The cell wall active mannuronan C5-epimerases in the model brown alga Ectocarpus: from gene context to recombinant protein

Richard Fischl, Kresten Bertelsen, Fanny Gaillard, Susana Coelho, Markus Klinger, Catherine Boyen, Mirjam Czjzek, Cécile Hervé

The cell wall active mannuronan C5-epimerases in the model brown alga *Ectocarpus*: from gene context to recombinant protein

Richard Fischl[1,2], Kresten Bertelsen[3], Fanny Gaillard[4], Susana Coelho[1,2], Gurvan Michel[1,2], Markus Klinger[3,5], Catherine Boyen[1,2], Mirjam Czjzek[1,2,6], Cécile Hervé[1,2,6]

[1]*Sorbonne Universités, UPMC Univ Paris 06, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, Roscoff, France;* [2]*CNRS, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, Roscoff, France;* [3]*DuPont Nutrition Biosciences ApS, Aarhus, Denmark,* [4]*FR2424, Station Biologique de Roscoff, CS 90074, Roscoff, France*

[5]Actual address: Markus Klinger, Novozymes A/S, Bagsværd, Denmark

[6]To whom correspondence should be addressed: e-mail: cecile.herve@sb-roscoff.fr; mirjam.czjzek@sb-roscoff.fr Tel: +33 298 29 23 23; Fax: +33 298 29 23 24;

Keywords: alginate/brown algae/*Ectocarpus*/mannuronan C5-epimerase

Running title: mannuronan C5-epimerases from *Ectocarpus*

List of Supplementary Data: Table S1 to S2 and Fig.S1 to S8

Abstract (182 words)

Mannuronan C5-epimerases (ManC5-Es) catalyse in brown algae the remodelling of alginate, a major cell wall component which is involved in many biological functions in these organisms. ManC5-Es are present as large multigenic families in brown algae, likely indicating functional specificities and specialisations. ManC5-Es control the distribution pattern of (1-4) linked β-D-mannuronic acid (M) and α-L-guluronic acid (G) residues in alginates, giving rise to widely different polysaccharide compositions and sequences, depending on tissue, season, age or algal species. As such they are also a source of powerful new tools for the biotechnological and enzymatic processing of alginates, to match the growing interest for food hydrocolloids and in biomedical and nanotechnological applications. We report here the first heterologous production of a ManC5-E of brown algal origin that is successfully refolded in an active form. The activity was measured by ^{1}H-NMR and by an indirect enzymatic assay using a known bacterial alginate lyase. The transcript expression as a function of the developmental program of the brown alga *Ectocarpus,* together with the bioinformatic analyses of the corresponding gene context of this multigenic family, are also presented.

Introduction

Alginate is a major cell wall constituent of brown algae (Phaeophyceae) also synthesized by some bacteria of the *Azotobacter* and *Pseudomonas* genera. This polysaccharide is of important commercial value with an annual worldwide production of about 45,000 tons, extracted from various types of brown seaweeds including species mostly from the orders of Laminariales and Fucales (Smidsrod and Draget 1996). In many different applications, alginates act as stabilizing, thickening, and/or gelling agents. The main uses are food additives. Alginates are increasingly used in cosmetics, health care and pharmaceuticals (Zhao et al. 2012) and lately received a renewed interest for their use in nanotechnology (Kovalenko et al. 2011) and in fuel and ethanol productions (Enquist-Newman et al. 2014, Wargacki et al. 2012, Wei et al. 2013).

Alginate is a linear copolymer of 1,4-linked uronic acid epimers differing only at C5: the β-D-mannuronic acid (M) and the α-L-guluronic acid (G). These residues are arranged in homopolymeric regions of MM- and GG-blocks, interspaced with random arrangements of both monomers (MG-blocks), each of which have different conformational preferences and, hence, different physico-chemical properties (Donati et al. 2005, Heyraud et al. 1996). While the MM-block regions do not associate in the presence of divalent cations, GG-blocks regions will form ‘egg-box’ junctions with calcium (Grant et al. 1973), bridging two antiparallel chains and leading to gel formation. The mechanical strength of an alginate gel thus correlates with the proportion of GG-blocks (Figure 1). Alternating structures take up rather disordered conformations, yielding more flexible gels.

G-rich alginates with a M/G ratio ≤ 1.0 constitute around 25% of the total alginate market, with a yearly growth of about 10%, mainly in the food market. This type of alginate is also widely used in the pharmaceutical area (Draget and Taylor 2011). As most seaweed

species have a M/G ratio of 1.2-1.6 in alginate (Haug et al. 1974, Smidsrod and Draget 1996), brown seaweeds with a naturally low M/G ratio of 0.7-0.5 are in high demand. Alginate with an M/G ratio of 0.4 is exclusively found in stipes of *Laminaria hyperborea* and *Lessonia trabeculata* (Smidsrod and Draget 1996). The M/G ratio of alginate from the whole *L. hyperborea* plant is 1.0 (Draget et al. 2005) and separation of fronds from stipes is laborious. Thus, novel ways to ensure an accessible and efficient production of alginate with low M/G ratios are desirable. Although chemical epimerisation can be performed on the polysaccharide (Pawar and Edgar 2012), the high temperature required for this process leads to undesirable depolymerisation. At lower temperature actual processes use chemical catalysts that are difficult to remove from polyanionic material with strong cationic complexing properties, such as alginates.

The alginate biosynthesis pathway is yet not fully resolved in brown algae but genes homologous to their bacterial counterparts have been identified (Michel et al. 2010). The final steps were proposed to be the production of polymannuronate and the subsequent epimerization on the polymer of the M residues into G residues; the latter being catalysed by putative mannuronan C5-epimerases (ManC5-Es) resembling bacterial ManC5-Es (Nyvall et al. 2003). The biochemical characterisation of native and hybrid ManC5-E of bacterial sources has shown their abilities to generate distinct alginate structures (Wolfram et al. 2014, Bjerkan et al. 2004, Hartmann et al. 2002). The presence of a large number of ManC5-E genes in brown algae also suggests that their gene products might yield polymers with different patterns of G and M distributions. Given that alginates in brown algae affect the texture of the cell wall in a similar manner as do pectins in the cell walls of land plants, this multigene situation is reminiscent of pectin methyl-esterases (PMEs) (Micheli 2001). By analogy to PMEs, which can yield partially methylesterified pectins with either random or blockwise distribution of the methyl groups, the different isoforms of ManC5-E might yield

alginates with different patterns of GG-block distribution. Moreover, the diversity of enzymes present in brown algae to remodel the algal cell wall polysaccharide, is much larger than that found in bacterial alginate synthetizing systems (Bjerkan et al. 2004). It can thus be assumed that these algal enzymes could be used to produce alternative “tailor-made alginates” with novel physico-chemical properties

Heterologous expression of algal genes is notoriously difficult and has presented a bottleneck in projects aiming at characterising enzymes of macroalgal origins (Groisillier et al. 2010, Kurland and Gallant 1996). Recent access to the first genome sequence of a brown alga (Cock et al. 2010) allowed the identification of all the genes shaping the ManC5-E family in *Ectocarpus*. By providing multiple targets, it opened a new door toward the expression of recombinant ManC5-Es. We took advantage of a medium throughput strategy we recently developed for protein expression (Groisillier et al. 2010), to heterologously express the ManC5-Es from *Ectocarpus*. Our work has generated the first recombinant and active ManC5-E from a brown alga. An additional comprehensive analysis of the gene context and transcript expression during the life cycle of *Ectocarpus* has provided new insights into the possible roles of ManC5-E genes in the brown algae.

Results

Ectocarpus contains a multigenic family of putative ManC5-Es

The functional annotation of the genome of *Ectocarpus* previously indicated the occurrence of 28 genes encoding putative mannuronan C5-epimerases (ManC5-Es) (Michel et al. 2010). A more detailed analysis led to the identification of additional genes that shape a final multigenic family of 31 members that accounts for 0.19% of the total predicted gene number in *Ectocarpus*. The gene names and the corresponding acronyms (which use a 5-characters code) are listed in Supplementary Table S1. For phylogenetic analyses, 10 genes were discarded due to doubtful predicted gene structure. The phylogenetic tree indicated that at least four groups of ManC5-Es can be identified in *Ectocarpus* (Figure 2A). Among them one group contains modular proteins that are predicted to contain additional modules appended to the catalytic ManC5-E domain: a schematic diagram showing the protein structure of the eight candidates is shown in Figure 2B and Supplementary Figure S1. Five of the eight encoded proteins are predicted to contain a signal peptide targeting the protein to the cell surface. In addition to the catalytic ManC5-E domain, three main supplementary domains are found: the Wall Sensing Component (WSC), the AGP-protein core domain and a module of unknown function. The WSC domains are known to be involved in cell-wall sensing in yeast. They also shape a large multigenic family in the phylogenetic-distant brown algae where they are expected to play similar functions (Cock et al. 2010, Hervé et al. 2016). Arabinogalactan-proteins (AGPs) are highly glycosylated, hydroxyproline-rich proteins found at the cell surface of plants, but they have also been recently investigated in brown algae (Hervé et al. 2016). In both systems they play key roles in developmental processes. The module of unknown function (MUF) was identified manually and up to now we could not detect any significant homology to other proteins or domains deposited in public databases. The MUF

sequence is highly conserved (Supplementary Figure S2) and is often found as repetitive motifs in the ManC5-E sequences under concern (i.e. MEP3, MEP1, MEP4) (Figure 2B). The protein motif consists of a sequence of 56 amino acids in average, and displays a particular amino acid composition (cysteine-, glycine- and serine-rich) (Supplementary Figure S2). Such particular amino acid patterns are often unstructured, i.e. linkers, or display propensity to self-assemble and we thus speculate that these domains might play a role in structuring these proteins within the cell wall.

The ManC5-E family includes genes that are differentially regulated during the life-cycle of *Ectocarpus*

Thirty of the 31 predicted ManC5-E gene family members are transcribed during the life cycle of *Ectocarpus* (http://bioinformatics.psb.ugent.be/orcae/overview/EctsiV2). The remaining gene that exhibits no expression in any of the conditions tested (MEP30), contains a large insertion of the *Ectocarpus* virus (EsV1-163) which may alter gene expression. The life cycle of *Ectocarpus* involves an alternation between two morphologically distinct generations, the sporophyte and the gametophyte (Cock et al. 2015). Microarray expression data (Coelho et al. 2011) obtained for a subset of the ManC5-E genes revealed an interesting expression pattern during the life cycle generation stages. While the majority of these twelve genes were indistinctively expressed in both generations of the life cycle, four ManC5-E genes (MEP6, MEP26, MEP27, MEP28) were preferentially expressed in the gametophyte generation (Figure 3, *t*-test $P<0.05$), suggesting a role for these genes in the development of the gametophyte. These four genes belong to different ManC5-E phylogenetic clusters. Therefore, within a common cluster of ManC5-E genes, some can be either mostly expressed in the gametophytes, or showing no difference in their expression pattern. We conclude that the expression of some of the ManC-Es is differentially regulated during the life cycle of

Ectocarpus and that the ManC5-Es do not cluster in the phylogenetic tree in relation to their expression patterns.

Gene structure analyses

Gene exon-intron structures analyses of the ManC5-E family were conducted. Figure 2C shows the mosaic structure of all the ManC5-E genes included in the tree. The average gene length of ManC5-E genes is larger than the total average gene length in *Ectocarpus* (Table 1A). In addition most ManC5-E genes possess more than 9 introns, with the average intron number being nearly twice that of the other *Ectocarpus* genes. The most common system for getting recombinant proteins used bacterial strains derived from *Escherichia coli*. The ManC5-E genes from *Ectocarpus* were examined regarding the codon usage preference: three of the 5 rare codons for *E. coli* show critical relative adaptiveness values (Table 1B). All these characteristics are likely to render heterologous expression of the ManC5-E genes difficult and we payed special attention during the process (described below) to overcome this problem.

Expression, purification and refolding of the mannuronan C5-epimerase MEP13

Although different strategies were used (e.g. *E. coli* strains, mammalian cells, in vitro assays), our previous attempts during the past fifteen years to express two putative ManC5-Es from *Laminaria digitata* (Nyvall et al. 2003), were proved unsuccessful. The genome sequence of *Ectocarpus* offers the opportunity to have access to a higher number of genes, and potentially proteins. In order to increase our chances of getting soluble expression of active and recombinant ManC5-Es, a medium throughput strategy was applied (Groisillier et al. 2010). The 31 putative ManC5-E genes were analysed and the corresponding encoded catalytic domains tentatively expressed. The medium throughput strategy rapidly showed that

obtaining soluble heterologous protein expression of this eukaryotic protein family in bacterial expression systems was challenging, if not impossible. Attempts to express some ManC5-E genes in soluble form in other organisms than *E. coli* (i.e. the bacterium *Lactococcus lactis*, the yeasts *Pichia pastoris* and *Hansenula polymorpha*, and insect cells) were not successful. The targeted catalytic module of MEP13 (MEP13-C5) was insolubly but highly expressed in *E. coli* (Supplementary Figure S3A), and our efforts in getting an active recombinant ManC5-E were subsequently focused on this specific target. In order to overcome a codon usage preference in *E. coli* strains, we used a codon-optimized sequence of *MEP13-C5*, the latter one being designated *MEP13-C5*$_{syn}$ (Table 1B and Supplementary Figure S4). None of the different media or strains used improved the soluble expression of MEP13-C5, and a high level of insoluble protein was obtained in all cases (Supplementary Figure S3A,C). Nevertheless, usage of different complex media such as TB resulted in a higher cell density and consequently in a higher amount of inclusion bodies. In order to increase our chances of getting recombinant proteins, we included two additional *E. coli* strains developed for the expression of toxic protein or rare codon containing genes. Usage of *E. coli* Rosetta-gami2(DE3)pLysS produced larger and more significant inclusion bodies than *E. coli* BL21CodonPlus(DE3)RIPL. This discrepancy might be due to the different growth rates of these strains. Rosetta-gami2(DE3)pLysS grows more slowly than BL21CodonPlus(DE3)RIPL and incubation overnight seems to be more suitable to obtain the targeted protein expression with this strain.

The heterologous expression of MEP13-C5 in *E. coli* Rosetta-gami2(DE3)pLysS yielded about 15 mg/L of insoluble enzyme. The His-tag containing protein was then successfully refolded using an on-column refolding procedure. The analysis of the SDS-PAGE gels and corresponding western-blots (Supplementary Figure S3A,B) showed that after eluting the refolded protein, the samples were essentially pure, also indicating that only large

aggregates, eluted in the flow-through, or refolded protein were present, yielding a final quantity of 6.5 mg of active MEP13-C5 enzyme. The identity of the protein was confirmed by peptide mass fingerprinting (Supplementary Figure S5). The DLS spectra of the protein solutions showed high monodispersity of the refolded samples, as compared to the unfolded protein samples (Supplementary Figure S6).

The refolded MEP13-C5 shows a mannuronan C5-epimerase activity

The activity of the refolded MEP13-C5 was monitored by measuring the change of M/G ratio in two different alginate solutions (having an initial M/G ratio of 1.5 and 1.7, respectively) using ^{1}H-NMR. In both cases a decrease of the M/G ratio indicated the action of MEP13-C5, epimerizing some mannuronate units into guluronate (Figure 4A). The change of blockiness could be analysed by integrating and measuring the relative surface of the NMR-peaks that appear in the range of 4.4.40 to 5.15 ppm (Figure 4B,C, Supplementary Figure S7), following the method described previously (Grasdalen et al. 1981, Grasdalen 1983, Hartmann et al. 2002). The increase of G-block structure (Figure 4B) is observed with both alginates but is only significant for the sample with a M/G-ratio of 1.5. Figure 4C illustrates that for both alginate samples, the repeated MM structures disappear while GG and GGG structures significantly increase. However, these data do not yet allow clear and full identification of which patterns are preferentially recognized and which are produced by the enzyme.

An additional activity assay, using a GG-specific alginate lyase (AlyA1; Thomas et al. 2013), was set up that allowed indirect evidence of ManC5-E activity. Indeed, under identical conditions, the MEP13-C5 pre-treated alginate samples produced up to 1.5-fold higher abundance of unsaturated sugar units upon incubation with the GG-specific alginate lyase than the initial untreated alginate samples (Figure 5).

Discussion

Functional validation of a putative mannuronan C5-epimerase gene from brown alga

To date, fully biochemically characterized mannuronan C5-epimerases have only been reported for enzymes of bacterial origin (Buchinger et al. 2014, Campa et al. 2004, Hartmann et al. 2002, Rozeboom et al. 2008, Wolfram et al. 2014). Nyvall et al. (2003) were the first to isolate full length cDNAs from a brown alga (*Laminaria digitata*) displaying similarity with bacterial mannuronan C5-epimerase genes. They proposed that these genes encoded functional ManC5-Es involved in the terminal step of alginate biosynthesis in brown algae. Analysis of the genomic sequence of *Ectocarpus* provided candidate genes for the whole alginate biosynthetic pathway, including the putative ManC5-E genes (Michel et al, 2010). Nonetheless, the predicted function of these brown algal genes has never been biochemically proven until now. Our study reports, for the first time, the heterologous production, refolding and purification of an active ManC5-E, namely MEP13-C5, of brown algal origin. At least for the here reported gene, this result thus validates the hypothesis formulated by Nyvall et al. (2003) twelve years ago. To our knowledge this is also the first characterized enzyme specific of the biosynthesis of a cell wall polysaccharide from brown algae.

Fine-tuning of alginate structure

In the context of industrial applications, based on various alginates with different physico-chemical properties, the use of specific ManC5-Es that could modulate rheological properties in a defined manner would be of precious biotechnological advantage. Moreover, since significant variations impact the hydrocolloid properties of alginates as a function of seasonal factors, tissue or species, having at hand specific ManC5-Es that could adjust the rheological properties after or during the industrial processing and extraction is also of large interest. In

the past, several ManC5-Es of bacterial origin have been biochemically characterized (Ertesvåg 2015, Wolfram et al. 2014, Campa et al. 2004, Hartmann et al. 2002) and even engineered to produce tailored alginates (Tøndervik et al. 2013, Bjerkan et al. 2004). To be useful in existing processes of commercial alginate production, the alginate ManC5-Es need to meet a number of requirements. For example, it is important that the enzymes are selected to work on polymeric substrate, not only on shorter oligosaccharides. This is the case for MEP13-C5, similarly to those of bacterial origin (Campa et al. 2004, Hartmann et al. 2002).

In our study, low but significant epimerization activity of the algal MEP13-C5 could be measured by ^{1}H-NMR and by the indirect measure of the activity of a GG-specific alginate lyase. However at this stage, it remains difficult to compare these preliminary analyses of activity to the mode of action of the bacterial enzymes. All alginate-producing bacteria expressed a periplasmic epimerase, AlgG, and the putative mannuronan C5-epimerases from *Ectocarpus* show highest sequence identity to these bacterial epimerases, with around 32% identity of MEP13-C5, for example, with AlgG from *P. aeruginosa* (Figure 6). Nevertheless, while all characterized bacterial enzymes of the AlgG-type have been reported to be rather calcium-independent (Ertesvåg 2015, Wolfram et al. 2014), the activity of MEP13-C5 could only be measured in presence of Ca^{2+}. This situation is reminiscent of the AlgE-type enzymes produced by bacteria of the *Azotobacter* genera, which are calcium-dependent mannuronan C5-epimerases (Wolfram et al. 2014, Hartmann et al. 2002). However the sequences of the algal mannuronan C5-epimerases are much more distantly related to these enzymes, with identities often less than 17%. Despite the lack of sequence similarity and different characteristics, the enzymes of bacterial origins that have been biochemically characterized all display a processive mode of action and common active site residues (Wolfram et al. 2014, Campa et al. 2004, Hartmann et al. 2002), but differences appear with respect to the preference of formation of MG or GG-blocks. Our algal enzyme seems to display a somewhat

different specificity than the bacterial enzymes of the AlgG-type, in that it appears to recognize the presence of GG-blocks to preferentially increase the GG block sizes. This observation might be deduced from the NMR study, where the enzyme increased GG-block structures in a substrate already rich in GG-block motifs, but this preference is subtle and not confirmed by the alginate lyase-based assays. A more exhaustive library of standards should therefore be used in future biochemical studies to determine the alginate patterns, which are more susceptible to be epimerized. If verified, this subtle difference in mode of action is not surprising in the view of the differences that can be identified at the primary sequence level, as illustrated in Figure 6, although the characteristic active site signature 'DPHD' and the regular features specific of the β-helix fold align well. In the structural sequence alignment one can notice that several residues directly adjacent to the active site, and identified as being involved in substrate binding and processing in AlgE4 and AlgG (Rozeboom et al. 2008, Wolfram et al. 2014), have a different nature in MEP13-C5 (blue triangles in Figure 6). Moreover, single residue insertions or deletions (green triangles in Figure 6) or large inserted or deleted regions within the catalytic domain (blue dashed lines in Figure 6) may influence the positioning of important residues and influence substrate specificity, which may differ among the different algal ManC-5Es.

Studies of the bacterial multigenic C5-epimerase family from *Azetobacter vinelandii* (Hartman et al. 2002, Buchinger et al. 2014) have highlighted the subtle differences in substrate specificities of different AlgE-type enzymes. In the biological context of brown algal cell wall, the presence of a large multigenic ManC5-E family thus points towards the fact that variability in substrate specificity is to be expected. Interestingly, besides the rather well conserved catalytic domain of about 400 amino acids, the modular organisation is highly variable among ManC5-Es from *Ectocarpus*. The functional role of additional R-modules in bacterial C5-epimerases has only recently been analysed (Buchinger et al. 2014) and showed

that they do influence alginate-pattern recognition and mode of action. Our study provides a first experimental indication that given their different modular organisation and fine structure at the primary sequence level, the mannuronan C5-Epimerases in brown algae are likely to recognise and/or generate distinct motifs, which are important for the physiological functions and the regulation of alginate synthesis and remodelling in the cell wall. Future structural and biochemical analyses will be necessary to elucidate in detail the potential of fine-tuning alginate structures by these algal ManC5-Es.

The roles of ManC5-Es in cell-wall remodelling in brown algae

By acting on a major component of the brown algal cell walls, the ManC5-Es are likely to be of importance in many physiological responses, and various studies are supporting the idea. For instance, in *Laminaria digitata*, some ManC5-E genes are differentially expressed during stress responses such as elicitation and protoplast generation (Tonon et al. 2008), and upon seasonal variations (Nyvall et al. 2003). Probably one of the most important (but the less explored) field of action of ManC5-Es, is during developmental processes. Variations of the alginate structure upon tissues are known in Laminariales species, and our analysis of microarray expression data highlighted the possible role for ManC5-E during the development of gametophyte versus sporophytes. Interestingly, almost all of the ManC5-E genes are constitutively expressed, strongly suggesting that ManC5-Es are major actors during *Ectocarpus* development.

In land plants the homogalacturonan motifs of pectins are secreted into the wall as highly methylesterified forms, subsequently modified by the multigene PME family. The PMEs can either act randomly, promoting the action of cell wall hydrolases and contributing to cell wall loosening, or linearly, giving rise to blocks of demethylesterified motifs, which contribute to cell-wall stiffening through calcium chelation (Micheli 2001). In brown algae,

the functional analogy of ManC5-Es to PMEs is tempting but challenges our view of their modes of action. In contrast to PMEs in plants, our present knowledge on their epimerase activity indicates a 'dead-end' reaction, leading to cell wall strengthening only. The *Ectocarpus* genome does not encode any known alginate lyases, but the presence of enzymes possibly acting on unsaturated oligosaccharides (i.e. GH88; Michel et al. 2010) suggests that new families of polysaccharide lyases are present, or alternatively that some ManC5-Es have the dual activity, although this is not the case of MEP13-C5 (i.e. no lyase activity; data not shown). Also, the analysis of the ManC5-E gene structure in *Ectocarpus* indicates additional modules such as the WSC domains and the AGP-core motifs, which might putatively influence the ManC5-E activity (Hervé et al. 2016). The knowledge of the mechanisms involved in cell wall remodelling in brown algae is still in its infancy, partly because expressing algal enzymes into heterologous systems remains challenging. Nevertheless, our success on this first heterologous and refolded algal C5-epimerase, as well as the access to recent genome sequences (Cock et al. 2010, Ye et al. 2015) should help future analysis on deciphering cell wall metabolism in brown algae.

Material and Methods

Identification and bioinformatical analyses of the ManC5-E sequences

Ectocarpus contains 31 genes of ManC5-Es, available at the genome website of *Ectocarpus siliculosus* at https://bioinformatics.psb.ugent.be/gdb/ectocarpus and listed in Supplementary Table S1. For phylogenetic analyses, 10 genes were discarded due to doubtful predicted gene structure. The remaining deduced amino acid sequences encoding the ManC5-E catalytic domains were aligned using the MAFFT program at http://bioinformatics.uams.edu/mafft/ (Supplementary Fig. S8). Unrooted maximum likelihood phylogenetic trees were conducted using MEGA6 (Tamura et al. 2007). The best-fit model among 48 different amino acid substitution models was the LG model. The tree with the highest log likelihood (-8620.9504) is shown in Figure 1A. Initial trees for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.1283)). All positions containing gaps and missing data were eliminated. Bacterial sequences were used as an outgroup (Michel et al. 2010). There were a total of 24 sequences and 253 positions in the final data set. The robustness of the branches was evaluated using 100 bootstrap replicates. The modularity of each deduced amino acid sequence was further examined using Blast queries against the UniProt database and domain searches against the InterPro server. Additional refinements were performed manually by examination of the individual HCA plots. Signal peptides and transmembrane domains were predicted using the SignalP and TMHMM servers, respectively. Predictions of chimeric AGP-motifs were performed as described (Hervé et al. 2016). The cluster of ManC5-Es containing

modular proteins is described in more details in Figure 1B. For each ManC5-E included in the phylogenetic tree, the gene structure of the full-length sequence is indicated (Figure 1C).

Microarray analysis of the ManC5-E mRNA abundances

We explored the microarray dataset that was obtained in Coelho et al. (2011) to carry out the analysis of changes in mRNA abundances of ManC5-E genes in the gametophyte and sporophyte generations (ArrayExpress accession no. E-MTAB-485).

Medium throughput screening for cloning and protein expression of the ManC5-Es

The ManC5-E genes were tentatively cloned and heterologously expressed using a medium throughput strategy in a microtiter-plate format, as previously described (Groisillier et al. 2010). In short, the open reading frames, trimmed to the catalytic domains only, were amplified by PCR and using primers incorporating specific restriction sites compatible with our ligation strategies (Supplementary Table S2). From 31 starting genes, 12 were discarded due to either doubtful predicted gene structure or incompatibility with our cloning strategy. The amplification products were cloned into the pFO4 vector (a vector modified from pET15b from Novagen, to be compatible with the BamHI/EcoRI ligation strategy) and the pGEX (GE Healthcare Life Science) vectors, encoding an N-terminal His_6-tag and a GST-tag, respectively. Recombinant plasmids were used to transform *E. coli* strains DH5α (Promega). The validated plasmids were used to transform appropriate *E. coli* strains BL21(DE3) and BL21 (Novagen), respectively. Screening of protein expression was performed at a small-scale range using a 24-well plate format and the auto-inducible ZYP5052 media as described (Groisillier et al. 2010). The plates were centrifuged at 1,200 g for 20 min at 4°C and the pelleted bacteria were resuspended in a lysis buffer containing 50 mM Tris pH 8, 300 mM NaCl, 1 mg/ml lysozyme, 0.1 mg/ml DNAse and an anti-proteases cocktail. The cell lysates

were further centrifuged at 12,000 g for 20 min. The supernatants contained the soluble expressed fractions. The remaining cell pellets, which contained the insoluble expressed fractions, were extracted with 6 M urea. All fractions were analysed by SDS-PAGE. All further steps were performed on the single insolubly but highly expressed target module of MEP13.

Codon optimisation and scale-up expression and purification of MEP13-C5

Codon-optimisation of the encoded catalytic domain of MEP13 (named MEP13-C5) was performed by GeneArt within the framework PolyModE (www.polymode.eu). The sequence was extended with terminal restriction sites for BamHI and EcoRI to be compatible with our cloning strategy (Supplementary Table S2). The codon-optimised sequence was delivered as being cloned into the shuttle vector pMA. To differentiate between the cDNA and the codon-optimized sequence, the latter one is designated *MEP13-C5*$_{syn}$, where *syn* indicates the synthetic and codon-optimised sequence. *MEP13-C5*$_{syn}$ was excised from the pMA vector using the BamHI and EcoRI restriction enzymes, purified, and further cloned into the pFO4 vector. In order to increase our chances of getting recombinant proteins, the codon-optimised sequence was also cloned into the pQE-80L vector (Qiagen). In addition to encode a His_6-tag, this vector contains a PT5 promoter and two *lac*-operators to allow tight regulation of basal level expression, which makes this vector suitable to express toxic genes. The codon-optimised sequence was amplified by PCR from the pMA vector using specific primers (Supplementary Table S2), purified and digested by BamHI and SacI before ligation into the pQE-80L vector. The recombinant plasmids were used to transform *E. coli* strains DH5α. The plasmid DNA of the candidate clones was purified using Wizard® Plus SV Minipreps DNA Purification Systems (Promega) and the correctness of cloning was confirmed by DNA sequencing.

In order to increase our chances of getting recombinant proteins, we decided to use diverse expression strains and included two additional *E. coli* strains developed for the expression of toxic or rare codon containing genes. We therefore transformed *E. coli* BL21(DE3), *E. coli* Rosetta-gami2(DE3)pLysS and *E. coli* BL21CodonPlus(DE3)RIPL cells (all from Novagen) as expression strains. The transformed strains were cultivated in different complex media, including LB and TB (Sambrook and Russel, 2001; Korf et al., 2005). The culture volume ranged from 200 mL to 1 L. To trigger formation of insoluble proteins, induction (1 mM IPTG final) has been carried out permanently at 37 °C overnight at 180 rpm. After 24 hrs, the cells were analysed by microscopy for the presence of inclusion bodies, which contain the insoluble proteins. Cells were harvested by centrifugation at 6,000 g for 30 min and mechanically lysed using a French press.

The cell lysate was clarified by ultracentrifugation at 100,000 g during 1 hour. The insoluble protein present in the cell pellet was purified and refolded on-column in a single step. For this purpose the pellet was solubilised using a tissue homogeniser and a refolding buffer containing 25 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 mM 2-mercaptoethanol, 10% glycerol, 8 M urea, 40 mM imidazole and 4 mM $CaCl_2$. The solubilised inclusion body-solution was centrifuged at 100,000 g for 30 min and the supernatant loaded onto a 5 mL HisTrap HP column charged with 100 mM $NiSO_4$. The refolding on-column was carried out by a slow gradient over two hours against the refolding buffer and using a buffer containing 25 mM Tris-HCl pH 7.8, 200 mM NaCl, 3 mM reduced glutathione, 0.3 mM oxidised glutathione, 10% glycerol and 4 mM $CaCl_2$. The bound proteins were finally eluted in a gradient with the elution buffer containing 25 mM Tris-HCl pH 7.8, 200 mM NaCl, 3 mM reduced glutathione, 0.3 mM oxidised glutathione, 10% glycerol, 500 mM imidazole and 4 mM $CaCl_2$. Protein purification was carried out on an Äkta Avant Chromatographic System from GE Healthcare. Columns and resins used were from GE Healthcare. All fractions

collected were analysed by SDS-PAGE. The targeted eluates containing the refolded MEP13-C5 were pooled and the volume was reduced by ultrafiltration to approximately 3.5 mL of a final protein concentration of 1.94 g/L. In the aim to separate the putative different folding intermediates, a second purification step was performed by size exclusion chromatography using a Superdex 200 column from GE Healthcare. Equilibration and running buffers were those that were used in the final step of the refolding procedure.

Western-blotting

The sequences of the catalytic domains of all the brown algal ManC5-Es were aligned with the sequence of the catalytic domain of AlgE4 from *Azetobacter vinelandii* in order to identify specific sequence stretches that, when reported onto the structural coordinates of AlgE4, appeared to be located at the surface of the protein. Two stretches (CDGRAKNEMGECRMDIINS and NHGIIASKRCNNVKIF) were identified and polyclonal antibodies raised in mice against these peptides were purchased from Eurogentech. The antibodies were tested by ELISA and further used to verify heterologous expression of algal ManC5-Es (Supplementary Figure S3).

Analysis of protein identity by peptide mass fingerprinting

The refolded and purified MEP13-C5 was analysis on a SDS-PAGE gel. Bands at 50 kDa, which correspond to the expected size of the protein, were excised and analysed by peptide mass fingerprinting (PMF). Coomassie blue-stained protein slices were cut into small pieces, washed with distilled water and de-stained using acetonitrile. The cysteine residues were reduced by 100 µl of 10 mM DTT at 56°C and alkylated by 150 µl of 55 mM iodoacetamide at room temperature. Iodoacetamide solution was replaced by 100 µl of 100 mM NH_4HCO_3 and gel dehydration was achieved with acetonitrile. After evaporation in Speed Vac

(Thermo), proteins were digested overnight at 37°C by a solution containing 0.9 µg of a modified bovine trypsin (Promega, Madison, WI, USA) prepared in 25 mM NH_4HCO_3. Finally, a double extraction was performed, first with 1% (v/v) trifluoro acetic acid solution, and subsequently with 100% (v/v) ACN. The resulting peptide mixture was extracted and dried under vacuum and re-suspended in 50% ACN, 0.1% TFA solution. PMF by MALDI-TOF MS was performed using a Voyager DE-STR MALDI-TOF mass spectrometer (Applied Biosystems). Each pooled tryptic digest (1 µL) was mixed with an equal volume of CHCA matrix (Sigma; 10 mg/mL in 50% ACN, 0.1% TFA) and spotted onto the MALDI target. Spectra were acquired in positive ion reflector mode under 20 kV accelerating voltage and a mass range of 450– 5000 Da. Internal calibration was performed using trypsin autolysis fragments at m/z 1433.70, 2163.05, and 2289.1. The samples resulting from attempts to produce soluble proteins mostly contained proteins from the *E. coli* host strains, while the samples containing insoluble proteins released eighteen peptides specific of MEP13-$C5_{syn}$, covering 52% of the protein sequence (Supplementary Figure S5), and there was no contamination with *E. coli* proteins. This result demonstrated that MEP13-C5 was indeed highly expressed in an insoluble form, whereas each attempt to produce the enzyme in its soluble form failed.

Dynamic light scattering (DLS)

DLS was performed to control the mono-dispersity of the refolded protein. 100 µL of sample proteins (unfolded; refolded fractions) were centrifuged at 20,817 rcf on an Eppendorf Centrifuge 5417R for 10 minutes. 60 µL of supernatant was transferred into a quartz cuvette and placed into a Zetasizer Nano S from Malvern. Measurement followed at 20°C during 20 minutes.

Enzyme activity assays

All alginate samples were provided by DuPont. 100 µg of sample proteins were incubated in presence of 0.5% sodium alginate (M/G ratio of 1.5 and 1.7) in a buffer containing 20 mM MOPS pH7.0, 4 mM $CaCl_2$, 100 mM NaCl, using a final volume of 1 mL. The reaction was incubated overnight at 30°C. To avoid signal-interference in subsequent NMR measurements, it was necessary to remove the $CaCl_2$ after the enzymatic reaction. Removal of $CaCl_2$ started with acidification of the samples by mixing them with equal volume (1 mL) of acidic water (300 µL of concentrated H_2SO_4 in 100 mL of water, pH 1.4) and incubation for one hour at room temperature on a shaker. The samples were centrifuged at 141 *g* for 5 minutes and the top phase was removed. The samples were washed with an equal volume of water (1 mL) and incubated in the same way as before. Each washing step with acidic water and normal water was done three times. In the final step, the top phase (water) was removed and the samples were neutralised with Na_2CO_3 (0.2 g in 10 g water) to a pH of 7.0. MEP13-C5 untreated alginates were submitted to the same procedure as the MEP13-C5 treated samples. The samples were lyophilized prior to measurements. Immediately before performing NMR measurements the samples were solubilized with D_2O. NMR-spectra were measured without depolymerisation of the samples.

Non-lyophilized samples, without calcium removal, after treatment with refolded MEP13-C5 were also analysed in a second, indirect, activity assay using the action of a G-specific alginate lyase known as AlyA1 from *Zobellia galactanivorans* (Thomas et al. 2013). The MEP13-C5 treated alginate samples were incubated at 30°C with 100 µg of AlyA1 in a solution containing 100 mM Tris-HCl buffer at pH 7.0. Alginate lyase activity was assayed by measuring the increase in absorbance at 235 nm (A_{235}) of the reaction products (unsaturated uronates) for 5 min in a 1 cm quartz cuvette containing 0.5 ml of reaction mixture in a

thermostated spectrophotometer. The difference in the AlyA1 activity between the MEP13-C5-treated and -untreated alginates was assigned to the MEP13-C5 activity.

NMR analysis

The untreated and enzyme-treated alginate samples were prepared as described above. All NMR experiments were performed on a Bruker Avance III spectrometer (Karlsruhe, Germany) equipped with a double-tuned direct-detection broadband 5 mm probe (and deuterium lock) and a 14.1 T magnet corresponding to a ^{1}H resonance frequencies of 600 MHz. Each spectrum was run with a pulse angle of 30º, 2.7 second acquisition and one second delay between each scan with a total of 64 scans. The data was zero filled to 65536 and a Gaussian apodiaztion function was added prior to the Fourier transform yielding a spectrum with a 20 ppm spectral width. Further processing such as deconvolution, was done using python scripting in combination with the nmrglue python library (Helmus et al. 2013). The chemical shifts are expressed in parts per million (ppm) downfield from the signal for 3-(trimethylsilyl)propanesulfonate.

For the alginate substrate with a M/G ratio of 1.7, four control samples were measured, giving an average M/G ratio of 1.68 and a standard deviation (sd) of 0.02. When the enzyme was added (4 replicates) the average M/G ratio decreased to 1.51 with an sd of 0.01. For the alginate substrate with a M/G ratio of 1.5, four control samples were measured, giving an average M/G ratio of 1.46 and a standard deviation (sd) of 0.01. When the enzyme was added (4 replicates) the average M/G ratio decreased to 1.33 with a sd of 0.02. The measure of block distribution in terms of “GG”, “MM”, ”GM = MG ”, ”GGG”, “MGM”, “GGM = MGG”, was undertaken by deconvolution or integration of the peaks in the region between 4.40 and 5.15 ppm following standard protocols (Grasdalen et al. 1981, Grasdalen 1983, Hartmann et al. 2002).

Acknowledgments

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement No. 222628 (www.polymode.eu). This work was also supported by the Region Bretagne and the Centre National de la Recherche Scientifique (CNRS).

Abbreviations

ManC5-E, mannuronan C5-epmimerase

MEP, mannuronan epimerase

M, β-D-mannuronic acid

G, α-L-guluronic acid

PME, pectin methylesterase

AGP, arabinogalactan protein

WSC, Wall Sensing Component

Tris, Tris(hydroxymethyl)aminomethane or 2-Amino-2-hydroxymethyl-propane-1,3-diol

MOPS, 3-(N-morpholino)propane-1-sulfonic acid

ACN, acetonitrile

TFA, trifluoroacetic acid

CHCA, α-cyano-4-hydroxycinnamic acid

DTT, dithiothreitol

IPTG, isopropyl β-D-1-thiogalactopyranoside

Sd, standard deviation

HCA, hydrophobic cluster analysis

PMF, peptide mass fingerprinting

MS, mass spectrometry

MALDI-TOF, matrix-assisted laser desorption/ionisation – time of flight spectrometry

ELISA, enzyme-linked immunosorbent assay

SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis

DLS, dynamic light scattering

NMR, nuclear magnetic resonance

Ppm, parts per million

Legends to figures

Fig. 1. Schematic representation of the effect of ManC5-Es on alginate structures. Alginate in brown algae is first formed as a polysaccharide chain containing mannuronic acid residues only; these are subsequently transformed by the ManC5-E into guluronic acid residues, generating distinct patterns arranged in regions of MM-, GG- and MG-blocks (top). Patterns containing large stretches of adjacent guluronic acid residues (GG-blocks) form structured inter-chain associations in the presence of Ca^{2+} ions (central). These inter-chain junctions have the so-called 'egg-box' conformation (bottom left) and are responsible for the gelling properties of alginate and cell-wall strengthening.

Fig. 2. Bioinformatical analyses of the ManC5-E family from *Ectocarpus*. (**A**) Phylogenetic tree. Numbers indicate the bootstrap values in the maximum likelihood analysis. The sequences belonging to the four main clusters are marked by distinctive symbols (black diamond, open circle, black square and black triangle). (**B**) A focus is made on the ManC5-E sequences marked by an open circle. Putative structure of the encoded proteins is shown on the right side with the signal peptide (black box), the catalytic domain ManC5-E, the WSC domains, the AGP protein-core featuring the glycan decoration (green "wattle blossom" shape), the module of unknown function (yellow oval) and a transmembrane domain (red box). Lengths are relatively scaled. (**C**) Corresponding gene intron-exon structure of the ManC5-Es shown in the tree.

Fig. 3. Microarray analysis of the abundance of ManC5-E transcripts in *Ectocarpus* sporophytes versus gametophytes, obtained from Coelho et al. (2011). The belonging to the clusters described in the phylogenetic tree shown in Figure 2A is indicated (bottom). Data are

means of three independent biological replicates ± SE. Asterisks indicate significant differences (*t*-test, P<0.05).

Fig. 4. Assessment of mannuronan C5-epimerase activity of MEP13-C5 by ^{1}H-NMR. (**A**) The average M/G ratio of two different industrial alginates (M/G ratios of 1.5 and 1.7) before and after treatment with the recombinant and refolded MEP13-C5 are displayed as histograms. (**B**) Average G-block length of the corresponding samples. (**C**) The relative occurrences of the distinct block structures of the alginate samples are shown. The scale of the x-axis corresponds to block structures derived from NMR data by integration and deconvolution of individual NMR peaks in the region of 4.40 to 5.15 ppm. In each case, data are means of four independent replicates ± SD. Asterisks indicate significant differences (*t*-test, P<0.01).

Fig. 5. Assessment of mannuronan C5-epimerase activity of MEP13-C5 by analysis of the degradation pattern using a GG-specific bacterial alginate lyase. The production of unsaturated sugar units produced by the action of AlyA1 (a strictly GG-specific alginate lyase from *Z. galactanivorans*) was followed during 5 minutes by photospectrometry at a wavelength of 235 nm on the two alginate samples (M/G ratios of 1.5 and 1.7). In each case, data are means of three independent replicates ± SD.

Fig. 6. Extract of a structural sequence alignment of selected mannuronan C5-epimerases from *Ectocarpus,* including MEP13, against the sequence and structural features of the bacterial AlgG (pdbcode 4NK6). One sequence from each cluster of the phylogenetic tree shown in Figure 2A (namely MEP5, MEP13, MEP25 and MEP29) were used and the ManC5-E catalytic sites were aligned against the sequence of AlgG (pdbcode 4NK6) from *Pseudomonas aeruginosa*. The alignment was edited using the program ESPRIPT (Robert and

Gouet, 2014). The colour code of the amino acid residues is according to their biochemical character. The secondary structure assignments and residue numbering above the sequences correspond to those of 4NK6. Regions displaying high sequence similarity are boxed. The red triangles below the sequences indicate the catalytic residues His319 and Asp320, the blue triangles below the sequences indicate two positions in this highly conserved pattern that are modified in the algal ManC5-Es, the green triangles highlight a single residue insertions/deletions close to the active site and the blue dashed lines highlight larger insertion/deletion between the algal clades and the bacterial sequence.

Tables

Table 1. Statistics for several features of the ManC5-E genes, including MEP13-C5. (**A**) Structure of genes compared to the *Ectocarpus* genome, (**B**) relative adaptiveness values (%) for 5 rare codons for *E. coli*, compared to the nuclear codon usage of *E. coli*

A

	genome (average)	ManC5-Es (average)	MEP13
Gene length (bp)	6,974	9,712	5,494
CDS length (bp)	1,607	2,083	1,899
Number of introns per gene	7.1	13.7	10.0

B

	AGA	AGG	ATA	CTA	CCC
E. coli (genome)	43	23	45	16	43
ManC5-Es (average)	40	60	9	11	79
MEP13-C5	33	100	4	0	91
MEP13-C5$_{syn}$	0	0	0	0	0

References

Bjerkan TM, Lillehov BE, Strand WI, Skjak-Braek G, Valla S, Ertesvag H. 2004. Construction and analyses of hybrid *Azotobacter vinelandii* mannuronan C-5 epimerases with new epimerization pattern characteristics. *Biochem J*. 381:813-821.

Buchinger E, Knudsen DH, Behrens MA, Pedersen JS, Aarstad OA, Tondervik A, Valla S, Skjak-Braek G, Wimmer R, Aachmann FL. 2014. Structural and functional characterization of the R-modules in alginate C-5 epimerases AlgE4 and AlgE6 from Azotobacter vinelandii. *J Biol Chem*. 289:31382-31396.

Campa C, Holtan S, Nilsen N, Bjerkan TM, Stokke BT, Skjak-Braek G. 2004. Biochemical analysis of the processive mechanism for epimerization of alginate by mannuronan C-5 epimerase AlgE4. *Biochem J*. 381:155-164.

Cock JM, Sterck L, Rouze P, Scornet D, Allen AE, Amoutzias G, Anthouard V, Artiguenave F, Aury JM, Badger JH, *et al.* 2010. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature*. 465:617-621.

Cock JM, Godfroy O, Strittmatter M, Scornet D, Uji T, Farnham G, Peters AF, Coelho SM. 2015. Emergence of *Ectocarpus* as a model system to study the evolution of complex multicellularity in the brown algae. In: Ruiz-Trillo I, Nedelcu AM editors. *Evolutionary Transitions to Multicellular Life*: Springer Netherlands. p. 153-162.

Coelho SM, Godfroy O, Arun A, Le Corguillé G, Peters AF, Cock JM. 2011. OUROBOROS is a master regulator of the gametophyte to sporophyte life cycle transition in the brown alga Ectocarpus. *PNAS*. 108:11518-11523.

Donati I, Holtan S, Morch YA, Borgogna M, Dentini M, Skjak-Braek G. 2005. New hypothesis on the role of alternating sequences in calcium-alginate gels. *Biomacromolecules*. 6:1031-1040.

Draget KI, Smidsrød O, Skjåk-Bræk G. 2005. Alginates from Algae. In: Steinbüchel A, Rhee SK editors. *Polysaccharides and Polyamides in the Food Industry: Properties, Production, and Patents.*: Wiley-VCH Verlag GmbH & Co. KGaA.

Draget KI, Taylor C. 2011. Chemical, physical and biological properties of alginates and their biomedical implications. *Food Hydrocolloids* 25:251-256.

Enquist-Newman M, Faust AME, Bravo DD, Santos CNS, Raisner RM, Hanel A, Sarvabhowman P, Le C, Regitsky DD, Cooper SR, *et al.* 2014. Efficient ethanol production from brown macroalgae sugars by a synthetic yeast platform. *Nature*. 505:239-243.

Ertesvåg H. 2015. Alginate-modifying enzymes: biological roles and biotechnological uses. *Frontiers in Microbiology* 6: 523.

Grant GT, Morris ER, Rees DA, Smith PJC, D T. 1973. Biological interactions between polysaccharides and divalent cations : the egg-box model. *FEBS letters* 32:195-198.

Grasdalen H, Larsen B, Smisrod O. 1981. ^{13}C-n.m.r. studies of monomeric composition and sequence in alginate. *Carbohydrate Research*. 89:179-191.

Grasdalen H. 1983. High-field, ^{1}H-N.M.R. spectroscopy of alginate: sequential structure and linkage conformations. *Carbohydr. Res.* 118, 255–260.

Groisillier A, Hervé C, Jeudy A, Rebuffet E, Pluchon PF, Chevolot Y, Flament D, Geslin C, Morgado IM, Power D, *et al.* 2010. MARINE-EXPRESS: taking advantage of high throughput cloning and expression strategies for the post-genomic analysis of marine organisms. *Microb Cell Fact*. 9:45.

Hartmann M, Holm OB, Johansen GA, Skjak-Braek G, Stokke BT. 2002. Mode of action of recombinant Azotobacter vinelandii mannuronan C-5 epimerases AlgE2 and AlgE4. *Biopolymers*. 63:77-88.

Haug A, Larsen B, Smidsrod O. 1974. Uronic acid sequence in alginate from different sources. *Carbohydrate Research*. 32:217-225.

Hervé C, Siméon A, Jam M, Cassin A, Johnson KL, ASalmean AA, Willats WGT, Doblin MS, Bacic A, Kloareg B. 2016. Arabinogalactan-proteins have deep roots in Eukaryotes: identification of genes and epitopes in brown algae and their role in *Fucus serratus* embryo development. *New Phytologist*. 209: 1428-1441.

Helmus J J, Jaroniec C P. 2013. Nmrglue: an open source Python package for the analysis of multidimensional NMR data. *Journal of Biomolecular NMR*. 55:355-67.

Heyraud A, Gey C, Leonard C, Rochas C, Girond S, Kloareg B. 1996. NMR spectroscopy analysis of oligoguluronates and oligomannuronates prepared by acid or enzymatic hydrolysis of homopolymeric blocks of alginic acid. Application to the determination of the substrate specificity of *Haliotis tuberculata* alginate lyase. *Carbohydrate Research*. 289:11-23.

Korf U, Kohl T, van der Zandt H, Zahn R, Schleeger S, Ueberle B, Wandschneider S, Bechtel S, Schnölzer M, Poustka A, *et al.* 2005. Large-scale protein expression for proteome research. *Proteomics* 5: 3571–3580.

Kovalenko I, Zdyrko B, Magasinski A, Hertzberg B, Milicev Z, Burtovyy R, Luzinov I, Yushin G. 2011. A Major Constituent of Brown Algae for Use in High-Capacity Li-Ion Batteries. *Science*. 334:75-79.

Kurland C, Gallant J. 1996. Errors of heterologous protein expression. *Curr Opin Biotechnol*. 7:489-493.

Michel G, Tonon T, Scornet D, Cock JM, Kloareg B. 2010. The cell wall polysaccharide metabolism of the brown alga *Ectocarpus siliculosus*. Insights into the evolution of extracellular matrix polysaccharides in Eukaryotes. *New Phytol*. 188:82-97.

Micheli F. 2001. Pectin methylesterases: cell wall enzymes with important roles in plant physiology. *Trends in Plant Science*. 6:414-419.

Nyvall P, Corre E, Boisset C, Barbeyron T, Rousvoal S, Scornet D, Kloareg B, Boyen C. 2003. Characterization of mannuronan C-5-epimerase genes from the brown alga *Laminaria digitata*. *Plant Physiol*. 133:726-735.

Pawar SN, Edgar KJ. 2012. Alginate derivatization: a review of chemistry, properties and applications. *Biomaterials*. 33:3279-3305.

Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucl. Acids Res.* 42: W320-W324.

Rozeboom HJ, Bjerkan TM, Kalk KH, Ertesvag H, Holtan S, Aachmann FL, Valla S, Dijkstra BW. 2008. Structural and mutational characterization of the catalytic A-module of the mannuronan C-5-epimerase AlgE4 from Azotobacter vinelandii. *J Biol Chem*. 283:23819-23828.

Sambrook J, Russell D. 2001. Molecular Cloning: A laboratory manual. Woodbury, USA, Cold Spring Harbor Laboratory Press; pp999.

Smidsrod O, Draget KL. 1996. Chemistry and physical properties of alginates. *Carbohyd. in Europ.* . 14:6-13.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*. 24:1596-1599.

Thomas F, Lundqvist LCE, Jam M, Jeudy A, Barbeyron T, Sandström C, Michel G, Czjzek M. 2013. Comparative characterization of two marine alginate lyases from Zobellia galactanivorans reveals distinct modes of action and exquisite adaptation to their natural substrate. *Journal of Biological Chemistry*.

Tøndervik A, Klinkenberg G, Aachmann FL, Svanem BI, Ertesvåg H, Ellingsen TE, Valla S, Skjåk-Bræk G, Sletta H. 2013. Mannuronan C-5 epimerases suited for tailoring of specific alginate structures obtained by high-throughput screening of an epimerase mutant library. *Biomacromolecules*. 14(8):2657-66.

Tonon T, Rousvoal S, Roeder V, Boyen C. 2008. Expression profiling of the mannuronan C5-epimerase multigenic family in the brown alga *Laminaria digitata* (Phaeophyceae) under biotic stress conditions. *Journal of Phycology*. 44:1250-1256.

Wargacki AJ, Leonard E, Win MN, Regitsky DD, Santos CN, Kim PB, Cooper SR, Raisner RM, Herman A, Sivitz AB, *et al.* 2012. An engineered microbial platform for direct biofuel production from brown macroalgae. *Science*. 335:308-313.

Wei N, Quarterman J, Jin Y-S. 2013. Marine macroalgae: an untapped resource for producing fuels and chemicals. *Trends in Biotechnology*. 31:70-77.

Wolfram F, Kitova EN, Robinson H, Walvoort MT, Codee JD, Klassen JS, Howell PL. 2014. Catalytic mechanism and mode of action of the periplasmic alginate epimerase AlgG. *J Biol Chem*. 289:6006-6019.

Ye N, Zhang X, Miao M, Fan X, Zheng Y, Xu D, Wang J, Zhou L, Wang D, Gao Y, *et al.* 2015. *Saccharina* genomes provide novel insight into kelp biology. *Nat Commun*. 6.

Zhao D, Zhuo R-X, Cheng S-X. 2012. Alginate modified nanostructured calcium carbonate with enhanced delivery efficiency for gene and drug delivery. *Molecular BioSystems*. 8:753-759.