UNIVERSITY OF COPENHAGEN

FACULTY OF SCIENSE

DEPARTMENT OF FOOD SCIENCE

Data, Models, and Meaning:

Structure–Function Insight in Carrageenan via PAT

Oksana Mykhalevych

PhD Thesis

2025

Data, Models, and Meaning:

Advancing Structure–Function Insight in Carrageenan via PAT

PhD Thesis

Oksana Mykhalevych

This thesis has been submitted to the PhD School of The Faculty of Science,

University of Copenhagen

University of Copenhagen

Department of Food Science,

Faculty of Science,

Rolighedsvej 26, DK-1958 Frederiksberg

Denmark

Title	Data, Models, and Meaning: Advancing Structure–Function Insight in Carrageenan via PAT
Author	Oksana Mykhalevych
Submission date	31 May 2025
PhD defense	29 August 2025
Principal supervisor	Rasmus Bro Professor Department of Food Science University of Copenhagen Denmark
Co-supervisor	Henrik Stapelfeldt Senior Manager, PhD, Tate & Lyle Denmark

Assessment Committee

Chairperson	Åsmund Rinan Associate Professor Department of Food Science University of Copenhagen Denmark
External	Ingrid Måge Senior Scientist, PhD NOFIMA Norway
External	Alessandra Biancolillo Fixed-term researcher type B (RTDB), PhD Department of Physical and Chemical Sciences The University of L'Aquila Italy

PREFACE

This Industrial PhD thesis is submitted in partial fulfilment of the requirements for the PhD degree at the University of Copenhagen. The project was carried out under the Industrial PhD program with financial support from Innovation Fund Denmark (grant number 104400009B).

The research presented herein was conducted at CP Kelco ApS (Tate & Lyle as of November 1, 2024), utilizing exclusively company-provided samples and equipment. The project reflects the collaborative framework between academia and industry aimed at advancing scientific knowledge with practical application.

ACKNOWLEDGEMENTS

Coming from a small village in western Ukraine, I could never have imagined that one day I would be pursuing a PhD at one of the world's top universities and collaborating with a leading global company in ingredient production. What once felt like an unattainable dream has become a reality, thanks to the incredible people I have met along the way. Their guidance, support, and encouragement have shaped my journey and brought me to where I am today. I would like to express my heartfelt gratitude to them.

First, I want to thank my principal supervisor, Professor Rasmus Bro. Thank you for your guidance, mentorship, and friendship. I truly appreciate how you allowed me to be independent in my work. I am grateful for all your time and I'm looking forward to future industrial collaboration.

I would also like to express my sincere gratitude to my company supervisor, Dr. Henrik Stapelfeldt. I have greatly enjoyed our professional discussions and learned so much from them. Thank you for your guidance, always constructive feedback, and for teaching me so much about carrageenan, QC and beyond.

A special thank you to my previous supervisors, Dr. Jan Larsen and Dr. Astrid Benie. Despite the inevitable frustrations and moments of confusion that come with a PhD, I always felt supported and motivated by your guidance. Thank you for generously sharing your deep knowledge and experience with me and for supporting the beginning of this work.

Furthermore, I want to thank my company advisor, Dr. Thomas Worm, for sharing your extensive knowledge of carrageenan and for our insightful conversations about career and life.

Finally, I extend my heartfelt thanks to everyone at Tate & Lyle in Skensved for giving me the opportunity to work on this project alongside some of the most knowledgeable experts in the field of ingredient production. It has been a privilege to collaborate with such an outstanding team.

I am also deeply grateful to my collaborators along the way. I would like to express my gratitude to Professor Federico Marini from the University of Rome "La Sapienza" for sharing valuable tools in data fusion and providing guidance on their application. Additionally, my thanks go to the NMR experts: Professor Thomas Vosegaard and Dr. Armin Afrough from Aarhus University, as well as Professor Moreno Lelli and Dr. Lorenzo Niccoli from the University of Florence, for their support, guidance, and innovative ideas in NMR experiments and data analysis. I also deeply appreciate the enriching learning experience during my visit to the Magnetic Resonance Center, University of Florence.

For me, the most valuable part of the PhD journey has been the people I've met—many of whom I hope will become lifelong friends. Thank you to the "DCB/IDT gang" for all the great times. A special thanks to Bea, Paul, Jason, Ilona, Maxime, Cecilie, Tjark, and Ching for our deep conversations about everything from publishing papers to parenthood.

Susie and Ingrid, they say that if a friendship lasts seven years, it will last forever—well, here we are at ten years and counting! I'm beyond grateful for both of you, for standing by my side through the highs and lows of this PhD journey. I'm looking forward to our trip to London!

Галю, дякую тобі за те, що завжди готова вирушити зі мною в нову пригоду. А тобі, Віко, дякую за те, що завжди поруч — варто лише подзвонити, і ти вже тут. Я безмежно вдячна вам обом!

And finally, to the most important people in my life: Alex, thank you for your unwavering support in everything I do, and for always fresh flowers in my office. My dear boy, Theo, thank you for your smiles and laughter during the writing period. I hope that one day, when you skim through this work, you will be a little bit proud of your mama.

ABSTRACT

Carrageenan is a complex natural polysaccharide extracted from red seaweed, widely used in food and pharmaceutical applications for its gelling and thickening properties. Due to its structural complexity, the relationship between carrageenan structure and functionality in food systems remains insufficiently understood. This Industrial PhD project aimed to develop chemometric models that link measurable molecular characteristics of carrageenan to its functional properties in food systems, thereby supporting the implementation of Process Analytical Technology and Quality by Design strategies in carrageenan production.

A large set of commercial carrageenan samples was analyzed using four complementary analytical techniques: size exclusion chromatography with multi-angle light scattering and viscometry (SEC-MALS) to measure molecular size and intrinsic viscosity and related parameters; Fourier-transform infrared spectroscopy (FT-IR) and proton nuclear magnetic resonance to capture functional group composition; and inductively coupled plasma mass spectrometry to quantify associated cations.

Carrageenan functionality was tested in three food-relevant systems: viscosity in aqueous solution, gel and breaking strength in milk and suspension stabilization. Viscosity was best predicted using SEC-MALS data coupled with Partial Least Squares regression, with intrinsic viscosity and hydrodynamic radius as primary predictors. Gel and breaking strength were most accurately predicted using FT-IR spectra combined with Support Vector Machine regression. The third functionality, suspension stabilization by carrageenan, was explored using design of experiment. It was applied to optimize carrageenan blend composition for stabilizing chocolate milk. The experimental results were analyzed, and a formulation tool was proposed to support blend design.

Multiblock data fusion techniques were used to combine data from the four analytical platforms. While these methods did not improve predictive performance, they offered insight into how each platform contributes uniquely or redundantly to modeling structure–function relationships.

Two predictive models were implemented in an industrial quality control setting: one for viscosity using SEC-MALS data, and one for milk-carrageenan gel and breaking strength using FT-IR spectra. These models demonstrated that carrageenan functionality can be predicted with sufficient accuracy to support faster and more efficient quality control. In addition, the study shows how chemometric modeling and combined analytical data can provide insight into the molecular drivers of carrageenan performance in food systems.

RESUME (DANISH ABSTRACT)

Carrageenan er et komplekst naturligt polysaccharid, der udvindes fra rødalger og anvendes bredt i fødevare- og medicinalindustrien på grund af dets evne til at danne gel og virke som fortykningsmiddel. På grund af dets strukturelle kompleksitet er sammenhængen mellem carrageenans struktur og funktionalitet i fødevaresystemer stadig utilstrækkeligt forstået. Dette industrielle ph.d.-projekt havde til formål at udvikle kemometriske modeller, der forbinder målbare molekylære egenskaber af carrageenan med dets funktionelle adfærd i fødevaresystemer og dermed understøtter implementeringen af Process Analytical Technology (PAT) og Quality by Design (QbD) i carrageenan-produktion.

Et stort datasæt af kommercielle carrageenan prøver blev analyseret ved hjælp af fire komplementære analytiske teknikker: størrelseseksklusionskromatografi med multi-vinkel lysspredning og viskometri (SEC-MALS) til måling af molekylstørrelse og intrinsisk viskositet; Fourier-transform infrarød spektroskopi (FT-IR) og proton nuklear magnetisk resonans (^{1}H NMR) til bestemmelse af funktionelle grupper og bindingstyper; samt induktivt koblet plasma massespektrometri til kvantificering af tilhørende kationer.

Carrageenans funktionalitet blev testet i tre fødevare-relevante systemer: viskositet i vandige opløsninger, gel- og brudstyrke i mælk samt stabilisering af suspensioner. Viskositet blev bedst forudsagt med SEC-MALS data kombineret med Partial Least Squares-regression, hvor intrinsisk viskositet og hydrodynamisk radius var de primære forklarende variabler. Gel- og brudstyrke blev mest præcist forudsagt ud fra FT-IR-spektre kombineret med Support Vector Machine-regression. Den tredje funktion, suspensionstabilisering med carrageenan, blev undersøgt ved hjælp af forsøgsplanlægning (Design of Experiments). Metoden blev anvendt til at optimere blandingssammensætningen af carrageenan til stabilisering af kakaomælk. De eksperimentelle resultater blev analyseret, og et formuleringsværktøj blev foreslået til støtte for blenddesign.

Multiblok-datasammenfletning blev anvendt til at kombinere data fra de fire analytiske platforme. Selvom disse metoder ikke forbedrede forudsigelsesnøjagtigheden, gav de indsigt i, hvordan hver platform bidrager unikt eller overlappende til modellering af struktur–funktionsforhold.

To prædiktive modeller blev implementeret i industriel kvalitetskontrol: én til viskositet baseret på SEC-MALS data og én til gel- og brudstyrke i mælkecarrageenan baseret på FT-IR-spektre. Disse modeller viste, at carrageenans funktionalitet kan forudsiges med tilstrækkelig nøjagtighed til at understøtte hurtigere og mere effektiv kvalitetskontrol. Desuden viser studiet, hvordan kemometrisk modellering og kombinerede analytiske data kan give indsigt i de molekylære mekanismer bag carrageenans funktion i fødevaresystemer.

АНОТАЦІЯ (UKRAINIAN PREFACE AND ABSTRACT)

Ця промислова докторська дисертація під назвою "Дані, моделі та значення: Структурно-функціональний аналіз карагінану за допомогою технології аналітичного контролю процесу" подана на здобуття наукового ступеня доктора філософії в Університеті Копенгагена. Проєкт був реалізований у рамках програми промислових докторських досліджень за фінансової підтримки Фонду інновацій Данії . Представлені дослідження були проведені на підприємстві з використанням виключно зразків та обладнання, наданих компанією CP Kelco (Tate & Lyle 1.11.2024). Проєкт відображає співпрацю між академічними колами та промисловістю, спрямовану на розвиток наукових знань із практичним застосуванням.

Каррагінан — це складний природний полісахарид, який отримують із червоних водоростей і широко застосовують у харчовій та фармацевтичній промисловості завдяки його здатності до гелеутворення та загущення. Через його складну структуру взаємозв'язок між будовою каррагінану та його функціональністю в харчових системах залишається недостатньо вивченим. Цей промисловий докторський проєкт мав на меті розробити хемометричні моделі, що пов'язують вимірювані молекулярні характеристики каррагінану з його функціональною поведінкою в харчових системах, тим самим сприяючи впровадженню технології аналітичного контролю процесів (PAT) та концепції «Проєктування якості» (QbD) у виробництві каррагінану.

Великий набір зразків комерційного каррагінану було проаналізовано за допомогою чотирьох взаємодоповнюючих аналітичних методів: гель-проникної хроматографії з багатокутовим розсіюванням світла та віскометрією (SEC-MALS) для визначення молекулярного розміру та внутрішньої в'язкості; інфрачервоної спектроскопії з перетворенням Фур'є (FT-IR) та протонного ядерного магнітного резонансу (^{1}H NMR) для аналізу складу функціональних груп і типів хімічних зв'язків; а також мас-спектрометрії з індуктивно зв'язаною плазмою (ICP-MS) для кількісного визначення супровідних катіонів. Ці методи надали структурні дескриптори, що мають значення для застосування каррагінану.

Функціональність каррагінану досліджувалася у трьох харчових системах: в'язкість у водному розчині, гелеутворення та міцність на розрив у молоці, а також стабілізація суспензій. В'язкість найкраще передбачалася за допомогою даних SEC-MALS у поєднанні з регресією часткових найменших квадратів (PLS), де основними предикторами були внутрішня в'язкість і гідродинамічний радіус. Гелеутворення та міцність на розрив найточніше прогнозувалися на основі спектрів FT-IR із використанням моделі опорних векторів (SVM). Третю функцію — стабілізацію суспензій — досліджували за допомогою планування експерименту. Метод було застосовано для оптимізації складу сумішей каррагінану для стабілізації шоколадного молока. Отримані експериментальні дані було проаналізовано, і запропоновано інструмент для проєктування таких сумішей.

Методи багатоблокового об'єднання даних використовувалися для комбінування інформації з чотирьох аналітичних платформ. Хоча ці методи не покращили точність прогнозування, вони надали розуміння того, як кожна платформа робить унікальний або частково перекривний внесок у моделювання взаємозв'язків між структурою та функціональністю.

Дві прогностичні моделі були впроваджені в промислову систему контролю якості: одну для в’язкості — на основі даних SEC-MALS, іншу — для гелеутворення та міцності на розрив каррагінану в молоці на основі спектрів FT-IR. Ці моделі показали, що функціональні властивості каррагінану можна достовірно передбачити для забезпечення швидшого та ефективнішого контролю якості. Дослідження також демонструє, як хемометричне моделювання та поєднання аналітичних даних дають уявлення про молекулярні чинники, що визначають ефективність каррагінану в харчових системах.

ABBREVIATIONS

[η](avg)	Uncertainty-weighted average intrinsic viscosity
[η]w	Weight average intrinsic viscosity
COW	Correlation Optimized Warping
dn/dc	specific refractive index increment
DoE	Design of Experiments
FCC	Food Chemical Codex
FT-IR	Fourier-transform infrared spectroscopy
HPLC	High-Performance Liquid Chromatography
HSQC	Heteronuclear single quantum coherence spectroscopy
ICP-MS	Inductively coupled plasma mass spectrometry
LWR	Locally Weighted Regression
MB-PLS	Multiblock PLS
M(avg)	Uncertainty-weighted average molar mass
Mn	Number average molecular mass
Mp	Molar mass at the top of the concentration peak
Mw	Weight-average molecular mass
Mz	Z-average molar mass
NMR	Nuclear magnetic resonance
PAT	Process Analytical Technology
PCA	Principal Component Analysis
PLS	Partial Least Squares
PO-PLS	Parallel Orthogonalized PLS
RBF	Radial Basis Function
rh(v)(avg)	Uncertainty weighted average hydrodynamic radius
rh(v)w	Weight average hydrodynamic radius
RMSE	Root Mean Square Error
RMSECV	Root Mean Square Error of Cross-Validation
RMSEP	Root Mean Square Error of Prediction
ROSA	Response Oriented Sequential Alteration
rw	Weight average mean square radius

SEC-MALS	Size-exclusion chromatography with multi-angle light scattering detection
SNV	Standard Normal Variate scaling
SO-PLS	Sequential and Orthogonalized PLS
SVM	Support Vector Machines
QbD	Quality by Design
UHT	Ultra-high-temperature process

Declaration of Generative AI Use

During the writing of this thesis, I used OpenAI's ChatGPT to help improve the readability and phrasing of certain sections. I reviewed and edited all suggestions myself and take full responsibility for the final content.

LIST OF PAPERS

Paper I

O. Mykhalevych, H. Stapelfeldt, R. Bro, *Chemometric approaches for polysaccharide viscosity profiling*, Carbohydrate Polymers **348** (2025) 122824.

https://doi.org/10.1016/J.CARBPOL.2024.122824

Paper II

O. Mykhalevych, H. Stapelfeldt, F. Marini, R. Bro, Chemometric insights into milk-carrageenan breaking and gel strength, Food Hydrocolloids **158** (2025) 110544.
https://doi.org/10.1016/J.FOODHYD.2024.110544

Paper III

O. Mykhalevych, H. Stapelfeldt, R. Bro, *Optimizing Carrageenan Blends for Chocolate Milk Using Combined Factorial and Mixture Design*, Submitted to LWT – Food Science and Technology.

Poster

Oksana Mykhalevych, Rasmus Bro, Astrid Benie, Henrik Stapelfeldt and Federico Marini, *Data fusion for prediction of gel and breaking strength for milk gels of carrageenan*, presented at the 18th Scandinavian Symposium on Chemometrics, Gothenburg, 2023.

Best Poster Award

Additional publications

Confidential industry reports:

- Characterization of a unit operation in the carrageenan process and proposal for a PAT-based QC solution
- Implementation of chemometric models in the QC laboratory with user guidelines
- Estimation of sampling error in production settings
- Method validation for carrageenan application tests

Table of Contents

1 GENERAL INTRODUCTION

This Ph.D. project was initiated by CP Kelco ApS (now part of Tate & Lyle), a global producer of carrageenan and other hydrocolloids. The company operates a modern production facility in Denmark and is committed to increasing both the efficiency and sustainability of its manufacturing processes. In line with this vision, this project was initiated to propose a Process Analytical Technology (PAT) framework specifically tailored to carrageenan production.

Traditionally, quality control of carrageenan products has relied heavily on laboratory testing of the final product. While effective in verifying product compliance, this quality-by-testing approach offers limited insight into the upstream factors that influence product quality. As a result, it provides little support for process control or proactive decision-making. To initiate the development of PAT in carrageenan manufacturing, this project began by focusing on at-line measurements (in laboratory on the production site) as a practical and accessible entry point.

1.1 AIMS AND OBJECTIVES OF THE STUDY

The overall aim of this project is to support the implementation of PAT and quality-by-design (QbD) principles in carrageenan manufacturing by developing advanced chemometric tools and as part of elucidating the structure-function relationship for carrageenan.

The work focuses on modeling the structure–function relationship of carrageenan products using chemometrics, with an emphasis on food technology applications, despite the broader industrial applicability of carrageenan.

Specifically, the objectives are to:

1. Comprehensively describe carrageenan structure using relevant molecular parameters
2. Relate structural characteristics to functional performance through predictive modeling, and
3. Implement predictive models for carrageenan quality control.

To address these objectives, the following research hypotheses are formulated:

1. Carrageenan structure can be comprehensively described by molecular characteristics measured by proton nuclear magnetic resonance (^{1}H NMR) spectroscopy and Fourier Transform Infrared (FT-IR) spectroscopy, inductively coupled with plasma mass spectrometry (ICP-MS) and Size-Exclusion Chromatography coupled with Multi-Angle Light Scattering (SEC-MALS). These parameters provide sufficient chemical and physical insight to capture structural variability in commercial products.
2. The functional properties of carrageenan can be modeled from the data obtained by defined analytical platform and modelled by chemometrics based techniques. These models allow quantification of the relationship between structure and functionality.
3. The obtained predictive models can be implemented for carrageenan quality control.

1.2 OUTLINE

This thesis describes the development of predictive models for carrageenan manufacturing, with a focus on understanding and modeling the relationship between molecular structure and functional performance. Published and in-review papers (Paper 1, Paper 2 and Paper 3), provided as supplementary material, are referenced throughout the text where relevant, and redundancy has been minimized. Theoretical background is included only when essential for understanding the results; referenced provided to detailed explanations of analytical instrumentation and chemometric theory.

In alignment with the vision of PAT, the thesis is structured as a progressive journey—from empirical measurement toward actionable understanding and practical implementation. This structure is inspired by the Knowledge Pyramid (also known as the DIKW hierarchy: Data, Information, Knowledge, Wisdom), a framework commonly used in systems and information sciences to describe the transformation of raw data into useful decision-making tools [1].

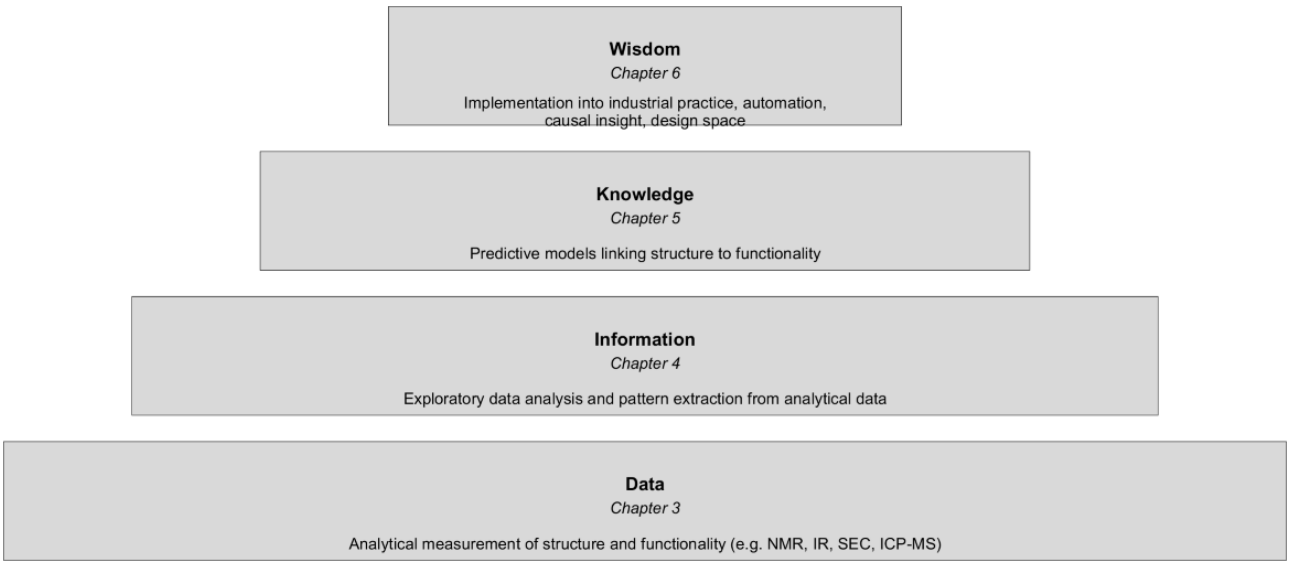


Figure 1 Thesis structure visualized in for of PAT knowledge pyramid.

In the context of PAT, this progression reflects the essential steps of the framework: data acquisition, through preprocessing and modeling, to the implementation of predictive control strategies. This structure not only mirrors the flow of research in this PhD thesis but also emphasizes the increasing complexity and value of insights at each stage. The adapted PAT-specific knowledge pyramid is illustrated in Figure 1 and defines the structure of this thesis as follows:

- Data – Selection and validation of analytical instrumentation to measure molecular structure and functionality of carrageenan.
- Information – Application of exploratory chemometric techniques to identify and explain meaningful patterns.
- Knowledge – Development of predictive models that link structural information to functionality.
- Wisdom – Integration of models into industrial practice enabling implementation of PAT frameworks.

2 BACKGROUND AND STATE-OF-THE-ART

2.1 CARRAGEENAN

Carrageenan is a complex sulfated polysaccharide extracted from red seaweed (Rhodophyta) that plays a fundamental role in maintaining structural integrity of the plant [2,3]. In the food industry, carrageenan is widely employed as a gelling, thickening, and stabilizing agent, making it a versatile hydrocolloid in various food systems.

Historically, the gelling properties of carrageenan were recognized centuries ago, particularly in the preparation of traditional foods using *Chondrus crispus* (Irish moss), a seaweed endemic to the Irish coast. The name "carrageenan" is derived from the village of Carragheen near Waterford, Ireland, where *Chondrus crispus* grows abundantly and has been utilized for its gelling properties [4,5]. An example of such early use is documented by Smith (1905), who provided a recipe in which dried Irish moss was soaked in water, boiled in milk, and flavored to produce a dessert gel [6]:

> *"Soak half a cup of dry moss in cold water for five minutes, tie in a cheesecloth bag, place in a double boiler with a quart of milk and cook for half an hour; add half a teaspoonful of salt or less, according to taste, strain, flavor with a teaspoonful of lemon or vanilla extract as desired, and pour into a mold or small cups, which have been wet with cold water; after hardening, eat with sugar and cream."* – Smith, 1905 [6].

Today, carrageenan is commercially extracted from a variety of red seaweed species in addition to *Chondrus crispus.* Major sources include *Kappaphycus alvarezii*, commonly known as Cottonii (trade name: *Eucheuma cottonii*), and Eucheuma denticulatum, referred to as Spinosum (trade name: *Eucheuma spinosum*). Other important sources include *Mastocarpus stellatus*, previously classified as *Gigartina stellata*, as well as various other species within the *Gigartina* and *Chondrus* genera [7–9]. The global carrageenan market is experiencing steady growth, driven by increasing interest in natural and functional ingredients for food applications [10].

2.1.1 Structure

The term *carrageenan* refers to a family of structurally related sulfated galactans, rather than a single, uniform polysaccharide. These compounds share a common linear backbone consisting of alternating units of 3-linked β-D-galactose and 4-linked α-D-galactose [11]. Disaccharide repeating unit structure of carrageenans are schematically depicted with letter codes assigned to each residue proposed by Knutsen et al. [12] in Figure 2 and the nomenclature is explained in Table 1. Classification of carrageenans traditionally employs Greek letter prefixes, which reflect structural differences such as the presence or absence of a 3,6-anhydrobridge at the 4-linked α-D-galactose

residue, as well as the number and position of sulfate ester groups. The term “anhydro” indicates the elimination of a water molecule between two hydroxyl groups, resulting in the formation of a cyclic ether [11]. The six carrageenan types depicted in Figure 2 are the most known but not all existing carrageenans [13–15]. In this work, the three commercial carrageenans ι-, κ-, and λ-carrageenans are of most importance as their structure defines their application in food.

Carrageenans are high-molecular weight polydisperse hydrocolloids with molecular weight ranging from 200 kDa to 2000 kDa and polydispersity of 1.2 – 3 [15–17]. Carrageenans negatively charged sulfate ester groups are balanced by positively charged ions most commonly by Ca^{2+}, K^{+}, Na^{+}, and Mg^{2+} [15,18,19]. This large and complex structure results in many structural variations and thereby in a broad range of functionalities.

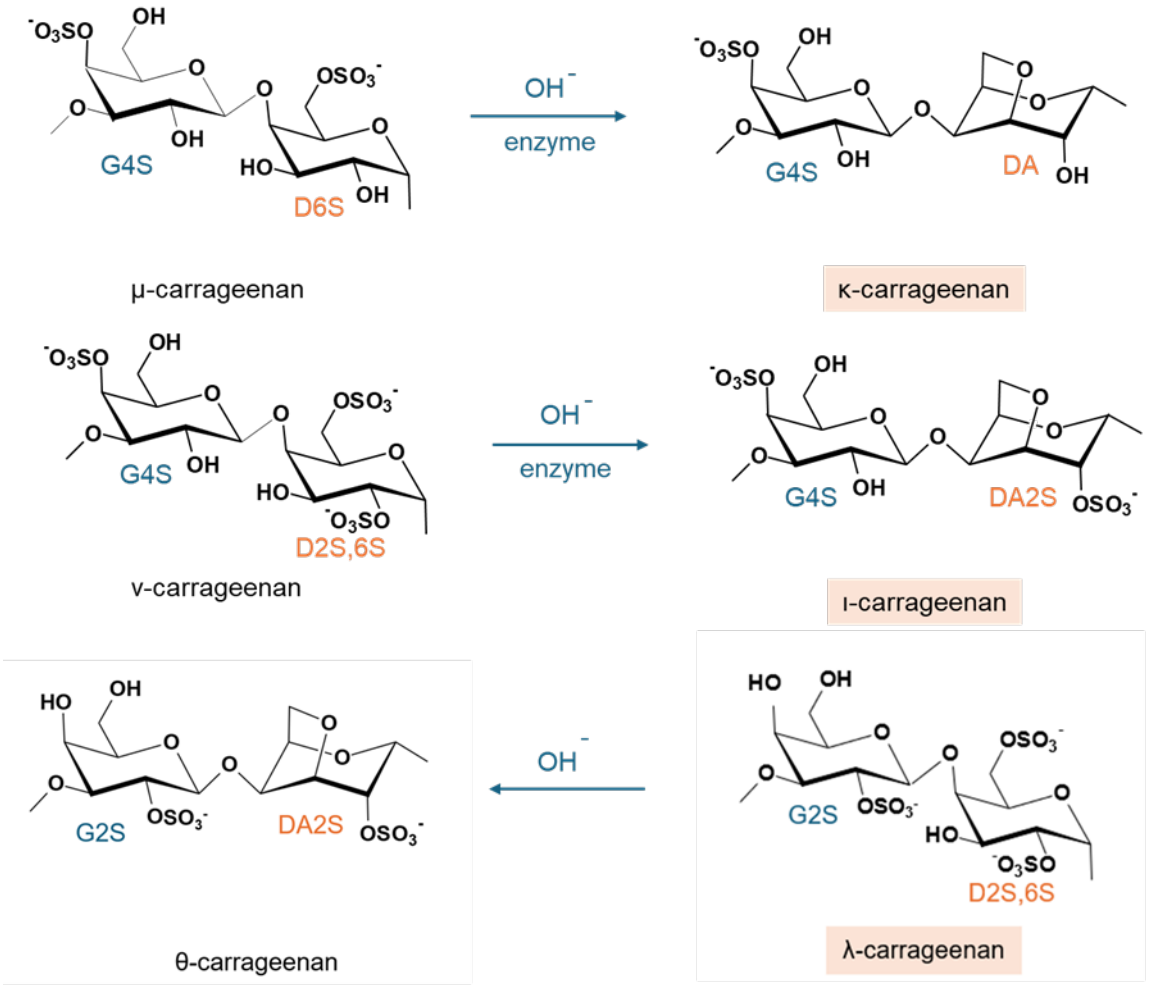


Figure 2 Reaction scheme of μ-carrageenan to κ-carrageenan, of ν-carrageenan to ι-carrageenan and of λ-carrageenan to θ-carrageenan, with idealized disaccharide repeating unit structures of carrageenans. μ- and ν-carrageenans are so-called precursors that found in seaweed before processing and small amount in commercial carrageenan. Θ-carrageenan is also not applied commercially [2,20].

Table 1 Carrageenan nomenclature explained [12,15,20].

Prefix	Letter Code	IUPAC names
μ	G4S-D6S	Carrageenose 4, 6′- disulfate
κ	G4S-DA	Carrageenose 4’- sulfate
ν	G4S-D2S,6S	Carrageenose 4, 2′,6′- trisulfate
ι	G4S-DA2S	Carrageenose 2,4’- disulfate
θ	G2S-DA2S	Carrageenose 2, 2′-sulfate
λ	G2S-D2S,6S	Carrageenose 2,6,2’ - trisulfate

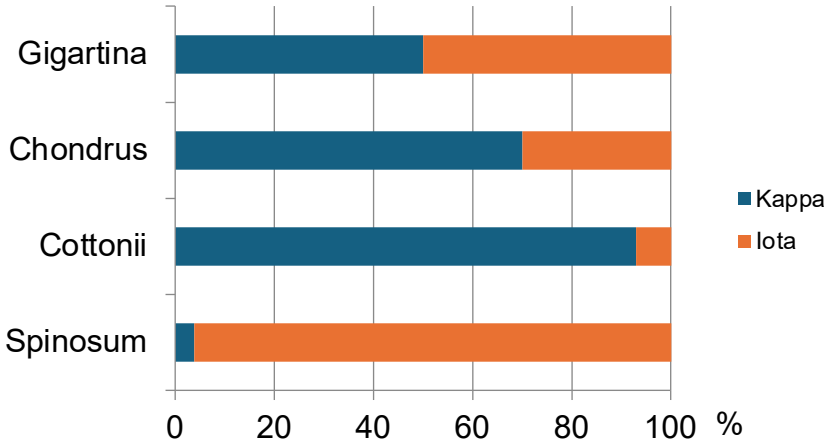


Figure 3 Approximate ration of κ- and ι-carrageenan within carrageenan extracted from species from Gigartina and Chondrus genera, Cottonii (Kappaphycus alvarezii) and Spinosum (Eucheuma denticulatum) [15,16].

The distribution of κ- and ι-units within carrageenan molecules is a critical determinant of the physical and rheological properties of κ/ι-hybrid carrageenans. This distribution varies depending on the algal species, life cycle stage, and extraction conditions. In addition to nearly homopolymeric carrageenans composed predominantly of either κ- or ι-units, κ/ι-hybrids also exist and may feature either random sequences or distinct blocks enriched in κ- or ι-disaccharides. The extent and pattern of such blockwise organization is commonly referred to as blockiness. The approximate κ/ι composition of various commercially relevant carrageenans is summarized in Figure 3. The most homogenous is ι-carrageenan extracted from *Eucheuma denticulatum* (Spinosum), which contains approximately 4% κ-units. This is followed by κ-carrageenan derived from *Kappaphycus alvarezii* (Cottonii). More structurally complex hybrids include carrageenans from *Chondrus crispus*, which contain 64–75% κ-units, and from various *Gigartina* species, with κ-unit content averaging around 50% depending on the specific taxon [19,21].

Despite these quantitative estimates, the precise sequence arrangement of κ- and ι-units in hybrid galactans remains unresolved. In κ/ι-hybrids, ι-units may occur sporadically as isolated residues or as short to long sequences embedded within predominantly κ-type chains [24]. For example, hybrids

from *Chondrus crispus* have been characterized by relatively random κ/ι distribution, whereas *Gigartina skottsbergii* has been associated with more defined domain structures comprising extended κ- or ι-enriched blocks [19,21].

2.1.2 Properties and Functionality

Carrageenan is water-soluble but insoluble in most organic solvents. Carrageenan's solubility in water depends on its type, ionic form, and environmental conditions. Kappa carrageenan is less soluble, especially in the potassium salt form, and typically requires heating to dissolve. Iota carrageenan also has limited solubility in its potassium form but swells in cold water. In contrast, lambda carrageenan is highly soluble in water across all salt forms due to its higher degree of sulfation and lack of 3,6-anhydrogalactose. Sodium salts generally enhance solubility compared to potassium salts [2].

Carrageenan forms thick solutions because its long, straight chains stay extended in water, driven by repulsion between their negative charges. These charges also draw in water molecules, creating a hydrated environment that resists flow. The viscosity of carrageenan solutions is influenced by several factors, including the type of carrageenan, molecular weight, concentration, temperature, and the presence of salts. Viscosity increases nearly exponentially with concentration and molecular weight, following the Mark-Houwink relationship. Viscosity decreases with presence of cations that reduce the electrostatic repulsion among sulfate groups [17,22–26].

When a hot carrageenan solution is cooled, it undergoes a solution–gel (sol-gel) transition. This process is influenced by several factors, including carrageenan type, ionic environment, and temperature [27]. Although the gelling mechanism is not fully understood, it is generally accepted that upon cooling, carrageenan chains adopt an ordered double-helix structure. This is facilitated by the presence of 3,6-anhydro bridges in the DA and DA2S units, which stabilize the chain in a $^{1}C_{4}$ conformation. These helices then aggregate into larger assemblies, forming junction zones that physically link polymer chains into a gel. The number and strength of these junctions determine the texture of the resulting gel. κ-Carrageenan tends to form strong, brittle gels due to tight aggregation, while ι-carrageenan forms more elastic and flexible gels [19,28–30].

The gelling ability of κ-carrageenan is enhanced by potassium and calcium ions, whereas sodium ions weaken gel formation but increase solubility. Potassium is the main gel-promoting ion for κ-carrageenan, while calcium plays this role for ι-carrageenan. The strongest κ-carrageenan gels are obtained in the presence of both potassium and calcium ions [2,31].

Carrageenan can also induce gelation in milk, where the mechanism resembles that in water but occurs at much lower concentrations due to specific electrostatic interactions between carrageenan and casein micelles, which contribute to network formation and gel stability [32]. The strength of these interactions depends on the electrostatic charge density of the carrageenan [33].

The different carrageenan types exhibit distinct interaction patterns with milk proteins: κ-carrageenan forms strong, brittle gels mainly through carrageenan–carrageenan interactions, with limited protein binding. In contrast, ι-carrageenan interacts more strongly with casein, leading to softer, more elastic gels. Hybrid carrageenans (κ/ι blends) show mixed, less predictable behavior, often resulting in weaker gels or flocculation depending on conditions. These effects are concentration-dependent and influenced by the carrageenan-to-protein ratio [34]. The presence of cations such as calcium and

potassium can significantly strengthen the carrageenan–milk gel network, but the extent of this effect depends on carrageenan type, ionic form, and concentration. Calcium is particularly effective in promoting ι-carrageenan–protein interactions, while potassium is essential for κ-carrageenan helix formation. Sodium and magnesium have weaker effects on gelation, mainly modulating ionic strength and charge shielding [35–38].

Carrageenan's gelling behavior is also affected by storage, pH, temperature, and milk composition. Over long storage periods, weak carrageenan-protein aggregates may further strengthen the gel via slow structural rearrangements [39].

2.1.3 Carrageenan application

These chemical and physical properties make carrageenan a versatile food ingredient used in a wide range of products, including dairy, meat, and bakery products, plant-based foods, 3D-printed foods, and beverages, where it functions as a gelling, thickening, and stabilizing agent [2,40].

The three-dimensional gel network formed by carrageenan and water upon cooling from a heated solution is utilized in meat and bakery products to retain water. In meat, carrageenan is known for increasing cooking yield and hardness [41–43], while in bakery applications, it strengthens dough by increasing its water-holding capacity and functions as an antistaling agent [44,45].

Carrageenan reduces ice crystal size in frozen desserts by limiting water mobility. In milk-based systems, it forms a protein–polysaccharide network through electrostatic interactions with casein, whereas in fruit-based systems such as sorbets, it forms a hydrocolloid matrix that traps free water. In both cases, this network inhibits ice crystal growth and improves texture during freezing and storage[46,47].

Carrageenan has a wide range of applications in dairy products, mostly due to its interactions with milk proteins. It is used in several dairy products, including cheese [48], chocolate milk [49] and milk pudding [50].

In beverages, carrageenan is used as a fining agent, as it interacts with proteins in wort and precipitates them out, improving beer's turbidity [51]

Carrageenan, particularly the ι-type, is used in multilayer emulsions to enhance physicochemical stability through electrostatic interactions with milk proteins. These emulsions support the development of label-friendly and potentially healthier low-fat products [52].

Several studies investigated the specific or synergic effect and addition of carrageenan into food matrices in combination with other gums such as oat gum [53], konjac gum [54], starch [55], pectin [56].

2.1.4 Carrageenan Manufacturing

Carrageenan is derived from seaweed through a series of processing steps that include harvesting, drying, washing, alkaline treatment, filtration, precipitation, drying, milling, and blending. The initial step involves harvesting and sun drying the seaweed to reduce its moisture content, thereby facilitating preservation and transportation. Once dried, the seaweed is washed to remove impurities.

The carrageenan content of commercially utilized seaweed varies significantly, ranging from 30% to 60% of dry weight, with some batches reaching up to 80%[15,57]. The washed seaweed is then treated with an alkaline solution, typically comprising NaOH, $Ca(OH)_2$, or KOH, under elevated temperatures. This alkaline treatment disrupts the seaweed cell walls, allowing the carrageenan to be released into the solution. The alkaline environment further promotes the conversion of precursor molecules into the desired carrageenan forms, resulting in the formation of 3,6-anhydrogalactose (reactions shown in Figure 2), a key structural component that enhances gel strength and functionality [58,59]. Furthermore, alkaline treatment maintains the molecular weight of carrageenan [15]. Following alkaline treatment, the carrageenan-rich solution is filtered to eliminate residual seaweed and insoluble materials. The filtrate is then concentrated by evaporating excess water. Carrageenan is subsequently precipitated from the concentrated solution using either alcohol (e.g., isopropanol) or potassium chloride (KCl) gel pressing. Alcohol precipitation induces coagulation of carrageenan, whereas gel pressing involves forming a gel that is pressed to remove water.

The precipitated carrageenan undergoes drying using air drying, drum drying, or freeze-drying techniques. Proper drying is critical to reducing moisture content, preventing microbial growth, and ensuring product stability. However, excessive drying temperatures may lead to decreased molecular weight, compromising the quality of the final product [15]. The dried carrageenan is then milled to a fine powder to meet specific particle size specifications. Due to natural variability in carrageenan composition, which comprises mixtures of different sulfated polysaccharides, the milled product may be blended with other carrageenan batches to achieve consistent functional properties [60]. Depending on the type of alkaline solution used, the final product may exhibit varying concentrations of calcium, potassium, or sodium. If the extraction process results in high sodium content, it can be exchanged with potassium via KCl precipitation to adjust the cationic composition [15]. The cationic content influences carrageenans' functionality but it is also important for the long-term chemical stability of carrageenan as incomplete neutralization may lead to sample acidification and subsequent degradation via autohydrolysis [18].

Carrageenan is classified under the EU food additive E-number E407, indicating its regulatory acceptance as a food additive. Manufacturing processes involve rigorous quality control to ensure consistency, safety, and functional integrity across production batches. Compliance with specifications for heavy metals, pesticides, and other contaminants is systematically verified. Functional testing, including gel strength, viscosity, and sulfate content assessments, confirms carrageenan type and its suitability for specific food applications, ensuring consistency with standard references.

2.2 STATE-OF-THE-ART

In the context of modern food manufacturing, the integration of advanced quality assurance frameworks such as Quality by Design (QbD) and Process Analytical Technology (PAT) has become increasingly relevant. These approaches align with industry efforts to enhance product consistency and regulatory compliance through data-driven methodologies. Methods within QbD promote the systematic design of processes with defined quality objectives, grounded in scientific knowledge and robust risk assessment, as advocated in ICH Q8(R2) and related guidelines [61]. Several steps are involved in successful implementation of QbD [62]. Complementarily, PAT enables real-time monitoring and control of critical quality attributes (CQAs), shifting quality assurance from end-

product testing to predictive process management [63,64]. To support these frameworks, analytical chemistry and chemometrics are essential tools in PAT implementation.

The implementation of PAT relies on a combination of analytical chemistry and chemometrics. Chemometric methods convert complex measurements into interpretable data, enabling both process understanding and quality prediction [65]. In the context of polysaccharides, PAT has been suggested, e.g., for pectin manufacturing [66] but for carrageenan it remains largely unexplored.

Numerous studies have investigated the functional properties of carrageenan in application systems [42,46,67–69]. These are typically observational: carrageenan is applied, and the resulting behavior is measured, with interpretations often relying on theoretical links to molecular structure. However, structure and function are frequently assessed in connection, limiting mechanistic insight.

As described in previous sections, carrageenan functionality is influenced by a range of molecular parameters, including molecular size and weight distribution, degree and pattern of sulfation, cationic composition, the presence of 3,6-anhydrobridges and other functional groups. These structural features determine carrageenan's solubility, gelling ability, viscosity behavior, and interactions with proteins or ions in formulation systems.

Among the most informative tools for probing molecular structure are proton nuclear magnetic resonance (^{1}H NMR) spectroscopy and Fourier Transform Infrared (FT-IR) spectroscopy. Proton NMR spectroscopy allows precise determination of carrageenan type by identifying chemical shifts related to κ- and ι-disaccharide units. It also provides semi-quantitative insights into the relative proportions of these units, enabling differentiation between pure, hybrid, or modified structures. In addition, NMR can detect methylation, sulfation at specific positions, and the presence of biosynthetic precursors such as μ- and ν-carrageenan. It has been employed for the structural elucidation of carrageenan for several decades [20,29,60,70–75]. It is also used in industry to screen seaweed as a potential source of carrageenan by analyzing spectra of extracts and for quality control, for example, to estimate the ratio between repeating units in hybrid carrageenans [60]. Furthermore, when combined with chemometric methods, ^{1}H NMR spectra can also be used for quantitative tasks [76,77].

Fourier Transform Infrared spectroscopy complements this by rapidly detecting functional groups, such as sulfate esters, 3,6-anhydrogalactose, and carboxyl or methyl groups, based on characteristic absorbance bands. This method has long been a pivotal tool in the structural analysis of carrageenans [78–83]. In combination with chemometric approaches, it has also been applied for quantification of carrageenan types and structural features [76,84,85].

Size-Exclusion Chromatography coupled with Multi-Angle Light Scattering (SEC-MALS) is the method of choice for measuring the molecular mass of carrageenan [86–89]. In addition to determining molecular weight distribution, SEC-MALS in combination with viscometry is also used to estimate intrinsic viscosity—a parameter that describes a polymer's contribution to solution viscosity, extrapolated to zero concentration [90]. Intrinsic viscosity reflects not only the molecular size but also the conformation of the polymer in solution.

The concentration and type of counterions significantly influence gelling and thickening behavior [91]. These are typically quantified using elemental analysis methods such as Inductively Coupled Plasma Atomic Emission Spectroscopy (ICP-AES) or Mass Spectrometry (ICP-MS)[18,92,93].

To measure the distribution of disaccharide units in hybrid κ/ι-carrageenan, the sample is enzymatically digested using κ-carrageenase and ι-carrageenase, which selectively cleave κ- and ι-units. The resulting fragments are analyzed using High-Performance Anion Exchange Chromatography (HPAEC), Size-Exclusion Chromatography with Multi-Angle Laser Light Scattering (SEC-MALLS), ^{1}H NMR spectroscopy, and Electrospray Ionization Mass Spectrometry (ESI-MS) to determine the sequence and distribution of κ and ι units along the polymer chain [94,95].

Several studies have compared the structure–function relationships of ι-, κ-, and λ-carrageenan, including blends of these [92,96,97] and hybrid forms [19,21,98,99]. These investigations typically utilize a combination of the aforementioned analytical platforms - ICP-AOE/MS, ^{1}H NMR, SEC-MALS, and FT-IR – to characterize molecular structure. However, to date, no study has comprehensively combined data from all these methods to model carrageenan functionality.

Some studies have attempted to predict functional properties from structural data [100,101]. For example, Anderson et al. (2002) developed a power-law model to predict viscosity of skim milk with low carrageenan concentration levels, supporting process design in thermal application [101]. Other studies have used Design of Experiments (DoE) to optimize formulations or understand the effects of processing parameters, including the formulation of skim milk-based desserts stabilized with carrageenan and other gums [102], and studies on milk–carrageenan interactions using factorial designs [32]. Bixler at co-authors [103–105] investigated the impact of source material and extraction methods on the functional and structural characteristics of commercial carrageenans.

Current practice in both academia and industry often focuses on one-to-one correlations between a single structural and a single functional parameter. This approach overlooks the complexity and potential interactions among structural features. Data fusion across multiple analytical platforms offers a means to integrate complementary information, providing a more complete characterization of carrageenan. Predictive data fusion approaches are increasingly used in food science and technology for food authenticity verification, quality control and process monitoring [106–111]. However, no applications were found for polysaccharide quality control, highlighting the relevance of this study.

3 DATA: FUNCTIONAL AND STRUCTURAL CHARACTERIZATION OF CARRAGEENAN

This chapter provides the experimental foundation for modeling the structure–function relationships for carrageenan. In alignment with the PAT knowledge pyramid (see Chapter 1), it represents the "Data" level, encompassing both application-based functional tests and molecular-level structural measurements.

This chapter explores the hypothesis that the molecular structure of carrageenan can be thoroughly characterized using four analytical platforms: SEC-MALS, ICP-MS, FT-IR spectroscopy, and ^{1}H NMR spectroscopy. While these techniques provide structural insights, carrageenan's functionality is assessed through application-based tests.

To maintain clarity and consistency throughout this thesis, specific terminology will be used to refer to carrageenan types and classes as well as data structures. These terms are defined in Box 1 below.

Box 1: Working Terminology

Carrageenan Classification Terms

Class	Carrageenan classified based on its source material: Spinosum class for *Eucheuma denticulatum*, Cottonii class for *Kappaphycus alvarezii*, Chondrus class for *Chondrus crispus*, and Gigartina class for carrageenan extracted from seaweed species within the *Gigartina* genus.
Type	Refers to the molecular structure of carrageenan: ι-, κ-, or κ/ι – carrageenan

Data Structure Terms

Data block	a single set of data collected from one instrument, structured as a matrix with rows (samples) and columns (variables).
Data set	the complete collection of all data relevant to a specific problem or analysis, encompassing multiple data blocks.

3.1 FUNCTIONALITY DATA

Given the broad application range of carrageenan, three functional tests were selected to capture its key performance characteristics and support model development: (i) apparent viscosity in aqueous solution (Paper 1), (ii) gel and breaking strength in milk systems (Paper 2), and (iii) suspension stability of cocoa particles in chocolate milk (Paper 3). These methods were chosen because of their different matrices and because of these were prioritized by the company.

To ensure that the functional data reflected the diversity of carrageenan performance, samples were selected from the company's broad internal sample base. These included a wide range of functionality levels and spanned approximately two years of production, thus capturing variability introduced by seasonal differences in seaweed harvesting and processing conditions.

3.1.1 Viscosity

Viscosity measurement of carrageenan is described in detail in Paper 1 [17]. This measurement is part of the quality control process, as it is required to confirm that the product meets the specification of a minimum apparent viscosity of 5 cP at 75 °C in a 1.5% solution, as outlined in the analytical procedure specified in the Food Chemicals Codex (FCC), using a rotational viscometer [112,113]. In addition, the aqueous viscosity of carrageenan provides insight into how the product will perform in application and behave during processing, particularly in production pipelines. Therefore, this measurement is considered critical for both compliance and functional assessment.

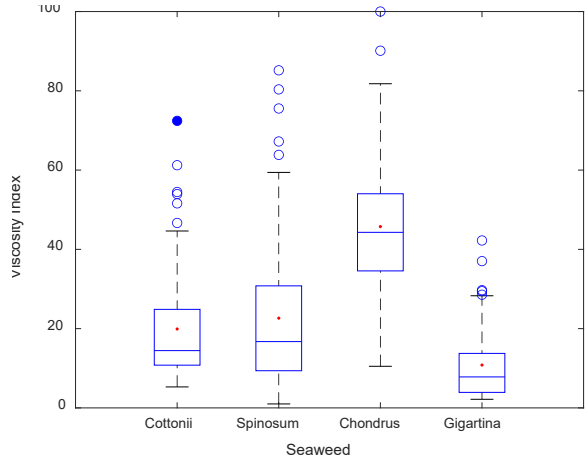


Figure 4 Boxplots showing the distribution of viscosity index (normalized values of viscosity originally measured in cP) for carrageenan samples extracted from each seaweed type. The dataset described in Paper 1.

Figure 4 shows distribution of viscosity for carrageenan samples extracted from four red seaweed species. Extreme samples were intentionally included to capture the variation in carrageenan functionality, and by extension, its presumed structural diversity. Chondrus had the highest median and mean viscosity index, with a broad interquartile range and several high outliers, indicating both high and variable viscosity performance. Spinosum also exhibited a wide distribution with multiple high values, suggesting inconsistent but potentially high viscosity. Cottonii showed lower central values and moderate variability. Gigartina had the lowest median and mean values, with a narrow interquartile range and fewer extreme values, indicating consistently low viscosity. The viscosity index follows: Chondrus > Spinosum > Cottonii > Gigartina, which matches the order of intrinsic viscosity values shown in Figure 12B. This supports that solution viscosity depends primarily on intrinsic viscosity rather than only molecular weight, consistent with the Mark–Houwink relation [23].

3.1.2 Gel and breaking strength for carrageenan-milk gels

Measurement of gel and breaking strength in carrageenan–milk gels is a standard practice in the industry, used to evaluate the gelling properties of carrageenan in dairy systems. These measurements serve two key purposes: (1) quality control, to ensure batch-to-batch consistency as

expected by customers, and (2) product development, to optimize the functional performance of carrageenan in target applications. The measurement methodology is described in detail in Paper 2. In brief, the procedure involves preparing standardized milk gels and evaluating their mechanical properties using a texture analyzer. The instrument applies a controlled force to the gel, recording the resistance at a predefined depth (gel strength) and identifying the point at which the gel structure ruptures (gel break).

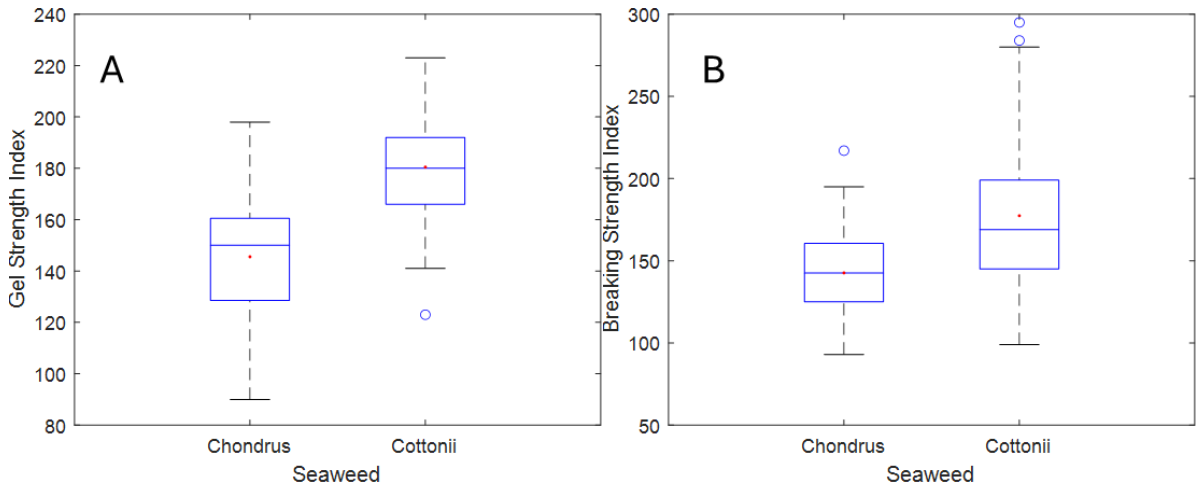


Figure 5 Boxplots showing the distribution of carrageenan-milk gel strength (panel A) and breaking strength (panel B) index for carrageenan extracted for Chondrus and Cottonii. The dataset described in Paper 2.

The distribution of gel strength indices for carrageenan-milk gels is presented in Figure 5, Panel A, while the breaking strength statistics are visualized in Figure 5, Panel B of the same figure. The data are shown for two carrageenan classes: Chondrus and Cottonii. In this method, only two carrageenan classes are included, as these are typically used of this application.

Both gel strength and breaking strength values are higher for Cottonii-derived carrageenan. This trend is consistent with previous findings, where an increased κ-carrageenan fraction within the carrageenan sample is associated with enhanced gel and breaking strength [21,114].

3.1.3 Suspension stabilization

Carrageenan is widely used at low concentrations to stabilize suspensions in dairy-based beverages such as chocolate milk. The objective is to enhance suspension stability without significantly altering the product's pourability or texture. This effect is achieved through a modest increase in viscosity, primarily due to weak gel-like networks formed via interactions between carrageenan and milk proteins. According to Stokes' law, higher viscosity and smaller density differences help slow down sedimentation, keeping particles evenly suspended [2].

As previously discussed in Section 2.1.4 (page 7), the final carrageenan product is typically formulated as a blend of several semi-finished types. Blending is one of the final and most critical unit operations in carrageenan manufacturing. Traditionally, these blends are developed using fixed recipes determined by the application development team. If the specified raw materials are available, the blend is produced and then tested in application—for example, by measuring the viscosity of

chocolate milk—to assess whether it effectively stabilizes cocoa particle suspensions. If the viscosity falls within the target range, the blend is approved for use.

However, this conventional approach depends on the consistent availability and quality of specific raw materials, which limits flexibility in manufacturing and can lead to inefficiencies. To address these limitations and improve both sustainability and resource efficiency, the principles of QbD were adopted. A key element of QbD is the development of a design space, defined as "the multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality" [61]. In Paper 3, Design of Experiments (DoE) approach [115] was used to systematically investigate how different carrageenan classes, individually and in combination, affect suspension stability. The primary response variable was the apparent viscosity of chocolate milk, which served as a proxy for evaluating the suspension stability of cocoa particles in the presence of various carrageenan blends. In this context, higher viscosity indicates a thicker beverage capable of effectively suspending cocoa particles and preventing sedimentation. The experimental design included both concentration and quality factors. Variables related to concentration comprised the levels of Cottonii, Chondrus, and Gigartina carrageenans, as well as the sucrose concentration.

As a quality factors carrageenan–milk gel breaking strength was selected based on preliminary screening designs (data not shown). Among different levels of carrageenan–milk gel strength and carrageenan viscosity in water, gel strength demonstrated a moderate correlation with chocolate milk viscosity. The experimental procedure combined mixture and factorial design methods, resulting in 159 unique formulations and corresponding viscosity measurements. Due to confidentiality, exact viscosity values are not disclosed. Detailed analysis is presented in Paper 3.

3.2 MOLECULAR STRUCTURE DATA FROM ANALYTICAL PLATFORMS

Based on the literature review, it was observed that four analytical platforms are frequently employed by various research groups to characterize carrageenan. These methods—Fourier Transform Infrared Spectroscopy (FT-IR), Proton Nuclear Magnetic Resonance (^{1}H NMR), Size-Exclusion Chromatography with Multi-Angle Light Scattering and Viscometry (SEC-MALS-dRI-Visco), and Inductively Coupled Plasma Mass Spectrometry (ICP-MS)—provide complementary structural information. Together, they cover a broad range of molecular features, from functional group composition and polymer conformation to ionic content and molar mass distribution. As such, they are included in first hypothesis of this study.

A total of 377 carrageenan samples were analyzed in duplicate using the four analytical platforms mentioned above. These samples were extracted from four seaweed sources, with the following distribution: 102 samples from Chondrus, 103 from Cottonii, 82 from Gigartina, and 90 from Spinosum. Detailed descriptions of the analytical procedures, data acquisition, and measurement protocols are described in the papers.

In the following sections, the data generated from each of the four analytical platforms are presented, along with a brief introduction to the corresponding analytical methods. Detailed descriptions of the measurement procedures can be found in Paper 1 and Paper 2. All datasets have been preprocessed according to the protocols outlined in these papers, with additional explanation

provided in Section 4.1 (page 26), while in this chapter the applied preprocessing methods are only mentioned.

3.2.1 Fourier Transform Infrared Spectroscopy

Fourier Transform Infrared (FT-IR) spectroscopy is used to identify the functional groups in carrageenan. Infrared light passes through the sample and is absorbed at specific wavelengths by the bonds in the molecule. As the molecules absorb IR radiation, their chemical bonds vibrate—stretching, bending, or twisting—depending on the energy of the light. The detector measures how much light is absorbed at each wavelength, and the resulting data is transformed into a spectrum using Fourier transform. The resulting spectrum shows peaks that correspond to characteristic vibrations of carrageenans functional group [116].

Figure 6 displays the spectra of carrageenan samples, color-coded according to their seaweed source: Spinosum, Cottonii, Gigartina, and Chondrus. The spectral fingerprint region is highlighted on the left. The spectra were preprocessed by applying Savitzky-Golay smoothing followed by Standard Normal Variate (SNV) transformation. Figure 7A presents the preprocessed spectra in the fingerprint region, while Figure 7B shows the mean spectra for each carrageenan type, grouped by their respective seaweed source. All spectral bands were assigned to functional groups characteristic of carrageenans based on established literature sources [78–83]. A summary of these assignments is provided in Table 2.

The intense band at 1220 cm^{-1} corresponds to the antisymmetric stretching vibration of the O=S=O group in the sulphate esters of carrageenan [78]. Its intensity reflects the overall sulphate content, which decreases from λ- to ι- to κ-carrageenan, in line with the declining number of sulphate ester groups (see Figure 2) [81,84]. Accordingly, a decreasing signal intensity is expected from Spinosum- to Gigartina-, Chondrus-, and Cottonii-derived carrageenan, in line with the respective concentrations of ι- to κ-carrageenan within each polymer (see Figure 3), as evidenced by the FT-IR spectra shown in Figure 7.

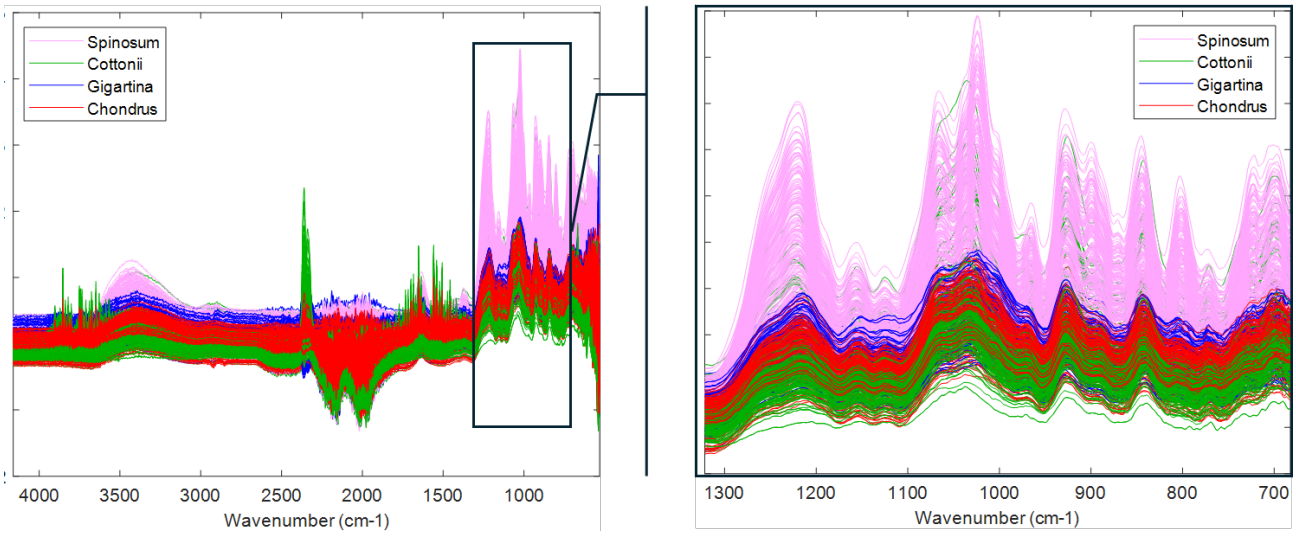


Figure 6 FT-IR spectra of 377 carrageenan samples extracted from Spinosum, Cottonii, Gigartina and Chondrus seaweed obtained in the range 4000 – 500 cm^{-1} in duplicates and on the right side is carrageenans FT-IR fingerprint in the range 1300-700 cm^{-1}.

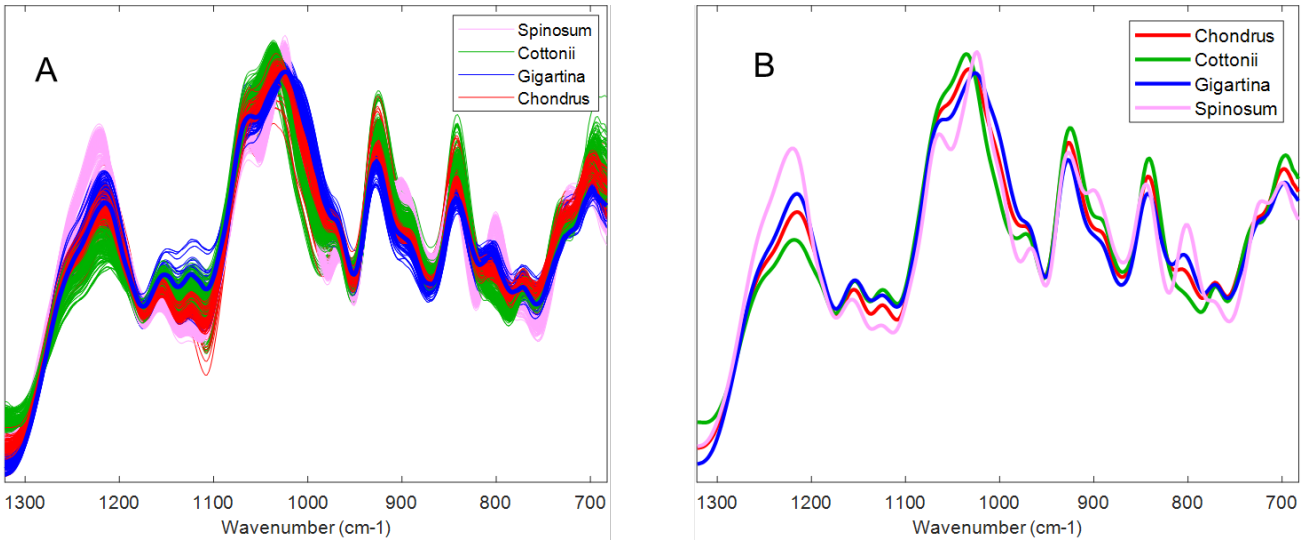


Figure 7 FT-IR spectra of carrageenan extracted from Spinosum, Cottonii, Gigartina and Chondrus seaweed. The spectra are color coded accordingly to the corresponding raw material. Figure 4.A. shows SNV filtered spectra. Figure 4.B. shows a mean spectra of carrageenan samples for each seaweed type.

Table 2 Assignment of FT-IR spectra bands of carrageenan and their correspondence to three carrageenan types[78–83].

Wavenumber (cm^{-1})	Functional group	κ	ι	λ
1220	O=S=O (asymmetric stretching)	+	+	+
1150	C-O-C (asymmetric stretching)	+	+	-
1120	Glycosidic bonds (asymmetric stretching)	+	+	+
1060	C-O and C-OH	+	+	+
1030	C-OH and S=O	+	+	+
970-1070	Glycosidic bonds	+	+	+
925	C-O-C (3,6 anhydrogalactose) (stretching)	+	+	+
890	C6, β-D-galactose	+	+	+
840	C4-O-S galactose (stretching)	+	+	-
805	C2-O-S in 3,6-anhydrogalactose	-	+	-
770	Skeletal bending of galactose ring	+	+	+
720	C-O-C α (1,3) (stretching)	+	+	-
700	Sulphates in C4, galactose	+	+	+

The “shoulder” of this broad band at 1250 cm^{-1} is due to interaction for carrageenan molecule with cations [82]. A similar effect is observed in the glycosidic bond region (1100–970 cm^{-1}), where both

shoulders and slight signal shifts can be detected [117]. Due to the limited number of studies addressing this phenomenon in detail, it is not currently possible to precisely describe how individual cations influence specific vibrational signals. One in-depth study exists [82]; however, its measurements were performed in aqueous solution and are therefore not directly comparable to the spectra obtained in the present work. The band at 805 cm^{-1} is attributed to a sulfate group at the C2 position of 3,6-anhydrogalactose, characteristic of ι-carrageenan and hybrid carrageenans. This band serves as a useful marker to distinguish hybrid and ι-carrageenan from κ-carrageenan, which lacks sulfation at C2 (see Figure 2). The signal intensity observed in the mean spectra in Figure 7B aligns with expectations based on carrageenan composition: the highest intensity appears in samples derived from Spinosum, followed by Gigartina and Chondrus, while the band is absent in Cottonii-derived carrageenan, consistent with its classification as pure κ-carrageenan (see Figure 3).

A weak band was observed around 770 cm^{-1}; however, there is limited reporting of this band in carrageenan-specific literature, possibly due to its relatively low intensity. It has been reported in agar, where it is associated with skeletal bending vibrations of the galactose ring [83]. Its presence in carrageenan may be attributed to similar ring deformation modes of galactopyranose backbone, which is common for both polysaccharides.

3.2.2 Proton NMR Spectroscopy

Proton nuclear magnetic resonance (^{1}H NMR) spectroscopy exploits the magnetic properties of hydrogen nuclei (^{1}H) in carrageenan. When placed in a strong magnetic field and exposed to a radiofrequency pulse, the hydrogen nuclei absorb energy and transition between spin states. As they relax back to their lower energy state, they emit radiofrequency signals that are detected and converted into a spectrum by Fourier transform. The resulting spectrum contains resonance peaks that depend on the electron environment of each proton [118]. The resulting spectrum displays resonance peaks that depend on the electronic environment of each proton. In carrageenan, the local structure, degree of sulfation, and presence of 3,6-anhydro bridges affect the chemical shifts of specific protons, especially those on the anomeric carbon (C1) of the galactose units. These shifts occur in a well-defined region (approximately 4.5–6.5 ppm), allowing the identification of characteristic proton signals for κ-, and ι-carrageenan.

However, due to a combination of spectral overlap, signal broadening and limited resolution, ^{1}H NMR alone is often insufficient for complete structural characterization. As a result, it is typically complemented with two dimensional techniques such as ^{1}H-^{13}C Heteronuclear Single Quantum Coherence (HSQC) spectroscopy [72]. This method correlates hydrogen and carbon nuclei through single-bond couplings, producing a two dimensional spectrum, helping resolve overlapping signals in complex mixtures.

Figure 8A presents the ^{1}H NMR spectra of carrageenan samples within the chemical shift range of 8 ppm to 0 ppm. A more detailed view of the region between 6 ppm and 3.8 ppm is shown in Figure 8B, where characteristic signals corresponding to carrageenan molecules are observed. In contrast, the region spanning approximately to 3.0 to 1.0 ppm primarily contains signals attributed to impurities likely introduced during the production process. Figure 8C illustrates the same spectra following spectral alignment [119] and SNV filtering, as described in Paper 2. Distinguishing clear patterns or similarities among different carrageenan types remains challenging when assessed visually. To facilitate interpretation, the mean spectra corresponding to each carrageenan raw material are also presented in Figure 8D, providing a comparative overview.

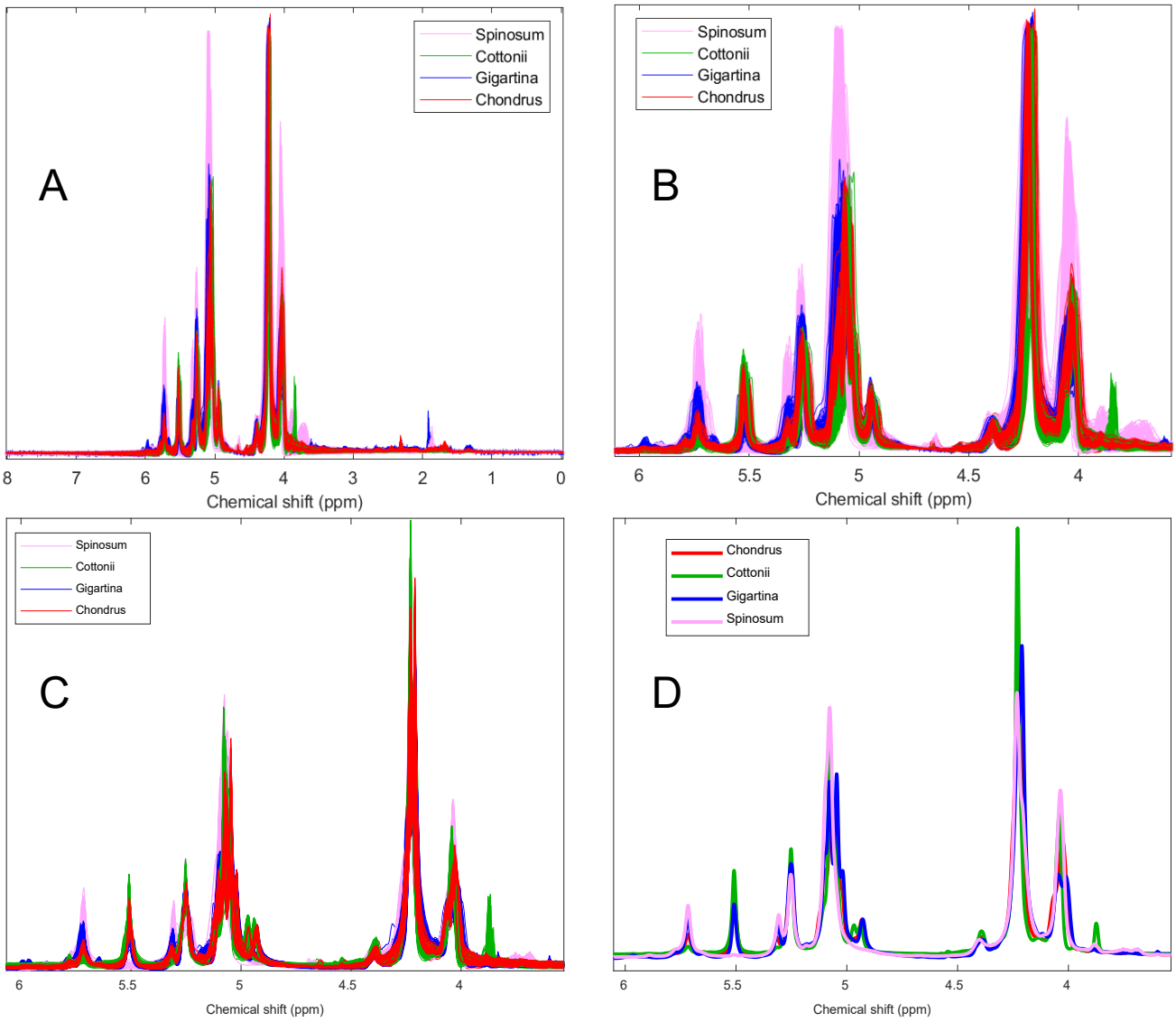


Figure 8 ^{1}H NMR spectra for 377 carrageenan extracted from Spinosum, Cottonii, Gigartina and Chondrus seaweed. The spectra are color coded accordingly to the corresponding raw material. Figure A shows raw NMR spectra. Figure B shows spectra region with signals corresponding to protons in carrageenan molecule. Figure C shows spectra after COW alignment and SNV filtering. Figure D shows mean spectra after preprocessing for each raw material class.

Table 3 1H-NMR chemical shifts of κ- and ι-carrageenan based on spectra in Figure 8, Figure 9 and Figure 10. Assignments were made using literature [20,60,77,120].

Type	Unit	H1	H2	H3	H4	H5	H6
κ–carrageenan	G4S	5.07	4.02	4.38	5.25	4.21	3.85
	DA	5.53	4.58	4.94	5.55	5.10	4.58, 4.46
ι-carrageenan	G4S	5.07	4.04	5.30	5.34	4.19	4.23
	DA2S	5.73	5.08	4.38	5.10	5.07	4.64, 4.52

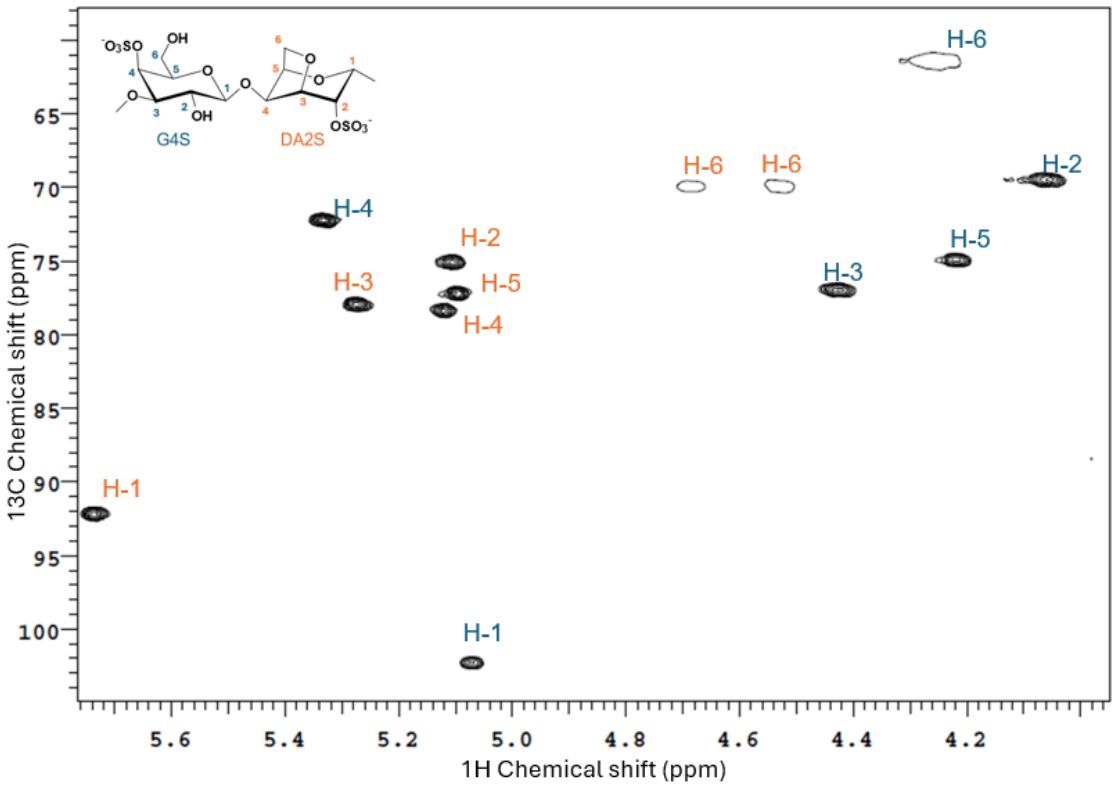


Figure 9 Heteronuclear 1H-13C chemical shift correlated spectrum of low molecular weight (50 kDa) ι-carrageenan. Signals corresponding to protons from the DA2S unit are highlighted in orange, while those from the G4S unit of the diad are shown in blue.

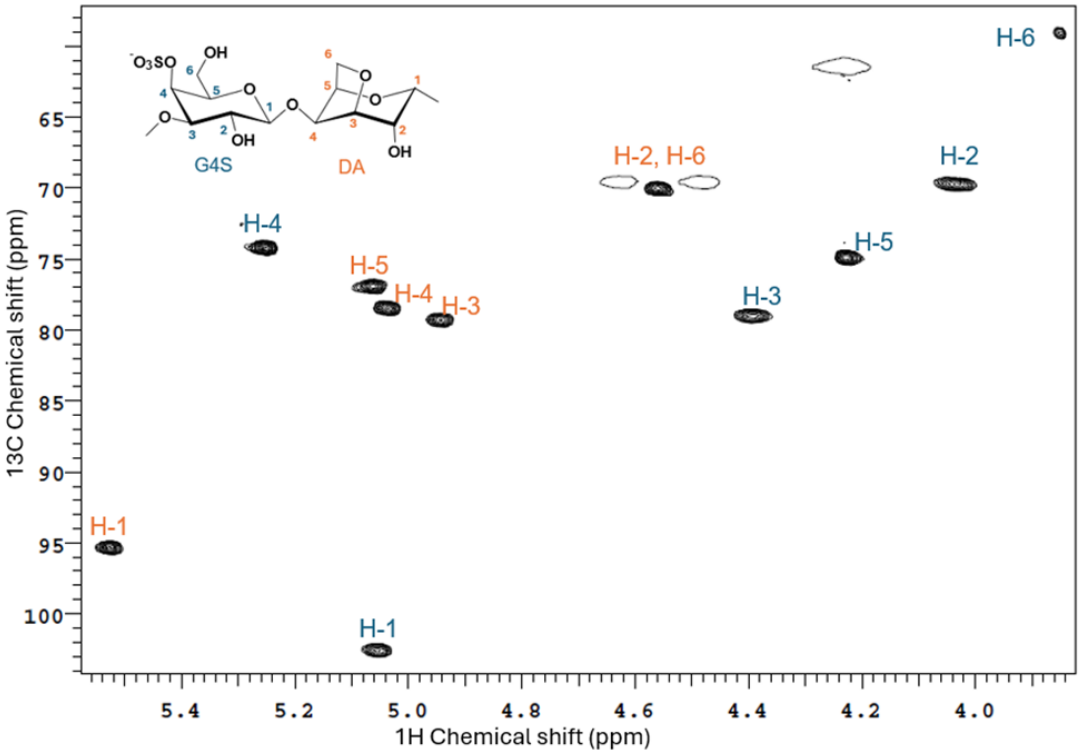


Figure 10 Heteronuclear 1H-13C chemical shift correlated spectrum of low molecular weight (50 kDa) κ-carrageenan. Signals corresponding to protons from the DA2S unit are highlighted in orange, while those from the G4S unit of the diad are shown in blue.

Chemical shifts assignments are summarized in Table 3, based on combination of relevant literature sources [20,60,77,120] and two-dimensional heteronuclear ^{1}H-^{13}C correlation (HSQC) spectra for pure low molecular weight ι- and κ-carrageenan (50 kDa), shown in Figure 9 and Figure 10, respectively. These well-resolved spectra enabled reliable assignment of proton signals, which were then used as a basis for interpreting the more complex spectra of carrageenan samples in Figure 8.

The distinct indicators of carrageenan types in the 1H NMR spectra (Figure 8) are the resonances of the anomeric protons at 5.53 ppm and 5.73 ppm, corresponding to κ- and ι-carrageenan respectively. The signal intensities in the mean spectra in Figure 8D appear to align with the expected κ/ι composition for each carrageenan type based on the raw material source. Specifically, the resonance at 5.73 ppm – characteristic of ι-carrageenan – shows a decreasing intensity in the order Spinosum > Gigartina > Chondrus > Cottonii, which is consistent with the typical ι-carrageenan content in these seaweed species. Another distinct signal appears at 3.85 ppm, originating from the two protons on the C-6 methylene group of G4S unit in κ-carrageenan. This well-separated resonance can serve as a useful indicator of κ-carrageenan presence; however, it tends to appear as a broad signal, which may complicated quantitative interpretation. The remaining signals largely consist of overlapping resonances, making it difficult to extract additional structural information without applying chemometric analysis. Nevertheless, the three well-separated signals mentioned above are commonly used for quantitative evaluations in carrageenan characterization [60]. The chemical shifts observed in ^{1}H NMR spectra of κ/ι-hybrid carrageenans are influenced not only by the type of disaccharide units (κ or ι), but also by their sequence context—namely, the distribution of adjacent units and proximity to reducing or non-reducing ends. Alternating or blockwise arrangements induce subtle shielding or deshielding effects, particularly in the anomeric proton region [72,121]. While this offers valuable structural insights, it requires high spectral resolution. Such information could not be extracted from the spectral data presented in Figure 8.

3.2.3 Inductively Coupled Plasma Mass Spectrometry

The ionic composition of commercial carrageenan samples was assessed using inductively coupled plasma mass spectrometry (ICP-MS). In this method the samples is firstly digested using strong acids and in a microwave system. This breaks down the organic matter and releases metals into solutions. The solution is then introduced into a high-temperature plasma, which ionizes the elements. These ions are then separated by their mass-to-charge ration in the mass spectrometer [122].

Figure 11 presents boxplots of potassium (K^{+}), calcium (Ca^{2+}), sodium (Na^{+}), magnesium (Mg^{2+}), and sulfur (S) concentrations across samples derived from Spinosum, Cottonii, Gigartina, and Chondrus. These elements were chosen due to their known roles in influencing carrageenan gelation, solubility, and rheological performance.

Potassium levels (Figure 11A) were highest in Spinosum and Cottonii carrageenans, with medians near 50 mg/g. Gigartina and Chondrus carrageenan exhibited significantly lower K content. An inverse trend is observed for calcium (Figure 11B), where concentrations decrease progressively from Gigartina to Chondrus, Cottonii, and finally Spinosum derived carrageenan. Since ester sulfate groups on the carrageenan backbone require charge compensation, the total concentration of counterions (K^{+} and/or Ca^{2+}) is expected to increase with higher degrees of sulfation, as seen in ι-carrageenan that is derived from Spinosum species. The observed ion distributions suggest that the

commercial samples were manufactured to optimize the levels of gelling-promoting cations—either potassium, calcium, or a combination thereof—based on the desired functional performance of the final product.

Sodium (Figure 11C) was most abundant in Spinosum-derived carrageenan, with values ranging up to 80 mg/g, likely reflecting use of sodium-based reagents during purification. Cottonii samples had consistently low sodium levels, while Gigartina and Chondrus displayed moderate concentrations. Carrageenan samples with elevated sodium levels also tend to have increased potassium concentrations, suggesting intentional adjustment of the cation profile during manufacturing to ensure optimal gelling performance, since sodium ions do not promote gelation.

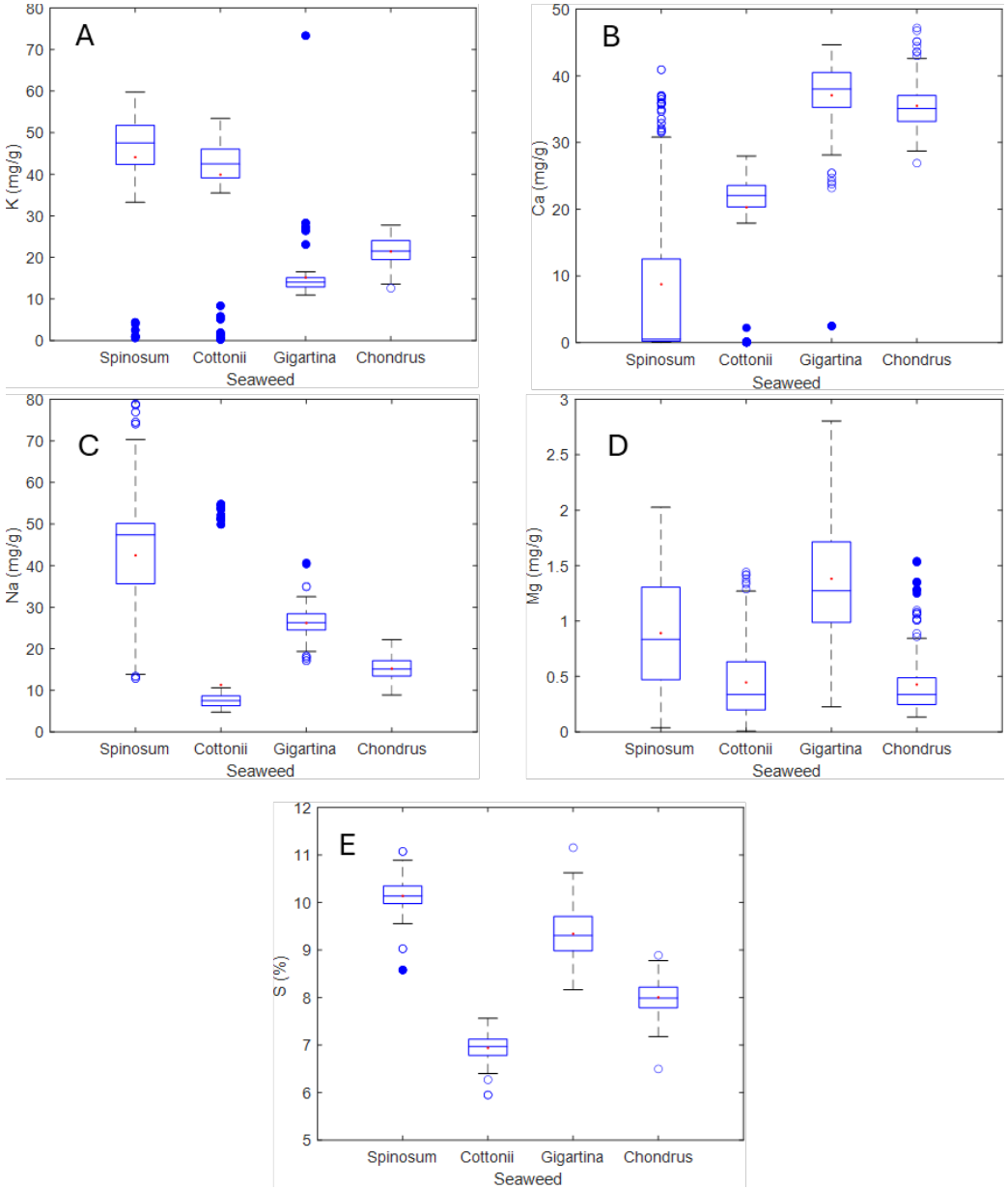


Figure 11 Boxplots showing counterion concentration for carrageenan derived from Spinosum, Cottonii, Gigartina and Chondrus of A) potassium; B) calcium; C) sodium; D) magnesium and E) sulfur.

Magnesium (Figure 11D) concentrations were low across all groups (generally <2 mg/g), consistent with the known minor role of this ion in processed carrageenan. Nonetheless, Mg remains relevant due to its ability to influence κ-carrageenan gelation even at low levels [36].

Sulfur content (Figure 11E), reported as a percentage, serves as a proxy for the degree of sulfation and thus helps distinguish carrageenan types. Higher sulfur levels were observed in Spinosum and Gigartina, supporting the information provided earlier in Figure 3Cottonii samples showed the lowest sulfur content, consistent with their predominance of κ-carrageenan, which contains fewer sulfate groups.

Several extreme values appear in the boxplots; these should not be regarded as outliers in the statistical sense but rather as a reflection of the inherent variability among commercial products. Including such samples ensures a more comprehensive representation of the range of ion compositions present in carrageenan.

3.2.4 Size exclusion Chromatography coupled with multi angle light scattering, Concentration Detector and Viscometry

Size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) was employed to characterize the molecular size distribution and hydrodynamic behavior of commercial carrageenan products. This technique provides absolute molar mass and intrinsic viscosity data, independent of calibration standards, and is well suited for analyzing the heterogeneous nature of processed polysaccharides like carrageenan [24,123–125]. In this work this analytical platform combines four parts: size exclusion chromatography (SEC), multi-angle light scattering (MALS), a differential refractive index detector (dRI), and a differential viscometer (see Paper 1 and Paper 2 for model and instrument producer). In the literature, the system is commonly abbreviated as SEC-MALS or SEC-MALS-visco, as used in Paper 1.

First, SEC separates carrageenan molecules by size. Large molecules cannot enter the pores in the column packing and come out first. Smaller molecules enter the pores and come out later. This separation process does not change the molecules but just sorts them by size.

As the molecules leave the SEC column, they pass through the MALS detector. MALS shines a laser on the molecules and measures how much light is scattered at different angles. From the amount and angle of scattered light, it calculates the molar mass and the radius of gyration, which describes how the molecular mass is distributed and the shape of the molecule in solution.

Then the sample passes through the dRI detector. This measures how much the solution's refractive index changes due to the sample. The change is proportional to the carrageenan concentration at that point in the flow. This concentration is needed to calculate the molar mass from the MALS signal.

After that, the sample enters the viscometer. This measures how much the carrageenan increases the viscosity of the solution. It calculates the specific and intrinsic viscosity. More detailed information on the instrument can be found in references [24,123,124,126–128].

Parameters calculated by SEC-MALS are the following: number average molecular mass (M_n), molar mass at the top of the concentration peak (M_p), weight average molecular mass (M_w), Z-average molar mass (M_z), uncertainty-weighted average molar mass ($M_{(avg)}$), polydispersity (M_w/M_n), weight

average mean square radius (*RMS* radius or *rw*), weight average hydrodynamic radius ($rh(v)_w$), uncertainty weighted average hydrodynamic radius (*rh(v)avg)*, mass recovery, weight average intrinsic viscosity $[\eta]_w$, uncertainty weighted average intrinsic viscosity $[\eta]_{(avg)}$. The way these were estimated is given in Paper 1 [17].

Figure 12 summarizes the distributions of weight-average molar mass (Mw), intrinsic viscosity ([η]w), and mass recovery for carrageenan samples from the four commercial sources. A detailed statistical overview is provided in Table 4.

Table 4 Statistics summary for each parameter obtained by SEC-MALS, grouped by seaweed source of carrageenan. The parameters are the same as in Paper 1 [17], where these are explained in details.

	Mn (kDa)	Mp (kDa)	Mw (kDa)	Mz (kDa)	M(avg) (kDa)	Polydispersity (Mw/Mn)	rw (nm)	rh(v)w (nm)	rh(v)(avg) (nm)	Mass recovery (%)	[η]w (mL/g)	[η](avg) (mL/g)
Chondrus												
Mean	521	782	900	1294	539	1.7	93	49	46	86	944	813
Std. dev	73	78	85	108	103	0.1	7	3	4	4	90	95
Min	332	540	588	862	271	1.4	65	37	34	70	647	472
Max	718	926	1054	1637	838	2.1	104	56	54	97	1171	991
Cottonii												
Mean	310	568	615	933	272	2.0	71	39	31	92	695	554
Std. dev	64	71	75	107	68	0.3	6	3	3	3	80	65
Min	179	419	463	689	144	1.5	53	33	25	84	500	256
Max	530	823	853	1257	507	3.1	90	48	43	99	940	708
Gigartina												
Mean	311	487	578	929	291	1.9	71	38	32	82	681	537
Std. dev	84	101	134	189	97	0.1	11	5	7	4	139	109
Min	188	330	378	593	163	1.5	51	29	22	71	453	327
Max	548	746	904	1394	609	2.3	96	51	50	93	1062	830
Spinosum												
Mean	604	1200	1126	1655	439	1.9	107	49	38	86	763	611
Std. dev	213	511	325	389	154	0.3	21	8	8	4	169	176
Min	218	413	470	778	167	1.5	62	29	24	73	334	236
Max	1121	2508	1830	2396	1006	3.1	146	65	62	95	1051	1013

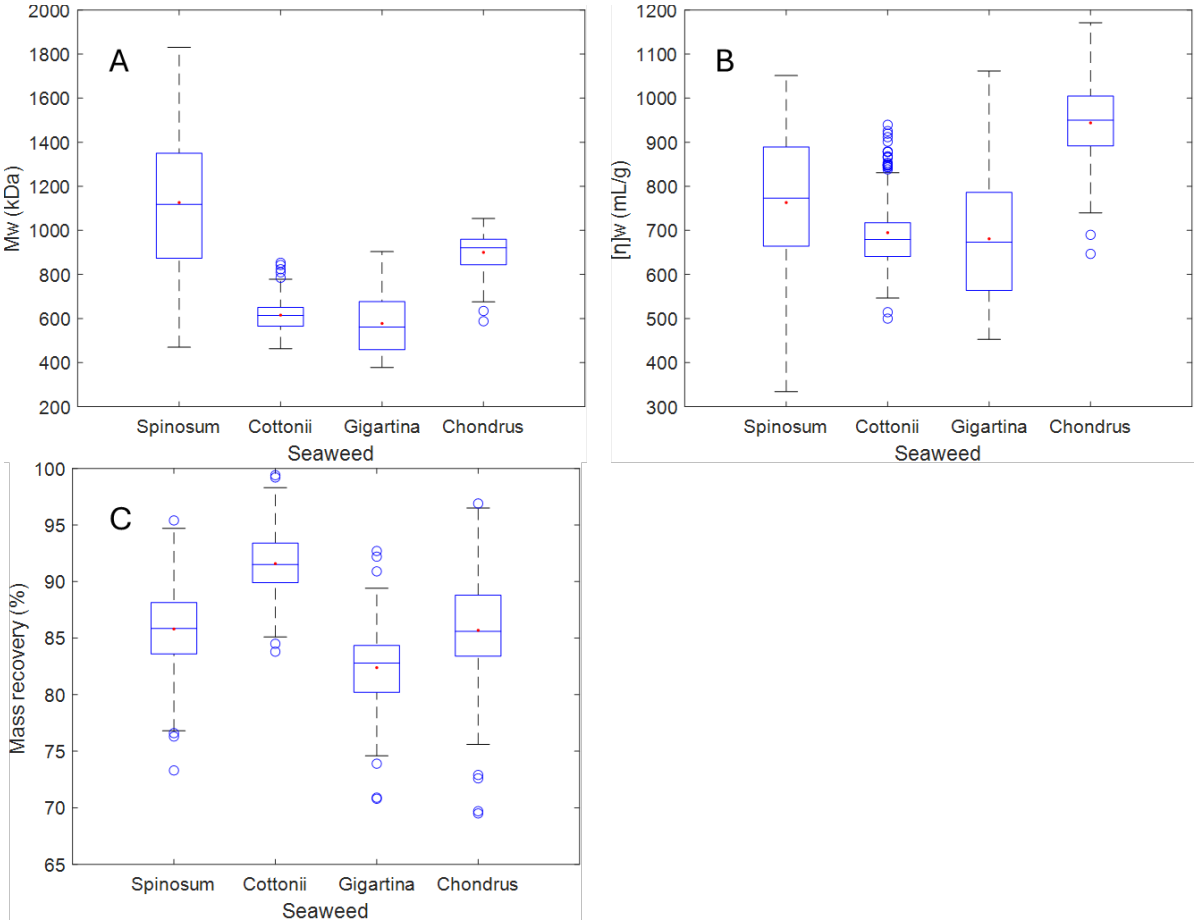


Figure 12 Boxplots showing A) the weight average molecular mass (Mw); B) the weight average intrinsic viscosity ([η]w and C) mass recovery for carrageenan samples extracted from Spinosum, Cottonii, Gigartina and Chondrus seaweed.

Out of the twelve parameters presented in the table below, the weight average molecular mass, M_w, is most often reported in the literature. In this dataset it lies in between 200 kDa and 2000 kDa in agreement with this data reported earlier [19,23]. Spinosum-derived carrageenan exhibited the highest values the weight average molecular (Figure 12 A) with a wide range extending from 470 to over 1800 kDa and a mean of 1126 kDa. In contrast, carrageenan extracted from Cottonii, Gigartina, and Chondrus samples showed narrower molecular weight distributions, with mean values between 600 and 900 kDa.

Intrinsic viscosity (Figure 12B), which reflects the coil size of the polymer in solution, was highest in carrageenan samples extracted from Chondrus and Spinosum, in agreement with their higher weight average molecular mass. The differences in viscosity also align with variation in sulfate content (see ICP-MS data), since sulfation affects the stiffness and solubility of the carrageenan chains [23].

Mass recovery (Figure 12C), used in this study as an indicator of polysaccharide purity (carrageenan concentration in carrageenan sample). Mass recovery was slightly higher for carrageenans from Cottonii (mean of 92%), while the other three sources yielded similar values, averaging around 86%. The higher mass recovery observed for Cottonii may reflect a lower degree of molecular hydration, resulting in less moisture retention during drying.

Based on Table 4, Chondrus carrageenan, despite a lower weight average molecular mass than Spinosum, exhibited comparable or larger radii of gyration (r_w) and higher hydrodynamic radii (r_h)

indicating more expanded coil structures. Spinosum showed the broadest molecular mass distribution and highest polydispersity, consistent with a wide size distribution of ι-carrageenan chains. Cottonii and Gigartina displayed lower weight average molecular mass and smaller hydrodynamic radii, suggesting either shorter chains or more compact conformations. The consistently lower hydrodynamic radii values in these samples point to tighter molecular packing or increased chain flexibility. Mean polydispersity of carrageenan is around two for all carrageenan classes, which is consistent with previously published data [129].

4 INFORMATION EXTRACTED BY PCA

In the context of PAT, transitioning from raw data to actionable insight requires more than just data acquisition – it demands the extraction of meaningful patterns that reflect underlying structural variation. This chapter corresponds to the "Information" tier of the PAT Knowledge Pyramid and aims to distill relevant features from complex analytical measurements of carrageenan using Principal Component Analysis (PCA) [130]. PCA serves as an essential exploratory tool in this thesis, allowing the identification of dominant variation patterns across multiple datasets. Furthermore, it reveals potential links between structural attributes and known functionality metrics, laying the groundwork for the predictive models presented in later chapters.

This chapter demonstrates how different analytical platforms provide complementary—sometimes overlapping—views of carrageenan's structural diversity, forming a critical step toward comprehensive structure–function understanding.

The dataset used is the same as in Section 3.2.

4.1 PREPROCESSING

All data blocks were preprocessed prior to analysis. The preprocessing methods used in this thesis are identical to those described in Paper 2. The main objective of preprocessing in chemometrics is to remove irrelevant variation (artefacts) and to focus on chemical information of interest [131].

For the ICP-MS and SEC-MALS data blocks, autoscaling was applied to ensure that each variable had a mean of zero and a standard deviation of one. This transformation equalizes the contribution of variables with different units or ranges and prevents dominance by those with large magnitudes during multivariate analysis.

The FT-IR spectra of carrageenan were preprocessed using Savitzky–Golay smoothing [132] to reduce high-frequency noise while preserving peak shapes. This was followed by Standard Normal Variate (SNV) transformation [133] to correct for scattering effects and baseline variation often caused by physical differences in sample packing or surface properties. Prior application of SNV, Correlation Optimized Warping algorithm [119,134] was applied to align 1H NMR spectra so the signals corresponding to one proton will not be confused of being from several protons due to slightly different chemical shift caused by e.g. differences in cationic environments. Both spectral data blocks were also mean centered to a zero mean.

Software used for data preprocessing and following analysis were MATLAB (Release R 2023a, version: 9.14.0; The MathWorks Inc.) and PLS_Toolbox 9.2 (2023) Eigenvector Research, Inc. T [135]

4.2 PCA OF CARRAGEENAN FT-IR SPECTRA

A six-component PCA model was constructed using preprocessed FT-IR spectra of 377 carrageenan samples. The model's score and loading plots, along with the assigned spectral bands listed in

Table 2 and data information in subsection 3.2.1, were utilized to extract relevant features for further analysis.

The score plot of the first two principal components is presented in Figure 13A, with the corresponding loading plot in Figure 13B. Along component one, a clear separation is observed between carrageenan samples derived from Cottonii and those from Spinosum. This axis is characterized by strong positive loadings at 1220 cm^{-1} and 805 cm^{-1}, associated with O=S=O asymmetric stretching and C2-O-S in 3,6-anhydrogalactose that are typical marker for ι-carrageenan.

The second principal component is defined by a pronounced loading at 997 cm-1, attributed to glycosidic bond vibrations. This component separates hybrid carrageenan samples, Gigartina and Chondrus classes, from the homopolymeric carrageenan extracted from Cottonii and Spinosum. This suggests that component reflects structural variations in the distribution of κ- and ι-carrageenan units, referred to as "blockiness." According to literature, hybrid carrageenan exhibits a random distribution of ι- and κ-type repeating units [21]. The arrangement of these structural units along the polysaccharide chain can influence local C–O–C and C–O vibrational modes, leading to variations in both intensity and peak position in the FT-IR spectrum. The broader distribution of Gigartina-derived samples along component two may be due to a higher ratio of ι- to κ-type sequences within the molecule, indicating more random distribution (Figure 3). Variation along second component is also observed among Cottonii-derived (κ-dominant) samples. This may be attributed to the presence of small amounts of ι-type disaccharide units within the polysaccharide (supported by information in Figure 3). Despite being present at low concentrations, these units can still influence the vibrational modes detected in this component.

Figure 13 C and Figure 13 D display the score and loading plots for components three and four, respectively. It is primarily Gigartina-derived samples are discriminated along component three, that is characterized by a positive loading at 1120 cm^{-1} – glycosidic bond asymmetric stretching and a negative loading at 840 cm^{-1}, which corresponds to C4-O-S vibrations. Although the exact structural feature driving this separation is not fully understood, it may relate to differences in block distribution.

Principal component four exhibits high loadings at approximately 1260 cm^{-1} and 1040 cm^{-1}, which correspond to cation effects on spectral bands (Table 2). This component primary explains the variation within Cottonii-derived samples, which are mostly spread along axis of component four.

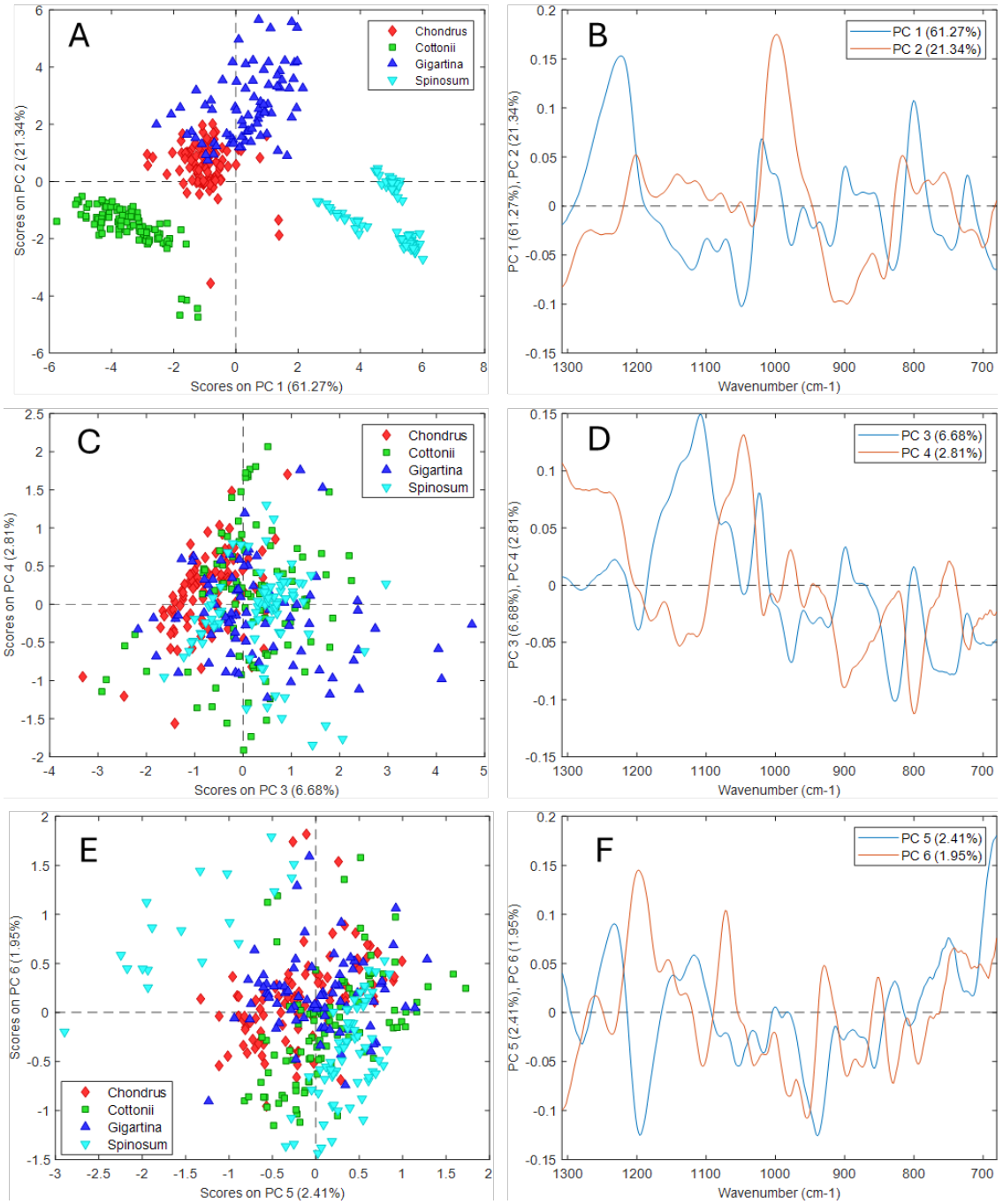


Figure 13 Score and loading plots for the PCA model of carrageenan FT-IR spectra. The score plots (A, C, E) are color-coded according to the seaweed class from which the samples were extracted. The corresponding loading plots (B, D, F) indicate spectral regions contributing to each principal component.

The fifth principal component separates carrageenan samples extracted from Spinosum into two groups, characterized by high loading values at 1235 cm^{-1} and 700 cm^{-1} (Figure 13C and Figure 13D). A closer examination of the data (Figure 14) revealed that this separation correlates with calcium concentrations in samples measured by ICP-MS, as visualized in Figure 14. Interestingly, the samples with lower scores correspond to higher calcium content. This suggests that the variation is not due to a direct spectral shift caused by calcium bindings such as those typically seen in response to divalent cation complexation—but may instead reflect another structural feature that is indirectly associated with calcium levels.

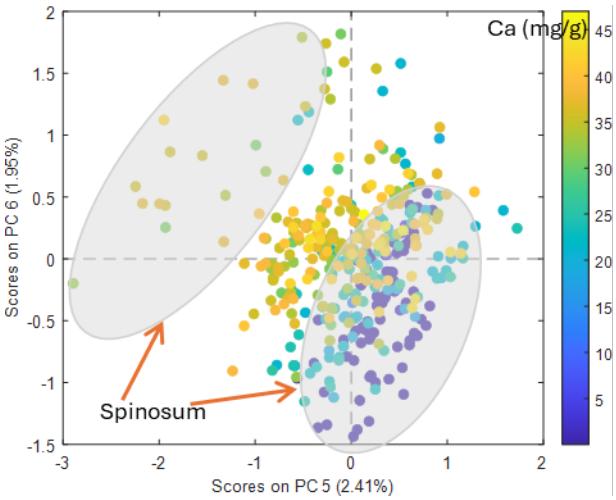


Figure 14 Scoreplot of PC6 vs. PC5 of PCA model for FT-IR spectra of carrageenan where scores are colored by calcium concentration values in corresponding samples measured by ICP-MS.

The sixth principal component is defined by high loadings at 1150, 1070 and 925 cm^{-1}, that together are likely to be an indication for 3,6-anhydrogalactose content. This structural feature is functionally important, as it contributes to the gelling properties of carrageenan. However, PC6 accounts for only a small part of the total variance. This likely reflects the fact that commercial carrageenan products are produced to maximize 3,6-anhydrogalactose content, resulting in minimal variability in this aspect of molecular structure across samples.

4.3 PCA of Carrageenan ^{1}H NMR Spectra

Figure 15 A and B present the score and loading plots for the first two components of a PCA model developed for carrageenan ^{1}H NMR spectra. Separation along component one reflects the relative κ-carrageenan content of the samples, following the expected trend based on species origin: Cottonii > Chondrus > Gigartina >> Spinosum, in agreement with the compositional data shown in Figure 3. This separation is primarily driven by variation in the intensities of the anomeric proton signals at 5.73 ppm (typical for ι-carrageenan) and 5.53 ppm (κ-carrageenan), which show negative and positive loadings, respectively.

The second principal component distinguishes hybrid carrageenans from homopolymeric types. It is defined by high loadings at 5.07 and 4.19 ppm, and low loadings at 4.02 and 4.21 ppm. The peaks with positive loadings correspond to H1 and H5 of the G4S (D-galactose-4-sulfate) residue, while those with negative loadings correspond to H2 and H5 of the same unit. These resonances are common to both κ- and ι-carrageenan, but their chemical environments—particularly near the glycosidic bond—can be influenced by the identity of neighboring disaccharide units. Thus, the second component likely reflects differences in block distribution within the polymer chain.

The third principal component (Figure 15C), with corresponding loadings shown in Figure 15D, is dominated by resonances characteristic of ι-carrageenan. This component does not describe overall

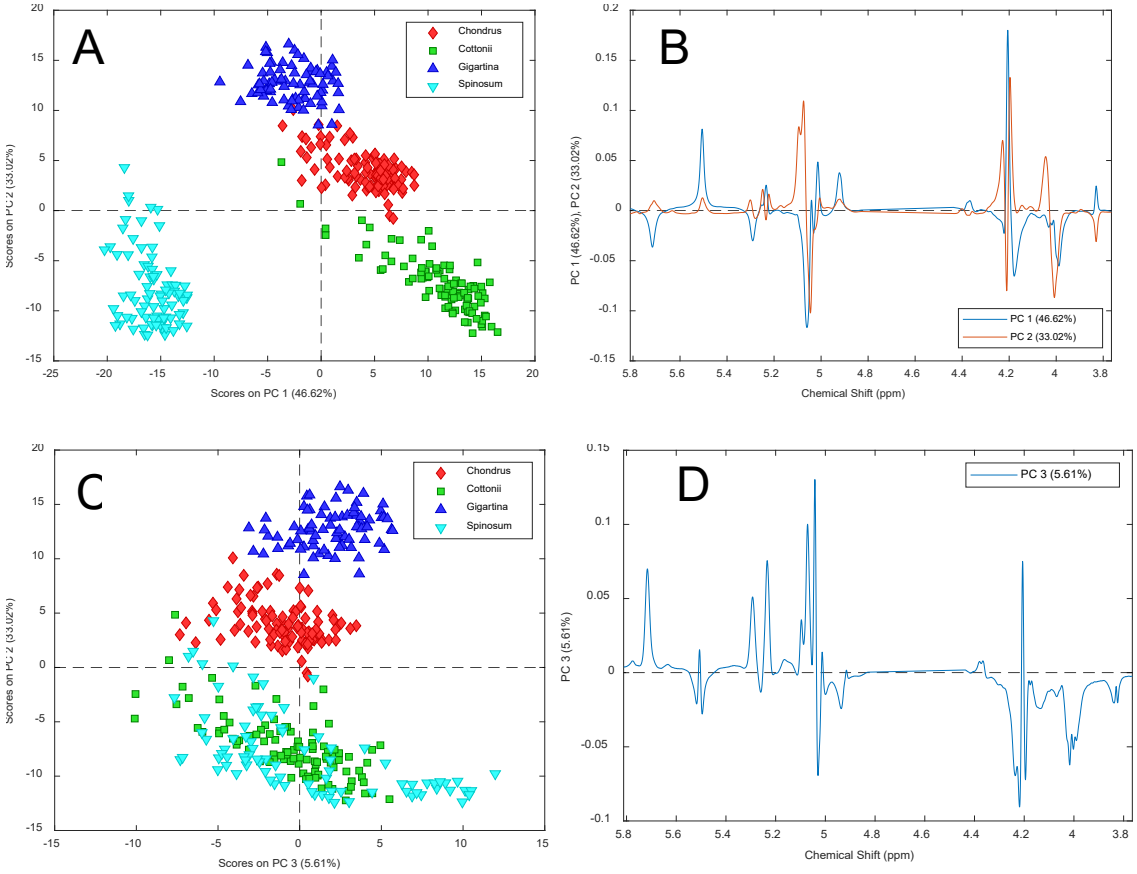


Figure 15 Score and loading plots for the PCA model of carrageenan 1H NMR spectra. The score plots (A, C) are color-coded according to the seaweed class from which the samples were extracted. The corresponding loading plots (B, D) indicate spectral regions contributing to each principal component.

ι-carrageenan concentration across all samples but rather distinguishes a specific subclass within the Spinosum group. These samples likely represent particularly pure ι-carrageenan extracted from Spinosum seaweed. Notably, this same subclass was found to have elevated calcium content in the PCA model based on FT-IR spectra (Figure 14), which is consistent with the known gel-promoting effect of calcium ions on ι-carrageenan [29] and might point to a different processing method of this sample group.

4.4 PCA OF ELEMENTAL COMPOSITION OF CARRAGEENAN

Principal component analysis was applied to elemental composition data for carrageenan acquired by ICP-MS. Figure 16 shows the PCA biplots that visualize distribution of samples along the first three principal components, with overlaid loading indication the contribution of each measured element (sodium, potassium, calcium, sulfur, and magnesium contents). In the component one versus component two plot, the samples are well separated by seaweed source. Cottonii and Gigartina cluster along component one with opposite signs, while Chondrus and Spinosum are more separated along component two. The biplot vectors indicate that Spinosum is characterized by higher levels of sodium and sulfur. Chondrus aligns closely with elevated calcium concentrations, while Cottonii is more associated with potassium. These trends are consistent with known cation preferences and accumulation patterns in different carrageenan-producing seaweeds. Potassium is a gel-promoting cation in κ- carrageenan, which is the dominant form in Cottonii, while sulfur content

is higher in Spinosum, Gigartina and Chondrus due to its greater proportion of ι-carrageenan, which contains more sulfate ester groups, that are neutralized by sodium and calcium.

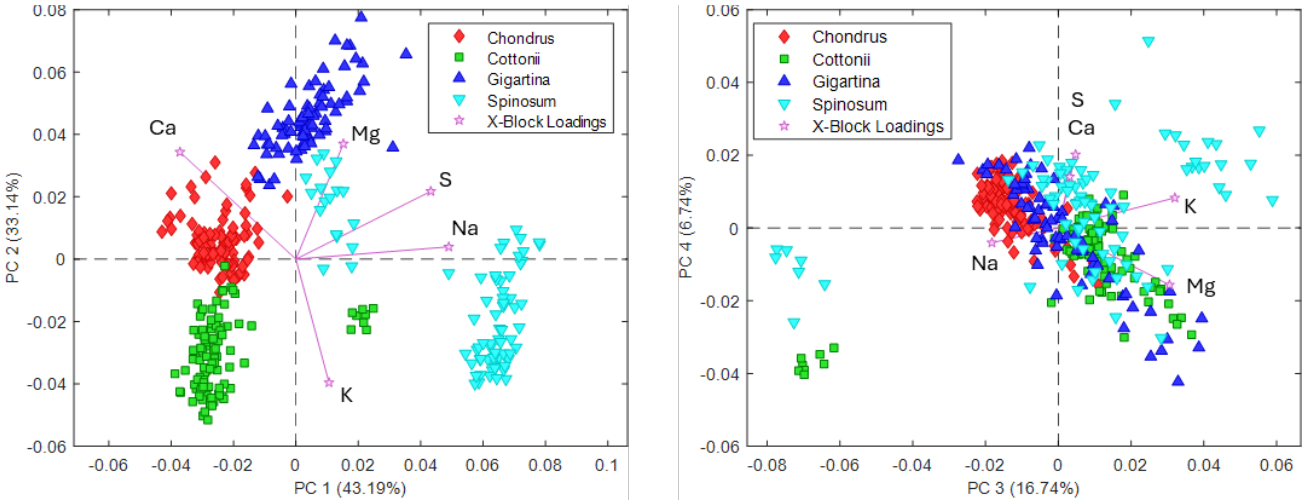


Figure 16 PCA biplots of cation composition of carrageenan extracted from the four seaweed sources. The left panel displays the score plot for PC1 vs. PC2. The right panel shows the score plot for PC3 vs. PC4.

Based on the component three vs. component four plot, Chondrus and Gigartina classes remain relatively clustered, while there are groups of samples that have higher concentrations sodium or potassium, presumably due to a different extraction methods for some batches of carrageenan extracted from Spinosum and Cottonii, as revealed by PCA of FT-IR and ^{1}H NMR spectra.

4.5 PCA OF CARRAGEENAN MEASUREMENTS VIA SEC-MALS

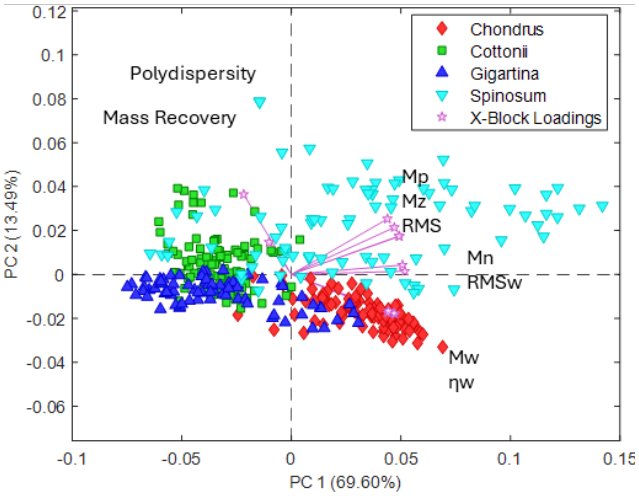


Figure 17 PCA biplots of SEC-MALS data of carrageenan extracted from the four seaweed sources.

Principal component analysis was also performed on the SEC-MALS data, with all twelve variables as an input (Table 4), and reduced to two principal components. The resulting biplot is shown in Figure 17. Spinosum-derived samples exhibit higher scores along component one and display

widespread, indicating substantial variability in the molecular weight among Spinosum-derived carrageenan. The loading vectors indicate that Chondrus-derived samples are associated with higher molecular weight values compared to carrageenan extracted from Gigartina.

Polydispersity and mass recovery show higher loadings along component two. The direction and magnitude of these loadings suggest that carrageenan extracted from Spinosum and Cottonii generally exhibits higher polydispersity and mass recovery compared to that from Chondrus and Gigartina. This is consistent with previous findings in analysis of raw data in Section 3.2.4.

4.6 SUMMARY

Principal component analysis of carrageenans FT-IR spectra, 1H NMR spectra, ICP-MS and SEC-MALS consistently separated samples by seaweed material these were extracted from. The separation was based on content of κ- and ι-carrageenan residues and was in agreement with theoretical ratio of κ- and ι-carrageenan in corresponding seaweed genera, summarized in Figure 3. This separation is likely driven by type-specific functional groups in FT-IR and 1H NMR spectra, cationic profiles in ICP-MS, and molecular size in SEC-MALS corresponding to seaweed class specific differences observed in raw data analysis in Section 3.2. Hybrid carrageenans were separated from homopolymeric types based on glycosidic bond variation, indicating differences in repeating unit distribution. These patterns support feature assignment in later modeling and aid interpretation of structure–function relationships.

5 KNOWLEDGE – STRUCTURE-FUNCTION MODELING

One of the hypotheses in this study is that combining all structural information obtained from the four analytical platforms and modelling it in relation to functionality will enhance our understanding of the relationship between carrageenan structure and functionality. To integrate the data into a unified analytical framework, multiblock data fusion techniques were applied. This approach combines complementary information from different instruments to improve both predictive performance and interpretability. It also allows assessment of whether the selected analytical techniques provide overlapping or distinct structural insights [136,137].

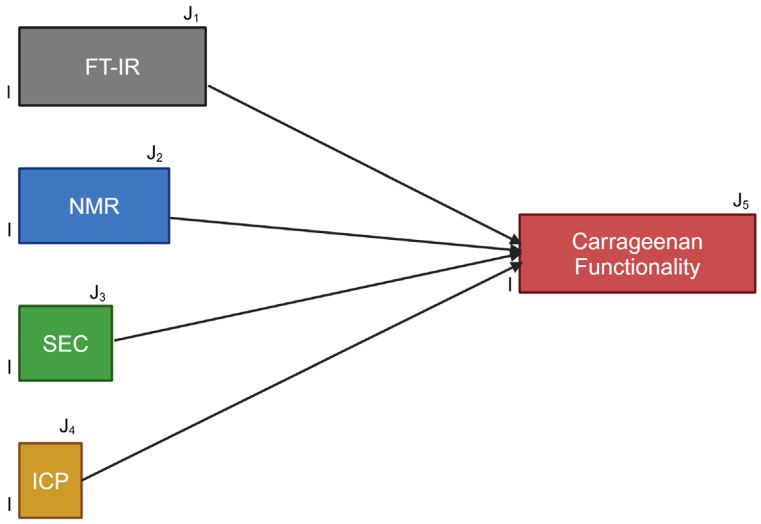


Figure 18 Multiblock set of five blocks, where FT-IR (variable set J_1), NMR (J_2), SEC (J_3), ICP (J_4), are used for prediction and Carrageenan Functionality (J_5) block is to be predicted. All five blocks share the same sample mode (I). The arrows in the figure refer to predictive relationships between four blocks and carrageenans functionality. While the sample mode is the same, the set of variables (J_{1-5}) is different and there are several different types of functionalities for carrageenan but shown here as one block since each functionality were treated as separated problems.

Data fusion approaches are commonly categorized into low-, mid-, and high-level fusion, depending on the stage of data integration. In low-level multiblock data fusion, raw data blocks are concatenated sample-wise into a single matrix, which is then treated as one unified data set for modeling. In mid-level multiblock data fusion, dimensionality reduction is first applied independently to each block—commonly through feature extraction techniques such as PCA. The resulting feature sets are then concatenated into a single array, which serves as the input for modeling. In high-level multiblock data fusion, separate models are developed independently for each block, and the outputs of these individual models are then combined to produce the final prediction [138,139]. In this work, mid-level multiblock data fusion was selected, as it offers a good balance between interpretability and predictive performance. Unlike low-level fusion, it avoids issues with high dimensionality and dominance of single data blocks, and unlike high-level fusion, it retains inter-block relationships [108,140].

Figure 18 schematically illustrates the mid-level data fusion setup used in this study, comprising four predictive blocks: FT-IR spectroscopy, ^{1}H NMR spectroscopy, SEC-MALS (SEC), and ICP-MS (ICP),

with functionality as the response variable. Two distinct functionalities were separately modeled: the apparent viscosity of carrageenan in water, and carrageenan-milk breaking and gel strength

A typical feature of mid-level data fusion is that feature selection or dimensionality reduction is applied before regression modelling. In this case, information from PCA, as discussed in the previous section, was used. Several data fusion methods were tested (described in the following subsections), aiming to address specific questions related to interpretability and to evaluate predictive performance. All blocks were preprocessed as described in Section 4.1, and each variable was averaged across two replicate measurements. The Kennard–Stone algorithm [141] was used to divide the dataset into a calibration set (80%) and a validation set (20%). The validation set was withheld until after model interpretation to ensure unbiased assessment. Model performance was evaluated using cross-validation and independent test sets, with predictive accuracy assessed by root-mean square error of cross-validation (RMSECV) and prediction (RMSEP) [142]. Cross-validation was performed by 10 folds cross-validation on randomized data without replicates. The optimal number of latent variables (=components) was chased based on RMSECV minimization. Variable selection was also tested, but as it did not improve prediction or interpretability, it was not applied. However, it should be noted that some data fusion methods include integrated approaches for variable selection [143,144].

Furthermore, individual Partial Least Squares (PLS) regression models [145] were developed from each block. These serve as benchmarks for comparison with the multiblock results and are discussed primarily in papers 1 and 2 [16,17].

5.1 UNDERSTANDING CARRAGEENANS VISCOSITY

Paper 1 describes a PLS regression model developed for viscosity prediction based on 503 SEC-MALS measurements. The input variables included the original twelve SEC-derived parameters from Table 4, along with their squared, cubic, quartic, and cross-product terms. Separate models were constructed for each carrageenan class. The highest predictive accuracy was achieved for Spinosum (RMSEP = 12.4), and the lowest for Cottonii (RMSEP = 24.0). The relationship between structure and function was reflected in the model variable selection, where carrageenans' molecular weight, intrinsic viscosity and hydrodynamic radius consistently emerged as the most informative predictors of apparent viscosity.

Size-exclusion chromatography combined with chemometric modeling offers a more efficient and sustainable alternative to traditional viscosity testing. However, its qualification as a true PAT method may be questioned, as it remains a destructive technique that requires sample preparation. This raises the question of whether any of the four analytical platforms evaluated could offer more suitable alternatives for quality control monitoring. Additionally, investigating the relationships between carrageenan structure and its viscosity remains essential for understanding how molecular characteristics influence functional performance.

To investigate this, multiblock data fusion methods applied to a dataset consisting of four analytical blocks and one response block (viscosity), covering carrageenan samples extracted from four seaweed sources: Chondrus, Cottonii, Gigartina, and Spinosum. The dataset comprised 362 samples in total, with 100, 100, 78, and 84 samples from each source, respectively. The number of samples differs from Paper 1 due to the reduced availability of matched measurements across all

five blocks. As the multiblock approach in common sample mode requires identical sample sets across blocks, only samples with complete data were included. The response variable (viscosity) was log-transformed prior to modeling.

5.1.1 Parallel Orthogonalized – PLS for carrageenan viscosity

Parallel Orthogonalized PLS (PO-PLS) regression was used to explore how data from four analytical platforms—ICP, SEC, IR, and NMR—relate to each other and to carrageenan functionality, specifically aqueous viscosity. The method was selected because it separates information into common and distinct components [136], which allows assessment of which features are shared across techniques, and which are unique. This distinction is important both for understanding structure–function relationships and for selecting models with better interpretability and predictive value [146]. This is achieved by applying canonical correlation analysis (CCA) [147] to dimension-reduced representations of each block (typically obtained via PCA) and identifying those linear combinations of variables in the blocks that are maximally correlated. Only those components with canonical correlations above a pre-defined threshold are retained as common. These shared components are first used to model the response. The residuals from this regression are then modeled using the block-specific unique components extracted by PLS [148].

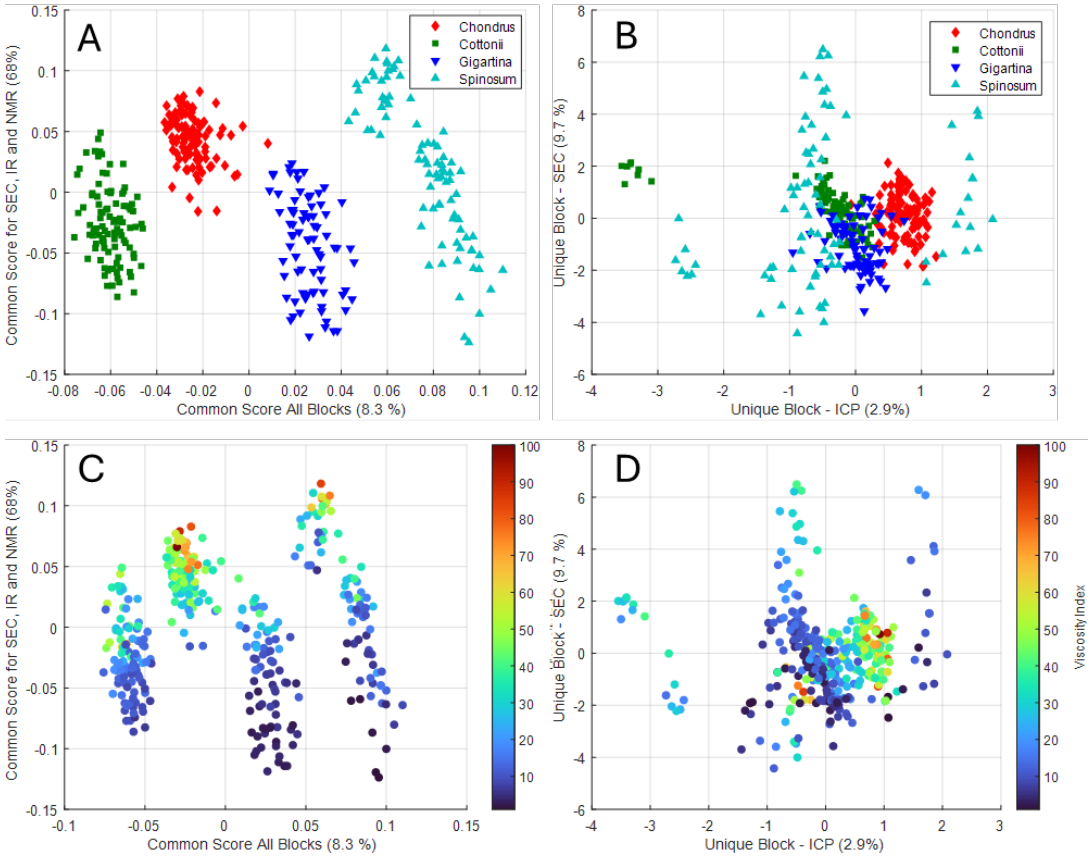


Figure 19 PO-PLS score plots for the two selected common and two distinct components. Subplots A and B display the sample distributions by class, while C and D use color gradients to reflect the viscosity index. The explained variance for the response by each component is given in parentheses.

In this study, PO-PLS was implemented using a MATLAB function and script developed by Dr. Ingrid Måge [149] adapted to fit the dataset structure. Each data block was initially modeled separately

using PLS regression. Generous numbers of components were retained in each case to avoid discarding potential signals that could later be identified as common across platforms. This resulted in models with five components for ICP, twelve for SEC, ten for IR, and ten for NMR. The variance explained in viscosity was 22% for ICP, 81% for SEC, 50% for IR, and 69% for NMR.

Eleven candidate sets of common components and four distinct block contributions were evaluated. The selection criteria for common components were a minimum canonical correlation of 0.90 and at least 3% explained variance in each contributing block. High canonical correlation combined with low explained variance was considered indicative of noise and such components were excluded, following standard recommendations [146,150]. The distinct blocks contributors were selected based on information obtained by PLS regression developed on unique part of the data that was left after removal of the common information from the original data by orthogonalization.

The final PO-PLS model included two common and two unique components, together explaining 89% of the variation in viscosity. One common component was shared across all four blocks, while the second was common to SEC, IR, and NMR. The two unique components represented variation specific to the ICP and SEC blocks, respectively. Figure 19 shows the associated score plots. In panels A and B, samples are colored by class. In panels C and D, a color gradient shows the viscosity values that clearly indicate that viscosity increases along second component. The second common component explains the largest share of response variance, 68% and captures within-class variation. The first common component explains only 8.3% of the response but clearly separates samples by seaweed species they were extracted from. It was retained because it explains substantial X-variance across all blocks (34% in ICP, 13% in SEC, 53% in IR, and 43% in NMR) and had a canonical correlation coefficient above the threshold. This component confirms that information about carrageenan type is embedded in all four chemical measurements, mostly pronounced in the spectral data and less in ICP and SEC-blocks. Inspection of the loadings for common component one in Figure 20 indicates that the variation captured relates to carrageenan type, where the IR and NMR spectra are particularly rich in this information. Loadings from IR, in panel C, resemble the principal component from PCA of the FT-IR spectra in Figure 13, with markers assigned to ι-carrageenan. The NMR loadings show dominant features corresponding to ι-carrageenan. Supporting this, loadings of the ICP block in panel A pictures typical for ι-carrageenan cation composition with a high sulfur content. The SEC loadings in panel B show moderate contribution in this component and resemble information extracted by the corresponding PCA model (Figure 17) that indicated the higher weight average molecular mass and lower recovery for a part of samples from the ι-type rich carrageenan from Spinosum. Together these patterns suggest that ι-type concentration is captured across all blocks but is only weakly predictive of viscosity.

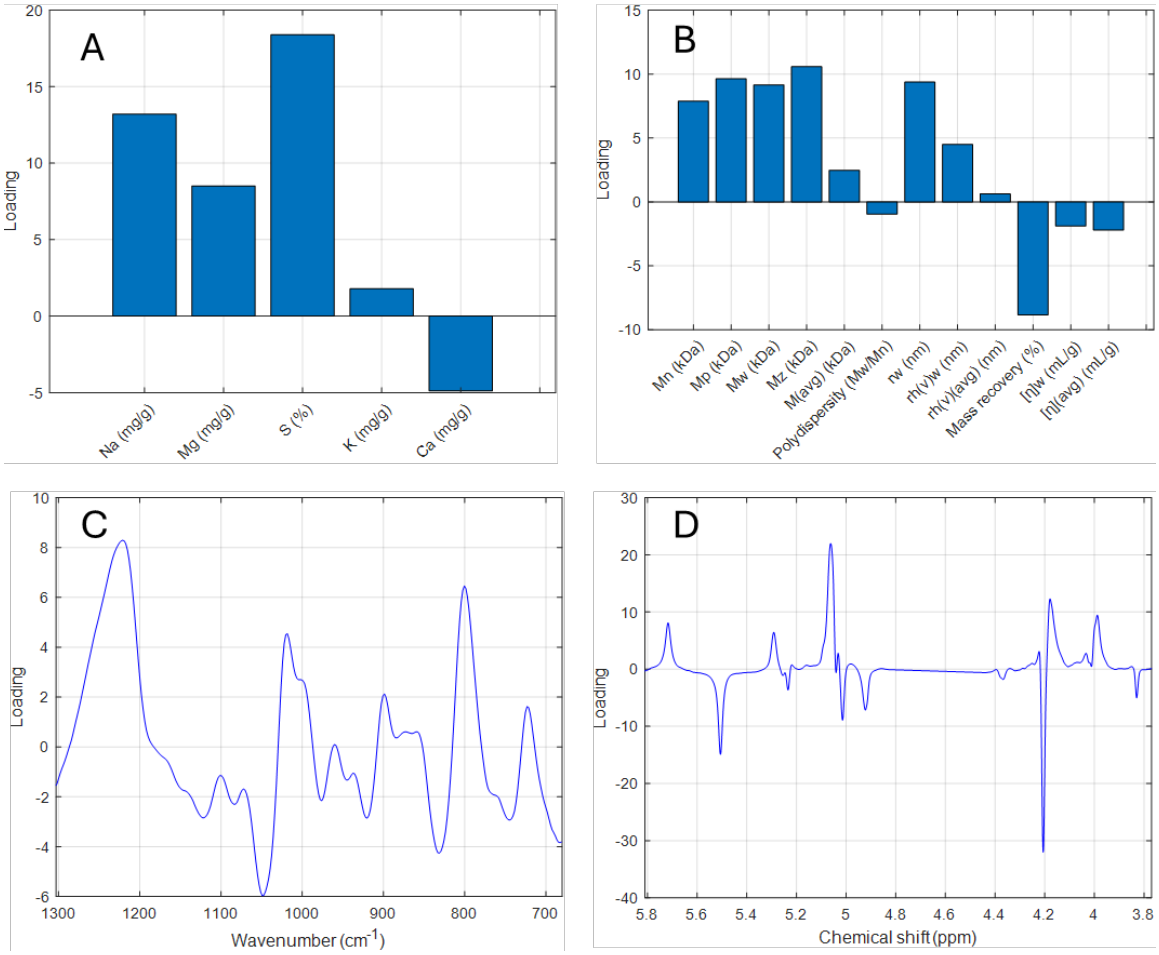


Figure 20 PO-PLS loading plots for component one including A) ICP-block; B) SEC-block; C) IR-block; and D) NMR-block.

Inspection of the loadings for common component one in Figure 20 indicates that the variation captured relates to carrageenan type, where the FT-IR and ^{1}H-NMR spectra are particularly rich in this information. Loadings from the IR block, in panel C, resemble the principal component from PCA of IR spectra in Figure 13, with markers assigned to ι-carrageenan. The NMR loadings show dominant features corresponding to ι-carrageenan. Supporting this, loadings of the ICP block in panel A pictures typical for iota-carrageenan cation composition with a high sulfur content. The SEC block loadings in panel B show moderate contribution in this component and resemble information extracted by the corresponding PCA model (Figure 17) that indicated the higher weight average molecular mass and lower recovery for a part of samples from the ι-type rich carrageenan from Spinosum. Together these patterns suggest that ι-type concentration is captured across all blocks but is only weakly predictive of viscosity.

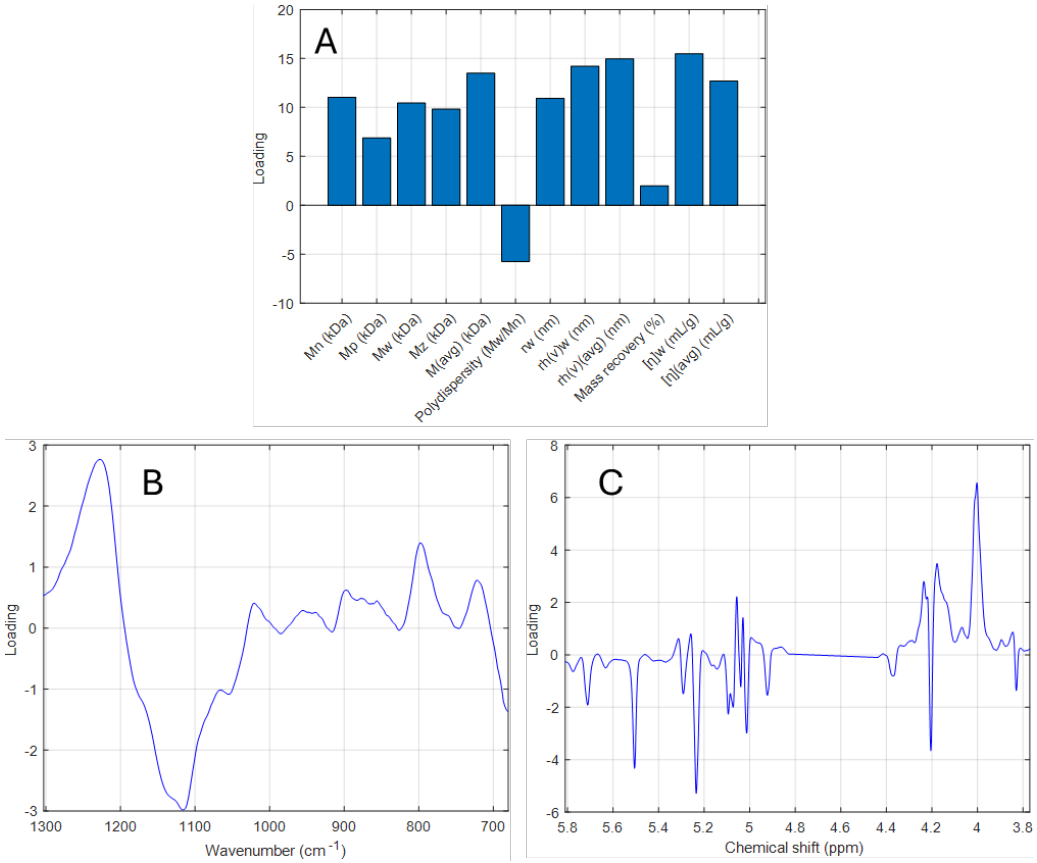


Figure 21 PO-PLS loading plots for component two including A) SEC-block; B) IR-block and C) NMR-block.

The second common component, responsible for most of the response variance, links features in three blocks, SEC, IR, and NMR, that explain 36, 9 and 4 % of viscosity, respectively. The loading plots for the second component are presented in Figure 21. In the SEC block, panel A, negative loadings are seen for polydispersity and mass recovery, while positive loadings are observed for the weight average intrinsic viscosity (η_w), weight-average molecular mass (M_w) and related parameters. This in combination with color gradient in Figure 19C, indicate that an increase in viscosity increases with second component, implies that samples with larger, more uniform molecules are associated higher viscosity, while high polydispersity is associated with reduction in it, likely due to a greater proportion of short chains. In the IR block, panel B, negative loadings appear around 1120 cm^{-1} and positive loadings near 1220 cm^{-1}, corresponding to glycosidic bonding and sulfate substitution, respectively. While the latter is indicative of overall carrageenan content, the former has previously been attributed to blockiness (see sections 2.1.1 and 4.2) and may also reflect chain ends or molecular heterogeneity, potentially correlating with SEC-derived polydispersity. The NMR loadings reinforce this interpretation, showing negative contributions near 5.25 ppm, corresponding to H-2 in G4S units of κ-carrageenan, and positive contributions around 4.00 ppm, which may be affected by neighboring molecules and change depending on blockiness [72,121]. These observations suggest that the second principal component captures key aspects of molecular size, abundance, and topology. Viscosity increases are associated with intrinsic viscosity increase, while specific forms of blockiness and elevated polydispersity are associated in reduction of it—likely due to a higher proportion of shorter chains. This hypothesis could be further evaluated through detailed SEC-MALS analysis of individual carrageenan fractions [151]; however, such an investigation lies beyond the scope of the present study.

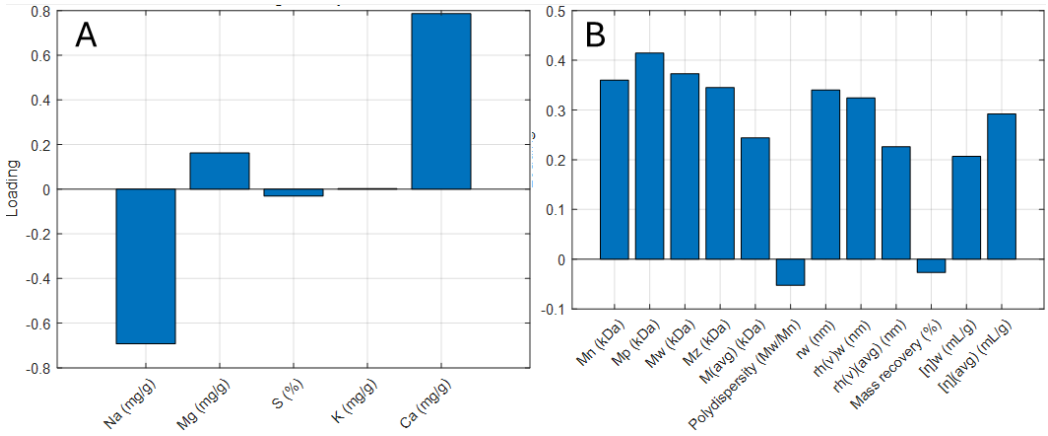


Figure 22 PO-PLS plots for two distinct components for A) ICP-block and B) SEC-block.

The model incorporates two distinct components, each capturing specific sources of variation in the response variable. The first, derived from the ICP block, accounts for 2.9% of the variation in viscosity, while the second, originating from the SEC block, explains 9.7%. The corresponding score plots are presented in Figure 19B, with the associated loading plots shown in Figure 22A for the ICP block and in Figure 22A for the SEC block.

For the ICP block the scores distribution in Figure 19B closely mirrors the clustering observed in the PCA score plots in Figure 16, particularly in separating Spinosum and Cottonii samples with elevated sodium content from the remaining dataset. The pattern supported by the loadings in Figure 22A, which indicates an importance on sodium and calcium content. Notably, a similar calcium-related pattern was also evident in PCA of the FT-IR spectra, Figure 14, yet the corresponding IR block was not retained as a distinct component in PO-PLS model for viscosity, suggesting that this signal was either insufficiently predictive or too diffuse when considered jointly.

A comparable situation is observed in the SEC block. The loading plot in Figure 22B, together with the score distribution in Figure 19B, highlights variation within the Spinosum-derived samples. These samples exhibit a broad range of in the weight average molecular mass (Figure 17).

Taken together, these findings suggest that the inclusion of these distinct components reflects variation introduced by group-specific differences. Therefore, their relevance appears limited to a small subset of samples, implying that these components may not be broadly generalized across the dataset.

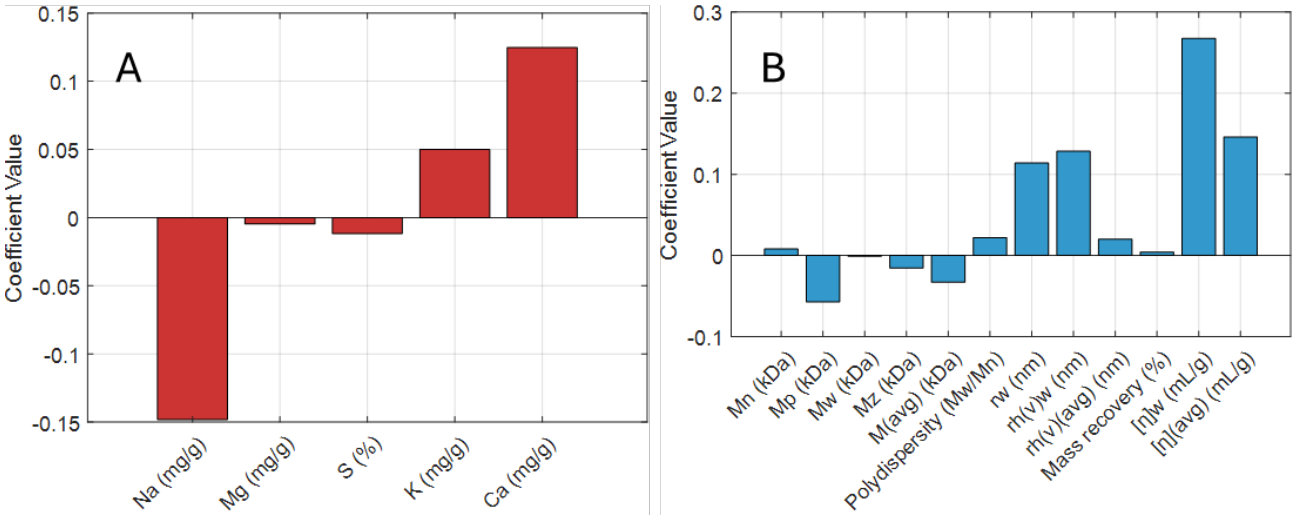


Figure 23 Regression coefficients for the unique components from ICP (left panel) and SEC (right panel) blocks of PO-PLS model predicting carrageenan viscosity.

Regression coefficients for the PO-PLS components from the ICP block shown in Figure 23A indicate that carrageenan with high potassium and calcium concentrations, but low sodium concentrations is are associated with increase viscosity. The regression coefficient from the SEC block indicates that viscosity increases with increasing values weight-average intrinsic viscosity, uncertainty-weighted average intrinsic viscosity, and weight-average mean square radius.

In summary, the most relevant block for predicting viscosity is SEC-MALS data, as it is included in all components. The information on carrageenan type, while clearly present in all blocks, does not appear to strongly influence viscosity directly but may be relevant for explaining other functional attributes. Notably, FT-IR and ^{1}H-NMR spectra are dominated by signals related to carrageenan type, making them suitable for structural classification, though not necessarily for viscosity prediction.

Several combinations of component numbers and cutoff thresholds were tested. Cross-validation yielded RMSECV values between 8 and 10. However, when tested on an independent validation set, the model gave a much higher RMSEP of 22. This result suggests overfitting and highlights the sensitivity of PO-PLS to model choices, including the number of components and canonical correlation threshold. Increasing the correlation threshold removed the ICP block as a contributor without substantially improving performance. Because of these limitations, the model is not considered suitable for industrial prediction of viscosity, but it provides a useful framework for examining structure–function relationships and for identifying which blocks contribute most to key properties.

5.1.2 Sequential and Orthogonalized PLS for carrageenan viscosity

Sequential and Orthogonalized Partial Least Squares (SO-PLS) is a multi-block regression method that models a response variable using multiple input data blocks, each entered the model sequentially. For each block, a PLS regression is performed, and any redundant information shared with previous blocks is removed by orthogonalization before modeling. This ensures that each block's contribution is interpreted as an additional one to what has already been modeled. The procedure works as follows: the first block is modeled using standard PLS against the response. The second block is orthogonalized with respect to the scores from the first block's PLS model, and then this orthogonalized second block is used in a new PLS model to fit the residuals from the first model. This process continues for all subsequent blocks. The final prediction is the sum of all block-specific predictions. This method allows individual optimization of the number of components for each block and enables interpretation of block-specific contributions [152–154]. Unlike PO-PLS, which treats all blocks equally, SO-PLS models them step by step, allowing interpretation of how much additional information each block provides beyond the previous ones. This is useful when there is a logical structure in the data, such as when moving from basic composition (ICP) to more detailed structural information (NMR). Furthermore, in the context of PAT, SO-PLS can help identify whether some blocks can be omitted without significant loss of predictive performance, supporting the selection of faster and more cost-effective measurement techniques.

The order in which data blocks are introduced in SO-PLS influences model structure, complexity, and predictive performance [153–155]. This is a direct result of the algorithm's sequential nature: each block is orthogonalized with respect to all previous blocks and can only explain residual variance not already captured. Consequently, blocks introduced earlier are prioritized in explaining the response, while those placed later have access only to the residual signal.

In this work, SO-PLS models were developed using a MATLAB function provided by prof. Federico Marini [156]. A total of 24 models were constructed, covering all possible permutations of the four analytical blocks. As summarized in Figure 25, the results clearly show that block order influences both prediction error and the number and origin of selected components. A consistent pattern across all models is the presence of multiple components from the SEC block, regardless of its position in the sequence. In contrast, the ICP block contributed components in only three models (models 2, 7, and 9), and only when it was placed first in the sequence; in all other permutations, no components were selected from the ICP block. Models beginning with the IR block consistently performed better, yielding lower prediction errors and requiring fewer components, while still incorporating the SEC block as the second block. Conversely, models starting with the NMR or SEC block often resulted in more complex structures with weaker predictive performance. The models with the lowest prediction errors typically included 8–10 components from IR and 4–6 components from SEC. The optimal model, defined as having both low prediction error and minimal complexity, was model 16, which followed the block order: NMR, ICP, IR, and SEC, and retained 0, 0, 10, and 6 components, respectively. Based on these results, it was decided to develop a model using only the IR and SEC blocks, focusing on how structural features measured by these techniques are associated with the viscosity of carrageenans.

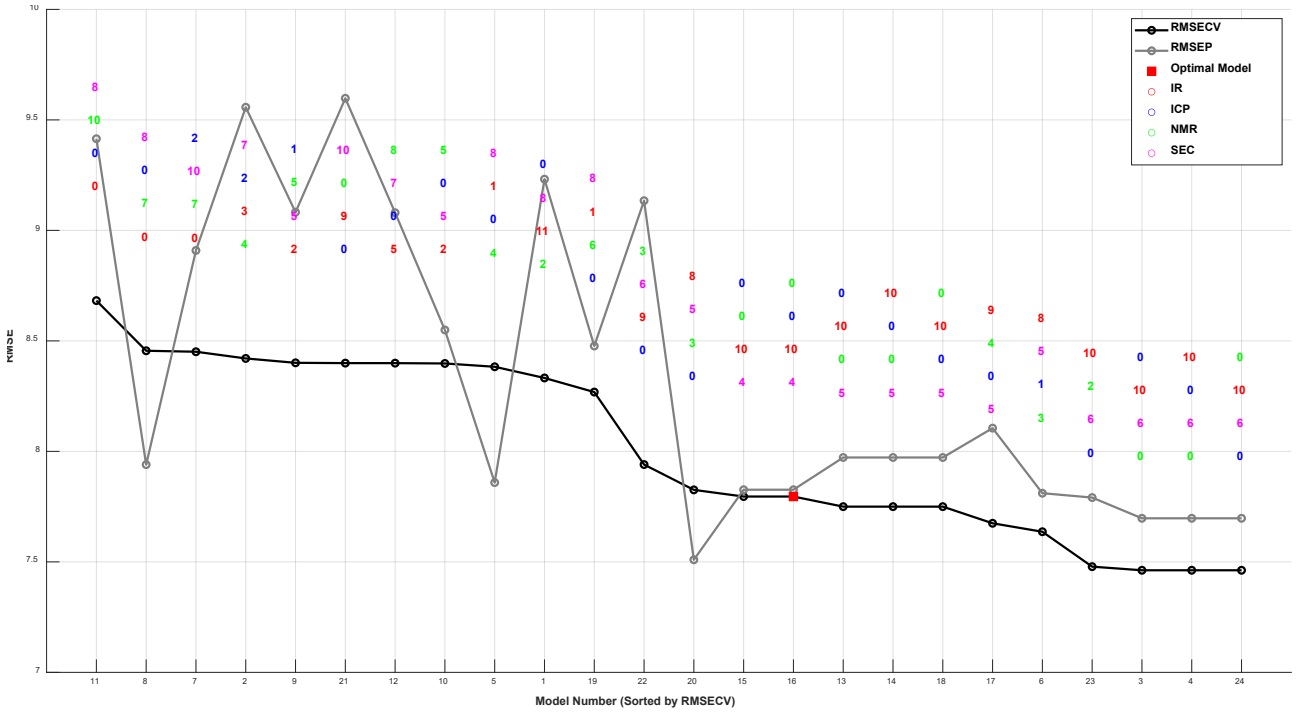


Figure 25 Performance of 24 SO-PLS models trained on viscosity data using four analytical blocks: ICP, SEC, IR and NMR. Each model uses a different block order. For each model, the optimal number of components extracted from each block is displayed above the model number, color-coded by analytical platform (red = IR, blue = ICP, green = NMR, magenta = SEC) and in the order these were included in the model. The red square highlights the optimal model with the lowest RMSECV, RMSEP, and modest model complexity (total number of components and number blocks used). This visualization approach is inspired by Paper 2 [16].

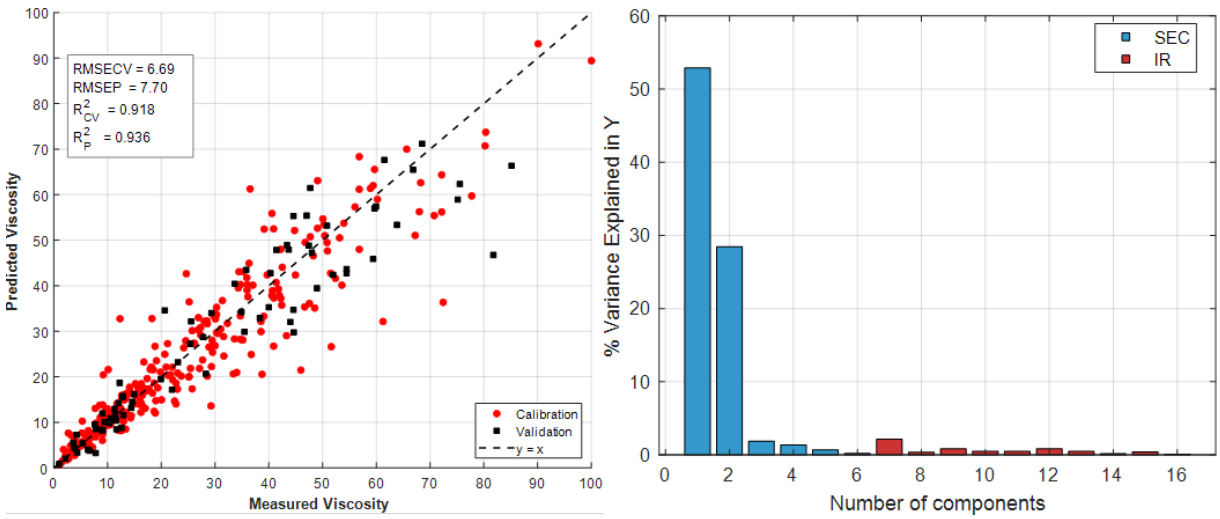


Figure 24 Left panel: Measured vs. predicted viscosity for the SO-PLS model developed using six components from the SEC block and ten components from the IR block. Calibration samples (n = 287) are shown in red, and validation samples (n = 75) in black. Performance metrics (RMSECV, RMSEP, R²CV, R²P) indicate strong predictive power and model robustness.
Right panel: Percentage of variance in viscosity explained by each component in the SO-PLS model, highlighting the dominant contribution of the SEC block in capturing the variance in viscosity, while IR components provide minimal explanatory power.

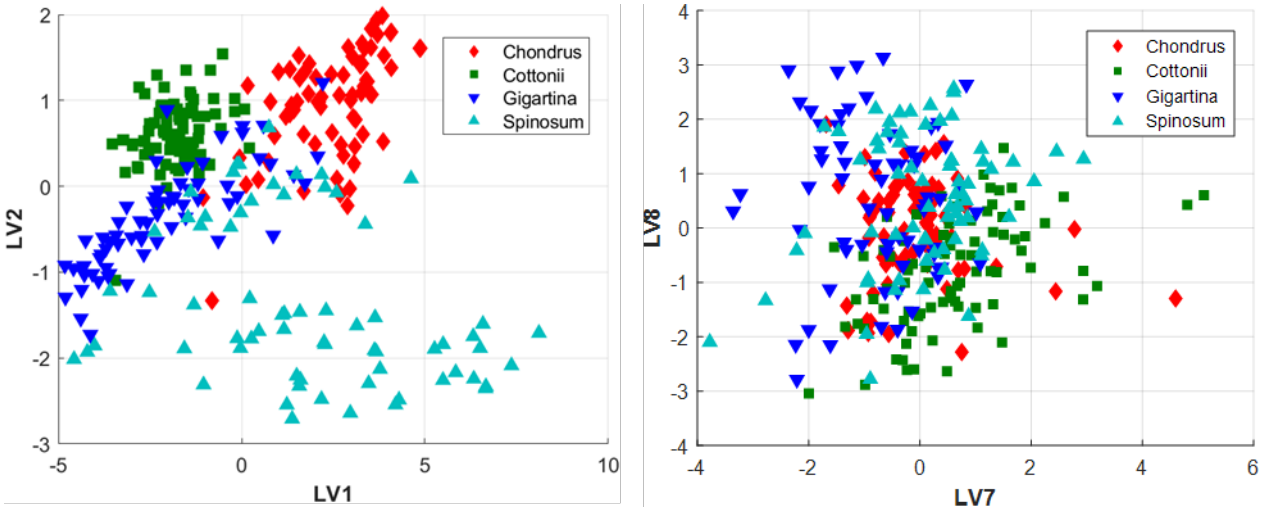


Figure 26 Left panel: Score plots of component 1 vs. component 2 (LV1 vs. LV2) for the SEC block for the SO-PLS model. Right panel: Score plots of component 7 vs. component 8 (LV7 vs. LV8) for the IR block for the SO-PLS model. Both plots are colored by the seaweed species from which carrageenan samples originated.

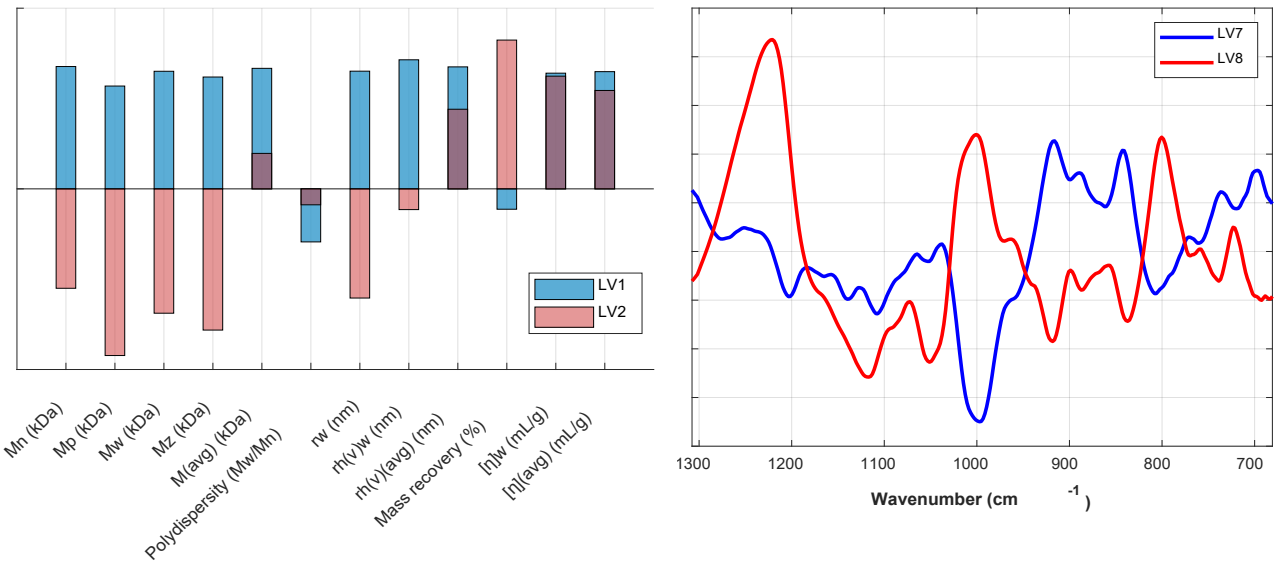


Figure 27 Left panel: Overlay of loadings for component 1 and component 2 (LV1 and LV2) for the SEC block in the SO-PLS model, indicating variable contributions to each component.
Right panel: Overlay of loadings for component 7 and component 8 (LV7 and LV8) for the IR block, highlighting spectral regions contributing to each component.

Figure 24 left panel shows measured versus predicted viscosity index by the SO-PLS model developed using the SEC and IR blocks, it demonstrates strong predictive performance in both calibration and validation datasets. The RMSECV of 6.69 and RMSEP of 7.70 indicate minimal overfitting, with comparable predictive accuracy in both calibration and validation datasets. The right panel illustrates the variance in viscosity explained by each component in the SO-PLS model. The first two components from the SEC block account for over 80% of the variance, emphasizing the dominant predictive contribution of SEC data. In contrast, the IR block components contribute minimally to the explained variance, each accounting for less than 5%. The score plots for the first two components, both from SEC-MALS, are shown in Figure 26, left panel. It illustrates a distinct gradient along the first component, with samples from Spinosum displaying a wide score distribution. Chondrus, Cottonii, and Gigartina samples are positioned from high to low scores, with Gigartina and Cottonii showing partial overlap. Score values corresponding to samples extracted from Spinosum seaweed are spread along component one, and then from high to low score values for samples extracted from Chondrus, Cottonii and Gigartina with the last partly overlapping. The corresponding loading plot in Figure 27, left panel, reveals that the first component is characterized by high positive loadings for intrinsic viscosity (η_w), the weight average molecular mass and related molecular weights parameters. Polydispersity and carrageenan content display negative loadings, indicating an inverse relationship with these molecular weight parameters. This pattern is consistent with the loading structure observed in the unique PO-PLS component for SEC-block in Figure 22 and the PCA biplot in Figure 17. The second component separates a subset of Spinosum samples, as seen in the score plot. The corresponding loadings are characterized by high positive values for mass recovery and intrinsic viscosity, while weight average molecular (*Mw*), number average molecular mass (*Mn*), molar mass at the top of the concentration peak (*Mp*), weight average molecular mass (*Mw*), Z-average molar mass (*Mz*) exhibit strong negative loadings. This pattern contrasts with the global common component in the SEC-block PO-PLS model in Figure 20B, where it was associated with iota content.

For the IR block, the score plot for components seven and eight in Figure 26 right panel does not show a clear trend based on carrageenans seaweed origin. The loading plot for the first component Figure 27, right panel, reveals high positive loadings at 805 cm^{-1}, 997 cm^{-1}, and 1230 cm^{-1}. These bands were previously characterized in the PCA loading plot (Figure 13B), where 805 cm-1 and 1230 cm^{-1} were associated with sulfate ester functionalities (O=S=O and C2-O-S in 3,6-anhydrogalactose), typical markers for ι-carrageenan. The band at 1000 cm^{-1}, associated with glycosidic bond vibrations, was noted as a potential marker for structural blockiness, although direct measurements of blockiness were not available. The second IR component is dominated by a strong negative loading at 1000 cm^{-1}, a region associated with blockiness, and positive loadings at 835 cm^{-1} and 925 cm^{-1}, which correspond to C4-O-S in galactose and 3,6-anhydrobridges. The 925 cm^{-1} band is linked to the $^1C_4$ chair conformation of 3,6-anhydrogalactose, a structural feature known to facilitate gel formation [19].

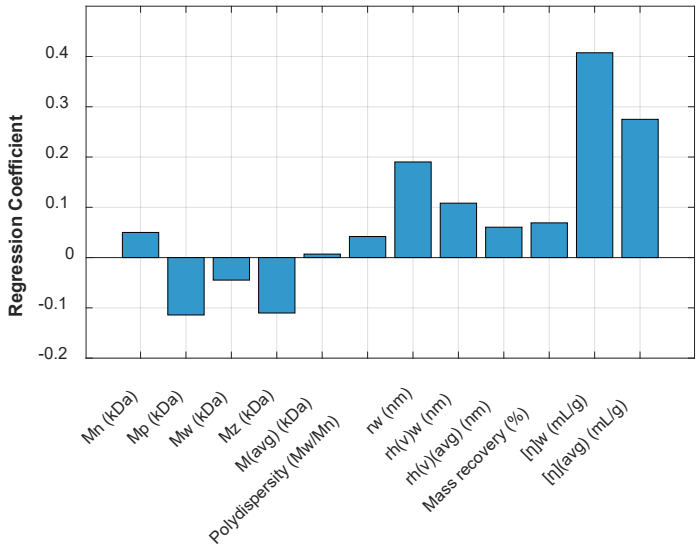


Figure 28 Regression coefficients for the SEC block in the SO-PLS model, illustrating the influence of each molecular parameter on the predicted viscosity.

Figure 28 presents values of regression coefficients for the SEC block in the SO-PLS model. The most prominent positive coefficients are associated with weight-average intrinsic viscosity, uncertainty-weighted average intrinsic viscosity, and weight-average mean square radius. These variables are informative as they encapsulate both molecular size distribution and cationic composition, providing a comprehensive measure of the molecular structure and its potential impact on viscosity [23]. These patterns are very similar to what was revealed by analysis of PO-PLS model.

The regression coefficients for the IR block are not presented because when all ten components were included, the resulting profile appeared overly complex, potentially indicating overfitting. However, the RMSEP (Figure 24) for the full model remains acceptable, suggesting that the model may still generalize reasonably well despite the challenging interpretation of the regression coefficients.

5.1.3 Response Oriented Sequential Alteration

In the response oriented sequential alternation (ROSA) blocks are used multiple times as instead of analyzing all data together each block is treated as a separate source of structural information. At each step in the development of the PLS regression model, components are calculated separately for each block. The component that best predicts the response variable (i.e., minimizes the residuals) is selected and added to the model. Once a component is selected, its information is removed from all data blocks by projecting them onto a space orthogonal to the selected component. This process ensures that each subsequent component is based on new, non-redundant information, resulting in a model that systematically integrates the most predictive and unique information from each data block [157]. In this work, ROSA models were developed using a MATLAB function provided by prof. Federico Marini [156]. This method was selected to explore which specific information within each block is most effective for predicting viscosity. While previous analyses indicated that the four blocks (SEC, ICP, IR, NMR) share considerable common information regarding viscosity, ROSA model allows to assess whether specific structural features are more predictive when considered independently. Thus, ROSA model can reveal whether a particular data source (e.g., SEC molecular

size, IR spectral bands) is more informative for certain components than others, shedding light on unique and dominant predictive features across and within blocks.

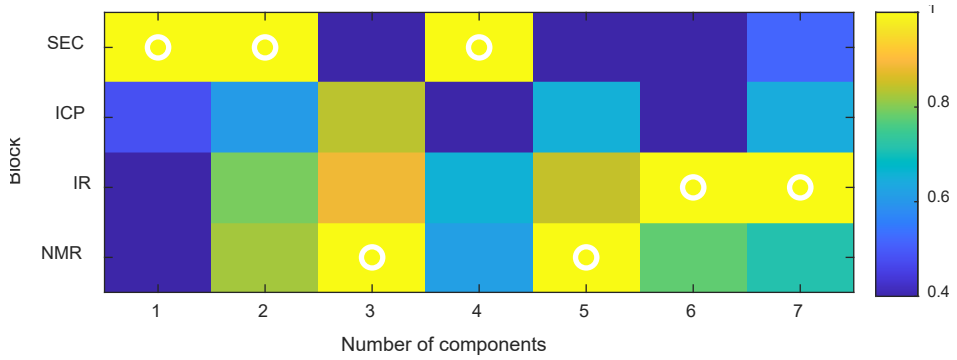


Figure 29 Correlation between candidate block scores and selected component scores across latent variables in the ROSA model. For each component (x-axis), the correlation between each block's candidate score and the final selected score is shown. White circles indicate the block selected by ROSA for that component. High correlation values suggest redundancy or similar predictive structure across blocks; low correlation highlights unique contribution from the selected block. The optimal number of latent variables was seven based on RMSECV minimization. Inspired by Smilde at al. [138].

Figure 29 illustrates the block selection process in ROSA and visualizes the correlation between the selected score vectors from the winning blocks and those from the remaining blocks. Across the seven components retained, the scores were selected from the SEC (three components), NMR (two components), and IR (two components) blocks, with no components selected from the ICP block.

Notably, the first two and fourth components were derived from the SEC block. These scores showed minimal correlation with those from other blocks, indicating that SEC captures unique information strongly associated with viscosity. This aligns with previous findings, where weight-average molar masses and intrinsic viscosity were identified as key predictors. The third component, however, originated from the NMR block and showed a significant correlation with both IR and ICP scores, suggesting overlapping information between these blocks.

Upon examining the weight plots for SEC, no new structural information was revealed beyond the expected dominance of weight-average molar masses and intrinsic viscosity, reinforcing the primary role of molar mass distribution in viscosity prediction already explained in previous sections and Paper 1. Similarly, the final two components from the IR-block that were characterized by information related to carrageenan concentration and structural blockiness, as indicated by spectral bands around 1180–1110 cm^{-1} (vibrations of glycosidic bonds and C-O-C) and 1225 cm^{-1} (O=S=O vibrations, influenced by sulfate content).

The NMR weight plot shown in Figure 30 revealed strong negative contributions from chemical shifts at 5.73, 5.07, and 4.19 ppm. These signals correspond to anomeric protons and the H-5 proton in the G4S unit of ι-carrageenan Table 3. Interestingly, ROSA prioritized this NMR-derived component over the correlated components from the IR and ICP blocks, suggesting that it captured the relevant structural variation more directly or effectively—particularly in relation to ι-carrageenan. The selection of this NMR-derived component is intriguing, as it also correlated with a component in the ICP-block, likely due to a distinct cationic profile associated with ι-carrageenan, and with the IR-block, specifically at 805 cm^{-1}, a band characteristic of ι-carrageenan.

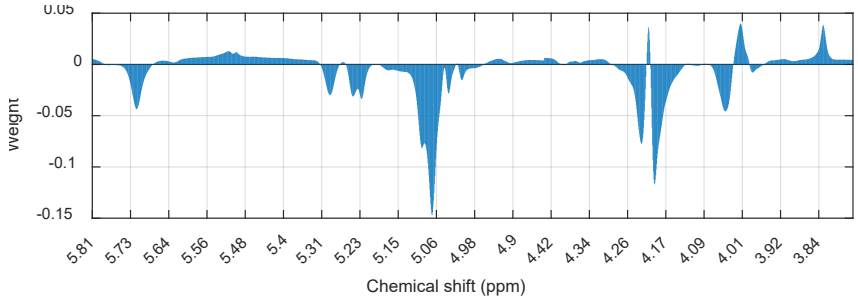


Figure 30 ROSA loading weights for component 3 from NMR-block.

Although NMR data for carrageenan are often difficult to interpret due to inherently low resolution, the clear resonance of the anomeric proton at 5.73 ppm provided a clear marker for ι-carrageenan, allowing the ROSA model to effectively isolate and prioritize this information.

Models' cross-validation performance reached a RMSECV of 8.6 with a corresponding R^2CV of 0.79. External prediction using a separate validation set yielded an RMSEP of 9.6 and an R^2P of 0.82, confirming that the model generalizes reasonably well.

5.1.4 Multiblock PLS Regression for carrageenan viscosity prediction

The multiblock PLS regression methodology was applied by concatenating four data blocks into a single matrix [158], followed by PLS modeling using the PLS_Toolbox [135]. Prior to concatenation, each block was dimensionally reduced using PCA, selecting two components each for the ICP and SEC blocks, and seven components each for IR and NMR. The dataset was autoscaled before proceeding with MB-PLS modeling, the majority of the principal components were interpreted and assigned to specific chemical features based on the PCA analysis discussed in Section 4.

The biplot for the MB-PLS model consistent of two components explaining 31% of the variance in viscosity is shown in Figure 31. The viscosity response is primarily aligned with weight-average molecular mass, polydispersity, and recovery, indicating that molecular weight distribution is the main predictor of viscosity in this dataset. Variables from the SEC block dominate this axis, underscoring the importance of size-related features in the structural characterization of carrageenan. A component from the IR block, associated with carrageenan blockiness and corresponding to a spectral band around 1000 cm^{-1}, shows a negative correlation with viscosity. This suggests that increased blockiness may reduce solution viscosity, possibly by limiting molecular entanglement.

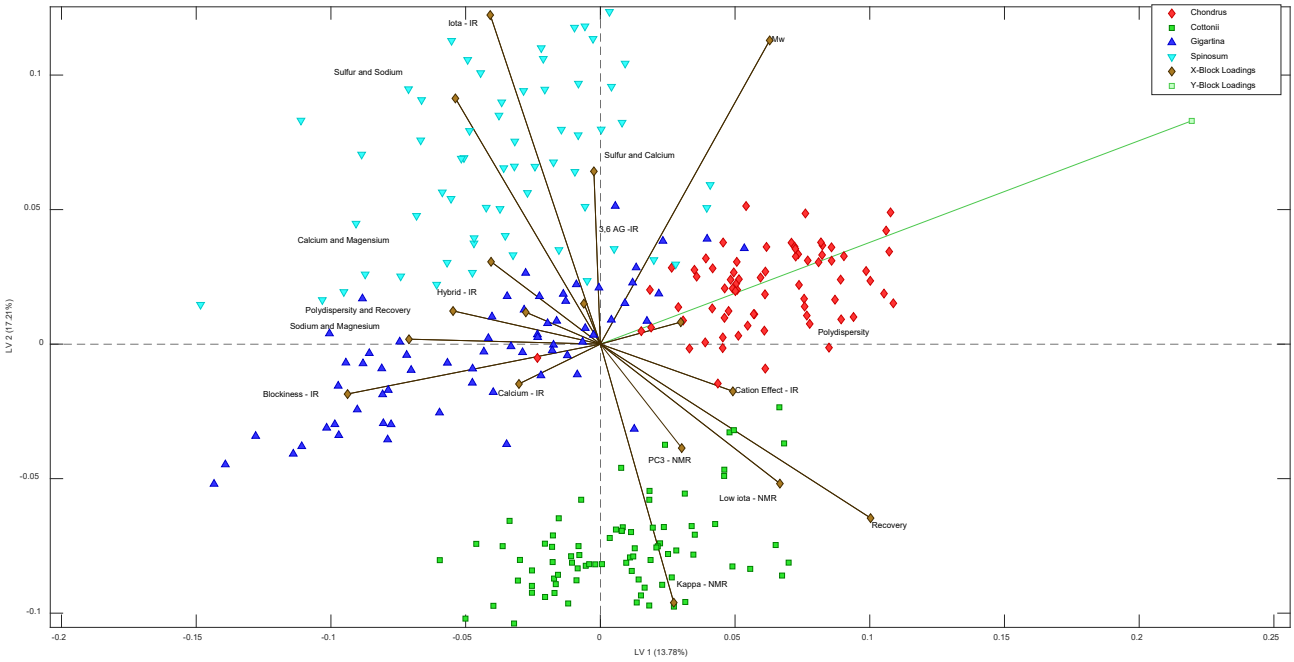


Figure 31 Biplot of the Multiblock-PLS model showing the relationship between structural features from the SEC, IR, NMR, and ICP data blocks and the viscosity response (Y-block loading).

In contrast, components derived from the NMR and ICP blocks appear nearly orthogonal to the viscosity response vector. This indicates that these variables capture structural or compositional information that is not directly predictive of viscosity.

The initial MB-PLS model, including all variables, yielded a relatively high RMSEP compared to RMSECV indication overfitting. To address this, interval partial least squares regression (iPLS) [159], was employed to investigate if variable reduction will explain the overfit and it did. Consistent with previous methodologies, the ICP block was excluded from the final model, and it improved RMSEP. This step was only done out of curiosity as the RMSEP value is not to be optimized as it solely to be used to gain realistic and robust estimation of prediction error.

5.1.5 Summary

Multiblock data fusion methods including, PO-PLS, SO-PLS, ROSA and MB-PLS, were used to investigate how carrageenan structure relates to its viscosity. The analysis included four data blocks: FT-IR, ^{1}H NMR, SEC-MALS and ICP-MS.

All blocks contained information about carrageenan type, but mostly by SEC-MALS structural features directly relevant for predicting viscosity were provided. This was confirmed by all multiblock methods. The most important predictors were weight-average intrinsic viscosity, RMS radius, and hydrodynamic radius.

One feature from the FT-IR data, identified by several methods is associated with glycosidic bond variation, presumably reflecting the distribution of κ- and ι-carrageenan units within molecules, appears to be relevant for viscosity, though not necessarily predictive.

5.2 UNDERSTANDING CARRAGEENAN-MILK GEL AND BREAKING STRENGTH

Multiblock data fusion analysis was also applied to understand the relationship between carrageenans' molecular structure and its gel and breaking strength when applied as a gelling agent in milk. As concluded in Paper 2, the use of the FT-IR spectra coupled with PLS was found to predict carrageenan-milk gel and breaking strength sufficiently well. However, merging the IR data with other data blocks demonstrated slightly better predictive capability. Specifically, for gel breaking strength, the combination of the SEC and IR blocks was optimal, while for the gel strength, the combination of the ICP and IR blocks was more predictive. This observation underscores the importance of investigating the specific contributions of information within these data blocks to the functional properties of carrageenans.

Below show SO-PLS models identified as optimal in Paper 2 but here developed using the complete dataset without partitioning it into calibration and validation sets. The aim is to make full use of the data for exploring and interpreting structure-function relationships comprehensively. Additionally, PO-PLS models were employed to further investigate the inter-block relationships and deepen the understanding of these functional relationships and value of adding more analytical platforms.

5.2.1 SO-PLS for carrageenan-milk gel and breaking strength

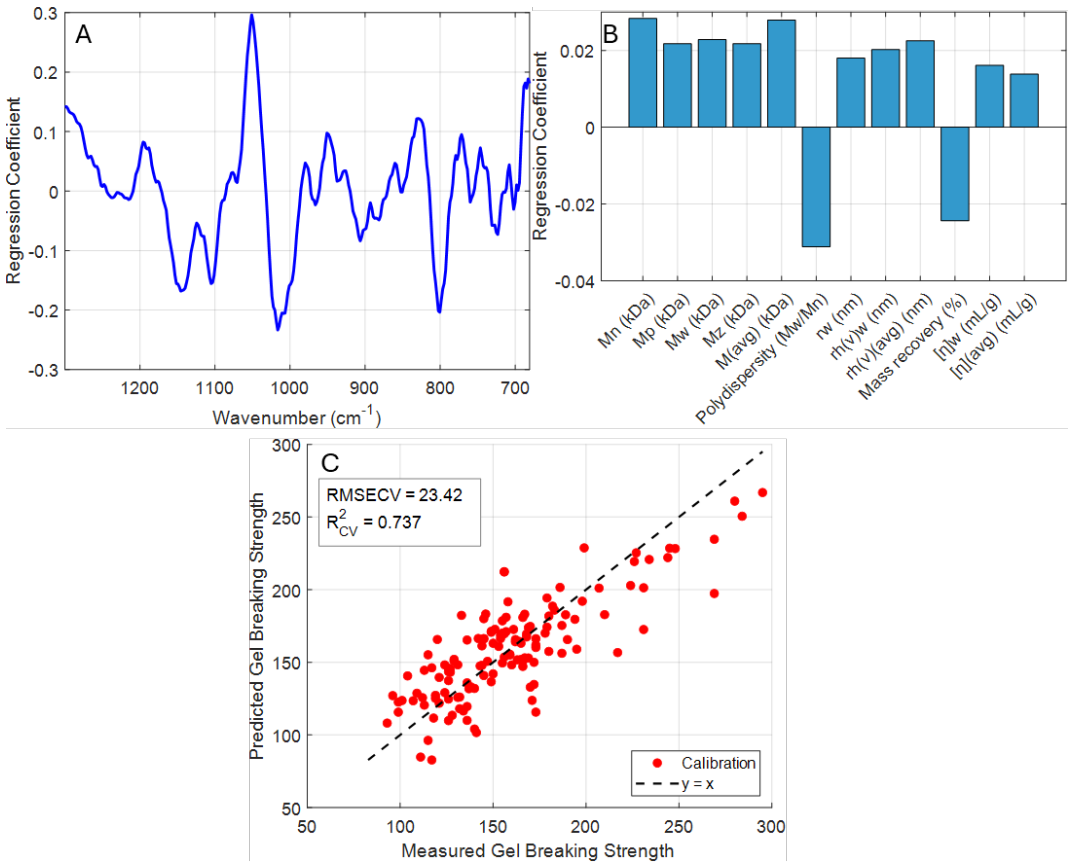


Figure 32 Results of SO-PLS model for prediction of carrageenan-milk gel breaking strength based on combination of the IR and SEC blocks. Panel A shows regression coefficients for the IR block. Panel B shows regression coefficients for the SEC block. Panel C shows cross-validated measured vs. predicted by SO-PLS carrageenan-milk gel breaking strength and model statistics.

The SO-PLS model developed for predicting gel breaking strength of carrageenan-milk gels included one component from the SEC block and five components from the IR block. The model achieved a moderate predictive performance with an RMSECV of 23.4 showed in Figure 32C. In the SEC block, Figure 32B, the regression coefficient plot indicates that the molecular weight distribution parameters exhibit the most significant positive association, suggesting the higher gel breaking strength is associated with higher molecular weight of carrageenan. In contrast, polydispersity presented a prominent negative coefficient, indicating that broader molecular weight distribution adversely relates to gel strength, likely due to the formation of less uniform gel networks. Additionally, mass recovery exhibited a negative association with gel breaking strength, which can be explained by lower hydrocolloid concentration which produces more rigid gels that are prone to fracturing under stress [21]. The IR block regression coefficient profile, in Figure 32A, reveals more complex structural associations. Notable positive contributions were observed at 1060 cm^{-1}, corresponding to C-O and C-OH stretching vibrations, while negative contributions were evident at 1030 cm^{-1}, associated with glycosidic bonds, and at 805 cm^{-1}, linked to C2-O-S in 3,6-anhydrogalactose, a structural motif characteristic of ι-carrageenan (Table 2). The negative coefficients suggest that the presence of ι-carrageenan decreases carrageenan-milk gel breaking strength, while κ-carrageenan, characterized by higher C-OH content due to none sulfate substitution at the C2 position in the DA unit (opposite to ι-carrageenan), contribute positively to gel breaking strength. The strongly negative coefficient at 1030 cm^{-1} may reflect differences in block distribution of the repeating units across carrageenan molecules.

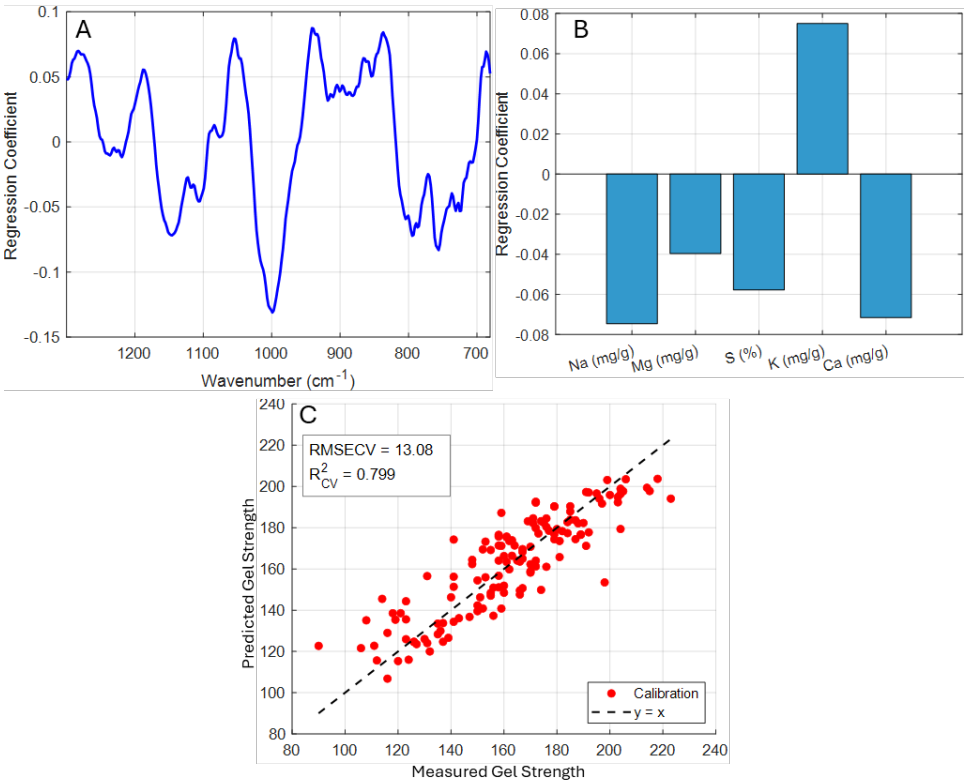


Figure 33 Results of SO-PLS model for prediction of carrageenan-milk gel strength based on combination of IR and ICP blocks. Panel A shows regression coefficients for IR block. Panel B shows regression coefficients for the ICP block. Panel C shows cross-validated measured vs. predicted by SO-PLS carrageenan-milk gel strength and model statistics.

Figure 33C, presents the measured versus predicted carrageenan-milk gel strength based on the SO-PLS model, comprising one component from the ICP block and three components from the IR block. The model achieved a moderate predictive performance with an RMSECV of 13 showed in Figure 33C. The regression coefficients in Figure 33B, for the ICP block, indicate that increasing potassium concentrations is associated with an increase in gel strength, while other cations exhibit a negative relation with it. While the effect of potassium ion is known to increase gel strength, the same would be expected from calcium [68], suggesting a more complex interaction in this system. The regression coefficient profile for the IR block Figure 33 A, reveals a dominant negative coefficient at 1000 cm^{-1}, and positive coefficients at 945 cm^{-1} and 1060 cm^{-1}. The negative contribution near 1000 cm^{-1} corresponds to glycosidic bond vibrations, which may reflect the distribution of κ- and ι-carrageenan units (see sections 2.1.1 and 4.2). The positive peak at 945 cm^{-1} is associated with the presence of 3,6-anhydrogalactose (typically observed near 925 cm^{-1}), a structural motif crucial for helix formation and gelation [29]. The positive coefficient at 1060 cm^{-1} is attributed to C-O and C-OH stretching, which are more prevalent in κ-carrageenan due to the absence of sulfate substitution at C2.

5.2.2 PO-PLS for milk-carrageenan gel strength

To explore which information was shared across blocks and which was distinct with respect to the prediction of carrageenan–milk gel and breaking strength, a PO-PLS model was developed. For the breaking strength, the model resulted in only the IR-block included in agreement with results in Paper 2. The PO-PLS modeling the gel strength resulted in the identification of two common components and one unique component. The first common component included information from all blocks, while the second did not include elemental data (ICP-block). The unique component was derived solely from the IR block. Overall, the PO-PLS model explained 80% of the variance in gel strength values, with 56.9% accounted for by the global component, 16.4% by the SEC-NMR component, and 6.7% by the IR-specific component with RMSECV of 12.

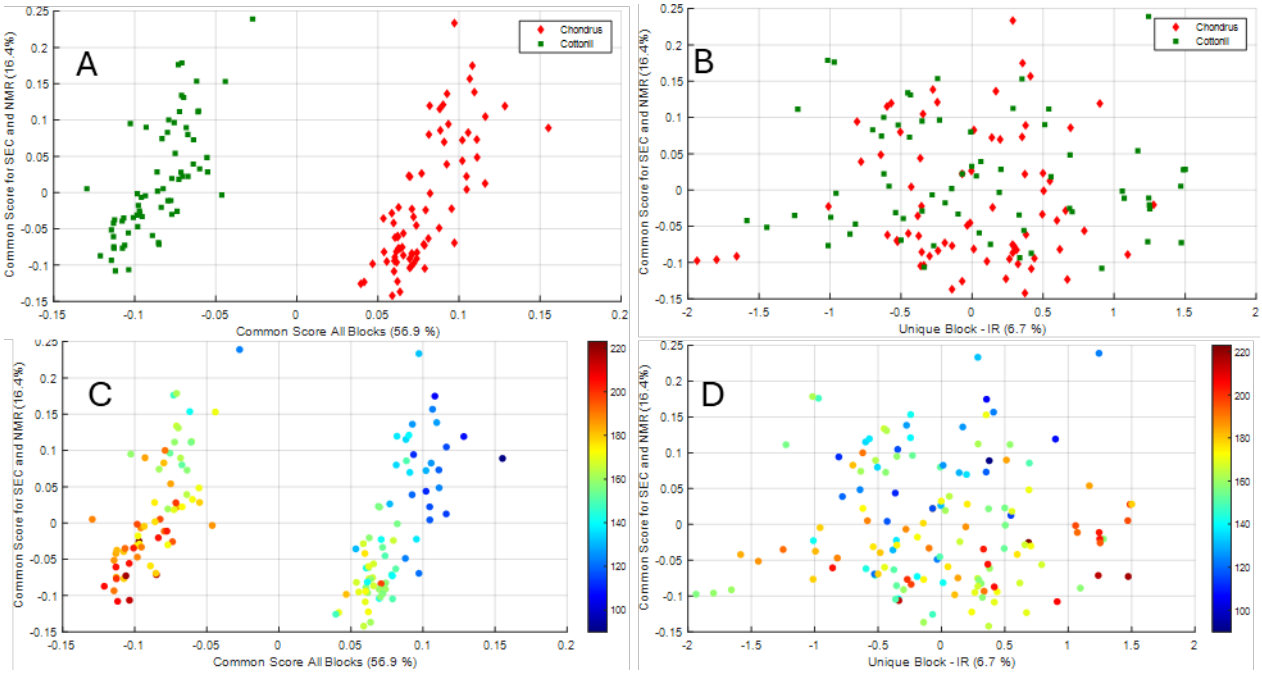


Figure 34 Score plots for the PO-PLS model for the common and unique components derived from the SEC, NMR, and IR blocks. Panels A and B show sample distribution based on carrageenan source (Chondrus vs. Cotoni), while panels C and D illustrate the color-coded gel strength, with warmer colors representing higher gel strength values.

The score plots in Figure 34 illustrate the separation of carrageenan samples based on the global component, clearly distinguishing between carrageenan derived from Cottonii (κ-carrageenan) and Chondrus (hybrid carrageenan). Since this separation is primarily driven by carrageenan type, the loading plots are not shown, as they mirror the previous results: hybrid carrageenans are associated with higher sulfate ester concentrations, whereas κ-carrageenans exhibit higher potassium content. When combined with the trends observed in the score plots (Figure 34, panel C), these findings suggest that lower concentrations of ι-carrageenan are associated with reduced gel strength, while higher concentrations of κ-carrageenan and potassium contribute to increased gel strength. These results are consistent with theoretical expectations (Section 2.1.2) and previous findings.

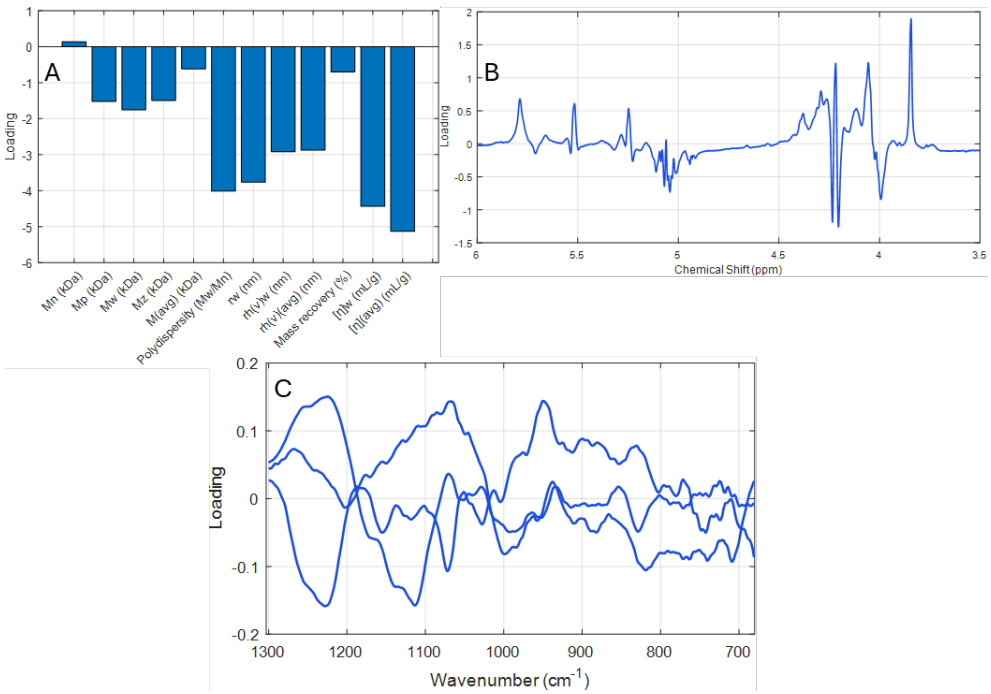


Figure 35 Loading plots for common component two of PO-PLS model including A) SEC-block; B) NMR-block; C) IR-block.

The second component was more challenging to interpret. As shown in Figure 35C, it exhibits strong loadings at 1225 cm^{-1} (O=S=O stretching), within the 1000-1100 cm^{-1} glycosidic bond region, and at 945 cm^{-1}, attributed to 3,6 anhydrogalactose. In addition to these FT-IR features, this component also shows notable contributions from intrinsic viscosity, polydispersity, and RMS radius in the SEC data (Panel A), as well as from the H6 signal of G4S (3.85 ppm) in κ-carrageenan observed in the ^{1}H NMR spectrum (Panel B). This component likely reflects the conformational state of the carrageenan molecules, as indicated by the high loadings for intrinsic viscosity (highly dependent on conformation) and 3,6-anhydrogalactose (necessary for coil-to-helix transition).

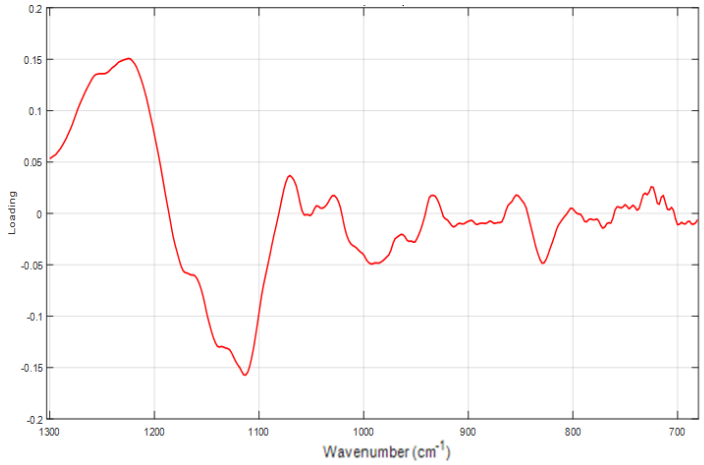


Figure 36 Loading plot for the IR block distinct component form PO-PLS model predicting carrageenan-milk gel strength.

The unique component derived from the IR block, as shown in the loading plot in Figure 36, closely resembles the loadings observed in Figure 21B from the PO-PLS model of carrageenan viscosity. In both cases, high loading values at 1230 cm^{-1} are associated with total carrageenan concentration. Additionally, the signal near 1100 cm^{-1} corresponding to glycosidic bonds might be due to differences in distribution of the repeating units within carrageenan molecule. This component captures variance in the FT-IR spectra that is uncorrelated with information in the SEC or NMR blocks. It reflects structural variation specific to the IR domain, likely related to heterogeneity in sulfate substitution or glycosidic linkage configuration, not accounted for by the shared components. The similarity to the analysis of aqueous carrageenan viscosity suggests that similar structural phenomena may underline both the gelling ability of carrageenan in milk systems and its thickening behavior in water.

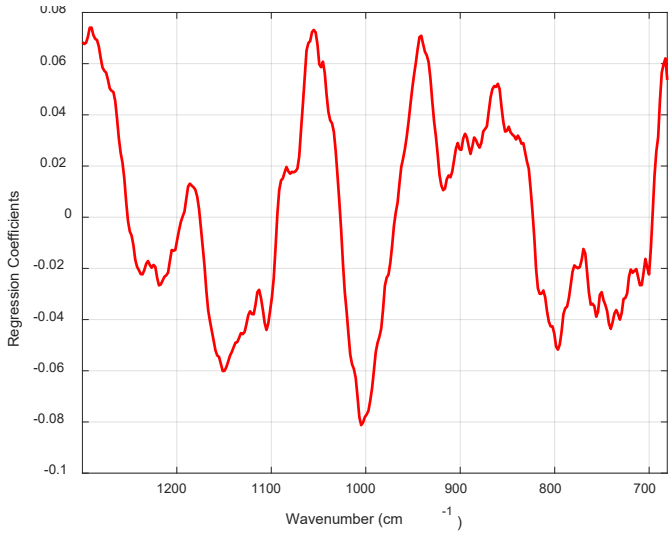


Figure 37 Regression coefficient for the IR-block in PO-PLS model for carrageenan-milk gel strength prediction.

Regression coefficients for the IR block in PO-PLS model are shown in Figure 37. These closely resemble the regression coefficients for the FT-IR block in the SO-PLS model for carrageenan–milk gel strength in Figure 33 A, with high negative values at 1000 cm^{-1}, and high positive values at 945

cm^{-1}, 1060 cm^{-1}. These bands have been previously associated with blockiness, 3,6-anhydrogalactose content, and κ-carrageenan concentration, respectively. The agreement between the PO-PLS and SO-PLS results strengthens confidence in the IR block's relevance for modeling carrageenan–milk gel properties. It also supports the conclusion in Paper 2 that FT-IR combined with PLS regression is suitable for predicting both gel strength and gel breaking point in carrageenan–milk systems.

5.2.3 Summary

The relation between carrageenan molecular structure and milk gel functionality was investigated using SO-PLS and PO-PLS models. For breaking strength, SO-PLS showed that higher molecular weight and lower polydispersity from SEC-MALS data correlated with stronger gels. Variations in carrageenan type showed that κ-carrageenan was linked with higher breaking strength, while ι-carrageenan was linked with lower values. Block distribution (blockiness) was relevant. Measurements of potassium the ICP-block along with spectral features related to 3,6-anhydrogalactose and glycosidic bond structure observed in FT-IR spectroscopy, were associated with increased gel strength.

Carrageenan type was consistently represented across all data blocks, revealed by a common PO-PLS component. The SEC block uniquely contributed intrinsic viscosity, while FT-IR provided unique information on 3,6-anhydrogalactose and blockiness.

Although adding more data blocks provided additional structural information, their inclusion did not improve the predictive accuracy beyond models based on FT-IR alone described in Paper 2.

5.3 FEASIBILITY OF SOLID-STATE NMR FOR SEAWEED CHARACTERIZATION

The current methodology for assessing the quality of seaweed for carrageenan extraction predominantly involves conducting the extraction process and subsequently evaluating the quality and yield of the extracted carrageenan. This approach primarily emphasizes the extraction performance and the suitability of the extraction method for a given seaweed type, rather than providing detailed insights into the chemical structure and composition of the seaweed itself. As a result, potential correlations between the carrageenan structure and the seaweed's carrageenan composition remain unexplored. To comprehensively understand the structure-function relationship in carrageenan, there is a clear need for a rapid, reliable analytical method capable of characterizing the chemical structure of seaweed, which is suitable for application in industrial settings.

Recent studies have indicated the potential applicability of solid-state NMR spectroscopy as a non-destructive analytical technique for evaluating seaweed and carrageenan compositions and estimating carrageenan concentration in seaweed samples [129,160,161].

In this study, high-resolution cross-polarization Magic Angle Spinning High Power Decoupling ^{13}C NMR spectroscopy (HR CP-MAS-HPD 13C NMR) was performed to obtain spectra from eight pairs of seaweed samples and their corresponding extracted carrageenan. The seaweed samples were dried, ground, and packed into 3.2 x 25 mm Si3N4 rotors and analyzed using an 850 MHz Bruker WB US2 spectrometer with MAS at 17.7 kHz. To further enhance resolution during proton evolution in fast spinning 2D ^{1}H - ^{13}C correlation spectroscopy, Frequency-switched Lee-Goldberg irradiation was applied [162].

Figure 38 (panel A and D) presents the CP-MAS ^{13}C NMR spectra for Spinosum seaweed and its derived carrageenan, as well as for Cottonii seaweed and its carrageenan extract. Notably, in the carrageenan samples, some distinct signals were observed, including a prominent signal at approximately 95 ppm corresponding to the C1 position in the DA2S unit of ι-carrageenan [20] However, the seaweed spectra exhibited significant signal overlap, with reduced resolution attributed to band broadening. This broadening effect can be ascribed to paramagnetic interactions caused by the presence of elements such as iron in seaweed matrices [163].

To estimate carrageenan concentration in seaweed, the approach outlined by Azevedo et al. (2022) [160], was applied. However, the estimated concentration was significantly overestimated, likely due to poor spectral resolution and the presence of overlapping signals from other seaweed constituents. To further explore the structural composition of seaweed, 2D 1H-13C spectra of Spinosum seaweed and its derived carrageenan (Figure 38, panels B and C), as well as Cottonii seaweed and its carrageenan extract (Figure 38, panels D and E), were obtained. A slight improvement in spectral resolution was observed, particularly in panel B, where additional structural information could be discerned.

Despite these improvements, the extended acquisition time of over 30 hours per measurement (27 hours for CP-MAS 13C NMR spectra alone) presents a considerable limitation, rendering this approach impractical for industrial applications where rapid analysis is essential. Consequently, further method optimization or alternative approaches are necessary to effectively characterize seaweed structure in a time-efficient manner suitable for industrial settings.

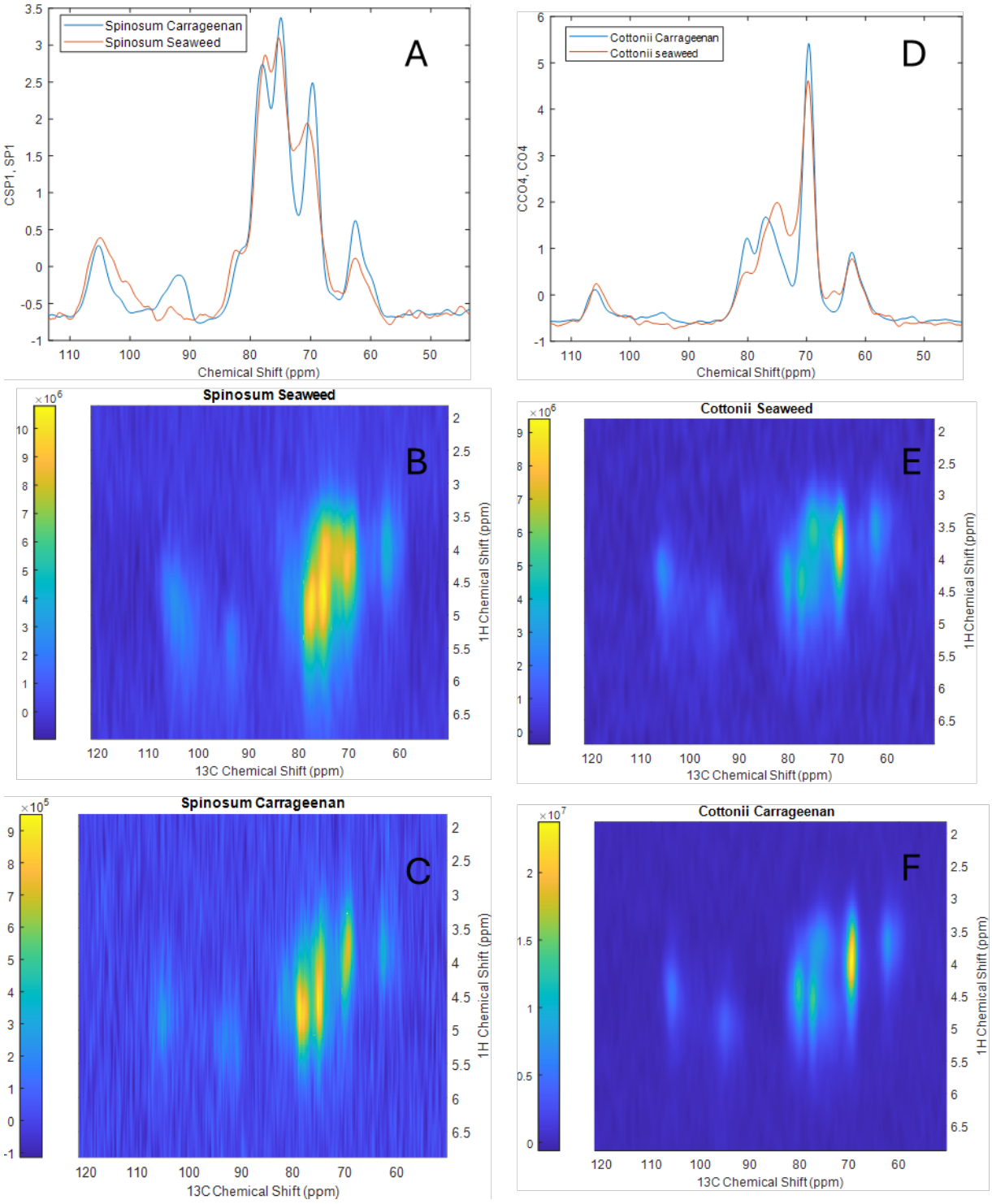


Figure 38 CP-MAS 13C NMR spectra of Spinosum seaweed and carrageenan extracted from it (panel A) and of Cottonii seaweed and carrageenan extracted from it. Panel B and C show ^{1}H-^{13}C NMR spectra for Spinosum seaweed and Spinosum-derived carrageenan, respectively. Panel E and F C show ^{1}H-^{13}C NMR spectra for Cottonii seaweed and Cottonii-derived carrageenan, respectively.

6 WISDOM: IMPLEMENTATION OF PREDICTIVE MODELS FOR CARRAGEENAN FUNCTIONALITY

6.1 IMPLEMENTATION OF PREDICTIVE MODELS FOR INDUSTRIAL QUALITY CONTROL OF CARRAGEENAN

Another hypothesis of this work was that predictive models developed for structure-functionality characterization of carrageenan could be employed in a quality control laboratory to predict the quality of carrageenan products. For both functionality tests, apparent viscosity and carrageenan-milk breaking and gel strength, predictive models were implemented using a commercially available prediction engine. For predicting apparent viscosity, the PLS models based on SEC-MALS data described in Paper 1 were employed. For predicting gel breaking strength, the PLS models based on IR spectra described in Paper 2 were implemented. The same implementation procedure was used for all models, and it will be explained here using the PLS models developed for predicting the gel strength.

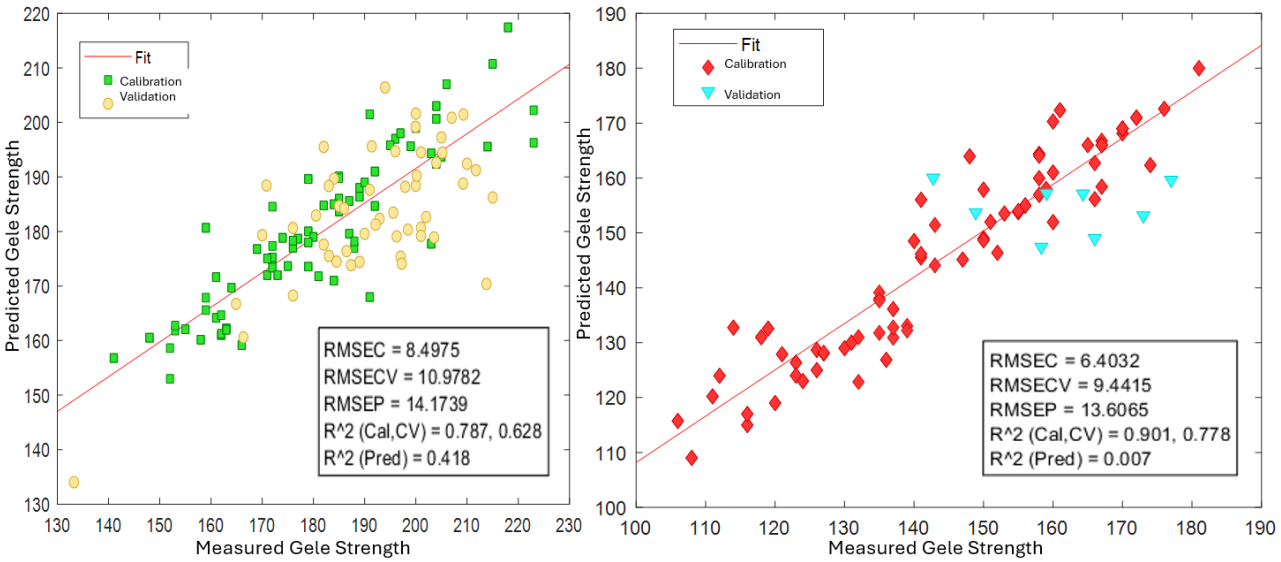


Figure 39 PLS regression models for predicting carrageenan-milk gel strength as implemented in an industrial quality control setting. The left panel shows the model for carrageenan extracted from Cottonii seaweed, while the right panel presents the model for carrageenan derived from Chondrus seaweed. Calibration samples include a subset of the samples from Paper 2, while validation samples represent carrageenan samples produced during six months of model testing

The predictive models were integrated into a prediction engine installed on the computer controlling the FT-IR instrument that collected spectral data. Over a six-month period, all milk-carrageenan gel strength measurements for products derived from Cottonii and Chondrus seaweed were conducted using FT-IR coupled to PLS and the original texture analyzer method. The results of this ongoing validation are presented in Figure 39, where the newly collected samples are labeled as validation samples. As the number of samples corresponds to real production dynamics, a perfect distribution across the model range is not expected, in contrast to a test set extracted from a complete dataset by a sampling algorithm.

In the case of Chondrus-derived carrageenan, only eight batches were produced during the validation period, resulting in a relatively clustered validation set (Figure 39, right panel). The RMSEP of 13.6 is considered acceptable given the limited number of samples and the constrained range.

For Cottonii-derived carrageenan, a larger number of batches were produced, providing a broader sample distribution (Figure 39, left panel). The resulting RMSEP of 14.2 is within acceptable limits; however, both RMSEP values are affected by the clustered nature of the validation samples. The final assessment of model performance will require additional samples covering the entire model range to ensure robust prediction accuracy. Nevertheless, the model can be utilized provided that its performance is deemed acceptable based on internal quality assessment protocols.

Despite a slightly better performance of Support Vector Machines model described in Paper 2 it was decided to implement the PLS regression due to higher explainability and thereby better options for maintenance and control.

A combination of two strategies was proposed for model maintenance. The first strategy is based on external validation samples, wherein both the reference method and the predictive model are applied periodically. The second strategy involves monitoring model residuals and leverage, which can be continuously evaluated using the same software that operates the predictive model, provided that appropriate action limits are established [164]. Additionally, consideration was given to the possibility of fusing the models to minimize maintenance time and streamline validation procedures [165].

6.2 Implementation of a Quality by Design Approach for Carrageenan Blends Development

Based on the results of work described in Paper 3, a predictive model for estimating the viscosity of carrageenan-based formulations based on ingredient composition was developed and implemented as a calculator tool. This tool was created using regression models that utilize the concentration of four key ingredients: carrageenan extracted from *Kappaphycus alvarezii*, *Chondrus crispus*, *Gigartina radula*, and sucrose. The model estimates viscosity using a reduced ANOVA model that includes main effects and selected two-way interactions, as detailed in Table 3, Paper 3.

The calculator tool allows users to input the percentages of each ingredient and provides a predicted viscosity value based on the established model. The calculation is performed using a linear regression equation that incorporates both individual ingredient concentrations and significant interaction terms. The structure of the regression equation is as follows:

$$Viscosity = I + \beta_1 A + \beta_2 B + \beta_3 C + \beta_4 D + \beta_5 (D \times B) + \beta_6 (D \times C) + \beta_7 (A \times B) + \beta_8 (A \times C) \quad \text{Eq.1}$$

Where:

- I - intercept
- β (1...i) – estimates
- A - *K. alvarezii concentration*
- *B - C. crispus concentration*
- *C – G. radula concentration*

- D – sugar concentration.

An example of calculation using ingredient concentrations provided in Table 5, Paper 3 is as follows:

$$\begin{aligned} Viscosity = {} & -1435 + 15.511 \times 5 + 15.679 \times 40 + 15.846 \times 26 + 14.377 \times 29 \\ & + (-0.026305) \times 29 \times 40 + (-0.036116) \times 29 \times 26 + (-0.031428) \times 5 \times 40 \\ & + (-0.027523) \times 5 \times 26 \end{aligned}$$

$$Viscosity \approx 31$$

The resulting estimated viscosity is approximately 31 A.U., which falls within the target viscosity range of 30 to 60 A.U., previously identified in Paper 3 as suitable for cocoa particle suspension in chocolate milk formulations.

The calculator tool was designed to provide a rapid, user-friendly method for optimizing ingredient ratios to achieve desired viscosity levels without extensive experimental testing. Additionally, the calculator can be easily integrated into existing software systems or experimental workflows to streamline the formulation process and reduce resource use. The model's applicability and accuracy have to be further assessed as new data is collected, and model updates can be implemented to incorporate additional interaction terms or quality metrics as required. Afterwords, the tool can be applied to define the design space for formulation of carrageenan blends.

7 CONCLUSION AND PERSPECTIVES

This Industrial PhD thesis demonstrated that chemometric modeling of carrageenan functionality, based on multiblock data fusion, can support the implementation of Process Analytical Technology and Quality by Design principles in carrageenan production. The research addressed three primary objectives: (i) comprehensive structural characterization of carrageenan, (ii) modeling of carrageenan functionality based on structural data, and (iii) implementation of predictive models for quality control. All objectives were addressed through experimental work on commercial carrageenan samples characterized by four analytical platforms—SEC-MALS, ICP-MS, FT-IR, and ^{1}H NMR—and in application-based functionality tests for viscosity (Paper 1), carrageenan-milk breaking and gel strength (Paper 2), and suspension stabilization (Paper 3). Chemometric tools were used to model these functionalities, beginning with PCA for data exploration, followed by multiblock data fusion and PLS regression to uncover structure–function relationships and develop predictive models.

For structural characterization, raw data analysis, PCA, and data fusion showed that information about carrageenan type is consistently present across all four analytical platforms. This was particularly evident in the spectral platforms—FT-IR and ^{1}H NMR—which both captured information on the content and positioning of sulfate esters, distinguishing κ- and ι-carrageenan. Cationic profiles obtained via ICP-MS also varied by carrageenan type, largely due to differences in manufacturing processes optimized to enhance functionality. Variation in the molecular mass and intrinsic viscosity among carrageenans was revealed by SEC-MALS, though separation based on origin was less distinct for this platform. Nevertheless, SEC-MALS made a unique contribution by providing direct information on molecular conformation in solutions, through intrinsic viscosity and hydrodynamic radius.

Proton NMR spectroscopy provided similar information to FT-IR, although with lower resolution and higher cost, as well as requiring more extensive sample preparation. From an industrial perspective, FT-IR spectroscopy offers a more practical and efficient option. Importantly, FT-IR consistently highlighted importance of spectral bands related to glycosidic bonds, which were relevant both for distinguishing structural differences among samples and for predicting functionality. In this work, these signals were linked to the distribution of disaccharide units of κ- and ι-carrageenan along the polysaccharide chain. Despite its importance, this remains a knowledge gap, as no rapid, industry-suitable method exists to measure blockiness directly. The current standard, enzymatic degradation of carrageenan into following by chromatography and mass spectrometry analysis [94], is labor-intensive and impractical for large sample sets.

Given the known impact of the distribution of disaccharide units on functionality [21], a promising direction would be to investigate a subset of samples using traditional blockiness assays and then measure the same samples by FT-IR. A chemometric model could then be developed to predict blockiness from FT-IR spectra. A similar approach was successfully applied to pectin using different instrumentation [166].

The second objective—modeling functionality—was addressed through interpretation of multiblock data fusion models and PLS regressions developed for apparent viscosity, milk–carrageenan gel strength, and breaking strength. The viscosity study showed that SEC-MALS parameters such as intrinsic viscosity, hydrodynamic radius, and molar mass were the most predictive. Increasing these values was consistently associated with increased viscosity (Paper 1 and Section 5.1).

For the breaking and gel strength, FT-IR spectroscopy alone provided comparable predictive performance to multiblock models and offered the highest interpretability. Carrageenan–milk breaking and gel strength correlated positively with κ-content, 3,6-anhydrogalactose, and potassium concentration. Relevant signals related to sulfate positioning and glycosidic bond types were captured in the FT-IR spectra. Blockiness—the distribution of κ- and ι-units—was a relevant as well. Parameters from SEC-MALS was weakly associated with to the gel strength and moderately with the breaking strength.

It must be emphasized that the structure–function relationships developed in this thesis are based on correlations between molecular structure parameters and functional properties for carrageenan. The analytical platforms applied here describe the structure of the final product but not the upstream raw material or intermediate processing states. To study causality, experimental intervention is necessary [167]. A key challenge in this project was identifying ways to intervene on carrageenan structure, since industrial production is designed to yield a consistent final product regardless of raw material variability. Therefore, a possible intervention would require measuring structural properties of the raw material, then manipulating them systematically to get new structures of carrageenan and observing the resulting functionality.

Although solid-state NMR was tested as a potential method for seaweed characterization, it proved unsuitable due to resolution limitations and extensive measurement time. As an alternative, a fast FT-IR method can be tested on a small set of seaweed samples [168]. It was tested on a small seaweed dataset (data not shown) but it was only modeled towards commercial carrageenan yield with very low variation in it. This approach may still be worth further investigation. Another strategy would be to vary carrageenan extraction conditions (e.g. time, temperature, alkali treatment) in a laboratory to produce structurally diverse carrageenans. Combining this with functionality testing, especially at the oligosaccharide level, could allow targeted analysis of how defined structural changes affect gel behavior.

Causal testing could also be conducted at the matrix level—for example, by varying milk composition or processing conditions and observing how this modifies carrageenan functionality. Such studies have been done previously, as noted in the state of the art (Section 2.2), and existing historical data may offer a resource for retrospective causal analysis.

The practical and measurable output of this project corresponds to the third objective: model implementation. Two predictive models were implemented to estimate carrageenan functionality. For prediction of milk–carrageenan gel and breaking strength, a Support Vector Machine trained on FT-IR spectra yielded the lowest prediction errors, as detailed in Paper 2. However, due to their high interpretability, FT-IR spectra based PLS regressions, also described in Paper 2, were implemented and partially validated in an industrial quality control workflow, demonstrating feasibility for routine use. Compared to traditional application tests, the method is significantly faster, more sustainable, and more cost-effective.

For carrageenan viscosity prediction, PLS regression was applied using polynomial expansions and interaction terms based on SEC-MALS data. The best-performing models were developed separately for carrageenan extracted each seaweed source. Although one my question the degree to which this method qualifies as PAT based method, it demonstrates the feasibility of instrument- and chemometrics based quality control monitoring. Real-time MALS instruments are now commercially available and could potentially be adapted for in-line monitoring of carrageenan

molecular size and weight distribution [169]. However, this would require known analyte concentration, which may be difficult to ensure in a production line. A combined approach using chemometrics could address this limitation and merits further investigation, starting at lab scale (as the company has already MALS instruments in the QC laboratory).

A formulation tool was also developed for carrageenan blends for chocolate milk using design of experiments and regression analysis, as described in Paper 3 and Section 6.2, to support blend optimization under raw material variability. However, among the tested formulation factors, total carrageenan concentration was the dominant determinant of viscosity, while the most interesting factor, the milk-carrageenan breaking strength, is less relevant. This finding limits the novelty of the model. It would be relevant to revisit the dataset and investigate whether additional quality parameters (from historical data) could explain differences in functional behavior. Furthermore, given the high structural informativeness of FT-IR, measuring spectra of the actual blends used in the study could offer new insights into structure–function relationships and help clarify the interactions identified as important in the statistical models. This represents a limitation of the current work: while some effects were statistically significant, their molecular mechanisms remain unclear.

This study marks a first step toward PAT transformation and offers a pathway to digitalization in the processing of natural hydrocolloids. It demonstrates that predictive models can be developed for the functionality of structurally complex polysaccharides like carrageenan. Its molecular complexity is a source of strength, enabling the broad functional versatility that has supported its industrial and culinary use for centuries.

BIBLIOGRAPHY

[1] M. Fricke, The Knowledge Pyramid: the DIKW Hierarchy, Knowledge Organization 46 (2019) 33–46. https://doi.org/10.5771/0943-7444-2019-1-33.

[2] CP Kelco, GENU ® Carrageenan Book, Kelco CP (2001) 4–23. http://foodsci.rutgers.edu/carbohydrates/Carrageenan Book.pdf.

[3] A. Chevenier, D. Jouanneau, E. Ficko-Blean, Carrageenan biosynthesis in red algae: A review, The Cell Surface 9 (2023) 100097. https://doi.org/10.1016/J.TCSW.2023.100097.

[4] M.E. Mitchell, M.D. Guiry, Carrageen: A local habitation or a name?, J Ethnopharmacol 9 (1983) 347–351. https://doi.org/10.1016/0378-8741(83)90043-0.

[5] J.R. Tanoeiro, D. Fortunato, J. Cotas, T. Morais, C. Afonso, L. Pereira, Different Chondrus crispus Aquaculture Methods and Carrageenan Extraction, Applied Sciences 13 (2023) 5466. https://doi.org/10.3390/app13095466.

[6] H.M. Smith, The utilization of seaweeds in the United States, in: Bulleting of the Bureau of Fisheries, 1905. https://spo.nmfs.noaa.gov/content/utilization-seaweeds-united-states (accessed July 13, 2024).

[7] R. Rupert, K.F. Rodrigues, V.Y. Thien, W.T.L. Yong, Carrageenan From Kappaphycus alvarezii (Rhodophyta, Solieriaceae): Metabolism, Structure, Production, and Application, Front Plant Sci 13 (2022) 859635–859635. https://doi.org/10.3389/fpls.2022.859635.

[8] R. Campbell, S. Hotchkiss, Carrageenan Industry Market Overview, in: A.Q. Hurtado, I.C. Neish, A.T. Critchley (Eds.), Tropical Seaweed Farming Trends, Problems and Opportunities, Springer International Publishing AG, Switzerland, 2017: pp. 193–205. https://doi.org/10.1007/978-3-319-63498-2_13.

[9] S. Genicot, A. Préchoux, G. Correc, N. Kervarec, G. Simon, J.S. Craigie, Carrageenans: New Tools for New Applications, in: S. La Barre, S. S. Bates (Eds.), Blue Biotechnology, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2018: pp. 371–416. https://doi.org/10.1002/9783527801718.ch12.

[10] K., A.K., Pulidinidi, Carrageenan Market Size, (n.d.). The global carrageenan market, valued at approximately USD 850 million in 2022, is projected to grow at a compound annual growth rate of over 6% through 2032, driven by increasing demand for natural and functional food ingredients. (accessed March 16, 2025).

[11] J.N. BeMiller, Chapter 13 in Carbohydrate chemistry for food scientists, in: Third edition., Woodhead Publishing, an imprint of Elsevier, Duxford, United Kingdom, 2019.

[12] S.H. Knutsen, D.E. Myslabodski, B. Larsen, A.I. Usov, A Modified System of Nomenclature for Red Algal Galactans, Botanica Marina 37 (1994) 163–170. https://doi.org/10.1515/botm.1994.37.2.163.

[13] T.T.T. Thành, Y. Yuguchi, M. Mimura, H. Yasunaga, R. Takano, H. Urakawa, K. Kajiwara, Molecular Characteristics and Gelling Properties of the Carrageenan Family, 1. Preparation of Novel Carrageenans and their Dilute Solution Properties, Macromol Chem Phys 203 (2002) 15–23. https://doi.org/10.1002/1521-3935(20020101)203:1<15::AID-MACP15>3.0.CO;2-1.

[14] T. Thanh, H. Yasunaga, R. Takano, H. Urakawa, K. Kajiwara, Molecular characteristics and gelling properties of carrageenan family 2. Tri-sulfated and tetra-sulfated carrageenans, Polymer Bulletin 47 (2001) 305–312. https://doi.org/10.1007/s289-001-8186-0.

[15] R. Tuvikene, Carrageenans, Handbook of Hydrocolloids (2021) 767–804. https://doi.org/10.1016/B978-0-12-820104-6.00006-1.

[16] O. Mykhalevych, H. Stapelfeldt, F. Marini, R. Bro, Chemometric insights into milk-carrageenan breaking and gel strength, Food Hydrocoll 158 (2025) 110544. https://doi.org/10.1016/J.FOODHYD.2024.110544.

[17] O. Mykhalevych, H. Stapelfeldt, R. Bro, Chemometric approaches for polysaccharide viscosity profiling, Carbohydr Polym 348 (2025) 122824. https://doi.org/10.1016/J.CARBPOL.2024.122824.

[18] M. Robal, T. Brenner, S. Matsukawa, H. Ogawa, K. Truus, B. Rudolph, R. Tuvikene, Monocationic salts of carrageenans: Preparation and physico-chemical properties, Food Hydrocoll 63 (2017) 656–667. https://doi.org/10.1016/j.foodhyd.2016.09.032.

[19] F. Van De Velde, A.S. Antipova, H.S. Rollema, T. V. Burova, N. V. Grinberg, L. Pereira, P.M. Gilsenan, R.H. Tromp, B. Rudolph, V.Y. Grinberg, The structure of κ/ι-hybrid carrageenans II. Coil–helix transition as a function of chain composition, Carbohydr Res 340 (2005) 1113–1129. https://doi.org/10.1016/J.CARRES.2005.02.015.

[20] F. Van De Velde, L. Pereira, H.S. Rollema, The revised NMR chemical shift data of carrageenans, Carbohydr Res 339 (2004) 2309–2313. https://doi.org/10.1016/J.CARRES.2004.07.015.

[21] F. van de Velde, Structure and function of hybrid carrageenans, Food Hydrocoll 22 (2008) 727–734. https://doi.org/10.1016/J.FOODHYD.2007.05.013.

[22] Food and Agriculture Organization of the United Nations (FAO), Carrageenan, in: Compendium of Food Additive Specifications – Addendum 6, FAO, Rome, 1998. https://www.fao.org/4/x5822e/x5822e05.htm (accessed March 4, 2025).

[23] L. Picullel, Gelling carrageenans, Food Polysaccharides and Their Applications (2006) 239–287. https://doi.org/10.1201/9781420015164-13.

[24] S. Podzimek, Light scattering, size exclusion chromatography, and asymmetric flow field flow fractionation powerful tools for the characterization of polymers, proteins, and nanoparticles, Wiley, Hoboken, NJ, 2011.

[25] C. Rochas, M. Rinaudo, S. Landry, Role of the molecular weight on the mechanical properties of kappa carrageenan gels, Carbohydr Polym 12 (1990) 255–266. https://doi.org/10.1016/0144-8617(90)90067-3.

[26] M. Marcotte, A.R.T. Hoshahili, H.S. Ramaswamy, Rheological properties of selected hydrocolloids as a function of concentration and temperature, Food Research International 34 (2001) 695–703. https://doi.org/10.1016/S0963-9969(01)00091-6.

[27] R. Tuvikene, K. Truus, A. Kollist, O. Volobujeva, E. Mellikov, T. Pehk, Gel-forming structures and stages of red algal galactans of different sulfation levels, in: Nineteenth International Seaweed Symposium, Springer Netherlands, Dordrecht, n.d.: pp. 77–85. https://doi.org/10.1007/978-1-4020-9619-8_11.

[28] G.A. De Ruiter, B. Rudolph, Carrageenan biotechnology, Trends Food Sci Technol 8 (1997) 389–395. https://doi.org/10.1016/S0924-2244(97)01091-1.

[29] V.L. Campo, D.F. Kawano, D.B. da Silva, I. Carvalho, Carrageenans: Biological properties, chemical modifications and structural analysis – A review, Carbohydr Polym 77 (2009) 167–180. https://doi.org/10.1016/J.CARBPOL.2009.01.020.

[30] F. Van De Velde, H.S. Rollema, N. V. Grinberg, T. V. Burova, V.Y. Grinberg, R. Hans Tromp, Coil-helix transition of ι-carrageenan as a function of chain regularity, Biopolymers 65 (2002) 299–312. https://doi.org/10.1002/bip.10250.

[31] V.T.N.T. Bui, B.T. Nguyen, T. Nicolai, F. Renou, Mixed iota and kappa carrageenan gels in the presence of both calcium and potassium ions, Carbohydr Polym 223 (2019) 115107. https://doi.org/10.1016/J.CARBPOL.2019.115107.

[32] D.D. Drohan, A. Tziboula, D. McNulty, D.S. Horne, Milk protein-carrageenan interactions, Food Hydrocoll 11 (1997) 101–107. https://doi.org/10.1016/S0268-005X(97)80016-1.

[33] V. Langendorff, G. Cuvelier, C. Michon, B. Launay, A. Parker, C.G. De Kruif, Effects of carrageenan type on the behaviour of carrageenan/milk mixtures, Food Hydrocoll 14 (2000) 273–280. https://doi.org/10.1016/S0268-005X(99)00064-8.

[34] D. Arltoft, R. Ipsen, F. Madsen, J. de Vries, Interactions between Carrageenans and Milk Proteins: A Microstructural and Rheological Study, Biomacromolecules 8 (2007) 729–736. https://doi.org/10.1021/bm061099q.

[35] V.J. Morris, G.R. Chilvers, Rheological studies of specific cation forms of kappa carrageenan gels, Carbohydr Polym 3 (1983) 129–141. https://doi.org/10.1016/0144-8617(83)90003-6.

[36] C. Rochas, M. Rinaudo, Activity coefficients of counterions and conformation in kappa-carrageenan systems, Biopolymers 19 (1980) 1675–1687. https://doi.org/10.1002/bip.1980.360190911.

[37] S.Y. XU, D.W. STANLEY, H.D. GOFF, V.J. DAVIDSON, M. LE MAGUER, Hydrocolloid/milk gel formation and properties, J Food Sci 57 (1992) 96–102. https://doi.org/10.1111/j.1365-2621.1992.tb05433.x.

[38] M.G. Lynch, D.M. Mulvihill, Rheology of ι-carrageenan gels containing caseins, Food Hydrocoll 10 (1996) 151–157. https://doi.org/10.1016/S0268-005X(96)80029-4.

[39] R.L.M. Tijssen, L.S. Canabady-Rochelle, M. Mellema, Gelation upon Long Storage of Milk Drinks with Carrageenan, J Dairy Sci 90 (2007) 2604–2611. https://doi.org/10.3168/jds.2006-854.

[40] G. Kalsi, U. Hazarika, L.D. Baruah, P.L. Bordoloi, M. Gogoi, Comprehensive review of carrageenan's multifaceted role in health and food systems, Discover Food 5 (2025) 1–15. https://doi.org/10.1007/s44187-025-00405-7.

[41] M.A. Ayadi, A. Kechaou, I. Makni, H. Attia, Influence of carrageenan addition on turkey meat sausages properties, J Food Eng 93 (2009) 278–283. https://doi.org/10.1016/J.JFOODENG.2009.01.033.

[42] D. Verbeken, N. Neirinck, P. Van Der Meeren, K. Dewettinck, Influence of κ-carrageenan on the thermal gelation of salt-soluble meat proteins, Meat Sci 70 (2005) 161–166. https://doi.org/10.1016/J.MEATSCI.2004.12.007.

[43] L. Marchetti, S.C. Andrés, A.N. Califano, Low-fat meat sausages with fish oil: Optimization of milk proteins and carrageenan contents using response surface methodology, Meat Sci 96 (2014) 1297–1303. https://doi.org/10.1016/J.MEATSCI.2013.11.004.

[44] C.M. Rosell, J.A. Rojas, C. Benedito de Barber, Influence of hydrocolloids on dough rheology and bread quality, Food Hydrocoll 15 (2001) 75–81. https://doi.org/10.1016/S0268-005X(00)00054-0.

[45] C. Onyango, S.K. Luvitaa, K. Lagat, L. K'osambo, Impact of carrageenan copolymers from two red seaweed varieties on dough and bread quality, J Appl Phycol 33 (2021) 3347–3356. https://doi.org/10.1007/s10811-021-02524-x.

[46] A. Kamińska-Dwórznicka, A. Janczewska-Dupczyk, A. Kot, S. Łaba, K. Samborska, The impact of ι- and κ-carrageenan addition on freezing process and ice crystals structure of strawberry sorbet frozen by various methods, J Food Sci 85 (2020) 50–56. https://doi.org/10.1111/1750-3841.14987.

[47] S. ADAPA, K.A. SCHMIDT, I.J. JEON, T.J. HERALD, R.A. FLORES, MECHANISMS OF ICE CRYSTALLIZATION AND RECRYSTALLIZATION IN ICE CREAM: A REVIEW, Food Reviews International 16 (2000) 259–271. https://doi.org/10.1081/FRI-100100289.

[48] B. Błaszak, G. Gozdecka, A. Shyichuk, Carrageenan as a functional additive in the production of cheese and cheese-like products, Acta Sci Pol Technol Aliment 17 (2018) 107–116. https://doi.org/10.17306/J.AFS.0550.

[49] M. Yanes, L. Durán, E. Costell, Rheological and optical properties of commercial chocolate milk beverages, J Food Eng 51 (2002) 229–234. https://doi.org/10.1016/S0260-8774(01)00061-9.

[50] H.-T.V. Lin, J.-S. Tsai, H.-H. Liao, W.-C. Sung, Effect of Hydrocolloids on Penetration Tests, Sensory Evaluation, and Syneresis of Milk Pudding, Polymers (Basel) 17 (2025) 300. https://doi.org/10.3390/polym17030300.

[51] M. Saluri, M. Robal, R. Tuvikene, Hybrid carrageenans as beer wort fining agents, Food Hydrocoll 86 (2019) 26–33. https://doi.org/10.1016/j.foodhyd.2017.12.020.

[52] M. Tippetts, S. Martini, Influence of ι-Carrageenan, Pectin, and Gelatin on the Physicochemical Properties and Stability of Milk Protein-Stabilized Emulsions, J Food Sci 77 (2012) C253–C260. https://doi.org/10.1111/j.1750-3841.2011.02576.x.

[53] P. Zarzycki, A.E. Ciołkowska, E. Jabłońska-Ryś, W. Gustaw, Rheological properties of milk-based desserts with the addition of oat gum and κ-carrageenan, J Food Sci Technol 56 (2019) 5107–5115. https://doi.org/10.1007/s13197-019-03983-4.

[54] X. Wang, D. Zhou, Q. Guo, C. Liu, Textural and structural properties of a κ-carrageenan–konjac gum mixed gel: effects of κ-carrageenan concentration, mixing ratio, sucrose and Ca2+ concentrations and its application in milk pudding, J Sci Food Agric 101 (2021) 3021–3029. https://doi.org/10.1002/jsfa.10936.

[55] G. Agoda-Tandjawa, C. Le Garnec, P. Boulenguer, M. Gilles, V. Langendorff, Rheological behavior of starch/carrageenan/milk proteins mixed systems: Role of each biopolymer type

and chemical characteristics, Food Hydrocoll 73 (2017) 300–312. https://doi.org/10.1016/J.FOODHYD.2017.07.012.

[56] D. Arltoft, F. Madsen, R. Ipsen, Relating the microstructure of pectin and carrageenan in dairy desserts to rheological and sensory characteristics, Food Hydrocoll 22 (2008) 660–673. https://doi.org/10.1016/j.foodhyd.2007.01.025.

[57] A. Naseri, S.L. Holdt, C. Jacobsen, Biochemical and Nutritional Composition of Industrial Red Seaweed Used in Carrageenan Production, Journal of Aquatic Food Product Technology 28 (2019) 967–973. https://doi.org/10.1080/10498850.2019.1664693.

[58] M. Ciancia, M.D. Noseda, M.C. Matulewicz, A.S. Cerezo, Alkali-modification of carrageenans: mechanism and kinetics in the kappa/iota-, mu/nu- and lambda-series, Carbohydr Polym 20 (1993) 95–98. https://doi.org/10.1016/0144-8617(93)90083-G.

[59] A.G. Viana, M.D. Noseda, M.E.R. Duarte, A.S. Cerezo, Alkali modification of carrageenans. Part V. The iota–nu hybrid carrageenan from Eucheuma denticulatum and its cyclization to iota-carrageenan, Carbohydr Polym 58 (2004) 455–460. https://doi.org/10.1016/J.CARBPOL.2004.08.006.

[60] F. Van De Velde, S.H. Knutsen, A.I. Usov, H.S. Rollema, A.S. Cerezo, 1H and 13C high resolution NMR spectroscopy of carrageenans: application in research and industry, Trends Food Sci Technol 13 (2002) 73–92. https://doi.org/10.1016/S0924-2244(02)00066-3.

[61] ICH, ICH Q8(R2): Pharmaceutical Development., (2009). https://database.ich.org/sites/default/files/Q8%28R2%29%20Guideline.pdf (accessed March 24, 2025).

[62] A.S. Rathore, G. Kapoor, Implementation of Quality by Design for processing of food products and biotherapeutics, Food and Bioproducts Processing 99 (2016) 231–243. https://doi.org/10.1016/J.FBP.2016.05.009.

[63] ICH, ICH Q12: Lifecycle Management., (2019). chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://database.ich.org/sites/default/files/Q12_Guideline_Step4_2019_1119.pdf (accessed March 24, 2025).

[64] FDA, Guidance for Industry: PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance., (2004). https://www.fda.gov/media/71012/download (accessed March 24, 2025).

[65] F. van den Berg, C.B. Lyndgaard, K.M. Sørensen, S.B. Engelsen, Process Analytical Technology in the food industry, Trends Food Sci Technol 31 (2013) 27–35. https://doi.org/10.1016/J.TIFS.2012.04.007.

[66] C.B. Zachariassen, J. Larsen, F. Van Den Berg, S.B. Engelsen, Use of NIR spectroscopy and chemometrics for on-line process monitoring of ammonia in Low Methoxylated Amidated pectin production, Chemometrics and Intelligent Laboratory Systems 76 (2005) 149–161. https://doi.org/10.1016/J.CHEMOLAB.2004.10.005.

[67] A.B. Rodd, C.R. Davis, D.E. Dunstan, B.A. Forrest, D. V. Boger, Rheological characterisation of ‘weak gel’ carrageenan stabilised milks, Food Hydrocoll 14 (2000) 445–454. https://doi.org/10.1016/S0268-005X(00)00024-2.

[68] F. Wang, W. Zhang, F. Ren, Effect of carrageenan addition on the rennet-induced gelation of skim milk: Corrigendum, J Sci Food Agric 97 (2017) 711–711. https://doi.org/10.1002/jsfa.8121.

[69] A. Puvanenthiran, S.J. Goddard, M.A. Augustin, Gelation of Mixed Gels Containing κ-Carrageenan and Skim Milk Components, J Food Sci 67 (2002) 573–577. https://doi.org/10.1111/j.1365-2621.2002.tb10640.x.

[70] M.A. Roberts, B. Quemener, Measurement of carrageenans in food: challenges, progress, and trends in analysis, Trends Food Sci Technol 10 (1999) 169–181. https://doi.org/10.1016/S0924-2244(99)00043-6.

[71] A.I. Usov, NMR spectroscopy of red seaweed polysaccharides: agars, carrageenans, and xylans, Botanica Marina 27 (1984) 189–202. https://doi.org/10.1515/botm.1984.27.5.189.

[72] M. Guibet, N. Kervarec, S. Génicot, Y. Chevolot, W. Helbert, Complete assignment of 1H and 13C NMR spectra of Gigartina skottsbergii λ-carrageenan using carrabiose oligosaccharides prepared by enzymatic hydrolysis, Carbohydr Res 341 (2006) 1859–1869. https://doi.org/10.1016/J.CARRES.2006.04.018.

[73] S.H. Knutsen, H. Grasdalen, Analysis of carrageenans by enzymic degradation, gel filtration and super(1)H NMR spectroscopy, Carbohydr Polym 19 (1992) 199–210.

[74] L. Pereira, F. Van De Velde, Portuguese carrageenophytes: Carrageenan composition and geographic distribution of eight species (Gigartinales, Rhodophyta), Carbohydr Polym 84 (2011) 614–623. https://doi.org/10.1016/J.CARBPOL.2010.12.036.

[75] H.N. Cheng, T.G. Neiss, Solution NMR Spectroscopy of Food Polysaccharides, Polymer Reviews 52 (2012) 81–114. https://doi.org/10.1080/15583724.2012.668154.

[76] M. Dyrby, R. V. Petersen, J. Larsen, B. Rudolf, L. Nørgaard, S.B. Engelsen, Towards on-line monitoring of the composition of commercial carrageenan powders, Carbohydr Polym 57 (2004) 337–348. https://doi.org/10.1016/J.CARBPOL.2004.05.015.

[77] E. Tojo, J. Prado, A simple 1H NMR method for the quantification of carrageenans in blends, Carbohydr Polym 53 (2003) 325–329. https://doi.org/10.1016/S0144-8617(03)00080-8.

[78] M. Sekkal, P. Legrand, A spectroscopic investigation of the carrageenans and agar in the 1500-100 cm−1 spectral range, Spectrochim Acta A 49 (1993) 209–221. https://doi.org/10.1016/0584-8539(93)80176-B.

[79] M. Şen, E.N. Erboz, Determination of critical gelation conditions of κ-carrageenan by viscosimetric and FT-IR analyses, Food Research International 43 (2010) 1361–1364. https://doi.org/10.1016/j.foodres.2010.03.021.

[80] J. Prado-Fernández, J.A. Rodríguez-Vázquez, E. Tojo, J.M. Andrade, Quantitation of κ-, ι- and λ-carrageenans by mid-infrared spectroscopy and PLS regression, Anal Chim Acta 480 (2003) 23–37. https://doi.org/10.1016/S0003-2670(02)01592-1.

[81] C. Rochas, M. Lahaye, W. Yaphe, Sulfate Content of Carrageenan and Agar Determined by Infrared Spectroscopy, Botanica Marina 29 (1986) 335–340. https://doi.org/10.1515/botm.1986.29.4.335.

[82] P.S. Belton, R.H. Wilson, D.H. Chenery, Interaction of group I cations with iota and kappa carrageenans studied by Fourier transform infrared spectroscopy, Int J Biol Macromol 8 (1986) 247–251. https://doi.org/10.1016/0141-8130(86)90035-8.

[83] L. Pereira, A. Sousa, H. Coelho, A.M. Amado, P.J.A. Ribeiro-Claro, Use of FTIR, FT-Raman and 13C-NMR spectroscopy for identification of some seaweed phycocolloids, Biomol Eng 20 (2003) 223–228. https://doi.org/10.1016/S1389-0344(03)00058-3.

[84] E. Tojo, J. Prado, Chemical composition of carrageenan blends determined by IR spectroscopy combined with a PLS multivariate calibration method, Carbohydr Res 338 (2003) 1309–1312. https://doi.org/10.1016/S0008-6215(03)00144-7.

[85] P. Volery, R. Besson, C. Schaffer-Lequart, Characterization of Commercial Carrageenans by Fourier Transform Infrared Spectroscopy Using Single-Reflection Attenuated Total Reflection, J Agric Food Chem 52 (2004) 7457–7463. https://doi.org/10.1021/jf040229o.

[86] D. Slootmaekers, J.A.P.P. van Dijk, F.A. Varkevisser, C.J.B. van Treslong, H. Reynaers, Molecular characterisation of κ- and λ-carrageenan by gel permeation chromatography, light scattering, sedimentation analysis and osmometry, Biophys Chem 41 (1991) 51–59. https://doi.org/10.1016/0301-4622(91)87209-N.

[87] A. Michna, W. Płaziński, D. Lupa, M. Wasilewska, Z. Adamczyk, Carrageenan molecule conformations and electrokinetic properties in electrolyte solutions: Modeling and experimental measurements, Food Hydrocoll 121 (2021) 107033. https://doi.org/10.1016/J.FOODHYD.2021.107033.

[88] A comprehensive study on the size exclusion chromatography of kappa-carrageenan for the identification of after-peaks, J Appl Polym Sci 127 (2013) 494.

[89] G. Marcelo, E. Saiz, M.P. Tarazona, Unperturbed dimensions of Carrageenans in different salt solutions, Biophys Chem 113 (2005) 201–208. https://doi.org/10.1016/J.BPC.2004.09.005.

[90] W. Kozicki, P.Q. Kuang, An alternative method for evaluation of intrinsic viscosity, Canadian Journal of Chemical Engineering 74 (1996) 429–432. https://doi.org/10.1002/cjce.5450740317.

[91] M. Ciancia, M.C. Matulewicz, R. Tuvikene, Structural Diversity in Galactans From Red Seaweeds and Its Influence on Rheological Properties, Front Plant Sci 11 (2020) 559986–559986. https://doi.org/10.3389/fpls.2020.559986.

[92] T. Brenner, R. Tuvikene, A. Parker, S. Matsukawa, K. Nishinari, Rheology and structure of mixed kappa-carrageenan/iota-carrageenan gels, Food Hydrocoll 39 (2014) 272–279. https://doi.org/10.1016/j.foodhyd.2014.01.024.

[93] R. Azevedo, A.R. Oliveira, A. Almeida, L.R. Gomes, Determination by ICP-MS of Essential and Toxic Trace Elements in Gums and Carrageenans Used as Food Additives Commercially Available in the Portuguese Market, Foods 12 (2023) 1408. https://doi.org/10.3390/foods12071408.

[94] M. Guibet, P. Boulenguer, J. Mazoyer, N. Kervarec, A. Antonopoulos, M. Lafosse, W. Helbert, Composition and Distribution of Carrabiose Moieties in Hybrid κ-/ι-Carrageenans Using Carrageenases, Biomacromolecules 9 (2008) 408–415. https://doi.org/10.1021/bm701109r.

[95] C. Rochas, M. Rinaudo, S. Landry, Relation between the molecular structure and mechanical properties of carrageenan gels, Carbohydr Polym 10 (1989) 115–127. https://doi.org/10.1016/0144-8617(89)90061-1.

[96] L.C. Geonzon, T. Enoki, S. Humayun, R. Tuvikene, S. Matsukawa, K. Mayumi, Understanding the rheological properties from linear to nonlinear regimes and spatiotemporal structure of mixed kappa and reduced molecular weight lambda carrageenan gels, Food Hydrocoll 150 (2024) 109752. https://doi.org/10.1016/J.FOODHYD.2024.109752.

[97] M.A.F. Monteiro, B. Faria, I.C.F. Moraes, L. Hilliou, Hybrid Carrageenans Versus Kappa-Iota-Carrageenan Blends: A Comparative Study of Hydrogel Elastic Properties, Gels 11 (2025) 157. https://doi.org/10.3390/gels11030157.

[98] L. Hilliou, Hybrid Carrageenans: Isolation, Chemical Structure, and Gel Properties, Adv Food Nutr Res 72 (2014) 17–43. https://doi.org/10.1016/B978-0-12-800269-8.00002-6.

[99] F. Van De Velde, H.A. Peppelman, H.S. Rollema, R.H. Tromp, On the structure of κ/ι-hybrid carrageenans, Carbohydr Res 331 (2001) 271–283. https://doi.org/10.1016/S0008-6215(01)00054-4.

[100] J.F. Dahl, M. Schlangen, A. Jan van der Goot, M. Corredig, Predicting rheological parameters of food biopolymer mixtures using machine learning, Food Hydrocoll 160 (2025) 110786. https://doi.org/10.1016/J.FOODHYD.2024.110786.

[101] A.D. Anderson, C.R. Daubert, B.E. Farkas, Rheological Characterization of Skim Milk Stabilized with Carrageenan at High Temperatures, J Food Sci 67 (2002) 649–652. https://doi.org/10.1111/j.1365-2621.2002.tb10654.x.

[102] F.R. Reis, A.B. de Pereira-Netto, J.L.M. Silveira, C.W.I. Haminiuk, L.M.B. Cândido, Apparent Viscosity of a Skim Milk Based Dessert: Optimization Through Response Surface Methodology, Food Nutr Sci 2 (2011) 90–95. https://doi.org/10.4236/fns.2011.22012.

[103] H.J. Bixler, K. Johndro, R. Falshaw, Kappa-2 carrageenan: structure and performance of commercial extracts: II. Performance in two simulated dairy applications, Food Hydrocoll 15 (2001) 619–630. https://doi.org/10.1016/S0268-005X(01)00047-9.

[104] R. Falshaw, H.J. Bixler, K. Johndrob, Structure and performance of commercial kappa-2 carrageenan extracts: I. Structure analysis, Food Hydrocoll 15 (2001) 441–452. https://doi.org/10.1016/S0268-005X(01)00066-2.

[105] R. Falshaw, H.J. Bixler, K. Johndro, Structure and performance of commercial κ-2 carrageenan extracts. Part III. Structure analysis and performance in two dairy applications of extracts from the New Zealand red seaweed, Gigartina atropurpurea, Food Hydrocoll 17 (2003) 129–139. https://doi.org/10.1016/S0268-005X(02)00045-0.

[106] A. Biancolillo, R. Boqué, M. Cocchi, F. Marini, Data Fusion Strategies in Food Analysis, Data Handling in Science and Technology 31 (2019) 271–310. https://doi.org/10.1016/B978-0-444-63984-4.00010-7.

[107] Y. Hong, N. Birse, B. Quinn, Y. Li, W. Jia, P. McCarron, D. Wu, G.R. da Silva, L. Vanhaecke, S. van Ruth, C.T. Elliott, Data fusion and multivariate analysis for food authenticity analysis, Nat Commun 14 (2023) 3309–3309. https://doi.org/10.1038/s41467-023-38382-z.

[108] E. Borràs, J. Ferré, R. Boqué, M. Mestres, L. Aceña, O. Busto, Data fusion methodologies for food and beverage authentication and quality assessment - A review, Anal Chim Acta 891 (2015) 1–14. https://doi.org/10.1016/j.aca.2015.04.042.

[109] E. Hayes, D. Greene, C. O'Donnell, N. O'Shea, M.A. Fenelon, Spectroscopic technologies and data fusion: Applications for the dairy industry, Frontiers in Nutrition (Lausanne) 9 (2023) 1074688–1074688. https://doi.org/10.3389/fnut.2022.1074688.

[110] M. Henningsson, K. Östergren, R. Sundberg, P. Dejmek, Sensor fusion as a tool to monitor dynamic dairy processes, J Food Eng 76 (2006) 154–162. https://doi.org/10.1016/j.jfoodeng.2005.05.003.

[111] J. Raeber, C. Steuer, Exploring new dimensions: Single and multi-block analysis of essential oils using DBDI-MS and FT-IR for enhanced authenticity control, Anal Chim Acta 1277 (2023) 341657. https://doi.org/10.1016/J.ACA.2023.341657.

[112] National Research Council, Carrageenan Monograph, in: Food Chemicals Codex, Third Edition, Washington, DC: The National Academies Press., 1981: pp. 73–75.

[113] Bourne Malcolm C, Viscosity Measurement, in: Food Texture and Viscosity - Concept and Measurement, 2nd Edition, Elsevier, 2002: pp. 1–1.

[114] A. Puvanenthiran, S.J. Goddard, M.A. Augustin, Milk-carrageenan gels, Australian Journal of Dairy Technology 55 (2000) 111.

[115] D.C. Montgomery, Design and analysis of experiments, Eighth edition., John Wiley & Sons, Inc., Hoboken, NJ, 2013.

[116] P.R. Griffiths, Fourier transform infrared spectrometry, 2nd ed., Wiley-Interscience, Hoboken, N.J, 2007.

[117] R.H. Wilson, B.J. Goodfellow, P.S. Belton, Fourier transform infrared spectroscopy for the study of food biopolymers, Food Hydrocoll 2 (1988) 169–178. https://doi.org/10.1016/S0268-005X(88)80015-8.

[118] James. Keeler, Understanding NMR spectroscopy, 2. ed., John Wiley and Sons, Chichester, 2010.

[119] T. Skov, F. van den Berg, G. Tomasi, R. Bro, Automated alignment of chromatographic data, J Chemom 20 (2006) 484–497. https://doi.org/10.1002/cem.1031.

[120] R.A. Rodríguez Sánchez, M.C. Matulewicz, M. Ciancia, NMR spectroscopy for structural elucidation of sulfated polysaccharides from red seaweeds, Int J Biol Macromol 199 (2022) 386–400. https://doi.org/10.1016/J.IJBIOMAC.2021.12.080.

[121] D. Jouanneau, M. Guibet, P. Boulenguer, J. Mazoyer, M. Smietana, W. Helbert, New insights into the structure of hybrid κ-/μ-carrageenan and its alkaline conversion, Food Hydrocoll 24 (2010) 452–461. https://doi.org/10.1016/J.FOODHYD.2009.11.012.

[122] R. Thomas, Practical Guide to ICP-MS and Other Atomic Spectroscopy Techniques : A Tutorial for Beginners, CRC Press, Boca Raton, Florida, 2023.

[123] S. Podzimek, Molar mass distribution by size exclusion chromatography: Comparison of multi-angle light scattering and universal calibration, J Appl Polym Sci 136 (2019) 47561-n/a. https://doi.org/10.1002/app.47561.

[124] S. Podzimek, Multi-Angle Light Scattering: An Efficient Tool Revealing Molecular Structure of Synthetic Polymers, Macromol Symp 384 (2019) 1800174-n/a. https://doi.org/10.1002/masy.201800174.

[125] S. Podzimek, Truths and myths about the determination of molar mass distribution of synthetic and natural polymers by size exclusion chromatography, J Appl Polym Sci 131 (2014) n/a. https://doi.org/10.1002/app.40111.

[126] J.B. Matson, A.Q. Steele, J.D. Mase, M.D. Schulz, Polymer characterization by size-exclusion chromatography with multi-angle light scattering (SEC-MALS): a tutorial review, Polym Chem 15 (2024) 127–142. https://doi.org/10.1039/d3py01181j.

[127] T. Hjerde, O. Smidsrød, B.E. Christensen, Analysis of the conformational properties of κ- and ι-carrageenan by size-exclusion chromatography combined with low-angle laser light scattering, Biopolymers 49 (1999) 71–80. https://doi.org/10.1002/(SICI)1097-0282(199901)49:1<71::AID-BIP7>3.0.CO;2-H.

[128] D. Lecacheux, R. Panaras, G. Brigand, G. Martin, Molecular weight distribution of carrageenans by size exclusion chromatography and low angle laser light scattering, Carbohydr Polym 5 (1985) 423–440. https://doi.org/10.1016/0144-8617(85)90003-7.

[129] L. Hilliou, I.C. Freitas Moraes, P.L. Almeida, From the seaweeds' carrageenan composition to the hybrid carrageenans' hydrogel elasticity: Identification of a relationship based on the content in iota-carrageenan, Food Hydrocoll 162 (2025) 111007. https://doi.org/10.1016/J.FOODHYD.2024.111007.

[130] R. Bro, A.K. Smilde, Principal component analysis, Analytical Methods 6 (2014) 2812–2831. https://doi.org/10.1039/C3AY41907J.

[131] P. Mishra, A. Biancolillo, J.M. Roger, F. Marini, D.N. Rutledge, New data preprocessing trends based on ensemble of multiple preprocessing techniques, TrAC - Trends in Analytical Chemistry 132 (2020) 116045. https://doi.org/10.1016/j.trac.2020.116045.

[132] A. Savitzky, M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Analytical Chemistry (Washington) 36 (1964) 1627–1639. https://doi.org/10.1021/ac60214a047.

[133] Å. Rinnan, F. van den Berg, S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, TrAC Trends in Analytical Chemistry 28 (2009) 1201–1222. https://doi.org/10.1016/J.TRAC.2009.07.007.

[134] N.-P.V. Nielsen, J.M. Carstensen, J. Smedsgaard, Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping, J Chromatogr A 805 (1998) 17–35. https://doi.org/10.1016/S0021-9673(98)00021-1.

[135] Eigenvector Research Inc., PLS_Toolbox, (2023). http://www.eigenvector.com. (accessed April 9, 2025).

[136] A.K. Smilde, I. Måge, T. Næs, T. Hankemeier, M.A. Lips, H.A.L. Kiers, E. Acar, R. Bro, Common and distinct components in data fusion, J Chemom 31 (2017) n/a. https://doi.org/10.1002/cem.2900.

[137] P. Mishra, J.M. Roger, D. Jouan-Rimbaud-Bouveresse, A. Biancolillo, F. Marini, A. Nordon, D.N. Rutledge, Recent trends in multi-block data analysis in chemometrics for multi-source

data integration, TrAC Trends in Analytical Chemistry 137 (2021) 116206. https://doi.org/10.1016/J.TRAC.2021.116206.

[138] A.K. Smilde, Multiblock data fusion in statistics and machine learning : applications in the natural and life sciences, John Wiley & Sons, Inc., Hoboken, New Jersey, 2022.

[139] A. Smolinska, J. Engel, E. Szymanska, L. Buydens, L. Blanchet, General Framing of Low-, Mid-, and High-Level Data Fusion With Examples in the Life Sciences, Data Handling in Science and Technology 31 (2019) 51–79. https://doi.org/10.1016/B978-0-444-63984-4.00003-X.

[140] L. Huang, J. Zhao, Q. Chen, Y. Zhang, Nondestructive measurement of total volatile basic nitrogen (TVB-N) in pork meat by integrating near infrared spectroscopy, computer vision and electronic nose techniques, Food Chem 145 (2014) 228–236. https://doi.org/10.1016/J.FOODCHEM.2013.06.073.

[141] R.W. Kennard, L.A. Stone, Computer Aided Design of Experiments, Technometrics 11 (1969) 137–148. https://doi.org/10.1080/00401706.1969.10490666.

[142] Eigenvector Research Inc., Eigenvector Research Documentation Wiki : Using Cross-Validation, (2023). https://www.wiki.eigenvector.com/index.php?title=Eigenvector_Research_Documentation_Wiki:General_disclaimer (accessed March 17, 2024).

[143] A. Biancolillo, K.H. Liland, I. Måge, T. Næs, R. Bro, Variable selection in multi-block regression, Chemometrics and Intelligent Laboratory Systems 156 (2016) 89–101. https://doi.org/10.1016/J.CHEMOLAB.2016.05.016.

[144] A. Biancolillo, F. Marini, J. Roger, SO-CovSel: A novel method for variable selection in a multiblock framework, J Chemom 34 (2020) n/a. https://doi.org/10.1002/cem.3120.

[145] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: A basic tool of chemometrics, Chemometrics and Intelligent Laboratory Systems 58 (2001) 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1.

[146] I. Måge, A.K. Smilde, F.M. Kloet, Performance of methods that separate common and distinct variation in multiple data blocks, J Chemom 33 (2019) n/a. https://doi.org/10.1002/cem.3085.

[147] B. Thompson, Canonical correlation analysis : uses and interpretation, Sage Publications, Beverly Hills, 1984.

[148] I. Måge, B.-H. Mevik, T. Næs, Regression models with process variables and parallel blocks of raw material measurements, J Chemom 22 (2008) 443–456. https://doi.org/10.1002/cem.1169.

[149] Ingrid Måge, Parallel Orthogonalizaed PLS, (n.d.). https://nofimadatamodeling.wordpress.com/software-downloads-list/ (accessed March 7, 2025).

[150] I. Måge, E. Menichelli, T. Næs, Preference mapping by PO-PLS: Separating common and unique information in several data blocks, Food Qual Prefer 24 (2012) 8–16. https://doi.org/10.1016/J.FOODQUAL.2011.08.003.

[151] V. Spichtig, S. Austin, Determination of the low molecular weight fraction of food-grade carrageenans, Journal of Chromatography B 861 (2008) 81–87. https://doi.org/10.1016/J.JCHROMB.2007.11.012.

[152] A.K. Smilde, J.A. Westerhuis, S. De Jong, A framework for sequential multiblock component methods, J Chemom 17 (2003) 323–337. https://doi.org/10.1002/cem.811.

[153] A. Biancolillo, T. Næs, The Sequential and Orthogonalized PLS Regression for Multiblock Regression: Theory, Examples, and Extensions, Data Handling in Science and Technology 31 (2019) 157–177. https://doi.org/10.1016/B978-0-444-63984-4.00006-5.

[154] T. Næs, R. Romano, O. Tomic, I. Måge, A. Smilde, K.H. Liland, Sequential and orthogonalized PLS (SO-PLS) regression for path analysis: Order of blocks and relations between effects, J Chemom 35 (2021) n/a. https://doi.org/10.1002/cem.3243.

[155] M. P. Campos, R. Sousa, M. S. Reis, Establishing the optimal blocks' order in SO-PLS: Stepwise SO-PLS and alternative formulations, J Chemom 32 (2018) n/a. https://doi.org/10.1002/cem.3032.

[156] F. Marini, Sequential and orthogonalized Partial Least Squares (SO-PLS), (2023). https://www.chem.uniroma1.it/romechemometrics/research/algorithms/so-pls/ (accessed March 5, 2024).

[157] K.H. Liland, T. Næs, U.G. Indahl, ROSA-a fast extension of partial least squares regression for multiblock data analysis, J Chemom 30 (2016) 651–662. https://doi.org/10.1002/cem.2824.

[158] J. Westerhuis, T. Kourti, J. MacGregor, Analysis of multiblock and hierarchical PCA and PLS models, J Chemom 12 (1998) 301–321. https://doi.org/10.1002/(SICI)1099-128X(199809/10)12:5<301::AID-CEM515>3.3.CO;2-J.

[159] L. Nørgaard, A. Saudland, J. Wagner, J.P. Nielsen, L. Munck, S.B. Engelsen, Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy, Appl Spectrosc 54 (2000) 413–419. https://doi.org/10.1366/0003702001949500.

[160] G. Azevedo, M.D. Torres, P.L. Almeida, L. Hilliou, Exploring relationships between seaweeds carrageenan contents and extracted hybrid carrageenan properties in wild and cultivated Mastocarpus stellatus, Chondrus crispus and Ahnfeltiopsis devoniensis., Algal Res 67 (2022) 102840. https://doi.org/10.1016/J.ALGAL.2022.102840.

[161] G. Azevedo, B. Domingues, H. Abreu, I. Sousa-Pinto, G. Feio, L. Hilliou, Impact of cultivation of Mastocarpus stellatus in IMTA on the seaweeds chemistry and hybrid carrageenan properties, Carbohydr Polym 116 (2015) 140–148. https://doi.org/10.1016/J.CARBPOL.2014.04.069.

[162] B.J. Van Rossum, H. Förster, H.J.M. De Groot, High-Field and High-Speed CP-MAS13C NMR Heteronuclear Dipolar-Correlation Spectroscopy of Solids with Frequency-Switched Lee–Goldburg Homonuclear Decoupling, Journal of Magnetic Resonance 124 (1997) 516–519. https://doi.org/10.1006/JMRE.1996.1089.

[163] Q. Wang, U.G. Nielsen, Applications of solid-state NMR spectroscopy in environmental science, Solid State Nucl Magn Reson 110 (2020) 101698. https://doi.org/10.1016/J.SSNMR.2020.101698.

[164] B.M. Wise, R.T. Roginski, A Calibration Model Maintenance Roadmap, IFAC-PapersOnLine 48 (2015) 260–265. https://doi.org/10.1016/J.IFACOL.2015.08.191.

[165] H.F.F. Halberg, A.Y. Holst, N. Kaufmann, R. Bro, Calibration model fusion, J Chemom 37 (2023) n/a. https://doi.org/10.1002/cem.3350.

[166] C.B. Zachariassen, J. Larsen, F. van den Berg, R. Bro, A. de Juan, R. Tauler, Multi-way analysis for investigation of industrial pectin using an analytical liquid dilution system, Chemometrics and Intelligent Laboratory Systems 84 (2006) 9–20. https://doi.org/10.1016/j.chemolab.2006.03.010.

[167] J. Peters, P. Bühlmann, N. Meinshausen, Causal inference by using invariant prediction: identification and confidence intervals, J R Stat Soc Series B Stat Methodol 78 (2016) 947–1012. https://doi.org/10.1111/rssb.12167.

[168] T. Chopin, E. Whalen, A new and rapid method for carrageenan identification by FT IR diffuse reflectance spectroscopy directly on dried, ground algal material, Carbohydr Res 246 (1993) 51–59. https://doi.org/10.1016/0008-6215(93)84023-Y.

[169] D. Some, RT-MALS: Real-time process analytics and control for vaccine nanoparticles and macromolecules, (n.d.). chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://wyattfiles.s3-us-west-2.amazonaws.com/literature/app-notes/rt-mals/Biotherapeutics+-+Real-Time+MALS+for+Vaccine+PAT.pdf (accessed May 10, 2025).

8 APPENDIX: PAPERS

Carbohydrate Polymers 348 (2025) 122824

Contents lists available at ScienceDirect

Carbohydrate Polymers

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

Chemometric approaches for polysaccharide viscosity profiling

Oksana Mykhalevych [a,b,*], Henrik Stapelfeldt [a], Rasmus Bro [b]

[a] CP Kelco ApS, 4623 Lille Skensved, Denmark
[b] Department of Food Science, University of Copenhagen, 1958 Frederiksberg, Denmark

ARTICLE INFO

Keywords:
Carrageenan
Viscosity
Chemometrics
SEC-MALS-visco
PLS

ABSTRACT

Traditional viscosity measurements for carrageenan are laborious, present practical and environmental challenges, and fail to provide structure-property understanding for application and manufacturing development. We hypothesize that integrating Size Exclusion Chromatography (SEC) with Multi-Angle Light Scattering (MALS) and online viscometry, combined with chemometric techniques, can develop a more efficient and environmentally friendly method for determining the apparent viscosity of carrageenan solutions.

To test this hypothesis, predictive chemometric models were developed using SEC-MALS data for carrageenan extracted from four different seaweed species. By integrating SEC-MALS with Partial Least Squares (PLS) regression, key molecular parameters such as hydrodynamic radius, intrinsic viscosity, and molecular mass were identified as significant influencers of viscosity. The model for carrageenan from *Eucheuma denticulatum* yielded the lowest prediction error (RMSEP 8.4), while those for carrageenan extracted from *Kappaphycus alvarezii* or from several species of the *Chondrus* genus showed higher errors due to κ-carrageenan sensitivity. For carrageenan extracted from seaweed of the *Gigartina* genus, incorporating the root mean square radius resulted in a low prediction error of 10.

This study concludes that integrating SEC-MALS with PLS regression effectively identifies key molecular parameters influencing carrageenan viscosity, enhancing structure-property understanding and providing a reliable analytical method for optimizing quality control and application in various industries.

1. Introduction

Carrageenan is a very complex sulfated polysaccharide commercially extracted from the cell walls and the intercellular matrix of red seaweeds (*Rhodophyta*). This polysaccharide is an ionic hydrocolloid used as a gelling and stabilizing agent in food and cosmetics. Carrageenans are sulfated galactans with a backbone of altering disaccharide repeating units of β-D-galactose (G-units) linked to position 3 and α-D-galactose (D-units) or 3,6-anhydro- α-D-galactose (DA-units) at position 4. Carrageenans have one or two sulphate groups esterified to a hydroxyl group at carbon atoms 2 or 6. Traditionally, six common carrageenan types are distinguished: Iota (ι-), Kappa (κ-), Lambda (λ-), Nu (η-), Mu (μ-) and Theta (θ) (Campo et al., 2009). Commercial ι- and κ- carrageenans are extracted from red seaweeds *Eucheuma denticulatum* (trade name *Eucheuma spinosum* or Spinosum) and *Kappaphycus alvarezii* (trade name *Eucheuma cottonii* or Cottonii), respectively, while hybrids ι/κ-carrageenans are extracted from several species of the *Gigartina* and *Chondrus* genera (De Ruiter & Rudolph, 1997; van de Velde, 2008). The different types vary in molecular structure, e.g., in the 3,6-anhydrogalactose and ester sulphate content. Carrageenan is a high molecular mass hydrocolloid with a high degree of polydispersity (Almeida et al., 2010). Therefore, there are many structural variations. The five mentioned types are idealized archetypes (De Ruiter & Rudolph, 1997; van de Velde & de Ruiter, 2002).

The functional characteristics of carrageenan, including viscosity, are significantly influenced by its structural variations. Viscosity is particularly crucial due to its impact on the processing and application of carrageenan in various industries and is subject to regulatory scrutiny by, e.g., the European Food Safety Authority (EFSA). Traditionally, viscosity measurements have been performed using methods outlined in the Food Chemicals Codex (FCC) (National Research Council, 1981) that are not only labor-intensive but also require substantial amounts of sample material and water, posing practical and environmental challenges, and do not give molecular informative knowledge that can be used to optimize the manufacturing process.

Advanced analytical techniques such as Size Exclusion

* Corresponding author at: CP Kelco ApS, 4623 Lille Skensved, Denmark.
E-mail address: oksana@food.ku.dk (O. Mykhalevych).

https://doi.org/10.1016/j.carbpol.2024.122824
Received 24 July 2024; Received in revised form 11 September 2024; Accepted 30 September 2024
Available online 1 October 2024

Chromatography (SEC) coupled with Multi-Angle Light Scattering (MALS), refractive index, and viscometry detectors (we use SEC-MALS-visco as abbreviation for the whole system) suggest an efficient alternative for carrageenan characterization. This method leverages the radii of gyration (*Rg*, or root mean square radius (RMS radius)), which represent the average distance from the center of gravity of a molecule in solution, to elucidate the structural properties of carrageenan that contribute to its functionality. The addition of a viscometer provides additional structural information that relates intrinsic viscosity and the molecular mass of a polymer via Mark-Houwink relation (Vreeman et al., 1980). A combination of MALS detector and refractive index detector, that provides polysaccharide concentration, gains measure of absolute molar masses of carrageenan. Over the past 30 years, many publications have focused on the physicochemical characterization of carrageenan through SEC-MALS (Marcelo et al., 2005; Nickerson & Paulson, 2004; Robal et al., 2017; Slootmaekers et al., 1991; Vreeman et al., 1980). In this study we combine this instrumentation with chemometrics for prediction of apparent viscosity of carrageenan.

The viscosity of a carrageenan solution is intrinsically linked to its molecular size and chemical structure, which dictate the molecule's spatial conformation in aqueous systems. Viscosity arises from the energy-consuming collisions and associations between molecules, which generate friction. Larger molecules, characterized by greater hydrodynamic volumes, are more likely to collide at equivalent concentrations, thus increasing viscosity. Besides molecular structure of the polymer it is also influenced by the temperature and polymer concentration. The relationship between apparent viscosity for non-Newtonian fluids, such as carrageenan solution, and the structure of hydrocolloids can be characterized by the power-law model (Eq. (1.3) derived from Eq. (1.1) and from power-law function 1.2) (Marcotte et al., 2001; Rao & Kenny, 1975; Speers & Tung, 1986). The viscosity for Newtonian fluids is a relation between shear stress and shear rate:

$$\eta = \frac{\sigma}{\gamma} \tag{1.1}$$

where η is the viscosity, σ is the shear stress, γ is the shear rate.

For non-Newtonian fluids the power law-model is used:

$$\sigma = m\gamma^n \tag{1.2}$$

where, m is consistency coefficient, and n is the flow behavior index for carrageenan. The two coefficients are usually experimentally calculated as slope (n) and intercept (m) of a logarithmic plot of shear stress vs shear rate (Marcotte et al., 2001; "Viscosity", 1989).

Relation between the apparent viscosity η_{app} and power-law function is derived from the two equations above, giving:

$$\eta_{app} = m\gamma^{n-1} \tag{1.3}$$

The coefficients m and n affected by structure and concentration of polysaccharides and are unique for each of them (Marcotte et al., 2001).

Apparent viscosity is not directly related to intrinsic viscosity. However, intrinsic viscosity reflects the ability of carrageenan to enhance solution viscosity, which depends on molecular size and shape (Podzimek, 2011). Therefore, intrinsic viscosity show correlation to apparent viscosity. Additionally, the Mark-Houwink-Sakurada equation (Eq. (1.4)) relates intrinsic viscosity to the molecular mass of carrageenan, providing insight into polymer chain conformation.

$$[\eta] = KM^a \tag{1.4}$$

where $[\eta]$ is intrinsic viscosity, K is a constant, M is molar mass, and a is a constant. The constants are affected by the molecular size and shape but also are specific for each solvent and temperature (Podzimek, 2011). Intrinsic viscosity is also used to derive hydrodynamic radius that is a measure of hydrodynamic volume and radius and provides insight into the size of the particle (molecule seen as a sphere) in a solution. Its value is influenced by the shape and surface of the molecule.

These insights provide a foundation for a more nuanced understanding of how carrageenan's molecular characteristics influence its functional properties. However, the data obtained by this measurement is not straightforwardly translatable to functionality (in this case – the viscosity measurement) of carrageenan. Therefore, in this study, SEC-MALS-visco is integrated with chemometrics to develop a predictive model for carrageenan's viscosity.

The present study aims to develop a chemometric model for predicting the viscosity of carrageenan derived from four different seaweed species. This innovative approach not only offers a significantly faster and more sustainable alternative to traditional viscosity measurements methods but also provides enhanced insights into the molecular structure of carrageenan. By leveraging chemometric techniques, we can achieve accurate and efficient viscosity predictions, thereby advancing understanding and application of carrageenan in various industries and provides the manufacturing industry with an understanding of the molecular test characteristics that need preservation in process optimizations.

2. Method and materials

2.1. Samples

Carrageenan samples were obtained from CP Kelco ApS (Lille Skensved, Denmark). The samples were chosen to cover a range of values for the viscosity.

The sample domain comprised 503 carrageenan samples, comprising 121 from *Kappphycus alvarezii (trade name Eucheuma cottonii or* Cottonii*)*, 128 from *Chondrus* genus, 156 from *Eucheuma denticulatum* (trade name *Eucheuma spinosum* or Spinosum), and 98 from several species of the *Gigartina* genus.

All chemicals utilized in this study were of analytical grade. Exclusively, ion-exchanged water was employed throughout the experiments.

2.2. Viscometry

Determination of the viscosity of an aqueous 1.5 % carrageenan solution was conducted as described by Food Chemical Codex (National Research Council, 1981). Carrageenan was dissolved in water under continuous stirring for 30 min before transferring the resulting dispersion to a water bath maintained at 85 °C, allowing the solution to reach a temperature of 80 °C (20–30 min). Water was added to adjust for loss by evaporation. Viscosity measurements were taken twice using a Brookfield LVF viscometer (Brookfield Engineering Laboratories, Inc. (Middleboro, Massachusetts, USA)) using preheated guard and spindle No.1, (19 mm in diameter and 65 in length), at a speed of 30 rpm for 30 s. The viscosity results correspond to the measure of apparent viscosity, originally recorded in centipoises, were normalized to a range of 1–100 for the purposes of this study.

The classical viscosity measurement method was evaluated for its environmental impact using the analytical greenness (AGREE) index (Pena-Pereira et al., 2020) achieving a score of 0.52 (Appendix A).

2.3. SEC – MALS - visco

SEC-MALS-visco analysis was conducted using an Agilent 1260 Infinity II HPLC system (Agilent Technologies Inc., Santa Clara, USA) equipped with a Shodex OHpak LB-806 M column (8.0 × 300 mm, ID x length) and a Shodex OHpak SB-G 6B guard column (6.0 × 50 mm, ID x length) (Resonac Europe GmbH, Wiesbaden, Germany). The system included a MALS detector, model DAWN 8, operating at a wavelength of 660 nm, a ViscoStar viscometer, and an Optilab differential refractometer (dRi) operating at 658 nm (Wyatt Technology, Santa Barbara, USA). Elution was performed using 0.1 M LiNO3 at a constant flow rate of 1 mL/min at a system temperature of 40 °C. The specific refractive index

increment (dn/dc) used was 0.11.

Samples were prepared by dissolving carrageenan in the eluent at a concentration of 2 mg/mL, followed by heating to 80 °C under stirring for 30 min. After cooling, 30 μL of the solution was injected into the column. The total run time for each measurement was 32 min.

Data analysis was carried out using ASTRA software (Wyatt Technology Corp.). The Zimm fit method (Podzimek, 2011) was applied for all samples. A comprehensive description of SEC-MALS and size characteristics may be found in Podzimek (2011).

In total, twelve predictor variables obtained by SEC-MALS-visco were included in this study: number average molecular mass (M_n), molar mass at the top of the concentration peak (M_p), weight average molecular mass (M_w), *Z*-average molar mass (M_z), uncertainty-weighted average molar mass ($M_{(avg)}$), polydispersity (M_w/M_n), weight average mean square radius (r_w), weight average hydrodynamic radius ($rh(v)_w$), uncertainty weighted average hydrodynamic radius $rh(v)_{avg}$, mass recovery (R%), weight average intrinsic viscosity $[\eta]_w$, uncertainty-weighted average intrinsic viscosity $[\eta]_{(avg)}$. The mentioned variables were automatically estimated using ASTRA software and the equations behind these calculations are summarized in the Appendix B.

SEC-MALS-visco method was evaluated for its environmental impact using the AGREE index (Pena-Pereira et al., 2020). It achieved a higher score of 0.64 compared to the classical viscometry method (see Appendix A), indicating a reduced environmental impact. This improvement is due to lower solvent and sample consumption, less waste generation, and the automation of the method.

2.4. Data analysis

In total, 503 carrageenan samples were measured by the SEC-MALS-visco on the reference method. The average across duplicates was used in this study. Prior to model development, the datasets were shuffled in row dimension and then divided into calibration and validation datasets using the Kennard-Stone algorithm (Kennard & Stone, 1969) keeping 80 % of the data in the calibration dataset.

Both the SEC-MALS-visco and the reference data were autoscaled to get zero as a mean and one as standard deviation for each variable. The reference data was normalized before preprocessing so all values were in 1–100 range.

Partial least squares (PLS) regression (Wold et al., 2001) was used as a linear regression model for predicting carrageenan viscosity. Predictive models were developed for four data classes based on the raw material (seaweed species) carrageenan was extracted from: spinosum model, included carrageenan extracted from *Eucheuma spinosum*, cottonii model – included Cottonii carrageenan samples data, chondrus model trained on data for carrageenan samples extracted from *Chondrus* genus and gigartina model for samples extracted from several seaweed species of the *Gigartina* genus.

The SEC-MALS-visco dataset consisting of twelve variables was transformed into a dataset with one hundred and fourteen variables by adding polynomials (squares, cubes, quartics) and crossterms. This approach was taken after observed nonlinear behavior in PLS models developed using the original twelve variables. By adding polynomial and cross-terms to the dataset, the feature space is expanded to capture nonlinear relationship between predictors and the response variables. Although a linear PLS model is applied, these transformations enable the model to behave in a nonlinear manner relative to the original variables (Berglund & Wold, 1997).

Logarithmic transformation of the response variable was tested as well, however adding polynomials and crossterms resulted in slightly better predictions.

Several non-models, including neural networks and support vector machines, were initially tested (data is not shown) and resulted in performance that was similar to or worse than the PLS regression model described in this study. However, PLS regression was chosen for this study due to its interpretability and ease of implementation in industrial settings.

The number of variables was reduced by using forward variable selection method (Nørgaard et al., 2000).

The average prediction error, Root Mean Square Error (RMSE), was used as a metric to evaluate model performance, and it was calculated in cross-validation (RMSECV). The quality of the models for feature predictions was determined from a validation set of data (RMSEP) that was firstly calculated after the final predictive model for each class was defined.

The optimal number of components for PLS was selected according to the lowest RMSECV by performing cross-validation with ten cancelation groups.

The data analysis was performed using MATLAB (Release R 2023a, version: 9.14.0; The MathWorks Inc.) and PLS_Toolbox 9.2 (2023). Eigenvector Research, Inc.

3. Results and discussion

3.1. Exploratory data analysis

The size exclusion chromatogram for carrageenan samples extracted from the four different seaweed types is presented in Fig. 1 (four randomly chosen chromatograms from the dataset). The light scattering (LS) signal from the MALS detector indicates the molar masses eluting from the SEC column during the elution time. The distinct peak between 5.90 and 10.20 min corresponds to the carrageenan signal. Signals observed between 10.20 and 14.30 min indicate ionic impurities (Sen et al., 2013), at the end of the chromatogram (not shown here) the broad peaks from 20.0 to 32.0 min are observed corresponding to breakthrough peaks caused by the sample passing through the delay column.

Table 1 summarizes statistics for the twelve molecular parameters obtained by SEC-MALS-visco measurements. Parameters presented in the table are correlated since, as it can be seen from the equations in Appendix B, they are often derived from each other. One of the most important parameters determined by SEC-MALS-visco is the weight average molecular mass (Mw) since it influences functionality of carrageenan such as viscosity. The mean weight average molecular mass is highest for the spinosum class with value of 1102 kDa, following by the chondrus class with 905 kDa, cottonii class with 621 kDa and the gigartina exhibit the lowest molecular mass average of 573 kDa. One could expect comparable order of the intrinsic viscosity among these carrageenan classes since the value of intrinsic viscosity depends on the molecular mass through the Mark-Houwink relation (Vreeman et al., 1980). However, the weight average intrinsic viscosity, $[\eta]_w$, is highest for chondrus class with its value of 947 mL/g, while the other classes exhibit the weight average intrinsic viscosity of approximately 700 mL/g. This may be explained by the fact that the squared molecular mass is used to calculate the value of the weight average molecular mass and therefore it is strongly influenced by the larger carrageenan molecules, i.e., this estimate is very close to the molecular mass of the largest molecules present in carrageenan sample. Furthermore, for polydisperse samples the value of intrinsic viscosity is also influenced by the concentration of each molecular size species (Vreeman et al., 1980). The value of polydispersity is over 1.5 for all carrageenan classes indicating highly polydisperse structure of the sample.

Root mean square radius (RMS, also known as radius of gyration) describes the distribution of mass around the center of gravity of the molecule, i.e., it corresponds to the size of the molecule in the solution. The value of this parameter given in Table 1, RMS (nm), is weight average RMS radius. From Table 1, it may be observed that molecules belonging to the spinosum class are larger in terms of their shape followed by chondrus carrageenan, while the other two classes exhibit similar shape size. This observation can be explained by the molecular structure of carrageenans extracted from *Eucheuma spinosum,* which yields ι-carrageenan that contains two sulfur groups per disaccharide unit compared to one sulfur group per disaccharide unit in

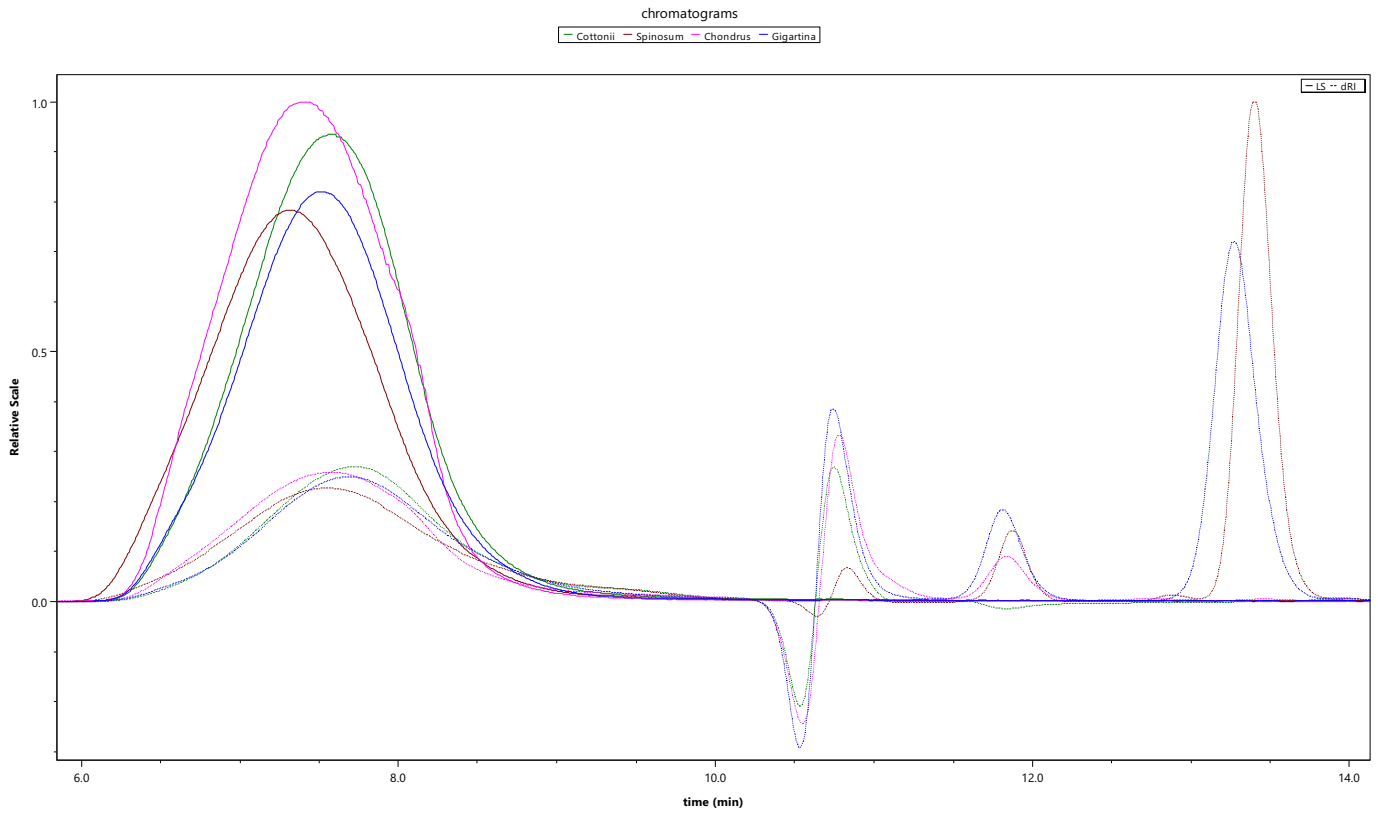


Fig. 1. Chromatograms of carrageenan extracted from four carrageenan types extracted from seaweed species of Chondrus genus (magenta), Cottonii (green), Gigartina genus (blue), Eucheuma spinosum (maroon). The signal from MALS is shown as a solid line, while dashed line shows signal from concentration detector.

Table 1
Mean and standard deviation values for each parameter obtained by SEC-MALS-visco. The data was grouped by raw material classes and used to develop the four corresponding predictive models towards carrageenans viscosity.

	M_n (kDa)	M_p (kDa)	M_w (kDa)	M_z (kDa)	$M_{(avg)}$ (kDa)	Polydispersity (Mw/Mn)	RMS (nm)	$Rh(v)_w$ (nm)	$Rh(v)_{(avg)}$ (nm)	Mass recovery (%)	$[\eta]_w$ (mL/g)	$[\eta]_{(avg)}$ (mL/g)
Spinosum												
Mean	572	1168	1102	1648	445	1.98	109	47	39	85	720	615
Std. Dev	192	456	289	342	128	0.23	19	7	6	6	150	141
Cottonii												
Mean	309	573	621	941	275	2.06	72	39	32	92	701	575
Std. Dev	62	67	75	108	66	0.32	7	3	3	3	77	86
Chondrus												
Mean	524	788	905	1296	549	1.74	94	49	47	86	947	829
Std. Dev	66	77	82	104	94	0.11	6	3	4	4	83	92
Gigartina												
Mean	309	483	573	918	290	1.87	71	37	32	83	676	542
Std. Dev	77	98	130	189	95	0.13	11	5	6	3	133	105

κ-carrageenan (De Ruiter & Rudolph, 1997). However, the fraction of ι-carrageenan in the hybrid carrageenan extracted from species of the *Gigartina* genus is higher compared to its part in hybrid carrageenans extracted from *Chondrus* genus (Falshaw et al., 2001; Van De Velde et al., 2005). Therefore, it was expected that RMS and R_h would be higher for gigartina class carrageenan compared to the chondrus class, but the opposite is observed. This discrepancy may be explained by differences in the distribution of ι- and κ- regions within the carrageenan molecules, which could influence the overall molecular conformation and, consequently, affect the RMS radius values.

Furthermore, the concentration of carrageenan, determined as recovery percentage, shows slight variations between carrageenan classes. The highest recovery percentage is observed in samples extracted from Cottonii seaweed, with a value of 92 ± 3 %. This is followed by samples

from spinosum class at 85 ± 6 %, and chondrus and gigartina classes at 86 ± 4 % and 83 ± 3 %, respectively. These percentages represent the portion of carrageenan powder that is pure carrageenan, while the remaining content may include salts, moisture, ash, or minor residuals from the extraction process. Consequently, although the mass of carrageenan powder used during viscosity measurement remains constant, the concentration of pure carrageenan varies between samples. This means that the variation in carrageenan concentration might influence the reproducibility of the reference values. It should be noted that other conditions, such as shear rate, temperature, pressure, and time of shearing, were kept constant throughout the measurements of all carrageenan samples as described in subsection 2.3.

Mark-Houwink-Sakurada plot in Fig. 2 shows molars mass vs intrinsic viscosity in a double logarithmic scale. This plot reveals additional information about the structure of the polymers. The slope of the curve is exponent a in Mark-Houwink-Sakurada equation, while the intercept is K, the Mark-Houwink constant (see Eq. (1.4)). The exponent a contains information about conformation of carrageenan in the solution. The exponent a given in Fig. 2 is in range 0.5–0.7 for all four polysaccharides indicating random-coil conformation (Hill et al., 1998) which corresponds well with earlier findings (Vreeman et al., 1980). It is noteworthy that the Mark-Houwink fit shows good agreement for intermediate molar masses, but there is variability in intrinsic viscosity at the same molar mass, particularly for the Spinosum and Cottonii samples. These variations can be explained by differences in the molecular conformation and flexibility of the carrageenan molecules. Structural features such as the degree of sulfation and the distribution of ι- and κ-regions likely affect how the molecules coil or expand in solution, even at the same molecular mass. More expanded or flexible molecules would lead to higher intrinsic viscosities, and this variation is particularly evident in Spinosum and Cottonii carrageenans. In Appendix C, more Mark-Houwink-Sakurada plots of carrageenan can be found.

Fig. 3 shows the relation between RMS radius vs. molar mass known as a conformation plot. It is usually used for characterization of polymers with respect to branching. The slopes around 0.58 indicates a not branched polymer and it will decrease with presence of branching (Podzimek, 2011). As it was expected for carrageenan molecular structure, the slope of around 0.55–0.71 indicates absence of branching. However, as seen in Fig. 3, some samples with similar molecular masses exhibit different RMS radii. This discrepancy can also be attributed to variations in the molecular structure, such as sulfate group distribution and molecular conformation.

3.2. Predictive modelling

Fig. 4 shows the predicted vs. measured viscosity plots for PLS models developed for each carrageenan class separately.

Figs. 5–8 depicting PLS regression coefficients for each model show the relative influence of each variable measured by SEC-MALS-visco on the predicted viscosity of carrageenan, with positive and negative coefficients indicating variables that increase or decrease viscosity, respectively. To interpret the chemistry behind the relationship established by the PLS between the SEC-MALS-visco data and carrageenan viscosity, the regression coefficients were analyzed. The coefficients with positive and negative values suggest variables that are positively or negatively correlated with carrageenan viscosity, respectively. However, it is important to note that these coefficients also reflect compensatory adjustments for correlations among predictor variables. This means that negative coefficients do not necessarily indicate a direct negative influence on viscosity but may also represent an adjustment for the presence of correlations with properties other than viscosity. Therefore, caution is warranted when making qualitative interpretations solely

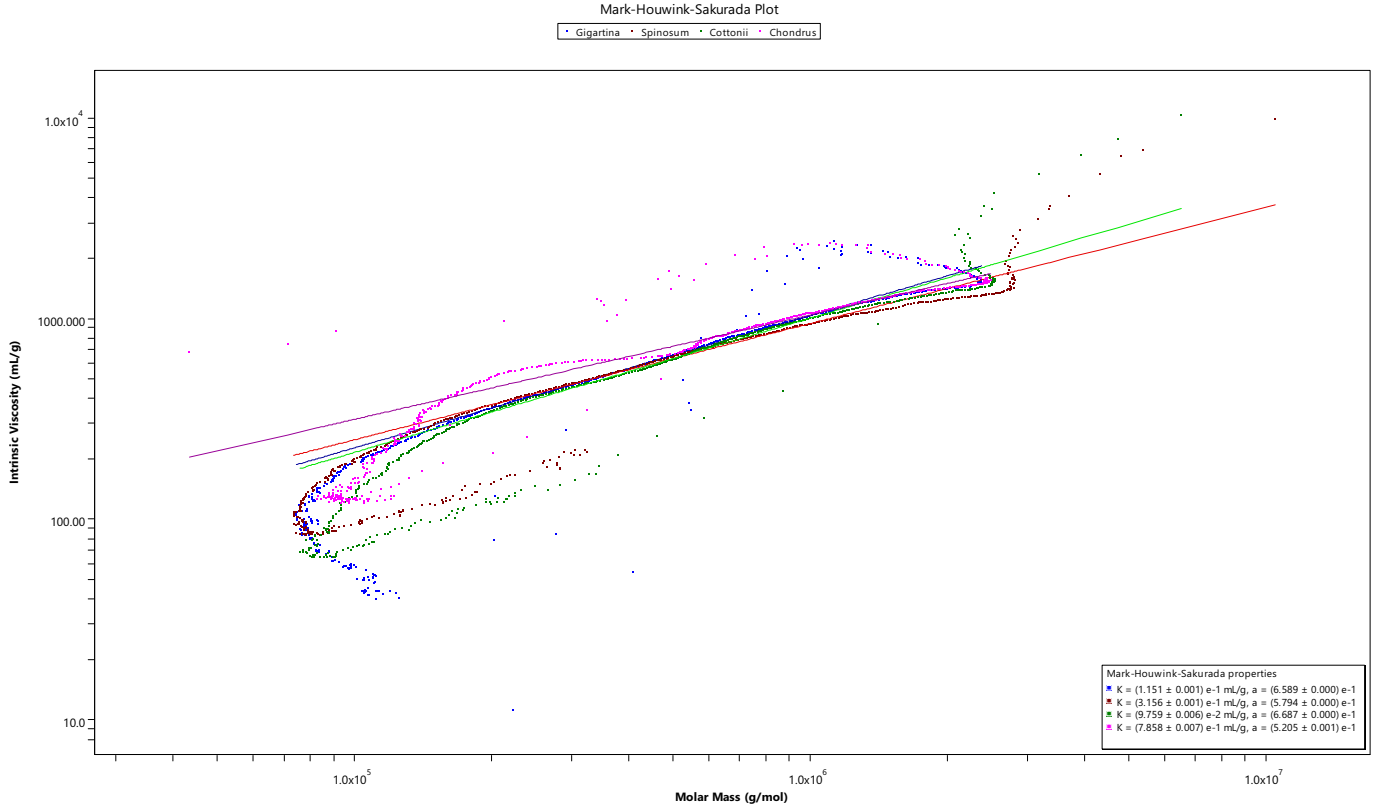


Fig. 2. Mark-Houwink-Sakurada Plot for four carrageenan types extracted from seaweed species of Chondrus genus (magenta), Cottonii (green), Gigartina genus (blue), Eucheuma spinosum (maroon). The Mark-Houwink-Sakurada exponent a was calculated for all four carrageenan types and is displayed in the right corner of the figure. It was 0.66, 0.58, 0.68, 0.52 for carrageenan extracted from seaweed species of Gigartina genus, Eucheuma spinosum, Cottonii, and from Chondrus genus, respectively.

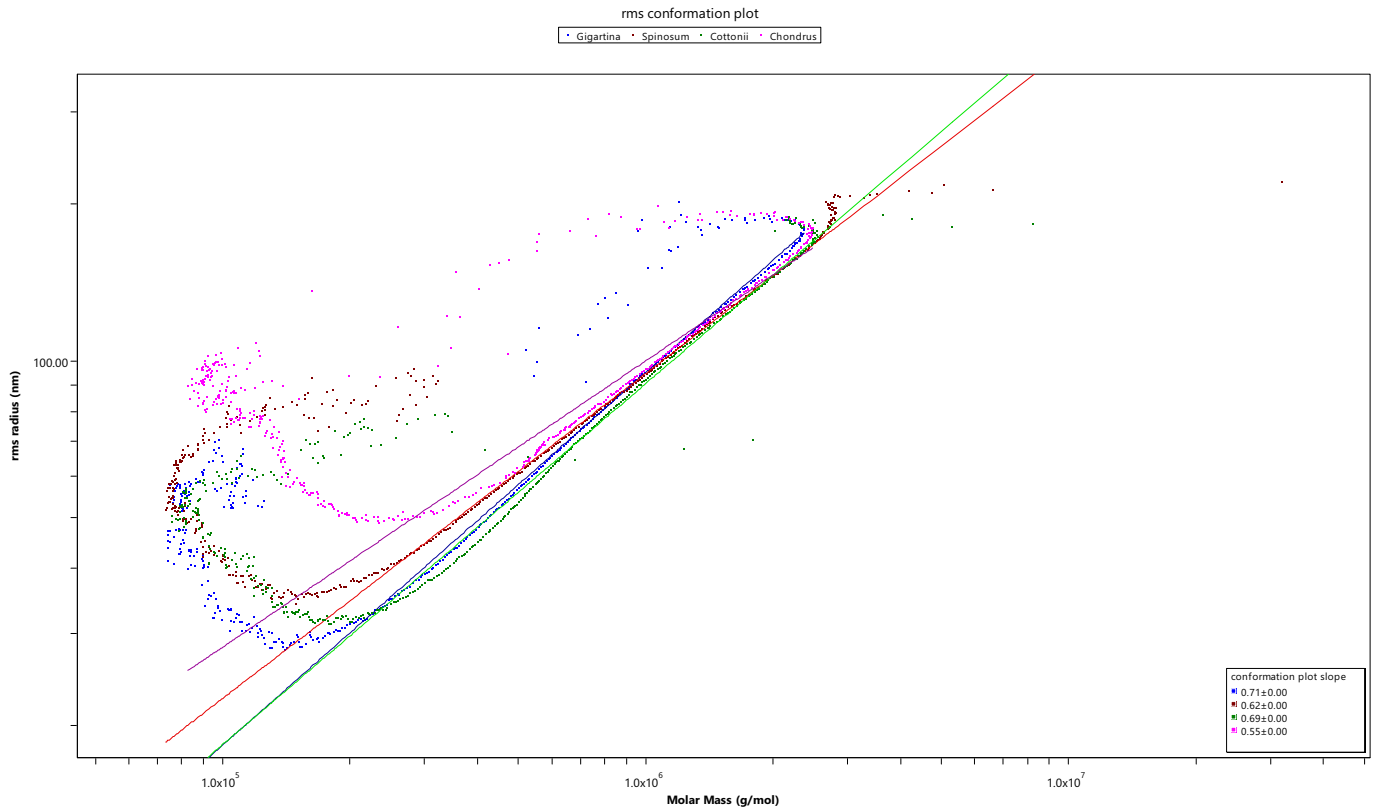


Fig. 3. Conformation plot for four carrageenan types extracted from seaweed species of Chondrus genus (magenta), Cottonii (green), Gigartina genus (blue), Eucheuma spinosum (maroon). The conformation plot slope was calculated for all four carrageenan types and is displayed in the right corner of the figure. It was 0.71, 0.62, 0.69, 0.55 for carrageenan extracted from seaweed species of Gigartina genus, Eucheuma spinosum, Cottonii, and from Chondrus genus, respectively.

based on the sign of these coefficients.

Despite this caution, analyzing the regression coefficients can still provide valuable insights into which variables are most important in influencing carrageenan viscosity and, importantly, any suggested observation can be verified by going back to the raw data. For instance, coefficients with larger absolute values may indicate variables that have a more substantial impact, either directly or indirectly, on viscosity. This understanding is crucial for identifying key molecular characteristics that need to be preserved or modified during the manufacturing process to achieve the desired viscosity. As always when interpreting empirically observed correlations causality claims must be backed by other evidence.

Based on data shown in Fig. 4, the lowest prediction error, RMSEP, of 8.4 was obtained for carrageenan samples extracted from *Eucheuma spinosum*, with four principal components (=latent variables (LV) in the figure). Out of the 114 variables originally included in the dataset, nine were selected during forward variable selection and included in the final model, as shown in Fig. 5, which displays the PLS regression coefficients for the model. Both the weight-averaged and uncertainty-weighted average hydrodynamic radius are significant for predicting apparent viscosity. These variables appear in multiple forms, including their cubes and quartics, and their interaction with carrageenan mass recovery. This indicates that not only the average size of the molecules but also the variability and higher-order interactions of the hydrodynamic radius are crucial for understanding the carrageenan solution's viscosity. Despite the correlation between the hydrodynamic radius and molecular mass, the quartic form of weight average molecular mass was also included in the model, suggesting a nonlinear influence on apparent viscosity, and indicating that the hydrodynamic radius provides additional information, such as its interaction with the solvent.

Furthermore, the positive regression coefficient for the weight average intrinsic viscosity indicates that carrageenan apparent viscosity increases with intrinsic viscosity. Higher intrinsic viscosity implies larger or more interactive molecules, leading to increased resistance to flow. The product of polydispersity and weight average molecular mass is another important predictor. This variable captures the combined effect of the distribution of molecular sizes (polydispersity) and the average molecular weight, both of which impact the rheological properties of the solution. A wide range of molecular masses can lead to complex interactions within carrageenan solution, affecting its flow characteristics.

The prediction error for viscosity for carrageenan extracted from Cottonii seaweed was 24, the highest among all four classes. Based on the PLS coefficients for the cottonii model, shown in Fig. 6, the product of carrageenan concentration and weight average intrinsic viscosity was one of the important variables, along with polydispersity to the fourth power and the product of polydispersity and weight average molecular mass. The concentration and weight average intrinsic viscosity product could be explained by the fact that higher concentrations and larger molecular weights enhance polymer entanglements and network formation, increasing viscosity. The term polydispersity to the fourth power emphasizes the impact of a broad molecular mass distribution, which affects the complexity and stability of the viscosity. Finally, the interaction between polydispersity and molecular weight captures the influence of both the size range and the magnitude of the molecular masses on the solution's rheological properties. The reason for selection of a product of e.g., polydispersity and weight average molecular mass, instead of taking these parameters independently, could be that the product is needed to capture the synergistic effect of having a broad molecular mass distribution combined with significant high molecular weight fractions.

The PLS model for carrageenan samples extracted from seaweed of the *Chondrus* genus had a prediction error of 17. Investigation of the PLS coefficients in Fig. 7 reveals the importance of parameters consistent

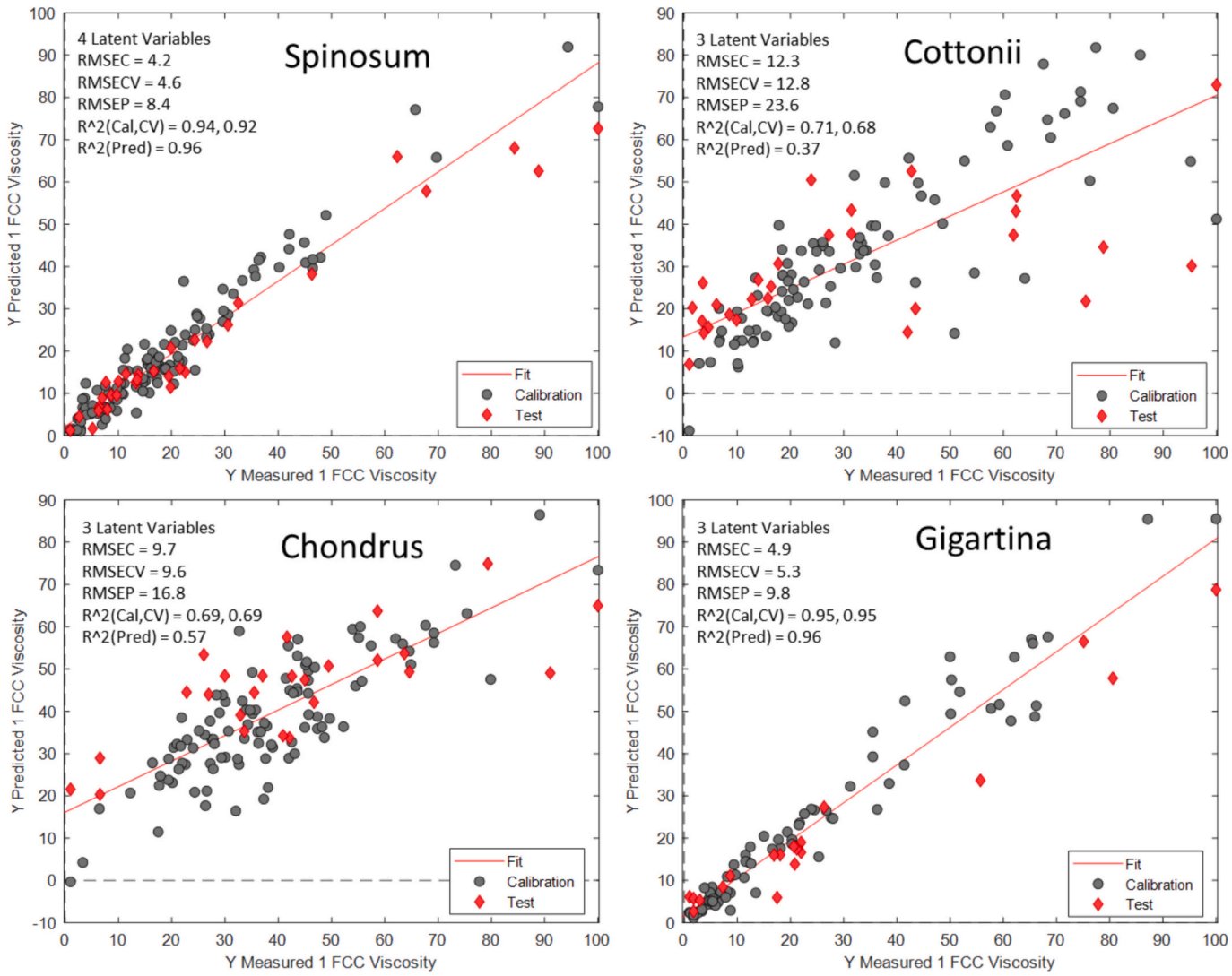


Fig. 4. Predicted vs measured viscosity (method defined by FCC) plots for PLS models developed on data for spinosum, cottonii, chondrus and gigartina classes. Both calibration and validation (test) data are shown, with gray data points representing calibration data and red data points representing validation data.

with those included in the model for the spinosum class, such as the hydrodynamic radius, intrinsic viscosity, and molecular mass.

The model developed for the Gigartina class exhibited a slightly different profile, incorporating the RMS radius instead of the hydrodynamic radius, likely due to the higher complexity of this carrageenan type's molecular structure. Despite this difference, the model still included intrinsic viscosity and molecular mass, achieving a prediction error of 10.

The higher prediction error for cottonii and chondrus classes could be attributed to the high percentage of κ-carrageenan in these carrageenan types. The lower number of sulfate groups in κ-carrageenan results in reduced hydration, making it more sensitive to sample preparation conditions such as mixing and temperature (Tye, 1988). This increased sensitivity leads to higher measurement errors in the reference method, which is labor-intensive and difficult to control or adjust specifically to each polysaccharide (e.g., room temperature, mixing speed, etc.). Furthermore, the correlation coefficients on cross-validation for the models developed for these classes were around 0.7, suggesting that a moderate portion of the variability in viscosity is explained by the SEC-MALS-visco data, but there is still substantial unexplained variability. Therefore, the influence of ion concentration and type was tested (data not shown); however, due to limited variation in these parameters in commercial samples, adding this information did not improve the models.

4. Conclusions

In this study, we developed chemometric models to predict the viscosity of carrageenan solutions derived from four different seaweed species using SEC-MALS-visco data. Our models demonstrate that key molecular parameters, including hydrodynamic radius, intrinsic viscosity, and molecular weight, are fundamental in prediction of apparent viscosity.

For carrageenan extracted from *Eucheuma spinosum*, the model achieved the lowest prediction error (RMSEP of 8.4), highlighting the importance of both weight-averaged and uncertainty-weighted average hydrodynamic radius. These variables, along with their higher-order interactions and quartic forms of weight average molecular mass, underscore the nonlinear relationship between molecular size and apparent viscosity.

The higher prediction errors observed for the cottonii (RMSEP of 24) and chondrus (RMSEP of 17) classes can be attributed to the high percentage of κ-carrageenan, which has fewer sulfate groups, resulting in reduced hydration and greater sensitivity to sample preparation conditions. This sensitivity likely leads to higher measurement errors in the reference method.

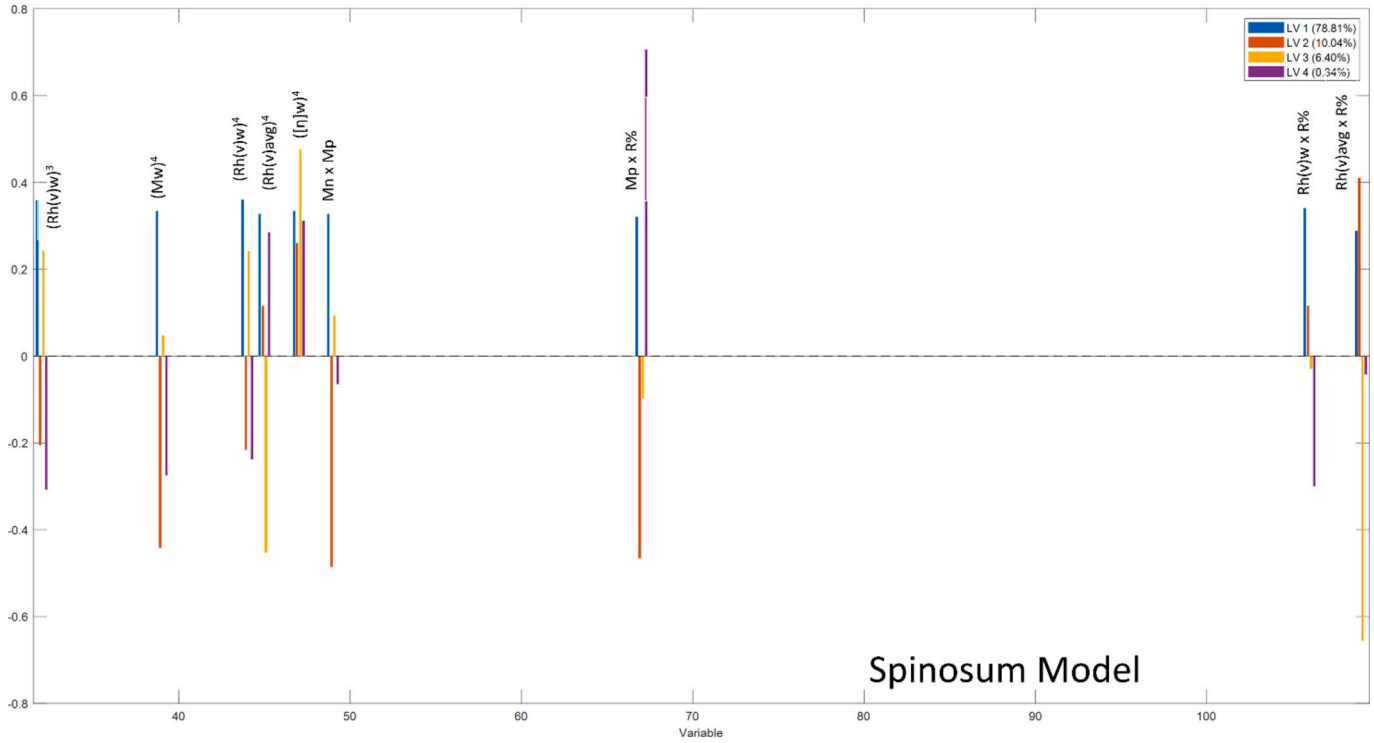


Fig. 5. PLS regression coefficients for the model for carrageenan viscosity prediction based on SEC-MALS-visco data for spinosum class.

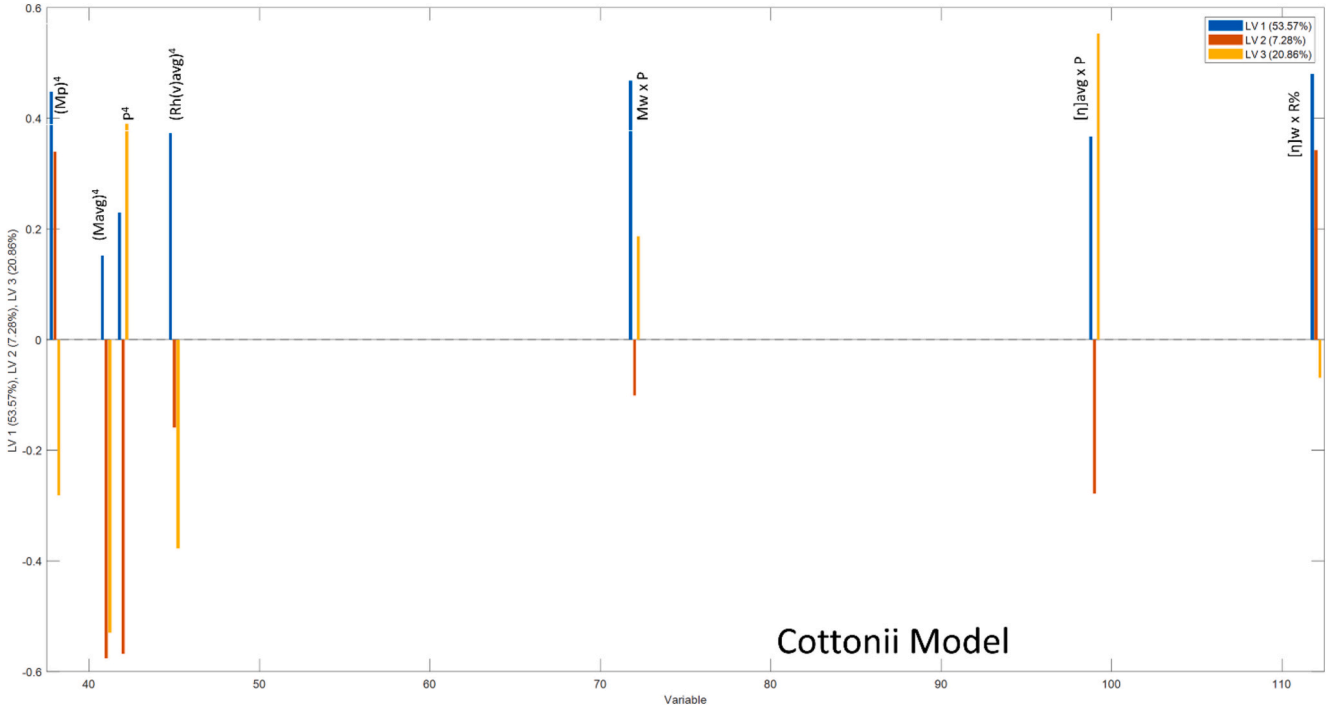


Fig. 6. PLS regression coefficients for the model for carrageenan viscosity prediction based on SEC-MALS-visco data for cottonii class.

The gigartina class model, with an RMSEP of 10, incorporated the RMS radius instead of the hydrodynamic radius, likely reflecting the higher complexity of molecular structure of carrageenan extracted from seaweed of *Gigartina* genus. This model also included intrinsic viscosity and molecular mass, demonstrating robust predictive performance despite the structural differences.

Overall, our study highlights the potential of integrating SEC-MALS-visco with chemometric techniques to provide a faster, more sustainable alternative to traditional viscosity measurements. The predictive models developed here offer enhanced insights into the molecular structure of carrageenan, advancing our understanding and application of this important hydrocolloid in various industries and the molecular characteristics that needs to be preserved during the manufacturing process to obtain the desired functionality. Our decision to use PLS regression

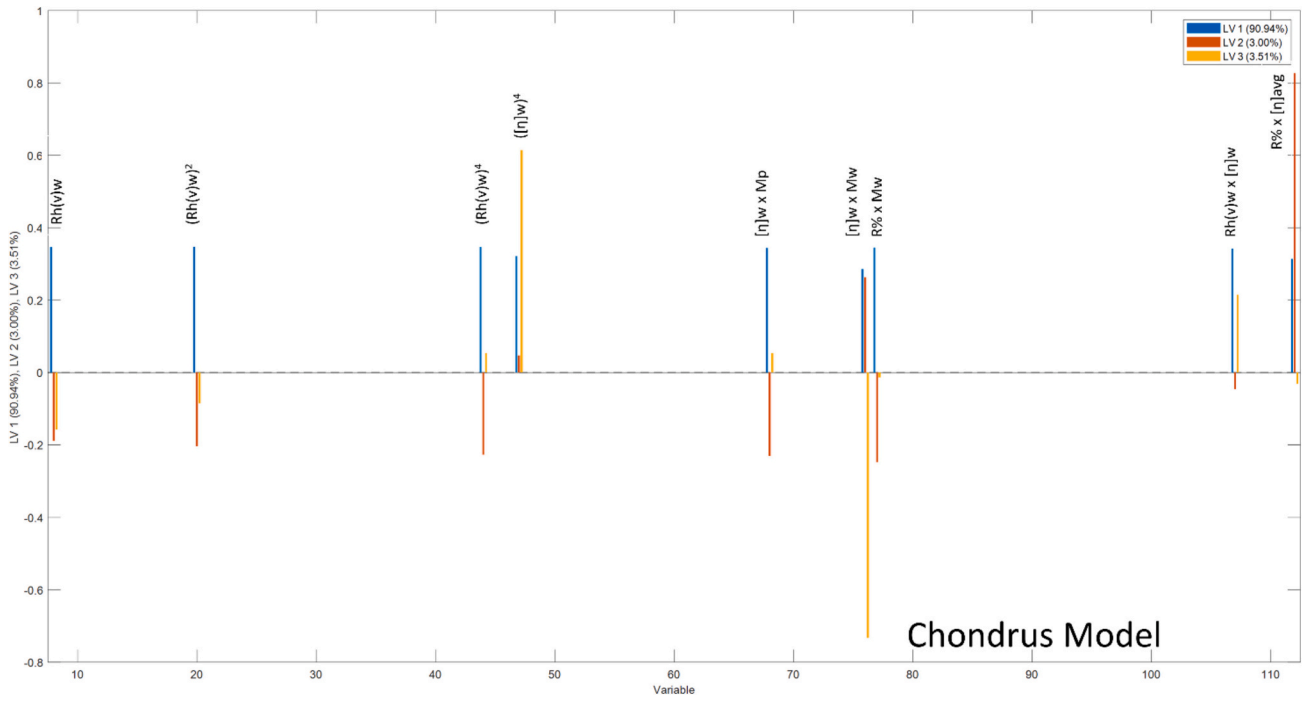


Fig. 7. PLS regression coefficients for the model for carrageenan viscosity prediction based on SEC-MALS-visco data for chondrus class.

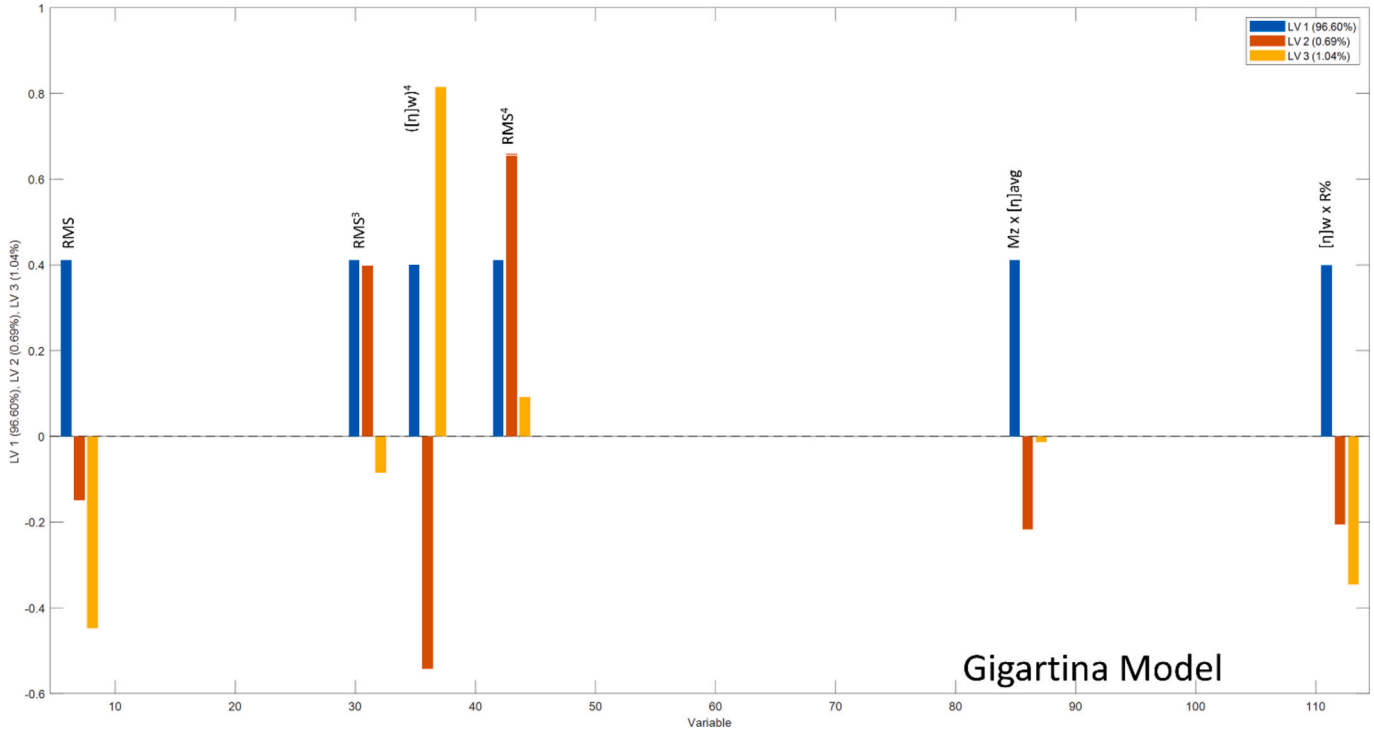


Fig. 8. PLS regression coefficients for the model for carrageenan viscosity prediction based on SEC-MALS-visco data for gigartina class.

with polynomial terms rather than more advanced nonlinear techniques was motivated by several practical considerations. PLS models offer significant advantages in terms of interpretability, which is crucial for industrial applications of the models discussed in this study, where understanding the influence of molecular parameters on viscosity is essential for process optimization and quality control. Moreover, PLS regression is widely supported by industry-standard software tools, making it easier to implement, use, and maintain across various applications and by professionals with diverse backgrounds. The simplicity and robustness of PLS models facilitate seamless integration into existing workflows, reducing the need for extensive retraining or adjustments. Future work could explore nonlinear models as complementary tools when interpretability is less critical. Additionally, testing the applicability of modified linear methods, such as the one proposed by Cook and Forzani (2021), for this data set might provide new insights or yield a simpler model (Cook & Forzani, 2021). Furthermore, future

work should focus on refining these models with additional data and exploring the impacts of other structural variables on carrageenan viscosity. By continuing to improve the accuracy and robustness of these predictive models, we can better meet the demands of industrial applications and regulatory standards.

CRediT authorship contribution statement

Oksana Mykhalevych: Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Henrik Stapelfeldt:** Writing – review & editing, Supervision, Resources, Project administration, Conceptualization. **Rasmus Bro:** Writing – review & editing, Supervision, Resources, Project administration, Conceptualization.

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.

Data availability

Data will be made available on request.

Acknowledgement

We extend our thanks to all colleagues at CP Kelco who shared their deep knowledge of carrageenan with us, especially Dr. Jan Larsen, Jette Andersen and the Analytical QC laboratory personnel for their assistance with method development for SEC-MALS-visco. Special thanks to Dr. Heidi Liva Pedersen for the great discussion of the advanced theory behind carrageenan rheology. Furthermore, Prof. Stepan Podzimek is kindly acknowledged for discussions and critical review of the manuscript. We thank Signe Fobian, B.Eng., internship student in the QC laboratory at CP Kelco, for sharing with us the SEC-MALS-visco measurements of 80 carrageenan samples from her bachelor's thesis project. Special recognition is given to the QC Carrageenan laboratory personnel for their assistance with viscosity measurements and equipment support.

Funding details

This work was supported by CP Kelco ApS as part of the industrial PhD program of the Innovation Fund Denmark (grant number: 104400009B).

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used OpenAI's ChatGPT in order to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Appendix A

AGREE calculator applied:

Criteria/Method	Proposed Method	Classical Method
Sampling procedure	Reduced number of steps	Large number of steps
Minimal Sample amount (g)	0.002	7.5
Positioning of the analytical device	At-line	At-line
No of steps in the sample preparation	1	2
Degree of automation/ sample preparation	Automatic/none or miniaturized	Manual/not miniaturized
Derivatization agent	None	None
Amount of waste per sample (mL)	50	500
Number of analytes per run/ Sample throughput	1/1.5	1/3
Use of energy (estimated by choosing a technique)	LC*	Solvent evaporation (10–150 min)*
Reagents	Some reagents are biobased	All reagents are biobased
Toxic reagents	No	No
Operators' safety	Safe	Safe
Total score	0.64	0.52

* Similar techniques were chosen since the exact techniques used (SEC-MALS, waterbath, viscometer) were not available in the software.

Appendix B

Following equations (Podzimek, 2011, 2014; Wyatt Technology, 2023) were used by the software to calculate all parameters included in the calibration dataset.

For each i_{th} elution volume (V_i) in SEC-MALS, the molar mass, M_i, and the mean square radius, R_i^2, is obtained. Concentration, c_i, in the *i*th elution volume measured by RI detector is calculated using the following equation.

$$c_i = \frac{\alpha(S_i - S_{i,baseline})}{dn/dc} \tag{6.1}$$

where:

dn/dc	specific RI increment
S_i	RI sample signal
$S_{i,baseline}$	Baseline signal

$$M_{n=}\frac{\sum_i c_i}{\sum_i \frac{c_i}{M_i}} \tag{6.2}$$

where:

M_n Number average molecular mass

$$M_{w=}\frac{\sum_i c_i M_i}{\sum_i c_i} \tag{6.3}$$

where:

M_w Weight-average molecular mass

$$M_z = \frac{\sum_i c_i M_i^2}{\sum_i c_i M_i} \tag{6.4}$$

where:

M_z *Z*-average molar mass

$$M_{avg} = \frac{\sum_i M_i \frac{1}{\sigma^2_{M_i}}}{\sum_i \frac{1}{\sigma^2_{M_i}}} \tag{6.5}$$

where:

M(avg) Uncertainty-weighted average molar mass

σ_{Mi} the uncertainty in the value Mi:

$$\sigma_{M_{avg}} = \sqrt{\frac{1}{\sum_i \frac{1}{\sigma^2_{M_i}}}} \tag{6.6}$$

$$\left\langle r_g^2 \right\rangle_w = \frac{\sum_i c_i \left\langle r_g^2 \right\rangle_i}{\sum_i c_i} \tag{6.7}$$

where:

r_w weight average mean square radius

$\left\langle r_g^2 \right\rangle_i$ is mean square radius of the i^{th} elution slice.

The specific viscosity obtained by viscometer is used to calculate the intrinsic viscosity.

$$n_{Sp} = \frac{\eta - \eta_0}{\eta_0} \tag{6.8}$$

where:

η_{sp} Specific viscosity

η Viscosity of polymer solution

η_0 Viscosity of solvent

$$[\eta] = \lim_{c \to 0} \frac{\eta_{Sp}}{c} \tag{6.9}$$

where:

[η] Intrinsic viscosity

$[\eta]_w$ Weight average intrinsic viscosity

[η](avg) Uncertainty-weighted average intrinsic viscosity

The viscosity hydrodynamic radius is calculated from the intrinsic viscosity.

Appendix C

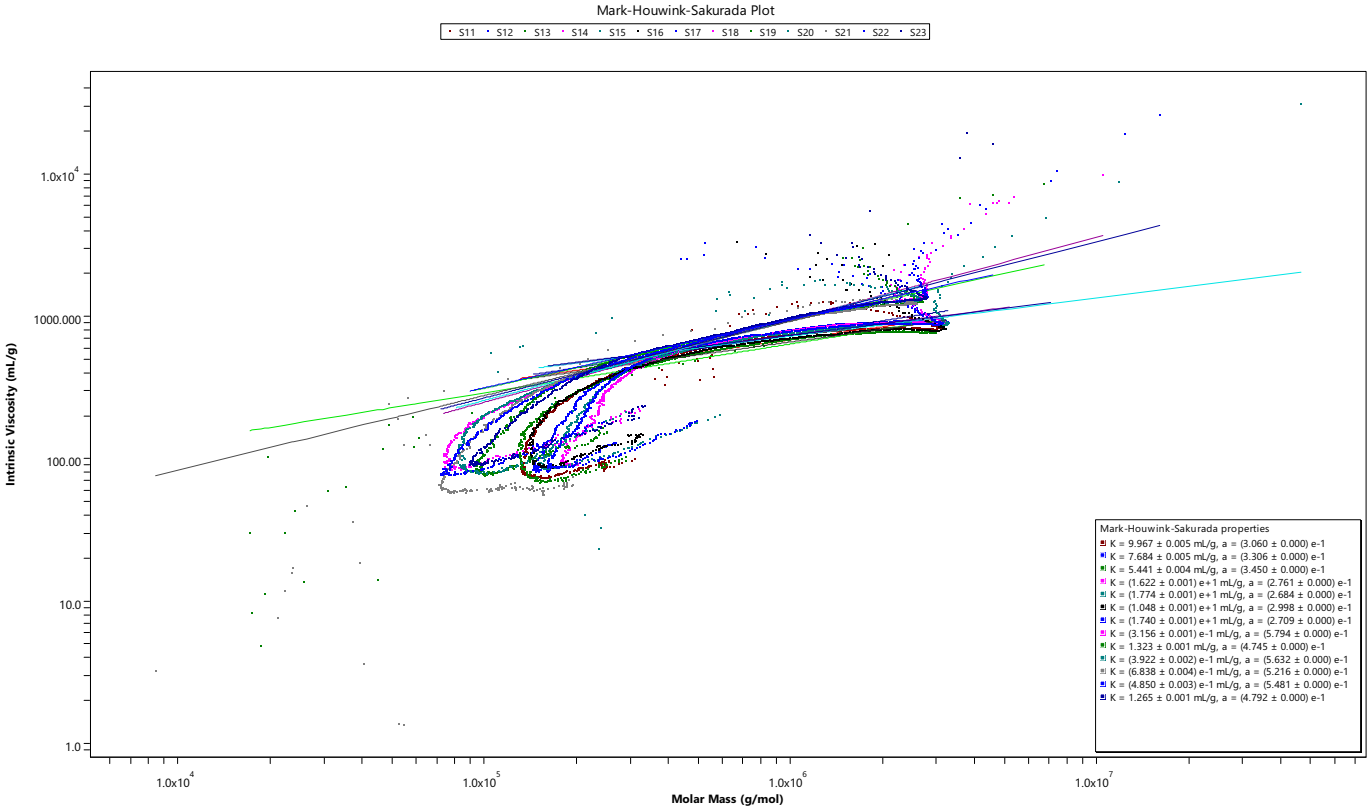


Fig. 9. Mark-Houwink-Sakurada Plot for 13 randomly chosen carrageenan samples extracted from Eucheuma spinosum. MHS a-parameter varies between the samples due to different processing conditions applied in extraction of carrageenan from this seaweed type. During experimental design a large variation of different samples produced by several extraction methods method were incorporated into the data collection aiming to represent all carrageenan types produced from Eucheuma spinosum.

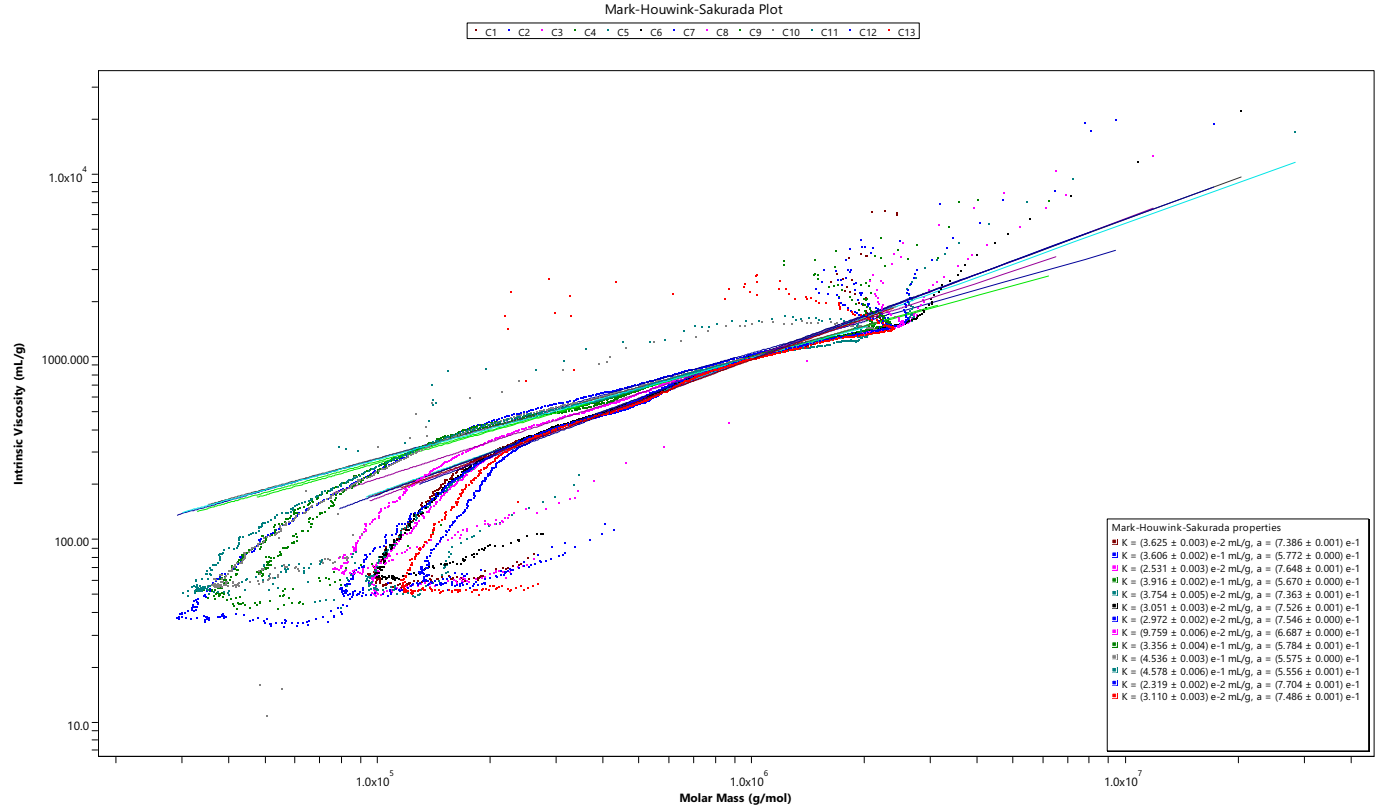


Fig. 10. Mark-Houwink-Sakurada Plot for 13 randomly chosen carrageenan samples extracted Cottonii carrageenans.

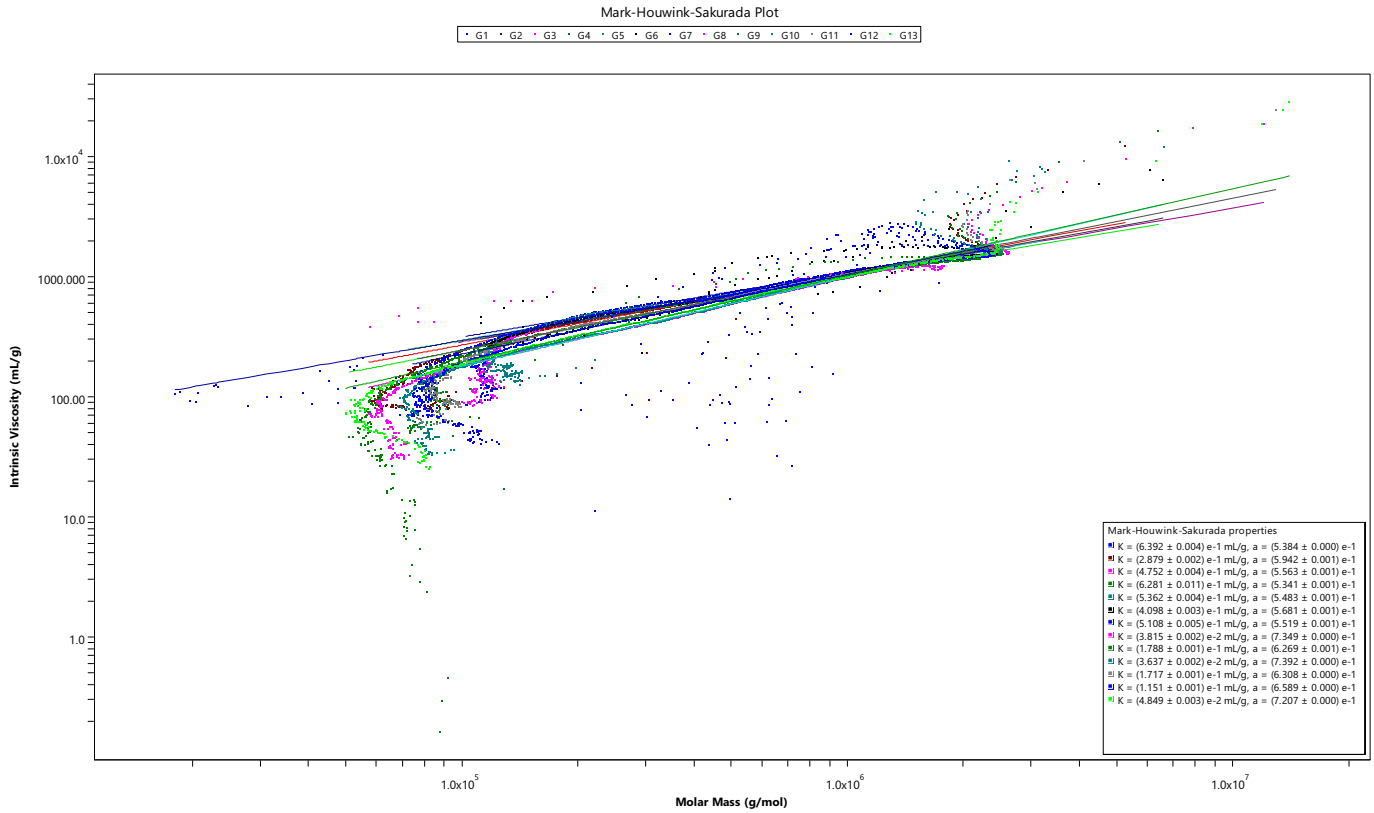


Fig. 11. Mark-Houwink-Sakurada Plot for 13 randomly chosen carrageenan samples extracted from different species of Gigartina genus.

References

Almeida, N., Hirschi, S., Mueller, A., & Rakesh, L. (2010). Viscoelastic properties of Delta k-carrageenan in saline solution [article]. *Journal of Thermal Analysis and Calorimetry, 102*(2), 647–652. https://doi.org/10.1007/s10973-010-0944-0

Berglund, A., & Wold, S. (1997). INLR, implicit nonlinear latent variable regression [article]. *Journal of Chemometrics, 11*(2), 141–156. https://doi.org/10.1002/(SICI)1099-128X(199703)11:2<141::AID-CEM461>3.0.CO;2-2

Campo, V. L., Kawano, D. F., da Silva, D. B., & Carvalho, I. (2009). Carrageenans: Biological properties, chemical modifications and structural analysis – A review. *Carbohydrate Polymers, 77*(2), 167–180. https://doi.org/10.1016/J.CARBPOL.2009.01.020

Cook, R. D., & Forzani, L. (2021). PLS regression algorithms in the presence of nonlinearity. *Chemometrics and Intelligent Laboratory Systems, 213*, Article 104307. https://doi.org/10.1016/J.CHEMOLAB.2021.104307

De Ruiter, G. A., & Rudolph, B. (1997). Carrageenan biotechnology. *Trends in Food Science and Technology, 8*(12), 389–395. https://doi.org/10.1016/S0924-2244(97)01091-1

Falshaw, R., Bixler, H. J., & Johndrob, K. (2001). Structure and performance of commercial kappa-2 carrageenan extracts: I. Structure analysis. *Food Hydrocolloids, 15*(4–6), 441–452. https://doi.org/10.1016/S0268-005X(01)00066-2

Hill, S. E., Ledward, D. A., & Mitchell, J. R. (1998). In S. E. Hill, D. A. Ledward, & J. R. Mitchell (Eds.), *Functional properties of food macromolecules* (2. ed.). Aspen Publishers.

Kennard, R. W., & Stone, L. A. (1969). Computer aided Design of Experiments [article]. *Technometrics, 11*(1), 137–148. https://doi.org/10.1080/00401706.1969.10490666

Marcelo, G., Saiz, E., & Tarazona, M. P. (2005). Unperturbed dimensions of Carrageenans in different salt solutions. *Biophysical Chemistry, 113*(3), 201–208. https://doi.org/10.1016/J.BPC.2004.09.005

Marcotte, M., Hoshahili, A. R. T., & Ramaswamy, H. S. (2001). Rheological properties of selected hydrocolloids as a function of concentration and temperature. *Food Research International, 34*(8), 695–703. https://doi.org/10.1016/S0963-9969(01)00091-6

National Research Council. (1981). Carrageenan Monograph. In *Food chemicals codex* (3rd ed., pp. 73–75). Washington, DC: The National Academies Press.

Nickerson, M. T., & Paulson, A. T. (2004). Rheological properties of gellan, κ-carrageenan and alginate polysaccharides: Effect of potassium and calcium ions on macrostructure assemblages. *Carbohydrate Polymers, 58*(1), 15–24. https://doi.org/10.1016/J.CARBPOL.2004.06.016

Nørgaard, L., Saudland, A., Wagner, J., Nielsen, J. P., Munck, L., & Engelsen, S. B. (2000). Interval partial least-squares regression (iPLS): A comparative Chemometric study with an example from near-infrared spectroscopy [article]. *Applied Spectroscopy, 54*(3), 413–419. https://doi.org/10.1366/0003702001949500

Pena-Pereira, F., Wojnowski, W., & Tobiszewski, M. (2020). AGREE analytical GREEnness metric approach and software [article]. *Analytical Chemistry (Washington), 92*(14), 10076–10082. https://doi.org/10.1021/acs.analchem.0c01887

Podzimek, S. (2011). *Light scattering, size exclusion chromatography, and asymmetric flow field flow fractionation powerful tools for the characterization of polymers, proteins, and nanoparticles [book]*. Wiley.

Podzimek, S. (2014). Truths and myths about the determination of molar mass distribution of synthetic and natural polymers by size exclusion chromatography [article]. *Journal of Applied Polymer Science, 131*(7), n/a. https://doi.org/10.1002/app.40111

Rao, M. A., & Kenny, J. F. (1975). Flow properties of selected food gums. *Canadian Institute of Food Science and Technology Journal, 8*(3), 142–148. https://doi.org/10.1016/S0315-5463(75)73766-5

Robal, M., Brenner, T., Matsukawa, S., Ogawa, H., Truus, K., Rudolph, B., & Tuvikene, R. (2017). Monocationic salts of carrageenans: Preparation and physico-chemical properties. *Food Hydrocolloids, 63*, 656–667. https://doi.org/10.1016/J.FOODHYD.2016.09.032

Sen, M., Yolacan, B., & Gueven, O. (2013). A comprehensive study on the size exclusion chromatography of kappa-carrageenan for the identification of after-peaks [article]. *Journal of Applied Polymer Science, 127*(1), 494–499. https://doi.org/10.1002/app.37758

Slootmaekers, D., van Dijk, J. A. P. P., Varkevisser, F. A., van Treslong, C. J. B., & Reynaers, H. (1991). Molecular characterisation of κ- and λ-carrageenan by gel permeation chromatography, light scattering, sedimentation analysis and osmometry [article]. *Biophysical Chemistry, 41*(1), 51–59. https://doi.org/10.1016/0301-4622(91)87209-N

Speers, R. A., & Tung, M. A. (1986). Concentration and temperature dependence of flow behavior of xanthan gum dispersions [article]. *Journal of Food Science, 51*(1), 96–98. https://doi.org/10.1111/j.1365-2621.1986.tb10844.x

Tye, R. J. (1988). The hydration of carrageenans in mixed electrolytes. *Food Hydrocolloids, 2*(1), 69–82. https://doi.org/10.1016/S0268-005X(88)80039-0

van de Velde, F. (2008). Structure and function of hybrid carrageenans. *Food Hydrocolloids, 22*(5), 727–734. https://doi.org/10.1016/J.FOODHYD.2007.05.013

Van De Velde, F., Antipova, A. S., Rollema, H. S., Burova, T. V., Grinberg, N. V., Pereira, L., … Grinberg, V. Y. (2005). The structure of κ/ι-hybrid carrageenans II. Coil–helix transition as a function of chain composition. *Carbohydrate Research, 340*(6), 1113–1129. https://doi.org/10.1016/J.CARRES.2005.02.015

van de Velde, F., & de Ruiter, G. A. (2002). Carrageenan. In E. J. Steinbüchel, A. De Baets, & S. Van Damme (Eds.), *Vol. 6. Biopolymers Vol. 6: Polysaccharides II: Polysaccharides from Eukaryotes* (pp. 245–274). Weinheim, Germany: Wiley-VCH.

Viscosity. (1989). *Rheology Series, 3*(C), 11–35. https://doi.org/10.1016/B978-0-444-87469-6.50006-8

Vreeman, H. J., Snoeren, T. H. M., & Payens, T. A. J. (1980). Physicochemical investigation of k-carrageenan in the random state [article]. *Biopolymers, 19*(7), 1357–1374. https://doi.org/10.1002/bip.1980.360190711

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems, 58*(2), 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1

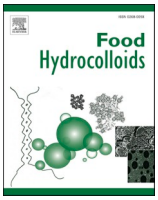
Wyatt Technology, L. L. C. (2023). ASTRA TM 8 software User's guide. https://wyatttechnology.zendesk.com/hc/en-us/article_attachments/24763817327507.

Food Hydrocolloids 158 (2025) 110544

Contents lists available at ScienceDirect

Food Hydrocolloids

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



Chemometric insights into milk-carrageenan breaking and gel strength

Oksana Mykhalevych [a,b,*], Henrik Stapelfeldt [a], Federico Marini [c], Rasmus Bro [b]

[a] CP Kelco ApS, 4623, Lille Skensved, Denmark
[b] Department of Food Science, University of Copenhagen, 1958, Frederiksberg, Denmark
[c] Department of Chemistry, University of Rome "La Sapienza", 00185, Rome, Italy

ARTICLE INFO

Keywords:
Carrageenan
Spectroscopy
Chemometrics
Data fusion
Breaking and gel strength
Milk-carrageenan gels

ABSTRACT

The relationship between the molecular structure and composition of carrageenan and its milk gelation properties has been studied using chemometric and machine learning tools. Carrageenan types—κ-carrageenan and κ/ι-hybrid— were analyzed for their gel and breaking strengths, crucial for their industrial application as food gelling and thickening agents. Four analytical platforms, Fourier-transform infrared spectroscopy (FT-IR), nuclear magnetic resonance (NMR) spectroscopy, size-exclusion chromatography with multi-angle light scattering detection (SEC-MALS), and inductively coupled plasma mass spectrometry (ICP-MS), were employed to characterize the molecular structure and composition aiming to predict milk-carrageenan breaking strength.

Both single and multi-block predictive modeling were applied to predict functionality, challenging the conventional approach that relies only on single analytical platforms for prediction.

Support Vector Machine (SVM) trained on FT-IR spectra, achieved the most accurate predictions, indicating its potential as an efficient alternative to traditional characterization methods by requiring only measurements directly on the carrageenan powder rather than the laborious functionality testing.

In examining multi-block modeling, particularly through Sequential and Orthogonalized PLS (SO-PLS), the study evaluated the added value of incorporating further analytical blocks. While adding SEC-MALS and ICP-MS data did not significantly improve prediction models, their inclusion enriched the causal understanding of carrageenan's structure-function relationship.

1. INTRODUCTION

Carrageenan, an essential hydrocolloid extracted from red seaweed (phylum *Rhodophyceae*), holds significant importance in the food industry due to its properties as a gelling and thickening agent (De Ruiter & Rudolph, 1997). Application of carrageenan to promote gelling of milk is one of the most important uses of this ingredient (Hurtado et al., 2017).

Carrageenan is a mixture of sulfated galactans composed of altering 3-linked β-D-galactopyranose and 4-linked α-D-galactopyranose or 4-linked 3,6-anhydrogalactose (De Ruiter & Rudolph, 1997). In this study, our focus is solely on κ-carrageenan, ι-carrageenan, and the carrageenan containing both κ- and ι-repeating units, termed κ/ι-hybrid carrageenan. These variations primarily differ in their sulphate content: κ-carrageenan is formed of alternating β-(1–3)-D-galactose 4-sulphate and α-(1–4) 3,6-anhydro-D-galactose units, while ι-carrageenan consist of alternating β-(1–3)-D-galactose 4-sulphate and α-(1–4)-3,6-anhydro-D-galactose-2-sulphate units. The κ-carrageenan is derived from *Kappaphycus alvarezii* (previously known as *Eucheuma cottonii*, or Cottonii), whereas *Chondrus* genus results in κ/ι-hybrid carrageenan (De Ruiter & Rudolph, 1997).

The molecular composition and structure of these carrageenan types significantly influence their functional properties in food systems, particularly their ability to form gels with milk proteins. The formation of milk-carrageenan gel occurs through the electrostatic interaction between carrageenans and milk proteins (Anderson et al., 2002; Langendorff et al., 2000; Lynch & Mulvihill, 1996). There are distinct differences in the gel-forming mechanism of κ-carrageenan and ι-carrageenan. Kappa-types form firm and brittle gels, iota types form soft and elastic gels, while gel formation mechanism for the hybrid types depends on the amount of iota residues present (Baeza et al., 2002; BeMiller, 2019, pp. 279–291; Bixler et al., 2001). Furthermore, the milk gelling mechanism with carrageenan is influenced by several parameters such as the molecular structure, concentration, conformation, the ionic environment, and the presence of other ingredients in the matrix (Drohan et al., 1997; Elfaruk et al., 2021; Hilliou, 2021; Jiang et al., 2021; Langendorff et al., 1999; Puvanenthiran et al., 2003; Sabadini

* Corresponding author. CP Kelco ApS, 4623 Lille Skensved, Denmark.
E-mail address: oksana@food.ku.dk (O. Mykhalevych).

https://doi.org/10.1016/j.foodhyd.2024.110544
Received 14 April 2024; Received in revised form 3 August 2024; Accepted 19 August 2024
Available online 22 August 2024

Abbreviations

[η](avg)	Uncertainty-weighted average intrinsic viscosity
[η]w	Weight average intrinsic viscosity
COW	Correlation Optimized Warping
dn/dc	Specific refractive index increment
DP	Differential pressure
FT-IR	Fourier-transform infrared spectroscopy
HPLC	High-Performance Liquid Chromatography
ICP-MS	Inductively coupled plasma mass spectrometry
LV	Latent Variable
LWR	Locally Weighted Regression
M(avg)	Uncertainty-weighted average molar mass
M_n	Number average molecular mass
M_p	Molar mass at the top of the concentration peak
M_w	Weight-average molecular mass
M_z	Z-average molar mass
NMR	Nuclear magnetic resonance
PCA	Principal component analysis
PLS	Partial Least Squares
RBF	Radial Basis Function
rh(v)(avg)	Uncertainty weighted average hydrodynamic radius
$rh(v)_w$	Weight average hydrodynamic radius
RMSE	Root Mean Square Error
RMSECV	Root Mean Square Error of Cross-Validation
RMSEP	Root Mean Square Error of Prediction
r_w	Weight average mean square radius
SEC-MALS	Size-exclusion chromatography with multi-angle light scattering detection
SNV	Standard Normal Variate scaling
SO-PLS	Sequential and Orthogonalized PLS
SVM	Support Vector Machines
UHT	Ultra-high-temperature

et al., 2006; Sedlmeyer et al., 2003).

In this study, it is hypothesized that all molecular characteristics that define the functionality of carrageenan as a milk gelling agent can be directly and/or indirectly measured by inductively coupled plasma mass spectrometry (ICP-MS), nuclear magnetic resonance (NMR) spectroscopy, Fourier-transform infrared spectroscopy (FT-IR), and size-exclusion chromatography with multi-angle light scattering detection (SEC-MALS). To distinguish between the two main carrageenan types, FT-IR spectroscopy is often the method of choice (Chopin & Whalen, 1993; Dyrby et al., 2004; Jacobsson & Hagman, 1993; Prado-Fernández et al., 2003; Tojo & Prado, 2003b; Černá et al., 2003). Identification of the types of carrageenans is possible due to signals corresponding to stretching vibrations of the sulphate esters and the 3,6-anhydro ring that give identifiable signals for each carrageenan type (Chopin & Whalen, 1993). ^{1}H NMR spectroscopy has also been applied with the same purpose to identify carrageenan type and for detailed elucidation of carrageenan structure (Rodríguez Sánchez et al., 2022; Tojo & Prado, 2003a; Van De Velde et al., 2002, 2004). However, ^{1}H NMR spectra can also be used to extract more detailed information about fractions of κ- and ι-carrageenan in a hybrid carrageenan using chemical shift and intensity of signal given by the anomeric protons (van de Velde, 2008; Van De Velde et al., 2001). The cation composition of carrageenan has previously been analyzed by flame ionization spectroscopy but is now mainly analyzed by inductively coupled plasma–atomic emission spectrometry (ICP–AES or ICP-MS) (Van De Velde et al., 2005). SEC-MALS in conjunction with a differential refractometer and viscometry detector is used for determination of parameters such as molecular weight, polydispersity index (PDI), concentration, root-mean-square radius of gyration, intrinsic viscosity [η], and other parameters (Agoda-Tandjawa et al., 2017; Berth et al., 2008; Michna et al., 2021; Robal et al., 2017; Slootmaekers et al., 1991).

Gelling properties are typically characterized using texture analyzers from which gel and breaking strength give valuable information for the manufacturers. However, this methodology requires laborious sample preparation and may be affected by seasonal variations in the compositions of other components, e.g., milk included in the gel.

Even though previous studies considering the structure of carrageenan include data from several analytical platforms for carrageenan samples, only single block chemometric methods were used for predictive modelling and were not applied for prediction of carrageenan functionality but rather for analysis of its structure (Dyrby et al., 2004). Data fusion methods can show improved predictive performance and enhanced information extraction compared to the traditional single block chemometric methods (Mishra et al., 2021). In this study, we will investigate the added potential of combining data sources using data fusion.

The purpose of this work is to determine the relationship between instrumentally measured chemical structure and milk gelation properties of carrageenan by applying advanced chemometric and machine learning tools. At a first level we will investigate if we can *predict* functional parameters of carrageenan as a milk gelling agent. If a predictive model can be obtained, such a model may also provide leads to causal relations that can help explain *why* a certain functionality is achieved and this can be important for designing products of a specified quality. We will pursue the use of so-called data fusion or multi-block (a block here defined as a dataset obtained from one data source or one analytical platform) methods and single-block methods for building the predictive models. Such a predictive model based on molecular characteristics is expected to be much faster than classical functionality tests, to be free from matrix variations, and more easily adjusted for changes of molecular structure of the hydrocolloid caused by deviations in production processes or quality of raw material.

2. Materials and methods

2.1. Samples

Carrageenan samples were obtained from CP Kelco (Lille Skensved, Denmark). The samples were chosen to cover a range of values for the distance to break (breaking strength) and gel breaking strength of milk gels made with carrageenan and characterized by a texture analyzer. Additionally, the samples were collected over approximately two years to account for seasonal variations in the raw material.

The sample domain comprised sixty-seven κ-carrageenan samples derived from *Kappaphycus alvarezii* (Cottonii), and seventy-two κ/ι-hybrid carrageenans extracted from *Chondrus* genus.

All chemicals utilized in this study were of analytical grade. Exclusively, ion-exchanged water was employed throughout the experiments.

2.2. Data collection

2.2.1. Breaking and gel strength

A milk gel system was prepared by dissolving a blend of carrageenan and sucrose in commercial ultra-high-temperature (UHT) processed milk (1.5 % fat (Arla Foods amba, Viby J, Denmark) under stirring at 80 °C for 10 min. The solution was then transferred to bloom glasses (70 × 40 mm, diameter × height) equipped with a tape to allow 1 cm above the brim. The samples were left to cool down for 2.5–3 h in a 5 °C water bath from which they were removed just prior to measurement. Before measuring on a TA.XTplus Texture Analyzer (Stable Micro Systems,

Godalming, UK), 1 inch plunger P/1R and speed 1.0 mm/s, the top of the gel was cut with a cheese cutter to provide an even penetration surface and a well-defined gel volume.

For each set of experiments, a control sample was included to monitor any batch-to-batch variations of UHT milk and their potential impacts were adjusted accordingly.

The gel strength was measured as the force (in g) at 2 mm, while the breaking strength was defined as the force (in g) at rupture. The gels were measured in triplicates and the average was used as final value. Fig. 1 illustrates characteristic texture analyzer curves for carrageenans extracted from the two raw material types.

Gel and breaking strengths indexes were used in this study as reference values. These were corrected to a carrageenan mass used for the same milk system, so, the reference values are independent of the carrageenan mass. This conversion was performed to correct differences between different carrageenan types included in this study since different carrageenan types require different concentration to produce a milk-carrageenan gel.

2.2.2. FT-IR

The FT-IR spectra of carrageenan powder samples were scanned using FT-IR MB3000 with ATR, ZnSe optical crystal (ABB Inc., Quebec, Canada). A total 16 scans were averaged for each carrageenan sample and the nominal resolution was 4 cm^{-1}. The spectra were recorded in the 4000-370 cm^{-1} region. The background was recorded at the beginning of the measurement series by exposing the crystal to air.

2.2.3. ICP-MS

Carrageenan samples (25 mg carrageenan/5.0 mL 65% HNO_3 and 1 mL 37 HCl) were solubilized by closed-vessel microwave-assisted acid digestion using a Milestone Ethos Up microwave oven (Milestone Srl, Sorisole, Italy). The solution (1 mL) was diluted in double-distilled water (9 mL). The samples were analyzed using Agilent 7800 ICP-MS (Agilent Technologies Inc., Santa Clara, USA) including autosampler. Calibration was performed by using a certified multi-element solution containing Na, Mg, K, Ca and S standards (Inorganic Ventures, Christiansburg, United States).

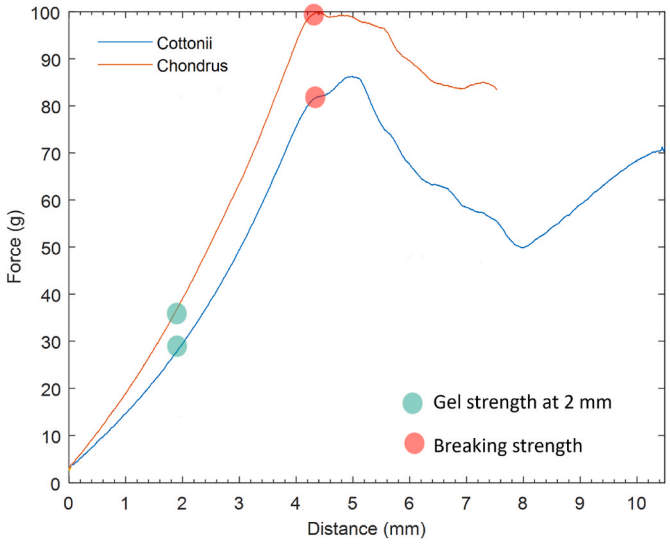


Fig. 1. Typical texture analyzer curves, Force (g) vs distance (mm), for milk gels prepared with carrageenan extracted from Kappaphycus alvarezii (blue line labeled "Cottonii") and Chondrus (orange line labeled "Chondrus") seaweed types. Gel strength is measured as force at 2 mm, offering insights into gels stability. Breaking strength is a measure of the firmness of the gel and is read at rupture as indicated.

2.2.4. NMR spectroscopy

The carrageenan samples were prepared as 5% (w/w) solutions in D_2O using a thermomixer Eppendorf™ F2 Model ThermoMixer™ (Eppendorf AG, Germany) at 80 °C and 1500 rpm. The solution was transferred to 5 mm NMR tubes. The ^1H NMR were obtained at Agilent Technologies 600 MHz spectrometer operating at 599.776 MHz equipped with 5 mm PFG OneNMR 7450 Probe (Agilent Technologies Inc., Santa Clara, USA). The spectra were acquired at 80 °C, with a 30° pulse sequence (pulse 90° 8.1 μs), with 5 s of relaxation delay and 128 scans, and 54 gain (dB). The spectra were recorded at 80 °C to obtain low viscosity solutions to avoid peak broadening. The choice of relaxation delay was based on experimentally determined T1 values that for the resonances of carrageenans were smaller than 3 s. A pulse experiment with water suppression was used. Total experiment time was 14.48 min. The spectra were referenced using the ^2H absorption frequency of the solvent. The spectrometer was controlled by OpenVnmrJ 3.1A software.

2.2.5. SEC-MALS

SEC-MALS analysis was performed using an HPLC instrument Agilent 1260 Infinity II (Agilent Technologies Inc., Santa Clara, USA) with Shodex OHpak LB-806M (8.0 × 300 mm, ID x length) column (Resonac Europe GmbH, Wiesbaden, Germany) and Shodex OHpak SB-G 6B (6.0 × 50 mm, ID × length) guard column (Resonac Europe GmbH, Wiesbaden, Germany). The HPLC was equipped with MALS detector, model DAWN 8, operating at wavelength 660 nm (Wyatt Technology, Santa Barbara, USA), viscometer ViscoStar (Wyatt Technology, Santa Barbara, USA), and differential refractometer - Optilab, operating at 658 nm wavelength (Wyatt Technology, Santa Barbara, USA). Elution was carried out using 0.1 M $LiNO_3$ at a constant flow rate of 1 mL/min at system temperature of 40 °C. The specific refractive index increment (dn/dc) was 0.11 for all measurements.

Samples were prepared by dissolving carrageenan in the eluent at concentration 2 mg/mL, and heating to 80 °C under stirring for 30 min. The solution was allowed to cool down and 30 μL was injected into the column. The total measurement time was 32 min.

Data analysis was performed using ASTRA (Wyatt Technology Corp.). The Zimm fit method (Podzimek, 2011, p. 220) was applied in data analysis for all samples.

A very detailed description of SEC-MALS and size characteristics may be found in a book by S. Podzimek (2011) (Podzimek, 2011) and accordingly not specified further.

2.3. Data analysis

In total, 139 carrageenan samples were measured on the four analytical platforms and the texture analyzer reference method. The average across duplicates measured on each analytical platform was used in the study. To identify and understand outliers, principal component analysis (PCA) (Bro & Smilde, 2014) in combination with fundamental knowledge of carrageenan structure were used. All five datasets were divided into calibration and test datasets: 112 samples in the calibration dataset and 27 in the test set. The test subset was selected by the Kennard-Stone algorithm (Kennard & Stone, 1969).

The ICP-MS, SEC-MALS, and the reference data were autoscaled so that each variable has a mean of zero and a standard deviation of one.

The FT-IR spectra were preprocessed by Standard Normal Variate transformation (SNV) (Rinnan et al., 2009) to correct for scatter effects and other sources of spectral variation not related to chemical information.

The NMR spectra were aligned using Correlation Optimized Warping (COW) algorithm (Nielsen et al., 1998). The spectra were aligned to the mean spectrum with the maximum shift (slack) of 16 and the segment length of 150 defined by an automated COW algorithm (Skov et al., 2006). After COW alignment, the spectra were preprocessed using SNV.

Other preprocessing methods were tested but the above-mentioned were chosen as preprocessing techniques for these datasets. Both the

NMR and FT-IR spectra were mean centered before analysis.

Both single and multi-block predictive modeling techniques were applied to investigate the gel and breaking strength of carrageenan-milk gels. Initially, single block models were developed using each data block independently. This approach was aimed at evaluating the predictive capabilities of individual analytical platforms, with the intent to gain insights into the predictive relationship between the chemical structure of carrageenan and its functionality. The analysis of each data block independently provided a basis for understanding the contribution of each analytical method to the overall prediction of carrageenans structure-function relationship. Predictive models were developed for three data classes based on the seaweed source of carrageenan: global model, included data for both carrageenan types, cottonii model – included *Kappaphycus alvarezii (*Cottonii*)* carrageenan samples data, and chondrus model was trained on data for carrageenan samples extracted from *Chondrus* genus.

Partial least squares (PLS) regression (Wold et al., 2001) was used as a linear regression model for predicting functionality. Several non-linear alternatives were also investigated; namely, locally weighted regression (LWR) (Bakshi & Chen, 2009) and support vector machines (SVM) with a radial basis function kernel (Chen, 2004). The average prediction error, Root Mean Square Error (RMSE), was used as a metric to evaluate model performance, and it was either calculated in cross-validation (RMSECV) or determined from a complete pristine prediction set of data (RMSEP).

The LWR models were optimized by investigating different numbers of local training points (5, 10, 15 … 35). The model with the lowest RMSECV was chosen as the optimal.

The optimal number of components for PLS and LWR was selected according to the lowest RMSECV by performing cross-validation with ten cancelation groups.

Cross-validation with ten cancelation groups was also employed in the training of SVM models to avoid overfitting.

Subsequently, multi-block models were explored, integrating data from multiple sources to assess the potential synergistic effects of combining data from different analytical platforms on the predictive power.

Several data fusion techniques were tested (multiblock PLS regression, parallel and orthogonalized PLS regression and Response Oriented Sequential Alteration (Smilde, 2022)), among which sequential and orthogonalized PLS regression (SO-PLS) (Næs et al., 2013) was selected for detailed exploration in this study.

In SO-PLS, the method extracts information from each data block one by one, starting from the first. Initially, the first data block is analyzed using a standard PLS approach. Then, each following block is orthogonalized to the scores of the previous one and then fitted to the residuals of the reference variable. In that way, any information that was already covered by the previous block(s) is removed, ensuring that only new, unique information from each subsequent block contribute to the model. The order in which these blocks are processed can affect the outcome of the final prediction (Biancolillo & Næs, 2019).

The optimal SO-PLS model was determined by training a series of models, each varying in the number of PLS components included from each data block. The goal was to find the combination that resulted in the lowest prediction error, indicating the best predictive performance. This requires establishing an upper number of PLS components to be considered from each data block. For the ICP and SEC blocks, the maximal number of components was equivalent to the number of variables in each data block, 5 and 12, respectively. For FT-IR and NMR spectra, a PCA was applied to define the maximum number of components that could be relevant. The maximum number of components was chosen to be nine for the FT-IR block, and ten for the NMR block.

All variables from both SEC and ICP blocks were employed in predictive modeling. It is noteworthy that parameters extracted from SEC exhibit significant correlation. However, due to limited understanding of the relationships between these data and functional properties of carrageenan, the decision was made to include all variables in the data analysis.

To investigate the effect of the block order in SO-PLS method, models for all combinations of block orders and block numbers were developed and analyzed.

The optimal number of components for SO-PLS was selected according to the lowest RMSECV by performing cross-validation with ten splits.

2.4. software for data analysis

The data analysis was performed using MATLAB (Release R 2023a, version: 9.14.0; The MathWorks Inc.) and PLS_Toolbox 9.2 (2023). Eigenvector Research, Inc. The code for SO-PLS was obtained from Rome Chemometrics (Marini, 2023).

3. Results and discussion

3.1. Data blocks

Table 1 presents the descriptive statistics of the gel and breaking strength index for each class (seaweed source) used as reference values in the development of the models. The reference values are characterized by their mean, standard deviation (std), minimum (min), and maximum (max). The gel and breaking strength indexes were measured in range 90–223 and 93–295, respectively. The mean values for *Kappaphycus alvarezii* carrageenan are higher than for *Chondrus* carrageenan, which may be explained by the molecular structure of these two carrageenan types: carrageenan extracted from *Kappaphycus alvarezii* seaweed exhibit κ-carrageenan molecular structure, while κ/ι-hybrid carrageenans were extracted from *Chondrus* genus. Kappa-types form firm and brittle gels, while iota types form soft and elastic gels (BeMiller, 2019, pp. 279–291) and the higher observed reference values for κ-carrageenan compared to these for hybrid carrageenan is thus expected.

Table 2 summarizes the descriptive statistic for the ICP-block. The statistical measures in the table offer insights into trends of carrageenans ionic composition that is influenced by the ionic binding between the negatively charged ester sulphate groups and cations. The sulfur content, which is directly proportional to the sulphate concentration, exhibits the higher mean percentage (8.0%) in carrageenans extracted from seaweed species within the *Chondrus* genus compared to *Kappaphycus alvarezii* carrageenans (6.8%). Considering the molecular composition, ι-carrageenan contains two sulfur groups per disaccharide unit, whereas κ-carrageenan comprises one sulfur group per disaccharide unit. This difference indicates a higher proportion of iota-carrageenan units within the κ/ι-hybrid carrageenan from *Chondrus* genus species compared to Cottonii carrageenan that are pure

Table 1
Descriptive statistics of the reference data, where Cottonii data includes measurements of carrageenan samples from Kappaphycus alvarezii, and Chondrus data includes measurements of carrageenan samples extracted from the Chondrus genus.

	GEL STRENGTH (index)	BREAKING STRENGTH (index)
Chondrus		
Mean	146	143
Std	22	25
Min	90	93
Max	198	217
Cottonii		
Mean	180	177
Std	19	46
Min	123	99
Max	223	295

Table 2
Ionic composition of carrageenan samples obtained from ICP-MS, where Cottonii data includes measurements of carrageenan samples from Kappaphycus alvarezii, and Chondrus data includes measurements of carrageenan samples extracted from the Chondrus genus.

	Na (mg/g)	Mg (mg/g)	S (%)	K (mg/g)	Ca (mg/g)
Chondrus					
Mean	15.3	0.45	8.0	21.5	35.7
Std	2.5	0.43	0.3	3.4	3.5
Min	9.1	0.13	7.4	13.1	29.0
Max	21.8	3.38	8.7	27.1	47.0
Cottonii					
Mean	7.2	0.50	6.8	43.3	22.1
Std	1.5	0.36	0.6	5.6	2.8
Min	3.8	0.05	3.8	22.7	11.6
Max	10.6	1.43	7.3	51.7	27.3

κ-carrageenans.

The K^+, Ca^{2+}, Na^+ ion concentrations shown in Table 2 are highly influenced by the manufacturing process, since potassium and calcium salts of the sulfated ester group of carrageenan are favorable for its gelation. Potassium concentration is highest for *Kappaphycus alvarezii* carrageenan suggesting potassium form of the polysaccharide, whereas calcium ions are predominant for the other carrageenan type, suggesting a calcium form of carrageenan. These two ionic forms of carrageenan, i.e., potassium for of κ-carrageenan and calcium form of κ/ι-carrageenan form the strongest gels (BeMiller, 2019, pp. 279–291). The concentration of magnesium is very low in both carrageenan types and its variation is mainly due to different natural occurrence.

The wavelength region showing carrageenans FT-IR fingerprint at 1300-700 cm^{-1} is shown in Fig. 2. A very detailed assignment and review of FT-IR spectra for carrageenan was published by Prado-Fernandez et al. (2003) (Prado-Fernández et al., 2003). The spectra and the absorption bands presented in Fig. 2 correspond to those earlier reported FT-IR spectra of carrageenan, however with some signals slightly shifted most likely due to presence of counter ions. The relevant absorption bands are discussed later when relevant for prediction of the gel and breaking strength.

The preprocessed NMR spectra of all samples are shown in Fig. 3. The chemical shifts observed are slightly different from those previously reported (Ciancia et al., 2020; Tojo & Prado, 2003a; Van De Velde et al., 2004) because the measurements were performed at higher temperature with a different referencing method. The chemical shifts of the anomeric protons of κ- and ι-carrageenans were 5.53 and 5.73 ppm, respectively.

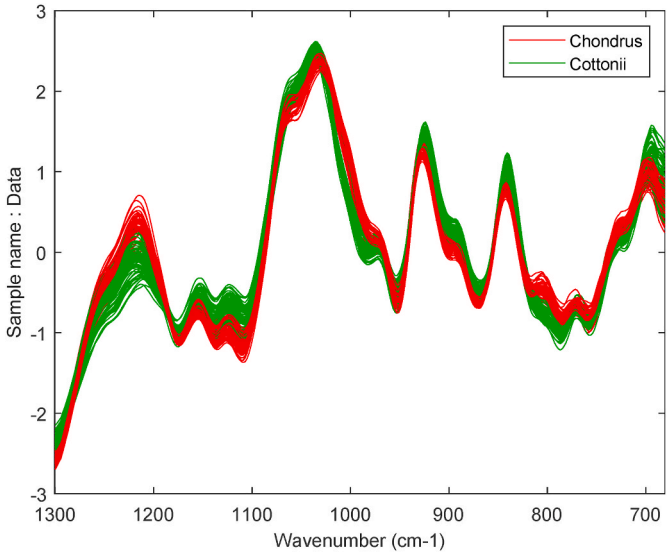


Fig. 2. SNV filtered IR spectra of carrageenan samples measured in replicates. Carrageenan were extracted from seaweeds belonging to Chondrus crispus (red spectra), or Kappaphycus alvarezii (green spectra (Cottonii)) species.

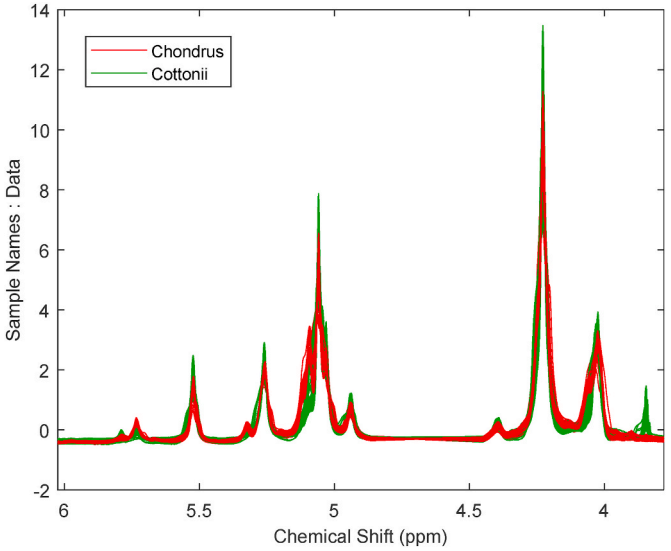


Fig. 3. Aligned and SNV filtered ^1H NMR spectra of carrageenan samples measured in replicates. Carrageenan was extracted from seaweeds belonging to Chondrus crispus (red spectra), or Kappaphycus alvarezii (green spectra) genera.

The size exclusion chromatogram of a carrageenan sample is shown in Fig. 4. The light scattering (LS) signal obtained by the MALS detector represents molar masses eluting form SEC column during elution time, the dRi (blue signal) represents the concentrations of the molecules eluting, while the differential pressure (DP) curve is the specific viscosity (η_{sp}) chromatogram measured by the viscometer. The defined peak at 5.90–10.20 min corresponds to carrageenan signal. The signals at 10.20–14.30 min correspond to ionic impurities (Sen et al., 2013), while the broad peaks at 20.0–32.0 min are so-called breakthrough peaks due to a sample passing the delay column.

Table 3 summarizes statistics for the respective molecular parameters obtained by SEC-MALS measurements. In total, twelve variables estimated using the data from SEC-MALS were included in this study: number average molecular mass (M_n), molar mass at the top of the concentration peak (M_p), weight average molecular mass (M_w), Z-average molar mass (M_z), uncertainty-weighted average molar mass (*M (avg)*), polydispersity (M_w/M_n), weight average mean square radius (r_w), weight average hydrodynamic radius ($rh(v)_w$), uncertainty weighted average hydrodynamic radius (*rh(v)avg*), mass recovery, weight average intrinsic viscosity ($[\eta]_w$), uncertainty-weighted average intrinsic viscosity (*[η](avg)*). PCA of the SEC-block data (not shown) revealed, as expected, correlation between several parameters since many are closely related. The mean weight-average molecular mass value of 909 kDa for carrageenan samples extracted from *Chondrus* genus species is significantly than that of *Kappaphycus alvarezii* carrageenans, which is 595 kDa. The average molecular mass of carrageenan samples from *Chondrus* genus species, at 909 kDa, is notably higher than the 595 kDa of *Kappaphycus alvarezii* carrageenans. This disparity stems from the inherent polymer sizes and molecular distinctions, such as κ/ι-hybrid carrageenan possessing an extra sulphate ester group relative to ι-carrageenan. Additionally, the molecular mass influences other metrics measured by SEC-MALS, such as the weight average mean square radius and intrinsic viscosity, which are subsequently lower for *Kappaphycus alvarezii* carrageenan.

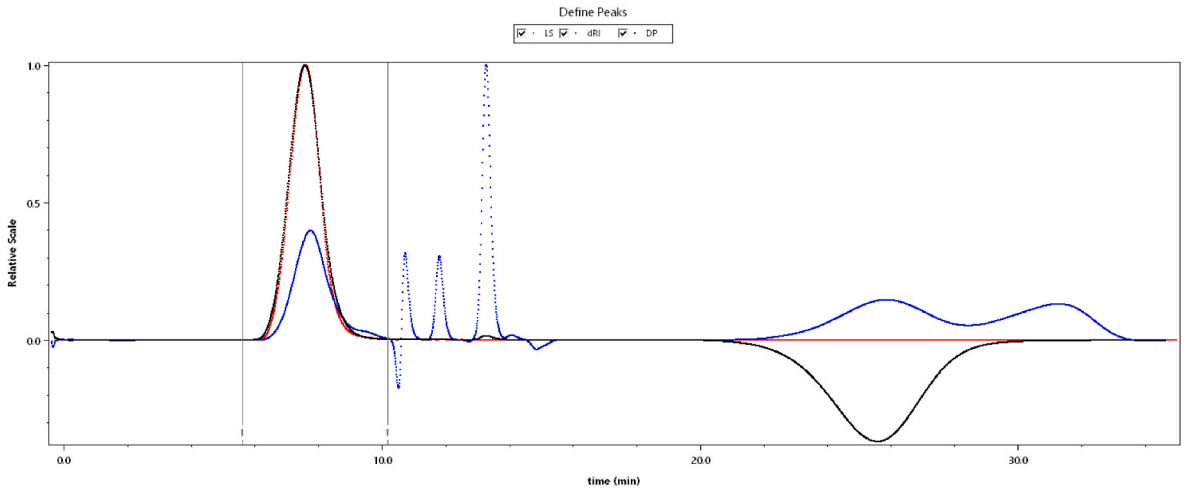


Fig. 4. Typical chromatograms of carrageenan form SEC coupled to MALS (red signal, LS), viscometer (black, DP) and a dRI (blue) detectors.

Table 3
Data obtained from SEC-MALS for the two carrageenan types, where Cottonii data includes measurements of carrageenan samples from Kappaphycus alvarezii, and Chondrus data includes measurements of carrageenan samples extracted from the Chondrus genus.

	M_n (kDa)	M_p (kDa)	M_w (kDa)	M_z (kDa)	M(avg) (kDa)	Polydispersity (Mw/Mn)	r_w (nm)	$rh(v)_w$ (nm)	rh(v)(avg) (nm)	Mass recovery (%)	$[\eta]_w$ (mL/g)	[η](avg) (mL/g)
Chondrus												
Mean	516.8	786.9	909.1	1312.6	554.7	1.8	95.0	49.1	47.6	87.0	925.8	834.5
Std	60.3	68.5	68.8	87.0	88.0	0.1	5.1	2.4	3.4	4.1	65.9	74.6
Min	369.1	622.6	738.7	1122.1	373.1	1.6	81.5	42.6	39.0	69.6	759.4	614.3
Max	646.1	913.8	1032.1	1564.6	774.8	2.0	104.1	53.3	53.6	97.4	1040.6	934.6
Cottonii												
Mean	306.5	546.9	594.9	911.6	271.5	2.0	69.4	37.8	30.5	92.3	662.2	570.2
Std	63.8	48.5	50.6	71.8	61.6	0.3	4.0	1.7	2.3	2.4	39.4	40.3
Min	197.3	441.3	486.1	751.3	150.1	1.5	60.0	34.0	26.3	85.2	565.3	455.9
Max	464.5	666.6	763.9	1191.5	425.2	2.9	81.4	43.9	39.8	97.6	804.7	680.4

3.2. single block models

Several models were developed to predict the gel and breaking strength of carrageenan-milk gels. The performance of these models, summarized in Table 4, was evaluated by training individual models on separate data blocks and assessing their performance using the RMSECV and the RMSEP. The analysis identified that models trained on FT-IR spectra—both linear and nonlinear—outperformed others. Hence, Table 4 focuses solely on FT-IR-based models for a comprehensive comparison with models that incorporate data from multiple blocks.

Comparing the RMSECV and RMSEP for gel and breaking strength, it is observed that both figures of merit are higher for the breaking strength index predictions. This could be attributed to an analytical error and the broader target range associated with the breaking strength index (Table 1). As shown in Fig. 1, the breaking strength index is determined using the point of gel rupture as defined by an operator and this manual assessment likely contributes to the error. Given the estimated reproducibility of the reference method (data not shown), which is approximately 5% for gel strength and 8% for breaking strength, a higher prediction error for breaking strength prediction was thus anticipated because the prediction error reflects the combined error stemming from the chemometric model *and* the reference method.

The RMSEP for the gel strength ranges between 11 and 18. Notably, the SVM model consistently delivers the lowest prediction error among all models: 11 for cottonii model, 13 for chondrus model and 12 for the global model. The linear approach, namely PLS, resulted in slightly lower performance: 12 for cottonii model, 14 for chondrus model and 15 for the global model. Among, the models developed for separate carrageenan types, the LWR model resulted in the highest RMSEP of 15 for cottonii, and 18 for chondrus model. For the global model, which encompasses both classes, RMSEP for LWR was 14.

For the breaking strength, the RMSEP is between 23 and 29. For models trained on FT-IR spectra for Cottonii carrageenans and for both carrageenan types together, the SVM model achieved the lowest prediction error of 24 and 25, respectively. Among the single-block models trained on FT-IR spectra of *Chondrus* carrageenans, the PLS model gave the lowest RMSEP value of 23. The worst performing single-block model for the breaking strength prediction was the LWR that resulted in the highest RMSEP for all three model classes.

The superior performance of the SVM models suggests that the underlying relationship between variables may be somewhat non-linear. However, the observed differences in prediction uncertainty are relatively modest. Moreover, the interpretability of SVM models is considerably low. Given that the objective of this study extends beyond achieving optimal prediction accuracy to also elucidating the structure-functionality relationship of carrageenan in milk applications, the limited explainability of SVM poses a significant drawback.

Consequently, the PLS model was chosen to elucidate the relationship between the molecular information captured by FT-IR and the gel and breaking strength of milk-carrageenan gels.

Fig. 5 shows loading vectors for all PLS models. The absorption bands at approximately 1110 cm^{-1}, 930 cm^{-1}, 840 cm^{-1}, 1230 cm^{-1}, and 1000 cm^{-1}, respectively, are important for both global models indicated by the high values in the loading vectors. Based on the score plots shown in Fig. 6, the mentioned absorption bands separate κ-carrageenans from hybrid carrageenan. In the PLS model score plots in Fig. 6, the scores

Table 4
Model performance summary for predicting gel and breaking strength in milk-carrageenan gels. The global model includes data for both carrageenan types. The cottonii model includes carrageenan samples from Kappaphycus alvarezii (Cottonii), and the chondrus model is trained on data from carrageenan samples extracted from the Chondrus genus. The number of PLS components stated in the table, e.g., [3IR 1ICP], indicates the number of PLS components from each data block included in the selected model, i.e., three PLS components from IR and one component from ICP.

Data Type	Model	PLS components number	RMSEC	RMSECV	RMSEP	R2 CV	R2 Pred
Gel Strength							
Global	PLS (IR)	3	12	13	15	0.80	0.67
	LWR (IR, 30 loc. points)	4	8	13	14	0.77	0.72
	SVM (IR)		7	12	12	0.79	0.76
	SO-PLS	[3IR 1ICP]	10	12	14	0.81	0.69
Chondrus	PLS	2	10	11	14	0.70	0.65
	LWR (15)	1	13	15	18	0.50	0.50
	SVM (IR)		9	12	13	0.68	0.70
	SO-PLS	[1ICP 2IR]	9	11	14	0.73	0.63
Cottonii	PLS	3	9	11	12	0.70	0.48
	LWR(IR, 20 loc. points)	4	5	12	15	0.62	0.46
	SVM (IR)		8	11	11	0.70	0.60
	SO-PLS	[4 IR]	9	10	12	0.72	0.51
Breaking Strength							
Global	PLS	4	22	24	26	0.65	0.52
	LWR (IR, 30 loc. points)	4	15	24	27	0.67	0.52
	SVM (IR)		15	22	25	0.72	0.55
	SO-PLS	[5IR 1SEC]	20	23	27	0.75	0.49
Chondrus	PLS	3	14	18	23	0.411	0.438
	LWR (IR, 15 loc. points)	1	17	20	27	0.245	0.413
	SVM (IR)		13	16	25	0.49	0.4
	SO-PLS	[5IR 1ICP]	11	14	24	0.61	0.23
Cottonii	PLS	3	20	24	25	0.75	0.57
	LWR(IR, 20 loc. points)	5	9	29	29	0.67	0.63
	SVM (IR)		15	27	24	0.68	0.58
	SO-PLS	[3 IR]	20	24	24	0.75	0.57

were colored based on the gel and breaking strength indices. A positive correlation between the scores of the first PLS component and the indices indicated that an increase in the first component's scores corresponds to higher gel and breaking strength. For the second PLS component, a similar trend is noted predominantly for breaking strength, where higher scores are associated with greater breaking strength values.

The absorption band around 1230 cm^{-1}, likely representing ester groups, results in higher intensity for κ/ι-carrageenans compared to the pure κ-carrageenan due to presence of an additional sulphate ester group on the iota unit as its intensity increases with the number of substitutions by sulphate groups (one for kappa to two for iota). This signal correlates with the absorption band around 805 cm^{-1} due to the sulphate group at the 3,6-anhydrogalactose-2-sulphate in ι-carrageenan (Prado-Fernández et al., 2003). Based on the score plots and the loading vectors, it is plausible to hypothesize that the amount of the sulphate groups (ι-carrageenan %) affects both reference values. A lower concentration of ι-carrageenan is indicative of increased gel and breaking. This suggests a complex interaction between ι-carrageenan concentration and the mechanical properties of the gel, highlighting the nuanced role of ι-carrageenan in modulating both gel and breaking strength. I-types carrageenans are known for forming soft and elastic gels, while κ-carrageenans form firm and brittle gel. The elasticity, provided by ι-carrageenan, allows the carrageenan-milk gel to stretch and compress without immediately fracturing, i.e., giving a higher breaking strength value at lower gel strength values (but not at low values of gel strength, since gel and breaking strength indexes are positively correlated).

The spectral range around 1100-1000 cm^{-1} is typical for polysaccharides and corresponds to glycosidic bonds (Prado-Fernández et al., 2003). However, it was suggested by Wilson et al., 1988, that the intensity of the S-O symmetric stretch observed at 1090 cm^{-1} might be increased by ion binding, while the symmetric stretch in spectra K^+ ion form of kappa-carrageen is observed as a shoulder (Wilson et al., 1988). Therefore, the observed effect of the functional group giving rise to the absorption band at 1110 cm^{-1} is suggested to originate from the interaction between K^+ and kappa-carrageen. Potassium carrageenan form firm and brittle gels. Based on the score and loading plots in Figs. 6 and 5, the gel strength value increase with increasing intensity of the signal at 1110 cm^{-1} while the breaking strength would slightly decrease. Potassium carrageenan form strong but more brittle gels resulting in a higher gel strength and a lower breaking strength.

The peaks at approximately 840 cm^{-1} and 930 cm^{-1} that have high values in the loading vector for the first PLS components were assigned to galactose 4-sulphate and C-O-C structure in 3,6-anhydrogalactose, respectively, present in both carrageenan types (Prado-Fernández et al., 2003). However, as may be seen in Fig. 2, the signals are more intense for the Cottonii carrageenans. Dyrby et al. (2004) suggested that the signal around 840-850 cm^{-1} might be affected by the counter ions. Despite investigating the correlation between the counter ion concentrations, measured by ICP-MS, and the score values, no significant correlation was found, possibly due to the homogeneous ionic composition of the commercial samples.

The intensity of the peak at 930 cm^{-1} indicates so-called degree of modification of carrageenan, i.e., 1,4 linked galactose-6-sulphate moieties modified to 3,6-anhydrogalactose moieties, essential for the gel forming properties of carrageenan (Van De Velde et al., 2002). A higher signal intensity in *Kappaphycus alvarezii* carrageenans, compared to *Chondrus* carrageenans, suggests a greater modification degree in the former. However, confirming this is challenging as commercial carrageenan aim for maximal modification degree.

The intensities of absorption bands at approximately 840 cm^{-1} and at 930 cm^{-1} are correlated and have high values in the loading vector for the first PLS component for models trained on the FT-IR spectra for *Chondrus* carrageenans (Fig. 5, middle images). The functional groups giving rise to these two signals are positively correlated to the breaking strength values, while the signals at 1110 cm^{-1} and 1000 cm^{-1} are

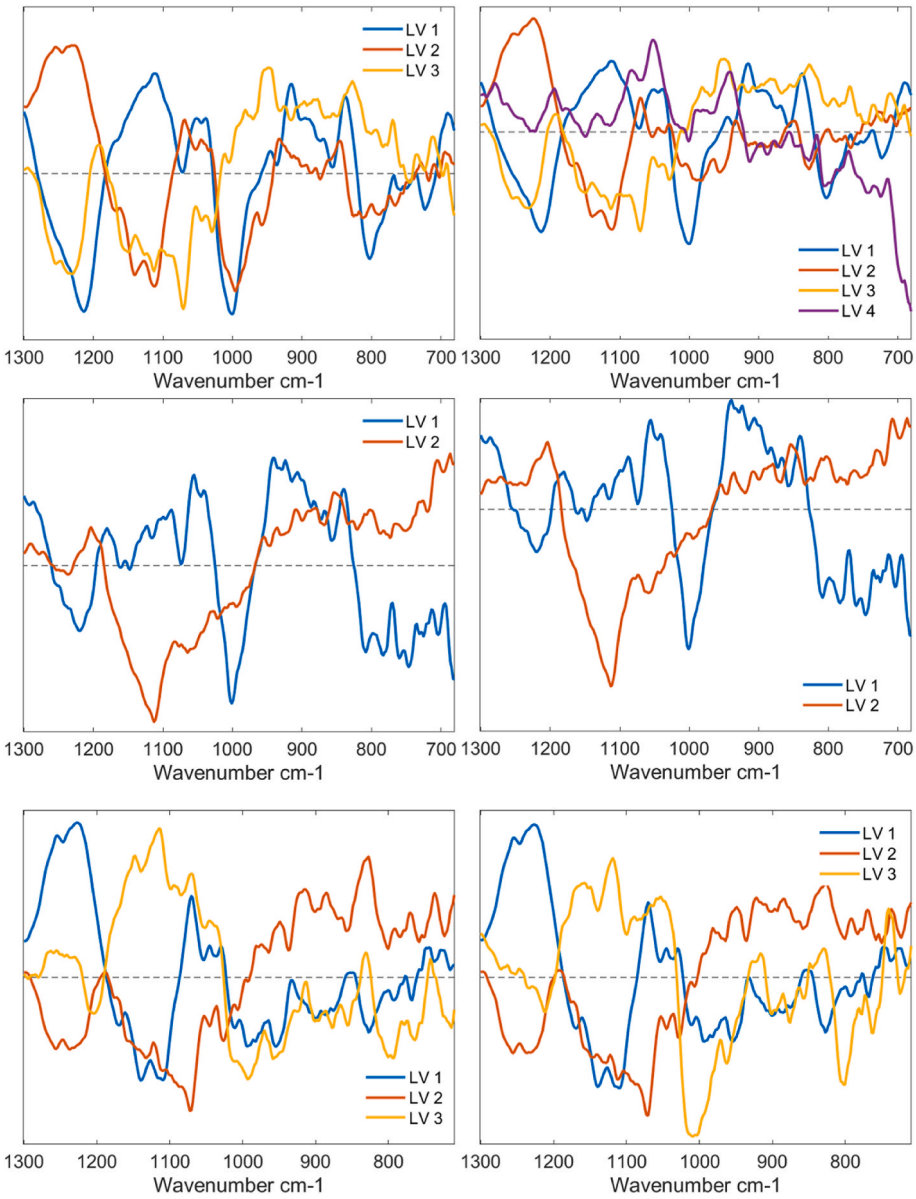


Fig. 5. Loading vectors of PLS models for the gel strength (left three), the breaking strength (right three) trained on FT-IR spectra for both carrageenan types (top two); only for Chondrus carrageenan samples (middle two); and only for Kappaphycus alvarezii carrageenan (bottom two).

negatively correlated with this parameter. The gel strength values increase with the higher concentrations of the functional groups linked to 840 cm^{-1}, 930 cm^{-1}, and 1110 cm^{-1} signals, but decrease with those at 1240 cm^{-1} and 1000 cm^{-1}. These observations exhibit similarities to the phenomena observed within the context of the global model.

The loading vectors for cottonii models reveal similar properties besides positive effect of the sulphate content on both functional properties and the signal 1110 cm^{-1} negatively correlated to both reference values. The sulfur content, S%, given in Table 2 and is 6.8 ± 0.6 (mean ± standard deviation) in range 3.8 %–7.3 % indicates high variation within *Kappaphycus alvarezii* carrageenan samples. Hence, it may be assumed that the observed high values in the loading vectors for the corresponding absorption band might be due to its correlation to the total carrageenan concentration in the sample since it would affect both properties, as higher carrageenan concentration would increase (not linearly) both reference values.

3.3. SO-PLS

Considering the SO-PLS models, different block combinations (numbers of blocks and their order) were tested, and the optimal model was chosen based on the lowest RMSECV and model complexity (total number of PLS components). To investigate the behavior of the RMSEP of the different models, it was calculated and displayed for all block combinations as well. Fig. 7 illustrates how RMSECV and RMSEP behave depending on the number and order of the blocks included in series of SO-PLS models predicting the breaking strength (top figure) and the gel strength (bottom figure).

Out of 24 different block combinations for training SO-PLS models to predict breaking strength, the second model including five PLS components form FT-IR and one from SEC, gained the lowest RMSECV and RMSEP, 23 and 27, respectively. However, the first SO-PLS models trained on the FT-IR spectra and on ICP data, and that only includes five PLS components from the FT-IR block, performs only slightly poorer, with RMSECV and RMSEP, of 24 and 27, respectively. The same holds for the third model developed on the FT-IR and NMR blocks.

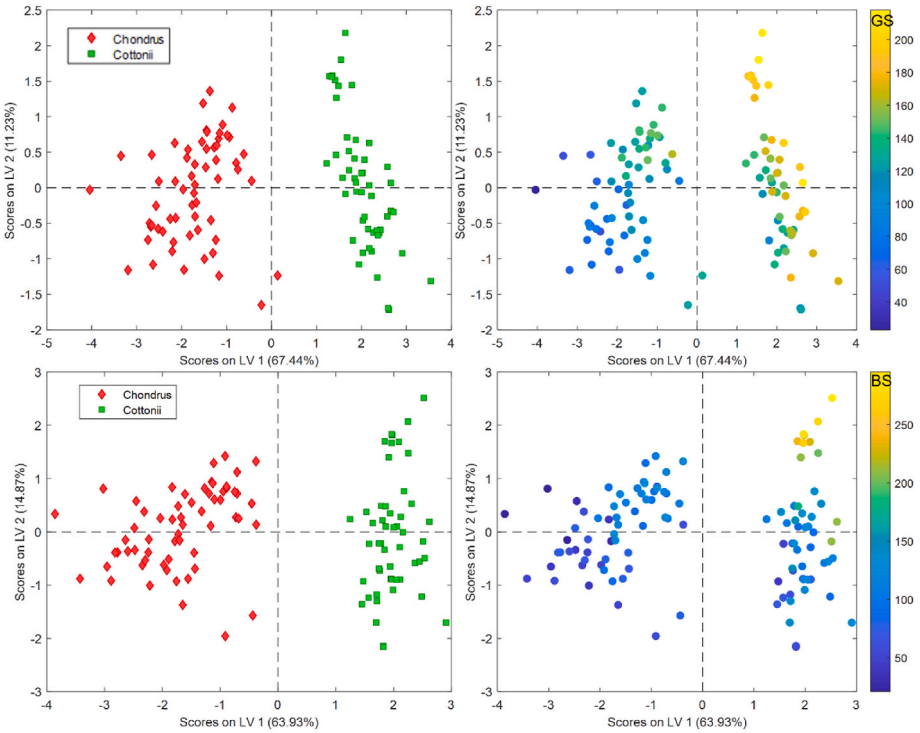


Fig. 6. Score plots for PLSR for global model for the gel strength (top), and the breaking strength (bottom). The scores on the left score plots are colored according to gel strength and breaking strength values for each sample.

The data from ICP-MS was not included in any of the models, most likely due to the limited low variation in the ionic composition of the commercial carrageenan thereby not showing effects of ions on carrageenans gelation properties of milk.

The NMR block was only included in the models that did not contain data from the FT-IR block indicating that these two data blocks contribute with similar information, as expected.

One or two PLS components from SEC were included in several SO-PLS models that also included information from the FT-IR blocks. These models show good performance. The data obtained from SEC includes information about molecular weight and concentration of carrageenan, among other parameters. The effect of the weight-average molecular mass on the breaking strength might be assumed due to the process of the coil-double helix transition during the carrageenan-milk gel formation that increases with the length of the carrageenan loops or tails involved (Snoeren, 1976). However, the studies of these relationship in milk matrices were not reported elsewhere in the literature.

The performance of the SO-PLS models for prediction of the gel strength are depicted in Fig. 7 (bottom part). The first model, trained on the FT-IR block and on the ICP block, resulted in the lowest RMSECV and RMSEP, of 12 and 14, respectively. The prediction error for the SO-PLS models is only slightly lower than the one obtained from PLS regression trained only on the FT-IR spectra (14 vs.15, Table 4).

Like the breaking strength models, the inclusion of the NMR block in the SO-PLS models did not improve the models, indicating that the FT-IR block alone contains all the necessary information for prediction of the gel strength as well. Although the SEC block is incorporated into several SO-PLS models, it does not enhance performance but contributes just to increased model complexity. In contrast, the ICP-block appears to be meaningful for predicting the gel strength, with one or two PLS components included in the most well performing models, probably due to well-known cation effects on carrageenan gelling properties. However, it does not outperform the single-block models trained only on the FT-IR models.

Considering the SO-PLS models trained on the two carrageenan types separately, their performance is summarized in Table 4 and their performance is comparable to the global models. Taking into the account the range of the gel and breaking strength values for each carrageenan, the model performance for global models is the same or better compared to the models developed on the individual carrageenan types. Hence, there is no need to split the data into different classes depending on the seaweed types they were extracted from, which simplifies model maintenance.

Comparing the multi-block model with other models presented in Table 4, the single block models utilizing FT-IR spectra still appeared as the most useful for predicting carrageenans functionality all things considered. The RMSEP for the SO-PLS model marginally surpasses that of the PLS model based on FT-IR spectra, with values of 14 and 15, respectively. However, this performance is slightly inferior to that of the SVM model, which achieved an RMSEP of 12. A similar trend is observed in the prediction of breaking strength, where single block models demonstrate comparable or superior predictive performance relative to the multi-block SO-PLS models. This observation is particularly insightful, since it challenges the initial assumption that a comprehensive analysis incorporating multiple analytical platforms would be necessary for predicting gel and breaking strength. Instead, the findings suggest that a singular analytical approach, specifically FT-IR spectroscopy, is adequate for predicting the functionality of carrageenan. However, a more profound causal understanding is facilitated by multi-block models, as demonstrated when adding the SEC or the ICP blocks to the IR block in SO-PLS. This approach emphasizes the significance of molecular information obtained through these techniques in comprehending the mechanical properties of the carrageenan-milk gel.

The explanatory power of the PLS models, detailed in Section 3.2, reveals that the predominant information captured by FT-IR spectra concerns the type of carrageenan, i.e., κ-carrageenan vs. κ/ι-carrageenan. The ability of the FT-IR spectrum-based model to effectively predict gel and breaking strength is likely attributable to the well-known gel-forming capabilities associated with different carrageenan types. Furthermore, it can be speculated that the lack of improvement in model performance with the inclusion of SEC-MALS or ICP-MS data may be attributed to the limited variation within each

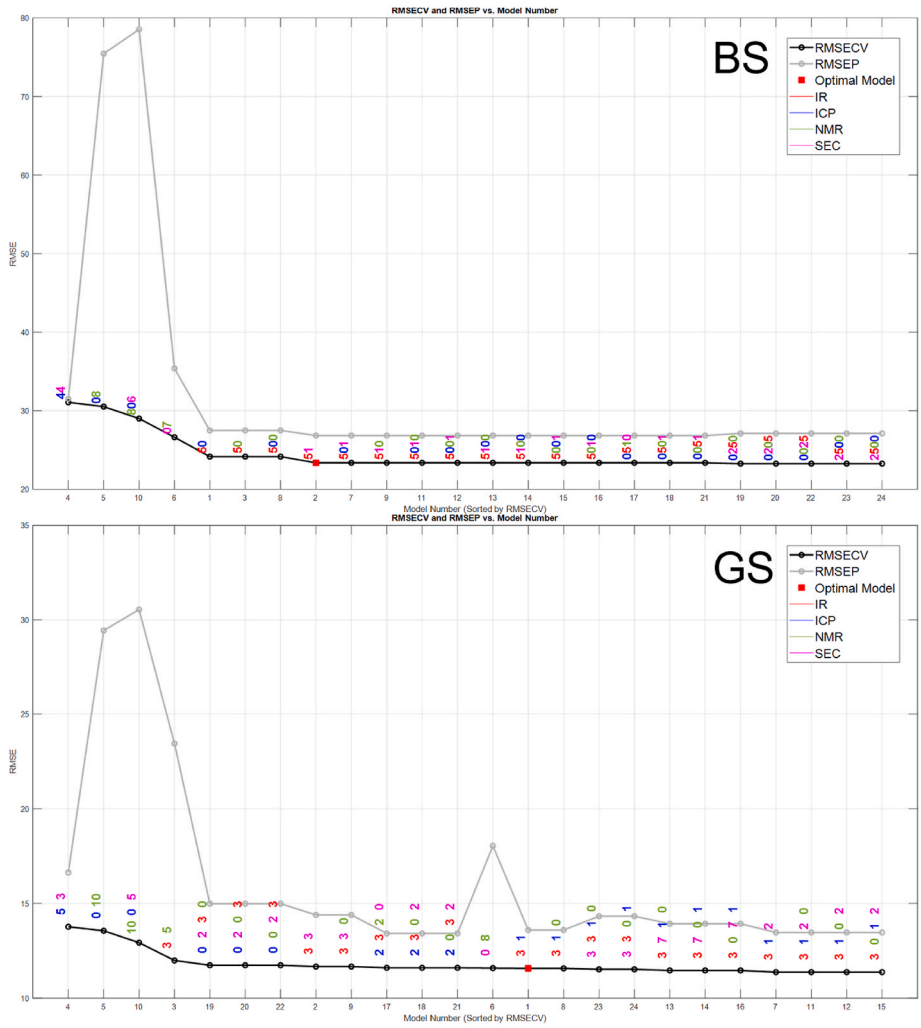


Fig. 7. SO-PLS models trained on data for both carrageenan types predicting the breaking strength (top) and the gel strength (bottom). Figure shows RMSECV and RMSEP for each model developed with different block combinations and orders given. The labels over each data point in RMSECV correspond to the optimal number of latent variables colored according to the data block they are from. The optimal model is defined as the one with minimal: RMSECV, RMSEP, RMSEP/RMSECV, and model complexity (total sum of PLS components).

dataset for each carrageenan type. For instance, commercial carrageenan samples are engineered to optimize performance, considering the effect of cations and molecular size, which results in a narrow range of variability. This homogeneity in the data could explain why additional analytical platforms, such as SEC-MALS and ICP-MS, do not significantly enhance the predictive models. The minimal variation within these datasets likely fails to introduce new, discriminative information that could refine the models' predictive power beyond what is achievable with FT-IR spectroscopy alone.

4. conclusion

In this study, single and multi-block modelling methods were applied to predict the gel and breaking strength for carrageenan-milk gels using data from four analytical platforms: FT-IR spectroscopy, NMR spectroscopy, SEC-MALS, and ICP-MS.

Among several single-block, linear or nonlinear methods the SVM trained on FT-IR spectra resulted in the lowest RMSEP of 12 for the gel strength and of 25 for the breaking strength. This method presents a viable alternative to the traditional approach used for characterizing the breaking and gel strength of milk-carrageenan mixtures. FT-IR spectroscopy, when combined with SVM, offers an efficient and rapid analytical technique that necessitates only carrageenan powder for measurement, bypassing the need for preparing carrageenan-milk gels and Texture Analyzer testing. By embedding the trained SVM model into the FT-IR system's software, manufacturers can achieve continuous monitoring of carrageenan quality.

To elucidate the relationship between the molecular structure of carrageenan, as revealed by FT-IR, and its functionality, PLS regression was employed for its capacity to enhance model explainability. The predictive capability for gel and breaking strength based on the FT-IR spectra of carrageenan is attributed to the distinct gelation mechanisms of κ- and ι-carrageenan types, along with the ability of FT-IR spectroscopy to differentiate between these carrageenan types.

Besides single-block chemometric methods, one multi-block method, SO-PLS, was employed. One of the aims of the performed multiblock data analysis was to investigate the value of adding another block. For this dataset, besides FT-IR spectra, an addition of other blocks did not improve the prediction power of the models. However, several data fusion models indicated the importance of the SEC-MALS and ICP-MS for causal understanding for carrageenan structure-functionality, while information provided by NMR spectra was comparable with evidence from FT-IR spectra. The limited improvement from integrating SEC-MALS or ICP-MS data is likely due to high homogeneity of commercial carrageenan, which limits the addition of discriminative information.

This study demonstrates the potential of combining spectroscopy

with machine learning for the quality control of hydrocolloids, specifically κ-carrageenan and κ/ι-hybrid carrageenan extracted from *Kappaphycus alvarezii* and *Chondrus* genus, using UHT milk as a model matrix. However, other hydrocolloids may present different spectral characteristics and exhibit distinct structure-functionality profiles in more complex or variable matrices. To expand the applicability of the proposed solution, it will be necessary to incorporate FT-IR spectra along with gel and breaking strength data for a wider range of hydrocolloid-milk gels into the model.

Funding details

This work was supported by CP Kelco ApS as part of the industrial PhD program of the Innovation Fund Denmark (grant number: 104400009B).

CRediT authorship contribution statement

Oksana Mykhalevych: Writing – original draft, Visualization, Validation, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Henrik Stapelfeldt:** Writing – review & editing, Validation, Supervision, Resources, Funding acquisition, Conceptualization. **Federico Marini:** Writing – review & editing, Software, Methodology, Conceptualization. **Rasmus Bro:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used OpenAI's ChatGPT in order to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

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.

Data availability

Data will be made available on request.

Acknowledgement

We thank all colleagues at CP Kelco who shared their deep carrageenan knowledge with us. Special credit is given to the QC laboratory personnel for help with measurements and equipment support.

References

Agoda-Tandjawa, G., Le Garnec, C., Boulenguer, P., Gilles, M., & Langendorff, V. (2017). Rheological behavior of starch/carrageenan/milk proteins mixed systems: Role of each biopolymer type and chemical characteristics [Article]. *Food Hydrocolloids, 73*, 300–312. https://doi.org/10.1016/j.foodhyd.2017.07.012

Anderson, A. D., Daubert, C. R., & Farkas, B. E. (2002). Rheological characterization of skim milk stabilized with carrageenan at high temperatures [article]. *Journal of Food Science, 67*(2), 649–652. https://doi.org/10.1111/j.1365-2621.2002.tb10654.x

Baeza, R. I., Carp, D. J., Pérez, O. E., & Pilosof, A. M. R. (2002). κ -Carrageenan—protein interactions: Effect of proteins on polysaccharide gelling and textural properties. *LWT, 35*(8), 741–747. https://doi.org/10.1006/FSTL.2002.0938

Bakshi, B., & Chen, H. (2009). Linear approaches for nonlinear modeling [bookitem]. *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis, 3*, 453–462. https://doi.org/10.1016/B978-044452701-1.00060-0

BeMiller, J. N. (2019). Carrageenans. *Carbohydrate chemistry for food scientists*. https://doi.org/10.1016/B978-0-12-812069-9.00013-3

Berth, G., Vukovic, J., & Lechner, M. D. (2008). Physicochemical characterization of carrageenans-A critical reinvestigation. *Journal of Applied Polymer Science, 110*(6), 3508–3524. https://doi.org/10.1002/app.28937

Biancolillo, A., & Næs, T. (2019). The sequential and orthogonalized PLS regression for multiblock regression: Theory, examples, and extensions. *Data Handling in Science and Technology, 31*, 157–177. https://doi.org/10.1016/B978-0-444-63984-4.00006-5

Bixler, H. J., Johndro, K., & Falshaw, R. (2001). Kappa-2 carrageenan: Structure and performance of commercial extracts: II. Performance in two simulated dairy applications. *Food Hydrocolloids, 15*(4–6), 619–630. https://doi.org/10.1016/S0268-005X(01)00047-9

Bro, R., & Smilde, A. K. (2014). Principal component analysis [Article]. *Analytical Methods, 6*(9), 2812–2831. https://doi.org/10.1039/C3AY41907J

Černá, M., Barros, A. S., Nunes, A., Rocha, S. M., Delgadillo, I., Čopíková, J., & Coimbra, M. A. (2003). Use of FT-IR spectroscopy as a tool for the analysis of polysaccharide food additives. *Carbohydrate Polymers, 51*(4), 383–389. https://doi.org/10.1016/S0144-8617(02)00259-X

Chen, N. (2004). In N. Chen (Ed.), *Support vector machine in chemistry*. World Scientific [Book].

Chopin, T., & Whalen, E. (1993). A new and rapid method for carrageenan identification by FT IR diffuse reflectance spectroscopy directly on dried, ground algal material. *Carbohydrate Research, 246*(1), 51–59. https://doi.org/10.1016/0008-6215(93)84023-Y

Ciancia, M., Matulewicz, M. C., & Tuvikene, R. (2020). Structural diversity in galactans from red seaweeds and its influence on rheological properties. *Frontiers in Plant Science, 11*, 559986. https://doi.org/10.3389/fpls.2020.559986 [Article], 559986.

De Ruiter, G. A., & Rudolph, B. (1997). Carrageenan biotechnology. *Trends in Food Science and Technology, 8*(12), 389–395. https://doi.org/10.1016/S0924-2244(97)01091-1

Drohan, D. D., Tziboula, A., McNulty, D., & Horne, D. S. (1997). Milk protein-carrageenan interactions. *Food Hydrocolloids, 11*(1), 101–107. https://doi.org/10.1016/S0268-005X(97)80016-1

Dyrby, M., Petersen, R. V., Larsen, J., Rudolf, B., Nørgaard, L., & Engelsen, S. B. (2004). Towards on-line monitoring of the composition of commercial carrageenan powders. *Carbohydrate Polymers, 57*(3), 337–348. https://doi.org/10.1016/J.CARBPOL.2004.05.015

Elfaruk, M. S., Wen, C., Chi, C., Li, X., & Janaswamy, S. (2021). Effect of salt addition on iota-carrageenan solution properties [Article]. *Food Hydrocolloids, 113*, Article 106491. https://doi.org/10.1016/j.foodhyd.2020.106491

Hilliou, L. (2021). Structure-elastic properties relationships in gelling carrageenans. *Polymers, 13*(23), 4120. https://doi.org/10.3390/polym13234120 [Article].

Hurtado, A. Q., Critchley, A. T., & Neish, I. C. (2017). Carrageenan industry market overview [bookitem]. In *Tropical seaweed farming trends, problems and opportunities* (Vol. 9, pp. 193–205). Springer International Publishing AG. https://doi.org/10.1007/978-3-319-63498-2_13.

Jacobsson, S. P., & Hagman, A. (1993). Chemical composition analysis of carrageenans by infrared spectroscopy using partial least squares and neural networks. *Analytica Chimica Acta, 284*(1), 137–147. https://doi.org/10.1016/0003-2670(93)80017-F

Jiang, J. L., Zhang, W. Z., Ni, W. X., & Shao, J. W. (2021). Insight on structure-property relationships of carrageenan from marine red algal: A review. *Carbohydrate Polymers, 257*, Article 117642. https://doi.org/10.1016/J.CARBPOL.2021.117642

Kennard, R. W., & Stone, L. A. (1969). Computer aided design of experiments. *Technometrics, 11*(1), 137–148. https://doi.org/10.1080/00401706.1969.10490666 [Article].

Langendorff, V., Cuvelier, G., Launay, B., Michon, C., Parker, A., & De Kruif, C. G. (1999). Casein micelle/iota carrageenan interactions in milk: Influence of temperature. *Food Hydrocolloids, 13*(3), 211–218. https://doi.org/10.1016/S0268-005X(98)00087-3

Langendorff, V., Cuvelier, G., Michon, C., Launay, B., Parker, A., & De Kruif, C. G. (2000). Effects of carrageenan type on the behaviour of carrageenan/milk mixtures. *Food Hydrocolloids, 14*(4), 273–280. https://doi.org/10.1016/S0268-005X(99)00064-8

Lynch, M. G., & Mulvihill, D. M. (1996). Rheology of ι-carrageenan gels containing caseins. *Food Hydrocolloids, 10*(2), 151–157. https://doi.org/10.1016/S0268-005X(96)80029-4

Marini, F. (2023). Sequential and orthogonalized partial least squares (SO-PLS). https://www.chem.uniroma1.it/romechemometrics/research/algorithms/so-pls/.

Michna, A., Płaziński, W., Lupa, D., Wasilewska, M., & Adamczyk, Z. (2021). Carrageenan molecule conformations and electrokinetic properties in electrolyte solutions: Modeling and experimental measurements [Article]. *Food Hydrocolloids, 121*, Article 107033. https://doi.org/10.1016/j.foodhyd.2021.107033

Mishra, P., Roger, J. M., Jouan-Rimbaud-Bouveresse, D., Biancolillo, A., Marini, F., Nordon, A., & Rutledge, D. N. (2021). Recent trends in multi-block data analysis in chemometrics for multi-source data integration. *TrAC, Trends in Analytical Chemistry, 137*, Article 116206. https://doi.org/10.1016/J.TRAC.2021.116206

Næs, T., Tomic, O., Afseth, N. K., Segtnan, V., & Måge, I. (2013). Multi-block regression based on combinations of orthogonalisation, PLS-regression and canonical correlation analysis [Article]. *Chemometrics and Intelligent Laboratory Systems, 124*, 32–42. https://doi.org/10.1016/j.chemolab.2013.03.006

Nielsen, N.-P. V., Carstensen, J. M., & Smedsgaard, J. (1998). Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *Journal of Chromatography A, 805*(1), 17–35. https://doi.org/10.1016/S0021-9673(98)00021-1 [Article].

Podzimek, S. (2011). *Light scattering, size exclusion chromatography, and asymmetric flow field flow fractionation powerful tools for the characterization of polymers, proteins, and nanoparticles [Book]*. Wiley.

Prado-Fernández, J., Rodríguez-Vázquez, J. A., Tojo, E., & Andrade, J. M. (2003). Quantitation of κ-, ι- and λ-carrageenans by mid-infrared spectroscopy and PLS

regression. *Analytica Chimica Acta, 480*(1), 23–37. https://doi.org/10.1016/S0003-2670(02)01592-1

Puvanenthiran, A., Goddard, S. J., Mckinnon, I. R., & Augustin, M. A. (2003). Milk-based gels made with κ-carrageenan. *Journal of Food Science, 68*(1), 137–141. https://doi.org/10.1111/j.1365-2621.2003.tb14129.x [Article].

Rinnan, Å., Berg, F. van den, & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC, Trends in Analytical Chemistry, 28*(10), 1201–1222. https://doi.org/10.1016/J.TRAC.2009.07.007

Robal, M., Brenner, T., Matsukawa, S., Ogawa, H., Truus, K., Rudolph, B., & Tuvikene, R. (2017). Monocationic salts of carrageenans: Preparation and physico-chemical properties. *Food Hydrocolloids, 63*, 656–667. https://doi.org/10.1016/j.foodhyd.2016.09.032 [Article].

Rodríguez Sánchez, R. A., Matulewicz, M. C., & Ciancia, M. (2022). NMR spectroscopy for structural elucidation of sulfated polysaccharides from red seaweeds. *International Journal of Biological Macromolecules, 199*, 386–400. https://doi.org/10.1016/J.IJBIOMAC.2021.12.080

Sabadini, E., Hubinger, M. D., & Cunha, R. L. (2006). The effects of sucrose on the mechanical properties of acid milk proteins-κ-carrageenan gels. *Brazilian Journal of Chemical Engineering, 23*(1), 55–65. https://doi.org/10.1590/S0104-66322006000100007 [Article].

Sedlmeyer, F., Daimer, K., Rademacher, B., & Kulozik, U. (2003). Influence of the composition of milk-protein κ/ι-hybrid-carrageenan gels on product properties. *Colloids and Surfaces B: Biointerfaces, 31*(1–4), 13–20. https://doi.org/10.1016/S0927-7765(03)00039-0

Sen, M., Yolacan, B., & Gueven, O. (2013). A comprehensive study on the size exclusion chromatography of kappa-carrageenan for the identification of after-peaks [Article]. *Journal of Applied Polymer Science, 127*(1), 494–499. https://doi.org/10.1002/app.37758

Skov, T., van den Berg, F., Tomasi, G., & Bro, R. (2006). Automated alignment of chromatographic data. *Journal of Chemometrics, 20*(11–12), 484–497. https://doi.org/10.1002/cem.1031 [Article].

Slootmaekers, D., van Dijk, J. A. P. P., Varkevisser, F. A., van Treslong, C. J. B., & Reynaers, H. (1991). Molecular characterisation of κ- and λ-carrageenan by gel permeation chromatography, light scattering, sedimentation analysis and osmometry. *Biophysical Chemistry, 41*(1), 51–59. https://doi.org/10.1016/0301-4622(91)87209-N [Article].

Smilde, A. K. (2022). In H. K, K. H. Liland, & T. Næs (Eds.), *Multiblock data fusion in statistics and machine learning : Applications in the natural and life sciences*. John Wiley & Sons, Inc [Book].

Snoeren, T. H. M. (1976). KAPPA-CARRAGEENAN a study on its PHYSICO-chemical properties, sol-gel transition and interaction with milk proteins. https://edepot.wur.nl/200249.

Tojo, E., & Prado, J. (2003a). A simple 1H NMR method for the quantification of carrageenans in blends [Article]. *Carbohydrate Polymers, 53*(3), 325–329. https://doi.org/10.1016/S0144-8617(03)00080-8

Tojo, E., & Prado, J. (2003b). Chemical composition of carrageenan blends determined by IR spectroscopy combined with a PLS multivariate calibration method. *Carbohydrate Research, 338*(12), 1309–1312. https://doi.org/10.1016/S0008-6215(03)00144-7

van de Velde, F. (2008). Structure and function of hybrid carrageenans. *Food Hydrocolloids, 22*(5), 727–734. https://doi.org/10.1016/j.foodhyd.2007.05.013

Van De Velde, F., Antipova, A. S., Rollema, H. S., Burova, T. V., Grinberg, N. V., Pereira, L., Gilsenan, P. M., Tromp, R. H., Rudolph, B., & Grinberg, V. Y. (2005). The structure of κ/ι-hybrid carrageenans II. Coil-helix transition as a function of chain composition. *Carbohydrate Research, 340*(6), 1113–1129. https://doi.org/10.1016/j.carres.2005.02.015

Van De Velde, F., Knutsen, S. H., Usov, A. I., Rollema, H. S., & Cerezo, A. S. (2002). 1H and 13C high resolution NMR spectroscopy of carrageenans: Application in research and industry. *Trends in Food Science & Technology, 13*(3), 73–92. https://doi.org/10.1016/S0924-2244(02)00066-3

Van De Velde, F., Peppelman, H. A., Rollema, H. S., & Tromp, R. H. (2001). On the structure of κ/ι-hybrid carrageenans. *Carbohydrate Research, 331*(3), 271–283. https://doi.org/10.1016/S0008-6215(01)00054-4

Van De Velde, F., Pereira, L., & Rollema, H. S. (2004). The revised NMR chemical shift data of carrageenans. *Carbohydrate Research, 339*(13), 2309–2313. https://doi.org/10.1016/J.CARRES.2004.07.015

Wilson, R. H., Goodfellow, B. J., & Belton, P. S. (1988). Fourier transform infrared spectroscopy for the study of food biopolymers. *Food Hydrocolloids, 2*(2), 169–178. https://doi.org/10.1016/S0268-005X(88)80015-8

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems, 58*(2), 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1

Optimizing Carrageenan Blends for Chocolate Milk Using Combined Factorial and Mixture Design

Oksana Mykhalevych[1,2], Henrik Stapelfeldt[1], and Rasmus Bro[2]

[1] CP Kelco ApS, 4623 Lille Skensved, Denmark

[2]Department of Food Science, University of Copenhagen, 1958 Frederiksberg, Denmark

oksana@food.ku.dk

Abstract

The impact of carrageenans from *Kappaphycus alvarezii*, *Chondrus crispus*, and *Gigartina radula* and their quality on the viscosity of chocolate milk for optimization of cocoa particle suspension were evaluated through a combined factorial and mixture design of experiments (DoE). The factorial component, assessing quality measured as carrageenan–milk gel breaking strength, showed no significant effect on viscosity, while the mixture design revealed strong concentration-dependent effects and key two-way interactions. Of the factors, the concentration of *Chondrus crispus*-derived carrageenan had the greatest influence on viscosity. A predictive model was developed and used to optimize carrageenan and sugar ratios for stable chocolate milk formulation. This model is to be used to support rapid and more sustainable formulation of chocolate milk stabilizing carrageenan mixtures without trial-and-error testing representing an excellent example of the use of Quality by Design.

Keywords Carrageenan, chocolate milk, design of experiments

1 INTRODUCTION

Carrageenan is a family of sulfated polysaccharides naturally found in the cell walls of red seaweeds (Rhodophyta). These hydrocolloids are widely used for stabilizing and texturing a variety of products. Chemically, carrageenans consist of repeating disaccharide units formed by β-D-galactose and either α-D-galactose or its 3,6-anhydro derivative, with varying positions of sulfate substitution that influence their ionic nature and functional properties (BeMiller, 2019a).

Several types of carrageenan exist, differing in their molecular structure and sulfate content, which determine their physical behavior and industrial use (Campo et al., 2009). The most common forms used commercially are κ-carrageenan, ι-carrageenan, and their hybrids. κ-carrageenan is composed of alternating β-(1→3)-D-galactose-4-sulfate and α-(1→4)-3,6-anhydro-D-galactose, while ι-carrageenan includes α-(1→4)-3,6-anhydro-D-galactose-2-sulfate units (De Ruiter & Rudolph, 1997).

Different types of carrageenan are typically derived from specific red seaweed genera. The *Kappaphycus* genus is a primary source of κ-carrageenan, yielding an almost pure form (with less than 10% ι-carrageenan). Hybrid κ/ι-carrageenans are extracted from various *Gigartina* and *Chondrus* species, though the proportion of each type varies depending on the seaweed source. In *Chondrus,* κ-carrageenan tends to predominate, while *Gigartina* species typically contain higher levels of ι-carrageenan in their copolymers—this ratio can differ based on the specific carrageenophyte. These variations provide additional flexibility in terms of functional properties and industrial applications (Bixler et al., 2001; Colusse et al., 2022; Falshaw et al., 2001).

One of the well-known applications of carrageenan is in chocolate milk preventing sedimentation of cocoa particles and improving mouthfeel of the drink. This stabilization is due the result of binding of κ-carrageenans to casein micelles through charge

interactions, creating a weak, thixotropic, non-thermally reversible, and pourable gel, while its shear-thinning behavior allows for easy pouring and swallowing (BeMiller, 2019b). However, the use of pure κ-carrageenan is limited due to its insufficient shear stability in chocolate milk formulations. Although hybrid carrageenan has weaker water-gelling properties compared to pure κ-carrageenan, it still maintains a sufficiently high milk reactivity, making it suitable for dairy applications. While pure κ-carrageenan provides adequate viscosity and suspension initially, its κ-carrageenan/κ-casein network is prone to breakdown during transport-induced agitation and does not readily reform under refrigeration which may lead to cocoa particle sedimentation. In contrast, hybrid carrageenan forms a network with κ-casein that can rapidly re-establish under cold conditions, thereby maintaining stability of the chocolate milk (Bixler et al., 2001).

Carrageenan products used industrially for stabilizing cocoa particles in chocolate milk typically consist of blends of different carrageenan types. These mixtures are often developed through trial-and-error rheological testing, which is resource-intensive, ineffective and eventually generates food waste. To promote more sustainable and efficient product development, an alternative approach is needed. Given that different carrageenan types—extracted from distinct seaweed species—vary in functionality, it is valuable to investigate how their individual and combined effects influence cocoa stabilization. In this study, we propose using Design of Experiments (DoE) (Montgomery, 2013) to systematically formulate and evaluate blends of carrageenans from various seaweed sources. By modeling the DoE data in relation to functional performance (i.e., cocoa particle stabilization), we can begin to uncover structure–function relationships for carrageenan. This approach not only defines a design space for optimal stabilization but also supports more sustainable and resource-efficient formulation by maximizing raw material use and reducing waste.

2 MATERIALS AND METHODS

2.1 SAMPLES

Carrageenan samples were provided by CP Kelco (Lille Skensved, Denmark) as semi-finished materials, representative of the product prior to final commercial blending. As such, no catalogue code is available.

Carrageenan samples were derived from three seaweed species: *Kappaphycus alvarezii, Chondrus Crispus,* and *Gigartina radula*. The samples were chosen based on their milk-carrageenan breaking strength measured using methods described by (Mykhalevych et al., 2025). Nine samples were selected – three per each seaweed specie used for carrageenan extraction – representing low (-1), medium (0), and high (1) breaking strength levels.

Table 1 Molecular weight (Mw), polydispersity index (Mw/Mn), recovery percentage, sulfur content (S%), and cation composition (Ca, K, Na in mg/g) of the carrageenan samples analyzed by SEC-MALS and ICP-MS.

Type	Chondrus			Cottonii			Gigartina		
Level	-1	0	1	-1	0	1	-1	0	1
M_w(kDa)	888	760	924	586	577	620	491	447	734
Polydispersity (M_w/M_n)	1.8	1.9	1.6	1.6	1.9	1.8	1.9	2	1.8
Purity %	87	87	85	91	92	91	77	79	80
S %	9	9	8	7	7	7	10	9.1	9.2
Ca (mg/g)	36	33	34	20	24	22	43	43	40
K (mg/g)	22	24	24	36	38	36	15	14	12
Na (mg/g)	18	18	16	10	10	7	29	25	28

The weight-average molecular weight (M_w), polydispersity index (Mw/Mn), purity percentage, sulfur content and cation profile for each sample are summarized in Table 1. Elemental analysis was performed by Inductively Coupled Plasma Mass Spectrometry (ICP-MS), where the sulfur content (S%) represents the weight percentage of elemental

sulfur in the sample and serves as a proxy for the degree of sulfation, assuming sulfate esters are the primary sulfur-containing groups. The recovery percentage, measured by Size Exclusion Chromatography coupled with Multi-Angle Light Scattering detection (SEC-MALS) reflects the proportion of light-scattering material detected relative to the amount of sample injected, serving as a proxy for total carbohydrate content. The methods used to collect the data are described by (Mykhalevych et al., 2025). Given sample comparable cation composition (Ca^{2+}, K^{+}, Na^{+}) in Table 1, it is reasonable to assume that the distribution of monosaccharide types within the hybrid carrageenan copolymer — particularly the ratio of κ- to ι-type units — is also similar across samples within each source group.

2.2 DATA COLLECTION

The response in this study is the apparent viscosity of chocolate milk prepared with the addition of different carrageenan blends to achieve stabilization of the suspension of cocoa particles.

2.2.1 Chocolate Milk Preparation

Chocolate milk was prepared by dissolving a mixture of carrageenan (0.025%), skim milk powder (Arla Foods amba, spray-dried; 8.75%), sucrose (6.00%), and cocoa (ADM cocoa, 10-12% fat, type D-11-MR; 1.20%) in deionized water. The solution was heated in a 50°C water bath with continuous stirring until fully dissolved. Once the temperature of the chocolate milk reached 68°C, it was maintained in the water bath for an additional 15 min.

Subsequently, the chocolate milk was cooled to 5-10 °C and stored overnight in a climate-controlled cabinet at 3-4 °C. Before the measurement the milk is transferred to cooled down to 5 ˚C viscosity glasses. The apparent viscosity is measured by Brookfield

Viscometer LVF or LVT, spindle no.1, and guard, 60 rpm, 30 sec at temperature of chocolate milk in the range of 4.5 – 5.5 ˚C.

For each experiment set, a control sample was incorporated to track any variations between batches of cocoa and skim milk powder, with adjustments made for their potential impacts.

2.2.2 Viscosity Measurement

Prior to measurement, the chocolate milk was transferred to viscosity glasses cooled to 5 °C. The apparent viscosity at 5.0 ± 0.5 °C was measured using a Brookfield Viscometer LVF or LVT (Brookfield Engineering Laboratories, Inc. (Middleboro, Massachusetts, USA)) with spindle No.1 (19 mm in diameter and 65 in length) and guard, at 60 rpm for 30 seconds.

The apparent viscosity values, originally recorded in centipoises, were normalized to a range of 1 – 100 arbitrary units (A.U). Based on the experience, the viscosity range 30 – 60 (A.U) was identified as sufficient to achieve a permanent suspension of cocoa particles, which is considered the desired quality range for the carrageenan blend.

All carrageenan blends, as defined by the DoE protocol, were prepared and used for chocolate milk preparation randomly by multiple operators over a period of approximately four months ensuring a completely randomized design.

This study exclusively employed analytical grade chemicals. Ion-exchanged water was consistently utilized throughout all experimental procedures.

2.3 DESIGN OF EXPERIMENTS

Table 2 Ingredients (factors), levels and mixture concentrations for each ingredient, for the design of experiments protocol.

Ingredient	Quality Levels	Concentration Range (%)
K. alvarezii	Low, Medium, High	0-20
C. crispus	Low, Medium, High	5-40
G. radula	Low, Medium, High	5-40
Sugar (Sucrose)	One Quality Level	15-35

The designed experiment was a combination of a factorial and a mixture design. Table 2 summarizes the factors and levels used to calculate treatments in the DoE protocol. Four ingredients were included in the experiments: sucrose and three carrageenan types extracted from *K. alvarezii, C. Crispus,* and *G. radula.* Each carrageenan type was evaluated at three quality levels (low (-1), medium (0), high (1), see section 2.1), while sucrose was included with one quality level, giving four categorical variables: three factors with three levels each and one factor with a single level, forming the factorial part of the design.

For the mixture design part of the DoE protocol, each carrageenan type was added to the blend at varying concentration. The concentration ranges for each carrageenan type and sugar in each carrageenan blend are given in Table 2. These concentrations were chosen based on industrial experience with carrageenan application in chocolate milk, ensuring coverage of the apparent viscosity of interest, as well as extending above and below this range. This results in four continuous variables for the mixture design part of the DoE protocol. The samples were also selected based on their molecular composition to block potential confounding effects from parameters such as ion composition or sample purity (see data in Table 1). These characteristics were kept as similar as possible within each group to ensure comparability.

The constraint for the design was that the sum of all concentrations for each treatment must be 100%.

The experimental design utilized was a combination of a factorial design (Montgomery Douglas C, 2013a) and a mixture design (Montgomery Douglas C, 2013b). This hybrid approach allowed for the exploration of both categorical factors (three ingredient quality levels) and continuous factors (ingredient concentrations) within a single experimental framework. The inclusion of a factorial design allowed for the examination of how the different quality levels of each carrageenan type interact with each other and how these interactions influence the overall viscosity.

The mixture design was chosen to address the constraint that the total concentration of ingredients must sum to 100%. This design allows for the exploration of the effects of changing proportions of ingredients within a defined range, providing insight into the optimal blend that achieves the target viscosity. Thereby answering question two and three above.

In total 159 treatments were designed with 27 combinations withing factorial design (categorical variables of ingredients quality) varying from five to seven treatments each covering different combinations of continuous variables. This may not give a perfect balance within categorical variables varying but is reasonably well-distributed to cover experimental space within continuous variables adequately. Due to a very high number of experiments needed to cover the experimental space it was not possible to perform replicated measurements and therefore only duplicate viscosity measurements were performed and the average value of these was used.

2.4 Data Analysis

The effect of ingredient quality levels and concentrations on the viscosity of chocolate milk was modeled using an interaction model (Eq. 1):

$$y = \beta_0 + \sum_{i=1}^{p} \beta_i X_i + \sum_{i=1}^{p} \sum_{j=i+1}^{p} \beta_{ij} X_i X_j + \epsilon \qquad \text{Eq.1}$$

Where:

- y – the response variable (viscosity of chocolate milk)
- X_i – the main effects (ingredient quality levels and concentrations)
- β_0 – the intercept
- β_i – the coefficients for the main effects
- β_{ij} – the coefficients for the two-way interactions
- β_{ijk} – the coefficients for the three-way interactions
- ε – random error.

To understand the relationship between the variables, we initially fitted a model that included all main effects, two-way interactions, and relevant three-way interactions, and evaluated using analysis of variance (ANOVA). The model was refined by identifying significant predictors by evaluating the *p-values* for each term in the model than had to be less than 0.05 to be considered statistically significant. For each significant interaction term, all relevant lower-order terms were also included to maintain the hierarchy of the model. Factors that were not statistically significant (*p-value* > 0.05) were removed resulting in a final reduced ANOVA model.

The goodness of fit for each model was evaluated using the R^2 and adjusted R^2 values. The fit was further assessed using residual plots including, residual vs. fitted values, histogram of residuals, to assure that the model assumption of normality and homoscedasticity were met.

Furthermore, ANOVA lack of fit test to use to evaluate if the model captures the structure. While F-statistic from ANOVA (*p-value* < 0.05) was used to assess overall significance of the model.

Furthermore, due to a slightly unbalanced design and limited number of replicates, ten-fold cross validation was performed for the final model to ensure robust predictive performance and to mitigate the potential effects of overfitting. All rows in the dataset were randomized before cross-validation.

2.5 RECIPE OPTIMIZATION

After establishing the ANOVA model, the next step was to use the model to optimize the recipe. The aim of this step was to maximize the proportion of sugar concentration within the mixture while ensuring that the resulting viscosity remains within a target range of 30 to 60 A.U.

The optimization problem was formulated to maximize sugar concentration limited by three constraints:

a) The sum of all ingredients must be 100%
b) The predicted viscosity, computed from the ANOVA model must lie within the acceptable range of 30 to 60 A.U.
c) Ingredient proportions must be within experimentally observed bounds (concentration ranges given in Table 1).

The optimization was implemented using MATLAB's "fmincon" function (The Mathworks, n.d.) , employing the Sequential Quadratic Programming (SQP) algorithm (Nocedal & Wright, 2006).

2.6 SOFTWARE FOR DATA ANALYSIS

The design of experiments was developed using JMP Pro 16.0.0, 2021 SAS Institute Inc. The data analysis was performed using MATLAB (Release R 2023a, version: 9.14.0; The MathWorks Inc.).

3 RESULTS AND DISCUSSION

3.1 DESIGN OF EXPERIMENT RESULTS

Table 3 Summary of estimated coefficients from the ANOVA model predicting viscosity. The table includes main effects and two-way interaction terms for ingredient percentages and categorical quality variables and p-values are reported for each term. Statistically significant interactions (p < 0.05) are highlighted in bold.

	Estimate	p-value
K. alvarezii	-1195	0.383
C. crispus	-3.9	0.096
G. radula	1166	0.397
Sugar %	14.4	0.204
K. alvarezii %	15.5	0.168
C. crispus %	15.7	0.164
G. radula %	15.8	0.160
K. alvarezii × Sugar %	12.0	0.381
***C. crispus* × Sugar %**	**0.2**	**0.053**
G. radula × Sugar %	-11.7	0.396
K. alvarezii × *K. alvarezii %*	11.8	0.389
G. radula × *K. alvarezii* %	-11.7	0.395
K. alvarezii × *C. crispus* %	11.9	0.385
G. radula × *C. crispus* %	-11.6	0.399
Sugar % × *C. crispus* %	**-0.026**	**0.038**
***K. alvarezii* % × *C. crispus* %**	**-0.031**	**0.014**
K. alvarezii × *G. radula* %	11.9	0.385
G. radula × *G. radula* %	-11.5	0.406
Sugar % × *G. radula* %	**- 0.036**	**0.003**
***K. alvarezii* % ×*G. radula* %**	**-0.027**	**0.034**

Table 4 Analysis of variance (ANOVA) model predicting viscosity. The table reports the sum of squares (SumSq), degrees of freedom (DF), mean squares (MeanSq), F-statistics (F), and p-values for the overall model, linear and nonlinear terms, and residuals. The model explains a significant proportion of the variance in viscosity ($p < 0.0001$). The lack-of-fit test indicates that the model's residuals do not significantly deviate from pure error ($p = 0.39$), suggesting an adequate fit to the data.

	Sum of Squares	DF	Mean Square	F-value	p-value
Total	66269	158	419		
Model	59782	20	2989	64	0
Main effects	56050	7	8007	170	0
Interaction effects	3732	13	287	6	0
Residual	6487	138	47		
Lack of fit	6474	137	47	3	0.39
Pure error	12	1	12		

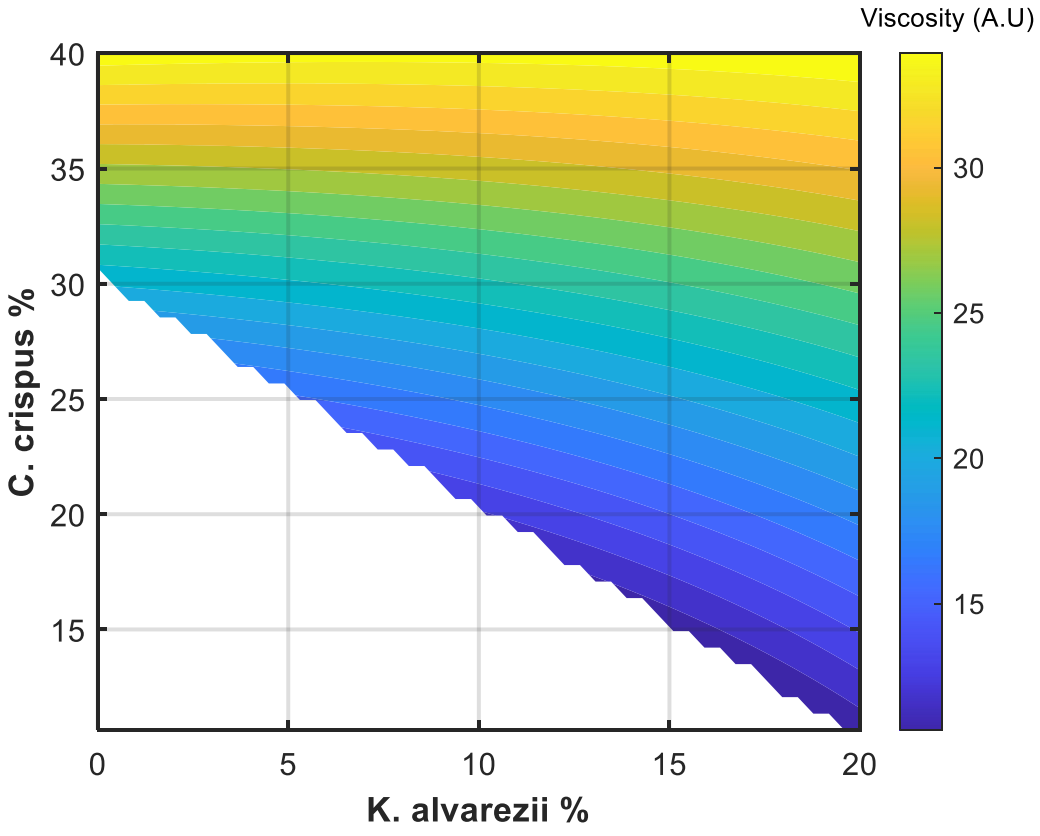


Figure 1 Contour plot illustrating the interaction between concentrations of carrageenan extracted from K. alvarezii and from C. Crispus on viscosity. Concentrations of the other two ingredients were adjusted proportionally to maintain a total mixture sum of 100%.

248

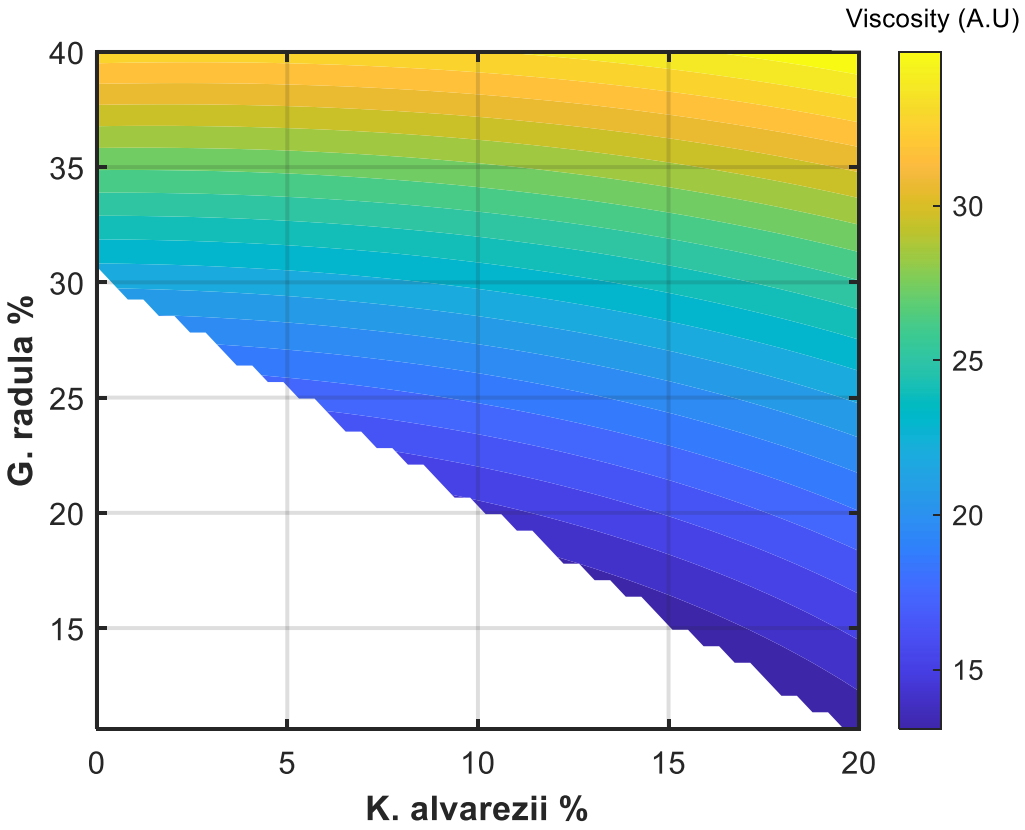


Figure 2 Contour plot illustrating the interaction between concentrations of carrageenan extracted from K. alvarezii and from G. radula on viscosity. Concentrations of the other two ingredients were adjusted proportionally to maintain a total mixture sum of 100%.

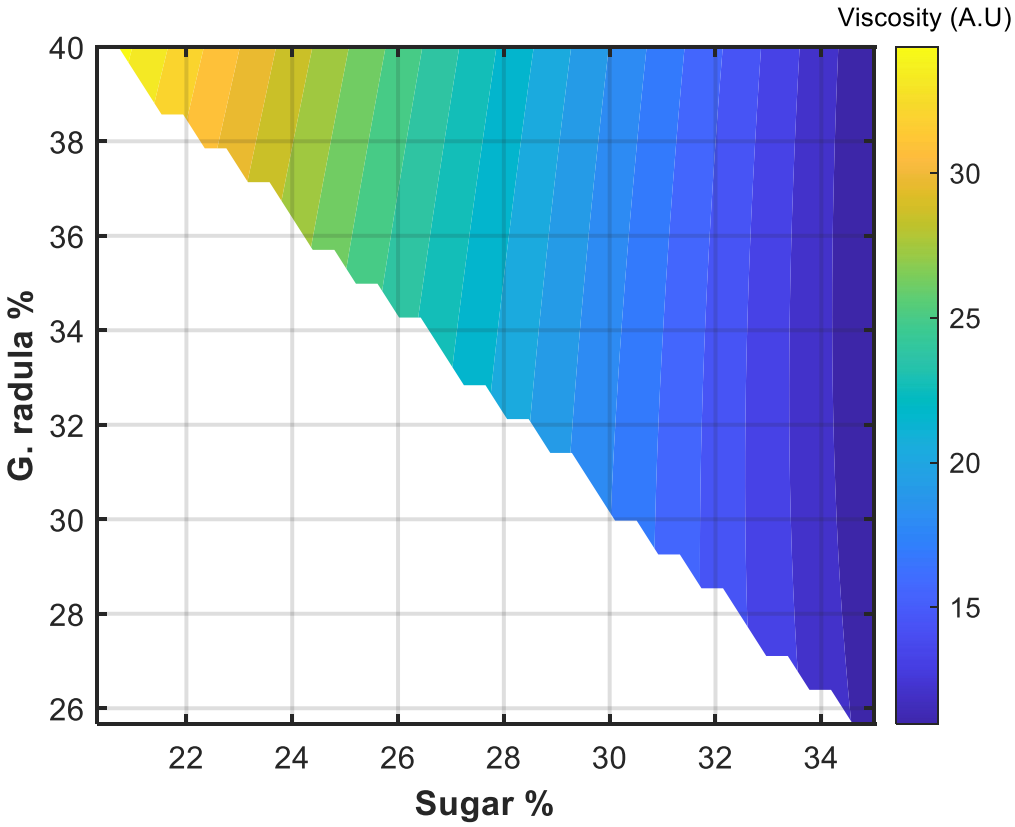


Figure 3 Contour plot illustrating the interaction between concentrations of sugar and carrageenan extracted from G. radula on viscosity. Concentrations of the other two ingredients were adjusted proportionally to maintain a total mixture sum of 100%.

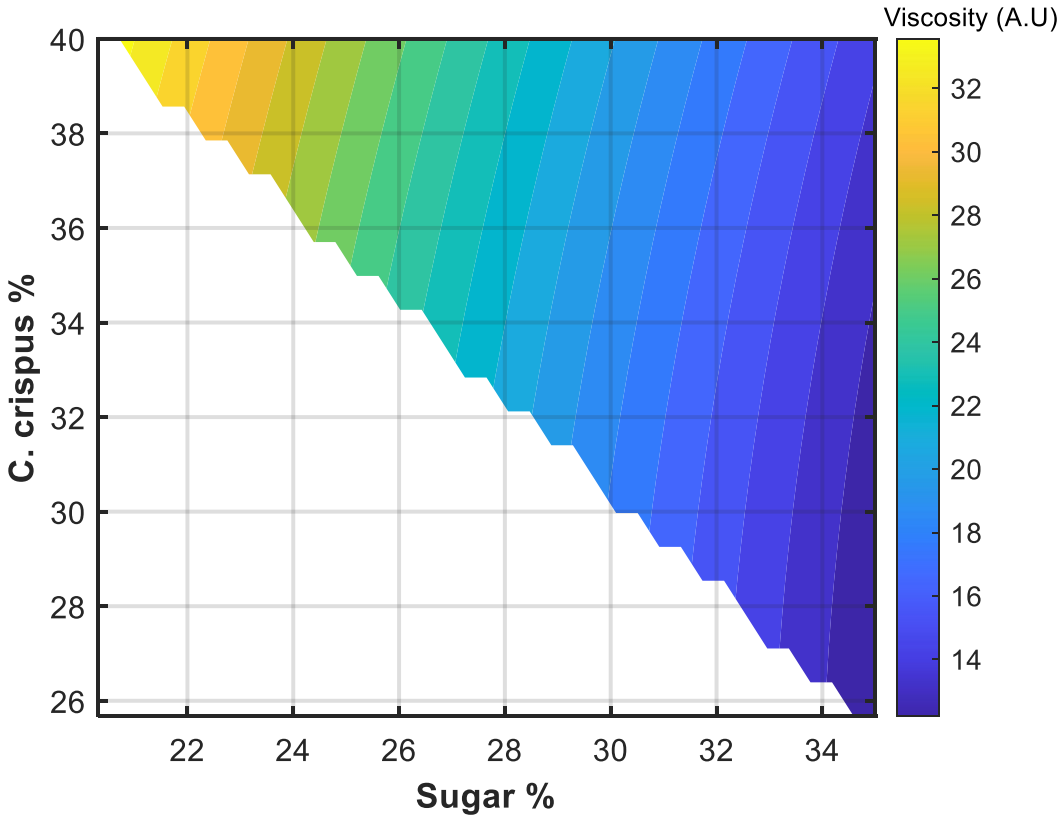


Figure 4 Contour plot illustrating the interaction between concentrations of sugar and carrageenan extracted from C. Crispus on viscosity. Concentrations of the other two ingredients were adjusted proportionally to maintain a total mixture sum of 100%.

All formulations and corresponding viscosity indices are given in appendix A. The viscosity of chocolate milk was measured and modeled as a response for each carrageenan formulation in ANOVA model including main effects and two-way interactions. The output of this model is summarized in appendix B. Estimated coefficients and model statistics from the reduced ANOVA model are presented in Table 3.The model included all seven main effects and thirteen interactions. The results of analysis of variance, summarized in the Table 4, show that both main effects and interactions of these (as each separate group) are statistically significant ($p < 0.05$). The limited replication in the experimental design resulted in a degree of freedom (DF) of only one for the pure error estimation, which makes the lack-of-fit test (p=0.39) less reliable. The lack-of-fit test is typically used to assess whether the model adequately describes

the data or if additional terms might be needed. Appendix C presents additional model statistics used to evaluate the goodness of fit of the final model. To further support the model's validity, cross-validation was conducted to assess its ability to generalize to unseen data. The results, also included in Appendix C, confirm the model's robustness and predictive performance. Despite, in general good model statistics, it must be acknowledged that the model is mostly suitable for the viscosity range of up to 70, as there is limited number of measurements above this value. This due to the limited assortment of the samples that would result in a higher viscosity without inducing gelation (Tijssen et al., 2007) that, for mouthfeel reasons, is undesirable in this application.

Notably, none of the main effects were statistically significant at the 0.05 level (Table 3). However, the relatively low p-value for quality of carrageenan extracted from *C. crispus* ($p = 0.096$) indicated that it may still be important in combination with other factors. The following interactions were statistically significant at the 0.05 level: "*K. alvarezii* % × *C. crispus* %", "*K. alvarezii* % × *G. radula* %", "Sugar % × *G. radula* %", "Sugar % × *C. crispus* %" and on the borderline of the significance level was – "*C. crispus* × Sugar %".

Carrageenan quality types, here defined as carrageenan-milk gel breaking strength, alone are not significant but influence viscosity when interacting with concentrations. It means that if we change the quality of carrageenan, but keep its concentration fixed, the viscosity does not change significantly. Quality alone does not significantly affect viscosity, but it influences how much the ingredient's concentration affects viscosity. So, quality is important—but only when one also considers the concentration.

It may appear surprising that the concentration of sugar in the formulation is not significant in this case while it is well known that it corresponds to lowering concentration of carrageenan, which results in lower viscosity. It might be a case in a simple formulation with only one carrageenan type of a constant quality. In these more complex formulations, an effect of a sugar concentration depends on the whole mixture as

294 “Sugar%” is involved in significant interaction terms such as “*C. crispus* × Sugar %”,
295 “Sugar % × *C. crispus* %” and “Sugar % × *G. radula* %”, meaning the effect depends on
296 other factors.

297 The coefficients for all significant interactions except the quality level of carrageenan
298 extracted from *C. crispus* in combination with sugar concentration were negative (Table
299 3, “Estimates”). These negative coefficients suggest that when both interacting factors
300 increase together, their combined effect reduces viscosity more than expected from the
301 individual parts.

302 The contour plot in figure 1 visualizes predicted viscosity values across the experimental
303 range of the “*K. alvarezii* % × *C. crispus* %” interaction. Although the interaction
304 coefficient between *K. alvarezii* % and *C. crispus* % is negative, the overall viscosity still
305 increases as both concentrations rise. This suggests that the combined effect of these
306 two ingredients is not purely additive; rather, they exhibit diminishing returns. While each
307 ingredient increases viscosity individually, their simultaneous increase does not enhance
308 viscosity as much as their independent contributions would predict. The interaction
309 between *K. alvarezii* and *C. crispus* suggests a balance between viscosity enhancement
310 and gelation. While carrageenan from *K. alvarezii* contributes to a more rigid network
311 that can lead to gelation, carrageenan from *C. crispus* introduces an elastic structure that
312 maintains viscosity without inducing gelation. The negative interaction term implies that
313 increasing both together results in diminishing viscosity gains, likely due to structural
314 competition. This interplay is crucial for applications requiring stable suspensions while
315 avoiding excessive gelation, ensuring a solution that remains flowable yet capable of
316 suspending dispersed particles.

317 The contour plot in Figure 2 visualizes predicted viscosity values across the experimental
318 range of the “*K. alvarezii* % × *G. radula* %” interaction. The predictions closely resemble
319 those of the previously discussed “*K. alvarezii %* ×*C. crispus %*” interaction, which aligns

with the similar interaction coefficients estimated by the ANOVA model (Table 3). This similarity may be attributed to the structural composition of carrageenan. *G. radula* seaweed produces hybrid carrageenans with a mix of κ- and ι-carrageenan structures, akin to the carrageenan from *C. crispus*. The comparable viscosity behavior suggested by the model implies that *G. radula* and *C. crispus* carrageenans may serve as functional substitutes in viscosity-driven applications. However, this is probably depended on the ratio of each carrageenan type within co-polymer.

The effects of "*G. radula* % × Sugar %" and "*C. crispus* % × Sugar %" interactions on viscosity are illustrated in Figures 3 and 4. These two contour plots exhibit strong similarities, which is expected due to the similar interaction estimates provided by the model (Table 3). Based on the predictions from the contour plots, low sugar concentration combined with high carrageenan concentration leads to higher viscosity. This aligns with the rheological behavior of carrageenan as a thickening and gelling agent.

However, the negative interaction coefficients for both interactions (Table 3) indicate that simultaneously increasing both sugar and carrageenan concentrations leads to a slightly lower-than-expected viscosity. This suggests that while increasing carrageenan concentration generally enhances viscosity, its effect is not as steep when sugar is also increased. This may be due to molecular interactions where high sugar concentrations reduce carrageenan hydration or alter its network formation, leading to diminished viscosity enhancement.

The positive coefficient for “*C. crispus* × Sugar %” suggests that increasing sugar concentration contributes more effectively to viscosity when combined with higher quality of carrageenan originated from *C. crispus*. However, if a lower quality of this carrageenan type is used, adding more sugar has a weaker effect on viscosity. This can be due to varying concentrations of ι-carrageenan and κ-carrageenan within co-polymer that

affects carrageenan–milk gel breaking strength (Mykhalevych et al., 2025) and thereby viscosity of chocolate milk.

3.2 OPTIMIZATION OF THE RECIPE AND IMPLEMENTATION

The optimization of the recipe was implemented using sequential quadratic programming algorithm-based function with an aim to maximize the sugar concentration. Viscosity predictions, constrained to the acceptable range of 30 to 60 A.U., were based on the ANOVA model estimates presented in Table 3. An example of an optimal formulation achieving this objective is provided in Table 5.

Table 5 Carrageenan blend recipe defined by sequential quadratic programming algorithm maximizing sugar concentration. The estimated viscosity for this recipe was 31.

Ingredient	Concentration Range (%)
K. alvarezii	5
C. crispus	40
G. radula	26
Sugar	29

The predicted viscosity for this recipe was 31 A.U. that is almost minimal allowed by the constraint (viscosity range of 30 to 60 A.U). That is expected as it was earlier shown that maximizing sugar concentration lowers solutions viscosity. The predicted value suggests that concentration of *C. crispus*-derived carrageenan is important when a high sugar concentration is desired. Importantly, the optimization algorithm operated within the same factor ranges as the original DoE, as extrapolation beyond the experimental domain is generally avoided. It is therefore reasonable to question whether the inclusion of *K. alvarezii*-derived carrageenan in the blend is driven solely by the imposed constraint requiring a minimum of 5% of this component. When this constraint is removed in the optimization model, the proportion of *K. alvarezii*-derived carrageenan drops to zero, with its 5% share being fully reallocated to *C. crispus*-derived carrageenan.

The 5% minimum for *K. alvarezii* was set based on practical experience regarding the stability of chocolate milk formulations. However, since stability was not directly measured in this study, the model suggests that greater flexibility in recipe optimization could be achieved if additional quality attributes were measured and broader experimental ranges explored. Nonetheless, given the scope of the current work, expanding the study further was beyond its feasible limits.

4 CONCLUSION

This study aimed to investigate the effects of carrageenan type and quality—defined here as carrageenan–milk gel breaking strength—on the viscosity of chocolate milk, with the goal of identifying formulation strategies that ensure stable cocoa suspension. A combined factorial and mixture design of experiments approach was used to evaluate both the qualitative and quantitative contributions of carrageenans extracted from *Kappaphycus alvarezii*, *Chondrus crispus*, and *Gigartina radula*, along with sugar concentration.

The factorial component, representing different quality levels of each carrageenan, did not yield statistically significant effects on viscosity within the tested conditions. This suggests that the breaking strength of milk–carrageenan gels, while relevant in other contexts, may not sufficiently capture the complexity of how carrageenan interacts in multicomponent systems like chocolate milk. The limited quality range and measurement uncertainty may have also contributed to the lack of significance. Consequently, within the studied ranges, using a single quality level for each carrageenan type may be sufficient for achieving desired viscosity outcomes.

In contrast, the mixture design revealed significant effects of ingredient concentrations and their interactions. Notably, the concentration of *C. crispus*-derived carrageenan had the strongest influence on viscosity, appearing in several significant two-way interaction

terms. While sugar concentration alone was not a significant predictor, its effect was strongly modulated by interactions with specific carrageenan types, particularly *C. crispus* and *G. radula*. These findings highlight the importance of considering synergistic interactions when formulating carrageenan-based suspensions.

Using the ANOVA model, an optimization routine was implemented to identify the ingredient composition that maximizes sugar content while maintaining viscosity within the target range (30–60 A.U.). The optimal blend also included the highest tested concentration of hybrid *C. crispus*-derived carrageenan, while *K. alvarezii* (a source of pure κ-carrageenan) was kept at its minimum constrained level of 5%. This result raises the question: under what structural characteristics of hybrid carrageenan can κ-carrageenan be excluded entirely, allowing for a simpler blend formulation? Although the current model performed well in predicting viscosity, it does not capture other quality parameters such as phase separation or long-term stability, therefore it is not possible to answer this question without a sensory study. Future work could extend this approach by incorporating three-way interaction terms, exploring additional quality descriptors (e.g., infrared spectroscopy-based structural features such as sulfate content), or modeling other functional outputs like yield stress, gel stability, and sensory performance. A smaller, targeted mixture design using alternative quality metrics in milk-like matrices could provide additional insight into carrageenan structure–function relationships.

Ultimately, this study demonstrates how DoE and predictive modeling can reduce trial-and-error in food formulation, enhance resource efficiency, and support more sustainable development of dairy suspensions stabilized by seaweed-derived hydrocolloids.

ACKNOWLEDGEMENT

We gratefully acknowledge our colleagues at CP Kelco for generously sharing their extensive expertise in carrageenan. Special thanks are extended to Dr. Heidi Pedersen, Wencke Dybvik Henriksen, Thomas Worm, and Soeren Biensoe for their valuable discussions on carrageenan chemistry, functionality, and applications. We also sincerely thank Prof. Franciscus Winfried J. van der Berg for his insightful input on experimental design strategies. Finally, we extend our appreciation to the QC Carrageenan laboratory personnel for their assistance with viscosity measurements and technical support.

FUNDING DETAILS

This work was supported by CP Kelco ApS as part of the industrial PhD program of the Innovation Fund Denmark (grant number: 104400009B).

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.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work the authors used OpenAI's ChatGPT in order to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

4.1 REFERENCES

BeMiller, J. N. (2019a). Carrageenans. *Carbohydrate Chemistry for Food Scientists*, 279–291. https://doi.org/10.1016/B978-0-12-812069-9.00013-3

BeMiller, J. N. (2019b). *Chapter 13 in Carbohydrate chemistry for food scientists* (Third edition.) [Book]. Woodhead Publishing, an imprint of Elsevier.

Bixler, H. J., Johndro, K., & Falshaw, R. (2001). Kappa-2 carrageenan: structure and performance of commercial extracts: II. Performance in two simulated dairy applications. *Food Hydrocolloids*, *15*(4–6), 619–630. https://doi.org/10.1016/S0268-005X(01)00047-9

Campo, V. L., Kawano, D. F., Silva, D. B. da, & Carvalho, I. (2009). Carrageenans: Biological properties, chemical modifications and structural analysis – A review. *Carbohydrate Polymers*, *77*(2), 167–180. https://doi.org/10.1016/J.CARBPOL.2009.01.020

Colusse, G. A., Carneiro, J., Duarte, M. E. R., Ranga Rao, A., Ravishankar, G. A., de Carvalho, J. C., & Noseda, M. D. (2022). Challenges and Recent Progress in Seaweed Polysaccharides for Industrial Purposes [Bookitem]. In G. A. Ravishankar & A. Ranga Rao (Eds.), *Sustainable Global Resources of Seaweeds Volume 2* (pp. 411–431). Springer International Publishing. https://doi.org/10.1007/978-3-030-92174-3_22

De Ruiter, G. A., & Rudolph, B. (1997). Carrageenan biotechnology. *Trends in Food Science and Technology*, *8*(12), 389–395. https://doi.org/10.1016/S0924-2244(97)01091-1

Falshaw, R., Bixler, H. J., & Johndrob, K. (2001). Structure and performance of commercial kappa-2 carrageenan extracts: I. Structure analysis. *Food Hydrocolloids*, *15*(4–6), 441–452. https://doi.org/10.1016/S0268-005X(01)00066-2

Montgomery, D. C. (2013). *Design and analysis of experiments* (D. C. Montgomery, Ed.; Eighth edition.) [Book]. John Wiley & Sons, Inc.

Montgomery Douglas C. (2013a). Introduction to Factorial Designs [Bookitem]. In *Design and Analysis of Experiments* (8th Edition, pp. 1–1). John Wiley & Sons.

Montgomery Douglas C. (2013b). Response Surface Methods and Designs [Bookitem]. In *Design and Analysis of Experiments* (8th Edition, pp. 1–1). John Wiley & Sons.

Mykhalevych, O., Stapelfeldt, H., Marini, F., & Bro, R. (2025). Chemometric insights into milk-carrageenan breaking and gel strength. *Food Hydrocolloids*, *158*, 110544. https://doi.org/10.1016/J.FOODHYD.2024.110544

Nocedal, J., & Wright, S. J. (2006). Sequential Quadratic Programming. In T. V. Mikosch, S. I. Resnick, & S. M. Robinson (Eds.), *Numerical Optimization* (Second Edition, pp. 529–561). Springer New York.

The Mathworks, I. (n.d.). *MATLAB function "fmincon."* Retrieved March 22, 2025, from https://se.mathworks.com/help/optim/ug/fmincon.html#responsive_offcanvas

Tijssen, R. L. M., Canabady-Rochelle, L. S., & Mellema, M. (2007). Gelation upon Long Storage of Milk Drinks with Carrageenan [Article]. *Journal of Dairy Science*, *90*(6), 2604–2611. https://doi.org/10.3168/jds.2006-854

5 APPENDIX A

Table 6 Factorial and mixture design for carrageenan formulation for stabilization of chocolate milk with viscosity as the response factor.

Quality Level			Concentration				
A	B	C	D%	A%	B%	C%	Vsicosity
1	-1	-1	20	0	40	40	57.53
-1	-1	-1	23	10	40	27	35.56
-1	-1	-1	35	0	25	40	11.40
0	-1	-1	15	20	25	40	61.92
-1	1	-1	35	0	40	25	22.67
1	1	1	29	20	40	11	22.38
1	-1	0	15	20	25	40	33.80
0	-1	1	25	20	40	15	24.72
0	1	0	23	8	40	29	40.98
-1	-1	0	19	20	40	21	38.78
1	0	1	23	10	40	27	35.56
0	0	-1	15	9	40	35	40.25
0	1	0	24	20	40	16	34.10
1	1	1	35	20	5	40	14.77
1	-1	-1	35	20	5	40	1.00
1	-1	-1	15	12	32	40	40.98
0	-1	1	35	20	26	19	15.06
0	1	-1	35	20	23	23	6.27
0	-1	0	35	20	5	40	8.32
1	-1	0	35	6	40	19	16.52
1	-1	0	35	0	25	40	15.94
1	0	0	35	20	40	5	7.74
-1	1	0	35	0	40	25	23.85
-1	0	0	15	20	25	40	39.66
-1	-1	-1	23	20	17	40	23.55
0	-1	-1	35	20	5	40	1.29
0	1	1	35	0	40	25	26.78
-1	-1	0	35	5	27	33	16.82
-1	1	1	15	5	40	40	82.43
-1	0	-1	15	20	25	40	41.42
0	1	0	35	0	25	40	20.33
0	0	0	35	6	19	40	13.01
1	1	1	15	16	30	40	61.92
-1	0	0	35	0	40	25	13.59

0	1	1	20	0	40	40	64.85
-1	0	0	28	12	40	20	25.02
-1	1	0	28	0	40	33	45.81
0	1	-1	32	0	40	28	34.68
0	0	1	35	20	40	5	3.93
1	0	0	29	0	40	31	23.55
0	1	-1	15	19	40	27	53.14
0	-1	1	34	16	10	40	16.23
-1	1	0	15	20	31	34	47.28
-1	0	1	35	0	40	25	21.21
1	0	-1	35	20	15	29	4.81
0	-1	0	25	8	27	40	38.49
1	1	0	35	6	19	40	18.28
-1	1	1	35	20	19	26	19.45
-1	1	1	25	20	15	40	44.93
1	1	-1	23	20	28	29	25.60
-1	1	1	35	20	40	5	23.26
1	-1	1	35	20	40	5	11.11
1	0	0	35	10	31	24	13.59
0	0	0	35	20	40	5	11.84
-1	1	-1	25	20	40	15	34.39
-1	0	1	35	20	5	40	20.04
0	-1	1	35	11	32	23	23.85
0	-1	-1	35	0	40	25	17.99
1	-1	-1	35	20	26	19	8.32
-1	-1	-1	35	20	40	5	20.92
1	-1	1	35	0	40	25	32.63
1	-1	0	27	17	19	37	20.92
1	-1	1	21	20	19	40	41.42
-1	-1	1	15	10	36	40	80.96
0	0	-1	21	20	19	40	24.72
0	-1	1	15	5	40	40	98.54
0	0	0	20	0	40	40	63.39
0	0	-1	35	20	5	40	3.05
-1	-1	1	15	20	25	40	65.44
-1	-1	-1	15	20	40	25	52.55
0	0	1	35	20	16	30	8.62
0	0	1	21	20	19	40	38.78
1	1	1	35	0	40	25	40.83
-1	0	0	35	20	26	19	7.59
-1	0	-1	15	5	40	40	43.62
1	0	1	35	0	25	40	15.64
0	1	1	35	20	40	5	23.85

-1	0	0	15	20	40	25	44.35
1	-1	1	35	20	12	33	12.13
0	0	1	35	4	21	40	16.82
-1	1	0	35	20	5	40	5.98
1	0	-1	35	0	40	25	5.98
1	0	1	35	20	5	40	10.08
1	0	-1	27	5	28	40	27.36
1	-1	-1	23	20	40	17	23.85
0	0	1	15	20	40	25	32.93
-1	1	-1	20	0	40	40	54.31
1	-1	-1	35	7	27	32	8.91
0	1	0	15	20	25	40	37.61
-1	0	1	15	20	40	25	42.88
1	0	1	15	20	25	40	50.21
-1	1	1	35	0	25	40	41.42
0	1	1	35	20	5	40	17.99
-1	0	-1	35	0	33	33	11.54
-1	0	0	35	0	25	40	14.77
-1	1	-1	35	10	32	23	14.33
-1	0	-1	35	20	40	5	8.18
-1	1	0	23	8	29	40	39.37
1	1	-1	35	20	40	5	14.47
0	0	-1	25	20	40	15	14.47
1	1	1	15	20	40	25	48.01
-1	-1	0	35	20	40	5	12.13
1	-1	0	35	20	5	40	3.93
-1	-1	-1	35	0	40	25	16.38
1	1	1	35	11	25	29	15.64
-1	1	0	35	20	40	5	16.23
-1	0	1	27	14	31	28	35.56
-1	-1	-1	35	20	19	26	5.83
0	1	0	35	20	5	40	6.13
-1	-1	1	25	20	30	25	33.66
1	1	-1	35	0	25	40	12.13
-1	-1	0	35	20	5	40	9.64
0	-1	-1	15	20	25	40	43.76
0	-1	0	35	0	40	25	27.36
0	-1	0	15	20	40	25	40.54
1	0	1	15	5	40	40	72.17
1	0	-1	35	20	40	5	12.13
1	1	0	15	20	40	25	44.35
0	0	1	31	0	40	30	38.49
1	1	-1	35	18	7	40	2.76

-1	0	1	35	20	40	5	16.52
0	0	0	15	20	25	40	44.93
1	1	-1	15	20	25	40	31.02
1	-1	0	35	20	40	5	7.15
0	-1	-1	15	5	40	40	61.92
0	1	-1	35	20	40	5	13.59
0	0	-1	35	8	40	17	7.15
1	1	0	20	0	40	40	61.92
1	-1	1	15	20	40	25	47.28
1	-1	0	15	5	40	40	62.51
0	1	0	35	6	40	19	22.38
-1	0	0	35	20	5	40	4.95
-1	-1	1	20	0	40	40	66.76
0	-1	1	34	16	10	40	21.21
-1	0	1	20	0	40	40	64.56
1	0	1	35	20	30	16	8.62
1	0	-1	15	20	40	25	35.56
-1	1	1	17	20	40	23	56.07
-1	-1	1	35	12	40	13	17.26
1	1	0	35	16	40	9	1.44
0	-1	-1	35	20	40	5	14.77
0	0	0	29	20	23	28	16.52
-1	1	-1	35	20	5	40	17.26
1	0	0	35	20	5	40	7.74
0	0	-1	35	0	25	40	8.62
1	1	1	24	0	35	40	56.07
0	1	1	15	20	40	25	58.99
1	1	-1	15	5	40	40	55.48
1	0	0	15	15	29	40	35.12
1	0	0	23	20	40	17	26.78
-1	-1	1	35	20	5	40	23.85
0	-1	0	35	20	40	5	11.54
1	1	0	35	20	23	23	7.44
-1	-1	0	15	5	40	40	100.00
0	1	-1	35	12	13	40	5.25
0	-1	1	35	0	25	40	26.63
0	1	-1	20	2	38	40	58.26
-1	0	-1	35	18	7	40	4.81
0	-1	0	35	20	23	23	13.59

487

7 APPENDIX B

Table 7 Estimated coefficients for ANOVA model with main effects and two-way interaction.

	Estimate	SE	tStat	p -value
Intercepts	-1445.800	1160.460	-1.246	0.215
Quality A	-1167.696	1415.634	-0.825	0.411
Quality B	-837.531	1389.542	-0.603	0.548
Quality C	1095.320	1418.012	0.772	0.441
D %	14.179	11.661	1.216	0.226
A %	15.031	11.602	1.295	0.197
B %	15.796	11.629	1.358	0.177
C %	15.962	11.659	1.369	0.173
Quality A : Quality B	0.207	0.845	0.245	0.807
Quality A : Quality C	-0.293	0.861	-0.341	0.734
Quality B : Quality C	0.181	0.839	0.216	0.829
Quality A : D %	11.746	14.164	0.829	0.408
Quality B : D %	8.495	13.902	0.611	0.542
Quality C : D %	-10.995	14.187	-0.775	0.440
Quality A : A %	11.527	14.156	0.814	0.417
Quality B : A %	8.345	13.898	0.600	0.549
Quality C : A %	-11.019	14.184	-0.777	0.439
D % : A %	0.012	0.023	0.523	0.602
Quality A : B %	11.648	14.164	0.822	0.412
Quality B : B %	8.369	13.901	0.602	0.548
Quality C : B %	-10.924	14.185	-0.770	0.443
D % :B %	-0.021	0.017	-1.254	0.212
A % : B %	-0.025	0.017	-1.430	0.155
Quality A : C %	11.636	14.165	0.821	0.413
Quality B : C %	8.311	13.907	0.598	0.551
Quality C : C %	-10.747	14.191	-0.757	0.450
D % : C %	-0.031	0.016	-1.902	0.059
A % : C %	-0.021	0.018	-1.194	0.235
B % : C %	-0.002	0.008	-0.237	0.813

Table 8 Estimated coefficients for reduced ANOVA model with main effects and some two-way interactions.

	Estimate	SE	tStat	p -value
Intercepts	-1435	1119	-1.28	0.202
K. alvarezii	-1195	1366	-0.87	0.383

C. crispus	-3.9	2.3	-1.67	0.096
G. radula	1166	1374	0.85	0.397
Sugar %	14.4	11.3	1.28	0.204
K. alvarezii %	15.5	11.2	1.39	0.168
C. crispus %	15.7	11.2	1.40	0.164
G. radula %	15.8	11.2	1.41	0.160
K. alvarezii : Sugar %	12.0	13.7	0.88	0.381
C. crispus : Sugar %	0.2	0.1	1.95	0.053
G. radula : Sugar %	-11.7	13.7	-0.85	0.396
K. alvarezii : K. alvarezii %	11.8	13.7	0.86	0.389
G. radula : *K. alvarezii* %	-11.7	13.7	-0.85	0.395
K. alvarezii : *C. crispus* %	11.9	13.7	0.87	0.385
G. radula : *C. crispus* %	-11.6	13.7	-0.85	0.399
Sugar % : *C. crispus* %	-0.026	0.013	-2.09	0.038
K. alvarezii % : *C. crispus* %	-0.031	0.013	-2.48	0.014
K. alvarezii : *G. radula* %	11.9	13.7	0.87	0.385
G. radula : *G. radula* %	-11.5	13.8	-0.83	0.406
Sugar % : *G. radula* %	- 0.036	0.012	-2.98	0.003
K. alvarezii % : *G. radula* %	-0.027	0.013	-2.14	0.034

492

8 APPENDIX C

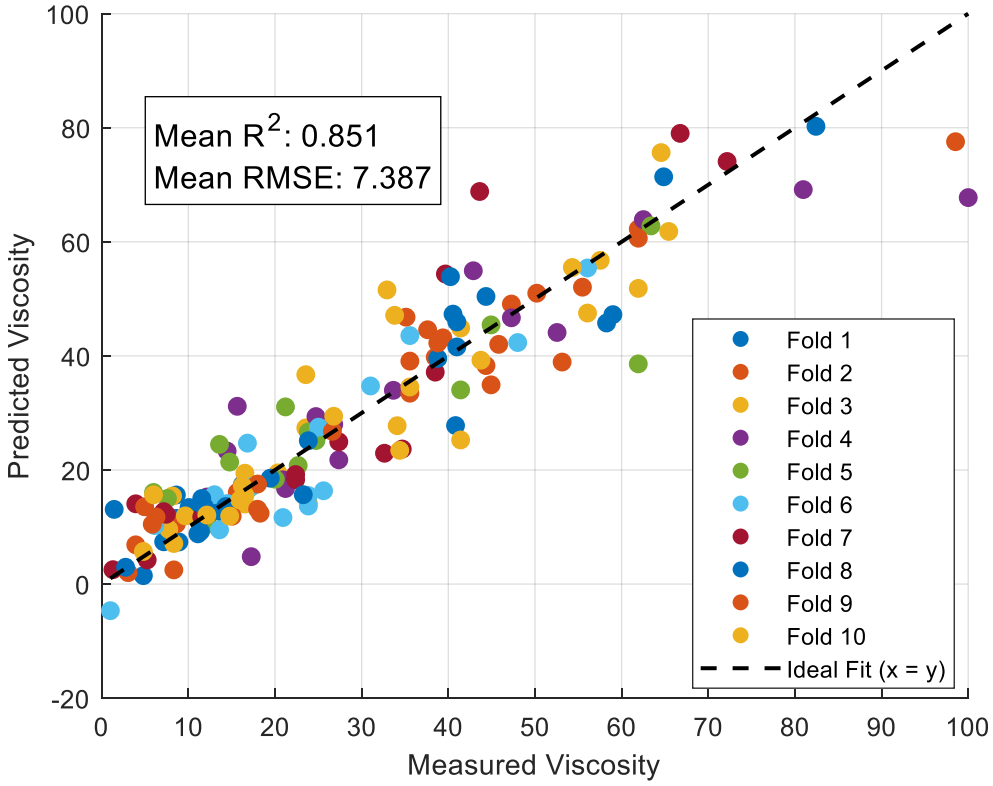


Figure 5 Plot of measured vs. predicted by the model viscosity values colored by cross-validation folds colors to show distribution of each value in each fold within the dataset. The coefficient of determination was 0.85 indicating that a large proportion of variance of chocolate milk viscosity can be adequately explained by carrageenan structure and concentration of different types in the formulations. The small RMSE of 7.4 compared to the viscosity range of 1-100 indicates an acceptable prediction accuracy.

.

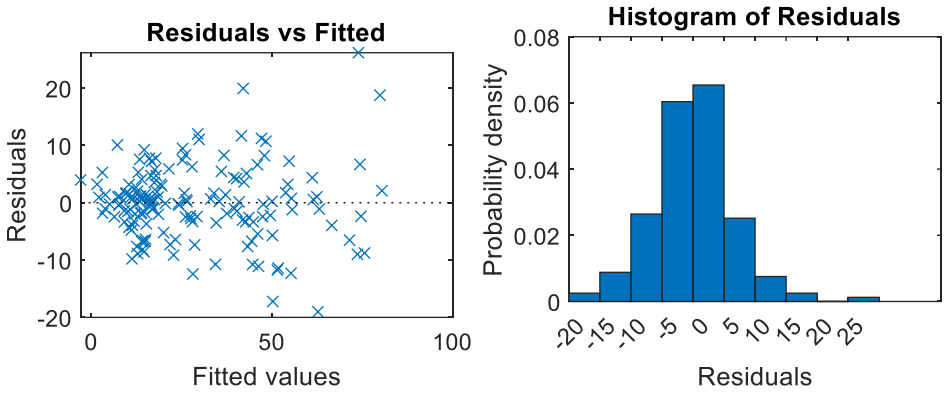


Figure 6 Selected plots for model statistics for the final reduced model show good model behavior, i.e. randomly distributed residuals centered around zero in the residual vs. fitted plot (left top plot); a normal distribution curve centered at zero (right plot).

CP Kelco
A HUBER COMPANY

UNIVERSITY OF COPENHAGEN
DEPARTMENT OF FOOD SCIENCE

Data fusion for prediction of gel and breaking strength for milk gels of carrageenan

Oksana Mykhalevych[1,2], Rasmus Bro[2], Astrid Benie[1], Henrik Stapelfeldt[1] and Federico Marini[3]

[1] CP Kelco, 4623 Lille Skensved, Denmark

[2]Department of Food Science, University of Copenhagen, 1958 Frederiksberg, Denmark

[3]Department of Chemistry, University of Rome "La Sapienza", 00185 Rome, Italy

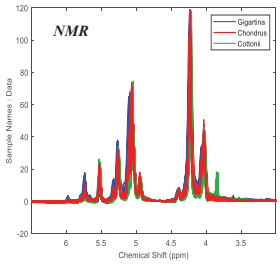
oksana@food.ku.dk

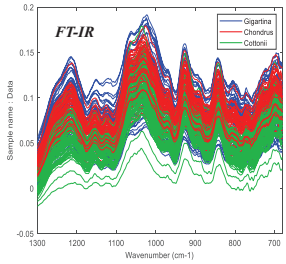
The aim

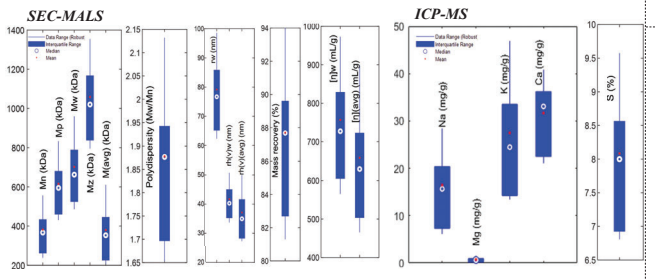
1. To discover a direct link between carrageenan's chemical structure and its effects on milk gel properties
2. To use it to develop a multi- or single block regression for prediction of the gel properties.

Method

Chemical structure of 200 carrageenans in 4 Data blocks

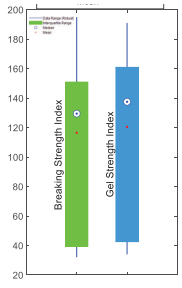






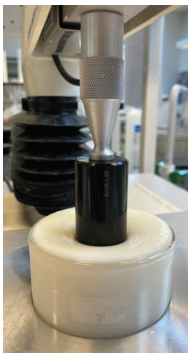
Single and multi-block analysis:

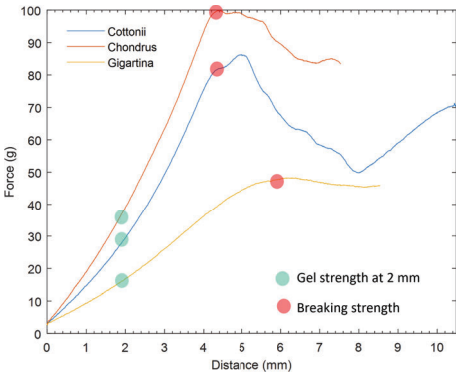
PLS
SO-PLS
MB-PLS
PO-PLS
LW-SO-PLS



Reference values

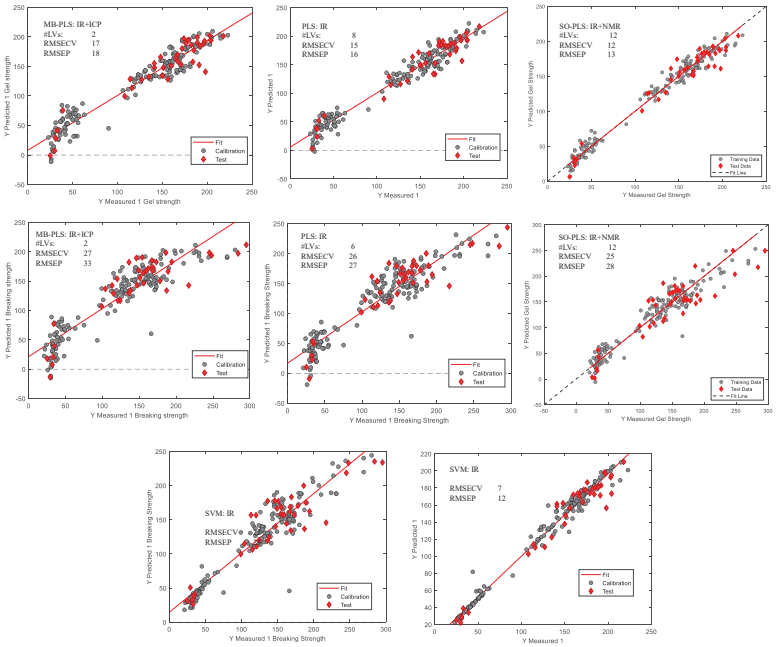
Background





Left Figure : Milk gel sample during characterization by a texture analyzer, TAXT. **Right Figure:** Typical texture analyzer curves, Force (g) vs distance (mm), for milk gels prepared with carrageenan extracted from Cottonii, Chondrus and Gigartina seaweed types. Gel strength is measured as force at 2 mm, offering insights into gels stability. While breaking strength is a measure for the gel's firmness and is read at rupture as indicated.

Results



Conclusion

1. The gel and breaking strength of carrageenan milk gels are influenced by specific structural characteristics, including the position of sulfate esters, the presence of an anhydrogalactose residue bridge, and the ionic composition.
2. PLSR based on FT-IR spectra can be employed to forecast gel properties and breaking strength, although nonlinear techniques tend to yield superior results.

Acknowledgements

The authors would like to gratefully acknowledge the Industrial PhD grant made available by the Innovation fund Denmark to partly cover the expenses. This PhD project is a collaboration between CP Kelco and the Department of Food Science, University of Copenhagen.

Innovation Fund Denmark