



Aalborg Universitet

AALBORG UNIVERSITY  
DENMARK

## Proteomic characterization of pilot scale hot-water extracts from the industrial carrageenan red seaweed *Eucheuma denticulatum*

Gregersen Echers, Simon; Pertseva, Margarita; Marcatili, Paolo; Holdt, Susan Løvstad; Jacobsen, Charlotte; García Moreno, Pedro Jesús; Hansen, Egon Bech; Overgaard, Michael Toft

*Published in:*  
Algal Research

*DOI (link to publication from Publisher):*  
[10.1016/j.algal.2021.102619](https://doi.org/10.1016/j.algal.2021.102619)

*Creative Commons License*  
CC BY 4.0

*Publication date:*  
2022

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

### *Citation for published version (APA):*

Gregersen Echers, S., Pertseva, M., Marcatili, P., Holdt, S. L., Jacobsen, C., García Moreno, P. J., Hansen, E. B., & Overgaard, M. T. (2022). Proteomic characterization of pilot scale hot-water extracts from the industrial carrageenan red seaweed *Eucheuma denticulatum*. *Algal Research*, 62, Article 102619. <https://doi.org/10.1016/j.algal.2021.102619>

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -



## Proteomic characterization of pilot scale hot-water extracts from the industrial carrageenan red seaweed *Eucheuma denticulatum*

Simon Gregersen<sup>a,\*</sup>, Margarita Pertseva<sup>b,c</sup>, Paolo Marcatili<sup>b</sup>, Susan Løvstad Holdt<sup>d</sup>,  
Charlotte Jacobsen<sup>d</sup>, Pedro J. García-Moreno<sup>d,e</sup>, Egon Bech Hansen<sup>d</sup>, Michael Toft Overgaard<sup>a</sup>

<sup>a</sup> Department of Chemistry and Bioscience, Aalborg University, Denmark

<sup>b</sup> Department of Bio and Health Informatics, Technical University of Denmark, Denmark

<sup>c</sup> Department of Biosystems Science and Engineering, ETH, Zurich, Switzerland

<sup>d</sup> National Food Institute, Technical University of Denmark, Denmark

<sup>e</sup> Department of Chemical Engineering, University of Granada, Spain

### ARTICLE INFO

#### Keywords:

*Eucheuma denticulatum*  
Hot-water protein extraction  
Quantitative proteomics  
de novo quantitative transcriptomics  
Bioinformatics  
Subcellular localization

### ABSTRACT

Seaweeds have a long history as a resource for polysaccharides/hydrocolloids extraction for use in the food industry due to their functionality as stabilizing agents. In addition to the carbohydrate content, seaweeds also contains a significant amount of protein, which may find application in food and feed. Here, we present a novel combination of transcriptomics, proteomics, and bioinformatics to determine the protein composition in two pilot-scale extracts from *Eucheuma denticulatum* (Spinosum) obtained via hot-water extraction. Although the quality of extracted protein appeared quite poor based on SDS-PAGE analysis, extracts were characterized by qualitative and quantitative proteomics using LC-MS/MS and a *de-novo* transcriptome assembly for construction of a suitable protein database. A novel concept of length-normalization for relative quantification of sub-optimal protein extracts with partial, non-specific digestion is introduced and validated against conventional methods for relative quantification of proteins. Despite a limited number of protein identifications due to poor protein quality, our data suggest that the majority of quantified protein in the extracts (>75%) is constituted by merely three previously uncharacterized proteins. Putative subcellular localization for the quantified proteins was determined by bioinformatic prediction using several predictors, and by correlating with the expected copy number from the transcriptome analysis, we find that the extracts appear highly enriched in extracellular proteins. This implies that the extraction method used predominantly extracts extracellular proteins, and thus appear ineffective for cellular disruption and subsequent release of intracellular proteins. Nevertheless, the highly abundant proteins may be potential substrates for targeted hydrolysis and release of bioactive peptides. Ultimately, this study highlight the potential of quantitative proteomics for characterization of alternative protein sources intended for use in foods and evaluating protein extraction process efficiency through novel combinations with bioinformatic analysis.

### 1. Introduction

Seaweeds are known to contain numerous compounds of interest, such as polysaccharides, proteins and other compounds with health beneficial properties such as anti-inflammatory, anti-oxidant, and anti-cancer [1,2]. The industry to produce hydrocolloids from seaweed is well established, and the hydrocolloids are used as e.g. stabilizing agents in toothpaste, canned whipped cream, and as meat glue. The production of red carrageenan accounts for 54,000 ton/year and constitutes the

majority of the total hydrocolloids sold worldwide (also incl. alginate and agar). Carrageenan is extracted from 212,000 ton dried seaweed, and brings in a value of 530 million USD [3]. *Eucheuma denticulatum* (commercially referred to as “*spinosum*”) is among the most cultivated and harvested red seaweed species for the carrageenan industry. However, at present carrageenan is extracted in a process, which targets this as the only compound of interest while protein and other compounds are not exploited. The most common industrial method to obtain carrageenan from *E. denticulatum* uses hot water at high pH (>12). If other

\* Corresponding author at: Frederik Bajers Vej 7H, DK-9220 Aalborg, Denmark.  
E-mail address: [sgr@bio.aau.dk](mailto:sgr@bio.aau.dk) (S. Gregersen).

<https://doi.org/10.1016/j.algal.2021.102619>

Received 7 July 2021; Received in revised form 8 December 2021; Accepted 21 December 2021

Available online 5 January 2022

2211-9264/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

compounds, such as protein, could be isolated prior to or as part of the industrial hot water extraction without compromising the yield and quality of carrageenan, this could be of interest, as the amount of available biomass is substantial. Proteins from *E. denticulatum* were shown to constitute only 3.8% of dry biomass, but were of high quality with respect to their amino acid profile and nutritional quality [4]. The protein is also comparable to beef in regards to content of branched chained amino acids (i.e. leucine, isoleucine, and valine), which are of specific interest due to their diverse metabolic and physiological functions [5].

In addition to the general health benefits from ingestion [6,7], seaweed may also be a source of bioactive peptides that could exhibit a direct biological purpose or be utilized as functional food ingredients. Such peptides may be released through bio-processing of proteins extracts using e.g. enzymatic hydrolysis or fermentation [8]. In the past decade, peptides derived from seaweed proteins with e.g. renin-inhibitory [9], ACE-inhibitory [10], antioxidant [11], and antidiabetic [12,13] activities have been identified. Common for all bioactive peptides is that they were identified in enzymatic hydrolysates by a non-targeted trial-and-error approach. This methodology, commonly employed in the food industry, requires numerous often costly and time-demanding steps of hydrolysis, separation, isolation, identification, and finally in vitro or in vivo verification of activity. In contrast, an orthogonal approach utilizing bioinformatic prediction of bioactive peptides, is gathering increased attention [14]. This method reduces cost and work load tremendously, and allows for targeted peptide release by enzymatic hydrolysis. With recent advances in bioinformatic prediction of peptide functionality [15–19], and the growing availability of peptide databases [20–23], the primary prerequisite for such analysis is the availability of protein sequences and quantitative information on protein composition. Recently, quantitative proteomics for identification of abundant proteins, followed by bioinformatic prediction, has been employed for identification of highly functional emulsifier [24] and antioxidant [19] peptides from potato. LC-MS/MS and bioinformatic data analysis has also been applied to identify probable emulsifier and antioxidant peptides in hydrolysates from fish processing side streams following processing [25–27]. Nevertheless, proteomic quantification of the biomass or protein substrate would in such a scheme be a prerequisite for rational design of an enzymatic process and ultimately to maximize the yield of bioactive peptides. Here, we present a proteomic characterization of two industrially relevant, pilot-scale extracts from *E. denticulatum* obtained by hot-water extraction. Protein identification is based on a de novo transcriptome assembly for creating a suitable reference protein database. Furthermore, we present a novel approach for quantifying proteins based on non-tryptic peptides, and correlate protein abundance with quantitative transcriptomics. Using bioinformatic prediction of protein subcellular origin, we are able to determine enrichment of certain protein classes in the extracts. Through such analysis, it is possible to evaluate not only protein composition but also protein quality with added depth, thereby facilitating a more data-driven and informed decision making process for application and/or further processing of such extracts in their current state.

## 2. Materials and methods

### 2.1. Materials

Two *Eucheuma denticulatum* protein extracts (obtained using hot-water extraction at near-neutral pH) were supplied as pilot-scale trial extracts by the global food ingredient provider CP Kelco. Protein extract A was obtained by dispersing the raw seaweed in deionized water (pH adjusted to 8.9 with sodium carbonate) and applying continuous stirring at 95 °C for 5 h (20 kg seaweed in 1000 L water). The slurry was subsequently filtered in a Büchner funnel followed by diafiltration using a 300 kDa MWCO membrane. The retentate was washed with three volumes of 0.9% sodium chloride in deionized water, diafiltrated, and all

permeates were subsequently pooled. The pooled permeate was concentrated using a 1 kDa MWCO membrane, and the retentate lyophilized to yield the final protein extract A (155 g total product, CP Kelco supplied information). Protein extract B was obtained similarly to extract A, but with stirring at 90 °C for 16 h before filtering, diafiltration, concentration, and lyophilization (approximately same yield, CP Kelco supplied information). Furthermore, the lyophilized retentate was dissolved in deionized water, the pH was adjusted to 2.9 with nitric acid, and the mixture was stirred at room temperature for 1 h. Precipitated protein was isolated by centrifugation and washed twice with isopropanol before air drying and lyophilization to yield the final protein extract B (6.7 g total product, CP Kelco supplied information). The total protein content of protein extracts A and B (by Kjeldahl-N) was 7.1% and 70% (w/w), respectively, using a nitrogen-to-protein conversion factor of 6.25 (CP Kelco supplied information). The remaining content of samples was expected to be predominantly residual carrageenan co-extracted with the protein. All chemicals used were of analytical grade.

### 2.2. Total soluble protein

Protein extracts A and B were solubilized to an estimated protein concentration of 2 mg/mL in ddH<sub>2</sub>O and in 200 mM NH<sub>4</sub>HCO<sub>3</sub> with 0.2% SDS (pH 9.5) for maximal solubilization compatible with the Qubit protein assay. Following solvent addition, samples were thoroughly mixed, vortexed for 30 s, sonicated for 30 min, and left overnight on a Stuart SRT6 roller mixer (Cole-Parmer, UK). The next day, samples were sonicated for 30 min, left on a roller mixer for 60 min, and centrifuged at 3095 RCF (ambient temperature) for 10 min in a 5810 R centrifuge (Eppendorf, Germany), to precipitate solids prior to aliquoting the supernatant. The total soluble protein content of the samples in both solvents, was quantified using Qubit protein assay (Thermo Scientific, Germany) according to the manufacturer guidelines.

### 2.3. 1D-SDS-PAGE and in-gel digestion

Protein extracts A and B were solubilized as described above. SDS-PAGE analysis was performed on precast 4–20% gradient gels (Gen-Script, USA) in a Tris-MOPS buffered system under reducing conditions according to manufacturer guidelines. Briefly, 20 µg protein/peptide was mixed with reducing (final DTT concentration 50 mM) SDS-PAGE sample buffer and subsequently denatured at 95 °C for 5 min prior to loading on the gel. As molecular weight marker, PIERCE Unstained Protein MW Marker P/N 26610 (ThermoFisher Scientific, USA) was used. Protein visualization was achieved by using Coomassie Brilliant Blue G250 staining (Sigma-Aldrich, Germany) and imaging with a ChemDoc MP Imaging System (Bio-Rad, USA).

Proteins were in-gel digested as previously described [24]. Briefly, each gel lane from the gradient gel was excised with a scalpel and divided into 6 fractions guided by the MW marker (<14 kDa; 14–25 kDa; 25–45 kDa; 45–66 kDa; 66–116 kDa; >116 kDa). Individual fractions were cut into 1 × 1 mm pieces before being subjected to washing, reduction with DTT, Cys alkylation with iodoacetamide, and digestion with sequencing grade modified trypsin (Promega, Madison, WI, USA). Following digestion, peptides were extracted, dried down by SpeedVac, and suspended in 0.1% (v/v) formic acid (FA), 2% acetonitrile (ACN) (v/v). Next, peptides were desalted using StageTips [28,29], dried down by SpeedVac, and finally suspended in 0.1% (v/v) FA, 2% ACN (v/v) for LC-MS/MS analysis.

### 2.4. De novo transcriptome assembly

The transcriptome of *E. denticulatum* was downloaded from the NCBI SRA database (<https://www.ncbi.nlm.nih.gov/sra/SRX2653634>). The raw reads were preprocessed by Trimmomatic software to filter short sequences (less than 36 bp) and to trim low-quality ends [30]. Processed reads were then assembled de novo into contigs using Trinity with

default parameters [31]. Overall, 9458 contigs were assembled with an average length of 1021 bp.

## 2.5. Transcript annotation, abundance estimation and protein database construction

The potential protein-coding sequences were predicted by TransDecoder based on the length of open reading frames and nucleotide composition [31]. Candidate sequences were annotated by BlastP and BlastX search against SwissProt database [32] with the cutoff *E*-value of 1E-5 as well as by HMMER [33] search against Pfam database [34,35]. An alignment *E*-value of 1E-5 means that a homology hit has a 1 in 100,000 probability of occurring by chance alone, therefore we chose this threshold to get only high-quality homologous proteins hits.

The abundance of the transcripts (transcripts per megabase, TPM) was calculated by re-aligning reads to the assembled contigs using the RNA-Seq by Expectation-Maximization (RSEM) estimation method included in Trinity software [31]. Obtained transcript abundance matrix was joined with Blastp-annotated transcripts to attain a list of highly expressed proteins.

## 2.6. Prediction of subcellular localization and signal peptides

For all proteins in the final database, subcellular localization was predicted using deepLoc [36] through the freely available web-tool (<http://www.cbs.dtu.dk/services/DeepLoc/index.php>). All searches were performed using the BLOSUM62 protein encoding to achieve a probability based subcellular localization for use in enrichment analysis on both transcriptome and protein level. For the most abundantly quantified proteins, predicted subcellular localization was validated using the plant-specific versions of YLoc [37], LocTree3 [38], and SCLpred [39].

For selected proteins, the presence of signal peptides was predicted using SignalP (v.5.0) [40] through the freely available web-tool (<http://www.cbs.dtu.dk/services/SignalP-5.0/index.php>). Predicted signal peptides were disregarded when calculating the experimental sequence coverage from proteomics analysis, as the coverage of the mature protein form is of higher interest than the pre-protein.

## 2.7. LC-MS/MS analysis

Tryptic peptides were analyzed by an automated LC-ESI-MS/MS consisting of an EASY-nLC system (Thermo Scientific, Bremen, Germany) on-line coupled to a Q Exactive mass spectrometer (Thermo Scientific) via a Nanospray Flex ion source (Thermo Scientific), as previously reported [24]. Separation of peptides was achieved by use of an Acclaim Pepmap RSLC analytical column (C18, 100 Å, 75 μm × 50 cm (Thermo Scientific)). Instrumental settings, solvents, flows, gradient, and acquisition method were identical to what was described previously [24].

## 2.8. Proteomics data analysis

Protein identification and quantification was performed using MaxQuant 1.6.10.43. [41,42] using the *de-novo* protein database, assembled from the transcriptomic analysis. Initially, standard settings were employed using specific digestion (Trypsin/P, 2 missed cleavages allowed, minimum length 7 AAs) and false discovery rate (FDR) of 1% on both peptide and protein level. FDR was controlled using reverse decoy sequences and common contaminants were included. Protein quantification was obtained with including both unique and razor peptides. Samples were analyzed as six fractions with boosted identification rates by matching between runs, second peptides, and dependent peptides enabled. The iBAQ algorithm [43] was used for relative in-sample protein quantification. iBAQ intensities were normalized to the sum of all iBAQ intensities after removal of reverse hits and

contaminants, to obtain the relative iBAQ (riBAQ), as previously described [24,44].

MS-data were furthermore analyzed using both semi-specific (tryptic *N*- or *C*-terminus) and nonspecific (no terminal restrictions) digestion in MaxQuant. All settings were maintained except for applying unspecific digestion with peptide length restrictions from 4 to 65 AAs. Additional unspecific searches with peptide and protein level FDR of 5% and 10% as well as semi-specific searches with peptide and protein level FDR of 5% was conducted to increase identification rates and sequence coverage for comparison and data quality assessment.

Relative quantification with iBAQ employs strict tryptic restrictions to peptide termini and consequently, this type of quantifications is not possible for semi-specific and unspecific searches. In order to compare and evaluate the semi-specific and nonspecific results, we introduced two additional quasi-quantitative relative metrics: i) relative intensity,  $I_{rel}$  and ii) length-normalized relative intensity,  $I_{rel}^L$ . They were defined as:

$$I_{rel}(n) = \frac{I_n}{\sum_{n=1}^p I_n} * 100\% \quad (1)$$

$$I_{rel}^L(n) = \frac{I_n/L_n}{\sum_{n=1}^p I_n/L_n} * 100\% \quad (2)$$

Where  $I_n$  is the intensity of protein  $n$  of  $p$  quantified proteins in a given sample and  $L_n$  is the length of protein  $n$ , based on the processed protein database. For evaluation of the two metrics, relative protein abundance was plotted as scatter plots between the different analysis conditions and the Pearson correlation coefficient (PCC) was calculated in Perseus [41].

For final protein quantification, MS data were analyzed as both tryptic and semi-tryptic digests using the following optimized search criteria: Peptides per protein  $\geq 2$  (razor and unique), protein FDR = 0.05, unmodified peptide score > 40, peptide FDR = 0.005. Match between runs and dependent peptides were both enabled. Increasing FDR to 5% for the tryptic analysis did not affect identification and quantification due to the applied score threshold.

## 2.9. Comparative analysis of transcriptomic and proteomic data

Comparative analysis was done on both the individual protein and subcellular compartmental levels. Transcript abundance was calculated as the relative TPM (rTPM) for the individual protein encoding gene, similarly to riBAQ on the protein level. As such, rTPM was calculated as:

$$rTPM(n) = \frac{TPM_n}{\sum_{n=1}^p TPM_n} * 100\% \quad (3)$$

where  $TPM_n$  is the transcript abundance of protein encoding gene  $n$  of  $p$  (1628) protein encoding genes in the final database.

Using the predicted subcellular localization, we then estimated the relative, subcellular distribution of proteins based on the transcriptome by grouping proteins according to predicted localization and summarizing rTPM compartment wise. Finally, we compared the transcriptome-based protein distribution with the actual protein distribution for the extracts in a relative, quantitative manner.

## 2.10. Method validation

To validate the analytical workflow and quantitative method presented, whole cell lysates from *Escherichia coli* (internal QC sample for instrument performance evaluation), was used. The biomass was prepared by cell lysis using ultra sonication, as previously described [45]. Following protein isolation by acetone precipitation, proteins were

subjected to conventional in-solution TCEP reduction, iodoacetamide alkylation, and tryptic digestion, before peptide purification and desalting using StageTips, as previously described [46]. One  $\mu\text{g}$  of sample was loaded on column in the start, middle, and end of the seaweed digest sequence, and data acquisition was performed as described above. Data analysis was performed identically to the seaweed samples and quantified using riBAQ as well as  $I_{\text{rel}}^{\text{L}}$  (using tryptic, semi-tryptic, an unspecific in silico digestion), as described above. The Uniprot reference proteome for *E. coli* K12 (UP000000625, downloaded October 10th 2017) was used as protein database in MaxQuant 1.6.0.43.

Robustness and data reproducibility was investigated by computing the protein-level relative standard deviation (RSD) across replicates for each condition as well as the RSD of means across the four quantitative approaches. RSD was calculated as unweighted (stochastic) and weighted means across all quantified proteins for each quantitative workflow. The weighted mean was defined as:

$$\overline{\text{RSD}}_{\text{mean}} = \sum_{n=1}^p \text{RSD}_n * \text{riBAQ}_n$$

where  $\text{RSD}_n$  is the RSD across the three replicates for protein  $n$  of  $p$  quantified proteins and  $\text{riBAQ}_n$  is the relative abundance (the weight) for protein  $n$ . Stochastic and weighted mean RSD was similarly computed for all three forms of  $I_{\text{rel}}^{\text{L}}$  (using tryptic, semi-tryptic, an unspecific in silico digestion).

Moreover, we also investigated the correlation between methods by computing the PCC across the four replicate means in the initial quantification, after filtering by requiring identification of a given protein in a given workflow in at least two of three replicates, as well as after a quality-based filtering (QBF, see Supplementary information), as previously described [47]. The impact of QBF was also evaluated through the quantitative loss, i.e. the relative abundance constituted by filtered proteins prior to the filtering step. Finally, we visualized the occurrence of each quantified protein (group) across replicates within each quantitative methods as well as all quantified protein (groups) between methods after the different data filtration steps using Venn diagrams.

### 2.11. Data analysis and visualization

Statistical and correlation analysis of transcriptome and MS data was performed in Perseus 1.6.1.3 [41,48]. Venn diagrams were plotted with jvenn [49] or Venny 2.1 (<https://bioinfogp.cnb.csic.es/tools/venny/index.html>). Additional data visualization was performed using Origin-Pro 8.5.0 SR1 (OriginLab Corporation, Northampton, MA, USA) and figures assembled in their final form using INKSCAPE version 0.92.3 (<https://inkscape.org/>).

## 3. Results and discussion

### 3.1. Transcriptome assembly, protein annotation, and subcellular localization

The transcriptome of *Eucheuma denticulatum* was de novo assembled using publicly deposited transcriptome data at NCBI SRA database. The quality of the assembly was estimated by basic contig statistics and percentage of the remapped reads. Both metrics indicated a high quality of the assembly with an N50 value of 1891 bp (Table A.1) and more than 90% of the reads mapped back to the contigs (Table A.2). Based on the transcriptomic information, an *E. denticulatum* protein database was constructed for subsequent mass-spectrometry (MS) analysis. First, the protein-coding sequences were predicted and identified by BlastX and BlastP search as well as their protein family by searching against Pfam database. Then the transcript expression level was calculated in terms of transcripts per kilo megabase (TPM) and removed proteins with TPM below 100, which resulted in 1628 proteins retained for the database. The TPM threshold was applied in order to filter out any potentially erroneous reads. Although this may in fact also filter some proteins with

low copy numbers from the database, the primary objective was to identify highly expressed proteins, and consequently this filtering is not considered to have substantial impact for the intended use of the database. A full list of protein accessions and their associated TPMs, rTPMs, Pfam functions, BlastX targets, and BlastP targets can be found in Table A.3 and in the linked Mendeley data repository [50]. The *de-novo* protein database for *E. denticulatum* can be found in.fasta format in Table A.4 as well as in the Mendeley data repository.

Although homology-inferred annotation using BLAST can indicate potential functions and localizations for the individual proteins, extraction of potential functions and subcellular localization on the proteome level is a tedious task. Additionally, as only verified Uniprot/Swiss-Prot proteins were included, the resulting annotations were of suboptimal quality (Table A.3) due to the lack of verified annotations on related and comparable species to *E. denticulatum*. Consequently, a bioinformatic prediction of subcellular localization on the individual protein level was used. This data type is easily binnable for large proteomes. The DeepLoc neural network was developed for eukaryotic proteins with little or no available homology data [36], and was therefore deemed highly suitable for the case of this study. Using DeepLoc, a localization probability of  $0.63 \pm 0.21$  (Fig. A.1) was achieved for the entire proteome.

### 3.2. 1D SDS-PAGE analysis and protein quality assessment

Both protein extracts display absence of distinct proteins bands and an apparent smear along the gel concentrating in the low MW range, as seen from 1D SDS-PAGE analysis in Fig. 1. This is in contrast to previous studies on *E. denticulatum* protein extracts [51], where distinct protein bands were observed and the low MW concentrated smear was absent. The significant difference in protein appearance by SDS-PAGE may be directly ascribed to the extraction method, as the authors here used a more elaborate protocol including organic (phenol) solvents as well as reducing conditions. Their approach may be significantly better for efficient extraction of higher quality protein from the whole seaweed, but is not feasible on an industrial scale, and hence was not attempted. Although analyzed extracts in this study were obtained from a scalable process, it appears to be at the cost of protein quality.

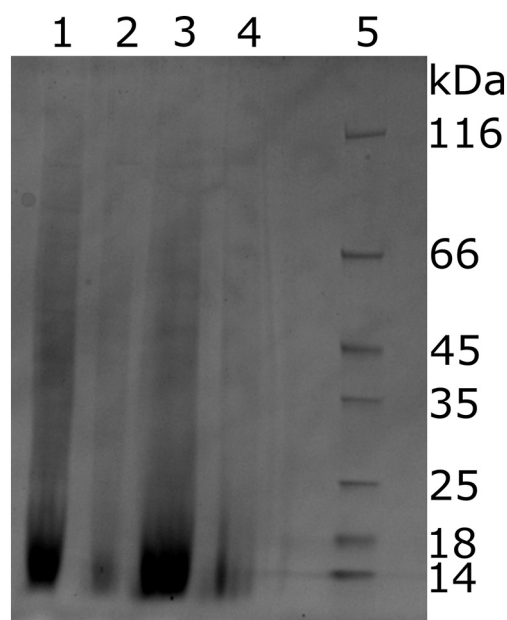


Fig. 1. SDS-PAGE of *E. denticulatum* protein extracts investigated in this study. Protein loading is based on supplied protein content of 7.1% and 70% for extract A and B, respectively. 1: Extract A, 100  $\mu\text{g}$ . 2: Extract A, 20  $\mu\text{g}$ . 3: Extract B, 100  $\mu\text{g}$ . 4: Extract B, 20  $\mu\text{g}$ . 5: MW Marker.

The overall appearance of both extracts analyzed, however, are quite similar. The lack of distinct protein bands could indicate co-extraction of other cellular moieties, which could interfere with electrophoresis and ultimately resulting in the observed smears. This has been reported for co-extracted lipids [52,53], carbohydrates [54,55], and DNA [56]. Further modification of proteins (e.g. glycoproteins) may also add to the smearing observed on SDS-PAGE [57–59]. Based on the excessive gelling in aq. solution and at higher sample concentrations particularly for extract A (data not shown), the samples likely contain significant amount of residual carrageenan, co extracted with the protein under the applied conditions as presumed. Nevertheless, partial hydrolysis during extraction using high temperature under alkaline conditions, as employed for both extraction methods, may also have an influence on the high density, low molecular weight pattern with no distinct bands.

In order to estimate the accuracy of the total protein by Kjeldahl-N analysis, we determined the soluble protein content in both aqueous solution and a slightly alkaline buffer with added detergent using Qubit protein assay (Table 1). From here, it is evident that the Kjeldahl-based total protein in fact correlates quite well with the soluble protein content – at least when solubilized in an alkaline buffer with detergent. A nitrogen-to-protein conversion factor of 6.25, the “Jones factor”, is commonly employed in food protein science and has been so for 90 years [60,61]. Nevertheless, the universal conversion factor has been subject to several investigations, and species-dependent conversion factors are commonly recommended [62]. For seaweeds in particular, the factor can still vary significantly, but as no factor is available for *E. denticulatum*, a general conversion factor of 5.0 can be applied [63]. By doing so, and thereby lowering the protein content by 20% (Table 1), the Kjeldahl-N method now underestimates the protein content compared to Qubit – in particular for extract B. In this respect, it is worth considering that the conversion factor is representative of the total organism and may not be directly suitable for the extracted proteome. Additionally, the non-protein nitrogen content of the extract is undetermined, and may also influence both the Kjeldahl-N and the Qubit outputs to some degree.

It is also evident that the aqueous solubility of the protein in the extracts is quite low (11–15% of the total protein), whereas a slightly alkaline buffer with a low amount of detergent practically fully solubilizes the protein (6-fold and 10-fold solubility increase for extract A and B, respectively). This also correlates well with the physical appearance of the solubilized extracts following centrifugation (Fig. A.2), where a significantly higher amount of solid precipitate is visible in the aqueous solutions. Nevertheless, smear and apparent lack of intact high MW protein from SDS-PAGE must be taken into consideration for protein quantification and in the evaluation of the protein extracts as source for further processing as well as potential release of bioactive peptides.

**Table 1**

General characteristics for the two *E. denticulatum* extracts analyzed in this work including total protein and soluble protein content.

Extract	Extraction method	Protein content (Kjeldahl-N * 6.25) <sup>1</sup>	Protein content (Kjeldahl-N * 5.0) <sup>2</sup>	Soluble protein (ddH <sub>2</sub> O)	Soluble protein (buffer)
A	Alkaline, hot-water extraction → ultracentrifugation → lyophilization	7.1%	5.7%	1.1%	6.2%
B	Alkaline, hot-water extraction → ultracentrifugation → lyophilization → acidic precipitation → lyophilization	70%	56%	7.3%	74.8%

<sup>1</sup> Total protein by Kjeldahl-N was supplied by CP Kelco.

<sup>2</sup> Calculated based on supplied protein content (<sup>1</sup>) using a conversion factor of 5.0 [63].

### 3.3. Identification and quantification of peptides and proteins by LC-MS/MS

Quantitative proteomics was carried out with the main objective of identifying high abundance proteins with a satisfactory level of evidence (i.e. score and sequence coverage). Identification of such proteins would potentially open for further processing of the extract in a targeted manner for e.g. release of potentially bioactive peptides. Initially, we applied an iterative process where different in silico digestion methods (i.e. specific, semi-specific, and unspecific digestion), peptide- and protein-level FDR, and number of identified peptides per protein were attempted. This was done not only to identify the optimal parameters for analysis, but also to investigate the feasibility of applying the two presented quantitative metrics. The iterative process was of utmost importance, as the sample quality and especially the number of identified peptides and proteins for the extracts (i.e. data quality) was low. In particular, using FDR-controlled thresholds for low quality data may be an unsuitable approach, as true FDR may be much higher. An indication hereof is also the identification of unexpected contaminants (e.g. milk proteins). This furthermore relates to the protein database not being experimentally validated against a high quality protein extract from the organism. Such a validation would strengthen the analytical pipeline by improving confidence in both the database and the experimental findings. These aspects were also part of the rationale for performing manual inspection and quality-based filtering of data following parameter optimization.

A low number of peptide identifications significantly affects protein identification and quantification via the impact on FDR-controlled thresholds. This is ultimately an inherent property of the peptide scoring algorithm. MaxQuant employs the Andromeda search engine, in which peptide score is not only based on PEP, but also on the intensity of a given feature [64–66]. Consequently, high intensity features with significant PEP (i.e. potential false positives), which in other studies may have been filtered out, will obtain a sufficiently high peptide score and be used in protein quantification. Ultimately, this leads to false identification of proteins with a significant relative abundance, which impairs further analysis. By applying more stringent thresholds on both peptide and protein level, this is alleviated to some extent. Nevertheless, it may be needed to inspect and evaluate PEPs rather than apply threshold filtering on peptide score alone, as PEP relies solely on PSM and sequence-dependent features. Furthermore, as charge +1 was excluded during acquisition, this may have limited the number of unspecific peptide significantly, ultimately making this approach somewhat unsuitable for analysis of these data. Exclusion of +1 peptides also favors peptides containing at least one Lys/Arg residues in the unspecific analysis (676 of 740 unfiltered peptides), thereby introducing bias. As sample and data quality was also poor for the analyzed extracts, we ultimately consider an unspecific database search too unreliable for this data. These aspects are further discussed and evaluated in Appendix A.

By applying the optimized search parameters, a total of 66 proteins across both extracts and analysis methods (tryptic and semi-specific) following filtering of trypsin and reverse hits (Stage 1) were identified and quantified (Table 2). Extract B is highly contaminated since 80% (based on  $I_{rel}^L$  for semi-tryptic analysis) of all identified proteins were constituted by common contaminants (Table 2 (Stage 1)), primarily keratin, whereas the content in extract A was 20%. Although common contaminants are usually filtered out prior to quantification as they may be introduced during sample preparation for MS analysis, the magnitude in Extract B is noteworthy. This particularly low data quality for this extract further highlights the low sample quality and indicates that additional processing, although increasing total protein content substantially, comes at a great cost. In total, 40 proteins were identified across both extracts and analysis conditions, following filtering of common contaminants and subsequent re-quantification (Table 2 (Stage 2), Tables A.5; A.6). Semi-specific analysis resulted in identification of four additional proteins (one in extract A and three in extract B),

**Table 2**

Relative protein abundance of *E. denticulatum* extracts A and B (after filtering of trypsin) following initial quantification (Stage 1) with optimized search parameters by  $I_{rel}^L$  and riBAQ for both tryptic and semi-specific analysis. Proteins are divided in common contaminants (Stage 1 filtered proteins), false positive identifications/contaminants (Stage 2 filtered protein), and final, verified proteins (Stage 3). Common contaminants are annotated using their UniProt accession number. NQ: Protein not quantified in the specific sample using the specific analysis method.

Stage 1								Stage 2							Stage 3						
Protein IDs	$I_{rel}^L$ A tryp	riBAQ A	$I_{rel}^L$ A semi	$I_{rel}^L$ B tryp	riBAQ B	$I_{rel}^L$ B semi	Contaminant ID	Protein IDs	$I_{rel}^L$ A tryp	riBAQ A	$I_{rel}^L$ A semi	$I_{rel}^L$ B tryp	riBAQ B	$I_{rel}^L$ B semi	Protein IDs	$I_{rel}^L$ A tryp	riBAQ A	$I_{rel}^L$ A semi	$I_{rel}^L$ B tryp	riBAQ B	$I_{rel}^L$ B semi
CON_ENSEMBL	NQ	NQ	NQ	0.1%	0.1%	NQ	Keratin	rflc10492_g1_i1. p1	NQ	NQ	NQ	0.1%	0.1%	NQ	rflc1505_g2_i1. p1	8.6%	11.8%	13.4%	1.2%	1.5%	1.7%
CON_O43790	NQ	NQ	NQ	0.7%	0.5%	0.6%	Keratin	rflc1275_g1_i1. p1	NQ	NQ	NQ	NQ	NQ	11.1%	rflc1613_g1_i1. p1	NQ	NQ	0.0%	0.0%	0.0%	0.0%
CON_P02533	0.1%	0.1%	0.0%	1.9%	1.5%	1.8%	Keratin	rflc1294_g1_i1. p1	0.0%	0.0%	NQ	0.0%	0.0%	NQ	rflc17304_g1_i1. p1	2.6%	2.3%	2.5%	0.3%	0.2%	0.2%
CON_P02662	NQ	NQ	NQ	0.1%	0.1%	NQ	$\alpha$ S1-casein	rflc1357_g1_i1. p1	NQ	NQ	NQ	0.0%	0.0%	NQ	rflc17615_g1_i1. p1	0.1%	0.1%	NQ	NQ	NQ	NQ
CON_P02666	0.0%	0.0%	0.0%	2.2%	3.8%	1.8%	$\beta$ -casein	rflc17161_g1_i1. p1	NQ	NQ	NQ	0.1%	0.0%	NQ	rflc231_g1_i1.p1	0.1%	0.1%	NQ	NQ	NQ	NQ
CON_P02754	6.5%	4.7%	6.9%	5.2%	3.7%	4.9%	$\beta$ -lactoglobulin	rflc17201_g1_i1. p1	0.0%	0.0%	NQ	NQ	NQ	NQ	rflc2364_g1_i1. p1	0.1%	0.1%	0.1%	0.5%	0.4%	0.4%
CON_P02768	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	Albumin	rflc17231_g1_i1. p1	0.0%	0.0%	0.0%	0.3%	0.3%	0.2%	rflc2556_g1_i1. p1	0.0%	0.0%	NQ	0.0%	0.0%	NQ
CON_P02769	0.0%	0.0%	0.0%	1.5%	1.1%	1.3%	Albumin	rflc2788_g1_i1. p1	NQ	NQ	NQ	NQ	NQ	0.2%	rflc3760_g1_i1. p1	NQ	NQ	0.7%	NQ	NQ	NQ
CON_P04264	5.5%	5.6%	4.6%	33.3%	33.1%	26.7%	Keratin	rflc3249_g1_i1. p1	0.0%	0.0%	NQ	NQ	NQ	NQ	rflc4090_g1_i1. p1	0.1%	0.1%	0.1%	0.0%	0.0%	NQ
CON_P08779	0.0%	0.0%	0.0%	0.3%	0.2%	0.2%	Keratin	rflc4757_g1_i1. p1	NQ	NQ	NQ	0.1%	0.1%	0.2%	rflc4354_g1_i1. p1	3.3%	4.1%	3.1%	0.0%	0.0%	0.0%
CON_P13645	6.2%	6.9%	5.5%	19.3%	21.1%	16.8%	Keratin	rflc4921_g1_i1. p1	NQ	NQ	NQ	0.0%	0.0%	0.1%	rflc4671_g1_i1. p2	0.3%	0.4%	0.2%	NQ	NQ	NQ
CON_P13647	0.2%	0.2%	0.2%	1.8%	1.6%	0.9%	Keratin	rflc5168_g1_i1. p1	NQ	NQ	NQ	0.0%	0.0%	0.1%	rflc5232_g1_i1. p1	0.9%	0.9%	0.8%	NQ	NQ	NQ
CON_P19013	NQ	NQ	0.0%	0.2%	0.1%	0.1%	Keratin	rflc5952_g1_i1. p1	NQ	NQ	NQ	NQ	NQ	0.1%	rflc6313_g1_i1. p1	27.2%	25.3%	25.2%	0.2%	0.2%	1.4%
CON_P35527	1.1%	1.3%	1.1%	17.2%	19.7%	16.2%	Keratin	rflc6797_g1_i1. p1	NQ	NQ	NQ	0.0%	0.0%	0.0%	rflc6373_g1_i1. p1	0.0%	0.0%	0.0%	0.0%	0.0%	NQ
CON_P35908	2.3%	1.8%	1.9%	6.1%	4.7%	5.9%	Keratin	rflc6825_g1_i1. p3	1.4%	0.9%	1.3%	2.8%	1.7%	2.1%	rflc6458_g1_i1. p1	1.1%	1.8%	1.4%	NQ	NQ	NQ
CON_P48668	0.0%	0.0%	0.0%	0.6%	0.6%	1.3%	Keratin	rflc6945_g1_i1. p2	0.0%	0.0%	0.0%	0.2%	0.2%	0.2%	rflc6656_g1_i1. p1	0.0%	0.0%	0.0%	0.3%	0.3%	0.3%
CON_P78386	0.0%	0.0%	NQ	0.1%	0.1%	0.1%	Keratin	rflc8389_g1_i1. p1	0.0%	0.0%	NQ	0.1%	0.0%	NQ	rflc6834_g1_i1. p3	0.0%	0.0%	NQ	0.0%	0.0%	NQ
CON_Q04695	NQ	NQ	0.0%	0.2%	0.1%	0.2%	Keratin								rflc6963_g2_i1. p1	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%
CON_Q14525	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	Keratin								rflc7052_g1_i1. p1	27.0%	25.1%	25.8%	1.6%	1.5%	1.5%
CON_Q5D862	0.0%	0.0%	NQ	0.0%	0.0%	NQ	Filaggrin-2								rflc7052_g1_i2. p1	3.5%	3.9%	3.2%	0.4%	0.4%	0.4%
CON_Q6KB66	NQ	NQ	NQ	0.0%	0.0%	NQ	Keratin								rflc7216_g1_i1. p1	0.3%	0.3%	0.4%	NQ	NQ	NQ
CON_Q9UE12	0.0%	0.0%	NQ	0.6%	0.5%	0.5%	Keratin								rflc8421_g1_i1. p1	1.2%	2.0%	1.2%	NQ	NQ	NQ
CON_Q9NSB2	NQ	NQ	NQ	0.1%	0.0%	0.0%	Keratin								rflc926_g1_i1.p1	0.0%	0.0%	0.0%	NQ	NQ	NQ
CON_Q7Z3Y8	NQ	NQ	NQ	0.0%	0.0%	0.0%	Keratin														
CON_Q86YZ3	NQ	NQ	NQ	0.0%	0.0%	0.0%	Hornerin														
CON_Q8IUT8	NQ	NQ	NQ	NQ	NQ	NQ	Keratin														

whereof one (c1275\_g1\_i1.p1) constitutes more than half of the Stage 2 protein by  $I_{rel}^L$  in extract B. Furthermore, 11 proteins were not identified by this approach (four in extract A, three in extract B and four identified in both extracts using tryptic conditions), but none of these were of high abundance. From plotting relative abundance by both riBAQ and  $I_{rel}^L$  (Fig. A.3), a correlation was seen within each extract (PPC = 0.99–1.0 for extract A; PPC = 0.19–0.95 for extract B), but the semi-specific analysis of extract B correlated poorly with the tryptic analysis. The correlation between extracts was even worse (PPC = 0.14–0.55), indicating that the stringent quality parameters applied for automatic filtering, were not fully capable of cleaning the data from bad peptide spectrum matches (PSMs) and dubious protein identifications. This illustrates the challenges of applying FDR control for quality-based filtering of low quality data from low quality protein samples.

Identified “outliers” (Tables A.5; A.6) that did not correlate between extracts (i.e. are suddenly highly enriched in extract B) may in fact be contaminants with some homology to the *E. denticulatum* proteome (further details are presented in the Appendix A). For instance, in the tryptic analysis of extract B, c6825\_g1\_i1.p3 is highly abundant but only identified by two peptides, which both map to histones from e.g. humans. Histone was also the BLASTX target (*Xenopus laevis* (African clawed frog) histone H2AX) as well as the predicted function by Pfam (Table A.3). Consequently, and because it was very low abundance in extract A, this was ascribed as contaminant to the extract and not originating from the seaweed. Although histones were bound to be identified in *E. denticulatum*, homologues from other organisms would bias quantification and it was consequently excluded. Furthermore, the highly abundant protein identified by semi-specific analysis of extract B only (c1275\_g1\_i1.p1), was also identified by only two peptides. As the protein score of 11.8 was very low (see Appendix A and Table A.6), and the posterior error probability (PEP) was significant (PEP > 0.05), these were regarded as bad PSMs and the protein ID was deemed false positive. Based on these observations, manual inspection and validation was performed in order to apply a final filtering step using the rationale described above. In the filtering, significant weight was put on evaluation of PEP rather than peptide score, as low scoring peptides (<40) were pre-filtered in the optimized search parameters (see Appendix A for further details). Filtered proteins, along with the rationale for their exclusion, can be found in Table A.7 and proteins are listed in Table 2 (Stage 2).

Following filtering, verified proteins were re-quantified (Stage 3) the list of identified proteins was reduced from 40 to 23 proteins across extracts and conditions (Table 3). The stringent parameters applied in data analysis, as well as requirements for inclusion in the final list, fully alleviated the problem of new and significantly abundant proteins showing up in extract B (see Table 2 and Fig. A.3), as no proteins exclusive for extract B, were observed (Fig. 2B). Nine proteins were observed exclusively in extract A, but this may be explained as loss during the extended processing for extract B. Extended processing may also be a likely explanation for the extract B exclusive peptides identified (Fig. 2A). Furthermore, all nine proteins are of somewhat low abundance ( $I_{rel}^L < 2\%$ ), and do not affect the overall protein distribution significantly. Interestingly, the nine proteins identified in both extracts using both analyses approaches, constituted >93% of the verified protein in extract A and > 99% of the verified protein in extract B (by  $I_{rel}^L$ ). In fact, three proteins (c6313\_g1\_i1.p1, c7052\_g1\_i1.p1, and c1505\_g2\_i1.p1) constitute more than 75% of the total protein identified in both extracts (Table 3). Furthermore, an isoform of c7052\_g1\_i1 (c7052\_g1\_i2), which only differs in the C-terminal region of the protein, was also identified in significant abundance. If included, the proteins constitute >80% of the verified seaweed-specific protein in both extracts. Regardless of using riBAQ or  $I_{rel}^L$ , the three (four) proteins constitute the vast majority of quantified protein in both extracts (Table 3). With MW in the range 16–24 kDa, all three (four) proteins correlated well with the observations from SDS-PAGE (Fig. 1), even though no clear protein bands were observed. This indicated that these

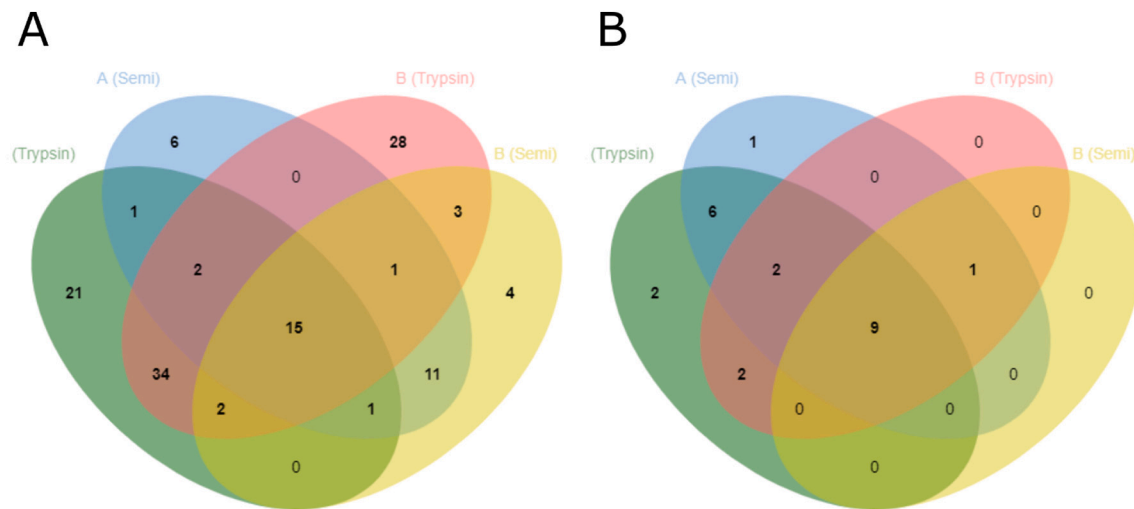
three (four) proteins in particular may be of certain interest as potential sources of e.g. bioactive peptides, however, the low sample quality may be a critical factor in terms of e.g. cleavage site accessibility in large scale processes. Protein sequences and experimental sequence coverage for the three major proteins are shown in Fig. 3. From BlastP against verified proteins in UniProtKB/Swiss-Prot (Table A.3), c7052\_g1\_i1.p1 (as well as the isoform) shows some homology to an immunogenic, periplasmic protein from *Brucella suis* (UniProt AC# POA3U9), whereas Pfam indicates it could be related to the DNA repair protein REV1. Neither c6313\_g1\_i1.p1 nor c1505\_g2\_i1.p1 matched any proteins from the Blast homology or Pfam protein families. Consequently, the nature, structure, and function of the three highly abundant proteins remains unknown.

Filtering resulted in improved correlation between the two extracts up to a PCC of 0.91 for relative abundances quantified by  $I_{rel}^L$  (Fig. 4). This indicates that in light of all the complications, the two protein extracts are in fact somewhat comparable, when all redundancy and contamination was addressed. Furthermore, the in-sample correlation between riBAQ and  $I_{rel}^L$  (PCC 0.87–1.0) indicated that  $I_{rel}^L$  may in fact be quite powerful analogue to riBAQ for non-standard (i.e. semi- or un-specific) analysis. As semi-specific in most cases increase both number of identified peptides as well as the sequence coverage on the individual protein level  $I_{rel}^L$  could be a powerful tool in the analysis of proteins where partial (non-specific) hydrolysis is observed, as this will include all peptide originating from the parent proteins rather than proteolytic peptides alone. Nevertheless, the lack of database validation and particularly low data quality means that any finding in such a dataset should be seen as merely indicative. Moreover, the presented approach for quantification via semi-tryptic peptides should indeed be systematically investigated and validated before the potential and applicability of such a method can be evaluated.

The tryptic analysis showed a significantly lower number of peptides and relative abundance for c6313\_g1\_i1.p1 compared to the semi-tryptic analysis of extract B. This could indicate that this particular protein is subject to partial hydrolysis during the additional processing, which again strengthens the use of the semi-specific analysis for this type of protein extract. That being said, non-tryptic hydrolysis may also originate endogenous proteases/peptidases or from associated organisms such as seaweed specific/associated bacteria or fungi, which may have been present prior to thermal inactivation [67–69]. There may also be a risk of missing identification of endogenous, non-proteinogenic peptides and peptidomimetics, the co-called ribosomally synthesized and post-translationally modified peptides (RiPPs) [70], which have been reported across different marine sources [71], using this type of encoded database search strategy. Process-induced amino acid modification such as side chain cross-linking (e.g. lysinoalanine formation [72]) and Asp/Glu deamidation [73], may also have occurred during alkaline heating. As such, the low number of identifications may have been increased by including other protein modifications or even performing an open modification search [74] to obtain insight on endogenous modifications and/or potential process-induced modifications. However, as no high quality protein extracts were available as baseline for comparison and thus discriminate between endogenous and process-induced modifications, and because the main objective of the study is to identify highly abundant proteins, this approach was regarded outside the scope of this study.

Revisiting the unfiltered peptide identifications from unspecific analysis (using FDR = 0.1), we find that of the 740 peptides identified, regardless of scorer or PEP, 624 (84%) had at least one tryptic terminus while 428 (58%) were in fact fully tryptic peptides (data not shown but accessible through the Mendeley repository). This observation could indicate that in spite of increasing the size of the decoy database using unspecific analysis [75], and in spite of the low sample and data quality complicating particularly unspecific analysis, process-induced decomposition may not be widespread in the extracts. Ultimately, this suggests that the effect of applied processes during extraction may more likely be





**Fig. 2.** 4-way Venn diagrams showing identified peptides (A) and proteins (B) with optimized parameters (5% FDR and minimum score threshold) and following filtering for extract A using tryptic analysis (green), extract A using semi-tryptic analysis (blue), extract B using tryptic analysis (red), and extract B using semi-tryptic analysis (yellow). List sizes (in the same order) for peptides (A) are 76, 37, 85, and 37 for a total of 129 identified peptides. List sizes (in the same order) for proteins (B) are 21, 19, 14, and 10 for a total of 23 identified proteins. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

```
>c1505_g2_i1.p1      MatureSequenceCoverage=37%
MSHKLLPSLLLLSFLLLLNFPPPTSSTPALSPHSPPHKNPNNALSPHLQSITEEDDDDLTA
PSTLRRRVFATPRVFENIIVRYRRFAPLALNADPETTSQIRAVLVVCLQQVRELQRDDNVRN
VRILLSSLVLLLLDVLVCLLN*
```

```
>c6313_g1_i1.p1      MatureSequenceCoverage=59%
MALIWLLSIVFALLTALGTTSAVNLNLPVLRVNANCRNDFPVRNNIRLVRVYVWVNDMQTD
LDTSTRFLGENVGFACSGSAQTYLSFEGDNTGRGEEVAIVEVGDARKDEAWRGTTTCIVL
KAQWFNSRNQGNIRVIVEIRNKGTNNLIRDPLEIVARPGVGDSCSMRLIATVVVDEDEGI
YLARAFNCPN*
```

```
>c7052_g1_i1.p1      MatureSequenceCoverage=58%
MTPLLLLPLLLALTTANTPHPTPRISVITGDASVAAEPDIATLTTEVVTLAPTAQAALTR
NNRLTSALFDLSEFNVSRRDIQTTTSFVSVPREFQRPDNDVTRIIGYTVRNSLRVTVRDL
SNLGLILDALVRAGSNLSLSRISFGISNEADLRDQARELAVKDAVRRATLLTKAAGTGLGK
VLSIREGGRSTGGFSAQVRARREAEVPIAPGELQVSARVTLIEIELVG*
```

related to thermal denaturation and aggregation, making solubility and co-extracted compounds the most pronounced challenge in relation to application the extracts. Nevertheless, exclusion of +1 charged peptides may introduce a significant bias towards peptides with tryptic C-termini as previously discussed. Adding to the low data quality, it is not possible to draw any finite conclusions in relation to process-induced hydrolysis.

### 3.4. Enrichment of extracellular proteins

In Fig. 5, the relative subcellular distribution of proteins predicted by DeepLoc, is presented. For the transcriptome analysis (Fig. 5C), a relative broad distribution of proteins (by rTPM) is observed with the majority of proteins being ascribed to the nucleus (24%), plastid (22%), cytoplasm (20%), mitochondria (18%), and extracellular (7%). This distribution does not correlate with the protein distribution established by LC-MS/MS, regardless of data analysis conditions employed. In fact, there is a very significant enrichment in extracellular proteins. For extract A (Fig. 5A), almost exclusively extracellular proteins are identified (97%) by  $I_{rel}^L$ . While extract B (Fig. 5B) has some content of plastid and cytoplasmic protein, the majority of identified proteins are extracellular (87%) by  $I_{rel}^L$ . The three primary proteins in both extracts are all

**Fig. 3.** Protein sequence and experimentally identified peptides across both extracts and analysis methods (highlighted in grey) for the three most abundant *E. denticulatum* proteins identified. Signal peptides (indicated in red text) were predicted by signalP-5.0 [40] and the calculated sequence coverage is for the mature form of the proteins. All three proteins passed final selection criteria (Stage 3) and accounted for 82.2% and 76.4% (quantified by  $I_{rel}^L$  using semi-specific analysis) of the verified, seaweed-specific proteins in extracts A and B, respectively. Including the isoform of c7052\_g1\_i1 (c7052\_g1\_i1 – not shown), the proteins account for 86.4% and 83.2%, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

classified as being extracellular (Table 3). At the individual protein level, the extracellular protein with the highest rTPM of 1.1%, c17304\_g1\_i1.p1 (see Table A.3), was determined to constitute 3.2–3.5% of the molar protein content. Although still significantly abundant, the three highly abundant extracellular proteins described above ( $I_{rel}^L$  17–33% each) merely constituted 0.02–0.29% on the transcript level, indicating that the extraction method may not be selective for extracellular proteins per se, but rather a few selected extracellular proteins. To ensure validity of our predicted localizations, we analyzed the six most abundant proteins (constituting the majority of all quantified proteins) using three additional plant/eukaryotic predictors, and at least three of four predictors (all four predictors for five of six proteins) confirmed extracellular/secreted as the most probable pathway (Table A.9).

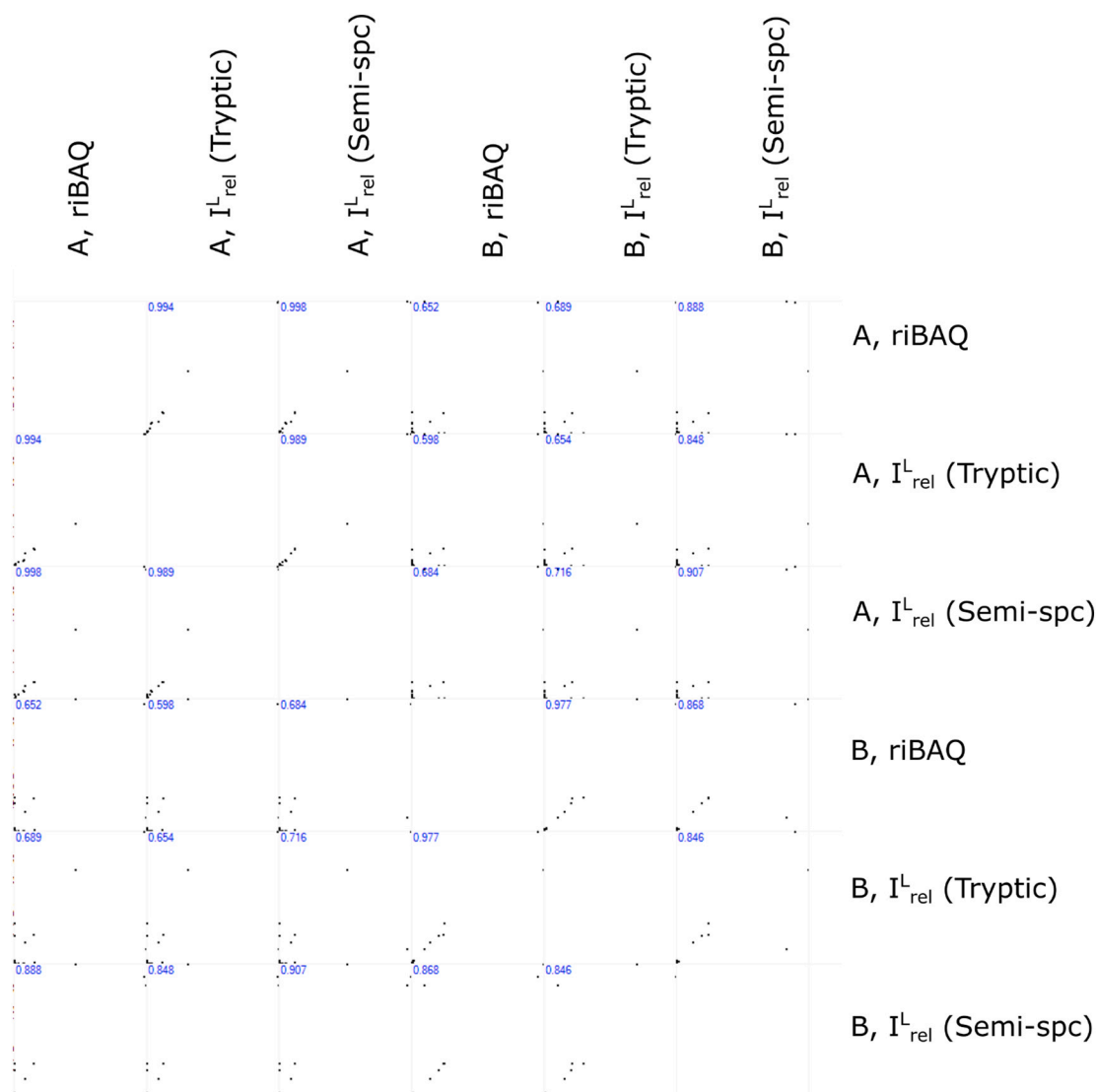
The fact that extracellular protein were almost exclusively identified in the extracts, is also very likely to explain the low extraction yields observed at the pilot plant (unpublished data from CP Kelco). From 20 kg of seaweed, 155 g material was obtained using a 1000 L extraction tank (Extract A). The protein content (by Kjeldahl-N and converted using the Jones factor) of 7.1% correspond to merely 11 g of protein following extraction corresponding to a protein yield of 0.055%. Further

**Table 3**

Summary of verified proteins following parameter optimization, manual inspection, and filtering (Stage 3) for *E. denticulatum* extracts A and B using both tryptic and semi-tryptic analysis. For each identified protein, the molecular weight, number of identified peptides, sequence coverage, protein score, riBAQ,  $I_{rel}^L$ , rTPM, subcellular localization, and localization probability. Semi-tryptic quantification by  $I_{rel}^L$  is highlighted in bold as the final quantitative estimation metric.

Protein ID	MW [kDa]	#Pep A tryp	#Pep B tryp	#Pep A semi	#Pep B semi	Seq. cov. A tryp [%]	Seq. cov. B tryp [%]	Seq. cov. A semi [%]	Seq. cov. B semi [%]	Score tryp	Score semi	riBAQ A tryp	$I_{rel}^L$ A tryp	$I_{rel}^L$ A semi	riBAQ B tryp	$I_{rel}^L$ B tryp	$I_{rel}^L$ B semi	rTPM	Subcellular localization <sup>1</sup>	Subcell score <sup>1</sup>
c6313_g1_i1.p1	21.153	7	1	13	5	36.3	7.4	47.9	17.4	323.3	323.3	32.3%	35.5%	<b>32.1%</b>	3.4%	3.8%	<b>23.3%</b>	0.29%	Extracellular	0.6985
c7052_g1_i1.p1	24.213	7	8	16	7	36.1	34.8	45.4	31.7	323.3	323.3	31.9%	35.2%	<b>33.0%</b>	31.9%	35.9%	<b>25.1%</b>	0.02%	Extracellular	0.9128
c1505_g2_i1.p1	15.778	4	4	7	5	30	30	30	30	323.3	323.3	15.1%	11.3%	<b>17.1%</b>	33.4%	25.4%	<b>28.0%</b>	0.14%	Extracellular	0.4441
c4354_g1_i1.p1	40.332	8	1	7	3	24.6	3.5	21.7	5.6	323.3	323.3	5.2%	4.3%	<b>4.0%</b>	0.2%	0.1%	<b>0.3%</b>	0.15%	Extracellular	0.8483
c7052_g1_i2.p1	23.965	6	5	15	5	30.8	25.1	40.1	25.1	163.3	140.2	4.9%	4.5%	<b>4.1%</b>	8.3%	7.8%	<b>6.8%</b>	0.06%	Extracellular	0.9431
c17304_g1_i1.p1	27.965	6	2	6	1	23	6.7	19.7	3	188.3	190.4	2.9%	3.4%	<b>3.2%</b>	4.8%	5.7%	<b>3.5%</b>	1.08%	Extracellular	0.5089
c8421_g1_i1.p1	59.681	4	0	4	1	10	0	10	2.3	323.3	323.3	2.6%	1.6%	<b>1.6%</b>	0.0%	0.0%	<b>0.0%</b>	0.04%	Membrane	0.9998
c6458_g1_i1.p1	46.381	3	0	7	0	10.8	0	17.5	0	303.3	145.4	2.2%	1.4%	<b>1.8%</b>	0.0%	0.0%	<b>0.0%</b>	0.12%	Extracellular	0.8601
c5232_g1_i1.p1	18.952	2	0	2	0	13.5	0	13.5	0	125.3	104.9	1.2%	1.1%	<b>1.1%</b>	0.0%	0.0%	<b>0.0%</b>	0.03%	Extracellular	0.6419
c4671_g1_i1.p2	29.874	4	0	3	0	19.9	0	17	0	52.5	31.4	0.5%	0.4%	<b>0.2%</b>	0.0%	0.0%	<b>0.0%</b>	0.03%	Plastid	0.995
c7216_g1_i1.p1	25.446	2	0	3	0	10.8	0	16.9	0	16.1	19.3	0.4%	0.3%	<b>0.4%</b>	0.0%	0.0%	<b>0.0%</b>	0.08%	Extracellular	0.8121
c17615_g1_i1.p1	27.973	2	0	0	0	7.7	0	0	0	13.9	0.0	0.2%	0.2%	<b>0.0%</b>	0.0%	0.0%	<b>0.0%</b>	0.06%	Extracellular	0.9751
c6963_g2_i1.p1	165.47	10	2	4	2	6.5	1.6	2.1	1.6	108.7	55.5	0.1%	0.1%	<b>0.1%</b>	0.5%	0.5%	<b>0.6%</b>	0.04%	Plastid	0.6933
c4090_g1_i1.p1	16.129	1	2	1	1	6.8	12.9	6.8	6.1	15.3	11.2	0.1%	0.1%	<b>0.1%</b>	0.5%	0.7%	<b>0.0%</b>	0.02%	Plastid	0.9815
c231_g1_i1.p1	18.006	2	0	0	0	10.2	0	0	0	11.7	0.0	0.1%	0.1%	<b>0.0%</b>	0.0%	0.0%	<b>0.0%</b>	0.05%	Extracellular	0.9998
c2364_g1_i1.p1	50.492	1	5	2	5	2.4	9.9	4.1	9.9	135.4	105.1	0.1%	0.1%	<b>0.1%</b>	8.4%	10.8%	<b>6.9%</b>	0.23%	Cytoplasm	0.7655
c926_g1_i1.p1	79.764	3	0	2	0	3.6	0	2.6	0	20.2	10.9	0.1%	0.1%	<b>0.1%</b>	0.0%	0.0%	<b>0.0%</b>	0.04%	Extracellular	0.9924
c6373_g1_i1.p1	119.64	4	1	2	0	4.1	0.9	2.3	0	79.2	40.8	0.1%	0.0%	<b>0.0%</b>	0.1%	0.1%	<b>0.0%</b>	0.06%	Extracellular	0.5704
c6656_g1_i1.p1	43.007	3	3	1	3	8.8	8.5	3.3	8.5	79.4	75.8	0.0%	0.0%	<b>0.0%</b>	7.0%	7.4%	<b>5.4%</b>	0.12%	Plastid	0.9982
c6834_g1_i1.p3	22.388	1	1	0	0	5.3	6.7	0	0	15.3	0.0	0.0%	0.0%	<b>0.0%</b>	0.5%	0.7%	<b>0.0%</b>	0.06%	Plastid	0.9985
c2556_g1_i1.p1	57.477	1	3	0	0	2	4.9	0	0	19.4	0.0	0.0%	0.0%	<b>0.0%</b>	0.8%	0.8%	<b>0.0%</b>	0.05%	Plastid	0.567
c1613_g1_i1.p1	32.099	0	2	1	2	0	7.8	5.4	9.5	35.0	31.5	0.0%	0.0%	<b>0.0%</b>	0.4%	0.4%	<b>0.2%</b>	0.09%	Plastid	0.9671
c3760_g1_i1.p1	32.533	0	0	3	0	0	0	6.6	0	0.0	253.1	0.0%	0.0%	<b>0.0%</b>	0.0%	1.0%	<b>0.0%</b>	0.06%	Lysosome	0.3775

<sup>1</sup> Subcellular localization and localization probability was computed using DeepLoc [36].



**Fig. 4.** Correlation of relative protein abundances between extracts (A and B), analysis conditions (tryptic and semi-specific), and quantification method (riBAQ and  $I_{rel}^L$ ) following manual validation, filtering, and re-quantification (Stage 3). Pearson Correlation Coefficients are shown in blue in the upper left corner of each subplot. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

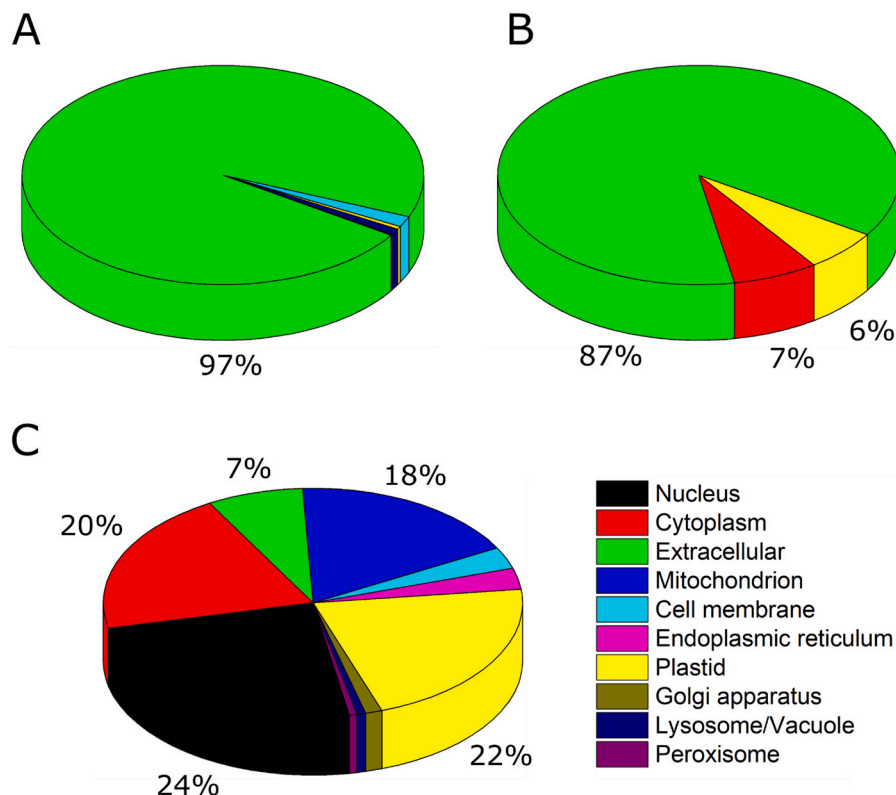
processing to concentrate protein by acid precipitation (Extract B) yielded 6.7 g of product with 71% protein corresponding to 4.8 g of protein and consequently a loss of 57% protein mass and thus an even lower yield (0.024%). These findings indicate that the hot-water extraction used, is not capable of disrupting cells and releasing intracellular proteins, although our findings are only indicative based on extract and data quality. Although our findings indicate how quantitative proteomics and bioinformatic analysis conceptually can be combined to obtain new insight into industrial processes, both the quantitative approach and the downstream bioinformatic pipeline need systematic validation to fully unleash the potential.

Low protein yields using simple aqueous extraction from *E. denticulatum* has previously been reported in the literature [76,77]. This in turn, implies that there is still a significant potential for improving protein extraction from the seaweed. In the past decade, numerous methods have been presented for extraction of protein from seaweed. Some of these may also be more suitable for *E. denticulatum*. These methods include pressurized and supercritical fluid extraction [78], addition of cofactors [79,80], microwave-assisted extraction [81], ultrasound-assisted extraction [82], or any combination thereof [83]. Enzyme assisted extraction (EAE) is an emerging technology for

seaweed protein extraction, showing great potential [84–87]. In a recent study, enzyme assisted extraction of *E. denticulatum* increased the protein yield up to 60% using Alcalase® or Viscozyme® (0.2% w/w) at pH 7 and room temperature [4]. The increased protein extraction efficiency was furthermore obtained without compromising the downstream carrageenan production. However, this method is not at present implemented by the carrageenan industry.

### 3.5. Method validation

Initially, a total of 629 protein groups were identified across analytical workflows (Table A.10), and 81–87% of all protein group IDs were found in all three sample replicates across analytical workflows (Fig. A.11). The quantitative variability between replicates, was evaluated by determining the  $RSD_{mean}$  for the four workflows and found to be highly comparable between tryptic and semi-specific workflows (29–30%), while the variability was slightly lower (25%) for unspecific  $I_{rel}^L$  (Table 4). As the variability for low abundance proteins may be markedly higher than for high abundance proteins, the weighted mean RSD was computed, where the relative abundance (for a given quantitative workflow) was used as weights on the individual protein group



**Fig. 5.** Relative subcellular protein distribution as predicted by DeepLoc [36] for A: Protein extract A. B: Protein extract B. For both protein extracts, relative abundance was estimated by  $I_L^{rel}$  through semi-tryptic analysis using optimized parameters, following manual inspection, validation, and filtering. C: Transcriptome analysis (by rTPM).

**Table 4**

Reproducibility within and across the four quantitative workflows using validation data from *E. coli* whole cell lysates. Quantitative variability between three technical replicates is evaluated by protein-level unweighted ( $RSD_{mean}$ ) and weighted ( $\overline{RSD_{mean}}$ ) mean relative standard deviation for riBAQ as well as tryptic (t), semi-tryptic (s), and unspecific (u) quantification with  $I_L^{rel}$ . RSDs are listed following initial quantification (following filtering of common contaminants and reverse hits), following filtration by requiring identification in at least two of three replicates (2rep), and following the quality based filtering (QBF). For each of the filtering steps, the relative quantitative loss (Loss) associated with the specific mode of data filtering is listed. Lastly, the method variability across all four workflows (Cross) is evaluated by RSD between triplicate means in the four workflows.

		riBAQ	$I_L^{rel}$ (t)	$I_L^{rel}$ (s)	$I_L^{rel}$ (u)	Cross
Initial	$RSD_{mean}$	29.7%	29.8%	29.2%	24.9%	14.7%
	$\overline{RSD_{mean}}$	9.1%	8.8%	8.6%	8.6%	12.2%
2rep	$RSD_{mean}$	29.7%	29.8%	29.2%	24.9%	12.9%
	$\overline{RSD_{mean}}$	9.1%	8.8%	8.6%	8.6%	12.1%
	Loss	0.05%	0.04%	0.07%	0.02%	N/A
QBF	$RSD_{mean}$	22.1%	22.1%	23.8%	22.0%	16.6%
	$\overline{RSD_{mean}}$	8.7%	8.5%	8.3%	8.5%	12.1%
	Loss	1.1%	0.96%	0.93%	0.53%	N/A

level. Not surprisingly, this significantly reduced overall variability (8.6–9.1%) and the four workflows perform equally in terms of reproducibility between replicates. Requiring quantification in at least two of three replicates for a given protein group in a given workflow had no effect on between replicate variability for any workflow (stochastic or weighted mean RSD), while the application of the quality based filtering (QBF) reduced mean RSD substantially (22–24%) with lesser impact on the weighted mean RSD (8.3–8.7%). These findings confirm that particularly low quality identifications have a negative impact on

quantitative reproducibility, that QBF positively affects this aspect, and that quantitative variability is generally higher in low abundance proteins.

While QBF positively affects the reproducibility, it reduces the number of identifications substantially (Table A.10). However, as QBF removes low quality identifications, it also substantially improves correlation of the identified protein groups between the quantitative workflows (Fig. A.12). Following initial quantification, 74% of protein identifications were identified by tryptic, semi-tryptic, and unspecific analysis, whereof ~7% were only identified by semi-tryptic and/or unspecific analysis. This in spite of lower identification rates in these workflows, likely due to increasing the size of the decoy database by applying such in silico approaches for producing the peptide databases, and even with increased FDR. Requiring two of three replicate identifications slightly reduced identification rates across all approaches and slightly improved the proportion of shared identifications across workflows to 77%. In contrast, QBF increased shared identifications to >99%, providing excellent comparability between the workflows (Fig. A.12). Furthermore, application of QBF did not affect relative quantification to any substantial degree, as the quantitative loss was found to be 0.5–1.1% across workflows (Table 4).

Having investigated reproducibility, correlation of identifications across workflows, and the impact of QBF, it is crucial to evaluate quantitative correlation between workflows. This was done by investigating variability (by stochastic and weighted mean RSD) and Pearson correlation between replicate means across the four workflows. While the stochastic mean RSD increases slightly from 15% to 17% following QBF, the weighted mean is largely unaffected at ~12% (Table 4). This indicates some degree of variability between the different quantitative workflows. Considering the quantitative correlation between replicate means across the four workflows (Figs. A.13,14,15), an overall satisfactory Pearson correlation is observed. It is noteworthy, that neither

requiring identification in two of three replicates nor QBF has any substantial influence on the cross-workflow correlation. However, it is also evident that riBAQ has the lowest correlation with other methods (PCC = 0.94–0.96), while all three  $I_{rel}^L$  workflows show excellent correlation (PCC = 0.99–1.0). This indicates that the method of normalization (tryptic peptides of 6–30 amino acids for iBAQ [43] and full protein length for  $I_{rel}^L$ ), is the main source for discrepancy.

Looking at merely the top-most abundantly quantified proteins, this does indeed appear the case. For instance, the most abundant protein group (P68066;C4ZYK2;B1XBQ6 – see Table A.11 and associated Mendeley data set [88]) has an average riBAQ of 7.8% while average  $I_{rel}^L$  across all workflows is substantially (approx. 1.5-fold) higher (11.7–12.2%). This is largely ascribed to the content and distribution of tryptic residues (Arg and Lys). P68066 contains 12.6% tryptic residues (Lys/Arg = 16; length = 127) resulting in 10 tryptic peptides used for iBAQ normalization (i.e. total protein intensity divided by protein iBAQ) and thus 12.7 amino acids per peptide. In contrast, the tryptic protein group 65 (P0A7N9;C4ZX8;B1X970) for instance has an extremely high content (27%) of tryptic peptides (Lys/Arg = 15; length = 55), resulting in only one peptide adhering to iBAQ length restrictions and thus 55 amino acids per peptide. This is also reflected in quantification, where the average tryptic  $I_{rel}^L$  is 2.9-fold higher than average riBAQ (2.7% vs. 0.9%). However, if such inconsistencies are inherent bias of riBAQ or  $I_{rel}^L$  cannot be determined without systematic investigation of both methods in model systems where the ground truth is explicitly known. Thus, to obtain full justification for quantification by  $I_{rel}^L$ , the workflows should be systematically benchmarked against riBAQ and other methods such as the total protein approach originally described by Wiśniewski [89–91] (see Supplementary information for further details), using model protein samples digested with a various proteases representing different specificities. Comparison of acquisition methods (e.g. DDA vs. DIA) is also needed, as the stochastic sampling of precursor ions in DDA may be somewhat biased [92]. Nevertheless, based on the good correlations between riBAQ and  $I_{rel}^L$  and the excellent correlation between the three  $I_{rel}^L$  workflows observed, we consider this approach for relative protein quantification a useful approach for identifying particularly abundant proteins in protein extracts.

#### 4. Conclusion

Using de novo transcriptome assembly, we were able to construct a novel protein database for *E. denticulatum*, which was used to characterize two pilot-scale, hot-water extracts from this red seaweed species. Although further processing (extract B) increased protein content significantly (compared to extract A), the aqueous solubility of both was quite low and both extracts appeared to be low quality, displaying a high degree of smear and a lack of distinct protein bands by SDS-PAGE. A slightly alkaline pH and addition of a small amount of detergent fully solubilized the protein assessed by agreement of solubilized protein by Qubit and total protein by Kjeldal-N. The extracts were analyzed using label-free, bottom-up proteomics and quantified as non-standard protein digests via a novel length-normalized relative abundance approach. Using length-normalization for relative quantification was validated against riBAQ for a whole cell *E. coli* lysate, showing good overall correlation. Furthermore, the concept of quality based filtering of low quality identifications was verified in the validation data, showing no substantial impact on correlation or quantitative loss across the different workflows.

Following QBF, the relative protein abundances in the two seaweed extracts correlated quite well. Nevertheless, the analysis yielded a limited number of identifications, and this low data quality also reflects the low protein quality in the extracts, which may limit their applicability. Using subcellular localization prediction, we determined that both extracts appear highly enriched in extracellular protein compared to the expected protein distribution from quantitative transcriptome analysis and estimated protein copy number. In fact, data suggest that

>75% of the quantified seaweed-specific proteins, were constituted by merely three proteins, which were predicted to be extracellular. This was confirmed using several subcellular predictors and observed regardless of using a tryptic or semi-specific quantitative strategy. Extracellular protein enrichment indicates that hot-water extraction is not capable of extracting intracellular proteins, but may be useful for isolation of extracellular protein content on large, industrial scale. In spite of the low data quality, this study illustrates the potential of quantitative proteomics for characterization of extracts to be used as potential sources of novel food protein or bioactive peptides. Furthermore, the results demonstrate the conceptual power of the methodology, particularly in combination with quantitative transcriptomics and bioinformatics, for evaluating extraction methods and for use as a guide in the development and optimization of industrial processes relating to both the extraction itself but also for potential downstream processing.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.algal.2021.102619>.

#### Funding

This work was supported by Innovation Fund Denmark (grant number 7045-00021B (PROVIDE)).

#### Data availability

MaxQuant output files (txt folder) and parameters file (mqpar) can be accessed through the linked Mendeley Data repository [50]. Raw MS data can be made available upon request. Validation data (raw and MaxQuant output files) can be found in a separate Mendeley Data repository [88].

#### CRediT authorship contribution statement

**S.G.:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft preparation, Writing – review and editing, Visualization. **M.P.:** Methodology, Formal analysis, Investigation, Writing – original draft preparation, Writing – review and editing. **P.M.:** Conceptualization, Methodology, Writing – review and editing, Supervision. **S.L.H.:** Writing – original draft preparation, Writing – review and editing. **C.J.:** Writing – original draft preparation, Writing – review and editing, Funding acquisition. **P.J.G-M.:** Methodology, Writing – review and editing. **E.B.H.:** Conceptualization, Writing – review and editing, Project administration, Funding acquisition. **M.T.O.:** Conceptualization, Writing – review and editing, Supervision.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

The authors would like to thank CP Kelco and in particular, Senior Scientist Jimmy Sejberg for supplying the seaweed protein extracts and for fruitful discussions related to analysis and manuscript preparation.

#### References

- [1] S.L. Holdt, S. Kraan, Bioactive compounds in seaweed: functional food applications and legislation, in: *Journal of Applied Phycology* 543–597, Springer, 2011, June 9, <https://doi.org/10.1007/s10811-010-9632-5>.
- [2] A. Leandro, D. Pacheco, J. Cotas, J.C. Marques, L. Pereira, A.M.M. Gonçalves, Seaweed's bioactive candidate compounds to food industry and global food security, in: *Life* 1–37, MDPI AG, 2020, August 1, <https://doi.org/10.3390/life10080140>.
- [3] H. Porse, *The Seaweed Processing Industry*, Seaweed Network of Denmark, 10th Anniversary, Copenhagen, 2018.

- [4] A. Naseri, C. Jacobsen, J.J.P. Sejberg, T.E. Pedersen, J. Larsen, K.M. Hansen, S. L. Holdt, Multi-extraction and quality of protein and carrageenan from commercial spinosum (*Eucheuma denticulatum*), in: *Foods* 9, MDPI AG, 2020, p. 1072, <https://doi.org/10.3390/foods9081072>.
- [5] S. Zhang, X. Zeng, M. Ren, X. Mao, S. Qiao, Novel metabolic and physiological functions of branched chain amino acids: a review, in: *Journal of Animal Science and Biotechnology* 1–12, BioMed Central Ltd., 2017, January 23, <https://doi.org/10.1186/s40104-016-0139-z>.
- [6] R. Peñalver, J.M. Lorenzo, G. Ros, R. Amarowicz, M. Pateiro, G. Nieto, Seaweeds as a functional ingredient for a healthy diet, in: *Marine Drugs* 18, NLM (Medline), 2020, p. 301, <https://doi.org/10.3390/md18060301>.
- [7] A. Gomez-Zavaglia, M.A. Prieto Lage, C. Jimenez-Lopez, J.C. Mejuto, J. Simal-Gandara, The potential of seaweeds as a source of functional ingredients of prebiotic and antioxidant value, in: *Antioxidants* 406, MDPI AG, 2019, September 1, <https://doi.org/10.3390/antiox8090406>.
- [8] H. Admassu, M.A.A. Gasmalla, R. Yang, W. Zhao, Bioactive peptides derived from seaweed protein and their health benefits: antihypertensive, antioxidant, and antidiabetic properties, in: *Journal of Food Science* 83, Blackwell Publishing Inc., 2018, pp. 6–16, <https://doi.org/10.1111/1750-3841.14011>.
- [9] C. Fitzgerald, L. Mora-Soler, E. Gallagher, P. O'Connor, J. Prieto, A. Soler-Vila, M. Hayes, Isolation and characterization of bioactive pro-peptides with in vitro renin inhibitory activities from the macroalga *Palmaria palmata*, in: *Journal of Agricultural and Food Chemistry* 60, American Chemical Society, 2012, pp. 7421–7427, <https://doi.org/10.1021/jf301361c>.
- [10] T. Furuta, Y. Miyabe, H. Yasui, Y. Kinoshita, H. Kishimura, Angiotensin I converting enzyme inhibitory peptides derived from phycobiliproteins of dulce *Palmaria palmata*, in: *Marine Drugs* 14, MDPI AG, 2016, p. 32, <https://doi.org/10.3390/md14020032>.
- [11] R.E. Cian, A.G. Garzón, D.B. Ancona, L.C. Guerrero, S.R. Drago, Hydrolyzates from *Pyropia columbina* seaweed have antiplatelet aggregation, antioxidant and ACE I inhibitory peptides which maintain bioactivity after simulated gastrointestinal digestion, in: *LWT - Food Science and Technology* 64, Academic Press, 2015, pp. 881–888, <https://doi.org/10.1016/j.lwt.2015.06.043>.
- [12] P.A. Harnedy, R.J. Fitzgerald, In vitro assessment of the cardioprotective, anti-diabetic and antioxidant potential of *Palmaria palmata* protein hydrolysates, in: *Journal of Applied Phycology* 25, Springer, 2013, pp. 1793–1803, <https://doi.org/10.1007/s10811-013-0017-4>.
- [13] P.A. Harnedy-Rothwell, C.M. McLaughlin, A.V. Le Gouic, C. Mullen, V. Parthasarathy, P.J. Allsopp, E.M. McSorley, R.J. Fitzgerald, F.P.M. O'Harte, In vitro and in vivo effects of *Palmaria palmata* derived peptides on glucose metabolism, in: *International Journal of Peptide Research and Therapeutics*, Springer Science and Business Media B.V., 2021, pp. 1–10, <https://doi.org/10.1007/s10989-021-10199-8>.
- [14] M. Tu, S. Cheng, W. Lu, M. Du, Advancement and prospects of bioinformatics analysis for studying bioactive peptides from food-derived protein: sequence, structure, and functions, in: *TrAC - Trends in Analytical Chemistry*, Elsevier B.V., 2018, August 1, pp. 7–17, <https://doi.org/10.1016/j.trac.2018.04.005>.
- [15] C. Mooney, N.J. Haslam, G. Pollastri, D.C. Shields, Towards the improved discovery and design of functional peptides: common features of diverse classes permit generalized prediction of bioactivity, in: L. Kurgan (Ed.), *PLoS ONE* 7, Public Library of Science, 2012, <https://doi.org/10.1371/journal.pone.0045012>.
- [16] C. Mooney, N.J. Haslam, T.A. Holton, G. Pollastri, D.C. Shields, PeptideLocator: prediction of bioactive peptides in protein sequences, *Bioinformatics* (Oxford, England) 29 (2013) 1120–1126, <https://doi.org/10.1093/bioinformatics/btt103>.
- [17] T. Panyayai, C. Ngamphiw, S. Tongsim, W. Mhuantong, W. Limsripraphan, K. Choowongkamon, O. Sawatdichai, PeptideDB: a web application for new bioactive peptides from food protein, in: *Heliyon* 5, Elsevier Ltd, 2019, <https://doi.org/10.1016/j.heliyon.2019.e02076>.
- [18] P.J. García-Moreno, C. Jacobsen, P. Marcatili, S. Gregersen, M.T. Overgaard, M. L. Andersen, A.D.M. Sorensen, E.B. Hansen, Emulsifying peptides from potato protein predicted by bioinformatics: stabilization of fish oil-in-water emulsions, in: *Food Hydrocolloids* 101, Elsevier B.V., 2020, <https://doi.org/10.1016/j.foodhyd.2019.105529>.
- [19] T.H. Olsen, B. Yesiltas, F.I. Marin, M. Pertseva, P.J. García-Moreno, S. Gregersen, M.T. Overgaard, C. Jacobsen, O. Lund, E.B. Hansen, P. Marcatili, AnOxPePred: using deep learning for the prediction of antioxidative properties of peptides, in: *Scientific Reports* 10, Nature Research, 2020, p. 21471, <https://doi.org/10.1038/s41598-020-78319-w>.
- [20] P. Minkiewicz, A. Iwaniak, M. Darewicz, BIOPEP-UWM database of bioactive peptides: current opportunities, in: *International Journal of Molecular Sciences* 20, NLM (Medline), 2019, <https://doi.org/10.3390/ijms20235978>.
- [21] C. Chen, Z. Li, H. Huang, B.E. Szezek, C.H. Wu, U. Consortium, A peptide match service for UniProt knowledgebase, *Bioinformatics* 29 (2013) 2808–2809. <https://academic.oup.com/bioinformatics/article/29/21/2808/196283>.
- [22] G. Wang, X. Li, Z. Wang, APD2: the updated antimicrobial peptide database and its application in peptide design, *Nucleic Acids Res.* 37 (2009) D933–D937. [https://academic.oup.com/nar/article/37/suppl\\_1/D933/1013124](https://academic.oup.com/nar/article/37/suppl_1/D933/1013124).
- [23] F. Liu, G. Baggerman, L. Schoofs, G. Wets, The construction of a bioactive peptide database in metazoa, in: *Journal of Proteome Research* 7, American Chemical Society, 2008, pp. 4119–4131, <https://doi.org/10.1021/pr800037n>.
- [24] P.J. García-Moreno, S. Gregersen, E.R. Nedamani, T.H. Olsen, P. Marcatili, M. T. Overgaard, M.L. Andersen, E.B. Hansen, C. Jacobsen, Identification of emulsifier potato peptides by bioinformatics: application to Omega-3 delivery emulsions and release from potato industry side streams, in: *Scientific Reports* 10, Nature Research, 2020, p. 690, <https://doi.org/10.1038/s41598-019-57229-6>.
- [25] A. Jafarpour, R.M. Gomes, S. Gregersen, J.J. Sloth, C. Jacobsen, A.D. Moltke Sørensen, Characterization of cod (*Gadus Morhua*) frame composition and its valorization by enzymatic hydrolysis, in: *Journal of Food Composition and Analysis* 89, Academic Press Inc., 2020, <https://doi.org/10.1016/j.jfca.2020.103469>.
- [26] A. Jafarpour, S. Gregersen, R.M. Gomes, P. Marcatili, T.H. Olsen, C. Jacobsen, M. Toft Overgaard, A.-D.M. Sørensen, Biofunctionality of enzymatically derived peptides from codfish (*Gadus Morhua*) frame: bulk in vitro properties, quantitative proteomics, and bioinformatic prediction, 2020, <https://doi.org/10.20944/preprints202011.0195.v1>. Preprints, Preprints.
- [27] A. Thaha, B.-S. Wang, Y.-W. Chang, S.-M. Hsia, T.-C. Huang, C.-Y. Shiau, D.-F. Hwang, T.-Y. Chen, Food-derived bioactive peptides with antioxidative capacity, xanthine oxidase and tyrosinase inhibitory activity, in: *Processes* 9, Multidisciplinary Digital Publishing Institute, 2021, p. 747, <https://doi.org/10.3390/pr9050747>.
- [28] C. Fernandez-Patron, M. Calero, P.R. Collazo, J.R. Garcia, J. Madrazo, A. Musacchio, F. Soriano, R. Estrada, R. Frank, L.R. Castellanoserra, E. Mendez, Protein reverse staining: high-efficiency microanalysis of unmodified proteins detected on electrophoresis gels, in: *Analytical Biochemistry* 224, Academic Press, 1995, pp. 203–211, <https://doi.org/10.1006/ABIO.1995.1031>.
- [29] J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips, in: *Nature Protocols* 2, Nature Publishing Group, 2007, pp. 1896–1906, <https://doi.org/10.1038/nprot.2007.261>.
- [30] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, in: *Bioinformatics* 30, Oxford University Press, 2014, pp. 2114–2120, <https://doi.org/10.1093/bioinformatics/btu170>.
- [31] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Maudeli, N. Hacohen, A. Gnirke, N. Rhind, F. Di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-seq data without a reference genome, in: *Nature Biotechnology* 29, Nature Publishing Group, 2011, pp. 644–652, <https://doi.org/10.1038/nbt.1883>.
- [32] T. Madden, The BLAST sequence analysis tool, in: *The NCBI Handbook [Internet]*, 2nd Edition, National Center for Biotechnology Information, US, 2013.
- [33] R.D. Finn, J. Clements, S.R. Eddy, HMMER web server: interactive sequence similarity searching, *Nucleic Acids Res.* 39 (2011) W29–W37, <https://doi.org/10.1093/nar/gkr367>.
- [34] R.D. Finn, J. Mistry, J. Tate, P. Coghill, A. Heger, J.E. Pollington, O. Luke Gavin, P. Gunasekaran, G. Ceric, K. Forslund, L. Holm, E.L. L. Sonnhammer, S.R. Eddy, A. Bateman, The Pfam protein families database, *Nucleic Acids Res.* 38 (2010) D211–D222, <https://doi.org/10.1093/nar/gkq985>.
- [35] S. El-Gebali, J. Mistry, A. Bateman, S.R. Eddy, A. Luciani, S.C. Potter, M. Qureshi, L.J. Richardson, G.A. Salazar, A. Smart, E.L.L. Sonnhammer, L. Hirsh, L. Paladın, D. Pivovarov, S.C.E. Tosatto, R.D. Finn, The pfam protein families database in 2019, *Nucleic Acids Res.* 47 (2018) 427–432, <https://doi.org/10.1093/nar/gky995>.
- [36] J.J. Almagro Armenteros, C.K. Sønderby, S.K. Sønderby, H. Nielsen, O. Winther, DeepLoc: prediction of protein subcellular localization using deep learning, in: J. Hancock (Ed.), *Bioinformatics* 33, 2017, pp. 3387–3395, <https://doi.org/10.1093/bioinformatics/btx431>.
- [37] S. Briesemeister, J. Rahnenföhrer, O. Kohlbacher, YLoc—an interpretable web server for predicting subcellular localization, in: *Nucleic Acids Research* 38, Oxford Academic, 2010, pp. W497–W502, <https://doi.org/10.1093/NAR/GKQ477>.
- [38] T. Goldberg, M. Hecht, T. Hamp, T. Karl, G. Yachdav, N. Ahmed, U. Altermann, P. Angerer, S. Ansoerge, K. Balasz, M. Bernhofer, A. Betz, L. Cizmadija, K.T. Do, J. Gerke, R. Greil, V. Joerdens, M. Hastreiter, K. Hembach, M. Herzog, M. Kalemamov, M. Kluge, A. Meier, H. Nasir, U. Neumaier, V. Prade, J. Reeb, A. Sorokoumov, I. Troshani, S. Vorberg, S. Waldraff, J. Zierer, H. Nielsen, B. Rost, LocTree3 prediction of localization, in: *Nucleic Acids Research* 42, Oxford Academic, 2014, pp. W350–W355, <https://doi.org/10.1093/NAR/GKU396>.
- [39] C. Mooney, Y. Wang, G. Pollastri, SCLpred: protein subcellular localization prediction by N-to-1 neural networks, in: *Bioinformatics* 27, Oxford Academic, 2011, pp. 2812–2819, <https://doi.org/10.1093/BIOINFORMATICS/BTR494>.
- [40] J.J.A. Armenteros, K.D. Tsirigos, C.K. Sønderby, T.N. Petersen, O. Winther, S. Brunak, G.von Heijne, H. Nielsen, SignalP 5.0 improves signal peptide predictions using deep neural networks, in: *Nature Biotechnology* 37, Nature Publishing Group, 2019, pp. 420–423, <https://doi.org/10.1038/s41587-019-0036-z>, 2019 37:4.
- [41] S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M.Y. Hein, T. Geiger, M. Mann, J. Cox, The Perseus computational platform for comprehensive analysis of (prote) omics data, in: *Nature Methods* 731–740, Nature Publishing Group, 2016, August 30, <https://doi.org/10.1038/nmeth.3901>.
- [42] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized P.P.B.-range mass accuracies and proteome-wide protein quantification, *Nat. Biotechnol.* 26 (2008) 1367–1372, <https://doi.org/10.1038/nbt.1511>.
- [43] B. Schwanhäusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, M. Selbach, Global quantification of mammalian gene expression control, in: *Nature* 473, Nature Publishing Group, 2011, pp. 337–342, <https://doi.org/10.1038/nature10098>.
- [44] J.B. Shin, J.F. Krey, A. Hassan, Z. Metlagel, A.N. Tauscher, J.M. Pagana, N. E. Sherman, E.D. Jeffery, K.J. Spinelli, H. Zhao, P.A. Wilmarth, D. Choi, L.L. David, M. Auer, P.G. Barr-Gillespie, Molecular architecture of the chick vestibular hair bundle, in: *Nature Neuroscience* 16, NIH Public Access, 2013, pp. 365–374, <https://doi.org/10.1038/nn.3312>.

- [45] H. Kjeldal, L. Pell, A. Pommerening-Röser, J.L. Nielsen, Influence of P-cresol on the proteome of the autotrophic nitrifying bacterium *Nitrosomonas europaea* C91, *Arch. Microbiol.* 3 (2014) 497–511, <https://doi.org/10.1007/s00203-014-0985-z>.
- [46] N.A. Zhou, H. Kjeldal, H.L. Gough, J.L. Nielsen, Identification of putative genes involved in bisphenol A degradation using differential protein abundance analysis of *Sphingobium* sp. BId32, in: *Environmental Science and Technology* 49, American Chemical Society, 2015, pp. 12231–12241, <https://doi.org/10.1021/acs.est.5b02987>.
- [47] S. Gregersen, A.-S.H. Kongsted, R.B. Nielsen, S.S. Hansen, F.A. Lau, J.B. Rasmussen, S.L. Holdt, C. Jacobsen, Enzymatic extraction improves intracellular protein recovery from the industrial carrageenan seaweed *Eucheuma denticulatum* revealed by quantitative, subcellular protein profiling: a high potential source of functional food ingredients, in: *Food Chemistry: X* 12, Elsevier, 2021, <https://doi.org/10.1016/j.FOCHX.2021.100137>.
- [48] S. Tyanova, J. Cox, Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research 133–148, Humana Press, New York, NY, 2018, [https://doi.org/10.1007/978-1-4939-7493-1\\_7](https://doi.org/10.1007/978-1-4939-7493-1_7).
- [49] P. Bardou, J. Mariette, F. Escudié, C. Djemiel, C. Klopp, Jvenn: an interactive Venn diagram viewer, in: *BMC Bioinformatics* 15, BioMed Central Ltd., 2014, p. 293, <https://doi.org/10.1186/1471-2105-15-293>.
- [50] S. Gregersen, E.Denticulatum quant BUP V2. Mendeley data, Mendeley Data V2 (2021), <https://doi.org/10.17632/y4kmb3tvcx.2>.
- [51] M. Rosni, A. Faisal, A. Azwan, F.Y. Chye, P. Matanjun, S. Mohd Rosni, A. Faisal, A. Azwan, F.Y. Chye, P. Matanjun, M. Rosni, A. Faisal, A. Azwan, F.Y. Chye, P. Matanjun, Crude proteins, total soluble proteins, total phenolic contents and SDS-PAGE profile of fifteen varieties of seaweed from Semporna, Sabah, Malaysia, *Int. Food Res. J.* 22 (2015) 1483–1493, <https://search.proquest.com/docview/1698633507/fulltextPDF/CBDD91AFF68E4050PQ/1?accountid=8144>.
- [52] A. Simões-Barbosa, J.M. Santana, A.R.L. Teixeira, Solubilization of delipidated macrophage membrane proteins for analysis by two-dimensional electrophoresis, in: *Electrophoresis* 21, John Wiley & Sons Ltd, 2000, pp. 641–644, [https://doi.org/10.1002/\(SICI\)1522-2683\(20000201\)21:3<641::AID-ELPS641>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1522-2683(20000201)21:3<641::AID-ELPS641>3.0.CO;2-H).
- [53] W. Wang, R. Vignani, M. Scali, E. Sensi, P. Tiberi, M. Cresti, Removal of lipid contaminants by organic solvents from oilseed protein extract prior to electrophoresis, *Anal. Biochem.* 329 (2004) 139–141, <https://doi.org/10.1016/j.ab.2004.02.044>.
- [54] A. Hashimoto, M.A. Pickard, Chloroperoxidases from *Caldariomyces* (= *Leptoxiphium*) cultures: glycoproteins with variable carbohydrate content and isoenzymic forms, *Microbiology* 130 (1984) 2051–2058, <https://doi.org/10.1099/00221287-130-8-2051>.
- [55] H. Chart, B. Rowe, Purification of Lipopolysaccharide from strains of *Yersinia enterocolitica* belonging to serogroups O3 and O9. *FEMS Microbiology Letters*, in: *FEMS Microbiology Letters* 77, Oxford University Press (OUP), 1991, pp. 341–346, <https://doi.org/10.1111/j.1574-6968.1991.tb04373.x>.
- [56] J. Park, K.J. Kim, K.S. Choi, D.J. Grab, J.S. Dumler, Anaplasma phagocytophilum Anka binds to granulocyte DNA and nuclear proteins, in: *Cellular Microbiology* 6, John Wiley & Sons Ltd, 2004, pp. 743–751, <https://doi.org/10.1111/j.1462-5822.2004.00400.x>.
- [57] H.J. Möller, J.H. Poulsen, Staining of glycoproteins/proteoglycans on SDS-gels, in: J.M. Walker (Ed.), *The Protein Protocols Handbook*. Springer Protocols Handbooks, Humana Press, Totowa, NJ, 2009, pp. 569–574, [https://doi.org/10.1007/978-1-59745-198-7\\_52](https://doi.org/10.1007/978-1-59745-198-7_52).
- [58] S. Elliott, J. Egrie, J. Browne, T. Lorenzini, L. Busse, N. Rogers, I. Ponting, Control of RHEUO biological activity: the role of carbohydrate, in: *Experimental Hematology* 32, Elsevier, 2004, pp. 1146–1155, <https://doi.org/10.1016/j.exphem.2004.08.004>.
- [59] K. Späbier, S. Koch, I. Kessler, T. Wenzel, M. Kostrzewa, Selective isolation of glycoproteins and glycopeptides for MALDI-TOF MS detection supported by magnetic particles, in: *Journal of Biomolecular Techniques* 16, The Association of Biomolecular Resource Facilities, 2005, pp. 407–411.
- [60] D.B. Jones, Factors for Converting Percentages of Nitrogen in Foods And Feeds Into Percentages of Protein, US Department of Agriculture, 1931. [https://books.google.dk/books?hl=en&lr=&id=MaM4Z5DjC&oi=fnd&pg=PA9&ots=QAQAi-eXWv&sig=edeemZYHONz\\_m89QRHtgI\\_94oGg&redir\\_esc=y#v=onepage&q&f=false](https://books.google.dk/books?hl=en&lr=&id=MaM4Z5DjC&oi=fnd&pg=PA9&ots=QAQAi-eXWv&sig=edeemZYHONz_m89QRHtgI_94oGg&redir_esc=y#v=onepage&q&f=false).
- [61] P.P. Salo-Väänänen, P.E. Koivisto, Determination of protein in foods: comparison of net protein and crude protein (N x 6.25) values, in: *Food Chemistry* 57, Elsevier Ltd, 1996, pp. 27–31, [https://doi.org/10.1016/0308-8146\(96\)00157-4](https://doi.org/10.1016/0308-8146(96)00157-4).
- [62] F. Mariotti, D. Tomé, P.P. Mirand, Converting nitrogen into protein - beyond 6.25 and Jones' factors, *Crit. Rev. Food Sci. Nutr.* 48 (2008) 177–184, <https://doi.org/10.1080/10408390701279749>.
- [63] A.R. Angell, L. Mata, R. de Nys, N.A. Paul, The protein content of seaweeds: a universal nitrogen-to-protein conversion factor of five, *J. Appl. Phycol.* 28 (2016) 511–524, <https://doi.org/10.1007/s10811-015-0650-1>.
- [64] J. Cox, N. Neuhauser, A. Michalski, R.A. Scheltema, J.V. Olsen, M. Mann, Andromeda: a peptide search engine integrated into the MaxQuant environment, in: *Journal of Proteome Research* 10, American Chemical Society, 2011, pp. 1794–1805, <https://doi.org/10.1021/pr101065j>.
- [65] S. Tyanova, T. Temu, J. Cox, The MaxQuant computational platform for mass spectrometry-based shotgun proteomics, in: *Nature Protocols* 11, Nature Publishing Group, 2016, pp. 2301–2319, <https://doi.org/10.1038/nprot.2016.136>.
- [66] S. Tiwary, R. Levy, P. Gutenbrunner, F. Salinas Soto, K.K. Palaniappan, L. Deming, M. Berndt, A. Brant, P. Cimermancic, J. Cox, High-quality MS/MS spectrum prediction for data-dependent and data-independent acquisition data analysis, in: *Nature Methods* 16, Nature Publishing Group, 2019, pp. 519–525, <https://doi.org/10.1038/s41592-019-0427-6>.
- [67] T. Indarmawan, A.Z. Mustopa, B.R. Budiarto, K. Tarman, Antibacterial activity of extracellular protease isolated from an algiculous fungus *Xylaria psidii* KT30 against gram-positive bacteria HAYATI, in: *Journal of Biosciences* 23, 2016, pp. 73–78, <https://doi.org/10.1016/J.HJB.2016.06.005>. No longer published by Elsevier.
- [68] L.A. Zinke, C. Glombitza, J.T. Bird, H. Røy, B.B. Jørgensen, K.G. Lloyd, J.P. Amend, B.K. Reese, Microbial organic matter degradation potential in Baltic Sea sediments is influenced by depositional conditions and in situ geochemistry, in: *Applied and Environmental Microbiology* 85, American Society for Microbiology, 2019, <https://doi.org/10.1128/AEM.02164-18>.
- [69] P. Naval, T.S. Chandra, Characterization of membrane vesicles secreted by seaweed associated bacterium *Alteromonas macleodii* KS62, in: *Biochemical and Biophysical Research Communications* 514, Academic Press, 2019, pp. 422–427, <https://doi.org/10.1016/J.BBRC.2019.04.148>.
- [70] P.G. Arnisson, M.J. Bibb, G. Bierbaum, A.A. Bowers, T.S. Bugni, G. Bulaj, J. A. Camarero, D.J. Campopiano, G.L. Challis, J. Clardy, P.D. Cotter, D.J. Craik, M. Dawson, E. Dittmann, S. Donadio, P.C. Dorrestein, K.D. Entian, M.A. Fischbach, J.S. Garavelli, U. Göransson, C.W. Gruber, D.H. Haft, T.K. Hemscheidt, C. Hertweck, C. Hill, A.R. Horswill, M. Jaspers, W.L. Kelly, J.P. Klinman, O. P. Kuipers, A.J. Link, W. Liu, M.A. Marahiel, D.A. Mitchell, G.N. Moll, B.S. Moore, R. Müller, S.K. Nair, I.F. Nes, G.E. Norris, B.M. Olivera, H. Onaka, M.L. Patchett, J. Piel, M.J.T. Reaney, S. Rebuffat, R.P. Ross, H.-G. Sahl, E.W. Schmidt, M. E. Selsted, K. Severinov, B. Shen, K. Sivonen, L. Smith, T. Stein, R.D. Süssmuth, J. R. Tagg, G.-L. Tang, A.W. Truman, J.C. Vederas, C.T. Walsh, J.D. Walton, S. C. Wenzel, J.M. Willey, W.A. van der Donk, Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature, in: *Natural Product Reports* 30, Royal Society of Chemistry, 2012, pp. 108–160, <https://doi.org/10.1039/C2NP20085F>.
- [71] R. Sable, P. Parajuli, S. Jois, Peptides, peptidomimetics, and polypeptides from marine sources: a wealth of natural sources for pharmaceutical applications, in: *Marine Drugs* 15, Multidisciplinary Digital Publishing Institute, 2017, p. 124, <https://doi.org/10.3390/MD15040124>.
- [72] M. Friedman, M.R. Gumbmann, P.M. Masters, Protein-alkali reactions: chemistry, toxicology, and nutritional consequences, in: *Advances in Experimental Medicine and Biology* 177, Springer, Boston, MA, 1984, pp. 367–412, [https://doi.org/10.1007/978-1-4684-4790-3\\_18](https://doi.org/10.1007/978-1-4684-4790-3_18).
- [73] X. Chen, W. Fu, Y. Luo, C. Cui, I. Supparorasatit, L. Liang, Protein deamidation to produce processable ingredients and engineered colloids for emerging food applications, in: *Comprehensive Reviews in Food Science and Food Safety* 20, John Wiley & Sons Ltd, 2021, pp. 3788–3817, <https://doi.org/10.1111/1541-4337.12759>.
- [74] A.T. Kong, F.V. Leprevost, D.M. Avtonomov, D. Mellacheruvu, A.I. Nesvizhskii, MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics, in: *Nature Methods* 14, Nature Publishing Group, 2017, pp. 513–520, <https://doi.org/10.1038/nmeth.4256>.
- [75] D.L. Swaney, C.D. Wenger, J.J. Coon, Value of using multiple proteases for large-scale mass spectrometry-based proteomics, *J. Proteome Res.* 9 (2010) 1323–1329, <https://doi.org/10.1021/pr900863u>.
- [76] J. Fleurence, C. Le Coeur, S. Mabeau, M. Maurice, A. Landrein, Comparison of different extractive procedures for proteins from the edible seaweeds *Ulva rigida* and *Ulva rotundata*, in: *Journal of Applied Phycology* 7, Kluwer Academic Publishers, 1995, pp. 577–582, <https://doi.org/10.1007/BF00003945>.
- [77] M. Bjarnadóttir, B.V. Aðalbjörnsson, A. Nilsson, R. Sliżyte, M.Y. Røleda, G.O. Hreggviðsson, Ö.H. Friðjónsson, R. Jónsdóttir, *Palmaria palmata* as an alternative protein source: enzymatic protein extraction, amino acid composition, and nitrogen-to-protein conversion factor, in: *Journal of Applied Phycology* 30, Springer, Netherlands, 2018, pp. 2061–2070, <https://doi.org/10.1007/s10811-017-1351-8>.
- [78] M. Herrero, A. del P. Sánchez-Camargo, A. Cifuentes, E. Ibáñez, Plants, seaweeds, microalgae and food by-products as natural sources of functional ingredients obtained using pressurized liquid extraction and supercritical fluid extraction, in: *TrAC - Trends in Analytical Chemistry*, Elsevier B.V., 2015, September 1, pp. 26–38, <https://doi.org/10.1016/j.trac.2015.01.018>.
- [79] P.A. Harnedy, R.J. FitzGerald, Extraction of protein from the macroalga *Palmaria palmata*, in: *LWT - Food Science and Technology* 51, Academic Press, 2013, pp. 375–382, <https://doi.org/10.1016/j.lwt.2012.09.023>.
- [80] H.K. Maehre, G.K. Edvinsen, K.E. Eilertsen, E.O. Ellevevold, Heat treatment increases the protein bioaccessibility in the red seaweed dulse (*Palmaria palmata*), but not in the brown seaweed winged kelp (*Alaria esculenta*), in: *Journal of Applied Phycology* 28, Springer, Netherlands, 2016, pp. 581–590, <https://doi.org/10.1007/s10811-015-0587-4>.
- [81] M. Magnusson, C.R.K. Glasson, M.J. Vucko, A. Angell, T.L. Neoh, R. de Nys, Enrichment processes for the production of high-protein feed from the green seaweed *Ulva ohnoi*, in: *Algal Research* 41, Elsevier B.V., 2019, <https://doi.org/10.1016/j.algal.2019.101555>.
- [82] S. Bleakley, M. Hayes, Algal proteins: extraction, application, and challenges concerning production, in: *Foods* 6, MDPI AG, 2017, p. 33, <https://doi.org/10.3390/foods6050033>.
- [83] M. Cermeño, T. Kleekayai, M. Amigo-Benavent, P. Harnedy-Rothwell, R. J. FitzGerald, Current knowledge on the extraction, purification, identification, and validation of bioactive peptides from seaweed, in: *ELECTROPHORESIS*, John Wiley & Sons Ltd, 2020, <https://doi.org/10.1002/elps.202000153>.

- [84] A. Naseri, G.S. Marinho, S.L. Holdt, J.M. Bartela, C. Jacobsen, Enzyme-assisted extraction and characterization of protein from red seaweed *Palmaria palmata*, *Algal Res.* 47 (2020), 101849, <https://doi.org/10.1016/j.algal.2020.101849>.
- [85] K. Hardouin, G. Bedoux, A.S. Burlot, C. Donnay-Moreno, J.P. Bergé, P. Nyvall-Collén, N. Bourgougnon, Enzyme-assisted extraction (EAE) for the production of antiviral and antioxidant extracts from the green seaweed *Ulva armoricana* (Ulvales, Ulvophyceae), in: *Algal Research* 16, Elsevier, 2016, pp. 233–239, <https://doi.org/10.1016/j.algal.2016.03.013>.
- [86] V. Vásquez, R. Martínez, C. Bernal, Enzyme-assisted extraction of proteins from the seaweeds *Macrocystis pyrifera* and *Chondracanthus chamissoi*: characterization of the extracts and their bioactive potential, in: *Journal of Applied Phycology* 31, Springer, Netherlands, 2019, pp. 1999–2010, <https://doi.org/10.1007/s10811-018-1712-y>.
- [87] N. Terme, K. Hardouin, H.P. Cortès, A. Peñuela, Y. Freile-Pelegrín, D. Robledo, G. Bedoux, N. Bourgougnon, Emerging seaweed extraction techniques: enzyme-assisted extraction a key step of seaweed biorefinery?, in: *Sustainable Seaweed Technologies* 225–256 Elsevier, 2020 <https://doi.org/10.1016/b978-0-12-817943-7.00009-3>.
- [88] S. Gregersen, E.Denticulatum quant method validation, in: *Mendeley Data*, 2021, <https://doi.org/10.17632/c8hkst76t4.1>.
- [89] J.R. Wiśniewski, Label-free and standard-free absolute quantitative proteomics using the “Total protein” and “Proteomic ruler” approaches, in: *Methods in Enzymology*, Academic Press Inc., 2017, pp. 49–60, <https://doi.org/10.1016/bs.mie.2016.10.002>.
- [90] J.R. Wiśniewski, C. Wegler, P. Artursson, Multiple-enzyme-digestion strategy improves accuracy and sensitivity of label- and standard-free absolute quantification to a level that is achievable by analysis with stable isotope-labeled standard spiking, *J. Proteome Res.* 18 (2018), acs.jproteome.8b00549, <https://doi.org/10.1021/acs.jproteome.8b00549>.
- [91] J.R. Wiśniewski, P. Ostasiewicz, K. Duś, D.F. Zielińska, F. Gnad, M. Mann, Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma, *Mol. Syst. Biol.* 8 (2012), <https://doi.org/10.1038/msb.2012.44>.
- [92] B. Van Puyvelde, S. Willems, R. Gabriels, S. Daled, L. De Clerck, S. Vande Castele, A. Staes, F. Impens, D. Deforce, L. Martens, S. Degroev, M. Dhaenens, Removing the hidden data dependency of DIA with predicted spectral libraries, in: *PROTEOMICS* 20, Wiley-VCH Verlag, 2020, p. 1900306, <https://doi.org/10.1002/pmic.201900306>.