First Principles and Molecular Dynamics Modelling of a Mucoid *Pseudomonas aeruginosa* Biofilm Extracellular Matrix

Oliver James Hills

A thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

Computational Chemistry, Biomaterials and Interfaces Modelling Group
School of Food Science & Nutrition
October 2022

Declaration

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others. To confirm, **Chapters 3**, **4**, **5** and **7** are based on work from jointly authored publications. For each of these publications (listed below), the work which is directly attributable to the candidate includes: conceptualization, methodology, investigation, visualisation, writing - original draft and writing - review and editing. The contributions of the other authors to this work include: project supervision, scientific discussion, software and writing - review and editing.

Chapter 3 contains work that was published in Hills OJ, Smith J, Scott AJ, Devine DA, Chappell HF, Cation complexation by mucoid *Pseudomonas aeruginosa* extracellular polysaccharide., *PLoS ONE*, 2021 **16(9)**, e0257026.

Chapter 4 contains work that was published in Hills, O.J., Yong, C.W., Scott, A.J., Devine, D.A., Smith, J., Chappell, H.F., Atomic-scale interactions between quorum sensing autoinducer molecules and the mucoid *P. aeruginosa* exopolysaccharide matrix., *Sci Rep*, 2022 **12**, 7724.

Chapter 5 contains work that was published in Hills, O.J., Yong, C.W., Scott, A.J., Smith, J., Chappell, H.F., Polyguluronate simulations shed light onto the therapeutic action of OligoG CF-5/20., *Bioorg. Med. Chem*, 2022, **72**, 116945.

Candidate signature: Oliver J. Hills **Date**: 02.10.2022

Acknowledgements

I would firstly like to thank the University of Leeds, School of Food Science & Nutrition for the fully funded 3-year PhD studentship. This PhD studentship has gone above and beyond to provide opportunities for me to develop as an early career research scientist, for which I am extremely grateful. By extension, I would like to thank my primary supervisor, Dr Helen Chappell, along-side all project co-supervisors, namely Dr James Smith, Dr Andrew Scott and Prof. Deirdre Devine, for their academic support and guidance. I have learned a great deal about Density-Functional Theory, Molecular Dynamics and bacterial biofilm biochemistry under their tutelage, which has further fueled my interest for computational chemistry and biophysical modelling, giving me a solid platform from which I aim to build a career in this area. I consider myself very lucky to have worked with you all.

I would like to offer thanks to Dr Chin Yong at the Science and Technology Facilities Council (STFC) for his support in rectifying software bugs in the DLFIELD support application for DLPOLY. His assistance was instrumental in the progression of my research project and I am extremely appreciative for his time and efforts. I would also like to extend thanks to Dr Jonathan Vernon for taking the time to read through my introduction chapter and offer helpful feedback.

Over the duration of my PhD, I have suffered with my mental health and if it were not for TalkWorks, a mental health cognitive behavioural therapy service offered by the Devon partnership NHS trust, I honestly believe I would have failed to complete my PhD. The TalkWorks service spent countless hours teaching me all the strategies I needed to cope with my obsessive compulsive disorder, anxiety and depression. I will always be forever grateful for their help and encourage all those who are struggling with their mental health, and feel like they need assistance, to seek help and support.

A most special thank you goes to my partner, Giuseppina Magri. You have always been there to support me, care for me and push me to reach my full potential. You are not only an inspiration to me, but are an inspiration to all those around you. I know you will continue to strive and achieve great things in your own PhD studies, and I could not be more

proud of you. I consider myself incredibly lucky to have you by my side.

I wish to thank my family for their support throughout my full academic journey, from a young age up until now. My mum and dad, Claire and Peter, have always done everything in their power to ensure that myself, as well as my siblings, never went without and always had access to opportunities both academic and non-academic. My siblings, Matthew, Amy and Katie, along-side my grandmothers, Molly and Sylvia, have always been there to care for me and pick me up when I am down. My mother- and father-in-law, Yvette and Vince, for offering me their home, care and support for the next chapter of my life. As a collective, you have all been there to support me in everything that I have ever done and I cannot fully express in words how thankful I am for this.

Lastly, I wish to express a heartfelt thank you to my grandfathers, Peter and Stanley, for cultivating my passion for science from such an early age. I am sorry you are both no longer here to see me progress throughout my scientific journey, but please know that you were both instrumental in forming the trajectory that has led me to where I am today. I hope I have made you both proud.

Abstract

Bacterial biofilms consist of bacterial populations enveloped and buried deep within an extracellular matrix (ECM). The ECM is a dense, tightly aggregated, scaffold of organic and inorganic material which behaves as a highly effective diffusion barrier, able to sequester incoming antimicrobials and functioning as a mode of adaptive antimicrobial resistance.

The Cystic Fibrosis (CF) lung is a prime infection site for mucoid *Pseudomonas aeruginosa* which, upon colonisation of the CF lung epithelia, establishes a biofilm matrix rich in acetylated, anionic, alginate exopolysaccharide (EPS). The EPS affords the enveloped bacteria resistance to host defenses and antimicrobial attack, manifesting a chronic infection state in the CF lung, which significantly shortens an already truncated life expectancy. As such, there is an urgent need to identify, at the atomic-scale, critical links between EPS biomolecular structure and pathogenic virulence. In particular, this includes elucidating the atomistic mechanisms behind mucoid biofilm stabilisation, penetration and eradication.

The aim of this research was to develop the first atomistic model of the mucoid *P. aeruginosa* biofilm EPS that was structurally representative of the EPS observed in the lungs of CF patients *in vivo*. Particular attention is given to the thermodynamically stable accommodation of key structural motifs unique to mucoid *P. aeruginosa*, as well as the EPS conformational change and intermolecular associations that drive cation-induced stabilisation, govern cell-to-cell signal distributions and facilitate mature matrix disruption.

The atomistic model critically implicated calcium ions in biofilm matrix chronicity, identified the molecular functionality pivotal for the interactions that govern molecular motion through the EPS, elucidated the atomistic mechanism underpinning mature EPS matrix disruption and extended the mechanistic understanding of exogenous gallium therapy. This information will prime pharmacologists with the understanding required to develop novel small molecule antimicrobials capable of treating chronic mucoid *P. aeruginosa* biofilm infections in the lungs of CF patients.

Contents

List of Figures

List of Tables

Abbreviations

3-oxo-C_{12}-HSL N-3-oxo-dodecanoyl-L-homoserine lactone

4-PolyM 4-chain system constructed from the stacking of two $PolyM_{ap}$ systems

4-PolyM* The most stable 4-PolyM stacking geometry

4-PolyMG 4-chain system constructed from the stacking of two $PolyMG_p$ systems

4-PolyMG_{MD} molecular model of a hydrated mucoid *P. aeruginosa* EPS matrix which exists at physiological temperature.

4-PolyMG* The most stable 4-PolyMG stacking geometry

algD Alginate biosynthetic gene cluster

aGM1/2 asialo-GM1/2

AHL Acetylated homoserine lactone

AMBER Assisted Model Building with Energy Refinement

BCC Body-centred cubic

C_4-HSL N-butanoyl-L-homoserine lactone

CASTEP Cambridge serial total energy package

CF Cystic Fibrosis

CFTR Cystic Fibrosis transmembrane conductance regulator

CFUs Colony forming units

CHARMM Chemistry at Harvard macromolecular mechanics

CL Cytoplasmic loop

CLSM Confocal laser scanning microscopy

CN Coordination number

COO^- Carboxylate group

DFT Density-Functional Theory

DNA Deoxyribonucleic acid

ECM Extracellular matrix

eDNA Extracellular deoxyribonucleic acid

EDS Energy dispersive spectroscopy

EPS Exopolysaccharide

EXAFS Extended X-ray absorption fine structure

FRET Fluorescence resonance energy transfer

FTIR Fourier transform infra-red

G α-L-guluronate

GGA Generalized gradient approximation

H3-T6SS H3 type VI secretion system

HCP Hexagonally closed packed

HGT Horizontal gene transfer

HHQ 2-heptyl-4-quinolone

HOMO Highest occupied molecular orbital

HS High-spin

HSL Homoserine lactone

ICP-MS Inductively coupled plasma mass spectroscopy

IL Interleukin

IV Intravenous

K Kelvin

LDA Local density approximation

LPS Lipopolysaccharide

LS Low-spin

LUMO Lowest unoccupied molecular orbital

M β-D-mannuronate

MD Molecular dynamics

MDR Multi-drug resistant

MIC Minimum inhibitory concentrations

MSD C-terminal membrane spanning domain

NBD N-terminal nucleotide-binding domain

NMR Nuclear magnetic resonance

OH Hydroxyl group

OMV Outer membrane vesicle

OPLS Optimised potentials for liquid simulations

PBE Perdew-Burke-Ernzerhof

PEGDA Poly(ethylene glycol) diacrylate

PES Potential energy surface

Phe Phenylalanine

PME Principle of minimum electrophilicity

PMH Principle of maximum hardness

PMN's Polymorphonuclear leukocytes

Poly-G_2 Poly-G_4 dimer

Poly-G_4 Guluronate quadramer model of OligoG CF-5/20

PolyM M1(C3) M3(C2) acetylated poly-β-D-mannuronate

PolyM_{ap} Two antiparallel associated PolyM chains

PolyMG M2(C3) acetylated poly-α-L-guluronate

PolyMG_p Two parallel associated PolyMG chains

PQS *Pseudomonas* quinoline signal

QQ Quorum quenching

QS Quorum sensing

QSAI Quorum sensing autoinducer

RMSD Root mean squared deviation

SEM Scanning electron microscopy

SPC Simple point charge

TBDR TonB-dependent membrane receptor protein

TEM Transmission electron microscopy

TNF Tumour necrosis factor

TS Tkatchenko and Scheffler dispersion correction scheme

vdW van der Waals

XANES X-ray absorption near edge structure

XC Exchange-correlation

XPS X-ray photoelectronic spectroscopy

Introduction

1.1 Pseudomonas aeruginosa

1.1.1 Gram-negative survivalist

Pseudomonas aeruginosa is a gram-negative, monoflagellated, rod bacterium that can incubate and multiply prolifically at temperatures ranging from 25°C to 37°C [1]. In fact, *P. aeruginosa* can populate and multiply in environments that reach temperatures of up to 42°C, which is a particular survival adaptation that distinguishes *P. aeruginosa* from other *Pseudomonas* species [1]. *P. aeruginosa* displays next to no nutritional requirements for its growth and survival, being able to obtain carbon and nitrogen from acetate and ammonia sources respectively [2]. These attributes give *P. aeruginosa* the ability to colonize and thrive in soil, water, vegetation, animals and human hosts [2]. The survival characteristics of *P. aeruginosa* allow it to successfully colonise environments such as the dry surfaces of hospital operating rooms and the surfaces of medical equipment, as well as sinks and showers [3]. In 2014, de Abreu *et al.* evaluated the occurrence of *P. aeruginosa* cells that persist on surfaces in a main hospital in Portugal over two years [3]. The authors reported that *P. aeruginosa* was repeatedly isolated from sinks, taps, bed-side tables, meal trays and medical equipment including oxygen flasks, stethoscopes and serum supports [3]. Persistence in hospital settings is causing an increase in the number of nosocomial infections such as urinary tract infections, surgical wound infections and Cystic Fibrosis (CF) lung infections [2]. The financial burden of nosocomial *P. aeruginosa* infections, specifically patients that have entered hospital for a non-*P. aeruginosa* related illness and are re-admitted due to *P. aeruginosa* infection acquired during the hospital visit, has been estimated to be approximately $70,000 per patient in the US alone as of 2017 [4].

Traditionally, *P. aeruginosa* infections have been treated using conventional antibiotics, but with their wide-spread use has come the development of antibiotic resistance. The discovery of new antibiotics accelerated on an upwards trajectory from the years 1928, following the discovery of penicillin, up until the 1960s [5]. Since then, novel discovery has

drastically slowed and is now failing to offer novel therapeutic agents capable of combating the increasing number of antibiotic resistant bacterial strains [5]. *P. aeruginosa* is one such strain that is capable of retarding antibiotic chemotherapy through three mechanisms: intrinsic resistance, acquired resistance and adaptive resistance [6]. These mechanisms are discussed in the following section.

1.1.2 Mechanisms of resistance

1.1.2.1 Low outer-membrane permeability

The outer-membrane of gram-negative bacteria is an asymmetrical lipid bilayer, synthesised inside the cell and assembled in the periplasmic space. It is composed of an inner leaflet of phospholipids and an outer leaflet of lipopolysaccharide (LPS) [7]. These combine to serve as a selective permeable barrier, protecting the cytoplasm in addition to regulating the excretion of waste products from inside the cell whilst at the same time regulating the transport of nutrients from outside the cell [8].

LPS encompasses a lipid A core, a short oligosaccharide bearing negatively charged phosphate groups that give the scaffold a net negative charge, with multiple fatty acid acyl chains branching off the central sugar residue (**Fig 1.1**) [8]. Specifically, *P. aeruginosa* synthesises a lipid A core that is a polymer of aminoarabinose, with a molecular weight ranging from 1400-1800 g/mol and terminated by negatively charged phosphate groups at carbon 1 and carbon 4' positions in the ring [9]. Anywhere between five and seven fatty acid acyl side chains, approximately sixteen carbon atoms in length, decorate the aminoarabinose oligomer [9]. In addition to the lipid A core, the LPS also harbours an O-antigen which is a long, repetitive glycan polymer [8]. *P. aeruginosa* produces two separate O-antigen structures, the A-band and the B-band [10]. The A-band lacks amino sugars [10] and is homopolymer of α-1-2 and α-1-3 glycosdidic linked rhamnose [11], whereas the B-band is more heteropolymeric in character, encompassing xylose, glucose, glucosamine and galactosamine, in addition to rhamnose [11]. In total, both the A- and B-bands account for approximately 15 % of total LPS scaffold architecture by molecular weight [10].

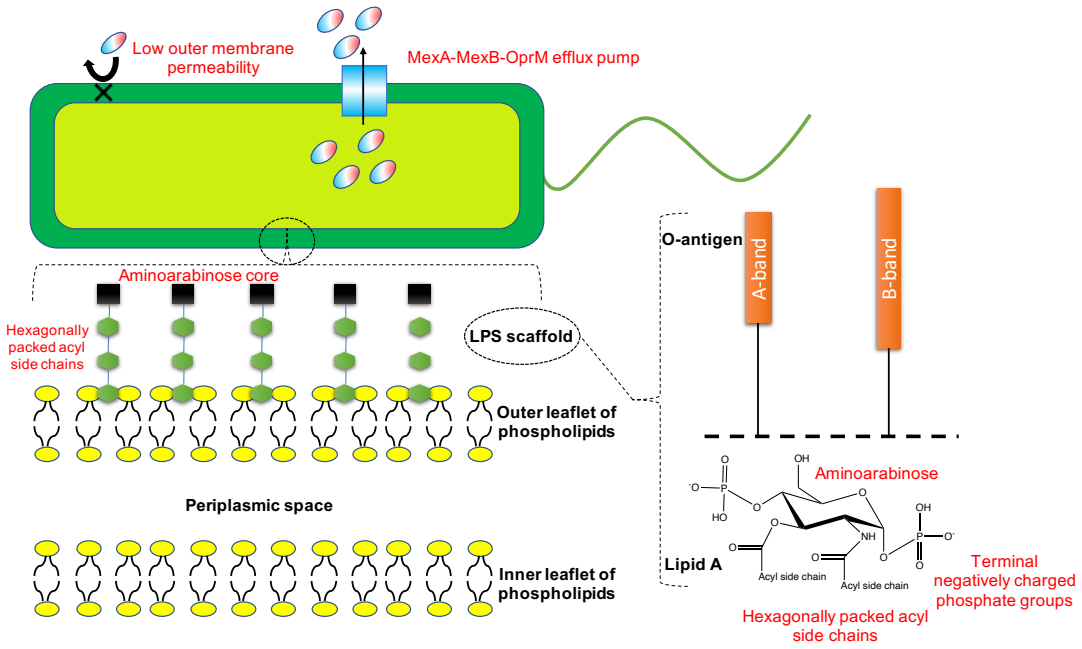


Figure 1.1: Graphical representation of *P. aeruginosa* antibiotic resistance mechanisms.

There is large structural heterogeneity in the monosaccharide composition of the B-band O-antigen structure [11]. In addition, there is also high cell-cell heterogeneity in O-antigen polymer length, the polymer length scales of both the A- and B-band structures varying anywhere between 1 and 100 monosaccharide residues [12]. Consequently, this makes it difficult to ascribe a single distinctive function to the O-antigen LPS component of *P. aeruginosa* [12]. That being said, some virulence functions have been linked to the O-antigen. For example, the O-antigen is able to modulate the hydrophobicity of the cell surface and assists in facilitating close cell-cell association and the subsequent formation of stacked, cohesive, multicellular aggregates [13]. These aggregates present with high cell densities and are capable of maturing to form bacterial biofilms [14]. Bacterial biofilms are a mode of adaptive extracellular resistance introduced in detail later in this chapter.

The lipid A component of the LPS bears influence over the mechanical stability of the LPS scaffold [8]. X-ray diffraction and conformational energy analysis have shown that hydrophobic interactions between the fatty acid acyl chains drive the lipid A to adopt a conformation in which its hydrophilic aminoarabinose backbone is oriented 5° relative to the membrane surface [15, 16]. This facilitates the low energy, highly ordered packing of the fatty acid acyl chains in a hexagonal lattice with an inter-chain separation of 4.9 Å [15, 16]. The lipid A phosphate groups offer suitable positions for binding divalent ions, such as calcium (Ca^{2+}) and magnesium (Mg^{2+}) [16], which electrostatically cross-link lipid A domains, establishing rigid, densely aggregated LPS scaffolds [17, 18]. The cross-linking divalent ions fix the inter-molecular separation between the anionic functionality on two lipid A domains,

decreasing the conformational flexibility and increasing the structural stability [18–20]. Recent coarse-grained molecular dynamics simulations have shown that the LPS is capable of withstanding surface tensions of up to 100 mN/m [21]. In contrast, the distribution of accessible finite temperature conformations for the O-antigen component is broad, indicative of conformational flexibility [20]. In turn, perhaps this suggests that conferring mechanical stability is not the primary role of the O-antigen LPS constituent.

Recently, meta-dynamics based free energy calculations have elucidated that the electrostatically cross-linked lipid A core provides a free energy barrier for the permeation of hydrophobic solutes, such as hexane, ethane and benzene, of 2-6 Kcal/mol [19]. Whereas, the hexagonally packed fatty acid acyl side chains provide a free energy barrier for the permeation of polar solutes, such as acetic acid and ethanol, of $\sim$ 5 Kcal/mol [19]. This clearly demonstrates how the two major domains of the LPS scaffold can work in tandem to provide the free energy barriers required to significantly reduce the permeation of both hydrophilic and hydrophobic solutes. In fact, the LPS is able to induce an approximate 4-fold reduction in the diffusion rates of hydrophilic and hydrophobic solutes, relative to their diffusion rates in water [19]. Regarding clinical chemotherapies, this translates to a significant reduction in the permeation of certain antibiotics, including nitrocefin, azithromycin, clarithromycin, roxithromycin and cephaloridine with polar functional groups [22, 23].

1.1.2.2 Chromosomally encoded efflux pumps

Chromosomally encoded efflux pumps are membrane proteins responsible for transporting, or *pumping*, antibiotics out of the cell [24]. In addition, these membrane proteins are also able to remove endotoxins and metabolites that build up in concentration as side products of metabolism. Efflux pumps are, therefore, critical in facilitating protection from molecules toxic to the bacterial cell [24].

Drug efflux is a process dependent upon protein-ligand binding affinity and protein conformational change. Most notably in *P. aeruginosa*, the membrane protein efflux pump that has the broadest antibiotic-spectrum activity, capable of pumping out a wide array of structurally diverse antibiotics, is the MexA-MexB-OprM efflux pump, encoded for by the *mexA-mexB-oprM* operon [25]. In fact, *mexA* and *oprM* null mutants, which fail to assemble the complete, functional, three-dimensional, MexA-MexB-OprM efflux transporter structure, possess no ability to efflux antimicrobial molecules such as tetracycline, chloramphenicol, norfloxacin and benzylpenicillin out of the cell [26]. In addition, transcriptional analysis of the MexA-MexB-OprM efflux pump biosynthesis genes has identified a large over-expression of the MexA-MexB-OprM efflux pump in *P. aeruginosa* isolates resistant to carbapenams [27], a form of β-lactam antibiotic. Taken collectively, this indicates that this particular efflux pump

confers resistance to both tetracycline and β-lactam based antibiotics [26, 27].

However, interestingly, the MexA-MexB-OprM efflux pump has a demonstrably malleable antibiotic specificity, conferring resistance to cationic aminoglycosides (in low ionic strength medium) [28], fluoroquinolones [29, 30] and, recently, carvacrol [31] - a herbal compound extracted from essential oils, which has shown promising bactericidal activity against *P. aeruginosa* [32]. The efflux mechanism of the MexA-MexB-OprM efflux pump has been visualised at the atomic-scale in recent equilibrium and non-equilibrium all-atom and coarse grained molecular dynamics simulations [33]. These simulations show that ten hydrogen bonding interactions occur between the α-hairpin of the MexA sub-unit and the OprM sub-unit, with the strongest of these interactions being the hydrogen bonds formed between aspartate (Asp-126), leucine (Leu-122) and glutamine (Gln-127) residues on the α-hairpin of MexA and lysine (Lys-149), tyrosine (Tyr-423) and alanine (Ala-220) residues on OprM [33]. This hydrogen-bonding contact interface lowers the conformational energy required for the OprM conformational change and stimulates an opening of an OprM aperture into the periplasmic space, measuring approximately 40 $Å^2$ in area [33]. This long tubular aperture permits the translocation of antibiotics, which occurs by passive diffusion through the full 15 nm distance of the periplasmic aperture and out of the cell [33]. In this work, the authors speculate that, as the translocation is achieved by diffusion, an increase in the drug concentration inside the cell - on the one side of the periplasmic aperture - would accelerate the rate at which the drug is excreted from the cell [33].

1.1.2.3 Horizontal gene transfer

Acquired resistance refers to antibiotic resistant genes acquired through horizontal gene transfer (HGT) [34]. This involves the passage of antibiotic resistance genes carried on plasmids (circular double-stranded DNA), transposons (transposable DNA sequences) and bacteriophages between bacteria of the same, or different, species [34]. HGT occurs through three mechanisms: conjugation, transduction and transformation [35]. Conjugation requires a close, direct, physical cell-cell contact between two bacterial cells. Such a cell-cell contact is referred to as the "sex pilus" [36] and allows antibiotic resistance genes on a plasmid to be transferred between the two cells. In contrast, for transduction, no such close direct physical contact is required and genetic material is transferred from one bacterium to another using bacteriophages as the delivery vectors [37]. Specifically, the bacteriophage gains entry to the donor bacterial cell, acquires genetic material and migrates to a recipient bacterial cell where it injects the foreign genetic material [35]. Direct cell-cell contact is also not required in transformation, that allows bacteria to take up fragments of genetic material that have been released into the extracellular environment [35], for example, as a result of cell lysis.

HGT has been implicated in the transferal of genes encoding aminoglycoside and β-lactam resistance, and such genes can accumulate to reach a prevalence of up to 43% and 76% respectively in *P. aeruginosa* colonies [38–40]. In fact, HGT is currently hypothesised to be the primary driver of aminoglycoside resistance in areas of the world such as Iran, where aminoglycosides have been heavily over-prescribed [38].

Adaptive resistance refers to biofilm formation and is the mechanism of *P. aeruginosa* resistance central to, and the primary focus of, this research and is introduced below.

1.2 Bacterial biofilms

1.2.1 Protection afforded through chemical cross-linking

Bacterial biofilms consist of bacterial cells enveloped and buried beneath an extracellular matrix (ECM), a hydrated mixture of polysaccharide, protein, DNA, (secondary) metabolites and ions [41] and these constituents are often referred to collectively as the "matrixome" [42]. Analogous to the proteome, the complete set of proteins expressed by an organism, and the genome, the entire set of DNA found within the nucleus of a cell, the matrixome constitutes all of the extracellular molecules and ions present within the ECM [42].

The distributions of these components vary from bacterial species and between phenotypes, however, the fundamental pathogenic, virulence, metabolic and cell signalling functions of these components remain common across species. Charged chemical functional groups of polysaccharides, DNA and proteins permit the formation of non-covalent electrostatic and charge-dipole interactions with inorganic ions and charged and polar constituents, leading to stable gel formation, which advances the cohesion and mechanical stability of the ECM [43]. In particular, cationic polysaccharides composed of cationic amino sugars, such as N-acetylgalactosamine and N-acetylglucosamine can electrostatically bind the anionic phosphate backbone of DNA and cross-link DNA fragments [44]. Polysaccharide-extracellular DNA (eDNA) associations are critical for governing biofilm matrix morphology. Specifically, polysaccharide-eDNA intermolecular cross-linking, facilitated through electrostatic and hydrogen bonding interactions, is responsible for establishing three-dimensional ECM scaffold architectures, assisting in cell-cell adhesion and the formation of a mature biofilm network [45, 46]. In addition, cationic polysaccharides composed of cationic amino sugars can undergo favourable electrostatic interactions with extracellular matrix protein adhesins [47], which offer protection from degradation caused by endogenous proteases [48]. However, it is also possible for these secreted extracellular protein adhesins to aggregate, in the absence of any exopolysaccharide or other substrate molecules, through the formation of protein-protein interactions [48]. In particular, these are hydrophobic interactions that bridge aliphatic

and aromatic sub-units, in addition to the electrostatic interactions that bridge the polar N- and O- containing sub-units, aggregating (flocculating) two different quaternary protein structures [49].

Neutral hydrophilic polysaccharides comprised of mannose, rhamnose and glucose can undergo charge-dipole and water-mediated hydrogen bonding interactions between their hydroxyl groups and secreted extracellular soluble lectins (glycoproteins) at sites encompassing threonine, asparagine and aspartate residues [50]. Polysaccharide-lectin interactions are believed to act as a *localising* interaction, tethering the bacterial cell surface to the ECM, conferring bacterial cell localisation and keeping bacteria retained within the ECM under laminar flow [51] and shear stress [50]. These same charge-dipole and water-mediated hydrogen bonding interactions facilitate the sorption of water molecules, forming hydrated polymer networks resistant to desiccation [43]. Non-destructive three-dimensional imaging by confocal Raman spectroscopy has quantified that up to 95% of the composition (by mass) of the ECM can be attributed to water molecules [52].

In addition to solidifying the ECM scaffold, these ECM constituents also facilitate nutrient storage [43]. DNA (eDNA) can sequester cations and cationic molecules through charge-dipole interactions facilitated through carbonyl oxygen lone pair electrons within cytosine-guanine (C-G) base pairs [53], but more predominantly, these electrostatic interactions are facilitated through the polyanionic phosphate backbone [54]. Extracellular anionic polysaccharides composed of carboxylate bearing sugars, such as mannuronate and guluronate (alginates), can also retain cations, such as Ca^{2+} and Mg^{2+}, at carboxylate sites through ionic interactions [55] with neutral polysaccharides utilising charge-dipole interactions at polar hydroxyl, ring and glycosidic sites [56]. Extracellular proteins can sequester diffusing ions by electrostatic attraction [57, 58] facilitated by residues such as aspartate for cations, asparagine for anions and, in addition, attracting aliphatic uncharged molecules through hydrophobic, π-π (if the uncharged molecule encompasses aromatic heterocyclic rings) and van der Waals (vdW) interactions [57], facilitated through residues such as tyrosine, tryptophan and phenylalanine.

Secreted extracellular constituents interact collectively to establish a dense, cohesive, complex biofilm ECM that is a highly effective physical (diffusion) barrier, protecting the enveloped bacteria from antimicrobial chemotherapies as well as attack from human host immune responses [59]. For antimicrobial molecules to penetrate the ECM, they are required to rupture these incredibly stable intermolecular (cross-linking) associations and, as such, the penetration of antimicrobials into the compact, mechanically stable, biofilm matrix is significantly impeded. Consequently, the minimum inhibitory concentrations of antibiotics to biofilm enveloped bacteria are up to 1000-fold higher in comparison to their

planktonic counterparts [41, 59]. As an illustration, for *P. aeruginosa* biofilm matrices, this corresponds to a 50-fold and 6-fold increase in resistance to the antibiotics tobramycin and ciprofloxacin respectively [60]. Although these molecules are not of direct relevance to this research, they offer suitable examples of the molecular functionality that can lead to molecular sequestration by the ECM. Tobramycin possesses multiple polar OH and NH_2 groups, expected to become electrostatically sequestered by the ECM, whereas ciprofloxacin encompasses conjugated aromaticity, as well as acidic functional groups, expected to be sequestered through hydrophobic, π-π and electrostatic interactions.

1.2.2 Cell surface attachment and biofilm formation

The chronology of biofilm formation begins with individual bacterial cells attaching to a surface. With regards to biotic (living) surfaces, such as epithelial cell surfaces within the human host, attachment is to the glycocalyx, the network of glycolipids protruding from the cell surface. Approximately 90% of the adhesion capability of *P. aeruginosa* cells to the glycocalyx on host cell surfaces during initial colonisation is reliant upon type IV pili-mediated adhesion [1]. Type IV pili are extremely thin, but very long (several microns in length), polymeric repeats of the pilin protein [61]. Type IV pili are classified into two main groups, type IVa and type IVb pili, with the former possessing repeat sequences 5-6 amino acids in length and the latter possessing repeat sequences 15-30 amino acids in length [61]. Type IVa pili are the primary pili adhesins used to adhere to the glycocalyx of epithelial cell surfaces and mucosal surfaces [1], achieved specifically through facilitating interactions with the cell surface glycolipids asialo-GM1 (aGM1) [62, 63] and asialo-GM2 (aGM2) [64, 65]. Type IV pili are considered to be critical in the attachment to epithelial cell surfaces, and medical device surfaces such as stainless steel, polystyrene and polyvinylchloride [66]. Unlike host cell attachment that relies on specific biophysical molecular recognition, abiotic surface attachment is governed by hydrogen bonding associations, with the primary pre-requisite being an energetically favourable, complementary steric and electrostatic molecular association [67].

Once adhered to host epithelial cells, the type IV pili initiate a twitching motility, a mechanism of repeated pilus contraction and extension capable of *pushing* the bacterial cell across the surface it has adhered to and can move at a rate of approximately 500 cell lengths per hour [68]. As such, the twitching motility is critical in facilitating colony aggregation and biofilm expansion [69–73]. *P. aeruginosa* cells harbouring mutations in the genes *pilA*, *pilB*, *pilC* and *pilY1*, which code for the biosynthesis of type IV pili, are unable to activate the twitching motility and cannot proliferate to form a biofilm [69, 70, 73]. Expansion over the host cell surface is accompanied by the synthesis and efflux of ECM material, proceeded by the subsequent maturation of the global biofilm architecture [74]. The process of differentiation from the planktonic form to the biofilm-forming phenotype

and the proliferation and maintenance of ECM architecture is regulated by the quorum sensing signalling system. This is introduced and discussed in more detail in **Section 1.4** (below). With enhanced resistance, the biofilm is able to flourish radially and the colony grows in size and population number, eventually forming a biofilm community where bacterial sub-populations are enveloped within a complex three-dimensional ECM scaffold architecture and these encased bacteria exist as a broad spectrum of phenotypes [75]. This is illustrated in**Figure 1.2**. Many bacteria form biofilms following the same developmental trajectory, independent of ecosystem and extracellular environment, and are even able to form in high shear locations such as industrial water systems and dental unit water lines [76]. When forming in high-shear locations, the resulting biofilm is smoother, more dense and possesses greater mechanical stability - often resistant to environmental deterioration [76].

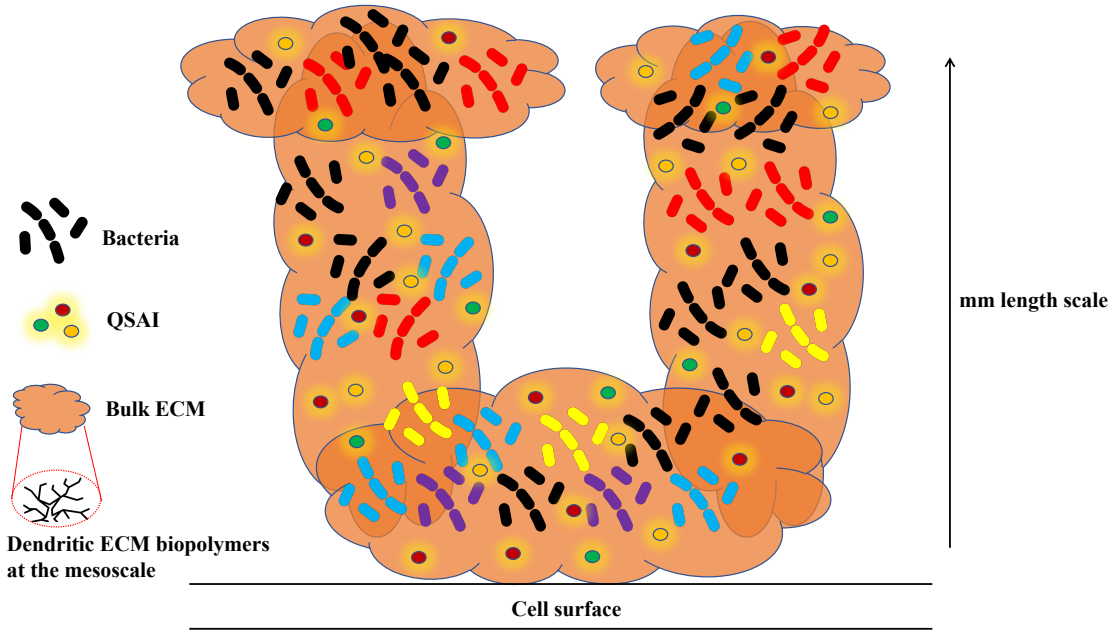


Figure 1.2: Illustration of a 3D biofilm matrix. The differently coloured bacteria represent a different phenotypic state. QSAI refers to quorum sensing autoinducer molecules which coordinate biofilm proliferation and maintenance (**Section 1.4**). At the mesoscale (μm) ECM biopolymers adopt a dendritic arrangement in 3D space. This is introduced, modelled and discussed in greater detail in **Section 4.4**.

In this thesis, the focus is on investigating the molecular basis and interactions occurring with the dominant ECM constituents and secondary metabolites, present within mucoid *P. aeruginosa* biofilm matrix. The context is the typical lung of a Cystic Fibrosis (CF) patient, introduced below.

1.2.3 Mucoid *P. aeruginosa* and the cystic fibrosis lung

Cystic Fibrosis (CF) is a genetic disease, acquired through autosomal recessive inheritance that is potentially life-threatening and prevalent in Caucasians but is most prevalent in the Caucasian populations of Northern Europe [77]. The origins of the disease lie with the defective *CF transmembrane conductance regulator* (*CFTR*) gene, a gene 250 kB in length residing on chromosome 7 and encoding the 1480 amino acid CFTR protein [78]. The function of the CFTR protein is to act as a chloride ion channel [79, 80] capable of regulating epithelial salt and fluid transport across epithelial membranes present within the lung, as well as other organs such as the pancreas, intestine and sweat glands [77]. The most common CF genetic mutation in Caucasians is the ΔF508-CFTR mutation and this accounts for 86% of all CF alleles occurring with a prevalence ranging from 1 in 1700 to 1 in 7700 live births in Europe [77, 78]. This mutation harbours a deletion of a phenylalanine (Phe) residue at position 508 from the N-terminal nucleotide-binding domain (NBD1) [77].

The Phe 508 residue resides on the surface of NBD1 and its aromatic side-chain establishes a hydrophobic interaction with leucine (Leu 1065), glycine (Gly 1069) and phenylalanine (Phe 1068 and Phe 1074) residues located within the cytoplasmic loop (CL4) in the C-terminal membrane spanning domain (MSD2), stabilising an NBD1-MSD2 intermolecular fold [81]. In addition, the hydrophobic interactions mediated by Phe 508 are also critical in facilitating the post-translational folding of NBD2 [82]. Therefore, without the Phe 508 residue, neither of these intermolecular conformational changes occur and, subsequently, there is a decrease in the folding yield of the CFTR protein [81], which reduces the quality of the CFTR protein [83]. Fewer correctly functioning CFTR proteins results in a failure to regulate chloride ion transport leading to airway-surface liquid dehydration and reduced mucociliary clearance [77]. As such, the phenotypic features (clinical symptoms) consistent with CF diagnosis include: sputum production, airway obstruction, chronic cough, persistent chest radiograph abnormalities and persistent bacterial infection [84]. Collectively, these features culminate in the progressive loss of lung tissue that leads to a severely reduced life expectancy [85]; the median life expectancy for patients suffering from Cystic Fibrosis is 31 years [86].

The CF lung is highly susceptible to chronic infection by *P. aeruginosa* [87]. This has been demonstrated in CF mice infection models, where CF lung epithelia expressing the ΔF508-CFTR mutation were 97 % more likely to contract *P. aeruginosa* infections compared to non-CF controls, a difference that is statistically significant ($p < 0.001$) [88]. Therefore, it is unsurprising that *P. aeruginosa* is the most prevalent pulmonary pathogen in CF patients and accounts for the majority of the mortality seen in CF disease [87]. Quantitative microbiological analysis of CF sputa isolated from 102 CF patients over a time duration of

22 months highlighted the much higher incidence of *P. aeruginosa* infection compared to infection by other commensal bacterial species [89]. Indeed, this study by Bauernfeind *et al.* demonstrated that 89 % of CF patients suffered from chronic *P. aeruginosa* infections and CF sputa microbial concentrations were highest for *P. aeruginosa*, specifically, greater than 10^9 colony forming units (CFUs) per ml of CF sputum [89]. This exceeds the measured CFUs of other bacterial species, such as *Staphylococcus aureus* ($\sim 10^6$ CFU/ml), *Staphylococcus epidermidis* ($\sim 10^6$ CFU/ml) and *Candida albicans* ($\sim 10^5$ CFU/ml), also found in CF sputum [89]. The hyper-susceptibility of the CF lung to chronic *P. aeruginosa* infections has been directly linked to the quantity of the CFTR protein produced [88]. Specifically, over-expression of CFTR facilitates the clearance of *P. aeruginosa* infections from the CF airway in far shorter time-scales [88] and this is missing in CF. Furthermore, there is a 9 % increase in the number of airway epithelial cells hosting aGM1 glycolipid in CF patients compared to non-CF controls [63]. Raised levels of aGM1 offer additional tethering points for *P. aeruginosa* type IVa pili, increasing the frequency of *P. aeruginosa* cellular adhesion, which increases *P. aeruginosa* prevalence in the CF lung [62].

Initial colonization in the CF lung is by the non-mucoid phenotype, but immune system attacks, alternating steep gradients of partial oxygen pressure, and competition with the host tissues for essential cations such as iron and zinc, drive the conversion to the mucoid phenotype [90]. The mucoid phenotype becomes the dominant variant in the CF patient lung [91–93] and is most responsible for chronic infection. The extent and stability of the infection is associated with its ability to form a dense biofilm able to retard pharmacological challenge [90]. Following mucoid biofilm formation, an overactive cellular immune response is triggered [94, 95]. Polymorphonuclear leukocytes (PMNs) are recruited into the CF airway to eliminate the infection but, as *P. aeruginosa* is protected within its biofilm matrix, these PMNs fail to clear the infection whilst damaging lung epithelial tissue as collateral [96]. Consequently, the occurrence of mucoid colonies in the CF lung leads to lung destruction and severely reduced pulmonary function in CF patients [93, 94, 97].

The genetic origin of the non-mucoid to mucoid conversion lies with the transcriptional activation of the alginate biosynthetic gene cluster (*algD*) [98] that encodes for alginate biosynthesis [59]. Sigma-factors are proteins that are capable of triggering the transcriptional activation of certain bacterial genes [99]. In contrast, anti-sigma factors are protein antagonists to sigma-factors [99]. The transcriptional activation of *algD* is achieved through a sequence of sigma-factor - anti-sigma factor associations. Ordinarily the membrane-bound anti-sigma factor, MucA, sequesters the sigma factor AlgU through forming an MucA-AlgU complex [59]. However, under envelope stress caused by perturbations to the outer-membrane, MucE is produced which binds the PDZ domain of the protease AlgW [100]. Perturbations could

include chemical perturbation, for example sharp gradients in nutrient and oxygen availability as well as attack from immune system molecules, and mechanical perturbation, for example the pressure force exerted on one bacterium as the colony expands and grows in population number. The MucE-AlgW association triggers the proteolytic activity of AlgW [101], which then proceeds to cleave the MucA-AlgU complex leaving the AlgU sigma-factor free to activate transcription of *algD* [59, 100].

The label "mucoid", therefore, is commonly interchanged with "alginate-overproducing" and mucoid *P. aeruginosa* in the CF lung produces an alginate rich biofilm where the extracellular matrix is predominantly anionic acetylated exopolysaccharide (EPS) alginate [1, 59, 93, 102, 103]. In fact, large abundances of alginate can be visibly seen in the autopsied lungs of CF patients who have died from chronic *P. aeruginosa* (biofilm) infections [93]. Indeed, chemical composition analysis of extracellular material from mucoid and non-mucoid *P. aeruginosa* strains grown *in vitro* indicates 5-32 times the amount of alginate in the mucoid biofilm compared to other extracellular constituents, such as proteins [104–106]. The structure of the mucoid *P. aeruginosa* EPS is described in detail in below.

1.3 Mucoid *P. aeruginosa* EPS matrices

1.3.1 Polyuronate backbone

The mucoid *P. aeruginosa* biofilm matrix is predominantly composed of a polysaccharide that is structurally a co-polymer of two uronate sugars: β-D-mannuronate (M) and its C5 epimer α-L-guluronate (G), linked via a C1–C4 glycosidic bond [103, 106, 107]. The **L** and **D** sterioisomers are the dominant enantiomers and the two anionic uronate sugar residues, along-side the Fischer projections for the acid forms, are displayed below in **Fig 1.3**. It is important to note here, that the C2, C3 and C4 carbon positions are also chiral centres, but here we are only considering the C5 epimer.

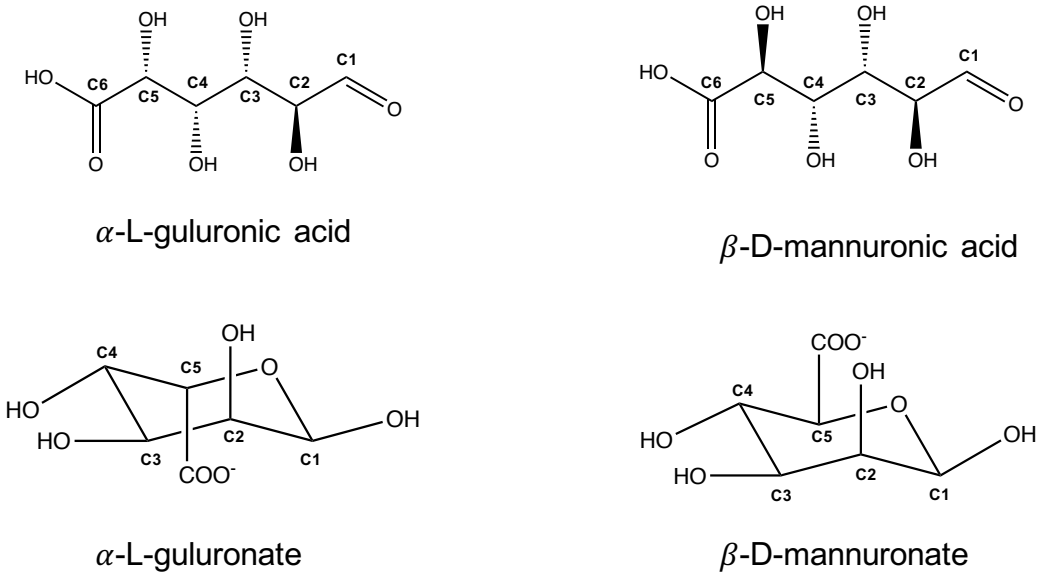


Figure 1.3: β-D-mannuronate (M) and α-L-guluronate (G) units along-side the Fischer projection for the acid forms.

The pKa values of mannuronic acid and guluronic acid are 3.38 and 3.65 respectively. Therefore, in the CF lung, which has a pH in the range of 6.85-7.65 [108, 109], the carboxylic acid groups are deprotonated and the polysaccharide chains are considered anionic. This gives the polysaccharide the ability to bind extracellular cations elevated in the CF lung, and form stable cation cross-linked scaffolds. These ion complexations and the structural changes are considered and discussed in detail later in this thesis.

Acetyl groups are also present, located solely on mannuronate residues [106, 110–112]. Specifically, ^{1}H-NMR studies have identified that the degree of M unit acetylation is diverse, ranging between 4-57 %, and the acetyl groups are bound at the C2 or C3 positions in the ring [111, 112]. This is illustrated in **Fig 1.4** below.

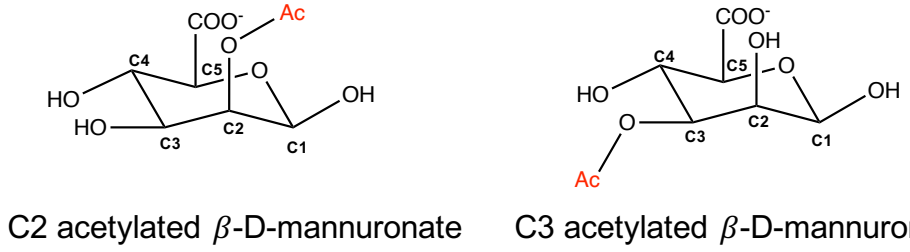


Figure 1.4: C2 and C3 acetylated β-D-mannuronate (M) units. The red Ac denotes an acetyl group (CH_3CO) which is attached through the carbon atom to the oxygen atom of the hydroxyl group located on the C2 or C3 carbon atom positions in the ring.

Acetylation involves the substitution of a H atom from the hydroxyl group for a CH_3CO

moiety. Extracellular enzymatic acetylation of mannuronate residues, at the hydroxyl groups located at C2 or C3 carbon positions, is encoded for by the *algI*, *algJ* and *algF* genes [113–115]. Activation of these genes encodes the production of the AlgI-AlgJ-AlgF protein complex, which spans the bacterial inner membrane and periplasmic space, and performs the acetylation [114, 115]. Intriguingly, the source of the acetyl group (CH_3CO), which allows the AlgI-AlgJ-AlgF protein complex to catalyse the acetylation, is unknown [116] and, as such, so is the acetylation mechanism. However, the mechanism may possibly follow the standard acetylation pathway of alcohols as is shown in **Fig 1.5**.

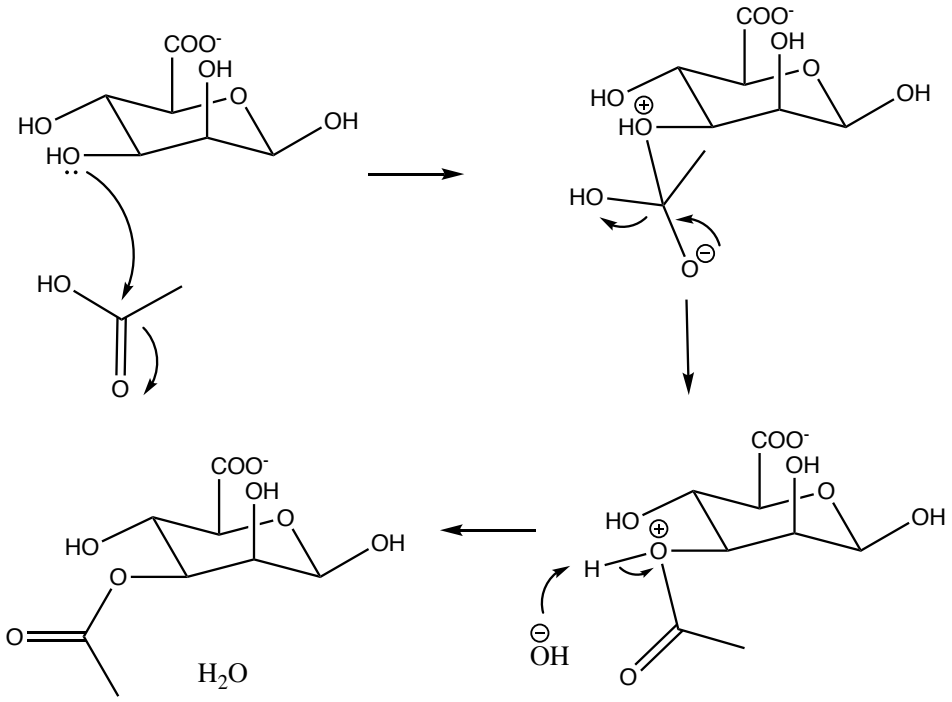
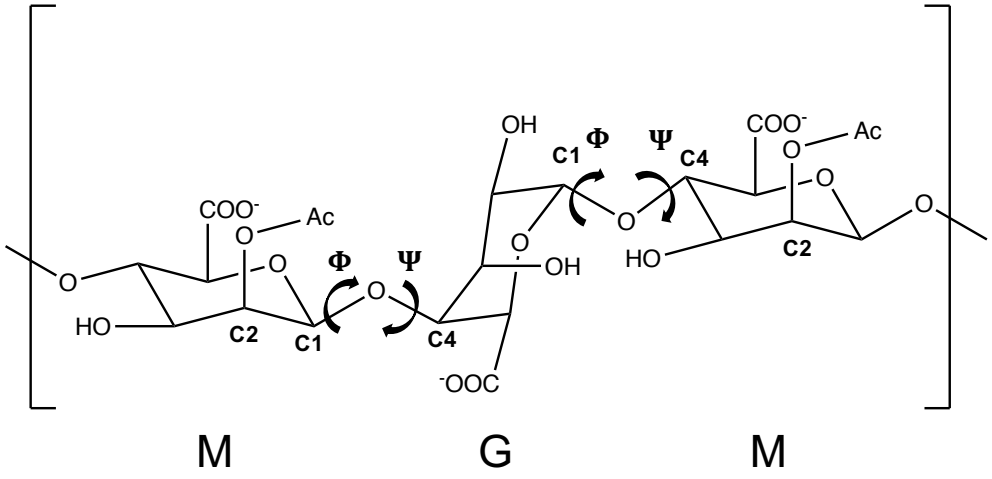


Figure 1.5: Alcohol acetylation reaction mechanism to give C3 acetylated β-D-mannuronate. Note, the source of the CH_3CO moiety in this scheme is acetic acid.

Acetylation of the alginate exopolysaccharide enhances cell-cell aggregation and micro-colony formation, giving rise to thicker, more viscous, biofilm matrices that are better able to retain cells within the matrix [117] and which, therefore, can encase higher cell densities [117, 118]. Additionally, the acetylated alginate exopolysaccharide produced by *P. aeruginosa* is resistant to challenge by opsonic phagocytosis [119] and alginate lyase [103, 112]. Acetylation is an advantageous phenotypic motif for *P. aeruginosa* cells that have colonised the CF airway, but it also prevents the mannuronate-to-guluronate epimerisation reaction from taking place. This means, (acetylated) M units cannot be epimerised to G units [113] and, consequently, polymer sequences of G units (G-blocks) do not occur throughout the entirety of the EPS scaffold [106, 111, 112]. In fact, guluronate inclusion peaks at 50 % [106], which is only satisfied when alternating M-G junctions are present throughout the EPS architecture.

Alongside the presence of acetyl groups, M-G junctions define structural motifs unique to mucoid *P. aeruginosa* and distinguishes bacterial alginate from its well-characterised algal counterpart [107, 120]. An example mucoid *P. aeruginosa* EPS structure is illustrated in **Fig 1.6**.



$$\Phi = \text{O5-C1-O-C4}$$

$$\psi = \text{C1-O-C4-C5}$$

Figure 1.6: An example mucoid *P. aeruginosa* MGM polysaccharide chain. This example illustrates a co-polymer of mannuronic (M) and guluronic (G) residues linked via a C1-C4 glycosidic bond with acetylation at the C2 of each M unit. The torsion angles about the glycosidic linkages are also illustrated.

The anionic carboxylate groups, in addition to hydroxyl and acetyl functional groups, are present along the length of the EPS scaffold and are the functional groups capable of forming ionic and hydrogen bonding interactions between separate EPS structures [121]. In addition, these same functional groups can accommodate inorganic cations, promoting polysaccharide aggregation and gel formation, giving rise to densely aggregated EPS scaffolds possessing the physical characteristics of a solid (viscoelastic) material, namely, a high storage modulus and low loss modulus [121–123]. The accommodation of extracellular cations present in the CF lung is discussed below.

1.3.2 Elevated cationic atmosphere in the CF lung

The CF lung is a host environment which possesses elevated concentrations of certain biological cations. In addition to the biologically ubiquitous sodium ion, which is the most significantly elevated ion in the CF lung [124], inductively coupled plasma mass spectroscopy (ICP-MS) analysis on samples of expectorated sputum from CF sufferers has shown significantly elevated

levels of magnesium (30 mg/L), calcium (102 mg/L), iron (797 μg/L) and zinc (1285 μg/L) cations compared to non-CF controls (magnesium: 4mg/L; calcium: 45 mg/L; iron: 0 μg/L; zinc 179 μg/L) [125]. Iron, magnesium and zinc had the largest increase compared to non-CF sputum samples, but when considering patients with *P. aeruginosa* infections specifically, it was only magnesium (2.5 times higher compared to non-CF controls) and calcium (7.5 times higher compared to non-CF controls) that showed significant elevation compared to samples from patients with other common CF-infections [125]. Intriguingly, the presence of calcium ions in CF sputum is linearly correlated with the presence of magnesium ions (Pearson correlation r-value = 0.89), suggesting that the disease process causing the raised calcium levels is also causing the raised magnesium levels [125].

The occurrence of these ions in the CF lung also shows a moderate positive correlation (Pearson correlation r-value = 0.42) with the occurrence of interleukin (IL)-8 [125]. (IL)-8 is a chemoattractant cytokine responsible for neutrophil recruitment into the CF airway, causing inflammatory damage even before the onset of bacterial infection [126, 127]. Upon bacterial colonisation, however, (IL)-8 levels are further raised, increasing in concentration as the microbial community grows in size [128, 129]. The positive correlation between the levels of sputum ions and levels of (IL)-8 in the CF airway supports one possible hypothesis that these elevated ion levels, which resemble the concentrations found in whole blood (calcium: 46.5 mg/L; magnesium: 22 mg/L), arise from cell necrosis during local tissue damage and bleeding [125].

Interactions between these cations and negatively charged functional groups on the EPS result in EPS stabilising physico-chemical cross-linking interactions [55]. In the absence of any interactions with cations, the EPS matrix is stabilised solely through entanglements [123]. However, as observed through ^{13}C-NMR spin-lattice relaxation measurements, upon exposure to cations such as calcium and magnesium, ion-chelate complexes are established between M-M and/or M-G junctions on neighbouring EPS chains [130]. This causes the mucoid *P. aeruginosa* EPS chains to engage in tight ionic cross-linking [123], establishing a permanent, direct, ionic bridging interaction between M-M and/or M-G junctions from opposing EPS chains [122]. This ionic interaction rigidifies the EPS network, which increases the overall mechanical strength of the biofilm [122, 123]. Now, within these rigid EPS networks, entanglements only offer minor contributions to overall EPS stability with electrostatic attractions (coulomb interactions) being the dominant stabilising inter-molecular interaction [123].

Ca^{2+} ions have been shown to cause a broadening of the CHOH-carbon chemical shifts, compared to Mg^{2+} ions, in solid-state ^{13}C-NMR studies on hydrated *P. aeruginosa*

biofilms [130]. The binding of calcium ions, but not magnesium ions, to form stable chelate complexes within M-M and M-G junctions induces EPS aggregation, slowing functional group rotational motion and increasing the anisotropy of the system [130]. This is direct spectroscopic evidence to suggest the binding of magnesium ions, relative to calcium ions, to the EPS is weaker [130]. Similar observations have been made for sodium ions, which induce weak, transient cross-linking throughout the polysaccharide, giving rise to softer, mechanically weaker, gel-like structures [122, 131].

Calcium ions are well-retained in the EPS matrix in experimental systems, in fact there is a strong, positive, linear correlation between the calcium content of the agar medium in which the mucoid EPS biofilm matrix has been grown *in vitro* and the calcium content of the mature EPS matrix [132]. The retention of calcium ions within the EPS matrix leads to thicker, more granular, mechanically stable biofilm matrices that are more difficult to detach *in vitro* [132, 133]. Specifically, the *P. aeruginosa* biofilm matrix thickens ten to twenty-fold upon exposure to calcium and this increase in biofilm thickness is attributed to the calcium cross-linked M-M and M-G junctions enhancing preservation of the EPS as a gel within the biofilm matrix in addition to up-regulating *algD* [133]. Calcium ions, therefore, are highlighted as key ions for *P. aeruginosa* virulence.

1.3.3 Computational studies of alginate-metal ion complexes

The binding of metal ions to algal alginates has been well studied using theoretical models based on classical molecular dynamics (MD). MD simulations studying poly-β-D-mannuronate and poly-α-L-guluronate 2-chain aggregation have shown that the carboxylate (COO^-) group is the only uronate functional group responsible for mediating the association of two alginate chains about calcium [134–136]. The calcium-algal alginate free energy landscape is "rough" [135], meaning it is a landscape encompassing widely varying gradients in the free energy [135]. This gives rise to many local minima corresponding to accessible meta-stable calcium-bound aggregates [135]. However, in the most stable calcium-alginate complexes, the COO^- group is the only functional group implicated in cation binding [135, 136].

Transition path sampling, using MD, has clarified the atomistic motions that facilitate the energetically favourable self-organisation and calcium induced cross-linking in poly-α-L-guluronate systems, which begins with hydroxyl groups forming hydrogen bond interactions, bringing two poly-α-L-guluronate chains into close proximity [137]. The Ca^{2+} ions then associate with the COO^- groups of a single chain, which is proceeded by the reorientation of the COO^-groups in the second chain and the establishment of COO^- - Ca^{2+} - COO^- bridges [137]. MD simulations studying the interaction between Na^+ ions and alginate chains, based on poly-β-D-mannuronate, poly-α-L-guluronate and copolymeric

β-D-mannuronate-α-L-guluronate decamers, have identified that Na^+ ions, unlike Ca^{2+} ions, are *unable* to mediate the aggregation of two alginate chains [136, 138, 139]. Specifically, in poly-α-L-guluronate systems, sodium ions weakly associate with COO^- groups, are not elevated in the chelate pocket and, consequently, sit too deep in G-G junctions to attract a second chain [139].

More accurate theoretical models based on quantum chemical Density-Functional Theory (DFT), studying the interaction between algal alginate and group 2 divalent cations, highlighted that hydroxyl, glycosidic and ring oxygen atoms can also participate in cation binding [140, 141]. Interactions from these particular functional groups are ionic in nature and increase in bond length (typically by 0.3 Å) as the ionic radius of the central divalent ion is increased from Mg^{2+} to Ca^{2+} to Sr^{2+} [140]. Additionally, bond lengths to the central M^{2+} cation are also insensitive to the presence of (water) solvent molecules [140, 142], meaning that explicit water molecules have negligible influence on the energetically stable chelated geometries adopted between M-M and G-G junctions [140, 142]. Coordination numbers (CN), principally the number of ionic bonds formed between the uronate oxygen functionality and the central ion, of five or six are achieved upon divalent cation complexation in M-M and G-G junctions [140, 141]. The COO^--M^{2+} interactions are, on average, 0.1 - 0.2 Å shorter compared to the interactions with other uronate oxygen functionality [140, 141], which is an indicator of increased bond strength. It can therefore be interpreted that the COO^- groups contribute most to the stability of the divalent ion chelation.

In a recent theoretical study, again based on DFT, the stability as defined by chemical hardness (η) of M-M, M-G, G-M and G-G cation-disaccharide complexes was predicted to decrease as the complexed ion species were incrementally changed from Mg^{2+} to Ca^{2+} and then to Na^+ [143]. Chemical hardness is a global reactivity index that estimates the (conformational) stability of the resulting complex, with the largest chemical hardness indicating the most energetically stable chelate geometry [144]. This particular trend in chemical hardness of the ion-algal alginate complexes, specifically Mg^{2+} being greater than Ca^{2+} which is greater than Na^+, is a trend inversely proportional to the ionic radius, suggesting the complexation event is purely electrostatic in origin [143]. That is, the more charge density on the cation, the more stable the complexation. However, it also suggests that Mg^{2+} ions form more stable (cross-linked) chelate geometries compared to Ca^{2+} ions, something that is neither observed in experimentally manufactured cation-algal alginate gels [145] nor in the ^{13}C-NMR spectroscopic characterisation of bacterial alginates *in vitro* [130].

Both experimental and theoretical studies have aimed to shed light on the contributions

specific cations make to the stability of alginate complexes upon the establishment of ionic cross-links. However, there is still disagreement as to which ions produce the most stable cross-linked alginate structures. Furthermore, theoretical investigations into cross-linking acetylated bacterial alginates are sparse and therefore little is known about key geometrical and chemical features that drive stable bacterial alginate aggregation, how this differs from the case of (classical) algal alginates and which ions are implicated in mucoid *P. aeruginosa* biofilm chronicity in the CF lung.

1.4 Virulence of the colony

1.4.1 Quorum sensing signalling pathways

Production and maintenance of vast biofilm architectures is reliant upon quorum sensing (QS) signalling, the ability of organisms to respond to diffusible metabolites that illicit intracellular signalling responses. QS pathways are responsible for the differentiation from planktonic individuals into biofilm forming communities through the production of cell-to-cell signals [146], called quorum sensing autoinducers (QSAIs) [147]. Bacterial communities which have inactivated QS systems (i.e., do not produce and transmit QSAIs) form flat, structurally homogeneous, undifferentiated biofilms, directly implicating QSAI production and transfer in the formation of morphologically differentiated, three-dimensional biofilm matrix structures [146].

As a mechanism of bacterial cell-to-cell communication, QS is dependent upon the bacterial population density. When a single bacterium releases a QSAI molecule, it is too low in concentration to be detected by neighbouring bacteria. However, when QSAI molecules are released collectively by enough bacteria, the concentration of these molecules increases past a threshold level, such that they are now detectable by the neighbouring bacteria. In response, the bacteria perceive a critical cell mass and coordinate the activation (or suppression) of certain genes [147, 148]. Specifically, for *P. aeruginosa* aggregates in a CF lung infection model, QSAI release from approximately 2000 cells was required to initiate a community-based QS [149]. However, larger aggregate sizes, greater than 5000 cells, can synthesise larger concentrations of QSAIs and can communicate with neighbouring aggregates at further distances, up to 180 μm away from source [149]. Activation of QS then proceeds to activate genes responsible for virulence factor expression, including bacterial aggregation and biofilm formation [146, 150], biosurfactant production [151] and pyocyanin production [152].

Biosurfactants increase the hydrophobicity of the bacterial cell surface which increases the adhesive capabilities of the bacteria and contributes to adhered micro-colony formation [153]. In addition, biosurfactants are able to micellise portions of the biofilm matrix [154] as well as

activate the type IV pili mediated twitching motility [153]. Collectively, these not only assist in the cellular migration to the biofilm surface to form the distinctive mushroom-like 3D architectures [153], but also, in the formation and maintenance of water channel networks [154].

Pyocyanin is a redox active phenazine, capable of producing reactive oxygen species, which impose oxidative stress on host cells, leading to an inflammatory response [155]. In addition, pyocyanin is toxic and functions as a form of *molecular defense*, neutralising competition from foreign bacteria [156]. The toxicity of pyocyanin also makes it an "auto-poisoning" phenazine, which initiates cell death and the subsequent release of bacterial DNA material into the biofilm matrix [157], which can then be acquired through transformation (a mechanism of HGT, see **Section 1.1.2**) or contribute to matrix three-dimensional scaffold stability (by chemical cross-linking, see **Section 1.2.1**). In addition, pyocyanin toxicity is increased when there is limited nutrient (carbon, phosphorous and sulphur) availability, as abstraction of these nutrients from the extracellular environment exacerbates the ability of pyocyanin to generate reactive oxygen species [157]. A possible interpretation for this increased toxicity may be that to sacrifice certain members of the colony preserves resources in the most efficient way possible [157].

P. aeruginosa and most other gram-negative bacteria utilise acetylated homoserine lactones (AHLs) as their primary QSAI molecules [147, 148], which are specific to the LasR and/or RhlR transcriptional activators [158]. The concentration of AHLs present within sputum isolated from CF patients is highly similar to the concentration present within *in vitro* cultured *P. aeruginosa* biofilms, indicating that QS is active in the lungs of CF patients infected with *P. aeruginosa* [159]. Specifically, the two primary AHL signals present in sputum isolated from CF patients are N-butanoyl-L-homoserine lactone (C_4-HSL) and N-3-oxo-dodecanoyl-L-homoserine lactone (3-oxo-C_{12}-HSL) [160].

C_4-HSL is specific to the RhlR transcriptional activator [161] and the concentration of C_4-HSL within *P. aeruginosa* biofilms increases over time in the biofilm life-cycle, with its concentration within the biofilm matrix approximately doubling every four hours [162]. Without the time-dependent increase in C_4-HSL concentration, the biofilm fails to proliferate [163] and produce virulence factors such as pyocyanin [162]. Taken collectively, this suggests that C_4-HSL is critical for biofilm maturation [162, 163]. 3-oxo-C_{12}-HSL is specific to the LasR transcriptional activator [164] and, in contrast, its concentration remains low and static over the full duration of the *P. aeruginosa* biofilm life cycle [162]. 3-oxo-C_{12}-HSL concentrations exceed C_4-HSL concentrations at the very inception of the biofilm, but once the biofilm begins proliferating and increasing in surface coverage and biomass, the concentration of 3-oxo-C_{12}-HSL falls dramatically below the concentration of C_4-HSL [162].

As such, it is believed that 3-oxo-C_{12}-HSL is required for the initial differentiation to the biofilm mode of life [146, 163], rather than facilitating matrix maturation. Notwithstanding, transcriptome analysis of the *P. aeruginosa* genome has revealed that, in combination, these two AHL-QS systems regulate approximately 6% of the total *P. aeruginosa* genome [165].

The *Pseudomonas* quinoline signal (PQS) is the third key QSAI molecule utilised by *P. aeruginosa* in the CF lung [166, 167]. RhlR and LasR activation are required for PQS production [166] and, as such, PQS is understood to behave as a "connecting signal" between the *las* and *rhl* systems [166, 168] and is produced maximally when cultures reach the late stationary phase of growth (between 30-42 hours) [168]. This is at a time where the cell density has already become stable and, therefore, unlike both C_4-HSL and 3-oxo-C_{12}-HSL, PQS is not a QSAI signal involved in perceiving cell density [168]. The occurrence of PQS at later stages in biofilm development explains why there are higher levels of PQS, relative to C_4-HSL and 3-oxo-C_{12}-HSL, present in the sputum isolated from the bronchoalveolar space of end-stage CF lungs [167]. Furthermore, the concentration of PQS present in sputum samples taken from CF patients positively correlates with the estimated *P. aeruginosa* cell density present within the sputum, in turn, establishing a connection between PQS signal production (and transmission) and chronic infection [167].

PQS is able to sustain the chronic infection state as it contributes to regulating biophysical processes that maintain biofilm matrix architecture. PQS regulates the production of rhamnolipids, which are compounds that behave as biosurfactants, able to micellise the EPS, creating cavity and void spaces, maintaining water channel structure [154]. Water channels ensure the continued flow of water soluble nutrients to cells buried deep beneath the matrix scaffold [169, 170]. In addition, PQS stimulates the autolysis of *P. aeruginosa* cells within the biofilm [171], permitting the release of intracellular material such as proteins and DNA that, as discussed in **Section 1.2.1**, can further stabilise the biofilm matrix through intermolecular interactions with other matrix constituents. Finally, PQS binds iron (Fe^{3+}) and assists in siderophore (secondary metabolites with high ferric binding affinities) uptake of Fe^{3+} into the cell [172–174]. This is discussed in further detail in **Section 1.5.2**.

1.4.2 QSAIs in the EPS matrix

Confocal laser scanning microscopy and fluorescent lectin binding studies have shown that surrounding naturally occurring bacterial sub-populations within *P. aeruginosa* biofilms, there exist three distinct ECM layers that combine to segregate both individual cells and individual sub-populations [175]. The first layer of ECM surrounds individual bacterial cells, the second (intercellular ECM) is a layer separating individual cells and, finally, the third (inter-micro-colony ECM) is a layer separating different sub-populations [175]. The

central implication is that the QSAI molecules must pass through ECM material to reach neighbouring bacteria. In fact, ecological bioreactor investigations have observed that QSAI molecules are able to partition from surrounding bulk liquid into the biofilm matrix, facilitating transmission between cells throughout the biofilm [150, 176]. Furthermore, *P. aeruginosa* biofilms grown in poly(ethylene glycol) diacrylate (PEGDA) chambers, can produce QSAI molecules able to diffuse throughout the hydrogel (diffusion coefficient, D, $\sim$ 0.08-0.13 mm^2 min^{-1}), creating a signal concentration gradient detected by *P. aeruginosa* cells up to 8 mm from the source [177]. Indeed, the genetic responses, and thus virulence factor expression, as well as the biofilm encased cell density, are directly positively correlated with the concentrations of QSAI molecules that successfully partition in the biofilm matrix; when these concentrations are low, the biofilm matrix collapses [150].

The molecular structures of the three primary QSAI molecules utilised by *P. aeruginosa* can be seen in **Fig 1.7**.

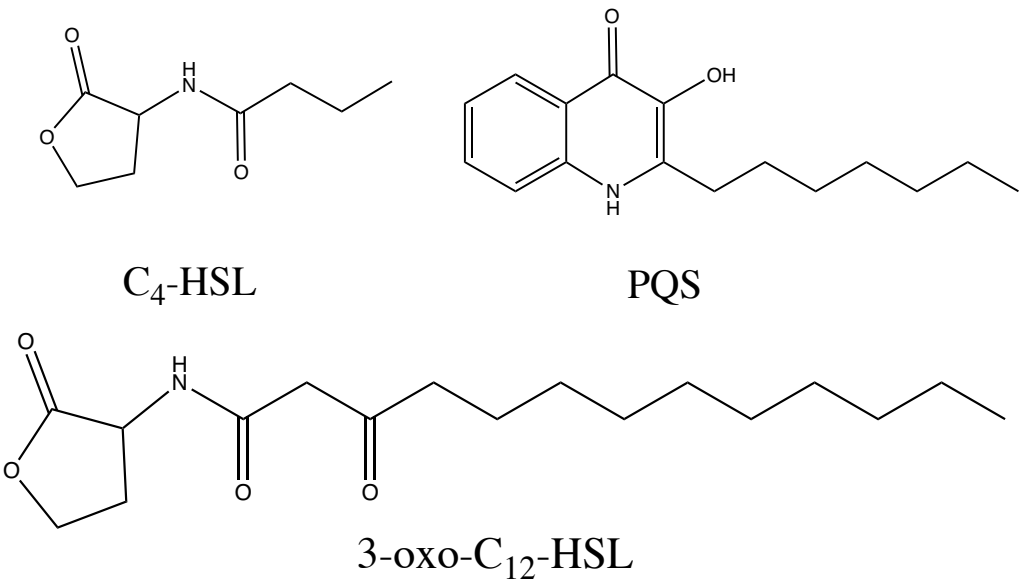


Figure 1.7: Molecular structures of C4-HSL and PQS. For reference, the structure of 3-oxo-C_{12}-HSL is also given.

Given that QSAI signals establish a signal concentration gradient as they migrate away from the source, biofilm cells in close proximity to a QSAI-emitting cell experience enhanced growth rates, whereas, bacterial cells further away receive lower concentrations of the QSAIs and experience slower growth [177]. However, intriguingly, the atomistic mechanism(s) behind the passage of QSAIs throughout the biofilm EPS matrix is poorly understood and the observed signal concentration gradients are currently understood to be due to simple diffusion [178, 179]. Such models may be insufficient to completely characterise, and even predict, QSAI signal distributions in the biofilm matrix as such matrices are structurally heterogeneous in

three-dimensions and will encounter completely different magnitudes of fluid (water) flow rate dependent on the extracellular environment [178, 179]. For example, the velocity of water flow through the biofilm matrix formed within the mucus of the CF lung will be completely different to flow rate through biofilm matrices formed in locations such as urinary catheters, dental unit water lines and kidney stones [179]. In light of these uncertainties, a more detailed atomic-scale understanding is required to accurately characterise, based on inter-molecular interactions, differences in QSAI motion through the biofilm matrix.

In summary, there is very little information on how both C_4-HSL and PQS interact with the EPS, and which molecular functional groups and molecular interactions dictate the motion of these molecules throughout the EPS. For example, it is unknown whether these molecules can interact with the EPS through chemisorption, which are direct, irreversible, interactions electrostatic or covalent in nature. Equally, it is unknown whether physisorption, reversible interactions that originate from vdW forces, also govern molecular motion. Finally, it is unknown whether disruption, the breaking of EPS-EPS molecular associations to create an aperture through which to move, is responsible for transfer of QSAI molecules between cells. Knowledge of this is critical for pharmaceutical design and EPS penetration, for example, understanding the functionality governing why certain QSAI molecules extend further afield (display slowly varying concentration gradients over longer ranges and are non-local signals) can guide the design of pharmaceutical agents aimed at deep biofilm matrix penetration.

1.5 Emerging therapeutics

Currently, the mucoid *P. aeruginosa* biofilm within the CF lung is able to evade pharmacological challenge by conventional antibiotics (see **Section 1.2.1**), exposing the need for a new breed of novel antimicrobials. This need is a quickly accelerating field of research [180]. Introduced in this section are two emerging therapeutic strategies, OligoG CF-5/20 and intravenous gallium therapy, which display examples of unique physiological mechanisms targeted at mucoid *P. aeruginosa* biofilm eradication within the CF lung.

1.5.1 OligoG CF-5/20 therapy

OligoG CF-5/20 is an anionic guluronate alginate oligomer, encompassing ≥ 85 % guluronate abundance, and which can be safely inhaled into the lungs of CF patients and excreted naturally by the GI tract [181]. An example OligoG CF-5/20 structure can be seen in **Fig 1.8** below.

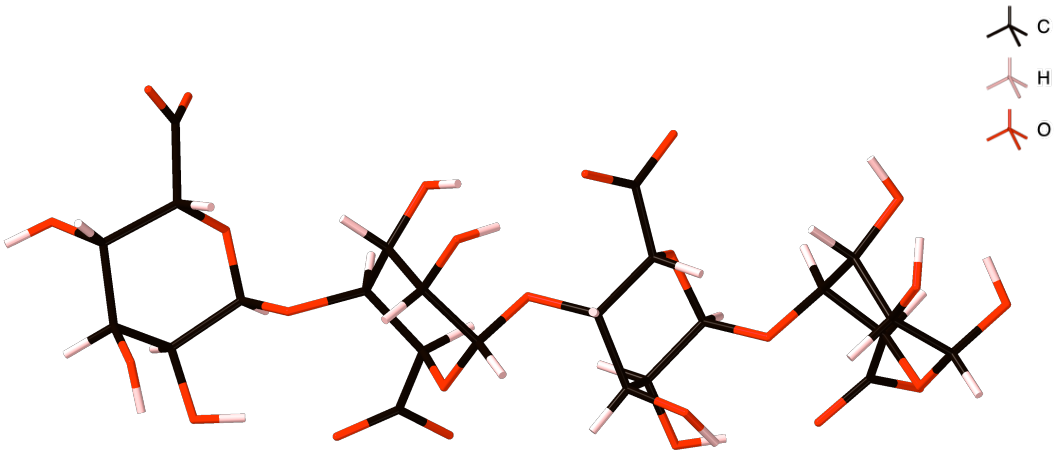


Figure 1.8: An example molecular structure of anionic OligoG CF-5/20 encompassing 100 % guluronate abundance.

Fourier transform infra-red (FTIR) spectroscopy, coupled with MD simulations, have provided evidence to suggest that OligoG CF-5/20 is able to bind to MUC5AC, the predominant CF airway mucin [181]. As well as being able to interact directly with the MUC5AC sulphated Lewis-X antigen [182], a sulphate moiety on the glycan residues of the mucin scaffold critical for leukocyte homing, OligoG CF-5/20 hydroxyl groups are able to establish hydrogen bonds to nitrogen and oxygen atoms on the MUC5AC peptide backbone [181, 182]. Most notably, these are the N and O atoms that form the amide linkages [181], potentially as these N and O atoms are sp^2 hybridised and contain lone-pair electrons free to participate in hydrogen bonding. Recall in **Section 1.3**, the pKa of a guluronate carboxyl group is 3.65 which means, just as with the EPS carboxyl groups, these groups are deprotonated at physiological pH, making the OligoG CF-5/20 molecule negatively charged. In addition, as mentioned above, the sp^2 oxygen atoms on the amide bond carbonyls have exposed electrons for hydrogen bonding, which means these atoms also encompass a partial negative charge. Therefore, the contribution of these two electronic structural features means that when binding to CF mucin, the net negative surface charge of CF mucin is increased [181]. The surface charge of CF mucin at pH 7 before OligoG CF-5/20 therapy is -16.7 mV, but after exposure to OligoG CF-5/20 it decreases to -23.1 mV, a difference that is statistically significant ($p<0.0001$) [181]. The increase in the net negative surface charge permanently changes the morphology of the mucin scaffold as the magnitude of the electrostatic (anionic) repulsion between mucin structures is increased, causing an increased separation between molecules, which increases the CF mucus pore size [181]. This lowers the viscosity of the CF mucus [181].

In addition, OligoG CF-5/20 has a high affinity for Ca^{2+} ions and can chelate Ca^{2+} ions through its carboxylate functionality [183]. As a consequence of favourable

carboxylate-calcium ionic interactions, diffusing OligoG CF-5/20 has the potential to out-compete mucin for the binding of Ca^{2+} ions [184], an ion that has the ability to electrostatically bind to, and cross-link, mucin to form large stable hexagonal gelled matrices [185]. In CF mouse ileum mucus model explant experiments, in which the predominant mucin is intestinal mucin MUC2N, OligoG CF-5/20 is therapeutically active at a range of different polymer length scales, from six guluronate units up to 20 guluronate units, and is able to out-compete MUC2N for the available extracellular calcium [184]. As such, OligoG CF-5/20 was able to clear intestinal mucus at concentrations as low 6.5 mmol/L [184]. In this particular study, the authors suggest that the encouraging therapeutic effects observed when MUC2N is exposed to OligoG CF-5/20 can be generalised to MUC5AC also [184], as MUC2N and MUC5AC share highly similar glycosylation patterns and thus physico-chemical properties and interactions [186], which highlights the potential role OligoG CF-5/20 could play in clearing stagnant mucus from the CF airway [184].

The therapeutic activity of OligoG CF-5/20 extends past mucociliary clearance, being the first alginate-based (guluronate-based) polymer therapeutic able to treat bacterial biofilm infections directly [187]. For example, when dispersed into *P. aeruginosa* biofilms, 10 % (v/v) OligoG CF-5/20 therapy was able to potentiate aztreonam, azithromycin and oxytetracycline, demonstrating 4-fold, 16-fold and 2-fold reductions respectively in the minimum inhibitory concentrations (MIC) of these antibiotic chemotherapies towards *P. aeruginosa* [187]. Furthermore, when combined with the antibiotic colistin, lower concentrations (1 % (v/v)) of OligoG CF-5/20 expressed sufficient potentiating activity to cause active disruption to *P. aeruginosa* biofilms grown on artificial CF sputum *in vitro*, dramatically increasing the dead : live cell ratio within these biofilms from 0.1 (in cultures exposed to no combined chemotherapy) to 0.5 [188]. Encouragingly, *P. aeruginosa* fails to develop resistance to OligoG CF-5/20 in the face of prolonged antibiotic-OligoG CF-5/20 combined therapy *in vitro* [187]. In particular, OligoG CF-5/20 has been observed to enforce no selection pressures, not triggering any form of resistance-based genomic mutation, in over 245 generations of *P. aeruginosa* cultures grown over a 45 day time period [189]. These are promising observations given the growing concern of antibiotic resistance.

There are two proposed mechanisms of OligoG CF-5/20 action against *P. aeruginosa* biofilms. The first mechanism of action relates to binding at the *P. aeruginosa* cell surface [187, 190]. OligoG CF-5/20 exposure is able to decrease the net negative charge of the *P. aeruginosa* cell surface from -30.9 mV to -47 mV within a pH range of 5.0 - 9.0 and an NaCl concentration range of 0.001 - 0.01 M, which is a difference that is statistically significant ($p<0.001$) [190]. Atomic force microscopy (AFM) has shown that this arises as a result of strong and irreversible binding between OligoG CF-5/20 and the *P. aeruginosa* LPS,

forming a stable complex which is resistant to hydrodynamic shear [190]. Recall, as discussed in **Section 1.1.2**, that the lipid A core of *P. aeruginosa* encompasses negatively charged phosphate groups at the terminal ends of the aminoarabinose backbone, giving the LPS scaffold a net negative charge and offers sites for calcium binding. As such, the mechanism of association between the LPS and OligoG CF-5/20 does not differ significantly from the mechanism against mucin, and the OligoG CF-5/20 molecule is capable of associating with the LPS through interaction with LPS-bound calcium [191]. Once OligoG CF-5/20 is tethered to the bacterial cell surface LPS, it causes a drastic reduction in bacterial motility and prevents biofilm proliferation [190]. It should be noted here, that OligoG CF-5/20 is not rupturing the cell surface, but rather undergoing a "non-permeabilizing" association with the cell surface LPS [187, 190] and it is hypothesised that this non-permeabilizing membrane-specific effect is the primary reason behind the failure of *P. aeruginosa* to develop resistance during prolonged exposure *in vitro* [187].

The second mechanism of action relates to biofilm structural weakening [183, 192]. OligoG CF-5/20 has been shown, through AFM and scanning electron microscopy (SEM) measurements, to visibly weaken mature *Pseudomonas* biofilms *in vitro* through disrupting the biofilm matrix architecture, giving rise to less densely packed biofilms that are more susceptible to hydrodynamic shear [192]. As a consequence, OligoG CF-5/20 exposed *P. aeruginosa* biofilm matrices are approximately 15-20 μm less thick and encase a significantly reduced cell density compared to non-OligoG CF-5/20 exposed controls [183]. The ability for OligoG CF-5/20 to induce these disruption effects stems from being able to readily partition and penetrate deep into mature *P. aeruginosa* biofilm matrices [183]. In mucoid *P. aeruginosa* biofilm murine lung infection models, 5 % (v/v) OligoG CF-5/20 caused biofilm matrix disruption over a 24 hour time period, resulting in a statistically significant ($p<0.001$) 2.5-log reduction in bacterial CFU [193]. In addition, this study further corroborates the observation that OligoG CF-5/20 is able to potentiate the antibiotic colistin against mucoid biofilms, as the minimum biofilm eradication concentration of colistin was lowered from 512 μg/ml to 4 μg/ml within 8 hours of OligoG CF-5/20 combined therapy [193].

Histopathology staining and confocal laser scanning microscopy (CLSM) clearly show how OligoG CF-5/20 can cause large scale disruptions to the biofilm matrix, creating clear cavities and void spaces within the scaffold, and it is this structural phenomenon that gives rise to the evident antibiotic synergy [183, 193]. But, it should also be noted here, that the observed matrix weakening effects under OligoG CF-5/20 exposure has also been attributed to OligoG CF-5/20 functioning as a QSAI antagonist [194]. In particular, OligoG CF-5/20 exposure causes a 4-fold and 2-fold reduction in the concentrations of C_4-HSL and 3-oxo-C_{12}-HSL respectively present within the mature *P. aeruginosa* biofilm matrix over a

12 hour time period [194]. This hinders the virulence activities involved in plasmid eDNA release into the biofilm matrix and, consequently, there is approximately a 2-fold reduction in the concentration of eDNA within the biofilm matrix, which significantly reduces the biofilms ability to proliferate [194].

Additionally, given that Ca^{2+} ion exposure is a large contributor to the increase in the Young's modulus, and thus mechanical stability at the macroscale, of mucoid *P. aeruginosa* biofilm matrices [122, 123, 132], it is also entirely possible that Ca^{2+} chelation chemistry is vital for the observed matrix weakening effects of OligoG CF-5/20. Recent MD simulations have implicated Ca^{2+} chelation, through the OligoG CF-5/20 carboxylate functionality, in preventing Ca^{2+} ion-induced ionic cross-linking of *P. aeruginosa* biofilm eDNA structures over a 50 ns time period [183]. In addition to sequestering free extracellular calcium in sub nanosecond time-scales, OligoG CF-5/20 is able to perturb the pre-established eDNA - Ca^{2+} - eDNA ionic cross-links to form more favourable eDNA - Ca^{2+} - OligoG CF-5/20 - Ca^{2+} - eDNA complexes, limiting direct Ca^{2+} ion mediated eDNA complexation [183].

Although the global matrix disruption effects of OligoG CF-5/20 have been observed (on bulk living biofilms) experimentally, the complexation of OligoG CF-5/20 at the atomistic and molecular level has only been tested through simulation against eDNA and MUC5AC structures [181, 183]. In fact, the disruption effect (mechanism) on mucoid EPS scaffolds has not been explored and, as such, is unknown at atomistic and molecular levels. This is true despite the EPS being the major component of mucoid *P. aeruginosa* CF biofilms [93, 103, 106, 107]. Knowledge of this mechanism will provide the necessary understanding required to potentially modify OligoG CF-5/20 to improve selectivity and efficacy when treating mucoid *P. aeruginosa* CF lung biofilm infections.

1.5.2 Novel gallium therapy

Intravenous (IV) gallium nitrate $Ga(NO_3)_3$ is emerging as another novel therapeutic strategy targeted at eradicating *P. aeruginosa* (biofilm) infections in the CF lung [180]. The structure of gallium nitrate, encompassing a single Ga^{3+} ion ionically coordinated to three bidentate NO_3^- ligands, is given in **Fig 1.9** below. Before discussing the therapeutic mode of action of gallium, it is first necessary to describe bacterial Fe^{3+} uptake mechanisms.

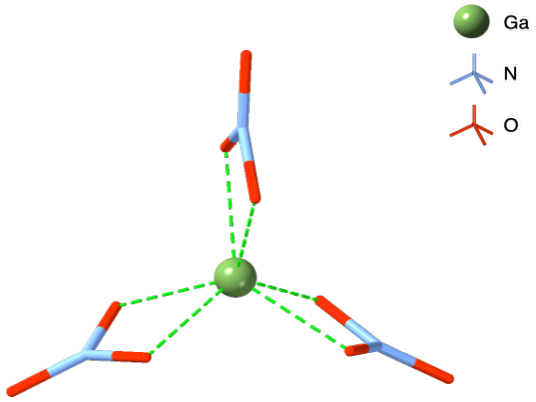


Figure 1.9: Molecular structure of gallium nitrate. Ionic bonds are shown with dashed green lines.

1.5.2.1 Ferric iron uptake in *P. aeruginosa*

CF airway cells expressing the ΔF508-CFTR mutation are present with increased extracellular iron concentrations [195]. These elevated extracellular iron concentrations are positively and linearly correlated with raised levels of the sputum inflammatory components TNF-α, (IL)-1β, (IL)-8 as well as sputum microalbumin protein, with Spearman's correlation r-values equal to 0.6, 0.7, 0.67 and 0.7 respectively [196, 197]. Increased sputum microalbumin protein concentrations suggest vascular leakage as part of the inflammatory process and bleeding in the CF lung [196, 197], contributing to the increased extracellular concentrations of iron from haem [197]. Moreover, the diffusion of lactoferrin, a protein that binds and transports iron, is significantly impeded by the dehydrated mucus present within the CF lung, reducing its ability to bind iron, and again, resulting in higher levels of extracellular iron [198]. Taken collectively, these factors contribute to iron levels in the CF airway exceeding normal serum levels, specifically 44.4 μM in the CF airway compared to 13 μM in the airway of non-CF controls [196].

There is a strong positive correlation (Spearman's correlation r-value equal to 0.67) between sputum iron levels and *P. aeruginosa* CFUs in the lungs of CF patients, which suggests that iron is implicated in survival and proliferation of *P. aeruginosa* infections [199]. Extracellular ferric iron functions as an extracellular signal, triggering bacterial aggregation followed by biofilm matrix formation and proliferation [200, 201]. Iron has been observed to be critical for both the establishment of mature *P. aeruginosa* biofilm matrices *in vitro* [201] and on cells expressing the ΔF508-CFTR mutation [195]. When extracellular iron levels are depleted, for example through iron chelation, *P. aeruginosa* biofilms cannot proliferate into mature, heterogeneous, three-dimensional structures *in vitro* and present with a significantly

reduced biofilm biomass [202]. This is because, in iron-deplete extracellular medium, the twitching motility is up-regulated [203]. Recall in **Section 1.2.2**, type IV pili initiate a twitching motility that facilitates colony aggregation and biofilm expansion. However, up-regulation of this twitching motility means it is not possible for *P. aeruginosa* cells to settle onto the substratum, aggregate, and form a biofilm matrix [195, 201, 203]. These observations further emphasize the role of ferric iron in *P. aeruginosa* growth, pathogenicity, biofilm formation and, consequently, virulence in the CF lung.

To perceive the extracellular ferric iron, low-molecular weight secondary metabolites encompassing catecholate, hydroxamate, or hydroxycarboxylate functionality are synthesised by *P. aeruginosa* and transported out of the cell into the extracellular environment to chelate iron with high affinity and return it to the bacterium [204]. These secondary metabolites are called *siderophores*. *P. aeruginosa* cells cannot obtain sufficient levels of iron to facilitate biofilm proliferation and maintenance by passive diffusion alone and siderophores, therefore, play a critical role in the establishment of mature, structurally heterogeneous, three-dimensional, biofilm matrices [201]. Pyoverdine and pyochelin are the two primary siderophores utilised by *P. aeruginosa*. Pyoverdine is deployed in severely iron limited environments and pyochelin is deployed when iron levels are not as scarce [201]. Pyoverdine has a higher binding affinity [201], arising from hydroxamate, carbonyl and catecholate functional groups coordinating the Fe^{3+} centre in a stable octahedral geometry [205] and this higher binding affinity rationalises its use when iron is limited. Given that iron is elevated, not limited, in CF sputum, pyochelin is deployed by *P. aeruginosa* upon colonisation of the CF lung. This has been corroborated through transcriptional analysis of pyochelin synthesis genes, which show high expression during the early stage of *P. aeruginosa* growth and biofilm formation in CF sputum [206].

The molecular structure of pyochelin in its iron binding conformation is given in **Fig 1.10**. Pyochelin encompasses phenolate, amine and carboxylate binding sites (marked with asterisks in **Fig 1.10**) which are involved in the coordination of a single ferric centre [207] and electrospray ionisation mass spectrometry has shown the existence of both the monochelate and bischelate ferric complexes ($[FeL]^+$ and $[FeL_2]^-$ respectively where L = pyochelin) at physiological pH [208]. In extracellular medium, however, the monochelate complex dominates over the bischelate complex [207] with the remaining bidentate coordination saturated by hydroxide and/or water molecules, giving an octahedral complex geometry [207].

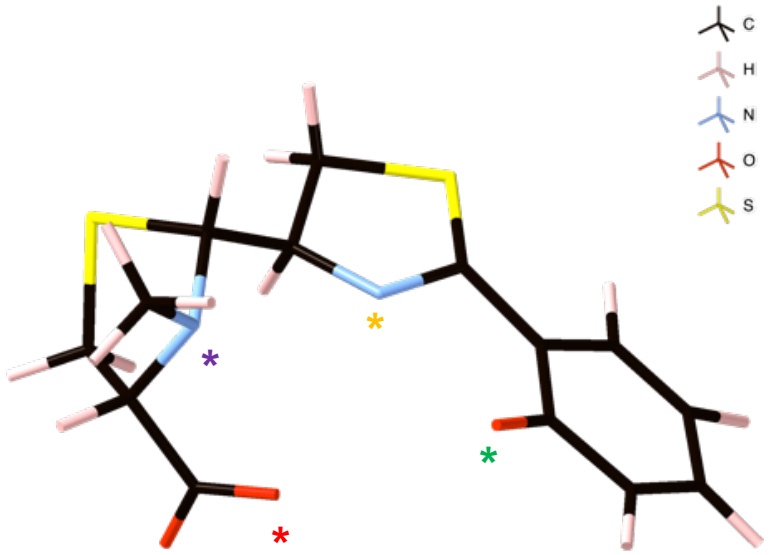


Figure 1.10: Molecular structure of pyochelin in its iron binding conformation isolated from its membrane receptor, FptA (PDB entry 1XKW [209]). The carboxylate group is indicated by the red asterisk, the phenolate group by the green asterisk, the sp^2 amine site by the orange asterisk and the sp^3 amine site by the purple asterisk.

Once pyochelin has complexed ferric iron, it is translocated into the bacterial cell through binding the TonB-dependent membrane receptor protein (TBDR), FptA. [210]. FptA is a 682 amino acid 75,993 g/mol protein that functions as the TBDR membrane receptor protein responsible for pyochelin-iron complex translocation [210]. This protein, as with other membrane receptor proteins of the TBDR family [211], encompasses a transmembrane 22-stranded β-barrel domain which encircles a globular plug domain [209]. The pyochelin binding site spans these two domains, comprising the hydrophobic side-chain amino acids phenylalanine (Phe 114) and leucine (Leu 116 and 177) within the plug domain and methionine (Met 271), tyrosine (Tyr 334), tryptophan (Trp 702) and glutamine (Gln 395) within the β-2, 3, 7 and 9 strands of the β-barrel domain [209]. This is visualised in **Fig 1.11**. In particular, fluorescence resonance energy transfer (FRET) investigations have identified that the tryptophan residue (Trp 702) stabilises the pyochelin-iron complex within the FptA binding site and, perhaps, is most critically implicated in molecular recognition [212].

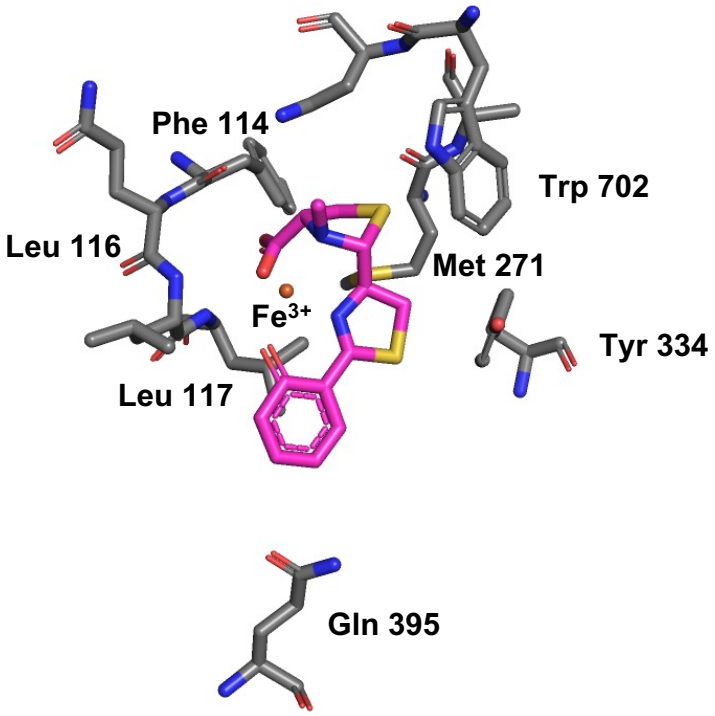


Figure 1.11: Molecular structure of pyochelin in its iron binding conformation bound within the ligand binding site of the membrane receptor, FptA (PDB entry 1XKW [209]). For clarity, the carbon backbone of the FptA ligand binding site is shown grey, whereas the carbon backbone of the pyochelin ligand is shown in pink. Oxygen atoms are shown in red, nitrogen in blue, sulphur in yellow and iron in brown. Hydrogen atoms are not shown.

Given that FptA is located on the cell surface of *P. aeruginosa*, it does not possess an energy source [213]. As such, upon the binding of the pyochelin-iron complex to the pyochelin binding site within FptA, the proton motive force of the cytoplasmic membrane, supplied by the inner membrane complex TonB-ExbB-ExbD [214], provides the necessary energy to facilitate the active transport of the pyochelin-iron complex into the cell [213, 215]. In particular, it is the conformational change stimulated when the pyochelin-iron complex binds at Phe 114, Leu 116 and Leu 117 that signals to TonB-ExbB-ExbD to energize FptA for the uptake of the pyochelin-iron complex [209]. It is interesting to note here that the hydroxide and/or water ligands that saturate the octahedral coordination geometry of the ferric centre [207], are not required for molecular recognition and cellular uptake by FptA [209]. In essence, the conformational change within the binding pocket is induced by the monochelate pyochelin alone.

1.5.2.2 Gallium as an iron mimic

Siderophores are critical for iron uptake and are, therefore, essential for *P. aeruginosa* growth and biofilm proliferation in the CF airway. Indeed, a recent systematic review, conducted by

Firoz *et al.*, evaluating 30 studies reporting data on novel iron-targeting therapies, concluded that therapeutic strategies that interfere with iron uptake have anti-*Pseudomonal* biofilm properties [216]. This is where IV gallium nitrate therapy has been having success in recent years.

Gallium nitrate is a source of Ga^{3+} ions that behave as Fe^{3+} mimics. Octahedral Ga^{3+} possesses an ionic radius of 0.620 Å, that is similar to the ionic radius of high-spin octahedral Fe^{3+} which measures 0.645 Å [217]. Consequently, Ga^{3+} is mistaken for Fe^{3+} by siderophores, which not only quenches the siderophores from the extracellular medium, but also results in the cellular up-take of gallium instead of iron. Both of these culminate in reduced cellular iron up-take and a significant disruption to iron homeostasis and metabolism [180]. Studies exploring siderophore activity in relation to gallium exposure *in vitro* observe the strongest negative correlation between *P. aeruginosa* growth and exogenous gallium concentration in *P. aeruginosa* cultures that can produce siderophores [218]. This is an observation that further underlines the association between gallium and siderophore quenching.

Recently, intravenous gallium nitrate therapy has been shown to function as a novel therapeutic strategy that is both safe and efficacious, inhibiting *P. aeruginosa* growth in the lungs of CF patients [219]. It has been shown that gallium concentrations as low as 1-5 μM can completely inhibit *P. aeruginosa* growth on CF sputum [219, 220] and on human serum extracted from CF patients [221]. In addition to slowing bacterial growth, gallium is effective against *P. aeruginosa* cells embedded within the biofilm matrix [220, 222], in addition to cells simply occupying the matrix surface layers, and is even able to completely obstruct *P. aeruginosa* biofilm proliferation *in vitro* [220]. Encouragingly, in a pilot phase 1b non-randomized study (ClinicalTrials.gov NCT01093521) in which intravenous gallium nitrate therapy was administered to CF patients suffering with chronic *P. aeruginosa* lung infections, there was a notable regain in lung function over 14 - 28 days with no adverse side effects [219]. In fact, the regain in lung function is comparable to that observed when using vigorous conventional antibiotic therapy, and only a very small number of cells developed gallium resistance genomic mutations [219]. Given that only a small number of cells develop resistance, this has little effect on the overall resistance profile of the gallium therapy [219].

The siderophore-quenching coordination chemistry has been explored using DFT based Gibbs free energy calculations [223, 224]. Specifically, the Gibbs free energy change when Ga^{3+} is substituted into the siderophore ferric ligating groups, phenolate, carboxylate, catecholate and hydroxamate, was calculated with the aim of identifying the molecular functionality present within pyoverdine and pyochelin that contributes most significantly to the iron-for-gallium exchange [224]. It was shown in this study that catecholate, hydroxamate

and phenolate functionality is slightly more selective to binding Ga^{3+} ions in preference to Fe^{3+} ions, giving rise to a negative (spontaneous) Gibbs free energy for the iron-for-gallium exchange [224]. An analogous DFT study investigating the substitution of iron for gallium into model binding sites of the iron metalloproteins ribonucleotide reductase and transferrin, elucidated that when there are at least three anionic COO^- bearing acidic binding site amino acid residues, such as aspartate and glutamate, the binding site becomes more selective for gallium over iron [223]. In addition, these binding sites become more selective for gallium when exposed to a high dielectric solvent medium, suggesting that polar solvent accessibility further improves the iron-for-gallium exchange spontaneity [223].

Although the siderophore quenching mechanism of gallium therapy has been well demonstrated experimentally, it has also been observed that *P. aeruginosa* mutants deficient in pyochelin and pyoverdine continue to display reduced biofilm proliferation *in vitro* [220]. This suggests that the therapeutic activity of gallium does not solely rest on the competition for siderophore binding. In fact, in mucoid *P. aeruginosa* CF clinical isolates, inactivation of the pyochelin siderophore uptake system did not have any effect on the resistance profile of exogenous gallium therapy [225]. In addition, there is a 35-fold reduction in the amount of gallium that can be translocated across the *P. aeruginosa* cell envelope through FptA-mediated uptake relative to iron [226]. The interpretation being that FptA-mediated cellular gallium uptake is far less efficient compared to the native iron uptake [226]. This entertains the possibility that, to achieve its anti-biofilm effects at μM doses, gallium may also be exploiting another entry mechanism that circumvents this particular low efficiency FptA pathway. Intriguingly, it is also unknown whether exogenous gallium has any direct influence on the biofilm matrix structural integrity, which would contribute to the observed reductions in biofilm biomass and proliferation upon exposure to gallium.

This raises the question of whether gallium has a direct effect on the EPS architecture, and whether it is weakening the biofilm structure. This is yet to be tested *in vitro*, yet, encouragingly, Ga^{3+} ions have been observed to bind in a thermodynamically stable fashion to algal alginate scaffolds in theoretical studies based on DFT [142] as well as in energy dispersive spectroscopy (EDS) mapping of gallium-algal alginate hydrogels [227]. Whether gallium could be accommodated by the mature mucoid *P. aeruginosa* EPS scaffold, displacing the calcium cross-links and weakening the overall biofilm structure, is yet to be tested. It has already been discussed in **Section 1.3.2** that calcium ions adopt permanent ionic cross-linking sites between EPS scaffolds, giving rise to mechanically stable network structures possessing high storage and Young's moduli [122, 132, 133]. Recently, it has been demonstrated that mucoid *P. aeruginosa* biofilm proliferation is, in fact, calcium-dependent, meaning that *P. aeruginosa* biofilms only require EPS alginate and calcium ions to establish

structurally heterogeneous three-dimensional biofilm matrices [102]. In essence, it is only the calcium-EPS inter-molecular associations that govern biofilm matrix scaffold morphology [102].

As already discussed, mucoid *P. aeruginosa* biofilm matrices have a high Young's modulus [122, 132, 133], and in fact stiffen and become more cohesive over time, transitioning from viscoelastic liquids from their inception to a viscoelastic gel as they mature [228]. These gelled matrices have a high yield stress, making them resistant to deformation and remodelling over time [228]. In contrast, gallium-alginate scaffolds cannot dissipate energy under deformation and, consequently, are hard and brittle [227]. In addition, gallium-alginate scaffolds are prone to degradation through gallium-for-sodium cation exchange and display fast degradation kinetics *in vitro* [222, 227]. Therefore, if gallium ions could be accommodated by the EPS, in preference to the native calcium ions, then potentially this would create a scaffold morphology more susceptible to physiological softening through sodium-ion exchange, thus reducing matrix proliferation.

PQS, the QSAI responsible for biofilm maturation and maintenance (see **Section 1.4**, **Fig 1.7**), also possesses a role as an Fe^{3+} chelator [172–174]. Exposing *P. aeruginosa* cells to exogenous PQS concentrations exceeding 10 μM leads to a statistically significant increase in the levels of the siderophores pyochelin ($p<0.001$) and pyoverdine ($p<0.001$) within the extracellular medium [172]. This suggests that PQS is able to chelate iron, reducing the concentration of free iron in the extracellular medium and inducing an iron-starvation response. In fact, this has been corroborated through iron-chelation assays, which observe a 0.29 mg/L reduction in the free (available) iron content within the extracellular medium of *P. aeruginosa* cultures upon exposure to 40 μM concentrations of exogenous PQS [173].

Structurally, at physiological pH, PQS is able to complex about a single Fe^{3+} centre in a 2:1 or 3:1 stoichiometry [173, 174]. PQS was initially considered to be unable to deliver iron into the bacterial cell, instead localising iron at the bacterial membrane interface. This observation suggested that PQS does not act as a siderophore, but rather an iron-trap, aiding in siderophore-mediated iron delivery [174]. In light of this observation however, metal-uptake experiments have shown that gallium uptake in pyoverdine and pyochelin-deficient mutants is approximately 10 pM per 10^8 cells, compared to 20 pM per 10^8 cells in siderophore-active control cultures, at gallium nitrate concentrations of 2 μM [220]. It is therefore possible that gallium is targeting another uptake system and circumventing siderophore uptake. Recently, it has been shown that siderophore-deficient *P. aeruginosa* cells can uptake Fe^{3+} utilising PQS as a delivery vessel [229], extending the role of PQS beyond storage. The H3-T6SS (type VI secretion system) produces a substrate, TseF, that can bind PQS-Fe^{3+} adducts and facilitate cellular iron uptake through binding the TBDR membrane siderophore

uptake protein, FptA [229]. Interestingly, when PQS-Fe^{3+} complexes are the only source of extracellular iron, *P. aeruginosa* mutants that do not possess the ability to produce TseF fail to uptake iron, demonstrating that the TseF substrate is critical for mediating iron acquisition from PQS [229]. Gallium substitution into the ferric-PQS uptake systems has not previously been explored theoretically or experimentally and could offer another reason for the reduced biofilm matrix proliferation in siderophore-deficient *P. aeruginosa* cells.

1.6 Research aims

Biofilms present a highly prevalent medical challenge and this research will focus on the biofilm EPS matrix produced by mucoid *P. aeruginosa* in the CF lung, with the aim of building a more comprehensive understanding of how the structure of the EPS relates to pathogenesis. In turn, this will provide data and guidance suitable for designing novel treatment strategies for mucoid *P. aeruginosa* biofilm infections in the CF lung.

Within computational chemistry, one seeks to construct atomistic (computer) models of molecules to accurately compute their energies, inter-atomic bonds and their stable conformations. Accurate prediction of molecular energies enables one to understand which molecular structures are thermodynamically stable and, therefore, likely to occur experimentally *in vitro* and *in vivo*. Coupled with an accurate quantification of the inter-atomic interactions, one can accurately implicate specific molecular functional groups in certain experimental (*in vitro* and *in vivo*) observations and biophysical processes. For example, this includes gaining an atomic-scale understanding of which functional groups are responsible for stabilising EPS matrix architecture, retarding pharmacological challenge and facilitating molecular movement throughout the biofilm. Therefore, the aim of this research was to use computational chemistry to provide connections, at the atomic-scale, between chemical structure, pathogenicity and bacterial virulence.

P. aeruginosa was chosen as the target pathogen because of its clinical prevalence in CF sufferers, its position at the epicentre of ongoing biofilm research, and because of its experimentally well-characterised biofilm matrix architecture. Mucoid *P. aeruginosa* produces a biofilm matrix, in the lungs of CF patients, that is composed primarily of one molecular constituent: acetylated, anionic, exopolysaccharide (EPS) alginate. Therefore, the atomistic model of the mucoid *P. aeruginosa* biofilm matrix was simplified through neglecting larger biopolymeric matrix constituents that are more scarcely observed, encompassing the EPS as the primary model component, without the loss of representability.

To that end, an atomistic model of the mucoid *P. aeruginosa* EPS was created. This EPS atomistic model encompassed, in a thermodynamically stable fashion, key structural

motifs unique to the mucoid *P. aeruginosa* biofilm matrix present within the CF lung. The model was created ensuring that it could validate spectroscopic/characterisation data and known cation thermochemistry, definitively elucidating which sputum ions, of those most elevated in the CF lung, contribute most significantly to EPS matrix stability. The model was then used to identify which molecular interactions (and molecular functional groups) are responsible for facilitating molecular movement throughout the biofilm EPS matrix. C_4-HSL and PQS are two QSAI molecules important for establishing mature biofilms and therefore contributing to biofilm chronicity. These molecules move throughout the EPS and were ideal candidates to study possible modes of interaction with the EPS matrix. Not only did this provide a physico-chemical rational for the movement of cell-to-cell signalling molecules throughout the biofilm, but also highlighted structural chemistry of extreme importance for the design of matrix penetrating (bio-)molecules.

Finally, the EPS atomistic model was deployed to study the therapeutic mechanisms of the two novel anti-*Pseudomonal* agents, OligoG CF-5/20 and intravenous gallium therapy. With regards to the former, these models captured the atomistic mechanisms behind the exothermic dispersal of OligoG CF-5/20 into the EPS matrix, elucidating the therapeutic mechanism against mucoid *P. aeruginosa* CF biofilms. With regards to the latter, these models studied the effect of novel intravenous gallium therapy on the EPS structure, investigating whether it could be accommodated by the EPS, in preference to CF sputum ions, or by the PQS bacterial iron uptake system. This offered insight into the therapeutic action of this therapy and, potentially, extends what is currently known about its therapeutic mechanism, shedding light onto new, previously unknown, gallium targets.

The use of quantum mechanical simulation in this research is necessary to quantitatively assess which molecular interactions contribute most significantly to pathogenicity and virulence, but, places a limit on model length scale. Therefore, this research embodies chemical prediction at the atomic (Å) and molecular (nm) scales, focusing on molecular interaction fine-structure, explicitly, studying what is a very small (localised) chemical niche within the mucoid *P. aeruginosa* biofilm EPS matrix, visualised in **Figure 1.12**, and extrapolating any observations to rationalise virulence at a bulk scale.

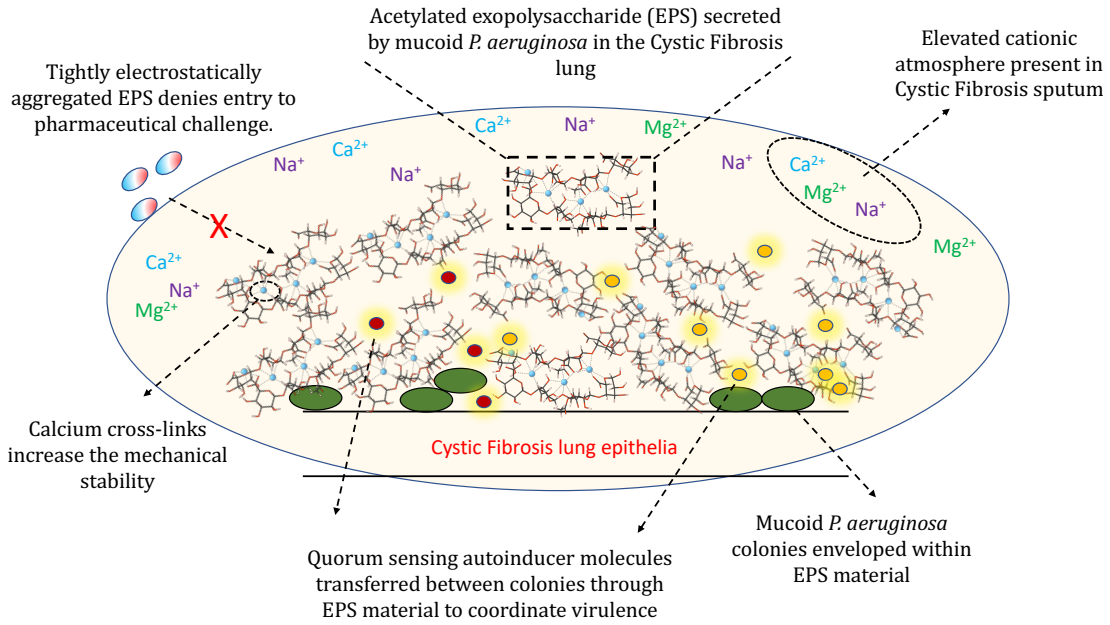


Figure 1.12: Molecular scale visualisation of a mucoid *P. aeruginosa* biofilm EPS matrix.

Theory

2.1 Introduction to computational chemistry

The principle aim of computational chemistry is to use quantum mechanical and forcefield simulation to explore the conformational energetics of molecules to establish quantitative and qualitative relationships between molecular structure, bonding, function and reactivity. This, in turn, helps elucidate the mechanisms governing a chemical or biophysical process at a scale not possible through experimental measurement. Simulations of small molecules can be achieved using quantum mechanical approaches, solving the time-independent Schrödinger equation to yield the ground state energy and wavefunction of the molecular system. If the total molecular energy can be calculated, then any chemical observable related to the molecular energy (for example, spectroscopic properties), or related to the difference in the relative molecular energy (for example, thermodynamic stabilities), can also be determined. In the case of large molecules, for example larger bio-macromolecular complexes, energy computation using quantum mechanics is not computationally feasible and, as such, forcefields are used. Forcefields are classical energetic expressions well-parameterised to reproduce the molecular energies, although not to the same accuracy as quantum mechanical computation, but to a sufficient enough accuracy to identify chemical trends. In this research, both quantum mechanical and forcefield simulations strategies were deployed.

2.2 The system energy

2.2.1 Separation of space and time variables

The molecular energy is the central output of computational chemistry and is the solution to the many-body Schrödinger equation. The time-dependent Schrödinger equation is given by

$$\hat{H}(\mathbf{r}_i, t)\Phi(\mathbf{r}_i, t) = i\frac{\partial\Phi(\mathbf{r}_i, t)}{\partial t} \tag{2.2.1}$$

where $\mathbf{r}_i$ and t are particle positions in 3D space and time respectively, and $\hat{H}(\mathbf{r}_i, \mathbf{R}_I, \mathrm{t})$ is the Hamiltonian operator which represents the sum of the kinetic (T) and potential (V) energies

of the particles in the system denoted by positions $\mathbf{r}$,

$$\hat{H}(\mathbf{r}_i, t) = T + V(\mathbf{r}_i, t) \tag{2.2.2}$$

and $\Phi(\mathbf{r}_i, t)$ is the many-body ground-state wavefunction which is dependent upon both the spatial and temporal coordinates of the particles. Molecules represent *bound* systems, meaning the nuclei and the electronic orbits about these nuclei remain localised in space [230]. Therefore, the potential energy in the Hamiltonian is independent of time

$$V(\mathbf{r}_i, t) = V(\mathbf{r}_i) \tag{2.2.3}$$

and, as such, the Hamiltonian becomes time-independent and yields the total energy (E) when acting on the wavefunction

$$\hat{H}(\mathbf{r}_i, t) = \hat{H}(\mathbf{r}_i) = T + V(\mathbf{r}_i) \tag{2.2.4}$$

$$\hat{H}(\mathbf{r}_i, t)\Phi(\mathbf{r}_i, t) = E\Phi(\mathbf{r}_i, t)\ . \tag{2.2.5}$$

By inserting this into the time-dependent Schrödinger equation

$$\hat{H}(\mathbf{r}_i, t)\Phi(\mathbf{r}_i, t) = E\Phi(\mathbf{r}_i, t) = i\frac{\partial\Phi(\mathbf{r}_i, t)}{\partial t} \tag{2.2.6}$$

and through solving the first-order differential with respect to time yields the following solution for the ground-state many-body wavefunction,

$$\Phi(\mathbf{r}_i, t) = \Phi(\mathbf{r}_i)e^{-iEt} \tag{2.2.7}$$

which shows that the time dependence can be written as a phase factor multiplied with the spatial wave function [230]. As already discussed, molecules represent bound states, their potentials are independent of time and the phase factor (e^{-iEt}) can be neglected. This returns the time-independent Schrödinger equation

$$\hat{H}(\mathbf{r}_i)\Phi(\mathbf{r}_i) = E\Phi(\mathbf{r}_i) \tag{2.2.8}$$

which is the starting point for the computation of molecular energies.

2.2.2 The molecular Hamiltonian and the variational principle

The Hamiltonian operator ($\hat{H}$) for molecular systems takes the form

$$\hat{H} = -\sum_i \frac{\hbar^2}{2m_i}\nabla_i^2 - \sum_k \frac{\hbar^2}{2m_k}\nabla_k^2 + \sum_{i<j} \frac{e^2}{4\pi\epsilon_0 \mathbf{r}_{ij}} + \sum_k \sum_l \frac{Z_k Z_l e^2}{4\pi\epsilon_0 \mathbf{r}_{kl}} - \sum_i \sum_k \frac{Z_k e^2}{4\pi\epsilon_0 \mathbf{r}_{ik}} \tag{2.2.9}$$

where i and j are labels denoting the electrons, k and l are labels denoting the nuclei, m denotes particle mass, ∇^2 is the Laplacian operator, Z is the atomic number and $\mathbf{r}_{ab}$ is the distance between particles a and b. The reaming terms include $\hbar$, which is Plank's constant divided

by 2π, e, which is the charge of the electron, and ϵ_0, which is the permittivity of free space. These latter terms are expressed in atomic units, $\hbar = e = 4\pi\epsilon_0 \equiv 1$, meaning the Hamiltonian is simplified to the following form

$$\hat{H} = -\sum_i \frac{1}{2}\nabla_i^2 - \sum_k \frac{1}{2m_k}\nabla_k^2 + \sum_{i<j} \frac{1}{\mathbf{r}_{ij}} + \sum_k \sum_l \frac{Z_k Z_l}{\mathbf{r}_{kl}} - \sum_i \sum_k \frac{Z_k}{\mathbf{r}_{ik}} \tag{2.2.10}$$

The first two terms represent the kinetic energy operators for the electrons and nuclei respectively and the remaining parts of the expression represent the Coulomb attraction and repulsion operators that make up the remainder of the potential energy. When the molecular Hamiltonian acts on the molecular wavefunction $\Phi(\mathbf{r}_i,\mathbf{R}_i)$, which explicitly encompasses both the electronic ($\mathbf{r}$) and nuclear ($\mathbf{R}$) positions, it yields the energy (E) of the molecular system [231]. As discussed in **Section 2.2.1**, this is an operation described by the time-independent Schrödinger equation (**Equation 2.2.8**) and is central to modern quantum mechanics, offering a theoretical framework for which the energy of any system of electrons and nuclei can be evaluated, given a suitable expression of the wavefunction $\Phi(\mathbf{r}_i, \mathbf{R}_i)$.

Given the exact (true) wavefunction for the molecular system (Φ_0), the exact ground-state energy expectation value of the Hamiltonian (E_0) is given as [232]

$$E_0 = \frac{\langle \Phi_0 | \hat{H} | \Phi_0 \rangle}{\langle \Phi_0 | \Phi_0 \rangle} \tag{2.2.11}$$

with $\langle \Phi_0 | \Phi_0 \rangle = 1$ for normalised wavefunctions. Given an approximate wavefunction (ϕ), which corresponds to any trial normalised representation of $\Phi(\mathbf{r}_i, \mathbf{R}_i)$, the energy expectation value of the Hamiltonian is given in an analogous fashion

$$E_\phi = \frac{\langle \phi | \hat{H} | \phi \rangle}{\langle \phi | \phi \rangle} \, . \tag{2.2.12}$$

The variational principle states that the energy expectation value of the Hamiltonian operator acting on an approximate wavefunction will always be higher than when acting on the true wavefunction

$$E_\phi \geq E_0 \tag{2.2.13}$$

for all approximate wavefunctions ϕ [231, 232]. Given an arbitrary representation λ of the trial wavefunction $\phi(\lambda)$, the energy $E_\phi(\lambda)$ will always exceed E_0 and, as such, this representation (λ) can be altered until $E_\phi(\lambda)$ reaches a minimum [232]

$$\frac{\partial E_\phi(\lambda)}{\partial \lambda} = 0 \, . \tag{2.2.14}$$

This energy ($E_\phi(\lambda)$) corresponds to the variational estimate of the exact ground-state energy and the approximate wavefunction ($\phi(\lambda)$) that gives rise to this energy is the variational estimate of the exact ground-state wavefunction.

2.3 The Born-Oppenheimer approximation

A fundamental approximation in quantum mechanics is the Born-Oppenheimer approximation which states, due to the large difference in mass between the electrons and nuclei, electronic relaxation is instantaneous relative to nuclear motion [231]. The electrons within a molecule will adopt a stable ("relaxed") configuration for any given stationary nuclear configuration. Nuclear kinetic energy ($\sum_k \frac{1}{2m_k}\nabla_k^2$) can therefore be neglected and the repulsive nuclear–nuclear potential energy term ($\sum_k \sum_l \frac{Z_k Z_l}{\mathbf{r}_{kl}}$) becomes an evaluated constant for a given nuclear configuration [232]. Now, under the Born-Oppenheimer approximation, the electronic Hamiltonian ($\hat{H}_e$)

$$\hat{H}_e = -\sum_i \frac{1}{2}\nabla_i^2 + \sum_{i<j} \frac{1}{\mathbf{r}_{ij}} - \sum_i \sum_k \frac{Z_k}{\mathbf{r}_{ik}} \tag{2.3.1}$$

acts on the electronic wavefunction (Ψ_e), a complete description of the molecular electronic structure at 0 K,

$$\Psi_e = \Psi_e(\mathbf{r}_i; \mathbf{R}_I) \tag{2.3.2}$$

yielding the electronic energy (E_e), an operation described by the time independent *electronic* Schrödinger equation

$$\hat{H}_e \Psi_e = E_e \Psi_e \ . \tag{2.3.3}$$

The electronic energy (E_e) is combined with the constant nuclear repulsion to obtain the total energy (E_{tot}) of the static (fixed nuclei) molecular system

$$E_{tot} = E_e + \sum_k \sum_l \frac{Z_k Z_l}{\mathbf{r}_{kl}} \ . \tag{2.3.4}$$

Notionally, the Born-Oppenheimer approximation has reduced the problem of expressing a wavefunction explicitly dependent on nuclear and electronic coordinates to expressing a wavefunction explicitly dependent on electronic coordinates alone. In turn, solving **Equation 2.2.8** is reduced to solving **Equation 2.3.3**. This is referred to as the *electronic problem*, obtaining the energy of a molecule from the molecular electronic structure, and forms the foundation of modern electronic structure theory.

2.4 The potential energy surface

As electronic and nuclear motions are separated, the electronic wavefunction (Ψ_e) and total energy (E_{tot}) depend explicitly on the electronic coordinates but parametrically on the nuclear coordinates. As such, the energy for each nuclear configuration $\mathbf{R}_I$ defines a Born-Oppenheimer potential energy surface (PES), which is a potential invoked by the electrons over which the nuclei can evolve.

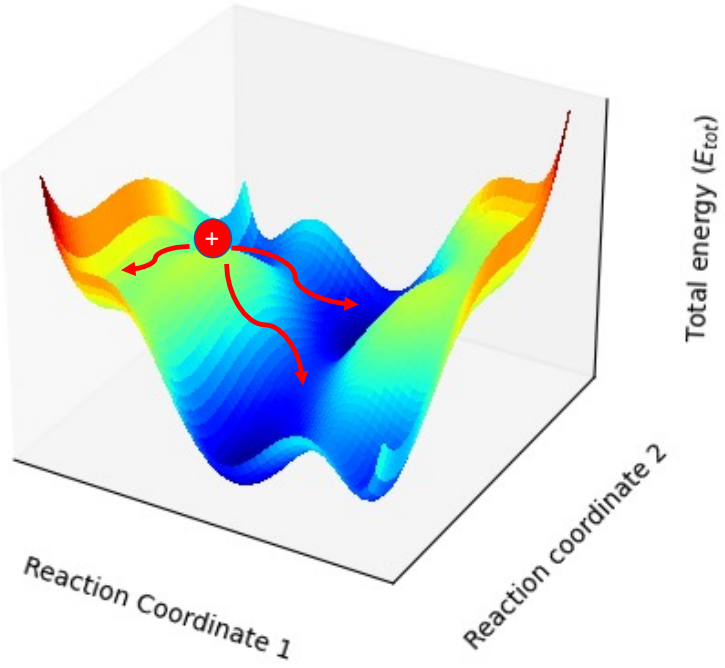


Figure 2.1: Schematic of a three-dimensional Born-Oppenheimer PES that provides a potential over which a nucleus, marked on the figure as a red ball with positive charge, can evolve.

Under the Born-Oppenheimer approximation (**Section 2.3**) the nuclei are, essentially, considered as heavy classical particles - their quantum effects are considered as negligible. If atomic nuclei displayed non-negligible quantum effects, the concept of a molecular structure, a unique nuclear configuration with a unique energy, would not have any meaning as the atomic nuclei would simply tunnel through energy barriers towards the global minimum [230]. Therefore, molecular geometries are distinguishable based on their relative energies, meaning unique molecular structures will have unique properties, functions and reactivity.

For molecular systems the PES is $3N-6$ dimensional, where N is the number of atomic nuclei. Conceptually, if the distance between the particles is fixed, then translating or rotating the full system will not alter the system energy and, as such, the 3 translational and 3 rotational degrees of freedom are removed giving rise to the "-6" term. Understandably, mapping the entire PES for molecules possessing more than three or four atoms is impossible. For example, calculating the energy of a molecular system encompassing four atoms every 0.1 Å over a 1 Å range gives a PES with 10^6 points [230]. As a consequence, it is often the preference of the simulator to restrict the search space to regions of the PES confined to local minima (the wells in the PES), as these regions encompass the most stable molecular geometries which are most likely to be observed experimentally.

In a practical sense, mapping the minimum energy regions of the PES, that is, identifying the

most stable structures, allows the relationships between molecular structure and molecular function to be established. Through predicting the stable molecular geometry adopted by a given system, one can assess the most stable relative orientation of key functional groups and observe site-directed interactions. This allows one to critically implicate specific molecular functionality in key chemical and biophysical processes. In effect, exploration of PES minima will elucidate the mechanism governing a chemical or biophysical process at atomic-scale resolution. Methods to explore the PES minima and identify these stable structures are introduced and discussed in more detail in **Section 2.12.1**.

2.5 Density Functional Theory

The electronic wavefunction (Ψ_e) is 4N dimensional, dictated by three spatial coordinates and one spin coordinate, meaning that expressing the electronic wavefunction soon becomes intractable as the molecules become larger and the number of electrons increases. Density Functional Theory (DFT) is an alternative approach, based on the premise that the energy of a molecular system can be obtained from the electron density [233]. The advantage of this is that the electron density is three-dimensional and is thus far easier to express compared to the 4N-dimensional electronic wavefunction.

The Hohenberg-Kohn theorem [234] provides a theoretical basis for the justification of using the electron density as the basic variable from which the energy of a (static) molecule can be computed. The theorem states that the ground-state energy of a molecular system is a functional of the ground-state electron density n($\mathbf{r}$)

$$E[n(\mathbf{r})] = \int d^3\mathbf{r}\; V_{ext}(\mathbf{r})\; n(\mathbf{r}) \; + F[n(\mathbf{r})] \tag{2.5.1}$$

where $V_{ext}(\mathbf{r})$ is the external potential and F[n($\mathbf{r}$)] is a *unique energy functional*. Definition of n($\mathbf{r}$) and F[n($\mathbf{r}$)] allows for the energy of any molecular system to be evaluated through **Equation 2.5.1**. If the unique energy functional is known *exactly* then the energy of the molecular system is known *exactly* and the electron density that minimizes the functional is the *exact* ground-state electron density. However, the nature of the unique functional is unknown.

The Kohn-Sham formalism [235] provides a strategy to define both n($\mathbf{r}$) and F[n($\mathbf{r}$)] by introducing the concept of a *non-interacting* system of electrons which preserves the electron density of the true interacting system. These independent, non-interacting, particle states are the *Kohn-Sham orbitals* $\phi_i(\mathbf{r})$ from which the electron density can be defined

$$n(\mathbf{r}) = \sum_{i}^{N} |\phi_i(\mathbf{r})|^2 \; . \tag{2.5.2}$$

The ground-state electron density can then be *inserted* into the energy functional to evaluate the ground-state energy. The unique functional from **Equation 2.5.1** for this non-interacting reference system is given as

$$F[n(\mathbf{r})] = T[n(\mathbf{r})] + E_H[n(\mathbf{r})] + E_{XC}[n(\mathbf{r})] \; . \tag{2.5.3}$$

The first term in this functional is the kinetic energy term

$$T[n(\mathbf{r})] = -\frac{1}{2}\sum_{i}^{N} \langle \phi_i | \nabla^2 | \phi_i \rangle \tag{2.5.4}$$

which is expressed in terms of the non-interacting Kohn-Sham orbitals and is evaluated exactly. The second term is the Hartree energy term, which is the classical electrostatic interaction of a charge density with the system's total electron density [236]

$$E_H[n(\mathbf{r})] = \frac{1}{2}\int\int \; d^3\mathbf{r} \; d^3\mathbf{r}' \; \frac{n(\mathbf{r})n(\mathbf{r'})}{|\mathbf{r} - \mathbf{r'}|} \tag{2.5.5}$$

although, this term does introduce an unphysical self-interaction error. The final term is the exchange-correlation term $\mathrm{E}_{XC}[\mathrm{n}(\mathbf{r})]$, that exists to bridge the gap between the universal functional and the functional employed in the Kohn-Sham formalism. Specifically, it is accounting for the differences in kinetic energy between the real system and the non-interacting system, the cancellation of the self-interaction energy, the exchange energy and the correlation energy. The exchange and correlation energies are the non-classical electrostatic contributions to the total energy and arise from the antisymmetric and Coulombic nature of electrons [233]. This is discussed in more detail in **Section 2.6.1**.

The Kohn-Sham formalism is an independent particle theory [237]. Nevertheless, it is a formalism that introduces a means by which the electron density, in principle, can be computed exactly and through the connecting functional so can the energy of an interacting (correlated) molecular system. Determining the ground-state energy is achieved through minimisation of the Kohn-Sham functional over a set of N Kohn-Sham orbitals $\phi_i(\mathbf{r})$. The energy functional is variationally minimised with respect to independent variations in these orbitals [233]. The set of Kohn-Sham orbitals that minimise the Kohn-Sham functional are given as the self-consistent solutions to the following pseudo-eigenvalue equation

$$\left(-\frac{1}{2}\nabla^2 + V_{ext}(\mathbf{r}) + V_H(\mathbf{r}) + V_{XC}(\mathbf{r})\right)\phi_i(\mathbf{r}) = \epsilon_i \phi_i(\mathbf{r}) \tag{2.5.6}$$

$$\hat{H}_{KS} \; \phi_i(\mathbf{r}) = \epsilon_i \phi_i(\mathbf{r}) \; . \tag{2.5.7}$$

The terms inside the brackets are the kinetic energy, external potential, Hartree energy and the exchange-correlation energy respectively and can be considered as a Hamiltonian (**Equation 2.5.7**). Conceptually, these equations represent a process where, given an initial

set of N Kohn-Sham orbitals, one constructs an initial (trial) electron density and independent variations are introduced into these orbitals such that electron density is gradually improved until *self-consistent*. Specifically, this iterative cycle stops when $|E^{(s)} - E^{(s-1)}| < \eta_E$ and $\int d^3r|n^{(s)} - n^{(s-1)}| < \eta_n$, where $E^{(s)}$ and $n^{(s)}$ represent the total energy and electron density at iteration s with η_E and η_n representing (user defined) tolerances. The final solution constitutes a set of N Kohn-Sham orbitals which return the ground-state electron density and ground-state energy of the molecular system.

2.6 Exchange-correlation energy

2.6.1 Exchange and correlation interactions

Electrons are fermions with spin equal to $\frac{1}{2}$ and, therefore, obey Fermi statistics, specifically, obeying Pauli exclusion. The Pauli exclusion principle states that it is not possible to constrain two electrons of the same spin to the same spatial orbital. Furthermore, since electrons are indistinguishable, the electron density calculated by squaring the modulus of the electronic orbitals ($\phi_i(\mathbf{r})$) cannot change when the electrons are interchanged between different orbitals. Essentially, the wavefunction for a system of identical fermions must be antisymmetric with respect to the exchange of any two fermions. This gives rise to the exchange energy

$$E_X = -\int\int d\mathbf{r}_1 d\mathbf{r}_2 \; \phi_i^*(\mathbf{r}_1)\phi_j^*(\mathbf{r}_2) \; \frac{1}{\mathbf{r}_{12}} \; \phi_i(\mathbf{r}_2)\phi_j(\mathbf{r}_1) \tag{2.6.1}$$

which, conceptually, is quantifying the drop in the inter-electronic Coulomb energy due to the separation of electrons with the same spin [233]. The exchange interaction, therefore, creates a volume element surrounding a given electron which prevents the access of another electron with the same spin. This volume element is referred to as the *exchange hole*.

As detailed in **Equation 2.5.5**, the Coulombic repulsion between electrons is being treated as the Coulombic repulsion between a single electron and average charge density. This average treatment means that the Coulomb repulsion between individual electrons is being neglected. Consequently, given the scenario where electrons of the same spin are allowed to fully avoid each other through Pauli repulsion, accurate quantification of individual electron-electron Coulomb interactions will lead to a further drop in the inter-electronic Coulombic energy. This energy is referred to as the correlation energy and is much harder to define compared to the exchange energy. Usually, the correlation energy is taken to be the difference between the exact many-body quantum mechanical energy, essentially the exact solution to the time-independent Schrödinger equation under the operation of the molecular Hamiltonian, and the energy predicted through DFT within the Kohn-Sham formalism [230, 232, 233].

Summation of the exchange and correlation energies constitutes the total energy of the exchange-correlation interaction(s) and is accounted for in the Kohn-Sham density-functional through the exchange-correlation term $\mathrm{E}_{XC}[\mathrm{n}(\mathbf{r})]$.

2.6.2 Approximations to the exchange-correlation energy

The exchange-correlation energy term ($\mathrm{E}_{XC}[\mathrm{n}(\mathbf{r})]$) is not known exactly and is consequently approximated. It is the definition of different exchange-correlation functionals that distinguish different DFT methods. The most basic approximation is the *local density approximation* (LDA), which assumes the exchange-correlation energy per electron ϵ_{XC} at a given point in space $\mathbf{r}$ in the molecular system is equal to the exchange-correlation energy per electron at a given point in space $\mathbf{r}$ in a homogeneous electron gas that has the same electron density [236]. Thus

$$\epsilon_{XC}(\mathbf{r}) = \epsilon_{XC}^{homo}[n(\mathbf{r})] \tag{2.6.2}$$

and

$$E_{XC}^{LDA}[n(\mathbf{r})] = \int d^3r\ \epsilon_{XC}^{homo}[n(\mathbf{r})]n(\mathbf{r})\ . \tag{2.6.3}$$

The exchange-correlation energy per electron can be broken up into individual exchange and correlation contributions

$$\epsilon_{XC}^{homo} = \epsilon_{X}^{homo} + \epsilon_{C}^{homo} \tag{2.6.4}$$

where the exchange contribution ϵ_X for a homogeneous electron gas can be evaluated analytically using the Hartree-Fock-Slater method, which assumes there exists a volume of space surrounding each electron that no other electron of the same spin is able to enter (the *exchange hole*) [233]

$$\epsilon_{X}^{homo} = -\frac{3}{4}\left(\frac{3}{\pi}\right)\int d^3\mathbf{r}\ n(\mathbf{r})^{\frac{4}{3}}. \tag{2.6.5}$$

The correlation energy ϵ_C^{homo}, although not known exactly, has been formulated based on accurate quantum Monte-Carlo studies on uniform electron gases of different electron densities [238].

LDA assumes the electron density is constant at all points in space, meaning it best captures the exchange-correlation energy for systems that closely resemble a uniform electron gas, such as metals. However, it is an unrealistic assumption in the case of molecules as variations in the electron density define chemical bonds. Therefore, it is common to employ (semi-local) gradient corrections $\nabla\mathrm{n}(\mathbf{r})$ referred to as *generalized gradient approximation* functionals (GGA). The GGA approach can be understood as applying corrections to the LDA correlation and exchange energy densities to account for areas of

electron density gradients [239]. The general form for the generalised gradient approximated exchange-correlation energy is given in **Equation 2.6.6**

$$E_{XC}^{GGA}[n(\mathbf{r})] = \int d^3\mathbf{r}\ f(n(\mathbf{r}), \nabla n(\mathbf{r}))\ . \tag{2.6.6}$$

The choice of $f(\mathrm{n}(\mathbf{r}),\ \nabla \mathrm{n}(\mathbf{r}))$ defines the variety of different GGA functionals and the best choice is largely open for debate. Perdew and co-workers [240] defined $f(\mathrm{n}(\mathbf{r}),\ \nabla \mathrm{n}(\mathbf{r}))$ as an analytic function known as the enhancement factor, $\mathrm{F}_{XC}[\mathrm{n}(\mathbf{r}),\ \nabla \mathrm{n}(\mathbf{r})]$, that directly modifies the LDA exchange-correlation energy per electron

$$E_{XC}^{GGA} = \int d^3\mathbf{r}\ \epsilon_{XC}^{homo}[n(\mathbf{r})]\ F_{XC}[n(\mathbf{r}), \nabla n(\mathbf{r})]\ . \tag{2.6.7}$$

This particular GGA functional constitutes the PBE-GGA approach [240] and this was the exchange-correlation functional employed in this investigation. In the PBE-GGA approach F_{XC} is defined as a variable dependent on the the Wigner-Seitz radius $\boldsymbol{r}_s = \left(\frac{3}{4\pi n(\boldsymbol{r})}\right)^{\frac{1}{3}}$, the spin polarization $\zeta = \frac{n(\boldsymbol{r})\uparrow n(\boldsymbol{r})\downarrow}{n(\boldsymbol{r})}$, which for closed shell systems with an even number of (paired) electrons has a 50% spin-up population and 50% spin down population, and the density gradient $\nabla \mathrm{n}(\mathbf{r})$. This results in a well balanced XC functional that describes molecular and solid-state properties equally as well [240].

Chemical accuracy is a term used throughout computational chemistry, especially in modelling the binding and thermochemistry of organic-organic and organic-inorganic complexes, and refers to the target that any predicted (binding) energy should possess a mean absolute deviation of 1 Kcal/mol [241]. The PBE functional, when corrected for dispersion (**see Section 2.11**), is able to evaluate energies of organic and inorganic complexes well within chemical accuracy [242–244]. Calculations relying on the use of (non-local) hybrid functionals, B3LYP for example, in a plane-wave pseudopotential framework (**see Section 2.8**) are computationally demanding and slowly converging [245–247] and do not make energy or geometrical predictions in (bio-)molecular systems significantly more accurate [248, 249]. In fact, the use of hybrid functionals in the plane-wave pseudopotential approach is limited to simulation cells encompassing a few dozen atoms [246], meaning the modelling of large biomolecular (exopolysaccharide) systems, as was the intention of this research, would be out of reach. The PBE functional offers a good compromise between computational expense and accuracy and has been used successfully to predict metal ion affinities in alginates [140], geometries of biomolecules [244, 248] and hydrogen bond strength and directionality in hydrogen-bonded systems [249, 250], which are all pertinent to the systems under investigation in this work. Therefore, PBE was used as the appropriate XC functional for this work.

2.7 Periodic systems

In periodic systems, Bloch's theorem can be employed [251], stating that the electronic wavefunctions can be written as a product of a wave-like part and lattice periodic part

$$\psi_{i\mathbf{k}}(\mathbf{r}) = u_{n\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \tag{2.7.1}$$

where $u(\boldsymbol{r})$ is a periodic function such that $u(\mathbf{r}) = \mathrm{u}(\mathbf{r}+\mathbf{R})$. The lattice periodic part can in turn be expanded in a basis of plane waves. This style of basis has many advantages [252]:

1. It is unbiased and complete, meaning that the basis functions span all space and, therefore, all space is treated equally.

2. Following from the previous point, given the basis is complete, any property can be represented to arbitrary accuracy, given there are enough members within the basis. In a practical sense, this means that the size of the plane-wave basis can be made suitable for any calculation simply by increasing the number of plane-waves until the property of interest, for example the total molecular energy, stops changing.

3. It does not depend on atomic positions.

Additionally, this style of basis set only presents with a few disadvantages, those being, that empty space is treated with equal representation, which can slow the speed of a DFT total energy computation, and the number of plane-waves needed is determined by the smallest oscillation/curvature in the wavefunction.

Within this research, the calculation of binding/formation energies was one of the main approaches taken to assess the thermodynamic stability of the biomolecular systems under investigation. These types of calculations can become inaccurate if small, atom-localised, basis expansions are employed, due to the basis set superposition error [253]. This is the idea that basis functions assigned to atom A borrow functions from atom B and *vice versa* and consequently this leads to over-binding estimations [253]. Using a plane-wave basis expansion, these functions are origin-less and all space is treated equally (atom independent), mitigating basis set contributions to over-binding.

2.8 Plane-waves, k-points and pseudopotentials

In a periodic system, the Kohn-sham orbitals can be expanded in a plane-wave basis [233]

$$\phi_{i\mathbf{k}}(\mathbf{r}) = \sum_{G} C_{i\mathbf{k}+\mathbf{G}} e^{i(\mathbf{G}+\mathbf{k})\cdot\mathbf{r}} \tag{2.8.1}$$

where $\mathbf{G}$ is the reciprocal lattice vector and plays the role of an exponent/frequency factor and higher $\mathbf{G}$ values indicate rapid oscillation [230]. $\mathbf{k}$ is a wave vector in the first Brillouin

zone, the primitive cell in reciprocal space. Different values of $\mathbf{k}$ can be considered as different points within reciprocal space where there is a unique solution to the Schrödinger equation [254]. The energies of the independent (non-interacting) one-electron Kohn-Sham orbitals are continuous functions of $\mathbf{k}$ and a continuous range of energies, for a given orbital ϕ_i, is referred to as a *band*. Accurate band structures require the definition of the one-electron Kohn-Sham orbitals over a range of $\mathbf{k}$-points.

The electron density is computed as an integral over the first Brillouin zone

$$n(\mathbf{r}) = \int_{BZ} d^3\mathbf{k} \, |\phi_{i\mathbf{k}}(\mathbf{r})|^2 \, . \tag{2.8.2}$$

The orbital energies ($\epsilon|\boldsymbol{k}|$) change slowly as a function of $\mathbf{k}$ (for isolated molecules there is virtually no change in orbital energy with respect to $\mathbf{k}$) and the integral can be replaced by a summation over $\mathbf{k}$ values

$$n(\mathbf{r}) = \sum_{i\mathbf{k}}^{BZ} |\phi_{i\mathbf{k}}(\mathbf{r})|^2 \tag{2.8.3}$$

which in this work is specified using an $\boldsymbol{k}_x$, $\boldsymbol{k}_y$, $\boldsymbol{k}_z$ Monkhorst-Pack grid [255], with the size of the grid representing the number of times the grid cuts the x, y and z-axes respectively. The size of the Monkhorst-Pack grid that is suitable for accurate energy computation is discussed in more detail in **Section 2.10**.

A complete description of the Kohn-Sham orbitals requires an infinite summation over $\mathbf{G}$-vectors which is not possible. The plane-wave coefficients $C_{i\boldsymbol{k}+\boldsymbol{G}}$ that describe plane-waves with a lower kinetic energy

$$\frac{(\mathbf{G}+\mathbf{k})^2}{2} \tag{2.8.4}$$

are more important to the description of the orbitals compared to plane-waves with higher kinetic energies. Consequently, the plane-wave basis can be truncated to avoid these unnecessary plane-waves by providing an energy cut-off

$$\frac{(\mathbf{G}+\mathbf{k})^2}{2} \leq E_{cut} \tag{2.8.5}$$

and the basis will exhibit systematic convergence with respect to this energy cut-off.

The core electrons are strongly localized around the nuclei and the valence wavefunctions oscillate rapidly in the core region to maintain orthogonality [236]. As a result, describing the core region requires a large set of rapidly oscillating functions i.e. a high cut-off energy. Consequently, when using a plane-wave basis, it is common to employ *pseudopotentials* (also called *effective core potentials*) to smear the nuclear charge and model the effect the core

electrons have on the valence wavefunctions. By employing a pseudopotential the need to maintain orthogonality is removed as the effect of the core electrons is incorporated into the new ionic potential. It has also been observed that relativistic effects are intimately related to the core electrons and such effects can be also be included in the potential [230]. Removing the core electrons should not effect the bonding or physical properties of the system, as these are dictated by the less tightly-bound valence electrons. Pseudopotentials that require a higher energy cut-off are referred to as *hard* whereas pseudopotentials that require a lower energy cut-off are referred to as *soft* [236].

The idea behind designing a pseudopotential is to replace the valence wavefunctions by a set of pseudo-wavefunctions that behave correctly in the outer region but without a nodal structure in the core region [230]. The core electrons are replaced by a pseudopotential parameterized such that solutions to the Schrödinger equation under this pseudopotential produce the desired pseudo-wavefunctions Φ_l with a charge density matching that of the all-electron wavefunctions Ψ_l outside of the core region ($\mathbf{r}_c$)

$$\langle \Phi_l | \Phi_l \rangle_{\mathbf{R}} = \langle \Psi_l | \Psi_l \rangle_{\mathbf{R}} \quad \text{for } \boldsymbol{R} \geq \boldsymbol{r}_c. \tag{2.8.6}$$

This is a property referred to as *norm-conservation* [256]. Now, the pseudo-wavefunctions can be expanded in a much smaller plane-wave basis making energy computations far more computationally tractable.

2.9 Inter-atomic forces

The electrons, confined to their respective Kohn-Sham orbitals, exert a force on the atomic nuclei. Therefore, the total energy is not the only parameter that can be used to assess the stability of static nuclear configurations, the net inter-atomic force can also be used. Specifically, the stable nuclear configurations adopted by molecules will encompass the smallest possible net inter-atomic force. A net inter-atomic force of zero would mean that there exists an equilibrium between the attractive and repulsive forces between the atoms in a molecule, the inference being that the atoms have adopted an equilibrium configuration.

The net inter-atomic force exerted on the nuclei by the electronic orbitals can be expressed as the negative of the gradient of the total energy with respect to the atomic positions

$$\boldsymbol{F}_{\boldsymbol{R}_I}^{elec} = -\frac{\partial E^{tot}}{\partial \mathbf{R}_I} \tag{2.9.1}$$

which is simply the slope of the Born-Oppenheimer PES (**Figure 2.1**). However, this would suggest that the total energy at two or more independent nuclear configurations needs to be known to evaluate $\boldsymbol{F}_{\boldsymbol{R}_I}^{elec}$.

Hellmann-Feynmann theorem [257] offers a more computationally tractable approach, which can obtain the net inter-atomic force for a single nuclear configuration, from the knowledge of only the wavefunction for that particular configuration. If the wavefunction Ψ is an exact wavefunction of the Hamiltonian $\hat{H}$ with energy E, then the expectation value of the derivative of the Hamiltonian constitutes the exact inter-atomic force

$$\boldsymbol{F}_{\boldsymbol{R}_I}^{elec} = -\langle\Psi|\frac{\partial\hat{H}}{\partial\mathbf{R}_I}|\Psi\rangle \ . \tag{2.9.2}$$

The Hellmann-Feynman force holds true for self-consistent solutions in the limit of a complete basis set. As such, for accurate computation of the inter-atomic force, the Kohn-Sham orbitals must be fully converged to their variational ground-state minima

$$\frac{\partial E[n(\mathbf{r})]}{\partial n(\mathbf{r})} = \frac{\partial E[n(\mathbf{r})]}{\partial C_{i\mathbf{k}+\mathbf{G}}} = 0 \ . \tag{2.9.3}$$

However, it should be noted here that if incomplete atom-centred basis sets are used, an inaccuracy is introduced into the Hellmann-Feynman force, referred to as the so-called Pulay force [258]. Atom-centred basis functions are not resistant to perturbation, for example caused by the relaxation of the nuclear configuration to its most stable geometry, and this increases the error in the computed force (**Equation 2.9.2**). However, when using a plane-wave basis set, which is not atom-centred, this inaccuracy vanishes in the limit of an increasing cut-off energy [230, 254].

2.10 Convergence testing

Truncation of the plane-wave basis set at a finite energy cut-off will lead to an error in the computed total energy. The magnitude of the error can be reduced by increasing the cut-off energy. The same is true for the **k**-point set - fewer **k**-points will lead to errors in the computed total energy. Therefore, the energy of the system should be converged with respect to the cut-off energy and the number of **k**-points [236]. In practice, the energy can be calculated as a function of increasing cut-off energy and **k**-point grid size and the point at which the energy converges within a tolerance outlined by the simulator, for example 0.05 eV, informs one on the appropriate cut-off energy and **k**-point grid size.

It is common practice to use a simplified (truncated) version of the true chemical system when convergence testing, with the only requirement being that the simplified system must encompass the same atomic species as the true system. In this work, convergence testing was performed on a single guluronic acid monosaccharide - a 23 atom molecule from which the mucoid EPS atomistic model was constructed. This structure can be seen in **Fig 2.2**.

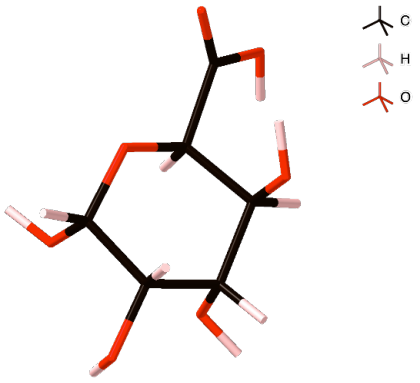


Figure 2.2: The 23 atom guluronic acid monosaccharide, the simplest building unit, possessing identical atom types, from which the mucoid EPS atomistic model was constructed.

The total energy as a function of increasing cut-off energy for this monosaccharide is displayed in **Fig 2.3**.

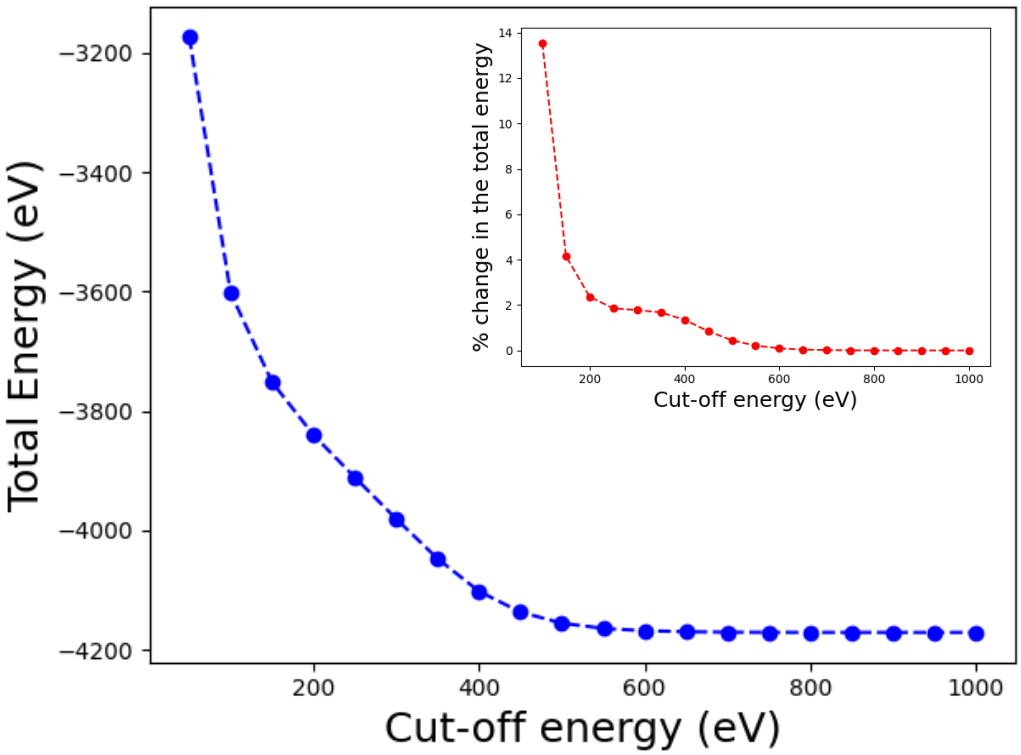


Figure 2.3: The total energy (eV) for a single guluronic acid monosaccharide as an increasing function of the cut-off energy. Inset, is the % change in the total energy as an increasing function of the cut-off energy.

As shown in **Fig 2.3**, and as expected, the total energy decreases monotonically as a function of increasing cut-off energy. By 900 eV, the total energy is converged within 0.01 eV along-side

the % change in the total energy (inset **Fig 2.3**) which is converged within 1×10^{-4} %. In addition, the magnitude of the total net inter-atomic force ($|\boldsymbol{F}^{elec}_{\boldsymbol{R}_I}|$) as a function of increasing cut-off energy is given in **Fig 2.4**.

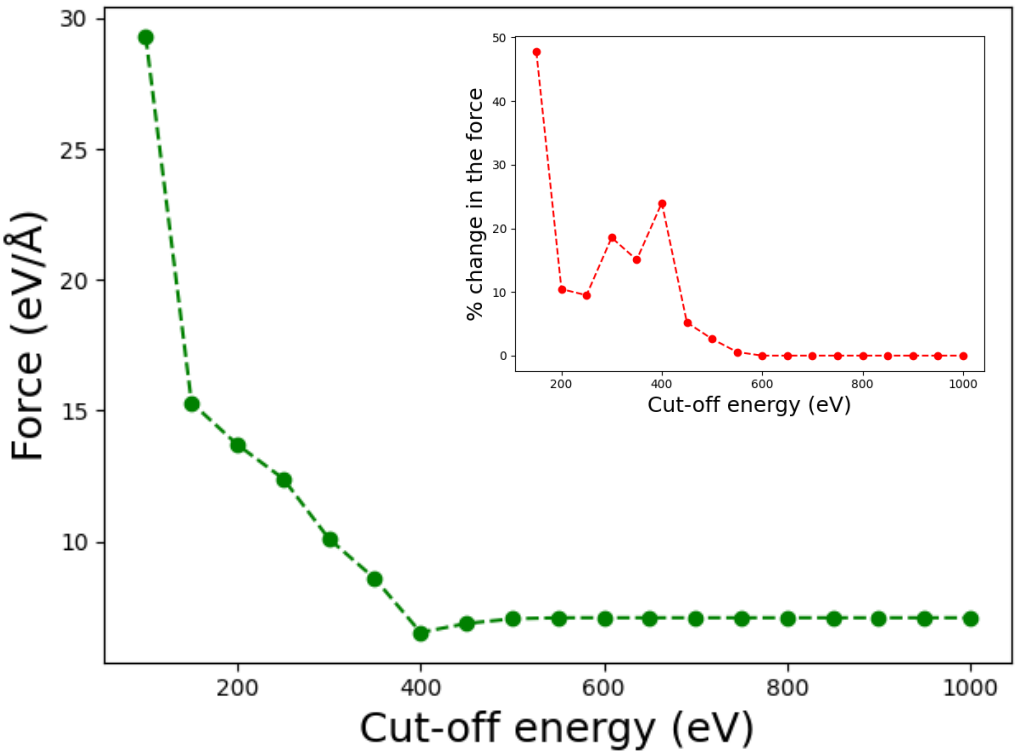


Figure 2.4: The magnitude of the total net inter-atomic force (eV/Å) for a single guluronic acid monosaccharide as an increasing function of the cut-off energy. The inset shows the % change in the magnitude of the total net inter-atomic force as an increasing function of the cut-off energy.

Likewise, the magnitude of the total net inter-atomic force decreases monotonically as a function of increasing cut-off energy, effectively converging within 0.001 eV/Å by 700 eV with the % change converging at the same cut-off. It is important to note here that, although the profile of the total force (and energy) displays a smooth response to the increasing cut-off energy, when interrogating the profile of the % change in the total force (or energy), it is not guaranteed to show the same smooth response. This is because, as the cut-off energy linearly increases, the accuracy of the force (or total energy) computation does not increase proportionately. This is illustrated in the above figures (**Fig 2.3** and **Fig 2.4**), where the magnitude of the negative gradient connecting the points at low cut-off energies fluctuates and it is this which causes the oscillatory behaviour in the % change.

It is not uncommon to observe the force converging at a lower cut-off energy compared to the total energy. Therefore, with regards to structure prediction (**Section 2.12.1**), there is

a suitable argument to use 700 eV as the cut-off energy in this work. However, it is also the focus of this research to quantify the thermodynamic stability of any predicted molecular structures. The thermodynamic stability of a molecular system is a property evaluated with the differences in relative molecular energies. An appropriate cut-off should be converged, therefore, with respect to both the force and the energy. This is achieved at 900 eV, but not at 700 eV (**Fig 2.3** and **Fig 2.4**). Considering this, 900 eV represents a highly suitable choice when deploying mucoid EPS atomistic models possessing the same atom types and was the cut-off energy chosen for this research.

The total energy as function of the number of **k**-points is given in **Fig 2.5**. It should be noted here that the profile of the total energy with respect to the number of **k**-points is not guaranteed to be monotonic, as it is for the cut-off energy. In fact, as is the case here, the profile of the total energy with respect to the number of **k**-points is oscillatory. However, the same protocol can be followed here and it is important to look for the number of **k**-points at which there is only a negligible change to the total energy. To best characterise this value, the % change in the total energy as a function of increasing number of **k**-points is also provided (inset **Fig 2.5**).

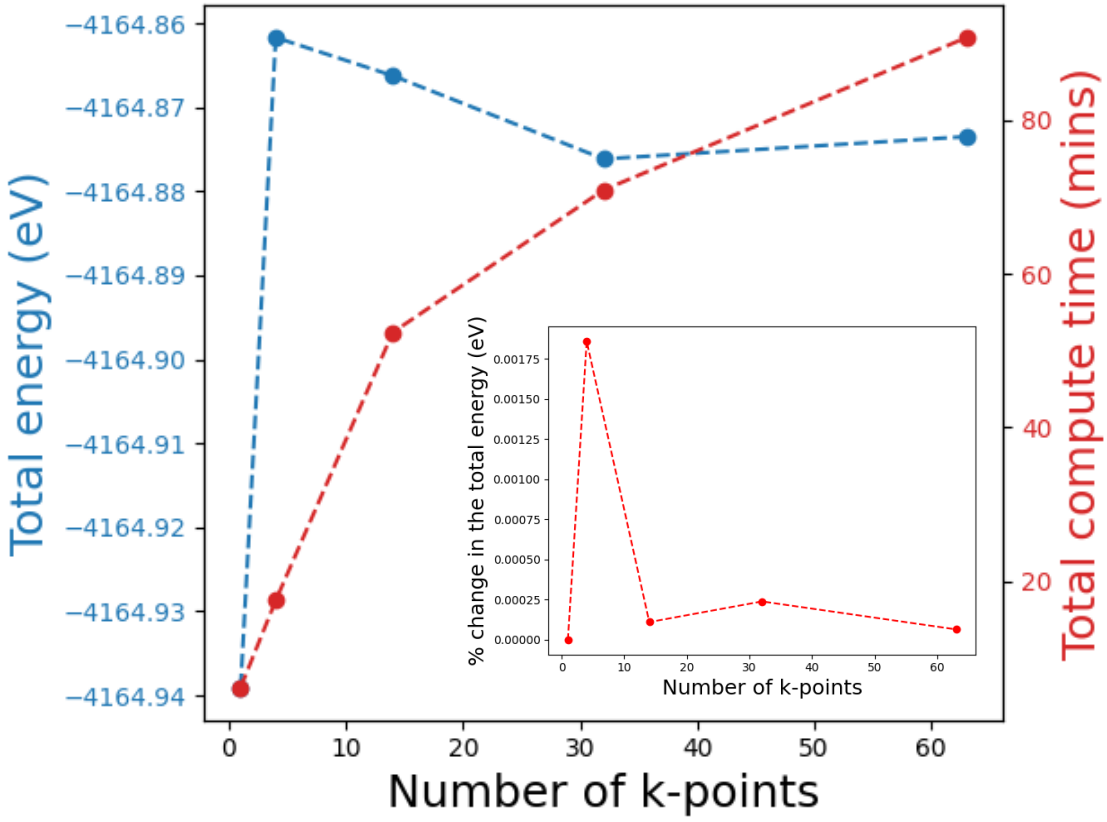


Figure 2.5: The total energy (eV), along-side the total compute time (mins), for a single guluronic acid monosaccharide as a function of increasing number of **k**-points. Inset, is the % change in the total energy as an increasing function of the number of **k**-points

Increasing the total number of **k**-points from 1 (1 × 1 × 1 Monkhorst-Pack grid to sample the Brillouin zone) to 63 (5 × 5 × 5 Monkhorst-Pack grid to sample the Brillouin zone) gives rise to extremely minor oscillations in the total energy, within 0.08 eV. As can be seen in **Fig 2.5**, increasing the number of **k**-points past 30 requires compute times over an hour for a single energy calculation on a system 22 times smaller than the final EPS atomistic model (**Chapter 4**). This represents a drastic computational bottleneck. Fortunately, as discussed in **Section 2.8**, the energy change as a function of **k** for isolated molecules is near negligible and, in fact, that is seen also here, with the % change in the total energy, as a function of **k** oscillating within 1.79 × 10^{-3} %. Taken collectively, a single **k**-point was chosen (1 × 1 × 1 Monkhorst-Pack grid) when deploying the mucoid EPS atomistic model, as it strikes a compromise between the accuracy of the computed total molecular energy and the compute time of the DFT total energy calculation.

2.11 Correcting for dispersion

Dispersion forces are long-range attractive forces between atoms and molecules that are not directly bonded, such as van der Waals interactions (vdW). Such interactions are referred to as *long-range correlation effects* [233]. Dispersive interactions arise from electromagnetic zero-point energy fluctuations in the vacuum that result in pseudo-charge densities which interact electrostatically [259]. In a more qualitative description, electron oscillations on one molecular fragment generate a transient dipole which, in turn, induces a temporary dipole on a second molecular fragment and the existence of two dipoles creates a net attractive electrostatic interaction [236]. This electrostatic interaction decays with intermolecular distance R^{-6}. The (semi-)local nature of approximate LDA and GGA exchange-correlation functionals means they are not able to describe the long-range (non-local) correlation effects from which the dispersion forces originate [233]. The omitted dispersion forces are rectified by adding a dispersion-like contribution to the total energy (referred to as the DFT-D approach) which takes the general form [260]

$$E_{disp} = -\frac{1}{2}\sum_{A,B} f_{damp}(R_{AB}, R_A^0, R_B^0) C_{6AB} \mathbf{R}_{AB}^{-6} \tag{2.11.1}$$

where $\boldsymbol{R}_{AB}$ is the interatomic distance, C_6 is the dispersion coefficient, $\boldsymbol{R}_A^0$ and $\boldsymbol{R}_B^0$ are the vdW radii. The $\boldsymbol{R}_{AB}^{-6}$ singularity at small distances is mitigated by the short-range dampening function $f_{damp}(\boldsymbol{R}_{AB}, \boldsymbol{R}_A^0, \boldsymbol{R}_B^0)$. The scheme employed in this research is that of Tkatchenko and Scheffler [261], from here on referred to as TS, which determines the dispersion coefficients and vdW radii from the ground-state electron density. Specifically, this method involves scaling the free-atom values to reflect how a given atom is modified by its environment. For example, this method will ensure that a carbon atom in benzene will have a different dispersion coefficient and vdW radius compared to a carbon atom in methane. As such, this scheme gives a dispersion energy representative of the given molecular system.

Coupled with its low computational complexity [262], when TS is combined with the PBE functional, it shows impressive improvements in the binding energy predictions within mixed hydrogen-bonded and vdW molecular dimers, encompassed within the S22 intermolecular interactions data set [261, 263]. Explicitly, an observed reduction in the mean absolute deviation (MAD) from 2.5 to 0.3 Kcal/mol upon combination of PBE and TS is observed for the S22 data set, with a larger reduction from 26.5 to 6.5 Kcal/mol observed for the S12L data set of supramolecular complexes [262, 264]. With regards to dispersion-dominated molecular extended systems, PBE-TS can accurately capture geometrical parameters, lattice constants, bulk moduli, and cohesive energies [265]. In addition, PBE-TS also accurately describes interaction distances in aromatic-π, aliphatic-π and non-π interacting complexes [266]. Therefore, overall, the TS scheme is a dispersion

scheme that is more than suitable for studying molecular complexes [267].

However, other dispersion schemes are commonly used for molecular systems, so these alternative schemes were tested to ensure the most suitable dispersion correction scheme was used. Low energy molecular geometries of an algal alginate disaccharide were predicted using PBE-TS, PBE-G06 and PBE-MBD, and compared against the crystal structure of an algal alginate disaccharide ligand complexed within AlgQ1 alginate-binding protein (PBD code 1Y3N) [268]. Comparisons of key geometrical distances are given in **Fig 2.6**.

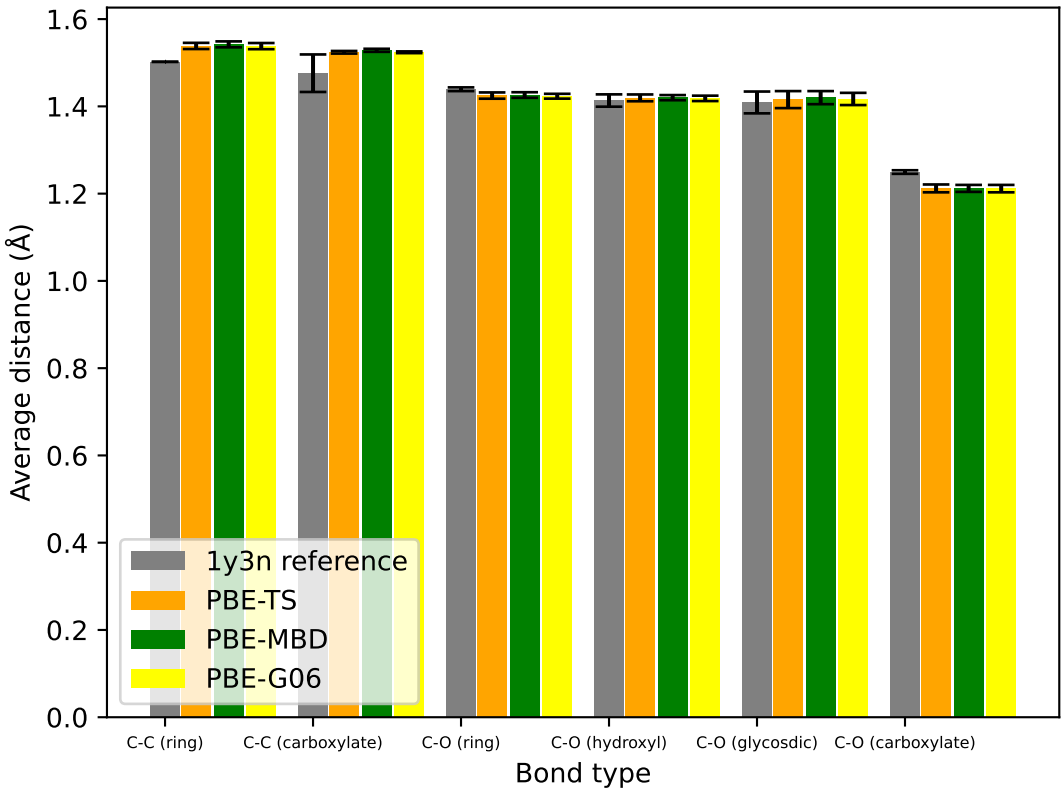


Figure 2.6: Comparisons of key geometrical distances ± 1 sdev between the dispersion corrected structures and crystal structure (PBD code 1Y3N [268]) of an algal alginate disaccharide.

As can be seen in **Fig 2.6**, all dispersion correction schemes predict the structural distances within one standard deviation of one another and no scheme appears superior upon comparison with the crystal structural reference (**Fig 2.6**). Encouragingly, these data correlate with observations made in work of Binns *et al.*, in which the performance of dispersion corrected density functional theory in optimizing molecular crystal structure parameters was assessed [269]. In this work, PBE-TS, PBE-G06 and PBE-MBD were all

shown to perform equally well in predicting C-C, C-O, O-H, C=O and C-C (aromatic) contact distances and, overall, it was concluded that PBE-TS marginally out-performed PBE-G06 and PBE-MDB for fixed and variable cell structure predictions of molecular crystals [269].

2.12 Structure and bonding prediction

2.12.1 Geometry optimisations

Given a framework that allows for the calculation of the energy of a static molecular system, it is possible to optimise the configuration of this system to identify the 3D low energy equilibrium structure from which structure-function hypotheses can be evaluated. This process is called a *geometry optimisation*.

Geometry optimisations locate molecular geometries that occupy minima (*wells*) on the ground-state Born-Oppenheimer PES. Specifically, these geometries occupy the *stationary points* on the PES. As outlined in **Fig 2.7**, stationary points can correspond to both minima and maxima. Locating a stationary point is achieved through making adjustments to the atomic positions ($\boldsymbol{R}_I$) such that the gradient in the energy (E_{tot}) with respect to each atomic coordinate is zero. To differentiate between minima and maxima, the second derivatives of the energy with respect to each atomic coordinate are evaluated, where, if all are positive, a (local) minimum has been located. If one second order energy stays negative, where all the others are positive, a saddle point, or (local) maximum, has been located.

Given the energy of the system is evaluated from the energy functional in DFT and the molecular system consists of one or more atoms $\boldsymbol{x}_1, \boldsymbol{x}_2, ...\boldsymbol{x}_i$, locating a minimum is formulated as [254]

$$\frac{\partial E_{tot}}{\partial \mathbf{x}_i} = 0; \quad \frac{\partial^2 E_{tot}}{\partial \mathbf{x}_i^2} > 0 \, . \tag{2.12.1}$$

The minima with the lowest energy is referred to as the *global* minimum and all other minima are referred to as *local* minima. **Equation 2.12.1** defines the general conditions on locating the minimum in the system energy and is further expanded below to give a more precise definition on how to locate minimum energy molecular structures.

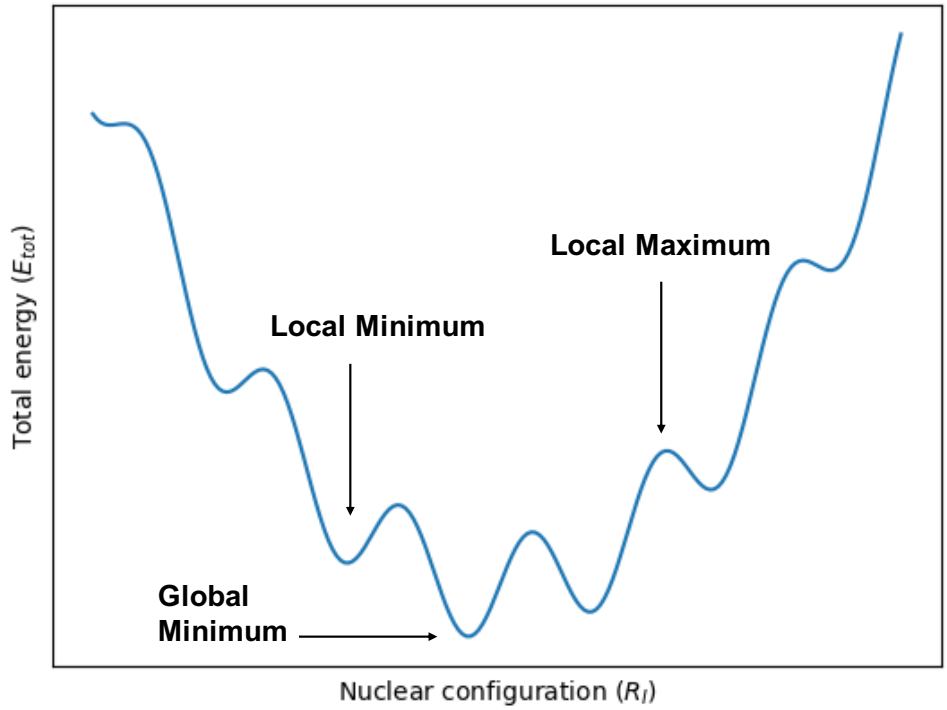


Figure 2.7: Schematic of a two-dimensional Born-Oppenheimer potential energy surface highlighting the presence of both local and global minima as well as maxima. Note, a two-dimensional surface is shown here, in preference to a three-dimensional surface as displayed in **Fig 2.1** for illustrative purposes.

Given that both first and second derivatives are needed to locate a minimum on the PES, second derivative methods are commonly employed. An example of such a method is the Newton-Raphson method. If the first derivative of the energy with respect to each atomic coordinate ($g^{(k)}$) and second derivative of the energy with respect to each atomic coordinate, which constitutes the Hessian matrix ($H^{(k)}$), are known for the current molecular geometry $\boldsymbol{q}^{(k)}$, then a new geometry $\boldsymbol{q}^{(k+1)}$ can be identified using **Equation 2.12.2** [232]

$$\mathbf{q}^{(k+1)} = \mathbf{q}^{k} - (H^{(k)})^{-1} g^{(k)}. \tag{2.12.2}$$

Equation 2.12.2 constitutes a single optimisation step and a stationary point is located when the following condition is satisfied

$$0 = g^{(k)} + H^{(k)}(\mathbf{q}^{(k+1)} - \mathbf{q}^{k}). \tag{2.12.3}$$

Calculating the full Hessian at each optimisation step is computationally expensive. The Hessian possesses n^2 elements, where n is the number of atomic coordinates in the molecular geometry vector ($\boldsymbol{q}^{(k)}$) [232]. Therefore, in systems comprised of a larger number of atoms, given the Hessian is evaluated at the first optimisation step, it is common to approximate the Hessian at each successive optimisation step, for example, from the finite differences

in the first derivatives from the previous few optimisation steps [232]. This method is a quasi-Newton-Raphson approach and is usually sufficient to navigate the PES to locate minima. For a complex PES, that is flat in many different directions, locating minima can become difficult without evaluating the Hessian at every optimisation step [232].

Geometry optimisations locate a *single* geometry typically in the *nearest* local minimum - in other words, the minimum that is closest to the region of the PES occupied by the starting structure. This can have its disadvantages, which are discussed in **Section 2.13**.

2.12.2 Mulliken population analysis

Geometry optimisations identify nuclear configurations corresponding to saddle point minima on the PES, essentially, identifying stable molecular geometries. As the energy of the system is evaluated from the DFT energy functional, the ground-state electron density at these particular nuclear configurations is also known. This means, type, directionality and strength of the inter-atomic bonds can be obtained, and provide an atomistic rationale behind a given molecular systems stability, reactivity and function.

Decomposition of the electron density into atomic contributions allows for the assignment of atomic charge and, what follows, is the approximation of the *nature* of the chemical bond between two atoms [231]. There is no quantum mechanical operator for this assignment of charge and, therefore, any partitioning scheme is arbitrary [254]. The scheme employed in this project is that of Mulliken [270]. Conceptually, electrons associated with a single basis function are thought of as being located entirely on the atom on which that basis function is centred. Electrons shared between basis functions are divided up evenly between the two atoms on which the basis functions are centred. This technique only works with localized (atom-centred) basis functions. With plane-waves, which are originless, unbiased, functions that represent all space equally, the basis must first be projected onto a localized basis [271]. The quality of the projection is gauged through the *spilling parameter*. The charge on atom A can now be defined through the Mulliken analysis

$$\rho_A(\mathbf{r}) = \sum_{\mathbf{k}} \omega_{\mathbf{k}} \sum_{\mu \in A} \sum_{\nu} D_{\mu\nu}(\mathbf{k}) S_{\mu\nu}(\mathbf{k}) \tag{2.12.4}$$

where $\omega_{\boldsymbol{k}}$ are the weights associated with the calculated $\mathbf{k}$-points in the first Brillouin zone, $D_{\mu\nu}$ is the density matrix and $S_{\mu\nu}$ is the overlap matrix. The overlap population between two atoms A and B is given by

$$N(AB) = \sum_{\mathbf{k}} \omega_{\mathbf{k}} \sum_{\mu \in A} \sum_{\nu \in B} 2 D_{\mu\nu}(\mathbf{k}) S_{\mu\nu}(\mathbf{k}). \tag{2.12.5}$$

The overlap population is also referred to as the *bond population* and large positive values indicate that the two atoms are bonded and large negative values indicate that the atoms are

in an antibonding state [272]. However, it is important to also introduce the potential issues with the Mulliken population analysis method. Firstly, the equal division of electrons between atoms is unrealistic in cases where there is an extremely large electronegativity difference between the bonded atoms and, secondly, the atomic charges are highly sensitive to the size of the basis set [254]. Furthermore, the basis set must be balanced, providing an equivalent number of basis functions for each atom. However, this latter problem is dismissed when in a plane-wave basis as these functions do not favour any particular atomic position. For these reasons, the magnitude of the assigned atomic charges have little meaning, although one is able to make comparisons between different molecular systems where the calculations have been performed with the same basis set size [272]. Furthermore, the magnitude of the Mulliken atomic charges can be used to evaluate an effective ionic valence. The effective ionic valence is the difference between the atomic charge predicted through the Mulliken analysis and the formal atomic charge [272]. The lower the value of the effective ionic valence, the more ionic character is present within the atom-atom bond [272]. Therefore, in partnership with the bond population, the effective ionic valence can be used to discriminate ionic interactions from covalent interactions.

2.13 Molecular dynamics

2.13.1 The need for a dynamical description

As discussed in **Section 2.12.1**, the process of finding a minimum energy geometry will often result in a single minimum energy geometry occupying a local minimum *closest* to the region of the PES occupied by the starting structure. However, although this may correspond to an equilibrium structure, it is insufficient to describe large chemical systems which will adopt many unique molecular configurations due to thermal motion. Specifically, a single, time-independent, structure is no longer representative of the chemical system as the probability distribution of possible structures is not compactly localised [232]. Specifically, at non-zero absolute temperature, there exists a distribution of structures found about different local minima on the PES [232]. This also applies to the prediction of molecular conformations adopted when two (or more) molecules interact. On account of this, to model this *multiple minima* phenomenon, it is advisable to compute a *molecular dynamics* trajectory. Molecular dynamics is a technique which propagates the chemical system forwards in time over the PES which, in turn, samples multiple different PES minima and, subsequently, chemical states/configurations [230]. Given a long enough molecular dynamics trajectory, the entire PES would have been accessed and the most stable configurations identified.

Notionally, chemical reactions can be described as the movement of the system's atomic nuclei between minima on the PES [230]. Therefore, molecular dynamics can be seen as performing a

computational experiment. The initial reactants (and solvent) are combined, heated to target temperature, and their time-evolution simulated. This will elucidate the final/product state of the system as well as the reaction pathway, namely the series of intermediate structures, taken to get there.

2.13.2 Simulating a system's time-evolution

Given an initial (starting) chemical system configuration, molecular dynamics (MD) computes a trajectory, successive configurations adopted by the chemical system linked in time, by solving Newton's equation of motion for each atom ($\boldsymbol{x}_i$) on the chemical system's PES [254]

$$\frac{d^2\mathbf{x}_i}{dt^2} = \frac{F_{\mathbf{x}_i}}{m_i} \; . \tag{2.13.1}$$

Solving **Equation 2.13.1** is achieved through numerical integration. The integration scheme used in this research is the velocity-Verlet scheme [273], which uses the position ($\mathbf{r}$), velocity ($\mathbf{v}$) and acceleration ($\mathbf{a}$) of each particle at time t to calculate the new position and velocity of each particle at time $t + \delta t$

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t \mathbf{v}(t) + \frac{1}{2}\delta t^2 \mathbf{a}(t) \tag{2.13.2}$$

$$\mathbf{v}(t + \delta t) = \mathbf{v}(t) + \frac{1}{2}\delta t[\mathbf{a}(t) + \mathbf{a}(t + \delta t)] \; . \tag{2.13.3}$$

The accuracy of this scheme is determined by the size of the time step δt. It is common practice to employ a time step an order of magnitude smaller than the fastest molecular motion in the system [232]. For organic systems this is a C-H vibration which occurs on the time scale of approximately 10 fs, meaning δt should be set to 1 fs. However, if C-H vibrations are constrained, then the integration time-step can be increased above 1 fs which permits the computation of longer trajectories in shorter compute times.

The force is calculated from the total energy of the system (E_{tot})

$$-\frac{\partial E_{tot}}{\partial \mathbf{x}_i} = F_{\mathbf{x}_i} \; . \tag{2.13.4}$$

The energy of the system is calculated through a classical approach as a function of the nuclear positions only and omitting the electrons completely. This approach describes the *forcefield* approach. The forcefield is a classical energetic expression possessing energy terms that describe stretching a bond between two bonded atoms, bending a bond between two bonded atoms, torsional rotation about a bond, electrostatic interaction between two non-bonded atoms and the vdW interaction between two non-bonded atoms [230]. The sum of all these energy terms provides the total energy of the system. The functional form for such a (general)

forcefield is as follows [254]

$$E_{tot}(\mathbf{r}^N) = \sum_{bonds} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{angles} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 + \sum_{torsions} \frac{V_n}{2}(1 + cos(n\omega - \gamma) + \sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(4\epsilon\left[(\frac{\sigma_{ij}}{\mathbf{r}_{ij}})^{12} - (\frac{\sigma_{ij}}{\mathbf{r}_{ij}})^{6}\right] + \frac{q_i q_j}{4\pi\epsilon_0 \mathbf{r}_{ij}}\right). \tag{2.13.5}$$

The functional form given in **Equation 2.13.5** describes how the total energy $E_{tot}(\boldsymbol{r}^N)$ which is a function of $\mathbf{r}$ positions and N particles is changing as the molecule is perturbed. The first two terms describe how the energy of a system changes as bonds are stretched and bent away from their reference values and are modelled using a harmonic potential. Using a Taylor-series expansion of the stretching and bending potential energy between two particles, it is possible to include cubic and quartic terms in the potential also, to better mimic the true Morse stretching and bending potential. However, these terms increase the computational complexity when evaluating these energy terms, whilst only benefiting the bonding description at extreme stretching and bending - which isn't applicable for bonds in organic and biological molecules which are stiff and remain close to equilibrium at physiological temperature [232, 254]. As such, the Taylor-series expansion is truncated at the quadratic term.

The third term in the forcefield describes how the energy of the system changes through rotation about a bond modelled using a torsional potential. Finally, the last two terms describe the non-bonded contributions, calculated between all pairs of atoms separated by at least three bonds, with the first being the Lennard-Jones potential and the latter being the Coulomb potential describing vdW and electrostatic interactions respectively. More specifically, the Lennard-Jones term is attempting to describe the distance-dependent dispersive attraction and repulsion (the dispersion energy was outlined in **Section 2.11**) between two particles. In the Lennard-Jones term, σ is the inter-particle separation that gives a value of zero for the dispersion potential and ϵ is the depth of the dispersion potential energy well. The Coulomb term is attempting to describe the classical distance-dependent electrostatic attraction and repulsion between two particles. In the Coulomb term, q denotes the atomic charges and ϵ_0 denotes the vacuum permittivity. Examples of these potential forms are given in **Fig 2.8**.

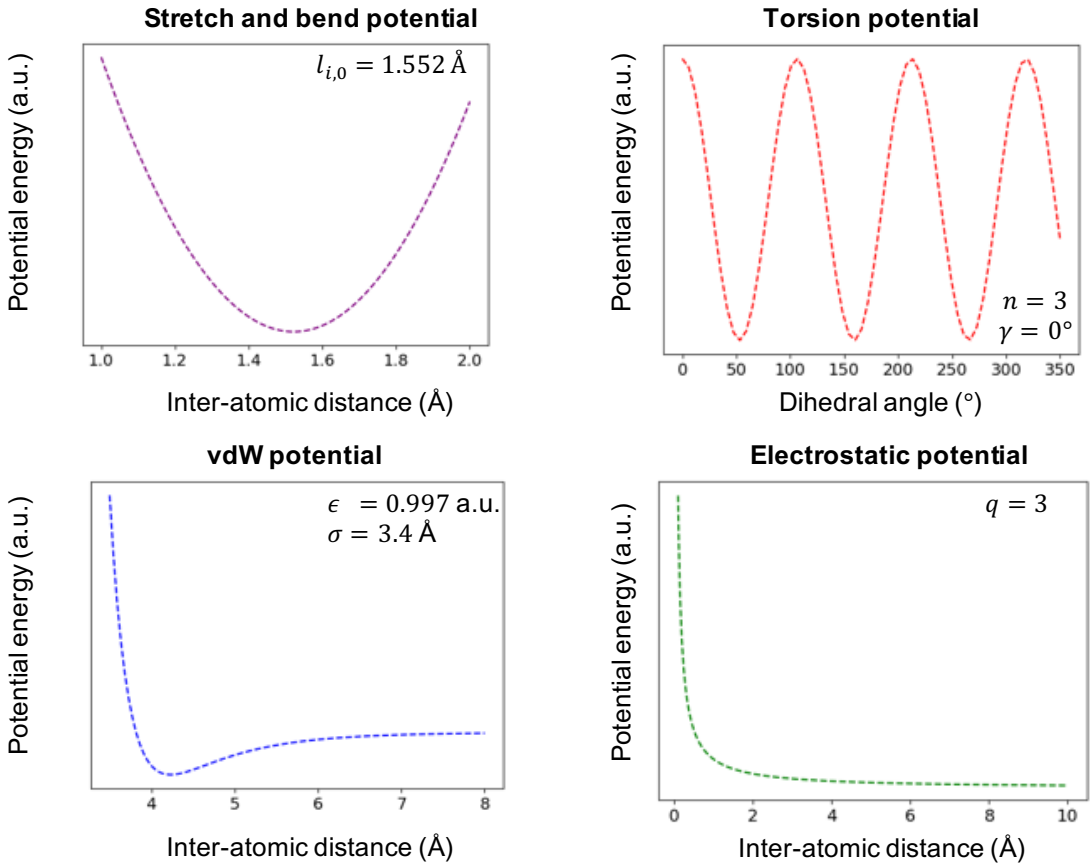


Figure 2.8: Illustrations of the forcefield potential forms given in **Equation 2.13.5**.

It is clear to see that, generically, semi-empirical forcefields are simply a collection of potentials and their associated constants and, by virtue, are highly parameterised objects. Forcefields, typically, are parameterised against structural and energetic data obtained either from experiment or high-level quantum chemical calculations [232]. The semi-empirical forcefield must be able to replicate structures and energies observed in experiment or quantum chemical calculations. The direct implication of forcefield parameterisation is that a level of *uniqueness* is inherent in each different forcefield making its performance (and accuracy) tailored to specific systems, in essence the systems that formed the dataset from which the forcefield was parameterised. Therefore, it is paramount that the right forcefield - the most suitable collection of potentials and constants - should be used to describe the system at hand. This research employed the OPLS2005 forcefield [274, 275]. This is a forcefield that extends the well-known AMBER forcefield [276], and is suitable for the simulation of biomacromolecules, allowing for accurate simulations of a wider array of different (highly solvated) biomolecules, specifically pharmaceutically relevant molecules and oligomers. The OPLS2005 forcefield includes bond stretch, bend, torsional and vdW potentials, parameterised against LMP2/cc-pVTZ(-f) and B3LYP/6-31G* quantum chemical calculations, which improve the accuracy of the conformational energetics and molecular

electrostatic potentials of organic (bio-)molecules [274, 275]. For clarity, these quantum chemical calculations are not periodic, plane-wave, DFT calculations performed using GGA functionals as is used in this research, but rather real-space, atom-localised, DFT calculations performed using double-hybrid (LMP2) and hybrid (B3LYP) functionals.

2.13.3 Sampling the canonical ensemble

Given high enough molecular energy, the classical propagation of the atomic system over the PES allows the system to overcome energy barriers and access multiple minima (stable states) as indicated in **Fig 2.9**. The energy of the molecular system is intimately related to the temperature. As most chemical phenomena occur under conditions whereby heat is exchanged with the environment, it is common to conduct MD simulations in the *canonical* ensemble - also referred to as the NVT ensemble. In this ensemble, the number of atoms (N), volume (V) and temperature (T) are all held constant during the simulation and the temperature is controlled using a thermostat [232]. For a thermally equilibrated system, the temperature of an NVT ensemble is related to the time average of the kinetic energy

$$\langle KE \rangle = \frac{3}{2} N k_b T \tag{2.13.6}$$

where N is the number of atoms, k_b is Boltzmann's constant and T is the temperature. Thus, for the thermostat to control the temperature, it must regulate the kinetic energy of the system. There are multiple different ways of doing this, where the most simple is to re-scale the velocities of each particle at each time step to maintain a constant temperature - the so-called velocity re-scale thermostats [277, 278]. The thermostat used in the molecular dynamics simulations in this work was the Langevin thermostat [279], which is a slightly more complex approach, accounting for the friction exerted on the particles by the solvent medium, slowing them and in turn regulating the system temperature. On that account, the Langevin thermostat performs well for systems encompassing molecules immersed in solvent medium.

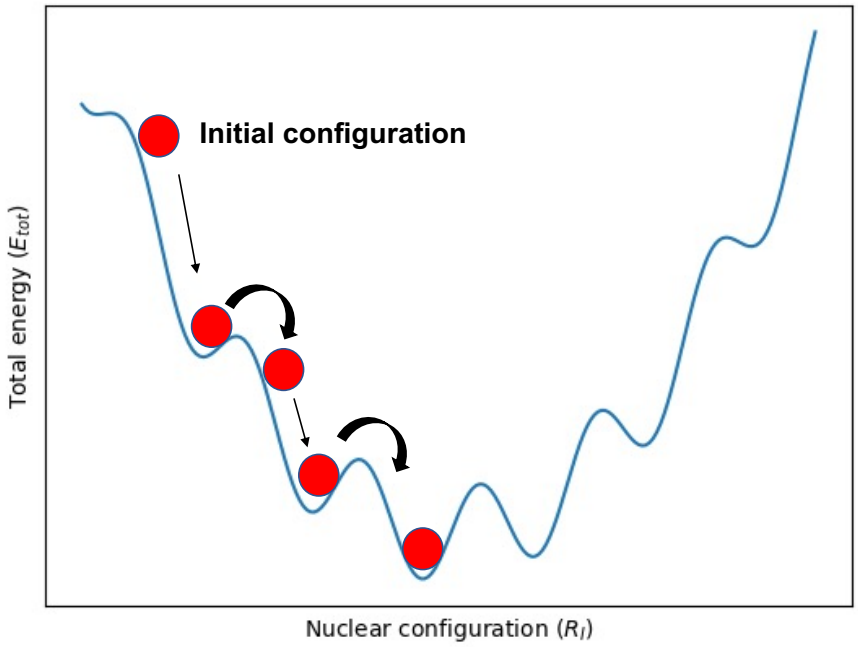


Figure 2.9: Molecular dynamics trajectory of a molecular system (red ball) over the Born-Oppenheimer potential energy surface. Each arrow represents a $t + \delta t$ classical propagation of the system. The chemical system's trajectory will also sample, and overcome, high energy configurations (barriers on the PES) in search for the lowest energy state.

To generate a molecular dynamics trajectory, all that is needed to begin with are the initial coordinates, velocities and forces of all atoms in the system. However, for a trajectory to sample the entire PES, impractically long simulation times are a requirement which, unfortunately due to the large computational expense incurred, in the majority of cases cannot be fulfilled. Therefore, along with the choice of an appropriate time step, a choice must be made on the simulation length and one must choose a given length that ensures the dynamical trajectory explores regions of the PES, exploring different molecular configurations, that are of most significance in describing the desired (reaction) chemistry.

The trajectory length chosen to study exopolysacchsride (EPS) conformational change, as well as the movement of quorum sensing autoinducer (QSAI) molecules through the EPS, in this research was 10 ns. This particular trajectory length was chosen as it ensured that the EPS system had reached a static, unchanging, equilibrium configuration from which relationships between structure, function and virulence could be established. But also, 10 ns was sufficient to allow the QSAI molecules to sample, and eventually settle into, preferential binding sites on the EPS scaffold offering key structural insights into the interactions that

dictate the movement of these molecules through the EPS matrix.

2.14 The CASTEP code

In this project, the first principles calculations are carried out using CASTEP [280], which is a plane-wave, pseudopotential, Density-Functional Theory (DFT), quantum mechanical code. The CASTEP code uses a computationally efficient protocol to compute the ground-state electronic structure, energy and geometry of molecular systems. The theoretical basis for the CASTEP code is to obtain the ground-state energy and electron density from solving the Kohn-Sham pseudo-eigenvalue equation (**Equation 2.5.6**) through a self-consistent field procedure. The electronic wavefunctions (Kohn-Sham orbitals) are expanded in a basis of plane-waves, defined by the use of periodic boundary conditions and Bloch's theorem [251], employing pseudopotentials [281] to describe the core electron-ion potential. Following the calculation of the energy for a static nuclear configuration, the positions of the nuclei are then optimised using quasi-Newton-Raphson methods to identify the most stable structure(s) adopted by the system of interest.

2.15 The DLPOLY4 code

DLPOLY4 [282] is a general purpose molecular dynamics (MD) simulation code, that allows for the computation of MD trajectories, using the OPLS2005 forcefield [274, 275], and can be used to explore the many conformations adopted by large solvated bio-polymeric systems within the canonical (NVT) ensemble. DLPOLY4 permits the control of temperature through Langevin thermostatting [279] and features the RATTLE algorithm [283], which constrains covalent bonds to hydrogen atoms, allowing for the use of larger integration time-steps. This, in turn, facilitates the simulation of molecular motion over long time-scales providing detailed insights into processes such as molecular recognition and binding.

2.16 Relative hydrophobicity and SwissADME

In this research, when investigating the molecular motions of quorum sensing autoinducer (QSAI) molecules through the mucoid exopolysaccharide, the relative differences in hydrophobicity were sought to assist in the explanation of solvent mediated dynamical behaviour. The hydrophobicity difference between two molecules is quantified through estimation of the $\log(P)$ for each molecule. P is a partition coefficient, describing the ratio of the neutral solute concentration in octanol and water at equilibrium

$$log(P) = log\left(\frac{[solute]^{non-ionized}_{octanol}}{[solute]^{non-ionized}_{water}}\right). \tag{2.16.1}$$

The larger the value for log(P), the more hydrophobic the solute. In practice, for an octanol/water system, log(P) is calculated as the Gibbs free energy change when transferring the neutral solute between the two solvents [284]

$$-RT \times log(P) = \Delta G^{o/w}_{transfer} \tag{2.16.2}$$

$$\Delta G^{o/w}_{transfer} = \Delta G^{o}_{solv} - \Delta G^{w}_{solv} \, . \tag{2.16.3}$$

It is possible to evaluate **Equation 2.16.3** from first principles, in the plane-wave pseudopotential Kohn-Sham DFT framework, firstly, through the calculation of the many particle partition function (Z) [285]

$$Z = \int d\mathbf{r} \, \exp\left[-\hat{H}/k_b T\right] . \tag{2.16.4}$$

Within this framework, simulations of molecules are performed *in vacuo* under constant volume, and the Helmholtz free energy (F) is evaluated as follows

$$F = -k_b T \, log(Z) \, . \tag{2.16.5}$$

The Helmholtz free energy is transformed into the Gibbs free energy through the following relation [285]

$$G = F + PV \tag{2.16.6}$$

where PV is the product of pressure and volume. However, the integral in **Equation 2.16.4** is over all particle coordinates ($\mathbf{r}$), i.e., over all possible atomic configurations, which is not computationally possible for large molecular systems as it would require exploration of the full PES.

SwissADME [286] is a software tool used to evaluate pharmacokinetic parameters of small molecules and provides a quicker, more computationally tractable, method to evaluate Log(P) through the *iLOGP* approach [284]. The *iLOGP* method is a multi-linear regression model trained on a drug-like dataset encompassing 11,993 different molecular structures. The electrostatic contribution to the Gibbs free energy in octanol and water ($\Delta\Delta G_{solv,elec} = \Delta G^{w}_{solv,elec} - \Delta G^{o}_{solv,elec}$) along-side the solvent accessible surface area (SASA) in octanol and water ($\Delta SASA = SASA^{w} - SASA^{o}$) were calculated for each one of the 11,993 molecules in the dataset using the CHARMM forcefield, and then used to train a multiple linear regression (MLR) model [284]

$$Log(P) = \alpha_1 \times \Delta\Delta G_{solv,elec} + \beta_1 \times \Delta SASA + C_1 \tag{2.16.7}$$

$$Log(P) = \alpha_2 \times \Delta\Delta G_{solv,elec} + \beta_2 \times \Delta SASA^{w} + C_2 \tag{2.16.8}$$

where α_1, α_2, β_1, β_2 are the regression coefficients and C_1 and C_2 are the regression constants. The MLR model was validated against the known experimental Log(P) values for the same

molecules within the dataset, giving a correlation coefficient $r = 0.72$ [284]. This is of suitable accuracy to qualitatively assess the difference in hydrophobicity between two independent molecular structures, explicitly, the difference in hydrophobicity between different QSAI molecules.

2.17 Summary

In summary, the energy of a chemical system can be obtained from the system's electron density, which is the basis of the very effective Density-Functional Theory approach. The energy of a molecule has parametric dependence on the molecular structure, which introduces the concept of a potential energy surface and locating the three-dimensional minimum energy structures becomes an optimisation problem - identifying the atomic arrangement that occupies a minimum on the potential energy surface. However, for chemical systems that are inherently *dynamic* in their ground-states, a *multiple minima problem* arises and molecular dynamics must be employed to sample the representative conformational ensemble.

In this chapter, plane-wave pseudopotential Density-Functional Theory and classical molecular dynamics approaches have been introduced, which represent the theoretical basis for the CASTEP electronic structure [280] and DL POLY molecular dynamics [282] codes respectively. These computational chemistry codes were utilised in this work to develop accurate, structurally representative, molecular models of a mucoid *P. aeruginosa* exopolysaccharide matrix. These molecular models were deployed to probe key structural chemistry underpinning matrix stability, permeability and destruction, with the hope of elucidating insights into novel therapeutic strategies.

Model Development

3.1 Scope of the chapter

This chapter describes the step-wise creation of two mucoid *P. aeruginosa* exopolysaccharide (EPS) matrix molecular models. These models were created to capture, in a thermodynamically favourable fashion, key structural motifs unique to mucoid *P. aeruginosa*, that serve to distinguish *bacterial alginate* from the well-known classical *algal alginate*. The molecular models were validated by ensuring experimentally determined metal ion affinities (gelation stability trends) could be recreated, defining the geometrical requirements for stable cation accommodation. Furthermore, these models assist in highlighting the relationship between structural chemistry and bacterial virulence, the rationale being that the virulence in cystic fibrosis (CF) sputum is a consequence of ion-induced cross-linking. Finally, the resulting models can be used to study cation substitution chemistry as well as small-molecule inhibitor delivery at the atomistic-scale.

3.2 Computational details

3.2.1 DFT simulation parameters

All geometry optimizations were performed using the plane-wave Density Functional Theory (DFT) code, CASTEP [280]. For all polyuronate and ion-complexation optimizations, a convergence tested cut-off energy of 900 eV was employed and a Monkhorst-Pack **k**-point grid of 1 × 1 × 1 was used to sample the Brillouin zone [255]. Simulations were conducted in an orthorhombic box of size 45 Å × 27 Å × 16 Å. On-the-fly ultrasoft pseudopotentials were used [281] alongside the PBE exchange-correlation functional [240]. The semi-empirical dispersion correction of Tkatchenko and Scheffler [261] was employed to account for intra- and intermolecular dispersive forces. The applicability of the PBE-TS scheme to studying molecular systems was discussed in **Section 2.11**. The SCF tolerance was set to 1×10^{-7} eV Atom^{-1} and the energy, force and displacement tolerances for the geometry optimisations were set to 1×10^{-5} eV Atom^{-1}, 0.03 eV Å^{-1} and 1 $\times10^{-3}$ Å respectively. Following each

geometry optimization, Mulliken bond populations [270] were calculated, following the projection of plane-waves into atomic orbitals [271], to classify the nature of bonding in each of the complexed structures.

For the determination of formation energies, chemical potentials for sodium, magnesium and calcium were calculated by their respective zero Kelvin (0 K) energy per atom, from the pure metals in their lowest energy configurations, namely HCP sodium, hexagonal magnesium and BCC calcium. These chemical potentials, appropriate for oxygen-poor conditions, can be seen in **Table 3.1**. Oxygen consumption by bacteria, neutrophils and CF airway cells effectively renders the thick CF mucus anaerobic [287, 288]. As such, CF sputum hosts very steep oxygen gradients and below the airway-surface interface, CF sputum is effectively hypoxic or anoxic [289]. Oxygen deprived conditions are further suggested through the observation of high levels of expression of the gene encoding OprF, a protein that is critical for anaerobic *P. aeruginosa* biofilm formation, and other gene clusters required for anaerobic respiration *in vivo* in the CF airway [288, 290]. Chemical potentials for hydrogen and ethanal were calculated from optimized single molecules. All calculations were conducted using the same PBE-TS functional at the same cut-off energy of 900 eV.

Table 3.1: Chemical potentials (eV) for the ionic species

Chemical potentials (eV)	
Na	−1310.33
Ca	−1004.88
Mg	−1690.79

It is important to note here, that the chemical potentials of the ionic species depend on their mole fractions within the ECM. As such, it is worth noting that the chemical potentials given in **Table 3.1** can be transformed into values that reflect their relative abundances within the ECM using **Equation 3.2.1**

$$\mu_A = \mu_A^0 + RT\ \ln(f_A) \tag{3.2.1}$$

where μ_A^0 is chemical potential of the ion calculated from the pure metal and f_A is the mole fraction of the ion within the environment, explicitly, the ratio of the number of moles of the ion to the total number of moles of all constituents within the full system. However, determining f_A for sputum ions in the ECM *in vivo* is highly non-trivial and, consequently, these values are currently unknown. As such, this research makes the approximation that the sputum ions are entering, and complexing into the system, from their bulk phase and f_A is taken as 1. Therefore, the chemical potential of the ionic species is taken solely as μ_A^0.

3.2.2 The omission of water

The mucoid *P. aeruginosa* biofilm extracellular matrix (ECM) architecture *in vivo* has been characterised using confocal laser scanning microscopy and fluorescent lectin binding analysis, revealing an open 3D network of matrix material [175, 291] encompassing void spaces where the vast majority of biofilm water is confined [170]. Pulsed field gradient NMR has highlighted that a very minor amount of water is entrapped, bound entropically, within the secondary structures of the polyuronates, but this is exchanged frequently with the bulk [170]. NMR observations on water transport through polysaccharide gels provide consistent observations, that water molecules near to the polysaccharide scaffold do not have a significantly reduced motion - in other words, water molecules do not bind specifically [292]. Furthermore, with reference to cation complexation, crystal structures of calcium-polysaccharide complexes highlight that stable cation accommodation occurs preferentially in positions where oxygen functionality is well-positioned to displace water from the coordination shell of the cation [293]. This is a requirement satisfied when placing the cations in-between polyuronate chains. Consequently, as well as for reduced computational expense, all optimisations were performed *in vacuo* with the omission of water molecules. This, in turn, allowed for large-scale DFT optimisations focusing solely on polyuronate-cation interactions (**Section 3.4**).

3.3 Acetylation configurations

Recall in **Section 1.3**, mucoid *P. aeruginosa* alginate is acetylated at the hydroxyl groups bound at the C2 or C3 carbon positions within the ring of the mannuronate residues, and is catalysed by the AlgI-AlgJ-AlgF protein complex [114, 115]. In effect, the acetylation reaction involves the abstraction of the H atom from the OH groups at the C2 or C3 positions coupled with the addition of a CH_3CO moiety in its place. The acetylation reaction mechanism may possibly follow the standard acetylation pathway of alcohols as is described in **Fig 1.5**. What follows the C2 or C3 acetylation is the down-regulation of epimerase activity. Failure to epimerise the (acetylated) mannuronate units to guluronate units gives rise to two separate, structurally distinguishable, EPS fractions. One where the EPS is composed exclusively of mannuronate polymer sequences and the other where the EPS is composed of mannuronate-guluronate copolymeric sequences. These represent fractions of the EPS corresponding to 0% and 50% (minimum and maximum) guluronate inclusion respectively.

For the molecular models to reflect this, two molecular templates were defined. The first was a poly-β-D-mannuronic acid quadramer and the second was a copolymeric β-D-mannuronic-α-L-guluronic acid quadramer. The former encompasses a single polyuronic acid chain of four M units linked via a 1–4 glycosidic bond and the latter encompasses a single polyuronic acid chain of two M units and two G units linked via a 1–4 glycosidic bond

in an alternating M-G pattern. It is important to note here that M-M-G-G uronate backbone configurations were not considered as this introduces a contiguous G-block, which is not observed in mucoid *P. aeruginosa* EPS scaffolds [106, 111]. These can be seen in **Fig 3.1**.

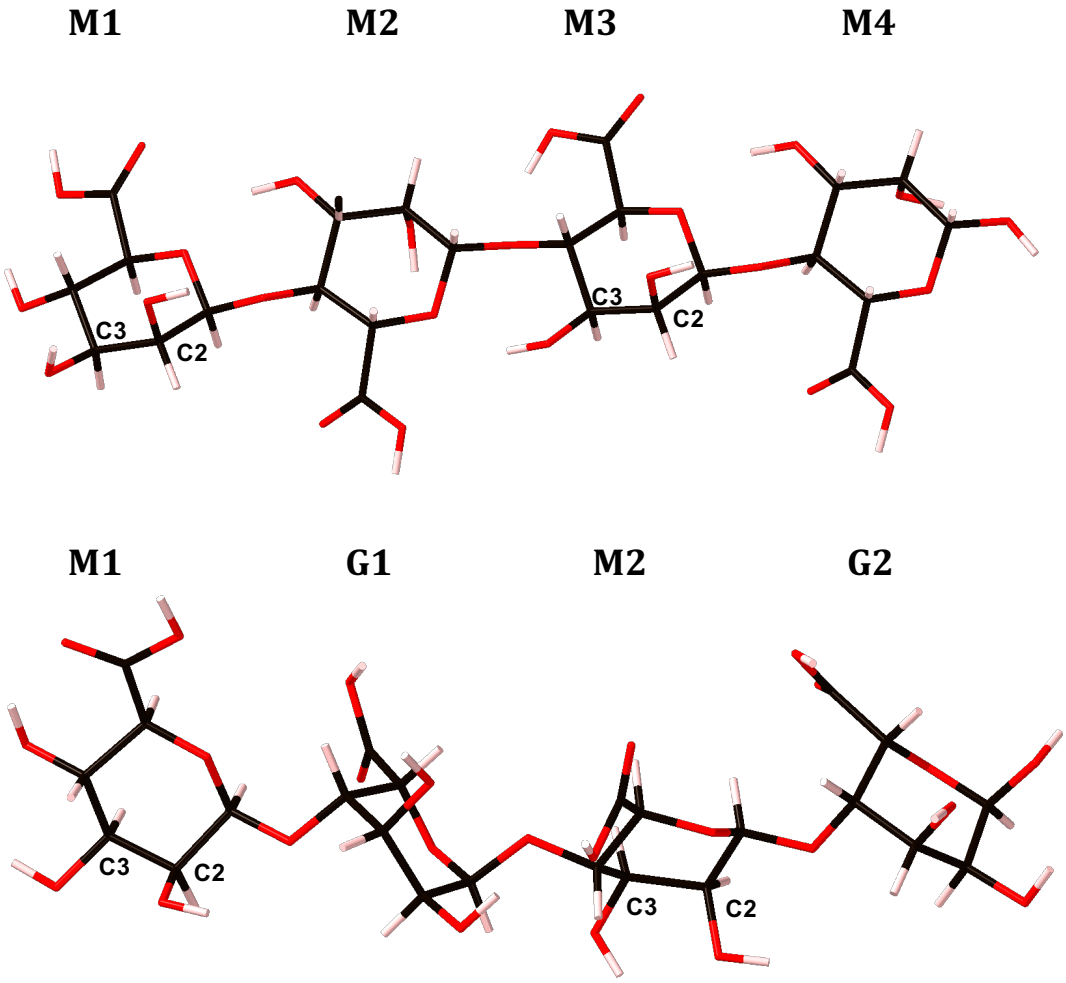


Figure 3.1: Poly-β-D-mannuronic acid (top) and copolymeric β-D-mannuronic-α-L-guluronic acid (bottom). Carbon atoms are shown in black, oxygen in red and hydrogen in pink. The uronate units are labelled and the carbon atom labels indicate the C2 and C3 carbon positions on the mannuronic acid structures which are subject to acetylation, defined as above.

The molecular templates were then acetylated at the C2 or C3 positions [111, 112] exclusively on the M units, as displayed in **Fig 3.1**, and subjected to a geometry optimisation. The degree of acetylation chosen was 50% - a value that falls within the $4\% - 57\%$ range determined experimentally [111, 112]. In assessing the thermodynamic stability of the resulting acetylated alginate structure, ethanal (acetaldehyde) was chosen as the source of the CH_3CO moiety and the acetylation modelled as the abstraction of H from the OH group and from the α-C from the ethanal molecule with the addition of the CH_3CO to the deprotonated hydroxyl at the

C2 or C3 positions within the ring. As such, the thermodynamic stability of each acetylation pattern was measured by means of the formation energy according to **Equation 3.3.1**

$$E_f = E_{final} - (E_{initial} + nE_{ethanal} - m\mu_H) \tag{3.3.1}$$

where E_{final} is the energy of the resulting acetylation configuration, $E_{initial}$ is the energy of the initial template, $E_{ethanal}$ is the energy of an ethanal molecule and μ_H is the chemical potential of a hydrogen atom. For the poly-β-D-mannuronic acid quadramer, $n = 2$ and $m = 4$ and for the copolymeric β-D-mannuronic-α-L-guluronic acid quadramer, $n = 1$ and $m = 2$.

It is worth noting here, that the formation energy is evaluating the thermodynamic favourability of forming a particular molecular structure, from a set of constituent reactant structures, purely enthalpically. DFT-based formation energy calculations ignore entropic contributions as they are performed at 0 K in vacuo. The same applies to all formation energy calculations performed throughout the remainder of this research.

The formation energies for the different acetylation configurations are shown in **Table 3.2**.

Table 3.2: Formation energies (eV) for the different acetylation configurations for each polyuronic acid template. The asterisk (*) indicates the most thermodynamically favourable acetylation configuration for the β-D-mannuronic acid and β-D-mannuronic-α-L-guluronic acid systems.

β-D-mannuronic acid	Energy (eV)	Relative energy (eV)	E_f (eV)
M1(C2) M2(C2)	-16881.26	0.08	-0.16
M1(C2) M2(C3)	-16881.20	0.14	-0.10
M1(C3) M2(C2)	-16881.28	0.06	-0.18
M1(C3) M2(C3)	-16880.90	0.44	0.20
M1(C2) M3(C2)	-16880.74	0.60	0.36
M1(C2) M3(C3)	-16880.64	0.70	0.46
M1(C3) M3(C2)*	-16881.34	0.00*	-0.24*
M1(C3) M3(C3)	-16881.24	0.10	-0.14
M1(C2) M4(C2)	-16881.02	0.32	0.09
M1(C2) M4(C3)	-16880.61	0.73	0.49
M1(C3) M4(C2)	-16881.09	0.25	0.01
M1(C3) M4(C3)	-16880.93	0.41	0.17
β-D-mannuronic-α-L-guluronic acid	Energy (eV)	Relative energy (eV)	E_f (eV)
M1(C2)	-16093.61	0.12	0.06
M1(C3)	-16093.56	0.17	0.11
M2(C2)	-16093.37	0.36	0.29
M2(C3)*	-16093.73	0.00*	-0.06*

As the acetylation configurations for the β-D-mannuronic and β-D-mannuronic-α-L-guluronic acid templates are structural isomers of one another (encompassing the same number of atoms, molecular weight and atom type) it is possible to evaluate the most stable acetylation configuration without the explicit need to evaluate E_f. In the circumstance of structural isomerism, the most stable conformer (acetylation pattern) can be obtained by interrogating the relative differences in the molecular energy. This is also given in **Table 3.2**. As an assurance, the correlation between the relative energy and formation energy has also been calculated by means of Pearson's correlation coefficient, given in **Fig 3.2**. As can be seen, the relative energy difference between acetylation configurations (acetylation structural isomers) is strongly, linearly, positively correlated with the formation energy. The lowest energy acetylation configuration based on comparison of the molecular energy differences is also the most thermodynamically favourable based on **Equation 3.3.1**.

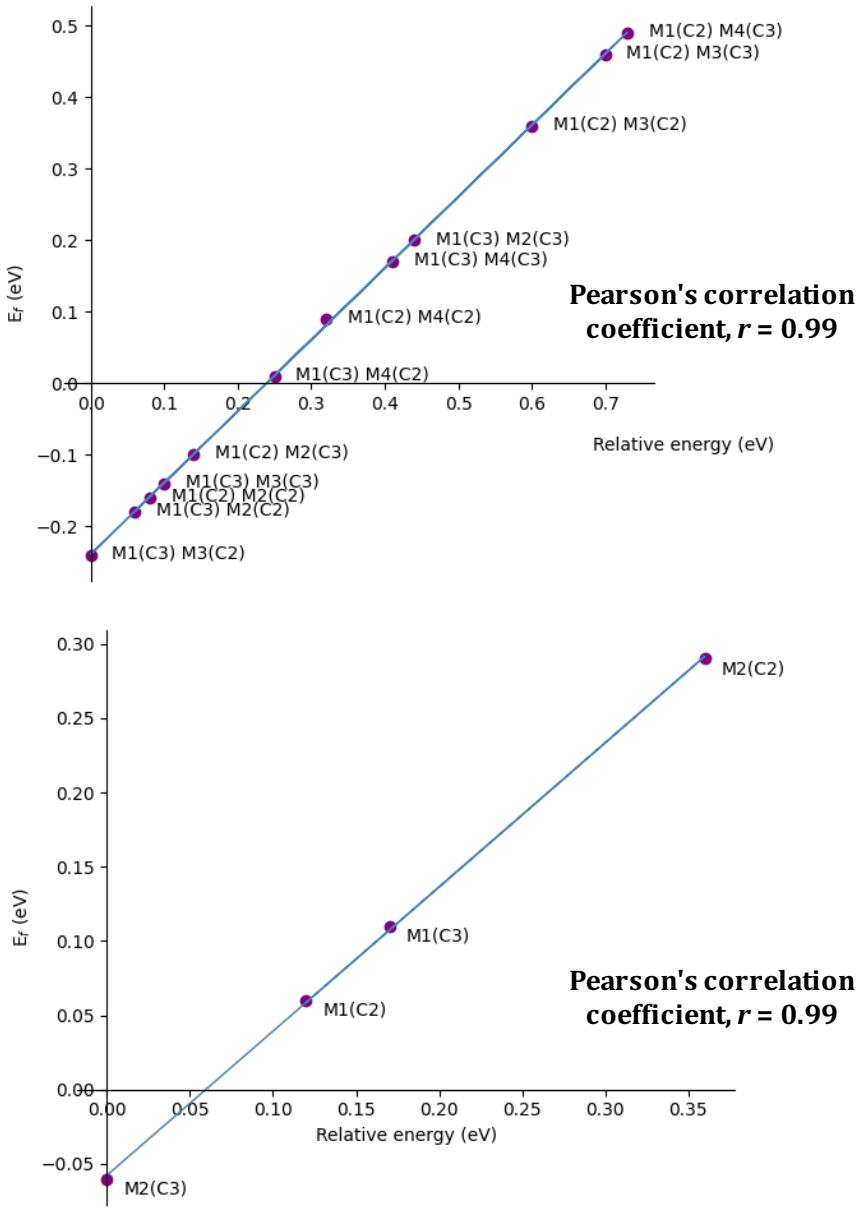


Figure 3.2: The correlation between the relative energy difference (eV) and formation energy (eV) of the different acetylation patterns for the β-D-mannuronic acid template (top) and β-D-mannuronic-α-L-guluronic acid template (bottom).

Therefore, from both **Table 3.2** and **Fig 3.2**, it is evident that the M1(C3) M3(C2) acetylated β-D-mannuronic acid and M2(C3) acetylated β-D-mannuronic-α-L-guluronic acid are the most stable acetylation patterns for EPS quadramers encompassing solely M-M and M-G junctions respectively. These structures are given in **Fig 3.3**.

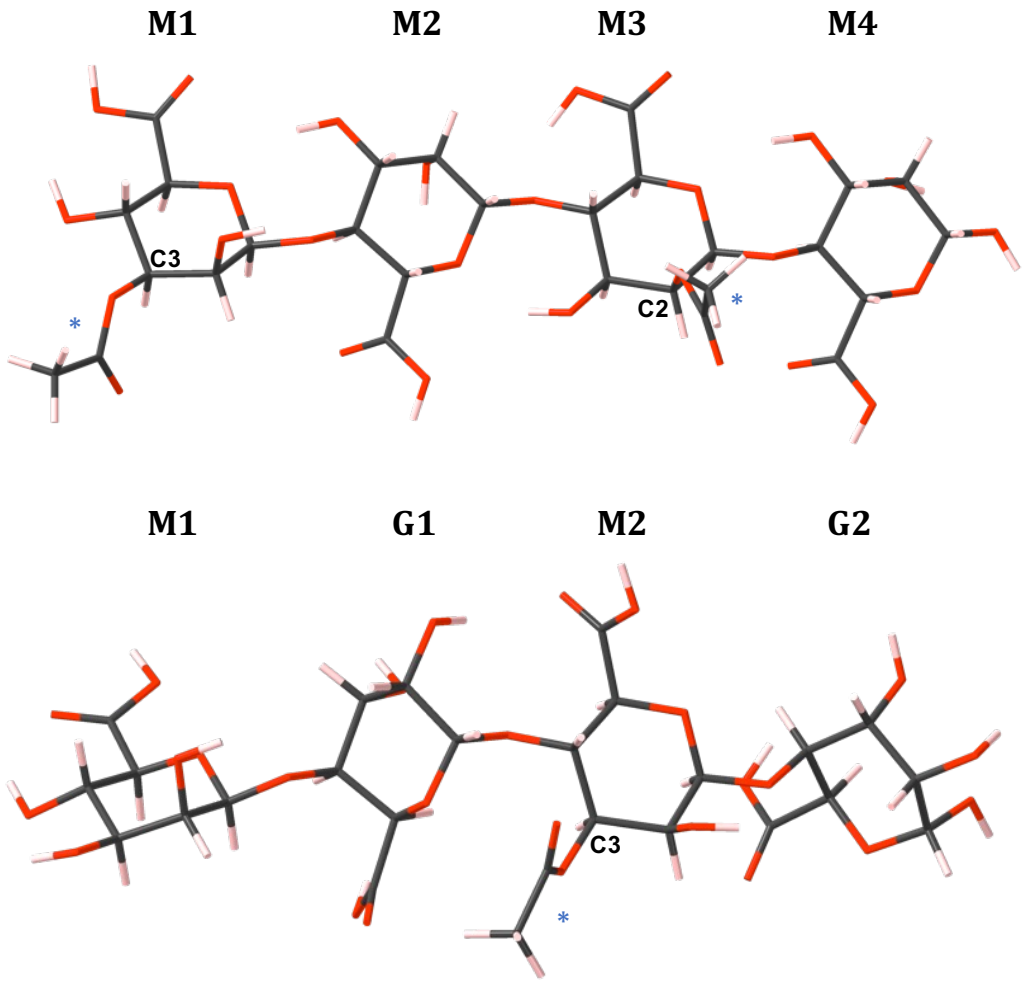


Figure 3.3: Most thermodynamically favourable acetylated poly-β-D-mannuronic acid (top) and copolymeric β-D-mannuronic-α-L-guluronic (bottom) configurations. Carbon atoms are shown grey, oxygen red and hydrogen pink. The acetyl groups are indicated by a blue asterisk and the acetylated carbon atom positions are labelled for clarity.

Although the most thermodynamically favourable acetylation configurations for each polyuronic acid system have been identified, other thermodynamically unfavourable acetylation configurations were found. These unstable acetyl positions indicate for our given systems, a degree of selectivity as to where an acetyl group could be attached. Only a single acetylation pattern was found to be thermodynamically favourable in the copolymeric β-D-mannuronate-α-L-guluronic acid system and this structure is less thermodynamically favourable relative to the most favourable acetylated poly-β-D-mannuronic acid configuration. However, it is important to note here that additional stable acetyl rotations, about the CO – $COCH_3$ bond, may exist at these carbon positions that haven't been located due to the fact that geometry optimisations identify local minima (see **Section 2.12.1**).

It is thermodynamically favourable to acetylate at adjacent carbon atom positions in

the poly-β-D-mannuronic acid system, for example, M1(C2) M2(C2) and M1(C3) M2(C2) configurations (**Table 3.2** and **Fig 3.2**), suggesting inter-acetyl steric repulsion does not destabilise the acetylation configuration and EPS backbone. In fact, ^{1}H-NMR confirms the occurrence, albeit at low abundance, of C2-C3 di-acetylated mannuronate structures and, even though di-acetylated configurations have been neglected here, this would also imply that steric bulk at one carbon position following acetylation does not prevent acetylation at the other adjacent carbon position [111]. In the same system, it is thermodynamically unstable to position the acetyl groups as far away from one another as possible (M1 M4 configurations), suggesting that acetyl-acetyl hydrophobic interactions are required to stabilise the acetylation configuration.

Overall, these models suggest a limit for acetyl steric accessibility, explicitly, having acetyl groups too far apart can give rise to thermodynamically *unstable* acetylation patterns. But on the other hand, directly adjacent acetyl groups do not correspond to the most stable positioning, suggestive that it is most stable to acetylate within an EPS region that balances intra-acetyl steric repulsion with acetyl-acetyl hydrophobic interactions. From here on, the lowest energy and most thermodynamically favourable acetylated poly-β-D-mannuronic acid (**Fig 3.3** top) and copolymeric β-D-mannuronic-α-L-guluronic acid (**Fig 3.3** bottom) configurations are referred to as PolyM and PolyMG respectively. These models represent mucoid *P. aeruginosa* EPS structures that possess the unique structural motifs observed *in vivo* and are the models carried forward to study cation accommodation and cross-linking by the EPS.

3.4 Cation cross-linked EPS scaffolds

3.4.1 Requirements for stable cation binding

The extracellular cationic atmosphere in the CF lung alters the chemical structure of the EPS through inducing ionic cross-links between EPS chains, increasing the mechanical stability of the EPS. As described, the consequence of increased structural rigidity makes the biofilm matrix more difficult to clear using conventional chemotherapies. To build a mucoid *P. aeruginosa* EPS molecular model, it is necessary to include cations. Sodium, calcium and magnesium ions are of particular importance, as these are most abundant in CF sputum and available to induce EPS cross-linking (magnesium: 30 mg/L, calcium: 102 mg/L, sodium: 10 g/L [124, 125]). Conditions for stable cation binding have been well-characterised, both experimentally and theoretically, for algal alginates (see **Section 1.3.2**). However, the oxygen functionality required for cation accommodation by acetylated bacterial alginates is unknown.

To understand how each acetylated polyuronate chain accommodates cations, geometry

optimisations of single sodium, calcium and magnesium ions were performed at different points along the length of a single chain axis in the PolyM and PolyMG systems, in the vicinity of the M-M and M-G junctions, as shown in **Fig 3.4**. These positions correspond to three unique positions along the length of each chain axis where the cation could potentially sustain contacts to multiple oxygen atoms.

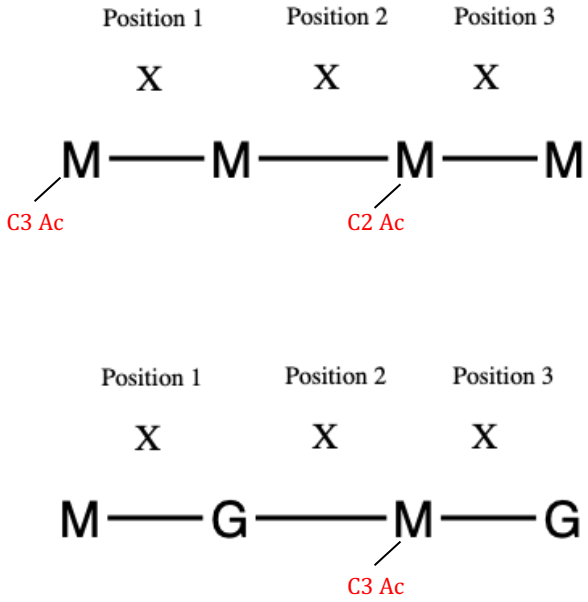


Figure 3.4: Schematic showing the three unique ion positions (denoted as X) along the length of the chain axis in the PolyM and PolyMG systems. The M-unit acetylation positions are also labelled to show that the three ionic positions are unique.

It was important to place the cations in positions within the short oligosaccharide models that were unbiased and opened up the possibility of all oxygen bearing functionality, namely the acetyl group, carboxylate group(s), hydroxyl group(s), ring oxygen(s) and glycosidic oxygen(s), engaging in the complexation event. In addition to the single cation optimisations, geometry optimisations were also performed where all four carboxylate groups of the PolyM and PolyMG chains were deprotonated, as is expected *in vivo*, and charge saturated (i.e., binding of four monovalent sodium ions and two divalent calcium or magnesium ions). The thermodynamic stability of each cation-bound position was gauged by means of evaluating a formation energy using **Equation 3.4.1**.

$$E_f = E_{final} - (E_{initial} + n_A\mu_A - m\mu_H) \tag{3.4.1}$$

where E_{final} is the final energy of the resulting complex, $E_{initial}$ is the energy of the polyuronate chain, μ_A is the chemical potential of the cationic species and μ_H is the chemical potential of a hydrogen atom. The most thermodynamically favourable binding position of single sodium, calcium and magnesium ions along the length of a single PolyM and PolyMG chain axis are shown in **Figs 3.5 - 3.7**. For reference, the optimized positions of the ions

at the other two (less stable) positions along the length of the chain axis are labelled and displayed as grey spheres. **Table 3.3** displays the formation energy for each single ion complex and charge saturated complexes.

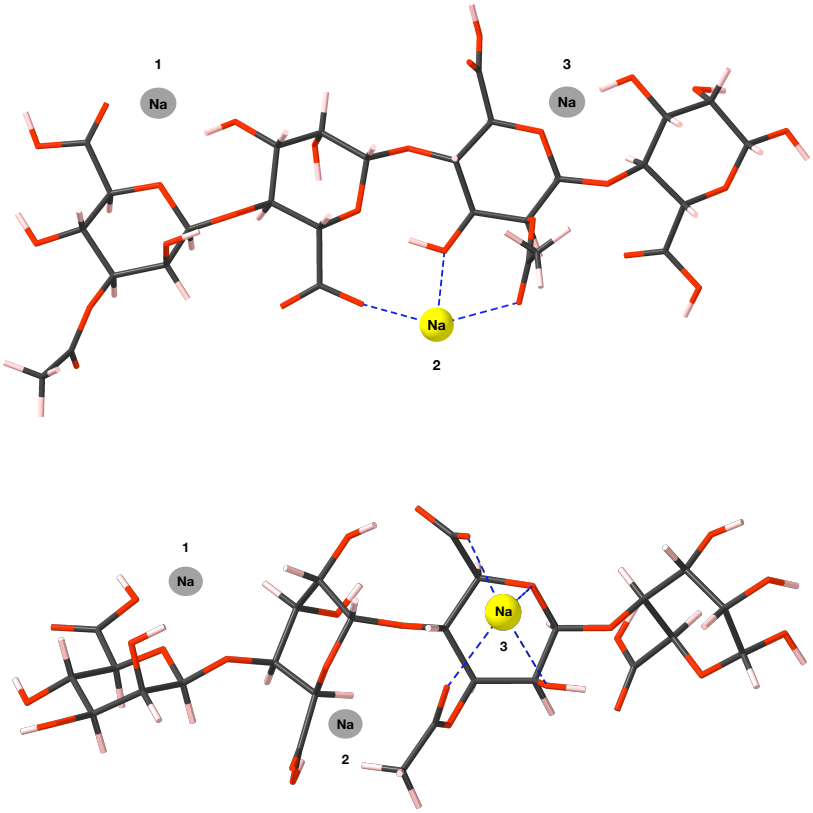


Figure 3.5: Most thermodynamically favourable binding position for a single sodium ion along the length of a single PolyM (top) and PolyMG (bottom) chain. Carbon atoms are shown in grey, oxygen in red, sodium as yellow spheres and hydrogen in pink. Ionic bonds to the sodium ion are labelled as blue dash lines. For reference, the optimized positions of the ions at the other two points along the length of the chain axis are labelled and displayed as grey spheres.

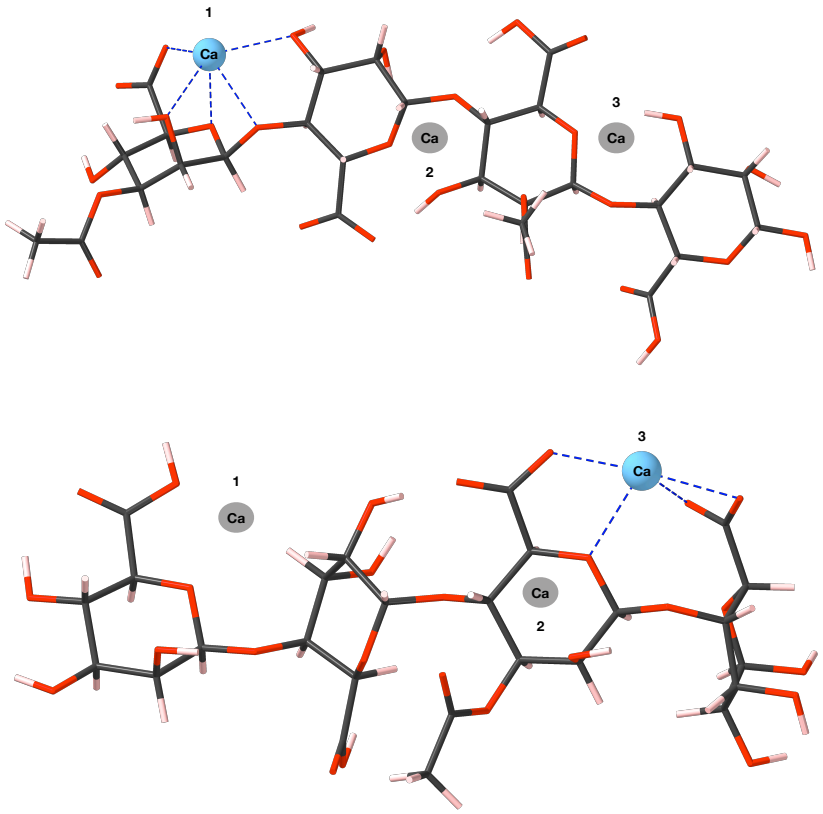


Figure 3.6: Most thermodynamically favourable binding position for a single calcium ion along the length of a single PolyM (top) and PolyMG (bottom) chain. Carbon atoms are shown in grey, oxygen in red, calcium as blue spheres and hydrogen in pink. Ionic bonds to the calcium ion are labelled as blue dash lines. For reference, the optimized positions of the ions at the other two points along the length of the chain axis are labelled and displayed as grey spheres.

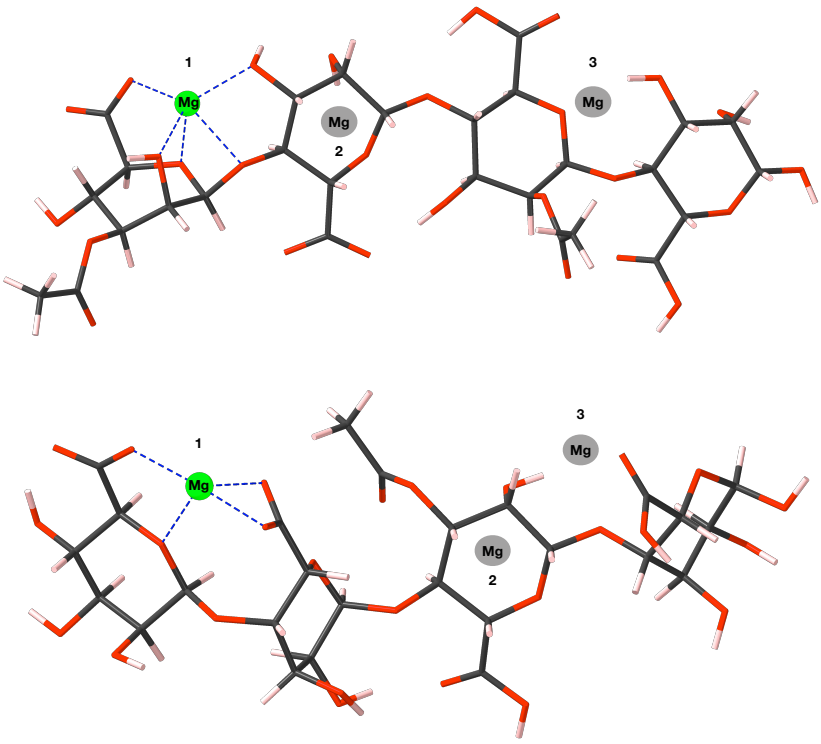


Figure 3.7: Most thermodynamically favourable binding position for a single magnesium ion along the length of a single PolyM (top) and PolyMG (bottom) chain. Carbon atoms are shown in grey, oxygen in red, magnesium as green spheres and hydrogen in pink. Ionic bonds to the magnesium ion are labelled as blue dash lines. For reference, the optimized positions of the ions at the other two points along the length of the chain axis are labelled and displayed as grey spheres.

Table 3.3: Formation energies (eV) for all the single ion binding positions along the length of a single PolyM and PolyMG chain. Formation energies for the fully charge-saturated single PolyM and PolyMG chains with respect to the four carboxylic acid groups are also given.

Cation	Position	E_f (eV) in PolyM complexes	E_f (eV) in PolyMG complexes
Na^+	1	-0.97	-0.89
	2	-1.15	-1.21
	3	-0.13	-1.35
	Charge saturated	-3.55	-4.58
Ca^{2+}	1	-1.81	-0.94
	2	-1.17	-0.84
	3	-1.26	-2.94
	Charge saturated	-3.04	-5.73
Mg^{2+}	1	-0.49	-1.85
	2	0.24	-0.10
	3	0.08	-1.84
	Charge saturated	-0.65	-3.53

Firstly, it is worth noting that nearly all single ion complexes in both the PolyM and PolyMG systems are thermodynamically favourable, meaning cations can be accommodated by the EPS at a single-chain level. In other words, cations can be accommodated by the EPS without the need of being entrapped in-between two chains. The only exceptions are the magnesium ion positions 2 and 3 in the PolyM system, which give slightly thermodynamically unstable structures, in turn highlighting a degree of selectivity for where magnesium ions can bind along the length of a single PolyM chain. It is most thermodynamically favourable to bind single cations within M-G junctions and within these structures, taken collectively, the cations form multiple oxygen contacts to all oxygen bearing functional groups. This aligns well with the calcium chelation geomtries observed in the crystal structures of calcium ions complexed by β-D-linked pyranoses, which are carbohydrates analogous to the EPS models reported here [293, 294]. For reference, and for comparison with cation-complexed geometries predicted in this work, the x-ray crystallographic structure of a calcium-pyranose complex, specifically, calcium complexed onto a single β-D-galacturonate chain, is given in **Figure 3.8**.

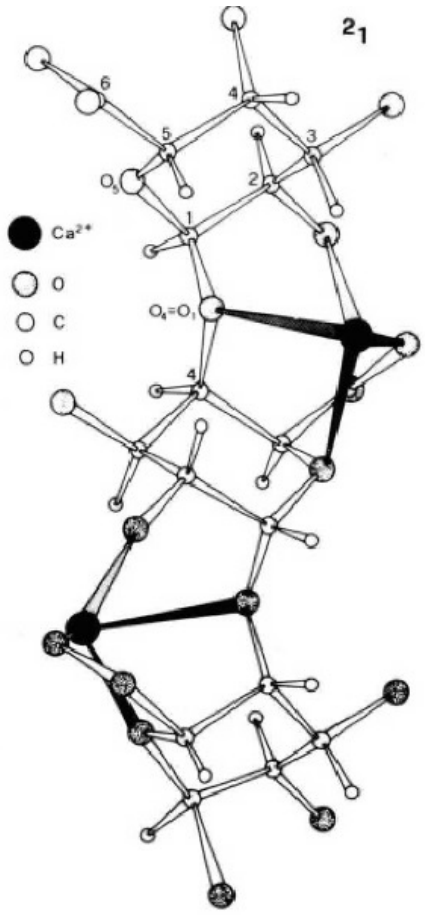


Figure 3.8: The calcium-β-D-galacturonate structure as determined through x-ray crystallography. Figure printed from reference [294]. The 2_1 annotation refers to the 2 repeat units per turn helical conformation of the poly-β-D-galacturonate.

Furthermore, observing cation-oxygen contacts to carboxylate, hydroxyl and acetyl functional groups is also consistent with high resolution solid-state ^{13}C-NMR spectra of *P. aeruginosa* biofilms that implicate these functional groups in the formation of electrostatic cross-links [121].

Similarly, CN of up to 5 are seen in the stable divalent ion positions, which matches DFT predictions of magnesium and calcium binding to M-M, M-G, G-M and G-G disaccharides [140]. The COO^- group is involved in cation binding in the most stable positions for all ions. The binding to this group is monodentate in the stable sodium positions, but bidenatate in the divalent ion positions when the number (and variety) of coordinating oxygen functionality is lower (i.e., bidentate in CN < 5 environments). Therefore, the preference for monodentate or bidentate COO^- binding modes at the single-chain level is determined by the degree of saturation in the chelate pocket, with higher degrees of saturation (higher CN) biasing towards the monodentate mode. Most importantly, charge saturating

the PolyM and PolyMG systems with respect to all available COO^- groups gave the most thermodynamically favourable structures for each cation. In other words, it is more stable for the bacterial alginate chain to complex cations in a fully deprotonated state, compared to situations where single uronic acid units remain in their protonated form; it is more stable to saturate the COO- group with a cation than a proton. Given that *in vivo* the mannuronate and guluronate carboxylic acid groups are deprotonated (the pKa values of mannuronic acid and guluronic acid are 3.38 and 3.65 respectively), it is more stable for these anionic scaffolds to saturate their charge using the available cationic atmosphere. This, in turn, justifies the placing of multiple, rather than single, cations in-between the 2 chains when studying 2-chain cross-linking (see **Section 3.4.3**).

Stable coordination environments for sodium, calcium and magnesium ions along the length of a single bacterial alginate chain have been identified, revealing that the origin of stable complexation arises from binding multiple oxygen atoms from different functional groups. This served as a guide when positioning the cations in-between 2-chains for obtaining models of EPS cross-linked structures.

3.4.2 2-chain arrangement in space

Before allowing two polyuronate chains to complex cations, the most thermodynamically favourable arrangement of two polyuronic acid chains in space was evaluated. The configurations tested were: parallel, parallel-inverted (where one polyuronate chain has been inverted 180°about the chain axis), antiparallel and antiparallel inverted. The formation energy for each spatial arrangement was evaluated using **Equation 3.4.2**.

$$E_f = E_{final} - (2E_{initial}) \tag{3.4.2}$$

where E_{final} is the final energy of the two-chain arrangement and $E_{initial}$ is the energy of a single PolyM or PolyMG chain. The formation energies and number of hydrogen bonds established for the two chain configurations are given in **Table 3.4**. The antiparallel arrangement of two PolyM structures and the parallel arrangement of two PolyMG structures were the most thermodynamically favourable spatial arrangements and are shown in **Fig 3.9**.

Table 3.4: Formation energies (eV) and number of hydrogen bonds established for the different spatial arrangements of two PolyM and PolyMG chains.

Spatial arrangements		
PolyM	E_f (eV)	N^o hydrogen bonds
Parallel	−2.27	2
Parallel-inverted	−0.93	1
Antiparallel	−2.29	4
Antiparallel-inverted	−1.66	2
PolyMG	E_f (eV)	N^o hydrogen bonds
Parallel	−2.87	6
Parallel-inverted	−1.92	3
Antiparallel	−1.90	4
Antiparallel-inverted	−2.65	4

All spatial arrangements are thermodynamically favourable for both the PolyM and PolyMG systems, meaning it is thermodynamically favourable for two bacterial alginate chains to associate in space in the absence of cations. Hydrogen-bonding between the polar oxygen containing functional groups of adjacent chains is responsible for mediating the association. The number of hydrogen bonds established between the two chains positively correlates with thermodynamic stability, with the spatial arrangements that adopt more hydrogen bonds being more stable. In general, the PolyMG spatial arrangements adopt a larger hydrogen-bonded network between chains and return the more stable structures. This is because, within M-G junctions, there is a wider variety of hydrogen-bonding functionality at a suitable orientation to sustain more hydrogen bonds. For example, oxygen functionality that participates in hydrogen bonding (i.e., an acceptor) is limited to carboxyl and hydroxyl functionality in the most stable (antiparallel) PolyM system. In the most stable (parallel) PolyMG system, the glycosidic oxygen also acts as a hydrogen-bonding acceptor. From here on, the antiparallel PolyM and parallel PolyMG arrangements are referred to as PolyM_{ap} and PolyMG_{p}, respectively.

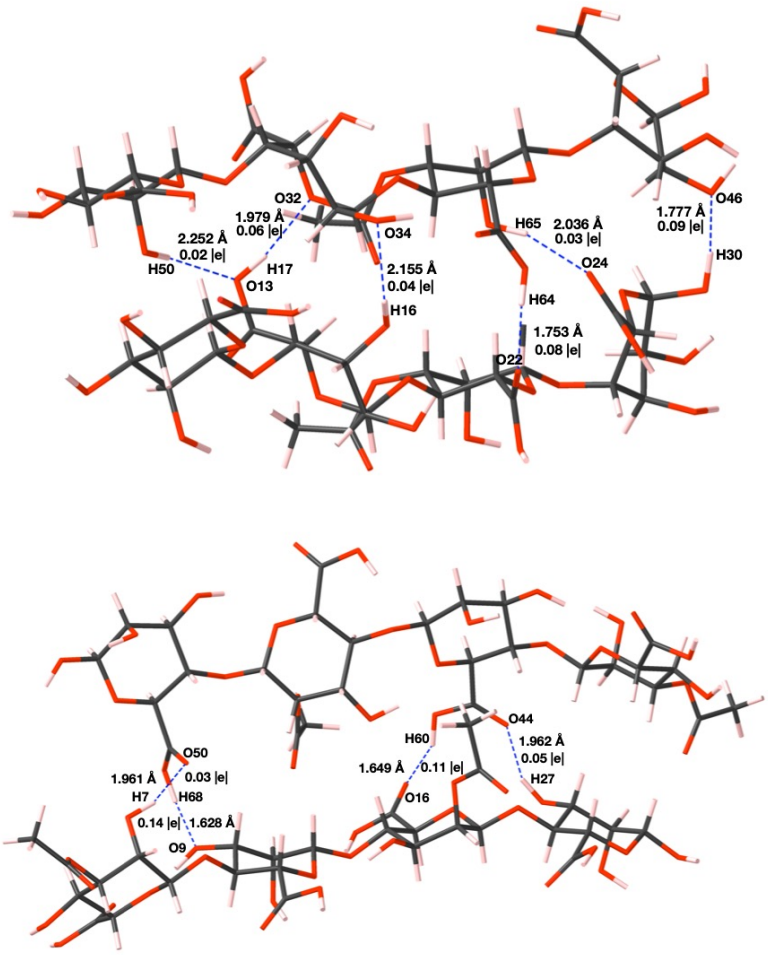
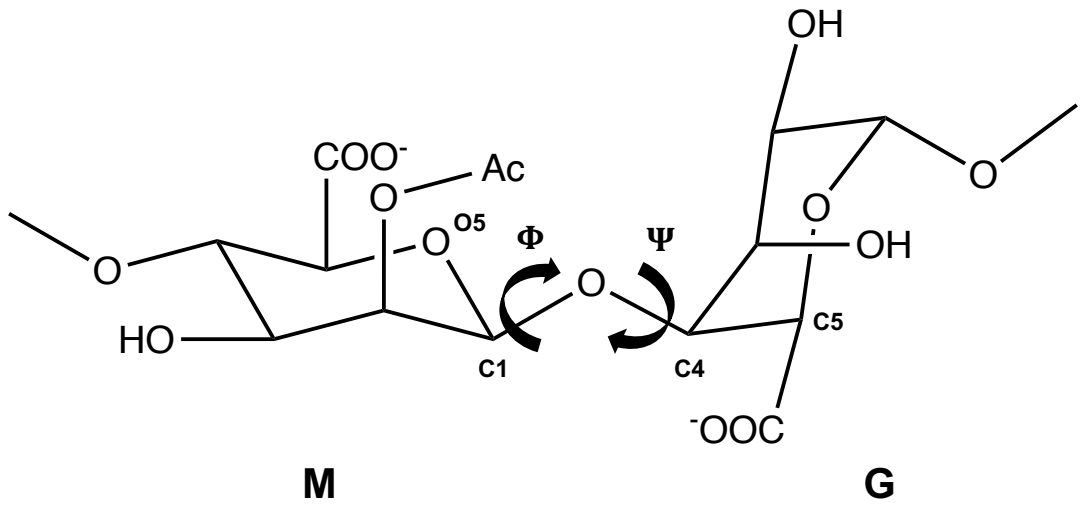


Figure 3.9: Geometry optimised antiparallel arrangement of two PolyM structures (top) and parallel arrangement of two PolyMG structures (bottom). Carbon atoms are shown grey, oxygen red and hydrogen pink. Hydrogen bonds between chains are shown as dashed blue lines and the Mulliken populations and lengths of these bonds are marked.

Table 3.5: Dihedral angles (ϕ, ψ) for the polyuronate systems. All angles are given in degrees (°). See **Fig 3.1** for labelling of the uronate units.

System	M1M2/M1G1	M2M3/G1M2	M3M4/M2G2
PolyM	(-76, -132)	(-97, -155)	(-85, -134)
PolyMG	(-68, -114)	(-95, -146)	(-74, -100)
PolyM$_{ap}$	Top chain: (-70, -125) Bottom chain: (-54, -130)	Top chain: (-70, -134) Bottom chain: (-94, -156)	Top chain: (-96, -151) Bottom chain: (-107, -123)
PolyMG$_{p}$	Top chain: (-87, -139) Bottom chain: (-67, -132)	Top chain: (-91, -137) Bottom chain: (-101, -144)	Top chain: (-74, -99) Bottom chain: (-47, -55)
Na-PolyM$_{ap}$	Top chain: (-61, -120) Bottom chain: (-81, -139)	Top chain: (-43, -121) Bottom chain: (-95, -136)	Top chain: (-78, -171) Bottom chain: (-111, -118)
Ca-PolyM$_{ap}$	Top chain: (-51, -99) Bottom chain: (-76, -121)	Top chain: (-51, -127) Bottom chain: (-93, -103)	Top chain: (-49, +139) Bottom chain: (-102, -108)
Mg-PolyM$_{ap}$	Top chain: (-49, -67) Bottom chain: (-81, -102)	Top chain: (-77, -125) Bottom chain: (-103, -142)	Top chain: (-72, +162) Bottom chain: (-85, -123)
Na-PolyMG$_{p}$	Top chain: (-85, -155) Bottom chain: (-49, -116)	Top chain: (-88, -140) Bottom chain: (-107, -143)	Top chain: (-113, -79) Bottom chain: (-48, -65)
Ca-PolyMG$_{p}$	Top chain: (-91, -127) Bottom chain: (-58, -122)	Top chain: (-94, -130) Bottom chain: (-89, -130)	Top chain: (-110, -78) Bottom chain: (-64, -44)
Mg-PolyMG$_{p}$	Top chain: (-104, -106) Bottom chain: (-50, -84)	Top chain: (-93, -128) Bottom chain: (-92, -137)	Top chain: (-123, -61) Bottom chain: (-55, -48)
Defect 1	Top chain: (-84, -147) Bottom chain: (-62, -131)	Top chain: (-97, -146) Bottom chain: (-96, -136)	Top chain: (-81, -98) Bottom chain: (-38, -85)
Defect 2	Top chain: (-97, -113) Bottom chain: (-59, -120)	Top chain: (-88, -133) Bottom chain: (-92, -127)	Top chain: (-111, -88) Bottom chain: (-49, -64)

The torsion angles (ϕ, ψ) for the PolyM$_{\text{ap}}$ and PolyMG$_{\text{p}}$ models, as defined in **Fig 3.10**, are given in **Table 3.5**. The torsion angles are also given for the 2-chain cation complexes, which are discussed in detail in **Section 3.4.3.2**.



$$\Phi = \text{O5-C1-O-C4}$$
$$\psi = \text{C1-O-C4-C5}$$

Figure 3.10: Definition of the torsion angles Φ and Ψ across an M-G junction in bacterial alginate. For reference, a torsion angle equal to ±0° indicates that O5 and C4, and C1 and C5, lie within the same plane; are synperiplanar or *cis*. A torsion angle equal to ±180° indicates these atoms are antiperiplanar or *trans*. Note, definition of Φ and Ψ is the same for an M-M junction.

It is firstly important to note that the geometry optimised 2-chain arrangements in space was studied using the protonated form of each polyuronate. The dihedral angles for both these arrangements correlate well, within 18°, with the theoretically derived dihedral angles present within the minimum energy conformations of poly-β-D-mannuronate and poly-α-L-guluronate systems also in their protonated forms [295]. Constrained conformational searching about the central glycosidic linkage in the structurally unique M-M, M-G, G-M and G-G disaccharides of algal alginates, performed using DFT/B3LYP and Hartree-Fock theoretical approaches, has identified multiple different minima corresponding to different values of (ϕ, ψ) [140, 143]. DFT/B3LYP calculations reported (ϕ, ψ) angles for the minimum energy conformations of M-M, M-G, G-M and G-G disaccharides of (312°, 92°), (57°, 248°), (269°, 202°) and (270°, 203°) respectively [143]. Hartree-Fock simulations produced (ϕ, ψ) angles for the minimum energy conformations of M-M and G-G disaccharides of (274°, 344°) and (305°, 292°) respectively [140]. It should be noted here, that the torsion angles reported in these works [140, 143] are quoted as obtuse angles, ranging from 0° to 360°. Standard nomenclature for reporting torsion profiles is in the acute form, ranging from −180° to +180°. Upon conversion to the acute form, DFT/B3LYP calculations reported (ϕ, ψ) angles of (−48°, 92°), (57°, −112°), (−91°, −158°) and (−90°, −157°) for the minimum energy conformations of

M-M, M-G, G-M and G-G disaccharides respectively [143]. Additionally, upon conversion to the acute form, Hartree-Fock simulations produced (ϕ, ψ) angles of $(-86°, -16°)$ and $(-55°, -68°)$ for the minimum energy conformations of M-M and G-G disaccharides respectively [140].

The dihedral angles reported in these theoretical works differ from one another and also differ from the dihedral angles present in the bacterial alginate models presented here at both the 1-chain and 2-chain level. Most significantly, the torsion profiles of the bacterial alginate models in this work encompass only negative values of both ϕ and ψ. This is different to the DFT/B3LYP results in the work of Bekri *et al.*, which report positive ϕ angles within M-M junctions and positive ψ angles within M-G junctions [143]. It must be noted that, the bacterial alginate molecular models in this work were constructed such that all atoms were assigned complete freedom and the ground-state structures were not identified by constrained conformational searching. Moreover, the bacterial alginate models possess sterically bulky acetyl groups that affect the axial conformational flexibility. In line with this, the bacterial alginate models are at the quadramer scale, not the disaccharide scale, which will also affect axial conformational flexibility. For example, torsion rotation about one M-M or M-G junction may be hindered because rotation about a neighbouring junction is also hindered. Finally, it is also critical to note that in these works (ϕ, ψ) rotations are not structurally defined [140, 143], meaning the authors in these works do not define which atoms are included the definition of ϕ and ψ. As such, deviation in torsion profiles may arise from (ϕ, ψ) angles calculated under different definitions. This may also explain why the angles in these works also differ from one another [140, 143].

Although the above discussion lists reasons why the the (ϕ, ψ) angles for the bacterial alginate models do not correlate highly with the M-M, M-G, G-M and G-G (ϕ, ψ) angles computed using the aforementioned DFT and Hartree-Fock methods, it is nevertheless a useful measure to compare polyuronate configurations in this study. For example, at a 1-chain level (PolyM and PolyMG systems), the largest similarity is the occurrence of oppositely displaced carboxyl groups across the M-M and M-G junctions, which is a structural feature also present in the minimum energy M-M, M-G and G-G disaccharides [140, 143]. The PolyM_{ap} and PolyMG_{p} models were carried forward to study cation cross-linking of EPS chains.

3.4.3 2-chain cross-linking

3.4.3.1 Thermodynamic stability trend

The most thermodynamically favourable two chain arrangements (PolyM_{ap} and PolyMG_{p}) were used to study EPS-ion aggregation. Eight Na^{+}, four Ca^{2+} and four Mg^{2+} ions were initially positioned in-between the two chains, in the vicinity of the charge neutralising

carboxylate group(s) and where as many oxygen contacts could be made as possible. These positions were determined by reference to the most thermodynamically favourable binding positions observed at the 1-chain level (see **Section 3.4.1**). Carboxylic acid groups were deprotonated to ensure charge balance, and the number of cations included represented a fully charge-saturated system with respect to the carboxylic acid groups. These structures were then subject to a geometry optimisation.

The formation energies (**Equation 3.4.1**) for the sputum ion cross-linked $PolyM_{ap}$ and $PolyMG_{p}$ systems are given in **Table 3.6**.

Table 3.6: Formation energies (eV) for the cation cross-linked $PolyM_{ap}$ and $PolyMG_{p}$ complexes.

System	E_f (eV)		
	Na^+	Mg^{2+}	Ca^{2+}
$PolyM_{ap}$	+0.62	-4.27	-9.53
$PolyMG_{p}$	-1.53	-5.95	-10.01

The negative formation energies indicate that it is thermodynamically favourable for sputum ions to induce the aggregation of the bacterial alginate structures. The thermodynamic stability trend follows the order $Ca^{2+} > Mg^{2+} > Na^+$ for both the $PolyM_{ap}$ and $PolyMG_{p}$ systems. This is a stability trend that parallels experimentally determined metal-alginate affinities [296] and, in addition, matches the stability trend of acetylated bacterial alginates from ion chromatography experiments [297]. $PolyMG_{p}$ complexes are more thermodynamically favourable compared to the $PolyM_{ap}$ complexes (by 0.5 to 1.5 eV) for each sputum ion. This highlights the preference for the sputum ions to favour complexation within M-G junctions, aligning with ^{13}C-NMR observations on the formation of cation chelate complexes within *P. aeruginosa* biofilms [130], theoretical investigations based on molecular mechanics [295] and is consistent with the understanding that elevated guluronate levels increase overall metal ion affinity in alginates [298]. Calcium ions form the most stable cross-links for both bacterial alginate structures, forming more stable complexes by approximately 9 and 5 eV relative to sodium and magnesium ions respectively. Calcium ions forming highly stable cross-linked bacterial alginate architectures are to be expected, as these ions have been implicated in inducing an increase in EPS polymer network stability in *P. aeruginosa* biofilms [123]. The stability trend also provides a thermodynamic rational for the increased gelation capability of calcium ions relative to sodium (and magnesium) ions and matches predictions made by previous theoretical models, based on molecular dynamics simulations, studying sodium-calcium ion competition in algal alginate trimers [139, 299].

A negative formation energy suggests the electrostatically stabilised (ionic cross-linked) EPS scaffold are more stable relative to EPS scaffolds bound only through hydrogen bonding interactions (**Fig 3.9**). This corroborates observations from rheological measurements that the upon the addition of cations, such as calcium, to the mucoid *P. aeruginosa* EPS matrix, there is a notable gain in biofilm mechanical stability relative to when cations were absent from the EPS matrix [123]. This implies that coulomb interactions become the dominant stabilising interaction in *P. aeruginosa* EPS matrices upon exposure to cations [123]. The most unstable system found is the Na^+ - $PolyM_{ap}$ complex. Sodium ions fail to induce stable cross-linking and aggregation of bacterial alginate structures possessing solely acetylated M-blocks. In this structure, it is more stable for bacterial alginate (PolyM) chains to associate by hydrogen bonding than to associate by cation-mediated ionic bonding. This correlates well with rheological experiments on the *P. aeruginosa* EPS which, in the presence of sodium, weakly aggregated networks are formed held together by temporary, transient, ionic cross-links and stabilised more-so through entanglements (of vdW and hydrogen bonds) than coulombic interactions [122]. The overall stability does not increase as the ionic radius of the cation decreases and the charge density on the cation increases. This suggests that the stability of the resultant cation chelate complex is not solely dependent on the strength of the ionic O-ion bonds. The thermodynamic predictions of these models, therefore, contrast with the predictions made recently, using theoretical models based on DFT. In these calculations, magnesium ions form the most stable complexes (induce more stable cross-links) compared to calcium ions in complexes with M-M, M-G, G-M and G-G algal alginate disaccharides [143]. Consequently, the bacterial alginate models developed here are closer to the predicted thermodynamic stability of the *in vitro* bacterial alginate structures.

3.4.3.2 Geometry of the chelation site

The sputum ion cross-linked $PolyM_{ap}$ and $PolyMG_p$ complexes (**Table 3.6**) are shown in **Figs 3.12 - 3.14**. All sputum ions bind preferentially in positions where they can maximise electronegative oxygen contacts, which is also observed in the crystal structures of calcium-β-D-linked pyranose and calcium-β-D-linked galacturonate complexes (**Figure 3.8**) [293, 294]. The complexation of calcium and magnesium ions between bacterial alginate chains, in both $PolyM_{ap}$ and $PolyMG_p$ systems, encompass two COO^- groups per ion, aligning with thermogravimetric measurements on divalent ion algal alginate complexes [300]. In the 2-chain divalent ion complexes, both mono- and bidentate COO^- binding modes were observed, but unlike the single chain complexes, where the bidentate coordination mode occurs for CN < 5 environments, the bidentate mode now occurs in CN > 5 environments. As such, the preference for monodentate or bidentate COO^- binding is no longer determined by the degree of saturation in the chelate pocket when the ions are complexed in-between two EPS chains.

The divalent ions are complexed in chelate pockets formed in-between adjacent chains and share similarities with the classical *egg-box* model of divalent ion complexation by algal alginates [301]. However, there are deviations from the model. Specifically, the egg-box model dictates that each divalent ion is bound by 4 uronate residues via 10 ionic contacts. Two uronate residues reside on the bottom chain and the other two uronate residues reside on the top chain. Each polyuronate chain contributes five ionic contacts to the central cation [301]. This is illustrated in **Fig 3.11** and the differences between our 2-chain cation cross-linked systems (**Figs 3.12 - 3.14**) and the egg-box model are discussed below.

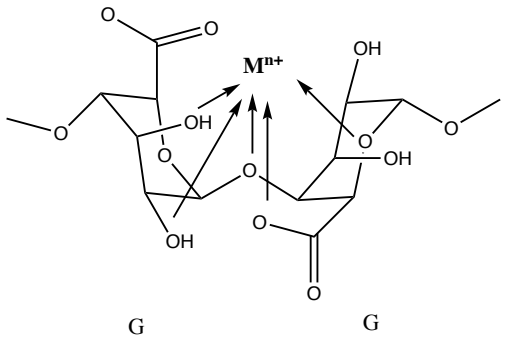


Figure 3.11: Illustration of half of the egg-box model, showing five contacts out of a possible ten [301]. In this illustration, only the bottom-half of the egg-box is displayed, notably, two uronate residues and five ionic bonds to the central ion.

In the 2-chain models (**Figs 3.12 - 3.14**), the coordination geometries display two to three uronate residues per divalent cation, originating from the bottom and top EPS chains, and CNs of up to six. This is below the egg-box predicted four uronate residues and ten ionic contacts per cation [301]. In a previous theoretical molecular mechanics investigation, a calcium probe was migrated over the vdW surfaces of poly-β-D-mannuronate and poly-α-L-guluronate hexamers to search for the different possible stable calcium chelation positions that could exist within a 15 Kcal/mol energetic window. The only requirement for stable calcium binding was a suitable orientation of oxygen atoms [295]. Similarly, deviations from the egg-box model in calcium cross-linked algal alginate networks have been observed theoretically, where MD simulations have shown that 2-chain complexation about calcium ions occurs solely with the COO^- groups [134, 135] and through perpendicular chain conformations [138, 141]. Overall, this introduces two alternative perspectives to cation-alginate complexation. The first is that bacterial alginates are capable of complexing into stable, low energy, cation chelate complexes with a reduced uronate ionic contribution - namely, the number of uronate residues and ionic bonds per cation fall below the egg-box prediction. The second is that the classical egg-box model is perhaps out-dated as an array

of alternative stable algal alginate disaccharide cation chelate complexes, which also have a reduced uronate contribution, have been predicted theoretically [134, 135, 138, 141, 295].

The charge neutralising stoichiometry of the sodium complexes gives rise to a very crowded 2-chain intra-complex space (**Fig 3.14**) and the sodium ions do not present defined inter-chain binding sites. As such, single COO^- groups are involved in monodentate binding to multiple sodium ions. The cooperative effect, the principle that the polyuronate backbone offers defined cation inter-chain binding positions, distributed in regular arrays, is negligible in the sodium cross-linked bacterial alginate systems and this contributes to its weaker stability (**Table 3.6**). Specifically, the instability of the Na^+ - $PolyM_{ap}$ system arises when two sodium ions (Na3 and Na7, **Fig 3.14a**) bind to outward facing COO^- groups (O17 and O18, O23 and O24) and do not contribute to aggregation of the two PolyM chains. Furthermore, Na8 fails to saturate a COO^- group, instead binding to terminal acetyl and hydroxyl groups (O25, O30 and O53) at the periphery of the terminal residue. The aggregation of poly-β-D-mannuronate decamers, induced by sodium ions, studied using MD, also captured sodium ions failing to bind to COO^- groups, forming only transient interactions with other oxygen functional groups and sodium induced cross-linking was also not observed [139].

In both the $PolyM_{ap}$ and $PolyMG_p$ systems, calcium and magnesium ions bind in-between chains forming ionic 2-chain associated complexes, displaying CNs of up to six. These CNs match exactly with previous DFT studies on divalent ion complexation by algal alginate [140, 141]. The divalent ion CNs are not increased compared to the 1-chain single ion complexes (**Figs 3.5 - 3.7**), which further underscores that 1-chain systems can present suitable sites for stable complexation. The maximum calcium ion CN of six is greater than that of the magnesium ion (CN=5), despite the higher charge density of the magnesium ion. Lower CNs about magnesium ions, compared to calcium ions, has also been seen in previous DFT works on M-G divalent cation complexes [143].

The lower CN is due to the smaller ionic radius of Mg^{2+} compared to Ca^{2+} and the smaller the radius, the less able the ion is to facilitate bonding to additional oxygen donors reducing the interaction [145]. As a consequence, the smaller ion can only facilitate ionic binding to oxygen donors in the immediate vicinity. For bonding at larger distances to occur, other structural rearrangements would be required, which may be thermodynamically unfavourable. In contrast, a larger ion could bind to oxygen donors at greater distances, without the requirement for large-scale global conformational change. For example, comparing Ca1 and Mg1 in the $PolyM_{ap}$ systems, it is evident that these adopt identical coordination environments apart from the interaction with O9, which is missing in the magnesium case. Similarly, Mg2 and Mg3 display highly similar coordination environments to Ca2 and Ca3

in the PolyMG$_p$ structure, however, the Mg2-O48 and Mg3-O49 interactions are absent. Subsequently, there are fewer ionic cross-links in the magnesium ion complexes, giving a weaker gelation capability relative to calcium ions [302]. It is worth noting that in a recent theoretical modelling investigation, DFT was used to predict more stable complexation of magnesium, compared to calcium, by algal alginate M-G disaccharides [143]. This highlights the importance of employing a model constructed from multiple M-M/M-G junctions (from opposing chains) to predict the effect that ionic radii has on uronate oxygen accessibility and to capture the correct gelation trends.

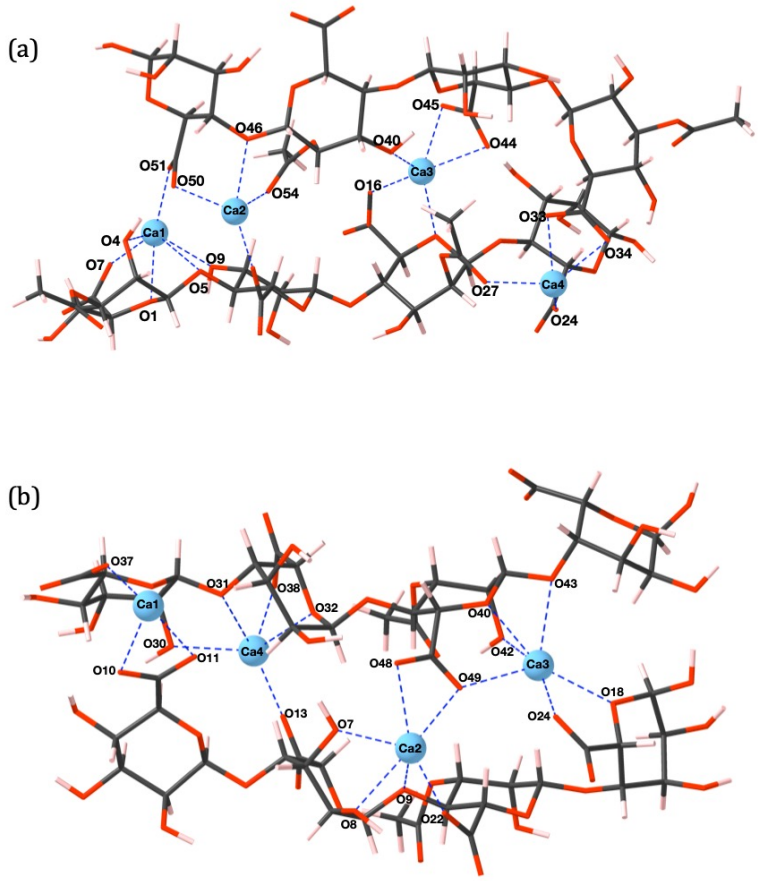


Figure 3.12: (a) PolyM$_{ap}$ (b) PolyMG$_p$ 2-chain calcium complexes. Carbon atoms are shown in grey, oxygen in red, calcium as blue spheres and hydrogen in pink. Oxygen atoms and cations involved in ionic bonding are labelled and the cation-oxygen bonds are shown with dashed blue lines.

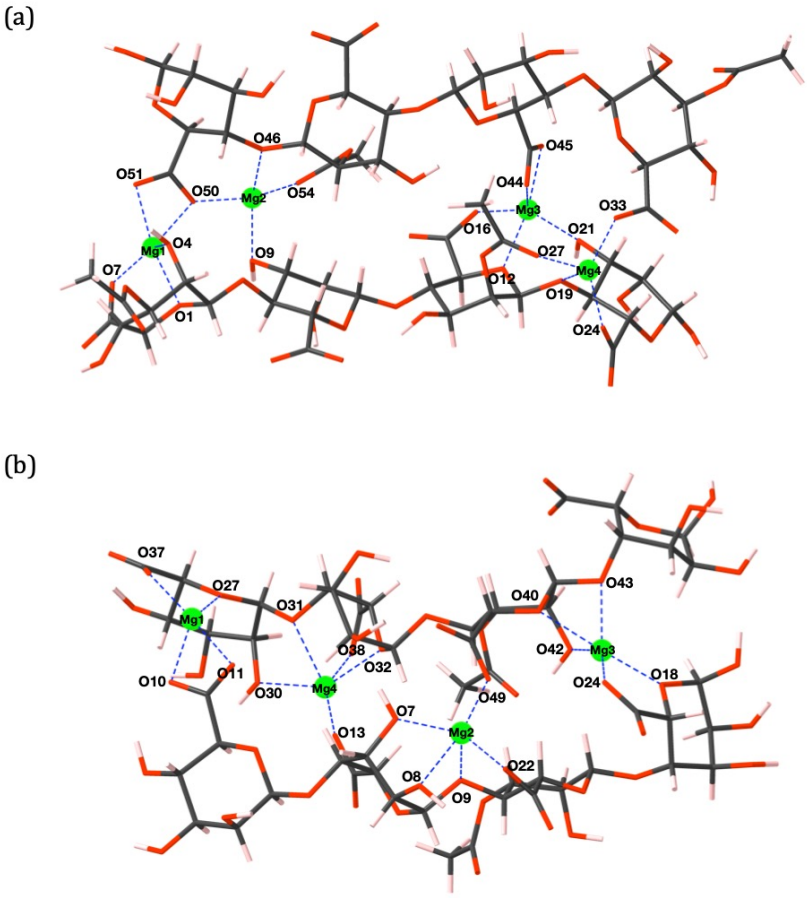


Figure 3.13: (a) PolyM$_{ap}$ (b) PolyMG$_p$ 2-chain magnesium complexes. Carbon atoms are shown in grey, oxygen in red, magnesium as green spheres and hydrogen in pink. Oxygen atoms and cations involved in ionic bonding are labelled and the cation-oxygen bonds are shown with dashed blue lines.

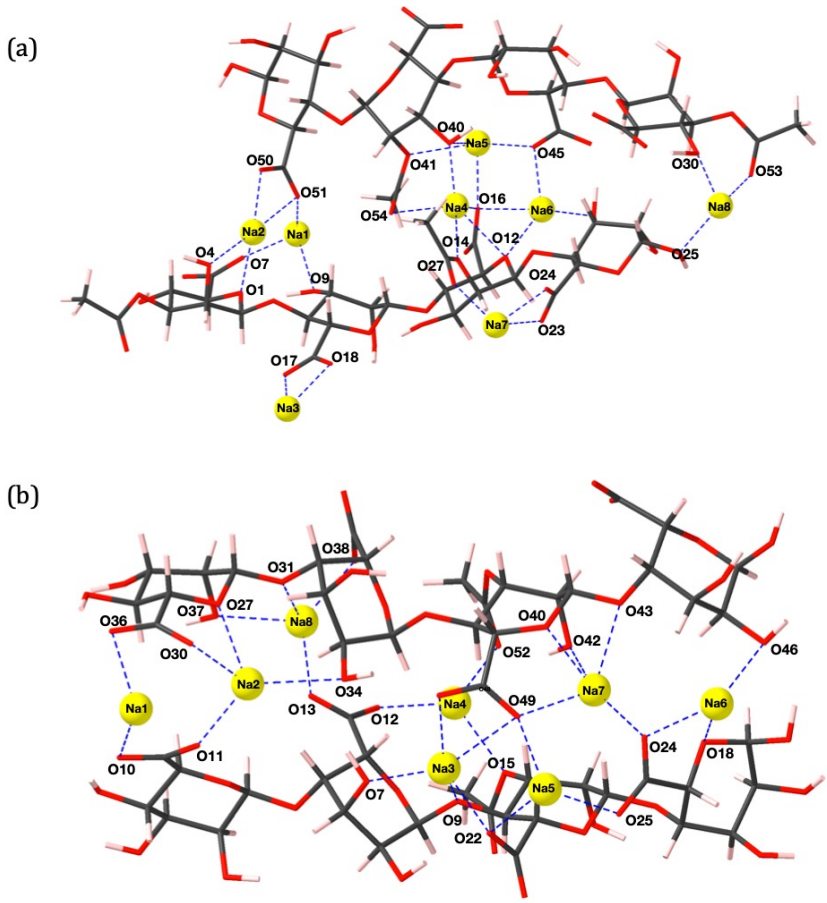


Figure 3.14: (a) PolyM$_{ap}$ (b) PolyMG$_p$ 2-chain sodium complexes. Carbon atoms are shown in grey, oxygen in red, sodium as yellow spheres and hydrogen in pink. Oxygen atoms and cations involved in ionic bonding are labelled and the cation-oxygen bonds are shown with dashed blue lines.

Table 3.7: Bond populations and lengths for the Na^+-oxygen contacts in the Na^+-PolyM_{ap} complex.

Sodium PolyM_{ap} complex		
Bond	Population (|e|)	Length (Å)
Na1-O7 (COO^-)	0.13	2.15
Na1-O9 (OH)	0.10	2.19
Na1-O51 (COO^-)	0.05	2.38
Na2-O1 (Ring O)	0.07	2.46
Na2-O4 (OH)	0.05	2.36
Na2-O50 (COO^-)	0.02	2.25
Na2-O51(COO^-)	0.03	2.46
Na3-O17 (COO^-)	0.04	2.24
Na3-O18 (COO^-)	0.05	2.24
Na4-O12 (Ring O)	0.07	2.63
Na4-O14 (Acetyl)	0.11	2.50
Na4-O16 (COO^-)	0.09	2.43
Na4-O40 (OH)	0.12	2.26
Na4-O54 (Acetyl)	0.06	2.63
Na5-O16 (COO^-)	0.07	2.34
Na5-O40 (OH)	0.01	2.97
Na5-O41 (Acetyl)	0.09	2.53
Na5-O45 (COO^-)	0.11	2.27
Na6-O12 (Ring O)	0.06	2.31
Na6-O16 (COO^-)	0.03	2.37
Na6-O21 (OH)	0.04	2.30
Na6-O45 (COO^-)	0.07	2.26
Na7-O23 (COO^-)	0.05	2.34
Na7-O24 (COO^-)	0.04	2.23
Na7-O27 (Acetyl)	0.06	2.29
Na8-O25 (OH)	0.05	2.33
Na8-O30 (OH)	0.14	2.16
Na8-O53 (Acetyl)	0.06	2.28

Table 3.8: Bond populations and lengths for the Na^+-oxygen contacts in the Na^+-PolyMG$_p$ complex.

Sodium PolyMG$_p$ complex		
Bond	Population (|e|)	Length (Å)
Na1-O10 (COO^-)	0.03	2.34
Na1-O36 (COO^-)	0.02	2.37
Na2-O11 (COO^-)	0.12	2.21
Na2-O27 (Ring O)	0.11	2.63
Na2-O34 (OH)	0.08	2.68
Na2-O37 (COO^-)	0.11	2.25
Na3-O7 (OH)	0.06	2.22
Na3-O9 (Glycosidic O)	0.12	2.53
Na3-O22 (COO^-)	0.11	2.25
Na3-O48 (COO^-)	0.05	2.30
Na4-O12 (COO^-)	0.13	2.16
Na4-O15 (Acetyl)	0.09	2.38
Na4-O52 (Acetyl)	0.08	2.28
Na5-O22 (COO^-)	0.07	2.29
Na5-O25 (COO^-)	0.09	2.30
Na5-O49 (COO^-)	0.06	2.32
Na6-O18 (Ring O)	0.08	2.44
Na6-O24 (COO^-)	0.08	2.21
Na6-O46 (OH)	0.06	2.37
Na7-O24 (COO^-)	0.08	2.22
Na7-O40 (Ring O)	0.05	2.29
Na7-O42 (OH)	0.11	2.43
Na7-O43 (Glycosidic O)	0.10	2.52
Na7-O49 (COO^-)	0.10	2.39
Na8-O13 (COO^-)	0.06	2.22
Na8-O30 (OH)	0.07	2.43
Na8-O31 (Glycosidic O)	0.08	2.49
Na8-O38 (COO^-)	0.10	2.24

Table 3.9: Bond populations and lengths for the Ca^{2+}-oxygen contacts in the Ca^{2+}-PolyM_{ap} complex.

Calcium PolyM_{ap} complex		
Bond	Population (|e|)	Length (Å)
Ca1-O1 (Ring O)	0.08	2.39
Ca1-O4 (OH)	0.09	2.48
Ca1-O5 (Glycosidic O)	0.04	2.68
Ca1-O7 (COO^-)	0.19	2.17
Ca1-O9 (OH)	0.09	2.46
Ca1-O51 (COO^-)	0.11	2.42
Ca2-O18 (COO^-)	0.21	2.15
Ca2-O46 (Glycosidic O)	0.12	2.46
Ca2-O50 (COO^-)	0.13	2.24
Ca2-O54 (Acetyl)	0.11	2.33
Ca3-O12 (Ring O)	0.11	2.41
Ca3-O16 (COO^-)	0.18	2.18
Ca3-O40 (OH)	0.09	2.46
Ca3-O44 (COO^-)	0.06	2.42
Ca3-O45 (COO^-)	0.09	2.34
Ca4-O24 (COO^-)	0.22	2.10
Ca4-O27 (Acetyl)	0.10	2.31
Ca4-O33 (COO^-)	0.08	2.32
Ca4-O34 (COO^-)	0.08	2.34

Table 3.10: Bond populations and lengths for the Ca^{2+}-oxygen contacts in the Ca^{2+}-PolyMG$_p$ complex.

Calcium PolyMG$_p$ complex		
Bond	Population (|e|)	Length (Å)
Ca1-O10 (COO^-)	0.10	2.23
Ca1-O11 (COO^-)	0.09	2.25
Ca1-O37 (COO^-)	0.20	2.08
Ca2-O7 (OH)	0.10	2.37
Ca2-O8 (OH)	0.06	2.46
Ca2-O9 (Glycosidic O)	0.10	2.46
Ca2-O22 (COO^-)	0.19	2.20
Ca2-O48 (COO^-)	0.05	2.52
Ca2-O49 (COO^-)	0.06	2.42
Ca3-O18 (Ring O)	0.12	2.52
Ca3-O24 (COO-)	0.21	2.13
Ca3-O40 (Ring O)	0.08	2.41
Ca3-O42 (OH)	0.10	2.41
Ca3-O43 (Glycosidic O)	0.08	2.59
Ca3-O49 (COO^-)	0.09	2.53
Ca4-O13 (COO^-)	0.24	2.11
Ca4-O30 (OH)	0.10	2.41
Ca4-O31 (Glycosidic O)	0.08	2.56
Ca4-O32 (Ring O)	0.09	2.44
Ca4-O38 (COO^-)	0.16	2.20

Table 3.11: Bond populations and lengths for the Mg^{2+}-oxygen contacts in the Mg^{2+}-PolyM_{ap} complex.

Magnesium PolyM_{ap} complex		
Bond	Population (|e|)	Length (Å)
Mg1-O1 (Ring O)	0.13	2.08
Mg1-O4 (OH)	0.16	2.14
Mg1-O7 (COO^-)	0.19	1.92
Mg1-O50 (COO^-)	0.06	2.14
Mg1-O51 (COO^-)	0.02	2.10
Mg2-O9 (OH)	0.15	2.04
Mg2-O46 (Glycosidic O)	0.21	2.09
Mg2-O50 (COO^-)	0.17	1.94
Mg2-O54 (Acetyl)	0.18	1.93
Mg3-O12 (Ring O)	0.12	2.12
Mg3-O16 (COO^-)	0.18	1.94
Mg3-O21 (OH)	0.10	2.06
Mg-O44 (COO^-)	0.09	2.04
Mg3-O45 (COO^-)	0.07	2.08
Mg4-O19 (Glycosidic O)	0.13	2.21
Mg4-O24 (COO^-)	0.21	1.91
Mg4-O27 (Acetyl)	0.12	1.96
Mg4-O33 (COO^-)	0.17	1.91

Table 3.12: Bond populations and lengths for the Mg^{2+}-oxygen contacts in the Mg^{2+}-PolyMG$_p$ complex.

Magnesium PolyMG$_p$ complex		
Bond	Population (|e|)	Length (Å)
Mg1-O10 (COO^-)	0.07	2.02
Mg1-O11 (COO^-)	0.09	2.00
Mg1-O27 (Ring O)	0.13	2.05
Mg1-O37 (COO^-)	0.16	1.91
Mg2-O7 (OH)	0.13	2.10
Mg2-O8 (OH)	0.07	2.17
Mg2-O9 (Glycosidic O)	0.17	2.16
Mg2-O22 (COO^-)	0.21	1.94
Mg2-O49 (COO^-)	0.15	1.96
Mg3-O18 (Ring O)	0.19	2.12
Mg3-O24 (COO^-)	0.22	1.89
Mg3-O40 (Ring O)	0.11	2.26
Mg3-O42 (OH)	0.16	2.04
Mg3-O43 (Glycosidic O)	0.15	2.12
Mg4-O13 (COO^-)	0.22	1.87
Mg4-O30 (OH)	0.13	2.15
Mg4-O31 (Glycosidic O)	0.15	2.32
Mg4-O32 (Ring O)	0.14	2.21
Mg4-O38 (COO^-)	0.21	1.96

3.4.3.3 Bonding analysis

Bond lengths (Å) and Mulliken populations (|e|) for all cation-oxygen contacts in all sputum ion cross-linked bacterial alginate structures are given in **Tables 3.7 - 3.12**. These tables correspond to the first theoretically derived dataset of O-ion interaction lengths (Å) and populations (|e|) for cation cross-linked mucoid *P. aeruginosa* EPS matrices. This geometric dataset can assist in future crystallographic and NMR structure solutions of mucoid *P. aeruginosa* EPS matrices isolated experimentally or *ex vivo* and also, for the first time, gives contact distances that portray the scale of the inter-molecular separations between EPS components of the mucoid ECM in the lungs of the CF patients. The latter accentuates the magnitude of the impermeability of EPS matrices which is facilitated through tight ionic cross-linking at distances less than 3 Å.

By considering coordination number, electronic spin state and cation-anion interatomic separation, Shannon and Prewitt-derived ionic radii are appropriate for systems where bonding of cations occurs to oxygen, across an array of different coordination environments [303]. As such, the trend in ionic radii for the sputum ions considered in this work follows the order

Na^+ (1.02 Å) ≈ Ca^{2+} (1.00 Å) > Mg^{2+} (0.72 Å). As expected, across both bacterial alginate systems, the average Na^+-oxygen bond length (2.36 Å) ≈ average Ca^{2+}-oxygen bond length (2.35 Å) > average Mg^{2+}-oxygen bond length (2.03 Å). These agree well (within 0.1 Å) with previous theoretical predictions, based on DFT, on algal alginate disaccharide complexes with sodium, calcium and magnesium ions [140, 143] and, specifically, the Ca^{2+}-oxygen bond lengths agree well (within 0.06 Å) with observed bond lengths in calcium-β-D-linked pyranose complexes [293].

All Mulliken bond populations for the cation-oxygen contacts are < 0.3 |e|, which indicates that these interactions are ionic in nature. Furthermore, cation-oxygen contacts with COO^- groups are shorter and occur with larger populations (**Tables 3.7 - 3.12**) for each sputum ion in each cross-linked bacterial alginate system relative to all other O-ion interactions. The average Na^+-COO^- populations (PolyM$_{ap}$ 0.06 |e|, PolyMG$_p$ 0.08 |e|) are smaller compared with the average Ca^{2+}-COO^- (PolyM$_{ap}$ 0.14 |e|, PolyMG$_p$ 0.14 |e|) and Mg^{2+}-COO^- (PolyM$_{ap}$ 0.13 |e|, PolyMG$_p$ 0.17 |e|) populations, indicating that calcium and magnesium ions bind more stably to COO^- groups compared to sodium ions. This is to be expected following charge density arguments and complements computed ion-oxygen radial distribution functions which show, in classical MD simulations of polyguluronate decamers binding calcium and sodium ions, that the density of the Ca^{2+}-COO^- contacts is six times greater than the density of the Na^+-COO^- contacts [139].

Alongside binding within M-G junctions, there is a clear stabilising effect of binding a COO^- group. Within the sodium and magnesium cross-linked structures, fewer COO^- groups are saturated in the PolyM$_{ap}$ system compared to the PolyMG$_p$ system. As a result, the stability difference (**Table 3.6**) between the two systems is approximately 2.2 eV for sodium and 1.7 eV for magnesium, with PolyMG$_p$ always being the more stable. In the calcium case, where the same number of COO^- groups are saturated in both cross-linked bacterial alginate structures, the stability difference is lower (approximately 0.5 eV). The affinity for, and stabilisation effect of, the COO^- group underlines the electrostatic nature of cation complexation in these bacterial alginate systems. Notably, cation binding at COO^- sites maximises electrostatic interactions leading to greater overall stabilisation of the resultant complex. However, the average ion-COO^- bond lengths and populations, for calcium and magnesium ions in both cross-linked systems, indicate the stronger association between Mg^{2+} and the COO^- group which, under purely electrostatic arguments, should give rise to more stable complexes. However, this is not what it is observed, and calcium ions are responsible for giving the most stable cross-linked complexes. This reinforces the observation that, although electrostatics drive cation accommodation, they are not the determining factor behind cation cross-linked 2-chain stability.

As discussed, the stoichiometry of a bacterial alginate 2-chain system, charge saturated by sodium ions, gives rise to ill-defined inter-chain binding positions. The consequence is that single COO^- groups are involved in coordination to multiple sodium ions, giving rise to a weak ionic architecture with Na^+-COO^- populations as low as 0.03 |e|. Furthermore, ill-defined inter-chain binding positions result in sodium ions binding externally, not in-between chains, and subsequently failing to establish cross-links between chains (**Fig 3.14a**). The divalent calcium and magnesium ions display more well-defined inter-chain binding sites, which arise because these ions are able to more significantly induce conformational change across the polyuronate backbone (relative to 2-chains in the absence of cations; **Fig 3.9**).

Upon complexing calcium and magnesium ions, global acute torsion changes ((ϕ, ψ); **Table 3.5**) reach +36° and +49° for calcium ((+36, +21)°) and magnesium ((+49, +38)°) respectively in the PolyMG$_p$ system, which are be ascribed to ϕ rotation about the top chain M2-G2 junction within Ca-PolyMG$_p$ and Mg-PolyMG$_p$ systems respectively. In contrast, these angles reach −70° and and −47° for calcium ((+47, −70)°) and magnesium ((+24, −47)°) respectively in the PolyM$_{ap}$ system. These particular large torsion changes measured in the PolyM$_{ap}$ system can be ascribed to the ψ reorientation of the glycosidic linkage at O46 (top chain M3-M4 junction) to face Ca2 and Mg2, establishing an ionic contact (**Figs 3.12a** and **3.13a**). There is an additional large global torsion change of +58° which is ascribed to the ψ reorientation across the top chain M1-M2 junction in the Mg-PolyM$_{ap}$ system ((+21, +58)°). This particular torsion change results from the reorientation of the carboxylate group O33 to face Mg4 bound within the 2-chain intra-complex space (**Fig 3.13a**). Accordingly, smaller torsion changes in the PolyMG$_p$ divalent ion 2-chain complexes arise because M-G junctions offer oxygen functionality whose spatial arrangement is suitable for saturating the coordination environment of the cation.

In contrast, complexing sodium ions leads to torsion changes which do not surpass 39° in the PolyM$_{ap}$ and PolyMG$_p$ 2-chain systems. Polyuronate accommodation is, therefore, improved by calcium and magnesium ions, which are able to induce conformational change to maximise the geometry of their respective chelation sites. Theoretical molecular mechanics studies, investigating the conformational flexibility of acidic polyuronates upon complexing sodium and calcium ions, drew similar conclusions [295]. The ability to induce conformational change, to create well defined binding sites, is a clear contributor to 2-chain complex stability. However, a balance exists, an ion must alter the conformation of the polyuronate backbone, but larger torsion changes lead to less stable complexes. This offers an explanation for the differences in stability between complexing divalent ions within M-M and M-G junctions (**Table 3.6**). Again, as previously discussed, capturing this steric effect is only possible when

employing large molecular models with multiple M-M and M-G junctions from opposing chains.

3.4.3.4 Physiological softening

The formation of calcium chelate complexes in *P. aeruginosa* EPS matrices has been observed in ^{13}C-NMR to be reversible when exposed to a cation exchange resin [130] and, in addition, destabilisation of calcium alginate gels is observed under exposure to (excess) sodium ions [304]. Given the ubiquity of biological sodium and its elevation in the CF lung, the calcium cross-linked PolyMG$_p$ system, the most thermodynamically favourable cation cross-linked EPS scaffold, was used as a model system to study calcium-for-sodium cation exchange. This choice allowed for the assessment of how these matrices soften under physiological conditions.

To this end, two substitution patterns (**Fig 3.15**), that maintained charge balance, were tested in the calcium cross-linked PolyMG$_p$ system:

- Defect 1: four calcium ions were directly substituted for four sodium ions at native calcium sites, with the addition of four protons to four COO^- groups.
- Defect 2: the four calcium ions were substituted for eight sodium ions, positioned in-between chains along the length of the chain axis.

Both these defect structures were subject to a geometry optimisation and the formation energies of each sodium substituted structure was evaluated using **Equations 3.4.3** and **3.4.4** for defect 1 and defect 2 respectively.

$$E_f = E_{final} - (E_{initial} + 4\mu_{Na} - 4\mu_{Ca} + 4\mu_H) \tag{3.4.3}$$

$$E_f = E_{final} - (E_{initial} + 8\mu_{Na} - 4\mu_{Ca}) \tag{3.4.4}$$

where E_{final} is the final energy of the sodium substituted structure and $E_{initial}$ is the energy of the calcium cross-linked PolyMG$_p$ structure. The formation energies for defect 1 and defect 2 are give in **Table 3.13**.

Table 3.13: Formation energies (eV) for the calcium-for-sodium substitutions in the Ca^{2+}-PolyMG$_p$ system.

Sodium substitution	E_f (eV)
Defect 1	+9.00
Defect 2	+8.88

Interestingly, both substitution patterns were found to be equally as unstable. This shows that the calcium chelate complexes are very stable, thus, it is thermodynamically unstable to liberate these ions from these binding positions regardless of substitution mechanism.

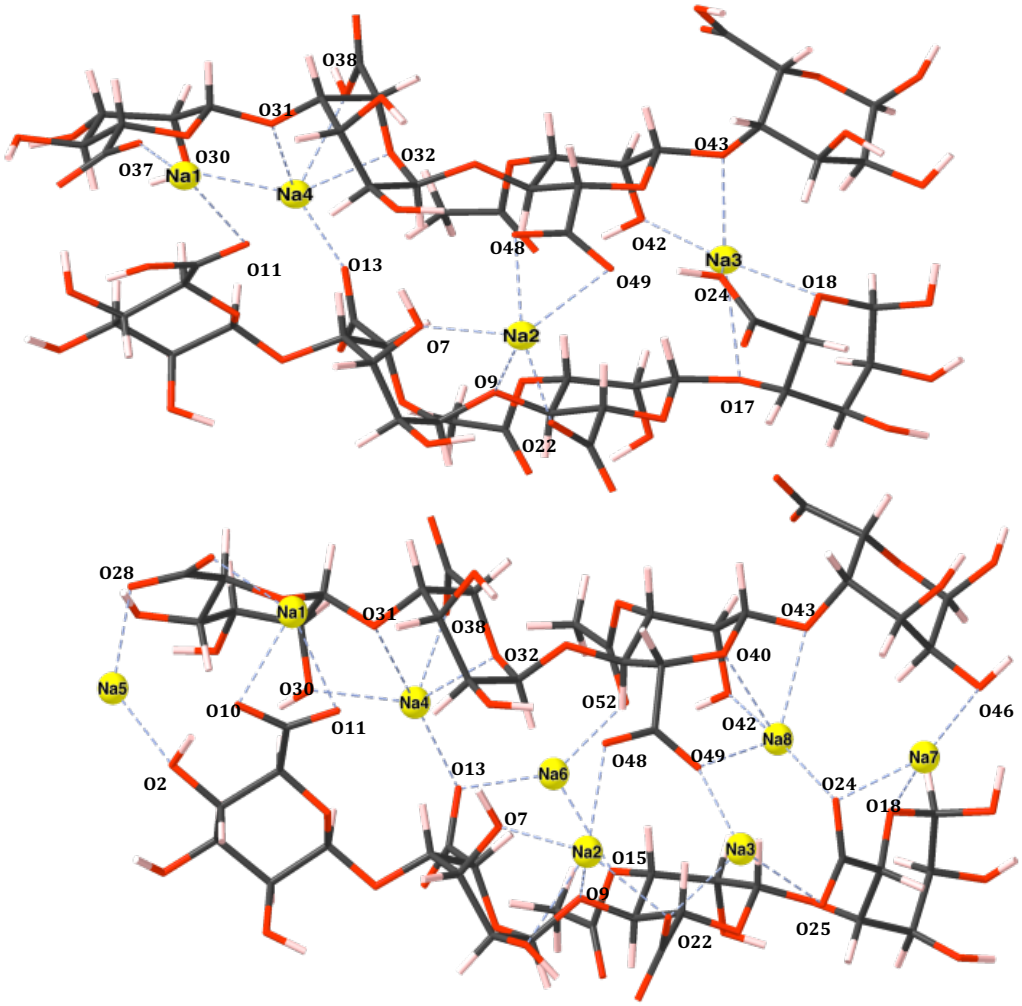


Figure 3.15: Sodium ion substitution patterns defect 1 (top) and defect 2 (bottom). Carbon atoms are shown in grey, oxygen in red, sodium as yellow spheres and hydrogen in pink. Oxygen atoms and cations involved in ionic bonding are labelled and the cation-oxygen bonds are shown with dashed blue lines.

The sodium ion positions in defect 1 (**Fig 3.15 top**) show analogies to native calcium coordination environments, with Na4 adopting an identical environment to Ca4, as well as Na1, Na2 and Na3 adopting identical environments to Ca1, Ca2 and Ca3 apart from interactions with O10, O8 and O40 respectively. Maximum CNs of 5 are observed in the sodium cross-linked bacterial alginate structures, which is a characteristic retained here also and, as a result, clearly the loss of the O10, O8 and O40 interactions in the defect 1 structure can be interpreted. In turn, this reinforces a potential upper-bound for sodium ion CNs when

complexed in-between two bacterial alginate chains. The ability of the sodium ion to display a similar coordination to the native calcium ions upon substitution can be attributed to the near equivalent ionic radii. The oxygen donors that were accessible to the native calcium ion are also accessible to the sodium ion. Despite this, however, the resulting sodium complex is thermodynamically more unstable.

Discussed in **Section 3.4.3.3** was the observation that the ionic interactions to COO^- functionality are weaker in the case of sodium compared to calcium and that strong bonding to the COO^- groups drives overall complex stability. The average Na^+-COO^- bond length and population for defect 1 are 2.45 Å and 0.06 |e| respectively, whereas for the native calcium these values are 2.27 Å and 0.14 |e|. As with the initial 2-chain complexes, the sodium ions bind in a weaker fashion to the COO^- groups, destabilising the resulting complex, despite being able to nearly mimic the native calcium chelate geometry. The defect 2 structure displays the same ill-defined inter-chain binding positions that were seen in the sodium cross-linked bacterial alginate structures and, therefore, for reasons also discussed previously in **Sections 3.4.3.2** and **3.4.3.3** it is unsurprising to observe a thermodynamically more unstable structure. Surprisingly, the torsion changes, relative to the calcium cross-linked $PolyMG_p$ structure, are larger in defect 1 despite the sodium ions adopting a similar coordination environment to the native calcium, with torsion changes reaching $(+29, -20)^\circ$ (top chain M2-G2 junction). For defect 2, the torsion changes are smaller, not surpassing $(+15, -20)^\circ$ (bottom chain M2-G2 junction) which suggests substitution patterns involving re-protonation of certain COO^- groups have a larger impact on the conformation of the polyuronate backbone.

Furthermore, it is thermodynamically unstable for sodium ions to substitute calcium ions from their binding pockets in the $PolyMG_p$ structure, which underscores the significant stability of these chelate complexes. The large positive formation energies would indicate that the calcium inter-chain binding positions within bacterial alginate structures are, potentially, stable over time and, in turn, corroborates the rheological observation that calcium inter-chain binding positions between algal alginate chains are permanent [122]. The ^{13}C-NMR observation that the formation of the calcium chelate complex in the *P. aeruginosa* EPS is reversible [130], is likely only achieved upon long exposure to large excesses of sodium ions, perhaps unachievable in biology.

It is therefore unlikely that physiological sodium softens the mucoid *P. aeruginosa* EPS scaffold to a point where it could be eradicated by the host. In fact, this conclusion is clearly justified through clinical observations [93, 102]. If mucoid biofilms softened under physiological conditions (physiological sodium), then perhaps mucoid EPS matrices would

not be long lasting in CF patients.

3.4.3.5 *In vivo* implications

It is clear that, despite their biological ubiquity, sodium ions are not implicated in inducing stable aggregation of the mucoid *P. aeruginosa* EPS and, therefore, are not implicated in stabilising mucoid *P. aeruginosa* biofilms in the CF lung. In contrast, calcium and magnesium ions, which are significantly elevated in the CF lung, are able to induce stable aggregation of EPS structures. Calcium ions form the most thermodynamically favourable cross-linked bacterial alginate structures and are, therefore, considered the most important sputum ion, out of those tested, for mucoid *P. aeruginosa* EPS stability and biofilm chronicity. The greater stability of the calcium 2-chain complexes helps in understanding the virulent consequences of the mucoid *P. aeruginosa* biofilms exposure to calcium ions, namely the development of thicker, more granular and rigid biofilms that are difficult to detach [132, 133, 305]. This work shows how the stability of these sputum ion cross-linked bacterial alginate systems (**Table 3.6**) can be greatly reduced if calcium can be removed as the cross-linking ion. However, removing calcium ions from their binding positions via calcium-for-sodium cation exchange is thermodynamically unfavourable and, as such, could only be achieved over long time-scales in large excesses of sodium ions, which potentially is unachievable in the host given the continual leakage of calcium from cell necrosis and bleeding. Therefore, more therapeutically impactful strategies to combat mucoid biofilm infection in the CF lung may include calcium chelation. Recently, a novel low molecular weight guluronate rich alginate polymer has been observed to disrupt established *P. aeruginosa* biofilms through chelating calcium within the biofilm matrix [183].

3.5 Summary

Two 2-chain bacterial alginate molecular models, structurally representative of the mucoid *P. aeruginosa* EPS observed in biofilms within the CF lung, were constructed using a step-wise approach to ensure the development of the most accurate mucoid EPS models constructed to date. Specifically, these models represented areas of 0% and 50% guluronate as well as encompassing acetyl groups, a structural motif unique to mucoid *P. aeruginosa*, at thermodynamically favourable acetylation positions. These models demonstrate how thermodynamically favourable accommodation of sodium, calcium and magnesium ions (ions significantly elevated in CF sputum) is electrostatic in origin, but the overall stability of the resulting 2-chain complexes is influenced more-so by the geometry of the chelation sites rather than simply the strength of the cation-oxygen bonds. Independent of guluronate content, calcium ions are the strongest cross-linker, forming more stable 2-chain complexes by approximately 9 eV and 5 eV compared to sodium and magnesium respectively. These

large differences show unequivocally how calcium and magnesium ions impact the stability of the EPS structure relative to the biologically ubiquitous sodium ion.

The divalent ions are able to induce conformational changes to the polyuronate, maximising the geometry of the chelation site. In regions of guluronate inclusion, more thermodynamically favourable cation cross-linked 2-chain complexes were obtained as M-G junctions offer oxygen functionality whose spatial arrangement is suitable for saturating the coordination environment of the cation and chelate complexes can be formed without the need for severely large torsion changes. This result, in particular, demonstrates the importance of using larger whole-chain models, rather than relying on smaller disaccharides to infer chemical nuance and clearly explains the significance of the guluronate units in the bacterial alginate chains with regard to bacterial virulence and EPS chronicity. Finally, it is clear that preventing calcium ions from mediating electrostatic cross-links between bacterial alginate structures will be critical in facilitating the disruption of the mucoid EPS and that using sufficiently detailed theoretical models is necessary for establishing meaningful chemical insight.

The models constructed in this chapter can be considered as quadramer "base-units" that will be layered to construct a larger EPS model system. This larger EPS model system will be used to study the molecular functionality and interactions that govern the movement of molecules through the mucoid *P. aeruginosa* EPS matrix. This is presented in the proceeding chapter, **Chapter 4**. In addition, these 2-chain models are also suitable to study cation exchange events and cation-induced EPS weakening through exposure to exogenous gallium therapy. This is presented in **Chapter 6**.

Molecular motion through the EPS

4.1 Scope of the chapter

The principle that mucoid *P. aeruginosa* biofilm cells utilise quorum sensing as a mechanism for cell-to-cell communication was introduced in **Section 1.4**. By sending quorum sensing autoinducer (QSAI) molecules between individual cells, through the EPS matrix, the bacterial (biofilm) population is able to collectively express specific genes, making quorum sensing crucial for biofilm maintenance and proliferation. Two key QSAI molecules utilised by mucoid *P. aeruginosa* in the CF lung are C_4-HSL and PQS. Little is known regarding the physicochemical interactions that facilitate the motion of these molecules through the EPS, given that the EPS is an extremely effective diffusion barrier. Understandably, knowledge of these interactions is critical for pharmaceutical design and EPS penetration.

This chapter explains how the PolyMG$_p$ and PolyM$_{ap}$ systems are extended using DFT modelling to stable four chain systems, introducing the key structural details required for stable exothermic association. Following this, the most stable 4-chain system is transformed using finite temperature, explicit solvent MD to give an EPS model more representative of the EPS that occurs under physiological conditions. This physiological structure is then used as the model system to explore modes of interaction that occur between the EPS and the two QSAIs, C_4-HSL and PQS.

4.2 Computational details

4.2.1 Simulation parameters

All Density Functional Theory (DFT) calculations were performed using the plane-wave Density Functional Theory (DFT) code, CASTEP [280]. A convergence tested cut-off energy of 900 eV was employed, as well as a Monkhorst-Pack **k**-point grid of 1 x 1 x 1 to sample the Brillouin zone [255]. On the-fly ultrasoft pseudopotentials were used [281] alongside the PBE exchange-correlation functional [240]. Intra- and intermolecular dispersive forces were

accounted for by applying the semi-empirical TS dispersion correction [261].

All molecular dynamics trajectories were computed using DLPOLY4 [282]. The conversion of all molecular models into DLPOLY input files was performed using DLFIELD [306]. All MD trajectories were computed using the OPLS2005 forcefield [274, 275] in the canonical (NVT) ensemble, where the RATTLE algorithm [283] was used to constrain covalent bonds to hydrogen atoms, meaning the integration time-step could be set to 2 fs. The temperature was held at 310 K (body temperature) using Langevin thermostatting [279]. Electrostatics were treated using the Smooth-Particle-Mesh-Ewald method [307] and the distance cut-offs for electrostatic and Leonard-Jones interactions were set to 1.2 nm.

4.2.2 Creation of an exothermic 4-chain system

To explore modes of interaction, which occur under physiological conditions, between the QSAIs and the EPS, a model larger than 2-chains was needed. As such, the 2-chain systems, PolyMG$_p$ and PolyM$_{ap}$, needed extending to form a 4-chain system. A system on the scale of four chains, cross-linked by Ca^{2+}, captures key morphological features present within the bulk that cannot be captured by 2-chain systems alone. For example, fluorescent amphiphilic carbon quantum dot imaging has captured a branched, dendritic, fractal like geometry adopted by mucoid EPS filaments at a multi-chain (mesoscopic) scale [308]. Furthermore, discontinuities in the mucoid *P. aeruginosa* EPS mesostructure, in the form of step-edges, open void spaces and water channels has been observed in transmission electron microscopy (TEM), fluorescent lectin binding analysis and pulsed field gradient NMR spectroscopy experiments performed *in vitro* and *in vivo* [170, 175, 309]. All these particular morphologies cannot be captured by considering two EPS chains alone and an expansion to a multi-chain system is required. The major limitation in not expanding to a multi-chain EPS system would be that the model is less structurally representative of the true chemical (*in vivo*) system, which may fail to capture accurate modes of interaction between the EPS matrix and secondary molecules which occur under physiological conditions.

All these listed morphological features are (partially) captured at a 4-chain scale and this is discussed in depth in **Section 4.3** and **Section 4.4**. Therefore, the 4-chain system offers a better representation of the bulk EPS whilst, importantly, retaining DFT tractability and permitting the quantification of thermodynamic stabilities, which is key to rationalising the dynamical behaviour of the QSAI molecules through the EPS scaffold.

Therefore, two PolyM$_{ap}$ and two PolyMG$_p$ models need to be combined to create two unique 4-chain systems. In considering the structure of these models, various experimental observations were taken into account. Calcium alginate gels present with a low scattering

exponent in SAXS measurements, indicating a fine fibrillar morphology which arises as a result of a large ionic cross-linking density, achieved when calcium ions promote gelation [300]. Additionally, fibre X-ray diffraction has shown that bulk aggregation of calcium alginate gels is facilitated by alginate chains associating about calcium ions through both parallel and antiparallel chain orientations - the so called "up-down" association [310]. In our models, these particular structural observations can be realised in the 4-chain systems by stacking the 2-chain complexes either parallel or antiparallel along the chain axis with the acetyl groups oriented either parallel or antiparallel, both inward and outward facing. The 4-chain system constructed from the stacking of two PolyMG$_p$ systems is referred to as 4-PolyMG and the 4-chain system constructed from the stacking of two PolyM$_{ap}$ systems is referred to as 4-PolyM. For the 4-PolyM system, which has alternating acetyl orientations, this gave four possible stacking arrangements. In the 4-PolyMG system, which has all of its acetyl groups in the same orientation, two additional stacking arrangements arose corresponding to the antiparallel acetyl groups both facing towards or away from the neighbouring stack. The stacking arrangements for the 4-PolyMG and 4-PolyM systems are shown in **Fig 4.1** and **Fig 4.2** respectively.

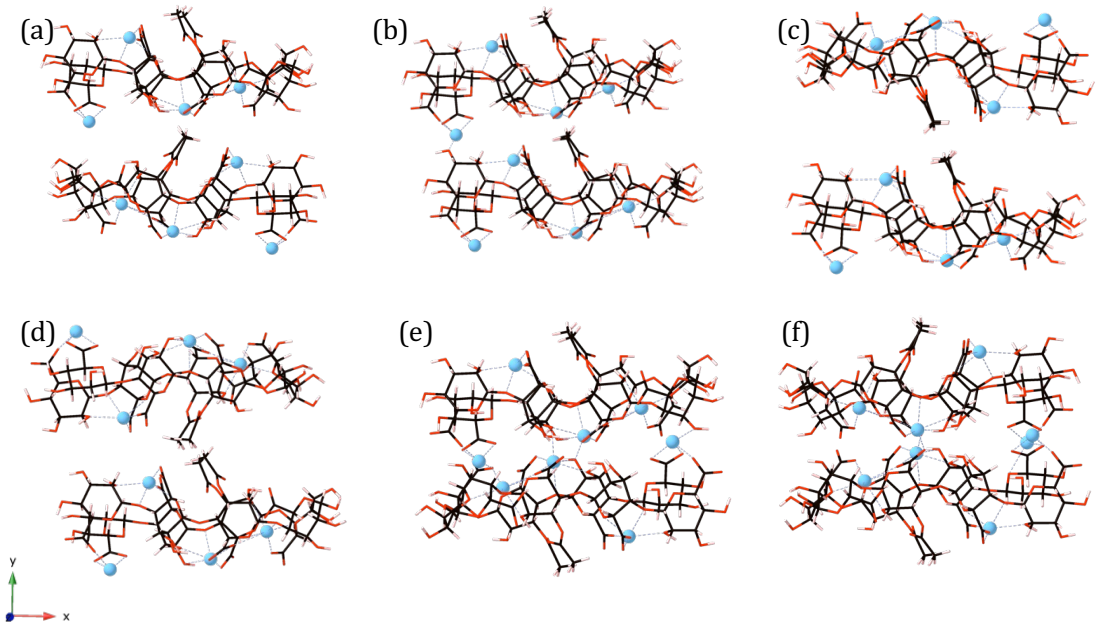


Figure 4.1: Different stacking arrangements tested when identifying the most thermodynamically favourable packing arrangement in the 4-PolyMG system: a) Chains antiparallel, acetyl parallel, b) Chains parallel, acetyl parallel, c) Chains antiparallel, acetyl antiparallel facing towards neighbouring stack, d) Chains parallel, acetyl antiparallel facing towards neighbouring stack, e) Chains antiparallel, acetyl antiparallel facing away from neighbouring stack & f) Chains parallel, acetyl antiparallel facing away from neighbouring stack. Carbon atoms are shown in black, oxygen in red, calcium as blue spheres and hydrogen in pink. Calcium-oxygen bonds are shown with dashed blue lines.

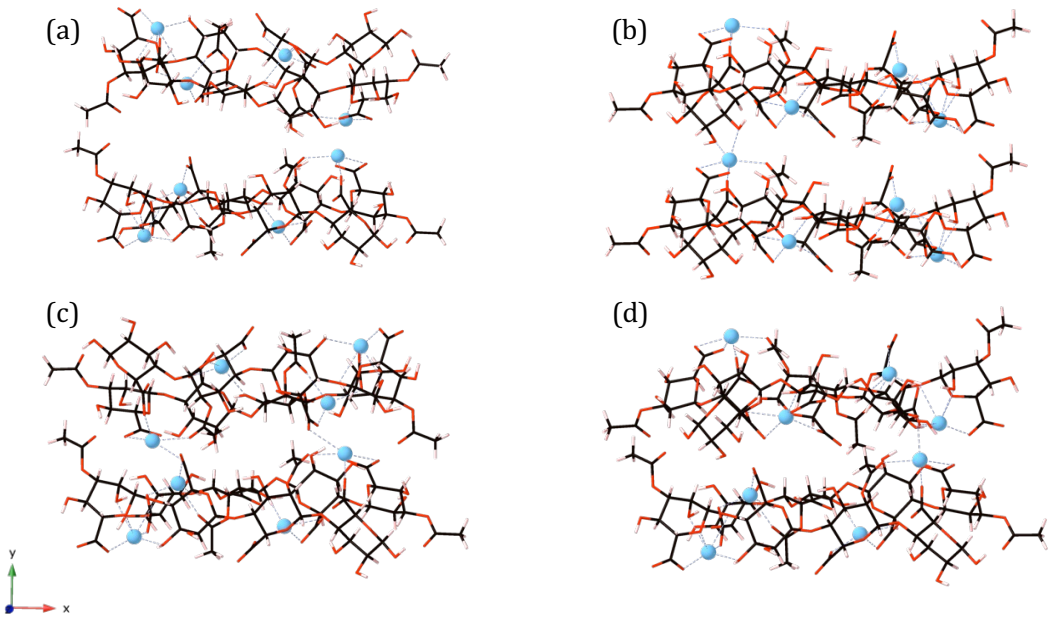


Figure 4.2: Different stacking arrangements tested when identifying the most thermodynamically favourable packing arrangement in the 4-PolyM system: a) Chains parallel, acetyl antiparallel, b) Chains parallel, acetyl parallel, c) Chains antiparallel, acetyl parallel & d) Chains antiparallel, acetyl antiparallel . Carbon atoms are shown in black, oxygen in red, calcium as blue spheres and hydrogen in pink. Calcium-oxygen bonds are shown with dashed blue lines.

Each stacking arrangement was subject to a geometry optimisation in a simulation cell measuring 42 Å × 29 Å × 45 Å. Due to the large number of different stacking arrangements, six and four for the 4-PolyMG and 4-PolyM arrangements respectively, an initial filtering, at lower tolerances, was performed to identify the most stable stacking arrangement for both the 4-PolyMG and 4-PolyM systems. During these filtering optimisations, the SCF tolerance was set to 1×10^{-5} eV Atom^{-1} and the energy, force and displacement tolerances for the geometry optimisations were set to 5×10^{-5} eV Atom^{-1}, 0.1 eV Å^{-1} and 5×10^{-3} Å respectively. The thermodynamic stability of each 4-chain stacking arrangement was measured by means of evaluating a formation energy as given in **Equation 4.2.1**

$$E_f = E_{4-chain} - 2(E_{2-chain}). \tag{4.2.1}$$

$E_{4-chain}$ is the energy of the optimised 4-PolyMG or 4-PolyM stacking arrangement and $E_{2-chain}$ is the energy of the initial PolyMG$_p$ or PolyM$_{ap}$ system.

The most stable 4-PolyMG and 4-PolyM stacking arrangements identified by the initial low tolerance filtering optimisations, labelled as 4-PolyMG* (**Fig 4.1f**) and 4-PolyM* (**Fig 4.2c**) respectively, were further optimised at finer tolerances to obtain better converged ground-state geometries. For these optimisations the SCF tolerance was set to 2×10^{-6}

eV Atom^{-1} and the energy, force and displacement tolerances for the optimisation were set to 2 $\times$ 10^{-5} eV Atom^{-1}, 0.05 eV Å^{-1} and 2 $\times$ 10^{-3} Å respectively. These finer tolerance calculations show the 4-PolyMG* structure to be the most thermodynamically favourable arrangement of four mucoid *P. aeruginosa* EPS chains complexed about calcium ions, and its optimised structure is given in **Fig 4.4**. The origins of its greater stability are discussed in more detail in **Section 4.3.1**.

4.2.3 Generation of a physiological structure

The above DFT optimisations were performed at zero Kelvin (0 K) in vacuo. Therefore, although the above optimisations allowed the quantification and identification of the most thermodynamically favourable arrangement of four mucoid *P. aeruginosa* EPS chains complexed about calcium ions, they fail to reflect possible conformational changes that could occur at physiological temperature and in the presence of water.

The most thermodynamically favourable structure obtained from the previous step (4-PolyMG*, **Fig 4.4**) was, therefore, used as a starting structure for a subsequent MD simulation to obtain a structure more representative of that observed at physiological temperature. This MD simulation was performed over 2 ns under periodic boundary conditions in a simulation cell measuring 60 Å $\times$ 60 Å $\times$ 60 Å and encompassing the 4-PolyMG* structure and SPC (Simple Point Charge) water. The thermally equilibrated 4-PolyMG* structure, labelled as 4-PolyMG$_{MD}$, possesses a discontinuous, dendritic, V-shaped morphology and the origins of these morphological features are discussed in more detail in **Section 4.4**. The 4-PolyMG$_{MD}$ model corresponds to a complete molecular model of a hydrated mucoid *P. aeruginosa* EPS matrix which exists at physiological temperature.

4.2.4 Simulating EPS-molecule binding

MD simulations were performed to sample modes of interaction between the 4-PolyMG$_{MD}$ structure and two QSAI molecules, C_4-HSL and PQS. To get physiologically representative conformations, these two molecules were equilibrated over 1 ns under periodic boundary conditions in a simulation cell measuring 40 Å $\times$ 40 Å $\times$ 40 Å, solvated with SPC water.

After obtaining thermally equilibrated conformations for the two QSAI molecules, they were individually combined with 4-PolyMG$_{MD}$ structure, positioned 6 Å away from the base of a V-shaped cleft that opened up during its thermal equilibration at 310 K (see **Section 4.4**). Two orientations of each QSAI molecule within the V-shaped cleft were tested: one with the hydrocarbon tail facing outwards towards the solvent atmosphere and the other with the hydrocarbon tail oriented inwards. When trying to isolate equilibrium configurations - to isolate EPS-QSAI adducts that occur at *physiological* equilibrium - simulations achieve the

greatest success when the initial starting position is close to the state(s) one is interested in sampling [254]. As such, these particular starting positions seek to maximise the potential for QSAI-EPS binding. This makes it possible to critically assess the molecular functionality inherent within the QSAIs and EPS that govern interactions between the two and dictate the motions of the QSAIs through the EPS. Furthermore, the dynamical trajectory over the potential energy surface (PES) is influenced by the starting configuration. A different starting configuration (or initial velocities) gives rise to a different, completely unique, trajectory over the PES. Therefore, performing additional simulations from different starting configurations enables greater sampling efficiency, leading to a more representative array of isolated structures. Therefore, starting from each of two unique QSAI initial orientations outlined above, all MD simulations were performed in triplicate.

MD trajectories were computed over 10 ns under periodic boundary conditions in a simulation cell measuring 60 Å × 60 Å × 60 Å solvated with SPC water. This trajectory length allows the molecule to sample preferential binding modes. Structures (EPS-molecule adducts) were isolated from the trajectories every 1 ns, as well as at each time-step corresponding to the minimum in the total forcefield, configurational and electrostatic energies, for subsequent DFT thermodynamic stability calculations.

If the isolated EPS-QSAI structures were translationally separated by an inter-molecular distance ≤ 6 Å, their thermodynamic stabilities were calculated by means of evaluating a formation energy given in **Equation 4.2.2**

$$E_f = E_{matrix-molecule\ adduct} - (E_{4-PolyMG_{MD}} + E_{molecule})\ . \tag{4.2.2}$$

$E_{matrix-molecule\ adduct}$ is the energy of an isolated EPS-molecule system, $E_{4-PolyMG_{MD}}$ is the energy of the thermally equilibrated 4-PolyMG$_{\mathrm{MD}}$ system and $E_{molecule}$ is the energy of a thermally equilibrated molecule, either C_4-HSL or PQS. The energy of each system (each term in **Equation 4.2.2**) was evaluated using a 0 K in vacuo DFT single-point energy calculation where the SCF tolerance was set to 2×10^{-6} eV Atom^{-1}. Finally, Mulliken bond populations [270] were calculated to classify the nature of bonding between the EPS and the molecule in each of the final EPS-molecule adducts. The EPS-QSAI trajectory that yielded the most thermodynamically favourable EPS-QSAI adducts (according to **Equation 4.2.2**) is reported and analysed in **Section 4.5**.

4.3 Exothermic association of four chains

4.3.1 Stability of the 4-chain stacks

The filtering optimisation of each staking arrangement in the 4-PolyMG and 4-PolyM systems yielded a thermodynamically favourable 4-chain system, indicating that independent of orientation, having two 2-chain systems in the vicinity of one another, and/or combining to form a larger 4-chain system, is thermodynamically favourable. The formation energies as well as the number of hydrogen and ionic bonds formed between the adjoining stacks for all stacking arrangements in 4-PolyMG and 4-PolyM systems are given in **Table 4.1** and **Table 4.2** respectively.

Table 4.1: Formation energies (eV), number of ionic bonds between adjoining stacks and number of hydrogen bonds between adjoining stacks for all stacking arrangements in the 4-PolyMG system.

System	E_f (eV)	Number of ionic bonds between adjoining stacks	Number of hydrogen bonds between adjoining stacks
Chains antiparallel, acetyls parallel	-1.75	1	0
Chains parallel, acetyls parallel	-1.81	2	1
Chains antiparallel, both acetyls antiparallel facing towards neighbouring stack	-0.17	0	0
Chains parallel, both acetyls antiparallel facing towards neighbouring stack	-0.52	0	0
Chains antiparallel, both acetyls antiparallel facing away from neighbouring stack	-5.11	5	1
Chains parallel, both acetyls antiparallel facing away from neighbouring stack	-6.88	8	3

Table 4.2: Formation energies (eV), number of ionic bonds between adjoining stacks and number of hydrogen bonds between adjoining stacks for all stacking arrangements in the 4-PolyM system.

System	E_f (eV)	Number of ionic bonds between adjoining stacks	Number of hydrogen bonds between adjoining stacks
Chains parallel, acetyl antiparallel	-1.44	1	1
Chains parallel, acetyl parallel	-2.84	3	0
Chains antiparallel, acetyl parallel	-6.20	7	3
Chains antiparallel, acetyl antiparallel	-2.83	3	1

The number of ionic bonds established between the adjoining stacks correlates well with overall thermodynamic stability, with the 4-chain systems that establish the largest and smallest number of ionic bonds corresponding to the most and least thermodynamically

favourable systems respectively. Stacking arrangements that establish an equal number of ionic bonds between adjoining stacks are comparatively as stable, independent of the number of hydrogen bonds also established. Therefore, exothermic association of bacterial alginate chains is driven primarily through the establishment of an ionic cross-linking network about calcium, rather than the establishment of hydrogen bonding interactions between the EPS chains. When considering calcium induced algal alginate gelation, classical molecular dynamics simulations have shown that hydrogen bonds can form between negatively charged alginate chains on timescales shorter than alginate-cation complexation, indicating that hydrogen bonds maybe initially required to bring the alginate chains into closer proximity before the cation complexation event [137]. However, hydrogen bonds are not necessarily observed in the final cation-complexed state [137, 138], in turn, highlighting their secondary contribution behind ionic bonding, driving the formation of, as well as influencing the final stability of, multi-chain alginate calcium complexes. Indeed, in previous DFT modelling, studying the association of multiple algal alginate disaccharides about divalent cations, aggregation is driven solely through ionic interactions [142]. This work has now extended this understanding to bacterial alginates - the mucoid *P. aeruginosa* EPS.

Using simple steric arguments, it is easy to explain why the thermodynamic stability of the 4-PolyMG system increases as the number of acetyl groups facing the adjoining stack decreases. Having both acetyl groups facing away from the adjoining stack minimises steric repulsion and facilitates the close association of the EPS chains about calcium ions, in turn, increasing the number ionic bonds between adjoining stacks, which correlates with greater stability. Acetyl interference with multi-EPS chain association is not observed in the 4-PolyM system, where all stacking arrangements led to ionic association of four chains. The 4-chain intracomplex space is saturated with ionic bonds rather than being an empty void and this is observed independent of stacking arrangement. Interestingly, it is thermodynamically favourable to have two PolyMG$_p$ chains in the same vicinity without interacting, for example, when both acetyl groups face towards the adjoining stack. When this occurs, the calcium ions retain their binding positions and coordination environments seen in the original 2-chain complex and the inter-stack steric repulsion imposed by the acetyl groups provides no perturbation to the stable 2-chain binding configurations. Effectively, inward facing acetyl groups provide a barrier to 4-chain complexation (and ionic bond formation), favouring the retention of the 2-chain complexes in their initial stable states. Similarly, the lack of calcium induced association of $>$ 2 poly-β-D-mannuronate and poly-α-L-guluronate decamers has been observed also in MD simulated annealing studies. Both these systems were shown to be capable of establishing an ionically cross-linked 2-chain system, with the third chain lying in close vicinity to the 2-chain system without directly binding, leading to empty inter-chain space [138, 139]. In our models, this is a feature, driven by acetyl steric hindrance, that

creates void spaces within the EPS. Structurally, *P. aeruginosa* biofilms can be described as open systems encompassing cells, extracellular matrix material and void spaces [291], where the latter act as channels allowing the flow of water throughout the biofilm [170]. All stacking arrangements led to close ionic association of four chains in the 4-PolyM system, therefore, the presence of acetylated mannuronate-guluronate blocks in the EPS provides a structural origin for void spaces and water channels in mucoid *P. aeruginosa* biofilms.

Shown in **Fig 4.3** is the least stable 4-PolyMG stacking arrangement, where acetyl groups face inwards towards the adjoining stack, alongside computed cavity centroids with a minimum radius of 2.0 Å, to represent the volume that could be occupied by water molecules. This visualisation clearly shows that acetyl interference and the lack of ionic association creates a void space in which water molecules can be accommodated. For the least stable 4-PolyMG system, where the chain lengths are 4 uronate residues long (quadramers), the void spaces allow at least five water molecules to exist in the inter-chain void space. This is in contrast to the most favourable 4-PolyMG configuration that has no void spaces suitable to hold water molecules. It is also important to note that the acetyl groups penetrating this void space create a hydrophobic environment which, in turn, will prevent water binding and assist in the flow of water through these channels.

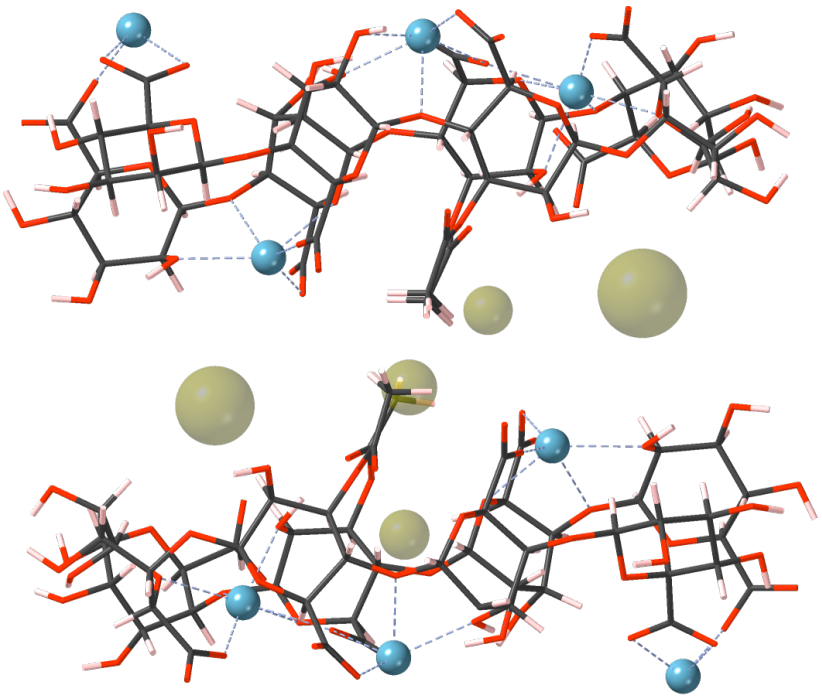


Figure 4.3: The least stable 4-PolyMG stacking arrangement, encompassing acetyl groups facing towards the adjoining stack, showing spherical cavities (minimum radius 2.0 Å) where water molecules could reside. Carbon atoms are shown in black, oxygen in red, calcium as blue spheres and hydrogen as pink. Spaces where single water molecules could reside are represented as dark yellow spheres.

4.3.2 The most stable mannuronate and mannuronate-guluronate 4-chain systems

The more accurate DFT simulations (at finer tolerances) of the most stable 4-PolyMG and 4-PolyM stacking arrangements yielded the more stable and better converged 4-PolyMG* and 4-PolyM* geometries, and can be seen in **Fig 4.4** and **Fig 4.5** respectively. These geometries correspond to the most thermodynamically favourable EPS structures, of those tested, encompassing four acetylated mannuronate-guluronate co-polymers and four acetylated mannuronate homopolymers about calcium ions.

The formation energies (according to **Equation 4.2.1**) are -6.90 eV for the 4-PolyMG* structure and -6.38 eV for the 4-PolyM* structure, meaning the association of two PolyMG$_p$ systems is slightly more exothermic compared to the association of two PolyM$_{ap}$ systems. Viewing each system down the x-axis, a central columnar distribution of calcium ions, responsible for facilitating the cross-linking, can be observed. In both systems, the most exothermic association of 4-chains leaves some calcium ions at the chain periphery, bound outside of the 4-chain intracomplex space. It can be concluded that not all calcium ions are

utilised in ionic association in EPS aggregates larger than 2-chains, which is also a feature captured in MD simulations of calcium alginate networks [134]. Fibre X-ray diffraction measurements on calcium alginate gels also observed a central column of calcium ions which facilitate chain complexation, giving rise to complexes with an inter-chain separation of 6.6 Å [310]. In these models, the inter-chain separation, as measured between the glycosidic functional groups of the top and bottom 2-chain complexes, is larger for the 4-PolyM* system (8.5 Å) compared to the 4-PolyMG* system (7 Å), with both differing from separations measured by fibre X-ray diffraction in alginate fibres by 1.5 Å and 0.4 Å respectively [310]. This is illustrated in **Figs 4.6 and 4.7**. In particular, the alginate fibres used in this fibre X-ray diffraction study possessed a G-G junction abundance of 55% and, given the 0.4 Å deviation in inter-chain separation compared to our 4-PolyMG* model, this might suggest that experimentally the PolyMG chains contributed most significantly to the observed diffraction pattern. The distribution of calcium ions is more homogeneous in the 4-PolyMG* system compared to the 4-PolyM* system, the latter possessing no calcium ions cross-linking the central uronate residues. As such, six calcium ions are responsible for facilitating 4-chain association in the 4-PolyMG* system, whereas only four are responsible in the 4-PolyM* system. Subsequently, this assists in promoting tighter ionic association in the 4-PolyMG* system, decreasing the inter-chain separation. This reinforces a recurring theme seen thus far in bacterial alginate cation complexation modelling, that guluronate inclusion is required for the tightest and most exothermic calcium complexation.

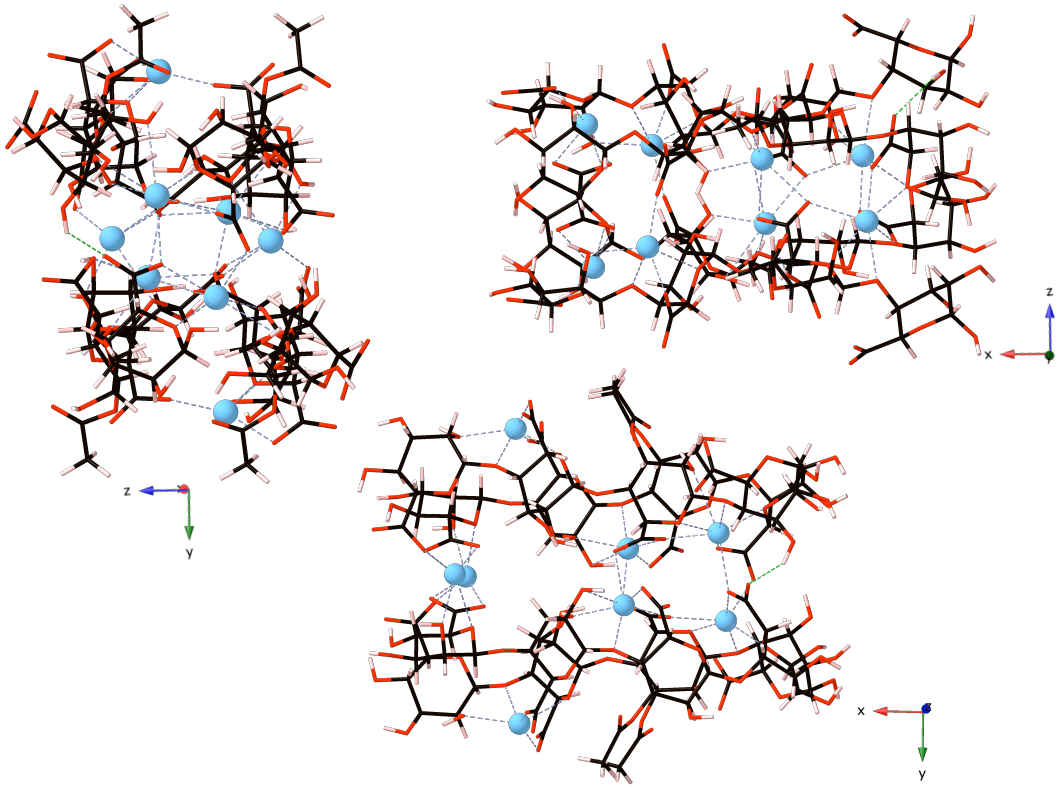


Figure 4.4: The 4-PolyMG* structure, corresponding to the most thermodynamically favourable arrangement of four EPS chains complexed about calcium ions, of those tested, viewed down the x, y and z axes. Carbon atoms are shown in black, oxygen in red, calcium as blue spheres and hydrogen in pink. Calcium-oxygen ionic bonds are shown with dashed blue lines and hydrogen bonds are shown with dashed green lines.

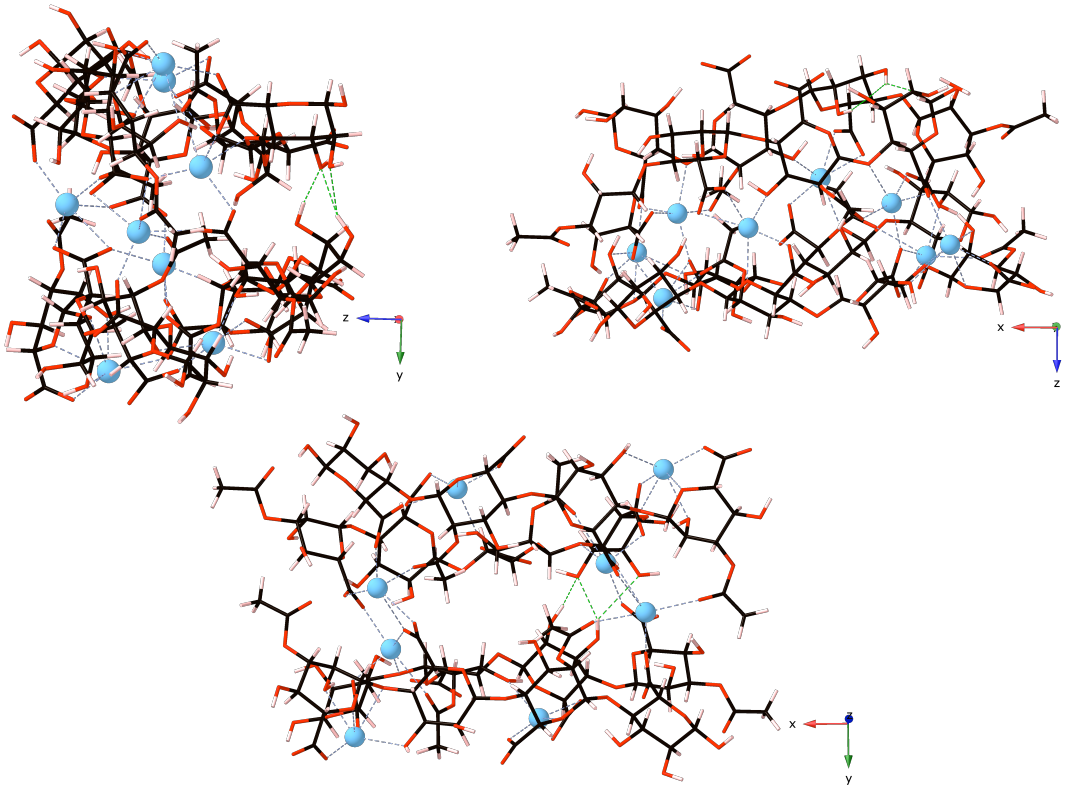


Figure 4.5: The 4-PolyM* structure viewed down the x, y and z axes. Carbon atoms are shown in black, oxygen in red, calcium as blue spheres and hydrogen in pink. Calcium-oxygen ionic bonds are shown with dashed blue lines and hydrogen bonds are shown with dashed green lines.

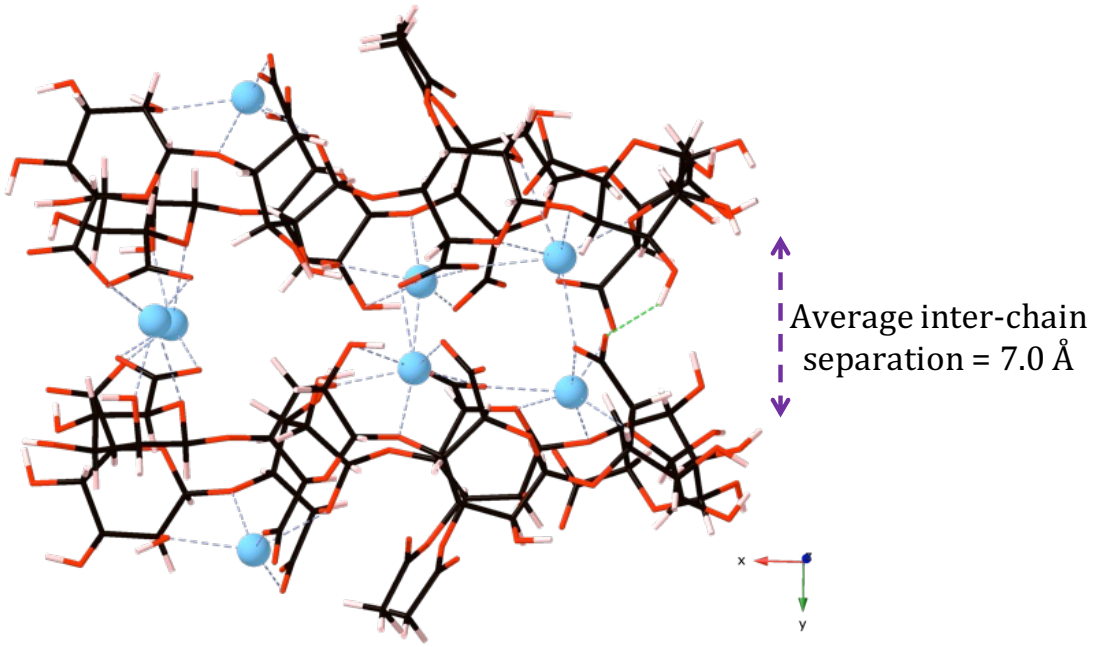


Figure 4.6: Average inter-chain spacing in the 4-PolyMG* structure. Carbon atoms are shown in black, oxygen in red, calcium as blue spheres and hydrogen in pink. Calcium-oxygen ionic bonds are shown with dashed blue lines and hydrogen bonds are shown with dashed green lines.

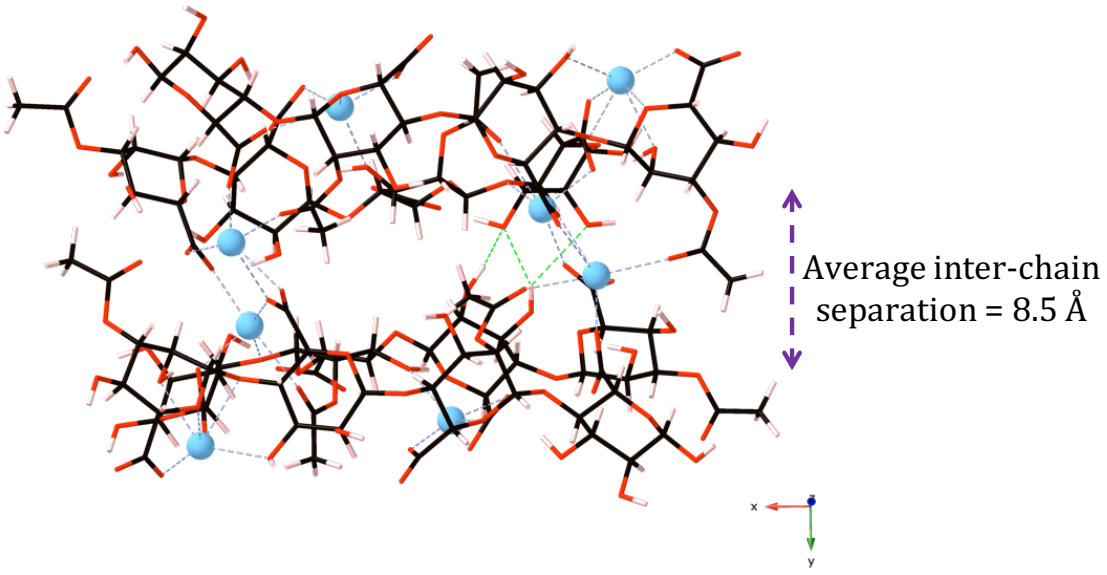


Figure 4.7: Average inter-chain spacing in the 4-PolyM* structure. Carbon atoms are shown in black, oxygen in red, calcium as blue spheres and hydrogen in pink. Calcium-oxygen ionic bonds are shown with dashed blue lines and hydrogen bonds are shown with dashed green lines.

The average oxygen-Ca^{2+} bond lengths and populations, for all coordinating polyuronate

oxygen functionality, for both the 4-PolyMG* and 4-PolyM* systems, are given in **Table 4.3**. The stability difference between the 4-PolyMG* and 4-PolyM* system is minor because these complexes share highly similar geometrical features. Specifically, the average O-Ca^{2+}, COO^{-}-Ca^{2+}, OH-Ca^{2+}, glycosidic O-Ca^{2+} and ring O-Ca^{2+} bond lengths in the 4-PolyMG* system are highly similar to those observed in the 4-PolyM* system.

Table 4.3: Average oxygen-Ca^{2+} bond lengths and populations for the 4-PolyMG* and 4-PolyM* systems.

4-PolyMG*		
Contact	Average length (Å)	Average population (|e|)
O-Ca^{2+}	2.41	0.10
COO^{-}-Ca^{2+}	2.33	0.11
OH-Ca^{2+}	2.50	0.09
Ring O-Ca^{2+}	2.52	0.09
Glycosidic O-Ca^{2+}	2.49	0.11
4-PolyM*		
Contact	Average length (Å)	Average population (|e|)
O-Ca^{2+}	2.39	0.11
COO^{-}-Ca^{2+}	2.31	0.13
OH-Ca^{2+}	2.50	0.09
Ring O-Ca^{2+}	2.48	0.09
Glycosidic O-Ca^{2+}	2.43	0.09
Acetyl O-Ca^{2+}	2.55	0.09

As already discussed, acetyl groups do not inhibit ionic association in the 4-PolyM system. In fact, the acetyl groups contribute to ionic association, engaging in ionic cross-linking, providing O-Ca^{2+} contacts at the longest range relative to other oxygen functionality. The acetyl O-Ca^{2+} contact is the only ionic contact present in the 4-PolyM* system that isn't present in the 4-PolyMG* system. The contribution of the acetyl group to ionic cross-linking is a geometric feature that distinguishes multi-EPS chain, exothermic, ionic association of acetylated mannuronate EPS fractions from acetylated mannuronate-guluronate EPS fractions. Acetyl groups are involved in ionic association in the former but are implicated in increased steric repulsion and preventing ionic association in the latter. The difference in stability (0.52 eV) between the 4-PolyMG* and 4-PolyM* systems can be interpreted by considering the ionic cross-linking density, maximum cation CNs and the number of calcium ions involved in the 4-chain association. Higher maximum CNs are observed in the 4-PolyMG* system (CN = 7), which also encompasses an additional ionic bond between the adjoining stacks, compared to the 4-PolyM* system (CN = 6). Furthermore, six calcium ions are involved in establishing ionic bonds between the adjoining stacks in the 4-PolyMG* system, whereas only four are involved in the 4-PolyM* system. Overall, these three factors

outline a larger 4-chain intracomplex ionic density in the 4-PolyMG* system, compared to the 4-PolyM* system, making it the more stable out of the two 4-chain systems.

As with the 2-chain complexes, across both the 4-PolyMG* and 4-PolyM* systems, all polyuronate oxygen functionality is involved in the cation cross-linking and, thus, 4-chain association. In both systems, the COO^{-}-Ca^{2+}contacts are the ionic bonds that are most stable (shortest of all O-Ca^{2+} bonds with the largest bond populations; **Table 4.3**) and these interactions occur most frequently (**Figs 4.4** and **4.5**). Therefore, it can be inferred that the COO^{-} group is the functional group most implicated in the formation of an exothermic 4-chain complex. Not only is this analogous to 2-chain calcium complexation, but also continues to align with previous findings in MD simulations of calcium ion induced algal alginate gelation [134, 137, 299]. Calcium ions with CN = 7 are coordinated to four different uronate residues, whereas when the CN = 5 or 6, the ions are only bound to three uronate residues. Coordination to four uronate residues per calcium ion is closer to the classical egg-box description of calcium chelation by alginates [301]. Calcium chelation by two bacterial alginate chains deviates from the egg-box model (**Fig 3.11**), so it is interesting to observe egg-box characteristics regained as the number of coordinating bacterial alginate chains increases. Even though coordination to four uronate residues is possible, no calcium ion is bound to more than three EPS chains. The implication of this is that 3-chain complexation is possible about a single calcium ion, but 4-chain complexation is not; ≥ 2 calcium ions are needed to establish 4-chain aggregation.

Finally, returning to the discussion on void spaces, fibrillar stacking of two 4-PolyMG* systems, to form an 8-chain system, would recover a stacking arrangement where acetyl groups face towards the adjoining stack, creating a void space. For reference, two 4-PolyMG* systems arranged in close proximity can be seen in **Fig 4.8**.

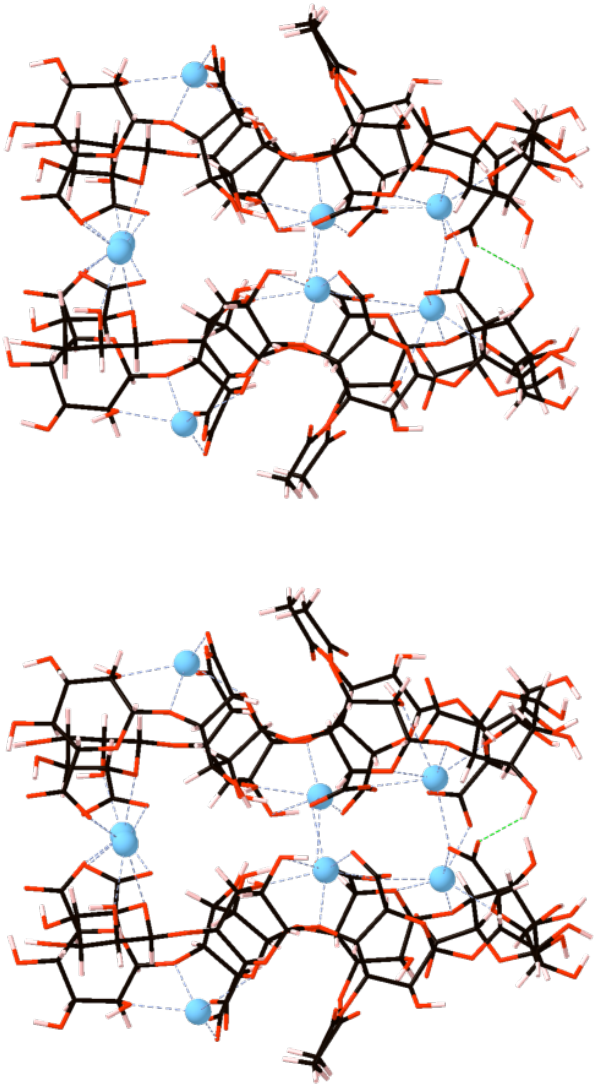


Figure 4.8: A visualisation showing how two 4-PolyMG* systems appear in close proximity, highlighting that, independent of which orientation these 4-PolyMG* systems associate to give a fibrillar morphology, acetyl groups will always face the neighbouring stack. Carbon atoms are shown in black, oxygen in red, calcium as blue spheres and hydrogen in pink. Calcium-oxygen bonds are shown with dashed blue lines and hydrogen bonds are shown with dashed green lines.

As can be seen, independent of orientation these two 4-PolyMG* systems associate to maintain a fibrillar morphology, whether that be parallel or antiparallel, a final 8-chain structure encompassing acetyl groups facing the adjoining stack cannot be avoided. Therefore, this further highlights, in acetylated mannuronate-guluronate EPS fractions, the formation of void spaces is inevitable and necessary.

4.4 Physiological equilibration

Fig 4.9 depicts the 4-PolyMG$_{MD}$ structure, created from the 4-PolyMG* structure using finite temperature explicit solvent MD. In turn, this structure corresponds to the first published [311] atomistic model of the hydrated mucoid *P. aeruginosa* EPS, which occurs at physiological temperature.

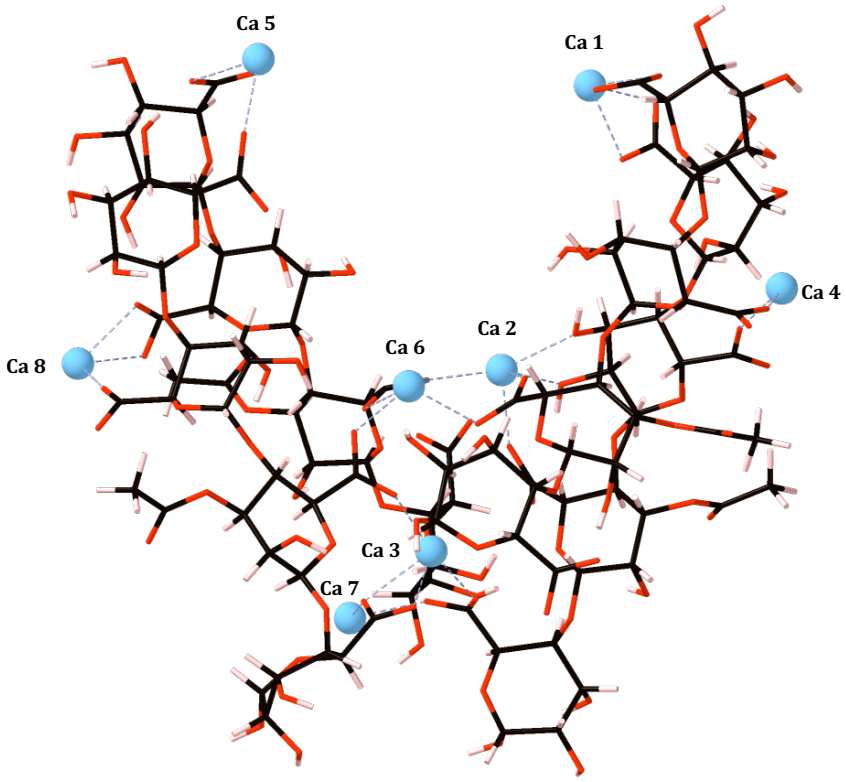


Figure 4.9: The 4polyMG$_{MD}$ structure, corresponding to the structure of the hydrated mucoid *P. aeruginosa* EPS at physiological temperature. Carbon atoms are shown in black, oxygen in red, calcium as blue spheres and hydrogen in pink. Calcium ions are labelled and calcium-oxygen ionic bonds are shown with dashed blue lines. Explicit water molecules are not shown.

Following equilibration under physiological conditions, the resultant structure shows a significant deviation from the neatly stacked EPS chains present in the 4-PolyMG* structure obtained from the DFT modelling (**Fig 4.4**). Specifically, the 4polyMG$_{MD}$ system has adopted an entangled, dendritic, V-shaped motif. This motif is not unexpected, as it has also been observed in MD simulations of calcium algal alginates [138]. Ca6 adopts an inter-chain chelate position at the base of the V-shaped cleft and the inter-chain V-angle, specifically the angle measured between Ca5-Ca6-Ca1, is 61°. This matches well with the 60° inter-chain angle measured in MD simulations of calcium ion induced aggregation of copolymeric β-D-mannuronate-α-L-guluronate decamers performed in the work of Stewart *et*

al. [138]. Indeed, in these MD simulations, all calcium ions are bound by a single carboxylate group, with the exception of a single COO^{-}-Ca^{2+}-COO^{-} inter-chain ionic cross-link which consequently favours the development of a V-shaped motif [138]. The carboxylate group dominating the calcium ion chelation geometry is observed here also.

Changes in dihedral angles, as defined in **Fig 3.10**, relative to the 4-PolyMG* system, are given in **Table 4.4**.

Table 4.4: Dihedral angles (ϕ, ψ) for the 4polyMG$_{MD}$ system. Changes relative to the 4-PolyMG* are also given. All angles are given in degrees (°). See **Figure 3.1** for labelling of the uronate units.

4polyMG$_{MD}$	(ϕ, ψ)°			Δ (ϕ, ψ)°		
	M1-G1	G1-M2	M2-G2	M1-G1	G1-M2	M2-G2
Bottom chain (1)	(-91, -167)	(-125, -164)	(-108, +50)	(-29, -51)	(-32, -38)	(-57, +100)
Bottom chain (2)	(-153, -151)	(-84, -95)	(-62, -91)	(-66, -22)	(+35, +38)	(+52, -22)
Top chain (1)	(-103, +178)	(-85, -146)	(-106, +175)	(-17, -52)	(+13, -17)	(+6, +144)
Top chain (2)	(-70, -114)	(-95, -138)	(+18, -98)	(-7, +1)	(-1, -13)	(+69, -46)

Table 4.4 lists large torsions and torsion changes reaching +69° and +144° respectively and highlighting the gain in conformational flexibility, which promotes large torsion changes to maximise the establishment of COO^{-}-Ca^{2+}interactions. As such, the frequency of COO^{-}-Ca^{2+}interactions (25 contacts) far exceeds the frequency of any other O-Ca^{2+} contact, with the only other calcium ion coordinating oxygen functional group being the hydroxyl group (OH-Ca^{2+}; 2 contacts). The glycosidic O-Ca^{2+} and ring O-Ca^{2+} interactions are eliminated upon thermal equilibration at 310 K to form the 4-PolyMG$_{MD}$ structure as a result of the EPS chains favouring the establishment of COO^{-}-Ca^{2+} interactions above all other O-Ca^{2+} interactions. Within the 4-PolyMG$_{MD}$ structure, the average O-Ca, COO^{-}-$Ca^{2+}$and OH-Ca^{2+} bond lengths (2.32 Å, 2.31 Å and 2.40 Å respectively) and Mulliken populations (0.13 |e|, 0.13 |e| and 0.09 |e| respectively) highlight the dominant contribution of the COO^{-}-Ca^{2+} interaction to the overall binding characteristics of the system. Their continued appearance across the whole equilibration trajectory suggests that the COO^{-}-Ca^{2+} interaction is most stable at 310 K in the presence of water. The carboxylate groups being the dominant contributor to the calcium ion chelation geometry has been observed extensively in finite temperature MD simulations of solvated calcium algal (non-acetylated) alginate networks [134, 137–139, 299]. The 4-PolyMG$_{MD}$ structure also extends this observation to acetylated bacterial alginate and definitively demonstrates how, under physiological conditions, the COO^{-} groups have the largest exposed surface area of electronegativity to offer near exclusive sites for cation binding. Accordingly, the 4-PolyMG$_{MD}$ structure provides pivotal structural insight into the morphology of the bulk

EPS *in vivo*, emphasising how the presence of temperature and solvent reduces the diversity in the geometry of the chelation site within calcium chelate pockets.

The only structural feature retained from the 4-PolyMG* system, is the acetyl groups facing away from the neighbouring stack, which supports the assumption the acetyl groups preferentially orient away from the main EPS chains. The acetyl groups are bound over the G1-M2 junction, in the centre of the polyuronate chain, and torsion changes about this junction are significantly smaller compared to changes about the M1-G1 and M2-G2 junctions (**Table 4.4**). The acetyl steric hindrance at this uronate position reduces the susceptibility of the junction to undergo large conformational change and, as such, there has not been a complete (180°) inversion about the G1-M2 junction, which keeps the acetyl groups facing away from the neighbouring EPS chains. Only Ca2 and Ca6 (**Fig 4.9**) are bound to three EPS chains – all other calcium ions are involved in coordination to two EPS chains. In the 4-PolyMG* structure, which possessed regained egg-box characteristics, calcium ions were bound to three EPS chains and, as such, it is clear that 4-chain intracomplex space is heavily reduced upon thermal equilibration at physiological temperature to form the 4-PolyMG$_{\mathrm{MD}}$ structure. Effectively, the exothermic 4-chain system (4-PolyMG*) has partitioned into two sets of 2-chain systems (2 $\times$ PolyMG$_p$ systems) with only Ca 2 and Ca 6 keeping the bacterial alginate network together. The reduction in the size of the 4-chain intracomplex space is the reason for the origin of the V-shaped cleft. The structural morphology of the 4-PolyMG$_{\mathrm{MD}}$ system correlates well with morphologies of calcium cross-linked mannuronate-guluronate heteropolymers observed in recent large-scale multi-chain implicit solvent MD simulations [312]. In these simulations, heteropolymers (copolymeric mannuronate-guluronate) have increased chain flexibility relative to homopolymers (polymannuronate) which leads to entanglement upon calcium cross-linking and giving discontinuous, V-shaped, morphologies possessing open clefts after aggregation [312].

The creation of a V-shaped cleft, which arises due to a reduction in the diversity of the calcium ion chelate geometry, introduces a structural discontinuity into the EPS. This assists in interpreting the complex, discontinuous, EPS structures observed in transmission electron microscopy (TEM) measurements on *P. aeruginosa* biofilms [309], as well as providing an origin for the large-scale branched/dendritic organisation of *P. aeruginosa* EPS scaffolds observed in fluorescent amphiphilic carbon quantum dot imaging [308]. The latter imaging technique is the first imaging technique able to resolve EPS chain organisation at the mesoscopic (μm) scale and, for the first time, reveals how the EPS chains localise around *P. aeruginosa* cells, adopting a dendritic fractal-like geometry [308]. Specifically, the multi-EPS associations adopt a V-shaped motif at their base and have a branching projection through 3D space [308]. The 4-PolyMG$_{\mathrm{MD}}$ system calculated here also encompasses a V-shaped motif

and perhaps, at the 4-chain molecular (nm) scale, represents the beginning, or base, of these dendritic morphologies observed experimentally at the mesoscale.

It is important to note here, that the carbon quantum dot imaging of the *P. aeruginosa* EPS biofilm matrix, performed in the work of Ritenberg and colleagues, definitively shows how temperature effects the magnitude of the dendritic scaffold morphology observed *in vitro* [308]. Explicitly, *P. aeruginosa* biofilms cultivated at temperatures of 20° formed an EPS network encompassing far fewer EPS branching (fractal) domains compared to *P. aeruginosa* biofilms cultured at 26° [308]. In turn, this offers two possible insights. Firstly, it definitively shows how temperature changes of 6° can effect, either accelerate or decelerate, the *in vitro* growth kinetics of *P. aeruginosa* biofilm matrices. Explicitly, the lower temperature growth environment is slowing *P. aeruginosa* growth and biofilm development. Secondly, and perhaps more interestingly, it showcases how temperature as a thermodynamic variable plays a role in determining EPS fine structure at the mesoscale. The 4-PolyMG$_{\mathrm{MD}}$ system, encompassing the V-shaped cleft, arises as temperature increases the conformational flexibility on the calcium cross-linked EPS network and reduces the diversity of the Ca^{2+} chelate geometry. Understandably, at lower temperatures, this conformational change is less prevalent and dendritic branched organisations of EPS chains may be far less frequent.

4.5 QSAI simulations

4.5.1 QSAI-EPS binding geometries

EPS-C_4-HSL and EPS-PQS structures and their formation energies are shown in **Fig 4.10** and **Fig 4.11** respectively. The PQS molecule binds early and remains bound to the EPS throughout the full time-scale of the trajectory and, therefore, EPS-PQS adducts are displayed at 2 ns intervals for brevity (**Fig 4.11**). In **Fig 4.10** the differences in the relative orientation of the x, y and z axes is also given to act as a reference when visualising the migration of the C_4-HSL molecule about the perimeter of the 4-PolyMG$_{\mathrm{MD}}$ structure. But note here that in all structures presented within **Fig 4.10** there has been no separation of the EPS chains and the V-shaped morphology is sustained throughout.

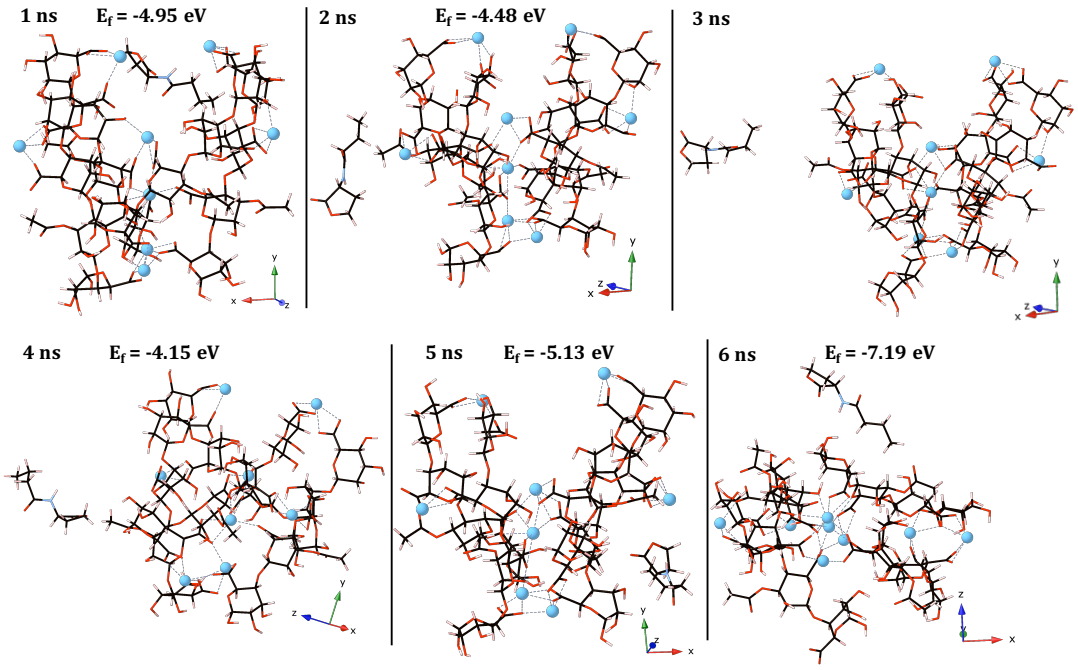


Figure 4.10: EPS-C_4-HSL structures and their formation energies (eV). Carbon atoms are shown in black, oxygen in red, calcium as blue spheres and hydrogen in pink. Calcium-oxygen ionic bonds are shown with dashed blue lines. Formation energies were not calculated if the molecule $>$ 6 Å away from the EPS.

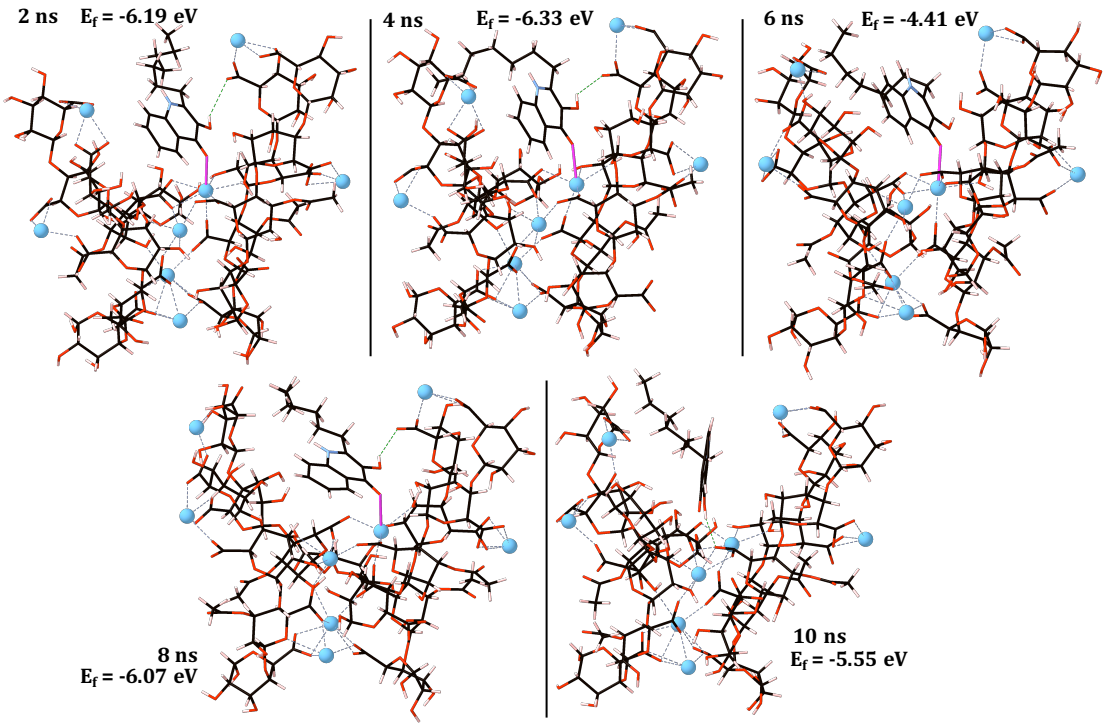


Figure 4.11: EPS-PQS structures and their formation energies (eV) displayed at 2 ns intervals. Carbon atoms are shown in black, oxygen in red, calcium as blue spheres and hydrogen in pink. Calcium-oxygen ionic bonds are shown with dashed blue lines and hydrogen bonds are shown with dashed green lines. Ionic bonds between the PQS and 4polyMG$_{MD}$ system are shown with bold pink lines.

Evident from **Fig 4.10** and **4.11** is the clear immobility of PQS, relative to C_4-HSL, when in the EPS vicinity. The C_4-HSL molecule liberates the cleft in the EPS and migrates over the EPS scaffold, but after 6 ns (**Fig 4.10**) becomes separated from the EPS and does not return. This supports the conclusion that C_4-HSL fails to show any (permanent immobilising) electrostatic (charge-dipole) or hydrogen bonding interaction with the EPS and as such does not become sequestered by the EPS scaffold. In stark contrast, the PQS molecule forms both an ionic interaction between its ketone oxygen and a single EPS Ca^{2+} ion and a hydrogen bonding interaction between its hydroxyl group and a single EPS carboxylate group (**Fig 4.11**). This results in the formation of a thermodynamically favourable ionic complex with the EPS, which is sufficient to keep the PQS molecule tethered in the cleft of the 4-PolyMG$_{MD}$ structure, overcoming hydrophobic repulsion between the hydrocarbon tail and EPS chains, for the full time-scale of the 10 ns trajectory.

After 8 ns, the PQS ionic ketone O-Ca^{2+} interaction is lost and is compensated for by a rearrangement in the single hydrogen bonding interaction, which keeps the PQS molecule irreversibly bound within the cleft. Given in **Table 4.5** are the bond lengths (Å) and

Mulliken populations (|e|) for the tethering contacts which occur in the PQS-EPS adducts.

Table 4.5: Bond lengths (Å) and Mulliken populations (|e|) for the tethering contacts which occur in the PQS-EPS adducts.

PQS ketone O-Ca^{2+} ionic bond		
Time step (ns)	Length (Å)	Mulliken population (|e|)
2	2.42	0.10
4	2.63	0.08
6	2.51	0.09
8	2.66	0.08
10	-	-
PQS OH-EPS hydrogen bond		
Time step (ns)	Length (Å)	Mulliken population (|e|)
2	2.72	0.01
4	2.47	0.01
6	-	-
8	2.56	0.01
10	1.87	0.04

It is evident from **Table 4.5** that the OH-EPS hydrogen bond is very weak relative to the ionic ketone O-Ca^{2+} interaction, as indicated by its greater length and smaller population. Explicitly, the Mulliken population of the OH-EPS hydrogen bonding interaction is approximately one 10^{th} of the ionic ketone O-Ca^{2+} interaction. On account of that, the hydrogen bond is not *static*, making it prone to rearrangement. The hydrogen bond formed at 10 ns, however, is the strongest hydrogen bond present in all of the isolated adducts indicating that the loss of the ketone O-Ca^{2+} interaction is required to drive the formation of a stronger OH-EPS hydrogen bond. The ionic ketone O-Ca^{2+} interaction is not able to undergo a change in directionality and, when lost at 8 ns, is not able to recover over the remainder of the 10 ns simulation. Note, however, this does not mean that this particular interaction may not reform on timescales greater than 10 ns. Recall, this is a rearrangement that lowers the overall stability of the EPS-PQS complex (**Fig 4.11**) and, therefore, further accentuates the role of the PQS OH group in mediating irreversible EPS adsorption under physiological conditions. Effectively, the hydrogen bond rearrangement is a compromise between overall complex stability and retaining a stable position adsorbed onto the EPS surface.

Interestingly, the EPS-PQS adduct at 6 ns, where the hydroxyl hydrogen bond is absent and only the ketone O-Ca^{2+} remains, is the least thermodynamically favourable configuration and, intriguingly, is less stable than the EPS-PQS adduct at 10 ns where only the hydroxyl

hydrogen bond exists. From **Table 4.5** it is clear to see that the OH-EPS hydrogen bond, following the hydrogen bond rearrangement, at 10 ns is a tighter tethering contact compared to the ketone O-Ca^{2+} interaction at 6 ns, which assists in explaining the greater complex stability at 10 ns. PQS dispersal into lipopolysaccharide (LPS) scaffolds has been previously investigated by FRET measurements, revealing that the hydroxyl group is of critical importance for mediating PQS-LPS interactions [313]. Therefore, critically, these simulations call definitive attention to the role of the PQS hydroxyl group in molecular binding. Despite the hydrogen bond rearrangement, highlighting the OH group as a focal point for EPS adsorption under physiological conditions, the most stable adsorbed EPS-PQS adducts occur when both the ionic ketone O-Ca^{2+} and OH-EPS hydrogen bonding interactions occur simultaneously (**Fig 4.11**). This, in turn, emphasizes that electrostatic interactions are what drives PQS adsorption onto the EPS surface. Such interactions significantly reduce the mobility of PQS and the C_4-HSL molecule, which forms no such interactions, is notably more mobile. This is true despite the C_4-HSL molecule sharing similar functional groups to PQS, namely, two ketone functional groups, one located on its "head" group and the other as part of an amide linkage attaching the hydrocarbon tail (**Fig 1.7**). However, unlike PQS, neither of these C_4-HSL ketone groups form any interaction with the EPS Ca^{2+} ions leading to the molecule being more mobile. The potential explanation for this is discussed in **Section 4.5.2**.

Recall in **Section 4.4**, the carboxylate group was the main facilitator of Ca^{2+} ion induced EPS aggregation under physiological conditions. As such, by analysing the EPS COO^--Ca^{2+}geometrical parameters within each of the EPS-QSAI adducts, it is possible to ascertain whether the QSAI molecules cause an active perturbation/disruption to the EPS ionic scaffold. Given in **Fig 4.12**, are the COO^--Ca^{2+} average bond lengths (Å) and populations (|e|), along side the average Ca^{2+} CNs, over the course of the trajectory for each of the isolated EPS + C_4-HSL and EPS + PQS adducts.

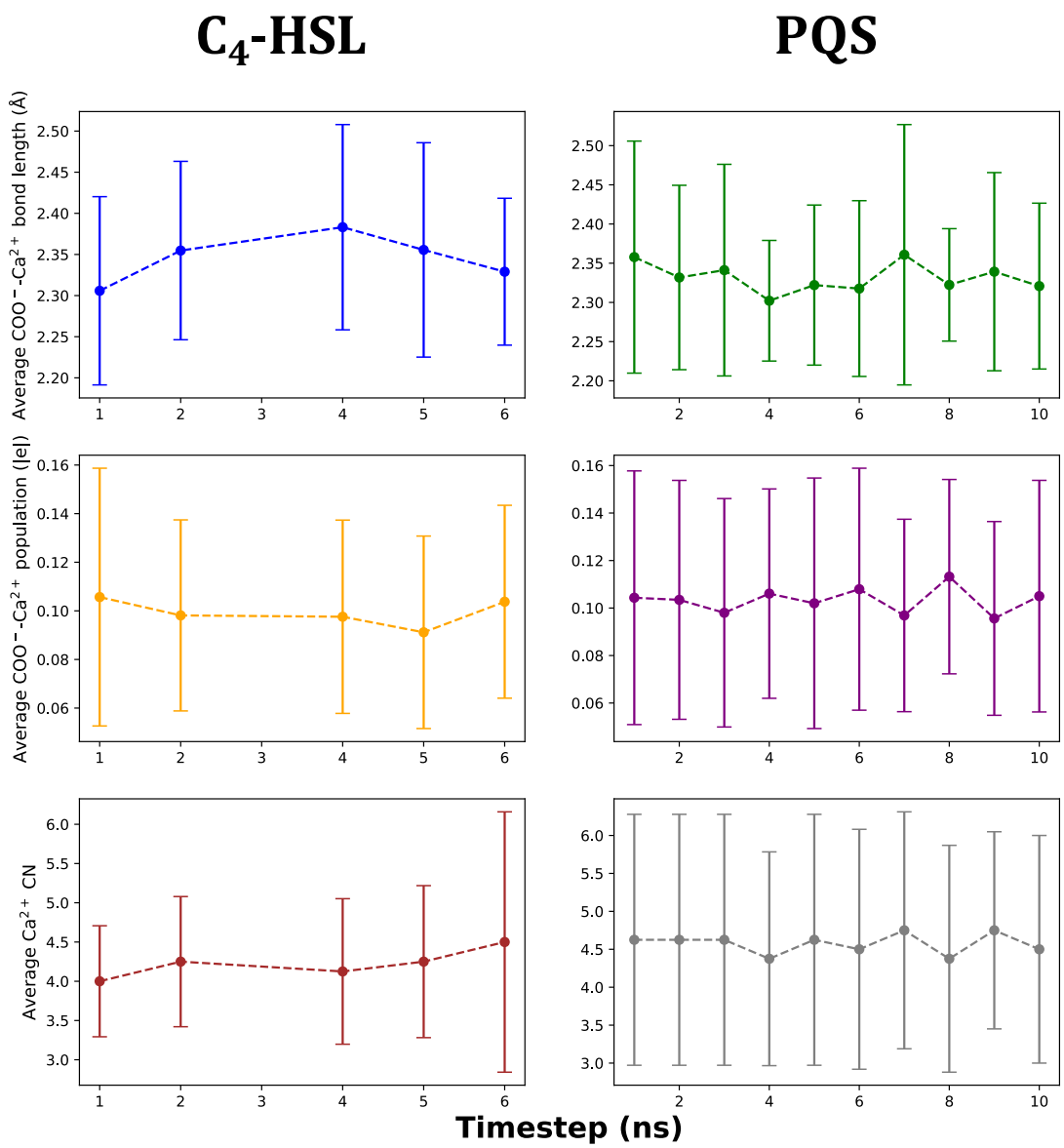


Figure 4.12: COO^{-}-Ca^{2+}average bond lengths (Å) and populations (|e|), along side the average Ca^{2+} CNs, over the course of the trajectory for the EPS + C_4-HSL and EPS + PQS adducts. All data points visualised as averages ± one standard deviation.

For both QSAI molecules, the average COO^{-}-Ca^{2+} bond lengths (Å) and populations (|e|) oscillate consistently within one standard deviation, as too does the Ca^{2+} ion CN, considering only the EPS donor atoms. Therefore, the presence of either QSAI molecule, bound or unbound, within the EPS vicinity does not give rise to a disrupted EPS matrix scaffold. Given that C_4-HSL shows no interactions with the EPS (**Fig 4.10**), and does not interfere with the EPS ionic architecture, namely, is not influenced by the cationic charge distribution in the EPS, it can be interpreted that this molecule exploits solely vdW interactions. This interaction is not able to render the molecule immobile at physiological temperature.

The EPS-C_4-HSL and EPS-PQS configurations that correspond to the minima in the total forcefield, configurational and electrostatic energies are shown, for reference, in **Fig 4.13**. Note, the same EPS-QSAI adduct structures exist in each of the three minima.

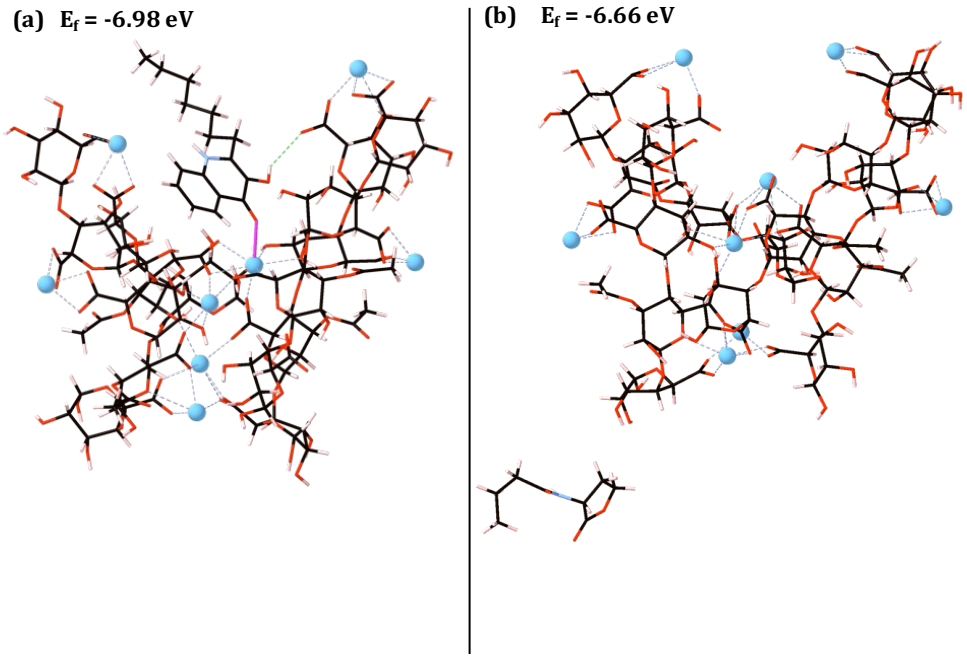


Figure 4.13: The EPS-PQS adduct (a) and EPS-C_4-HSL adduct (b) that correspond to the minima in the total forcefield, electrostatic and configurational energies. The associated formation energies are also given. Carbon atoms are shown in black, oxygen in red, calcium as blue spheres and hydrogen in pink. Calcium-oxygen ionic bonds are shown with dashed blue lines. The ionic bond between the PQS and 4-PolyMG$_{MD}$ system is shown with a bold pink line and the hydrogen bond shown as a dashed green line.

Firstly, considering C_4-HSL (**Fig 4.13b**), the distances between the molecule and the nearest EPS acetyl and EPS-bound Ca^{2+} ion are $\sim$ 16 Å and $\sim$ 19 Å respectively and, as such, the molecule can be considered well separated from the EPS in this configuration. It is important to note here, considering all of the isolated EPS-C_4-HSL configurations, that the most stable isolated EPS-C_4-HSL systems do not encompass the C_4-HSL molecule in the closest vicinity to the nearest acetyl group, nor do the least stable structures encompass the C_4-HSL molecule furthest away from the nearest acetyl group. The implication is that acetyl group proximity, whether this maximises hydrophobic interactions or maximises repulsion, does not dictate the overall thermodynamic stability of the EPS-C_4-HSL complexes. In light of the above, it is evident that the C_4-HSL molecule shows a clear disposition to separate from the EPS and migrate into solvent. In all of the thermodynamically favourable structures where C_4-HSL is within 6 Å of the EPS scaffold, no ionic or hydrogen bonding occurs and C_4-HSL binds

solely through vdW interactions. Importantly, there is a *thermodynamic rationale* behind the apparent unperturbed motion of this molecule through the EPS scaffold.

As can be seen in **Fig 4.13a**, the EPS-PQS adduct occupying the minima in the total forcefield, electrostatic and configurational energies encompasses the same ionic and hydrogen bonding molecular binding modes to the EPS as is present in other isolated adducts (**Fig 4.11**). From this structure, it is clear to see that the minima occurs prior to the H-bond rearrangement which occurs after 8 ns. This, in turn, reinforces the observation that, although critical to keep the PQS molecule irreversibly tethered under physiological conditions, the H-bond rearrangement lowers the overall stability of the EPS-PQS complex. This adduct also corresponds to the most thermodynamically favourable EPS-PQS system, further accentuating this molecule's significantly larger propensity to bind to the EPS and, importantly, underscoring the significant role electrostatic interactions play in driving EPS-PQS binding. It is important to note also, in this configuration, as well as all isolated configurations from the trajectory (**Fig 4.11**), that the PQS π system does not align in a configuration that ensures a Ca^{2+} is positioned above its respective plane. As such, the inference is that a quadrupole π-cation interaction is not a molecular interaction contributing significantly to the overall stability of the EPS-PQS complex. To better understand the length scale for a PQS quadrupole π-Ca^{2+} interaction, thus identifying its applicability in forming stable EPS-PQS adducts, the PQS molecule and a single Ca^{2+} ion were combined in isolation, with the Ca^{2+} ion positioned directly above the PQS π system at different separations, and a DFT single point energy calculation was performed using the same SCF tolerances used in **Section 4.2.1**. The PQS π-Ca^{2+} binding energy profile is given in **Fig 4.14**.

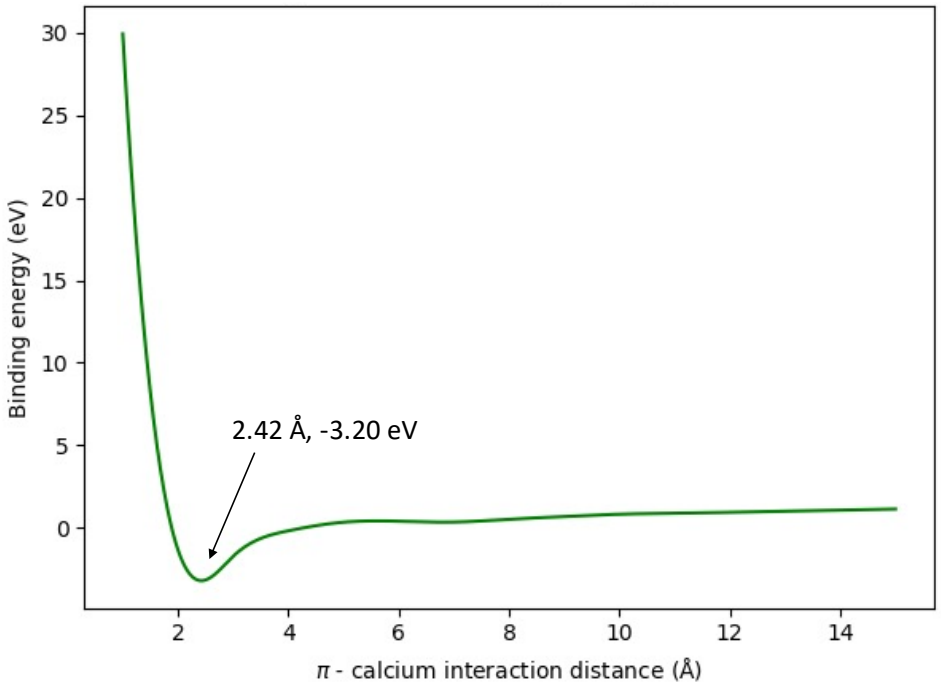


Figure 4.14: PQS π-Ca^{2+} binding energy profile with an annotation at the π-Ca^{2+} distance which gives rise to most stable binding.

As can be seen from the binding energy profile (**Fig 4.14**), the π-Ca^{2+} interaction is only stable at distances ≤ 4 Å and is most stable at 2.42 Å. The most thermodynamically favourable EPS-PQS structure has a π-Ca^{2+} separation of 5.2 Å, meaning that it is not an (electrostatic) interaction that contributes to the overall stability of the complex. Indeed, previous DFT estimations of π-cation interactions, across an array of different aromatic systems, have shown that N-heterocyclic aromatics and the presence of electron withdrawing groups, of which both are present in PQS, lower π-cation binding affinities [314]. This, in turn, helps to rationalise the observation that the PQS quadrupole π-Ca^{2+} interaction becomes unfavourable at relatively short length scales (≥ 4 Å).

4.5.2 Rationalising QSAI signal distributions

4.5.2.1 Solvent disposition

The relative hydrophobicity of C_4-HSL is smaller than PQS, the latter of which possesses conjugated aromatic rings and a larger hydrocarbon tail. The difference in hydrophobicity can be quantified by considering the $\log(P)$ for each molecule. Discussed in **Section 2.16** was how the *iLOP* multi-linear regression model [284], as implemented in SwissADME [286], offers a computationally efficient approach to approximate the $\log(P)$ values of small organic molecules.

The *iLOGP* predicted log(P) values of C_4-HSL and PQS are 1.47 and 3.28 respectively. It is, therefore, no surprise that the C_4-HSL molecule partitions more readily, compared to PQS, into the solvent space surrounding the EPS. Given that biofilms are heavily hydrated systems, with the major matrix component being water [315], this result suggests that C_4-HSL is better able to migrate into solvent filled void spaces (water channels) rather than maintaining interaction with the EPS. Now, in this circumstance, its transport is assisted by convection [316] and its diffusibility increases as a result of the decreased tortuosity of the water channels [317].

In light of this, additional long time-scale simulations of a *caged* C_4-HSL molecule, namely a C_4-HSL molecule entrapped within the cavity created by two inter-locking 4-PolyMG$_{MD}$ structures, visualised in **Fig 4.15**, were computed to test if permanent EPS-C_4-HSL interactions could be achieved from an extreme starting position within the EPS.

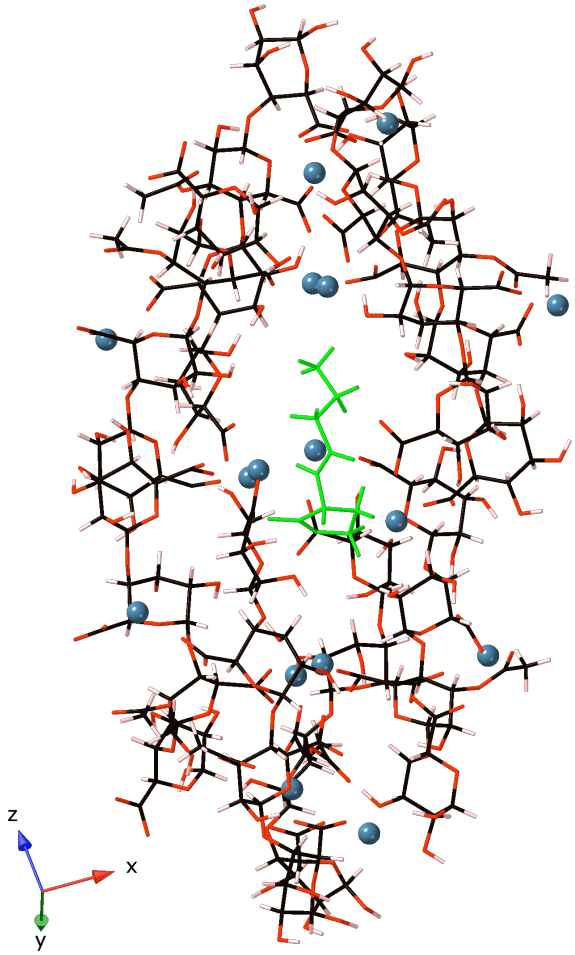


Figure 4.15: C_4-HSL (shown in green) entrapped within the cavity created by two inter-locking 4-PolyMG$_{MD}$ structures. Carbon atoms are shown in black, oxygen in red, calcium as blue spheres and hydrogen in pink.

MD simulations were performed under the same simulation regime outlined in **Section 4.2.1** in a simulation cell measuring 80 Å × 80 Å × 80 Å for 20 ns. Configurations isolated at 15 ns and 20 ns are given in **Fig 4.16**.

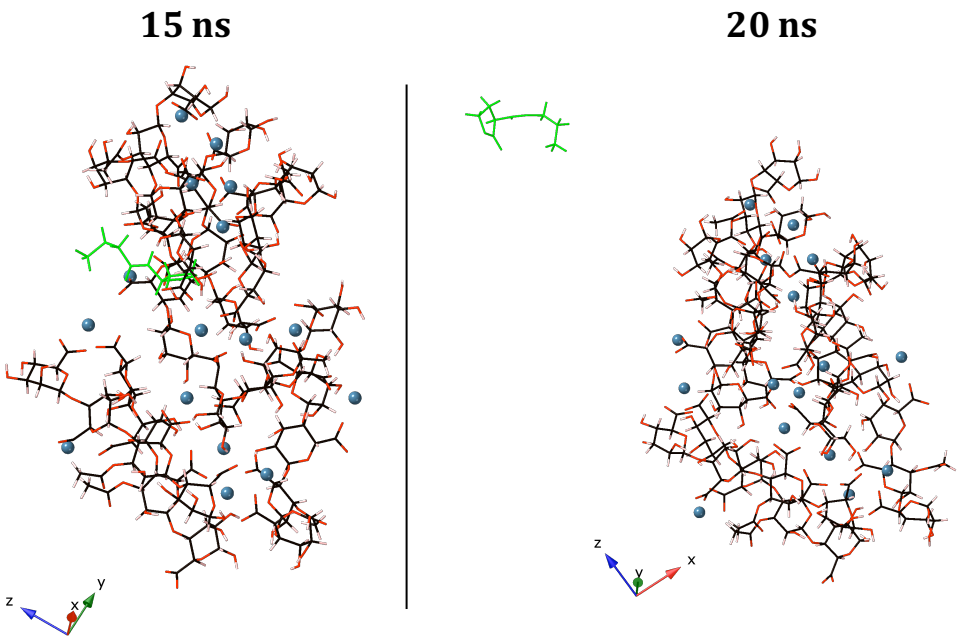


Figure 4.16: The 15 ns (left) and 20 ns (right) configurations isolated from the caged C_4-HSL trajectory. The C_4-HSL molecule is displayed in green for clarity. Carbon atoms are shown in black, oxygen in red, calcium as blue spheres and hydrogen in pink. Water molecules are not shown.

At 15 ns, the C_4-HSL molecule is beginning to breach the inter-locking EPS structures and at 20 ns is completely separated from the EPS. In the final configuration, the Ca^{2+} ions are retained at their native binding sites and, therefore, the C_4-HSL molecule has not liberated the EPS network through active disruption of the native ionic architecture. This, by virtue, is an extension of the observation made in the previous section, regarding the fact that the C_4-HSL molecule is not influenced by the cationic charge distribution within the EPS. Rather, the caged C_4-HSL molecule induces a widening of the EPS V-shaped motif, increasing the Ca5-Ca6-Ca1 angle (**Fig. 4.9**) which defines the V-shaped motif. This creates an aperture through which the C_4-HSL molecule can escape. Explicitly the EPS V-angle for the top and bottom EPS chains (**Fig 4.15**) widen by 29° and 20° respectively over the 20 ns trajectory. Again, this molecule does not tether to the EPS through electrostatic interactions, which as previously discussed, renders the molecule mobile in the EPS vicinity allowing the molecule to escape into the solvent space.

Given that the C_4-HSL molecule establishes solely vdW interactions with the EPS scaffold, the ring oxygen, ketone and amide functional groups are left free to participate in hydrogen bonding to the solvent atmosphere. To gauge whether these specific functional groups could indeed participate in stable hydrogen bonding interactions, the water binding energy at the available hydrogen bonding sites was evaluated for the C_4-HSL molecule as

well as the PQS molecule for reference. To perform these calculations, water molecules were positioned at the hydrogen bonding sites on the respective QSAI molecule and optimised to locate the low energy hydrogen bonded configurations. The water binding energy (E_b) for each optimised hydrogen bonded complex was evaluated using the following equation

$$E_b = E_{QSAI-water} - (E_{QSAI} + nE_{water}) \, . \tag{4.5.1}$$

The water bound QSAI complexes are given in **Fig 4.17** and the water binding energy (E_b), as well as the water binding energy per hydrogen bond, is given in **Table 4.6**.

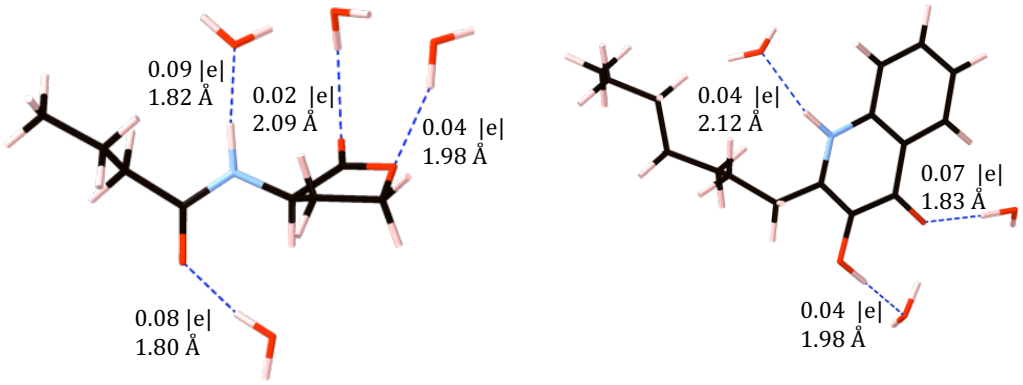


Figure 4.17: C_4-HSL water complex (left) and the PQS water complex (right). Carbon atoms are shown in black, nitrogen in blue, oxygen in red and hydrogen in pink. The bond lengths (Å) and Mulliken populations (|e|) for each hydrogen bond is also displayed

Table 4.6: The water binding energy (E_b) as well as the water binding energy per hydrogen bond for the C_4-HSL and PQS hydrogen bonded complexes.

System	E_b	E_b/H-bond
C_4-HSL	-2.57	-0.64
PQS	-2.37	-0.79

Both QSAI molecules return negative water binding energies and, as such, it is thermodynamically favourable for both molecules to establish hydrogen bonding interactions at their respective hydrogen bonding sites. The PQS forms slightly stronger hydrogen bonds with water (more negative E_b per H-bond) which can be attributed to its greater capacity to interact electrostatically. As already discussed, PQS is able to establish electrostatic interactions to the EPS in scaffold regions where C_4-HSL cannot. This is explored in greater detail in **Section 4.5.2.2**. Overall, the C_4-HSL molecule complexes in a more exothermic fashion with the surrounding water as it possesses more available hydrogen bonding sites. The C_4-HSL molecule returns a negative water binding energy, indicating that this molecule

is capable of using its ketone, ring and amide functional groups to bind water. In light of this, therefore, as these groups do not interact electrostatically with the EPS scaffold, they remain free to hydrogen bond to the solvent atmosphere to form thermodynamically favourable solvated complexes. This offers an additional thermodynamic rationale for why C_4-HSL migrates into the solvent space in such short time-scales.

Overall, the *caged* C_4-HSL simulations, along-side the water binding energy calculations, further exemplify this molecule's proclivity to migrate into water channels. This is a strong indicator that C_4-HSL can utilise water channels for long-range cell-to-cell signalling.

4.5.2.2 Global reactivity indices

Given in **Fig 4.18** are the DFT calculated Mulliken atomic charges for the ketone functional groups in the C_4-HSL and PQS structures following their equilibration under physiological conditions. The ring (head group) ketone functional group has a less negative atomic charge on the oxygen atom in C_4-HSL, compared to the ring ketone functional group on the PQS. This indicates a stronger potency for the ketone group in PQS to form ionic interactions relative to C_4-HSL. However, the C_4-HSL amide ketone group is as negative as the PQS ketone group which suggests an equal ionic binding potency in this case. EPS-C_4-HSL ionic binding facilitated by this particular ketone group however is not observed and, in fact, the EPS-C_4-HSL system is stabilised solely through vdW contacts.

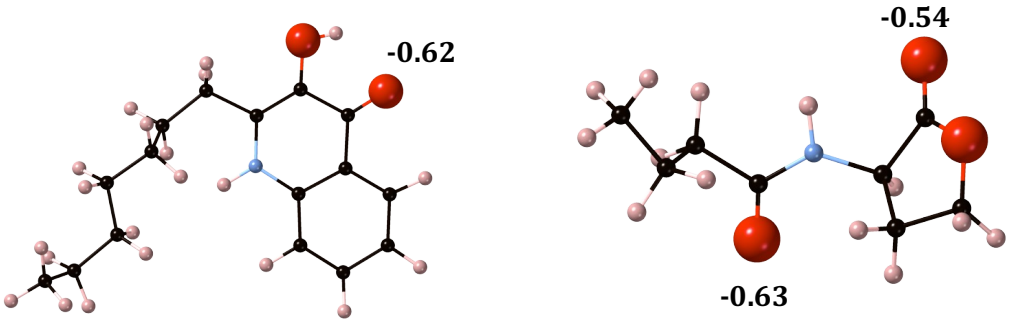


Figure 4.18: Structures and ketone atomic charges for the PQS (left) and C_4-HSL (right) structures following their equilibration under physiological conditions. Carbon atoms are shown in black, nitrogen in blue, oxygen in red and hydrogen in pink. Ketone Mulliken atomic charges (|e|) are also given.

To better interrogate this discontinuity, global reactivity indices for these two molecules were also computed. These are *molecular metrics*, stemming from the energy gap between the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) that allow inferences to be drawn regarding molecular reactivity. The first of these reactivity

indices is the chemical hardness

$$\eta = \frac{E_{LUMO} - E_{HOMO}}{2} \tag{4.5.2}$$

where, in terms of molecular orbital theory, the highest value of η reflects the highest possible energy gap between the LUMO and HOMO orbitals and this value correlates with the stability of a system [318]. A chemical system at a given temperature will evolve to a configuration of maximum absolute hardness, provided that the potential due to the nuclei, plus any external potential and the electronic chemical potential, remain constant. This introduces the principle of maximum hardness (PMH): for equilibrium geometries, the more stable isomers have greater chemical hardness values [319]. Interestingly, this is an observation that can be extended to reaction trajectories; during a reaction, hardness increases and, as such, molecules already *chemically hard* are less likely to undergo reactions [144, 320].

The second of these reactivity indices is the electronic chemical potential

$$\mu = \frac{E_{HOMO} + E_{LUMO}}{2} \tag{4.5.3}$$

which characterises the escaping tendency of electrons from the equilibrium system; electron density will flow from the region where μ is positive to regions where μ is negative until $\mu = \mu^0$ everywhere [321]. Conceptually, considering two systems A and B, if $\mu_B > \mu_A$ then electrons will flow from B to A. Finally, utilising both the chemical hardness and electronic chemical potential, a final reactivity index can be defined - the global electrophilicity index

$$\omega = \frac{\mu^2}{2\eta}. \tag{4.5.4}$$

The global electrophilicity index (ω) measures the capacity of a species to accept electrons [322]. In fact, as reactions have been observed to proceed to states of larger chemical hardness, the inverse is true for the electrophilicity index in which reactions tend to lower electrophilicity indices [323, 324]. This is referred to as the principle of minimum electrophilicity (PME) [325].

The global reactivity indices (**Equations 4.5.2 - 4.5.4**) for the thermally equilibrated C_4-HSL and PQS conformations are given in **Table 4.7**.

Table 4.7: Computed global reactivity indices for C_4-HSL and PQS.

Reactivity index (eV)	C_4-HSL	PQS
E_{HOMO}	-5.49	-4.68
E_{LUMO}	-1.79	-2.03
$\Delta E_{LUMO-HOMO}$	3.7	2.65
η	1.85	1.32
μ	-3.63	-3.36
ω	3.56	4.26

Immediately, it is evident that the energetic separation between HOMO and LUMO for PQS is smaller than for C_4-HSL. This is not unsurprising as the degree of unsaturation is larger in PQS, offering chromophoric properties. Indeed, PQS shows an intense UV-Vis absorption at 245 nm attributed to π to π^* transitions with an electronic transition dipole moment aligned along the axis of the aromatic ring system [326]. As a consequence, PQS is predicted to be more *chemically soft* and, thus, more polarizable. In other words it is more susceptible to changes in its electron density (and charge transfer). In addition, PQS possesses a larger electronic chemical potential (μ) that may give rise to a larger ionic binding potency, facilitating binding to EPS Ca^{2+} ions. As discussed in **Section 4.5.1**, it was observed that C_4-HSL is not effected by the cationic charge distribution within the EPS scaffold and, perhaps, the greater chemical hardness, lower electronic chemical potential and lower global electrophilicity index diminish the ability of C_4-HSL to form permanent electrostatic interactions with the EPS. Doubtlessly, the lower reactivity of C_4-HSL, compared to PQS, is not in violation of, rather is well described by, the principles of maximum hardness and minimum electrophilicity.

Potentially, the lower overall reactivity of C_4-HSL, compared to PQS, is responsible for establishing solely vdW contacts with the EPS in its most thermodynamically favourable configuration. The preference for establishing vdW, coupled with the molecules proclivity to migrate into water channels, provides a stronger motive force behind the C_4-HSL dynamics in the vicinity of the EPS. Overall, by forming no covalent or electrostatic interactions with the EPS and readily partitioning into water channels, C_4-HSL presents with inherent non-local behaviour, giving the molecule the ability to travel further distances throughout the biofilm.

4.5.2.3 Critical structural features

By forming electrostatic interactions with the EPS, the PQS molecule shows a strong preference for remaining in the EPS vicinity. PQS regulates virulence within mucoid *P. aeruginosa* CF biofilm infections through binding the transcriptional activator pqsR (**Fig 4.19**).

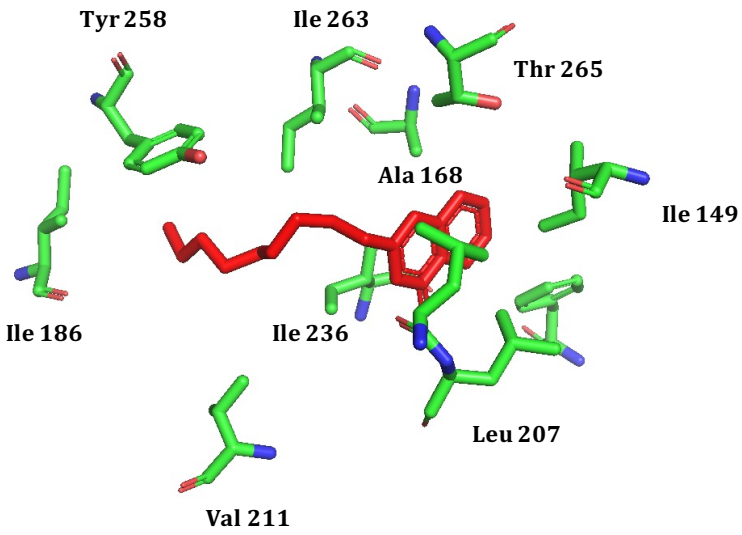


Figure 4.19: pqsR binding site pocket for PQS extracted from PDB entry 4JVD [327]. Note in this structure that NHQ, an alkyl quinolone structural analogue of PQS which lacks the OH group, and not PQS is complexed and highlighted red. The carbon backbone of the pqsR ligand binding site is shown green and oxygen atoms are shown in red and nitrogen in blue. Hydrogen atoms are not shown.

Structurally, PQS-pqsR binding occurs through two primary structural-chemical interactions: firstly, the PQS alkyl chain is buried within a hydrophobic pocket and, secondly, OH-Leucine (L207) hydrogen bonding [327]. Significantly, site directed mutagenesis, introducing mutations to the L207 region of the active site, inactivated PQS binding, confirming that the OH-Leucine (L207) hydrogen bond is critical for protein binding and bioisosteric substitutions at the OH site reduce PQS-pqsR binding affinity [327]. The simulations in this work also call critical attention to the OH group, in addition to the ketone group, for mediating molecular electrostatic interactions with the EPS and, furthermore, these simulations would suggest that PQS is a short range QSAI signal not received community-wide.

In fact, this analysis complements recent secondary ion mass spectroscopy and confocal Raman microscopy studies into spatial distributions of alkyl quinolones in *P. aeruginosa* biofilms, which also expose PQS as a local QSAI signal [328, 329]. AQNO is an alkyl quinolone structural analogue of PQS in which the ketone and hydroxyl groups are absent and which is also present in late stage *P. aeruginosa* biofilm colonies [330]. Compared to PQS, AQNO is capable of distributing further afield from the source bacteria, specifically reaching ranges of 400 μm compared to the 150 μm range achieved by PQS [329]. The larger spatial distribution of AQNO, which lacks the key binding functionality responsible for PQS tethering, offers a critical link, reinforced by our simulations, between the PQS ketone and hydroxyl groups and reduced movement throughout the EPS. Indeed, local PQS distribution

patterns have been similarly observed using confocal Raman spectroscopy, which enabled the measurement of 2-D xy spatial PQS distributions, identifying that PQS is localised near its bacterial source [331]. In fact, the ability of the PQS molecule to localise and tether to the EPS has survival advantages where the localisation of PQS at the boundary between infected and non-infected sub-populations, within viral bacteriophage-infected *P. aeruginosa* biofilms, behaves as a warning signal allowing non-infected bacteria in the immediate neighbourhood to avoid danger [332]. The survival applications of PQS cannot be extended to C_4-HSL as only PQS can be selectively retained by the EPS and, as these simulations outline, the two oxygen bearing functional groups are of paramount importance to facilitate this.

The C_4-HSL and PQS simulations, whilst exposing PQS as immobile relative to C_4-HSL over the same time period, allow for estimations on the diffusion coefficients of these two molecules relative to one another. It is important to note here, that to accurately capture diffusion coefficients from molecular dynamics simulation, one would require simulation length scales on the order of μs and not ns. That being said, it can be proposed that the diffusion coefficient for PQS through the mucoid *P. aeruginosa* biofilm EPS matrix would be smaller than that of C_4-HSL due to its greater EPS binding affinity. What this translates to *in vitro*, is the ability for C_4-HSL to travel distances on the order of mm away from the QSAI emitting bacterial source [177], whereas PQS only sustains travel on the order of μm away from the QSAI emitting bacterial source [328–331], in *P. aeruginosa* biofilms cultured over 48 h. Evidently, PQS cannot travel as far away from the source as C_4-HSL due to its far higher tendency to interact electrostatically with the EPS.

Although, on the onset it should appear deceptively simple to compare *in vitro* diffusion coefficients for these two molecules through the EPS matrix, however, diffusion coefficients are scarcely reported in the literature. A possible explanation for this, is that determination and comparison of diffusion coefficients can be made difficult due to the inability to control QSAI concentration in the biofilm matrix as well as biofilm matrix molecular composition. A potential experiment to address this limitation would be to supplement a polymeric matrix of solely EPS material of a known viscosity with exogenous QSAI molecules of an exact and identical concentration, monitoring the concentration of QSAI which appears in a particular region of the matrix system. What follows, using Fick's first law of diffusion, is the relationship between the amount of QSAI that has permeated the matrix and time [333]

$$\frac{Q}{A} = \frac{DK_pCt}{h} \ . \tag{4.5.5}$$

Equation 4.5.5 is taken from the works of Sriamornsak *et al.,* when studying the diffusion of theophylline, a phosphodiesterase inhibitor used in the treatment of respiratory diseases, through calcium alginate gels [333]. In **Equation 4.5.5**, Q is the amount of QSAI which

has diffused per unit area A (mg cm^{-2}), D is the diffusion coefficient (cm^2 s^{-1}), K_p is the partition coefficient, C is the concentration of QSAI in the a particular region of the matrix system (mg cm^{-3}), t is time (s) and h is the distance travelled through the matrix (cm). Given how **Equation 4.5.5** can be used to quantify diffusion of organic molecules in calcium alginate gels, it is understandable that it may also have applicability to studying the motion of QSAI molecules through the EPS.

Quorum quenching (QQ) is emerging as a novel strategy to limit biofilm proliferation through reducing the concentration of QSAI molecules in the biofilm matrix, for example, via enzymatic degradation [334]. For QQ to be effective, the QSAI molecules must be immobile and easy to sequester and, as such, QQ fails to have any impact on QSAI molecules encompassed within an aqueous phase where the QSAI molecule's motion is driven by convection [316]. Therefore, given the immobility of PQS relative to C_4-HSL, which can partition into the solvent medium (void spaces), this work proposes PQS as a viable QQ target. In fact, enzymatic PQS deactivation, for example, through exogenous supplementation of 2,4-dioxygenase, has proved to be a viable strategy for eliminating PQS and PQS related virulence from *P. aeruginosa* biofilms *in vitro* [335, 336].

In addition, the electrostatically induced tethering to the EPS, facilitated by the ketone and hydroxyl functionality, gives PQS the inability to propagate throughout the EPS matrix and, thus, underscores the requirement for this molecule to be packaged in outer membrane vesicles (OMVs) if it is to maximise its effectiveness as a (long-range) cell to cell signal. Despite this, it is important to note that not all PQS molecules will be encompassed within OMVs as it must mediate its own OMV packaging [313, 337] and is ineffective at performing its virulence functions whilst packaged; the biological activity of PQS is associated with the free molecule and not the packaged form [337]. Indeed, OMVs have been implicated in the transportation of PQS, but not the transportation of C_4-HSL [337] and these MD simulations answer the question, with regards to intermolecular interactions at the atomic-scale, as to why this is the case.

Finally, it is interesting to note that the most thermodynamically favourable EPS-C_4-HSL system, encompassing C_4-HSL held in the EPS vicinity solely through vdW interactions, and PQS-EPS system, encompassing PQS ionically tethered to the EPS, are comparable in thermodynamic stability.

4.6 Summary

This chapter probed the origin of molecular interactions between the two quorum sensing autoinducer (QSAI) molecules and the mucoid *P. aeruginosa* exopolysaccharide (EPS)

matrix, with the aim of rationalising which molecular interactions govern molecular motion throughout the EPS matrix. To accurately model EPS-molecule interactions, a combined molecular dynamics (MD) and Density-Functional Theory (DFT) approach was needed. The former allows for the identification of unique modes of interaction between the EPS and the invading molecule, which occur under physiological conditions, with the latter allowing for the quantification of the stabilities of these unique interaction modes.

The aim was to capture EPS-molecule interactions at longer EPS molecular length scales which, naturally, begins with using DFT to develop a large 4-chain EPS molecular model. In so doing, it was possible to identify key geometrical features that most stably facilitate exothermic multi-EPS chain association. Specifically, these calculations identified that at least two Ca^{2+} ions are needed for the aggregation of four EPS chains and that acetylated copolymeric β-D-mannuronate-α-L-guluronate 4-chain structures are able to facilitate tight complexation about Ca^{2+} ions when the acetyl groups are oppositely displaced. Furthermore, this structure can be distinguished from the less stable acetylated poly-β-D-mannuronate 4-chain analogue through the absence of acetyl contributions to the Ca^{2+} ion chelation geometry. The most stable Ca^{2+} cross-linked 4-chain EPS model possessed egg-box characteristics which were absent at the 2-chain level.

The most stable Ca^{2+} cross-linked 4-chain EPS model was transformed using finite temperature explicit solvent MD, to a molecular model more representative of that observed at physiological temperature and in the presence of water. Physiological temperature and water reduced the diversity in the Ca^{2+} EPS chelation geometries, promoting the carboxylate group to being the dominant contributor to the Ca^{2+} coordination shell over all other oxygen functionality. The reduction in the diversity in the Ca^{2+} EPS chelation geometries gave rise to a reduced 4-chain intracomplex space, influencing the adoption of a discontinuous, V-shaped, dendritic morphology. This *physiologically representative* molecular EPS model was combined with C_4-HSL and PQS, in a combined MD-DFT theoretical approach to provide atomic-scale chemical insight into the functional groups required for exothermic EPS adsorption.

The C_4-HSL molecule exploits only vdW interactions with the EPS, which are insufficient to render the molecule immobile at physiological conditions and, coupled with its lower hydrophobicity relative to PQS, enable the molecule to readily partition into bulk solvent and void spaces. In stark contrast, electrostatic interactions are a critical driving force behind PQS adsorption onto the EPS scaffold. The PQS molecule has the ability to form thermodynamically favourable ionic complexes with EPS-bound Ca^{2+} as well as establishing a hydrogen bond directly to a single EPS chain. The PQS hydroxyl group is central for mediating irreversible binding to the EPS under physiological conditions and

the PQS molecule is rendered immobile through EPS binding. As such, these simulations support the observation that outer membrane vesicles are required to maximise the effectiveness of PQS as a long range cell to cell signal. Finally, with significantly reduced EPS mobility, the PQS molecule is identified as a potential target for quorum quenching (QQ).

Overall, it is clear that electrostatic interactions play a vital role in driving exothermic EPS adsorption. Therefore, the following two considerations can be put forward regarding therapeutic development:

- When considering deep matrix penetration, designing molecules that are not excessively hydrophobic, exploit solely vdW interactions with the EPS scaffold and can exploit solvent disposition may lead to more active therapeutic outcomes.
- When considering the design of quorum quenchers, creating a QQ agent that is selective to PQS would be more efficient, as C_4-HSL has the molecular characteristics that give rise to increased matrix mobility and, as such, are more capable of avoiding QQ agents.

Finally, the model created and deployed here represents the major mucoid *P. aeruginosa* CF lung biofilm matrix component. As such, the molecular interactions between the QSAI's and the EPS, which occur at physiological equilibrium, captured in these simulations, would also occur in models of larger sizes. In conclusion, these models and simulations provide critical molecular insight into molecular motion through the EPS.

The physiological EPS model (4-PolyMG$_{\mathrm{MD}}$) constructed in this chapter is a suitable model system to probe the therapeutic mechanism of OligoG CF-5/20 against the mucoid *P. aeruginosa* EPS at the atomistic-scale. This mechanism is currently unknown at the atomistic and molecular levels and the following chapter will detail the mechanistic insights gained through 4-PolyMG$_{\mathrm{MD}}$ + OligoG CF-5/20 combined simulations.

Therapeutic mechanism of OligoG CF-5/20

5.1 Scope of the chapter

In **Section 1.5.1** OligoG CF-5/20 was introduced as an emerging, novel, anti-Pseudomonal agent targeted at disrupting *P. aeruginosa* biofilms. Although biofilm matrix disruption effects on bulk living biofilms *in vitro* and *in vivo* have been observed, the disruption mechanism has only been explored at an atomistic level on eDNA structures. The EPS, rather than eDNA, is the primary constituent of the mucoid *P. aeruginosa* CF lung biofilm matrix and, as such, the atomistic disruption mechanism against such biofilms is incompletely unknown.

The aim of this chapter is to use the previously defined physiological EPS atomistic model (4-PolyMG$_{\mathrm{MD}}$) to study the interactions established between OligoG CF-5/20 and the EPS, elucidating the OligoG CF-5/20 therapeutic mechanism against the mature mucoid EPS scaffold at the atomic-scale.

5.2 Computational details

5.2.1 Simulation parameters

All molecular dynamics trajectories were computed using DLPOLY4 [282] and the conversion of the molecular models into DLPOLY input files was performed using DLFIELD [306]. MD trajectories were computed using the OPLS2005 forcefield [274, 275] in the canonical (NVT) ensemble, where the RATTLE algorithm [283] was used to constrain covalent bonds to hydrogen atoms, which allowed the integration time-step to be set to 2 fs. The temperature was held at 310 K (body temperature) using Langevin thermostatting [279]. Electrostatics were treated using the Smooth-Particle-Mesh-Ewald method [307] and the distance cut-offs for electrostatic and Leonard-Jones interactions were set to 1.2 nm.

All Density Functional Theory (DFT) calculations were performed using the plane-wave Density Functional Theory (DFT) code, CASTEP [280]. A convergence tested cut-off energy of 900 eV was employed, as well as a Monkhorst-Pack **k**-point grid of 1 × 1 × 1 to sample the Brillouin zone [255]. On the-fly ultrasoft pseudopotentials were used [281] alongside the PBE exchange-correlation functional [240]. Intra- and intermolecular dispersive forces were accounted for by applying the semi-empirical TS dispersion correction [261]. Zero Kelvin (0 K) in vacuo DFT single-point energy calculations were performed using an SCF tolerance of 2 × 10^{-6} eV $Atom^{-1}$. Mulliken bond populations [270] were calculated to classify the nature of bonding between the EPS and the Poly-G_4 in each of the final EPS-Poly-G_4 adducts. The thermodynamic stability of the isolated EPS-Poly-G_4 adducts was evaluated by means of evaluating a formation energy (**Equation 5.2.1**)

$$E_f = E_{EPS-Poly-G_4} - (E_{EPS} + E_{Poly-G_4}) \tag{5.2.1}$$

where $E_{EPS-Poly-G_4}$ is the energy of the EPS-Poly-G_4 adduct, E_{EPS} is the energy of the thermally equilibrated 4polyMG$_{MD}$ system and E_{Poly-G_4} is the energy of the thermally equilibrated Poly-G_4 molecule (**Fig 5.1**).

5.2.2 Defining an OligoG CF-5/20 molecular model

Here, a guluronate quadramer (Poly-G_4) was chosen as a model to represent OligoG CF-5/20. It possesses the correct guluronate abundance, is on a 1:1 length scale with the EPS molecular model, and allows for tractability when calculating the thermodynamic stabilities from DFT. The Poly-G_4 molecule was thermally equilibrated, under the same conditions as C_4-HSL and PQS (see **Section 4.2.4**), to obtain a representative *in vivo* conformation. The equilibrated Poly-G_4 (**Figure 5.1**) encompasses a torsion profile (mean values: ϕ = -73°, ψ = -149°) in good agreement with the torsion profile predicted for the ground state conformation of acidic β-1-4 linked guluronate disaccharides (ϕ = -82°, ψ = -155°), obtained from MM3 forcefield based molecular mechanics minimisations, in the work of *Braccini et al.,* [295]. Reassuringly, both the MM3 and OPLS2005 forcefields are able to accurately reproduce the geometries of biomacromolecules and small organic molecules [338]. The thermally equilibrated Poly-G_4 molecule was positioned 6 Å away from the base of the V-shaped cleft in the EPS structure, a distance sufficient to prevent any initial hydrophobic and/or electrostatic interactions between the Poly-G_4 molecule and the EPS. Subsequent EPS-Poly-G_4 simulations were performed in SPC water, under periodic boundary conditions, in a simulation cell measuring 60 Å × 60 Å × 60 Å, over 10 ns. This trajectory length was sufficient to allow the Poly-G_4 molecule to sample preferential binding sites with the EPS and adopting an equilibrium binding configuration.

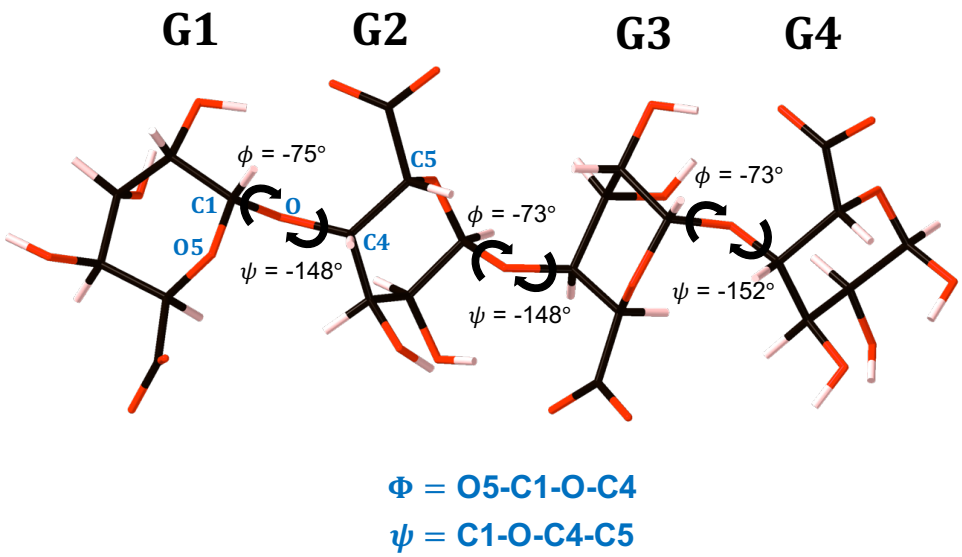


Figure 5.1: The structure, guluronate unit nomenclature and torsion parameters of the Poly-G_4 system after 1 ns equilibration under physiological conditions. Carbon atoms are shown in black, oxygen in red and hydrogen in pink. Definition of ϕ and ψ angles are also given.

5.3 Exothermic OligoG CF-5/20 dispersal into the mucoid EPS

The Poly-G_4 system binds to the EPS scaffold, in a thermodynamically favourable fashion, early (within 2 ns) in the trajectory and remains bound for the full time-scale of the trajectory. As such, thermodynamically favourable EPS-Poly-G_4 adducts are displayed, for brevity, at 2 ns intervals in **Fig 5.2**.

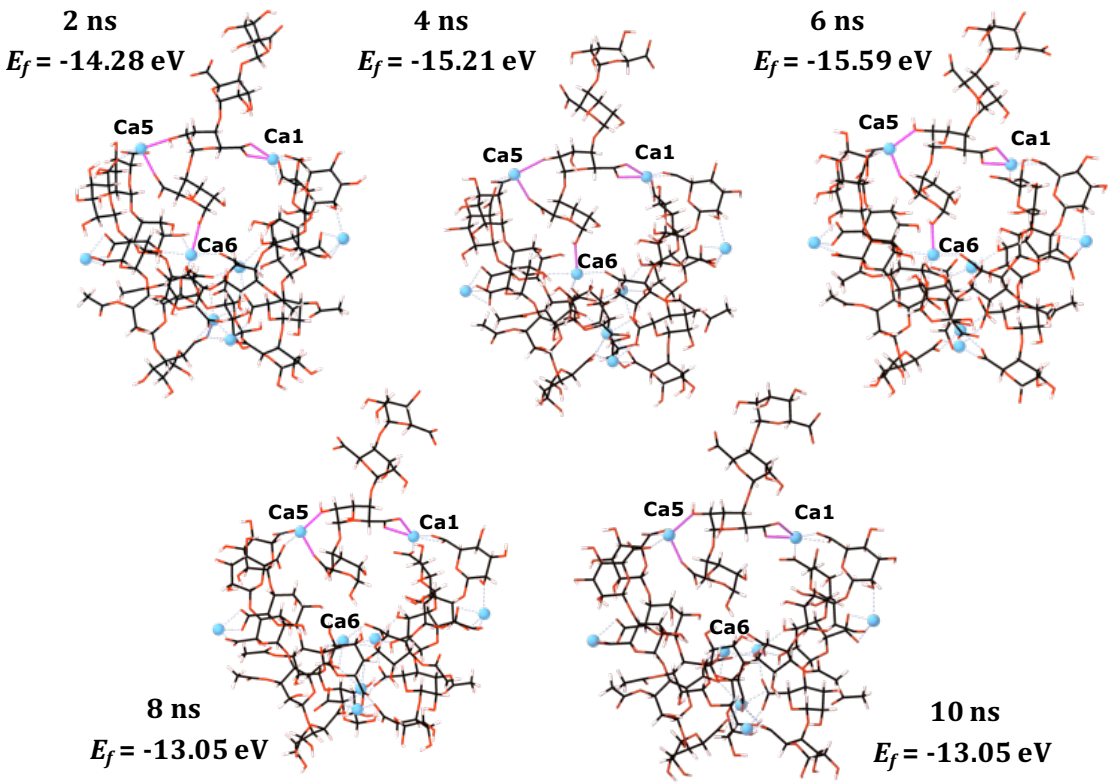


Figure 5.2: Bound EPS-Poly-G_4 adducts, along with their formation energies, displayed every 2 ns. Carbon atoms are shown in black, oxygen in red, calcium as blue spheres and hydrogen in pink. Calcium-oxygen ionic bonds are shown with dashed blue lines and calcium ions implicated in binding of Poly-G_4 are labelled. Ionic bonds between the Poly-G_4 and EPS system are shown with bold pink lines.

As can be seen, for each adduct (**Fig. 5.2**), **Equation 5.2.1** returns a negative formation energy, indicating that the Poly-G_4 dispersal into the EPS is exothermic. The exothermic nature of Poly-G_4 binding is driven through ionic bonding to EPS-bound Ca^{2+} through Poly-G_4 carboxylate and hydroxyl functional groups. The interaction between OligoG CF-5/20 and mucin, as observed in previous MD simulations, is also mediated through the hydroxyl functional group. Specifically, the hydroxyl group exhibits hydrogen bonding interactions to N and/or O atoms within the amide linkages of the mucin peptide backbone [181] highlighting the role of the hydroxyl group in EPS tethering. However, in contrast, the Poly-G_4 molecule in our simulations does not show any hydrogen bonding interaction with the EPS chains and all interactions are facilitated by bridging Ca^{2+} ions. With regards to mucoid *P. aeruginosa* CF matrixome components (eDNA and LPS), FTIR and circular dichroism measurements have also implicated Ca^{2+} ions as the main facilitator of OligoG CF-5/20 binding [183, 191].

Upon EPS tethering, the Poly-G_4 molecule remains tethered over time, does not liberate the cleft and only experiences minor conformational changes ($\leq 5°$) to its uronate backbone. This

indicates that ionic association to the Ca^{2+} ions in the EPS significantly reduces the molecule's mobility and flexibility. Essentially, upon binding to EPS bound Ca^{2+} ions, the Poly-G_4 system is rendered irreversibly bound and immobile. The relative immobility of the tethered Poly-G_4 molecule is best exemplified by considering the global RMSD of all adducts, following structural alignment, as well as considering the torsion changes in ϕ and ψ angles about the G3-G4 junction (**Figure 5.1**), the guluronate junction involved in calcium binding and EPS tethering. Through alignment of the 2 and 10 ns adducts, the global RMSD is calculated to be 1.18 Å and, interestingly, considering the Poly-G_4 backbone in isolation of the EPS, the 2-10 ns RMSD is also 1.18 Å. The change in ϕ and ψ angles across the G3-G4 junction, where the ionic O-Ca^{2+} tethers are established, are 5° and 4° respectively over the same time period. As a comparison, the global RMSD following the alignment of the pre-equilibrated and post-equilibrated Poly-G_4 structures is 4.24 Å, with maximum torsion changes in ϕ and ψ angles of 43° and 44° respectively - visibly highlighting the larger flexibility of the free molecule.

The average COO^{-}–Ca^{2+} and HO-Ca^{2+} bond lengths and Mulliken populations of the EPS-Poly-G_4 tethering contacts, displayed at 2ns intervals, are given in **Tables 5.1** and **5.2** along side their graphical representation in **Fig 5.3**. These data highlight the sustained stability of the EPS-Poly-G_4 tethering contacts over the time course of the 10 ns trajectory.

Table 5.1: Average EPS-Poly-G_4 O-Ca^{2+} bond lengths (Å) calculated at 2 ns intervals. Note, from 8 ns only a single Poly-G4 HO-Ca^{2+}+ interaction exists.

Time (ns)	Average COO-Ca^{2+} length (Å)	(±) std	Average OH-Ca^{2+} length (Å)	(±) std
2	2.29	0.01	2.42	0.06
4	2.23	0.04	2.52	0.06
6	2.23	0.04	2.38	0.05
8	2.26	0.08	2.37	-
10	2.29	0.08	2.42	-

Table 5.2: Average EPS-Poly-G_4 O-Ca^{2+} Mulliken populations (|e|) calculated at 2 ns intervals. Note, from 8 ns only a single Poly-G_4 HO-Ca^{2+}+ interaction exists.

Time (ns)	Average COO-Ca^{2+} population (|e|)	(±) std	Average OH-Ca^{2+} population (|e|)	(±) std
2	0.11	0.05	0.08	0.01
4	0.10	0.05	0.09	0.02
6	0.11	0.06	0.09	0.01
8	0.12	0.06	0.06	-
10	0.11	0.06	0.08	-

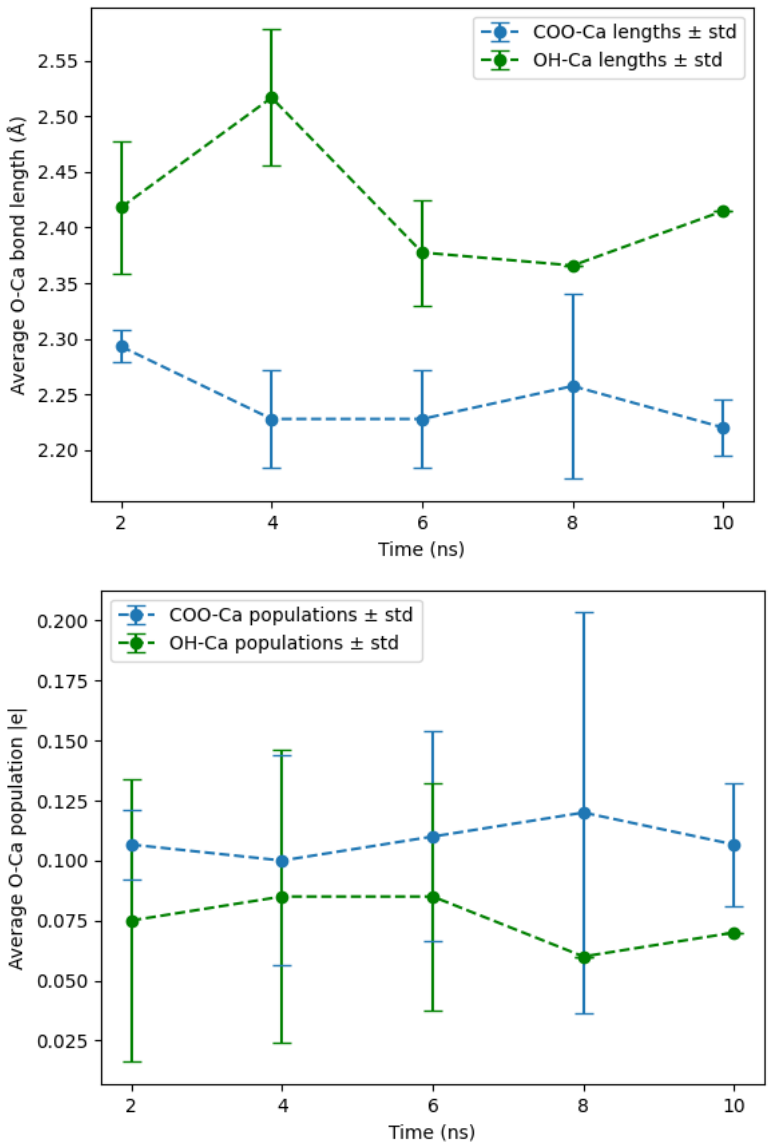


Figure 5.3: A graphical representation of the geometrical data encompassed within **Tables 5.1** and **5.2**. Top, the average O-Ca^{2+} bond lengths (Å), and bottom, the average O-Ca^{2+} Mulliken populations (|e|), for the Poly-G_4 hydroxyl and carboxylate tethering contacts over the course of the trajectory. Error bars correspond to ± one standard deviation. Note, the standard deviation for the OH tethers at 8 and 10 ns is zero as only a single Poly-G4 OH-Ca^{2+} interaction exists.

With Mulliken populations ≤ 0.12 |e|, it is clear to see that, within each structure, all O-Ca^{2+} tethers are ionic in nature. The Poly-G_4 carboxylate groups establish ionic interactions with EPS-bound calcium and possess shorter average bond lengths and larger average populations, relative to the hydroxyl tethers, indicating that these groups are implicated in establishing the tightest and most stable ionic interactions with the EPS-bound calcium. The tighter binding of the carboxylate tethers is a structural characteristic retained over the 10 ns trajectory

where, considering all COO^{-}-Ca^{2+} and HO-Ca^{2+} tethers averaged over the full course of the trajectory, the carboxylate tethers are 0.17 Å shorter (2.25 Å) compared to the hydroxyl tethers (2.42 Å). Furthermore, both monodentate and bidentate carboxylate coordination modes are observed during the Poly-G_4 tethering and these respective modes, specifically the monodentate COO^{-}-Ca5 and bidentate COO^{-}-Ca1, remain in place over the course of the trajectory (**Fig. 5.2**). Both the COO^{-}-Ca^{2+} and HO-Ca^{2+} tethers retain consistency in their lengths and Mulliken populations, with their respective averages fluctuating within one standard deviation over the 10 ns trajectory (**Fig 5.3**). The tethering data show that upon Poly-G_4 binding, established hydroxyl and carboxylate tethers retain their stability over 10 ns, the only exception being HO-Ca6, which is present throughout the first 6 ns, but afterwards is lost and not present in the final structure. Between 6 and 8 ns, Ca6 sinks deeper into the V-shaped cleft, losing its interaction with the invading Poly-G_4 OH group (**Fig 5.4**). The depression of Ca6 into the V-shaped cleft increases the HO-Ca6 distance by 1.63 Å, positioning the Ca6 out of the hydroxyl ionic bonding range. The depression of Ca6 is rationalised by considering changes to the EPS V-angle over the course of the trajectory, explicitly, the Ca5-Ca6-Ca1 angle (**Fig. 5.2**) which defines the V-shaped motif given in **Table 5.3**.

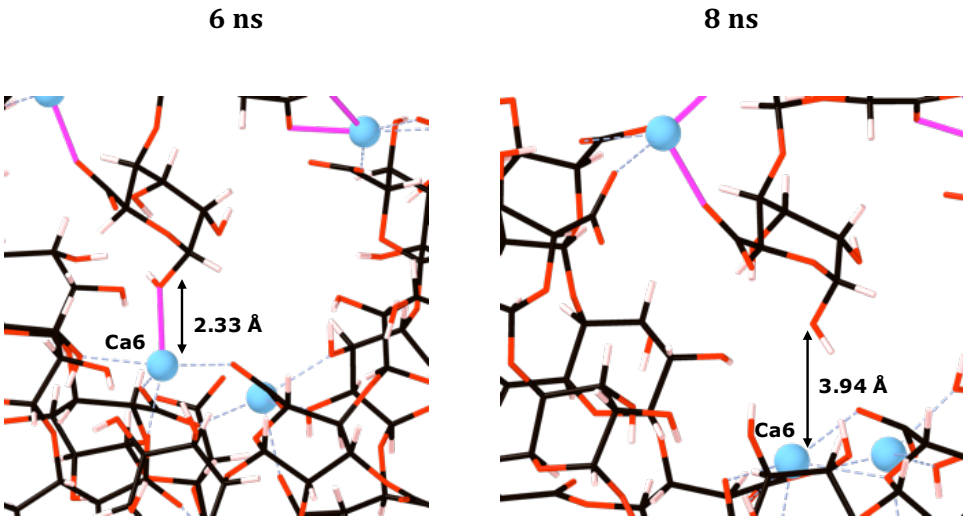


Figure 5.4: A close up of the bound EPS-Poly-G_4 adducts at 6 and 8 ns, highlighting the depression of Ca6 into the EPS cleft. Carbon atoms are shown in black, oxygen in red, calcium as blue spheres and hydrogen in pink. Calcium-oxygen ionic bonds are shown with dashed blue lines and ionic bonds between the Poly-G_4 and EPS system are shown with bold pink lines.

Table 5.3: The EPS V-angle, namely, the angle formed between Ca5-Ca6-Ca1, which defines the EPS V-shaped motif, calculated at 2 ns intervals.

Time (ns)	EPS V-angle (°)
2	66
4	69
6	68
8	60
10	60

From **Table 5.3** it can be seen that the EPS V-angle only exhibits minor variations between 2 and 6 ns, fluctuating between 66°-69°. However, post 6 ns, there is a contraction in the V-angle to 60°, effectively further encapsulating the Poly-G_4 and, as a consequence, allowing Ca6 to sink deeper into the cleft, moving away from the invading Poly-G_4 OH group. The HO-Ca5 interaction does survive the full time-scale of the trajectory, however, it only does so because Ca5 is held in its position through the invading monodentate COO^--Ca5 interaction - recall, the tethering contacts retain their stability over the 10 ns (**Tables 5.1** and **5.2**). From this observation, it can be inferred that the hydroxyl tethering contacts are not primarily responsible for keeping OligoG CF-5/20 irreversibly bound to the mucoid *P. aeruginosa* biofilm EPS. Their weakness, relative to the carboxylate tethers, means that these contacts are less robust and less likely to survive global conformational change, such as the contraction of the EPS V-angle. Rather, it is the carboxylate group, more so than the OH group, that facilitates the primary molecular mechanism of action over longer time-scales.

The formation energies (-13.05 eV), Poly-G_4 binding geometry and EPS V-angle (60°) at 8 ns and 10 ns are identical. These particular configurations encompass the same site-directed calcium chelation modes observed both experimentally and theoretically when OligoG CF-5/20 is dispersed within CF airway mucin [181, 184] and eDNA [183], as well as *P. aeruginosa* EPS scaffolds [339]. This is a strong indicator that the Poly-G_4 system has adopted an equilibrium binding configuration with the solvated EPS at 310 K within 10 ns. This matches the time-scale required to identify an equilibrium binding geometry between OligoG CF-5/20 and CF airway mucin detailed in previous MD simulations [181]. In addition, the minor deviation in RMSD (1.18 Å) between the 2 ns and 10 ns bound states indicates, potentially, that the Poly-G_4 molecule has already adopted a binding configuration close to equilibrium within 2 ns. In fact, low energy, conformationally static, equilibrium calcium binding configurations adopted upon complexation with a single polyguluronate chain have been shown to occur within time-scales as low as 0.8 ns [138, 139] and 6 ns [312, 340] for 10-mer and 30-mer uronate length scales respectively.

The formation energies for the EPS-Poly-G_4 adducts at 2, 4 and 6 ns indicate that these structures are more stable relative to the adducts at 8 and 10 ns. The difference between these is the presence of the OH-Ca6 interaction. Therefore, although HO-Ca6 bond is an ionic interaction that does not survive the full time-scale of the trajectory, it assists in stabilising the overall EPS-Poly-G_4 binding geometry. Therefore, with regards to the structural chemistry that underpins exothermic Poly-G_4 dispersal into the EPS, the hydroxyl group, although not modulating irreversible binding, modulates overall complex stability. In previous MD simulations, investigating the modes of EPS and eDNA binding over a 50 ns time period, the participation of the hydroxyl group in binding biofilm Ca^{2+} ions was absent [183]. These simulations, therefore, highlight for the first time the role of the hydroxyl group in mediating exothermic penetration into the EPS scaffold.

Finally, it is important to note that as the formation energy, Poly-G_4 binding geometry and EPS V-angle at 8 and 10 ns are identical - an equilibrium complex configuration has been reached. It is, therefore, anticipated that this particular EPS conformational arrangement impedes the kinetic behaviour of the OH-Ca6 interaction, specifically, preventing the re-establishment of this tethering contact post 10 ns.

5.4 Testing the influence of uronate length scale

As the Poly-G_4 molecule is on a 1:1 length scale with the EPS model, it is possible that its immobility in the EPS vicinity may arise as a result of its comparatively large size. In essence, the Poly-G_4 molecule is stationary (immobile) in the EPS vicinity because it cannot diffuse past the EPS scaffold, which is of the same length in our simulations. To test whether the immobility of the Poly-G_4 system, upon EPS binding, is due to its size, the Poly-G_4 system size was halved to a dimer: Poly-G_2. This truncated system was thermally equilibrated in SPC water under the same conditions and combined with the EPS in the same fashion as the Poly-G_4 system (see **Section 5.2.2**). The thermodynamic stability of the EPS-Poly-G_2 adducts was assessed using **Equation 5.2.1**, now using the energy of a Poly-G_2 dimer. The EPS-Poly-G_2 adducts, displayed at 2 ns intervals, are given in **Fig 5.5** along with their formation energies.

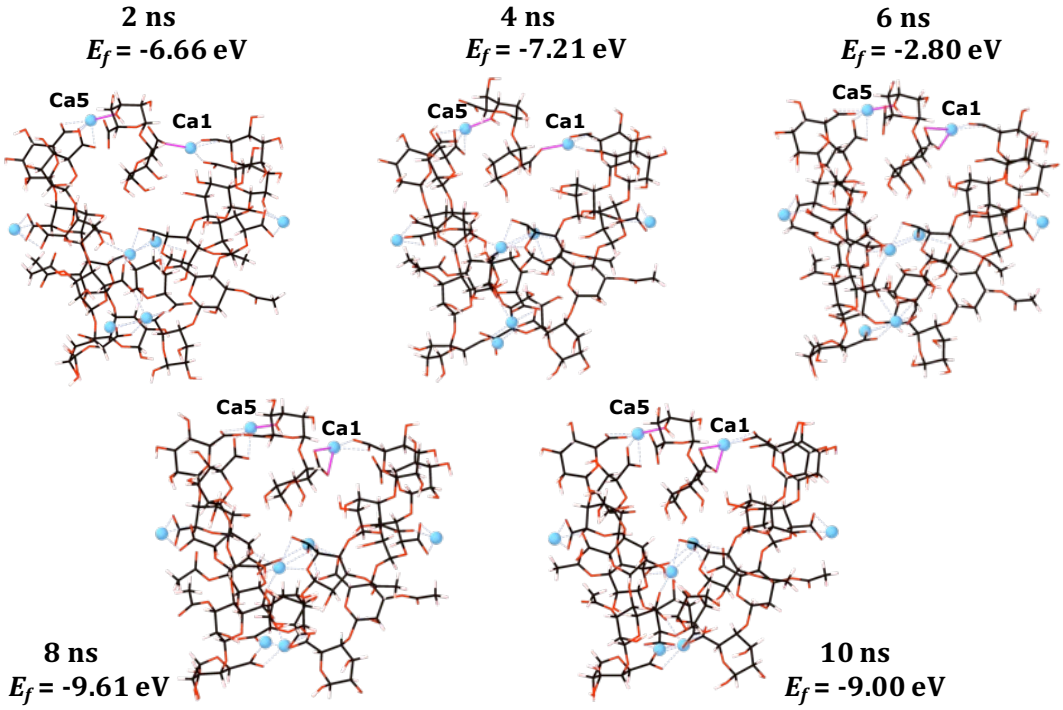


Figure 5.5: Bound EPS-Poly-G_2 adducts, along with their formation energies, displayed every 2 ns. Carbon atoms are shown in black, oxygen in red, calcium as blue spheres and hydrogen in pink. Calcium-oxygen ionic bonds are shown with dashed blue lines and calcium ions implicated in binding of Poly-G_2 are labelled. Ionic bonds between the Poly-G_2 and EPS system are shown with bold pink lines.

In the same fashion as the Poly-G_4 molecule, the Poly-G_2 dimer binds to the EPS in a thermodynamically favourable fashion within the V-shaped cleft, through hydroxyl and carboxylate ionic tethering contacts to Ca^{2+} ions. Similarly, no hydrogen bonding interactions between the Poly-G_2 dimer and the EPS occur and all interactions are facilitated through the EPS Ca^{2+} ions only. Encouragingly, the smaller size of the Poly-G_2 does not lead to any increased mobility compared to the Poly-G_4 system and the dimer remains bound to the EPS within the cleft for the full 10 ns. Therefore, it can be interpreted that the ionic tethering between the EPS and the Poly-G_4 system is primarily responsible for the immobility of the Poly-G_4 molecule in the EPS vicinity, rather than any spurious size-related effects. The thermodynamic stabilities, as defined by the formation energies, are approximately half those of the EPS-Poly-G_4 systems, which is intriguing, as it would suggest that matching the polyguluronate and EPS length scales results in more exothermic penetration. However, an inconsistency in the stability trend occurs in EPS-Poly-G_2 adduct at 6 ns. There is a gain in instability (+ 4.41 eV) when propagating from the conformation at 4 ns to the conformation at 6 ns, which arises because Ca1 is undergoing a transition from the monodentate EPS carboxylate O24 site to settle in the more stable bidentate Poly-G2 carboxylate O12 + O13

binding site (**Fig 5.6**).

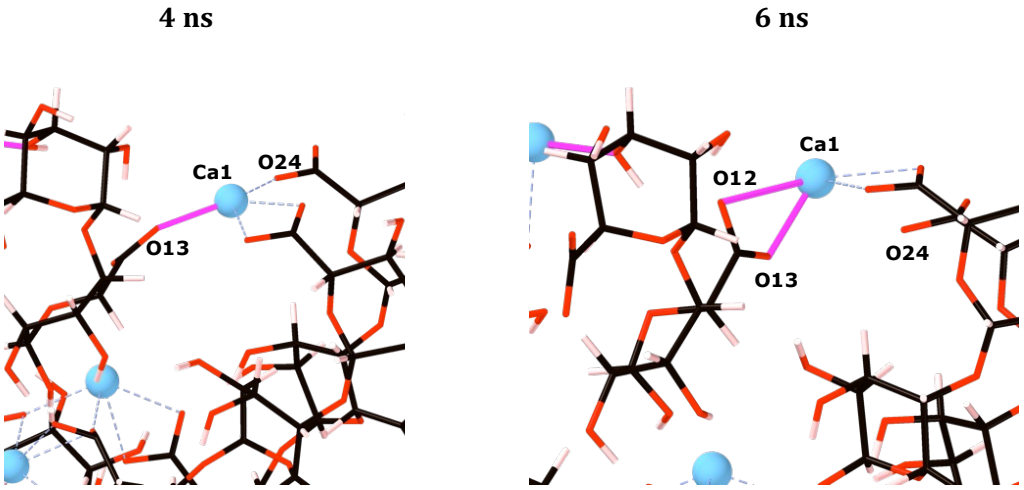


Figure 5.6: A close up of the bound EPS-Poly-G_2 adducts at 4 and 6 ns, highlighting the transition of Ca1 between the monodentate carboxylate O24 site to the more stable bidentate carboxylate O12 + O13 binding site. Carbon atoms are shown in black, oxygen in red, calcium as blue spheres and hydrogen in pink. Calcium-oxygen ionic bonds are shown with dashed blue lines and ionic bonds between the Poly-G_2 and EPS system are shown with bold pink lines.

The newly established bidentate carboxylate binding mode to Ca1 gives rise to more stable EPS-Poly-G_2 adducts by $\sim$ 3 eV, relative to complexes reliant upon the initial monodentate mode (**Fig 5.5**). Furthermore, this new carboxylate binding configuration is stable over the remainder of the 10 ns trajectory. This suggests that Ca^{2+} ions have the ability to transition between carboxylate binding sites, upon Poly-G_2 invasion, although, it also shows that the Ca^{2+} ion (Ca1) is still retained in a chelate complex with the EPS and is not liberated from the EPS chains.

Ca5 and Ca1 are very clearly implicated in facilitating the interaction(s) of both the Poly-G_4 and Poly-G_2 molecules with the EPS scaffold. To test whether these gate-keeper calcium ions are critical in the Poly-G_2 binding or whether the Poly-G_2 would still be bound without them, some additional simulations were undertaken. This was done to ensure that this favourable tethering was not an artifact of our particular model architecture, which represents only one small section of the EPS. Ca5 and Ca1 were, therefore, removed from the EPS-Poly-G_2 starting structure, and the MD simulation repeated for another 10 ns, to gauge the response of the Poly-G_2 dimer within the V-shaped cleft. The EPS-Poly-G_2 adduct occurring at 10 ns in this trajectory can be seen in **Fig 5.7**.

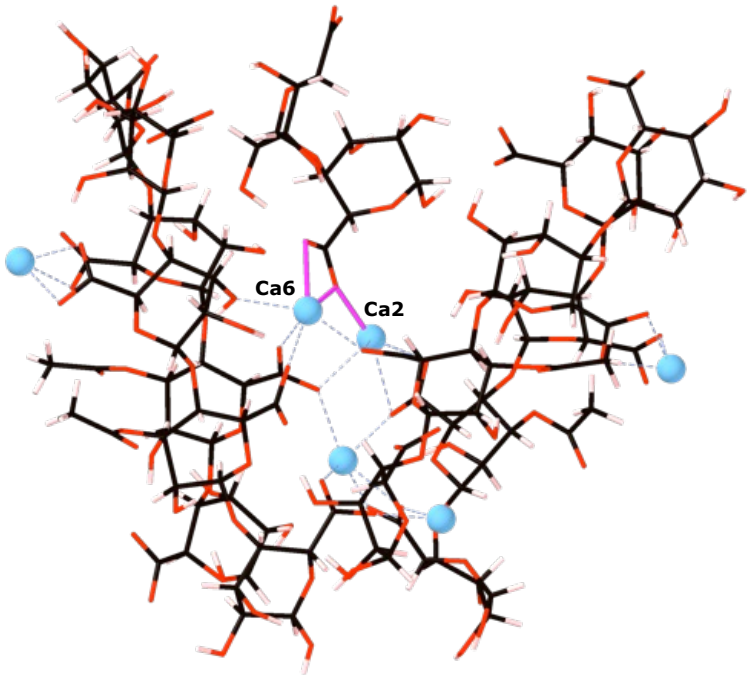


Figure 5.7: The EPS-Poly-G_2 adduct 10 ns following removal of Ca1 and Ca5 from the starting structure. Carbon atoms are shown in black, oxygen in red, calcium as blue spheres and hydrogen in pink. Calcium-oxygen ionic bonds are shown with dashed blue lines and ionic bonds between the Poly-G_2 and EPS system are shown with bold pink lines.

Even when removing the structural bias from the EPS, the Poly-G_2 molecule responded to the absence of Ca5 and Ca1 by sinking deeper into the cleft and reorienting to a position where a single carboxylate group faced Ca6 and Ca2. In turn, the Poly-G_2 dimer establishes three ionic tethering contacts to these ions, in preference to liberating the cleft. It is clear that the cationic charge density still present in the EPS, despite the removal of Ca5 and Ca1, is sufficient to retain the Poly-G_2 dimer in the EPS vicinity and subsequently to bind electrostatically. This, therefore, further highlights the authority Ca^{2+} ions have in mediating interactions between polyguluronate structures and the EPS chains. Additionally, this exemplifies the large (thermodynamic) affinity that polyguluronate structures have towards the EPS scaffold.

5.5 OligoG CF-5/20 molecular mechanism of action

Collectively, the EPS-Poly-G_4 and Poly-G_2 simulations highlight that the Ca^{2+} ions are not liberated from the EPS chains by the invading OligoG CF-5/20 and native EPS bound Ca^{2+} ions retain their chelate geometries bound to EPS oxygen functionality upon OligoG CF-5/20 invasion. These ions are not pulled away from the EPS chains and the invading OligoG CF-5/20 causes no disruption to the EPS ionic scaffold. As discussed in **Section 4.4**, the carboxylate group was most implicated in calcium induced EPS chain association under physiological

conditions and, as shown in **Fig 5.8**, the EPS COO^{-}-Ca^{2+} geometrical parameters remain stable throughout the trajectory.

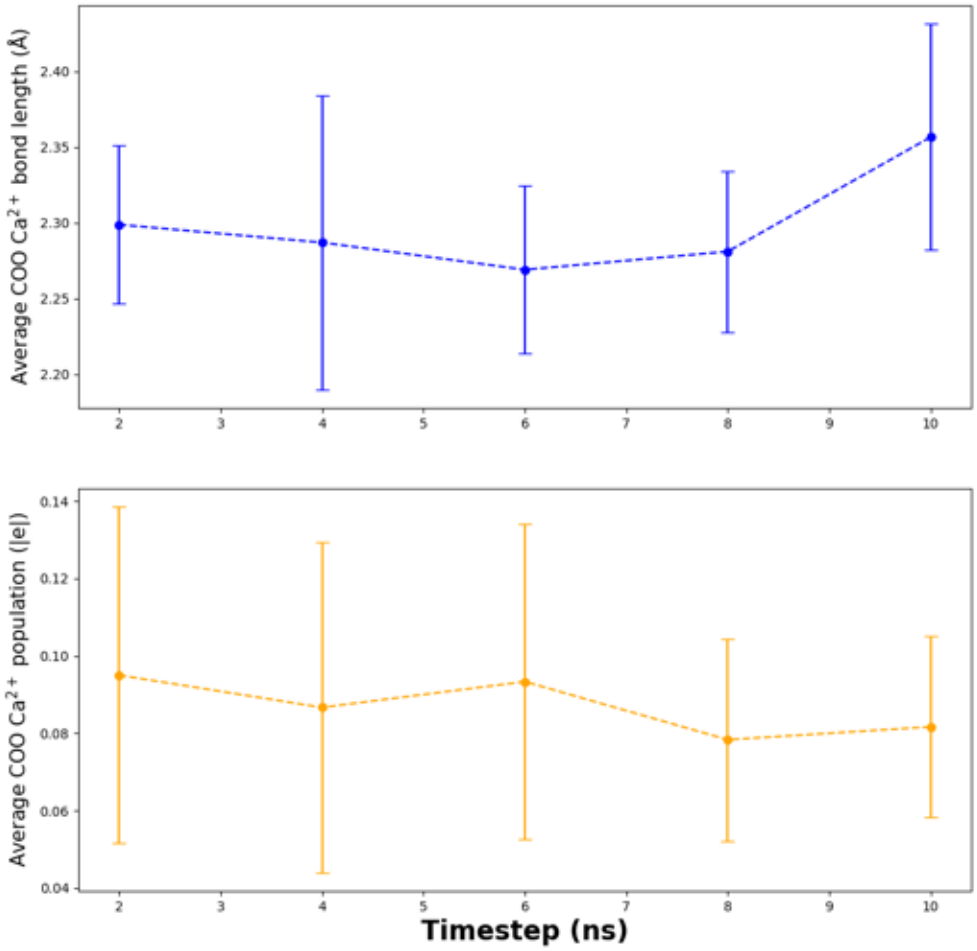


Figure 5.8: Average EPS COO-Ca^{2+} bond lengths (Å) and populations (|e|) ± one standard deviation for the EPS-Poly-G_4 complexes.

The existing EPS COO^{-}-Ca^{2+} interactions are unaffected by the invading Poly-G_4 molecule with their bond lengths and Mulliken populations fluctuating within one standard deviation over the entire course of the trajectory. Furthermore, over the full 10 ns of both the Poly-G_4 and Poly-G_2 trajectories, the orientation of the EPS carboxylate groups remains unchanged, continuously facing the Ca^{2+} ions. This matches exactly with circular dichroism measurements on systems of OligoG CF-5/20 dispersed in *P. aeruginosa* biofilms, that also observe no change in bacterial alginate carboxylate orientation [339]. As the invading polyguluronate molecule does not liberate Ca^{2+} ions from the EPS chains, the overall V-shaped morphology of the model is maintained throughout the full time-scale of the trajectory and there is no separation of the 4-chain EPS structure into two sets of 2-chains or any other smaller units. The EPS V-angle (**Table 5.3**) decreases from 69° to 60° over the 10 ns, showing that, in fact, the EPS chains become closer together as opposed to further apart

upon Poly-G_4 binding, increasing the extent of the encapsulation of the invading Poly-G_4.

In light of all the observations above, the proposed mechanism of disruption, when CF mucoid *P. aeruginosa* biofilm matrices are exposed to OligoG CF-5/20, involves a non-disruptive invasion of the Ca^{2+} cross-links. This stimulates OligoG CF-5/20 capture which, in turn, prevents both the sequestration of free (unbound) Ca^{2+} ions by the EPS, as well as further Ca^{2+} mediated EPS aggregation. This mode of action draws significant comparisons with the proposed mode of action on eDNA exposed to OligoG CF-5/20, which was elucidated through previous MD simulations [183].

5.6 Summary

This chapter investigated the mechanism of action, at the atomic-scale, of OligoG CF-5/20 against mature mucoid *P. aeruginosa* EPS scaffolds. OligoG CF-5/20 represents an exciting, novel anti-*Pseudomonas* therapeutic targeted at treating chronic mucoid *P. aeruginosa* biofilm infections in the CF lung. However, at the atomic scale, the global matrix disruption effects of the OligoG CF-5/20 are unknown, even though the molecule has been shown to have clinical efficacy *in vivo*. To gain understanding into this molecule's mode of action against the EPS, Poly-G_4 and Poly-G_2 - quadramer and dimer molecular models of the OligoG CF-5/20 system - were combined with the physiological EPS structure ($4PolyMG_{MD}$) and propagated through 10 ns of dynamical trajectories to sample modes of interaction.

These simulations have identified that carboxylate and hydroxyl functional groups are responsible for establishing thermodynamically favourable EPS-Poly-$G_4$and Poly-G_2 adducts, although the former is most likely responsible for irreversible binding and, by extension, for the therapeutic effects that are observed over long time-scales. Furthermore, these simulations have highlighted that Ca^{2+} ions are critically important in mediating the interactions between OligoG CF-5/20 and the EPS chains with only electrostatic interactions driving the binding events. Although Ca^{2+} behaves as an OligoG CF-5/20 binding facilitator, these ions are not liberated from the EPS chains and the effectively benign effects on the EPS ionic architecture exerted by the invading OligoG CF-5/20 allows the EPS to retain its V-shaped conformation, with no EPS chain separation being observed.

Simulations of the Poly-G_2 dimer led to thermodynamically favourable EPS adducts, showing large similarities in binding modality to the Poly-G_4 system. The removal of key Ca^{2+} ions, which acted as tethering points for the invading guluronate, did not prevent Poly-G_2 binding, as the dimer was able to sink deeper into the discontinuous cleft in the EPS architecture and reorient a carboxylate group to sustain new ionic O-Ca^{2+} interactions. The greater thermodynamic stability of the EPS-Poly-G_4 adducts, relative to the Poly-G_2 adducts,

would suggest that OligoG CF-5/20 molecules on a similar length scale to the EPS chains are capable of more exothermic EPS dispersal. The similarities between the Poly-G_4 and Poly-G_2 simulations suggest that it is Ca^{2+} ion affinity that keeps the guluronate molecule tethered to the EPS, rather than its size, and further highlights that the EPS bound Ca^{2+} ions give impetus to OligoG CF-5/20 dispersal. Overall, the matrix disruption mechanism posed by OligoG CF-5/20 involves capture by the EPS, which prevents the sequestration of free (unbound) Ca^{2+} ions by the EPS, as well as preventing further Ca^{2+} mediated EPS aggregation.

Exogenous gallium accommodation by the EPS

6.1 Scope of the chapter

As discussed in **Section 1.5.2**, intravenous (exogenous) gallium nitrate therapy is emerging as a novel therapeutic strategy to combat mucoid *P. aeruginosa* CF biofilm infections. The therapeutic activity of gallium is attributed to the high degree of structural similarity between Ga^{3+} and Fe^{3+} ions, allowing the former to bind siderophores, to gain access to the bacterial cell and interfere with physiological processes dependent upon Fe^{3+}. With regards to *P. aeruginosa* bacterial cells, this culminates in either completely blocking or reducing biofilm proliferation.

However, the anti-biofilm effects of gallium exposure are retained in *P. aeruginosa* mutants deficient in siderophores and, in fact, siderophore-mediated gallium uptake is a slow and inefficient process. This raises the question of whether gallium is having a direct effect on the EPS architecture, for example, displacing the calcium cross-links, causing matrix weakening and, subsequently, reduced matrix proliferation. The aim of this chapter is to investigate whether Ga^{3+} ions could be directly accommodated by the mature mucoid *P. aeruginosa* EPS scaffold by means of assessing the thermodynamic feasibility of the calcium-for-gallium exchange.

6.2 Computational Details

6.2.1 DFT simulation parameters

All Density-Functional Theory (DFT) calculations were performed using the plane-wave pseudopotential code, CASTEP [280], using a convergence tested cut-off energy of 900 eV, and a Monkhorst-Pack **k**-point grid of 1 x 1 x 1 to sample the Brillouin zone [255]. On-the-fly ultrasoft pseudopotentials [281] and the PBE-TS exchange-correlation functional were

employed [240, 261]. The SCF tolerance for the electronic minimisations was set to 1×10^{-7} eV Atom^{-1} and the energy, force and displacement tolerances for the geometry optimisations were set to 1×10^{-5} eV Atom^{-1}, 0.03 eV Å^{-1} and 1×10^{-3} Å respectively. Following each geometry optimization, Mulliken bond populations [270] were calculated to classify the nature of bonding in each of the complexed structures. All atomistic models were created and visualised using CrystalMaker [341]. Chemical potentials for sodium, calcium and gallium were calculated by their respective zero Kelvin (0 K) energy per atom from the pure metals in their low energy configurations, namely, hexagonal (HCP) sodium, cubic (BCC) calcium and orthorhombic (alpha) gallium.

6.2.2 Identifying EPS functionality with high gallium affinity

In **Section 3.4.3**, thermodynamically favourable cation cross-linked PolyMG$_p$ and PolyM$_{ap}$ EPS scaffolds were constructed, using a ground-up approach, starting with 1- and 2-chain EPS systems in the absence of cations and identifying stable ionic cross-linking geometries between PolyMG$_p$ and PolyM$_{ap}$ structures. Calcium ions established the most thermodynamically favourable cross-linked aggregates out of all sputum ions tested (**Table 3.6**), offering a thermodynamic explanation for why calcium ions will out-compete other sputum ions for 2-chain cross-linking sites on the EPS scaffold. In addition, the particular calcium cross-linking sites are not susceptible to physiological softening, specifically, the substitution of calcium out the EPS scaffold in preference to sodium ions (**Table 3.13**). It is assumed therefore, for exogenous gallium (Ga^{3+}) to be accommodated by the mature EPS, these ions will be competing for EPS binding against pre-bound, *native*, calcium ions complexed within the Ca-PolyM$_{ap}$ and Ca-PolyMG$_p$ scaffolds.

The calcium cross-linking sites within each of the Ca-PolyM$_{ap}$ and Ca-PolyMG$_p$ EPS scaffolds are unique, each displaying different oxygen functional groups and CNs. Therefore, to better target the calcium-for-gallium substitutions, initial DFT calculations were performed on separate chains - the charge saturated calcium complexed acetylated poly-β-D-mannuronate chain and the charge saturated calcium complexed acetylated copolymeric β-D-mannuronate-α-L-guluronate chain. From here on these systems are labelled as Ca-PolyM and Ca-PolyMG respectively. They were constructed according to **Section 3.4.1** (E_f given in **Table 3.3**), corresponding to anionic (-4 charged) EPS quadramers, saturated by two Ca^{2+} ions (**Fig 6.1**). As discussed in **Section 3.4.1**, the calcium ion positions in **Fig 6.1** correspond to the most thermodynamically favourable 1-chain chelation sites and offer suitable substitution sites (starting positions) for gallium exchange onto a single EPS chain.

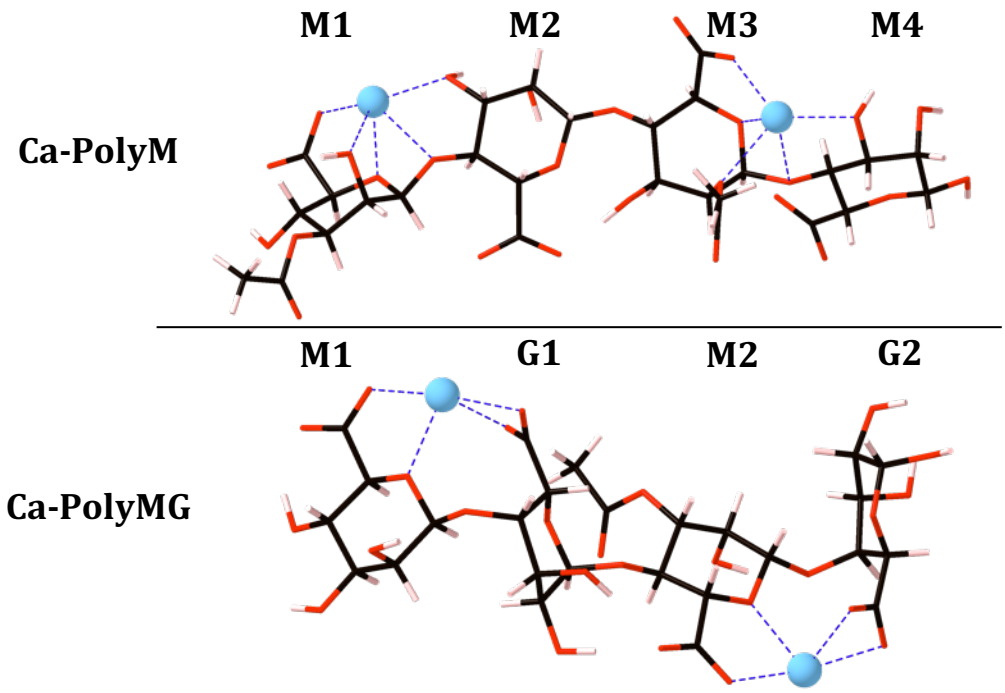


Figure 6.1: Mucoid EPS molecular models of acetylated poly-β-D-mannuronate (Ca-PolyM; top) and acetylated copolymeric β-D-mannuronate-α-L-guluronate (Ca-PolyMG; bottom). Carbon atoms are shown in black, oxygen in red, hydrogen in pink. The native calcium ions are shown as blue spheres and bonds to the calcium ions are also shown in blue. Uronate nomenclature is also given.

For the gallium substitutions into the Ca-PolyM/MG systems, the following calcium-for-gallium charge-balanced substitution was considered:

$$\textbf{Scheme 1} : 2 \times \mathrm{Ca}^{2+}(+4) \Longrightarrow 1 \times \mathrm{Ga}^{3+} + 1 \times \mathrm{Na}^{+}(+4)\ . \tag{6.2.1}$$

6.2.3 Calcium-for-gallium cation exchange 2-chain models

After identifying EPS sites that display the highest gallium affinities at the 1-chain level, targeted calcium-for-gallium cation exchanges were performed into the Ca-PolyM$_{ap}$ and Ca-PolyMG$_p$ EPS scaffolds, substituting the calcium ions preferentially occupying any of the predicted binding site positions with affinity for gallium.

The initial Ca-PolyM$_{ap}$ and Ca-PolyMG$_p$ EPS scaffolds were charge neutral models and, therefore, calcium-for-gallium substitutions were charge-balanced to ensure the final complexed states were also charge neutral. In the 2-chain model systems, two charged-balanced substitution mechanisms were considered:

$$\textbf{Scheme 2} : 3 \times Ca^{2+}(+6) \Longrightarrow 2 \times Ga^{3+}(+6) \; ; \tag{6.2.2}$$

$$\textbf{Scheme 3} : 4 \times Ca^{2+}(+8) \Longrightarrow 2 \times Ga^{3+} + 2 \times Na^{+}(+8) \; . \tag{6.2.3}$$

The calcium ions removed during the targeted calcium-for-gallium exchange, modelled using **Schemes 2** and **3**, are labelled in **Fig 6.4**. Note here, that **Scheme 2** retains a single native calcium ion within the EPS scaffold, whereas **Scheme 3** represents full cation exchange. The thermodynamic stability of the exchange reactions was evaluated by means of calculating a formation energy according to **equation 6.2.4**

$$E_f = E_{gallium-EPS} - (E_{EPS} + l\mu_{Na} + m\mu_{Ga} - n\mu_{Ca}). \tag{6.2.4}$$

Within **Equation 6.2.4**, $E_{gallium-EPS}$ represents the final energy of the gallium-EPS complex (for either the Ga-PolyM$_{ap}$ or Ga-PolyMG$_p$ system), E_{EPS} represents the energy of the initial EPS system (for either the Ca-PolyM$_{ap}$ or Ca-PolyMG$_p$ system), μ_{Na} represents the chemical potential of sodium, μ_{Ga} represents the chemical potential of gallium and μ_{Ca} represents the chemical potential of calcium.

It is important to note here that **Equation 6.2.4** is considering the exchange event form a thermodynamic stand-point, it does not provide any information on the kinetics for the exchange directly. However, the relative kinetics of the exchange can be determined from **Equation 6.2.4** *indirectly* using the Eyring relationship [342, 343]. This is discussed in more detail in **Section 7.4**, within the application of designing a novel gallium delivery vessel.

6.3 Accommodation of gallium by the mucoid EPS

6.3.1 1-chain gallium complex stability

The gallium exchange was performed at both unique calcium sites in both the Ca-PolyMG and Ca-PolyM scaffolds (**Fig 6.1**), yielding two independent gallium-sodium co-substitutions for each 1-chain EPS scaffold. The resulting structures are given in **Fig 6.2** and **6.3**, with their formation energies calculated according to **Equation 6.2.4**.

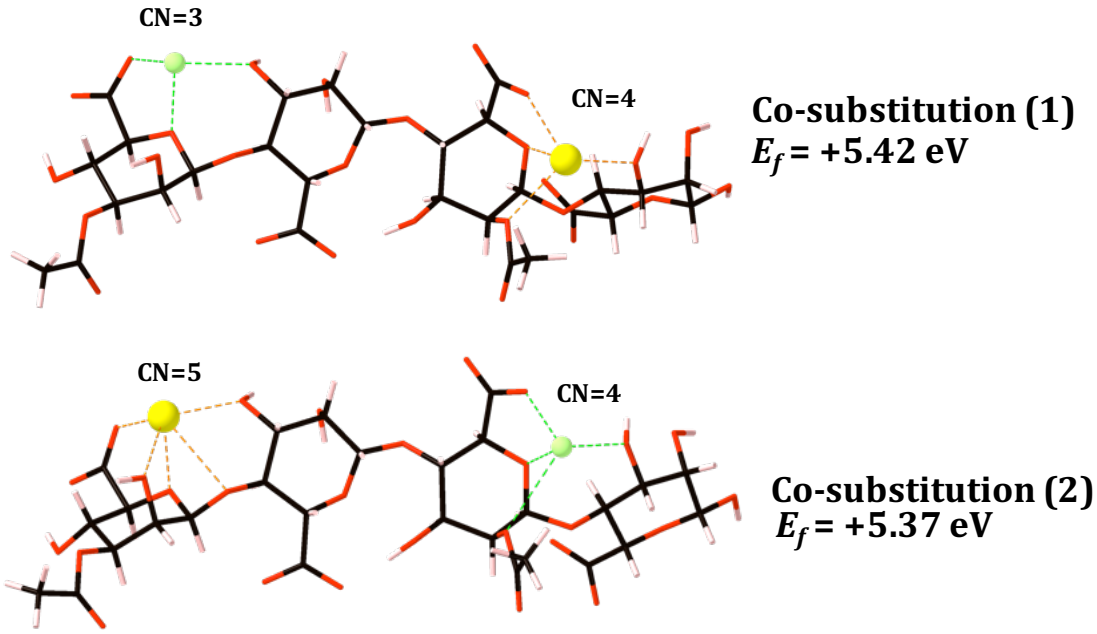


Figure 6.2: Optimised Ga-PolyM EPS complexes, along-side their formation energies (eV), following substitution into the Ca-PolyM EPS scaffold. Carbon atoms are shown in black, oxygen in red, hydrogen in pink, gallium as green spheres and sodium as yellow spheres. Bonds to the gallium and sodium ions are shown as green and orange dashed lines respectively. The cation CNs are also displayed.

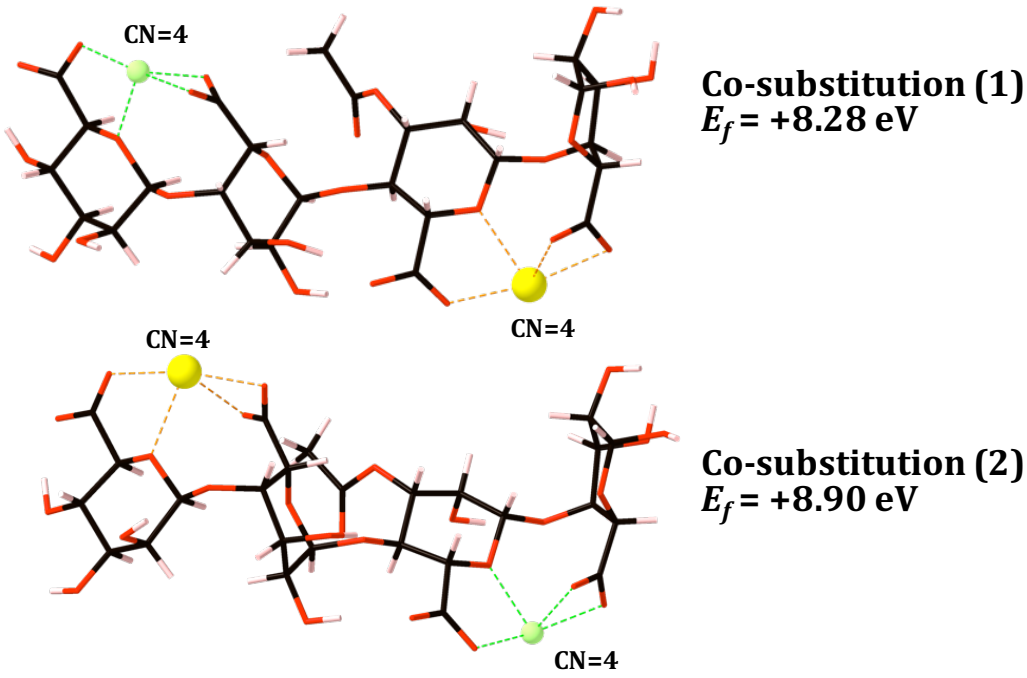


Figure 6.3: Optimised Ga-PolyMG EPS complexes, along-side their formation energies (eV), following substitution into the Ca-PolyMG EPS scaffold. Carbon atoms are shown in black, oxygen in red, hydrogen in pink, gallium as green spheres and sodium as yellow spheres. Bonds to the gallium and sodium ions are shown as green and orange dashed lines respectively. The cation CNs are also displayed.

Firstly, it is evident that calcium-for-gallium substitutions into both 1-chain EPS scaffolds are thermodynamically unfavourable (**Equation 6.2.4** returning a positive formation energy), indicating that gallium cannot substitute the native calcium ions from their chelate complexes within 1-chain EPS scaffolds. The structural chemistry underpinning this instability is discussed in more detail throughout the remainder of the chapter, but this does not neglect the principle aim of these initial calculations - namely, to identify EPS functionality that displays the largest gallium affinity. Even though the resulting 1-chain complexes are thermodynamically unfavourable, it is still possible to interrogate the molecular structure and identify the oxygen functional groups that most stabilise the gallium chelate geometry.

The geometrical parameters for these complexes are given in **Table 6.1** and offer two critical geometrical insights. Firstly, gallium ion binding is possible with a lowered CN when complexed into the mannuronate rich scaffold, Ca-PolyM (**Fig 6.1**). Compared to the native calcium, the gallium CN drops by two for co-substitution 1, losing interactions with hydroxyl and glycosidic oxygen atoms, and by one for co-substitution 2, losing interaction with the glycosidic oxygen atom (**Fig 6.1**). Compared to calcium, which has an ionic radius measuring 0.99 Å, gallium has a smaller ionic radius measuring 0.62 Å [217, 303]. Consequently, it is not

surprising there is a drop in CN. This is analogous structural coordination chemistry to that which was used to explain the lower CN of cross-linking Mg^{2+} ions compared to cross-linking Ca^{2+} ions observed in the PolyMG$_p$ and PolyM$_{ap}$ 2-chain scaffolds (**Figs 3.12** and **3.13**) which, in turn, explained the lower thermodynamic stability of the magnesium complexes relative to the calcium complexes.

Intriguingly, gallium does not experience this same reduction in CN when complexed into the Ca-PolyMG scaffold (**Fig 6.3**). The primary structural feature distinguishing gallium (and calcium) complexation within single M-M and M-G junctions on a single EPS chain is the occurrence of two COO^- groups per cation in the latter compared to only one COO^- group per cation in the former. It is apparent that the gallium ion can complex into a low energy chelate geometry with a lower CN when more COO^- groups are available in the complexation event. This leads to the second geometrical insight, that the average COO^--Ga bond lengths are shorter, and Mulliken populations larger, than all other O-Ga contacts present within both the 1-PolyMG/M gallium complexes (**Table 6.1**). Explicitly, the COO^--Ga lengths are 0.16 - 1 Å shorter compared to all other O-Ga contacts and Mulliken populations indicate, on average, 1.76 times the charge transfer from COO^- donors relative to hydroxyl, ring and acetyl donors. This is in agreement with previous DFT calculations, predicting the complex geometries of trivalent metal cations into non-acetylated algal alginate disaccharides, that resulted in twice the charge transfer from COO^- groups relative to OH groups [142]. It is also possible to discriminate single Ga^{3+} ion complexation at single M-M junctions from M-G junctions, as the monodentate COO^--Ga contact is more stable in the latter and, additionally, the bidentate COO^--Ga mode occurs solely in M-G junctions.

Note, the Mulliken populations for the bidentate COO^--Ga coordination modes, observed in the Ga-PolyMG gallium complexes, do not exceed the ring O-Ga populations. However, the lower populations for the bidentate coordination mode arise from the distribution of the charge transfer from the COO^- group to the Ga^{3+} ion, which, in this circumstance, is distributed over two separate bonds. As such, a better representation of the COO^- to Ga^{3+} charge transfer for bidentate modes can be gained by doubling these respective populations. Upon doing so, the original functional group stability trend, namely COO^--Ga $>$ O-Ga, is retained.

Table 6.1: Average O-Ga bond lengths and Mulliken populations for the gallium 1-chain EPS complexes.

Ga-PolyM co-substitution (1)		
Bond	Average length (Å)	Average population (|e|)
COO^--Ga^{3+} (monodentate)	2.01	0.19
OH-Ga^{3+}	2.43	0.07
Ring O-Ga^{3+}	2.20	0.13
Ga-PolyM co-substitution (2)		
COO^--Ga^{3+} (monodentate)	1.97	0.24
OH-Ga^{3+}	2.36	0.07
Ring O-Ga^{3+}	2.19	0.14
Acetyl O-Ga^{3+}	2.95	0.04
Ga-PolyMG co-substitution (1)		
COO^--Ga^{3+} (monodentate)	1.84	0.36
COO^--Ga^{3+} (bidentate)	1.98	0.19
Ring O-Ga^{3+}	2.00	0.27
Ga-PolyMG co-substitution (2)		
COO^--Ga^{3+} (monodentate)	1.90	0.25
COO^--Ga^{3+} (bidentate)	2.14	0.11
Ring O-Ga^{3+}	2.20	0.17

Collectively this suggests the geometrical principle that COO^- groups are most implicated in gallium complexation into the EPS scaffold, offering the most stable interactions with the alien gallium. This is similar to observations made during divalent sputum ion complexation by the $PolyM_{ap}$ and $PolyMG_p$ scaffolds. In the case of the divalent ions Ca^{2+} and Mg^{2+}, the COO^- group also offers the most stable mode of interaction with the divalent ion, giving COO^--M^{2+} Mulliken populations that are larger than all other oxygen donors and indicating the largest degree of charge transfer from these groups. However, the Mulliken populations for these contacts indicate that the degree of charge transfer between the COO^- and M^{2+} ion is 1.21 times the degree of charge transfer from each of the other oxygen donors (**Table 3.10** and **3.11**), which is lower than the magnitude of the charge transfer difference observed in the case of gallium (1.76 times). Effectively, the COO^- group is a larger contributor to gallium chelate stability than divalent ion chelate stability and gives rise to a larger discrepancy between the stability of the COO^--Ga^{3+} interactions compared to other oxygen donors.

This observation compliments recent DFT Gibbs free energy calculations, evaluating the spontaneity of substituting Ga^{3+} into Fe^{3+}-metalloproteins, which also showed a strong

stabilisation effect offered by carboxylate groups [223]. Specifically, the COO^- bearing amino acid residues, such as aspartate and glutamate, greatly increase Ga^{3+} selectivity and these particular acidic binding site residues display a larger preference for complexing gallium in comparison to iron [223]. In addition, our finding corroborates thermogravimetric analyses that identified that the gallium metal content of Ga^{3+}-loaded alginate is 1.10 mmol/g (10.6 % w/w) [344]. This result indicates gallium saturation based on the number of carboxylate groups along the length of the alginate scaffold and the interpretation is that the gallium ions are complexed by alginate COO^- groups alone [344]. This is not dissimilar to Fe^{3+} complexation by alginate fibres, where FTIR and X-ray photoelectronic spectroscopy (XPS) showed that the COO^- group is the only functional group responsible for direct binding to Fe^{3+} [345].

Therefore, from the O-Ga interactions present within 1-chain gallium complexes, it is suggested that the sites where gallium could potentially be complexed most favourably within the 2-chain EPS structure is at the calcium sites that encompass the most COO^- donors and have the largest CNs. These particular calcium ions are indicated in **Fig 6.4** and it is these ions that are substituted for gallium in **Schemes 2** and **3**.

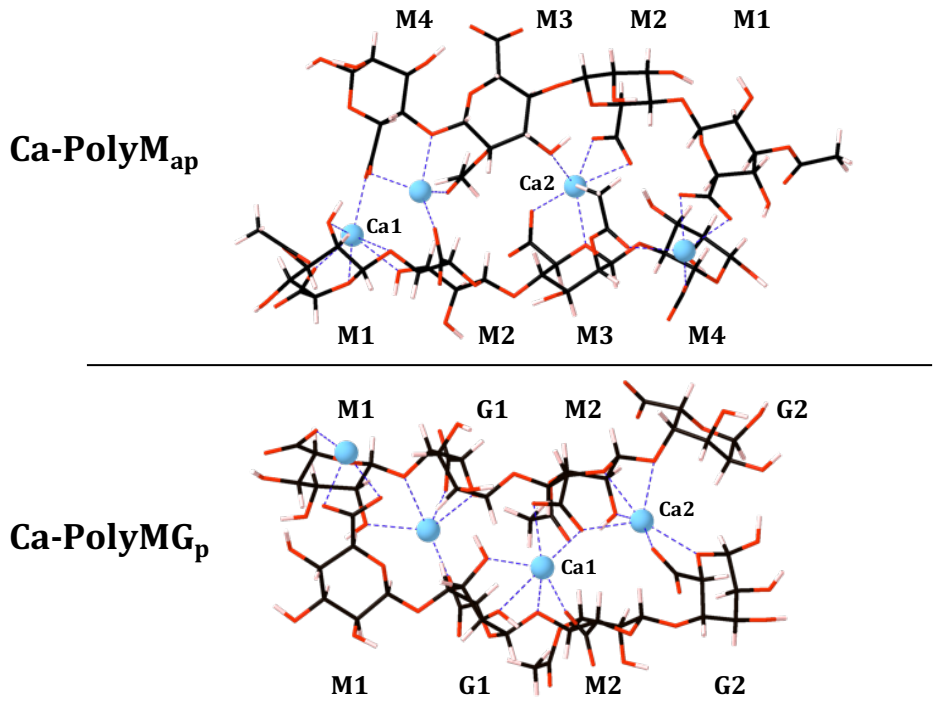


Figure 6.4: PolyM$_{ap}$ (top) and PolyMG$_p$ (bottom) 2-chain calcium complexes constructed in **Section 3.4.3**. Carbon atoms are shown in black, oxygen in red, hydrogen in pink. The native calcium ions are shown as blue spheres and bonds to the calcium ions are also shown in blue. Uronate nomenclature is also given as well as the calcium ions substituted for gallium through **Schemes 2** and **3**.

6.3.2 2-chain gallium complex stability

The optimised gallium 2-chain EPS complexes, in which gallium has been substituted at the calcium positions that occupy the higher gallium affinity sites, along-side their formation energies, for the Ca-PolyM$_{ap}$ and Ca-PolyMG$_{p}$ EPS scaffolds, are given in **Fig 6.5** and **6.6**.

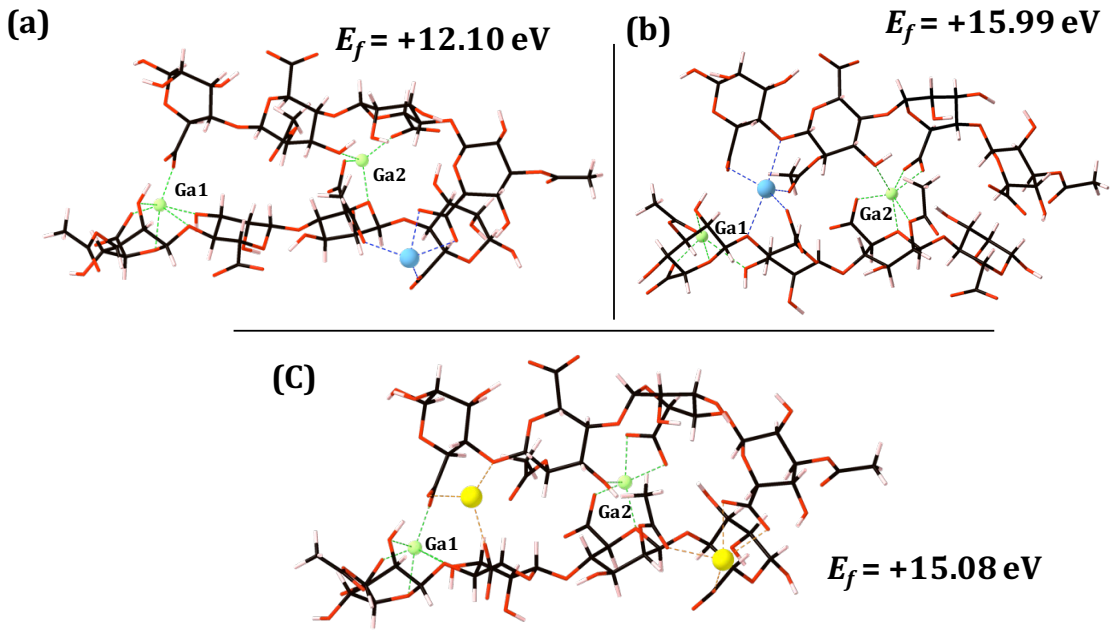


Figure 6.5: Optimised Ga-PolyM$_{ap}$ complexes, along-side their formation energies (eV), following substitution into the Ca-PolyM$_{ap}$ EPS scaffold. Carbon atoms are shown in black, oxygen in red, hydrogen in pink, gallium as green spheres and sodium as yellow spheres. Bonds to the gallium and sodium ions are shown as green and orange dashed lines respectively. The gallium ions are labelled and the native calcium ions are shown as blue spheres with bonds to these calcium ions also shown in blue. (a) and (b) correspond to **Scheme 2** and (c) corresponds to **Scheme 3**.

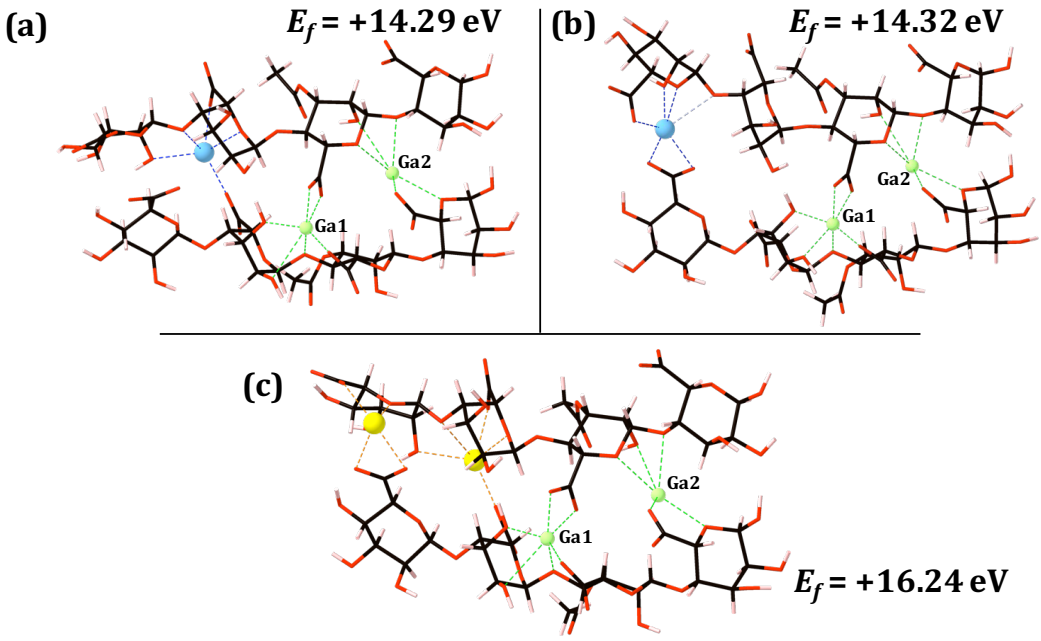


Figure 6.6: Optimised Ga-PolyMG$_p$ complexes, along-side their formation energies (eV), following substitution into the Ca-PolyMG$_p$ EPS scaffold. Carbon atoms are shown in black, oxygen in red, hydrogen in pink, gallium as green spheres and sodium as yellow spheres. Bonds to the gallium and sodium ions are shown as green and orange dashed lines respectively. The gallium ions are labelled and the native calcium ions are shown as blue spheres with bonds to these calcium ions also shown in blue. (a) and (b) correspond to **Scheme 2** and (c) corresponds to **Scheme 3**.

Within the Ga-PolyM$_{ap}$ and Ga-PolyMG$_p$ scaffolds, the charge balanced calcium-for-gallium cation exchange, modelled according both **Schemes 2** and **3**, is thermodynamically unfavourable in all cases. Increasing the number of accessible oxygen atoms, through the coordination of two EPS chains, does not increase the stability of the substitutions, as it does for the divalent cations. Specifically, substitutions into the 2-chain EPS systems are more unfavourable by 7-10 eV for the PolyM$_{ap}$ scaffold and 6-8 eV for the PolyMG$_p$ scaffold relative to their 1-chain analogues (**Fig 6.2** and **6.3**). Therefore, taken along-side the thermodynamic predictions of the 1-chain gallium complexes, it can be inferred that exogenous gallium will not substitute the native calcium ions from their binding sites within M-M and M-G (cross-linking) junctions.

Overall, there is a larger enthalpic barrier to the substitution in the Ga-PolyMG$_p$ system, which displays more positive formation energies relative to the Ga-PolyM$_{ap}$ system. A possible reason for this maybe that, despite retaining charge neutrality over the entire molecular frame, substituting gallium ions into the native calcium sites (according to **Schemes 2** and **3**) creates a charge separation. This involves co-localising two (+3) gallium ions on one half

of the scaffold - leaving two (+1) sodium ions, or one (+2) calcium ion with one vacant (-2) binding pocket on the other side of the scaffold. This is illustrated in **Fig 6.7**.

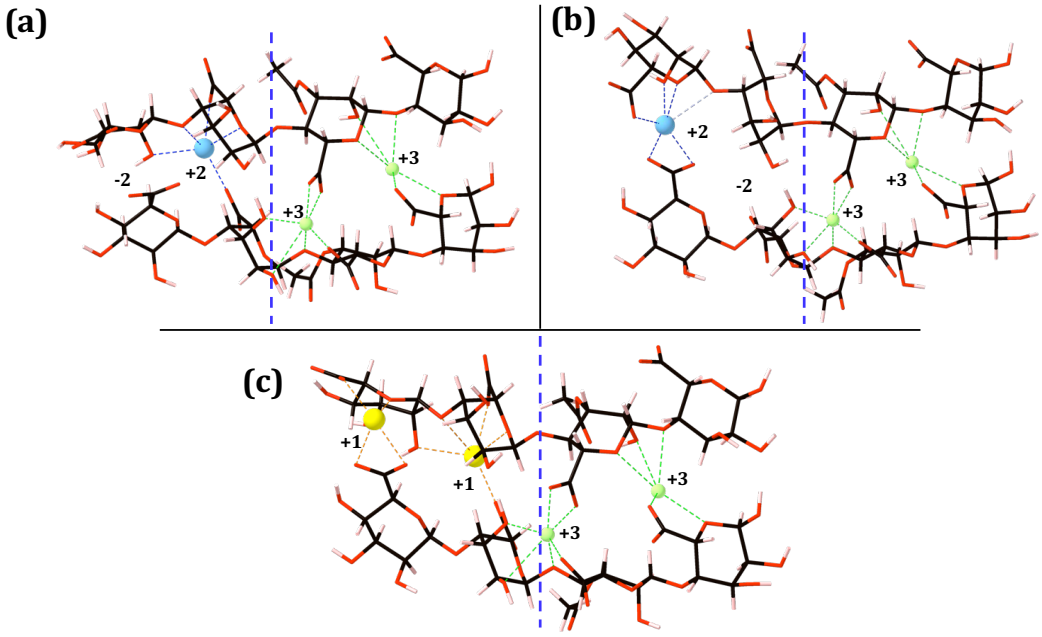


Figure 6.7: Optimised Ga-PolyMG$_p$ complexes, along-side their formation energies (eV), following substitution into the Ca-PolyMG$_p$ EPS scaffold which highlights the heterogeneous cationic charge distribution over the EPS scaffold. Carbon atoms are shown in black, oxygen in red, hydrogen in pink, gallium as green spheres and sodium as yellow spheres. Bonds to the gallium and sodium ions are shown as green and orange dashed lines respectively. The gallium ions are labelled and the native calcium ions are shown as blue spheres with bonds to these calcium ions also shown in blue. (a) and (b) correspond to **Scheme 2** and (c) corresponds to **Scheme 3**.

To investigate whether this is the cause of the increased instability of the Ga-PolyMG$_p$ complexes, an additional substitution scheme was investigated, whereby one Ga^{3+} and one Na^{+} were substituted at the calcium positions which occupy the higher gallium affinity sites as outlined in the previous section (**Fig 6.4**), leaving two Ca^{2+} ions retained within the scaffold. This ensured a +4 charge is partitioned on each side of the Ga-PolyMG$_p$ scaffold. The optimised geometry for this particular complex, along with its formation energy, is given in **Fig 6.8**.

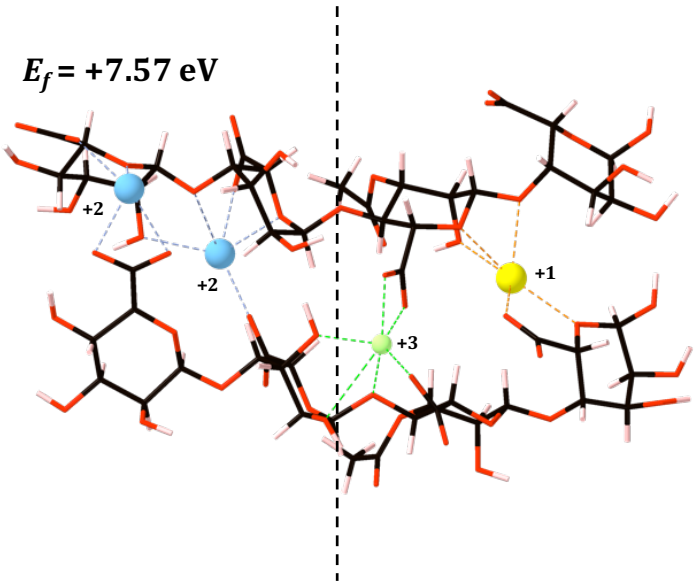


Figure 6.8: Optimised Ga-PolyMG$_p$ complex, along-side its formation energy (eV), following substitution into the Ca-PolyMG$_p$ EPS scaffold. This substitution pattern corresponds to $1 \times Ga^{3+} + 1 \times Na^{+}$ substituted at the high affinity sites leaving $2 \times Ca^{2+}$ retained within the scaffold which ensures +4 charge is partitioned on each half of the Ca-PolyMG$_p$ scaffold. Carbon atoms are shown in black, oxygen in red, hydrogen in pink, gallium as green spheres and sodium as yellow spheres. Bonds to the gallium and sodium ions are shown as green and orange dashed lines respectively. The native calcium ions are shown as blue spheres with bonds to these calcium ions also shown in blue.

This particular gallium substitution is thermodynamically unfavourable (E_f = + 7.57 eV), however, it is 5-8 eV and 7-9 eV more stable than the Ga-PolyM$_{ap}$ and Ga-PolyMG$_p$ complexes respectively (**Fig 6.5** and **6.6**). This suggests that an uneven distribution of positive charge across the EPS scaffold does contribute to the thermodynamic instability of the gallium substitution. It would appear that this particular calcium-for-gallium substitution, which equates the cationic charge distribution on each half of the EPS scaffold, leads to a 50 % increase in stability. Nevertheless, it is not possible to attribute the gain in stability solely to this phenomenon, as this particular complex (**Fig 6.8**) remains thermodynamically unfavourable. This substitution retains two native Ca^{2+} ions within the EPS scaffold, rather than one (**Scheme 2**) or none (**Scheme 3**). Therefore, it must also be taken into consideration that, as fewer calcium ions are removed as part of the exchange, the enthalpic barrier for calcium removal is reduced, making the substitution more favourable. This is discussed in more detail in **Section 6.3.5**.

In addition, the gallium substitution patterns of the Ga-PolyM$_{ap}$ and Ga-PolyMG$_p$ systems (**Fig 6.5** and **6.6**) are accommodating a +3 ion in, effectively, a −2 binding pocket. To assess whether gallium can be more favourably complexed within a vacant −3 binding pocket,

additional DFT calculations were performed, complexing Ga^{3+} into a tri-polyguluronate (non-acetylated algal alginate) quadramer scaffold. The gallium ions were bound exclusively at COO^- sites, as thermogravimetric analysis has identified that algal alginate complexes Ga^{3+} ions at COO^- sites alone [344]. The optimised gallium tri-polyguluronate complex is depicted in **Fig 6.9** along-side the formation energy, evaluated according to **Equation 6.3.1**

$$E_f = E_{Gallium-Polyguluronate} - (3E_{Polyguluronic\,acid} + 2\mu_{Ga} - 6\mu_H)\ . \tag{6.3.1}$$

In **Equation 6.3.1**, $E_{Gallium-Polyguluronate}$ represents the energy of the gallium tri-polyguluronate scaffold, $E_{Polyguluronic\,acid}$ represents the energy of poly-α-L-guluronic acid, μ_{Ga} is the chemical potential of gallium and μ_H is the chemical potential of a hydrogen atom. In this structure (**Fig 6.9**), to maintain charge balance, only six protons were removed from the six COO^- groups oriented into the intra-complex space, to create two -3 charge pockets to accommodate the two Ga^{3+} ions.

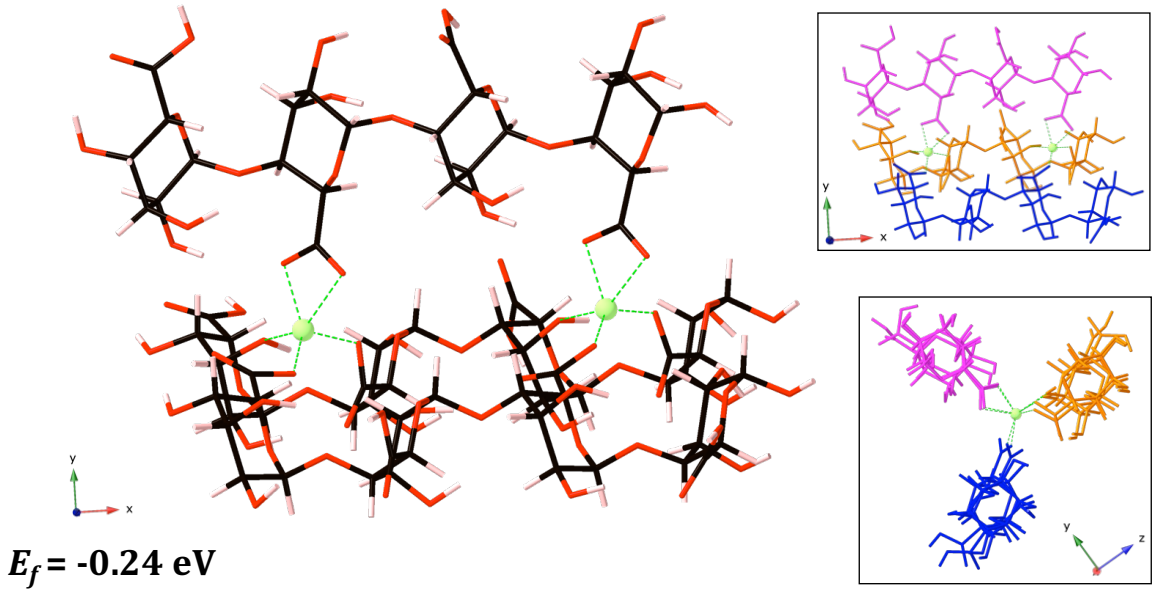


Figure 6.9: Gallium accommodation by a tri-polyguluronate (non-acetylated algal alginate) quadramer scaffold with the associated formation energy (eV). Carbon atoms are shown in black, oxygen in red, hydrogen in pink and gallium as green spheres. Bonds to the gallium ions are shown as green dashed lines respectively. Included within the insets is the same structure, but each polyguluronate chain can be better distinguished based on colour.

This system is minorly thermodynamically favourable ($E_f = -0.24$ eV) indicating that Ga^{3+} ions can favourably complex into vacant -3 binding pockets present within alginate trimers. In fact, this agrees with previous DFT observations that show that it is thermodynamically favourable for Ga^{3+} ions to bind within vacant (-3) pockets present within multi-chain

algal alginate disaccharides and hexasaccharides [142]. Therefore, the -2 charge pocket generated upon liberation of the native calcium from in-between two EPS chains, offers insufficient counter-charge to complex the alien Ga^{3+} ion in a stable fashion, despite retaining charge neutrality over the entire scaffold. The unsuitability of the 2-chain chelation sites for gallium ions is further explored in **Section 6.3.4**, but at this stage, we begin to see that 2-chain cross-linking sites do not provide suitably oriented and/or suitably accessible negatively charged (acidic) functionality to accommodate the Ga^{3+} ion in a thermodynamically favourable chelate pocket in-between two EPS chains. However, it is also important to note finally, that this complexation event (**Fig 6.9**) does not involve the removal of any native, pre-bound, cations. In other words, there is no enthalpic barrier associated with any form of cation exchange, which may also offer another reason for the occurrence of stable complexation.

6.3.3 Gallium-EPS interactions

The geometrical parameters for the 2-chain gallium-EPS complexes are given in **Tables 6.2** and **6.3**.

Table 6.2: Average O-Ga bond lengths and Mulliken populations for the Ga-PolyM_{ap} complexes.

Ga-PolyM_{ap} substitution 1 (**Fig 6.5a**)		
Bond	Average length (Å)	Average population (|e|)
COO^{-}-Ga^{3+} (monodentate)	1.92	0.37
OH-Ga^{3+}	2.01	0.25
Ring O-Ga^{3+}	2.07	0.24
Glycosidic O-Ga^{3+}	2.29	0.16
O-Ga^{3+} (all O donors)	1.98	0.31
O-Ca^{2+} (all O donors)	2.27	0.12
Ga-PolyM_{ap} substitution 2 (**Fig 6.5b**)		
COO^{-}-Ga^{3+} (monodentate)	1.94	0.31
COO^{-}-Ga^{3+} (bidentate)	2.03	0.22
OH-Ga^{3+}	2.57	0.10
Ring O-Ga^{3+}	2.20	0.18
Acetyl O-Ga^{3+}	2.37	0.19
O-Ga^{3+} (all O donors)	2.24	0.19
O-Ca^{2+} (all O donors)	2.37	0.13
Ga-PolyM_{ap} co-substitution (**Fig 6.5c**)		
COO^{-}-Ga^{3+} (monodentate)	1.87	0.40
COO^{-}-Ga^{3+} (bidentate)	1.99	0.23
OH-Ga^{3+}	2.08	0.24
Ring O-Ga^{3+}	2.10	0.22
Glycosidic O-Ga^{3+}	2.17	0.18
O-Ga^{3+} (all O donors)	2.02	0.27
O-Ca^{2+} (all O donors)	-	-

Table 6.3: Average O-Ga bond lengths and Mulliken populations for the Ga-PolyMG$_p$ complexes.

Ga-PolyMG$_p$ substitution 1 (**Fig 6.6a**)		
Bond	Average length (Å)	Average population (|e|)
COO^--Ga^{3+} (monodentate)	1.92	0.34
COO^--Ga^{3+} (bidentate)	2.05	0.20
OH-Ga^{3+}	2.58	0.13
Ring O-Ga^{3+}	2.65	0.06
Glycosidic O-Ga^{3+}	2.45	0.16
O-Ga^{3+} (all O donors)	2.35	0.17
O-Ca^{2+} (all O donors)	2.35	0.13
Ga-PolyMG$_p$ substitution 2 (**Fig 6.6b**)		
COO^--Ga^{3+} (monodentate)	1.92	0.34
COO^--Ga^{3+} (bidentate)	2.05	0.19
OH-Ga^{3+}	2.35	0.16
Ring O-Ga^{3+}	2.60	0.06
Glycosidic O-Ga^{3+}	2.39	0.17
O-Ga^{3+} (all O donors)	2.27	0.18
O-Ca^{2+} (all O donors)	2.41	0.09
Ga-PolyMG$_p$ co-substitution (**Fig 6.6c**)		
COO^--Ga^{3+} (monodentate)	1.91	0.34
COO^--Ga^{3+} (bidentate)	2.06	0.20
OH-Ga^{3+}	2.50	0.14
Ring O-Ga^{3+}	2.59	0.06
Glycosidic O-Ga^{3+}	2.40	0.17
O-Ga^{3+} (all O donors)	2.31	0.18
O-Ca^{2+} (all O donors)	-	-

As with the 1-chain complexes, the COO^- group remains the EPS functional group offering the most stable mode of interaction with the alien gallium ion. Independent of the substitution position for both EPS scaffolds, the COO^--Ga^{3+} interactions are most stable, experiencing average lengths up to 0.5 Å shorter, and average Mulliken populations up to 0.28 |e| larger, relative to the other O-Ga contacts across both 2-chain scaffolds (**Tables 6.2** and **6.3**). With the exception of the carboxylate donors, all O-Ga interactions, across the Ga-PolyM$_{ap}$ and Ga-PolyMG$_p$ complexes, possess Mulliken populations within the range of 0.06 - 0.25 |e|, explicitly indicating the smaller degree of covalent character in these interactions relative to the COO^--Ga interactions. More specifically, the carboxylate donors possess far larger

Mulliken populations, reaching 0.40 |e| and 0.34 |e| in the Ga-PolyM$_{ap}$ and Ga-PolyMG$_{p}$ complexes respectively, that exceed those of the COO^{-}-Ga^{3+} populations observed in the 1-chain complexes (**Table 6.1**). The COO^{-}-Ga^{3+} interactions are 0.15-0.43 Å and 0.43-0.68 Å shorter than the remaining O-Ga interactions in the Ga-PolyM$_{ap}$ and Ga-PolyMG$_{p}$ complexes respectively, not too dissimilar from the shortening observed at the 1-chain scale (**Section 6.3.1**). These shortenings agree well with previous DFT predictions, highlighting that COO^{-}-Ga^{3+} bond lengths are 0.1-0.4 Å shorter compared to OH-Ga bond lengths when gallium is complexed within 1-, 2- and 3-chain non-acetylated algal alginate [142]. The shorter COO^{-}-Ga^{3+} interactions, relative to the remaining O-Ga interactions, can be attributed to the larger degrees of covalency. Specifically, the larger degree of orbital overlap between carboxylate oxygen $2p$ orbitals and the vacant gallium ion $4s$ orbital. This serves to establish more stable interactions.

Also given in **Tables 6.2** and **6.3** are the average O-Ga and O-Ca bond lengths and populations, encompassing all oxygen donors for each respective chelate pocket. Analysis of the average O-ion Mulliken populations gives an understanding of covalent character in the cation chelate complex as a whole. For all complexes, the O-Ga populations are 0.04 |e| - 0.19 |e| larger relative to the O-Ca populations, indicating 1.31 - 2.5 times the charge transfer from the oxygen donors to the alien gallium ions relative to the native calcium ion. Effectively, the gallium chelate complexes, overall, possess greater covalent character.

The larger degree of covalency present within the gallium chelate pockets can be further exemplified by considering the effective ionic valence of Ga^{3+} cations bound within the Ga-PolyM$_{ap}$ and Ga-PolyMG$_{p}$ complexes. The effective ionic valence (described in **Section 2.12.2**) is the difference between the atomic charge predicted through the Mulliken analysis and the formal atomic charge [272]. The closer the value to zero, the smaller the discrepancy and the greater the ionic character exists within the chelate pocket. As such, in these systems, the larger the effective ionic valence, the greater the degree of charge transfer from the EPS scaffold oxygen donors to the complexed cation. The formal charges are +3 and +2 for gallium and calcium ions respectively and the Mulliken atomic charges for Ga1, Ga2 and Ca (**Fig 6.5** and **6.6**), along-side the effective ionic valances, are given in **Table 6.4**. In the Ga-PolyM$_{ap}$ and Ga-PolyMG$_{p}$ complexes the effective ionic valence for the gallium ions exceeds that of the native calcium ion by 2-4 times. This roughly correlates with the degree of charge transfer expected from analysis of the O-Ga and O-Ca Mulliken populations averaged over all oxygen donors (**Tables 6.2** and **6.3**). This is further evidence of the presence of substantially greater charge transfer from the EPS scaffold to the gallium ion, reflective of the far larger degrees of covalency in the gallium chelate pockets relative to the native calcium chelate pockets.

Table 6.4: Mulliken populations and effective ionic valances for the Ga^{3+} and Ca^{2+} ions bound within the Ga-PolyM_{ap} and Ga-PolyMG_p EPS scaffolds.

Ga-PolyM_{ap} substitution 1 (**Fig 6.5a**)		
	Mulliken atomic charge	Effective ionic valence
Ga 1	1.78	1.22
Ga 2	1.87	1.13
Ca	1.45	0.55
Ga-PolyM_{ap} substitution 2 (**Fig 6.5b**)		
	Mulliken atomic charge	Effective ionic valence
Ga 1	0.80	2.20
Ga 2	1.80	1.20
Ca	1.50	0.50
Ga-PolyM_{ap} co-substitution (**Fig 6.5c**)		
	Mulliken atomic charge	Effective ionic valence
Ga 1	1.74	1.26
Ga 2	1.84	1.16
Ga-PolyMG_p substitution 1 (**Fig 6.6a**)		
	Mulliken atomic charge	Effective ionic valence
Ga 1	1.78	1.22
Ga 2	0.84	2.16
Ca	1.45	0.55
Ga-PolyMG_p substitution 2 (**Fig 6.6b**)		
	Mulliken atomic charge	Effective ionic valence
Ga 1	1.77	1.23
Ga 2	0.90	2.10
Ca	1.48	0.52
Ga-PolyMG_p co-substitution (**Fig 6.6c**)		
	Mulliken atomic charge	Effective ionic valence
Ga 1	1.77	1.23
Ga 2	0.89	2.11

On account of this, it is possible to distinguish uronate oxygen bonding to trivalent metal ions from uronate oxygen bonding to divalent metal ions by the larger degrees of covalency in O-ion contacts present within the former. The occurrence of O-Ga interactions with larger degrees of covalency is not specific to polysaccharide systems alone. Gallium has also been observed to bind to regions of the gallium binding peptide C3.15 (NYLPHQSSSPSR) where there exists donor lone pairs, for example, on backbone carbonyl groups, to establish

larger magnitudes of charge transfer [346]. In particular, MD simulations investigating the binding mechanism between Ga^{3+} ions and C3.15 suggest the most energetically favoured gallium-bound structure occurs when gallium binds to the backbone carbonyl residues linking leucine and proline (Leu3-Pro4) residues [346], suggesting that suitably oriented sp^2 oxygen atoms with donor lone pair electrons are critical for gallium chelate stability. Although the authors observed gallium to bind stably at Leu3-Pro4 [346], it is important to note that backbone carbonyl groups are not sequence specific, meaning that other low energy gallium complexed structures may exist at different regions along the length of the peptide chain.

Although the alien gallium ions complex most strongly to COO^- groups in the ECM, they do participate in binding to all types of oxygen donors, a property also observed in gallium-chitosan complexes and in gallium-alginate bio-glasses [347–349]. Contributions from the acetyl groups to the gallium chelate geometry are observed in the Ga-PolyM$_{ap}$ complexes (**Fig 6.5** and **6.6**), but not the Ga-PolyMG$_p$ complexes. This is a distinguishing feature of the calcium cross-linked 2-chain and 4-chain acetylated poly-β-D-mannuronate scaffolds (**Sections 3.4.3.2** and **4.3.1**) and it is maintained here upon calcium-for-gallium exchange. In addition to acetyl contributions, Ga-PolyM$_{ap}$ complexation can be further differentiated from Ga-PolyMG$_p$ complexation by analysis of the hydroxyl, glycosidic and ring O-Ga^{3+} interactions. Within the Ga-PolyM$_{ap}$ complexes, the average OH-Ga^{3+} and ring O-Ga^{3+} contacts are 0.07 - 0.28 Å shorter relative to the glycosidic O-Ga^{3+} interactions. Whereas, this is inverted within the Ga-PolyMG$_p$ complexes with the average glycosidic O-Ga^{3+} contacts now being 0.1 - 0.21 Å shorter relative to the OH-Ga^{3+} and ring O-Ga^{3+} interactions (**Tables 6.2** and **6.3**). Therefore, the glycosidic oxygen atoms play a greater role in gallium complex stability in the Ga-PolyMG$_p$ scaffolds.

For each cation chelate complex, there exists an average O-ion length as well as a number of satisfactory interactions (CN). As such, the volume of a given chelate pocket, which is not spherical, can be approximated by the average O-ion length divided by the CN. Although this measurement is a normalised length scale and not a strict volume, it offers a good approximation to the change in volume upon the exchange.

Table 6.5: The normalised length scale (Å) within the native calcium and alien gallium chelation pockets for Ga-PolyM$_{ap}$ and Ga-PolyMG$_{p}$ complexes. The number of carboxylate groups bound to each gallium ion is reported also. Ca1 and Ca2 are outlined **Fig 6.4**.

Normalised length scale (Å) within the chelate pocket (Å)				
	O-Ca1	O-Ga1	O-Ca2	O-Ga2
Ga-PolyM$_{ap}$ substitution 1	0.41	0.34 2 COO-/Ga	0.47	0.47 2 COO-/Ga
Ga-PolyM$_{ap}$substitution 2		0.61 1 COO-/Ga		0.35 2 COO-/Ga
Ga-PolyM$_{ap}$ co-substitution		0.34 2 COO-/Ga		0.40 2 COO-/Ga
Ga-PolyMG$_{p}$ substitution 1	0.40	0.35 2 COO-/Ga	0.41	0.52 1 COO-/Ga
Ga-PolyMG$_{p}$ substitution 2		0.34 2 COO-/Ga		0.43 1 COO-/Ga
Ga-PolyMG$_{p}$ co-substitution		0.35 2 COO-/Ga		0.43 1 COO-/Ga

Although the O-Ga interactions occur at shorter distances on average (**Tables 6.2** and **6.3**), it is interesting to observe that not all the gallium chelate pockets are contracted relative to the native calcium. In fact, contraction is only observed when two carboxylate groups are bound to the gallium ion. In this circumstance, the normalised length scale of the chelate pocket is contracted by up to 0.12 Å across both the Ga-PolyM$_{ap}$ and Ga-PolyMG$_{p}$ complexes. In contrast, when only a single carboxylate group is bound, the alien gallium chelation pocket is expanded relative to the native calcium chelate pocket and the normalised length scale is lengthened by up to 0.2 Å across both complexes, without an increase in CN. This further accentuates the role that COO^- groups play in tight gallium complexation into the EPS scaffold, underscoring the need to establish two or more COO^- groups per gallium ion to maximise gallium affinity.

Finally, it also interesting to note, that when only a single COO^- group is bound to the gallium ion, the effective ionic valence of the ion is larger compared to when two COO^- groups are bound (**Table 6.4**). Intriguingly, this suggests that, although the COO^- group is very clearly implicated in the largest degree of charge transfer to the alien Ga^{3+} ion (and primarily contributing to the covalent nature of the gallium chelate pocket: **Tables 6.2** and **6.3**), increasing the number of coordinating anionic COO^- groups from one to two increases the *ionic* character of the chelate pocket. This can be further illustrated by considering the Mulliken charges and effective ionic valances for the gallium ions complexed

within the gallium tri-polyguluronate scaffold (**Fig 6.9**). The Mulliken charges for the two Ga^{3+} ions are 1.87 and 1.81, giving effective ionic valances of 1.23 and 1.29. The Ga^{3+} ions here are bound within -3 binding pockets encompassing three COO^- groups per gallium and with effective ionic valances highly similar to that observed when two COO^- groups are bound per gallium ion (1.13-1.23, **Table 6.4**). This reiterates the principle that there is more ionic character in the chelate pocket when more anionic functionality is involved in the coordination, despite the larger charge donating capacity of this functionality.

Analysis of the gallium-EPS interactions alone would suggest that substitution into the EPS scaffold, in principle, should be stable (viable) as the gallium ions complex with far greater covalent character relative to the native calcium ions. However, as outlined in **Sections 6.3.1** and **6.3.2** these substitutions are destabilised by lack of accessibility to negatively charged functionality when complexing in-between 2-chains as well as, potentially, the +3 charge localization issue identified above. These geometrical features, however, are not the sole contributor to the unfavourability of the cation exchange. Further insights can be gained through interrogating the torsion parameters, as well as the enthalpic barrier for calcium removal, which are discussed in the following sections.

6.3.4 Gallium induced conformational change

When complexed in-between two EPS chains, the gallium ion is more susceptible, compared to 1-chain scaffolds (**Fig 6.3**), to a drop in CN (**Fig 6.5** and **6.6**). The average reduction in the gallium CN, relative to the native calcium CN, is 1.33 and 1 for the Ga-PolyM$_{ap}$ and Ga-PolyMG$_p$ complexes respectively. The reduction in gallium CN is less for the PolyMG$_p$ scaffolds as mannuronate-guluronate junctions offer more suitably oriented oxygen atoms to bind the smaller gallium ion. Torsion angles (ϕ, ψ) across the mannuronate(M)-mannuronate(M) and mannuronate(M)-guluronate(G) junctions, present within the Ga-PolyM$_{ap}$ and Ga-PolyMG$_p$ complexes, are given in **Table 6.6** and **6.7** respectively.

Table 6.6: Torsion angles (ϕ, ψ) across the mannuronate(M)-mannuronate(M) junctions in the Ca-PolyM_{ap} EPS systems and Ga-PolyM_{ap} gallium complexes. Uronate nomenclature is given in **Fig 6.4** and definition of (ϕ, ψ) is given in **Fig 3.10**. All angles are given in degrees (°)

System	M1-M2 (ϕ, ψ)°	M2-M3 (ϕ, ψ)°	M3-M4 (ϕ, ψ)°
Ca-PolyM_{ap} top chain	(-51, -99)	(-51, -127)	(-49, +140)
Ca-PolyM_{ap} bottom chain	(-76, -121)	(-94, -104)	(-102, -109)
Ga-PolyM_{ap} substitution 1 top chain	(-53, -105)	(-63, -123)	(-73, -175)
Ga-PolyM_{ap} substitution 1 bottom chain	(-72, -107)	(-85, -154)	(-94, -112)
Ga-PolyM_{ap} substitution 2 top chain	(-53, -93)	(-59, -98)	(-57, +144)
Ga-PolyM_{ap} substitution 2 bottom chain	(-76, -127)	(-84, -102)	(-106, -108)
Ga-PolyM_{ap} co-substitution top chain	(-48, -119)	(-73, -70)	(-23, +136)
Ga-PolyM_{ap} co-substitution top chain	(-80, -120)	(-91, -119)	(-99, -98)

Table 6.7: Torsion angles (ϕ, ψ) across the mannuronate(M)-mannuronate(G) junctions in the Ca-PolyMG_p EPS systems and Ga-PolyMG_p gallium complexes. Uronate nomenclature is given in **Fig 6.4**. All angles are given in degrees (°)

System	M1-G1 (ϕ, ψ)°	G1-M2 (ϕ, ψ)°	M2-G2 (ϕ, ψ)°
Ca-PolyMG_p top chain	(-91, -127)	(-94, -130)	(-110, -78)
Ca-PolyMG_p bottom chain	(-58, -122)	(-89, -130)	(-64, -44)
Ga-PolyMG_p substitution 1 top chain	(-84, -129)	(-86, -120)	(-101, -86)
Ga-PolyMG_p substitution 1 bottom chain	(-55, -130)	(-98, -147)	(-62, -49)
Ga-PolyMG_p substitution 2 top chain	(-105, -103)	(-90, -144)	(-102, -78)
Ga-PolyMG_p substitution 2 bottom chain	(-58, -117)	(-89, -144)	(-56, -47)
Ga-PolyMG_p co-substitution top chain	(-97, -104)	(-86, -128)	(-99, -86)
Ga-PolyMG_p co-substitution top chain	(-56, -120)	(-97, -148)	(-62, -50)

M-G junctions are more resistant to uronate backbone conformational change upon calcium-for-gallium cation exchange, with global (acute) changes in ϕ and ψ for gallium-complexed M-G junctions not exceeding 16° and 24° respectively. Whereas these changes exceed 22° and 57° respectively for gallium-complexed M-M junctions within the Ga-PolyM_{ap} complexes. This also reinforces the observation that the M-G junctions offer suitably oriented oxygen functionality to bind the incoming gallium ion without the need for conformational rearrangement, explaining why the gallium CN for the Ga-PolyMG_p complexes drops less severely compared to the Ga-PolyM_{ap} complexes.

The larger gallium-induced conformational responses in the M-M junctions within the Ga-PolyM_{ap} complexes leads to two destabilising structural alterations. Firstly, the

Ga-PolyM$_{ap}$ substitution 1 complex hosts a large ψ reorientation across the top chain M3-M4 junction equal to $+45°$. This particular torsion change is attributed to the vacant 2-chain cross-linking site exposed through the removal of a calcium ion. This leaves the glycosidic oxygen site vacant, which reorients to face away from the molecular frame. This is visualized in **Fig 6.10**.

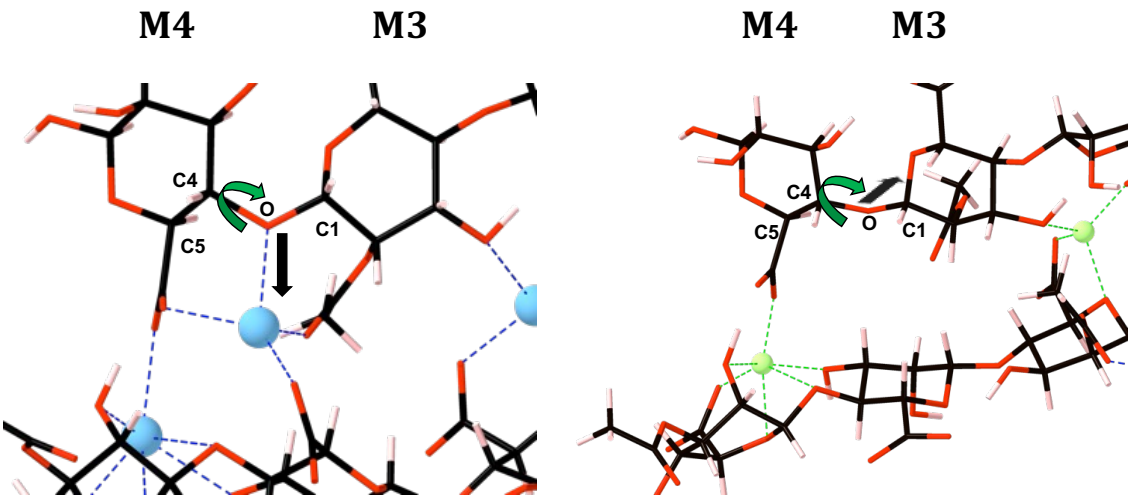


Figure 6.10: Close up perspective of the native calcium chelate site (left) and the vacant coordination site created through substitution **Scheme 1** in the optimised Ga-PolyM$_{ap}$ complex displayed in **Fig 6.5a** (right). The atom labels and ψ angle about the glycosidic linkage are displayed (green arrow) along-side the orientation of the glycosidic oxygen atom relative to the molecular frame (black arrow). Uronate unit nomenclature is also given - note the antiparallel chain arrangement. Carbon atoms are shown in black, oxygen in red, hydrogen in pink, gallium as green spheres and calcium as blue spheres. Bonds to the gallium and calcium ions are shown as green and blue dashed lines respectively.

Recall in **Section 3.4.3.3**, within the calcium ion cross-linked PolyM$_{ap}$ system, the ψ reorientation about the glycosidic oxygen within the top-chain M3-M4 junction, which reoriented the glycosidic oxygen to face the calcium within the intra-complex space, was $-70°$. The vacant binding site generated upon the calcium-for-gallium exchange into the Ca-PolyM$_{ap}$ system stimulates the glycosidic oxygen to *relapse* to its initial orientation facing away from the molecular frame (**Fig 6.10**) and it is this relapse which gives the $\Delta\psi = +45°$ rotation. Effectively, the $\Delta\psi = +45°$ rotation has "reversed" the $\Delta\psi = -70°$ rotation which was stimulated by the initial calcium complexation event.

Given that the large glycosidic torsion rearrangement served to destabilise the Ca-PolyM$_{ap}$ and Mg-PolyM$_{ap}$ cross-linked systems relative to their PolyMG$_p$ analogues, the inference is that the large glycosidic torsion rearrangement upon relapse to its initial non-complexed orientation also offers additional instability. As such, it is evident that the removal of

the calcium from the EPS scaffold (the creation of an empty binding site), as opposed to binding gallium to the EPS scaffold, induces a larger conformational rearrangement in the uronate backbone. The fact that gallium can only induce minor torsion change across the uronate backbone, principally occupying an unaltered native calcium chelate pocket, means it is unable to bring about conformational change required to mitigate against the drop in its CN.

Secondly, the Ga-PolyM$_{ap}$ substitution 2 (**Fig 6.5b**) and co-substitution (**Fig 6.5c**) complexes host ψ reorientations across the top chain M2-M3 junction equal to +29° and +57° respectively, the latter representing the largest global torsion change observed across both EPS scaffolds. Compellingly, this occurs because the M2 uronate residue undergoes a pseudorotation, namely the inter-conversion from the energetically stable chair ($^4C_1$) conformation to the less energetically stable twisted boat conformation, upon gallium exposure. This is visualised in **Fig 6.11**.

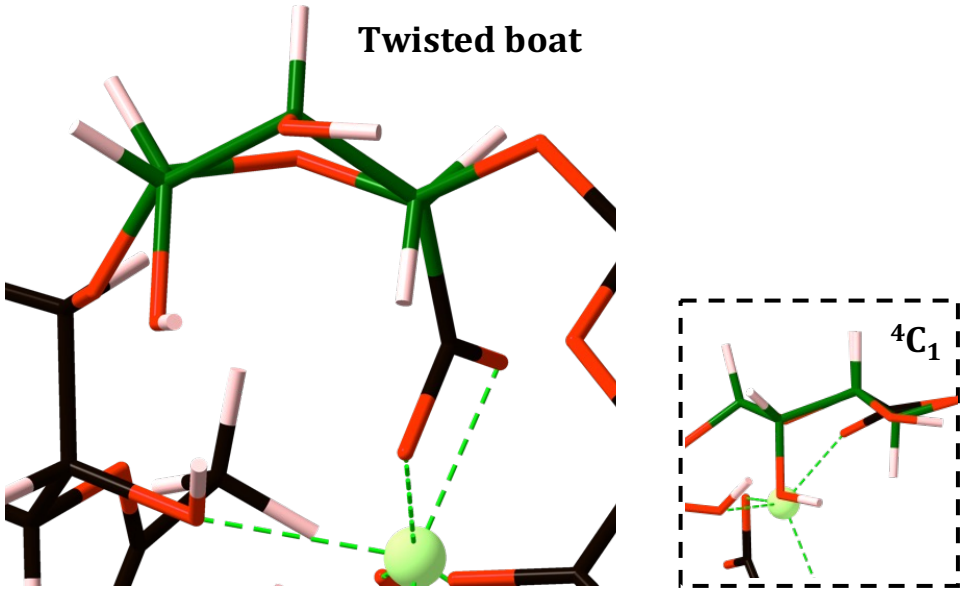


Figure 6.11: Less energetically favoured twisted-boat configuration adopted by the mannuronate residue (M2 top chain) within the Ga-PolyM$_{ap}$ co-substitution gallium complex. Carbon atoms are shown in black, oxygen in red, hydrogen in pink and gallium as green spheres. Bonds to gallium are shown in green. The twisted-boat uronate residue backbone is shown as dark green for clarity. For comparison, the $^4C_1$ chair conformation is also given.

This monosaccharide conformational change has also been observed in DFT structure predictions of trivalent metal ion complexes with non-acetylated algal alginates [142]. At the single monosaccharide scale, this particular (pyranose) configuration lies approximately 5-8 Kcal/mol (0.22-0.35 eV) above the ground-state chair minima [350]. Therefore, it is plausible that this particular conformational change serves to destabilize the gallium-EPS complex at a polysaccharide scale. To ascertain a more representative understanding of this

energetic penalty, the twisted boat conformation at M2 was introduced to the PolyM system (**Fig 3.3 top**) and the energy of the PolyM building unit recomputed. By comparing the relative energy difference, the twisted boat conformation contributes a 4.58 eV instability to the quadramer conformation. It is, therefore, no surprise to observe that the Ga-PolyM$_{ap}$ substitution 2 (**Fig 6.5b**) and co-substitution (**Fig 6.5c**) gallium complexes have reduced stability, approximately 3 eV more unstable relative to the Ga-PolyM$_{ap}$ substitution 1 complex (**Fig 6.5a**). 8 Kcal/mol (0.35 eV) falls well below 4.58 eV and, as such, the mannuronate pseudorotation causes greater destabilization at the polysaccharide (quadramer) scale relative to the monosaccharide scale. This is an EPS structural response that is unique to gallium complexation and was not seen in the divalent ion-EPS complexation events (**Section 3.4.3**).

6.3.5 The primary calcium-for-gallium enthalpic barrier

Despite complexing with large degrees of covalency, the gallium-EPS complexes remain highly thermodynamically unfavourable relative to the native calcium-EPS complexes. Thus far, poorly accessible acidic (anionic) functionality, the +3 charge localization and gallium induced conformational re-arrangements have all been shown to contribute to the instability of the calcium-for-gallium exchange into the EPS scaffold. However, the gallium-induced destabilising conformational re-arrangements are specific to the Ca-PolyM$_{ap}$ scaffold, yet, interestingly, it remains more thermodynamically unfavourable to perform calcium-for-gallium cation exchanges across the Ca-PolyMG$_p$ scaffold.

Calcium ions display a 0.5 eV higher thermodynamic affinity for complexing within M-G junctions on the EPS scaffold (**Table 3.6**) and it is therefore harder to liberate calcium ions from their chelate geometries within these junctions, relative to M-M junctions. This suggests that the stability of the initial calcium complex is the primary influencer of the favourability of the calcium-for-gallium substitution rather than the stability of the resulting gallium complex.

To further exemplify this point, calcium-for-gallium substitutions were introduced into the 4-chain PolyMG EPS system (4-PolyMG*, **Fig 4.4**). To maintain charge balance, the calcium-for-gallium substitution was performed such that all six cross-linking calcium ions were exchanged for four gallium ions. To keep continuity with **Scheme 2** and **3**, the four gallium ions were substituted at the calcium sites which encompassed large CN environments (CN=7 or CN=6) bound to three COO^- groups per ion, as well as to four uronate residues per ion. This is in contrast to the three uronate residues per ion present within the Ca-PolyMG$_p$ EPS system. The formation energy, evaluated according to **Equation 6.2.4**, for the charge balanced exchange of all six cross-linking calcium ions for four gallium ions is +26.83 eV. This structure can be seen in **Fig 6.12**.

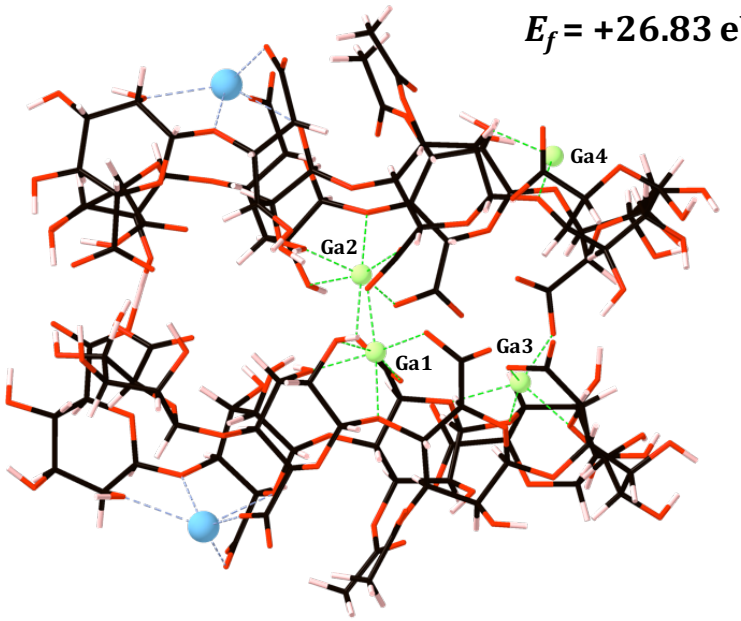


Figure 6.12: Optimised gallium 4-PolyMG* complex, along-side the formation energies (eV). Carbon atoms are shown in black, oxygen in red, hydrogen in pink, gallium as green spheres and calcium as blue spheres. Bonds to the gallium and calcium ions are shown as green and blue dashed lines respectively.

Ga1, Ga2 and Ga3 experience a drop of one in CN, which correlates with the average reduction in CN experienced within the Ga-PolyMG$_p$ scaffolds. These ions retain binding to their COO^- sites. However, Ga4 experiences a drop of five in its CN, as a result of its complexing in an elevated chelate geometry. Explicitly, Ga4 is raised by 2.24 Å relative to the native calcium (**Fig 4.4**), binding at the periphery of the 4-chain intra-complex space, losing all interaction with any COO^- functionality.

The formation energy (+26.83 eV) indicates an extremely unfavourable substitution and gives rise to the following stability trend: no substitution > 1-chain EPS gallium complex > 2-chain EPS gallium complex > 4-chain EPS gallium complex. As the number of calcium cross-linked EPS chains increases, the calcium ions become increasingly encapsulated within the EPS network and, as such, harder to substitute for gallium. This, finally, gives rise to the observation that removal of calcium ions from their native (cross-linked) chelate pockets offers a large enthalpic barrier for the calcium-for-gallium cation exchange.

Given this observation, a final substitution scheme was investigated, whereby the gallium ions were substituted for the calcium ions occupying the sites of the lowest gallium affinity. These sites included calcium ions with the lowest CN and bound to the least number of COO^- groups - specifically, the sites that were not substituted in **Schemes 2** and **3**. The

substitution was performed under **Scheme 2**, coupled with the removal of the remaining calcium that possessed the lowest CN. These structures can be seen in **Fig 6.13**.

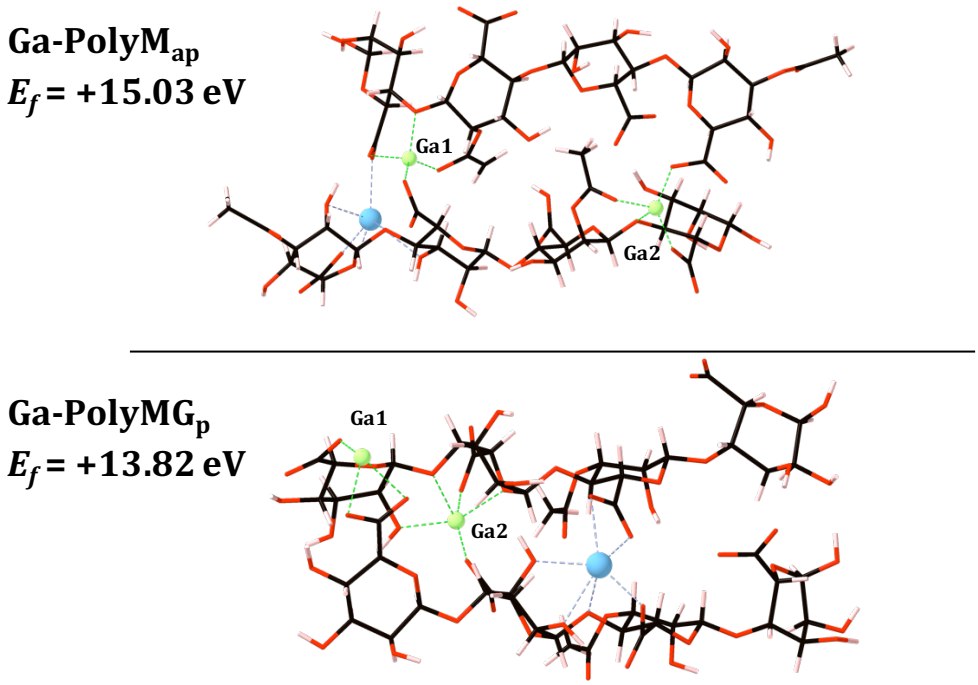


Figure 6.13: Optimised Ga-PolyM$_{ap}$ (top) and Ga-PolyMG$_p$ (bottom) complexes, along-side their formation energies (eV), following substitution into the Ca-PolyM$_{ap}$ and Ca-PolyMG$_p$ EPS scaffolds at the low affinity sites. Carbon atoms are shown in black, oxygen in red, hydrogen in pink, gallium as green spheres and calcium as blue spheres. Bonds to the gallium and calcium ions are shown as green and blue dashed lines respectively.

Given the clear enthalpic barrier to calcium removal from the EPS scaffold, the gallium complex presented in **Fig 6.13** aimed to minimise this barrier by substituting the calcium ions that were most weakly bound by the EPS scaffold. As evident from **Fig 6.13**, this substitution is equally as unfavourable as the complexes presented in **Fig 6.5** and **6.6**. It is becoming increasingly evident that substitution at any calcium cross-linking site in-between two EPS chains is thermodynamically unfavourable. Taken collectively, the structural and thermochemical predictions of the 1-, 2- and 4-chain EPS gallium complexes indicate that exogenous gallium cannot be accommodated by the mature mucoid EPS through substitution of native calcium from its binding site on the mucoid EPS scaffold. This, therefore, suggests that intravenous gallium nitrate therapy (exogenous gallium) does not hinder mucoid *P. aeruginosa* CF biofilm proliferation through inducing calcium-for-gallium cation exchange, and does not therefore create an EPS morphology susceptible to physiological softening.

6.4 Summary

Intravenous gallium therapy is emerging as a novel, non-antibiotic, therapeutic capable of inhibiting *P. aeruginosa* biofilm proliferation through siderophore quenching and represents a viable treatment option to CF patients with chronic *P. aeruginosa* biofilm lung infections. The nature of any interactions between exogenous gallium ions and the native mucoid EPS scaffold has yet to be determined either experimentally or theoretically. Understanding such interactions will assist in elucidating an unknown mechanism of gallium therapy that assists in keeping gallium active against siderophore-deficient cells. To study gallium accommodation by the mature mucoid *P. aeruginosa* EPS, Density-Functional Theory (DFT) was deployed to ascertain whether gallium substitutions into the EPS scaffold in favour of the native calcium ions, which are responsible for facilitating and maintaining stable EPS architecture, were thermodynamically favourable.

Upon substitution for the native calcium ions, the alien gallium ions receive larger degrees of charge transfer from EPS oxygen donors, forming interactions with larger degrees of covalency, relative to the native calcium ions. The COO^- group facilitates the largest degree of charge transfer to the alien gallium ion, forming the most stable interactions, and thus providing sites on the EPS scaffold with the largest gallium affinity. The increase in the covalent character of the oxygen-ion bonds is a key structural factor that allows trivalent ion complexation by the EPS to be distinguished from divalent ion complexation.

However, despite the gain in covalency upon substitution, it was shown that a charged balanced calcium-for-gallium cation exchange into the EPS scaffold was thermodynamically unfavourable at a 1-, 2- and 4-chain EPS scale. Poorly accessible acidic (anionic) functionality, the +3 charge localization and gallium-induced destabilising conformational re-arrangements (the drop in CN, relapse in the glycosidic orientation and mannuronate pseudorotation) have all been shown to contribute to the instability of the calcium-for-gallium exchange into the EPS scaffold. Despite these destabilising structural motifs, the largest enthalpic barrier to the calcium-for-gallium substitution is the removal of the native calcium ions from the EPS scaffold. Overall, the stability of the native calcium chelate geometry influences the favourability of the substitution more so than the stability of the resulting gallium complex. The thermodynamic predictions suggest that it is more stable for gallium to exist as an unbound ion in the EPS vicinity, which, encouragingly, highlights its retained availability for siderophore binding.

Although gallium remains free for siderophore binding in the vicinity of the mucoid *P. aeruginosa* EPS, this observation offers no explanation for why gallium remains active in

siderophore-deficient *P. aeruginosa* cells. In the following chapter, the thermodynamic feasibility of iron-for-gallium substitutions into ferric-PQS complexes will be investigated, as gallium interfering with PQS-mediated ferric iron uptake could offer a potential explanation for the observed sustained therapeutic activity in siderophore-deficient mutants. In addition, the following chapter will also detail proof-of-concept calculations behind the design of a novel gallium delivery system.

Gallium-PQS complexes and delivery vector design

7.1 Scope of the chapter

Gallium ions are not accommodated by the mature, calcium cross-linked, mucoid *P. aeruginosa* EPS matrix and, therefore, remain free for siderophore binding. Described in **Section 1.5.2** was the novel insight that PQS, the alkyl quinolone QSAI molecule, can shuttle Ga^{3+} ions into the bacterial cell. PQS is able to sequester Fe^{3+} and forms an adduct with TseF, a substrate produced by the H3-Type IV Secretion System, gaining access to the bacterial cell by binding the TBDR siderophore uptake protein, FptA. It is currently unknown whether gallium can out-compete iron for PQS binding which, in turn, offers a route of entry for Ga^{3+} ions that circumvents siderophore uptake. As such, this chapter considers the thermodynamic feasibility of gallium substitutions into ferric-PQS complexes. If gallium can substitute iron within ferric-PQS complexes, this would offer a another rationale for gallium remaining active against siderophore-deficient *P. aeruginosa* mutants. Finally, proof-of-concept calculations behind the design of a novel gallium delivery vessel are introduced, with the aim to call attention to the use of OligoG CF-5/20 as a potential gallium delivery vessel for CF patients harbouring mucoid *P. aeruginosa* infections.

7.2 Computational details

7.2.1 DFT simulation parameters

All Density-Functional Theory (DFT) calculations were performed using the plane-wave pseudopotential code, CASTEP [280], using a convergence tested cut-off energy of 900 eV, and a Monkhorst-Pack **k**-point grid of 1 x 1 x 1 to sample the Brillouin zone [255]. On-the-fly ultrasoft pseudopotentials [281] and the PBE-TS exchange-correlation functional were employed [240, 261]. The SCF tolerance for the electronic minimisations was set to 1×10^{-7} eV $Atom^{-1}$ and the energy, force and displacement tolerances for the geometry optimisations

were set to 1×10^{-5} eV Atom^{-1}, 0.03 eV Å^{-1} and 1×10^{-3} Å respectively. Following each geometry optimization, Mulliken bond populations [270] were calculated to classify the nature of bonding in each of the complexed structures. All atomistic models were created and visualised using CrystalMaker [341].

Chemical potentials for sodium, calcium and gallium were calculated by their respective zero Kelvin (0 K) energy per atom from the pure metals in their low energy configurations, namely, hexagonal (HCP) sodium, cubic (BCC) calcium and orthorhombic (alpha) gallium. Respectively, the chemical potential of iron was calculated from its respective 0 K energy per atom from the pure metal in its low energy (ferromagnetic) cubic (BCC; alpha) packing configuration.

7.2.2 Gallium complexation by PQS

PQS sequestration by gallium could offer another plausible explanation for why gallium remains active against siderophore deficient mutants. To consider this, iron-for-gallium cation exchanges within PQS complexes were computed and the stability of the resulting gallium complex measured according to **Equation 7.2.1**

$$E_f = E_{gallium-PQS} - (E_{iron-PQS} + \mu_{Ga} - \mu_{Fe}) \, . \tag{7.2.1}$$

Within **Equation 7.2.1**, $E_{gallium-PQS}$ represents the final energy of the gallium substituted PQS complex, $E_{iron-PQS}$ represents the energy of the initial ferric-PQS complex, μ_{Ga} represents the chemical potential of gallium and μ_{Fe} represents the chemical potential of iron.

7.3 PQS sequestration by gallium

7.3.1 Circumventing siderophore uptake

As identified in the previous section, gallium ions cannot be accommodated by the mature mucoid EPS. This underscores the relatively benign effects that exogenous gallium ions have on mucoid EPS architecture. From a therapeutic standpoint, however, this is encouraging, as it suggests that gallium should remain free and available to compete with iron for siderophore binding in the EPS vicinity. However, this does not entirely explain the discontinuity that there is no change in the gallium resistance profile allowing for a continued reduction in biofilm proliferation in *P. aeruginosa* isolates deficient in siderophore uptake systems [220, 225]. It is entirely possible, therefore, that gallium is targeted by another uptake system and circumventing siderophore uptake. Explicitly, gallium uptake in pyoverdine and pyochelin-deficient mutants is approximately 10 pM per 10^8 cells, compared to 20 pM per 10^8 cells in the wild type, at gallium nitrate concentrations of 2 μM [220]. Potentially, this

indicates that another gallium delivery vector is being utilised.

As discussed in **Section 1.5.2**, PQS has been implicated in Fe^{3+} uptake into *P. aeruginosa* cells. PQS can complex Fe^{3+} and these ferric-PQS complexes can be escorted into the cell by binding TseF, a substrate secreted by the H3-T6SS (type VI secretion system) [229]. It is currently unknown, both experimentally and theoretically, whether gallium ions can substitute iron from ferric-PQS complexes, potentially exploiting this particular uptake system and circumventing siderophore-mediated delivery.

7.3.2 Control models of siderophore competition

To effectively compare gallium uptake into ferric-PQS complexes to the traditional siderophore uptake, control calculations were performed on iron-for-gallium substitutions into ferric-pyochelin. Gallium has been shown theoretically, through DFT free energy binding calculations [223, 224], and experimentally, through X-ray absorption near edge structure (XANES) and extended X-ray absorption fine structure (EXAFS) spectroscopies [351], to complex within the pyochelin active site in preference to native iron. This particular binding event leads to the siderophore quenching therapeutic activity of Ga^{3+} ions [218].

The stoichiometry of ferric-pyochelin complexes in extracellular medium was discussed in **Section 1.5.2**. In particular, the monochelate complex dominates and the remaining bidentate coordination space is saturated by hydroxide and/or water molecules, giving an overall octahedral complex geometry [207]. To maintain charge balance, giving charge neutral ferric-pyochelin complexes, one water molecule and one hydroxide moiety were chosen to saturate the remaining bidentate coordination space. Pyochelin can chelate Fe^{3+} in either an octahedral Low-Spin (S=1/2; doublet) or octahedral High-Spin (S=5/2; sextet) geometry and these optimised complexes are given in **Fig 7.1**.

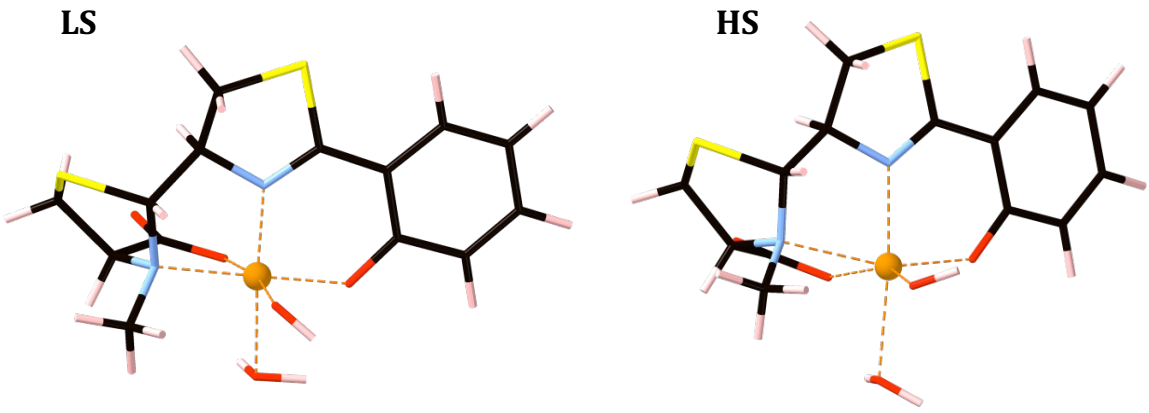


Figure 7.1: Low-spin (left) and high-spin (right) ferric-pyochelin complexes. Carbon atoms are shown in black, oxygen in red, nitrogen in blue, hydrogen in pink, sulphur in yellow and iron as brown spheres. Bonds to the iron centre are shown in brown.

The geometrical parameters for the LS and HS ferric-pyochelin complexes (**Fig 7.1**) are given in **Table 7.1**.

Table 7.1: Bond lengths (Å) and Mulliken populations (|e|) for all pyochelin-Fe^{3+} interactions present within the LS and HS ferric-pyochelin complexes.

LS ferric-pyochelin		
bond	Bond length (Å)	Mulliken population (|e|)
sp2 N-Fe^{3+}	1.84	0.22
Hydroxide O-Fe^{3+}	1.84	0.34
Phenolate O-Fe^{3+}	1.88	0.3
COO^- O-Fe^{3+}	1.95	0.26
sp3 N-Fe^{3+}	2.06	0.21
Water O-Fe^{3+}	2.10	0.02
Average O-Fe^{3+}	1.94	0.23
Average N-Fe^{3+}	1.95	0.22
Average for all donors	1.95	0.23
HS ferric-pyochelin		
bond	Bond length (Å)	Mulliken population (|e|)
Hydroxide O-Fe^{3+}	1.89	0.33
Phenolate O-Fe^{3+}	1.94	0.27
COO^- O-Fe^{3+}	2.00	0.25
sp2 N-Fe^{3+}	2.08	0.21
Water O-Fe^{3+}	2.31	0.02
sp3 N-Fe^{3+}	2.37	0.13
Average O-Fe^{3+}	2.03	0.22
Average N-Fe^{3+}	2.22	0.17
Average for all donors	2.10	0.20

The hydroxide and phenolate contacts establish the most tightly bound interactions with the central Fe^{3+}; these interactions occur with the shortest bond lengths and largest Mulliken populations (**Table 7.1**). The water molecule, which establishes the weakest interactions with the iron centre in both the LS and HS complexes, is evidently a spectator ligand, establishing transient interactions with the central Fe^{3+} ion. In contrast, the hydroxide ion is critical for facilitating ferric-pyochelin complex stability and, therefore, is a key ligand in mediating Fe^{3+} sequestration by pyochelin. It is clear from **Table 7.1** that the charge transfer, measured from the Mulliken populations, from oxygen and nitrogen donors to the central Fe^{3+} is equal in the LS state and, consequently, the average O-Fe^{3+} and N-Fe^{3+} lengths are equivalent. However, charge transfer from nitrogen donors is reduced in the HS state, which generates asymmetry in the average O-Fe^{3+} and N-Fe^{3+} bond lengths, leading to a distorted octahedral complex geometry in the HS state.

The average length of the ferric chelate pocket, or pocket size (Å), measured as the average donor-Fe^{3+} bond length in the LS and HS ferric-pyochelin complexes are shown in **Table 7.2**. Also shown, is how these distances deviate from the average ferric chelate pocket sizes determined experimentally through X-ray diffraction [209, 352] and EXAFS [351] performed on the ferric-pyochelin crystal structure isolated from *P. aeruginosa* cultures. Note, in these experimental measurements, no information on the spin state of the iron centre is obtained.

Table 7.2: Deviations in the size of the ferric-pyochelin chelate complex in these models from experimental structures.

Complex	Pocket size (Å)	Deviation from experiment (Å)	% difference
HS ferric-pyochelin	2.10	0.08 [209]	3.68
		0.04 [352]	1.96
		0.09 [351]	4.50
LS ferric-pyochelin	1.95	0.16 [209]	7.36
		0.23 [352]	11.28
		0.06 [351]	3.00

From **Table 7.2** it is clear that there is better agreement between the predicted HS ferric-pyochelin structure and the ferric-pyochelin observed experimentally. Explicitly, the average % difference for the predicted HS ferric-pyochelin complex is 3.46% compared to 7.21% in the LS case. This is strong evidence to suggest that, even without the need of comparing the relative energies of the two spin states, the HS ferric-pyochelin complex is the one that occurs experimentally (*in vitro*). For reference, the HS is indeed 0.57 eV more stable than the LS state.

Iron-for-gallium substitutions were therefore performed on the HS ferric-pyochelin complex. The gallium substituted pyochelin complex is given in **Fig 7.2** along-side the formation energy calculated according to **Equation 7.3.1**

$$E_f = E_{gallium-pyochelin} - (E_{iron-pyochelin} + \mu_{Ga} - \mu_{Fe}) \, . \tag{7.3.1}$$

Within **Equation 7.3.1**, $E_{gallium-pyochelin}$ represents the final energy of the gallium-pyochelin complex following the substitution and $E_{iron-pyochelin}$ represents the energy of the HS ferric-pyochelin complex (**Fig 7.1**).

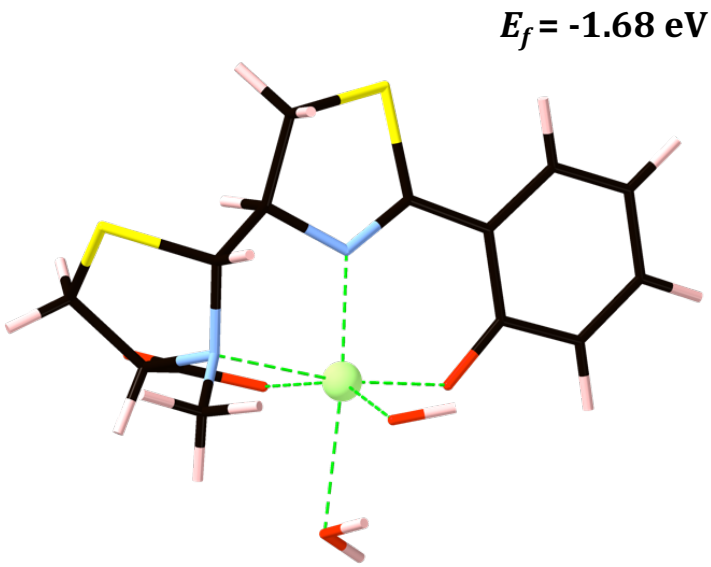


Figure 7.2: Gallium-pyochelin complex following iron-for-gallium exchange into the HS ferric-pyochelin complex, along-side the formation energy (eV). Carbon atoms are shown in black, oxygen in red, nitrogen in blue, hydrogen in pink, sulphur in yellow and gallium as green spheres. Bonds to the gallium centre are shown in green.

Gallium substitution into ferric-pyochelin gives rise to the siderophore quenching therapeutic activity of Ga^{3+} ions [218] and it is encouraging to see the theoretical approach utilised here returns a negative formation energy for this process, confirming that the endpoint of this reaction is thermodynamically feasible. The computed formation energy given in **Fig 7.2** is now the *control* against which the feasibility of gallium-PQS complexation can be compared against.

7.3.3 Gallium-PQS complexation events

At physiological pH, three PQS molecules chelate a single ferric (Fe^{3+}) centre through a bidentate coordination mode facilitated by its ketone and OH functional groups [174]. In fact, 2-heptyl-4-quinolone (HHQ), a structural analogue of PQS that lacks the OH group, cannot bind iron, highlighting the critical role the OH group plays in iron sequestration [174]. Given that the phenolate group offers a key coordination mode in pyochelin (**Fig 1.10**) [209] as well as in ferric hydroxyquinolone [353], it is likely the phenolate group, rather than the phenol group, is involved in the ferric complexation event giving an overall charge neutral complex $[Fe(PQS)_3]$.

Three PQS bidentate ligands can chelate Fe^{3+} [173, 174] in either an octahedral low-spin (S=1/2; doublet) or octahedral high-spin (S=5/2; sextet) geometry. Initial DFT geometry optimizations using CASTEP show the high-spin $[Fe(PQS)_3]$ complex to be 0.057 eV

(5.50 KJ/mol) more stable than the low-spin complex and, as such, can be considered as the dominant spin-polarisation. Therefore, the high-spin [Fe(PQS)$_3$] complex was used as the model system to study iron-for-gallium substitutions into ferric-PQS. The greater stability of the HS system can be understood through analysing the LS and HS complex geometries, given in **Fig 7.3**.

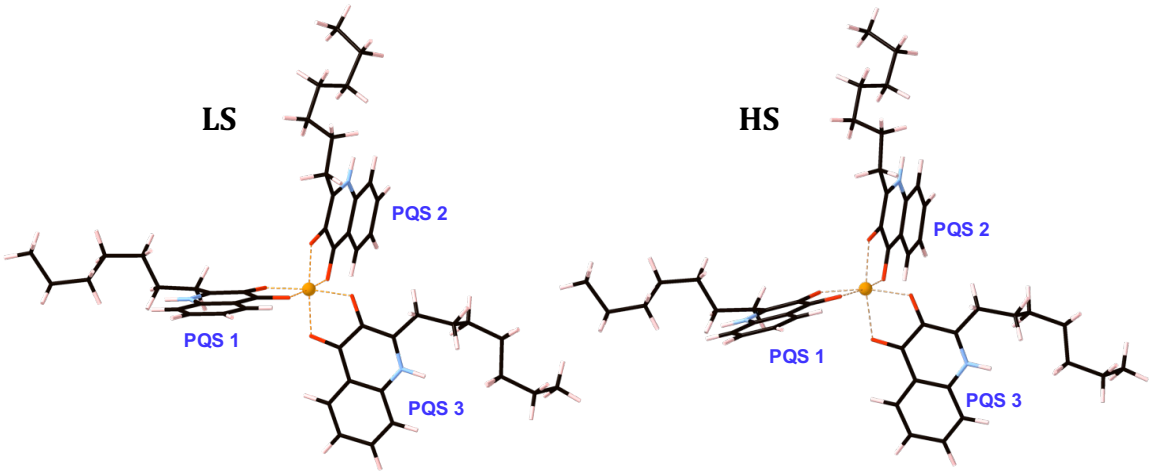


Figure 7.3: Low-spin (left) and high-spin (right) ferric-PQS complexes. Carbon atoms are shown in black, oxygen in red, nitrogen in blue, hydrogen in pink and iron as brown spheres. Bonds to the iron centre are shown in brown. Note the tilted orientation of the PQS1 in the HS state.

In both the HS and LS ferric-PQS complexes (**Fig 7.3**), the Fe^{3+} centre adopts an octahedral geometry, which is distorted in the HS state. This distortion causes PQS 1 to tilt its orientation relative to PQS 3. This is better visualised in **Fig 7.4**, in which a plane has been drawn through the centre of the octahedral iron centre. From this, it can be seen that PQS 1 in the LS state lies almost perfectly in this plane, deviating away from the plane by 1°. In contrast, PQS 1 in the HS state deviates by 32° away from this plane.

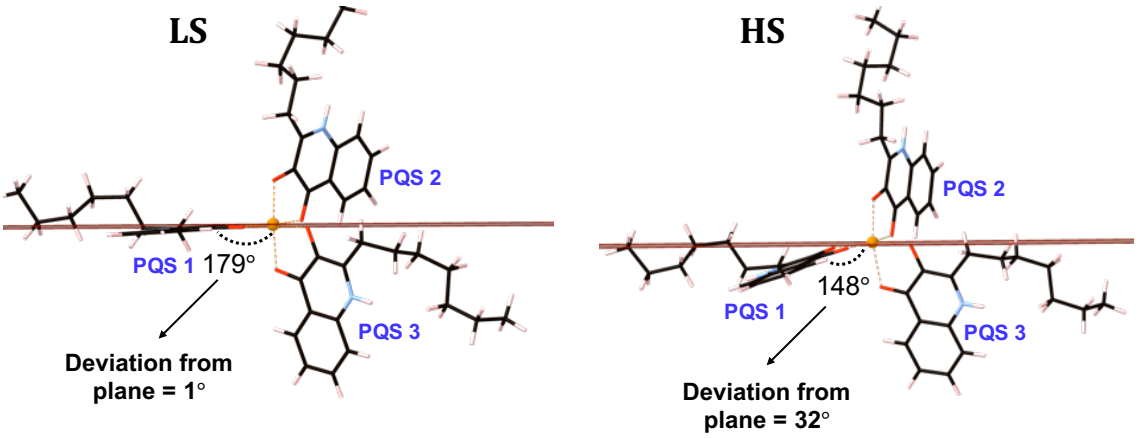


Figure 7.4: Low-spin (left) and high-spin (right) ferric-PQS complexes. Carbon atoms are shown in black, oxygen in red, nitrogen in blue, hydrogen in pink and iron as brown spheres. Bonds to the iron centre are shown in brown. A plane has been drawn through the iron centre to visualise the difference in relative PQS orientation between the two spin states.

Recall, the HS ferric centre within 6-coordinate pyochelin complexes also resembled a distorted octahedron, leading to the observation that this particular distortion is, potentially, a structural motif unique to HS Fe^{3+} centres within *P. aeruginosa* ferric uptake systems consisting of oxygen (and nitrogen) donors.

The distortion in the HS ferric centre is better characterised through analysis of the phenolate O-Fe^{3+}-O ketone bite angle, where a bite angle closer to 90° would indicate a smaller deviation away from a pure octahedral geometry. In the LS ferric-PQS complex, the average phenolate O-Fe^{3+}-O ketone bite angle is 85.67°, a deviation of 4.33° away from 90°. In contrast, within the HS ferric-PQS complex, the average bite angle is 79.67°, a deviation of 10.33° away from 90°. Coupled with the larger deviation away from 90° in the average bite angle, the HS spin complex encompasses an expansion of the ferric chelate pocket, relative to the LS complex. The average O-Fe^{3+} bond length in the LS complex is 1.93 Å, where the average O-Fe^{3+} bond length is 2.06 Å in the HS complex. The comparatively more expanded chelate pocket in the HS ferric-PQS complex giving bite angles with larger deviations away from 90°, ultimately resulting in the adoption of a minorly distorted octahedral geometry, serves to better separate the PQS aromatic systems as visualised in **Fig 7.5**. In turn, this reduces the overall inter-PQS steric repulsion and further stabilises the overall complex geometry.

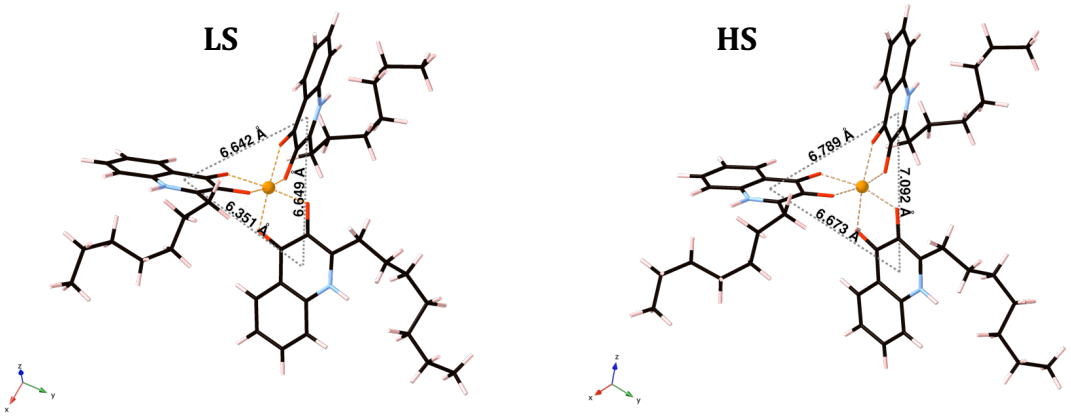


Figure 7.5: The separations (Å) between the PQS aromatic systems in the LS and HS complexes. Carbon atoms are shown in black, oxygen in red, nitrogen in blue, hydrogen in pink and iron as brown spheres. Bonds to the iron centre are shown in brown.

The gallium PQS complex, following the iron-for-gallium cation exchange into HS $[Fe(PQS)_3]$, is shown in **Fig 7.6**.

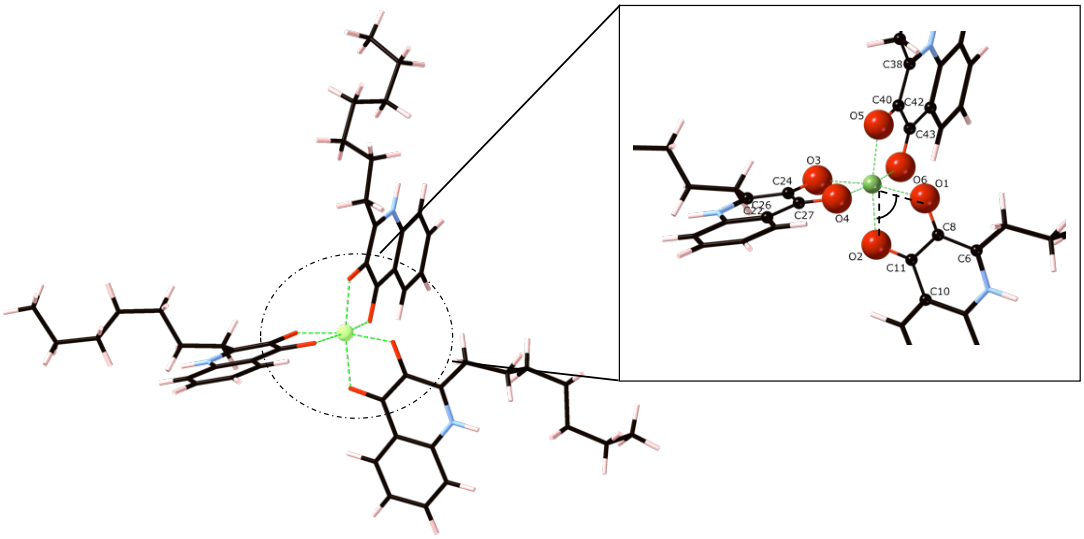


Figure 7.6: Gallium PQS complex following the iron-for-gallium cation exchange into the high-spin ferric-PQS complex. Carbon atoms are shown in black, oxygen in red, hydrogen in pink and gallium as green spheres. Bonds to the gallium centre are shown in green. An expanded view of the gallium coordination sphere is also displayed, along-side an illustration of the phenolate O-Ga^{3+}-O ketone bite angle.

Encouragingly, this substitution is thermodynamically stable, with **Equation 7.2.1** returning a negative formation energy of −1.81 eV. As such, exogenous gallium can substitute iron from its native binding site within PQS complexes. Additionally, this formation energy closely resembles the formation energy of the iron-for-gallium substitution into ferric-pyochelin

($E_f = -1.68$ eV, **Fig 7.2**), from which it can be inferred that these processes are equally as feasible and, in principle, likely to occur *in vivo*.

Within the gallium-PQS complex (**Fig 7.6**), the phenolate establishes the tightest and most stable interaction with the central gallium ion (average bond length and population equal to 1.99 Å and 0.32 |e| respectively) relative to the ketone group (average bond length and population equal to 2.05 Å and 0.27 |e| respectively). The 0.06 Å shortening of the O-phenolate interaction matches within 3 % of those observed in other gallium-organic ligand systems, namely the crystal structures of gallium-chrysin complexes and DFT optimised gallium-5-hydroxyflavone complexes [354, 355]. The average phenolate O-Ga^{3+}-O ketone bite angle (**Fig 7.6**) within the gallium-PQS complex is 82.83°, also agreeing within 3 % with the same bite angles observed within gallium-chrysin complexes and DFT optimised gallium-5-hydroxyflavone complexes [354, 355]. As can be seen, an average bite angle of 82.83° deviates from 90° indicating that the geometry of the gallium chelate pocket diverges slightly from pure octahedral coordination. This is further supported by analysing the Ga^{3+}-O-C-C dihedral angles, which are given in **Table 7.3**.

Table 7.3: Ga^{3+}-O-C-C dihedral angles present within the gallium-PQS complex following iron-for-gallium substitution into the high-spin ferric-PQS complex.

Plane	Dihedral angle (°)
Ga^{3+}-O1-C8-C6	+172.90
Ga^{3+}-O2-C11-C10	-172.54
Ga^{3+}-O3-C24-C22	-172.25
Ga^{3+}-O4-C27-C26	+173.50
Ga^{3+}-O5-C40-C38	-176.42
Ga^{3+}-O6-C43-C42	+179.40

Ga^{3+}-O-C-C dihedral angles equal to ±180° would indicate that the donating oxygen atoms, from a single bidentate PQS ligand, lie within the same plane. As shown in **Table 7.3**, these angles deviate by up to 7.75° away from 180°, further highlighting that the gallium-PQS complex, is a distorted octahedron with the PQS ligands adopting a quasi-planar coordination geometry.

The retention of this particular metal complex geometry gives rise to the very low global (all atom) RMSD value of 0.062 Å following alignment of the high-spin [$Fe(PQS)_3$] and [$Ga(PQS)_3$] complexes. The extremely high degree of structural similarity between the between the high-spin ferric-PQS complex and the gallium PQS complex truly exemplifies the Trojan Horse mechanism exhibited by Ga^{3+} ions. That is, the principle that gallium

ions are able to establish near identical complex geometries within ferric uptake systems. The inference is that given the near identical geometries of the high-spin [$Fe(PQS)_3$] and [$Ga(PQS)_3$] complexes, it is entirely conceivable that TseF-PQS-Ga^{3+} adducts could trigger gallium uptake in a similar fashion to iron uptake [229].

Finally, one additional DFT calculation was performed to determine whether gallium could sequester *free* PQS ligands from the extracellular medium, in addition to competing for native ferric-PQS complexes. The thermodynamic stability of this gallium free-PQS complexation event was evaluated according to **Equation 7.3.2**

$$E_f = E_{gallium-PQS} - (3E_{PQS} + \mu_{Ga} - 3\mu_H) \tag{7.3.2}$$

where E_{PQS} is the energy of the neutral PQS molecule. **Equation 7.3.2** returned a formation energy of -5.08 eV, indicating that a complexation event between free gallium ions and free PQS ligands is thermodynamically feasible. Moreover, not only is this substitution more favourable compared to ferric-PQS, but it also is more favourable than in the HS ferric-pyochelin complexes. This is to be expected given the enthalpic barrier associated with removing the native ion. This is very similar to the gallium-EPS models. Furthermore, this identifies a more favourable mechanism of gallium uptake. The lowest energy route for gallium to gain access to *P. aeruginosa* bacterial cells is to complex with free PQS and, if this is not possible, gallium can proceed to invade both ferric-pyochelin and ferric-PQS uptake systems.

Overall, gallium is predicted to target the ferric-PQS uptake system on two fronts: by displacing Fe^{3+} from its native binding site within the HS [$Fe(PQS)_3$] complex and by sequestering free PQS ligands from the extracellular medium, thereby reducing the number of active iron sequestering molecules from the extracellular medium.

7.4 Designing a novel gallium delivery vessel

A thermodynamically stable gallium tri-polyguluronate scaffold is shown in **Fig 6.9**. Initially, this model was deployed to test whether Ga^{3+} ions could complex favourably into vacant -3 charge pockets. However, given the success of recent novel gallium-polysaccharide complexes, such as gallium carboxymethyl cellulose, alginate and pectin scaffolds, at delivering gallium to treat *P. aeruginosa* infections *in vitro* [222, 227, 344], this model was further utilised to assess its suitability of a novel gallium-polyguluronate delivery vector. Gallium ion release from polysaccharide delivery scaffolds is stimulated through enzymatic degradation [344] and hydrolysis [222], as well as through cation exchange with the surrounding cationic atmosphere [227]. It is the latter mechanism of release that can be tested using the gallium

tri-polyguluronate scaffold model (**Fig 6.9**).

OligoG CF-5/20 was introduced in **Section 1.5.1** and its effects on the mature mucoid EPS modelled in **Section 5.3**. The therapeutic mechanism behind OligoG CF-5/20 disruption of mature *P. aeruginosa* biofilms is critically dependent upon calcium (Ca^{2+}) ions, with the OligoG CF-5/20 scaffold forming highly thermodynamically stable complexes with EPS bound calcium (**Fig 5.2**). Given the high thermodynamic affinity shown by this molecule towards calcium, which, as discussed, is elevated in CF sputum, the question arises as to whether OligoG CF-5/20 could be used as a gallium delivery vessel. The gallium tri-polyguluronate complex (**Fig 6.9**) is constructed by complexing two Ga^{3+} ions in-between three OligoG CF-5/20 chains and is, therefore, a suitable model system for studying OligoG CF-5/20 as a possible gallium delivery vector. More specifically, could the gallium tri-polyguluronate scaffold undergo gallium-for-calcium exchange, with calcium coming from the elevated cationic atmosphere, stimulating the release of Ga^{3+} ions? In turn, this creates a bi-functional therapeutic approach to treat mucoid *P. aeruginosa* CF lung biofilm infections, as the liberated OligoG CF-5/20 and gallium ions would be free to perform their separate therapeutic functions.

Gallium-for-calcium substitutions into the gallium tri-polyguluronate delivery scaffold were performed according to **Scheme 4**

$$\textbf{Scheme 4} : 2 \times Ga^{3+}(+6) \Longrightarrow 3 \times Ca^{2+}(+6)\ . \tag{7.4.1}$$

The formation energy for this substitution scheme is given in **Equation 7.4.2**

$$E_f = E_{calcium-3G} - (E_{gallium-3G} + 3\mu_{Ca} - 2\mu_{Ga}) \tag{7.4.2}$$

where $E_{calcium-3G}$ represents the final energy of the calcium tri-polyguluronate system and $E_{gallium-3G}$ represents the energy of the initial gallium tri-polyguluronate scaffold.

Two calcium ions are substituted into the native gallium sites, complexed within the -3 binding pockets, with the additional calcium ion positioned in the vacant G-G junction. The final calcium tri-polyguluronate complex, along-side its formation energy, is given in **Fig 7.7**.

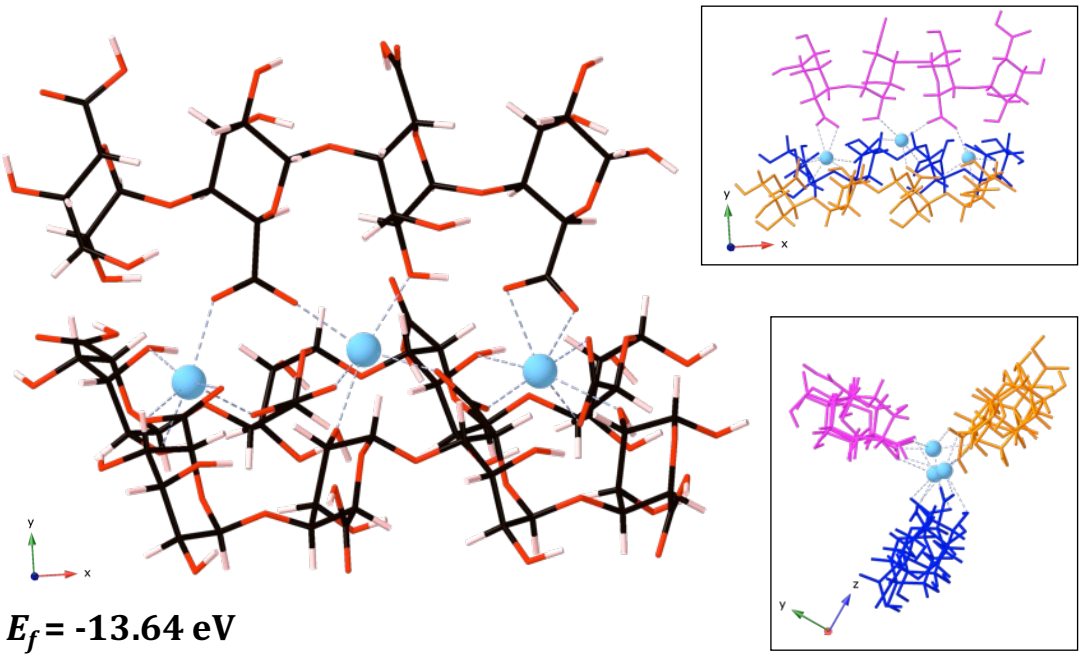


Figure 7.7: Calcium tri-polyguluronate system, along-side the formation energy (eV), following calcium substitution into the gallium tri-polyguluronate. Carbon atoms are shown in black, oxygen in red, hydrogen in pink and calcium as blue spheres. Bonds to the calcium ions are shown in blue. Included within the inset is the same structure, but each polyguluronate chain can be better distinguished based on colour.

Immediately it is evident that this substitution is highly stable thermodynamically. The formation energy, outlining the feasibility of generating the gallium tri-polyguluronate scaffold, although favourable, was small ($E_f = -0.24$ eV) suggesting the native Ga^{3+} ions were only weakly bound to the scaffold. In turn, this gives rise to a small enthalpic barrier associated with the removal of gallium and, along-side the large thermodynamic affinity that guluronate scaffolds (OligoG CF-5/20) have for calcium ions, as has been critically illustrated throughout this work. This offers substantial energetic potential for the gallium-for-calcium exchange.

It is interesting to consider that the formation energy for this substitution (**Fig 7.7**) is highly similar to the formation energies calculated for the EPS-OligoG CF-5/20 equilibrium binding configurations ($E_f = -13.05$ eV, **Fig 5.2 8 and 10 ns**). The calcium induced gallium release from the gallium tri-polyguluronate scaffold is as thermodynamically favourable as the binding of OligoG CF-5/20 to EPS-bound calcium. As such, it would be natural to make the inference that the calcium available within the EPS matrix could, potentially, act as the calcium initiators behind gallium release. However, recall from **Section 5.5**, that the OligoG CF-5/20 molecule is unable to liberate calcium from its native binding site on the EPS scaffold. Therefore, rather than calcium dissociating from the EPS to initiate gallium release

from the tri-polyguluronate scaffold, it is more likely that the initiating calcium ions will come from the *free* calcium present within the elevated cationic atmosphere. Encouragingly, sequestration of free calcium, preventing further calcium-induced EPS aggregation, is one of the therapeutic modes of action of OligoG CF-5/20 against the mature mucoid *P. aeruginosa* EPS [183, 184] and, therefore, the gallium-for-calcium exchange into the gallium tri-polyguluronate scaffold highlights the bi-functional nature of this delivery system; the release of gallium ions which do not interfere with the native EPS scaffold, remain free for siderophore binding, and the sequestration of calcium ions by OligoG CF-5/20 preventing further calcium-induced EPS aggregation.

To gauge the exothermic nature of gallium release from the gallium-OligoG CF-5/20 scaffold against current gallium delivery vectors, the formation energy of gallium-for-calcium exchange was measured for gallium nitrate ($Ga(NO_3)$) and gallium maltolate delivery vectors. Gallium nitrate and gallium maltolate have been demonstrated to possess minimum inhibitory concentrations (MIC) of 16 and 8 μM respectively against *P. aeruginosa* cultures grown in culture medium complimented with 10 % human serum *in vitro* [356, 357]. Both these therapies, hopefully, represent a non-antibiotic therapeutic alternative for treating *P. aeruginosa* infections. The optimised gallium nitrate and maltolate structures can be seen in **Fig 7.8**.

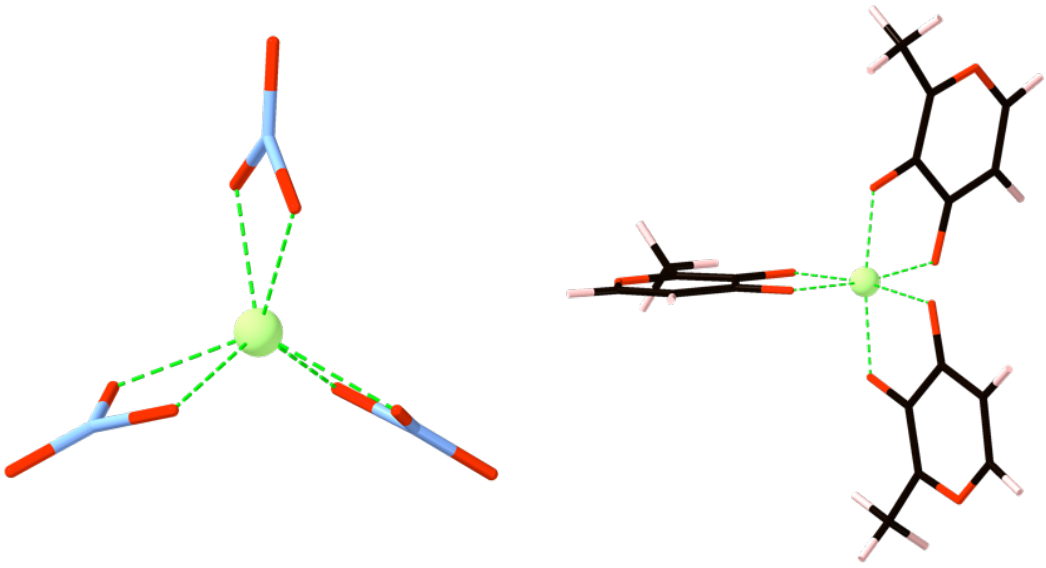


Figure 7.8: Gallium delivery vectors: gallium nitrate (left) and gallium maltolate (right). Carbon atoms are shown in black, oxygen in red, hydrogen in pink, nitrogen in blue and gallium as green spheres. Bonds to the gallium ion are shown in green.

The substitution into the gallium nitrate and maltolate delivery vectors was performed

according to **Scheme 5**

$$\textbf{Scheme 5} : 1 \times \mathrm{Ga}^{3+}(+3) \Longrightarrow 1 \times \mathrm{Ca}^{2+} + 1 \times \mathrm{H}\ (+3)\ . \tag{7.4.3}$$

In **Scheme 5**, the addition of the proton was to a single nitrate or the phenolate group of a single maltolate molecule. The formation energies for calcium induced gallium release according to **Scheme 5** for gallium nitrate and gallium maltolate (**Fig 7.8**) are calculated according to **Equation 7.4.4**

$$E_f = E_{calcium-complex} - (E_{gallium-delivery} - \mu_{Ga} + \mu_{Ca} + \mu_H)\ . \tag{7.4.4}$$

In **Equation 7.4.4**, $E_{calcium-complex}$ represents the final energy of the calcium substituted complex and $E_{gallium-delivery}$ represents the energy of the initial gallium delivery complex - either gallium nitrate or gallium maltolate. The formation energies for the calcium induced gallium release from the gallium nitrate and gallium maltolate delivery vectors are -3.21 eV and -1.96 eV respectively. Although favourable, these values indicate that calcium-initiated release from the tri-polyguluronate scaffold (**Fig 7.7**) is far more thermodynamically stable and provides a theoretical justification for the idea that a tri-polyguluronate scaffold maybe a better mode of gallium delivery as it offers a more exothermic release.

The rate of a chemical process is related to the energy change between reactants and products (ΔE), which in this case is essentially the formation energy (E_f). This is given by the Eyring relationship [342, 343]

$$\text{rate} \propto \exp \frac{-\Delta E}{K_b T}\ . \tag{7.4.5}$$

From **Equation 7.4.5**, given a constant temperature of 310 K, the more negative the formation energy (E_f), effectively the more exothermic the conversion of reactant(s) to product(s), the faster the rate of the chemical change. The release of gallium from the tri-polyguluronate scaffold is more exothermic compared to the nitrate and maltolate scaffolds, in turn, suggesting that the tri-polyguluronate scaffold may give rise to faster release kinetics *in vivo* at 310 K. However, rate estimation (**Equation 7.4.5**) is better evaluated using using the energy of activation to reach the transition state (E_a), rather than E_f, which follows from transition state theory. This would require knowledge of both the energy and structure of the transition state accessed along the gallium-for-calcium exchange reaction coordinate which, although possible through DFT computation, would be computationally demanding for a system of this size (**Fig 7.7**).

This work has already highlighted a firm affinity for the EPS scaffold shown by OligoG CF-5/20 and, now, has identified favourable calcium-induced gallium release from a gallium complexed tri-polyguluronate. Given the large thermodynamic affinity shown by OligoG CF-5/20 for the EPS, this, potentially, may translate into precise targeting of gallium release in the mature mucoid *P. aeruginosa* EPS in the CF lung *in vivo*. Therefore, with regards to therapeutic development, there is a strong thermodynamic rationale behind the applicability of gallium-tri-polyguluronate delivery scaffolds that may be able to exploit a bi-functional therapeutic mechanism - namely, the release of gallium ions stimulated through the sequestration of free calcium ions.

7.5 Summary

As gallium is not able to interact with the EPS, it is free to interfere with ferric uptake systems. For the first time, it has been demonstrated that gallium ions can bind PQS, a QSAI molecule that can facilitate iron (Fe^{3+}) uptake through binding TseF - a substrate secreted by H3-T6SS. This offers another mechanism of gallium uptake that can be extended to *P. aeruginosa* cells that are deficient in the siderophores pyochelin and pyoverdine. Gallium is able to complex free PQS ligands, as well as substitute the native ferric centre out of the PQS chelate pocket, in a thermodynamically stable fashion. The very high structural similarity upon substitution into the native ferric-PQS complex potentially suggests that this gallium-PQS complex could undergo adduct formation with TseF in a similar fashion to the native ferric-PQS complex. This, in turn, offers a route into *P. aeruginosa* bacterial cells that circumvents siderophore uptake completely.

Finally, a novel gallium delivery system, based on a tri-polyguluronate complex (two Ga^{3+} ions complexed in-between three OligoG CF-5/20 chains) was designed and proof-of-concept calculations demonstrate a highly exothermic gallium release under exposure to calcium ions. Release of gallium from the tri-polyguluronate scaffold was far more exothermic relative to release from the nitrate and maltolate scaffolds currently deployed, providing a theoretical justification for the idea that a tri-polyguluronate scaffold maybe a better vector for gallium delivery. Utilising OligoG CF-5/20 as the delivery vector could give rise to precise targeting in the mature mucoid *P. aeruginosa* EPS, in addition to exploiting a bi-functional therapeutic mechanism - the release of gallium ions stimulated through the sequestration of free calcium ions. This would offer therapeutic synergism in one therapeutic agent.

Discussion, future work and conclusions

8.1 Atomistic model development

8.1.1 Elucidating biofilm matrix structural chemistry

The establishment of mature mucoid *P. aeruginosa* biofilm infections in the lungs of cystic fibrosis (CF) patients leads to lung destruction, severely reduced pulmonary function and, ultimately, death. The mucoid *P. aeruginosa* biofilm matrix is able to retard conventional antibiotic therapy and, as such, presents an extreme pharmacological challenge. Not only are conventional antibiotic strategies ineffective, but also unsustainable given the plight of the growing antibiotic resistance crisis. Therefore, there is an urgent need to better understand the key biofilm structural chemistry that gives rise to pathogenic phenomena *in vivo*, most notably, the development of highly stable biofilm scaffold architectures that reduce antimicrobial penetration. This will serve to offer new therapeutic insights that can guide the development of new antimicrobial chemotherapies.

Atomistic models are designed to predict the stable microstates of a chemical system, as these are most likely to occur experimentally. From these states, it is then possible to isolate key molecular functionality critical to biophysical function or, specifically in the case of this research, bacterial virulence and chronicity. The structural observations made through atomistic modelling offer deeper structural-chemical insight than can be afforded by experimental (macroscopic) measurement. Predicting the state(s) adopted by the chemical system can be done through first principles modelling, finding the stable molecular structures that provide the variational minima in the electron density, or by molecular dynamics modelling, sampling an array of structures that occur at finite temperatures in the presence of solvent. Both were utilised in this research.

Highlighting the structural-chemical pre-requisites behind mucoid *P. aeruginosa* biofilm matrix stabilisation, along-side how molecules move through matrix architecture, will help pharmacologists and drug designers to develop novel small molecule antimicrobials

with increased selectivity and efficacy, targeted at matrix eradication and/or deep matrix penetration. This is something that has not before been probed experimentally or theoretically. Structurally similar atomistic models of algal alginate have been developed previously [134, 135, 137, 140, 142, 143, 295, 312], and along-side the microscopic and spectroscopic characterisation of the mucoid *P. aeruginosa* biofilm matrix [93, 102, 111, 112, 122, 123, 130, 175, 308, 309], serve as a useful comparison when evaluating the structural-chemical predictions of the models constructed in this work. The remainder of the chapter will discuss where the model predictions both align with and deviate from the results of others and how the combination of our findings and the findings of others can contribute new insight.

8.1.2 Models of the mucoid *P. aeruginosa* ECM

8.1.2.1 Model representability

The largest systematic error intrinsic to any atomistic model is the representability. One seeks to develop a model as close to the true chemical system (experimental or *in vivo*) as possible. However, in nearly all cases, such modelling is non-trivial. The reason for this stems from computational tractability, explicitly, it is not computationally feasible to include every particle present within the bulk system. This would lead to models that have atom numbers on the order of Avogadro's number, that would be impossible to deploy in simulation on current high-performance computing architectures. To that end, all models require simplification, with the only requirement being that one must ensure not to oversimplify the model to the point where it has poor or no degree of representability at all.

The aim of this research was to develop the first atomistic model of the main component of the mucoid *P. aeruginosa* biofilm matrix, bacterial alginate polysaccharide, and deploy this model in simulations to unpick some of the structural chemistry (structure-directed interactions) that govern both virulence and molecular motion through the ECM. Mucoid *P. aeruginosa* establishes a biofilm matrix in the lungs of CF patients that is primarily composed of a single ECM constituent - anionic, acetylated, alginate exopolysaccharide (EPS) [1, 93, 103, 107]. In fact, in the lungs of recently deceased CF patients who died from chronic mucoid *P. aeruginosa* lung biofilm infections, the EPS is the only detectable extracellular material present, regardless of intensive, long-term antibiotic regimes [93]. Indeed, it has been reported that established mucoid biofilms encompass anywhere between 5 and 32 times the amount of polysaccharide material compared to other ECM components such as protein or nucleic acids [104, 105]. Immediately, focusing on the polysaccharide simplifies the computational model, as an atomistic model of the EPS possesses a high degree of representability when describing the molecular composition of mucoid biofilms. In this work, the development of the first atomistic model of a mucoid *P. aeruginosa* biofilm (EPS) matrix was achieved using a *ground-up* approach. This involved defining smaller EPS building blocks, from which the

larger atomistic model was developed.

It is important to recognise here the value of the existing research and data when constructing the atomistic EPS model. Without acetyl abundances and distributions determined through NMR spectroscopy [111, 112], the alginate-cation chelation geometries determined experimentally [122, 123, 130] and theoretically [134, 135, 137–139, 142, 143, 295, 300], the mechanism of alginate chain-chain association determined theoretically and through fibre X-ray diffraction and SAXS measurements [137, 300, 310], and finally the bulk EPS morphologies visualised using advanced imaging techniques [102, 175, 308, 309], it would not have been possible to construct an atomistic model that captures as much *in vivo* representability. How these works, listed above, have offered critical insight pertinent to model design will be discussed in more detail throughout the chapter.

8.1.2.2 EPS backbone

The EPS is bacterial alginate, which effectively is a structurally modified form of algal alginate. As such, to begin defining an atomistic model of the EPS that exists within the CF lung, an algal alginate template was required, which could then undergo acetylation to form the EPS. The ^{1}H-NMR and chromatography characterisation works performed by SkjÅk-Bræk *et al.,* and Davidson *et al.,* were critical in defining structurally representative algal alginate templates [110–112]. Their characterisations elucidated that the mucoid *P. aeruginosa* EPS can be distinguished from classical (algal) alginate with unique *in vivo* structural motifs. The first motif is the apparent lack of homopolymeric guluronate blocks (G-blocks), meaning the EPS is solely comprised of homopolymeric mannuronate blocks (M-blocks) or heteropolymeric mannuornate-guluronate blocks (M-G blocks) [111, 112]. The second distinguishing motif is that 50% of the mannuronate (M) units are acetylated at the C2 or C3 carbon positions [111, 112] and this acetylation is exclusive to mannuronate units only [110]. In light of these observations, the atomistic model development began with the definition of two molecular templates: a poly-β-D-mannuronic acid quadramer and a copolymeric β-D-mannuronic-α-L-guluronic acid quadramer. The former encompasses a single chain of four M units linked via 1–4 glycosidic bonds and the latter a single chain of two M units and two G units linked via 1–4 glycosidic bonds in an alternating M-G pattern. These two EPS templates capture the correct uronate distributions observed *in vivo* [111, 112].

What followed was the determination of the most thermodynamically stable acetylation configurations given the aforementioned experimentally derived acetyl % abundances and localisation. The thermodynamic stability of all possible acetylation patterns, at all M-unit hydroxyl sites at the C2 or C3 positions, was assessed. The results from the thermodynamic stability analysis highlighted that stable acetylation configurations suggested a limit for

proximity. In other words, acetylation patterns that incurred the largest and smallest acetyl-acetyl separation returned the least stable structures. In fact, it is most stable to acetylate within an EPS region that permits a balancing of intra-acetyl steric repulsion with intra-acetyl hydrophobic interactions. The most thermodynamically stable acetylated EPS configurations correspond to a M1(C3) M3(C2) acetylated poly-β-D-mannuronic acid quadramer (PolyM) and an M2(C3) acetylated copolymeric β-D-mannuronate-α-L-guluronic acid quadramer (PolyMG).

8.1.2.3 Cation-induced cross-linking

The ICP-MS and spectrophotometric characterisations of CF sputum, outlined in the works of Smith *et al.,* and Grandjean Lapierre *et al.,* identified the presence of an elevated cationic atmosphere in the CF lung [124, 125], which raised the question about "how" the mucoid EPS would respond to this environment and whether its response was critical for virulence. Relative to non-CF controls, sodium (Na^+), calcium (Ca^{2+}) and magnesium (Mg^{2+}) ions are most elevated in concentration [124, 125]. The occurrence of these ions in the CF lung is positively correlated with the occurrence of interleukin (IL)-8, a chemoattractant cytokine responsible for neutrophil recruitment into the CF airway. This supports the hypothesis that these elevated ion levels, which resemble the concentrations found in whole blood, arise from cell necrosis during local tissue damage and bleeding [125].

Rheological measurements performed by Wloka *et al.* and Körstgens *et al.* identified that exposure of the EPS to calcium increases the mechanical stability of the EPS *in vitro* through ionic cross-linking and makes the EPS harder to clear [122, 123, 132]. It is understandable, therefore, that these ions should possess a similar chemical influence over the EPS scaffold *in vivo*, making them an environmental condition that had to be accommodated in the EPS atomistic model(s). Despite the aforementioned rheological properties, the effects these ions have on the structure and morphology of the mucoid EPS scaffold was not definitively understood. Intriguingly, it was also unknown which of the ions elevated in the CF lung was most responsible for driving stable EPS aggregation, contributing most to biofilm matrix stabilisation and, thus, chronic infection in the CF lung.

Plazinski *et al.* showed using theoretical modelling based on molecular dynamics that the association of two alginate chains, in parallel or antiparallel chain arrangements, bound through hydrogen bonds alone proceeds the cation cross-linking event [137]. Indeed, the rheological measurements made in the work of Wloka *et al.* also underscore that only hydrogen bonds and entanglements stabilise extended EPS networks in the absence of cations [123]. As such, to mimic this association in our EPS model systems, in the absence of cations, the thermodynamic stability of two PolyM and PolyMG chains was assessed when associated

either parallel or antiparallel. In support of the observations of both Plazinski *et al.* [137] and Wloka *et al.* [123], it was seen that hydrogen bonding interactions are the primary stabilising interaction between neighbouring EPS chains in the absence of cations. These hydrogen bonding interactions stabilise the PolyM system in an antiparallel arrangement ($PolyM_{ap}$) and the PolyMG system in a parallel arrangement ($PolyMG_p$). M-G junctions offer more suitably oriented oxygen functional groups and establish a larger hydrogen bonding network relative to M-M junctions and, as such, the $PolyMG_p$ system is stabilised more so than the $PolyM_{ap}$ system.

When cross-linking about the cations elevated within CF sputum, as these are the ions most likely to have the largest cross-linking influence, it was shown that the ionic cross-links stabilise EPS architecture more so than hydrogen bonds alone. This was in agreement with aforementioned works of Plazinski *et al.*and Wloka *et al.*, as well as the NMR characterisations of Mayer *et al.* and Lattner *et al.*, which outlined that upon exposure to cations, the ionic cross-linking between neighbouring EPS chains becomes the dominant stabilising interaction; the stabilising influence of hydrogen bonds becomes insignificant [121–123, 130, 137]. Calcium ions are the more potent ionic cross-linkers, forming more stable 2-chain complexes by approximately 9 eV and 5 eV compared to sodium and magnesium ions respectively (**Fig 3.12 - 3.14**). Encouragingly, this is in agreement with the theoretical predictions of Stewart *et al.* and Agulhon *et al.*, along with with the experimental observations of Kaklamani *et al.* and Wloka *et al.*, who all showed the very weak cross-linking ability of sodium and magnesium ions relative to calcium ions [122, 138–140, 145]. Not only do these large predicted stability differences show, unequivocally, how calcium and magnesium ions impact the stability of the EPS structure relative to the biologically ubiquitous sodium ion, but also demonstrate that calcium is the sputum ion most implicated in thermodynamically stable matrix aggregation and stabilisation.

However, it is important to note here that our model predictions differ from the DFT-based model predictions of Bekri *et al.*, which suggest magnesium ions form the most stable complexes within M-M and M-G junctions of algal alginate [143]. By using the chemical hardness (η) as an indicator of cation complex stability, magnesium ions were demonstrated to complex within M-M and M-G junctions with larger values of η relative of calcium and sodium ions, indicating that the resulting magnesium complexes are most stable [143]. A possible reason for this calculated difference is that our EPS models are larger in scale compared to the M-M and M-G disaccharides employed by Bekri and colleagues [143], meaning they are more capable of capturing the influence sterics play in stable cation-induced EPS cross-linking event. The smaller ionic radius of magnesium reduces the number of accessible oxygen atoms, lowering the cation coordination number and ionic cross-linking

density and, therefore, the overall stability of the resulting complex. This is a structural effect captured only in multi-EPS chain models.

Sodium ions fail to aggregate mannuronate rich fractions of the EPS as stoichiometric (charge neutralising) amounts of sodium complex in-between EPS chains at ill-defined inter-chain binding sites which, in some regions of the EPS, fail to establish any cross-links at all. These binding sites give rise to a weak, transient, ionic architecture that leads to unstable associations when complexed in-between PolyM$_{ap}$ EPS chains and weakly stable associations when complexed in-between PolyMG$_{p}$ EPS chains. This same mechanistic principle behind the failure of sodium to aggregate algal alginate chains was also observed in the theoretical models of Stewart and colleagues, which showed that the binding site adopted by sodium in the G-G junctions of algal alginate decamers was too far separated from the neighbouring alginate chain to facilitate cross-linking [138, 139]. Calcium and magnesium ions display better defined inter-chain binding sites, as these ions are able to more greatly influence the conformation, more specifically the torsion profile, of the uronate backbone, in turn, influencing the orientation of oxygen functional groups to face the cation centre. This, coupled with the ability of these ions to form stronger ionic bonds with the EPS oxygen donors, explains the greater complex stability when compared to sodium ions.

Calcium ions, out of all sputum ions tested in this research, form the most thermodynamically stable, ionically cross-linked 2-chain EPS matrices. As a consequence, they can be assumed to out-compete other sputum ions for EPS binding *in vivo*. In addition, substitution of the native calcium ions from their cross-linking geometries within the 2-chain EPS scaffold using exogenous sodium was thermodynamically unfavourable. This indicates that the calcium cross-linked 2-chain EPS matrices would be resistant to physiological softening, supportive of the rheological observation that the storage and loss moduli of calcium cross-linked alginate chains are stable over time and, therefore, that the calcium cross-links established between neighbouring M-M, M-G and G-G junctions are permanent *in vivo* [122, 228].

The cation binding predictions generated by the models constructed in this work are the first to offer mechanistic insight into the stabilisation of bacterial alginate (EPS) pertinent to the CF lung in particular. Combined with the works of others detailed above, our models have shown that the influence calcium ions have over the stability of algal alginates can also be extended to bacterial alginates (the EPS). But our models help to go one step further, offering nuanced insight into the role sterics, in addition to electrostatics, contribute to the cation cross-linking event and suggest that calcium ions are critical for virulence. Calcium ions can cross-link the EPS, stabilising the matrix scaffold more so than other ions elevated in concentration in the CF lung, and are the ions most implicated in the structural rigidity

of EPS matrices within the CF lung. This corroborates the work of Sarkisova *et al.*, in which it was shown that calcium ions can cross-link the mucoid EPS, and this structurally stable ionic network was better able to retain the EPS as a gel within the biofilm matrix [133]. In turn, this leads to the development of larger 3D biofilm matrices *in vitro* [133]. Therefore, the cation cross-linked PolyM$_{ap}$ and PolyMG$_{p}$ 2-chain EPS models, identify calcium as key ions for mucoid *P. aeruginosa* chronicity and virulence within the CF lung.

In light of this, calcium ions pose as a key target for EPS matrix eradication. Removing the calcium cross-links could destabilise EPS architecture giving rise to biofilm matrices that are easier to clear. In fact, recently this has prompted research into OligoG CF-5/20 therapy, which is an guluronate rich alginate oligomer targeted at disrupting the calcium cross-links within mucoid *P. aeruginosa* biofilm matrices [183]. Powell *et al.* demonstrated through molecular dynamics simulations and CLSM imaging, that when OligoG CF-5/20 sequesters calcium ions, which prevents the formation of calcium cross-links between ECM components, the *P. aeruginosa* biofilm matrix which forms has a marked reduction in biofilm biomass [183]. However, it must be mentioned here that these works make the assumption that eDNA is the primary matrix component of *P. aeruginosa* biofilms [183], not the EPS as is assumed in this research. As such, it is possible that the matrix disruption mechanism proposed by Powell *et al.* is not explicitly representative of the EPS also. Indeed, this research has investigated the impact that OligoG CF-5/20 has on the calcium cross-linked EPS morphology. This is discussed in **Section 8.2.2**.

8.1.3 Improving the physiological representability of the EPS model

The 2-chain cation cross-linked PolyM$_{ap}$ and PolyMG$_{p}$ models offered critical insights into the atomistic mechanism governing cation induced aggregation and stability, definitively showing how the calcium component, more so than the other ionic components, of the elevated cationic atmosphere contributes to chronic mucoid *P. aeruginosa* biofilm infection. These systems, however, are not of sufficient size to accurately characterise *bulk* matrix morphology. Explicitly, the amphiphilic carbon quantum dot visualisations of the mucoid *P. aeruginosa* EPS performed by Ritenberg *at al.* have shown the EPS to possess a branched and dendritic macrostructure [308], which is a particular morphology absent at the 2-chain level. To this end, the calcium cross-linked 2-chain PolyM$_{ap}$ and PolyM$_{p}$ systems (Ca-PolyM$_{ap}$ and Ca-PolyMG$_{p}$ respectively) were extended to a stable 4-chain system using finite temperature, explicit solvent, molecular dynamics. This gave an EPS model more representative of the EPS which occurs under physiological conditions.

Extension from 2-chains to 4-chains enables the identification of key geometrical features that most stably facilitate exothermic multi-EPS chain association. Accordingly, it is important

to recognise the SAXS and fibre X-ray diffraction works of Agulhon *at al.* and Sikorski *et al.*, which demonstrate that multi-algal alginate chain association (> 2-chain association) about calcium ions proceeds via fibrillar stacking [300, 310]. Specifically, calcium alginate gels present with a high surface area and a low small angle X-ray scattering exponent indicative of a fine fibrillar morphology [300] and the X-ray diffraction pattern suggest multi-chain packing about a central columnar distribution of calcium ions [310]. These results offered suitable detail when scaling to the 4-chain models, with a fibrillar stacking regime to form a 4-chain system in which four EPS chains are associated about a central columnar distribution of calcium.

Under a fibrillar stacking regime, the Ca-PolyM$_{ap}$ and Ca-PolyMG$_{p}$ systems were stacked on top of one another, in an array of unique stacking geometries, to assess the most thermodynamically stable association geometry between two Ca-PolyM$_{ap}$ and two Ca-PolyMG$_{p}$ systems to form a 4-PolyM or 4-PolyMG system. It was most thermodynamically stable to associate two Ca-PolyM$_{ap}$ and two Ca-PolyMG$_{p}$ systems when a dense ionic intracomplex space was established, further accentuating the role that electrostatics play in the overall stability of matrix architecture. Within the most stable 4-chain PolyM and PolyMG stacking arrangements, labelled as 4-PolyM* and 4-PolyMG* respectively (**Fig 4.5** and **4.4**), the calcium ions are tightly bound within the 4-chain network. The matrix-bound calcium ions established a large ionic intracomplex space, bound to four uronate residues per calcium ion, within CN=5, 6 and 7 environments.

It was encouraging to observe that this stacking regime returned calcium complexed multi-EPS chain geometries that were thermodynamically stable and consistent with the geometries observed through SAXS and fibre X-ray diffraction characterisation [300, 310]. Furthermore, within these stable structures, the number of ionic cross-links, and not the number of hydrogen bonds, correlated with overall complex stability. This suggested that hydrogen bonding interactions do not govern the geometry of multi-EPS chains associated about calcium ions. Indeed, this also complements the recent theoretical modelling performed by Hecht *at al.*, in which the morphologies of multi-algal alginate chains 30 monomers in length cross-linked about calcium were predicted using implicit solvent molecular dynamics [312]. In this work, it was shown that the association of twenty algal alginate chains was facilitated by calcium ions alone, which bind within the M-G and M-M junctions. This facilitated a "zipping" mechanism, cross-linking a single chain end first, then propagating to sequentially cross-link the remaining M-G and M-M junctions until the other chain end is reached [312]. Furthermore, Braccini *et al.* performed energy partitioning analysis to identify which inter-molecular interactions most govern the stability of calcium cross-linked algal alginate chains and found that cation-mediated electrostatic interactions stabilised the resultant

complexes, with no contributions from hydrogen bonding [358]. Overall, these works, combined with our predicted stable stacking geometries, strongly suggest that the calcium cross-links govern the stability and resultant morphology of EPS architecture.

Interestingly, inter-penetrating acetyl groups prevent 4-PolyMG* complexation about calcium but contribute to the calcium chelate geometry in 4-PolyM* complexation. This suggests that acetyl contribution to the calcium chelate geometry can, effectively, distinguish exothermic ionic association of acetylated mannuronate EPS fractions from acetylated mannuronate-guluronate EPS fractions. But, more compellingly, the inter-penetrating acetyl groups, that lead to a lack of ionic association within mannuronate-guluronate EPS fractions, create a void space that water molecules could potentially occupy. We suggest that acetylated mannuronate-guluronate EPS fractions are a possible origin for water channels in the mature mucoid biofilm matrix.

The stability of the 4-PolyMG* structure exceeded the 4-PolyM* structure, which can be attributed to the larger intracomplex space established within the former. Therefore, the 4-PolyMG* structure was taken forward and transformed into an atomistic model more representative of the physiological morphology through an explicit solvent molecular dynamics trajectory, equilibrating the structure at 310 K (**Fig 4.9**). This is a similar procedure to the simulation protocol employed by Hecht *et al.* when studying the bulk morphology of twenty calcium cross-linked algal alginate chains [312], the only difference being the use of an implicit solvent in the work of Hecht and colleagues, whereas an explicit water solvent was used here. Implicit solvents in atomistic modelling are designed to represent the continuous dielectric medium imposed by water without the need to include any water molecules in the simulation. This helps to minimise the total atom number and improve the computational tractability of the simulation. The choice behind the use of explicit, rather than implicit, solvent in this work ensured the EPS morphology following the dynamical simulation was representative of an EPS that experiences the "flow" of water, effectively the translation of water molecules across the EPS interface, as is evidently present *in vivo* [169]. The pulsed field gradient NMR investigation conducted by Vogt *et al.*, has identified that the vast majority of biofilm bulk water is confined to, and flows through, channels within the EPS scaffold [170]. In fact, when water flows through polysaccharide gels, the molecules do not have their motions significantly reduced, resulting in a continuous flow of water, and a lack of bound water, at the polysaccharide interface [292]. This observation alludes to the geometric principle that a boundary between the EPS and bulk water exists, and water flows across the EPS interface. Using an explicit solvent in the dynamical simulations, representing the flow of the surrounding water and the EPS-water collisions that unavoidably follow, was therefore seen as an essential simulation requirement.

Upon the physiological equilibration of the 4-PolyMG* structure, using explicit solvent molecular dynamics, there is a large reduction in the diversity of the calcium chelation geometries, with COO^- groups dominating the calcium coordination environment over all EPS oxygen functionality. Hence, under physiological conditions, the COO^- offers the dominant mode of aggregation of EPS chains about calcium. The reduction in the diversity of the calcium chelation geometries reduces the number of 4-chain ionic cross-links, effectively reducing the density of the intracomplex space. Intriguingly, the acetyl groups remain oppositely displaced, preferring to face the solvent rather than face the neighbouring EPS chains. Taken collectively, both these geometrical features, that occur under physiological conditions, influence the adoption of a discontinuous, V-shaped, dendritic morphology in the bulk. This is a critical observation, as this model has, for the first time, provided an atomistic explanation for the origin of such morphologies observed in TEM and amphiphilic carbon quantum-dot visualisations of *P. aeruginosa* EPS scaffolds [308, 309]. Therefore, despite the limitations imposed on the model, explicitly the construction of a model that encompasses the EPS and calcium alone, at a 4-chain (400-atom) scale, a model of this size and composition is able to replicate key aspects of the *bulk* morphology observed *in vitro*. This increases confidence in, not only the model's representability, but also the value of the model predictions.

8.1.4 Model applications

This model development pipeline has permitted the construction of two primary mucoid EPS atomistic models. The first is a calcium cross-linked 2-chain EPS model(s), which is well defined to study cation exchange processes within the mature EPS architecture. This model was deployed to investigate whether exogenous gallium ions could be accommodated within the mature EPS scaffold, in preference to the native calcium cross-linking ions. This allowed an investigation into whether gallium ions could initiate the liberation of calcium from its well-defined inter-chain binding geometry, which would potentially result in biofilm matrices that were more susceptible to physiological softening, reducing biofilm proliferation.

The second model pertains to the physiological 4-chain EPS model, obtained following physiological equilibration of the 4-PolyMG* system, which captured key structural aspects of *in vitro* bulk atomistic morphology [308, 309]. As such, this particular model was a suitable atomistic model to deploy in simulations to probe the structural-chemical origins that govern how quorum sensing autoinducer molecules move throughout the EPS scaffold between bacteria. In addition, the physiological 4-chain EPS model was also used to elucidate, for the first time, the therapeutic mechanism of OligoG CF-5/20, at the atomic-scale, against the mature mucoid EPS scaffold.

8.2 Atomistic model predictions

8.2.1 Biophysical relationships between EPS structure and virulence

Calcium-mediated electrostatic interactions are clearly implicated in mucoid EPS matrix stabilisation. At a 2-chain and 4-chain EPS scale, the density of the ionic interactions between chains correlates well with the thermodynamic stability. The conformational flexibility of the EPS uronate backbone in the presence of calcium ions creates well defined, permanent, inter-chain binding sites and calcium ions are, therefore, potentially playing a role in matrix chronicity by virtue of it being an electrostatically stabilised matrix that will be resistant to perturbation. Clinically, this translates into the emergence of a mechanically stable matrix [122, 123], which is harder to clear and penetrate with antimicrobials [132, 133].

At a 4-chain scale and in response to physiological conditions, the mucoid EPS adopts a dendritic, V-shaped, morphology and this organisation arises as a response to the reduction in the calcium chelation geometries that again accentuates calcium's role in governing EPS macrostructure. The predicted V-shaped morphology suggests that the mucoid *P. aeruginosa* biofilm (EPS) matrix is structurally heterogeneous encompassing open clefts and cavity spaces. This offers another explanation for the origin of water channels throughout the biofilm matrix, which permits the transport of nutrients to bacterial cells buried deep beneath the EPS [169]. Also, these open clefts give rise to regions of the EPS with an increased binding surface area which may serve to maximise structure-directing interactions with invading antimicrobials, effectively sequestering these molecules and preventing them from reaching the biofilm-enveloped bacteria.

8.2.2 Weakening mucoid EPS architecture

Calcium ions are critical for facilitating stable EPS aggregation, therefore, removal of these ions offers a viable strategy for weakening, and potentially eradicating, mature EPS architecture. OligoG CF-5/20 is an emerging anti-pseudomonal therapeutic targeted at mature mucoid *P. aeruginosa* biofilm matrix disruption. The mechanism of action against non-mucoid biofilms has been probed both theoretically, using molecular dynamics simulations, and experimentally, using CLSM, AFM, FTIR and circular dichroism and suggest that biofilm eradication relies on calcium chelation chemistry [182–184, 191]. However, the mechanism of action against mucoid biofilm EPS matrices is currently unknown, but is assumed to be the same, specifically, removing calcium from the biofilm matrix and weakening the stability of the matrix [192, 339].

For the first time, the key structural chemistry and physico-chemical interactions that facilitate, at least in part, the mechanism of action of OligoG CF-5/20 against the mucoid EPS have been determined. As with non-mucoid biofilm matrices, calcium chelation is

critical for exothermic OligoG CF-5/20 dispersal into the mucoid EPS matrix. Calcium ions offer tethering points to the EPS surface for the invading OligoG CF-5/20 molecule, which establishes COO^{-}-Ca^{2+} and OH-Ca^{2+} interactions. No other form of interaction, covalent or non-covalent, with the EPS chains is established and all interactions are facilitated by the native EPS calcium ions. These modes of interaction are not size dependent, being observed at smaller and larger OligoG CF-5/20 length scales (**Fig 5.5**). Although calcium ions are an OligoG CF-5/20 binding facilitator, they are not liberated from their native EPS binding sites upon OligoG CF-5/20 binding. This suggests that the OligoG CF-5/20 mechanism of action against the mature mucoid EPS does not involve abstraction of calcium ions from the EPS scaffold. In fact, the OligoG CF-5/20 stimulates a contraction in the EPS V-shaped motif, capturing the OligoG CF-5/20 within the EPS scaffold.

The clinical significance of this, is that the OligoG CF-5/20 molecule is not causing biofilm weakening through calcium abstraction, but rather through preventing further calcium induced EPS aggregation. Encouragingly, this is highly complementary to the mechanism of action against eDNA matrices proposed by Powell *at al.*, where molecular dynamics, FTIR and CLSM experiments suggested that sequestration of free calcium ions from the extracellular medium was primarily responsible for the eradication of mature non-mucoid biofilm matrices [183]. It has already been highlighted in depth, throughout this research, how calcium induced EPS aggregation contributes to increased matrix stability. Understandably, preventing this chemical change would serve to weaken mature EPS architecture. Therefore, this research offers a change in perception when it comes to mucoid biofilm matrix weakening, rather than being a process reliant upon calcium abstraction, EPS scaffold weakening is a process that is achieved through preventing multi-EPS chain association, as well as sequestering free calcium from the elevated cationic atmosphere. A schematic of this is given in **Fig 8.1**.

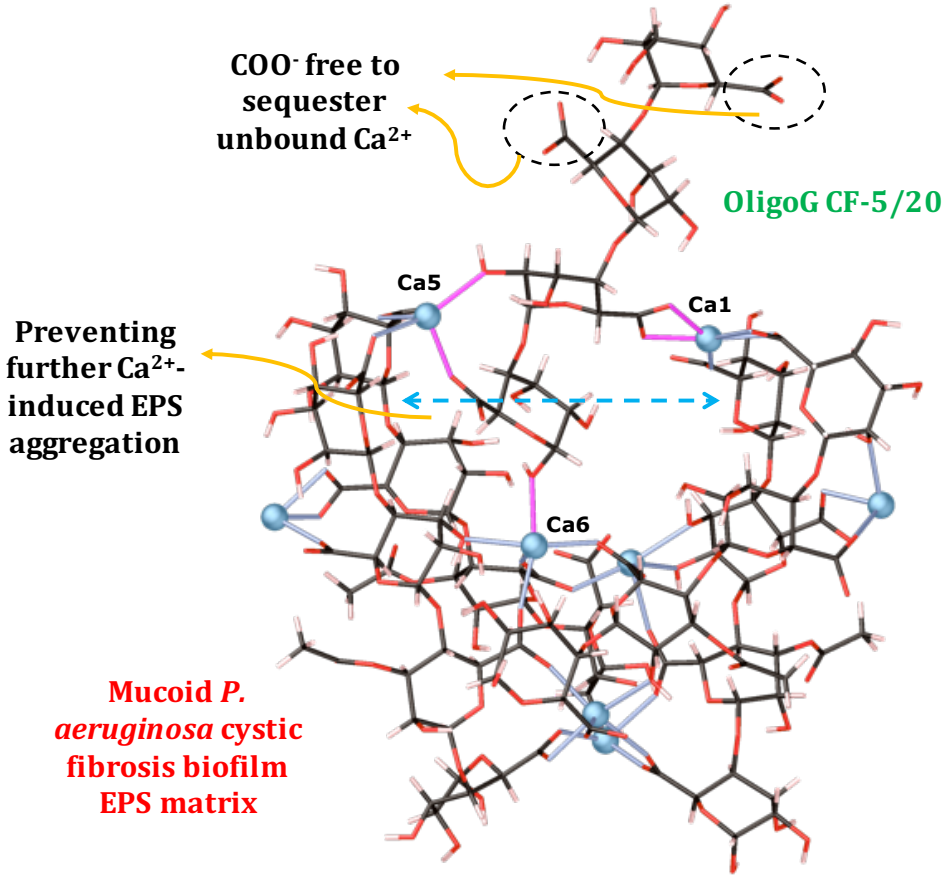


Figure 8.1: A schematic detailing the potential mechanism of OligoG CF-5/20 against the mucoid *P. aeruginosa* EPS.

8.2.3 Interactions that govern motion through the mucoid EPS

Understanding the molecular functionality that governs how molecules move through the mucoid EPS scaffold architecture is a question with obvious connections to pharmaceutical design. If a particular functionality could be identified, then it could supplement any novel small antimicrobial chemotherapy, enabling more effective deep matrix penetration.

To study molecular motion through the EPS, it was considered appropriate to interrogate how quorum sensing autoinducer (QSAI) molecules pass between bacteria throughout the EPS scaffold. These are small signalling molecules that pass between bacteria, allow the bacteria to communicate and regulate population-wide genetic expression. Mucoid *P. aeruginosa* utilises QSAI's to regulate the progression from planktonic individuals, to initial biofilm forming communities [146, 148], to mature biofilm architectures [162, 163]. In light of this, one can limit the investigation to QSAI molecules that are definitively implicated in mid-to-late stage biofilm proliferation, as it is these molecules which will have to disperse throughout the EPS between bacteria. The mucoid *P. aeruginosa* QSAI signals that are present within CF lung

infections and fit this brief are C_4-HSL and PQS [160, 166, 167].

The physiological (4-chain) EPS model was deployed in molecular dynamics and DFT simulations to sample the modes of interaction that occur between the EPS and the QSAI molecules, C_4-HSL and PQS, under physiological conditions (**Fig 4.10** and **4.11**). C_4-HSL is a molecule that exploits solely vdW interactions with the EPS scaffold. These interactions stem from the reluctance of C_4-HSL to undergo changes in its electron density, as inferred from this molecule's *chemically hard* reactivity index. Consequently, the C_4-HSL molecule is unaffected by the cationic charge distribution within the EPS scaffold, and binds solely through vdW interactions which are not sufficient to render the molecule immobile at the EPS surface and allows the molecule to migrate into water channels within the EPS scaffold. Taken collectively, this offers a thermodynamic rationale as to why this particular QSAI is non-local and can be received community-wide.

The PQS molecule, however, is the opposite. This molecule should be able to facilitate irreversible electrostatic interactions with the EPS scaffold. The PQS molecule has the ability to form thermodynamically stable ionic complexes with EPS-bound Ca^{2+} as well as establishing a hydrogen bond between the OH group and the COO^- group on a single EPS chain. In fact, it is the re-arrangement in this hydrogen bonding interaction that facilitates irreversible binding over long time-scales. Conversely, the PQS molecule is rendered immobile through EPS binding and, due to its larger hydrophobic nature, shows no proclivity to migrate into biofilm water channels. The reduced HOMO-LUMO separation present within this molecule increases the *chemical softness*, heightens the polarizability and makes this molecule more susceptible to changes in its electron density and thus, charge transfer. As a consequence, the PQS molecule has a higher ionic binding potency. Despite forming direct electrostatic interactions with the EPS, the PQS molecule causes no active disruption to the native ionic architecture and, along-side C_4-HSL, has relatively benign effects on overall matrix morphology. The irreversible, electrostatic, *chemisorption* of the PQS molecule onto the EPS scaffold means this molecule is predicted to be a local QSAI signal, not received community-wide. In fact, this corroborates and, for the first time, offers an atomistic explanation behind the confocal Raman and secondary ion mass spectrometry studies that show that PQS localizes close to its source and is not a QSAI received community-wide [328, 329, 331].

Overall, this further calls attention to the role electrostatics play in biofilm structural chemistry. Not only are these interactions critical for matrix stabilisation, but they hinder motion throughout the EPS scaffold. Naturally, what follows as a wider generalisation for the *de novo* design of novel antimicrobials specialised at deep matrix penetration, it

may be more beneficial to design molecules that possess similar chemical characteristics to C_4-HSL. Explicitly, this could relate to molecules that are weakly hydrophobic, specifically encompassing Log(P) values ≤ 1.47, and that exploit solely vdW interactions when binding to the EPS which, potentially, is a property dictated by the HOMO-LUMO separation.

Finally, the proclivity of the C_4-HSL molecule to migrate into the solvent space was further exemplified through *caging* this molecule in-between two physiological 4-chain EPS models. Again, this molecule exploits solely vdW interactions with the EPS and, within 20 ns, manages to breach the EPS scaffold successfully migrating into the solvent space. Establishing solely vdW interactions leaves the C_4-HSL hydrogen bonding donor/acceptor groups free to form hydrogen bonding interactions with the water solvent. It is thermodynamically stable for hydrogen bonding interactions to be established at these groups and this, along-side its hydrophilic character, contributes to the solvent proclivity. From a clinical standpoint, this would suggest that quorum quenching (QQ) agents, molecules that degrade QSAI signals, maybe less effective against C_4-HSL. This is because in a recent *in vitro Escherichia coli* biofilm flow reactor study, it was shown that QQ agents fail to degrade QSAI molecules freely diffusing through an aqueous phase [316]. Therefore, given the immobility of PQS relative to C_4-HSL, this work proposes PQS as a more realistic and viable target for QQ.

8.2.4 Identifying novel gallium targets and delivery vectors

Exogenous gallium therapy is a strategy used to combat *P. aeruginosa* biofilm matrix proliferation by competitively inhibiting ferric iron (Fe^{3+}) binding to the the ferric iron uptake proteins, pyochelin and pyoverdine [218–222]. However, the anti-biofilm effects, induced through exogenous gallium exposure, remain active against *P. aeruginosa* mutants deficient in the siderophores pyochelin and pyoverdine [220, 225]. This raised the question of whether gallium is having a direct effect on the EPS architecture, for example, by displacing the calcium cross-links, and it is this that is weakening the biofilm structure leading to reduced proliferation in siderophore-deficient cells. Mechanistically, if gallium ions could be accommodated by the EPS, in preference to the native calcium ions, then potentially this would create a scaffold morphology more susceptible to physiological softening through sodium-ion exchange. This could serve to reduce matrix proliferation.

Calcium-for-gallium cation exchange was introduced into the calcium cross-linked 2-chain EPS models (**Fig 6.5** and **6.6**). Upon substitution for the native calcium ions, the alien gallium ions complex with larger degrees of covalency, driven through the large degree of charge transfer (large degree of orbital overlap) between the COO^- $2p$ orbitals and the vacant Ga^{3+} $4s$ orbital. The increased covalent character of the gallium chelate complexes, relative to the native calcium chelate complexes, is a key structural factor that allows trivalent gallium

ion complexation by the EPS to be distinguished from divalent sputum ion complexation. The increased covalency in the chelate geometries infers more stable interactions, however, it is thermodynamically unstable to complex gallium in place of the native calcium ions.

There is a large array of unique structural chemistry that contributes to the instability of the cation exchange. Firstly, the 2-chain EPS scaffold provides suitably accessible acidic (anionic) functionality to bind divalent sputum ions, but such functional group orientation is not suitable to bind the trivalent gallium ion. The +3 gallium ion is, effectively, complexing within a −2 binding pocket which does not satisfy the requisite degree of anionic functionality in the Ga^{3+} coordination sphere, failing to facilitate thermodynamically stable chelate complex formation. Additionally, calcium-for-gallium substitutions into high affinity calcium chelate pockets gave rise to a +3 charge localisation on one side of the EPS scaffold. Despite retaining charge neutrality over the molecular frame, the +3 charge localisation on one side of the EPS scaffold leads to an imbalanced charge separation and a sharp cationic density gradient, which injects further instability into the system (**Fig 6.7**).

The gallium ions introduce large, unstable, conformational rearrangements into the EPS uronate backbone. Namely, driven by the drop in CN relative to the native calcium ions, gallium ions induce a $\Delta\psi = +45°$ relapse in the glycosidic orientation to its non-cation-bound orientation away from the molecular frame (**Fig 6.10**), inducing large torsion rearrangements in the EPS uronate backbone. Recall, with divalent ion complexation, large torsion rearrangements serve to destabilise the system and this observation extends to trivalent gallium ions. Furthermore, gallium is able to induce mannuronate pseudorotation, that is the inter-conversion from the energetically stable chair ($^4C_1$) to the less energetically stable twisted boat conformation (**Fig 6.11**). This particular conformational change, at the quadramer scale, increases the instability of the system by 4.58 eV.

When performing calcium-for-gallium substitutions into the EPS scaffold, a clear trend emerges: the fewer calcium ions removed from the scaffold, the more favourable the substitution. This, therefore, suggests that the largest barrier to calcium-for-gallium substitution is the enthalpic barrier associated with calcium removal from the EPS. The thermodynamical predictions suggest that it is more stable for gallium to exist as an unbound ion in the EPS vicinity.

Exogenous gallium will remain free and, therefore, available for competition into ferric uptake systems. PQS is a QSAI molecule that can facilitate ferric binding [172–174] and uptake, through binding TseF - a substrate secreted by the H3-T6SS [229]. This process assists in cellular Fe^{3+} uptake in the absence of siderophores and iron-for-gallium substitution

into ferric-PQS complexes would offer a plausible explanation for the ability of gallium ions to retain anti-biofilm activity in siderophore-deficient mutants. Gallium is able to complex free PQS ligands, as well as substitute the native ferric centre out of the PQS chelate pocket (**Fig 7.6**) in a thermodynamically stable fashion. In fact, the favourability of the iron-for-gallium substitution into the native ferric-PQS complex is highly similar to the favourability of the iron-for-gallium substitution into the native ferric-pyochelin complex. From which, it can be inferred that these processes are equally feasible *in vivo*. The complex geometry formed upon gallium-PQS binding shows a very high degree of structural similarity with the native ferric-PQS complex. This suggests that the gallium-PQS complex could, potentially, undergo adduct formation with TseF in a similar fashion to the native ferric-PQS complex, in turn, offering a route into *P. aeruginosa* cells that circumvents siderophore uptake completely.

In the absence of all other cations, gallium ions can complex into vacant -3 binding pockets created between three OligoG CF-5/20 scaffolds. A gallium-OligoG CF-5/20 (gallium tri-polyguluronate) scaffold was designed with the aim of exploiting a bi-functional therapeutic mechanism: the binding of OligoG CF-5/20 to native EPS calcium, causing EPS scaffold contraction and inhibition of further calcium mediated EPS aggregation, coupled with the release of gallium ions which would be free to compete for ferric uptake systems. Proof-of-concept calculations, explicitly gallium-for-calcium substitutions into the gallium-OligoG CF-5/20 delivery vessel, were highly exothermic (**Fig 7.7**) and, encouragingly, far more exothermic relative to release from the nitrate and maltolate gallium delivery scaffolds deployed currently [356, 357]. Clinically, this far greater exothermic release could translate into more favourable release kinetics *in vivo* and, in addition, utilising OligoG CF-5/20 as the delivery vector could give rise to precise targeting of the mature mucoid *P. aeruginosa* EPS.

Aside from OligoG CF-5/20, there have been recent research efforts to design a gallium delivery system based on the natural polysaccharides, such as alginate, carboxymethyl cellulose and pectin. Gallium ions are able to homogeneously distribute across these scaffolds and their release is prompted by alginate lyase digestion, hydrolysis and cation exchange [222, 227, 344]. These aforementioned delivery vessels have been shown to inhibit *P. aeruginosa* biofilm growth activity *in vitro* [222, 344], possessing an MBC of < 5 mg ml [344] with the ability to induce a 1.5-log reduction in biofilm CFU over 48 hours [222]. These results show promise behind the use of novel gallium delivery vessels to treat *P. aeruginosa* infections.

8.3 Critical analysis and future work

8.3.1 Quantum mechanical tractability

The aim of this research was to use atomistic modelling to predict EPS molecular structure and bonding to implicate key molecular functionality and site-directed interactions in pathogenicity and virulence. Accurate structure and bonding calculation, in addition to thermodynamic stability analysis, requires quantum mechanical computation. However, this comes at the cost of deploying reduced scale models. Small scale models are pivotal in understanding atomic fine structure and inter-atomic interactions, but can fail to explain how such interactions translate, at a bulk scale, to the ultrastructure and morphologies that can be observed in microscopy images or even by eye.

This can be easily illustrated by contrasting the general appearance of mucoid *P. aeruginosa* biofilm matrices to the dendritic, V-shaped, EPS morphology predicted in this research. *P. aeruginosa* biofilm colonies construct a 3D structurally heterogeneous biofilm ECM with a floccular, mushroom-like, morphology [43], which has been observed extensively through microscopy [52, 93, 102, 175, 201]. As the models constructed in this research are only capable of making fine structure predictions, i.e. the dendritic V-shaped motifs observed at a 4-chain EPS quadramer scale, the models are not capable of identifying how these V-shaped architectures combine (arrange) to form the ultrastructure morphology observed microscopically. Furthermore, the aperture created at the top of the V-shaped motif is 10.51 Å in diameter - far shorter than the size of an individual *P. aeruginosa* cell for example, which measures approximately 0.5 μm by 1.5 μm. As such, the observation of the V-shaped motif cannot explain how the EPS matrix facilitates a diffuse 3D cell arrangement [102] in its macroscopic state. In fact, a current unanswered question in bacterial biofilm microbiology is how the physico-chemical properties of the EPS matrix modulate the formation of micro-domains and pathogenic niches in polymicrobial biofilm communities [42]. Taken collectively, there will always be a gap between the fine structure and the ultrastructure, brought into focus through limitations in computational resources and quantum mechanical tractability.

Scale limitations also effect the systematic design of the models. Mass spectroscopy analysis of expectorated CF sputum has quantified the relative abundances of certain cations in the CF lung, specifically, that the concentration of Na^+ is greater than the concentration of Ca^{2+}, which is greater than the concentration of Mg^{2+} [125]. In this research, EPS models were cross-linked about a mono-species cationic distribution and, based on the calculated thermodynamic stabilities, an assumption has been made regarding which ions will effectively out-compete the other ions for EPS binding. Although in principle this is

not incorrect, an alternative development pipeline would be to cross-link the EPS chains accommodating all sputum ions in a ratio that satisfies their relative abundances observed experimentally. This would require EPS chains larger than the quadramer scale. For example, there is approximately four times the amount of calcium to magnesium in CF sputum [125], meaning that to accommodate four calcium ions and one magnesium ion one would need a minimum of an EPS pentamer, and to accommodate eight calcium ions and two magnesium ions one would need a minimum of an EPS decamer. Unfortunately, for ground-up model development, this would quickly make quantum mechanical computation intractable and, as such, this particular scale limitation has enforced a particular systematic design protocol. As mixed ionic-species models were not investigated, this introduces the possibility that a more stable cross-linked scaffold may exists through mixed species cross-linking. Understandably, these particular cross-linked scaffolds could give rise to a morphology that deviates from the mono-species 2-chain cross-linked scaffold geometries predicted in this work. However, it is important to note here that the predicted cation cross-linked EPS molecular structures in this work do not differ significantly from the *P. aeruginosa* EPS atomic-scale morphology observed *in vitro* through NMR spectroscopy, TEM and amphiphilic carbon quantum dot imaging [111, 112, 130, 308, 309]. As such, if any morphological differences arise through mixed species cross-linking, the deviation away from the structures obtained in this work is not expected to be large. Additional discussions of the potential atomic-scale morphological alterations, elicited by a different model design approach, are given in the **Section 8.3.2** below.

8.3.2 Matrix morphology

During the cation-induced 2-chain cross-linking, the PolyM and PolyMG models were cross-linked separately. The cation mediated cross-linking of one PolyM chain with one PolyMG chain was not investigated, although it has been shown to occur spectroscopically when mucoid *P. aeruginosa* biofilm EPS matrices are exposed to calcium and magnesium ions *in vitro* [130]. This could potentially lead to an architecture that is equally or more stable with a unique morphology. As the model development was ground-up, defining smaller building units from which a larger model was constructed, a "domino effect" emerges, whereby these alterations in the design at the 2-chain scale could induce unrealistic changes at the 4-chain scale, compounding small inaccuracies. For example, with regards to QSAI motion through the EPS scaffold, the PQS molecule may not tether to the EPS interface through electrostatic interactions, rendering itself immobile, if the cationic charge distribution across the molecular frame were different as a result of mixed chain cross-linking. By extension, the mechanism behind exothermic OligoG CF-5/20 dispersal into the EPS, as well as the mechanism behind unstable gallium accommodation, may be different for the same reason.

Although the EPS is the dominant component of mucoid *P. aeruginosa* biofilm matrices, that is not to suggest that other ECM components which occur in lower abundances are not contributing to matrix macrostructure and stability. *P. aeruginosa* cells within mature, chronic, biofilm matrices produce pyocyanin which promotes cell autolysis [155, 157]. This will serve to release cystolic constituents such as proteins, in addition to DNA, into the biofilm matrix. Similar to the EPS, eDNA is capable of chelating calcium, which can increase the viscosity and mechanical stability of the biofilm matrix [359]. Indeed, preventing calcium-induced DNA aggregation, for example by administering OligoG CF-5/20 therapy, has been shown to disperse mature *P. aeruginosa* biofilm matrices *in vitro* [183]. Therefore, a more representative model may encompass eDNA biopolymeric structures, scaled to achieve the same EPS : eDNA molecular weight ratio at the 4-chain EPS quadramer scale, that is observed *in vitro* or *in vivo*, cross-linked with the EPS about calcium. As discussed above, this novel cross-linking interaction could alter the matrix scaffold morphology, potentially altering the physico-chemical model predictions.

8.3.3 Matrix permeation

When studying the molecular functionality that governs the motion of QSAI molecules through the EPS scaffold, 3-oxo-C_{12}-HSL was neglected from the investigation as it is primarily responsible for the initial differentiation into the biofilm mode of life and, therefore, is not definitively associated with movements in-between the bacteria within the mature biofilm [146, 162, 163]. However, given that 3-oxo-C_{12}-HSL is still present at 0.5 ppm in 24 hour old *in vitro P. aeruginosa* biofilms [162], it will be moving through the matrix to some degree. As such, an argument could be made for studying 3-oxo-C_{12}-HSL dynamics in addition to PQS and C_4-HSL. Studying the motion of 3-oxo-C_{12}-HSL in the proximity of the EPS under physiological conditions, would further shed light onto the interplay between hydrophobicity, vdW interactions and ionic binding potency, that dictate the motion of these molecules throughout the EPS scaffold. 3-oxo-C_{12}-HSL is structurally similar to C_4-HSL, the only difference being a longer hydrocarbon tail harbouring an additional ketone group. New simulations should be able to ascertain whether the increased hydrophobicity, attributed to the longer hydrocarbon chain, reduces the tendency of the molecule to migrate into the solvent space and whether the exploitation of vdW interactions with the EPS, perhaps, is a chemical modality specific to homoserine lactones.

In addition to studying the dynamics of 3-oxo-C_{12}-HSL, it would also be appropriate to study the dynamics of small organic molecules that function as quorum sensing inhibitors (QSIs) that are effective against mature *P. aeruginosa* biofilms, blocking the binding of C_4-HSL to the RhlR transcriptional regulator. Such a QSI molecule must be able to permeate the EPS matrix and, therefore, it would be interesting to observe whether this

molecule exploits the same inter-molecular interactions as C_4-HSL, underlining the molecular functionality and inter-molecular associations that govern matrix motion. Appropriate QSI test molecules should encompass a head group that retains the oxygen hetero-atom, a carbonyl group adjacent to the carbon atom of the amide linkage and a hydrocarbon tail with no more than four carbon atoms, as these structural descriptors have been shown to promote the largest activity against the RhlR system of *P. aeruginosa* [360]. The same simulation protocol as used with C_4-HSL could be applied here also, and the conformational dynamics interrogated to gauge whether vdW interactions are the sole EPS contacts, in addition to the molecule displaying a large solvent proclivity. The electronic structure, specifically the HOMO-LUMO separation, could also be examined in detail to compute the chemical reactivity indices, to ascertain whether the lack of conjugated aromaticity increases chemical hardness, lowering the ionic binding potency. All this information would further elucidate the observation that these particular structural motifs and site-directed interactions promote motion of molecules through the matrix.

8.3.4 Gallium targets

The thermodynamic stability of iron-for-gallium substitutions into ferric-PQS was interrogated to elucidate another possible mode of cellular gallium uptake in bacterial cells deficient in siderophore uptake systems. However, *P. aeruginosa* can also acquire iron (Fe^{3+}) using other siderophore-independent pathways, such as the ferric citrate [201]. In fact, when deficient in the siderophores pyochelin and pyoverdine, addition of exogenous ferric citrate restored the ability of *P. aeruginosa* to form structurally heterogeneous 3D biofilm matrix scaffolds [201]. Therefore, a more thorough examination of novel gallium targets should include gallium substitutions into ferric citrate complexes. The molecular structure of ferric citrate is given in **Fig 8.2**.

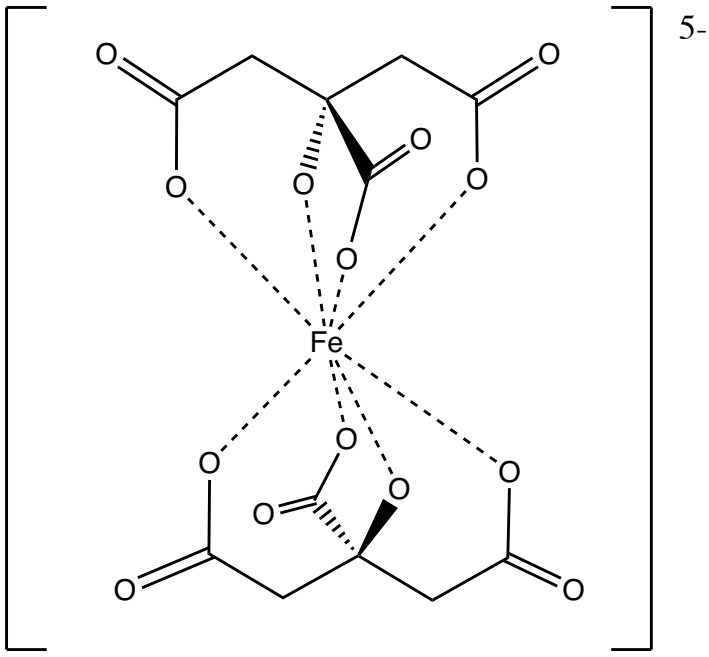


Figure 8.2: Molecular structure of ferric citrate.

When coordinated to a single Fe^{3+} centre, the hydroxyl group is deprotonated and the citrate adopts a net charge of -4 [361]. The mono-iron dicitrate species $[Fe(Cit)_2]^{5-}$ is most prevalent in aqueous medium at physiological pH [361] and at physiological body temperature (310 K) the chelated citrate ligands enforce a high-spin (sextet; S=5/2) spin state on the central iron cation [362]. The thermodynamic stability of the iron-for-gallium exchange into ferric citrate could be probed in identical fashion to the protocol applied to study substitutions into ferric PQS. This would further extend understanding into how gallium gains entry, in a siderophore-independent fashion, into *P. aeruginosa* bacterial cells.

8.3.5 Airway mucin

Mucoid *P. aeruginosa* biofilm colonies, within the Cystic Fibrosis (CF) lung, are found within the mucus, intraluminally [93]. Mucins are heavily glycosylated, large molecular weight, polymeric proteins which are capable of forming gelled structures that function as protective barriers and lubricants. The predominant mucin in the CF airway, contributing the bulk of the sticky, hard to clear, airway mucus is MUC5AC [363]. *P. aeruginosa* biofilm exoproducts are capable of interacting with CF mucin, which causes cellular immobilisation and increases bacterial adherence to the CF airway, in turn, promoting biofilm matrix proliferation [364]. As such, to a good approximation, it is very likely that *in vivo* the EPS matrix will be in complex with the surrounding mucin.

In light of this, an appropriate extension to the model would be to layer the 1-, 2- and/or 4-chain EPS models onto a truncated model of MUC5AC, for example a glycosylated

quadramer polypeptide of the central motif of MUC5AC. A similar protocol was followed in the work of Pritchard *et al.* when investigating the interactions established between a 12-mer OligoG CF-5/20 molecular model and MUC5AC, specifically, a glycosylated 100-mer polypeptide of the central motif of MUC5AC [181]. Note, in our case, a quadramer length polypeptide would have to be used as the EPS chains are quadramers. As discussed above, this is yet another example of a scale limitation of quantum mechanical computation.

Using DFT modelling, one could elucidate the molecular functionality pertinent for thermodynamically stable EPS-mucin interactions and tethering, testing whether the EPS can establish direct inter-molecular associations with CF airway mucin, or whether any tethering is facilitated by the elevated cationic atmosphere. Thermodynamically stable EPS-mucin interactions will anchor the EPS to the surrounding mucus, which, as a consequence, will serve to increase the resistance of the EPS to clearance, either through pharmaceutical intervention or through host defense mechanisms. Understanding the type, directionality and strength of EPS-mucin interactions can provide insights into which functionality to target, through pharmaceutical intervention, to best assist in clearance of the mucoid EPS matrix from the CF lung.

8.3.6 Electrostatic complementarity

Electrostatic interactions, governed by the cationic charge distribution and the orientation (and accessibility) of charge donor functional groups, have proved critical for dictating matrix stabilisation and morphology, but also these interactions dictate the motion of molecules throughout the matrix scaffold. When no electrostatic interactions are sustained, molecular mobility is high. Potentially, therefore, being able to quickly predict the strength of any electrostatic interactions that could occur between a target molecule and EPS scaffold could offer insight into whether this molecule is capable of EPS penetration without the need to perform long molecular dynamics simulations. Quick predictions on EPS-molecule (electrostatic) interactions then permit the analysis of a larger number of molecules in a shorter time frame. Such an approach is commonly referred to as virtual screening or *molecular docking*. Unfortunately, the use of Density-Functional Theory (DFT) to compute the molecular structures and the directionality of the electrostatic interactions, although accurate, would be too computationally slow to permit the screening of large numbers of molecules. Therefore, in the interest of assessing complementary based primarily on electrostatic interactions, polarizable forcefields, such as the polarizable atomic multipole AMOEBA forcefield [365], can be utilised.

Traditionally, virtual screening and molecular docking is used across computational drug design to predict the conformations of ligands that give rise to the largest protein-ligand

binding affinities [366–368]. In this proposed approach, and unconventionally, the physiological 4-chain EPS scaffold (4-PolyMG$_{MD}$) would act as the template system in which multiple different ligands (drug candidates) could be *docked* and the sites of the electrostatic interactions, along-side the EPS-ligand binding affinity, identified using AMOEBA. Drug candidates predicted to interact with the EPS through a large volume of different electrostatic interactions, would likely become sequestered by the EPS and not be able to reach the biofilm bacteria enveloped deep within the EPS matrix. This particular screening approach would not be possible if it were not for the DFT and MD models developed in this work which, not only provide a template model representative the mucoid EPS architecture observed *in vivo*, but also have elucidated that electrostatics are key for governing molecular sequestration by the EPS.

8.3.7 Coarse-grained simulations

The simulations in this research have identified molecular-level details, specifically, the molecular interactions and atomistic mechanisms critical for mucoid *P. aeruginosa* EPS matrix stabilisation, permeation and disruption. However, the molecular models generated in this research are too small in scale to study the time-scales required for deep penetration of molecules into the EPS network. This particular phenomenon is of critical importance when testing as to whether a novel antimicrobial is capable of penetrating the EPS to reach its bacterial protein target. Understandably, novel antimicrobial molecules that cannot achieve this would be perform poorly during *in vitro* testing and, as such, it is advantageous to filter these molecules from the drug discovery pipeline before resources are spent on experimental testing.

To develop such an *antimicrobial virtual screening platform*, coarse-graining is necessary. In coarse-grained simulations, collections of atoms are represented as a single "bead" [254]. As such, a molecule can be represented as a small collection of connected beads and, as a result, molecular systems encompassing many molecules can be modelled as system of thousands of connected beads instead of millions of connected atoms. What this now facilitates, is the the construction of a molecular model of hundreds of EPS chains of varying length scales, cross-linked about Ca^{2+} ions, forming a cationic stabilised polymeric matrix. Thermal equilibration of this system at physiological temperature in the presence of water would generate a model of the EPS network now at the mesoscale and should, therefore, be able to capture the dendritic fractal-like EPS scaffold architecture with the branching projection though 3D space.

With a coarse-grained model, one is now in a position to study molecular penetration into the calcium cross-linked polymeric matrix. Conceptually, a possible way to deploy this

model in simulation would be to position the target molecule above the polymeric matrix and simulate the forwards time propagation into the μs regime. Novel antimicrobials that show good permeation properties will pass deeper into the matrix scaffold, in comparison to molecules with poor permeation properties, over the same time-period.

8.4 Concluding remarks

In conclusion this thesis has achieved the aim of elucidating mucoid *P. aeruginosa* biofilm structural chemistry that, in future, can serve as the critical first step when considering the design of novel antimicrobial agents targeted at mature matrix eradication or deep matrix penetration. First principles and molecular dynamics modelling have been employed to develop the first atomistic model of the mature mucoid *P. aeruginosa* EPS (biofilm) matrix present within the cystic fibrosis lung, which was deployed to ascertain the physico-chemical interactions that influence matrix stability, matrix weakening and how molecules are permitted to move through the matrix. The combination of DFT and MD proved vital for computation of accurate thermochemistry and modes of inter-molecular interaction, allowing for connections to be established between EPS atomic structure, biofilm chronicity and bacterial virulence. The applicability of the model(s) and theoretical approach extend also to the identification of novel pharmaceutical targets in addition to molecular functionality that may give rise to more efficacious antimicrobials.

References

(1) W. Wu, Y. Jin, F. Bai and S. Jin, *Molecular Medical Microbiology*, Academic Press, Boston, 2015.

(2) P. Pachori, R. Gothalwal and P. Gandhi, *Genes Dis.*, 2019, **6**, 109–119.

(3) P. M. de Abreu, P. G. Farias, G. S. Paiva, A. M. Almeida and P. V. Morais, *BMC Microbiol.*, 2014, **14**, 118.

(4) C. M. Blanchette, J. M. Noone, G. Stone, E. Zacherle, R. P. Patel, R. Howden and D. Mapel, *Med. Sci.*, 2017, **5**, 20.

(5) C. Nathan and O. Cars, *N. Engl. J. Med.*, 2014, **371**, 1761–1763.

(6) Z. Pang, R. Raudonis, B. R. Glick, T. Lin and Z. Cheng, *Biotechnol. Adv.*, 2019, **37**, 177–192.

(7) K. L. May and M. Grabowicz, *Proc. Natl. Acad. Sci. U.S.A.*, 2018, **115**, 8852–8854.

(8) H. Nikaido, *Microbiol. Mol. Biol. Rev.*, 2003, **67**, 593–656.

(9) R. K. Ernst, A. M. Hajjar, J. H. Tsai, S. M. Moskowitz, C. B. Wilson and S. I. Miller, *J. Endotoxin Res.*, 2003, **9**, 395–400.

(10) M. Rivera, L. E. Bryan, R. E. Hancock and E. J. McGroarty, *J. Bacteriol.*, 1988, **170**, 512–521.

(11) S. Yokota, S. Kaya, S. Sawada, T. Kawamura, Y. Araki and E. Ito, *Eur. J. Biochem*, 1987, **167**, 203–209.

(12) S. M. Huszczynski, J. S. Lam and C. M. Khursigara, *Pathogens*, 2019, **9**, 6.

(13) S. Azimi, J. Thomas, S. E. Cleland, J. E. Curtis, J. B. Goldberg and S. P. Diggle, *mBio*, 2021, **12**, e00860–21.

(14) K. Murphy, A. J. Park, Y. Hao, D. Brewer, J. S. Lam and C. M. Khursigara, *J. Bacteriol*, 2014, **196**, 1306–1317.

(15) H. Labischinski, G. Barnickel, H. Bradaczek, D. Naumann, E. T. Rietschel and P. Giesbrecht, *J. Bacteriol.*, 1985, **162**, 9.

(16) M. Kastowsky, A. Sabisch, T. Gutberlet and H. Bradaczek, *Eur. J. Biochem*, 1991, **197**, 707–716.

(17) D. Naumann, C. Schultz, A. Sabisch, M. Kastowsky and H. Labischinski, *J. Mol. Struct*, 1989, **214**, 213–246.

(18) S. Obst, M. Kastowsky and H. Bradaczek, *Biophys. J.*, 1997, **72**, 1031–1046.

(19) T. S. Carpenter, J. Parkin and S. Khalid, *J. Phys. Chem. Lett.*, 2016, **7**, 3446–3451.

(20) D. S. Patel, S. Re, E. L. Wu, Y. Qi, P. E. Klebba, G. Widmalm, M. S. Yeom, Y. Sugita and W. Im, *Biophys. J.*, 2016, **110**, 930–938.

(21) C. A. López, H. Zgurskaya and S. Gnanakaran, *Biochim. Biophys. Acta - Biomembr.*, 2020, **1862**, 183151.

(22) D. S. Snyder and T. J. McIntosh, *Biochemistry*, 2000, **39**, 11777–11787.

(23) M. Vaara, *Antimicrob. Agents Chemother.*, 1993, **37**, 354–356.

(24) J. M. Blair, M. A. Webber, A. J. Baylay, D. O. Ogbolu and L. J. Piddock, *Nat. Rev. Microbiol.*, 2014, **13**, 42–51.

(25) J. R. Aeschlimann, *Pharmacotherapy*, 2003, **23**, 916–924.

(26) X. Li, H. Nikaido and K. Poole, *Antimicrob Agents Chemother.*, 1995, **39**, 1948–1953.

(27) D. Choudhury, A. Das Talukdar, M. Dutta Choudhury, A. P. Maurya, D. Paul, D. Dhar Chanda, A. Chakravorty and A. Bhattacharjee, *PLoS ONE*, 2015, **10**, e0133842.

(28) L. Xian-Zhi, P. Keith and N. Hiroshi, *Antimicrob Agents Chemother.*, 2003, **47**, 27–33.

(29) P. Dupont, D. Hocquet, K. Jeannot, P. Chavanet and P. Plésiat, *J Antimicrob Chemother.*, 2005, **55**, 518–522.

(30) N. Masuda and S. Ohya, *Antimicrob Agents Chemother.*, 1992, **36**, 1847–1851.

(31) P. V. Pesingi, B. R. Singh, P. K. Pesingi, M. Bhardwaj, S. V. Singh, M. Kumawat, D. K. Sinha and R. K. Gandham, *Front. Microbiol.*, 2019, **10**, 2664.

(32) M. R. Tapia-Rodriguez, A. Hernandez-Mendoza, G. A. Gonzalez-Aguilar, M. A. Martinez-Tellez, C. M. Martins and J. F. Ayala-Zavala, *Food Control*, 2017, **75**, 255–261.

(33) C. A. López, T. Travers, K. M. Pos, H. I. Zgurskaya and S. Gnanakaran, *Sci. Rep.*, 2017, **7**, 16521.

(34) E. B. M. Breidenstein, C. de la Fuente-Núñez and R. E. W. Hancock, *Trends Microbiol.*, 2011, **19**, 419–426.

(35) Z. Pang, R. Raudonis, B. R. Glick, T. Lin and Z. Cheng, *Biotechnol. Adv.*, 2019, **37**, 177–192.

(36) F. Alain, *J. Bacteriol.*, 2010, **192**, 3243–3245.

(37) W. Arber, *Life*, 2014, **4**, 217–224.

(38) L. Ahmadian, Z. Norouzi Bazgir, M. Ahanjan, R. Valadan and H. R. Goli, *Biomed Res. Int.*, 2021, **2021**, 7077344.

(39) M. A. Adam and W. I. Elhag, *BMC Infectious Diseases*, 2018, **18**, 668.

(40) J. Yan, P. Hsueh, J. Lu, F. Chang, W. Ko and J. Wu, *J. Antimicrob. Chemother.*, 2006, **58**, 530–536.

(41) J. W. Costerton, Z. Lewandowski, D. E. Caldwell, D. R. Korber and H. M. Lappin-Scott, *Annu. Rev. Microbiol.*, 1995, **49**, 711–745.

(42) L. Karygianni, Z. Ren, H. Koo and T. Thurnheer, 2020, **28**, 668–681.

(43) H. Flemming and J. Wingender, *Nature Reviews Microbiology*, 2010, **8**, 623–633.

(44) L. K. Jennings, K. M. Storek, H. E. Ledvina, C. Coulon, L. S. Marmont, I. Sadovskaya, P. R. Secor, B. S. Tseng, M. Scian, A. Filloux, D. J. Wozniak, P. L. Howell and M. R. Parsek, *Proc. Natl. Acad. Sci. U.S.A.*, 2015, **112**, 11353–11358.

(45) N. Peng, P. Cai, M. Mortimer, Y. Wu, C. Gao and Q. Huang, *BMC Microbiol.*, 2020, **20**, 115.

(46) S. Wang, X. Liu, H. Liu, L. Zhang, Y. Guo, S. Yu, D. J. Wozniak and L. Z. Ma, *Environ. Microbiol. Rep.*, 2015, **7**, 330–340.

(47) C. Reichhardt, H. M. Jacobs, M. Matwichuk, C. Wong, D. J. Wozniak and M. R. Parsek, *J. Bacteriol.*, 2020, **202**, e00216–20.

(48) C. Reichhardt, C. Wong, D. Passos da Silva, D. J. Wozniak and M. R. Parsek, *mBio*, 2018, **9**, e01376–18.

(49) J. Janin, R. P. Bahadur and P. Chakrabarti, *Q. Rev. Biophys.*, 2008, **41**, 133–180.

(50) D. Passos da Silva, M. L. Matwichuk, D. O. Townsend, C. Reichhardt, D. Lamba, D. J. Wozniak and M. R. Parsek, *Nat. Commun.*, 2019, **10**, 2183.

(51) S. P. Diggle, R. E. Stacey, C. Dodd, M. Cámara, P. Williams and K. Winzer, *Environ. Microbiol.*, 2006, **8**, 1095–1104.

(52) X. Liu, S. Guo, T. Bocklitz, P. Rösch, J. Popp and H. Yu, *Water Res.*, 2022, **210**, 117973.

(53) B. Pergolese, A. Bonifacio and A. Bigotto, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3610–3613.

(54) W. Jiang and S. Ran, *J. Chem. Phys.*, 2018, **148**, 205102.

(55) T. Wang, S. Flint and J. Palmer, *Biofouling*, 2019, **35**, 959–974.

(56) S. Yu, Q. Wei, T. Zhao, Y. Guo and L. Z. Ma, *Appl. Environ. Microbiol.*, 2016, **82**, 6403–6413.

(57) S. Patil, A. Sandberg, E. Heckert, W. Self and S. Seal, *Biomaterials*, 2007, **28**, 4600–4607.

(58) Q. Xiao, S. Huang, Z. Qi, B. Zhou, Z. He and Y. Liu, *Biochim Biophys Acta Proteins Proteom*, 2008, **1784**, 1020–1027.

(59) E. Maunders and M. Welch, *FEMS Microbiol. Lett.*, 2017, **364**, fnx120.

(60) G. Lise and T. Tolker-Nielsen, *Antimicrob Agents Chemother.*, 2017, **61**, e02696–16.

(61) L. Craig, M. E. Pique and J. A. Tainer, *Nat. Rev. Microbiol.*, 2004, **2**, 363–378.

(62) L. Imundo, J. Barasch, A. Prince and Q. Al-Awqati, *Proc. Natl. Acad. Sci. U.S.A.*, 1995, **92**, 3019.

(63) L. Saiman and A. Prince, *J Clin Invest.*, 1993, **92**, 1875–1880.

(64) H. C. Krivan, V. Ginsburg and D. D. Roberts, *Arch. Biochem. Biophys.*, 1988, **260**, 493–496.

(65) R. Ramphal, C. Carnoy, S. Fievre, J. C. Michalski, N. Houdret, G. Lamblin, G. Strecker and P. Roussel, *Infect. Immun.*, 1991, **59**, 700–704.

(66) C. L. Giltner, E. J. Van Schaik, G. F. Audette, D. Kao, R. S. Hodges, D. J. Hassett and R. T. Irvin, *Mol. Microbiol.*, 2006, **59**, 1083–1096.

(67) R. Bos, H. C. van der Mei and H. J. Busscher, *FEMS Microbiol. Rev.*, 1999, **23**, 179–230.

(68) L. L. Burrows, *Annu. Rev. Microbiol.*, 2012, **66**, 493–520.

(69) A. Semmler, C. B. Whitchurch and J. S. Mattick, *Microbiology*, 1999, **145**, 2863–2873.

(70) G. A. O'Toole and R. Kolter, *Mol. Microbiol.*, 1998, **30**, 295–304.

(71) L. M. Otton, M. da Silva Campos, K. L. Meneghetti and G. Corção, *Arch. Microbiol.*, 2017, **199**, 677–682.

(72) G. Horna, K. Quezada, S. Ramos, N. Mosqueda, M. Rubio, H. Guerra and J. Ruiz, *Int. Microbiol.*, 2019, **22**, 131–141.

(73) M. Klausen, A. Heydorn, P. Ragas, L. Lambertsen, A. Aaes-Jørgensen, S. Molin and T. Tolker-Nielsen, *Mol. Microbiol.*, 2003, **48**, 1511–1524.

(74) P. Stoodley, K. Sauer, D. G. Davies and J. W. Costerton, *Annu. Rev. Microbiol.*, 2002, **56**, 187–209.

(75) M. Jacques, V. Aragon and Y. Tremblay, *Anim. Health Res. Rev.*, 2010, **11**, 97–121.

(76) R. M. Donlan and J. W. Costerton, *Clin. Microbiol. Rev.*, 2002, **15**, 167–193.

(77) B. Lubamba, B. Dhooghe, S. Noel and T. Leal, *Clin. Biochem.*, 2012, **45**, 1132–1144.

(78) P. B. Davis, *Am. J. Respir. Crit. Care Med.*, 2006, **173**, 475–482.

(79) M. P. Anderson, D. P. Rich, R. J. Gregory, A. E. Smith and M. J. Welsh, *Science*, 1991, **251**, 679–682.

(80) C. E. Bear, C. Li, N. Kartner, R. J. Bridges, T. J. Jensen, M. Ramjeesingh and J. R. Riordan, *Cell*, 1992, **68**, 809–818.

(81) A. Serohijos, T. Hegedűs, A. Aleksandrov, L. He, L. Cui, N. V. Dokholyan and J. R. Riordan, *Proc. Natl. Acad. Sci. U.S.A.*, 2008, **105**, 3256–3261.

(82) K. Du, M. Sharma and G. L. Lukacs, *Nat. Struct. Mol. Biol*, 2005, **12**, 17–25.

(83) M. J. Welsh and A. E. Smith, *Cell*, 1993, **73**, 1251–1254.

(84) P. M. Farrell, B. J. Rosenstein, T. B. White, F. J. Accurso, C. Castellani, G. R. Cutting, P. R. Durie, V. A. LeGrys, J. Massie, R. B. Parad, M. J. Rock and P. W. Campbell, *J. Pediatr.*, 2008, **153**, S4–S14.

(85) H. Heijerman, *J. Cyst. Fibros.*, 2005, **4**, 3–5.

(86) *The Cystic Fibrosis Trust*, https://www.cysticfibrosis.org.uk, (accessed June 2022).

(87) J. B. Lyczak, C. L. Cannon and G. B. Pier, *Microbes. Infect.*, 2000, **2**, 1051–1060.

(88) F. T. Coleman, S. Mueschenborn, G. Meluleni, C. Ray, V. J. Carey, S. O. Vargas, C. L. Cannon, F. M. Ausubel and G. B. Pier, *Proc. Natl. Acad. Sci. U.S.A.*, 2003, **100**, 1949.

(89) A. Bauernfeind, G. Hörl, R. Jungwirth, C. Petermüller, B. Przyklenk, C. Weisslein-Pfister, R. M. Bertele and K. Harms, *Infection*, 1987, **15**, 270–277.

(90) E. Rossi, R. La Rosa, J. A. Bartell, R. L. Marvig, J. A. J. Haagensen, L. M. Sommer, S. Molin and H. K. Johansen, *Nat. Rev. Microbiol.*, 2021, **19**, 331–342.

(91) N. Høiby, *Acta Pathol. Microbiol. Scand. B*, 1975, **83B**, 549–552.

(92) S. S. Pedersen, *APMIS Suppl.*, 1992, **28**, 1–79.

(93) T. Bjarnsholt, P. Ø. Jensen, M. J. Fiandaca, J. Pedersen, C. R. Hansen, C. B. Andersen, T. Pressler, M. Givskov and N. Høiby, *Pediatr. Pulmonol.*, 2009, **44**, 547–558.

(94) S. S. Pedersen, N. Høiby, F. Espersen and C. Koch, *Thorax*, 1992, **47**, 6–13.

(95) R. Morten, J. P. Østrup, N. C. Henrik and T. Tolker-Nielsen, *Infect. Immun.*, 2020, **89**, e00631–20.

(96) B. Rada, *Pathogens*, 2017, **6**, 10.

(97) C. A. Demko, P. J. Byard and P. B. Davis, *J. Clin. Epidemiol.*, 1995, **48**, 1041–1049.

(98) C. E. Chitnis and D. E. Ohman, *Mol. Microbiol.*, 1993, **8**, 583–590.

(99) M. Y. Galperin, *Environ. Microbiol.*, 2018, **20**, 4221–4229.

(100) D. Qiu, V. M. Eisinger, D. W. Rowen and H. D. Yu, *Proc. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 8107–8112.

(101) B. O. Cezairliyan and R. T. Sauer, *Mol. Microbiol.*, 2009, **72**, 368–379.

(102) H. M. Jacobs, L. O'Neal, E. Lopatto, D. J. Wozniak, T. Bjarnsholt and M. R. Parsek, *J. Bacteriol.*, 2022, **204**, e00568–21.

(103) L. R. Evans and A. Linker, *J. Bacteriol*, 1973, **116**, 915–924.

(104) S. Grobe, J. Wingender and H. G. Trüiper, *J. Appl. Microbiol.*, 1995, **79**, 94–102.

(105) O. Orgad, Y. Oren, S. L. Walker and M. Herzberg, *Biofouling*, 2011, **27**, 787–798.

(106) V. Sherbrock-Cox, N. J. Russell and P. Gacesa, *Carbohydr. Res.*, 1984, **135**, 147–154.

(107) A. Linker and R. S. Jones, *J. Biol. Chem*, 1966, **241**, 3845–3851.

(108) D. McShane, J. C. Davies, M. G. Davies, A. Bush, D. M. Geddes and E. Alton, *Eur. Respir. J*, 2003, **21**, 37–42.

(109) H. Fischer, *WIREs Membr Transp Signal*, 2012, **1**, 247–258.

(110) I. W. Davidson, I. W. Sutherland and C. J. Lawson, *J. Gen. Microbiol.*, 1977, **98**, 603–606.

(111) G. SkjÅk-Bræk, H. Grasdalen and B. Larsen, *Carbohydr. Res.*, 1986, **154**, 239–250.

(112) G. SkjÅk-Bræk, S. Paoletti and T. Gianferrara, *Carbohydr. Res.*, 1989, **185**, 119–129.

(113) M. J. Franklin and D. E. Ohman, *J. Bacteriol.*, 1993, **175**, 5057–5065.

(114) M. J. Franklin and D. E. Ohman, *J. Bacteriol.*, 2002, **184**, 3000–3007.

(115) W. Chanasit, Z. J. Gonzaga and B. H. Rehm, *Appl Microbiol Biotechnol*, 2020, **104**, 2179–2191.

(116) I. D. Hay, Z. U. Rehman, M. F. Moradali, Y. Wang and B. H. Rehm, *Microb. Biotechnol.*, 2013, **6**, 637–650.

(117) P. Tielen, M. Strathmann, K.-E. Jaeger, H. C. Flemming and J. Wingender, *Microbiol. Res.*, 2005, **160**, 165–176.

(118) D. E. Nivens, D. E. Ohman, J. Williams and M. J. Franklin, *J. Bacteriol.*, 2001, **183**, 1047–1057.

(119) G. B. Pier, F. Coleman, M. Grout, M. J. Franklin and D. E. Ohman, *Infect. Immun.*, 2001, **69**, 1895–1901.

(120) N. J. Russell and P. Gacesa, *Mol. Aspects Med.*, 1988, **10**, 1–91.

(121) C. Mayer, R. Moritz, C. Kirschner, W. Borchard, R. Maibaum, J. Wingender and H. Flemming, *Int. J. Biol. Macromol.*, 1999, **26**, 3–16.

(122) M. Wloka, H. Rehage, H. Flemming and J. Wingender, *Colloid Polym. Sci.*, 2004, **282**, 1067–1076.

(123) M. Wloka, H. Rehage, H. Flemming and J. Wingender, *Biofilms*, 2005, **2**, 275–283.

(124) S. Grandjean Lapierre, M. Phelippeau, C. Hakimi, Q. Didier, M. Reynaud-Gaubert, J. C. Dubus and M. Drancourt, *Medicine*, 2017, **96**.

(125) D. J. Smith, G. J. Anderson, S. C. Bell and D. W. Reid, *J. Cyst. Fibros*, 2014, **13**, 289–295.

(126) O. Tabary, J. M. Zahm, J. Hinnrasky, J. P. Couetil, P. Cornillet, M. Guenounou, D. Gaillard, E. Puchelle and J. Jacquot, *Am. J. Pathol.*, 1998, **153**, 921–930.

(127) T. P. Dean, Y. Dai, J. K. Shute, M. K. Church and J. O. Warner, *Pediatr. Res.*, 1993, **34**, 159–161.

(128) G. Majka, H. Mazurek, M. Strus, M. Ciszek-Lenda, R. Szatanek, A. Pac, E. Golińska and J. Marcinkiewicz, *Clin. Exp. Immunol.*, 2021, **205**, 391–405.

(129) D. L. Frey, S. Boutin, S. A. Dittrich, S. Y. Graeber, M. Stahl, S. Wege, F. J. Herth, O. Sommerburg, C. Schultz, M. A. Mall and A. H. Dalpke, *J. Cyst. Fibros.*, 2021, **20**, 754–760.

(130) D. Lattner, H. Flemming and C. Mayer, *Int. J. Biol. Macromol.*, 2003, **33**, 81–88.

(131) M. A. LeRoux, F. Guilak and L. A. Setton, *J. Biomed. Mater. Res.*, 1999, **47**, 46–53.

(132) V. Körstgens, H. Flemming, J. Wingender and W. Borchard, *Water Sci. Technol*, 2001, **43**, 49–57.

(133) S. Sarkisova, M. A. Patrauchan, D. Berglund, D. E. Nivens and M. J. Franklin, *J. Bacteriol.*, 2005, **187**, 4327–4337.

(134) W. Plazinski, *J. Comput. Chem.*, 2011, **32**, 2988–2995.

(135) W. Plazinski and W. Rudzinski, *Struct. Chem.*, 2012, **23**, 1409–1415.

(136) Z. J. Li, S. Srebnik and O. J. Rojas, *Biomacromolecules*, 2022, **23**, 1949–1957.

(137) W. Plazinski and M. Drach, *J. Comput. Chem.*, 2012, **33**, 1709–1715.

(138) M. B. Stewart, S. R. Gray, T. Vasiljevic and J. D. Orbell, *Carbohydr. Polym.*, 2014, **112**, 486–493.

(139) M. B. Stewart, S. R. Gray, T. Vasiljevic and J. D. Orbell, *Carbohydr. Polym.*, 2014, **102**, 246–253.

(140) P. Agulhon, V. Markova, M. Robitzer, F. Quignard and T. Mineva, *Biomacromolecules*, 2012, **13**, 1899–1907.

(141) M. Zhang, H. Hong, H. Lin, L. Shen, H. Yu, G. Ma, J. Chen and B. Q. Liao, *Water Research*, 2018, **129**, 337–346.

(142) C. Menakbi, F. Quignard and T. Mineva, *J. Phys. Chem. B.*, 2016, **120**, 3615–3623.

(143) L. Bekri, M. Zouaoui-Rabah, M. Springborg and M. S. Rahal, *J. Mol. Model.*, 2018, **24**, 312.

(144) S. Liu, *J. Chem. Sci.*, 2005, **117**, 477–483.

(145) G. Kaklamani, D. Cheneler, L. M. Grover, M. J. Adams and J. Bowen, *J. Mech. Behav. Biomed. Mater.*, 2014, **36**, 135–142.

(146) D. G. Davies, M. R. Parsek, J. P. Pearson, B. H. Iglewski, J. W. Costerton and E. P. Greenberg, *Science*, 1998, **280**, 295–298.

(147) T. R. De Kievit and B. H. Iglewski, *Infect. Immun.*, 2000, **68**, 4839–4849.

(148) R. S. Smith and B. H. Iglewski, *Curr. Opin. Microbiol.*, 2003, **6**, 56–60.

(149) S. E. Darch, O. Simoska, M. Fitzpatrick, J. P. Barraza, K. J. Stevenson, R. T. Bonnecaze, J. B. Shear and M. Whiteley, *Proc. Natl. Acad. Sci. U.S.A.*, 2018, **115**, 4779–4784.

(150) C. H. Tan, K. S. Koh, C. Xie, M. Tay, Y. Zhou, R. Williams, W. J. Ng, S. A. Rice and S. Kjelleberg, *ISME J.*, 2014, **8**, 1186–1197.

(151) U. A. Ochsner and J. Reiser, *Proc. Natl. Acad. Sci. U.S.A.*, 1995, **92**, 6424–6428.

(152) A. Latifi, M. K. Winson, M. Foglino, B. W. Bycroft, G. Stewart, A. Lazdunski and P. Williams, *Mol. Microbiol.*, 1995, **17**, 333–343.

(153) S. J. Pamp and T. Tolker-Nielsen, *J. Bacteriol.*, 2007, **189**, 2531–2539.

(154) M. E. Davey, N. C. Caiazza and G. A. O'Toole, *J. Bacteriol.*, 2003, **185**, 1027–1036.

(155) B. Rada and T. L. Leto, *Trends Microbiol.*, 2013, **21**, 73–81.

(156) L. S. Pierson and E. A. Pierson, *Appl. Microbiol. Biotechnol.*, 2010, **86**, 1659–1670.

(157) L. A. Meirelles and D. K. Newman, *Mol. Microbiol.*, 2018, **110**, 995–1010.

(158) M. Whiteley and E. P. Greenberg, *J. Bacteriol.*, 2001, **183**, 5529–5534.

(159) P. K. Singh, A. L. Schaefer, M. R. Parsek, T. O. Moninger, M. J. Welsh and E. P. Greenberg, *Nature*, 2000, **407**, 762–764.

(160) D. L. Erickson, R. Endersby, A. Kirkham, K. Stuber, D. D. Vollman, H. R. Rabin, I. Mitchell and D. G. Storey, *Infection and Immunity*, 2002, **70**, 1783–1790.

(161) J. P. Pearson, L. Passador, B. H. Iglewski and E. P. Greenberg, *Proc. Natl. Acad. Sci. U.S.A.*, 1995, **92**, 1490–1494.

(162) A. B. Alayande, M. M. Aung and I. S. Kim, *Curr. Microbiol.*, 2018, **75**, 787–793.

(163) S. Favre-Bonté, T. Köhler and C. Van Delden, *J Antimicrob Chemother.*, 2003, **52**, 598–604.

(164) J. P. Pearson, K. M. Gray, L. Passador, K. D. Tucker, A. Eberhard, B. H. Iglewski and E. P. Greenberg, *Proc. Natl. Acad. Sci. U.S.A.*, 1994, **91**, 197–201.

(165) M. Schuster, C. P. Lostroh, T. Ogi and E. P. Greenberg, *J. Bacteriol.*, 2003, **185**, 2066–2079.

(166) E. C. Pesci, J. B. Milbank, J. P. Pearson, S. McKnight, A. S. Kende, E. P. Greenberg and B. H. Iglewski, *Proc. Natl. Acad. Sci. U.S.A.*, 1999, **96**, 11229–11234.

(167) D. N. Collier, L. Anderson, S. L. McKnight, T. L. Noah, M. Knowles, R. Boucher, U. Schwab, P. Gilligan and E. C. Pesci, *FFEMS Microbiol. Lett.*, 2002, **215**, 41–46.

(168) S. L. McKnight, B. H. Iglewski and E. C. Pesci, *J. Bacteriol.*, 2000, **182**, 2702.

(169) D. de Beer, P. Stoodley and Z. Lewandowski, *Biotechnol. Bioeng.*, 1994, **44**, 636–641.

(170) M. Vogt, H. Flemming and W. S. Veeman, *J. Biotechnol*, 2000, **77**, 137–146.

(171) D. A. D'Argenio, M. W. Calfee, P. B. Rainey and E. C. Pesci, *J. Bacteriol.*, 2002, **184**, 6481–6489.

(172) R. Popat, F. Harrison, A. C. da Silva, S. A. Easton, L. McNally, P. Williams and S. P. Diggle, *Proc. R. Soc. B.*, 2017, **284**, 20170200.

(173) F. Bredenbruch, R. Geffers, M. Nimtz, J. Buer and S. Häussler, *Environ. Microbiol.*, 2006, **8**, 1318–1329.

(174) S. P. Diggle, S. Matthijs, V. J. Wright, M. P. Fletcher, S. R. Chhabra, I. L. Lamont, X. Kong, R. C. Hider, P. Cornelis, M. Cámara and P. Williams, *Chem. Biol.*, 2007, **14**, 87–96.

(175) J. R. Lawrence, G. D. Swerhone, U. Kuhlicke and T. R. Neu, *Can. J. Microbiol.*, 2007, **53**, 450–458.

(176) T. S. Charlton, R. De Nys, A. Netting, N. Kumar, M. Hentzer, M. Givskov and S. Kjelleberg, *Environ. Microbiol.*, 2000, **2**, 530–541.

(177) S. T. Flickinger, M. F. Copeland, E. M. Downes, A. T. Braasch, H. H. Tuson, Y. J. Eun and D. B. Weibel, *J. Am. Chem. Soc.*, 2011, **133**, 5966–5975.

(178) M. R. Parsek and E. P. Greenberg, *Trends Microbiol.*, 2005, **13**, 27–33.

(179) D. P. da Silva, M. C. Schofield, M. R. Parsek and B. S. Tseng, *Pathogens.*, 2017, **6**, 51.

(180) W. D. Smith, E. Bardin, L. Cameron, C. L. Edmondson, K. V. Farrant, I. Martin, R. A. Murphy, O. Soren, A. R. Turnbull, N. Wierre-Gore, E. W. Alton, J. G. Bundy, A. Bush, G. J. Connett, S. N. Faust, A. Filloux, P. S. Freemont, A. L. Jones, Z. Takats, J. S. Webb, H. D. Williams and J. C. Davies, *FEMS Microbiol. Lett.*, 2017, **364**, 121.

(181) M. F. Pritchard, L. C. Powell, G. E. Menzies, P. D. Lewis, K. Hawkins, C. Wright, I. Doull, T. R. Walsh, E. Onsøyen, A. Dessen, R. Myrvold, P. D. Rye, A. H. Myrset, H. N. Stevens, L. A. Hodges, G. MacGregor, J. B. Neilly, K. E. Hill and D. W. Thomas, *Mol. Pharm.*, 2016, **13**, 863–872.

(182) M. F. Pritchard, J. L. Oakley, C. D. Brilliant, P. D. Rye, J. Forton, I. J. Doull, I. Ketchell, K. E. Hill, D. W. Thomas and P. D. Lewis, *Vib. Spectrosc.*, 2019, **103**, 102932.

(183) L. C. Powell, M. F. Pritchard, E. L. Ferguson, K. A. Powell, S. U. Patel, P. D. Rye, S. Sakellakou, N. J. Buurma, C. D. Brilliant, J. M. Copping, G. E. Menzies, P. D. Lewis, K. E. Hill and D. W. Thomas, *NPJ biofilms and microbiomes*, 2018, **4**, 13.

(184) A. Ermund, C. V. Recktenwald, G. Skjåk-Braek, L. N. Meiss, E. Onsøyen, P. D. Rye, A. Dessen, A. H. Myrset and G. C. Hansson, *Clin. Exp. Pharmacol. Physiol.*, 2017, **44**, 639–647.

(185) D. Ambort, M. E. Johansson, J. K. Gustafsson, H. E. Nilsson, A. Ermund, B. R. Johansson, P. J. Koeck, H. Hebert and G. C. Hansson, *Proc. Natl. Acad. Sci. U.S.A.*, 2012, **109**, 5645–5650.

(186) J. M. Holmén Larsson, K. A. Thomsson, A. M. Rodríguez-Piñeiro, H. Karlsson and G. C. Hansson, *Am J Physiol Gastrointest Liver Physiol*, 2013, **305**, 357–363.

(187) S. Khan, A. Tøndervik, H. Sletta, G. Klinkenberg, C. Emanuel, E. Onsøyen, R. Myrvold, R. A. Howe, T. R. Walsh, K. E. Hill and D. W. Thomas, *Antimicrob Agents Chemother.*, 2012, **56**, 5134.

(188) J. Stokniene, M. Varache, P. D. Rye, K. E. Hill, D. W. Thomas and E. L. Ferguson, *Sci. Rep.*, 2022, **12**, 4986.

(189) J. L. Oakley, R. Weiser, L. C. Powell, J. Forton, E. Mahenthiralingam, P. D. Rye, K. E. Hill, D. W. Thomas and M. F. Pritchard, *mSphere*, 2021, **6**, e01216–20.

(190) L. C. Powell, M. F. Pritchard, C. Emanuel, E. Onsøyen, P. D. Rye, C. J. Wright, K. E. Hill and D. W. Thomas, *Am. J. Respir. Cell Mol. Biol.*, 2014, **50**, 483–492.

(191) M. F. Pritchard, L. C. Powell, S. Khan, P. C. Griffiths, O. T. Mansour, R. Schweins, K. Beck, N. J. Buurma, C. E. Dempsey, C. J. Wright, P. D. Rye, K. E. Hill, D. W. Thomas and E. L. Ferguson, *Sci. Rep.*, 2017, **7**, 1–12.

(192) L. C. Powell, A. Sowedan, S. Khan, C. J. Wright, K. Hawkins, E. Onsøyen, R. Myrvold, K. E. Hill and D. W. Thomas, *Biofouling*, 2013, **29**, 413–421.

(193) W. Hengzhuang, Z. Song, O. Ciofu, E. Onsøyen, P. D. Rye and N. Høiby, *Antimicrob. Agents Chemother.*, 2016, **60**, 2620–2626.

(194) A. A. Jack, S. Khan, L. C. Powell, M. F. Pritchard, K. Beck, H. Sadh, L. Sutton, A. Cavaliere, H. Florance, P. D. Rye, D. W. Thomas and K. E. Hill, *Antimicrob. Agents Chemother*, 2018, **62**, e02318–17.

(195) S. Moreau-Marquis, J. M. Bomberger, G. G. Anderson, A. Swiatecka-Urban, S. Ye, G. A. O'Toole and B. A. Stanton, *Am. J. Physiol. Lung Cell Mol. Physiol.*, 2008, **295**, 25–37.

(196) D. W. Reid, Q. T. Lam, H. Schneider and E. H. Walters, *Eur. Respir. J.*, 2004, **24**, 286–291.

(197) A. J. Ghio, V. L. Roggli, J. M. Soukup, J. H. Richards, S. H. Randell and M. S. Muhlebach, *J. Cyst. Fibros.*, 2013, **12**, 390–398.

(198) H. Matsui, V. E. Wagner, D. B. Hill, U. E. Schwab, T. D. Rogers, B. Button, R. M. Taylor, R. Superfine, M. Rubinstein, B. H. Iglewski and R. C. Boucher, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **103**, 18131–18136.

(199) D. E. Reid, V. Carroll, C. O'May, A. Champion and S. M. Kirov, *Eur. Respir. J*, 2007, **30**, 286–292.

(200) F. Berlutti, C. Morea, A. Battistoni, S. Sarli, P. Cipriani, F. Superti, M. G. Ammendolia and P. Valenti, *Int. J. Immunopath. Ph.*, 2005, **18**, 661–670.

(201) E. Banin, M. L. Vasil and E. P. Greenberg, *Proc. Natl. Acad. Sci. U.S.A.*, 2005, **102**, 11076–11081.

(202) K. Mettrick, K. Hassan, I. Lamont and D. Reid, *Antibiotics*, 2020, **9**, 144.

(203) P. K. Singh, M. R. Parsek, E. P. Greenberg and M. J. Welsh, *Nature*, 2002, **417**, 552.

(204) V. Vinuesa and M. J. McConnell, *Int. J. Mol. Sci.*, 2021, **22**, 1–19.

(205) R. Tappe, K. Taraz, H. Budzikiewicz, J. M. Meyer and J. F. Lefèvre, *J. Prakt. Chem*, 1993, **335**, 83–87.

(206) N. J. Hare, C. Z. Soe, B. Rose, C. Harbour, R. Codd, J. Manos and S. J. Cordwell, *J. Proteome Res.*, 2012, **11**, 776–795.

(207) C. Tseng, A. Burger, G. L. Mislin, I. J. Schalk, S. S. Yu, S. I. Chan and M. A. Abdallah, *J. Biol. Inorg. Chem.*, 2006, **11**, 419–432.

(208) J. Brandel, N. Humbert, M. Elhabiri, I. J. Schalk, G. L. Mislin and A. Albrecht-Gary, *Dalton Trans.*, 2012, **41**, 2820–2834.

(209) D. Cobessi, H. Celia and F. Pattus, *J. Mol. Biol.*, 2005, **352**, 893–904.

(210) R. G. Ankenbauer and H. N. Quan, *J. Bacteriol.*, 1994, **176**, 307–319.

(211) N. Noinaj, M. Guillier, J. Barnard Travis and S. K. Buchanan, *Annu. Rev. Microbiol.*, 2010, **64**, 43–60.

(212) B. Yang, F. Hoegy, G. L. Mislin, P. J. Mesini and I. J. Schalk, *J. Inorg. Biochem.*, 2011, **105**, 1293–1298.

(213) J. R. Sheldon, H. A. Laakso and D. E. Heinrichs, *Microbiol. Spectr.*, 2016, **4**, 4.2.05.

(214) C. Bradbeer, *J. Bacteriol.*, 1993, **175**, 3146–3150.

(215) Y. Zhang, X. Pan, L. Wang and L. Chen, *J. Drug Target*, 2021, **29**, 249–258.

(216) A. Firoz, M. Haris, K. Hussain, M. Raza, D. Verma, M. Bouchama, K. S. Namiq and S. Khan, *Cureus.*, 2021, **13**, e13716.

(217) R. D. Shannon, *Acta Crystallogr. A*, 1976, **32**, 751–767.

(218) A. Ross-Gillespie, M. Weigert, S. P. Brown and R. Kümmerli, *Evol. Med. Public Health.*, 2014, **2014**, 18–29.

(219) C. H. Goss, Y. Kaneko, L. Khuu, G. D. Anderson, S. Ravishankar, M. L. Aitken, N. Lechtzin, G. Zhou, D. M. Czyz, K. McLean, O. Olakanmi, H. A. Shuman, M. Teresi, E. Wilhelm, E. Caldwell, S. J. Salipante, D. B. Hornick, R. J. Siehnel, L. Becker, B. E. Britigan and P. K. Singh, *Sci Transl Med.*, 2018, **10**, eaat7520.

(220) Y. Kaneko, M. Thoendel, O. Olakanmi, B. E. Britigan and P. K. Singh, *J Clin Invest.*, 2007, **117**, 877.

(221) C. Bonchi, E. Frangipani, F. Imperi and P. Visca, *Antimicrob. Agents Chemother.*, 2015, **59**, 5641–5646.

(222) S. P. Valappil, H. H. Yiu, L. Bouffier, C. K. Hope, G. Evans, J. B. Claridge, S. M. Higham and M. J. Rosseinsky, *Dalton Trans.*, 2013, **42**, 1778–1786.

(223) V. Nikolova, S. Angelova, N. Markova and T. Dudev, *J. Phys. Chem. B*, 2016, **120**, 2241–2248.

(224) N. Kircheva and T. Dudev, *Inorg. Chem.*, 2020, **59**, 6242–6254.

(225) A. Tovar-García, V. Angarita-Zapata, A. Cazares, R. Jasso-Chávez, J. Belmont-Díaz, V. Sanchez-Torres, L. E. López-Jacome, R. Coria-Jiménez, T. Maeda and R. García-Contreras, *Arch. Microbiol.*, 2020, **202**, 617–622.

(226) A. Braud, M. Hannauer, G. L. Mislin and I. J. Schalk, *J. Bacteriol.*, 2009, **191**, 3517–3525.

(227) J. Qin, M. Li, M. Yuan, X. Shi, J. Song, Y. He, H. Mao, D. Kong and Z. Gu, *ACS Appl. Mater. Interfaces*, 2022, **14**, 22426–22442.

(228) E. S. Gloag, G. K. German, P. Stoodley and D. J. Wozniak, *Scientific Reports*, 2018, **8**, 9691.

(229) J. Lin, W. Zhang, J. Cheng, X. Yang, K. Zhu, Y. Wang, G. Wei, P. Qian, Z. Luo and X. Shen, *Nat. Commun.*, 2017, **8**, 14888.

(230) F. Jensen, *Introduction to computational chemistry*, John Wiley & Sons, West Sussex, 2nd edn, 2007.

(231) A. Szabo and N. S. Ostlund, *Modern quantum chemistry: introduction to advanced electronic structure theory*, Dover Publications Inc, New York, 1989.

(232) C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models*, John Wiley & Sons Ltd, West Sussex, 2nd edn, 2004.

(233) W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory*, Wiley-VCH, Weinheim, 2nd edn, 2002.

(234) P. Hohenberg and W. Kohn, *Phys. Rev*, 1964, **136**, B864.

(235) W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133.

(236) D. Sholl and J. A. Steckel, *Density functional theory: a practical introduction*, John Wiley & Sons, New Jersey, 2009.

(237) A. D. Becke, *J. Chem. Phys.*, 2014, **140**, 18A301.

(238) D. M. Ceperley and B. J. Alder, *Phys. Rev. Lett.*, 1980, **45**, 566.

(239) K. I. Ramachandran, G. Deepa and K. Namboori, *Computational chemistry and molecular modeling: principles and applications*, Springer, Heidelberg, 2008.

(240) J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865.

(241) K. A. Peterson, D. Feller and D. A. Dixon, *Theor. Chem. Acc.*, 2012, **131**, 1–20.

(242) N. Mardirossian and M. Head-Gordon, *Mol. Phys.*, 2017, **115**, 2315–2372.

(243) L. Goerigk, H. Kruse and S. Grimme, *ChemPhysChem*, 2011, **12**, 3421–3433.

(244) A. Karton and P. R. Spackman, *J. Comput. Chem.*, 2021, **42**, 1590–1601.

(245) A. Sorouri, W. M. Foulkes and N. D. Hine, *J. Chem. Phys.*, 2006, **124**, 064105.

(246) I. Carnimeo, S. Baroni and P. Giannozzi, *Electron. Struct.*, 2019, **1**, 015009.

(247) P. Broqvist, A. Alkauskas and A. Pasquarello, *Phys. Rev. B.*, 2009, **80**, 085114.

(248) R. Kaschner and D. Hohl, *J. Phys. Chem. A*, 1998, **102**, 5111–5116.

(249) K. S. Thanthiriwatte, E. G. Hohenstein, L. A. Burns and C. D. Sherrill, *J. Chem. Theory Comput.*, 2010, **7**, 88–96.

(250) J. Ireta, J. Neugebauer and M. Scheffler, *J. Phys. Chem. A*, 2004, **108**, 5692–5698.

(251) F. Bloch, *Zeitschrift für physik*, 1929, **52**, 555–600.

(252) M. D. Segall, P. J. D. Lindan, M. J. Probert, C. J. Pickard, P. J. Hasnip, S. J. Clark and M. C. Payne, *J. Phys. Condens. Matter.*, 2002, **14**, 2717–2744.

(253) H. Kruse, L. Goerigk and S. Grimme, *J. Org. Chem.*, 2012, **77**, 10824–10834.

(254) A. R. Leach, *Molecular Modelling: Principles and Applications*, Addision Wesley Longman, Essex, 2nd edn, 1997.

(255) H. J. Monkhorst and J. D. Pack, *Phys. Rev. B.*, 1976, **13**, 5188–5192.

(256) D. Marx and J. Hutter, *Ab initio molecular dynamics : basic theory and advanced methods*, Cambridge University Press, Cambridge, 2009.

(257) R. P. Feynman, *Phys. Rev.*, 1939, **56**, 340–343.

(258) P. Pulay, *Mol. Phys.*, 1969, **17**, 197–204.

(259) S. Grimme, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2011, **1**, 211–228.

(260) S. Grimme, *J. Comput. Chem.*, 2004, **25**, 1463–1473.

(261) A. Tkatchenko and M. Scheffler, *Phys. Rev. Lett.*, 2009, **102**, 73005.

(262) S. Grimme, A. Hansen, J. G. Brandenburg and C. Bannwarth, *Chem. Rev.*, 2016, **116**, 5105–5154.

(263) P. Jurečka, J. Šponer, J. Černý and P. Hobza, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1985–1993.

(264) S. Grimme, *Chem. Eur. J.*, 2012, **18**, 9955–9964.

(265) W. A. Al-Saidi, V. K. Voora and K. D. Jordan, *J. Chem. Theory Comput.*, 2012, **8**, 1503–1513.

(266) Y. Cho, W. J. Cho, I. S. Youn, G. Lee, N. J. Singh and K. S. Kim, *Acc. Chem. Res.*, 2014, **47**, 3321–3330.

(267) L. Kronik and A. Tkatchenko, *Acc. Chem. Res.*, 2014, **47**, 3208–3216.

(268) K. Momma, Y. Mishima, W. Hashimoto, B. Mikami and K. Murata, *Biochemistry*, 2005, **44**, 5053–5064.

(269) J. Binns, M. R. Healy, S. Parsons and C. A. Morrison, *Acta. Crystallogr. B*, 2014, **70**, 259–267.

(270) R. S. Mulliken, *J. Chem. Phys.*, 1955, **23**, 1833–1840.

(271) D. Sanchez-Portal, E. Artacho and J. M. Soler, *Solid State Commun.*, 1995, **95**, 685–690.

(272) M. D. Segall, C. J. Pickard, R. Shah and M. C. Payne, *Mol. Phys.*, 1996, **89**, 571–577.

(273) W. C. Swope, H. C. Andersen, P. H. Berens and K. R. Wilson, *J. Chem. Phys.*, 1982, **76**, 637–649.

(274) W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.

(275) J. L. Banks, H. S. Beard, Y. Cao, A. E. Cho, W. Damm, R. Farid, A. K. Felts, T. A. Halgren, D. T. Mainz, J. R. Maple, R. Murphy, D. M. Philipp, M. P. Repasky, L. Y. Zhang, B. J. Berne, R. A. Friesner, E. Gallicchio and R. M. Levy, *J. Comput. Chem*, 2005, **26**, 1752–1780.

(276) W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 2002, **117**, 5179–5197.

(277) H. J. Berendsen, J. P. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1998, **81**, 3684.

(278) D. J. Evans and O. P. Morriss, *Comput. Phys. Rep.*, 1984, **1**, 297–343.

(279) S. A. Adelman and J. D. Doll, *J. Chem. Phys*, 1976, **64**, 2375.

(280) S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. Probert, K. Refson and M. C. Payne, *Z. Kristallogr.*, 2005, **220**, 567–570.

(281) D. Vanderbilt, *Physical Review B*, 1990, **41**, 7892–7895.

(282) I. Todorov, W. Smith, K. Trachenko and M. Dove, *J. Mater. Chem.*, 2006, **16**, 1911–1918.

(283) H. C. Andersen, *J. Comput. Phys.*, 1983, **52**, 24–34.

(284) A. Daina, O. Michielin and V. Zoete, *J. Chem. Inf. Model.*, 2014, **54**, 3284–3301.

(285) G. J. Ackland, *J. Phys. Condens. Matter.*, 2002, **14**, 2975–3000.

(286) A. Daina, O. Michielin and V. Zoete, *Sci. Rep.*, 2017, **7**, 42717.

(287) M. Kolpen, C. R. Hansen, T. Bjarnsholt, C. Moser, L. D. Christensen, M. van Gennip, O. Ciofu, L. Mandsberg, A. Kharazmi, G. Döring, M. Givskov, N. Høiby and P. Ø. Jensen, *Thorax*, 2010, **65**, 57–62.

(288) S. S. Yoon, R. F. Hennigan, G. M. Hilliard, U. A. Ochsner, K. Parvatiyar, M. C. Kamani, H. L. Allen, T. R. DeKievit, P. R. Gardner, U. Schwab, J. J. Rowe, B. H. Iglewski, T. R. McDermott, R. P. Mason, D. J. Wozniak, R. E. Hancock, M. R. Parsek, T. L. Noah, R. C. Boucher and D. J. Hassett, *Dev. Cell*, 2002, **3**, 593–603.

(289) E. S. Cowley, S. H. Kopf, A. LaRiviere, W. Ziebis and D. K. Newman, *mBio*, 2015, **6**, e00767–15.

(290) M. S. Son, W. J. Matthews Jr, Y. Kang, D. T. Nguyen and T. T. Hoang, *Infect. Immun.*, 2007, **75**, 5313–5324.

(291) J. R. Lawrence, D. R. Korber, B. D. Hoyle, J. W. Costerton and D. E. Caldwell, *J. Bacteriol.*, 1991, **173**, 6558–6567.

(292) P. S. Belton, *Int. J. Biol. Macromol*, 1997, **21**, 81–88.

(293) M. L. Dheu-Andries and S. Pérez, *Carbohydr. Res.*, 1983, **124**, 324–332.

(294) L. Alagna, T. Prosperi, A. A. Tomlinson and R. Rizzo, *J. Phys. Chem.*, 1986, **90**, 6853–6857.

(295) I. Braccini, R. P. Grasso and S. Pérez, *Carbohydr. Res.*, 1999, **317**, 119–130.

(296) D. Cozzi, P. G. Desideri, L. Lepri and G. Ciantelli, *J. Chromatogr. A*, 1968, **35**, 405–415.

(297) J. L. Geddie and I. Sutherland, *J. appl. bacteriol.*, 1993, **74**, 467–472.

(298) A. Haug and O. Smidsrød, *Acta. Chem. Scand.*, 1965, **19**, 341–351.

(299) Y. Xiang, Y. Liu, B. Mi and Y. Leng, *Langmuir*, 2014, **30**, 9098–9106.

(300) P. Agulhon, M. Robitzer, L. David and F. Quignard, *Biomacromolecules*, 2012, **13**, 215–220.

(301) G. T. Grant, E. R. Morris, D. A. Rees, P. J. Smith and D. Thom, *FEBS. lett.*, 1973, **32**, 195–198.

(302) F. Topuz, A. Henke, W. Richtering and J. Groll, *Soft Matter*, 2012, **8**, 4877–4881.

(303) R. D. Shannon and C. T. Prewitt, *Acta Crystallogr. B Struct. Sci. Cryst. Eng. Mater.*, 1969, **25**, 925–946.

(304) N. K. Sachan, S. Pushkar, A. Jha and A. Bhattcharya, *J. Pharm. Res*, 2009, **2**, 1191–1199.

(305) M. H. Turakhia and W. G. Characklis, *Biotechnol. Bioeng.*, 1989, **33**, 406–414.

(306) C. W. Yong, *J. Chem. Inf. Model.*, 2016, **56**, 1405–1409.

(307) U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1998, **103**, 8577.

(308) M. Ritenberg, S. Nandi, S. Kolusheva, R. Dandela, M. M. Meijler and R. Jelinek, *ACS Chem. Biol.*, 2016, **11**, 1265–1270.

(309) R. C. Hunter and T. J. Beveridge, *J. Bacteriol.*, 2005, **187**, 7619.

(310) P. Sikorski, F. Mo, G. Skjåk-Bræk and B. T. Stokke, *Biomacromolecules*, 2007, **8**, 2098–2103.

(311) O. J. Hills, C. W. Yong, A. J. Scott, D. A. Devine, J. Smith and H. F. Chappell, *Sci. Rep.*, 2022, **12**, 7724.

(312) H. Hecht and S. Srebnik, *Biomacromolecules*, 2016, **17**, 2160–2167.

(313) L. Mashburn-Warren, J. Howe, P. Garidel, W. Richter, F. Steiniger, M. Roessle, K. Brandenburg and M. Whiteley, *Mol. Microbiol.*, 2008, **69**, 491–502.

(314) S. Mecozzi, A. P. West and D. A. Dougherty, *Proc. Natl. Acad. Sci. U.S.A.*, 1996, **93**, 10566–10571.

(315) H. C. Flemming, J. Wingender, U. Szewzyk, P. Steinberg, S. A. Rice and S. Kjelleberg, *Nat. Rev. Microbiol.*, 2016, **14**, 563–575.

(316) C. H. Tan, H. S. Oh, V. M. Sheraton, E. Mancini, S. C. Joachim Loo, S. Kjelleberg, P. M. Sloot and S. A. Rice, *Environ. Sci. Technol.*, 2020, **54**, 6730–6740.

(317) J. Sankaran, N. J. Tan, K. P. But, Y. Cohen, S. A. Rice and T. Wohland, *NPJ Biofilms Microbiomes.*, 2019, **5**, 1–10.

(318) R. G. Pearson, *J. Chem. Sci.*, 2005, **117**, 369–377.

(319) R. G. Pearson, *Acc. Chem. Res.*, 1993, **26**, 250–255.

(320) P. K. Chattaraj and B. Maiti, *J. Am. Chem. Soc.*, 2003, **125**, 2705–2710.

(321) R. G. Pearson, *J. Chem. Educ.*, 1987, **64**, 561.

(322) R. G. Parr, L. V. Szentpály and S. Liu, *J. Am. Chem. Soc.*, 1999, **121**, 1922–1924.

(323) P. K. Chattaraj, U. Sarkar and D. R. Roy, *Chem. Rev.*, 2006, **106**, 2065–2091.

(324) R. A. Miranda-Quintana and P. W. Ayers, *J. Chem. Phys.*, 2018, **148**, 196101.

(325) C. Morell, V. Labet, A. Grand and H. Chermette, *Phys. Chem. Chem. Phys.*, 2009, **11**, 3417–3423.

(326) F. Zsila, M. Ricci, I. C. Szigyártó, P. Singh and T. Beke-Somfai, *Front. mol. biosci.*, 2021, **8**, 899.

(327) A. Ilangovan, M. Fletcher, G. Rampioni, C. Pustelny, K. Rumbaugh, S. Heeb, M. Cámara, A. Truman, S. R. Chhabra, J. Emsley and P. Williams, *PLoS Pathog.*, 2013, **9**, e1003508.

(328) N. F. Baig, S. J. Dunham, N. Morales-Soto, J. D. Shrout, J. V. Sweedler and P. W. Bohn, *Analyst*, 2015, **140**, 6544–6552.

(329) N. Morales-Soto, S. J. Dunham, N. F. Baig, J. F. Ellis, C. S. Madukoma, P. W. Bohn, J. V. Sweedler and J. D. Shrout, *J. Biol. Chem.*, 2018, **293**, 9544–9552.

(330) N. Morales-Soto, T. Cao, N. F. Baig, K. M. Kramer, P. W. Bohn and J. D. Shrout, *Microbiol. Insights.*, 2018, **11**, 1178636118817738.

(331) T. Cao, A. A. Weaver, S. Baek, J. Jia, J. D. Shrout and P. W. Bohn, *J. Chem. Phys.*, 2021, **154**, 204201.

(332) J. L. Bru, B. Rawson, C. Trinh, K. Whiteson, N. M. Høyland-Kroghsbo and A. Siryaporn, *J. Bacteriol.*, 2019, **201**, e00762–17.

(333) P. Sriamornsak and R. A. Kennedy, *Int. J. Pharm.*, 2008, **358**, 205–213.

(334) E. Paluch, J. Rewak-Soroczyńska, I. Jędrusik, E. Mazurkiewicz and K. Jermakow, *Appl. Microbiol. Biotechnol.*, 2020, **104**, 1871–1881.

(335) C. Pustelny, A. Albers, K. Büldt-Karentzopoulos, K. Parschat, S. R. Chhabra, M. Cámara, P. Williams and S. Fetzner, *Chem. Biol.*, 2009, **16**, 1259–1267.

(336) A. Arranz San Martín, J. Vogel, S. C. Wullich, W. J. Quax and S. Fetzner, *Biomolecules*, 2022, **12**, 170.

(337) L. M. Mashburn and M. Whiteley, *Nature*, 2005, **437**, 422–425.

(338) C. A. Stortz, G. P. Johnson, A. D. French and G. I. Csonka, *Carbohydr. Res*, 2009, **344**, 2217–2228.

(339) M. F. Pritchard, L. C. Powell, A. A. Jack, K. Powell, K. Beck, H. Florance, J. Forton, P. D. Rye, A. Dessen, K. E. Hill and D. W. Thomas, *Antimicrob. Agents Chemother.*, 2017, **61**, e00762–17.

(340) Z. J. Li, S. Srebnik and O. J. Rojas, *Biomacromolecules*, 2021, **22**, 4027–4036.

(341) *CrystalMaker*, http://www.crystalmaker.com, (accessed October 2019).

(342) H. Eyring, *J. Chem. Phys.*, 1935, **3**, 107–115.

(343) M. G. Evans and M. Polanyi, *Trans. Faraday Soc.*, 1935, **31**, 875–894.

(344) M. G. Best, C. Cunha-Reis, A. Y. Ganin, A. Sousa, J. Johnston, A. L. Oliveira, D. G. Smith, H. H. Yiu and I. R. Cooper, *ACS Appl. Bio Mater.*, 2020, **3**, 7589–7597.

(345) B. Li, Y. Dong, C. Zou and Y. Xu, *Ind. Eng. Chem. Res.*, 2014, **53**, 4199–4206.

(346) N. Schönberger, C. Taylor, M. Schrader, B. Drobot, S. Matys, F. L. Lederer and K. Pollmann, *J. Hazard. Mater.*, 2021, **414**, 125366.

(347) V. Mouriño, P. Newby and A. R. Boccaccini, *Adv. Eng. Mater.*, 2010, **12**, B283–B291.

(348) M. A. Akhtar, Z. Hadzhieva, K. Ilyas, M. S. Ali, W. Peukert and A. R. Boccaccini, *Pharmaceutics*, 2021, **13**, 1702.

(349) M. A. Akhtar, Z. Hadzhieva, I. Dlouhý and A. R. Boccaccini, *Coatings*, 2020, **10**, 483.

(350) M. K. Dowd, A. D. French and P. J. Reilly, *Carbohydr Res*, 1994, **264**, 1–19.

(351) F. Porcaro, C. Bonchi, A. Ugolini, E. Frangipani, G. Polzonetti, P. Visca, C. Meneghini and C. Battocchio, *Dalton Trans.*, 2017, **46**, 7082–7091.

(352) K. Schlegel, J. Lex, K. Taraz and H. Budzikiewicz, *Z. Naturforsch. C*, 2006, **61**, 263–266.

(353) T. Huyen Vu, N. Serradji, M. Seydou, É. Brémond and N. Ha-Duong, *J. Inorg. Biochem.*, 2020, **203**, 110864.

(354) E. Halevas, B. Mavroidi, O. Antonoglou, A. Hatzidimitriou, M. Sagnou, A. A. Pantazaki, G. Litsardakis and M. Pelecanou, *Dalton Trans.*, 2020, **49**, 2734–2746.

(355) A. Munteanu, M. Badea, R. Olar, L. Silvestro, M. Mihaila, L. I. Brasoveanu, M. G. Musat, A. Andries and V. Uivarosi, *Appl. Organomet. Chem.*, 2018, **32**, e4579.

(356) S. Hijazi, D. Visaggio, M. Pirolo, E. Frangipani, L. Bernstein and P. Visca, *Front. Cell. Infect. Microbiol.*, 2018, **8**, 316.

(357) D. Visaggio, E. Frangipani, S. Hijazi, M. Pirolo, L. Leoni, G. Rampioni, F. Imperi, L. Bernstein, R. Sorrentino, F. Ungaro and P. Visca, *ACS Infect. Dis.*, 2022, **8**, 78–85.

(358) I. Braccini and S. Pérez, *Biomacromolecules*, 2001, **2**, 1089–1096.

(359) T. Das, S. Sehar, L. Koop, Y. K. Wong, S. Ahmed, K. S. Siddiqui and M. Manefield, *PloS ONE*, 2014, **9**, e91935.

(360) M. E. Boursier, J. D. Moore, K. M. Heitman, S. P. Shepardson-Fungairino, J. B. Combs, L. C. Koenig, D. Shin, E. C. Brown, R. Nagarajan and H. E. Blackwell, *ACS Chem. Biol.*, 2018, **13**, 2655–2662.

(361) A. M. Silva, X. Kong, M. C. Parkin, R. Cammack and R. C. Hider, *Dalton Trans.*, 2009, 8616–8625.

(362) F. Biaso, C. Duboc, B. Barbara, G. Serratrice, F. Thomas, D. Charapoff and C. Béguin, *Eur. J. Inorg. Chem.*, 2005, **2005**, 467–478.

(363) D. J. Thornton, K. Rousseau and M. A. McGuckin, *Annu. Rev. Physiol.*, 2008, **70**, 459–486.

(364) R. M. Landry, D. An, J. T. Hupp, P. K. Singh and M. R. Parsek, *Mol. Microbiol.*, 2006, **59**, 142–151.

(365) P. Ren and J. W. Ponder, *J. Comput. Chem.*, 2002, **23**, 1497–1506.

(366) L. Pinzi and G. Rastelli, *Int. J. Mol. Sci.*, 2019, **20**, 4331.

(367) P. H. Torres, A. C. Sodero, P. Jofily and F. P. Silva Jr., *Int. J. Mol. Sci.*, 2019, **20**, 4574.

(368) J. Fan, A. Fu and L. Zhang, *Quant. Biol.*, 2019, **7**, 83–89.