Carbohydrate Polymers 333 (2024) 121975

Contents lists available at ScienceDirect

Carbohydrate Polymers

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

On the origin of optical rotation changes during the κ-carrageenan disorder-to-order transition

B.P. Westberry [a,b,c,*], M. Rio [d,e], M.R. Waterland [a,b], M.A.K. Williams [a,b]

[a] School of Natural Sciences, Massey University, Palmerston North, New Zealand
[b] MacDiarmid Institute for Advanced Materials and Nanotechnology, Wellington, New Zealand
[c] Fonterra Research and Development Centre, Private Bag 11029, Dairy Farm Road, Palmerston North, New Zealand
[d] National Institute of Water and Atmospheric Research (NIWA), Wellington, New Zealand
[e] Science Infrastructure (NeSI), Auckland, New Zealand

ARTICLE INFO

Keywords:
Carrageenan
Optical rotation
Molecular dynamics
Density-functional theory
Machine learning

ABSTRACT

It is well established that solutions of both polymeric and oligomeric κ-carrageenan exhibit a clear change in optical rotation (OR), in concert with gel-formation for polymeric samples, as the solution is cooled in the presence of certain ions. The canonical interpretation – that this OR change reflects a 'coil-to-helix transition' in single chains – has seemed unambiguous; the solution- or 'disordered'-state structure has ubiquitously been assumed to be a 'random coil', and the helical nature of carrageenan in the solid-state was settled in the 1970s. However, recent work has found that κ-carrageenan contains substantial helical secondary structure elements in the disordered-state, raising doubts over the validity of this interpretation. To investigate the origins of the OR, density-functional theory calculations were conducted using atomic models of κ-carrageenan oligomers. Changes were found to occur in the predicted OR owing purely to dimerization of chains, and - together with the additional effects of slight changes in conformation that occur when separated helical chains form double-helices - the predicted OR changes are qualitatively consistent with experimental results. These findings contribute to a growing body of evidence that the carrageenan 'disorder-to-order' transition is a cooperative process, and have further implications for the interpretation of OR changes demonstrated by macromolecules in general.

1. Introduction

Carrageenans are a class of polysaccharides characterized by a linear primary structure of repeating (1→3)-linked β-D-galactopyranose residues ('G' in this work) and (1→4)-linked α-D-galactopyranose residues ('A' in this work), as shown in Fig. 1. Different types of carrageenan are differentiated according to their degree of sulfation, with κ-carrageenan (oligomers of which are the focus of this study) carrying one negatively charged sulfate per 3-linked galactose reside, and a 3,6-anhydro bridge on each 4-linked galactose reside.

Carrageenan solutions find useful applications in the food, pharmaceutical, and medical industries due to their tunable viscoelasticity (Dea, 1993; Yegappan et al., 2018). The viscoelastic properties that carrageenans impart to their solutions and gels depend on their conformational state, which at high temperatures and/or low ionic strengths is typically described as a 'disordered-state', to distinguish from the 'ordered-state' existing at low temperatures in the presence of certain ions. These two states have historically been assigned as a 'random coil' and 'helix' respectively, based on optical rotation (OR) measurements (McKinnon et al., 1969) and comparison to the 'coil-to-helix' transition known at the time to occur in many polypeptides (Yang & Doty, 1957).

OR is the tendency for a solution of chiral molecules to rotate the plane of incident linearly polarized radiation as it traverses the sample. It is usually measured in degrees and normalized for path length and either the solute's molar mass (as is the case in this study), or molar concentration. The measurement is sensitive to molecular conformation (Cahn et al., 1966), and for carrageenan it has been used to monitor structural changes as a function of temperature (Bosco et al., 2005; Mangione et al., 2003; McKinnon et al., 1969; Rees et al., 1969; Takemasa & Nishinari, 2004), often while also varying: salt concentration (Nilsson & Piculell, 1989; Paoletti et al., 1985; Slootmaekers et al., 1988; Snoeren & Payens, 1976) and/or salt species (Austen et al., 1988; Ciancia et al., 1997; Morris et al., 1980; Rochas & Rinaudo, 1980); polymer concentration (Bryce et al., 1982); solvent type (Cardoso &

* Corresponding author at: School of Natural Sciences, Massey University, Palmerston North, New Zealand.
E-mail address: ben.westberry@fonterra.com (B.P. Westberry).

https://doi.org/10.1016/j.carbpol.2024.121975
Received 10 December 2023; Received in revised form 8 February 2024; Accepted 20 February 2024
Available online 24 February 2024

Sabadini, 2010; Ramakrishnan & Prud'homme, 2000); chain composition (Dea et al., 1972; Rochas et al., 1989; van de Velde et al., 2005, 2002); molecular weight (Hjerde et al., 1998; Rees et al., 1982; Rochas et al., 1983; Viebke et al., 1994); and time after ion introduction (Norton et al., 1983). While the use of different wavelengths of incident light and varied unit conventions complicate the quantitative comparison of the entire dataset of existing results, two universal conclusions can be made: both the disordered- and ordered-states are dextrorotatory (i.e. exhibit positive OR), and the OR of the ordered-state is higher than that of the disordered-state. Measurements carried out on κ-carrageenan solutions that are made using the sodium D-line ($\lambda = 589$ nm) find OR values ($[\alpha]_D$) of approximately 45 and 73 $deg.\cdot mL\cdot g^{-1}dm^{-1}$ for the disordered- and ordered-states, respectively (Grasdalen & Smidsrød, 1981; Mangione et al., 2003; Smidsrød et al., 1980).

Clearly, OR is central to the carrageenan story, and attempts have been made to understand its dependence on the molecular conformation of the polysaccharide for many decades. Such an attempt was first made by Rees et al. (1970), where the OR for ι-carrageenan was modelled empirically as the sum of contributions from monomers plus additional terms related to the orientation of glycosidic linkages. More recently, Schafer and Stevens (1995) built a coupled-oscillator model to calculate the circular dichroic rotational strengths for a given molecular geometry, from which OR at a specific wavelength can be obtained via a Kramers-Kronig transform. This model was applied to a small set of unsulfated carrageenan conformers, and found that double-helices and more tightly-wound single-helices produced higher OR values compared to single, more elongated chains. Herein, with the use of molecular dynamics (MD) and density-functional theory (DFT) calculations utilizing atomic coordinates of κ-carrageenan oligomers, we investigate whether the OR changes observed during the disorder-to-order transition are consistent with a molecular model of the transition being two single-helices forming a double-helix, as suggested by recent x-ray scattering and MD studies (Westberry et al., 2022; Westberry et al., 2023). To understand which molecular features most influence the OR calculation, a battery of machine learning (ML) models are then trained on a set of OR values calculated using the atomic coordinates of a large set of conformers.

2. Methods

Due to the thermal fluctuations inherent in molecular structures, the ensemble-averaged OR over many conformers was calculated. Each conformer was drawn from one of the three 1 μs-long MD trajectories, each containing two κ-carrageenan oligomers. During MD simulations, the two chains either remained as separate chains (SC) or spontaneously formed double-helices (DH) as described in detail elsewhere (Westberry et al., 2023, 2022). In the case of the simulation of SC, the two chains maintained a separation greater than 3 nm, with this simulation being used to model the disordered-state. The two other simulations yielded DH arrangements, and were used to model the ordered-state. Towards the end of each of the three MD simulations, the atomic coordinates of both κ-carrageenan oligomers within the simulation box were extracted at 1 ns intervals, and subsequently used as input for Gaussian 09 quantum chemistry software (Frisch et al., 2009) in order to calculate the OR, as described below.

Due to the computationally expensive nature of DFT calculations, it was not possible to calculate the OR from the entirety of both simulated oligomers (676 atoms). Instead, the coordinates for fragments from each of the two oligomers within each simulation box were extracted from the larger structures. Specifically, glycosidic linkages were severed *in-silico*, with hydroxyl moieties added at the bare reducing ends, and equatorial hydrogens at the non-reducing ends of the oligomeric fragments. To understand the ramifications of approximating the OR of the polymer by calculations performed using the atomic coordinates of fragments, calculations were performed for various fragment lengths. These ranged in length from individual residues, up to 7-residue long fragments in the case of SC simulations and 5-residue long fragments in the case of DH simulations (10 residues in total).

Many DFT calculations were performed simultaneously on both fragments extracted from the two oligomers within each MD simulation, as shown in Fig. 2(a) and (b) for SC and DH calculations, respectively. This was necessary to capture any interaction between fragments, but had the added benefit of maintaining an even ratio of residue and linkage types. However, some DFT calculations were performed on each of the two fragments in isolation (Fig. 2(c)). In these cases, the OR values calculated for the two fragments both in isolation were used to form a mass-weighted average, which could be compared to the OR value calculated using both fragments simultaneously.

Atomic coordinates for the fragment(s) were loaded into Gaussian 09 input files for use in the calculation of OR using the 'Polar = OptRot' keyword and a perturbing frequency of 589 nm. The majority of calculations were performed using the popular 6-31+G(d,p) basis set and B3LYP hybrid functional. For comparison, a small subset of calculations using the aug-cc-pVDZ basis set and CAM-B3LYP hybrid functional were also conducted. Solvation effects were accounted for using the Polarizable Continuum Model with water set as the solvent. The OR calculations carried out using the 6-31+G(d,p)/B3LYP combination are summarized in Table 1.

3. Results

Fig. 3 shows that the ensemble-averaged OR calculated for both SC and DH structures of all fragment lengths was dextrorotatory (with the exception of 3-residue SC fragments whose mean OR was zero).

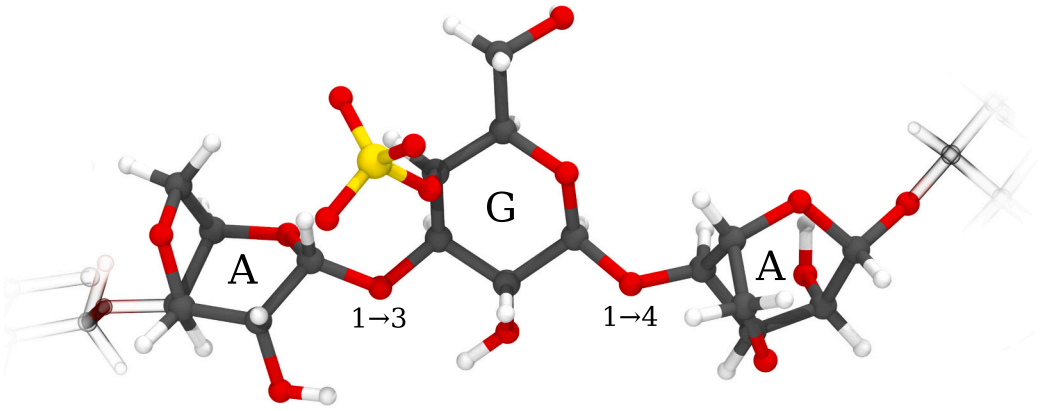

Fig. 1. Idealised κ-carrageenan structure, with α-D-galactopyranos labelled 'A' and β-D-galactopyranose labelled 'G'. 1→3 and 1→4 glycosidic linkages are also appropriately labelled.

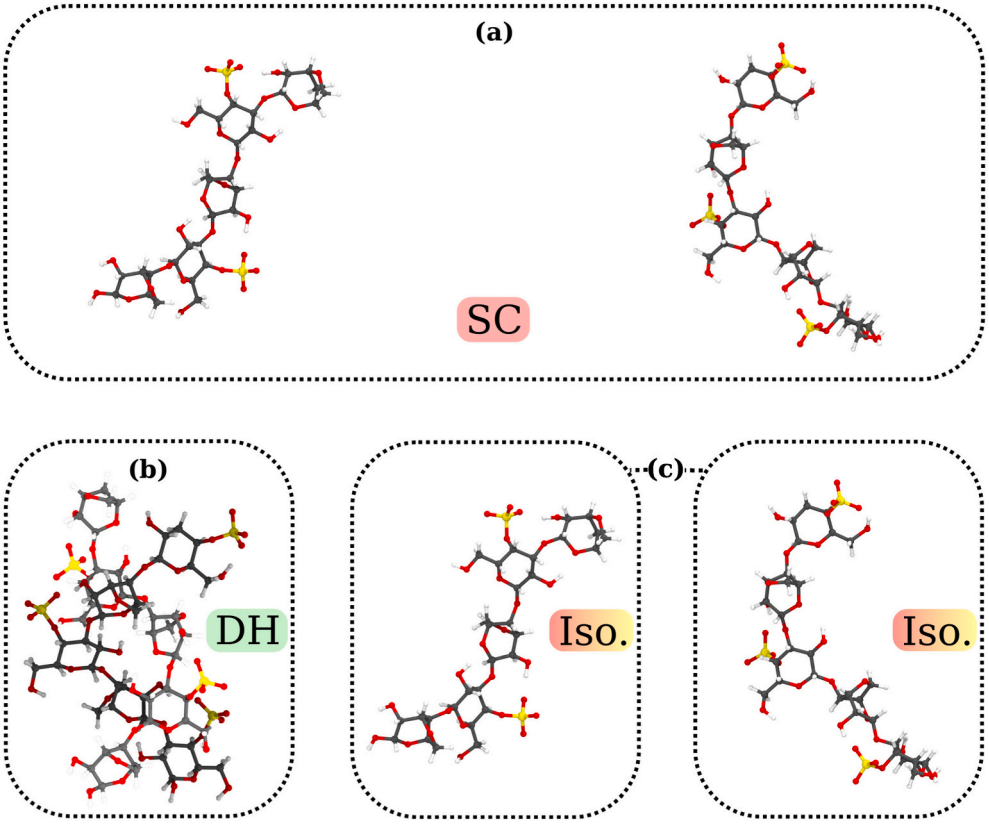

Fig. 2. Density-functional theory calculations were performed on (a) two fragments from separate, non-interacting chains (SC), (b) two fragments from double-helices (DH), or (c) either SC or DH where the calculation was performed on each fragment in isolation (Iso; figure shows calculation performed on both chains in (a) separately).

Table 1
The number of optical rotation calculations performed for double-helices (DH) and separate chains (SC) for various fragment lengths (measured by number of residues) using the B3LYP hybrid functional and 6-31+G(d,p) basis set.[a] DH calculations each involve two fragments, and hence twice as many residues as specified by the length.[b] SC calculations are a mix of calculations performed simultaneously on both fragments and calculations made for each fragment in isolation, with the number of calculations specifying the number of fragment pairs.

Length	DH[a]	SC[b]
1	100	151
2	51	71
3	42	573
4	16	46
5	11	41
6	0	21
7	0	12

Furthermore, the OR calculated for DH fragments was significantly higher than that for SC fragments at all lengths. This was irrespective of which of the two distinct DH simulations was used, and whether SC calculations were performed on fragments in isolation, or simultaneously on both fragments. For fragments longer than one residue, the dependence of the mean calculated OR on fragment length was approximately sigmoidal, with a fit to this functional form plateauing for longer fragment lengths at approximately 18 and 92 $deg.\cdot mL\cdot g^{-1}dm^{-1}$ for SC and DH structures, respectively. At a fragment length of one residue, SC and DH differ only in the proximity between the two fragments, and yet their mean calculated OR differs by 27 $deg.\cdot mL\cdot g^{-1}dm^{-1}$.

Most OR calculations for DH were performed simultaneously on both of the fragments, in order to capture any interaction between them.

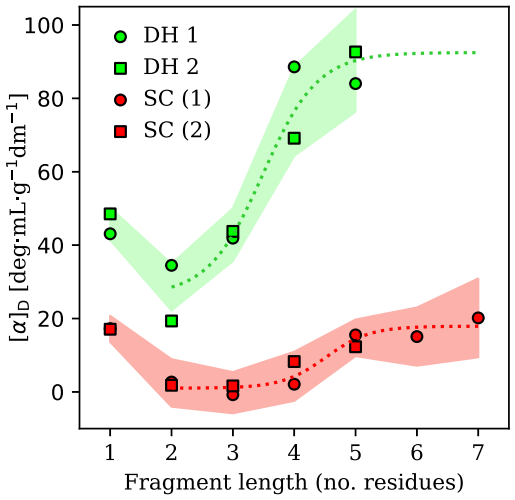

Fig. 3. Mean calculated optical rotation versus fragment length for both double-helix simulations (DH 1 and DH 2), as well as one simulation of separate chains where calculations were performed on fragments in isolation, or on both fragments simultaneously (SC (1) and SC (2), respectively). Red and green shaded areas represent 95 % confidence intervals for all SC and DH calculations, respectively. Dashed lines indicate sigmoidal fits to the data at fragment lengths above single residues. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

However, additional calculations were performed using a subset of 3-residue long DH fragments in isolation - such that they each maintained the conformation assumed while in the DH, but without the proximity of the neighboring fragment. Using the mass-weighted

average OR for each pair, the mean calculated OR for DH without inter-fragment interactions present was 25 ± 6 $deg.\cdot mL\cdot g^{-1}dm^{-1}$. This was reduced relative to the mean of 43 ± 7 $deg.\cdot mL\cdot g^{-1}dm^{-1}$ calculated for DH with inter-fragment interactions present, but still significantly higher than the mean of 0 ± 6 $deg.\cdot mL\cdot g^{-1}dm^{-1}$ calculated for SC fragments (Fig. 4).

OR for one pair of 3-residue long SC fragments and one pair of 3-residue long DH fragments was also recalculated using different combinations of functional (B3LYP and CAM-B3LYP) and basis set (6-31+G(d,p) and aug-cc-pVDZ) in an attempt to assess the robustness of the results. While some quantitative differences were found in the results when employing different basis sets and functionals, the results shown in Table 2 demonstrate that the main findings – that the OR calculated for both SC and DH is dextrorotatory, and that the OR for DH is higher than that for SC – still hold regardless. The difference between the OR values calculated for SC fragments versus those found for the DH configuration was largest when using the combination of B3LYP and 6-31+G(d,p) ($\Delta[\alpha]_D = 44.51\, deg.\cdot mL\cdot g^{-1}dm^{-1}$), and smallest when using the combination of CAM-B3LYP and aug-cc-pVDZ ($\Delta[\alpha]_D = 34.17\, deg.\cdot mL\cdot g^{-1}dm^{-1}$). However, in all cases the difference between the results obtained from SC and DH calculations was much larger than the variation found between OR values calculated for the same structure using different combinations of functional and basis set.

4. Analysis

In an endeavor to understand the structural attributes exerting the greatest impact on the calculated OR, multiple supervised ML models were devised. These included elastic net linear regression models (EN) (Zou & Hastie, 2005), random forest models (RF) (Breiman, 2001), extremely randomized trees models (ERT) (Geurts et al., 2006), and multilayer perceptron with rectified linear units and dropout (MLP) (Srivastava et al., 2014) models. All models were implemented using the Scikit-Learn Python package (Pedregosa et al., 2011). Each ML model used conformers of 3-residue long fragments (fragment 1 or 2 referring to fragments with a central G or A residue, respectively), with the atomic coordinates used to derive an array of molecular features. These features were either Coulomb matrices (Rupp et al., 2012), or a variant of the bag-of-bonds (BoB) representation including the values of every bond length, bond angle, and dihedral angle present in the conformer (Huang & Von Lilienfeld, 2016), which will simply be referred to hereafter as BoB. OR values for each conformer, as calculated on each fragment in isolation using DFT, served as the target variable for each model. For each combination of ML algorithm and type of molecular features, separate models were built for fragments 1 and 2. In total, 16 ML models were built (4 ML algorithms × 2 types of molecular features × 2 fragment types).

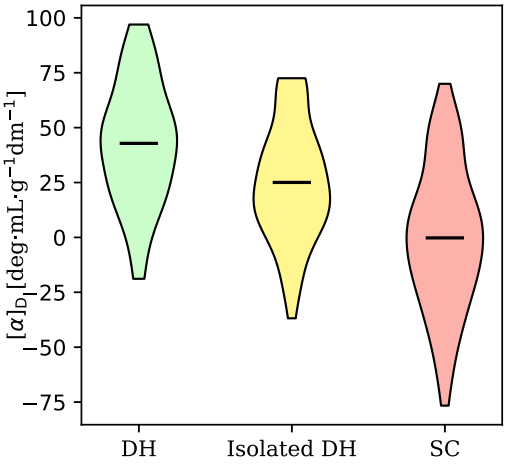

Fig. 4. Distribution of optical rotation values calculated for 3-residue double-helix structures (DH), the same set of double-helix structures but with the calculation performed on each of the two fragments in isolation (Isolated DH), and for 3-residue fragments from simulations of separate chains (SC). Horizontal lines indicate distribution means.

Table 2
Optical rotation (OR) values calculated using various combinations of functional and basis set for one pair of 3-residue fragments of separate chains (SC), and one pair of 3-residue double-helical fragments (DH).

Structure	Method	Basis set	OR
SC	B3LYP	6-31+G(d,p)	0.35
SC	B3LYP	aug-cc-pVDZ	4.17
SC	CAM-B3LYP	6-31+G(d,p)	3.97
SC	CAM-B3LYP	aug-cc-pVDZ	5.92
DH	B3LYP	6-31+G(d,p)	44.86
DH	B3LYP	aug-cc-pVDZ	43.96
DH	CAM-B3LYP	6-31+G(d,p)	38.83
DH	CAM-B3LYP	aug-cc-pVDZ	40.09

Key hyper-parameters of each model were optimized using a random search strategy with a five fold cross-validation scoring scheme, and a budget of 100 configurations. We increased this budget to 250 configurations for MLP models. Input features (as Coulomb matrices or BoB) were rescaled using a z-score normalization for the linear models and the MLP models. Similarly, for these models, the target OR values were also rescaled with a z-score normalization. Due to RF and ERT models being scale invariant, they did not require these data preprocessing steps.

To compare predictive performances across models for a given fragment type, we employed a five fold cross-validation scheme, with the dataset for each of the fragment types being partitioned into five folds. Models were trained using four of the folds and evaluated based on their prediction on the remaining fold, for all combinations of four training folds. The resulting performance measures were then averaged to give the cross-validated measure. We compared models based on their mean absolute error between OR values predicted by the ML model and those values calculated using DFT. The best performing models were the EN models applied to both types of features, i.e. the Coulomb matrices and the BoB features, as illustrated in Table 3.

Features with the largest influence on OR were determined based on each feature's coefficient for EN models, or each feature's mean absolute SHAP (SHapley Additive exPlanations) values (Lundberg & Lee, 2017) for RF, ERT, and MLP models. The shrinkage property of the elastic net penalty applied to the linear model simplifies the interpretation of the coefficients as feature importance. Indeed, this penalization forces coefficients of irrelevant features to be set to zero, effectively providing a feature selection mechanism for the linear model. For the non-linear models, the SHAP values, an approach based on game theory, provided a way to assign an importance value to each feature, in the context of a prediction for a given conformer. We evaluated the overall importance of each input features by averaging the absolute SHAP values over

Table 3
Comparison of machine learning model performances. Reported for both fragments is the average±standard deviation of cross-validated mean absolute error for either bag-of-bonds (BoB) or Coulomb matrix (CM) feature sets. Units are $deg.\cdot mL\cdot g^{-1}dm^{-1}$, and average values are approximately 1/3 of the standard deviation of either fragment's distribution of OR values.

Features	Model	Fragment 1	Fragment 2
BoB	EN	15 ± 1	17 ± 1
	ERT	20 ± 2	19 ± 1
	MLP	21 ± 9	20 ± 2
	RF	21 ± 3	21 ± 2
CM	EN	15 ± 2	16 ± 1
	ERT	19 ± 1	17 ± 2
	MLP	20 ± 10	17 ± 1
	RF	20 ± 1	21 ± 2

all test samples for each test set of the cross-validation scheme.

Between the models using the Coulomb matrix representation, there was little consistency in the features deemed important. Conversely, when the BoB representation was used, a small subset of the features consistently emerged as the most important. For all models and both fragment types, bond lengths and bond angles were relatively unimportant relative to a small set of dihedral angles. Specifically, the most influential dihedral angles were those defining the rotation about the bonds forming the glycosidic linkages. Additionally, all models for fragment 2 were influenced by the dihedral angle specifying the rotation of the hydroxymethyl group on the reducing-end G residue. This, along with the glycosidic bond angles, is shown in Fig. 5.

5. Discussion

It has recently been shown that the differences between the experimental wide-angle x-ray scattering profiles obtained from polymeric κ-carrageenan solutions in the disordered- and ordered-states could be explained by a transition from single 'looser' helices to 'stabilized' double-helices, respectively (Westberry et al., 2023). Here we have shown that oligomeric forms of both of these structures are dextrorotatory, and that the OR for double-helices is higher than that for single helices; qualitatively in agreement with all experimental results available in the literature. Assuming that polymeric chains would show the same trend, this suggests that the differences between OR measurements in the disordered- and ordered-states can also be explained by the same structural transition measured by x-ray scattering. That is; the long-reported OR changes do not rely on the idea of a uni-molecular coil-to-helix transition and that the more recent model, far from being at odds with the canonical observation of OR change, provides a consistent explanation of the phenomena.

Quantitatively, for longer fragment lengths the plateau value of 18 $deg.\cdot mL\cdot g^{-1}dm^{-1}$ calculated for SC as shown in Fig. 2 is lower than experimental measurements of about 45 $deg.\cdot mL\cdot g^{-1}dm^{-1}$ for polymeric κ-carrageenan in the disordered-state. Similarly, the plateau value of 92 $deg.\cdot mL\cdot g^{-1}dm^{-1}$ predicted here for longer DH fragments is higher than experimental measurements of about 73 $deg.\cdot mL\cdot g^{-1}dm^{-1}$ for polymeric κ-carrageenan in the ordered-state (Grasdalen & Smidsrød, 1981; Mangione et al., 2003; Smidsrød et al., 1980). Possible explanations for this are the approximations introduced by the basis set and functional selected; indeed when using a more complete basis set (aug-cc-pVDZ) and treating electron correlation more rigorously (CAM-B3LYP), calculated OR increased for SC conformers and decreased for DH conformers (Table 2). Alternatively, better quantitative agreement between calculation and experiment is possible if the disordered- and ordered-states are not assumed to comprise entirely of SC and DH, respectively. That is to say; a small fraction of DH are stable enough to exist under disordered-state conditons, and under ordered-state conditions some sections of chain remain undimerized due to kinetic trapping during network formation. Using the calculated values of 18 and 92 $deg.\cdot mL\cdot g^{-1}dm^{-1}$ for SC and DH as an example, this on its own would require the disordered- and ordered-states to contain approximately 36 % and 74 % DH, respectively, to be in quantitative agreement with experiment. However, these percentages seem unlikely given the limited flexibility permitted for consistency with the experimental wide-angle x-ray scattering profiles (Westberry et al., 2023).

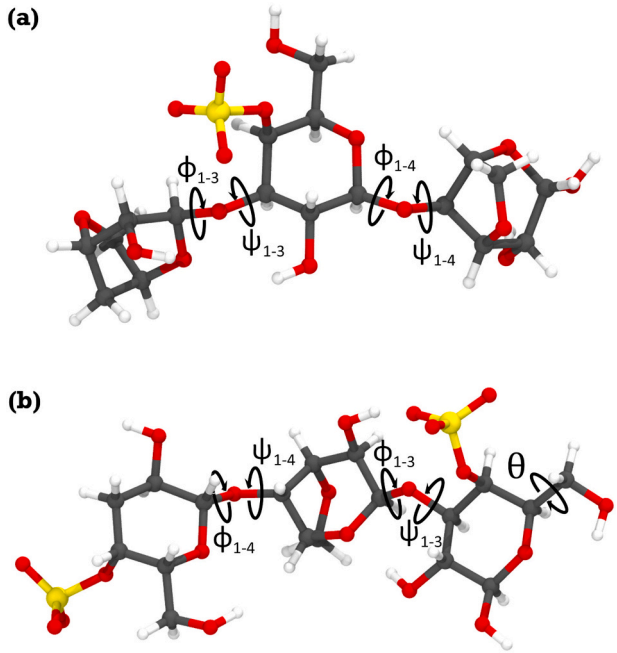

Fig. 5. Bag-of-bonds features determined important by all machine learning models of fragment 1 (a) and fragment 2 (b). ϕ_{1-3}, ψ_{1-3}, ϕ_{1-4}, and ψ_{1-4} are defined in Westberry et al. (2023), whereas θ is defined as the rotation about the bond connecting the G residue's C5 atom to the hydroxymethyl group.

Regarding an explanation for the differences between the OR in the disordered- and ordered-states, chain-chain interactions have been noted when calculating OR for both agarose and carrageenan using coupled-cluster theory (Schafer & Stevens, 1995, 1996), and are also thought to be responsible for the OR changes demonstrated by other double-helix forming polymers (Charney, 1979). From Fig. 4, the reduction in OR when calculated for DH fragments in isolation (each fragment having the conformation reflecting its presence in the DH but *without* the actual proximity of its neighbor) versus for DH *with* interactions between the fragments, suggests that chain-chain interactions are in this case also associated with increased OR. However, caution should be exercised when DFT calculations are performed with close proximity between DH fragments, as this does raise the possibility of basis set superposition errors, to which OR calculations have been shown to be susceptible to (Bursch et al., 2022; Peyton & Crawford, 2019). Although Counterpoise corrections were applied (Boys & Bernardi, 1970), these errors cannot be ruled out entirely as a contributing factor to the apparent chain-chain interaction.

In the absence of chain-chain interaction between DH fragments, it is still clear from Fig. 4 that these conformations alone do produce higher ORs compared to SC. This is despite the helical parameters of SC and DH being similar; albeit with a smaller and less variable helical pitch in the case of DH (Westberry et al., 2023). The association between higher OR values and a more tightly-wound helix has been suggested before (Schafer & Stevens, 1996), and highlights the importance of the dihedral angles across the glycosidic linkages (which determine the helical pitch) to the value of the calculated OR. Despite the fact that predictions by the developed ML models often deviated somewhat from observations (Table 3), using the models as tools for interpreting which molecular features most influence the OR was successful, and reaffirmed the strong sensitivity of the OR on the orientation of glycosidic bonds.

Thus assuming the findings presented here for oligomeric κ-carrageenan extrapolate to polymeric forms, OR changes measured upon the disorder-to-order transition appear to be explained by two single helices forming a double-helix. Here it was found that OR increased due to the combined effects of the reorientation of glycosidic linkages - and an interaction between chains - upon dimerization. This is a significant departure from recent claims that the disorder-to-order transition involves a large uni-molecular structural change (Schefer et al., 2015), and is a return to earlier arguments that the disordered-state contains significant secondary structure (Le Questel et al., 1995), and that the transition is a cooperative process (Austen et al., 1988; Norton et al., 1983). These findings are also relevant to the wider macromolecular community, demonstrating that the origins of OR changes may be nuanced and need not necessarily be associated with large uni-

molecular conformational changes.

6. Conclusion

Herein it has been shown that the differences between the OR measured for κ-carrageenan solutions in the disordered- and ordered-states can be explained qualitatively by molecular models of separate helical chains and double-helices, respectively. A combination of slight changes to chains' conformation and a chain-chain interaction appear to be responsible for the higher OR in the ordered-state. Crucially, the long-reported OR changes upon the disorder-to-order transition do not rely on the idea of a uni-molecular coil-to-helix transition. Indeed the currently presented model - where salt- and temperature-induced conformational changes are a consequence of double-helix formation rather than a prerequisite to it - provides a rational explanation of the phenomena for the canonical observation of OR change.

CRediT authorship contribution statement

B.P. Westberry: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **M. Rio:** Writing – original draft, Formal analysis, Data curation. **M.R. Waterland:** Methodology. **M.A.K. Williams:** Writing – original draft, Supervision, Project administration, Funding acquisition, 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.

Acknowledgements

Dr. Tyson Dais is acknowledged for the many helpful discussions regarding running DFT calculations. The authors also wish to acknowledge the use of New Zealand eScience Infrastructure (NeSI) high performance computing facilities as part of this research. New Zealand's national facilities are provided by NeSI and funded jointly by NeSI's collaborator institutions and through the Ministry of Business, Innovation and Employment's Research Infrastructure programme. Dr. Bradley Mansel is also thanked for the generous compute time.

References

Austen, K. R. J., Goodall, D. M., & Norton, I. T. (1988). Anion effects on equilibria and kinetics of the disorder-order transition of κ-carrageenan. *Biopolymers, 27*, 139–155.
Bosco, M., Segre, A., Miertus, S., Cesàro, A., & Paoletti, S. (2005). The disordered conformation of κ-carrageenan in solution as determined by NMR experiments and molecular modeling. *Carbohydrate Research, 340*, 943–958.
Boys, S., & Bernardi, F. (1970). The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Molecular Physics, 19*, 553–566.
Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.
Bryce, T. A., Clark, A. H., Rees, D. A., & Reid, D. S. (1982). Concentration dependence of the order-disorder transition of carrageenans. Further confirmatory evidence for the double Helix in solution. *European Journal of Biochemistry, 122*, 63–69.
Bursch, M., Mewes, J., Hansen, A., & Grimme, S. (2022). Best-practice DFT protocols for basic molecular computational chemistry. *Angewandte Chemie, 134*, Article e202205735.
Cahn, R. S., Ingold, C., & Prelog, V. (1966). Specification of molecular chirality. *Angewandte Chemie International Edition in English, 5*, 385–415.
Cardoso, M. V. C., & Sabadini, E. (2010). The gelling of κ-carrageenan in light and heavy water. *Carbohydrate Research, 345*, 2368–2373.
Charney, E. (1979). *The molecular basis of optical activity*. John Wiley and Sons, Inc.
Ciancia, M., Milas, M., & Rinaudo, M. (1997). On the specific role of coions and counterions on kappa-carrageenan conformation. *Int. J. Bio. Macromol., 20*, 35–41.
Dea, I. C. (1993). Chapter 2 - conformational origins of polysaccharide solution and gel properties. In R. L. Whistler, & J. N. Bemiller (Eds.), *Industrial gums* (third edition, pp. 21–52). London: Academic Press. (Third edition ed.).
Dea, I., McKinnon, A., & Rees, D. (1972). Tertiary and quaternary structure in aqueous polysaccharide systems which model cell wall cohesion: Reversible changes in conformation and association of agarose, carrageenan and galactomannans. *Journal of Molecular Biology, 68*, 153–172.
Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., … Fox, D. J. (2009). *Gaussian 09*.
Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning, 63*, 3–42.
Grasdalen, H., & Smidsrød, O. (1981). Iodide-specific formation of κ-carrageenan single helices. 127I NMR spectroscopic evidence for selective site binding of iodide anions in the ordered conformation. *Macromolecules, 14*, 1842–1845.
Hjerde, T., Smidsrød, O., Stokke, B. T., & Christensen, B. E. (1998). Acid hydrolysis of κ- and ι-carrageenan in the disordered and ordered conformations: Characterization of partially hydrolyzed samples and single-stranded oligomers released from the ordered structures. *Macromolecules, 31*, 1842–1851.
Huang, B., & Von Lilienfeld, O. A. (2016). Communication: Understanding molecular representations in machine learning: The role of uniqueness and target similarity. *The Journal of Chemical Physics, 145*, Article 161102.
Le Questel, J.-Y., Cros, S., Mackie, W., & Pérez, S. (1995). Computer modelling of sulfated carbohydrates: Applications to carrageenans. *International Journal of Biological Macromolecules, 17*, 161–175.
Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *vol. 30. Advances in neural information processing systems*. Curran Associates, Inc.
Mangione, M. R., Giacomazza, D., Bulone, D., Martorana, V., & Biagio, P. L. S. (2003). Thermoreversible gelation of κ-carrageenan: Relation between conformational transition and aggregation. *Biophysical Chemistry, 104*, 95–105.
McKinnon, A. A., Rees, D. A., & Williamson, F. B. (1969). Coil to double Helix transition for a polysaccharide. *Journal of the Chemical Society D: Chemical Communications, 1*, 701–702.
Morris, E. R., Rees, D. A., & Robinson, G. (1980). Cation-specific aggregation of carrageenan helices: Domain model of polymer gel structure. *J. Mol. Bio., 138*, 349–362.
Nilsson, S., & Piculell, L. (1989). Helix-coil transitions of ionic polysaccharides analyzed within the Poisson-Boltzmann cell model. 2. Effects of salt concentration on the thermal transition. *Macromolecules, 22*, 3011–3017.
Norton, I. T., Goodall, D. M., Morris, E. R., & Rees, D. A. (1983). Equilibrium and dynamic studies of the disorder–order transition of kappa carrageenan. *Journal of the Chemical Society, Faraday Transactions, 1* (Physical Chemistry in Condensed Phases, 79, 2489).
Paoletti, S., Delben, F., Cesàro, A., & Grasdalen, H. (1985). Conformational transition of κ-carrageenan in aqueous solution. *Macromolecules, 18*, 1834–1841.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.
Peyton, B. G., & Crawford, T. D. (2019). Basis set superposition errors in the many-body expansion of molecular properties. *The Journal of Physical Chemistry A, 123*, 4500–4511.
Ramakrishnan, S., & Prud'homme, R. K. (2000). Behavior of κ-carrageenan in glycerol and sorbitol solutions. *Carbohydrate Polymers*, 327–332.
Rees, D. A., Steele, I. W., & Williamson, F. B. (1969). Conformational analysis of polysaccharides. III. The relation between stereochemistry and properties of some natural polysaccharide sulfates (1). *Journal of Polymer Science Part C: Polymer Symposia, 28*, 261–276.
Rees, D. A., Scott, W. E., & Williamson, F. B. (1970). Correlation of optical activity with polysaccharide conformation. *Nature, 227*, 390–392.
Rees, D. A., Williamson, F. B., Frangou, S. A., & Morris, E. R. (1982). Fragmentation and modification of ι-carrageenan and characterisation of the polysaccharide order-disorder transition in solution. *European Journal of Biochemistry, 122*, 71–79.
Rochas, C., & Rinaudo, M. (1980). Activity coefficients of counterions and conformation in kappa-carrageenan systems. *Biopolymers, 19*, 1675–1687.
Rochas, C., Rinaudo, M., & Vincendon, M. (1983). Spectroscopic characterization and conformation of oligo kappa carrageenans. *International Journal of Biological Macromolecules, 5*, 111–115.
Rochas, C., Rinaudo, M., & Landry, S. (1989). Relation between the molecular structure and mechanical properties of carrageenan gels. *Carbohydrate Polymers, 10*, 115–127.
Rupp, M., Tkatchenko, A., Müller, K.-R., & Von Lilienfeld, O. A. (2012). Fast and accurate modeling of molecular atomization energies with machine learning. *Physical Review Letters, 108*, Article 058301.
Schafer, S. E., & Stevens, E. S. (1995). A reexamination of the double-helix model for agarose gels using optical rotation. *Biopolymers, 36*, 103–108.
Schafer, S. E., & Stevens, E. S. (1996). The optical rotation of ordered carrageenans. *Carbohydrate Polymers, 31*, 19–22.
Schefer, L., Usov, I., & Mezzenga, R. (2015). Anomalous stiffening and ion-induced coil–helix transition of carrageenans under monovalent salt conditions. *Biomacromolecules, 16*, 985–991.
Slootmaekers, D., Jonghe, C. D., Reynaers, H., & Varkevisser, F. A. (1988). Static light scattering from κ-carrageenan solutions. *Int. J. Bio. Macromol., 10*, 160–168.
Smidsrød, O., Andresen, I.-L., Grasdalen, H., Larsen, B., & Painter, T. (1980). Evidence for a salt-promoted "freeze-out" of linkage conformations in carrageenans as a prerequisite for gel-formation. *Carbohydrate Research, 80*, C11–C16.

Snoeren, T., & Payens, T. (1976). On the sol-gel transition in solutions of kappa-carrageenan. *Biochimica et Biophysica Acta - General Subjects, 437*, 264–272.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research, 15*, 1929–1958.
Takemasa, M., & Nishinari, K. (2004). The effect of the linear charge density of carrageenan on the ion binding investigated by differential scanning calorimetry, dc conductivity, and kHz dielectric relaxation. *Colloids and Surfaces B: Biointerfaces, 38*, 231–240.
van de Velde, F., Rollema, H. S., Grinberg, N. V., Burova, T. V., Grinberg, V. Y., & Tromp, R. H. (2002). Coil-Helix transition of ι-carrageenan as a function of chain regularity. *Biopolymers, 65*, 299–312.
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*, 1113–1129.
Viebke, C., Piculell, L., & Nilsson, S. (1994). On the mechanism of gelation of Helix-forming biopolymers. *Macromolecules, 27*, 4160–4166.
Westberry, B. P., Mansel, B. W., Ryan, T. M., Lundin, L., & Williams, M. (2022). X-ray scattering and molecular dynamics simulations reveal the secondary structure of κ-carrageenan in the solution state. *Carbohydrate Polymers, 296*, Article 119958.
Westberry, B. P., Mansel, B. W., Lundin, L., & Williams, M. (2023). Molecular dynamics simulations and x-ray scattering show the κ-carrageenan disorder-to-order transition to be the formation of double-helices. *Carbohydrate Polymers, 302*, Article 120417.
Yang, J. T., & Doty, P. (1957). The optical rotatory dispersion of polypeptides and proteins in relation to configuration. *Journal of the American Chemical Society, 79*, 761–775.
Yegappan, R., Selvaprithiviraj, V., Amirthalingam, S., & Jayakumar, R. (2018). Carrageenan based hydrogels for drug delivery, tissue engineering and wound healing. *Carbohydrate Polymers, 198*, 385–400.
Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology), 67*, 301–320.