Enzyme and Microbial Technology 182 (2025) 110528

Contents lists available at ScienceDirect

Enzyme and Microbial Technology

journal homepage: www.elsevier.com/locate/enzmictec

Targeted metagenomics – Enrichment for enzymes active on sulfated polysaccharides from seaweeds

Bjorn Thor Adalsteinsson [a,b,*], Hörður Guðmundsson [a], Andrius Jasilionis [c], Morten Schiøtt [d], Maria Dalgaard Mikkelsen [d], Elísabet Eik Guðmundsdóttir [a], Pavithra Sivakumar [c], Annika Malmgren [c], Tushar Kaushik [c], Erik Apelqvist [c], Signe Vangsgaard [d], Rébecca Leblay [a], Ólafur Friðjónsson [a], Anne S. Meyer [d], Eva Nordberg Karlsson [c], Guðmundur Óli Hreggviðsson [a,b]

[a] Matis ohf, Vínlandsleið 12, Reykjavík 113, Iceland
[b] University of Iceland, Sæmundargata 2, Reykjavík 102, Iceland
[c] Division of Biotechnology, Department of Chemistry, Lund University, PO Box 124, Lund SE-221 00, Sweden
[d] Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads, Building 221, Kongens Lyngby DK-2800, Denmark

ARTICLE INFO

Keywords:
Ulvan lyase
Fucoidanase
Seaweed
Polysaccharide
Thermophile

ABSTRACT

Seaweeds (macroalgae) are an attractive resource for diverse microbial- and enzymatic production processes. They are abundant, underutilized, cheap, and rich in carbohydrates, and therefore have the potential to be used as a source of mono- or oligosaccharides, and as substrates for industrial fermentation processes. Many seaweed polysaccharides, including the sulfated polysaccharides ulvan and fucoidan, are however complex and heterogenous in structure, and there are currently few enzymes available to modify them, and understanding of their enzymatic depolymerization remains limited. The present study aimed to identify and characterize robust fucoidanases and ulvan lyases. Metagenomes were obtained from microbial enrichments from an intertidal hot-spring, genes identified that expressed putative fucoidanases and ulvan lyases, and following gene cloning and expression, the respective enzymes were screened for enzymatic activity. Consistent with their origin, the identified protein sequences were considerably divergent from previously characterized enzymes, with a 44 % average maximal sequence identity. In total, the study resulted in the characterization of 10 new fucoidanases (GH107 and GH168 families) and 8 new ulvan lyases (PL24, PL25 and PL40 families). Notably, the new fucoidanases appeared to have functional specificity towards fucoidan containing α-1,3 linked L-fucosyl and several functioned at high temperature. The study contributes a metagenomics-based approach to identify new seaweed polysaccharide degrading enzymes and an increased understanding of the diversity of such enzymes, which may have implications for the realization of biotechnology based valorization of seaweed biomass.

1. Introduction

Macroalgae, i.e., seaweeds, are photosynthetic, aquatic eukaryotic organisms that are commonly classified in three groups, as brown (Phaeophyceae), red (Rhodophyta) and green (Chlorophyta) seaweeds. Among other features, the groups are distinguished by the production of characteristic pigments used for photosynthesis, and by possessing specific polysaccharides in their cell walls and storage organs that differ profoundly from those in terrestrial plants. Well known polysaccharides in brown seaweeds are laminarin, which is the primary carbohydrate storage substance, and alginate and fucoidan, which are unique structural cell wall polysaccharides, in red seaweeds they are agar and carrageenan, with some containing unique xylan structures, and in green seaweeds the most well-known polysaccharide is ulvan [1,2].

Carbohydrate content in seaweeds is generally high, often constituting well over half their dry-weight [3]. Some seaweed polysaccharides pose challenges, as well as opportunities, as industrial feedstock because of heterogeneous sugar composition, sulfation and other modifications. Besides sugars known from lignocellulosics, e.g. hexoses (glucose, mannose and galactose), and pentoses (xylose and arabinose), appreciable amounts of "rare sugars" (sugar alcohols, deoxy sugars and sugar acids (uronates)) are present in macroalgae [4].

* Corresponding author at: Matis ohf, Vínlandsleið 12, Reykjavík 113, Iceland.
E-mail address: bjornth@matis.is (B.T. Adalsteinsson).

https://doi.org/10.1016/j.enzmictec.2024.110528
Received 2 September 2024; Received in revised form 14 October 2024; Accepted 14 October 2024
Available online 16 October 2024

Macroalgal polysaccharides such as agar, carrageenan, and alginate are well known hydrocolloids that have been used in biotechnology, pharma, and in the food industry for decades [5], whereas fucoidan and ulvan are relatively new to industry but may have very high-value applications as biomedical or nutri-functional products [6].

Fucoidan and ulvan are sulfated polysaccharides that have heterogenous chemical composition and complex structure. Ulvan is composed of a backbone of repeating disaccharide units that vary in frequency between species. The most common disaccharides consist of β-D-glucuronic acid (1,4)-linked to α-L rhamnose-3-sulfate; α-L-iduronic acid (1,4)-linked to α-L-rhamnose 3-sulfate; and α-L-rhamnose-3-sulfate linked to either β-D-xylose or β-D-xylose-2-sulfate [7]. Fucoidans are generally composed of a backbone of either 1,3-α-L-fucose or alternating 1,3-α and 1,4-α L-fucose that can be substituted with sulfate, acetate, and/or fucose or uronic acids – depending on seaweed species; a third type, known as galactofucans, have a backbone containing both L-fucosyl and D-galactosyl residues[8].

About 36 million tons of seaweeds are harvested globally per year, by vast majority from aquaculture in Asia [9], primarily for use as food or for extraction of hydrocolloids. The recent emergence of seaweed farming in the Northern hemisphere, including North America, France, and notably in the North Atlantic, is poised to drive a new wave of industrial development of high-value products and applications based on tailormade biorefining of seaweeds, with enzymatic bioprocessing playing a main role [6]. For example, seaweed carbohydrates can in principle be used as feedstock for industrial microbiology processes, to yield products such as organic acids, alcohols and ketones, which in turn can be used as biofuels, as building blocks for bioplastics, and other products [4]. Further, diverse products could be extracted from the seaweed biomass, including [10] proteins and omega-3 fatty acids that could be used in food and feed, phycobiliproteins for use as pigments, carotenoids for use as antioxidant, and mannitol for use as sweetener, but due to the high levels of unique polysaccharides in the seaweeds a main focus remains on producing oligosaccharides with unique bio-functional effects. The realization of these processes is dependent on, or would benefit from, the availability of diverse enzymes that break down or selectively modify the seaweed biomass polysaccharides.

The complexity of fucoidan and ulvan structures makes them challenging for enzymatic digestion. To meet the request of producing homogenous products, a diverse panel of enzymes is necessary to expand the toolbox of enzymes that break down or selectively modify seaweed polysaccharides.

In the Carbohydrate-Active enZYmes Database (CAZy), published carbohydrate active enzymes are categorized into families based on structural similarities – and hence enzymatic mechanisms [11]. In the database, fucoidanases are categorized in glycoside hydrolase (GH) families GH107 and GH168, as well as the recently established families GH174, GH187 and in polysaccharide lyase (PL) family PL43, and a total of 24 enzymes are listed as functionally characterized [11]. Ulvan lyases are distributed in polysaccharide lyase families PL24, PL25, PL28, PL37, and PL40, and a total of 15 enzymes are listed as functionally characterized [11].

The present study aimed to identify and characterize novel fucoidanases and ulvan lyases to expand the functional diversity within their respective enzyme families. Further, the study aimed to identify robust enzymes applicable for use in industrial settings to facilitate their use in advanced processes, such as in biorefineries, to aid in increased valorization possibilities of seaweeds. In that context, the study aimed to identify enzymes that function optimally in the presence of environmental stressors such as extreme heat or salinity, which are frequently encountered in industrial settings. To achieve this, a microbial enrichment for seaweed degrading microorganisms was performed on seaweed substrates in an intertidal hot-spring, the respective metagenomes were isolated and sequenced, and fucoidanases and ulvan lyases identified in the metagenomes were characterized.

2. Materials and methods

2.1. Preparation and processing of in situ enrichment

An enrichment substrate was prepared using mixed green and brown seaweeds sourced from the Icelandic coast (out of Reykjanes peninsula and in Miðfjörður fjord, respectively) and a carrageenan gel containing fucoidan. The gels were composed of 1 g carrageenan polysaccharide (C1013, Merck), 6 g KCl, and approximately 30 mg each of sugars L-fucose (F2252, Merck), fucoidan from *Fucus vesiculosus* (F8190, Merck), fucoidan from *Ascophyllum nodosum* (GF102, Glycomix) and fucoidan from *Laminaria* sp. (PSA13, Glycomix) that were dissolved together in 100 mL of boiling water and cast to form a gel. The enrichment substrates were placed in bags of finely meshed nylon and incubated in an intertidal hot-spring called Skarðshver, located in Miðfjörður in northern Iceland, for 5 weeks. After incubation, the enrichment substrates were placed in 5 L containers with water collected from the hot-spring, brought to the laboratory, and kept at 4 °C until processing. The remaining enrichment substrates, mainly seaweed debris, were processed with the aim of harvesting bacterial cells that were putatively enriched on the substrates. Seaweed substrates were placed in a few liters of sterile water and agitated by gentle shaking and stirring to release cells. Each sample was spun in a centrifuge at 500×*g* to pellet the 'spent' enrichment substrates. The supernatant was vacuum filtered through Whatman Grade 4 Qualitative Filter Paper (Cytiva) with frequent replacement of filter paper to prevent clogging. The flow-through was collected and spun in a centrifuge at 16,000×*g* to pellet the cells contained in the liquid. The supernatant was discarded. Harvested cell pellets were resuspended in 10–20 mL of sterile water.

2.2. DNA isolation, sequencing, and assembly

For selection of DNA extraction method, DNA was extracted separately from two *in situ* enriched cell pellets using four different kits: MasterPure Complete DNA and RNA Purification Kit (LGC Biosearch Technologies), MasterPure Gram Positive DNA Purification Kit (LGC Biosearch Technologies), DNeasy PowerSoil Kit (QIAGEN), and FastDNA SPIN Kit for Soil (MP Biomedicals). All kits were used according to the manufacturers´instructions with the exception that in the MasterPure kits, buffer TE was replaced with 10 mM Tris-HCl buffer, while for the FastDNA kit cell pellets were disintegrated using a vortex instead of recommended bead-based homogenizer. For each extraction, 2 mL of enrichment cell fractions were used, corresponding to approximately 0.16–0.29 g of cell pellets. DNA concentration and purity was determined with spectrophotometry and fluorometry using a Nanodrop ND-1000 instrument (Thermo Fisher Scientific) and a Qubit fluorometer with Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific).

For assessment of microbial diversity, the 16S rRNA gene was amplified for all DNA extraction samples using the primer pairs S-D-Bact-0341-b-S-17, 5′-CCTACGGGNGGCWGCAG-3′/S-D-Bact-0785-a-A-21, 5′-GACTACHVGGGTATCTAATCC-3 targeting the V3–V4 hypervariable regions of the 16S rRNA gene for bacteria and S-D-Arch-0519-a-S-15, 5′-CAGCMGCCGCGGTAA-3′/S-D-Arch-1041-a-A-18, 5′- GGCCATGCACCWCCTCTC-3′ targeting the V4-V6 hypervariable regions for archaea. Amplification was performed using Q5 High-Fidelity DNA Polymerase (New England Biolabs) and the following thermocycling conditions: 98 °C for 30 s; 35 cycles of 98 °C for 10 s, 52 °C for 30 s, 72 °C for 30 s; 72 °C for 2 min; hold at 10 °C. Amplicons were purified using HighPrep PCR Clean-up System (MagBio Genomics) and indexed with Nextera DNA CD indices (Illumina) in a 5 cycle amplification using Q5 High-Fidelity DNA Polymerase. The indexed amplicons were purified and normalized with SequalPrep Normalization Plate Kit (Thermo Fisher Scientific) and sequenced on the MiSeq System (Illumina) using the MiSeq Reagent Kit v3 (600-cycle) (Illumina). Data were analyzed with the DADA2 v1.26.0, phyloseq v1.42.0, dplyr v1.1.2, and ggplot2 v3.4.2 packages in RStudio (Posit) with SILVA v138.1 taxonomy

database. Alpha diversity was compared between DNA isolation methods by the Observed number of taxonomic groups and the Chao1, Shannon and Simpson indices of sample richness and evenness.

For metagenome sequencing, DNA was extracted using the DNeasy PowerSoil Kit and sequencing libraries made for all enrichment samples using the Nextera DNA Flex Library Preparation Kit (Illumina) and Nextera DNA CD indices (Illumina). The libraries were sequenced on the MiSeq system using MiSeq Reagent Kit v3 (600-cycle), resulting in 5–15 million raw reads for each sample and a total of 37.8 million raw reads.

For taxonomic characterization of metagenomes, the trimmed Illumina MiSeq reads were assigned taxonomy using Kraken 2 v2.1.3 and the Standard Database of RefSeq indices v20240112 and relative abundance calculated by Bracken v2.9 and phyloseq v1.44.0. Results were plotted in RStudio using ggplot2 v3.5.0.

For assembly of individual metagenomes, the sequencing reads were trimmed with Trimmomatic v0.39 [12] (using options HEADCROP:15, LEADING:3, TRAILING:3, SLIDINGWINDOW:5:20, MINLEN:100). About 50–60 % of the bases were of sufficient quality to pass the trimming step. Assembly was performed with SPAdes v3.14.0 in metagenome mode, and contigs shorter than 1000 bp omitted from further analysis (about 90 % of all contigs). Open reading frames (ORFs) in the remaining contigs were detected using Prodigal v2.6.2.

Since the different enrichment samples were obtained from an almost identical environment, their genomic content could be very similar. As a result, it was reasoned that it might be possible to obtain longer contigs by performing a pooled assembly using the sequencing data from all four enrichments, as compared to each individual enrichment. Raw reads were quality assessed using FastQC v0.11.7 [13] and trimmed using Trimmomatic v0.39, resulting in a total of 22.7 million reads and 6.1×10^9 bases for all samples. Trimmed and filtered reads for all libraries were assembled in one assembly using MEGAHIT v1.2.9 [14] producing 608,962 contigs, of which 515,626 contigs were shorter than 1000 bp and were excluded from further analysis. The remaining 93,336 contigs were polished with the trimmed Illumina MiSeq reads using Bowtie2 v2.4.2 and Pilon v1.23 and ORFs subsequently detected using Prodigal v2.6.2.

For improved assembly and gene identification, all metagenomic samples were sequenced by Oxford Nanopore long-read sequencing technology. Libraries were prepared using the same DNA as for Illumina sequencing and sequencing libraries prepared using the SQK-LSK109 kit (Oxford Nanopore Technologies) and barcodes EXP-NBD104 (Oxford Nanopore Technologies) according to the 'Native Barcoding Genomic DNA' protocol with the optional selection for longer fragments. The barcoded libraries were pooled and sequenced on an R9 flow cell in a MinION portable sequencing device (Oxford Nanopore Technologies). Base-calling and demultiplexing was performed using Guppy v3.5.2 (Oxford Nanopore Technologies) and hybrid assemblies of the Oxford Nanopore MinION reads, and trimmed Illumina MiSeq reads for each sample library generated using SPAdes v3.14.0 in metagenome mode. Contigs shorter than 1000 bp were excluded, the trimmed Illumina MiSeq reads mapped on to the remaining contigs by Bowtie2 v2.4.2, the draft metagenomes polished using Pilon v1.23 and ORFs identified using Prodigal v2.6.2.

2.3. Identification and analysis of CAZymes

ORFs identified in the four metagenomes assembled with Illumina Miseq data were analyzed for the presence of CAZy motifs using dbCAN2 server [15]. The analysis was performed in November 2020 using HMMER, DIAMOND and Hotpep tools. All sequences identified with one or more tool as containing GH107, GH168, PL24, PL25, PL28, PL37 or PL40 family motif – the genes of interest (GOI) - were collated and analyzed further. The respective protein sequences were compared to one another using blastp with default settings in Geneious R9.1.8 (Dotmatics) to identify homologs. A total of 14 sequences were $\geq$98 % identical to another protein and hence one sequence in the respective pair omitted from further analysis. When the pair included sequences of unequal length, the longer sequence was retained. The sequence identity of the translated GOI were further compared with sequences deposited in the CAZy database [11] using blastp in Geneious with default settings. For the analysis, protein sequences from the CAZy database were downloaded from the web-server hosting dbCAN2 (filename CAZyDB.07312020.fa [15]). The translated GOI were compared against the NCBI nr database in November 2020 with blastp implemented in NCBI's BLAST suite [16] with default parameters. Phylogenetic trees were constructed using the Clustal Omega multiple sequence alignment program [17] with default parameters. Data were extracted from the Clustal Omega output and trees constructed with Geneious Prime (Dotmatics). Reference sequences used to construct the trees were obtained from the CAZy database, downloaded from the web-server hosting dbCAN2. Boxplots of percent identity of translated GOI compared with different databases were constructed using R v3.4.0 with data obtained from the alignment analyses. For the analysis of 'protein domain architecture' InterProScan was used [18]. The domains were detected using the 'Find Protein domains with interproscan' function in Geneious Prime. Where GH107, GH168, PL24, PL25, PL28, PL37 or PL40 homology is indicated, the boundaries of the homology were determined through HMMER or DIAMOND analysis on the dbCAN3 web-server [19], or through analysis using CUPP [20].

2.4. Cloning and primary determination of production yield

GOI were amplified from each respective metagenome DNA using PCR with primers (Supplementary file 1) that included overhangs compatible for DNA assembly into vector pHWG1106. Protein sequences encoded by the GOI were analyzed for the presence of signal peptides and secretion signals using InterProScan sequence analysis [18] and the respective primers were designed to omit these sequences. pHWG1106 is an in-house expression vector containing high-copy-number ColE1/pMB1/pBR322/pUC origin of replication; a gene encoding β-lactamase for ampicillin resistance; and a construct, under regulation of a rhamnose-inducible promoter, enabling cloning of a GOI to generate fusion proteins with I) MalE-Smt3 on the N-terminus, II) hexa histidine-tag (His-tag) on the C-terminus, III) MalE-Smt3 and His-tag fusion, or IV) with no fusion. MalE is maltose/maltodextrin-binding periplasmic protein (also referred to as maltose binding protein, MBP) and Smt3 is *Saccharomyces cerevisiae* ubiquitin-like protein Smt3 motif. MalE facilitates protein folding and affinity purification while Smt3 enables proteolytic removal of the MalE fusion with Ulp1 protease cleavage [21]. When a C-terminal His-tag was desired without N-terminal fusion, the vector was digested with *Nde*I and *Bam*HI restriction enzymes, and when an N-terminal MalE-Smt3 fusion and C-terminal His-tag was desired, the vector was digested with *Kas*I and *Bam*HI restriction enzymes. After restriction, the vector was agarose gel purified (Monarch DNA Gel Extraction Kit, New England Biolabs) and bead purified using HighPrep PCR Clean-up System according to manufacturer's instructions. PCR amplicons were also bead purified. DNA assembly was performed with NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs) according to the manufacturer's instructions and *Escherichia coli* NEB10-beta competent cells (New England Biolabs) transformed with the assembly products. The cells were plated on LB medium (Miller: 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) containing 100 μg/mL ampicillin and grown at 37 °C overnight. A few clones were cultured overnight in liquid LB medium containing 100 μg/mL ampicillin, and plasmids isolated with Monarch Plasmid Miniprep Kit (New England Biolabs). The fidelity of the cloning was estimated by restriction analysis of the plasmids with *Eco*RI-HF and *Eco*RV-HF restriction enzymes (New England Biolabs) followed by agarose gel electrophoresis.

Primary determination of production yield was performed at 30 mL expression scale. *E. coli* NEB10-beta clones transformed with expression plasmids were cultivated overnight at 37 °C, 200 RPM, in LB-Miller

media supplemented with 100 μg/mL ampicillin. On the following day, 200 μL of the culture were inoculated in 30 mL of fresh media with ampicillin and incubated at 37 °C. When the culture's optical density at 600 nm (OD_{600}) reached 0.7 the incubation temperature was lowered to room temperature. When the culture's OD_{600} reached 0.9, L-rhamnose was added to a final concentration of 0.15 % (w/v) to induce GOI expression and the culture incubated at room temperature overnight. The cells were pelleted by centrifugation at 4000 RPM for 15 min and lysed by sonication after resuspension in a lysis buffer (50 mM sodium citrate pH 6.1, 200 mM NaCl, and 15 % (v/v) glycerol), using a volume of lysis buffer corresponding to four-fold wet weight of the pelleted cells. The lysate was spun in a centrifuge at 15,000 × g for 20 min at 4 °C and the soluble (clarified cell lysate) and resuspended insoluble fractions analyzed with SDS-PAGE on 4–20 % gradient Mini-PROTEAN TGX Stain-Free Precast gels (Bio-Rad). Bioinformatic prediction of soluble protein yield was performed using SoluProt v1.0 server [22]. Amino acid sequences of the enzymes produced by each of the successfully cloned constructs are listed in Supplementary file 2.

2.5. Primary activity screening

For putative fucoidanases, primary activity screening was performed by incubating clarified cell lysates of GOI expression clones with fucoidan substrates followed by carbohydrate polyacrylamide gel electrophoresis (C-PAGE). Four fucoidan polysaccharide substrates were tested from Phaeophyceae algae. They were from *Fucus vesiculosus* (FUC100, ELICITYL), *Ascophyllum* (YF09363, Glycomix), *Laminaria digitata* (YF09361, Glycomix), and NovaMatrix PROTASEA fucoidan from *Laminaria hyperborea* (DuPont Nutrition). For the reaction, the cell lysates were combined with fucoidan at a final concentration of 0.7 % (w/v) in a reaction buffer (25 mM sodium citrate or sodium phosphate depending on pH desired, 1 % w/v NaCl, and 1 mM $CaCl_2$). For each enzyme, the screening tests were performed at two to three pH values in the range of pH 5.7—7.5 in an overnight reaction at 30 °C. The reaction products were separated with C-PAGE and stained to visualize the hydrolysis products of fucoidan [23]. Where activity was detected, further reactions were performed at a wider range of reaction times, temperatures and pH levels.

For putative ulvan lyases, primary activity screening was performed by incubating clarified cell lysates of GOI expression clones with ulvan substrates, followed by thin-layer chromatography (TLC). A substrate containing a mixture of ulvan polysaccharides from the Chlorophyta algae *Ulva lacinulata* (formerly *Ulva armoricana*) (native grade) (ULV100, ELICITYL) and from *Ulva intestinalis* (native grade) (ULV101, ELICITYL) was used for the assays, at a final concentration of 0.5 % (w/v) each. The reactions were performed in a reaction buffer (25 mM sodium citrate or 25 mM sodium phosphate, depending on pH tested, 1 % w/v NaCl, and 1 mM $CaCl_2$). For each enzyme, the screening tests were performed at two to three pH values in the range of pH 5.7—7.5, at one to two temperatures in the range of 30—40 °C, and incubated overnight. The reaction mixtures were separated with TLC and stained to visualize the degradation products. Briefly, reaction products (2 μL) were applied on a TLC Silica gel 60 F_{254} plates (Merck) and resolved using a mobile phase solution composed of n-butanol, acetic acid and water at a 2:1:1 vol ratio, respectively. The developed plates were dried and products visualized using orcinol/sulfuric acid staining (100 mg orcinol dissolved in 5 mL sulfuric acid and 95 mL methanol), followed by heating at 125 °C for a few minutes. Where activity was detected, further reactions were performed at a wider range of reaction times, temperatures and pH levels, and reaction products detected with TLC or DNS (3,5-dinitrosalicylic acid) [24], using dilutions of freshly dissolved D-glucose as a standard and measuring absorbance at 540 nm (A_{540}) using a microplate reader.

2.6. Production upscaling of fucoidanases

For fucoidanases, recombinant enzyme production upscaling was performed at a 500 mL expression scale in LB-Lennox media (10 g/L tryptone, 5 g/L yeast extract, and 5 g/L NaCl) supplemented with 100 μg/mL of ampicillin and 34 μg/mL of chloramphenicol, in accordance with selection markers of the expression vectors. The expression media was inoculated with 2 % (v/v) overnight culture of *E. coli* BL21 (Thermo Fisher Scientific) containing the pGro7 plasmid (Takara Bio) encoding the GroES and GroEL chaperonins under regulation of an arabinose-inducible promoter and subsequently transformed with the GOI expression plasmids and cultivated at 37 °C. Chaperonin expression was induced with 0.05 % (w/v) of L-arabinose after inoculation. When culture OD_{600} reached 0.8—1.0, GOI expression was induced with 0.15 % (w/v) of L-rhamnose and the cultivation temperature was lowered to 20 °C and incubated overnight. The expression cultures were harvested by centrifugation and lysed by sonication after resuspension in extraction buffer (20 mM Tris-HCl pH 7.4, 20 mM imidazole, 250 mM NaCl supplemented with 0.2 mg/mL of lysozyme). Insoluble protein fraction was separated by centrifugation at 19,000×*g* for 15 min at 4 °C. The protein fractions obtained were analyzed with 12 % glycine-SDS-PAGE (Bio-Rad) stained with Bio-Safe Coomassie G250 Stain (Bio-Rad).

2.7. Purification of fucoidanases

Enzymes were purified from soluble protein fractions by applying nickel affinity chromatography with HisPur Ni-NTA resin (Thermo Fisher Scientific). One mL of resin was equilibrated with equilibration buffer (20 mM Tris-HCl pH 7.4, 250 mM NaCl) before adding soluble protein fraction after filtration. Protein was bound to affinity resin under slow agitation for 1 h in an ice bath, and subsequently the soluble fraction was separated by centrifugation at 2,000×*g* for 2 min at 4 °C and discarded. The affinity resin with bound proteins was then transferred to a gravity flow column after washing with wash buffer (20 mM Tris-HCl, pH 7.4, 20 mM imidazole, 250 mM NaCl). Proteins bound to affinity resin were eluted in stepwise manner with nickel affinity chromatography elution buffers (20 mM Tris-HCl, pH 7.4, 250 mM NaCl and 100, 150, 200, 250 or 300 mM imidazole). Elution fractions containing enzymes were pooled. Elution buffers were exchanged applying group separation with PD-10 desalting columns (Cytiva) to fucoidanase sample buffer (20 mM Tris-HCl pH 7.4, 250 mM NaCl). The purified proteins were stored in sample buffer at 4 °C after filtration. The enzyme purity and integrity were assessed with SDS-PAGE. Western blot analysis was also applied to assess enzyme purification. The proteins fractionated with 12 % glycine-SDS-PAGE (Bio-Rad) were transferred to the membrane of a Trans-Blot Turbo Mini 0.2 μm Nitrocellulose Transfer Pack (Bio-Rad) blotting in a Trans-Blot Turbo Transfer System (Bio-Rad) in accordance with the manufacturer's turbo protocol. The membrane was subsequently blocked in skim milk buffer (1X TBS, 2 % (w/v) skim milk) for 1 h at room temperature, next it was incubated in skim milk buffer containing anti-His-tag antibodies (Merck) at a 1:10,000 vol ratio for 1 h at room temperature, and finally washed three times in skim milk buffer for 15 min. The membrane was stained using an AEC Staining Kit (Merck) according to the manufacturer's protocol. Protein concentration was measured spectrophotometrically with the Bradford assay using dilutions of BSA as a standard.

2.8. Activity determination of purified fucoidanases

Activity of purified fucoidanases was tested using fucoidan substrates. In the tests, the reaction mixtures had a final volume of 20 μL and contained 0.9 % (w/v) fucoidan and 9 μg fucoidanase in a fucoidanases reaction buffer (9 mM Tris-HCl pH 7.4, 112.5 mM NaCl, and 10 mM $CaCl_2$). For all purified enzymes, an initial activity test was performed on *Ld* and *Sl* fucoidan (Table 1) by incubation overnight at 30 °C with mixing at 500 RPM. Reactions were inactivated by heating for 10 min at

Table 1
Overview of the fucoidans used for substrate specificity determination of fucoidanases.

Fucoidan origin	Abbreviation	Source
Alaria esculenta	*Ae*	YF157138, Glycomix
Ascophyllum nodosum	*An*	YF09363, Glycomix
Ecklonia radiata	*Er*	YF157166, Glycomix
Fucus distichus subsp. evanescens (C.Agardh) H.T.Powell *(formerly Fucus evanescens)*	*Fe*	in-house purification [25]
Fucus vesiculosus	*Fv*	F8190, Merck
Laminaria digitata	*Ld* (commercial)	YF09361, Glycomix
Laminaria digitata (Huds.) J.V.Lamour	*Ld*	in-house purification [25]
Saccharina latissima (L.) C.E.Lane, C.Mayes, Druehl & G.W.Saunders	*Sl*	in-house purification [26]
Sargassum mcclurei Setch.	*Sm*	in-house purification [26]
Sargassum oligocystum Mont.	*So*	in-house purification [26]
Sargassum polycystum C. Agardh	*Sp*	in-house purification [26]
Sargassum serratum Nguy Huu Dai	*Ss*	in-house purification [26]
Turbinaria ornata (Turner) J.Agardh	*To*	in-house purification [26]

95 °C, with mixing at 500 RPM. For selected enzymes, further tests were conducted, using the same reaction parameters at a range of reaction times (up to 24 h), a range of reaction temperatures (from 10–65 °C or 70 °C in 5 °C or 10 °C intervals) and other fucoidan substrates (Table 1). C-PAGE [23] was used as a qualitative investigation of fucoidanase activity in each of the fucoidanase assays. C-PAGE samples were prepared by mixing 20 μL of inactivated reaction mixture with 20 μL of loading buffer (20 % (v/v) glycerol, 0.02 % (w/v) phenol red). A solution containing 1 % (w/v) of the relevant fucoidan served as a corresponding negative substrate control. Reaction products of fucoidanase Fhf1 [23] on *Fe* was used as a C-PAGE standard.

2.9. Production upscaling of ulvan lyases

For ulvan lyases, recombinant enzyme production upscaling was performed at a 200 mL expression scale in LB-Lennox media supplemented with 100 μg/mL of ampicillin. The expression media was inoculated with 1 % (v/v) overnight culture of *E. coli* JM109 (Promega) transformed with the GOI expression plasmids and cultivated at 37 °C. When the culture's OD_{600} reached 0.4—0.6 the cultivation temperature was lowered to 28 °C and after temperature equilibration, GOI expression was induced with 0.15 % (w/v) of L-rhamnose for 4 h at 28 °C. When enzyme production was optimized, induction with 0.05 % (w/v) or 0.15 % (w/v) L-rhamnose was performed at 16 °C or 28 °C in all possible combinations. The expression cultures were harvested by centrifugation and lysed by sonication after resuspension in nickel affinity chromatography binding buffer (50 mM Tris-HCl pH 7.4, 50 mM imidazole, 500 mM NaCl, and 10 % (v/v) glycerol) or amylose affinity chromatography binding buffer (50 mM Tris-HCl pH 7.4, 10 % (v/v) glycerol). Insoluble protein fraction was separated by centrifugation at 12,000×*g* for 15 min at 4 °C. The protein fractions obtained were analyzed with 4–15 % glycine-SDS-PAGE (Bio-Rad). GelDoc Go imaging system (Bio-Rad) was used for gel photography, performing produced enzyme relative abundance quantification in protein fraction SDS-PAGE profile by densitometry [27] using Image Lab v6.1 program (Bio-Rad). Total protein concentration was measured spectrophotometrically assuming A_{280} of 1 as corresponding to 1 mg/mL.

2.10. Purification of ulvan lyases

The enzymes were purified from soluble protein fraction applying nickel affinity chromatography with HisTrap HP 1 mL (7×25 mm) column (Cytiva) or amylose affinity chromatography with MBPTrap HP 1 mL (7×25 mm) column (Cytiva). Proteins bound to affinity resin were eluted with nickel affinity chromatography elution buffer (50 mM Tris-HCl pH 7.4, 500 mM imidazole, 500 mM NaCl, and 10 % (v/v) glycerol) or amylose affinity chromatography elution buffer (50 mM Tris-HCl pH 7.4, 10 mM maltose, and 10 % (v/v) glycerol) after extensive washing with binding buffers. MalE-Smt3 domains in fusion proteins were cleaved with Ulp1 protease [28]. Following cleavage reactions, the enzymes were purified applying nickel affinity chromatography. Elution buffer was exchanged applying group separation with HiTrap Desalting 5 mL (16×25 mm) columns (Cytiva) to ulvan lyases sample buffer (50 mM Tris-HCl pH 7.4, 10 % (v/v) glycerol). The purified proteins were stored in sample buffer at 4 °C after filtration. The enzyme purity and integrity were assessed with SDS-PAGE. Protein concentrations were measured spectrophotometrically considering calculated extinction coefficients.

2.11. Activity determination of purified ulvan lyases

Activity of purified ulvan lyases was tested against a panel of ulvan substrates (Table 2). In the tests, reaction mixtures had a final volume of 250 μL and contained 0.5 % (w/v) ulvan and 1 μg ulvan lyase in a ulvan lyases reaction buffer (50 mM Tris-HCl pH 7.4, 1 % (w/v) NaCl, 1 mM $CaCl_2$, and 10 % (v/v) glycerol). Activity was determined after incubation for 3 h at 30 °C with mixing at 500 RPM. Reactions were inactivated by heating for 10 min at 95 °C with mixing at 500 RPM. Further, tests were conducted using the same reaction parameters with incubation times in the range of 3–24 h and with incubation temperatures in the range of 15–55 °C with 10 °C intervals.

The activity of purified ulvan lyases was determined with the DNS assay and by measuring A_{235} spectrophotometrically, while ulvan degradation products were analyzed with TLC. One unit (U) of ulvan lyase activity, when determined with the DNS assay, was defined as the amount of enzyme required to produce 1 μmol of D-xylose reducing sugar equivalent per minute. Unsaturated ulvan degradation products measured at A_{235} were quantified considering extinction coefficient of the double bond. The enzyme inactivated by heating for 10 min at 100 °C served as a corresponding control. One unit (U) of ulvan lyase activity, when determined by measuring unsaturated products, was defined as the amount of enzyme required to produce 1 μmol of products per min.

3. Results

3.1. In situ enrichment

The study aimed to obtain metagenomes enriched in diverse, novel, and potentially thermostable CAZymes that degrade fucoidan and ulvan.

Table 2
Overview of the ulvans used for substrate specificity determination of ulvan lyases.

Ulvan origin	Abbreviation	Source
Ulva lacinulata (formerly *Ulva armoricana)*	*Ua* (native grade)	ULV100, ELICITYL
Ulva lacinulata (formerly *Ulva armoricana)*	*Ua* (fine grade)	ULV102, ELICITYL
Ulva lactuca L.	*Ul*	in-house purification [29]
Ulva papenfussii P.H.Hô	*Up*	in-house purification [29]
Ulva sp.	*Usp.*	ALGAplus

To that end an *in situ* enrichment was carried out for microbial cells in an intertidal hot-spring. The selected hot-spring, called Skarðshver, is located in Miðfjörður fjord on the coast of northern Iceland, close to the village Hvammstangi. At low tide, a pool forms around the source of the hot-spring with a water temperature of approximately 50 °C, while at high tide the pool is submerged by seawater. The shore surrounding the hot-spring has a natural dense growth of brown seaweed (Fig. 1**A**). Further, green seaweeds grow in smaller quantities near the spring, primarily in a cluster approximately 5—10 m from the source. Therefore, it can be expected that the hot-spring and surrounding sea-water is enriched in microbial communities that utilize this biomass for growth, and hence harbor enzymes that break down fucoidan and ulvan.

The enrichment substrate contained a combination of mixed green and brown seaweeds (a source of ulvan and fucoidan, respectively) together with a carrageenan gel supplemented with fucose and fucoidan. The green seaweed was sourced from the Icelandic coast out of Reykjanes peninsula. The brown seaweed was collected from the hot-spring and was partially degraded. Carrageenan gels were placed in two small nylon mesh bags (approximately 10×10 cm) that were filled with the seaweeds. They were each placed in larger mesh bags (approximately 50×50 cm) that were filled with the seaweeds. The two bags were held in place in the hot-spring by placing them under large rocks at different locations in the pool and were incubated for five weeks. No trace of the carrageenan gels was left at retrieval, possibly due to limited thermal resilience. The four samples, i.e., the two large bags, and two small bags, were given sample names SH1S, SH1L, SH2S and SH2L (for Skarðshver; bag number; Large/Small). Using a combination of filtration and centrifugation in several steps, the cells and the spent enrichment substrate were separated, and the cells from each sample suspended in sterile water.

3.2. *Metagenome DNA isolation, sequencing, and assembly*

Several methods for DNA isolation from the cellular fractions obtained from the enrichment samples were tested and compared to ensure sufficient DNA quality for library preparation and sequencing, and to ensure faithful representation of the genetic diversity of the enrichment community. The tests were conducted on cell fractions from the two large bags, SH1L and SH2L, using the DNA purification kits MasterPure Complete, MasterPure Gram Positive, DNeasy PowerSoil and FastDNA SPIN. The apparent quantity of isolated DNA was several-fold greater from isolation using the two MasterPure series kits compared with DNeasy PowerSoil and FastDNA SPIN kits (Table 3). DNA purity was however suboptimal in DNA isolated with MasterPure series and

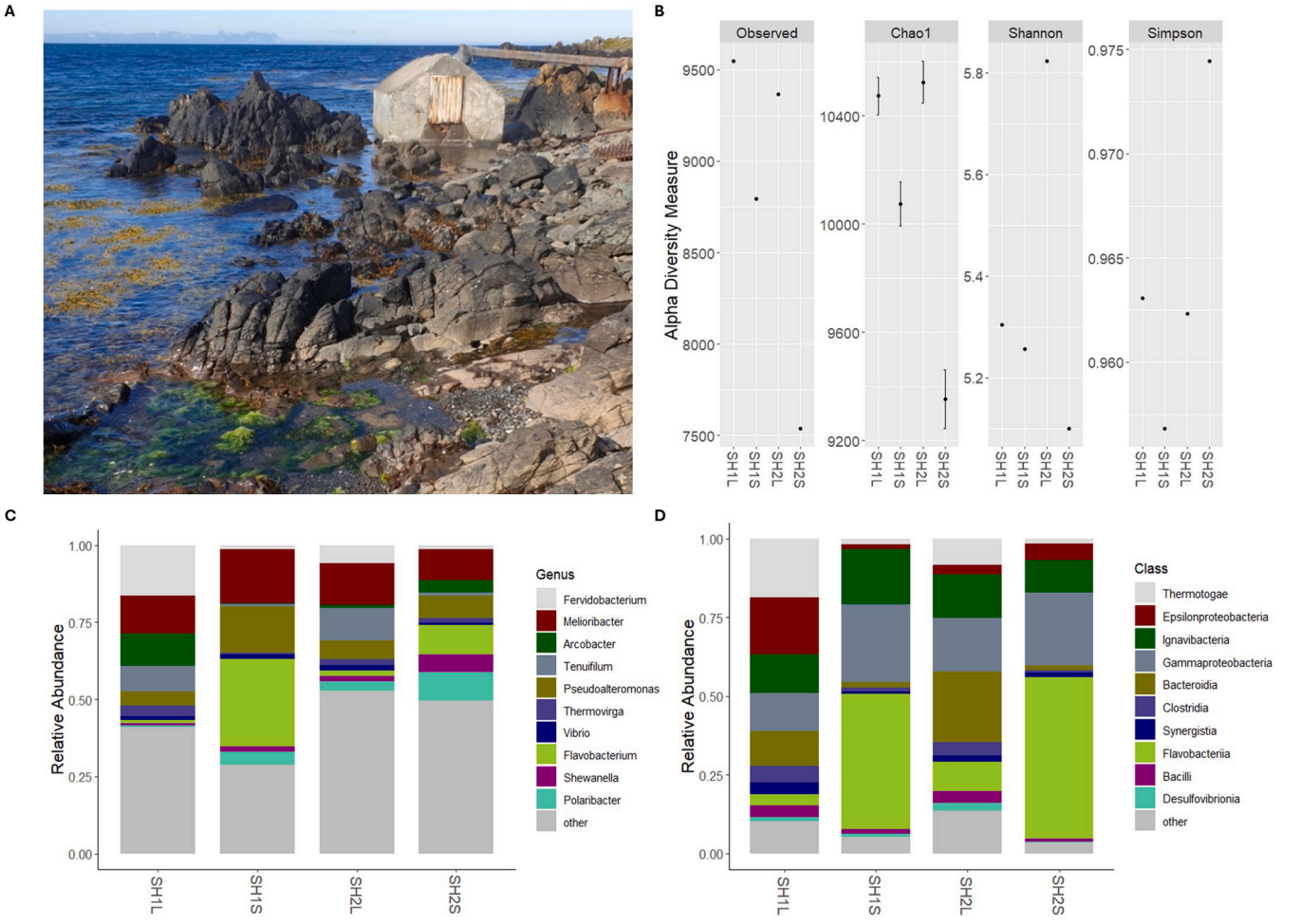


Fig. 1. Skarðshver intertidal hot-spring, the site of *in situ* enrichment, and species richness and abundance in metagenomes. A: The source of the hot-spring is located inside a concrete building (for scale, the building is approximately 1 m high). The building was presumably erected to collect the hot water for use in a nearby farm, but is not currently in use and hot water seeps freely out into the surrounding and collects in a pool under the side of the structure that directly opposes the viewer. The photo was taken when the tide was high enough for seawater to flow over the pool. The shore surrounding the pool is rich in seaweeds. Note the green seaweeds in the foreground and brown seaweeds to the left. B: Alpha diversity of the four metagenomes collected from the Skarðshver enrichment cultures, SH1L, SH1S, SH2L and SH2S. Differences in alpha diversity observed between samples were not significant. C and D: Relative abundance at the Genus and Class levels, respectively, showing taxa making up 0.5 % or more relative abundance individually and taxa with lower abundance are grouped as "other".

Table 3
Quantity and quality of DNA isolated from enrichment cultures.

		Spectrophotometry			Fluorometry	
	Sample	*ng/μL	260/280**	260/230**	ng/μL***	μg dsDNA***
MasterPure Complete	SH1L	3086	1.69	1.08	1250	121
	SH2L	986	1.74	1.05	540	52
MasterPure Gram Positive	SH1L	1988	1.81	1.22	1220	118
	SH2L	910	1.76	0.97	540	52
DNeasy PowerSoil	SH1L	296	1.95	2.10	222	22
	SH2L	153	1.90	2.05	98	9
FastDNA SPIN	SH1L	299	1.93	0.77	200	19
	SH2L	147	1.96	0.27	95	9

* Total nucleotide concentration;
** absorbance ratios at indicated wavelengths (nm);
*** concentration and total quantity of double stranded DNA.

FastDNA SPIN kits (Table 3**,** as indicated by low 260/230 nm absorbance ratios). In the case of MasterPure series kits, this was consistent with the visible appearance of the resuspended DNA, which had light brown color (Supplementary file 3). Fluorometry analysis of double-stranded DNA concentration showed lower concentration than nanodrop results for all samples, indicating presence of single stranded DNA and/or free nucleotides in all samples. DNA isolated with the DNeasy PowerSoil kit was of the highest purity (Table 3), with 260/230 nm absorbance ratio >2, and a satisfactory final yield of double stranded DNA, although lower that that obtained with the MasterPure series kits. DNA isolated with all kits appeared of high molecular weight with no significant signs of degradation (Supplementary file 3).

For each DNA isolation, the 16S rRNA gene was amplified, individually barcoded and sequenced on a MiSeq system. For each sample, approximately 50,000—100,000 sequences were obtained. A combined bacterial and archaeal diversity analysis via DADA2 revealed that both the observed number of taxonomic groups and the Shannon index of species richness was comparable or higher in samples isolated with the PowerSoil method compared with other DNA isolation methods (Supplementary file 3). Combined with the superior quality of the DNA from this isolation method, the PowerSoil DNA isolation was chosen for the preparation of sequencing libraries for metagenome sequencing of all samples.

Metagenome sequencing libraries were prepared with DNA from microbial cell fractions isolated from samples SH1L, SH1S, SH2L and SH2S. The libraries were sequenced on a MiSeq system, contigs were assembled using the sequencing data and ORFs were identified on the contigs. After exclusion of contigs shorter than 1000 bp, the metagenomes contained 27,073 contigs on average, with an average total length of 77,220,025 bp per sample. The longest contigs were on average 443,436 bp, the N50 contig length was on average 4012 bp and nucleotide coverage was 10.7 on average (Table 4). The average number of predicted ORFs was 88,609 per sample. A large proportion of the called ORFs were truncated at one or both ends which precluded cloning of some GOI. To obtain additional metagenomic data to extend the truncated GOI and acquire their complete sequence, two approaches were used. First, a pooled assembly using the combined sequencing data from all four enrichments was performed. In the pooled assembly, a total of 93,336 contigs were obtained that were 1000 bp or larger. The total metagenome length was 252,235,582 bp, its largest contig was 358,046 bp, N50 contig length was 3296 bp and it contained a predicted total of 295,509 ORFs (Table 4). The second approach involved long-read sequencing of the metagenomes by Oxford Nanopore technology and a hybrid assembly using the MinION and MiSeq data combined. The basic principle involves assembling MinION data to provide a scaffold of long contigs with a relatively high sequence error that are overlaid with MiSeq data of higher quality, hence reducing the probability of sequence error. After exclusion of contigs shorter than 1000 bp, the hybrid assemblies produced on average 23,384 contigs and an average total length of 123,859,067 bp per sample. Their longest contig was on average 1074,460 bp, N50 contig length was 12,831 bp on average and the nucleotide coverage was 7.5 on average (Table 4). The average number of predicted ORFs was 148,089 per sample.

Taxonomic characterization of metagenomes was performed using the trimmed Illumina MiSeq reads. Calculation of alpha diversity indices by phyloseq indicated that the highest sequence richness (Observed/Chao1) was found in the two enrichments in large bags (SH1L and SH2L) and the lowest species richness in the SH2S sample (Fig. 1**B**). The Shannon index showed that the highest diversity was found in sample SH2L (Shannon), and the lowest microbial diversity was found in sample SH2S (Simpson). Overall, the differences in alpha diversity among samples were small and none of these diversity indices were found to be significantly different between the two sampling sites or the two enrichment methods by pairwise Wilcoxon tests (p-values 0.33—1).

Analysis of relative abundance at the species level showed that the facultatively anaerobic, thermophilic and cellulolytic *Melioribacter roseus* [30] was the most abundant species in all samples, with a relative abundance of 10.1 % in SH2S to 17.4 % in SH1S. This was mirrored at the genus level where *Melioribacter* was the first or second most abundant genus in all samples (Fig. 1**C**). Other genera above 0.5 % relative abundance can be grouped into either marine and soil microbes, often adapted to cold environments, such as *Arcobacter*, *Pseudoaltaeromonas*, *Vibrio* and *Polaribacter* or thermophilic or moderately thermophilic facultative anaerobes including *Fervidobacterium*, *Tenuifilum* and *Thermovirga*. This mixture clearly represents the sampling environment of hot-springs found in the intertidal zone and flooded by cold seawater, and the samples demonstrate an adaptation to salinity, high temperature and reduced oxygen availability. Overall, the four samples had a similar composition of genera but at the class level both samples extracted from the small enrichment bags (SH1S and SH2S) had high abundance of *Flavobacteriia* while in the large enrichment bags (SH1L and SH2L) *Thermotogae* and *Bacteroidia* were relatively more abundant, and no single class was as dominating (Fig. 1**D**). All samples contained a similar abundance of *Ignavibacteria*, containing the most abundant species *M. roseus*, and *Gammaproteobacteria*. Together these results showed that

Table 4
Summary of yield and quality measures of metagenome sequencing.*

Assembly	Total contig count	Longest contig (bp)	Total length (bp)	GC (%)	N50	Depth	ORF count
SH1L Illumina	24,214	426,499	86,334,558	43.2	6311	18.8	95,073
SH1S Illumina	29,308	307,396	76,144,344	41.0	3015	7.3	88,619
SH2L Illumina	23,343	220,022	71,328,004	42.6	4147	10.9	80,126
SH2S Illumina	31,427	819,827	75,073,195	37.8	2576	5.7	90,619
Pooled assembly	93,336	358,046	252,235,582	39.5	3296	14.2	295,509
SH1L Hybrid	17,944	2028,128	105,441,868	42.9	19,211	13.4	116,440
SH1S Hybrid	25,353	576,003	131,080,673	41.6	9896	5.3	159,518
SH2L Hybrid	17,058	854,815	112,534,839	41.7	15,142	7.7	131,088
SH2S Hybrid	33,182	838,895	146,378,886	39.0	7073	3.5	185,311

* After exclusion of contigs shorter than 1000 bp.

the samples from the two sampling sites (SH1 vs. SH2) had a very similar microbial composition both at the genus and class level but a greater difference was seen between the two enrichment approaches (i.e., in small and large enrichment bags), perhaps through determining access to environmental nutrients and oxygen.

3.3. Identification of genes encoding CAZymes

In metagenomes from the *in situ* enrichments (SH1L, SH1S, SH2L and SH2S), assembled with Illumina MiSeq reads only and without pooled assembly data, a total of 354,435 ORFs were identified. To predict for presence of CAZy motifs all ORFs were compared against the CAZy database [11] using the dbCAN2 server [15] with HMMER, DIAMOND and Hotpep. Families GH174, GH187 and PL43 are recent additions to the CAZy database and they were not included in the screening. In total, after exclusion of homologs identified in different samples, genes encoding 38 putative fucoidanases and 71 putative ulvan lyases were identified Table 5) – a total of 109 GOI. They were given identifiers reflecting the sample of origin (SH1L, SH1S, etc.), appended with "Fuc" for fucoidanase or "Ulv" for ulvan lyases, and a running number – e.g., SH1L_Fuc1, SH1L_Fuc2, etc.

To assess the sequence diversity of the GOI, the respective protein sequences were compared to one another using blastp. The mean maximal identity was 61.7 % (Fig. 2**A**), suggesting that the identified sequences are quite diverse. The maximum amino acid sequence identity between two separate translated GOI was 92 %. Further, for assessment of the novelty of the encoded enzymes, a sequence alignment (blastp) was performed against three databases: the NCBI nr database; the entire CAZy database; and the subset of the CAZy database that includes only enzymes with characterized activity. The enzymes encoded by GOI identified in the metagenomes had a mean maximal identity of 72.7 % with sequences in the NCBI nr database, of 61.5 % with sequences in the entire CAZy database, and of 44.2 % with characterized enzymes in CAZy database (Fig. 2**A**), suggesting they are quite divergent from previously studied CAZymes and hence may have previously undiscovered characteristics. Accordingly, the sequences are dispersed across their respective phylogenetic trees (Fig. 2**B,** Fig. 2**C**, Supplementary file 4 and Supplementary file 5).

In the metagenome data many of the ORFs were truncated at one or both ends. Of the 109 GOI, 84 of the respective protein sequences were truncated at the N-terminus, C-terminus, or both. For the ORFs of interest with truncations, a search was performed for the respective full-length sequences in the pooled assembly and in the hybrid assembly. Of the 84 truncated target ORFs, 42 respective full-length sequences (gene with start and stop codon) were obtained from the hybrid assembly, and 3 from the pooled assembly. A further 14 sequences were found in the hybrid and pooled assemblies that were longer than the original, yet still lacking a start or a stop codon.

3.4. Cloning and domain architecture analysis of genes encoding CAZymes

Of the 109 GOI identified, a subset were cloned and expressed to obtain soluble enzymes for functional characterization. Briefly, the methodology involved PCR amplification of target genes, assembly into an expression vector, transformation of *E. coli* with the assembled plasmids, culture of transformants, and evaluation of enzyme production with SDS-PAGE. An in-house expression vector, pHWG1106, was used, that allows for generation of different fusion variants of target enzymes. Initially GOIs were cloned to generate proteins with no fusion on N-terminus and only with C-terminal His-tag, and the respective enzymes referred to as 'native' variants. If expression of the GOI in the resulting clone resulted in low enzyme yield and/or insolubility, the gene was re-cloned to generate an enzyme with N-terminal MalE-Smt3 fusion and a C-terminal His-tag, and the respective enzymes referred to as 'MalE' variants. An additional strategy to increase enzyme yields and/or solubility involved cloning truncated versions of long GOIs. The GOIs were cloned to generate truncated enzymes with C-terminal His-tag only and the respective enzymes referred to as 'S' variants, or they were cloned to generate truncated enzymes with additional MalE-Smt3 fusion and referred to as 'S+MalE' variants.

In total, cloning of 45 GOI from the metagenomes (Table 5, Table 6) was attempted. The genes and corresponding protein sequences were submitted to the NCBI GenBank database and are available under the 45 identifiers in the range of PP839817—PP839861. In total, 42 GOI were successfully cloned, generating a total of 68 clones (counting all constructs; sequences of successfully cloned constructs are presented in Supplementary file 2). Amplification of two GOIs was unsuccessful, hindering their cloning, and for one GOI assembly of the respective amplicon into the expression vector failed.

For all GOIs selected for cloning, protein domains in the respective translated sequences were analyzed (Fig. 3). The putative GH107 enzymes contained diverse domains: Besides regions of homology with the GH107 family, three enzymes contained no detectable protein domain, three enzymes contained only a secretion signal (domain identifier IPR026444), one enzyme contained a secretion signal and a domain associated with carbohydrate binding (G3DSA:2.60.120.260), two contained only one or more immunoglobulin-like or Cadherin-like domains (IPR013783, IPR015919, and PF05345), one contained immunoglobulin-like and Cadherin-like domains and a repeat domain (IPR044060), and three enzymes contained domains associated with glycosidase activity (IPR017853 and G3DSA:3.20.20.80) in combination with a secretion signal, immunoglobulin domain, or a Carboxypeptidase regulatory domain-like (IPR008969). The domain architecture of putative GH168 enzymes was more homogenous: they all contained domains associated with glycosidase activity (IPR029455, IPR017853 and G3DSA:3.20.20.80).

In the putative PL24 enzymes, besides regions of homology with the PL24 family, one enzyme contained no detectable protein domain, one contained only an immunoglobulin-like domain (IPR013783), six contained the poorly defined 'BNR repeat-containing family member' domain (PF15892) only and a further two contained this domain in addition to immunoglobulin-like domains (IPR013783 and PF17957). In 7 of the 9 putative PL25 enzymes the 'BNR repeat-containing family member' domain (PF15892) was detected, in five cases overlapping with a Sialidases domain (IPR036278) that is implicated in the hydrolysis of sialic acid linkages. Two of the putative PL25 enzymes carried domains associated with polysaccharide degradation (IPR025975, G3DSA:2.60.120.200). The putative PL37 enzymes all carried a 'Six-hairpin glycosidase' domain (IPR008928) which is associated with diverse pathways including ulvan degradation, and/or a 'Chondroitin AC/alginate lyase' domain (IPR008929) which is associated with polysaccharide degradation functions, including of ulvan. All putative PL40 enzymes also contained this domain and in addition they all contained a domain associated with heparin- and alginate lyases (G3DSA:2.70.98.70). One putative PL40 enzyme carried a domain that currently is not associated with EC classification (G3DSA:2.60.40.2700), and one putative PL40 enzyme carried an immunoglobulin-like domain (IPR013783) and a putative carbohydrate binding domain (IPR008979,

Table 5
Putative fucoidanases and ulvan lyases identified, cloned, and solubly produced.

		Identified	Cloning Attempted	Cloning Successful	Produced
Fucoidanases	**GH107**	19	13	13	10
	GH168	19	6	6	6
Ulvan lyases	**PL24**	22	10	8	7
	PL25	11	9	8	8
	PL28	0	0	0	0
	PL37	5	4	4	3
	PL40	33	3	3	3
Total		109	45	42	37

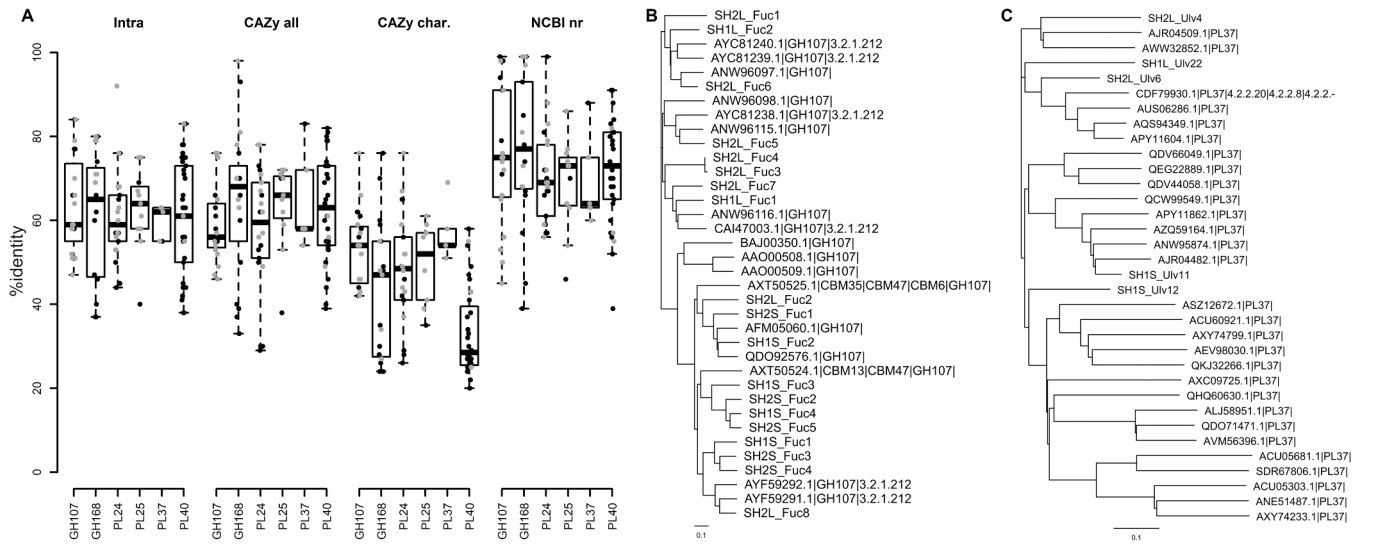


Fig. 2. GOI sequence identity and phylogeny compared with CAZy and NCBI nr databases. A: Sequence identity of proteins encoded by GOI compared with one another (intra), with the entire CAZy database (CAZy all), with the subset of enzymes in CAZy with characterized activity (CAZy char.) and with the NCBI nr database (NCBI nr). In each boxplot datapoints of GOI selected for cloning are displayed in grey and other sequences, identified in the metagenome but not selected for cloning, are in black. B and C: Phylogenetic trees of GH107 sequences and PL37 sequences, respectively. The trees include all protein sequences encoded by GOI of the respective enzyme family, along with all respective protein sequences from the CAZy database. In the latter case, EC classification is indicated for sequences with confirmed activity, while no EC classification is displayed for sequences that are included in the CAZy database on basis of sequence similarity alone (i.e., the respective activity has not been characterized).

cd04080), associated with diverse functional pathways, including ulvan degradation.

A total of 7 truncated constructs were cloned, where C-terminal sequences of the respective translated GOI were excluded (Fig. 3). The truncations removed: C-terminal immunoglobulin-like domains in four enzymes (SH1L_Fuc2, SH2L_Fuc1, SH1L_Ulv11, SH1L_Ulv12); and a 'Carboxypeptidase regulatory domain-like' in one enzyme (SH2L_Fuc2). In two enzymes that contained two apparently separate domains involved in polysaccharide degradation (SH1L_Ulv31 and SH1L_Ulv34), the truncations removed one of the two domains.

3.5. Primary determination of enzyme production yield and activity

For all cloned constructs, a primary determination of enzyme production was performed in a 30 mL culture volume in *E. coli*. After an overnight cultivation, during which expression of the GOI was induced, the cells were harvested, sonicated, spun in a centrifuge and the soluble and insoluble cell lysate fractions analyzed by SDS-PAGE. Soluble enzyme production was observed for 37 of the respective GOI, in one or more genetic constructs (Table 5 and Table 6). Where MalE fusion was included, it increased protein yields in most cases compared with yields of the same protein without MalE fusion (Table 6). In contrast, truncation alone did not result in increased protein yields (Table 6). The observed effects of MalE fusion and truncations on protein yields are consistent with bioinformatic predictions (Supplementary file 6), as determined by SoluProt [22].

For all constructs that yielded soluble proteins, a primary determination of enzymatic activity was performed by incubation of clarified cell lysates with fucoidan or ulvan substrates. Putative fucoidanases were incubated with different types of fucoidan polysaccharides isolated from *Fucus vesiculosus, Ascophyllum, Laminaria digitata* and *Laminaria hyperborea* and reaction products analyzed with C-PAGE. Putative ulvan lyases were incubated with ulvan polysaccharides from *Ulva lacinulata (formerly Ulva armoricana)* and *Ulva intestinalis* and the products analyzed with TLC. For the initial screening for activity, the reactions were incubated overnight at one to two temperatures and at two to three pH levels. Where activity was detected, further reactions were performed at a wider range of reaction times, temperatures, and pH levels and products analyzed with C-PAGE, TLC and/or DNS assay. In total, enzymatic activity was detected in five GH107 fucoidanases, two PL24 ulvan lyases, five PL25 ulvan lyases, and one PL40 ulvan lyase (Table 6, Fig. 4 and Supplementary file 7).

3.6. Production upscaling and purification of fucoidanases

Production upscaling was performed for all 19 fucoidanases in a total of 23 expression constructs (Supplementary file 8). Production upscaling was successfully established for all constructs. Overall, upscaled production levels corresponded to levels observed in primary testing. However, in several cases the fucoidanases were produced at substantially higher levels after upscaling. Enzymes with MalE fusion were produced more efficiently than native variants, although the observed trend was not as pronounced as in the primary determination of production yield. Purification applying nickel affinity chromatography was successfully established for all the produced fucoidanase sequence variants (Supplementary file 8). Proteolytic removal of MalE fusion in respective enzymes was not attempted.

3.7. Activity of purified fucoidanases

For purified fucoidanases, activity tests were performed on a range of fucoidan substrates. As reducing sugar assays had previously been unsuccessful for fucoidanases [31], activity was exclusively investigated using C-PAGE. Ten fucoidanases displayed activity on one or more of the fucoidans tested (Table 7). Optimal temperature and incubation time were subsequently determined for the enzymes, in each case using a substrate that the enzyme degraded most efficiently. The fucoidanases were mostly active on fucoidan from *S. latissima*, *L. digitata*, and *L. hyperborea*. The fucoidan from these species is characterized as type (I) fucoidan, consisting mainly of α-1,3-L-fucose, suggesting that the fucoidanases are specific towards the α-1,3-L-fucososyl linkages [32,33]. One fucoidanase, SH2L_Fuc5, was active on fucoidan from *Fucus distichus subsp. evanescens*, a type (II) fucoidan consisting of L-fucose residues linked by alternating α-1,3 and α-1,4 linkages [32]. Two fucoidanases, SH1S_Fuc3 and SH1L_Fuc3, were active on fucoidan from *Ecklonia radiata*, the structure of which is not characterized. However,

Table 6
Summary of cloning and primary determination of enzyme yield and activity.

GOI ID	CAZy family*		Cloned constructs**	Soluble protein yields***				Active ****
	dbCAN	CUPP		native	MalE	S	S+MalE	
SH1L_Fuc2	GH107	GH107:1.1	native, S+MalE	-			++(+)	-
SH1S_Fuc1	GH107	GH107:2.1	native, MalE	+	++(+)			Y
SH1S_Fuc2	GH107	GH107:2.1	native	+(+)				Y
SH1S_Fuc3	GH107	GH107:2.1	native, MalE	-	+			Y
SH1S_Fuc4	GH107	GH107:2.1	native	-				-
SH2L_Fuc1	GH107	GH107:1.1	native, S, S+MalE	-		-	-	-
SH2L_Fuc2	GH107	GH107:2.1	S			++		Y
SH2L_Fuc3	GH107	GH107:1.1	native	-				-
SH2L_Fuc4	GH107	GH107:3.1	native, MalE	(+)	+++(+)			-
SH2L_Fuc5	GH107	GH107:3.1	native	+				-
SH2L_Fuc6	GH107	GH107:1.1	native, MalE	-	+(+)			-
SH2L_Fuc8	GH107	GH107:2.1	MalE		++			Y
SH2S_Fuc2	GH107	GH107:2.1	native, MalE	-	++			-
SH1L_Fuc3	GH168	GH168:1.1	native	++				-
SH1S_Fuc8	GH168	GH168:3.1	native	+++				-
SH1S_Fuc9	GH168	GH168:6.1	native	++				-
SH1S_Fuc11	GH168	GH168:1.1	native	++(+)				-
SH1S_Fuc13	GH168	GH168:1.1	native	++(+)				-
SH2S_Fuc6	GH168	GH168:6.1	native, MalE	-	++(+)			-
SH1L_Ulv2	PL24	PL24:1.1	native, MalE	-	+(+)			-
SH1L_Ulv3	PL24	PL24:1.1	native, MalE	-	+++			-
SH1L_Ulv4	PL24	PL24:1.1	native	++				-
SH1L_Ulv5	PL24	PL24:1.1	native	-				-
SH1L_Ulv6	PL24	PL24:1.1	failed					
SH1L_Ulv10	PL24	PL24:1.1	failed					
SH1L_Ulv11	PL24	PL24:1.1	native, S, S+MalE	(+)		-	++++	Y
SH1L_Ulv12	PL24	PL24:1.1	native, S, S+MalE	-		-	+++	-
SH1L_Ulv15	PL24	PL24:1.1	native, MalE	-	++			Y
SH1S_Ulv4	PL24	PL24:1.1	native, MalE	-	(+)			-
SH1L_Ulv16	PL25	PL25:3.1	failed					
SH1L_Ulv17	PL25	PL25:3.1	native	++(+)				-
SH1L_Ulv18	PL25	PL25:1.1	native, MalE	-	++++			-
SH1L_Ulv19	PL25	-	native, MalE	(+)	++			Y
SH1L_Ulv20	PL25	PL25:1.1	native	++++				Y
SH1L_Ulv21	PL25	PL25:1.1	native	++++				Y
SH1S_Ulv6	PL25	-	native, MalE	(+)	+++(+)			Y
SH1S_Ulv9	PL25	PL25:1.1	native, MalE	(+)	++(+)			-
SH2L_Ulv3	PL25	PL25:1.1	native	++++				Y
SH1L_Ulv22	PL37	PL37:1.1	native	++				-
SH1S_Ulv12	PL37	PL37:2.1	native, MalE	-	-			-
SH2L_Ulv4	PL37	PL37:1.1	native	++				-
SH2L_Ulv6	PL37	PL37:1.1	MalE		+++			-
SH1L_Ulv31	PL40	PL40:1.1	native, S, S+MalE	-		-	+(+)	-
SH1L_Ulv34	PL40	PL40:5.1	native, S, S+MalE	+++		-	++++	-
SH2S_Ulv5	PL40	PL40:1.1	native	+(+)				Y

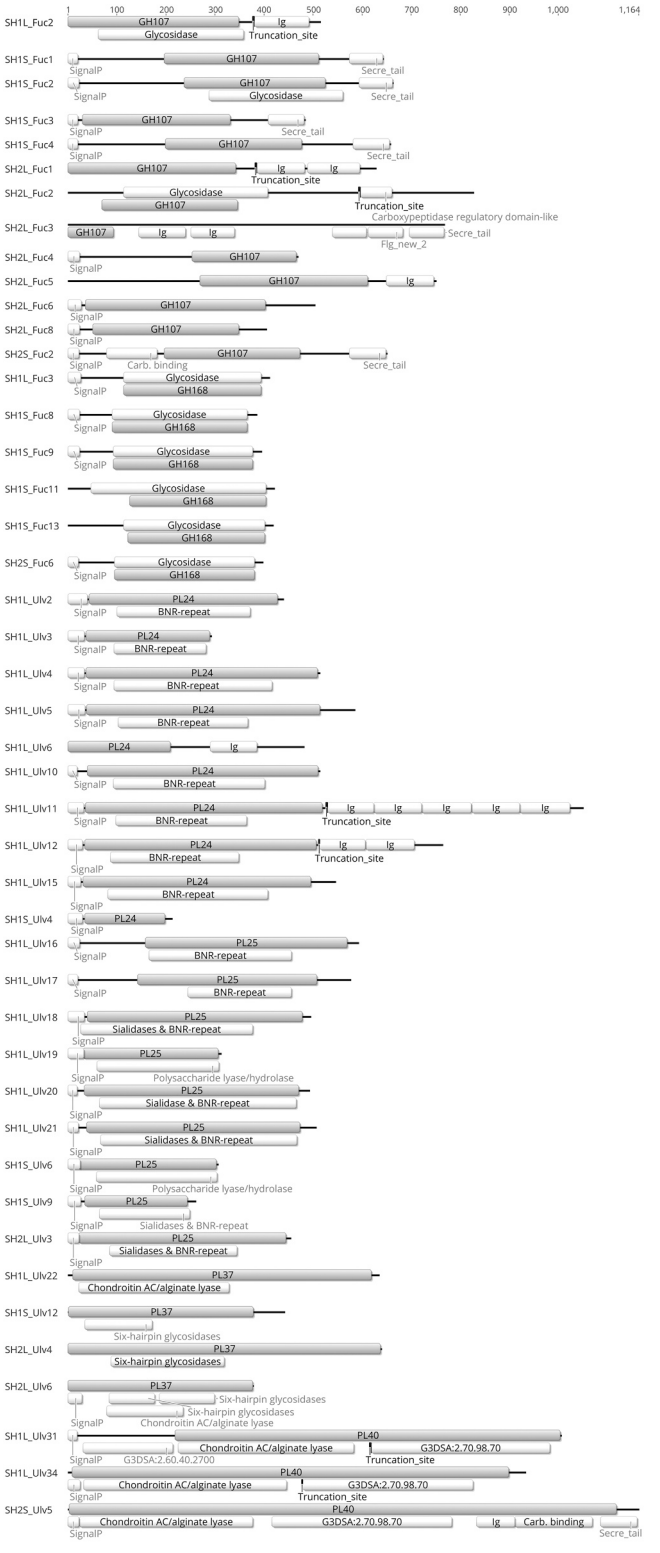
* Predicted CAZy family designation using dbCAN2 and CUPP tools, '-' indicates not detected;
** 'Failed' denotes GOI where cloning was attempted but was unsuccessful;
*** Quantity of protein yields estimated with SDS-PAGE, where '-' indicates no yield, '(+)' indicates lowest yields and '++++' highest yield; empty cells where not applicable;
**** Summary of results from primary enzyme activity tests, where values are: 'Y' where activity was detected in all constructs; '-' where activity was not detected in any construct; empty cells where not applicable.

type (II) fucoidan is present in the related *Ecklonia maxima* [34]. Two of the fucoidanases, SH1S_Fuc1 and SH2L_Fuc2, demonstrated very high activity and had almost depleted the substrate after few minutes of incubation, whereas four others, SH1S_Fuc2, SH1S_Fuc4, SH2L_Fuc5 and SH2L_Fuc8, had much lower activity, requiring hours before significant hydrolysis could be detected (Fig. 5**A**). Three enzymes, SH1S_ Fuc1, SH1S_Fuc3 and SH2L_Fuc2, were active at 60 °C. SH1S_Fuc3 and SH2L_Fuc2 displayed optimal activity at a wide range of temperatures, 15–60 °C and 10–60 °C respectively, while SH1S_Fuc1 displayed optimal activity at lower temperatures, but retained activity at 60 °C.

3.8. Production upscaling and purification of ulvan lyases

Primary enzyme production evaluation led to prioritization of 21 target ulvan lyases for production upscaling in a total of 28 expression constructs. The upscaling was successfully established for all prioritized constructs. Upscaled production levels obtained for 10 out of 15 ulvan lyase native variants were comparable to levels observed during primary production testing, while 5 native variants were produced at lower levels after upscaling (Supplementary file 9). Production optimization was performed for the native variants produced at lower levels (Supplementary file 9). Optimization was however only successful for SH1L_Ulv11 native variant, which was produced slightly more efficiently at lower expression induction temperature. Upscaled production levels obtained for all 13 ulvan lyases with MalE fusion were comparable to primary production evaluation levels (Supplementary file 9). Enzyme production upscaling and optimization revealed that heterologous expression induction with 0.15 % (w/v) L-rhamnose for 4 h at 28 °C resulted in the highest level of ulvan lyases production.

Purification applying nickel affinity chromatography was successfully established for all 15 ulvan lyases native variants (Supplementary file 9). The enzymes were purified to near homogeneity with high purification recovery level. However, aggregation of 12 of the enzymes was observed immediately after elution. Only SH1L_Ulv21, SH2L_Ulv3, and



(caption on next column)

Fig. 3. Domain architecture of proteins encoded by cloned GOI. Protein domains and features predicted by InterProScan are displayed with white annotation boxes and regions of homology with proteins in the CAZy database are displayed with grey annotation boxes. For GOI where cloning of truncated gene sequence was attempted, the location of boundary of the truncation is indicated as 'Truncation_site'. The scale at the top of the image indicates the number of amino-acid residues in the protein sequences. Raw results from InterProScan have been simplified here for clarity. *BNR-repeat* is 'BNR repeat-containing family member' (PF15892); *Carb. binding* is 'Galactose-binding domain-like' (G3DSA:2.60.120.260), 'Galactose-binding-like domain superfamily' (IRP008979), and/or Carbohydrate Binding Module 6 (cd04080); *Chondroitin AC/alginate lyase* is IPR008929; *CPRDL* is 'Carboxypeptidase-like, regulatory domain superfamily' (IPR008969); *Flg_new_2* is 'Bacterial repeat domain' (IPR044060); *Glycosidase* is 'Glycosidases' (G3DSA:3.20.20.80), 'Glycoside hydrolase superfamily' (IPR017853), and/or 'Hypothetical glycosyl hydrolase family 15' (IPR029455); G3DSA:2.60.40.2700 and G3DSA:2.70.98.70 are unnamed database entries; *Ig* is 'immunoglobulin-like fold' (IPR013783), 'Cadherin-like superfamily' (IPR015919), 'Putative Ig domain' (PF05345), and/or Bacterial Ig domain (PF17957); *Polysaccharide lyase/hydrolase* is 'Polysaccharide lyase' (IPR025975), and/or an unnamed database entry G3DSA:2.60.120.200; *Secre_tail* is 'Secretion system C-terminal sorting domain' (IPR026444); *Six-Hairpin glycosidases* is 'Six-hairpin glycosidase superfamily' (IPR008928); *Sialidases* is 'Sialidase superfamily' (IPR036278); *SignalP* is a signal peptide.

SH2L_Ulv4 native variants remained soluble in nickel affinity chromatography elution buffer as well as stored in ulvan lyases sample buffer. Purification applying amylose affinity chromatography was successfully established for all 13 ulvan lyases with MalE fusion, resulting in purification to near homogeneity (Supplementary file 9). Proteolytic removal of MalE with Ulp1 protease was successfully established for all ulvan lyases with MalE fusion. However, only SH1L_Ulv18 remained soluble after MalE removal. C-terminal His-tag ensured SH1L_Ulv18 purification from cleavage reaction to near homogeneity with high recovery level. SH1L_Ulv18 was stable and non-aggregation-prone during final purification, remaining soluble in nickel affinity chromatography elution buffer as well as stored in ulvan lyases sample buffer.

3.9. Activity of purified ulvan lyases

For the four ulvan lyases that remained soluble after purification, enzyme activity was determined on a range of ulvan substrates. Ulvan lyase activity was observed for SH2L_Ulv3 (Fig. 5**B**), but not for SH1L_Ulv21, SH2L_Ulv4, and SH1L_Ulv18. The lack of activity of SH1L_Ulv21 was surprising, since the enzyme displayed activity in primary screening. Most likely, SH1L_Ulv21 did not display catalytic activity after purification due to aggregation after buffer exchange to ulvan lyases sample buffer that was not visually observable. SH2L_Ulv3 demonstrated broad substrate specificity, degrading all ulvan substrates tested (Table 2). It was active at a broad temperature range, 15—55 °C, and optimally at 20—25 °C. The enzyme degraded ulvan from *Ulva lacinulata* (formerly *U. armoricana*) (fine grade) slightly more efficiently than the other ulvans tested. When incubated with *U. lacinulata* (fine grade) at pH 7.4 and 30 °C for 3 h, the enzyme's activity was 5.35 ± 0.09 U/mg and 4.12 ± 0.12 U/mg, as estimated by DNS assay and absorbance of products at 235 nm, respectively.

4. Discussion

In the present study, fucoidanases and ulvan lyases were sourced

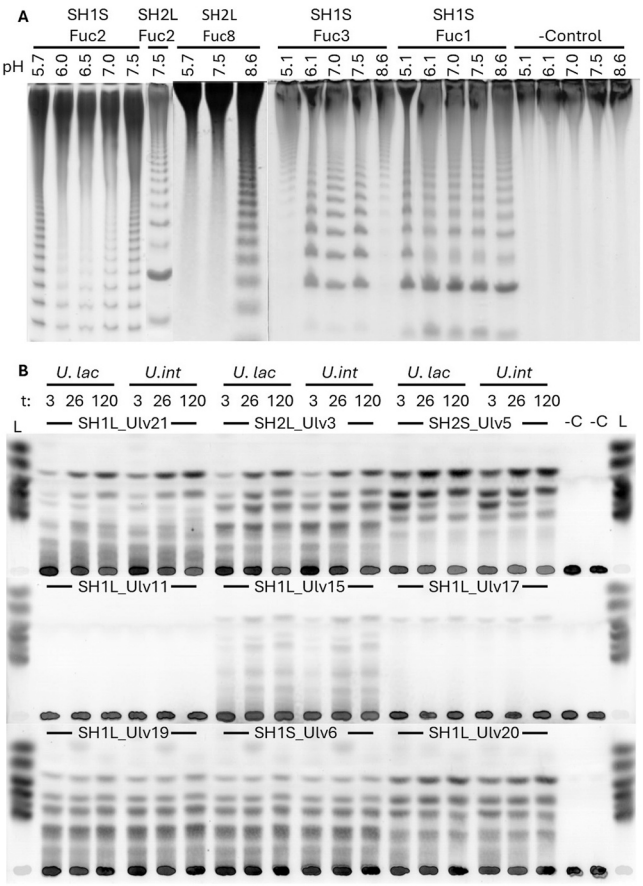


Fig. 4. Primary screening for enzyme activity. A: Examples of results from screening for enzymatic activity of GH107 enzymes using clarified cell lysates. In the tests displayed, fucoidan from *Laminaria hyperborea* was used as a substrate. Reactions were incubated at 30 °C overnight at pH levels indicated at the top of the figure. SH1S_Fuc1, SH1S_Fuc3 and SH2L_Fuc8 MalE variants, SH2L_Fuc2 S variant and SH1S_Fuc2 native variant were used in the tests, along with a lysate of the expression strain without fucoidanase expression plasmid (-Control). The image is a composite of photos taken of four C-PAGE gels that were cropped and further processed to adjust color scheme and intensity. B: Examples of results from screening for enzymatic activity of PL24, PL25 and PL40 enzymes using clarified cell lysates. In the tests displayed, ulvan from *Ulva lacinulata (U. lac) and U. intestinalis (U. Int)* was used as a substrate and reactions performed for 3, 26 or 120 h, as indicated at top of image. Reactions were incubated at 30 °C overnight at pH 7.5, except SH1L_Ulv19 which was incubated at pH 7, and SH1S_Ulv6 and SH1L_Ulv26 which were incubated at pH 6.5. The enzymes tested were native variants, except SH1L_Ulv11 S+MalE variant and SH1L_Ulv15, SH1L_Ulv19 and SH1S_Ulv6 MalE variants, along with a lysate of the expression strain without ulvan lyase expression plasmid (-C). The image is a composite of three TLC plates that were cropped and processed to adjust color and intensity. Absence of observed activity of SH1L_Ulv11 may have been due to a technical error, as activity of this enzyme was detected in a previous assay (supplementary file 7) and was later confirmed in other assays (data not shown).

from an intertidal hot-spring, an environment that is rare globally. To our knowledge, fucoidanases and ulvan lyases have previously not been identified and characterized from marine geothermal metagenomes, nor from microbial species isolated from such environments. In this context, the study provides a blueprint for the discovery of CAZymes that degrade seaweed carbohydrates from metagenomes, in that it demonstrates: the value of careful selection of DNA isolation methods to ensure high nucleic acid quality and representation of sequence diversity; the value of the combination of next-generation and long-read sequencing and combined assembly; and strategies for obtaining soluble recombinant enzymes, and their purification.

Consistent with their unique origin, the primary structure of enzymes identified in the study is divergent from previously characterized fucoidanases and ulvan lyases. Though the organismal origin of enzymes identified in the study cannot be determined, the metagenomes are enriched in DNA from microbial classes *Thermotogae, Epsilonproteobacteria, Ignavibacteria, Gammaproteobacteria, Bacteroidia, Clostridia, Synergistia, Flavobacteriia, Bacilli* and *Desulfovibrionia* – and the enzymes therefore likely originate from one or more of these classes. All previously characterized GH107 fucoidanases and PL24, PL25, PL37 and PL40 ulvan lyases are from bacteria classified in *Gammaproteobacteria* and *Flavobacteriia* classes [11].

At present, seventeen GH107 fucoidanases have been characterized [11]. They are Mef2 [26] and Mef1 [35] from *Muricauda eckloniae,* OUC-FaFcn1 from *Flavobacterium algicola* [36], FFA1 and FFA2 from *Formosa algae* [37,38], Fhf1 and Fhf2 from *Formosa haliotis* [23,39], FcnA from *Mariniflexile fucanivorans* [40], P5AFcnA and P19DFcnA from *Psychromonas* spp. [41] metagenome derived enzymes Fp273, Fp277 and Fp279 [42], and FWf1, FWf2, FWf3 and FWf4 from *Wenyingzhuangia fucanilytica* [43,44]. Overall, the previously characterized fucoidanases function optimally at neutral to slightly alkaline conditions and at temperatures approximately in the range of 25—40 °C. Mef2 catalyzes hydrolysis of endo-α-1,3 glycosidic bonds in fucoidan (EC 3.2.1.211) while all other characterized fucoidanases are specific towards α-1,4 linkages in fucoidan (EC 3.2.1.212). In contrast, the GH107 enzymes SH1S_Fuc1, SH1S_Fuc2, SH1S_Fuc3, SH1S_Fuc4, SH2L_Fuc2, SH2L_Fuc8, and SH2S_Fuc2, described in the present study, appear to have a preference for class I fucoidan enriched in α-1,3 linked L-fucose found in *Saccharina* and *Laminaria* species [33,45] and do not degrade galactofucan from *Sargassum* species [46] or fucoidan with mixed α-1,3 and α-1,4 linkages found e.g., in *Fucus* species [47]. This likely reflects a specificity for hydrolysis of α-1,3 linkages in the substrate. Enzymes SH1S_Fuc1, SH1S_Fuc3 and SH2L_Fuc2 are active at 60 °C, considerably higher than that observed for previously characterized GH107 enzymes. Only one GH168 enzyme, FunA, has so far been characterized and is isolated from *W. fucanilytica* and active on α-1,3-L-fucose linkages (i.e. type (I) fucoidan) [48]. Likewise, the GH168 enzymes SH1L_Fuc3 and SH1S_Fuc8 identified in the present study mainly demonstrated activity on this type of fucoidan, although SH1L_Fuc3 also demonstrated activity on fucoidan from *E. radiata*, which is expected to have alternating α-1,3 and α-1,4 linkages (i.e. type (II) fucoidan) as this type is known to be present in the related *E. maxima* [34]. It is, however, possible that the enzyme targets the α-1,3 linkages in type (II) fucoidan.

At present, 16 ulvan lyases have been characterized in families PL24, PL25, PL28, PL37 and PL40. In family PL24 they are UllA, LOR_107 and LOR_61 from *Alteromonas* spp. [49,50], Uly1 from *Catenovulum maritimum* [51], and PLSV_387 and PLSV_3925 from *Pseudoalteromonas* sp. PLSV [50]. In family PL25 they are ALT3695 and LOR_29 from *Alteromonas* spp. [52,53], NLR_492 from *Nonlabens ulvanivorans* [53], PLSV_3936 from *Pseudoalteromonas* sp. PLSV [54] and TsUly25B from *Thalassomonas* sp. LD5 [55]. In family PL28 they are FA2219 from *Formosa agariphila* [56] and NRL42 and NRL48 from *Nonlabens ulvanivorans* [57,58]. In PL37 and PL40, they are Cdf79930 and P10_PLnc, respectively, from *Formosa agariphila* [59,60]. Where reported, optimal reaction conditions for previously characterized ulvan lyases are at temperatures of 40–60 °C and at neutral to slightly alkaline pH. For example, enzymes PLSV_3875, ALT3695, NRL42, NRL48 were all reported to function optimally at 50 °C and TsUly25B at 60 °C. This is somewhat surprising, given that the enzymes are sourced from marine organisms that grow optimally at considerably lower temperatures. Their thermal stability may however be limited at elevated temperatures, as indicated for NRL42 and NRL48, both of which rapidly lose catalytic function after brief incubation at 50 °C. In contrast, the PL25 enzyme SH2L_Ulv3, described in the present study, was optimally active at modest temperature, 20–25 °C.

Further characterization of enzymes identified in the study is in progress or is planned. For enzymes that are of industrial interest, e.g.,

Table 7
Substrates and optimal reaction conditions for purified fucoidanases.*.

Enzyme	CAZy family	Variant	Substrates**	Optimal reaction conditions		Activity score***
				Time	Temperature (°C)	
SH1L_Fuc2	GH107	native	N/A	18 h	37	0
		S+MalE	N/A	18 h	37	0
SH1S_Fuc1	GH107	native	*Er, Ld, Lh, Sl*	18 h	37	+++
		MalE	*Sl, Ld, Lh*	1 min to 24 h	20–40	+++
SH1S_Fuc2	GH107	native	*Sl, Lh*	2 h - 18 h	30–45	+++
SH1S_Fuc3	GH107	MalE	*Ld, Sl, Er, Lh*	30 min - 24 h	15–60	++
SH1S_Fuc4	GH107	native	*Sl, Ld*	2–24 h	30	+
SH2L_Fuc1	GH107	native	N/A	18 h	37	0
		S+MalE	N/A	18 h	37	0
SH2L_Fuc2	GH107	S	*Sl, Ld, Lh*	3 min - 24 h	10–60	+++
SH2L_Fuc3	GH107	native	N/A	18 h	37	0
SH2L_Fuc4	GH107	MalE	N/A	18 h	37	0
SH2L_Fuc5	GH107	native	*Sl, Fe*	4–24 h	15–30	++
SH2L_Fuc6	GH107	MalE	N/A	18 h	37	0
SH2L_Fuc8	GH107	MalE	*Sl, Ld*	2–24 h	30	+++
SH2S_Fuc2	GH107	MalE	*Ld, Lh, Sl*	30 min - 24 h	10–30	+
SH1L_Fuc3	GH168	native	*Er, Ld, Lh, Sl*	18 h	37	+
SH1S_Fuc8	GH168	native	*Lh, Sl*	18 h	37	++
SH1S_Fuc9	GH168	native	N/A	18 h	37	0
SH1S_Fuc11	GH168	native	N/A	18 h	37	0
SH1S_Fuc13	GH168	native	N/A	18 h	37	0
SH2S_Fuc6	GH168	native	N/A	18 h	37	0
		MalE	N/A	18 h	37	0

* Optimal under the conditions tested; in cases where no activity was detected, the incubation temperature and time indicate the conditions used for the activity testing.
** N/A for not applicable, where enzyme activity was not detected; Er = *Ecklonia radiata*, Fe = *Fucus distichus* subsp. *evanescens*, Ld = *Laminaria digitata*, Lh = *Laminaria hyperborea*, Sl = *Saccharina latissima*.
*** semi-quantitative enzyme activity with +++ being the highest score and 0 being no observable activity.

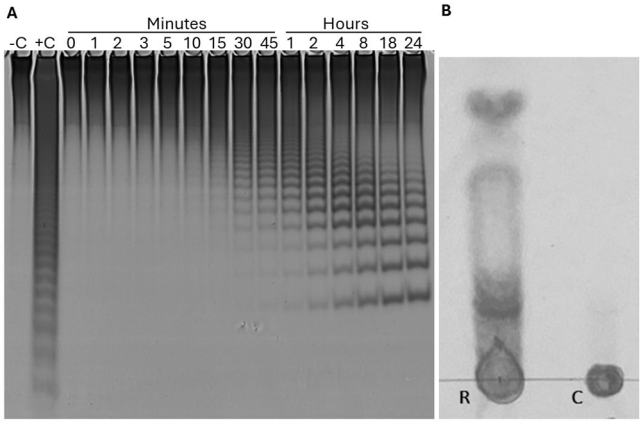


Fig. 5. Examples of enzyme activity screening of purified fucoidanase and ulvan lyase. A: Fucoidan hydrolysis by purified SH1S_Fuc4 fucoidanase, native variant, visualized by C-PAGE, where –C is undigested fucoidan from *Laminaria digitata*; +C is a size standard consisting of *Fucus distichus subsp. evanescens* fucoidan digested with Fhf1 fucoidanase; and in lanes 3 – 17 (labelled '0 minutes' through to '24 hours') *F. evanescence* fucoidan digested with SH1S_Fuc4 fucoidanase at different time periods. B: Ulvan degradation by purified SH2L_Ulv3 ulvan lyase, native variant, visualized by thin layer chromatography. Ulvan from *Ulva lacinulata* (fine grade) was incubated at pH 7.4 with (R) and without (C) the enzyme at 30 °C for 3 h and the products separated by TLC.

those displaying optimal activity at high temperatures, thermostability and resistance to other environmental extremes will be determined. Further, a considerable number of enzymes identified in the study were produced and soluble, but no catalytic activity observed. Various reasons may explain this result, such as mis-folding or mutation of the respective genes during PCR amplification. However, given the number of enzymes in question, the structural diversity of fucoidan and ulvan, and the substrate specificity of fucoidanases and ulvan lyases, it is reasonable to assume that, at least in some cases, the enzymes have a yet undiscovered substrate specificity and that the substrate panel used for characterization in the present study was incompatible for their function. As a result, these enzymes will be studied in further detail to ascertain their potential unique catalytic functions. Many successfully produced enzymes were aggregation-prone after MalE fusion removal and/or purification. Some of these enzymes displayed activity in clarified cell lysates, and they may therefore nevertheless be relevant for implementation in industrial processes. Further development of soluble production of aggregation-prone enzymes and formulation after purification to increase solubility remains possible.

Funding sources

This work was supported through the BlueBio ERA-NET (project Marikat, supported by The Technology Development Fund, contract 2012273–0612 and project MARIKAT-BCOM, supported by The Technology Development Fund, contract 2322501-0611), Horizon Europe (project SeaMark, contract 101060379) and the Icelandic Research Fund (project CAZyme-X, contract 228860–051). The funders had no role in relation to the study design; in collection, analysis or interpretation of data, writing and decision to submit the article for publication. This project was funded by MARIKAT JPI Cofund Blue BioEconomy Project grant No. 9082-00021B, by Innovation Fund Denmark and by the Technical University of Denmark.

CRediT authorship contribution statement

Guðmundur Óli Hreggviðsson: Supervision, Project administration, Funding acquisition, Conceptualization. **Maria Dalgaard Mikkelsen:** Methodology, Investigation. **Eva Nordberg Karlsson:** Supervision, Funding acquisition. **Morten Schiøtt:** Writing – original draft, Visualization, Methodology, Investigation. **Pavithra Sivakumar:** Investigation. **Elísabet Eik Guðmundsdóttir:** Writing – original draft, Visualization, Methodology, Investigation. **Tushar Kaushik:** Investigation. **Annika Malmgren:** Investigation. **Erik Apelqvist:** Investigation.

Rébecca Leblay: Investigation. **Bjorn Thor Adalsteinsson:** Writing – original draft, Visualization, Supervision, Project administration, Methodology, Investigation. **Signe Vangsgaard:** Investigation. **Anne S. Meyer:** Supervision, Funding acquisition. **Andrius Jasilionis:** Writing – original draft, Visualization, Methodology, Investigation. **Ólafur Friðjónsson:** Funding acquisition. **Hörður Guðmundsson:** Visualization, Methodology, Investigation.

Acknowledgements

Vector pHWG1106 was a gift from Hildegard Watzlawick and Josef Altenbuchner (University of Stuttgart).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.enzmictec.2024.110528.

Data Availability

Data that is not included in the manuscript and the associated supplementary material will be made available upon request, where possible/permitted

References

[1] D.J. McHugh, A guide to the seaweed industry., ⟨https://www.fao.org/4/y4765e/y4765e00.htm⟩, 2003 (accessed 30 August 2024).
[2] M. Jönsson, L. Allahgholi, R.R.R. Sardari, G.O. Hreggviðsson, E. Nordberg Karlsson, Extraction and modification of macroalgal polysaccharides for current and next-generation applications, Molecules 25 (2020) 930.
[3] S. Kraan, Algal polysaccharides, novel applications and outlook, in: Carbohydrates-Comprehensive Studies on Glycobiology and Glycotechnology, IntechOpen, 2012.
[4] Y. Sasaki, Y. Yoshikuni, Metabolic engineering for valorization of macroalgae biomass, Metab. Eng. 71 (2022) 42–61.
[5] N. Rhein-Knudsen, A.S. Meyer, Chemistry, gelation, and enzymatic modification of seaweed food hydrocolloids, Trends Food Sci. Technol. 109 (2021) 608–621.
[6] M. Schiøtt, M. Jönsson, L. Allahgholi, E.N. Karlsson, L. Lange, G.O. Hreggviðsson, A.S. Meyer, Biorefining of brown seaweeds catalyzed through innovative enzyme processes, Ind. Biotechnol. 20 (2024) 127–136.
[7] J.T. Kidgell, M. Magnusson, R. de Nys, C.R.K. Glasson, Ulvan: A systematic review of extraction, composition and function, Algal Res. 39 (2019) 101422.
[8] M.T. Ale, A.S. Meyer, Fucoidans from brown seaweeds: An update on structures, extraction techniques and use of enzymes as tools for structural elucidation, RSC Adv. 3 (2013) 8131–8141.
[9] J. Cai, A. Lovatelli, A. Stankus, X. Zhou, Seaweed revolution: where is the next milestone?, FAO Aquaculture Newsletter (2021) 13–16.
[10] I. Poblete-Castro, S.-L. Hoffmann, J. Becker, C. Wittmann, Cascaded valorization of seaweed using microbial cell factories, Curr. Opin. Biotechnol. 65 (2020) 102–113.
[11] E. Drula, M.-L. Garron, S. Dogan, V. Lombard, B. Henrissat, N. Terrapon, The carbohydrate-active enzyme database: functions and literature, Nucleic Acids Res 50 (2022) D571–D577.
[12] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (2014) 2114–2120.
[13] S. Andrews, FastQC A Quality Control tool for High Throughput Sequence Data [Online][WWW Document], URL ⟨https://www.bioinformatics.babraham.ac.uk/projects/download.html#fastqc⟩. Bioinformatics. Babraham. (Accessed 11.19. 20) (2010).
[14] D. Li, C.-M. Liu, R. Luo, K. Sadakane, T.-W. Lam, MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph, Bioinformatics 31 (2015) 1674–1676.
[15] H. Zhang, T. Yohe, L. Huang, S. Entwistle, P. Wu, Z. Yang, P.K. Busk, Y. Xu, Y. Yin, dbCAN2: a meta server for automated carbohydrate-active enzyme annotation, Nucleic Acids Res 46 (2018) W95–W101.
[16] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.
[17] F. Madeira, M. Pearce, A.R.N. Tivey, P. Basutkar, J. Lee, O. Edbali, N. Madhusoodanan, A. Kolesnikov, R. Lopez, Search and sequence analysis tools services from EMBL-EBI in 2022, Nucleic Acids Res 50 (2022) W276–W279.
[18] T. Paysan-Lafosse, M. Blum, S. Chuguransky, T. Grego, B.L. Pinto, G.A. Salazar, M. L. Bileschi, P. Bork, A. Bridge, L. Colwell, InterPro in 2022, Nucleic Acids Res 51 (2023) D418–D427.
[19] J. Zheng, Q. Ge, Y. Yan, X. Zhang, L. Huang, Y. Yin, dbCAN3: automated carbohydrate-active enzyme and substrate annotation, Nucleic Acids Res 51 (2023) W115–W121.
[20] K. Barrett, C.J. Hunt, L. Lange, A.S. Meyer, Conserved unique peptide patterns (CUPP) online platform: peptide-based functional annotation of carbohydrate active enzymes, Nucleic Acids Res 48 (2020) W110–W115.
[21] S.-J. Li, M. Hochstrasser, A new protease required for cell-cycle progression in yeast, Nature 398 (1999) 246–251.
[22] J. Hon, M. Marusiak, T. Martinek, A. Kunka, J. Zendulka, D. Bednar, J. Damborsky, SoluProt: prediction of soluble protein expression in *Escherichia coli*, Bioinformatics 37 (2021) 23–28.
[23] M. Vuillemin, A.S. Silchenko, H.T.T. Cao, M.S. Kokoulin, V.T.D. Trang, J. Holck, S. P. Ermakova, A.S. Meyer, M.D. Mikkelsen, Functional characterization of a new GH107 endo-α-(1, 4)-fucoidanase from the marine bacterium *Formosa haliotis*, Mar. Drugs 18 (2020) 562.
[24] G.L. Miller, Use of dinitrosalicylic acid reagent for determination of reducing sugar, Anal. Chem. 31 (1959) 426–428.
[25] T.T. Nguyen, M.D. Mikkelsen, V.H.N. Tran, V.T.D. Trang, N. Rhein-Knudsen, J. Holck, A.B. Rasin, H.T.T. Cao, T.T.T. Van, A.S. Meyer, Enzyme-assisted fucoidan extraction from brown macroalgae *Fucus distichus* subsp. *evanescens* and *Saccharina latissima*, Mar. Drugs 18 (2020) 296.
[26] V.H.N. Tran, T.T. Nguyen, S. Meier, J. Holck, H.T.T. Cao, T.T.T. Van, A.S. Meyer, M.D. Mikkelsen, The endo-α (1, 3)-fucoidanase Mef2 releases uniquely branched oligosaccharides from *Saccharina latissima* fucoidans, Mar. Drugs 20 (2022) 305.
[27] S.M. Alonso Villela, H. Kraiem, B. Bouhaouala-Zahar, C. Bideaux, C.A. Aceves Lara, L. Fillaudeau, A protocol for recombinant protein quantification by densitometry, Microbiologyopen 9 (2020) 1175–1182.
[28] H. Motejadded, J. Altenbuchner, Construction of a dual-tag system for gene expression, protein affinity purification and fusion protein processing, Biotechnol. Lett. 31 (2009) 543–549.
[29] V.H.N. Tran, M.D. Mikkelsen, H.B. Truong, H.N.M. Vo, T.D. Pham, H.T.T. Cao, T. T. Nguyen, A.S. Meyer, T.T.T. Thanh, T.T.T. Van, Structural characterization and cytotoxic activity evaluation of ulvan polysaccharides extracted from the green algae *Ulva papenfussii*, Mar. Drugs 21 (2023) 556.
[30] O.A. Podosokorskaya, V.V. Kadnikov, S.N. Gavrilov, A.V. Mardanov, A.Y. Merkel, O.V. Karnachuk, N.V. Ravin, E.A. Bonch-Osmolovskaya, I.V. Kublanov, Characterization of *Melioribacter roseus* gen. nov., sp. nov., a novel facultatively anaerobic thermophilic cellulolytic bacterium from the class *Ignavibacteria*, and a proposal of a novel bacterial phylum *Ignavibacteriae*, Environ. Microbiol 15 (2013) 1759–1771.
[31] V.H.N. Tran, V. Perna, M.D. Mikkelsen, T.T. Nguyen, V.T.D. Trang, A. Baum, H.T. T. Cao, T.T.T. Van, A.S. Meyer, A new FTIR assay for quantitative measurement of endo-fucoidanase activity, Enzym. Micro Technol. 158 (2022) 110035.
[32] A. Cumashi, N.A. Ushakova, M.E. Preobrazhenskaya, A. D'Incecco, A. Piccoli, L. Totani, N. Tinari, G.E. Morozevich, A.E. Berman, M.I. Bilan, A comparative study of the anti-inflammatory, anticoagulant, antiangiogenic, and antiadhesive activities of nine different fucoidans from brown seaweeds, Glycobiology 17 (2007) 541–552.
[33] G. Kopplin, A.M. Rokstad, H. Mélida, V. Bulone, G. Skjåk-Bræk, F.L. Aachmann, Structural characterization of fucoidan from *Laminaria hyperborea*: Assessment of coagulation and inflammatory properties and their structure–function relationship, ACS Appl. Bio Mater. 1 (2018) 1880–1892.
[34] C.D. Daub, B. Mabate, S. Malgas, B.I. Pletschke, Fucoidan from *Ecklonia maxima* is a powerful inhibitor of the diabetes-related enzyme, α-glucosidase, Int J. Biol. Macromol. 151 (2020) 412–420.
[35] M.D. Mikkelsen, V.H.N. Tran, S. Meier, T.T. Nguyen, V.H.N. Holck, H.T.T. Cao, T.T. T. Van, P.D. Thinh, A.S. Meyer, J.P. Morth, Structural and functional characterization of the novel endo-α (1, 4)-fucoidanase Mef1 from the marine bacterium *Muricauda eckloniae*, Acta Crystallogr D. Struct. Biol. 79 (2023) 1026–1043.
[36] Y. Qiu, H. Jiang, Y. Dong, Y. Wang, H.I. Hamouda, M.A. Balah, X. Mao, Expression and biochemical characterization of a novel fucoidanase from *Flavobacterium algicola* with the principal product of fucoidan-derived disaccharide, Foods 11 (2022) 1025.
[37] A.S. Silchenko, A.B. Rasin, M.I. Kusaykin, A.I. Kalinovsky, Z. Miansong, L. Changheng, O. Malyarenko, A.O. Zueva, T.N. Zvyagintseva, S.P. Ermakova, Structure, enzymatic transformation, anticancer activity of fucoidan and sulphated fucooligosaccharides from *Sargassum horneri*, Carbohydr. Polym. 175 (2017) 654–660.
[38] A.S. Silchenko, N.E. Ustyuzhanina, M.I. Kusaykin, V.B. Krylov, A.S. Shashkov, A. S. Dmitrenok, R.V. Usoltseva, A.O. Zueva, N.E. Nifantiev, T.N. Zvyagintseva, Expression and biochemical characterization and substrate specificity of the fucoidanase from *Formosa algae*, Glycobiology 27 (2017) 254–263.
[39] V.T.D. Trang, M.D. Mikkelsen, M. Vuillemin, S. Meier, H.T.T. Cao, J. Muschiol, V. Perna, T.T. Nguyen, V.H.N. Tran, J. Holck, The endo-α (1, 4) specific fucoidanase Fhf2 from *Formosa haliotis* releases highly sulfated fucoidan oligosaccharides, Front Plant Sci. 13 (2022) 823668.
[40] S. Colin, E. Deniaud, M. Jam, V. Descamps, Y. Chevolot, N. Kervarec, J.-C. Yvin, T. Barbeyron, G. Michel, B. Kloareg, Cloning and biochemical characterization of the fucanase FcnA: definition of a novel glycoside hydrolase family specific for sulfated fucans, Glycobiology 16 (2006) 1021–1032.
[41] C. Vickers, F. Liu, K. Abe, O. Salama-Alber, M. Jenkins, C.M.K. Springate, J. E. Burke, S.G. Withers, A.B. Boraston, Endo-fucoidan hydrolases from glycoside hydrolase family 107 (GH107) display structural and mechanistic similarities to α-l-fucosidases from GH29, J. Biol. Chem. 293 (2018) 18296–18308.
[42] M. Schultz-Johansen, M. Cueff, K. Hardouin, M. Jam, R. Larocque, M.A. Glaring, C. Hervé, M. Czjzek, P. Stougaard, Discovery and screening of novel metagenome-derived GH 107 enzymes targeting sulfated fucans from brown algae, FEBS J. 285 (2018) 4281–4295.
[43] A.O. Zueva, A.S. Silchenko, A.B. Rasin, M.I. Kusaykin, R.V. Usoltseva, A. I. Kalinovsky, V.V. Kurilenko, T.N. Zvyagintseva, P.D. Thinh, S.P. Ermakova, Expression and biochemical characterization of two recombinant fucoidanases

from the marine bacterium *Wenyingzhuangia fucanilytica* CZ1127T, Int J. Biol. Macromol. 164 (2020) 3025–3037.
[44] A.O. Zueva, A.S. Silchenko, A.B. Rasin, O.S. Malyarenko, M.I. Kusaykin, A. I. Kalinovsky, S.P. Ermakova, Production of high-and low-molecular weight fucoidan fragments with defined sulfation patterns and heightened in vitro anticancer activity against TNBC cells using novel endo-fucanases of the GH107 family, Carbohydr. Polym. 318 (2023) 121128.
[45] M.I. Bilan, A.A. Grachev, A.S. Shashkov, M. Kelly, C.J. Sanderson, N.E. Nifantiev, A.I. Usov, Further studies on the composition and structure of a fucoidan preparation from the brown alga *Saccharina latissima*, Carbohydr. Res 345 (2010) 2038–2047.
[46] P.Duc Thinh, R.V. Menshova, S.P. Ermakova, S.D. Anastyuk, B.M. Ly, T. N. Zvyagintseva, Structural characteristics and anticancer activity of fucoidan from the brown alga *Sargassum mcclurei*, Mar. Drugs 11 (2013) 1456–1476.
[47] M.I. Bilan, A.A. Grachev, A.S. Shashkov, N.E. Nifantiev, A.I. Usov, Structure of a fucoidan from the brown seaweed *Fucus serratus* L, Carbohydr. Res 341 (2006) 238–245.
[48] J. Shen, Y. Chang, Y. Zhang, X. Mei, C. Xue, Discovery and characterization of an endo-1, 3-fucanase from marine bacterium *Wenyingzhuangia fucanilytica*: A novel glycoside hydrolase family, Front Microbiol 11 (2020) 1674.
[49] C. He, H. Muramatsu, S. Kato, K. Ohnishi, Characterization of an *Alteromonas* long-type ulvan lyase involved in the degradation of ulvan extracted from *Ulva ohnoi*, Biosci. Biotechnol. Biochem 81 (2017) 2145–2151.
[50] M. Kopel, W. Helbert, Y. Belnik, V. Buravenkov, A. Herman, E. Banin, New family of ulvan lyases identified in three isolates from the Alteromonadales order, J. Biol. Chem. 291 (2016) 5871–5878.
[51] F. Xu, F. Dong, X.-H. Sun, H.-Y. Cao, H.-H. Fu, C.-Y. Li, X.-Y. Zhang, A. McMinn, Y.-Z. Zhang, P. Wang, Mechanistic insights into substrate recognition and catalysis of a new ulvan lyase of polysaccharide lyase family 24, Appl. Environ. Microbiol 87 (2021) e00412-21.
[52] J. Gao, C. Du, Y. Chi, S. Zuo, H. Ye, P. Wang, Cloning, expression, and characterization of a new PL25 family ulvan lyase from marine bacterium *Alteromonas* sp. A321, Mar. Drugs 17 (2019) 568.
[53] E. Foran, V. Buravenkov, M. Kopel, N. Mizrahi, S. Shoshani, W. Helbert, E. Banin, Functional characterization of a novel "ulvan utilization loci" found in *Alteromonas* sp. LOR genome, Algal Res 25 (2017) 39–46.
[54] T. Ulaganathan, M.T. Boniecki, E. Foran, V. Buravenkov, N. Mizrachi, E. Banin, W. Helbert, M. Cygler, New ulvan-degrading polysaccharide lyase family: structure and catalytic mechanism suggests convergent evolution of active site architecture, ACS Chem. Biol. 12 (2017) 1269–1280.
[55] D. Wang, Y. Li, L. Han, C. Yin, Y. Fu, Q. Zhang, X. Zhao, G. Li, F. Han, W. Yu, Biochemical properties of a new polysaccharide lyase family 25 ulvan lyase TsUly25B from marine bacterium *Thalassomonas* sp. LD5, Mar. Drugs 20 (2022) 168.
[56] A. Salinas, C.E. French, The enzymatic ulvan depolymerisation system from the alga-associated marine flavobacterium *Formosa agariphila*, Algal Res 27 (2017) 335–344.
[57] P.N. Collén, J.-F. Sassi, H. Rogniaux, H. Marfaing, W. Helbert, Ulvan lyases isolated from the flavobacteria *Persicivirga ulvanivorans* are the first members of a new polysaccharide lyase family, J. Biol. Chem. 286 (2011) 42063–42071.
[58] T. Ulaganathan, E. Banin, W. Helbert, M. Cygler, Structural and functional characterization of PL28 family ulvan lyase NLR48 from *Nonlabens ulvanivorans*, J. Biol. Chem. 293 (2018) 11564–11573.
[59] V.R. Konasani, C. Jin, N.G. Karlsson, E. Albers, A novel ulvan lyase family with broad-spectrum activity from the ulvan utilisation loci of *Formosa agariphila* KMM 3901, Sci. Rep. 8 (2018) 14713.
[60] L. Reisky, A. Prechoux, M.-K. Zühlke, M. Bäumgen, C.S. Robb, N. Gerlach, T. Roret, C. Stanetty, R. Larocque, G. Michel, A marine bacterial enzymatic cascade degrades the algal polysaccharide ulvan, Nat. Chem. Biol. 15 (2019) 803–812.