is encoded by the *yjeA* gene, which is part of the YycFG regulon involved in regulation of cell wall metabolism, but *yjeA* does not appear to play a prominent role in cell autolysis (18).



subsite	-3	-3	-2	-2	-2	-2	-1	-1	-1	-1	-1	0	0	0
*Bs*PdaC	Y318	Y319	N289	-	R315	S314	Y430	I429	L312	S339	R405	D286	H427	D285
*Sp*PgdA	G308	N309	N279	-	N305	K304	H420	I419	L302	S329	K395	D276	H417	D275
*Bs*PdaA	D106	Q107	N78	Y76	F103	H102	S225	V224	T100	H127	K195	N74	H222	D73

subsite	0	0	0	0	0	0	+1	+1	+1	+1	+1	+1	+2	+3
*Bs*PdaC	H336	H340	S337	W402	Y377	L425	P376	K403	R374	P375	D398	W395	L394	-
*Sp*PgdA	H326	H330	S327	W392	Y367	L415	P366	K393	R364	P365	D388	W385	M384	Y459
*Bs*PdaA	H124	H128	S125	W194	R166	L220	P165	-	R163	P164	A190	W187	F186	F58

Figure 7. Structural superposition and sequence alignment of solvent-exposed amino acids surrounding the active site: *Bs*PdaC (*cyan*) and *Sp*PgdA (*yellow*) (*A*) and *Bs*PdaC (*cyan*) and *Bs*PdaA (*purple*) (*B*). The catalytic center represented by the metal coordination site is in the *middle*. Amino acid labels correspond to *Bs*PdaC. Subsites are defined by analogy to chitin deacetylases crystal structures in complex with ligands. Amino acid side chains of *Bs*PdaC are *labeled* in the structures.

PGN hydrolysis by host lysozymes is an initial event for triggering the innate immune responses. Besides deacetylation of polymeric PGN to confer lysozyme resistance, other modifications of PGN fragments (muropeptides) may impair the detection by host immune receptors. *N*-Acetylmuramyl dipeptide (MDP) was shown to be the minimal PGN fragment to directly interact with Nod2 receptors, ultimately resulting in activation of inflammatory response via the NF-κB and mitogen-activated protein kinase signaling pathways (30). Recently, it has been reported that *N*-deacetylated MDP did not activate NF-κB, thus suggesting that the MurNAc acetyl group is an important feature for recognition by the Nod2 signaling cascade (31). Whether MurNAc deacetylases act only on polymeric PGN or also on PGN fragments after the action of lytic enzymes, including lysozyme, is unknown. *Bs*PdaC is currently the only characterized MurNAc deacetylase active on peptide-substituted PGN chains, yet the minimal size of its PGN substrates is unknown. We hypothesize that PdaC-type MurNAc deacetylase activity might be present in pathogenic bacteria, but not yet identified because of the lack of a clear sequence signature to distinguish GlcNAc and MurNAc deacetylases. This activity may still be hidden in the large number of annotated putative polysaccharide deacetylases, some proposed to be GlcNAc deacetylases but not biochemically characterized. Identification and characterization of homologues of this new subclass of MurNAc deacetylases in other bacteria, especially in pathogenic bacteria, will certainly contribute to unravel the complex degradation-modification–signaling mechanisms in bacteria-host interactions.

Conclusion

*Bs*PdaC is a specific MurNAc deacetylase on peptidoglycan, but it also has GlcNAc deacetylase activity on COS. We have determined the X-ray 3D structure of the CE4 catalytic domain of *Bs*PdaC at 1.54 Å resolution and analyzed the mode of action on COS substrates. The catalytic site retains the conserved Asp-His-His metal-binding triad characteristic of CE4 enzymes acting on GlcNAc residues, which differs from other MurNAc deacetylases that lack the metal-coordinating Asp. On COS, the enzyme follows a multiple-chain (or distributive) mechanism that starts at the internal GlcNAc units and further leads to deacetylation of all but the reducing-end GlcNAc residues. *Bs*PdaC is not associated with sporulation, and its biological function, other than conferring lysozyme resistance, remains to be elucidated. We propose that PdaC is the first member of a new subclass of peptidoglycan MurNAc deacetylases based on these differential functional and structural characteristics.

Materials and methods

Cloning of BsPdaC full-length (FL) and catalytic domain (CD) constructs

A synthetic and codon-optimized gene coding for *Bs*PdaC (amino acid residues 27–467, without signal peptide and Nt-

transmembrane region) (GeneArt® gene synthesis service, Thermo Fisher Scientific) was cloned into a pET22b(+) vector containing a C-terminal Strep-tag sequence (WSHPQFEK). Both the synthetic gene and the vector were digested with the restriction enzymes NdeI and SacI (New England Biolabs) and ligated with T4 DNA ligase (Bio-Rad). From the generated construct (pET22b-*Bs*PdaC-FL), the catalytic domain (from Glu^{270} to Lys^{467}) was amplified by PCR with primers CD_fwd and CD_rvs and cloned into the pET22b(+) vector following the same strategy (list of primers in Table S5). An N terminal Strep-tagged variant of the catalytic domain (BsPdaC-CD-NtStrep) was also prepared. The same sequence corresponding to the catalytic domain (from Glu^{270} to Lys^{467}) was amplified by PCR with primers CD-Nt_fwd (containing a KasI restriction site and the tobacco etch virus proteolysis site) and CD-Nt_rvs (containing a PstI restriction site) and cloned into a pPR-IBA2 vector containing an N-terminal Strep-tag sequence. Both the amplified gene and the vector were digested with the restriction enzymes KasI and PstI (New England Biolabs) and ligated with T4 DNA ligase (Bio-Rad).

The inactive mutants D285S of all constructs and *Bs*PdaC-CD-W402A and Y430A were generated by site-directed mutagenesis (QuikChange SDM, Agilent) with the mutagenic primers listed in Table S5. The final plasmids were verified by gene sequencing. *E. coli* BL21(DE3) cells were transformed with each plasmid for protein expression.

Protein expression and purification

E. coli BL21(DE3) cells harboring the pET expression plasmid pET22b-*Bs*PdaC-FL (full-length protein), pET22b-*Bs*PdaC-CD (catalytic domain with Strep-tag at the C terminus), pPR-IBA2-*BsPdaC*-CD-NtStrep (catalytic domain with Strep-tag at the N terminus) and the inactive mutants pET22b-*Bs*PdaC-CD-D285S and pPR-IBA2-*Bs*PdaC-CD-NtStrep-D285S were grown in an autoinduction medium (32) containing 100 μg/ml ampicillin at 25 °C, 170 rpm for 48 h. The cells were harvested by centrifugation, suspended in PBS buffer (50 mM phosphate, 300 mM NaCl, pH 7) with 1 mM serine protease inhibitor phenylmethylsulfonyl fluoride, and disrupted by sonication. After centrifugation, the protein in the supernatant was purified by affinity chromatography with a Strep Trap column (5 ml; GE Healthcare). The protein was eluted with 2.5 mM D-desthibiotin in PBS, and, after concentration with an Amicon Ultra-15 centrifugal filter (Millipore), it was loaded onto a Superdex 200 size-exclusion chromatography column (GE Healthcare) and eluted with PBS, displaying different oligomeric forms (Fig. S2). Fraction F1 is composed of soluble oligomers, fraction F2 corresponds to dimers, and fraction F3 is the monomeric enzyme. The relative intensity of the different fractions was slightly different for both *Bs*PdaC-FL and *Bs*PdaC-CD as well as between different batches of the same protein. The monomeric fraction (F3) resulted in being the most active for both *Bs*PdaC-FL and *Bs*PdaC-CD proteins, and it was the one used for kinetic characterization and crystal screening. Analysis of the eluted proteins by SDS-PAGE showed the expected apparent molecular masses, which were assessed by MALDI-TOF MS (52.7 kDa for the FL enzyme and 24.4 kDa for the CD protein). The thermal stability of both the FL and CD proteins were similar, with a melting temperature (T_m) of about 60 °C, as determined by a thermal shift assay (Fig. S4). Protein concentration was determined with the BCA protein assay kit (Thermo Fisher Scientific).

Thermal stability

Purified protein (*Bs*PdaC-FL or *Bs*PdaC-CD; 4 μM) in PBS (50 mM phosphate, 300 mM NaCl, pH 7) was mixed with the commercial dye Sypro Orange (Thermo Fisher Scientific; 1:5000 dilution according to the manufacturer's protocol) in a final volume of 25 μl. The samples were subjected to a thermal gradient in a thermocycler (Rotogene 3000, Corbett Research), consisting of 1 min at 25 °C followed by 1 °C increments (30 s at each temperature) up to 95 °C. The fluorescence was measured (λ_{ex} = 483 nm, λ_{em} = 560 nm), and T_m was determined by fitting the data to a Boltzmann sigmoidal equation using Prism (GraphPad Software, La Jolla, CA).

Enzyme activity of gel filtration fractions

All fractions collected from the size-exclusion chromatography were evaluated with 4-methylumbelliferyl acetate (AcOMU) as nonspecific esterase substrate. 10 μl of sample were incubated with 0.5 mM AcOMU in a final volume of 200 μl in PBS buffer, pH 7.0. Enzyme activity was determined by monitoring the fluorescence increase (λ_{ex} = 340 nm, λ_{em} = 460 nm) due to 4-methylumbelliferone (MU) release and expressed as initial rates (μM/s) using MU standards in the same buffer.

Enzyme activity on peptidoglycan

Enzyme activity toward *B. subtilis* peptidoglycan (from Sigma-Aldrich) was determined by monitoring the release of acetic acid with a commercial acetate determination kit (Acetic Acid Assay Kit, Megazyme). Reactions were performed by incubating the peptidoglycan substrate (1 mg/ml) and purified *Bs*PdaC-CD or *Bs*PdaC-FL (0.25–5 μM) in PBS (50 mM phosphate, 300 mM NaCl) buffer, pH 7.0, at 37 °C with shaking for 30 min. At different time intervals, the reaction samples were heated at 100 °C for 5 min to inactivate the enzyme and then centrifuged. The supernatants were collected and used for measurement of the acetate released. Specific activity was expressed as μM acetate/min/μM enzyme (specific activity in min^{-1}).

To assess that *Bs*PdaC-CD maintains the same MurNAc deacetylase activity as the full-length enzyme, end-point reactions with both *Bs*PdaC-CD and *Bs*PdaC-FL (0.5 μM enzyme, 1 mg/ml PGN in PBS buffer, pH 7.0, 37 °C for 48 h with agitation) were set up, and the release of acetic acid was determined as described above. After dialysis against PBS buffer to remove the free acetate from the samples, *Bs*Pdac-CD was added to the sample of the initial *Bs*PdaC-FL reaction, and *Bs*PdaC-FL was added to the sample of the initial *Bs*PdaC-CD reaction. After a 48-h incubation at 37 °C, the released acetate was determined. Finally, the GlcNAc peptidoglycan deacetylase *Sp*PgdA (0.8 μM) was added to the final reactions to monitor acetate release. As reference, the activity of *Sp*PgdA on *B. subtilis* PGN was determined.

Enzyme kinetics by the fluorescamine assay

Enzyme activity was determined by fluorescamine labeling of the amino groups generated in the deacetylation reaction using COS substrates (11). All reactions were performed by incubating $(GlcNAc)_n$, $n = 4$ or 5, substrate (2 mM), and purified *Bs*PdaC-CD (0.15–1.3 μM) in PBS (50 mM phosphate, 300 mM NaCl), pH 7, at 37 °C in a final volume of 200 μl in microtiter plates. For determination of the pH optimum, the pH was varied from 6.0 to 10.6 at 37 °C. For optimum temperature determination, the temperature was varied from 25 to 75 °C at pH 7.0 (Fig. S3). At different time intervals, aliquots (20 μl) were withdrawn and added to 90 μl of water/1-propanol (1:1) to stop the reaction. Then 20 μl of 2 mg/ml fluorescamine in DMF were added, and the reaction was incubated for 10 min at room temperature. The labeling reaction was terminated by adding 150 μl of DMF/water (1:1), and fluorescence was quantified (λ_{ex} = 340 nm, λ_{em} = 460 nm). A set of glucosamine standards were treated following the same protocol.

Enzyme kinetics by HPLC-MS

The formation of products with different degrees of acetylation from $(GlcNAc)_n$, $n = 3$–5, substrates was monitored by HPLC-MS (16) (Agilent 1260 HPLC-MS, electrospray ionization (ESI+), single-quadrupole MS detector) using an XBridge BEH Amide 2.5-μm, 3.0 × 100-mm XP column (Waters) in combination with an XBridge BEH Amide Guard Cartridge (2PK) precolumn (2.5 μm, 4.6 × 20 mm; Waters). Reactions were performed by incubating the $(GlcNAc)_n$ substrate (2 mM) and purified *Bs*PdaC-CD (5 μM) in PBS, pH 7.0, at 37 °C and stopped at different time intervals as described above. 5-μl samples were injected into the system and eluted at 60 °C with acetonitrile/water (65:35, v/v), 1% formic acid at a flow rate of 0.4 μl/min. MS detection was performed on both SIM mode (for monitoring $[M + H]^+$ of substrate and deacetylation products) and SCAN mode (for total ion monitoring, 250–1100 m/z scan range). The m/z values monitored in SIM mode were as follows (A for GlcNAc, D for $GlcNH_2$): (*a*) reactions with tetraacetylchitotetraose (A4) substrate: 831.4 (A4), 789.4 (A3D1), 747.4 (A2D2), and 353.3 (A1D3); (*b*) reactions with pentaacetylchitopentaose (A5) substrate: 1035.4 (A5), 992.4 (A4D1), 950.4 (A3D2), 908.4 (A2D3), and 867.3 (A1D4). Specific activities for the initial monodeacetylation reactions with $(GlcNAc)_n$ substrates were performed as above, using the corresponding monodeacetylated product standards for quantification (prepared by reaction of the corresponding $(GlcNAc)_n$ substrate with *Vibrio cholera* CDA as reported (16)). Michaelis–Menten parameters were determined for the initial monodeacetylation reaction with $(GlcNAc)_4$, varying the substrate concentration from 0.5 to 17.5 mM with 3.7 μM enzyme in PBS buffer, pH 7.0, 37 °C.

Determination of the deacetylation pattern

To determine the pattern of acetylation of the different products formed during the time course monitoring, preparative reactions with 2 mM $(GlcNAc)_4$ and $(GlcNAc)_5$ substrates and 0.1 mg of purified *Bs*PdaC-CD in PBS, pH 7.0, at 37 °C, in a final volume of 1 ml, were incubated for different reaction times. The freeze-dried samples were analyzed for pattern of acetylation by MS (MALDI-TOF-MS/MS analysis) after previous ^{18}O labeling and re-*N*-acetylation with $[^2H_6]$acetic anhydride according to the protocol by Cord-Landwehr *et al.* (20), and the relative amount of each sequence in the mixtures was calculated based on the peak intensities compared for each ion type.

Effect of metal ions and EDTA

Purified protein (*Bs*PdaC-CD) in PBS (50 mM phosphate, 300 mM NaCl, pH 7.0) was dialyzed against PBS with EDTA (200 mM) and urea (2 or 7 M). During the refolding step, EDTA and urea were removed by dialysis (first with PBS and then with water). Activity of the treated proteins and the native enzyme with and without the addition of metal ions (Zn^{2+}) was determined by HPLC-MS. The metal content of the apoenzyme (after the refolding step from 7 M urea) was $[Zn^{2+}]$ (μM)/$[E]$ (μM) = 0.1 by inductively coupled plasma MS (NexION 300XX, PerkinElmer Life Sciences), meaning that ~90% of the bound metal was removed following this procedure.

Effect of ligands on thermal stability

A J-810 CD spectropolarimeter (Jasco Corp., Tokyo, Japan) equipped with a Peltier thermal device was used. Measurements were carried out in a range from 20 to 90 °C at 222 nm by using quartz cuvettes with a 1-mm optical path length. Temperature was increased stepwise at 1 °C/min. Samples were 10 μM *Bs*PdaC-CD-NtStrep-tag in 50 mM Tris buffer, 150 mM NaF, pH 7.5. The effect of substrate on the thermal stability was tested under the same conditions, using 1 mM $(GlcNAc)_2$, $(GlcNAc)_4$, and $(GlcNAc)_5$. Transitions were fitted according to a Boltzmann sigmoidal equation using the GraphPad software.

Crystallographic structure

BsPdaC-CD crystallization and data collection—Crystals of *Bs*PdaC-CD (WT and D285S mutant) were obtained by mixing 0.25 μl of the protein in 50 mM Tris, pH 7.0, and 150 mM NaCl at 5.6 or 10 mg ml^{-1}, respectively, with 0.25 μl of a mother liquor containing 0.2 M ammonium tartrate dibasic, pH 6.6, 20% (w/v) PEG 3.350 or 0.4 M ammonium phosphate monobasic by sitting-drop vapor-diffusion crystallization. Crystals appeared after 8–10 days and grew as prisms, reaching 0.1–0.8 × 0.5–0.4 × 0.05–0.07 mm. Single crystals of *Bs*PdaC-CD and *Bs*PdaC-CD-D285S were cryocooled in liquid nitrogen by using a cryoprotectant solution of the mother liquor supplemented with 15% (v/v) ethylene glycerol or 30% (v/v) glycerol, respectively. X-ray diffraction data were collected on a PILATUS 6MF pixel array detector at the microfocus I02 beamline (λ = 0.9796 Å, Diamond Light Source) and integrated with XDS (33) following standard procedures.

Structure determination and refinement—Structures of *Bs*PdaC-CD *and Bs*PdaC-CD-D285S were solved using as a template the previously reported *S. pneumoniae* peptidoglycan deacetylase *Sp*PgdA (PDB code 2C1G) and molecular replacement methods implemented in Phaser (34) and the PHENIX suite (35). Model rebuilding was carried out with Buccaneer (36) and the CCP4 suite (37). The final manual building was performed with Coot (38) and refinement with phenix.refine (35). The structures were validated by MolProbity (39). Data

collection and refinement statistics are presented in Table S4. Atomic coordinates and structure factors have been deposited with the Protein Data Bank, accession codes 6H8L and 6H8N corresponding to *Bs*PdaC-CD and *Bs*PdaC-CD-D285S crystal structures, respectively. Molecular graphics and structural analyses were performed with the UCSF Chimera package (40).

Modeling of BsPdaC, BsPdaA, and SpPgdA structures in complex with substrates

The 3D structures of the enzyme-ligand complexes of *Sp*PgdA, *Bs*PdaA, and *Bs*PdaC were modeled with AUTODOCK 4.2 and AUTODOCK VINA (41, 42). Protein structures were directly taken from the Protein Data Bank with PDB accession codes 2C1G (*Sp*PgdA), 1W1B (*Bs*PdaA), and 6H8L (*Bs*PdaC, this work). Ligand structures covering the different range of accepted substrates were extracted from other protein structures with PDB accession codes 6AVE (GlcNAc), 1TWQ (MurNAc), 9LYZ (GlcNAcβ(1→4)MurNAc and MurNAc β(1→4)GlcNAc), and 1LZC (tetraacetylchitotetraose). Both the protein and ligand structures were first parametrized with AutoDockTools (41): polar hydrogens were added, AutoDock4.2 atom typing was used, and Gaisteger partial charges were computed. All rotable bonds of the ligands were considered free during the docking calculations, except amide and ring bonds, whereas the whole protein structure was kept fixed. Tetraacetylchitotetraose, $(GlcNAc)_4$, was docked to the *Bs*PdaC structure with AUTODOCK VINA (42). The docking search space was confined in a 19.5 × 27.75 × 36.75 Å^3 box centered in the active site. Exhaustiveness level was set to 24, and 20 binding modes of the ligands were generated. Only low-energy binding poses were considered for analysis. Distances and surfaces were measured, and pictures of the complexes were generated with VMD (43). Docking of mono- and disaccharides with AUTODOCK 4.2 was used as a probe for the discovery of putative binding subsites along the enzymatic cavity of *Sp*PgdA, *Bs*PdaA, and *Bs*PdaC structures. A grid box of 31.5 × 25.5 × 28.5 Å^3 with a grid point spacing of 0.375 Å, centered at the active site, was used as the search space for docking. For each ligand docking, 1000 rounds of the genetic algorithm implemented in AUTODOCK 4.2 were performed. For each round, an initial population of 150 members was considered, with randomized initial position and orientation coordinates and randomized conformations of the substrate flexible bonds. The genetic algorithm was extended up to 27,000 offspring generations, with a maximum of 2,500,000 energy evaluations. Only low-energy and repetitive binding poses were considered for analysis.

Sequence alignment and phylogenetic analysis of CE4 enzymes

The sequence of *Bs*PdaC catalytic domain (CE4 domain) was retrieved from Uniprot (accession code O34798, aa 278–456) and incorporated into the multiple-sequence alignment guided by the structural superimposition of characterized CE4 enzymes (12, 16) by hidden Markov model comparisons using HMMER (44). *Bs*PdaC was analyzed together with the following sequences (shown as name (organism, Uniprot accession code)) (Fig. S11): (*a*) chitin deacetylases: *Mr*CDA (*Mucor rouxii*, P50325), *Cl*CDA (*Colletotrichum lindemuthianum*, Q6DWK3), *An*CDA (*Aspergillus nidulans*, Q5AQQ0), *Pes*CDA (*Pestalotiopsis* sp., A0A1L3THR9), *Pa*CDA (*Podospora anserine*, B2AAQ0), *Pgt*CDA (*Puccinia graminis*, E3K3D7), *Vc*CDA (*V. cholera*, Q9KSH6), *Vp*CDA (*Vibrio parahemeolyticus*, A6P4T5) *Rm*NodB (*Rhizobium meliloti*, P02963), *Ar*CE4 (*Arthrobacter* sp., A0A2C8C1T7); (*b*) peptidoglycan deacetylases: *SpPgdA* (*S. pneumoniae*, Q8DP63), *Sm*Pgda (*Streptococcus mutants*, Q8DV82), *Bc*Pgd (*B. cereus*, Q81EK9), *Er*Pgd (*Eubacterium rectale*, C4ZEZ9), *Bs*PdaA (*B. subtilis*, O34928), *Ba*CE4 (*B. anthracis*, Q81Z49); (*c*) unknown: BC0361 (*B. cereus*, Q81IM3), BA0330 (*B. anthracis*, Q81ZD9), BA0150 (*B. anthracis*, Q81VP2), ECU11_0510 (*Encephalitozoon cuniculi*, Q8SU65), (*d*) acetylxylan esterases: *Sl*AxeA (*Streptomyces lividans*, Q54413), *Ct*AxeA (*Clostridium thermocellum*, O87119); and (*e*) poly-β1,6-GlcNAc deacetylases: *Ec*PgaB (*E. coli*, a P75906), *Ad*IcaB (*Ammonifex degensii*, C9RCK9), *Bb*BpsB (*Bordetella bronchiseptica*, A0A0C6P1R7), *Aa*PgaB (*Aggregatibacter actinomycetemcomitans*, A5HJW8).

The phylogenetic relationships were inferred by using the maximum likelihood method based on the JTT matrix-based model (45) computed from the multiple-sequence alignment. Bootstrap analysis consisted of 500 replicates. These evolutionary analyses were conducted with MEGA7 (46), and the output dendrogram is shown in Fig. 5.

Author contributions—L. G.-R., M. A. S.-P., D. A.-J., M. E. G., X. B., and A. P. investigation; L. G.-R. and A. P. writing-original draft; M. E. G. and A. P. writing-review and editing; A. P. conceptualization; A. P. funding acquisition.

Acknowledgments—We gratefully acknowledge Diamond Light Source (proposal mx15304), ALBA, and iNEXT (proposal numbers 1618/2538) for granting access to synchrotron radiation facilities. We thank Dr. Hugo Aragunde for the sequencing of deacetylated COS and Prof. Bruno Moerschbacher and Dr. Stefan Cord-Landwehr (University of Münster) for advice on MS/MS sequencing analysis. Montserrat Murcia is acknowledged for performing some of the docking simulations.

References

1. Araki, Y., Nakatani, T., Hayashi, H., and Ito, E. (1971) Occurrence of non-*N*-substituted glucosamine residues in lysozyme-resistant peptidoglycan from *Bacillus cereus* cell walls. *Biochem. Biophys. Res. Commun.* **42,** 691–697 CrossRef Medline
2. Zipperle, G. F., Jr., Ezzell, J. W., Jr., and Doyle, R. J. (1984) Glucosamine substitution and muramidase susceptibility in Bacillus anthracis. *Can. J. Microbiol.* **30,** 553–559 CrossRef Medline
3. Wolf, A. J., and Underhill, D. M. (2018) Peptidoglycan recognition by the innate immune system. *Nat. Rev. Immunol.* **18,** 243–254 CrossRef Medline
4. Vollmer, W. (2008) Structural variation in the glycan strands of bacterial peptidoglycan. *FEMS Microbiol. Rev.* **32,** 287–306 CrossRef Medline
5. Gilmore, M. E., Bandyopadhyay, D., Dean, A. M., Linnstaedt, S. D., and Popham, D. L. (2004) Production of muramic δ-lactam in Bacillus subtilis spore peptidoglycan. *J. Bacteriol.* **186,** 80–89 CrossRef Medline
6. Fukushima, T., Yamamoto, H., Atrih, A., Foster, S. J., and Sekiguchi, J. (2002) A polysaccharide deacetylase gene (pdaA) is required for germination and for production of muramic δ-lactam residues in the spore cortex of *Bacillus subtilis*. *J. Bacteriol.* **184,** 6007–6015 CrossRef Medline

7. Fukushima, T., Tanabe, T., Yamamoto, H., Hosoya, S., Sato, T., Yoshikawa, H., and Sekiguchi, J. (2004) Characterization of a polysaccharide deacetylase gene homologue (pdaB) on sporulation of *Bacillus subtilis*. *J. Biochem.* **136,** 283–291 CrossRef Medline
8. Kobayashi, K., Sudiarta, I. P., Kodama, T., Fukushima, T., Ara, K., Ozaki, K., and Sekiguchi, J. (2012) Identification and characterization of a novel polysaccharide deacetylase C (PdaC) from *Bacillus subtilis*. *J. Biol. Chem.* **287,** 9765–9776 CrossRef Medline
9. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42,** D490–D495 CrossRef Medline
10. John, M., Röhrig, H., Schmidt, J., Wieneke, U., and Schell, J. (1993) Rhizobium NodB protein involved in nodulation signal synthesis is a chitooligosaccharide deacetylase. *Proc. Natl. Acad. Sci. U.S.A.* **90,** 625–629 CrossRef Medline
11. Blair, D. E., Schüttelkopf, A. W., MacRae, J. I., and van Aalten, D. M. F. (2005) Structure and metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal virulence factor. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 15429–15434 CrossRef Medline
12. Aragunde, H., Biarnés, X., and Planas, A. (2018) Substrate recognition and specificity of chitin deacetylases and related family 4 carbohydrate esterases. *Int. J. Mol. Sci.* **19,** E412 CrossRef Medline
13. Balomenou, S., Fouet, A., Tzanodaskalaki, M., Couture-Tosi, E., Bouriotis, V., and Boneca, I. G. (2013) Distinct functions of polysaccharide deacetylases in cell shape, neutral polysaccharide synthesis and virulence of *Bacillus anthracis*. *Mol. Microbiol.* **87,** 867–883 CrossRef Medline
14. Fukushima, T., Kitajima, T., and Sekiguchi, J. (2005) A polysaccharide deacetylase homologue, PdaA, in *Bacillus subtilis* acts as an *N*-acetylmuramic acid deacetylase *in vitro*. *J. Bacteriol.* **187,** 1287–1292 CrossRef Medline
15. Blair, D. E., and van Aalten, D. M. F. (2004) Structures of *Bacillus subtilis* PdaA, a family 4 carbohydrate esterase, and a complex with *N*-acetylglucosamine. *FEBS Lett.* **570,** 13–19 CrossRef Medline
16. Andrés, E., Albesa-Jové, D., Biarnés, X., Moerschbacher, B. M., Guerin, M. E., and Planas, A. (2014) Structural basis of chitin oligosaccharide deacetylation. *Angew. Chem. Int. Ed. Engl.* **53,** 6882–6887 CrossRef Medline
17. Oberbarnscheidt, L., Taylor, E. J., Davies, G. J., and Gloster, T. M. (2007) Structure of a carbohydrate esterase from *Bacillus anthracis*. *Proteins* **66,** 250–252 CrossRef Medline
18. Bisicchia, P., Noone, D., Lioliou, E., Howell, A., Quigley, S., Jensen, T., Jarmer, H., and Devine, K. M. (2007) The essential YycFG two-component system controls cell wall metabolism in *Bacillus subtilis*. *Mol. Microbiol.* **65,** 180–200 CrossRef Medline
19. Fukushima, T., Furihata, I., Emmins, R., Daniel, R. A., Hoch, J. A., and Szurmant, H. (2011) A role for the essential YycG sensor histidine kinase in sensing cell division. *Mol. Microbiol.* **79,** 503–522 CrossRef Medline
20. Cord-Landwehr, S., Ihmor, P., Niehues, A., Luftmann, H., Moerschbacher, B. M., and Mormann, M. (2017) Quantitative mass-spectrometric sequencing of chitosan oligomers revealing cleavage sites of chitosan hydrolases. *Anal. Chem.* **89,** 2893–2900 CrossRef Medline
21. Grifoll-Romero, L., Pascual, S., Aragunde, H., Biarnés, X., and Planas, A. (2018) Chitin deacetylases: structures, specificities, and biotech applications. *Polymers* (*Basel*) **10,** 352 CrossRef Medline
22. Vollmer, W., and Tomasz, A. (2000) The pgdA gene encodes for a peptidoglycan *N*-acetylglucosamine deacetylase in *Streptococcus pneumoniae*. *J. Biol. Chem.* **275,** 20496–20501 CrossRef Medline
23. Boneca, I. G., Dussurget, O., Cabanes, D., Nahori, M.-A., Sousa, S., Lecuit, M., Psylinakis, E., Bouriotis, V., Hugot, J.-P., Giovannini, M., Coyle, A., Bertin, J., Namane, A., Rousselle, J.-C., Cayet, N., *et al.* (2007) A critical role for peptidoglycan *N*-deacetylation in Listeria evasion from the host innate immune system. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 997–1002 CrossRef Medline
24. Arnaouteli, S., Giastas, P., Andreou, A., Tzanodaskalaki, M., Aldridge, C., Tzartos, S. J., Vollmer, W., Eliopoulos, E., and Bouriotis, V. (2015) Two putative polysaccharide deacetylases are required for osmotic stability and cell shape maintenance in *Bacillus anthracis*. *J. Biol. Chem.* **290,** 13465–13478 CrossRef Medline
25. Psylinakis, E., Boneca, I. G., Mavromatis, K., Deli, A., Hayhurst, E., Foster, S. J., Vårum, K. M., and Bouriotis, V. (2005) Peptidoglycan *N*-acetylglucosamine deacetylases from *Bacillus cereus*, highly conserved proteins in *Bacillus anthracis*. *J. Biol. Chem.* **280,** 30856–30863 CrossRef Medline
26. Bui, N. K., Turk, S., Buckenmaier, S., Stevenson-Jones, F., Zeuch, B., Gobec, S., and Vollmer, W. (2011) Development of screening assays and discovery of initial inhibitors of pneumococcal peptidoglycan deacetylase PgdA. *Biochem. Pharmacol.* **82,** 43–52 CrossRef Medline
27. Giastas, P., Andreou, A., Papakyriakou, A., Koutsioulis, D., Balomenou, S., Tzartos, S. J., Bouriotis, V., and Eliopoulos, E. E. (2018) Structures of the peptidoglycan *N*-acetylglucosamine deacetylase Bc1974 and its complexes with zinc metalloenzyme inhibitors. *Biochemistry* **57,** 753–763 CrossRef Medline
28. Balomenou, S., Koutsioulis, D., Tomatsidou, A., Tzanodaskalaki, M., Petratos, K., and Bouriotis, V. (2018) Polysaccharide deacetylases serve as new targets for the design of inhibitors against *Bacillus anthracis* and *Bacillus cereus*. *Bioorg. Med. Chem.* **26,** 3845–3851 CrossRef Medline
29. Popham, D. L., Helin, J., Costello, C. E., and Setlow, P. (1996) Muramic lactam in peptidoglycan of *Bacillus subtilis* spores is required for spore outgrowth but not for spore dehydration or heat resistance. *Proc. Natl. Acad. Sci. U.S.A.* **93,** 15405–15410 CrossRef Medline
30. Moreira, L. O., and Zamboni, D. S. (2012) NOD1 and NOD2 signaling in infection and inflammation. *Front. Immunol.* **3,** 328 CrossRef Medline
31. Melnyk, J. E., Mohanan, V., Schaefer, A. K., Hou, C. W., and Grimes, C. L. (2015) Peptidoglycan modifications tune the stability and function of the innate immune receptor Nod2. *J. Am. Chem. Soc.* **137,** 6987–6990 CrossRef Medline
32. Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41,** 207–234 CrossRef Medline
33. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 125–132 CrossRef Medline
34. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 CrossRef Medline
35. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221 CrossRef Medline
36. Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62,** 1002–1011 CrossRef Medline
37. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 235–242 CrossRef Medline
38. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501 CrossRef Medline
39. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 12–21 CrossRef Medline
40. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 CrossRef Medline
41. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009) Software news and updates AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* **30,** 2785–2791 CrossRef Medline
42. Trott, O., and Olson, A. J. (2010) Software news and update AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31,** 455–461 CrossRef Medline

43. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.* **14,** 33–38, 27–28 CrossRef Medline
44. Eddy, S. R. (1998) Profile hidden Markov models. *Bioinformatics* **14,** 755–763 CrossRef Medline
45. Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992) The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* **8,** 275–282 CrossRef Medline
46. Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33,** 1870–1874 CrossRef Medline
47. Varki, A., Cummings, R. D., Aebi, M., Packer, N. H., Seeberger, P. H., Esko, J. D., Stanley, P., Hart, G., Darvill, A., Kinoshita, T., Prestegard, J. J., Schnaar, R. L., Freeze, H. H., Marth, J. D., Bertozzi, C. R., *et al.* (2015) Symbol nomenclature for graphical representations of glycans. *Glycobiology* **25,** 1323–1324 CrossRef Medline

Structure-function relationships underlying the dual *N*-acetylmuramic and *N*-acetylglucosamine specificities of the bacterial peptidoglycan deacetylase PdaC

Laia Grifoll-Romero, María Angela Sainz-Polo, David Albesa-Jové, Marcelo E. Guerin, Xevi Biarnés and Antoni Planas

J. Biol. Chem. 2019, 294:19066-19080.
doi: 10.1074/jbc.RA119.009510 originally published online November 5, 2019

Supporting Information

Structure-function relationships underlying the dual N-acetylmuramic and N-acetylglucosamine specificities of the bacterial peptidoglycan deacetylase PdaC

Laia Grifoll-Romero[‡], María Angela Sainz-Polo[§], David Albesa-Jové[§,¶], Marcelo E. Guerin[§,¶], Xevi Biarnés[‡], and Antoni Planas[‡,*]

From the [‡]Laboratory of Biochemistry, Institut Químic de Sarrià, University Ramon Llull, 08017 Barcelona, Spain, [§]Structural Biology Unit, CIC bioGUNE, Bizkaia Technology Park, Ed. 801A, 48160 Derio, Spain, [¶]IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain

1. Supporting Tables

2. Supporting Figures

3. Supporting References

Table S1. CE4 enzymes with known 3D structure (October 2019).

ACTIVITY	ENZYME	ORGANISM	PDB	REF
Chitin DA	*Cl*CDA	*Colletotrichum lindemutianum*	2IW0	Blair et al., 2006 (1)
	*An*CDA	*Aspergillus nidulans*	2Y8U	Liu et al., 2017 (2)
	*Vc*CDA	*Vibrio cholerae*	4NY2	Andrés et al., 2014 (3)
	*Vp*CDA	*Vibrio parahaemolyticus*	3WX7	Hirano et al., 2015 (4)
	*Ar*CE4A	*Arthrobacter sp.*	5LFZ	Tuveng, 2017 (5)
PG MurNAc DA	*Bs*PdaA	*Bacillus subtilis*	1W17	Blair et al., 2004 (6)
	*Ba*CE4 (BA0424)	*Bacillus anthracis*	2J13	Oberbarnscheidt et al., 2006 (7)
	*Bs*PdaC	*Bacillus subtilis*	6H8L	**This work**
PG GlcNAc DA	*Sp*PgdA	*Streptococcus pneumoniae*	2C1G	Blair et al, 2005 (8)
	*Sm*PgdA	*Streptococcus mutants*	2W3Z	Deng et al., 2009 (9)
	*Bc*Pgd (BC1960)	*Bacillus cereus*	4L1G	Tsalafouta et al., 2008 (10)
	BC1974	*Bacillus cereus*	5N1J	Giastas et al. 2018 (11)
	*Er*Pgd	*Eubacterium rectale*	5JMU	--
PPda (unk)[a]	BC0361	*Bacillus cereus*	4HD5	Fadouloglou et al. 2013, (12)
	BA0330	*Bacillus anthracis*	4V33	Arnaouteli et al. 2015 (13)
	BA0150[b]	*Bacillus anthracis*	4M1B	Strunk et al. 2014 (14)
	ECU11_0510[b]	*Encephalitozoon cuniculi*	2VYO	Urch et al. 2009 (15)
	BA3943	*Bacillus antracis*	6HM9	--
	Bp2533	*Burkholderia pseudomallei*	3S6O	--
	PA1517	*Pseudomonas aeruginosa*	1Z7A	--
	SpPdi	*Streptococcus pyogenes*	6DQ3	-
Acetylxylan esterase	*Sl*AxeA	*Streptomyces lividans*	2CC0	Taylor et al., 2006 (16)
	*Ct*AxeA	*Clostridium thermocellum*	2C71	Taylor et al., 2006 (16)
Poly-ß-1,6-GlcNAc DA	*Ec*PgaB	*Escherichia coli*	3VUS	Nishiyama et al., 2013 (17)
	*Ad*IcaB	*Ammonifex degensii*	4WCJ	Little et al., 2014 (18)
	*Bb*BpsB	*Bordetella bronchiseptica*	5BU6	Little et al., 2015 (19)
	*Aa*PgaB	*Aggregatibacter actinomycetemcomitans*	4U10	Parthiban et al., 2015 (20)

[a] PPda (unk): putative polysaccharide deacetylases with unknown activity. [b] Presumably inactive (no metal coordination). [c] Inactive (lack of Asp general base and His metal-binding).

Table S2. Deacetylation activity of *Bs*PdaC-FL and *Bs*PdaC-CD on *Bacillus subtilis* PGN.

Reaction	Enzyme/Substrate	Acetate released ($\mu M\ AcO^-/\mu M$ E) at 48 h
1	*Bs*PdaC-CD + PGN	3252
2	*Bs*PdaC-FL + PGN	4010
3	Reaction 1 + *Bs*PdaC-FL	53.6
4	Reaction 2 + *Bs*PdaC-CD	18.4
5	*Sp*PgdA + PGN	3358
6	Reaction 3 + *Sp*PgdA	2680 (80%)[1]
7	Reaction 4 + *Sp*PgdA	2434 (72%)[1]

Conditions: 0.5 µM enzyme, pH 7.0, 37ºC. After 48h incubation with the first enzyme (reactions 1 and 2), the sample is dialyzed and the second enzyme is added (reactions 3 and 4). Likewise, after 48h incubations of reactions 3 and 4, samples are dialyzed and *Sp*PgdaA is added (reactions 6 and 7).

[1]Percentatge of *Sp*PgdA activity relative to activity on intact PGN (reaction 5).

Table S3. MALDI-TOF-MS/MS analysis of the deacetylation products from $(GlcNAc)_4$ and $(GlcNAc)_5$ by *Bs*PdaC. PA: pattern of acetylation, A: GlcNAc, R: reacetylated GlcN with [$^2H_6$]-acetic anhydride. Procedure according to Cord-Landwerh *et al*. (21).

$(GlcNAc)_4$

Sample	Reaction time	PA	m/z $[M+H]^+$	Abundance (%)	Products (%) per type [a]
DP4.1	30 min	AAAA	831.335	61.0	
		AARA	834.354	19.6	50.3
		ARAA		19.4	49.7
DP4.2	5 h	AAAA	831.335	12.3	
		AARA	834.354	36.0	56.9
		ARAA		27.3	43.1
		ARRA	837.373	24.4	
DP4.3	24 h	AARA	834.354	7.2	27.9
		ARAA		18.7	72.1
		ARRA	837.373	58.8	88.8
		RARA		7.4	11.2
		RRRA	840.392	7.9	
DP4.4	96 h	ARAA	834.354	9.0	
		ARRA	837.373	55.3	89.6
		RRAA		6.4	10.4
		RRRA	840.392	29.3	

[a] % of isobaric products in each sample.

$(GlcNAc)_5$

Sample	Reaction time	PA	m/z $[M+H]^+$	Abundance (%)	Products (%) per type [a]
DP5.1	30 min	AAAAA	1034.415	22.9	
		AAARA	1037.434	32.9	50.7
		AARAA		23.0	35.5
		ARAAA		9.0	13.8
		AARRA	1040.452	12.1	
DP5.2	5 h	AARRA	1040.452	54.3	77.4
		ARARA		15.8	22.6
		ARRRA	1043.471	29.9	
DP5.3	24 h	AARRA	1040.452	14.2	45.6
		ARARA, ARRAA[b]		9.2	29.5
		ARRAA, RARAA[b]		7.7	24.9
		ARRRA	1043.471	46.7	81.0
		RARRA		11.0	19.0
		RRRRA	1046.490	11.2	
DP5.4	96 h	ARRAA, RARAA[b]	1040.452	10.8	50.4
		ARARA, ARRAA[b]		10.7	49.6
		ARRRA	1043.471	43.2	
		RRRRA	1046.490	35.2	

[a] % of isobaric products in each sample. [b] Any of both structures, or both.

Table S4. *Bs*PdaC-CD (wt and D285S mutant) x-ray data collection and refinement statistics.

	PdaC-CD	PdaC-CD-D285S
Wavelength	0.9796 Å	0.9796 Å
Resolution range	56.07 - 1.541 (1.596 - 1.541)	21.88 - 1.26 (1.305 - 1.26)
Space group	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1
Unit cell	55.346 61.916 132.226 90 90 90	55.42 61.8 130.99 90 90 90
Total reflections	440385 (44864)	690848 (29571)
Unique reflections	67783 (6689)	120462 (10846)
Multiplicity	6.5 (6.7)	5.7 (2.8)
Completeness (%)	100 (100)	99 (89)
Mean I/sigma(I)	11.39 (2.04)	11.01 (1.74)
Wilson B-factor	17.77	12.29
R-merge	0.0877 (0.8135)	0.06804 (0.4831)
R-meas	0.0957 (0.8819)	0.07441 (0.595)
CC1/2	0.998 (0.713)	0.999 (0.707)
CC*	0.999 (0.912)	1 (0.91)
Reflections used in refinement	67782 (6689)	120628 (10846)
Reflections used for R-free	3327 (333)	5936 (546)
R-work	0.1749 (0.2463)	0.1600 (0.3166)
R-free	0.1944 (0.2646)	0.1745 (0.3127)
CC(work)	0.951 (0.841)	0.967 (0.816)
CC(free)	0.958 (0.760)	0.959 (0.830)
Number of non-hydrogen atoms	3697	3830
macromolecules	3277	3334
ligands	22	29
Protein residues	408	408
RMS(bonds)	0.005	0.007
RMS(angles)	0.79	0.98
Ramachandran favored (%)	98	97
Ramachandran allowed (%)	1.9	1.7
Ramachandran outliers (%)	0.49	0.96
Rotamer outliers (%)	0.56	1.1
Clashscore	1.07	2.08
Average B-factor	21.24	17.16
macromolecules	19.78	15.32
ligands	16.44	22.63
solvent	33.59	29.89

Table S5. Primers used for cloning and side directed mutagenesis.

Cloning primers	
CD_fwd	5'-TAGTGGCATATGGAAGAAACAGTTGATCCGAATC-3'
CD_rvs	5'-GACCATTGTCGACGGAG-3'
CD-Nt_fwd	5'-TAGTGGCGCC**GAAAACCTGTATTTTCAGGGC**CAGAAAGTTATT-GCCCTGACCTTTGATGATGG-3' (KasI restriction site underlined, TEV proteolysis site in bold)
CD-Nt_rvs	5'- TATCCTGCAGCTATTTTGCTTCACGCTG TTTTTTAAC-3' (PstI restriction site underlined)
Mutagenic primers (mutated bases in bold)	
D285S_fwd	5'-CTGACTTT**TC**TGATGGTCCGAATCCGCAA CC-3' and;
D285S_rvs	5'-GGATTCGGACATCA**GA**AAAGGTCAGGGCAATAAC-TTTC-3'.
W402A_fwd	5' – GGATCCGGAAGAT**GCG**AAAGATCGTAACAAAAAAACC - 3'
W402A_rvs	5' – CGATCTTT**CGC**ATCTTCCGGATCCACATCCC - 3'
Y430A_fwd	5' – TGATATT**GCG**CGTACCAGCGCAGATGC - 3'
Y430A_rvs	5' – GCTGGTA**CGC**GCAATATCATGAATCAGAATGGTACG - 3'

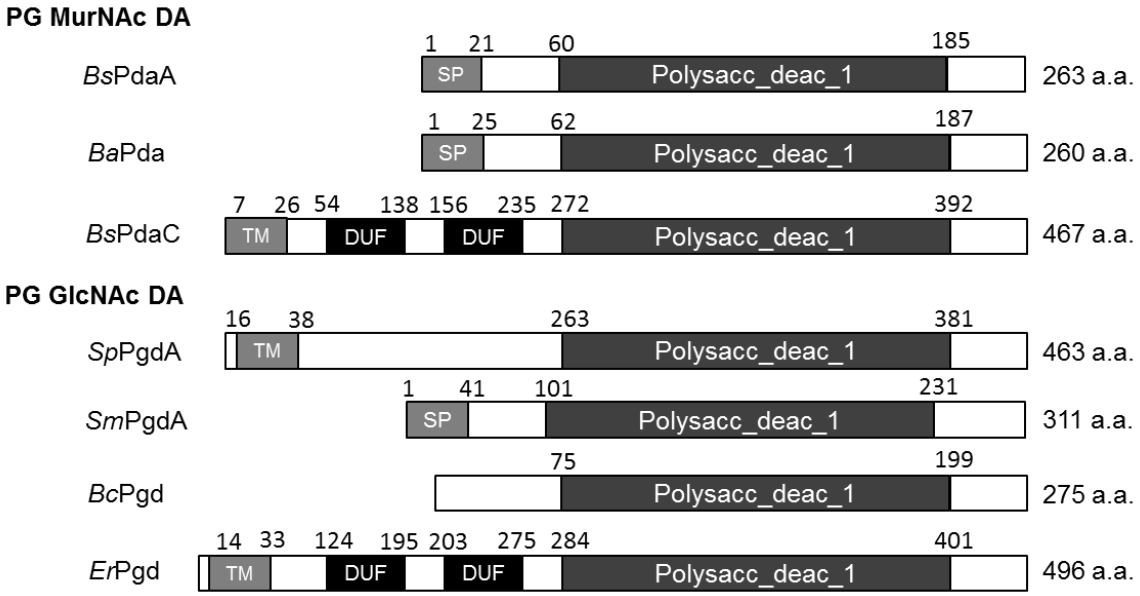


Figure S1. Domains organization of peptidoglycan deacetylases. Numbering according to pfam (22). SP: signal peptide. TM: transmembrane. DUF: domain of unknown function.

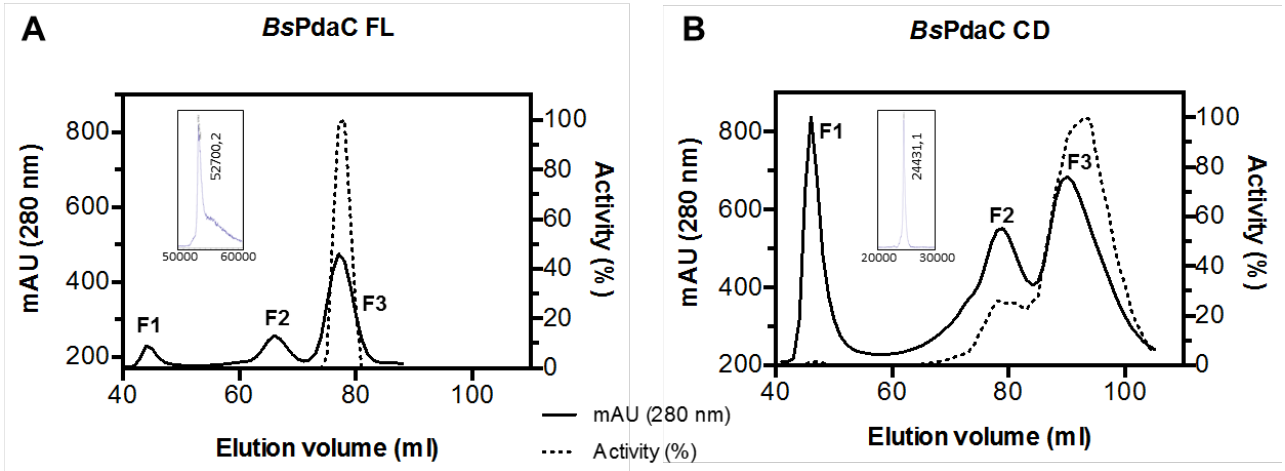


Figure S2. Size exclusion chromatography of purified (A) *Bs*PdaC-FL, (B) *Bs*PdaC-CD. Fractions corresponding to different oligomeric forms: F1 high molecular mass soluble oligomers, F2: dimers,, F3: monomeric form. Elution profile by absorbance at 280 nm (—), activity with AcOMU (⋯). Insets, MALDI-TOF MS of fraction F3.

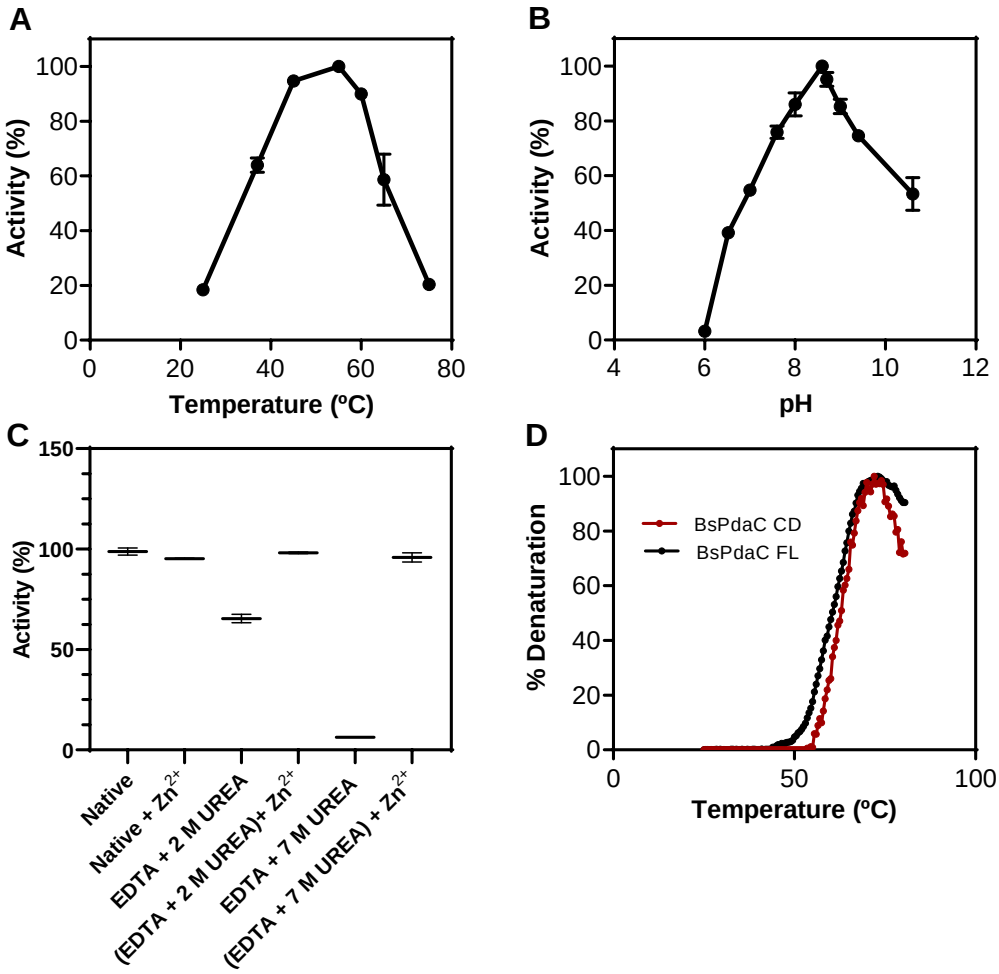


Figure S3. Biochemical characterization of *Bs*PdaC. *A* and *B*. Temperature and pH and profiles of *Bs*PdaC-CD. Activity determined with $(GlcNAc)_5$ in 50 mM phosphate, 300 mM NaCl, at 37°C (pH profile), pH 7.0 (temperature profile). *C.* Metal dependence of *Bs*PdaC activity. Residual activity upon treatment with EDTA (200 mM) and urea (2 or 7 M), and reconstitution with Zn^{2+}cation. *D.* Thermal stability of *Bs*PdaC full length (-FL) and catalytic domain (-CD) by the thermal shift assay with Sypro Orange dye and fluorescence monitoring (λex 483 nm, λem 560 nm).

k_{cat} = 0.93 ± 0.03 s^{-1}

K_M= 6.25 ± 0.47 mM

k_{cat}/K_M = 149 $M^{-1}s^{-1}$

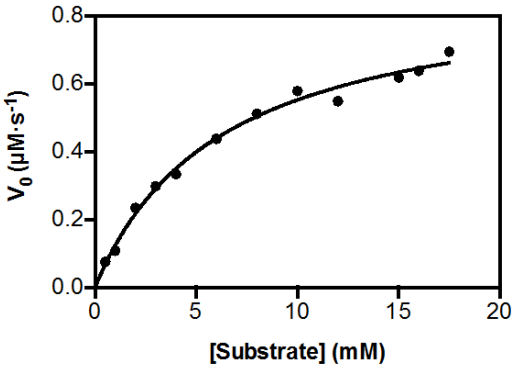


Figure S4. Michaelis-Menten kinetics of *Bs*PdaC-CD with $(GlcNAc)_4$ substrate in 50 mM phosphate, 300 mM NaCl, pH 7.0, 37°C. Initial rates of monodeacetylation were determined by the HPLC-MS method.

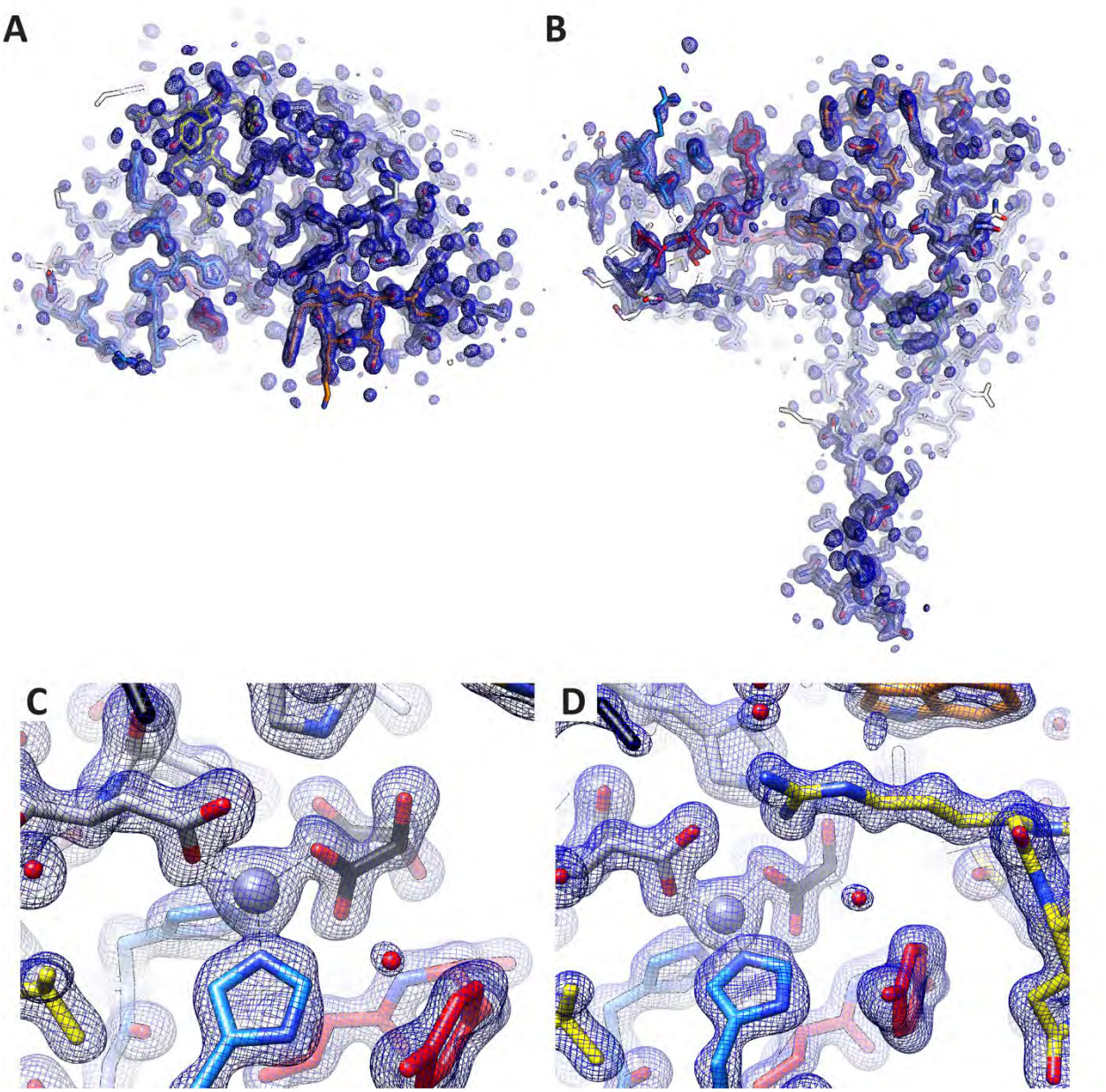


Figure S5. Electron density map of the refined *Bs*PdaC-CD. *A-B*. Two views of the final electron density maps (2mFo-DFc contoured at 1σ) corresponding to the overall structure of *Bs*PdaC-CD. *C-D*. Two views of the final electron density maps (2mFo-DFc contoured at 1σ) corresponding to the active site of *Bs*PdaC-CD.

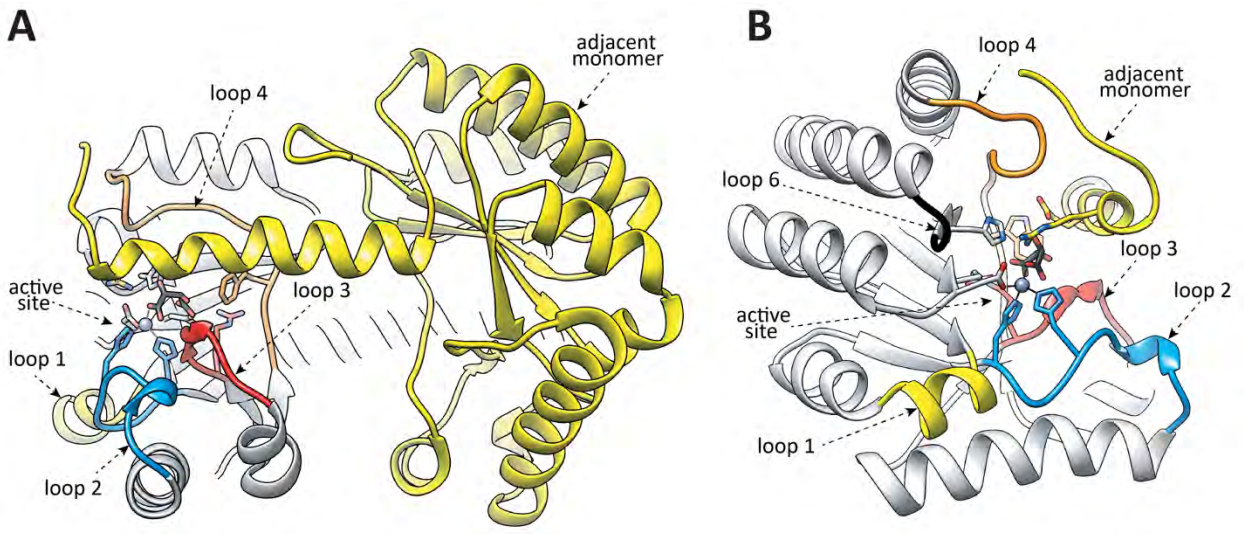


Figure S6. *A.* Crystallographic dimer, one monomer in yellow and the other in grey. *B.* Detail of the dimer interface, in which the side chain of R471 in the C-terminus α-helix from one monomer is inserted into the active site of the adjacent monomer.

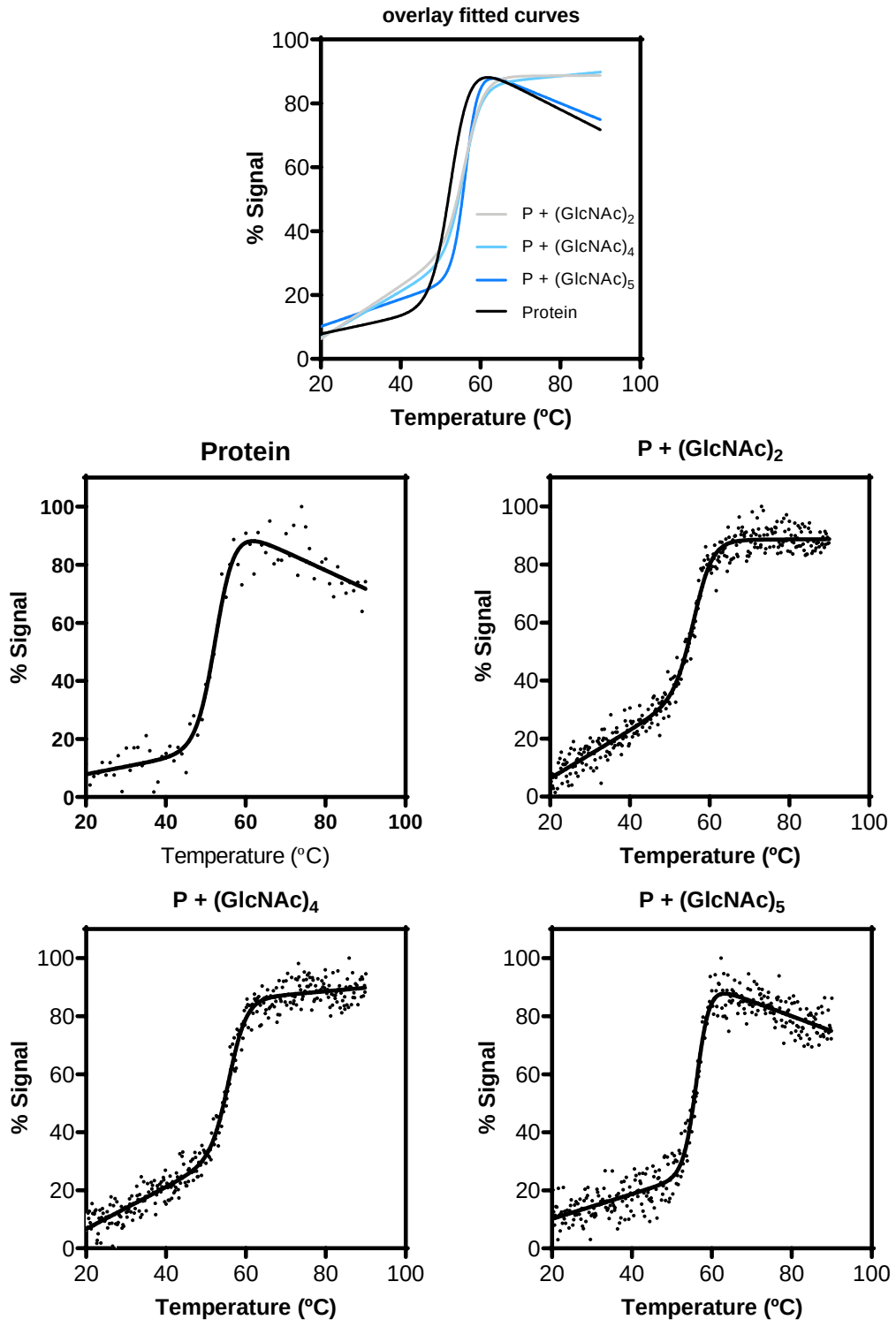
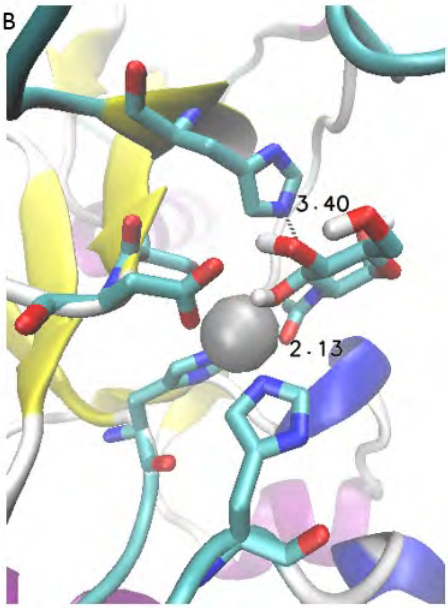


Figure S7. Effect of ligand binding on the thermal stability of *Bs*PdaC-CD-NtStrep by circular dichroism. CD signal data were fitted to a 2-state unfolding model (23):
$\%Signal = ((\alpha_F+\beta_F \cdot T) + (\alpha_U+\beta_U \cdot T) \cdot \exp(m \cdot (T-T_m)) / (1 + \exp(m \cdot (T-T_m))$
where α_F and β_F are the intercept and slope of the base line for the folded protein and α_U and β_U for the unfolded protein, *m* the cooperativity of unfolding and T_m the melting temperature
T_m values: protein: 52.6 ± 0.4 ºC; P+$(GlcNAc)_2$: 56.3 ± 0.2 ºC; P+$(GlcNAc)_4$: 55.9 ± 0.2 ºC; P+$(GlcNAc)_4$: 55.9 ± 0.2 ºC; P+$(GlcNAc)_5$: 56.3 ± 0.2 ºC.

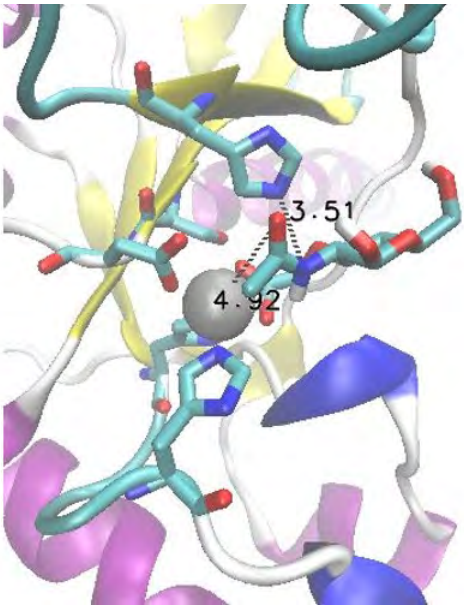
A.1 GlcNAc



C=O···Zn^{2+}= 2.13 Å; NH···His = 3.40 Å

Productive

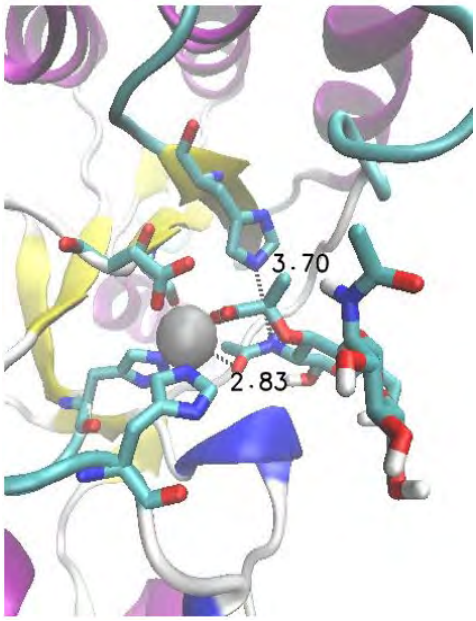
A.2 MurNAc



C=O··· Zn^{2+}= 4.92 Å; NH···His = 3.51 Å

Non productive

B GlcNAc-MurNAc

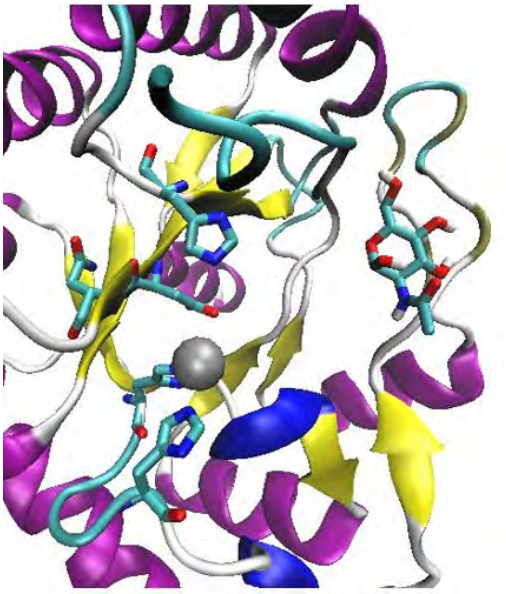


C=O··· Zn^{2+}= 2.83 Å; NH···His = 3.70 Å

Productive, GlcNAc (-1) - MurNAc (0)

Figure S8. Ligand docking simulations on the X-ray structure of *Bs*PdaC. A) Monosaccharides GlcNAc (A1) and MurNAc (A2). B) Disaccharide GlcNAcβ(1→4)MurNAc.

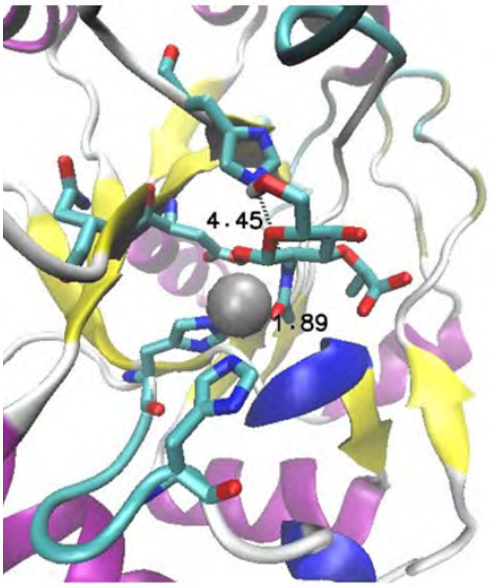
A.1 GlcNAc



C=O··· Zn^{2+}= --; NH···His = --

Non productive

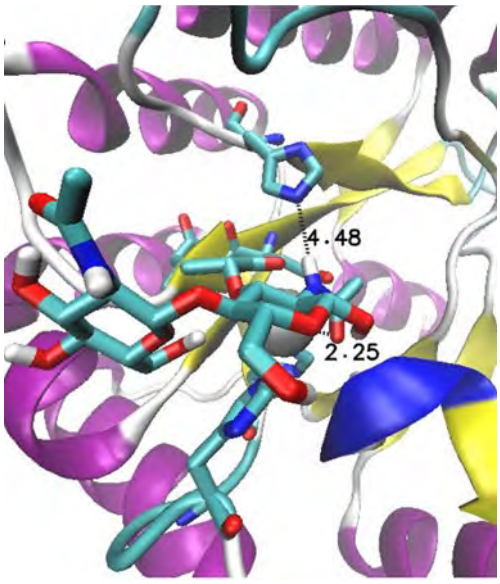
A.2 MurNAc



C=O··· Zn^{2+}= 1.89 Å; NH···His = 4.45 Å

Productive

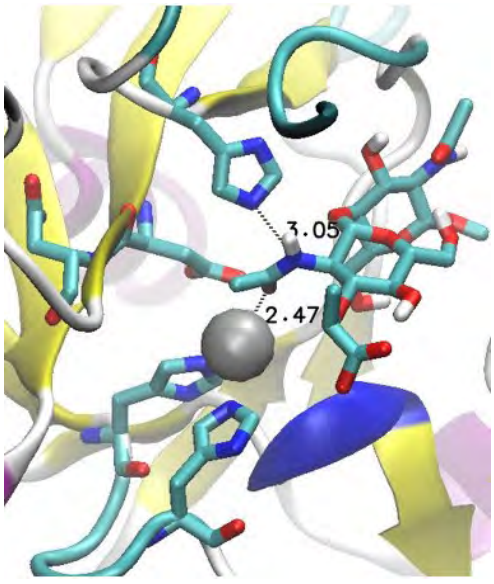
B.1 GlcNAc-MurNAc



C=O··· Zn^{2+}= 2.25 Å; NH···His = 4.482 Å

Productive, GlcNAc (-1) - MurNAc (0)

B.2 MurNAc-*GlcNAc*

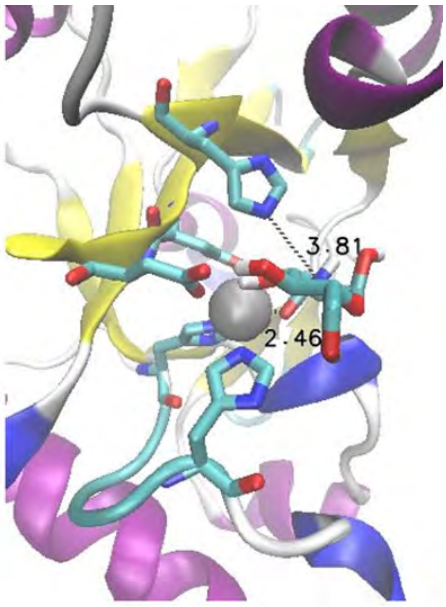


C=O··· Zn^{2+}= 2.47 Å; NH···His = 3.05 Å

Productive, MurNAc (0) - GlcNAc (+1)

Figure S9. Ligand docking simulations on the X-ray structure of *Bs*PdaA MurNAc deacetylase. *A.* Monosaccharides GlcNAc (A1) and MurNAc (A2). *B.* Disaccharides GlcNAcβ(1→4)MurNAc (B1) and MurNAcβ(1→4)GlcNAc (B.2).

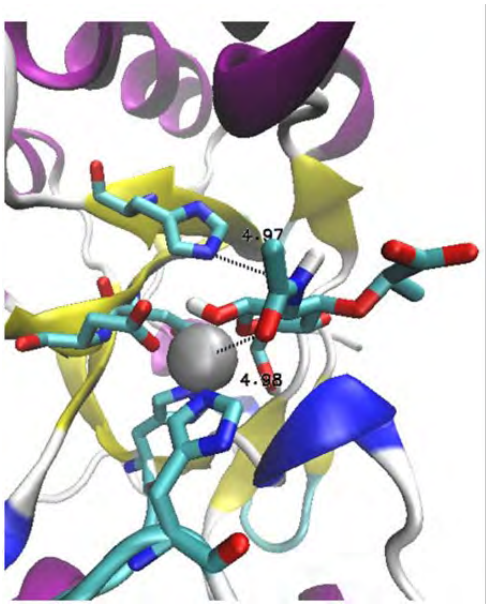
A.1 GlcNAc



C=O··· Zn^{2+}= 2.46 Å; NH···His = 3.81 Å

Productive

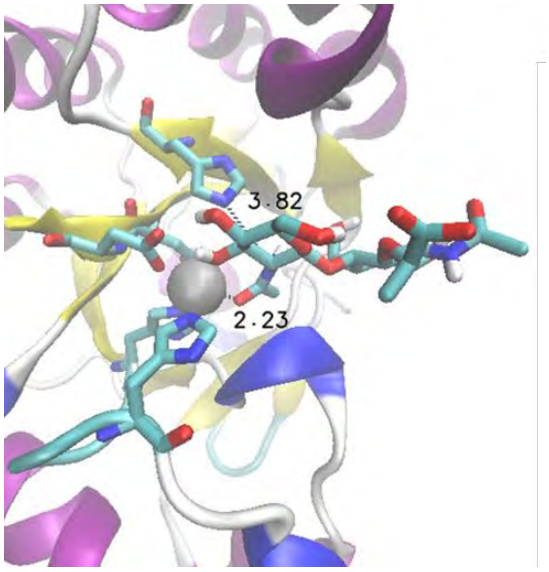
A.2 MurNAc



C=O··· Zn^{2+}= 4.98 Å; NH···His = 4.97 Å

Non productive

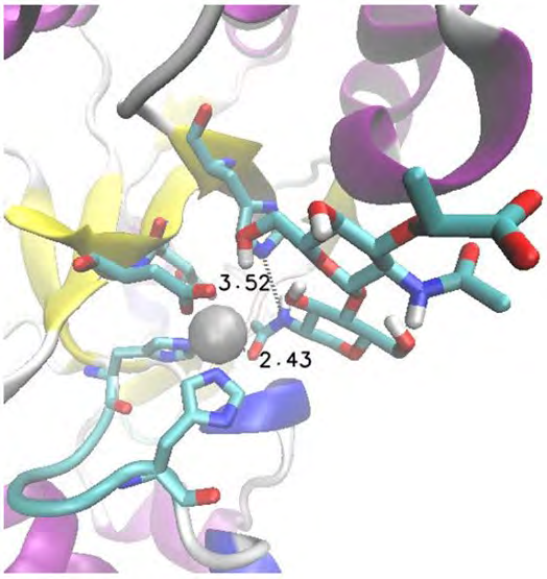
B.1 GlcNAc-MurNAc



C=O··· Zn^{2+}= 2.23 Å; NH···His = 3.82 Å

Productive, GlcNAc (0) - MurNAc (+1)

B.2 MurNAc-GlcNAc



C=O··· Zn^{2+}= 2.43 Å; NH···His = 3.52 Å

Productive, MurNAc (-1) - GlcNAc (0)

Figure S10. Ligand docking simulations on the X-ray structure of *Sp*PgdA GlcNAc deacetylase. *A.* Monosaccharides GlcNAc (A1) and MurNAc (A2). *B.* Disaccharides GlcNAcβ(1→4)MurNAc (B1) and MurNAcβ(1→4)GlcNAc (B.2).

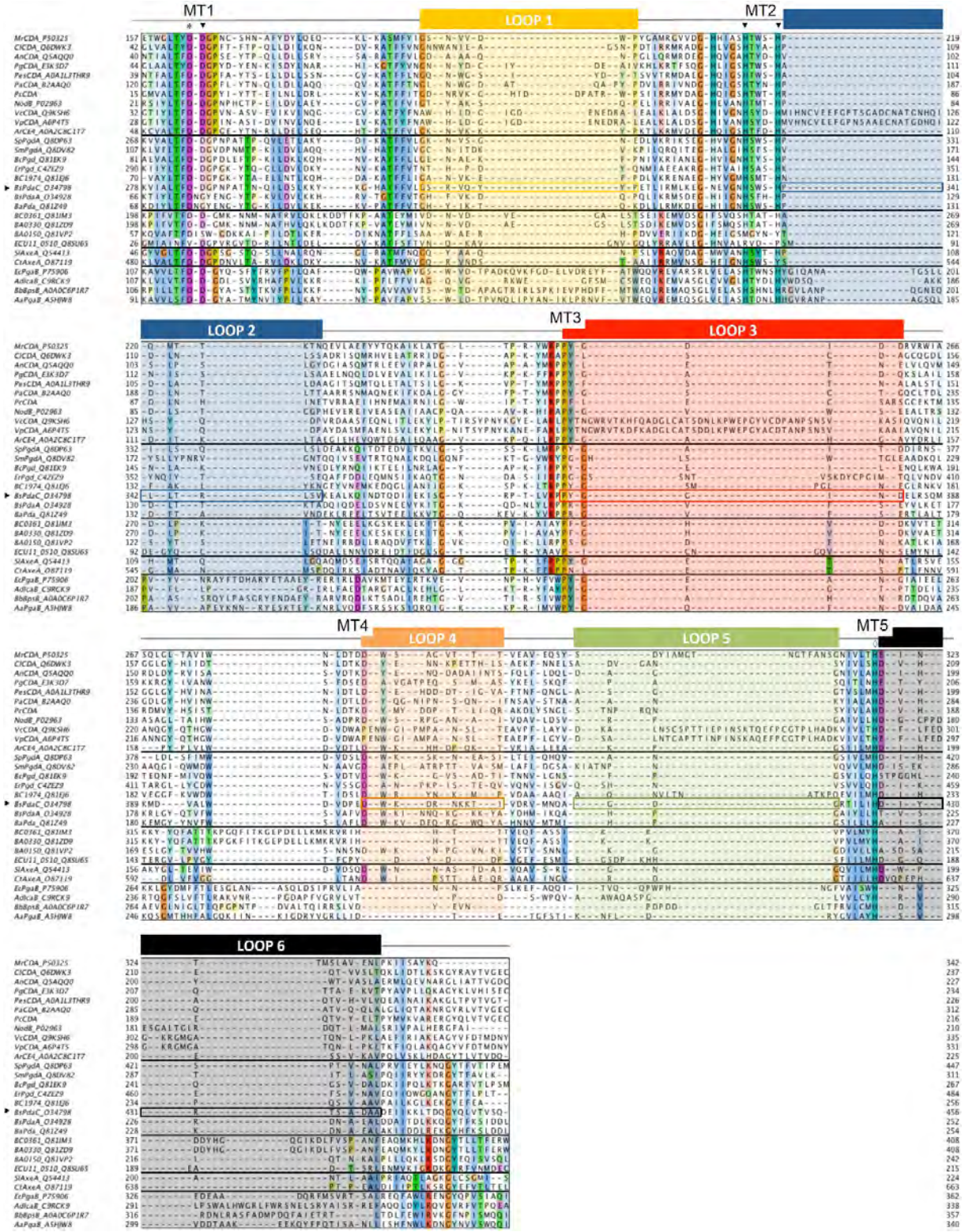


Figure S11. Multiple sequence alignment of characterized CE4 enzymes, grouped by enzyme activity (separated by a line, chitin deacetylases, peptidoglycan deacetylases, annotated polysaccharide deacetylases, acetylxylan esterases, and poly-β-1,6-glucosamine deacetylases). Labels: * general base, ◊ general acid, ▼ metal-binding triad; conserved motifs MT1 (TFDD), MT2 (H(S/T)xxH), MT3 (RxPY), MT4 (DxxD(W/Y), MT5 (I(V/I)LxHD). Loops 1 to 6, colored according to (3).

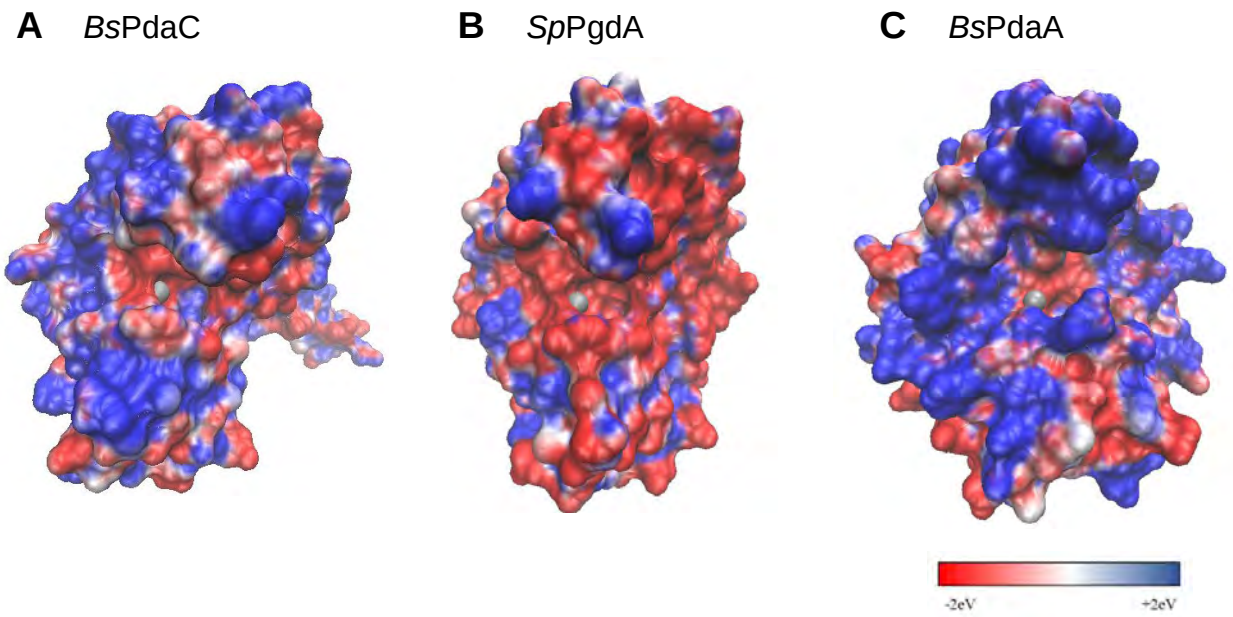


Figure S12. Surface electrostatic potential of *Bs*PdaC (*A*), *Sp*Pgda GlcNAc deacetylase (*B*), and *Bs*PdaA MurNAc deacetylase (*C*).

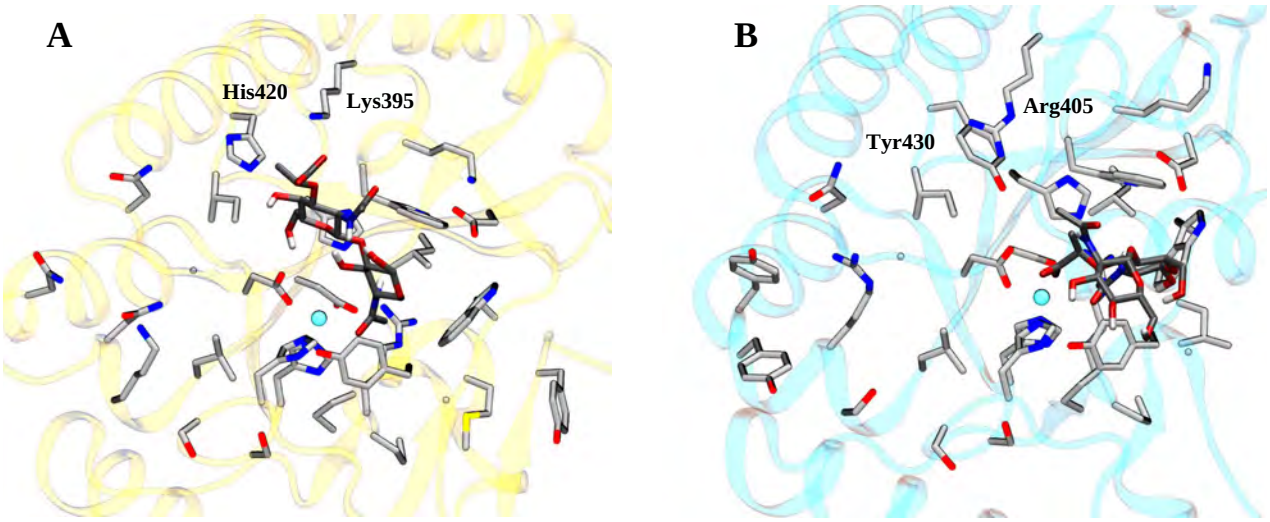


Figure S13. Ligand docking of A) MurNAcβ(1→4)GlcNAc on the X-ray structure of *Sp*PgdA GlcNAc deacetylase (in subsites -1, 0) and B) GlcNAcβ(1→4)MurNAc on the X-ray structure of *Bs*PdaC MurNAc deacetylase (in subsites 0, +1). Solvent-exposed amino acids surrounding the active site are represented. Catalytic center represented by the metal coordination site is in the middle.

REFERENCES

1. Blair, D. E., Hekmat, O., Schüttelkopf, A. W., Shrestha, B., Tokuyasu, K., Withers, S. G., and van Aalten, D. M. (2006) Structure and mechanism of chitin deacetylase from the fungal pathogen Colletotrichium lindemuthianum. *Biochemistry*. **45**, 9416–9426
2. Liu, Z., Gay, L. M., Tuveng, T. R., Agger, J. W., Westereng, B., Mathiesen, G., Horn, S. J., Vaaje-Kolstad, G., van Aalten, D. M. F., and Eijsink, V. G. H. (2017) Structure and function of a broad-specificity chitin deacetylase from Aspergillus nidulans FGSC A4. *Sci. Rep.* **7**, 1746
3. Andrés, E., Albesa-Jové, D., Biarnés, X., Moerschbacher, B. M., Guerin, M. E., and Planas, A. (2014) Structural basis of chitin oligosaccharide deacetylation. *Angew. Chemie - Int. Ed.* **53**, 6882–6887
4. Hirano, T., Sugiyama, K., Sakaki, Y., Hakamata, W., Park, S. Y., and Nishio, T. (2015) Structure-based analysis of domain function of chitin oligosaccharide deacetylase from Vibrio parahaemolyticus. *FEBS Lett.* **589**, 145–151
5. Tuveng, T. R., Rothweiler, U., Udatha, G., Vaaje-Kolstad, G., Smalås, A., Eijsink, V. G. H., and 1 (2017) Structure and function of a CE4 deacetylase isolated from a marine environment. *PLoS One*. **12**, e0187544
6. Blair, D. E., and Van Aalten, D. M. F. (2004) Structures of Bacillus subtilis PdaA, a family 4 carbohydrate esterase, and a complex with N-acetyl-glucosamine. *FEBS Lett.* **570**, 13–19
7. Oberbarnscheidt, L., Taylor, E. J., Davies, G. J., and Gloster, T. M. (2007) Structure of a carbohydrate esterase from Bacillus anthracis. *Proteins Struct. Funct. Genet.* **66**, 250–252
8. Blair, D. E., Schuttelkopf, A. W., MacRae, J. I., and van Aalten, D. M. F. (2005) Structure and metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal virulence factor. *Proc. Natl. Acad. Sci.* **102**, 15429–15434
9. Deng, M. D., Urch, J. E., Ten Cate, J. M., Rao, V. A., Van Aalten, D. M. F., and Crielaard, W. (2009) Streptococcus mutans SMU.623c codes for a functional, metal-dependent polysaccharide deacetylase that modulates interactions with salivary agglutinin. *J. Bacteriol.* **91**, 394–402
10. Tsalafouta, A., Psylinakis, E., Kapetaniou, E. G., Kotsifaki, D., Deli, A., Roidis, A., Bouriotis, V., and Kokkinidis, M. (2008) Purification, crystallization and preliminary X-ray analysis of the peptidoglycan N-acetylglucosamine deacetylase BC1960 from Bacillus cereus in the presence of its substrate (GlcNAc)6. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **64**, 203–205
11. Giastas, P., Andreou, A., Papakyriakou, A., Koutsioulis, D., Balomenou, S., Tzartos, S. J., Bouriotis, V., and Eliopoulos, E. E. (2018) Structures of the peptidoglycan N-acetylglucosamine deacetylase Bc1974 and its complexes with Zinc metalloenzyme inhibitors. *Biochemistry*. **57**, 753–763
12. Fadouloglou, V. E., Kapanidou, M., Agiomirgianaki, A., Arnaouteli, S., Bouriotis, V., Glykos, N. M., and Kokkinidis, M. (2013) Structure determination through homology modelling and torsion-angle simulated annealing: Application to a polysaccharide deacetylase from Bacillus cereus. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 276–283
13. Arnaouteli, S., Giastas, P., Andreou, A., Tzanodaskalaki, M., Aldridge, C., Tzartos, S. J., Vollmer, W., Eliopoulos, E., and Bouriotis, V. (2015) Two putative polysaccharide deacetylases are required for osmotic stability and cell shape maintenance in Bacillus anthracis. *J. Biol. Chem.* **290**, 13465–13478
14. Strunk, R. J., Piemonte, K. M., Petersen, N. M., Koutsioulis, D., Bouriotis, V., Perry, K., and Cole, K. E. (2014) Structure determination of BA0150, a putative polysaccharide deacetylase from Bacillus anthracis. *Acta Crystallogr. Sect. F, Struct. Biol. Commun.* **70**, 156–9
15. Urch, J. E., Hurtado-Guerrero, R., Brosson, D., Liu, Z., Eijsink, V. G. H., Texier, C., and Van Aalten, D. M. F. (2009) Structural and functional characterization of a putative polysaccharide deacetylase of the human parasite Encephalitozoon cuniculi. *Protein Sci.* **18**, 1197–1209

16. Taylor, E. J., Gloster, T. M., Turkenburg, J. P., Vincent, F., Brzozowski, A. M., Dupont, C., Shareck, F., Centeno, M. S. J., Prates, J. A. M., Puchart, V., Ferreira, L. M. A., Fontes, C. M. G. A., Biely, P., and Davies, G. J. (2006) Structure and activity of two metal ion-dependent acetylxylan esterases involved in plant cell wall degradation reveals a close similarity to peptidoglycan deacetylases. *J. Biol. Chem.* **281**, 10968–10975
17. Nishiyama, T., Noguchi, H., Yoshida, H., Park, S. Y., and Tame, J. R. H. (2013) The structure of the deacetylase domain of Escherichia coli PgaB, an enzyme required for biofilm formation: A circularly permuted member of the carbohydrate esterase 4 family. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 44–51
18. Little, D. J., Bamford, N. C., Pokrovskaya, V., Robinson, H., Nitz, M., and Howell, P. L. (2014) Structural basis for the De-N-acetylation of poly-β-1,6-N-acetyl-D-glucosamine in gram-positive bacteria. *J. Biol. Chem.* **289**, 35907–35917
19. Little, D. J., Milek, S., Bamford, N. C., Ganguly, T., Difrancesco, B. R., Nitz, M., Deora, R., and Howell, P. L. (2015) The protein BpsB is a poly-β-1,6-N-acetyl-D-glucosamine deacetylase required for biofilm formation in Bordetella bronchiseptica. *J. Biol. Chem.* **290**, 22827–22840
20. Parthiban, C., Varudharasu, D., Shanmugam, M., Gopal, P., Ragunath, C., Thomas, L., Nitz, M., and Ramasubbu, N. (2017) Structural and functional analysis of de-N-acetylase PgaB from periodontopathogen Aggregatibacter actinomycetemcomitans. *Mol. Oral Microbiol.* **32**, 324–340
21. Cord-Landwehr, S., Ihmor, P., Niehues, A., Luftmann, H., Moerschbacher, B. M., and Mormann, M. (2017) Quantitative mass-spectrometric sequencing of chitosan oligomers revealing cleavage sites of chitosan hydrolases. *Anal. Chem.* **89**, 2893–2900
22. Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., and Bateman, A. (2016) The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–D285
23. Pace, C. N. (1986) Determination and Analysis of Urea and Guanidine Hydrochloride Denaturation Curves. *Methods Enzymol.* **131**, 266–280