CHARGED POLYSACCHARIDES AS MODEL POLYELECTROLYTES: COMPUTATIONAL STUDIES OF TRANSPORT AND CONFORMATIONAL PROPERTIES

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.

AMIR HOSSEIN IRANI

2018

Contents

List of Tables

List of Figures

Abstract

Homogalacturonans (HGs) are polysaccharide co-polymers of galacturonic acid and its methylesterified counterpart, that play a crucial role in the mechanobiology of the cell walls of all land plants. When extracted, in solution, at pH values above the pKa, the carboxyl groups carried by the unmethylesterified residues endow the polymer chains with charge, making these systems interesting polyelectrolytes. The inter- and intra-molecular distributions of the non-charged methylesterifed residues and their charged unmethylesterifed counterparts are vital behaviour-determining characteristics of a sample's structure. Previous work has led to the development of techniques for the control of the amount and distribution of charges, and with these tools and samples available in different degrees of polymerisation, including small oligomers, the system offers a flexible test-bed for studying the behaviour of biological polyelectrolytes.

This thesis is rooted in exploring the use of computational approaches, in particular molecular dynamics, to calculate the conformation of such polyelectrolytes in solution and to describe their transport properties in electric fields. The results of simulations are, in all cases, compared with the results of experimental work in order to ground the simulations.

First, in chapter 2, these simulations are applied to calculate the free solution electrophoretic mobilities of galacturonides, charged oligosaccharides derived from digests of partially methylesterified HGs. The simulations are compared with experiment and were found to correctly predict the loss of resolution of electrophoretic mobilities for fully-charged species above a critical degree of polymerisation (DP), and the ionic strength dependence of the electrophoretic mobilities of different partially charged oligosaccharides.

Next, in chapter 3, molecular dynamics (MD) simulations are used to calculate the electrophoretic mobilities of HGs that have different amounts *and distributions* of charges placed along the backbone. The simulations are shown to capture experimental results

well even for samples that possess high charge densities. In addition they illuminate the role that *local* counterion condensation can play in the determination of the electrophoretic mobility of *heterogeneous* blocky polyelectrolytes that cannot be adequately described by a single chain-averaged charge spacing.

Finally, in chapter 4, the last part of the research focusses on the configurations of these polyelelectrolyes in dilute solution, and on how the interactions between several chains can lead to the spatially heterogenous nature of polyelectrolyte solutions. Such questions are of long standing interesting in the polyelectrolyte field and the results are compared with results from Small Angle X-Ray Scattering(SAXS).

Overall the work demonstrates how state of the art MD approaches can provide insights into experimental results obtained from fundamentally interesting and biologically relevent polyelectrolytes.

Acknowledgements

I would like to thank God first to giving me an opportunity to have an experience of living in this beautiful world and have a chance to travel around the globe to see the glories of Mother nature. After that I owe a big thank you to my supervisor Prof Bill Williams not only as a great supportive supervisor but also as a friend. His passion for science and novel ideas helped me to accomplish my PhD project. His efforts brought me to a level of a scientific researcher which will be a stepping stone for my next career. I also acknowledge my co-supervisor Dr Fu-Guang Cao for his helpful advises. A special thanks goes to Dr Davide Mercadante for his help and guidance in computational methods. Dr Bradley W. Mansel is thanked for performing experimental SAXS work. The Institute of Fundamental Sciences is appreciated for the award of a Graduate Assistantship position and financial support. Dr Stephen Keen and all other staff in the physics department are appreciated for their friendly collaborations.

I wish to acknowledge the contribution of NeSI high-performance computing facilities to the results of this project. NZ's national facilities are provided by the NZ eScience Infrastructure and funded jointly by NeSI's collaborator institutions and through the Ministry of Business, Innovation & Employment's Research Infrastructure programme.

Last but not the least, I would like to thank my beloved wife for her supports, cares and all times that I had to spent for her but I used to work on this project. Thank you for your patience and kindness. I would like to thank my parents, my elder brother and younger sister. This project probably wouldn't have been done without all of their supports.

Chapter 1

Introduction

1.1 Polyelectrolytes

Polyelectrolytes are polymers with a charged unit in the repeating chain [7]. They can be totally or partially charged. Upon dissolution charged parts will interact with the solution (if it is polarizable) and with other charged particles in the solution such as ions. There are a lot of polyelectrolyte examples in biological systems, such as polypeptides, glycosaminoglycans, and DNA, where indeed the interaction of charged parts of polyelectrolytes with the present ions and solution itself is a key point of their function [7, 8]. Polyelectrolytes can be categorized as weak or strong, like acids. Strong polyelectrolytes are totally charged at most pH values, while weak polyelectrolytes are partially charged. The fraction of charged residues in weak polyelectrolytes can be controlled by the pH and ion content of the solution.

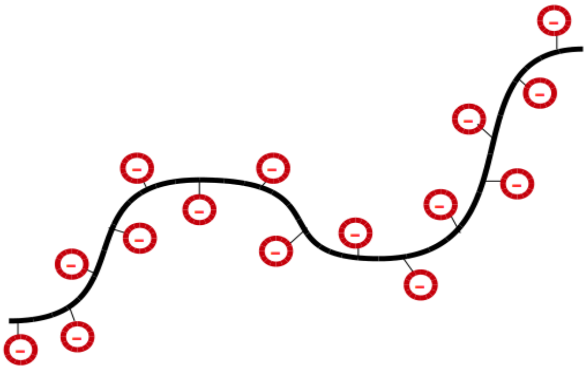


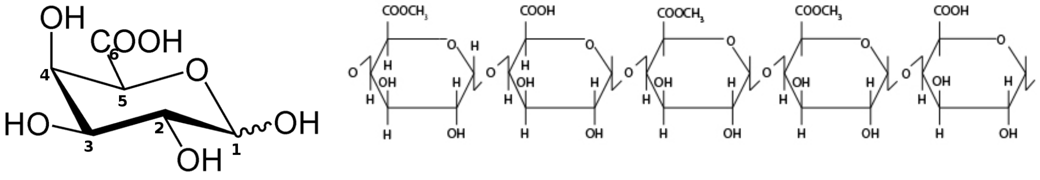
Figure 1.1: Schematic representation of a charged polymer.

Polyelectrolytes have a host of applications [9] amongst which they are used as thickeners, emulsifiers, conditioners, clarifying agents, and even drag reducers [7]. *Pectin* is one biological polyelectrolyte that is widely used in the food, pharmaceutical, cosmetic, and biomedical industries and has a crucial role in-vivo in land plants.

1.1.1 Pectin

Pectin, a macromolecular complex of the plant cell wall, can consist of some 17 different types of sugar residue, and has been called 'probably the most complex macromolecule in nature' [10]. While the details of the arrangements of certain predominant motifs, and the nature and positioning of the attachment of these to each other, and to other elements of the plant cell wall, are not yet settled, it is largely accepted that the structure consists of an assembly of distinct, structurally well-characterized domains [11, 12, 13]. These domains are homogalacturonan(HG), Xylogalacturonan(XGA), rhamnogalacturonan I (RGI), rhamnogalacturonan II (RGII), arabinan, arabinogalactan I (AGI) and arabinogalactan II (AGII).[14]. The most popular model of the overall structure posits that the domains are arranged such that the macromolecular structure has a main backbone consitsing of galacturonic acid(GalA) sugar rings (HG) (fig.1.2a) which are connected to each other by axial 1-4 carbon connections(fig.1.2b). There are some kinds of branches of neutral sugars that may be connected to the rhamnose residues of RG1, also found in the backbone [15, 16, 17]. Pectin has an important role in plant cell walls [18, 19, 20, 21, 22, 23] and in its technological applications as a gelling agent.[24, 25, 26, 27]. Many commercially available pectins are predominantly homogalacturonan (HG) and it is this part of the polymer that is largely responsible for its polyelectrolytic nature and its well known ion-binding properties.

The ubiquitous occurrence of the polysaccharide pectin in the cell walls of land plants, taken together with the large number of pectin-modifying enzymes encoded in the genomes of plants and their pathogens, clearly points to the importance of pectin, and pectin derived compounds, in diverse aspects of plant physiology. Indeed, the products of pectin digestion, the oligogalacturonides, play a key role in triggering plant defence mechanisms [28]. Pectin consists of an assembly of distinct, structurally well-characterised domains [11, 12, 13]. Homogalacturonan (HG), an extensive linear region of 1-4 linked galacturonic acid residues, that can each exist in a state of methylesterification or not, is probably the most studied domain and is most associated with the mechanical functionality of pectin, both in the cell wall [29, 30, 31, 32] and in its technological applications as a gelling agent. The degree of methylesterification (DM)



(a) GalA monomer.

(b) Homogalacturonan chain.

Figure 1.2: Schematic representation of galacturonic acid.

and the patterning of the methylester groups (and thus charged residues) can be expected to play a key role in determining local polymer mechanics [33], in controlling the propensity of the chains to self-assemble into functional networks mediated by ions, hydrogen bonding and hydrophobicity [24, 25, 26, 27, 34], and in mediating the binding of proteins [35, 36, 37, 38].

1.1.2 Homogalacturonan

Homogalacturonan is an extensive linear region of 1-4 linked galacturonic acid residues (figure 1.2), each of which can exist in a state of methylesterification or not. It is probably the most studied domain and is most associated with the mechanical functionality of pectin [29, 30, 31, 32]. Certainly the ability of this portion to assemble in the presence of divalent ions or in acidic conditions is key to its functionality *in-muro*, and typically pectin manufacturers seek to maximise the proportion of HG in commercially available pectins (typically achieving between 60 and 85%). More than half of the mass of pectins in the plant cell walls is homogalacturonan (HG) [39, 40]. In citrus, sugar beet, and apple pectin the minimum estimated length of this HG backbone is about 72-100 GalA residues [41]. The methyl-esterification of galacturonic acid residues effects the physical properties of pectin [42, 43]. The methyl-esterifed GalA residues are uncharged, so the charge fraction on HGs, at a certain pH, can be controlled by modifying the number of methyl-esterifed GalA residues in the chain.

1.1.3 The Importance of Patterning in Biological Polyelectrolytes

There are many examples of how the spatial patterning of substituent groups displayed along biological-polymer backbones is exploited in nature, including the iconic replication of DNA [44]. But the nucelotide base sequence not only stores information in its 1-D sequence, it also determines the local mechanical and assembly properties of the chains, as exemplified by the emergence of such biologically relevant structures as

G-quadruplexes [45, 46] and the exploitation of DNA as a construction material in bionanotechnology [47, 48]. Homogalacturonans offer another example of the importance of the patterning of substituent groups and of how 1-D sequence information can be manifest in 3-D mechanical function. The patterning of the methylester groups (and thus charge) plays a key role in determining local polymer mechanics [33], in controlling the propensity of the chains to self assemble into functional networks mediated by ions, hydrogen bonding and hydrophobicity [24, 25, 26, 27], and in mediating the binding of proteins [35, 36, 37, 38].

1.2 Computational Methods

Nowadays with the incredibly fast growth of computational capability, there is a vast interest in using computational methods to glean insights into natural phenomena. The development of GPUs and the availability of high speed CPUs provides the power to face simulating large molecules and study their behaviour in different conditions in a reasonable time frame. In the mean time we have to keep this in mind that experimental studies can not replaced by the computational simulations. In fact, computational studies help us to have a better understanding and more clear view of what is happening in the molecular and atomic level.

Sometimes solving of an analytical equation that describes a phenomena is straight forward but sometimes it is not. In most cases, there are too many variables and it becomes a complicated problem to solve real problems exactly. In these cases computational methods can often illuminate the salient points and reveal the most important underlying physics.

In this research, we use several computational methods such as quantum mechanical (QM) methods and molecular dynamics (MD) to simulate the conformational and transport behaviour of homogalacturonan in solution. Both an Intel(R) Xeon(R) 3.50GHz desktop computer and also High Performance Computing (HPC) facilities provided by New Zealand eSicence Infrastructure (NeSI) were utilised.

1.3 Quantum Chemistry

Ab initio calculation, (calculation from the beginning or from first principles), is a powerful computational method based on Quantum Mechanical knowledge to obtain the electronic properties of matter. The importance of this method led to the awarding of the Nobel Prize in chemistry in 1998 to Walter Kohn for his development of Density Functional Theory(DFT) and John Pople for his development of computational methods in Quantum Chemistry.

In this method, the main equation is the time-independent Schrödinger equation:

$$\widehat{H}\psi = E\psi \tag{1.1}$$

where ψ is the wavefunction, $\widehat{H}$ is the Hamiltonian operator and E is the total energy of the system. Hamiltonian operator($\widehat{H}$) can be written in this form,

$$\widehat{H} = \widehat{T} + \widehat{V} \tag{1.2}$$

where $\widehat{T}$ is the kinetic energy operator and $\widehat{V}$ is the potential operator. The kinetic energy can be written as

$$\widehat{T} = \widehat{T}_e + \widehat{T}_n \tag{1.3}$$

where $\widehat{T}_e = -\sum\limits_i \frac{\hbar^2}{2m_e}\nabla^2_{\mathbf{r}_i}$ is the kinetic energy of electrons and $\widehat{T}_n = -\sum\limits_i \frac{\hbar^2}{2M_i}\nabla^2_{\mathbf{R}_i}$ is the kinetic energy of the nucleus. The Laplace operator is:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

The potential operator($\widehat{V}$) has three components:

$$\widehat{V} = \widehat{V}_{en} + \widehat{V}_{ee} + \widehat{V}_{nn} \tag{1.4}$$

where $\widehat{V}_{en}$ is the potential energy operator between electrons and nuclei, $\widehat{V}_{ee}$ is the potential energy operator for Coulombic electron-electron repulsion and $\widehat{V}_{nn}$ is the potential energy operator for Coulombic nuclei-nuclei repulsion, which are:

$$\begin{aligned}
\widehat{V}_{en} &= -\sum_i \sum_j \frac{Z_i e^2}{4\pi\epsilon_0 |\mathbf{R}_i - \mathbf{r}_j|} \\
\widehat{V}_{ee} &= \frac{1}{2}\sum_i \sum_{j\neq i} \frac{e^2}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{r}_j|} = \sum_i \sum_{j>i} \frac{e^2}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{r}_j|} \\
\widehat{V}_{nn} &= \frac{1}{2}\sum_i \sum_{j\neq i} \frac{Z_i Z_j e^2}{4\pi\epsilon_0 |\mathbf{R}_i - \mathbf{R}_j|} = \sum_i \sum_{j>i} \frac{Z_i Z_j e^2}{4\pi\epsilon_0 |\mathbf{R}_i - \mathbf{R}_j|}
\end{aligned} \tag{1.5}$$

Solving Schrödinger's equation analytically is not easy for complicated systems. In fact if the system is more complicated than the hydrogen atom or H_2, approximations are required to solve the system. Moreover, we have to keep in mind that in our

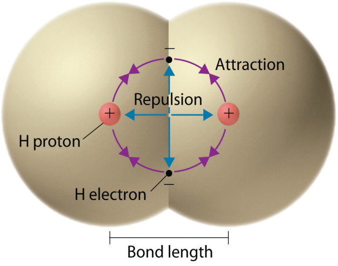


Figure 1.3: H_2 molecule.

approximations relativity effects are neglected.

1.3.1 Hartree-Fock theory

Hartree-Fock theory[49] is the basis of Molecular Orbit (MO) theory and in this theory there are several approximations such as the Born-Oppenheimer approximation[50] and that the electron's motion can be described in a way which is independent from other electron's motions. This theory considers molecular orbitals as a linear combination of atomic orbitals.

1.3.1.1 The Born-Oppenheimer approximation

In this approximation the kinetic energy of nuclei is ignored($\widehat{T}_e = 0$). This is justified because the nuclei's mass is ≈ 2000 times of electron's mass, so the movement of nuclei is much less than electron's movement. It is possible to consider that the nuclei has a fixed position and $\widehat{V}_{nn}$ is just a constant. With this assumption we can rewrite the wave equation as follow:

$$\psi(\mathbf{R}, \mathbf{r}) = \psi_n(\mathbf{R})\psi_e(\mathbf{r}; \mathbf{R}) \tag{1.6}$$

where $\mathbf{R}$ determines the coordinate of nucleus and $\mathbf{r}$ determines the coordinate of electrons around the nucleus and the ";" means this dependence is just a parametric dependence.

Here to have a better view of Hartree-Fock theory, the H_2 molecule is considered (Fig 1.3). The total wavefunction $\psi = \psi(\mathbf{R}_1, \mathbf{R}_2, \mathbf{r}_1, \mathbf{r}_2)$ can be written, considering of Born-Oppenheimer approximation, as

$$\psi = \psi_n(\mathbf{R}_1, \mathbf{R}_2)\psi_e(\mathbf{r}_1, \mathbf{r}_2) \tag{1.7}$$

In this model particles are independent and so it is possible to write the total wavefunction as a Hartree product: $\psi(\mathbf{r}_1, \mathbf{r}_2) = \varphi_1(\mathbf{r}_1)\varphi_2(\mathbf{r}_2)$ where each φ_i is a molecular orbital. But there is a problem here, we know that the total wavefunction must be antisymmetric with respect to the interchange of any set of space-spin coordinates based on the Pauli principle.

In the following we will change our notation to the spinorbital $(\varphi(\mathbf{x}))$, where $\mathbf{x} = \{\mathbf{r}, \omega\}$ is the set of space-spin coordinates and ω is the spin coordinate. To have an antisymetric wavefunction we have to add all signed permutations;

$$\psi(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{\sqrt{2}}[\varphi_1(\mathbf{x}_1)\varphi_2(\mathbf{x}_2) - \varphi_1(\mathbf{x}_2)\varphi_2(\mathbf{x}_1)] \tag{1.8}$$

This antisymmetrized wavefunction is called the Hartree-Fock wavefunction and can also can be written as a determinant:

$$\psi(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{\sqrt{2}}\begin{vmatrix} \varphi_1(\mathbf{x}_1) & \varphi_2(\mathbf{x}_1) \\ \varphi_1(\mathbf{x}_2) & \varphi_2(\mathbf{x}_2) \end{vmatrix} \tag{1.9}$$

If we have more than 2 electrons, we can generalize this determinant for N electrons, which is called a Slater determinant[51]:

$$\psi = |\psi_1, \psi_2, \cdots \psi_N >= \frac{1}{\sqrt{N!}}\begin{vmatrix} \varphi_1(\mathbf{x}_1) & \varphi_2(\mathbf{x}_1) & \cdots & \varphi_N(\mathbf{x}_1) \\ \varphi_1(\mathbf{x}_2) & \varphi_2(\mathbf{x}_2) & \cdots & \varphi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_1(\mathbf{x}_N) & \varphi_2(\mathbf{x}_N) & \cdots & \varphi_N(\mathbf{x}_N) \end{vmatrix} \tag{1.10}$$

This is an antisymmetric wavefunction and it is a combination of one-electron Molecular Orbitals(MO)[1]. Now, each one-electron wavefunction in the molecule is a combination of atomic orbitals and can be written as:

$$\varphi_i(\mathbf{x}_i) = \sum_\mu C_{\mu i} X_\mu(\mathbf{x}_i) \tag{1.11}$$

where X_μ is the atomic orbital and $C_{\mu i}$ is the molecular orbital coefficient. For a normalized wavefunction we can write the expectation value for the energy of the system as:

$$E =< \psi|\widehat{H}|\psi > \tag{1.12}$$

In the HF method this energy is: $E_{HF} = \sum\limits_i H_i + \frac{1}{2}\sum\limits_{ij}(J_{ij} - K_{ij})$. This has two

---

[1] Molecular Orbital(MO) is a mathematical function describing an electron in a molecule which has a wave-like behaviour. The probability of finding an electron and some other chemical and physical properties can calculated by using of this function.

parts: first part(H_i) is for one-electron terms and the second one(J_{ij} and K_{ij}) is for two-electron terms. H_i is from the kinetic energy of the electrons and the nuclear attraction. J_{ij} is the Coulomb repulsion between the electrons and K_{ij} is the exchange energy. By definition

$$\begin{aligned} H_i &=< \varphi_i|\widehat{H}_{1e}|\varphi_i > \\ &= \sum_{\mu\nu} C_{\mu i}C_{\nu i} < X_\mu|\widehat{H}_{1e}|X_\nu > \end{aligned} \tag{1.13}$$

It is possible to carry out this calculation to find out the expectation values for the J_{ij} and K_{ij} and find the MO coefficients.

A theorem called the Variational Theorem can help us to find the MO coefficients. The **Variational Theorem** states that: *The energy determined from any approximate wavefunction will always be greater than the energy for the exact wavefunction.* So, to find the exact wavefunction we have to find the minimum energy of the system.

To solve the HF equation computationally, there is a algorithm called "the Self-Consistent Field" method that can help to find the answer. The Self-Consistent Field(SCF) process is as follow:

1- Guess a set of MOs, $C_{\mu i}$
2- Use MOs to compute H_i, J_{ij} and K_{ij}
3- Solve the HF equations for the energy and new MOs
4- Are the new MOs different?(if Yes go to (2), if No go to (5))
5- Self-consistent field converged

1.3.2 Density Functional Theory

In the HF method the wavefunction is the basic quantity used to find the ground state energy. It can be complicated to find a wavefunction for a molecule with several atoms and n electrons. Consequently, methods which work with wavefunctions such as HF method are not sufficient for attacking large molecules. To overcome this problem Density Functional Theory(DFT) works with the electron density($\rho(\mathbf{r})$) instead of wavefunction. Regardless of how many electrons exist in the system, the density is a 3 dimensional parameter[52].

The electron density in $\mathbf{r}_1$ is given by:

$$\rho(\mathbf{r}_1) = \int \cdots \int \psi^*(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N)\psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N) d\mathbf{r}_2 \cdots d\mathbf{r}_N \tag{1.14}$$

and we know that $\rho(\mathbf{r})d\mathbf{r}$ gives us the probability of finding an electron in the volume element $d\mathbf{r}$.

To understand Density Functional Theory better we have to know meaning of "functional". A function takes a number and returns another number, for example: if $f(x) = 2x^2$ therefore $f(3) = 18$. An operator takes a function and returns another function, for example: $\widehat{O}(f) = \frac{df}{dx}$, therefore $\widehat{O}(2x^3+5) = 6x^2$. A functional takes a function and returns a number, for example $F[f] = \int_0^1 f(x)dx$ therefore $F[x^2 - 1] = -2/3$.

In 1964 Hohenburg and Kohn provided the two theorems that lead to the basic statements of density functional theory.

The first theorem is:
1-*If two electronic systems have potentials of* $V_1(\boldsymbol{r})$ *and* $V_2(\boldsymbol{r})$ *and also have the same ground-state density then the difference of these two potentials must be constant,*$V_1(\boldsymbol{r}) - V_2(\boldsymbol{r}) = Const.$,
In other word this theorem says that *The electron density determines the external potential (to within an additive constant).*

The second theorem is based on the variational principle;
2-*Any calculated energy based on a trial density function is greater than or equal to the exact ground-state energy.*

The energy functional can be written in three parts. The kinetic energy, the interaction with the external potential and the electron-electron interaction;

$$E[\rho] = T[\rho] + V_{ext}[\rho] + V_{ee}[\rho] \tag{1.15}$$

The interaction with the external potential is $V_{ext} = \int \widehat{V}_{ext}\rho(\mathbf{r})d\mathbf{r}$, but the kinetic and electron-electron interaction parts of the energy functional are unknown.

Kohn and Sham assumed a system of n non-interacting electrons to find an approximation for the kinetic and electron-electron part of energy functional. By this assumption the kinetic energy is;

$$T_s[\rho] = -\frac{1}{2}\sum_i^N < \varphi_i|\nabla^2|\varphi_i > \tag{1.16}$$

For the electron-electron interaction, the classical Coulomb interaction is the major component,so;

$$V_H[\rho] = \frac{1}{2} \int \int \frac{\rho(\mathbf{r}_1 \rho(\mathbf{r}_2))}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \tag{1.17}$$

and then the energy functional looks like:

$$E[\rho] = T_s[\rho] + V_{ext}[\rho] + V_H[\rho] + E_{xc}[\rho] \tag{1.18}$$

where E_{xc} is the exchange-correlation functional.

To solve this equation, an initial guess is used in order to define electron density(ρ) and with this initial density, Kohn-Sham(KS) equations can be solved to get KS orbitals. These calculated orbitals are used to find electron density and again the new electron density can be put in the KS equation. Here also the variation principle is used as a guide to find the ground sate.

1.4 Molecular Dynamics

Molecular Dynamics(MD) is an atomistic simulation technique that can investigate the behaviour of many-atom systems [53]. By sampling of the phase space of a system, it can calculate the statistical parameters and also the dynamical behaviour of the system in equilibrium. In fact, this method aims to explore all microscopic properties of a system by calculation based on the interaction between atoms.

Imagine a system with n atoms $\{\mathbf{r}_i\}, i = 1 \cdots N$ interacting with the interatomic potential $V(\{\mathbf{r}_i\})$. In this case, by solving of Newton's equation of motion we can find the net force exerted on each atom at each site i.

$$m_i \ddot{\mathbf{r}}_i = \mathbf{F}_i \equiv -\frac{\partial V}{\partial \mathbf{r}_i} \tag{1.19}$$

So for many atoms with pairwise interaction potentials, the total interaction potential on site i is;

$$V(\{\mathbf{r}_i\}) = \frac{1}{2} \sum_{i}^{N} \sum_{j \neq i}^{N} V_{LJ}(r_{ij}) \tag{1.20}$$

where $V_{LJ}(r_{ij})$ is the Lennard-Jones potential between pairs of atoms;

$$V_{LJ}(r) = -4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right] \tag{1.21}$$

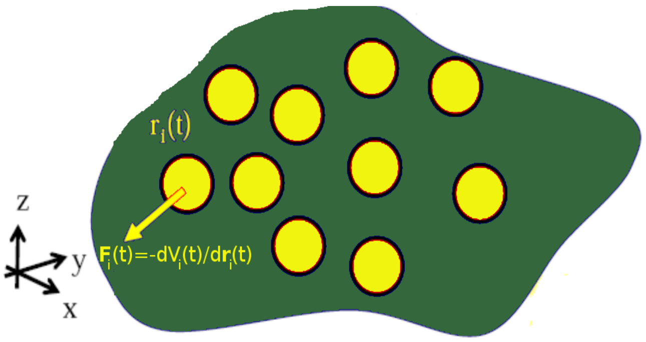


Figure 1.4: Felt force from interatomic potential at each each time.

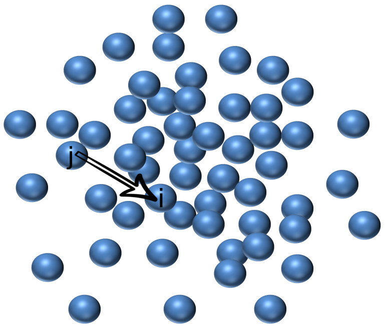


Figure 1.5: A cluster of atoms interact with each others.

Now, by definition of force($\mathbf{F}_i = -\frac{\partial V}{\partial \mathbf{r}_i}$) the force acting on the atom i is;

$$\mathbf{F}_i = -\frac{1}{2}\sum_{i}^{N}\sum_{j\neq i}^{N} \left.\frac{dV_{LJ}(r_{ij})}{dr}\right|_{r=r_{ij}} \frac{\mathbf{r}_{ij}}{r_{ij}} \tag{1.22}$$

Molecular Dynamics is usually used for systems with interactions over the length-scale between 10(Å) and 50-60(Å), and is typically run for time-scales between 1ps and ∼50ns. However, by improving computational hardware MD simulations on larger systems are becoming possible (for example 15-20nm for 100s of ns).At smaller scales Quantum Statistical methods need to be applied (Beyond the scope of this work).

1.4.1 Statistical mechanics in MD

In MD the goal is to understand macroscopic properties of the system under study, such as temperature, pressure and so on. Fortunately these macroscopic properties can be derived from the microscopic parameters of system. The position and velocity of each particle can be used to find out those macroscopic parameters. In this case we have to introduce ”the Phase Space”.

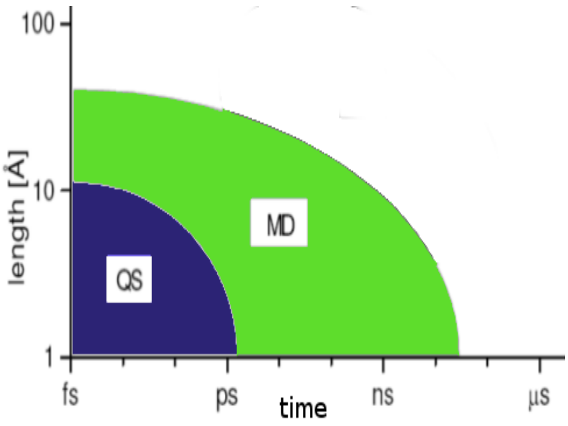


Figure 1.6: MD and QS acting region.

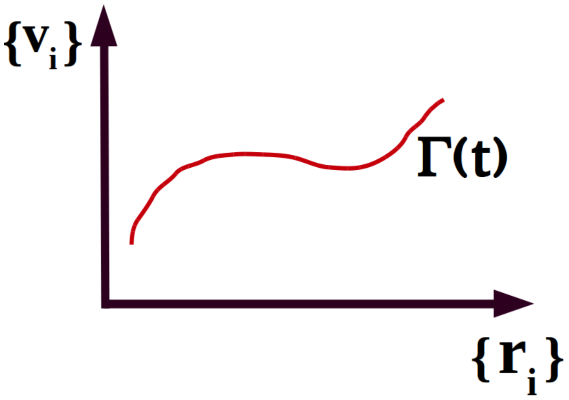


Figure 1.7: Schematic figure of the phase space at time t.

1.4.1.1 Phase Space

The phase space for a system with N particles is the 6N dimensional space of all possible states of the system. When we talk about states it means that we have to include positions and velocities of particles (3N dimensions for the positions and 3N dimensions for the velocities). The coordinates of a system in phase space(Γ) depend on time, with the trajectory in phase space (fig. 1.7) reflecting the dynamics of the system.

$$\Gamma(t) \equiv \{\mathbf{r}_i(t), \mathbf{v}_i(t)\}_{i=1\cdots N}$$

A set of macroscopic parameters are needed to describe the state of a macroscopic system. The macroscopic parameters such as temperature (T), pressure (P), number of particles (N), free energy (G), chemical potential (μ) can be derived from the phase space curve (or better to say 6N dimensional surface)

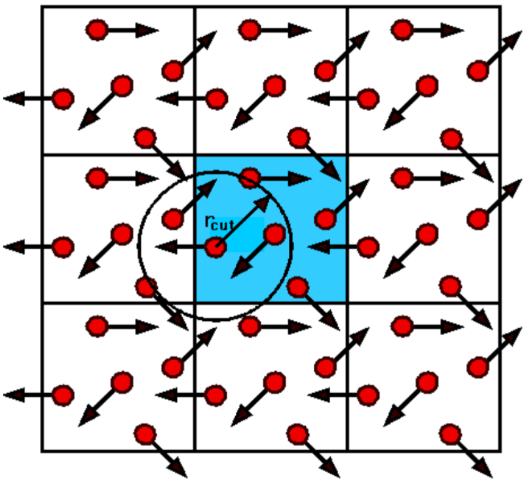


Figure 1.8: An illustration of periodic boundary conditions.

1.4.1.2 Connection to Statistical Mechanics

As we discussed in the previous section, each macroscopic observable quantity, Λ, is dependent on the microstates of the system, so;

$$\Lambda(t) \equiv \Lambda(\Gamma(t)) \tag{1.23}$$

and the average of instantaneous values of the Λ gives us the macroscopic value of the quantity Λ.

$$\overline{\Lambda} \equiv \langle \Lambda(t) \rangle = \lim_{T\to\infty} \frac{1}{T} \int_0^T \Lambda(t)dt \tag{1.24}$$

In ergodic[2] systems, the time average of a observable Λ converges to a macroscopic average of the quantity Λ.

In order to simulate a bulk solid or fluid, instead of considering a huge cell, periodic boundary conditions are applied and if any particles travels out of the simulation box then another one enters the box exactly from the other side. By increasing the distance between atoms, the interaction between them decreases quickly, so it is acceptable to calculate interactions between atoms just for a specific region within a radius called the **cutoff** radius(r_{cut}). By increasing r_{cut} the computation time will increase but a more accurate answer will be achieved. It is important to have a r_{cut} balancing the requirements of calculation time and accuracy.

---

[2] In the ergodic system, the time average of the quantity is the same as average over phase space. That is: the system can explore all the phase space in a finite time

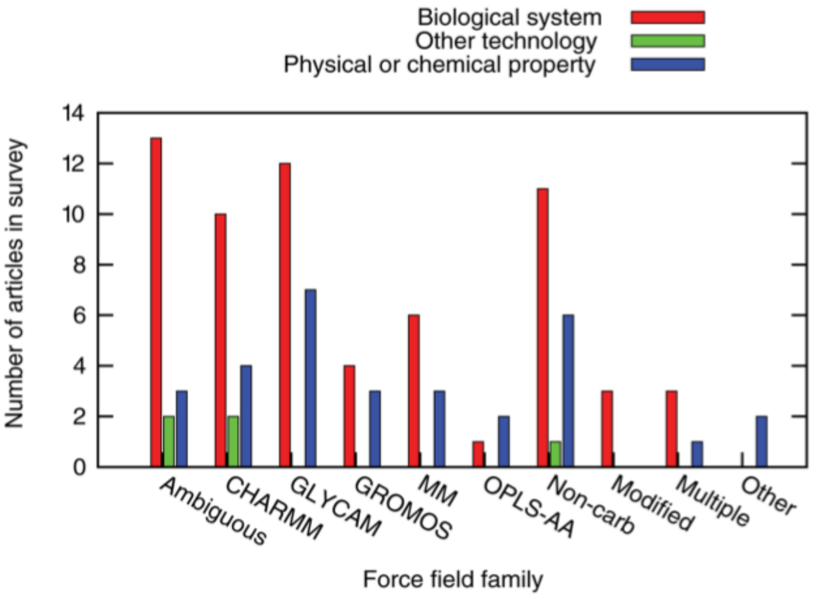


Figure 1.9: Frequency of force fields in different class of studies. Adapted from [1].

1.4.2 MD Algorithm

Here the MD algorithm is introduced by examining an MD simulation step by step. To carry out an MD simulation an initial configuration for all the atoms composing the system under study is required. This configuration can come from experimental data or previous computational works, but basically a so-called pdb file which contains the coordinates and types of atoms must be supplied. Once a pdb file is obtained a force field is required by the MD simulator package. The MD force field, which needs to be chosen carefully, contains all the information about how the potentials between different atoms depend on the length of atomic bonds, angles between bonds, mass of particles and so on. Given the initial coordinates and a mechanism to calculate the force acting on any atom imposed by the other atoms in the simulation all that is required in to start the system in motion and iterate Newton's equations of motion.

There are several available force fields to use in molecular dynamics simulations [54]. To choose an appropriate force field for the system under study it is needed to look at the literatures and find which force field used to investigate which system. Figure 1.9 (adapted from [1]) shows the application of force fields in different systems reported in the literatures. It can be seen that GLYCAM is widely used to study the biological systems and basically this forcefield is developed for carbohydrates base on Amber force field. Therefore, GLYCAM [55] was selected to be the main force field in this project.

The initial velocity of the atoms are selected from a Maxwell-Boltzmann distribution

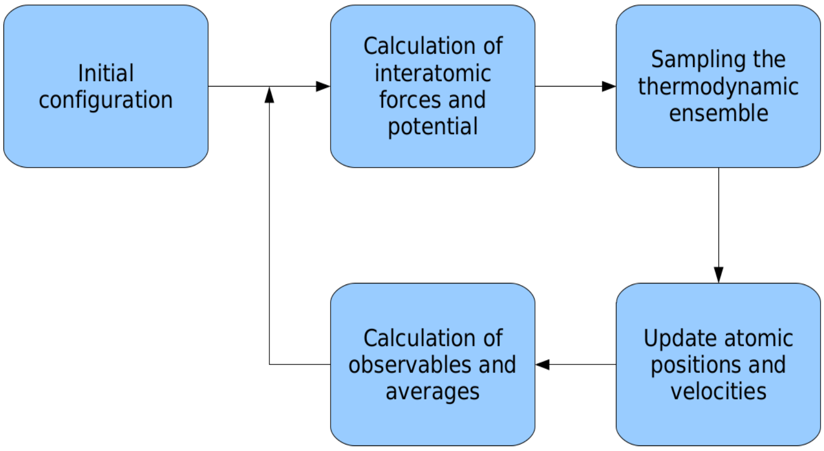


Figure 1.10: MD algorithm.

at the appropriate temperature.

$$P(v_{i,\alpha}) = \left(\frac{m_i}{2\pi k_B T}\right)^{1/2} exp\left[-\frac{m_i v_{i,\alpha}^2}{2k_B T}\right] \qquad \alpha = \{x, y, z\} \tag{1.25}$$

After determining of positions and velocities of all atoms we can calculate interatomic forces. The general form for the potential at atom i is;

$$V(\{\mathbf{r}_i\}) = \frac{1}{2}\sum_{i}^{N}\sum_{j\neq i}^{N} V(r_{ij}) \tag{1.26}$$

and there are several descriptions of interatomic pair interaction potentials that can be used

$$\begin{aligned} V_{LJ}(r) &= -4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \\ V_{Morse}(r) &= D\left[e^{-2\alpha(r-r_0)} - 2e^{-\alpha(r-r_0)}\right] \\ V_{BMH}(r) &= A_{\alpha,\beta}e^{-b_{\alpha,\beta}r} - \frac{Q_\alpha Q_\beta}{r} - \frac{C_{\alpha,\beta}}{r^6} + \frac{D_{\alpha,\beta}}{r^8} \end{aligned} \tag{1.27}$$

(there are also three-body and many-body potentials). Knowing the potential at each atom, it is straight forward to calculate the force by taking the derivative ($F_i(t) = \frac{dV_i(t)}{dr_i}$), and subsequently this allows the positions to be updated.

Choosing the appropriate thermodynamic ensemble relevant for the simulation is also important. The calculated forces in a system which conserves total(kinetic+potential) energy generate trajectories in the **microcanonic** ensemble(NVE) in which the total number of atoms(N), the volume of the system, and also energy of the system are constant. Where energy changes or volume changes are possible during updating of potentials it is important that the system must be coupled to a thermostat and/or to a barostat. The most common techniques are velocity rescaling, the Nose-Hoover thermostat[56] and the Andersen thermostat[57].

Once the thermo- and baro-statting of the ensemble have been considered, the updating of the atomic positions and velocities can be applied. The most commonly used time integration methodology in MD method called the Verlet algorithm that easily updates the position and velocity of atoms. By a Taylor expansion we can write expression for how the co-ordinates at time t are related to those in prior or following time increments;

$$r(t+\Delta t) = r(t) + v(t)\Delta t + (1/2)a\Delta t^2 + (1/6)b(t)\Delta t^3 + O(\Delta t^4)$$

$$r(t-\Delta t) = r(t) - v(t)\Delta t + (1/2)a\Delta t^2 - (1/6)b(t)\Delta t^3 + O(\Delta t^4)$$

and by adding these two expressions;

$$r(t+\Delta t) = 2r(t) - r(t-\Delta t) + a\Delta t^2 + O(\Delta t^4) \tag{1.28}$$

The Taylor expansion for the velocity is $v_i(t+\Delta t) = v_i(t) + \frac{\Delta t}{2m_i}[f_i(t) + f_i(t+\Delta t)]$, and using the same methodology for the velocity as for the position gives:

$$v_i(t) = \frac{r_i(t+\Delta t) - r_i(t-\Delta t)}{2\Delta t} + O(\Delta t^2) \tag{1.29}$$

Now information about positions and velocities of atoms and how they are updated throughout the course of the simulation have been described. At this stage the MD method can calculate other observables such as, for example, instantaneous temperature and instantaneous pressure. In fact, by knowing of position and velocity of atoms it is possible to figure out the kinetic energy ($K = \frac{1}{2}\sum_i m_i|\mathbf{v}_i|^2$) and potential energy ($V = V(\{\mathbf{r}_i\})$) and finally the total energy $E = K + V$

$$\begin{aligned} T_i &= \frac{2K}{N_f K_B} \\ P_i &= \frac{1}{3} Tr\Pi \end{aligned} \tag{1.30}$$

where Π is the stress tensor. After these calculations, the MD simulation again calculates the interatomic potentials and forces using the updated set of positions. This loop will continue for the number of defined time-steps.

1.4.3 Water Molecules in MD

Water, the most important solvent, has a key role in the biological studies. With only three atoms, it looks really simple molecules, however because of its high impact on the most natural studies there are good number of investigations and models about water molecules [58, 59, 60]. In molecular dynamics studies it is important to chose an appropriate water molecule depends on needs of study. Water molecules can be considered as rigid or flexible molecules which have different computational cost regarding to complexity of molecules. Flexible water molecule models have more accurate results and are more reliable, however they dramatically increase the computational calculations as the most of molecules in the simulation box are water molecules. For some specific research fields, it is necessary to use flexible models for water molecules, for example at high temperature and high pressure [61].

In contrast to the flexible models of water molecules, there are several rigid water molecules models that can be used at atmospheric pressure and room temperature with a good agreement with experiment. Using rigid molecules reduce a high amount of calculations and therefore reduce the simulation time very well. In this research, as there was no need of extreme conditions of temperature and pressure, it is decided to use rigid model for water molecules. There are several rigid models for water molecules with different parameters in terms of bond lengths, atomic partial charges and angle between atoms. Table 1.1 shows the parameters considered in different rigid water model. In this research SPC/E (extended simple point charge) [62, 63] model for water molecule has been used because of its simplicity and accurate results.

1.5 Poisson-Boltzmann Equation

When thinking about the behaviour of polyelectrolytes in ionic solutions the Poisson-Boltzmann Equation gives a convenient starting point. The Poisson equation gives the potential (Ψ) as a function of the volume charge density(ρ) at any point in an electrolyte solution to be:

$$\nabla^2\Psi = -\frac{\rho}{\epsilon\epsilon_0} \tag{1.31}$$

	TIPS [64]	SPC [65]	TIP3P [66]	SPC/E [62]
r_0^{OH}(Å)	0.9572	1.0	0.9572	1.0
θ_0^{HOH}(deg)	104.52	109.47	104.52	109.47
q^O (e)	0.80	0.82	0.834	0.8476
q^H (e)	+0.40	+0.41	+0.417	+0.4238
C_{OO}^{12}(kcal Å^{12}/mol)	580.0	629.4	582.0	109.47
C_{OO}^{6}(kcal Å^{6}/mol)	525.0	625.5	595.0	625.5

r_0^{OH}: Equilibrium distance between Oxygen and Hydrogen.
θ_0^{HOH}: Equilibrium angle for H-O-H.
q^O: Partial charge of Oxygen.
q^H: Partial charge of Hydrogen.
C_{OO}^{12} & C_{OO}^{12}: Lennard-Jones parameters for the interaction between the oxygen atoms.

Table 1.1: Parameters for some water molecule models.

where the ϵ is the dielectric constant of the medium and ρ is the density of counter-ions and co-ions in the medium. The ions will have a Boltzmann distribution in the potential $\Psi(r)$. The Boltzmann distribution gives the probability of finding a particle (P) at the distance r as proportional to:

$$P \propto e^{-\frac{E(r)}{K_B T}} \tag{1.32}$$

By this definition, in a solution that contains n_s molecules of salt, the number of counter-ions with charge Z_1 at distance r from the origin is given by :

$$\begin{gathered} n_1 = n_s e^{\frac{Z_1 e \Psi(r)}{k_B T}} \\ (E(r) = Z e \Psi(r)) \end{gathered} \tag{1.33}$$

since Z_1 and Ψ are of opposite signs, and the number of co-ions(n_2)is:

$$n_2 = n_s e^{-\frac{Z_2 e \Psi(r)}{k_B T}} \tag{1.34}$$

and Z_2 and Ψ have the same sign. Therefore, since $\rho = -n_1 e + n_2 e$ then;

$$\rho(r) = n_s e \left[- e^{\frac{Z_1 e \Psi(r)}{k_B T}} + e^{-\frac{Z_2 e \Psi(r)}{k_B T}} \right] \tag{1.35}$$

and with a 1:1 salt, $Z_1 = Z_2 = 1$, So;

$$\rho(r) = -2 n_s e \sinh\left(\frac{e \Psi(r)}{k_B T} \right) \tag{1.36}$$

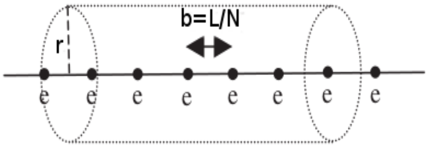


Figure 1.11: Schematic figure of a charge line with uniform charge distribution[2].

and by substitution of equation(1.36) in the equation(1.31), the Poisson-Boltzmann equation is arrived at;

$$\nabla^2\Psi = \frac{2n_s e}{\epsilon\epsilon_0}\sinh\left(\frac{e\Psi}{k_B T}\right) \tag{1.37}$$

This equation can only be solved analytically for symmetric systems such as plane or charged line with the uniform charge distribution. Next the solution for mobile charges around a charged line is investigated in order to provide some insight into the behaviour of linear polyelectrolytes.

1.5.1 Counter-ion Atmosphere Around a Charged Line

A uniform charged line is a first approximation to a charged polymer. It is a line with length (L) covered by the uniform distribution of N charged points with distance $b = \frac{L}{N}$ (Fig 1.11) between them. The electric field(E) around this charge line in the continuum solvent with a uniform dielectric constant(ϵ) can be calculated by the Gauss' theorem;

$$E(r) = \frac{-e/b}{2\pi\epsilon\epsilon_0 r} \tag{1.38}$$

and the potential(Ψ) is given by;

$$\Psi(r) = -\int E dr \tag{1.39}$$

so,

$$\Psi(r) = A + \frac{e}{2\pi\epsilon\epsilon_0 br}\ln(r) \tag{1.40}$$

The potential energy($E_p(r)$) for a monovalent counterion with the unit charge of e at the distance r from the charge line is:

$$E_p(r) = e\Psi(r) \tag{1.41}$$

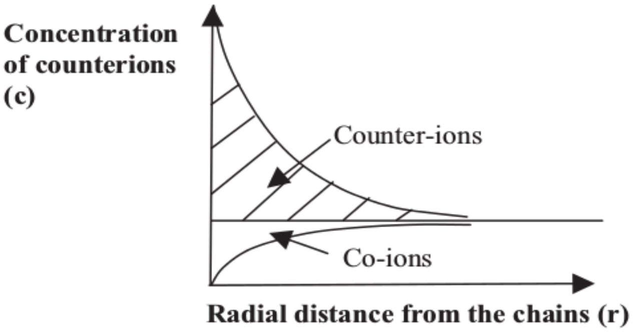


Figure 1.12: Radial distribution of the concentration of counterions and co-ions around a charge line in the solution[2].

and with this potential energy for the counterions in the medium, the Boltzmann distribution of counterions(D) can be written as:

$$D = e^{-\frac{E_p}{k_B T}} = D_0 r^{-\frac{e^2}{2\pi\epsilon\epsilon_0 b k_B T}} \tag{1.42}$$

where $D_0 = e^{-\frac{eA}{k_B T}}$. To simplify of this equation, it is possible to rewrite it as a function of charge parameter(ξ), where;

$$\xi = \frac{l_B}{b} \tag{1.43}$$

and l_B is the Bjerrum length which defines a distance between two opposite charges in an ion pair such that the electrostatic energy of attraction is equal to $k_B T$, therefore the Bjerrum length(l_B) is:

$$l_B = \frac{e^2}{4\pi\epsilon\epsilon_0 k_B T} \tag{1.44}$$

So, by these definitions, equation(1.42) can be rewritten as:

$$D = D_0 r^{-2\xi} \tag{1.45}$$

The number of counterions inside a cylinder with the radius r_0 is then proportional to the integral:

$$\int_0^{r_0} D_0 r^{-2\xi} 2\pi r dr = 2\pi D_0 \int_0^{r_0} r^{1-2\xi} dr \tag{1.46}$$

It has to be noted that if $\xi > 1$ then this integral diverges at the origin($r = 0$) which is typically interpreted as an unstable situation.

1.5.2 Counter-ion Condensation

According to equation(1.40 and 1.41), the potential energy of counter-ions in the solution is:$E_p(r) = eA + \frac{2l_B}{b}\ln(r)$. Now suppose the counter-ion tries to escape from the charge line moving from distance, r_1, to a larger distance from the charge line, r_2. The difference between potential energy of these two point that the counter-ion has to pay to move is:

$$\Delta E_p = E_p(r_2) - E_p(r_1) = \frac{2l_B}{b}\ln(\frac{r_2}{r_1}) \tag{1.47}$$

The entropy of a single ion in volume V is $S = k_B N \ln(\frac{V}{N\lambda_T^3}) + \frac{5}{2}$ where $\lambda_T = (\frac{3Nh^2}{4\pi mU})^{1/2}$. When the ion (N=1) moves from the smaller cylindrical region to the larger region the change in entropy is:

$$\Delta S = S(r_2) - S(r_1) = k_B \ln\left(\frac{r_2^2}{r_1^2}\right) = 2k_B \ln(\frac{r_2}{r_1}) \tag{1.48}$$

So, finally the change in the free energy will be:

$$\frac{\Delta F}{k_B T} = \Delta E_p - \frac{\Delta S}{k_B} = 2\left(\frac{l_B}{b} - 1\right)\ln(\frac{r_2}{r_1}) \tag{1.49}$$

Here, two possible behaviours emerge, one for a line of low charge density ($b > l_B$) and one for a line of high charge density line($b < l_B$). For the low density case, $\Delta F < 0$ and the counter ion is free to move, but for the higher charge density, $\Delta F > 0$ and the free-energy cost is much higher than that gained through an entropy increase. In this case the the counter-ion is essentially condensed on the line, always staying in close proximity.

To reiterate, if $\xi > 1$ then some counter-ions in a polyelectrolyte solution will be trapped in the vicinity of charge line and a condensed layer of counter-ions will be generated around the charge line, called "counter-ion condensation". These counter-ions reduce the effective charge density of the charge line until the density of charge on the line reduces from e/b to e/l_b.

According to the above, when ξ is greater than one(high density), the effective charge felt by the any further un-condensed counter-ions in the solution is e/l_B. So, that an amount of $(e/b - e/l_B)$ is neutralised per length. Therefore, the fraction of the neutralised charges on the charged line (f_N) is:

$$f_N = \frac{e/b - e/l_B}{e/b} = 1 - \frac{b}{l_B} = 1 - \frac{1}{\xi} \tag{1.50}$$

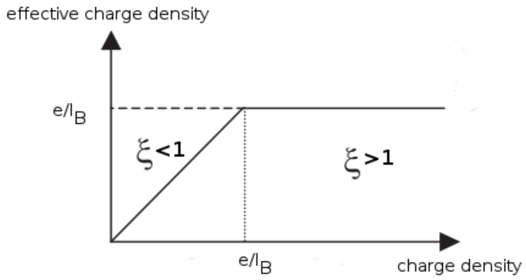


Figure 1.13: effective charge density felt by the counter-ions as a function of charge density of line[2].

or in the multivalent case;

$$f_N = 1 - \frac{1}{Z\xi} \tag{1.51}$$

For example, for DNA with $b = 0.33nm$, $l_B = 0.7nm$ and $Z = 2$ about 76% of the DNA charges are neutralised.

1.6 Aim of the thesis

In this thesis, by performing MD simulations (armed with the partial charges calculated by DFT), we aim to:

- Study the electrophoretic mobilities of oligosaccharides with different charge patterning, providing insights into the experimental dependences on length and ionic strength .
- Study the effect of counter-ions in modifying the electrophoretic mobilities of charged polysaccharides, investigating the possibility that counterion condensation can explain experimental results on HGs.
- Investigate the effect of patterning (blockiness or randomness of the charge distribution) along the chain on the ionic environment around the chain and its electrophoretic mobility.
- Study the conformation of single chains and the arrangement of multiple chains in solutions of differing ionic strength, and compare the results with SAXS experiments.

Chapter 2

Transport properties of Oligogalacturonides

2.1 Introduction

Biopolymers are crucial components of biological systems, and are all, at some point, disassembled, typically during recycling, remodelling, or digestion, either in-vivo or in a chemical processing plant. Resulting fragments not only play key physiological roles in their own right but also contain information regarding the structure of the pre-digested substrate at the time of processing. Particularly for polysaccharides, which are routinely remodelled after biosynthesis, and where the molecular biology tools that have so advanced nucleotide and protein sequencing are simply not available, using in-vitro digestion and picking up the pieces generated by a controlled dis-assembly offers hope as a useful tool in polymeric fine structure determination.

Both for understanding the behaviour and biological function of oligosaccharides, and in order to exploit fragmentation as a tool for polysaccharide fine structure elucidation, the detailed measurement of the structure of such pieces is a pre-requisite. Some progress has been made using several different chromatographic methods in an attempt to separate and identify different, but similar, oligosaccharide species. However, the lack of a quantitative predictive framework for the transport behaviour of different oligomers in the presence of different external fields and with different solvent conditions and interacting stationary phases, has hampered the rapid assignment of species. Herein, we report the results of molecular dynamics (MD) simulations that have been performed in order to calculate the free solution electrophoretic mobilities of oligogalacturonides, and compare these with the existing results of experiments carried out in our research group using capillary electrophoresis (CE) .

For DPs of larger than around 25 the symmetrical scaling of charge and hydrodynamic friction coefficient means that pectin and HG chains, regardless of their DP, elute according to their average charge density[67]. For high charge densities it should be noted that this electrical-transport-determining charge might be modified from that which would be naively calculated from the structure, by counterion condensation [3]. Nevertheless, this independence of electrophoretic mobilty on DP for HGs means that each migration time marks species with a unique degree of methylesterification and peak shapes thus reflect the intermolecular methylesterification distribution (the distribution of degree of methylesterification among chains) of the sample [68]. However, it has been predicted theoretically [69], and previously detailed experimentally for single stranded DNA [70] and for oligogalacturonides [67], that this scaling symmetry breaks down at smaller DP values. This symmetry breaking at low DP (DP¡25) enables the separation of distinct fragments obtained from digests of pectic substrates, regardless of their same average charge density [71].

To help understand and predict the transport behaviour of oligogalacturonides molecular dynamics simulations have been carried out. The GLYCAM force field [72] was predominantly used, owing to its popularity and development in the field of polysaccharides. Occasional simulations were repeated with another force field (Amber) in order to assess the robustness of the results. In addition to co-ordinates and force field, the partial charges of the atoms in system required which were calculated using Quantum Mechanical calculations, specifically DFT.

2.2 Partial charge distribution of atoms

It is known that the electric charge has a quanta and fractional charges are not physical. However, partial charges exhibit electronegative or electropositive nature in molecules. That is; shared electrons spend more time in surrounding of more electronegative atoms and atoms are assigned partial negative or positive charges because electrons in the bonds spend more time close to some atom's cores than others [73].

Schematic representations of a single galacturonic acid residue and its methyl-esterified counterpart are shown in figure 2.1, which displays the atomic nomenclature used hereafter. All homogalacturonan (HG) substrates of interest are simple co-polymers, with

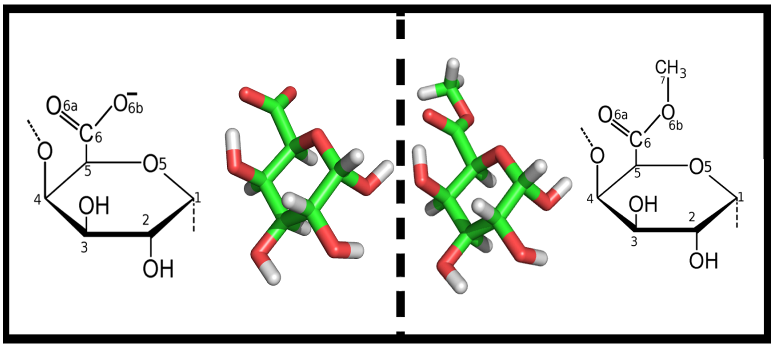


Figure 2.1: Galacturonic acid (left) and methylesterified galacturonic acid (right).

different lengths, and different arrangements of these two kinds of residues; galacturonic acid (G): negatively charged in our simulation, and uncharged methylesterified galacturonic acid (M).

In order to consider the effects of the chemical identity (G or M) of the nearest neighbour residues on the partial charges of a central (G or M) residue, we performed exhaustive calculations on the 8 possible trimer configurations. Figure 2.2 shows a fully charged trimer. The non-reducing end of chain, the left hand side residue as drawn, is the residue where the anomeric carbon (the carbon that has single bonds to two oxygen atoms) is involved in the glycosidic bond (the covalent bond joining carbohydrate residues). The reducing end, the right hand side residue as drawn, is the residue where the anomeric carbon is not involved in the glycosidic bond.

DFT calculations on these substrates were performed at the B3LYP/6-31G* level of theory in water using Gaussian09 [74]. Figure 2.3 shows the change of total energy during optimisation steps for a trimer using DFT calculations. When the total energy optimised, the final configuration was used to obtain the partial charges.

Figure 2.4 shows the electron density cloud around a fully charged galacturonan acid trimer showing there is a high electron density around the carboxyl group, where the oxygen atoms are deprotonated. It can also be seen that the negative charge density reduces slightly from non-reducing end (left) to the reducing end (right).

DFT calculations were also performed for a single galactronic acid residue and its dimer. Once the DFT calculation is complete then RESP fitting [75] was used to derive the partial charges on each atom. Table 2.1 shows the partial charges of atoms obtained for monomer and dimer.

Brief examination of (the considerably more computationally expensive than trimers) pentamers suggested that neglecting the effects of the second nearest neighbours was

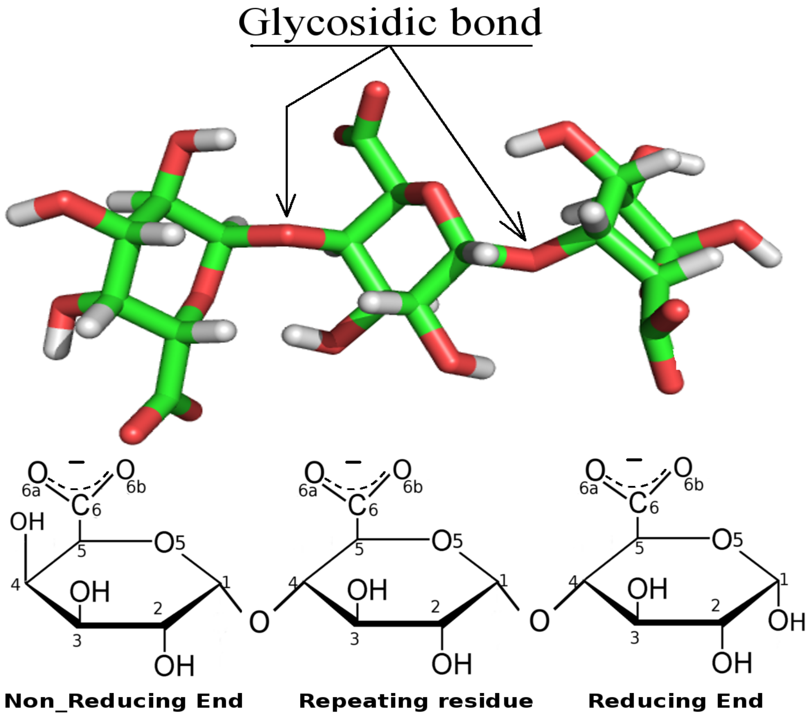


Figure 2.2: Fully charged galacturonan acid trimer.

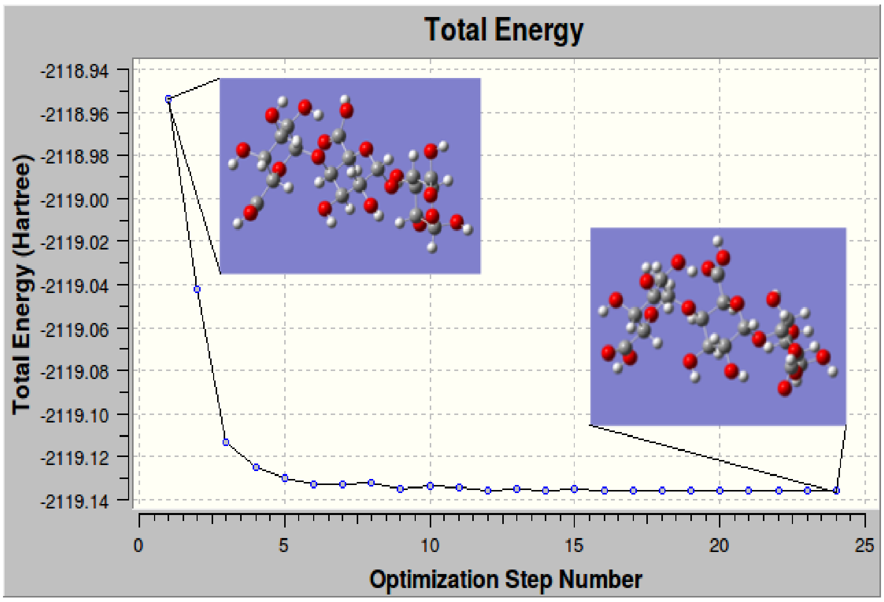


Figure 2.3: Change of total energy during optimisation steps using DFT calculation.

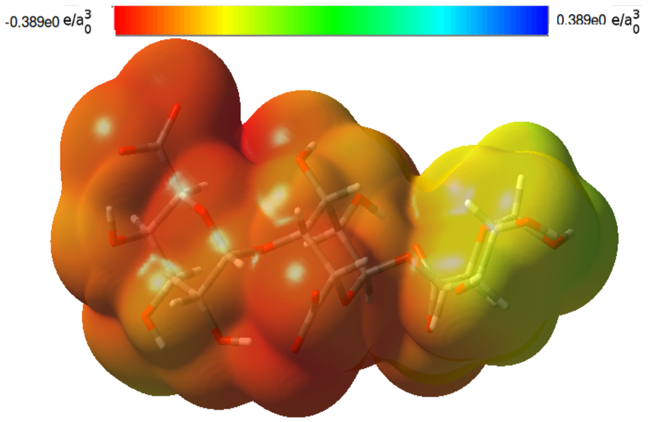


Figure 2.4: Electron Density around atoms in a fully charged galacturonan acid trimer.

Atoms	Dimer		Monomer
	Non-reducing end	Reducing end	
C_1	0.160060	0.213697	0.107820
C_2	0.069492	0.101241	0.247408
C_3	0.239617	0.319374	0.182562
C_4	0.148244	−0.127350	0.124866
C_5	0.184672	0.022523	−0.049232
C_6	0.692344	0.741737	0.661662
O_1		−0.631480	−0.625848
O_2	−0.675782	−0.624604	−0.641497
O_3	−0.659510	−0.610483	−0.621077
O_4	−0.645786	−0.307443	−0.631079
O_5	−0.409474	−0.396290	−0.338341
O_{6a}	−0.764069	−0.734936	−0.701024
O_{6b}	−0.764069	−0.734936	−0.701024
H_{C_1}	0.112488	0.085417	0.098691
H_{C_2}	0.060292	0.122966	0.070715
H_{C_3}	0.017875	0.020179	0.028932
H_{C_4}	0.022626	0.137723	0.053035
H_{C_5}	−0.028486	0.070984	0.062914
H_{O_1}		0.465160	0.472662
H_{O_2}	0.444359	0.414597	0.414232
H_{O_3}	0.437460	0.395064	0.400317
H_{O_4}	0.414509		0.383306

Table 2.1: The partial charge of atoms calculated for mono and di-meric galacturonic acid using RESP fitting after DFT calculations.

Atoms	From trimer	From pentamer
C_1	0.2934	0.2395
C_2	0.0021	0.0278
C_3	0.3618	0.3297
C_4	−0.0023	−0.0164
C_5	0.0580	0.0418
C_6	0.7601	0.7515
O_2	−0.6681	−0.6654
O_3	−0.6230	−0.6130
O_4	−0.3389	−0.3469
O_5	−0.4384	−0.4158
O_{6a}	−0.7543	−0.7474
O_{6b}	−0.7543	−0.7474
H_{C_1}	0.0751	0.0891
H_{C_2}	0.0868	0.0925
H_{C_3}	−0.0235	−0.0029
H_{C_4}	0.0945	0.1055
H_{C_5}	−0.0335	0.0406
H_{O_2}	0.4360	0.4343
H_{O_3}	0.4015	0.4029

Table 2.2: Comparison between the partial charge of atoms calculated for the central residue from a fully charged trimer and pentamer.

a reasonable assumption. Table 2.2 shows the atomic partial charges for the central residue for fully charged trimer and pentamer. The partial charges used for the first and last residues of chains of interest in MD were assigned using the results obtained for the atoms in the terminal residues of the trimer, shown in figure 2.2. These calculations were performed on a Intel(R) Xeon(R) 3.50GHz desktop computer and took around 18 hours to complete for each configuration.

Due to the affect of neighbour residues, the total charges calculated by RESP fitting for each individual sugar residue might not be an exact integer. However, this slight divergence from residues being completely neutral (for methylesterifed residues) or completely negatively charged (for galacturonic acid) is non-physical. Especially for long chains, a small amount of non-integer charge on each residue can lead to a non-integer charge on the entire chain which is a problem in MD. In particular if this chain is placed into a simulation box we cannot make the box neutral, and then owing to the periodic boundary condition (PBC) the system will have a huge amount of non neutralised charge. To deal with this issue we add all partial charges belonging to a specific residue and then, if this diverges from an integer number a small amount of charge is spread over the all atoms in order to achieve an integer charge (0 or 1) for each individual residue.

Tables 2.3 and 2.4 show the partial charges of atoms calculated respectively for the central and end residues for the different trimer configurations. M labels methylesterified galacturonic acid residues; and G, galacturonic acid residues; while non-r labels the non-reducing end of the oligomer and r, the reducing end. Superscript shows the type of adjacent residue(M or G). The electrophoretic mobility simulations were carried out in phosphate buffer, in order to match the experimental conditions as closely as possible, and as such the partial charges of the electrolyte atoms were also calculated and are shown in Table 2.5.

Table 2.3: The partial charges of atoms for the central residue in different trimers

Atoms	Configuration of trimers							
	GGG[a]	GGM[b]	GMG	GMM	MGG	MGM	MMG	MMM
C_1	0.2934	0.0251	0.3762	0.2096	0.2497	0.2114	0.3549	0.2337
C_2	0.0021	0.2084	−0.0142	0.1407	0.0083	0.0484	−0.0418	0.0892
C_3	0.3618	0.2529	0.2484	0.1085	0.3369	0.2957	0.3820	0.2477
C_4	−0.0023	0.0877	0.0823	0.3604	−0.0146	−0.0340	−0.0059	0.1733
C_5	0.0580	−0.1718	−0.0660	−0.6495	0.0487	−0.0988	−0.0959	−0.5053
C_6	0.7601	0.8362	0.8383	0.9940	0.7550	0.7999	0.8564	0.9869
C_7			−0.0383	−0.1049			0.0774	−0.0018
O_2	−0.6681	−0.6295	−0.6483	−0.5819	−0.6558	−0.5699	−0.6254	−0.5715
O_3	−0.6230	−0.6077	−0.6461	−0.6341	−0.5981	−0.5880	−0.5921	−0.5819
O_4	−0.3389	−0.3492	−0.2712	−0.3381	−0.3534	−0.3361	−0.3201	−0.3420
O_5	−0.4384	−0.3120	−0.4274	−0.2592	−0.4215	−0.3848	−0.4226	−0.3176
O_{6a}	−0.7543	−0.7594	−0.5664	−0.5771	−0.7478	−0.7577	−0.5607	−0.5769
O_{6b}	−0.7543	−0.7594	−0.3971	−0.3682	−0.7478	−0.7577	−0.4226	−0.4252
H_{C_1}	0.0751	0.1043	0.0763	0.1171	0.0926	0.1068	0.0878	0.1167
H_{C_2}	0.0868	0.1054	0.0989	0.1096	0.0921	0.1176	0.1168	0.1196
H_{C_3}	−0.0235	−0.0042	0.0383	0.0686	−0.0131	−0.0033	−0.0108	0.0257
H_{C_4}	0.0945	0.0998	0.0906	0.0718	0.0993	0.1223	0.1207	0.1075
H_{C_5}	0.0335	0.1076	0.0798	0.2093	0.0442	0.0891	0.1141	0.2148
$H_{C_{7a,b,c}}$			0.0912	0.1040			0.0610	0.0784
H_{O_2}	0.4360	0.3805	0.4365	0.3919	0.4315	0.3714	0.4221	0.3810
H_{O_3}	0.4015	0.3853	0.4358	0.4195	0.3938	0.3677	0.3930	0.3909

[a] G: Galacturonic acid residue (charged),

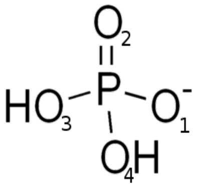
[b] M: Methyl-esterified galacturonic acid residue (uncharged).

Table 2.4: The partial charges of atoms for terminal residues

Atoms	G^G_{Non-r}	G^M_{Non-r}	G^G_r	G^M_r	M^G_{Non-r}	M^M_{Non-r}	M^G_r	M^M_r
C_1	0.0998	−0.1850	0.2180	0.3060	0.2146	0.0705	0.3257	0.3364
C_2	0.1423	0.3324	0.1264	0.0601	0.0615	0.1717	0.0403	0.0178
C_3	0.2769	0.2001	0.2759	0.3938	0.3587	0.2256	0.3176	0.3683
C_4	0.1711	0.3093	−0.0240	−0.1537	0.2350	0.3562	−0.0739	−0.1121
C_5	0.0154	−0.2706	−0.0795	0.0550	−0.2200	−0.5071	−0.1120	−0.0981
C_6	0.6745	0.7165	0.7856	0.7369	0.8367	0.9563	0.8306	0.8249
C_7					−0.0867	−0.1068	0.0944	0.0927
O_1			−0.6383	−0.6427			−0.6278	−0.6277
O_2	−0.6920	−0.6566	−0.6294	−0.6199	−0.6774	−0.6112	−0.6042	−0.5993
O_3	−0.6434	−0.6401	−0.6186	−0.6346	−0.6688	−0.6573	−0.5982	−0.6040
O_4	−0.6565	−0.6998	−0.2640	−0.2789	−0.6588	−0.6592	−0.1594	−0.1822
O_5	−0.3692	−0.2356	−0.3769	−0.4291	−0.3794	−0.3062	−0.4094	−0.4142
O_{6a}	−0.7168	−0.7090	−0.7478	−0.7335	−0.5588	−0.5764	−0.5502	−0.5453
O_{6b}	−0.7168	−0.7090	−0.7478	−0.7335	−0.3795	−0.3918	−0.4400	−0.4363
H_{C_1}	0.1003	0.1323	0.0827	0.0623	0.0934	0.1096	0.0759	0.0761
H_{C_2}	0.0477	0.0639	0.1146	0.1308	0.0921	0.1128	0.1403	0.1438
H_{C_3}	−0.0135	−0.0081	0.0926	0.0068	−0.0186	0.0257	0.0326	0.0287
H_{C_4}	0.0214	0.0135	0.1193	0.1443	0.0248	0.0304	0.1403	0.1513
H_{C_5}	0.0237	0.1319	0.0926	0.0622	0.1131	0.1841	0.1378	0.1374
$\mathsf{H}_{C_{7a,b,c}}$					0.1054	0.1097	0.0562	0.0579
H_{O_1}			0.4669	0.4614			0.4644	0.4647
H_{O_2}	0.4418	0.4009	0.4159	0.4065	0.4397	0.3932	0.4057	0.4050
H_{O_3}	0.4114	0.3978	0.3995	0.3998	0.4252	0.4158	0.4009	0.3984
H_{O_4}	0.3819	0.4152			0.4370	0.4350		

Table 2.5: The partial charges for atoms of the phosphate ions used in the MD simulations.

$H_2PO_4^-$		HPO_4^{-2}	
atoms	partial charges	atoms	partial charges
P	1.008824	P	0.910761
O_1,O_2	−0.759102	O_1,O_2,O_3	−0.856866
O_3, O_4	−0.664733	O_4	−0.714562
H_{O_3}, H_{O_4}	0.419422	H_{O_4}	0.374399



2.3 MD simulation details

Armed with reliable partial charges obtained from the quantum mechanical calculations of a set of trimer sequences, larger HGs were constructed *in-silico* and MD simulations carried out. SPC/E water molecules [62] were used. The temperature was maintained at 298 K using a Berendsen thermostat [76], all bonds were kept constrained, and the pressure was kept fixed at 1.0 bar using the Parrinello-Rahman algorithm [77]. Periodic boundary conditions were applied and the integration step was set to 2 fs. Coordinates were recorded every 500 steps for the first nanosecond in order to calculate a molecular diffusion coefficient, and subsequently an external electric field was applied, with coordinates recorded every 5000 steps in this part of the simulation. The Particle Mesh Ewald summation method [78] was used in order to treat the long-range electrostatics. The simulations were performed using supercomputing facilities available through New Zealand eScience Infrastructure(NeSI). Between 20 and 250 ns of simulation could be carried out per day depending on the length of the saccharide moiety.

The initial configurations of the oligogalacturonides were generated based on those available at glycam.org, with bonded and Lenard-Jones parameters retrieved from the GLYCAM force field [72], which has been specifically developed to simulate the dynamics of sugars and sugar-like molecules. PyMol software [79] was used to perform methyl-esterification of selected galacturonic acid residues. MD visualization was performed using VMD [80]. Phosphate buffer was simulated by mixing $HPO_4{}^{-2}$ and $H_2PO_4{}^-$ anions to obtain a specific ionic strength, and finally the whole system was neutralised by Na^+ cations. 200ps of simulation was performed for both the NVT and NPT ensembles, which are canonical and Isothermalisobaric ensembles respectively, prior to the start of the simulation proper, which was sufficient to achieve equilibrium of temperature, pressure and density. After 1 ns of simulation without the presence of any electric field, simulations were run in which external electric fields were applied, typically with $E = 0.01$, 0.02, 0.03 and 0.04 Vnm^{-1}.

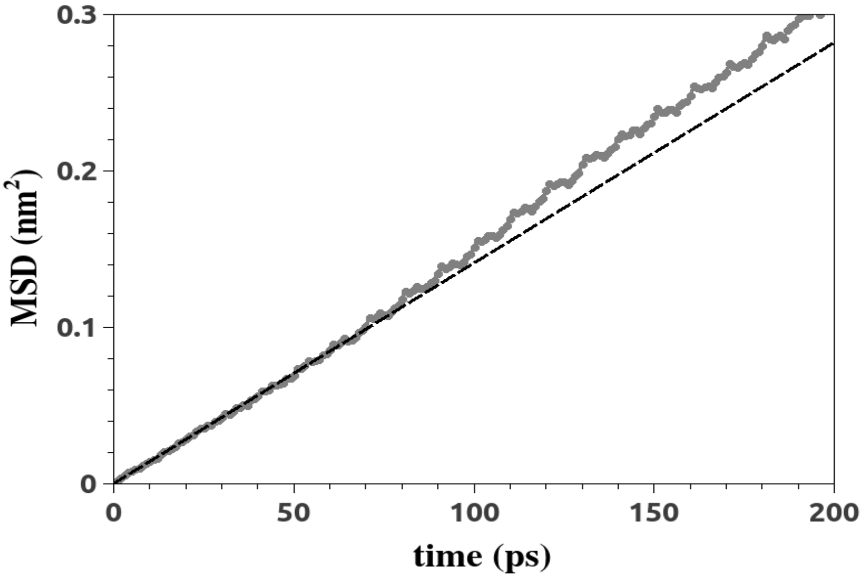


Figure 2.5: The mean square displacement, MSD, of the centre of mass of a pentameric oligogalacturonic acid molecule versus time. The diffusion coefficient is extracted from the gradient.

2.3.1 MD Box Size correction

Figure2.5 shows the mean squared displacement of the centre of mass for the simulated molecule versus time, from which self-diffusion coefficients, D, were extracted from the slope of the linear regime (Equation 2.1).

$$D = \lim_{t \longrightarrow \infty} \frac{\partial}{\partial t} \frac{\langle |\mathbf{r}(t) - \mathbf{r}(0)|^2 \rangle}{6} \tag{2.1}$$

Using periodic boundary conditions are essential in MD, but these restrictions effect on the calculated diffusion coefficient of particles inside the system. Figure 2.6 displays how the calculated diffusion coefficients scale with box size, and how, by applying the equation 2.2, values calculated with any size of box can be corrected to give a physically realistic value (that would obtained from a very large box).

$$D_{md}(L) = D_0 - \frac{k_B T \xi_{EW}}{6\pi\eta L} \tag{2.2}$$

where D_{md} is the diffusion coefficient calculated from MD simulation, D_0 is the diffusion coefficient, T is the absolute temperature,η is the solvent viscosity, $\xi_{EW} = 2.83729$ [81] is the self-term for a cubic lattice and L is the size of simulation box [82, 83].

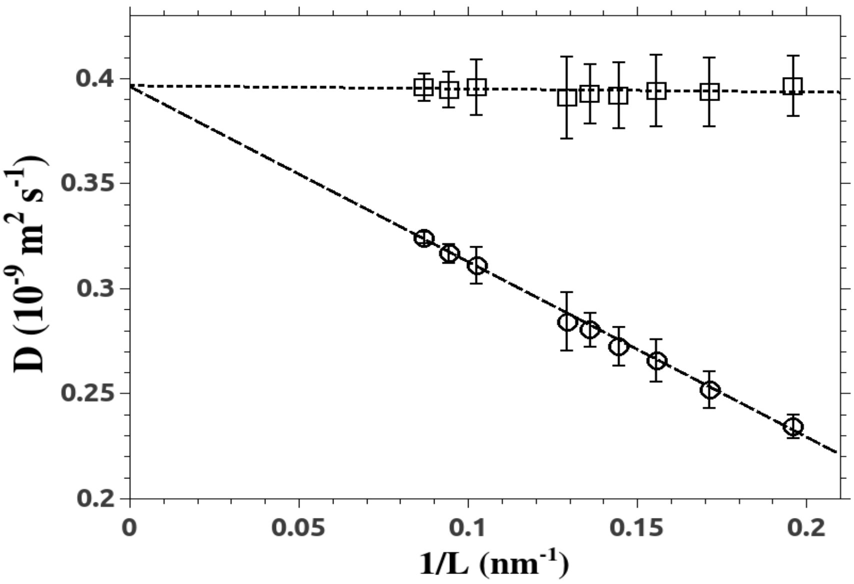


Figure 2.6: The diffusion coefficient for a fully charged pentamer , D, versus reciprocal box length; circles show the raw data, and squares the data corrected for finite size affects as described in the text.

The diffusion coefficient of chains and their mobility has a linear relation described by the Nernst-Einstein equations;

$$D = \frac{k_B T}{Q_{eff}} \mu \tag{2.3}$$

where Q_{eff} is the effective charge on the chain. This relation can be used for finding effective charge on the chain using known mobility and diffusion coefficients of a particular particle. Combining equation 2.3 and equation 2.2 leads to a similar box size correction for the electrophoretic mobility of chains, which can be rewritten as;

$$\mu_{md}(L) = \mu_0 - \frac{Q_{eff}\xi_{EW}}{6\pi\eta L} \tag{2.4}$$

2.3.2 Effect of Viscosity of SPC/E Water

In addition it is well known that while SPC/E water performs well in MD in many regards, it does not exactly reproduce its viscosity and, as such, when transport properties are of interest, the calculated mobility must be scaled by the ratio of the viscosity of SPC/E water and the viscosity of the actual physical solution in which the experiments being modelled were carried out.

$$D_{\eta} = \frac{\eta_{SPC/E}}{\eta_{H_2O}} D_0 \tag{2.5}$$

The scaling relationship can equally provide a reliable method of correcting the results of the electrophoretic mobility simulations for the effects of finite box size and viscosity correction.

2.3.3 Effect of External Electric Field on the Ionic Distribution around the Chain

It should be noted that, owing to the short lengths of time for which MD can be realistically simulated, to observe reasonable displacements of the molecules under study requires that the electric fields applied in MD are significantly greater than those applied in the experimental situation. It is therefore pertinent to ensure that the field used in the simulations is not so large that it perturbs the ion atmosphere significantly during its application. Figure 2.7 shows the radial distribution function of sodium counter-ions around an HG molecule comprising of 25 charged residues upon application of external electric fields with different strengths. It can be seen that while the application of fields larger than 0.1 V nm^{-1} substantially modify the distribution of the ions, smaller fields can be safely applied with minimum perturbation.

2.4 Electrophoretic mobility as a function of degree of polymerisation

To determine the electrophoretic mobilites of a chain, after doing simulation for 1 ns without any external field to calculate the diffusion coefficient, we turn on the external field to monitor the movement of chain during the simulation time. An external electric field is applied in the positive Z direction and the negatively charged chains move in the opposite direction. Figure 2.8 shows the motion of centre of mass for a fully charged trimer under the influence of different external electric fields. As it is shown in the figure 2.8, there are a linear regions that we can use to extract the drift velocity of chains for specific external electric fields.

There is a linear relationship between the drift velocity of a charged particle and the applied electric field, with a constant of proportionality which is the electrophoretic mobility (Equation 2.6);

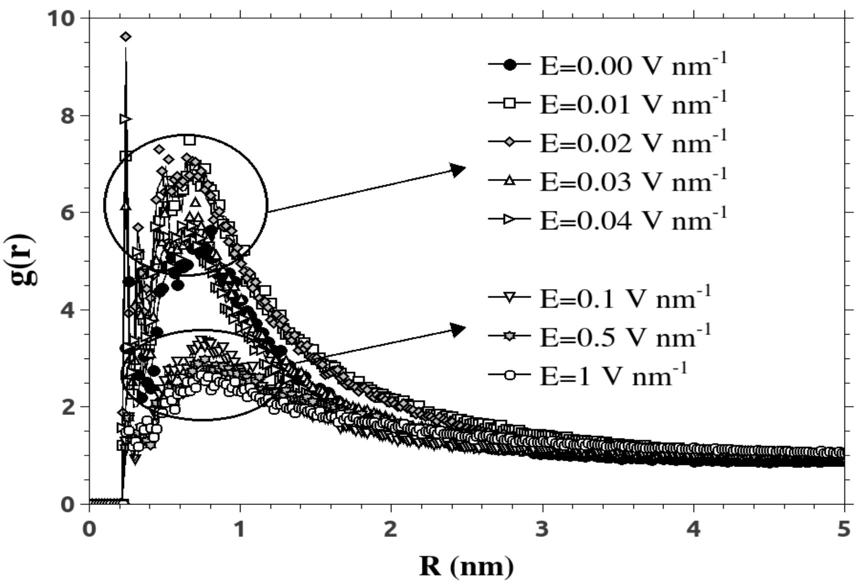


Figure 2.7: Radial distribution function of counter-ions around a fully charged chain(DP25) in present of external electric field with different strengths, showing the effect of an applied electric field on the form of the ion distribution around the polyelectrolyte.

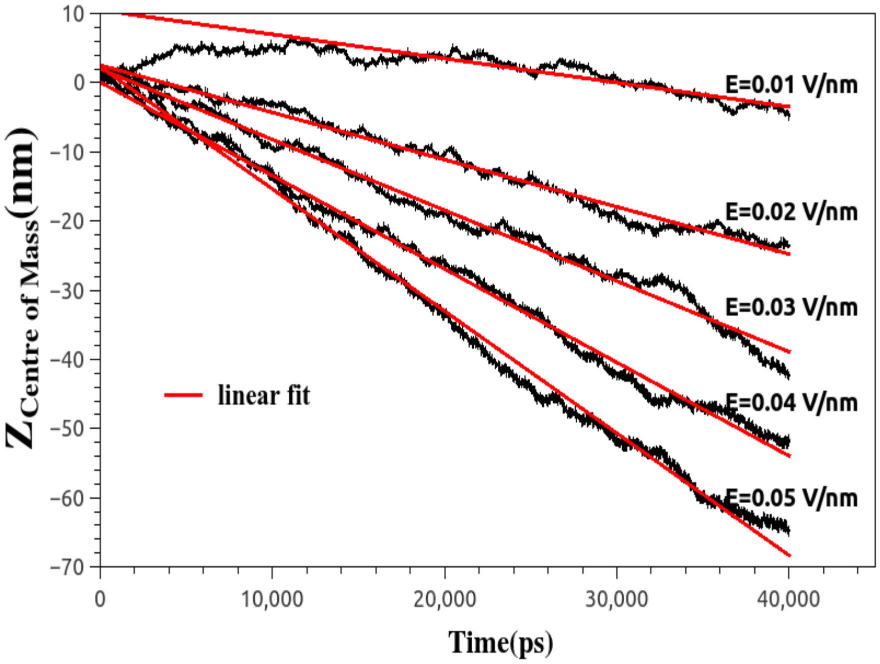


Figure 2.8: Displacement of centre of mass for a fully charged trimer in the affect of different electric field.

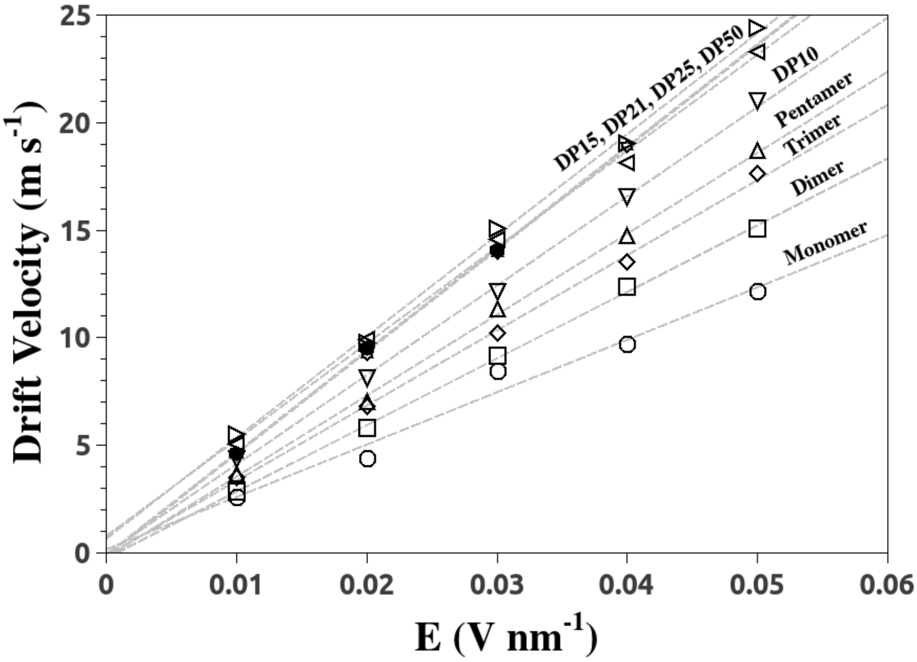


Figure 2.9: Drift velocity of fully charged oligogalacturonides with different DPs versus electric field.

$$v = \mu E \tag{2.6}$$

Having a sequence of drift velocities corresponding to different external electric field strengths, we can plot a linear graph of drift velocity versus external electric field strength. By looking at equation 2.6 we can see that the gradient of this graph is the electrophoretic mobility of the particle under study. Figure 2.9 shows the raw drift velocity versus electric field data for a series of fully charged oligogalacturonides with DPs ranging from 1 to 50 calculated using molecular dynamics simulations as described herein. The latter simulation takes around one week to complete as opposed to several hours for the shorter chains.

Figure 2.10 displays how the extracted electrophoretic mobilities of different species scale with DP in 50 mM sodium phosphate ionic strength background electrolyte. It is shown that after about DP25 the electrophoretic mobility become length independence. For the short chains (DPs less than 25) the electrophoretic mobility scales logarithmically with DP (See Inset in Figure). This behaviour has been previously reported experimentally for other charged biopolymers[84]. This DP independence after a certain DP allows us to compare results of DP25 with the experimental data which are usually for long chains (~DP 50-100).

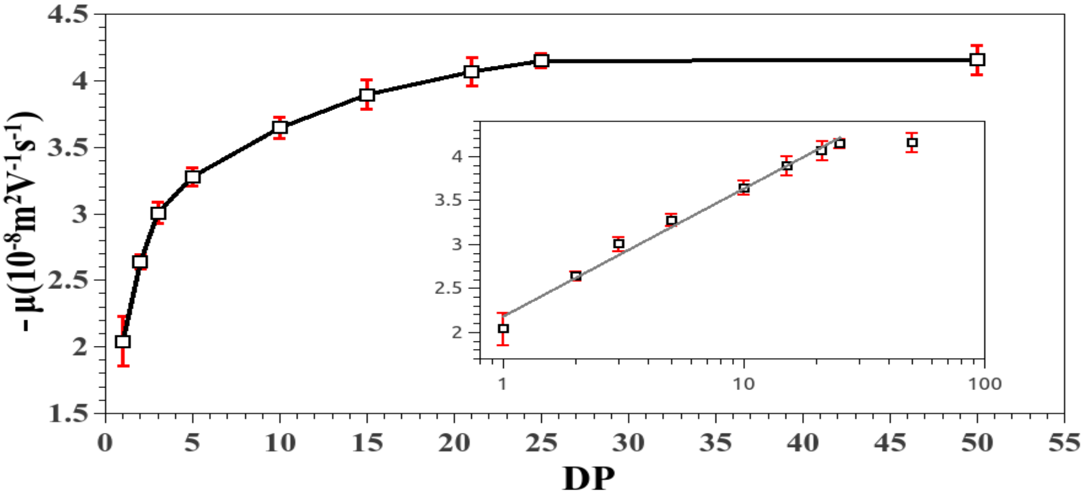
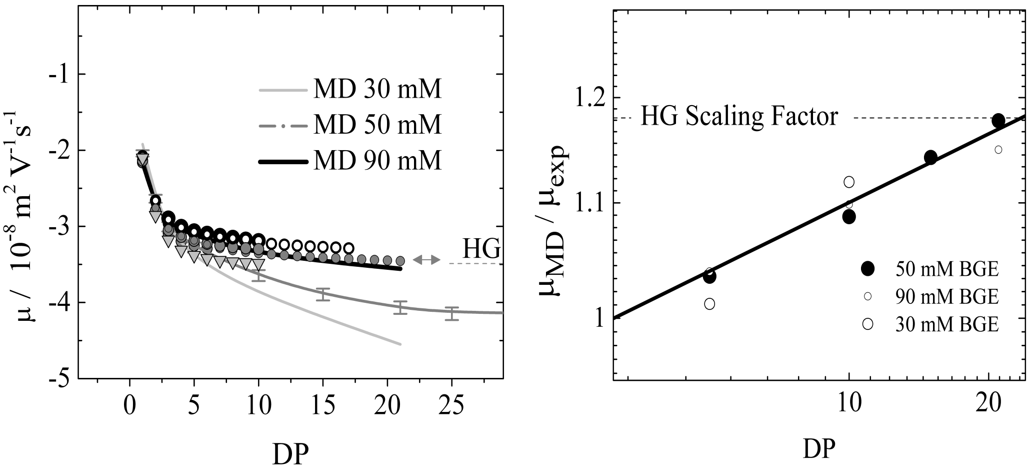


Figure 2.10: Electrophoretic mobility of fully charged chains versus degree of polymerisation in 50mM of phosphate buffer. Fitted line in inset is $\mu = b + a\ln(DP)$, where $b = 2.18 \pm 0.006$ and $a = 0.63 \pm 0.03$.

Figure 2.11a displays how the MD predictions (Solid Lines) of the electrophoretic mobilities of these different species scale with DP, and includes calculations performed in 30, 50 and 90 mM ionic strength sodium phosphate background electrolyte(BGE). Also shown are several experimental datasets points available in the literature. It can be seen that for oligogalactronic acids of DP 5 or less the MD simulations do an excellent job of quantitatively predicting the electrophoretic mobility, capturing both the DP and ionic strength dependence.

Above DP 5 the data (at all ionic strengths) can be seen to begin to deviate increasingly from the MD predictions (Figure 2.11a) . Also shown in the figure as a dotted line against the right hand axis is the experimental result obtained in 50 mM BGE from longer (DP ∼ 100) homogalacturonans [3]. It can be seen that the experimental data at the same ionic strength as a function of DP indeed approaches that value as the DP increases as expected. It is speculated that the discrepancy between the experimental work and MD data found for longer DPs may arise from an increase in viscosity above that of water. Given that the predictions of DP5 fit well without adjustment and that in previous work the DP 100 HGs required a scaling of a factor around 1.18 to map the MD predictions onto the experimental data it is not perhaps surprising that this scaling factor is a function of DP. This scaling factor is shown in figure 2.11b, and pragmatically allows the MD results to be mapped to an experimental prediction.



(a) (b)

Figure 2.11: a) The electrophoretic mobilities of different species scale with DP in 30, 50 and 90 mM ionic strength sodium phosphate BGE. b) Scaling factor of MD results to be mapped to an experimental prediction. Lines in (a) and (b) are guides to the eye.

2.5 Electrophoretic mobility for n^m Oligomers

In the previous section it was shown that the ionic strength of the background electrolyte has an important impact on the electrophoretic mobilities of charged polysaccharides particularly for the long chains. In this section the effect of phosphate buffer on the mobility of short oligomers with different charge patterns is investigated. Figure 2.12 shows the configuration of charged GalA residues, (open circles), and uncharged methylestified residues, (solid black circles), for different species with the notation of n^m where n and m are degree of polymerisation and number of methyester group on them respectively. As it is not observed in the enzymatic digests used to generate these species experimentally that terminal residues of oligomers (neither reducing end nor non-reducing end) can be methylesterified[85], the simulated species also do not have either of the two ends methylesterified.

Figure 2.13 shows how the electrophoretic mobilities of specific oligogalcturonides, of varying DP, n, and number of methylesterified residues, m, that have been selected owing to their prevalence in enzymatic digests of pectin and relevance to plant biology, vary with ionic strength. The vast majority of the species examined here (and indeed those most easily characterized experimentally in digests) have a DP of less than six and, as such, the comparison of the MD predictions with experiment is undertaken without the scaling required for longer-DP fragments as described in the previous section.

Figures 2.14a and 2.14b compare the electrophoretic mobilities predicted by MD to

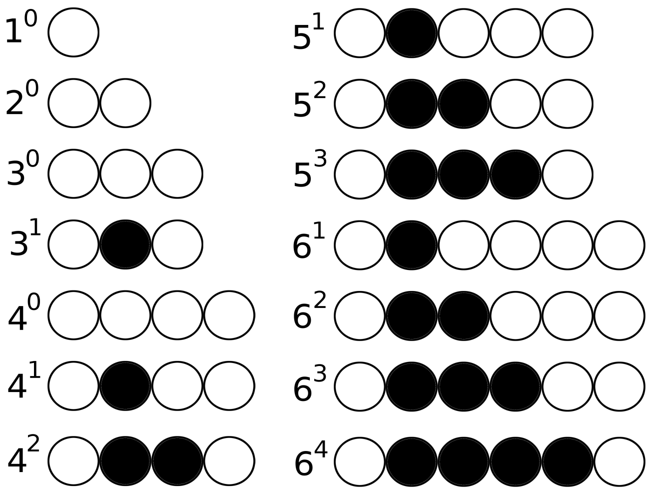


Figure 2.12: Configuration of charged residues,open circles, and methyl-esterified residues (uncharged), solid circles for n^m oligomers.

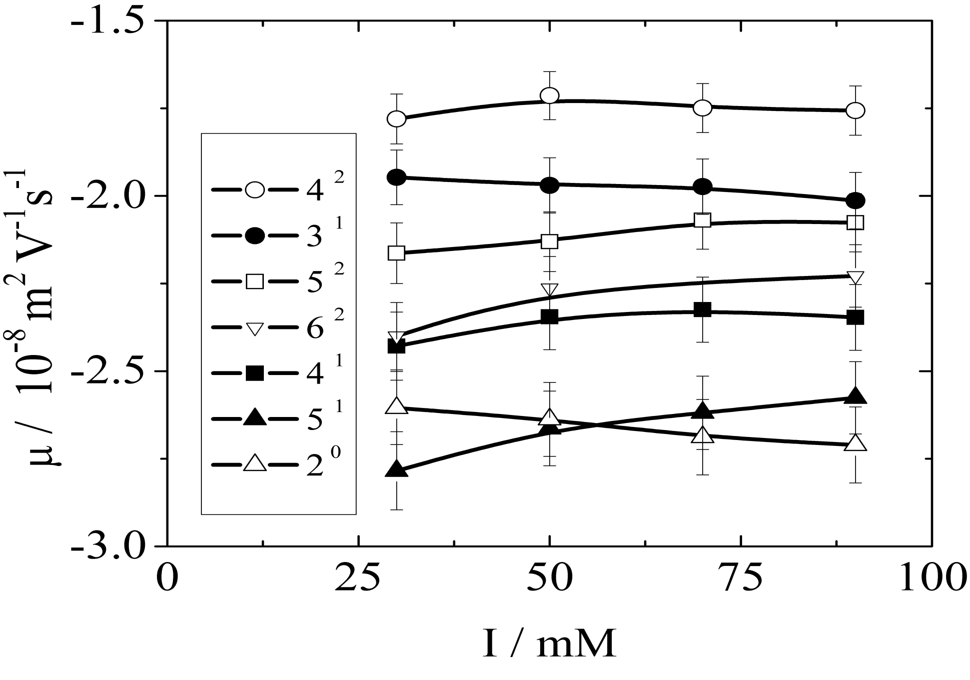
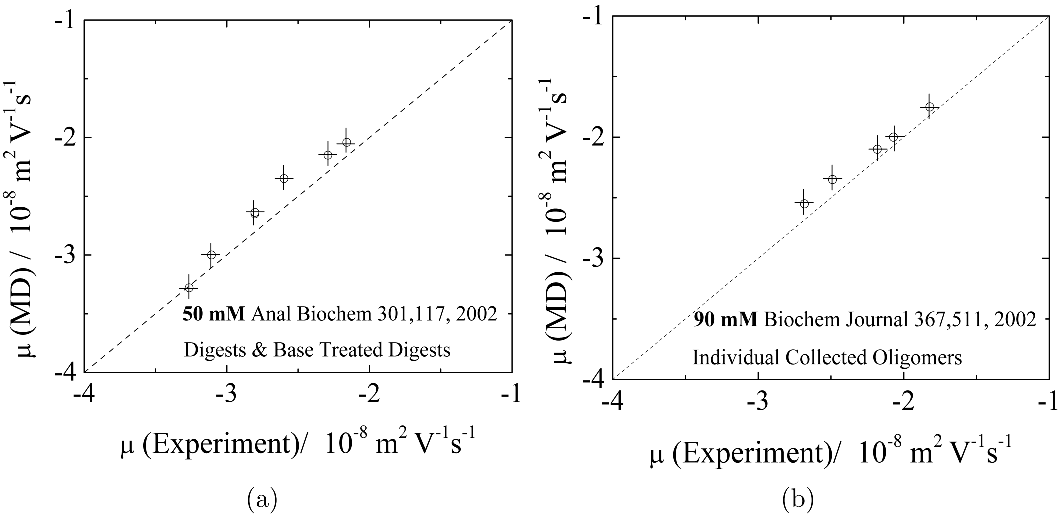


Figure 2.13: Electrophoretic mobility of n^m oligomers in different phosphate buffer concentration(I).



(a) (b)

Figure 2.14: The electrophoretic mobilities predicted by MD to available experimental data for a set of partially methylesterified oligomers, at a) 50mM, b) 90mM ionic strengths.

available experimental data for a set of partially methylesterified oligomers, at two ionic strengths, where indeed the agreement can be seen to be reasonable.

This information should be helpful in designing experimental conditions that optimize the resolution of particular species, and allows for the prediction of electropherograms. Of particular interest for the optimization of the experimental electrophoresis conditions (such as the ionic strength of the BGE) and the preliminary identification of species is to be able to predict the order of migration of the different oligogalacturonides. This is particularly valuable for species where standard samples are not available or easily obtained.

Table 2.6 shows the order of migration of galacturonides obtained experimentally (from an endo-PG II digest of a partially-methylesterified pectin (DM 31% randomly distributed)) and in particular how the order of some of the species changes as the ionic strength is varied between 30 and 90 mM. Also shown in the table 2.6 are the results of our simulations which can be seen to correctly predict the order of migration of the oligogalacturonide species (and the co-elution of some species) as a function of ionic strength.

Furthermore, Figure 2.15 shows how the electrophoretic mobilities of species of varying DP, n, vary as the number of methylesterifed groups, m increases from 0 to n-2. It is shown that by adding consecutive methylester groups to a chain the mobility of the oligomer reduces linearly, (with a gradient dependent on DP). The gradient of mobility

30 mM		50 mM		90 mM	
Exp	MD	Exp	MD	Exp	MD
4^2	4^2	4^2	4^2	4^2	4^2
1^0	1^0	1^0	1^0	3^1 5^2	3^1 V Close
3^1	3^1	3^1	3^1	3^1 5^2	5^2 V Close
5^2	5^2	5^2	5^2	1^0	1^0
6^2	6^2	6^2	6^2	6^2	6^2
4^1	4^1	4^1	4^1	4^1	4^1
2^0	2^0	2^0 5^1	2^0 V Close	5^1	5^1
5^1	5^1	2^0 5^1	5^1 V Close	2^0	2^0
3^0	3^0	3^0	3^0	3^0	3^0

Table 2.6: The migration order of different fully and partially charged oligogalacturonides in free solution capillary electrophoresis with different ionic strength BGE. Experimental data compared with the predictions of MD. V Close, denotes that the results of the mobility calculations for the two species are very close to each other.

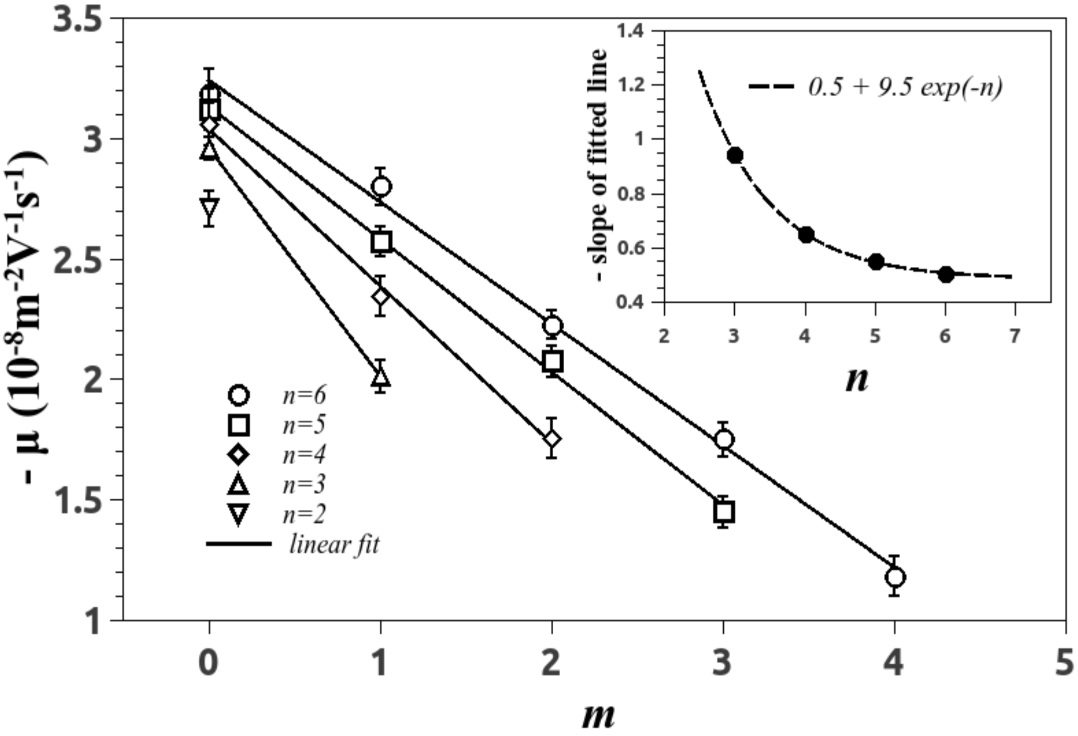


Figure 2.15: The electrophoretic mobilities of species of varying DP, n, vary as the number of methylesterifed groups,m increases from 0 to *n-1*, the ionic concentration is 90mM.

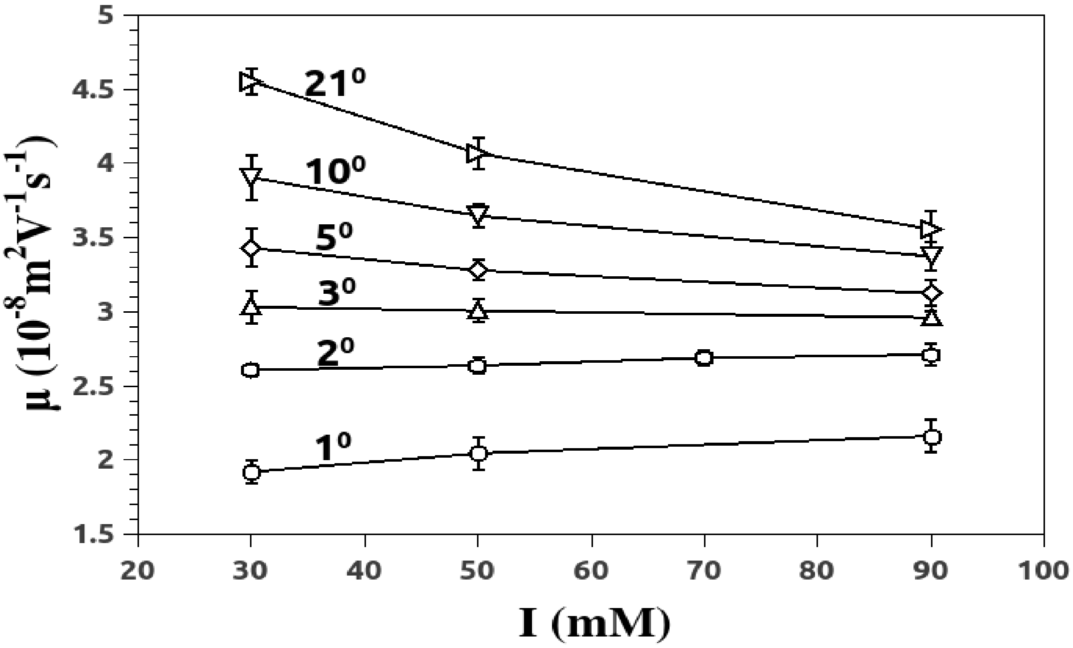


Figure 2.16: Electrophoretic mobility of n^0 oligomers (fully charged) as a function of ionic strength.

change for short oligomers is higher than longer oligomers, and as it is plotted in the inner graph in figure 2.15 this gradient scales with DP, n, as an exponential decay. This figure should be of help to experimentalists in predicting the electrophoretic mobility of particular species.

2.6 Dependence of electrophoretic mobility of n^0-mers on ionic strength

Figure 2.16 shows how the electrophoretic mobilities of specific DP oligogalcturonides, is calculated to vary with ionic strength. This figure can be divided to two parts, part1 for oligomers below DP3 and part2 for oligomers above DP3. It is possible to see that the electrophoretic mobility of oligomers with DP below DP3 increases with ionic strength and above DP3 the mobility decreases with ionic strength. For long chains increasing the ionic strength generates more counter-ions interacting strongly with the chain, leading to more neutralisation and less mobility (Counterion condensation is discussed extensively in the next chapter). But for DP less than 3 the small size of fragment and less possible room for counter-ions to interact with the oligomers means this condensation compensation mechanism is not active.

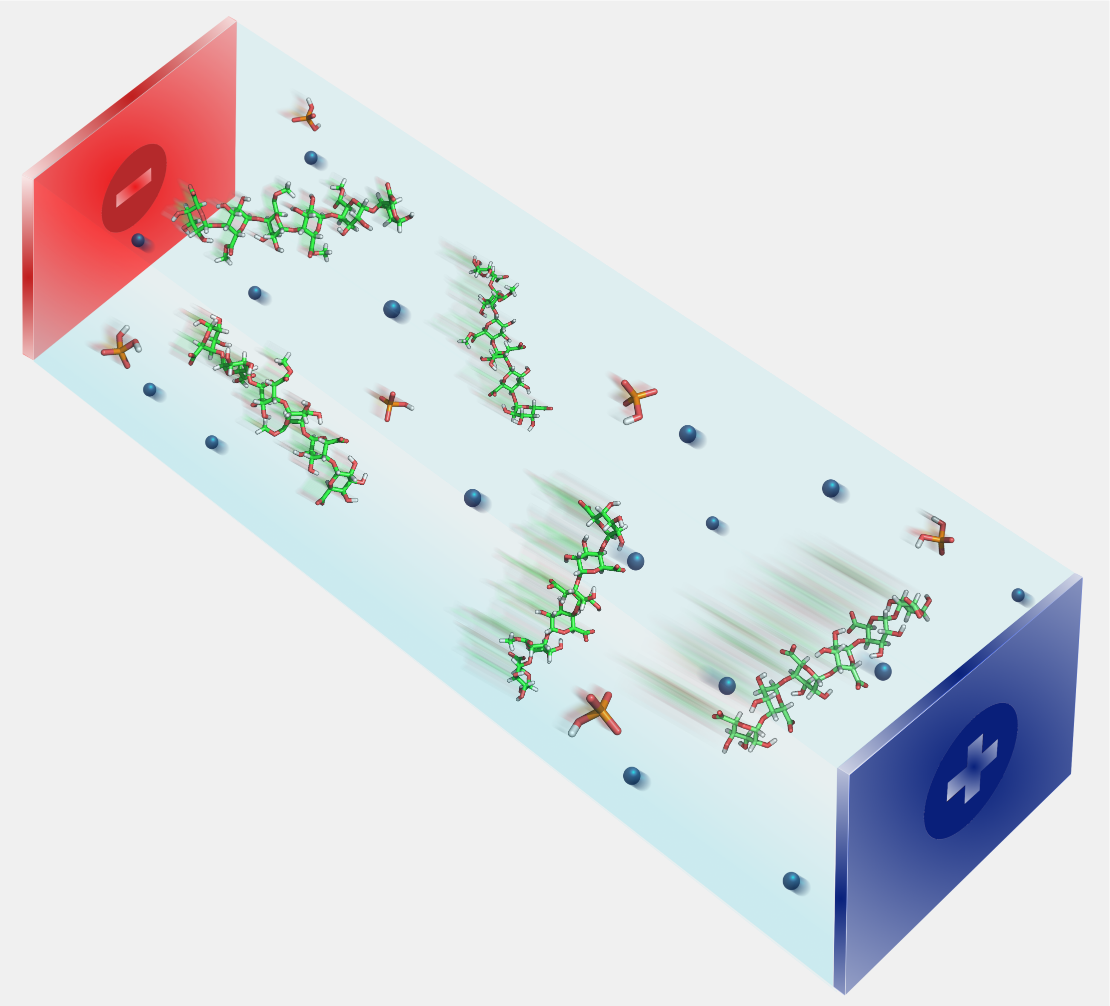


Figure 2.17: Proposed Cover Picture to the ELECTROPHORESIS journal which represents the high mobility of fully charged hexamer compared to partially charged species.

Chapter 3

Electrophoretic Transport of HGs

3.1 Introduction

The biological function of many polysaccharides, including several that are important in mammalian physiology, from the chondroitin and dermachondan sulphate strings that hold collagen fibrils in perfect register [86, 87, 88, 89] to the multi-functional heparin [90, 91, 92] depends on the patterning of charged residues along the backbone. Detailed structure-function understanding then depends on the reliable measurement of the *distribution* of charge-carrying groups. This means that for HGs (the focus of this work) the sequence that denotes the state of methylesterification of the individual residues should be ascertained. But with polysaccharide fine structure being routinely re-modelled to optimise its function, as a function of its position or time (*in-muro* or in the processing plant) , each HG chain in a sample presented for analysis can possess a different pattern of methylesterification.

It has previously been argued that measuring the intermolecular distribution of degree of methyl-estrification (DM) (the relative numbers of chains possessing slightly different chain-averaged DM) has some advantages in sample characterisation over the more traditionally used fragmentation approaches. It pursues information regarding intramolecular sequences and that by measuring and modelling this intermolecular distribution, significant progress in generating faithful ensembles of chains can be made [93, 94]. The success of this methodology clearly relies on having an experimental technique that can faithfully measure the intermolecular distribution of DM. This has been successfully demonstrated using HG samples that had been randomly demethylesterified from a high-DM starting substrate to obtain a series of samples with different

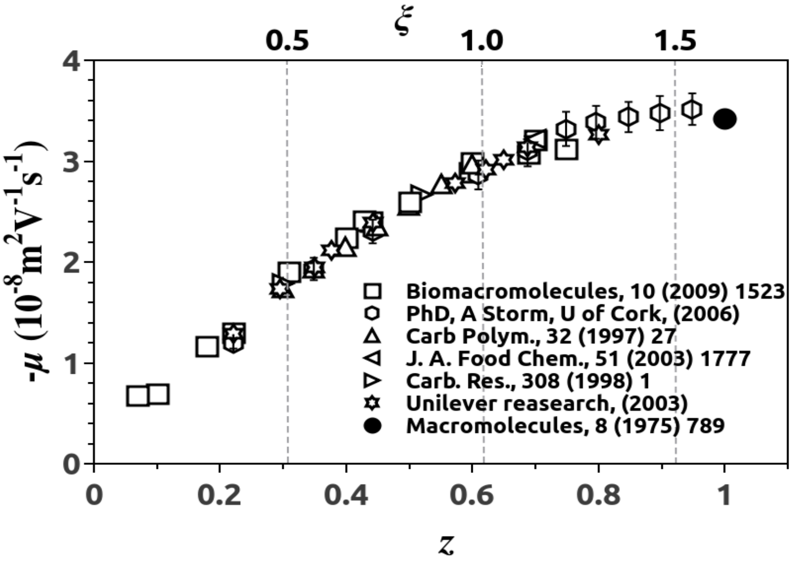


Figure 3.1: Electrophoretic mobilities, μ, measured for pure HGs and a number of pectin samples, gathered from the literature, as a function of the fraction of the sugar rings charged, z, and the dimensionless polyelectrolytic charge density parameter, ξ, described in the text.

sample-averaged DM values [94]. The agreement obtained, not just for the measured average DM values, but also for the known form of the DM distributions, attests to the fact that the *empirical* relationship used between the mobility and the fraction of charged residues in a chain (Figure 3.1) is a reasonable description of the actual physical relationship.

It is clear from the data in figure 3.1 that, in the region in which the majority of previous work in the literature has been carried out, corresponding to fractional charge density (the number of charged residues divided by the total number of residues comprising the chain), $z \sim 0.1$ - 0.6, there is a well-defined consistent relationship, wherein simple linear regression analysis based on the mobility of surrounding standards can yield a reasonable estimate of the DM of an unknown sample from its electrophoretic mobility [95]. However it is also abundantly clear that at higher vales of z (lower DM) the dependence of electrophoretic mobility on the fractional charge (as calculated naively from the structure as 1-(DM/100)) is modified. In this regime chains migrate consistently with lower charge densities than might be naively anticipated from their structures if an extrapolation of the linear relationship is assumed, reminiscent of the postulates of counter-ion condensation theory [84, 96, 97, 98, 99]. It should be noted that although in detail both the chain conformation and the pK_a of the galacturonic acid residues will change with DM the effect of these changes on the electrophoretic mobility are expected to be minor owing to the relatively high intrinsic stiffness of the uncharged chains (in contrast with proteins) and the experimental pH respectively [100].

Conceptually, counter-ion condensation in polyelectrolyte solutions can be understood by considering the distribution of ions around a charged line as described by Poisson-Boltzmann theory.In this model the fine structure of the polyelectrolyte (the precise patterning of the charged and uncharged residues) is simply mapped onto a charged-line model by setting a uniform spacing of charges on the line that corresponds to the average spacing found on the polyelectrolyte. The abstracted charged lines are subsequently characterized by a so-called polyelectrolyte charge parameter(as described in chapter 1), ξ, defined by $\xi = \frac{l_B}{b}$, where l_B is the Bjerrum length and b is the uniform distance between charges (Described in more detail in Chapter1). Physically the Bjerrum length represents the distance at which the electrostatic interaction energy between two single charges is equivalent to the thermal energy, and in water at 298 K is around 0.7 nm. The distance between charges placed on the line is defined as described as $b =< l_i >$ where l_i is distance between the centre of mass of two neighbouring charges on the polyelectrolyte.

The electric field, E, and concomitant potential around such a uniformly charged line in a continuum solvent with a uniform dielectric constant ϵ can be calculated by Gauss' theorem. Subsequently, with the position-dependent potential-energy in hand, the expected probability distribution of counter-ions around such a uniformly charged line can be described as a function of radial distance from the line, r, simply using the Boltzmann equation

$$P = P_0 r^{-2\xi} \tag{3.1}$$

where P_0 is a constant that ensures the cumulative probability sums to 1.

In the case where the spacing between charges on the line used to model the polyelectrolyte is significantly greater than the Bjerrum length then ξ can be extracted from equation 3.1 if the distribution of the counter-ions about a polyelectrolyte is known. Such strategies have been pursued experimentally by techniques such as anomalous small angle x-ray scattering (ASAXS) that aim to measure the ion distribution directly [101, 102, 103, 104]. The basic argument for the existence of more complex behaviour at increased charge densities, and the proposal of the so-called *condensation* of counter-ions, originates from attempting to integrate equation 3.1 in order to calculate the number of counter-ions inside a cylinder of fixed radius and unit length. In this case the integrand contains the radius to the exponent (1-2ξ) so that when $\xi > 1$ (that

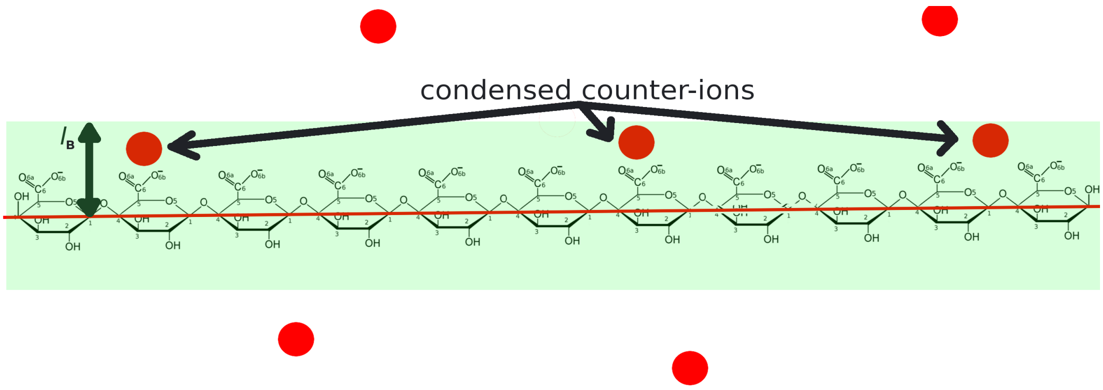


Figure 3.2: Schematic representation of counter-ion condensation when the charge spacing becomes less than the Bjerrum length(l_B).

is; the charge spacing becomes less than the Bjerrum length, figure 3.2) the integral diverges at the charged line, signalling that such a form for the counter-ion distribution is unstable under these conditions, and suggesting that in reality a tight binding of a certain number of counter-ions to the chain effectively limits ξ to values less than 1. Indeed, under these conditions, when a single average charge spacing, b, can realistically be used to define ξ then it has been shown that a fraction of counter-ions, 1-$(1/\xi)$, become *condensed* [99].

It is noteworthy that in figure 3.1 the simple linear relationship between electrophoretic mobility and naively-calculated charge density indeed appears to break down at a polyelectrolyte charge parameter close to that of 1. Manning's counterion condensation theory has previously been found to be useful for the description of the behaviour of a number of polyelectrolytes of both synthetic and biological origin,[105, 106, 107, 108] and while more complicated and rigorous models do exist, [109, 110, 111] these do not presently appear to be able to capture the available experimental data. In previous studies using HGs as model polyelectrolytes the key relationship that was used to map a chain's electrophoretic mobility to the fraction of methylesterified residues it contained, (1-z) where z is the fraction of charged groups, was obtained empirically from the data shown in figure 3.1 [94]. Hypothesizing that the form of the relationship shown in figure 3.1 does indeed signal that counter-ion condensation is important at low degrees of methylesterification, raises the question of what role *local* condensation might play in determining the mobility of samples with high degrees of methylesterification, but with non-random, *blocky*, fine structures. How many singly-charged sugar residues can be placed next to one another, sandwiched between neutral sugar residues, before local condensation effects modify the ion distribution and effect the measured electrophoretic mobility?

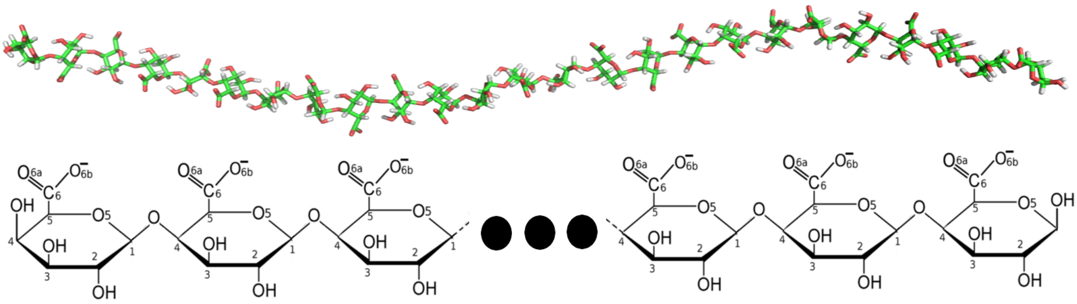


Figure 3.3: Snapshot of fully charged 25-mer after 5ns MD simulation.

Herein it has focussed on existing experimental data which compares the electrophoretic mobilities of just such *locally blocky* samples with sister-samples of similar average DM but with random DM distributions. It is shown that Molecular Dynamics (MD) simulations can not only explain the differences found in this case, but can also provide a prediction of the form of figure 3.1, for random patterns, that agrees well with the experimental data, both illuminating the role of counterion condensation and paving the way to allowing fine structure predictions for non-random samples to be mapped to predicted experimental electrophoretic mobility distributions.

3.2 MD on HGs

Following the same protocol of chapter 2 for setting up the simulations, larger HGs were constructed *in-silico* (figure 3.3) and MD simulations carried out. 200 ps of simulation was performed for both the NVT(canonical) and NPT(Isothermalisobaric) ensemble prior to the start of the simulation proper, which was sufficient to achieve equilibrium of temperature, pressure and density(fig.3.4 a,b and c). After 1 ns of simulation without the presence of any electric field to have an equilibrium distribution of atoms, simulations were run in which external electric fields were applied, typically with E = 0.01, 0.02, 0.03 and 0.04 V nm^{-1}. Additional details regarding the MD simulations are provided in table3.1.

Having established the equilibration conditions, the strength of electric fields that could safely be applied in the simulations, and how the results could be corrected to account for the effects of the finite box size and the underestimated viscosity of the simulated water (as discussed in detail in the previous chapter), simulations of the dependence of the electrophoretic mobility of HGs on fine structure were carried out. Figure 3.5 shows

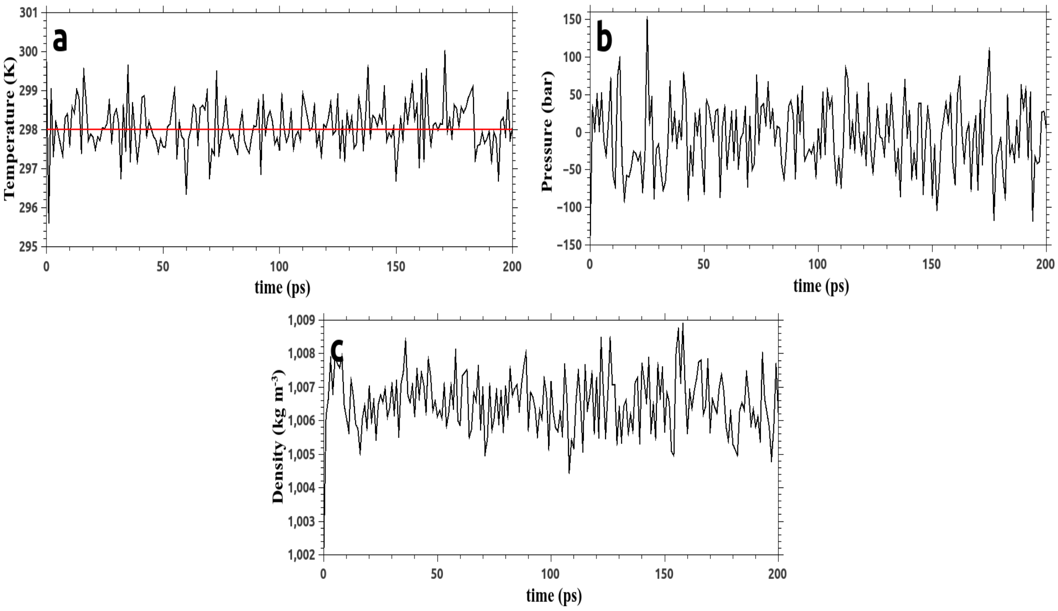


Figure 3.4: a) Temperature, b) pressure and c) density profile for a simulation of a fully charged 25mer.

Table 3.1: MD simulations details.

DM%	$\#_{HG}$	$\#_{HPO_4^{-2}}$	$\#_{H_2PO^-}$	$\#_{H_2O}$	$\#_{Na^+}$	length of box(nm)
20	1	36	36	79312	128	13.372
40	1	36	36	79309	123	13.372
48	1	36	36	79302	120	13.372
60	1	36	36	79300	118	13.372
80	1	36	36	79305	113	13.372

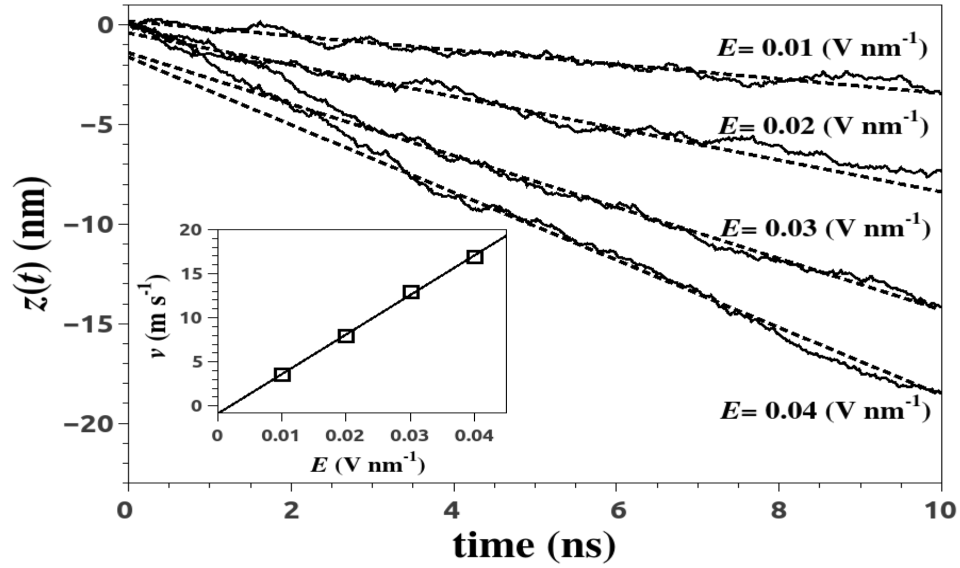


Figure 3.5: Illustration of the simulation procedure: Drift velocities, v, for molecules of interest are obtained from the slope of plots of the centre-of-mass coordinate, z, movement versus time, and plotted against applied field strength, E, (Inset), from which the electrophoretic mobility was obtained.

how a typical simulation was performed by monitoring the position of the centre-of-mass of the molecule in question as a function of time, extracting a drift velocity, and repeating the simulation while incrementing the strength of the applied electric field. From this electrophoretic mobilities were extracted (using $\mu = \frac{v}{E}$), and subsequently scaled as described to produce quantitative experimental predictions.

3.3 Electrophoretic mobility as a function of the charge patterning

Here the amount and distribution of galacturonic acid and methyl-esterified galacturonic acid residues along the chain are modified and the effect on the transport behaviour observed. Having different charge densities on the chain will change the ionic environment around the HG and it is expected that at some point, when the average distance between charge residues is less than Bjerrum length, counter-ion condensation will occur. This counter-ion coating is expected to reduce the electrophoretic mobility .

HG samples, which provide a substantial amount of the prior experimental data [94], typically have a degree of polymerisation(DP) of around 100 residues, while pectin samples (comprising of several HG domains that are linked together) are larger still.

Molecules this large are currently challenging to study *in-silico* becoming prohibitively time consuming to simulate as their size increases dramatically and it is not possible to do simulation with a very long chain. However, experimentally it is known that above a length of around 20-25 residues fully charged oligosaccharides with different degrees of polymerisation exhibit experimentally indistinguishable electrophoretic mobilities[67]. Moreover, the MD simulations carried out in the last chapter also show that the electrophoretic mobilities of oligoglacturonides are not dependent on DP above around 25. Simulating 25-mers then is a good compromise between investigating species containing enough residues for patterning to be investigated while gleaning results relevant to even larger HGs in a reasonable simulation time.

3.3.1 Methylesterified Substrates

The distribution of charged (G) and uncharged (M) residues can be randomly or blocky. Table 3.2 shows the fine-structure of several (DP=25) HGs, each with different specific sequences of methylesterification, whose electrophoretic transport was investigated using MD. Simulations were first carried out for several different molecules possessing the same chain-averaged DMs: typically two different sequences that had been generated *randomly* and one with a regular placement of methylesterification were investigated (Shown in Table 3.2).

Table 3.2: The sequence of residues for different DM substrates on which MD simulations were carried out (G; charged, M; uncharged).

DM%	Sample	Sequence of random residues																								
20 %	Sample 1	G	G	G	G	G	G	M	G	G	G	G	G	G	G	M	G	G	M	G	M	M	G	G	G	G
	Sample 2	M	G	M	G	M	G	G	G	M	G	G	M	G	G	G	G	G	G	G	G	G	G	G	G	G
	Regular	G	G	M	G	G	G	G	M	G	G	G	G	M	G	G	G	G	M	G	G	G	G	M	G	G
40 %	Sample 1	M	M	M	M	G	G	G	G	G	G	M	G	M	G	M	G	G	M	G	M	G	G	G	M	G
	Sample 2	G	G	M	G	M	G	M	G	M	G	G	G	M	M	M	G	G	G	G	M	G	M	G	G	M
	Regular	G	M	G	G	M	G	M	G	G	M	G	M	G	G	M	G	M	G	G	M	G	M	G	G	M
48 %		M	M	M	G	M	M	M	M	G	G	G	G	M	G	G	M	G	M	M	M	G	M	G	G	G
60 %	Sample 1	M	G	G	M	G	G	G	M	M	M	G	M	M	M	M	G	M	M	G	M	G	G	M	M	M
	Sample 2	G	G	G	M	M	M	G	G	M	M	M	M	M	G	G	M	G	M	M	G	G	M	M	M	M
	Regular	M	G	M	M	G	M	G	M	M	G	M	G	M	M	G	M	G	M	M	G	M	G	M	M	G
80 %	Sample 1	G	M	M	G	G	M	M	G	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	G	M
	Sample 2	G	M	M	M	G	M	M	G	M	G	M	M	M	M	M	M	M	M	M	M	M	G	M	M	M
	Regular	M	M	G	M	M	M	M	G	M	M	M	M	G	M	M	M	M	G	M	M	M	M	G	M	M

DM%	Sample	Sequence of blocky residues																								
48 %		M	M	G	G	G	G	G	G	G	G	G	G	G	M	G	M	M	M	M	M	M	M	G	M	M
60 %		M	M	G	M	M	M	M	M	M	M	M	M	G	M	G	G	G	G	G	G	G	G	M	M	M

Figure 3.6 shows the results of these molecular dynamics simulations superimposed on experimental data[94] of electrophoretic mobility versus fractional charged residues, z, as shown in Figure 3.6. The quantitative agreement obtained with only one variable parameter is remarkably good, and suggests that the MD simulations capture the important physics of the transport of these polyelectrolytes well. (This variable parameter can be incorporated as a scaled viscosity and in this case would correspond to an increase compared to water to a value of around 1.06 mPa s, some 20% larger than that of pure water, but eminently reasonable considering the local concentration of polymer present in the sample zone of the separation.) Motivated by this observation we examined more closely the spatial distribution of ions around the HGs and their dynamics in order to illuminate how these changes ultimately yield the observed transport behaviour particularly for systems with $z > 0.6$ where there is a possibility of having counterion condensation around the chain.

Figures 3.7 (a and b) show the number density of counterions as a function of the radial distance from the chain, R, obtained for several HGs. In all cases, a regime is observed where the ion concentration varies as a 'diffuse layer', that is; as predicted by the Poisson-Boltzmann (PB) equation described in chapter 1 . Fitting the data in this region to the PB model allows the extraction of a polyelectrolyte parameter, ξ, which gives the ratio of the Bjerrum length(l_B) to an average charge spacing along the backbone(b)($\xi = \frac{l_B}{b}$). In cases where the HG samples were highly methylesterfied (>40%) the average charge spacing predicted in this fashion was indeed found to be close to that which would have been predicted simply from counting every unmethylesterified carboxyl group as a single electronic charge and using inter-charge spacings obtained directly from the molecular structure.

However, as the putative charge spacing calculated in this way decreases and becomes less than the Bjerrium length, the ions in the diffuse regime appear from their distribution to be surrounding a polymer of significantly lower charge density than expected. These more sparsely methylesterifed samples have an chain-averaged polyelectrolyte parameter of greater than 1 and as such are indeed predicted to exhibit counterion condensation yielding an ion-decorated chain with reduced charge density compared to the naked polyelectrolyte. Details of the results of these simulations are given in Table 3.3, including a comparison of the linear charge density, λ, (electronic charges per nm) both naively calculated from the structure and extracted from the fits to the simulation results as described.

It has been suggested previously that, by monitoring a point of inflection in a plot of the fraction of ions found within cylinders of certain radii as a function of the radii,

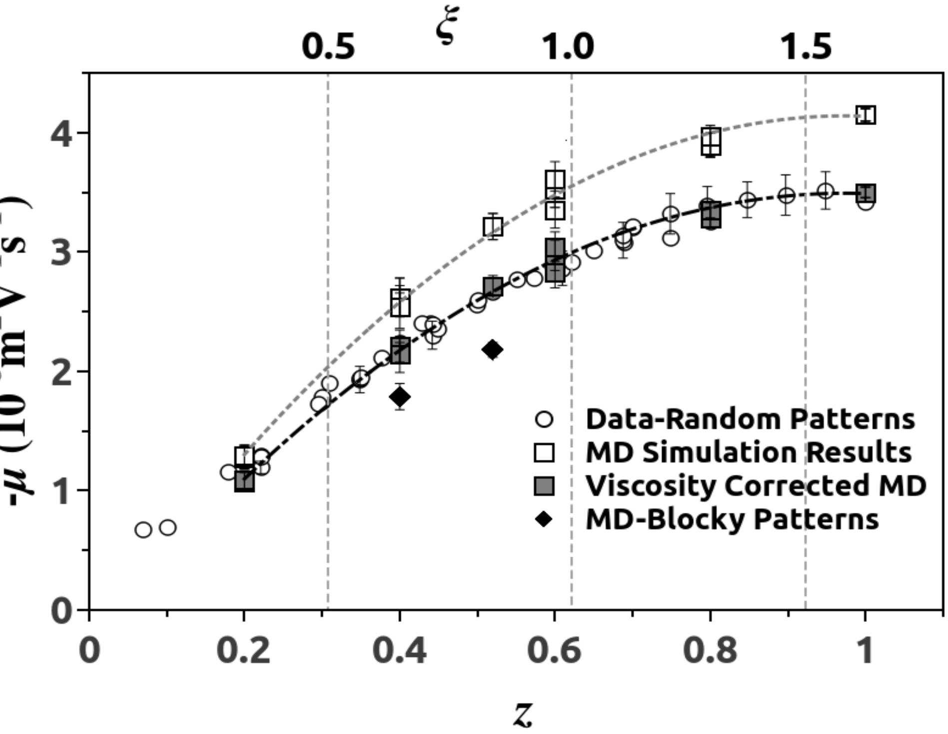


Figure 3.6: Electrophoretic mobilities, μ, measured for HGs and a number of pectin samples, as a function of the fraction of the sugar rings charged, z, compared with the results of MD simulations described herein. The lines are splines, added as a guide to the eye.

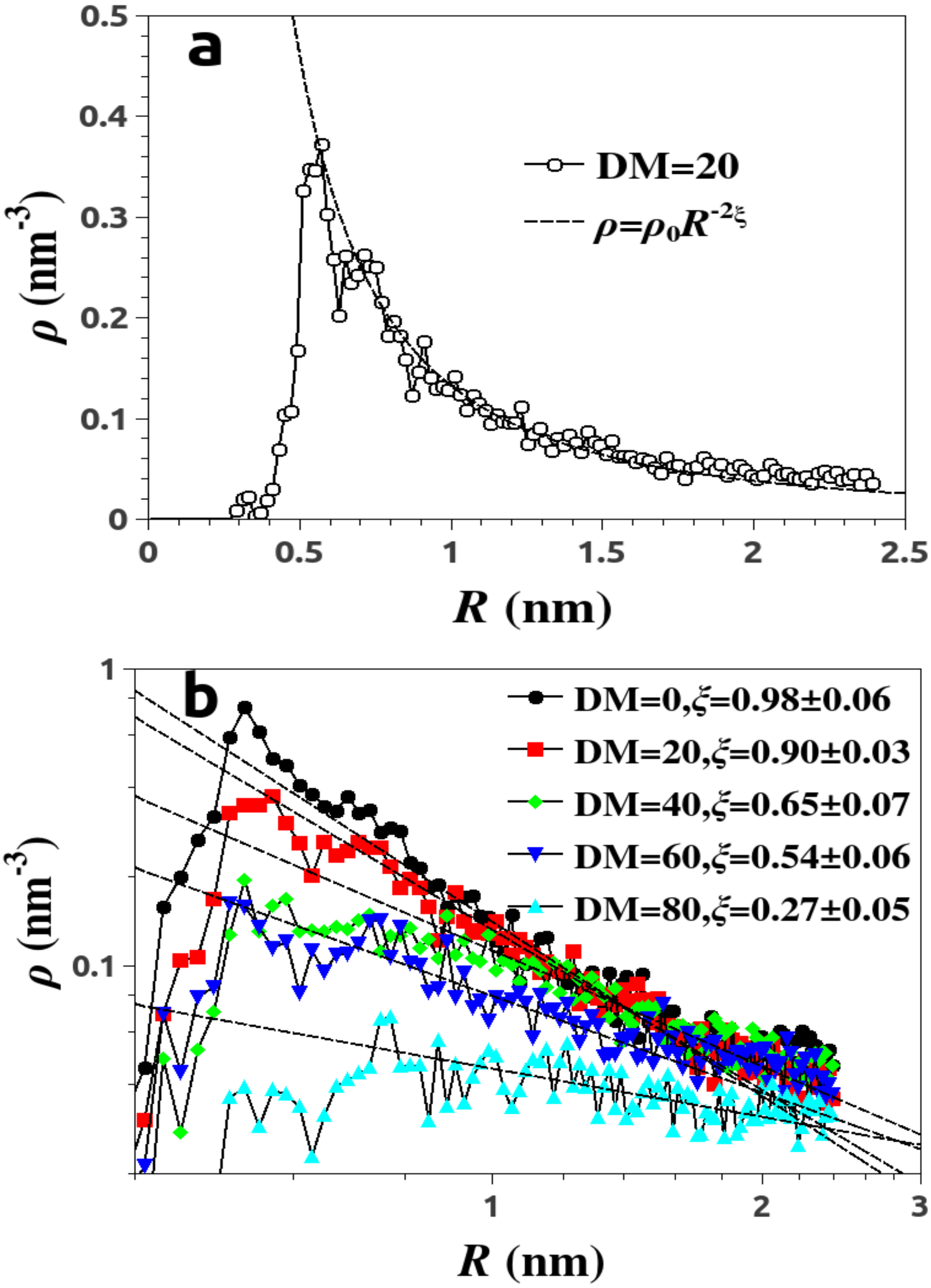


Figure 3.7: Number density (ρ) distributions for counterions as a function of closest distance from the chain, R, obtained for several randomly methylesterifed HGs with different DMs. a), linear; b), logarithmic. Lines show fits to equation $C = C_0 R^{-2\xi}$ allowing the extraction of ξ.

DM%	Sample	λ	$\xi = l_B\lambda$	E_{ext}	ξ_{eff}	$\lambda_{eff} = \frac{\xi_{eff}}{l_B}$	$\lambda_{Q_{ions}}$
0 %	-	2.016	1.608	0	0.98±0.06	1.20±0.08	0.727
				0.01	0.95±0.04	1.19±0.05	0.755
20 %	Sample1	1.605	1.280	0	0.91±0.04	1.14±0.05	0.437
				0.01	0.91±0.05	1.14±0.06	0.482
	Sample2	1.597	1.273	0	0.90±0.03	1.13±0.04	0.486
				0.01	0.89±0.07	1.12±0.09	0.496
	Regular	1.608	1.282	0	0.99±0.04	1.24±0.05	0.357
				0.01	0.90±0.05	1.13±0.06	0.436
40 %	Sample1	1.182	0.942	0	0.65±0.07	0.81±0.09	0.292
				0.01	0.64±0.03	0.80±0.04	0.310
	Sample2	1.187	0.947	0	0.65±0.07	0.82±0.09	0.327
				0.01	0.70±0.04	0.88±0.05	0.296
	Regular	1.223	0.975	0	0.65±0.04	0.82±0.05	0.294
				0.01	0.71±0.07	0.89±0.09	0.302
48 %	-	1.021	0.814	0	0.64±0.03	0.80±0.04	0.189
				0.01	0.64±0.03	0.80±0.04	0.181
60 %	Sample1	0.79	0.630	0	0.55±0.06	0.69±0.08	0.116
				0.01	0.53±0.05	0.64±0.06	0.147
	Sample2	0.788	0.628	0	0.54±0.06	0.67±0.08	0.140
				0.01	0.49±0.09	0.6±0.1	0.172
	Regular	0.753	0.601	0	0.41±0.07	0.51±0.09	0.057
				0.01	0.51±0.05	0.65±0.06	0.146
80 %	Sample1	0.369	0.294	0	0.30±0.02	0.38±0.03	-0.024
				0.01	0.29±0.07	0.36±0.09	0.031
	Sample2	0.357	0.285	0	0.27±0.05	0.34±0.06	0.041
				0.01	0.28±0.06	0.35±0.08	-0.009
	Regular	0.355	0.283	0	0.29±0.07	0.36±0.09	-0.020
				0.01	0.27±0.03	0.34±0.04	-0.017

Table 3.3: Linear charge density calculated from the structure $\lambda = \frac{e}{b}$ (e nm^{-1}), polyelectrolyte charge parameter, ξ, external electric field applied during calculation in V nm^{-1}, effective charge parameter, ξ_{eff}, calculated by fitting $\rho_0 R^{-2\xi}$ to the simulation results in the Gouy region, corresponding effective linear charge density, λ_{eff}, and the linear charge density of 'condensed' ions within the Bjerrum length of the chain.

a so-called Manning radius can be defined within which ions are taken as condensed [99]. However, in these simulations, which were primarily concerned with the local environment around charged regions of increasing but limited extents, such an approach is complicated by the heterogeneity of the polyelectrolytes of interest. In addition the HG systems investigated here have a maximum ξ of around 1.6, and are exclusively for monovalent ions. Under these conditions the inflections clearly seen at higher values of ξ or for multivalent ions [99] are difficult to locate. For these reasons we turn to the examination of the restricted nature of the thermal fluctuations of the ions, specifically the space they explore relative to the Berjum length, as a pragmatic demarcation between those that are free or condensed. In order to provide further evidence that simulations capture counterions condensing onto the more highly charged HGs, and that it is this phenomena that explains both the changes in the electrophoretic mobility and the modified distribution of ions observed in the diffuse layers, the mobilities of ions at different proximities from the chain were examined.

Figure 3.8 (a-c) shows how the radial distance of ions from the chain backbone changes over the timecourse of the simulations; the data points represent the average, and the bars represent the standard deviation of the fluctuations in the position. When the degree of methylesterification is high (c: 80%) (the charge density is low) all ions appear relatively mobile and none of them spend long times within a Bjerrium length of the HG backbone (indicated by the dashed line and the shaded section). In contrast, as the degree of methylesterification decreases (b: 40% and a: 0%) (so that the charge density increases and the naively calculated polyelectrolyte parameter approaches or passes one respectively) the motion of several counterions become constrained. In particular for the completely charged chain several ions are now found only in the proximity of the chain (closer than the Bjerrium length) over the entire length of the simulation. These are identified as the *condensed* ions. While this assignment is somewhat pragmatic it finds justification by observing that the linear charge density of ions selected in this manner closely resembles the amount by which the linear charge density of the polyelectrolyte, λ, is found to be modified (table 3.3). That is; selecting condensed ions according to this definition and using their linear charge density to calculate the appropriate value for the *ion-decorated* chain is consistent with the reduced value of linear charge density extracted using the simulated ion distributions.

3.3.2 Substrates Exhibiting Blockwise Demethylesterification

Finally attention is focused on results from samples where, although the overall chain-averaged degree of methylesterifcation is not low enough for significant condensation

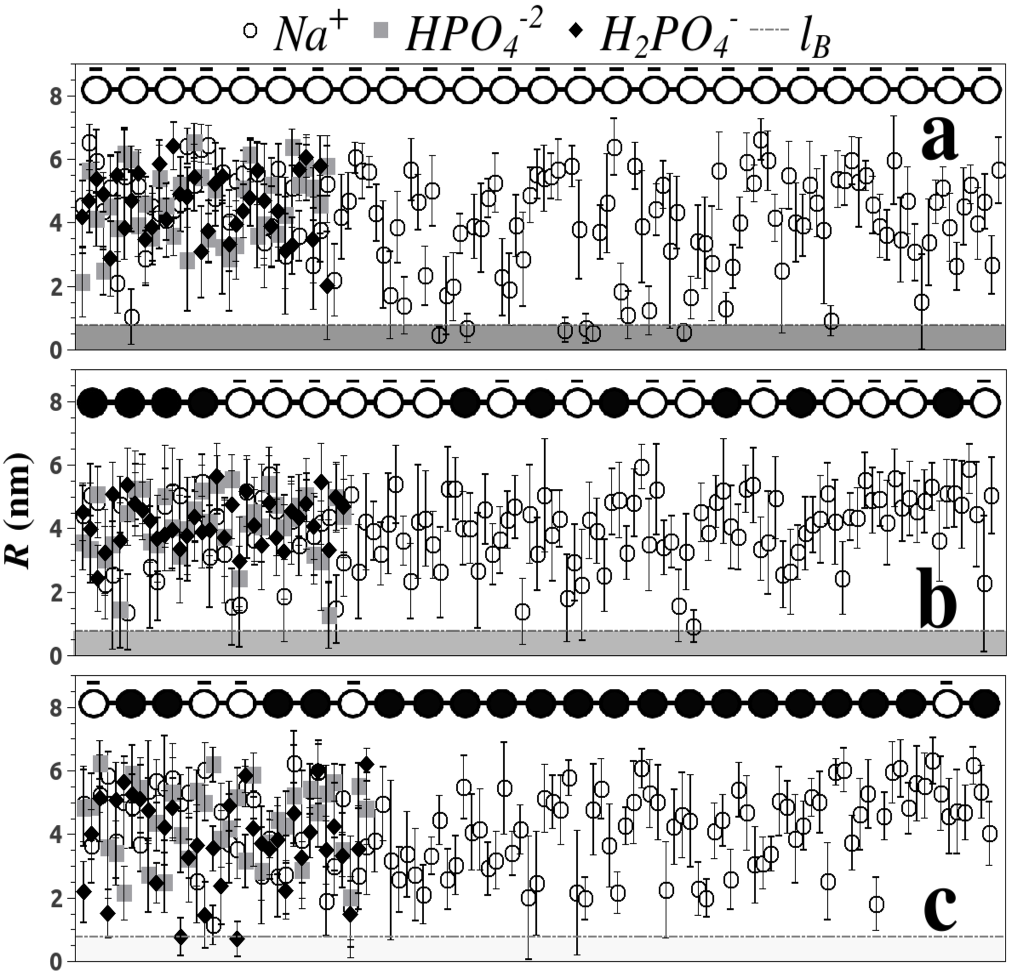


Figure 3.8: The radial distance of each ion from the chain backbone over the timecourse of the simulations (a, b and c; DM 0, 40 and 80 % respectively). Data points show the average position, and the bars represent the standard deviation of the fluctuations in the position over whole simulation. The dashed line denotes the Bjerrum length, and the greyscale-darkness of the region of space in closer proximity to the chain represents the fraction of the backbone charge effectively neutralised by the counterions.

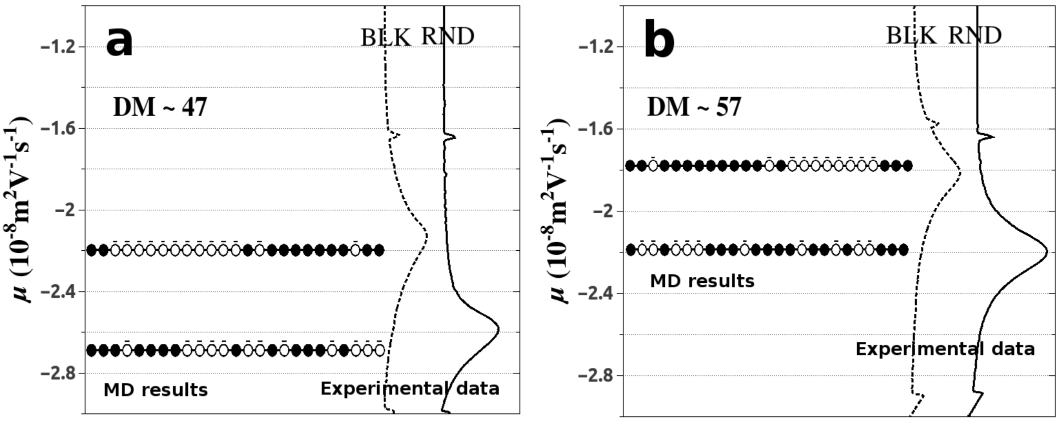


Figure 3.9: Measured electrophoretic mobility distributions of DM ∼ 47(a) and DM ∼ 57% (b), along with pictures of fine-structures that have been investigated in MD, placed at a position corresponding to the resulting calculated mobility. Experimental data from [3]. BLK and RND refer to blocky and random distribution respectively.

to take place *if* the distribution of the charged groups along the chain was random, high localized charge densities can nevertheless occur. In this case these locally blocky charge distributions have been introduced into the fine structure in the experimental samples by a processive enzyme [112].

As can be seen in Table 3.2, such sequences do possess local blocks of unmethylesterified regions. Figure 3.9 (a) and (b) shows the comparison of MD simulation result (this work) and experimental output[3], with the time-axis transformed to electrophoretic mobility, overplayed with several different fine structures that have been placed in the figure according to their predicted mobilities that have been obtained through MD simulation. It is clear that in the DM range examined the differences in the electrophoretic mobility of samples possessing random or blockwise intramolecular distributions of charge can indeed be explained by local counterion condensation, and is captured by the MD. The results of the MD simulations performed on these blocky fine structures are also shown in figure 3.6.

So, how blocky does an HG have to be to induce *local* counterion condensation, thereby modifying the *local* distribution of ions enough to affect the measured electrophoretic mobility of the chain? Figure 3.10 shows how the positions of ions (relative to the charged block) change over the time-course of the simulations, specifically in the proximity of charged blocks of different lengths sandwiched between two runs of methylesterified residues(with 4 methylesterified residues at each side). Once again the points show the average position of ions, while bars show the standard deviations of the fluctuations. Note that in previous figures R denotes the closest distance between an ion and any backbone atom while here r denotes the closest distance between an ion and any atom *within the charged block*.

It should be noted that, in contrast to the simulations carried out previously, in some of these cases partially-condensed species could be observed; that is, ions that spend a considerable amount of time within a Bjerrum length of the chain before escaping, as shown in figure 3.11. In such cases the degree of charge that such a block contributes to that reflected by the electrophoretic mobility of the species is approximated by subtracting the charge of the condensed ions averaged over the timecourse of the simulation away from that of the residues themselves.

Finally, figure 3.12 shows the fraction of the galacturonic acid residues within a contiguous block that host a condensed ion, as the local block-length increases. Once the block-length reaches around ten residues the ion-decorated block only manifests some 65% of the nominal charge of the galacturonic residues, consistent with the behaviour observed for longer galacturonic oligomers, and in line with the $1 - (\frac{1}{\xi})$ predicted [99]. In addition an empirical fit to the data below block lengths of ten enables an approximate prediction of the electrophoretic mobility of any fine structure, from its block length distribution, taking counterion condensation into account without the necessity of carrying out further MD simulations.

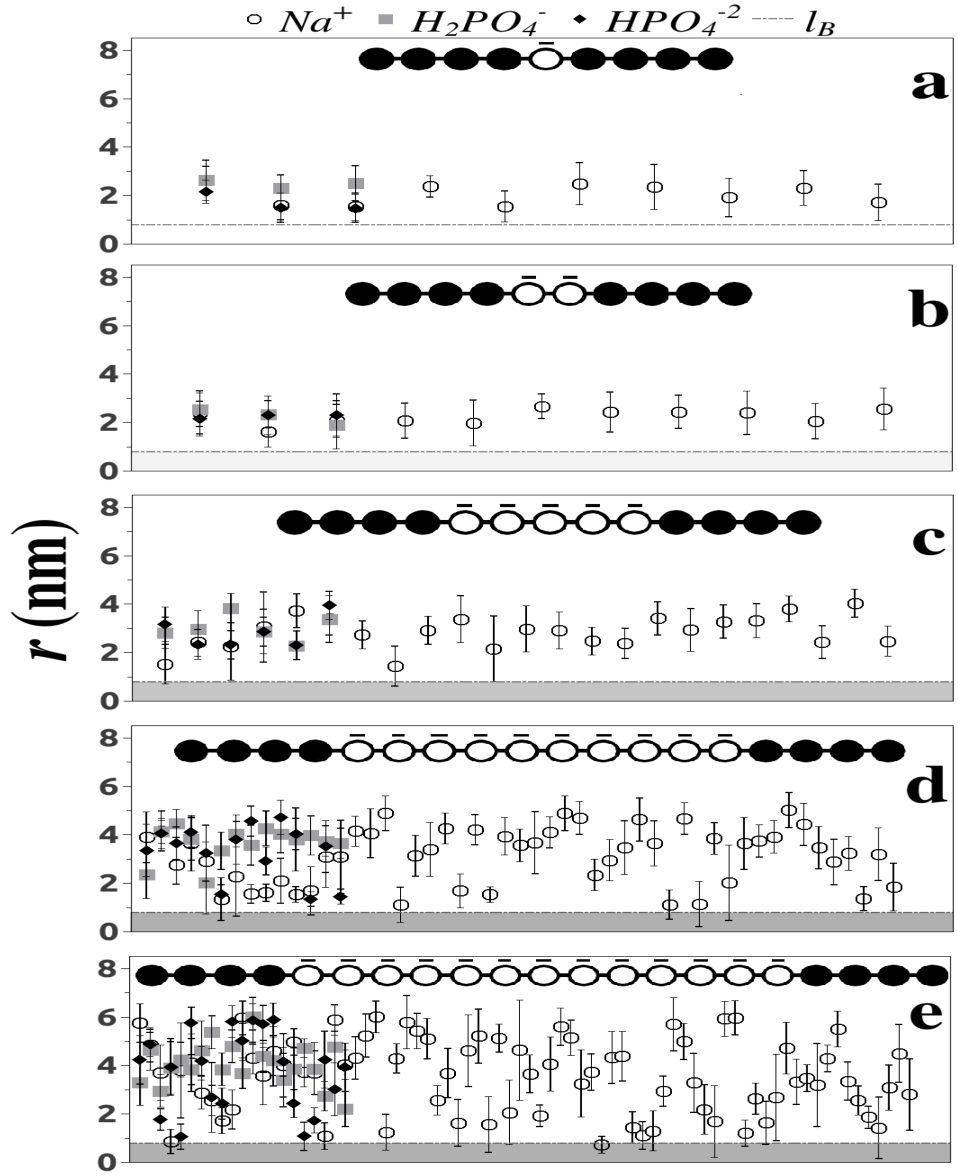


Figure 3.10: Representative results of the distance of counter-ions and co-ions, r, from charged blocks of length: 1 (a), 2 (b), 5 (c), 10 (d), and 13 (e), sandwiched between small methylesterfied blocks as shown. (Note that in previous figures R denotes the closest distance between an ion and any backbone atom while here r denotes the closest distance between an ion and any atom *within the charged block*). The dashed line denotes the Bjerrum length, and the greyscale-darkness of the region of space in closer proximity to the chain represents the fraction of the backbone charge effectively neutralised by the counterions.

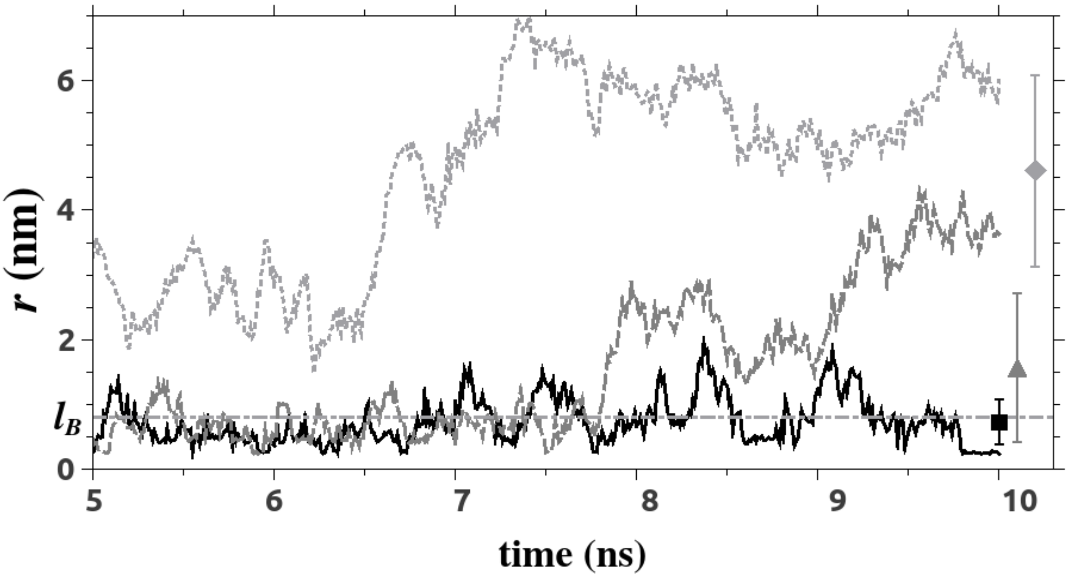


Figure 3.11: The behaviour of 'free', 'partially condensed' and 'condensed' ions over the course of the simulation shown in figure 3.10 (e) (gray, dark gray and black respectively). Data points show the average position, and the bars represent the standard deviation of the fluctuations in the position over the presented time.

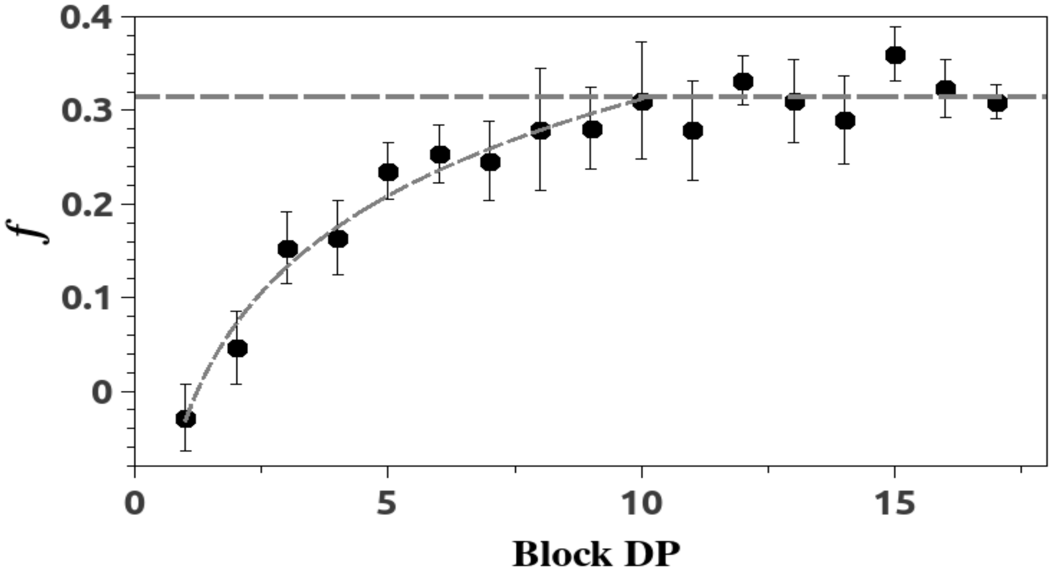


Figure 3.12: The fraction of charge of galacturonic blocks that is negated by counterion condensation (the fraction of the galacturonic acid residues within a continguous block that host a condensed ion) as a function of block length, together with a fitted empirical relationship to data for block lengths less than 10 residues (A Log(Bx) where A = 0.34 ± 0.02 and B = 0.81 ± 0.07), dashed line is $1 - \frac{1}{\xi}$.

Chapter 4

Conformation of polygalacturonic acid

In the previous chapters the electrophoretic mobility of individual HG chains with diverse charge patterns and different degrees of polymerisation was studied. In this chapter it is focussed on examination of chains' conformation and the way that they arrange themselves in solution at different concentrations and under different ionic concentrations. Specifically these studies have been undertaken with a view to understanding the results of SAXS experiments carried out on these polyelectrolyte solutions.

One of the key parameters that can be used to describe the conformation of HG chains is the dihedral angles between sugar residues, represented in figure 4.1 and defined as:

$$\begin{aligned} \phi &= O'_5 - C'_1 - O_4 - C_4 \\ \psi &= C'_1 - O_4 - C_4 - C_5. \end{aligned} \tag{4.1}$$

Whether the individual sugar residues are charged or uncharged has an influence on these angles and indeed can thereby change the preferred conformation of the chains, the extremes being two-folded(2^1) or three-folded(3^1) helical conformations. (For the two-folded conformations after traversing two residues along the chain the same orientation of sugar residue is recovered, whereas for three folded conformation three residues must be traversed to get the same orientation of the following sugar residue to the initial one). These conformations can be recognised by examining the direction of C_5-C_6 carbons relative to each other in adjacent sugar residues. In the case of the two-folded conformation(2^1) the angle between C'_5-$C_{6'}$ and C_5-C_6 is about 180 degrees , while for the three-folded conformation this angle is around 120 degrees. In highly

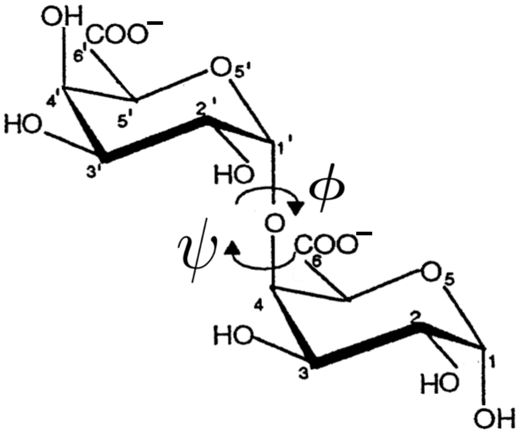


Figure 4.1: Schematic representation of disaccharide along with the dihedral angles ϕ and ψ.

charged chains the electrostatic forces between carboxyl groups in adjacent residues favour a maximum distance between charged parts, achieved by a 180 degree angle between them, and so favour a two-folded conformation. With the residues uncharged however, the three folded conformation is preferred. Having different chain conformations can change the interaction of chains with each other and with the surrounding environment and thus is a crucial consideration in the detailed molecular description of these macromolecules[113, 114, 115].

4.1 Conformation of dimers

The distribution of dihedral angles as described in equation 4.1 for a charged and protonated dimer of galacturonic acid is shown in figure 4.2. 20 ns of simulation has been performed (with the first ns considered as a equilibration time and not included into the calculations). Frequencies with which particular dihedral angles are present in the thermal fluctuations, for charged or protonated dimers were calculated and are shown in figure 4.2. It can be seen that indeed there are two main configurations, for charged or for protonated dimers, which the fluctuations are centred about, and that these are those referred to as two- and three-folded conformations respectively. It should be noted that these results hold regardless of the starting conformation i.e. if the initial conformations of the dimers were 3^1 then the protonated dimer stays in this configuration, (black area in the figure 4.2) while the charged dimer quickly changes its conformation (red area) and vice versa.

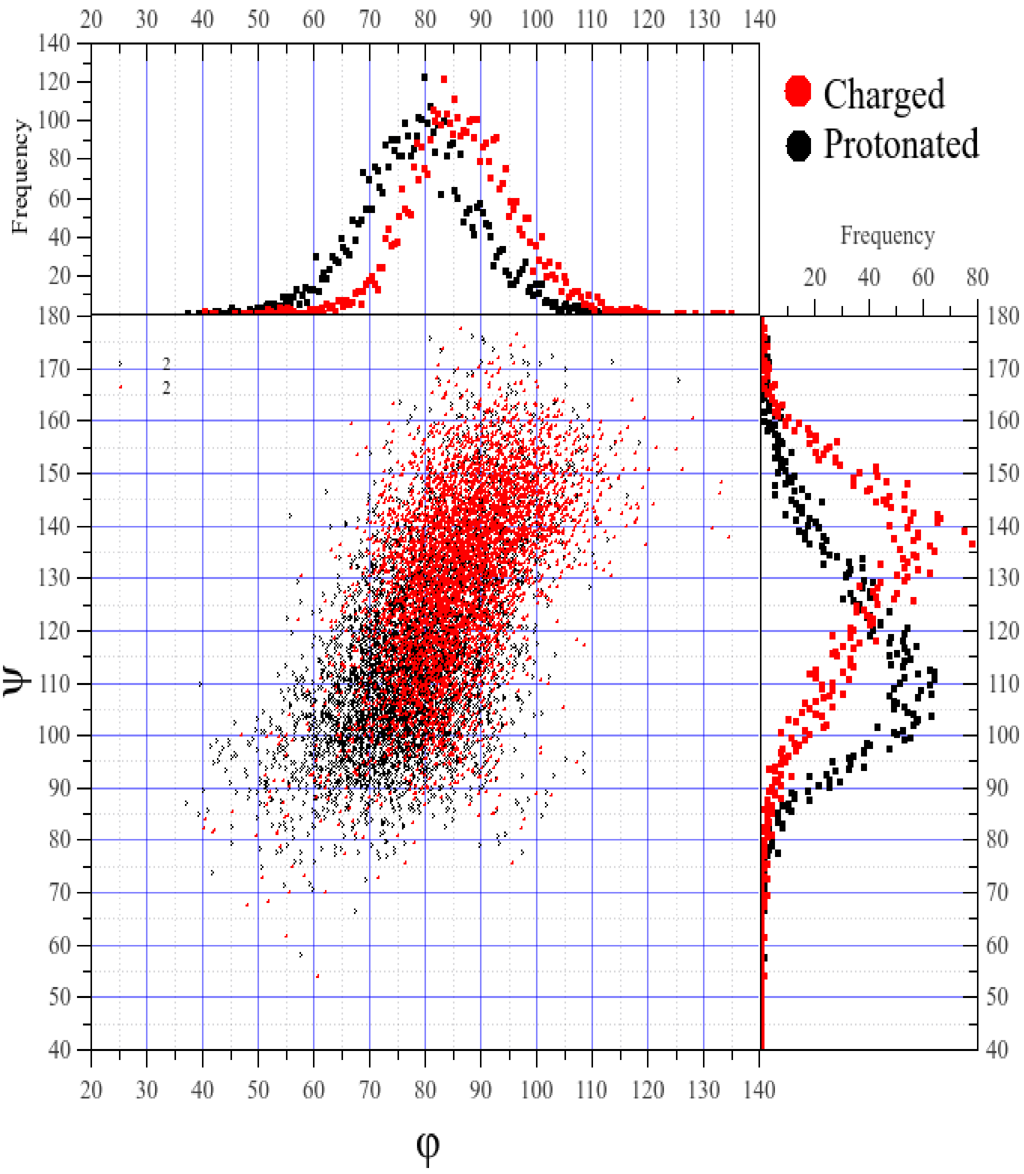


Figure 4.2: ϕ vs ψ plot for Charged and Protonated dimer in the simulation and their frequency.

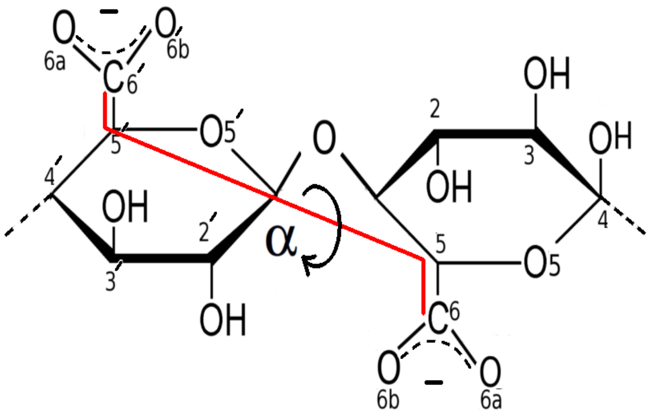


Figure 4.3: Schematic representation of pseudo-dihedral angle α.

4.2 Conformation of Longer Chains

As it is discussed in the previous section there is a major difference between helical conformation of charged and uncharged HGs, which basically results from the competition of intra-residue electrostatic repulsion and hydrogen bonding. In this section the helical conformation of dp6 and dp25 chains is examined. As well as calculating dihedral angles around the glycosidic bond in dimer discussed in the previous section, to investigate the orientation of the residues along the chain in higher DPs and to help with the interpretation we introduce a pseudo-dihedral angle α as;

$$\alpha = C_6' - C_5' - C_5 - C_6 \tag{4.2}$$

for the longer chains as it is shown in figure 4.3. In fact if you look at the chain through the main axis this pseudo-dihedral angle, α, is the angle between C_5-C_6 bond in a residue with C_5-C_6 in the neighbouring residue.

Defining α in this way helps to distinguish between a predominantly two-folded conformation, with almost 180 degrees for α, and a three-folded conformation with α=120 degrees. To study the effect of initial conformation, simulations were performed for both charged and protonated chains of DP 6 with different initial helical conformations where 3^1 conformation taken from carbohydrate builder provided in www.glycam.org and the 2^1 conformation came from ref. [116]. PyMOL software [79] used for protonation of chains.

As can be seen in figure 4.4, regardless of initial conformation, the normalised distribution of the absolute value of α in each ns for the protonated chain is centred around 120 (again the favoured conformation for a protonated chain is 3^1). In contrast for charged chains and there are a good chance of the chain being in the 180 configuration, although it should be noted that the α distribution for the charged chain is spread over a large range of angles showing a large flexibility originating largely from the terminal residues.

The same calculation of α was carried out for DP 25 chains and the comparison with DP 6, is shown in Figure 4.5. Indeed for the longer chain there is an increased chance of the 180 configuration and less flexibility to move into other areas of configuration space.

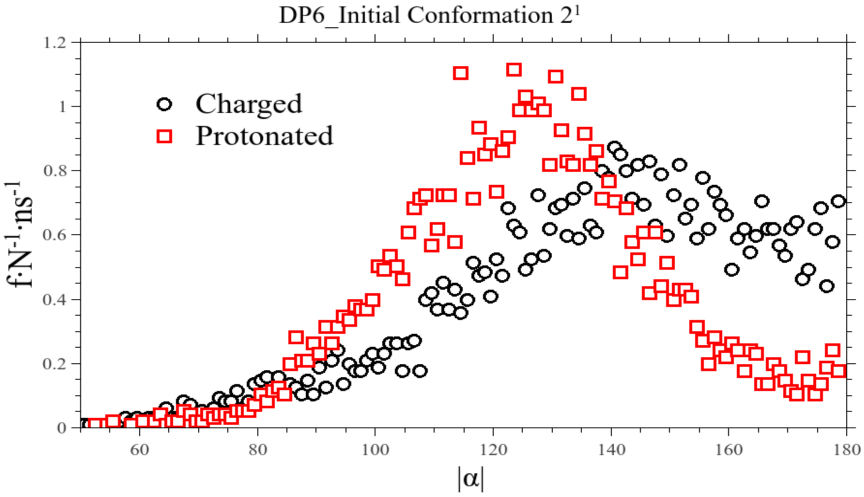
The conformations of HG chains and their dependence on the charge distribution is exemplified by observing them in real space. Figure 4.6 shows the 3^1 conformation, together with a view through the central axis, and the 120 degree conformation of carboxyl group can clearly be seen. For a charged chain this conformation quickly changes during the MD simulation and arrives at a 2^1 conformation, as it is shown in figure 4.7. (It should be remembered that especially for smaller oligomers there is substantial flexibility as discussed - See Figure 4.4).

Figure 4.8 shows the converse, that taking the the 2^1 conformation as the starting configuration for a protonated chain it quickly reverts to is preferred 3^1 conformation (figure 4.9).

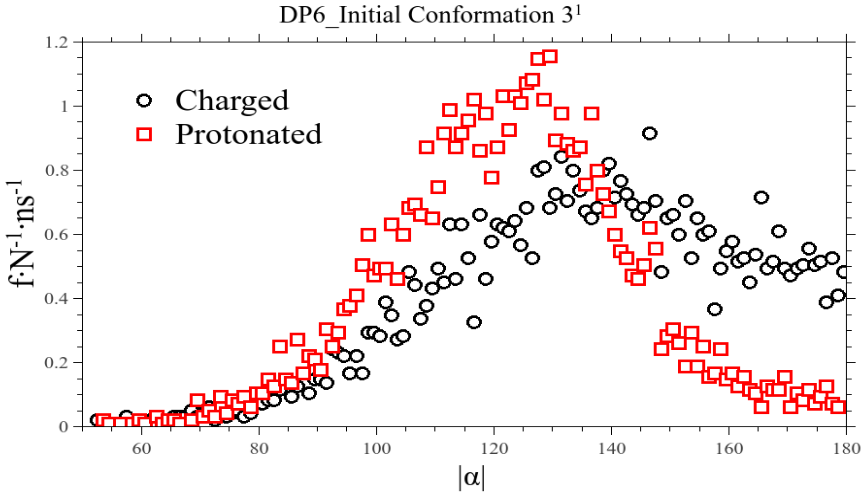
Experimentally one good method for studying the shape of molecules is Small Angle X-Ray Scattering(SAXS) which can give an idea of the geometry of molecule at atomic resolution. In the next section MD simulations are investigated as a methodology for producing configurations that can be used to calculate SAXS profiles from HG chains, both individually in a dilute solution, as well as at higher concentrations - with multiple HG chains in the simulation box.

4.3 Introduction to SAXS

Small Angle X-Ray Scattering(SAXS) is a common experimental methodology for studying the structure of molecules in solution and it is widely used to do research on the shape and conformation of scatterers. But like all experimental investigations, the usefulness of SAXS is restricted by the accuracy of the experimental apparatus. Reducing noise in the SAXS profile for some systems is not an easy job and sometimes is impossible. Computational studies can help to understand what the scattering of an ideal system would look like.



(a) Initial 2^1



(b) Initial 3^1

Figure 4.4: Distribution of pseudo-dihedral angle, α, for charged and protonated DP 6 with different initial conformations. N = 5, is the number of residues linkages in the chain.

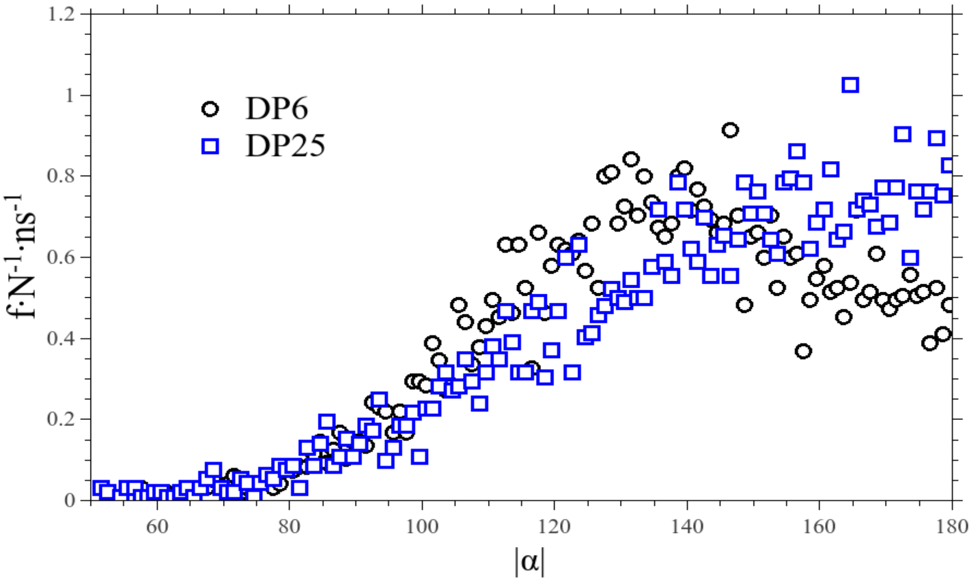


Figure 4.5: Comparison of normalised frequency of α for dp25 and dp6.

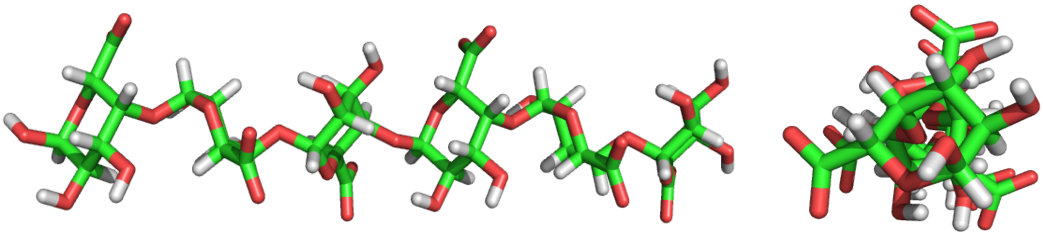


Figure 4.6: 3^1 conformation considered as an initial conformation for fully charged DP6.

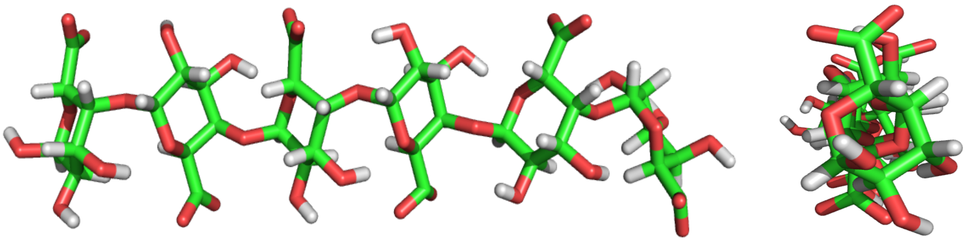


Figure 4.7: The common conformation of fully charged DP 6 after equilibration time of simulation.

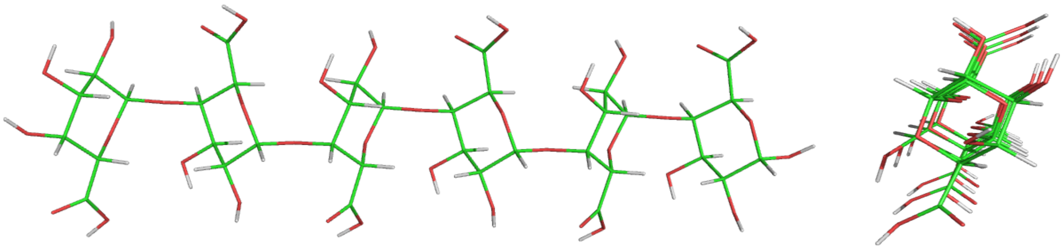


Figure 4.8: 2^1 conformation considered as an initial conformation for protonated DP6.

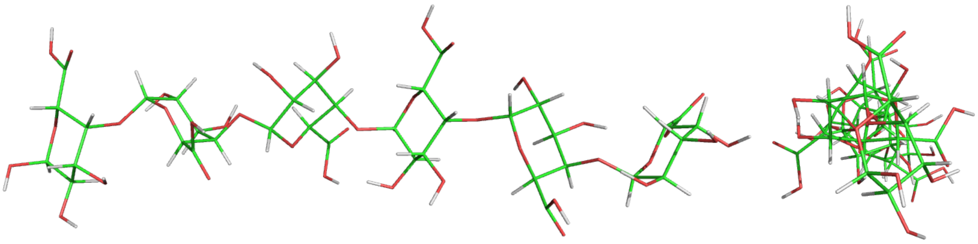


Figure 4.9: The common conformation of protonated charged dp6 after equilibration time of simulation.

Many theoretical studies, such as the electrostatic blob model, proposed by De Gennes et. al. [117], present a complex structure at small length-scales. These studies primarily focus on flexible, model polyelectrolytes, although how relevant such models are for common, highly charged, and semi-flexible polyelectrolytes is unclear. Knowledge of the solution state structure of macromolecules is important, not only from a fundamental perspective, but also for the design of foods, pharmaceuticals and even the often heralded organic electronics. Molecular Dynamics(MD) can help to understand more about the solution state structure of these macromolecules.

Molecular Dynamics(MD) is a method which applies to molecular level studies, calculating simple mechanical potentials on each atom to find atomic trajectories. Doing MD on a particular system will make it possible to calculate SAXS profiles and compare with the experimental data. To calculate SAXS profile from MD trajectory after 1ns of simulation (which considered as an equilibrium time) a series of atomic coordinate files(pdb files) were made skipping some steps to reduce the analysing time and then feed these pdb files to CRYSOL software [118] to calculate SAXS profile of each individual snapshots of MD. Finally an average over all calculated SAXS profiles was made to get the SAXS profile of a particular solution. MD simulations are becoming more accessible with help from High Performance Computing facilities (like NeSI). In this research systems with a size of up to about 20nm cubed were studied.

Figure 4.10 schematically shows how a scattering process works. An incident beam

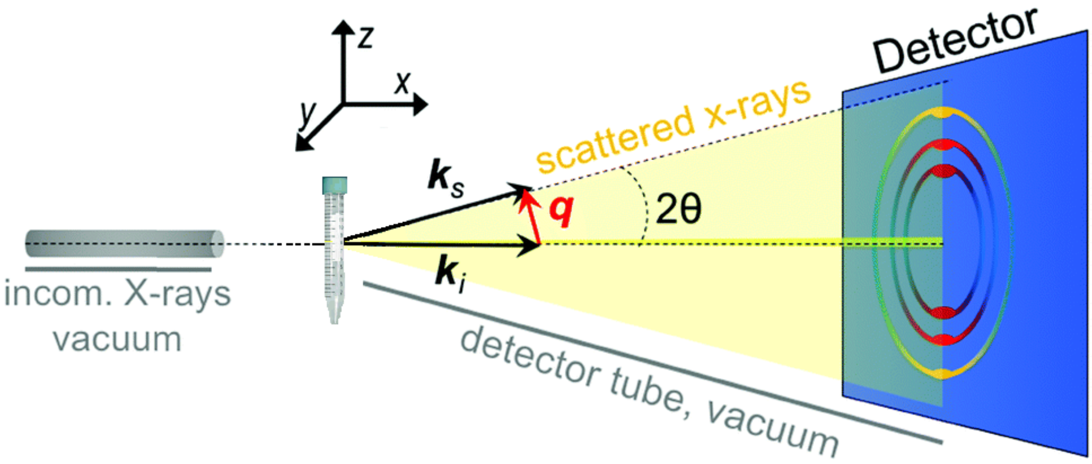


Figure 4.10: Schematic view of scattering set-up (Adapted from ref.[4]).

(indicated by $\mathbf{k}_i$) scatters: hitting the sample and being deflected (shown by $\mathbf{k}_s$). Then, the scattering vector, $\mathbf{q}$, is;

$$\mathbf{q} = \mathbf{k}_s - \mathbf{k}_i \tag{4.3}$$

Since the magnitude of incident and scattered wave vector are equal, $\frac{2\pi}{\lambda}$, and $|\mathbf{q}| = 2|\mathbf{k}|\sin\theta$,(Figure 4.10) then;

$$|q| = \frac{4\pi\sin\theta}{\lambda} \tag{4.4}$$

where θ is the half of angle between incident beam and scattered beam, and λ is the wavelength. Using Bragg's equation, $\lambda = 2d\sin\theta$,:

$$|q| = \frac{2\pi}{d} \tag{4.5}$$

where d is the distance scale in real space between scatterers inside the solution and q is the scattering vector.

Depending on the shape and size of objects the scattering pattern on the detector will change. Scattering patterns are usually displayed using radial integrals around the central point as intensity versus scattering vector. Figure 4.11 shows typical x-ray scattering profiles of different objects in nano-meter size range. While there is a mostly linear dependence of the SAXS pattern for a rod shaped object, there are resonances for more complicated objects such as spheres that depend on the size of the objects. These graphs are helpful in order to compare SAXS profile of samples to get an idea of configuration of particles inside it. For example having a SAXS profile with a slope of one shows that in real space there are rod shaped objects inside sample, or it is possible to compare the spherical SAXS profile with some SAXS data of some folded globular

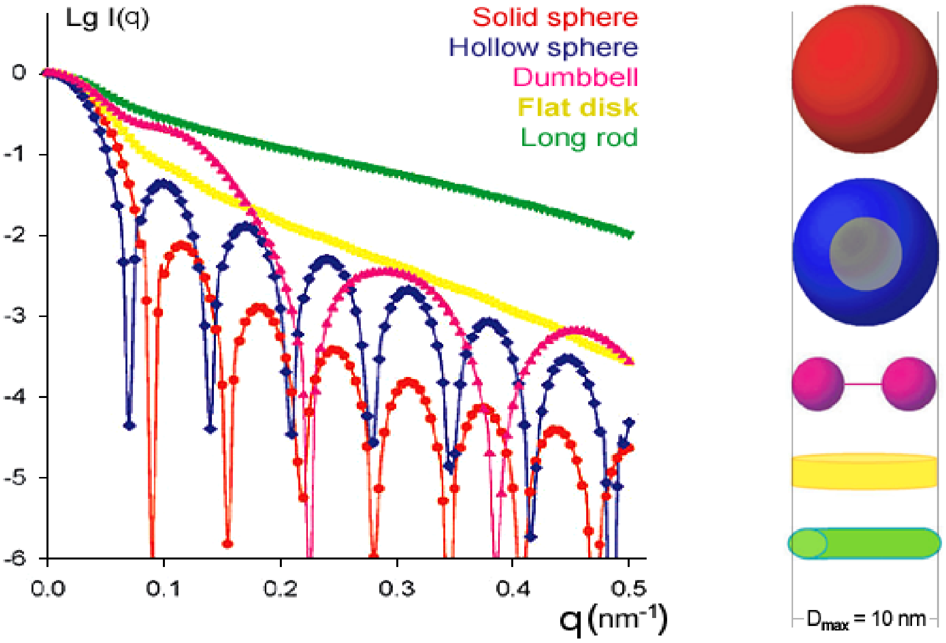


Figure 4.11: Scattering profile of different objects in nano-meter scale. Adapted from ref. [5].

proteins [5, 119].

4.3.1 Single Chain

In this section SAXS profiles of single chains are calculated, using trajectories from MD simulations, and compared with experimental data obtained by this research group at the Australian Synchrotron. Theoretical SAXS profiles are generated from the trajectories using CRYSOL software[118] described in the previous section.

Figure 4.12 shows the experimental SAXS profile for a fully charged DP 6 oligomer along with theoretical calculations based on MD simulations performed using two different force fields, GLYCAM and AMBER. In this dilute regime the scattering profile is conveniently discussed in two main sections: a low q region ($q \lesssim 1Å^{-1}$) and a high q region (q vales $\gtrsim 1Å^{-1}$) indicated in the figure by a shaded area. In the low q region an excellent agreement between the experimental data and theoretical calculation based on MD simulation is found (using either force fields).

However in the high q region (See inset in figure 4.12), there are clear differences in the SAXS profiles taken from MD simulations that used AMBER or GLYCAM force fields. This is perhaps not surprising given that at these extremely small length

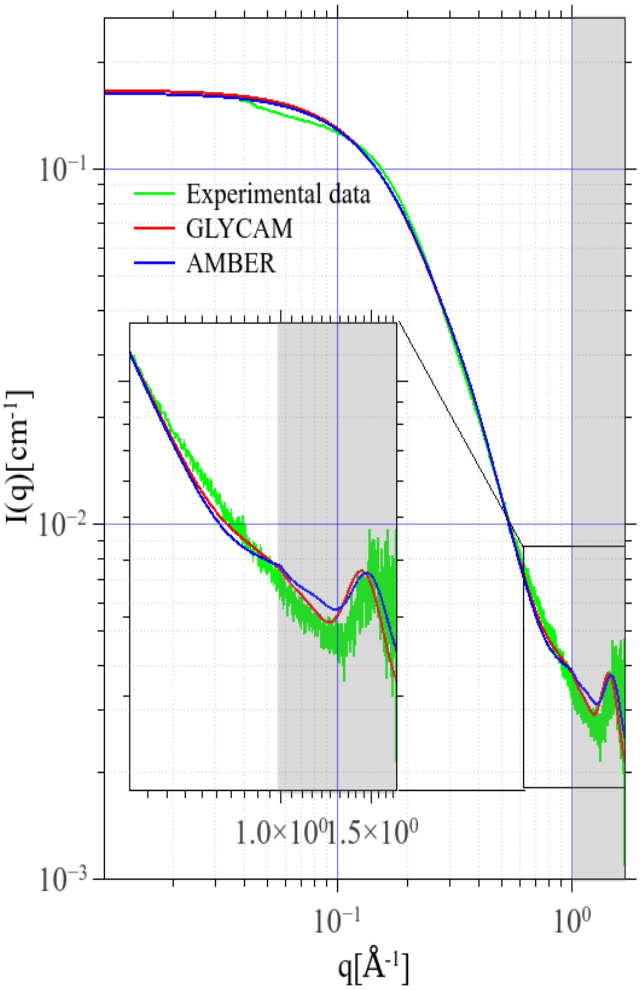


Figure 4.12: SAXS profile of fully charged DP 6, comparison between experimental, GLYCAM and AMBER force filed.

scales (that typically have not been accessed with experiments) the SAXS profile is highly dependent on the exact position of atoms within the sugar rings, which in turn is dependent on the force field. With improvements in Synchrotron technology this offers an exciting possibility that such experiments might be used in the future to help optimise force-fields.

Figure 4.13 compares profiles generated from single snapshots of the atomic positions (pdb files) from an MD trajectory and an average over many coordinate sets (representing fluctuations in the equilibrated state). It is striking that due the fact that the sugar residues in the HG chains are joined by glycosidic bonds, and have intrinsically rather low flexibility, that there is not a significance difference between SAXS profile taken from each snapshot and the average (figure 4.13) especially in the low q region. Therefore, in the case of HG chains in the area of q$\lesssim$ 1Å^{-1} the result averaged over SAXS profiles from many configurations is essentially the same as the SAXS profile of any individual configuration of that chain (including the starting configuration (4.15)). The real space conformation of chains in figure 4.14 shows that while the chains clearly change curvature, differences are relatively minor. In contrast with proteins for examples, where large-scale folding process might take place which change the shape of molecule significantly and lead to different SAXS profiles, the glycosidic bond in polysaccharides keeps their overall shape similar and it gives a similar SAXS profile for each individual snapshot of MD in the low q region.

Figure 4.15 shows the results of the simulated SAXS profile for DP 25. The effect of the limited flexibility of the chains(figure 4.14), particularly at lower qs as described above, is investigated further by comparing the result to that which would be obtained from a rigid rod. The SAXS profile of a rod with a length of 125 Åand radius of 2.5 Å(See Section on Rg fro Estimation of these numbers) was calculated using SasView software (http://www.sasview.org) and is shown in figure 4.15 along with a SAXS profile calculated for a fully charged single chain of DP 25. It can be seen that below q$\sim$ 0.2Å^{-1}, corresponding to distance over $d = 3.14$nm in real space, the chain indeed looks like a rod and the SAXS profiles match well.

By way of summarising the isolated chain work, figure 4.16, shows the modelled SAXS profile for DP 6, DP 10 and DP 25 in comparison with the experimental data.

Finally, it is important to return to the high q data, where some variance in the predicted scattering is seen depending on force-field, and between simulation and experiment. In order to demonstrate the spirit of how the data might be used for force-field optimisation and parametrisation the dependence of the simulation on the torsion potential was explored. The torsion potentials are one of the key parameters effecting the

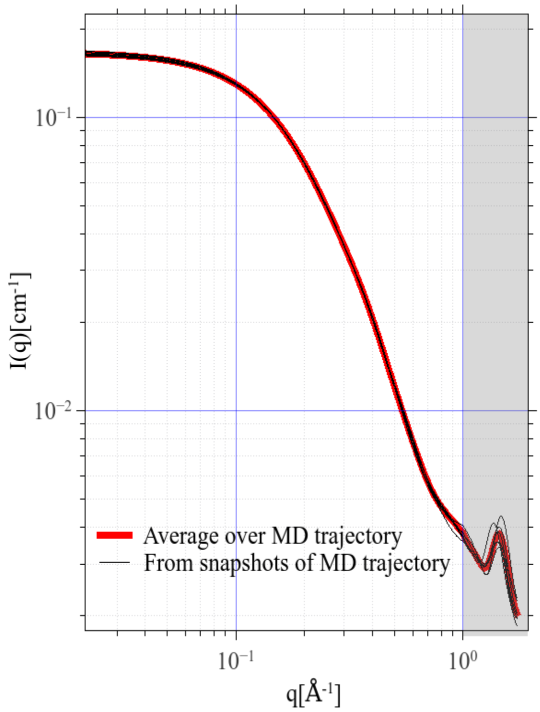


Figure 4.13: Average SAXS profile of single chain DP6 along with the individual SAXS profiles calculated from snapshots of MD trajectory.

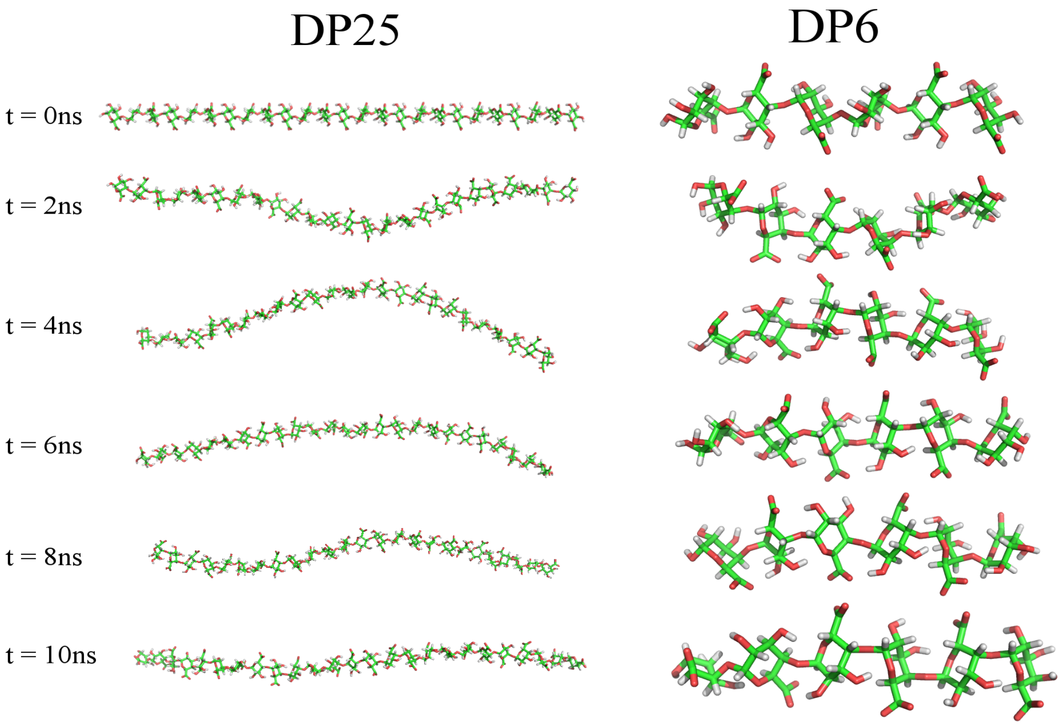


Figure 4.14: Real space view of DP25 and DP6 along the MD simulation.

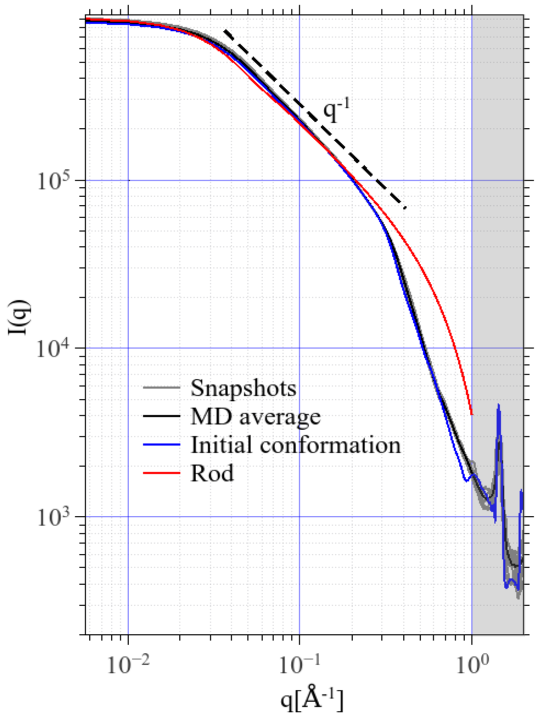


Figure 4.15: SAXS profile of a single chain DP 25 and a rod (size as same as the chain using SasView software).

conformation of the HG chains. Figure 4.17 shows how the SAXS profile changes with a doubling of the height of the torsion potentials. As expected such a change effects the SAXS profile in high q region, supporting this contention that in the future such data might be useful for force field development. However, it can also be seen in figure 4.17 that the other part of SAXS profile ($q < 1\text{Å}^{-1}$) are not significantly altered indicating a robustness to such changes at a larger length-scale.

A key parameter that can be useful to understand the stiffness of a molecule is calculating the radius of gyration for the molecule during the simulation time. In the next section the radius of gyration is calculated for HG species to have a better understanding of stiffness of a HG chains.

4.4 Radius of Gyration

Radius of gyration, R_g,is defined as a distance at which a spherical shell having the entire mass of an object would be placed away from the centre of mass so that moment

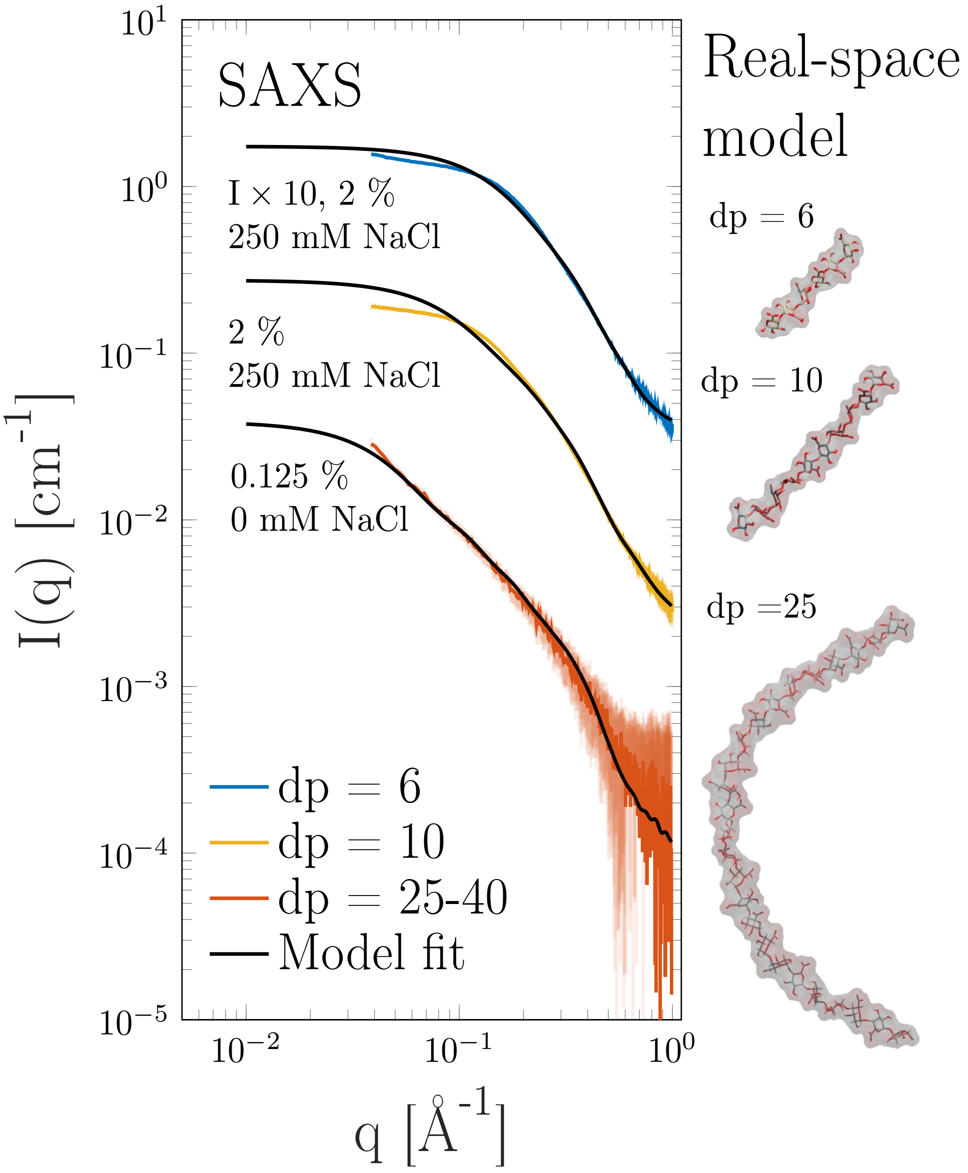


Figure 4.16: Comparison between SAXS profile calculated from MD and experimental data for HGs with DP6, DP10 and DP25. pH=7 for the experimental samples.

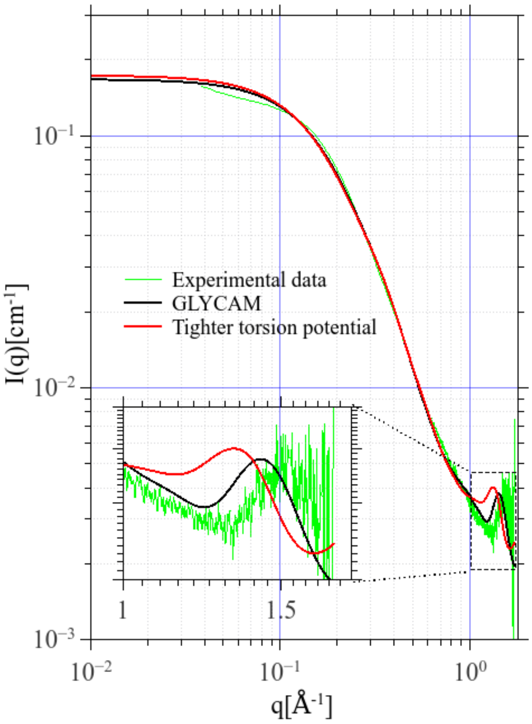


Figure 4.17: Comparison of SAXS profile calculated using GLYCAM force field with normal torsion potential and double height for torsion potential barrier.

of inertia of such a shell and the object would be the same, and can be written as;

$$R_g = \sqrt{\frac{I}{M}} \tag{4.6}$$

where M is the total mass of object and I is the moment of inertia around a certain axis.

In polymer studies, this parameter tells us something about the size of a polymer configuration. For a particular flexible polymer, a high value of R_g means the polymer has an extended conformation, compared to a compact configuration that would exhibit a small value for the radius of gyration. Mathematically, the radius of gyration for a polymer can be calculated as;

$$R_g = \sqrt{\frac{1}{N}\sum_{i}^{N}\vec{S}_i^2} \tag{4.7}$$

where S_i represents the distance of each monomer to the centre of mass of the polymer and N is the total number of monomers along the polymer chain. The distance between

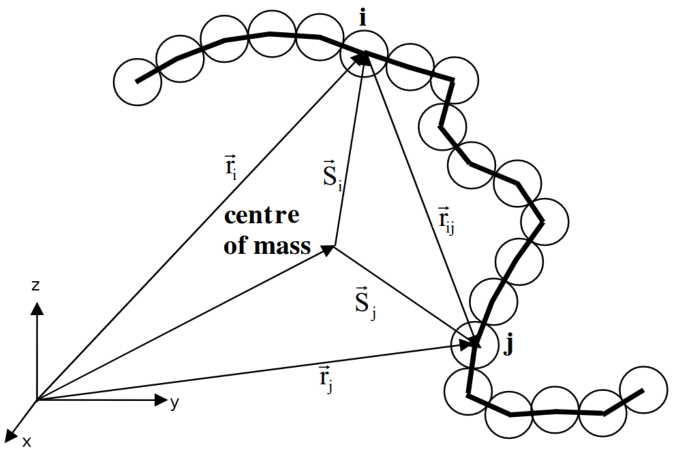


Figure 4.18: Schematic representation of a monomers in a polymer at site i and j with a distance of r_{ij}. Adapted from [6].

two monomers can be calculated as $\vec{r_{ij}} = \vec{S}ij = \vec{S_j} - \vec{S_i}$ and it is known that;

$$\sum_{i,j}^{N} \vec{S_{ij}^2} = N\sum_{i}^{N} \vec{S_i^2} + N\sum_{j}^{N} \vec{S_j^2} - 2\sum_{i,j}^{N} \vec{S_i} \cdot \vec{S_j} \tag{4.8}$$

where the last term is null [6], as S_i and S_j are independent vectors from the centre of mass, so it can be rewrite as 4.8 as

$$\sum_{i}^{N} S_i^2 = \frac{1}{2N}\sum_{i,j}^{N} S_{i.j}^2 = \frac{1}{2N}\sum_{i,j}^{N} r_{i.j}^2$$

and combining with equation 4.7 gives

$$R_g = \sqrt{\frac{1}{2N^2}\sum_{i,j}^{N} r_{i,j}^2} \tag{4.9}$$

This gives a general formula to calculate the radius of gyration for a polymer.

Using equation 4.6 and the moment of inertia for a cylinder around the axis through the centre perpendicular to the cylinder axis, $I = \frac{r^2}{4} + \frac{l^2}{12}$, the radius of gyration(R$_g$) for a rod with a length of l and radius of r can be calculated by;

$$R_g = \sqrt{\frac{r^2}{4} + \frac{l^2}{12}}, \tag{4.10}$$

using this equation for a rod with a radius of 0.25nm and length of 4.8nm, the radius

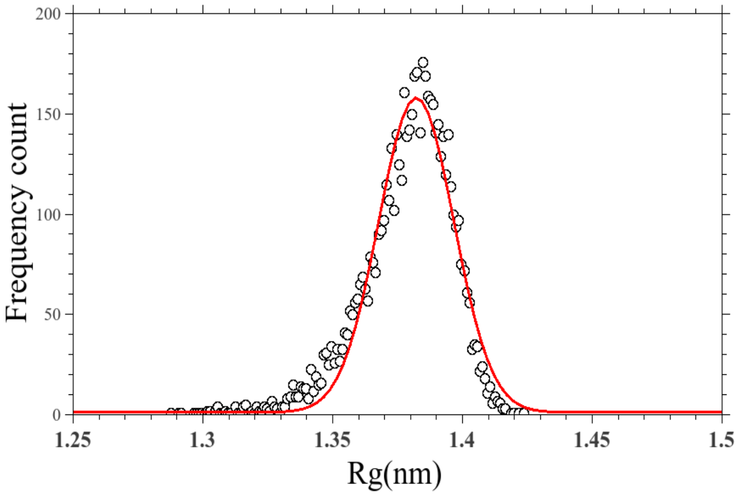


Figure 4.19: Distribution of R$_g$ calculated during the MD trajectory for DP10.

of gyration will be 1.39nm. This calculated R$_g$ for a rod is in the range of R$_g$s for a HG with DP 10 given in figure 4.19. Figure 4.19 shows the frequency of calculated R$_g$ for a chain with 10 residues during the MD trajectory. The fitted red line is showing the Gaussian distribution of the value around the mean.

It is known that the radius of gyration for a random-coil polymer scales power-law with the length of molecule[120]. R$_g \propto N^\nu$, where N represent the length of polymer. For an ideal polymer this power factor, ν, is 0.5[120]. However when the stiffness of the polymer increases then this exponent will increase as well, to 1 for a completely rigid linear polymer.

It worth to say that in the literatures, the power-low behaviour is observed for the long enough polymers (in the range of 20-500)[120] which is really time consuming for MD to cover this range, specially for long chains. Here the radius of gyration for a list of HGs in a range of DP2-DP21 is reported in figure 4.20. A linear relation can be seen in this range between R$_g$ and DP demonstrating a rigid behaviour of HGs in this range. However, fluctuations of R$_g$, σ_{R_g}, which is related to the stiffness of the molecule do increase with the length. The inset in figure 4.20 shows that the standard deviation of R$_g$ for longer chains increases as they have more freedom of flexibility.

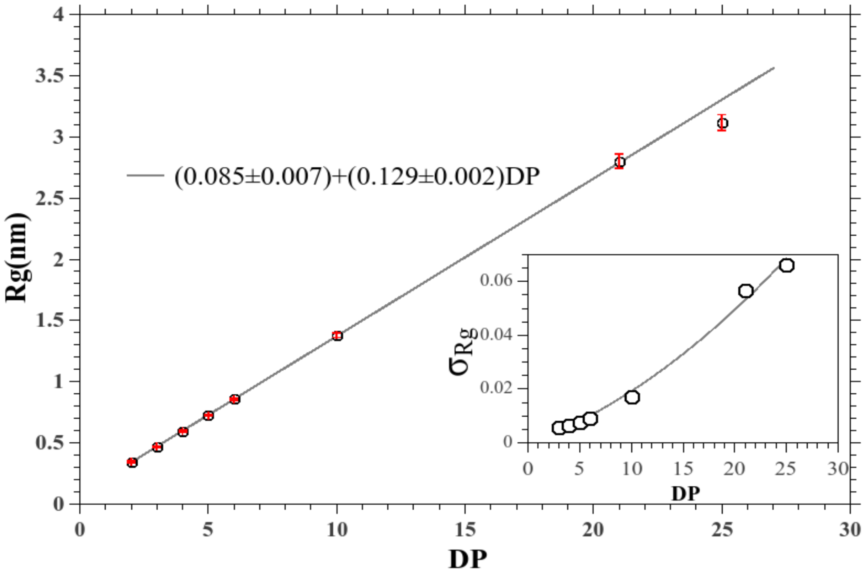


Figure 4.20: Radius of Gyration as a function of chain length. Error bars are the standard deviation calculated during the entire MD simulation.

4.4.1 Interacting Chain Regime: Multi-Chain Simulations

SAXS studies can be used to investigate how molecules are arranged in space relative to one another. It particular it has been known for decades that polyelectrolyte solutions are far from homogeneous and that the long range electrostatic interactions lead to a favoured length-scale between chains, reflected in a structure factor known as the "polyelectrolyte peak". For isotropic particles analytical expressions exist that predict the form of structure factors, but for longer chains there are no such expressions. Furthermore, in the case of polymers with heterogeneous charge distributions (such as blocky HG samples) it is difficult to image how a useful analytical expression would be derived. Here multi-chain MD simulations are carried out in order to investigate whether they can capture the interactions of the the HG chains at different concentrations and in different ionic environments. To reach this goal, MD simulations were performed in such a way that several chains are introduced into a single simulation box filled by water molecules. Chain concentrations (controlled for comparison with experiment) were calculated as a percentage ratio of the molecular mass of HG chains to the total molecules mass in the simulation box. E.g. 10 fully charged DP 25 HG chains in a simulation box with 120582 water molecules is equivalent to a 2% solution (each chain has a mass of 4396 amu and each water molecules has a mass of 18 amu).

The MD Protocol is essentially the same as described in the previous chapters, with charged HG neutralized by Na^+, and runs carried out for 20ns to 40ns depending

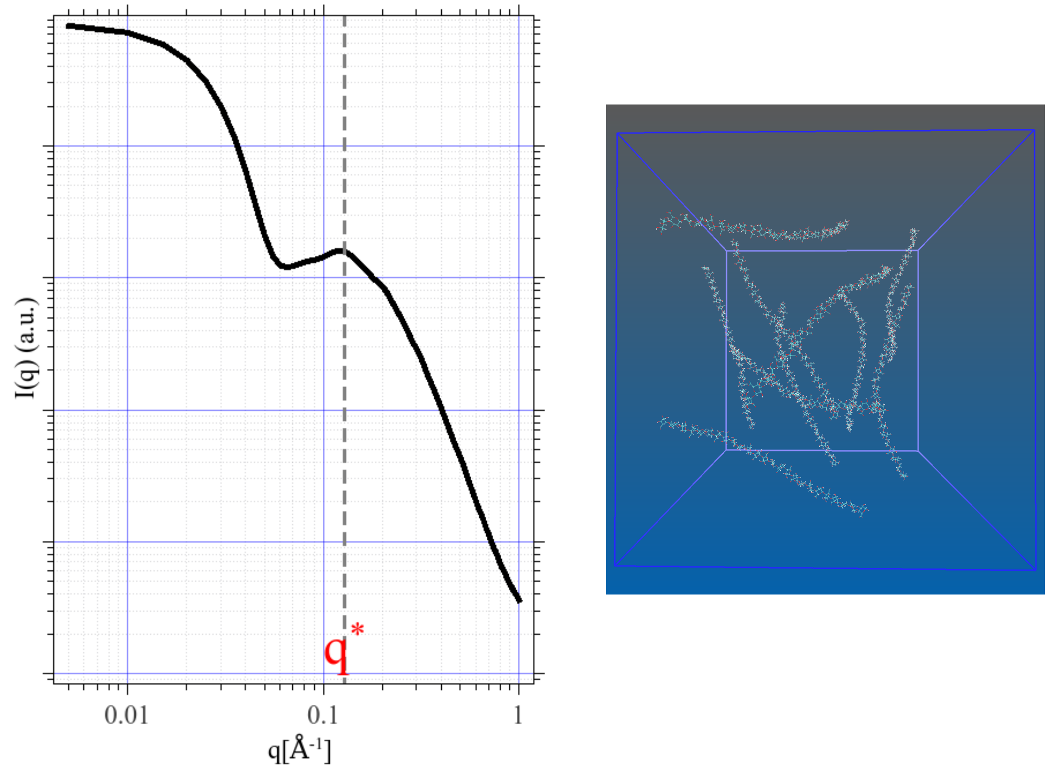


Figure 4.21: SAXS profile for DP25 with concentration of 2% with indication of structure factor peak (q*) and a snapshot of MD trajectory.

on the size of system. The first 5 ns are ignored in the analysis (to be confident of equilibration).

Figure 4.21 shows the calculated SAXS profile for a C=2.0% solution of fully charged DP 25 HG chains. The emergence of a structure factor peak at a specific scattering vector shown by q*, corresponding to the average distance between chains in real space can clearly be seen as electrostatic repulsions drive the charged chains to be as far from each other as possible given the concentration of chains in the box.

Figure 4.22 shows the SAXS profile for two higher concentration systems. It should be recognised that the form of these curves contains features that are artefacts from the density step of the simulation box itself (the SAXS profile calculated for a cube with same size of simulation box is also shown). It can be seen that the first peak in two SAXS profiles and also the second one at C=8.2% occur in the same place as those calculated for the box scattering pattern. In figure 4.22 it is also shown (in real space) that a simulation box with C=8.2%(top) is almost filled with chains (so that zooming out the simulation space has the appearance of a cube), in contrast with the low concentration C=0.58%(bottom) - explaining why this is only an issue with

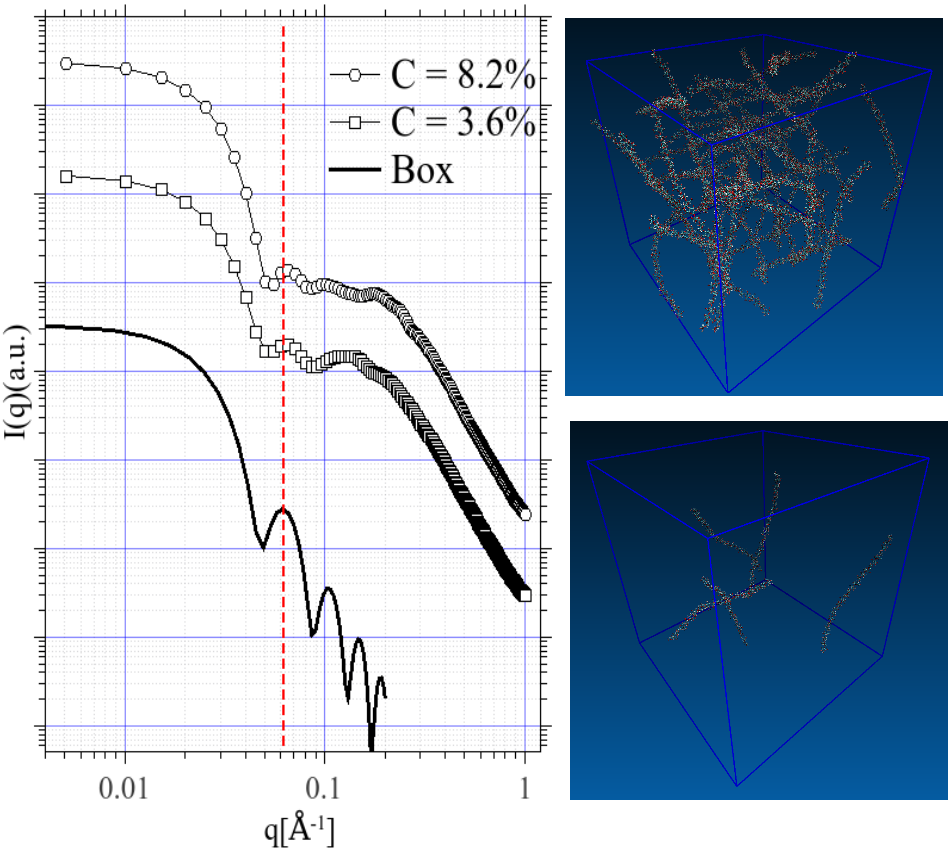


Figure 4.22: Box effect on SAXS profile calculation at high chain concentrations. SAXS profile for DP25 at high concentrations. Solid line is the scattering pattern of a cube in size of MD simulation box. Right, high and low concentration chain configuration in the box.

simulating more concentrated systems.

At low concentration, charged chains have a chance to be further away each other compared to at higher concentrations. So, the structure factor peak should depend on the concentration of chains in the box (indeed there are analytical forms for the scaling discussed in due course). This dependence was studied here by running a series of MD simulations with different HG concentrations. It can be seen in figure 4.23 that, as expected, the position of the structure factor peak varies with the concentration of HGs. The red arrows indicate the position, q*s, of the structure factor peak.

Figure 4.24 shows the dependence of q^* on the concentration and a fit to $(0.086 \pm 0.002)C^{(0.34 \pm 0.02)}$ confirming a scaling behaviour with concentration that is predicted in the dilute regime of $C^{1/3}$ [121, 122].

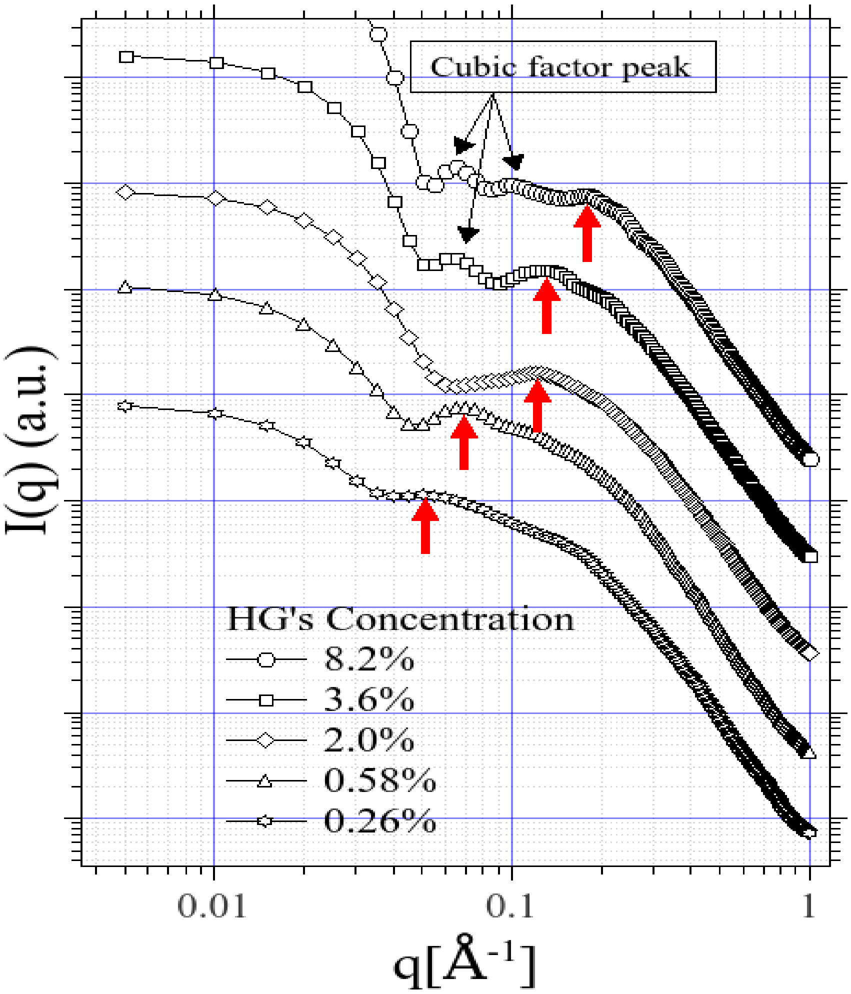


Figure 4.23: SAXS profile of DP25 fully charged chains with different concentration. Red arrows point the position of structure factor peaks.

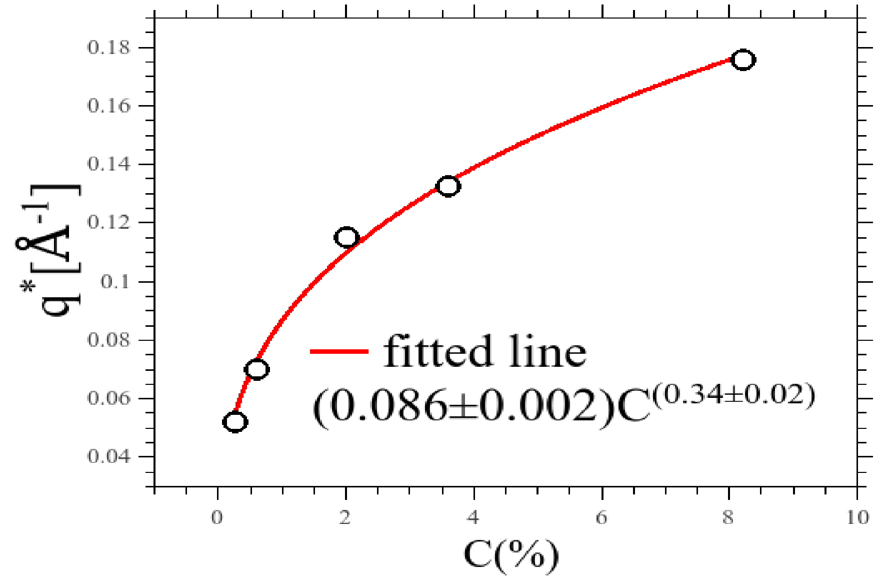


Figure 4.24: Structure factor peak position (q*) as a function of chain concentration for DP25(fully charged). Errors represent standard deviation.

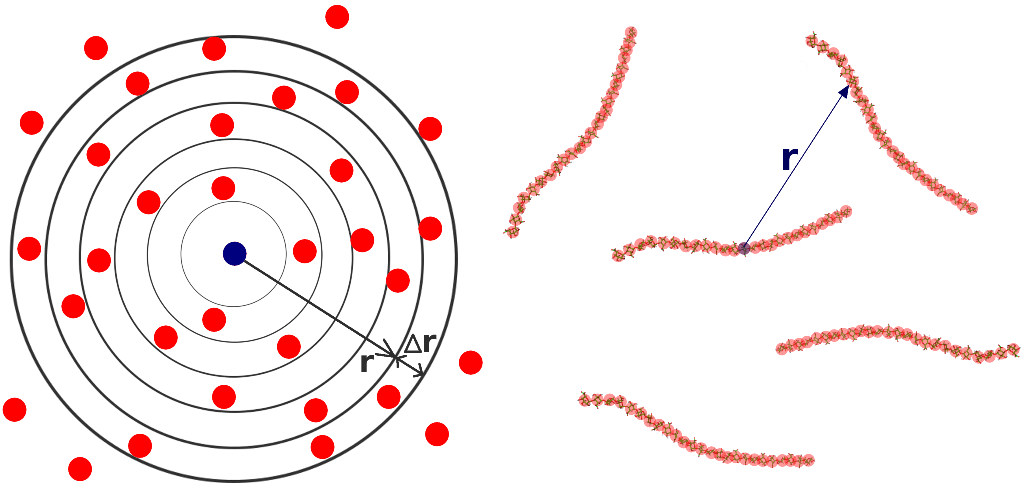
4.5 The Radial Distribution Function of Interacting HG's

The Radial Distribution Function (RDF) describes how the molecules (on average) are located around each other in the system. It is a good way to see how the molecules interact with each other in order to minimise the inter-molecular potential. For a liquid phase system as particles have a more freedom to move, the RDF should be calculated for a reasonable simulation time to be able to compare with the experiment.

To calculate the RDF, first it is needed to chose a particle and then consider spheres with the constant distance between them ($\Delta\mathbf{r}$) centred on the first chosen particle (see figure 4.25a). Then the number of particles in the specific spherical shell at distance $\mathbf{r}$ needs to be counted and divided by the volume of the shell and also the average density of particles in the system. The mathematical formula for calculating the RDF at distance $\mathbf{r}$ is given by;

$$g(\mathbf{r}) = \frac{n(\mathbf{r})}{\rho 4\pi^2 \Delta\mathbf{r}} \tag{4.11}$$

where, n($\mathbf{r}$) in the mean value of the number of particles in the sell at distance $\mathbf{r}$ from the centre, ρ is the average number density of particles in the entire of system. To calculate the RDF (g($\mathbf{r}$)), this process should be continued over all particles to have a better average of particle's arrangement inside the system and for an MD simulation, this process goes over all steps of the (once equilibrated) simulation trajectory.



(a) (b)

Figure 4.25: (a) Radial Distribution of particles around a particular particle at distance r in a layer of Δ(r). (b) A presentation of how RDF for residues in the chains were calculated.

Calculating g(**r**) for system of chains is not as trivial as performing the calculation for atomic systems or spherical molecules. For the chains some residues can be close to the residues on the other chain while other residues may be far from it. To solve this issue g(r) was calculated taking each sugar residue as the centre of the sphere, and ignoring intra-molecular residues. Using this method the radial distribution function for HG chains with DPs of 6, 10 and 25 with different concentration and ionic environments have been investigated.

Experimentally, the radial distribution function, g(r), was calculated from the inverse Fourier transform of the structure factor, $S(q)$. Basically, the measured scattered intensity, $I(q)$, is related to S(q) and to the form factor, $P(q)$ through: $I(q) \propto S(q)P(q)$. $P(q)$ was taken from scattered intensity in dilute regime. As the HG chains are not highly flexible to change their form factor considerably, it was assumed that $P(q)$ will remain the same for all studied chain concentrations. Using this assumption, $S(q)$ was extracted from $I(q)$ and then g(r) was calculated from the inverse Fourier transform of $S(q)$ through [123]:

$$g(r) = 1 + \frac{1}{2\pi^2}\int_0^\infty dq \frac{S(q)-1}{\rho} q^2 \frac{\sin(qr)}{qr} \tag{4.12}$$

where ρ is a parameter related to the density.

4.5.1 Hexamer

Figure 4.26 shows the radial distribution function calculated as described above for solutions of DP 6 HG with different chain concentration simulated by MD. The peak of g(r) graph shows the most highly populated distance between chains in real space. As it is expected, when the chain concentration decreases, chains have more space to push each other away mediated by the electrostatic repulsion between them, so the most highly populated distance between chains (the peak in the g(r) graph) increases.

The arrows in figure 4.26 indicates the position of the peak in g(r) graph for each chain concentration. The positions of the peak maxima were identified by numerical differentiation of the smoothed data and the inset in the figure shows how the peak position, which we call it ζ, the correlation length, scales with the concentration. The fitted curve shows that ζ scales with the concentration in a power-law form with the power of -(0.30±0.02). This behaviour is expected for a three dimensional dilute system where by changing of concentration, charged particles try to fill the box as much as possible and distance between particles scales with the power of -dimension^{-1}. As it is well known that in dilute regime, below the critical (or overlap) concentration, the polyelectrolyte peak scales with the concentration as $C^{1/3}$ [121, 122] and as there is an inverse relation between scattering vector(q) and real space distance(r) it is expected to have a scaling of $C^{-1/3}$ for ζ respect to the concentration of chains.

The effects of counter-ions concentration were investigated by running the simulations with different salt amounts of added ions (Na^+ and Cl^-) (Simulations in this system to date have used the minimum amount of ions required to neutralise the charges on the chains). Figure 4.27 shows g(r) calculated for DP 6 HG oligomers from the MD simulations with a chain concentration of 0.46% and different ionic strength solutions. It can be seen in figure 4.27 when 130mM of salt was added in the system after neutralisation, the probability of finding a chain in closer distance is higher,although the most populated distance between chains appears to remain the same as that found the neutralised system.

4.5.2 Decamer

In this section simulating HGs is continued with DP10. First simulations were run with two different HG concentrations (figure 4.28a), and as expected in the high concentration systems chains are on average closer in comparison with the low concentration.

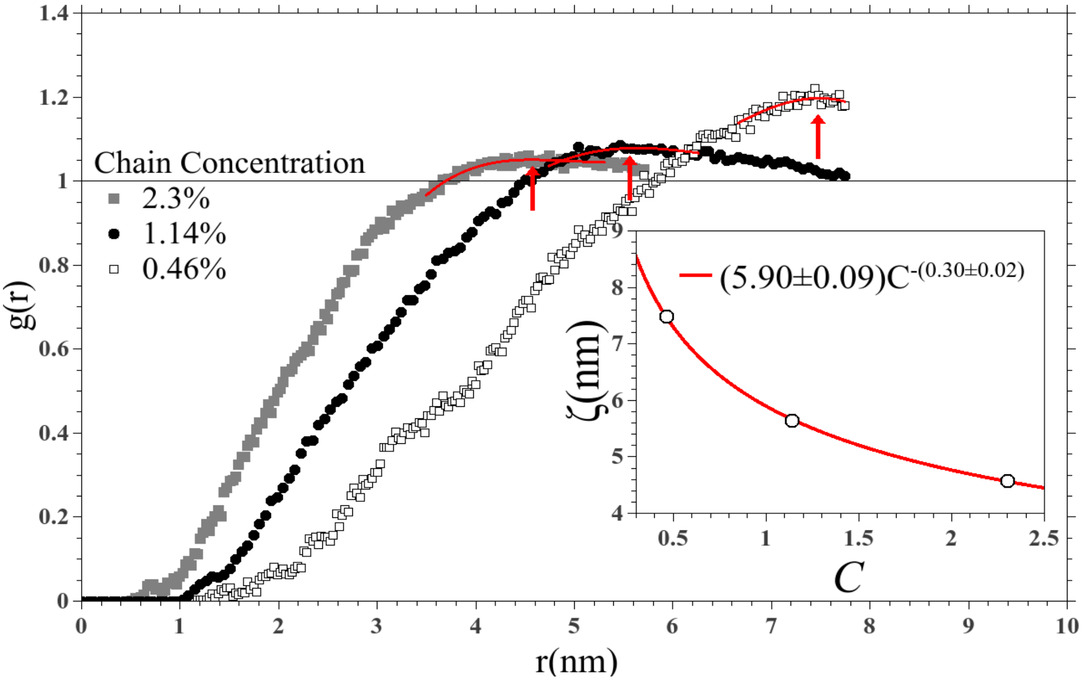


Figure 4.26: Radial Distribution Function for DP6 (fully charged) with different concentration. Red arrows indicate ζ. Inset: ζ as a function of chain concentrations. Errors represent standard deviation.

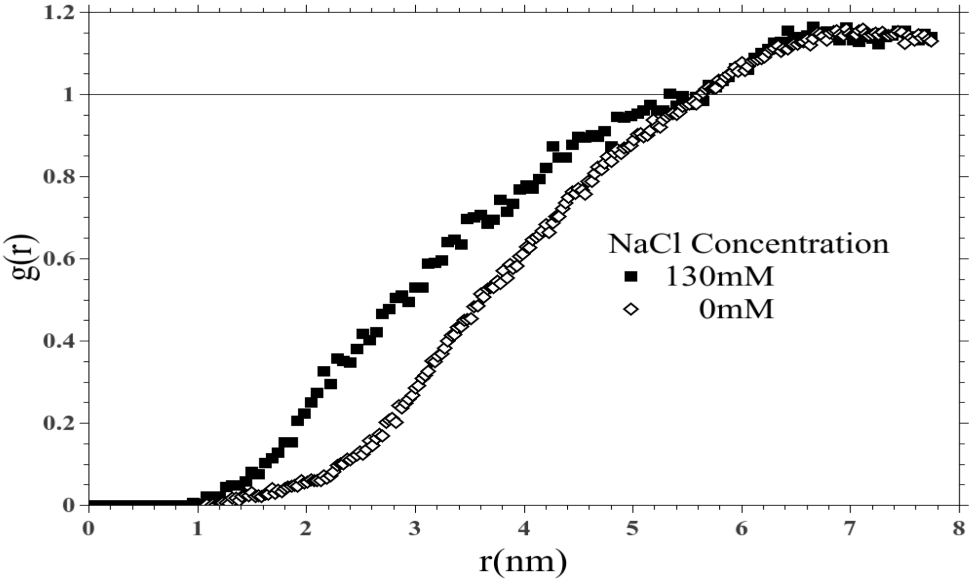
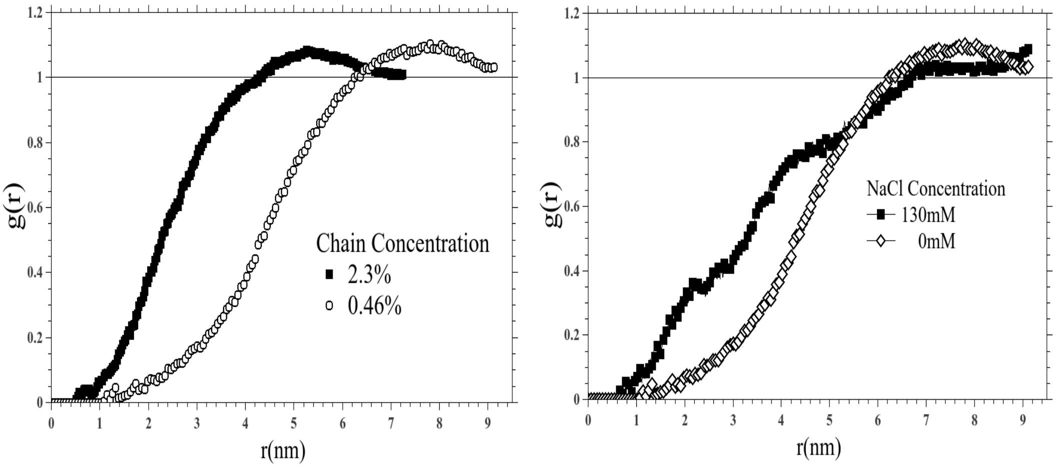


Figure 4.27: Radial Distribution Function for DP6 (fully charged, C=0.46%) with different amount of salt in the solution.



(a) (b)

Figure 4.28: (a) Radial Distribution Function for DP10 (fully charged) with two concentrations. (b) RDF for DP10 (C=0.46%) with different amount of salt.

Figure 4.28b shows the g(r) graphs of DP 10 with the concentration of 0.46% with different amounts of added ions and again as expected, like DP 6 in the previous section, chains in higher salt concentration are able to come closer together.

Figure 4.29 shows a comparison between experimental data from this research group and the results taken from MD simulation for the radial distribution function of DP 10 and DP 6 at 2.3% of chain concentration. It can be seen that the peak positions at g(r) graphs for DP 10 are shifted to higher distance in comparison with the DP 6 owing to the increased charge.The peak position and the difference of the position between the different oligomers is captured well by the MD simulation. At the smallest length scales however there is some discrepancy, likely owing to the failure of some of the assumptions in the MD force-field (for example the breakdown of continuum electrostatics, such as the assumption of the constant dielectric constant, or lack of proper consideration of hydration waters).

4.5.3 DP 25

Finally DP 25, the longest chain that has been investigated in multi-chain simulations on in a reasonable computational time is investigated. Figure 4.30 shows how the g(r) graphs change with increasing of chain concentration in a similar fashion to found for DP 6 an DP 10. Again the positions of the maxima in the g(r) curves (shown with arrows) were extracted. In the case of the very low concentration simulation (0.26%) an extrapolation was required to estimate the value of ζ given box size restraints. The

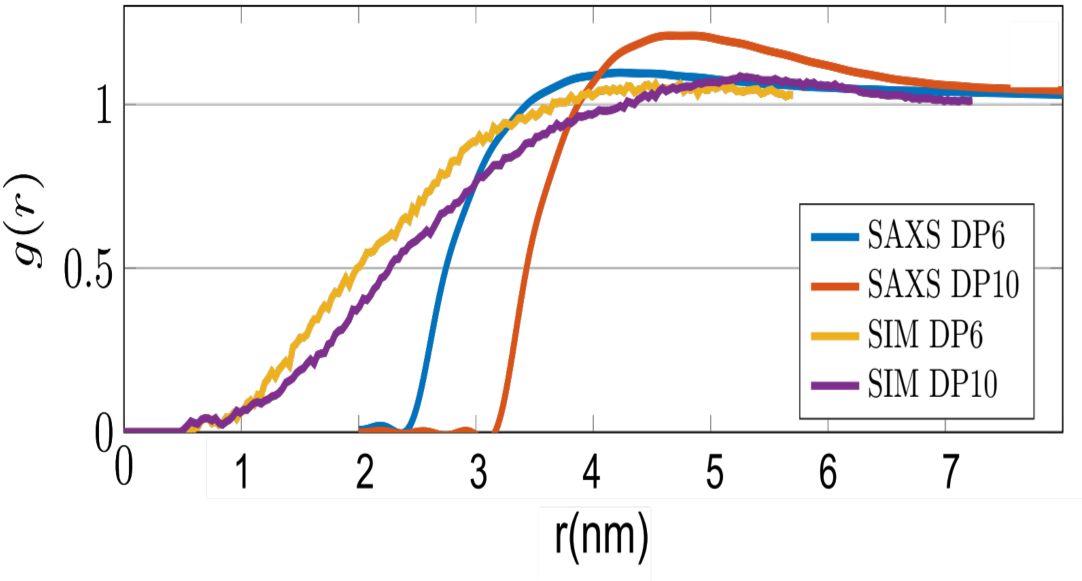


Figure 4.29: Experimental and theoretical calculation of rdf for Decamer and Hexamer (fully charged) in 2.3% of chain concentration. pH=7 for the experimental samples.

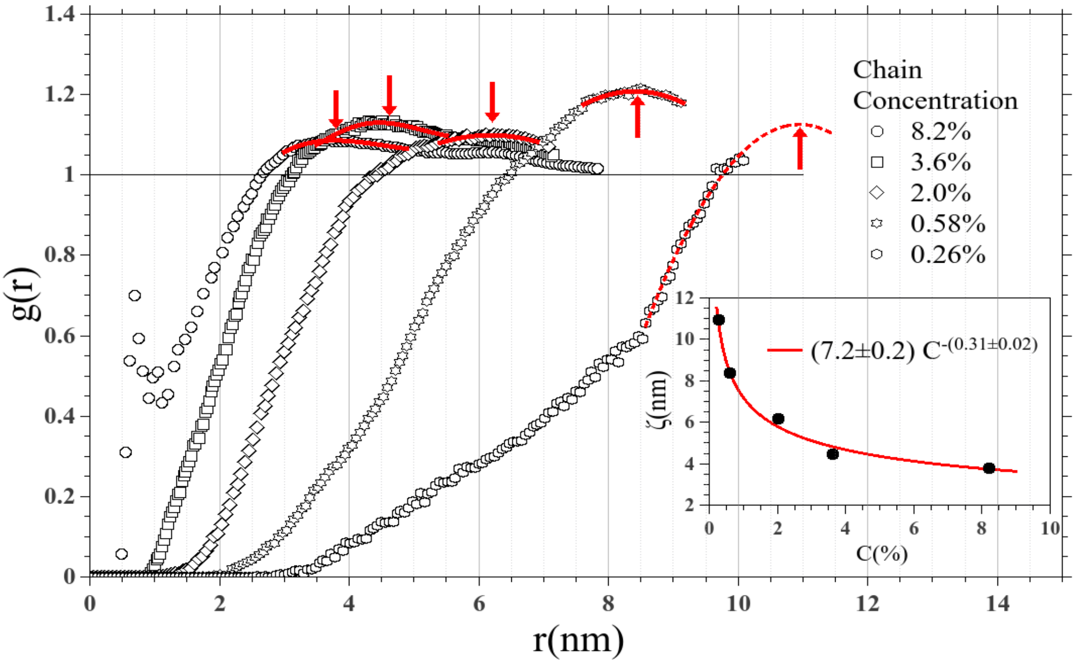


Figure 4.30: Radial Distribution Function for DP25 (fully charged) with different concentration. Red arrows indicate ζ. Inset: Power-law behaviour of ζ as a function of chain concentrations. Errors indicate the standard deviation.

inset in figure 4.30 shows how ζ varies with the concentration, and once again the fitted line agrees well with the power law scaling of around -1/3 (0.31±0.02).

Along with the MD simulation results there were some other available experimental data in this research group. The results are shown in figure 4.31 where the calculated correlation lengths at different concentrations for a variety of different degrees of polymerization (DP) are shown.

For the SAXS measurements the DP varies from 6 to around 500 (for both the blocky (B48) and random (R47) samples. Blocky and random refer to the charge distribution pattern on the chain). For the samples with highest DP, namely B48 and R47, there is a power law scaling of the correlation length, ζ, with concentration with an exponent approaching 0.5, (the insert shows power law fit exponents versus DP). This suggests that these systems are no longer dilute. At intermediate DP values a sample named HG40 with $DP \approx 100$ and a blocky charge pattern with a ratio of 60 % of the chain having charged monomers has been investigated. This sample reveals a power-law between the two predicted values of 1/3 and 1/2. Interestingly, this is similar to the smaller sample labeled as DP25, which has DP distribution of 25 to 50, and 100% unmethylesterified homogalacturonic acid. For the lowest DP samples, those with a DP of 6, the scaling laws were only obtained at multiple concentrations using simulation, as limited amounts of material were available for experimental work. However, the SAXS results for the DP6 and DP10 samples at 23 g/L(2.3%) are included for comparison. The simulation data indeed captures the scaling laws well and agrees well with the SAXS data for the oligomer samples.

As the g(r) graph and SAXS profile both address the arrangement of molecules in the solution therefore a connection between parameters taken from these two methods can be found. Based on the definition of ζ, this real space distance is related to the q* seen in reciprocal space and more directly measured in scattering experiments (discussed in the beginning of this chapter) by:

$$q^* = \frac{2\pi}{\zeta}. \tag{4.13}$$

Using this equation it is possible to calculate the predicted position of the structure factor peak that would be observed in SAXS and compare these with calculations in reciprocal space and ultimately with experiments. Table 4.1 shows the comparison between q* calculated from equation 4.13 and the q* taken from figure 4.24 where the structure factor was directly calculated. There is good agreement between the values obtained from different methods and this gives a confidence not only in the soundness

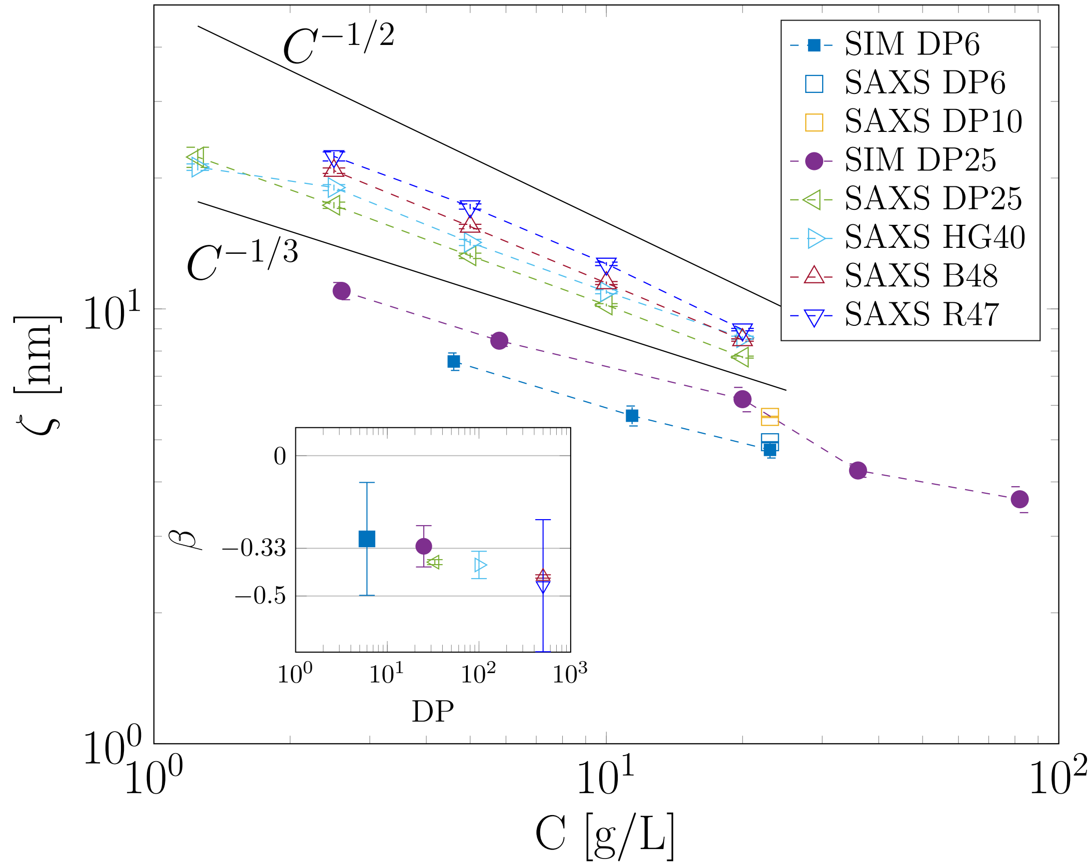


Figure 4.31: Correlation lengths, ζ, extracted from both SAXS and MD simulations (fully charged) as a function of concentration. The insert shows extracted exponents from power law fits as a function of degree of polymerization (DP). Error bars represent 95 % confidence intervals. pH=7 for the experimental samples.

Concentration	q* calculated from $\frac{2\pi}{\zeta}[\text{Å}^{-1}]$	q* from SAXS profiles$[\text{Å}^{-1}]$
0.26%	0.057	0.053
0.58%	0.074	0.070
2.0%	0.101	0.115
3.6%	0.141	0.132
8.2%	0.172	0.175

Table 4.1: Comparison between structure factor peak position (q*) calculated using ζ taken from g(r) graphs and the value taken from the SAXS profile graphs.

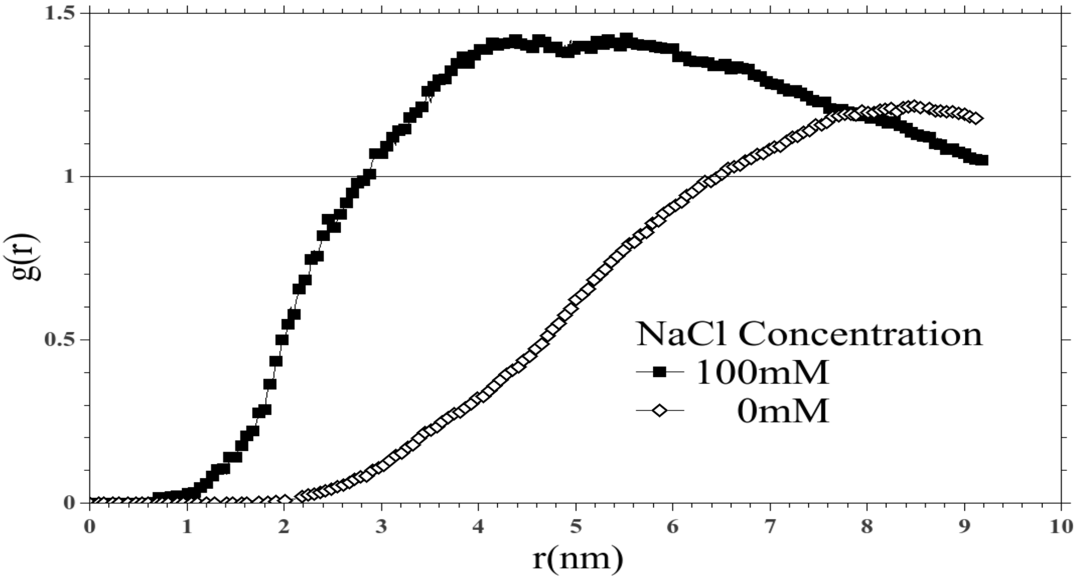


Figure 4.32: Radial Distribution Function for DP25(fully charged) (C=0.46%) with different amount of salt in the solution.

of calculations but also in assigning peaks that are from systems of interest and not the concentration"box".

Once again the effect of the ionic environments was studied, this time for DP25, and figure 4.32 shows the change of g(r) for these HGs at a chain concentration of 0.58%. As expected, adding more salt leads to a higher probability of the chains being found closer together.

4.6 Uncharged HGs

At present there is no experimental SAXS data for completely methylesterifed or protonated systems, (these would be interesting as these conditions potentially remove polyelectrolytic effects). However, it is known that protonated chains interact with each other and so it was deemed interesting to perform some theoretical investigations here.

The radial distribution function was calculated for totally uncharged systems, either fully protonated or fully methylesterified chains. Both system are completely uncharged and in the protonated system there is a high chance for hydrogen bonds to form (the favoured 3^1 conformation of the uncharged chains producing a propensity to form intermolecular H bonds).

To carry out the MD simulations for the fully protonated chain it was first necessary to perform DFT and RESP fitting in order to obtain the relevant partial charges, as described previously for the charged residues. Results for partial charges obtained from a protonated trimer are given in table 4.2. It can be seen that as expected these are slightly different from those applicable to the fully charged trimer (given in chapter 2), particularly for the carbonyl groups.

Figure 4.33b shows a snapshot of the final steps in an MD trajectory for fully protonated chains at C=0.58% tangling due to the formation of hydrogen bonds. In fact, in the absence of electrostatic repulsion the probability of establishing hydrogen bond junctions between chains is hugely increased and given enough time the chains completely stick together during the simulation. The number of intra- and inter-molecular hydrogen bonds in the system during the simulation time can be seen in figure 4.34a. It is clear that at the beginning of simulation where there are no connections between chains the number of hydrogen bonds here indicates the number of intra-molecular hydrogen bonds (a). After about 10 ns of simulation, chains start to touch each other and the number of hydrogen bonds starts to increase (b). With time the number of hydrogen bonds reaches double its starting value (c) with this increase related to appearance of inter-molecular hydrogen bonds (See real-space representations).

The multiple states in the radial distribution function represented in figure 4.33a that show a repeating distance are consistent with the size of individual sugar residues (about 0.5nm) indicating multi-chain associations. It should also be noted that such associations are not prominent in completely methylesterifed species despite the removal of the stabilising charge. It could be because of geometric configuration of completely methylesterifed species that prevents chains to be as close as protonated case and this leads to have less hydrogen bound between species.

Atoms	Non-Reducing	Central	Reducing
C_1	−0.07420	0.03340	0.21920
C_2	0.20010	0.14000	0.08250
C_3	0.16150	0.31260	0.32530
C_4	0.18940	−0.09190	−0.20670
C_5	−0.03200	−0.03660	0.06340
C_6	0.75100	0.71520	0.67700
O_1			−0.62290
O_2	−0.61740	−0.61070	−0.61040
O_3	−0.63510	−0.57590	−0.59050
O_4	−0.62780	−0.32970	−0.18280
O_5	−0.29620	−0.27860	−0.37030
O_6	−0.58490	−0.56720	−0.55620
O_7	−0.57690	−0.55700	−0.54220
H_{C_1}	0.16010	0.16070	0.10610
H_{C_2}	0.09540	0.13160	0.13660
H_{C_3}	0.05150	0.01490	0.04490
H_{C_4}	0.05980	0.16240	0.17960
H_{C_5}	−0.08040	0.13810	0.11850
H_{O_1}			0.47170
H_{O_2}	0.40310	0.40010	0.41570
H_{O_3}	0.42560	0.39340	0.39640
H_{O_4}	0.42300		
H_{O_7}	0.44360	0.44520	0.44510

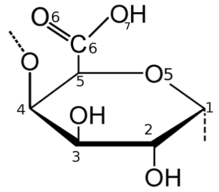
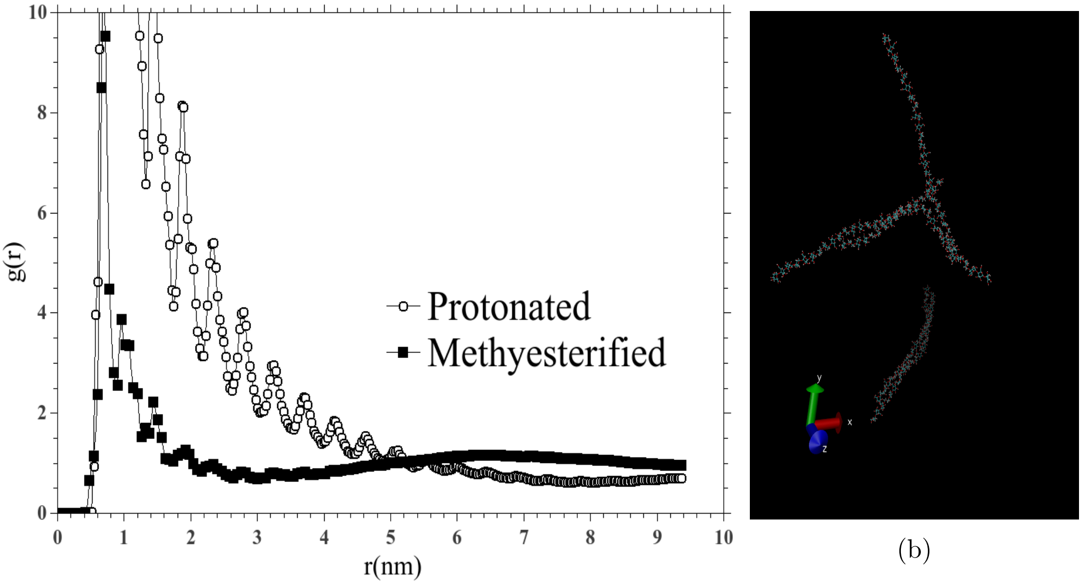


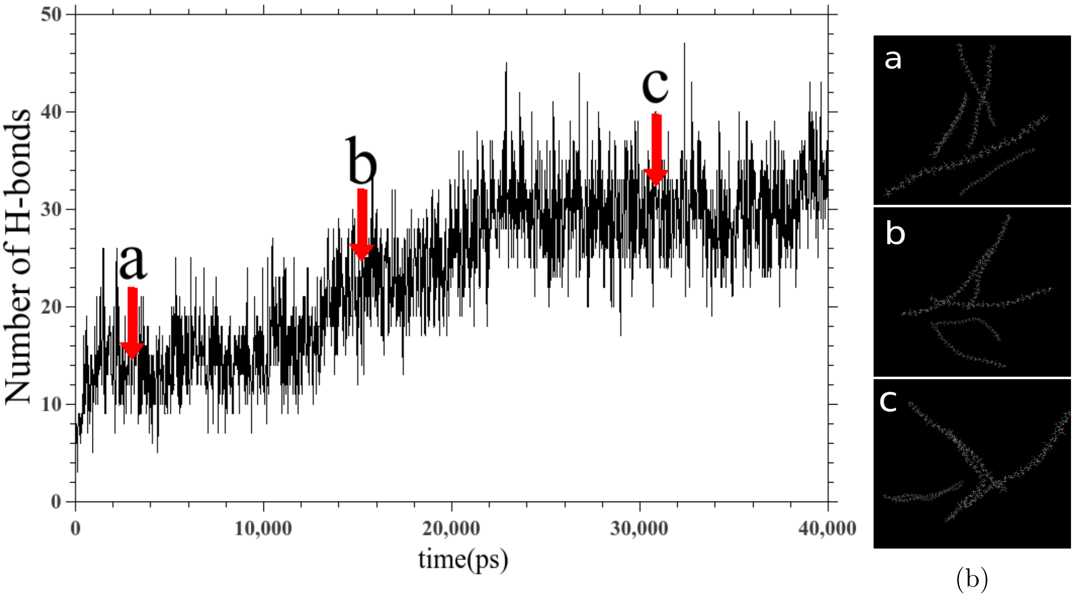
Table 4.2: The partial charge of atoms calculated for residues in a protonated trimer and a schematic presentation of the repeating protonated residue.



(a)

(b)

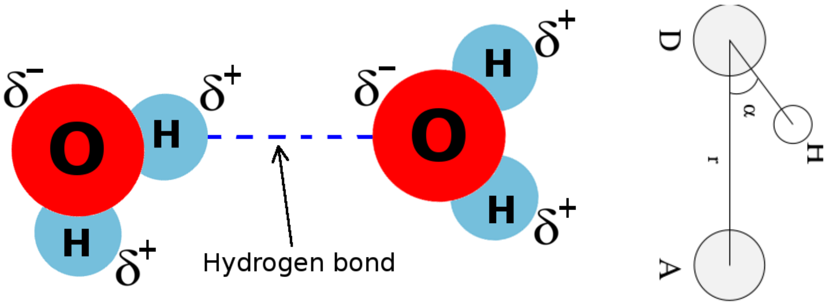
Figure 4.33: (a) Radial Distribution Function for fully protonated and fully methyl-esterified HG solutions of DP 25. (b) A snapshot of MD trajectory showing the protonated HGs stick together via hydrogen bonds.



(a)

(b)

Figure 4.34: (a) Number of Hydrogen bonds in protonated system during the simulation time. (b) Arrangement of chains corresponding to time a, b and c.



(a) Hydrogen bond between oxygen and hydrogen in different water molecules.

(b) Geometrical Hydrogen bond criterion adapted Gromacs manual [124].

Figure 4.35: Schematic representation of Hydrogen bond

4.7 Hydrogen bonds in HG clusters

Hydrogen bonding is the well known attraction between a hydrogen atom that is bonded to a highly electronegative atom like oxygen via a covalent bound, with another electronegative atom around the hydrogen atom. One of the most famous hydrogen bonds is between water molecules (figure 4.35a) where the oxygen atom draws electrons from hydrogen becoming partially negative (δ^-) and leaving hydrogen atoms partially positive (δ^+). The electrostatic attraction between δ^- and δ^+ makes a bond between two water molecules.

In Gromacs, to determine hydrogen bonds between donors (D) and acceptors (A), figure 4.35b, following geometrical criterion are used;

$$\begin{aligned} r &\leq r_{HB} = 0.35nm \\ \alpha &\leq \alpha_{HB} = 30^\circ \end{aligned} \tag{4.14}$$

where r and α represented in figure 4.35b and $r_{HB} = 0.35nm$ is taken from the first minimum of radial distribution function of SPC water.

As there are OH groups on HGs with δ^-, δ^+ charges on them there is a possibility of intramolecular hydrogen bonds within a particular chain or intermolecular hydrogen bonds with the other chains. Red dots in figure 4.36a shows a hydrogen bond between hydrogen in hydroxyl group of a residue with O_5 from the neighbouring residue, this is an example of intra-molecule hydrogen bond in the system, while figure 4.36b, 4.36c and 4.36d show the inter-molecule hydrogen bonds between hydroxyl group and O_5, two hydroxyl groups and hydroxyl-carboxyl group respectively. The hydrogen bond length is between 0.3 and 0.35nm. Hydrogen bond categorised as Strong if the h-bond

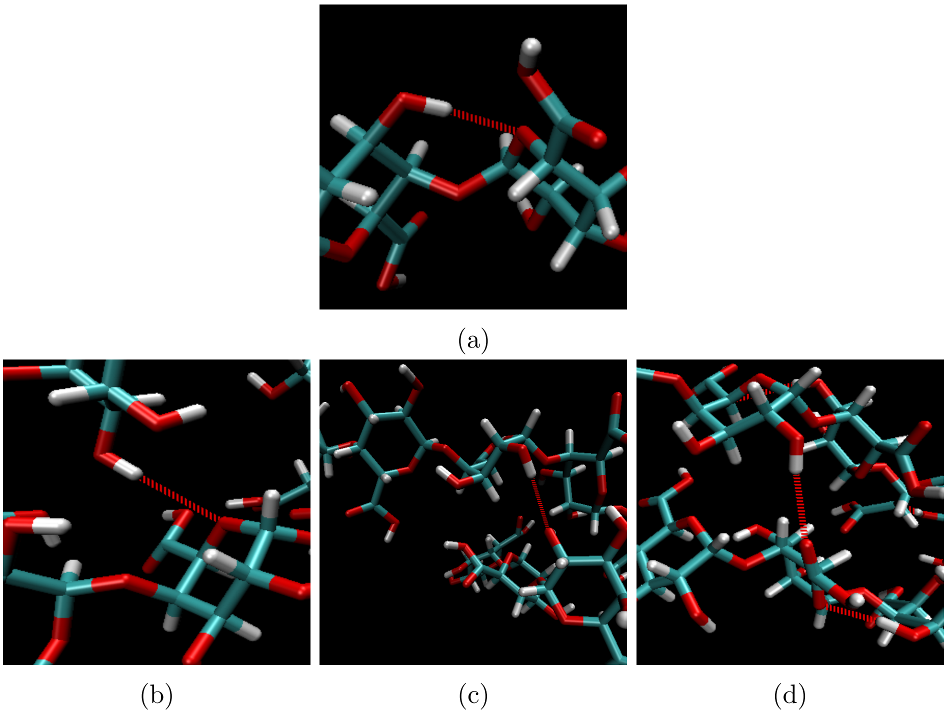


Figure 4.36: (a) Intra-molecule hydrogen bond. (b), (c) and (d) inter-molecule hydrogen bond in the protonated HG chains, DP25.

length is between 0.22nm to 0.25nm, Moderate when h-bond length is in the range of 0.25nm to 0.32nm and as a week bond when it is between 0.32nm and 0.4nm [125]. Therefore, hydrogen bonds here are in the moderate category. Figure 4.37 shows the distribution of hydrogen bond length found in a simulation of 0.58% fully protonated DP 25.

These simulations suggest that qualitatively the MD simulations capture the long-known association of HG chains in acidic conditions, and can form the basis of interesting further SAXS studies on protonated samples.

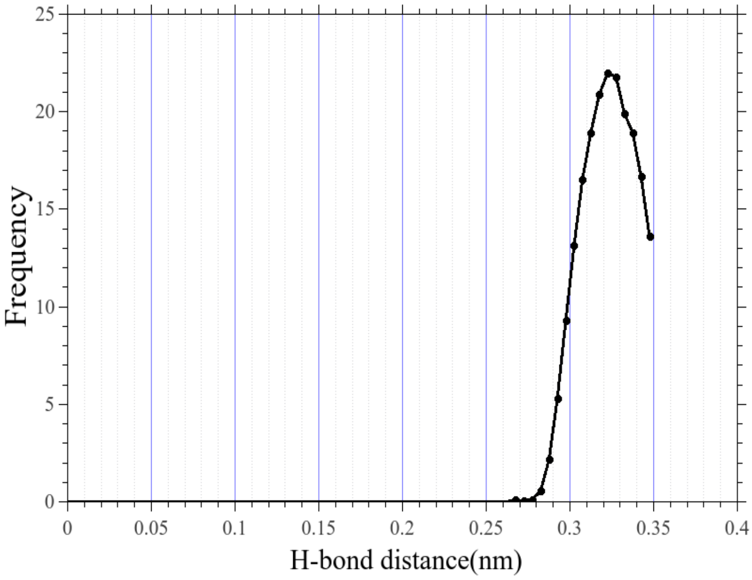


Figure 4.37: Distribution of hydrogen bonds for fully protonated DP25, Chain concentration = 0.58%.

Chapter 5

Conclusion and Future Work

Overall the thesis demonstrates how MD simulations have been able to illuminate some key areas of polysaccharide biophysics.

The work in chapter 2 starts by considering electrical transport phenomena and specifically provides a sound methodology for the calculation of the electrophoretic mobilities of partially-charged oligosaccharides as a function of charge patterning and degree of polymerisation. The methods and data presented in this project allow for the preliminary identification of fragments without the time consuming necessity of purifying standards, and allow experimental conditions that optimize the resolution of particular species to be ascertained and understood. This approach should be useful to scientists interested in structure-function relationships of oligo-polyelectrolytes per-se, particularly the oligogalacturonides, but also to researchers pursuing fragmentation of polysaccharides as a way of ascertaining information on pre-digested substrates. If the fragmentary processess used are well understood then modelling the form of the resulting electropherogram gives a direct mapping between simulation and experiment, and promises to help simulate fine structure ensembles that properly model the substrate. Alternatively, if substrates with well-controlled patterns of methylester groups are utilised then comparison of experiment with predictions can reveal detailed information about the fragmentary processes themselves (such as enzyme action patterns).

In chapter 3 the work is extended to the study of larger polymeric molecules, still focussing on transport. By using carefully designed anionic polysaccharides, with different degrees and patterns of charge, and performing MD simulations of their respective electrophoretic mobilities, the role that counterion condensation plays in determining the electrical transport properties of these polyelectrolytes was investigated in some detail. When the DM of HGs whose charged groups are randomly positioned along the polymer backbone becomes low enough counterion condensation takes place. In addition, HGs with non-random distributions of charged groups can exhibit local counterion condensation *regardless of the average charge spacing* if the patterning of charged groups becomes blockwise enough. In these cases both the number density distribution of mobile diffuse ions and the measured electrophoretic mobility reflect the charge density of the *ion-decorated* polyelectrolyte. These simulations not only give a clearer understanding of the physics in play but also permit that, if independent information is available from other measurements regarding the charge patterning motif of the sample (for example something is known of the demethylesterfication process), then the number of charges on the pre-condensed substrate can be ascertained from the charge density of the ion-decorated species. That is; even in cases where counterion condensation is prominent, (for samples of low DM or a blocky nature) experiments that measure the electrophoretic mobility can be still be useful indicators of the polyelectrolyte fine structure.

Finally, in chapter 4 the conformations of isolated HGs and the interaction between chains with different lengths and charge patterns were investigated. SAXS profiles and radial distribution functions for different chain concentrations were calculated and compared where possible with experimental data. The MD simulations match the experimental data well for isolated chains. Perhaps most importantly it was demonstrated that multi-chain MD simulations were capable of capturing the evolution of a structure factor (the so-called polyelectrolyte peak). In addition the behaviour of the peak in different ionic strength solutions and the predicted power-low scaling of the peak position with concentration were observed. This gives confidence that the technique can be usefully applied to situations where analytical expressions are unavailable, such as complex charge patterning of the chains.

In addition, further work ripe for study includes the further examination of how SAXS experiments at high q values (which are at the edge of the envelope in current Synchrotron experiments) might reveal the details of chain conformations in solution (previously the exclusive realm of fibre and crystal diffraction). Our simulations suggest that discrimination of conformations based on such experiments might be possible for the first time. In addition further simulations and experiments addressing the hydrogen bonding of uncharged chains (methylesterified and protonated chains) promise to reveal how local changes in charge, mediated by conformational changes, are reflected in different degrees of aggregation. These systems make excellent targets for further experimentation.

Appendix A

Glycam to Gromacs Unit Conversion

To be able to do MD simulation using Gromacs package we need to use an appropriate force field according to the chosen system. There are some force fields built in the Gromacs ready to use for MD simulations, however there are some more force fields not included in Gromacs. For example, GLYCAM has developed for carbohydrates, but it is not included in Gromacs package and we need to implement it in a way that can be used by Gromacs.

The latest GLYCAM parameters can be downloaded from `http://glycam.org/docs/forcefield/all-parameters/`, but they have different units compare to the force fields built in the Gromacs. Here we are doing a process for comparison between those parameters and will try to find a conversion factors for GLYCAM parameters to be used in Gromacs.

A.1 Bonded Parameters

A.1.1 Bond stretching

The first essential parameter in force fields is the bond parameter. In Gromacs, the bond stretching potential between atom i and j which are covalently bonded in a molecule is given by equation A.1 and presented in figure A.1.

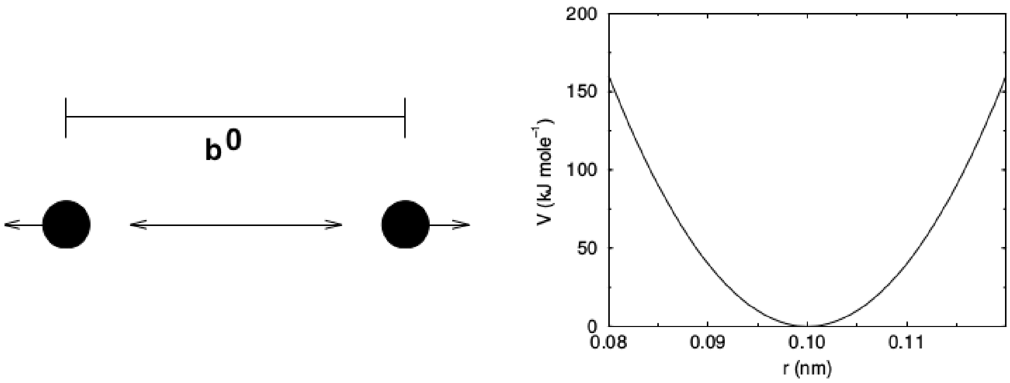


Figure A.1: schematic representation of bond stretching and its potential (Adopted from Gromacs manual).

$$V^b_{ij} = \frac{1}{2}k^b_{ij}(r_{ij} - b_{ij})^2 \tag{A.1}$$

where the $k^b_{ij}$is the spring constant and b_{ij} is the equilibrium distance between atom i and j. The unit of potential energy in Gromacs is KJ/mol and distance unit is nm.

In the other hand, the bond stretching potential used in GLYCAM is a bit different and it is given by;

$$V^b_{ij} = \mathbf{K}^B_{ij}(r_{ij} - b_{ij})^2 \tag{A.2}$$

where K^B is a constant value. The unit of potential energy in GLYCAM is KCal/mol and unit for distance is Angstroms. So in terms of using values given in the GLYCAM parameters file we do need a unit conversion as follow,

$$\begin{gathered}
V^b_{ij}(GLYCAM) = V^b_{ij}(Gromacs) \\
\mathbf{K}^B_{ij}(\frac{kCal}{mol \cdot \text{Å}^2})(r_{ij}(\text{Å}) - b_{ij}(\text{Å}))^2 = \frac{1}{2}k^b_{ij}(\frac{r_{ij}}{10}(\text{nm}) - \frac{b_{ij}}{10}(\text{nm}))^2 \\
\mathbf{K}^B_{ij}(\frac{4.184kJ}{mol \cdot \text{Å}^2})(r_{ij}(\text{Å}) - b_{ij}(\text{Å}))^2 = \frac{1}{2}k^b_{ij}(\frac{r_{ij}}{10}(\text{nm}) - \frac{b_{ij}}{10}(\text{nm}))^2 \\
therefore \\
836.8\mathbf{K}^B_{ij}(\frac{kJ}{mol \cdot \text{nm}^2}) = k^b_{ij}
\end{gathered} \tag{A.3}$$

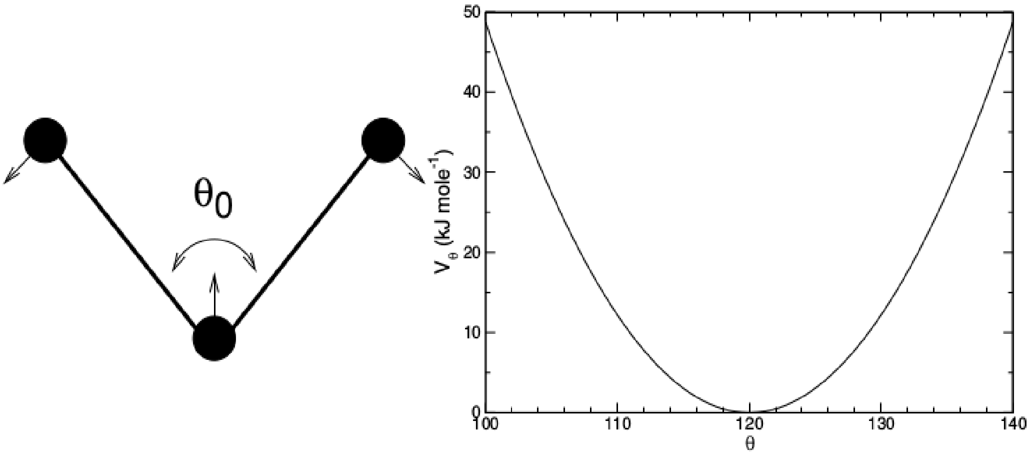


Figure A.2: schematic representation of angle vibration and its associated potential (Adopted from Gromacs manual).

So the conversion factor for bond stretch constant from GLYCAM to Gromacs is **836.8**, which means that it should be multiplied to the GLYCAM parameters to be suitable for Gromacs software.

A.1.2 Harmonic angle potential

For the angle potential we have the same sort of process to get GLYCAM parameters meaningful for Gromacs. Start from harmonic angle potential used in Gromacs we have;

$$V_{ij}^{a} = \frac{1}{2}k_{ijk}^{\theta}(\theta_{ijk} - \theta_{ijk}^{0})^{2} \tag{A.4}$$

wile in the GLYCAM the potential angle expression is;

$$V_{ij}^{a} = \mathbf{K}_{ijk}^{\theta}(\theta_{ijk} - \theta_{ijk}^{0})^{2} \tag{A.5}$$

considering a different energy unit and same angle unit for CLYCAM and Gromacs

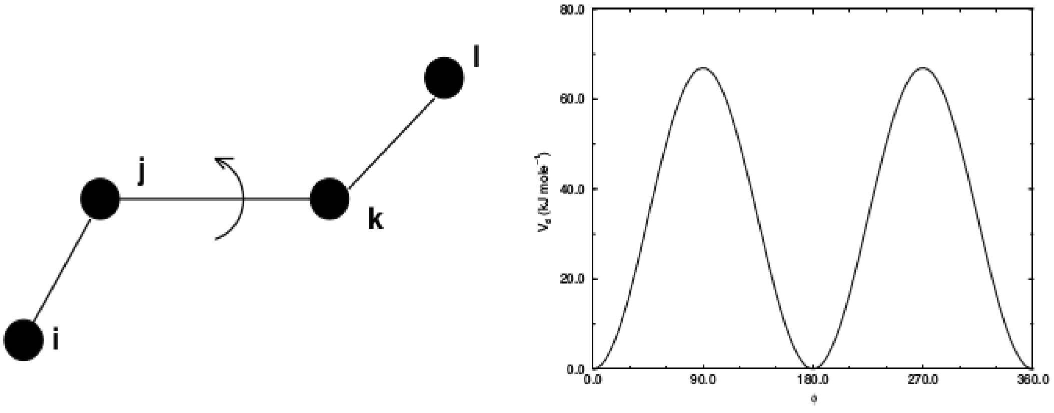


Figure A.3: schematic representation of a dihedral angle and its associated potential (Adopted from Gromacs manual).

leads to;

$$\begin{aligned}
V^a_{ijk}(GLYCAM) &= V^a_{ijk}(Gromacs)\\
\mathbf{K}^\theta_{ijk}(\frac{kCal}{mol \cdot rad^2})(\theta_{ijk}-\theta^0_{ijk})^2 &= \frac{1}{2}k^a_{ijk}(\theta_{ijk}-\theta^0_{ijk})^2\\
\mathbf{K}^\theta_{ijk}(\frac{4.184kJ}{mol \cdot rad^2})(\theta_{ijk}-\theta^0_{ijk})^2 &= \frac{1}{2}k^a_{ijk}(\theta_{ijk}-\theta^0_{ijk})^2\\
&therefore\\
8.368\mathbf{K}^\theta_{ijk}(\frac{kJ}{mol \cdot rad^2}) &= k^a_{ijk}
\end{aligned} \tag{A.6}$$

So in the case of angle potential the conversion factor from GLYCAM to Gromacs will be **8.368**.

A.1.3 Proper dihedrals

Finally the last part of bond's parameter which is important to describing the conformation of molecule is dihedral angle. Figure A.3 represents a dihedral angle which is an angle between ijk and jkl planes. The proper dihedral angle potential in Gromacs defines as;

$$V^d_{ijkl} = k^\phi_{ijkl}(1+\cos(n\phi-\phi_s)) \tag{A.7}$$

and in GLYCAM[1];

$$V^d_{ijkl} = [V/2]^\phi_{ijkl}(1+\cos(n\phi-\phi_s)) \tag{A.8}$$

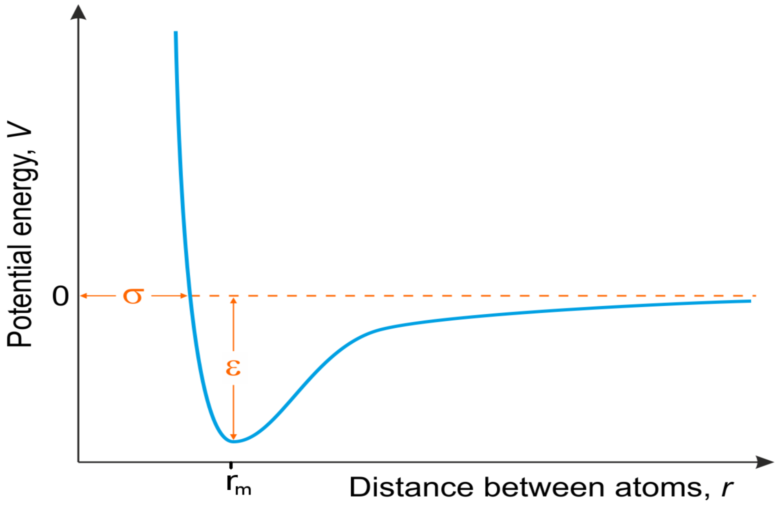


Figure A.4: Representation of Lennard-Jones potential versus distance between atoms.

The provided values in the proper dihedral part of GLYCAM parameters are belong to (V/2)[55], so the conversion process will be like;

$$\begin{gathered} V^{d}_{ijkl}(GLYCAM) = V^{d}_{ijkl}(Gromacs) \\ [V/2]^{\phi}_{ijkl}(\frac{kCal}{mol})(1+\cos(n\phi-\phi_s)) = k^{\phi}_{ijkl}(\frac{kJ}{mol})(1+\cos(n\phi-\phi_s)) \\ [V/2]^{\phi}_{ijkl}(\frac{4.184kJ}{mol})(1+\cos(n\phi-\phi_s)) = k^{\phi}_{ijkl}(\frac{kJ}{mol})(1+\cos(n\phi-\phi_s)) \\ therefore \\ 4.184[V/2]^{\phi}_{ijkl} = k^{\phi}_{ijkl} \end{gathered} \tag{A.9}$$

then the conversion factor for the proper dihedral potential is **4.184**.

A.2 Non-Bonded Parameters

In the non-bonded part of total potential acting on an atom, there are an electrostatic and Lennard-Jones potential. in the electrostatic potential $(k\frac{q_i q_j}{r^2_{ij}})$ there is no constant that needs any conversion from GLYCAM to Gromacs but in the Lennard-Jones potential we need to do some conversion.

The Lennard-Jones potential can be formulated in two ways as follow;

$$V_{LJ} = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{A.10}$$

where ϵ is the depth of potential and σ is the distance where the total potential is zero.

Parameter	**Conversion Factor**
Bonded Parameters	
Bond string constant	**836.8**
Angle potential constant	**8.368**
Proper dihedral constant	**4.184**
Non-Bonded Parameters(LJ)	
ϵ	**4.184**
σ **from** r_{VDW}	**0.17818**

Table A.1: Conversion coefficient for GLYCAM parameters to be suitable for Gromacs

This form of LJ potential is used in Gromacs. The other form of LJ potential which is used in GLYCAM can be written as;

$$V_{LJ} = \epsilon \left[\left(\frac{r_m}{r} \right)^{12} - 2 \left(\frac{r_m}{r} \right)^{6} \right] \tag{A.11}$$

where r_m is the distance between two atoms when the net force is zero and they are in equilibrium distance. Half of this equilibrium distance called Van der Waals radius, therefore $r_{VDW} = \frac{r_m}{2}$.

The Parameters given in the non-bonded part of GLYCAM file is for r_{VDW} and ϵ [55]. The unit of ϵ in GLYCAM is kCal/mol, so to convert the unit of ϵ to kJ/mol and be ready to use by Gromacs it is enough to times the given value by **4.184**. The other parameter that we need in Gromacs is σ. The relation between σ and r_m is;

$$r_m = 2^{\left(\frac{1}{6}\right)} \sigma \tag{A.12}$$

however the parameter given in GLYCAM is r_{VDW} which is half of r_m so;

$$r_{VDW} = \frac{r_m}{2} = \frac{2^{\left(\frac{1}{6}\right)} \sigma}{2} = 2^{\left(\frac{-5}{6}\right)} \sigma \tag{A.13}$$

by rearranging of equation A.13 and as the unit for length in GLYCAM is Angstroms then we have;

$$\sigma = 2^{\frac{5}{6}} \frac{r_{VDW}}{10} = \mathbf{0.17818} r_{VDW} \tag{A.14}$$

Finally, the conversion coefficient of GLYCAM raw parameters to be applicable parameters for the Gromacs can be summarized in following table;

By multiplying these conversion factors to the GLYCAM raw parameters we can get appropriate values can be used by Gromacs.

Bibliography

[1] B Lachele Foley, Matthew B Tessier, and Robert J Woods. Carbohydrate force fields. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(4):652–697, 2012.

[2] Helmut Schiessel. *Biophysics for beginners: a journey through the cell nucleus.* CRC Press, 2013.

[3] Amir Hossein Irani, Jessie L Owen, Davide Mercadante, and Martin AK Williams. Molecular dynamics simulations illuminate the role of counterion condensation in the electrophoretic transport of homogalacturonans. *Biomacromolecules*, 2017.

[4] Bruno FB Silva. Saxs on a chip: from dynamics of phase transitions to alignment phenomena at interfaces studied with microfluidic devices. *Physical Chemistry Chemical Physics*, 19(35):23690–23703, 2017.

[5] Dmitri I Svergun and Michel HJ Koch. Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics*, 66(10):1735, 2003.

[6] $www.ncnr.nist.gov/staff/hammouda/distance_learning/chapter\_26.pdf$.

[7] S Lankalapalli and VRM Kolapalli. Polyelectrolyte complexes: A review of their applicability in drug delivery technology. *Indian J. Pharm. Sci.*, 71(5):481, 2009.

[8] Michael Bouix, Jérôme Gouzi, Bernadette Charleux, Jean-Pierre Vairon, and Philippe Guinot. Synthesis of amphiphilic polyelectrolyte block copolymers using ”living” radical polymerization. application as stabilizers in emulsion polymerization. *Macromol. Rapid Commun.*, 19(4):209–213, 1998.

[9] Dmitriy Berillo, Linda Elowsson, and Harald Kirsebom. Oxidized dextran as crosslinker for chitosan cryogel scaffolds and formation of polyelectrolyte complexes between chitosan and gelatin. *Macromol. Biosci.*, 12(8):1090–1099, 2012.

[10] Jean-Paul Vincken, Henk A Schols, Ronald JFJ Oomen, Maureen C McCann, Peter Ulvskov, Alphons GJ Voragen, and Richard GF Visser. If homogalacturonan

were a side chain of rhamnogalacturonan I. Implications for cell wall architecture. *Plant Physiol.*, 132(4):1781–1789, 2003.

[11] Beda M Yapo. Pectic substances: From simple pectic polysaccharides to complex pectins a new hypothetical model. *Carbohydr. Polym.*, 86(2):373–385, 2011.

[12] Andrew N Round, Neil M Rigby, Alistair J MacDougall, and Victor J Morris. A new view of pectin structure revealed by acid hydrolysis and atomic force microscopy. *Carbohydr. Res.*, 345(4):487–497, 2010.

[13] Beda Marcel Yapo, Patrice Lerouge, Jean-François Thibault, and Marie-Christine Ralet. Pectins from citrus peel cell walls contain homogalacturonans homogenous with respect to molar mass, rhamnogalacturonan I and rhamnogalacturonan II. *Carbohydr. Polym.*, 69(3):426–435, 2007.

[14] Debra Mohnen. Pectin structure and biosynthesis. *Curr. Opin. Plant Biol.*, 11(3):266 – 277, 2008. Physiology and Metabolism - Edited by Markus Pauly and Kenneth Keegstra.

[15] Alexander Oosterveld, Gerrit Beldman, Henk A. Schols, and Alfons G.J. Voragen. Arabinose and ferulic acid rich pectic polysaccharides extracted from sugar beet pulp. *Carbohydr. Res.*, 288(0):143 – 153, 1996.

[16] Henk A. Schols, Esther Vierhuis, Edwin J. Bakx, and Alfons G.J. Voragen. Different populations of pectic hairy regions occur in apple cell walls. *Carbohydr. Res.*, 275(2):343 – 360, 1995.

[17] Yu.S. Ovodov, R.G. Ovodova, O.D. Bondarenko, and I.N. Krasikova. The pectic substances of zosteraceae : Part iv. pectinase digestion of zosterine. *Carbohydr. Res.*, 18(2):311 – 318, 1971.

[18] Christoph Rosenbohm, Inge Lundt, ToveM.I.E. Christensen, and NiallW.G. Young. Chemically methylated and reduced pectins: preparation, characterisation by 1h {NMR} spectroscopy, enzymatic degradation, and gelling properties. *Carbohydr. Res.*, 338(7):637 – 649, 2003.

[19] MaxwellS. Bush, Mazz Marry, MaxI. Huxham, MichaelC. Jarvis, and MaureenC. McCann. Developmental regulation of pectic epitopes during potato tuberisation. *Planta*, 213(6):869–880, 2001.

[20] Susana M. Cardoso, Artur M.S. Silva, and Manuel A. Coimbra. Structural characterisation of the olive pomace pectic polysaccharide arabinan side chains. *Carbohydr. Res.*, 337(10):917 – 924, 2002.

[21] Fernando Dourado, Pedro Madureira, Vera Carvalho, Ricardo Coelho, Manuel A. Coimbra, Manuel Vilanova, Manuel Mota, and Francisco M. Gama. Purification, structure and immunobiological activity of an arabinan-rich pectic polysaccharide from the cell walls of prunus dulcis seeds. *Carbohydr. Res.*, 339(15):2555 – 2566, 2004.

[22] Marie-Christine Ralet, Gwénaëlle André-Leroux, Bernard Quéméner, and Jean-François Thibault. Sugar beet (beta vulgaris) pectins are covalently cross-linked through diferulic bridges in the cell wall. *Phytochemistry*, 66(24):2800–2814, 2005.

[23] WilliamG.T. Willats, Lesley McCartney, ClareG. Steele-King, SusanE. Marcus, Andrew Mort, Miranda Huisman, Gert-Jan van Alebeek, HenkA. Schols, AlphonsG.J. Voragen, Angélique Le Goff, Estelle Bonnin, Jean-François Thibault, and J.Paul Knox. A xylogalacturonan epitope is specifically associated with plant cell detachment. *Planta*, 218(4):673–681, 2004.

[24] Bradley W Mansel, Che-Yi Chu, Andrew Leis, Yacine Hemar, Hsin-Lung Chen, Leif Lundin, and Martin AK Williams. Zooming in: Structural investigations of rheologically characterized hydrogen-bonded low-methoxyl pectin networks. *Biomacromolecules*, 16(10):3209–3216, 2015.

[25] Randall G Cameron, Yang Kim, Ashley L Galant, Gary A Luzio, and Jason TC Tzen. Pectin homogalacturonans: Nanostructural characterization of methylesterified domains. *Food Hydrocolloids*, 47:184–190, 2015.

[26] Ali Assifaoui, Adrien Lerbret, Huynh TD Uyen, Fabrice Neiers, Odile Chambin, Camille Loupiac, and Fabrice Cousin. Structural behaviour differences in low methoxy pectin solutions in the presence of divalent cations (ca 2+ and zn 2+): a process driven by the binding mechanism of the cation with the galacturonate unit. *Soft matter*, 11(3):551–560, 2015.

[27] Anna Ström, Pascual Ribelles, Leif Lundin, Ian Norton, Edwin R Morris, and Martin AK Williams. Influence of pectin fine structure on the mechanical properties of calcium-pectin and acid-pectin gels. *Biomacromolecules*, 8(9):2668–2674, 2007.

[28] Pär Davidsson, Martin Broberg, Tarja Kariola, Nina Sipari, Minna Pirhonen, and E Tapio Palva. Short oligogalacturonides induce pathogen resistance-associated gene expression in arabidopsis thaliana. *BMC plant biology*, 17(1):19, 2017.

[29] Alexis Peaucelle, Siobhan Braybrook, and Herman Höfte. Cell wall mechanics and growth control in plants: the role of pectins revisited. *Front. Plant. Sci.*, 3(121):1–6, 2012.

[30] Siobhan A Braybrook, Herman Hofte, and Alexis Peaucelle. Probing the mechanical contributions of the pectin matrix: insights for cell growth. *Plant Signaling Behav.*, 7(8):1037–1041, 2012.

[31] Gabriel Levesque-Tremblay, Jerome Pelloux, Siobhan A Braybrook, and Kerstin Müller. Tuning of pectin methylesterification: consequences for cell wall biomechanics and development. *Planta*, 242(4):791–811, 2015.

[32] Sebastian Wolf and Steffen Greiner. Growth control by cell wall pectins. *Protoplasma*, 249(2):169–175, 2012.

[33] Martin AK Williams, Aaron T Marshall, Padmesh Anjukandi, and Richard G Haverkamp. Investigation of the effects of fine structure on the nanomechanical properties of pectin. *Phys. Rev. E*, 76(2):021927, 2007.

[34] Romano Lapasin. *Rheology of industrial polysaccharides: theory and applications*. Springer Science & Business Media, 2012.

[35] Lisa M Kent, Trevor S Loo, Laurence D Melton, Davide Mercadante, Martin AK Williams, and Geoffrey B Jameson. Structure and properties of a non-processive, salt-requiring, and acidophilic pectin methylesterase from aspergillus niger provide insights into the key determinants of processivity control. *J. Biol. Chem.*, 291(3):1289–1306, 2016.

[36] Charith A Hettiarachchi, Laurence D Melton, Duncan J McGillivray, Simon M Loveday, Juliet A Gerrard, and Martin AK Williams. β-lactoglobulin nanofibrils can be assembled into nanotapes via site-specific interactions with pectin. *Soft matter*, 12(3):756–768, 2016.

[37] Amy Y Xu, Laurence D Melton, Geoffrey B Jameson, Martin AK Williams, and Duncan J McGillivray. Structural mechanism of complex assemblies: characterisation of beta-lactoglobulin and pectin interactions. *Soft matter*, 11(34):6790–6799, 2015.

[38] Yunqi Li, Qin Zhao, and Qingrong Huang. Understanding complex coacervation in serum albumin and pectin mixtures using a combination of the boltzmann equation and monte carlo simulation. *Carbohydr. Polym.*, 101:544–553, 2014.

[39] William G. T. Willats, Caroline Orfila, Gerrit Limberg, Hans Christian Buchholt, Gert-Jan W. M. van Alebeek, Alphons G. J. Voragen, Susan E. Marcus, Tove M. I. E. Christensen, Jørn D. Mikkelsen, Brent S. Murray, and J. Paul Knox. Modulation of the degree and pattern of methyl-esterification of pectic

homogalacturonan in plant cell walls: Implications for pectin methyl esterase action, matrix properties, and cell adhesion. 276(22):19404–19413, 2001.

[40] Malcolm O'Neill, Peter Albersheim, and Alan Darvill. 12 the pectic polysaccharides of primary cell walls. *Carbohydrates*, 2:415, 1990.

[41] Catherine MGC Renard and Jean-François Thibault. Structure and properties of apple and sugar-beet pectins extracted by chelating agents. *Carbohydr. Res.*, 244(1):99–114, 1993.

[42] Piet JH Daas, Brigitte Boxma, Anja MCP Hopman, Alphons GJ Voragen, and Henk A Schols. Nonesterified galacturonic acid sequence homology of pectins. *Biopolymers*, 58(1):1–8, 2001.

[43] Alphons GJ Voragen, Gerd-Jan Coenen, René P Verhoef, and Henk A Schols. Pectin, a versatile polysaccharide present in plant cell walls. *Struct. Chem.*, 20(2):263–275, 2009.

[44] James D Watson and Francis HC Crick. Molecular structure of nucleic acids. *Nature*, 171(4356):737–738, 1953.

[45] Jaime Gómez-Márquez. Dna g-quadruplex: structure, function and human disease. *FEBS J.*, 277(17):3451–3451, 2010.

[46] Bo Sun and Michelle D Wang. Single-molecule perspectives on helicase mechanisms and functions. *Crit. Rev. Biochem. Mol. Biol.*, pages 1–11, 2015.

[47] Paul WK Rothemund. Folding dna to create nanoscale shapes and patterns. *Nature*, 440(7082):297–302, 2006.

[48] Arun Richard Chandrasekaran. Programmable dna scaffolds for spatially-ordered protein assembly. *Nanoscale*, 8(8):4436–4446, 2016.

[49] John C Slater. A simplification of the hartree-fock method. *Phys. Rev.*, 81(3):385, 1951.

[50] Michael Baer. *Beyond Born-Oppenheimer: electronic nonadiabatic coupling terms and conical intersections*. John Wiley & Sons, 2006.

[51] KI Ramachandran, Gopakumar Deepa, and Krishnan Namboori. *Computational chemistry and molecular modeling: principles and applications*. Springer Science & Business Media, 2008.

[52] F Matthias Bickelhaupt and Evert Jan Baerends. Kohn-sham density functional theory: Predicting and understanding chemistry. *Rev. Comput. Chem.*, 15:1–86, 2007.

[53] Dennis C Rapaport. *The art of molecular dynamics simulation.* Cambridge university press, 2004.

[54] Andreas Kukol. *Molecular modeling of proteins*, volume 443. Springer, 2008.

[55] 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, 2008.

[56] Denis J Evans and Brad Lee Holian. The nose–hoover thermostat. *J. Chem. Phys.*, 83(8):4069–4074, 1985.

[57] EA Koopman and CP Lowe. Advantages of a lowe-andersen thermostat in molecular dynamics simulations. *J. Chem. Phys.*, 124(20):204103, 2006.

[58] JS Medina, Rita Prosmiti, Pablo Villarreal, Gerardo Delgado-Barrio, G Winter, B González, JV Aleman, and C Collado. Molecular dynamics simulations of rigid and flexible water models: Temperature dependence of viscosity. *Chemical Physics*, 388(1-3):9–18, 2011.

[59] Pekka Mark and Lennart Nilsson. Structure and dynamics of the tip3p, spc, and spc/e water models at 298 k. *The Journal of Physical Chemistry A*, 105(43):9954–9960, 2001.

[60] Qing Ji, Roland J-M Pellenq, and Krystyn J Van Vliet. Comparison of computational water models for simulation of calcium–silicate–hydrate. *Computational Materials Science*, 53(1):234–240, 2012.

[61] Carlos Pinilla, Amir H Irani, Nicola Seriani, and Sandro Scandolo. Ab initio parameterization of an all-atom polarizable and dissociable force field for water. *The Journal of chemical physics*, 136(11):114511, 2012.

[62] HJC Berendsen, JR Grigera, and TP Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91(24):6269–6271, 1987.

[63] Swaroop Chatterjee, Pablo G Debenedetti, Frank H Stillinger, and Ruth M Lynden-Bell. A computational investigation of thermodynamics, structure, dynamics and solvation behavior in modified water models. *The Journal of chemical physics*, 128(12):124511, 2008.

[64] William L Jorgensen. Quantum and statistical mechanical studies of liquids. 10. transferable intermolecular potential functions for water, alcohols, and ethers. application to liquid water. *Journal of the American Chemical Society*, 103(2):335–340, 1981.

[65] Herman JC Berendsen, James PM Postma, Wilfred F van Gunsteren, and Jan Hermans. Interaction models for water in relation to protein hydration. In *Intermolecular forces*, pages 331–342. Springer, 1981.

[66] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

[67] Martin AK Williams, Tim J Foster, and Henk A Schols. Elucidation of pectin methylester distributions by capillary electrophoresis. *J. Agric. Food Chem.*, 51(7):1777–1781, 2003.

[68] Huijuan-J Zhong, Martin A K Williams, David M Goodall, and Mette E Hansen. Capillary electrophoresis studies of pectins. *Carbohydr Res*, 308(1):1–8, 1998.

[69] AR Völkel and J Noolandi. On the mobility of stiff polyelectrolytes. *J Chem Phys*, 102(13):5506–5511, 1995.

[70] A R Völkel and J Noolandi. Mobilities of labeled and unlabeled single-stranded dna in free solution electrophoresis. *Macromolecules*, 28(24):8182–8189, 1995.

[71] Martin A K Williams, Gaëlle MC Buffet, and Tim J Foster. Analysis of partially methyl-esterified galacturonic acid oligomers by capillary electrophoresis. *Anal Biochem*, 301(1):117–122, 2002.

[72] Robert J Woods, Raymond A Dwek, Christopher J Edge, and Bert Fraser-Reid. Molecular mechanical and molecular dynamic simulations of glycoproteins and oligosaccharides. 1. glycam_93 parameter development. *J. Phys. Chem.*, 99(11):3832–3846, 1995.

[73] Frank Jensen. *Introduction to computational chemistry.* John wiley & sons, 2017.

[74] 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, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian 09 Revision D.01. Gaussian Inc. Wallingford CT 2009.

[75] Christopher I Bayly, Piotr Cieplak, Wendy Cornell, and Peter A Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the resp model. *J. Phys. Chem.*, 97(40):10269–10280, 1993.

[76] Herman JC Berendsen, J Pl M Postma, Wilfred F van Gunsteren, ARHJ DiNola, and JR Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684–3690, 1984.

[77] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, 52(12):7182–7190, 1981.

[78] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092, 1993.

[79] The PyMOL Molecular Graphics System, Version 1.7.2.1, Schrödinger, LLC. The PyMOL molecular graphics system, version 1.3r1. August 2010.

[80] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: visual molecular dynamics. *J. Mol. Graphics*, 14(1):33–38, 1996.

[81] In-Chul Yeh and Gerhard Hummer. Diffusion and electrophoretic mobility of single-stranded rna from molecular dynamics simulations. *Biophys. J.*, 86(2):681–689, 2004.

[82] In-Chul Yeh and Gerhard Hummer. System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions. *The Journal of Physical Chemistry B*, 108(40):15873–15879, 2004.

[83] Othonas A Moultos, Yong Zhang, Ioannis N Tsimpanogiannis, Ioannis G Economou, and Edward J Maginn. System-size corrections for self-diffusion coefficients calculated from molecular dynamics simulations: The case of co2, n-alkanes, and poly (ethylene glycol) dimethyl ethers. *The Journal of Chemical Physics*, 145(7):074109, 2016.

[84] Burcu Baykal Minsky, Anand Atmuri, Igor A Kaltashov, and Paul L Dubin. Counterion condensation on heparin oligomers. *Biomacromolecules*, 14(4):1113–1121, 2013.

[85] Gerrit Limberg, Roman Körner, Hans Christian Buchholt, Tove MIE Christensen, Peter Roepstorff, and Jørn Dalgaard Mikkelsen. Analysis of different de-esterification mechanisms for pectin by enzymatic fingerprinting using endopectin lyase and endopolygalacturonase ii from a. niger. *Carbohydrate Research*, 327(3):293–307, 2000.

[86] Richard G Haverkamp, Martin AK Williams, and John E Scott. Stretching single molecules of connective tissue glycans to characterize their shape-maintaining elasticity. *Biomacromolecules*, 6(3):1816–1818, 2005.

[87] John E Scott and Alison M Thomlinson. The structure of interfibrillar proteoglycan bridges ('shape modules') in extracellular matrix of fibrous connective tissues and their stability in various chemical environments. *J. Anat.*, 192(03):391–405, 1998.

[88] John E Scott. Cartilage is held together by elastic glycan strings. physiological and pathological implications. *Biorheology*, 45(3-4):209–217, 2008.

[89] SA Piripi, MAK Williams, and KG Thompson. On the sulfation pattern of polysaccharides in the extracellular matrix of sheep with chondrodysplasia. *Cartilage*, 2(1):36–39, 2011.

[90] Burcu Baykal Minsky, Thuy V Nguyen, Shelly R Peyton, Igor A Kaltashov, and Paul L Dubin. Heparin decamer bridges a growth factor and an oligolysine by different charge-driven interactions. *Biomacromolecules*, 14(11):4091–4098, 2013.

[91] Emek Seyrek and Paul Dubin. Glycosaminoglycans as polyelectrolytes. *Adv. Colloid Interface Sci.*, 158(1):119–129, 2010.

[92] A Basak Kayitmazer, Bonnie Quinn, Kozue Kimura, Gillian L Ryan, Angela J Tate, David A Pink, and Paul L Dubin. Protein specificity of charged sequences in polyanions and heparins. *Biomacromolecules*, 11(12):3325–3331, 2010.

[93] Martin AK Williams, Aurelie Cucheval, Abrisham Tanhatan Nasseri, and Marie-Christine Ralet. Extracting intramolecular sequence information from intermolecular distributions: highly nonrandom methylester substitution patterns in Homogalacturonans generated by pectinmethylesterase. *Biomacromolecules*, 11(6):1667–1675, 2010.

[94] Martin AK Williams, Aurélie Cucheval, Anna Ström, and Marie-Christine Ralet. Electrophoretic behavior of copolymeric galacturonans including comments on the information content of the intermolecular charge distribution. *Biomacromolecules*, 10(6):1523–1531, 2009.

[95] H-J Zhong, MAK Williams, RD Keenan, DM Goodall, and C Rolin. Separation and quantification of pectins using capillary electrophoresis: a preliminary study. *Carbohydr. Polym.*, 32(1):27–32, 1997.

[96] Gerald S Manning. Limiting laws and counterion condensation in polyelectrolyte solutions I. colligative properties. *J. Chem. Phys.*, 51(3):924–933, 1969.

[97] Marcia O Fenley, Gerald S Manning, and Wilma K Olson. Approach to the limit of counterion condensation. *Biopolymers*, 30(13-14):1191–1203, 1990.

[98] Gerald S Manning and Jolly Ray. Counterion condensation revisited. *J. Biomol. Struct. Dyn.*, 16(2):461–476, 1998.

[99] Markus Deserno, Christian Holm, and Sylvio May. Fraction of condensed counterions around a charged rod: Comparison of poisson-boltzmann theory and computer simulations. *Macromolecules*, 33(1):199–206, 2000.

[100] Marie-Christine Ralet, Vincent Dronnet, Hans Christian Buchholt, and Jean-François Thibault. Enzymatically and chemically de-esterified lime pectins: characterisation, polyelectrolyte behaviour and calcium binding properties. *Carbohydr. Res.*, 336(2):117–125, 2001.

[101] Suzette A Pabit, Kenneth D Finkelstein, and Lois Pollack. Using anomalous small angle x-ray scattering to probe the ion atmosphere around nucleic acids. *Methods Enzymol.*, 469:391–410, 2009.

[102] Suzette A Pabit, Steve P Meisburger, Li Li, Joshua M Blose, Christopher D Jones, and Lois Pollack. Counting ions around dna with anomalous small-angle x-ray scattering. *J. Am. Chem. Soc.*, 132(46):16334–16336, 2010.

[103] Klaus Huber and Ulrich Scheler. New experiments for the quantification of counterion condensation. *Curr. Opin. Colloid Interface Sci.*, 17(2):64–73, 2012.

[104] B Guilleaume, J Blaul, M Ballauff, M Wittemann, M Rehahn, and G Goerigk. The distribution of counterions around synthetic rod-like polyelectrolytes in solution. *Eur. Phys. J. E*, 8(3):299–309, 2002.

[105] David A Hoagland, David L Smisek, and Derek Y Chen. Gel and free solution electrophoresis of variably charged polymers. *Electrophoresis*, 17(6):1151–1160, 1996.

[106] Donald J Winzor, Lyle E Carrington, Marcin Deszczynski, and Stephen E Harding. Extent of charge screening in aqueous polysaccharide solutions. *Biomacromolecules*, 5(6):2456–2460, 2004.

[107] Qian Dong, Earle Stellwagen, John M Dagle, and Nancy C Stellwagen. Free solution mobility of small single-stranded oligonucleotides with variable charge densities. *Electrophoresis*, 24(19-20):3323–3329, 2003.

[108] Ivan M Peric, Bernabe L Riavs, Amalta Pooly, Elizardo Rifffo, and Luis A Basaez. Separation of copolymers of variable compositions by capillary zone electrophoresis. *Bol. Soc. Chil. Quim.*, 44(3):345–350, 1999.

[109] Robert L Cleland. Electrophoretic mobility of wormlike chains. 2. theory. *Macromolecules*, 24(15):4391–4402, 1991.

[110] Dirk Stigter. Electrophoresis of highly charged colloidal cylinders in univalent salt solutions. 1. mobility in transverse field. *J. Phys. Chem.*, 82(12):1417–1423, 1978.

[111] DA Hoagland, E Arvanitidou, and C Welch. Capillary electrophoresis measurements of the free solution mobility for several model polyelectrolyte systems. *Macromolecules*, 32(19):6180–6190, 1999.

[112] Davide Mercadante, Laurence D Melton, Geoffrey B Jameson, and Martin AK Williams. Processive pectin methylesterases: the role of electrostatic potential, breathing motions and bond cleavage in the rectification of brownian motions. *PloS one*, 9(2):e87581, 2014.

[113] Isabelle Braccini, Robert P Grasso, and Serge Pérez. Conformational and configurational features of acidic polysaccharides and their interactions with calcium ions: a molecular modeling investigation. *Carbohydrate Research*, 317(1-4):119–130, 1999.

[114] DA Rees. Polysaccharide conformation in solutions and gelsrecent results on pectins. *Carbohydrate Polymers*, 2(4):254–263, 1982.

[115] MD Walkinshaw and Struther Arnott. Conformations and interactions of pectins: I. x-ray diffraction analyses of sodium pectate in neutral and acidified forms. *Journal of molecular biology*, 153(4):1055–1073, 1981.

[116] MD Walkinshaw and Struther Arnott. Conformations and interactions of pectins: Ii. models for junction zones in pectinic acid and calcium pectate gels. *Journal of Molecular Biology*, 153(4):1075–1085, 1981.

[117] PG De Gennes, P Pincus, RM Velasco, and F Brochard. Remarks on polyelectrolyte conformation. *Journal de physique*, 37(12):1461–1473, 1976.

[118] Dibckmhj Svergun, Claudio Barberato, and Michel HJ Koch. Crysol–a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of applied crystallography*, 28(6):768–773, 1995.

[119] Michel HJ Koch, Patrice Vachette, and Dmitri I Svergun. Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Quarterly reviews of biophysics*, 36(2):147–227, 2003.

[120] Jonathan E. Kohn, Ian S. Millett, Jaby Jacob, Bojan Zagrovic, Thomas M. Dillon, Nikolina Cingel, Robin S. Dothager, Soenke Seifert, P. Thiyagarajan, Tobin R. Sosnick, M. Zahid Hasan, Vijay S. Pande, Ingo Ruczinski, Sebastian Doniach, and Kevin W. Plaxco. Random-coil behavior and the dimensions of chemically unfolded proteins. *Proceedings of the National Academy of Sciences*, 101(34):12491–12496, 2004.

[121] A Patkowski, Erdogan Gulari, and B Chu. Long range trna–trna electrostatic interactions in salt-free and low-salt trna solutions. *The Journal of Chemical Physics*, 73(9):4178–4184, 1980.

[122] Hideki Matsuoka, Norio Ise, Tsuneo Okubo, Shigeru Kunugi, Hiroshi Tomiyama, and Yukihiro Yoshikawa. orderedstructure in dilute solutions of biopolymers as studied by small-angle x-ray scattering. *The Journal of chemical physics*, 83(1):378–387, 1985.

[123] Kai Zhang. On the concept of static structure factor. *arXiv preprint arXiv:1606.03610*, 2016.

[124] MJ Abraham, D Van Der Spoel, E Lindahl, and B Hess. The gromacs development team gromacs user manual version 5.0. 4. *J. Mol. Model.*, 2014.

[125] George A Jeffrey and George A Jeffrey. *An introduction to hydrogen bonding*, volume 12. Oxford university press New York, 1997.

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
TO DOCTORAL THESIS CONTAINING PUBLICATIONS

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: Amir Irani

Name/Title of Principal Supervisor: M.A.K. (Bill) Williams

Name of Published Research Output and full reference:

On the electrophoretic mobilities of partially-charged oligosaccharides as a function of charge patterning and degree of polymerisation IN PRESS IN ELECTROPHORESIS

In which Chapter is the Published Work: 2

Please indicate either:

- The percentage of the Published Work that was contributed by the candidate: 75

  and / or

- Describe the contribution that the candidate has made to the Published Work:

  The candidate performed all the MD simulations and made a large contribution to the preparation of the paper, and was deservedly first author.

Amir Irani
Digitally signed by Amir Irani
Reason: I am the author of this document
Email: a.h.irani@massey.ac.nz
Date: 2018/05/12 18:06:54 +12'00'

Candidate's Signature

12/05/18

Date

D453D7AC-00B4-4793
-9747-45E26973BACE
Digitally signed by D453D7AC-00B4-4793-9747-45E26973BACE
DN: cn=D453D7AC-00B4-4793-9747-45E26973BACE, o=Massey University
Date: 2018.05.12 15:26:47 +12'00'

Principal Supervisor's signature

12/05/18

Date

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
TO DOCTORAL THESIS CONTAINING PUBLICATIONS

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: Amir Irani

Name/Title of Principal Supervisor: M.A.K. (Bill) Williams

Name of Published Research Output and full reference:

Molecular Dynamics Simulations Illuminate the Role of Counterion Condensation in the Electrophoretic Transport of Homogalacturonans
By: Irani, Amir H.; Owen, Jessie L.; Mercadante, Davide; et al.
BIOMACROMOLECULES Volume: 18 Issue: 2 Pages: 505-516 Published: FEB 2017

In which Chapter is the Published Work: 3

Please indicate either:

- The percentage of the Published Work that was contributed by the candidate: 75

  and / or

- Describe the contribution that the candidate has made to the Published Work:

  The candidate performed all the MD simulations and made a large contribution to the preparation of the paper, and was deservedly first author.

Amir Irani
Digitally signed by Amir Irani
Reason: I am the author of this document
Email: a.h.irani@massey.ac.nz
Date: 2018/05/12 18:05:29 +12'00'

Candidate's Signature

12/05/18

Date

D453D7AC-00B4-4793
-9747-45E26973BACE
Digitally signed by D453D7AC-00B4-4793-9747-45E26973BACE
DN: cn=D453D7AC-00B4-4793-9747-45E26973BACE, o=Massey University
Date: 2018.05.12 15:26:47 +12'00'

Principal Supervisor's signature

12/05/18

Date