

Carrageenan desulfation and depolymerization by the marine isolate
Pseudoalteromonas sp. PS47

by

John Andrew Hettle
BSc (Hons), University of Victoria 2012

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Biochemistry and Microbiology

© John Andrew Hettle, 2018
University of Victoria

All rights reserved. This dissertation may not be reproduced in whole or in part,
by
photocopy or other mean, without the permission of the author.

Supervisory Committee

Carrageenan desulfation and depolymerization by the marine isolate
Pseudoalteromonas sp. PS47

by

John Andrew Hettle
BSc (Hons), University of Victoria 2012

Supervisory Committee

Dr. Alisdair B. Boraston (Department of Biochemistry and Microbiology)

Supervisor

Dr. Francis E. Nano (Department of Biochemistry and Microbiology)

Departmental Member

Dr. John E. Burke (Department of Biochemistry and Microbiology)

Departmental Member

Dr. Tom Fyles (Department of Chemistry)

Outside Member

Abstract

Carrageenans are sulfated polysaccharides found in the cell walls of red algae with 20 – 30 % of the dry weight coming from sulfate esters. The understanding of how heterotrophic bacteria desulfate and depolymerize carrageenan has become a rather arduous endeavor as there are 15 different classes of carrageenan distinguished by the degree of sulfation and the presence or absence of a unique galactose derivative, the 3,6-anhydro-D-galactose. The depolymerization of carrageenan requires the removal of the sulfate substituents, a role fulfilled by sulfatases, which hydrolyze sulfate esters playing a key role in the regulation of sulfation states that determine the function of sulfated biomolecules. Through structural, mechanistic, and sequence-based studies a highly conserved sulfate-binding motif has been identified among sulfatases; however, the molecular determinants for substrate specificity remain largely speculative. Additionally, the largest sulfatase family S1, requires a unique catalytic residue resulting from a post-translationally modified cysteine in order to be functional thus making them difficult to study in vitro. Using a strain of *Pseudoalteromonas* sp. PS47 isolated in the Boraston Lab I show that the depolymerization of carrageenan is dependent on the degree of sulfation and that recognition of the leaving group is the driving force behind S1 specificity. With little information on the recognition of sulfated biomolecules, the X-ray crystal structures of the three sulfatases from PS47; PsS1_19A, PsS1_19B, and PsS1_NC in complex with their biological substrates provides a deeper understanding of how carbohydrate specific sulfatases recognize their cognate substrate and how this recognition of the leaving group can be extended to other S1 sulfatase families. Furthermore, I show that an *exo*-acting glycoside hydrolase (PsGH42) requires desulfation of carrageenan oligosaccharides before it can hydrolyze the β -glycosidic linkage, a new specificity of family 42. This research demonstrates how carrageenan depolymerization is entirely dependent on the functionality and specificity of the sulfatases found within the carrageenan utilization locus.

Supervisory Committee

Dr. Alisdair B. Boraston (Department of Biochemistry and Microbiology)

Supervisor

Dr. Francis E. Nano (Department of Biochemistry and Microbiology)

Departmental Member

Dr. John E. Burke (Department of Biochemistry and Microbiology)

Departmental Member

Dr. Tom Fyles (Department of Chemistry)

Outside Member

Table of Contents

Table of Contents	iv
List of Tables	vi
List of Figures	vii
Acknowledgements	ix
Chapter 1: Introduction	1
1.1 Carbon cycling	1
1.2 Macroalgae-associated microorganisms	2
1.3 Macroalgae: major carbohydrate components of the green, the red, and the brown	4
1.3.1 Unique galactans of red algae	5
1.3.2 Carrageenan	7
1.4 Carbohydrate Metabolism	10
1.4.1 Glycoside hydrolases	10
1.4.2 Glycoside hydrolase subsite nomenclature	13
1.5 Sulfatases	15
1.5.1 Family S1 sulfatases	17
1.5.2 FGly generation and family S1 sulfatase maturation	17
1.5.3 Family S1 sulfatase architecture	20
1.6 Carrageenan degradation	24
1.7 PS47: a marine polysaccharide-degrading strain of <i>Pseudoalteromonas</i> sp.	30
1.7.1 The PS47_2012-2044 gene cluster encodes putative carrageenan-specific enzymes.....	30
1.8 Hypothesis and Research objectives	33
Chapter 2: Materials and Methods	35
Chapter 3: Identification of an <i>endo</i>-4S-κ-carrageenan sulfatase, PsS1_19A	50
3.1 Introduction	50
3.2 Results	53
3.2.1 PsS1_19A is a 4S- κ / ι -carrageenan sulfatase	53
3.2.2 Architecture of an <i>endo</i> -4S- κ / ι -carrageenan sulfatase.....	56
3.2.3 The structure of PsS1_19A in complex with intact κ -carrapentaose.....	60
3.2.4 The recognition of ι -carrageenan by PsS1_19A	63
Chapter 4: Identification of two <i>exo</i>-acting sulfatases from PS47: PsS1_19B and PsS1_NC	79
4.1 Introduction	79
4.2 Results	81
4.2.1 Identification of an <i>exo</i> -4S- κ -carrageenan sulfatase.....	81
4.2.2 The structure of PsS1_19B	83
4.2.3 The structure of PsS1_19B in complex with κ -NC2	87
4.2.4 The recognition of ι -NC4 by PsS1_19B forms an inactive complex.....	90
4.2.5 Identification of a third PS47 sulfatase.....	92
4.2.6 Presence of a dead-end intermediate in PsS1_NC maturation	96

4.2.7 PsS1_NC: an <i>exo</i> -2S-I-carrageenan sulfatase	100
4.2.8 PsS1_NC in complex with κ -NC2 and HEPES	103
4.3 Discussion	105
Chapter 5: The identification of an <i>exo</i>-β-carrabiosidase	112
5.1 Introduction	112
5.2 Results	114
5.2.1 Identification of an <i>exo</i> - β -carrabiosidase	114
5.2.2 Architecture of an <i>exo</i> - β -carrageenase from GH family 42	116
5.2.3 Metal binding site of the GH42 family	119
5.3 Discussion	122
Chapter 6: General discussion	127
6.1 PS47 contains the machinery to desulfate and depolymerize κ - and I-carrageenan	127
6.2 Recognition of the leaving group is critical for S1 sulfatase activity	132
6.2.1 The C-terminal subdomain of family S1 sulfatases plays an intricate role in substrate recognition	132
6.2.2 The recognition of 3,6-anhydro-D-galactose by carrageenan active enzymes	139
References	145

List of Tables

Table 1: Oligonucleotide primer sequences used for gene amplification.....	37
Table 2: Crystallization condition for all X-ray crystal structures presented.....	44
Table 3: X-ray data collection and structure statistics.....	46

List of Figures

Figure 1. Select marine polysaccharides from macroalgae.	6
Figure 2. The smallest neocarrabiose repeating unit of four different classes of carrageenan.	9
Figure 3. Glycoside hydrolase common topologies.	12
Figure 4. Mechanisms employed by glycoside hydrolases.	14
Figure 5. Schematic representation for maturation of prokaryotic sulfatases.	19
Figure 6. Possible mechanism of formylglycine formation.	19
Figure 7. X-ray crystal structures of S1 family sulfatases.	21
Figure 8. Identification of the 10 conserved polar residues among the S1 family sulfatases.	23
Figure 9. Known sulfatases in complex with their biological substrate.	25
Figure 10. Carrageenan depolymerization pathways.	28
Figure 11. Genomic organization of the putative PS47 carrageenan PUL.	32
Figure 12. <i>Pseudoalteromonas</i> sp. U2A growth on ι -NC4.	32
Figure 13. Known carrageenan PULs with characterized components.	52
Figure 14. Activity of PsS1_19A from <i>Pseudoalteromonas</i> sp. PS47.	54
Figure 15. Molecular details driving trimerization of PsS1_19A.	57
Figure 16. Structural architecture of PsS1_19A.	59
Figure 17. PsS1_19A C78S in complex with κ -C5.	62
Figure 18. PsS1_19A C78S in complex with polygeenan of different sulfation patterns.	64
Figure 19. Proposed nomenclature for sugar binding subsites in carbohydrate active sulfatases.	70
Figure 20. BT1596 ^{2S-sulf} and BT4656 ^{6S-sulf} complexes labeled using the proposed carbohydrate active sulfatase sugar binding subsite nomenclature.	72
Figure 21. Residue conservation from S1_19 that are involved in substrate recognition.	74
Figure 22. Comparison of ι -carrageenan double helical structure with the structures of the coordinated κ - ι - κ -NC6 and ι - κ -C4 oligosaccharides.	76
Figure 23. Cartoon schematic of products resulting from PsS1_19A sulfate ester hydrolysis.	78
Figure 24. Known carrageenan PULs with characterized components.	80
Figure 25. Activity of PsS1_19B from <i>Pseudoalteromonas</i> sp. PS47.	82
Figure 26. Molecular details driving dimerization of PsS1_19B.	84
Figure 27. X-ray crystal structure of PsS1_19B.	86
Figure 28. PsS1_19B C77S in complex with κ -NC2.	88
Figure 29. Catalytically inactive complex of PsS1_19B C77S with ι -NC4.	91
Figure 30. Molecular details driving dimerization of PsS1_NC.	94
Figure 31. X-ray crystal structure of PsS1_NC.	95
Figure 32. X-ray crystallographic evidence for the identification of amino acid residue number 84 in PsS1_NC.	98
Figure 33. Formylglycine generation of a dead-end intermediate.	99
Figure 34. PsS1_NC C84S in complex with ι -NC4.	101
Figure 35. PsS1_NC C84S in complex with κ -NC2 and HEPES.	104

Figure 36. Conservation of residues involved in substrate recognition from family S1_19.	108
Figure 37. Cartoon schematic of products resulting from PsS1_19B and PsS1_NC sulfate ester hydrolysis.	111
Figure 38. Known carrageenan PULs with characterized components.	113
Figure 39. Activity of PsGH42 from <i>Pseudoalteromonas</i> sp. PS47.	115
Figure 40. Structural features of BovGH42.	117
Figure 41. Structural comparison of BovGH42, Bca- β -gal, and A4- β -gal.	121
Figure 42. Conservation of residues among the family GH42.	124
Figure 43. Cartoon schematic of products from PsGH42 hydrolysis.	126
Figure 44. Saccharification of κ - and ι - carrageenan by PS47.	128
Figure 45. The importance of C-terminal subdomain residues in S1 family sulfatase substrate recognition.	134
Figure 46. C-terminal subdomain comparison of known X-ray crystal structures from family S1 sulfatases.	137
Figure 47. The molecular interactions of 3,6-anhydro-D-galactose recognition	140
Figure 48. Geometries of carbohydrate-protein CH- π interactions.	143

Acknowledgements

Firstly, I would like to express my deepest gratitude to my supervisor, Professor Boraston for the continuous support throughout my Ph.D research, for his knowledge, motivation, advice, and patience during my time here. Working in Dr. Boraston's Lab has been an honour and a privilege I will not soon forget. I could not have imagined a better advisor, lab, or community for my Ph. D studies. I would also like to thank the rest of my thesis committee: Professor Fran Nano, Professor John Burke, and Professor Tom Fyles, for their insightful comments and encouragement. I thank my fellow labmates, both past and present, for the lengthy discussions and fun research environment they helped to create making this a truly memorable experience. I owe a special thank you to Dr. Craig Robb and Dr. Jo Hobbs for their inspiration and input that contributed immensely to my studies and the improvement of this thesis. Last but certainly not least I would like to thank my family: my parents and my brothers for supporting me throughout my studies and the writing of this thesis.

Chapter 1: Introduction

1.1 Carbon cycling

Approximately one half of all primary production occurs in non-terrestrial environments (Field, 1998). As primary production is the main entry point of organic material into an ecosystem, often it occurs through carbon dioxide (CO₂) fixation by autotrophic organisms, such as chemosynthetic microorganisms, photosynthetic bacteria, and plants or algae. The flux of this energy through these ecosystems is largely dependent on heterotrophic organisms, which acquire their metabolic energy through the consumption and degradation of live or dead organic matter (Azam and Malfatti, 2007; Falkowski, 1998). The control of this net flux of CO₂ between the atmosphere and the ocean is dependent on uptake by autotrophic photosynthesizers, respiration by heterotrophic microorganisms, or the export of dissolved carbon to the oceanic depths. In a broad stroke, the movement of carbon through the carbon cycle can be equated as:

$$\text{Export to the ocean depths} = \text{Primary Production} - \text{Microbial Respiration}$$

For the most part, many of these components of export and primary production are fairly well defined in open ocean areas (Rivkin and Legendre, 2001); however, it has proved quite difficult to measure the contribution from bacterial respiration, and thus its input has been based upon bacterial growth efficiency (Paul and Jonathan, 1998). Bacterial growth efficiency has historically shown large variations (Falkowski, 1998; Paul and Jonathan, 1998), imparting some uncertainty on the flux of carbon through these

ecosystems. It has been reported that temperature and bacterial growth efficiency have an inverse relationship, and that this temperature observation is able to account for 54 % of the variation in bacterial growth efficiency (Rivkin and Legendre, 2001). According to this relationship, the lower the temperature the more efficient bacterial growth; thus cooler locations (such as the polar regions) would experience higher activity of bacterial-based food webs. Bacterial heterotrophs are not the only heterotrophic organisms involved in the global carbon cycle, fungi and archaea also participate; however, bacteria play a crucial role in the movement of carbon as well as occupying a key position in the food web in the colder regions of the globe. The large amount of polysaccharides synthesized by marine algal species, and the heterotrophic microbes that colonize the marine algae and subsequently digest the polysaccharides liberating photosynthetically fixed carbon, represents one of the largest and fastest biotransformations on Earth and a key component in the global carbon cycle (Azam and Malfatti, 2007; Hedges et al., 2001).

1.2 Macroalgae-associated microorganisms

The surfaces of marine eukaryotes, such as invertebrates, microalgae, and macroalgae, offer a unique habitat for colonization by microorganisms. Although microorganisms such as yeasts, fungi, protists, and bacteria in the marine environment have been identified as surface-associated on algae to an abundance of $>1.1 \times 10^8$ microorganisms/cm² (Martin et al., 2014) the majority of reports in the literature pertaining to this algal colonization focus on bacteria, with 10^6 - 10^9 bacteria/cm² of algal surface area. Both gram-positive and gram-negative bacteria have been shown to

colonize macroalgal surfaces (Gontang et al., 2007); however, the latter is the more predominant of the two types with the phyla of *Bacteroidetes* and *Proteobacteria* being the most abundant. That being said, there are still large dynamic changes in the colonizing bacterial population, depending on the algal species in question and time of year. At the level of bacterial genus and species, there is another degree of variation as the macroalgae-specific physicochemical constraints (such as cell wall composition and matrix polysaccharides) and any defense mechanisms will influence the diversity in the distribution of the colonizing populations. The factors that influence the composition of algal-associated bacterial populations are quite diverse. As proposed by Martin et al. (2014), they are: (i) the age and health of the thallus; (ii) algal defense mechanisms; (iii) presence/absence of attractants; and (iv) the season when tissues were sampled. However, the most intriguing influencing factor is the algal polysaccharide composition, which selects for the similar functional genetic profiles of colonizing communities found on different specimens. This suggests that the complement of functional genes, rather than bacterial species, may explain the diversity of the bacterial epibionts (Martin et al., 2014).

The major cell wall polysaccharide composition and identity that contributes to the selection of bacterial populations varies between the three different taxa of algal species: Chlorophyta (green algae), Rhodophyta (red algae), and Phaeophyceae (brown algae). Brown, green, and red macroalgae produce a diverse and complex range of polysaccharides which act as essential components of their cell walls, matrix polysaccharides and, in some cases, storage polysaccharides. As described above, the range of macroalgal polysaccharides plays a crucial role in determining the diversity of

macroalgae-associated microorganisms. Macroalgal polysaccharides constitute a major component of the photosynthetically fixed carbon found in the marine environment and, as such, are an important and abundant source of fixed carbon in the overall global carbon cycle. Interestingly, terrestrial polysaccharides are predominantly lignocellulosic, consisting of recalcitrant cellulose, lignin, and hemicellulose, whereas macroalgal biomass is non-lignocellulosic. Macroalgal polysaccharides are also often decorated with functional groups, such as sulfate esters, carboxylic acids and methyl esters, thus giving rise to a high degree of complexity and diversity. These modifications also impart different biochemical and biomechanical properties on the polysaccharides that aid in its environmental interactions.

1.3 Macroalgae: major carbohydrate components of the green, the red, and the brown

The three main macroalgae species – green algae, red algae, and brown algae – constitute a very diverse group of primitive photosynthetic eukaryotes inhabiting marine coastal ecosystems across the globe. The diversity of algal cell wall composition between these species is vast and, as alluded to previously, can be a major driving force behind the population compositions of macroalgae-associated microorganisms. This photosynthetically fixed carbon represents the dominant carbohydrate of the extracellular matrix of the different algal species (Popper et al., 2011).

Terrestrial plant cell walls contain high percentages of cellulose and other polysaccharides that confer rigidity as they battle gravity and grow higher in order to outcompete neighboring plants for sunlight. The marine environment offers other biotic

and abiotic stresses that marine macroalgae must overcome for survival, such as tidal changes and oceanic currents, which requires a polysaccharide composition allowing for more flexibility. There are some extracellular matrix components that are shared amongst macroalgal species, but the differences in the species-specific polysaccharides are quite drastic: pectins and ulvans are found in green algae; agars, carrageenans, and porphyrans are found in red algae; and alginates and fucoidans are found in brown algae (Figure 1A-C) (Popper et al., 2014). This is by no means an exhaustive list of marine macroalgae polysaccharides; however, it not only shows the diversity and complexity employed by different macroalgal species to interact with their environment, but it also highlights that in order for microorganisms to survive on these complex algal polysaccharides there must be an equally diverse and complex array of carbohydrate degrading enzymes.

1.3.1 Unique galactans of red algae

Red algae constitute the oldest group of marine macrophytes and comprise one of the largest components in terms of biomass and species diversity found within intertidal and littoral areas (Usov, 2011). The composition of red algal cell walls consists of rigid and structural glycans (such as cellulose, mannans and xylans) found within a diverse array of matrix polymers with more flexible characteristics. Of these matrix polymers, the most abundant are the sulfated galactans (Domozych et al., 2011), which are found across a broad range of red seaweeds. Variation in the building blocks of these sulfated galactans are often seasonally dependent (Dawes et al., 1973; Freile-Pelegrín and Robledo, 2008), and the various hydroxyl groups can be methylated,

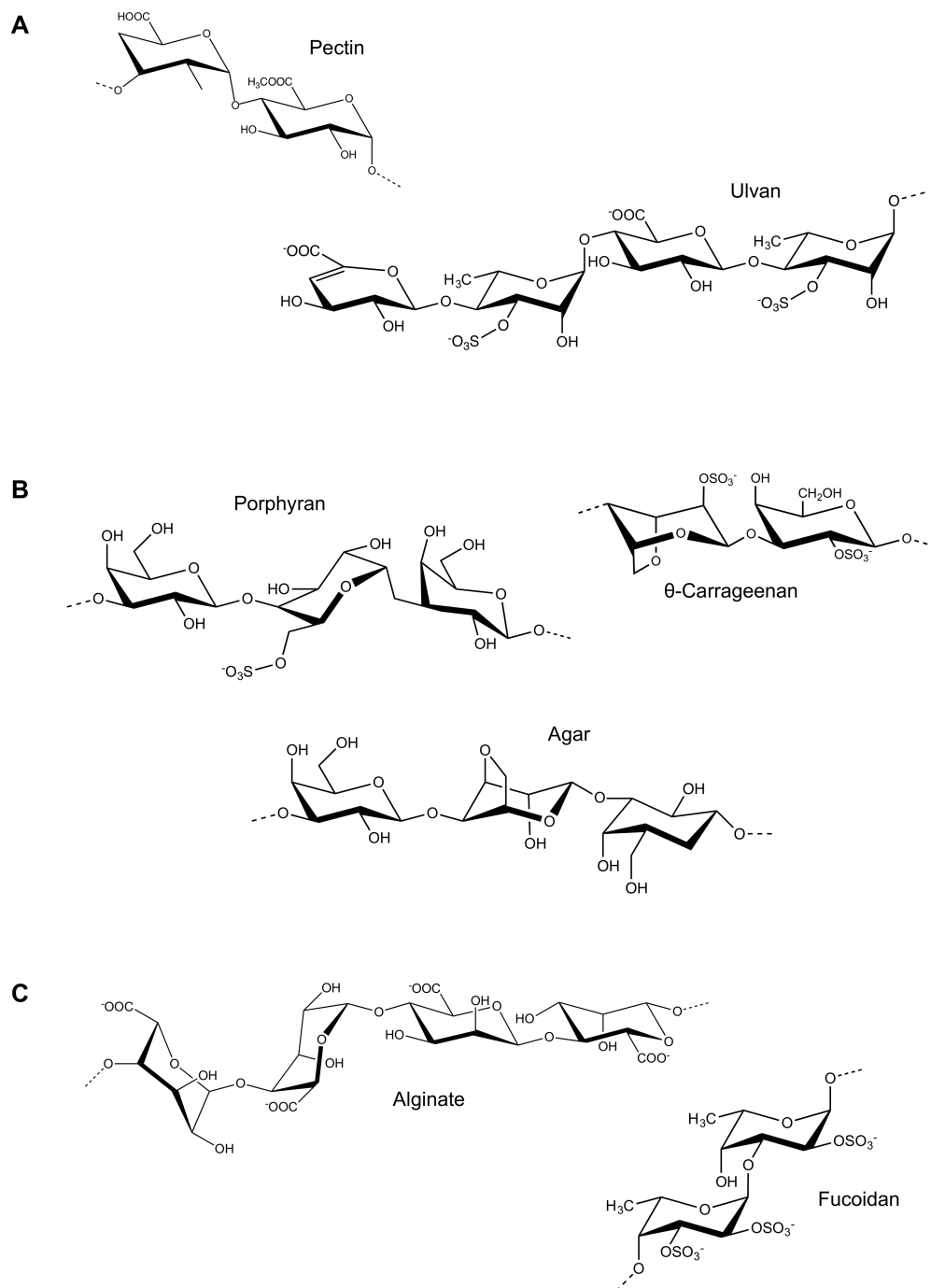


Figure 1. Select marine polysaccharides from macroalgae.

(A) Polysaccharides from green algae.

(B) Polysaccharides from red algae.

(C) Polysaccharides from brown algae.

glycosylated (by single monosaccharides) (Craigie and Jurgens, 1989; Karamanos et al., 1989), or, in some cases, a pyruvic acid derivative can be attached (Painter, 1983). The main class division of these matrix polysaccharides is dependent on whether the 4-linked α -galactose residue is of the L-series (agar group) or of the D-series (carrageenan group). This stereochemical difference confers differences in the physicochemical properties of these polysaccharides, leading to significant industrial applications in both the pharmaceutical and food industries.

1.3.2 Carrageenan

Carrageenans and agarose share many structural and chemical characteristics; however, the degradation of agarose by microorganisms is quite well understood (Hehemann et al., 2012, 2014; Pluvinau et al., 2013) but little is known about degradation and utilization of the different carrageenan classes. Carrageenans are unbranched high molecular weight polymers whose backbone is comprised of alternating α -1,3 linked β -D-galactopyranose and β -1,4 linked α -D-galactopyranose residues. The α -D-galactose (G) residue may exist as a 3,6-anhydro-D-galactose derivative (DA), and varying degrees of sulfate esters can decorate either galactose unit within the disaccharide building blocks. These sulfate esters can be found on carbon-2 (C2), carbon-4 (C4) or carbon-6 (C6) of the β -D-galactose unit (G2S, G4S or G6S, respectively) and with various combinations of the three, and also in some cases on C2 of the 3,6-anhydro-D-galactose unit (DA2S). Based on these characteristics it has been theorized that there are 42 possible repeating structures for carrageenan (Stortz and Cerezo, 1992); however, only 15 classes are naturally observed. These classes are

based on the degree and pattern of sulfation, and the presence or absence of a DA unit within the underlying disaccharide structure. Some attempts have been made to group the different classes of carrageenan into four major clans based upon the sulfation state of the 3-linked β -D-galactose residue: unsulfated (beta [β]-carrageenan), or bearing sulfate esters on C2 (lambda [λ]-carrageneen), C4 (kappa [κ]-carrageen) or C6 (omega [ω]-carrageenan). This naming system has not been fully adopted in the literature, as the high occurrence of hybrid structures of both the β - and κ -carrageenan families found in the full polymer and small oligosaccharides make classification by this nomenclature confounding.

More recently, Knutsen et al. (1994) proposed a nomenclature based upon the 4-linked α -D-galactose residue of the repeating disaccharide block: carrageenan, when the residue is in the pyranose form (Knutsen et al., 1994); and carrageenose, when the 3,6-anhydro derivative is present. This system is similar to the nomenclature accepted for agars. Finally, as also used in the naming of agar oligosaccharides, the linkage found at the non-reducing end of carrageenan oligosaccharides distinguishes between carrageenan (C; β -glycosidic linkage) and neo-carrageenan (NC; α -glycosidic linkage) oligosaccharides. Examples of these naming systems and any related derivatives after carrageenan remodeling are shown in Figure 2. An abbreviation to describe specific carrageenan oligosaccharides will be used, for example κ -neocarrabiose, κ -carrabiose and ι -neocarragbiose will be referred to as κ -NC2, κ -C2, and ι -NC2, respectively (Figure 2E), this abbreviation system can be applied to longer carrageenan oligosaccharides.

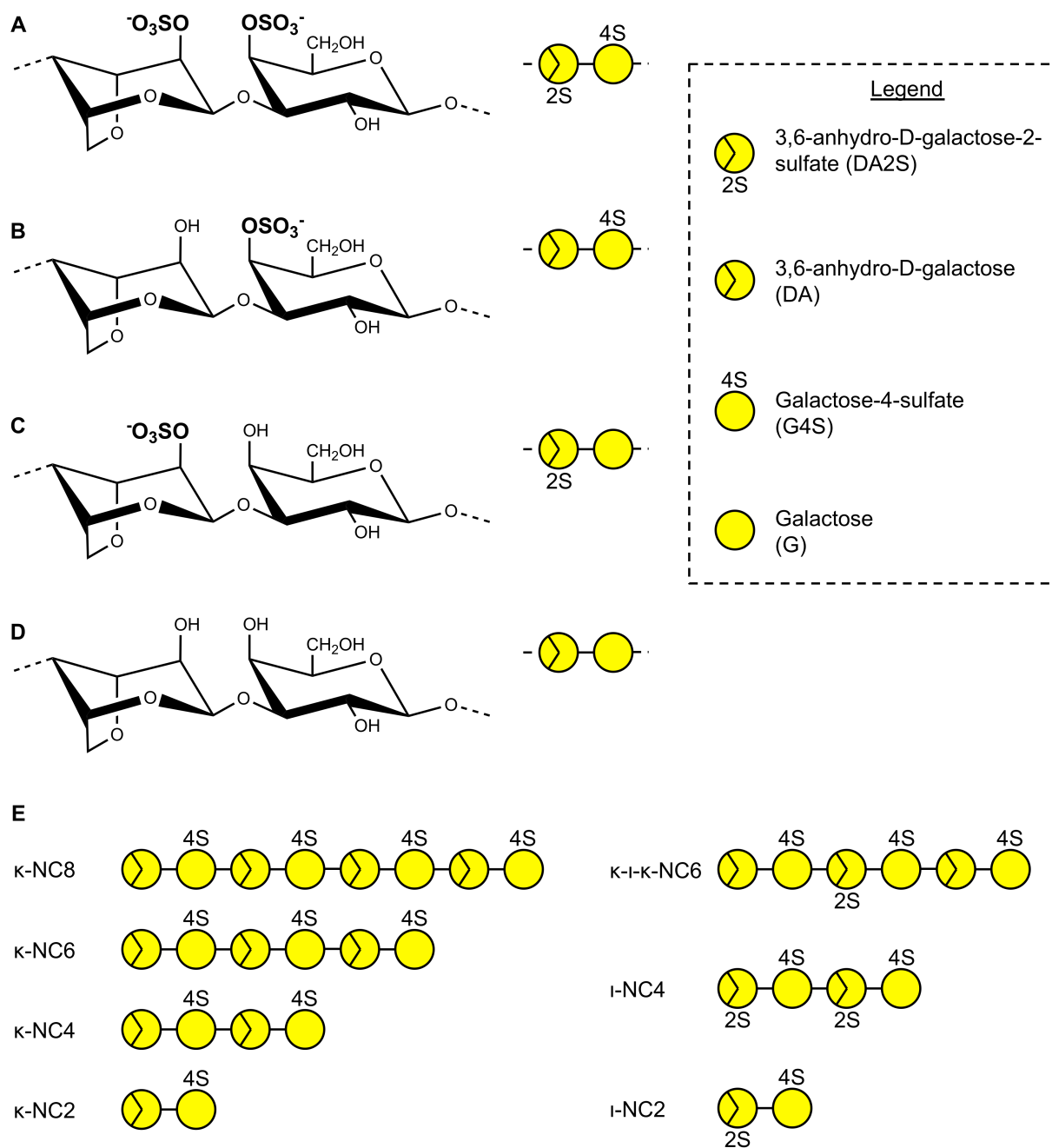


Figure 2. The smallest neocarrabiose repeating unit of four different classes of carrageenan.

(A) ι-carrageenan

(B) κ-carrageenan

(C) α-carrageenan

(D) β-carrageenan

(E) Example schematic and nomenclature of carrageenan oligosaccharides

1.4 Carbohydrate Metabolism

Carbohydrates constitute the most abundant compound on the globe and are found as monosaccharides, disaccharides, oligosaccharides (composed of three to nine monosaccharide residues), and polysaccharides. The diversity in the composition of higher order carbohydrate structures means that they play diverse roles in nature, including nutrient storage, structural components of cell walls, and can be involved in cell-to-cell communication (Dhoot et al., 2017). Many organisms on earth have dedicated large portions of their genomes to proteins and enzymes involved in carbohydrate anabolism, catabolism, and transport. The anabolic component of the carbohydrate active enzymes (CAZymes) found in red algae and other auxotrophs that are used to synthesize these functional carbohydrates, such as glycosyl transferases (GTs), have a catabolic counterpart in heterotrophs, namely glycoside hydrolases (GHs), polysaccharide lysases (PLs) and other CAZymes involved in the hydrolysis of carbohydrate functional group decorations such as methyl groups, carboxylic acids, and sulfate esters. These catabolic CAZymes are the tools utilized in the saccharification of marine carbohydrates, reducing the polymeric structure to its monomeric components, allowing for the resulting monosaccharides to enter metabolism and ultimately liberate the photosynthetically-fixed carbon through catabolic processes.

1.4.1 Glycoside hydrolases

The enzyme super families of GHs and PLs cleave the glycosidic bond, via hydrolysis or β -elimination, respectively, linking a monosaccharide to lipids, proteins, or other monosaccharides and carbohydrates. They are found in all domains of life and are

involved in a diverse array of biological processes. Currently there are 28 PL families and 153 GH families consisting of 15,401 and 538,416 modules classified based on amino acid sequence similarities, respectively, in the CAZy database (www.CAZy.org) (Lombard et al., 2014), indicating GHs to be by far the most prevalent catabolic CAZyme. Although GHs are classified based on sequence similarities, different families often share common structural elements and catalytic mechanism. Because of this, the establishment of an additional classification system based on conserved secondary structural motifs and conserved catalytic machinery has been accepted, organizing GHs into 14 different clans (clans A-N) (www.CAZy.org). The majority of GHs possess a TIM barrel or a $(\beta/\alpha)_8$ fold; however, β -propeller, β -jelly, and $(\alpha/\alpha)_6$ organizations are also commonly observed.

In addition to the family and clan classifications of GHs, there are further sub-classifications depending on their mode of action: *endo*-acting versus *exo*-acting, and the mechanism resulting in either retention or inversion of the anomeric configuration (Davies and Henrissat, 1995; Koshland, 1953). In *endo*-acting GHs, the glycosidic linkage being cleaved is an internal linkage of the polymer; this mode of action is typically seen in enzymes containing an active site groove or tunnel (Figure 3A). *Exo*-acting GHs require molecular interactions with either the non-reducing end or the reducing end of the polymer and they cleave the terminal glycosidic linkage. This mode of action is often seen in GHs containing an active site pocket (Figure 3B). The retaining or inverting mechanisms can be utilized by both *endo*- and *exo*-acting enzymes. Historically, it was believed that the distance between the catalytic residues was indicative of the mechanism utilized by the enzyme, e.g. ~ 5.5 Å for retaining

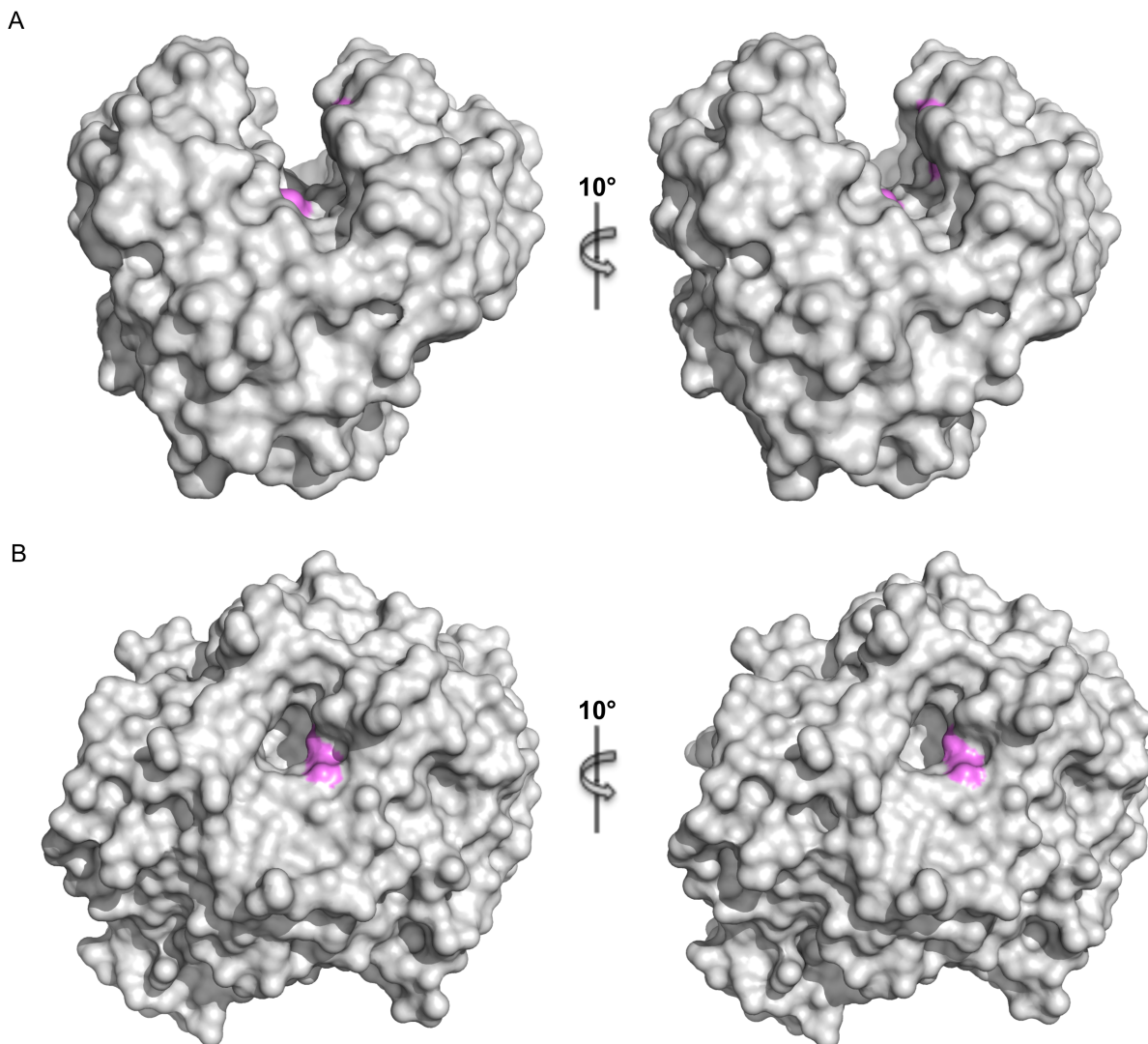


Figure 3. Glycoside hydrolase common topologies.

Solvent accessible surfaces of a GH6 (A) from *Thermobifida fusca* (PDB ID: 4AVN) (Wu et al., 2013) and a GH15 (B) from *Aspergillus Awamori* (PDB ID: 1AGM) (Aleshin et al., 1994) highlighting the structural differences between *endo*- and *exo*-acting enzymes, respectively. In both panels the solvent accessible surface is shown in gray with the catalytic residues coloured violet.

mechanisms and ~ 10 Å for inverting mechanisms (Davies & Henrissat, 1995; McCarter & Stephen Withers, 1994), these values stemmed from the spatial arrangement of the catalytic residues to each other and to the glycosidic linkage. However, this no longer holds explicitly true, as the relevant catalytic residues have been observed in perpendicular orientations to the targeted glycosidic linkage (Br  x et al., 2006), and thus the only true identifier of mechanism is the configuration of the anomeric carbon (Figure 4A and 4B).

1.4.2 Glycoside hydrolase subsite nomenclature

Polysaccharides exhibit huge structural diversity based upon the stereochemistry of the glycosidic linkage (α - or β -), the respective carbons involved in the linkage from each contributing monosaccharide, the identity of the monosaccharides involved, the number of monosaccharide units in the polymer, and how they are subsequently linked together. This leads to a massive number of possible carbohydrate structures as evidenced by the theoretical possibility of 10^{12} isomers for a reducing hexasaccharide (Laine, 1994). Because of this diversity GH active sites contain subsites that employ different molecular specificities in order to discriminate between different glycosidic linkages and monosaccharides. These subsites coordinate individual monosaccharides within the larger polymer, allowing for a higher degree of selectivity between carbohydrate structures (Davies et al., 1997). Moreover, a subsite nomenclature aids in the ease of communicating structure-function relationships. GH subsites are numbered depending on whether the sugar residues occupying them are on the reducing or non-reducing end with respect to the linkage being cleaved. They are numbered as

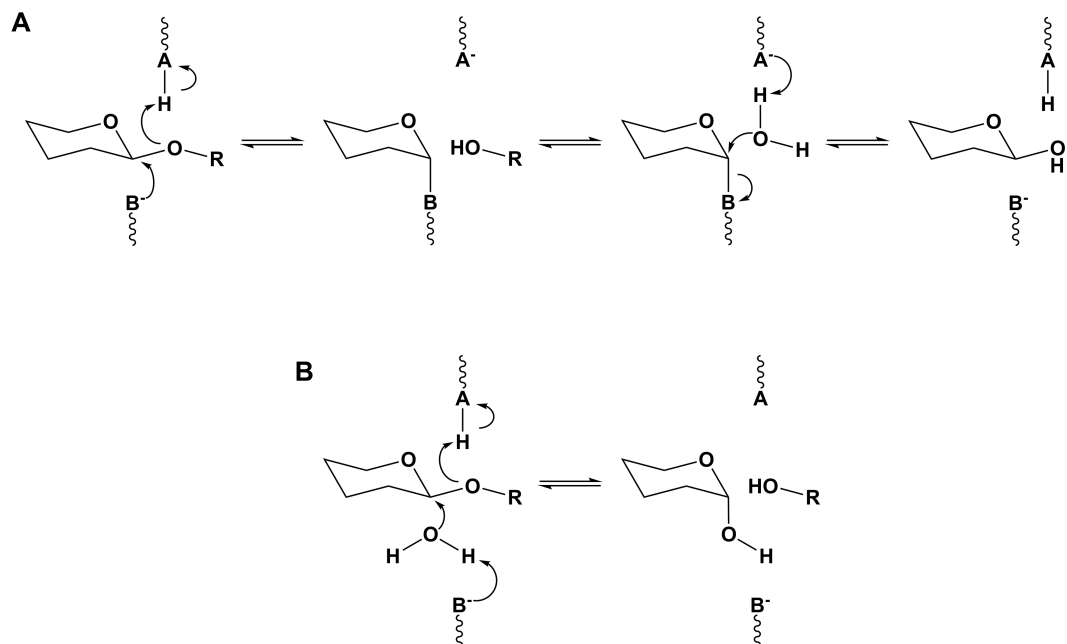


Figure 4. Mechanisms employed by glycoside hydrolases.
(A) Retaining mechanism
(B) Inverting mechanism

increasingly positive towards the reducing end (+1, +2, +3, etc.) and increasingly negative towards the non-reducing end (-1, -2, -3, etc.) (Davies et al., 1997).

The large number of GH and PL families highlights the complexity and diversity observed in carbohydrate structures. As mentioned previously, polysaccharides found in the marine environment often bear decorations, which have an impact on the functionality of GHs and PLs and can contribute to the molecular requirements for recognition. These functional groups are either required for recognition by GHs and PLs, or they must be removed in order for accessibility to the glycosidic linkage. This type of dependency is observed in the depolymerization of pectin, a polygalacturonate, which often comes acetylated and methylated. In *Erwinia* spp. the activity of pectin specific PLs was dependent on the deacetylation and demethylation of pectin (Abbott and Boraston, 2008). Pectin is not exclusive to the marine environment; however, it can serve as an example for the hypothesis that depolymerizing enzymes are dependent on the state of functional group decorations.

1.5 Sulfatases

Sulfatases constitute a class of enzyme that are essential in catalyzing the hydrolytic cleavage of sulfate esters (ROSO_3^-) and sulfamates (RN(H)SO_3^-), and they often play major roles in the regulation and degradation of sulfated biomolecules in biological systems. Sulfatases have historically received little research attention largely due to the difficulty in identifying physiological substrates, as well as the existence of a very limited set of biochemical tools to assess the efficiency of catalysis. The identification of sulfatases within genomes has become possible, given the high degree

of sequence similarity between sulfatase signature motifs and the frequent observation of activity against synthetic substrates; however, functional characterization of sulfatases is still extremely challenging and poses the greatest hurdle to sulfatase research. The recent surge in sulfatase interest is mainly due to the appreciation of the roles they play in eukaryotic systems, such as hormone regulation (Prost et al., 1984), cell signaling (Dhoot et al., 2017), and glycosaminoglycan modifications (Bielicki and Hopwood, 1991). However, appreciation for the importance of sulfatases in prokaryotic systems has started to gain some traction with an increased need for deeper understanding of the role they play in bacterial pathogenesis, sulfate scavenging, and polysaccharide biomass turnover.

The growing interest in sulfatases has led to a system for amino acid sequence-based classification into four major families (Barbeyron et al., 2016) consisting of over 35,000 putative sulfatases as of April 2018. The largest of these families is the S1 family composed of 30,727 putative sulfatases, which are characterized by the need for post-translational modification of a cysteine or a serine into a unique C α -formylglycine residue (FGly) that is used to catalyze the hydrolysis of sulfate esters (Hanson et al., 2004). This family is further broken down into 72 subfamilies based on sequence similarities, of which only 11 subfamilies have a known activity. The other three remaining families are the S2 family (1,431 entries), members of which belong to the non-heme iron (II) α -ketoglutarate-dependent dioxygenase superfamily; the family S3 (2,838 entries) encompassing only alkyl sulfatases; and the S4 family (95 entries) whose members all contain the hexapeptide signature sequence (T/S)-H-X-H-X-D

which forms a catalytic zinc-binding site as found in metallo- β -lactamases (Davison et al., 1992; Hagelueken et al., 2006).

1.5.1 Family S1 sulfatases

Across eukaryotic, prokaryotic and archael species, S1 sulfatases have a high degree of sequence, structure and mechanistic conservation (Hanson et al., 2004). This high degree of sulfatase sequence similarity makes for their easy bioinformatic identification based upon the highly conserved pentapeptide signature sequence (C/S-X-P-X-R), which is accepted as the S1 signature sulfatase motif. This motif is found within a longer conserved 12 amino acid sequence (C/S-X-P-S/X-R-X-X-X-L/X-T/X-G/X-R/X) (Dierks et al., 1999) that is necessary for directing the machinery responsible for the post-translational modification of the proto-catalytic cysteine or serine residue into the family S1 catalytic FGly residue (Knaust et al., 1998). As such, family S1 sulfatases can be split into Cys-type or Ser-type sulfatases (Knaust et al., 1998). The maturation of the catalytic FGly residue employs one of several specific post-translational modification systems.

1.5.2 FGly generation and family S1 sulfatase maturation

There are currently two known and one hypothesized but unidentified prokaryotic systems responsible for the post-translational modification of Cys-type and Ser-type sulfatases into matured FGly-sulfatases. Under anaerobic conditions, the oxygen-independent anaerobic sulfatase-maturing enzyme (anSME) is able to mature both Cys- and Ser-type sulfatases (Berteau et al., 2006), whereas in an aerobic environment the

machinery involves an oxygen-dependent formylglycine-generating enzyme (FGE) that is specific for the maturation of Cys-type sulfatases only (Dierks et al., 2003). It has been shown that *Escherichia coli* possesses an additional, unidentified maturation system that differs from the FGE system but is also specific for the maturation of Cys-type sulfatases (Figure 5) (Benjdia et al. 2007).

As sulfatase maturation is a necessity in order to observe catalytic function, research has focused on understanding the biochemistry behind the post-translational maturation processes. For the FGE maturation system, structural work has elucidated the utilization of a cysteine disulfide bridge as a major component in FGly generation (Carlson et al., 2008). A proposed mechanism utilizes an intramolecular disulfide bridge in the FGE, which undergoes an interchange with the C-X-P-X-R recognition sequence of the targeted unmaturing S1 sulfatases creating an intramolecular disulfide intermediate (Figure 6) (Dierks et al., 2005). In this proposed monooxygenase mechanism, there is the generation of a cysteinesulfenic acid intermediate of the FGE cysteine that does not participate in the intermolecular disulfide bridge. The hydroxyl group from the cysteinesulfenic acid is transferred to the sulfatase substrate cysteine residue causing a regeneration of the intramolecular disulfide bridge releasing the sulfatase. The substrate sulfatase cysteinesulfenic acid residue then spontaneously eliminates the hydroxyl group creating a thioaldehyde intermediate, which then spontaneously hydrolyzes to form the FGly residue (Figure 6). Through X-ray crystal structure and biochemical analysis, the proposal of this monooxygenase mechanism is the currently accepted model as the activity is shown to be dependent on an external reducing factor; however, there is the acknowledgment of a possible oxidase

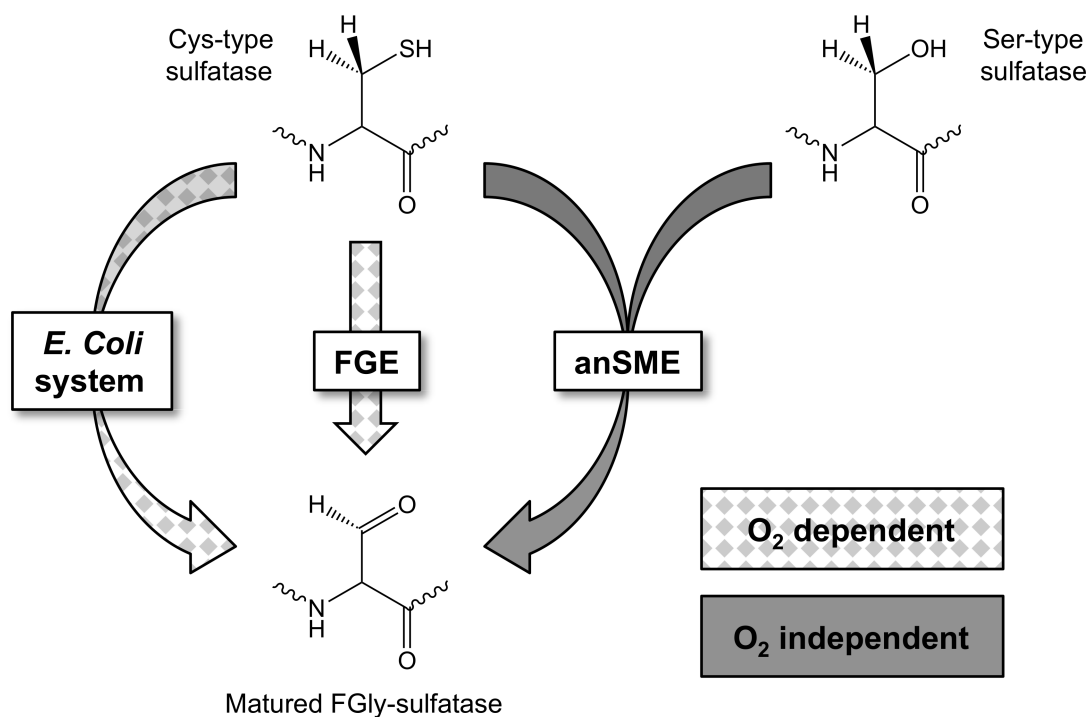


Figure 5. Schematic representation for maturation of prokaryotic sulfatases. Anaerobic sulfatase maturing enzymes (anSME) are able to mature both Cys-type and Ser-type sulfatases in an oxygen-independent manner. In an oxygen-dependent manner there are two non-homologous systems: the formylglycine generating enzyme (FGE) and an unknown system present in *E. coli* (*E. coli* system), which are only able to mature Cys-type sulfatases. (adapted from Benjida et al. 2007).

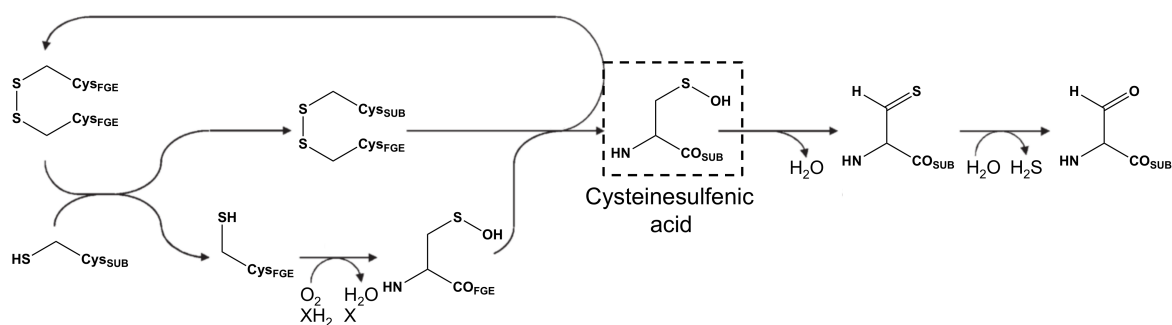


Figure 6. Possible mechanism of formylglycine formation. Possible mechanism proposed by Dierks et al. (2005) utilizing cysteinesulfenic acid as a mechanistic intermediate (adapted from Dierks et al. (2005))

mechanistic model that produces a cysteinesulfinic acid intermediate (-SOOH) (Peng et al., 2015). This model would ultimately involve the elimination of H_2O_2 due to the necessity for an internal reducing factor such as O_2 , yet without the detection of this intermediate the oxidase model remains largely speculative.

1.5.3 Family S1 sulfatase architecture

The fold adopted by members of the S1 family has been known since the late 1990s and consists of a spherical globular monomer of mixed α/β topology. This is divided into two subdomains: the large N-terminal subdomain and the smaller C-terminal subdomain (Figure 7A) (Boltes et al., 2001; Bond et al., 1997; Cartmell et al., 2017; Hernandez-Guzman et al., 2003; Lukatela et al., 1998; Rivera-Colón et al., 2012; Sidhu et al., 2014). This fold is conserved in all S1 subfamilies, which have representation from all domains of life. The larger N-terminal subdomain consists of solvent exposed α -helices surrounding a meandering β -sheet of which the number of strands composing this can vary between sulfatases. The smaller C-terminal subdomain consists of a four stranded anti-parallel β -sheet surrounded by solvent exposed α -helices (Figure 7B). The four-stranded anti-parallel β -sheet is highly conserved; however, the number and placement of the solvent exposed loops and α -helices shows some variability. It has been proposed that the major contributor to sulfatase substrate discrimination is the C-terminal subdomain of the S1 family as it has the highest degree of structural variation amongst S1 sulfatases (Figure 7C) (Hanson et al., 2004).

Through X-ray crystallographic and mutagenic studies of family S1 sulfatases, 10 highly conserved, polar residues were initially identified leading to the hypothesis of a

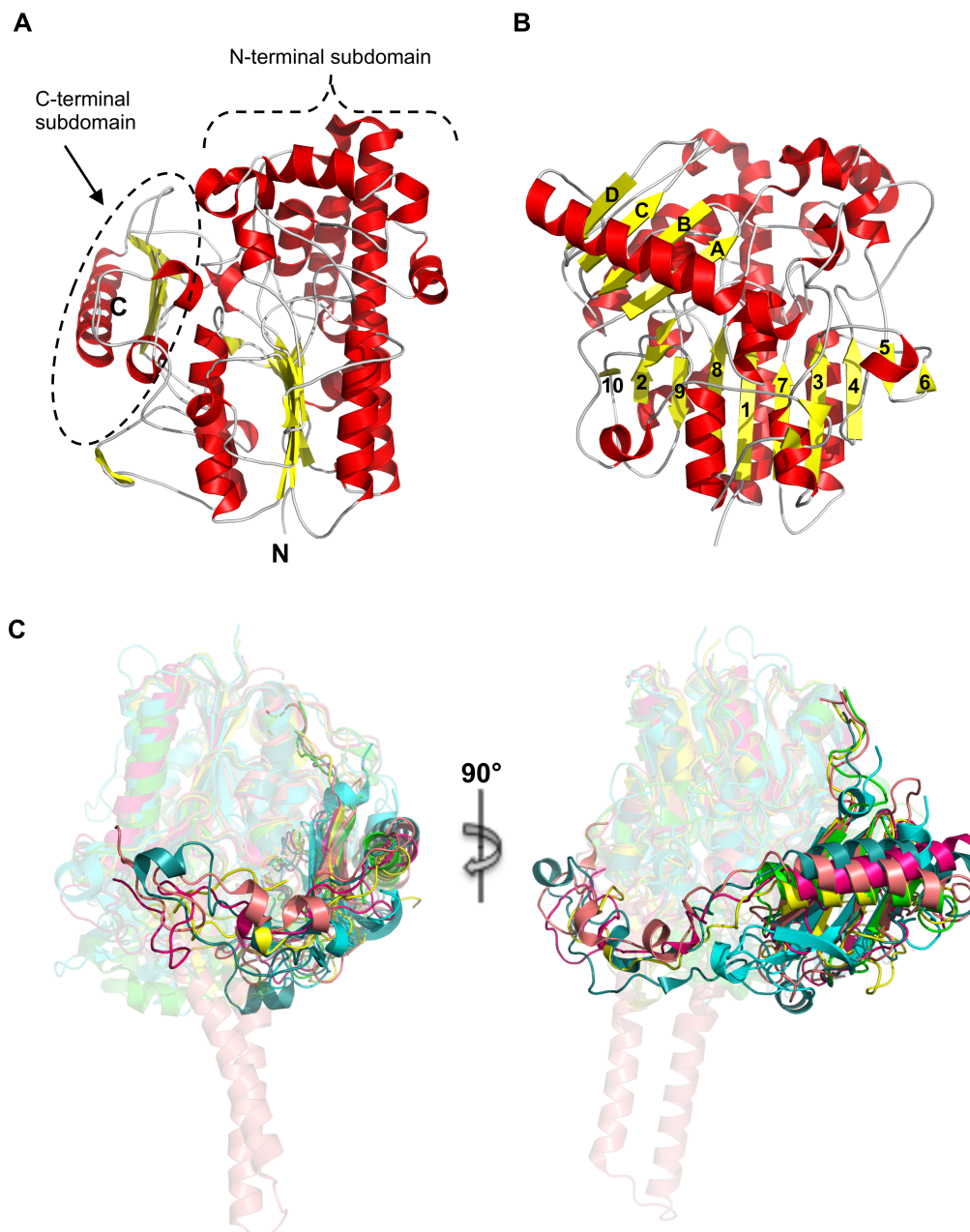


Figure 7. X-ray crystal structures of S1 family sulfatases.

The X-ray crystal structure of *Pseudomonas aeruginosa* PAS showing (A) the two subdomains with mixed α/β topology, and (B) rotated 90° showing the large meandering β -sheet of the N-terminal subdomain (numbered 1-10) and the small anti-parallel β -sheet of the smaller C-terminal subdomain (lettered A-D).

(C) Overlay of S1 sulfatase structures highlighting the differences of the C-terminal subdomain between sulfatases.

Colouring: *P. aeruginosa* PAS (PDB ID: 1HDH) in green, *H. sapiens*: iduronate-2-sulfatase (PDB ID: 5FQL) in cyan, N-acetylgalactosamine-4-sulfatase (PDB ID: 1FSU) in yellow, estrone sulfatase (PDB ID: 1P49) in salmon, galactosamine-6-sulfatase (PDB ID: 4FDI) in deepcyan, arylsulfatase A (PDB ID: 1AUK) in pink.

conserved mechanism for sulfate ester hydrolysis. These residues include the FGly, and three Asp and either an Asn or Gln, which coordinate a metal cation; in rare instances His has been observed within this quartet (Figure 8A). Proximal to this metal coordinating site, the pocket is lined with highly conserved charged residues; a single Arg, two His and two Lys are involved in FGly stabilization, sulfate binding and activation (Boltes et al., 2001; Bond et al., 1997; Demydchuk et al., 2017; Hernandez-Guzman et al., 2003; Lukatela et al., 1998; Rivera-Colón et al., 2012), likely by drawing electron density away from the sulfur center (Figure 8B). These residues that are involved in sulfate coordination and activation appear to lack a role in recognition of the leaving group, indicating that the discrimination between sulfated biomolecules can be attributed to residues other than those described here. Several structural studies have crystallized sulfatases with synthetic substrates, such as *p*-nitrocatechol sulfate (pNCS) bound in the active site (von Bülow et al., 2001). These structures showed the tightly bound sulfate coordinated in the pocket with the covalently attached phenyl ring showing a high degree of disorder. Historically, these small aryl substrates (which are often used as generic synthetic substrates for sulfatases) were initially used in the identification of sulfatases, and thus many have been classified as “arylsulfatases” based on this observed activity.

The Bolam research group has recently reported the structures of two S1 sulfatases (BT1596^{2S-sulf} [S1_9] and BT4656^{6S-sulf} [S1_11]), from *Bacteroides thetaiotaomicron* in complex with their biological substrates, which are components of heparin (Δ 4,5UA2S β 1–4GlcNS6S, an unsaturated uronic acid disaccharide, and N,O6-disulfo-glucosamine, respectively; PDB ID: 5G2T and 5G2V) (Cartmell et al., 2017).

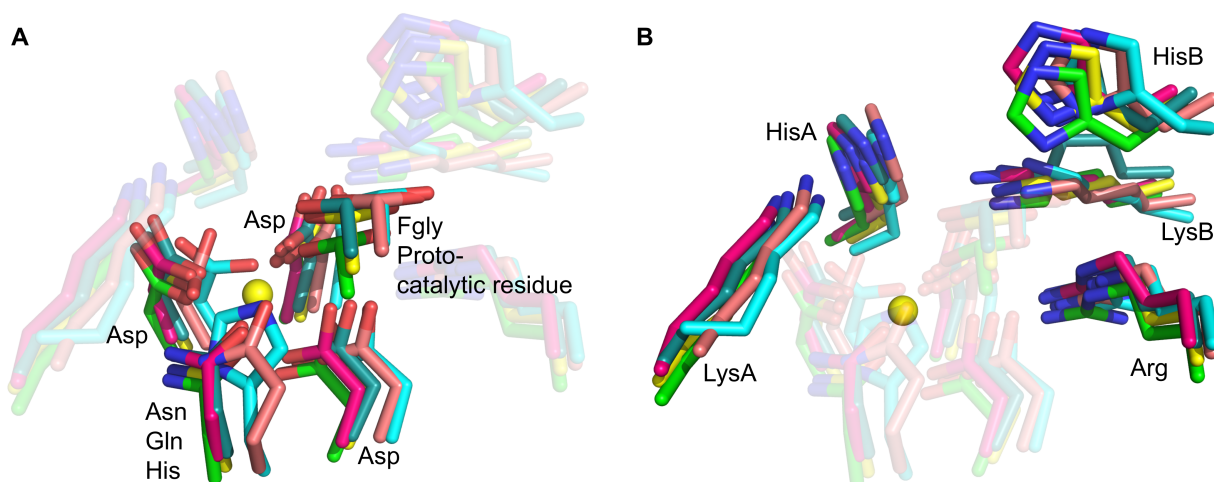


Figure 8. Identification of the 10 conserved polar residues among the S1 family sulfatases.

Overlay of X-ray crystal structures showing (A) the catalytic Fgly (or protocatalytic residue) and the metal coordinating residues and (B) the residues involved in sulfate binding, sulfate activation, and stabilization of Fgly.

Colouring: *P. aeruginosa* PAS (PDB ID: 1HDH) in green, *H. sapiens* iduronate-2-sulfatase (PDB ID: 5FQL) in cyan, N-acetylgalactosamine-4-sulfatase (PDB ID: 1FSU) in yellow, estrone sulfatase (PDB ID: 1P49) in salmon, galactosamine-6-sulfatase (PDB ID: 4FDI) in deepcyan, arylsulfatase A (PDB ID: 1AUK) in pink.

These sulfatases share the expected overall architecture of the S1 family (Figure 9A and 9B); however, at the active site location there are some differences among the conserved residues, for example BT1596^{2S-sulf} does not coordinate a metal ion but instead utilizes histidine residues to fulfill the role of sulfate activation (Figure 9C). The residues involved in substrate recognition in these complexes use an extensive hydrogen bond network to coordinate the non-reducing ends of the ligands in an *exo-acting* manner (Figure 9C and 9D); this agrees with the notion that the recognition of the leaving group is vital for substrate turnover. Given the vast and diverse pool of sulfated biomolecules, it is likely that these interactions are exclusive to the recognition of heparin by heparin-active S1_9 and S1_11 sulfatases and cannot be translated to the recognition of other sulfated polysaccharides.

1.6 Carrageenan degradation

The discovery and characterization of individual enzymes involved in carrageenan degradation and metabolism has often proved difficult due to its recalcitrance and the complex structural features of the different carrageenan families. To date only a few enzymes specific for the degradation of this sulfated galactan have been reported: κ -carrageenases (GH16) from *Pseudoalteromonas carrageenovora* (Michel et al., 2001a) and a marine Cytophaga-like bacterium (Potin et al., 1991); ι -carrageenases (GH82) from *Alteromonas fortis* (Michel et al., 2001b) and *Z. galactanivorans* (Ficko-Blean et al., 2017a); a λ -carrageenase (unclassified) from *P. carrageenovora* (Guibet et al., 2007); and endo-4S- κ -carrageenan sulfatases and endo-4S- ι -carrageenan sulfatases from *Pseudoalteromonas atlantica*

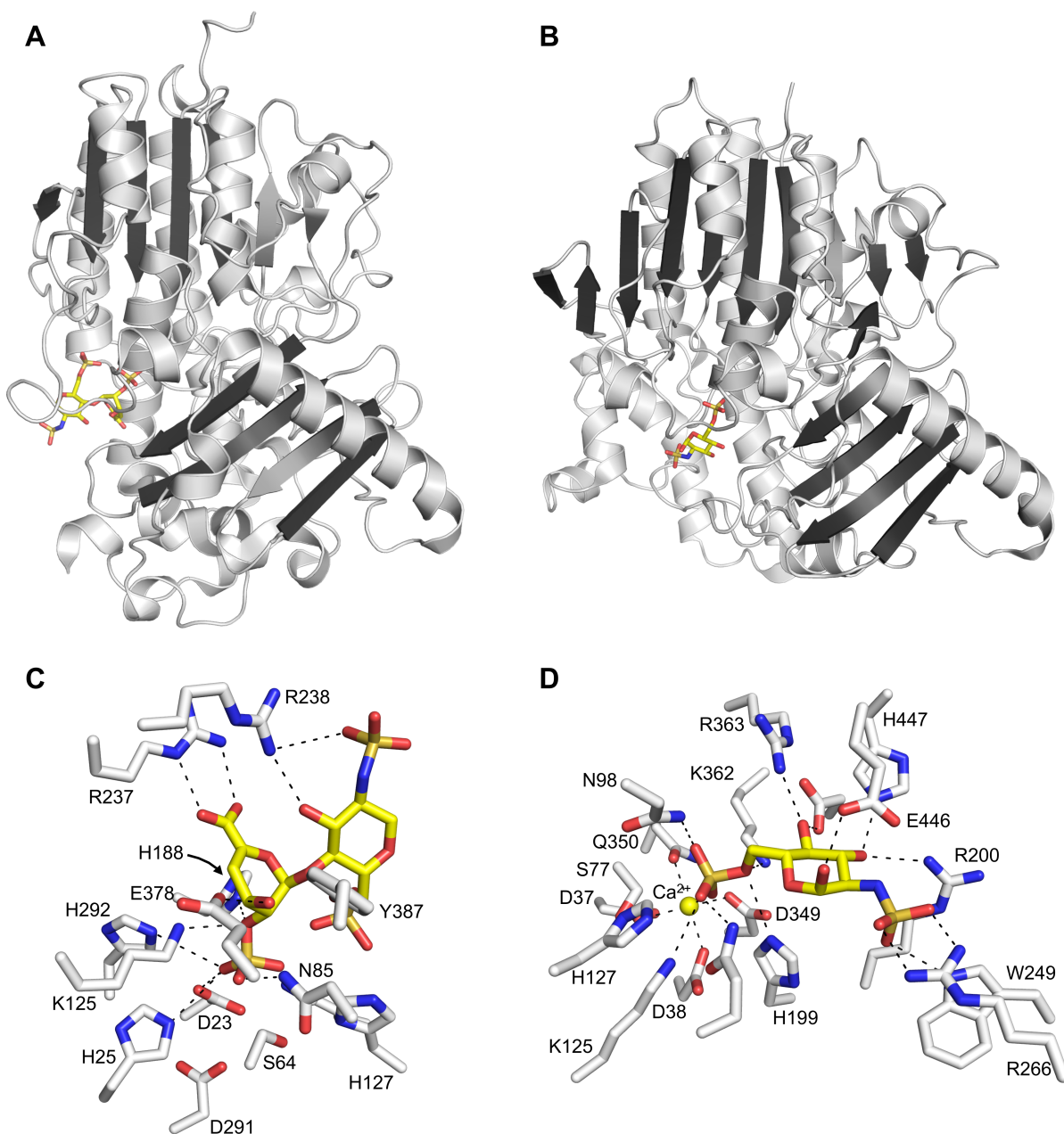


Figure 9. Known sulfatases in complex with their biological substrate.

(A) Cartoon structure of BT1596^{2S-sulf} in complex with $\Delta 4,5\text{UA}2\text{S}\beta 1\text{-}4\text{GlcNS6S}$.

(B) Cartoon structure of BT4656^{6S-sulf} in complex with GlcNS6S.

(C) and (D) Stick representations of BT1596^{2S-sulf} and BT4656^{6S-sulf} in complex with their biological substrates $\Delta 4,5\text{UA}2\text{S}\beta 1\text{-}4\text{GlcNS6S}$ and with GlcNS6S, respectively.

For both panels (C) and (D) amino acid side chains are shown in white, the sugars are shown in yellow, hydrogen bonds are represented as dashed lines, and the calcium ion is shown as a yellow sphere.

(Préchoux et al., 2013, 2016) and *P. carrageenovora* (Genicot et al., 2014), respectively. All of these enzymes displayed *endo*-acting activity. The identification of these enzymes shows that even though the differences between different classes of carrageenan can be quite minor, the enzymes employed for their degradation are drastically different at the sequence and structural levels, and often show carrageenan class specificity. For the characterized GHs, the target linkage is the β -glycosidic bond, and resulting substrate turnover yields a pool of mixed short carrageenan oligosaccharides (polygeenan) ranging in degree of polymerization, with the most abundant being tetrasaccharides and disaccharides. This suggests that the shortest oligosaccharide that satisfies the molecular requirements of substrate recognition for the GHs is a hexasaccharide. In order for this pool of short polygeenan to be further degraded, a subset of *exo*-acting CAZymes are required; these enzymes would have to be specific for not only the α - and β -linkages, but also any sulfate ester decorations. Furthermore, the specificity for different carrageenan classes observed among the *endo*-acting GHs may also be a specificity characteristic observed among *exo*-acting carrageenan enzymes. Interestingly, the first structure of a carrageenan degrader was the GH16 from *P. carrageenovora* in 2001 (Michel et al., 2001), and only recently have downstream components of carrageenan degradation been elucidated.

Genes encoding for CAZymes are often grouped together in the genome, along with transporters and sensory proteins. The term PUL (Polysaccharide Utilization Loci) has historically been reserved for bacteria belonging to the phylum *Bacteroidetes*, which contains the starch utilization system (SUS). However, this term has now been adopted across all bacteria to indicate regions of the genome that contain co-localized gene

clusters encoding for enzymes and proteins (often containing a high degree of CAZyme encoding genes) involved in the transport and saccharification of carbohydrates. As the complexity of a polysaccharide increases, the complexity of the respective PUL often increases as well.

The enzymatic saccharification of κ - and ι -polygeenan by the *Z. galactanivorans* carrageenan PUL showed specificity for carrageenan containing the 3,6-anhydro-D-galactose unit (predominantly κ -, ι -, and β -carrageenans) (Ficko-Blean et al., 2017a). Some of the enzymes encoded in this carrageenan PUL were biochemically characterized, including the necessary *exo*-acting GHs and sulfatases, though the characterization on the latter was limited to identifying the regioselectivity of the sulfatases. This work reported on the first 2S-sulfatase specific for α -carrageenan motifs, along with structural and functional characterization of the 3,6-anhydro-D-galactose processing enzymes. The degradative pathway proposed for κ - and ι -carrageenan by *Z. galactanivorans* firstly involves the polymer being hydrolyzed into a pool of polygeenan by the *endo*-acting GH16 and GH82 enzymes, respectively. Subsequently, the polygeenan is then desulfated by *exo*-2S- and *exo*-4S-sulfatases, before undergoing further depolymerization with alternating hydrolysis of the non-reducing end α - and β -glycosidic linkages by the *exo*-carrageenases GH127 and GH2, respectively (Figure 10A). In contrast, *P. carrageenovora* seems to utilize two different degradation pathways depending on the class of carrageenan being depolymerized (κ - or ι -carrageenan). For κ -carrageenan, the proposed degradative mechanism for *P. carrageenovora* utilizes a GH16 *endo*- β -carrageenase like that of *Z. galactanivorans* creating products of different degrees of polymerization. This is followed by subsequent

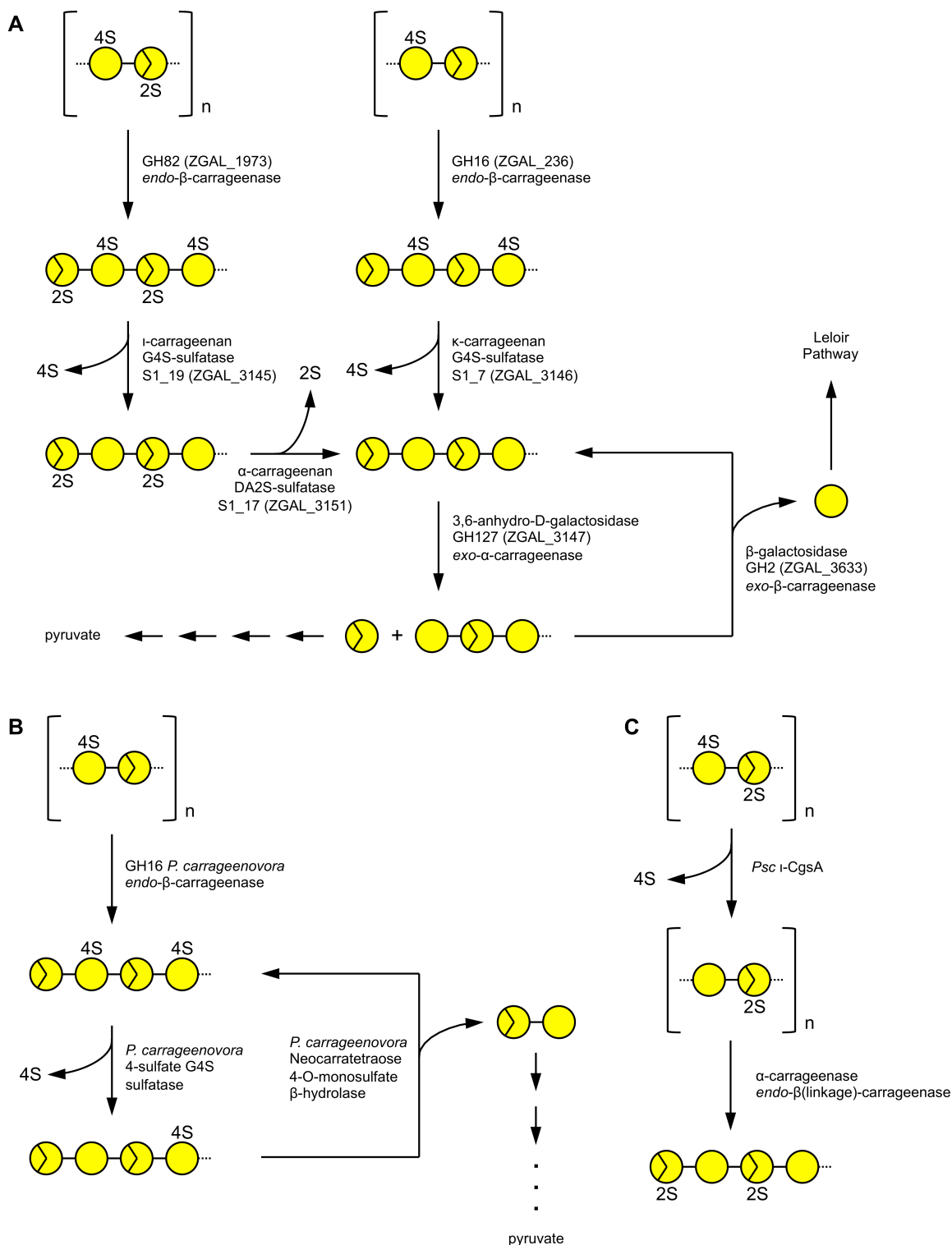


Figure 10. Carrageenan depolymerization pathways.

(A) κ - and ι -carrageenan degradation by *Z. galactanivorans*

(B) κ -carrageenan and ι -carrageenan degradation by *P. carrageenovora*

Representation of carrageenan components the same as in Figure 2.

alternating steps of an *exo*-4S- κ -carrageenan sulfatase removing the 4-sulfate group and a non-reducing end *exo*- β -carrageenase producing a β -carrageenan disaccharide (Figure 10B). However, the proposed degradative mechanism for *I*-carrageenan involves an *endo*-4S-*I*-carrageenan sulfatase (*Psc I-CgsA*) (Genicot et al., 2014), which, in this scenario, acts as the initiator and causes a class switch by remodelling *I*-carrageenan into α -carrageenan. The α -carrageenan is then hydrolyzed by an *endo*-carrageenase, and the resulting polygeenan pool is putatively degraded by as-yet unidentified *P. carrageenovora* *exo*-carrageenases and *exo*-carrageenan sulfatases (Figure 10C). It is possible that *P. carrageenovora* utilizes an *exo*-2S- α -carrageenan sulfatase hydrolyzing the 2-sulfate group at the non-reducing end of the polymer resulting in depolymerization of the short oligosaccharides similarly to κ -carrageenan depolymerization by *Z. galactanivorans*; however, this sulfatase is yet to be identified. These differing mechanisms of carrageenan degradation suggest that the order of events may not only be carrageenan class specific, but also prokaryotic species specific. As suggested earlier, the functional gene clusters between different species of bacteria influences the colonization population distribution of macroalgae-associated microorganisms, yet how the targeted polymer is degraded can differ between these colonizing species.

The gaps in knowledge surrounding carrageenan degradation are due not only to the issues attributed to the differing physicochemical properties characterizing the different carrageenan families, but also the difficult undertakings involved in sulfatase research: identifying the physiological substrate(s), identifying regioselectivity of sulfatases when substrates contain multiple sulfate esters, assays that appropriately

characterize the kinetics of sulfate ester hydrolysis, and the maturation of sulfatases under laboratory conditions.

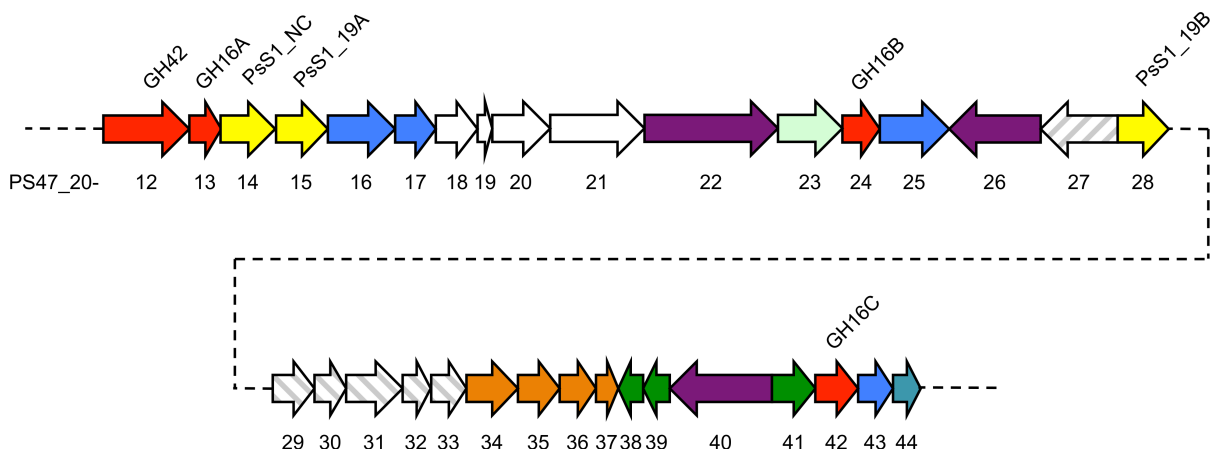
1.7 PS47: a marine polysaccharide-degrading strain of *Pseudoalteromonas* sp.

A new strain of *Pseudoalteromonas* sp. named PS47 was isolated by the Boraston Lab from the marine environment and selected for further study due to its ability to utilize seaweed polysaccharides as a carbon source. Its genome was sequenced resulting in 4.8 million base pairs present in 268 contigs that contained 4,159 putative coding sequences. The complete list of predicted amino acid sequences was submitted to dbCAN (DataBase for automated Carbohydrate-active enzyme ANnotation) (Yin et al., 2012) to identify regions of the genome that may encode for proteins that participate in carbohydrate decomposition. The sequences of these highlighted regions were further analyzed to identify possible PULs or PUL-like regions that utilize Ton-B dependent transporters to transport polysaccharides as opposed to the SUS system in *Bacteroidetes*. Analysis of the CAZome of PS47 revealed several putative PULs likely involved in the saccharification of agarose, porphyran, marine pectin, alginate, and carrageenan.

1.7.1 The PS47_2012-2044 gene cluster encodes putative carrageenan-specific enzymes

Of the putative marine polysaccharide PULs identified within the PS47 genome, the PUL encoded by ORFS 2012-2044 was of interest in relation to carrageenan

degradation as this region contains putative carrageenan CAZymes. This PUL encodes for three GH16 enzymes (PS47_2013, PS47_2024, and PS47_2042), one enzyme assigned to the family GH42 (PS47_2012), three hypothetical proteins with predicted β -propeller folds (PS47_2016, PS47_2017, and PS47_2025; a fold observed in *exo*-acting agarases), and three putative sulfatases (PS47_2014, PS47_2015, and PS47_2028). The amino acid sequences of PS47_2014, PS47_2015, and PS47_2028 fall into family S1_NC (Not Classified) for the former, and family S1_19 for the latter two. To distinguish between the sulfatases, the frequently used systematic nomenclature used for naming GHs will be adapted to suit the sulfatases. Thus, the three enzymes will be referred to as PsS1_NC, PsS1_19A, and PsS1_19B (Ps for *Pseudoalteromonas* sp., NC for the unclassified subfamily, and S1_19 for the family and sub-family, with the suffix “A” or “B” indicating the order of the S1_19 enzymes within the PUL). Collectively, the predicted composition of this gene cluster strongly suggests that it is an algal galactan PUL specific for carrageenan (Figure 11). This 33-gene locus also encodes for three Ton-B dependent transporters (PS47_2022, PS47_2026, and PS47_2040), and four enzymes putatively involved in the intracellular processing of the monosaccharides resulting from carrageenan degradation (PS47_2034, PS47_2035, PS47_2036, and PS47_2037). These latter four enzymes have homologs in *Z. galactanivorans* (ZGAL_3155, ZGAL_3156, ZGAL_3154, and ZGAL_3153, respectively) that have been shown to participate in the metabolism of 3,6-anhydro-D-galactose metabolism and are conserved across carrageenolytic bacteria (Ficko-Blean et al., 2017a). The PS47 PUL is conserved in another strain of *Pseudoalteromonas* sp. isolated by the Boraston Lab, U2A (PSU2A). All components of the PUL share > 95 % amino acid sequence identity



Legend:

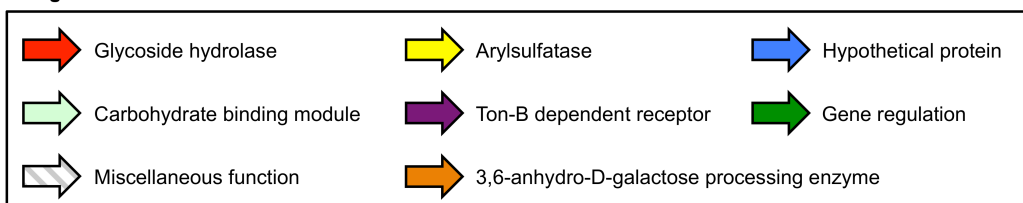


Figure 11. Genomic organization of the putative PS47 carrageenan PUL. Genes coloured according to protein function as in the legend.

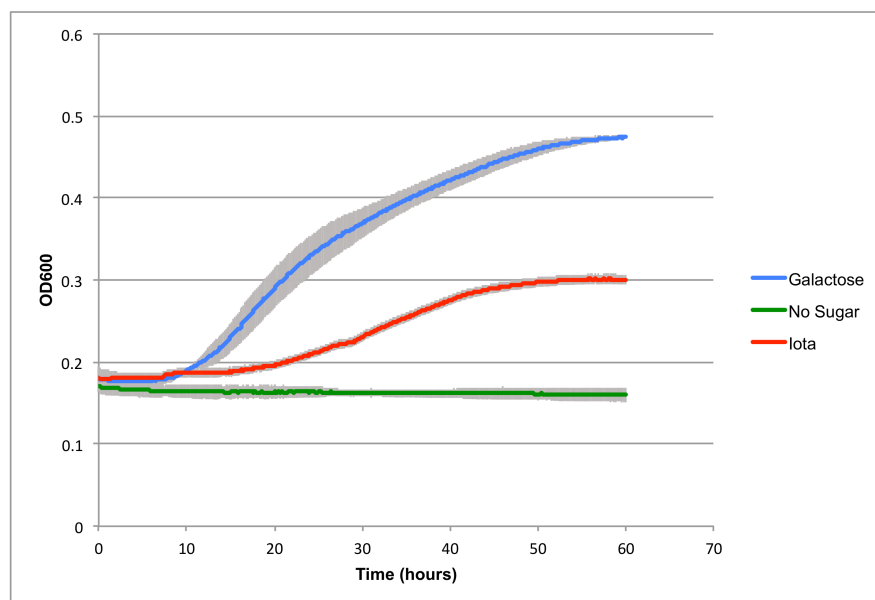


Figure 12. *Pseudoalteromonas* sp. U2A growth on i-NC4. Error bars show the standard deviation of measurements made in triplicate.

with their counterparts in PS47. PS47 proved difficult to culture in the laboratory, but PSU2A exhibited growth on i-NC4 as the sole carbon source (Figure 12; results kindly provided by Dr. Joanne Hobbs). The ability of PSU2A to utilize i-carrageenan for growth is consistent with the presence of the putative carrageenan PUL that it shares with PS47.

1.8 Hypothesis and Research objectives

Macroalgae playing a major role in the marine environment carbon cycle by the fixation of inorganic carbon into polysaccharides, yet the understanding of how this carbon is returned to the environment via microbial saccharification and metabolism of marine carbohydrates remains in large part unknown. Carrageenans constitute a massive amount of red algae biomass in marine ecosystems, and with carrageenan sulfation levels ranging from 20 – 30 % (Domozych et al., 2011) it becomes apparent that in order for PS47, and other marine heterotrophic bacteria to utilize this polysaccharide as a carbon source they must possess the machinery needed to overcome the metabolic barrier that the sulfate groups create. These sulfate groups have ultimately led to the relatively poor understanding of carrageenan processing due to the difficulties of working with the class of enzymes that removes them, ie. sulfatases.

I have identified a putative PUL in *Pseudoalteromonas* sp. PS47 that contains three possible sulfatases, PsS1_19A, PsS1_19B, and PsS1_NC (see section 1.7). Given the gene content of this PUL I **hypothesize that its encoded machinery targets carrageenan**. Furthermore, because this PUL incorporates three distinct sulfatase genes, I **hypothesize that each sulfatase will have a unique specificity in**

carrageenan or polygeenan desulfation and that these will be critical to enabling the complete breakdown of the carbohydrate. With the high degree of catalytic residue conservation amongst the S1 sulfatase family and the huge diversity of sulfated biomolecules, **I hypothesize that recognition of the leaving group is what drives specificity for S1 family sulfatases.** Along with these three putative sulfatases I also propose that an additional GH from PS47, PsGH42 non-homologous to any *Z. galactanivorans* GHs, plays a pivotal role in the final steps of κ - and ι -polygeenan saccharification.

I will address my hypotheses with the objectives of **demonstrating activity of the PUL components against carrageenan, elucidating the order in which the enzymes function in the depolymerization of carrageenan, and providing molecular insight into how PsS1_19A, PsS1_19B, and PsS1_NC target different sulfate substituents.** My goal is to provide a deeper understanding of how marine bacteria cope with sulfated galactans in terms of carbohydrate degradation and nutrient acquisition. Furthermore, providing insight into how sulfatases (specifically, carbohydrate-active sulfatases) bind and interact with their biological substrates through X-ray crystallographic structures of sulfatases in complex with their cognate substrate will deepen the understanding of how S1 sulfatases function.

Chapter 2: Materials and Methods

Materials

κ-neocarrabiose, κ-neocarratetraose, κ-neocarrahexaose, κ-ι-κ-neocarrahexaose, and κ-neocarraoctaose were obtained from V-Labs (Covinton, LA). κ-carrapentaose was obtained from Qingdao BZ Oligo Biotech Co. (Qingdao, China). All reagents, chemicals and other carbohydrates were purchased from Sigma unless otherwise specified.

Methods

Isolation, genome sequencing and annotation of Pseudoalteromonas sp. PS47 – *Pseudoalteromonas sp. PS47* was isolated from *Macrocystis sp.* growing in the intertidal zone near Victoria British Columbia during spring 2010 by Jan-Hendrick Hehemann, a previous PostDoc in the Boraston Lab. A marine seaweed broth was prepared from 1 % (w/v) dried *Fucus sp.* in artificial seawater. This broth autoclaved and then inoculated with fragments of freshly collected, partially decomposed *Macrocystis sp.* and incubated at 20 °C and 200 rpm for 36 hours, after which the culture was diluted in artificial seawater and streaked for isolated colonies on seaweed broth agar plates, which were selected and re-streaked twice to ensure purity. Isolated colonies were selected and further screened for their ability to use marine polysaccharides as the sole carbon source by plating on minimal marine medium containing 0.25 % (w/v) of polysaccharide and vitamin mix. Genomic DNA was extracted using the QIAamp DNA mini kit (Qiagen) after culture in marine broth 2216 (Difco) at 20 °C and 200 rpm for 24 hours and Genome sequencing was performed by the Genome Sciences Centre of the British Columbia Cancer Agency. 4.8 mega base pairs of sequence were assembled in

268 contigs. Open reading frames were identified and annotated using the automated RAST annotation pipeline resulting in 4,159 coding sequences. The complete FASTA sequence list was then submitted to the webserver and database dbCAN (database for Carbohydrate-active enzyme Annotation) (Yin et al., 2012) of which identified putative CAZymes were further annotated using BLASTp and PHYRE2 (Kelley and Sternberg, 2009) looking for putative carrageenan active enzymes.

Growth of U2A on I-NC4 - U2A was grown overnight in Zobell marine broth at 25 °C with 180 rpm, pelleted and washed with complex minimal marine medium (MMM) (Therkildsen et al., 1997). I-NC4 was prepared in MMM at 0.5 %, inoculated with washed cells in a 96-well microplate and incubated in a SpectraMax M5 plate reader at 25 °C for 60 hours with shaking and OD_{600nm} readings every 10 minutes. Control wells contained uninoculated I-NC4, carbohydrate-free MMM inoculated with U2A, and MMM containing 0.5 % D-galactose inoculated with U2A. Data shown are the mean of two replicates.

Cloning and mutagenesis – The genes encoding for PsS1_19A, PsS1_19B, and PsS1_NC were amplified from *Pseudoalteromonas* sp. PS47 genomic DNA using the primers PsS1_19A fwd/rev, PsS1_19B fwd/rev, and PsS1_NC fwd/rev (Table 1), respectively, PCR products were cleaned-up (Qiagen PCR Purification Kit) and ligated into pET28a between the *Nhe*I and *Xho*I restriction enzyme cut sites using T4DNA ligase to create constructs encoding proteins with a thrombin cleavable N-terminal histidine affinity tag (His-Tag). The gene encoding for PsGH42 was amplified and cloned into pET28a using the PIPE (Polymerase Incomplete Primer Extension) method and the primers PsGH42 fwd/rev and Vector_primer1/2 (Table 1). Briefly, the insert is

Table 1: Oligonucleotide primer sequences used for gene amplification

PsS1_NC_fwd	5'-CATATGGCTAGCATCAAAAAGCCAAATGTGC -3'
PsS1_NC_rev	5'-GTGTGTCTCGAGCTATAAAAAGCTTTCATGTGGTG-3'
PsS1_NC_fwd_C84S	5'-CAGGCTTATCCTCACCTTCGCGC-3'
PsS1_NC_rev_C84S	5'-AGGTGAGGATAAGCCTGTAGATG-3'
PsS1_19A_fwd	5'-CATATGGCTAGCGAACAAAAACCCAATATTATTCTTATTG-3'
PsS1_19A_rev	5'-GTGTGTCTCGAGTTACTTGGTTGTTTAAACGTTTTATC-3'
PsS1_19A_fwd_C78S	5'-CACCCATTCTCTGGCCCAAGCCGAG-3'
PsS1_19A_rev_C78S	5'-CTTGGGCCAGAGAATGGGTGCGCTAC-3'
PsS1_19B_fwd	5'-CATATGGCTAGCAAACCTAATATTGTTTTAATTTTTGC-3'
PsS1_19B_rev	5'-GTGTGTCTCGAGTTAAGGTTGAGTTGCTGGTAAAC-3'
PsS1_19B_fwd_C77S	5'-GTGATTCTACTTCTGGTCCTTCACGCGCAG-3'
PsS1_19B_rev_C77S	5'-GTGAAGGACCAGAAGTAGAATCACTCACGTAAC-3'
PsGH42_fwd	5'-aaggagatataccatgTCGACGAAGAAAGCG-3'
PsGH42_rev	5'-gtggtggtggtggtgTTGAACAGTAACCTCAACAAGC-3'
Vector_primer1	5'-GCTTCTTGCTGCGACATGGTATATCTCCTTCTTAAAGTTAAAC-3'
Vector_primer2	5'-GAAGTTACTGTTCAACACCACCACCACCACC-3'
BovGH42_fwd	5'-CATATGGCTAGCGGTATCAACTACCTGTA-3'
BovGH42_rev	5'-GTGTGTCTCGAGTCATTCCAGGCTCACTT-3'

Restriction enzyme recognition sequences underlined

Mutated codon sequence shaded gray

pET28 vector sequence represented as lowercase

amplified with the PsGH42 fwd/rev primers containing 5'- sequence complementary to amplified vector, which was amplified with the Vector_primer1/2 primers containing 5'- sequence complementary to amplified insert. All amplification was done using Phusion High-Fidelity DNA polymerase. The constructs for PsS1_19B and PsS1_NC were transformed into *Escherichia coli* BL21 (DE3) Star cells (Invitrogen) and were grown in LB broth containing 50 $\mu\text{g mL}^{-1}$ kanamycin sulfate. The PsS1_19A construct was transformed into *E. coli* BL21 (DE3) Star Codon Plus cells (Invitrogen) and grown in LB broth containing 50 $\mu\text{g mL}^{-1}$ kanamycin and 50 $\mu\text{g mL}^{-1}$ chloramphenicol. For activity assays requiring matured sulfatases, the PsS1_19A, PsS1_19B, and PsS1_NC constructs were co-transformed with pBAD/myc-his A Rv0712 (FGE) (pBAD/myc-his A Rv0712 (FGE) was a gift from Carolyn Bertozzi (Addgene plasmid #16132)) (Carrico et

al., 2007) as previously described with the addition of 100 $\mu\text{g mL}^{-1}$ ampicillin. Catalytically inactive mutants were created by site-directed mutagenesis (QuikChange Site-Directed Mutagenesis kit), replacing the catalytic cysteine with a serine using the primers PsS1_19A_C78S fwd/rev, PsS1_19B_C77S fwd/rev, and PsS1_NC_C84S fwd/rev (Table 1). The mutant plasmids were sequenced to confirm the mutation had occurred in the correct location and no other mutations had been introduced. These mutant plasmids were transformed into *E. coli* BL21 (DE3) Star or *E. coli* BL21 (DE3) Star Codon Plus as described above.

E. coli optimized synthetic genes - Optimized synthetic genes for *Bacteroides ovatus* GH16 (BovGH16; accession number HMPREF1069_02099) and *B. ovatus* GH42 (BovGH42; accession number HMPREF1069_02044) in the vector pUC57-KAN were synthesized (GeneScript) and transformed into *E. coli* BL21 (DE3) Star cells. Cultures were grown in LB broth containing 50 $\mu\text{g mL}^{-1}$ kanamycin sulfate at 37 °C overnight, and the plasmids were extracted using the QIAprep Miniprep kit (Qiagen). The BovGH16 gene was excised using the restriction enzymes NheI and XhoI, gel extracted using the QIAquick Gel Extraction kit (Qiagen), ligated into pET28a between the NheI and XhoI restriction enzyme cut sites and transformed into *E. coli* BL21 (DE3) Star cells. The BovGH42 gene was amplified using the primers BovGH42 fwd/rev (Table 1), the resulting PCR product was cleaned-up, cleaved with the restriction enzymes NheI and XhoI, ligated into pET28a between the NheI and XhoI restriction enzyme cut sites, and transformed into *E. coli* BL21 (DE3) Star cells as previously described.

Protein expression and purification – Cells were grown at 37 °C with agitation at 180 rpm until cell density reached an OD₆₀₀ of ~ 0.5 at which time the temperature was dropped to 16 °C, at this time cells containing the FGE plasmid had FGE expression induced with 0.02 % L-arabinose. After ~ 2 hours, sulfatase expression was induced with a final concentration of 0.5 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) and the culture incubated for a further 16 hours. Cultures were centrifuged at 8,000 × *g* for 10 min. Cells were chemically lysed by resuspension in 35 % (w/v) sucrose, 1 % (w/v) deoxycholate, 1 % (v/v) Triton X-100, 500 mM NaCl, 10 mg lysozyme and 0.2 μg mL⁻¹ DNase in 20 mM Tris (pH 8.0). Following centrifugation at 16,000 × *g* for 30 min, the supernatant was applied to a gravity flow nickel affinity chromatography column (Bio-Rad) and eluted with a stepwise increase in imidazole concentration of 5, 10, 15, 20, 40, 50, 100, and 500 mM in 0.5 M NaCl, 20 mM Tris (pH 8.0). Elution fractions were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), and samples containing the protein of interest were concentrated with an Amicon ultrafiltration cell (EMD Millipore) fitted with a 10 kDa MWCO membrane. Further purification was performed using a HiPrep 16/60 Sephacryl S-200 HR size exclusion chromatography column (GE Healthcare) equilibrated with 0.5 M NaCl, 20 mM Tris (pH 8.0). Pure samples, as assessed by SDS-PAGE, and free of soluble aggregate were pooled and concentrated for use. All proteins used for crystallization were subjected to a further purification step cleaving the N-terminal His-Tag via overnight incubation with thrombin in thrombin cleavage buffer: 500 mM NaCl, 20 mM Tris-HCl (pH8.0), and 2.5 mM CaCl₂ prior to size exclusion chromatography.

Oligomeric state determination using size exclusion chromatography – Elution volumes (V_e) for protein molecular weight standards of 10 mg mL⁻¹ albumin (MW: 66 kDa), 3 mg/mL conalbumin (MW: 75 kDa), 5 mg/mL alcohol dehydrogenase (MW: 150 kDa), and 4 mg mL⁻¹ beta-amylase (MW: 200 kDa) were utilized to calibrate a HiPrep 16/60 Sephacryl S-300 HR column, 16 mm × 1600 mm (GE Healthcare). Blue dextran (MW: 2,000 kDa) was used to determine the void volume (V_o) and a standard curve was created plotting molecular mass vs. V_e/V_o for each respective protein standard. Samples of PsS1_19A, PsS1_19B, and PsS1_NC at 10 mg/ml were run and their V_e/V_o values plotted against the standard curve. All samples and standards were run at a flow rate of 0.5 mL/min in 500 mM NaCl and 20 mM Tris (pH 8.0).

Oligosaccharide production and purification – For oligosaccharide production, ι-carrageenan was resuspended in 20 mM Tris (pH 8.0) and 1 mM dithiothreitol (DTT) to a final concentration of 2 %. Excess BovGH16 was added and incubated at 37 °C for 72 hours. The sample was frozen, lyophilized, resuspended in 2 mL dH₂O and passed over a size exclusion chromatography column containing Bio-Gel P-2 Gel (Bio-Rad) in 0.1 M ammonium bicarbonate. Fractions were analyzed for the presence and size of oligosaccharides by thin layer chromatography (TLC) in the running buffer 8:4:1 formic acid:butanol:water. TLC plates were developed by spraying with 5 % H₂SO₄ in ethanol and incubating at 100 °C for 10 min. Fractions containing pure oligosaccharides were pooled and lyophilized.

pH-based sulfatase activity assay – PsS1_19A and PsS1_19B were expressed and purified as detailed above, with the exception that the size exclusion buffer was replaced with assay buffer (0.5 mM NaCl, 1 mM 3-(N-morpholino)propanesulfonic acid

(MOPS) (pH 7.16)). The hydrolysis of sulfate esters by sulfatases releases protons, which were detected spectrophotometrically in a pH-based assay using a para-nitrophenol (pNP)/MOPS indicator/buffer pairing at pH 7.16. Each reaction contained 0.5 mM NaCl, 1.0 mM MOPS (pH 7.16), 0.8 mM pNP (500 mM stock prepared in DMSO), 1 μ M sulfatase, and 0.01 – 3.3 mM or 0.05 – 2.5 mM carrageenan substrate for PsS1_19A and PsS1_19B, respectively. All dH₂O used to prepare assay buffers was treated to remove any dissolved carbonate through boiling for 30 minutes followed by cooling at 4 °C under nitrogen. Initial rates of sulfate ester hydrolysis were determined by monitoring a decrease in absorbance at 405 nm at 25 °C for 1 hour. Standard curves were prepared by substituting carrageenan substrates with titrated amounts of HCl up to 1 mM. To confirm the change in absorbance was due to sulfate ester hydrolysis, samples with omitted pNP, carrageenan oligosaccharide, or sulfatase were also monitored. Rates were determined as previously described (Janes et al., 1998), which establishes a linear relationship between protons produced and absorbance; a correction factor was also used to account for buffer and indicator concentrations using equations (1) and (2) as a previously described method (Janes et al., 1998).

$$(1) Q = \frac{C_B}{C_{In}} \times \frac{1}{\Delta\epsilon_{403nm}l}$$

$$(2) rate \left(\frac{\mu mol}{min} \right) = \frac{dA}{dt} \times Q \times volume \times 10^6$$

Where C_B and C_{In} are the MOPS and pNP concentrations, respectively, $\Delta\epsilon_{403nm}$ is the difference between the protonated and deprotonated forms of the extinction coefficient at 403 nm, and l is the path length of the cuvette. Rate vs. substrate concentration data were then fit to the Michaelis-Menten equation and kinetic parameters determined using

nonlinear regression in GraphPad Prism. Values and errors represent the means and standard deviations, respectively, from experiments performed in triplicate.

Fluorophore-assisted carbohydrate electrophoresis (FACE) – Carrageenan oligosaccharides were resuspended to a final concentration of 1 mM in 500 mM NaCl, 20 mM Tris (pH 8.0) and incubated at room temperature overnight with 1 μ M PsS19A or PsS1_19B. Following incubation, samples were dried using a speed vacuum until completely dry, then resuspended in equal volumes of 0.02 M 8-aminonaphthalene-1,3,6-trisulfonic acid disodium salt (ANTS) in 3:17 acetic acid:dH₂O and 0.1 M NaCNBH₃ in DMSO and incubated at 37 °C overnight. The samples were then dried again in the speed vacuum and resuspended in FACE loading dye (0.014 % bromophenol blue, 10 % glycerol, 62 mM Tris (pH 6.8)) and stored at -20 °C. Samples were separated in a 35 % polyacrylamide native gel run at 100 V for 30 min followed by 300 V for 1 hour and visualized under UV light (Robb et al., 2017).

NMR spectroscopy – Full length κ -carrageenan was resuspended to a final concentration of 0.1 % in sodium phosphate buffer (pH 7.4), 500 mM NaCl and heated at 90 °C for 15 minutes. After cooling, samples were incubated at room temperature with gentle mixing for 48 hours with 2 μ M PsS1_19A. Following incubation samples were lyophilized to dryness. Deuterium oxide (D₂O; 2 mL) was added to the resultant solids and incubated at 4 °C for 30 min to solubilize excess salts, which were removed along with the supernatants following centrifugation at 3800 rpm. The remaining carrageenan was re-suspended in 10 mM sodium phosphate buffer in D₂O (pD = 8.0) and sonicated for 3 h at 20 – 40 °C. ¹H NMR data were collected on a Bruker 400 MHz spectrometer equipped with an inverse probe at 70 °C. Chemical shifts were calibrated

using 0.02 % sodium 3-(trimethylsilyl)-1-propanesulfonate (DSS) as the internal standard.

Dynamic light scattering (DLS) – DLS was performed using solutions of PsS1_19A protein at 0.6 and 1.2 mg/mL in 0.5 M NaCl, 20 mM Tris (pH 8.0). DLS data were collected at 25 °C in triplicate, averaging 10 acquisitions of 10 seconds using a DynaPro plate reader (Wyatt Technology). The data were then analyzed using the Dynamics version 7.1 software. The molecular weight was determined based on globular protein model.

Thin layer chromatography (TLC) – β - κ -NC4 and κ -NC4 oligosaccharides were resuspended in to a final concentration of 1 mM in 500 mM NaCl, 20 mM Tris (pH 8.0) and incubated at room temperature overnight with 1 μ M PsGH42 or BovGH42 with and without 1 μ M PsS1_19B. Following incubation, samples were spotted on a silica gel plate and allowed to air dry. The silica gel plate is then placed in a chamber containing the running buffer of formic acid, butanol, and water at a ratio of 8:4:1. The silica gel plate is then dried and visualized using 5 % H₂SO₄ in ethanol and heated at 110 °C for 15 minutes.

Crystallization – All PsS1_19A, PsS1_19B, and PsS1_NC crystals were grown at 18 °C by hanging drop vapour-diffusion and BovGH42 crystals were grown at 18 °C by sitting drop vapour-diffusion, both with equal volumes (1 μ L) of protein (3 – 20 mg/mL) and crystallization solution. Crystallization conditions can be found in Table 2. For cryoprotection, 25 - 32 % ethylene glycol was added to the respective crystallization solution. Crystals were then flash-cooled in liquid nitrogen and kept at 100 K for diffraction data collection.

Table 2: Crystallization condition for all X-ray crystal structures presented

Construct	Precipitant	Buffer	Salt	Additive
PsS1_19A (all)	19 – 23 % PEG 3350	0.1 M citric acid (pH 3.5)	-	0.1 M arginine
PsS1_19B wild type	25 % PEG 3350	0.1 M sodium acetate (NaOAc) trihydrate (pH 4.5)	-	-
PsS1_19B C77S ι- NC4 and κ-NC2 complexes	20 – 21 % PEG 3350	-	0.02 – 0.05 M ZnCl ₂	0.1 M gly-gly 7.5 % glycerol
PsS1_NC wild type	25 % PEG 5K MME	0.1 M Bis Tris (pH 6.5)	-	0.1 M 2-Methyl-2,4- pentanediol (MPD)
PsS1_NC C84S ι-NC2 and κ-NC2 complexes	20 – 21 % PEG 4000	0.1 M HEPES (pH 7.5)	0.05 – 0.075 M NaOAc	0.1 M arginine 10 % glycerol
BovGH42 wild type	30 % PEG 400	0.1 M Tris (pH 8.5)	0.2 M MgCl ₂	-
BovGH42 iodine derivative	25 % PEG 4000	0.1 M tri-sodium citrate (pH 5.6)	0.2 M ammonium sulfate	-

Structure determination – For all data collected “in-house” (beamline identification in Table 3); diffraction data were collected at a wavelength of 1.54187 Å using a Rigaku R-Axis IV++ area detector coupled to a MM-002 X-ray generator with Osmic “blue” optics and an Oxford Cryostream 700, and an “in-house” beam comprising a Pilatus 200K 2D detector coupled to a MicroMax-007HF X-ray generator with a VariMax™-HF ArcSec Confocal Optical System and an Oxford Cryostream 800. For all data collected at the Canadian Light Source (CLS, Saskatoon, Saskatchewan) on beamline 08ID-1 were processed using MOSFLM and SCALA (Powell, 1999; Project, 1994). For all data collected “in-house” or at the Stanford Synchrotron Radiation Lightsource (SSRL, Stanford, California) on beamline BL9-2 or BL11-1, diffraction data were integrated, scaled and merged using HKL2000 (Otwinowski and Minor, 1997) and converted to a file with the extension *mtz* with SCALEPACK2MTZ (Project, 1994). The structure of wild-type PsS1_19A was solved by molecular replacement using a sulfatase

from *Pseudomonas aeruginosa* (PDB ID code: 1HDH) (Boltes et al., 2001) as a search model, and the structures of wild-type PsS1_19B and PsS1_NC were solved by molecular replacement using PsS1_19A (PDB ID code: 6BIA) (Hettle et al., 2018) as a search model and the program PHASER (McCoy et al., 2007) for both solutions. An iodide derivative of BovGH42 was obtained by soaking the crystal in crystallization solution containing 25 % (v/v) ethylene glycol and 1 M sodium iodide for 10 minutes prior to data collection. Processed data for the BovGH42 wild type and BovGH42 iodide derivative were combined using CAD (Dodson, Winn, & Ralph, 1997) and a heavy atom substructure was determined using SHELXC and SHELXD, followed with initial phasing done with SHELXE (Schneider and Sheldrick, 2002). The resulting phases were of high enough quality for an initial model to be built using ARPw/ARP (Morris, 2002). This iodide derivative BovGH42 model was then used to solve the high-resolution structure of BovGH42 by molecular replacement using PHASER. Starting with these initial phases for; PsS1_19A, PsS1_19B, PsS1_NC, and Bov42, BUCCANEER (Cowtan, 2006) was used to build working models, which were finished by manual building with COOT (Emsley and Cowtan, 2004) and refinement with REFMAC (Murshudov et al., 1997). These complete models were used as search models to solve all mutant complexes. All data processing and model refinement statistics are shown in Table 3. For all structures, the addition of water molecules was performed in COOT with FINDWATERS and manually checked after refinement. In all datasets, refinement procedures were monitored by flagging 5 % of all observations as “free” (Brünger, 1992). Model validation was performed with MOLPROBITY (Davis et al., 2007).

Table 3: X-ray data collection and structure statistics

	PsS1_19A	PsS1_19A C78S	PsS1_19A C78S
	Native	κ -C5 complex	κ -I- κ -NC6 complex
PDB ID	6BIA	6B0K	6B0J
Data Collection			
Beamline	In-house	CLS 08ID-1	SSRL BL 9-2
Wavelength (Å)	1.54187	0.97949	0.97946
Space Group	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2
Cell Dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	133.83, 133.83, 223.26	133.07, 133.07, 223.94	123.39, 123.39, 240.11
Resolution (Å)	30.00-2.80 (2.89-2.80)	29.76-2.15 (2.19-2.15)	39.02-2.50 (2.64-2.50)
<i>R</i> _{merge}	0.143 (0.587)	0.168 (1.145)	0.174 (0.780)
<i>R</i> _{pim}	0.074 (0.356)	0.059 (0.434)	0.067 (0.311)
CC1/2	0.993 (0.912)	0.998 (0.878)	0.992 (0.778)
<I/σI>	13.7 (2.3)	14.9 (2.3)	9.2 (2.5)
Completeness (%)	99.9 (99.8)	97.3 (99.1)	97.8 (97.7)
Redundancy	7.4 (5.0)	17.1 (15.3)	7.2 (6.9)
No. of reflections	377,673	1,816,109	452,618
No. Unique	50,781	106,480	63,124
Refinement			
Resolution (Å)	2.80	2.15	2.50
<i>R</i> _{work} / <i>R</i> _{free}	0.24/0.27	0.24/0.27	0.21/0.23
No. of atoms			
Protein	3567 (A), 3561 (B), 3543 (C)	3564 (A), 3565 (B), 3555 (C)	3534 (A), 3566 (B), 3553 (C)
Ligand	3 (Ca), 39 (Citrate), 12 (EDO)	3 (Ca), 51 (κ-C5-A), 66 (κ-C5-B), 66 (κ-C5-C), 13 (Citrate), 12 (ARG), 20 (EDO), 1 (Cl)	3 (Ca), 80 (κ-I-κ-NC6-A), 80 (κ-I-κ-NC6-B), 80 (κ-I-κ-NC6-C), 30 (κ-I-κ-NC6-A*), 55 (κ-I-κ-NC6-B*), 16 (TRIS), 24 (ARG), 8 (EDO)
Water	61	360	239
<i>B</i> -factors			
Protein	44.4 (A), 43.5 (B), 55.5 (C)	31.0 (A), 30.3 (B), 38.9 (C)	28.4 (A), 31.1 (B), 27.0 (C)
Ligand	45.4 (Ca), 65.9 (Citrate), 61.6 (EDO)	49.3 (Ca), 42.8 (κ-C5-A), 51.6 (κ-C5-B), 49.3 (κ-C5-C), 66.5 (Citrate), 47.2 (ARG), 42.3 (EDO), 24.6 (Cl)	41.3 (Ca), 33.9 (κ-I-κ-NC6-A), 33.8 (κ-I-κ-NC6-B), 31.8 (κ-I-κ-NC6-C), 51.9 (κ-I-κ-NC6-A*), 57.6 (κ-I-κ-NC6-B*), 59.0 (TRIS), 66.4 (ARG), 48.3 (EDO)
Water	25.8	28.2	19.5
r.m.s.d			
Bond lengths (Å)	0.008	0.008	0.008
Bond angles (°)	1.237	1.342	1.391
Ramachandran (%)			
Preferred	96.7	97.1	96.5
Allowed	3.3	2.9	3.5
Disallowed	0	0	0
C-beta outliers	0	0	0

	PsS1_19A C78S	PsS1_19B	PsS1_19B C77S
	i-NC4 complex	Native data set	κ-NC2 complex
PDB ID	6B1V		
Data Collection			
Beamline	In-house	CLS 08ID-1	CLS 08ID-1
Wavelength (Å)	1.54187	0.9794	0.9794
Space Group	P4 ₃ 2 ₁ 2	P1	P3 ₁
Cell Dimensions			
a, b, c (Å)	130.44, 130.44, 229.75	48.51, 88.58, 106.6	93.05, 93.05, 300.21
Resolution (Å)	30.00-2.84 (2.89-2.84)	47.60-2.50 (2.57-2.50)	48.15-1.95 (1.98-1.95)
R _{merge}	0.181 (0.571)	0.109 (0.334)	0.279 (1.178)
R _{pim}	0.071 (0.262)	0.104 (0.317)	0.299 (1.278)
CC1/2	0.988 (0.872)	0.988 (0.864)	0.996 (0.828)
<I/σI>	10.7 (2.2)	8.0 (2.9)	12.4 (3.2)
Completeness (%)	99.7 (99.7)	97.9 (96.4)	99.7 (99.1)
Redundancy	7.3 (6.3)	3.8 (3.3)	11.6 (11.6)
No. of reflections	345,931	220,129	2,444,195
No. Unique	47,557	14,846	211,234
Refinement			
Resolution (Å)	2.84	2.50	1.61
R _{work} /R _{free}	0.23/0.26	0.24/0.27	0.19/0.19
No. of atoms			
Protein	3556 (A), 3574 (B), 3571 (C)	3523 (A), 3526 (B), 3517 (C), 3519 (D)	3537 (A), 3526 (B), 3542 (C), 3517 (D), 3546 (E), 3529 (F)
Ligand	3 (Ca), 59 (i-NC4-A), 59 (i-NC4-B), 59 (i-NC4-C), 59 (i-NC4-A*), 59 (i-NC4-B*), 4 (EDO)	4 (Ca)	6 (Ca), 156 (κ-NC2), 9 (Cl), 6 (Zn), 20 (EDO)
Water	65	105	891
B-factors			
Protein	46.4 (A), 46.7 (B), 51.4 (C)	33.7 (A), 32.7 (B), 32.9 (C), 35.1 (D)	20.6 (A), 23.1 (B), 20.0 (C), 23.1 (D), 22.0 (E), 18.5 (F)
Ligand	41.2 (Ca), 53.8 (i-NC4-A), 45.1 (i-NC4-B), 55.0 (i-NC4-C), 71.5 (i-NC4-A*), 61.0 (i-NC4-B*), 63.7 (EDO)	36.6 (Ca)	29.6 (Ca), 18.4 (κ-NC2), 19.5 (Cl), 22.4 (Zn), 28.2 (EDO)
Water	24.8	19.0	20.2
r.m.s.d			
Bond lengths (Å)	0.009	0.019	0.019
Bond angles (°)	1.402	1.965	1.975
Ramachandran (%)			
Preferred	95.0	96.6	96.0
Allowed	5.0	3.4	4.0
Disallowed	0	0	0
C-beta outliers	0	1	0

	PsS1_19B C77S	PsS1_NC	PsS1_NC C84S
	i-NC4 complex	Native data set	i-NC4 complex
Data Collection			
Beamline	In-house	CLS 08ID-1	In-house
Wavelength (Å)	1.5418	0.97869	1.5418
Space Group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Cell Dimensions			
a, b, c (Å)	57.72, 102.44, 171.21	57.47, 102.09, 170.78	57.72, 102.44, 171.21
Resolution (Å)	30.00-2.00 (2.03-2.00)	49.72-1.45 (1.47-1.45)	50.25-2.25 (2.29-2.25)
R _{merge}	0.050 (0.087)	0.115 (0.284)	0.096 (0.292)
R _{pim}	0.023 (0.062)	0.067 (0.167)	0.042 (0.154)
CC1/2		0.995 (0.947)	
<I/σI>	25.6 (9.2)	10.8 (5.4)	14.6 (3.5)
Completeness (%)	99.6 (98.8)	100.0 (100.0)	95.3 (94.9)
Redundancy	3.7 (2.4)	7.3 (7.5)	4.0 (3.5)
No. of reflections	261,729	1,304,617	189,172
No. Unique	70,964	178,310	46,858
Refinement			
Resolution (Å)	2.00	1.45	2.25
R _{work} /R _{free}	0.19/0.23	0.15/0.17	0.21/0.25
No. of atoms			
Protein	3562 (A), 3547 (B)	3810 (A), 3813 (B)	3789 (A), 3793 (B)
Ligand	2 (Ca), 2 (Cl), 59 (i-NC4 A), 59 (i-NC4 B)	2 (Ca), 6 (Cl), 44 (EDO)	2 (Ca), 30 (i-NC4 A), 30 (i-NC4 B)
Water	346	748	203
B-factors			
Protein	23.8 (A), 27.1 (B)	10.8 (A), 9.7 (B)	32.1 (A), 32.5 (B)
Ligand	17.2 (Ca), 28.6 (Cl), 30.1 (i-NC4 A), 30.4 (i-NC4 B)	6.3 (Ca), 13.9 (Cl), 19.9 (EDO)	36.4 (Ca), 47.2 (i-NC4 A), 60.1 (i-NC4 B)
Water	25.1	17.5	28.6
r.m.s.d			
Bond lengths (Å)	0.019	0.022	0.013
Bond angles (°)	1.986	2.073	1.594
Ramachandran (%)			
Preferred	96.5	98.0	97.6
Allowed	3.5	2.0	2.4
Disallowed	0	0	0
C-beta outliers	4	4	0

	PsS1_NC C84S	BovGH42	BovGH42
	κ -NC2 complex	Native data set	Iodide derivative
Data Collection			
Beamline	SSRL B11-1	SSRL B11-1	SSRL B11-1
Wavelength (Å)			
Space Group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Cell Dimensions			
a, b, c (Å)	57.69, 101.79, 170.54	60.25, 109.68, 126.1	60.54, 110.16, 125.61
Resolution (Å)	39.33-1.66 (1.75-1.66)	39.21-1.95 (2.00-1.95)	39.14-2.38 (2.47-2.38)
R _{merge}	0.076 (0.603)	0.134 (0.866)	0.238 (1.199)
R _{pim}	0.028 (0.244)	0.054 (0.417)	0.066 (0.382)
CC1/2	0.999 (0.814)	0.997 (0.546)	0.994 (0.689)
<I/σI>	20.5 (3.1)	11.7 (1.8)	10.3 (2.1)
Completeness (%)	99.9 (99.2)	99.6 (94.8)	99.6 (96.4)
Redundancy	8.3 (6.9)	7.1 (5.0)	14.0 (10.2)
No. of reflections	989,254	436,620	479,400
No. Unique	119,231	61,382	34,182
Refinement			
Resolution (Å)	1.66	1.95	
R _{work} /R _{free}	0.18/0.18	0.18/0.21	
No. of atoms			
Protein	3806 (A), 3804 (B)	5826	
Ligand	2 (Ca), 26 (κ -NC2 A), 26 (κ -NC2 B), 8 (EDO), 15 (EPE)	28 (EDO)	
Water	669	415	
B-factors			
Protein	19.3 (A), 17.1 (B)	21.7	
Ligand	13.1 (Ca), 20.4 (κ -NC2 A), 17.5 (κ -NC2 B), 25.1 (EDO), 31.5 (EPE)	34.9 (EDO)	
Water	23.3	24.5	
r.m.s.d			
Bond lengths (Å)	0.022	0.012	
Bond angles (°)	2.034	1.482	
Ramachandran (%)			
Preferred	97.7	96.4	
Allowed	2.3	3.6	
Disallowed	0	0	
C-beta outliers	4	0	

Values for highest resolution shells are shown in parenthesis

Chapter 3: Identification of an *endo*-4S-I-carrageenan sulfatase, PsS1_19A

Adapted from: The Molecular Basis of Polysaccharide Sulfatase Activity and a Nomenclature for Catalytic Subsites in this Class of Enzyme

Hettle, A.G., Vickers, C., Robb, C.S., Liu, F., Withers, S.G., Hehemann, J-H., Boraston, A.B. (2018). Structure 26, 1-12.

Contributed all research except isolation of PS47 and genomic DNA extraction of PS47, and ¹H-NMR data.

3.1 Introduction

Sulfatases from *P. atlantica*, *Z. galactanivorans*, and *P. carrageenovora* have been reported as having *endo*-4S-k/I-carrageenan sulfatase (Préchoux et al., 2013, 2016) activity from the former and *endo*-4S-I-carrageenan sulfatase (Ficko-Blean et al., 2017; Genicot et al., 2014) activity from the latter two organisms, respectively. These *endo*-acting carrageenan sulfatases are the only ones reported in the literature to date. However, the size and sequence similarity of these three sulfatases, *P. atlantica*, *Z. galactanivorans*, and *P. carrageenovora* with that of PsS1_19A varies quite drastically at 65.6 %, 52.7 %, and 21.6 % identity, respectively. This suggests not only structural differences but also differences in the molecular interactions involved in substrate recognition even though the two substrates of these enzymes only differ structurally by a single sulfate substituent. When BLASTed against the SulfAtlas database, the sulfatases from *P. atlantica* and *Z. galactanivorans* both align with sulfatases in the S1_19 subfamily; however, the sulfatase from *P. carrageenovora* does not return any significant hits. To date there are no structures of carrageenan active sulfatases in

complex with their natural substrate leaving the identity of the residues involved in substrate recognition largely unknown.

In this chapter I focus on elucidating the role of the second sulfatase in the PS47 genome, PsS1_19A, in the degradation of carrageenan and to draw comparisons to the sulfatase homologs from *P. atlantica* and *Z. galactanivorans*. These enzymes will be referred to as PaS1_19A and ZgS1_19, respectively, as these characterized sulfatases share the highest degree of sequence similarity to PsS1_19A (Figure 13). Based on the high degree of sequence conservation with PaS1_19A, a known *endo*-carrageenan sulfatase (Préchoux, Genicot, Rogniaux, & Helbert, 2013), I hypothesize PsS1_19A participates in the early stages of carrageenan depolymerization. I will use a combination of enzyme activity assays and mutagenesis to demonstrate that PsS1_19A is an *endo*-4S-I-carrageenan sulfatase. Using this enzymatic information I will use X-ray crystallography to determine the molecular interactions that confer this activity and how this recognition of the leaving group drives specificity. Using the information gathered from the structural studies I propose a subsite nomenclature for S1 family sulfatases.

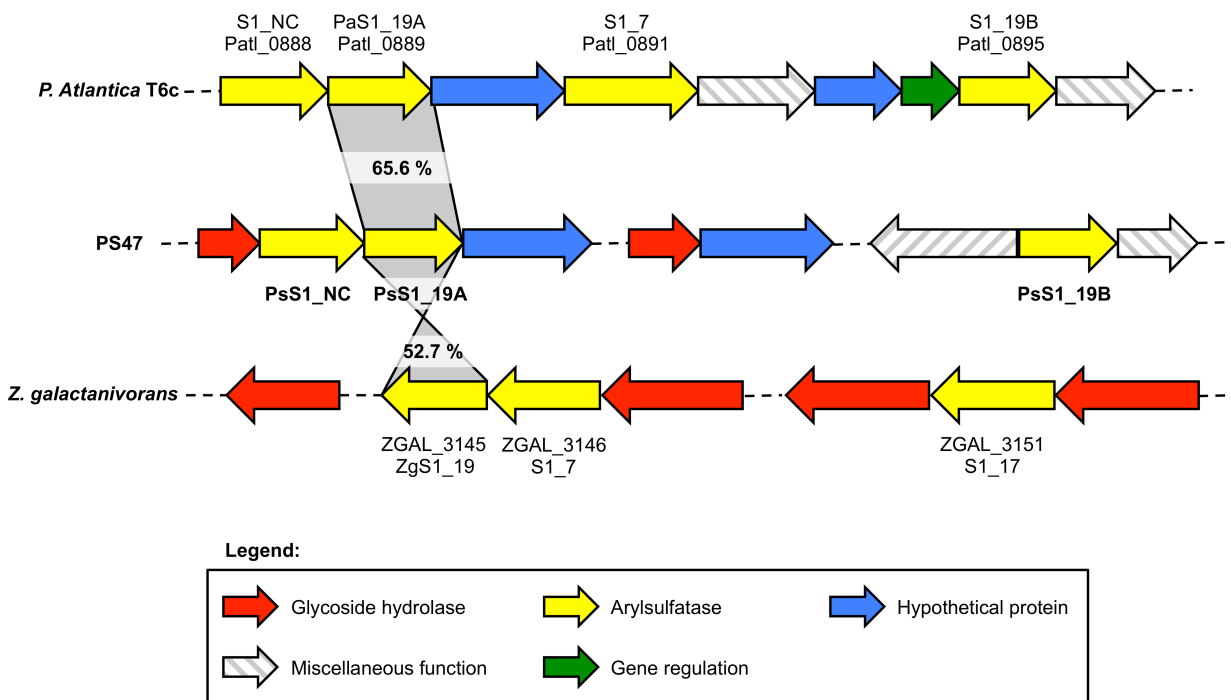


Figure 13. Known carrageenan PULs with characterized components. Genes are coloured according to protein function as per legend and percent identity of homologous genes with PsS1_19A are highlighted in gray.

3.2 Results

3.2.1 PsS1_19A is a 4S- κ /I-carrageenan sulfatase

Given the high sequence similarity between PsS1_19A and the published *endo*-acting I-carrageenan S1_19 sulfatases, PaS1_19A and ZgS1_19, a truncated construct omitting the N-terminal signal peptide of PsS1_19A was initially tested for activity against 4-nitrophenyl sulfate, a common synthetic substrate used for the identification of sulfatases; however, no activity was detected against this molecule. A library of commercially available κ -oligosaccharides, κ -I-hybrid oligosaccharides and in-house produced I-carrageenan oligosaccharides were also assayed for activity. In order to ensure that PsS1_19A was matured when overexpressed in *E. coli*, it was co-expressed with the FGE from *Mycobacterium tuberculosis* (Carrico et al., 2007). Sulfate removal from these oligosaccharides by PsS1_19A was assayed and visualized by fluorophore-assisted carbohydrate electrophoresis (FACE). Following treatment with PsS1_19A, the differently sulfated carrageenan oligosaccharides were labeled with a charged fluorophore and separated in a high percentage acrylamide gel (according to both their size and charge). The effect of PsS1_19A on different sized κ -oligosaccharides resulted in a heterogeneous population of differently sulfated products (Figure 14A). PsS1_19A displayed activity against κ -NC8, κ -NC6 and κ -NC4, as evidenced by the production of a second band in lanes 3, 5 and 7, respectively. This banding pattern was the result of the hydrolysis of an ester sulfate leading to a mixture of substrate and product oligosaccharides with differing charge to mass ratios and ultimately different migration distances. It was inconclusive using FACE whether PsS1_19A was active on κ -NC2 as it appeared that any resultant secondary band migrated with the background ANTS

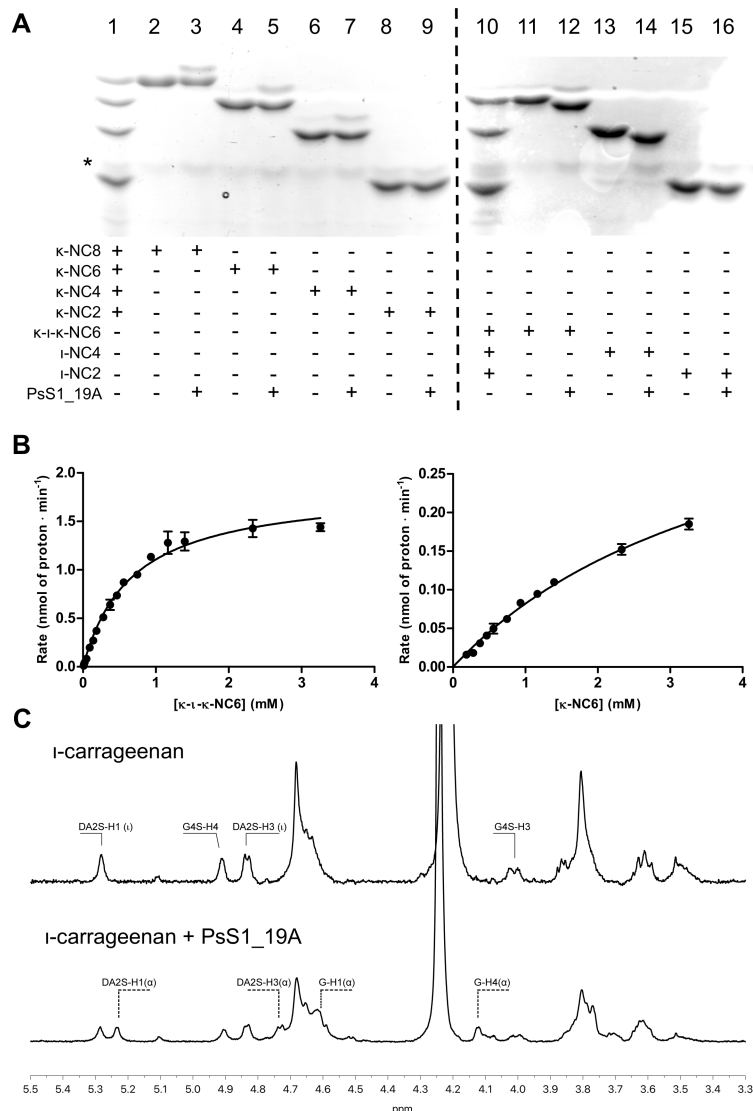


Figure 14. Activity of PsS1_19A from *Pseudoalteromonas* sp. PS47. (A) Face analysis of PsS1_19A on κ -NC8, κ -NC6, κ -NC4, κ -NC2, κ - ι - κ -NC6, ι -NC4, and ι -NC2. The presence or absence of polygeenan and enzyme are indicated below the gels. The asterisk indicates the position of free ANTS label (B) Kinetic analysis of PsS1_19A activity on κ - ι - κ -NC6 and κ -NC6 oligosaccharides at 25 °C using a pH-based colourimetric assay to detect the release of protons which accompanies sulfate ester hydrolysis. Error bars show the standard deviations of measurements made in triplicate. (C) ^1H -NMR spectra of full-length ι -carrageenan untreated (top panel) or treated with PsS1_19A (bottom panel). Signals with NMR chemical shifts corresponding with those distinctive of ι -carrageenan and α -carrageenan are labeled.

fluorophore and therefore could not be convincingly visualized. PsS1_19A not only showed activity on these relatively simple carrageenan oligosaccharides but also on more complex and heavily sulfated carrageenan oligosaccharides: κ -I- κ -NC6 and I-NC4. The resulting banding pattern with the more complex oligosaccharides showed a shift down upon PsS1_19A activity as seen in lanes 12 and 14 (Figure 14A) as opposed to the double banding patterns seen with the κ -oligosaccharides. There was no noticeable shift down in the I-NC2 + PsS1_19A lane or, if the result were to be a double banding pattern as seen with the κ -oligosaccharides, the secondary band would again be indistinguishable from the ANTS background. Therefore, the possibility of I-NC4 as a potential substrate was inconclusive using this method.

Due to the release of a proton upon sulfate ester hydrolysis, this reaction can be detected spectrophotometrically using a pH indicator. Using this pH sensitive colourimetric assay, we determined the K_M and apparent V_{max} values for PsS1_19A on κ -I- κ -NC6 to be $670.9 \pm 55.8 \mu\text{M}$ and $3.2 \times 10^{-5} \pm 1.0 \times 10^{-6} \mu\text{mol s}^{-1}$, respectively, while on κ -NC6 the K_M and apparent V_{max} values were $4,300 \pm 500 \mu\text{M}$ and $0.7 \times 10^{-5} \pm 0.01 \times 10^{-5} \mu\text{mol s}^{-1}$, respectively (Figure 14B). It must be noted that the post-translational modification efficiency of the *M. tuberculosis* FGE has only been assessed for recognition sequences when cloned onto protein constructs at either the N- or C-terminus (Carrico et al., 2007), and not for internal recognition sequences as found in PsS1_19A. Thus, we were unable to determine the concentration of matured PsS1_19A and therefore unable to report k_{cat} values. However, because the same protein preparation of PsS1_19A was used for both kinetic experiments the magnitudes of

$V_{\max}/K_M$ would be equal to the ratio of k_{cat}/K_M for the two substrates and showed a preference for κ -I- κ -NC6 roughly 30-fold higher than that of κ -NC6.

The FACE and pH assays both confirm activity of PsS1_19A on κ - and I-oligosaccharides. However, with the library of substrates being sulfated either internally or at the non-reducing end, we cannot with any degree of certainty say whether PsS1_19A is acting via an *exo*- or *endo*-mode of action. To address this question the activity of PsS1_19A was assessed on I-carrageenan by ^1H -NMR to provide insight into the mode of action, as well as the regioselectivity of the targeted sulfate group. The treatment of I-carrageenan with PsS1_19A produced new proton resonance signals indicating the formation of α -carrageenan when compared to reference signals of α -carrageenan (Figure 14C; results kindly provided by Feng Liu) (Préchoux et al., 2013). This result indicates that PsS1_19A is specific for the 4-sulfate group from the G4S residues in an *endo*-acting manner on full-length I-carrageenan, resulting in an α /I-carrageenan hybrid.

3.2.2 Architecture of an endo-4S- κ /I-carrageenan sulfatase

In order to elucidate the structural requirements for the specificity of PsS1_19A, we determined the X-ray crystal structure of the wild type PsS1_19A enzyme. PsS1_19A was solved at 2.80 Å resolution by molecular replacement using the sulfatase from *P. aeruginosa* (PAS) (PDB ID code: 1HDH) as a search model (Boltes et al., 2001). The final refined model contained three molecules in the asymmetric unit with each monomer being related via C3 non-crystallographic symmetry (Figure 15A) and each polypeptide chain making complimentary interactions with each of the

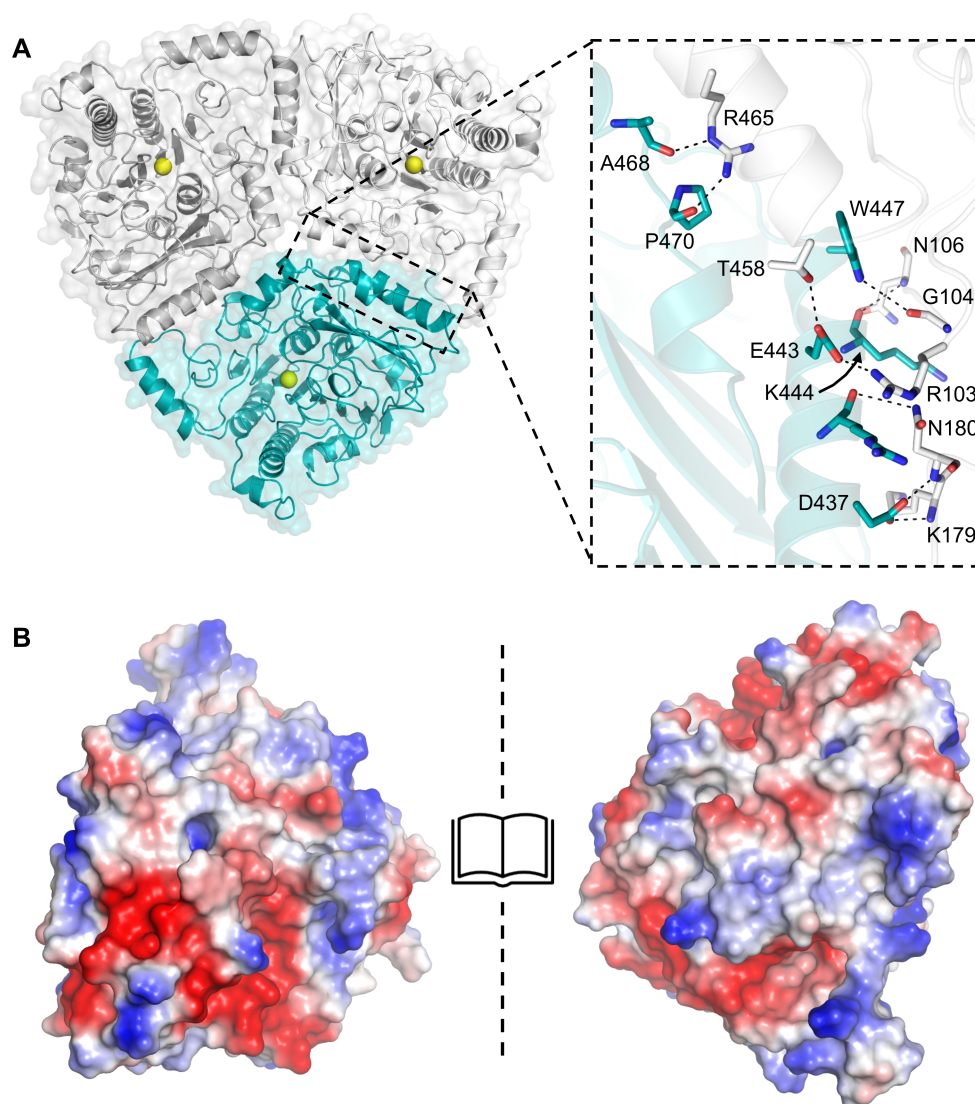


Figure 15. Molecular details driving trimerization of PsS1_19A.
 (A) Cartoon representation of the PsS1_19A trimer with each monomer composing the trimer coloured teal, white, and gray⁷⁰. Calcium ions are shown as yellow spheres. Inset showing the molecular interactions at the monomer interfaces. Amino acid residues represented as sticks and coloured according to their respective chains, hydrogen bonds represented as dashed lines.
 (B) Electrostatic surface representation of the interface between chains opened along the vertical axis (shown by dashed line) to expose the interacting surfaces.

neighbouring chains. The majority of the residues involved create an extensive hydrogen bond network. Slightly removed from this main interface, R465 makes a bidentate interaction with the carboxyl groups of A468 and P470. The buried surface areas at these interfaces when assessed by PISA analysis (Krissinel and Henrick, 2007) are $\sim 910 \text{ \AA}^2$ suggesting this may be a biologically relevant trimeric quaternary structure. When coloured by electrostatic potential this interface highlights the complementarity of these regions (Figure 15B). To assess this oligomeric state of PsS1_19A in solution dynamic light scattering analysis of PsS1_19A in solution was used. Two different concentrations of PsS1_19A (0.6 and 1.2 mg/mL) had molecular weights of $161 \pm 13 \text{ kDa}$ and $163 \pm 3 \text{ kDa}$, respectively, supporting the trimeric quaternary structure seen as being biologically relevant. Each monomer within the trimer contains the domain architecture typical of S1 sulfatases, all of which have a coordinated calcium ion found near the active site cleft. The larger N-terminal sub-domain consists of an α/β topology containing a meandering 8-stranded β -sheet in which all strands are parallel except strand $\beta 7$. All strands are surrounded by α -helices (Figure 16A). The smaller C-terminal sub-domain contains a four-stranded anti-parallel β -sheet tightly packed against the N-terminal domain by three α -helices.

Given the lack of crystallographic complexes of sulfatases and their respective physiological substrates, previous studies have relied heavily on sequence alignments and limited apo-enzyme structural work to identify 10 highly conserved polar residues involved in sulfate recognition, FGly activation, and sulfate ester hydrolysis (Hanson et al., 2004). We utilized a similar approach to help identify these 10 conserved residues in PsS1_19A. Structural alignments of PsS1_19A with PAS in complex with a sulfate ion

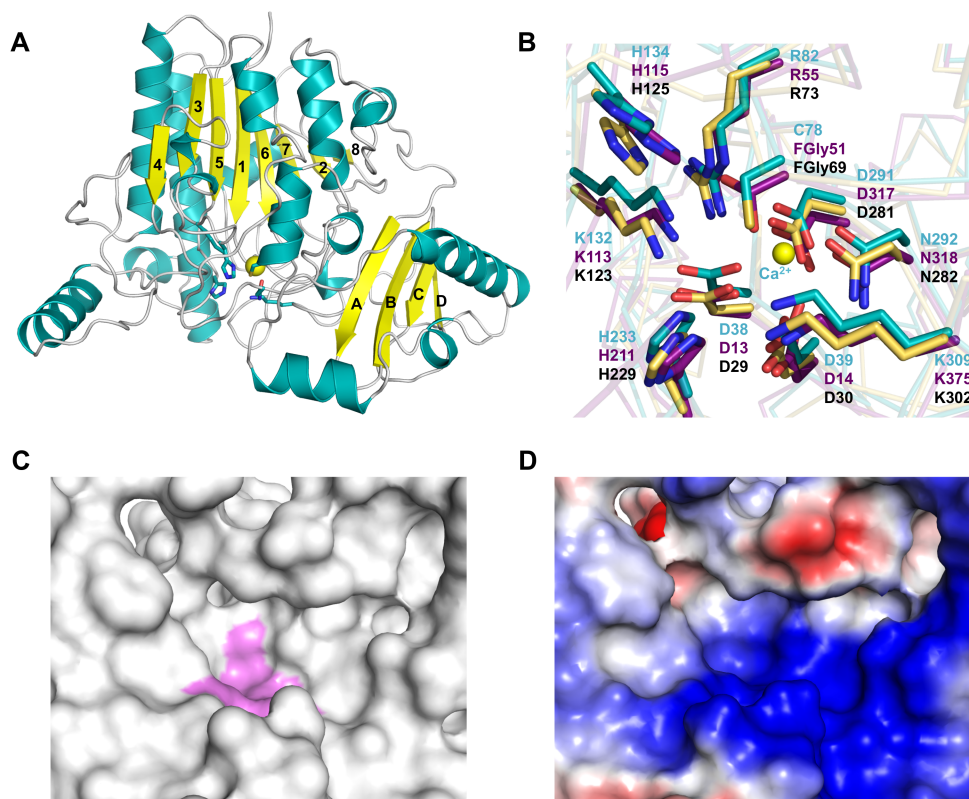


Figure 16. Structural architecture of PsS1_19A.

(A) PsS1_19A as characterized by the two subdomains of mixed α/β topology (α helices coloured teal and β sheets coloured yellow). N-terminal meandering β sheets are number 1-8 and the C-terminal subdomain anti-parallel β -sheet lettered A-D. Residues comprising the S-subsite are represented as teal sticks and the calcium ion as a yellow sphere.

(B) Structural overlay of PsS1_19A (teal) with both PAS (1HDH) (deep purple) and ASA (1AUK) (yellow orange) highlighting the ten conserved residues found within the S1 family. Only the calcium ion from PsS1_19A structure is represented, shown as a yellow sphere.

(C) and (D) A close-up of the solvent accessible surface of PsS1_19A active site groove showing the location of the catalytic center of the enzyme coloured violet (C) or as the electrostatic surface representation (D), where blue and red colouring represent basic and acidic patches, respectively.

(Boltes et al., 2001) and ASA (Human sulfatase A, PDB ID code: 1AUK) in complex with pNCS (von Bülow et al., 2001) were produced (Figure 16B). The PsS1_19A protein that was used for crystallization was produced in the absence of FGE co-expression and the proto-catalytic residue C78 did not show any evidence of maturation into the FGly residue. This indicates that the unidentified third sulfatase maturation system found in *E. coli* was unable to post-translationally modify and mature this protein. These alignments do, however, suggest that C78 is the proto-catalytic residue in PsS1_19A that undergoes post-translational modification to the catalytically active FGly. The other core-conserved residues in PsS1_19A are D38, D39, D291 and N292, which are involved in metal coordination; R82 and H134, which help stabilize FGly; and K132, H233 and K309, which are involved in sulfate recognition and activation (Figure 16B). This region identified by the overlays indicates the location of the active site, which is composed of a region within an extended cleft on the surface of each monomer (Figure 16C). This cleft shows largely basic characteristics, as is expected for recognition of a highly charged, sulfated substrate (Figure 16D).

3.2.3 The structure of PsS1_19A in complex with intact κ -carrapentaose

In an effort to achieve a structure of PsS1_19A in complex with intact substrate we sought to inactivate the enzyme by site-directed mutagenesis. Based on the structural data described above, C78 was mutated to a serine to ensure that it could not be post-translationally modified to the catalytic FGly. We proceeded to crystallize PsS1_19A C78S and obtained the same crystal form as for the apoenzyme. These mutant crystals were then soaked with κ -C5. The 2.15 Å resolution structure showed a

single intact κ -C5 molecule with continuous density spanning the catalytic machinery in each of the monomers composing the PsS1_19A C78S trimer (Figure 17A), with the 4-sulfate on the internal galactose unit coordinated by the amino acid residues conserved among S1 sulfatases (Figure 17B). The serine mutation at C78 mimics in size what would be the FGly residue in matured S1 sulfatases and sits ~ 3.4 Å beneath the targeted sulfate atom. H233, being ~ 3.2 Å from the scissile bond, is the catalytic acid protonating the oxygen of the sulfate ester bond. Interestingly, PsS1_19A utilizes N99 in sulfate coordination and activation, an extra residue that was not conserved amongst the sulfatases originally used to identify core polar residues involved in sulfate ester catalysis.

The bound κ -C5 substrate adopted a curved shape and this helical conformation was complemented by the active site groove (Figure 17C). Each sugar unit of the substrate makes interactions with the active site, both proximally and distally to the targeted sulfate ester. The coordination of the κ -C5 substrate is composed of a large network of direct and water-mediated hydrogen bonds (Figure 17D). Even though the G4S residues at the terminus of the substrate were sulfated, neither of these sulfates participated in any interactions with the protein suggesting that recognition of this substrate is independent of any sulfate esters on distal G4S residues.

Protein-carbohydrate recognition quite often utilizes CH- π interactions between aromatic amino acid side chains and the carbohydrate ring structures. Coordination of the κ -C5 substrate makes some of these common protein-carbohydrate interactions; for example, the reducing end terminal G4S unit sits above the phenyl ring of F182. However, the presence of DA units causes some unusual interactions. The side chain of

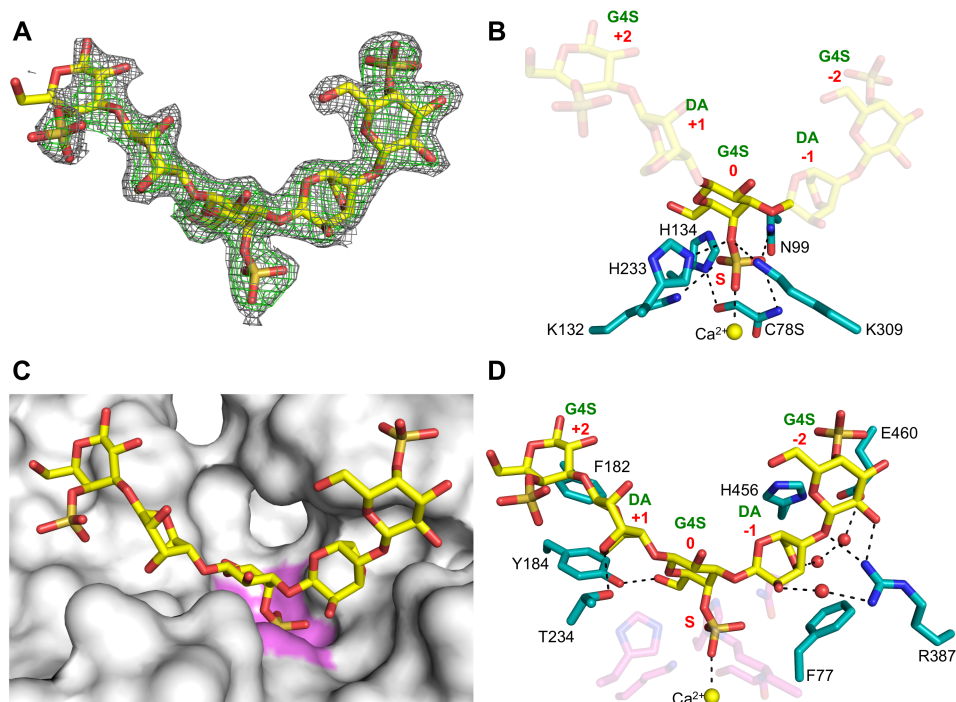


Figure 17. PsS1_19A C78S in complex with κ -C5.

(A) Representative electron density for κ -C5 modeled into the active site. The gray mesh shows the electron density map as a maximum likelihood/ σ_a -weighted $2F_o - F_c$ map contoured at 1.0σ . The green mesh shows the electron density map as a maximum likelihood/ σ_a -weighted $F_o - F_c$ map (contoured at 3.0σ) produced by refinement with the κ -C5 atoms omitted.

(B) The interaction of the central G4S residue in κ -C5 with the sulfate-binding S-subsite.

(C) The conformational complementarity of κ -C5 with the extended active site of the PsS1_19A C78S mutant. The solvent accessible surface is coloured in gray with the area comprising the S-subsite coloured in violet.

(D) The specific interactions of κ -C5 with the extended active site. Residues forming the S-subsite are represented as transparent violet sticks and residues comprising the remaining subsites are represented as teal sticks.

In all panels, κ -C5 is shown as yellow sticks. In (B) and (D), the calcium ion is represented as a yellow sphere, water molecules as red spheres, and hydrogen bonds are shown as dashed lines. Subsites are labeled in red, and sugar residues labeled in green. Relevant interacting side chains on the protein are shown as teal sticks.

Y184 coordinates only C5 (carbon-5) and C6 of the DA unit preceding the reducing end G4S unit (Figure 17D). The coordination of the other DA unit, at the non-reducing end of the substrate, makes some more uncommon interactions with the protein. The 3,6-anhydro ring of the DA unit is coordinated by H456 and F77 with C5 and C6 being coordinated by the ring of H456, while O6 and C3 are positioned by the phenylalanine ring of F77. These interactions are thus creating a cradle for the 3,6-anhydro ring of the DA unit (Figure 17D). Along with H456, residues R387 and E460 participate in coordination of the non-reducing end DA2S-G4S residues either directly or through water bridges showing participation of the C-terminal subdomain in substrate recognition and allowing for interactions with longer polymers.

3.2.4 The recognition of ι -carrageenan by PsS1_19A

The complementarity between the active site groove and the conformation of the κ -C5 substrate highlights the molecular interactions utilized by PsS1_19A to target the 4-sulfate group on G4S that is characteristic of the κ - and ι -carrageenan families. However, as seen in the kinetic assay and FACE experiments, PsS1_19A showed greater activity when the ι -moiety was present, namely a DA2S unit. To investigate the molecular interactions that contribute to this preference, we solved the x-ray crystal structure of PsS1_19A in complex with κ - ι - κ -NC6 at 2.50 Å resolution by employing the same techniques as for the κ -C5 complex. The κ - ι - κ -NC5 was coordinated with continuous density spanning the catalytic machinery, with the internal 4-sulfate on the G4S residue poised for nucleophilic attack in much the same manner as seen with the κ -C5 complex (Figure 18A).

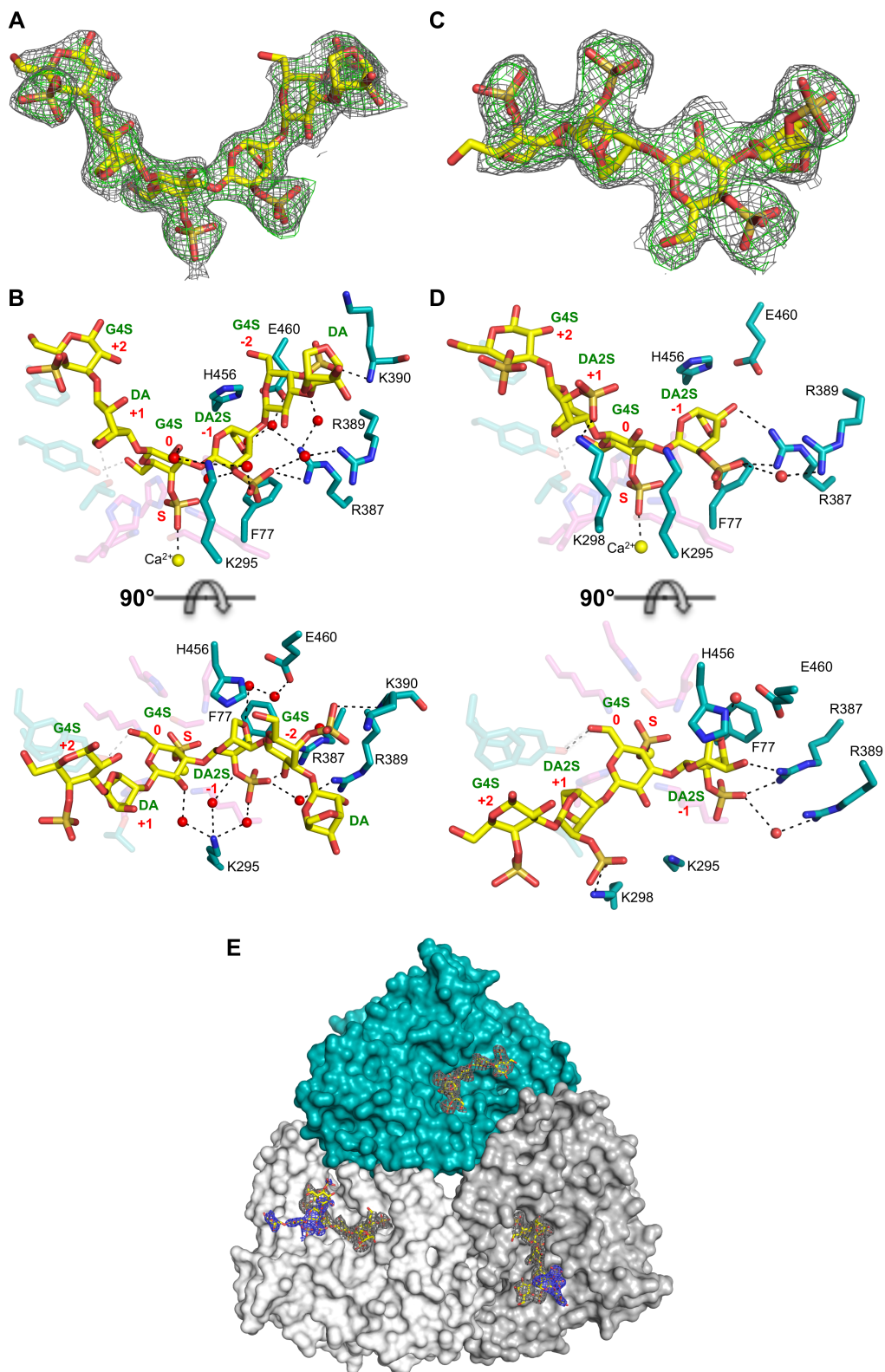


Figure 18. PsS1_19A C78S in complex with polygeenan of different sulfation patterns.
Figure legend on following page.

Figure 18. PsS1_19A C78S in complex with carrageenan oligosaccharides with different patterns of sulfation.

(A) Representative electron density for κ -I- κ -NC6 modeled into the active site.

(B) The κ -I- κ -NC6 bound in the active site highlighting the DA2S unit in the -1 subsite.

(C) Representative electron density for I-NC4 modeled into the active site.

(D) The I-NC4 bound in the active site of PsS1_19A C78S highlighting the DA2S unit in the -1 and +1 subsites.

In (A) and (C), the gray mesh shows the electron density map as a maximum likelihood/ σ_a -weighted $2F_o-F_c$ map contoured at 1.0 σ . The green mesh shows the electron density map as a maximum likelihood/ σ_a -weighted F_o-F_c map (contoured at 3.0 σ) produced by refinement with the κ -I- κ -NC6 and I-NC4 atoms omitted. In (B) and (D), the bound κ -I- κ -NC6 and I-NC4 are shown as yellow sticks, interacting amino acid side chains as teal sticks, residues in the S-subsite as magenta sticks. The calcium ion is shown as a yellow sphere, water molecules as red spheres, and hydrogen bonds are represented as dashed lines. Subsites are labeled in red and sugar residues are labeled in green.

(E) X-ray crystal structure of PsS1_19A C78S trimer in complex with κ -I- κ -NC6. The solvent accessible surface represented in teal, white, and gray80 depending on the monomer within the trimer, κ -I- κ -NC6 and I-NC4 represented as yellow sticks. The gray and blue mesh represents the $2F_o-F_c$ maps produced from phases generated by refinements with the κ -I- κ -NC6, I-NC4, and I-NC2 included at 1.0 σ , respectively.

The two main differences between κ -I- κ -NC6 and κ -C5 are the presence of a DA2S unit four residues from the reducing end of the oligosaccharide, and an extra DA unit at the non-reducing end. This non-reducing end DA unit did not interact with any amino acid side chains but the natural conformation of the carrageenan helix kept the DA unit ordered enough to model into electron density. PsS1_19A C78S coordinates the internal κ -I- κ -NC6 DA2S unit, the defining unit in I-carrageenan, in the same way as the DA unit of κ -C5; however, the sulfate group of the DA2S unit interacts directly with the side chain of R387 and indirectly with K295 and R389 via a water bridge (Figure 18B). The extra sulfate recognition of the DA2S unit has an impact on enzyme kinetics and suggests that even though PsS1_19A does not interact with any other 4-sulfate of G4S units outside of the targeted substituent, it may have a further dependency on 2-sulfate groups on other DA2S units.

To elucidate any potential sulfatase-sulfate interactions with 2-sulfate substituents of I-carrageenan DA2S units not already accounted for by the κ -I- κ -NC6 complex, we enzymatically produced I-carrageenan oligosaccharides using a GH16 *endo*-carrageenase from *Bacteroides ovatus* (BovGH16). The resulting I-NC4 product was purified and soaked into the catalytic mutant crystals and we determined the structure of PsS1_19A C78S in complex with I-NC4 at a resolution of 2.84 Å. A single intact I-NC4 molecule with continuous density occupied the active site of each monomer within the PsS1_19A trimer (Figure 18C). The molecular interactions utilized in the coordination of the G4S-DA2S motif at the non-reducing end were identical to that in the coordination of the analogous G4S-DA2S motif found internally in the κ -I- κ -NC6 complex (Figure 18D). The galactose backbone of the reducing end G4S-DA2S motif of

I-NC4 is coordinated using the F182 and Y184 aromatic platform similarly to the reducing end G4S-DA motif of κ -I- κ -NC6. The extra 2-sulfate of this G4S-DA2S motif is within ~ 3.2 Å of the terminal nitrogen of the K298. It is difficult to determine the importance of this interaction, given the lower resolution of the dataset and some disorder of the amino acid side chain; however, when comparing this G4S-DA2S to κ -carrageenan motifs, which lack the 2-sulfate, the possibility of an interaction cannot be disregarded and may possibly play a role in the observed increased PsS1_19A activity.

3.3 Discussion

The identification of FGly-dependent sulfatases has become possible based on the prediction of conserved signature motifs being quite reliable for the S1 family sulfatases. However, with the large diversity of sulfated biomolecules, one of the major challenges has proven to be matching the sulfatase to its respective biological substrate. The use of synthetic aryl-sulfate substrates to confirm sulfatase activity has historically led to the generic label of 'arylsulfatase' for these enzymes. As PsS1_19A showed no observable activity against pNP-sulfate, indicating that the recognition of the sulfate group is not enough to confer catalysis, and, therefore recognition of the leaving group is necessary for substrate specificity and substrate turnover. PsS1_19A showed activity on κ - and ι -polygeenan with a degree of polymerization larger than two. Through ^1H NMR analysis of PsS1_19A activity on ι -carrageenan we were able to show the specificity of PsS1_19A as an 4S- ι -carrageenan sulfatase; the same specificity that was reported for the PaS1_19A homolog (Préchoux et al., 2013). PsS1_19A displayed activity on κ -carrageenan motifs, yet this activity was greatly enhanced in the presence of the 2-sulfate group on neighbouring DA2S residues in ι -carrageenan motifs. The X-ray crystal structures of PsS1_19A C78S in complex with κ - ι - κ -NC6, κ -C5, and ι -NC4 all show coordination of the 4-sulfate of the G4S residue found internally in the differing oligosaccharides. These functional and structural studies of PsS1_19A support an *endo*-mode of sulfatase activity; furthermore, the extensive network of interactions involved in the recognition of the galactan backbone further support the notion that recognition of the leaving group, in this case carrageenan, is crucial in discriminating between sulfated biomolecules.

With the recent report of sulfatase structures in complex with their physiological substrates, there is a need to establish a binding site nomenclature in order to communicate accurately and concisely the nature of the substrate-enzyme interactions. Here we propose a sugar binding subsite nomenclature for sulfated carbohydrates that is adapted from that used to describe the binding sites of GHs, in which the numbering is dependent on which glycosidic linkage is being hydrolyzed (Davies et al., 1997). Like with the GH subsite nomenclature, ours is dependent on the linkage being cleaved, either the sulfate ester or sulfamate groups; for simplicity, we will focus on sulfatases active on sulfate esters. We propose that the subsite accommodating the targeted sulfate group be known as the “S subsite” (For Sulfate-binding subsite). The carbohydrate portion to which the sulfate ester is bound and that is the target for hydrolysis we propose to be the 0 (zero) subsite. With the 0 subsite being a point of reference, additional subsites will be numbered increasing negatively towards the non-reducing end (-1, -2, -3, etc.) and increasing positively towards the reducing end (+1, +2, +3, etc.). This nomenclature cannot only be utilized for *endo*-acting carbohydrate sulfatases (like PsS1_19A), but also *exo*-acting carbohydrate sulfatases that are specific for regions found at the termini of the carbohydrate polymers and sulfatases that recognize other sulfated and sulfamated biomolecules (Figure 19).

The two sulfatases from *B. thetaiotaomicron* are both *exo*-acting and specific for the non-reducing end sulfate group of heparin. As an example, our proposed subsite nomenclature can be applied to these two recently characterized *exo*-acting sulfatases for which X-ray crystal structures in complex with their physiological substrates have been reported. According to our sub-site nomenclature system, BT1596^{2S-sulf} (PDB ID

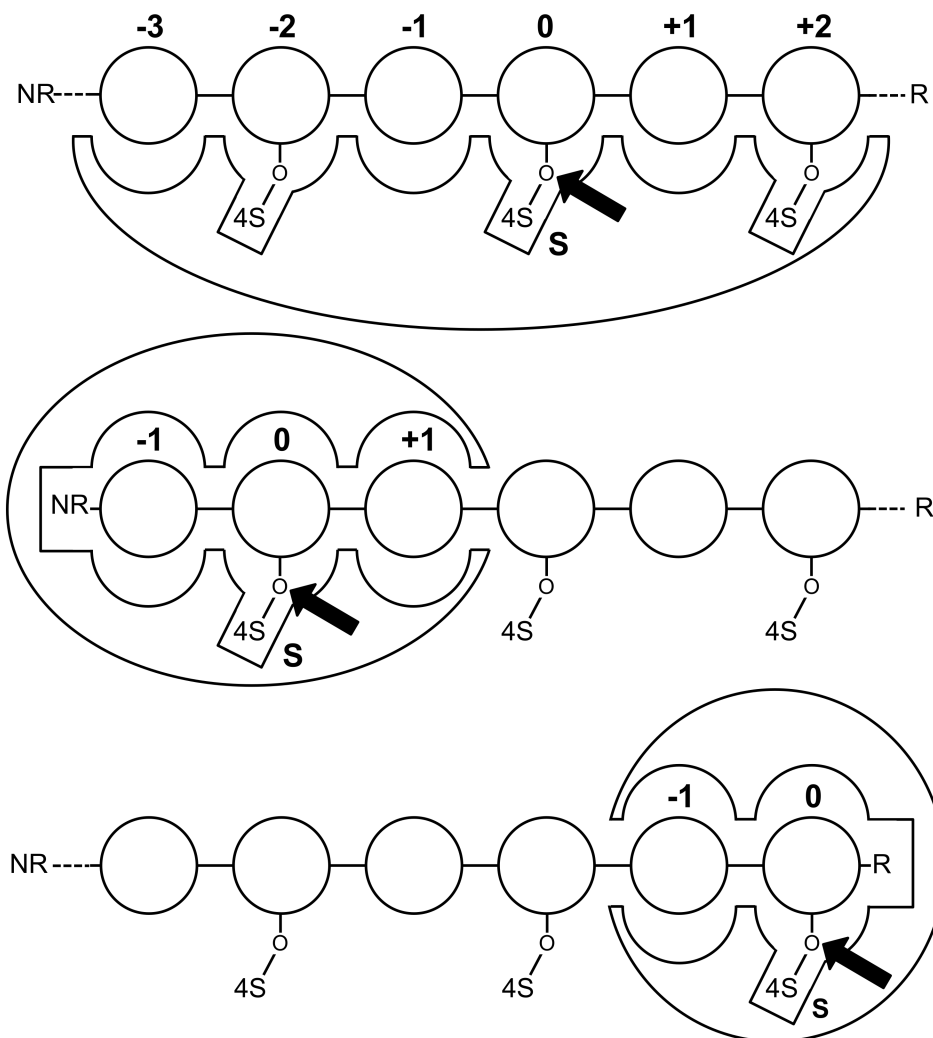


Figure 19. Proposed nomenclature for sugar binding subsites in carbohydrate active sulfatases.

Examples for hypothetical *endo*-acting sulfatases (top), non-reducing end *exo*-acting sulfatases (middle), and reducing end *exo*-acting sulfatases (bottom). Individual sugar residues are represented as sphere with the non-reducing (NR) and reducing (R) ends labeled. The S subsite is highlighted with an arrow indicating the scissile bond targeted, here it is represented as a sulfate ester. The 0 subsite coordinates the sugar residue with the targeted sulfate group substituent. Additional subsites are numbered increasingly positive toward the reducing end and increasingly negative toward the non-reducing end.

code: 5G2T) in complex with $\Delta 4,5\text{UA}2\text{S}\beta 1\text{-}4\text{GlcNS}6\text{S}$, with the 2-sulfate uronic acid at the non-reducing end coordinated in the active site, shows occupation of the S, the 0, and the +1 subsites (Figure 20A). Similarly, BT4656^{6S-sulf} is in complex with N,O6-disulfo-glucosamine (PDB ID code: 5G2V) coordinating the sulfate ester in the S subsite and with the glucosamine occupying the 0 subsite (Figure 20B). Both the *B. thetaiotaomicron* sulfatases have the pocket architecture of *exo*-acting enzymes recognizing the non-reducing end of the carbohydrates, and therefore there is no presence of any minus (-) subsites. It remains unclear as to whether BT1596^{2S-sulf} and BT4656^{6S-sulf} possess +2 and +1 subsites, respectively, based on the complexes reported.

As for PsS1_19A in complex with the $\kappa\text{-I-}\kappa\text{-NC6}$ oligosaccharide, we can define six subsites by this proposed nomenclature, including the S subsite. These subsites are occupied by the sulfate of a G4S unit in the S subsite with the G4S unit in the 0 subsite; DA and G4S residues occupy the +1 and +2 subsites, respectively. A DA2S residue and a G4S residue occupy the -1 and -2 subsites, respectively (Figure 20C). This active site architecture is typical of the *endo*-acting class of enzyme, in that it recognizes internal components of the polymer without interacting with the reducing or non-reducing ends; as seen in the X-ray crystal structure, the non-reducing end DA unit of the $\kappa\text{-I-}\kappa\text{-NC6}$ oligosaccharide is not making any specific interactions with PsS1_19A amino acid residues. This non-reducing end DA unit in the $\kappa\text{-I-}\kappa\text{-NC6}$ oligosaccharide does alter the overall conformation of the polysaccharide when compared to the adopted conformation of the $\kappa\text{-C5}$ oligosaccharide when in complex with PsS1_19A; however, this minor difference between the G4S residues in the -2 subsite does not alter the interactions

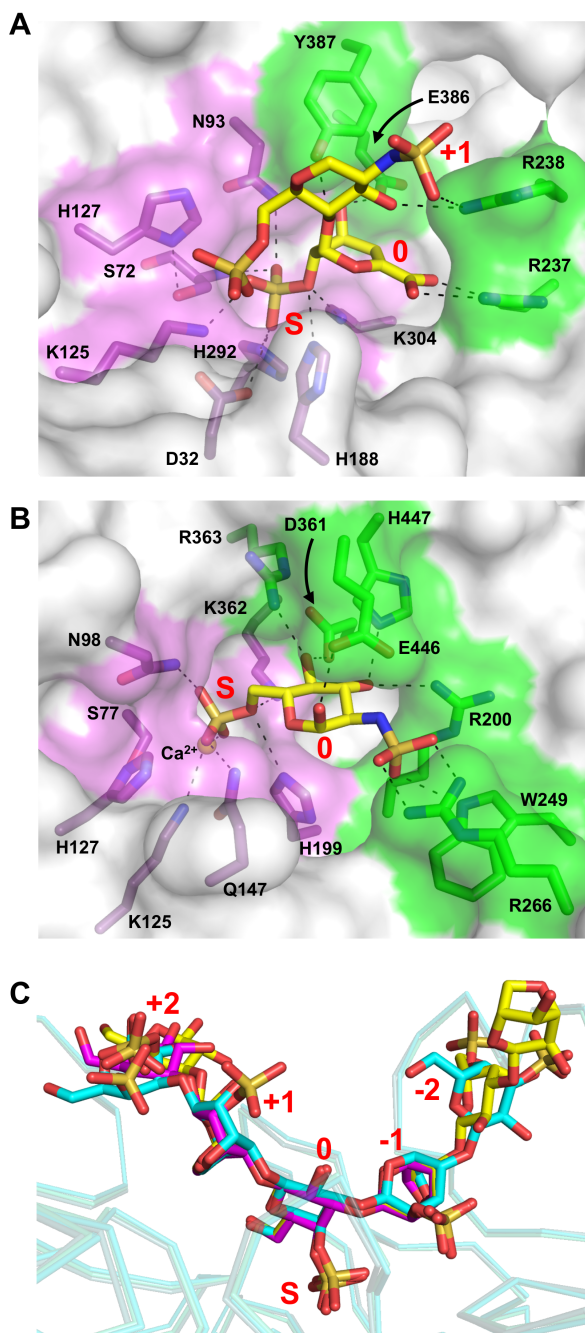


Figure 20. BT1596^{2S-sulf} and BT4656^{6S-sulf} complexes labeled using the proposed carbohydrate active sulfatase sugar binding subsite nomenclature.

(A) BT1596^{2S-sulf} in complex with $\Delta 4,5\text{UA}2\text{S}\beta 1\text{-4GlcNS6S}$ showing the 2-sulfate in the S-subsite.

(B) BT4656^{6S-sulf} in complex with N,O6-disulfo-glucosamine showing the 6S in the S-subsite.

In (A) and (B) $\Delta 4,5\text{UA}2\text{S}\beta 1\text{-4GlcNS6S}$ and N,O6-disulfo-glucosamine are represented as yellow sticks, residues comprising the S-subsites are represented as violet sticks, residues comprising the integer subsites are represented as green sticks. The solvent accessible surfaces of BT1596^{2S-sulf} and BT4656^{6S-sulf} are represented in gray apart from the surfaces composing the subsites, which are coloured as previously mentioned.

(C) Structural overlay of the PsS1_19A C78S complexes: $\kappa\text{-C5}$, $\kappa\text{-I-}\kappa\text{-NC6}$, and I-NC4 are shown as cyan sticks, yellow sticks, and magenta sticks, respectively. Subsites are labeled in red.

involved in coordination of the galactose pyranose ring. Furthermore, when assessing the conformations of each individual carrageenan oligosaccharide coordinated in the active site of PsS1_19A, it becomes apparent that the necessary degree of polymerization for catalysis must be greater than two as supported by the apparent disorder of carbohydrate residues found outside the +1, 0, and -1 subsites (Figure 20C), in conjunction with the biochemical kinetics data.

The SulfAtlas database classifies PsS1_19A into the subfamily S1_19 of which the only other two characterized members are PaS1_19A and ZgS1_19; both of these are *endo*-4S-I-carrageenan sulfatases. Homology models of PaS1_19A and ZgS1_19 were generated using the structure of PsS1_19A as a template with the Web server Phyre2 (Kelley and Sternberg, 2009). This indicates a high level of conservation in the active site subsites given the sequence similarities of 65.6 % and 52.7 % for PaS1_19A and ZgS1_19A to PsS1_19A, respectively (Figure 21A). However, a sequence alignment of 55 representative S1_19 sequence entries was used to try to identify conserved residues within the wider S1_19 sub-family by mapping the amino acid residue conservation onto the structure of PsS1_19A using the ConSurf Web server (Goldenberg et al., 2009). At first glance the active site groove of PsS1_19A compared to the S1 family does not seem to show any significant conservation of residues outside of the S subsite (Figures 21B and 21C). With very little information on the molecular requirements for sulfatase-substrate interactions, this residue conservation colouring highlights how little we know about sulfatases in general. The variability of residues in the subsites that we predict substrate specificity likely indicates the family is poly-specific, even within a single sulfatase subfamily, suggesting removal of other sulfate

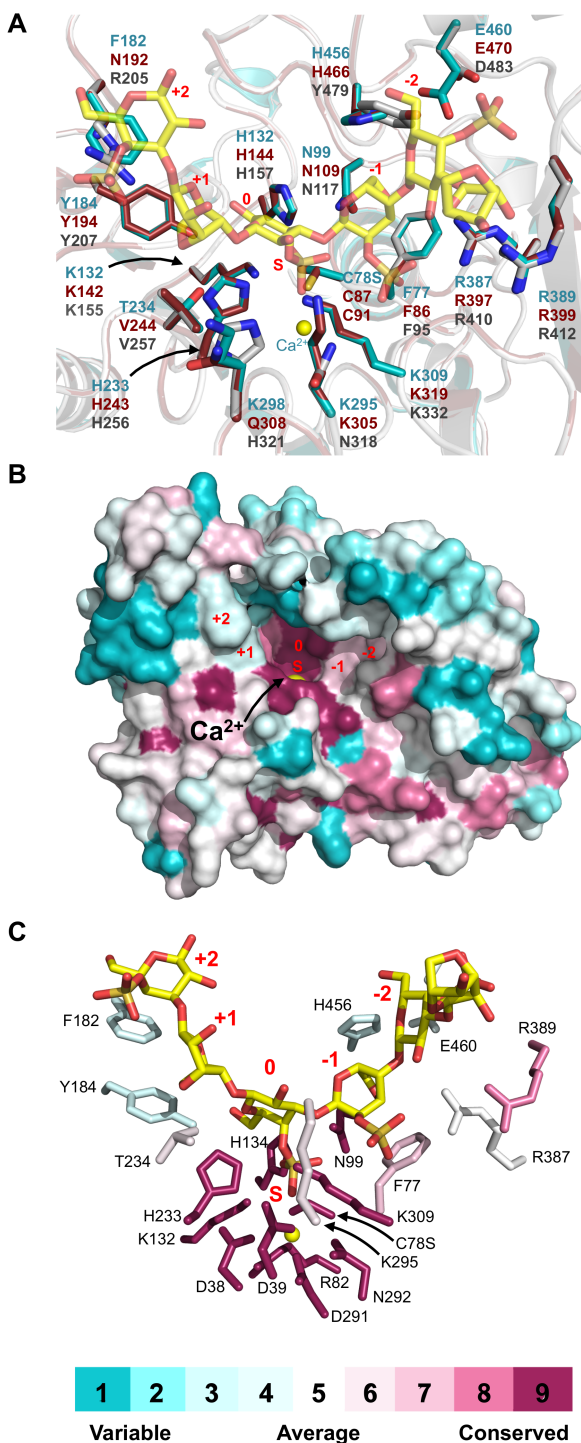


Figure 21. Residue conservation from S1_19 that are involved in substrate recognition.

(A) Homology model of PaS1_19A (ruby) and ZgS1_19 (white) based on the PsS1_19A template, overlaid with the PsS1_19A C78S (teal) κ -I- κ -NC6 substrate complex. The κ -I- κ -NC6 is represented as transparent yellow sticks.

(B) ConSurf model of S1_19 sequences (n=55) mapped onto the structure of PsS1_19A C78S in complex with κ -I- κ -NC6 (κ -I- κ -NC6 not shown).

(C) The conservation of active site residues of S1_19 as calculated with ConSurf represented as sticks.

In all panels the calcium atom is shown as a yellow sphere and subsites are labeled in red. The colour scheme representing degree of residue conservation for (B) and (C) is shown below the panels.

groups from carrageenan, the utilization of *exo*- instead of *endo*-lytic activity, or possibly activity on different sulfated polysaccharides, in agreeance with the hypotheses proposed by Barbeyron et al. (2016).

Carrageenans naturally occurring in nature exhibit tertiary and quaternary structures in a thermally reducible gel in the presence of counter ions. κ - and ι -carrageenans adopt right-handed double helices (Arnott et al., 1974), which aggregate upon cooling or in the presence of counter ions such as calcium and potassium. Serendipitously, the structure of PsS1_19A C78S in complex with κ - ι - κ -NC6 not only contained the hexasaccharide spanning the active site but we could also model an ι -NC2 and an ι - κ -NC4 stacked on top of the κ - ι - κ -NC6 bound in the active site of two of the monomers within the asymmetric unit (Figure 18E). Whether or not the active site κ - ι - κ -NC6 or the neighbouring ι - κ -NC4 molecule would be components of the same carrageenan double helix or neighbouring helices is unknown; however, these two oligosaccharides allow us to look at the differences in pitch of carrageenan double helix strands when their conformation is driven by physicochemical properties of the polysaccharide structure or if it is influenced by protein interactions and active site conformations. Remarkably, the conformation of the loosely bound ι - κ -NC4 when compared to the structure of ι -carrageenan reported in the literature shows a high degree of similarity (Figure 22A). This ι - κ -NC4 molecule can be extrapolated and the resulting coil would have a pitch of 26.30 Å, this value being very similar to the pitch of ι -carrageenan of 26.56 Å, suggesting that the serendipitously bound ι - κ -NC4 adopts the conformation expected for that of carrageenan.

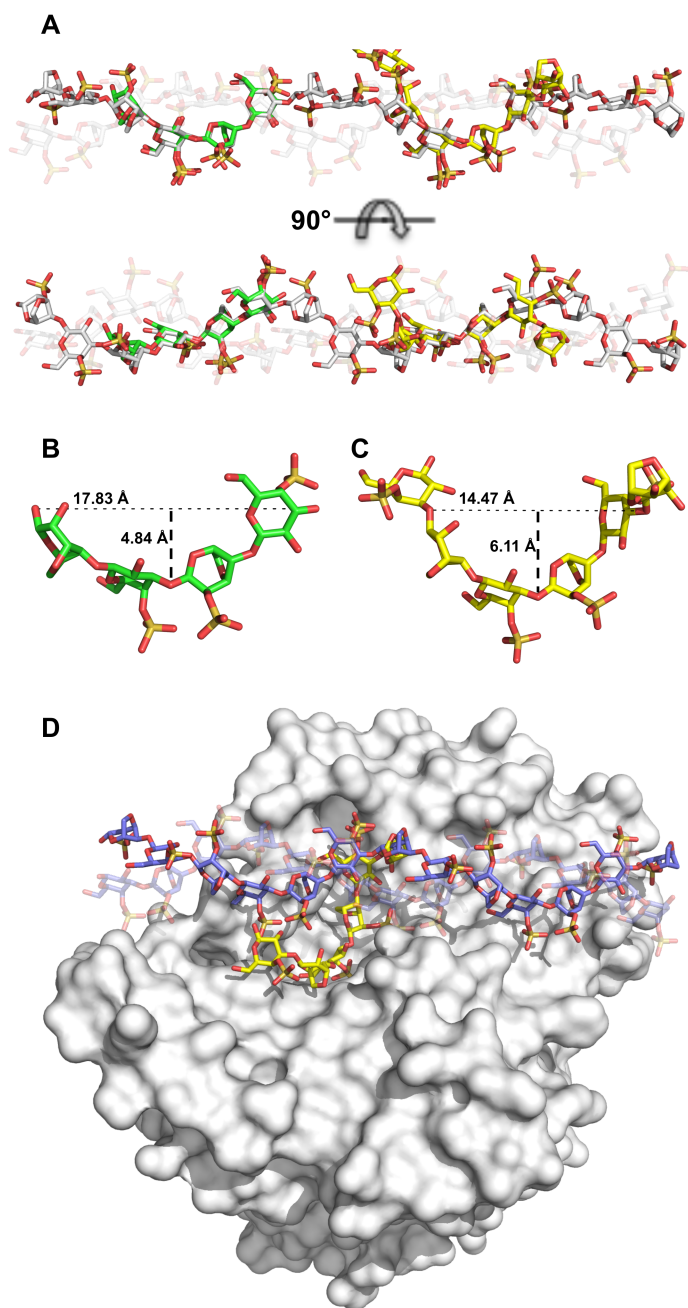


Figure 22. Comparison of ι -carrageenan double helical structure with the structures of the coordinated κ - ι - κ -NC6 and ι - κ -C4 oligosaccharides.

(A) Overlay of the active site bound κ - ι - κ -NC6 and the secondary ι - κ -C4 on the ι -carrageenan double helix. Extended model of ι -NC6 double helix (PDB ID code: 1CAR) is shown as grey sticks with one strand shown as transparent. ι - κ -C4 is shown as green sticks and κ - ι - κ -NC6 as yellow sticks.

(B) The ι - κ -C4 with distances between relevant points shown as dashed lines.

(C) The κ - ι - κ -NC6 with distances between relevant points shown as dashed lines.

(D) The solvent accessible surface of PsS1_19A shown in gray with the bound κ - ι - κ -NC6 represented as yellow sticks from the κ - ι - κ -NC6 PsS1_19A C78S substrate complex. The double-helical structure of ι -carrageenan is shown as purple sticks and was placed near the active site groove while avoiding clashes with the surface of PsS1_19A. The coordinated κ - ι - κ -NC6 shows a displacement into the active site relative to the modeled ι -carrageenan double helix.

The helical conformations adopted by the κ -I- κ -NC6 in the active site groove are quite different and when extrapolated the estimated pitch is ~ 18 Å, which is significantly shorter than that of I-carrageenan (Figure 22A). In comparing the distances between O1 of the DA unit in the +1 subsite and O3 of the G4S residue in the -2 subsite of the bound κ -I- κ -NC6 molecule to the distance between the analogous atoms in the loosely bound I- κ -NC4 molecule, we can see a compression of ~ 3 Å supporting the idea that the shorter pitch and increased curvature of the oligosaccharide results from the influence of the architecture of the PsS1_19A active site groove (Figure 22B and 22C). This notion is further supported by the compression of ~ 1.6 Å between the same atoms of the κ -C5 molecule in the PsS1_19A C78S in complex with κ -C5. Without this conformational distortion of the I-carrageenan helix it is unlikely that the targeted 4-sulfate groups of the G4S residues would be able reach the catalytic machinery given the architecture and depth of the active site groove (Figure 22D).

The biochemical characterization of PsS1_19A by FACE and H^1 -NMR, in conjunction with the structural characterization, indicates an *endo*-mode of action when desulfating and remodeling κ - and I-carrageenan. These results and conclusions point to the role of PsS1_19A within carrageenan degradation being in the initial steps of depolymerization. The ability of PsS1_19A to desulfate full-length carrageenan, as well as polygeenan, suggests that it either performs catalysis as the first enzyme to act in carrageenan degradation or in parallel with an *endo*-acting carrageenan GH from the PS47 carrageenan PUL, likely one or more of the three putative GH16s. One, or both, of these proposed steps would result in a differently-sulfated pool of α -, I-, β -, and κ

polygeenan oligosaccharides (Figure 23). These resulting pieces would go on to be further depolymerized by *exo*-acting sulfatases and *exo*-acting GHs.

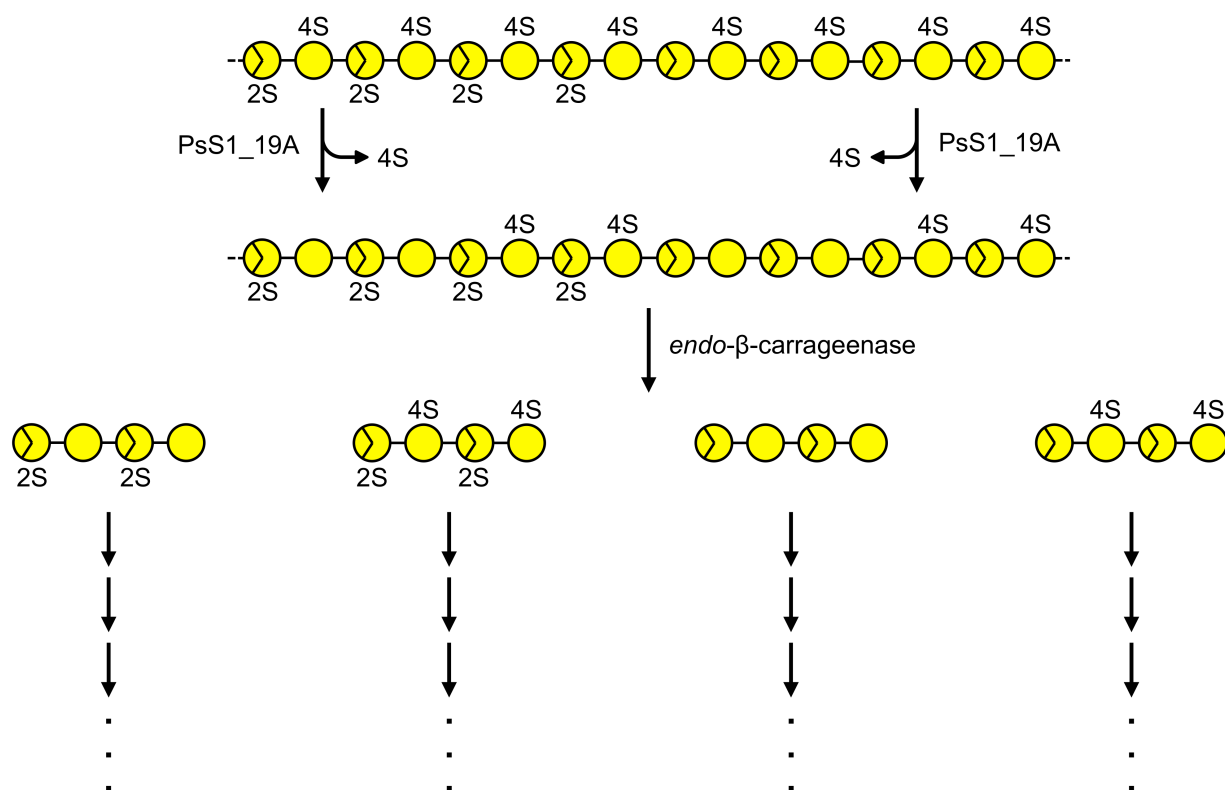


Figure 23. Cartoon schematic of products resulting from PsS1_19A sulfate ester hydrolysis.

Representation of carrageenan components are the same as in Figure 2.

Chapter 4: Identification of two *exo*-acting sulfatases from PS47: PsS1_19B and PsS1_NC

4.1 Introduction

In 1966, Weigle and Yaphe (1966) identified an intracellular sulfatase from *P. carrageenovora* that exhibited activity against κ -NC2 yet showed no observable activity on a galactose-4-sulfate monosaccharide, the G4S unit of κ -NC2. This enzyme was further characterized and determined to be an *exo*-4S- κ -carrageenan sulfatase specific for a single 4-sulfate group at the non-reducing end of κ -carrageenan (McLean and Williamson, 1979). This is unfortunately where research into *exo*-acting carrageenan sulfatases slowed significantly. Aside from the experimental difficulties involved with sulfatase research, the lack of research into *exo*-acting sulfatases is, in part, due to the fact that the rheological properties of carrageenan are only present in the full-length polymer, meaning there is little industrial interest or application for the small molecular weight polygeenan. However, from an environmental and carbon cycling standpoint, the understanding of how polygeenan is desulfated is critical to understanding how the photosynthetically fixed carbon trapped in carrageenan enters metabolism and is ultimately returned back into the global carbon cycle.

As described in Chapter 3, PsS1_19A is an *endo*-4S- κ /I-carrageenan sulfatase that, together with *endo*-acting GHs, would produce a pool of differently-sulfated polygeenan bearing 2- and/or 4-sulfate groups. Therefore, to continue the degradation of carrageenan and allow the resultant monosaccharides to enter metabolism, it would be expected that the PS47 carrageenan PUL encodes for one or more *exo*-acting sulfatases. The remaining uncharacterized sulfatases from the PS47 PUL (PsS1_19B

and PsS1_NC) share high amino acid sequence identity (>60 %) with sulfatases from the *P. atlantica* PUL, and lower identities with those from *Z. galactanivorans* (Figure 24). The sulfatases from *P. atlantica* have not been characterized; however, the *Z. galactanivorans* sulfatases ZgS1_17 and ZgS1_7 have been reported with specificities for ι-NC4 and κ-NC2, respectively (Ficko-Blean et al., 2017).

In this chapter, I focus on identifying the role PsS1_19B and PsS1_NC play in the desulfation of the polygeenan pool resulting from PS47 *endo*-carrageenase and PsS1_19A activity. I hypothesize that PsS1_19B and PsS1_NC perform their activity in an *exo*-active manner and will use enzymatic activity assays to demonstrate this. Furthermore, I will use X-ray crystallography to identify the molecular requirements of PsS1_19B and PsS1_NC specificity and how these interactions relate to the mode of activity.

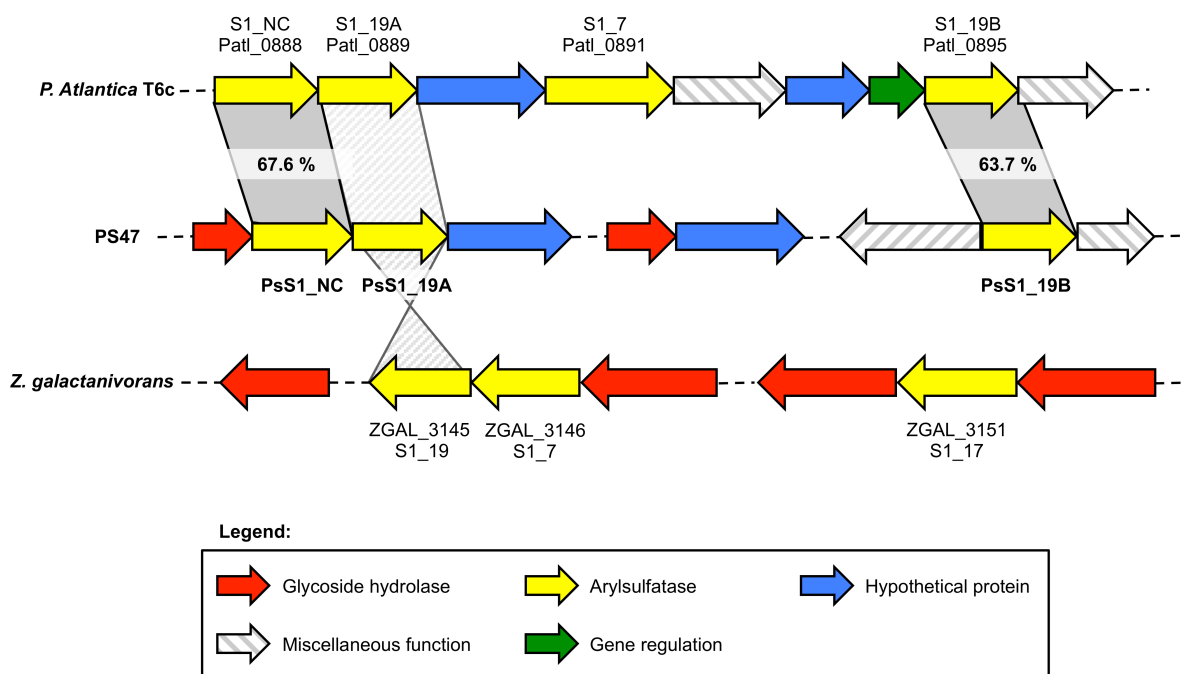


Figure 24. Known carrageenan PULs with characterized components. Genes are coloured according to protein function as per legend. Percent identity of PsS1_19B and PsS1_NC homologous genes are highlighted in gray. Light gray pattern showing homologous genes with PsS1_19A

4.2 Results

4.2.1 Identification of an κ -4S- κ -carrageenan sulfatase

In order to test our hypothesis that PsS1_NC and PsS1_19B are *exo*-acting carrageenan sulfatases, we cloned truncations of both genes, lacking the predicted signal peptides, into an expression vector and co-expressed them in *E. coli* with the FGE from *M. tuberculosis* as was done for PsS1_19A (Chapter 3). Following expression and purification, we assessed these enzymes for sulfatase activity against commercially available κ -NC2, κ -NC4, and in-house produced ι -NC2 and ι -NC4, by FACE. Enzymatic treatment of κ -NC2 and κ -NC4 with PsS1_19B resulted in products of different electrophoretic mobility as well as a reduction in the intensity of the substrate band (Figure 25A); however, PsS1_19B did not appear to have any activity on ι -NC2 or ι -NC4. Interestingly, PsS1_NC did not demonstrate activity on any carrageenan or polygeenan substrates (data not shown).

To explore the kinetics of PsS1_19B on κ -NC2, we used the same pH-based assay as for PsS1_19A. The K_M and apparent V_{max} values for PsS1_19B against κ -NC2 were $292.6 \pm 20.2 \mu\text{M}$ and $128.3 \pm 2.8 \mu\text{mol s}^{-1}$, respectively (Figure 25B). As we were unable to confirm the fraction of matured enzyme we cannot report k_{cat} . The K_M of PsS1_19B, being a concentration independent parameter, can be compared to the K_M of PsS1_19A ($670.0 \pm 55.8 \mu\text{M}$); these two sulfatases show the same sulfate group regioselectivity with the K_M PsS1_19B being roughly half that exhibited by PsS1_19A.

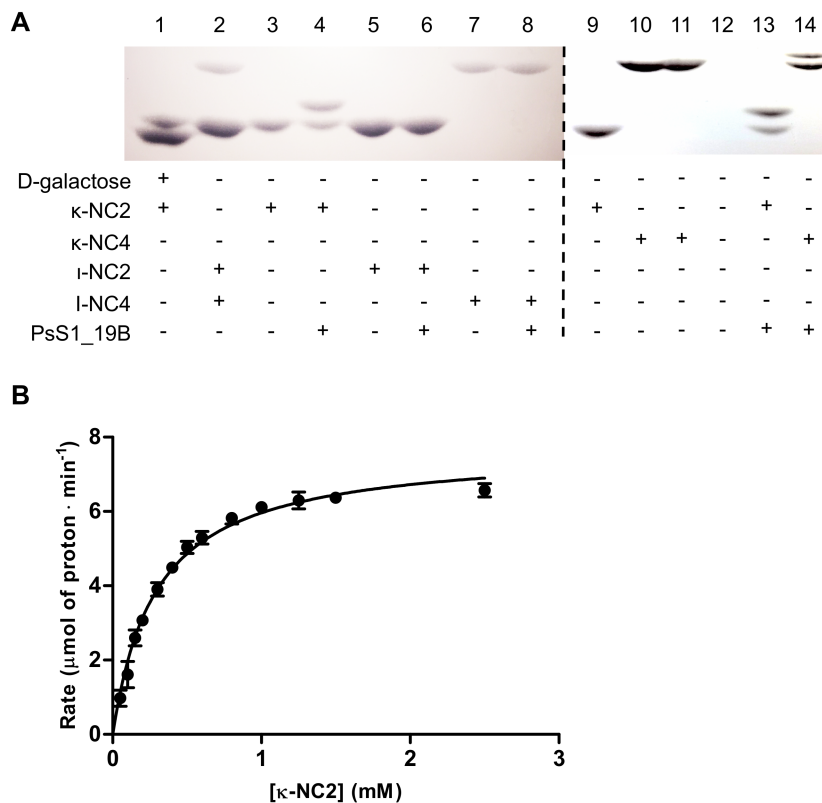


Figure 25. Activity of PsS1_19B from *Pseudoalteromonas* sp. PS47. (A) FACE analysis of PsS1_19B on κ -NC2, κ -NC4, I-NC2, and I-NC4. Presence or absence of oligosaccharides and enzyme are indicated below the gel. (B) Kinetic analysis of PsS1_19B activity on κ -NC2 at 25 °C using a pH-based colorimetric assay to detect proton release that accompanies sulfate ester hydrolysis.

4.2.2 The structure of PsS1_19B

To gain insight into the PsS1_19B mode of action, whether it be *endo*- or *exo*-acting, and to understand the molecular basis of its specificity, we determined the X-ray crystal structure of the wild type PsS1_19B enzyme at 2.50 Å resolution using PsS1_19A from PS47 (PDB ID code: 6BIA) (Hettle et al., 2018) as the search model. The final refined model contained four molecules in the asymmetric unit arranged as a tetramer (Figure 26A). PISA analysis of the PsS1_19B structure predicts the tetramer to be composed of two dimers, with buried surface areas of $\sim 1460 \text{ Å}^2$ between the chains participating in the dimer interface (chains A and B, or C and D), and only $\sim 362 \text{ Å}^2$ between chains from the different dimers (i.e. between chains A and C). The molecular interactions of the dimer interfaces shows an extensive and direct hydrogen bond network (Figure 26A inset). When the electrostatic potential is mapped onto the PsS1_19B homodimer interface there are two predominant areas with complimentary features that act as anchor points holding the dimer together (Figure 26B). R446 from chain A fits into a complementary pocket created in chain B formed by the residues N107, L451, R453, E87, and E97, of which the latter sits at the bottom of this pocket and coordinates R446 (Figure 26C). To assess the oligomeric state of PsS1_19B in solution, the protein was passed through a calibrated size exclusion chromatography column. The elution volume for PsS1_19B corresponds to an experimental molecular weight of 91.3 kDa. This is between the expected molecular weight of the PsS1_19B monomer (55.828 kDa) and dimer (112 kDa), but is much closer to that of a dimer (Figure 26D). The PISA analysis and the elution volume results support the dimer as being the biological unit for PsS1_19B and the tetramer is generated by non-crystallographic symmetry.

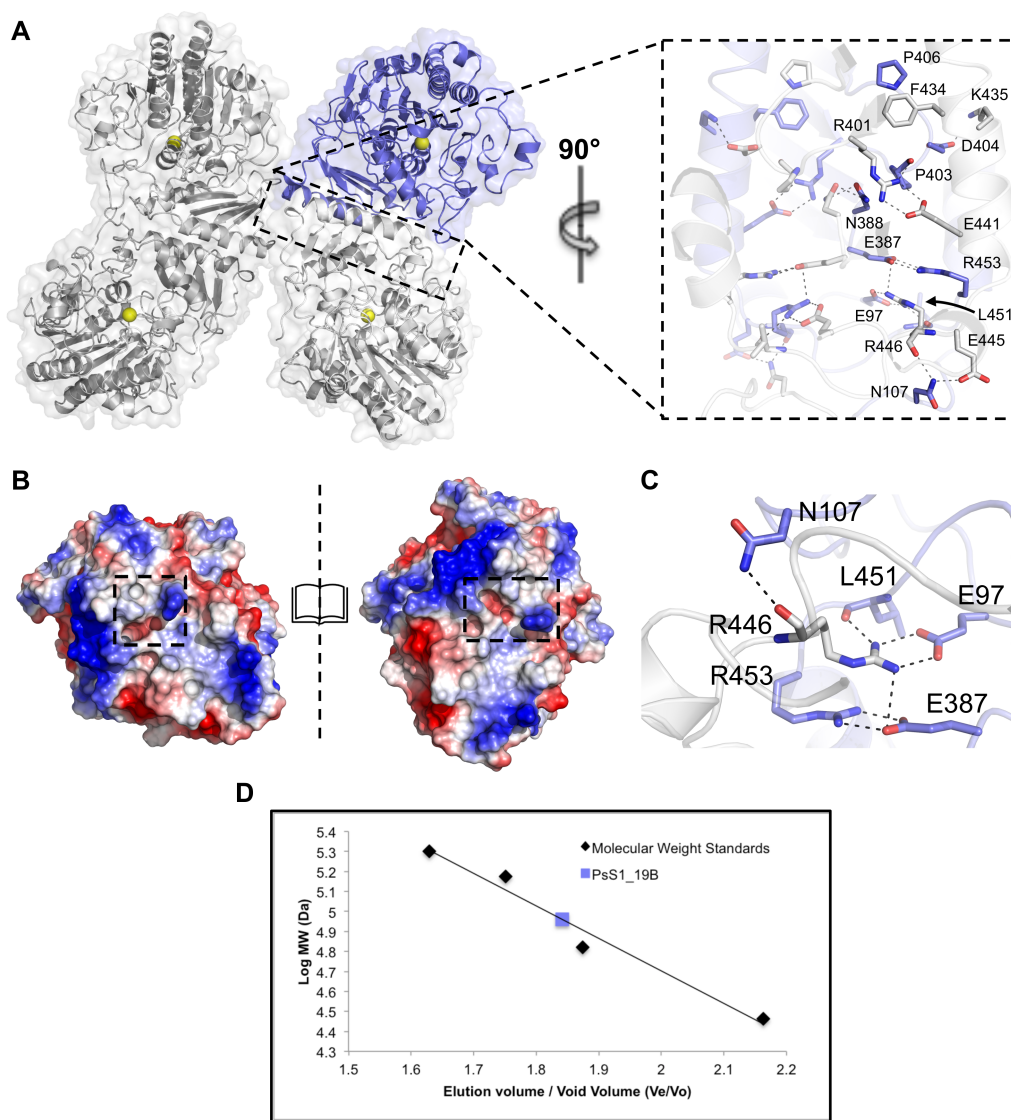


Figure 26. Molecular details driving dimerization of PsS1_19B.

(A) Cartoon representation of the PsS1_19B tetramer with each monomer composing the tetramer coloured slate, gray80, gray90, and white. Calcium ions are shown as yellow spheres. The inset shows the molecular interactions at the dimer interfaces. Amino acid residues are represented as sticks and coloured according to their respective chains; mirrored interactions are shown as transparent amino acid residues, and hydrogen bonds are represented as dashed lines.

(B) Electrostatic surface representation of the interface between chains opened along the vertical axis (shown by dashed line) to expose the interacting surfaces. Blue and red colouring represent basic and acid patches, respectively.

(C) Arginine anchor the dimer interface highlighted by dashed square in panel (B). Amino acid residues are coloured slate or white depending on contributing chain, chain A or chain B, respectively. Hydrogen bonds are represented as dashed lines.

(D) Oligomeric state determination of PsS1_19B based on elution volume when compared to S-300 calibration curve of protein markers ranging from 29 kDa to 200 kDa. PsS1_19B represented as a purple square.

PsS1_19B has the expected domain architecture typical of S1 sulfatases with each monomer containing a coordinated calcium ion in the active site pocket. Like PsS1_19A, the larger N-terminal sub-domain consists of an α/β topology containing a meandering 8-stranded β -sheet in which all strands are parallel except strand $\beta 7$, and all are surrounded by α -helices (Figure 27A). The smaller C-terminal sub-domain contains a four-stranded anti-parallel β -sheet tightly packed against the N-terminal domain by a single α -helix. As mentioned previously, FGly-dependent sulfatases possess ten highly conserved polar residues involved in sulfate recognition and sulfate ester hydrolysis. The PsS1_19A sulfatase structure was used to identify these residues in PsS1_19B. The PsS1_19B protein used for crystallization was produced in the absence of FGE co-expression resulting in similar observations regarding the proto-catalytic residue C77 to that of the analogous residue in PsS1_19A; i.e. no evidence was observed to suggest maturation into the nucleophilic FGly residue. However, C77 from PsS1_19B overlapped with the proto-catalytic residue C78 from PsS1_19A (Figure 27B), indicating that C77 is the likely proto-catalytic residue in PsS1_19B that undergoes post-translational modification to the catalytically active FGly. The overlay of PsS1_19B with PsS1_19A highlights the other core-conserved residues of PsS1_19B: D37, D38, D290 and N291, involved in metal coordination; R81 and H142, which help stabilize the FGly; and K140, H233 and K309, which are involved in sulfate recognition and activation of the FGly nucleophile (Figure 27B). Residue N99 from PsS1_19A, which is used in sulfate coordination and is a component of the S subsite, is not conserved in PsS1_19B. Residue H233 from PsS1_19A overlays with the proposed catalytic acid histidine from PsS1_19B, H233. This region, indicating the location of the

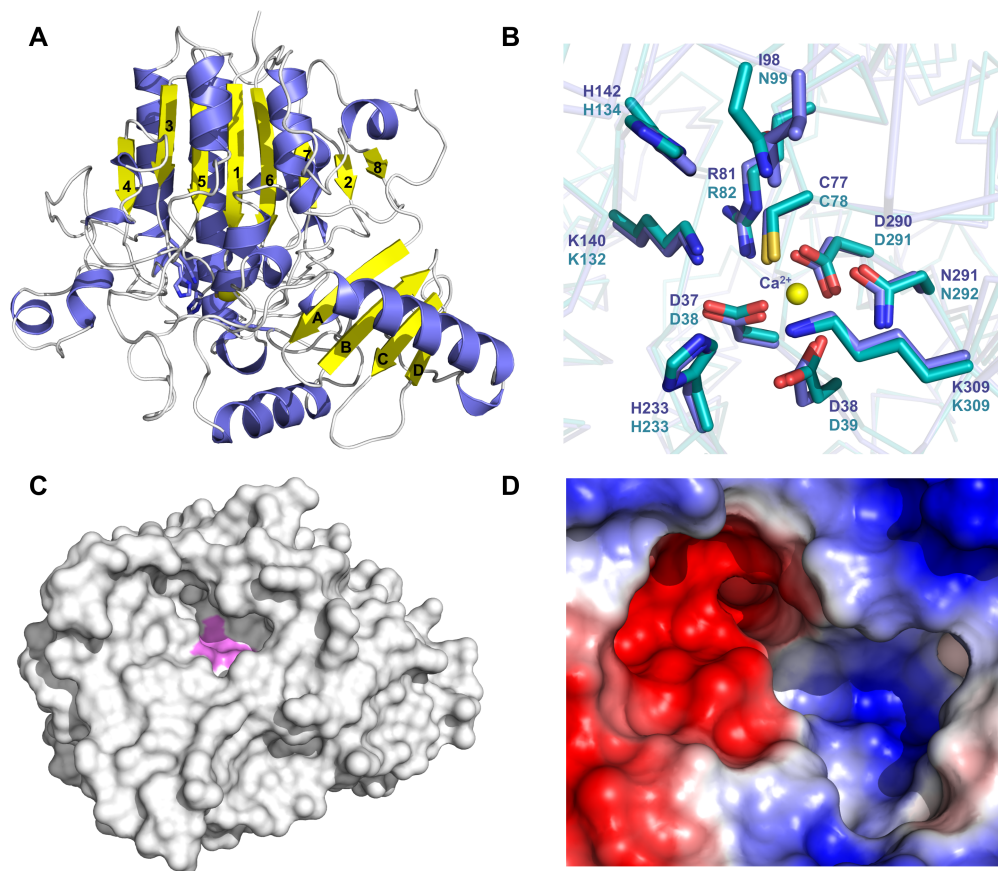


Figure 27. X-ray crystal structure of PsS1_19B.

(A) PsS1_19B is characterized by two subdomains of mixed α/β topology (α -helices coloured slate, β -sheets coloured yellow). N-terminal subdomain meandering β -sheets are numbered 1-8 and the C-terminal subdomain anti-parallel β -sheets are lettered A-D. Residues comprising the S-subsite are represented as slate sticks and the calcium ion as a yellow sphere.

(B) Structural overlay of PsS1_19B (slate) with PsS1_19A (teal) highlighting the conserved residues found within the S1_19 family. Only the calcium ion from the PsS1_19B structure is represented, shown as a yellow sphere.

(C) The solvent accessible surface of the PsS1_19B active site pocket. The violet coloured surface shows the location of the catalytic center of the enzyme.

(D) A close up of the solvent accessible surface of PsS1_19B active site pocket coloured as electrostatic potential, coloured blue and red representing basic and acidic patches, respectively.

active site, resides at the bottom of the pocket present on the surface of each monomer creating what is likely the S subsite (Figure 27C). The solvent accessible surface, with electrostatic potentials mapped onto the binding pocket of PsS1_19B, showed a highly charged pocket, which is divided between two distinct areas of basic and acidic patches (Figure 27D). This pocket topography, in conjunction with the FACE data, suggests that PsS1_19B likely utilizes an *exo*-mode of action.

4.2.3 The structure of PsS1_19B in complex with κ -NC2

In order to assess the mode of action of PsS1_19B and the molecular interactions it makes, we sought the structure of PsS1_19B in complex with intact substrate. As with PsS1_19A, we produced an inactive mutant of PsS1_19B by site-directed mutagenesis of C77 to a serine to ensure that it could not be post-translationally modified to the catalytic FGly residue. PsS1_19B C77S was incubated with excess κ -NC2 and re-screened for crystallization conditions to generate a complex. Co-crystallization was successful and the resulting 1.95 Å resolution structure revealed a hexamer, more specifically a trimer of non-crystallographically related dimers. The interface between the chains within a single dimer contains the same interactions seen in the PsS1_19B apo enzyme structure, further supporting this dimer being physiologically relevant. Each PsS1_19B C77S monomer contains a single intact κ -NC2 molecule with continuous density (Figure 28A) occupying the active site pocket and spanning the catalytic machinery (Figure 28B). The 4-sulfate, present on the G4S unit in the O subsite is coordinated in the S subsite by the amino acid sidechains conserved in S1 sulfatases. Again, the cysteine to serine mutation prevented maturation to the

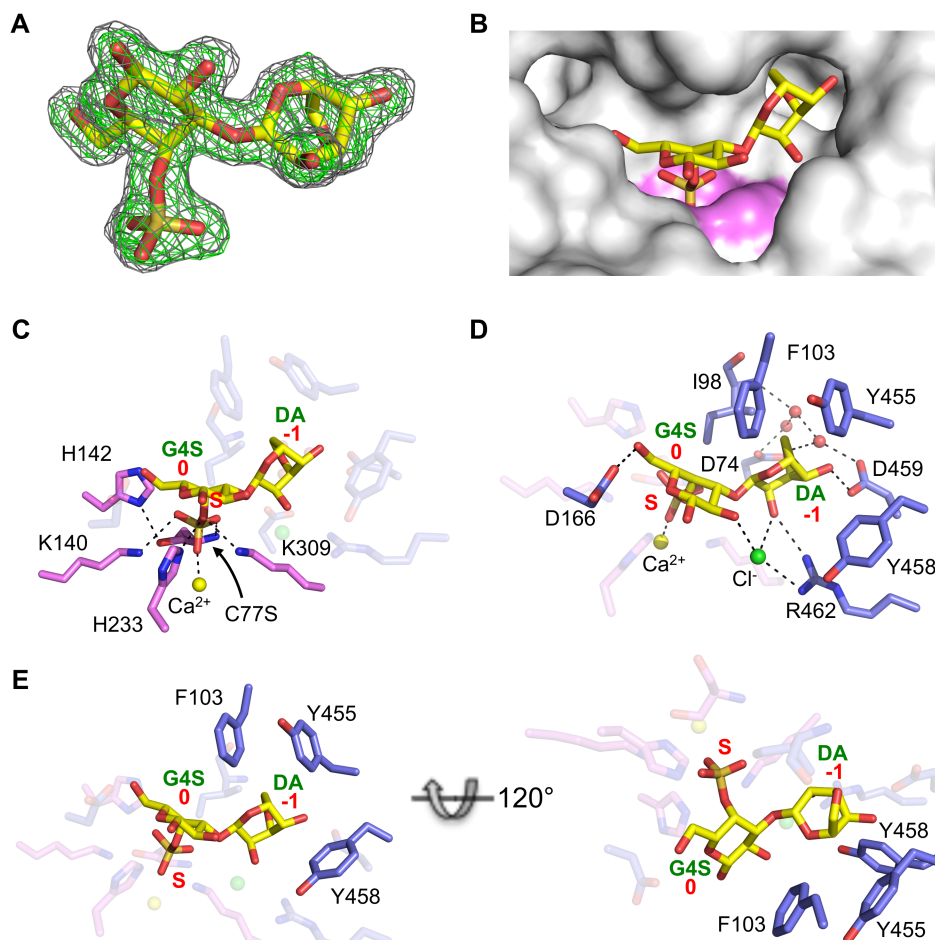


Figure 28. PsS1_19B C77S in complex with κ -NC2.

(A) Representative electron density for κ -NC2 modeled into the active site. The gray mesh shows the electron density map as a maximum likelihood/ σ_a -weighted $2F_o - F_c$ map contoured at 1.0σ . The green mesh shows the electron density map as a maximum likelihood/ σ_a -weighted $F_o - F_c$ map (contoured at 3.0σ) produced by refinement with the κ -NC2 atoms omitted.

(B) The conformational complementarity of κ -NC2 within the active site pocket of the PsS1_19B C77S mutant. The solvent accessible surface is coloured in gray with the area comprising the S-subsite coloured in violet.

(C) The interactions of the targeted sulfate ester of κ -NC2 with the sulfate binding S-subsite.

(D) The specific interactions of κ -NC2 with the active site pocket 0 and -1 subsites.

(E) Interactions with the DA unit in the -1 subsite highlighting the aromatic shelf. In panels (C) – (E), residues forming the S-subsite are represented as violet sticks and residues comprising the remaining subsites are represented as slate sticks. Calcium and chloride ions are shown as yellow and green spheres, respectively; hydrogen bonds are represented by dashed lines, and κ -NC2 is shown as yellow sticks. Sugar residues and subsites are labeled in green and red, respectively.

catalytic FGly residue, and C77S was positioned ~ 3.5 Å below the targeted sulfate ester mimicking the position of the FGly residue. The catalytic acid, H233, is positioned ~ 3.1 Å from the scissile bond oriented to protonate the ester oxygen (Figure 28C).

PsS1_19B interacted with the non-reducing end DA in the -1 subsite through a direct and water mediated hydrogen bond network interacting with the hydroxyl groups of carbons C2 and C4 and with the oxygen of the 3,6-anhydrous bridge, excluding solvent from the coordination of the non-reducing end indicating *exo*-recognition. F103, Y455 and Y458 create an “aromatic shelf” aiding in the positioning of the DA unit and allowing for the endocyclic oxygen to remain solvent exposed (Figure 28D). The respective pyranose rings of the G4S and DA units were oriented $\sim 70^\circ$ with respect to one another, allowing for the DA unit C2 hydroxyl group in the -1 subsite to interact with R462 directly and through a bridged chlorine ion (Figure 28D).

There are no common CH- π interactions between aromatic amino acid side chains and carbohydrate ring structures observed within the PsS1_19B C77S κ -NC2 complex as were observed in the PsS1_19A C78S complexes. The molecular interactions involved in coordination of the DA unit occupying the -1 subsite of PsS1_19B C77S involves aromatic amino acid side chains and utilize a mode of binding not commonly seen, different still from the cradle observed in the -1 subsite of PsS1_19A C78S in complex with κ -C5 or κ -I- κ -NC6. Of the three aromatic residues, only Y455 is oriented with the face of the phenol group towards the DA unit, whereas the side chain carbons CD2 and CE2 of F103 and Y458 are oriented towards the endocyclic oxygen of the pyranose ring and towards the underside of the DA unit, respectively (Figure 28E).

4.2.4 The recognition of ι-NC4 by PsS1_19B forms an inactive complex

PsS1_19B did not show any activity against ι-carrageenan moieties when tested by FACE, yet the incorporation of a chloride ion in the κ-NC2 complex suggests that PsS1_19B may be able to accommodate other negatively charged species at this location. Furthermore, with the reducing end being solvent exposed in the PsS1_19B C77S κ-NC2 complex, PsS1_19B may possess (+) subsites extending out from the active site pocket. To probe these hypotheses, the structure of PsS1_19B C77S in complex with ι-NC4 was solved at 2.00 Å resolution using the same co-crystallization approach as for the κ-NC2 complex; however, here PsS1_19B C77S crystallized as a dimer, again with the same dimer interface interactions as observed in both the apo enzyme and κ-NC2 complex structures. As with the κ-NC2 complex, the analogous DA2S-G4S at the non-reducing end of ι-NC4 compliments the active site pocket spanning the catalytic machinery (Figure 29A). The longer degree of polymerization of ι-NC4 highlights the extra subsites of the PsS1_19B catalytic pocket: the DA2S unit in the +1 subsite is coordinated via D166 and N181, with N181 coordinating the sulfate group of the DA2S unit, and the G4S occupying the +2 subsite is coordinated by K187 and E176 through a water bridge (Figure 29B). PsS1_19B is not only able to accommodate longer carrageenan oligosaccharides, but it is also able to accommodate a sulfate group in the -1 subsite. Interestingly, the recognition of this sulfate causes a shift in the ι-NC4 allowing for a counter ion to sit in between the coordinated calcium ion and the targeted sulfate in the S subsite leading to a catalytically inactive complex (Figure 29C). This movement positions the scissile bond ~4.38 Å away from H233, as opposed to ~3.02 Å as seen in the κ-NC2 complex, thus preventing its action as the catalytic acid (Figure 29D). Not only does this counter ion cause a proximity change between H233 and the

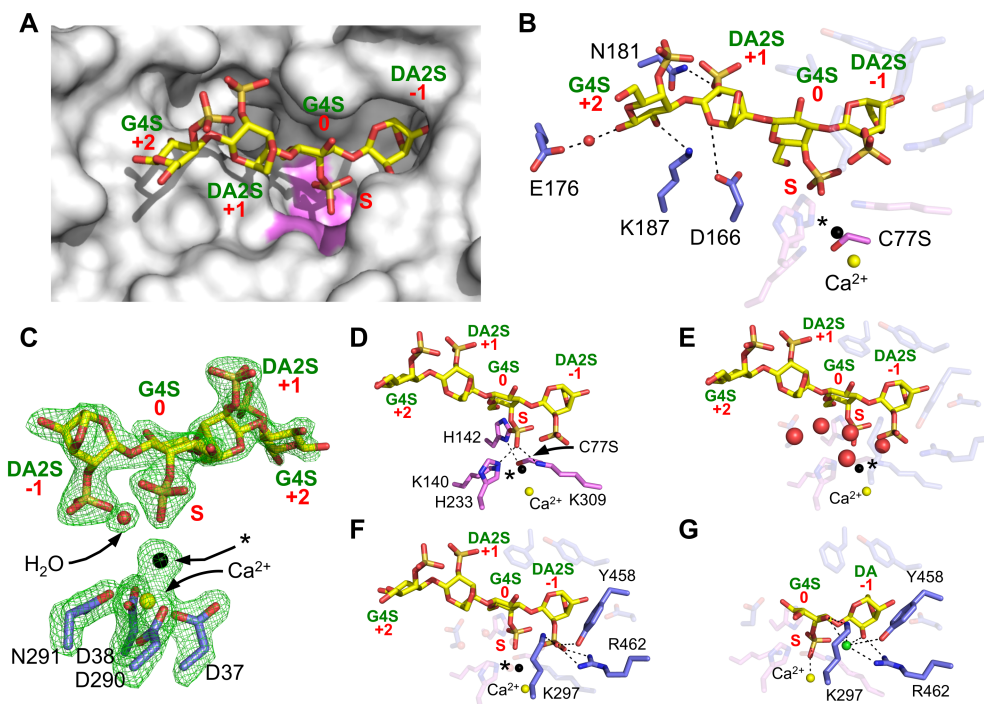


Figure 29. Catalytically inactive complex of PsS1_19B C77S with ι -NC4. (A) Close up of PsS1_19B C77S active site pocket with bound ι -NC4. The solvent accessible surface is coloured in gray with the area comprising the S-subsite coloured in violet. (B) The ι -NC4 bound in the active site with a focus on the DA2S and G4S residues in the +1 and +2 subsites, respectively. (C) Representative electron densities for ι -NC4, calcium ion, calcium binding residues, and the counter ion modeled into the active site. The green mesh shows the electron density map as a maximum likelihood/ σ_a -weighted $F_o - F_c$ map (contoured at 3.0 σ) produced by refinement with the corresponding residues omitted. Four views of the ι -NC4 bound in the active site with a focus on (D) interactions in the S-subsite, the coordinated waters in the active site (E), and a comparison of the coordination of the DA2S unit in the -1 subsite (F) with the coordination of the DA unit in the -1 subsite from the PsS1_19B C77S κ -NC2 complex. In all panels, ι -NC4 and κ -NC2 are shown as yellow sticks. In panels (B) – (G), the calcium ion is represented as a yellow sphere, water molecules as red spheres, chlorine molecules as green spheres, the counter ion as black spheres (designated as asterisk), and hydrogen bonds are represented as dashed lines. Subsites are labeled in red, and sugar residues labeled in green. Relevant interacting side chains on the protein are shown as violet (S-subsite) or slate (integer subsites and calcium binding residues).

scissile bond, it also allows for water molecules to be incorporated between the glycan and the residues comprising the binding pocket (Figure 29E). The major driving force behind this shift is the recognition of the DA2S unit in the -1 subsite where K297, R462, and Y458 form hydrogen bonds with the 2-sulfate group occupying the space that the chlorine ion did in the κ -NC2 complex. (Figures 29F and 29G). The recognition of the ι -moiety creating an inactive complex is supported by the lack of activity exhibited by PsS1_19B against both ι -NC2 and ι -NC4.

4.2.5 Identification of a third PS47 sulfatase

The PsS1_19B inactive complex and lack of activity on ι -carrageenan oligosaccharides suggest that the non-reducing end 2-sulfate group of ι -carrageenan would need to be removed in order for PsS1_19B to perform its catalytic function, and ultimately for the complete saccharification of ι -carrageenan. For this either the third PS47 sulfatase, PsS1_NC, participates in ι -carrageenan degradation or PS47 is unable to utilize ι -carrageenan as a carbon source; however, sharing the same carrageenan PUL as PSU2A and PsS1_19A's preference for ι -moieties, it is unlikely that the latter holds true. As PsS1_NC did not show any activity on any sulfated substrates, either synthetic or naturally occurring, and without biochemical data to suggest what the biological substrate for PsS1_NC is, structural analysis was performed in an attempt to provide insight into the physiological role of PsS1_NC. We determined the X-ray crystal structure of the wild type PsS1_NC enzyme at 1.45 Å resolution phased by molecular replacement using the PsS1_19A wild type structure as the search model. The final refined model contains two molecules in the asymmetric unit arranged as a homodimer

with C2 symmetry (Figure 30A). PISA analysis of this structure predicts the dimer to be a stable oligomeric state with ~1500 Å² of buried surface area at the dimer interface. This interface is composed of an extensive and symmetrical hydrogen bond network (Figure 30A inset). When looking at the electrostatic potential of the interface there are areas of complementarity between acidic and basic patches (Figure 30B). The residue R134 likely plays a key role in this dimer interface by acting as an anchor between the two chains when docked in the respective niche of the complementary strand (Figure 30C). The large buried surface area lends strength to the conclusion that this dimerization is physiological and is further supported by size exclusion chromatography analysis of PsS1_NC, whose elution volume corresponded to an experimental molecular weight of 102.3 kDa. With an expected molecular weight of 56.613 kDa for a monomer and 113.2 kDa for a dimer, this experimental value is more consistent with that of a dimer rather than a monomer (Figure 30D).

Like both PsS1_19A and PsS1_19B, each monomer of PsS1_NC has the domain architecture typical of FGly-dependent S1 sulfatases along with the coordinated calcium ion in the active site pocket. The PsS1_NC larger N-terminal sub-domain consists of α/β topology with a 7-stranded β -sheet, of which all strands are parallel except strand β_6 , and are surrounded by α -helices (Figure 31A). The smaller C-terminal sub-domain contains a four-stranded anti-parallel β -sheet packed against the N-terminal sub-domain by two α -helices.

Aligning PsS1_NC with a known carrageenan sulfatase (PsS1_19A), the ten conserved polar residues involved in sulfate recognition and sulfate ester hydrolysis are readily identified (Figure 31B). The residue C84 overlaps with the proto-catalytic residue

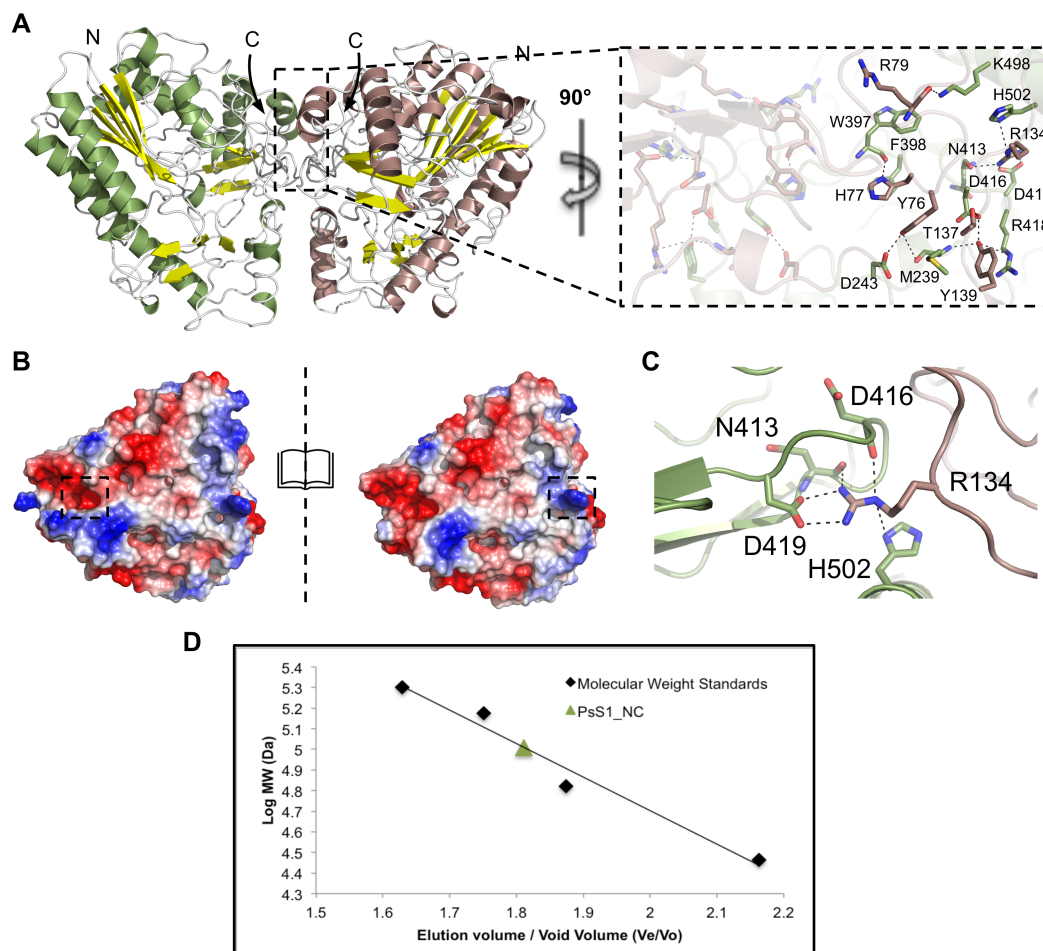


Figure 30. Molecular details driving dimerization of PsS1_NC

(A) Cartoon representation of the PsS1_NC dimer. The α -helices are coloured according to either chain A or chain B as green or brown, respectively. β -sheets are coloured yellow and polypeptide termini are labeled (N) or (C). The inset shows the molecular interactions at the dimer interface. Amino acid residues are represented as sticks and coloured according to their respective chains; mirrored interactions are shown as transparent amino acid residues and hydrogen bonds are represented as dashed lines.

(B) Electrostatic surface representation of the interface between chains opened along the vertical axis (shown by dashed line) to expose the interacting surfaces. Surface coloured blue and red representing basic and acidic patches, respectively.

(C) Specific interactions at the arginine anchoring location of the dimer interface as highlighted by dashed square in panel (B). Amino acid residues are coloured slate or brown depending on contributing chain of interface, chain A or chain B, respectively. Hydrogen bonds are represented as dashed lines.

(D) Oligomeric state determination of PsS1_19B based on elution volume when compared to S-300 calibration curve of protein markers ranging from 29 kDa to 200 kDa. PsS1_NC represented as a green triangle.

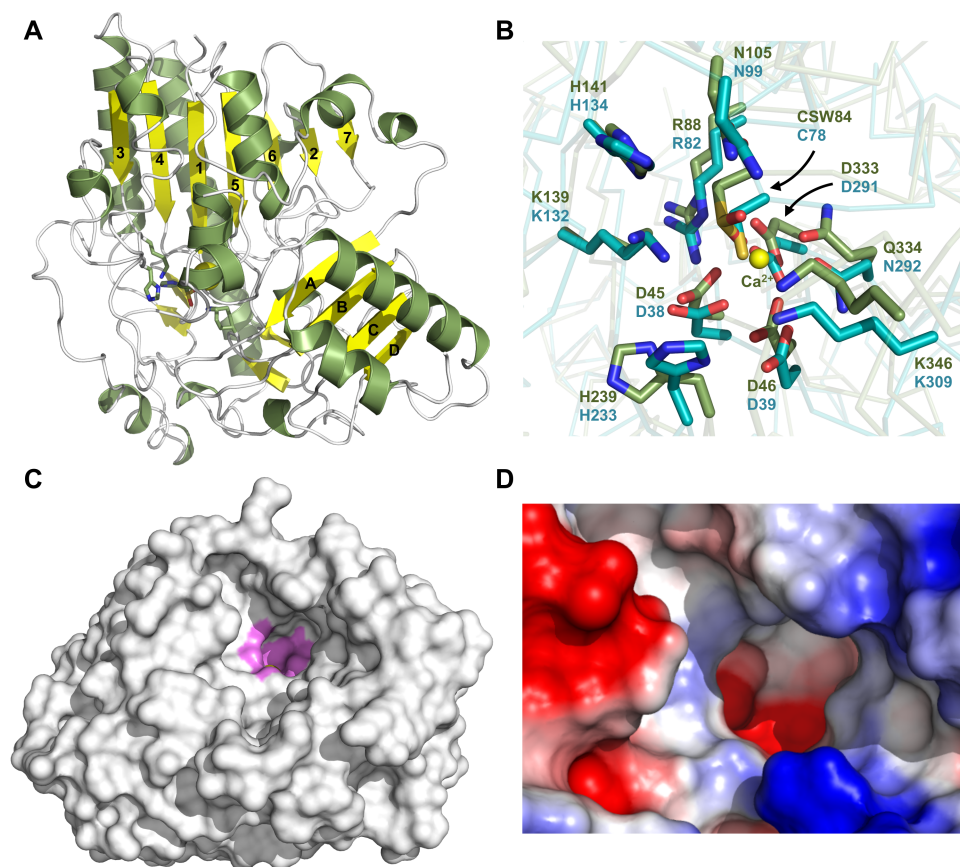


Figure 31. X-ray crystal structure of PsS1_NC.

(A) PsS1_NC is characterized by two subdomains of mixed α/β topology (α -helices coloured green, β -sheets coloured yellow). N-terminal subdomain meandering β -sheets are numbered 1-7 and the C-terminal subdomain anti-parallel β -sheets are lettered A-D. Residues comprising the S-subsite are represented as green sticks and the calcium ion as a yellow sphere.

(B) Structural overlay of PsS1_NC (green) with PsS1_19A (teal) highlighting the 10 conserved residues found within the S1_19 family. Only the calcium ion from the PsS1_NC structure is represented, shown as a yellow sphere.

(C) The solvent accessible surface of PsS1_NC active site pocket. The violet coloured surface shows the location of the catalytic center of the enzyme.

(D) A close up of the solvent accessible surface of PsS1_NC active site pocket coloured as electrostatic potential, blue and red represent basic and acidic patches, respectively.

of PsS1_19A, indicating that C84 is the proto-catalytic residue in PsS1_NC that undergoes post-translational modification to the catalytically active FGly. The other residues show almost complete conservation with those from PsS1_19A apart from one of the metal coordinating residues, Q334 (Figure 31B). H239, like H233 from PsS1_19B, overlays with the proposed catalytic acid histidine. This region, indicating the location of the active site, resides at the bottom of the pocket present on the surface of each monomer creating what is likely the S subsite (Figure 31C). The solvent accessible surface with electrostatic potentials mapped onto the binding pocket of the PsS1_NC show an acidic bottom, while basic residues line the entrance to this pocket (Figure 31D).

4.2.6 Presence of a dead-end intermediate in PsS1_NC maturation

When the interconnected polar residue composition is compared between PsS1_NC and the two PsS1_19 sulfatases, there are only two major differences. First, PsS1_NC utilizes three aspartate residues and one glutamate residue, as opposed to three aspartate residues and one asparagine residue, to coordinate the calcium ion. The residues involved in calcium ion coordination may be sulfatase family specific and therefore the use of asparagine may be exclusive to S1_19 sulfatases. Second, the PsS1_NC wild type protein used for crystallization was produced without FGE co-expression and, therefore, the predicted proto-catalytic residue C84 should not show any evidence of maturation to the FGly residue. No structural evidence of maturation was seen with the other PS47 sulfatases expressed in *E. coli* in the absence of FGE co-expression, however, in the PsS1_NC wild type X-ray crystal structure the electron

density around residue 84 was quite unique (Figure 32A). In the case of PsS1_NC, the proto-catalytic residue was neither the matured FGly residue or the unmatured cysteine, as the F_o-F_c electron density maps when modeled as either side chain showed an undesirable hydroxyl group or unsatisfied density neighbouring the thiol group, respectively (Figure 32B and 32C). This leads to the hypothesis that the residue here may be an intermediate of the maturation process. The general consensus model for sulfatase maturation, as introduced earlier, is a monooxygenase mechanism resulting in a cysteinesulfenic acid intermediate (Dierks et al., 2005); however, the density where the proto-catalytic cysteine would reside in the PsS1_NC wild type structure is not satisfied by cysteinesulfenic acid in conformation-1 (directed towards the calcium ion), conformation-2 (directed away from the calcium ion) or in both conformations each having 50 % occupancy (Figure 32D – 32F). Instead, the density at this residue location is more accurately described by a cysteinesulfinic acid residue (Figure 32G). The B-factors associated with cysteinesulfinic acid do not deviate largely from neighbouring atoms or other residues when modeled into this site. With this observation not being seen in either PsS1_19A or PsS1_19B, it is possible that the FGE recognition motif in PsS1_NC differs just enough from that found in PsS1_19A and PsS1_19B that it causes weak interactions between PsS1_NC and the *M. tuberculosis* FGE or the *E. coli* maturation machinery, leading to oxidation of the sulfhydryl group of the cysteinesulfenic acid intermediate into cysteinesulfinic acid and the creation of a dead-end intermediate resulting in a sulfatase that can no longer be matured (Figure 33).

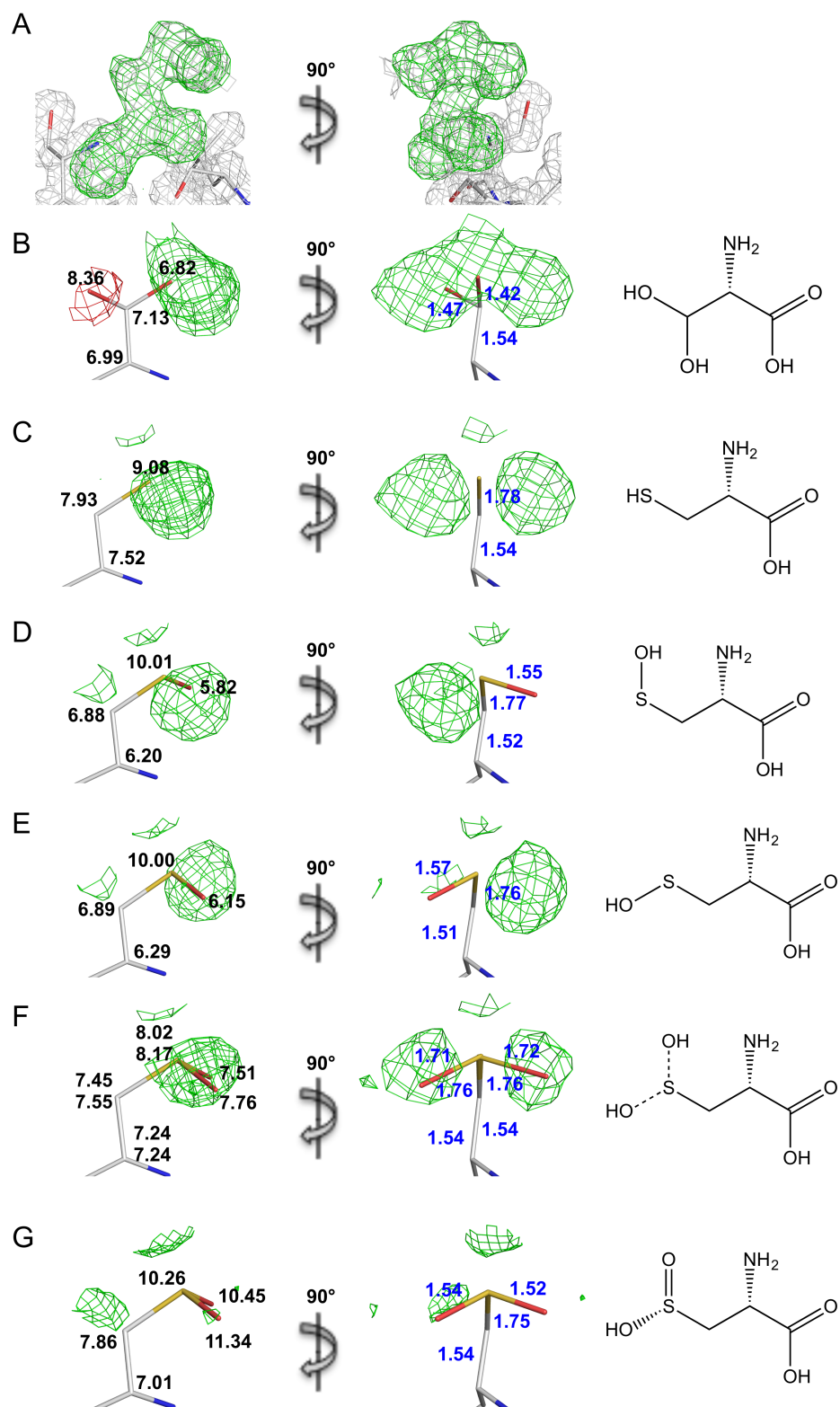


Figure 32. X-ray crystallographic evidence for the identification of amino acid residue number 84 in PsS1_NC. Figure legend on following page.

Figure 32. X-ray crystallographic evidence for the identification of amino acid residue number 84 in PsS1_NC.

(A) Representative electron density for PsS1_NC active site residues. The gray mesh shows the electron density map as a maximum likelihood/ σ_a -weighted $2F_o - F_c$ map contoured at 1.0σ . The green mesh shows the electron density map as a maximum likelihood/ σ_a -weighted $F_o - F_c$ map (contoured at 1.0σ) produced by refinement with residue 84 atoms omitted. Residues are coloured as gray lines.

In panels (B) – (G), representative electron density for PsS1_NC residue 84 is shown when modeled as FGly (B), cysteine (C), cysteinesulfenic acid conformation-1 (D), cysteinesulfenic acid conformation-2 (E), cysteinesulfenic acid in both conformations each having 50 % occupancy (F), and cysteinesulfenic acid (G). The green mesh is the same as in (A) and produced by refinement with atoms present. All colouring for panels (B) - (G) are the same as in (A). Respective B-factors and bond lengths are labeled in black and blue, respectively.

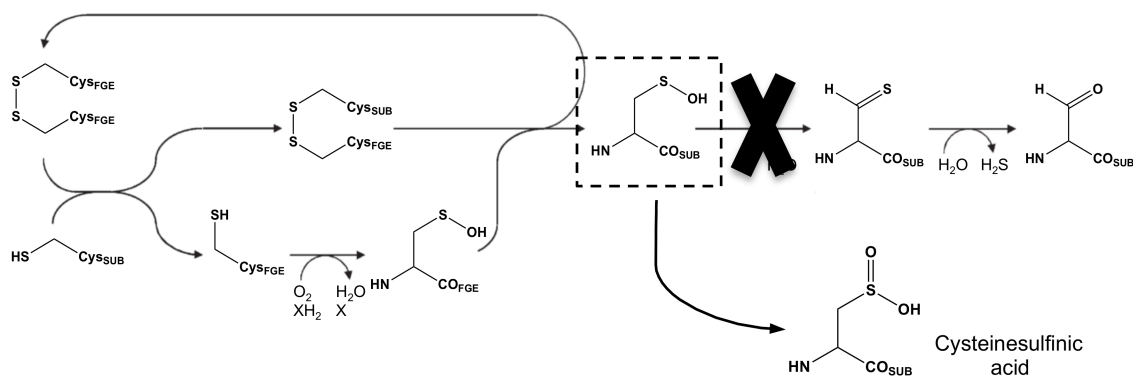


Figure 33. Formylglycine generation of a dead-end intermediate.

Formylglycine generation mechanism with cysteinesulfenic acid substituted in for cysteinesulfenic acid (adapted from Dierks et al. (2005)). X designates the end of maturation in the case of PsS1_NC within *E. coli* resulting in a dead-end intermediate.

4.2.7 PsS1_NC: an exo-2S-I-carrageenan sulfatase

When co-expressed in *E. coli* with the *M. tuberculosis* FGE, PsS1_NC showed no activity against any carrageenan oligosaccharides that were available, purchased or produced in-house. With the apparent presence of a cysteinesulfinic acid residue in the active site of PsS1_NC wild type when expressed without FGE, this lack of activity may be due to the partial maturation of this enzyme by the *E. coli* maturation system, thus leaving it unrecognizable by the *M. tuberculosis* FGE and functionally inactive. With the lack of any activity data to inform our structural studies, we pursued substrate complexes of PsS1_NC by soaking small carrageenan oligosaccharides into catalytically inactive mutants of PsS1_NC in the hopes of structurally identifying a candidate physiological substrate. As the cysteinesulfinic acid residue is larger than a FGly, we considered that this may pose a steric problem and potentially interfere with ligand binding; therefore, we mutated C84 to serine in order ensure that it could not be post-translationally modified. Crystals of PsS1_NC C84S were obtained in the same crystal form as the apo-enzyme and I-NC4 was subsequently soaked into these crystals. The X-ray crystal structure of PsS1_NC C84 in complex with I-NC4 was solved at 2.25 Å resolution revealing a dimer with only representative density for the non-reducing end disaccharide of I-NC4 occupying the active site pocket of PsS1_NC C84S, this I-NC2 represented in the crystal structure spans the catalytic machinery (Figure 34A and 34B). The 2-sulfate of the non-reducing end DA2S unit in the 0 subsite was coordinated by the amino acid sidechains that are conserved amongst the S subsite of S1 sulfatases (Figure 34C). The C84S mutation prevented the post-translational modification to the cysteinesulfinic acid residue, and was positioned ~3.2 Å directly

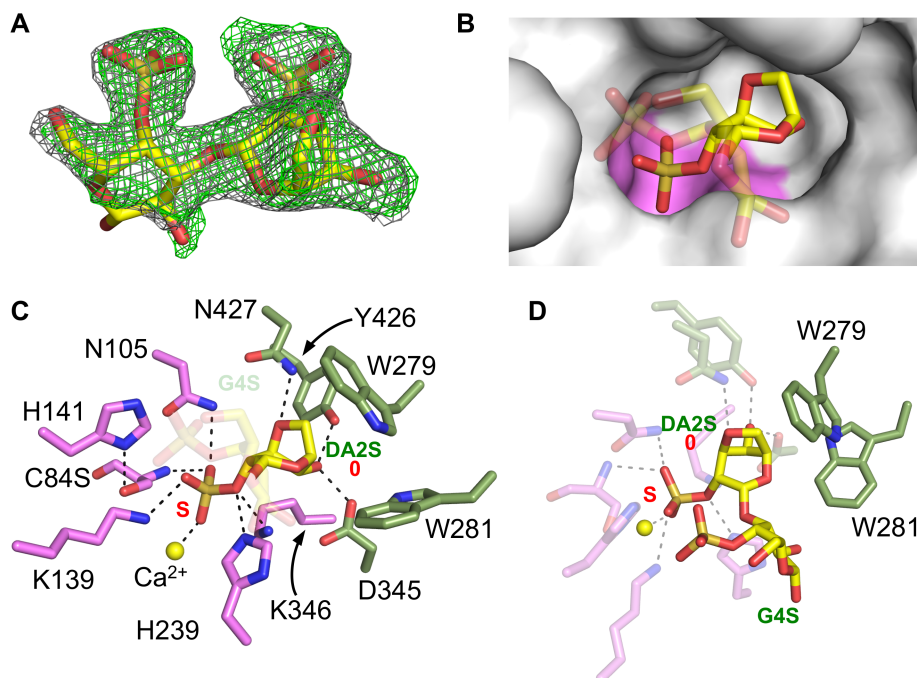


Figure 34. PsS1_NC C84S in complex with l-NC4.

(A) Representative electron density for l-NC4 modeled into the active site. The gray mesh shows the electron density map as a maximum likelihood/ σ_a -weighted $2F_o - F_c$ map contoured at 1.0σ . The green mesh shows the electron density map as a maximum likelihood/ σ_a -weighted $F_o - F_c$ map (contoured at 3.0σ) produced by refinement with the l-NC4 atoms omitted.

(B) The conformational complementarity of l-NC4 within the active site pocket of the PsS1_NC C84S mutant. The solvent accessible surface is coloured in gray with the area comprising the S-subsite coloured in violet.

(C) The interactions of the targeted sulfate ester of l-NC4 with the sulfate binding S-subsite and the DA2S unit with the residues of the 0 subsite.

(D) The specific interactions of DA2S in the 0 subsite highlighting the tryptophan cradle.

In panels (C) and (D) residues comprising the S-subsite are shown as violet sticks and those comprising the 0 subsite are shown as green sticks. Calcium ions are shown as a yellow sphere, hydrogen bonds as dashed lines, and l-NC4 as yellow sticks. Sugar residues and subsites are labeled in green and red, respectively.

beneath the sulfate atom mimicking where the FGly would reside, while the catalytic acid H239 was positioned ~ 2.7 Å from the scissile bond oxygen of the sulfate ester.

The bound α -NC2 substrate was coordinated with the non-reducing end fitting into the active site pocket, with the planes of each pyranose sugar rotated $\sim 66^\circ$ relative to each other. The DA2S unit in the 0 subsite is complemented by the shape of the active site pocket while excluding solvent. The DA2S unit in the 0 subsite is the only sugar component of the α -NC2 substrate making interactions with residues in the active site pocket of PsS1_NC. The molecular requirements for recognition of the DA2S unit were mainly focused on the non-reducing end hydroxyl group and the 3,6-anhydrous bridge of this residue. Though the G4S residue is represented in the crystal structure, PsS1_NC makes no direct or water-mediated interactions with this sugar residue, or with the G4S sulfate ester substituent, suggesting that the minimum recognition requirement for Ps_S1_NC is a single non-reducing end DA2S unit.

Again, as was previously seen in the other PS47 sulfatase complexes, the protein-carbohydrate interactions between PsS1_NC and the DA2S unit are not those commonly seen in carbohydrate specific enzymes. PsS1_NC utilizes a tryptophan cradle composed of W479 and W481 that interacts with the face of the 3,6-anhydro-D-galactose unit created by the endocyclic oxygen, C5, and C6 (Figure 34D). These interactions in the 0 subsite are crucial in positioning the sulfate ester near the proto-catalytic residue, thereby suggesting that PsS1_NC functions as an *exo*-2S- α -carrageenan sulfatase.

4.2.8 PsS1_NC in complex with κ -NC2 and HEPES

While pursuing PsS1_NC substrate complexes, we were also successful in achieving the structure of a pseudo-product complex by soaking κ -NC2 into PsS1_NC C84S crystals. The PsS1_NC C84S κ -NC2 complex was solved at 1.66 Å resolution and revealed a dimer with a single intact κ -NC2 molecule in the active site pocket that was coordinated in an almost identical manner to that of ι -NC2. The tryptophan cradle performed the same function, coordinating the DA unit in the 0 subsite; however, without the sulfate substituent present on the DA unit occupying the 0 subsite, the residues normally involved in both sulfate recognition and sulfate activation, K139 and H239, are now coordinating the G4S sulfate ester, thus creating a +1 subsite that was not seen in the ι -NC2 complex (Figure 35A).

Along with the coordination of κ -NC2, there is a single molecule of HEPES from the crystallization condition, with continuous electron density across the entirety of the molecule (Figure 35B), coordinated in a groove found on a face of PsS1_NC opposite to the active site pocket (Figure 35C). This highly charged molecule had quite an extensive direct electrostatic and water-mediated hydrogen bond network in comparison to its size, with the majority of this network contributing to the recognition of the HEPES ethanesulfonic acid group (Figure 35D). This HEPES molecule may imitate the binding of an inorganic sulfate ion resulting from sulfate ester hydrolysis, and thus indirectly identifying a putative regulatory allosteric site on PsS1_NC.

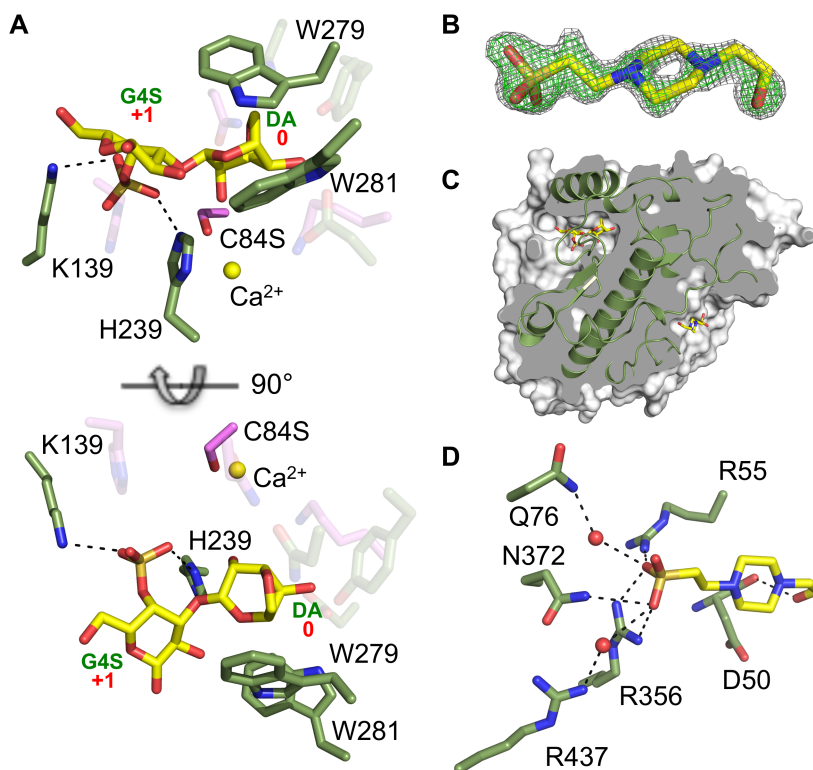


Figure 35. PsS1_NC C84S in complex with κ -NC2 and HEPES.

(A) Specific interactions highlighting the tryptophan cradle and the +1 subsite. Residues comprising the S-subsite and the integer subsites are coloured in violet and green, respectively, with the calcium ion represented as a yellow sphere, κ -NC2 as yellow sticks and hydrogen bonds as dashed lines. Sugar residues and subsites are labeled in green and red, respectively.

(B) Representative electron density for HEPES modeled into the backside groove. The gray mesh shows the electron density map as a maximum likelihood/ σ_a -weighted $2F_o - F_c$ map contoured at 1.0 σ . The green mesh shows the electron density map as a maximum likelihood/ σ_a -weighted $F_o - F_c$ map (contoured at 3.0 σ) produced by refinement with the HEPES atoms omitted.

(C) Cutaway of PsS1_NC C84S showing the spatial relationship between the active site and HEPES binding site. PsS1_NC is shown as a green cartoon and the solvent accessible surface is shown in white. Gray represents the interior of the enzyme and κ -NC2 is shown as yellow sticks.

(D) Molecular interactions involved in HEPES coordination. Amino acid side chains are shown as green sticks, water molecules as red spheres, and hydrogen bonds are represented as dashed lines.

In panels (B) – (D) HEPES is shown as yellow sticks.

4.3 Discussion

The diversity amongst sulfated biomolecules is vast, even within the context of carrageenan. With the large number of carrageenan classes observed in nature, the number of different possible sulfation patterns as sulfate esters are removed resulting in class remodeling, and with the redundant systems employed by organisms to deal with sulfate groups in *endo*- and *exo*-acting manners, the possible number of sulfatase-carrageenan pairs needed becomes enormous. Furthermore, the molecular requirements to accommodate the sulfate groups can become quite diverse even when the regioselectivity becomes redundant. PsS1_19B functions to remove the same 4-sulfate group as PsS1_19A; however, PsS1_19B acts in an *exo*-acting manner as evidenced by the biochemical and structural data presented here, and again it is the recognition of the leaving group (β -NC2) that provides this specificity.

Like PsS1_19A, PsS1_19B did not show any activity on pNP-sulfate but did display activity on oligosaccharides with a non-reducing end κ -NC2 building block. The X-ray crystal structure of the PsS1_19B C77S in complex with κ -NC2 shows, in each monomer, κ -NC2 occupies the active site with the G4S residue engaging the catalytic machinery. Yet when ι -NC2 occupies the analogous space in the PsS1_19B C77S ι -NC2 complex, the G4S residue is situated at a distance from the catalytic machinery that would not allow engagement by the necessary components. This interaction provides an explanation of the molecular details behind the lack of activity on polygeenan where the non-reducing end disaccharide component is ι -NC2, and by extension, other highly sulfated polygeenan classes. Together, the biochemical and

structural data supports the assignment of PsS1_19B as an *exo*-4S- κ -carrageenan sulfatase.

The growth of PSU2A on ι -NC4 (Figure 12; Section 1.7) indicates that there must be a gene product from the carrageenan PUL shared between PSU2A and PS47 that is able to cleave the 2-sulfate group ester from ι -NC4 DA2S units, and through extension other carrageenans containing DA2S units. The lack of PsS1_NC activity observed on any carrageenan class can be attributed to the dead-end cysteinesulfinic acid intermediate resulting from the production of unmaturred PsS1_NC. Without any biochemical activity data it cannot be stated unequivocally that PsS1_NC is an *exo*-2S- ι -carrageenan sulfatase. However, the X-ray crystal structure of PsS1_NC C84S in complex with ι -NC2 clearly shows the 2-sulfate group from the DA2S unit engaging the catalytic machinery. This alone is not enough to conclude the specificity of PsS1_NC; however, in conjunction with the PsS1_NC C84S – κ -NC2 pseudo-product complex, which shows the exact same recognition of both the non-reducing end DA unit in κ -NC2 and the analogous DA2S unit from ι -NC2, strongly suggests that the role of PsS1_NC is an *exo*-2S- ι -carrageenan sulfatase.

It has been reported that high sulfate concentrations will inhibit hydrolytic sulfatase activity (Dodgson et al., 1982; Nicholls and Roy, 1971; Recksiek et al., 1998) as well as inhibition by inorganic phosphate, which shares a tetrahedral molecular geometry with inorganic sulfate (Genicot et al., 2014; Recksiek et al., 1998). The inhibition of the enzymes reported in the literature by inorganic sulfate may be due to competitive inhibition or by an allosteric regulatory site. The PsS1_NC C84S in complex with HEPES suggests a possible allosteric regulatory site. The coordination of the

sulfonic acid functional group of HEPES may mimic the tetrahedral geometry of inorganic sulfate and inorganic phosphate coordination of the S=O or S-OH bonds by charged residues, possibly resulting in feedback inhibition. Unfortunately with the inability to produce matured PsS1_NC confirming this hypothesis with biochemical support was not possible.

The four PS47 sulfatase complexes reported here, especially the catalytically inactive PsS1_19B C77S – i-NC4 complex, further supports the hypothesis that it is recognition of the leaving group that drives specificity and substrate turnover for S1 sulfatases, especially in the case of carbohydrate specific sulfatases. It was argued by Barbeyron et al. (2016) that there are polyspecificities within a single sulfatase subfamily; however, it is also possible that different subfamilies may possess similar specificities, as evidenced by the characterized ZgS1_19 and ZgS1_7 sulfatases which are both specific for the 4-sulfate group of G4S units. *Z. galactanivorans* uses two sulfatases from differing subfamilies with the same regioselectivity, whereas PS47 utilizes two S1_19 sulfatases to release the 4-sulfate group. Homology models of ZgS1_7 and ZgS1_19 were generated using PsS1_19B as a template with the Web server Phyre2 (Kelley and Sternberg, 2009), and these revealed a conservation of the S subsite but no conservation among the other subsites (Figure 36A and 36B). These three sulfatases have the same regioselectivity for the 4-sulfate group; however, the PsS1_19B residues involved in recognition of the leaving group share no similarities with those of ZgS1_7 or ZgS1_19. The PsS1_19B residues Y455, Y458, D459 and R462 located on the C-terminal α -helix may be unique to PsS1_19B.

As with PsS1_19A, we can map residue conservation onto the PsS1_19B

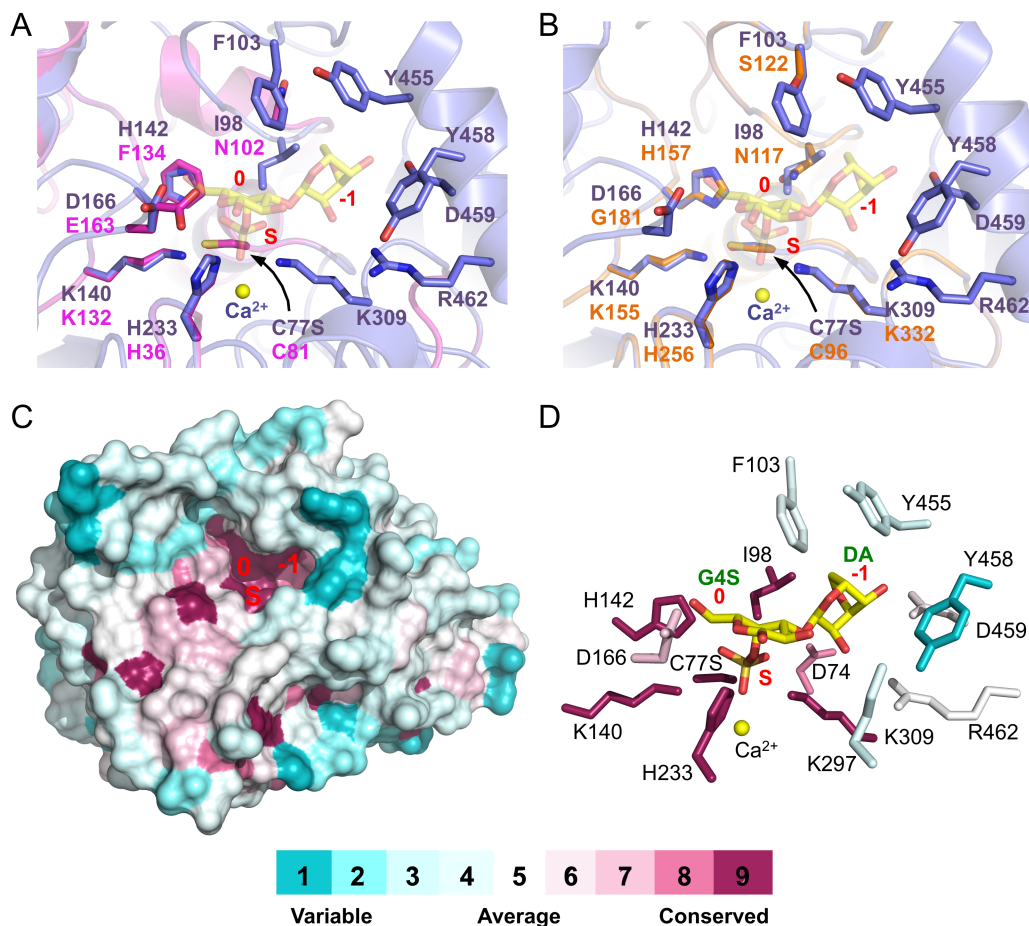


Figure 36. Conservation of residues involved in substrate recognition from family S1_19.

Homology model of ZgS1_7 (pink) and ZgS1_19 (orange) based on the PsS1_19A template, overlaid with the PsS1_19B C77S (slate) κ -NC2 substrate complex in panels (A) and (B), respectively. κ -NC2 is represented as transparent yellow sticks, with subsites labeled in red.

(C) ConSurf model of S1_19 sequences (n=430) mapped onto the structure of PsS1_19B C77S in complex with κ -NC2 (κ -NC2 not shown).

(D) The conservation of active site residues of S1_19 as calculated with ConSurf represented as sticks. Calcium atom is shown as a yellow sphere, sugar residues are labeled in green.

In panels (C) – (D) the subsites are labeled in red. The colour scheme representing degree of residue conservation for panels (C) and (D) is shown below the panels.

structure using alignments of PsS1_19B with sequences from the S1_19 subfamily using the webserver ConSurf (Goldenberg et al., 2009). The active site pocket of PsS1_19B shows conservation, yet moving further away from the S subsite the residue variability increases (Figure 36C and 36D). As expected, the S subsite is highly conserved amongst this family and, like with PsS1_19A, the residue composition of the subsites that coordinate the glycan are more variable. D116, I98, and D74 are quite well conserved; however, D459, Y458, and R462 show significant variability, which is interesting as the former three coordinate the non-reducing end hydroxyl group and the latter two coordinate the DA2S sulfate group of α -NC4 resulting in the inactive complex, respectively (Figure 36D). If residues Y458 and R462 drive the specificity for recognition of the sulfated DA2S of α -NC4 resulting in an inactive complex, it would be expected that they have a higher degree of conservation as they are involved in a very specific sulfate ester interaction. This low conservation could suggest that PsS1_19B is unique in this family, and the specialized role of these two residues is exclusive to PsS1_19B.

Recently it has been demonstrated that for *exo*-acting carrageenan GHs to show activity, the carrageenan must be stripped of its sulfate esters (Ficko-Blean et al., 2017). For ZGAL_3145 to remove the non-reducing end 2S-sulfate on the DA2S unit, the 4S-sulfate on the G4S unit (a single sugar residue in from the non-reducing end) must first be removed (Ficko-Blean et al., 2017). In the case of PS47, if we assume that when properly matured PsS1_NC exhibits *exo*-2S- α -carrageenan sulfatase activity, desulfation by PS47 would proceed as follows: the 2S-sulfate group of DA2S would be removed by PsS1_NC before PsS1_19B can hydrolyze the 4S-sulfate of G4S from the non-reducing end, as supported by the catalytically inactive PsS1_19B C77S α -NC2 complex and the

lack of any observed activity of PsS1_19B on ι -NC2 and ι -NC4 oligosaccharides. This leads to the proposal that PS47 desulfates κ - and ι -polygeenan in a sequential manner. For simplicity, I will discuss this decomposition in regards to tetrasaccharides and disaccharides; however, this holds true for larger degrees of polymerization of polygeenan. Starting from the sulfated polygeenan products discussed in chapter 2 (κ -NC4, α -NC4, and ι -NC4), desulfation occurs sequentially with PsS1_NC removing the non-reducing end 2-sulfate group (when present), followed by the removal of the 4-sulfate group by PsS1_19B resulting in a hybrid mixture of tetrasaccharides (β - κ -NC4, β - α -NC4, and β - ι -NC4; Figure 37A). The specificity of both PsS1_19B and PsS1_NC is dependent on the recognition of the non-reducing end hydroxyl group, and therefore, the β -moiety of the hybrids must first be removed, either by the hydrolysis of the α -glycosidic linkage by an *exo*- α -carrageenase followed by hydrolysis of the newly created non-reducing end β -glycosidic linkage by an *exo*- β -carrageenase (resulting in the products κ -NC2 or ι -NC2, 3,6-anhydro-D-galactose and D-galactose monosaccharides); or, alternatively, first by cleavage of the internal β -glycosidic linkage resulting in a heterogeneous mixture of products (β -NC2, α -NC2, κ -NC2, and ι -NC2). The desulfation of these resulting disaccharides would also be sequential resulting in the production of a homogenous mixture of β -NC2 (Figure 37B).

In support of the first proposed PS47 degradation pathway, an *exo*- α -carrageenase was recently identified in *Z. galactanivorans*, ZGAL_3152 (Ficko-Blean et al., 2017), and showed activity on the non-reducing end α -glycosidic linkage when the non-reducing end disaccharide moiety was unsulfated (i.e. β -). However, the PS47 genome does not contain any gene products with significant homology to the *exo*- α - and

exo- β -carrageenases from *Z. galactanivorans*. Therefore, we can hypothesize that PS47 possesses a non-homologous pathway for cleavage of the polygeenan glycosidic linkages in order for the resulting monosaccharides to enter metabolism.

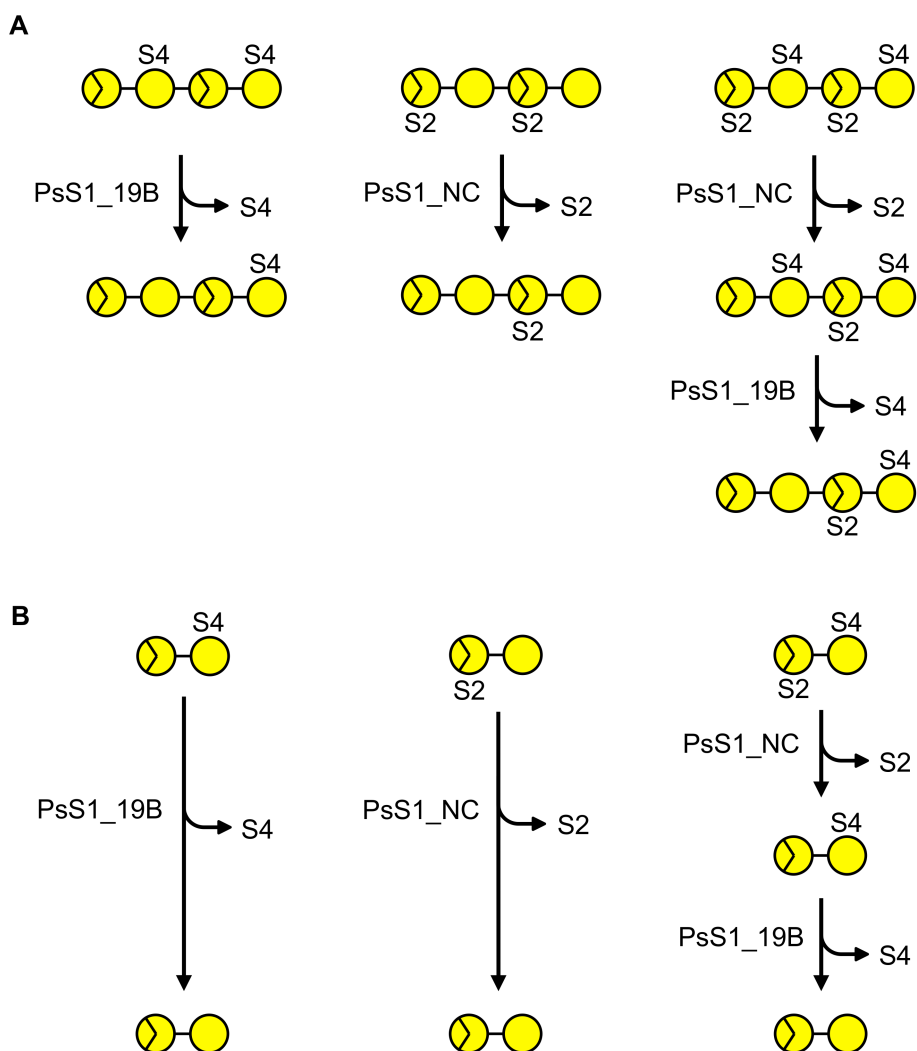


Figure 37. Cartoon schematic of products resulting from PsS1_19B and PsS1_NC sulfate ester hydrolysis.
 (A) Products from ester sulfate hydrolysis of κ -NC4, α -NC4, and *i*-NC4
 (B) Products from ester sulfate hydrolysis of κ -NC2, α -NC2, and *i*-NC2
 Representation of carrageenan components are the same as in Figure 2.

Chapter 5: The identification of an α -D-glucosidase

5.1 Introduction

In order for the complete saccharification of κ - and ι -carrageenan by PS47 there is the requirement of α -acting GHs to hydrolyze the bonds of the α - and β -glycosidic linkages of the polygeenan pool. The product of the PS47 *endo*-carrageenase and PsS1_19A will create the substrates for PsS1_19B and PsS1_NC as was presented in Chapters 3 and 4. The X-ray crystal structures of PsS1_19B and PsS1_NC in complex with their biological substrates show coordination of the non-reducing end indicating, with respect to oligosaccharides with higher degrees of polymerization than two, that for the remaining sulfate substituents to become accessible, one or more α -acting carrageenase must act next in the order of events.

As discussed previously, *Z. galactanivorans* does possess these α -acting carrageenases; ZgGH127 (ZGAL_3147) and ZgGH129 (ZGAL_3152) active on the non-reducing end α -glycosidic linkage. This hydrolysis creates the substrate for ZgGH2 (ZGAL_3633), the newly formed non-reducing end β -glycosidic linkage. None of these GHs are conserved in the PS47 carrageenan PUL indicating a non-homologous mechanism for polygeenan degradation as discussed in section 3.3. The PS47 carrageenan PUL does, however, encode for a putative GH42, the only other bioinformatically identified GH within this carrageenan PUL. There are 65 characterized family 42 entries in CAZy, with 64 showing β -galactosidase activity of which most are α -acting specific for the non-reducing end. Unlike with *Z. galactanivorans*, PsGH42 does have a homolog within the *P. atlantica* carrageenan PUL, sharing 73.9 %

sequence similarity (Figure 38), indicating a difference in polygeenan degradation between marine phyla.

In this chapter I focus on elucidating the role PsGH42 fulfills in the degradation of carrageenan by PS47 and to determine the impact the degree of sulfation has on PsGH42 catalysis. I hypothesize that PsGH42 plays a role in the final stages of carrageenan depolymerization, ie polygeenan depolymerization, and that like PsS1_19A and PsS1_19B, the sulfation pattern affects catalysis. I use linked enzyme activity assays to demonstrate the activity of PsGH42, and furthermore, to identify the molecular determinants behind substrate recognition I use X-ray crystallography for structural analysis of PsGH42.

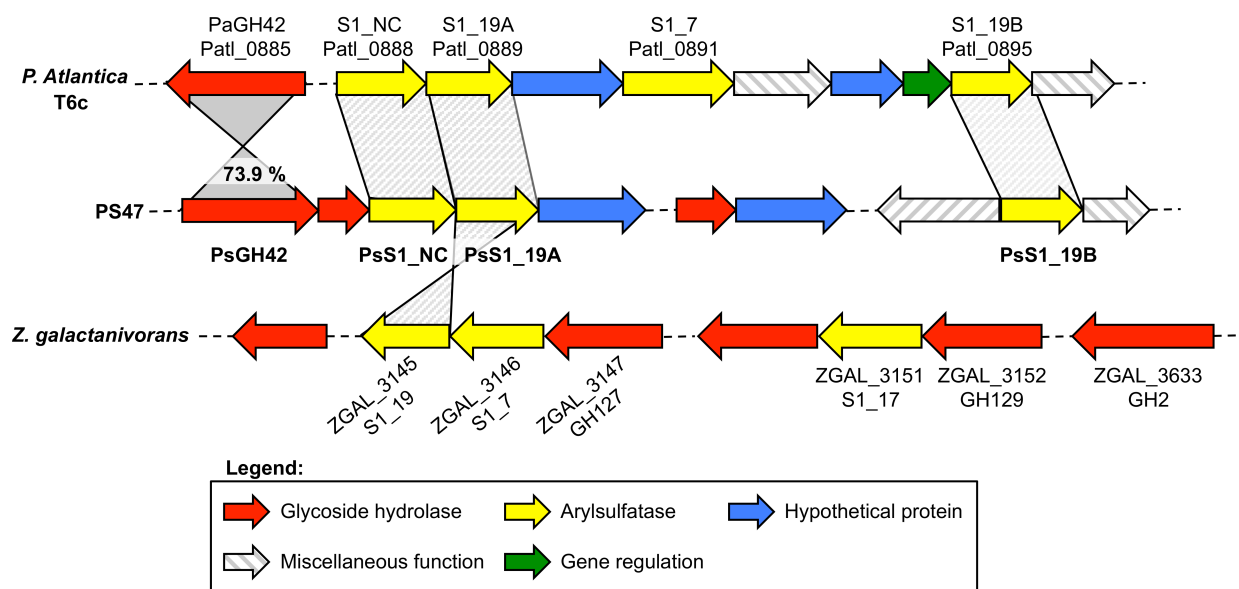


Figure 38. Known carrageenan PULs with characterized components. Genes are coloured according to protein function as per legend. Percent identity of PsGH42 homologous genes are highlighted in gray. Light gray pattern showing homologous genes with PsS1_19A, PsS1_19B, and PsS1_NC.

5.2 Results

5.2.1 Identification of an *exo*- β -carrabiosidase

As PS47 does not contain any enzymes with homology to the *Z. galactanivorans* enzymes utilized in the final stages of carrageenan depolymerization involving *exo*-acting glycoside hydrolases, we tested the hypothesis that PsGH42 fills the role of either *exo*- α - or *exo*- β -carrageenan hydrolysis. In order to test our hypothesis we cloned a truncation of PsGH42 lacking the predicted signal peptide into an expression vector and expressed it in *E. coli*. Following expression and purification, we attempted to gain insight on linkage preference by using pNP- α - and pNP- β -galactose as substrates; however, PsGH42 did not show any activity against these synthetic substrates suggesting that the molecular requirements for PsGH42 involve the occupation of more than one subsite. To address this, we tested PsGH42 for activity against commercially available β - κ -NC4 and κ -NC4 by TLC as these are the simplest carrageenan tetrasaccharides with accessible non-reducing end α - and β -glycosidic linkages. Enzymatic treatment of β - κ -NC4 with PsGH42 resulted in products of different migration distances, one of these produced bands migrated the same distance as the κ -NC2 standard (Figure 39; figure produced with the assistance of Dr. Benjamin Pluvinau). Interestingly, PsGH42 showed no activity against κ -NC4 suggesting the extra sulfate group is enough to inhibit activity. To confirm this we utilized a linked assay involving PsGH42 and PsS1_19B on κ -NC4. As discussed in Chapter 4, PsS1_19B is specific for the sulfate group on the G4S unit of the non-reducing end κ -moiety, thus creating the β - κ -NC4 substrate for PsGH42. Again the products of this linked assay are a heterogeneous mixture with different migration distances (Figure 39), indicating PsGH42

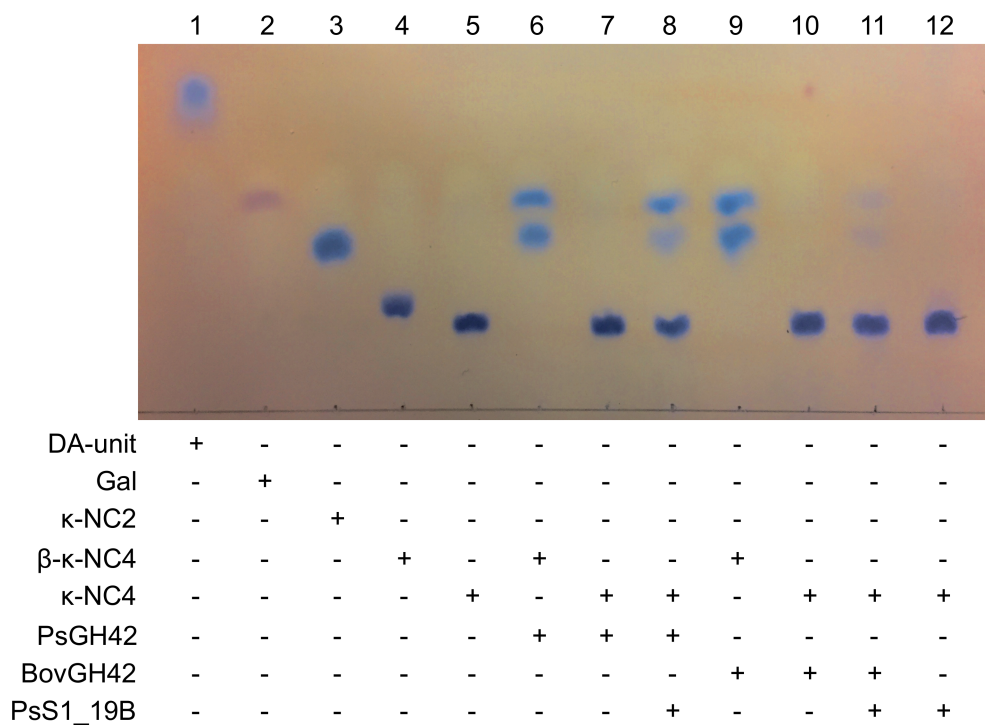


Figure 39. Activity of PsGH42 from *Pseudoalteromonas* sp. PS47. TLC analysis of PsGH42 and BovGH42 on β - κ -NC4 and κ -NC4 with and with out the presence of PsS1_19B.

is sulfate state dependent.

The heterogenous mixture created by PsGH42 activity produced κ -NC2 and a secondary band that appeared to migrate the same distance as the galactose standard. For galactose to be a product PsGH42 would have to be an *exo*- α -carrageenase; however, the other product of the *exo*- α -carrageenase activity, κ -C3, was not produced. The top most bands in lanes 6 and 8 (Figure 39) migrated the same distance as the galactose monosaccharide, however, this band is likely β -NC2 produced from the cleavage of the β -glycosidic linkage in β - κ -NC4. This strongly suggests PsGH42 is an *exo*- β -carrabiosidase, fitting with the overall model of *exo*-activity within family 42.

5.2.2 Architecture of an *exo*- β -carrageenase from GH family 42

In order to elucidate the structural requirements for the specificity of PsGH42 we sought to determine the X-ray crystal structure of the wild type PsGH42 enzyme. Unfortunately, PsGH42 did not crystalize in any of the crystallization conditions available, and therefore, we utilized a homolog from *Bacteroides ovatus* (BovGH42), which showed the same sulfate dependency (see Figure 39), to investigate the molecular determinants for activity on β - κ -NC4.

We determined the X-ray crystal structure of the wild type BovGH42 enzyme and solved the structure at 1.95 Å resolution by obtaining an iodide derivative and using SIRAS (Single Isomorphous Replacement with Anomalous Scattering) for phasing. The final refined structure contained one molecule of BovGH42 in the asymmetric unit (Figure 40A), which consisted of three domains as often seen within the GH42 family. Domain A, the catalytic domain, adopts an $(\alpha/\beta)_8$ fold with an extra α -helical bundle at

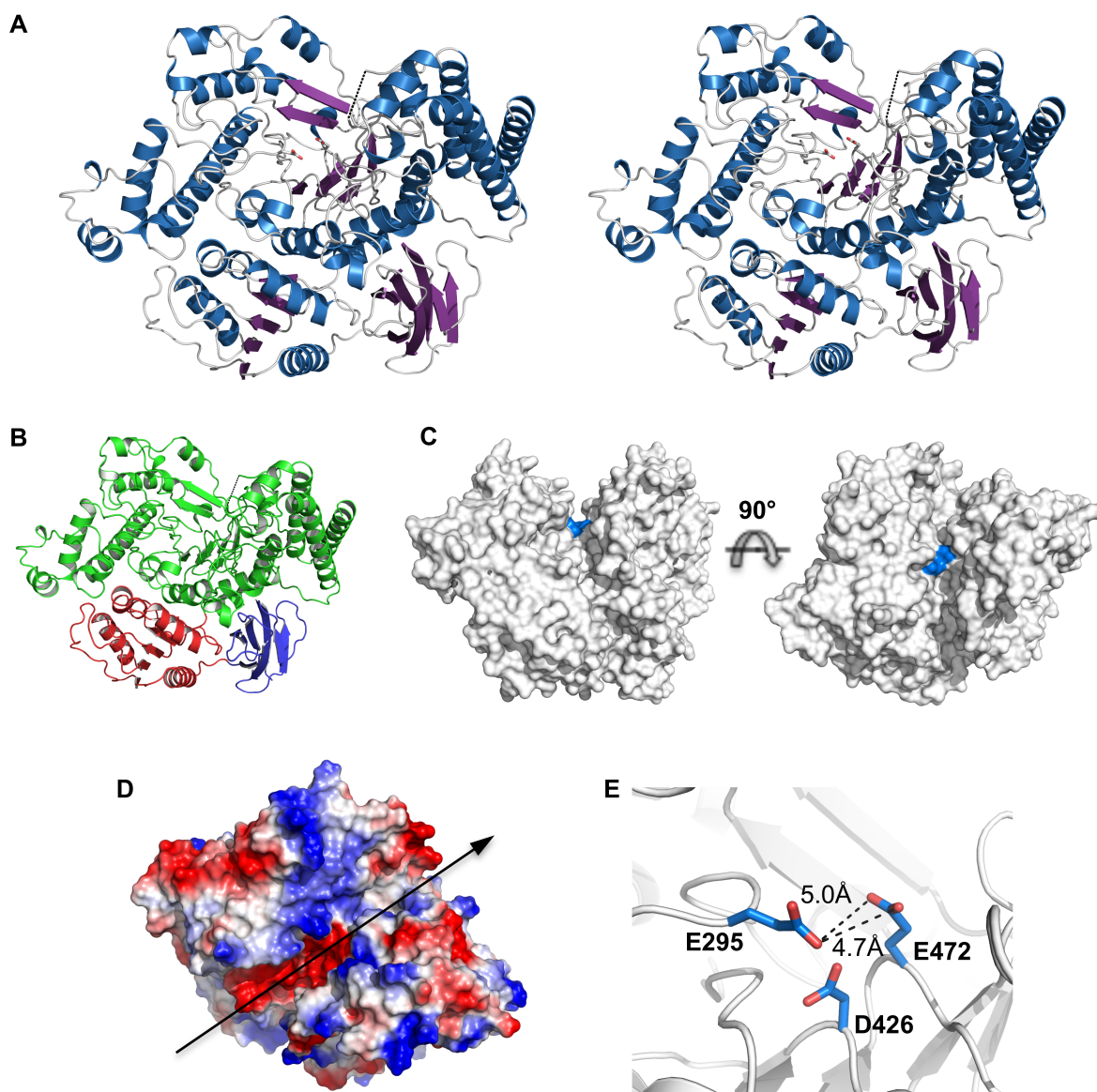


Figure 40. Structural features of BovGH42. Figure legend on following page.

Figure 40. Structural features of BovGH42.

(A) Divergent stereo view of BovGH42 showing the secondary structure elements of the $(\alpha/\beta)_8$ fold. α -helices are coloured marineblue, β -sheets coloured purple, and loops are coloured white.

(B) Cartoon representation of BovGH42 highlighting the different domains. Domains A, B, and C are coloured green, red, and blue, respectively.

(C) Solvent accessible surface showing the catalytic groove and the location of the catalytic residues. Surface coloured in white with the location of the catalytic residues coloured in blue.

(D) Solvent accessible surface of BovGH42 represented as electrostatic potential. Arrow identifies the active site groove. Surface coloured in blue and red representing basic and acidic patches, respectively.

(E) A close up view of the BovGH42 catalytic residues. BovGH42 represented as cartoon coloured white with the catalytic residues coloured in blue. Dashed lines represent distance between respective atoms.

the N-terminus not observed in other GH42 structures. Domain B is composed of a 5-stranded parallel β -sheet surrounded by α -helices, and domain C, the smallest domain, has an anti-parallel β -sandwich structure comprising two β -sheets (Figure 40B). The catalytic machinery is located centrally in a long groove comprised of residues from domain A (Figure 40C). Mapping the electrostatic potential onto the surface of BovGH42 shows a separation of the groove created by the catalytic residues of which the entrance being highly charged and the exit of the groove is more neutral (Figure 40D). Through sequence and structural alignments E295 and E472 were identified as the acid/base and the nucleophile, respectively, and are separated by a distance of ~ 5.0 Å (Figure 40E). It has been suggested that the aspartate proximal to the catalytic residues plays an important role substrate turnover (Figure 40E) (Maksimainen et al., 2012).

5.2.3 Metal binding site of the GH42 family

The overall architecture of the three domains found within the GH42 family is quite well conserved. Within domain A there is, in some instances, a metal binding site involving four cysteine residues coordinating a zinc ion as seen in Bca- β -gal from *Bacillus circulans* sp. *alkalophilus* (PDB ID: 3TTS) (Maksimainen et al., 2012) and A4- β -Gal from *Thermus thermophilus* A4 (PDB ID: 1KWG) (Hidaka et al., 2002). It is believed that the zinc plays no role in enzymatic catalysis of the enzymes but does play a stabilizing role within the catalytic domain as the B-factors for the zinc atoms in these structures are low, suggesting tight binding of the metal.

Overlaying the Bca- β -gal and A4- β -gal structures onto BovGH42, both with an r.m.s.d. value of 2.9 Å, highlights the high degree of structural conservation with the only major differences being; an extra helical bundle at the N-terminus of BovGH42 that is not found in the other GH42 structures, and a β -barrel near the C-terminal end of domain B found in Bca- β -gal and A4- β -gal but is not present in BovGH42 (Figure 41A). As mentioned, the catalytic domain of Bca- β -gal and A4- β -gal contain a metal binding site that uses four cysteines to coordinate a zinc atom. This metal binding site is not conserved in BovGH42, however the secondary structural elements are and at the equivalent location of the zinc binding cysteine residues, BovGH42 contains the residues D259, P309, S311, and A314 (Figure 41B). These BovGH42 residues share little functionality to those involved in zinc binding and may allow for domain A of BovGH42 to have more dynamic ability within the catalytic domain.

Moving towards the N-terminus from this equivalent site, BovGH42 has an extended loop that is not observed in either the Bca- β -gal or A4- β -gal structures. Extending away from C155 and C149 towards the N-terminus in the Bca- β -gal and A4- β -gal structures shows how this zinc-binding site is directly linked to the catalytic center of the enzymes; however, in the BovGH42 structure the peptide backbone initially extends away from the catalytic center creating a loop composed of residues K303, T304, and W305 that becomes one of the walls of the active site groove (Figure 41C and 41D). Without BovGH42 in complex with its biological substrate it is hard to say with any certainty that these residues are involved in the molecular recognition of β - κ -NC4.

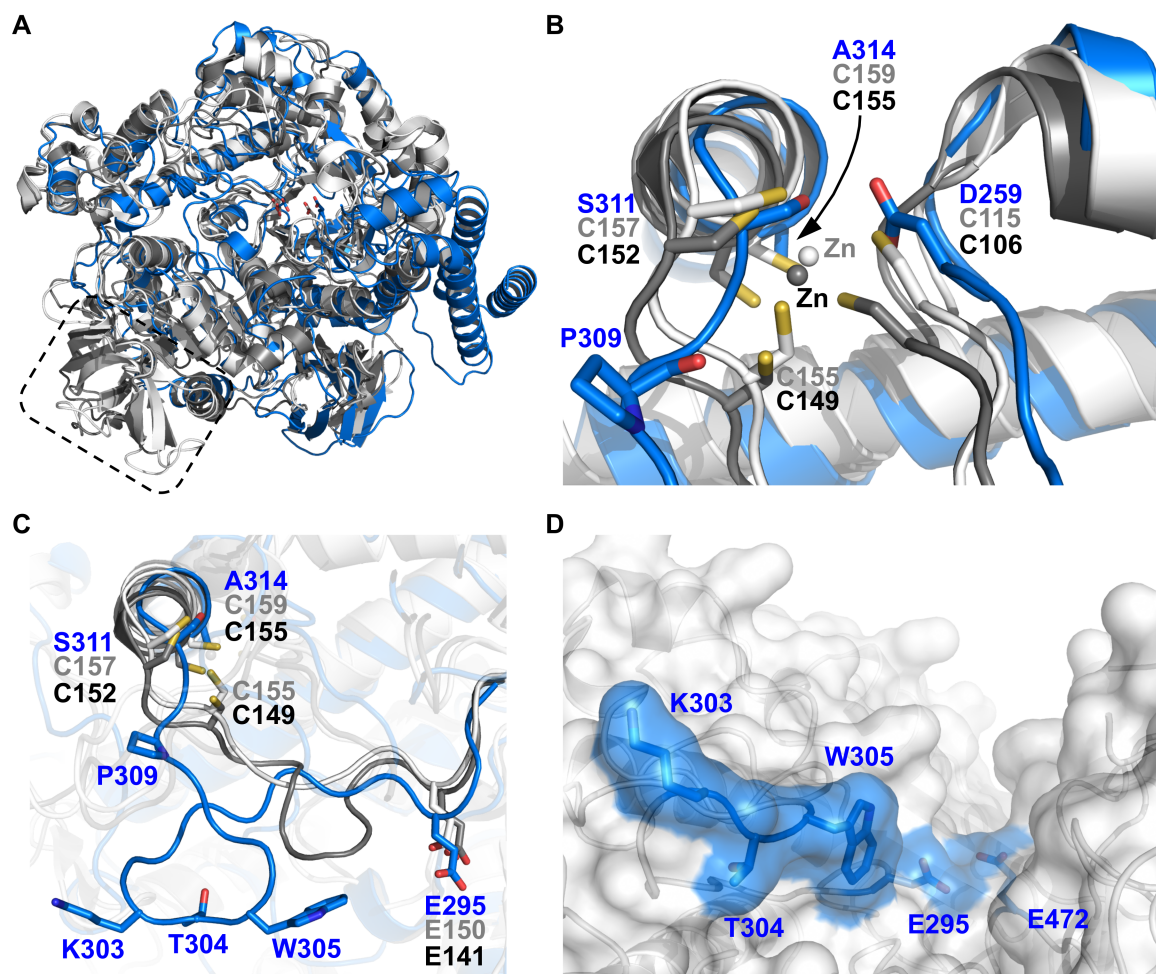


Figure 41. Structural comparison of BovGH42, Bca-β-gal, and A4-β-gal.

(A) Structural overlay of BovGH42, Bca-β-gal (PDB ID: 3TTS), and A4-β-gal (PDB ID: 1KWG). Dashed box highlighting conserved larger domain B between Bca-β-gal and A4-β-gal.

(B) Overlay of BovGH42, Bca-β-gal, and A4-β-gal at the metal binding site conserved among most GH42 enzymes.

(C) Looping secondary structure of BovGH42, Bca-β-gal, and A4-β-gal extending away from the metal binding site.

(D) Solvent accessible surface representation of the active site groove of BovGH42 highlighting the residues extending away from the metal binding site that line the active site groove.

In panels (A) – (C) BovGH42, Bca-β-gal, and A4-β-gal are coloured blue, white, and gray50, respectively. Atom labels are coloured the same as their cartoon representations

5.3 Discussion

Carrageenolytic bacteria with sequenced genomes contain genes encoding for *endo*-acting GH16 and/or GH82 enzymes, specific for the β -1,4-glycosidic linkage, along with sulfatases, have allowed for the bioinformatic identification of carrageenan PULs. For example, *Z. galactanivorans* utilizes both GH16s and GH82s to *endo*-lytically digest κ - and ι -carrageenan, respectively, whereas PS47s carrageenan PUL only contains GH16s even though the PsS1_19A and PsS1_NC data strongly suggests that PS47 is an ι -carrageenan specialist. Whichever combination of enzymes the organism employs the resulting polygeenan pool will require *exo*-acting α - and β -carrageenases. These *exo*-acting carrageenases have for a long time been hypothesized to exist in nature but only very recently been identified. Ficko-Blean et al. (2017) identified two *exo*-acting α -carrageenases (ZgGH127 and ZgGH129) and one *exo*-acting β -carrageenase (ZgGH2) (Ficko-Blean et al., 2017), none of which have a homolog within the PS47 carrageenan PUL. The *exo*- β -carrageenase activity of PsGH42 fills this role within the PS47 degradative pathway as opposed to its counterpart in *Z. galactanivorans*, ZgGH2. In the CAZy database both GH2 and GH42 families contain characterized entrants with β -galactosidase activity, therefore, it is not surprising that PS47 and *Z. galactanivorans* utilize these GH families given the glycan backbone of carrageenan is composed of galactose or galactose derivatives; however, neither family has been updated with the activity of β -carrageenase.

The activity of PsGH42 by TLC shows cleavage of the β - κ -NC4 hybrid into its disaccharide components of β - and κ -NC2. These products support the *exo*- β -carrageenase conclusion yet they do not fully address the sulfate pattern prerequisites

for hydrolysis. The lack of activity on κ -NC4 suggests that the recognition of a single extra sulfate group can inhibit catalysis, an observation similar to that of PsS1_19B on ι -NC4. Including PsS1_19B in a linked assay with PsGH42 on κ -NC4 shows that PsGH42 is dependent on PsS1_19B activity removing the 4-sulfate group on the reducing end disaccharide κ -moiety. Again, like with PsS1_19B, there is a dependency on the pattern of sulfation for catalysis.

To elucidate the molecular interactions imparting this specificity, a homology model of PsGH42 was generated and using the structure of BovGH42 with the Web server Phyre2 as a template reveals some interesting structural characteristics in the regions lining the active site groove. The looping region that lines the BovGH42 active site groove composed of residues K303, T304, and W305 has an extended counterpart in the PsGH42 homology model. This extended loop has a similar charged and aromatic characteristic composed of residues W297, T298, Q299, K300, and F301 (Figure 42A). The residues Y255 and F247 from the BovGH42 and PsGH42 structures, respectively, are situated proximal to the catalytic residues, possibly also playing a role in β - κ -NC4 recognition (Figure 42A).

Recently Viborg et al. (2017) applied a phylogenetic and sequence analysis to 150 enzymes containing the GH42 catalytic domain from different bacterial species and subspecies of various phyla. This analysis separated the sequences into two groups; subfamily A showing α -L-arabinopyranoside specificity of which all contained a unique sequence motif Trp – X₁₂ – Trp, and subfamily G showing β -galactosidase specificity containing the sequence motif Phe – X₁₂ – His. Interestingly, both BovGH42 and PsGH42 contain the unique Trp – X₁₂ – Trp motif suggesting they belong to subfamily A;

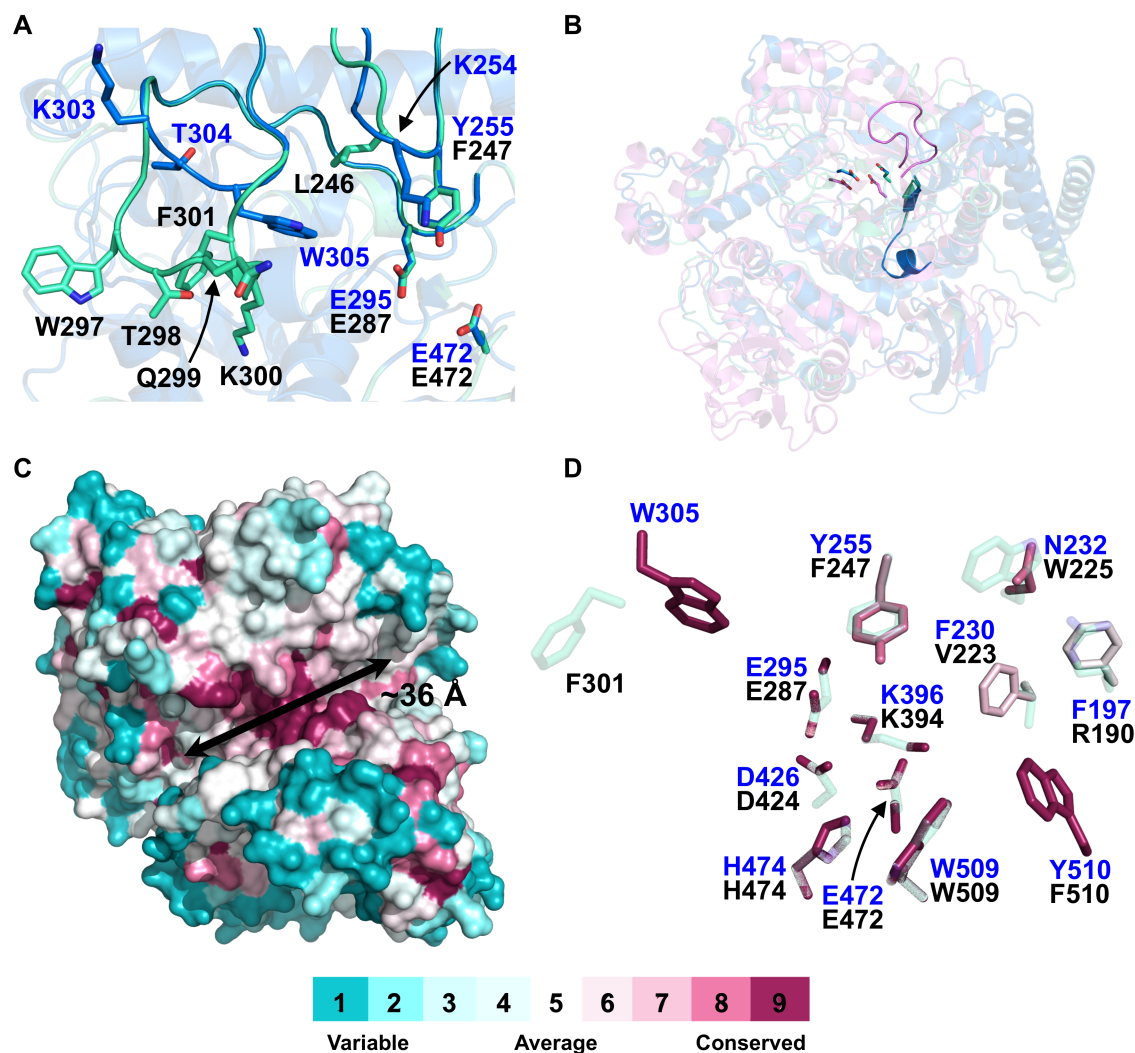


Figure 42. Conservation of residues among the family GH42.

(A) Homology model of PsGH42 (greencyan) based on the BovGH42 template, overlaid with the BovGH42 wild type structure.

(B) *BIARap42B* (pink) structure overlaid with the BovGH42 (marineblue) wild type and PsGH42 (greencyan) homology structures, highlighting the Trp – X₁₂ – Trp subfamily A sequence motif. All structures shown as transparent cartoon representation except for the sequence motifs. Catalytic residues shown as sticks coloured according to structure.

(C) ConSurf model of GH42 sequences (n = 150) mapped onto the structure of BovGH42.

(D) The conservation of the active site groove residues from family GH42 as calculated with ConSurf represented as sticks, overlaid with the PsGH42 homology model active site groove residues represented as transparent greencyan sticks.

In panels (A) and (D) BovGH42 and PsGH42 residues labeled in marineblue and black, respectively. The colour scheme representing degree of residue conservation for (C) and (D) is shown below the panels.

however, they exhibit β -carrageenase (galactosidase) activity suggesting a belonging to subfamily G. When BovGH42 and PsGH42 are overlaid onto *BlArap42B* from *Bifidobacterium* sp. (Figure 42B) it becomes apparent that this 14 amino acid motif in *BlArap42B* (PDB ID: 5XB7) (Viborg et al., 2017) lines the active site groove and plays a role in specificity, whereas in the BovGH42 and PsGH42 structures this motif is internal, likely playing more of a structural role.

Mapping the amino acid sequence conservation of 150 representative GH42 sequences onto the structure of BovGH42 using the ConSurf Web server shows a high level of sequence identity around the catalytic residues and the residues lining the active site groove; however, outside of this active site groove the residues conservation is quite low (Figure 42C). Highlighting the residues in the active site groove with the highest degree of conservation and overlaying the PsGH42 homology model shows more variability in the residues the more distally located from catalytic residues, E295 and E472, they are (Figure 42D). It is likely that some of these residues would interact with the sulfate group on the non-reducing end G4S residue of κ -NC4 inhibiting hydrolysis, however, without BovGH42 in complex with either the κ -NC4 or the β - κ -NC4 hybrid molecule it is hard to determine which residues would be involved in recognition.

Even though the β - κ -NC4 and κ -NC4 complexes were not obtained, the TLC activity data shows the dependency of PsGH42 on the polygeenan degree of sulfation and indicates the PsGH42 activity likely only releases an unsulfated β -NC2. This function would come in the final stages of polygeenan depolymerization. In keeping with the example of sulfated tetrasaccharides, the resulting polygeenan pool from the PS47 *endo*- β -carrageenase, PsS1_19A, PsS1_19B, and PsS1_NC activity would be β - κ -NC4,

β - α -NC4, and β -I-NC4. The activity of PsGH42 on these tetrasaccharides would produce a heterogeneous pool of β -NC2, κ -NC2, α -NC2, and I-NC2 (Figure 43) ultimately with an enriched amount of the β -NC2 disaccharide.

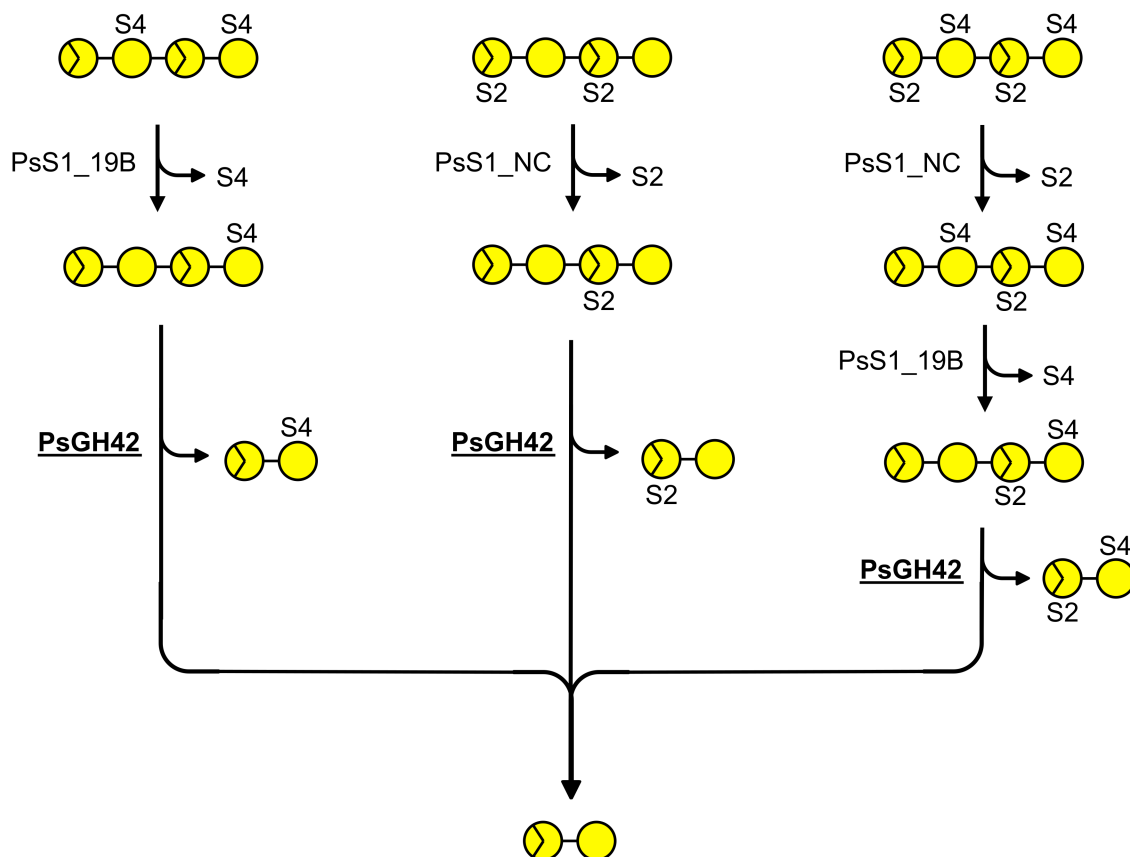


Figure 43. Cartoon schematic of products from PsGH42 hydrolysis. Representation of carrageenan components are the same as in Figure 2.

Chapter 6: General discussion

6.1 PS47 contains the machinery to desulfate and depolymerize κ - and ι -carrageenan

Through enzymatic and structural studies presented in Chapters 3, 4, and 5 the genes encoding carrageenan active enzymes; PsS1_19A an endo-4S- ι -carrageenan sulfatase, PsS1_19B and PsS1_NC are exo-4S- κ -carrageenan and exo-2S- ι -carrageenan sulfatases, respectively, and PsGH42 an exo- β -carrabiosidase, are identified and discussed. The study of these four enzymes confirms the PS47 carrageenan PUL contains the machinery necessary to desulfate and depolymerize κ - and ι -carrageenan with each component being specific for a particular carrageenan substructure.

The alluded to promiscuity of the *endo*- β -carrageenase activity of the GH16s is likely how PSU2A, and through extension PS47, is able to grow on ι -carrageenan even though its genome does not encode for any carrageenases from family 82. This leads to the notion that the GH16s in the PS47 genome are likely able to hydrolyze 3,6-anhydro-D-galactose containing carrageenan families regardless of the degree of sulfation, meaning α -, β -, κ -, and ι -carrageenan are all potential substrates. The possibility of the PsGH16s and PsS1_19A not being dependent on one another would lead to a pool of polygeenan with differing degrees of sulfation (Figure 44A) creating different hybrid oligosaccharides. Again for simplicity, I will discuss this hybrid polygeenan pool as all having a degree of polymerization of four, however, again the overall process of

Figure 44. Saccharification of κ - and ι - carrageenan by PS47.
Figure legend on following page.

Figure 44. Saccharification of κ - and ι - carrageenan by PS47.

(A) Initial desulfation and depolymerization of full-length κ/ι -hybrid carrageenan by PsS1_19A and PsGH16.

(B) Desulfation of α -, ι -, and κ -NC4 by PsS1_19B and PsS_NC.

(C) Final desulfation and depolymerization of polygeenan by PsS1_19B, PsS1_NC, and PsGH42 resulting in β -NC2.

In all panels the representation of carrageenan components the same as in Figure 2.

degradation of polygeenan oligosaccharides can be applied to greater degrees of polymerization. As seen in the structural and/or functional data, both PsS1_NC and PsS1_19B are *exo*-acting with specificity for sulfate esters proximal to the non-reducing end, and therefore the coordination of the non-reducing end becomes crucial for substrate turnover. Because of this coordination the desulfation of these oligosaccharides cannot go any further once the non-reducing end disaccharide moiety is remodeled into β -carrageenan (Figure 44B). The lack of activity observed for PsS1_19B on I-NC2 and I-NC4 in conjunction with the PsS1_19B C77S I-NC4 catalytically inactive complex strongly suggests that the desulfation of I-moieties at the non-reducing end occurs sequentially and that PS47s break down of I-carrageenan is solely dependent on the activity of PsS1_NC whose product would be the non-reducing end κ -moiety, the substrate for PsS1_19B.

As these hybrid tetrasaccharides can no longer be desulfated there is the necessity to make the reducing end sulfated moieties accessible for PsS1_NC and PsS1_19B. As seen in Chapter 4, PsGH42 shows *exo*- β -carrageenase activity specific for unsulfated, non-reducing end β -carrageenan moieties cleaving the β -glycosidic linkage releasing β -NC2. In the tetrasaccharide example this would result in a heterogenous mixture of β -NC2, and I-NC2, κ -NC2, or α -NC2 carrageenan disaccharides (Figure 44C). Once these are produced PsS1_NC and PsS1_19B are able to perform their roles again resulting in the production of a homogenous pool of β -NC2. The resulting enriched β -NC2 disaccharide pool would then be hydrolyzed by the *exo*- α -carrageenase, resulting in the production of D-galactose and 3,6-anhydro-D-

galactose monosaccharides completing saccharification of carrageenan, however, this *exo*- α -carrageenase is yet to be identified within PS47.

The enzymes that function in the final step of degradation creating the monosaccharides that would enter metabolism are typically accepted as the keystone enzymes in a degradative pathway, as in a genetic knockout of this gene would be fatal and result in the organisms inability to grow on the carbohydrate as a sole carbon source. This role being fulfilled by the PS47 *exo*- α -carrageenase, however, in respect to *i*-carrageenan metabolism PsS1_NC should also be considered a pseudo-keystone enzyme as the *i*-polygeenan pool would have the non-reducing ends decorated with the 2-sulfate group. If PsS1_NC were knocked out there would also prove to be fatal, as any limited growth on full-length *i*-carrageenan would likely be do to sparse hybrid regions of κ -carrageenan of which would be exhausted as a carbon source rather quickly. This sulfatase knockout fatal phenotype would likely only be the case for PsS1_NC as PS47 contains a regioselective redundant system to account for the 4-sulfate groups of κ - and *i*-carrageenan, the *endo*-4S-*i*-carrageenan sulfatase and the *exo*-4S- κ -carrageenan sulfatase, PsS1_19A and PsS1_19B, respectively. This would likely lead to PS47 limping along if only one of these genes were knocked out.

The data presented here suggest that the saccharification of carrageenan by PS47 proceeds with two bottleneck points, the removal of the 2-sulfate group by PsS1_NC and the hydrolysis of the β -NC2 α -glycoside linkage by the unidentified PS47 *exo*- α -carrageenase. This is largely consistent with the conclusion that the PS47 carrageenan PUL enzymes are dependent on the carrageenan degree of sulfation and on each other, thus indicating a pathway that functions in a sequential manner. The

proposed pathway by *Z. galactanivorans* for κ - and ι -carrageenan degradation utilizes GH16s and GH82s along with three sulfatases in a redundant system. However, these three sulfatases belong to different S1 subfamilies two of which are proposed to be *endo*-acting. The PS47 carrageenan degradation pathway shares a strong resemblance to that proposed for *P. carrageenovora*. To date there have been few biochemical characterizations of *P. carrageenovora* sulfatases, giving rise to the possibility that the PS47 and *P. carrageenovora* pathways diverge at the sulfatase level. With the lack of homology between the carrageenan degradation machinery of *Zobellia* and *Pseudoalteromonas*, the notion that it is the function of the PUL genes responsible for polysaccharide degradation that dictates population distribution of algal-associated microorganisms is strongly supported.

6.2 Recognition of the leaving group is critical for S1 sulfatase activity

6.2.1 The C-terminal subdomain of family S1 sulfatases plays an intricate role in substrate recognition

Structural studies of sulfatases in complex with their physiological substrates have only very recently been achieved (Cartmell et al., 2017; Hettle et al., 2018). Here, the structural analysis of three PS47 sulfatases in complex with their physiological substrates shows very specific molecular interactions with monosaccharides comprising the substrate occupying subsites outside of the conserved S subsite. The X-ray crystal structures were determined of; PsS1_19A in complex κ - ι - κ -NC6, κ -C5, ι -NC4, PsS1_19B in complex with κ -NC2 and ι -NC4, and PsS1_NC in complex with ι -NC2 and κ -NC2 all showing unique molecular recognition of the leaving group supporting the

conclusion that these interactions drive substrate turnover and specificity for family S1 sulfatases.

The early structural studies of wild type apo-sulfatase structures highlighted the two distinct subdomains based, as well as the structural motifs contributing to these subdomains. As mentioned previously, these early studies identified that the larger N-terminal subdomain aids in positioning the catalytic machinery and conserved polar residues near the base of the catalytic pocket, but the role of the smaller C-terminal subdomain remained speculative likely playing a role in substrate discrimination as this domain possess most of the structural diversity among S1 sulfatases. Analyzing the residues involved in substrate recognition for both PsS1_NC and PsS1_19B, it becomes apparent that the C-terminal subdomain contributes few molecular interactions in substrate recognition, but these few interactions are likely very important and crucial for specificity and substrate turnover. In the case of PsS1_NC, almost the entire catalytic pocket for the recognition of ι -NC2 is composed of N-terminal subdomain residues apart from N427 and Y426 (Figure 45A). Both of these residues are situated at the very bottom of the active site pocket. PsS1_19B also contains C-terminal subdomain residues that are likely crucial for the substrate recognition of κ -NC2 and κ -NC4 through the coordination of the non-reducing end DA unit in the -1 subsite by residues D459 and R462 (Figure 45B), the latter of which is also involved in the recognition of ι -NC4 forming the catalytically inactive complex.

The contribution by C-terminal subdomain residues in the PsS1_NC and PsS1_19B complexes shows that this domain is crucial in carrageenan-active exo-sulfatases. Analyzing the contribution of C-terminal subdomain residues from the

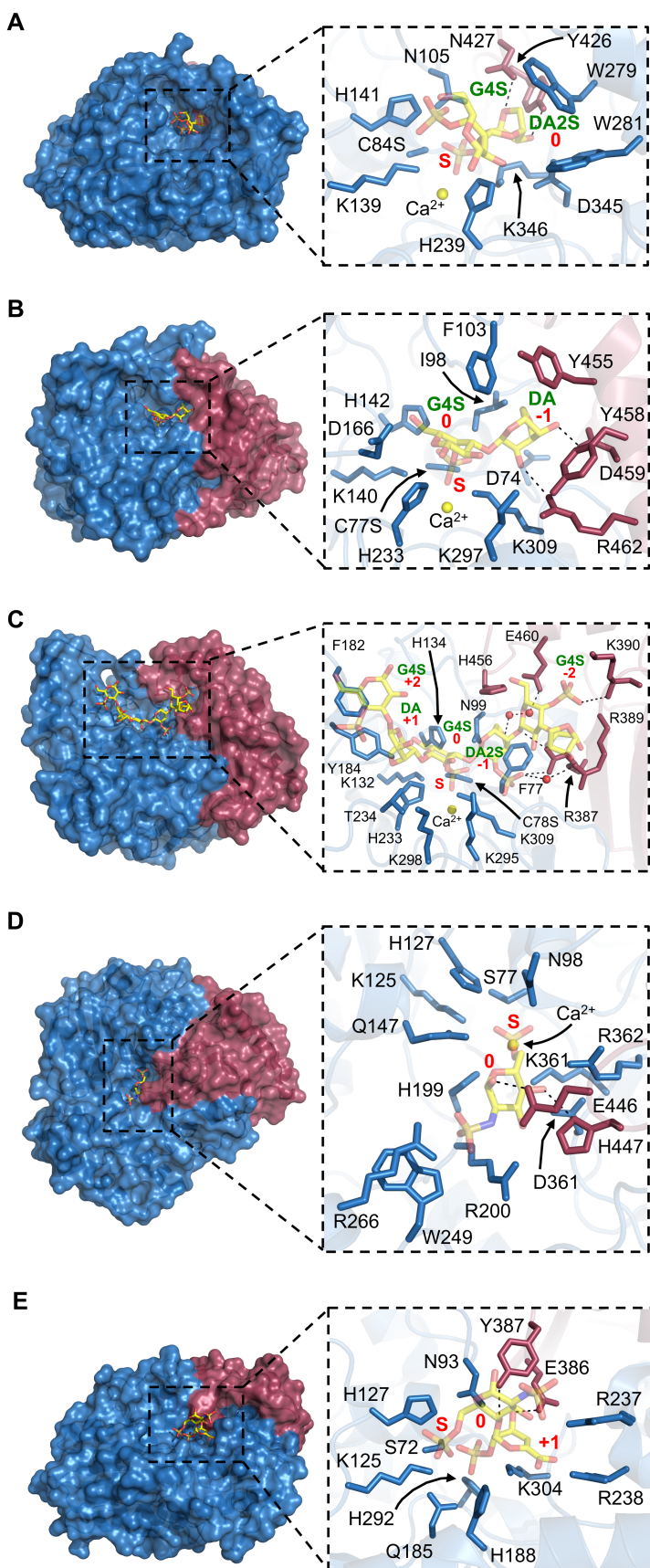


Figure 45. The importance of C-terminal subdomain residues in S1 family sulfatase substrate recognition.

(A) PsS1_NC C84S in complex with i-NC4.

(B) PsS1_19B C77S in complex with κ-NC2.

(C) PsS1_19A C78S in complex with κ-i-κ-NC6.

(D) BT4656^{6S-sulf} in complex with N,O6-disulfo-glucosamine.

(E) BT1596^{2S-sulf} in complex with Δ4,5UA2Sβ1-4GlcNS6S.

In all panels the solvent accessible surface for the N-terminal subdomain is coloured blue and for the C-terminal subdomain in ruby, all ligands are shown as yellow sticks. For insets residues from the N-terminal subdomain are coloured in blue, for the C-terminal subdomain in ruby, cartoon representations are transparent and coloured according to their respective subdomain, and ligands are coloured as partially transparent yellow sticks. Subsites are numbered in red.

PsS1_19A C78S κ -I- κ -NC6 complex it becomes apparent that this observation is not exclusive to *exo*-acting sulfatases. The recognition of the DA2S unit by the cradle in the -1 subsite is composed of F77 and H456 having each subdomain contributing a residue, while the recognition of the 2-sulfate of the DA2S unit in the -1 subsite and the residues comprising the -2 subsite are exclusively C-terminal subdomain residues (Figure 45C). This open C-terminal subdomain architecture in the *endo*-acting carrageenan sulfatase is what allows for the accommodation of an extended carbohydrate substrate.

The importance of the minimal C-terminal subdomain residues involved in recognition are not exclusive to carrageenan-active sulfatases, but applies to the broader S1 family as apparent in the recent publication of carbohydrate specific sulfatases in complex with sulfated glycosaminoglycans (Cartmell et al., 2017). The structures of *B. thetaiotaomicron* sulfatases in complex with their respective biological substrates also support the observation of C-terminal residue importance in substrate discrimination. In each of the BT4656^{6S-sulf} and BT1596^{2S-sulf} complexes there are two C-terminal subdomain residues involved in substrate recognition. E446 and H447 of the BT4656^{6S-sulf} GlcNS6S complex (Figure 45D), and residues E386 and Y387 of the BT1596^{2S-sulf} Δ UA2S-GlcNS6S complex (Figure 45E) are the contributing residues from the C-terminal subdomain, respectively.

In the case of the four reported *exo*-acting sulfatase complexes (PsS1_19B, PsS1_NC, BT4656^{6S-sulf} and BT1596^{2S-sulf}), the vital C-terminal subdomain residues are almost always proximal to each other within the primary polypeptide sequence: PsS1_NC residues Y426 and N427, PsS1_19B residues D459 and R462, (S1_11) BT4656^{6S-sulf} residues E446 and H447; and finally (S1_9) BT1596^{2S-sulf} residues E386

and Y387. Not only are these residues almost always neighbouring, they are, in all cases, within 100 residues of the C-terminus. The participating residues from the C-terminal subdomain in PsS1_19A are, like the exo-acting sulfatase residues, neighbouring, however, in this *endo*-acting example there are separate pairs of residues: R387 with R389 and H456 with E460.

These five examples are the only structures of S1 sulfatases in complex with their biological substrates, yet there are a total of 14 S1 subfamilies that do have an apoenzyme structure associated with them (including the S1_9, S1_11, and S1_19 subfamilies mentioned previously, however, not the S1_NC family). When looking at the active sites of these reported structures it can be noted that the C-terminal subdomain contributes, in all cases, at least one residue that is either in close proximity to the catalytic residue or lines the boundary of the active site pocket (Figure 46A-J). As these structures do not have their biological substrates bound, it is yet to be determined whether these residues do actually participate in substrate recognition. The residues forming the loop between the C-terminal subdomain β -sheet components β A and β B, seem to be crucial in substrate recognition for all but two S1 subfamilies, S1_3 and S1_5 (Figure 46C and 46E, respectively). All other families reported have residues from this loop contributing to either substrate recognition, PsS1_NC (Figure 45A), PsS1_19A (Figure 45C), S1_11 (Figure 45D), and S1_9 (Figure 45E); or are involved in forming the active site pocket, S1_1, S1_2, S1_4, S1_7, S1_8, S1_12, S1_23, and S1_27 (Figure 46A, 46B, 46D, 46F, 46G, 46H, 46I, and 46J, respectively). The S1_19 family in one complex shows this loop is involved in substrate recognition and in the other complex from this subfamily the loop is not involved in specificity, PsS1_19A (Figure

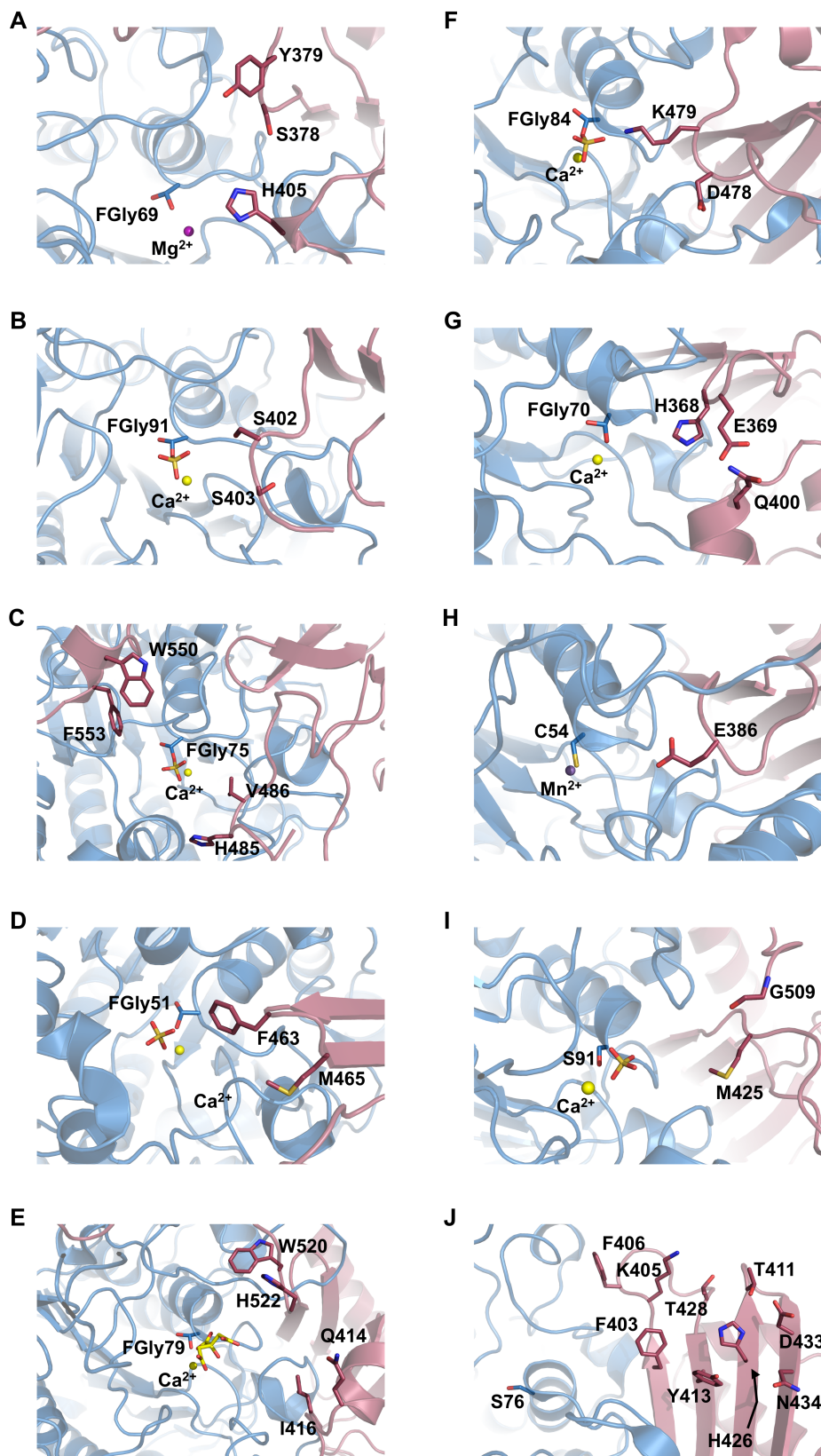


Figure 46. C-terminal subdomain comparison of known X-ray crystal structures from family S1 sulfatases. Figure legend on following page.

Figure 46. C-terminal subdomain comparison of known X-ray crystal structures from family S1 sulfatases.

(A) *H. sapiens* cerebroside sulfatase, subfamily S1_1 (PDB ID: 1AUK) (Lukatela et al., 1998), (B) *H. sapiens* N-acetylgalactosamine-4-sulfatase, subfamily S1_2 (PDB ID: 1FSU) (Bond et al., 1997), (C) *H. sapiens* steryl-sulfatase, subfamily S1_3 (PDB ID: 1P49) (Hernandez-Guzman et al., 2003), (D) *P. aeruginosa*, arylsulfatase, subfamily S1_4 (PDB ID: 1HDH) (Boltes et al., 2001), (E) *H. sapiens* N-acetylgalactosamine-6-sulfatase, subfamily S1_5 (PDB ID: 4FDI) (Rivera-Colón et al., 2012), (F) *H. sapiens* iduronate-2-sulfatase, subfamily S1_7 (PDB ID: 5FQL) (Demydchuk et al., 2017), (G) *H. sapiens* N-sulfoglucosamine sulfohydrolase, subfamily S1_8 (PDB ID: 4MIV) (Sidhu et al., 2014), (H) *S. melliloti* choline sulfatase, subfamily S1_12 (PDB ID: 4UG4) (van Loo et al., 2018), (I) *E. coli* putative arylsulfatase, subfamily S1_23 (PDB ID: 3ED4) (unpublished), (J) *B. fragilis* putative arylsulfatase yidJ, subfamily S1_27 (PDB ID: 2QZU) (unpublished).

In all panels cartoon representation and associated amino acid side chains are coloured according to subdomain, N-terminal subdomain in blue and C-terminal subdomain in ruby. In panels (B) – (G) and (I) calcium ions are represented as yellow spheres, in panel (A) magnesium ion is represented as a pink sphere, and in panel (H) manganese ion as a purple sphere. In panels (B) – (D) and (F) sulfate ions and citrate, panel (E), are shown as yellow sticks.

45C), S1_11 (Figure 45D), and S1_9 (Figure 45E); or are involved in forming the active site pocket, S1_1, S1_2, S1_4, S1_7, S1_8, S1_12, S1_23, and S1_27 (Figure 46A, 46B, 46D, 46F, 46G, 46H, 46I, and 46J, respectively). The S1_19 family in one complex shows this loop is involved in substrate recognition and in the other complex from this subfamily the loop is not involved in specificity, PsS1_19A (Figure 46C) and PsS1_19B (Figure 46B), respectively. As seen with PsS1_19A and PsS1_19B the C-terminal subdomain aids in coordinating the substrate leaving group indicating this recognition confers specificity.

6.2.2 The recognition of 3,6-anhydro-D-galactose by carrageenan active enzymes

The C-terminal subdomain provides a component of substrate recognition of the structurally characterized sulfatases in complex with their biological substrates, and through the analysis of the remaining S1 subfamilies with a known structure it becomes apparent that the C-terminal subdomain likely plays a critical role in substrate specificity. For the case of the PS47 sulfatases this subdomain not only participates, but it provides the residues that specifically interact with the DA or DA2S units, and thus this recognition of the 3,6-anhydro-D-galactose residue is necessary for substrate specificity and catalytic efficiency of the three PS47 sulfatases. PsS1_NC utilizes a tryptophan cradle created by W279 and W281 to correctly position the sulfate ester for cleavage (Figure 47A). PsS1_19A uses F77 and H456 to create a cradle that coordinates the 3,6-anhydrous bridge of the DA unit in the -1 subsite (Figure 47B). Finally, PsS1_19B uses an aromatic shelf composed of the residues F103 and Y455 to coordinate the DA unit leaving the C2 hydroxyl group relatively solvent exposed (Figure 47C), allowing for the

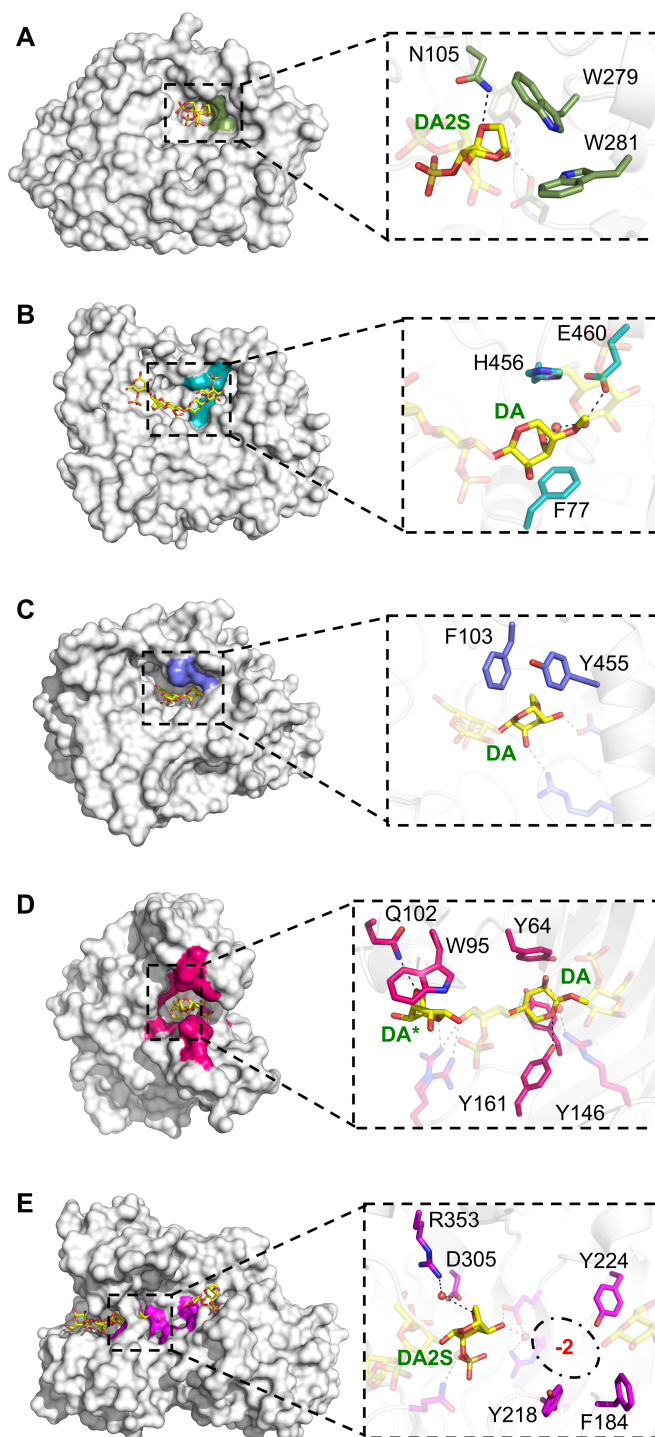


Figure 47. The molecular interactions of 3,6-anhydro-D-galactose recognition

(A) PsS1_NC C84S in complex with I-NC2, (B) PsS1_19A C78S in complex with κ -I- κ -NC6 (PDB ID: 6B0J), (C) PsS1_19B C77S in complex with κ -NC2, (D) *P. carrageenovora* GH16 in complex with κ -NC4 (PDB ID: 5OCQ), (E) *A. fortis* GH82 in complex with I-carrageenan fragments (PDB ID: 1KTW).

In all panels enzymes are represented as solvent accessible surfaces in gray with the relevant 3,6-anhydro-D-galactose binding sites coloured according to the amino acid side chain colouring represented as sticks in the inset of each panel; green (A), teal (B), slate (C), hotpink (D), magenta (E), dirtyviolet. The 3,6-anhydro-D-galactose residues are coloured as yellow sticks and labeled in green, non-interacting sugar residues are coloured partially transparent yellow sticks, water molecules as red spheres, hydrogen bonds are represented as dashed lines, and the asterisks identify 3,6-anhydro-D-galactose residue proximal to the non-reducing end. In panel (E) the -2 subsite is labeled in red and highlight by a dashed circle.

recognition of a DA2S unit. All the sulfatases utilize aromatic residues in a non-traditional way to interact with the DA and DA2S units of carrageenan. These sulfatases are not the first reports of carrageenan active enzymes in complex with their biological substrates, and therefore, provides an opportunity to analyze the mechanisms employed by carrageenan active enzymes in 3,6-anhydro-D-galactose recognition.

The first structures in complex with carrageenan oligosaccharides were *endo*-acting GHs from families 16 and 82. The GH16 from *P. carrageenovora* in complex with κ -NC4 (PDB ID: 5OCQ) (Matard-Mann et al., 2017) shows some of the similar interactions as seen with sulfatases when coordinating DA units. At the non-reducing end W95 is creating a platform that interacts with the C2-C3-O6 face of the DA unit similar to the coordination by PsS1_NC W279, however, on the opposite face of the residue. The other DA unit is also cradled, here, by the residues Y64 and Y146 (Figure 47D). The other family with known carrageenan activity, family GH82 specific for *i*-carrageenan moieties also has a known structure from *A. fortis* in complex with *i*-NC4 and *i*-NC2 fragments (PDB ID: 1KTW) (Michel et al., 2003). The DA2S unit at the non-reducing end of the *i*-NC4 is predominantly coordinated by a direct and water mediated hydrogen bond network. If this were the intact substrate there would be an *i*-disaccharide moiety connecting the two fragments of the complex which would create a continuous *i*-NC8 molecule bound in the active site. The residues composing what would be the -2 subsite are aromatic residues, and may coordinated a DA2S with a cradle as seen with the other carrageenan active enzymes (Figure 47E). If this held true, than DA and DA2S units of carrageenan would predominantly be coordinated by aromatic cradle motifs.

In almost all cases there is the coordination of the 3,6-anhydrous bridging oxygen via hydrogen bonding, and interestingly, there are no instances of an aromatic residue creating a platform for the pyranose ring of the residue. For the recognition of DA units the aromatic cradle allows for the C2 hydroxyl group to remain in positions that could allow for a 2-sulfate ester substituents, the difference between κ - and ι -carrageenan. This allows for the differences in catalysis observed among the sulfatases and, possibly aids in some GH promiscuity allowing for activity on different carrageenan classes.

It has been alluded to that the recognition of sugars with a 3,6-anhydrous bridge by sulfatases utilize an uncommon mechanism for interacting with these sugar units and through extension with carrageenan active GHs. The three main motifs observed with monosaccharide units are; when all hydroxyl groups are equatorial which positions the *C-H* hydrogens axial allowing for the typical CH- π “stacking” interactions with aromatic side chains (indole groups from tryptophan or benzene groups from tyrosine or phenylalanine). These hydrogen positions allow for the aromatic side chain to interact with either top face hydrogens on C2, C4, and C6, or the bottom face hydrogens on C1, C3, and C5 (Figure 48A). In cases where hydroxyl groups occupy axial positions, the aromatic side chains can interact with hydrogen patches created by hydrogens on C3, C4, C5, and C6 (Figure 48B) or on carbons C1, C2, and C3 (Figure 48C). In the case of 3,6-anhydro-D-galactose units these commonly observed CH- π interactions are inhibited due to the steric hindrance of the 3,6-anhydrous bridge and the axial hydroxyl group of C2. Because of these features enzymes have adopted different means of coordinating these monosaccharide residues. The aromatic residue side chains of these enzymes appear to predominantly interact with the hydrogen patch composed of

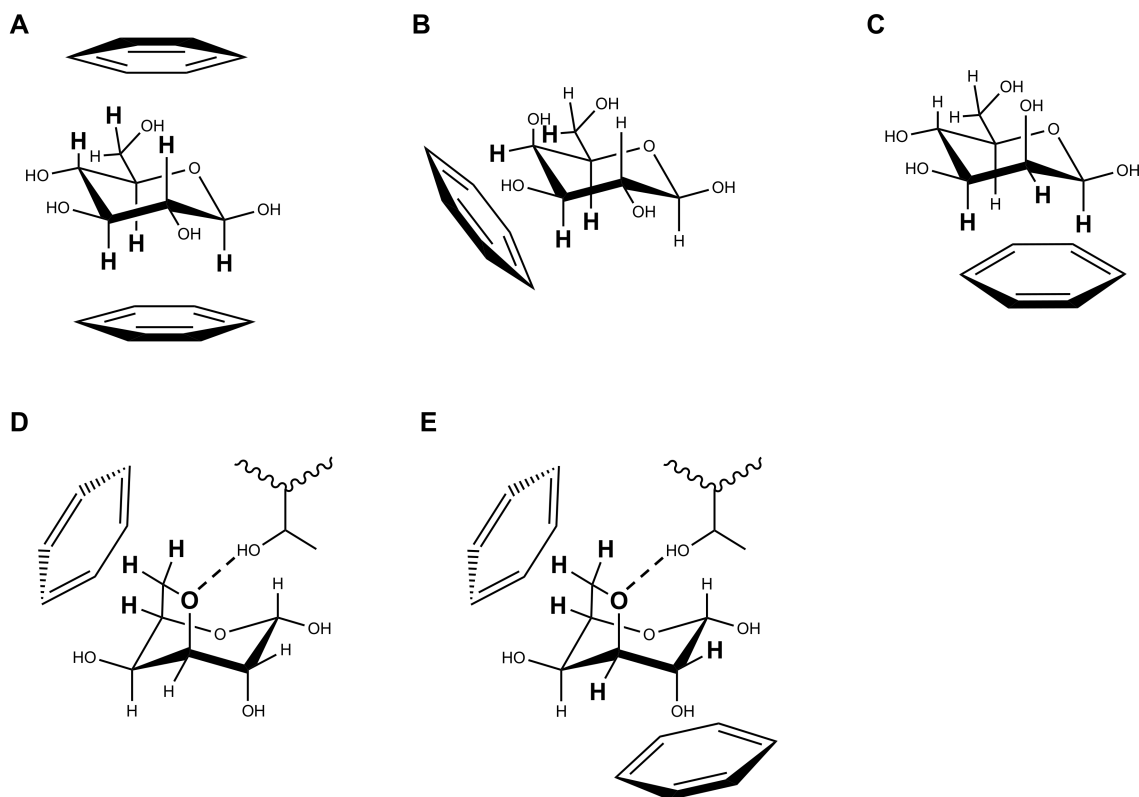


Figure 48. Geometries of carbohydrate-protein CH- π interactions (adapted from Spiwok et al. (2017)).

(A) β -D-glucopyranose can interact via both top and bottom faces, either hydrogens on carbons C1, C3, and C5 or on carbons C2, C4, and C6.

(B) α - or β -galactose interacting via hydrogens on carbons C3, C4, C5, and C6.

(C) β -D-mannose interacts via hydrogens on carbons C1, C2, and C3.

(D) α - or β -3,6-anhydro-D-galactose interacts via hydrogens on carbons C5 and C6.

(E) α - or β -3,6-anhydro-D-galactose interacts with hydrogens as in panel (D) in conjunction with a second aromatic system interacting with hydrogens on carbons C2 and C3.

(F) α -3,6-anhydro-L-galactose interacts via hydrogens on carbons C1 and C6.

The oxygen of the anhydrous bridge is often participating in hydrogen bonds, represented as dashed lines, with polar or charged amino acid side chains. Threonine is depicted here, however, arginine or aspartic acid, are also observed. In all panels, hydrogens participating in CH/ π interactions shown in boldface lettering, aromatic systems (indole of tryptophan or benzene rings of tyrosine and phenylalanine) are shown as benzene rings.

hydrogens on C5 and C6 (Figure 48D) and sometimes in conjunction with the coordination of the equatorial hydrogens of C2 and C3 (Figure 48E). The coordination of these hydrogens are often accompanied by one or more polar residues, such as threonine, glutamate, or arginine hydrogen bonding with the oxygen of the 3,6-anhydrous bridge (Figure 48D and 48E). These interactions involved in the recognition of 3,6-anhydro-D-galactose are quite unique and for the PsS1_19A, PsS1_19B, and PsS1_NC recognition aid in driving catalysis.

This research not only furthers the understanding of how carrageenan is depolymerized so that the photosynthetically fixed carbon can be returned to the global carbon cycle, but it also provides a much needed and deeper understanding of how carbohydrate active sulfatases bind and interact with their biological substrates. These findings strongly suggest that it is the recognition of the leaving group that drives S1 sulfatase specificity and the sulfatase subsite nomenclature proposed will allow for precision and ease when communicating the structural findings of sulfatases in complex with their cognate substrates.

References

- Abbott, D.W., and Boraston, A.B. (2008). Structural Biology of Pectin Degradation by *Enterobacteriaceae*. *Microbiol. Mol. Biol. Rev.* 72, 301–316.
- Aleshin, A.E., Firsov, L.M., and Honzatko, R.B. (1994). Refined structure for the complex of acarbose with glucoamylase from *Aspergillus awamori* var. X100 to 2.4-angstrom resolution. *J. Biol. Chem.* 269, 15631–15639.
- Arnott, S., Scott, W.E., Rees, D.A., and McNab, C.G.A. (1974). α -Carrageenan: Molecular structure and packing of polysaccharide double helices in oriented fibres of divalent cation salts. *J. Mol. Biol.* 90, 263–267.
- Azam, F., and Malfatti, F. (2007). Microbial structuring of marine ecosystems. *Nat. Rev. Microbiol.* 5, 782–791.
- Barbeyron, T., Brillet-Guéguen, L., Carré, W., Carrière, C., Caron, C., Czjzek, M., Hoebeke, M., and Michel, G. (2016). Matching the Diversity of Sulfated Biomolecules: Creation of a Classification Database for Sulfatases Reflecting Their Substrate Specificity. *PLoS One* 11.
- Berteau, O., Guillot, A., Benjdia, A., and Rabot, S. (2006). A new type of bacterial sulfatase reveals a novel maturation pathway in prokaryotes. *J. Biol. Chem.* 281, 22464–22470.
- Bielicki, J., and Hopwood, J.J. (1991). Human liver N-acetylgalactosamine 6-sulphatase Purification and characterization. *Biochem. J* 279, 515–520.
- Boltes, I., Czapinska, H., Kahnert, A., von Bülow, R., Dierks, T., Schmidt, B., von Figura, K., Kertesz, M.A., and Usón, I. (2001). 1.3 Å structure of arylsulfatase from *Pseudomonas aeruginosa* establishes the catalytic mechanism of sulfate ester cleavage in the sulfatase family. *Structure* 9, 483–491.
- Bond, C.S., Clements, P.R., Ashby, S.J., Collyer, C.A., Harrop, S.J., Hopwood, J.J., and Guss, J.M. (1997). Structure of a human lysosomal sulfatase. *Structure* 5, 277–289.
- Brünger, A.T. (1992). Free R-Value - A Novel Statistical Quantity for Assessing the Accuracy of Crystal-Structures. *Nature* 355, 472–475.
- Brüx, C., Ben-David, A., Shallom-Shezifi, D., Leon, M., Niefind, K., Shoham, G., Shoham, Y., and Schomburg, D. (2006). The structure of an inverting GH43 β -xylosidase from *Geobacillus stearothermophilus* with its substrate reveals the role of the three catalytic residues. *J. Mol. Biol.* 359, 97–109.
- von Bülow, R., Schmidt, B., Dierks, T., von Figura, K., and Usón, I. (2001). Crystal structure of an enzyme-substrate complex provides insight into the interaction between human arylsulfatase A and its substrates during catalysis. *J. Mol. Biol.* 305, 269–277.
- Carlson, B.L., Ballister, E.R., Skordalakes, E., King, D.S., Breidenbach, M.A., Gilmore, S.A., Berger, J.M., and Bertozzi, C.R. (2008). Function and Structure of a Prokaryotic Formylglycine-generating Enzyme. *J. Biol. Chem.* 283, 20117–20125.
- Carrico, I.S., Carlson, B.L., and Bertozzi, C.R. (2007). Introducing genetically encoded aldehydes into proteins. *Nat. Chem. Biol.* 3, 321–322.
- Cartmell, A., Lowe, E.C., Baslé, A., Firbank, S.J., Ndeh, D.A., Murray, H., Terrapon, N., Lombard, V., Henrissat, B., Turnbull, J.E., et al. (2017). How members of the human gut microbiota overcome the sulfation problem posed by glycosaminoglycans. *Proc. Natl. Acad. Sci. U. S. A.* 114, 7037–7042.

- Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 62, 1002–1011.
- Craigie, J.S., and Jurgens, A. (1989). Structure of agars from *Gracilaria tikvahiae* rhodophyta: Location of 4-O-methyl-I-galactose and sulphate. *Carbohydr. Polym.* 11, 265–278.
- Davies, G., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure* 3, 853–859.
- Davies, G.J., Wilson, K.S., and Henrissat, B. (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem. J.* 321 (Pt 2, 557–559.
- Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., Snoeyink, J., Richardson, J.S., et al. (2007). MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35, 375–383.
- Davison, J., Brunel, F., Phanopoulos, A., Prozzi, D., and Terpstra, P. (1992). Cloning and sequencing of *Pseudomonas* genes determining sodium dodecyl sulfate biodegradation. *Gene* 114, 19–24.
- Dawes, C.J., Lawrence, J.M., Cheney, D.P., and Mathiesow, A.C. (1973). Ecological studies of Floridian *Eucheuma* (Rhodophyta, Gigartinales). III. Seasonal variation of carrageenan, total carbohydrate, protein, and lipid. *Bull. Mar. Sci.*
- Demydchuk, M., Hill, C.H., Zhou, A., Bunkóczi, G., Stein, P.E., Marchesan, D., Deane, J.E., and Read, R.J. (2017). Insights into Hunter syndrome from the structure of iduronate-2-sulfatase. *Nat. Commun.* 8.
- Dhoot, G.K., Gustafsson, M.K., Ai, X., Sun, W., David, M., Jr, C.P.E., Sun, W., Standiford, D.M., and Emerson, C.P. Jr., (2017). Regulation of Wnt Signaling and Embryo Patterning by an Extracellular Sulfatase. *Science* (80-.). 293, 1663–1666.
- Dierks, T., Lecca, M.R., Schlotterhose, P., Schmidt, B., and von Figura, K. (1999). Sequence determinants directing conversion of cysteine to formylglycine in eukaryotic sulfatases. *EMBO J.* 18, 2084–2091.
- Dierks, T., Schmidt, B., Borissenko, L.V., Peng, J., Preusser, A., Mariappan, M., and von Figura, K. (2003). Multiple sulfatase deficiency is caused by mutations in the gene encoding the human Cα-formylglycine generating enzyme. *Cell* 113, 435–444.
- Dierks, T., Dickmanns, A., Preusser-Kunze, A., Schmidt, B., Mariappan, M., von Figura, K., Ficner, R., and Rudolph, M.G. (2005). Molecular basis for multiple sulfatase deficiency and mechanism for formylglycine generation of the human formylglycine-generating enzyme. *Cell* 121, 541–552.
- Dodgson, K.S., White, G.F., and Fitzgerald, J.W. (1982). Mechanism of enzymic and nonenzymic hydrolysis of sulfate esters. *Sulfatases of Microbial Origin, Volume 2* (CRC Press, Boca Raton, USA), pp. 43-123.
- Dodson, E.J., Winn, M., and Ralph, A. (1997). Collaborative computational project number 4: providing programs for protein crystallography. *Methods in Enzymology, Volume 277* (Elsevier, New York, USA), pp. 620-633.
- Domozych, D.S., College, S., Springs, S., and York, N. (2011). Algal Cell Walls.
- Emsley, P., and Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 60, 2126–2132.
- Falkowski, P.G. (1998). Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Science* (80-.). 281, 200–206.

- Ficko-Blean, E., Préchoux, A., Thomas, F., Rochat, T., Larocque, R., Zhu, Y., Stam, M., Génicot, S., Jam, M., Calteau, A., et al. (2017). Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria. *Nat. Commun.* 8.
- Field, C.B. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281, 237–240.
- Freile-Pelegrín, Y., and Robledo, D. (2008). Carrageenan of *Eucheuma isiforme* (Solieriaceae, Rhodophyta) from Yucatan, Mexico. II. Seasonal variations in carrageenan and biochemical characteristics. *J. Appl. Phycol.* 20, 537–541.
- Genicot, S.M., Groisillier, A., Rogniaux, H., Meslet-Cladière, L., Barbeyron, T., and Helbert, W. (2014). Discovery of a novel iota carrageenan sulfatase isolated from the marine bacterium *Pseudoalteromonas carrageenovora*. *Front. Chem.* 2, 67.
- Goldenberg, O., Erez, E., Nimrod, G., and Ben-Tal, N. (2009). The ConSurf-DB: Pre-calculated evolutionary conservation profiles of protein structures. *Nucleic Acids Res.* 37, 323–327.
- Gontang, E.A., Fenical, W., and Jensen, P.R. (2007). Phylogenetic diversity of gram-positive bacteria cultured from marine sediments. *Appl. Environ. Microbiol.* 73, 3272–3282.
- Guibet, M., Colin, S., Barbeyron, T., Genicot, S., Kloareg, B., Michel, G., and Helbert, W. (2007). Degradation of λ -carrageenan by *Pseudoalteromonas carrageenovora* λ -carrageenase: a new family of glycoside hydrolases unrelated to κ - and ι -carrageenases. *Biochem. J.* 404, 105–114.
- Hagelueken, G., Adams, T.M., Wiehlmann, L., Widow, U., Kolmar, H., Tu, B., Heinz, D.W., and Schubert, W. (2006). The crystal structure of SdsA1, an alkylsulfatase from *Pseudomonas aeruginosa*, defines a third class of sulfatases. 103, 7631–7636.
- Hanson, S.R., Best, M.D., and Wong, C.-H. (2004). Sulfatases: structure, mechanism, biological activity, inhibition, and synthetic utility. *Angew. Chem. Int. Ed. Engl.* 43, 5736–5763.
- Hedges, J.I., Baldock, J.A., Gélinas, Y., Lee, C., Peterson, M., and Wakeham, S.G. (2001). Evidence for non-selective preservation of organic matter in sinking marine particles. *Nature* 409, 801–804.
- Hehemann, J.-H., Smyth, L., Yadav, A., Vocadlo, D.J., and Boraston, A.B. (2012). Analysis of keystone enzyme in agar hydrolysis provides insight into the degradation of a polysaccharide from red seaweeds. *J. Biol. Chem.* 287, 13985–13995.
- Hehemann, J.-H., Boraston, A.B., and Czjzek, M. (2014). A sweet new wave: structures and mechanisms of enzymes that digest polysaccharides from marine algae. *Curr. Opin. Struct. Biol.* 28C, 77–86.
- Hernandez-Guzman, F.G., Higashiyama, T., Pangborn, W., Osawa, Y., and Ghosh, D. (2003). Structure of human estrone sulfatase suggests functional roles of membrane association. *J. Biol. Chem.* 278, 22989–22997.
- Hettle, A.G., Vickers, C., Robb, C.S., Liu, F., Withers, S.G., Hehemann, J.-H., and Boraston, A.B. (2018). The Molecular Basis of Polysaccharide Sulfatase Activity and a Nomenclature for Catalytic Subsites in this Class of Enzyme. *Structure* 26, 1-12
- Hidaka, M., Fushinobu, S., Ohtsu, N., Motoshima, H., Matsuzawa, H., Shoun, H., and Wakagi, T. (2002). Trimeric crystal structure of the glycoside hydrolase family 42 β -galactosidase from *Thermus thermophilus* A4 and the structure of its complex with galactose. *J. Mol. Biol.* 322, 79–91.

- Janes, L.E., Löwendahl, A.C., and Kazlauskas, R.J. (1998). Quantitative Screening of Hydrolase Libraries Using pH Indicators: Identifying Active and Enantioselective Hydrolases. *Chem. - A Eur. J.* 4, 2324–2331.
- Karamanos, Y., Ondarza, M., Bellanger, F., Christiaen, D., and Moreau, S. (1989). The linkage of 4-O-methyl-L-galactopyranose in the agar polymers from *Gracilaria verrucosa*. *Carbohydr Res.* 187, 93–101.
- Kelley, L., and Sternberg, M.J.E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* 4, 363–371.
- Knaust, A., Schmidt, B., Dierks, T., von Bülow, R., and von Figura, K. (1998). Residues critical for formylglycine formation and/or catalytic activity of arylsulfatase A. *Biochemistry* 37, 13941–13946.
- Knutsen, S.H., Myslabodski, D.E., Larsen, B., and Usov, A.I. (1994). A modified system of nomenclature for red algal galactans. *Bot. Mar.* 37, 163–170.
- Koshland, D.E.J. (1953). Stereochemistry and the mechanism of enzymatic reactions. *Boil. Rev. Cambridge Philos. Soc.* 28, 416–436.
- Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774–797.
- Laine, R.A. (1994). A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05×10 structures for a reducing hexasaccharide: The Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology* 4, 759–767.
- 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, D490-5.
- van Loo, B., Schober, M., Valkov, E., Heberlein, M., Bornberg-Bauer, E., Faber, K., Hyvönen, M., and Hollfelder, F. (2018). Structural and mechanistic analysis of the choline sulfatase from *Sinorhizobium melliloti*: a class I sulfatase specific for an alkyl sulfate ester. *J. Mol. Biol.* 430, 1004–1023.
- Lukatela, G., Krauss, N., Theis, K., Selmer, T., Gieseimann, V., von Figura, K., and Saenger, W. (1998). Crystal structure of human arylsulfatase A: The aldehyde function and the metal ion at the active site suggest a novel mechanism for sulfate ester hydrolysis. *Biochemistry* 37, 3654–3664.
- Maksimainen, M., Paavilainen, S., Hakulinen, N., and Rouvinen, J. (2012). Structural analysis, enzymatic characterization, and catalytic mechanisms of β -galactosidase from *Bacillus circulans* sp. *alkalophilus*. *FEBS J.* 279, 1788–1798.
- Martin, M., Portetelle, D., Michel, G., and Vandenbol, M. (2014). Microorganisms living on macroalgae: diversity, interactions, and biotechnological applications. *Appl. Microbiol. Biotechnol.* 98, 2917–2935.
- Matard-Mann, M., Bernard, T., Leroux, C., Barbeyron, T., Larocque, R., Préchoux, A., Jeudy, A., Jam, M., Nyvall-Collén, P., Michel, G., et al. (2017). Structural insights into marine carbohydrate degradation by family GH16 κ -carrageenases. *J. Biol. Chem.* 292, DOI: 10.1074
- McCarter, J.D., and Withers, S.G. (1994). Mechanisms of enzymatic glycoside hydrolysis. *Curr. Opin. Struct. Biol.* 4, 885–892.
- 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.

- McLean, M.W., and Williamson, F.B. (1979). Glycosulphatase from *Pseudomonas carrageenovora*. Eur. J. Biochem. 101, 497-505.
- Michel, G., Chantalat, L., Duee, E., Barbeyron, T., Henrissat, B., Kloareg, B., and Dideberg, O. (2001). The κ -carrageenase of *P. carrageenovora* features a tunnel-shaped active site: a novel insight in the evolution of Clan-B glycoside hydrolases. Structure 9, 513–525.
- Michel, G., Chantalat, L., Fanchon, E., Henrissat, B., Kloareg, B., and Dideberg, O. (2001). The ι -carrageenase of *Alteromonas fortis*. A beta-helix fold-containing enzyme for the degradation of a highly polyanionic polysaccharide. J. Biol. Chem. 276, 40202–40209.
- Michel, G., Helbert, W., Kahn, R., Dideberg, O., and Kloareg, B. (2003). The structural bases of the processive degradation of ι -carrageenan, a main cell wall polysaccharide of red algae. J. Mol. Biol. 334, 421–433.
- Morris, R.J. (2002). ARP/wARP's model-building algorithms. I. The main chain. Acta Crystallogr. Sect. D Biol. Crystallogr. 58, 968–975.
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. Sect. D Biol. Crystallogr. 53, 240–255.
- Nicholls, R.G., and Roy, A.B. (1971). 2 Arylsulfatases. In The Enzymes, Volume 5, P.D. Boyer, ed. (Academic Press), pp. 21–41.
- Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326.
- Painter, T.J. (1983). 4: Algal Polyssacharides. In The Polysaccharides, pp. 195–285.
- Paul, A., and Jonathan, J. (1998). Bacterial growth efficiency in natural aquatic systems. Annu. Re. Ecol. Syst. 29, 504-541
- Peng, J., Alam, S., Radhakrishnan, K., Mariappan, M., Rudolph, M.G., May, C., Dierks, T., von Figura, K., and Schmidt, B. (2015). Eukaryotic formylglycine-generating enzyme catalyses a monooxygenase type of reaction. FEBS J. 282, 3262–3274.
- Pluinage, B., Hehemann, J.H., and Boraston, A.B. (2013). Substrate recognition and hydrolysis by a family 50 exo- β -agarase, aga50D, from the marine bacterium *Saccharophagus degradans*. J. Biol. Chem. 288, 28078–28088.
- Popper, Z.A., Michel, G., Hervé, C., Domozych, D.S., Willats, W.G.T., Tuohy, M.G., Kloareg, B., and Stengel, D.B. (2011). Evolution and Diversity of Plant Cell Walls: From Algae to Flowering Plants. Annu. Rev. Plant Biol. 62, 567–590.
- Popper, Z.A., Ralet, M.C., and Domozych, D.S. (2014). Plant and algal cell walls: Diversity and functionality. Ann. Bot. 114, 1043–1048.
- POTIN, P., SANSEAU, A., LE GALL, Y., ROCHAS, C., and KLOAREG, B. (1991). Purification and characterization of a new κ -carrageenase from a marine Cytophaga-like bacterium. Eur. J. Biochem. 201, 241–247.
- Powell, H.R. (1999). The Rossmann Fourier autoindexing algorithm in MOSFLM. Acta Crystallogr. Sect. D Biol. Crystallogr. 55, 1690–1695.
- Préchoux, A., Genicot, S., Rogniaux, H., and Helbert, W. (2013). Controlling carrageenan structure using a novel formylglycine-dependent sulfatase, an endo-4S- ι -carrageenan sulfatase. Mar. Biotechnol. (NY). 15, 265–274.
- Préchoux, A., Genicot, S., Rogniaux, H., and Helbert, W. (2016). Enzyme-Assisted Preparation of Furcellaran-Like κ - β -Carrageenan. Mar. Biotechnol. 18, 133–143.

- Project, C.C. (1994). The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50**, 760–763.
- Prost, O., Turrel, M.O., Dahan, N., Craveur, C., and Adessi, G.L. (1984). Estrone and Dehydroepiandrosterone Sulfatase Activities and Plasma Estrone Sulfate Levels in Human Breast Carcinoma. *Cancer Res.* **44**, 661–664.
- Recksiek, M., Selmer, T., Dierks, T., Schmidt, B., and von Figura, K. (1998). Sulfatases, Trapping of the Sulfated Enzyme Intermediate by Substituting the Active Site Formylglycine. *J. Biol. Chem.* **273**, 6096–6103.
- Rivera-Colón, Y., Schutsky, E.K., Kita, A.Z., and Garman, S.C. (2012). The structure of human GALNS reveals the molecular basis for mucopolysaccharidosis IV A. *J. Mol. Biol.* **423**, 736–751.
- Rivkin, R.B., and Legendre, L. (2001). Biogenic Carbon Cycling in the Upper Ocean : Effects of Microbial Respiration. **291**, 2398–2400.
- Robb, M., Hobbs, J.K., and Boraston, A.B. (2017). Separation and Visualization of Glycans by Fluorophore-Assisted Carbohydrate Electrophoresis. *Methods Mol. Biol.* **1588**, 215–221.
- Schneider, T.R., and Sheldrick, G.M. (2002). Substructure solution with SHELXD. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58**, 1772–1779.
- Sidhu, N.S., Schreiber, K., Pröpper, K., Becker, S., Usón, I., Sheldrick, G.M., Gärtner, J., Krätzner, R., and Steinfeld, R. (2014). Structure of sulfamidase provides insight into the molecular pathology of mucopolysaccharidosis IIIA. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70**, 1321–1335.
- Stortz, C.A., and Cerezo, A.S. (1992). The ¹³C NMR spectroscopy of carrageenans: calculation of chemical shifts and computer-aided structural determination. *Carbohydr. Polym.* **18**, 237–242.
- Therkildsen, M.S., Isaksen, M.F., and Lomstein, B.A. (1997). Urea production by the marine bacteria *Delacya venusta* and *Pseudomonas stutzeri* grown in a minimal medium. *Aquat. Microb. Ecol.* **13**, 213–217.
- Usov, A. (2011). Polysaccharides of the red algae. *Adv. Carbohydr. Chem. Biochem.* **116**.
- Viborg, A.H., Katayama, T., Arakawa, X.T., Hachem, X.M.A., Leggio, X.L. Lo, Kitaoka, X.M., Svensson, B., and Fushinobu, X.S. (2017). Discovery of α -L-Arabinopyranosidases from human gut microbiome expands the diversity within glycoside hydrolase family 42. *J. Biol. Chem.* **292**, 21092–21101.
- Wu, M., Bu, L., Vuong, T. V., Wilson, D.B., Crowley, M.F., Sandgren, M., Ståhlberg, J., Beckham, G.T., and Hansson, H. (2013). Loop motions important to product expulsion in the *Thermobifida fusca* glycoside hydrolase family 6 cellobiohydrolase from structural and computational studies. *J. Biol. Chem.* **288**, 33107–33117.
- Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012). DbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **40**, 445–451.