The Disordered- and Ordered-State Structures of κ-Carrageenan: an X-ray Scattering, Molecular Dynamics, and Density-Functional Theory Study

A thesis presented in partial fulfilment of the requirements for the degree of
Doctor of Philosophy
in
Physics
at Massey University, Palmerston North,
New Zealand.

MASSEY
UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

Benjamin P. Westberry

2022

Contents

List of Tables

List of Figures

List of Acronyms

AFM	Atomic force microscopy
COM	Center of mass
CV	Collective variable
DFT	Density-functional theory
DH	Double-helical (specifically simulations initiated as such)
DP	Degree of polymerization
DSC	Differential scanning calorimetry
FES	Free energy surface
HB	Hydrogen bond
MD	Molecular dynamics
MM	Molecular mechanics
NMR	Nuclear magnetic resonance
OR	Optical rotation
PBC	Periodic boundary condition
QM	Quantum mechanics
SAXS	Small-angle X-ray scattering
SC	Separate chains (specifically simulations initiated as such)
WAXS	Wide-angle X-ray scattering
XRD	X-ray diffraction

Abstract

κ-carrageenan is a biopolymer extracted from marine algae. It exists in aqueous solution, at high temperatures and/or low salt concentrations as a 'disordered-state', and at low temperatures and in the presence of certain salts as an 'ordered-state'. The transition between disordered- and ordered-states involves molecular structural changes, which are essential to its interesting viscoelastic properties that are routinely exploited in a plethora of applications. Despite this, the molecular conformations of the disordered- and ordered-states, as well as the details of the transitional pathway connecting them, remain a source of contention. While decades of research have amassed a vast trove of information on the disorder-order transition, an atomistic understanding of the structure in solution has remained elusive. This study takes advantages of recent advances in computational capabilities in order to simulate κ-carrageenan solutions on length scales of $\sim$10 nm over µs time scales, and thus develop atomistic models of the disordered- and ordered-states. Both models are used to calculate wide-angle X-ray scattering profiles, and these are subsequently validated by comparison to data obtained at a synchrotron facility. The models will be further explored using density-functional theory to calculate their expected optical rotation behaviour, which finds that the formation of double-helices from single chains is able to explain the increase in optical rotation measured experimentally when transitioning from the disordered- to ordered-state. Structural analysis of both experimentally-verified models find the disordered-state to have a significant amount of residual helical secondary-structure, whereas the ordered-state is mostly double-helical. Crucially, simulations show that the ordered-state arises spontaneously from the so-called disordered-state at a rate dependant on salt concentration, *without* prior uni-molecular changes. The findings of this research are the most detailed model of the disorder-order transition to-date, and demonstrate that the existing paradigm of a 'coil-to-helix' transition is in need of revision.

Acknowledgements

A lot can happen in the years it takes to complete a PhD, and there are many people and events – both personal and global – that shape its course. Of those people, my good friend Professor Bill Williams has influenced my work the most, sharing his love for soft materials and experimental science while giving me the space to explore my own interests. Bill's previous student Dr Brad Mansel - whom I often felt I was following in the footsteps of - deserves mention for all his help with X-ray scattering and use of his workstations throughout my PhD. With these two friends I travelled to the Australian Syncrotron where I serendipitously met Dr Leif Lundin and his Swedish team, who initially got me interested in carrageenan systems (look what you've done!). In the years following, both Dr Davide Mercadante and Dr Amir Irani have both been of great help in developing my understanding of molecular dynamics.

I was fortunate to be involved with The MacDiarmid Institute during my PhD, to which I am very grateful; not only for funding, but also for serving as a shining example of how to do good science in New Zealand. I'm also very grateful for the many friendships I've made along the way at Massey and within MESA – in particular Doctors Susav Pradhan, Sashi Ramamirtham, and Samuel Brooke – with whom I've been on many great adventures.

Much of what I've been able to achieve in life is due to my parents Patrick and Dr Nicola Westberry, who encouraged me to be curious as a child. My sister (soon to be Dr) Abigail Westberry, who inspires through her hard work, also deserves an acknowledgement. My daughter Madison, as well as her brother due next year, will undoubtedly lead their generation to be better than ours and are also great sources of inspiration. Of all the people that have provided inspiration, however, is my amazing wife Emma. We have shared so many experiences over the years, and she has had to make most of the sacrifices that I have during this PhD. For this, and everything else, I will always be in gratitude.

Chapter 1

Introduction

The development of knowledge is analogous to urban development, often being established at the intersection between existing hubs and growing according to the opportunities present and the difficulty of the terrain. Like knowledge, new roads must be built off from existing ones, heavily used road are upgraded to main arteries, and unused roads eventually become derelict and forgotten. Regression is discouraged; work should only be commissioned if it allows access to somewhere new, or improves an existing route in some way. Within scientific disciplines this analogy is strongly encouraged; and yet in the study of the polysaccharide carrageenan, this network of knowledge has become somewhat of a labyrinth in places. Some areas are serviced by multiple redundant routes, which are sometimes built on unstable ground and used by people confused by the map. This work aims to alleviate some of this confusion in one particular area of carrageenan research: the secondary structures of the disordered- and ordered-states of κ-carrageenan.

Carrageenans are a class of linear polysaccharides that serve structural functions in marine algae (Percival, 1979; Domozych, 2011), characterised by a primary structure of disaccharide repeating units containing $(1 \rightarrow 3)$-linked β-D-galactopyranose residues and $(1 \rightarrow 4)$-linked α-D-galactopyranose residues (Figure 1.1). Types of carrageenan are categorised by their degree of sulfation, with κ-carrageenan carrying one negatively charged sulfate per 3-linked galactose reside (this residue being termed 'G' in this work) and possessing a 3,6-anhydro bridge on each 4-linked galactose reside (this residue being termed 'A' in this work). Being a polyanion, κ-carrageenan is commercially available as a salt, often of sodium or potassium. Polymer solutions are polydisperse, usually having molecular weights in the range of 200-400 kDa after extraction and before any treatment (Stanley, 1990). Although often idealised as entirely one type of repeating unit, other types of carrageenan (i.e. disaccharide with a different degree of sulfation) and even kinks in the chain are typically present (Rees, 1972).

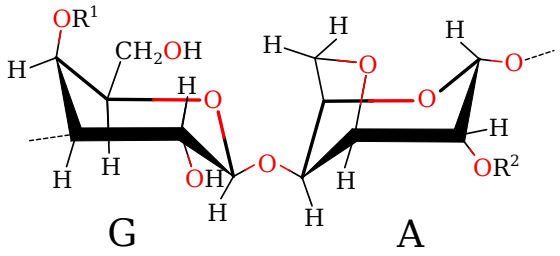

Figure 1.1: Primary chemical structure of the carrageenan disaccharide repeating unit. For κ-carrageenan, R^1=SO3$^-$ and R^2=H. For ι-carrageenan, R^1=SO3$^-$ and R^2=SO3$^-$.

The viscoelastic properties, ingestibility, and inexpensive production of carrageenan have lead to its widespread application; mainly in the food industry, however increasingly in medical, pharmaceutical, and cosmetic applications (Whistler and BeMiller, 1993; Yegappan et al., 2018). It can exist as a viscous solution, adding body to and/or stabilizing sauces and milk products, or as a gel, for instance in puddings and oral extended-release tablets (Stanley, 1990; Li et al., 2014). This tunable viscoelasticity is largely due to inter-molecular interactions which are affected by the polysaccharide's secondary structure – itself sensitive to the ionic environment and temperature.

At high temperatures and/or low ionic strengths, most carrageenans exist in the disordered-state, which was initially assigned by McKinnon et al. (1969) as a 'random coil', i.e. a polymer chain devoid of correlation between monomeric units (Rubinstein and Colby, 2003) and lacking any secondary structure. Although heavily dependant on the species of co-ion and counter-ion, temperatures above ~75°C at most salt concentrations – or salt concentrations of less than ~0.01 M at room temperature – are usually sufficient conditions to achieve the disordered-state (Ciancia et al., 1997). The initial assignment of the state to a random coil by McKinnon et al. was made based on the low optical rotation (OR) measured for the state (relative to the ordered-state) and a comparison to the 'coil-to-helix' transition known at the time to occur for polypeptides (Yang and Doty, 1957). Owing to the difficulties in obtaining structural information at small length scales for such amorphous samples, the term 'random coil' has since been used nearly unanimously, without much further justification to describe this state. Following the developments of cryo-electron microscopy, Hermansson (1988) attempted to resolve the disordered-state at the nm length scale, but was unable to obtain useful structural information. Using solution-state Small-Angle X-ray Scattering (SAXS), Turquois et al. (1995) was able to conclude that κ-carrageenan chains in the disordered-state adopted a 'wormlike conformation' over nm length scales. However, a series of molecular modelling and molecular dynamics (MM and MD respectively) studies conducted around the same time found that the carrageenan glycosidic linkages – whose

preferred conformations have the largest impact on secondary structure – are quite restricted energetically in their movement, rather than randomly oriented (Urbani et al., 1993; Le Questel et al., 1995; Stortz and Cerezo, 2000; Ueda et al., 2001; Stortz, 2002). This finding was later supported by Bosco et al. (2005), who found using MM and Nuclear Magnetic Resonance (NMR) that the disordered-state was better described as a 'loose helix' containing a 'substantial amount of residual order' on the sub-nm length scale. Most recently, atomic force microscopy (AFM) images have appeared to show κ-carrageenan in the disordered-state to be a random coil on the nm length scale (Schefer et al., 2015a,b; Diener et al., 2019), however care must be taken when assuming that the polyanions adsorbed to positively charged surfaces in these studies represent the true solution-state conformation.

In the same paper assigning the disordered-state, McKinnon et al. (1969) assigned the ordered-state secondary structure to a helix. Relative to the disordered-state secondary structure, the ordered-state structure has since received much more attention, and the evidence supporting its assignment is more convincing. However, contention surrounds whether the helical structure exists as a single- (Paoletti et al., 1984) or double- (Morris et al., 1980a) helix. The most decisive way to test this – whether the molecular weight differs by a factor of two between the disordered- and ordered-states – is complicated by the tendency for κ-carrageenan to aggregate in the presence of many salts (Morris et al., 1980a). While this can be suppressed using iodide salts (Takemasa and Nishinari, 2004), contradictory results still exist showing both that the molecular weight remains unchanged (Grasdalen and Smidsrød, 1981a; Ciancia et al., 1997; Slootmaekers et al., 1988), or that it increases (Hjerde et al., 1998, 1999) in the ordered-state. Although most calorimetric studies assume the ordered-state is dimeric, differential scanning calorimetry (DSC) data can be fitted by modelling the ordered-state as a double-helix (Viebke et al., 1998) or as side-by-side aggregated single-helices (Paoletti et al., 1984). On the other hand, molecular modelling based on fiber X-ray diffraction (XRD) experiments conclusively shows that κ-carrageenan in the condensed ordered-state exists as a right-handed, three-fold double-helix (Millane et al., 1988), with side-by-side models giving poorer fits to the data. Indeed the confusion continues to the present day, with the XRD results apparently being refuted by AFM images in which a doubling of the chain's height in the ordered-state has been interpreted as evidence for formation of single-helices (Schefer et al., 2015a,b; Diener et al., 2019), although no attempt is made to explain the influence of the surface on conformation in those studies.

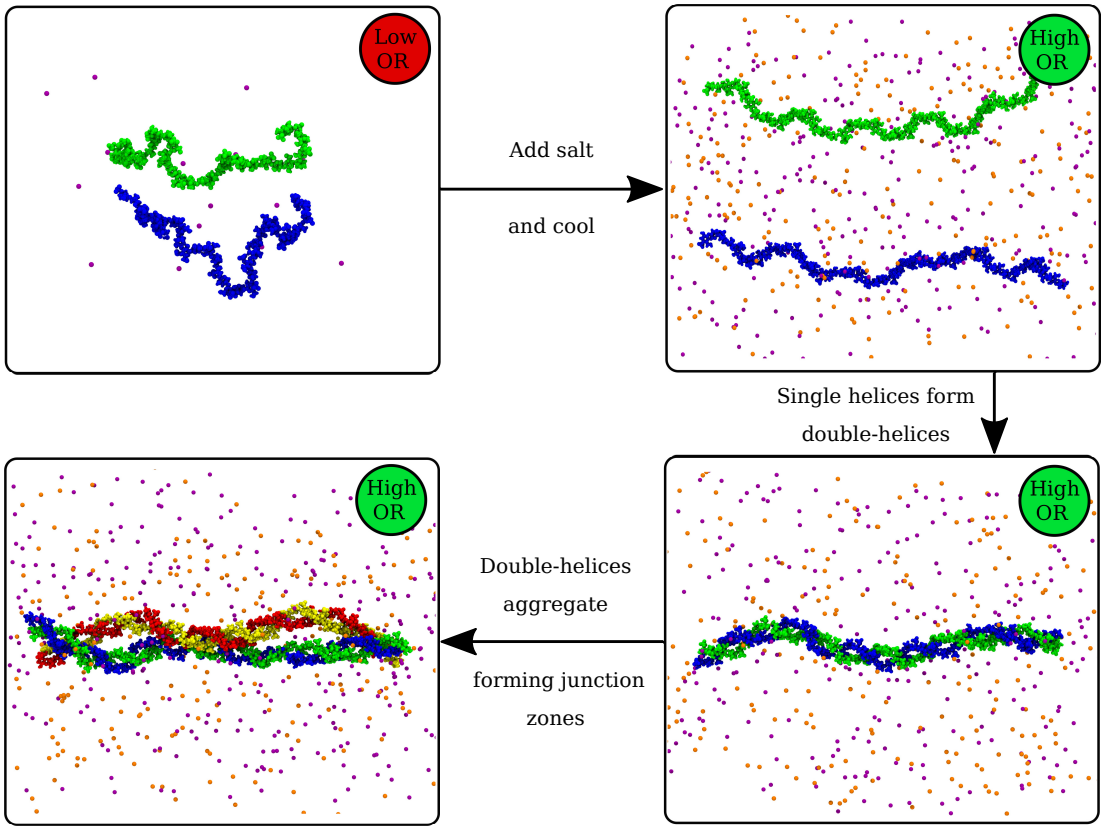

Figure 1.2: The most popular existing model for κ-carrageenan gelation.

Returning to the analogy, knowledge of the disordered-state has been hindered by difficult terrain, whereas knowledge of the ordered-state is an easy place to get lost due to the many superfluous routes. Despite this, most lines of reasoning pass through a common point on the carrageenan 'map'. That is, that the increase in measured OR upon the disorder-order transition, which is usually induced by cooling in the presence of added salt, is due to the secondary structure of individual chains becoming helical. In models where dimerization occurs, this newly acquired helical secondary structure then facilitates the dimerization of single helices (Morris et al., 1980a; Paoletti et al., 1984). The most popular model – that of Morris et al. as shown in Figure 1.2 – then sees the aggregation of double-helices forming the cross-links of the gel network in the presence of some salts.

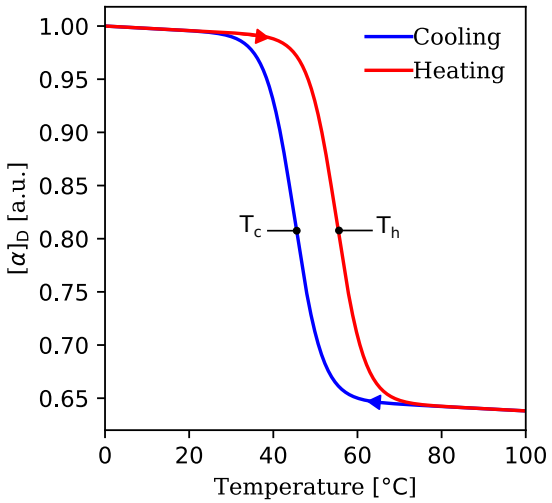

Figure 1.3: A hypothetical optical rotation profile as a function of temperature, showing hysteresis.

Since the initial report of OR changes in carrageenan solutions by McKinnon et al., OR has been used as an indicator of helical content to understand the dependence of the carrageenan disorder-order transition on: the added salt type and concentration (Rochas and Rinaudo, 1980; Morris et al., 1980a; Grasdalen and Smidsrød, 1981a; Austen et al., 1988; Borgström et al., 1996; Ciancia et al., 1997; Piculell, 1997); polymer concentration (Morris et al., 1980c; Bryce et al., 1982); solvent type (Cardoso and Sabadini, 2010); chain composition (Rees et al., 1982; van de Velde et al., 2002, 2005); and chain size (Rochas et al., 1983) to name a few. Of these studies which induce the transition thermally, two temperatures are important to note: the temperature at which the OR is at the midpoint between its values in the disordered- and ordered-states upon cooling T_c, and upon heating T_h (Figure 1.3). These temperatures are dependant on the species of cation present, and increase in the order (Rochas and Rinaudo, 1980):

$$Li^+ < Na^+ < N(CH_3)_4^+ < NH_4^+ < K^+ < Cs^+ < Rb^+,$$

and dependant on the species of anion present, with transition temperatures increasing in the order (Austen et al., 1988):

$$F^- < Cl^- < NO_3^- < Br^- < I^-.$$

The hysteresis (the difference between OR profiles upon cooling and heating) is particularly pronounced for κ-carrageenan in the presence of K^+, which has been interpreted as evidence for the stabilization of double-helices by extensive aggregation (Figure 1.2,

Morris et al. (1980a)). Where OR has been used to study the dynamics of the disorder-order transition, the results strongly favour a dimeric ordered-state (Norton et al., 1983; Austen et al., 1988), which has been shown (again using OR) to occur even for small oligomers (Rochas et al., 1983). Clearly on our carrageenan 'map', the OR is a major highway between different areas.

Given the importance of OR to the understanding of the carrageenan disorder-order transition, it is unsurprising that attempts have been made to understand its origin. The earliest attempt at a quantitative description of the OR was by Rees et al. (1970), who applied an empirical model where the observed specific OR for ι-carrageenan arose from the sum of two terms: one related to the OR of individual residues and the other a function of the chain conformation. Using this approach to calculate the OR for a small set of sterically-allowed single- and double-helical models, Rees et al. found a number of double-helices for which the calculated OR was close to the observed ordered-state value. A more rigorous treatment of the OR came nearly two decades later, when Stevens and Sathyanarayana (1987) found qualitative agreement with experimental data when calculating the OR for a set of monosaccharides, by expressing their OR as the sum of contributions from small fragments of their structures. This approach was extended to show the α-anomer of methyl-D-galactopyranose to have a higher OR than the β-anomer (Stevens et al., 1989), and then ultimately to calculate the OR of unsulfated carrageenan chains (Schafer and Stevens, 1996). In the latter study, MM was used to find a set of probable single- and double-helical models, for which OR calculations from tightly wound double-helices best matched experimental values for the ordered-state. As far as the carrageenan 'map' is concerned, this is a good starting point for further development.

Since the explosive development in carrageenan research during the final decades of the last century, experimental and computational techniques have advanced considerably. Solution-state wide-angle X-ray scattering (WAXS) beamlines at synchrotron facilities can now probe matter on sub-nm length scales, MD can now be used to simulate systems 10s of nm in size over μs time scales, and quantum mechanics (QM) software packages can now make *ab initio* calculations of molecular properties for systems containing 100s of atoms. This thesis aims to apply these advancements to three important areas of the κ-carrageenan 'map': (a) push into unexplored territory by modelling the disordered-state structure of κ-carrageenan using a combination of WAXS and atomistic MD simulations; (b) link the disordered-state structure to a refined double-helical model of the ordered-state, providing an atomistic understanding of the disorder-order transition using WAXS and MD simulation; and (c) build on the early attempts to calculate the OR using modern MD and QM approaches, providing evidence that the disordered- and ordered-state models proposed in this thesis are indeed robust.

Chapter 2

Theory

2.1 Molecular Dynamics

2.1.1 The Fundamentals

To develop a theoretical basis for this thesis, it would be helpful to first lay the foundations for a molecular understanding of biomolecules. However, the amorphous, heterogeneous, and often dynamic nature of soft materials makes an analytical description of the structures and interactions contained within a difficult task. For this reason, little attempt is made in this work to apply traditional treatments such as those of polymer physics (Rubinstein and Colby, 2003), whose degree of abstraction is often beneficial but will not help to develop an atomistic understanding. Rather, the molecular structures adopted are directly computed using MD simulations, whose sole assumptions are the specific chemical structure and a set of interaction parameters derived from *ab-initio* QM calculations.

The goal of most MD simulations is to propagate a system containing N particles (hereby called atoms) from some initial coordinates $\boldsymbol{X}(0) = \{\boldsymbol{x}_1(0), \boldsymbol{x}_2(0), ..., \boldsymbol{x}_N(0)\}$, through a series of discrete time steps Δt to some final coordinates $\boldsymbol{X}(t)$ at a later time t, based on the laws of classical mechanics. This is shown graphically in Figure 2.1.

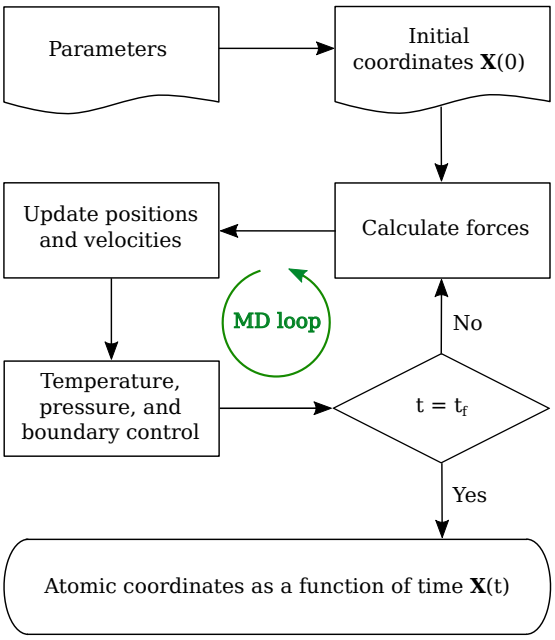

Figure 2.1: The central algorithm for most classical molecular dynamics codes.

At the core of the MD algorithm is the calculation of the force $\boldsymbol{F}_i(t)$ experienced by the i^{th} atom, which in classical mechanics is given by the gradient of the energetic potential $V_i(\boldsymbol{X}[t])$ within which the atom resides:

$$\boldsymbol{F}_i(\boldsymbol{X}[t]) = -\nabla V_i(\boldsymbol{X}[t]) \tag{2.1}$$

after which the atom's coordinates and velocity are updated according to classical mechanics:

$$\boldsymbol{x}_i(t + \Delta t) = \boldsymbol{x}_i(t) + \boldsymbol{v}_i(t)\Delta t + \frac{1}{2m_i}\boldsymbol{F}_i(t)\Delta t^2 \tag{2.2}$$

$$\boldsymbol{v}_i(t + \Delta t) = \boldsymbol{v}_i(t) + \frac{1}{m_i}\boldsymbol{F}_i(t)\Delta t \tag{2.3}$$

Although it has been omitted for brevity in (2.2), Equation (2.1) shows that the force $\boldsymbol{F}_i(t)$ depends on the locations of all atoms in the system $\boldsymbol{X}(t)$. For an atom existing within a molecule, the energetic potential can be split into contributions arising from the geometry of its covalent bonds (here called bonded interactions) and contributions arising from through-space (non-bonded) interactions:

$$V(\boldsymbol{X}[t]) = V_B(\boldsymbol{X}[t]) + V_{NB}(\boldsymbol{X}[t]) \tag{2.4}$$

The parameter set used to determine the energetic terms $V_B(\boldsymbol{X}[t])$ and $V_{NB}(\boldsymbol{X}[t])$ is called the force field, which can either be suitable for general purpose MD (for example CHARMM (Vanommeslaeghe et al. (2009))), or specialised for protein (AMBER, (Cornell et al. (1995))) or carbohydrate (GLYCAM,(Kirschner et al. (2008a))) systems to name a few.

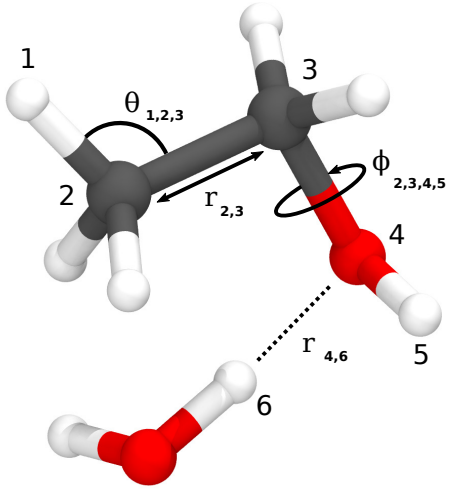

Figure 2.2: The energetic potential for atom 2 has bonded interaction contributions with dependence on $r_{2,3}$, $\theta_{1,2,3}$ and $\phi_{2,3,4,5}$. The energetic potential for atom 4 has non-bonded interaction contributions dependant on $r_{4,6}$, as well as on all other radii $r_{4,i}$.

2.1.2 Parameters

Bonded parameters have three contributions: a two-body term from bond stretching; a three-body term from bond angle bending; and a four-body term due to torsion of bonds (Figure 2.2). Dropping the explicit statement of coordinate dependence $\boldsymbol{X}[t]$ for brevity:

$$V_B = V_{bond} + V_{angle} + V_{torsion} \tag{2.5}$$

The functional form of the three terms in (2.5) can be derived by fitting *ab initio* quantum mechanical calculations for small molecular fragments (Zhou and Liu (2022)), as is the case for the GLYCAM force field used in this work. For this force field, the functional forms of the three terms in (2.5) are (Cornell et al. (1995):

$$V_{bond} = \frac{1}{2}k_{a,b}(r_{a,b} - r^{'}_{a,b})^2 \tag{2.6}$$

$$V_{angle} = \frac{1}{2}k_{a,b,c}(\theta_{a,b,c} - \theta^{'}_{a,b,c})^2 \tag{2.7}$$

$$V_{torsion} = k_{a,b,c,d}(1 + \cos(\phi_{a,b,c,d} - \phi^{'}_{a,b,c,d})) \tag{2.8}$$

where constants k and equilibrium values $r^{'}_{a,b}$, $\theta^{'}_{a,b,c}$, and $\phi^{'}_{a,b,c,d}$ are specific to the types of atoms a, b, c and d.

Non-bonded parameters have two contributions: one from electrostatic forces, and the other from Pauli and dispersion forces as modelled by a Lennard-Jones potential (Figure 2.2):

$$V_{NB} = V_{elec.} + V_{LJ} \tag{2.9}$$

The functional form of the electrostatic contribution on atom a is given simply by the Coulomb interaction:

$$V_{elec.} = \frac{1}{4\pi\epsilon_0}\sum_{i\neq a}^{N}\frac{q_a q_i}{r_{a,i}} \tag{2.10}$$

where the partial charges of each atom q_i are treated as point charges and assigned once at the start of the simulation. These are found by integrating the electron density around the atom (see section 2.3.2) and summing the result with the nucleus' charge. The partial charges computed for a small oligomer can be seen in Figure (2.3). Due to the high computational cost, the summation in (2.10) rarely runs over all atoms in the system N. Instead, this runs over atoms within a cut-off radius ($\sim$ 1 nm), with the potential due to long-ranged electrostatics calculated using the particle-mesh Ewald method, which sums long-ranged electrostatic energies in Fourier rather than real space. The functional form of the Lennard-Jones contribution is given by:

$$U_{LJ} = \sum_{i\neq a}^{N}\frac{k_{a,i}}{r_{a,i}^{12}} + \frac{c_{a,i}}{r_{a,i}^{6}} \tag{2.11}$$

where the constants $k_{a,i}$ and $c_{a,i}$ are specific to the types of atoms a and i. As with the electrostatic potential term, the summation rarely runs over all atoms in the system N but rather those within a cut-off radius.

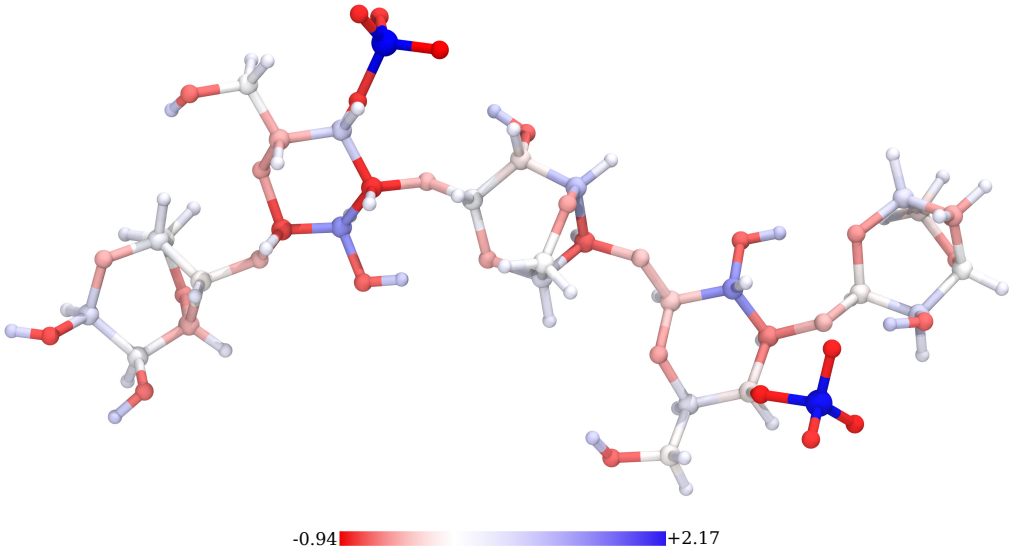

Figure 2.3: Partial charges for a κ-carrageenan oligomer. The scale bar is in units of elementary charge.

2.1.3 Metadynamics

Often one is interested in understanding a process that takes too long for an ordinary MD simulation to complete, for example the unfolding of a protein or the formation of a crystal. These final states may be energetically favourable, but the simulation may get 'stuck' in local energetic minima of the free energy surface (FES) on their way to them. If the process of interest is able to be parameterized by some collective variable (CV) $\boldsymbol{s}(\boldsymbol{X}[t])$, such as the end-to-end distance of the protein or the radius of gyration of a polymer chain, metadynamics (Laio and Parrinello (2002)) can be used to 'dislodge' the system from local energetic minima.

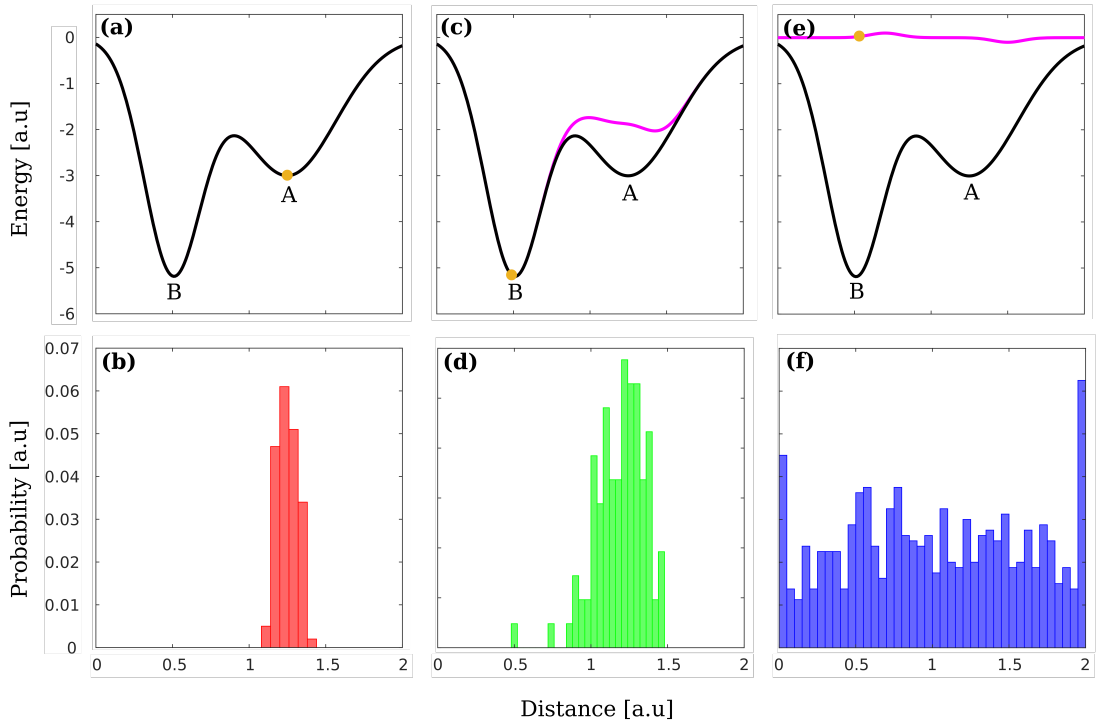

Figure 2.4: The fictitious free energy surface (black line) for a pair of atoms as a function of separation (the collective variable), and the instantaneous value of the collective variable (orange point). In an unbiased simulation (a), the value of the collective variable remains stuck in well A, as shown by the distribution of measured inter-atomic distances (b). When metadynamics is applied (c), the free energy surface is modulated by additional Gaussian kernels (purple line), causing the system to explore more of the space (d). As more Gaussian kernels are added, their sum eventually counteracts the underlying free energy surface (e), and the system freely diffuses through the collective variable space (f).

Consider a fictitious example of atoms 1 and 2, where the CV $\boldsymbol{s}(\boldsymbol{x}_1[t], \boldsymbol{x}_2[t]) = |\boldsymbol{x}(t)_2 - \boldsymbol{x}(t)_1|$ has a FES such as that shown in Figure (2.4 (a)). Due to the energetic barrier separating wells A and B, the system in an ordinary MD simulation will spend a long time in well A before making the jump to B. However, if the locations in the CV space where the system remained stuck became progressively less favourable energetically, by modulating the shape of the potential, the system would eventually move away. This is the core premise of metadynamics – at regular intervals small Gaussian kernels are placed in the CV space at $\boldsymbol{s}(\boldsymbol{X}[t])$, slowly building up an an auxiliary, history-dependant bias to counteract the underlying FES and force the system away

from local minima (Figure 2.4 (b)). At a time t, the bias $U_{bias}(\boldsymbol{s}[t])$ (where the dependence on $\boldsymbol{X}$ has been omitted for brevity) is the sum of all energetic kernels placed beforehand $U_{bias}(\boldsymbol{s}(t-\Delta t))$ plus an additional kernel at the current location in CV space $\boldsymbol{s}(t)$:

$$U_{bias}(\boldsymbol{s}[t]) = U_{bias}(\boldsymbol{s}[t-\Delta t]) + \omega \exp\Big(-\frac{1}{2}\Big|\frac{\boldsymbol{s}-\boldsymbol{s}[t]}{\boldsymbol{\sigma}}\Big|^2\Big) \tag{2.12}$$

The force on an atom within the system is then (2.1) plus the additional force arising from the bias:

$$\boldsymbol{F}_i(\boldsymbol{X}[t]) = -\nabla U(\boldsymbol{X}[t]) - \frac{\partial U_{bias}(\boldsymbol{s}[\boldsymbol{X}(t)])}{\partial \boldsymbol{s}}\Big|_{\boldsymbol{s}(t)} \frac{\partial \boldsymbol{s}[\boldsymbol{X}(t)]}{\partial \boldsymbol{X}(t)} \tag{2.13}$$

This process of building up an auxiliary bias to counteract the FES continues until it is determined that the effects of the underlying FES have been completely neutralised, that is:

$$U_{bias}(\boldsymbol{s}) = -U_{FES}(\boldsymbol{s}) \tag{2.14}$$

The point at which (2.14) is true (up to an additive constant) is when the system freely diffuses through the CV space (Figure 2.4 (e-f)). Under isobaric and isothermal conditions, the negative of the auxiliary bias $-U_{bias}(\mathbf{s})$ provides a convenient estimate of the Gibbs FES as a function of the collective variable(s), without having to explicitly calculate changes to the system's entropy.

2.1.4 Control Algorithms

By conducting MD simulations, the goal is to replicate as closely as possible an experimental setting, in this case a sample containing N atoms at a constant pressure P and temperature T. This is the isothermal-isobaric (NPT) ensemble, where volume can fluctuate (after all, our sample could expand) and internal energy is not necessarily conserved (due to thermal contact with the environment). As controlling the number of atoms in an MD simulation is a trivial matter, only methods to control the state variables T and P – as well as simulating a system without artificial boundaries – will be briefly discussed in the following paragraphs.

When simulating a system in equilibrium it is important to ensure that the distribution of atomic velocities matches the theoretical Maxwell-Boltzmann distribution for velocities at a temperature T (Zhou and Liu (2022)):

$$P(v_{i,\alpha}) = \Big(\frac{m_i}{2\pi k_B T}\Big)^{1/2} \exp\Big(-\frac{m_i v_{i,\alpha}^2}{2k_B T}\Big) \tag{2.15}$$

where $P(v_{i,\alpha})$ represents the probability distribution of velocities (after measuring for a

sufficiently long time) in the α Cartesian direction measured for the i^{th} atom with mass m_i, and k_B is the Boltzmann constant. This distribution has an ensemble-averaged square velocity in the α direction related to the temperature by:

$$\langle v_{i,\alpha}^2 \rangle = \frac{k_B T}{m_i} \tag{2.16}$$

At any instant t, the atom's instantaneous velocity in the α direction $v_{i,\alpha}(t)$ might define an 'instantaneous temperature' $T(t)$ in the same manner as in (2.16), and hence the ratio of the instantaneous to ensemble-averaged velocities could be written as:

$$\frac{v_{i,\alpha}(t)}{\langle v_{i,\alpha} \rangle} = \sqrt{\frac{T(t)}{T}} \tag{2.17}$$

Thus a simple attempt to control the temperature might involve rescaling velocities by dividing by the right-hand side of (2.17), however in practice this does not lead to the correct distribution of velocities (Zhou and Liu, 2022).

In this work a modified version of the Berendsen algorithm is used (Bussi et al. (2007)), where the difference between the system's instantaneous total kinetic energy $K(t)$ and the average kinetic energy at the desired temperature $\bar{K}$ is used to determine how much heat should be injected into the system during the timestep Δt in the form of velocity rescaling:

$$dK(t) = [\bar{K} - K(t)]\frac{\Delta t}{\tau_T} + 2\sqrt{\frac{K(t)\bar{K}}{N_f}}\frac{dW}{\sqrt{\tau}} \tag{2.18}$$

where N_f is the number of degrees of freedom and dW is a small stochastic term. The coupling parameter τ_T controls the speed at which deviations from the desired average kinetic energy are corrected, and is roughly analogous to the thermal conductivity of and imaginary vessel separating the sample from a thermal bath. Using the velocity scaling parameter derived from (2.18), the correct distribution of velocities is obtained.

To control the simulated system's pressure, a simple attempt might resemble (2.18), where the difference between the system's calculated instantaneous and desired pressures is used to determine how much the pressure should be adjusted during the time step Δt:

$$dP(t) = [\bar{P} - P(t)]\frac{\Delta t}{\tau_P} \tag{2.19}$$

and indeed the Berendsen barostat operates in this way. However, this barostat is not capable of of reproducing the NPT ensemble (Zhou and Liu (2022)). In this work a modified version of the Andersen barostat is employed (Andersen (1980), Parrinello and Rahman (1981)), whereby the atomic coordinates $\boldsymbol{x}_i$ are scaled by a factor inversely

proportional to the cube root of the system's volume to produce $\boldsymbol{\rho}_i = \boldsymbol{x}_i/V^{1/3}$. The equations of motion for the system are then solved using the scaled coordinates (Zhou and Liu (2022)), where two extra terms are added to the system's total kinetic and potential energies. In the case of the new total kinetic energy this becomes:

$$K(\boldsymbol{\rho}_i) = \frac{1}{2}V^{2/3}\sum_{i=1}^{N} m_i\dot{\boldsymbol{\rho}}_i \cdot \dot{\boldsymbol{\rho}}_i + \frac{1}{2}M\dot{V}^2 \tag{2.20}$$

where $\boldsymbol{\rho}_i$ is the velocity in scaled coordinates, $\dot{V}$ is the time derivative of the (now variable) system volume, and M can be thought of as the mass of an imaginary piston applied to the system to allow it to expand/contract. In the case of the system's new total potential energy, this becomes:

$$U(\boldsymbol{\rho}_i) = \sum_{i=1}^{N} U(V^{1/3}\boldsymbol{\rho}_i) + PV \tag{2.21}$$

where P is the system's desired pressure. Once the equations of motion are solved using the scaled coordinates and extra terms, the physical coordinates are recovered by $\boldsymbol{r}_i = \boldsymbol{\rho}_i V^{1/3}$. This treatment allows the simulation box to subtly expand or contract in response to pressure deviations from the desired pressure, with a larger choice of M causing a slower response as would be expected from a more massive imaginary piston.

Comparing the scale of a MD simulation (∼nm) to that of sample (∼mm), its clear that boundary conditions must be treated carefully. After all, while the vast majority of molecules within a real solution are unaffected by the influence of the container walls, if an MD simulation was modelled with rigid walls at its boundaries a sizeable fraction of molecules might be within a distance where the influence of the wall is significant. For this reason periodic boundary conditions (PBCs) are used in this work, whereby a molecule exiting the simulation box at one boundary reappears with the same velocity at the opposite boundary (Figure 2.5). What's more, a molecule near a boundary interacts with the periodic 'image' of the molecules near the opposite boundary via Equation (2.4) as if the entire system were replicated at the face of the boundary. This treatment has the benefit that the system's molecules never 'see' a rigid container wall, however care must be taken that the simulation box is of sufficient size that a molecule is never able to 'see' its periodic image (Figure 2.5).

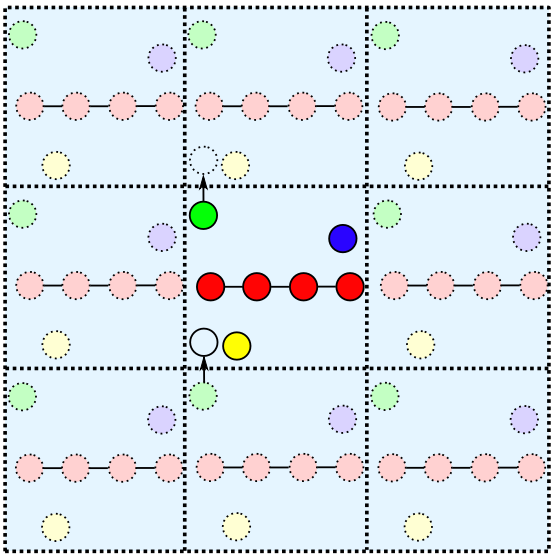

Figure 2.5: Two dimensional periodic boundary conditions. The blue particle interacts with the red chain and the periodic image of the green particle. If the green particle were to cross the upper boundary, it would reappear at the lower boundary with the same velocity. Care must be taken that the simulation box is of sufficient size that a molecule cannot interact with its own periodic image, as is the case for the red molecule.

2.2 X-ray Scattering Theory

2.2.1 Electric Fields in Matter

There are a number of challenges that must be addressed in structural studies of soft matter, including; the large range of length scales on which matter is organised, the amorphicity inherent to soft matter, and the need for non-intrusive assays. Theories governing the elastic scattering of light and X-rays, along with the development of coherent sources of radiation during the 20^{th}century, have made possible scattering techniques which overcome these challenges. In order to understand the application of scattering to the study of soft matter, some basic aspects of the theory will now be covered.

For a simple scattering experiment, typically the aim is to gain structural information for a solute – such as a biopolymer – in a solvent such as water. In this simple example, the polymer concentration in structural regions will be higher than that of the bulk solution. If the dielectric properties of the polymer differ to those of water, the permittivity will also vary between these regions. Defining the local permittivity of the sample as $\epsilon(\boldsymbol{r}, t)$ and the bulk permittivity as $\bar{\epsilon}$, the former depends on the latter plus a fluctuation from the bulk value $\delta\epsilon(\boldsymbol{r}, t)$:

$$\epsilon(\boldsymbol{r}, t) = \bar{\epsilon} + \delta\epsilon(\boldsymbol{r}, t) \tag{2.22}$$

The electric field produced by radiation passing through the sample will induce electric dipoles in the dielectric (Griffiths (2014)), which then re-radiate (scatter) in a different direction. Assuming the scattered field $\boldsymbol{E}_s$ is much smaller than the incident field $\boldsymbol{E}_i$, the total field $\boldsymbol{E}$ in the sample is then the sum:

$$\boldsymbol{E} = \boldsymbol{E}_i + \boldsymbol{E}_s \tag{2.23}$$

The electric field in the sample causes slight charge separation, inducing a 'displacement field' given by:

$$\boldsymbol{D} = \epsilon(\boldsymbol{r}, t)\boldsymbol{E} \tag{2.24}$$

Noting that an expression similar to (2.23) exists for the displacement field, and combining (2.23) and (2.24), the displacement field due to scattering is given by:

$$\boldsymbol{D}_s = \bar{\epsilon}\boldsymbol{E}_s + \delta\epsilon(\boldsymbol{r}, t)\boldsymbol{E}_i + \delta\epsilon(\boldsymbol{r}, t)\boldsymbol{E}_s \tag{2.25}$$

The last term in (2.25) is negligible, being the product of two small terms. Solving for the scattered electric field:

$$\boldsymbol{E}_s = \frac{\boldsymbol{D}_s - \delta\epsilon(\boldsymbol{r}, t)\boldsymbol{E}_i}{\bar{\epsilon}} \tag{2.26}$$

From (2.26) it is becoming clear that the scattered electric field – and therefore as will be seen the intensity of scattered radiation – depends on the incident field $\boldsymbol{E}_i$ and the local fluctuation in the sample's permittivity $\delta\epsilon(\boldsymbol{r}, t)$. To proceed, Berne and Pecora (1976) use a combination of Maxwell's equations, yielding a wave equation:

$$-\frac{\delta^2 \boldsymbol{D}_s}{\delta t^2} = \frac{c^2}{\bar{\epsilon}}\nabla^2(\delta\epsilon(\boldsymbol{r}, t)\boldsymbol{E}_i - \boldsymbol{D}_i) \tag{2.27}$$

which is first solved for the displacement field, then restated in terms of the electric field. From this, the magnitude of the field due to scattering at a detector located at a position $\boldsymbol{R}$ relative to the sample is:

$$E(\boldsymbol{R}, t) = \frac{\exp\left(i(kR - \omega t)\right)}{R}\int_V \frac{k^2 E_0}{4\pi}\left[\frac{\delta\epsilon(\boldsymbol{r}, t)}{\bar{\epsilon}}\right]\exp\left(-i\boldsymbol{q}\cdot\boldsymbol{r}\right)d^3r \tag{2.28}$$

The above equation can be broken into three terms. The first term:

$$\frac{\exp\left(i(kR - \omega t)\right)}{R}$$

is simply a spherical wave with frequency ω and wave vector k, emanating from an induced dipole located at the origin and detected at a radial distance R. The second term:

$$\frac{k^2 E_0}{4\pi}\left[\frac{\delta\epsilon(\boldsymbol{r},t)}{\bar{\epsilon}}\right]d^3r$$

Defines the strength of the dipole induced in the volume element d^3r, due to the interaction between a fluctuation in permittivity $\delta\epsilon(\boldsymbol{r},t)$ and an incident electromagnetic wave with amplitude E_0. The third term

$$\exp\left(-i\boldsymbol{q}\cdot\boldsymbol{r}\right)$$

simply shifts the phase of the spherical wave to account for the fact that the dipole may be at some position $\boldsymbol{r}$ relative to the origin. Here $\boldsymbol{q}$ is the 'scattering vector' – the most important variable in most scattering experiments. Also known as the momentum transfer vector, it is defined as the difference between incident and scattered wave vectors (Figure 2.6). Where n is the refractive index of the sample and θ the angle at which scattered radiation is detected, the scattering vector is defined as:

$$\boldsymbol{q} = \boldsymbol{k}_s - \boldsymbol{k}_i \tag{2.29a}$$

$$|\boldsymbol{q}| = \frac{4\pi n}{\lambda}\sin\left(\frac{\theta}{2}\right) \tag{2.29b}$$

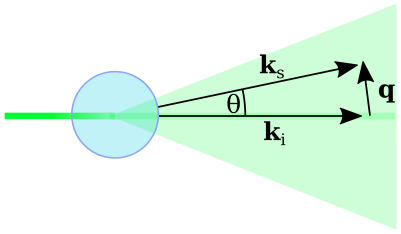

Figure 2.6: The scattering vector q is defined by the difference between incident and scattered wave vectors $\boldsymbol{k}_i$ and $\boldsymbol{k}_s$, respectively.

As Equation (2.28) indicates, the field due to scattering at a point $\boldsymbol{R}$ is due to contributions from all volume elements in the sample d^3r. The utility of (2.28) is that *the electric field due to scattering is the Fourier transform of the fluctuations in*

permittivity present in the particle. The spatial fluctuation in permittivity can be thought of as a 'scattering contrast', and is given by (Pusey (2002)):

$$\Delta\rho(\boldsymbol{r},t) = \frac{k^2}{4\pi}\left[\frac{\delta\epsilon(\boldsymbol{r},t)}{\bar{\epsilon}}\right] \tag{2.30}$$

Furthermore, the Fourier transform in (2.28) is called the particle's 'scattering length':

$$b(\boldsymbol{q},t) = \int_V \Delta\rho(\boldsymbol{r},t)\exp\left(-i\boldsymbol{q}\cdot\boldsymbol{r}\right)d^3r \tag{2.31}$$

2.2.2 Discrete Scattering Particles

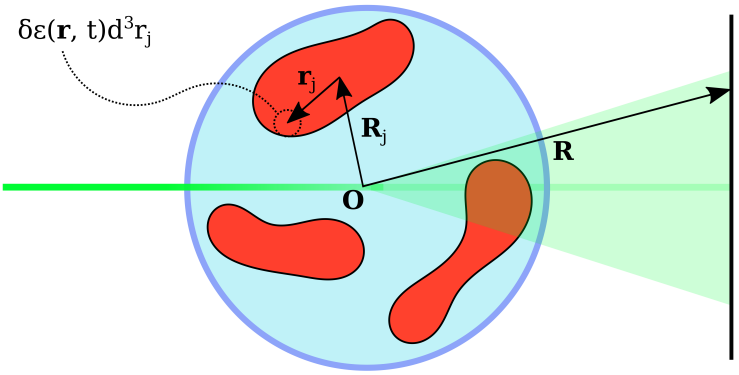

Figure 2.7: X-ray scattering from a volume element $\mathrm{d}^3\mathrm{r}_j$ within a sample at the origin **O** to a position on a detector at **R**. The j^{th} scattering particle is located at a position $\mathbf{R}_j$ relative to the origin and the volume element at a position $\mathbf{r}_j$ relative to the particle's center of mass.

Rather than describing the field scattered from a single particle with its center of mass (COM) located at the origin, (2.28) can be modified to describe a set of particles at arbitrary positions $\{\boldsymbol{R}_j\}$. To do this, the summation of (2.28) for N particles is taken, with a volume element's position relative to the origin now being the vector sum of it's COM and location of volume element d^3r relative to the COM $\boldsymbol{r} = \boldsymbol{R}_j + \boldsymbol{r}_j$ (Figure 2.7):

$$E(\boldsymbol{R},t) = -E_0 \frac{\exp\left(i(kR-\omega t)\right)}{R} \sum_{j=1}^{N} \int_{V_j} \frac{k^2}{4\pi} \left[\frac{\delta\epsilon(\boldsymbol{r}_j,t)}{\bar{\epsilon}}\right] \exp\left(-i\boldsymbol{q}\cdot(\boldsymbol{R}_j+\boldsymbol{r}_j)\right) d^3 r_j \tag{2.32}$$

where the integral is carried out over the j^{th} particle's volume V_j. Inserting (2.31) into (2.32) gives:

$$E(\boldsymbol{R},t) = -E_0 \frac{\exp\left(i(kR-\omega t)\right)}{R} \sum_{j=1}^{N} b(\boldsymbol{q},t) \exp\left(-i\boldsymbol{q}\cdot\boldsymbol{R}_j\right) \tag{2.33}$$

In an experiment, the physical quantity detected is the intensity of scattered light $I(\boldsymbol{q},t) = |E(\boldsymbol{R},t)|^2$. Taking the modulus of (2.33):

$$I(\boldsymbol{q}) = \frac{E_0^2}{R^2} \sum_{j=1}^{N} \sum_{k=1}^{N} b_j(\boldsymbol{q},t) b_k^*(\boldsymbol{q},t) \exp\left(-i\boldsymbol{q}\cdot(\boldsymbol{R}_j-\boldsymbol{R}_k)\right) \tag{2.34}$$

Two things should be noted to avoid any confusion in going from (2.33) to (2.34): the exponential prefactors in (2.33) have cancelled due to the nature of the modulus of a complex value, and the independent variable has changed from $\boldsymbol{R}$ to $\boldsymbol{q}$ as this quantity is of more significance experimentally. To proceed, note that the typical experimental acquisition time is much longer than both the time taken for the positions of the particles $\{\boldsymbol{R}_j\}$ to change significantly, and the period of field oscillation $2\pi/\omega$, and hence the spatial and temporal average of (2.34) is desired. The spatial average of the exponential term in (2.34) $\langle \exp\left(-i\boldsymbol{q}\cdot(\boldsymbol{R}_j-\boldsymbol{R}_k)\right)\rangle = \sin(qr_{jk})/qr_{jk}$ where $r_{jk} = |\boldsymbol{R}_j - \boldsymbol{R}_k|$, and so the expression for intensity due to scattering reduces to:

$$I(q) = \frac{E_0^2}{R^2} \sum_{j=1}^{N} \sum_{k=1}^{N} b_j(q) b_k(q) \frac{\sin(qr_{jk})}{qr_{jk}} \tag{2.35}$$

which is the famous Debye scattering equation (Debye (1915)).

2.2.3 Non-Interacting Particles

When there is no interaction between particles, their positions relative to one another are completely uncorrelated. The obvious exception to this is when $j = k$ – i.e. every particles' position is completely correlated with itself. The consequence of this when considering the spatially averaged scattering pattern is that when $j \neq k$ the average phase shift $\langle \exp\left(-i\boldsymbol{q}\cdot(\boldsymbol{R}_j-\boldsymbol{R}_k)\right)\rangle = 0$, further reducing (2.34) to;

$$I(q) = \frac{E_0^2}{R^2} \sum_{j=1}^{N} \langle b_j(\boldsymbol{q}) b_j^*(\boldsymbol{q})\rangle \tag{2.36}$$

Defining the 'form factor' as the modulus of the particle's scattering length, normalised by its value at $q = 0$:

$$P(q) = \frac{\langle |b(\boldsymbol{q})|^2 \rangle}{\langle |b(0)|^2 \rangle} \tag{2.37}$$

and assuming identical particles, (2.36) reduces to:

$$I(q) = N \frac{E_0^2}{R^2} \langle |b(0)|^2 \rangle P(q) \tag{2.38}$$

For most particle shapes the evaluation of form factor is difficult and often achieved using numerical techniques. Furthermore, knowledge of a particle's form factor only gives general information about its shape and lacks any information about a sample on length scales larger than the particle itself.

2.2.4 Interacting Particles

In samples where interactions between particles are present, matter is correlated on length scales larger than a single particle and the form factor alone will fail to describe the intensity of scattered light. To proceed recall Equation (2.34), which describes the scattering intensity without assuming non-interacting particles. Assuming identical particles, Equation (2.34) becomes:

$$I(q) = N \frac{E_0^2}{R^2} \langle |b(0)^2| \rangle P(q) \frac{1}{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \langle \exp\left(-i\boldsymbol{q} \cdot \boldsymbol{R}_{jk}\right) \rangle \tag{2.39}$$

where $\boldsymbol{R}_{jk} = \boldsymbol{R}_j - \boldsymbol{R}_k$. The double sum in (2.39) is called the 'structure factor' $S(q)$:

$$S(q) = \frac{1}{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \langle \exp\left(-i\boldsymbol{q} \cdot \boldsymbol{R}_{jk}\right) \rangle \tag{2.40}$$

As the name implies, the structure factor contains information on the inter-particle structure of the sample. Substituting (2.40) into (2.39) allows the scattering intensity to be written simply as:

$$I(q) = N \frac{E_0^2}{R^2} \langle |b(0)^2| \rangle P(q) S(q) \tag{2.41}$$

From (2.41) it is seen that for interacting particles, the resulting scattering intensity is a product of a function containing information about the individual particle's structure $P(q)$, and a function containing information about the structure on length scales larger than individual particles $S(q)$. Due to the inverse relationship between the size of the scattering vector q and the length scale of the scattering object d given by:

$$q = \frac{2\pi}{d} \tag{2.42}$$

one can expect that on length scales larger than individual particles, the form factor $P(q)$ takes on a constant value. Likewise, on length scales smaller than the typical inter-particle separation, the structure factor $S(q)$ takes on a constant value.

2.3 Optical Rotation Theory

2.3.1 A Molecule's Effect on the Polarization of Light

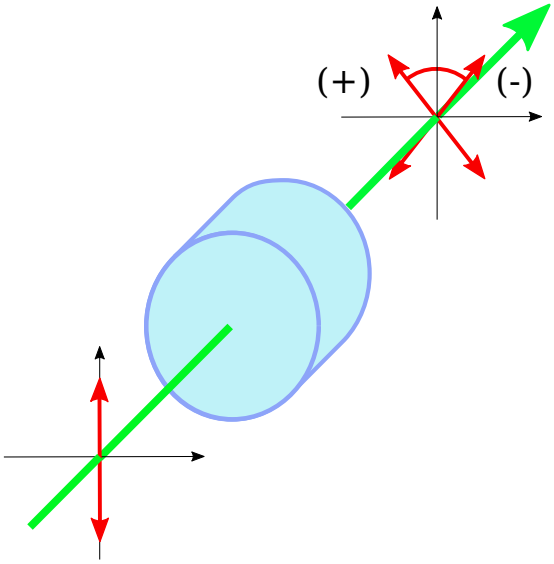

Figure 2.8: After a polarized beam passes through a chiral sample, anti-clockwise rotation of the polarization (when looking in the direction of propagation) is assigned positive optical rotation, whereas clockwise rotation is assigned negative.

Chiral materials – those either containing molecules or crystal structures for which the mirror image cannot be superimposed on the original – have been known to rotate the plane of polarized light since the 1800s (Herschel (1820)). When looking in the direction the light propagates through a material (Figure 2.8), materials that rotate the plane of polarization anti-clockwise are called dextrorotatory (right-handed, positive OR) and those that rotate it clockwise are called levorotatory (left-handed, negative OR). This phenomenon has proven useful in the pharmaceutical industry where the enantiomers of a drug – molecules with the same connectivity but different 'handedness' – often have very different affinities (H. Brooks et al. (2011)), and also in the field of biochemistry, where molecules such as proteins, nucleic acids and sugars are often chiral (Banik and

Nandi (2012); Hein and Blackmond (2012)). In this section the aim is to connect the observed OR of a molecule to its chemical structure and particular conformation. Only the essential details to do so are covered here – for a comprehensive treatment the reader is referred to Barron (2004), which the following derivation loosely follows.

Begin by considering an isotropic sample, through which a plane monochromatic wave with a frequency ω far from any resonant frequencies traverses the z direction. The field at an on-axis detector at a distance R_0 from the sample has unscattered and forward-scattered contributions:

$$E_\alpha = E_\alpha^{(0)} + iE_\alpha^s \tag{2.43}$$

Here a switch to tensor notation has been made to describe the electric field (with the use of Greek subscripts) for convenience. The second term in (2.43) is related to the unscattered wave $E_\alpha^{(0)}$ and the length of sample traversed $\mathrm{d}z$ via the scattering tensor $a_{\alpha\beta}$ (Barron (2004), (3.4.4a)):

$$E_\alpha^s = \frac{1}{2} N\omega\mu_0 c a_{\alpha\beta} E_\beta^{(0)} \mathrm{d}z \tag{2.44}$$

where N is the molecular number density, μ_0 and c are the familiar physical constants, and the scattering tensor $a_{\alpha\beta}$ depends on the molecules' response to the incident field. Among other things, $a_{\alpha\beta}$ is determined by a molecule's mixed electric dipole - magnetic dipole polarizability $G'_{\alpha\beta}(\omega)$ given by (Leszczynski et al. (2017), (159)):

$$G'_{\alpha\beta}(\omega) = -\frac{2\omega}{\hbar} \sum_{n\neq 0} \frac{\mathrm{Im}[\langle\Psi_0|\mu_\alpha|\Psi_n\rangle\langle\Psi_n|m_\beta|\Psi_0\rangle]}{\omega_n^2 - \omega^2} \tag{2.45}$$

where μ_α and m_β are the electric and magnetic dipole moment operators given by (Crawford (2006)):

$$\mu_\alpha = \sum_i q_i r_{i,\alpha} \tag{2.46}$$

$$m_\beta = \sum_i \frac{q_i}{2m_i} r_{i,\beta} \times p_{i,\beta} \tag{2.47}$$

and ω_n is the excitation frequency associated with the molecule's $0 \rightarrow n$ transition. Hence through its dependence on the scattering tensor $a_{\alpha\beta}$ (2.44), the field at the detector is affected by the molecule's physical properties (2.45). Of particular interest is the effect on the field's polarization direction relative to the x axis, which forms an angle given by (Stokes (1852)):

$$\theta = \frac{1}{2}\tan^{-1}\left(\frac{-(E_x E_y^* + E_y E_x^*)}{E_x E_x^* - E_y E_y^*}\right) \tag{2.48}$$

where use of asterisks denote complex conjugation and subscripts x and y denote components in those directions. While the change in (2.48) $d\theta$ due to traversing a small length of sample $\mathrm{d}z$ in general involves many terms (Barron (2004) (3.4.16b)), in the case of an isotropic sample this reduces to one term involving the trace of (2.45) $G'_{\alpha\alpha}(\omega)$:

$$\mathrm{d}\theta \approx -\frac{1}{3}\mu_0 N\omega G'_{\alpha\alpha}(\omega)\mathrm{d}z \tag{2.49}$$

which is easily seen to be the famous Rosenfeld equation (Rosenfeld (1928)) upon integration over the length of the sample. The experimental value most often reported, however, is the 'specific rotation' $[\alpha]$, which is essentially (2.49) normalised for sample thickness (in decimeters) and solute density (in g·cm^{-3}). For a given theoretical structure, computational chemistry programs such as Gaussian (Frisch et al. (2009)) can be used to calculate (2.45), or the related OR tensor $\beta_{\alpha\beta} = -G'_{\alpha\beta}(\omega)/\omega$. Accounting for normalization and converting to SI units, (2.49) relates the trace of the OR tensor β (in atomic units as output from Gaussian) to the specific rotation $[\alpha]$ via:

$$[\alpha] = -\frac{1 \times 10^3 \mu_0 N_A \hbar\omega}{\pi M}\beta_{\alpha\alpha} \tag{2.50}$$

where μ_0, N_A, and $\hbar$ are the familiar physical constants, ω is the frequency of the incident radiation, and M the molar mass of the solute – all in SI units.

2.3.2 Calculating Molecular Properties Using Density-Functional Theory

As seen in the previous section, and in particular (2.45), the problem of calculating the specific OR for a theoretical structure involves finding the excited states $|\Psi_n\rangle$ and their associated frequencies ω_n. As will be shown, by applying linear response theory this can be obtained via a ground-state density-functional theory (DFT) calculation. A comprehensive treatment of the current methods used to achieve this would require much more time and space than allowed for here, and hence only the essential points required to form a conceptual understanding are covered.

Following (Ullrich (2016)), begin by considering the time-independent Schrödinger equation for a system containing N electrons, where it is assumed the atomic nuclei are fixed in space at positions $\{\boldsymbol{r}_i\}$ (Born and Oppenheimer (1927)):

$$\hat{H}_j\Psi(\boldsymbol{r}_1, ..., \boldsymbol{r}_N) = E_j\Psi(\boldsymbol{r}_1, ..., \boldsymbol{r}_N). \tag{2.51}$$

In the above, the operator $\hat{H}$ is the Hamiltonian of the system given by:

$$\hat{H}(t) = \hat{T} + \hat{V} + \hat{W} \tag{2.52}$$

Here $\hat{T}$ is the operator for kinetic energy, $\hat{W}$ the operator for inter-electron interactions, and $\hat{V}$ the operator for the potential energy of the N electrons within a constant external field:

$$\hat{V} = \sum_{j=1}^{N} v(\boldsymbol{r}_j) \tag{2.53}$$

The complexity in solving the many-body Equation (2.51) was circumvented by Kohn and Sham (Kohn and Sham (1965)) where the inter-electron interactions embodied by $\hat{W}$ are instead bundled into an effective potential (2.53), which for a single electron becomes:

$$v_s(\boldsymbol{r}) = v_0(\boldsymbol{r}) + v_{e-e}(\boldsymbol{r}) \tag{2.54}$$

where $v_0(\mathbf{r})$ in the simplest case is due to the nuclei and $v_{e-e}(\mathbf{r})$ approximates inter-electron interactions. This simplifies solving Equation (2.51) to solving N Kohn-Sham equations for single electrons j:

$$\left[\hat{T} + v_s(\boldsymbol{r})\right]\psi_j(\boldsymbol{r}) = \epsilon_j \psi_j(\boldsymbol{r}) \tag{2.55}$$

To give the single-electron orbitals $\psi_j(\boldsymbol{r})$ in (2.55) a functional form, they are expanded as a linear combination of basis functions $\phi_i(\boldsymbol{r})$ with coefficients c_{ji}:

$$\psi_j(\boldsymbol{r}) = \sum_i c_{ji}\phi_i(\boldsymbol{r}) \tag{2.56}$$

The wave function is then expressed via the Slater determinant of single-electron orbitals:

$$\Psi_s(\boldsymbol{r}_1, ..., \boldsymbol{r}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(\boldsymbol{r}_1) & \psi_2(\boldsymbol{r}_1) & \dots & \psi_N(\boldsymbol{r}_1) \\ \psi_1(\boldsymbol{r}_2) & \psi_2(\boldsymbol{r}_2) & \dots & \psi_N(\boldsymbol{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(\boldsymbol{r}_N) & \psi_2(\boldsymbol{r}_N) & \dots & \psi_N(\boldsymbol{r}_N) \end{vmatrix} \tag{2.57}$$

It must be noted that Ψ_s above does *not* represent Ψ in (2.51), however it *does* produce the same electron density, which in the Kohn-Sham scheme is given by:

$$n(\boldsymbol{r}) = \sum_{j=1}^{N} |\psi_j(\boldsymbol{r})|^2 \tag{2.58}$$

The computational method of DFT relies on the fact that by modelling the electron-electron interaction instead as an electron-potential interaction, one is still able to reproduce the correct electron density (Ullrich (2016)). Furthermore, *all physical observables are functionals of the electron density*, including the single electron potential (2.54). This forms a loop, with the Kohn-Sham Equation (2.55) both requiring the electron density (through $v_s(\boldsymbol{r})$) and defining it (it's solutions being used in (2.58)). Solution for the ground-state involves solving (2.55) iteratively using a self-consistent field treatment, as shown in Figure 2.9.

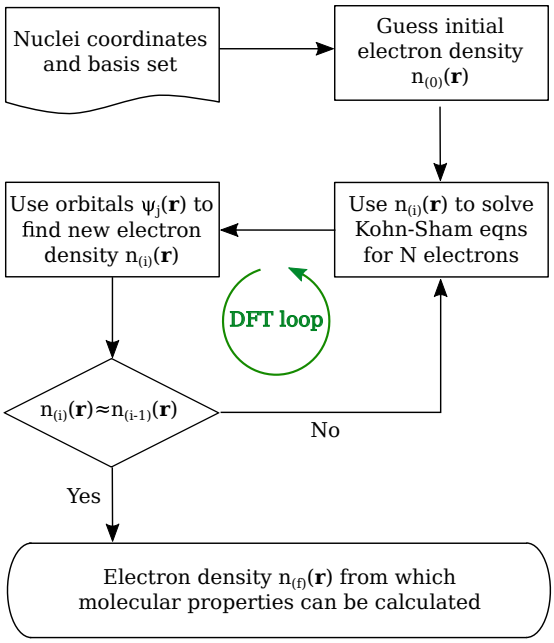

Figure 2.9: Density-functional theory algorithm. Here $n_{(i)}(\boldsymbol{r})$ denotes the electron density on the i^{th} iteration of the algorithm, whereas $\psi_j(\boldsymbol{r})$ denotes the j^{th} orbital.

If the potential in (2.54) were to include a time-dependent term $v_1(\boldsymbol{r}, t)$, e.g. due to an oscillating electromagnetic field such that the potential became:

$$v_s(\boldsymbol{r}, t) = v_0(\boldsymbol{r}) + v_1(\boldsymbol{r}, t) + v_{e-e}(\boldsymbol{r}, t) \tag{2.59}$$

we would instead have to solve the time-dependant Kohn-Sham equation:

$$\left[\hat{T} + v_s(\boldsymbol{r}, t)\right]\psi_j(\boldsymbol{r}, t) = i\frac{\partial}{\partial t}\psi_j(\boldsymbol{r}, t) \tag{2.60}$$

With some changes to the algorithm described above for the self-consistent field solution to the time-independent case, the time-dependant Kohn-Sham Equation (2.60) can also be solved in an iterative manner, propagating the electron density from the initial

ground-state density $n_0(\boldsymbol{r}, t_0)$ towards some final density $n(\boldsymbol{r}, t)$.

In practice, propagating an electron density from its ground state to some final state under the influence of a time-dependant potential is computationally expensive, and fortunately alternative DFT methods exist to find the mixed electric dipole-magnetic dipole polarizability (2.45). In particular, when the time-dependant term $v_1(\boldsymbol{r}, t)$ in (2.59) can be considered a small a perturbation, linear-response theory can be applied. This models the change in electron density (the response) due to a time-dependant perturbation $v_1(\boldsymbol{r}, t)$ as an expansion in powers of the perturbation, with only the first-order term being considered:

$$n(\boldsymbol{r}, t) - n_0(\boldsymbol{r}) \approx \int dt' \int d\boldsymbol{r}' \chi(\boldsymbol{r}, t, \boldsymbol{r}', t') v_1(\boldsymbol{r}, t') + ... \tag{2.61}$$

where χ is the response function, simply the derivative of the ground-state electron density with respect to a perturbation. When Fourier-transformed to express the response in terms of the perturbation frequency ω instead of t, singularities appear at the excitation frequencies Ω, which are found by solving a set of eigenvalue equations most succinctly described by the Casida equation (Casida (1995)):

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = \Omega \begin{bmatrix} -\mathbf{I} & 0 \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \tag{2.62}$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{X}$ and $\mathbf{Y}$ all contain information about the ground-state orbitals and electron interactions, and $\mathbf{I}$ is an identity matrix of the same size as the sub-matrices $\mathbf{A}$ and $\mathbf{B}$. This equation ultimately leads to the response function $G'_{\alpha\beta}$ for a perturbation with frequency ω via (Crawford (2006)):

$$G'_{\alpha\beta}(\omega) = -\mathrm{Im}\left(\mu_\alpha \begin{bmatrix} \mathbf{A} - \omega\mathbf{I} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} + \omega\mathbf{I} \end{bmatrix}^{-1} m_\beta \right) \tag{2.63}$$

Thus the OR, which is proportional to the trace $G'_{\alpha\alpha}$ (2.50), is determined by the structure of the molecule.

Chapter 3

κ-Carrageenan in the Disordered-State

3.1 Introduction

To understand the disorder-order transition known to occur for solutions of κ-carrageenan upon the addition of salt, an understanding of the structure of the disordered-state is first sought, as well as what uni-molecular structural changes might occur upon the addition of salt. Although a wealth of structural studies exist focusing on ι-carrageenan in the ordered-state (Anderson et al. (1969); Arnott et al. (1974); Janaswamy and Chandrasekaran (2002)), κ-carrageenan in the disordered-state has been primarily studied using NMR experiments (Grasdalen and Smidsrød, 1981a,b; Rochas et al., 1983; Paoletti et al., 1985; Stortz and Cerezo, 1992; Zhang et al., 1992; Ueda et al., 2001; Bosco et al., 2005; Zhang et al., 2010). The majority of these NMR studies have sought information about the interaction between mobile ions and the polymer chain, however those that have sought structural information about the polymer itself have reported a 'loose' right-handed three-fold helix in the disordered-state (Bosco et al., 2005). Recently, a series of AFM studies investigating the conformations adopted by carrageenan chains adsorped to functionalised surfaces under disordered-state ionic conditions have concluded that the disordered-state represents a 'random coil' (Schefer et al., 2015a,b; Diener et al., 2019). Furthermore, during AFM experiments performed under ordered-state ionic conditions, the dimensions of single chains appeared to change, which was interpreted to support the model of a uni-molecular disorder-order transition previously proposed based on OR experiments (Grasdalen and Smidsrød, 1981a; Smidsrød and Grasdalen, 1984; Paoletti et al., 1984; Ciancia et al., 1997; Bongaerts et al., 1999). However, in addition to the natural questions that arise when comparing an adsorped molecule's conformation to its conformation in solution, there is also considerable evidence to suggest the disorder-order transition is a cooperative process (Anderson et al.,

1969; Bryce et al., 1982; Norton et al., 1983; Austen et al., 1988; Hjerde et al., 1998, 1999). The few MD simulation studies carried out for carrageenan systems have generally supported the findings of the NMR study, reporting a restricted conformational space as opposed to a random coil (Ueda et al., 1998, 2001; Brzyska et al., 2018; Perez and Claudio, 2020), however these modelling studies to-date have only involved short carrageenan oligomers.

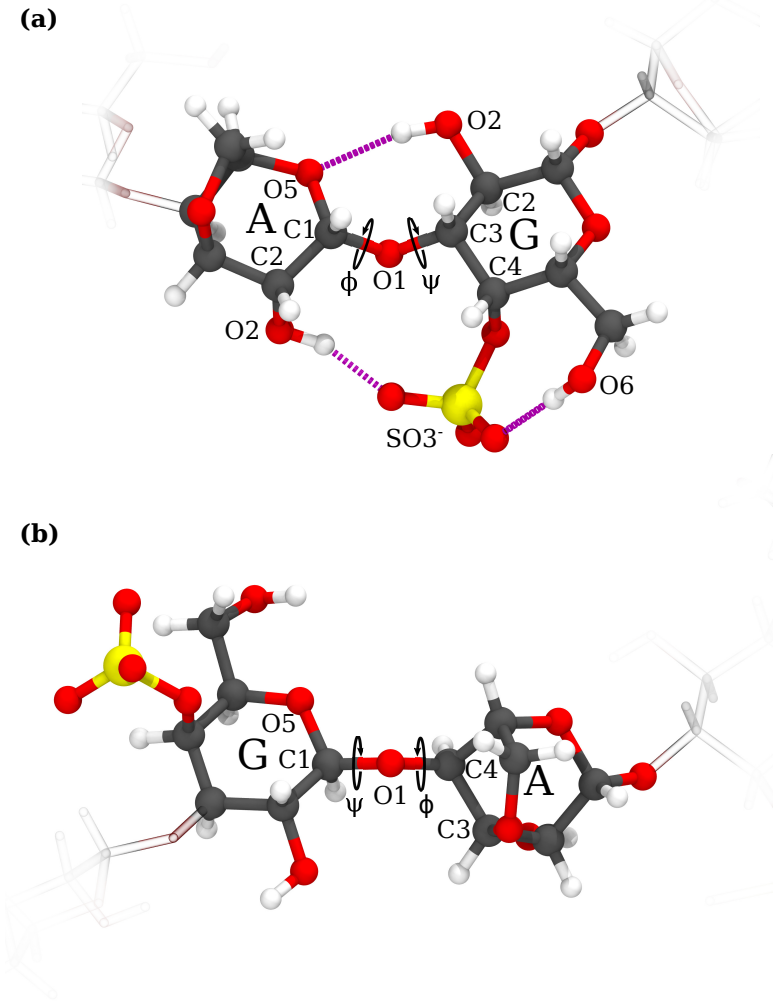

Figure 3.1: Glycosidic (a) $1 \rightarrow 3$ linkage and (b) $1 \rightarrow 4$ linkage. Anhydrogalactopyranose residues are labelled 'A' and galactopyranose residues 'G'. Relevant atoms or groups are also labelled, along with dihedral angles (ϕ and ψ), and hydrogen bonds (dashed magenta lines).

Considering the incongruity in the literature regarding the κ-carrageenan disordered-state, and the need to describe this state for a full understanding of the disorder-order

transition so central to carrageenan's physical properties, an investigation using state-of-the-art computational and experimental techniques is timely. The computing power of current generation supercomputers allow large-scale atomistic MD simulations to be run in parallel in days – what would have taken months previously (Hollingsworth and Dror, 2018). Similarly, recent advances in synchrotron-based WAXS now allow structural information to be obtained on length scales less than 0.5 nm, previously only possible for (semi-)crystalline materials (Mansel et al., 2020). The combination of these two techniques – which allows high resolution structural information to be obtained for solution-state samples (Hub, 2018; Mansel et al., 2019)– will hereby be used to understand: (1) the secondary structure of κ-carrageenan chains in the disordered-state, and (2) what uni-molecular structural changes might occur upon the addition of NaI. The choice of this salt is made as its cation is known to induce OR-detected structural changes, whereas its anion is thought to prevent network formation (Grasdalen and Smidsrød, 1981a; Ciancia et al., 1997), allowing any uni-molecular structural changes to be isolated from those that might arise due to the formation of higher order structures.

3.2 Experimental X-ray Scattering

3.2.1 κ-Carrageenan

κ-carrageenan was obtained as a dried powder from Danisco (journal ID 2654-164-2). The carrageenan content was 97% κ-carrageenan, with the remaining 3% being primarily ι-carrageenan, as determined by the supplier. The ion composition, as cited by the supplier, is shown in Table 3.1. Samples were prepared without modification to the carrageenan, as detailed in the following subsection.

Na^+	K^+	Mg^{2+}	Ca^{2+}
4.94 %	0.77 %	0.07 %	0.02 %

Table 3.1: Ionic composition of κ-carrageenan samples as a percentage of total dry weight.

3.2.2 Sample Preparation

κ-carrageenan solutions were prepared in the disordered-state by slowly dissolving dried κ- carrageenan powder into deionised water (18.2 MΩ cm) at a polymer concentration of 1% (wt/wt). Polymer concentrations of 0.5% and 2% were also prepared, however much of the following analysis focuses on the 1% sample. Mixing continued for a number of hours to allow full dissolution, followed by heating to 75 °C and filtration

(pore size 0.2 μm) directly into cells used for the WAXS instrument sample auto-loader. To obtain the excess scattering due to the polymer, scattering from solutions without polymer were also collected for background subtraction. These were prepared in a similar manner as detailed above, bar the polymer dissolution step.

3.2.3 X-ray Scattering Experiments

The SAXS/WAXS beamline at the Australian Synchrotron (Kirby et al., 2013) was used for all scattering experiments, with the wavelength of the monochromatic beam of X-rays set to $\lambda = 0.83$Å. To avoid damage due to the beam, samples flowed continuously through a capillary tube at a rate of 5 $\mu L\, s^{-1}$ during data acquisition. Following the sample stage, the scattered X-rays passed through a vacuum vessel, before reaching a Pilatus 1M detector (Dectris, Switzerland) housed at a sample-to-detector distance of 0.738 m. Multiple 1 s acquisitions of the scattering profile were made, which were subsequently corrected and integrated to form 1D scattering profiles using the beamline's scatterBrain software. The excess scattering was obtained by subtracting the polymer-free scattering from the scattering for polymer solutions, which was then normalised using the scattering from pure water as a standard.

3.3 Molecular Dynamics Simulations

3.3.1 Initial Configurations

The atomic coordinates for double-helical ι-carrageenan obtained using XRD (Arnott et al. (1974): Protein Data Bank entry 1CAR) were used as a starting point for the construction of κ-carrageenan chains. The conversion to single κ-carrageenan helices involved first removing one chain, then replacing the C2 sulfates on remaining G residues with hydroxyl moieties (Figure 3.1). These structures were then used to calculate the partial charges (Figure 2.3) for atoms within central residues – as well as the two residues at the chain's termini – using the B3LYP DFT (Section 2.3.2, Lee et al. (1988)) and Hartree-Fock (Roothaan (1951)) methods with the 6-31+G(d,p) basis set (Ditchfield et al. (1971)). The AM1-BCC charge model was also used (Jakalian et al. (2002)), which was found to be a good compromise between the two former methods, and hence was chosen for use.

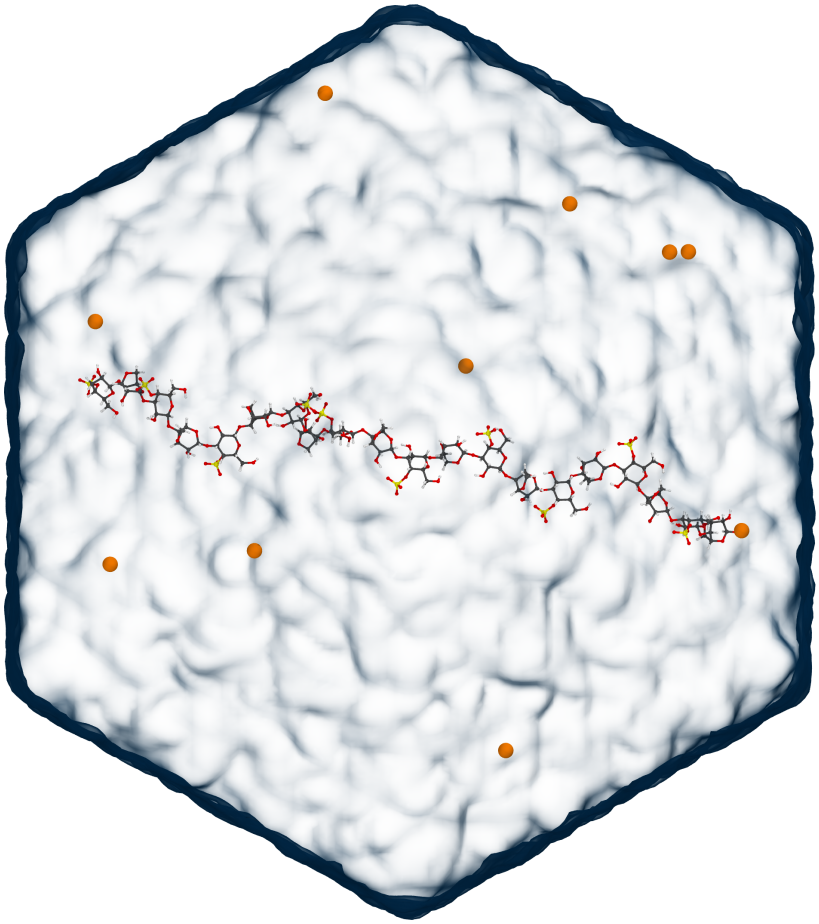

Figure 3.2: A dodecahedron simulation box containing explicit solvent (rendered as a surface), Na^+ ions, and one κ-carrageenan chain.

Partial charges and internal coordinates for each of the two central and terminal residue types, along with parameters from the GLYCAM-06j force field, were then passed to the AmberTools program LEaP to produce structure and topology files for chains with a degree of polymerisation (DP) of 10 (i.e. 20 sugar residues). This force field – which was chosen for its performance modelling carbohydrate systems (Kirschner et al., 2008b) – is the main force field used in this work, however topology files were also produced using the General Amber Force Field (GAFF) parameters to test the results' sensitivity to the parameters used.

Initial configurations were constructed by centering the DP 10 chain in a dodecahedron box with PBCs and dimensions sufficient to avoid interaction with the molecule's periodic image. The simulation box was then filled with 21173 SPC/E water molecules (Berendsen et al., 1987), a set of parameters which are computationally efficient and

replicate experimental behaviour when using the GLYCAM force field for carbohydrate systems (Irani et al. (2017); Mansel et al. (2019); Williams et al. (2020)). Water molecules were then replaced at random with either: 10 Na^+ ions to neutralise the polymer charge (Figure 3.2), or 49 Na^+ and 39 I^- ions to give a NaI concentration of 100 mM. Simulations with only neutralising Na^+ ions and those with both additional Na^+ and I^- are referred to in this as well as subsequent chapters as systems without NaI, and systems with NaI, respectively.

3.3.2 Equilibration and Production Simulations

Equilibration of solvated systems began by energetically minimizing the chain structure such that the maximum force on any atom was less than 1000 $kJ{\cdot}mol^{-1}{\cdot}nm^{-1}$, using the GROMACS (Pronk et al., 2013) steepest descent algorithm. The polymer's coordinates were then restrained while solvent underwent 100 ps of isochoric equilibration while maintaining a temperature of 20 °C using the V-scale thermostat (Bussi et al., 2007), followed by 100 ps of unrestrained isobaric equilibration maintaining a pressure of 1.00 bar using the Parrinello-Rahman barostat (Parrinello and Rahman, 1981).

GROMACS 2020.6 was then used to carry out ten productions simulations (five of each salt type), each of 200 ns duration and with velocities and coordinates being updated every 2 fs. Temperatures of 20 °C and pressures of 1.00 bar were maintained using the aforementioned algorithms, while the Particle-Mesh Ewald method was used to treat electrostatic interactions (Essmann et al., 1995). Two production simulations using GAFF parameters (one of each salt type) were also run in the same manner as described above, for a total of twelve unbiased simulations.

Additional simulations were run while using well-tempered metadynamics to explore the free energy surface of rotating dihedral angles ϕ and ψ (Figure 3.1) about the glycosidic linkages, using the following definitions:

$$\begin{aligned} \phi_{1\to3} &= \mathrm{AC2} - \mathrm{AC1} - \mathrm{AO1} - \mathrm{GC3} \\ \psi_{1\to3} &= \mathrm{AC1} - \mathrm{AO1} - \mathrm{GC3} - \mathrm{GC4} \\ \phi_{1\to4} &= \mathrm{GC1} - \mathrm{GO1} - \mathrm{AC4} - \mathrm{AC3} \\ \psi_{1\to4} &= \mathrm{GO5} - \mathrm{GC1} - \mathrm{GO1} - \mathrm{AC4} \end{aligned} \tag{3.1}$$

One simulation for each of the four combinations of salt condition (with/without NaI) and glycosidic linkage ($1 \to 3$- or $1 \to 4$-linked) was run, during which the PLUMED 2 plug-in for GROMACS (Laio and Parrinello, 2002; Tribello et al., 2014) was used to deposit Gaussian potentials with a height of 1.0 $kJ{\cdot}mol^{-1}$ and width of 0.2 radians ($\sim$11.5°) every 1 ps in the CV space defined by (3.1). As with the unbiased simulations,

an additional four metadynamics simulations using GAFF parameters were run to test the sensitivity of the FES to parameters used, for a total of eight metadynamics simulations.

3.4 Results

3.4.1 Experimental Wide-Angle X-ray Scattering

Experimental WAXS for 1% (wt/wt) κ-carrageenan solutions at 20°C in the disordered-state is presented in Figure 3.3. Four distinct features appear across the q range measured, corresponding to prevalent length scales within the sample, whose characteristic size is related to the location of the peak via (2.42). At $q_1 = 0.50\,\mathrm{nm}^{-1}$, this corresponds to a prevalent length scale of 12.67 nm. The broad shoulder at around $q_2 = 5.70\,\mathrm{nm}^{-1}$ and a subtle peak at $q_3 = 9.91\,\mathrm{nm}^{-1}$ correspond to length scales of 1.10 nm and 0.63 nm, respectively. Finally, at the highest accessible q values, another pronounced peak is found at $q_4 = 14.6\,\mathrm{nm}^{-1}$, indicative of a length scale of 0.43 nm. Between q_1 and $q \approx 8\,\mathrm{nm}^{-1}$, the scattering intensity $I(q)$ scales with q with an exponent of -1.386 ± 0.005 (here, as elsewhere in the text unless specified, the uncertainty is the 95% confidence interval).

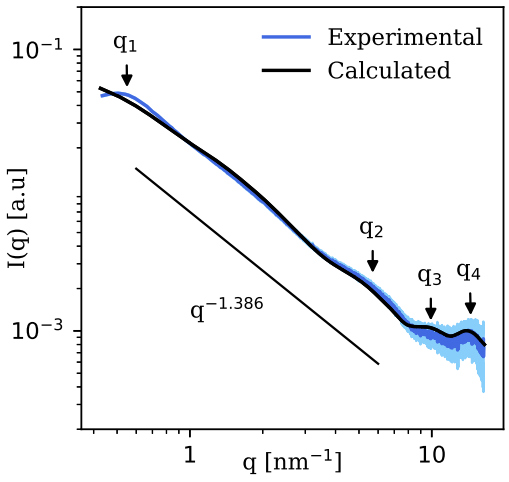

Figure 3.3: WAXS for 1% (wt/wt) κ-carrageenan solutions without NaI at a temperature of 20°C, with the lighter shaded region showing the 95% confidence interval. The solid black line shows the mean of scattering profiles calculated from simulated systems without NaI. Points of interest are labelled q_{1-4}, and a line showing $I(q) \propto q^{-1.386}$ is added as a guide.

As Figure 3.4 (a) shows, the shape of WAXS profiles from disordered-state solutions of κ-carrageenan with polymer concentrations of 0.5 and 2.0 % (wt/wt) differ from the profile from 1.0 % solutions only in location of the peak q_1, which shifts to higher q (corresponding to smaller length scales) as the concentration is increased. The locations of peaks q_{2-4} – as well as the scaling relationship $I(q) \propto q^{-1.386}$ – remain the same.

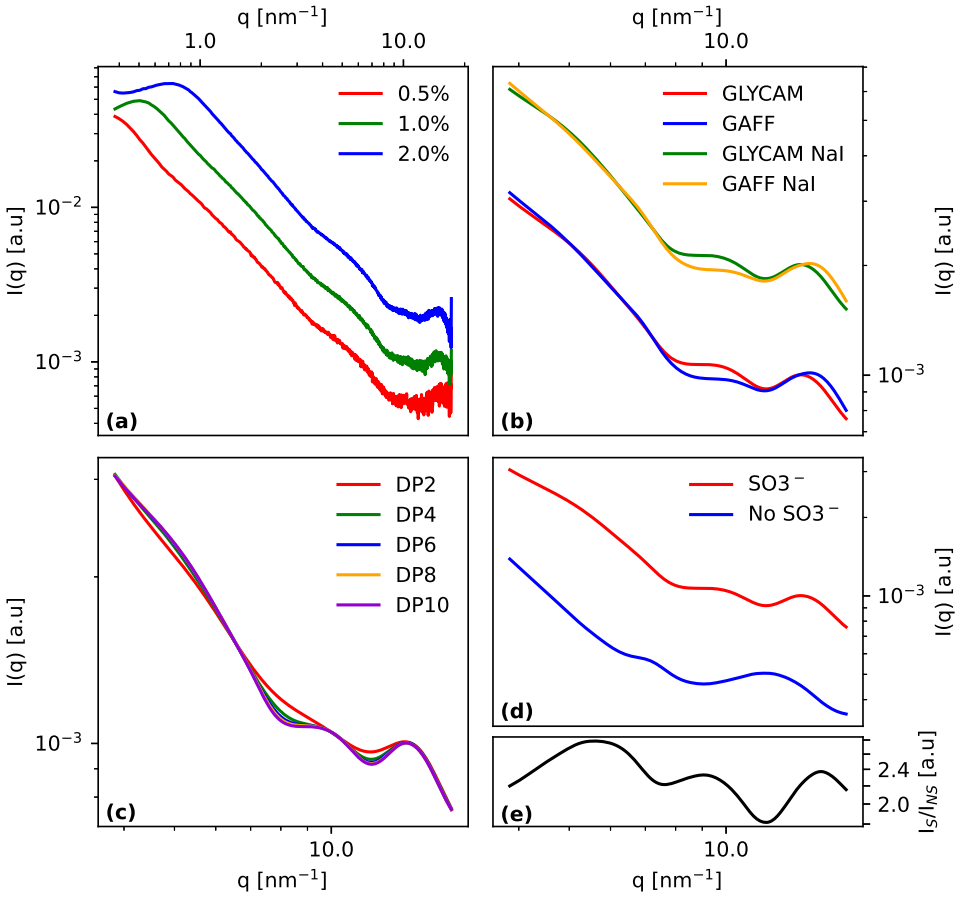

Figure 3.4: (a) Experimental scattering for various κ-carrageenan concentrations (wt/wt) in salt-free solutions. (b) Calculated scattering for simulated chains 10 disaccharides in length (DP10) with the GLYCAM and GAFF force fields, either without additional salt or with 100 mM NaI. The low-q scattering has been truncated as there were no significant differences between the four profiles. (c) Calculated scattering for fragments of various lengths extracted from one GLYCAM DP10 simulation without additional salt. (d) Scattering from GLYCAM DP10 simulations without additional salt, calculated either with all atomic coordinates, or atomic coordinates excluding sulfate atoms. (e) Ratio of the scattering profiles calculated with/without SO3$^-$ atoms shown in (d).

3.4.2 Simulated Wide-Angle X-ray Scattering

Simulated atomic coordinates and CRYSOL 3 (Franke et al., 2017) were used to calculate WAXS from the second half of each unbiased simulation at intervals of 10 ps, generating 50,000 WAXS profiles for each of the two salt conditions. These were then averaged to produce two WAXS profiles representative of the ensemble of conformations observed with/without NaI. The averaged profiles were then fitted to the experimental data above q_1 by minimizing the following measure of fitness:

$$\chi^2 = \frac{1}{N_q} \sum_{i=1}^{N_q} \frac{(\log[fI_c(q_i)] - \log[I_s(q_i) - \nu I_b(q_i) - c])^2}{\sqrt{(fSE_c)^2 + SE_s^2 + (\nu SE_b)^2}} \tag{3.2}$$

where a χ^2 closer to unity indicates a better fit. Here N_q is the number of data points above q_1 and I_s, I_b, and I_c are the solution, buffer, and calculated scattering intensities, respectively, whereas the standard errors of these three values appear in the denominator. The fitting parameters f, ν, and c denote intensity scaling, displaced buffer correction, and any small amount of background radiation present, respectively. To ensure a good fit across multiple orders of magnitude, the fit is made using a logarithmic scale. The profile calculated for atomic coordinates from simulations without NaI is shown in Figure (3.3), whose best fit (χ^2=0.71) to the experimental data was when values of ν and c were close to 1 and 0 (respectively) and f was varied.

The ensemble-averaged WAXS profile calculated using atomic coordinates from simulations with NaI was indistinguishable to that calculated from simulations without NaI (Figure 3.4 (b)). Furthermore, WAXS profiles calculated from simulations using the GAFF parameters only showed modest changes in the size and location of peaks located at q_3 and q_4. These findings show that the simulated polymer conformation was insensitive to salt concentration, and only mildly sensitive at small length-scales to the parameter set used.

To understand the origin of the features in WAXS profiles at q_{2-4}, profiles were calculated using chain fragments of various lengths extracted from a simulation without NaI (Figure 3.3 (c)). As the fragment length increases, so too did the definition of the three peaks, suggesting they have their origins in some linear repeat distance. When the scattering was calculated using the atomic coordinates of a simulated full-length chain minus sulfate atoms (Figure 3.3 (d)), the scattering intensity was considerably lower at q_{2-4} relative to the scattering from intact chains (Figure 3.3 (e)).

3.4.3 Simulated Intra-Molecular Distances

To assign the features in WAXS profiles to prevalent length scales within the disordered-state structure proposed here, the simulated coordinates were used to calculate the

Group	**d [nm]**	$\boldsymbol{\sigma}$ **[nm]**
A-G_{+1}	0.52	0.02
G-A_{+1}	0.66	0.02
A-A_{+2}	0.98 *	0.04
G-G_{+2}	1.12	0.05
GSO3-G$'$	0.41	0.01
GSO3-G_{+2}SO3	1.35	0.09

Table 3.2: Intra-molecular distances and their standard deviations (σ) for simulated systems without NaI. 500,000 measurements spanning a cumulative 500 ns of simulation time were made for each intra-molecular distance. G$'$ indicates the G residue excluding sulfate group atoms. * Second minor mode at 1.03$\pm$0.02 nm.

intra-molecular distances between electron-rich groups, which contribute strongly to scattering. The notation used to specify positions follows the format: R_NS, where R denotes the residue type, S a set of atoms within that residue, and N (when spanning a pair of residues) indicates the number of residues closer to the reducing or non-reducing end of the chain for positive or negative values of N, respectively. When a set of atoms is specified, their position is taken as the COM, and when no set is specified the COM of the residue is used. The ensemble-averaged intra-molecular distances between electron-rich groups for simulations without NaI are shown in Table 3.2. Compared to these, there were no significant differences for simulations with NaI.

The insensitivity of the molecular structure to NaI concentration is made clear in Figure 3.5, showing the change in every intra-molecular distance upon the addition of NaI. This involved first calculating the intra-molecular distance between every pair of polymer atoms, then averaging each of these values over all simulations with/without NaI, and finally taking the difference between the two averages for each intra-molecular pair. As expected, the largest changes occurred between atoms near opposite ends of the chain; for instance atoms 7 and 351 became on average 0.47 nm further apart in the presence of NaI. However, even the largest changes between simulations with/without NaI were only $\sim 5\%$ of the average chain end-to-end distance, which is small relative to the changes between individual simulations of the same salt condition. Furthermore, intra-molecular distance increases were balanced out by decreases, meaning that overall there was no net expansion or contraction of the polymer when NaI was present. Similarly, there were no significant differences between simulations using GAFF parameters with/without NaI.

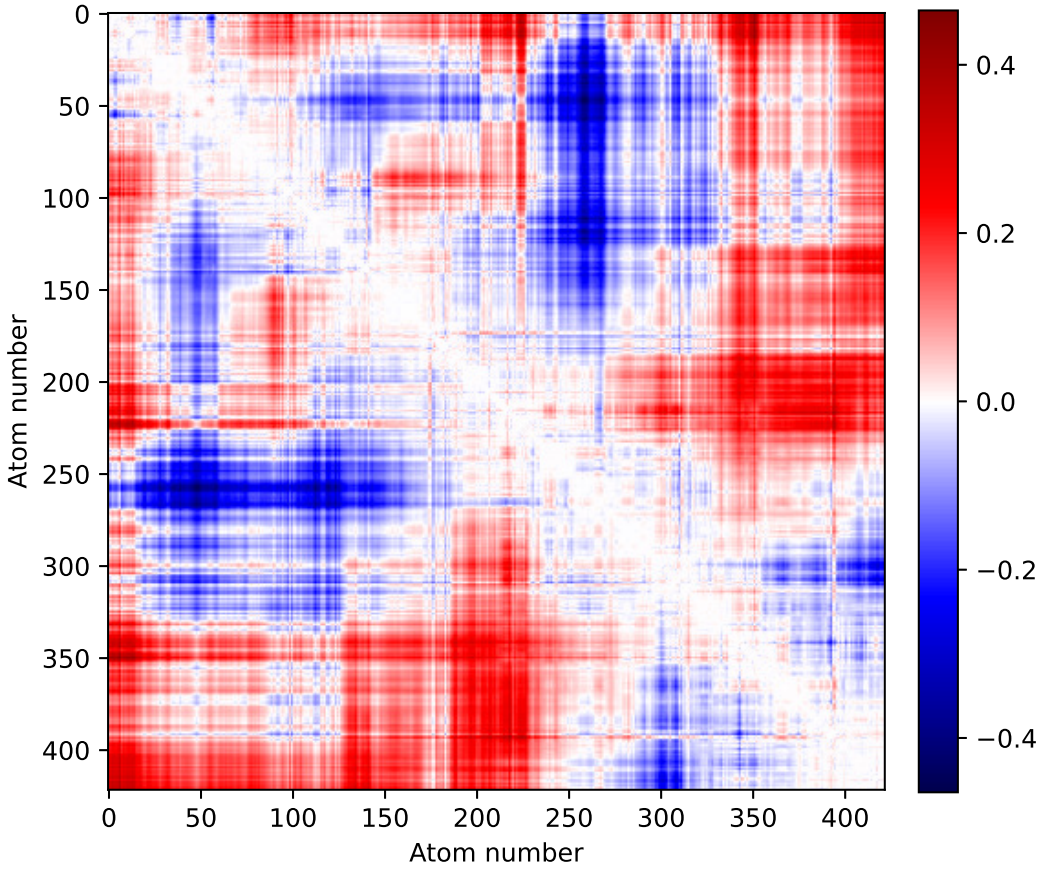

Figure 3.5: Average difference in whole-molecule inter-atomic distances between GLYCAM-simulated systems with and without NaI. The polysaccharide chain during one simulation without NaI formed a persistent bend and hence has been excluded as an outlier. Colorbar units are in nm, where red/blue indicates the inter-atomic distance between a pair of atoms (averaged over all simulations of the same salt condition) increased/decreased during simulation of NaI systems compared to systems without additional salt.

The single chain's structural insensitivity to NaI concentration may be in part due to the way Na^+ and I^- ions interact with the chain. As can be seen in the spatial distribution of I^- ions during simulations with NaI shown in Figure 3.6 (a), the ions are depleted in the region immediately surrounding the polymer. In contrast, an atmosphere of Na^+ forms nearby the chain (Figure 3.6 (b)), being most concentrated in a

region 0.24 nm from $SO3^-$ groups (Figure 3.6 (c)). This is coincident with the exterior boundary of the first hydration shell (defined as the first minimum in solvent-$SO3^-$ RDF (Laage et al., 2017)). Despite Figure 3.6 showing clear differences due to the electrostatic interaction between ions and the negatively charged chain, there is nothing to suggest site binding of ions occurred.

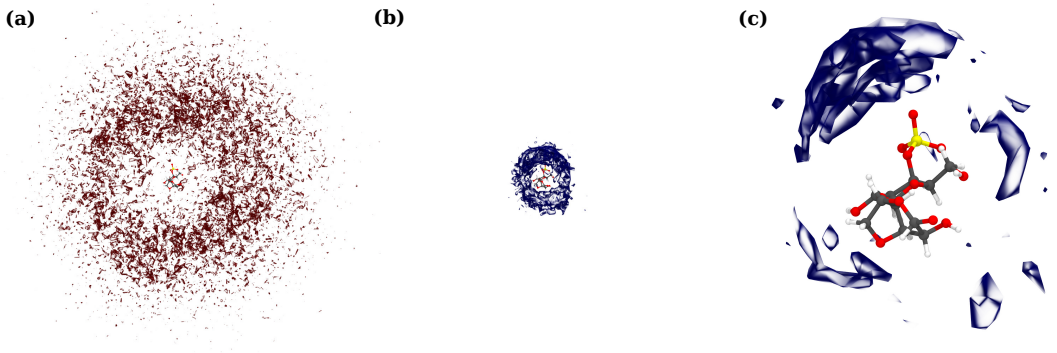

Figure 3.6: Spatial distribution of ions in a 100 mM NaI solution for (a) I^- ions and (b) Na^+ ions surrounding a κ-carrageenan chain. Shown in (c) is the same distribution as that in (b) but zoomed in and rendered at an iso-probability value 1.5 times that of (a) and (b). In all three images the view is down the axis of the chain, however only two central residues have been rendered for clarity.

3.4.4 Glycosidic Linkage Dihedral Angles

Of all the parameters affecting a linear polysaccharide's conformation, the preferred dihedral angles across the glycosidic linkages make the largest contribution, and these have been the subject of much study within the carrageenan community. The free energy surfaces of rotation about the ϕ and ψ dihedral angles for both linkages (Figure 3.1) – generated from the metadynamics simulations without NaI described in Section 3.3.2 – can be seen in Figure 3.7, along with contours representing the probability density of a particular combination of ϕ and ψ during unbiased simulations.

A summary of the energetic minima identified for each of the FESs can be found in Table 3.3. Relative to the deepest minimum (A) for the $1 \rightarrow 3$ linkage shown in Figure 3.7 (a), minima B and C were separated by potential barriers of ∼4.1 and ∼10.6 kJ·mol^{-1} respectively. Relative to the deepest minimum (Y) for the $1 \rightarrow 4$ linkage shown in Figure 3.7 (b), minima X and Z were separated by potential barriers of ∼3.3 and ∼10.2 kJ·mol^{-1} respectively. For all of these potential barriers, the height was much larger than the statistical error in the region of the barrier (Figure 3.8), which was less than 1.0 kJ·mol^{-1} in most cases. The FES generated with NaI were very similar to those generated without, however moderate differences were observed

when GAFF parameters were used (Figure 3.9). In particular, for a given linkage using GLYCAM parameters simulations populated two main energetic minima, whereas using GAFF parameters simulations preferred one of the two main minima observed using GLYCAM.

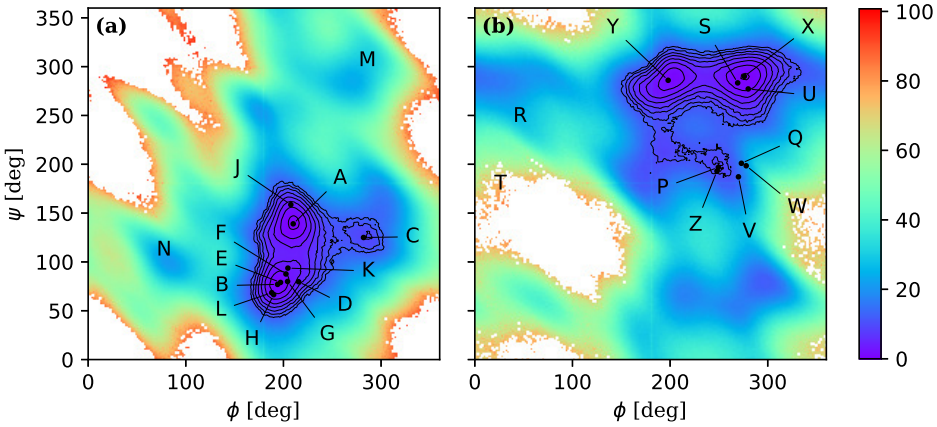

Figure 3.7: Free energy surfaces for glycosidic linkage dihedral angles generated from metadynamics simulations for the (a) $1 \rightarrow 3$ and (b) $1 \rightarrow 4$ glycosidic linkages in simulations without NaI, where the colorbar is in units kJ·mol^{-1} relative to the global minimum and white indicates sterically forbidden regions. The statistical error in free energy values is shown in Figure 3.8. The overlaid contour map is proportional to the probability density from unbiased simulations, where successive contours indicate a doubling of the probability density. For the $1 \rightarrow 3$ linkage, points A, B, and C are the energetic minima identified in this study, whereas minima D (Anderson et al., 1969), E (Arnott et al., 1974), F (Urbani et al., 1993), G (Le Questel et al., 1995), H (Ueda et al., 1998), J-K (Stortz and Cerezo, 2000), L-M (Ueda et al., 2001), and N (Bosco et al., 2005) are those identified by other studies of κ-carrageenan and its related forms. For the $1 \rightarrow 4$ linkage, points X, Y, and Z are minima identified in this study, whereas P (Anderson et al., 1969), Q (Arnott et al., 1974), R-S (Urbani et al., 1993), T (Le Questel et al., 1995), U-V (Stortz, 2002), and W (Bosco et al., 2005) are those identified by other studies of κ-carrageenan and its related forms. For many of the minima identified in previous studies, a different set of atoms than that shown in Equations (3.1) were used to define dihedral angles, and in these situations a conversion factor was found to enable comparison.

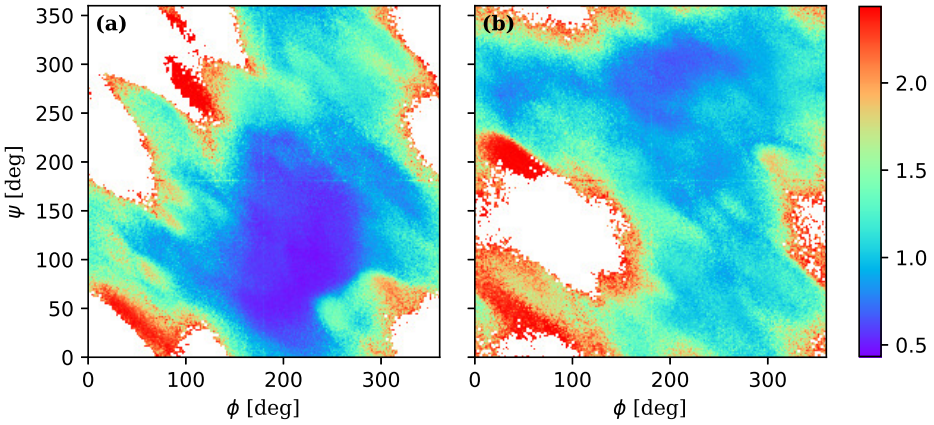

Figure 3.8: Error in energetic values shown in Figure 3.7 for the (a) $1 \rightarrow 3$ and (b) $1 \rightarrow 4$ glycosidic linkages. As with Figure 3.7, the colorbar is in units kJ·mol^{-1}.

Minimum	ϕ,ψ	$\boldsymbol{\Delta F}$ **[kJ·mol^{-1}]**	**Occupancy [%]**
A	210°,137°	0	54.5
B	193°,77°	0.5	43.9
C	287°,123°	8.4	1.6
X	274°,283°	0	54.8
Y	201°,289°	0.6	43.2
Z	252°,195°	8.4	2.0

Table 3.3: Energetic minima for the $1 \rightarrow 3$ (well A-C) and $1 \rightarrow 4$ (well X-Z) linkages. ΔF indicates the free energy difference between the well and the global minimum for the respective linkage. The percentage of trajectory time spent in each energetic minimum is given in the Occupancy column. The average duration between jumps to a new minimum was 159 ps and 167 ps for the $1 \rightarrow 3$ and $1 \rightarrow 4$ linkages, respectively.

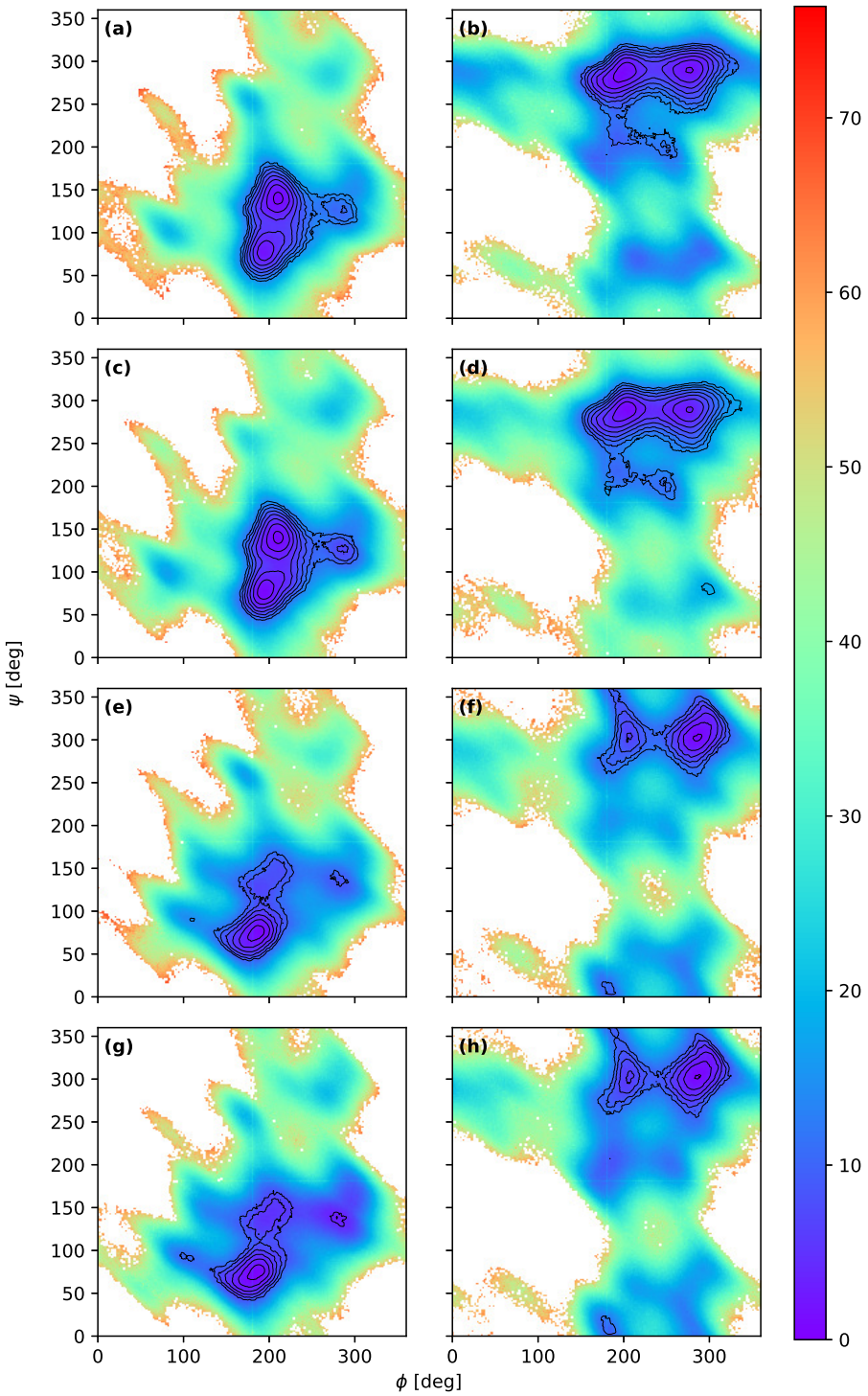

Figure 3.9: Free energy surfaces for glycosidic linkages using GLYCAM parameters and simulating the: (a) $1 \rightarrow 3$ and (b) $1 \rightarrow 4$ linkages both without NaI; and (c) the $1 \rightarrow 3$ and (d) $1 \rightarrow 4$ linkages both with NaI. Using GAFF parameters and simulating the: (e) $1 \rightarrow 3$ and (f) $1 \rightarrow 4$ linkages both without NaI; and (g) the $1 \rightarrow 3$ and (h) $1 \rightarrow 4$ linkages both with NaI. Colorbar units are in kJ·mol^{-1} and white indicates sterically forbidden regions. Contours are proportional to the probability density from unbiased simulations, where successive contours indicate a doubling of the probability density.

3.4.5 Intra-Molecular and Chain-Solvent Hydrogen Bonding

Three common intra-molecular hydrogen bonds (HBs) were identified in all simulations, as well as copious HBing between polar groups and solvent. The existence of a HB was inferred based on the following geometric criteria: a donor-acceptor distance $d_{D-A} < 0.3$ nm, and a hydrogen-donor-acceptor angle $\angle HDA < 20°$. Of the intra-molecular HBs measured during simulations without NaI, the most prevalent existed across the $1 \rightarrow 3$ linkage between the hydroxyl group at AO2 and the sulfate group $G_{+1}SO3$. Also across the $1 \rightarrow 3$ linkage was a HB between the hydroxyl group at GO2 and the ring oxygen O5 in A residues $A_{-1}O5$. The least common intra-molecular HB was between O6 in the hydroxymethyl group and the sulfate group of G residues ($GO6 \cdots GSO3$). The three HBs were not mutually exclusive, with all three HBs occasionally seen in conjunction as shown in Figure 3.1 (a). A summary of the HB occupancies and formation rates can be found in Table 3.4, and the dihedral angles most favourable for HBs spanning the $1 \rightarrow 3$ linkage can be seen in Figure 3.10.

Hydrogen bond	**Occupancy**	**Rate [ns^{-1}]**
$AO2 \cdots G_{+1}SO3$	0.452	14
$GO2 \cdots A_{-1}O5$	0.372	11
$GO6 \cdots GSO3$	0.037	2
$SOL \cdots GSO3$	4.09	28
$GO6 \cdots SOL$	1.27	13
$GO2 \cdots SOL$	0.84	21
$AO2 \cdots SOL$	0.66	22

Table 3.4: Intra-molecular and chain-solvent (SOL) hydrogen bonding during simulations without NaI. Occupancy denotes the time-average number of hydrogen bonds per group and the rate denotes the average number of new hydrogen bonds formed by each group per ns.

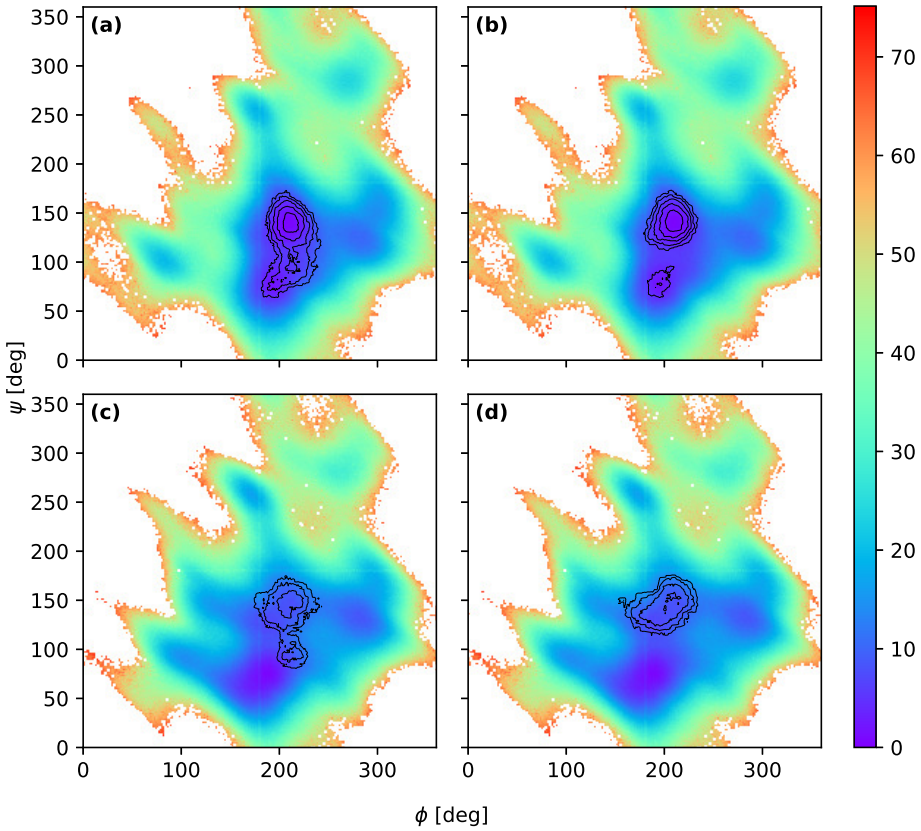

Figure 3.10: Dihedral angles which support hydrogen bond formation across the $1 \rightarrow 3$ glycosidic linkage in simulations without NaI. Overlaid onto the free energy surface for the $1 \rightarrow 3$ linkage simulated using GLYCAM parameters are the hydrogen bond densities for the (a) $\mathrm{AO2} \cdots \mathrm{G}_{+1}\mathrm{SO3}$, and (b) $\mathrm{GO2} \cdots \mathrm{A}_{-1}\mathrm{O5}$ hydrogen bonds. Overlaid onto the free energy surface for the $1 \rightarrow 3$ linkage simulated using GAFF parameters are the hydrogen bond densities for the (c) $\mathrm{AO2} \cdots \mathrm{G}_{+1}\mathrm{SO3}$, and (d) $\mathrm{GO2} \cdots \mathrm{A}_{-1}\mathrm{O5}$ hydrogen bonds. The contour maps are proportional to the number of hydrogen bonds observed across the $1 \rightarrow 3$ linkage when in that combination of ϕ and ψ over the course of simulations, where successive contours denote a doubling of the number of occurrences.

The polar groups consisting of the hydroxyls at O2 within both A and G residues, as well as hydroxymethyl and sulfate groups within G residues, also spent considerable time in HBs with solvent molecules. During simulations without NaI, the hydroxyl groups at AO2 and GO2 regularly formed HBs with solvent, with both groups acting

as the donor in ∼1/3 of the bonds formed. Hydroxymethyl group GO6 and sulfate group GSO3$^-$ also made frequent HBs to solvent, with the former acting as the donor in ∼1/2 of these bonds. Considering both intra-chain and chain-solvent HBing, groups AO2, GO2, and GO6 acted as the donor in 60%, 54%, and 51% of the HB that they participated in, respectively. A summary of the chain-solvent HBing can be found in Table 3.4.

When HB occupancies and formation rates were calculated for simulations with NaI, these were within 1% of those shown in Table 3.4, however when the same values were calculated for simulations using GAFF parameters, intra-molecular HB occupancies were 20-34 % lower. The latter point is likely due to glycosidic linkages during simulations with GAFF parameters preferring a combination of dihedral angles unfavourable for HBs across the linkage (Figure 3.10 (c-d)). The decrease in intra-molecular HBing observed during simulations using GAFF parameters was balanced out by increases in chain-solvent HBing of 14-37%.

3.4.6 Secondary Structure

The restricted flexibility of glycosidic linkages, as evident in Figure 3.7, raises the question of whether a random coil or loose helix is a better descriptor of the disordered-state. Two structural indicators are used to address this: (a) the chain's directional correlation as a function of contour length, which depends solely on the persistence length in the case of a random coil, and; (b) the chain's helical pitch, whose distribution over the course of a trajectory should average to zero in the case of a random coil and be centered on a non-zero value in the case of a helix.

Correlation between vectors tangential to the chain at two points was calculated as a function of the intervening contour length between the points. This is expected to decay exponentially as the distance between the points increases (Li et al., 2021), with the decay constant being the chain's persistence length L_P (Rubinstein and Colby, 2003). That is; for two points on the chain j and $j+i$, the bond-angle autocorrelation function is given by:

$$\left\langle \frac{\mathbf{b}_j \cdot \mathbf{b}_{j+i}}{|\mathbf{b}_j| \cdot |\mathbf{b}_{j+i}|} \right\rangle = e^{-i\sigma/L_p} \tag{3.3}$$

where σ is the average distance between consecutive points j and $j+1$, and the product $i\sigma$ is the contour length L separating points i residues apart. Vectors $\mathbf{b}$ are the virtual bonds defining the chain's local direction, and the averaging $\langle \cdots \rangle$ is taken for all measurements at a particular separation i. The virtual bonds here are taken to be the vectors connecting neighbouring glycosidic oxygen atoms, with the path formed by these approximating the chain's contour. From the plots of bond-angle

correlation versus contour length shown in Figure 3.11 (a), persistence lengths of (6 ± 1) nm were calculated according to (3.3) for simulations both with and without NaI, in good agreement with the value of 6.8 nm measured using solution-state light scattering (Slootmaekers et al., 1987).

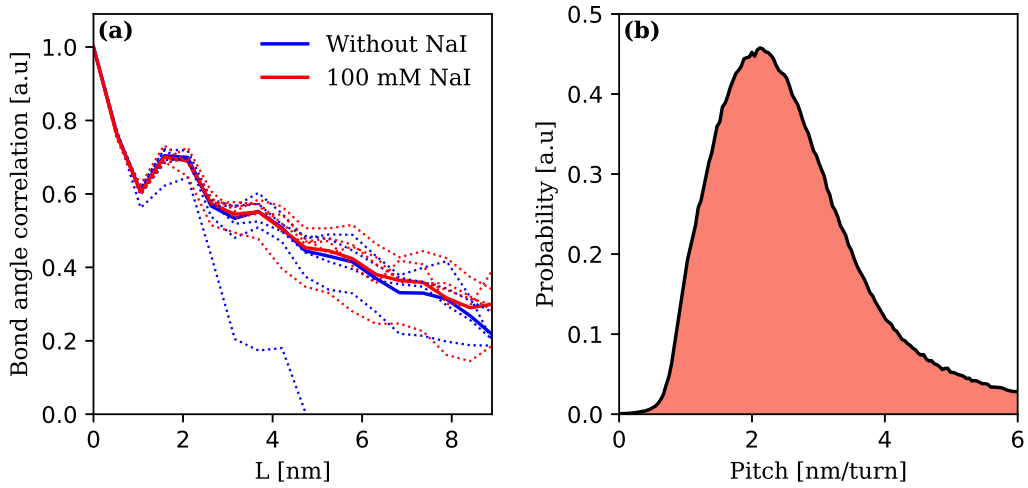

Figure 3.11: Secondary structure indicators for simulated chains. (a) Correlation in simulated chain direction (defined by vectors connecting neighboring glycosidic oxygens), as a function of contour interval L along simulated chains. Oscillations in the directional correlation every ~ 2 nm indicate the chain's direction recorrelates over this length. Dashed lines indicate individual runs whereas solid lines indicate the average over runs. (b) Distribution of helical pitch measurements for systems simulated without NaI, using the locations of glycosidic oxygens as reference. Positive values indicate right-handed helices.

The algorithm developed to calculate helical pitch from simulated trajectories can be found in Appendix A, which also used the locations of glycosidic oxygens to approximate the chain's contour. For simulations without NaI, 16 measurements of helical pitch were made for the conformation every 10 ps, creating the distribution of 800,000 measurements shown in Figure 3.11 (b). The distribution has a peak located at 2.17 nm/turn, with the positive sign indicating right-handed helices. When measurements were repeated for simulations with NaI, the distribution was found to be insensitive to the presence of salt.

3.5 Discussion

The agreement between experimental WAXS and that calculated from the model – as shown in Figure 3.3 and confirmed by the χ^2 value of 0.71 – is excellent over values of q above the peak at q_1. While the calculated scattering is slightly overestimated at higher q, the locations and relative heights of the features at q_{2-4} agree well with the experimental scattering. While not in agreement, the experimental and calculated scaling exponents of -1.386 ± 0.005 and -1.42 ± 0.01 (respectively) between q_1 and $q \approx 8\,\mathrm{nm}^{-1}$ are close, and the use of a narrower range of q values would likely reduce the discrepancy. Both scaling exponents are indicative of an extended conformation (Hammouda, 2010) over length scales from ∼12 down to about ∼1 nm (based on the q range over which the scaling is observed), which is sensible considering this range is similar in magnitude to the model's persistence length of (6 ± 1) nm.

That the model produces WAXS consistent with experiment gives confidence that the GLYCAM force field is capable of accurately modelling the ensemble of solution-state conformations adopted by κ-carrageenan on length scales below ∼10 nm, an example of which is shown in Figure 3.12. At larger length scales (smaller values of q), the experimental and calculated scattering diverge for two reasons; the first simply being the models finite size of ∼10 nm. Even if the model was made larger however, the peak at q_1 that is ubiquitous in polyelectrolyte solutions (Muthukumar, 2017; Mansel et al., 2019) originates from the inter-chain electrostatic potential giving rise to a structure factor, which is a feature the simulation could obviously never reproduce without multiple polymer molecules.

The length scales corresponding to the high-q features at q_2 and q_3 suggest these originate from disaccharide-repeat and monomer-monomer distances, respectively (Table 3.2). The claim that these features are due to linear repeat distances is supported by the observation that as calculations are made from longer chain fragments (Figure 3.4 (c)), these features become more defined. The length scale of 1.10 nm associated with q_2 falls among MD measurements of A-A_{+2} separation (0.98 or 1.03 nm) and G-G_{+2} separation (1.12 nm), whereas the length scale of 0.63 nm associated with q_3 falls among MD measurements of A-G_{+1} separation (0.52 nm) and G-A_{+1} separation (0.66 nm). By recalculating the scattering from conformers for which sulfate groups were removed post-simulation, no peak at q_4 was observed (Figure 3.4 (d-e)), and indeed the associated length scale of 0.43 nm is close to the 0.41 nm separation between the electron rich regions of the sulfate and its host G residue. The width of all three features q_{2-4} is related to the fluctuation of the associated intra-molecular distances (Table 3.2) with larger fluctuations causing wider, less defined features, further supporting this assignment.

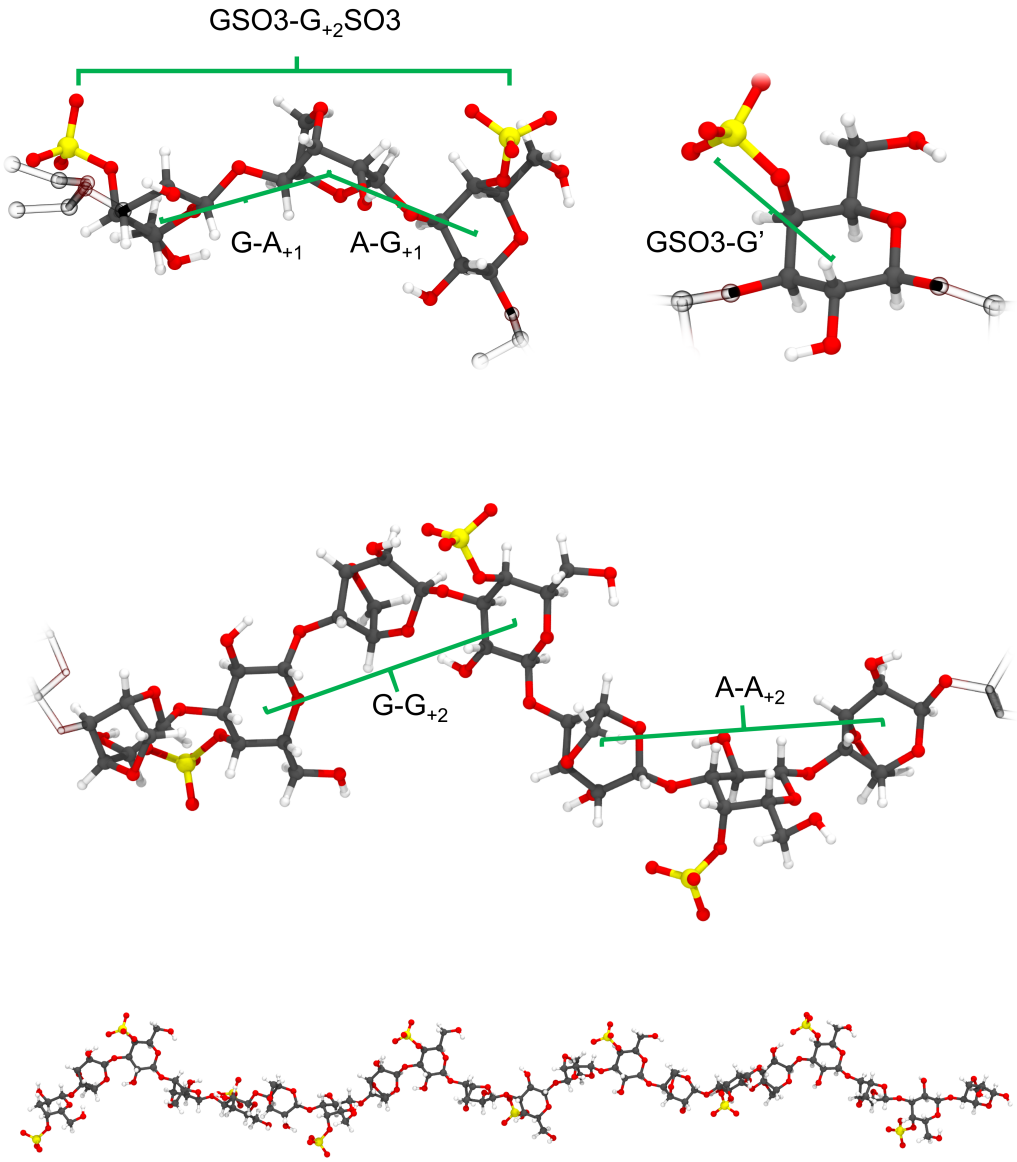

Figure 3.12: One conformation adopted by κ-carrageenan during simulations without NaI, with important intra-molecular distances indicated. Counter-ions and solvent have not been rendered in the image for clarity.

As well as providing a powerful means of validating models against experimental data, comparing WAXS profiles computed from different models allows for a convenient way to check for structural differences, which can be difficult to spot visually over the course of long and fluctuating trajectories. However, as Figure 3.4 (b) shows, for a given

set of parameters there were no significant differences in the scattering profiles – and hence the molecular structure – of simulations with or without NaI. Given that at 20 °C, κ-carrageenan exists in the disordered-state without NaI and in the ordered-state at a NaI concentration of 100 mM (Ciancia et al., 1997), one can expect there to be large structural differences between the two salt concentrations. Hence there are only two conclusions that can be drawn, either: the simulation cannot replicate the ordered-state due to a limitation in its design; or it is incapable of replicating the ordered-state due to inaccuracies in the parameter set used. In the next chapter it will be shown that it is a case of the former rather than the latter.

The large-scale molecular structure of κ-carrageenan is determined by the flexibility of the $1 \rightarrow 3$ and $1 \rightarrow 4$ glycosidic linkages. As the FESs in Figure 3.7 show, the preferred linkage orientations agree with the majority of past studies on the topic. This is surprising given the various carrageenan types used and approaches employed, from fibre XRD of ι-carrageenan double-helices (points D and P), to MM modelling of κ-carrageenan oligomers (points F,G,J,K,N and R-W) and ι-carrageenan oligomeric double-helices (points E and Q), to MD simulations of κ- (points L and M) and β-carrageenan (point H) oligomers using the CHARMM force field. The only outliers occur for MM modelling (points N,R and T) and MD simulations (point M) of κ-carrageenan oligomers. Given that the current results as well as those in previous studies indicate that the glycosidic linkages are confined to a relatively small window of orientations, the term 'random coil' – that is, a conformation bereft of secondary structure – cannot be used to describe κ-carrageenan in the disordered-state. In fact, the glycosidic linkages are restricted so much so that a persistent helical pitch of $\sim$2.17 nm/turn is measured.

Considering the inherent secondary structure possessed by the κ-carrageenan disordered-state, and its insensitivity to changes upon NaI concentration, the plausibility of a salt-induced uni-molecular disorder-order transition is called into question. Although a uni-molecular transition has often been assumed in prior studies, there is very little in the way of a proposed mechanism. A change to the intra-molecular HBing has been suggested by Schefer et al. (2015b) as one mechanism (presumably due to a change in the dielectric properties of the solvent), and intuitively the binding of counter-ions to the chain would seem like another. Although the interpretation of Figure 3.10 is that the $1 \rightarrow 3$-linkage – and hence the conformation – is stabilized to some degree by the HBs $\mathrm{AO2} \cdots \mathrm{G}_{+1}\mathrm{SO3}$ and $\mathrm{GO2} \cdots \mathrm{A}_{-1}\mathrm{O5}$, NaI was found to have negligible effect on these HBs' occupancies or rates as shown in Table 3.4. Furthermore, although the spatial distributions of Na^+ and I^- ions show the effects of electrostatic interaction with the negatively charged chain (Figure 3.6), counter-ion condensation is energetically disfavoured (and indeed absent) due to the relatively low linear charge

density of a single κ-carrageenan chain (Stroble, 2007). The lack of an obvious impetus for a uni-molecular κ-carrageenan disorder-order transition in simulated systems, and the indifference of all measured structural indicators to the concentration of NaI, strongly suggests that salt-induced transitions observed in κ-carrageenan systems are not uni-molecular events. In the following chapter, the hypothesis that this transition is in fact a cooperative process will be tested.

Chapter 4

κ-Carrageenan in the Ordered-State

4.1 Introduction

In the previous chapter MD simulations were used to elucidate the disordered-state structure of κ-carrageenan, which was validated via the comparison between the experimental WAXS and that calculated from the model. The disordered-state was in fact found to contain substantial secondary structure in the form of restricted glycosidic linkage flexibility leading to a 'loose' helical structure with a helical pitch of $\sim$2.17 nm/turn. These findings are in agreement with the majority of previous modelling (Urbani et al., 1993; Le Questel et al., 1995; Stortz and Cerezo, 2000; Ueda et al., 2001; Stortz, 2002) and the NMR experiments of Bosco et al. (2005), however challenge the pervasive use of the term 'random coil' to describe the state. Furthermore, no evidence was found to support the idea of a uni-molecular disorder-order transition upon the addition of NaI, incongruous with the interpretation of recent AFM images (Schefer et al., 2015b). In this chapter the combination of MD and WAXS is applied to understand the structure of κ-carrageenan in the ordered-state by: (1) generating an ensemble of multi-chain structures by running many MD simulations in parallel, (2) selecting the subset of the structures whose calculated WAXS best fits experimental data, and (3) further simulating the subset to understand the long time scale dynamics.

4.2 Experimental X-ray Scattering

Dried κ-carrageenan (as described in Section 3.2.1) was slowly dissolved into deionised water (18.2 MΩ cm) to yield a 2% (wt/wt) solution, followed by a number of hours of mixing to ensure full dissolution of the polymer. A 200 mM NaI solution was then prepared and added to the polymer solution in equal parts while mixing at 75 °C,

yielding a sample with concentrations 1% (wt/wt) polymer and 100 mM NaI. This solution was then hot-filtered (pore size 0.2 μm) directly into cells used for the WAXS instrument sample auto-loader. Preparation of solutions for background scattering followed a similar course, with the NaI solution diluted to 100 mM NaI using deionised water. After samples had sufficiently equilibrated at 20 °C, WAXS data was collected in a similar manner as described in Section 3.2.3, with the only difference being that a well plate was used to contain the sample as opposed to a flow cell.

4.3 Molecular Dynamics Simulations

4.3.1 Initial Configurations

The preparation of MD simulations followed the same procedure for single κ-carrageenan chains (Chapter 3), up to the point of creating structure and topology files for large chains. Here, the AmberTools program LEaP was used to create these files for a single chain of a DP of 8 (16 sugar residues), using the GLYCAM-06j force field. Two such chains were then placed in a dodecahedron box with PBCs and dimensions larger than twice the length of a single chain. The chains were placed relative to one another in two ways; either as separate chains (SC, Figure 4.1 (a)), or a pre-prepared double-helix (DH, Figure 4.1 (b)). DH configurations were prepared from two parallel chains by first shifting one along the chain's long axis by 0.45 nm, then rotating the chain about this axis by 120° to produce a double-helix. SC configurations were prepared by separating the two chains within the double-helix by 1.6 nm along a vector perpendicular to the long axis, then rotating one of the chains about this vector by 45°. Both configurations were then solvated by adding 33890 SPC/E molecules to the simulation boxes, which were then randomly replaced with either: 16 Na^+ ions to neutralise the polymer charge, or 79 Na^+ and 63 I^- ions to give a NaI concentration of 100 mM. This resulted in four initial configurations: SC configurations with and without NaI, and DH configurations with and without NaI.

Configuration	**NaI**	**100 ns**	**1μs**
DH	None	12	2
DH	0.1 M	12	2
SC	None	12	2
SC	0.1 M	12	2

Table 4.1: Summary of two-chain simulations.

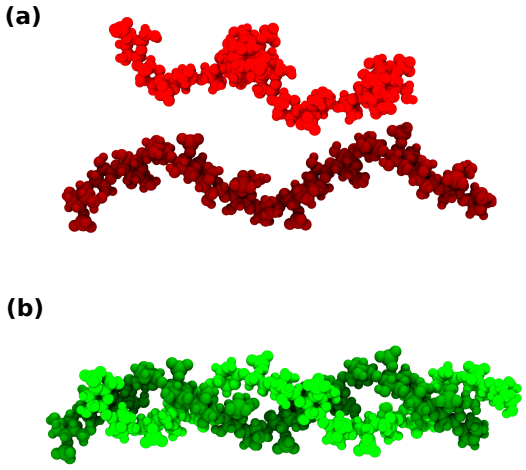

Figure 4.1: Initial configurations for MD simulations of (a) separate chains (SC) and (b) double-helices (DH).

4.3.2 Equilibration and Production Runs

Configurations were energetically minimized using the GROMACS 2020.6 steepest descent algorithm such that the maximum force on any atom was less than 1000 $kJ \cdot mol^{-1} \cdot nm^{-1}$. Twelve replicas of each of the four initial configurations were then created and assigned unique initial velocities at 20 °C, which then underwent isochoric and isobaric equilibration in the same manner as described in Section 3.3.2. Production simulations 100 ns in duration were then made for the 48 systems, after which a subset of systems was selected for further simulation to 1 μs (for details on the chosen subset, see the results section) as summarised in Table 4.1.

4.4 Results

4.4.1 Differences Between Simulations With and Without NaI

As with the simulation of single κ-carrageenan chains in Chapter 3, there were no detectable uni-molecular differences between simulated SC configurations with and without NaI, however SC configurations with NaI had higher rates of dimerization versus those without NaI. Of the twelve SC configurations simulated for each salt condition, only five without NaI dimerized, versus eleven with NaI. To understand the nature of the dimerization, every residue in the final conformation of each simulation was classified as either isolated, non-specifically aggregated, or double-helical, based on its minimum distance from – and orientation relative to – the other chain. Residues were

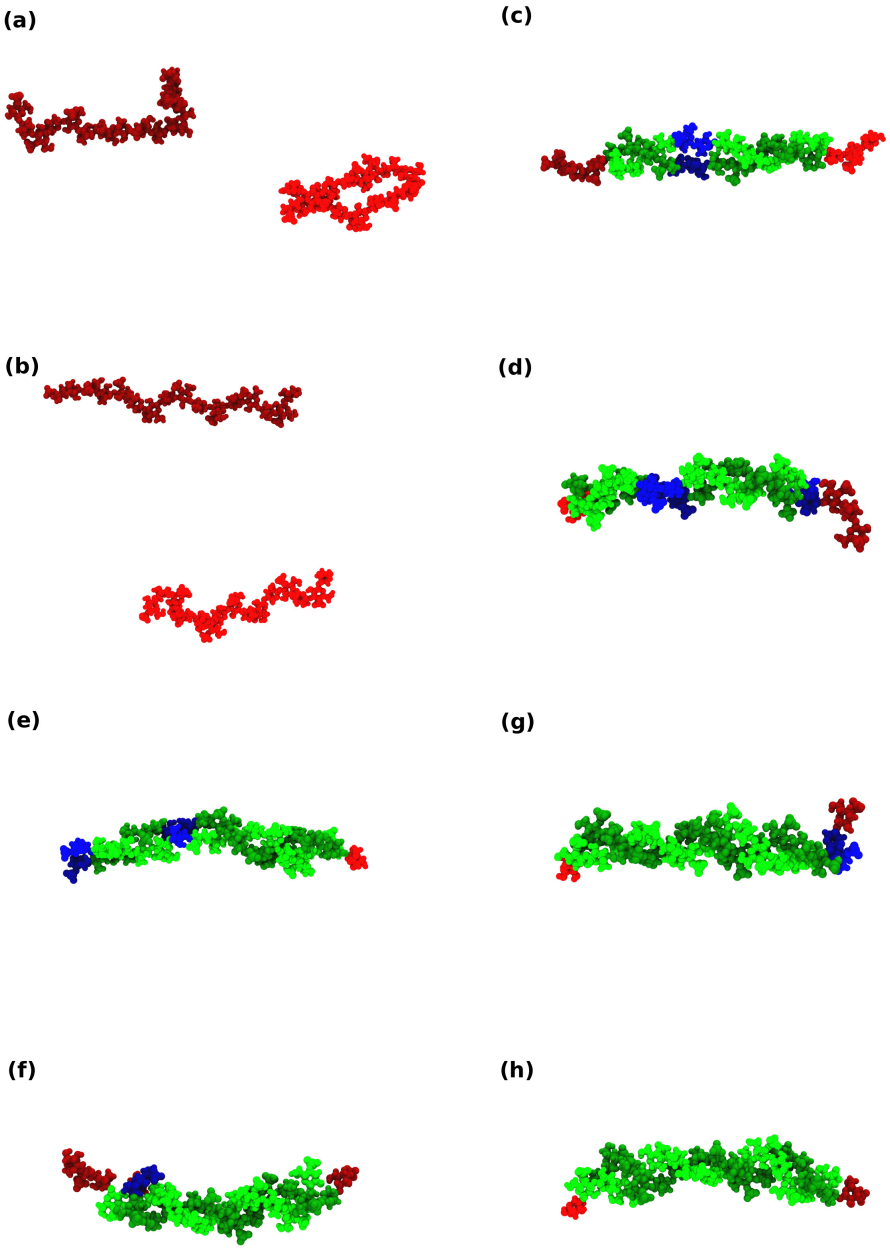

Figure 4.2: Simulated conformations after 1 μs when chains were initiated as SC without NaI (a-b) and with 0.1 NaI (c-d), and when chains were initiated as DH without NaI (e-f) and with 0.1 NaI (g-h). Residues are categorised as isolated (red), non-specifically aggregated (blue), or double-helical (green). For clarity, one chain is rendered lighter than the other.

considered double-helical when their COM was less than 0.5 nm from a residue on the other chain and the two chains were clearly wrapped around a common axis. Non-specifically aggregated residues were those whose inter-chain distance was less than 0.5 nm but did not meet the wrapping condition of double-helical residues, and isolated residues were those whose inter-chain distance was more than 0.5 nm. Conformations displaying all three categories can be seen in Figure 4.2 (c-h).

For each of the four initial configurations, the percentage of residues in each of the three categories after 100 ns simulation is shown in Figure 4.3. During simulations initiated as SC without NaI, 10.7% of residues ended up in double-helical sections of chain whereas 3.9% ended up in non-specifically aggregated sections. This is compared to simulations initiated as SC with NaI, for which 24.7% and 15.4% ended up in double-helical and non-specifically aggregated sections, respectively. The majority of simulations initiated as DH both with and without NaI remained mostly double-helical after 100 ns, with little difference between the two salt conditions.

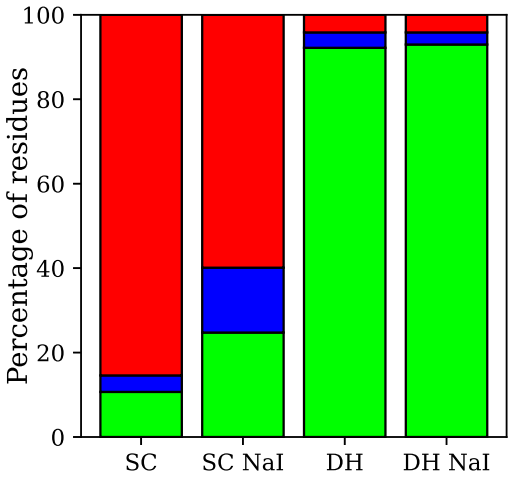

Figure 4.3: The percentage of residues in double-helical (green), non-specifically aggregated (blue), and isolated (red) sections of chain after twelve 100 ns simulations of initially separated chains without and with NaI ('SC' and 'SC NaI' respectively), and of initially double-helical chains without and with NaI ('DH' and 'DH NaI' respectively).

4.4.2 Experimental and Calculated Wide-Angle X-ray Scattering

Experimental WAXS for a 1% (wt/wt) κ-carrageenan NaI solution in the ordered-state is shown in Figure 4.4, along with the WAXS for a 1% (wt/wt) κ-carrageenan solution in the disordered-state as presented in Chapter 3. The significant difference between the two profiles indicates the disordered- and ordered-states are structurally dissimilar.

Apart from an elbow at $q \approx 5\,\mathrm{nm}^{-1}$ and a small peak at $q \approx 8\,\mathrm{nm}^{-1}$, the ordered-state WAXS is relatively featureless, and thus the simulated model will be used to interpret the results.

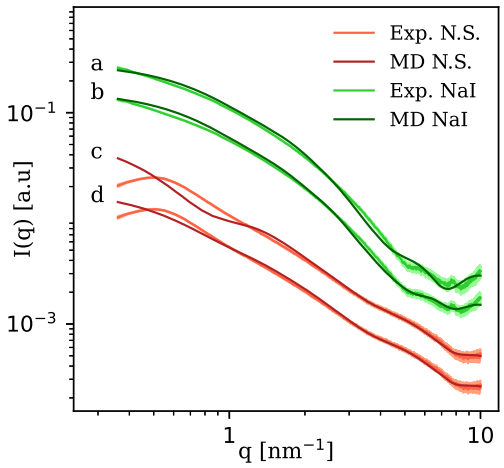

Figure 4.4: Experimental WAXS for 0.1 M NaI solutions of κ-carrageenan in the ordered-state compared to that calculated from simulations with NaI initiated as DH (a), and initiated as SC which subsequently dimerized (b). Reproduced from Figure 3.3 is experimental WAXS for native solutions of κ-carrageenan in the disordered-state compared to that calculated from simulations without NaI initiated as SC (c), and as single chains (d) (also reproduced from the cited publication). Calculated curves (a), (b), and (c) are the ensemble-averaged scattering profiles for the simulations resulting in the structures shown in Figure 4.2 (g-h), (c-d), and (a-b), respectively. The shaded areas represent the 95% confidence interval for experimental scattering.

Using the atomic coordinates of 1000 conformers visited by each of the 48 short simulations, WAXS profiles were calculated using CRYSOL 3, with explicit hydrogens and 50 spherical harmonics used in each calculation. The 1000 WAXS profiles were then used to produce one ensemble-averaged profile for each simulation, which was used to test the simulation's validity against whichever of the experimental profiles corresponded to the simulated salt condition. A simulation's validity was based on its minimum value of Equation 3.2, where the same descriptions of fitting parameters apply here.

To test whether WAXS profiles calculated from the atomic coordinates of larger multimers might fit the experimental scattering better, the same procedure described above was applied to simulations of three- and four-chain multimers. As Figure 4.5

shows, the ensemble-averaged WAXS profiles were sensitive to multimer size. In particular, the scaling of $I(q)$ becomes steeper in the region of $q \sim 2$ nm^{-1} and a peak at $q \sim 6.5$ nm^{-1} grows with increasing size. Differences between profiles for two non-specifically aggregated chains versus mostly double-helical chains are also apparent, with an elbow at $q \sim 5.5$ nm^{-1} being more pronounced for double-helices.

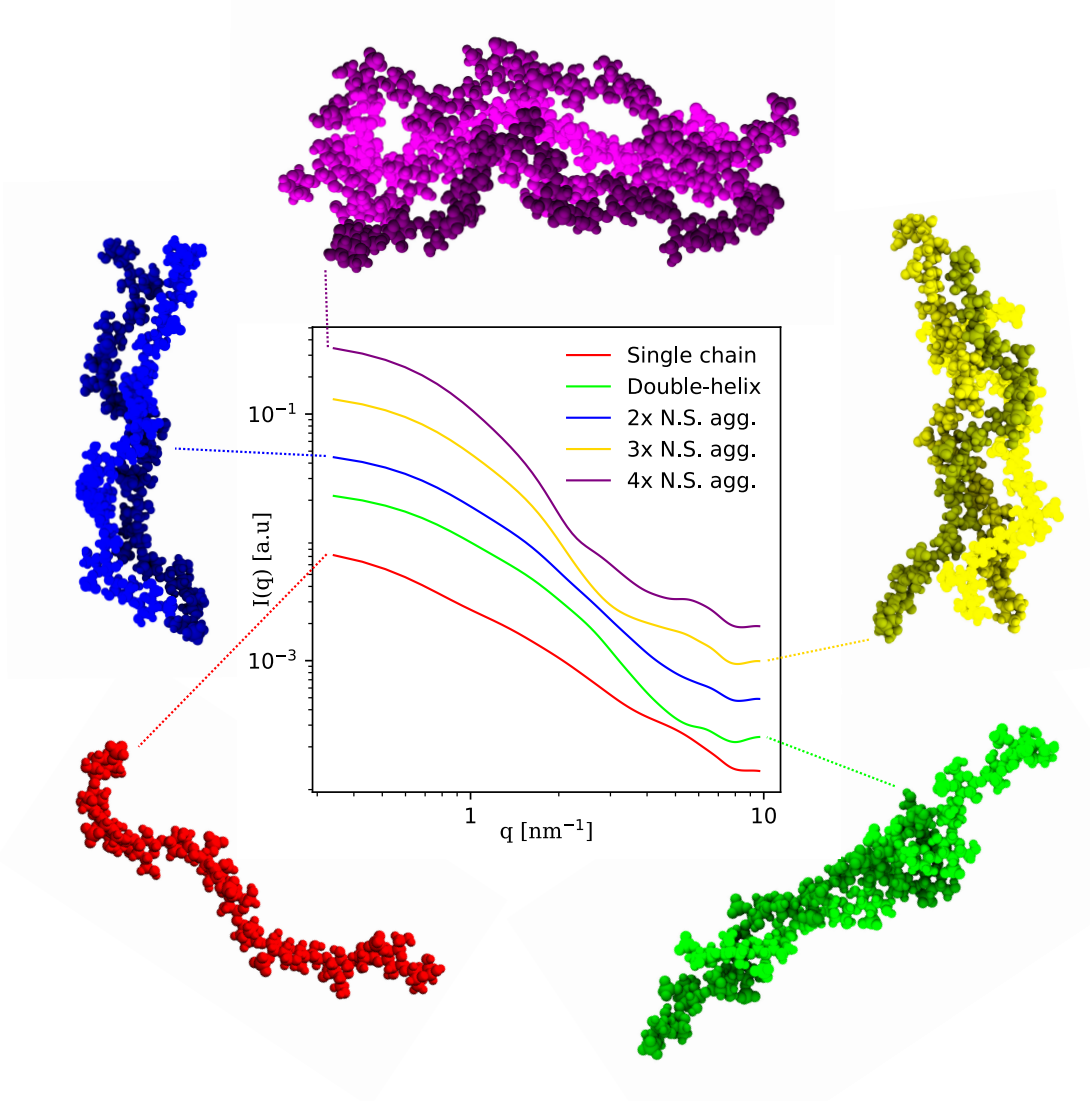

Figure 4.5: Wide-angle X-ray scattering profiles calculated for simulated single chains and spontaneously formed double-helices, as well as two, three, and four non-specifically aggregated parallel chains. Profiles are normalised by each structure's mass, then scaled by multiples of 2 to aid visualisation.

The two-chain simulations initiated as SC without NaI whose calculated WAXS profiles best fitted the experimental data for disordered-state solutions were chosen to further simulate to 1 μs. These were both simulations where the chains had remained

separated after 100 ns; the conformations after further simulation to 1 μs, which can be seen in Figure 4.2 (a-b), show that these remained separated (although one chain looped back on itself in the final stages of the simulation). Likewise, the two simulations initiated as SC with NaI whose calculated WAXS profiles best fitted the experimental data for ordered-state solutions were chosen to further simulate to 1 μs. These were both simulations where the chains had dimerized after 100 ns, and as Figure 4.2 (c-d) shows, these had become mostly double-helical after 1 μs. To test the long time scale stability of configurations initiated as DH, two of each salt condition were chosen to further simulate, and as Figure 4.2 (e-h) shows these remained mostly double-helical after 1 μs.

From the final stages of these longer simulations, the atomic coordinates of 10,000 conformers were used to calculate an ensemble-averaged WAXS profile for each, according to the process described above for short simulations. As Figure 4.4 shows, the WAXS calculated from dimers spontaneously formed during SC simulations with NaI (profile (b), $\chi^2 = 1.00$, conformers (c-d) in Figure 4.2) fitted the experimental profile for ordered-state solutions better than the profile calculated from simulations initiated as DH with NaI (profile (a), $\chi^2 = 2.25$, conformers (g-h) in Figure 4.2). The disordered-state experimental WAXS data is fitted better by the scattering calculated from single chains (profile (d), χ^2=0.71, Figure 3.12) than that calculated from simulations initiated as SC without NaI (profile (c), χ^2=1.67, conformers (a-b) in Figure 4.2). Although both single-chain and SC simulations without NaI reproduced the experimental scattering for the disordered-state solution well at q above $\sim 2\,\mathrm{nm}^{-1}$, the introduction of a second chain resulted in a structure factor appearing in the calculated scattering at $q \approx 1.5\ \mathrm{nm}^{-1}$. The scattering calculated for simulations initiated as DH without NaI was found to be indistinguishable from profile (a), and hence not shown in Figure 4.4.

4.4.3 Double-Helical and Non-Specifically Aggregated Dimers

Having used the batch of 48 short simulations to show that NaI causes more chains to dimerize, the following detailed structural analysis will focus on the 1 μs simulations, which were shown to best reproduce the experimental WAXS data. Particular emphasis will be put on the structural differences between isolated, non-specifically aggregated, and double-helical sections of chain. The reason for this is that although there were no *uni-molecular* structural differences between isolated sections of chain in simulations with or without NaI, the structure within non-specifically aggregated and double-helical sections – whose frequency *was* was dependant on NaI – did differ from that of isolated sections.

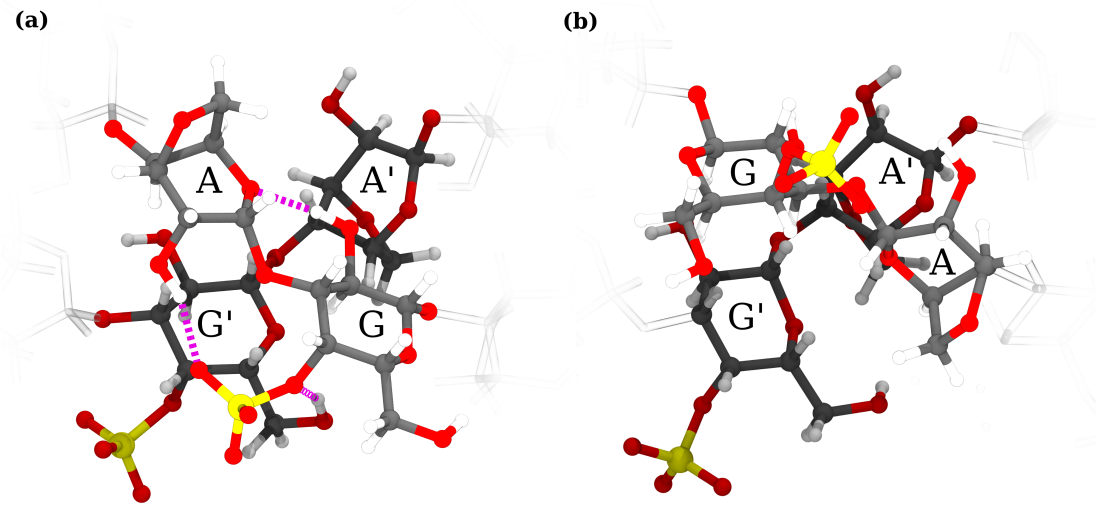

Figure 4.6: Double-helical sections of κ-carrageenan with parallel chains (a) and anti-parallel chains (b), formed spontaneously during simulations initiated as SC.

To begin the structural analysis, it is important to clarify the nature of the dimerized chains beyond the simple definition given in Section 4.4.1. Both non-specifically aggregated and double-helical sections of chain were twice as likely to have chains aligned parallel to each other (i.e. reducing and non-reducing ends of each chain at the same ends of the dimer) versus anti-parallel, and hence the analysis focuses more on the former. As well as not meeting the wrapping condition of double-helical sections, non-specifically aggregated sections (as the name implies) did not contact the other chain in a specific manner, as opposed to double-helical sections which made very specific inter-molecular contacts. The specific inter-molecular contacts made within double-helical sections were periodic, with the fundamental unit – consisting of one $1 \rightarrow 3$-linked and one $1 \rightarrow 4$-linked diad (Figure 4.6) – being repeated along a double-helical section. To understand the ensemble-averaged inter-molecular contact points, contact between a pair of atoms was determined based on the value of a sigmoidal function

$$s(r_{i,j}) = \frac{1}{1 + \exp(\beta(r_{i,j} - \lambda_{i,j})} \tag{4.1}$$

Here $r_{i,j}$ is the distance between a pair of atoms, $\lambda_{i,j}$ is the sum of their van-der Walls radii, and β is a constant set arbitrarily to unity. Equation 4.1 evaluates to < 0.5 for a pair of atoms whose separation is greater than the sum of their van-der Walls radii, and > 0.5 for a pair whose separation is less than this sum.

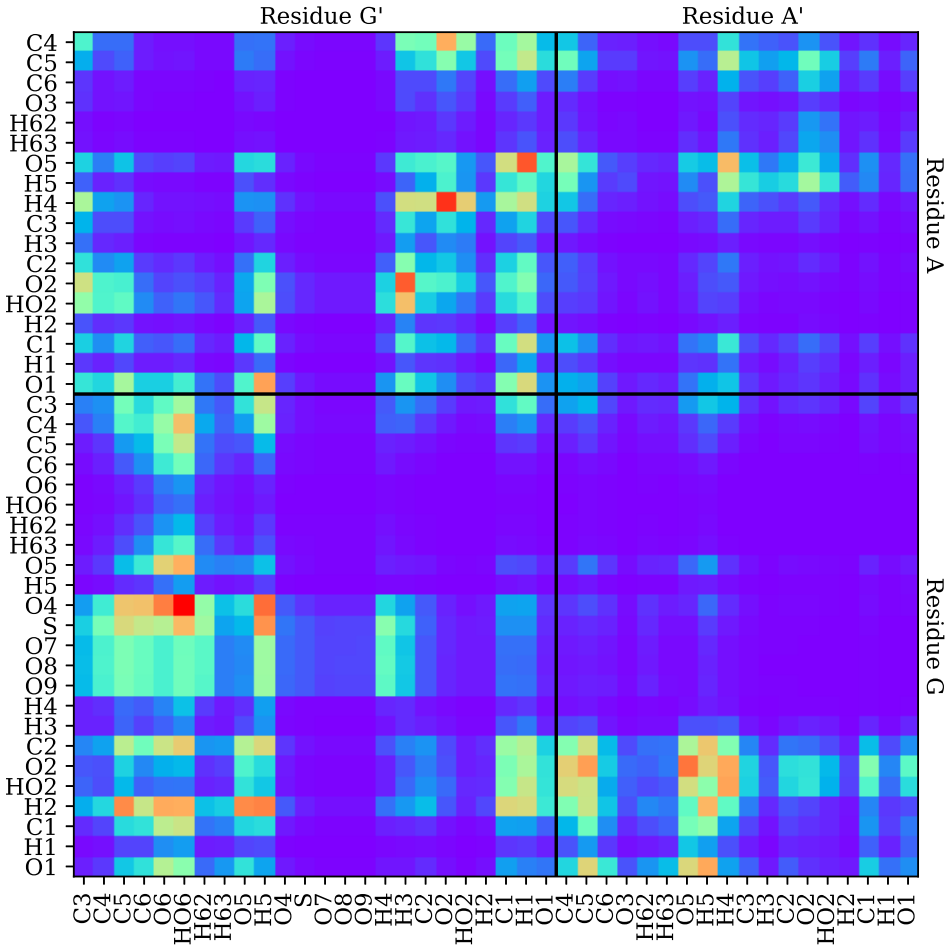

Figure 4.7: Inter-molecular contacts between atoms in double-helical sections (Figure 4.6 (a)) formed during simulations initiated as SC with NaI. The color map ranges from violet to red and indicates low to high values of Equation 4.1, i.e. no inter-molecular contact to persistent contact.

Figure 4.7 shows a map of inter-molecular contacts – evaluated for the double-helical unit shown in Figure 4.6 (a) using Equation (4.1) – averaged over 240,000 conformers drawn from five double-helical sections formed during 1 μs simulations initiated as SC with NaI. Between A and G′ residues, most contact involved axial hydrogen atoms from one residue contacting oxygen atoms on the other, whereas between G and A′ residues, contact primarily involved the G residue's O2 hydroxyl contacting the A′ residue in the vicinity of C4, C5, and O5. There was little contact between A and A′ residues, however between the G′ and G residues the former's hydroxymethyl group made regular

contact with the latter's sulfate (a HB). When inter-molecular contact was measured for double-helical sections during simulations initiated as DH (of either salt condition), these were nearly identical to that for double-helical sections formed spontaneously during simulations initiated as SC with NaI (Figure 4.8).

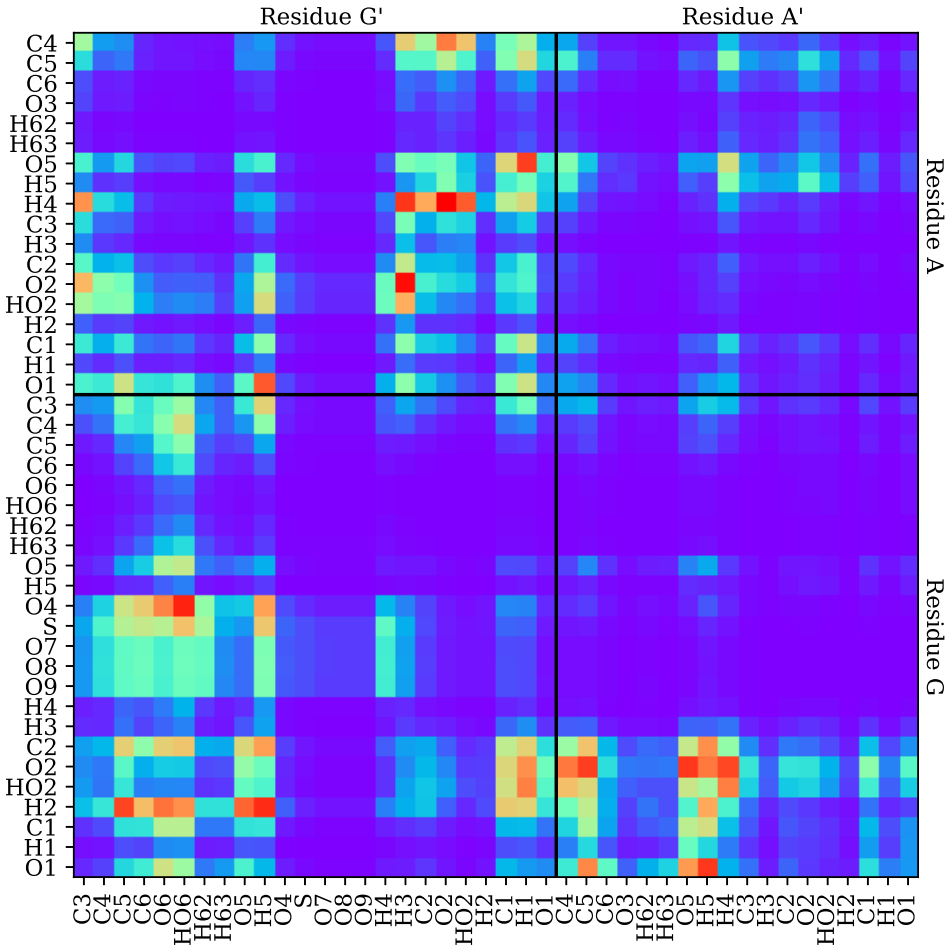

Figure 4.8: Inter-molecular contacts between atoms in double-helical sections (Figure 4.6 (a)) of chain during simulations initiated as DH. The color map ranges from violet to red and indicates low to high values of Equation 4.1, i.e. no inter-molecular contact to persistent contact.

To understand the evolution of κ-carrageenan dimers once formed, residues were categorized as isolated, non-specifically aggregated, or double-helical as function of time using the method described in Section 4.4.1 for one 1 µs SC simulation with NaI.

As Figure 4.9 shows, the two chains initially form a non-specifically aggregated dimer within 10 ns, although the speed at which this happens will be highly dependant on the initial distance and orientation between the chains. During the following 50 ns, parts of the dimerized section fluctuate between being non-specifically aggregated and double-helical, before finally locking into a predominantly double-helical arrangement. Although the general trend is for isolated sections to non-specifically aggregate and then rearrange to become double-helical, there are many examples either where isolated sections directly associate as a double-helix, or where double-helices revert to either non-specifically aggregated or isolated sections. Furthermore, despite the double-helix

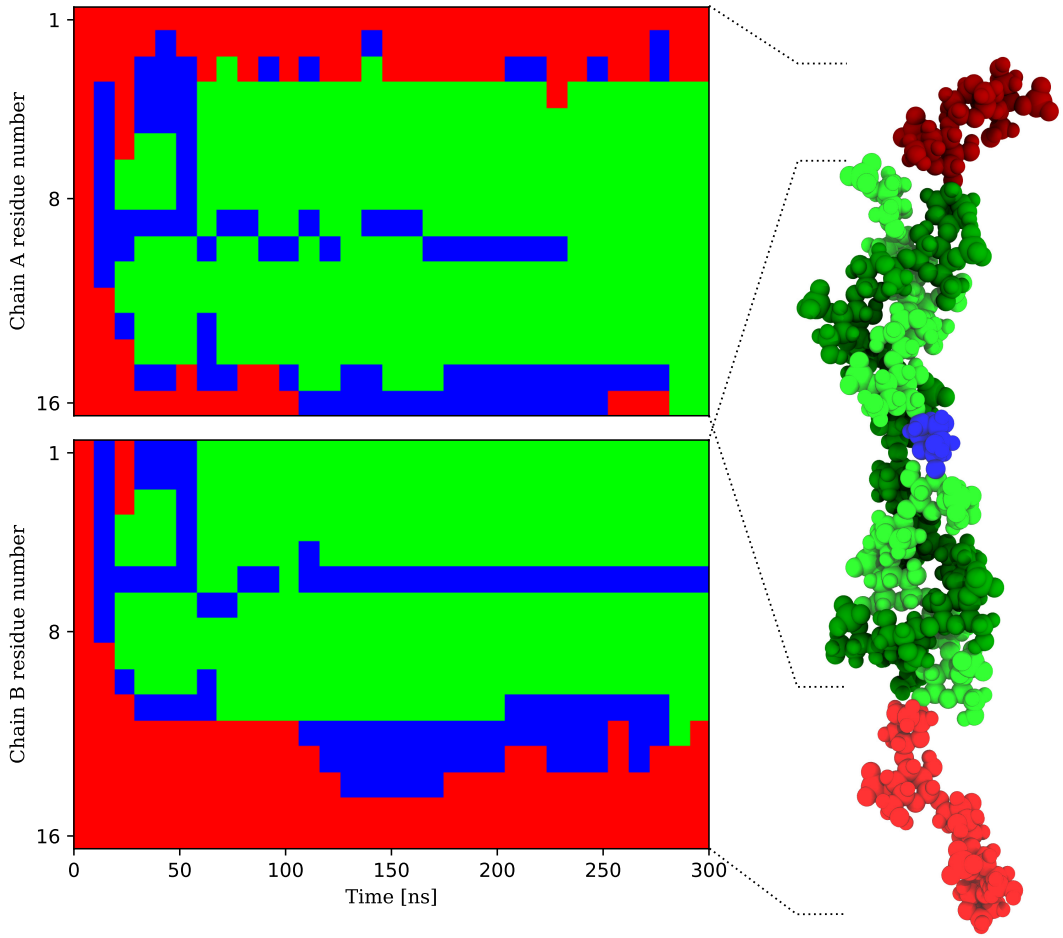

Figure 4.9: Time evolution of a dimer formed during one 1 μs simulation initiated as SC with NaI. Each residue is classified as residing in either an isolated (red), non-specifically aggregated (blue), or double-helical (green) section of chain.

appearing to be the most stable and therefore energetically favourable arrangement, residues may become kinetically trapped in non-specifically aggregated arrangements if surrounded by double-helical sections, as seen in the center of the molecular structure shown in Figure 4.9.

4.4.4 Secondary Structure

The dihedral angles adopted by glycosidic linkages – and hence the secondary structures of individual chains – were found to depend on which of the three categories residues belonged to. For each of the two glycosidic linkages, atomic coordinates from ∼217,000 linkages in isolated sections, ∼408,000 linkages in non-specifically aggregated sections, ∼616,000 linkages in double-helical sections from SC simulations, and ∼1.5M linkages in double-helical sections from DH simulations were used – along with the definition of dihedral angles defined in (3.1) – to generate the dihedral maps shown in Figure 4.10.

Rather than compute the FES for each of the glycosidic linkages using metadynamics (as was done in Chapter 3), here energetic minima are inferred based on the locations of high dihedral angle probability density, as shown in Figure 4.10. The minima for isolated and non-specifically aggregated sections' $1 \rightarrow 3$ linkages were found to be similar, with the ψ angle oscillating between two values (Figure 4.10 (a)), whereas the minima at higher ψ became deeper for $1 \rightarrow 3$ linkages belonging to double-helical sections (either those initiated as DH or those that spontaneously formed from SC, Figure 4.10 (c)). The ϕ angle for $1 \rightarrow 4$ linkages belonging to isolated sections oscillated between two minima of approximately equal depth, whereas the minimum at higher ϕ became deeper in $1 \rightarrow 4$ linkages belonging to non-specifically aggregated sections (Figure 4.10 (b)). For $1 \rightarrow 4$ linkages belonging to double-helical sections (either those initiated as DH or those that spontaneously formed from SC) this was reversed, with the minimum at lower ϕ becoming deeper and the barrier between the two minima becoming higher (Figure 4.10 (d)). Every linkage made regular exchanges between the energetic minima available to it, with the most stable linkage – the $1 \rightarrow 4$ linkage in double-helical sections – jumping between minima Y and Z on average every 0.5 ns. The locations of each of the minima identified in this study are summarised in Table 4.2.

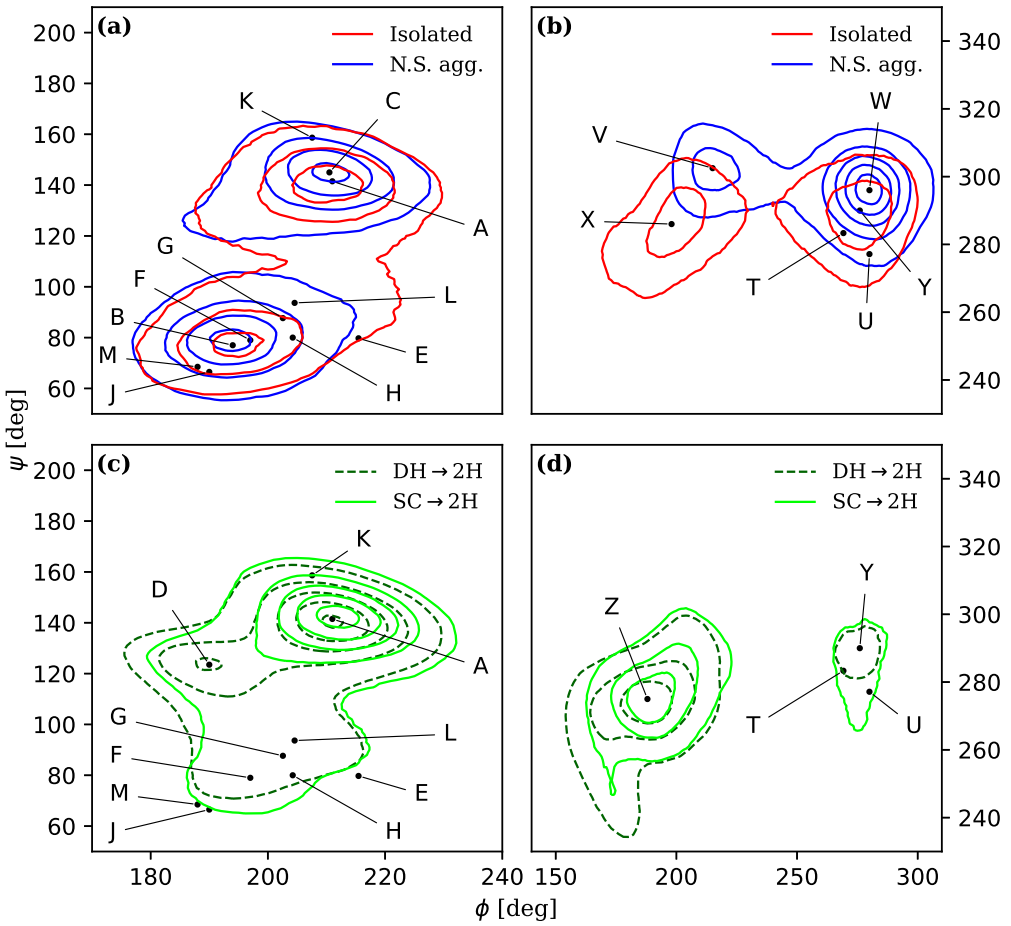

Figure 4.10: Dihedral angles measured in sections of isolated and non-specifically aggregated residues when chains were initiated as SC with NaI for $1 \rightarrow 3$ (a) and $1 \rightarrow 4$ (b) linkages, and those measured in double-helical sections (2H) when chains were either initiated as SC or DH with NaI for $1 \rightarrow 3$ (c) and $1 \rightarrow 4$ (d) linkages. Contours correspond to probability density, and points A-D and V-Z are energetic minima identified in this study, as summarised in Table 4.2. Also included are the following minima identified in past studies: E (Anderson et al., 1969), F (Arnott et al., 1974), G (Urbani et al., 1993), H (Le Questel et al., 1995), J (Ueda et al., 1998), K-L (Stortz and Cerezo, 2000), M (Ueda et al., 2001), T (Urbani et al., 1993), and U (Stortz, 2002).

Min.	ϕ	ψ
A	211°	141°
B	193°	77°
C	211°	145°
D	190°	124°
V	215°	303°
W	280°	296°
X	198°	285°
Y	276°	290°
Z	188°	275°

Table 4.2: Locations of the energetic minima for the $1 \rightarrow 3$ (minima A-D) and $1 \rightarrow 4$ (minima V-Z) linkages, as shown in Figure 4.10.

To determine whether isolated chains (which were found to be 'loose helices' in Chapter 3) became more or less helical after dimerizing as non-specifically aggregated or double-helical sections, helical pitch calculations were made using the atomic coordinates of the conformers used to generate Figure 4.10 and the algorithm described

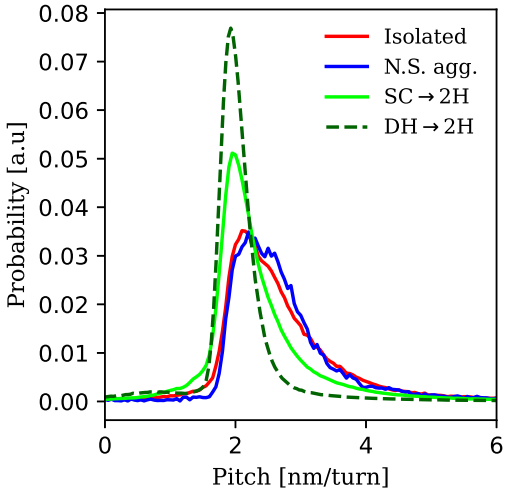

Figure 4.11: Distributions of helical pitch measured during SC simulations with NaI, for sections of isolated, non-specifically aggregated, and double-helical (SC→2H) sections of chain, and during DH simulations with NaI for double-helical sections of chain (DH→2H).

in Appendix A. As Figure 4.11 shows, the distribution of helical pitch values for non-specifically aggregated sections closely resembles that for isolated sections, with a similar width and a peak located at 2.27 nm/turn. Within double-helical sections this distribution became narrower and shifted to tighter helices, with a peak located at 1.95 nm/turn for sections formed spontaneously during simulations initiated as SC, and at 1.92 nm/turn for sections initiated as DH.

4.4.5 Intra- and Inter-Molecular Hydrogen Bonding

Inter-molecular HBs in non-specifically aggregated and double-helical sections, as well as intra-molecular HBs in all three categories, were identified based on the same geometric criteria used in Section 3.4.5, using the same conformers as in Section 4.4.4. The most common inter-molecular HB, as well as the three most common intra-molecular HBs, are detailed in Table 4.3. High HB occupancies within double-helices were due to the particular orientation of G residues in those sections (Figure 4.6 (a)), which simultaneously allowed the G′ hydroxymethyl access to the G residue sulfate, and promoted the two intra-molecular HBs spanning the glycosidic linkage.

Hydrogen bond	**Double-helix**		**N.S. aggregate**		**Isolated**	
	Occ.	Rate	Occ.	Rate	Occ.	Rate
G′O6···GSO3	0.19	8	0.03	2	N/A	N/A
AO2···GSO3	0.54	17	0.40	13	0.44	16
GO2···AO5	0.54	13	0.38	9	0.35	12
GO6···GSO3	0.04	2	0.05	2	0.04	2

Table 4.3: Ensemble-averaged hydrogen bond occupancies and formation rates [ns^{-1}] per A-G diad in double-helical, non-specifically aggregated, and isolated sections of chain.

4.4.6 Interaction Between Double-Helical Sections and Ions

To understand whether ions' interaction with the double-helix differed to their interaction with isolated chains, ions' spatial distributions surrounding a double-helical section of chain were calculated from 1 μs NaI simulations initiated as SC. As with the results from simulations of isolated chains (Figure 3.6), I^- in the vicinity of the chain was depleted (Figure 4.12 (a)) whereas Na^+ was most concentrated in a sheath surrounding the chain (Figure 4.12 (b-c)), as to be expected for mobile ions electrostatically interacting with an anionic chain. Na^+ concentration about double-helices peaked in a region 0.37 nm from the sulfate groups, whereas for isolated chains the concentration peaked at a distance of 0.24 nm from sulfates, suggesting that sulfate groups in double-helical

sections of chain are better solvated than those groups in isolated sections. However, in the vicinity of double-helices extending out to ∼1.7 nm, Na^+ concentration was approximately double that in the same region surrounding isolated chains, reflecting the increased linear charge density of the double-helix.

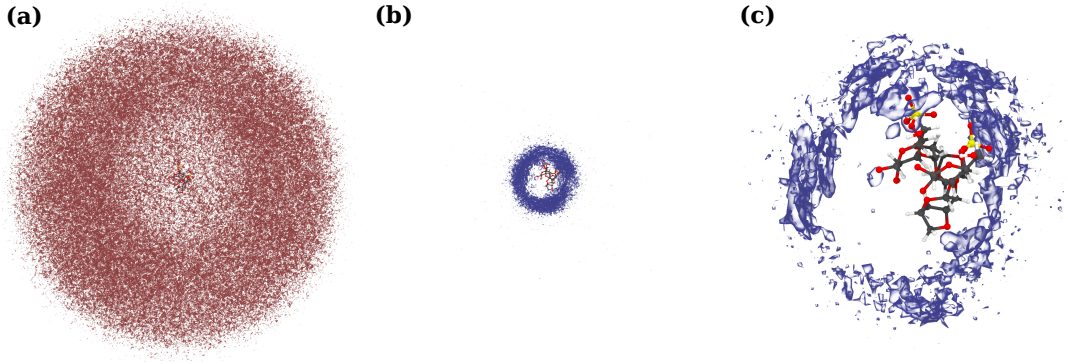

Figure 4.12: Spatial distribution of ions surrounding a double-helical section of chain in a 100 mM NaI solution for (a) I^- and (b) Na^+ ions. The same Na^+ distribution is rendered in (c) at an isovalue of 1.5× that of (a-b). Only two diads (one from each chain) have been rendered for clarity, and the view in all three images is down the chain.

4.5 Discussion

While the atomistic picture of the κ-carrageenan disorder-order transition presented in this chapter (which to the author's knowledge is the most detailed to-date) has been experimentally verified, the MD simulations have multiple limitations which should be recognised. Of primary concern is an overabundance of dimerization during some simulations initiated as SC without NaI; 15 % of residues participated in some sort of dimer in these simulations after 100 ns (Figure 4.3). Based on the large difference between the two experimental WAXS profiles for κ-carrageenan with and without NaI, and their close agreement with those calculated from mostly dimerized and entirely isolated chains respectively (Figure 4.2 (c-d) and (a-b) respectively), one would expect that the fraction of dimerized chains in disordered-state solutions (without NaI) to be low. Combined with the fact that simulations initiated as DH without NaI remained stable after 1 μs, this indicates that at short ranges the non-bonded attractive forces between polymer atoms are erroneously large relative to the repulsive electrostatic forces, which for AMBER family force fields is a known issue (Yoo and Aksimentiev, 2018).

By the same reasoning, one would expect that the fraction of isolated chains in ordered-state solutions (with NaI) to be low, and yet despite eleven of twelve simulations

initiated as SC with NaI resulting in dimers, 60 % of residues remained isolated after 100 ns. This is due to the the MD simulation being limited to just two oligomers, which in many instances form a dimer with a central double-helical and/or non-specifically aggregated section and two isolated ends. Without a third molecule, and with chains too short to form an ouroboros-like structure, these isolated ends remain for the lifetime of the simulated dimer. Within a real solution containing fractions of a mole of polymer chains, however, neither the number of molecules nor the length of chains limit the length of dimerized sections of chain.

The finite size of MD simulations also limits the accuracy of WAXS profiles calculated from their atomic coordinates. This is most apparent for WAXS calculated from simulated pairs of isolated molecules (Figure 4.4 profile (c)), which does not fit the experimental data for disordered-state κ-carrageenan solutions as well as that calculated from single molecule simulations (profile (d)). This is due to the emergence of a structure factor in profile (c); a consequence of the electrostatic potential between chains causing a preferred inter-molecular distance (Muthukumar, 2017; Mansel et al., 2019). While the disordered-state solution (without NaI) also demonstrates this, with a structure factor peak located at $q = 0.5$ nm^{-1}, the simulated profile's structure factor is artificially shifted to a higher q (corresponding to smaller preferred inter-molecular distance) due to the finite size of the simulation box.

Despite the limitations of MD simulations, the agreement between WAXS calculated from simulations and experimental data shown in Figure 4.4 profile (b) gives high confidence that this is a good model for the ordered-state of κ-carrageenan. As shown in Figure 4.2 (c-d), the structure is predominantly double-helical, with short sections of non-specifically aggregated and isolated residues. This heterogeneity has been suggested before based on the results of DSC of κ-carrageenan in Me_4NI solutions (Austen et al., 1988)), and may explain why theoretical predictions assuming an all-or-nothing random coil $\rightarrow$ double-helix transition overestimate the enthalpic change (Paoletti et al., 1984). The presence of non-specifically aggregated sections is explained by kinetic trapping, whereby adjacent double-helical sections prevent non-specifically aggregated sections from finding a more preferable conformation, which would be difficult to escape in a real sample of polymeric chains where longer stretches of double-helices would be possible. Short isolated sections could also be expected in a real sample, for example where chain A associates with chains B and C, with an isolated section of A between B and C.

The inter-molecular arrangement in the model for the ordered-state presented here, with chains primarily parallel and one residue out of register, is consistent with the findings of fibre-XRD experiments for both ι- and κ-carrageenan (Arnott et al., 1974; Millane et al., 1988). However, within double-helical sections the orientation of G

residues was different to that of the XRD models, reducing the inter-molecular HBs by one but increasing the intra-molecular HBs by two in the simulated model (Figure 4.13). The finding that the structure of double-helical sections was independent of whether simulations were initiated as SC or DH (Figures 4.7 and 4.8) indicates that the discrepancy between the scattering profiles (a) and (b) in Figure 4.4 is due to the relative amounts of non-specifically aggregated and isolated sections of chain. However, linear combinations of double-helical and isolated chain WAXS found that no combination fitted the WAXS better than the spontaneous double-helix.

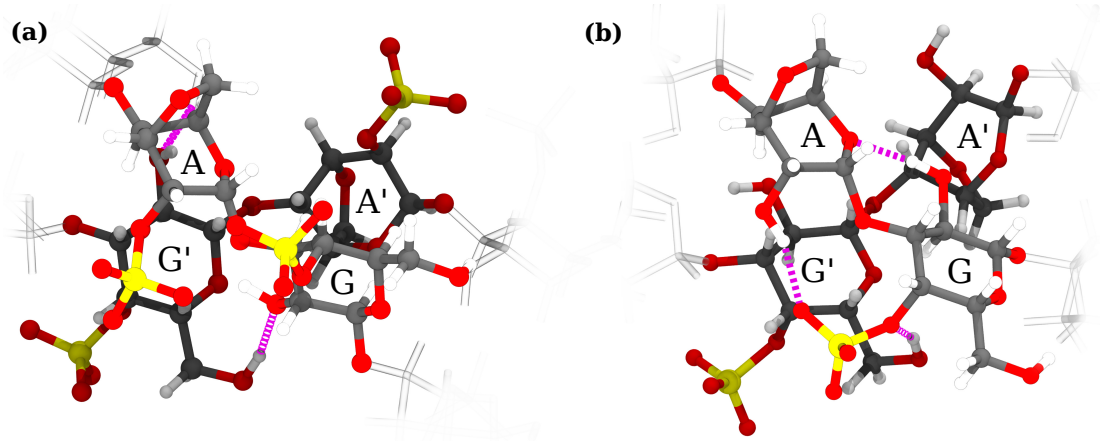

Figure 4.13: Double-helical sections of (a) ι-carrageenan obtained experimentally in the condensed ordered-state (Arnott et al., 1974), and (b) κ-carrageenan after forming spontaneously during MD simulation. Apart from the obvious lack of sulfate residues on A and A′ residues, the structures differ in the orientation of G residues. Dashed magenta lines indicate potential hydrogen bonds.

Simulated double-helical sections are stabilized by the inter-molecular HB G′O6· · ·GSO3 and short-ranged non-bonded forces between the set of atoms making inter-molecular contact (Figure 4.7). Intra-molecular HBs O2· · ·GSO3 and GO2· · ·AO5, which occur at higher rates in double-helical sections, also serve to stabilize the double-helix. These stabilizing interactions all serve to limit the flexibility of chains within double-helical sections, effectively 'locking in' the helical motif which is already present in isolated chains.

In Chapter 3 the simulated κ-carrageenan disordered-state – as validated by experimental WAXS data – was shown to be best described as a 'loose helix', supporting the findings of earlier NMR experiments (Bosco et al., 2005). The simulated ordered-state presented in this chapter – also validated using experimental data – shows unequivocally that the ordered-state is primarily double-helical, and that the individual chains within

double-helices have similar (albeit more stable) helical parameters to the disordered-state (Figure 4.11). The OR-detected disorder-order transition known to occur upon the addition of salt has to-date been assigned as a uni-molecular 'coil-to-helix' transition (Schefer et al., 2015a,b). While reports that the OR-detected transition temperature is independent of polymer concentration support this idea (Smidsrød et al., 1980), more recent findings cast doubt on this hypothesis (Bryce et al., 1982). The current modelling suggests that the structural changes are instead due to the increased number of specifically aggregated pairs of chains when salt is present (Figure 4.3).

The simple explanation for the salt-dependence of dimerization is the screening of the electrostatic interaction between negatively charged chains, which is supported by the fact that all salts induce some degree of OR-detected structural change (Rochas and Rinaudo, 1980; Ciancia et al., 1997). This is, however, only one of many ways in which salt could modulate the inter-molecular potential, which can also have contributions such as hydrophobic effects, HBing, and other interactions which may all be sensitive to the types of ions present. This is most apparent in the variance of transition temperatures for different ion types at a fixed normality (Ciancia et al., 1997), which cannot be explained simply by electrostatic screening alone.

Regardless of the specific interplay between ion-type and inter-molecular potential, what is clear is that once chains can come into contact, short-ranged inter-molecular interactions provide the impetus necessary for double-helix formation. This is in agreement with studies which have found the disorder-order transition to be a cooperative process (Norton et al., 1983; Austen et al., 1988). Crucially, the MD demonstrates a pathway between the monomeric and dimeric states which does not involve a uni-molecular conformational change. In this manner, *salt induces a conformational transition that is a consequence of dimerization, rather than a prerequisite to it.* In the following chapter, this new model for the disorder-order transition will be tested against experimental OR data.

Chapter 5

Modelling κ-Carrageenan Optical Rotation

5.1 Introduction

The sensitivity of OR measurements to an optically active molecule's structure has been known for two centuries (Cahn et al., 1966), and in the study of biomolecules proves to be a valuable non-invasive tool to obtain stereochemical information. The temperature-dependence of a carrageenan solution's OR was first reported in the 1960s (McKinnon et al., 1969), and its interpretation – that of a random coil becoming more helical at low temperatures – drew inspiration from earlier work on polypeptides (Yang and Doty, 1957). Since then it has been used in conjunction with an array of different techniques to classify this biopolymer's disorder-order transition, often being used to estimate the fraction of polymer chains in the ordered conformation.

A review of studies measuring the OR of κ-carrageenan solutions is complicated by the use of light sources with various wavelengths (which the OR is a non-linear function of) and the use of different unit conventions (often either the rotation angle, or rotation angle normalised for path length and either the mass or molar concentration). Despite the differences, two uncontroversial results are found from all κ-carrageenan OR studies: both the disordered- and ordered-states are dextrorotatory (positive OR); and the OR of the fully ordered-state is $\sim$1.7$\times$ that of the fully disordered-state (Morris et al., 1980a; Smidsrød et al., 1980; Rochas and Rinaudo, 1980; Bryce et al., 1982; Rochas et al., 1983; Rochas and Landry, 1987; Semenova et al., 1988; Ciancia et al., 1997; van de Velde et al., 2005; Cardoso and Sabadini, 2010).

The OR in a given temperature range for solutions without a disorder-order transition (usually due to insufficient salt concentration) shows a weak linear dependence on temperature, whereas for solutions with a disorder-order transition a reverse-sigmoidal function is superimposed on the linear relationship (see Figure 1.3). Here the fully

ordered-state is considered to be the state for which the OR is linearly dependant on the temperature at temperatures below the sigmoidal transition region. In contrast, the fully disordered-state is considered to be that for which the OR is linearly dependant on the temperature at temperatures *above* the sigmoidal transition region, or simply at all accessible temperatures if a transition does not exist. The change in OR due to the disorder-order transition is taken as the constant difference between the two extrapolated linear regions which describe the temperature dependence of each state.

In general, the linear dependence of the OR on temperature in disordered-state regions is similar regardless of salt type or concentration (Austen et al., 1988; Morris et al., 1980b), suggesting the disordered-state conformation proposed in Chapter 3 generalises to the disordered-state at any salt concentration. Assuming that the same structural change causes the OR changes for various salts – and that the OR shift from the disordered-state value simply indicates the amount of chains that have undergone a structural change – the ordered-state conformation proposed in Chapter 4 should be able to explain the behaviour of the polymer in the presence of other salts. In the specific case of κ-carrageenan solutions with NaI concentrations of 100 mM, the OR temperature dependence indicates the polymer is in the fully-ordered state at 20 °C, exhibiting a transition centered at 33.4 °C (Ciancia et al., 1997). Unfortunately no units were given for the OR in that study, however based on three studies which report the specific rotation using the sodium D-line wavelength (all of which share similar ordered-state OR values at 20 °C), one would expect a specific rotation $[\alpha]_D$ of ∼45 and ∼73 degrees·mL·g^{-1}dm^{-1} for the disordered- and ordered-states, respectively (Smidsrød et al., 1980; Grasdalen and Smidsrød, 1981a; Mangione et al., 2003).

Given the central importance of OR to the carrageenan story, any model for the disorder-order transition should include an attempt to explain the OR changes known to occur. In this chapter, DFT is used to understand whether the disordered- and ordered-state structures proposed in Chapters 3 and 4 can explain the OR experimentally measured for these states. Particular emphasis is put on understanding the (a) intra-molecular, and (b) inter-molecular structural changes that have the largest effect on the OR computed from the MD simulated and WAXS validated atomic coordinates.

5.2 Calculation Methods

In order to calculate the OR, the atomic coordinates towards the end of the 1 μs simulations whose calculated WAXS best fitted experimental data in Chapter 4 were saved every 1 ns. For the ordered-state, both the double-helices that formed spontaneously during NaI simulations (Figure 4.2 (c-d)) were used, whereas for the disordered-state only one of the simulations that remained as isolated chains when simulated without NaI was used (Figure 4.2 (b)). The use of both double-helical structures ensured the

results for the proposed ordered-state were somewhat independent of the specific way in which each associates, whereas one simulation with isolated chains was deemed sufficient for modelling the disordered-state due to its efficient exploration of most of its conformational space within the time frames involved.

Due to the computationally expensive nature of DFT calculations (relative to MD calculations), the calculation time using all 32 simulated residues would have been prohibitive. Instead, the coordinates for fragments from each of the two chains were extracted from the larger structures, severing at glycosidic linkages and replacing the severed ends with hydroxyl moieties at the reducing ends, and equatorial hydrogens at the non-reducing ends of the fragments, mimicking the polymer's real terminal residues. The placement of hydrogen atoms was made by replacing the following C1 atom with a hydrogen – with the O–H distance scaled to the average length of an O–H bond – in the case of reducing ends, or by replacing the preceding glycosidic oxygen with a hydrogen – with the C–H distance scaled to the average length of an C–H bond – in the case of non-reducing ends.

Length	**Double-helix**[a]	**Isolated**[b]
1	100	151
2	51	71
3	42	91
4	16	46
5	11	41
6	0	21
7	0	12

Table 5.1: Number of optical rotation calculations run for double-helices and isolated chains 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 involving two fragments and calculations done for individual fragments, however the number of runs specifies the number of pairs.

To gain a sense of the validity in approximating the OR of a polymer by calculations performed using the atomic coordinates of oligomers, calculations were done with varying fragment lengths. These ranged from individual residues, up to fragments with seven residues from the simulation of isolated chains (152 atoms) and fragments with five residues from the double-helical simulations (ten residues in total, 214 atoms). To

be consistent with the DH calculations – which had to be calculated with both fragments to capture any interaction between chains (Figure 5.1 (a)) – calculations for isolated chains were also calculated with fragments from both chains (Figure 5.1 (c)). However, to speed up calculations and also allow longer fragments to be used, some calculations for isolated chains were made using individual fragments (Figure 5.1 (b)), based on the assumption that there would be negligible interaction between the two chains which during simulation were separated by ∼3 nm. Regardless of whether calculations were made for double-helices, two isolated chains simultaneously, or isolated chains calculated individually, an even mixture of terminal residue types was ensured (e.g. in Figure 5.1 there is an even mixture of five-residue fragments with A/G residues at the reducing/non-reducing ends).

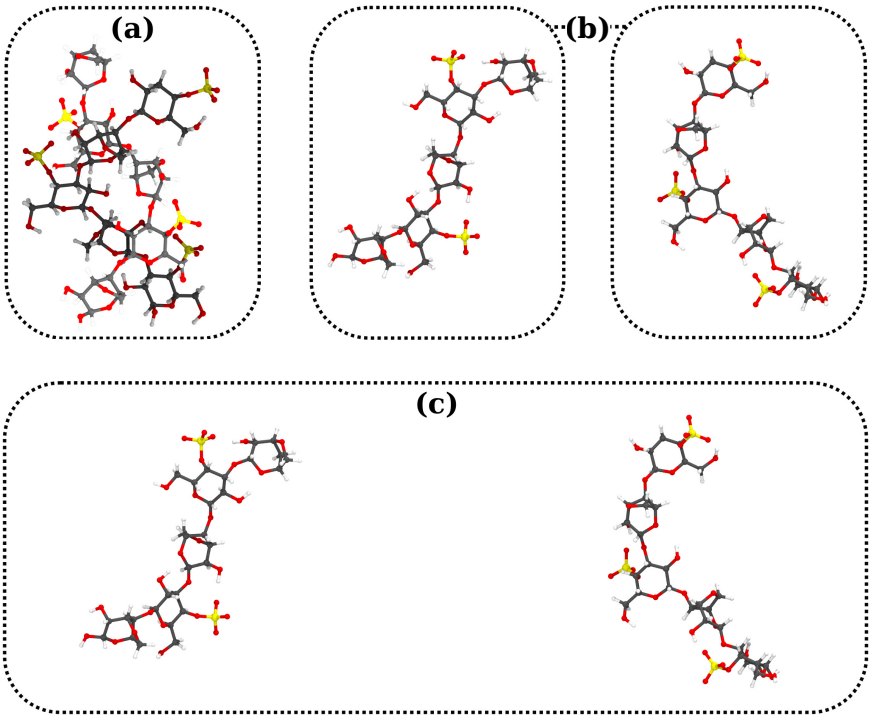

Figure 5.1: κ-carrageenan five-residue fragments extracted from MD simulations of two chains each with 16 residues. Optical rotation calculations were conducted for (a) double-helical fragments formed spontaneously during simulation, (b) chains which remained isolated during simulation and for which optical rotations calculations were performed on individual chains, and (c) the same isolated chains but with the calculation containing both chains simultaneously.

The atomic coordinates for each calculation were then loaded into input files for

Gaussian 09 quantum chemistry software (Frisch et al., 2009). No further geometry optimization was applied to coordinates, as the purpose of this exercise is partly to validate the MD simulations' ability to do this. Calculation of the fragment(s) mixed electric dipole - magnetic dipole polarizability was done using the Polar method with the OptRot keyword (Gaussian, Inc., 2018) and a perturbing frequency corresponding to the sodium D-line at 589.6 nm. In all calculations, an implicit solvation model was applied by using the SCRF keyword, with water set as the solvent option (Gaussian, Inc., 2021). The majority of calculations were performed using the B3LYP hybrid functional (Lee et al., 1988) and 6-31+G(d,p) basis set, and a summary of the OR calculations using this combination can be seen in Table 5.1. The results' dependence on the choice of functional and basis set was tested by repeating a small subset of calculations using the CAM-B3LYP hybrid functional (Yanai et al., 2004) and aug-cc-pVDZ basis set (Woon and Dunning, 1993).

5.3 Results

5.3.1 Functional and Basis Set Dependence

To test the dependence of the calculated OR on the choice of functional and basis set, one simulation frame for a double-helical pair of three-residue fragments and one frame for an isolated pair of three-residue fragments were randomly chosen for analysis. For both, the four combinations of functional (B3LYP and CAM-B3LYP) and basis set (6-31+G(d,p) and aug-cc-pVDZ) were used to calculate the OR, with the results shown in Table 5.2. The difference between OR calculated for the isolated chains versus for the double-helix was largest when using the combination of B3LYP and 6-31+G(d,p) ($\Delta[\alpha]_D = 44.51\,\text{deg.} \cdot \text{mL} \cdot \text{g}^{-1}\text{dm}^{-1}$), and smallest when using the combination of CAM-B3LYP and aug-cc-pVDZ ($\Delta[\alpha]_D = 34.14\,\text{deg.} \cdot \text{mL} \cdot \text{g}^{-1}\text{dm}^{-1}$). In all cases the difference between isolated and double-helical calculations was much larger than the variation between OR calculated using different combinations of functional and basis set for the same structure. Considering this – and that calculations using the B3LYP and 6-31+G(d,p) combination were an order of magnitude faster than other combinations – the B3LYP/6-31+G(d,p) combination was chosen for further use.

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

Table 5.2: The optical rotation values calculated using various combinations of functional and basis set for one pair of three-residue isolated fragments, and one pair of three-residue double-helical fragments.

5.3.2 Interaction Between Isolated Chains

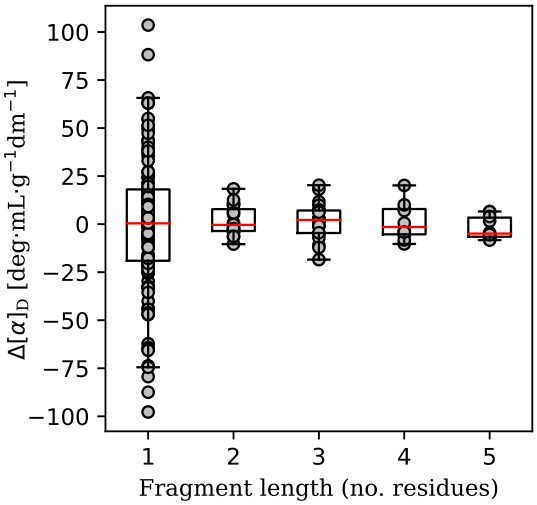

Figure 5.2: The difference between the optical rotation calculated for isolated chains of κ-carrageenan when the calculation for each simulation frame was performed with both chains simultaneously, and when it was performed on the chains individually. Horizontal red bars represent the median value.

Due to the nonlinear scaling of DFT algorithms with the number of atoms, it was advantageous to run calculations for isolated chains individually versus with both chains simultaneously (Figure 5.1 (b) versus (c)). Fragments extracted from the isolated chain

simulation were on average ∼3 nm apart; a distance which when separated by solvent should prevent interaction between chains. To check that this was the case and that the results were independent of the two aforementioned approaches, a subset of calculations performed using both chains simultaneously was repeated for the chains individually. As the OR is normalised for mass concentration, it was important to take the molecular weighted mean of the OR for the two individual fragments when comparing to the OR calculated for both simultaneously. As Figure 5.2 shows, the difference between the OR calculated for chains individually vs simultaneously averages to approximately zero for all chain lengths. While a curious anomaly, the large variation for fragments containing single residues is not an issue for the analysis that follows, as this is concentrated on the results of the larger fragments.

5.3.3 Double-Helices Versus Isolated Chains

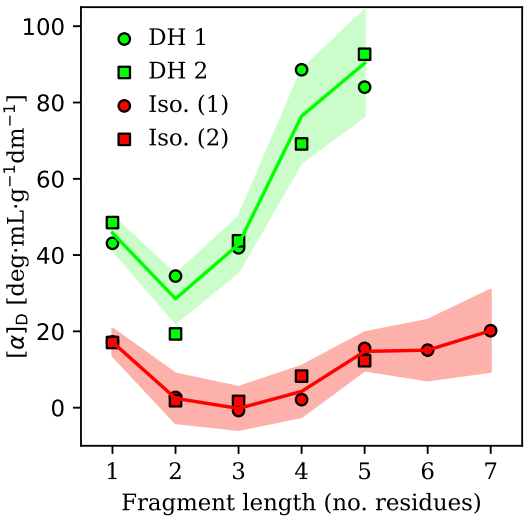

Figure 5.3: Mean optical rotation values calculated for κ-carrageenan fragments as a function of fragment length, with the results for the two double-helical simulations, and two approaches to isolated chain simulations (i.e. individual calculations (1) or simultaneous calculations (2)) shown separately. The solid lines represents the mean value for all double-helices or isolated chains, and the shaded areas represent their 95% confidence intervals.

From Figure 5.3 it is clear that at all fragment lengths the mean OR calculated for both double-helices and isolated chains is dextrorotatory (except isolated three-residue fragments, whose mean OR was zero), and that the mean OR for double-helices is higher than that for isolated chains. The latter point is true regardless of which of the two simulations the double-helical fragments were drawn from, with the mean OR

calculated from both being similar at all fragment lengths. After initially decreasing, the mean calculated OR increases with increasing fragment length, before showing signs of plateauing for both double-helices and isolated chains at longer fragment lengths.

The calculated OR plotted in Figure 5.4 shows significant variation between frames for the same fragment length (particularly for isolated chains) owing to the exploration of the conformational space due to thermal motion. Ordinarily, Boltzmann-weighting the OR values based on the calculated energy for the conformation would be wise when producing a mean, however in this case this is already achieved by the MD, which naturally biases towards energetically favourable conformations leading to more of these being included in the OR mean.

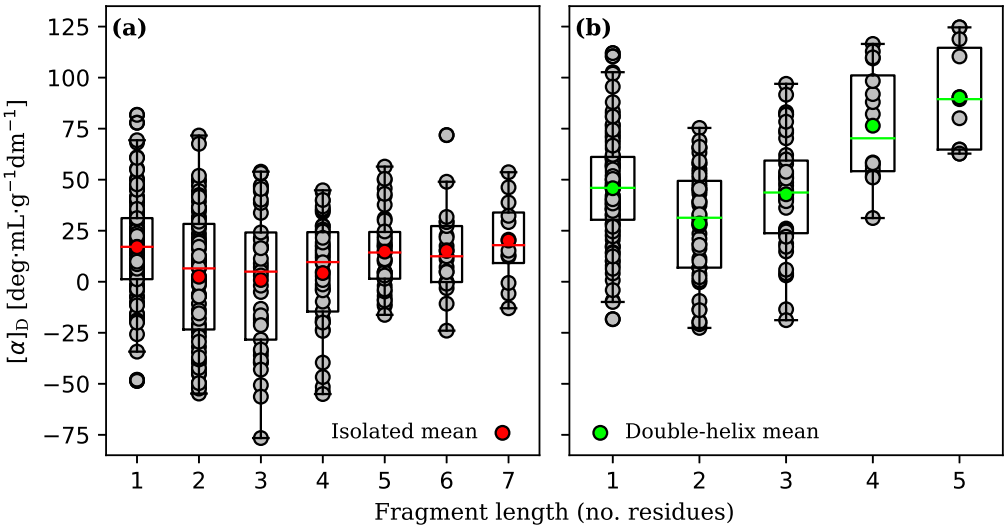

Figure 5.4: Calculated optical rotation distributions for κ-carrageenan (a) isolated chains and (b) double-helices, as a function of fragment length. Red and green horizontal bars represent median values.

5.3.4 Inter-Chain Interaction Effects Versus the Effect of Conformation

To test whether the differences in the OR calculated for isolated chains versus double-helices as shown in Figure 5.3 was due simply to the proximity of the two chains, or due to some difference in the conformation of individual chains when in double-helices as opposed to when isolated, a subset of three-residue double-helices were artificially separated. This consisted of displacing one of the double-helical fragment's COM by ∼3 nm, without changing the conformation of the individual fragments. The OR was then recalculated for the artificially separated pair and compared with the original OR calculated for the double-helix. As Figure 5.5 shows, when double-helical fragments

were separated while maintaining the individual chains' conformation, the OR reduced from (43 ± 7) to (25 ± 6) deg.·mL·g^{-1}dm^{-1}. This is compared to simulations of isolated chains, whose calculated OR for three-residue fragments was (0 ± 6) deg.·mL·g^{-1}dm^{-1}.

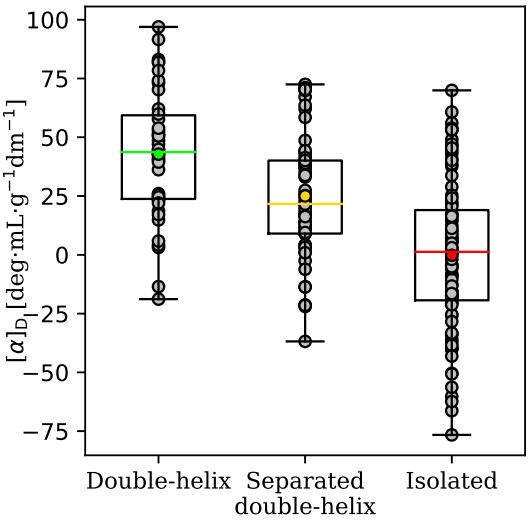

Figure 5.5: Optical rotation calculations for κ-carrageenan three-residue fragments. Double-helical and separated double-helical calculations use the same structures, with the separated double-helical fragments simply displaced by ∼3 nm while maintaining the conformation of individual chains.

5.3.5 Terminal Residues and Their Effect on Optical Rotation

A significant end-effect was observed when OR was calculated for individual isolated chains for all fragment lengths, with the arbitrarily labelled 'fragment A' having a higher mean OR (Figure 5.6). This was the fragment that always had an A residue at the non-reducing end, and is not explained by a different A/G residue number ratio or number of atoms between the two fragments (after all, when containing an even number of residues, both fragments have the same number of each residue). In contrast, the reducing end residue did not affect the OR in the same way; if it did have any effect one would expect the OR difference between the two fragments with an odd number of residues (G reducing ends), and between the two fragments with an even number of residues (A reducing ends) to differ.

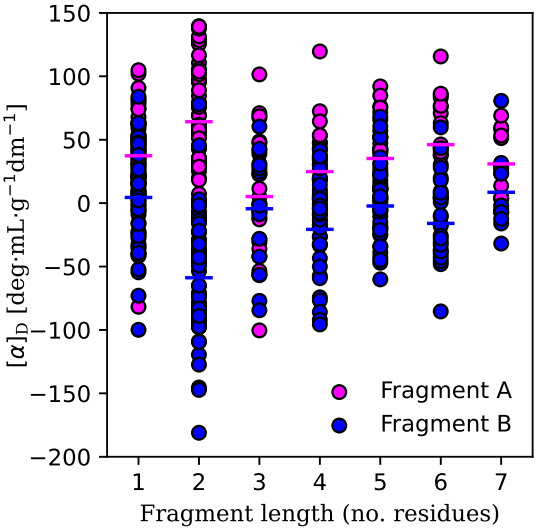

Figure 5.6: Optical rotation calculations for isolated κ-carrageenan fragments when the calculation was performed on the fragments individually and the fragment had an A residue at the non-reducing end (Fragment A) versus a G residue at the non-reducing end (Fragment B). Horizontal coloured lines indicate the distribution means.

5.3.6 Residue Orientations and Their Effect on the Optical Rotation

In an attempt to assign changes to the calculated OR to specific structural changes, the degree to which the OR was correlated with changes to the relative orientation of residues was considered. To do this, the plane of each residue's sugar ring was used to define its orientation, with this vector pointing in the direction of the anhydro ring in the case of A residues and the direction of the sulfate group in the case of G residues (Figure 5.7). The angle formed between each pair of residues' direction vectors was calculated for each conformation and this value's correlation with the calculated OR was determined.

As there is no reason to believe the relationship between residues' orientations and OR might be linear, a simple Pearson correlation coefficient was of limited value. Two other measures of correlation were tested; the Spearman rank correlation coefficient (Zar, 2005), and a non-parametric maximal information coefficient (Reshef et al., 2011). Of the three measures, the Spearman coefficient was found to be the most useful for detecting non-linear association between variables (Figure 5.8), and hence was chosen for use.

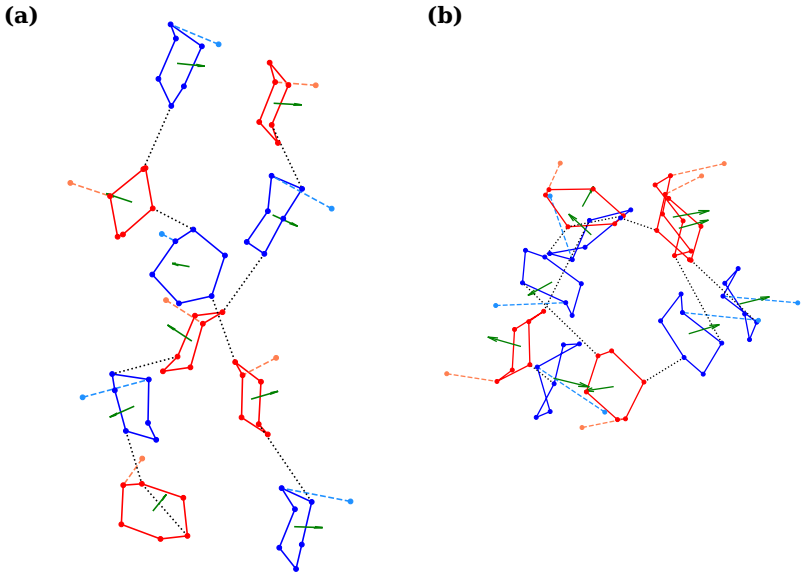

Figure 5.7: Residue sugar rings (solid blue for A residues, solid red for G residues) define a plane whose normal (green arrows) points in the direction of the anhydro ring (dotted blue lines) in the case of A residues or sulfate groups (dotted red lines) in the case of G residues. Shown is a five-residue κ-carrageenan double-helix looking perpendicular to the helical axis (a) and down the helical axis (b).

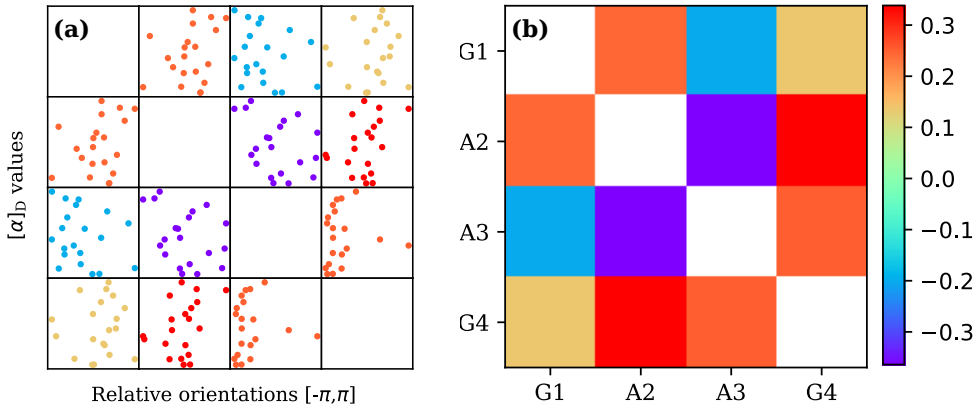

Figure 5.8: (a) Each subplot plots the optical rotation calculated for a two-residue double-helical κ-carrageenan fragment against the relative orientation between a pair of residues within the fragments. (b) The degree to which the calculated OR is correlated to each residue pair's relative orientation is calculated using the Spearman rank correlation coefficient, which in this case ranges from weak negative correlation (between residues A2 and A3) to weak positive correlation (between residues A2 and G4).

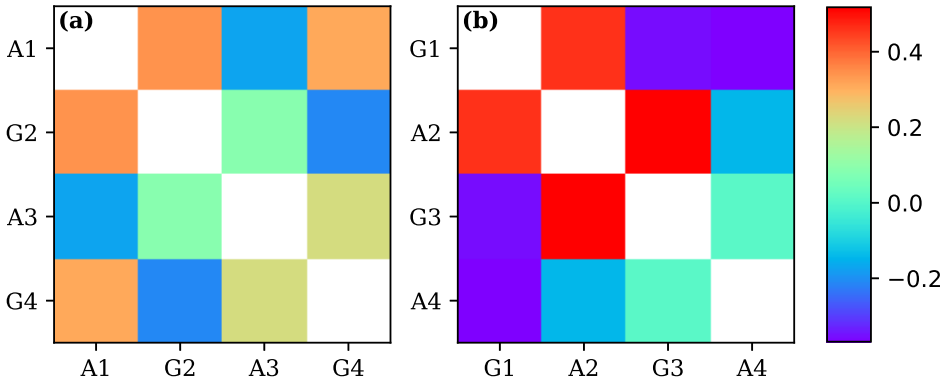

Figure 5.9: Correlation between residues' (A and G correspond to the residue type, numbering is from the non-reducing end) relative orientation and the optical rotation calculated for; (a) four-residue fragment A, and (b) four-residue fragment B, extracted from simulations of isolated κ-carrageenan chains. The colorbar gives the Spearman correlation coefficient.

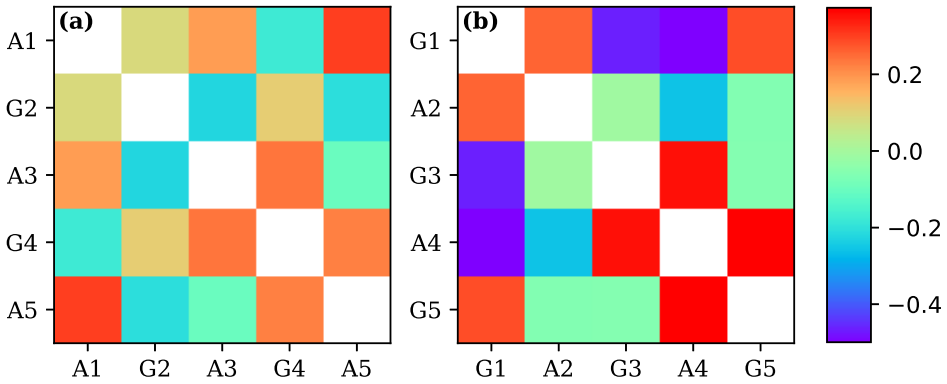

Figure 5.10: Correlation between residues' (A and G correspond to the residue type, numbering is from the non-reducing end) relative orientation and the optical rotation calculated for; (a) five-residue fragment A, and (b) five-residue fragment B, extracted from simulations of isolated κ-carrageenan chains. The colorbar gives the Spearman correlation coefficient.

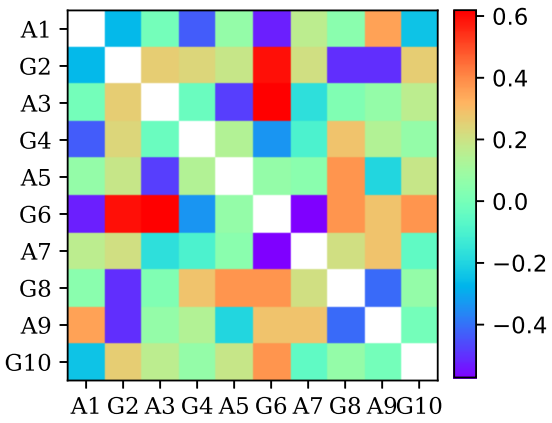

Figure 5.11: Correlation between residues' relative orientation and the optical rotation calculated for five-residue fragments extracted from simulations of double-helical κ-carrageenan chains. A and G correspond to the residue type, and residues 1-5 and 6-10 correspond to fragments A and B, respectively. The colorbar gives the Spearman correlation coefficient.

For smaller fragments from simulations of isolated chains, the correlation between the fragment's calculated OR and the relative orientation of nearest-neighbor residues tended to either be positive, or zero, whereas this correlation for the relative orientation of next-nearest-neighbor residues tended to be negative or zero (Figure 5.9). However, as fragments became larger, this trend became weaker (Figure 5.10). For fragments from simulations of double-helical chains, there was no pattern to which residue pairs were strongly correlated to the calculated OR, although strong correlations for certain pairs was observed (Figure 5.11). Results similar to these were also found for smaller double-helical fragments, but are not shown for brevity.

5.3.7 Inter-Atomic Distances and Their Effect on Optical Rotation

In a catch-all approach to assigning the changes in calculated OR to structural changes, the degree to which the OR was correlated with the distance between every pair of atoms was calculated, again using the Spearman coefficient. For isolated chains, OR calculated simultaneously on both fragments could not be used (as the OR from both fragments is mixed into one OR value), whereas for double-helical chains all intra- and inter-molecular distances were assessed. For isolated chains, the OR was particularly correlated with the intra-molecular distances between residue G2 and neighbour residue A3 – as well as next-neighbour G4 – in fragment A. Fragment B also showed strongest

correlation between the OR and intra-molecular distances between nearest neighbour (G1-A2, G3-A4, A4-G5) residues and next-nearest neighbours (G1-G3 and G3-G5, Figure 5.12 (b)), although there was also a net negative correlation between the OR and distances between A2 and G5. For both fragments the calculated OR showed no net correlation with distances between terminal residues, e.g. A1-A5 for fragment A and G1-G5 for fragment B.

For double-helical fragments, fragment A showed no net correlation between the calculated OR and distances between terminal residues A1-A5. However, for fragment B increases in the OR calculated for the double-helical dimer was correlated with increases in the distance between terminal residues G6-G10, as well as between the non-reducing end G6 and residue A9. In terms of inter-molecular distances, increases in the calculated OR were correlated with increases in the distance between residue G6 in fragment B, and residues G4 and A5 near the reducing end of fragment A. There was also a net negative correlation between the calculated OR and increases to the inter-molecular distance between residues G6 and A1, as well as between residues G10 and A5. The calculated OR was also strongly correlated with structural changes involving residue G8, which was tightly locked in the structure shown in Figure 4.13 (b) with residues G2 and A3 during one of the double-helical simulations The two prominent bands within G8 shown in Figure 5.13 are due to movements of the hydroxymethyl and sulfate groups.

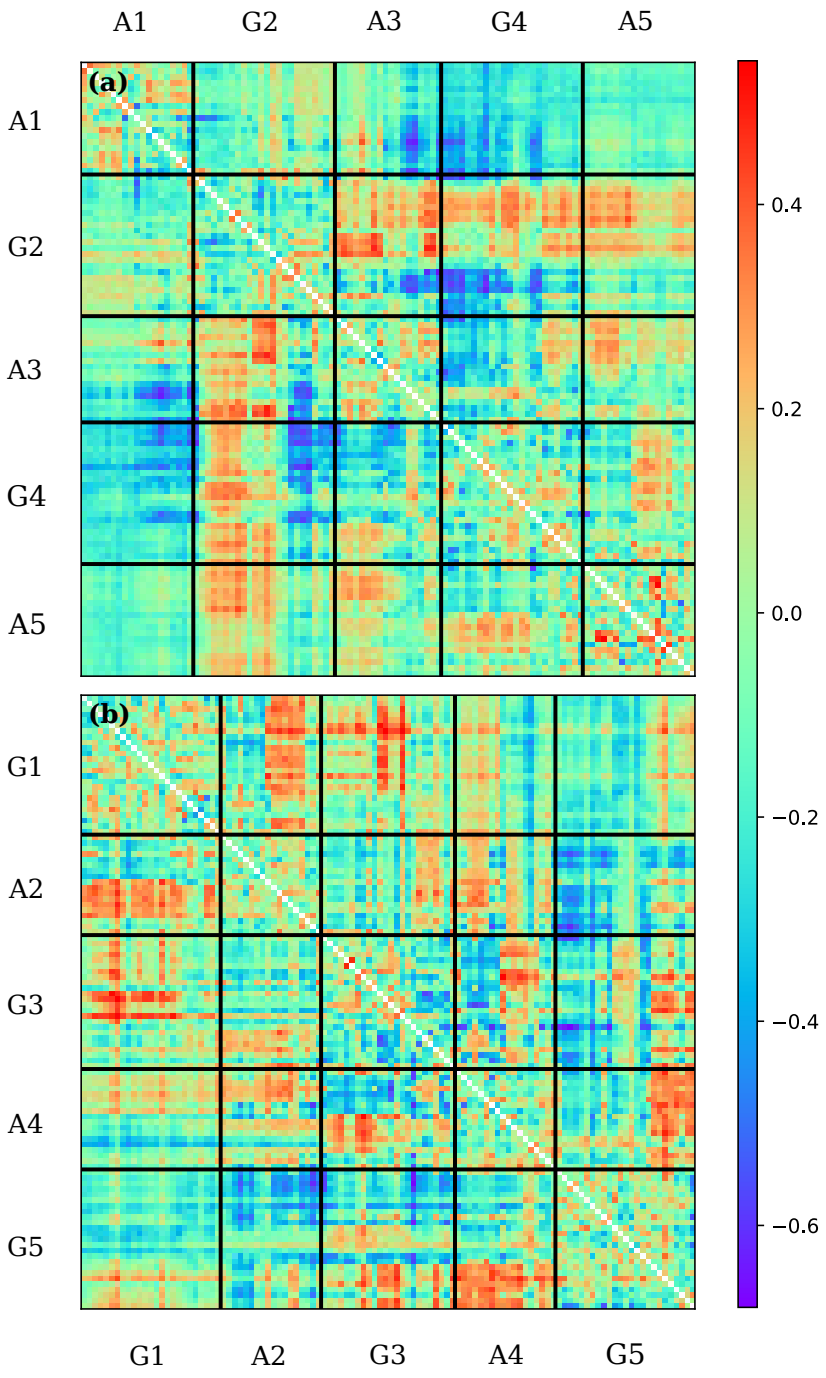

Figure 5.12: Correlation between inter-atomic distances within (a) fragment A, and (b) fragment B, and their calculated optical rotation. Both five-residue fragments are extracted from simulations of isolated κ-carrageenan chains. A and G correspond to the residue type, and numbering begins from the non-reducing ends. The colorbar gives the Spearman correlation coefficient.

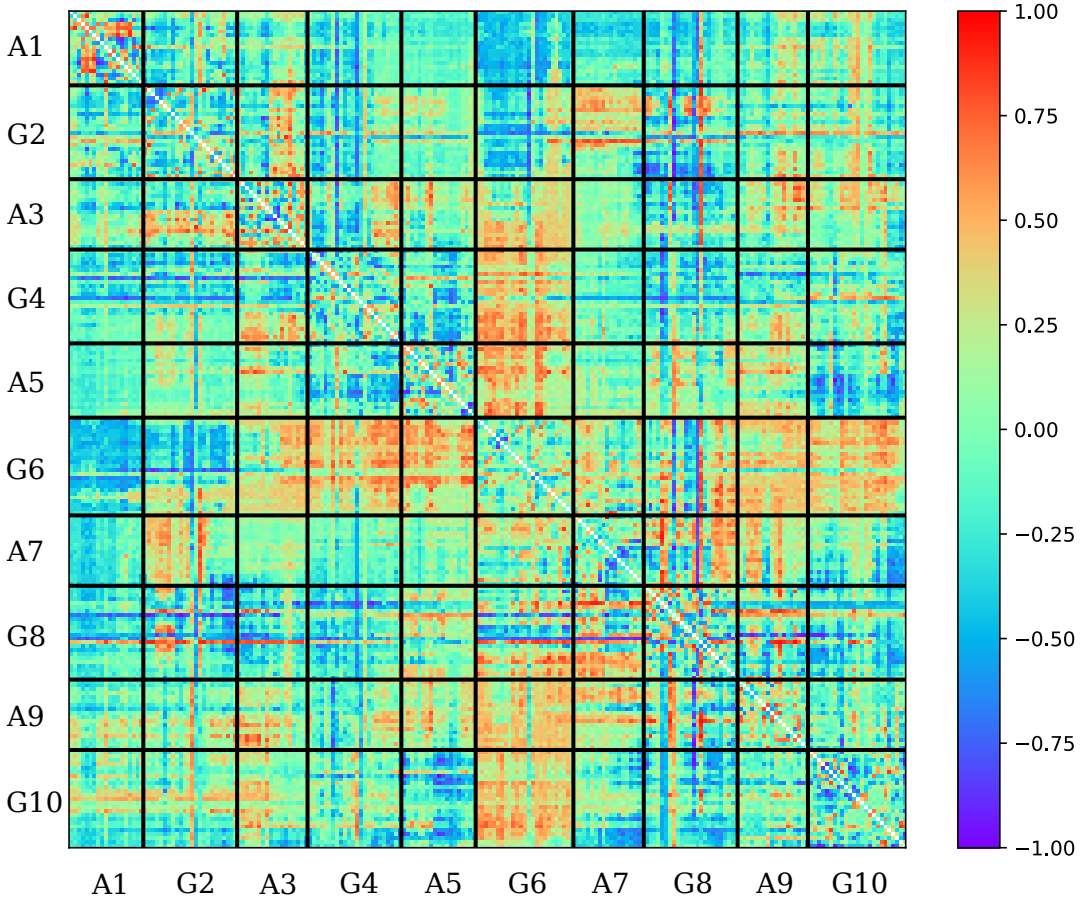

Figure 5.13: Correlation between inter-atomic distances and calculated optical rotation values for five-residue fragments extracted from simulations of double-helical κ-carrageenan chains. A and G correspond to the residue type, and residues 1-5 and 6-10 correspond to fragments A and B, respectively. The colorbar gives the Spearman correlation coefficient.

5.4 Discussion

Two choices were made in an effort to speed up calculations; the use of the B3LYP/6-31+G(d,p) combination of functional and basis set, and the calculation of OR for individual fragments from simulations of isolated chains. The validity of these two choices will now be discussed. As Figure 5.4 shows, for each fragment length the calculated OR varied widely due to the chains' thermal fluctuations, and hence many calculations were required. Although an estimated 70,000 CPU core-hours were used for OR calculations alone, only a handful of calculations for smaller fragments would have been possible if the CAM-B3LYP/aug-cc-pVDZ combination of functional and basis set were used. Although the increased level of theory did make modest changes

to the calculated OR values (Table 5.2) – and indeed this was expected based on the findings of other studies (Cheeseman et al., 2000; Pedersen et al., 2004; Parsons et al., 2022) – better statistics were clearly more advantageous.

By calculating OR for isolated fragments individually rather than simultaneously (Figure 5.1 (b) versus (c)), the approximate halving of the number of atoms reduced the calculation time taken for each pair of fragments to $\sim$ 10% of the time taken to calculate simultaneously. However, for this to be a valid approach it was assumed that the interaction between chains separated by $\sim$ 3 nm was negligible, and indeed Figure 5.2 shows that was true in most cases for fragments with more than one residue. In these cases the interaction between residues was much larger than the interaction between the two fragments. However, for single-residue fragments where inter-residue interactions were absent, the small interaction between single-residue fragments became important. This is the likely explanation for the large deviations from zero for single-residue fragments in Figure 5.2, however as most of the analysis is focused on larger fragments, this wasn't considered an issue.

Artificially separating double-helical fragments effectively removed any chain-chain interactions, and as Figure 5.5 shows, this reduced the calculated OR by 18 deg. $\cdot$ mL $\cdot$ $g^{-1}dm^{-1}$. This was still higher than the OR calculated for conformers simulated as isolated chains, indicating the increase in OR calculated for double-helical fragments had contributions from both intra-molecular structural changes, and chain-chain interactions. The latter has been noted when calculating the OR for both agarose and carrageenan fragments using coupled-cluster theory (Schafer and Stevens, 1995, 1996), and is thought to be responsible for OR changes for other polymers upon double-helix formation (Charney, 1979).

From Figure 5.13 it is clear that the calculated OR is correlated with both intra- and inter-molecular structural changes, however as per the age-old adage, this does not imply causation. For instance, changes to the OR calculated for five-residue double-helical fragments might hypothetically be largely due to the orientation of the G6 residue (as suggested by Figure 5.11), however this could manifest as an apparent link between the OR and inter-atomic distances between G6 residue atoms and atoms within other residues (as seen in Figure 5.13).

Earlier empirical treatments seeking a link between polysaccharide conformation and OR considered the latter to have contributions from the individual residues, and a term dependant on the value of glycosidic dihedral angles (Rees et al., 1970). More recently for sucrose, the OR was found to be highly sensitive to the glycosidic linkage conformation (Stevens and Duda, 1991). However, no strong link between the OR calculated for isolated chains and the dihedral angles adopted by their glycosidic linkages was observed in the current study, which is reflected by the weak correlation between

OR and nearest-neighbour residue-residue orientations (Figure 5.10). Indeed Stevens (1992) observed that the OR of an optically active molecule cannot be reduced to a few simple parameters – a fact reflected by the correlations between intra-molecular distances for isolated chains and the calculated OR as shown in Figure 5.12. The myriad of strong correlations make it impossible to assign changes in the calculated OR for isolated chains to a small subset of structural changes, however in general inter-residue distances (Figure 5.12 points within off diagonal blocks) tended to be more correlated with the OR than intra-residue distances (Figure 5.12 points within blocks along the diagonal). The most obvious exception to this was the weak correlation between the calculated OR and the distance between terminal residues (blocks furthest from the diagonal). This suggests that the OR of separate chains is not dependant on chain extension, which is at odds with the findings of semi-empirical approaches applied to unsulfated carrageenan (Schafer and Stevens (1996)).

For five-residue double-helical fragments, the extension of fragment B – and the distance between opposite ends of fragments A and B – was positively correlated with changes to the calculated OR (Figure 5.13). Combined with a net negative correlation between the calculated OR and the distance between the same ends of fragments A and B (i.e. the reducing ends A5-G10 and non-reducing ends A1-G6), this suggests that parallel chains in close proximity are associated with increases in OR. Another particularly active region in Figure 5.13 is that which involves the residue G8, with the calculated OR being particularly correlated with movements of the hydroxymethyl and sulfate groups. This residue is locked in a very stable double-helical conformation – and is in fact residue G in Figure 4.6 (a) – suggesting that this conformation contributes to the OR changes upon double-helix formation.

To be physically plausible, the calculated OR should not increase without bound as the fragment length is increased, and indeed Figure 5.3 shows that both the isolated and double-helical chains' mean OR begin to plateau at longer fragment lengths. However, the increase in calculated OR values for isolated chains as a function of fragment length intially seems to disagree with the findings of Rochas et al. (1983), who found that for oligomers less than eight residues long in the disordered-state, the observed OR *increased* with decreasing chain length. The explanation for this discrepancy could lie in the enzymatic hydrolysis of the polymer to create oligomers, and the effect of terminal residues on the OR – which has been found to be important when applying theoretical treatments to the OR of agarose (Schafer and Stevens, 1995). The hydrolysis occurs preferentially on the 1→4 linkage (Figure 3.1 (a), Rochas et al. (1983); Rees (1972); Guibet et al. (2007)), guaranteeing one of the products a G residue at the reducing end and the other an A residue at the non-reducing end. While the nature of the reducing end was found to have little effect on the calculated OR, Figure 5.6 shows

that fragments with A residue non-reducing ends had significantly higher OR. Thus because oligomers were more likely to have A residue non-reducing ends, the OR of the solution is increased with respect to the OR of the polymeric solution.

Qualitatively, the results are consistent with the experimental observation that the OR is lower in the disordered-state and higher in the ordered-state (Morris et al., 1980a; Smidsrød et al., 1980; Rochas and Rinaudo, 1980; Bryce et al., 1982; Rochas et al., 1983; Rochas and Landry, 1987; Semenova et al., 1988; Ciancia et al., 1997; van de Velde et al., 2005; Cardoso and Sabadini, 2010), with the smallest difference between isolated chains and double-helices being calculations for two-residue fragments ((2 ± 6) versus (29 ± 6) deg. $\cdot$ mL $\cdot$ g^{-1}dm^{-1}) and the largest being calculations for five-residue fragments ((15±5) versus (90±10) deg. $\cdot$ mL $\cdot$ g^{-1}dm^{-1}). Quantitatively, the calculated OR of 20 ± 10 deg. $\cdot$ mL $\cdot$ g^{-1}dm^{-1} for seven-residue isolated chains is considerably lower than the observed disordered-state value of $\sim$45 deg. $\cdot$ mL $\cdot$ g^{-1}dm^{-1}, and the calculated OR of 90 ± 10 deg. $\cdot$ mL $\cdot$ g^{-1}dm^{-1} for five-residue double-helical chains is somewhat higher than the observed ordered-state value of $\sim$73 deg. $\cdot$ mL $\cdot$ g^{-1}dm^{-1} (Smidsrød et al., 1980; Grasdalen and Smidsrød, 1981a; Mangione et al., 2003). This is easily explained if – instead of assuming an all-or-nothing model – it is assumed that the disordered- and ordered-states simply contain different fractions of double-helical and isolated chains. To understand this, a Monte-Carlo routine (Appendix B) was written to combine OR values calculated for isolated chains and double-helices in different amounts, and then take the weighted mean of the combination of calculated values. As Figure 5.14 (a) shows, the observed disordered-state OR values can be explained by the current OR calculations for the longest chains if it is assumed $\sim$35% of chains exist as double-helices and $\sim$65% exist as isolated chains. As for the observed ordered-state OR values, these can be explained by the current OR calculations if it is assumed $\sim$75% of chains exist as double-helices and $\sim$25% exist as isolated chains (Figure 5.14 (b)). In Chapter 4 it was argued that the disordered-state must be mostly isolated chains based on the comparison between the experimental and calculated WAXS, casting doubt on the explanation of the disordered-state OR as arising from 35% of chains existing as double-helices. This discrepancy would likely be remedied by the use of longer fragments and a higher level of theory, which were both found to increase the OR calculated for isolated chains, thus reducing the percentage of double-helices required to explain the experimentally observed disordered-state OR value.

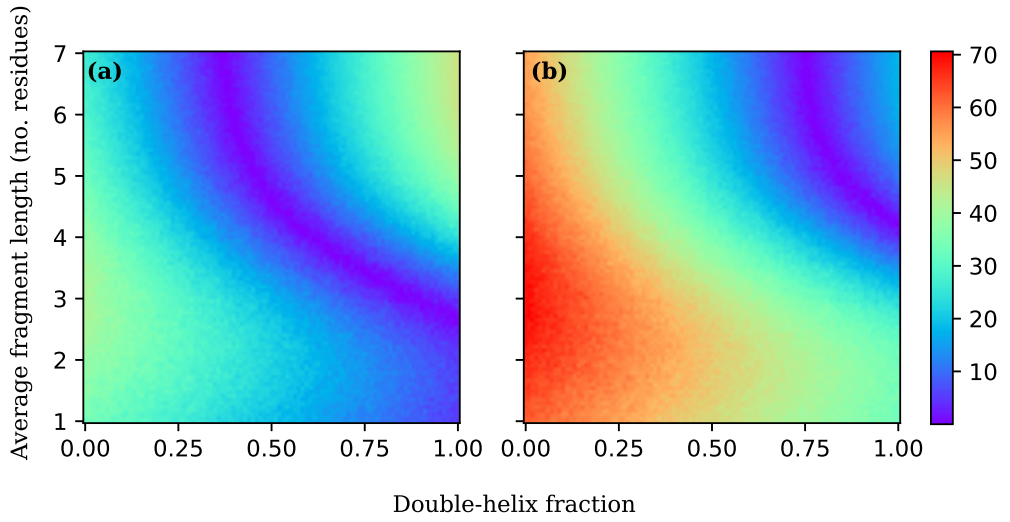

Figure 5.14: The root mean squared difference difference (colorbar units deg. · mL · $g^{-1}dm^{-1}$) between the weighted mean of a given fraction of isolated and double-helical OR calculations for a given fragment length, and the experimentally observed OR for (a) the disordered-state ($[\alpha]_D$=45 deg. · mL · $g^{-1}dm^{-1}$), and (b) the ordered-state ($[\alpha]_D$=73 deg. · mL · $g^{-1}dm^{-1}$).

Chapter 6

Conclusions and Future Directions

It is important for new models to adhere to the landscape dictated by experimental observables, just as it is important for a civil engineer to consider the terrain when designing a new road. In this work, this is achieved by the comparison of WAXS calculated from the proposed disordered and ordered-state models to WAXS obtained experimentally. The excellent agreement between these data sets assures that the models are built on solid ground, and validates the use of the GLYCAM force field for simulating carrageenan systems. As important as it is to ensure a new model is built on solid ground, so to is it important to plan for future work, which will be enumerated in the following for convenience. Based on the success of the current modelling, the first suggestion for future work is to:

1. Expand the modelling to other carrageenan systems, such as ι- and λ-carrageenan. For ι-carrageenan, this would only require calculating partial charges for the available XRD-based structure. For other carrageenans, slight modifications to the structure such as those outlined in Chapter 3 would be required prior to partial charge calculation.

Although the modeled disordered-state glycosidic linkages were found to be restricted to one of two energetic wells, which suggests that a random coil approximation of the chain would be inappropriate, the calculated persistence length of (6 ± 1) nm was nevertheless in agreement with the experimentally observed value of 6.8 nm. However, this experimental value may well be affected by the introduction of kinks in the chain, and thus future work should:

2. Model kinks by randomly replacing MD simulated 3,6-anhydro-D-galactopyranose residues with D-galactopyranose residues.

The restriction of the glycosidic linkages lead to the emergence of a helical secondary structure, which was found to be insensitive to NaI concentration and which had a pitch of 2.17 nm/turn, close to that found for XRD-based models for the condensed ordered-state. This is a major challenge to the long-held belief that carrageenan exists as a random coil in the disordered-state, and that the addition of salt induces a unimolecular structural change.

Recent AFM imaging results have been interpreted as evidence for a random coil structure in the disordered-state and a single-helix structure in the ordered-state, and thus an entire future MD project could aim to:

3. Model the AFM results using MD. This would involve finding or developing parameters for the APTES molecule with which the mica surface is functionalised to immobilise the carrageenan chain in those studies. A surface could then be simulated using a two-dimensional array of such molecules, with the triethoxysilane end immobilised using position restraints and the amino ends in contact with a solvated carrageenan system. Careful thought would have to be made as to how to define the 'height' of simulated carrageenan chains adsorped onto the surface as an AFM probe would see it, but in theory this approach may offer some insight into the single-molecule height changes interpreted as single-helix formation upon the addition of salt as seen in those studies.

Instead of a coil-helix transition, the disorder-order transition was shown to be the dimerization of inherently helical chains, which happens spontaneously at a rate that is dependant on salt concentration. The resulting ordered-state structure was found to be mostly double-helical, with the majority of double-helices having parallel chains approximately one residue out of register and with a helical pitch of 1.92 nm/-turn, resembling the most recent XRD-based double-helical models for the condensed ordered-state. However, the excellent agreement between the calculated and experimental WAXS data was found to be in part due to the existence of imperfections such as isolated and non-specifically aggregated sections of chain. This result could offer an explanation for the inability of early XRD-based modelling of the ordered-state to differentiate between double-helical and side-by-side models, and for the failure of the double-helical model to reproduce experimental DSC results. A promising avenue for further development could therefore be to:

4. Compare experimental DSC results to the calculated enthalpy change upon dimerization, using the currently proposed model and metadynamics. This would in fact involve the opposite process – calculating the theoretical amount of energy required to separate a dimer such as that shown in Figure 4.2 (c-d). To do so, the dimerization process would need to be parameterised by a small number of CVs,

which would then be biased in a similar fashion to the dihedral angles in Chapter 3. Suitable CVs would be the distance between chains, and a CV describing the degree to which the two chains were wrapped around each other. The first CV is trivial to implement, but the implementation of the second would require careful consideration.

The experimental validation of both the disordered- and ordered-state models, combined with the MD simulation trajectories which show a direct path between the two states, paints a complete picture of the disorder-order transition. Once chains are able to come into contact, short-ranged forces lock them into the (mostly double-helical) ordered-state structure. Crucially, there is no uni-molecular transition that precedes this; *the structural changes are a consequence of dimerization rather than a prerequisite to it.* Thus the most interesting unanswered question of this work – and indeed the most promising avenue for an entire future MD project – could be to:

5. Understand the mechanism by which salt modulates the inter-chain potential, increasing the rate of dimerization. This could be achieved by repeating the metadynamics calculation of the enthalpy change upon dimerization as proposed in point 4, instead varying the salt concentration and/or type. To understand the differences between the later, however, would likely require using a polarizable force field such as AMOEBA.

The smoking gun in this work is the calculation of an increased OR upon the disorder-order transition. Considering the central importance of OR to decades of carrageenan research, any new model should understandably raise questions about whether it could explain the disorder-order OR changes. Using the model proposed in this work and DFT, the mean OR calculated for the ordered-state (e.g. (90±10) deg. · mL · $g^{-1}dm^{-1}$ for five-residue double-helices) was always higher than for the disordered-state (e.g. (15±5) deg.·mL·$g^{-1}dm^{-1}$ for five-residue isolated chains). These findings were qualitatively in agreement with experimental data for the disordered- and ordered-states, and quantitative agreement was possible for the disordered-state based on five-residue OR calculations if it was assumed to contain ∼65% isolated and ∼35% double-helical sections of chain. Likewise, quantitative agreement with experimental results was possible for the ordered-state based on five-residue OR calculations if it was assumed to contain ∼25% isolated and ∼75% double-helical sections of chain. These ratios are not necessarily predictions of the true disordered- and ordered-state compositions – particularly for the disordered-state, which based on WAXS data is unlikely to contain 35% double-helical chains – but rather further highlight that all-or-nothing models may be an oversimplification. A number of suggestions were made for future work to bring the calculated OR values into quantitative agreement, such as

using longer fragment lengths and a higher level of theory in DFT calculations. Future work might also:

6. Further optimize the MD-optimized structure using DFT prior to each OR calculation, and take into consideration how a small number of anti-parallel and non-specifically aggregated chains might affect the calculated OR. Although an implicit solvent model was used for all OR calculations, a shell of proximal explicit water molecules could be also included in each calculation using the Gaussian ONIOM procedure.

Attempts to correlate the OR changes with changes to the molecular structure were met with difficulty due to the many degrees of freedom. However, it was possible to conclude that the higher OR calculated for the ordered-state was due to a combination of changes to each chain's conformation upon double-helix formation, and an interaction between chains. A more comprehensive future approach to the problem of understanding the origins of the OR changes would be to:

7. Run time-dependant DFT calculations for a subset of the structures used in Chapter 5. Although this would be much slower than the current approach, this would calculate the electric and magnetic dipole transition moments for every electronic transition, allowing the most significant contributions to the rotational strength to be traced back to the structure from which they originate.

This work has explored only a small area of the carrageenan 'map', and yet it has revealed a topic full of interesting and unsolved problems. I would like to think that what has been presented has explored some new ground, as well as further developed the most stable ground covered previously by others. Just like work on new infrastructure benefits from existing road access, I have benefited greatly from those who have built the synchrotrons, as well as MD and DFT codes that I have used. Although I'm well aware some aspects of the proposed model as shown in Figure 6.1 may be met with resistance – after all people like sticking to familiar territory – I hope there will be parts that others choose to further develop in pursuit of a better understanding of the behaviour of this intriguing family of polysaccharides.

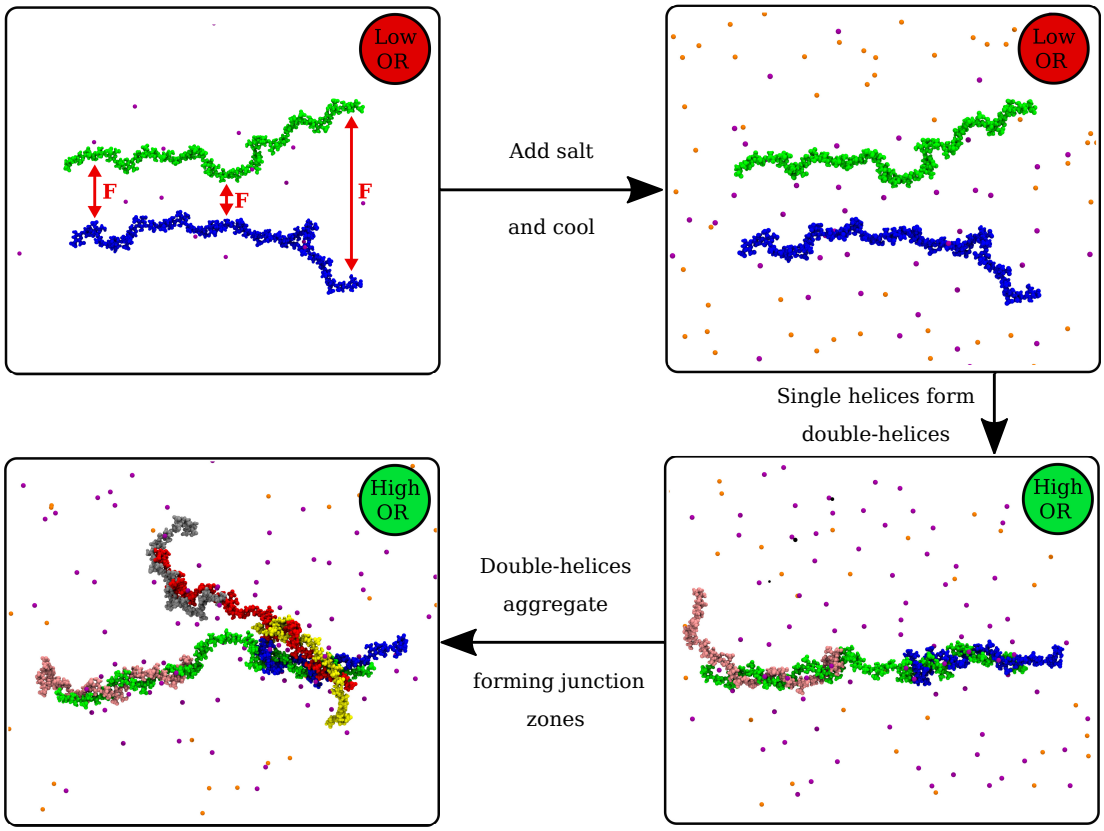

Figure 6.1: The model for κ-carrageenan gelation proposed in this work.

Appendix A

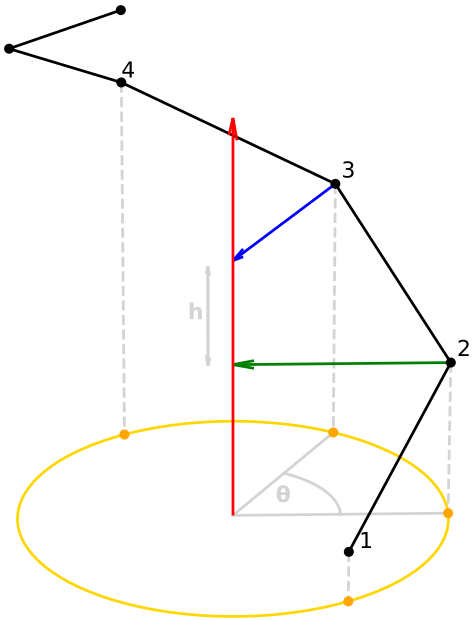

Figure A.1: Points along a model helix, shown in black, where a calculation of helical pitch is performed using a consecutive set of points points 1-4. Green and blue bisector vectors are for the vectors $\mathbf{v}_{2,1}$ and $\mathbf{v}_{2,3}$, and $\mathbf{v}_{3,2}$ and $\mathbf{v}_{3,4}$, respectively. The direction of the local helical axis for the points 1-4 is shown in red, and is perpendicular to the two bisector vectors. The axial translation per point h is found via projection of the points' coordinates onto the helical axis, and the rotation about the helical axis per point θ is found via projecting the points' coordinates onto the plane perpendicular to the helical axis and fitting to a circle (yellow).

To quantify how helical simulated structures were, a program to calculate helical pitch by locating helical axes using a vector-based approach (Kahn, 1989) was designed, sharing a similar procedure to that of the HELANAL program (Bansal et al., 2000). Given a set of atoms defining the path of a helix (Figure SA.1), the program operates in the following manner:

For every set of four consecutive points $\mathbf{r}_i$:

1. Find the bisector vector $\mathbf{b}_{1\text{-}3}$ to the vectors connecting point 2 to points 1 and 3, $\mathbf{v}_{2\text{-}1}$ and $\mathbf{v}_{2\text{-}3}$ respectively
2. Find the bisector vector $\mathbf{b}_{2\text{-}4}$ to the vectors connecting point 3 to points 2 and 4, $\mathbf{v}_{3\text{-}2}$ and $\mathbf{v}_{3\text{-}4}$ respectively
3. Find the direction of the local helix axis $\mathbf{a}$ by taking the cross product of the two bisector vectors
$$\mathbf{a} = \mathbf{b}_{1\text{-}3} \times \mathbf{b}_{2\text{-}4}$$
4. Find two orthonormal vectors $\mathbf{u}_x$ and $\mathbf{u}_y$ to helix axis $\mathbf{a}$ via the Gram-Schmidt process
5. Project points 1-4 onto helical axis $\mathbf{a}$ to find the axial translation:
$$\mathrm{h}_i = \mathbf{a} \cdot (\mathbf{r}_i - \mathbf{r}_1)$$
6. Project points $\mathbf{r}_i$ onto the plane perpendicular to the helix axis defined by $\mathbf{u}_x$ and $\mathbf{u}_y$:
$$\mathrm{r}_{x,i} = \mathbf{u}_x \cdot \mathbf{r}_i$$
$$\mathrm{r}_{y,i} = \mathbf{u}_y \cdot \mathbf{r}_i$$
7. Fit a circle to the points projected in into the perpendicular plane, $[\mathrm{r}_{x,i}, \mathrm{r}_{y,i}]$, finding the coordinates of the circle center $[\mathrm{c}_x, \mathrm{c}_y]$ in the perpendicular plane
8. Use the fitted circle center $[\mathrm{c}_x, \mathrm{c}_y]$ and the projected points $[\mathrm{r}_{x,i}, \mathrm{r}_{y,i}]$ to find the rotation per point θ_i:
$$\theta_i = \cos^{-1}\left(\frac{[r_{x,i} - c_x, r_{y,i} - c_y] \cdot [r_{x,1} - c_x, r_{y,1} - c_y]}{|[r_{x,i} - c_x, r_{y,i} - c_y]||[r_{x,1} - c_x, r_{y,1} - c_y]|}\right)$$
9. Calculate the pitch for each point:
$$p_i = h_i/\theta_i$$

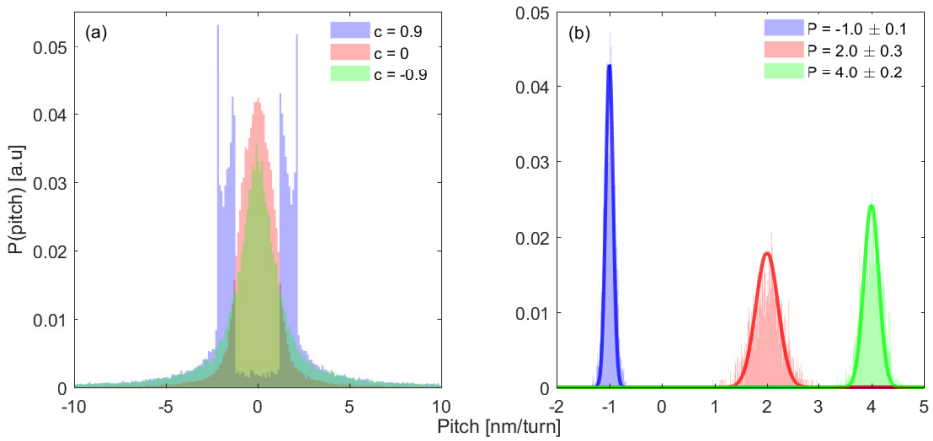

Figure A.2: Measured pitch distributions for (a) random coils with bond angle correlation of 0.9, 0, and -0.9 (1000 generated structures each), and (b) pitch measurements (shaded) of helices generated from distributions (solid lines) $P = -1.0 \pm 0.1$ (left-handed), $P = 2.0 \pm 0.3$, and $P = 4.0 \pm 0.2$ nm/turn (1000 generated structures each).

For polymer chains with a degree of polymerisation N, where each residue defines one point along the helical axis, this procedure yields N-4 values of helical pitch. The program was thoroughly tested, first by checking that non-helical structures were measured as such. To do this, random coils consisting of 30 points were generated, either with no correlation between successive bond vectors (truly random coils), correlation of 0.9 between successive bond vectors (nearly linear chains), or correlation of -0.9 between successive bond vectors (nearly collapsed chains). For each degree of bond vector correlation, 1000 coils were randomly generated, and it was found that in all cases the distribution of pitch (Figure SA.2 (a)) was centered on zero (i.e. not helical).

The program was then tested on helical structures with 30 points, where 1000 structures were randomly generated with a distribution of helical pitches. In almost all scenarios the program was able to determine the correct mean pitch and standard deviation from the set of generated structures (Figure SA.2 (b)), only failing when the separation between consecutive points approached the limit of precision supported by the file type (PDB in this case). The program proved to be robust against random noise and bending of the helix, only beginning to fail when the random noise or bending radius approached the same order of magnitude as the helical radius.

Appendix B

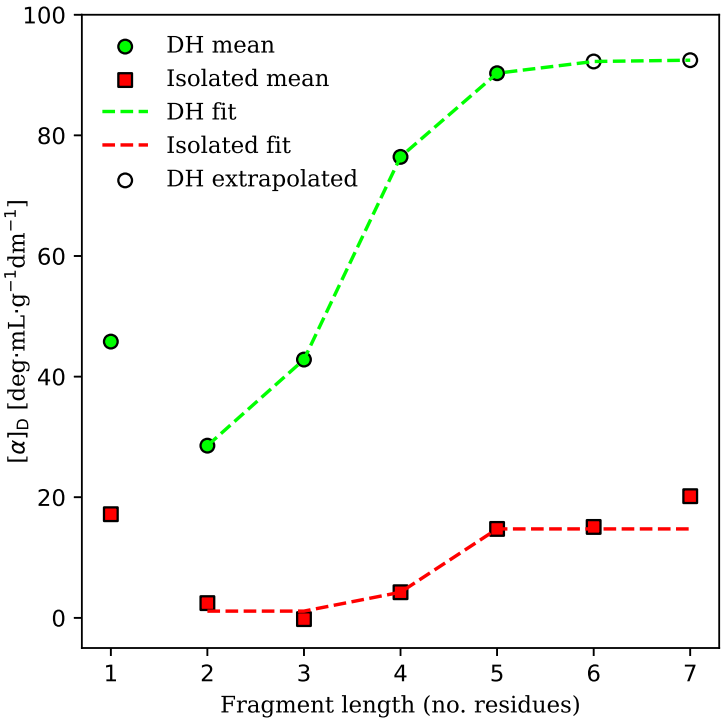

Figure B.1: Fitting of calculated optical rotation mean values as a function of fragment length. Double-helical values between fragment lengths of 2 and 5 were fit exceptionally well by a sigmoidal function, whereas the fit to isolated chain values between lengths of 2 and 7 was adequate.

Assuming the experimentally observed OR values of ∼45 and ∼73 deg.·mL·g^{-1}dm^{-1} for the disordered- and ordered-states (respectively) can be explained as a linear combination of OR values calculated for isolated and double-helical chains, a Monte-Carlo program was written to randomly select from the distributions of calculated values to

form a mean. Before the main loop runs, the mean values for double-helical OR values as a function of fragment length are fitted to a sigmoidal function (Figure B.1) in order to find extrapolated values for fragment lengths of 6 and 7 residues. Drawing from the distributions of OR values at each fragment length for isolated and double-helical chains – as well as the two extrapolated values for longer double-helical fragments – the program operates in the following manner (shown pictorially in Figure B.2):

1. Pick a double-helical fraction x and average fragment length L.
2. Decide how to partition N randomly selected OR values into $N_h = N \cdot x$ double-helical values and $N_s = N(1 - x)$ isolated chains values.
3. Partition the N_h double-helical values into $N_{h,l}$ values so that the number of randomly drawn values from each fragment length l approximates a Gaussian distribution centered on the average fragment length L and with a variance of 1 residue. Repeat for the N_s isolated chain values, again forming a fragment length distribution centered on L and with 1 residue variance, but in most cases a different total number of randomly drawn values N_s (unless $x = 0.5$).
4. Take root mean squared difference between the mean OR value of the combination of all N_h and N_s randomly selected OR values, and the experimentally observed value.

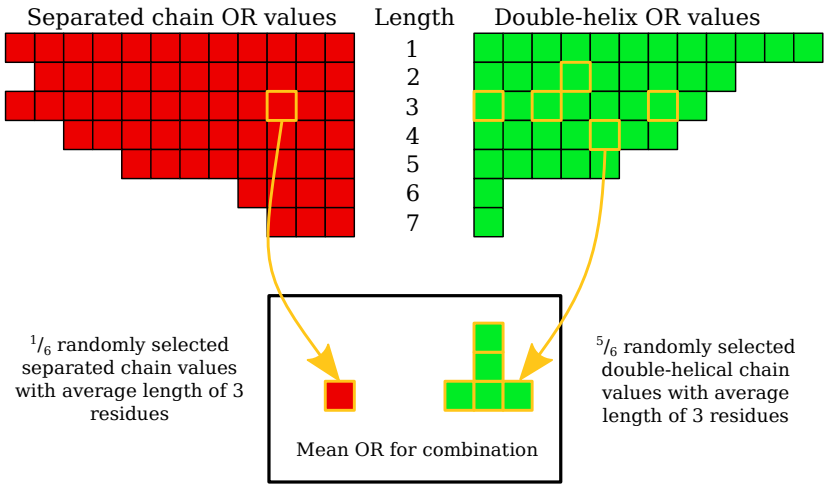

Figure B.2: Routine to randomly select OR values from calculated distributions. The example draws $N = 6$ values with an average fragment length of 3 and a double-helical fraction of 5/6.

Bibliography

Hans C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72(4):2384–2393, February 1980.

N S Anderson, J W Campbell, M M Harding, D A Rees, and J W B Samuel. X-ray Diffraction Studies of Polysaccharide Sulphates : Double Helix Models for κ-and ι-Carrageenans. *Journal of Molecular Biology*, 45:85–99, 1969.

S Arnott, W E Scott, D A Rees, and C G A McNabb. ι-Carrageenan: Molecular Structure and Packing of Polysaccharide Double Helices in Oriented Fibres of Divalent Cation Salts. *Journal of Molecular Biology*, 90:253–267, 1974.

Kevan R. J. Austen, David M. Goodall, and Ian T. Norton. Anion Effects on Equilibria and Kinetics of the Disorder-Order Transition of κ-Carrageenan. *Biopolymers*, 27(1): 139–155, January 1988.

Sindrila Dutta Banik and Nilashis Nandi. Chirality and Protein Biosynthesis. In Pedro Cintas, editor, *Biochirality*, volume 333, pages 255–305. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. Series Title: Topics in Current Chemistry.

Manju Bansal, Sandeep Kumart, and R. Velavan. HELANAL: A Program to Characterize Helix Geometry in Proteins. *Journal of Biomolecular Structure and Dynamics*, 17(5):811–819, April 2000.

L. D. Barron. *Molecular Light Scattering and Optical Activity.*, volume 2nd ed., rev. and enl. Cambridge University Press, 2004.

H J C Berendsen, J R Grigera, and T P Straatsma. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.*, 91:6269–6271, 1987.

Bruce J. Berne and Robert Pecora. *Dynamic light scattering : with applications to chemistry, biology, and physics.* Wiley, 1976.

Karin Bongaerts, Harry Reynaers, Flavio Zanetti, and Sergio Paoletti. On the Molar

Mass of κ-Carrageenan in the Course of Conformational Transition from the Disordered to the Fundamental Ordered Form. *Macromolecules*, 32(3):675–682, February 1999.

J. Borgström, L. Piculell, C. Viebke, and Y. Talmon. On the structure of aggregated kappa-carrageenan helices. A study by cryo-TEM, optical rotation and viscometry. *International Journal of Biological Macromolecules*, 18(3):223–229, April 1996.

M. Born and R. Oppenheimer. Zur Quantentheorie der Molekeln. *Annalen der Physik*, 389(20):457–484, 1927.

M Bosco, A Segre, S Miertus, A Cesàro, and S Paoletti. The disordered conformation of κ-carrageenan in solution as determined by NMR experiments and molecular modeling. *Carbohydrate Research*, 340:943–958, 2005.

Thomas A. Bryce, Allan H. Clark, David A. Rees, and David S. Reid. Concentration Dependence of the Order-Disorder Transition of Carrageenans. Further Confirmatory Evidence for the Double Helix in Solution. *European Journal of Biochemistry*, 122 (1):63–69, February 1982.

A. Brzyska, W. Płaziński, and K. Woliński. Force-induced structural changes in non-sulfated carrageenan based oligosaccharides – a theoretical study. *Soft Matter*, 14 (30):6264–6277, 2018.

Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity-rescaling. *The Journal of Chemical Physics*, 126(1):1–8, January 2007.

R. S. Cahn, Christopher Ingold, and V. Prelog. Specification of Molecular Chirality. *Angewandte Chemie International Edition in English*, 5(4):385–415, April 1966.

Marcus Vinícius Cangussu Cardoso and Edvaldo Sabadini. The gelling of κ-carrageenan in light and heavy water. *Carbohydrate Research*, 345(16):2368–2373, November 2010.

Mark E. Casida. Time-Dependent Density Functional Response Theory for Molecules. In *Recent Advances in Computational Chemistry*, volume 1, pages 155–192. Word Scientific, November 1995.

Elliot Charney. *The Molecular Basis of Optical Activity.* John Wiley and Sons, Inc, 1979.

James R. Cheeseman, Michael J. Frisch, Frank J. Devlin, and Philip J. Stephens. Hartree-Fock and Density Functional Theory ab Initio Calculation of Optical Rotation Using GIAOs: Basis Set Dependence. *The Journal of Physical Chemistry A*, 104(5):1039–1046, February 2000.

M Ciancia, M Milas, and M Rinaudo. On the specific role of coions and counterions on kappa-carrageenan conformation. *Int. J. Bio. Macromol.*, 20:35–41, 1997.

Wendy D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, and Peter A. Kollman. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, May 1995.

T. Daniel. Crawford. Ab initio calculation of molecular chiroptical properties. *Theoretical Chemistry Accounts*, 115(4):227–245, April 2006.

P. Debye. Zerstreuung von röntgenstrahlen. *Annalen der Physik*, 351(6):809–823, 1915.

Michael Diener, Jozef Adamcik, Antoni Sánchez-Ferrer, Florian Jaedig, Larissa Schefer, and Raffaele Mezzenga. Primary, Secondary, Tertiary and Quaternary Structure Levels in Linear Polysaccharides: From Random Coil, to Single Helix to Supramolecular Assembly. *Biomacromolecules*, 20(4):1731–1739, April 2019.

R. Ditchfield, W. J. Hehre, and J. A. Pople. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *The Journal of Chemical Physics*, 54(2):724–728, January 1971.

David S Domozych. Algal Cell Walls. In John Wiley & Sons, Ltd, editor, *eLS*, pages 1–11. John Wiley & Sons, Ltd, Chichester, UK, September 2011.

Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen. A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, 103(19):8577–8593, November 1995.

D. Franke, M. V. Petoukhov, P. V. Konarev, A. Panjkovich, A. Tuukkanen, H. D. T. Mertens, A. G. Kikhney, N. R. Hajizadeh, J. M. Franklin, C. M. Jeffries, and D. I. Svergun. *ATSAS 2.8* : a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of Applied Crystallography*, 50(4):1212–1225, August 2017.

M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov,

R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian 09. 2009.

Gaussian, Inc. Polar. 2018. URL `www.gaussian.com/polar/`. Accessed: 19-07-2022.

Gaussian, Inc. SCRF. 2021. URL `www.gaussian.com/scrf/`. Accessed: 19-07-2022.

H Grasdalen and O Smidsrød. Iodide-Specific Formation of κ-Carrageenan Single Helices. ^{127}I NMR Spectroscopic Evidence for Selective Site Binding of Iodide Anions in the Ordered Conformation. *Macromolecules*, 14:1842–1845, 1981a.

Hans Grasdalen and Olav Smidsrød. ^{113}Cs NMR in the Sol-Gel States of Aqueous Carrageenan. Selective Site Binding of Cesium and Potassium Ions in κ-Carrageenan Gels. *Macromolecules*, 14(1):229–231, 1981b.

David J. Griffiths. *Introduction to electrodynamics.* Pearson, 2014.

Marion Guibet, Sébastien Colin, Tristan Barbeyron, Sabine Genicot, Bernard Kloareg, Gurvan Michel, and William Helbert. Degradation of λ-carrageenan by *Pseudoalteromonas carrageenovora* λ-carrageenase: a new family of glycoside hydrolases unrelated to κ- and ι-carrageenases. *Biochemical Journal*, 404(1):105–114, May 2007.

W. H. Brooks, W. C. Guida, and K. G. Daniel. The Significance of Chirality in Drug Design and Development. *Current Topics in Medicinal Chemistry*, 11(7):760–770, April 2011.

Boualem Hammouda. Analysis of the Beaucage model. *Journal of Applied Crystallography*, 43(6):1474–1478, December 2010.

Jason E. Hein and Donna G. Blackmond. On the Origin of Single Chirality of Amino Acids and Sugars in Biogenesis. *Accounts of Chemical Research*, 45(12):2045–2054, December 2012.

Anne-Marie Hermansson. Rheological and Microstructural Evidence for Transient States During Gelation of Kappa-Carrageenan in the Presence of Potassium. *Carbohydrate Polymers*, 10:163–181, 1988.

J.F.W. Herschel. On the rotation impressed by plates of rock crystal on the planes of polarization of the rays of light, as connected with certain peculiarities in its crystallization. *Transactions of the Cambridge Philosophical Society*, 1:43–51, 1820.

T Hjerde, O Smidsrød, B T Stokke, and B E Christensen. 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, 1998.

T Hjerde, O Smidsrød, and B E Christensen. Analysis of the Conformational Properties of ι- and κ-Carrageenan by Size-Exclusion Chromatography Combined with Low-Angle Laser Light Scattering. *Biopolymers*, 49:71–80, 1999.

Scott A. Hollingsworth and Ron O. Dror. Molecular Dynamics Simulation for All. *Neuron*, 99(6):1129–1143, September 2018.

Jochen S Hub. Interpreting solution X-ray scattering data using molecular simulations. *Current Opinion in Structural Biology*, 49:18–26, April 2018.

Amir H. Irani, Jessie L. Owen, Davide Mercadante, and Martin A. K. Williams. Molecular Dynamics Simulations Illuminate the Role of Counterion Condensation in the Electrophoretic Transport of Homogalacturonans. *Biomacromolecules*, 18(2):505–516, February 2017.

Araz Jakalian, David B. Jack, and Christopher I. Bayly. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of Computational Chemistry*, 23(16):1623–1641, October 2002.

Srinivas Janaswamy and Rengaswami Chandrasekaran. Effect of calcium ions on the organization of ι-carrageenan helices: an X-ray investigation. *Carbohydrate Research*, 337(6):523–535, March 2002.

Peter C. Kahn. Defining the axis of a helix. *Computers & Chemistry*, 13(3):185–189, January 1989.

Nigel M. Kirby, Stephen T. Mudie, Adrian M. Hawley, David J. Cookson, Haydyn D. T. Mertens, Nathan Cowieson, and Vesna Samardzic-Boban. A low-background-intensity focusing small-angle X-ray scattering undulator beamline. *Journal of Applied Crystallography*, 46(6):1670–1680, December 2013.

Karl N. Kirschner, Austin B. Yongye, Sarah M. Tschampel, Jorge González-Outeiriño, Charlisa R. Daniels, B. Lachele Foley, and Robert J. Woods. GLYCAM06: A generalizable biomolecular force field. Carbohydrates: GLYCAM06. *Journal of Computational Chemistry*, 29(4):622–655, March 2008a.

Karl N. Kirschner, Austin B. Yongye, Sarah M. Tschampel, Jorge González-Outeiriño,

Charlisa R. Daniels, B. Lachele Foley, and Robert J. Woods. GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *Journal of Computational Chemistry*, 29(4):622–655, March 2008b.

W. Kohn and L. J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review*, 140(4A):A1133–A1138, November 1965.

Damien Laage, Thomas Elsaesser, and James T. Hynes. Water Dynamics in the Hydration Shells of Biomolecules. *Chemical Reviews*, 117(16):10694–10725, August 2017.

A. Laio and M. Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, October 2002.

Jean-Yves Le Questel, Soizic Cros, William Mackie, and Serge Pérez. Computer modelling of sulfated carbohydrates: Applications to carrageenans. *International Journal of Biological Macromolecules*, 17(3-4):161–175, January 1995.

Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37(2):785–789, January 1988.

Jerzy Leszczynski, Anna Kaczmarek-Kedziera, Tomasz Puzyn, Manthos G. Papadopoulos, Heribert Reis, and Manoj K. Shukla, editors. *Handbook of Computational Chemistry*. Springer International Publishing, Cham, 2017.

Liang Li, Rui Ni, Yang Shao, and Shirui Mao. Carrageenan and its applications in drug delivery. *Carbohydrate Polymers*, 103:1–11, March 2014.

Zezhong John Li, Simcha Srebnik, and Orlando J. Rojas. Revisiting Cation Complexation and Hydrogen Bonding of Single-Chain Polyguluronate Alginate. *Biomacromolecules*, 22(9):4027–4036, September 2021.

M R Mangione, D Giacomazza, D Bulone, V Martorana, and P L San Biagio. Thermoreversible gelation of κ-Carrageenan: relation between conformational transition and aggregation. *Biophysical Chemistry*, 104:95–105, 2003.

Bradley W. Mansel, Amir Hossein Irani, Timothy M. Ryan, Duncan J. McGillivray, Hsin-Lung Chen, and Martin A. K. Williams. Resolving solution conformations of the model semi-flexible polyelectrolyte homogalacturonan using molecular dynamics simulations and small-angle x-ray scattering. *The European Physical Journal E*, 42 (2):1–10, February 2019.

Bradley W. Mansel, Timothy M. Ryan, Hsin-Lung Chen, Leif Lundin, and Martin A.K. Williams. Polysaccharide conformations measured by solution state x-ray scattering. *Chemical Physics Letters*, 739:1–5, 2020.

A. A. McKinnon, D. A. Rees, and F. B. Williamson. Coil to Double Helix Transition for a Polysaccharide. *Journal of the Chemical Society D: Chemical Communications*, 1:701–702, 1969.

Rick P. Millane, Rengaswami Chandrasekaran, Struther Arnott, and Iain C.M. Dea. The molecular structure of kappa-carrageenan and comparison with iota-carrageenan. *Carbohydrate Research*, 182(1):1–17, October 1988.

Edwin R Morris, David A Rees, and Geoffrey Robinson. Cation-specific Aggregation of Carrageenan Helices : Domain model of polymer gel structure. *J. Mol. Bio.*, 138: 349–362, 1980a.

Edwin R. Morris, David A. Rees, Geoffrey Robinson, and Graham A. Young. Competitive inhibition of interchain interactions in polysaccharide systems. *Journal of Molecular Biology*, 138(2):363–374, April 1980b.

R Morris, David A Rees, Ian T Norton, and David M Goodall. Calorimetric and Chiroptical Evidence of Aggregate-Driven Helix Formation in Carrageenan Systems. *Carbohydrate Research*, 80:317–323, 1980c.

M. Muthukumar. 50^{th} Anniversary Perspective : A Perspective on Polyelectrolyte Solutions. *Macromolecules*, 50(24):9528–9560, December 2017.

Ian T. Norton, David M. Goodall, Edwin R. Morris, and David A. Rees. 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(10):2489, 1983.

Sergio Paoletti, Olav Smidsrød, and Hans Grasdalen. Thermodynamic stability of the ordered conformations of carrageenan polyelectrolytes. *Biopolymers*, 23(9):1771–1794, September 1984.

Sergio Paoletti, Franco Delben, Attilio Cesàro, and Hans Grasdalen. Conformational Transition of κ-Carrageenan in Aqueous Solution. *Macromolecules*, 18(10):1834–1841, October 1985.

M Parrinello and A Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, 1981.

Taylor Parsons, Ty Balduf, James R. Cheeseman, and Marco Caricato. Basis Set Dependence of Optical Rotation Calculations with Different Choices of Gauge. *The Journal of Physical Chemistry A*, 126(11):1861–1870, March 2022.

Thomas Bondo Pedersen, Alfredo M. J. Sánchez de Merás, and Henrik Koch. Polarizability and optical rotation calculated from the approximate coupled cluster singles and doubles CC2 linear response theory using Cholesky decompositions. *The Journal of Chemical Physics*, 120(19):8887–8897, May 2004.

Elizabeth Percival. The polysaccharides of green, red and brown seaweeds: Their basic structure, biosynthesis and function. *British Phycological Journal*, 14(2):103–117, June 1979.

Ser John Lynon P. Perez and Gil C. Claudio. Molecular dynamics simulations of two double-helical hexamer fragments of iota-carrageenan in aqueous solution. *Journal of Molecular Graphics and Modelling*, 98:1–7, July 2020.

L Piculell. Organisation and association of κ-carrageenan helices under different salt conditions. *International Journal of Biological Macromolecules*, 21:141–153, 1997.

Sander Pronk, Szilárd Pall, Roland Schulz, Per Larsson, Par Bjelkmar, Rossen Apostolov, Michael R. Shirts, Jeremy C. Smith, Peter M. Kasson, David van der Spoel, Berk Hess, and Erik Lindahl. Gromacs 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):43–56, 2 2013.

P. N. Pusey. Introduction to scattering experiments. In Peter Lindner and Thomas Zemb, editors, *Neutrons, X-rays and light : scattering methods applied to soft condensed matter*. North-Holland, 2002.

D A Rees. Shapely Polysaccharides. The Eighth Colworth Medal Lecture. *Biochemical Journal*, 126(2):257–273, January 1972.

D. A. Rees, W. E. Scott, and F. B. Williamson. Correlation of Optical Activity with Polysaccharide Conformation. *Nature*, 227(5256):390–392, July 1970.

David A. Rees, Frank B. Williamson, Sarah A. Frangou, and Edwin R. Morris. Fragmentation and Modification of ι-Carrageenan and Characterisation of the Polysaccharide Order-Disorder Transition in Solution. *European Journal of Biochemistry*, 122(1):71–79, February 1982.

David N Reshef, Yakir A Reshef, Hilary K Finucane, Sharon R Grossman, Gilean McVean, Peter J Turnbaugh, Eric S Lander, Michael Mitzenmacher, and Pardis C Sabeti. Detecting Novel Associations in Large Data Sets. *Science*, 334:1518–1524, December 2011.

C Rochas and S Landry. Molecular Organization of Kappa Carrageenan in Aqueous Solution. *Carbohydrate Polymers*, 7:435–447, 1987.

C. Rochas and M. Rinaudo. Activity coefficients of counterions and conformation in kappa-carrageenan systems. *Biopolymers*, 19(9):1675–1687, September 1980.

Cyrille Rochas, Marguerite Rinaudo, and Marc Vincendon. Spectroscopic characterization and conformation of oligo kappa carrageenans. *International Journal of Biological Macromolecules*, 5(2):111–115, April 1983.

C. C. J. Roothaan. New Developments in Molecular Orbital Theory. *Reviews of Modern Physics*, 23(2):69–89, April 1951.

L. Rosenfeld. Quantenmechanische theorie der natürlichen optischen aktivität von flüssigkeiten und gasen. *Zeitschrift für Physik*, 52:161–174, 1928.

Michael Rubinstein and Ralph H Colby. *Polymer Physics.* Oxford University Press, 2003.

Steven E. Schafer and Eugene S. Stevens. A reexamination of the double-helix model for agarose gels using optical rotation. *Biopolymers*, 36(1):103–108, July 1995.

Steven E. Schafer and Eugene S. Stevens. The optical rotation of ordered carrageenans. *Carbohydrate Polymers*, 31(1-2):19–22, September 1996.

Larissa Schefer, Jozef Adamcik, Michael Diener, and Raffaele Mezzenga. Supramolecular chiral self-assembly and supercoiling behavior of carrageenans at varying salt conditions. *Nanoscale*, 7(39):16182–16188, 2015a.

Larissa Schefer, Ivan Usov, and Raffaele Mezzenga. Anomalous Stiffening and Ion-Induced Coil–Helix Transition of Carrageenans under Monovalent Salt Conditions. *Biomacromolecules*, 16(3):985–991, March 2015b.

M G Semenova, I G Plashchina, E E Braudo, and V B Tolstoguzov. Structure Formation in Sodium κ-Carrageenan Solutions. *Carbohydrate Polymers*, 9:133–145, 1988.

C Slootmaekers, C De Jonghe, H Reynaers, F A Varkevisser, and C J Bloys va Treslong. Static light scattering from κ-carrageenan solutions. *Int. J. Biol. Macromol.*, 10:160–168, 1987.

C Slootmaekers, C De Jonghe, H Reynaers, F A Varkevisser, and C J Bloys va Treslong. Static light scattering from κ-carrageenan solutions. *Int. J. Biol. Macromol.*, 10:160–168, 1988.

O Smidsrød and H Grasdalen. *Conformations of κ-carrageenan in solution.* Springer Netherlands, Dordrecht, 1984.

Olav Smidsrød, Inger-lill Andresen, Hans Grasdalen, Bjørn Larsen, and Terence Painter. Evidence for a salt-promoted "freeze-out" of linkage conformations in carrageenans as a prerequisite for gel-formation. *Carbohydrate Research*, 80(1):C11–C16, 1980.

Norman F. Stanley. Carrageenans. In Peter Harris, editor, *Food Gels*, pages 79,119. Springer Netherlands, Dordrecht, 1990.

Eugene S. Stevens. Solution conformation of maltose from optical rotation: A procedure for evaluating carbohydrate force fields. *Biopolymers*, 32(11):1571–1579, November 1992.

Eugene S. Stevens and Christopher A. Duda. Solution conformation of sucrose from optical rotation. *Journal of the American Chemical Society*, 113(23):8622–8627, November 1991.

Eugene S. Stevens and Banlacore K. Sathyanarayana. A Semiempirical Theory of the Optical Activity of Saccharides. *Carbohydrate Research*, 166:181–193, 1987.

Eugene S. Stevens, Bangalore K. Sathyanarayana, and Edwin R. Morris. Optical rotatory dispersion of saccharides: testing a theory. *The Journal of Physical Chemistry*, 93(9):3434–3436, May 1989.

G.G. Stokes. On the composition and resolution of streams of polarized light from different sources. *Transactions of the Cambridge Philosophical Society*, 9:399, 1852.

Carlos A Stortz. Potential energy surfaces of carrageenan models: carrabiose, β-(1$\rightarrow$4)-linked D-galactobiose, and their sulfated derivatives. *Carbohydrate Research*, 337: 2311–2323, 2002.

Carlos A. Stortz and Alberto S. Cerezo. The ^{13}C NMR spectroscopy of carrageenans: calculation of chemical shifts and computer-aided structural determination. *Carbohydrate Polymers*, 18(4):237–242, January 1992.

Carlos A. Stortz and Alberto S. Cerezo. Conformational Analysis of Neocarrabiose and its Sulfated and/or Pyruvylated Derivatives using the MM3 Force-Field. *Journal of Carbohydrate Chemistry*, 19(9):1115–1130, January 2000.

Gert Stroble. *The Physics of Polymers*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

Makoto Takemasa and Katsuyoshi Nishinari. 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(3-4):231–240, November 2004.

Gareth A. Tribello, Massimiliano Bonomi, Davide Branduardi, Carlo Camilloni, and Giovanni Bussi. PLUMED 2: New feathers for an old bird. *Computer Physics Communications*, 185(2):604–613, February 2014.

Tristan Turquois, Cyrille Rochas, Francois-R Taravel, Jean Louis Doublier, and Monique-A-V. Axelos. Small-angle x-ray scattering of κ-carrageenan based systems: Sols, gels, and blends with carob galactomannan. *Biopolymers*, 36(5):559–567, November 1995.

K. Ueda, A. Imamura, and J. W. Brady. Molecular Dynamics Simulation of a Double-Helical β-Carrageenan Hexamer Fragment in Water. *The Journal of Physical Chemistry A*, 102(17):2749–2758, April 1998.

Kazuyoshi Ueda, Mariko Saiki, and John W. Brady. Molecular Dynamics Simulation and NMR Study of Aqueous Neocarrabiose 4^1-Sulfate, a Building Block of κ-Carrageenan. *The Journal of Physical Chemistry B*, 105(36):8629–8638, September 2001.

Carsten A. Ullrich. *Time-dependent density-functional theory: concepts and applications*. Oxford graduate texts. Oxford University Press, Oxford New York, reprinted, with corrections edition, 2016.

R Urbani, A Di Blas, and A Cesàro. Conformational features of carrabiose polymers: I. Configurational statistics of κ-carrageenan. *Int. J. Biol. Macromol.*, 15:24–29, 1993.

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, and V Y Grinberg. The structure of κ/ι-hybrid carrageenans II. Coil–helix transition as a function of chain composition. *Carbohydrate Research*, 340:1113–1129, 2005.

Fred van de Velde, Harry S. Rollema, Natalia V. Grinberg, Tatiana V. Burova, Valerij Ya. Grinberg, and R. Hans Tromp. Coil-Helix Transition of ι-Carrageenan as a Function of Chain Regularity. *Biopolymers*, 65(4):299–312, November 2002.

K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry*, 31(4):671–690, 2009.

Christer Viebke, Johan Borgström, Ingemar Carlsson, Lennart Piculell, and Peter Williams. A Differential Scanning Calorimetry Study of κ-Carrageenan in the NaCl/NaI/CsI/CsCl Systems and Analysis by Poisson-Boltzmann Calculations. *Macromolecules*, 31(6):1833–1841, March 1998.

Roy Lester Whistler and James N. BeMiller, editors. *Industrial Gums: Polysaccharides and Their Derivatives*. Academic Press, San Diego, 3rd ed edition, 1993.

Martin A. K. Williams, Valérie Cornuault, Amir H. Irani, V. Vaughan Symonds, Jenny Malmström, Yiran An, Ian M. Sims, Susan M. Carnachan, Christine Sallé, and Helen M. North. Polysaccharide Structures in the Outer Mucilage of *Arabidopsis* Seeds Visualized by AFM. *Biomacromolecules*, 21(4):1450–1459, April 2020.

David E. Woon and Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *The Journal of Chemical Physics*, 98(2):1358–1371, January 1993.

Takeshi Yanai, David P Tew, and Nicholas C Handy. A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chemical Physics Letters*, 393(1-3):51–57, July 2004.

Jen Tsi Yang and Paul Doty. The Optical Rotatory Dispersion of Polypeptides and Proteins in Relation to Configuration. *Journal of the American Chemical Society*, 79(4):761–775, February 1957.

Ramanathan Yegappan, Vignesh Selvaprithiviraj, Sivashanmugam Amirthalingam, and R. Jayakumar. Carrageenan based hydrogels for drug delivery, tissue engineering and wound healing. *Carbohydrate Polymers*, 198:385–400, October 2018.

Jejoong Yoo and Aleksei Aksimentiev. New tricks for old dogs: improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions. *Physical Chemistry Chemical Physics*, 20(13):8432–8449, 2018.

Jerrold H. Zar. Spearman Rank Correlation. In Peter Armitage and Theodore Colton, editors, *Encyclopedia of Biostatistics*, pages 1–6. John Wiley & Sons, Ltd, Chichester, UK, July 2005.

W Zhang, M Liu, P E Hansen, G Yu, B Yang, and X Zhao. Sequence and Structural Analysis of κ-Carrageenan-Derived Oligosaccharides by Two-Dimesional Nuclear Magnetic Resonance. *Journal of Phycology*, 46:831–838, 2010.

Wei Zhang, Lennart Piculell, and Svante Nilsson. Effects of Specific Anion Binding on the Helix-Coil Transition of Lower Charged Carrageenans. NMR Data and Conformational Equilibria Analyzed Within the Poisson-Boltzmann Cell Model. *Macromolecules*, 25(23):6165–6172, 1992.

Kun Zhou and Bo Liu. *Fundamentals of classical molecular dynamics simulation*. Elsevier, 2022.