Polymer Structure In Solution For Drug Delivery From Simulation and Experiment

Amnah Hadadi

A thesis submitted to

Cardiff University

in accordance with the requirements for the degree of

Doctor of Philosophy

School of Chemistry

Cardiff University

United Kingdom

May 2024

Acknowledgements

From the principle that whoever does not thank people does not thank God, I would like to take the opportunity to express my gratitude to everyone who has helped or made any contribution to this thesis.

I would like to start by sincerely thanking my supervisors Professor Jamie Platts and Doctor Alison Paul for allowing me to join their research groups and for all of their help and advice during my degree. Without their guides, help, support and efforts, nothing of this work would have been possible.

A huge thank to Doctor Loizos Savva sharing his expertise so willingly and helping at the beginning of my journey, where it was the challenging stage.

To Doctor Tom Headen and Doctor Tristan Youngs at the Rutherford Appleton Laboratory UK, with whom I collaborated with for the neutron experiments (NIMROD) and Dissolve software. They provided a great deal of assistance, direction, and advice from the beginning until the end of the project.

I would also like to thank my mentors, Doctor Benjamin Ward and Doctor Michaela Serpi, for all of the useful feedback they have provided throughout the various stages of my PhD. Doctor Benjamin also provided the help and guidance in the analysis of polymer tacticity. To Doctor Elaine Ferguson at Cardiff Dental School, who attributed in the GPC measurement.

Thanks should also go the Supercomputing of Wales for the access to do the calculations, and to Jose Munoz Criollo for his technical advice. To the Science and Technology Facilities Council for the access to the ISIS Data Analysis as a Service, where all Gudrun and Dissolve software's data analysis were done.

I am eternally grateful to the great love and support from my family, husband, kids, mother, sisters and brothers. To the soul of my father, who was encouraging my desire to continue acquiring knowledge.

Finally, I would never be here were it not for the support and generosity of the government of Saudi Arabia; including Umm Al-Qura University, where I am currently employed.

Abstract

To achieve accurate determination of macromolecular structures, combination of experimental and computational efforts often requires. Wide-angle neutron scattering (WANS) opens up avenues to accurately measure the positions of atoms within a material, enabling the precise mapping of atomic structures. In this thesis, molecular dynamic simulations beside small-angle neutron scattering (SANS) and Near and InterMediate Range Order Diffractometer (NIMROD) at ISIS facility were used to investigate polyethylene glycol (PEG) and Polyvinylpyrrolidone (PVP) in dilute solutions. This involves validation of force fields (OPLS-AA, CHARMM, AMBER) for both (molecules and water) and timescale.

A good agreement in the radius of gyration value was achieved between the experimental SANS, $\approx 26 \pm 2.5$ Å, and MD, $\approx 22 \pm 3.2$ Å, data for PEG (13,000 g/mol) using OPLS-AA/TIP3P models. A further improvement in the radius of gyration, $\approx 25 \pm 6$ Å, was obtained using a simulation box with similar experimental concentration. However, the discrepancy of the radius of gyration value for PVP (10,000 g/mol) between the SANS, $\approx 19 \pm 2.5$ Å, and MD, $\approx 14 \pm 0.5$ Å data was observed even with the simulation box applied the experimental concentration.

This resulted in the last chapter of this thesis using the SANS-driven MD methods, which can overcome the inaccuracies of the force fields and account for the hydration layer that affects the radius of gyration value. A better SANS fit was obtained for PVP than for PEG. This is due to the chain flexibility movement for PEG, which demanded a long simulation period to investigate several conformations while simultaneously calculating and fitting the scattering curve into the experimental data. The higher number of water molecules in the solvation shell of PEG indicates that it is more soluble in water than PVP. The application of the SANS-driven MD method confirms that the hydration layer can indeed have an impact on the radius of gyrations, particularly for PVP.

Using MD simulation in the Dissolve program, the total neutron scattering (NIMROD) for the both mentioned polymers was examined. This study represents the initial endeavor to utilize the Dissolve to determine highly localised structure of polymers in solution from the total neutron scattering data. After using the Empirical Potential Structural Refinement method (EPSR) and adjusting the water geometry, a good agreement was obtained between the

calculated and experimental total weighted neutron structure F(q) and weighted neutron radial distribution function G(r), especially for the deuterated solvent samples. The position of atoms within a material, including bond lengths and angles between atoms, can be determined using Dissolve. However, Dissolve is currently under development and requires improvements in various areas, particularly the timestep.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

General introduction

The primary objective of this study is to pave the way to develop methods that enable the accurate analysis of polymer structure and dynamics in dilute solutions by combining experimental neutron scattering (low and high resolutions) and molecular dynamic simulations. The main goal of current structural biophysics is to understand how molecules undergo dynamic reorganization to exhibit their activities. Frequently, acquiring structures of molecules and ensembles necessitates integrating data from multiple experiments. However, the experimental data may not contain enough information to fully determine all the important characteristics of the molecule being studied. Hence, to generate a set of molecular structures without overfitting, a combination of experimental data and computational sampling is required. This necessitates validating many aspects of computational processes such as force fields, solvents, and timescales.

For small-angle neutron scattering (SANS) data, a comparison was made with sampling data to validate the modelling approach for determining solution conformation using GROMACS, and this is done by two different methods: fitting the SANS curves for the free MD (chapter 3 for PEG and chapter 4 for PVP) and fitting the restraint-driven SANS MD that takes into account the solvation shell (chapter. 5). For wide-angle neutron scattering (NIMROD) data, a comparison was made between the experimental and calculated weighted total structural factor F(q) and the total weighted radial distribution function G(r) using Dissolve software that implemented the Empirical Potential Structural Refinement (EPSR) method (chapters 3 and 4).

The study focused on examining two categories of polymers in solutions: (a) a simple polymer chain (polyethylene glycol) that consists of only an ether monomer as a backbone; and (b) a polymer chain that contains a vinyl backbone with a side chain of pyrrolidone (poly-vinylpyrroldinone).

The thesis is organized as follows: Chapter 1. provides background information prior to discussing the methods in Chapter 2 and the details of specific results in Chapters 3, 4, and 5.

1.1 The Importance of Polymer Study

The history of humanity has been greatly impacted by the abundance of resources. History is categorized into several periods known as eras, each called after the predominant materials utilized at that time: the Stone Age, the Bronze Age, and the Iron Age. In a similar vein, we might declare that the twentieth century marked the beginning of the Polymer Age.[1]

Polymers are commonly used advanced substances that are present in nearly every item employed in our daily life. The word "polymer" comes from classical Greek, where "poly" means "many" and "meres" means "parts". Polymers are a diverse group of materials, natural and synthetic, composed of several monomers, which are small molecules that are connected together to create long chains. A covalent bond typically holds multiple structural units of the polymer molecule together, giving it a high molecular weight. Polymers can be constructed from a wide variety of monomers, each of which has special qualities that are helpful in various contexts. Polymers are synthesized by chemically reacting monomers. Monomers have the capacity to undergo a chemical reaction with another molecule of the same or a different type under certain conditions, resulting in the formation of a polymer chain. Natural polymers are formed by natural process, while synthetic polymers are created by humans[2,3]

For many years, humans have been utilizing polymers in their daily lives, however their understanding of this material was limited until the end of World War II. There were a limited number of materials accessible for the production of the item required for a civilized existence. The construction materials mostly consisted of steel, glass, wood, stone, brick, and concrete. For clothing or fabric production, cotton, wood, jute, and a few other agricultural items were utilized.[4]

The introduction of new materials is caused by the sharp rise in demand for produced goods. These newly developed substances are polymers, and their influence on the current way of life is considerable. Polymers are ubiquitous in our daily lives, manifesting in various forms such as synthetic fiber clothes, polyethylene cups, fiberglass, nylon bearings, plastic bags, polymer-based paints, epoxy glue, polyurethane foam cushions, silicone heart valves, and Teflon-coated cookware. The list is extensive.[4]

Materials made of polymers have a variety of chemical, mechanical, and physical characteristics. These characteristics have sparked a great deal of investigation and

advancement into the possible uses of polymeric biomaterials. In recent years, there have been considerable advancements in the realm of biomedical applications regarding the development of polymeric materials that possess the properties of being biodegradable and biocompatible.[5] Polymeric biomaterials are extensively utilized in the field of medicine, making them the most commonly employed materials among synthetic and hybrid options.[6]

Polymers exhibit advanced functionality, leading to unique features used in several fields such as drug delivery,[7] tissue engineering,[8] super-soft materials,[9] and antibacterial surfaces[10] etc. Thus, it is unsurprising that the majority of material scientists and many chemists and chemical engineers, as well as a significant number of physicists, textile technologists, mechanical engineers, and other scientific professionals, are engaged in research and development attempts related to polymers.[4,11]

1.2 Polymers used in drug delivery

Polymers are becoming essential components of drug delivery systems because of their enhanced pharmacokinetic characteristics.[12,13] They possess superior circulation time compared to traditional small molecules, allowing them to target tissues more precisely. Significant use of polymers has been observed in the field of polymer treatments and nano medicines.[14] Reservoir-based drug delivery systems have made significant advancements through the use of polymers, particularly in the development of hydrogels and liposomes. Polymers are being investigated for use in diffusion-based drug delivery systems and solvent-activated drug delivery systems. When exposed to an aqueous environment, solvent-activated systems such as hydrogels swell and release the medication. They are inherently hydrophilic.[15] Biocompatible polymers provide a safe way for medication distribution because of their carefully designed molecular structure that aligns with the changes in the biological process mechanisms. Biodegradable polymers are broken down by cleaving covalent connections, while bioerodible polymers erode by the dissolution of connecting chains without altering the chemical structure of the molecule. Polymers used in vivo as drug carriers must be water-soluble, non-toxic, and non-immunogenic. They passively minimize medication breakdown and enhance circulation time. Another crucial concern is the safe elimination of the medication. For nondegradable polymers, it is important to prevent accumulation in the body. Degradable polymers should break down into components that are below the renal threshold level, non-toxic, and do not trigger an immune response. Polymers that imitate biological systems can

react to external stimuli like pH or temperature, leading to changes in properties such as solubility, hydrophobic/hydrophilic balance, biomolecule release, and conformation.[12,14]

Traditional drug delivery systems administer drug doses in the form of capsules or tablets created through compression, coating, and encapsulation of bioactive drug molecules. Polymers have several functions in traditional formulations, acting as binding agents in capsules, film coating agents in tablets, and viscosity enhancers in emulsions and suspensions. Polymers such as cellulose derivatives, poly (N-vinyl pyrrolidone), and poly (ethylene glycol) PEG are commonly used with bioactive pharmacological compounds.

1.3 Systems studied

To examine the all-atom model for polymer solutions based on the experimental neutron scattering data (SANS and WANS), we selected the simplest and well-studied polymers, PEG and PVP in order to proof the method for further new complex systems.

1.3.1 Polyethylene glycol (PEG)

Polyethylene glycol (PEG) is a multipurpose polyether that finds use in many fields, including chemical industries and medicine. PEG is also known as polyethylene oxide (PEO). Ethylene oxide macromolecules with molecular weights below 20,000 g/mol with hydroxyl endings are referred to as PEG, while those with weights exceeding 20,000 g/mol are called PEO. It is a linear synthetic polyether comprised of a nonpolar ethylene group (C_2H_4) and a polar oxygen atom (Figure. 1.1) that can be synthesized from epoxyethane via anionic polymerization or by ring-opening polymerization.[16–18]

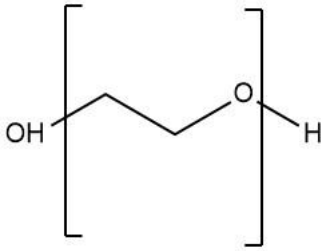


Figure 1. 1: The monomer structure of poly(ethylene glycol) (PEG)

Polyethylene glycol is biocompatible, synthetic, non-toxic, water-soluble, and highly flexible; these properties make PEG suitable for a wide range of applications especially in medical and pharmaceutical industry.[19] Since the 1950s, PEGs have been extensively studied and utilized

for many purposes such as separation and purification aids, matrices for embedding, anti-freeze agents, lubricants for medical devices, food additives, and as vehicles in dermatological applications, suppositories, parenteral, tablets, and pills. Colloidal science has revived PEG as a traditional steric stabilizer that can evade the host's immune defense mechanism. Abuchowski and Davis established the concept of PEGylation in the late 1970s for delivering proteins and medicines.[20] The initial PEGylated conjugate, Adagen, was launched in 1990 for the treatment of immunodeficiency illness and achieved clinical success.[21] Human safety is a crucial factor to be taken into account while developing excipients and final formulations. Although alternatives such sodium alginate, dextran, and poloxamers have been developed in the field of polymers, PEG remains the preferred polymer for many applications.[16]

1.3.2 Polyvinylpyrrolidone (PVP)

Polyvinylpyrrolidone (PVP), sometimes known as polyvidone or povidone, is a water-soluble polymer derived from the monomer N-vinylpyrrolidone, as shown in Figure 1.2. Dry PVP is a delicate, powdery substance that easily absorbs water, taking in up to 40% of its weight. When dissolved, it has superb wetting characteristics and easily creates films, making it ideal for use as a coating or an additive in coatings. Polyvinylpyrrolidone (PVP) can be synthesized using free-radical polymerization using its monomer N-vinylpyrrolidone and azobisisobutyronitrile (AIBN) as an initiator.[22–24]

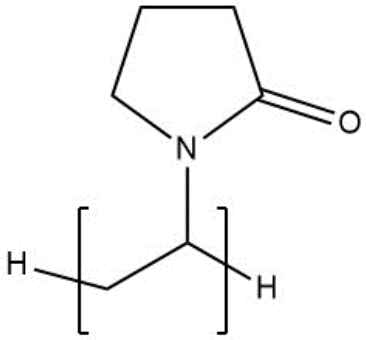


Figure 1. 2: The monomer structure of poly(vinylpyrrolidone) (PVP)

Walter Reppe, a chemist from BASF, discovered PVP in 1938 as one of the many compounds resulting from acetylene chemistry.[23] The monomer N-vinylpyrrolidone was synthesized by reacting acetylene with formaldehyde to produce 1,4-butyne diol, which was subsequently hydrogenated to butanediol. Butyrolactone is produced through oxidative cyclization. When 1,4-butyne diol reacts with ammonia and water is removed, pyrrolidone is generated.

PVP has an exceptional combination of physical and chemical characteristics, such as biocompatibility, non-toxicity, chemical stability, good solubility in water and various organic solvents, and the ability to interact with both hydrophobic and hydrophilic substances, have rendered it suitable for use as a biomaterial in a wide range of important medical and non-medical fields, including the pharmaceutical industry, medicine, optical and electrical, membranes, adhesives, ceramics, paper, coatings and inks, household, industrial, and institutional, lithography and photography, fibers and textiles, and environmental applications.[25,26]

PVP has been utilized in the pharmaceutical and biomedical sectors to provide several drug delivery systems, including oral, topical, transdermal, and ocular administration.[27] Additionally, PVP can be utilized for gene delivery[28–32] or combined with metal particles for regenerative medicine[33–35] and targeted -delivery[36,37] Thus, PVP proves to be a highly adaptable polymer. Diverse morphologies utilizing PVP as the polymeric carrier have been suggested for drug delivery.[27] PVP facilitates achieving a regulated medication release, enhancing the absorption of pharmaceuticals with limited water solubility, safeguarding the active ingredient from environmental factors including pH, temperature, and oxygen, and concealing undesirable odors and tastes. Various active chemicals from diverse categories have been integrated into PVP microparticles and nanoparticles. PVP-based particles have been created using many methods, ranging from conventional procedures like spray drying to more advanced methods such as supercritical fluids-assisted techniques.[38–42] PVP has also been used to create fibers that are packed with different active components.[43–47] Some articles have suggested the use of PVP-based hydrogels and oral tablets.[48–51] Moreover, PVP demonstrates outstanding film-forming capabilities.[52–56]

PVP's versatility and unique features make it a polymer with significant potential for pharmaceutical formulation manufacture. The utilization of PVP in the developing pharmaceutical and biomedical sectors can provide beneficial solutions that are still needed in the pharmaceutical market.

1.4 Polymer characterization

Understanding a material's physical properties is frequently necessary for designing polymers for particular uses. The characteristics of polymeric systems are mostly determined by the chemical identity of the monomers. The microstructure of the polymer, or how the atoms are arranged along the chain that is established during the polymerization process, is another important component. Polymers can be produced using several techniques, including ring-opening polymerization, ionic polymerization, chain-growth polymerization, and radical polymerization. Catalysts are essential in polymer synthesis as they have a significant impact on the speed of polymerization, the composition of the final polymer, its stereochemistry, and the distribution of molecular weights. In a vinyl polymer, if all the carbon atoms forming the backbone are arranged in a zig-zag manner within a single plane, the adjacent monomers can have their R group positioned either on the same side or on opposite sides of this plane. The polymer's tacticity describes the diversity in stereoisomerism of this kind. A vinyl polymer is considered isotactic when all of its R groups are positioned on the same side of the chain. Conversely, if the R groups alternate in a regular pattern, the polymer is considered syndiotactic. An alternative scenario is that the positioning of the R groups is entirely arbitrary, resulting in a polymer that is considered atactic.[1] Further elucidation on the polymer tacticity can be located in section 2.14.1.

Materials formed from polymeric chains can exist in two forms: highly organized, known as crystalline, or highly disordered, known as amorphous. Polymer solids exhibit a limited occurrence of complete crystallinity or amorphousness, with partly crystalline characteristics being frequently encountered. The molecular arrangement has a profound impact on the physical characteristics, and examining the structure is frequently a crucial aspect of analyzing any polymer system.[57]

1.5 Polymer conformation in solution

Polymer conformation refers to the ability of polymers to assume many conformations in a solution, such as random coils, stretched chains, and folded structures. The configuration of the polymer chains is influenced by variables such as the quality of the solvent, temperature, and molecular weight. The interaction of the polymer atoms with solvent determines the conformation of polymers in solution.[1] In the absence of ionic groups, the solvent quality is classified as either good, theta, or poor[58] (see Figure. 1.3). Polymer-solvent interactions exhibit

greater strength than polymer-polymer interactions in good solvents. In solvents with a theta condition, the interactions between polymers are equivalent to the interactions between polymers and solvents. In theta solvents, flexible polymers adopt a Gaussian coil shape, while in good solvents, they swell. When placed in a poor solvent, flexible chains of the polymer collapse because the interactions between polymer molecules are more powerful than the interactions between the polymer and the solvent.[1]

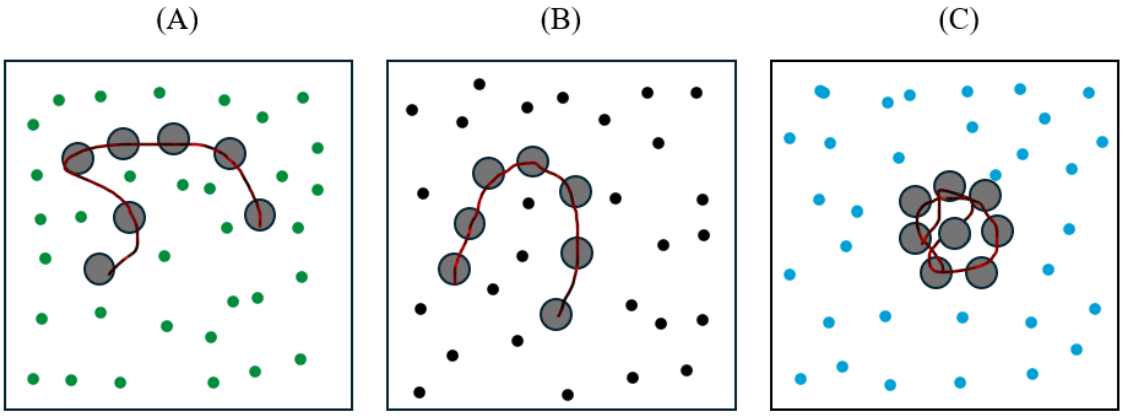


Figure 1. 3: Schematics of a flexible polymer chain in a (A) good, (B) theta and (C) poor solvents. green: good solvent molecules, black: theta solvent molecules, blue: poor solvent molecules, gray: monomers in polymer chain.

The degree of interaction between the polymer and solvent components is quantified using a simple but powerful mathematical model Flory-Huggins parameter, χ.[59,60] This theory is obtained by a straightforward lattice model, where each monomer is confined to a separate lattice site, and the same applies to solvent molecules. The parameter χ is expressed as:

$$\chi = \left(\frac{z}{k_B T}\right)\left[\varepsilon_{ms} - \frac{1}{2}(\varepsilon_{mm} + \varepsilon_{ss})\right] \qquad (1.1)$$

where z represents the number of nearest neighbors per unit monomer, k_B denotes the Boltzmann constant, and T signifies the temperature. The energy associated with the interactions between a monomer and a solvent, between two monomers, and between two solvents are represented by ε_{ms}, ε_{mm}, and ε_{ss} correspondingly.

When χ is positive, it indicates that the monomer-solvent interaction energies are larger than those of the monomer-monomer interaction. A negative χ indicates that the contacts between

the monomer and the solvent are stronger than the interactions between monomers, which facilitates the dissolution of the polymer.

The solvent quality determines the size of the polymer chain. Radius of gyration (R_g) is one way to quantify the size of a polymer.

$$R_g = R_0 N^{v} \qquad (1.2)$$

where N is the number of repeating units, v is a solvent factor, and R_0 is the monomer unit's radius of gyration. For the theta solvent, the scaling exponent v is $\frac{1}{2}$, while for good solvents is $\frac{3}{5}$, and the poor solvents is $\frac{1}{3}$. The Flory approximation in both solvents, implicit and explicit, were applied for PEG and PVP polymer chains (see chapters 3 and 4).

Statistical analysis of polymers focuses on universal parameters that define the size and shape of a macromolecule in a solvent at equilibrium[58,61] Many physical qualities are directly influenced by conformation, such as hydrodynamic properties of molecule fluids,[62] folding dynamics, and catalytic activity of molecules[63] etc. The size of a single macromolecule is often determined by measuring the mean square radius of gyration R_g^2 in static scattering experiments.[64–66] The radius of gyration is the distance from the center of mass of a body where the entire mass might be concentrated without altering its moment of rotational inertia around an axis passing through the center of mass. This refers to the root-mean-square distance of the segments of a polymer chain from its center of mass, given in equation 1.3.

$$R_g^2 = \frac{1}{N}\sum_{j=1}^{N}\left(r_j - r_{mean}\right)^2 \qquad (1.3)$$

Where N represents the number of repeating units, r is the centre position of each monomer, and r_{mean} denotes the average of these places. The radius of gyration measures the extent of the random coil form commonly adopted by synthetic polymers in solution or in the amorphous bulk state. A polymer's radius of gyration (R_g) in solution is determined by the macromolecule's molecular weight, structure (branched or linear), and solvent-induced swelling.[67] A more detailed discussion of polymer size and shape, including the radius of gyration data, can be found in Chapters 4 and 5 for PEG and PVP, respectively.

1.6 Studying Polymers in Liquid Phase

Crystallography utilizes the inherent propensity of molecules to arrange themselves in symmetrical patterns and maintain a structured organization. These units are usually included within a lattice of unit cells, where each unit cell represents a small volume that contains all the structural information needed to construct the overall picture when translated.[68] The primary method for acquiring structural information about biological macromolecules has therefore been through protein X-ray crystallography. The majority of the time, motion is caused by small temperature variations. In the case of liquids and indeed for highly flexible molecules in solution, however, the repetition of structure is limited to a maximum length of 1-2 nm,[69,70] which adds complexity to the task of acquiring a clear understanding of the structural characteristics. Therefore, the utilization of a unit cell technique is not possible. Hence it is more difficult to characterize the structural properties of polymers in liquid phase, as atoms are constantly moving about in liquids, there is no such thing as "structure" in the sense of crystallography. This means the least squares method applied to crystal structure analysis is less effective.

The radial distribution function, g(r), can be used to gain structural information in an experiment investigating liquid structure. Radial distribution functions (RDFs) quantify the spatial variation in particle density relative to a reference particle as a function of distance. The term "particles" in this context refers to individual atoms or molecules. The RDF, or radial distribution function, is a valuable tool for characterizing the typical arrangement of particles in disordered systems.[70] To calculate an RDF, one must create a sequence of concentric spheres around a designated particle, with each sphere being spaced apart by a distance Δr, as illustrated in Figure. 1.4. The number of particles is counted, and this process is then repeated for other particles in the system. By calculating the mean number of particles in each shell, denoted as n(r), it is feasible to generate a histogram illustrating the mean number of atoms detected within each volume segment as the distance increases. Every peak is characterized as a coordination shell, and the size of each peak corresponds to the number of particles present in that shell surrounding the reference particle. The final function is obtained by normalizing this by the product of the shell volume, $4\pi r^2\Delta r$, and the particle number density, ρ. The g(r) can be stated as:

$$g(r) = \frac{n(r)}{4\pi\rho r^2\Delta r} \qquad (1.4)$$

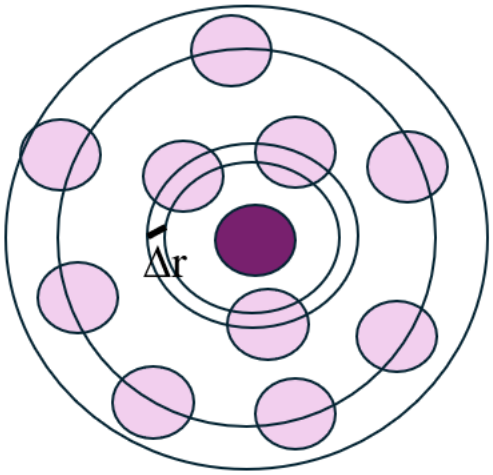


Figure 1. 4: The radial distribution function, g(r), which can be obtained by calculating the number of particles in each sphere within a distance Δr for a reference particle.

The RDF can indicate the most probable intermolecular separation between atoms/molecules, which provides a comprehensive understanding of the structure of a system when it is in a liquid condition. One can extract the function g(r) from experiments, e.g., diffraction data. Moreover, utilizing the g(r) function derived from experimental data allows for assessing the precision of molecular modeling.

1.7 Computational Techniques

Molecular modelling and simulations are highly valuable tools for the polymer science and engineering community. These computational methods allow for the prediction and interpretation of empirically observed characteristics of large molecules, including their structure, dynamics, thermodynamics, and microscopic and macroscopic material properties. Polymer simulations can inform, guide, and complement *in vitro* macromolecular materials design and discovery efforts in a synergistic way thanks to recent increases in computing power. In order to effectively utilize the increasing influence of simulations and obtain significant outcomes, it is crucial to test the accuracy and replicability of these simulations.[71]

Polymers and polymer solutions are a type of intricate fluids that pose distinct challenges to computational scientists. They display fascinating and significant phenomena across a wide range of sizes, starting from individual monomers (measured in Angstroms) to the overall size of the polymer (measured in nanometers). Additionally, these phenomena occur over various

time scales, ranging from femtoseconds to seconds or even years in the case of glasses. Several computational issues have been resolved through advancements in software, algorithms, and computer hardware, although not all of them have been tackled. The increasing number of polymer simulation studies, made possible by advancements in hardware and software, offer insightful analyses of both new and well-established macromolecular materials. Through forecasts, these studies encourage polymer chemists to discover synthetic pathways for the creation of highly promising new materials.[71]

1.7.1 Characterization of Single-Chain Behavior

In general, the majority of models for classical molecular simulations of polymers can be classified as either atomistic or coarse-grained. To represent the molecules in a system, one can use coarse-grain or all-atom models, depending on the particular study objective. Both models possess benefits and drawbacks in their utilization. Our particular research interests are in the study of polymer solutions using all-atom simulations, by employing atomistic models for investigating systems in which the chain configurations under specific conditions (such as temperature, pressure, and concentration) are already known beforehand from prior neutron experimentations.

1.7.1.1 Atomistic Models

All-atom model representations provide a precise and complete picture of a molecule's structure and dynamic behaviour. It is advisable to utilize atomistic models to analyze systems in which the chain configurations under specific conditions (such as temperature, pressure, and concentration) are predetermined through prior calculations or experiments.[71] The simulations offer useful insights into a range of physical, chemical, and biological phenomena. Atomistic simulations can provide insights into the structural properties of materials, essential for comprehending their structural behaviors. Moreover, it enables researchers to study the dynamic behavior of molecules and materials over time. This involves molecular movement, structural alterations, dispersion, and many dynamic activities. Models can offer insights into transport phenomena as diffusion coefficients, permeability, and conductivity. They may simulate the behavior of molecules in a solution and clarify the impact of solvation on molecular structure, dynamics, and reactivity, crucial for comprehending biological processes and medication interactions. Atomistic simulations can investigate the molecular

conformational space, offering insights into the folding mechanisms, stability, and dynamics of molecules.

1.7.1.2 Coarse-Grained (CG) Models

Although all-atom (AA) molecular dynamics (MD) is capable of precisely replicating the physical, chemical, and thermodynamic properties of materials, its effectiveness is limited by the characteristic length and time scales it operates on, as well as the available computer resources.[72,73] In the fields of computational chemistry and chemical engineering, simulations of processes such as agglomeration, phase transitions, and self-assembly are generally computationally intensive and costly. This is because they involve enormous time scales and system sizes, with the majority of the system being the bulk solvent.[74–77] An approach to addressing this problem involves employing implicit solvent models, in which the solvent is treated as a uniform and structureless medium, and the interactions between the solute and solvent are characterized by parameters that inherently account for the influence of solvent molecules.[78–80] Nevertheless, these models are inadequate in representing crucial structural and physical characteristics at the interface between the solute and solvent. They are employed in simulations when a comprehensive understanding of the interactions between the solute and solvent is not necessary.[81,82] Computer simulations of soft-material systems, including gels, colloids, polymers, and their structures, are frequently enormous in scale and need significant computational resources.[83–85]

The primary objective for researchers in the field of molecular dynamics (MD) is to create advanced models that can effectively capture intricate atomistic details without the need for computationally intensive and expensive brute force atomistic simulations. These models should also be capable of replicating macroscopic experimental properties under varying conditions such as temperature and solvent mixtures.[86] The difficulties noted above have prompted more work to be done on coarse-grained (CG) modelling, which is a method of depicting atomistic systems with fewer degrees of freedom.[87,88]

The position of CG beads is determined based on the atomistic structure using a mapping approach.[89] Mapping entails the conversion of clusters of atoms from AA simulations into individual sites (CG beads), thereby allowing the full molecule to be simulated. The degree of coarse-graining is a term used to describe the total amount of heavy atoms that are integrated into a single CG bead. Although there is no definitive method to assign a certain set of atoms to a coarse-grained description, it cannot be selected entirely at random. When selecting a

mapping scheme and determining the degree of coarse-graining, it is important to examine the fundamental physics and chemistry of the material being represented, as well as the level of detail that the coarse-grained (CG) model needs to maintain in order to effectively study a specific subject of interest.[90,91]

1.8 Neutron Scattering

Neutron scattering has been a crucial tool for characterizing polymeric materials throughout the past two decades.[92,93] This method has contributed significantly to the growing understanding of the characteristics of polymeric systems, including the development of novel theoretical techniques.[58,94]

Neutron scattering is unique among methods for examining materials at small scales because the neutron interacts with an atom's nucleus. This characteristic, along with neutrons being electrically neutral, enables researchers to investigate matter in ways that would otherwise be unattainable. Neutron scattering affords an analysis of materials that consist of a combination of light and heavy components. Alternative scattering methods utilize photons (such as visible light and X-rays) and electrons. Photons and electrons are effective at detecting heavy elements due to their interaction with an atom's electron shell, which contains numerous electrons, similar to how a dental X-ray operates. Photons and electrons scatter off an atom's electron cloud, while neutrons penetrate to an atom's nucleus, providing distinct information. Oak Ridge National Laboratory (ORNL) physicist Hassina Bilheux stated[95] that a significant amount of scientific research is conducted based on the distinctive sensitivity of neutrons to light elements. X-rays penetrate bones while neutrons are sensitive to light elements and can detect soft tissue around the bone in a tissue sample. Neutrons, being electrically neutral, often pass through a substance without undergoing any interactions. This benefits researchers and poses a challenge to neutron scattering facilities.

Neutrons interact via short-range nuclear forces. They are highly penetrating and do not thermally degrade material. Neutron scattering lengths exhibit random variation with atomic number and are not influenced by momentum transfer Q. This is utilized to benefit in deuterium labelling because of the significant difference in scattering lengths between hydrogen (b_H = -3.739×10^{-13} cm) and deuterium ($b_D = 6.671 \times 10^{-13}$ cm). (The negative sign preceding b_H indicates that the scattered neutrons' wavefunction is in antiphase with the incident neutrons' wavefunction.)[96]

Neutron scattering experiments utilize contrast variation to improve the visibility of certain components inside a sample by selectively adjusting the scattering length density (SLD) of certain regions of the sample. Contrast variation (CV) is a technique that involves separating the neutron scattering intensity from a multi-component complex by using hydrogen-deuterium (H-D) substitution in the complex and/or the solvent. This method is valuable for analyzing intricate materials including biological macromolecules, polymers, and colloidal systems, where various components may have comparable scattering length densities, making them hard to differentiate using traditional neutron scattering techniques.[97]

The scattering intensity, when viewed as a function of the scattering vector q, is the Fourier transform of the pair distribution of the scattering centers, indicating that q space and r space are conjugate. One can transition from the scattering object in space r to its diffusion pattern in space q by a Fourier transformation. Small q values correspond to large r values, and vice versa (Figure. 1.5 (lower panel)). Examining a scattering diagram in terms of q is comparable to observing the scattering object via a variable-powered magnifying glass and subsequently performing a Fourier transform. As the power rises, the field diminishes and reaches a magnitude about q^{-1}. We will focus on analyzing a dilute solution or a system where the scattering pattern is fully defined by the form structure factor F(q) (see section 2.10).

Working at low q is akin to using a low-power magnifying lens. Since every molecule in this scenario is essentially a point, the only thing that can be done is count the number of points, which is a measure of the molecular weight (Figure. 1.5(A)). For a highly concentrated solution, neutron scattering can demonstrate $\overline{\Delta N^2}$ that, in an ideal solution, is equal to N. Summarily, at q = 0, only thermodynamic characteristics of the solution are measured, and no information about its structure is obtained.

Increasing q causes the molecules to no longer resemble point-like structures, with q^{-1} being about equal to the radius of gyration, as depicted in diagram (B). Details of the shape and structure of the molecules are not visible, only their dimensions are observable. Guinier demonstrated that in this area, the radius of gyration is measured.[93]

If we increase q more, we will reach either domain (C) or (D), depending on the concentration. Diagram (C) represents a dilute solution, while diagram (D) represents a moderately concentrated solution referred to as semi-dilute by de Gennes.[58] Diagram (C) shows that just a portion of a molecule is visible, with scattering independent of molecular weight. It provides insights into the chain's statistical properties and persistence length. When the solution is

concentrated (diagram (D)), various sections from distinct chains are visible, providing information about the length of the chain separating two contact locations, known as the correlation length. Increasing q will result in only a single segment of a chain remaining in the field. If the lateral dimensions of the chain are significantly smaller than the longitudinal dimensions, it will resemble the chain in statistical mechanics as shown in picture (E). When the persistence length is significantly greater than q^{-1} and the diameter of the cylinder enclosing the side groups is sufficiently small, a behavior characteristic of a rigid rod will be evident.

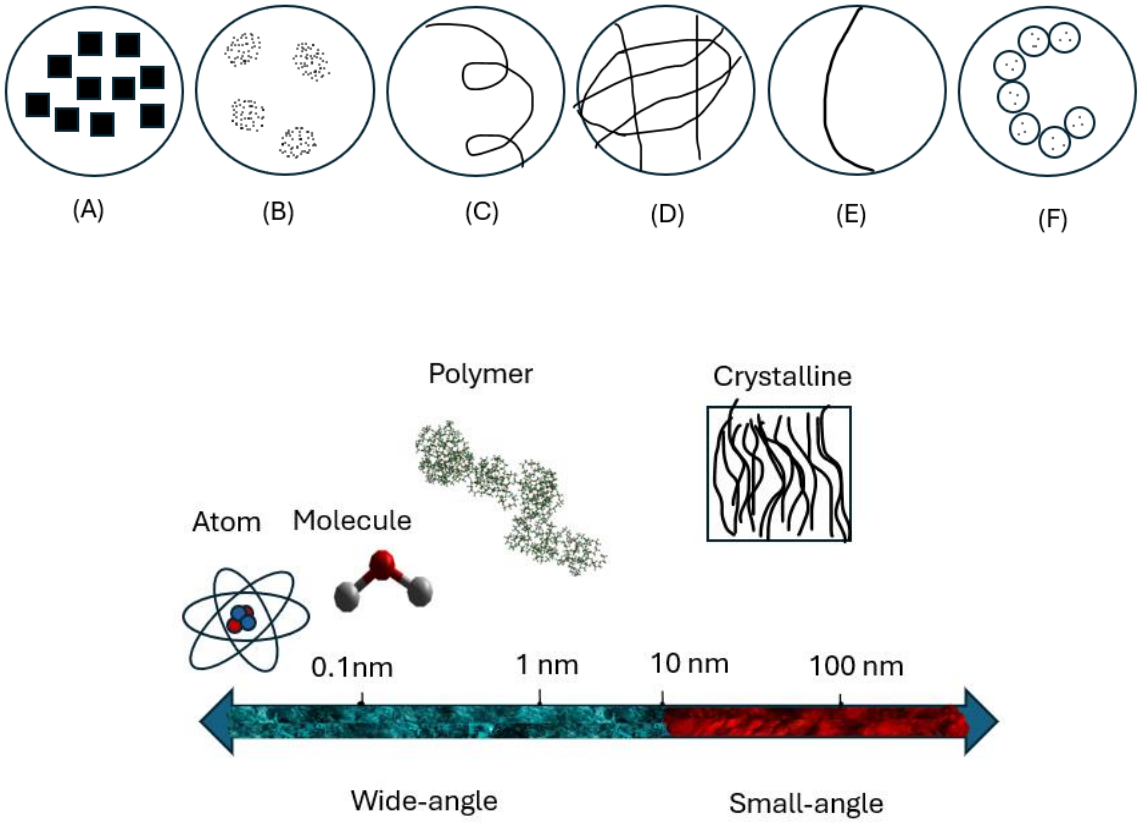


Figure 1. 5: Schematic illustration of different q domains (upper panel) and different space resolution of the scattering techniques (lower panel)

For higher values of q in picture (F), as q^{-1} approaches the length of the chemical bonds, the local structure of the chain starts to have an impact. One is outside the small angle scattering range an unresolved issue is determining the exact range of q values.[93] At that q range (F), wide-angle neutron scattering, detailed information about the atomic and molecular structure of materials, including the arrangement of atoms and chemical bonds can be obtained.

1.8.1 Near and InterMediate Range Order Diffractometer (NIMROD)

Near and InterMediate Range Order Diffractometer (NIMROD)[98] at the ISIS facility, located at the second target station UK, is designed to analyze disordered materials and liquids throughout a continuous length scale ranging from atomic to over 30 nm, with atomic distance resolution. This enables the instrument to measure the total structure factor F(q) and gather data on nanoscale systems and processes that are directly connected to the local atomic and molecular structure of the materials being studied. The capacity to conduct quantitative research on the structure of liquids and contemporary materials across a continuous length scale, ranging from the atomic to the nanometric, presents several novel prospects for enhancing comprehension of significant emerging fields in contemporary pure and applied science.[98] The instrument's innovative capability to gather precise structural data at atomic levels presents new challenges for traditional data analysis methods.[99–101] Classical and quantum atomistic simulations have significantly contributed to increasing the understanding in various fields of current condensed matter science. The problem posed by NIMROD is the need to construct and enhance atomistic models that consist of millions of atoms.

Dissolve is a simulation program,[102] which has been developing by Dr Tristan Youngs and his team, used to analyze total neutron scattering data, obtained from NIMROD instrument. The method expands on the principles introduced in the Empirical Potential Structure Refinement (EPSR) method[100] developed by Soper. It utilizes atomistic models to approximate experimental data and enable in-depth analysis. Dissolve software overcomes the limitations found in EPSR and offers a solution for handling bigger system sizes. Dissolve is used in our investigations of atomic resolution of polymer structure in solution (see chapter 3 and 4).

Existing research does not fully address the study of polymer solutions at such a high-resolution continuous length scale. Analysis of disordered materials has encountered numerous problems in both experimental and computational aspects. Neutron scattering data processing requires the removal of the undesired inelastic component, which can be difficult to eliminate and may lead to peaks or noise in the analysis. Furthermore, the huge size of the polymer system in all-atom simulations requires lengthy time scales to guarantee that the system reaches its equilibrated state.

The data obtained from the total neutron diffraction, contains the information of the total pairs correlations in the study system, which can be expressed as G(r), and this can include the correlations between solvent-solvent, polymer-polymer, and polymer-solvents. By fitting the

total structural function F(q) for the modelled system to the F(q) for the total neutron scattering data, G(r) can be obtained from the Fourier transform of F(q). Individual pair correlations then can be calculated from the modelled system. It should be noted that the calculated RDF obtained from Dissolve is in the weighted form, so it should not be surprised to see the weighted hydrogenated RDF lays in the negative part of the histogram due to the small coherent scattering length for hydrogen atom (b_{coh}) –3.74 fermis (fm).

1.9 Polydispersity

Characterizing polymer solutions can be a difficult task. Historically, the study of the arrangement of colloidal dispersions has depended on theoretical and experimental evaluations using model systems consisting of particles that are uniformly sized (monodisperse particles). Monodispersity is an assumption that all particles of the solute are the same in terms of their size and shape. In recent years, however, a large number of theoretical studies have been published that attempt to incorporate the impacts of polydispersity, size-polydispersity, in attractive interactions on the equilibrium structure of dispersions of spherical particles.[103] Polydispersity refers to the presence of a range of molecular weights within a sample of polymers. This adds complexity to the characterization procedure, as the properties of the solution are affected by the presence of molecules with varying sizes and weights.

It should be noted that scattering systems are typically characterized by averaging particle size distributions, irrespective of their sizes and shapes.[96] Thus, the experimental global averaging particle size aids in the study and understanding of polymer structural behaviors computationally represented as a single chain. However, varying chain lengths, which can impact the physical properties of the polymers, make a complete match between the estimated and actual results challenging.

1.10 Solvation Shell (Hydration layer)

The solvation shell is a layer of solvent molecules that surrounds a solute molecule when it is dissolved in a solvent. The shell is created by the interactions between the solute and solvent molecules. Water, the most prevalent solvent in biological and chemical processes, forms a solvation shell known as the hydration shell.

Hydration shell dynamics are believed to have a significant impact on various biological processes. The scientific community broadly agrees that studying the biomolecular hydration shell is essential for a comprehensive knowledge of the functions of proteins and nucleic acids.[104] Proteins and other biomolecules' conformational dynamics and structural stability are greatly influenced by the hydration layer.[105–108] The effective size of the biomolecule can be affected by the hydration layer's thickness. A thick hydration layer might cause the molecule to appear bigger, resulting in an increased radius of gyration.

Comparing the experimental small-angle scattering data (SAS) with the predicted data for a certain model presents another challenge. Since the experimental investigations are conducted in solutions, it is unknown how the solvent affects the radius of gyration. Using explicit-solvent SAS calculations to compute SAS curves from MD simulations, taking the explicit water molecules in the hydration shell into account, helps to quantify the effect of the hydration shell on the R_g of biomolecules. This method has been recently developed and successfully applied to biomolecules (see more information about this method in Section 2.12).

Polymer atoms in solution interact with their environment through hydration shells following the same idea for proteins or biomolecules. Thus, the mentioned methods were used to investigate the discrepancy between the predicted R_g in explicit water (for PVP and PEG) and SANS data. More detail on the result discussion can be found in Chapter 5.

1.11 History of Computational Modelling of PEG and PVP

1.11.1 PEG

The first computer models of PEG were developed in the early 1990s.[109] Smith et al.[110] created a force field for 1,2-dimethoxyethylene, which was then utilized to construct a model of PEG in water solution.[111,112] Early research included building united atom (UA) models of PEG to showcase its characteristics as a polymer electrolyte.[113,114] The PEG monomer, consisting of a nonpolar $(CH_2)_2$ group and an electronegative oxygen atom, will attach to cations present in the solution. Initial research with UA models demonstrated that PEG molecules surround cations, with four to five PEG oxygen atoms interacting with each ion. The initial simulations primarily examined the interaction between Li^+ ions and PEG in a solution containing dissolved LiI. The salt LiI was selected due to an interest in exploring the potential application of PEG as an electrolyte in lithium-ion batteries. Additional early simulations using the original models looked at the elastic properties of a single PEG chain,[115] the interaction between PEG and water

in diluted solution,[116] PEG dissolved in the nonpolar solvent benzene,[117] polymer melts of amorphous PEG,[118,119] crystalline PEG,[120] and PEG surfaces.[121]

Since then, a number of additional groups have created their own potential sets for PEG.[122–125] PEG force fields compatible with the CHARMM[126] and OPLS-AA[127,128] potential sets have been recently constructed. Additionally, a polarizable forcefield for PEG has been created.[118,129,130] Additionally, coarse-grained (CG) potentials for PEG have been produced alongside AA potentials. CG potential involves replacing atoms with interacting particles that represent atom groups. Greater access to time and length scales is possible with this method, but many subtleties unique to atomistic level interactions—like H bonds, for instance—are lost in the process.[131,132] An integrated Monte Carlo (MC) simulation of a PEGylated membrane with an experimental investigation of the interaction between fluorescent acylated PEG and PEGylated liposomes has also been conducted.[133]

Initial coarse-grained models of PEG utilized implicit solvent models, where the solvent's impact is simulated by modifying the interactions between solute particles. The results obtained for parameters such as aggregation number, chain diameters, and critical micelle concentration for PEG based surfactants were consistent with experimental findings.[134,135] The application of implicit solvent models is limited due to excessive approximation, however recent efforts have been made to merge implicit and explicit solvent models of PEG.[136] This led to the development of a CG model of PEG with an explicit solvent, which was subsequently utilized to describe the synthesis of diblock copolymers and their interactions with lipid bilayers.[137–139] An established method for creating coarse-grained interaction potentials for molecules is the widely used "MARTINI" coarse-grained force field.[140,141] Four separate MARTINI force fields were created for PEG[142–145] and utilized to simulate the creation of micelles, bicelles, and liposomes in mixed systems including PEGylated and conventional phospholipids. A CG force field was recently developed to combine two PEG monomers into one particle.[146]

1.11.2 PVP

Molecular modelling investigations of polypyrrolidone (PVP) structures incorporated in different biological systems were historically performed. An investigation was conducted to examine the potential effects of PVP on the characteristics of liquid and water in clathrate hydrate.[147] This study concludes that the PVP monomer had a favorable attachment to the hydrate surface. Further study for the same group compared the potential kinetic inhibitors of PVP, polyvinylcaprolactam (PVCap), and its modifications in a multiphase water–hydrate

system, which concluded PVCap outperforms PVP as a kinetic hydrate inhibitor.[148] Another group synthesized their own PVP derivatives and used molecular simulation to study the inhibition process of the hydrate growth mechanism.[149] Moreover, molecular simulation was used to clarify the interacting groups of PVP and chitosan blend polymers, which are used in various biomedical applications and in advanced drug delivery systems.[150] Similar studies were performed to understand the celecoxib-PVP amorphous systems[151], the solubility and hydrogen bonding interactions of amorphous indomethacin-PVP[152], PAN/PVP blends,[153], $CoFe_2O_4$/PVP,[154], structural evolution of PVP/Ag nanowires,[155] the dissolution process of solid dispersions of blend polymers[156,157], and the evaluation of amorphous simvastatin-PVP solid dispersions[158].

Molecular modeling of the conformational behavior of oligomer PVP in solution regarding stereochemistry, iso-, syndio-, and atactic, were previously investigated.[159] Additional study of PVP behavior adsorbed onto a hydrate aimed to analyse the interaction of PVP on a hydrate surface and distinguish the differences between the polymeric and monomeric systems.[160] A model was developed to explain the process of evaporation and drying of the polymer PVP in water.[161] A computational modelling of readily available PVP as a quenching fluorescent polymer for nitrite, nitrate, iodide, and thiocyanate in water-based solutions was investigated.[162] Very recently (2023), a combination of X-ray diffraction and the Empirical Structural Refinement (EPSR) model was used to quantify the nature and strength of hydrogen bonding between the absorbed water molecule and a 5-monomer unit of PVP.[163]

1.12 Current Challenges in Modelling Polymer Scattering

Studies of PVP[164,165] and PEG[166–171] solution structures using scattering measurements have been investigated in the literature. PEG molecules of molecular weights less than 10,000 g/mol were observed by Rubinson and Krueger[167] to be packed in flat plate-shaped chains at low concentrations. However, other studies[168,169] confirmed that PEG molecules behave like random chains, and more recently,[170] like wormlike chains. In addition, the PVP structure was examined at various concentrations.[164] At low concentrations, a flexible coil shape was seen, whereas at high concentrations, it transitioned towards a rodlike conformation. However, another study[165] showed that PVP exhibits behaviour similar to that of a chain with excluded volume. Moreover, the hydration analysis of PEG using molecular dynamic (MD) simulations (up to 50 monomer units)[172] and using differential scanning calorimetry (DSC) for different molecular weights[173] were discussed. Additionally, the study using ^{2}H NMR measurements[174]

for PVP hydration were investigated. However, there is a lack of research in the field that combines small-angle neutron scattering (SANS) and molecular dynamics (MD) simulations with a similar average Mw to understand the behaviour of the polymer chain and its hydration.

Even with long walk computational modelling investigations mentioned above, including the biopolymers of interest, PEG and PVP, into the biological systems, the combination study of atomistic molecular simulation and experimental total neutron scattering for water-soluble polymers to capture the critical polymer-water interactions, producing accurate polymer chain dimensions that influence by solvation shell, is not fully understood or covered in the literature. This in turn necessitates a force field (FF) validation study, for both the polymer and water, to ensure that the models being used are exploring the correct conformational space before being applied to assist in designing these systems. Measuring the dimensions of a polymer chain in water, like R_g, can serve as an initial screening step before conducting FF investigations for polymer-water systems. R_g is a valuable metric for validating against experimental data due to its common occurrence in experimental studies and its ability to provide precise polymer chain dimensions in solution, which serve as a reliable indicator of various interactions. Aside from determining the accurate chain dimensions, many properties can be assessed in polymer-solvent systems that are tailored to a specific purpose. Hydration shells and solvent structure analysis can be derived from radial distribution functions, g(r), between polymer atoms and water molecules.[175–178]

Many synthesised polymers do not conform to a singular, clearly defined structure in solution; rather, they exhibit diverse ensembles. A crucial factor in modelling success is the choice of model resolution; on the other hand, realistic interaction potentials and the availability of experimental data to build starting systems are prerequisites. Small-angle and wide-angle neutron scattering in solution (SANS, WANS) are precise methods for gathering information on molecular structures, ensembles, and time-resolved dynamics under conditions close to their natural state. Yet, analysing the solution scattering data using computational methods is challenging due to the limited information in the data, and the influence of scattering from the hydration layer. Molecular dynamics (MD) simulations could provide guidance on how to interpret scattering data in addition to helping to overcome these difficulties. Atomistic force fields provide detailed physical information that enhances the limited information from experimental data. Explicit-solvent molecular dynamics can be utilised to forecast solvent scattering, and the sampling methods connected to molecular dynamics can assist in refining the structure based on scattering data. Established methods[179,180] integrate experimental

SAXS/SANS data on-the-fly into MD simulation, leading to conformational sampling in agreement with the data. With this approach, the SAS curve can be predicted using explicit solvent calculations, incorporating atomistic models of the hydration layer and the excluded solvent, thus eliminating the need for solvent-related fitting factors. This approach has been used in numerous successful protein and biomolecular studies, which were most likely helped by the crystal structure's availability as a starting point.[179–183] Yet, the application of this method to synthesised (bio) polymers in solution for predicting the SAS curve and achieving thorough enhanced sampling, while considering the solvation shell, is still far off. Similarly, a refined atomistic MD modelling method based on total neutron scattering data (NIMROD), a high-resolution methodology in atomic level, has been developed using Dissolve software. Several molecule structures have been investigated[184–191] using the combination of the later methods, however, water soluble polymers, PEG and PVP, were not been examined.

References

1 M. Rubinstein and R. H. Colby, *Polymer Physics*, Oxford University Press, 2003.

2 F. W. Billmeyer, *Textbook of Polymer Science*, John Wiley & Sons, New York, 3nd ed., 1984.

3 M. N. Belgacem and A. Gandini, *Monomers, Polymers and Composites from Renewable Resources*, Elsevier, 2011.

4 H. Namazi, Polymers in Our Daily Life, *BioImpacts*, 2017, **7**, 73–74.

5 W. He and R. Benson, in *Handbook of Polymer Applications in Medicine and Medical Devices*, Elsevier, 2014, pp. 55–76.

6 B. Love, in *Biomaterials*, Academic Press, 2017, pp. 205–238.

7 J. Li and D. J. Mooney, Designing Hydrogels for Controlled Drug Delivery, *Nat. Rev. Mater.*, 2016, **1**, 1–17.

8 K. Y. Lee and D. J. Mooney, Hydrogels for Tissue Engineering, *Chem. Rev.*, 2001, **101**, 1869–1880.

9 W. F. M. Daniel, J. Burdyńska, M. Vatankhah-varnoosfaderani, K. Matyjaszewski, J. Paturej, M. Rubinstein, A. V Dobrynin and S. S. Sheiko, Solvent-Free , Supersoft and Superelastic Bottlebrush Melts and Networks, *Nat. Mater.*, 2016, **15**, 183–190.

10 B. Y. Zhou, W. Huang, J. Liu, X. Zhu and D. Yan, Self-Assembly of Hyperbranched Polymers and Its Biomedical Applications, *Adv. Mater.*, 2010, **22**, 4567–4590.

11 O. Pillai and R. Panchagnula, Polymers in drug delivery, *Curr. Opin. Chem. Biol.*, 2001, **5**, 447–451.

12 A. Srivastava, T. Yadav, S. Sharma, A. Nayak and A. Kumari, Polymers in Drug Delivery, *J. Biosci. Med.*, 2016, **4**, 69–84.

13 P. Bhatt, S. Trehan, N. Inamdar, V. K. Mourya and A. Misra, in *Applications of Polymers in Drug Delivery*, Elsevier, 2nd ed., 2021, pp. 1–42.

14 D. Schmaljohann, Thermo- and pH-Responsive Polymers in Drug Delivery, *Adv. Drug Deliv. Rev.*, 2006, **58**, 1655–1670.

15 W. B. Liechty, D. R. Kryscio, B. V Slaughter and N. A. Peppas, Polymers for Drug

Delivery Systems, *Annu. Rev. Chem. Biomol. Eng.*, 2010, **1**, 149–173.

16 A. D. Anisha and R. Shegokar, Polyethylene glycol (PEG): A versatile Polymer for Pharmaceutical Applications, *Expert Opin. Drug Deliv.*, 2016, **13**, 1257–1275.

17 P. Zarrintaj, M. R. Saeb and S. H. Jafari, in *Compatibilization of Polymer Blends*, Elsevier, 2020, pp. 511–537.

18 S. Mehdipour-ataei and M. Mohammadi, in *Applications of Unsaturated Polyester Resins*, Elsevier, 2023, pp. 141–153.

19 J. Milton Harris and R. B. Chess, Effect of Pegylation on Pharmaceuticals, *Nat. Rev. Drug Discov.*, 2003, **2**, 214–221.

20 A. Abuchowski, J. R. Mccoy, N. C. Palczuk, T. VAN Es and F. F. Davis, Effect of Covalent Attachment of Polyethylene Glycol on Immunogenicity and Circulating Life of Bovine Liver Catalase, *J. Biol. Chem.*, 1977, **252**, 3582–3586.

21 R. Webster, E. Didier, P. Harris, N. Siegel, J. Stadler, L. Tilbury, D. Smith and W. E. T. Al, PEGylated Proteins : Evaluation of Their Safety in The Absence of Definitive Metabolism Studies, *Drug Metab. Dispos.*, 2007, **35**, 9–16.

22 M. Y. Kariduraganavar, A. A. Kittur and R. R. Kamble, *Polymer Synthesis and Processing*, Elsevier, 1st edn., 2014.

23 F. Haaf, A. Sanner and F. Straub, Polymers of N-Vinylpyrrolidone: Synthesis, Characterization and Uses, *Polym. J.*, 1985, **17**, 143–152.

24 F. Fischer and B. Stephan, An All Rounder in The Chemistry Polyvinylpyrrolidone, *Chemie unserer Zeit*, 2009, **43**, 376–383.

25 M. Teodorescu and B. Maria, Poly (vinylpyrrolidone) – A Versatile Polymer for Biomedical and Beyond Medical Applications, *Polym. Plast. Technol. Eng.*, 2015, **54**, 923–943.

26 X. Zhi, H. Fang, C. Bao, G. Shen, J. Zhang, K. Wang, S. Guo, T. Wan and D. Cui, The Immunotoxicity of Graphene Oxides and The Effect of PVP-Coating, *Biomaterials*, 2013, **34**, 5254–5261.

27 P. Franco and I. De Marco, The Use of Poly(N-vinyl pyrrolidone) in The Delivery of Drugs: A review, *Polymers (Basel).*, 2020, **12**, 1114.

28 X. Zheng, T. Zhang, X. Song, L. Zhang, C. Zhang, S. Jin, J. Xing and X.-J. Liang, Structural Impact of Graft and Block Copolymers Based on Poly(N-vinylpyrrolidone) and Poly(2- dimethylaminoethyl methacrylate) in Gene Delivery, *J. Mater. Chem.*, 2015, **3**, 4027–4035.

29 L. Zhang, Y. Liang, L. Meng and C. Wang, Characterization of Complexation of PVP Copolymer with DNA, *Polym. Adv. Technol.*, 2009, **20**, 410–415.

30 Y. Song, T. Zhang, X. Song, L. Zhang, C. Zhang, J. Xing and X. J. Liang, Polycations with Excellent Gene Transfection Ability Based on PVP-g-PDMAEMA with Random Coil and Micelle Structures as Non-viral Gene Vectors, *J. Mater. Chem.*, 2015, **3**, 911–918.

31 A. Saxena, S. Mozumdar and A. K. Johri, Ultra-Low Sized Cross-Linked Polyvinylpyrrolidone Nanoparticles as Non-Viral Vectors for in Vivo Gene Delivery, *Biomaterials*, 2006, **27**, 5596–5602.

32 S. Sheu, L. Chou, Y. Bee, J. Chen, H. Lin, P. Lin, H. Lam and M. Tai, Suppression of Choroidal Neovascularization by Intramuscular Polymer-Based Gene Delivery of Vasostatin, *Exp. Eye Res.*, 2005, **81**, 673–679.

33 M. Hu, C. Li, X. Li, M. Zhou, J. Sun, F. Sheng and L. Lu, Zinc Oxide/Silver Bimetallic Nanoencapsulated in PVP / PCL Nanofibres for Improved Antibacterial Activity, *Artif. Cells, Nanomedicine, Biotechnol.*, 2018, **46**, 1248–1257.

34 M. Hecold, R. Buczkowska, A. Mucha, J. Grzesiak, O. Rac-rumijowska, H. Teterycz and K. Marycz, The Effect of PEI and PVP-Stabilized Gold Nanoparticles on Equine Platelets Activation : Potential Application in Equine Regenerative Medicine, *J. Nanomater.*, 2017, **2017**, 1–11.

35 G. G. de Lima, D. W. de Lima, M. J. de Oliveira, A. B. Lugão, M. T. Alcântara, D. M. Devine and M. J. de Sá, Synthesis and in Vivo Behavior of PVP/CMC/Agar Hydrogel Membranes Impregnated with Silver Nanoparticles for Wound Healing Applications, *ACS Appl. Bio Mater.*, 2018, **1**, 1842–1852.

36 V. Ramalingam, K. Varunkumar, V. Ravikumar and R. Rajaram, Target Delivery of Doxorubicin Tethered with PVP Stabilized Gold Nanoparticles for Effective Treatment of Lung Cancer, *Sci. Rep.*, 2018, **8**, 3815.

37 P. A. Rose, P. K. Praseetha, M. Bhagat, P. Alexander, S. Abdeen and M. Chavali, Drug Embedded PVP Coated Magnetic Nanoparticles for Targeted Killing of Breast Cancer Cells, *Technol. Cancer Res. Treat.*, 2013, **12**, 463–472.

38 C. Bothiraja, M. B. Shinde, S. Rajalakshmi and A. P. Pawar, Evaluation of Molecular Pharmaceutical and in-Vivo Properties of Spray-Dried Isolated Andrographolide – PVP, *J. Pharm. Pharmacol.*, 2009, **61**, 1465–1472.

39 P. Gupta and A. K. Bansal, Spray Drying for Generation of a Ternary Amorphous System of Celecoxib , PVP , and Meglumine, *Pharm. Dev. Technol.*, 2008, **10**, 273–281.

40 P. J. Dowding, R. Atkin, B. Vincent and P. Bouillot, Oil Core / Polymer Shell Microcapsules by Internal Phase Separation from Emulsion Droplets . II : Controlling the Release Profile of Active Molecules, *Langmuir*, 2005, **21**, 5278–5284.

41 K. E. Wu, J. Li, W. Wang and D. A. Winstead, Formation and Characterization of Solid Dispersions of Piroxicam and Polyvinylpyrrolidone Using Spray Drying and Precipitation with Compressed Antisolvent, *J. Pharm. Sci.*, 2009, **98**, 2422–2431.

42 V. Prosapio, I. De Marco and E. Reverchon, PVP/Corticosteroid Microspheres Produced by Supercritical Antisolvent Coprecipitation, *Chem. Eng. J.*, 2016, **292**, 264–275.

43 M. Rasekh, C. Karavasili, Y. Ling, N. Bouropoulos, M. Morris, D. Armitage, X. Li, D. G. Fatouros and Z. Ahmad, Electrospun PVP–Indomethacin Constituents for Transdermal Dressings and Drug Delivery Devices, *Int. J. Pharm.*, 2014, **473**, 95–104.

44 D. G. Yu, X. Shen, X. Zhang, C. Branford-White and L. Zhu, Preparation and Characterization of Fast-Dissolving Electrospun Drug-Loaded PVP Nanofiber Membranes, *Acta Polym. Sin.*, 2009, **11**, 1170–1174.

45 X.-Y. Dai, W. Nie, Y.-C. Wang, Y. Shen, Y. Li and S.-J. Gan, Electrospun Emodin Polyvinylpyrrolidone Blended Nanofibrous Membrane: A Novel Medicated Biomaterial for Drug Delivery and Accelerated Wound Healing, *J. Mater. Sci. Mater*, 2012, **23**, 2709–2716.

46 I. Sriyanti, D. Edikresnha, M. M. Munir, H. Rachmawati and K. Khairurrijal, Electrospun Polyvinylpyrrolidone (PVP) Nanofiber Mats Loaded by Garcinia

Mangostana L. Extracts, *Mater. Sci. Forum*, 2017, **880**, 11–14.

47 D. G. Yu, X. F. Zhang, X. X. Shen, C. Brandford—White and L. . Zhu, Ultrafine Ibuprofen-Loaded Polyvinylpyrrolidone Fiber Mats Using Electrospinning, *Polym. Int.*, 2009, **58**, 1010–1013.

48 M. Rapado Raneque and C. Rodríguez Rodríguez, A. Peniche Covas, Hydrogel Wound Dressing Preparation at Laboratory Scale by Using Electron Beam and Gamma Radiation, *Nucleus*, 2013, **53**, 24–31.

49 K. Wlodarski, L. Tajber and W. Sawicki, Physicochemical Properties of Direct Compression Tablets with Spray Dried and Ball Milled Solid Dispersions of Tadalafil in PVP-VA, *Eur. J. Pharm. Biopharm.*, 2016, **109**, 14–23.

50 B. Janugade, S. Patil, S. Patil and P. Lade, Effect of PVP Concentration on Lag Time of Press-Coated Montelukast Sodium Tablet, *J. Pharm. Res*, 2009, **2**, 502–506.

51 E. Karavas, E. Georgarakis and D. Bikiaris, Application of PVP/HPMC Miscible Blends with Enhanced Mucoadhesive Properties for Adjusting Drug Release in Predictable Pulsatile Chronotherapeutics, *Eur. J. Pharm. Biopharm*, 2006, **64**, 115–126.

52 I. D. Del Consuelo, F. Falson, R. H. Guy and Y. Jacques, Ex vivo Evaluation of Bioadhesive Films for Buccal Delivery of Fentanyl, *J. Control. Release*, 2007, **122**, 135–140.

53 L. Perioli, V. Ambrogi, F. Angelici, M. Ricci, S. Giovagnoli, M. Capuccella and C. Rossi, Development of Mucoadhesive Patches for Buccal Administration of Ibuprofen, *J. Control. Release*, 2004, **99**, 73–82.

54 V. Mohabe, R. Akhand and A. K. Pathak, Preparation and Evaluation of Captopril Transdermal Patches, *Bull. Pharm. Res*, 2011, **1**, 47–52.

55 R. Sadashivaiah, B. Dinesh, U. A. Patil and K. Raghu, Design and in Vitro Evaluation of Haloperidol Lactate Transdermal Patches Containing Ethyl Cellulose-Povidone as Film Formers, *Asian J. Pharm*, 2014, **2**, 43–49.

56 R. Gupta and B. Mukherjee, Development and in Vitro Evaluation of Diltiazem Hydrochloride Transdermal Patches Based on Povidone–Ethylcellulose Matrices, *Drug Dev. Ind. Pharm*, 2003, **29**, 1–7.

57 D. Campbell, R. A. Pethrick and J. R. White, *Polymer Characterization: Physical Techniques*, CRC press., 2nd ed., 2000.

58 P.-G. De Gennes, *Scaling Concepts in Polymer Physics*, Cornell university press, 1979.

59 P. J. Flory, Thermodynamics of High Polymer Solutions, *J. Chem. Phys.*, 1942, **10**, 51–61.

60 M. L. Huggins, Solutions of Long Chain Compounds, *J. Chem. Phys.*, 1941, **9**, 440.

61 J. des Cloizeaux and G. Jannink, *Polymers in Solution Their Modelling and Structure*, Clarendon Press, Oxford, 1990.

62 G. D. La Torre, J. Llorca, J. Carrascosa and J. Valpuesta, Hydromic: Prediction of Hydrodynamic Properties of Rigid Macromolecular Structures Obtained from Electron Microscopy Images, *Eur. Biophys. J.*, 2001, **30**, 457–462.

63 O. Zheng and J. I. E. Liang, Predicting Protein Folding Rates from Geometric Contact and Amino Acid Sequence, *Protein Sci.*, 2008, **17**, 1256–1263.

64 F. Ferri, M. Greco and R. Mattia, On The Determination of The Average Molecular Weight, Radius of Gyration, and Mass/Length Ratio of Polydisperse Solutions of Polymerizing Rod-Like Macromolecular Monomers by Multi-Angle Static Light Scattering., *Macromol. Symp.*, 2000, **162**, 23–44.

65 D. Smilgies and E. Folta-stogniew, Molecular Weight – Gyration Radius Relation of Globular Proteins : A Comparison of Light Scattering , Small-Angle X-ray Scattering and Structure-Based Data, *J. Appl. Crystallogr.*, 2015, **48**, 1604–1606.

66 K. Haidukivska, V. Blavatska and J. Paturej, Universal Size Ratios of Gaussian Polymers with Complex Architecture: Radius of Gyration vs Hydrodynamic Radius, *Sci. Rep.*, 2020, **10**, 14127.

67 A. Rudin and C. Phillip, *Practical Aspects of Molecular Weight Measurements*, The elements of polymer science & engineering, 2013.

68 C. Hammond, *The Basics of Crystallography and Diffraction*, Oxford University Press Oxford, 2001.

69 B. T. M. Willis and C. J. Carlile, *Experimental Neutron Scattering*, Oxford University

Press, New York, 2015.

70 D. S. Sivia, *Elementary Scattering Theory For X-ray and Neutron Users*, Oxford University Press, 2011.

71 T. E. Gartner and A. Jayaraman, Modeling and Simulations of Polymers : A Roadmap, *Macromolecules*, 2019, **52**, 755−786.

72 K. Y. Sanbonmatsu and C. Tung, High Performance Computing in Biology : Multimillion Atom Simulations of Nanoscale Systems, *J. Struct. Biol.*, 2007, **157**, 470–480.

73 S. Y. Joshi and S. A. Deshmukh, A Review of Advancements in Coarse-Grained Molecular Dynamics Simulations, *Mol. Simul.*, 2021, **47**, 786–803.

74 J. F. Kraft, M. Vestergaard, B. Schiøtt and L. Thøgersen, Modeling the Self-Assembly and Stability of DHPC Micelles Using Atomic Resolution and Coarse Grained MD Simulations, *J. Chem. Theory Comput.*, 2012, **8**, 1556–1569.

75 P. Brocos, P. Mendoza-Espinosa, R. Castillo, J. Mas-Oliva and J. Pineiro, Multiscale Molecular dynamics Simulations of Micelles: Coarse-Grain for Self-Assembly and Atomic Resolution for Finer Details, *Soft Matter*, 2012, **8**, 9005–9014.

76 J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, M. L. Klein, T. Procter, G. Company, M. V. Laboratories and P. O. Box, A Coarse Grain Model for Phospholipid Simulations, *J. Phys. Chem. B*, 2001, **105**, 4464–4470.

77 H. Leontiadou, A. E. Mark and S. J. Marrink, Antimicrobial Peptides in Action, *J. Am. Chem. Soc.*, 2006, **128**, 12156–12161.

78 I. Shub, E. Schreiber and Y. Kliger, Saving Significant Amount of Time in MD Simulations by Using an Implicit Solvent Model and Elevated Temperatures, *Int. Sch. Res. Not.*, 2013, **2013**, 1–5.

79 M. Feig, Ed., *Modeling Solvent Environments: Applications to Simulations of Biomolecules*, John Wiley & Sons, 2010.

80 S. Wang, Z. Li and W. Pan, Implicit-Solvent Coarse-Grained Modeling for Polymer Solutions Via Mori-Zwanzig Formalism, *Soft Matter*, 2019, **15**, 7567–7582.

81 M. Brieg, J. Setzler and W. Wenzel, Generalized Born Implicit Solvent Models for

Small Molecule Hydration Free Energies, *Phys. Chem. Chem. Phys.*, 2017, **19**, 1677–1685.

82 J. Kleinjung and F. Fraternali, Design and Application of Implicit Solvent Models in Biomolecular Simulations, *Curr. Opin. Struct. Biol.*, 2014, **25**, 126–134.

83 S. A. Deshmukh, S. K. R. S. Sankaranarayanan, K. Suthar and D. C. Mancini, Role of Solvation Dynamics and Local Ordering of Water in Inducing Conformational Transitions in Poly (N -isopropylacrylamide) Oligomers through the LCST, *J. Phys. Chem. B*, 2012, **116**, 2651–2663.

84 C. Peter and K. Kremer, Multiscale Simulation of Soft Matter Systems – From the Atomistic to The Coarse-Grained Level and Back, *Soft Matter*, 2009, **5**, 4357–4366.

85 G. Milano, G. Santangelo, F. Ragone, L. Cavallo and A. Di Matteo, Gold Nanoparticle / Polymer Interfaces : All Atom Structures from Molecular Dynamics Simulations, *J. Phys. Chem. C*, 2011, **115**, 15154–15163.

86 S. O. Nielsen, C. F. Lopez, G. Srinivas and M. L. Klein, Coarse Grain Models and The Computer Simulation of Soft Materials Soft Materials, *J. Phys. Condens. Matter*, 2004, **16**, R481.

87 Z. Jiang, J. He, S. A. Deshmukh, P. Kanjanaboos, G. Kamath, Y. Wang, S. K. R. S. Sankaranarayanan, J. Wang, H. M. Jaeger and X. Lin, Subnanometre Ligand-Shell Asymmetry Leads to Janus-Like Nanoparticle Membranes, *Nat. Mater.*, 2015, **14**, 912–917.

88 K. K. Bejagam, S. K. Singh, R. Ahn and S. A. Deshmukh, Unraveling the Conformations of Backbone and Side Chains in Thermosensitive Bottlebrush Polymers, *Macromolecules*, 2019, **52**, 9398–9408.

89 H. Al Karimi-Varzaneh and F. Muller-Plathe, in *Multiscale Molecular Methods in Applied Chemistry*, Multiscale molecular methods in applied chemistry, 2012, pp. 295–322.

90 M. Dallavalle and N. F. A. Van Der Vegt, Evaluation of Mapping Schemes for Systematic Coarse Graining of Higher Alkanes, *Phys. Chem. Chem. Phys.*, 2017, **19**, 23034–23042.

91 V. A. Harmandaris, D. Reith, N. F. A. Van Der Vegt and K. Kremer, Comparison

Between Coarse-Graining Models for Polymer Systems : Two Mapping Schemes for Polystyrene, *Macromol. Chem. Phys.*, 2007, **208**, 2109–2120.

92 J. S. Higgins and H. C. BenoÎt, *Polymers and Neutron Scattering*, Clarendon Press, Oxford, 1994.

93 H. Benor, On The Scattering of Neutrons by Polymers at High Scattering Angles, *Pure Appl. Chem.*, 1997, **69**, 131–142.

94 J. des Cloizeaux and G. Jannink, Polymers in Solution, *J. Phys. Condens. Matter*, 1990, **2**, 1–24.

95 L. Williams, ORNL Review, https://www.ornl.gov/blog/what-makes-neutron-scattering-unique, (accessed 25 February 2024).

96 B. Hammouda, *Probing Nanoscale Structures-The SANS Toolbox*, National Institute of Standards and Technology Center for Neutron Research, 2008.

97 S. Krueger, Small-Angle Neutron Scattering Contrast Variation Studies of Biological Complexes : Challenges and Triumphs, *Curr. Opin. Struct. Biol.*, 2022, **74**, 102375.

98 D. T. Bowron, A. K. Soper, K. Jones, S. Ansell, S. Birch, J. Norris, L. Perrott, D. Riedel, N. J. Rhodes, S. R. Wakefield, A. Botti, M. A. Ricci, F. Grazzi and M. Zoppi, NIMROD: The Near and InterMediate Range Order Diffractometer of The ISIS Second Target Station, *Rev. Sci. Instrum.*, 2010, **81**, 033905.

99 A. K. Soper, Empirical Potential Monte Carlo Simulation of Fluid Structure, *Chem. Phys.*, 1996, **202**, 295–306.

100 A. K. Soper, Partial Structure Factors from Disordered Materials Diffraction Data: An Approach Using Empirical Potential Structure Refinemen, *Phys. Rev. B*, 2005, **72**, 104204.

101 R. L. McGreevy, Reverse Monte Carlo Modelling, *J. Phys. Condens. Matter*, 2001, **13**, R877.

102 T. Youngs, Dissolve: Next Generation Software for The Interrogation of Total Scattering Data by Empirical Potential Generation, *Mol. Phys.*, 2019, **117**, 3464–3477.

103 S. Paul and R. Rajagopalan, *Polydispersity in Colloids: Implications to Static Structure and Scattering*, Advances in Colloid and Interface science, 1993, vol. 43.

104 A. C. Fogarty, E. Duboué-Dijon, F. Sterpone, J. T. Hynes and D. Laage, Biomolecular Hydration Dynamics: A Jump mModel Perspective, *Chem. Soc. Rev.*, 2013, **42**, 5672–5683.

105 D. Laage, T. Elsaesser and J. T. Hynes, Water Dynamics in The Hydration Shells of Biomolecules, *Chem. Rev.*, 2017, **117**, 10694–10725.

106 S. K. Pal and A. H. Zewail, Dynamics of Water in Biological Recognition, *Chem. Rev.*, 2004, **104**, 2099–2124.

107 Y. Levy and J. N. Onuchic, Water Mediation in Protein Folding and Molecular Recognition, *Annu. Rev. Biophys. Biomol. Struct*, 2006, **35**, 389–415.

108 P. Ball, Water as An Active Constituent in Cell Biology, *Chem. Rev.*, 2008, **108**, 74–108.

109 A. Bunker, in *Computational Pharmaceutics Application of Molecular Modeling in Drug Delivery*, Computational Pharmaceutics, 2015, pp. 217–233.

110 G. D. Smith, R. L. Jaffe and D. Y. Yoon, Force Field for Simulations of 1, 2-Dimethoxyethane and Poly (Oxyethylene) Based upon Ab Initio Electronic Structure Calculations on Model Molecules, *J. Phys. Chem.*, 1993, **97**, 12752–12759.

111 D. Bedrov and G. . Smith, Anomalous Conformational Behavior of Poly(Ethylene Oxide) Oligomers in Aqueous Solutions. A Molecular Dynamics Study, *J. Chem. Phys.*, 1998, **109**, 8118–8123.

112 G. D. Smith, D. Bedrov and O. Borodin, Conformations and Chain Dimensions of Poly(Ethylene Oxide) in Aqueous Solution: A Molecular Dynamics Simulation Study, *J. Am. Chem. Soc.*, 2000, **122**, 9548–9549.

113 F. Muller-Plathe and W. F. Van Gunsteren, Computer Simulation of A Polymer Electrolyte: Lithium Iodide in Amorphous Poly(Ethylene Oxide), *J. Chem. Phys.*, 1995, **103**, 4745–4756.

114 K. Laasonen and M. L. Klein, Molecular Dynamics Simulations of The Structure and Ion Diffusion in Poly(Ethylene Oxide), *J. Chem. Soc. Faraday Trans.*, 1995, **91**, 2633–2638.

115 B. Heymann and H. Grubmu¨ller, Elastic Properties of Poly(Ethylene-Glycol) Studied

by Molecular Dynamics Stretching Simulations, *Chem. Phys. Lett.*, 1999, **307**, 425–432.

116 K. Tasaki, Poly(Oxyethylene)-Water Interactions: A Molecular Dynamics Study, *J. Am. Chem. Soc.*, 1996, **118**, 8459–8469.

117 K. Tasaki, Conformation and Dynamics of Poly(Oxyethylene) in Benzene Solution: Solvent Effect from Molecular Dynamics Simulation, *Macromolecules*, 1996, **29**, 8922–8933.

118 O. Borodin and G. D. Smith, Molecular Dynamics Simulations of Poly(Ethylene Oxide)/LiI Melts 1. Structural and Conformational Properties, *Macromolecules*, 1998, **31**, 8396–8406.

119 S. Neyertz and D. Brown, A Computer Simulation Study of The Chain Configurations in Poly(Ethylene Oxide)-Homolog Melts, *J. Chem. Physics,* 1995, **102**, 9725–9735.

120 S. Neyertz, D. Brown and J. O. Thomas, Molecular Dynamics Simulation of Crystalline Poly(Ethylene Oxide), *J. Chem. Phys.*, 1994, **101**, 10064–10073.

121 A. Aabloo and J. Thomas, Molecular Dynamics Simulations of A Poly(Ethylene Oxide) Surface, *Comput. Theor. Polym. Sci.*, 1997, **7**, 47–51.

122 H. Dong, J. K. Hyun, C. Durham and R. A. Wheeler, Molecular Dynamics Simulations and Structural Comparisons of Amorphous Poly(Ethylene Oxide) and Poly(Ethylenimine) Models, *Polymer (Guildf).*, 2001, **42**, 7809–7817.

123 J. Fischer, D. Paschek, A. Geiger and G. Sadowski, Modeling of Aqueous Poly(Oxyethylene) Solutions: 1. Atomistic Simulations, *J. Phys. Chem.*, 2008, **112**, 2388–2398.

124 E. A. Tritopoulu and I. G. Economou, Molecular Simulation of Structure and Thermodynamic Properties of Pure Tri- and Tetra-Ethylene Glycols and Their Aqueous Mixtures, *Fluid Phase Equilib.*, 2006, **248**, 134–146.

125 M. Winger, A. H. de Vries and W. F. van Gunsteren, Force-Field Dependence of The Conformational Properties of α,ω-DimethoxyPolyethylene Glycol, *Mol. Phys.*, 2009, **107**, 1313–1321.

126 I. Vorobyov, V. M. Anisimov and S. et al. Greene, Additive and Classical Drude

Polarizable Force Fields for Linear and Cyclic Ethers, *J. Chem. Theory Comput.*, 2007, **3**, 1120– 1133.

127 A. Maciejewski, M. Pasenkiewicz-Gierula and O. et al. Cramariuc, Refined OPLS All-Atom Force Field for Saturated Phosphatidylcholine Bilayers at Full Hydration, *J. Phys. Chem. B*, 2014, **118**, 4571–4581.

128 M. Stepniewski, M. Pasenkiewicz-Gierula and T. et al. Róg, Study of PEGylated Lipid Layers as a Model for PEGylated Liposome Surfaces: Molecular Dynamics Simulation and Langmuir Monolayer Studies, *Langmuir*, 2011, **27**, 7788–7798.

129 O. Borodin, R. Douglas and G. D. et al. Smith, MD Simulations and Experimental Study of Structure Dynamics and Thermodynamics of Poly(Ethylene Oxide) and its Oligomers, *J. Phys. Chem. B*, 2003, **107**, 6813–6823.

130 O. Borodin and G. D. Smith, Development of Quantum Chemistry-Based Force Fields for Poly(Ethylene Oxide) with Many-Body Polarization Interactions, *J. Phys. Chem. B*, 2003, **107**, 6801–6812.

131 T. Murtola, A. Bunker and I. et al. Vattulainen, Multiscale Modeling of Emergent Materials: Biological and Soft Matter, *Phys. Chem. Chem. Phys.*, 2009, **11**, 1869–1892.

132 C. Chen, P. Depa and V. G. et al. Sakai, A Comparison of United Atom, Explicit Atom, and Coarse-Grained Simulation Models for Poly(Ethylene Oxide), *J. Chem. Phys.*, 2006, **124**, 234901.

133 S. Rex, M. J. Zuckermann, M. Lafleur and J. R. Silvius, Experimental and Monte Carlo Simulation Studies of The Thermodynamics of PolyethyleneGlycol Chains Grafted to Lipid Bilayers, *Biophys. J.*, 1998, **75**, 2900–2914.

134 D. Bedrov, C. Ayyagari and G. D. Smith, Multiscale Modeling of Poly(Etheylene Oxide)-Poly(Propylene Oxide)-Poly(Ethylene Oxide) Triblock Copolymer Micelles in Aqueous Solution, *J. Chem. Theory Comput.*, 2006, **2**, 598–606.

135 J. Fischer, D. Paschek, A. Geiger and G. Sadowski, Modeling of Aqueous Poly(Oxyethylene) Solutions: 2. Mesoscale Simulations, *J. Phys. Chem. B*, 2008, **112**, 13561–13571.

136 A. Juneja, J. Numata, L. Nilsson and E. W. Knapp, Merging Implicit with Explicit

Solvent Simulations: Polyethylene Glycol, *J. Chem. Theory Comput.*, 2010, **6**, 1871–1883.

137 G. Srinivas, J. C. Shelley, S. O. Nielsen, D. E. Discher and M. L. Klein, Simulation of Diblock Copolymer Self-Assembly, Using A Coarse-Grain Model, *Mol. Phys.*, 2004, **108**, 8153–8160.

138 G. Srinivas, D. E. Discher and M. L. Klein, Self-Assembly and Properties of Diblock Copolymers by Coarse-Grain Molecular Dynamics, *Nat. Mater.*, 2004, **3**, 638–644.

139 G. Srinivas and M. L. Klein, Coarse Grain Molecular Dynamics Simulations of Diblock Copolymer Surfactants Interacting with A Lipid Bilayer, *J. Phys. Chem.*, 2004, **102**, 8153–8160.

140 S. J. Marrink, A. H. de Vries and A. E. Mark, Coarse Grained Model for Semiquantitative Lipid Simulations, *J. Phys. Chem. B*, 2004, **108**, 750–760.

141 S. J. Marrink, H. J. Risselada and S. et al. Yefimov, The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations, *J. Phys. Chem. B*, 2007, **111**, 7812–7824.

142 H. Lee, A. H. de Vries, S. J. Marrink and R. W. Pastor, A Course-Grained Model for Polyethylene Oxide and Polyethylene Glycol: Conformation and Hydrodynamics, *J. Phys. Chem. B*, 2009, **113**, 13186–13194.

143 G. Rossi, P. F. J. Fuchs, J. Barnoud and L. Monticelli, A Course-Grained MARTINI Model of Polyethylene Glycol and of Polyoxyethylene Alkyl Ether Surfactants, *J. Phys. Chem. B*, 2012, **116**, 14353–14362.

144 S. C. Yang and R. Faller, Pressure and Surface Tension Control Self-Assembled Structures in Mixtures of PEGylated and Non-PEGylated Lipids, *Langmuir*, 2012, **28**, 2275–2280.

145 E. Choi, J. Mondal and A. Yethiraj, Coarse-Grained Models for Aqueous Polyethylene Glycol Solutions, *J. Phys. Chem. B*, 2014, **118**, 323–329.

146 Q. Wang, D. J. Keffer and D. M. Nicholson, A Coarse-Grained Model for Polyethylene Glycol Polymer, *J. Chem. Physics,* 2011, **135**, 214903.

147 B. B. Kvamme, G. Huseby and O. K. Forrisdahl, Molecular Dynamics Simulations of

PVP Kinetic Inhibitor in Liquid Water and Hydrate/Liquid Water Systems, *Mol. Phys.*, 1996, **90**, 979–992.

148 B. Kvamme, T. Kuznetsova and K. Aasoldsen, Molecular Dynamics Simulations for Selection of Kinetic Hydrate Inhibitors, *J. Mol. Graph. Model.*, 2005, **23**, 524–536.

149 Z. Li, F. Jiang, H. Qin, B. Liu, C. Sun and G. Chen, Molecular Dynamics Method to Simulate The Process of Hydrate Growth in The Presence/Absence of KHIs, *Chem. Eng. Sci.*, 2017, **164**, 307–312.

150 K. Suknuntha, V. Tantishaiyakul, V. Vao-Soongnern, Y. Espidel and T. Cosgrove, Molecular Modeling Simulation and Experimental Measurements to Characterize Chitosan and Poly (Vinyl Pyrrolidone) Blend Interactions, *J. Polym. Sci. Part B Polym. Phys.*, 2008, **46**, 1258–1264.

151 P. Gupta, R. Thilagavathi, A. K. Chakraborti and A. K. Bansal, Role of Molecular Interaction In Stability of Celecoxib− PVP Amorphous Systems, *Mol. Pharm.*, 2005, **2**, 384–391.

152 T. X. Xiang and B. D. Anderson, Molecular Dynamics Simulation of Amorphous Indomethacin-Poly (VinylPyrrolidone) Glasses: Solubility and Hydrogen Bonding Interactions, *J. Pharm. Sci.*, 2013, **102**, 876–891.

153 M. E. I. Linyu, W. A. N. G. Xigang, L. I. U. Yaqing and W. A. N. G. Junyuan, Computer Simulation of PAN/PVP Blends Compatibility and Preparation of Aligned PAN Porous Nanofibers Via Magnetic-Field-Assisted Electrospinning PAN/PVP Blends, *Mater. Sci.*, 2019, **25**, 54–59.

154 S. Mirzaee, S. Mahdavifar and S. H. Hekmatara, Synthesis, Characterization and Monte Carlo Simulation of CoFe2O4/Polyvinylpyrrolidone Nanocomposites: The Coercivity Investigation, *J. Magn. Magn. Mater.*, 2015, **393**, 1–7.

155 C. Hao, J. Su, P. Pan, L. Sun, Z. He and B. Tang, Structural Evolution of PVP@ Ag Nanowires Induced by Focused Electron Beam Irradiation: The Passivation Effect of PVP, *J. Mater. Res. Technol.*, 2024, **29**, 1467–1475.

156 Y. Shahzad, S. Sohail, M. S. Arshad, T. Hussain and S. N. H. Shah, Development of Solid Dispersions of Artemisinin for Transdermal Delivery, *Int. J. Pharm.*, 2013, **457**, 197–205.

157 T. Chan and D. Ouyang, Investigating The Molecular Dissolution Process of Binary Solid Dispersions by Molecular Dynamics Simulations, *J. Pharm. Sci.*, 2018, **13**, 248–254.

158 P. Kapourani, A. Chatzitheodoridou, M. Kontogiannopoulos, K.N. Barmpalexis, Experimental, Thermodynamic, and Molecular Modeling Evaluation of Amorphous Simvastatin-Poly (VinylPyrrolidone) Solid Dispersions, *Mol. Pharm.*, 2020, **17**, 2703–2720.

159 T. Flebbe, R. Hentschke, E. Hädicke and C. Schade, Modeling of PolyvinylPyrrolidone and Polyvinylimidazole in Aqueous Solution, *Macromol. theory simulations*, 1998, **7**, 567–577.

160 T. J. Carver, M. G. Drew and P. M. Rodger, Configuration-Biased Monte Carlo Simulations of Poly(vinylpyrrolidone) at a Gas Hydrate Crystal Surface, *Ann. N. Y. Acad. Sci.*, 2000, **912**, 658–668.

161 S. R. Gopireddy and E. Gutheil, Modeling and Simulation of Water Evaporation from a Droplet of Polyvinylpyrrolidone (PVP) Aqueous Solution, *ICLASS*, 2012, 1–7.

162 H. C. Kam, D. T. Ranathunga, E. R. Payne, R. A. Smaldone, S. O. Nielsen and S. C. Dodani, Spectroscopic Characterization and in Silico Modelling of PolyvinylPyrrolidone as an Anion-Responsive Fluorescent Polymer in Aqueous Media, *Supramol. Chem.*, 2019, **31**, 514–522.

163 C. J. Benmore, S. R. Benmore, S. K. Wilke, V. Menon, S. R. Byrn and J. K. . Weber, X-ray Diffraction of Water in Polyvinylpyrrolidone, *Mol. Pharm.*, 2023, **20**, 3645–3652.

164 L. Sapir, C. B. Stanley and D. Harries, Properties of polyvinylpyrrolidone in a deep eutectic solvent, *J. Phys. Chem. A*, 2016, **120**, 3253–3259.

165 A. Aschi, M. M. Jebari and A. Gharbi, Investigation of Poly (vinyl pyrrolidone) in methanol by dynamic light scattering and viscosity techniques, *e-Polymers*, 2007, **7**, 020.

166 K. L. Linegar, A. E. Adeniran, A. F. Kostko and M. A. Anisimov, Hydrodynamic radius of polyethylene glycol in solution obtained by dynamic light scattering, *Colloid J.*, 2010, **72**, 279–281.

167 K. A. Rubinson and S. Krueger, Poly (ethylene glycol) s 2000–8000 in water may be planar: A small-angle neutron scattering (SANS) structure study, *Polymer (Guildf).*, 2009, **50**, 4852–4858.

168 L. Almásy, O. P. Artykulnyi, V. I. Petrenko, O. I. Ivankov, L. A. Bulavin, M. Yan and V. M. Haramus, Structure and intermolecular interactions in aqueous solutions of polyethylene glycol, *Molecules*, 2022, **27**, 2573.

169 P. Thiyagarajan, D. J. Chaiko and R. P. Hjelm Jr, A neutron scattering study of poly (ethylene glycol) in electrolyte solutions, *Macromolecules*, 1995, **28**, 7730–7736.

170 A. Ryoki, F. Watanabe, T. Okudaira, S. Takahashi, T. Oku, K. Hiroi, R. Motokawa and Y. Nakamura, Contrast dependence of scattering profiles for poly (ethylene glycol) in water: Investigation by small-angle neutron scattering with 3He spin filter and small-angle x-ray scattering, *J. Chem. Phys.*, 2024, **160**, 114907.

171 K. Devanand and J. C. Selser, Asymptotic behavior and long-range interactions in aqueous solutions of poly (ethylene oxide), *Macromolecules*, 1991, **24**, 5943–5947.

172 A. M. Rozza, D. E. Vanpoucke, E. M. Krammer, J. Bouckaert, R. Blossey, M. F. Lensink, M. J. Ondrechen, I. Bakó, J. Oláh and G. Roos, Hydration sphere structure of architectural molecules: polyethylene glycol and polyoxymethylene oligomers, *J. Mol. Liq.*, 2023, **384**, 122172.

173 L. Huang and K. Nishinari, Interaction between poly (ethylene glycol) and water as studied by differential scanning calorimetry, *J. Polym. Sci. Part B Polym. Phys.*, 2001, **39**, 496–506.

174 M. Nakada, H. Ishida and Y. Furushima, Structural and dynamical characterisation of intermediate water interacting polyvinyl pyrrolidone, *Materialia*, 2020, **12**, 100743.

175 E. W. Bucholz, J. B. Haskins, J. D. Monk, C. W. Bauschlicher Jr and J. . Lawson, Phenolic Polymer Solvation in Water and Ethylene Glycol, I: Molecular Dynamics Simulations, *J. Phys. Chem. B*, 2017, **121**, 2839–2851.

176 T. E. de Oliveira, C. M. Marques and P. A. Netz, Molecular Dynamics Study of The LCST Transition in Aqueous Poly (Nn-Propylacrylamide), *Phys. Chem. Chem. Phys.*, 2018, **20**, 10100–10107.

177 Q. Wei, Y. Wang, W. Chai, T. Wang and Y. Zhang, Effects of Composition Ratio on

The Properties of Poly (Vinyl Alcohol)/Poly (Acrylic Acid) Blend Membrane: A Molecular Dynamics Simulation Study, *Mater. Des.*, 2016, **89**, 848–855.

178 S. J. Rukmani, G. Kupgan, D. M. Anstine and C. M. Colina, A Molecular Dynamics Study of Water-Soluble Polymers: Analysis of Force Fields from Atomistic Simulations, *Mol. Simul.*, 2019, **45**, 310–321.

179 L. Chatzimagas and J. S. Hub, Predicting Solution Scattering Patterns with Explicit-Solvent Molecular Simulations, *Methods Enzymol.*, 2022, **677**, 433–456.

180 P. Chen and J. S. Hub, Interpretation of Solution X-Ray Scattering by Explicit-Solvent Molecular Dynamics, *Biophysj*, 2015, **108**, 2573–2584.

181 P. C. Chen, R. Shevchuk, F. M. Strnad, C. Lorenz, L. Karge, R. Gilles, A. M. Stadler, J. Hennig and J. S. Hub, Combined Small-Angle X-ray and Neutron Scattering Restraints in Molecular Dynamics Simulations, *J. Chem. Theory Comput.*, 2019, **15**, 4687–4698.

182 W. He, A. Henning-knechtel and S. Kirmizialtin, Visualizing RNA Structures by SAXS-Driven MD Simulations, *Front. Bioinforma.*, 2022, **2**, 1–19.

183 L. Chatzimagas, J. S. Hub and T. Physics, Structure and Ensemble Refinement Against SAXS Data : Combining MD Simulations with Bayesian Inference or with The Maximum Entropy Principle, 2022, 1–23.

184 Y. Shen, Y. Xiao, R. M. Edkins, T. G. Youngs, T. L. Hughes, J. Tellam and K. Edkins, Elucidating The Hydrotropism Behaviour of Aqueous Caffeine and Sodium Benzoate Solution Through NMR and Neutron Total Scattering Analysis, *Int. J. Pharm.*, 2023, **647**, 123520.

185 T. F. Headen, D. Mino, G. A. Youngs and A. J. Clancy, The Structure of Liquid Thiophene from Total Neutron Scattering, *Phys. Chem. Chem. Phys*, 2023, **25**, 25157–25165.

186 G. R. Mitchell, D. Bowron, A. Mateus, P. Bartolo, T. Gkourmpis, K. Phomphrai, D. Lopez and F. Davis, SANS/WANS Time-resolving Neutron Scattering Studiesof Polymer Phase Transitions Using NIMROD, *MRS Online Proc. Libr.*, 2013, **1528**, 1–6.

187 A. M. Jubb, L. F. Ruppert, T. G. Youngs and T. F. Headen, Exploring methane

behavior in Marcellus Shale micropores via contrast matching neutron scattering. , , pp., *Energy & Fuels*, 2020, **34**, 10926–10932.

188 C. R. Hill, C. Mitterdorfer, T. G. Youngs, D. T. Bowron, H. J. Fraser and T. Loerting, Neutron scattering analysis of water's glass transition and micropore collapse in amorphous solid water, *Phys. Rev. Lett.*, 2016, **116**, 215501.

189 M. Falkowska, S. Chansai, H. G. Manyar, L. F. Gladden, D. T. Bowron, T. G. Youngs and C. Hardacre, Determination of Toluene Hydrogenation Kinetics with Neutron Diffraction, *Phys. Chem. Chem. Phys.*, 2016, **18**, 17237–17243.

190 A. K. Soper and K. J. Edler, Coarse-grained empirical potential structure refinement: Application to a reverse aqueous micelle, *Biochim. Biophys. Acta (BBA)-General Subj.*, 2017, **1861**, 1652–1660.

191 E. J. Nilsson, V. Alfredsson, D. T. Bowron and K. J. Edler, A Neutron Scattering and Modelling Study of Aqueous Solutions of Tetramethylammonium and Tetrapropylammonium Bromide, *Phys. Chem. Chem. Phys.*, 2016, **18**, 11193–11201.

Chapter 2

Theory and Methods

This chapter introduces the basic MD simulation theory and approach, including the force fields and water models, and covers the basic experimental techniques used in this study.

2.1 Molecular dynamics simulations

Molecular dynamics simulation is a computer approach that can be applied to study a diversity of various systems, including polymer solutions. There are two types of molecular dynamics simulations: classical and quantum. In quantum simulation, the change in the system is calculated by solving a time-dependent Schrödinger equation, which is impractical and computationally very costly for large systems. However, classical simulations use Newtonian mechanics as forcefields to represent the interaction between particles, and it is applicable and low-cost to model large system timescales. The study here focuses mostly on classical MD as we are dealing with polymers in solutions, which usually contain a large number of atoms.

A trajectory is generated, which contains details of the initial starting positions of each atom and is obtained by integrating Newton's laws of motion (Eq. 2.1). This trajectory specifies how the particles in the system vary with time.

$$F = ma = \frac{dp}{dt} = m\frac{dv}{dt} \qquad (2.1)$$

Where F, m, a, and p are force, mass, acceleration, momentum, and velocity, respectively.

The terms force (F) or the derivative of potential energy (V) with respect to the position (r) (or coordinates) can describe the interaction between particles (Eq. 2.2).

$$F(r) = -\frac{\partial V}{\partial r} \qquad (2.2)$$

The right-hand side is the negative of the energy gradient. Since the acceleration is the second derivative of the position r with respect to time, a differential form might be written as follows (Eq. 2.3).

$$-\frac{\partial V}{\partial r} = m\frac{\partial^2 r}{\partial t^2} \qquad (2.3)$$

By solving this equation, we can obtain the coordinates of particles as a function of time or the time behavior of the system. Estimating the stability of the results requires conducting many different simulations and considering various starting conditions.[1]

Before running the MD, several input sets are required to prepare the study system. Start by creating the initial coordinates of all atoms in the system and setting a box size and shape, which are determined by the three basis dimensions of the periodic box (x, y, and z). After the box system is defined, it is filled with solvent (water in this case). Users must provide the topology for the molecule(s) of interest, which includes all the necessary information to define the molecule within a simulation. This information includes nonbonded parameters (atom types and charges) as well as bonded parameters (bonds, angles, and dihedrals). Then, the system can be minimized, equilibrated, and used to carry out the final MD productions.

2.2 Energy minimization

Energy minimization is an important first step whose goal is to determine the system's minimal energy structure, and without determining the lowest energy, it can result in unstable high-energy systems. During the minimization step, the atoms change their original positions and move until they find the lowest energy structure. This lowest energy structure is typically known as a local minimum, which is a place where the value is at its lowest around that area and not necessarily the absolute lowest value over the entire domain. GROMACS offers various algorithms for minimization. The steepest descent method[2,3], also known as gradient descent, which is the oldest technique for minimizing a general nonlinear function, is used here because it is more resilient and simple to use for a large system than other GROMACS methods. Thanks to the energy minimization step, which allows us to start the MD with a reasonable structure in terms of geometry and solvent orientation.

One of the primary thermodynamic ensembles used in molecular dynamics (MD) simulations is the microcanonical ensemble, often known as the NVE ensemble, in which the number of particles, N, volume, V, and the total energy, E, are conserved and the temperature and pressure fluctuate. Alternative ensembles, such as a canonical ensemble (NVT) or an isobaric isothermal ensemble (NPT), that can be sampled offer outcomes that are more pertinent to the circumstances of the experiment. Examining the velocities and forces acting on the atoms in the simulation allows for the consideration of temperature and pressure in an MD run. Thus, NVT and NPT ensembles are used to specify the system's attributes: temperature T (thermostats) and pressure P (barostats) in this work.

2.3 Canonical ensemble (NVT)

For all explicit MD systems presented here, the NVT was performed, in which the velocities of particles are re-selected from the Maxwell-Boltzmann distribution and the velocities of particles are regulated to achieve the desired temperature (298 K). Usually, the velocity rescaling,[4] Nose-Hoover,[5] or Berendsen algorithms[6] are used to perform temperature coupling. Velocity scaling can be employed to establish the temperature in a molecular dynamics (MD) simulation due to the direct correlation between temperature and the average kinetic energy of the atoms. The velocity rescaling was used in this thesis to carry out the system equilibrations, which is appropriate for a wide range of systems due to its accuracy and stability.

In the NVT ensemble, the system maintains a constant temperature by exchanging energy (heat) with its surroundings, while keeping the volume (V) and particle number (N) fixed. This ensemble shows a system that may exchange energy (heat) with its surroundings to maintain a constant temperature. During the NVT process, the solvents equilibrate and orientate around polymers until the system reaches the proper temperature.

2.4 Isothermal–isobaric ensemble (NPT)

The NPT ensemble operates by setting the number of atoms, temperature, and pressure as constant values and controlling the temperature and pressure with a heat bath and a barostat, which allows the box volume to vary in order to obtain the correct pressure. In GROMACS, various barostat algorithms are available to keep pressure constant, and we used Parrinello-Rahman algorithms[7] to regulate pressure coupling in the system to reach the targeted pressure at 1 bar. This method is suitable for extended ensembles and guarantees the correct NPT simulation. Solvents are permitted to equilibrate and orient around polymers during the NPT simulation without causing any changes to the structural integrity of the polymers, and the volume of the simulation box is allowed to fluctuate until the system reaches the desired pressure.

After completing the two equilibration phases, the system is now well-equilibrated at the appropriate temperature and pressure. The position restrictions can now be lifted, and production MD can begin collecting data.

2.5 Solving as a Function of Time (MD)

For doing MD simulations, numerous integration algorithms are available to solve the Newton's equations of motion. The Newton dynamical equation is a differential equation including the derivative of the position vector with regard to time. In simple two-particle systems, the interaction between particles (potential energy V) can be determined systematically by giving r(t) in the mathematical functions. While in complicated systems (with more than two particles), the positions are calculated by a numerical sequence of small, finite-time steps that correspond to a Taylor series with time.

This Taylor expansion can be done in a number of ways, the most widely used being the leap-frog method, which is implemented as the main integrator in the GROMACS program package.[8] The leap-frog[9] is a modification of the Verlet integrator[10] that is also used in GROMACS but not all option sets have fully included it yet, so the leap-frog is used in this work. As a variation on the Verlet integrator, the leap-frog method is highly effective in generating locations and velocities that change over time and offer programming simplicity and convenience. The leap-frog technique, as its name implies, allows *r* and *v* to leapfrog over one another by generating the velocity (*v)* at half integer time steps (Δt) and the location (*r)* at integer time steps. The equations used in the Leapfrog algorithm are:

$$r\,(t\,+\,\Delta t)\,=\,r(t)\,+\,v\left(t\,+\,\frac{1}{2}\,\Delta t\right)\Delta t \qquad (2.4)$$

$$v\left(t\,+\,\frac{1}{2}\,\Delta t\right)\,=\,v\left(t\,-\,\frac{1}{2}\,\Delta t\right)\,+\,a(t)\,\Delta t$$

Where a(t), Δt, and r(t) are acceleration, integration timestep, and position, respectively. In order to evaluate the kinetic energy at the same time as other physical characteristics, the velocity at integer time step is derived as an average.

$$v\,(t)\,=\,\frac{1}{2}\left[v\left(t\,+\,\frac{1}{2}\,\Delta t\right)\,+\,v\left(t\,-\,\frac{1}{2}\,\Delta t\right)\right] \qquad (2.5)$$

For the fastest motions, the time step Δt must be small enough to avoid integration mistakes. One femtosecond time step would be necessary for an explicit simulation of bond vibrations. A time step of 2 fs is made possible by constricting the bond vibrations (h-bonds) using the

SETTLE algorithm[11] for water molecules and the LINCS method[12] for the other molecules (polymers) because they scarcely link to the macromolecule's global motion, so the time step used in this work is 2 fs.

2.6 Force-field

In the context of molecular modeling, a classical force field is a set of parameter expressions together with numerical coefficients that describe the potential energy of particles in a system as a function of the particles' positions, ignoring the behaviour of the individual electrons.[13] Based on the assumption of the Born-Oppenheimer approximation, the motions of nuclei and electrons in a molecule can be separated, so this makes it possible to write the energy as a function of the nuclear coordinates. Bonded interactions for covalently bound atoms (bond length E_{bond}, bond angle E_{angle}, and dihedral potentials E_{dih}) and nonbonded potentials $E_{nonbond}$ for electrostatic and van der Waals forces are typically included in a force field:[13–15]

$$E_{total} = E_{bond} + E_{angle} + E_{dih} + E_{nonbond} \qquad (2.6)$$

These parameters describe how a molecule will interact based on its position and surroundings. The sum of these forces yields energy, from which the force can be computed. Multiple commonly employed families of force fields exist. Polymers are commonly represented using Optimized Potentials for Liquid Simulations in All Atoms (OPLS-AA), the Chemistry at Harvard Macromolecular Mechanics (CHARMM), and GROMOS force fields. It is crucial to acknowledge that the terms of Equation (2.6) may vary slightly among different force fields. As most of our analysis is done using OPLS-AA force field, its potential energy function E is described by Eqs. (2.7 −2.10).

$$E_{bond} = \sum_{r} K_r (r - r_0)^2 \qquad (2.7)$$

$$E_{angle} = \sum\nolimits_{angle} K_0 (\theta - \theta_0)^2 \qquad (2.8)$$

$$E_{dih} = \sum_{i} \frac{V_1^i}{2} [1 + cos\varphi_i] + \frac{V_2^i}{2} [1 - cos\, 2\varphi_i] + \frac{V_3^i}{2} [1 + cos 3\varphi_i] \qquad (2.9)$$

$$E_{nonbond} = \sum\nolimits_{i} \sum_{j>i} \left\{ \frac{q_i q_j e^2}{r_{ij}} + 4\,\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}} \right) \right\} \qquad (2.10)$$

K_r and K_θ represent the force constants that relate to bond lengths and bond angles, respectively. The variable r denotes the distance between two atoms, whereas θ represents the angle formed by the bond. The terms V_1 etc. correspond to the coefficients of the Fourier series in Eq. (2.9). The partial atomic charges, Lennard-Jones radii, Lennard-Jones well-depth, and distance between one pair of non-bonded atoms, i and j, are represented by the variables q, σ, ϵ, and r_{ij}, respectively. Ultimately, the combining rules are such that $\sigma_{ij} = (\sigma_{ii}\sigma_{jj})^{1/2}$ and $\epsilon_{ij} = (\epsilon_{ii}\epsilon_{jj})^{1/2}$.

2.6.1 Electrostatic interactions

A list of all atom pairs is needed to calculate non-bonded interactions. The non-bonded interactions scale quadratically with the number of particles N in the system if all conceivable interactions are explicitly examined. For systems with more than 10^4 atoms, the precise calculation of every LJ and Coulomb interaction becomes unaffordable. As a result, the LJ interactions often end at a distance of 1.0–1.4 nm.[16] Nevertheless, cutting off the Coulomb potential results in notable artefacts because of its long-range nature (r^{-1}).[17,18] As a result, the Particle-Mesh Ewald (PME) approach is frequently used to represent long-range Coulomb interactions. This method mimics long-range electrostatics by allocating charges to a grid and computing the potential as a simple sum in reciprocal space. PME scales as N log N and benefits from quick computations of the Fourier transform in reciprocal space.[19,20]

2.7 Solvation Models

An essential part of computational chemistry involves assessing the impact of the environment, such as a solvent. Atomistic molecular dynamics (MD) simulation techniques can be classified into two primary categories: explicit solvent MD, which explicitly includes the solvent molecules, and implicit solvent MD, which does not explicitly include the solvent molecules, all of which have their own advantages and disadvantages. Combinations can be achieved by explicitly considering the first solvation shell and applying a continuum model to the other components.[21] The molecular conformation and chemical reaction viability are determined by the solvent model selected. Thus, assessing the accuracy of structural dynamics requires exploring different existence models.

2.7.1 Implicit Solvent Models

The popular implicit solvent model replaces the discrete molecules of a real water environment with an infinite continuous medium that possesses the dielectric characteristics of water.[22–29] The formation of a cavity in the medium requires energy, which leads to destabilization. On the other hand, dispersion interactions between the solvent and solute contribute to stabilization, namely the van der Waals energy between them. The dispersion term, often known as dispersion/repulsive, which may include a repulsive component in certain cases. The solvation energy can be expressed as shown in equation (2.11).

$$\Delta G_{solvation} = \Delta G_{cavity} + \Delta G_{dispersion} + \Delta G_{elec} \qquad (2.11)$$

Multiple implicit solvent models exist, and they can vary in their parameters and level of accuracy. A brief description of the implicit models that are mentioned in this work, which were done on Molecular Operator Environment (MOE) software.

2.7.1.1 Generalized Born Model

The generalized Born model (GB) has gained popularity in the implicit solvent framework due to its balanced trade-off between accuracy and speed, as well as the wide range of options it offers in prominent molecular modelling software programs. GB models assess the electrostatic component of solvation free energy by calculating the total of pairwise interactions between atomic charges. In the case of aqueous solvation of molecules with an interior dielectric of 1, these interactions are estimated. A function with analytical properties that was introduced by Still et al.[29]:

$$\Delta G_{sol} = -\frac{1}{2}\left(1 - \frac{1}{\epsilon_w}\right)\sum_{ij}\frac{q_i q_j}{\sqrt{r_{ij}^2 + R_i R_j \, exp\left(-\frac{r_{ij}^2}{4R_i R_j}\right)}} \qquad (2.12)$$

The variable r_{ij} represents the distance between atoms i and j, whereas q_i and q_j represent the partial charges. The value of ϵ_w, which is greater than 1, represents the dielectric constant of the solvent. The crucial variables in the GB function are the effective Born radii of the interacting atoms, R_i and R_j, which indicate the level of burial of each atom within the solution. Presently, numerous iterations of the GB model are accessible in a multitude of molecular modelling software packages. Almost all of these models are based on the same foundation,

which is represented by Equation (2.12). However, there can be significant variations in the methods used to calculate the effective radii.[28]

2.7.1.2 Distance-Dependent Dielectric Model

In the past, MD simulations commonly used the distance-dependent dielectric model[30] to consider solvation effects. In this method, the influence of electrostatic effects is represented using Coulomb's law, where the dielectric is a predetermined function of the distance between charges. For example, in the simplest version of the model, the dielectric function is defined as $\epsilon(r) = r$. Despite being widely considered less accurate than the GB model,[31] the model is in continuous usage today due to its extreme simplicity and computing efficiency.[32,33]

2.7.1.3 Reaction-Field Model

The introduction of this concept was initiated by Barker and Watts.[34,35] The purpose of this design was to simulate the impact of the solvent by incorporating a reaction field that considers the dielectric screening of electrostatic interactions, and modifies the paired potentials to account for the dielectric screening effect of the solvent.

2.7.2 Explicit Solvent Models

Explicit solvent models directly account for the presence and behavior of solvent molecules. Such models provide a more intuitively realistic representation where there are direct and distinct interactions between a solvent and a solute, as opposed to continuum models. Therefore, it is crucial to explicitly incorporate water molecules in order to provide a relatively precise description of solute-solvent interactions.[36,37] Several explicit water models have been developed in order to encompass all of the physicochemical characteristics of water.[38–47] Typically, these can be differentiated based on three criteria: (i) the quantity of interaction points referred to as sites, (ii) whether the model is inflexible or adaptable, (iii) whether the model incorporates polarization effects. These water models mainly vary on the description of their parameters, which includes bond lengths, angles, force constants for flexible models, atomic partial charges, and Lennard-Jones sigma and epsilon values that characterize van der Waals interactions.[48]

Transferable Intermolecular Potential with n Points (TIPnP) and Simple Point Charge (SPC) derivatives models are widely employed for biomolecular applications. In this work, the most analysis was done using TIP3P in GROMACS and simple point charge flexible water

(SPC/Fw) in Dissolve, along with the initial investigation used for SPC, TIP4P, and TIP5P on the polymer chain behavior. It is important to mention that while these explicit solvent models have been developed to accurately represent the physical and chemical properties of bulk water, specifically the interactions between water molecules, the accuracy of the interactions between solute and solvent in these models is uncertain and needs additional investigation.[36]

2.8 The Neutron Diffraction Experiment

Neutron generators rely on nuclear reactions to release neutrons that are trapped within atomic nuclei. These neutrons can be generated either in reactors or using accelerators.

2.8.1 Neutron Sources

2.8.1.1 Nuclear Reactors

The process of nuclear fission is employed in a nuclear reactor to generate a continuous stream of neutrons with varying wavelengths, using fuel components enriched with ^{235}U. The technique relies on James Chadwick's Nobel Prize-winning experiment in 1935, in which a beryllium target is hit with α particles to generate neutrons through the reaction:

$$^{9}Be + \alpha \rightarrow C \rightarrow {}^{12}C + {}^{1}n \qquad (2.13)$$

Using monochromators or velocity selectors, a range of wavelengths can then be picked correspondingly with the proper diffraction technique.

2.8.1.2 Neutron Spallation

Neutron spallation sources use the acceleration of proton beams to high energies, which are subsequently directed towards a heavy metal target for bombardment. The contact between protons and nuclei in the target results in the nuclei becoming excited, which triggers a cascade effect that leads to the ejection of high-energy particles, such as neutrons, from the nuclei. Spallation sources can exist in two forms: continuous or pulsed. Continuous spallation sources, like the SINQ facility in Switzerland, generate a steady stream of neutrons through the spallation process, which subsequently requires either monochromation or pulsing. At pulsed neutron sources, such as the SNS in the USA or ISIS in the UK, the protons that start the process of spallation come in discrete groups. As a result, the neutrons that are expelled have a pulsating

quality, which enables the measurement of their time-of-flight. Their wavelength can be determined using the de Broglie equation:

$$\lambda = \frac{h}{mv} = \frac{ht}{mL} \qquad (2.14)$$

Where h and m are the planck's constant and mass respectively. The variable t represents the time it takes for an object to travel a certain distance, while L represents the distance the object travels (both of which are calibrated using instruments). Pulsed neutron sources can access a variety of different wavelengths without the need for monochromation by measuring both the scattering angle, 2θ, and the time-of-flight, t, instead of only the scattering angle.

2.9 Instrumentation

The neutron scattering measurements conducted in this thesis were obtained through the utilisation of two instruments at the ISIS spallation neutron pulsed source located in the United Kingdom, the Rutherford Appleton Laboratory: Small Angle Neutron Scattering (SANS2d)[49] and NIMROD[50] (Near- and InterMediate Range Order Diffractometer).

2.9.1 Small-Angle Neutron Scattering (SANS2d) and Near- and InterMediate Range Order Diffractometer (NIMROD)

Small-angle neutron scattering (SANS2d Fig. 2.1(A)) experiments are designed to investigate structures in solution within the range of tens to thousands of Ångstroms. As a result, SANS is capable of detecting overall structural characteristics rather than the individual atomic components of the system. The SANS2d instrument utilises neutrons with wavelengths ranging from 2 to 14 Å. These neutrons are detected by a two-dimensional detector with an area of 0.96 cm^2, positioned either 8 m or 12 m away from the material. This setup allows for the measurement of a scattering vector in the range of 0.0045 to 0.4 $Å^{-1}$.The results of these experiments are presented in Chapter 4 (PEG) and Chapter 5 (PVP).

The Near- and InterMediate Range Order Diffractometer (NIMROD), which bridges the gap between small-angle and wide-angle neutron scattering by covering small- and wide-angle wavelength (see Fig. 2.1(B)), is specifically aimed for conducting structural investigations on materials that include hydrogen, with a special focus on liquids. The system utilises high-energy neutrons with an energy range of 1 meV < E > 30,000 meV. These neutrons are produced using a combination of liquid water and liquid hydrogen moderator assembly. The

wavelengths of these neutrons fall between 0.05 and 10 Å. The instrument utilises short-wavelength neutrons to explore high-Q regions and obtain precise structural resolution. It also employs long-wavelength neutrons to investigate low-Q regions and examine correlations on greater length scales.

Neutron detection necessitates the absorption of the neutron by a nucleus located inside the detector. NIMROD utilises ZnS scintillation detectors that are doped with silver and lithium. These detectors are equipped with a wide-angle detector bank, which can measure scattering angles between 3.5° and 40°, and a low angle detector bank, which can measure scattering angles from 0.5° to 2.2°. The detection angle (and wavelength) range corresponds to a Q-range of approximately 0.02 $Å^{-1} < Q < 50\ Å^{-1}$.[50]

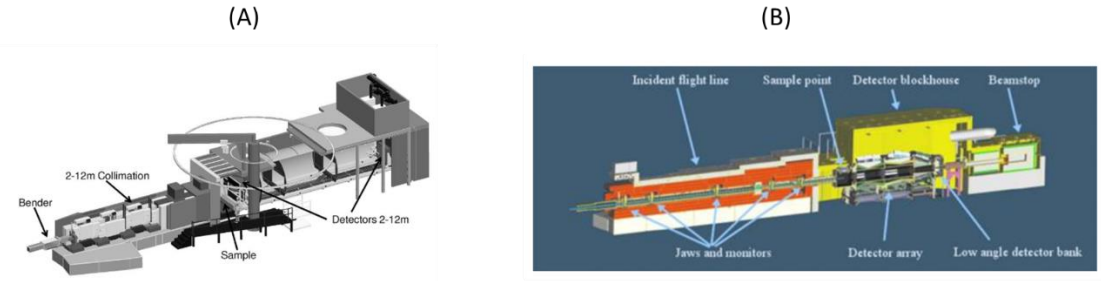


Figure 2. 1: A schematic of the (A) SAND2d and (B) NIMROD instruments used to complete measurements. Picture taken from reference[50,51].

2.10 Basic concept of Neutron Scattering Technique

The scattering length density differences in the sample determine the strength and shape of the dispersed signal in scattering techniques. Since neutrons interact with atomic nuclei, in the case of neutron scattering, the scattering pattern reflects both the distribution of various isotopes within the sample and their physical spatial arrangements. By counting the amount of scattered/reflected neutrons as a function of the wave vector, Q, neutron techniques investigate various length scales in reciprocal space as opposed to real space, therefore, the inverse distance is what determines the units of Q. The approximate relationship between Q and the distance between scattering neutrons (d) is as follows:[51–53]

$$Q \approx \frac{2\pi}{d} \qquad (2.15)$$

The Q is also related to the scattering angle, θ, and the wavelength of the incident neutron beam, λ:

$$Q = \frac{4\pi \sin\left(\frac{\theta}{2}\right)}{\lambda} \qquad (2.16)$$

Small angle neutron scattering (SANS) is a low resolution technique which has historically enabled the study of material's structural properties, including size and shape in larger length-scale (d) measurements for polymers or proteins, whose individual molecules fall into the SANS size range 1-100 nm. The 2D scattering pattern is radially averaged to give 1D plot intensity, I(Q) vs. Q, which contains the structural information and can be written as:

$$I(Q) = N_p V_p^2 (\Delta\rho)^2 F(Q) S(Q) + B_{inc} \qquad (2.17)$$

Where, N_p is the number of scattering particles, V_p is the volume of the scattering particles, $\Delta\rho$ is the difference in scattering length density (contrast) in the system between the scattering length density of particles and solvent, F(Q) is the form factor and provides intra-particle size and shape information, whilst the structure factor S(Q) provides inter-particle interactions information and is assumed to be equal to one in dilute systems, and B_{inc} is incoherent background signal.

Such data can be fitted using known approximations and mathematical modelling to reveal details like the size and shape of the system under study, which is designated by the total structure factor F(Q).

In contrast to the SANS technique, NIMROD can be used to study structural materials in small length scales 1Å<d<300Å and large values of Q which is a critical requirement for probing structural correlations in the sample with high atomic resolution. For a system with M different types of atoms, the total structure factor, F(Q), can be stated as:[50]

$$F(Q) = \sum_{i \le j}^{M} (2 - \delta_{ij}) c_i c_j b_i b_j [S_{ij}(Q) - 1] \qquad (2.18)$$

c_i and c_j are the concentrations of atoms i and j in the sample, b_i and b_j are their corresponding bound coherent scattering lengths, S_{ij} is the partial structure factor which represents the pair correlations between atoms i and j, and δ_{ij} is the Kronecker delta that prevents double counting interactions between like-atom pairings.

To understand the details of the partial structure factor for the large amounts of information from the experiment side alone is extremely difficult. The flexibility and computational

efficiency of MD simulations provide a viable way to gain insight into this information, but the long process from building the model and examining it against the experimental data to obtain a comparable one is another a significant challenge.

2.11 SANS data analysis

2.11.1 SasView

SasView was initially created by the University of Tennessee as a component of the Distributed Data Analysis of Neutron Scattering Experiments (DANSE) project, which received funding from the US National Science Foundation (NSF). However, it is now being developed as an Open Source project on GitHub and overseen by an association of scattering facilities (www.sasview.org/). The objective of the fitting process is to accurately align the data with an appropriate model and, for the systems studied here, to determine the polymer's radius of gyration. This can offer insights into the nature of the sample's structures. A background-subtracted scattering profile can be fitted into a suitable model provided with scattering vector (Q) vs intensity data I(Q) fit and parameters contained within the model will be displayed. An evaluation of the goodness-of-fit between a selected model and the data, based on a specific set of model parameters is represented by reduced Chi-squared (X^2, Chi2):

$$X^2_{Reduced} = \sum((I(Q)_{meas} - I(Q)_{calc})^2 / E(Q)^2) / (N_{pts} - N_{params}) \qquad (2.19)$$

The scattering intensity is represented by the variable I, while the error on the intensity value is denoted by E. As $X^2 \rightarrow 1$, the model fit improves.

The radius of gyration (R_g) can be determined directly from the scattering curve using the Guinier approximation.[54]

$$I(Q) = I(0) \exp\left(-Q^2 \frac{R_g^2}{3}\right) \qquad (2.20)$$

An alternative approach to determine R_g involves the computation of the space distance distribution function, P(r).[55,56]

2.11.1.2 Model fitting in SasView (poly_gaussian_coil)

The most suitable model fit in our data (PEG and PVP in dilute solutions) is poly_gaussian_coil model, which describes the empirical model characterises the scattering phenomenon that

occurs when polydisperse polymer chains are present in theta solvents or polymer melts. The model assumes a molecular weight distribution of the Schulz-Zimm type. The fitting parameters of this model is described in equation (2.21).[57]

$$I(q) = \text{scale} \cdot I(0) \cdot P(q) + \text{background} \qquad (2.21)$$

where

$$I_0 = \phi_{poly} \cdot V \cdot (\rho_{poly} - \rho_{solv})^2$$

$$P(q) = 2[(1+UZ)^{-1/U} + Z - 1] / [(1+U)Z^2]$$

$$Z = [(qR_g)^2] / (1+2U)$$

$$U = (Mw/Mn) - 1$$

$$V = M/(N_A \delta)$$

Scale is scale factor, I_0 intensity at q=0 cm^{-1}, ϕ_{poly} is the volume fraction of polymer, V is the volume of a polymer coil, ρ_{poly} is scattering length density of polymer (SLD), ρ_{solv} is scattering length density of solvent, P(q) is structural factor, R_g radius of gyration Å, Mw/Mn is polydispersity ratio, N_A is Avogadro's Number, δ is the bulk density of the polymer.

2.11.1.3 P(r) Inversion

The real space distance distribution function P(r) represents the probability distribution of distances in real space that characterises the spatial relationships within the system under analysis; the function's form can be utilised to deduce the general shape and interior structure of the scattering objects (see Fig. 2.2). Essentially, the distance distribution function P(r) and the scattering intensity I(q) convey the same information. However, the real space representation of P(r) is more intuitive and allows for easier visual analysis, often enabling the deduction of information on the shape of the particle.[58] It is mostly utilised for analysing small-angle scattering data obtained from monodisperse protein solutions, although it can also be applied to data from other types of systems.

The relationship between the scattering intensity in reciprocal space, I(q), and the real space function P(r) is as follows:

$$I(q) = 4\pi \int_0^{Dmax} P(r) \frac{sin\,(qr)}{qr}\, dr \qquad (2.22)$$

where

$$P(r) = r^2\,\gamma(r) \qquad (2.23)$$

Spherically, D_{max} is the largest dimension at which the probability function P(r) becomes zero, whereas r denotes distances in physical space. In practice, the function $P(r) = r^2\gamma(r)$ is commonly employed to represent the distribution of distances between volume elements within the particle, taking into account the excess density distribution.

$$\gamma(r) = \langle \int \Delta\rho(u)\,\Delta\rho(u+r)\,du \rangle_\omega \qquad (2.24)$$

Where γ(r) is the autocorrelation function of the excess scattering density, which is averaged spherically. The functions ρ(u) and ρ(u+r) represent the scattering densities at two specific points, u and u+r, respectively, within the particle. ω represents the volume over which the averaging is done.

The P(r) is calculated by performing an inverse[56] approach of Fourier transform of the scattering intensity, I(q).

$$P(r) = \frac{r^2}{2\pi^2} \int_0^\infty q^2 I(q) \frac{sin\,(qr)}{qr}\,dr \qquad (2.25)$$

The scattering intensity's behaviour at extremely low ($q \to 0$) and extremely high ($q \to \infty$) momentum transfer values is closely linked to the particle's overall properties.

The radius of gyration (R_g) represents the normalized second moment of the distance distribution of a particle around the centre of its scattering length density distribution.

$$Rg = \int_0^{Dmax} r^2\,P(r)\,dr \left[2 \int_0^{Dmax} P(r)\,dr\right]^{-1} \qquad (2.26)$$

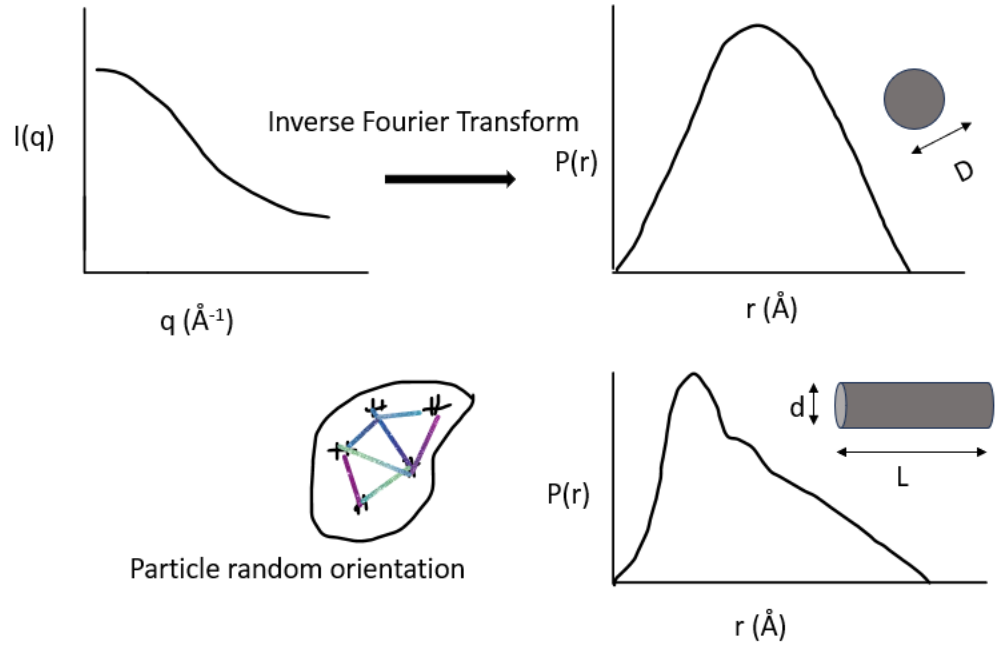


Figure 2. 2: Schematic of the real space distance distribution function P(r) using inverse scattering intensity I(q).

2.12 Methods for integrating experimental data with molecular dynamics (MD) simulations

The nature of MD simulations and experimental data determines how well they work together.[59,60] Simulations address the forward problem—creating a model that can be compared with observed data, while experimental data are transformed into a consistent model by solving the inverse problem. Researchers have proposed various methodologies to effectively integrate experimental data with molecular dynamics (MD) simulations. A recent successful method involves comparing experimental data with data derived from a back-calculated simulation.[59,61–64] If the back-calculated results are in agreement with the actual data, it provides evidence to suggest that the simulation accurately represents the experimental conditions. In such instances, the examination of the molecular dynamics simulation offers supplementary details regarding the investigated macromolecule. While this technique may seem logical, it is likely to face numerous challenges. Experimental data and molecular dynamics (MD) simulations frequently exhibit discrepancies, primarily due to the limitations of simulations such as inaccuracies in force fields and sampling. Furthermore, even if the data obtained by back-calculation aligns with the experimental data, there is no assurance that the

simulation accurately represents the conditions observed in the experiment. This problem is especially evident when dealing with experiments that yield time and ensemble-averaged data, such as SAS, and NMR.

Experimental data can be incorporated directly into a simulation as a constraint. The SANS-driven MD method was pursued in Chapter 5 of this thesis. Calibrating MD simulations based on ensemble-averaged experimental data might result in overfitting, particularly when dealing with flexible molecules such as PEG.

2.12.1 Computing SANS curve from all-atom MD simulation

The SANS curve predictions were calculated in two different ways: using all-atom MD trajectories[65] and using Chen and Hub approach[62] with (SANS-driven MD) and without including the experimental SANS data during the computing SANS spectra. The first approach is implemented as a programmed module within the GROMACS MD package, which called gmx_sans. This module enables calculating the average of the spectrum across the full trajectory or a specific portion of it, and we used this method to calculate the spectra for the polymer chains only (no water molecules counts in the solvation shell) along with construction of the histogram of the pair correlation function P(r).

Calculating the scattering intensity I(q) can be obtained using Eq. 2.27.

$$\mathrm{I(q)} = \sum_r P(r)\ \frac{sinq(r)}{q(r)} \qquad (2.27)$$

Where the P(r) is the pair correlation function, which is written in the form:

$$\mathrm{P(r)} = \frac{\sum_i G_i(r)}{\sum_i \int_0^\infty G_i(r)dr} \qquad (2.28)$$

i denote for runs through all presented conformations, and G(r) is represented in the form of the sum over particles using the coherent scattering lengths bi and bj as coefficients[65,66]:

$$\mathrm{G(r)} = \sum_{i,j} bi\ bj \qquad (2.29)$$

2.12.2 SANS-Driven Simulation

We have examined the application of molecular dynamics simulations to compute the SANS curves. If the curve obtained by back-calculating from the molecular dynamics (MD) simulation aligns with the experimental curve, it can be inferred that the simulation accurately

represents the atomic structure of the macromolecule under investigation. Nevertheless, there is often a lack of concurrence between experimental and computed curves, occasionally because of experimental difficulties, but more frequently as a result of simulation deficiencies. Several approaches have been devised to address the limitations of simulations by including small-angle scattering (SAXS/SANS) data as a constraint in molecular dynamics (MD) simulations.[63,67–69] The techniques devised by Chen and Hub (Ref. 63) and Hermann and Hub (Ref. 64) have been utilised in Ch 5 of this thesis. Conceptually, SANS is very similar to SAXS, and the mathematical foundation of neutron scattering is essentially the same as X-ray scattering, with the only difference being the substitution of X-ray atomic factors with neutron scattering lengths.[66,70,71]

Further information regarding this approach, along with its verification, may be located in Reference 62. Essentially, the connection between the MD simulation and the desired (target) curve I_{exp} is established through a hybrid energy.

$$E_{hybrid}(R; I_{exp}) = E_{MD}(R) + E_{exp}^{(1)}(R; I_{exp}) \qquad (2.30)$$

$E_{MD}(R)$ represents the energy obtained from the force field for the conformation of the molecule R. The expression $E_{exp}^{(1)}(R; I_{exp})$ represents an energy bias that is determined from experimental calculated via:

$$E_{exp}^{(1)}\left(R; I_{exp}\right) = \frac{k_r k_B T}{n_q} \sum_{i=1}^{n_q} \frac{[\, I_c(q_i, R) - I_{exp}(q_i)]^2}{\sigma_i^2} \qquad (2.31)$$

The force constant is represented by k_r, the Boltzmann constant by k_B, and the temperature is indicated by T. The quantity of intensity points distributed across the q-range is represented by n_q, and $I_c(q_i, R)$ represents the back-calculated scattering intensity using simulation coordinates R. The uncertainty expressed by σ_i encompasses experimental error, statistically determined errors, and systematic errors arising from the uncertainty of the buffer density.

2.13 NIMROD Data Analysis

2.13.1 Elastic Scattering

Scattering refers to the phenomenon in which radiation is redirected or dispersed from its original path due to irregularities in the medium it passes through. During a neutron scattering experiment, a small portion of the neutron beam that hits the sample is scattered, while the

remaining portion passes through the sample. This transmission may involve attenuation (inelastic), which refers to the absorption of neutrons by the sample, or multiple scattering, which occurs when the radiation is scattered more than once before leaving the sample.

Each element, as well as every isotope of each element, possesses two distinct scattering lengths: a coherent scattering length b_{coh}, an elastic, and an incoherent scattering length b_{inc}, inelastic. The coherent scattering refers to the process in which incident waves interact with atoms or molecules as a whole, resulting in constructive interference and maintaining the phase relationships. This interference, (which includes the interferences from intra and intermolecular scattering), provides the structural information for the sample. While for the incoherent scattering, the incident waves interact with individual electrons within the atom or molecule, resulting in random or destructive interference. The coherent scattering is evaluated in diffraction measurements, but the signal is additionally accompanied by an unstructured background due to incoherent scattering. In this work, the focus is on the utility of diffraction due to elastic, coherent scattering to understand aqueous solutions.

2.13.1.1 Correction and Calibration of Raw Data

In order to gain structural information about a system, it is important to extract the distinct component of the cross-section. It necessitates the examination of the elastic, coherent scattering of neutrons. Furthermore, the raw data of a diffraction experiment may exhibit various scattering effects as mentioned above, necessitating the need for correction.

In the examination of solid samples with a constant atomic structure, the static approach is used. This approximation assumes that the effects of attenuation, multiple scattering, and inelastic scattering are insignificant. The research conducted in this thesis focuses on liquid samples, so it is necessary to adjust the raw diffraction data (the original data quantified as the count rate per microsecond recorded by each detector and each count corresponds to the detection of one neutron) to account for these contributions in order to obtain the structural factor. These must be taken into consideration when merging the data from the detectors.

The datasets are adjusted using the Gudrun software.[72] The Gudrun routines utilise the algorithms from the ATLAS suite of programmes.[73] Diffraction measurements must be obtained from the following sources to guarantee that the required corrections may be made: the empty instrument, the empty cells, a vanadium standard cell, and each cell containing the sample. The correction procedure followed using Gudrun is briefly listed below:

- All datasets have been adjusted to account for detector dead-time, or the period of time between two events where no recording takes place. Detectors are grouped together and their standard deviations are compared. By categorising detectors and analysing the standard deviations within each category, detectors that have low standard deviation (indicating weak counting) or detectors that produce noisy data (indicated by big standard deviation) are eliminated. Subsequently, every dataset is normalised to eliminate the influence of fluctuating incident neutron flux levels.
- Attenuation (inelastic) effects have been corrected for all datasets. This includes applying Soper and Egelstaf's absorption corrections[74] and subtracting multiple scattering from datasets for both empty cells and cells containing samples, which can be approximated using the theory described here.[75]
- To convert to an absolute scale of barns $atom^{-1}$ sr^{-1}, datasets for cells and datasets of sample containing cells are normalised to the vanadium standard. Vanadium is utilised for equipment calibration since its scattering is nearly totally nuclear spin incoherent and its differential cross section can be determined with high accuracy.
- Subtracting the corrected and calibrated scattering from its cell in order to get the scattering from the sample alone.
- Correction for the background caused by scattering of particles with themselves and the effects of inelastic collisions. The self-scattering background is usually observed as unphysical peaks at low r, which are further compounded by truncation waves caused by the finite value of Q_{max} following Fourier transform of the data. In Gudrun, Soper[74] has devised an alternative correction that utilises a dynamical scattering law based on a harmonic oscillator model.

2.13.2 Dissolve-software (total neutron scattering simulation)

After concluding a neutron diffraction experiment and correcting the raw data, the data must be analysed in order to retrieve detailed atomic-level information about the system. It is impossible to acquire all partial structural elements for a given system just through experimental data. To acquire these results, the solution lies in employing computational modelling, which may provide a simulated structure that accurately replicates the one depicted in the data. Dissolve is a classical simulation algorithm[76] designed to provide representative atomic configurations that are consistent with reference data provided, aiding in the analysis of experimental scattering (corrected raw diffraction) results. The software is currently under

development and utilises a methodology similar to that employed in the Empirical Potential Structure Refinement (EPSR) code developed by A. K. Soper.[77] However, it has been constructed from the ground up with a specific focus on accommodating complex and large target systems. Although for more complicated systems can also benefit from its application, its main application lies in disordered systems, such as those made up of or containing liquids and glasses**.** This work is the first attempt at using Dissolve to determine highly localised structure of polymers in solution from total neutron scattering data.

2.14 Nuclear Magnetic Resonance (NMR)

NMR spectroscopy is a unique method for examining the precise stereochemical structure of a polymer, including its tacticity, at a microstructural level.[78–80] Proton nuclear magnetic resonance (^{1}H NMR) is frequently employed to analyse the structure of organic compounds, including polymers. However, when it comes to determining the stereochemical structures of polymers, ^{13}C nuclear magnetic resonance (^{13}C NMR) spectroscopy is often the preferable method. In ^{13}C NMR investigations, advanced decoupling techniques (coupling between adjacent atoms) can be used to simplify the spectra and improve the visibility of carbon signals. This can help assign and understand peaks more easily, especially in polymers with complicated structures where overlapping signals may arise. In this thesis we used ^{13}C NMR to confirm the tacticity type for PVP sample.

2.14.1 Tacticity

Vinyl monomers with an attached substituent group (e.g. -CH_2-CH_2R-) display a phenomenon known as “pseudo-asymmetric” behaviour along the polymer chain’s backbone, as shown in Figure 2.3. Consequently, the presence of these “pseudo-asymmetric” centres in the individual units of the polymer results in the polymer having a stereoregular nature. The three main stereoregularities of the monomeric units classify polymers as either isotactic, syndiotactic, or atactic.[81]

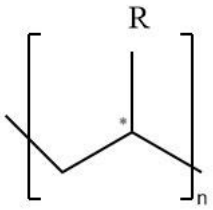


Figure 2. 3: Vinyl monomeric units with the pseudo-asymmetry

Isotactic polymers are characterised by having a single sort of monomeric configuration, where the side groups are all aligned on one side of the polymer backbone. This is seen in Figure 2.4. Due to the alignment of the pendant side groups on one side of the chain, the polymer chains can adopt a helical shape, enabling the neighbouring polymer chains to orderly pack together. The feature mentioned is the primary factor that enables purely isotactic polymers to easily form crystals.

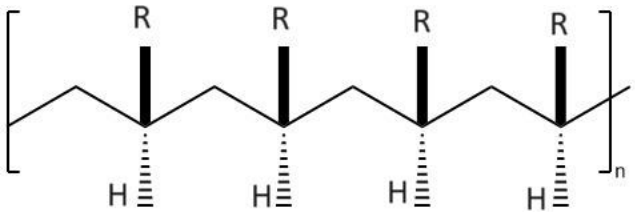


Figure 2. 4: An isotactic polymer

Syndiotactic polymers are characterised by monomeric units that have pendent side chains positioned on alternating sides of the polymeric backbone. This is illustrated in Figure 2.5. Like isotactic polymers, syndiotactic polymers can easily crystallise because their chains can form helices, allowing for effective packing of adjacent polymer chains.

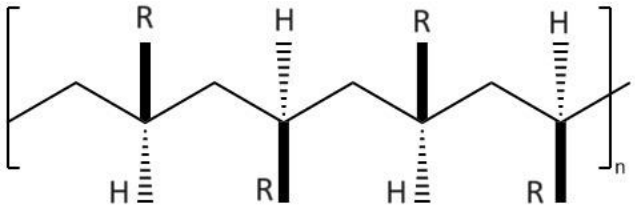


Figure 2. 5: A syndiotactic polymer

Atactic polymers differ significantly from isotactic and syndiotactic polymers as they lack any regularity in the arrangement of monomer sequences inside the polymer chain (Figure 2.6). Atactic polymers lack the stereoregular monomeric units seen in isotactic and syndiotactic polymers, resulting in a reduced propensity to crystallise. As a result, atactic polymers are often described as amorphous.

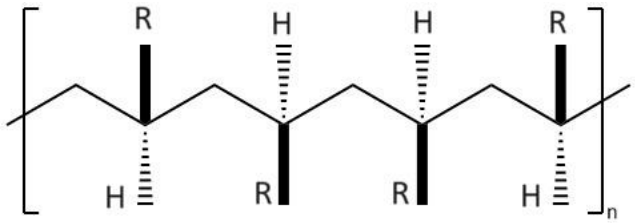


Figure 2. 6: An Atactic polymer

Typically, the utilisation of anionic and cationic catalysts in polymerizations is preferred for the production of polymers that exhibit stereoregularity and have the ability to form crystals. In contrast, free radical polymerizations typically yield amorphous polymers with atactic structure.

Polymers can also create stereo-blocks inside their chains. Within a polymer, there can be monomeric blocks that display isotactic properties, where the side groups align on the same side of the polymer chain. At the same time, other monomeric blocks in the same polymer chain can exhibit syndiotactic properties, where the adjacent monomeric units have side groups aligned on the opposite side of the chain.

A diad refers to the smallest blocks inside the polymeric chain, consisting of two neighbouring monomeric units. Subsequently, the subsequent monomeric block in size is a triad, which is then followed by a tetrad, and so on. Diads can be categorised as meso (m) or racemo (r). Meso diads refer to two neighbouring monomeric units in a polymer chain where the side groups are positioned on the same side. In contrast, racemo diads are two adjacent monomeric units with side groups positioned on opposite sides of the polymer chain.

To elaborate on this idea, triad tacticity sequences refer to three consecutive monomeric units in a polymer chain that can be designated in a similar way as diad sequences (Figure 7). The mm and rr sequences can alternatively be denoted as isotactic and syndiotactic blocks, respectively (Figure 2.7 (A) and (B)). The rm and mr triads are classified as atactic (or heterotactic) blocks (Figure 2.7 (C) and (D)).

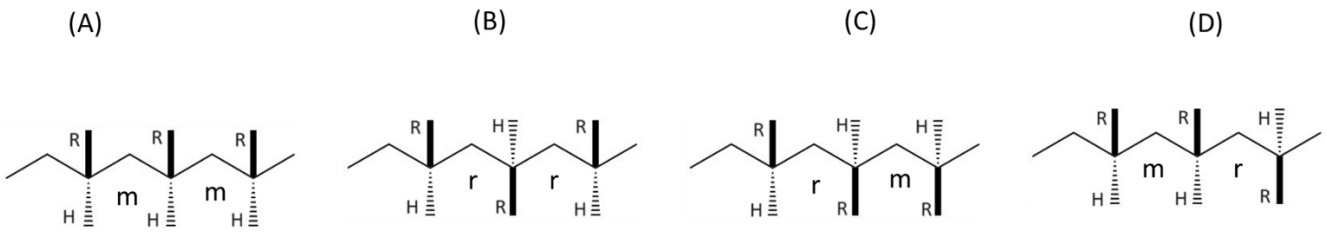


Figure 2. 7: Triad sequences: (A) isotactic block, (B) syndiotactic block, (C) and (D) atactic blocks.

Information about the structural tacticity of the PVP polymer can be derived from the number and intensities of the split resonance signals of the carbon nuclei (^{13}C NMR), assuming certain assumptions about the chain growth mechanism.[82]

2.15 Gel Permeation Chromatography

When characterising polymers, it is crucial to take into account their size distribution and molecular dispersity, as well as their average molecular weight. Polymers can be characterised using many criteria to determine their molecular weight. Gel permeation chromatography (GPC) has been employed to ascertain the mean molecular weight (Mw) of polyvinyl pyrrolidone (PVP), assuring accurate model representation of the polymer chain with respect to the obtained experimental value. During the passage of the sample through the column, the polymer molecules are separated according to their size, with larger molecules eluting prior to smaller ones. Elution times are compared to that for known standard polymers, in order to obtain MW and polydispersity information.

References

1 F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Sons Ltd, England, 2nd edn., 2007.

2 A. Cauchy, M´ethode g´en´erale pour la r´esolution des syst`emes d'´equations simultan´ees, *Comptes Rendus Hebd. Seances Acad. Sci*, 1847, **25**, 536–538.

3 J. C. Meza, Steepest Descent, *Wiley Interdiscip. Rev. Comput. Stat.*, 2010, **2**, 719–722.

4 G. Bussi, D. Donadio and M. Parrinello, Canonical sampling through velocity rescaling, *J. Chem. Phys.*, 2007, **126**, 014101.

5 W. G. Hoover, Canonical dynamics: Equilibrium phase-space distributions, *Am. Phys. Soc.*, 1985, **31**, 1695–1697.

6 H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola and J. R. Haak, Molecular dynamics with coupling to an external bath, *Am. Inst. Phys.*, 1984, **81**, 3684–3690.

7 M. Parrinello and A. Rahman, Polymorphic transitions in single crystals : A new molecular dynamics method, *J. Appl. Phys.*, 1981, **52**, 7182–7190.

8 M. James, T. Murtola, R. Schulz, J. C. Smith, B. Hess and E. Lindahl, High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX*, 2015, **2**, 19–25.

9 R. W. Hockney, the Potential Calculation and Some Applications, *Methods Comput. Phys.*, 1970, **9**, 135–211.

10 W. C. Swope, H. C. Andersen, P. H. Berens and K. R. Wilson, A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules : Application to small water clusters, *J. Chem. Phys.*, 1982, **76**, 637–649.

11 S. Miyamoto and P. A. Kollman, SETTLE : An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models, *J. Comput. Chem.*, 1992, **13**, 952–962.

12 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, LINCS : A Linear Constraint Solver for Molecular Simulations, *J. Comput. Chem.*, 1997, **18**, 1463–1472.

13 J. Leszczynski, *Handbook of computational chemistry*, Springer Science & Business

Media, 2012.

14 M. Karplus and G. A. Petsko, Molecular dynamics simulations in biology, *Nature*, 1990, **347**, 631–639.

15 A. D. Mackerell, Empirical Force Fields for Biological Macromolecules : Overview and Issues, *J. Comput. Chem.*, 2004, **25**, 1584–1604.

16 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations, *J. Comput. Chem.*, 1983, **4**, 187–217.

17 M. Saito, Molecular dynamics simulations of proteins in solution: Artifacts caused by the cutoff approximation, *J. Chem. Phys*, 1994, **101**, 4055–4061.

18 M. Saito, Molecular Dynamics / Free Energy Study of a Protein in Solution with All Degrees of Freedom and Long-Range Coulomb Interactions, *J. Phys. Chem*, 1995, **99**, 17043–17048.

19 T. Darden, D. York and L. Pedersen, Particle mesh Ewald : An Nlog (N) method for Ewald sums in large systems, *J. Chern. Phys*, 1993, **98**, 10089–10092.

20 T. Darden, H. Lee and L. G. Pedersen, A smooth particle mesh Ewald method, *J. Chem. Phys*, 1995, **103**, 8577–8593.

21 C. J. Cramer and D. G. Truhlar, Implicit Solvation Models : Equilibria , Structure , Spectra , and Dynamics, *Chem. Rev.*, 1999, **99**, 2161−2200.

22 B. Honig and A. Nicholls, Classical Electrostatics in Biology and Chemistry, *Science (80-.).*, 1995, **268**, 1144–1149.

23 P. Beroza and D. A. Case, Calculations of Proton-Binding Thermodynamics in Proteins, *Methods Enzymol.*, 1998, **295**, 170–189.

24 J. D. Madura, M. E. Davist, K. Michael, R. C. Wades, B. A. Luty and J. A. Mccammon, Biological Applications of Electrostatic Calculations and Brownian Dynamics Simulations, *Comput. Chem.*, 1994, **5**, 229–67.

25 M. K. Gilson, Theory of electrostatic interactions in macromolecules, *Curr. Opin. Struct. Biol*, 1995, **5**, 216–223.

26 M. Scarsi, J. Apostolakis and A. Caflisch, Continuum Electrostatic Energies of

Macromolecules in Aqueous Solutions, *J. Phys. Chem*, 1997, **5639**, 8098–8106.

27 R. A. Y. Luo, L. David and M. K. Gilson, Accelerated Poisson – Boltzmann Calculations for Static and Dynamic Systems, *J. Comput. Chem*, 2002, **23**, 1244–53.

28 A. Onufriev, *Implicit Solvent Models in Molecular Dynamics Simulations: A Brief Overview*, Annual Reports in Computational Chemistry, 2008.

29 W. C. Still, A. Tempczyk, R. C. Hawley and T. Hendrickson, Semianalytical treatment of solvation for molecular mechanics and dynamics, *J. Am. Chem. Soc.*, 1990, **112**, 6127–9.

30 J. Ramstein and R. Laveryt, Energetic coupling between DNA bending and base pair opening, *Proc. Nail. Acad. Sci. USA*, 1988, **85**, 7231–7235.

31 L. David, R. A. Y. Luo and M. K. Gilson, Comparison of Generalized Born and Poisson Models : Energetics and Dynamics of HIV Protease, *J. Comput. Chem*, 2000, **21**, 295–309.

32 L. Wang, B. E. Hingerty, A. R. Srinivasan, W. K. Olson and S. Broyde, Accurate Representation of B-DNA Double Helical Structure with Implicit Solvent and Counterions, *Biophys. J.*, 2002, **83**, 382–406.

33 A. E. Simulations, K. M. Kosikov, A. A. Gorin, X. Lu, W. K. Olson and G. S. Manning, Bending of DNA by Asymmetric Charge Neutralization: All-Atom Energy Simulations, *J. Am. Chem. Soc*, 2002, **124**, 4838–4847.

34 J. A. Barker and R. O. Watts, Monte Carlo studies of the dielectric properties of water-like models, *Mol. Phys.*, 1973, . **26**, 789–792.

35 R. O. Watts, Monte Carlo studies of liquid water by, *Mol. Phys.*, 1974, **28**, 1069–1083.

36 P. B. Petra Florova, Petr Sklenovsky and M. Otyepka, Explicit Water Models Affect the Specific Solvation and Dynamics of Unfolded Peptides While the Conformational Behavior and Flexibility of Folded Peptides Remain Intact, *J. Chem. Theory Comput*, 2010, **6**, 3569–3579.

37 M. Orozco and F. J. Luque, Theoretical Methods for the Description of the Solvent Effect in Biomolecular Systems, *Chem. Rev*, 2000, **100**, 4187−4225.

38 G. Bertrand, A reappraisal of what we have learnt during three decades of computer

simulations on water, *J. Mol. Liq.*, 2002, **3**, 219–260.

39 M. W. Mahoney and W. L. Jorgensen, A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions, *Chem. Phys.*, 2000, **112**, 28–32.

40 M. W. Mahoney and W. L. Jorgensen, Diffusion constant of the TIP5P model of liquid water, *Chem. Phys.*, 2001, **114**, 2001–2004.

41 P. G. Kusalik and M. Svishchev, The Spatial Structure in Liquid Water, *Science (80-.).*, 1994, **265**, 1219–1222.

42 C. Vega and J. L. F. Abascal, Relation between the melting temperature and the temperature of maximum density for the most common models of water, *Chem. Phys. 123,* 2005, **123**, 144504.

43 C. Vega, E. Sanz and J. L. F. Abascal, The melting temperature of the most common models of water, *Chem. Phys. 122,* 2005, **122**, 114507.

44 L. A. Baez and P. Clancy, Existence of a density maximum in extended simple point charge water, *J. Chern. Phys*, 1994, **101**, 9837–9840.

45 C. Vega, C. Mcbride, E. Sanz and J. L. F. Abascal, Radial distribution functions and densities for the SPC/E, TIP4P and TIP5P models for liquid water and ices I, *Phys. Chem. Chem. Phys.*, 2005, **7**, 1450–1456.

46 P. Mark and L. Nilsson, Structure and Dynamics of the TIP3P , SPC , and SPC / E Water Models at 298 K, *J. Phys. Chem. A*, 2001, **105**, 9954–9960.

47 J. Zielkiewicz, Structural properties of water: Comparison of the SPC, SPCE, TIP4P, and TIP5P models of water, *Chem. Phys.*, 2005, **123**, 104501.

48 S. P. K. Pathirannahalage, N. Meftahi, A. Elbourne, C. G. Alessia, T. C. Le, Q. A. Besford and A. J. Christofferson, A systematic comparison of the structural and dynamic properties of commonly used water models for molecular dynamics simulations, *J. Chem. Inf. Model.*, 2021, **61**, 4521–4536.

49 R. K. Heenan, S. M. King, D. S. Turner and J. R. Treadgold, SANS2D at the ISIS Second Target Station, *Proc. ICANS-XVII*, 2006, 780–785.

50 D. T. Bowron, A. K. Soper, K. Jones, S. Ansell, S. Birch, J. Norris, L. Perrott, D.

Riedel, N. J. Rhodes, S. R. Wakefield, A. Botti, M. A. Ricci, F. Grazzi and M. Zoppi, NIMROD: The Near and InterMediate Range Order Diffractometer of The ISIS Second Target Station, *Rev. Sci. Instrum.*, 2010, **81**, 033905.

51 L. A. Feigin and D. I. Svergun, *Structure Analysis by Small-Angle X-Ray and Neutron Scattering*, Plenum press, 1987.

52 K. J. Edler and D. T. Bowron, Combining wide-angle and small-angle scattering to study colloids and self-assembly, *Curr. Opin. Colloid Interface Sci.*, 2015, **20**, 227–234.

53 S. E. Rogers, A. E. Terry, M. J. Lawrence, J. Eastoe, J. T. Cabral and A. Chan, Soft matter at ISIS, *Mater. Today*, 2009, **12**, 92–99.

54 A. Guinier, G. Fournet, C. B. Walker and G. H. Vineyard, Small-Angle Scattering of X-Rays, *Phys. Today*, 1956, **9**, 38–39.

55 O. Glatter, A New Method for the Evaluation of Small-Angle Scattering Data, *J. Appl. Cryst.*, 1977, **10**, 415–421.

56 P. B. Moore, Small-angle scattering. Information content and error analysis, *J. Appl. Crystallogr.*, 1980, **13**, 168–175.

57 poly_gauss_coil, https://www.sasview.org/docs/user/models/poly_gauss_coil.html#poly-gauss-coil, (accessed 30 August 2021).

58 D. I. Svergun and M. H. J. Koch, Small-angle scattering studies of biological macromolecules in solution, *Reports Prog. Phys.*, 2003, **66**, 1735–1782.

59 S. Bottaro and K. Lindorff-larsen, Biophysical experiments and biomolecular simulations: A perfect match?, *Science (80-.).*, 2018, **361**, 355–360.

60 R. D. Schaeffer, A. Fersht and V. Daggett, Combining experiment and simulation in protein folding : closing the gap for small model systems, *Curr. Opin. Struct. Biol*, 2008, **18**, 4–9.

61 J. S. Hub, Interpreting solution X-ray scattering data using molecular simulations, *Curr. Opin. Struct. Biol.*, 2018, **49**, 18–26.

62 P. C. Chen and J. S. Hub, Validating solution ensembles from molecular dynamics

simulation by wide-angle X-ray scattering data, *Biophys. J.*, 2014, **107**, 435–447.

63 M. R. Hermann and J. S. Hub, SAXS-Restrained Ensemble Simulations of Intrinsically Disordered Proteins with Commitment to the Principle of Maximum Entropy, *J. Chem. Theory Comput*, 2019, **15**, 5103–5115.

64 M. Bonomi, C. Camilloni, A. Cavalli and M. Vendruscolo, Metainference : A Bayesian inference method for heterogeneous systems, *Sci. Adv.*, 2016, **2**, e1501177.

65 A. V. Shvetsov, A. E. Schmidt, D. V. Lebedev and V. V. Isaev-Ivanov, Method for calculating small-angle neutron scattering spectra using all-atom molecular dynamics trajectories, *J. Surf. Investig.*, 2013, **7**, 1124–1127.

66 V. F. Sears, Neutron scattering lengths and cross sections Special Feature Neutron scattering lengths and cross sections, *Neutron news*, 1992, **3**, 26–37.

67 P. Chen and J. S. Hub, Interpretation of Solution X-Ray Scattering by Explicit-Solvent Molecular Dynamics, *Biophysj*, 2015, **108**, 2573–2584.

68 D. Kimanius, I. Pettersson, G. Schluckebier, E. Lindahl and M. Andersson, SAXS-Guided Metadynamics, *J. Chem. Theory Comput.*, 2015, **11**, 3491–3498.

69 A. Bjo, S. Niebling, M. Marcellini, D. Van Der Spoel and S. Westenho, Deciphering Solution Scattering Data with Experimentally Guided Molecular Dynamics Simulations, *J. Chem. Theory Comput.*, 2015, **11**, 780–787.

70 P. Chen, R. Shevchuk, F. M. Strnad, C. Lorenz, L. Karge, R. Gilles, A. M. Stadler, J. Hennig and J. S. Hub, Combined Small-Angle X - ray and Neutron Scattering Restraints in Molecular Dynamics Simulations, *J. Chem. Theory Com*, 2019, **15**, 4687−4698.

71 L. A. Feigin and D. I. Svergun, *Structure Analysis by Small-Angle X-Ray and Neutron Scattering*, Springer, 1987.

72 A. K. Soper, *GudrunN and GudrunX: Programs for Correcting Raw Neutron and X-ray Diffraction Data to Differential Scattering Cross Section*, 2011.

73 A. K. Soper, W. S. Howells and A. C. Hannon, *Analysis of Time-of-Flight Diffraction Data from Liquid and Amorphous Samples*, Oxon, UK, 1989, vol. 445543.

74 A. K. Soper, Inelasticity corrections for time-of-flight and fixed wavelength neutron

diffraction experiments, *Mol. Phys.*, 2009, **107**, 1667–1684.

75 A. K. Soper, Multiple scattering from an infinite plane slab, *Nucl. Instruments Methods Phys. Res.*, 1983, **212**, 337–347.

76 T. Youngs, Dissolve: Next Generation Software for The Interrogation of Total Scattering Data by Empirical Potential Generation, *Mol. Phys.*, 2019, **117**, 3464–3477.

77 A. K. Soper, Partial Structure Factors from Disordered Materials Diffraction Data: An Approach Using Empirical Potential Structure Refinemen, *Phys. Rev. B*, 2005, **72**, 104204.

78 L. Dong, D. J. T. Hill, J. H. O. Donnell and A. K. Whittaker, Configurational Assignments for Poly (methacrylonitrile) Using Double-Quantum-Filtered Phase-Sensitive COSY and Proton-Detected 1H-13C Shift-Correlated NMR Spectroscopies, *Macromolecules*, 1994, **27**, 1830–1834.

79 A. Bulai, M.-L. J. Jimeno and J. S. Román, Stereochemical Structure of Poly (cyclohexyl acrylate) Studied by One-Dimensional and Two-Dimensional 13C-1H Spectroscopy, *Macromolecules*, 1995, **28**, 7363–7369.

80 M. Suchopárek and J. Spéváéek, Characterization of the Stereochemical Structure of Poly(ieri-butyl acrylate) by One- and Two-Dimensional NMR Spectroscopy, *Macromolecules*, 1993, **26**, 102–106.

81 R. B. Clark and R. L. McMullen, in *Handbook of Pyrrolidone and Caprolactam Based Materials: Synthesis, Characterization and Industrial Applications*, 2021, vol. 6, pp. 1469–1534.

82 V. I. Dubrovin and V. P. Panov, Carbon-13 NMR investigation of microtactic structure of polyvinylpyrrolidone, *J. Appl. Spectrosc.*, 1980, **33**, 754–757.

Chapter 3

Synergies Between All-Atom Molecular Dynamic (MD) Simulations and Small-and Wide-Angle Neutron Scattring (SANS/WANS) for Polyethylene Glycol (PEG) in Dilute Solution

3.1 Introduction

The utilization of scattering diffractions for the production of very detailed images of organized single crystals of proteins, polymers and other macromolecules, in the solid state, has tremendously helped modern biophysics. Molecules in the solid form fill the space with small spaces either in a regular array (crystalline) or at random (amorphous). The molecules are in close proximity to other molecules. In contrast, each molecule in a solution is surrounded by solvent molecules, and it is sensitive to these local environments. The number of approaches for studying the structure and dynamics of polymeric solution systems has not increased at the same rate as the number of reports of proteins innovative breakthroughs. Neutron scattering is considered to be a complementary experimental tool for analyzing such structures, due to the lack of radiation damage, the potential for contrast variation by solvent exchange (H_2O/D_2O) or selective deuteration, and its sensitivity to the light elements, like hydrogen, which enable the determination and understanding of the structure of polymeric solutions.[1–3]

Small and wide-angle neutron scattering techniques are proven accurate methods for characterizing solution structures in different length scales. The SANS generally contains information about the overall polymer structure but does not solve for individual chain conformation, while WANS provides more detailed information about atomic and molecular arrangements. The interpretation of a polymeric structure solution from these experimental techniques is very challenging due to the low data content. However, the molecular dynamics (MD) simulation approach can be a viable tool for predicting conformational ensembles to understand the dynamic and structural behavior in aqueous solutions. In order to predict accurate MD structural properties that can match the experimental data requires lots of validation process which includes solvent model, backbone force field and simulation timescale.

Polyethylene glycol (PEG) is a simple enough polymer to attempt the MD validation against neutron scatting data hoping to obtain a verified MD which can help to understand the dynamics and structural behavior of polymers in aqueous solutions. In this study, we first constructed the PEG model with different numbers of repeat units and examined the behavior of the polymer chains using different force fields in implicit solvents. Additionally, explicit MD simulations for PEG (6-292 mer) in water were performed on GROMACS. The radius of gyration was validated against the SANS derived value for PEG292 with OPLS-AA force field in two water models, SPC and TIP3P. Two different MD simulations (for PEG292), one performed using the default box size in GROMACS and the second with a box sized to mimic the experimental concentration, were carried out and their outputs analyzed to enable better insights into the conformational ensembles. SANS curves were predicted from the MD for selected conformations with different R_g (min, max, average) and also averaged across all analyzed MD frames, and compared to the experimental SANS patterns. Furthermore, here we have presented the implementation of analysis of disordered deuterated/hydrogenated-PEG (M_w ≈13,000 g/mol) in aqueous solutions (D_2O and H_2O) using the Dissolve analysis software,[4] making it the first use of the software for the analysis of dilute polymer solution scattering. Dissolve has been designed to investigate the integration of total neutron scattering data from the Near and InterMediate Range Order Diffractometer (NIMROD) and applies the Empirical Potential Structure Refinement (EPSR) to match simulated MD to the experimental data.

3.2 Sample preparation and neutron scattering experiment

Protonated polyethylene glycol (h-PEG, Mw=13,000 g/mol) and deuterated polyethylene glycol (d-PEG, Mw=14,000 g/mol) were purchased from Agilent and PolymerSource, respectively. The polymers were dissolved in water and deuterated water, and prepared by mass, corrected for solvent density of D_2O, to give same concentration ≈ 4% w/v, which is below the overlap concentration, see Table.3.1 for the full sample preparation.

SANS measurements utilising the SANS2D[5] equipment were conducted at ISIS Neutron Facility, RAL, STFC, UK. Five samples were tested, h/D, d/H, d/DH, dh/D and dh/DH. Hellma quartz cuvettes in the shape of a banjo were filled with the samples. The data were measured on a 1 m^2 detector to give a Q -range of 0.0045-0.7 $Å^{-1}$ with wavelength $1.75Å \leq \lambda \leq 16.5Å$. Using a traditional data reduction process[6] that adjusted for solvent, empty cell, and

transmission measurements at room temperature, the raw data was compiled into the data used for analysis.

Using NIMROD[7] (RB1910412), seven samples—h/D, h/H, d/H, d/D, d/DH, dh/D, and dh/DH—were examined. The samples were contained in Ti/Zr flat plate cell containers with about 1 mm-thick window, and a 30 x 30 mm incident beam size was used. The scattering vector (Q) range of 0.02-50 $Å^{-1}$ was achieved with broad wavelength ($0.04Å \leq \lambda \leq 10.0Å$).

Table 3. 1: Details of samples prepared for neutron scattering experiments, for h/d-polyethylene glycol in H_2O or D_2O water.

PEG/solution Isotopologue	Mass polymer (g)	Mass solvent (g)	Total Mass (g)	wt%	SLD solvent (10^{10} cm^{-2})	SLD dry polymer (10^{10} cm^{-2})
d/H	0.12	2.88	3.00	4.02	-0.56	5.75
d/D	0.12	3.20	3.32	3.60	6.34	5.75
h/D	0.12	3.21	3.33	3.62	6.34	0.56
h/H	0.12	2.88	3.00	4.04	-0.56	0.56
	Soln1 (ml)	Soln2 (ml)				
d/D(50)H(50)	0.72	0.83	1.5474		2.89	5.75
dh/D(50)H(50)	0.74	0.82	1.5523		2.89	3.16
dh/D	0.82	0.84	1.6588		6.34	3.16

H and D: H_2O and D_2O water; DH: mixture of H_2O and D_2O water; h and d: protonated and deuterated PEG.

3.3 Results and Discussion

3.3.1 Small Angle Neutron Scattering Fitting and Analysis

Changes in the experimental SANS intensity I(q) on the linear axes vs Q (1/Å) on the Log_{10} axes for 4 wt % h-PEG and d-PEG varying with deuterium content in water is shown in Fig. 1 (A) (background subtracted scattering examples are given for h/D, d/H, d/DH, dh/DH and d/DH). Measurements were made for the 4 wt% polymer solution (for a solution, therefore, that has a concentration below the overlap concentration)[8], so that interparticle interactions would be minimal and S(Q) can be taken as 1. The overlap concentration is the concentration at which polymer chains start to aggregate and overlap and gives either a peak in the scattering (for charge systems) or an upturn at low Q (for uncharged and aggregating systems), which

affects revealing the dimension and the R_g of the chain. The h-PEG in H_2O is not presented here due to the scattering from H_2O overwhelms the SANS signal as there is low contrast between the polymer and the solvent, resulting in flat and structurally meaningless SANS data. In the SANS data, which is sensitive only to the overall size and shape (R_g) of the polymer chains, the change in deuteration across the contrast series results in changes to the intensity of the observed scattering.[9] One can see from Fig.1(A) that the d/H sample scatters more than the h/D because of the different contrast conditions and the increased incoherent scattering from the H_2O solvent, compared to D_2O. Moreover, the scattering intensity for d/H and d/DH appears differently due to the contrast variation between the d-PEG in D_2O/H_2O and d-PEG in H_2O giving the latter more intense scattering. The dh/DH and dh/D, contains two different polymers and show similar curve pattern with small variation in contrast. In this study, the analysis will be focused on the polymer sample, h/D to validate MD.

The SANS data analysis was conducted using SasView5.05, a publicly accessible small-angle scattering analysis software package.[10] The estimated R_g for the h-PEG and d-PEG were obtained using shape-independent poly_guass_coil model which is selected as a simplest model (see section 2.11.1.2). The fitted curves are shown in Fig. 3.1 (C), (D) and (E) on the upper graphs and the lower graphs show the normalized residuals, the number of standard deviations between the measured value and the calculated value, for h/D, d/H and d/DH respectively. Table 3.2 summarizes the parameter values of the fitting results for these samples. The good fit obtained for the h/D with the R_g 26.2 $\pm$ 2.5 Å within Q range 0.007-3 $Å^{-1}$. This value (26 Å for PEG 13000 g/mol) is consistent with the R_g obtained using the same technique (SANS) by Almasy et al[11] (15 Å for PEG with Mw 10000 g/mol), but appears to be lower than the R_g obtained by Rubinson and Kreuger[12] (36 Å for a PVP with Mw 8000 g/mol). For the d-polymer, we tried to estimate the R_g in two different contrast samples d/H and d/DH as it was a tough fit, and the polydispersity illustrates an effect on the fitted model. The sensible fit gained for d-PEG along with R_g 23.3 $\pm$ 2.5 Å amongst Q range 0.013-0.24 $Å^{-1}$ in both samples.

The real space distance distribution function, P(r), was determined for h/D, d/H and d/DH data using the inversion option in SasView, as shown in Fig. 3.1 (B). This distribution can provide information about the maximum particle size, shape and average radius of gyration. The P(r) suggests a Gaussian coil structure for h-PEG and d-PEG in water solutions (Fig. 3.1 (B)), which is consistent with the previous study[13]. The P(r) for h/D has two non-symmetrical peaks and extensible part, with the maximum distance (D_{max}) within the system ~ 82 Å. D_{max} in the P(r) is the point at which P(r) goes to zero and represents the largest distance in the system, which

is the overall diameter of the spherical particles. The derivation of the R_g gives 24.2 Å, in agreement with the R_g determined from the shape-independent poly_guass_coil fitting above (considered $\pm$ 2.5 Å). The P(r) for d/H and d/DH seems identical with the same D_{max} 74.4Å, and the existence of H_2O increases an overall intensity of the P(r). The R_g obtained from the P(r) for d/H and d/DH are 23.3Å and 22.8Å respectively, which are in a good agreement with the fitted model above. Apparently, the dimensions of d-PEG in both samples, d/D and d/HD have a D_{max} of 74.4 Å which is less than the h-PEG dimension, which has a D_{max} 82 Å. This confirms that the dimensions of d-PEG and h-PEG are consistent with the R_g obtained in both methods.

As a particle travels through a fluid, its effective size is described by the hydrodynamic radius (R_h), and it is commonly determined through techniques like dynamic light scattering (DLS). From DLS and static light scattering (SLS), Devanand and Selser[14] reported the measurements of R_h and R_g of polyethylene oxide (PEO). By utilising their established relationship ratio between R_g and R_h, we may derive the R_h from what we know about Rg. Applying that relation for PEG we expect the $R_h \cong 16$ Å, as $R_h/R_g = 0.6$. This result is in agreement with the Linegar[15] finding ($R_h/R_g \cong 0.7$) and with theoretical value $R_h = 0.64\ R_g$[16].

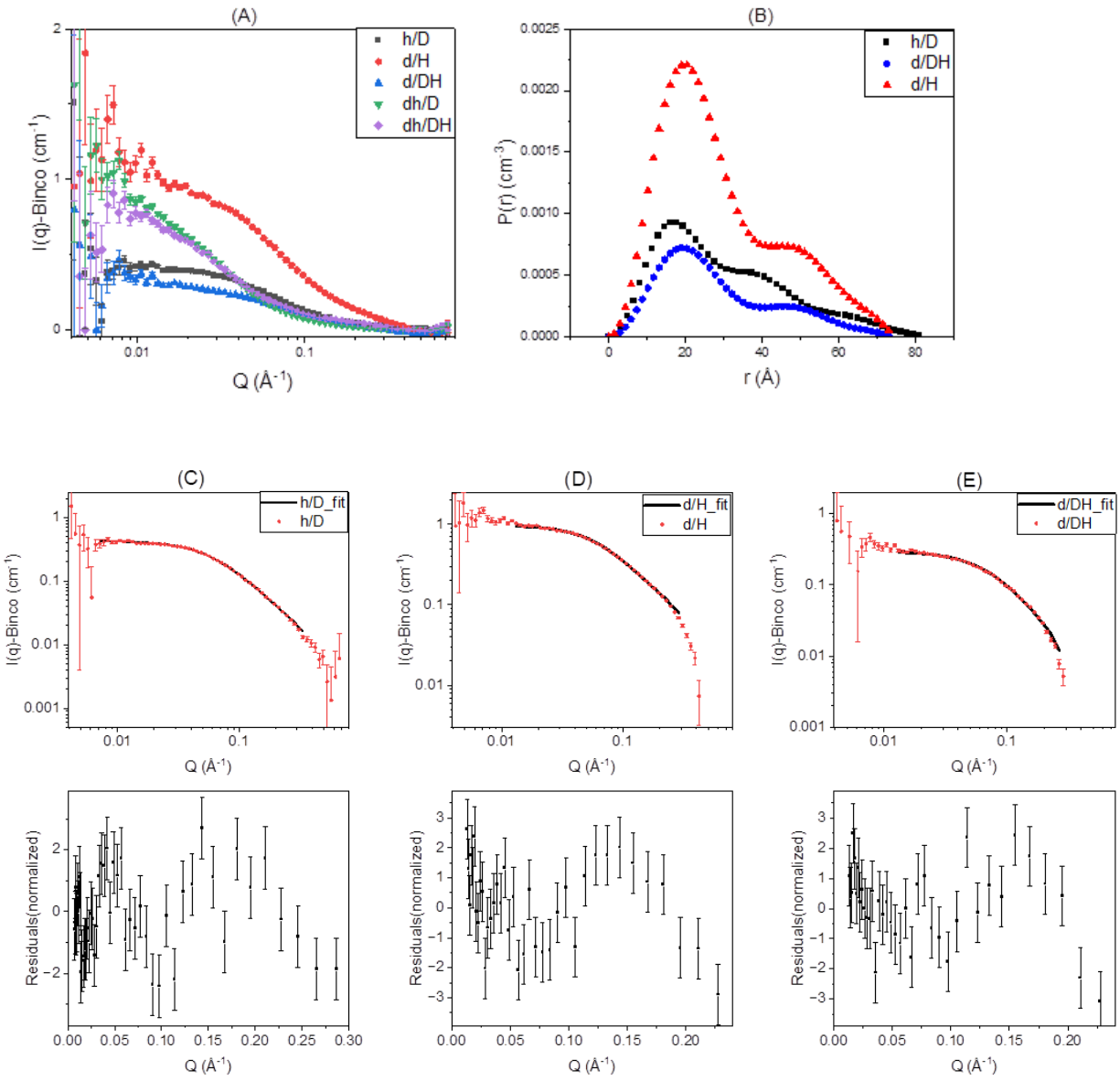


Figure 3. 1: (A) The intensity variation of the background subtracted SANS for 4 wt % h-PEG and d-PEG in varying deuterium content in water h/D, d/H, d/DH, dh/DH and d/D samples, (B) the P(r) function, the real space distance distribution function, for h/D, d/H and d/DH, (C), (D) and (E) are the fitted curves models to the SANS data at the upper graphs (fits show as solid lines) and lower graphs are the normalized residuals current model for h/D, d/H and d/DH respectively. Usually, residuals in SASView are normalized to the uncertainty, which depends on a number of things, such as the quality of the data, i.e., the signal to the noise ratio, the q-range probed, or the number of parameters we have in the model.

Table 3. 2: The summary parameters of the shape-independent poly_gaussian_coil model fit results for h/D, d/H, and d/DH.

sample	h/D	d/H	d/HD
scale	1	1	1
bckgrd	0.0031	0.0398	-0.0055
bckgrd_err	0.0006	0.0014	0.0009
I(0)	0.429	0.942	0.302
I(0)_err	0.0027	0.0034	0.0019
R_g	26.2	23.8	23.3
R_g_err	0.296	0.134	0.237
polydisp	1.46	1.4	1.4
polydisp_err	0.094	n/a	n/a
reduce chi2	1.75	1.99	1.73

Scale: scale factor or volume fraction; bckgrd : source background (cm^{-1}); I(0): intensity at q=0 (cm^{-1}); R_g: radius of gyration (Å); poly_disp: polymer Mw/Mn; reduced Chi2: a statistical evaluation of the degree to which a selected model, given a set of model parameters, fits the data; errors: the parameters' uncertainties.

3.3.2 Total neutron scattering (NIMROD)

The raw NIMROD scattering data were treated for sample holder background and instrument subtractions, and normalised using Gudrun software[17]. The process for sample scattering extraction and correction in Gudrun, can be found in the Gudrun manual[18]. The corrected raw differential cross-section (DCS) data for 4 wt% h-PEG and d-PEG with varying deuterium content in water samples reported in Fig. 3.2 (A) Q 0.1–50 $Å^{-1}$, illustrates that there are three main DCS intensities and features observed, depending mainly on the solvent. In the case of hydrogenated solvent samples, d/H and h/H, the diffraction intensity is large due to the high inelastic scattering contribution. However, the deuterated solvent ones, d/D, h/D, and dh/D, have a lower inelastic contribution, and in between these, the mix DH solvent is presented.

d/D, h/D and dh/D samples have similar diffraction pattern features (Fig. 3.2 (D)). Alike feature peaks for dh/DH and d/DH (Fig. 3.2 (C)) and for d/H and h/H (Fig. 3.2 (B)) are presented. It can be seen from Fig. 3.2 (B) that the intensity pattern for the d/H is higher than the h/H due to the contrast variation that leads the d/H to the high scatter as seen in the SANS section. Similarly, this can be explained by what is observed in Fig. 3.2 (C) and (D).

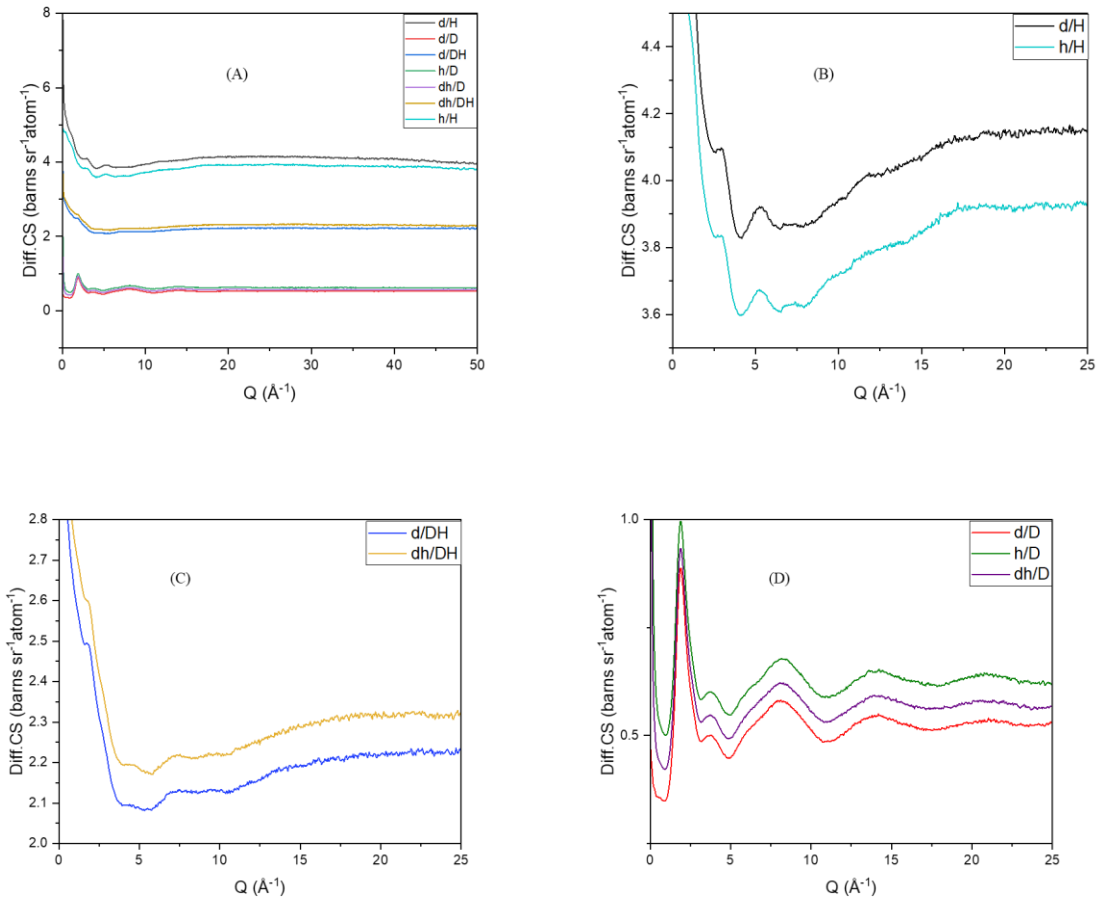


Figure 3. 2: The corrected raw diffraction cross section (DCS) data for (A) 4 wt% h-PEG and d-PEG in varying deuterium content in water (h/D, d/D, d/H, h/H, d/DH, dh/DH, dh/D). In (B), (C) and (D) expansion of relevant region for h/H, d/H, d/DH, dh/DH, d/D, h/D, and dh/D respectively.

As we are dealing with the dilute solutions, most diffraction come from the solvent which presents a challenge for probing structural dynamics of the polymer. Fig. 3.3 (A) shows the total structure factor F(q) for the h/D system from NIMROD and for pure D_2O from SANDALS, which is a comparable instrument used to investigate the structure of liquids and amorphous materials (data provided by Dr Tristan Youngs, STFC). The two patterns, h/D and D_2O, seem very similar. However, the main distinction is the rise at the lowest Q values for

h/D, which results from a contrast between the solvent and the polymer, as shown clearly in Fig. 3.3 (B). Moreover, looking at the first peak carefully, which is related to water-water molecular correlations, the peak shifts down to low Q in the presence of polymer. This indicates the presence of the PEG, and water molecules that are no longer "bulk" water but rather are forming the first solvation shell (and possibly beyond) around the PEG. This can confirm the overlap correlations for inter-atomic (between water and the polymer) or even the backbone correlations, which could happen within the same presented solvent peaks.

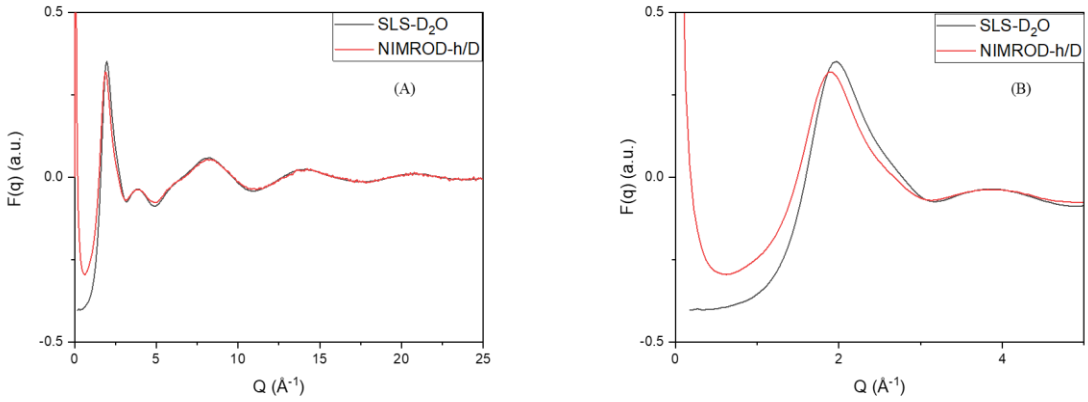


Figure 3. 3: The total weighted-neutron structural factor F(q) for NIMROD-h/D and SANDALS-D_2O in the Q range (A) 0-25 $Å^{-1}$ (B) 0-5 $Å^{-1}$.

It might be challenging to get neutron scattering experimental data, but it can be even more difficult to understand the data, mainly because the experimental signals have little information content for the target structure especially in dilute solutions. The total scattering data from neutrons can provide information on every pair of atoms in a sample, however, getting to that level of details needs molecular modelling. Here, the implementation of the Dissolve software that apply structural refinement for neutron scattering data, has been employed (sections 3.4.3 and 3.5.3).

After knowing the R_g value from experimental SANS, now we need to go to the modelling side to physically visualise the chain atoms and develop the analysis. First, we validated the R_g of the PEG in different length chains up to the near-experimental SANS data using the available force fields and implicit solvent models first in MOE and then, go further to the explicit solvent in GROMACS.

3.4 Computer Modelling Method

3.4.1 Constructing PEG Structure in MOE

The structures of PEG were constructed with different repeat units starting with 4 up to 300 monomers, and the conformations search of PEG were preformed using Stochastic method with RMS gradient ranges between 0.005-0.010 Å, RMSD limit 0.25-0.5 Å and energy window 7-1000 kcal/mol in MOE 2014.0901.[19] The 6-mer PEG was examined using different force fields, MMFF94s, MMFF94x, MMFF94,[20] AMBER12, AMBER10, AMBER94, AMBER99,[21] OPLS-AA,[22] CHARMM27,[23] and implicit solvation models, which include the distance dependent dielectric,[24] Reaction field (R-field)[25] and Generalized Born models[26].

3.4.2 GROMACS Simulation

The all atom explicit simulations were carried out using GROMACS 2018.229 software package[27,28]. The OPLS-AA force field was used in two water models, TIP3P[29] and SPC,[30] and the topology parameters can be acquired by CGM group service[31] up to 1787 atoms. The selection of the OPLS-AA FF was made after validating the chain behaviour, R_g, in multiple FFs with several implicit solvents in MOE (section 3.5.1). Three independent runs, each with the same starting coordinates but with different initial velocities, taken randomly from Maxwell-Boltzmann distributions, were used for the MD simulations – the simulation protocol is described below.

The starting structure of PEGs were obtained from the ensemble of conformations in the OPLS-AA/R-Field model from MOE as described above. A single PEG chain was first solvated in a cubic box that extended 1.0 nm beyond any PEG atom in the small and medium PEG chains (from 4 to 250 monomers) and 2.0 nm in the PEG292 chain. Two box sizes were examined for PEG292. First, the default box size in GROMACS $(123\ Å)^3$ fills with 60734 water molecules, which gives the concentration of 1 wt% PEG in water. A second box was attempted to mimic the experimental concentration used in neutron scattering, 4 wt%, with size of $(82\ Å)^3$ and fill with 17518 water molecules. Next, system energy minimization was done in 1241 steps for the long chain to remove bad contacts and obtain the converged system potential energy using a steepest descent algorithm. After minimization, two equilibration phases were applied. First the system was subjected to MD simulation utilizing the canonical ensemble NVT (0.1-10 ns)

at a cutoff of 10 Å and 2 fs timestep, and the system was heated to reach room temperature using modified Berendsen thermostat[32]. After reaching the desired room temperature, the second equilibration phase, the NPT ensemble, was conducted additionally (0.1-10 ns) to balance the 1 atm pressure using Berendsen barostat[33] at the same cutoff and timestep as NVT. In both phases, the restrictions on the conformational mobility of the backbone atoms were imposed and holonomic constraints were imposed with LINCS to constrain h-bonds. The leapfrog approach[34] was used to integrate the equations of motion. Temperature and pressure were coupled individually to two groups in both phases with the same values. The purpose of the equilibration phases is to allow the water molecules to equilibrate around the polymer without any structural backbone changes, and to reach the proper temperature and pressure before the final production. Finally, the restrictions of the backbone were removed, and the MD production runs were performed for a total of 1,500 ns, for PEG292, with Parrinello-Rahman barostat[35], omitting the first each 50 ns for further system equilibration.

3.4.3 The implementation of PEG solutions in Dissolve Software

Dissolve is a program for classical simulation using Monte Carlo methods, which helps analysing experimental scattering measurements through the generation of representative atomic configurations, consistent with experimental reference data. The atomic structure can be refined via the Empirical Potential Structure Refinement (EPSR) methodology,[36] which has been further developed by Dr. Tristan Youngs and his group at the ISIS Facility. More details about Dissolve, building and parameterising systems, as well as examples on implementation, can be found on the home web page[4,37]. The initial coordinates for simulations of PEG were obtained from MOE, as described in the MOE section above, by choosing the conformer that exhibited an R_g value close to the ones determined by analysis of the Small Angle Neutron Scattering (SANS) data. The model was uploaded in an xyz format, into Dissolve, and the classical OPLS-AA force field was implemented for the fragment of PEG using the LigParGen server[38] and copied the parameters into the long chain after adding extra definition for the terminal atoms. The water model was constructed in Dissolve and minimized using the simple point charge force field, SPC/Fw[39].

It is believed that Dissolve is capable of tackling large system sizes.[4] Thus, refinement of structure of PEG in solution has been attempted, to perform the prediction of empirical potential structure refinement using Dissolve, making this the first use of the software for analysis of dilute polymer solution scattering. Here, seven samples of different isotopic

components have been investigated simultaneously. The main workflows of applying PEG in Dissolve are illustrated below. Fig. 3.4 shows the starting atomic conformation applied in this study.

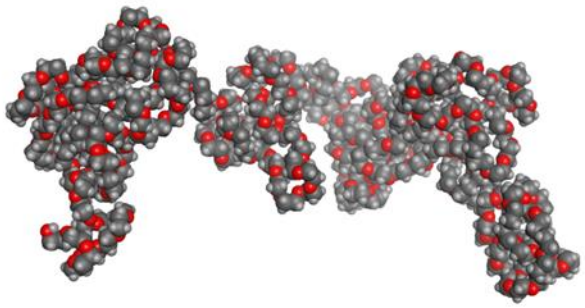


Figure 3. 4: Representation of initial atomic chain conformation of PEG structure obtained from MOE with $R_g \approx 25$ Å (292 monomer units).

3.4.3.1 Setting up a simulation box

A simple random mix configuration of PEG and water was created in a cubic box with lengths of 104 Å in each of the x, y, and z directions. The box contained of 2-polymer chains of PEG (4,080 atoms Mw ≈ 26, 000 g/mol) and 36,000 water molecules (108000 atoms) that meets the experimental NIMROD concentration, 26,000/(36,000 × 18) = 4 wt%, and the density (ρ) of 0.1 atom/Å^3 at 300 K. The density here was calculated by dividing the number of atoms in the system (4,080 + 108,000) by the volume of the box (104 Å)^3.

In the scatting data, the total structural factor represents the data across all polymer chains in the solution. Comparing the modelling system with the real data needs a representation of many chains in the water that meets the experimental concentration which needs a huge data memory and very long time to finish and is therefore unpractical. However, it is beneficial to allow for some different conformations to be represented. It has been proposed to utilise 2-polymer chains instead of a single-polymer chain in the simulation box of Dissolve program. This could represent the calculated backbone correlations included in the total structural factor better than including just calculated correlations for one polymer chain. This also allows for the possibility of chain-chain interaction (even though none is expected at the selected experimental concentration). A snapshot of the aforementioned system is shown in Fig. 3.5.

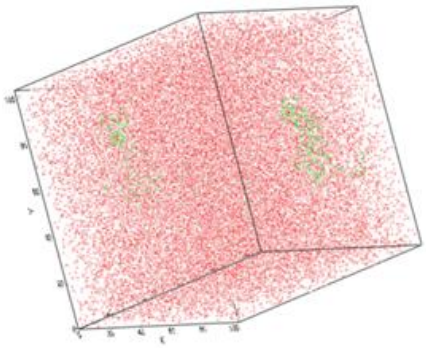


Figure 3. 5: A snapshot of randomized 2-PEG chains and 36,000 water molecules configurations box (4 wt%).

3.4.3.2 Evolution of the configurations

The evolution of the system in Dissolve has two modules, Monte Carlo (MC) and Molecular Dynamics (MD). The Standard MC simulation via the Molshake module, is the principal way to perform and evolve the moves of individual molecules in each iteration. For every 5th iteration run, MD were performed, to enable the relaxation and evolution of the intramolecular degrees of freedom and ensure a sensible sampling model. The MD was set to only run if the configuration's total energy is stable. The default number of MD steps to perform is 50, but it was increased to 300, as an attempt to improve the movement of PEG. All simulations were run on ISIS Data Analysis as a Service (IDAaaS).

Before doing any comparative calculations with the scattering data, the system needed to be stabilized first, in order to get a realistic energy model. The total energy was calculated every 10 steps and monitored during the simulation, until almost a steady value was obtained. As the simulation started running, the energy dropped (Fig. 3.6 (A)) from the first iteration (5×10^{18} kJ/mol) and fluctuated until it eventually reached a stable value at about -1.38×10^{6} kJ/mol for the total energy, -1.65×10^{6} kJ/mol intramolecular and 2.72×10^{5} kJ/mol intramolecular, after about 6,370 iterations (Fig. 3.6 (B)).

The abrupt decrease in energy in Dissolve occurs because the molecular dynamics (MD) is not activated during the early equilibration stages, and instead, only Monte Carlo motions are employed. The molecular dynamics (MD) simulation is initiated only after the energy has achieved a specific level of stability. During the initial MD run, significant effort is dedicated to adjusting the intramolecular geometry, which remains unchanged during the Monte Carlo simulation. As a result, the energy experiences a dramatic fall followed by a subsequent levelling off.

The system temperature is controlled using a simple velocity scaling algorithm. The algorithm adjusts the timestep based on the method developed by Marks and Robinson[40]. The magnitude of the timestep is determined by the maximal intramolecular force and is limited to a maximum value of 5×10^{-4} ps.

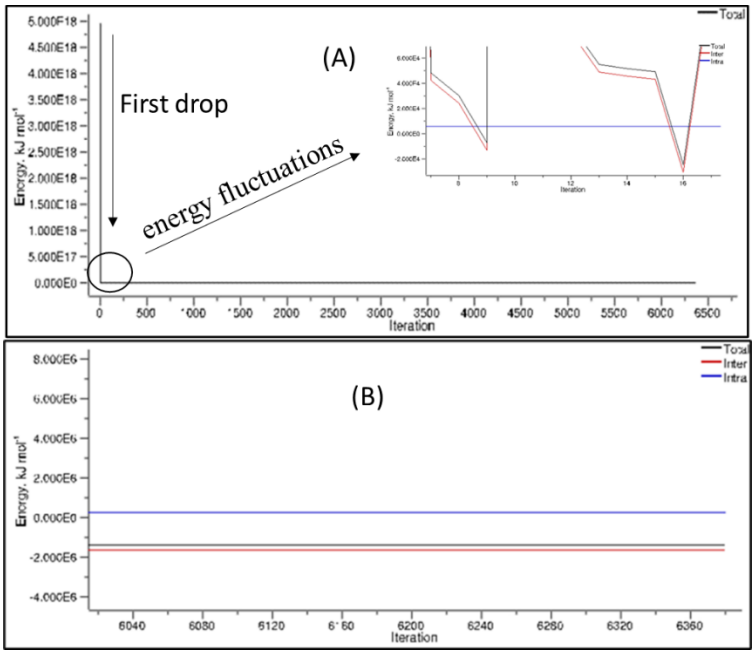


Figure 3. 6: Snapshot of (A) the dropping and fluctuating energy in the early stage of the simulation (B) the total (black line), interatomic (red line) and intramolecular (blue line) energy at the end of the minimization stage.

3.5 Results and Discussion

3.5.1 MOE

3.5.1.1 R_g Validation in different Force Fields and solvents

Nine force-fields were tested using MOE to evaluate the behaviour of 6-mers of PEG in a distance-dependent dielectric implicit solvent, by running a conformational search using the Stochastic method; these models were evaluated based on the radius of gyration (R_g). The data obtained (Table. 3.3), shows the conformations ranging from high to low R_g, based on the force field used: CHARMM-27>MMFF>OPLS-AA>AMBER. The most compact R_g range is

present in the AMBER-simulated conformations, as expressed by the standard deviation (STDEV) values, going as low as 0.1 Å. This family of forcefields were therefore not considered further, as they fail to represent the known flexibility of this system by generating too compact conformations. The CHARMM27 force field appears to have the higher deviation between max and min R_g values (2.9 Å) also expressed by the highest STDEV (0.8 Å), compared to the rest. The OPLS-AA displays the conformations with a moderate deviation (1.4 Å), with a STDEV of 0.2 Å. Thus, CHARMM and OPLS-AA were selected, seeing as how they are compatible with the recent computational studies for PEG[41–46]. This led to the comparison of these fields using the R-field and Born solvation models.

Table 3. 3: The radius of gyration (R_g) for 6-mers in different force fields/distance solvation/Stochastic method.

Force-field	Mean (Å)	Max (Å)	Min (Å)	STDEV (Å)	Max-Min (Å)
MMFF94x	3.6	4.9	3.0	0.4	2.0
MMFF94s	3.6	5.9	3.1	0.5	2.9
MMFF94	3.6	5.9	3.1	0.7	2.8
AMBER12	3.3	3.9	2.9	0.2	1.0
AMBER10	3.1	3.2	3.1	0.1	0.1
AMBER94	3.4	4.2	3.0	0.3	1.3
AMBER99	3.2	4.0	2.9	0.2	1.1
OPLS-AA	3.2	4.2	2.8	0.2	1.4
CHARMM27	3.9	6.0	3.0	0.8	2.9

Gradually increasing the size of PEG permitted the evaluation of these chains using the CHARMM27 and OPLS-AA force fields, and Born and R-Field implicit solvent models. The search ran for each of the chain lengths (from 4 to 300-mer), with the data shown in Tables 3.4 –3.7. The first small molecular weights seem to behave similarly in these models, and start to change when the weight gets larger. The conformation and the size of the polymer depend mainly on the solvent; with a well-solvated environment permitting the expansion of the polymer chain, and a bad description of this restricting the solvent-accessible-surface-area.[43,47] To examine the Born and R-Field solvents, the Flory approximation approach was used, whereby $R_g = R_{g0} N^{v}$, with the R_{g0} depicting the radius of gyration of the monomer, N the number of repeat units, and v the solvent exponent. As seen in Fig.3.7, the correlation function of R_g increases as the number of repeat units increases, under both of these solvents. Adding

the power trendline to get the same equation as Flory, the solvent value in CHARMM/R-Field model appears to have slightly higher value (0.44), while the CHARMM/Born has a lower value (0.40) and OPLS-AA/Born and OPLS-AA/R-Field have a similar value (0.43). The R_g obtained from this study is smaller compared to some previous studies[15,41,48]. To select an appropriate model, more accurate tests need to be performed, thus, molecular dynamics simulations in explicit water solvent were examined.

The MD simulations in explicit TIP3P water solvent were performed for different length chains using the OPLS-AA FF. The exponent solvent value that is related to the R_g, was achieved from the explicit water (0.5) which is closer enough to the ideal chain, compared to the implicit solvents, see Fig. 3.7. Subsequently, the choice of the OPLS-AA force field (FF) and explicit model was made, to run the MD simulations and predict the SANS curve for PEG292.

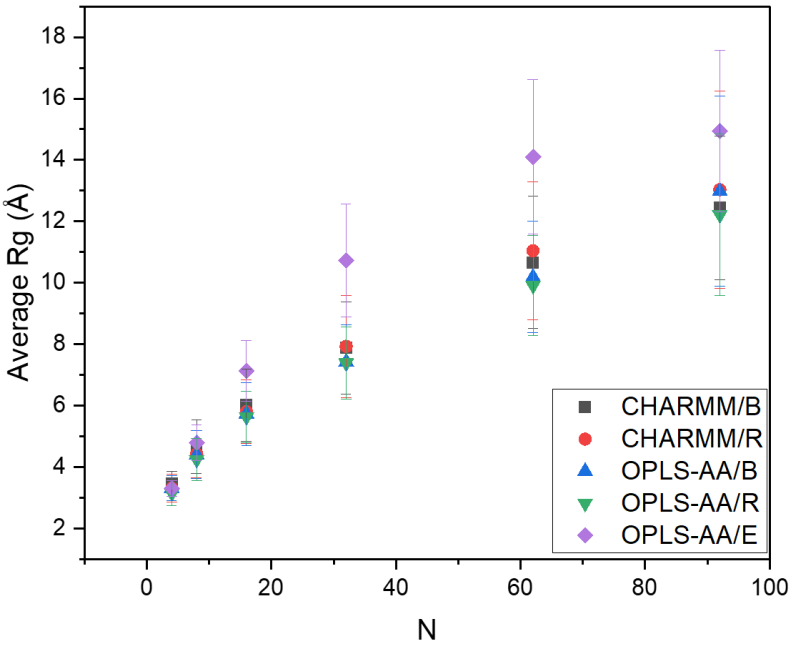


Figure 3. 7: R_g Comparison between the model of OPLS-AA in explicit TIP3P(OPLS-AA/E), OPLS-AA in R-field (OPLS-AA/R) and in Born (OPLS-AA/B), CHARMM FF in Born (CHARMM/B) and in R-Field (CHARMM/R). The error bars represent the standard deviations of the calculated R_g.

Table 3. 4: The R_g in OPLS-AA/Born/Stochastic

Number of monomer	Mean (Å)	Max (Å)	Min (Å)	STDEV.P (Å)	Max-Min (Å)
4	3.3	4.6	2.4	0.4	2.2
6	3.9	6.1	2.9	0.6	3.3
8	4.4	7.6	3.1	0.8	4.5
12	5.1	10.7	4.0	0.9	6.7
16	5.7	11.4	4.4	1.0	7.0
24	6.6	12.3	5.2	1.0	7.1
32	7.4	12.3	6.1	1.2	6.2
42	8.4	13.9	6.6	1.5	7.3
52	9.2	15.8	7.3	1.6	8.7
62	10.2	15.6	7.9	1.8	7.7
72	10.8	16.5	8.4	2.0	8.2
82	11.7	17.5	8.9	2.4	8.6
92	13.0	25.6	8.9	3.1	16.7
100	13.0	19.0	9.7	2.3	9.3
150	16.0	27.1	11.4	3.7	15.7
200	17.0	23.6	12.0	3.2	11.6
250	20.3	30.2	14.5	4.5	15.7
300	23.6	46.3	13.8	6.2	32.5

Table 3. 5: The R_g in CHARMM27/Born/Stochastic

Number of monomer	Mean (Å)	Max (Å)	Min (Å)	STDEV.P (Å)	Max-Min (Å)
4	3.5	4.5	2.5	0.4	2.0
6	4.2	6.1	3.0	0.7	3.1
8	4.7	7.8	3.3	0.9	4.5
12	5.5	10.0	4.1	1.1	6.0
16	6.0	11.2	4.6	1.2	6.5
24	6.9	13.8	5.4	1.3	8.4
32	7.9	15.1	6.1	1.5	9.0
42	8.8	16.6	7.0	1.7	9.6
52	9.3	16.3	7.4	1.9	8.8
62	10.7	18.0	7.8	2.2	10.3
72	11.4	17.9	8.5	2.3	9.4
82	12.0	18.7	9.0	2.2	9.7
92	12.4	17.8	9.2	2.4	8.6
100	13.0	19.6	9.7	2.6	9.9
150	17.9	33.6	11.6	5.3	22.0
200	18.9	26.5	14.9	3.2	11.7
250	23.0	36.9	14.4	6.1	22.4
300	26.7	42.3	16.0	7.3	26.3

Table 3. 6: The R_g in CHARMM/R-Field /Stochastic

Number of monomer	Mean (Å)	Max (Å)	Min (Å)	STDEV.P (Å)	Max-Min (Å)
4	3.3	4.5	2.4	0.5	2.0
6	3.9	6.3	3.0	0.7	3.3
8	4.4	7.5	3.4	0.8	4.1
12	5.3	9.9	4.1	1.0	5.9
16	5.8	10.6	4.6	1.0	5.9
24	7.0	12.9	5.3	1.2	7.7
32	7.9	15.3	6.1	1.7	9.3
42	8.6	15.7	6.7	1.5	9.1
52	9.9	15.1	7.5	1.9	7.6
62	11.0	20.2	7.9	2.3	12.3
72	11.1	18.2	8.5	2.2	9.7
82	12.3	23.3	8.7	3.2	14.6
92	13.0	20.9	9.4	3.2	11.5
100	13.6	26.7	9.6	3.2	17.1
150	16.5	24.1	11.5	3.2	12.5
200	18.6	27.5	13.6	3.4	13.8
250	23.4	31.0	14.5	4.7	16.6
300	21.7	33.1	14.0	5.2	19.1

Table 3. 7: The R_g in OPLS-AA/R-Field /Stochastic

Number of monomer	Mean (Å)	Max (Å)	Min (Å)	STDEV.P (Å)	Max-Min (Å)
4	3.2	4.4	2.4	0.4	2.0
6	3.8	5.7	2.9	0.6	2.8
8	4.2	7.0	3.3	0.7	3.7
12	5.0	8.4	4.0	0.8	4.4
16	5.6	9.4	4.5	0.8	4.8
24	6.6	10.0	5.2	1.1	4.8
32	7.4	12.2	5.9	1.2	6.3
42	8.2	14.1	6.6	1.3	7.6
52	8.9	12.9	7.2	1.3	5.7
62	9.9	13.5	7.9	1.6	5.7
72	10.9	17.0	8.2	2.2	8.9
82	11.4	15.9	8.9	1.8	7.0
92	12.2	19.3	9.1	2.6	10.2
100	12.1	21.5	9.1	2.5	12.4
150	15.4	23.1	11.0	3.3	12.1
200	17.5	28.8	13.0	4.4	15.8
250	22.4	31.9	13.0	6.2	18.9
300	25.1	46.5	15.5	7.0	31.0

3.5.2 GROMACS

3.5.2.1 Validating the R_g for PEG292 Explicitly in Two Water Models

The MD simulation in explicit water was successfully performed for PEG292, with the OPLS-AA FF, using GROMACS. First, the two water models, TIP3P and SPC, were tested in the default box size $(123\ Å)^3$, which is 1% PEG in water, in order to examine the performance of PEG chains in these conditions. Second, the MD of 4 wt% of PEG $(82.7\ Å)^3$ in TIP3P water was generated, which is the percentage used in the neutron scattering experiment for further analysis. All MD calculations were run for a total of 1,500 ns in each condition; each 500 ns run has an independent starting velocity with the same starting coordinates, and the initial 50 ns were removed from each run to account for additional equilibration in the analysis part.

The first important criterion in our purpose is to look at the R_g of the polymer. Fig. 3.8, illustrates the running average of the R_g versus time (every 25 ns to simplify the picture) for PEG292 in (A) the SPC default box, (B) the TIP3P default box and (C) the TIP3P with 4%, and the calculations of the average of R_g, max, min and SD were obtained in all models and shown in the Table. 3.8. The PEG chains display a greater expansion in the SPC than in the TIP3P when looking at the average of R_g values, which are ≈ 42 and ≈ 22 Å with SD 13 and 3 Å, respectively; the higher SD indicates that there is an increased range of conformations sampled in the SPC than in the TIP3P model. However, the TIP3P values display a more sensible R_g average value (≈ 22 Å) compared to the experimental SANS value (≈ 26 Å) than SPC (42 Å). TIP3P is generally considered to be more rigid than the SPC model, derived from TIP3P's ability to control chain growth to a suitable range and enhance PEG sampling. Therefore, for this purpose, the TIP3P model was chosen in this study, which has been used for PEG in previous MD studies[49–51].

The average R_g value (≈ 25 Å) obtained from the 4% PEG- MD was even closer to the SANS data with max ≈ 55, min ≈ 16, and SD ≈ 7 Å than in the default box. Consequently, decreasing the box size by 40 Å in each direction further improved the range of conformations sampled, which means that the ensembles of the conformations found in that MD trajectory are consistent with the SANS data. Increasing the box size means increasing the number of water molecules, and this can significantly affect the mobility and thus the properties of chain atoms. It can be noted how changing the concentration and the box size can play a role in the mobility of polymer atoms even with the same water model. Reducing the box size, which results in reducing the water particles in the box, allows the chain more freedom to expand regarding the PEG-concentration, which changed from 1% to 4%.

However, the average R_g value between the 4% MD simulation (≈ 25 Å) and the SANS data (≈ 26 Å) is not quite matched yet, as the calculated value is slightly lower than the experimental, but it is still considered to be within the standard deviation SANS fitted model (± 2.5 Å). This difference in the R_g is compatible with some protein studies[52,53] that found that common water models miscalculate the dispersion interactions with the protein and result in a slightly thinner hydration layer, and hence a slightly smaller R_g than anticipated from the experiment[54]. The calculation of the number of water molecules in the hydration layer found within 4 Å distance from the backbone can be found in the RDF section below.

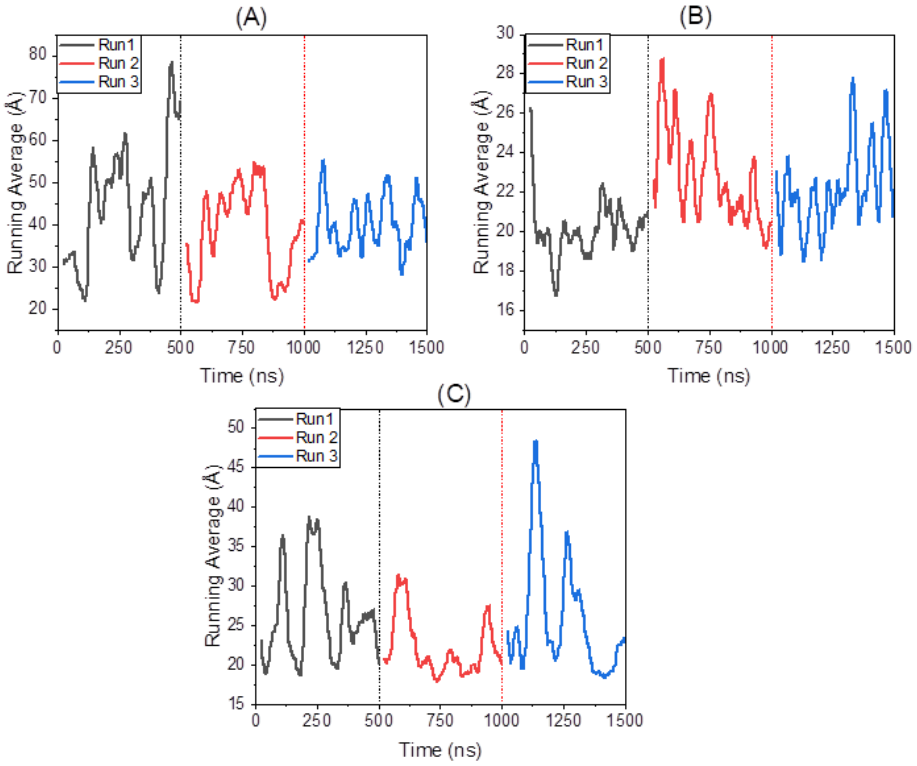


Figure 3. 8: The averaging R_g for PEG292 versus time in (A) the default box SPC (B) the default box TIP3P (C) 4% PEG in TIP3P.

Table 3. 8: The calculations of the average R_g, max, min and SD for PEG292 in SPC, TIP3P in the default box and in the 4% TIP3P.

Models	Mean R_g (Å)	Max (Å)	Min (Å)	SD (Å)
SPC	41.8	89.3	17.9	13.4
TIP3P-defult box 1 %	21.5	37.8	15.1	3.2
TIP3P-4 %	24.9	55.2	15.9	6.9

3.5.2.2 Computed SANS curves and Fitting with the experimental SANS data

Another reason for implementing explicit solvent when looking at the PEG system was to allow for calculation of the SANS curve from a given MD trajectory that has more similar R_g to SANS than the trajectory obtained from the implicit model, and hence to obtain more information about the distribution of polymer structural sizes in explicit solvent that might be

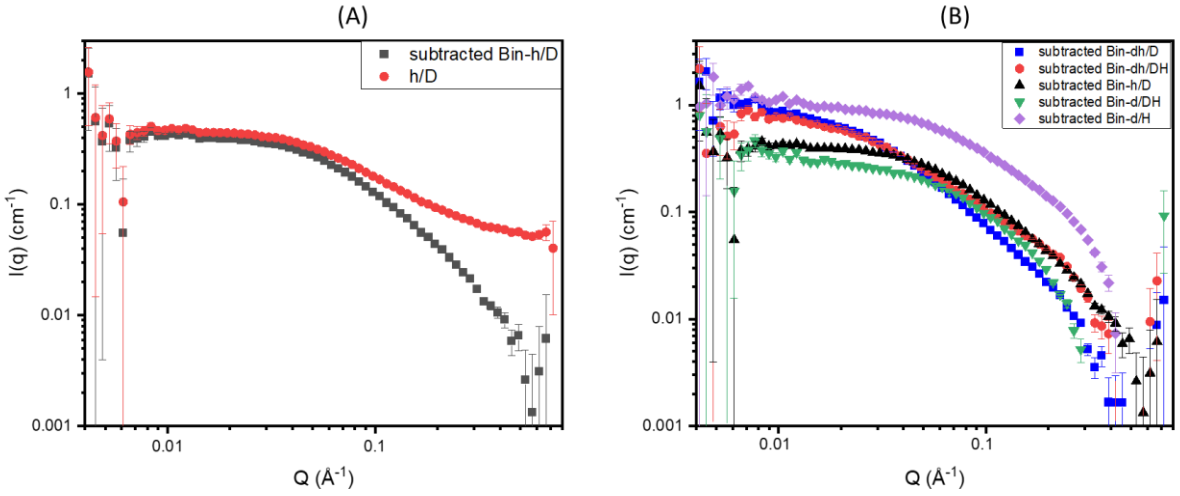
included in the scattering data. It is important to calculate the SANS curve from MD trajectories that contain several ensembles of conformations in order to interpret the SANS data. Numerous quick and well-liked software programmes, such as Cryson,[55] Sassie,[56,57] Pepsi-SANS,[58] and SASIM[59], can predict SANS spectra corresponding to given models. However, the spectra produced in this manner do not account for the internal conformational dynamics of macromolecules as a whole and generally raise the possibility of limiting the amount of structural data that can be taken from the data. Moreover, it is known that the hydration layer density might affect the R_g, as found in many protein studies,[60,61] where the R_g value obtained from solution was larger than the one obtained from the coordinates alone. Thus, the MD simulations can be utilised in order to account for the contribution of dynamic processes occurring in the structures of macromolecules in a solution, hoping to provide a more accurate prediction of SANS scattering curves.

It is important to keep in mind that the intensity scaling of the SANS curve in GROMACS simulations is an arbitrary unit, and the SANS spectrum is affected by all size macromolecule conformations in the MD trajectory, which may depend on a number of variables, including the simulation box's size, polymer shape, and the force field parameters selected. The calculated small-angle neutron scattering intensity curves for PEG292 from the sum of all frames in 1 and 4% MD output trajectories independently (135,000 frames each) and the maximum, minimum and average R_g conformations from both MDs were obtained and fitted over the Q range 0–0.25 Å using the sans tool within GROMACS, as shown in Fig. 3.9 (E), (C), and (D), respectively. It should be said that the box size was increased from 82 Å into 123 Å, which is the default size, in the 4% MD frames to avoid the corruptions in the extended frames during the calculations. In order to make a fair comparison with the experimental curve, it may be necessary to perform additional scaling steps to match the experiment. The experimental SANS signal consists of a contribution of both coherent and incoherent scattering, which generates a number of practical problems that can complicate contrast matching experiments. The predicted SANS curve contains only coherent information. Thus, the incoherent background (0.05) was subtracted from the experimental I(q), resulting in a decrease in the scattering intensity pattern as shown in Fig. 3.9 (A). To fit the experimental h/D pattern into the predicted curve, multiply the experimental I(q) by a factor of 2.44 to achieve the better-matched intensity.

One can note that the predicted SANS pattern is sensitive to the overall structural factors, which include the information of the shape and size of the backbone. We fitted the subtracted

experimental h/D data into the theoretical patterns. It can be recognised from looking into the extended structures: maximum R_g 38 Å in 1% MD and 55 Å in 4% MD, are shifted to the low Q, while the compacted structures, minimum R_g 15 Å in 1% MD and 16 Å in 4% MD, are shifted to the higher Q, especially in the 0.01-0.1 $Å^{-1}$ range, which are expected results. Surprisingly, the average R_g conformation in 1% MD, 22 Å, seems to match the experimental pattern better than the average R_g obtained from 4% MD, R_g 25 Å. Furthermore, the curve obtained across 4% MD and 1% MD appears similarly in good agreement with the experiment data shifting pattern a bit to the smaller size in the 1% MD and slightly to the bigger size in the 4% MD. This observation can explain the highly extended conformations found in 4% MD than in 1% MD. This is because all size conformations in the MD trajectory have an impact on the SANS spectrum, as stated above; when looking at the maximum R_g found in the 4% MD, it is about 55 Å, while 38 Å in the 1% MD, subsequently shifting the 4% MD curve into lower Q. The trend pattern for h/D seems to mediate between 1 and 4% MD, with a slight tendency more towards the 1% MD curve than 4% MD.

The discrepancy between the computed and experimental patterns across 1 and 4% MD, is presented after 0.1 $Å^{-1}$, and this could be for many reasons. One is that the contrast variations between the backbone and the solvent that play a significant role in the experimental production pattern. Difference in SANS patterns can be observed for the same backbone with different isotopic solvents; e.g., in Fig. 3.9 (B), the SANS patterns in the dh/D and dh/DH datasets start with a similar style at low Q and then differ after 0.1 $Å^{-1}$, which is the Q region that is inconsistent between the experimental and theoretical curves. The contrast variation term is not included in the computed SANS calculation used by this method.



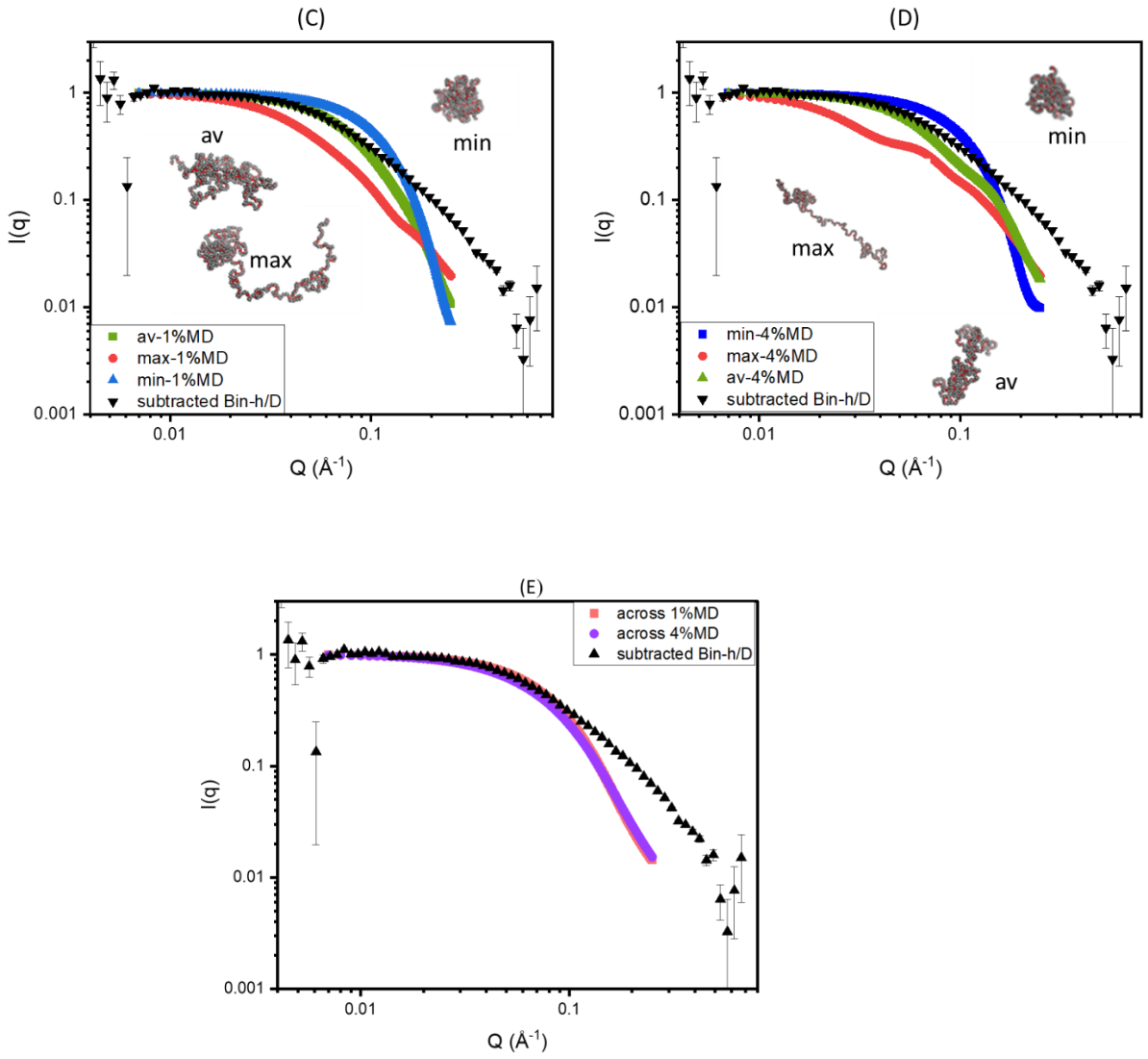


Figure 3. 9: (A) The experimental h/D SANS data before (red) and after (back) subtracted incoherent background (Q 0.007-0.7 Å^{-1}), (B) the subtracted incoherent background experimental dh/D (blue), dh/DH (red), d/H (purple), d/DH (green) and h/D (black), (C) the subtracted incoherent background experimental fitted h/D data (black) against maximum (red), minimum (blue) and average (green) R_g conformations obtained from 1%MD (D) the subtracted incoherent background experimental fitted h/D data (black) against maximum (red), minimum (blue) and average (green) R_g conformations obtained from 4%MD, (E) the subtracted incoherent background experimental fitted h/D data (black) against computed across 1 (pink) & 4 (purple) %MD curves. The experimental data have been scaled to match the intensity of the simulated data at the lowest Q data point

The real space distribution function P(r) is an important tool for examining and comprehending how physical systems' objects or particles are arranged in space. Fig. 3.10 (A) and (B) displays the predicted P(r) for PEG292 initial, maximum, minimum, average R_g conformations and across the analysed frames in 1 and 4% MDs, which were obtained from GROMACS, along

with intensity I(q), fitted into the experimental h/D obtained from inversion analysis in SasView. The calculated P(r) was scaled by 0.002 multiplication to match the experimental intensity. This allowed us to investigate the backbone arrangement by looking at the Gaussian formations. The Gaussian shapes provide a reasonable indication along with the maximum distance, D_{max}. For instance, the minimum R_g conformations (15 and 16 Å in both MDs) function as a globular shape with D_{max} around 50 Å. This is consistent with the backbone dimensions in those compacted structures, and it is about one third away distance from the experimental D_{max} and does not have a resembling shape to the targeted data. Moreover, the maximum R_g structures (38 and 55 Å) show the high D_{max} endings (145 and 170 Å), which are also far away in dimensional distance from the experimental distributions. The D_{max} for 1% MD across the analysed frames and the average R_g conformer in that MD have very close endings (82 Å) to the experiment while losing the overall shape features, making it globularly looking. Similarly, this is what is seen in the formation of P(r) across the 4% MD, but with a slightly higher ending than seen in the 1% MD. On the other hand, after testing many different individual frames, from the 4% MD, that have similar R_g to the experimental average, the particle arrangement most closely resembling the experimental Gaussian distribution, is for the average R_g structure which is shown in Fig. 3.10 (B).

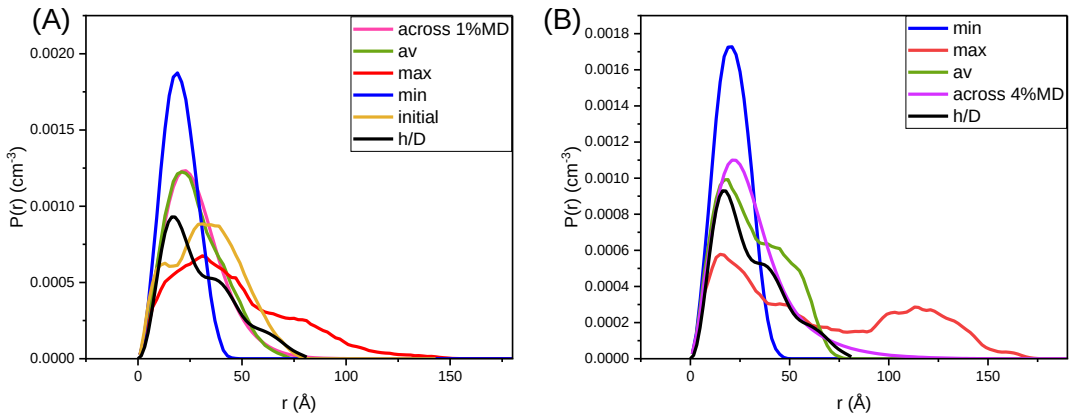


Figure 3. 10: The real space P(r) distribution of the experiment h/D (black line) with the calculated across 1% MD, (A) (pink), and across 4% MD, (B) (purple), with the maximum (red), minimum (blue), and average (green) R_g conformations in each MD and the initial conformation (yellow).

3.5.2.3 Radial Distribution Function

The RDF can provide insights into the structural organisation of molecules in the system. First, the RDF was generated for PEG backbone averaging across the 1% and 4% MD frames (Fig. 3.11 (A)), as well as individual frame sizes such as maximum, minimum, and average R_g structures (Fig. 3.11 (B)). All RDFs were calculated using a bin size of 0.025 nm. Generally, the higher intensity of the g(r) in the 1% MD indicates the more compacted structures found than in the 4% MD, as already seen in the R_g above. From the observation of the shapes of the RDFs obtained from individual R_g sizes of PEG, the two general features are identical regarding the position of the peaks, which are related to the covalent bond regions: the first peak is located at the C-H and the O-H regions, which are the shortest distances found in the PEG structure as calculated by all pairs to all pairs, followed by the second peak, which is located at the C-O and C-C areas. The inconsistency of the peak shapes appears between 2 and 2.5 Å, which is related to correlation between atoms within the backbone of the same monomer, e.g., oxygen to the hydrogen positioned at the adjacent carbon O... HC_1, oxygen to the hydrogen positioned at the adjacent-adjacent carbon O...HC_2, oxygen to the far carbon adjacent to the carbon O...C_2, or between atoms within different adjacent monomers. This is expected as we do have different ranges of folding conformations, as shown apparently in Fig. 3.11 (B). The feature peaks located between 4 – 5 Å are presented similarly in all frames but with slightly different g(r) intensities. This points out that the atoms in the two adjacent monomers are closer to each other in the minimum R_g frame, resulting in an increase in the possibility of detecting the particles. The accumulation of particles depletes after 6 Å, which might describe the random coil orientation of PEG monomers.

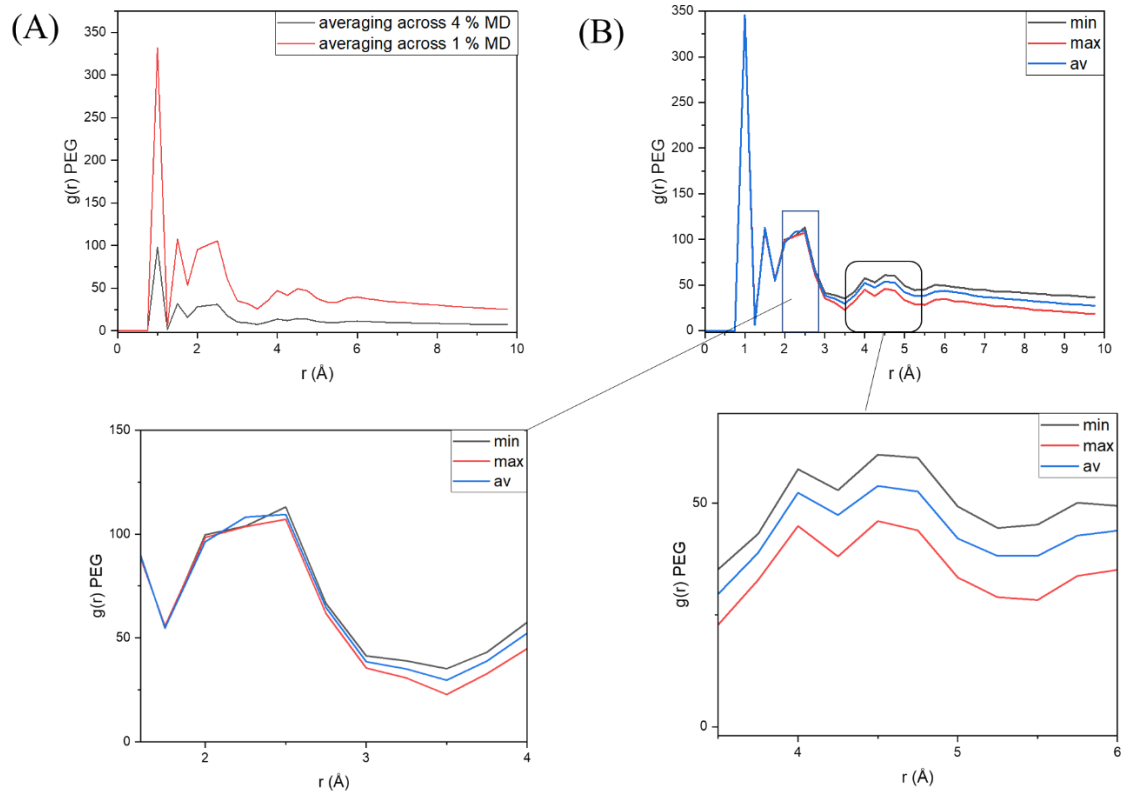


Figure 3. 11: The g (r) for (A) the all atoms-all atoms PEG averaging across 1% (red line) and 4% (black line) MDs and (B) the all atoms-all atoms PEG for the maximum (red line), minimum (black line) and average (blue line) R_g structures.

Additionally, the RDF can reveal the nature of interactions between the two molecules and can show the solvation shell or the distribution of solvent molecules around the reference molecule. Second, the RDF calculation between the PEG molecule set as a reference position to any water atom was made averaging across 1 and 4% MDs as well as the individual maximum, minimum and average R_g structures as shown in Fig. 3.12 (A) and (B) respectively. The detection of the water species starts to appear at early distance around 2 Å, which is less intense peak than seen between 3 – 6 Å before getting to the bulk water. The g (r) seems to be statistically identical in both MDs with the exception of a slight increase in the g(r) for 4% MD due to the broader range of conformations seen in the 4% MD. The area under the curve up to 4 Å can be considered to be the water within hydration layer that is associated directly to the PEG (see section 3.5.3.7). It is known that a strong hydrogen bond can be formed when the distance between H-bond donor and acceptor is in the range 2.2–2.5, 2.5–3.2 as moderate, and 3.2–4 Å as weak. Consequently, we calculated the number of water molecules from any atom of PEG to any OW to avoid any double counting for HW, e.g.H_1W or H_2W, that belong to the same

OW within 4 Å distance from the PEG as a function of time (ns) as shown in the Fig. 3.12 (C). The density and position of the water around the PEG might vary depending on the adapted PEG conformations in the molecular dynamic study. Table. 3.9 displays the mean, maximum, minimum and SD of the number of water molecules found within 4 Å in 1 & 4% MDs. The average number of water is similar across both MDs, about 1200 (about 4 water molecules per monomer unit), the maximum, 1687 and 1639, and the minimum, 924 and 821, in 1 & 4% respectively. This makes more sense that the structures observed in trajectories with a higher number of water molecules (Fig. 3.12 (D) max) appear to be more extended due to the space occupied by water. On the other hand, the more compact structures (Fig. 3.12 (D) min) contain less water. The coiled structures with an average number of water molecules (Fig. 3.12 (D) av) are surrounded by an average number of water molecules. Furthermore, this should be noticed when looking at the g (r) for PEG-water in the three different size structures in Fig 3.12 (B), the possibility of finding the surrounding water species is higher in the max structure follow by the av and then min. The validation of the anticipated quantity of water molecules present in the hydration shell associated with the backbone SANS scattering can be elucidated through the utilisation of the SANS-driven MD method, which takes into account the solvation layer as discussed in Chapter 5.

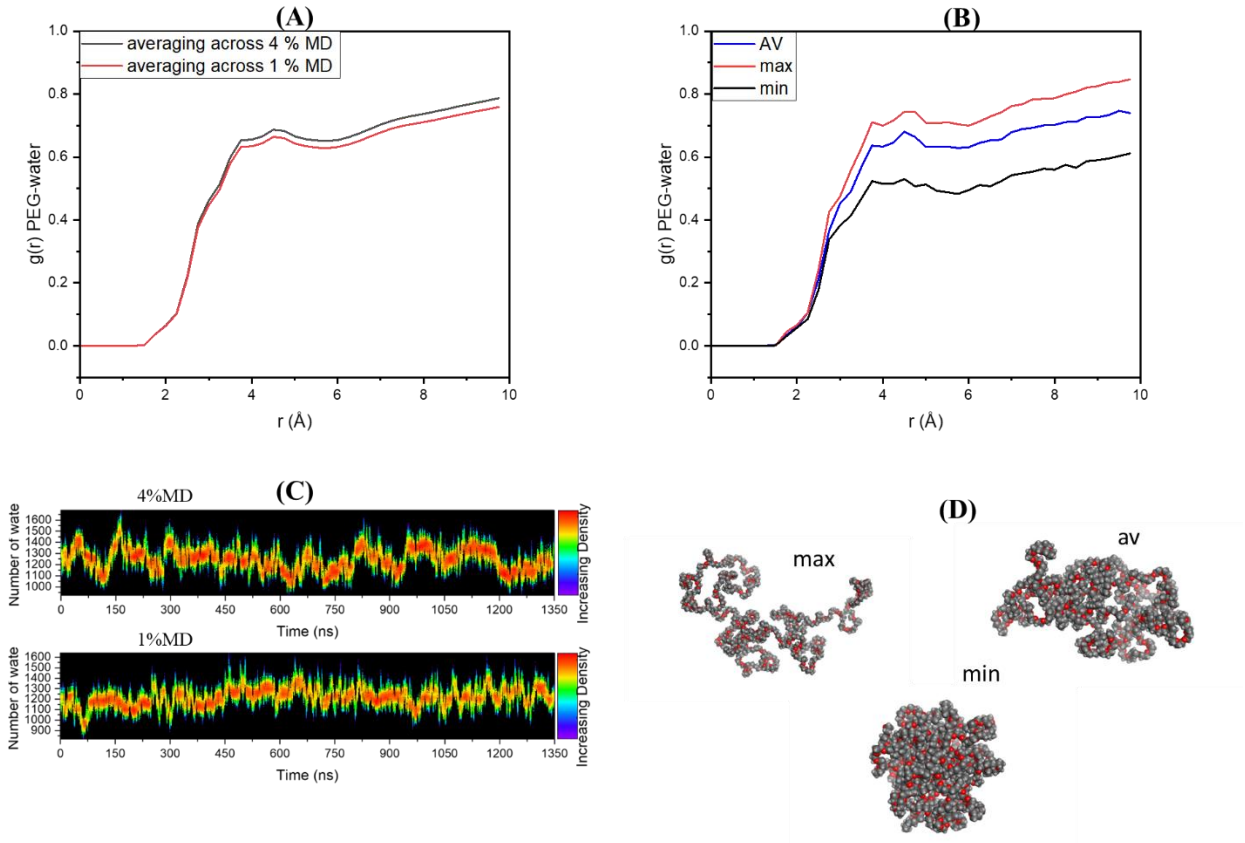


Figure 3. 12: The calculated RDF-PEG-water (A) averaging across the 1 and 4% MDs, (B) for maximum, minimum and average R_g structures; (C) the number of water molecules calculated within 4 Å of PEG as a function of time (ns) for 1 and 4% MDS (D) The examples of snapshots show the structures contain a higher, max, average, av, and lower number of water molecules within 4 Å.

Table 3. 9: The calculation of the mean, maximum, minimum and SD of water molecules within 4 Å from PEG in 1 and 4% MDs

Number of water molecules	Mean	Max	Min	SD
1% MD	122	1639	821	103
4% MD	125	1687	924	110

3.5.3 Dissolve Initial Data Fitting

After the system was minimised and equilibrated, the calculated simulation is now ready to be compared with the experimental data. The "NeutronSQ" modules were generated for all seven samples individually. The additional setting for the H isotope, ^{2}H, e.g., adding the scattering lengths of 6.671 fm for ^{2}H that meets the scattering lengths of -3.7406 fm for H, was made properly to mimic the experimental datasets. In other words, the deuterated samples, h/D, d/H, d/D, d/DH, dh/DH and dh/D, were specified in terms of isotopologues computationally. To begin with, a basic comparison means looking at the total neutron-weighted structure factor F(Q) and total neutron-weighted radial distribution function G(r) for the PEG solution.

3.5.3.1 Total Neutron Weighted Structure Factor F(Q)

In Fig. 3.13, the comparison between experimental data and the fitted calculated total neutron F(Q) from Dissolve of the seven samples, h/H, d/H, h/H, d/D, d/DH, dh/DH,dh/D, is shown. Despite the fact that the Q-range was measured experimentally from 0.01 to 50 $Å^{-1}$, the charts only display data up to 25 $Å^{-1}$, to highlight key aspects. Most of the scattering features in the samples are related to the solvent, as we are dealing with a polymer in dilute solution (the PEG accounts only for 4% of the mass concentration), and obviously, from looking at Fig. 3.13, the three aspects can be classified based on the solvents H_2O, D_2O and H_2O/D_2O. The h/H and d/H have a very similar pattern with two symmetrical peaks, while the d/DH and dh/DH have one distinct peak followed by small hills, and finally the h/D, d/D and dh/D have a sharp peak accompanied by small broad peaks. Overall, a reasonable agreement in the number of peaks exists between experimental data and the simulated model; in most of the samples, however, there are discrepancies at the exact peak positions, especially at low Q regions.

The leading peak in the hydrogen-containing (solvent) samples, d/H and h/H, occurs at position 3 $Å^{-1}$, and the first peak in the deuterated samples, d/D, h/D, dh/D, and mixture samples, dh/DH, d/DH, are close to 2 $Å^{-1}$, with a slight difference in the order of nanoscale structure. At low Q regions, there is an aspect in the H_2O solvent cases, d/H and h/H, that could be related to the processing of the experimental data and might link to the effect of inelastic scattering due to the light H that has vibrational degrees of freedom,[62] and that is less visible in the DH and D cases, or could be related to the backbone correlations or inter-atomic correlations, as discussed in detail in the partial section below. Now, we need to see the correlations of these solutions in the real-space distance.

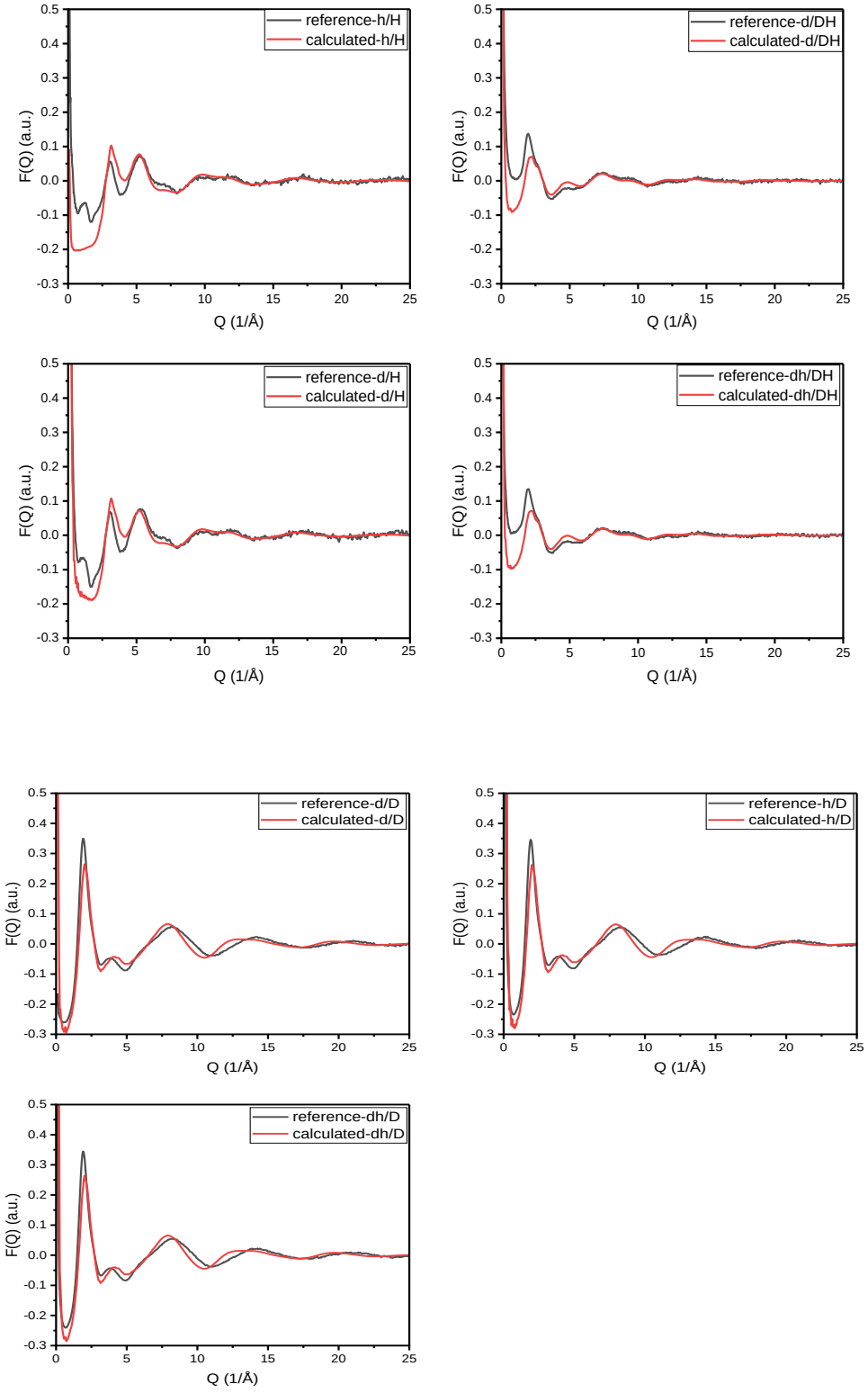


Figure 3. 13: Comparison of the experimental total neutron weighted structure factor F(Q) (black line) and the equilibrated calculated (red line) for 4 wt% PEG solutions d/H, h/H, d/dh, dh/DH, d/D, h/D and dh/D, Q range 0-25 $Å^{-1}$, in Dissolve before fixing the water geometry and applying the EPSR.

3.5.3.2 Total Neutron Weighted Radial Distribution Function G(r)

There are no clearly defined equilibrium points in the case of an amorphous substance. Rather than having well-defined atomic positions, one must refer to the atomic distribution as a function of distance. The radial distribution functions, in this case the weighted G(r), can provide a useful indicator of the quality of the total intra- and intermolecular fits for the whole system. This leads to validation of the modelling against the high experimental resolution technique in many aspects. It is worth noting that the representation of the calculated fitted G(r) in the graphs below is obtained by doubling the Fourier of the original calculated G(r), taking it to Q-space, and back again in order to get a balance comparison since the experimental data is subjected to the Fourier transform.

Fig. 3.14 shows the G(r) of experimental and neutron-weighted calculated data for h-PEG and d-PEG in the mixture of H_2O and D_2O solvents. The existence of high oscillating noise in the experimental samples containing hydrogen (in the solvent), d/H, h/H, d/DH, and dh/DH, might be related to the insufficient corrections and subtraction of the inelastic scattering, which measures the energy lost or gained by the neutron during the collision, from the average differential cross-section levels in Gudrun. These noises are not seen in deuterated solvent cases, h/D, d/D, or dh/D, which provide clear evidence for inelasticity effects being the cause of the oscillations. The intramolecular broadening was applied in the form of a Gaussian (fwhm 0.18) as a default to obtain an effectively intense peak to match the reference data. However, the mismatch in peak intensity still persisted in most of the cases. For example, the second peaks in the deuterated calculated samples (D_2O solvent) have higher intensities than their references, as does the first peak of the protonated calculated sample (h/H). This problem might either be related to the intra-broadening parameters or be related to the miscalculation of coherent scattering in Gudrun for the reference data. It is a great sign that Dissolve is able to distinguish between different contrast mixtures when looking at the small features present after the hydrogen-oxygen region at 1.5 Å in the experimental mixture samples, dh/DH and d/DH, correlating to a small broad peak in the calculated data.

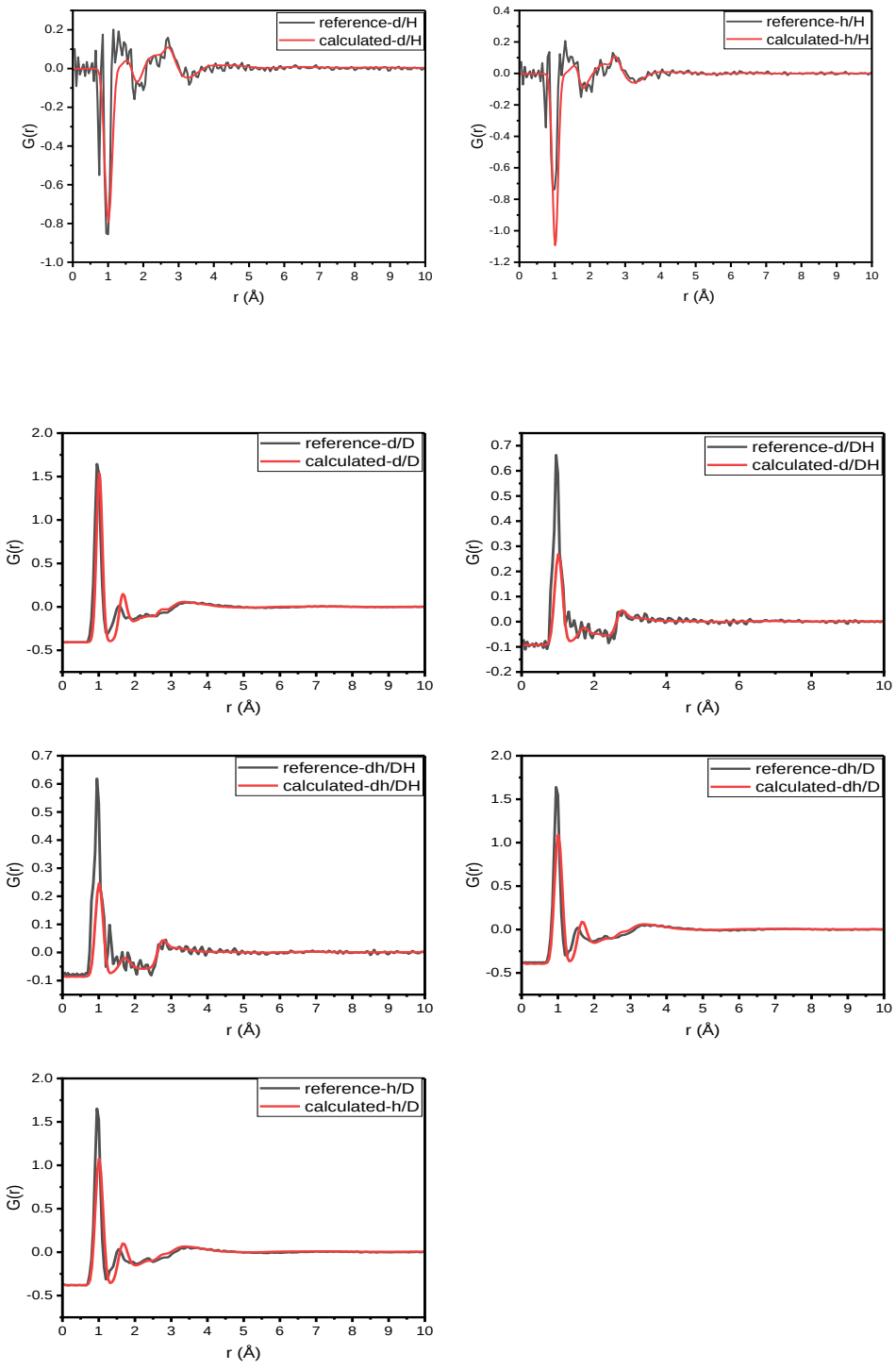


Figure 3. 14: Comparison of the experimental total weighted radial distribution function G(r) (black line) and the equilibrated calculated (red line) for 4 wt% PEG solutions d/H, h/H, d/dh, dh/DH, d/D, h/D and dh/D in Dissolve before fixing the water geometry.

The mismatch of peak position at 1 Å is observed in all representative data, which appears to be slightly shifted to the right in the calculated cases. It is significant to analyse oxygen-hydrogen (O-H) correlations since this gives information on the H-bonding network in water; the first O-H peak reflects the directly neighbouring H-bonding geometry. Subsequently, the O-H distance in the water molecule was modified from 1 Å to 0.976 Å and the angle of HÔH from 113.24° to 107.134°. Those values have been investigated by the neutron experts at ISIS, Dr. Tristan Young and Dr. Thomas Headen.[37] The simulation was run for another 1,000 iterations to adjust the intramolecular geometry structure of water, with the improvement in the results shown in Fig. 3.15, which now closely matched the experimental peak positions. These are the results, F(q) in Fig. 3.15 and G(r) in Fig. 3.16, obtained after the scattering correction raw data were repeated multiple times in Gudrun to minimise inelastic scattering effects and apply the EPSR model.

3.5.3.3 Application of the EPSR Module

The standard Empirical Potential Structural Refinement EPSR is applied to calculate the modification of the inter-atomic interactions for the PEG in aqueous solution using the EPSR function within Dissolve. The main purpose of using EPSR here was to apply an additional interatomic potential on top of those already existing in order to drive the simulation closer to experimental data agreement. The EPSR can approximate the partial structure, S(Q), with an appropriate function that has a precise Fourier transform in r-space.

The simulation was run in the default setting with Q-Max (30 $Å^{-1}$) and Q-Min (0.5 $Å^{-1}$), which determine the data range across which to create the potential in Q-space. The magnitude or strength of the generated interatomic potentials was thus determined by the value of EReq. It describes the limit of the quantity of the additional potential need for any one pair of potentials, and was set to 3.0 as a default. This number can be increased until a good fit can be obtained. In this study, we have tried some values ranges of 3-98 and the better fit was achieved at the 12 value, especially for the deuterated datasets. Unfortunately, there is no straightforward way to determine the optimal value for better fits. This is because the datasets are fitted simultaneously, but the optimal Ereq value may not be the same for every contrast, and a comprise must be made to select the Ereq value that gives best global fit for the data overall. In practise this means that the Ereq value must be redetermined for every experiment (it is not always 12).

Generally, we present good agreement between the calculated and experimental total structure factor F(Q) and total weighted radial distribution function G(r) (Figs. 3.15 and 3.16), showing an improvement compared to the unrefined data. Dissolve enhances the overall appearance of the spectra even though the simulation timestep is very low (2×10^{-5} ps), which seems to be designed for reaching the steady energy state of the configuration rather than undergoing the structural conformation change for the long chain seen in other MD tools such as GROMACS. Dissolve development is still under way, with the aim to reach a higher performance and upgrade the simulations to match the neutron scattering in different Q length scales. Our hope is to eventually drive the simulations towards conformations compatible with experimental data, as Dissolve is updated to support studies involving liquid water, liquid benzene, argon, bulk silica and many others.[37]

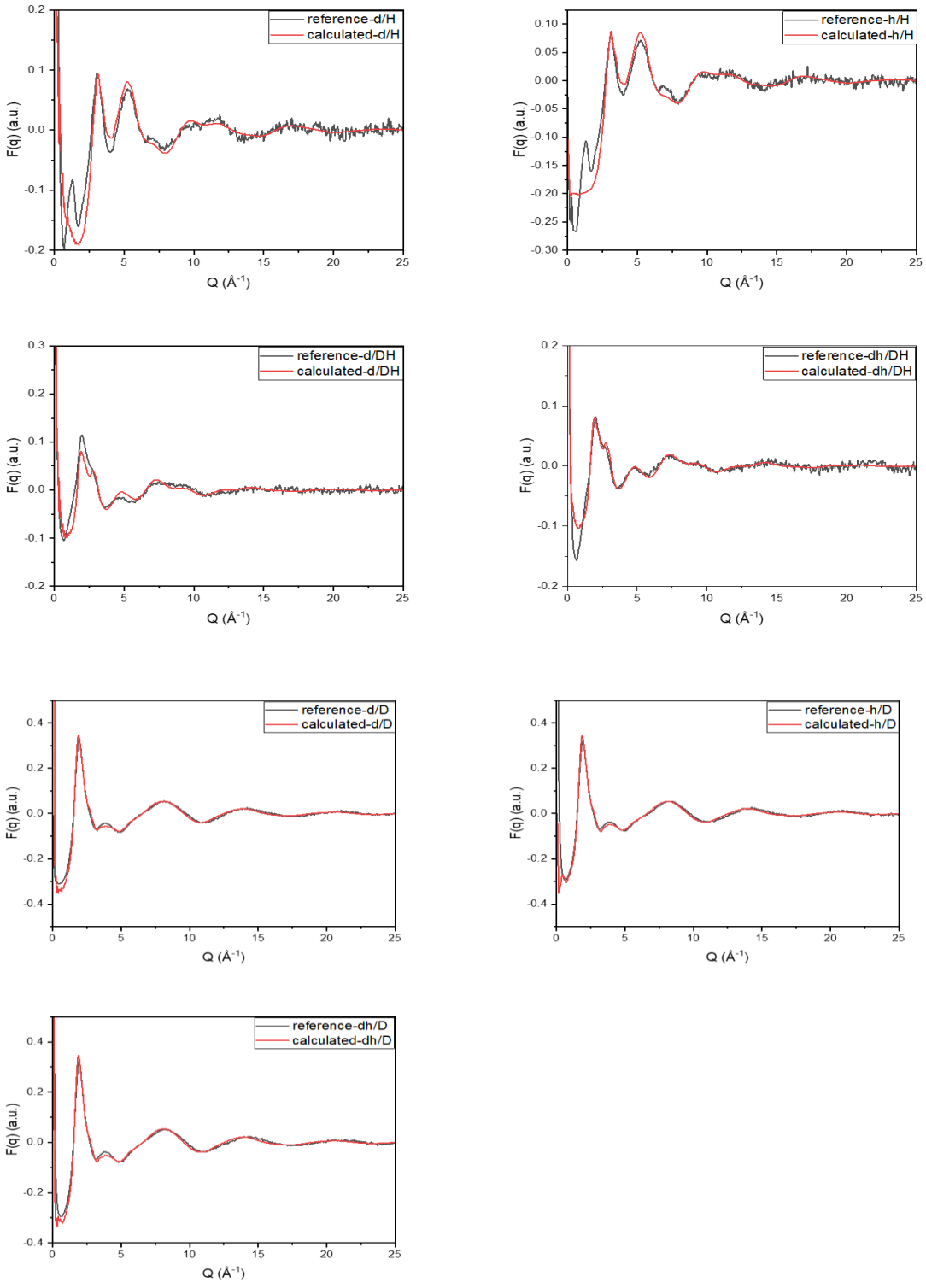


Figure 3. 15: Comparison of neutron experimental (black line) and calculated refinement (red line) total weighted structure factor F(q) after fixing the water geometry, applying the EPSR and re-correcting the NIMROD raw data in Gudrun for 4 wt% PEG solutions d/H, h/H, d/dh, dh/DH, d/D, h/D and dh/D.

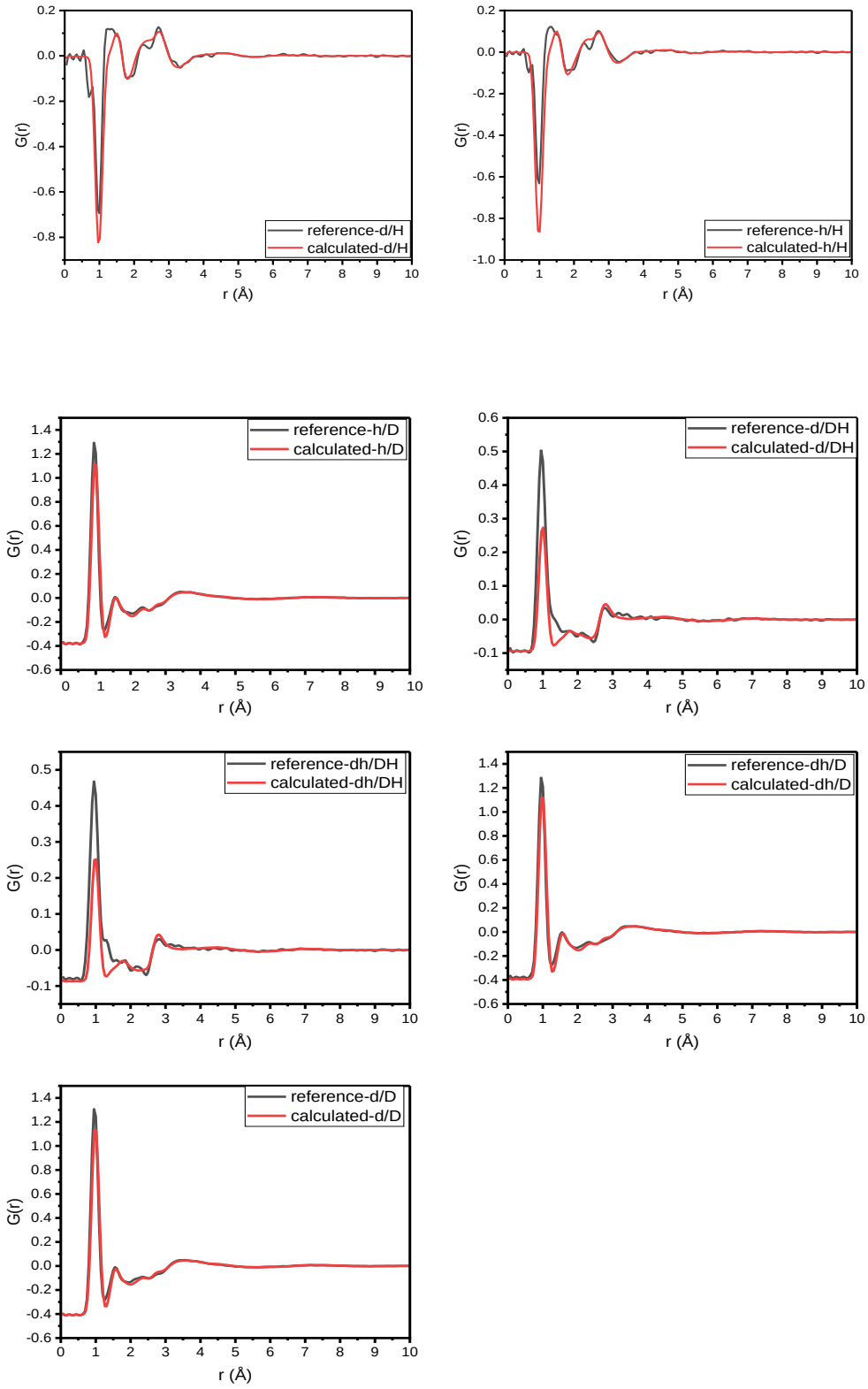


Figure 3. 16: Comparison of neutron experimental (black line) and calculated refinement (red line) total radial weighted distribution function G(r) after fixing the water geometry, applying the EPSR and re-correcting the NIMROD raw data in Gudrun for 4 wt% PEG solutions d/H, h/H, d/dh, dh/DH, d/D, h/D and dh/D.

3.5.3.4 Prediction of all partial pairs F(q) and G(r) in Dissolve

As we deal with PEG in dilute solutions, the major correlation features seen are related to the solvent-solvent relationship, as evidently observed from looking at the scaling of the sum partial corelations for solvent, PEG, and inter-atomic PEG-SOL in Fig. 3.17 (A). However, the overlap correlations for PEG solutions can happen within the same area. The main problem here is that the experimental data have the information about both the solvent and the polymer that cannot be separated, and aiming to investigate the model of polymer in aqueous solutions needs information about both solvent and PEG, which raises the challenge of testing the model. However, Dissolve can predict each pair of correlations for each component in the study system separately, which might help to determine where the issues could lie. If we take the h/D sample as an example to show the inconsistency in Q location between the experimental and the calculated F(q), Q range 0.5 – 25 $Å^{-1}$, it can be seen that the second peak located at 4 $Å^{-1}$ is less pronounced in the calculated data (Fig. 3.17 (B)). In order to examine the difference between the experimental and calculated F(q), we plotted the sum of all predicted pair correlations individually for the PEG, solvent, and inter-atomic PEG-SOL, as seen in Fig. 3.17 (A). Apparently, from inserting a straight line at the 4 $Å^{-1}$ position, that peak could be a contribution of any of the PEG, solvent, and inter-atomic correlations as long as the detected correlation is presented at that position in all of them. Thus, one of these reasons might cause this to happen: inaccuracy of the force field for PEG or even water; the conformation of PEG used in this study could not be precisely the proper model; the inefficiency of the performance of inter-atomic interactions between solvent-solvent or PEG-solvent, whether hydrogen bonding or Van der Waals forces, in the presented calculated data.

Obviously, the calculated total weighted-neutron F(q) for PEG-SOL seems to have two features that resemble the first two peaks that are seen in the SOL F(q) but with very low intensity values. The first peak starts at a very low Q value, 1.86 $Å^{-1}$, even before the first SOL peak, 1.93 $Å^{-1}$, and can overlap with it within the same region in the total F(q) system, which indicates the presence of the first solvation shell that includes the water molecules that are no longer bulk water (as mentioned earlier, section 3.3.2) when comparing the total F(q) for D_2O and h-PEG/D. The second small broad peak, which is unlike the narrow first peak, displays at 4 $Å^{-1}$ which is the Q location that shows the inconsistency with the experimental data. That small peak that exists at that position interfaces with the position of the structural factor OW-HW in the SOL as well as structural PEG correlations. Furthermore, it should be recognised that the

main correlations of the backbone end at high Q, 15 Å^{-1}, which reflects the small distance, contrary to the PEG-SOL correlations, which end at low Q, 4 Å^{-1}, which reflects the high distance. More investigation of these correlations in real space can be found in the partial weighted neutron g(r) below.

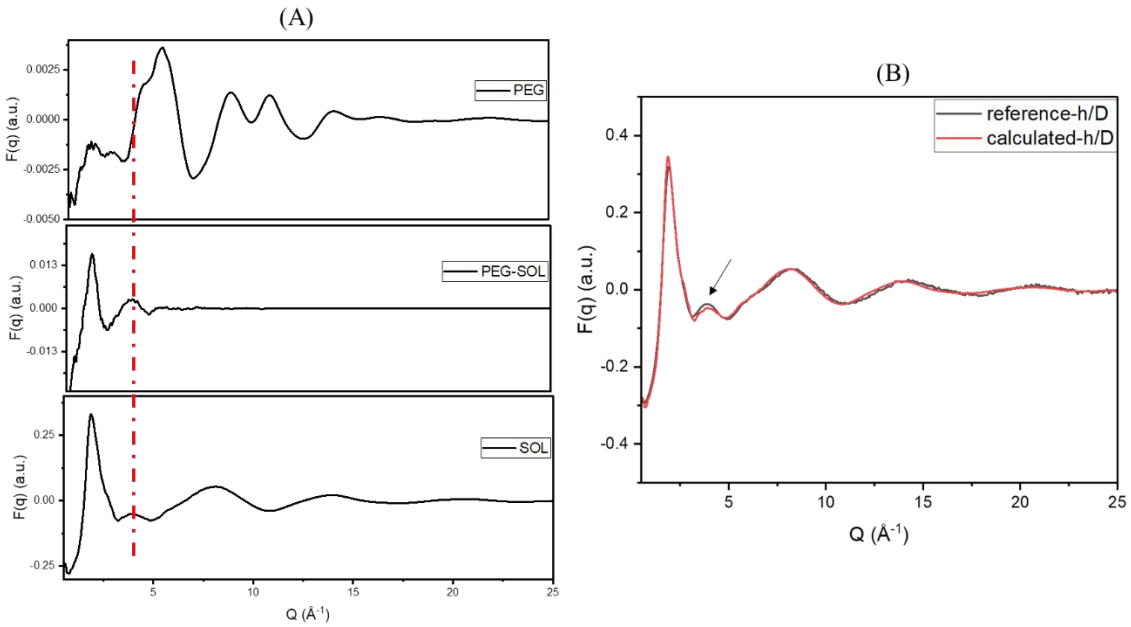


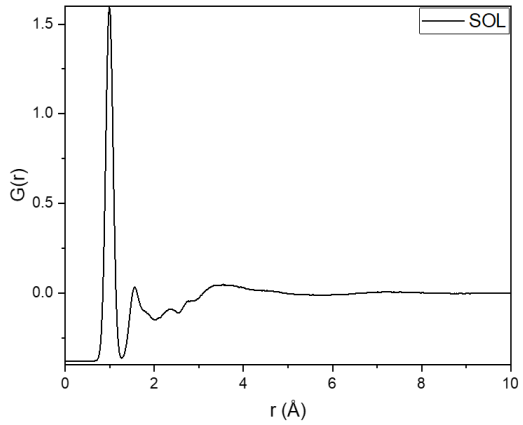
Figure 3. 17: (A) the predicted sum of all partial correlations weighted-neutron F(q) for PEG, solvent, and inter-atomic PEG-SOL (B) comparison of experimental (black line) and calculated refinement (red line) total weighted-neutron structure factor F(q) for 4 wt% h/D in the Q range 0.5-12 Å^{-1}.

Similarly, in Dissolve, the software can predict the partial pair correlations and the sum of these correlations for PEG, solvent, and PEG-SOL weighted G(r) in real space. The predicted results are shown in Figure 3.18 (A) for the sum and (B) for the partial. The representation of the PEG-atoms' names can be found in Fig. 3.18 (D). When taking the h/D as an example and comparing the experimental and the total calculated G(r) in Fig. 3.18 (C), it appears that we have a very good level of agreement between them in terms of overall existing peaks and their positions mentioned above, but it is not perfect. The first calculated peak, located at 1 Å, shifts slightly to the right of the reference, and searching for the existence of a peak at that position in the partial pairs, the OW-HW and C801-H813 (see the representation of theses atoms on Fig. 3.18 (D)) both occupy that location but with opposite intensity in the weighted G(r) for the latter.

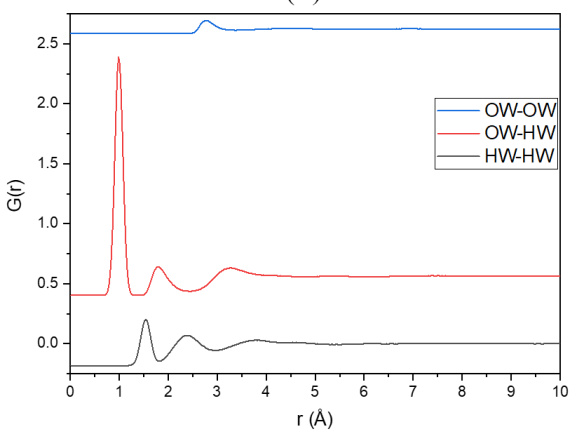
After fixing the water geometry, the next try was on the C-H bond length, which has been modified many times with lower values, e.g., 1.08, 1.07, and 1.06, that replaced the original 1.09 Å to examine any improvement, but there was no enhancement achieved. The decline of the calculated first peak does not perfectly match the reference, which affects the starting point of the second calculated peak, which starts at 1.31 Å instead of 1.25 Å in the reference data. Moreover, the discrepancy at 1.8 – 3 Å can be noticed, which is the area that could have the pair overlap of all combinations, e.g., solvent OW-HW, HW-HW, PEG H813-O803, H813-H813, PEG-SOL HW-O803. This can broaden the possibilities of testing the system, including the backbone sizes or even parameters, e.g., angles and torsions.

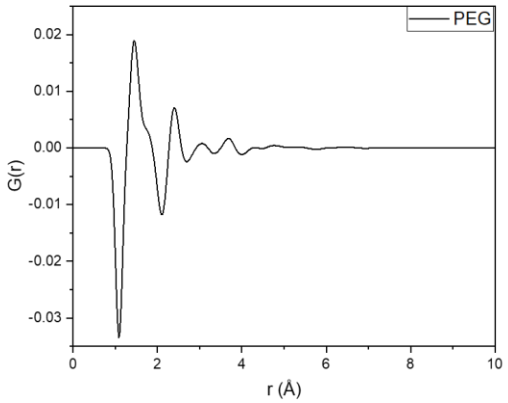
If considering the interactions between the PEG and the solvent, the sum of weighted-partial g(r) inter-atomic PEG-SOL can provide a good feature representation. The most obvious features are located around 2 Å that indicates accumulations of HW-O830, the first hydration shell, around 2.7 Å, the second layer presented in the OW-H813 zone, and around 4 Å, which belongs to the OW-C801 and HW-C801 areas (Fig. 3.18 row 3). In a non-ionised system, the water layer could be within 4 Å or even less regarding the strong hydrogen binding interactions, e.g., HW-O803 and OW-H813 as demonstrated from PEG-SOL correlations.

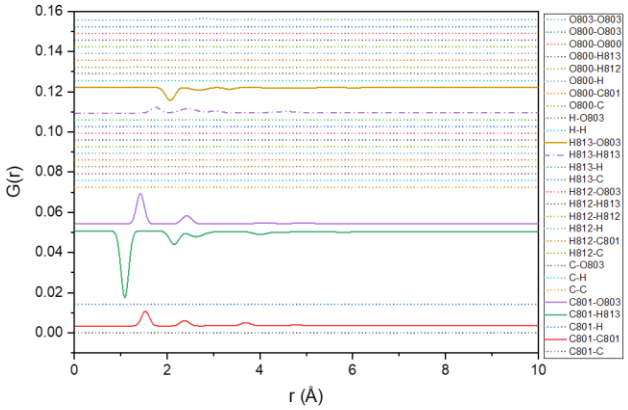
(A)

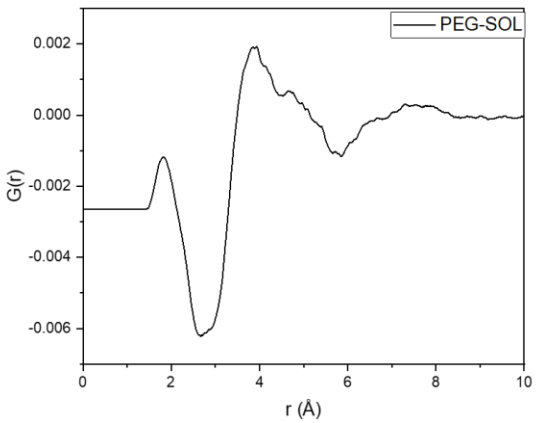

SOL
G(r)
r (Å)

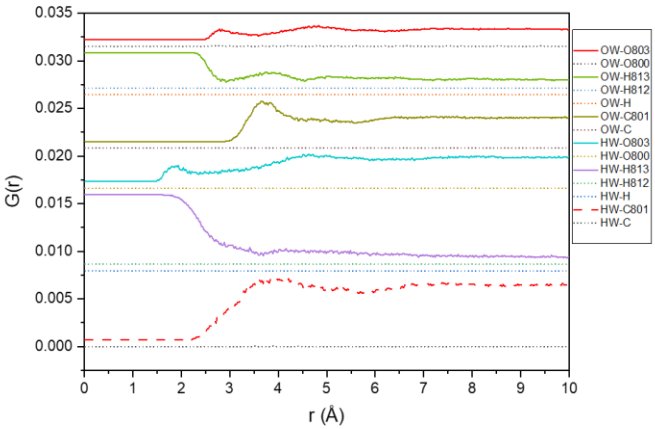
(B)


OW-OW
OW-HW
HW-HW
G(r)
r (Å)


PEG
G(r)
r (Å)


G(r)
r (Å)


PEG-SOL
G(r)
r (Å)


OW-O803
OW-O800
OW-H813
OW-H812
OW-H
OW-C801
OW-C
HW-O803
HW-O800
HW-H813
HW-H812
HW-H
HW-C801
HW-C
G(r)
r (Å)

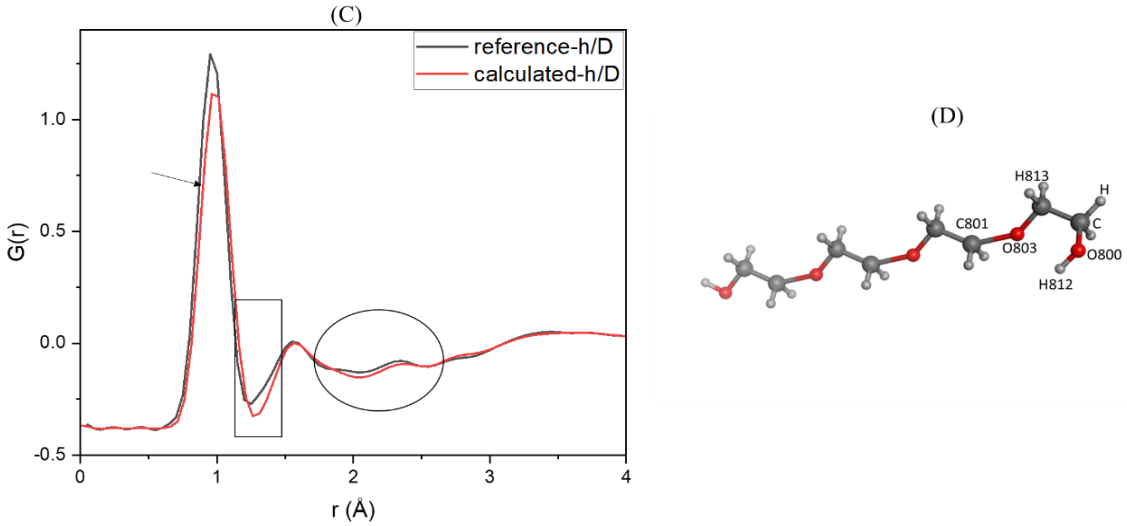


Figure 3. 18: (A) The predicted sum of all pair correlations neutron-weighted partial g(r) for solvent, first row, PEG, second row, and inter-atomic PEG-SOL, third row, (B) the predicted of all pair correlations neutron-weighted partial g(r) for solvent, first row, PEG, second row, and inter-atomic PEG-SOL, third row, (C) The comparison of the experimental h/D total neutron-weighted G(r) with the predicted total neutron-weighted G(r) from the summation of the weighted partial g(r) (D) the representation of atomic names used in this study.

3.5.3.5 Examining the different R_g conformations of PEG in Dissolve

The different R_g configurations of PEG were investigated in Dissolve to study if there is any effect on the overall F(q) or G(r) when using various conformational sizes. The starting structures were those generated in the MOE conformational search, as detailed above. Three different R_g conformations, 15 Å, 25 Å and 38 Å, were examined individually following the steps mentioned previously.

Fig. 3.19 displays the fitting F(q) and G(r) for the three R_g conformations before doing the refinement process. The main obvious difference is located at a very low Q of 0.5 $Å^{-1}$ and less obvious difference is around 5 $Å^{-1}$. At low Q, the high fluctuations in the F(q) are presented in the compacted 15 structure and less of them are seen in the 25 structure, while the extended conformation 38 shows a comparable smooth line to what was seen at that region in the experiment. This can lead to the conclusion that the elongated coil polymer can be found in the real data. There is no significant change that can be detected in the three total G(r).

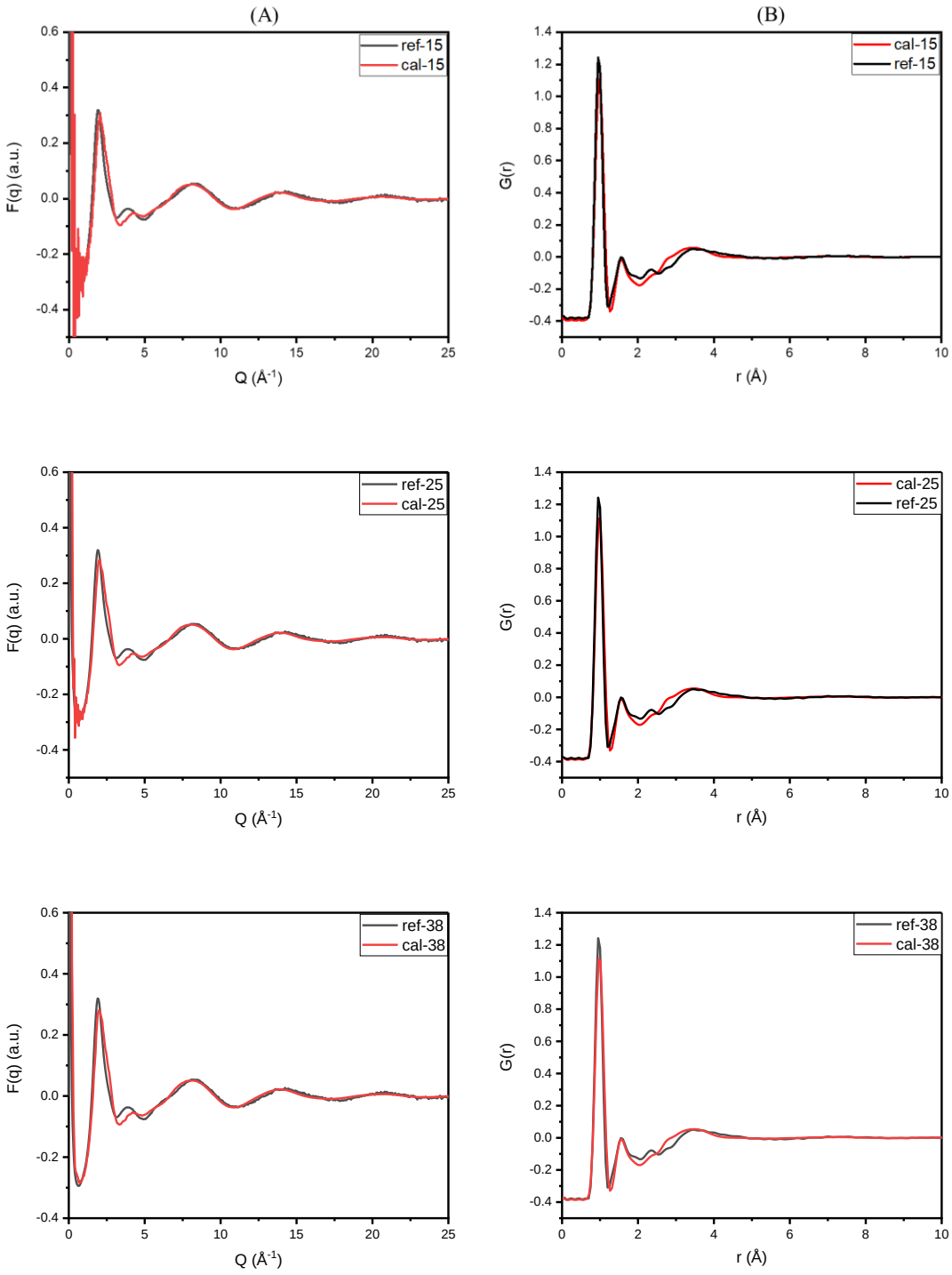


Figure 3. 19: The comparison between the experimental and calculated (A) total weighted neutron F(q) and (B) the total weighted neutron G(r) for PEG solution, h/D, in different R_g conformations (15, 25 and 38 Å). The arrows indicates common mismatch Q region in all three structures described in the text.

The EPSR process was applied to the three conformations, and the results of the F(q) are shown in Fig. 3.20. There was no remarkable change noticed in the EPSR fitting at low Q 0.5 $Å^{-1}$, where the most pronounced difference was seen between the three R_{gs}, which could indicate that the internal PEG backbone was not changing. However, looking at the expansion pictures in the low panel, the improved fitting can be observed in the 15 system more than 25 or 38 Å at Q range 3-7 $Å^{-1}$, which is within the area where the backbone correlations were detected in the partial F(q) (see partial F(q) part). Moreover, the peak shape at Q = 4 $Å^{-1}$ (indicating by the arrows in the lower panel of Fig. 3.20) differs in the three structures after EPSR, and the 15 shape is more close to the reference than the 25 or 38 shapes. In this case, we could say that internal backbone conformation may have some effects on the total F(q) spectra, and it is very hard to make any judgment about the internal backbone based on this.

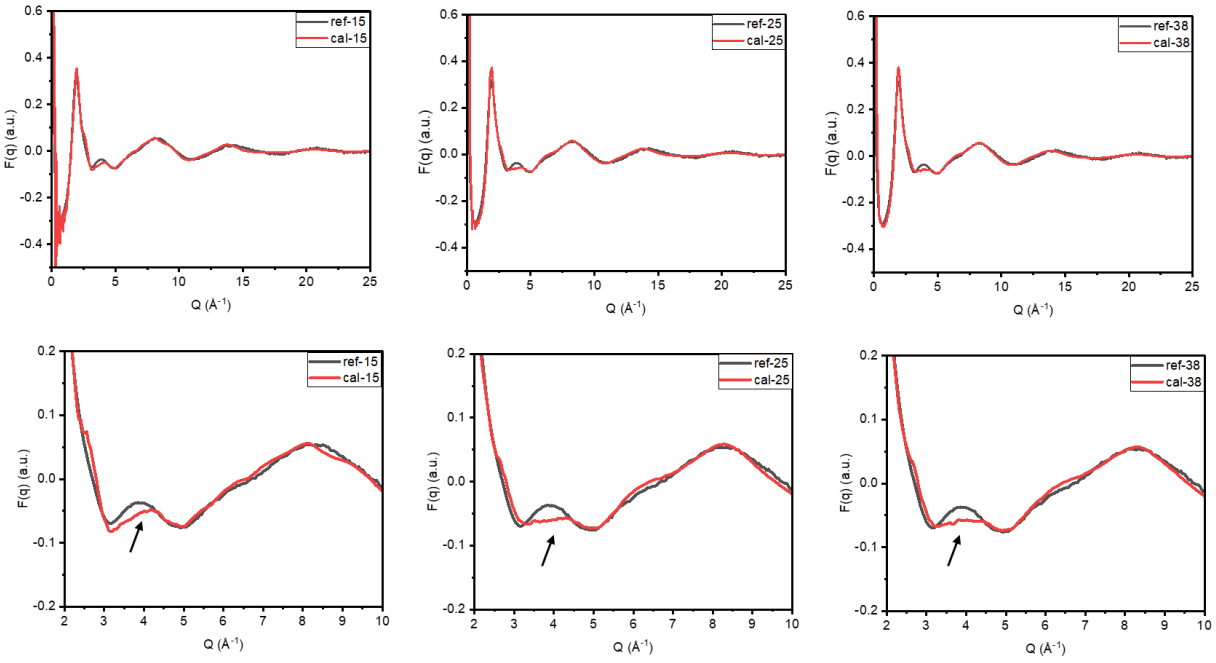


Figure 3. 20: The comparison between the experimental and calculated total weighted-neutron F(q) for PEG solution, h/D, in different R_g conformations (15, 25 and 38 Å) after EPSR was applied.

3.5.3.6 Importing trajectory from GROMACS into Dissolve

Randomly selected conformations across 4% MD, 9 frames, were imported in Dissolve to examine if any improvement could be achieved in the total F(q) when using multiple conformations simultaneously. Fig. 3.21 displays the results for F(q) and G(r) for PEG solution h/D before (A) and after (B) fixing the water geometry and applying the EPSR. Oscillations in the curve at low Q (0.5 $Å^{-1}$) in the calculated F(q) that points to the measurement of high length of the backbone as already discussed in the previous section, as well as the head peak shifting slightly to the right compared to the experiment. The selected water model can have a significant effect on the total F(q) spectra in terms of peak positions compared to the experiment. This can be confirmed when comparing the F(q) for h/D before EPSR was applied in the three water models: SPC in Fig. 3.13, after the water fixed in Fig. 3.19, and the TIP3P that was used in the GROMACS simulation, which has the H-O bond length 0.9572 Å and HÔH angle 104.52° in Fig. 3.21 (A). Furthermore, the first and second peak positions in the G(r) are shifting to the left of the experimental distances in the TIP3P model, as seen in Fig. 3.21 (A). This can confirm that the early suggestion parameters for water models, provided by Dr Tristan and Dr Tom, produce a better fit in the first, second, and even the rest of the feature peak positions. However, the first peak's intensity in TIP3P shows a better fit.

Fig. 3.21 (B) presents the results of F(Q) and G(r) for the imported frames after running the MD with fixed water parameters and EPSR applied. The improvement was achieved in both properties, the real and inverse distances. The first and second peaks were moved to the correct positions, balanced with the experiment in the total F(q) and G(r). However, the discrepancy still exists between the reference and the calculated F(q) in the same locations that have been noticed in all our previous attempts. In contrast to the reference F(q), the somewhat delayed first computed peak F(q) and the less pronounced second peak were observed. Both these represented deviations can be due to the miscalculation of inter-atomic interactions between the polymer and the solvent, as these features exist in the partial correlations PEG-SOL F(q) mentioned in the prediction partial section above. Additionally, polymer conformations can have some contributions at these locations, as noted previously especially when seeing the variations in the F(q) at low Q (0.5 $Å^{-1}$) and in the shape of the second small peak between 4 – 5 $Å^{-1}$. Introducing multiple conformations simultaneously in Dissolve results in a similar outcome as seen in the individual conformation study and no enhancement, particularly in the

discrepancy Q locations. Finally, the revealed deviations can also be relative to an insufficiently classical PEG force field or even a water model.

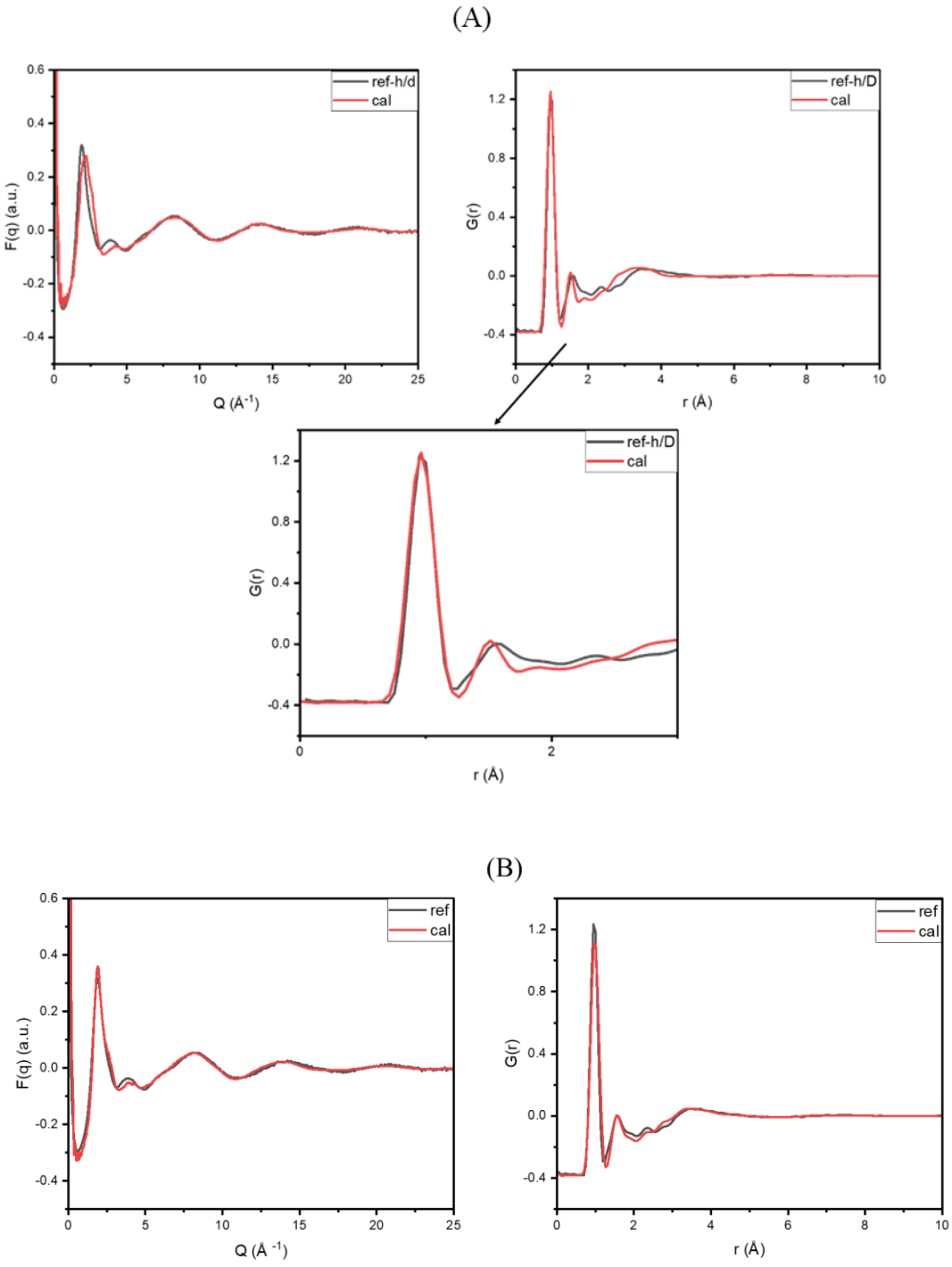
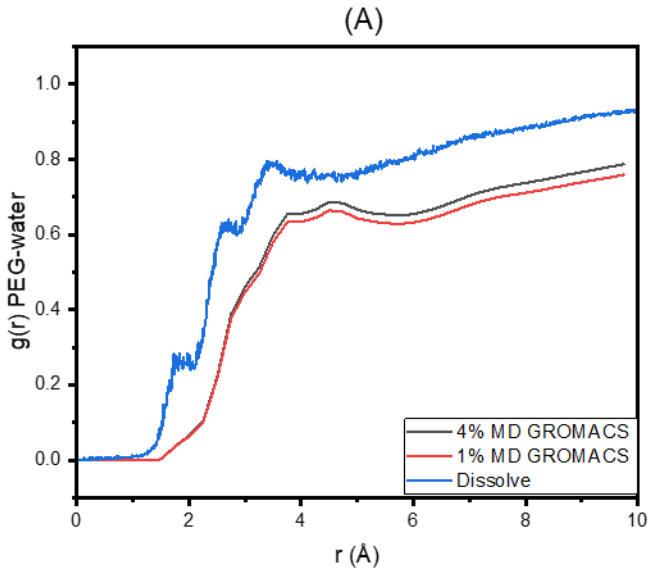


Figure 3. 21: The comparison between the experimental and calculated total weighted neutron F(q) and G(r) for PEG solution h/D from imported GROMACS frames before (A) and after (B) fixing the water geometry and applying the EPSR.

3.5.3.7 Comparison of g(r) PEG-water obtained from Dissolve and GROMACS

Figure 3.22 (A) presents a visual representation of the comparison between the g(r) PEG-water derived from GROMACS and Dissolve. The g(r) obtained using Dissolve exhibits more pronounced and distinct three feature peaks in comparison to the g(r) obtained from GROMACS. In addition, the initial solvation peak occurs at a shorter distance compared to the GROMACS function. This suggests that Dissolve may facilitate more effective intermolecular interactions between the polymer chains and water. This is attributed to the usage of driven MD with EPSR in Dissolve. In order to verify this, it is necessary to compare the quantity of water molecules present in the solvation shells seen using GROMACS vs. Dissolve. Dr Tristan employed Aten software to calculate the water population surrounding PEG using the output trajectory from Dissolve, as it is still in the development stage. In order to ensure a fair comparison, the computation was performed by utilising the imported trajectory from GROMACS (section 3.5.3.6) that has a single PEG chain. This trajectory was then imported into Dissolve to fix the water geometry and apply the EPSR. Both calculations consistently detected the same number of water molecules within a 4 Å distance from the PEG chain in each frame. Therefore, the aforementioned statement cannot be verified based on the existing findings. Figure 3.22 (B) depicts the PEG chain along with its solvation layer within a 4 Å distance in a set of randomly chosen frames, and these images were generously provided by Dr Tristan Youns.



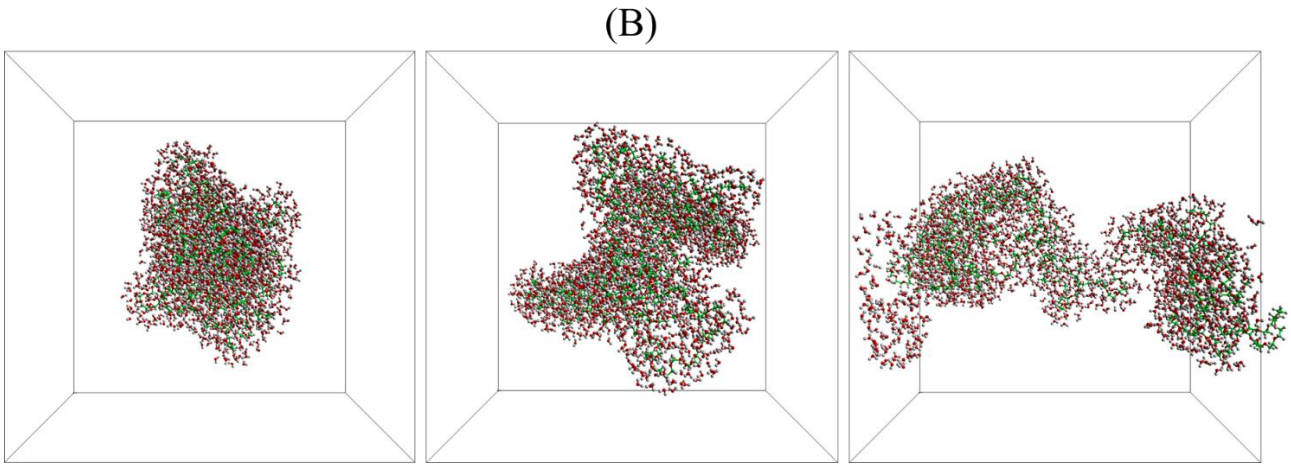


Figure 3. 22: (A) The comparison of g(r) PEG-water obtained from Dissolve (blue line) and GROMACS, 4% MD (black line) and 1% MD (red line). (B) The representation of the PEG chain with its solvation layer in some randomly selected frames (These pictures were provided by Dr Tristan Youngs).

3.6 Conclusion

In this study, the structural behaviour of several PEG chains of different lengths was studied using different force fields and solvents in MOE. The OPLS-AA in explicit TIP3P water models was selected to perform the MD simulations in GROMACS after examining the R_g in various force fields in implicit solvents in MOE and testing the SPC against TIP3P in the default box sizes (1% PEG-MD). The TIP3P simulations lead to a more sensible R_g average value (≈ 22 Å) compared to the experimental SANS value (≈ 26 Å) than SPC (42 Å). A further enhancement in the average R_g across the MD was achieved when mimicking the experimental concentration (4% MD). The calculated SANS intensity curves for PEG292 from the sum of all frames in 1 and 4% MD output trajectories independently (135,000 frames each) and the maximum, minimum and average R_g conformations from both MDs were obtained and fitted over the Q range 0–0.25 Å using the gmx_sans tool within GROMACS to investigate the sensitivity of the calculated curves on the R_g compared to the experiment. The scattering pattern for h/D seems to mediate between 1 and 4% MD, with a slight tendency more towards the 1% MD curve than 4% MD. The discrepancy between the computed and experimental patterns across 1 and 4% MD is presented after 0.1 $Å^{-1}$ due to the non-accountable solvation layer and contrast variation parameters in this calculated method. The experimental real space distribution function P(r) for h/D reveals a sensible Gaussian arrangement particle shape with the average

R_g, 25 Å, conformer in the 4% MD with a similar D_{max} of 82 Å. The RDF PEG-SOL reveals that the closest solvent are located 1.75 Å from the PEG chain. The number of water molecules was estimated within 4 Å as a function of time simulation, taking into account a non-ionised system and strong interactions.

Furthermore, the implementation of disordered deuterated/hydrogenated-PEG (Mw ≈ 13,000-14,000 g/mol) in aqueous solutions (D_2O and H_2O) using the Dissolve method, which investigates the integration of total neutron scattering data from NIMROD, introducing the first use of the software for analysis of dilute polymer solution scattering, was attempted. Dissolve enhanced the overall look of the spectra for the total F(q) and G(r), especially the peak positions for O-H, after water geometry was modified and EPSR was applied, even though the simulation timestep is very low for such a polymer size. The information of critical length scales obtained from the total scattering (NIMROD) is the key to understand the transformations in the structural features studied, yet getting the right model for the whole Q lengths has proven to be a very challenging task and needs lots of validating processes.

References

1 S. Hina, Y. Zhang and H. Wang, Characterization of polymeric solutions: A brief overview, *Rev. Adv. Mater. Sci.*, 2014, **36**, 165–176.

2 P. E. Mason, G. W. Neilson, D. Price, M. L. Saboungi and J. W. Brady, Simulation and Neutron Diffraction Studies of Small Biomolecules in Water, *Food Biophys.*, 2011, **6**, 210–216.

3 M. V. Petoukhov and D. I. Svergun, Joint use of small-angle X-ray and neutron scattering to study biological macromolecules in solution, *Eur. Biophys. J.*, 2006, **35**, 567–576.

4 T. Youngs, Dissolve: Next Generation Software for The Interrogation of Total Scattering Data by Empirical Potential Generation, *Mol. Phys.*, 2019, **117**, 3464–3477.

5 R. K. Heenan, S. M. King, D. S. Turner and J. R. Treadgold, SANS2D at the ISIS Second Target Station, *Proc. ICANS-XVII*, 2006, 780–785.

6 R. K. Heenan, J. Penfold and S. M. King, SANS at pulsed neutron sources: present and future prospects, *J. Appl. Crystallogr.*, 1997, **30**, 1140–1147.

7 D. T. Bowron, A. K. Soper, K. Jones, S. Ansell, S. Birch, J. Norris, L. Perrott, D. Riedel, N. J. Rhodes, S. R. Wakefield, A. Botti, M. A. Ricci, F. Grazzi and M. Zoppi, NIMROD: The Near and InterMediate Range Order Diffractometer of The ISIS Second Target Station, *Rev. Sci. Instrum.*, 2010, **81**, 033905.

8 N. Zębacz, S. A. Wieczorek, T. Kalwarczyk, M. Fiałkowski and R. Hołyst, Crossover regime for the diffusion of nanoparticles in polyethylene glycol solutions: Influence of the depletion layer, *Soft Matter*, 2011, **7**, 7181–7186.

9 A. H. Larsen, Y. Wang, S. Bottaro, S. Grudinin, L. Arleth and K. Lindorff-Larsen, Combining molecular dynamics simulations with small-angle X-ray and neutron scattering data to study multi-domain proteins in solution, *PLoS Comput. Biol.*, 2020, **16**, 1–29.

10 Sasview, https://www.sasview.org/, (accessed 31 August 2022).

11 L. Almásy, O. P. Artykulnyi, V. I. Petrenko, O. I. Ivankov, L. A. Bulavin, M. Yan and V. M. Haramus, Structure and intermolecular interactions in aqueous solutions of

polyethylene glycol, *Molecules*, 2022, **27**, 2573.

12 K. A. Rubinson and S. Krueger, Poly (ethylene glycol) s 2000–8000 in water may be planar: A small-angle neutron scattering (SANS) structure study, *Polymer (Guildf).*, 2009, **50**, 4852–4858.

13 G. Lancz, M. V. Avdeev, V. I. Petrenko, V. M. Garamus, M. Koneracká and P. Kopčanský, SANS study of poly(ethylene glycol) solutions in D2O, *Acta Phys. Pol. A*, 2010, **118**, 980–982.

14 K. Devanand and J. C. Selser, Asymptotic behavior and long-range interactions in aqueous solutions of poly (ethylene oxide), *Macromolecules*, 1991, **24**, 5943–5947.

15 K. L. Linegar, A. E. Adeniran, A. F. Kostko and M. A. Anisimov, Hydrodynamic radius of polyethylene glycol in solution obtained by dynamic light scattering, *Colloid J.*, 2010, **72**, 279–281.

16 I. Teraoka, *Polymer Solutions: An Introduction To Physical Properties*, New York: Wiley, 2002.

17 Gudrun - Routines for reducing total scattering data, https://www.isis.stfc.ac.uk/Pages/Gudrun.aspx, (accessed 25 August 2021).

18 A. K. Soper, *GudrunN and GudrunX: Programs for Correcting Raw Neutron and X-ray Diffraction Data to Differential Scattering Cross Section*, 2011.

19 C. Molecular Operating Environment (MOE), 2014.090 Chemical Computing Group ULC, 1010 Sherbrooke St. W., Montreal, QC H3A 2R7, 2014.

20 T. A. Halgren, MMFF VI. MMFF94s option for energy minimization studies, *J. Comput. Chem.*, 1999, **20**, 720–729.

21 D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. DeBolt, D. Ferguson, G. Seibel and P. Kollman, AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules, *Comput. Phys. Commun.*, 1995, **91**, 1–41.

22 W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic

liquids, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.

23 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, *J. Comput. Chem.*, 1983, **4**, 187–217.

24 E. W. Castner, G. R. Fleming, B. Bagchi and M. Maroncelli, The dynamics of polar solvation: Inhomogeneous dielectric continuum models, *J. Chem. Phys.*, 1988, **89**, 3519–3534.

25 R. F. Method, Multicavity Reaction Field Method for the Solvent Effect Description in Flexible Molecular Systems, *J. Phys. Chem.*, 1993, 11901–11907.

26 B. N. Dominy and C. L. Brooks, Development of a generalized born model parametrization for proteins and nucleic acids, *J. Phys. Chem. B*, 1999, **103**, 3765–3773.

27 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindah, Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX*, 2015, **1**–**2**, 19–25.

28 J. A. Lemkul, GROMACS Tutorials, http://www.mdtutorials.com/gmx/index.html, (accessed 1 June 2021).

29 W. L. Jorgensen, Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water, *J. Am. Chem. Soc.*, 1981, **103**, 335–340.

30 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.*, 1983, **79**, 926–935.

31 Group of Characterization of Materials. GCM, https://gcm.upc.edu/en/members/luis-carlos/molecular-dynamics/from-opls-to-gromacs, (accessed 10 July 2021).

32 G. Bussi, D. Donadio and M. Parrinello, Canonical sampling through velocity rescaling, *J. Chem. Phys.*, , DOI:10.1063/1.2408420.

33 H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola and J. R. Haak, Molecular dynamics with coupling to an external bath, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

34 R. W. Hockney, the Potential Calculation and Some Applications, *Methods Comput. Phys.*, 1970, **9**, 135–211.

35 M. Parrinello and A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, *J. Appl. Phys.*, 1981, **52**, 7182–7190.

36 A. K. Soper, Tests of the empirical potential structure refinement method and a new method of application to neutron diffraction data on water, *Mol. Phys.*, 2001, **99**, 1503–1516.

37 Dissolve/Home/Examples, https://www.projectdissolve.com/, (accessed 30 September 2021).

38 LigParGen,OPLS/CM1A Parameter Generator for Organic Ligands, http://zarbi.chem.yale.edu/ligpargen/, (accessed 11 November 2021).

39 Y. Wu, H. L. Tepper and G. A. Voth, Flexible simple point-charge water model with improved liquid-state properties, *J. Chem. Phys.*, 2006, **124**, 024503.

40 N. . Marks and M. Robinson, Variable timestep algorithm for molecular dynamics simulation of non-equilibrium processes., *Nucl. Instruments Methods Phys. Res. Sect. B Beam Interact. with Mater. Atoms*, 2015, **352**, 3–8.

41 H. Lee, R. M. Venable, A. D. MacKerell and R. W. Pastor, Molecular dynamics studies of polyethylene oxide and polyethylene glycol: Hydrodynamic radius and shape anisotropy, *Biophys. J.*, 2008, **95**, 1590–1599.

42 I. Vorobyov, V. M. Anisimov, S. Greene, R. M. Venable, A. Moser, R. W. Pastor and A. D. MacKerell, Additive and classical drude polarizable force fields for linear and cyclic ethers, *J. Chem. Theory Comput.*, 2007, **3**, 1120–1133.

43 U. R. Dahal and E. E. Dormidontova, The dynamics of solvation dictates the conformation of polyethylene oxide in aqueous, isobutyric acid and binary solutions, *Phys. Chem. Chem. Phys.*, 2017, **19**, 9823–9832.

44 A. Maciejewski, M. Pasenkiewicz-Gierula, O. Cramariuc, I. Vattulainen and T. Rog, Refined OPLS all-atom force field for saturated phosphatidylcholine bilayers at full hydration, *J. Phys. Chem. B*, 2014, **118**, 4571–4581.

45 M. Stepniewski, M. Pasenkiewicz-Gierula, T. Rog, R. Danne, A. Orlowski, M.

Karttunen, A. Urtti, M. Yliperttula, E. Vuorimaa and A. Bunker, Study of PEGylated lipid layers as a model for PEGylated liposome surfaces: Molecular dynamics simulation and langmuir monolayer studies, *Langmuir*, 2011, **27**, 7788–7798.

46 M. M. Hoffmann, M. D. Too, N. A. Paddock, R. Horstmann, S. Kloth, M. Vogel and G. Buntkowsky, On the behavior of the ethylene glycol components of polydisperse polyethylene glycol PEG200, *J. Phys. Chem. B*, 2023, **127**, 1178–1196.

47 G. Melesheko, Cardiff University, 2015.

48 H. Lee, A. H. De Vries, S. J. Marrink and R. W. Pastor, A coarse-grained model for polyethylene oxide and polyethylene glycol: Conformation and hydrodynamics, *J. Phys. Chem. B*, 2009, **113**, 13186–13194.

49 L. Xie, K. Y. Chan and N. Quirke, Poly(ethylene glycol) (PEG) in a Polyethylene (PE) Framework: A Simple Model for Simulation Studies of a Soluble Polymer in an Open Framework, *Langmuir*, 2017, **33**, 11746–11753.

50 P. Dalhaimer and K. R. Blankenship, All-Atom Molecular Dynamics Simulations of Polyethylene Glycol (PEG) and LIMP-2 Reveal That PEG Penetrates Deep into the Proposed CD36 Cholesterol-Transport Tunnel, *ACS Omega*, 2022, **7**, 15728–15738.

51 S. A. Oelmeier, F. Dismer and J. Hubbuch, Molecular dynamics simulations on aqueous two-phase systems - Single PEG-molecules in solution, *BMC Biophys.*, 2012, **5**, 1–13.

52 S. Piana, A. G. Donchev, P. Robustelli and D. E. Shaw, Water dispersion interactions strongly influence simulated structural properties of disordered protein states, *J. Phys. Chem. B*, 2015, **119**, 5113–5123.

53 R. B. Best, W. Zheng and J. Mittal, Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association, *J. Chem. Theory Comput.*, 2014, **10**, 5113–5124.

54 P. C. Chen, R. Shevchuk, F. M. Strnad, C. Lorenz, L. Karge, R. Gilles, A. M. Stadler, J. Hennig and J. S. Hub, Combined Small-Angle X-ray and Neutron Scattering Restraints in Molecular Dynamics Simulations, *J. Chem. Theory Comput.*, 2019, **15**, 4687–4698.

55 D. I. Svergun, S. Richard, M. H. J. Koch, Z. Sayers, S. Kuprin and G. Zaccai, Protein

hydration in solution: experimental observation by x-ray and neutron scattering, *Proc. Natl. Acad. Sci.*, 1998, **95**, 2267–2272.

56 S. J. Perkins, D. W. Wright, H. Zhang, E. H. Brookes, J. Chen, T. C. Irving, S. Krueger, D. J. Barlow, K. J. Edler, D. J. Scott and N. J. Terrill, Atomistic modelling of scattering data in the Collaborative Computational Project for Small Angle Scattering (CCP-SAS), *J. Appl. Crystallogr.*, 2016, **49**, 1861–1875.

57 M. C. Watson and J. E. Curtis, Rapid and accurate calculation of small-angle scattering profiles using the golden ratio, *J. Appl. Crystallogr.*, 2013, **46**, 1171–1177.

58 S. Grudinin, M. Garkavenko and A. Kazennov, Pepsi-SAXS: an adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles, *Acta Crystallogr. Sect. D Struct. Biol.*, 2017, **73**, 449–464.

59 F. Merzel and J. C. Smith, Calculation of small angle scattering intensities from molecular dynamics simulation, *Cell. Mol. Biol. Lett.*, 2002, **7**, 134–135.

60 F. Merzel and J. C. Smith, Is the first hydration shell of lysozyme of higher density than bulk water?, *Proc. Natl. Acad. Sci.*, 2002, **99**, 5378–5383.

61 P. C. Chen and J. S. Hub, Validating solution ensembles from molecular dynamics simulation by wide-angle X-ray scattering data, *Biophys. J.*, 2014, **107**, 435–447.

62 V. Tudisca, F. Bruni, E. Scoppola, R. Angelini, B. Ruzicka, L. Zulian, A. K. Soper and M. A. Ricci, Neutron diffraction study of aqueous Laponite suspensions at the NIMROD diffractometer, *Phys. Rev. E*, 2014, **90**, 032301.

Chapter 4

Synergies Between All-Atom Molecular Dynamic (MD) Simulations and Small- and Wide-angle Neutron Scattring (SANS/WANS) for Polyvinylpyrrolidone (PVP) in Dilute Solution

4.1 Introduction

Polyvinylpyrrolidone (povidone, polyvidone, PVP) is one of the synthetic polymers widely used in pharmaceutical formulations.[1–4] PVP consists of 1-vinyl-2-pyrrolidinone repeating chain, which has a chiral center, the carbon holding the ring, resulting in different tacticities (atactic, syndiotactic and isotactic). Because of its water solubility, biocompatibility, film-forming ability, and complexation properties, PVP is considered to be an important substance for improving medication delivery and patient outcomes. PVP, as a polymer, has a significant impact on many drug-related concerns that cannot be addressed by changing the chemistry or formulation methods. PVP aids in the conjugation of poorly soluble medicines, which can increase bioavailability and even introduce the intended swollen tract for prolonged release or control. A recent study emphasized the significant role of PVP in the development and clinical trials to evaluate therapeutic efficacy against COVID-19.[5]

In the drug delivery application, polymers used as carriers need to be studied in aqueous media. The full characterization of these polymers and their hydration processes is still limited. A combination of an experimental study and molecular dynamics simulation can probe the structural dynamics and its solvation accurately and at higher resolution (e.g., atomistic detail), than can be obtained from standard solution characterization e.g., NMR, diffusion measurements, Xray, light scattering, or viscosity measurements. However, in order to reach that, and accurate model needs to be provided, which includes the computational validation processes required such as polymer force fields, water models, simulation time and even the stereochemical configurations.

In this study, we examined the polymer structure of PVP, which is more complex than PEG (discussed in the previous chapter) due to the inclusion of side chains and stereochemistry properties. We used MD simulations and neutron scattering techniques, as well as NMR, to

analyse this. Using computational methods, we examined the behaviour of three distinct stereochemical configurations separately. These studies were performed to try and improve agreement between the R_g calculated from the MD simulations and the R_g derived through analysis of the associated SANS data. Furthermore, it is important to consider the impact of side chain effects on the properties of the polymer in comparison to the PEG chain as well as the stereochemistry.

Here, the experimental small angle neutron scattering (SANS) and the near and intermediate range order diffractometer (NIMROD) were obtained for PVP at 6 wt% in aqueous solution. These results helped to validate all atom-MD and provide insight into the behavior of the structure in diluted solution and its interactions with the solvent and solvation layer. AA-MD simulations were performed for PVP using various water models, including TIP4P, TIP5P, and SPC, as well as various stereochemical configurations, including syndiotactic, atactic, and isotactic. The computed SANS curves, along with the space distribution function P(r), were calculated using the GROMACS sans tool (gmx sans) for the three different stereochemical MDs and fitted into the experimental curve for the h-PVP in D_2O data, where the analysis here is focusing on refining the MD. Using the Dissolve software, the projected total neutron-weighted F(q) and G(r) from all-atom MD were produced and compared with experimental datasets of four distinct isotopic concentrations. Eventually, all partial pairs weighted F(q) and G(r) were studied, with a focus on the hydrogen bonding interactions between solvent and PVP.

4.2 Method

The overall methodology employed is comparable to that of the PEG study, although not identical. Details of each stage are included below for clarity of presentation.

4.2.1 Neutron scattering Experiments

SANS measurements were carried out at ISIS Neutron Facility, RAL, STFC, UK, using the equipment SANS2D. The intensity of SANS scattering, I(q), was measured as a function of momentum transfer. PVP polymer was purchased from Sigma-Aldrich, and its Mw distribution measured by gel permeation chromatography (GPC see section 4.3.1.1). Then, for the scattering experiments PVP samples were prepared at the concentration of 6 wt.% in D_2O, which is below the overlap concentration for PVP,[6] to avoid chain aggregation that can affect the radius of gyration (R_g) measurement. The samples were filled in Hellma quartz banjo-shaped cuvettes. The data utilised for analysis was collected from the original data using a

conventional data reduction procedure[7] that corrected for solvent, empty cell, and transmission measurements at room temperature.

Wide Q-range neutron total scattering measurements were collected using the Near and InterMediate Range Order Diffractometer (NIMROD)[8] instrument at the ISIS Neutron and Muon Source (Harwell, UK). As stated in Table 4.1, four isotope substitution liquid PVP samples were prepared and transferred into a flat-plate null scattering TiZr cell with 1 mm path length and wall thickness. Scattering data were also obtained from the empty instrument, the empty sample container, and a standard slab of 3 mm thickness made of Vanadium-Niobium alloy for data correction and calibration. The instrument has a wide-angle range of $2\theta = 0.5 - 40°$, and the momentum transfer (Q) range is from 0.02 to 50 $Å^{-1}$, which is calculated from $Q = 4\pi/(\lambda\sin\theta)$ with the neutron wavelengths $\lambda = 0.05 - 14$ Å. GUDRUN[9] was utilized to execute background subtraction, multiple scattering, absorption, and normalization correction processes, ultimately yielding the interference differential scattering cross-section for every isotopically distinct sample. The study of PVP was conducted using a solution concentration of 6% w/v, which was selected to be somewhat lower than the overlap concentration (c*).

Table 4. 1: Details of samples prepared for neutron scattering experiments, for h-polyvinylpyrrolidone in H_2O or D_2O water.

PVP solution prepared	Mass polymer (g)	Mass solvent (g)	Total Mass (g)	wt% PVP	
PVP/H_2O (A)	0.6272	9.2059	10.0023	6.81	
PVP/D_2O (B)	0.6339	10.4818	11.1157	6.05	
Sample-instrument preparation	PVP/H_2O (ml)	PVP/D_2O (ml)	Total (ml)	SLD solvent (10^{10} cm^{-2})	SLD dry polymer (10^{10} cm^{-2})
PVP/H_2O (A)	1.5	-	1.5	-0.56	1.162
PVP/D_2O (B)	-	1.5	1.5	6.34	1.162
Mixed-50(D):50(H)	0.800 (A)	0.800 (B)	1.6	2.89	1.162
Mixed-36(D):64(H)	0.360 (B)	0.640 (A)	1.0	1.92	1.162

4.2.2 Nuclear Magnetic Resonance (NMR)

PVP solution was prepared in D_2O, with roughly 6 wt% polymer, and ^{13}C NMR was acquired at 85°C on the Bruker 400MHz NMR and data recorded with 256 scans. The chosen solvent, D_2O, was based on the previous study, which included ethylene glycol, benzene, pyridine, 1,1,2,2-tetrachloroethane, and D_2O, and the best solvent for the tacticity spectra was determined to be D_2O. Since a higher temperature seems to improve signal resolution and enable the acquisition of comprehensive tacticity information, the NMR experiment was conducted at 84 °C.[10]

4.2.3 Computational Method

4.2.3.1 MOE-GROMACS

All the PVP chains were constructed (for PVP91 the different stereochemical configurations were built: atactic, syndiotactic, and isotactic) and the stochastic conformational search performed using three different force fields OPLS-AA,[11] CHARMM27,[12] and AMBER99[13] in Born solvent using MOE2020.0901 software package[14]. The average R_g structures from each search were then used as initial configurations to run the all-atom explicit solvent MD simulations.

All-atom explicit conventional MDs were carried out in GROMACS 2018.229 software[15]. The initial configurations were solvated in the two different boxes, default box (90 $Å^3$) and in the unit cell ≈ 64 $Å^3$ that meets the experimental concentration 6 wt.%. PVP91 was selected based on the observed Mw = 10062 g/mol, and the Mw for the monomer of 111 g/mol. Then, the system were minimized to eliminate any steric conflicts or bad geometric features, within OPLS-AA in the SPC[16] and the different TIP family of solvent models, TIP3P,[17] TIP4P,[18] TIP5P[19]. After the energy minimization, the following step is the NVT equilibration with random velocity generation from Maxwell distribution rescaling temperature coupling[20] at 298K. The NPT equilibration came after and the velocity continuation from NVT, regulated with Parrinello-Rahman pressure[21] coupling to 1 bar. These NVT and NPT equilibration simulations were run for 10 ns using 2 fs time steps with position restraint applied for the chain in both ensembles. The final MD simulations were performed after removing the polymer restriction to solve for motion-equations with the leapfrog approach[22].

A further set of simulations using the CHARMM36 forcefield used the CHARMM-GUI platform[23] to set up the linear-atactic PVP91 structure (in the relative box size to the one used in the OPLS-AA MD) and solvated with TIP3P water model, and the system was minimized and equilibrated. The production of the equilibrated structure with its topology were used to run the conventional MD in GROMACS, using CHARMM36 FF,[24] following the same process mentioned in the OPLS-AA MD above.

4.2.3.2 Dissolve simulation

Based on the neutron total scattering data, the three-dimensional atomic structure of the PVP solution sample is simulated using the Dissolve programme[25]. The structure of water molecules are created in Dissolve (with the modified parameters mentioned in the PEG section on the SPC force field); the PVP structure, taken from stochastic search minimized in MOE, was imported into Dissolve. The molecules' initial reference potentials are the classical standard OPLS-AA forcefield taken from the LigParGen service of Yale University,[26,27] which yields details about their bonds, angles, and torsion angles in addition to their Lennard-Jones potentials. A cubic simulation box of 80.6417 $Å^3$ was populated with 16448 water molecules and 2-PVP chains matching the experimental concentration 6 wt%, that corresponding atomic density of 0.1 atoms/$Å^3$. The MolShake module was utilised to minimise the initial system through the application of a 300-step MD simulation with a timestep of 2×10^{-5} ps applied every fifth Monte Carlo cycle, together with whole-molecule Monte Carlo "shakes." After more than 6,000 iterations, the system reached stabilization energy. After the system was stabilized, the EPSR module of Dissolve was introduced in order to improve the intermolecular potential match between the simulated-box and the four experimental neutron scattering datasets.

4.3 Result and Discussion

4.3.1 Experimental section

4.3.1.1 Gel Permeation Chromatographic (GPC) for PVP

The molecular weights and molecular weight distributions were measured by gel permeation chromatography (GPC) and the data provided by Dr Elaine Ferguson at Cardiff Dental School,

using Ultrahydrogel 250 x2 columns and Phosphate-buffered saline (PBS) at 30 °C. This measurement was made to double check the distribution and average Mw for the purchased polymer, which can help in getting a clear indication of molecular weight, size distributions, and constructing the chain length for the modeling side. Fig. 4.1 shows the PVP has one broad peak that belongs to the PVP populations with slightly wide-size distributions, followed by a narrow peak that belongs to the mobile phase (solvent). The obtained average Mw and polydispersity values of the PVP are 10,062 g/mol and Mw/Mn 1.635, respectively. This corresponds to an average of 91-monomer units per chain.

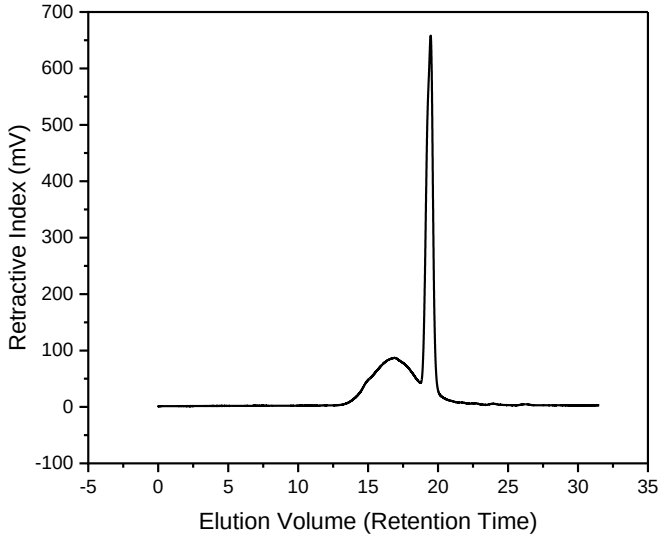


Figure 4. 1: GPC curve, RI vs retention time, that indicates the molecular weight distribution for the purchased PVP.

4.3.1.2 SANS fitting (SasView)

The R_g for PVP, 10,062 Da, was determined using poly_gauss_coil model[28] in SasView5.0.5[29]. For polydisperse data, as the polydispersity for the purchased polymer is 1.635, employing a polydisperse model or incorporating chain size distribution within the software might assist to enhance the accuracy of the R_g estimation. The experimental scattering pattern (Fig. 4.2, black symbols) represents a typical polymer scattering where the sample has some polydisperse particles. The fitting model, in Q range 0.013-0.5 $Å^{-1}$ (avoiding the noise points), along with standard deviation are shown in Fig. 4.2, and fitting parameters displayed in Table. 4.2. A good fit was achieved with R_g value of 18.6 ± 2.5 Å, however, the effect of the polydispersity on

reducing chi-square X^2, a statistical test that determines how well a selected model fits the data with a given set of model parameters, was noticed during the fitting process, which reduced more with higher polydispersity value, indicating the disperse sample. The molecular weight distribution obtained from the GPC results can also confirm that by looking into the broad shape of the curve around the average Mw rather than a sharp peak. The radius of gyration determined in this study, which is 18.6 $\pm$ 2.5 Å, is smaller than the radius of gyration (25-30 Å) found in a recent study[30]. It is worth to remember that the fitted R_g obtained from SasView for the experimental SANS data takes into account the polydispersity parameter and its effect on the R_g calculation and fitting the model. This should be recalled in the modeling analysis part, especially when examining the R_g of the single chain MD productions.

When comparing the radius of gyration (R_g) acquired in this investigation to the hydrodynamic radius (R_h) measured using dynamic light scattering (DLS) measurements in a recent study[31] with a similar molecular weight (Mw, 10000 g/mol), our value is significantly smaller than the reported one (46 Å).

Fundamentally, the information included in the scattering intensity I(q) is the same as the distance distribution function P(r), but the latter function is a more simple and intuitive visual representation.[32] To explore the shape, geometry and spatial intraparticle relations, the real-space function P(r) for PVP 6 wt.% was gained from the intensity scattering curve using the inversion approach[33]. The P(r) demonstrates the shouldering peak, ~ 40 Å, with extended tail, 40-67.8 Å, that decays slowly to a D_{max} 67.8 Å, as displayed in Fig. 4.2 (lower panel). The derived R_g obtained along with P(r) function, 19.1 Å, is in agreement with the value obtained by the poly_gauss_coil model above, 18.6 $\pm$ 2.5 Å. The elongated features scattering curve and the distance distribution function with the unsymmetrical shouldering peak suggest an elongated rod particle shape.

The aforementioned rod shape suggestion prompted us to re-evaluate the intensity curve using the cylinder model. The cylinder fit to the experimental intensity curve is shown in Fig. 4.2 (bottom right panel), where a good fit is observed. The dimensions of the cylinder model are 7.5 Å for a radius and 58 Å for a length. However, the lower statistical quality of the fit when considering the reduced Chi2 value, 4.45, is obtained compared to the coil fit model, which has a less value 2.59. This suggests that the coil model is a more sensible fit to the PVP scattering data than the cylinder model, which is unlikely to be entirely accurate for a PVP system. Although it does not perfectly fit into the cylinder model, the observed rod shape in the

P(r) function may be attributed to the existence of side chain pyrrolidone rings, which supports some rigidity in the model. Similar findings were found in an earlier investigation[34] on the HPMA copolymers containing aromatic amines at two different substitution levels s, 5 and 10 mol% with respect to the monomer repeat units; the worm-like model is the best fit describing the rigidity for 10 mol % while the random coil is best for describing the 5 mol %. Hydrophobic factors that reduce unfavourable hydrocarbon water interactions are what cause these structural alterations, and this explains why the smaller conjugate for the same HPMA backbone, the conjugate with 5 mol% is a coil and changes to a more rigid chain for conjugates with 10 mol% modification. Because of the diluted solution we are using, the hydrophobicity effect may not be able to cause a change in the morphology as a complete rigid formation.

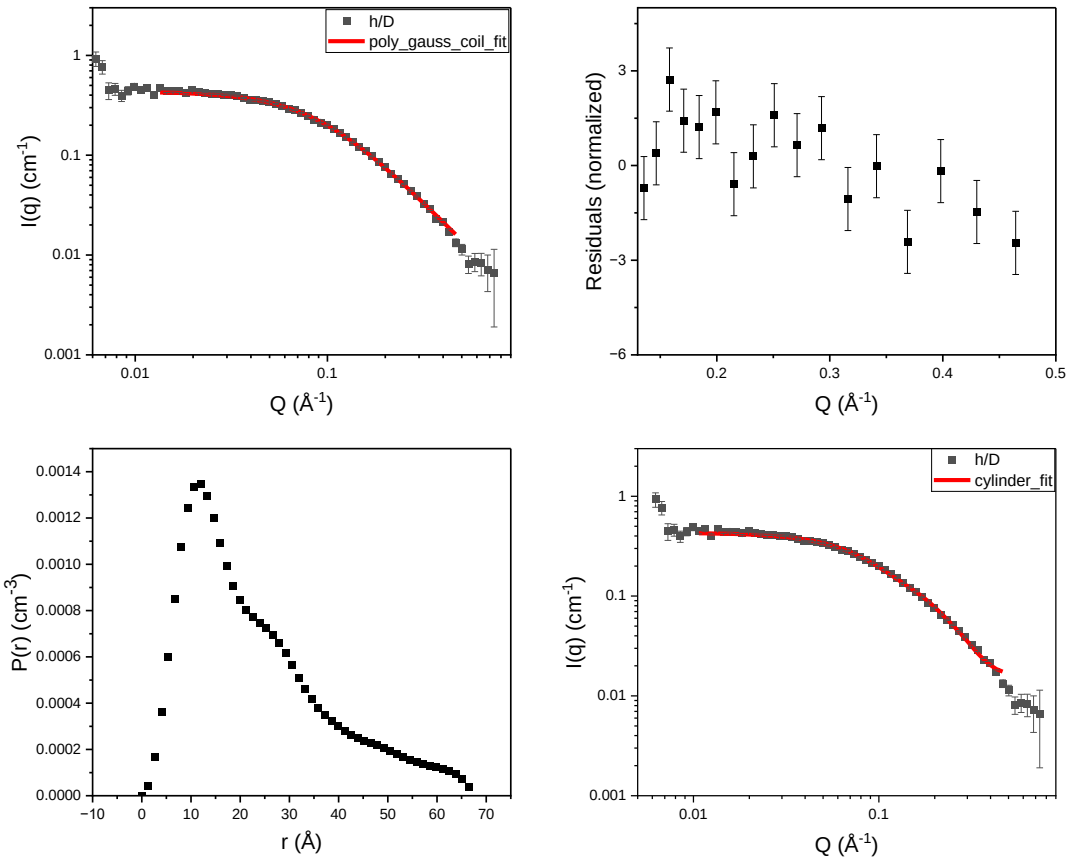


Figure 4. 2: The fitting model, poly_gauss_coil, (red) into the experimental SANS pattern, I(q) versus Q, for the PVP 6 wt. % (black) within the Q range 0.013- 0.5 Å^{-1} (upper left), and the residuals (normalized), which represents the standard deviation (upper right), the inverse-space distribution P(r) obtained by the inversion analysis for the SANS data (lower left panel) and the fitting model, cylinder, (red) into the experimental SANS pattern, I(q) versus Q (lower right panel).

Table 4. 2: The summary of the fit results for PVP, 10,062 Da, 6 wt % in D_2O using SasView

model	poly_gaussian_coil
scale	1
bckgrd	0.0005
bckgrd_err	0.0004
I(0)	0.433
I(0)_err	0.0016
R_g	18.6
R_g_err	0.089
polydisp	1.635
polydisp_err	0.089
reduce chi2	2.59

Scale: scale factor or volume fraction; bckgrd : source background (cm^{-1}); I(0): intensity at q=0 (cm^{-1}); R_g: radius of gyration (Å); poly_disp: polymer Mw/Mn; reduced Chi2: a statistical evaluation of the degree to which a selected model, given a set of model parameters, fits the data; errors: the parameters' uncertainties.

4.4 Modeling Section

4.4.1 Examining the evolutions of the R_g (atactic)

To approximate thermodynamic properties of polymers in solutions, computational approaches based on either explicit or implicit solvent models can be used. Here the R_g of PVP chains were predicted for chains involving 6, 12, 24, 48 and 91 monomers (Mw≈ 666-10,000 g/mol), in two water models, Generalized Born (implicit), and TIP3P (explicit), with OPLS-AA force field was evaluated, and used to fit Flory approximation solvent exponent. The choice of the OPLS-AA was based on our previous successful evaluation for PEG in TIP3P, that showed good sampling with a close average R_g to the experimental SANS value. The growing chain was assessed in those models and used to plot the average R_g of each monomer versus the number of repeat units. The average R_g for the small monomer, 6 and 12 have similar values in both solvent models, 4.8, 4.7 and 6.7, 6.5 Å respectively as shown in Fig. 4.3 (A). The average R_g value for the 24 monomer in TIP3P has a slightly higher value 9.6 Å compared to the Born 9.0 Å. Surprisingly, the explicit solvent model produced smaller average R_g for 48

and 91 monomers, 10.7 and 13.6 Å, than the implicit Born, 11.8 and 16.2 Å, which gave better R_g values for those chains, when compared to the experimental value.

The Flory exponent value was measured experimentally[35] with a value of 0.521 by a dynamic light scattering technique for different polymer molar masses at 25◦C. The trendline as a power was added into the line of each model to determine the solvent exponent that describes the behavior of the chain in these solvents. The value obtained for the Born is higher, 0.438, than the TIP3P, 0.384. The unexpected chain behavior in the explicit model generates many questions about the tested model. The expectation is when individual solvent molecules are explicitly present, permitting direct interactions with the polymer can lead to swelling or expansion of the polymer, resulting in increasing the polymer's R_g. On the other hand, the continuum implicit solvent can neglect some hydrogen bonding interactions, resulting in lower R_g. However, this does not happen here, and it seems the repulsive solute–solvent interactions increase when using the explicit TIP3P solvent for PVP, which stabilizes the solute, increasing compacted structures. The unfavorable solute-solvent interactions can alter the solute's behavior and characteristics, and this can be due to many different reasons such as the hydrophobic backbone effect, and the intramolecular bonding in the internal structure which may prevent the solute-solvent interactions.

The experimental R_g measurement we have for PVP with Mw ≈ 10,062 g/mol, for the 91 repeat units, was estimated by SANS technique at 18.6 Å ± 2.5, as discussed in experimental SANS section, which is closer to the value obtained from the Born, 16.2 Å than the TIP3P, 13.6 Å. Fig. 4.3 (B) displays the running average R_g vs time, running average each 25 ns, for PVP91 in OPLS-AA/TIP3P with three parallel MDs, 500 ns each, with three different starting velocities in the default box size (90 $Å^3$). The R_g decreases in a few nanoseconds in all three runs, and the molecular conformation evolution suggests the trend to form more compact structures. The first run has a lower mean R_g value of 13.0 Å with SD 0.04 Å, followed by the second run with mean R_g of 13.6 Å and SD 0.04 Å and finally the third run with slightly higher mean R_g, 14.5 Å and SD 0.74 Å, than the previous two runs. This could indicate that the starting velocity may have an effect on the PVP molecule performance and need longer time equilibrate/reach the experimental values. A similar result for the PVP compact structures is also discussed in the literature[36], the simulation for PVP 40,000 g/mol, 408 monomers, carried out with the GROMOS53a4 united-atom force field in TIP3P and resulted in compact conformations with R_g of 22.6 Å.

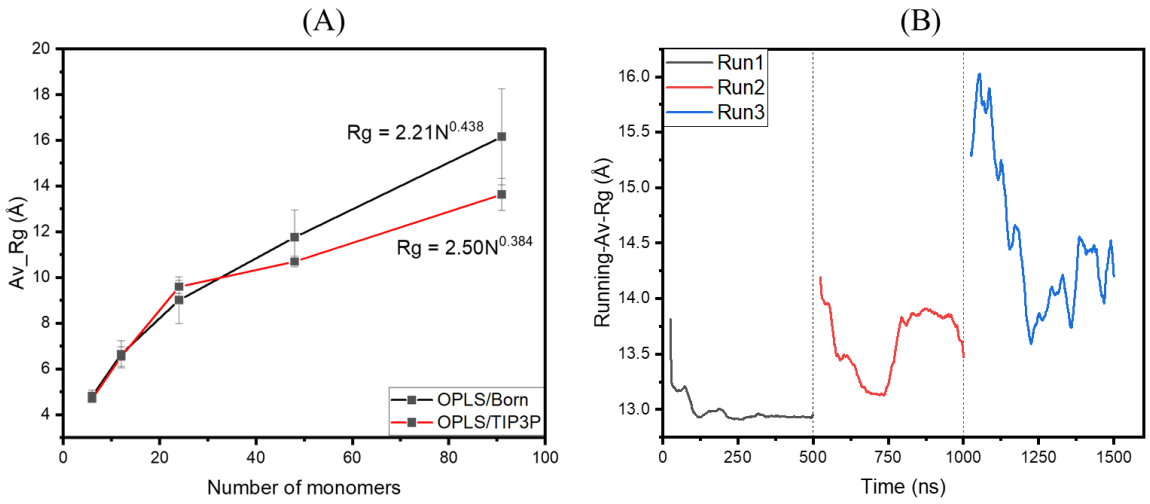


Figure 4. 3: (A) Comparison of average R_g for PVP, 6, 12, 24, 48, and 91 monomers, in OPLS-AA/explicit TIP3P (red line) and OPLS-AA/implicit Generalized Born solvent (black line) (B) running average R_g for atactic PVP91 in OPLS-AA/TIP3P versus simulation time (ns) in the default box size, 90 $Å^3$.

Beside the system of interest, the choice of the explicit solvent model can play a crucial role in the study of the solute behavior in solution.[37] Significant number of explicit water models have been constructed in an attempt to incorporate all of the physicochemical features of water. However, the study of their solute-solvent interactions requires additional investigation.[38] Further MDs were carried out for PVP91 using OPLS-AA in different explicit water models, TIP4P, TIP5P and SPC, to investigate which reliable water models that can provide accurate description for the PVP91 sampling. The 100 ns evolution of the radius of gyration for PVP91 in four different water models, TIP3P (right lower panel), TIP4P (left upper panel), TIP5P (right upper panel), and SPC (left lower panel) in the cubic size $\approx$ 64 $Å^3$ are shown in Fig. 4.4. The results show that the R_g of the molecule drops after 40 ns in all four models and does not increase again to the original one. In the TIP4P, the molecule resists to be expanded in the first 40 ns but soon alters and drops, while in the TIP3P, TIP5P and SPC drops from the first few iterations. This indicates that there is no significant difference in the overall description of the R_g for PVP91 in all tested water models. The overall average R_g across the TIP3P MD was noticed to be slightly better in the 64 box, 14.3 Å, compared to the TIP3P MD in the 90 box, 13.7 Å, which was then used in all tested water models.

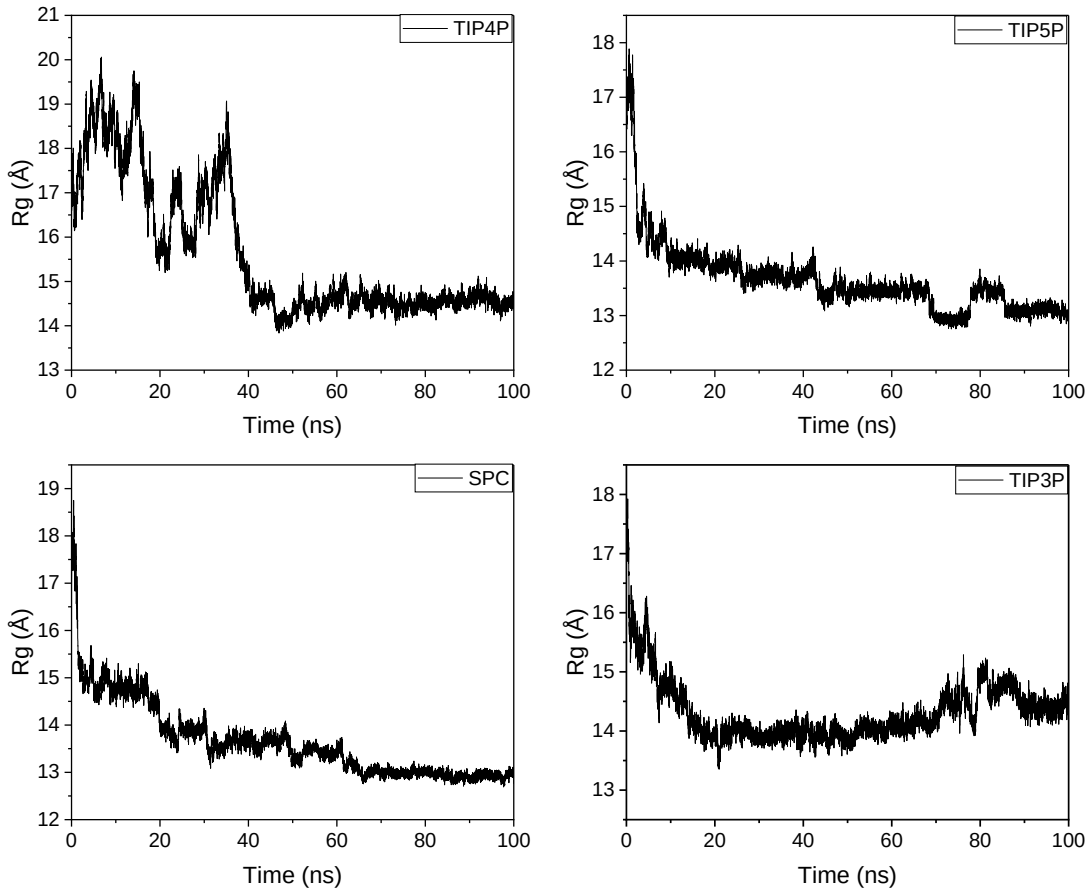


Figure 4. 4: Evolution of the radius of gyration versus simulation time (ns) for atactic PVP91 using OPLS-AA in four different water models, TIP3P, TIP4P, TIP5P, and SPC in the box size ≈ 64 Å^3.

After observing similar R_g results in the different solvation models for PVP91, the next attempt was to go back and examine the chain behavior using the two different widely used force fields, CHARMM and AMBER, in order to obtain the Flory exponent and see if there is any improvement in that value. The results from testing different length chains of PVP, 6, 12, 24, 48, and 91, using CHARMM27 and AMBER99 in the Born model, were used to plot the average R_g versus the number of monomers with the production of the solvent exponent for each models shown in Fig. 4.5. The results demonstrate there is no significant difference between these tested force fields in the chain attitude, as they behave very consistently in the implicit model. The value obtained from CHARMM27 is 0.459, which is very similar to the AMBER value 0.457. These values are better than the OPLS-AA value 0.438 in Born,

compared to the measured one, 0.521. Further action is needed to examine the performance of the PVP molecule in the explicit solvent using at least one of these evaluated forcefields.

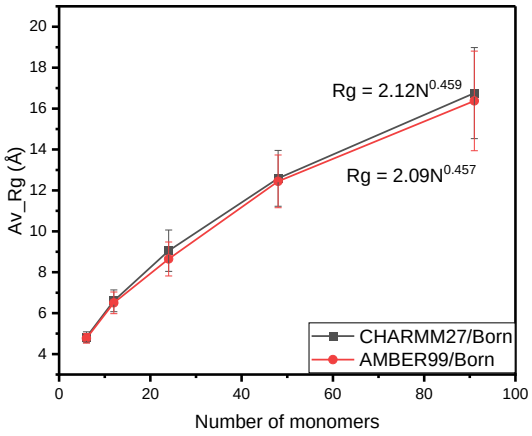


Figure 4. 5: Comparison of the R_g versus number of monomers between CHARMM27 and AMBER99 in implicit Born model with production of the solvent exponent for each model for atactic PVP.

A further 300 ns all-atom explicit MD was carried out using CHARMM36, for PVP91, in TIP3P, in order to explore the evolution of the R_g in the explicit water using this FF. The initial structure was built by CHARMM/GUI service (see Fig. 4.6 (B)). The variation in R_g is observed through the MD with maximum, average and minimum values of 32, 24, and 17 Å, respectively, after the removal of the initial 50 ns, as shown in Fig. 4.6 (A). When comparing the R_g for PVP91 in Born and TIP3P in CHARMM FF, there are more extended conformations pronounced in the explicit than the implicit model that could allow more hydrogen bonding to associate with the polymer, which resulted in sensible and expected results and seem that we achieved a good sampling with flexible chain movement during the MD. However, the average R_g, about 23 Å, obtained across the MD trajectory appears to be much higher than the experimental SANS value, 18.6 Å. While the average R_g obtained from the OPLS-AA/explicit MDs is about 14 Å. This led us to do more investigations on the internal structure to study the problem behind the compact conformations seen in the OPLS-AA/explicit water.

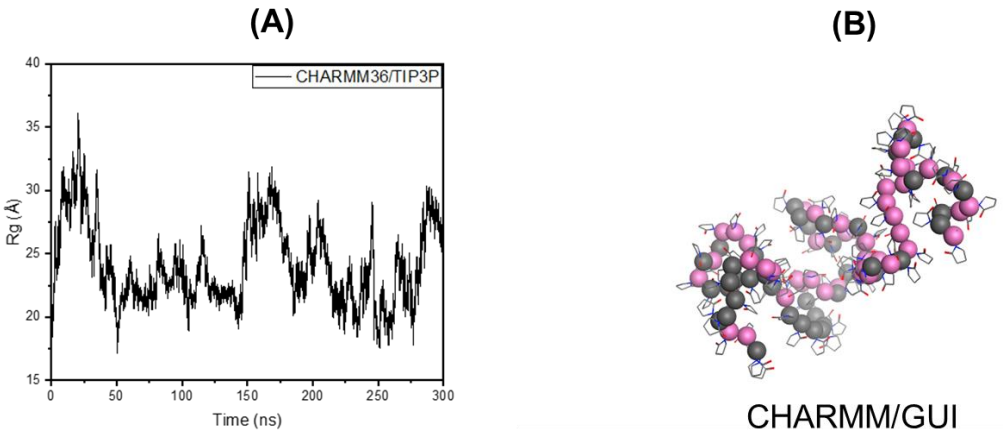


Figure 4. 6: (A) Evolution of the radius of gyration versus time simulation (ns) for atactic PVP91 using CHARMM36 in TIP3P water models in the corresponding OPLS-AA simulation box size, (B) the initial structure was built in CHARMM/GUI the pink and dark gray colors represent R and S centers, with the overall polymer being atactic.

4.4.2 Validation of R_g for atactic-PVP91 model constructed by CHARMM-GUI in several independent MD runs

The next attempt was to use the model constructed by the CHARMM/GUI service mentioned above to test the random atactic model's accuracy and see whether the chain behavior utilizing the OPLS-AA/TIP3P models could be improved. The 6 independent MDs were carried out for 500 ns each, and the evolution of the R_g versus simulation time (ns) is shown in Fig. 4.7, and the overall statistical analysis of R_g is shown in Table. 4.3. Obviously, the results demonstrate that the obtained average R_g across all MDs is 13.7 Å, with a maximum of 15.1 Å, a minimum of 13.0 Å, and a standard deviation of 0.3 Å. Therefore, there is no improvement observed in the R_g value. This indicates that the randomized construction of the atactic PVP, whether in MOE or CHARMM/GUI, does not have a significant effect on the R_g calculation.

After all attempts above to investigate the MD for PVP91 based on the derived scattering data, now we need to examine the chain behavior in different stereo configurations.

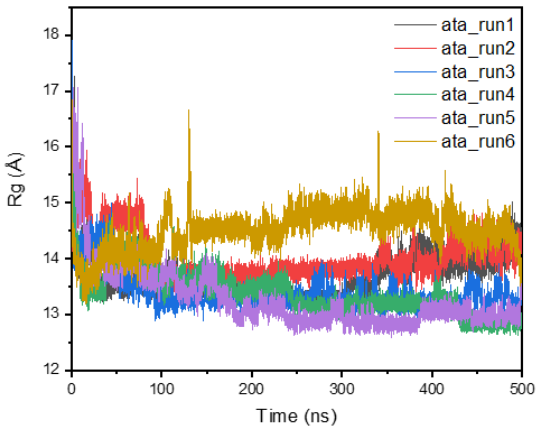


Figure 4. 7: Evolution of the radius of gyration versus time simulation (ns) for the 6-independent MDs (500-ns each) for atactic PVP91 model built in CHARMM/GUI services using OPLS-AA/TIP3P

Table 4. 3: Statistical analysis of R_g from averaging across all 6 independent 500 ns MDs for constructed atactic PVP91 model by CHARMM/GUI, using OPLS-AA/TIP3P.

Mean R_g (Å)	Max (Å)	Min (Å)	SD (Å)
13.7	15.1	13.0	0.3

Tacticity is one of the most essential characteristics of stereoregular polymers, influencing their morphology as well as their conformational, thermal, rheological and mechanical properties.[39–45] Additionally, tacticity can have an effect on the polymer water solubility as found in several studies.[46–48] As we are dealing with the polymer that has a chiral center in its monomer unit, the next examination was carrying out MDs for syndiotactic and isotactic structures, in addition to the previous atactic PVP91, to further investigate the behavior of the internal structure in the explicit water in different stereochemical configurations. The initial structures for different stereo configurations were obtained using MOE. The 300 ns MDs were performed for each model using OPLS-AA/TIP3P in GROMACS. The initial configurations used have a very similar R_g value ≈16.4 Å. Fig. 4.8 displays the results of the R_g evolution versus simulation time (ns) (upper left) and the statistical analysis of R_g Table. 4.4, radial distribution function g (r) versus distance (Å) (lower panel), and the calculated number of water molecules found within 4 Å (upper right) in all tested systems. The isotactic chain has higher R_g values during

the time simulation with the mean 17.9 Å and higher conformational range SD 1.1 Å than the systems of syndiotactic, which has mean 13.6 Å and SD 0.3 Å, and atactic, which has the mean 14.3 Å and SD 0.3 Å. This analysis was done after removing the initial 50 ns from each MD for equilibration purpose. In general, from the acquired single chain modeling data, the chain behavior appears to be similar in the atactic and syndiotactic systems and prefers the compacted configurations, whilst slightly higher R_g conformations are observed in the atactic structure. The isotactic chain seems to be more extended and flexible to adapt higher range of conformations than the two stereo-configurations.

Moreover, the g(r)-PVP-water starts to emerge at 2 Å and raises to the bulk water after 4 Å and have resembling appearance in all stereo structures. However, the intensity of the water accumulation is more noticeable in the isotactic-PVP (can be observed starting from 2.75 Å), which indicates the higher solvation prediction can be found around the structure. To confirm that, we calculated the number of water molecules found within 4 Å distance from PVP as a function of simulation time (ns) for each of them. The higher number of water molecules were in the isotactic, at a mean of 573, compared to the atactic, at a mean of 532, and the syndiotactic, 519, structures. These results correspond to the R_g seen for the isotactic, which allows the extended configurations to have a higher chance to interact with water resulting in larger solvation layer. The existence of consistent side groups orientation may help water molecules to interact with polymer chain more effectively to decrease the repulsion between the cyclic pyrrolidone groups that lay at the same side of the vinyl backbone in the isotactic PVP. While the random and alternating arrangements of the side groups in the atactic and syndiotactic configurations could be the reason for hindering the interactions with water molecules, leading to the lower R_g. These results are compatible with the modeling study for PVP 20-mers using AMBER FF and SPC/E model[49], which found the syndiotactic conformations are more rigid than isotactic conformations.

Based on the presented theoretical calculations of the R_g for PVP91 in OPLS-AA/TIP3P models, until now, it can be said the isotactic structure is the closer model that represents the experimental data based on the SANS value of 18.6 ± 2.5 Å, than syndiotactic or atactic configurations. It is possible that the purchased sample is dominated by atactic fractions, with a relatively high content of isotactic ones. Further independent MDs are needed to investigate the stability of MD for each stereo configuration as well as the experimental conformation of the tacticity using ^{13}C NMR.

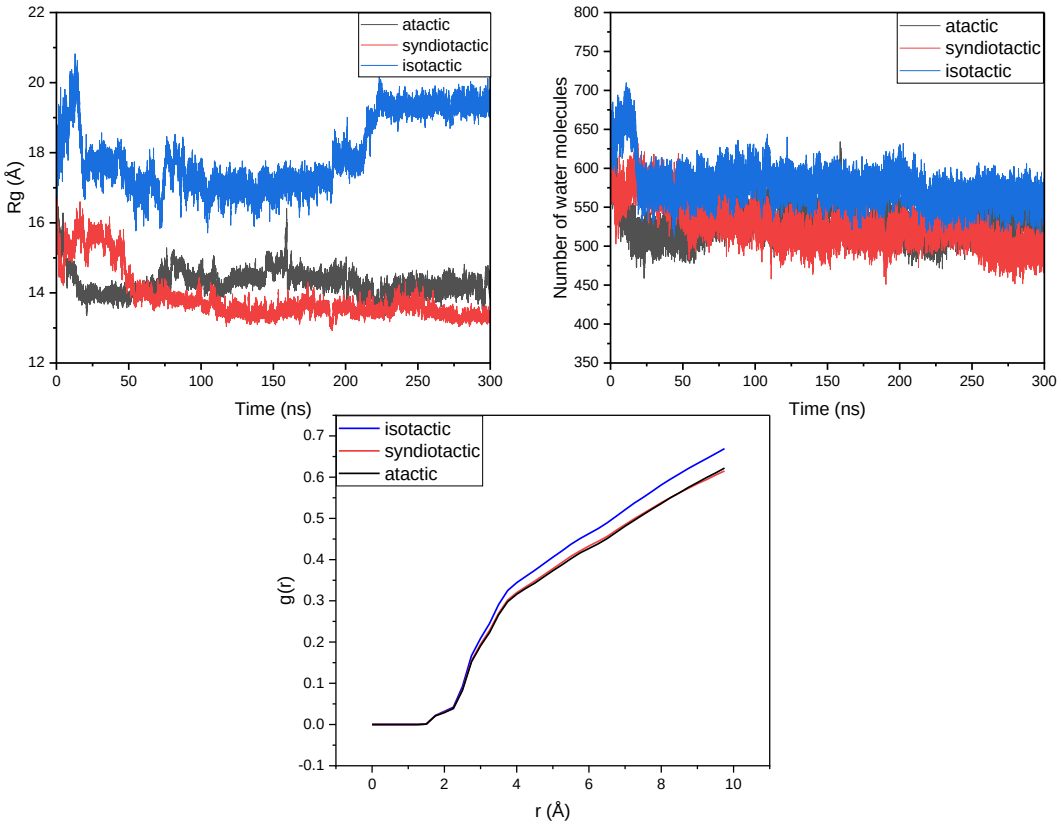


Figure 4. 8: The evolution of the radius of gyration versus simulation time (ns) (upper left), the representation of the number of water molecules versus time simulation (ns) within 4 Å (upper right), and the g (r) for PVP-water versus distance (Å) (lower panel) for atactic (black line), syndiotactic (red line), and isotactic (blue line) PVP91 in OPLS-AA/TIP3P.

Table 4. 4: Statistical analysis of R_g from 300 ns MD for PVP91 atactic, syndiotactic, and isotactic using OPLS-AA/TIP3P.

System	Mean R_g (Å)	Max (Å)	Min (Å)	SD (Å)
atactic	14.3	16.4	13.3	0.3
syndiotactic	13.6	14.8	12.9	0.3
isotactic	17.9	20.2	15.7	1.1

The various R_g results observed from the initial MD runs for the three stereo-structures (PVP91) led us to compare the predictions of the solvent exponent for iso and syndio-PVP,

additionally to the early-tested atactic, using OPLS-AA in TIP3P. Fig. 4.9 shows the theoretical representation of R_g versus the number of monomers along with the solvent exponent for each stereo-configuration. It is clear that the solvent value in the isotactic, 0.499, has a better value compared to the experimental reported value, 0.521, than syndio, 0.384, or atactic, 0.368. This means the theoretical (MD) prediction of the chain behavior in the solvent for iso-structure is close to the experimental value (0.521), followed by the syndio and atactic chains.

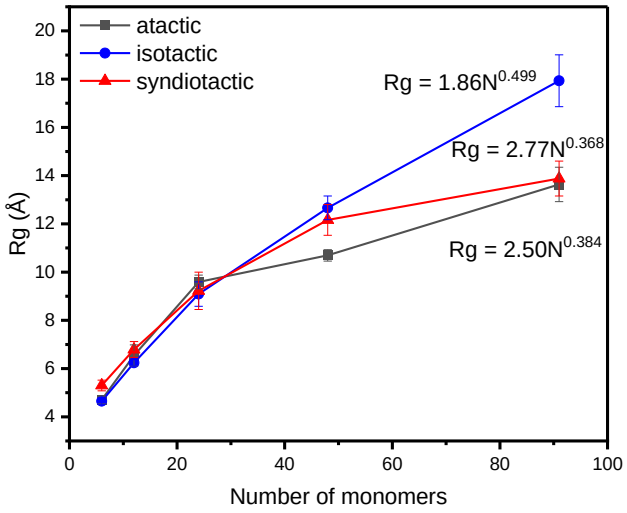


Figure 4. 9: Comparison of the R_g versus number of monomers between iso(blue), syndio (red), and atactic (black) PVP with production of the solvent exponent for each model in OPLS-AA/TIP3P.

4.4.3 R_g for isotactic, syndiotactic, and atactic-PVP91 in several independent MD runs

The dynamic simulations for isotactic, syndiotactic, and atactic PVP91 structures were carried out in six independent MDs (500 ns for each run) to obtain statistical sampling. Fig. 4.10 displays the R_g evolution results versus time for all 6 MD runs in each stereo structure, and Table. 4.5 shows the statistical analysis averaging across all trajectories.

It is obvious that the average R_g changes a lot in isotactic conformation. Run1 shows a slightly decreasing trend up until 200 ns (with an average of 18 Å), at which point it starts to grow and levels off at around 19 Å. Its fluctuations are around roughly 16 Å in run 6, but they are around a similar in value (14 Å) in runs 2, 3, 4, and 5. In MD, the production of slightly different paths with identical initial conditions can be seen due to the randomness of particle motion. The

reason for the different trajectories and behaviors of the PVP chain (even though they use the same type of stereo structure) can be due to using different initial velocities, which shows an impact on the R_g results. Running simulations with various random seeds can successfully explore different areas of the configurational structures, and thermodynamic averages of structural features can provide statistically meaningful results. Thus, the iso-PVP model seems to have a slightly different MD behavior than other stereochemistry.

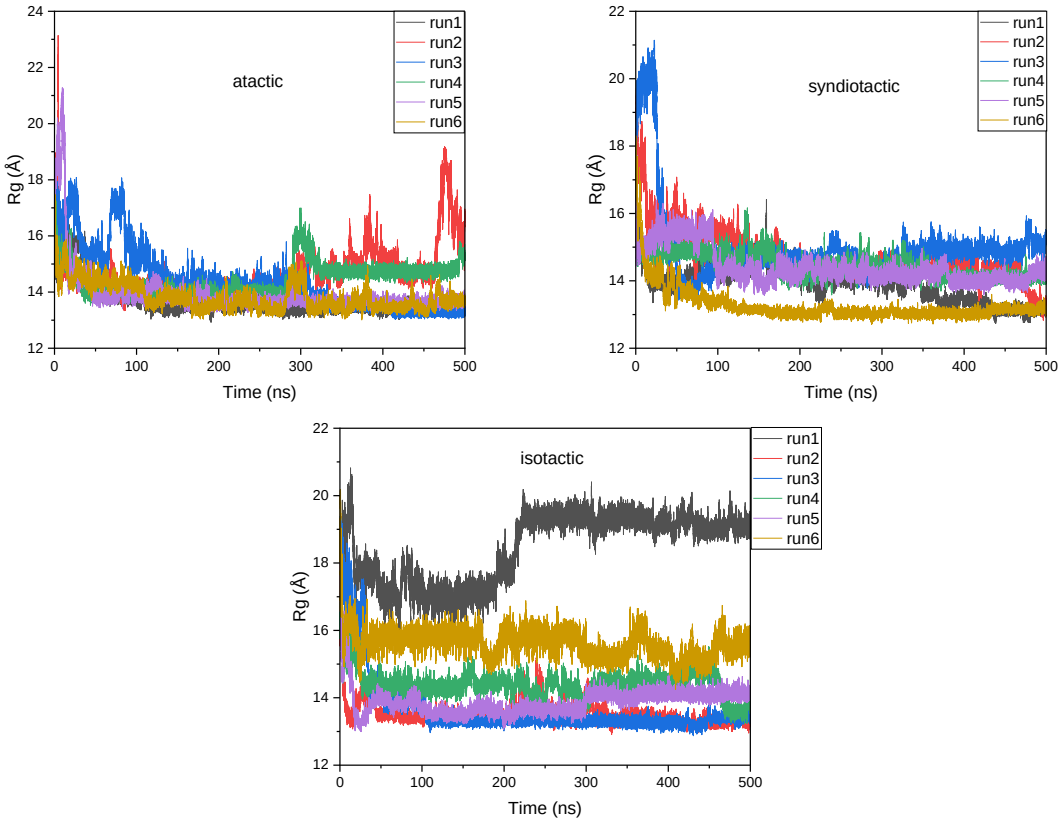


Figure 4. 10: The evolution R_g results versus time (ns) for atactic, syndiotactic, and isotactic PVP91 structure in 6 independent MD runs (500 ns each), run 1(black), 2 (red), 3 (blue), 4 (green), 5 (purple), and 6 (muddy yellow) in OPLS-AA/TIP3P.

Table 4. 5: Statistical analysis of R_g averaging overall 6- MDs for isotactic, syndiotactic, and atatctic-PVP91using OPLS-AA/TIP3P.

System	Mean R_g (Å)	Max (Å)	Min (Å)	SD (Å)
atactic	14.2	16.6	12.7	0.4
syndiotactic	14.0	19.2	12.9	0.5
isotactic	14.9	20.4	12.9	0.4

4.5 Computed SANS curves and Fitting with the experimental SANS data

To interpret the SANS data from MD trajectories that contain many ensembles, we calculated the scattering intensity for iso, syndio and atactic PVP using the sans tool in GROMACS, as shown in Fig. 4.11 (A). The experimental scattering intensity was fitted into the computed curves by multiplying the experimental I(q) by 2.14 to obtain the better-matched intensity. Obviously, there is disagreement between the experimental and calculated SANS curves for all stereoregular polymers. Examining the fitting data makes it evident that the experimental SANS curve actually indicates an extended conformation, while computed curves demonstrate the model structures are too compacted, more noticeable in the syndio and atactic curves (blue and green), which are shifting into high q (around 0.1 $Å^{-1}$) indicating smaller size. The iso curve is slightly shifting into the lower q resulting in a better fit around that area, as the iso MD shows somewhat open conformations compared to the other stereo MDs (see the section above). After the 0.1 q $Å^{-1}$ region, all stereo SANS curves display similarly, with a clear discrepancy to the experiment. These results are similar to what we have seen in the PEG fitting (see previous chapter), which indicates the polymer chain is not well solvated due to inadequacies in the applied force fields.

The predicted real space distribution function P(r), which was acquired alongside SANS curve calculations, was investigated and fitted into the experimental function, which was obtained from SANS inversion analysis in SasView (see Fig. 4.11 (B)). The Gaussian forms, which indicate the geometry and the relationship between the internal particles, show the deviated D_{max} for the syndio, 51 Å, and atactic, 47 Å, models from the measured D_{max}, 67.8 Å. However, the D_{max} for iso form, 65 Å , presents a close value to the experiment while losing the overall resemblance to the featured shape.

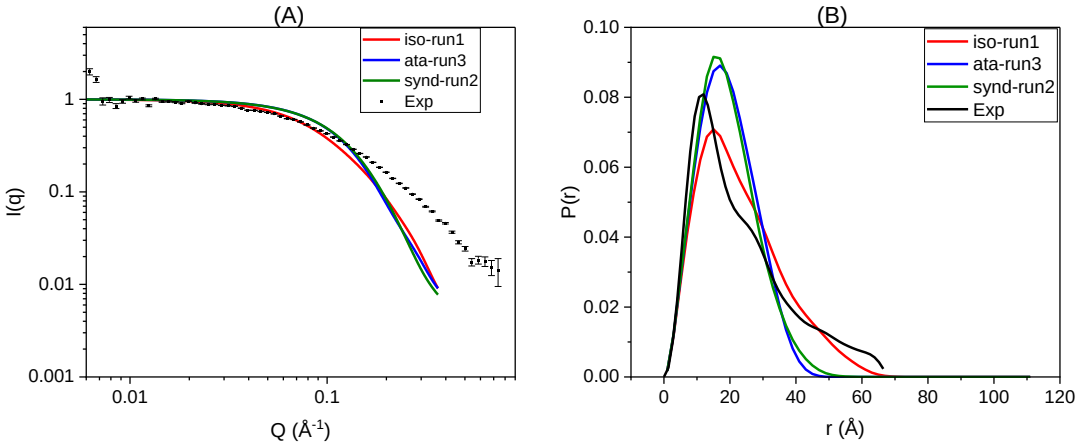


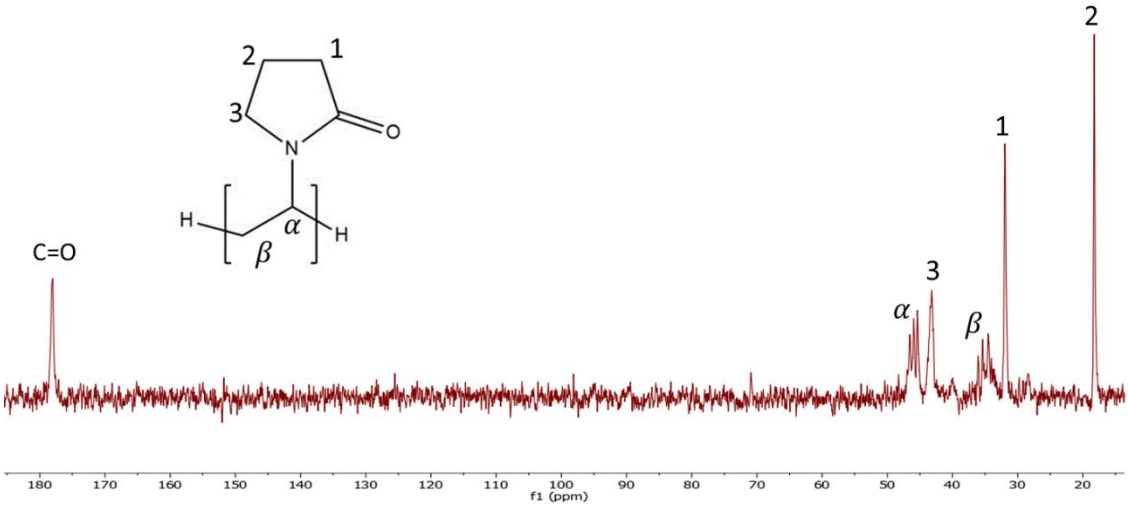
Figure 4. 11: (A) The experimental subtracted incoherent background h/D SANS data (dash black) fitted into the calculated MDs SANS curves for iso (red), syndio (green), and atactic (blue), (B) The experimental (h/D black) and calculated real space distribution function P(r) from MD trajectories for iso (red), syndio (green) and atactic (blue) PVP.

4.6 NMR

Different polymerization strategies can produce variations in the polymer tacticity.[10,50–53] For example, PVPs generated by thermally induced radical polymerization (FRP) are predominantly atactic,[50] and their tacticity differs depending on the polymerization solvent. The PVP's tacticity synthesized in aqueous solution differs from the one prepared in organic media.[10] Moreover, some increase in syndiotacticity and isotacticity can be noticed during vinyl pyrrolidone polymerization in the presence of fluoroalcohols/anionic surfactants and Lewis acid catalysts respectively, and these effects are mostly caused by electron pair donor-acceptor interactions between VP and the supplied additives.[54,55]

The proton and ^{13}C NMR are two different techniques that have been used in literature to study polymer tacticity. However, ^{13}C NMR is often preferred over proton NMR because it eliminates the complications of signals generated by proton-proton interaction. It allows for a more straightforward examination and offers a clearer understanding of the polymer's tacticity. Splitting or multiplicity (rather than a singlet) in the ^{13}C NMR spectrum indicates stereochemical anisotropy, which can be syndiotactic, isotactic, or atactic configurations.[56] The multiplicity in the spectrum appears if different carbon atoms of the same type in the polymer have distinct local chemical environments and are not magnetically equivalent.

The conventional method to corroborate the tacticity chemical shift assignments in an NMR spectrum is to connect the signal intensity to statistics or the probability of one repeat unit joining to the end of the polymer chain during the propagation step.[56] The ^{13}C NMR study of PVP in D_2O was used to analyze the PVP's tacticity. It should be noted that D_2O solvent was used since it was reported[50] that water was found to be the best solvent for the tacticity study, and at 85 °C was used since the signals were clearly separated (especially α-methine carbon). A full representative ^{13}C NMR spectrum for the purchased PVP is given in Fig. 4.12 (upper panel). The α-methine carbon region has three split resolved peaks at 46.5, 45.9 and 45.3 ppm, which indicates the polymer has an irregular micro-tactic structure, while the splitting signals for carbonyl and β-methylene are not well resolved. Thus, the α-methine is our focus here, and the expansion of the splitting methine peaks with the intensity fitting is shown in Fig.4.12 (lower panel). The analysis of these signals can provide information about the stereochemical configurations of the PVP chain. We reported the peak intensities according to the determination of triad sequences (mm, mr, rr)[50] in Table. 4.6. The conclusion from the derived results is that the utilized PVP in this study is atactic sample.



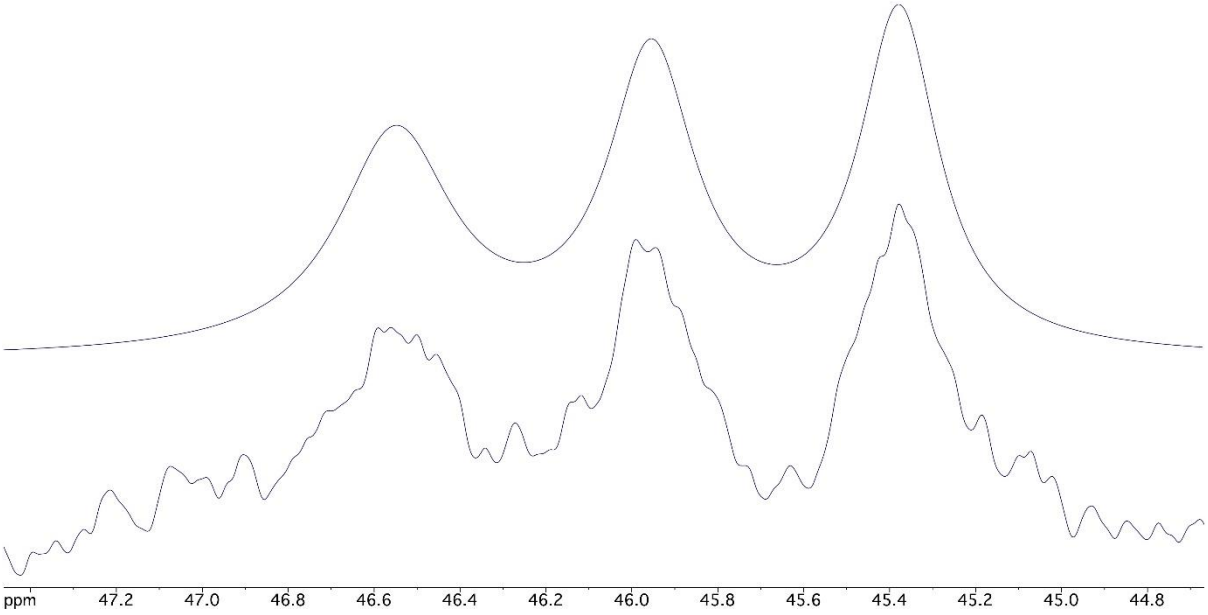


Figure 4. 12: Upper panel: ^{13}C NMR spectra for PVP in D_2O at 85 °C; Lower panel: expansion of the α-methine resonance region with the intensity fitting above. (lower panel picture obtained from Dr Ben Ward using iNMR).

Table 4. 6: Assignments and relative integration of intensities of polyvinylpyrrolidone (PVP) α-methine

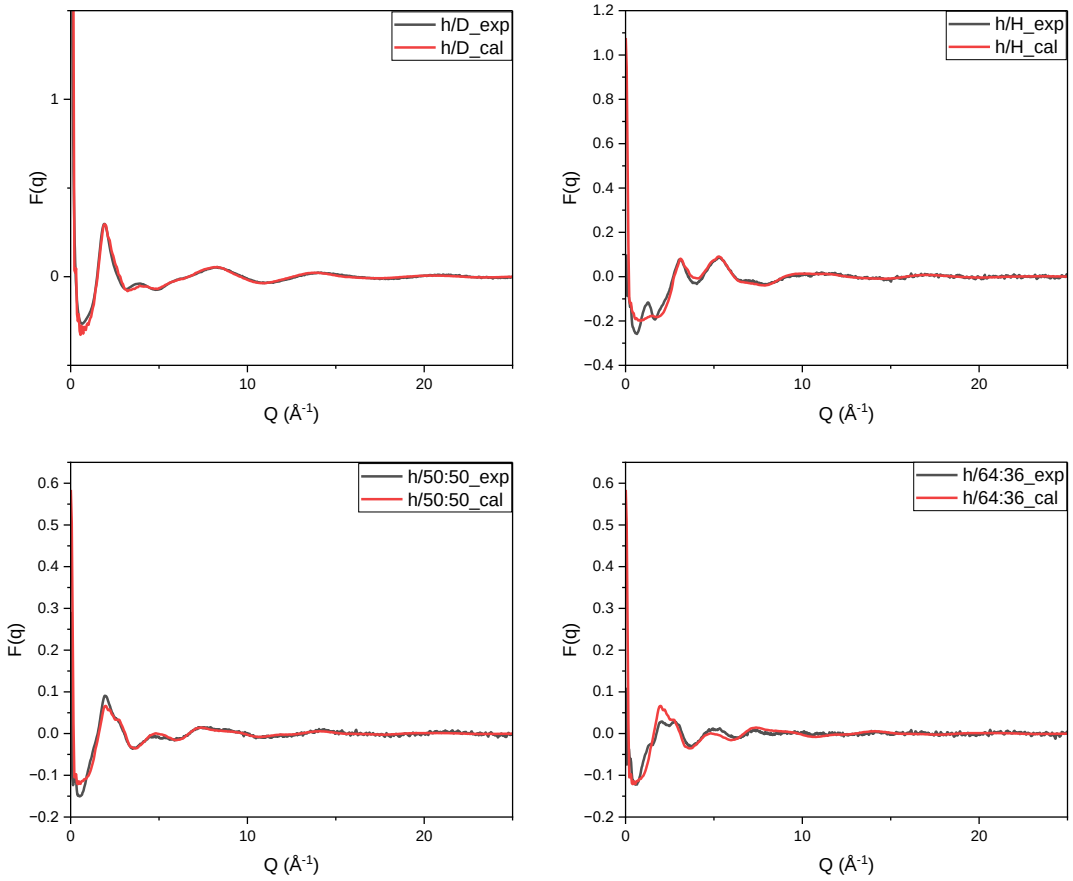
Chemical shift	Intensity-integration	Sequence Assignment[50]
46.5	33.23	mm
45.9	33.71	mr
45.3	35.19	rr

4.7 Neutron total scattering fitting (Dissolve)

To gain further insight into a full 3-dimensional understanding of PVP solutions, the software package Dissolve (with several developing versions from 0.9 and above), which implements the empirical potential structure refinement (EPSR) method as a refinement of classical MD and analysis tools for scattering data, was employed and fitted to the experimental data (NIMROD). Once the system has stabilized, the total structure factors F(q) for each isotopically substituted sample were predicted from weighted radial distribution functions G(r) obtained from the simulation over several iterations, as this technique has been successfully applied to diverse amorphous systems[57–60]. Due to the mismatch between experimental and simulated

data, an additional empirical potential was then calculated and added on to the reference potential to get the best possible fit between the simulation and experiment after repeating this process iteratively. The final comparison of the predicted to the experimental interference differential cross section F(Q) and weighted G(r) for the four datasets are shown in Fig. 4.13 (A) and (B), respectively. All samples show good agreement with the experimental F(q) data with a little discrepancy at low Q (< 2.5 $Å^{-1}$) for the hydrogenous samples (h/H, 50:50, 64:36) due to the subtraction of inelastic and incoherent scattering for these samples, which is known to be difficult to apply in an unambiguous manner[61]. Furthermore, the features of the experimental total G(r) data are well described by the fit on the radial distribution function for the simulated-box, especially for the h/D and h/H samples. However, the other two samples (50:50 and 64:36) show a discrepancy in intensity due to the inelastic corrections. Even though the EPSR module in Dissolve is capable of modifying the intermolecular potentials to better match the simulated-box scattering data (as seen from the fitting), the polymer conformational changes are omitted (see previous chapter).

(A)



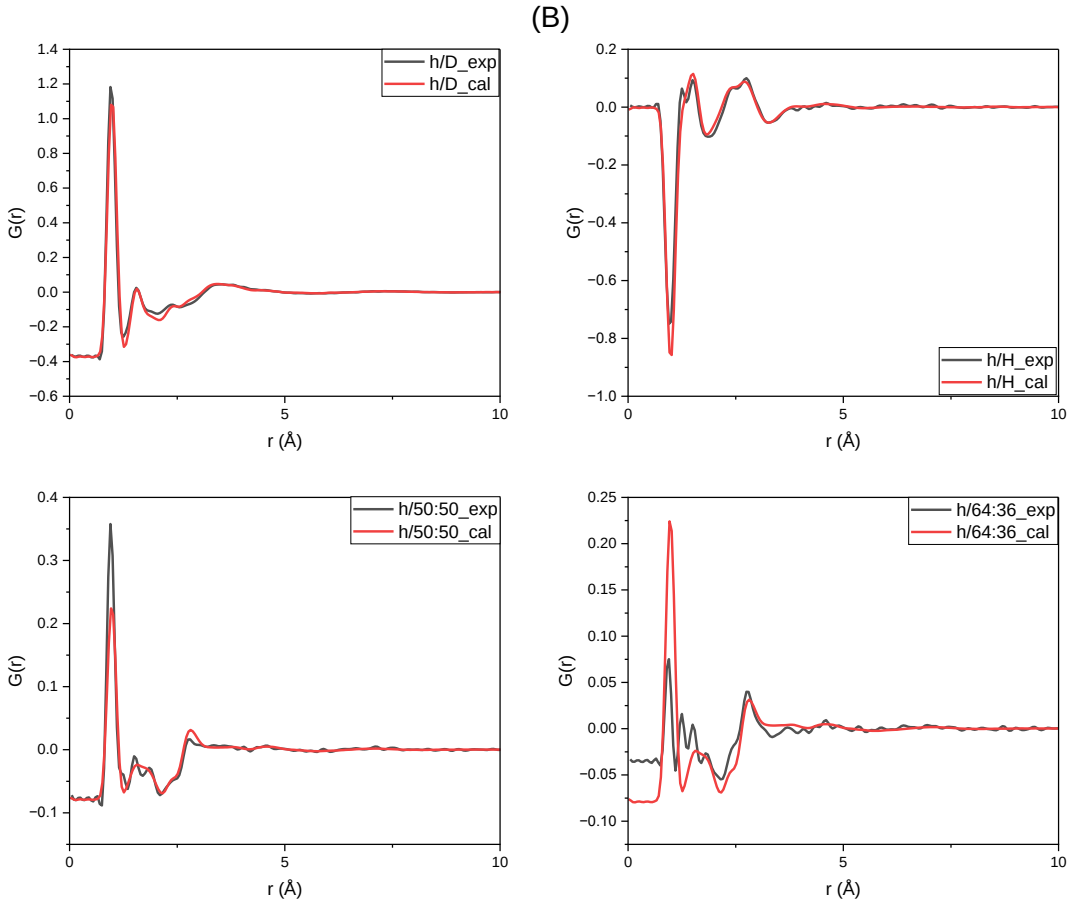


Figure 4. 13: Comparison of the experimental (A) total neutron weighted structure factor F(Q) (black line) and the equilibrated calculated (red line) Q range 0-25 $Å^{-1}$, (B) total weighted radial distribution function G(r) (black line) and the equilibrated calculated (red line) in Dissolve for PVP solutions d/H, h/H, h/H(50):D(50), and h/H(64):D(36) after applying the EPSR.

4.8 Prediction of the weighted-neutron partial pair correlations in Dissolve (F(q) and G(r)

It is now possible to evaluate the experimentally driven simulation (by the EPSR) in depth to understand the structure of PVP in solution. Our interpretation of the structural model (from Dissolve) is particularly interested in intermolecular interactions between polymer chains and solvents, and before that, we need first to show the prediction of total pair correlations for each component individually to test the capability of Dissolve to detect that. Figure 4.14 plots the total of all expected pair correlations F(q) separately for the PVP, solvent (SOL), and intermolecular PVP-SOL taken from one of the isotopic substituted samples (h/D), which

shows the better fit. The scaling of the total partial correlations for SOL, PVP, and PVP-SOL in Fig. 4.14 clearly shows that the main correlation features observed when dealing with PVP in diluted solutions are connected to the solvent-solvent corelations. On the other hand, PVP-SOL overlap correlations may occur in the same region. Evidently, two characteristics that are similar to the first two peaks in the SOL F(q) but have much lower intensity values can be detected in the computed total weighted-neutron F(q) for PVP-SOL, which indicates the Dissolve can capture these interactions. Moreover, it is logical to see that the PVP-SOL main correlations are located at very low q (0.5-6 $Å^{-1}$) which is consistent with the intermolecular interaction regions. PVP pair correlations are detected at both low (intramolecular interactions with different monomers) and high Q regions (intramolecular interactions within the same monomer) 1-16 $Å^{-1}$.

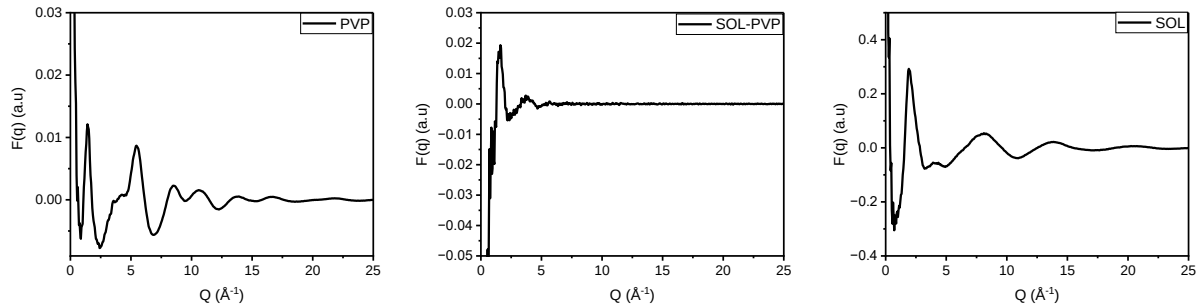
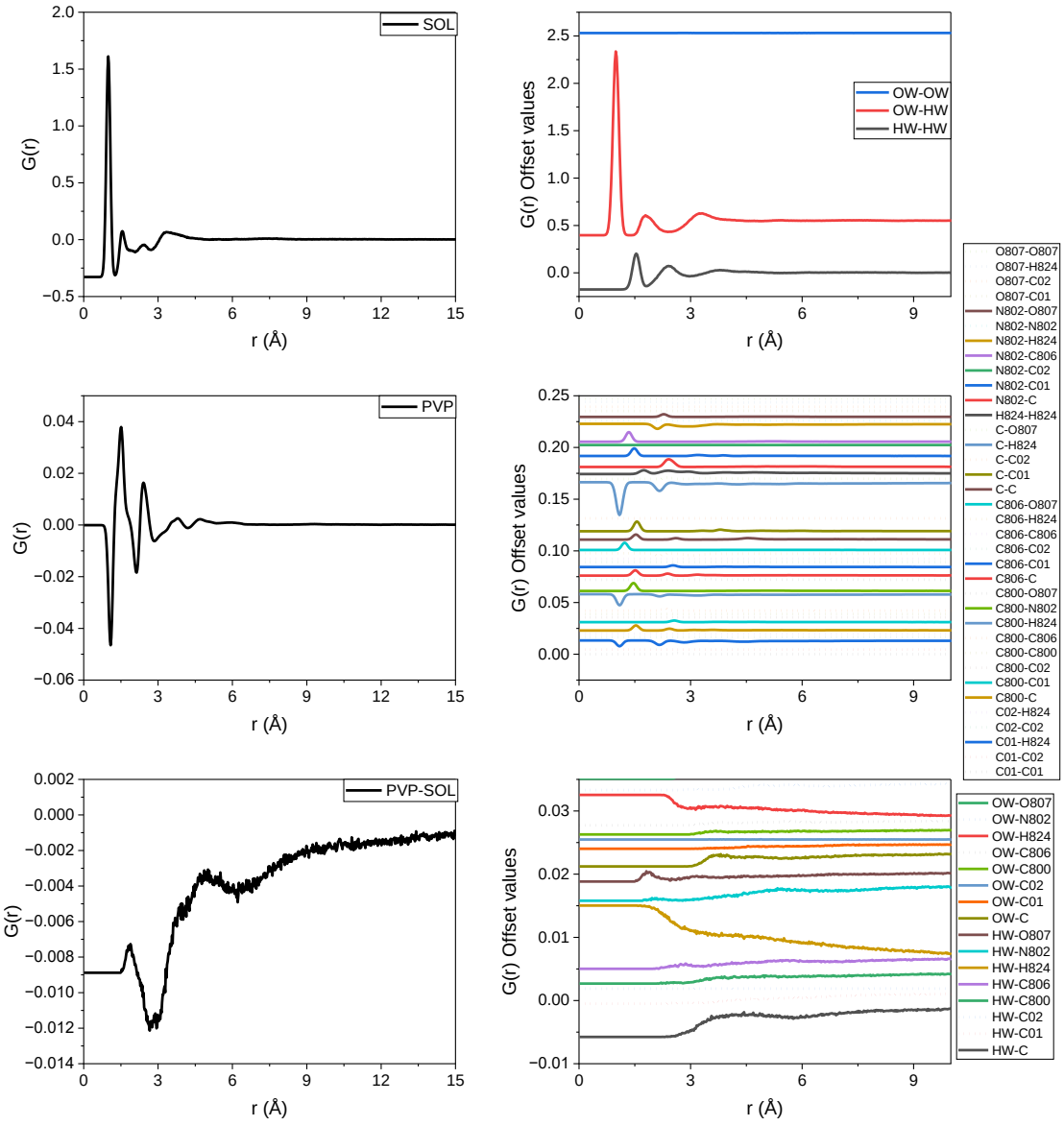


Figure 4. 14: The predicted sum of all partial correlations weighted-neutron structure factor F(q) separately for PVP, solvent(SOL), and intermolecular PVP-SOL for h/D sample.

The prediction of the all partial pairs weighted-neutron distribution functions G(r) for each component, SOL, PVP, and PVP-SOL, is shown in Fig. 4.15 (A) (the total pairs for each component are shown in the left-hand side and the individual pair correlation for each one is shown in the right-hand side, and the Fig. 4.15 (C) shows the names of the PVP-atoms represented). It is apparent from looking at Fig. 4.15 (A) that the primary abundant atom correlations are related to the solvent-solvent, as seen in F(q) above. The peaks at short distances starting from ≈ 1 Å and above are related to the intramolecular correlations in SOL (OW-HW, HW-HW) and PVP (C-H824, C-C, C-O807, N802-C, and N802-H824; note the C denotes any C atoms in the system. As mentioned above, the study of intermolecular interactions between solvent and polymer chains is of special relevance to our understanding of the structural model from Dissolve. Hence, as illustrated in Fig. 4.15 (B), the expected

hydrogen bonding, donor and acceptor, between OW-H824 and HW-O807 is the most significant association that can be examined. Obviously, the primary feature peaks lie between 2 and 3 Å, after which the features begin to diminish. The first peak at 2 Å represents the hydrogen bond interactions between HW and O807 with very small g(r) values. As the concentration (in form of atom fraction) for the carbonyl oxygen is fairly low, this is what gives the small numbers in the g(r). Furthermore, the second feature at 3Å relates to the interactions between OW and H824, which exhibits an opposing trend as a result of weighted H824 being determined from weighted F(q).

(A)



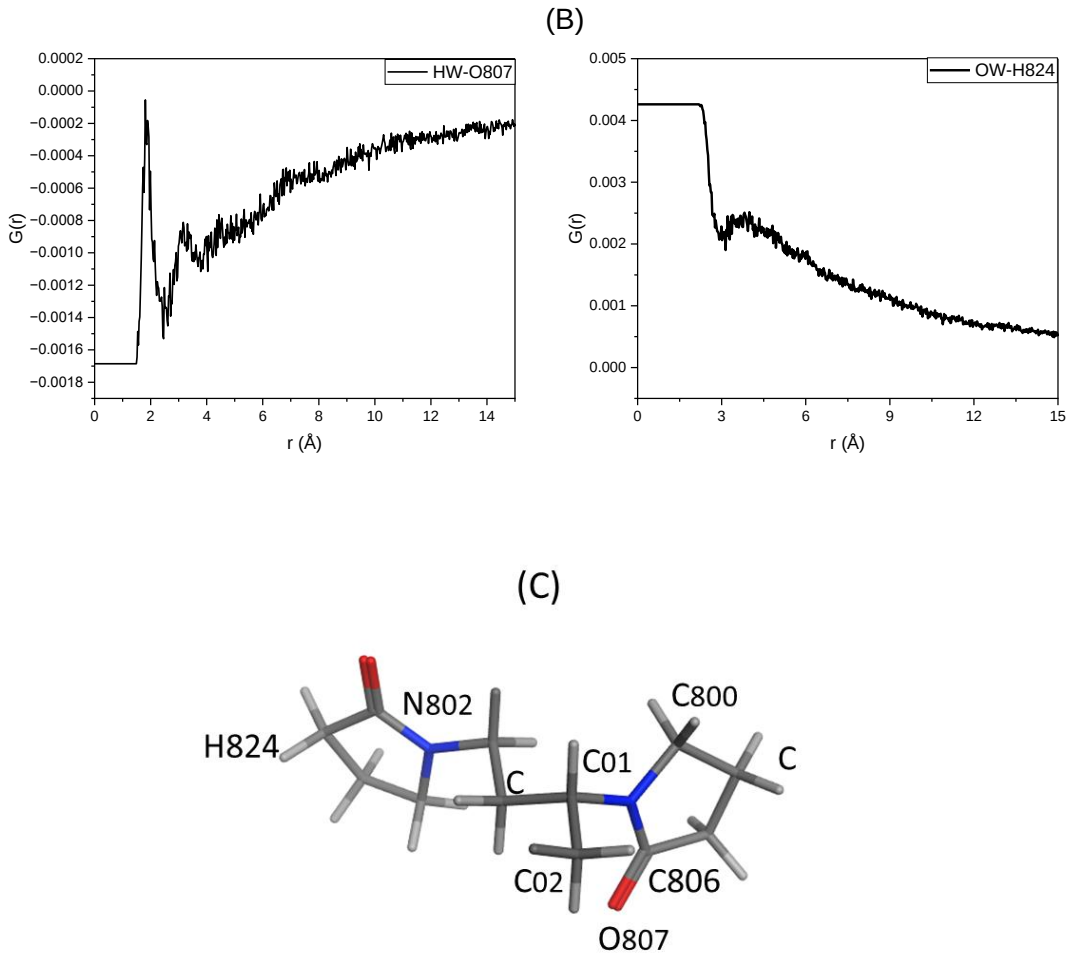


Figure 4. 15: (A) The predicted sum of all partial correlations weighted-neutron radial distribution functions [G(r) -1] separately for PVP, solvent(SOL), and PVP-SOL for 6 wt% h/D sample, (B) the weighted intermolecular pairs [G(r)-1] for HW-O807 and OW-H824 (C) the representation of atomic names used in this study.

4.9 Conclusion

The experimental SANS data were obtained for PVP, 10,062 g/mol, to help understanding the solute behavior and solute-solvent interactions. We examined the Flory approximation for atactic PVP using OPLS-AA FF in implicit (Born) and explicit (TIP3P) solvents (default box) and found that the implicit solvent showed a better value than the TIP3P. This led us to test the performance of the PVP91 in different water models, TIP3P, TIP4P, TIP5P and SPC (in the cubic unit cell that has a close concentration to the experimental data). Similar chain performance was noticed in all tested solvent models. The following attempt was to test the R_g

in the CHARMM36/TIP3P and found that the average R_g obtained across the trajectory was higher than the experimental SANS value. Using OPLS-AA/TIP3P, additional investigation was conducted to investigate the chain behavior for PVP91 in three distinct stereochemical configurations: atactic, isotactic, and syndiotactic (each with 300 ns MDs). The isotactic configuration displayed the better results in first MD run with: average R_g compared to the SANS value, conformational range throughout the MD, solvation layer, than other stereochemical structures. Additional 6- independent MDs were then carried out (500-ns each) to test the MD stability for each stereo model resulting in a slightly high average R_g across all 6-MDs in the iso model. The atactic PVP sample was verified by the ^{13}C NMR tacticity test. The SANS data fit into the coil model with higher quality than the cylinder model, despite the P(r) function indicating some rod shape signals. The PVP solution was studied using a combination of total neutron scatting experiments with isotopic substitution and molecular simulations via Dissolve. The hydrogen bond lengths, donor and accepter (HW-O, OW-H), are compatible with the SANS-driven MD results in the solvation layer thickness (see the next chapter).

References

1 P. Franco and I. De Marco, The Use of Poly(N-vinyl pyrrolidone) in The Delivery of Drugs: A review, *Polymers (Basel).*, 2020, **12**, 1114.

2 C. Bothiraja, M. B. Shinde, S. Rajalakshmi and A. P. Pawar, Evaluation of Molecular Pharmaceutical and in-Vivo Properties of Spray-Dried Isolated Andrographolide-PVP, *J. Pharm. Pharmacol.*, 2009, **61**, 1465–1472.

3 M. Rasekh, C. Karavasili, Y. L. Soong, N. Bouropoulos, M. Morris, D. Armitage, X. Li, D. G. Fatouros and Z. Ahmad, Electrospun PVP-indomethacin constituents for transdermal dressings and drug delivery devices, *Int. J. Pharm.*, 2014, **473**, 95–104.

4 X. Zheng, T. Zhang, X. Song, L. Zhang, C. Zhang, S. Jin, J. Xing and X. J. Liang, Structural Impact of Graft and Block Copolymers Based on Poly(N-vinylpyrrolidone) and Poly(2-Dimethylaminoethyl Methacrylate) in Gene Delivery, *J. Mater. Chem.*, 2015, **3**, 4027–4035.

5 M. Kurakula and G. S. N. K. Rao, Pharmaceutical assessment of polyvinylpyrrolidone (PVP): As excipient from conventional to controlled delivery systems with a spotlight on COVID-19 inhibition, *J. Drug Deliv. Sci. Technol.*, 2020, **60**, 102046.

6 T. Yu, Surfactant Assisted Dispersion of Single-Walled Carbon Nanotubes in Polyvinylpyrrolidone Solutions, *Colloid Surf. Sci.*, 2017, **2**, 96–106.

7 R. K. Heenan, J. Penfold and S. M. King, SANS at pulsed neutron sources: present and future prospects, *J. Appl. Crystallogr.*, 1997, **30**, 1140–1147.

8 D. T. Bowron, A. K. Soper, K. Jones, S. Ansell, S. Birch, J. Norris, L. Perrott, D. Riedel, N. J. Rhodes, S. R. Wakefield, A. Botti, M. A. Ricci, F. Grazzi and M. Zoppi, NIMROD: The Near and InterMediate Range Order Diffractometer of The ISIS Second Target Station, *Rev. Sci. Instrum.*, 2010, **81**, 033905.

9 Gudrun - Routines for reducing total scattering data, https://www.isis.stfc.ac.uk/Pages/Gudrun.aspx, (accessed 25 August 2021).

10 J. R. Ebdon, T. N. Huckerby and E. Senogles, The influence of polymerization conditions on the tacticity of poly(N-vinyl-2-pyrrolidone), *Polymer (Guildf).*, 1983, **24**, 339–343.

11 W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.

12 N. Foloppe and A. D. MacKerell, All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data, *J. Comput. Chem.*, 2000, **21**, 86–104.

13 D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. DeBolt, D. Ferguson, G. Seibel and P. Kollman, AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules, *Comput. Phys. Commun.*, 1995, **91**, 1–41.

14 Molecular Operating Environment (MOE), 2022.02 Chemical Computing Group ULC, 910-1010 Sherbrooke St. W., Montreal, QC H3A 2R7, 2023.

15 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindah, Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX*, 2015, **1**–**2**, 19–25.

16 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.*, 1983, **79**, 926–935.

17 W. L. Jorgensen, Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water, *J. Am. Chem. Soc.*, 1981, **103**, 335–340.

18 W. L. Jorgensen and J. D. Madura, Temperature and size dependence for Monte Carlo simulations of TIP4P water, *Mol. Phys.*, 1985, **56**, 1381–1392.

19 M. W. Mahoney and W. L. Jorgensen, A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions, *J. Chem. Phys.*, 2000, **112**, 8910–8922.

20 G. Bussi, D. Donadio and M. Parrinello, Canonical sampling through velocity rescaling, *J. Chem. Phys.*, 2007, **126**, 014101.

21 M. Parrinello and A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, *J. Appl. Phys.*, 1981, **52**, 7182–7190.

22 R. W. Hockney, the Potential Calculation and Some Applications, *Methods Comput. Phys.*, 1970, **9**, 135–211.

23 A. Allouche, Software News and Updates Gabedit — A Graphical User Interface for Computational Chemistry Softwares, *J. Comput. Chem.*, 2012, **32**, 174–182.

24 J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller and A. D. MacKerell, CHARMM36: An improved force field for folded and intrinsically disordered proteins, *Biophys. J.*, 2017, **112**, 175a-176a.

25 T. Youngs, Dissolve: Next Generation Software for The Interrogation of Total Scattering Data by Empirical Potential Generation, *Mol. Phys.*, 2019, **117**, 3464–3477.

26 L. S. Dodda, I. C. De Vaca, J. Tirado-rives and W. L. Jorgensen, LigParGen web server : an automatic OPLS-AA parameter generator for organic ligands, 2017, 1–6.

27 LigParGen, https://zarbi.chem.yale.edu/ligpargen/, (accessed 25 August 2021).

28 poly_gauss_coil, https://www.sasview.org/docs/user/models/poly_gauss_coil.html#poly-gauss-coil, (accessed 30 August 2021).

29 Sasview, https://www.sasview.org/, (accessed 31 August 2022).

30 L. Sapir, C. B. Stanley and D. Harries, Properties of polyvinylpyrrolidone in a deep eutectic solvent, *J. Phys. Chem. A*, 2016, **120**, 3253–3259.

31 A. Aschi, M. M. Jebari and A. Gharbi, Investigation of Poly (vinyl pyrrolidone) in methanol by dynamic light scattering and viscosity techniques, *e-Polymers*, 2007, **7**, 020.

32 D. I. Svergun and M. H. J. Koch, Small-angle scattering studies of biological macromolecules in solution, *Reports Prog. Phys.*, 2003, **66**, 1735–1782.

33 P. B. Moore, Small-angle scattering. Information content and error analysis, *J. Appl. Crystallogr.*, 1980, **13**, 168–175.

34 C. James, Cardiff University, 2011.

35 M. Guettari, R. Gomati and A. Gharbi, Determination of the flory exponent by study of steady shear viscosity, *J. Macromol. Sci. Part B Phys.*, 2012, **51**, 153–163.

36 Y. G. Grodaa, V. S. Vikhrenkoa, A. H. Poghosyanb, P. K. Hakobyanb, L. H. Arsenyanb and A. A. Shahinyanb, Conformation and diffusion properties of polyvinyl alcohol and polyvinylpyrrolidone molecules, *Electron. J. Nat. Sci.*, 2015, **375**, 60–67.

37 D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, Academic Press, 1996.

38 P. Florová, P. Sklenovský, P. Banáš and M. Otyepka, Explicit water models affect the specific solvation and dynamics of unfolded peptides while the conformational behavior and flexibility of folded peptides remain intact, *J. Chem. Theory Comput.*, 2010, **6**, 3569–3579.

39 P. N. Tzounis, D. V. Argyropoulou, S. D. Anogiannakis and D. N. Theodorou, Tacticity Effect on the Conformational Properties of Polypropylene and Poly(ethylene-propylene) Copolymers, *Macromolecules*, 2018, **51**, 6878–6891.

40 E. Von Meerwall, N. Waheed and W. L. Mattice, Effect of stereochemistry on diffusion of polypropylene melts: Comparison of simulation and experiment, *Macromolecules*, 2009, **42**, 8864–8869.

41 L. Chang and E. M. Woo, Tacticity effects on glass transition and phase behavior in binary blends of poly(methyl methacrylate)s of three different configurations, *Polym. Chem.*, 2010, **1**, 198–202.

42 K. Geng and O. K. C. Tsui, Effects of Polymer Tacticity and Molecular Weight on the Glass Transition Temperature of Poly(methyl methacrylate) Films on Silica, *Macromolecules*, 2016, **49**, 2671–2678.

43 S. Negash, Y. B. Tatek and M. Tsige, Effect of tacticity on the structure and glass transition temperature of polystyrene adsorbed onto solid surfaces, *J. Chem. Phys.*, 2018, **148**, 134705.

44 V. Arrighi, D. Batt-Coutrot, C. Zhang, M. T. F. Telling and A. Triolo, Effect of tacticity on the local dynamics of polypropylene melts, *J. Chem. Phys.*, 2003, **119**, 1271–1278.

45 S. M. Lippow, X. H. Qiu and M. D. Ediger, Effect of tacticity on the segmental dynamics of polypropylene melts investigated by 13C nuclear magnetic resonance, *J. Chem. Phys.*, 2001, **115**, 4961–4965.

46 E. Chiessi and G. Paradossi, Influence of Tacticity on Hydrophobicity of Poly(N-isopropylacrylamide): A Single Chain Molecular Dynamics Simulation Study, *J. Phys. Chem. B*, 2016, **120**, 3765–3776.

47 B. Ray, Y. Okamoto, M. Kamigaito, M. Sawamoto, K. I. Seno, S. Kanaoka and S. Aoshima, Effect of tacticity of poly(N-isopropylacrylamide) on the phase separation temperature of its aqueous solutions, *Polym. J.*, 2005, **37**, 234–237.

48 Y. Katsumoto, N. Kubosaki and T. Miyata, Poly (N -isopropylacrylamide): Solubility of Dimer Model Compounds in Water, *J. Phys. Chem. B*, 2010, **114**, 13312–13318.

49 T. Flebbe, R. Hentschke, E. Hädicke and C. Schade, Modeling of PolyvinylPyrrolidone and Polyvinylimidazole in Aqueous Solution, *Macromol. theory simulations*, 1998, **7**, 567–577.

50 H. N. Cheng, T. E. Smith and D. M. Vitus, Tacticity of poly(N-vinyl pyrrolidone), *J. Polym. Sci. Polym. Lett. Ed.*, 1981, **19**, 29–31.

51 K. Dutta and A. S. Brar, Poly(vinylpyrrolidone): configurational assignments by one- and two-dimensional NMR spectroscopy, *J. Polym. Sci. Part A Polym. Chem.*, 1999, **37**, 3922–3928.

52 V. I. Dubrovin and V. P. Panov, Carbon-13 NMR investigation of microtactic structure of polyvinylpyrrolidone, *J. Appl. Spectrosc.*, 1980, **33**, 754–757.

53 D. Heczko, B. Hachuła, P. Maksym, K. Kamiński, A. Zięba, L. Orszulak, M. Paluch and E. Kamińska, The Effect of Various Poly (N-vinylpyrrolidone) (PVP) Polymers on the Crystallization of Flutamide, *Pharmaceuticals*, 2022, **15**, 971.

54 D. Wan, K. Satoh, M. Kamigaito and Y. Okamoto, Xanthate-mediated radical polymerization of N-vinylpyrrolidone in fluoroalcohols for simultaneous control of molecular weight and tacticity, *Macromolecules*, 2005, **38**, 10397–10405.

55 K. Satoh and M. Kamigaito, Stereospecific Living Radical Polymerization: Dual Control of Chain Length and Tacticity for Precision Polymer Synthesis, *Chem. Rev.*, 2009, **109**, 5120–5156.

56 R. B. Clark and R. L. McMullen, in *Handbook of Pyrrolidone and Caprolactam Based Materials: Synthesis, Characterization and Industrial Applications*, 2021, vol. 6, pp. 1469–1534.

57 T. F. Headen, C. A. Howard, N. T. Skipper, M. A. Wilkinson, D. T. Bowron and A. K. Soper, Structure of π - π Interactions in Aromatic Liquids, *J. Am. Chem. Soc.*, 2010, **132**, 5735–5742.

58 T. F. Headen, Temperature dependent structural changes in liquid benzene studied using neutron diffraction neutron diffraction, *Mol. Phys.*, 2019, **117**, 3329–3336.

59 T. F. Headen, P. L. Cullen, R. Patel, A. Taylor and N. T. Skipper, The structures of liquid pyridine and naphthalene: the effects of heteroatoms and core size on aromatic interactions, *Phys. Chem. Chem. Phys.*, 2018, **20**, 2704–2715.

60 T. F. Headen, D. Mino, G. A. Youngs and A. J. Clancy, The Structure of Liquid Thiophene from Total Neutron Scattering, *Phys. Chem. Chem. Phys*, 2023, **25**, 25157–25165.

61 D. T. Bowron and S. Moreno, The structure of a concentrated aqueous solution of tertiary butanol : Water pockets and resulting perturbations, *J. Chem. Phys.*, 2002, **117**, 3753–3762.

Chapter 5

SANS-Driven All-Atom Molecular Dynamic (MD) Simulations for Polyvinylpyrrolidone (PVP) and Polyethylene Glycol (PEG) in Solutions

5.1 Introduction

Water molecules are essential when it comes to protein folding, molecular recognition, enzyme activity, and proton transport.[1,2] The hydration shell that envelops molecules in solutions has a different structure from the bulk solvent. The hydration layer can influence a solute's conformation and size, which can, in turn, affect its radius of gyration.[3–8] The hydration shell is widely recognized to be important for many biological processes, and the ways in which the hydration shell is regulated by molecular shape have been studied for many biomolecules, especially proteins. Molecular dynamics (MD) simulations with explicit solvent provide insight into molecule hydration in many protein studies[9,10] However, the interpretation of the solution scattering data by the MD simulation, especially scattering contributions from the hydration layer and how the solvation layer can influence the properties of polymer, is still lacking.[11,12] The dynamics of the water molecules surrounding the polymer should also be taken into account while validating the all-atom MD in order to acquire accurate characterization of the physical properties of the polymers.

The accuracy of MD simulations can be validated based on a suitable experimental methods, in order to achieve realistic models. One of the most well-known characterization techniques for soft matter studies, in a variety of materials, is small-angle neutron scattering (SANS).[13] A combination of SANS and MD simulation methods can provide a comprehensive understanding of polymer solutions.

Here, we examined the precision of all-atom MD simulations based on the experimental SANS data for polyvinyl pyrrolidone (PVP) and polyethylene glycol (PEG) in dilute aqueous solution. These were thus used to predict the SANS curves from MD simulations using a method that has been recently demonstrated successfully to provide a protein-specific footprint for several structural proteins with their solvation layer, to validate the MD against SANS data. The calculations use precise physical models for the hydration layer and the excluded solvent because all SANS predictions were based on explicit-solvent models.[14,15] This eliminates the

need for any solvent-related fitting parameters and allows for highly predictive structural modelling. However, the explicit-solvent SANS pattern predictions are non-trivial, because the patterns are affected by the force field imperfection, which includes the intermolecular interactions between the polymer chain and solvent. The general force fields that exist could produce physical models that do not precisely match the experimental measurements. To overcome this obstacle, the SANS-driven MD method was applied by incorporating the SANS data as an energy constraint into our MD simulations. Thus, we were able to improve our MD for PVP solution in comparison to scattering data and estimate the thickness of the solvation shell as well as the number of water molecules that tightly interact with the polymer chain and contribute to the SANS data.

5.2 Method

5.2.1 MD simulations and SANS curve predictions

Two separate MD simulations, pure solvent (TIP3P) and polymer solutions (OPLS-AA/TIP3P), were conducted in cubic boxes with dimensions (200 × 200 × 200 Å). The pure solvent is needed here because the electron density of the solvated polymer and the density of water are compared in order to anticipate SANS curves from an MD simulation. The polymer chain was placed with a sufficient distance from the box edge (20 Å) avoiding periodic boundary effects ($\approx\ 0.2\ wt\%$). The big boxes are required to avoid any systematic errors, with regards to the chain extension during the MD or/and not fitting the envelope into the unit cell while computing the SANS curves. The MD was done after the minimization and two equilibration processes, NVT and NPT. The experimental SANS curve of PVP (Mw$\approx$ 10,000 g/mol) and PEG (MW $\approx$ 13,000 g/mol)in D_2O was measured at the ISIS Neutron Facility, RAL, STFC, UK, using the equipment SANS2D. The sample preparation and simulation system were explained in detail in the previous chapters.

For the computed SANS curves, the envelope was constructed around the polymer conformations (from multiple simulation frames) at a distance of d = ranging from 1−6 Å (individual MD test) from the polymer atoms (see Fig. 5.1). The envelope contains the solvation layer and every conformational state of the polymer. A spatial envelope determines the thickness of the hydration layer that is considered in the SANS computation. The excluded solvent avoids any solvent-related fitting parameters during the SANS curve calculations and removes the unwanted water molecules (Fig. 5.1(C)). The force field fully defines the hydration

layer structure, so a free fitting parameter is not needed. The process of solvent subtraction involves calculating the difference in scattering intensity between the polymer with the envelope-enclosed solvent and the pure-solvent simulation, which is the identical volume obtained using the same envelop.[16] SANS curves were computed using the explicit-solvent SAXS/SANS calculations, according to the previously reported method,[17,18] [11,16] based on 1735 snapshots of the MD simulations.

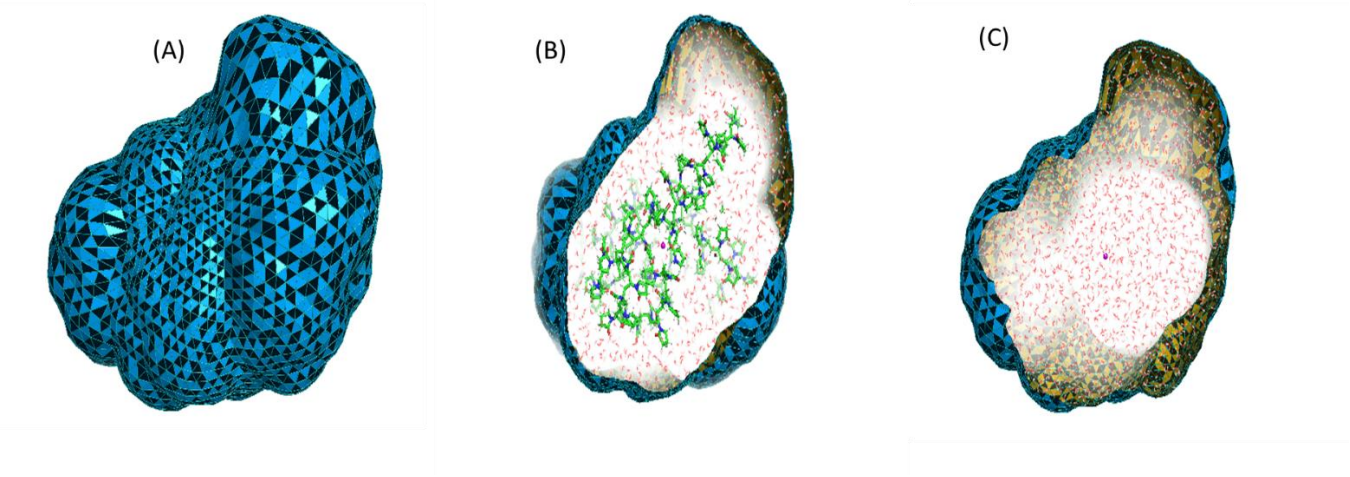


Figure 5. 1: (A) The polymer's spatial envelope (PVP91), which divides the solvation layer and the polymer chain from the bulk water described by multiple frames. (B) The polymer chain with the solvation layer, and the envelope at a distance of 6 Å from the polymer atoms. (C) example snapshot of the excluded solvent droplet within the envelope, used to compute the SANS curves.

5.2.2 SANS-driven MD simulations

MD simulations, driven by the SANS data, were performed to calculate the SANS curves from atomistic explicit-solvent MD simulations, using an extension of the GROMACS modified for SAXS/SANS calculations (same method as before, but with the experimental SANS data used as an MD guide, and the authors' website (http://cmb.bio.uni-goettingen.de/) provides access to the updated GROMACS source code used for SAXS/SANS predictions and SAXS/SANS-driven MD simulations). The SANS curves were computed on the fly during the SANS-driven simulations, and the polymer conformations were adjusted so that the computed curve more closely matches the experimental curve. Multiple SANS-driven MDs, with varying force constant values (1–10), were performed in order to find a value that works well for our systems and produces reliable conformational changes. To generate these calculations, the same spatial envelope and pure-solvent MD were used as before. The SANS curves were computed from

an average of 4,000, 12,000, 20,000 and 28,000 spectrum calculation frames for the force magnitude examination, and for the two extended MDs, 5 and 7 ns, with a force constant of 5 for PVP system. More details on how to compute the SANS curves from a given MD simulation with a template parameter file can be found in the ref[19].

One 3 ns SANS-driven MD simulations of PVP solution, on a modern compute node using both GPU and CPU, took 3 days due to the box size needed.

5.3 Result and Discussion

5.3.1 Fitting the experimental SANS data into the predicted SANS curves from free MD simulations (PVP system)

To validate the PVP solution ensembles against SANS data, first, we carried out MD using OPLS-AA/TIP3P in a box of 200 Å^3. The average R_g ($\approx 14 \pm 0.8$ Å) obtained from the output trajectories is similar to our previous investigation (see previous chapter), which is too small compared to the experimental value (18.6 ± 2.5 Å). It is known that the solvation shell influences the conformations and the R_g.[20,21] Explicit-solvent SANS calculations from all-atom MD simulations, provide a more accurate prediction of SANS scattering curves, because they can accommodate any solvent-related fitting parameter (see Fig. 5.2 (B)). To examine the incompatibility of the experimental and calculated R_g value, the SANS curve was calculated from free MD (1735 frames that did not restrict to the experimental data yet) and fitted the intensity of the experimental curve into the computed curve for the comparison as shown in Fig. 5.2(A). The experimental SANS curve must be scaled to the computed curve because experimental SANS data are normalised to the intensity per solute in cm^{-1}, while the computed curves are scaled to the electron squared (e^2) (the experimental intensity curve was multiplied by a factor of 1.1e+8 that match the intensity of the calculated curve). The intensity comparison between calculated and computed curves is not important here, as we are not comparing different contrast samples, instead resolving the size of the polymer. The experimental curve and the curve calculated from the model structure (using free MD) deviate significantly, as Fig. 5.2 (A) makes clear. Additionally, there is a difference in the R_g (14.8 against 18.6 Å) between the experimental and predicted curves, as determined by the Guinier analysis. This shows that the experimental SANS curve represents an expanded, more open state, and that the model structure is too confined. The discrepancy between the calculated and experimental curves, due to the imperfections of the applied OPLS-AA force-field and the insufficient intermolecular

interactions between the solvent and the polymer chain, resulting in compacted conformations. Therefore, to identify MD structures that are compatible with the SANS curve, we perform a SANS-driven MD (or SANS refinement) simulation in the next step.

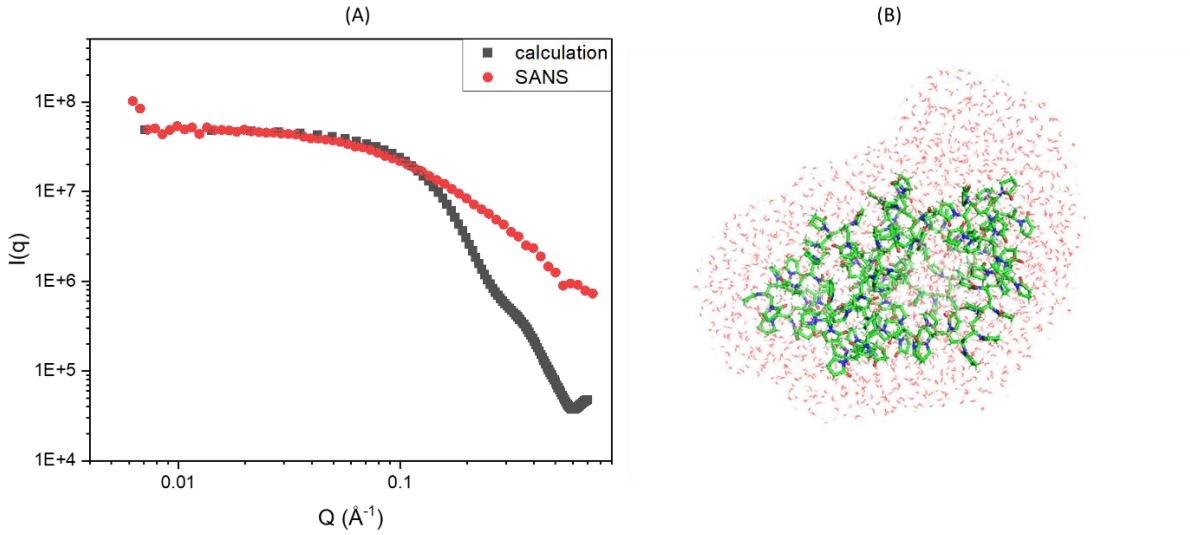


Figure 5. 2: (A) Comparison between SANS curve calculated from free MD and experimental SANS curve for PVP solution (B) the representation of the polymer chain (green) with its solvation layer (red and white).

5.3.2.1 SANS-driven MD simulations (PVP system)

Next, we used the SAXS/SANS-driven MD simulation method, in order to overcome force-field inaccuracies. By combining real SANS data with all-atom MD simulations, we were able to develop atomic models of PVP solution by incorporating the SANS data as an energetic restraint into MD simulations. This allows the refinement of MD against scattering data, and improved the MD ensembles.

Multiple calculations of SANS-driven MDs with different force constants were attempted to explore appropriate values that provide a reasonable Guinier fit with the SANS data. Force constant, which is the magnitude of force used to fit the MDs to the SANS data, ranging 1-10 kJ mol^{-1} nm^{-2}, were performed for 3 ns. Fig. 5.3 demonstrates the comparison between the calculated and experimental curves with different force constant values. It is apparent that all computed curves, from the different force constant values, agree well with the experimental curve at high q, 2.66-4.94 nm^{-1}. However, the discrepancy becomes apparent at low q values, specifically between 0.23 and 2.50 nm^{-1}, with varying degrees of inconsistency in each

individual fit. Amongst all attempts, using the value 5 kJ mol^{-1} nm^{-2} corresponding with the Guinier fit, 19.5 Å (see Table. 5.1), the nearest value to the SANS data acquired from all other SANS-driven MDs, resulted in a superior match at both low and high q combined. MDs with 1 and 10 kJ mol^{-1} nm^{-2} resulted in too small, 16.3 Å, and too high, 25.8 Å, Guinier fit respectively. Interestingly, despite the high R_g value in comparison to the actual data, the computed 10 kJ mol^{-1} nm^{-2} SANS curve shows a good fit to the experimental curve at both low and high q. Generally, the estimated R_g does not exhibit a progressive increase as the force constant value increases, as observed in Table 5.1.

This is clear evidence that the solvation layer affects the size and structure of the polymer when examining the radius of gyration for the molecule alone and for the Guinier fit, which counts the contributions of the solvation layer into the R_g calculations (see Table 5.1). The predicting value of how the solvation layer can affect the molecules' R_g, according to the Guinier fit, varies and depends on their shapes, sizes, and the used force constant. For example, the Guinier fit is about 8 Å away from the molecular radius of gyration when using the force constant 10 kJ mol-1 nm^{-2}, while the Guinier fit is about 2 and 4 Å when using the force constant 1 kJ mol^{-1} nm^{-2} and 5 kJ mol^{-1} nm-2, respectively. Moreover, the number of water molecules that contributes to the scattering pattern can be predicted from this calculation. The average number of water molecules within the solvation layer is about 1,800 in all presented MDs, which may contribute to the calculation of the polymer R_g (see Table. 5.1).

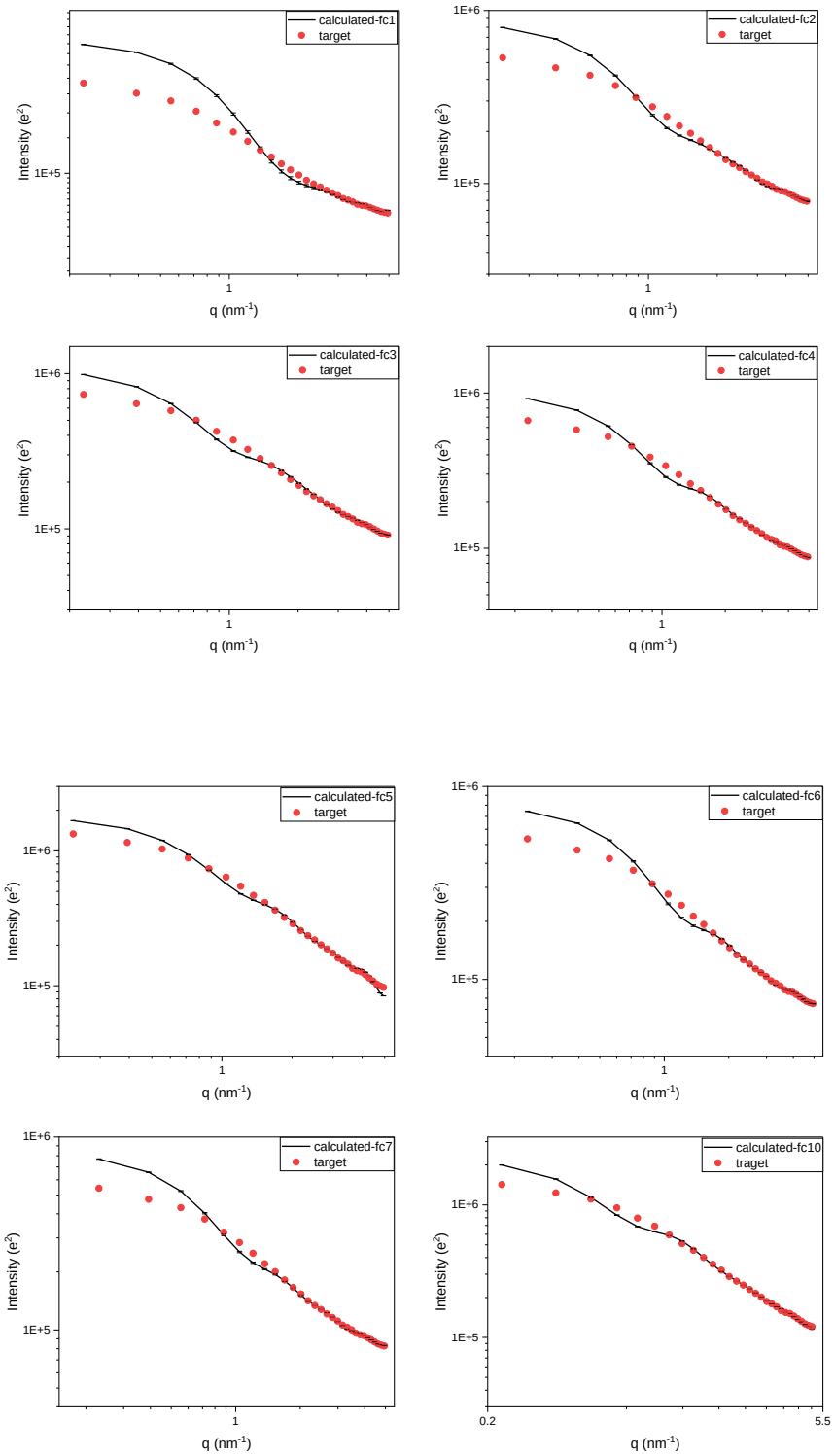


Figure 5. 3: Comparison between calculated (solid black) and experimental SANS (red points) curves for PVP solution with different force constant values (1-10 kJ mol^{-1} nm^{-2}) for 3 ns MD simulations.

Table 5. 1: Statistical analysis of the SANS-driven MDs, solute-R_g, Guinier fit, and number of water molecules in the solvation shell, for PVP solution in different force constants (kJ mol^{-1} nm^{-2}).

Force const (kJ mol^{-1} nm^{-2})	1	2	3	4	5	6	7	10
Mean R_g (solute only, electron-weighted) Å	14.6	15.2	16.5	16.1	15.4	15.7	15.5	17.1
Standard deviation Å	0.3	0.9	1.4	1.2	0.9	1.5	1.0	2.6
R_g (Guinier fit) Å	16.3	20.6	21.8	21.4	19.5	19.8	20.8	25.8
Mean number of waters in solvation shell	1759	1770	1779	1779	1788	1766	1771	1796

In an effort to obtain better fits, more SANS-driven MDs with a force constant of 5 kJ mol^{-1} nm^{-2} were run for longer simulation times – 5 and 7 ns. The intensity fit curves for the 5 and 7 ns MDs are displayed in Fig. 5.4, and the estimated water molecules in the solvation shell, along with the statistical analysis of the radius of gyration, are included in Table 5.2. When the MD is extended for 5 ns, a better Guinier value of 18.8 Å is produced, which is extremely close to the experimental value of 18.6 Å. Still, there is a discrepancy between the computed and calculated patterns at low q, up to 1.69 nm^{-1}. On the other hand, the 7 ns MD Guinier fit results in a high number, 21.1 Å, and high deviation at low q, up to 0.88 nm^{-1}. Both MDs have around 1800 water molecules in the solavtion layer, which is the same amount as in all previous MDs. The MD ensembles for PVP solution were effectively improved to achieve a very good radius of gyration, in comparison to the SANS data. The polydispersity of the real sample made the fitting procedure more difficult, as evidenced by the discrepancy in the intensity fit between the calculated and experimental curves at very low q.

Furthermore, the thickness of the solvation shell has been determined with SANS-driven MDs. The study of solvated layer thickness, helps to understand the effects of the nature and strength of interactions between the polymer and the solvent, which reflects the solubility and stability of the polymer in the solution. Thicker solvation layer may indicate strong polymer-solvent interactions and vice versa. About 90% of the solvated frames, in the 5 and 7 ns SANS-driven MDs, are found to be less than 3 Å solvation thickness. This thickness is propably a reasonable magnitude for a such PVP polymer, which is a neutral molecule and has a hydrophobic backbone structure. Thus, the main contribution to the PVP-water interactions is the hydrogen bonding between the pyrrolidone ring side chain, carbonyl C=O, and water hydrogens.

Additionaly, the diameter of the bounding envelope was estimated for the PVP solution from the latter SANS-driven MDs. This can gives a measure and be indicative of the overall

conformational size for polymer solution. The calculated D_{max}, 66.5 Å, is in accordance with the value seen from the experimental D_{max}, 67.8 Å, in the P(r) function, obtained from the inversion approach for scattering SANS curve. This indicates that a valid MD trajectory was successfully achieved, after restraining the MDs to the SANS data, which allowed for refinement of the MDs.

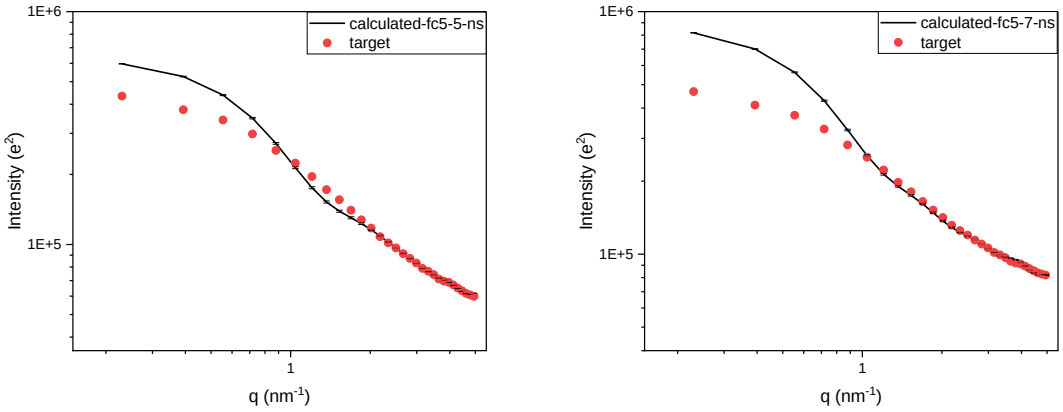


Figure 5. 4: Comparison between calculated (solid black) and experimental SANS (red points) curves for PVP solution with force constant 5 kJ mol^{-1} nm^{-2} for 5 and 7 ns MD simulations.

Table 5. 2: Statistical analysis of the SANS-driven MDs, solute-R_g, Guinier fit, and number of water molecules in the solvation shell, for PVP solution with a force constant of 5 kJ mol^{-1} nm^{-2}, in 5 and 7 ns MD simulations.

Force constant (5 kJ mol^{-1} nm^{-2})	5 ns	7 ns
Mean R_g (solute only, electron-weighted) Å	16.1	17.0
Standard deviation Å	0.8	1.1
R_g (Guinier fit) Å	18.8	21.1
Mean number of waters in solvation shell	1779	1785

As illustrated in Fig. 5.5, the actual space distribution functions P(r) were computed and contrasted with the experimental function. By examining the Gaussian formations, the functions were calculated for the polymer chain from the 5 and 7 ns MD trajectories, after applying the 5 kJ mol^{-1} nm^{-2} force constant. The size and shape of the newly acquired trajectories were assessed using the gmx_sans tool (see section 3.5.2.2). Overall, the computed particle arrangement from both MDs seems sensible and agrees with the experimental shape,

along with the maximum distance, D_{max}, of 67.8 Å. A slightly extending tail is presented in the experimental P(r) resulting in better agreement with the 7 ns MD conformations, especially from 28-42 Å distance. This may suggest the shorter MD misses some rare but important conformations. It should be noted that the experimental P(r) function was obtained from inversion analysis of the I(q) in SasView,[22] which estimates the arrangement of the molecule within the solvation layer. While the computed P(r) functions, presented here, were calculated considering the PVP chain only. As a result, the P(r) shape for the 5 ns MD, the nearest average R_g compared to the experiment, shows a slight deviation around the aforementioned distance.

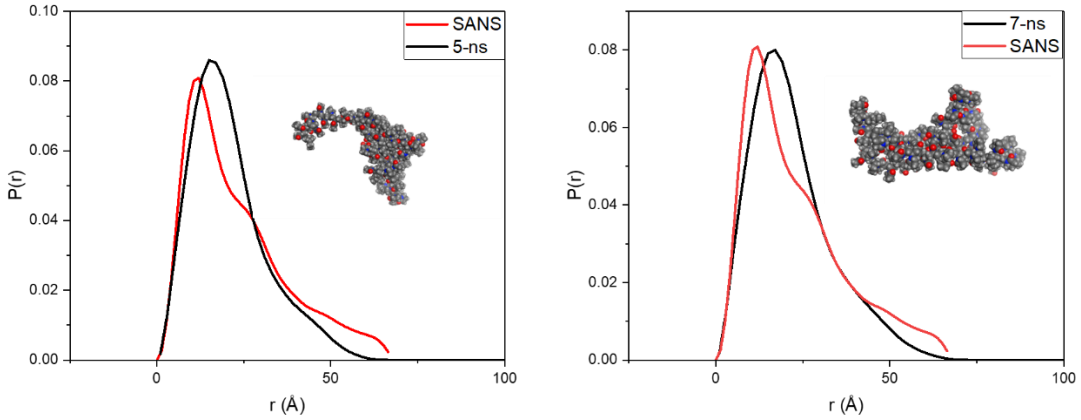


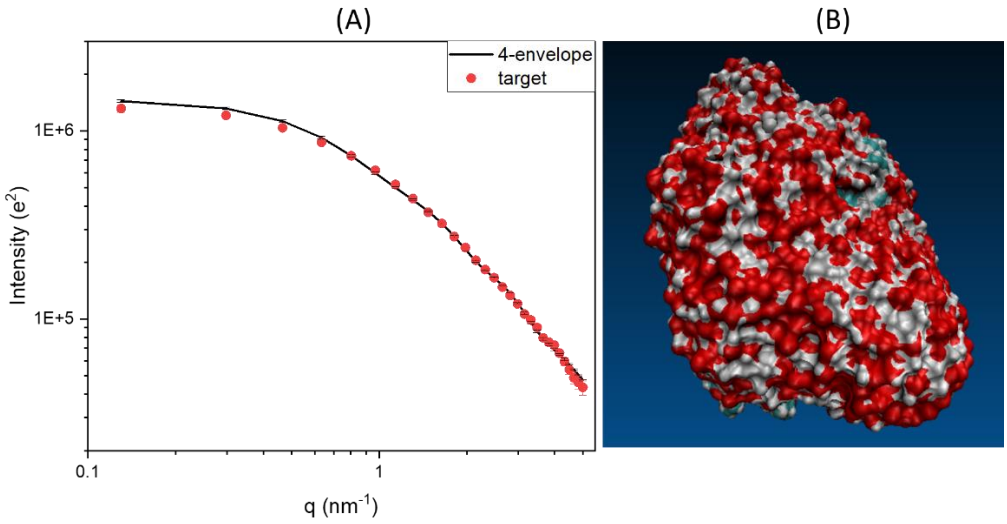
Figure 5. 5: Comparison of the P(r) function between calculated (black line) and experiment (red line) for PVP in solution calculated from the output trajectory obtained with a force constant of 5 kJ mol^{-1} nm^{-2} for 5 and 7 ns MD simulations.

5.3.2.2 SANS-driven MD simulations with less envelope size

To confirm our understanding of the solvation layer thickness, an extra SANS-driven MD attempt was made with a decrease in the distance between the polymer atoms and the envelop edge from 6 to 4 Å, having seen previously that the thickness of the solvation shell was mostly less than 3 Å, to examine any potential improvement in the SANS fitting at low q. It seems that the envelope size plays a critical role in the low q fitting by adding or reducing the electron density around the polymer surface, resulting in determination of the calculated scattering pattern and the Guinier fit. As a result, when the envelope size was reduced, as Fig. 5.6 (A) shows, a more favourable Guinier fit, 18.7 vs 18.6 Å, as well as an improved SANS pattern fit. The polymer chain's average R_g is 17.4 Å (SD 0.7 Å), and the solvation layer expands the chain's R_g by roughly 1 Å. Compared to the most recent MDs, there were more estimated water

molecules (2,000) observed in the solvation shell. Consequently, when the size of the envelope dropped, the SANS pattern and R_g value of the PVP solution exhibited a closer resemblance to the actual data, particularly at low Q. Getting the envelope in the PVP system smaller helps guide the internal PVP structure to the desired shape. This makes it easier for water molecules and the polymer chain to interact with each other. This leads to a higher count of water molecules in the solvation layer. Moreover, decreasing the distance of the envelope from the polymer atoms can verify the PVP solution's estimated solvation layer thickness is less than 3 Å and show the appropriate solvation thickness in relation to the internal PVP structure (see Fig. 5.6 (B)). The results for these distances are compatible with what we have seen in the Dissolve's results, which show the g(r) for the hydrogen bonding between polymer chain and water (HW-O807) has a featuring peaks up to 3 Å, and that is a sensible finding for the non-ionised system (see section 4.8 in the previous chapter).

We then computed the P(r) functions for the later trajectory using gmx_sans (see section 4.5) and plotted them against the experimental function, as illustrated in Fig. 5.6 (C). The purpose of this investigation was to investigate the Gaussian shape of the PVP, as the method used in this chapter (SANS-driven MD) does not have the capability to predict Gaussian shapes. The projected form P(r) and maximum distance (D_{max}) closely approximate the experimental data, but with minor discrepancies. These deviations arise from the utilisation of a distinct computational approach that does not account for the presence of water molecules surrounding the polymer chain. Additionally, the experimental analysis was conducted on a relatively polydisperse sample.



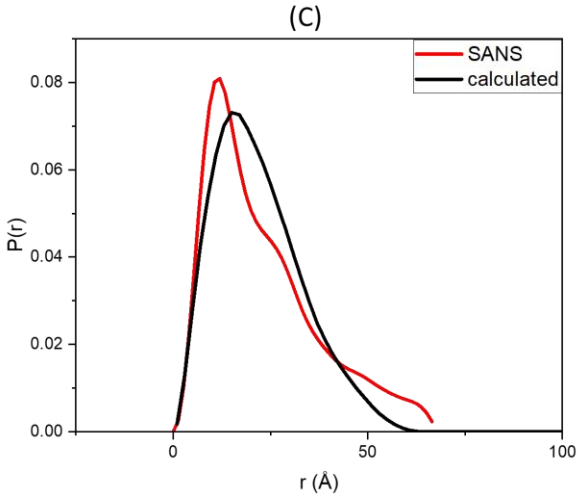


Figure 5. 6: (A) Comparison between calculated (solid black) and experimental SANS (red points) curves for PVP solution (with force constant 5 kJ mol^{-1} nm^{-2}) when decreasing the envelope size to 4 Å (B) The representation of the PVP structure surrounded with about 2,000 water molecules in the solvation shell, which produced the better fit at low q and Guinier fit compared to the SANS data (C) Comparison of the P(r) function between calculated (black line), calculated from sans tool, and experiment (red line) for PVP in solution calculated from the output trajectory obtained when decreasing the envelope size from 4 to 6 Å with a force constant of 5 kJ mol^{-1} nm^{-2}.

5.4 SANS-Driven MD for PEG Solution

5.4.1 Examination of different force constant values

The SANS-driven MD was applied to the PEG solution in an attempt to get an acceptable match. First, a look at each MD's different force constant values, which varied from 1 to 10 kJ mol^{-1} nm^{-2}. Fig. 5.7 and Table. 5.3 compare the experimental and estimated SANS fit curves and display the quantity of water molecules in the solvation shell, the Guinier fit, and the expected R_g for the polymer chain. The closest Guinier approximation, 28.0 Å, and better SANS fit to the scattering data were achieved with a force of 6 kJ mol^{-1} nm^{-2}. However, other attempts do not yield a realistic result or fit. Even though the obtained R_g is somewhat larger than the experimental R_g, it may fall within the experimental error ranges of $\pm$ 2.5 Å.

Despite the fact that the PEG system has a simpler structure than PVP, the PEG solution faces a tough fit. This is because the PEG chain is flexible and can adopt a wide range of conformations that can be explored over extended simulation times. This has an impact on the fitting results, which can require a lengthy fitting process that is difficult to complete in one

sitting for this kind of investigation. In contrast, the PVP structure can accommodate fewer conformations than PEG because of its side chain, which reduces flexibility movement and speeds up the process of achieving a good PVP fit.

The solvation shell of the PEG system contains about 7,500 water molecules (about 26 water molecules per monomer unit), according to Table 5.3, considering the improved fit, R_g 28 Å. This may indicate that the PEG structure is slightly more soluble in water than PVP, which has about 2000 water molecules in its solvation shell (about 22 water molecules per monomer unit). It can be noted the Mw for both system is not exactly the same, but they are not so far apart. The Mw for PEG and PVP are nearly 13,000 and 10,000 g/mol, respectively. There are more water molecules in the solvation shell of the PEG system due to its hydrophilic backbone, allowing an increase in intermolecular interactions between the chain and water molecules, than in the PVP system, which has a hydrophobic backbone that does not attract a high number of water molecules.

The successful application of the driven MD method can be confirmed by comparing the number of water molecules found within 4 Å of the polymer chains. In the regular MD simulation (refer to Chapter 3 and Chapter 4), roughly 1200 (for PEG) and 550 (for PVP) water molecules were observed; however, in the driven-MD simulation, 7500 (26 per monomeric PEG unit) and 2000 water molecules (22 per monomeric PVP unit) were present at the same distance, respectively. The quantity of water molecules identified in this study is larger compared to previous studies that used ^{2}H NMR measurements for PVP (6-9 water molecules per monomeric unit)[23] and Differential Scanning Calorimetry (DSC) for PEG (3.3 per ethylene glycol unit)[24]. This suggests the presence of a favourable intermolecular potential between the polymer chains and solvent, which enables a greater concentration of solvents in close proximity to the experiment. Therefore, this method may be used to produce a very accurate model of the polymer structure in solutions by taking into account the water molecules in the solvation shell that have an impact on the prediction of R_g.

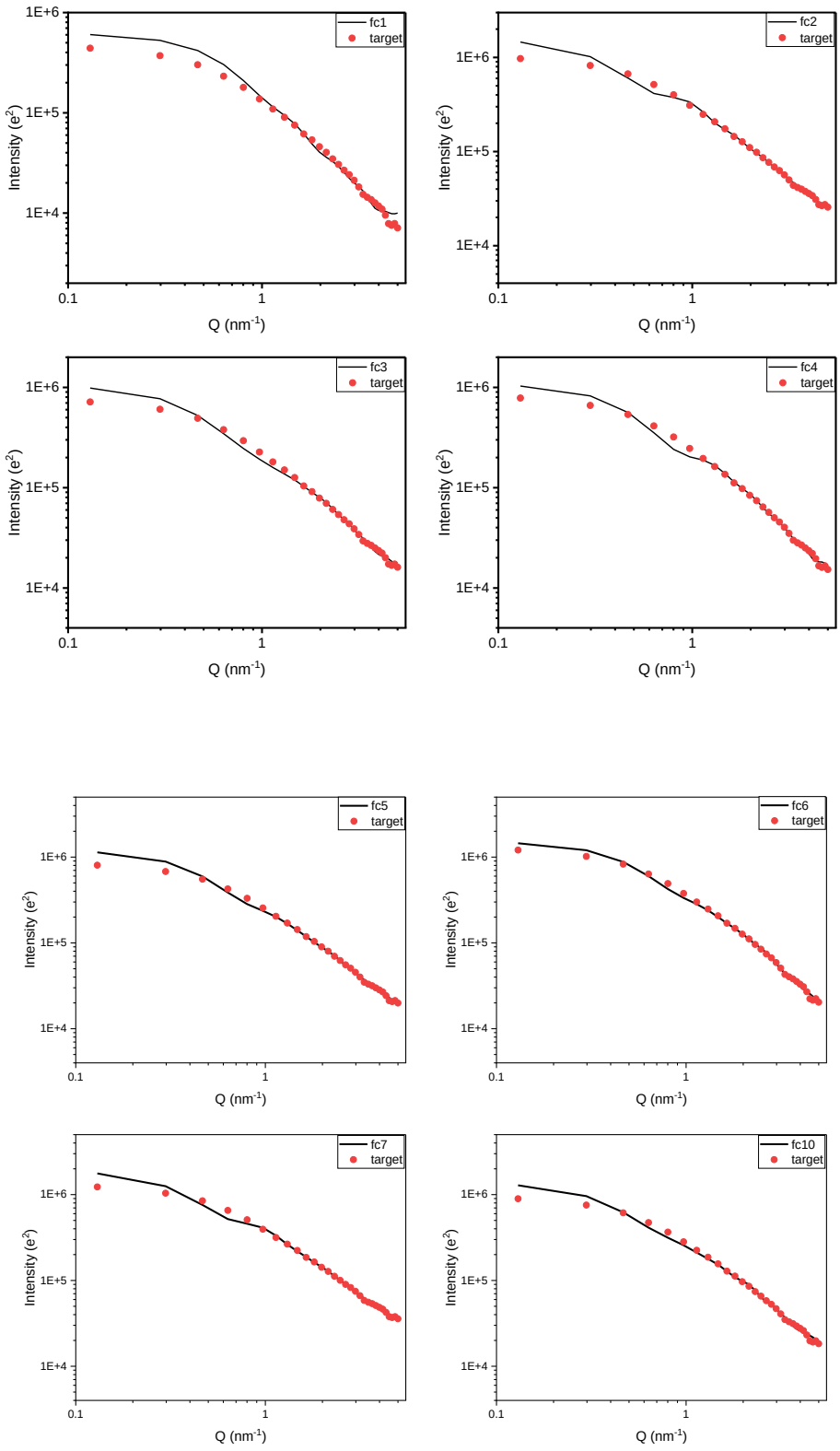


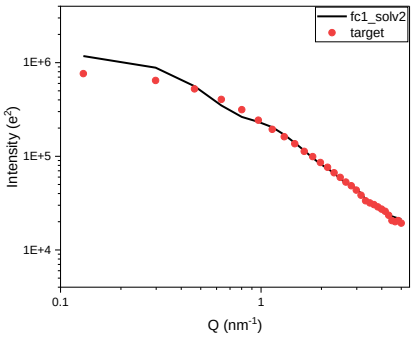
Figure 5. 7: Comparison between calculated (solid black) and experimental SANS (red points) curves for PEG solution with different force constant values (1-10 kJ mol^{-1} nm^{-2}).

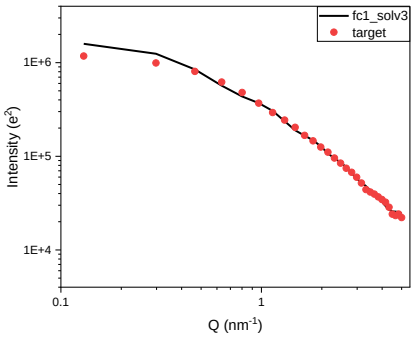
Table 5. 3: Statistical analysis of the SANS-driven MDs, solute-R_g, Guinier fit, and number of water molecules in the solvation shell, for PEG solution in different force constants (kJ mol^{-1} nm^{-2}) with 4 Å envelope size.

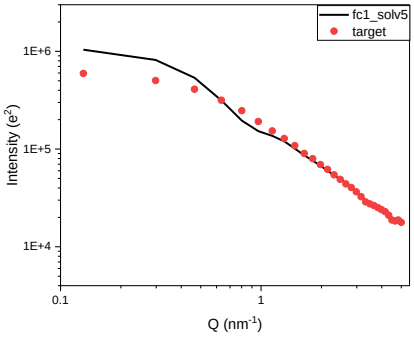
Force constant (kJ mol^{-1} nm^{-2})	1	2	3	4	5	6	7	10
Average radius of gyration (solute only, electron-weighted) Å	27.5	30.1	28.2	27.8	29.0	26.1	30.4	30.6
Standard deviation Å	1.1	2.6	0.8	0.7	1.6	1.8	1.5	1.7
radius of gyration (Guinier fit) Å	23.5	38.7	32.0	30.6	32.2	28.0	37.9	34.8
Average number of water molecules in the solvation shell	7507	7521	7519	7515	7517	7529	7539	7529

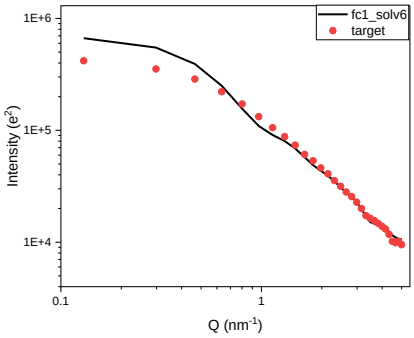
5.4.2 Examination of different envelope sizes

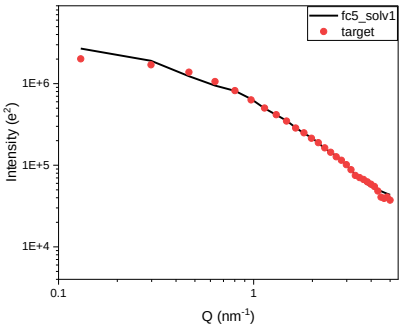
The attempts of examination of different envelope sizes of PEG system can be found in Fig. 5.8 and Table. 5.4 below. None of these attempts provide a good fit with a reasonable Guinier approximation. This suggests the above selected size, 4 Å, was a suitable envelope size for both PVP and PEG systems.

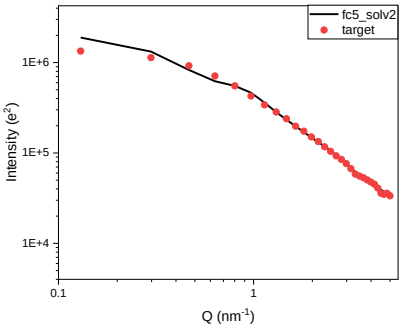

fc1_solv2
target
Intensity (e^2)
1E+6
1E+5
1E+4
0.1
1
Q (nm^{-1})

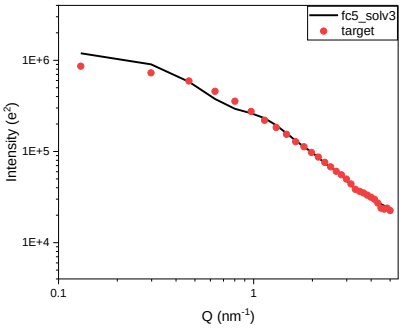

fc1_solv3
target
Intensity (e^2)
1E+6
1E+5
1E+4
0.1
1
Q (nm^{-1})

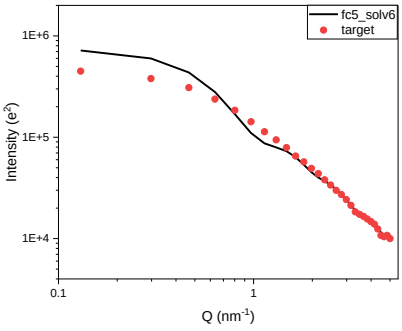

fc1_solv5
target
Intensity (e^2)
1E+6
1E+5
1E+4
0.1
1
Q (nm^{-1})


fc1_solv6
target
Intensity (e^2)
1E+6
1E+5
1E+4
0.1
1
Q (nm^{-1})


fc5_solv1
target
Intensity (e^2)
1E+6
1E+5
1E+4
0.1
1
Q (nm^{-1})


fc5_solv2
target
Intensity (e^2)
1E+6
1E+5
1E+4
0.1
1
Q (nm^{-1})


fc5_solv3
target
Intensity (e^2)
1E+6
1E+5
1E+4
0.1
1
Q (nm^{-1})


fc5_solv6
target
Intensity (e^2)
1E+6
1E+5
1E+4
0.1
1
Q (nm^{-1})

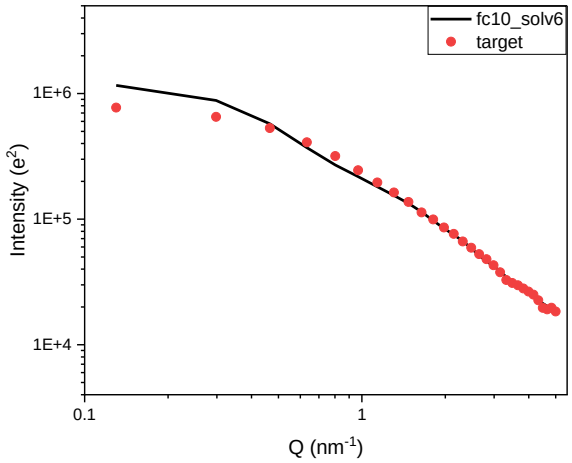


Figure 5. 8: Comparison between calculated (solid black) and experimental SANS (red points) curves for PEG solution with different force constant values (1, 5 and 10 kJ mol^{-1} nm^{-2}) with different envelope sizes.

Table 5. 4: Statistical analysis of the SANS-driven MDs, solute-R_g, Guinier fit, and number of water molecules in the solvation shell, for PEG solution in different force constants (1, 5 and 10 kJ mol^{-1} nm^{-2}) with different envelope sizes.

Force constant (kJ mol^{-1} nm^{-2})-envelope size	1-2	1-3	1-5	1-6	5-1	5-2	5-3	5-6	10-6
Average radius of gyration (solute only, electron-weighted) Å	29.0	29.9	29.0	29.0	29.7	29.2	28.2	27.7	29.9
Standard deviation Å	0.9	1.3	0.5	0.6	1.4	0.9	0.9	0.7	1.3
radius of gyration (Guinier fit) Å	34.5	32.0	31.8	28.5	38.0	38.6	34.2	27.5	34.0
Average number of water molecules in the solvation shell	5715	6626	7519	8344	4722	5728	6617	9362	9375

Conclusion

One effective way to improve both the accuracy and resolution of simulations is to fit them to experimental data. In a simulation, the driven-MD should be both computationally efficient and correct when applying a constraint. Our study presents the validation of the MDs against SANS data, which results in more compact structures with a lower R_g value. To overcome the

inaccuracy of the force field and obtain realistic MD corresponding with experimental data, the SANS-driven MD method was utilized. The improved Guinier fit was achieved at a force constant of 5 kJ mol^{-1} nm^{-2}. At both low and high q values, a superior fit between the computed and experimental SANS curves and the Guinier fit, 18.7 vs. 18.6 Å, was obtained when reducing the envelope size to 4 Å. This concludes the prediction that roughly 2,000 water molecules could participate within the solvation layer, which has a thickness of less than 3 Å in all calculated frames for the PVP system. The PEG system, on the other hand, seemed to be hard to match the experimental curve because of the chain flexibility movement, which required a lengthy simulation time to explore various conformations while both computing and fitting the scattering curve into the experimental data. The PEG system's acceptable fit and R_g were obtained with a minor increase of 1.5 Å from the experimental R_g, which can fall within the experimental error, ± 2.5 Å. PEG has greater water solubility compared to PVP, as evidenced by the increased quantity of water molecules found in the solvation shell.

Reference

1 S. K. Pal and A. H. Zewail, Dynamics of Water in Biological Recognition, *Chem. Rev*, 2004, **104**, 2099–2123.

2 P. Ball, Water as An Active Constituent in Cell Biology, *Chem. Rev.*, 2008, **108**, 74–108.

3 J. Linse and J. S. Hub, Scrutinizing the protein hydration shell from molecular dynamics simulations against consensus small-angle scattering data, *Commun. Chem.*, 2023, **6**, 272.

4 I. Hassan, F. Ferraro and P. Imhof, Effect of the Hydration Shell on the Carbonyl Vibration in the, *Molecules*, 2021, **26**, 2148.

5 D. Laage, T. Elsaesser, J. T. Hynes and E. Normale, Water Dynamics in the Hydration Shells of Biomolecules, *Chem. Rev*, 2017, **117**, 10694−10725.

6 A. C. Fogarty and D. Laage, Water dynamics in protein hydration shells: The molecular origins of the dynamical perturbation, *Phys. Chem. B*, 2014, **118**, 7715–7729.

7 Y. Levy and N. Onuchic, Water Mediation in Protein Folding and Molecular Recognition, *Annu. Rev. Biophys. Biomol. Struct.*, 2006, **35**, 389–415.

8 A. C. Fogarty, E. Duboué-Dijon, F. Sterpone, J. T. Hynes and D. Laage, Biomolecular hydration dynamics: a jump model perspective, *Chem. Soc. Rev.*, 2013, **42**, 5672–5683.

9 R. Harada, Y. Sugita and M. Feig, Protein Crowding Affects Hydration Structure and Dynamics, *Am. Chem. Soc.*, 2012, **134**, 4842−4849.

10 F. Merzel and J. C. Smith, Is the first hydration shell of lysozyme of higher density than bulk water ?, *Proc. Natl. Acad. Sci.*, 2002, **99**, 5378–5383.

11 J. S. Hub, Interpreting solution X-ray scattering data using molecular simulations, *Curr. Opin. Struct. Biol.*, 2018, **49**, 18–26.

12 T. E. Gartner and A. Jayaraman, Modeling and Simulations of Polymers : A Roadmap, *Macromolecules*, 2019, **52**, 755−786.

13 T. I. Imae, T. Kanaya, M. Furusaka and N. T. Torikai, *Neutrons in Soft Matter*, John

Wiley & Sons, 2011.

14 P. Chen, R. Shevchuk, F. M. Strnad, C. Lorenz, L. Karge, R. Gilles, A. M. Stadler, J. Hennig and J. S. Hub, Combined Small-Angle X - ray and Neutron Scattering Restraints in Molecular Dynamics Simulations, *J. Chem. Theory Com*, 2019, **15**, 4687−4698.

15 W. He, A. Henning-knechtel and S. Kirmizialtin, Visualizing RNA Structures by SAXS-Driven MD Simulations, *Front. Bioinforma.*, 2022, **2**, 1–19.

16 L. Chatzimagas and J. S. Hub, Predicting Solution Scattering Patterns with Explicit-Solvent Molecular Simulations, *Methods Enzymol.*, 2022, **677**, 433–456.

17 D. Mirandela, G. Tamburrino, M. T. Ivanovic, F. M. Strnad, O. Byron, T. Rasmussen, P. A. Hoskisson, J. S. Hub, U. Zachariae, F. Gabel and A. Javelle, Merging In-Solution X - ray and Neutron Scattering Data Allows Fine Structural Analysis of Membrane − Protein Detergent Complexes, *Phys. Chem.*, 2018, **9**, 3910−3914.

18 M. R. Hermann and J. S. Hub, Chapter 9 Interpreting SAXS / WAXS Data with Explicit-Solvent, *Biophys. Membr. Proteins Methods Protoc.*, 2020, **2168**, 199–215.

19 SAXS/WAXS/SANS calculations using all-atom MD simulations, https://biophys.uni-saarland.de/grenoble-tut/, (accessed 1 February 2022).

20 J. N. Dahanayake and K. R. Mitchell-koch, How Does Solvation Layer Mobility Affect Protein Structural Dynamics ?, *Front. Mol. Biosci.*, 2018, **5**, 65.

21 J. Hernández-lima, K. Ramírez-gualito, B. Quiroz-garcía, A. L. Silva-portillo, E. Carrillo-nava and F. Cortés-guzmán, How solvent determines the molecular reactive conformation and the selectivity : Solvation spheres and energy, *Front. Chem.*, 2022, **10**, 1012769.

22 Sasview, https://www.sasview.org/, (accessed 31 August 2022).

23 M. Nakada, H. Ishida and Y. Furushima, Structural and dynamical characterisation of intermediate water interacting polyvinyl pyrrolidone, *Materialia*, 2020, **12**, 100743.

24 L. Huang and K. Nishinari, Interaction between poly (ethylene glycol) and water as studied by differential scanning calorimetry, *J. Polym. Sci. Part B Polym. Phys.*, 2001, **39**, 496–506.

Chapter. 6

Conclusions and Outlook

This thesis has introduced a methodology for determining the structural characteristics of polymers in dilute solutions. Combining experimental and computational approaches can allow for a better understanding of polymer solutions, which in turn helps to develop novel materials and processes with improved performance and characteristics.

Here, we begin in Chapter 3 with an evaluation of the ability of different force fields and implicit solvents to reproduce a good range of polymer conformations for the simplest polymer structure (PEG). The OPLS-AA force field was selected with an acceptable range that can be used in the further explicit MD study. This choice of forcefield is consistent with that used in previous studies[1–3] on PEG. All-atom MD simulations were then assessed using two different solvents and box sizes for PEG292. The MD with the TIP3P model and applied experimental concentration box displayed a good agreement in the radius of gyration value compared to SANS data. The SANS intensity curves for PEG292, produced by summing the output trajectories of all MD frames (1% and 4% polymer concentrations independently), were fitted over the Q range of 0–0.25 $Å^{-1}$ using the sans tool in GROMACS. This was done to examine the similarity between the estimated curves and the experimental results in terms of their sensitivity to the R_g. The observed scattering pattern appears to be influenced by the R_g range of conformations found between 1 and 4% MD, with a little inclination towards the 1% MD curve rather than the 4% MD curve. The disparity between the calculated and SANS patterns becomes apparent at 0.1 $Å^{-1}$ in both MDs as a result of the unaccounted solvation layer in this particular computational approach. A sensible agreement between the experimental and calculated P(r), the Gaussian arrangement particle shape, was obtained, particularly for the average R_g, 25 Å, conformer (in the 4% MD) with a similar D_{max}.

Examining another polymer (PVP) behaviour in Chapter 4 using the above-mentioned force fields, OPLS-AA/TIP3P, the mismatch between the experimental and calculated R_g values was clearly observed in the different solvent models (TIP4P, TIP5P, SPC) and even with the experimental concentration applied. This led to testing the CHARMM force field, which produced a very high R_g. The tacticity characteristic that the PVP structure possessed raised the question of whether that could have some effects on the R_g calculation. Three independent MD simulations with three stereochemical configurations: atactic, isotactic, and syndiotactic,

were carried out explicitly, and out of all stereochemical structures, isotactic conformation showed superior results in the first MD attempted run in terms of average R_g in comparison to the SANS value, conformational range over the MD, and number of water in the solvation layer. Subsequently, six more independent MDs were conducted to assess the stability of the MD for each model, while the experimental sample was confirmed as atactic by the ^{13}C NMR spectrum.

The compact structure (PVP) with a lower R_g value observed during the validation of the MD against SANS data guided us in Chapter 5 to utilize the SAS-driven MD method, which was successfully reported for many protein studies. This is an effective way to improve the simulation results and overcome the imperfection of the applied force fields by restricting the MD using the experimental scattering data. This method can also take into account the solvation layer, which has been proven to have an effect on the R_g value, by conducting an envelope around the molecule. This method was applied to the two studied polymers, PVP and PEG. After many tests for different force constant values and envelope sizes, excellent SANS curve and Guinier fits were achieved for PVP structure at 5 kJ mol^{-1} nm^{-2} and 4 Å, respectively. The enhancement of polymer-water interactions results in a higher number of water molecules found within 4 Å compared to the free MD calculations in Chapter 4. However, the flexible chain PEG hinders obtaining a good match in both the Guinier and SANS fits with the current available timescale for this type of analysis.

Introducing the first use of the Dissolve software, which integrates the total neutron scattering obtained from NIMROD for the analysis of dilute polymer solutions, is presented here for both polymers (PEG in Chapter 3 and PVP in Chapter 4). For both solutions, the predicted and experimental total weighted neutron structure F(q) and weighted neutron radial distribution function G(r) showed good agreement, particularly for the deuterated solvent samples, which have less noise. It should be noted that the main correlation features observed while working with diluted solutions are clearly related to the solvent-solvent connection, as can be shown by examining the predicted scaling of the sum partial correlations for polymers, solvent, and polymer-solvent. Thus, finding the appropriate model over the entire range of Q lengths has proven to be a difficult effort that requires numerous validation procedures.

A flowchart (Fig. 6.1) is created to provide a comprehensive overview of the work conducted in this study.

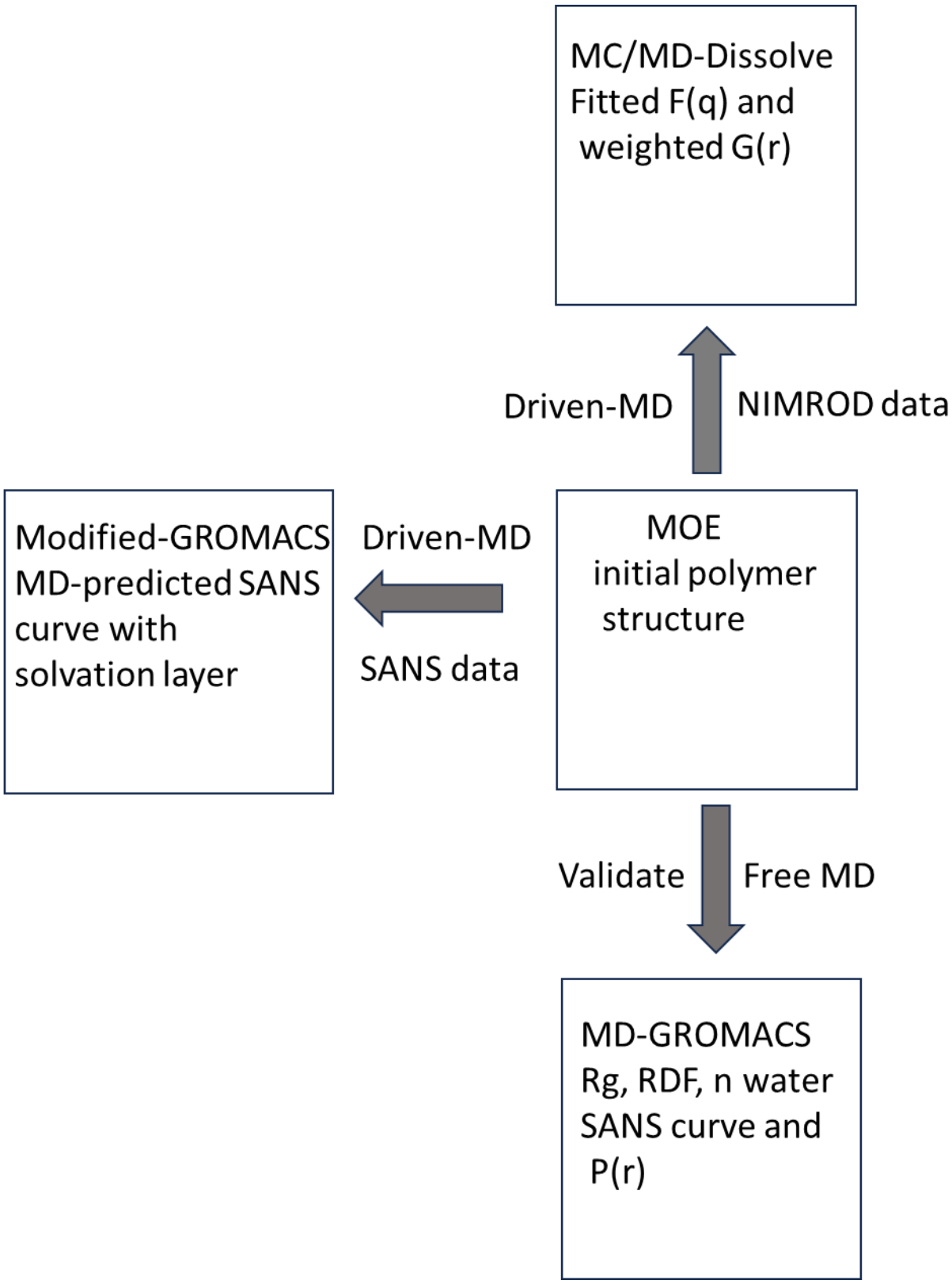


Figure 6. 1: A Flowchart depicting a brief overview of work conducted in this thesis.

Much of this work on validating the MD simulations against scattering data for pure polymer systems is a critical step for the subsequent study of exploring new generations of polymer materials. As seen from this thesis, the driven MD simulations is the key to understand the dynamics of polymers in solutions, but many parts remain unexplored and can be studied further. For example, in the Dissolve study (section 3.5.3.7), the number of water molecules within a 4 Å distance from polymer chains can be determined by analyzing the Dissolve MD data before and after applying EPSR. This analysis can be conducted to assess any improvements and confirm the capability of the Dissolve program to capture accurate intermolecular interactions between polymers and water. Additionally, in the NIMROD data, which has the total scattering information, looking at a specific Q range independently to study the validated details of solvent-solvent, polymer-solvent, polymer-polymer, and even atom-atom interactions in the high q region would be valuable to draw a final concluded model that covers all aspects of affirmation parameters.

Developing a software that has the ability to drive SAS/WAS MD while mimicking the reality of polymers in solutions (including polydispersity) using different chain length models that can be verified by matrix-assisted laser desorption/ionization (MALDI). This is indeed a very difficult task, especially for polymers that contain many units, which needs a timescale consideration that could be ignored while using GC models that have a successful mapping that provides beads with hydrogen bindings donor/acceptor. Even trying to simulate an all-atom solution of identical chain length in one box is not working here, as evidenced by Dissolve MD (for polymers) and the extra work we did in GROMACS. In Appendix A, the R_g results for attempted simulations with 2 identical PEG chains using GROMACS can be found, showing that the polymer chains are restricted for free movement, unlike the R_g results seen in Chapter 3 using one polymer chain in the box.

Preliminary investigations were conducted on pure polymer solution systems, and further comprehensive research might be undertaken on mixed systems to elucidate novel insights into the behavior of polymer-drug solutions. With the help of developed techniques, new systems can be created and investigated to learn more about the contributions made by different aspects to the behavior of the solution.

NMR spectroscopy offers structural validation, conformational constraints, force field parameterization, tacticity insights and dynamic behavior, making it an invaluable experimental tool for enhancing the reach of MD simulations. By combining experimental

NMR data with MD simulations, scientists can take advantage of both methods' advantages and learn more about the dynamics and structure of molecular systems.

Reference

1 U. R. Dahal and E. E. Dormidontova, The dynamics of solvation dictates the conformation of polyethylene oxide in aqueous, isobutyric acid and binary solutions, *Phys. Chem. Chem. Phys.*, 2017, **19**, 9823–9832.

2 M. Stepniewski, M. Pasenkiewicz-Gierula, T. Rog, R. Danne, A. Orlowski, M. Karttunen, A. Urtti, M. Yliperttula, E. Vuorimaa and A. Bunker, Study of PEGylated lipid layers as a model for PEGylated liposome surfaces: Molecular dynamics simulation and langmuir monolayer studies, *Langmuir*, 2011, **27**, 7788–7798.

3 M. M. Hoffmann, M. D. Too, N. A. Paddock, R. Horstmann, S. Kloth, M. Vogel and G. Buntkowsky, On the behavior of the ethylene glycol components of polydisperse polyethylene glycol PEG200, *J. Phys. Chem. B*, 2023, **127**, 1178–1196.

Appendix A

Chapter 3

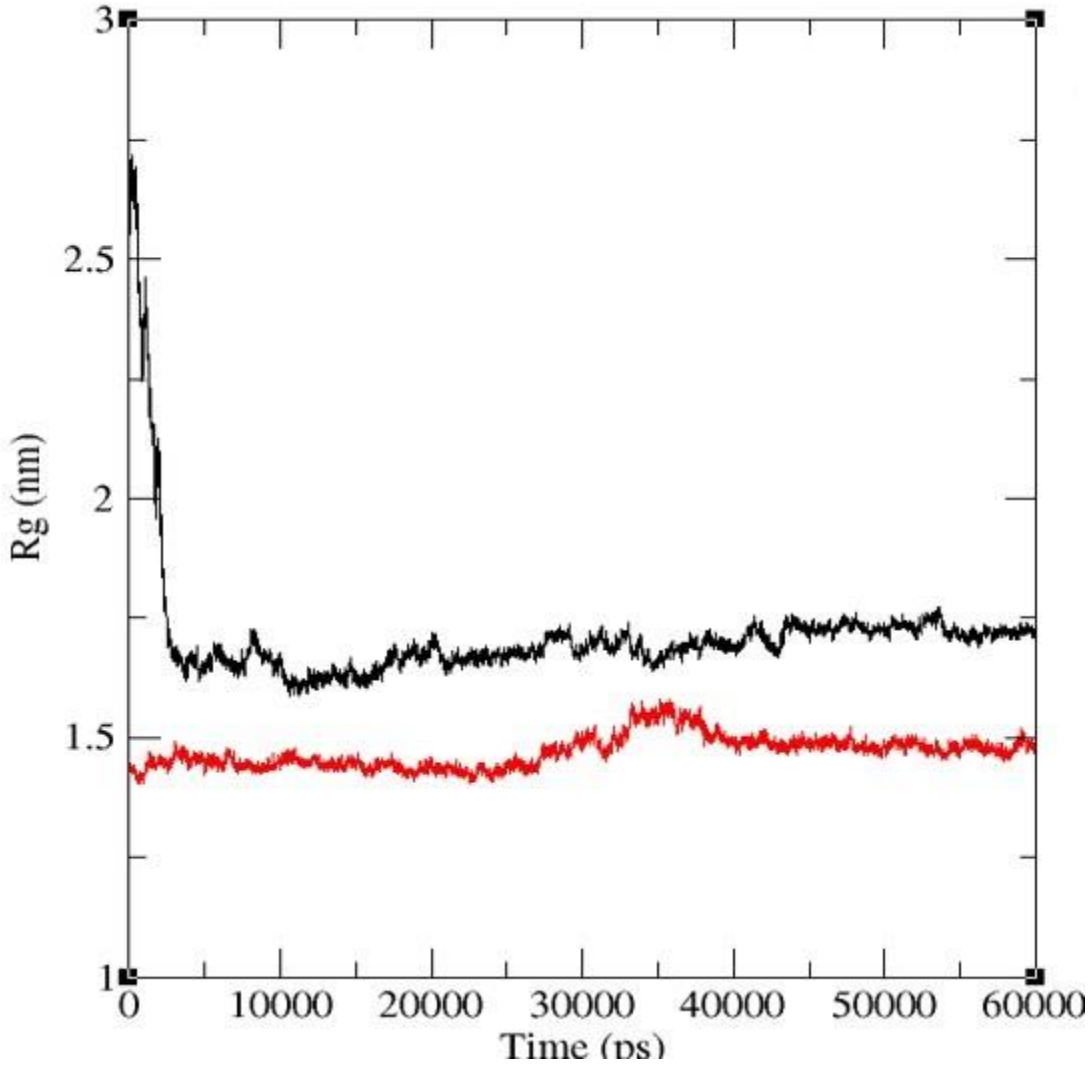


Figure A. 1: R_g (nm) plot versus time (ps) for 2-identical PEG chains in a box size 163 $Å^3$ (1%) using OPLS-AA/TIP3P models.

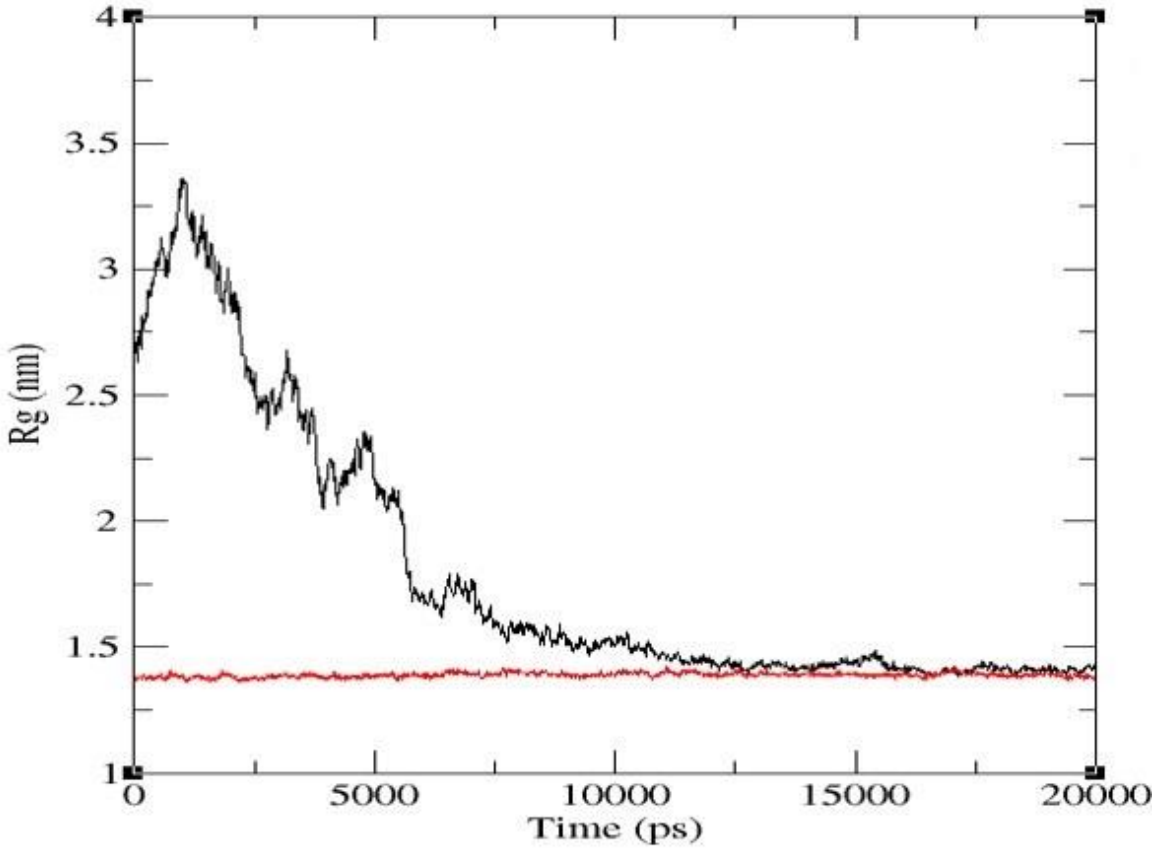


Figure A. 2: R_g (nm) plot versus time (ps) for 2-identical PEG chains in a box size 104 $Å^3$ (4% experimental concentrations) using OPLS/TIP3P models.