ALL-ATOM AND COARSE-GRAINED MOLECULAR DYNAMIC STUDY OF POLYMERS

HEMANT SINGH KHOBA

DISSERTATION FOR MASTER

Supervisor FEUP: Prof. Alexander Afonso

Co-supervisor: Dr. Sacha Mould

MASTER'S IN COMPUTATIONAL MECHANICS

2020

System Administrator 26 Oct 2020 at 13:08:02 18:7e:b9:0a:6c:66

System Administrator 26 Oct 2020 at 13:08:02 18:7e:b9:0a:6c:66

Confidentiality Clause

This research work was generated at ABBt Team Schwaneberg, Lehrstuhl für Biotechnologie, RWTH Aachen university under the supervision of Dr. Mehdi D. Davari and contains confidential data of HICAST.

This work may only be made available to the first and second reviewers and authorized members of the board of examiners. Any publication and duplication of this research report - even in part - is prohibited.

An inspection of this work by third parties requires the expressed permission of HICAST.

Alexandre Afonso Sacha T. Mould

ALL-ATOM AND COARSE-GRAINED MOLECULAR DYNAMIC STUDY OF POLYMERS

Abstract

The formulation of laundry detergents is very complex, reflecting the diverse requirements of the application. In general detergents consists of different classes of raw materials such as surfactants, builders, bleaching agents, enzymes, water softeners, fragrances and others[1]. All these various components have different functions in the washing solution but all of them are essential for a good performance.

Recent studies revealed that polymers do not only work as builders in laundry detergents but also interact with proteases, which increases the washing performance. As molecular dynamics (MD) simulation have become a versatile tool in the last years to study macro- and biomolecules, a computational protocol is developed to gain insights into polymer-ion interaction.

Several polymers have shown interaction with enzymes in detergents and thus affecting the performance during the washing process. However, the interaction mechanism of polymers and enzymes in laundry detergents is not yet explored. In order to investigate these interactions with molecular dynamics (MD) simulation, the used polymers must be parameterized first, as the standard force fields in MD simulations are made for biomolecules. Therefore, the main objective of this thesis is to understand the local, monomer-level arrangements, fluctuations, intermolecular interactions within the chain and between the chain and ions interactions, electrostatic interactions, effect of salt concentration on polymer systems at hand.

In salt-free polymer solutions the electrostatic interactions between charged groups on the polymer backbone result in a strong chain elongation with chain size scaling almost linearly with the chain degree of polymerization. Because of this strong dependence of the chain size on the chain degree of polymerization, the crossover to semi dilute polyelectrolyte solution regime occurs at much low polymer concentrations than in solutions of neutral polymers. The main contribution to the osmotic pressure in polyelectrolyte solutions comes from the ionic component. Polyelectrolyte conformations are sensitive to the solvent quality for the polymer backbone. In poor solvent conditions for the polymer backbone a polyelectrolyte chain forms an unusual necklace-like structure of dense polymeric beads connected by strings of monomers. Similar necklace-like structure can be formed in hydrophobically modified polyelectrolytes in which a hydrophobic side chains are attached to the polyelectrolyte backbone. Addition of salt leads to screening of the electrostatic interactions between ionized groups reducing the polyelectrolyte effect[2]

Due to computational expense, atomistic models are usually limited to length scales of 1- 100 Å and time scales of 1 fs (femtoseconds) – 100 ns (nanoseconds). Due to the limitations of all atom simulations we used coarse grained (CG) models to probe phenomena's involving polymer chain-level rearrangements that may occur over larger length and time scales. The sizes of CG beads and the bonded and the nonbonded potentials between all CG beads were chosen to reproduce the relative stiffness of various parts of the monomer and the placement of charges as seen in atomistic simulations and to ensure that electrostatic interactions are the dominant driving force in this problem. To ensure that this growing power of simulation is harnessed correctly, and meaningful results are achieved, special care was taken to ensure the validity and reproducibility of these simulations.

To reach this aim, we developed a computational protocol for the parameterization of new molecules for all-atom and coarse-grained molecular dynamics simulations with the AMBER force field and MARTINI force field, respectively. Further, the conformational parameters of polyelectrolytes in salt free and excess salt in synthetic lab water solution was studied.

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

This work focuses on two different model polymers, polyacrylic acid (PAA) and poly-γ-l-glutamic acid (γ-L-PGA). Both polymers are fully deprotonated and have a chain length of 20 and 62 monomers at basic pH. Simulations were carried out using two different methodologies of molecular modelling, all-atomistic (AA) and coarse-grained (CG) for 100 ns and 500 ns respectively using GROMACS 2019.1 molecular dynamics package with AMBER99SB force field and MARTINI force field, respectively.

MD simulations of PAA and γ-L-PGA in water with monovalent and divalent counter ions, showed that PAA and γ-L-PGA exhibit a similar hydration behaviour in water, although the polymer chain of PAA has a rather stretched conformation compared to γ-L-PGA caused by the strong electrostatic repulsion of the carboxylate groups. The radial distribution function between carboxylate oxygen atoms and calcium ions shows a slightly higher peak for γ-L-PGA, indicating stronger attraction between γ-L-PGA and calcium ions compared to PAA. γ-PGA MARTINI model indicates more fluctuation for radius of gyration compared to PAA, increase in deviation of Radius of Gyration for 62-mer martini model of PAA from AA and published results and less coiled structure obtained from MARTINI model for γ-PGA in comparison AA.

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

Acknowledgment

I would like to thank Prof. Dr. Ulrich Schwaneberg for the possibility of carrying out this work at the ABBt Team Schwaneberg, Lehrstuhl für Biotechnologie and and Mr Renato Natal for empowering and supporting me to take on this project. Many thanks to Dr. Mehdi D. Davari for the direct supervision and excellent technical support and the provision of the topic of this work. Thank you for the helpful suggestions and the personal attention in recent months.

I would like to specially thank Prof. Alexander Afonso and Mr. Sacha Mould for their continous supervision, help and support during this period. They accompanied me through all the ups and downs during the tenure of the project and always helped me with words and deeds.

A special thank you to Dr. Carlos Gonzalez Lopez for providing countinous valuable expertise, support and feedback in the development of this work.

I also thank Subrata Pramanik, Niklas Siedhoff, Farid Mehri Sofiani, and Fabio Diaz Cruz for the stimulating discussions and thier always positive attitude. Furthermore, I thank all colleagues and former colleagues who have accompanied me during my time and were friendly to my side. Thanks for the many funny hours that made all stress and effort forgotten.

I wholeheartedly thank my parents and grandparents for their inexhaustible support throughout my studies, both morally and financially.

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

Índice de Conteúdos

Abbreviations

α-PGA	Poly-α-glutamic acid
γ-L-PGA	Poly-γ-L-glutamic acid
AA	All-Atomistic
Ca_{strong}	Calcium ion in strong binding site of subtilisin protease
Ca_{weak}	Calcium ion in weak binding site of subtilisin protease
CG	Coarse-grained
CMC	Critical micelle concentration
COM	Centre of mass
e	Eccentricity
FF-1	Force field based on the GROMOS53a6 parameter set
FF-2	Force field based on the ENCAD parameter set
GAFF	General AMBER force field
I	Moments of inertia along the principle axes
LAS	Linear alkyl benzene sulfonate
MD	Molecular Dynamics
NPT	Constant number of particles, pressure and temperature
NVT	Constant number of particles, volume and temperature
PAA	Poly (acrylic acid)
PME	Particle Mesh Ewald
R	End-to-end distance
r	Distance
RDF	Radial distribution function
R_g	Radius of gyration
RMSD	Root-mean-square deviation
RMSF	Root-mean-square fluctuation
SASA	Solvent accessible surface area
VNS	Verlet neighbour search
W	Central interaction site of a water bead

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

1 INTRODUCTION

Molecular dynamics simulation is the methodology for detailed microscopic modelling on the molecular scale. The nature of matter can be well understood by studying the structure and motion of its constituent building blocks, solution to the N-body problem providing the dynamics of the matter. Molecular dynamics involves resolving the classical many-body problem in virtue of studying the matter at the atomistic level. Due to the lack of alternative approaches available to comprehend with problems at the required level of details over a broad range, molecular dynamics methods have been indispensable in field of research, pure or applied. Classically, the trajectories of atoms and molecules are calculated by solving Newton's equation of motion for a system consisting of interacting particles. The interacting forces between the particles and the potential energy are calculated through interatomic or molecular mechanics force fields.

The following work presents molecular dynamic studies of deprotonated polyacrylic acid and poly-l-gamma glutamic acid of 20 and 62 monomer chain length solvated in synthetic lab water having concentration of 15 dh at 300k. The study was conducted through performing all atomistic and coarse-grained simulations. Effects on the conformation of the chain are studied by varying salt and salt concentration in the solution.

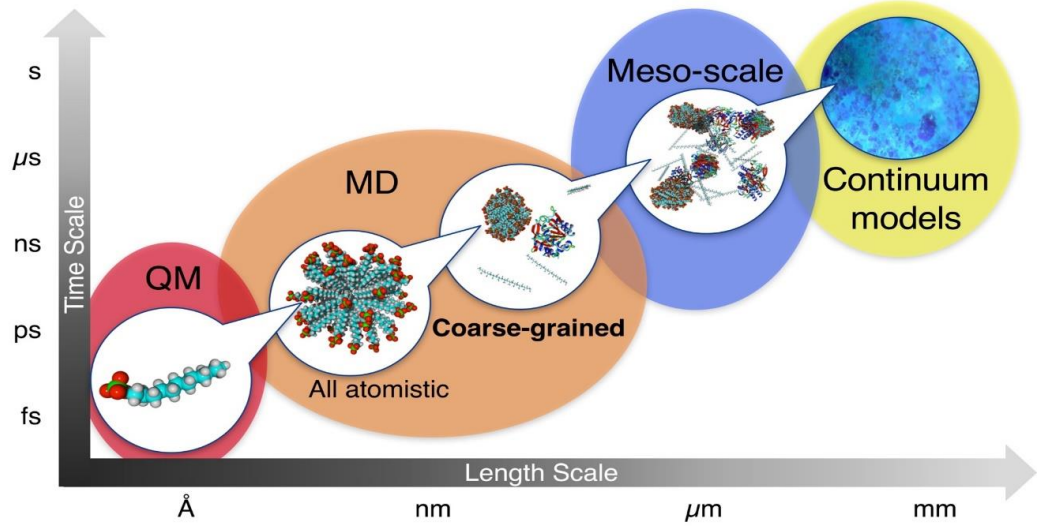


Figure 1 Scheme of time and length scales in the present system. With increasing system complexity and sampled time, methods become less accurate due to computational limitations.

1.1 MOTIVATION

Complex fluids like polymer present unique challenges to computational scientist, exhibiting interesting and important phenomena's over broad range of length and time scales. Not every of these computational challenges have been addressed over the years with numerous developments in software and algorithms. With the help of the advancements, many simulation studies of polymer have been done to provide valuable insight into the new and exciting macromolecular materials and inspire polymer chemists to find synthetic routes to design new materials that show tremendous promise. Polymer models and simulations lies somewhere between that of a chemist (preferring to maintain atomistic details) and a physicist (who prefers to design general models for universal behaviour). One big advantage of molecular simulation is the ability to visualise the real-space arrangements of

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

macromolecules at length scales inaccessible to experimental microscopy techniques. Although many important scattering geometries/techniques can also be calculated from simulation trajectories, including reflectivity,[4, 5] Small angle,[6, 7] and wide-angle[8, 9] scattering, diffraction,[10, 11] dynamic structure factor[12] and others[13-15].

Atomistic models are a useful tool for understanding of the local monomer-level (re)arrangements, fluctuations, or interactions, within a disordered or ordered polymer system. Atomistic simulations are useful to probe where and how counterions interact with charged polymers and/or conformations of the charged species within the polymer chains. However, atomistic models are usually limited to length scales of 1-100 Å and time scales of 1 fs – 100 ns as a result these models cannot probe phenomena's involving polymer chain-level rearrangements that may occur over larger length and time scale. To counter with this limitations, coarse-grained (CG) models are the best option to predict the structure/morphology of polymeric systems at a broad range of conditions or polymer design parameters as CG models reduces some of the degree of freedom in the system by grouping the selected atoms in a monomer/ a whole monomer/ a group of monomers/ Kuhn segments together into a single CG bead.

1.2 OBJECTIVES

The following work presents the computational simulation of the polymers using two different methodology. The main objective of the work presented is to characterize the conformational properties of Poly acrylic (PAA) acid and Poly-γ-L-glutamic (γ-L-PGA) acid in the presence of monovalent and divalent ions for different chain lengths (20-mer & 62-mer), study the size and flexibility of the polymer, interaction with water and surrounding ions using all atomistic modelling and coarse-grained modelling in order to answer prevailing scientific questions regarding the polymers under study in this work like effects of salt concentration on the polymer chain conformation, counterion condensation, effect on the polymer structure in the presence of monovalent and divalent ions in salt free and excess salt solutions, bridging effect of polymer in presence of multivalent ions, solubility and provide general insight into the structural properties of the polymers for future work on qualitative analysis of Kuhn length, persistent length, ion bridging effects on the electro-viscosity of polymer, polymer chain flocculation and concentration effects in polymer solutions.

The secondary objective of this report is to develop a computational protocol for the parameterization of new molecules for MD simulations with the AMBER and MARTINI force field and present the validation and reliability of the currently under development coarse-grained modelling technique of polymers using MARTINI force field through AAMD and experimental results which is beneficial for simulating polymers of higher chain lengths (beyond 1000 monomers) and for longer time period.

1.3 THESIS DISPOSITION

The aim of this thesis is to provide an in-depth knowledge of simulation of the polymers. Hence, it has been structured in a way to provide a complete insight into procedure followed from the beginning i.e. creation of polymer models, parametrization, visualization, calculation of dissociation degree, generation of force fields, mapping of polyelectrolytes using MARTINI force fields till system preparation and simulation of polymers in presence of various ions in different quantities and results extraction.

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

The present investigations address two distinct modelling levels of linear polymer chains: all-atom (AA) and coarse-grained (CG). Specifically, we employed an all-atom AMBER[16] force field for polyelectrolytes, based on symmetric residues and with the force field (FF) parameters optimized relative to high-quality ab initio calculations on a set of model polymers. Unlike other AA FF's for polyelectrolytes reported in the literature [17-19], we consistently optimized not only the dihedral parameters, but also the bond and angle parameters, along with the partial atomic charges. We carried out comprehensive MD simulations and present a detailed analysis of the structural and dynamical behaviour for Poly acrylic acid and Poly-γ-l-glutamic acid deprotonated chains of various lengths and protonation patterns (gyration radius, end-to-end distance, solvent accessible surface). From AA trajectories we constructed as part of the present work probability distributions for the distances, angles and dihedrals formed by the residues, which we used to parametrize a MARTINI force field[20] for CG beads (identified with entire residues). The Boltzmann inversion methodology[21] used to generate the CG bonded parameters was complemented with an additional model to combine the regression parameters resulted for the multi-peak probability distributions. We performed CG simulations for polymer chains of sizes and protonation patterns similar to the atomistic ones and, finally, fine-tuned the CG FF by matching the obtained CG structural and dynamical properties with their AA counterparts. The quality of the CG FF is validated by the fair agreement between the simulated diffusion coefficients and the experimental evidence. In conjunction with existing AMBER and MARTINI FFs for polymers, the elaborated AA and CG FFs for polymers are suitable for realistic large-scale simulations of solvated polyelectrolytes

It is an effort towards providing the reader understanding of the computational chemistry indulged in the field of bioengineering for simulation of polymers used in various fields and day-to-day life for a better understanding of their behaviour in a solution and therefore obtaining crucial results about their atomistic properties.

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

2 LITERATURE REVIEW

2.1 POLYELECTROLYTES

Polyelectrolytes are macromolecules with repeating units bearing an ionic or ionizable group when dissolved in suitable polar solvent. Polyelectrolytes can be either synthetic or natural. Polyelectrolytes properties are like both electrolytes and polymers. Properties of a polyelectrolyte such as solubility in water, polar and hydrogen-bonding liquids (alcohols etc.), electrical conductivity, and solution viscosity are dependent on the characteristic properties like ionic group, counter ion and structure of the repeating unit. Unlike non-ionic polymers, these properties strongly depend on the pH and salt content. Polyelectrolytes are not only of industrial use, but many biological molecules are also polyelectrolytes. For instance, polypeptides and DNA are polyelectrolytes. Nucleic acids, proteins, teichoic acids, some polypeptides, and some polysaccharides are examples of natural polyelectrolytes.

2.1.1 Polyacrylic Acid

Repeating unit of protonated Polyacrylic acid (PAA) is a synthetic homopolymer of acrylic acid (see Fig. 2). In water, at neutral pH value PAA is partially deprotonated. Due to the electrostatic repulsion of the negatively charged carboxylate ions, PAA is not coiled but rather has a stretched conformation allowing bigger distance between the charge groups, which causes a high viscosity. The formation of covalent bonds between the single polymer chains is the reason of the hygroscopic behavior of PAA and application as a superabsorbent polymer.[22]

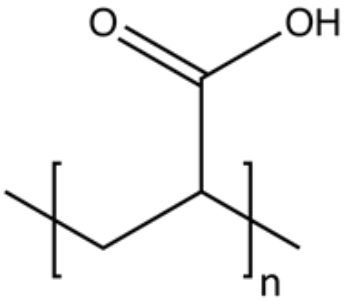


Figure 2 Repeating unit of protonated polyacrylic acid

With the increase in pH of the solution, the conformation of PAA transits from a coiled to an extended conformation due to the electrostatic repulsion of carboxylate ions.[23-30]. This is due the rise in degree of dissociation. The structure of PAA chains is also strongly influenced by the presence of divalent ions like calcium[31-40], copper[35], lead[41], silver[42] and strontium[37, 43] ions. Compared to monovalent ions, divalent ions interact much stronger with the carboxylate ions of PAA[33, 40], which screens the intermolecular electrostatic repulsion of the carboxylate groups. Thus "like-charge attractions"[31, 36] can be observed, which induce coil shrinking of the PAA chain[32, 35, 37-40]. If the number of divalent ions in the solution exceeds the binding capacity of PAA, insoluble PAA salts are formed.[38, 39]

Polyacrylic acid is a widely studied polyelectrolyte in the literature, using both experimental[23, 27, 33, 44-46] and simulation[23-25, 28, 30, 47] techniques. Biermann *et al.*[25] simulated one atactic oligomer strand with 23 repeating units in simple point charge (SPC) water with GROMOS96 force field. The carboxylic acid groups were fully deprotonated, 23 sodium ions were added as counter ions. The system was equilibrated at 333.15 K and 1 atm for 1 ns and simulated for about 4.5 ns. Following the approach of Biermann *et al.*, Reith *et al.*[28] studied fully deprotonated, atactic PAA with 8, 12 and 23 monomers with atomistic and

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

coarse-grained model using GROMOS force field. Most recently, Sulatha *et al.*[30] investigated difference in structural behavior of polyacrylic acid and poly(methacrylic acid) by MD simulations. They focused on atactic PAA chain with 20 monomer units and different degree of deprotonation. The MD simulations were carried out for 15 ns at 300 K and 1 atm with two different force fields, one based on the GROMOS53a6 parameter set (FF-1), the other one based on ENCAD parameter set (FF-2). Tjipangandjara *et al.*[48] presented experimental results by studying the conformational characteristic of PAA adsorbed on alumina by absorbing 10g of mineral in 194ml of 0.03 k mol/l^3 NaCl solution. Using fluorescence spectroscopy, it was observed polymeric species conformation are a major controlling factor for the determination of the stability factor in a suspension consisting of alumina and kaolinite. pH shifting offered a new means of controlling the conformation of the polymer in the absorbed state. PAA conformation changed from coiled to stretched due to increase in pH.

2.1.2 Poly Glutamic Acid

Poly glutamic acid (PGA) is naturally occurring anionic polymer that is water-soluble, biodegradable, edible and non-toxic towards human and environment biopolymer produced by Bacillus subtilis. It has multifarious potential applications in foods, pharmaceuticals, healthcare, water treatment and other fields. The production of PGA has already been established on the industrial scale. Various studies regarding the fermentative production, downstream processing and characterization of PGA have been reported in the literature. Hence, it has been suggested to be a good candidate for various industrial applications including thickener, bitterness reliving agent, cryoprotectant, sustained release material, drug carrier, curable biological adhesive, biodegradable fibres, highly water absorbable hydrogels, biopolymer flocculants, and heavy metal absorbers.[49] Apart of being a major constituent of the traditional Japanese food, 'natto', PGA also has industrial use in the bioremediation of contaminated water, in drug delivery systems[50] for cancer treatment, tissue engineering[51], biological adhesives for tissue adhesion, cosmetics, agriculture and as a functional food ingredient.[52-54]

Polyglutamic acid (PGA) exists in two different conformations called α-PGA and γ-PGA (see Fig. 3). γ-PGA is a biopolymer formed by bacterial fermentation and therefore biodegradable and non-toxic. Polyglutamic acid are good biodegradable polycarboxylates and possess the required functionality to be considered as a substitute for a builder system in a detergent.

Originally, pentasodium triphosphate (STP) was used however, due to its eco-toxicity the usage has been eliminated in the last 20 years. The application of polycarboxylates, like acrylic acid homo- and copolymers, displays advantages over STP due to their ability to bind calcium ions and re-dissolve calcium salt precipitates.[55] However, poly(acrylic acids) are not biodegradable and therefore enrich in the oceans when they are used as builders in detergents. To avoid further accumulation, poly(amino acids) with free carboxylic groups can be used, such as poly(aspartic acid) and poly(glutamic acid), which have both good biodegradability and necessary functionality similar to polyacrylic acid[56].

Poly-γ-glutamic acid (γ-PGA) is a naturally occurring biopolymer made up of repeating units of l-glutamic acid, d-glutamic acid or both. In pure water, when pH is increased, this macromolecule changes from right handed α-helix to conformations compatible with other secondary structures like left-handed polyproline II (pPII) helix.[57] α-PGA is only accessible synthetically by nucleophile-initiated polymerization of the C-protected N-carboxy-anhydride of L-glutamic acid.[53]

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

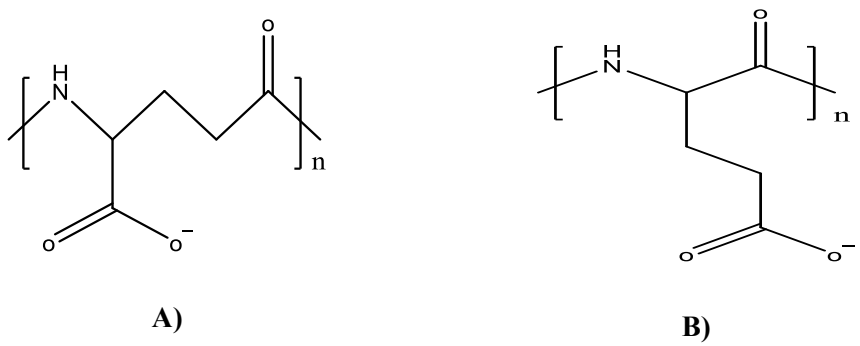


Figure 3 Repeating units of deprotonated poly (glutamic acid) (PGA) **a)** γ -PGA and **b)** α -PGA

PGA chains can form highly ordered structures like, α-helices and β-sheets, but also random coil conformations, depending on the solvent conditions.[58-60] The two conformations of PGA have been studied under various solvent conditions using both experimental[60-63] and theoretical[57, 64-67] methods. With increasing pH, corresponding to a higher degree of ionization, PGA changes from □-helical to extended structures.[59, 68-70] Moreover, circular dichroism measurements have shown that the presence of monovalent and divalent cations drastically increases the helical content of highly deprotonated PGA chains[59, 69, 71, 72], which has been verified in recent years using all-atom molecular dynamics (AAMD) simulations[57, 64, 65].

Aleman *et al.*[66] performed MD simulations of 20 – and 10 – residue chains of un-ionized PGDGA at different temperatures. $PGDGA_{10}$ at both 300 and 350 K provided very similar results. In both cases the initial helical conformation quickly transforms into a random coil. For $PGDGA_{20,}$ value of RMSD at 350 K increased considerably with respect to that of the same compound at 300 K. The results indicated the stability of the helical conformation of PGDGA is dependent on the polypeptide chain. Further, Marchand *et al.*[57] performed MD analysis to simulate the structural dynamics of PGA in different solvents. Qualitatively different results were obtained from modifying the Lennard jones (LJ) parameters in the standard pair-potentials of the solvents. Investigated parameter set led to hydration structures for the Na^+ ions. No major structural changes were obtained using equally standard parameters for solvents like KCl solution and water.

Zanuy *et al.*[73] investigated helical γ-PGA for the *α*-helix structure through studying the conformation of the structure. Considerable differences were revealed depending on the degree of ionization of the polymer. The un-ionized acid was reported to poses a helical conformation whereas ionized salt behaved like a random coil state. Zanuy *et al.*[66] performed molecular dynamics simulation in aqueous solution to investigate the effect of the chain length and temperature on the helical conformation of un-ionized poly(γ-d-glutamic acid). Results presented revealed the helix is not stable for small number of the residues and is independent of the temperature. Temperature induced a conformational transition from helical to random coil.

Ogasawara *et al.*[74] Studied the unfolding dynamics of a 20-residue PGA having protonated side chains in an acidic environment using an α-helical conformation as the initial structure. Simulation were performed for various time period observing various pathways of unfolding

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

process. PGA unfolding was majorly seen to be initiated by denaturation of the termini, followed by propagation of the coil conformation toward the opposite side.

2.2 METHODOLOGY

Molecular modelling is the process of defining chemical systems in terms of a realistic atomic model, with the aim of understanding & predicting macroscopic properties based on available descriptive knowledge at the atomic scale. It is vividly used for designing of new materials necessitating physical properties of realistic systems to be predicted precisely.

Physical properties at macroscopic level are distinguished by static equilibrium properties[75] compromising and dynamic or non-equilibrium properties[76]. The following chapter shortly explains the theory behind MD simulations and difference between all-atomistic and coarse-grained methods. The MARTINI force field including the new polarizable water model is introduced.

2.2.1 Molecular Dynamics (MD) Simulations

Molecular dynamics was originally designed to study relaxations accompanying various non-equilibrium phenomena of hard sphere within the theoretical physics by Alder and Wainwright in 1950s. In today's fast-growing world of science, molecular dynamics plays a very crucial role in various fields of chemistry, physics, material science and biophysics.

In layman's term, molecular dynamic is computer simulation technique used to study the physical movements of atoms and molecules. In molecular dynamics atoms are typically represented as single point masses inside van der walls potentials and what were the 6-12 potentials typically used to represent such things. This means that they come out mostly as hard spheres. Partial charges are then calculated by some form of quantum method to represent the distribution of electronic charge on those molecules for polar molecule like water. These partial charges can be quite large with the best part of a full formal negative charge in the oxygen atom. In molecular dynamics, bonds are typically represented as a simple harmonic oscillator as are the angle constraints.

To start the simulation a range of velocities is chosen that matches the Boltzmann distribution[77] for that temperature and then those velocities are assigned randomly to each atom. Similar methods are then used to put energy into the bonds, angles and dihedrals. The simulation is then run according to the Newtonian physics with the molecules transferring both energy and momentum to one another via electrostatics and van der walls interactions. In these simulations water is essentially a hard sphere with three-point charges in it. Surprisingly such a simple method for modelling these systems gives surprisingly good results and numerous physical properties that can be calculated maybe the best of these is the structure. If we plot the average probability of finding an oxygen in the unit volume with the distance out from a specific eater molecule, we get what is called a radial distribution function.

It turns out these structural functions can be measured directly by experimental methods such as neutron scattering[78]. These simple methods do very reliable job of modelling the structure of water. We can also do this with other solids in water. One molecular dynamics of relatively simple systems like water can be fairly successful it really is the more complicated system such as protein substrate interactions and how species translocate across

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

membranes and so forth that are the real pertinent questions in biochemistry and while the simulations are fantastic way of getting insights into these problems and visualizing the answers we get will only be as good as the parameterization of the model, its 'garbage in garbage out' and that's the ever-present caveat associated with molecular dynamics.

While molecular dynamics is a very solid player when it comes to getting insight into the behaviour of such systems it turns out in many biological systems the devil is in the details. The refinement of these details is critical to the development of this field and with the appropriate structural measurements usually from neutron scattering we can refine the details of the model and subsequently elucidate which are the important interaction that are critical in many biological systems.

Molecular dynamics simulates the natural motion of the molecular system. The atoms and molecules interact for a given period providing the dynamic transformation of the system. This transformation is analysed by analysing the trajectory of the atoms and molecules which is determined by numerically solving classical Newton's equation of motion for a group of atoms.

$$F_i = m_i a_i \qquad \text{Eq. 1}$$

Where.

i = 1……. N

F_i: Force on atom/molecule i

m_i: mass of atom/molecule i

a_i: acceleration of atom i.

Using the second derivate of the atomic position r_i with respect to time for the acceleration leads to

$$F_i = m_i \frac{\partial^2 r_i}{\partial t^2} \qquad \text{Eq. 2}$$

The force acting on every atom, their initial position and velocity need to be defined in order to integrate the equation of motion (Eq.). Atomic force fields are used to describe the forces on every atom based on the negative derivate of the potential energy functions

$$F_i = -\frac{\partial V_i}{\partial r_i} \qquad \text{Eq. 3}$$

where V_i represents the potential energy of *N* interacting atoms as a function of their positions $r_i = (x_i, y_i, z_i)$. A typical All Atomistic (AA) force field for biomolecules uses potential energy functions for non-bonded and bonded interactions between atom pairs r_{ij}.

$$V_i = V_{non-bonded}(r_{ij}) + V_{bonded}(r_{ij}) \qquad \text{Eq. 4}$$

$$V_i = \sum_{\text{atom pairs}} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{\text{atom pairs}} k\frac{q_i q_j}{r_{ij}} + \sum_{\text{bonds}} \frac{a_i}{2}(l_i - l_{i0})^2 + \sum_{\text{angles}} \frac{b_i}{2}(\theta_i - \theta_{i0})^2 + \sum_{\text{torsions}} \frac{c_i}{2}[1 + \cos(n\omega_i - \gamma_i)] \quad \text{Eq. 5}$$

The first 2 terms in Eq. 5 (taken from Meller[79]) sum up all non-bonded interactions between all pairs of atom separated by distance $r_{ij} = |r_i - r_j|$. The last three terms are used for the summation of bonded interactions defined by bonds, angles and torsion angles defining the covalent structure of the system. The first term describes the repulsive and attractive interatomic forces in form of 12-6 Lennard-Jones potential with the van-der-Waals parameters ε_{ij} and σ_{ij} The second term gives the Coulomb interaction between two atoms with partial charges q_i and q_j. The energies of the deformation of bond length l_i and bond angles θ_i from their respective equilibrium values l_{i0} and θ_{i0} are describe in term 3 and 4. The last term describes rotations around the chemical bond, which are characterized by periodic energy terms.

Most available force fields are designed for biomolecules like proteins and not for polymers, therefore the potential energy functions are missing. To investigate the interaction between polymers and the ions each polymer building block has to be parameterized according to the chosen force field, herein AMBER99SB[80] and MARTINI.

2.2.2 MARTINI Force Field

The MARTINI force field, developed by Marrink *et al.*[81], was originally designed for the simulation of small lipids. Nowadays, it is extended to a variety of molecules, lipids like cholesterol, proteins and polymers.[82-84]

The MARTINI fore field is not parameterized to match structural details perfectly, instead it aims for a broader range of application without the need to reparametrize the model every time. Therefore, chemical building blocks were calibrated against thermodynamic data, in particular oil/water partitioning coefficients.[85]

In accordance with the MARTINI philosophy of simplicity, usually a four-to-one mapping scheme is applied[86], i.e. four heavy atoms are represented by one MARTINI bead (see Fig. 4). To represent the geometry of small groups and ring-like structures more precisely, a mapping of resolution of two or three atoms to CG bead is applied (S type beads).[81] The same applies to water and ions: four water molecules are represented by one CG bead and also ions in solution are mapped to one bead including their first hydration shell.

MARTINI force field defines four main types of interaction sites: polar (P), non-polar (N), apolar (C), and charged (Q). The hydrogen bonding capability of each main type is denoted by a letter: donor (d), acceptor (a), both (da) and none (0). Furthermore, subtype numbers are used to indicate the polarity from 1 (low polarity) to 5 (high polarity).

The non-bonded interactions of pairs *i* and *j* at distance r_{ij} are described via a 12-6 Lennard-Jones (LJ) potential

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

$$V_{LJ} = 4\varepsilon_{ij}\left[(\sigma_{ij})^{12} - (\sigma/r_{ij})^{6}\right] \qquad \text{Eq. 6}$$

where ε_{ij} is the well-depth and determines the strength of the interaction. The value of ε_{ij} depends on the bead type and ranges from ε_{ij} = 5.6 kJ/mol for interactions of strongly polar groups to ε_{ij} = 2.0 kJ/mol for interactions between apolar and polar groups. The effective size of the particles is defined by LJ parameter σ = 0.47 nm for normal and σ = 0.43 nm for smaller bead types and ring-like structures (S bead type). Additionally, charged beds (Q type) interact via Coulombic energy function using a relative dielectric constant of ε_r = 15 for explicit background screening:

$$V_{el} = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}} \qquad \text{Eq. 7}$$

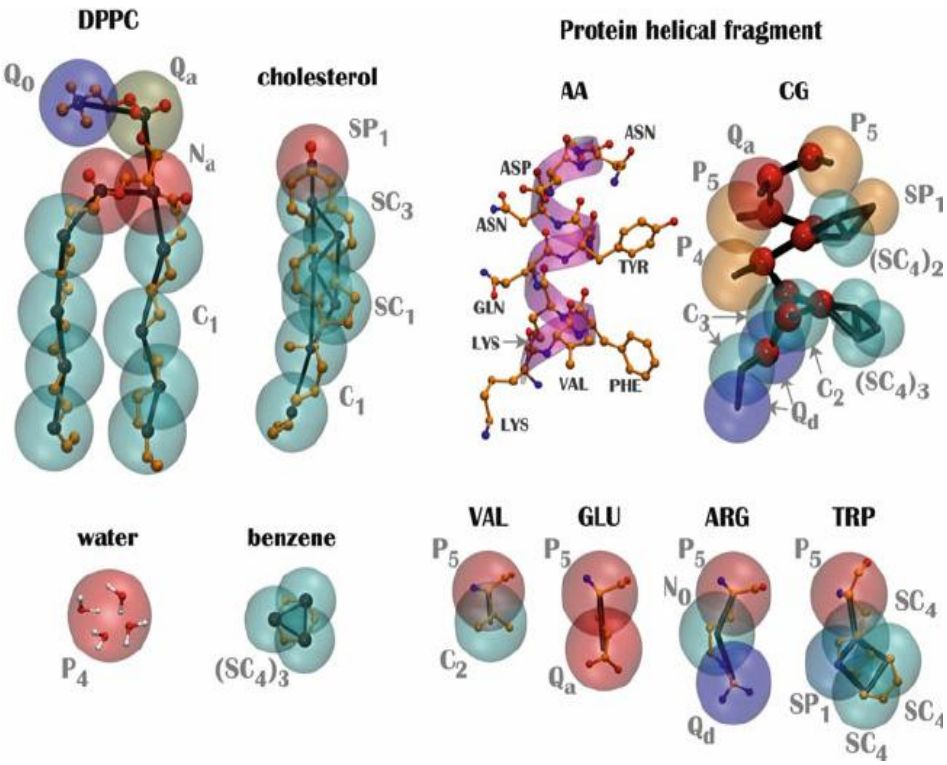


Figure 4 Examples for MARTINI mapping from AA level of chemical structure to the CG model. The CG beads are shown as transparent van-der-Waals spheres. The different types of CG bead types determine their interaction level. The prefix S denotes to ring-like structures or other smaller groups, which where mapped three/two-to-one. Picture take from Periole *et al.*74

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

3 COMPUTATIONAL METHODS

The following section describes the computational methods and programs used for the force field parameterization of building blocks, MD simulation and trajectory analysis of PAA and Poly-γ-l-glutamic acid (γ-L-PGA) MD simulations.

3.1 ATOMISTIC SYSTEMS

3.1.1 Model Generation

Polymers like PAA have different properties in dependence of the degree of dissociation. Whether the carboxylic acid group in the side chain of PAA is protonated (see Fig. 5a) or deprotonated (see Fig.5b) depends on the pH of the solution. According to the Henderson-Hasselbalch-Equation (See Eq. 8) the degree of dissociation at a given pH can be calculated.

$$pH = pKa + \log\frac{c(A^-)}{c(HA)} \qquad \text{Eq. 8}$$

With substitution of $c(A^-)$ by the dissociation degree α of the carboxylic acid groups follows

$$pH = pK_a + \log\frac{\alpha}{1-\alpha} \qquad \text{Eq. 9}$$

$$\alpha = (1 + 10^{pk_a - pH})^{-1} \qquad \text{Eq. 10}$$

The dissociation of monomer acrylic acid as a model system has been well studied and a pK_a value of 4.756 has been determined. Nevertheless, it is known that the electrostatic repulsion along the polymer chain has a significant effect on the acid-base properties in polymers and therefore it is expected that the pK_a value depends on the chain length. Laguecir *et al.*[27] studied the dissociation behaviour of PAA with different molecular weights by potentiometric titration[87]. They investigated that high molecular weight PAA is more difficult to deprotonate due to the increasing total electrostatic potential on each monomer. For PAA with a chain length of 25 monomers the experimental pK_a value was 5.98. Based

a) b)

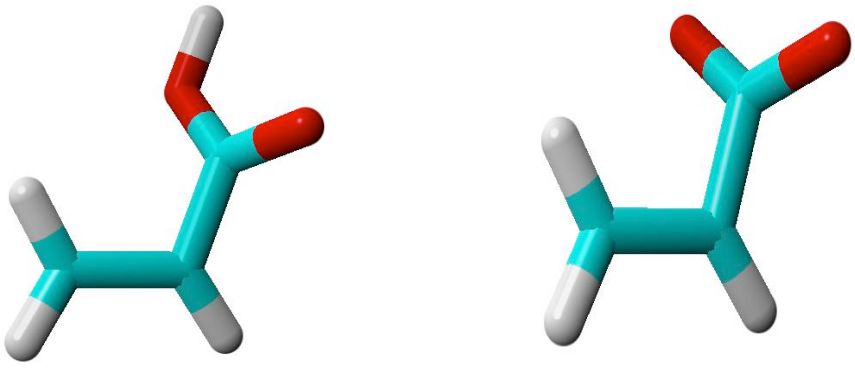


Figure 5 Protonated form of acrylic acid **a)** and deprotonated form of acrylic acid monomer unit **b)** as stick model with carbon in green, hydrogen in with and oxygen in red

on this value, the dissociation degree of our 20-mer model of PAA is 0.999 at pH 9, which means PAA is fully deprotonated.

Polybuild is used to build polymer chains of different chain lengths, weight or tacticity (see appendix 9.1). Homopolymers as well as copolymers with a statistic distribution (equiprobable or Bernoulli) are possible. The generated structure can be optimized using TINKER. A download link for polybuild is provided at *http://chembytes.wikidot.com/polybuild*

The experimentally used PAA (ACUSOL445N) has a molecular weight of 4500 Da that equals a polymer chain length of 62 repeating units. At pH 9, PAA has a dissociation degree of 0.999 according to the Henderson-Hasselbalch-Equation[88], which means PAA is fully deprotonated. In order to save computational time and to compare the results with previous studies, a model system must be created. Both PAA and γ-L-PGA structural models are generated with polybuild using 20 & 62 repeating units and a syndiotactic structure. The smile format output obtained from polybuild is converted to *.pdb* format by using YASARA Structure version 16.3.8[89].

In order to create a polymer with variable chain length and tacticity polybuild is used. A modified library *libpolymers.xml* that contains the structure of PAA, and γ-L-PGA is created. In addition to the default polymers the new polybuild library included

- protonated PAA isotactic and syndiotactic (paa-iso and paa-syndio)
- deprotonated PAA isotactic and syndiotactic (pab-iso and pab-syndio)
- mixture of protonated and deprotonated PAA syndiotatic (paapab-syndio)
- protonated α -PGA syndiotactic (alpha-pga-syndio)
- deprotonated α-PGA syndiotactic (alpha-pgb-syndio)
- deprotonated γ-PGA syndiotactic (gamma-pgb-syndio)
- mixture of protonated and deprotonated α-PGA syndiotatic (alpha-pgapgb-syndio)

To create fully deprotonated PAA with a chain length of 20 polybuild must be executed by entering

python polybuild –p pab–syndio –l 20 –f mm2 --smilesonly

–p	name of the polymer	*–f*	force field (always mm2)
–l	length of polymer chain	*--smilesonly*	only *.smiles* file as output

Polybuild creates a file called *pab-syndio.smiles* that contains the polymer structure in smiles code. The library entry for deprotonated PAA can be found in the appendix 9.2.

To simulate the polymer with GROMACS the smiles file must be converted into a .pdb file by opening YASARA and loading the file pab-syndio.smiles (file > load > other file format). Chose smiles as format and select the correct file. Once YASARA loaded the structure, it can be saved as .pdb file (file > save as > PDB file).

Since the .pdb file for PAA was created manually from a smiles file, YASARA names all residues as UNK for unknown residue type. This means that there is no residue name (ResName) and number (ResID) in the .pdb of polymer by default. In order to rename the .pdb file of protonated or deprotonated 20-mer of PAA we can use the script rename_PAA.py. The script can be found in the appendix 9.3.

We copy the file to our working directory, open it and edit the entry in line 5 to define the .pdb file that must be renamed. In order to rename all the PAA residues in a .pdb file

according to the residue names in the AMBER99SB force field, execute rename_PAA.py by entering

python rename_PAA.py

The script will create a copy of the original file and a the renamed .pdb file (e.g. paa_renamed.pdb)

3.1.2 Polymer Parameterization

The geometry optimization and charge parameterization are carried out with Gauss View version 5.0 and Gaussian09[90]. Density functional theory with B3LYP function[91-93] and 6-31G(d,p)[94] basis set is used for the geometry optimization for PAA. The 20-mer structural model of PAA is split up into head group, tail group and repeating units, all caring one negative charge. Head and tail group are capped with one methyl group, the building block with two methyl groups. The charge parameterization is performed at Hartree-Fock level of theory with 6-31G(d) basis set[95, 96] to calculate the electrostatic potential (ESP). The antechamber software package[97, 98] is used to get the AMBER99 compatible RESP[99-101] charges.

The ACPYPE[102] software is used to get a GROMACS[103] (Groningen Machine for Chemical Simulations) compatible topology for PAA with AMBER99 atom types.

3.1.3 Adding Polymer Residues (Building Blocks) To GROMACS Topology Library

As AMBER99Sb is designed for amino acids there is no entry in the GROMACS library for polyelectrolytes like PAA. However, it is possible to simulate PAA within AMBER99SB[80] but therefore PAA has to be parameterized in a manner consistent with how the rest of the force field was originally derived. This means that all missing parameters for bonds, charges and dihedrals must be added to the *aminoacids.rtp*, *ffbonded.itp*, *ffnonbonded.itp* and *atomtypes.itp* files of AMBER99SB. To get this information we must parameterize PAA manually. The charge parameterization consists of 3 steps. The first step is to optimize the geometry of the monomer, which is needed to generate an electrostatic potential (ESP) and finally get the restrained electrostatic potential (RESP) charges.

3.1.4 Geometry Optimization

Since the building blocks of polymers are connected to the next one, the simple monomer needs to be modified with regard to these two extra bonds in the polymer. Therefore, the monomer is capped with a suitable group. For the deprotonated acrylic acid, a simple methyl group is used for both ends (see Fig.6).

The *.pdb* file can be easily created by modifying the *.pdb* of the monomer with YASARA. Afterwards, the modified *.pdb* file need to be opened with Gaussian and saved as a Gaussian input file (*.com*). This file has to be adjusted by adding the following lines, which define the method of the geometry optimization, to the head of the file

```
%chk=PAA_Meth.chk
%mem=5000mb
```

%nprocshared=4

#P opt b3lyp/6-31g** geom=connectivity int=grid=ultrafine scf=tight

Title Card Required

-1 1

The -1 is for the charge, as deprotonated PAA the building block is negatively charged. Protonated PAA is neutral and therefore the line would start with 0.

The geometry optimization of the capped PAA will be performed at B3LYP/6-31G** level of theory by executing

g09 <PAA_Meth.com> PAA_Meth.out

3.1.5 Calculation of ESP Charges

The Gaussian output file *PAA_Meth.out* is used to get a new input file for the ESP charge calculation:

gv PAA_Meth.out

Gaussian will open the newly created file and we can save it as a new input file (*PAA_Meth_esp.com*). Again, this file must be adjusted by adding the following lines to the head of the file

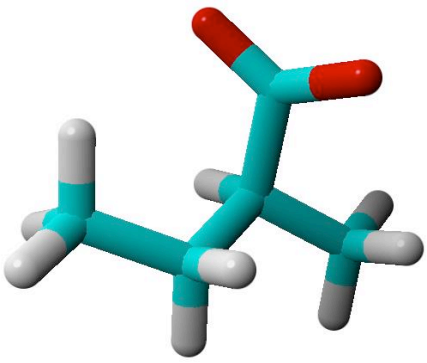


Figure 6 Stick model of deprotonated acrylic acid monomer with methyl caps on both sides. Carbon atoms are shown in green, oxygen atoms in red and hydrogen atoms in white.

%chk=PAA_Meth_esp.chk

%mem=5000MB

%nprocshared=4

#p hf/6-31g* geom=connectivity scf=(tight,maxcycle=800) pop=mk iop(6/33=2,6/41=10,6/42=17)

Title Card Required

-1 1

By entering,

g09 <PAA_Meth_esp.com> PAA_Meth_esp.out

The calculation of the electrostatic potential at HF/6-31G* level of theory will be executed.

3.1.6 Calculation of RESP Charges

In order to calculate the RESP charges with antechamber we need to covert the Gaussian output file into *.ac* format with

antechamber –i PAA_Methyl_esp.out –fi gout –o PAA_Meth_esp.ac –fo ac

As *resp* only reads ESP data in specific file format we need to extract the ESP data from the Gaussian output file and convert them into *.esp* file, which has the RESP input format.

espgen –i PAA_Meth_esp.out –o PAA_Meth.esp

The RESP fit is usually carried out in two stages. During the first stage all atoms are allowed to vary. However, in the second stage all degenerated hydrogens are constrained to have equal charge. Therefore, two input files are needed which have to generated first:

respgen –i PAA_Meth_esp.ac –o PAA_Meth.respin1 –f resp1

respgen –i PAA_Meth_esp.ac –o PAA_Meth.respin2 –f resp2

Before running the 2-stage RESP fit the generated input files have to be adjusted because we have to make sure that PAA will get the right charge. Open the file *PAA_Meth.respin1*. It consists of two parts. The first part looks like this

Resp charges for organic molecule

&cntrl

nmol = 1,
ihfree = 1,
ioutopt = 1,
qwt = 0.00050,

&end

This part defines the number of molecules (*nmol*), hydrogen restraints (*ihfree*), restart info for ESP charges (*ioutopt*) and RESP stage number including force field information (*qwt*). For further information concerning this part see antechamber manual (http://upjv.q4md-forcefieldtools.org/RED/resp/)

In the second part all atoms of PAA are specified:

1.0

Resp charges for organic molecule

-1 16

6 0

1 0

1 0

6 0

1 0

1 0

1 0

6 0

1 0

6 0

1 0

1 0

1 0

6 0

8 0

8 15

The first value (1.0) is a weight factor for multi-conformational fits. Since we only have one optimized structure this does not have to be modified for our molecule. The next line ('RESP charges….') is a subtitle for the structure followed by a line specifying the total charge and the number of atoms in the molecule (-1 16). Each of the following lines characterizes one atom in the same order as in the *.ac* or *.pdb* file. The first column is the atomic number, and second column is the fitting number N. N can either be -99, 0 or a positive integer. If N = -99 then *resp* will read the charge from a *.qin* file and constrain to this value. For N = 0 the charge is free to vary during the fitting. With a positive N two atoms are specified to have the same charge, which is used in the second stage fitting to constrain all generated atoms. This means if we want atom 3 to have the same charge as atom 2 we set the N value for atom 3 to 2. This will not fix the charge to a certain value, but both atom charges will vary as a pair.

As the charge of the PAA building block without cap has to be -1, we have to define the charge of the caps. We don't know the exact charge distribution on the methyl groups, but the charge should sum up to 0. Because of this, we don't have to use an additional *.qin* file but we have to add an additional section to *PAA_Meth.respin1* and *PAA_Meth.respin2*

4 0.000000

1 4 1 5 1 6 1 7

4 0.000000

1 10 1 11 1 12 1 13

This extra part defines the group constraints. While the first line determines the number of atoms (4) and the charge (0) they must sum to, the second line specifies the molecule number (1) and atom number of each atom (4, 5, 6,7) involved in the group.

Note: Not more than 16 values per line!

Then we can run the first stage of the RESP fit

resp –O –i PAA_Meth.respin1 –o PAA_Meth.respout1 –e PAA_Meth.esp –t qout_stage1

Note: Always check qout_stage1 if the values seem reasonable

Before we can start the second stage of the fitting, we have to open the *PAA_Meth.respin2* file and add the section for the group constraints. The first part of *PAA_Meth.respin2* looks like this

```
Resp charges for organic molecule
 &cntrl


 nmol = 1,
 ihfree = 1,
 ioutopt = 1,
 iqopt = 2,
 qwt =  0.00100,
```

Here we see a slight difference to the input file for the first stage fitting, the extra line ‘iqopt = 2,’ is added. This means that the RESP fit will use the charges of the marked atoms from an additional input file, which is *qout_stage1*.

```
 &end
  1.0
 Resp charges for organic molecule
  -1   16
   6    0
   1    0
   1    2
   6    0
   1    0
   1    5
   1    5
   6  -99
   1  -99
   6    0
   1    0
   1   11
```

1 11

6 -99

8 -99

8 -99

As mentioned above, for all atoms with fitting number N = -99 the charge will be read from a *.qin* file which in this case is *qout_stage1*. This concerns all non-methyl carbons and hydrogens.

To run the second stage of the RESP fit we need to execute:

resp –O –i PAA_Meth.respin2 –o PAA_Meth.respout2 –e PAA_Meth.esp –q qout_stage1 –t qout_stage2

In order to generate a mol*2* output file (other types are possible as well) with the calculated RESP charges we use

antechamber –i PAA_Meth_esp.ac –fi ac –o PAA_Meth_resp.mol2 –fo mol2 –c rc –cf qout_stage2 –a amber

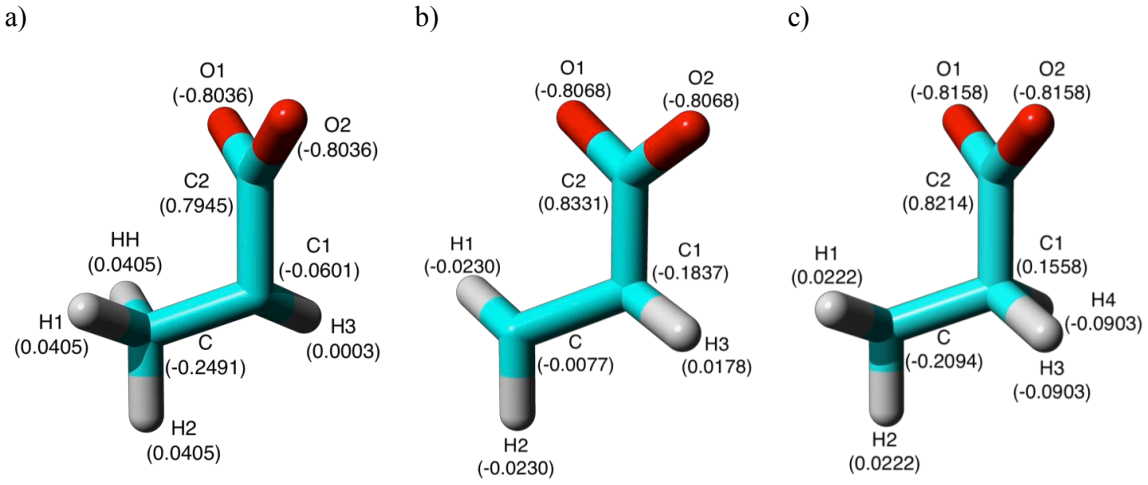


Figure 7 RESP charges of PAA head group (a), repeating unit (b) and tail group (c) with corresponding atom names. PAA is shown as stick model with carbon colored in cyan, hydrogen in white and oxygen in red. Equivalent atoms have the same charge

The whole procedure has to be repeated for the head and the tail group of PAA capped with methyl groups. In Fig.7, the obtained RESP charges for PAA head group, repeating unit and tail group without capping groups are shown with the corresponding atom names.

3.1.7 Creating GROMACS Topology for Capped Groups and Conversion of *.rtp*

Now we have the RESP charges, but we still need the topology in GROMACS format. This can be done by using ACPYPE

acpype.py –i PAA_Meth_resp.mol2 –n -1 –a amber –c user –n net charge –c charges defined by user –a atom type amber (AMBER99SB)

After successful execution, ACPYPE creates a folder that contains several files in different formats for the various MD programs. We can find the correct GROMACS topology in

PAA_Meth_resp_GMX.itp. The first part if this file defines the atom types of PAA repeating unit:

;name	bond_types	mass	charge	ptype	Sigma	epsilon		Amb	
CT	CT	**0.00000**	**0.00000**	**A**	**3.39967e-01**	**4.57730e-01**	;	1.91	0.1094
HC	HC	**0.00000**	**0.00000**	**A**	**2.64953e-01**	**6.56888e-02**	;	1.49	0.0457
C	C	**0.00000**	**0.00000**	**A**	**3.39967e-01**	**3.59824e-01**	;	1.91	0.0860
O2	O2	**0.00000**	**0.00000**	**A**	**2.95992e-01**	**8.78640e-01**	;	1.66	0.2100

Open *ffnonbnded.itp* and *atomtypes.itp*, which can be found among the force field files. We have to make sure that the correct values for every atom type (here: CT, HC, C, O2) exist in both files. As ACPYPE uses the AMBER99SB atom types for every know atom, there is nothing to do here for PAA. Note that if we want to simulate an 'exotic species' like metal ions, the required values have to be added to the list. To keep the right format, the correct mass and atom type number for every new atom is necessary.

The second part of PAA_Meth_resp_GMX.itp defines atoms and charges of the repeating unit of deprotonated PAA:

[atoms]

;nr	type	resi	res	atom	cgnr	charge	Mass	;	qtot	bond_type
1	**CT**	1	PAB	**C1**	**1**	**-0.007678**	12.01000	;	qtot	-0.008
2	**HC**	1	PAB	**H1**	**2**	**-0.022983**	1.00800	;	qtot	-0.031
3	**HC**	1	PAB	**H2**	**3**	**-0.022983**	1.00800	;	qtot	-0.054
4	CT	1	PAB	C2	4	0.178790	12.01000	;	qtot	0.125
5	HC	1	PAB	H3	5	-0.059597	1.00800	;	qtot	0.066
6	HC	1	PAB	H4	6	-0.059597	1.00800	;	qtot	0.006
7	HC	1	PAB	H5	7	-0.059597	1.00800	;	qtot	-0.054
8	**CT**	1	PAB	**C3**	**8**	**-0.183658**	12.01000	;	qtot	-0.237
9	**HC**	1	PAB	**H6**	**9**	**0.017789**	1.00800	;	qtot	-0.220
10	CT	1	PAB	C4	10	0.198342	12.01000	;	qtot	-0.021
11	HC	1	PAB	H7	11	-0.066114	1.00800	;	qtot	-0.087
12	HC	1	PAB	H8	12	-0.066114	1.00800	;	qtot	-0.153
13	HC	1	PAB	H9	13	-0.066114	1.00800	;	qtot	-0.220
14	**C**	1	PAB	**C5**	**14**	**0.833072**	12.01000	;	qtot	0.614
15	**O2**	1	PAB	**O1**	**15**	**-0.806779**	16.00000	;	qtot	-0.193
16	**O2**	1	PAB	**O2**	**16**	**-0.806779**	16.00000	;	qtot	-1.000

Only the atoms highlighted in bold belong to the repeating unit, the others are the capping groups, which are not needed anymore. For the deprotonated PAA a new residue type called PAB must be defined in aminoacids.rtp

The new entry should correspond to the previous ones in the file. The following part is added to the end of the aminoacids.rtp file:

```
[PAB]
 [atoms]
     C       CT      -0.007678       1
     H1      HC      -0.022983       2
     H2      HC      -0.022983       3
     C1      CT      -0.183658       4
     C2      C        0.833072       5
     H3      HC       0.017789       6
     O1      O2      -0.806779       7
     O2      O2      -0.806779       8
```

The next section of PAA_Meth_resp_GMX.itp defines the bond and bond types:

[bonds]

;ai	aj	funct	R	K		
1	2	1	1.09E+03	2.85E+09	;	**C1 - H1**
1	3	1	1.09E+03	2.85E+09	;	**C1 - H2**
1	4	1	1.53E+03	2.59E+09	;	C1 - C2
1	8	1	1.53E+03	2.59E+09	;	**C1 - C3**
4	5	1	1.09E+03	2.85E+09	;	C2 - H3
4	6	1	1.09E+03	2.85E+09	;	C2 - H4
4	7	1	1.09E+03	2.85E+09	;	C2 - H5
8	9	1	1.09E+03	2.85E+09	;	**C3 - H6**
8	10	1	1.53E+03	2.59E+09	;	C3 - C4
8	14	1	1.52E+03	2.65E+09	;	**C3 - C5**
10	11	1	1.09E+03	2.85E+09	;	C4 - H7
10	12	1	1.09E+03	2.85E+09	;	C4 - H8
10	13	1	1.09E+03	2.85E+09	;	C4 - H9
14	15	1	1.25E+03	5.49E+09	;	**C5 - O1**
14	16	1	1.25E+03	5.49E+09	;	**C5 - O2**

The two right columns define the atom connectivity. As this part includes several bonds, which only occur in the capped repeating unit but not in the final polymer the entry for the bonds in the aminoacids.rtp file has to be done mostly by hand. The bonds highlighted in bold define the repeating unit. Additionally, there will be bonds from C1 and C3 to the previous, respectively next building block, which have to be defined in the aminoacids.rtp as well. The section we have to add to aminoacids.rtp should look like this:

[bonds]	
C	-C1
C	H1
C	H2
C	C1
C1	H3
C1	C2
C1	+C
C2	O1
C2	O2

The rest of PAA_Meth_resp_GMX.itp describes bonded interactions, like bond angles and dihedrals. As mentioned before, PAA does not have any unusual bonds, angles or dihedrals. Therefore, it is nothing to do in the *ffbonded.itp* file, all bonded interactions for PAA are already defined. In case of an 'exotic species' new bond, angle and dihedral types have to be added to *ffbonded.itp*.

For the head and tail group of the γ-L-PGA a new entry to the *aminoacids.rtp* file has to be added likewise. In the appendix 9.4 the full entry for deprotonated PAA including head group (PBH), tail group (PBT) and the repeating unit (PAB) can be found.

3.2 COARSE-GRAINED SYSTEMS

The next part focuses on the generation of two polymers MARTINI models and single chain simulations of short and long-chain models of each polymer. Here, the applicability of the MARTINI model for PAA is verified based on other simulation studies[25, 26, 28, 30], experimental studies[38, 39] and the previously performed all-atomistic MD simulations. For the γ-L-PGA model only the all-atomistic MD simulations are available as a reference for

validation. Besides size and flexibility of the polymer coils, the interaction with water and surrounding ions is also determined.

3.2.1 MARTINI Mapping of PAA

In order to build the MARTINI topology of a new molecule, a MARTINI model must be parameterized from target atomistic data. First, a mapping scheme must be established and applied to the target data. Secondly, reference distributions for bonds and angles of the coarse-grained mapped target data must be recorded. Thirdly, a MARTINI topology must be created based on the bond and angle distributions of the previous step. Fourthly, this MARTINI topology is used for a first MD simulation. The bond and angle distributions of this CG simulation have to be recorded and compared to the reference distribution. The final task is to reproduce these distributions. Therefore, it might be necessary to modify the original topology, which leads to an iterative process of MD simulations and comparison of the respective distributions with the reference distributions.

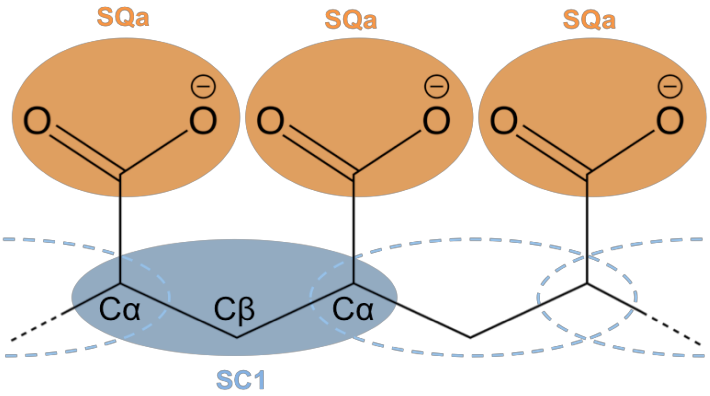


Figure 8 Mapping of PAA atomistic structure to MARTINI beads: Carboxylate groups (COO–) and aliphatic backbone atoms are represented by SQa beads (orange) and SC1 beads (light blue), respectively

Based on the MARTINI philosophy of simplicity, usually a four-to-one mapping scheme is applied[86], i.e. 4 heavy atoms are represented by one MARTINI bead. However, this is not a suitable approach for the mapping of PAA. The first idea was to use a slightly coarser mapping, where each monomer is represented by one bead, as done before by Reith *et al.*[28, 47]. Unfortunately, the non-bonded interactions (here, radius of gyration and end-to-end distance were chosen as a reference) could not be fitted to the results from all-atomistic simulations.

Therefore, another mapping strategy was developed according to the approach published by Rossi *et al.*[20] and Vögele *et al.*[104], who generated MARTINI models for polystyrene and poly(styrene sulfonate), respectively. This second strategy includes two beads per PAA monomer, as illustrated in Fig.8. For the aliphatic polymer backbone one SC1 bead is placed at the centre of mass of Cα-Cβ-Cα, in a position between two consecutive carboxylate groups, as proposed by Rossi *et al.*[20]. The bead type SC1 refers to apolar beads with reduced van-der-Waals diameter of 0.43 nm and low polarity. For the carboxylate groups SQa beads are placed at the centre of mass of each group. According to Vögele *et al.*[104], the SQa type was chosen as these beads carry a negative charge and have hydrogen bond acceptor properties. Additionally, due to the 3-to-1 mapping, a reduced bead size is used for the carboxylate group. As the Cα atoms are shared between neighbouring beads only one bond type (SQa-SC1) must be defined. However, the different structure of the two terminal groups

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

must be considered, resulting in different bond lengths and force constants for the terminal and core SQa-SC1 bonds.

Table 2 Parameters of bonded interactions between SQa (carboxylate group) and SC1 (aliphatic backbone atoms) beads of PAA MARTINI model. Bonds are defined by the equilibrium bond length req and the elastic constant of the harmonic bond potential kbond. Angles are defined by the equilibrium angle θeq and the elastic constant of the harmonic angle potential kθ.

Atoms	Type	Bond	r_{eq} (nm)	k_{bond} (kJ mol^{-1} nm^{-2})	Angle	θ_{eq} (°)	k_{θ} (kJ mol^{-1})
COO	SQa	SQa-SC1 (terminal)	0.298	46000	SQa-SC1-SQa (terminal)	133	210
CCC	SC1	SQa-SC1 (core)	0.273	35000	SQa-SC1-SQa (core)	142	70
					SC1-SQa-SC1	43	250

The same applies for the angles SQa-SC1-SQa, as seen in Table 2. In order to reproduce the reference radius of gyration and end-to-end distance with this MARTINI model, the non-bonded interactions of bonded neighbouring beads are excluded.

3.2.2 MARTINI Mapping of γ-L-PGA

The MARTINI mapping of γ-L-PGA is based on the model published by the 2015 iGEM Team from the University of Groningen[105]. Again, the standard four-to-one mapping is not suitable. Instead, a three-to-one mapping scheme is applied for most of the atoms (see Fig. 9). The MARTINI model provided by the iGEM Team had to be refined to match the previously performed all-atomistic simulations with AMBER99SB parameter set as it is not mentioned which force field they used for their atomistic simulation.

First of all, smaller (S) bead types are used to account for the three-to-one mapping (see Fig. 9). For the repeating units, the neutral bead types SNd and SNa are used to represent the hydrogen bond donor and acceptor properties of polymer backbone. Again, the SQa type is used for the carboxylate group. Considering the unlike chemistry of the head and tail group they have to be mapped in different ways. For the head group, containing the positively charged amino group, SQa, SQd and SNa beads are used to represent the carboxylate ion, the positively charged amino group and the carbonyl group of the peptide bond, respectively.

The tail group is mapped similar to the repeating unit except for the second carboxylate group. Here, a normal sized Qa bead type with 0.47 nm van-der-Waals diameter is used according to the four-to-one mapping.

Furthermore, the bonded values are modified to match the reference bond and angle distributions from all-atomistic simulations. In order to reproduce several narrow bond distributions, these bonds had to be constrained. The final parameters for the bond lengths and angles with the corresponding force constants are shown in Table 3**.** Again, non-bonded interactions are excluded between two beads directly connected by a bond.

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

Table 3 Parameters of bonded interactions between SQd (amino group), SQa/Qa (carboxylate group), SNa and SNd (the last two forming the peptide bond) beads of γ-L-PGA MARTINI model. Constraints are used for narrow bond distributions.

Atoms	Type	Bond	r_{eq} (nm)	k_{bond} (kJ mol^{-1} nm^{-2})	Angle	θ_{eq} (°)	k_{θ} (kJ mol^{-1})
NCC	SQd	SQd-SQa	0.240	constr.	SQa-SQd-SNa	108	50
COO	SQa	SQd-SNa	0.349	4000	SQd-SNa-SNd	140	45
CCO	SNa	SNa-SNd	0.288	33000	SNa-SNd-SQa	100	80
NCHC	SNd	SNd-SQa	0.240	constr.	SNa-SNd-SNa	124	30
CCOO	Qa	SNd-SNa	0.336	7500	SNd-SNa-SNd	121	45
		SNd-Qa	0.356	30500	SQa-SNd-SNa	110	45
					SNa-SNd-Qa	126	45
					SQa-SNd-Qa	135	330

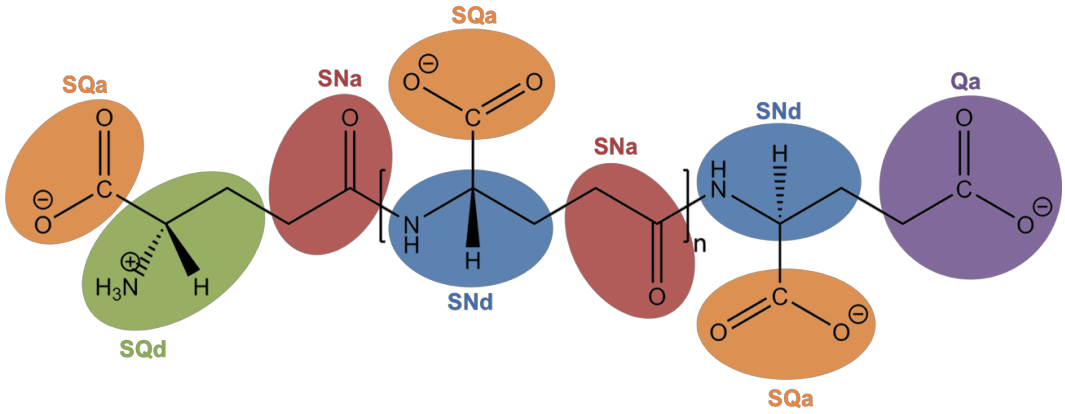


Figure 9 Mapping of γ-L-PGA atomistic structure to MARTINI beads: For the repeating units, the carboxylate groups (COO–) are represented by SQa beads (orange). The peptide bonds are modelled using SNa (red) and SNd (blue) beads. As the head and tail group of γ- L-PGA differ from the repeating units, a slightly different mapping scheme is applied. For the head group bead containing the positively charged amino group (NH3+), an SQd bead (green) and for the last bead of the tail group (including COO– group), a Qa bead (violet) are used

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

3.2.3 Polarizable MARTINI Water Model

In the original MARTINI force field, the water model has no charges. The Coulomb interaction are shifted to zero within 1.2 nm cut off radius and totally neglects long-range electrostatic interactions. Instead a uniform screening is applied using a background dielectric constant of e_r = 15, which is a reasonable approximation for bulk water but seriously underestimates the interactions in highly charged systems. As this work focuses on polyelectrolytes, the classical MARTINI water model cannot be used. Fortunately, a

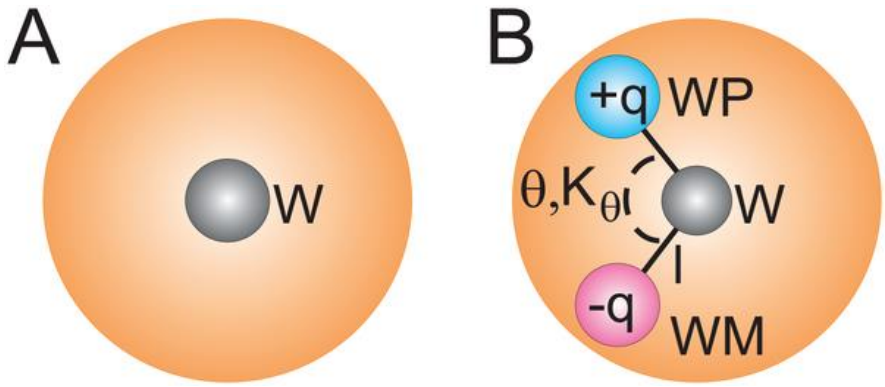


Figure 10 MARTINI water models. a) Standard model; b) polarizable model with LJ interaction site W and Coulomb interaction sites WP and WM. Orange spheres show the van-der-Waals radii. Picture taken from Yesylevskyy *et al*.80

polarizable water model[106] has been published and recently refined[105] for the use with long-range electrostatics.

Compared to the classical water model, two charged interaction sites WP (positive) and WM (negative) are added to the neutral LJ interaction site W (see Fig.10). The added interactions sites only represent dipole effects thus, they do not interact via LJ potential. Moreover, interaction within the same water bead is excluded.

3.3 Simulation Details

All molecular dynamics simulations are performed with GROMACS[107] version 2019.1 within AMBER99SB[80] force field on Elerion and on the cluster using parallelization with OpenMPI[108].

In order to obtain a concentration of 2.5 g/l in MD simulation, a cubic box of 941.192 nm^3 is created for 20-mer PAA and 1728 nm^3 for 20-mer γ-L-PGA with lab water simulations, respectively. Periodic boundary conditions are applied. The SPC/E water model[109] is used to solvate the systems. In order to neutralize the system charge and to simulate a water hardness of 15°dH water molecules are replaced by 20 sodium and 7 chlorine ions are used for the simulation of the system having only NaCl ions. The final system for both polymers with deionized water simulations are presented in Table 4.

Both systems are first energy minimized using steepest descent algorithm to keep the maximum force on any atom in the system to values < 1000 kJ mol^{-1} nm^{-1} and < 10 KJ mol^{-1} for all atomistic and coarse-grained simulation respectively. After heating, the systems are equilibrated for 1 ns and 10 ns for AA and CG simulations respectively at 300 K using Berendsen thermostat[110] in NVT ensemble and at 1 atm using Parrinello-Rahman barostat[111]

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

in NPT ensemble. Three independent production runs, each 100 ns and 500 ns, for every AA and CG system respectively are performed in NPT ensemble using cutoff of 10 Å for AA and 20 Å for CG for both short-range electrostatic interactions and van der Waals interactions. The Particle Mesh Ewald (PME) method[112] is used for the long-range electrostatic interactions and bond lengths are held constant using the LINCS algorithm[113].

Table 4 System setup for the PAA and γ-l-PGA with deionized water having concentration of 2.5 g/l

	All Atomistic			
	PAA		*γ-L-PGA*	
	20-mer	62-mer	20-mer	62-mer
Box size	941.192 nm^3	2958.984 nm^3	1728 nm^3	5832 nm^3
Water atoms	~94 800	~90 300	~56 400	~71 400
Total atoms	94 949	92 700	69 691	72 540
	Coarse Grained			
Polarized Water Beads	2363	7238	4303	1443
Total beads	2402	7361	4343	14629

The complete GROMACS workflow with all used commands can be found in the appendix 9.7.

3.3.1 Simulation of Polymer

For the simulation of polyglutamic acid, γ-L-PGA is chosen since it as an amino acid and the topology can be generated through the similar process used for the generation of topology for PAA.

For PAA and γ-L-PGA, polybuild software was used to build structural models of fully deprotonated single polymer chains in atactic configuration with 20 and 62 repeating units. The structural models of PAA and γ-L-PGA monomers are shown in Fig. 1 & 2. The models were parameterized by using Gaussian09[114] software package and R.E.D. server, respectively. For parameterization, the structural model of γ-L-PGA was split up into head group, tail group and repeating units. In the γ-L-PGA model, the negatively charged repeating unit was capped with a zwitteriinoic head group and double negatively charged tail group. RESP charges were derived using the R.E.D. server with charge constraints on the capping groups. Similarly, the zwitterionic head and the twice negatively charged tail group were parametrized: Both were capped with one ACE and one NME group each. However, for calculating charges a second molecule is necessary, i.e. in case of the head group methylammonium (CH3NH3+) is used together with intermolecular charge constraints between methylammonium and the ACE-NH group of the head group. Additionally, an intramolecular charged constraint has to be applied to the NME cap. For the tail group, acetate (CH3COO–) is used together with intermolecular charge constraints between acetate and the CO-NME group of the tail group. Here, the intramolecular charge constraint must be applied to the ACE cap. Each group carries one negative charge resulting in a net charge of –20. Geometry optimizations were performed by applying density functional theory with B3LYP function[91-93] and 6-31G(d,p)[94] basis set. In order to assign partial charges, electrostatic potential (ESP) charges were calculated using the Hartree-Fock

level of theory with 6-31G(d) basis set[95, 96]. The antechamber software package[97, 98] is used to get the AMBER99 compatible RESP[99-101] charges. The ACPYPE[102] software is used to get a GROMACS compatible topology for PAA and γ-L-PGA with AMBER99 atom types.

All molecular dynamics simulations are performed with GROMACS[107] version 2019 within AMBER99SB[80] force field on Elerion and on the cluster using parallelization with OpenMPI[108]. A cubic box of 1728 nm^3 for γ-L-PGA simulations of small chain is used. Periodic boundary conditions are applied. SPC/E water model is used to solvate the polymers. In order to neutralize the system charge and for simulation of water hardness of 15°dH water molecules are replaced by 26 sodium and 7 chlorine ions for 20-mer system of PAA and γ-PGA simulations.

All systems are first energy minimized using steepest descent algorithm to keep the maximum force on any atom in the system to values < 1000 kJ mol-1 nm-1. After heating, the systems are equilibrated for 1 ns at 300 K using Berendsen thermostat[110] in NVT ensemble and at 1 atm using Parrinello-Rahman barostat[111] in NPT ensemble. Three independent production runs, each 100 ns, for every all-atomistic system are performed in NPT ensemble using 10 Å cut-off for both short-range electrostatic interactions and van der Waals interactions. The Particle Mesh Ewald (PME) method[112] is used for the long-range electrostatic interactions and bond lengths are held constant using the LINCS algorithm[113].

In this study, GROMACS[103] (Groningen Machine for Chemical Simulations) developed by biophysical department of University of Groningen and designed to simulate proteins, lipids, and nucleic acid is used to study the molecular behaviour of polyelectrolytes. GROMACS is a molecular modelling engine belonging to the realm of computational chemistry ranging from quantum mechanics of molecules to dynamics of large complex molecular aggregates, used for performing molecular dynamics and energy minimization.

3.3.2 Creating A GROMACS Topology for A 20-Mer Model of PAA by Running pdb2gmx

The modified GROMACS library that contains the topology of the deprotonated PAA is available with the institute of biotechnology, RWTH Aachen. If we want to define the library used by GROMACS, we need to export the library by

export GMXLIB=/media/storage/b/data/PAA_models/top2itp/top

In order to get the topology for the 20-mer model of deprotonated PAA in GROMACS format we have to execute pdb2gmx

pdb2gmx –f paa_renamed.pdb –o 20PAA.gro

–f input in .pdb format –o output in .gro format

When prompted, choose ‘2’ for AMBER99sb.ff and ‘6’ for SPC/E water.

The function pdb2gmx creates the following files

- 20PAA.gro: Like a .pdb file a .gro file is a fixed-column coordination file. Files with the .gro file extension contain a molecular structure in Gromos87 format, which can be used with GROMACS.
- topol.top: The .top file extension stands for the system topology which contains all necessary information to define all molecules within the simulation. This information includes nonbonded parameters (atom types and charges) as well as bonded parameters (bonds, angles and dihedrals).

- posre.itp: The porse.itp file is used during the equilibration to restrain the positions of the heavy atoms. This means, movement is permitted, but only after overcoming a substantial energy penalty. Therefore, the position restraint allows to equilibrate the solvent around our polymer, but there will be no structural changes in the polymer.

3.3.3 System Preparation

To start the simulation, we need to build box around the polymer chain. The size of the box was calculated according to the experimental concentration of 2.5 g/l. In order to obtain this concentration for 20-mer on PAA chain a box with a volume of 941.192 nm^3 is needed. Box volumes of other simulation are available in Table 4. We create the box by executing *gmx editconf*

gmx editconf –f 20PAA.gro –o 20PAA_box.gro –c –box 9.85 –bt cubic

–f	Input	*–box*	box size
–o	output	*–bt*	box type
–c	centre molecule in box		

To solvate the box the GROMACS tool *gmx solvate* is used

gmx solvate –cp 20PAA_box.gro –cs spc216.gro –o 20PAA_solv.gro –p topol.top

–cp	input (coordination of protein)	*–o*	Output
–cs	input (coordination of solvent)	*–p*	process topology file (.top)

This function will fill the defined box with ~94000 molecules of water. The calculated density of our system is 996.521 g/l.

Since our polymer is charged negatively, we need to add ions to neutralize our system. However, the final washing solution will have a water hardness of 15° dH. Therefore, our system needs additional ions. For a box size of 941.192 nm^3, 20 Na^+ ions have to be added to neutralize the charge of the polymer for a salt free solution as well as 7 sodium and 7 chorine ions to simulate the synthetic water hardness for a system containing only monovalent ions.

First, we use *gmx grompp* to process the coordinate file and to generate an atomic-level input (*.tpr*). The *.tpr* file contains all the parameters for every atom in the system. To produce a *.tpr* file with *gmx grompp*, we will need an additional input file, called molecular dynamics parameter file (*.mdp*). *gmx grompp* will assemble the parameters specified in the *.mdp* file with the coordinates and topology information to generate a *.tpr* file.

Here we need the *ions.mdp* file as input. This file and all other *.mdp* files needed later on, are mentioned in the appendix 9.10 – 9.14.

To execute *gmx grompp* enter the following in the terminal

gmx grompp –f ions.mdp –c 20PAA_solv.gro –p topol.top –o ions.tpr

–f	input in *.mdp* format	*–p*	process topology file (*.top*)

–c	input of atom coordination in *.gro* format	*–o*	output in *.tpr* format

The GROMACS tool for adding ions is called *genion*, which reads through the topology and replace water molecules with the ions that the user specifies. For the synthetic lab water different kind of salt with different charges were used. However, within AMBER forcefield it is difficult to simulate multiple charged ions. For this reason, we use NaCl and $CaCl_2$ to simulate the same ionic strength as in the lab water. As GROMACS chooses NaCl by default we do not have to specify this.

gmx genion –s ions.tpr –o 20PAA_solv_ions.gro –p Na -nn Cl topol.top –np 27 –nn 7

–s	input in *.rtp* format	*–np*	number of cations
–o	output	*–nn*	number of anions
–p	adds entries to *topol.top*		

GROMACS will ask for which kind of atoms it should replace with ions, here we enter '6' for solvent.

Note: Furthermore, *genion* can be used to add a specified concentration (*–conc* …) of ions as well as to simply neutralize the system (*–neutral*)

3.3.4 Energy Minimization

Our system is now solvated and neutralized but before starting molecular dynamics (MD) we must perform an energy minimization. The process of solvating the polymer usually introduces some close contacts that result in very high repulsive energies. If we would start the MD simulation right now, the system would be unstable because of these high energies.

First, we must use *gmx grompp* again to process *mimim.mdp* and assemble the structure, topology, and simulation parameters into a binary input file (*em.tpr*), which is needed in *gmx mdrun.*

gmx grompp –f minim.mdp –c 20PAA_solv_ions.gro –p topol.top –o em.tpr

–f	input in *.mdp* format	*–p*	process topology file (*.top*)
–c	input of atom coordination in *.gro* format	*–o*	output in *.tpr* format

The tool *gmx mdrun*, the main computational chemistry engine within GROMACS, is used for running MD simulations, but it can also perform stochastic dynamics, test particle insertion, (re)calculation of energies and steepest descent energy minimization (EM).

gmx mdrun –v –deffnm em

–v	prints mdrun progress to screen
–deffnm	define the file names of the input and output, here it uses 'em' as the base filename for all options

gmx mdrun will produce the following files:

- *em.log*: ASCII-text log file of the EM process
- *em.edr*: Binary energy file
- *em.trr*: Binary full-precision trajectory

- *em.gro*: Energy-minimized structure

In order to determine if the energy minimization was executed successful there are two criteria to appraise. The first is the potential energy E_{pot} which should be negative and in the order of 10^5-10^6, depending on the system size and number of water molecules (see Fig.11). For our system, 20-mer model of fully deprotonated PAA with water and ions, we get a potential energy $E_{pot} = -1.51 \times 10^6$. The second criterion is the maximum force F_{max}, which is define as 'emtol = 1000.0' in *minim.mdp*. Therefore, F_{max} should not be greater than 1000 kJ mol^{-1} nm^{-1}. After the EM of our system we got a $F_{max} = 9.40 \times 10^2$. As *em.edr* contains all the energies GROMACS calculated during EM, we can analyse the curve of the potential energy using GROMACS tools.

gmx energy –f em.edr –o potential.xvg

–f input in *.edr* format **–*o*** output in *.xvg* format

At the prompt, type '9 0' to select Potential (9) and to terminate the input (0).

Note: for usage of xmgrace: On German operating systems, the comma is the default decimal separator. To use the dot separator instead, declare LC_NUMERIC=C

alias xmgrace="LC_NUMERIC=C xmgrace"

The system is now at the energetic minimum, but before starting the real MD simulation we have to equilibrate the system.

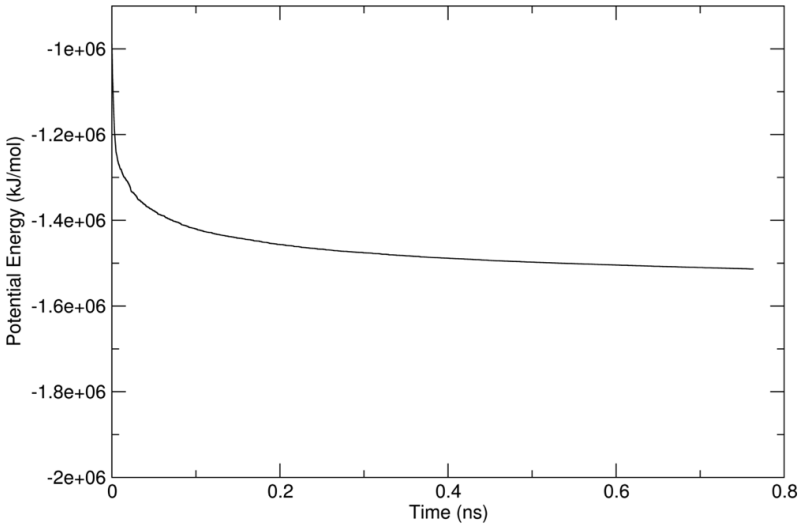


Figure 11 Potential energy curve after energy minimization (Fmax= 1000 kJ mol-1 nm-1) for 20-mer model of PAA in water with NaCl ions

3.3.5 System Equilibration

If we were to attempt unrestrained dynamics at this point, the system may collapse. The reason is that the solvent is mostly optimized within itself, and not necessarily with the solute. Equilibration is needed to further relax the system. The water molecules can then rearrange around the polymer. As structural changes in the polymer during the equilibration phase should not appear, the heavy atoms will be restrained with harmonic forces to its initial position. This is done by using *posre.itp.*

In order to define different temperature coupling groups, which are more accurate than one group, we have to create an index file.

gmx make_ndx –f 20PAA_solv_ions.gro

We spilt the whole system in three parts, called 'Polymer', 'Ion' and 'SOL'. Every atom that belongs to the residues PAB, PBH or PBT will be in group 'Polymer', every water molecule in SOL and ions will be in 'Ion'. This can be done here by simply renaming existing groups:

name 1 Polymer

q

Equilibration is carried out in two phases. The first phase is called NVT or 'canonical' equilibration. During this phase, the number of particles, volume, and temperature is set to constant value. We will use *gmx grompp* to process *nvt.mdp* and *gmx mdrun* to perform the first step of the equilibration.

gmx grompp –f nvt.mdp –c em.gro –p topol.top –o nvt.tpr –n index.ndx

–f	*input in .mdp format*	*–p*	*process topology file (.top)*
–c	*input of atom coordination in .gro format*	*–o*	*output in .tpr format*
–n	*index file for group index of polymer*		

The names of the temperature coupling groups in *nvt.mdp* have to match the names in the *index.ndx* file, otherwise there will be an error.

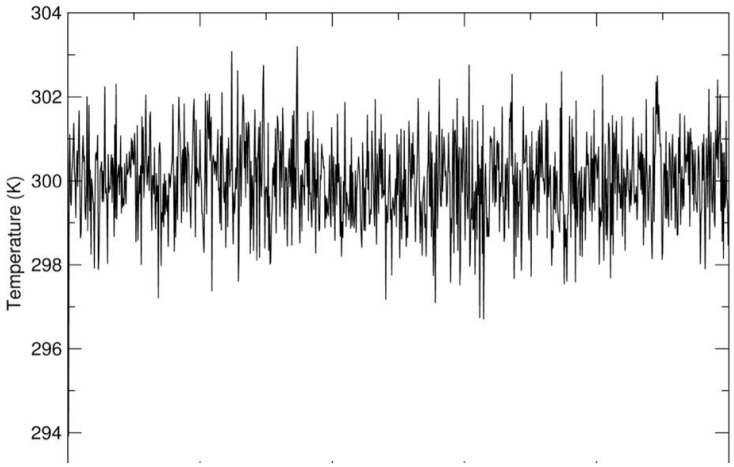


Figure 12 Temperature profile of NVT equilibration for 20-mer model of PAA with water and NaCl ion

gmx mdrun –v –deffnm nvt

To analyse temperature progression, *gmx energy* is used again:

gmx energy –f nvt.edr –o temp.xvg

Enter '14 0' to choose temperature analysis when prompted.

The previous step, *NVT* equilibration, stabilized the temperature of the system (see Fig.12). In the second phase the pressure and thus also the density of the system will be stabilized, which is conducted under an *NPT* ensemble. During *NPT* equilibration the number of particles, pressure, and temperature are all constant

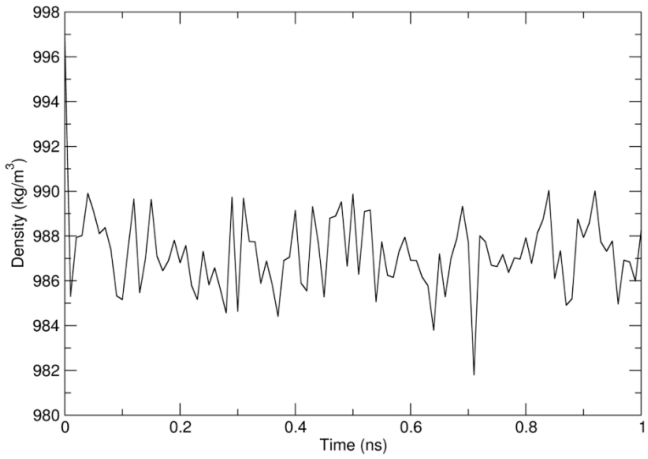


Figure 13 Density profile during NPT equilibration for 20-mer model of PAA with water and NaCl ions

gmx grompp –f npt.mdp –c nvt.gro –t nvt.cpt –p topol.top –o npt.tpr –n index.ndx

gmx mdrun –v –deffnm npt

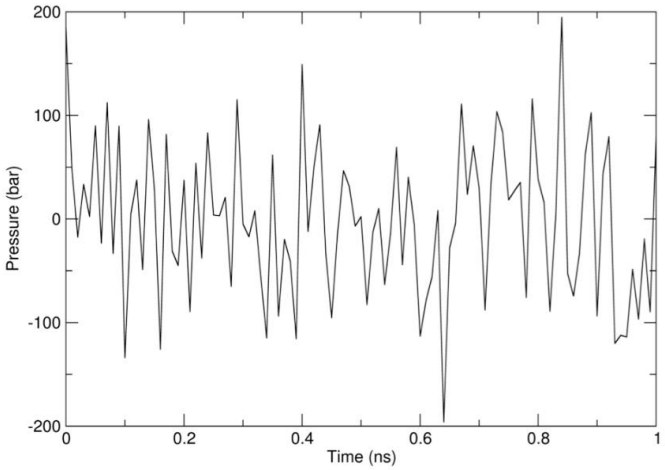


Figure 14 Pressure profile during NPT equilibration for 20-mer model of PAA with water and NaCl ions

To analyse pressure progression, *gmx energy* is used again:

gmx energy –f npt.edr –o pressure.xvg

Enter ’15 0’ to choose pressure analysis when prompted

The density of the system can be analysed as follows:

gmx energy –f npt.edr –o density.xvg

Enter ’21 0’ to choose density analysis when prompted.

Now the system is in a state with average temperature of 300 K, density is at 987 kg m^{-3} after equilibration, average pressure of 0.5 bar (see Fig.13 and Fig.14).

3.3.6 Running MD Simulation (Production Run)

After completing the two steps of equilibration the system is well equilibrated at the desired temperature and pressure. The position restraints can now be released, and we can run the production MD. As before, we will use gmx grompp and gmx mdrun to perform a 100 ns MD simulation for all-atomistic system.

gmx grompp –f md.mdp –c npt.gro –t npt.cpt –p topol.top –o md_0_1.tpr –n index.ndx

gmx mdrun –deffnm md_0_1

3.3.7 Running an MPI job on Cluster

A lot of computation time is needed to perform a 100 ns MD simulation. A single CPU would need several weeks to calculate the trajectory for a small system. In order to simulate bigger system in appropriate time the calculation has to be divided among several CPUs. This can be done by running a simulation on the cluster with MPI. By using gmx_mpi instead of gmx the simulation run will be parallelized to a pre-defined number of CPUs.

gmx_mpi mdrun –deffnm md_0_1

3.3.8 Visualisation of MD Trajectories

VMD[51] is a useful tool for visualization of MD trajectories. High quality snapshots can easily be saved with Tachyon rendering by defining the resolution with the *–res* flag. For example, we can use this render command to create snapshots with a resolution of 1024 x 1024 pixels:

"/Users/username/abc" –aasamples 12 %s –format TGA –res 1024 1024 –o %s.tga

4 Development of a protocol for analysis of trajectories from MD simulations with VMD and GROMACS

For the development of a protocol that would be followed for all the simulation models and system, 20-mer PAA all atomistic (AA) simulation was considered for the purpose of validation of the system parameters, model and protocol.

The following analysis protocol will start with a visual analysis with VMD of the 100 ns trajectory and ensure the quality of the simulation, followed by the calculation of important system parameters like radius of gyration, end-to-end distance, solvent accessible surface area, distribution of hydrogens bonds and radial distribution functions. The complete analysis protocol can be found in the appendix 9.5.

4.1 VISUALIZATION OF STRUCTURES WITH VMD

First, we use gmx trjconv, a post-processing tool to strip out coordinates, correct for periodicity, or manually alter the trajectory (time units, frame frequency, etc.). We will use gmx trjconv to correct the periodic boundary conditions of our system in order to visualize the system with VMD

gmx trjconv –s md_0_1.tpr –f traj_comp.xtc –o traj_noPBC.xtc –centre –pbc mol –ur compact

-s	*input of structure in .tpr format*
-o	*output of trajectory (.xtc)*
-f	*input of trajectory in .xtc format*
-pbc	*option mol puts centre of mass in box*
-ur	*option compact puts all atoms at the closest distance from the centre of the box*
-centre	*puts selected group in the centre of the box*

When prompted, select 0 (‘System’) for output.

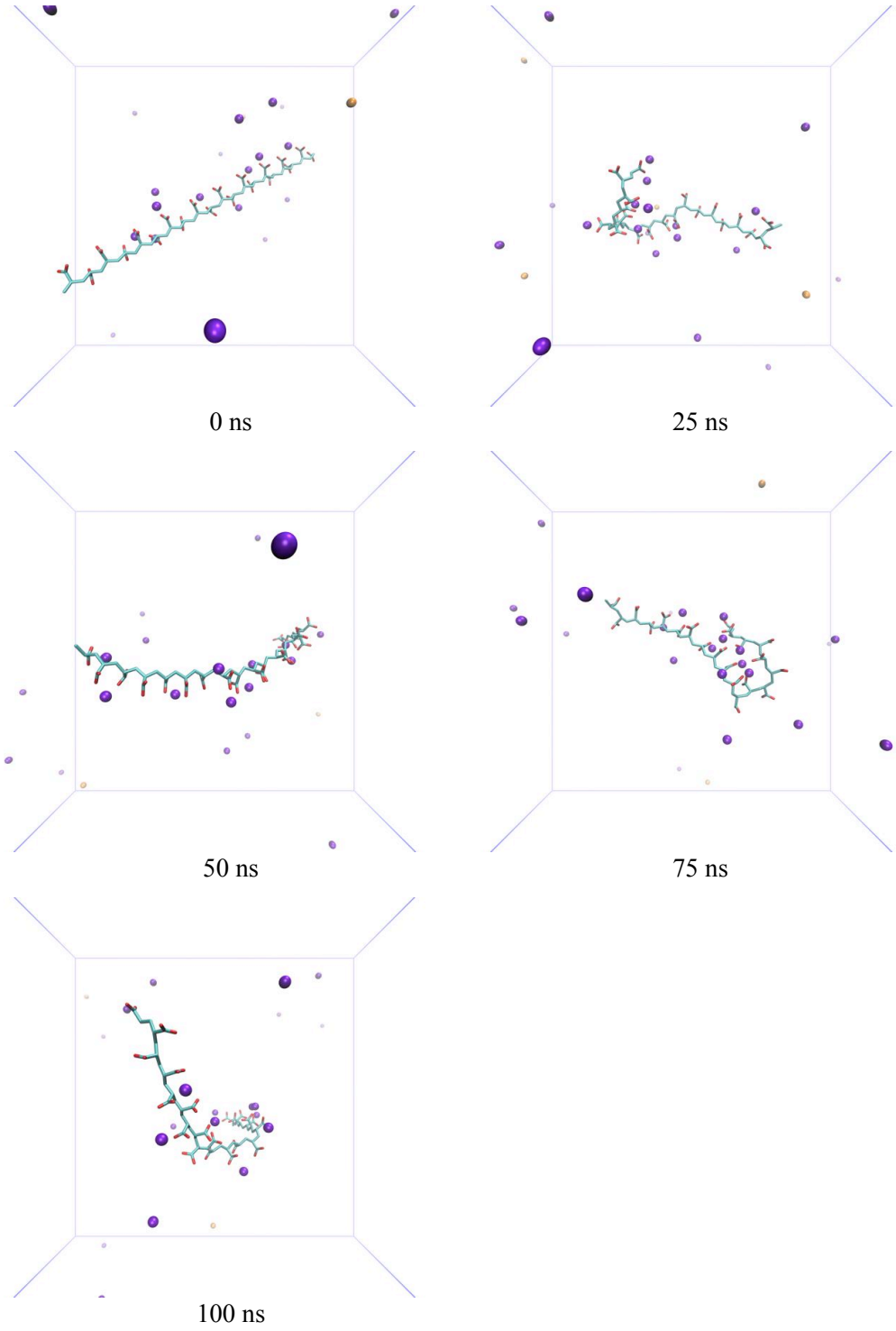


Figure 15 Snapshots taken from MD trajectory for run_2. PAA is shown as sticks with carbon (cyan) and oxygen (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: sodium ions (violet) and chlorine ions (orange)

Snapshots of polymer during the 100 ns simulation are shown in Fig. 15 there are some representative snapshots for one production run for 20-mer model of fully deprotonated PAA. Immediately, the initial structure starts to fold slightly due to the electrostatic interaction with the surrounding sodium ions. The snapshot of PAA after 75 ns of simulation time shows that PAA folds around the sodium ions and is able to enclose them. As this might not be the most stable conformation of PAA, the polymer is able to relax back in a more stretched conformation, as shown in the last snapshot after 100 ns simulation time.

4.2 QUALITY ASSURANCE

As a quality assurance we will have a look at the temperature, pressure, and energy of the system again to check if the system is not only well equilibrated with itself but also with its environment. This can be done by using GROMACS tool *gmx energy* as follows:

gmx energy –s md_0_1.tpr –f ener.edr –o temperature.xvg

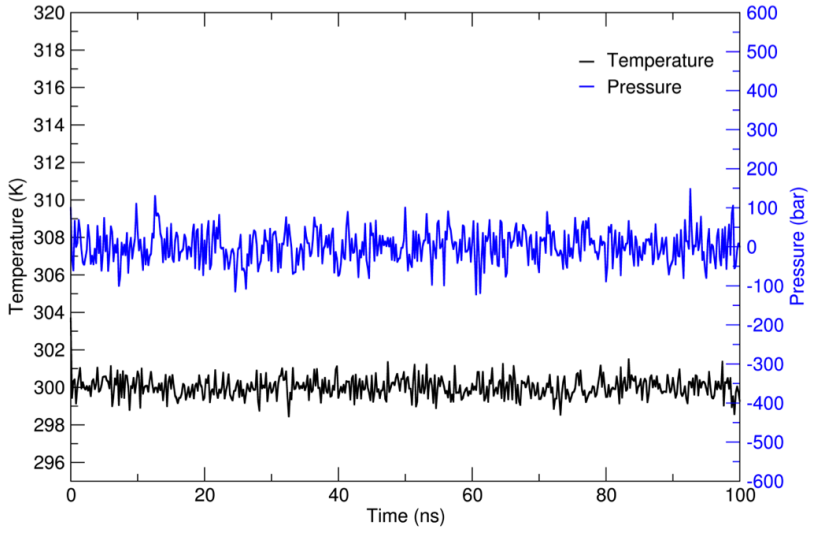


Figure 16 Average temperature and pressure profile of 3 independent production runs (100 ns) for 20-mer model of fully deprotonated PAA with NaCl ions and water

Enter '12 0' to choose temperature analysis when prompted.

gmx energy –s md_0_1.tpr –f ener.edr –o pressure.xvg

Enter '14 0' to choose pressure analysis when prompted.

gmx energy –s md_0_1.tpr –f ener.edr –o energies.xvg

When prompted enter '9 10 11 0' to choose the analysis of potential energy, kinetic energy and total energy.

The average temperature and pressure of all 3 production runs is 300.0 ± 0.5 K, respectively 0.5 ± 42 bar (see Fig.16). The average temperature seems to be quite stable compared to the

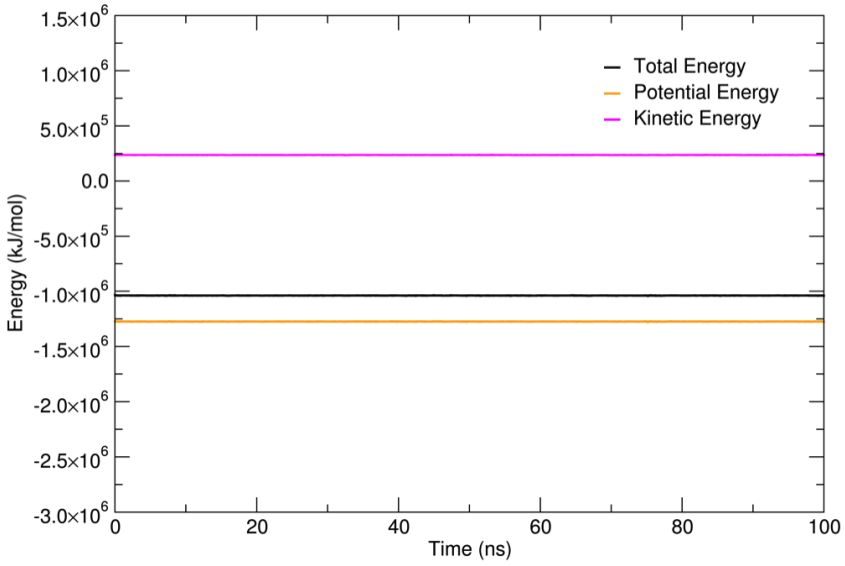


Figure 17 Average energy profiles of 3 independent production runs (100 ns) for 20-mer model of fully deprotonated PAA with NaCl ions and water

average pressure. However, big fluctuations in pressure are expected since pressure is a macroscopic property and thus can only be measured properly as a time average, while here we use a microscopic time scale.

The average energies (see Fig.17) of our system are constant, the kinetic energy is $2.4\text{x}10^5 \pm 4.3\text{x}10^2$ kJ/mol, the potential energy is $-1.3\text{x}10^6 \pm 6.5\text{x}10^2$ kJ/mol and the total energy is $-1.0\text{x}10^6 \pm 7.9\text{x}10^2$ kJ/mol.

As the system properties are stable, we can now be sure that the system was well equilibrated, and the simulation has been run successfully.

4.3 RADIUS OF GYRATION

The radius of gyration is usually used for proteins as a measure of its compactness, but it is also useful for polymers and polyelectrolytes as it gives a sense of the size of a polymer coil. For stably folded proteins and polymers the value of R_g will maintain steadily. But if a protein unfolds or a polymer changes its conformation, the radius of gyration will change over time.

gmx gyrate –s md_0_1.tpr –f traj_comp.xtc –o gyrate.xvg

In order to compare our simulation with the literature[25, 28, 30] an average curve from 3 single production runs is calculated and plotted in every following figures.

Biermann *et al.*[25] simulated one atactic oligomer strand with 23 repeat units in simple point charge (SPC) water with GROMOS96 forcefield. The carboxylic acid groups were fully deprotonated, and 23 sodium ions were added as counter ions. The system was equilibrated at 333.15 K and 1 atm for 1 ns and simulated for about 4.5 ns. They determined the R_G value of 1.27 nm for 23-mer. Following the approach of Biermann *et al.*, Reith *et al.*[28] studied fully deprotonated, atactic polyacrylic acid with 8, 12 and 23 monomers with atomistic and coarse grain model using GROMOS forcefield. The R_g values calculated from MD simulations at

333.15 K can be extrapolated for a chain length of 20 polyacrylic

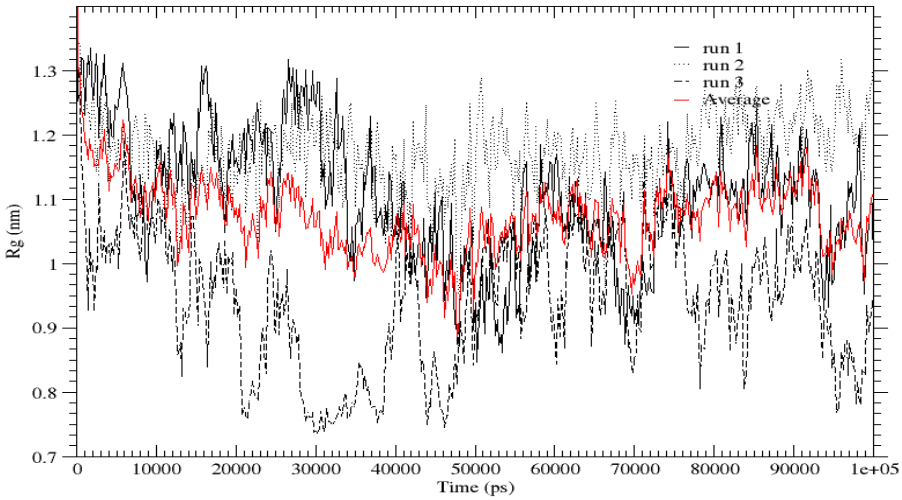


Figure 18 Change of the radius of gyration Rg over time for all 3 independent production runs (black) of PAA (fully deprotonated) simulation and average curve from these 3 runs (red)`

acid repeat units. This leads to a R_g value of 1.14 nm based on atomistic model, respectively 1.06 nm based on coarse grain model. Most recently, Sulatha *et al.*[30] investigated difference in structural behaviour of polyacrylic acid and polymethacrylic acid by MD simulations. They focused on atactic polyacrylic acid chain with 20 monomer units and different degree of deprotonation. The MD simulations were carried out for 15 ns at 300 K and 1 atm with 2 different force fields, one based on the GROMOS53a6 parameter set (FF-1), the other one based on ENCAD parameter set (FF-2). For fully deprotonated polyacrylic acid R_g values of 0.994 nm for FF-1 and 0.997 nm for FF-2 were determined.

The time-averaged radius of gyration, calculated from the average curve, is 1.22 ± 0.09 nm, which agrees reasonable with the previously reported simulation studies.[25, 28, 30]

As we can see from Fig.8 the radius of gyration stays stable for 1 independent production run. For the run 2 and 3 the system shows high instability for the entire simulation period. In run 3 the polymer chain folds after 13 ns and stays in the folded conformation for next 33 ns. The behaviour shown in the third run is very strange and needs to be inspected further. For run_1 (solid black line) conformation of the polymer does not show drastic change and in run_2 (dotted black line) there is a sharp decrease after 30 ns time. The radius of gyration of PAA during run_3 increases again after another 47 ns of simulation time and reaches its former value. During run_2 PAA undergoes a conformational change and remains in this conformation until the end of the simulation (see Figure 19. Snapshots taken from MD trajectory for run_2. PAA is shown as sticks with carbon (cyan) and oxygen (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: sodium ions (violet) and chlorine ions (orange)Root-mean-square dev8)

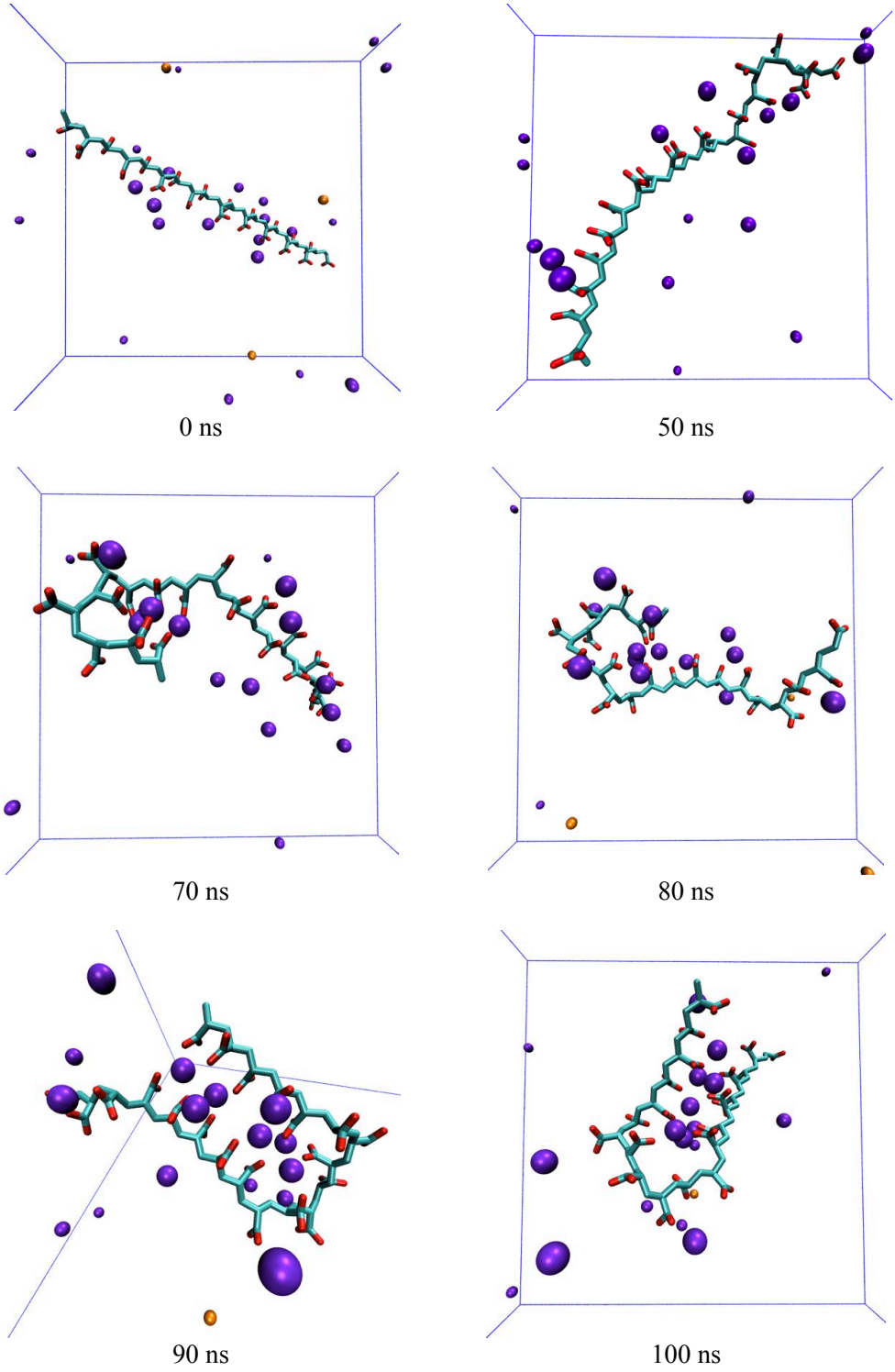


Figure 19. Snapshots taken from MD trajectory for run_2. PAA is shown as sticks with carbon (cyan) and oxygen (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: sodium ions (violet) and chlorine ions (orange)Root-mean-square dev

4.4 ROOT-MEAN-SQUARE DEVIATION

The root-mean-square deviation (RMSD) is a measure of the average distance between the atoms of superimposed proteins. RMSD calculation is usually used to measure the similarity of 3-dimensional structure and to analyse the structural stability of proteins.

The GROMACS tool *gmx rms* is used calculate the RMSD of the backbone atoms of PUR.

gmx rms –s md_0_1.tpr –f traj_comp.xtc –o rmsd_backbone.xvg –tu ns

Enter '4' (backbone) for both the least squares fit and the group for RMSD calculations.

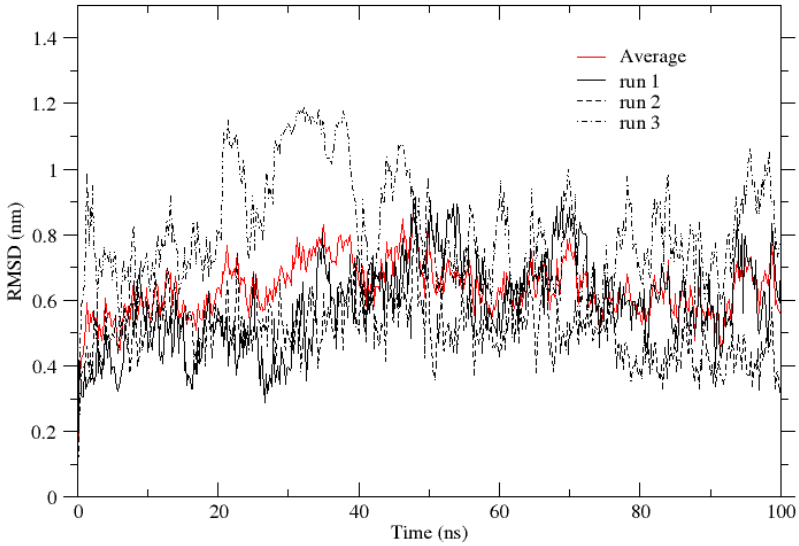


Figure 20 Change of the backbone RMSD over time for all 3 independent production runs (black) of PGA simulation and average curve from these 3 runs (green)

Fig.20 shows the RMSD of the backbone relative to the starting structure of the simulation. The time average RMSD for PAA is 0.073 ± 0.006 nm, indicating a stable polymer structure.

4.5 END–TO–END DISTANCE

The distance *R* between head and tail group of a polymer can easily be calculated by using *gmx mindist.* This GROMACS tool can not only compute the distance between one group and several other groups but also count the number of contacts within a given distance. With -or, minimum distances to each residue in the first group are determined and plotted as a function of residue number.

gmx mindist –s md_0_1.tpr –f traj_comp.xtc –od mindist.xvg

To calculate the distance between head and tail group of PAA enter '2' for PBH and '4' for PBT.

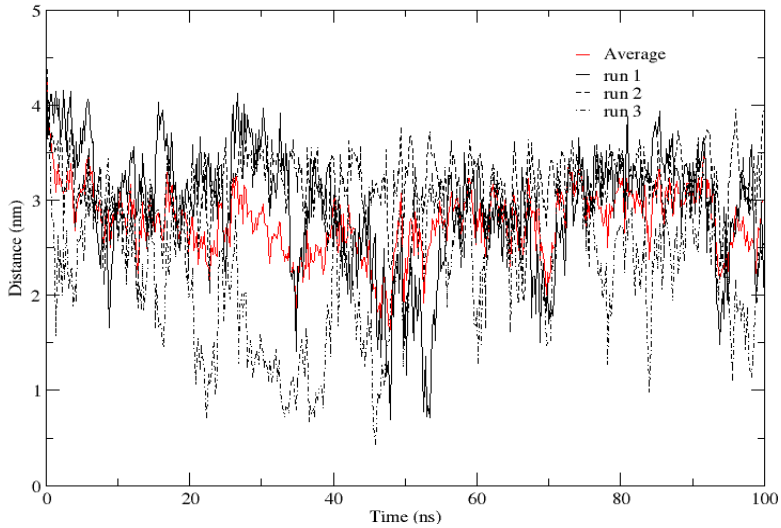


Figure 21 Change of the end-to-end distance R over time for all 3 independent production runs (black) of PAA (fully deprotonated) simulation and average curve from these 3 runs (red)

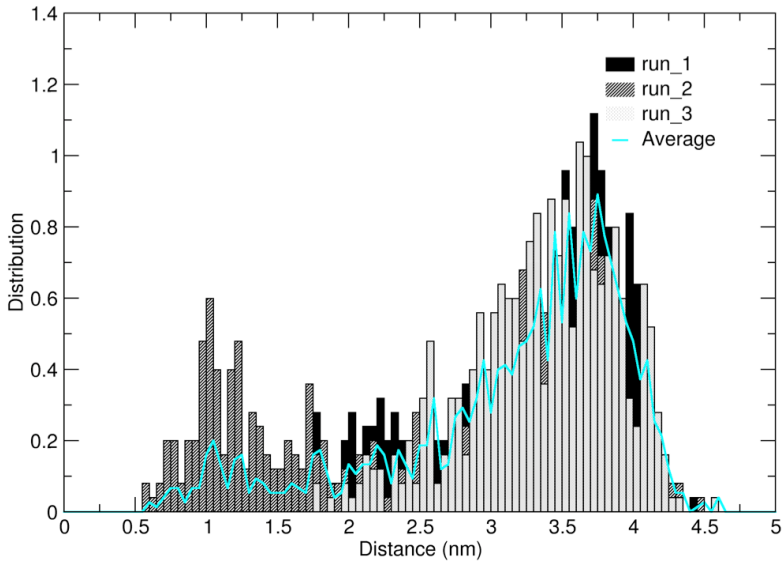


Figure 22 Distribution of end-to-end distance R for all 3 independent production runs of PAA (fully deprotonated) simulation (run_1: solid black; run_2: striped black; run_3: spotted black) and average distribution curve (cyan)

As reported by Biermann *et al.*[25] the average *R* is about 3.6 nm for a 23-mer of PAA. Sulatha *et al.*[30] investigated a difference in polymer chain folding for their two used force fields. While the polyacrylic acid with 20 monomers exhibits a steady increase in *R* with increasing charge density simulated with FF-2, an abrupt decrease after reaching a plateau at 0.8 charge density (80% of carboxylic acid groups are deprotonated) appears with FF-1. This indicates that the chain ends tend to fold inwards, leading to a reduction in *R* above a critical charge density. They found an average end-to-end distance of 2.1 nm for FF-1 and 2.8 nm with FF-2.

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

The time averaged *R* for our simulated 20-mer model of PAA is 3.12 ± 0.52 nm (see 21), which is in between the previously reported values for *R*.

The change of the distance from head to tail group of PAA behaves similar to the change of the radius of gyration. For the first 68 ns there is no big difference in the 3 production runs, they fluctuate around 3.4 nm. While the average end-to-end distance for run_3 remains constant until the end of the simulation, a sudden decrease in the other 2 production runs occurs because PAA folds (see Figure 19. Snapshots taken from MD trajectory for run_2. PAA is shown as sticks with carbon (cyan) and oxygen (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: sodium ions (violet) and chlorine ions (orange)Root-mean-square dev21). As PAA relaxes back to a more stretched conformation during run_1 after another 10 ns the end-to-end distance reaches its former value. The end-to-end distance during run_2 decreases steadily until the end of the simulation. Because of that, the distribution of the end-to-end distance shown in Fig.22 broadens for run_2 and two local maxima appear at 1.1 nm and 3.5 nm. For run 1 and run_2 there is only one maximum at 3.7 nm, respectively at 3.6 nm.

4.6 SOLVENT ACCESSIBLE SURFACE AREA

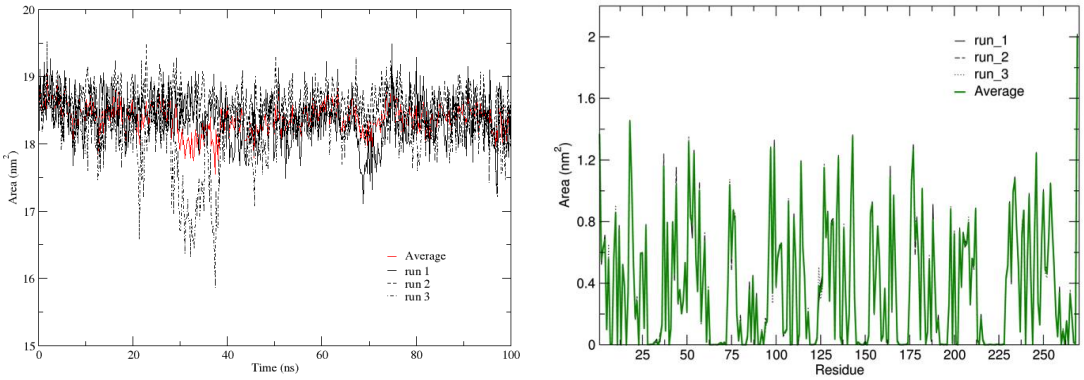


Figure 23 (a) Change of SASA over simulation time and (b) SASA per residue for all 3 independent production runs (black) of PAA simulation and average curve from these 3 runs (green)

The solvent accessible surface area (SASA) is used as a parameter in characterization of protein folds. SASA was first described by Lee and Richards[115] as a sphere, representing a solvent molecule, rolling over the van der Waals surface of a protein and by tracing the centre of the sphere.

gmx sasa –s md_0_1.tpr –f traj_comp.xtc –o sasa.xvg

The PAA has a time average SASA of 18 ± 1.19 nm. In Fig.23 the SASA is shown per residue.

4.7 HYDROGEN BONDS

As our polyacrylic acid is fully deprotonated, hydrogen bonds can only be formed between the carboxylate oxygen atoms and water hydrogen atoms, provided the two oxygen atoms are within a distance of 0.35 nm and the O–H–O angle is not greater than 130°. In order to study the hydration behaviour of PAA we use the GROMACS tool *gmx hbond*.

gmx hbond –s md_0_1.tpr –f traj_comp.xtc –n index.ndx –num hbond.xvg

–s	*input of structure in .tpr format*	*–n*	*index file for group index of polymer*
–f	*input of trajectory in .xtc format*	*–num*	*number of hydrogen bonds as function of time*

Enter ‘1’ for others (PAA) and ’16’ for water and ions at the prompt.

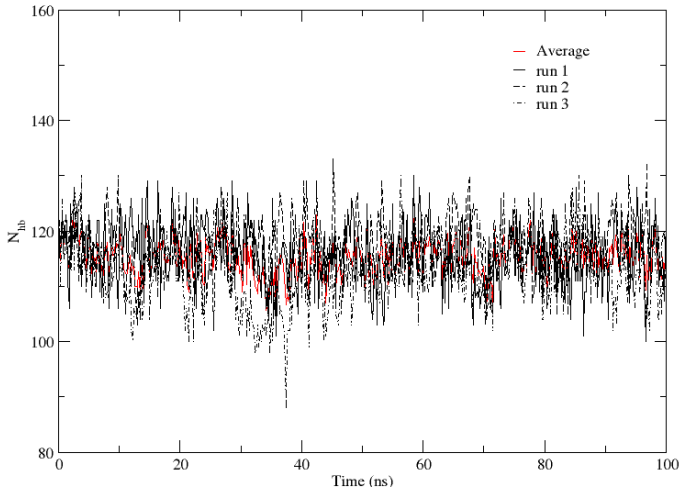


Figure 24 Change of number of hydrogen bonds Nhb over time for all 3 independent production runs (black) of PAA (fully deprotonated) simulation and average curve from these 3 runs (red)

Again, the conformational change happening during run_3 appears in Fig. 25 as a decrease

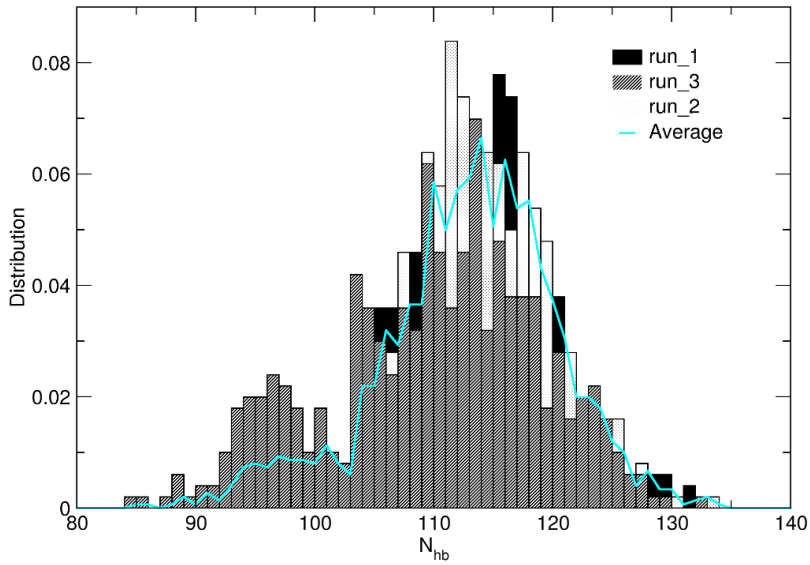


Figure 25 Change of number of hydrogen bonds Nhb over time for all 3 independent production runs (black) of PAA (fully deprotonated) simulation and average curve from these 3 runs (red)

after 35 ns simulation time. As the ends of the chain tend to fold in attendance of sodium ions, the number of free carboxylate ions decreases and thus the number of hydrogen bonds. The number of hydrogen bonds (N_{hb}) for all the runs fluctuate around 114 during the whole

simulation. Further analysis regarding the average number of hydrogen bonds or distribution of hydrogen bonds can be done with the tool *gmx analyse* or simply by using xmgrace and the built-in histogram function.

gmx analyse –f hbonds.xvg –dist hbonds_dist.xvg

The average N_{hb}, calculated as time average, is 114. As we can see from the RDF between carboxylate oxygens and water hydrogen atoms , every oxygen atom is surrounded by nearly 3 hydrogen atoms, so there should be around 120 hydrogen bonds. Bierman *et al.*[25] analysed the hydration behaviour of 23-mer of atactic, fully deprotonated polyacrylic acid. They observed a uniform hydration shell with 136 hydrogen bonds between the carboxylate oxygens and water hydrogens.

4.8 RADIAL DISTRIBUTION FUNCTIONS

The radial distribution function (RDF) between two kinds of particles A and B describes the frequency with which a particle of type B can be found at a distance *r* from a particle of type A, based on the frequency that two particles of an ideal solution are in this distance. Therefore, it can provide a better understanding of the hydration shell.

We want to calculate 4 different RDFs: between the centre of mass of PAA and the water oxygen, between the oxygens of the carboxylate ions of PAA and the water oxygen, the water hydrogens and last, the sodium ions.

First, two different groups of particles have to be defined to calculate an RDF with GROMACS. While PAA, water and ions appear by default in the index file, other groups have to be added manually. This can be done by creating a new index.ndx file.

gmx make_ndx –f npt.gro

In order to create one group for all water oxygen atoms enter ‘a OW’. This will find all oxygen atoms belonging to a water molecule. Similarly, the other missing groups are added. For the hydrogen atoms of water enter ‘a HW1 | a HW2‘. This will add 2 hydrogen atoms per water molecule in a new group. For the oxygen atoms of the carboxylate ions enter ‘a O1 | a O2’. Close and save the new index file by entering ‘q’.

4.8.1 RDF between the centre of mass of the PAA residues and the oxygen atoms of water molecules

After creating a new *index.ndx* file the GROMACS tool to calculate the radial distribution function can by executed by

gmx rdf –s md_0_1.tpr –f traj_comp.xtc –o rdf_PAA_OW.xvg –n index.ndx –rdf res_com

When prompted to select ‘1’ for others, which means all polyacrylic acid residues (PAB) including head (PBH) and tail group (PBT). For group 2 enter ‘17’ which is the newly added OW group (water oxygens).

The sharp peak at 0.370 ± 0.002 nm in indicates a well-structured solvation shell with ordered distribution of water molecules around the polymer chain. This in good agreement with the results of Sulatha *et al.*[30] They observed a first peak at 0.37 nm for FF-1 and 0.35

nm for FF-2. Biermann *et al.*[25] reported a first peak at 0.35 nm.

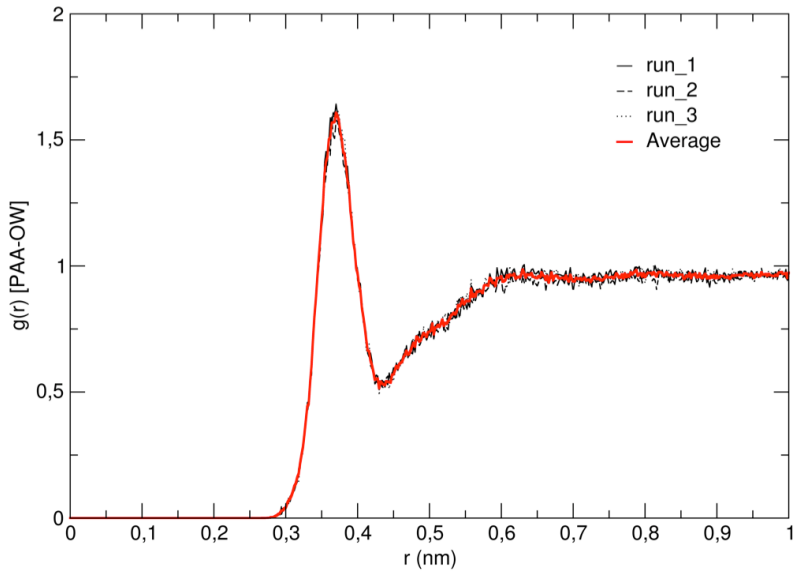


Figure 26 RDF between PAA (center of mass) and water oxygens (OW) for all 3 independent production runs (black) of PAA (fully deprotonated) simulation and average curve (red)

4.8.2 RDF between oxygen atoms of carboxylate group of PAA and the oxygen atoms of water molecules

In order to calculate an RDF between the oxygen atoms of the carboxylate ions of our 20-mer model of PAA and the water oxygens execute GROMACS tool *gmx rdf*

gmx rdf –s md_0_1.tpr –f traj_comp.xtc –o rdf_COO_OW.xvg –n index.ndx

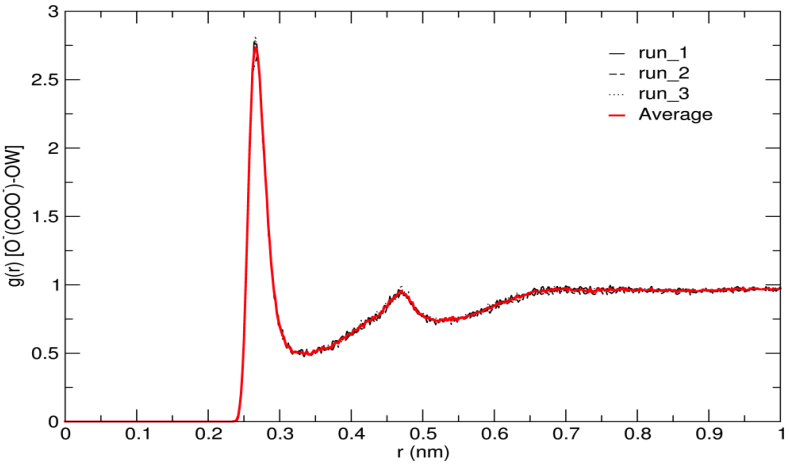


Figure 27 RDF between the carboxylate oxygens of PAA and water oxygens (OW) for all 3 independent production runs (black) of PAA (fully deprotonated) simulation and average curve (red)

For group 1 enter '19' for the newly created O1_O2 group, which includes 40 oxygen atoms belonging to the carboxylate ions of PAA, and for group 2 enter '17' for water oxygens.

The RDF between the carboxylate oxygen atoms and water oxygen atoms in show two peaks at 0.266 ± 0.002 nm and 0.466 ± 0.002 nm. Again, the results of Sulatha *et al.*[30] match ours. There RDFs show two 2 peaks at 0.28 and 0.48 nm, respectively 0.25 and 0.45 nm based on the simulation with FF-1 and FF-2.

4.8.3 RDF between oxygen atoms of carboxylate group of PAA and the hydrogen atoms of water molecules

The RDF between the oxygen atoms of the carboxylate ions and the water hydrogens is analysed in the same way

gmx rdf –s md_0_1.tpr –f traj_comp.xtc –o rdf_COO_HW.xvg –n index.ndx

Only the input at the prompt is different. For group 1 we choose '19', whereas for group 2 we enter '18' for hydrogen atoms of water (HW1_HW2**).**

The RDF between of the oxygen atoms of the carboxylate group with respect to water hydrogens shows a strong peak at 0.174 ± 0.002 nm (see 8). This is similar to the hydrogen bonding distance in bulk water. A shoulder peak at 0.304 ± 0.002 nm follows. Sulatha *et al.*[30] RDFs show a peak at 0.18 nm, followed by a shoulder peak at 0.33 nm for FF-1. The RDF calculated from FF-2 simulations has corresponding peaks at 0.15 and 0.30 nm, which is much shorter than the observed distance in bulk water.

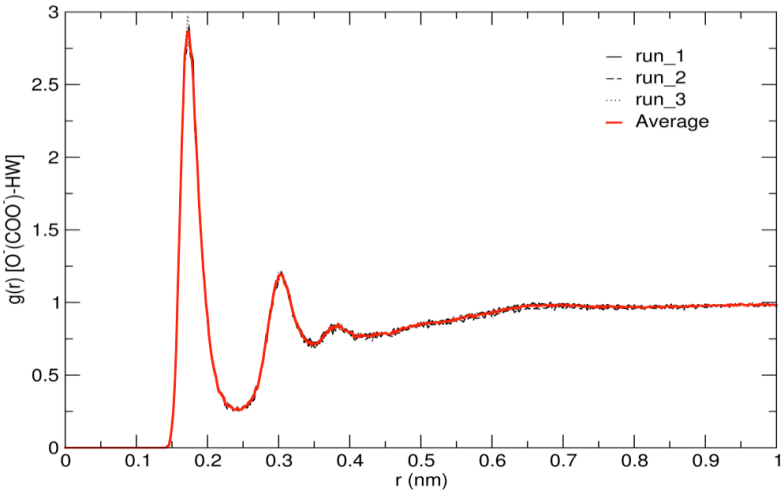


Figure 28 RDF between the carboxylate oxygens of PAA and water hydrogens (HW) for all 3 independent production runs (black) of PAA (fully deprotonated) simulation and average curve (red)

4.8.4 RDF between oxygen atoms of carboxylate group of PAA and the sodium ions

The last RDF is calculated between the oxygen atoms of the carboxylate and the sodium ions by executing *gmx rdf* and entering '19' and '14' for NA (sodium ions) at the prompt

gmx rdf –s md_0_1.tpr –f traj_comp.xtc –o rdf_COO_Na.xvg –n index.ndx

.

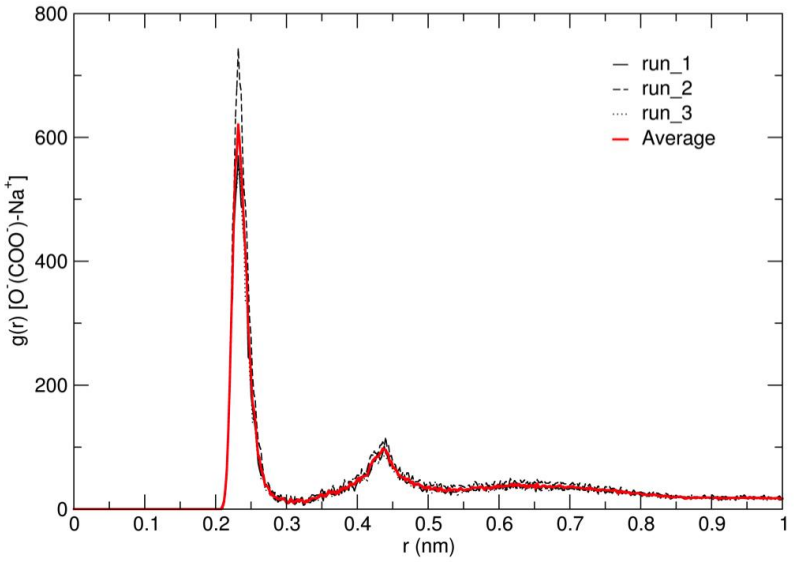


Figure 29 RDF between the carboxylate oxygens of PAA and sodium ions for all 3 independent production runs (black) of PAA (fully deprotonated) simulation and average curve (red)

As shown by Fig.29, the RDF between the carboxylate oxygens and sodium counterions exhibits a strong first peak at 0.232 ± 0.002 nm, suggesting a preferred condensation of sodium ions on to the chain backbone of PAA, and a weaker shoulder peak at 0,438 ± 0.002 nm. Sulatha *et al.*[30] discovered a similar behaviour for simulations with FF-2 parameter set. Their RDFs show a first peak at ~0.21 nm. However, the RDF received from FF-1 simulations predicts a completely different behaviour. There are two major peaks, one at 0.46 nm and a broader second peak at ~1 nm. As reported by Raman scattering spectroscopy studies on dilute aqueous solution of polyacrylic acid[44], sodium ions are more scattered and not bound to carboxylate groups. Therefore, FF-2 and thus our AMBER99SB parameter set are not a good choice to simulate a system with PAA and sodium counter ions. However, the experimental system does not only have sodium ions but also calcium and magnesium ions. Polymers like PAA show higher affinity towards divalent counter ions as shown by Raman spectroscopy[33] and simulation[31, 36]. Therefore, PAA simulated in AMBER99SB force field is regarded is a good model to describe the experimental results. The topology of PAA and the AMBER99SB force field will be used in the following simulations

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

5 REULTS

The final aim of this thesis is to get insights into the interaction mechanisms of polymers within salt free and excess salt solution systems containing monovalent, and divalent ion. The following chapter provides detailed results obtained from the AA and CG trajectories of PAA and γ-L-PGA for chain length of 20 and 62 monomers using AMBER99SB and MARTINI force fields simulated for 100 ns and 500 ns respectively.

5.1 ANALYSIS OF TRAJECTORIES FROM ALL ATOMISTIC MD SIMULATIONS OF FULLY DEPROTONATED 20-MER MODEL OF POLYACRYLIC ACID (PAA)

The simulations are validated with the results from literature and experimental results.

Table 5 Number of ions in Na & Ca (salt free) and NaCl & $CaCl_2$ (excess salt) systems of 20-mer PAA

Salt	***Number of ions***	
	Na/Ca	***Cl***
Na	20	-
NaCl	27	7
Ca	10	-
$CaCl_2$	17	14

The four generated boxes each include one PAA model short-chain model. Periodic boundary conditions were applied in all directions. For PAA simulations, cubic boxes with 9.8 nm edge length were used for the short-chain models. The new refined SPC/E water model was used to solvate the system. In order to neutralize the system charge of the polymer, water molecules were replaced by ions according to the number of monomers in the polymer chain (see Table 5). In order to account for the ionic strength of the synthetic tap water (15°dH) used in experiments additional water molecules were exchanged by sodium and chloride ions. The setup for each simulated system is summarized in Table.5.

All simulations were performed on the RWTH Compute Cluster using GROMACS 2019 simulation package[116]. All systems were first energy minimized using steepest descent algorithm with 50,000 integration steps or until the maximum force on any atom in the system did not exceed a value of 1000.0 kJ/mol/nm. For neighbor searching, Verlet cutoff-scheme was used having short range Van der Waals cut-off of 1.0 with periodic boundary conditions. For electrostatic forces, Particle-mesh Ewald method of order 8 was used having short range cut-off of 1.2 keeping relative dielectric constant and relative dielectric constant of the reaction field equal to 1. For Van der waals forces, twin range cut-offs with neighbor list cut-off and VdW cut-off are used. For the treatment of long-range electrostatic interaction and long-range dispersion, corrections for energy and pressure was applied. Grid dimensions are controlled with Fourier spacing of 0.15. The relative strength of the Ewald-shifted direct potential at rcoulomb is given by 'ewald-rtol'. For doing PME for VdW-interactions, ewald-rtol-lj is used to control the relative strength of the dispersion potential at rvdw. Both the values are kept default 1e-5 and 1e-3 respectively.

After heating, the systems were equilibrated for 140 ps each at 300 K using the velocity rescale thermostat (coupling time 1 ps) according to Bussi *et al.*[117] in NVT ensemble and at 1 bar using Berendsen barostat[118, 119] (coupling time 2 ps and compressibility 4.5×10^{-5} bar^{-1})

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

in NPT ensemble. During the equilibration, position restraints on every backbone atom of the PAA model and on every polymer bond were activated. Three independent production runs, 100 ns each, were performed for every system in NPT ensemble using a Verlet cutoff scheme with a cutoff distance of r_{cut} = 1.2 nm for the Lennard-Jones interactions according to a recent publication[120]. Consistently, a time step of 2 fs was used. The neighbor list was updated every 10 steps using the Verlet neighbor search (VNS) algorithm[120]. Bond lengths were held constant by the LINCS algorithm[113, 121]. The final box sizes are shown in Table.4.

A chain length of 20 monomer of PAA (fully deprotonated) is simulated in a system containing water and different quantity of monovalent and divalent ions (sodium, calcium, sodium chloride and calcium chloride) (see Table 5). The system is equilibrated at 300k and 1 bar of pressure. The system is simulated for 100 ns with a time step of 2 femtoseconds. System configuration for production run is provided in appendix 9.10 - 9.13. The simulation results are validated with the results from literature and experimental results.

5.1.1 Visualization of PAA interaction with Ions

Some representative snapshots of the MD simulations of the PAA model with ions are shown in Fig.30-33. The snapshots were taken after 100 ns of MD simulation.

Snapshots of the simulated system consisting of PAA, 20-mer model of fully deprotonated PAA with Na, Ca2+, NaCl and $CaCl_2$ ions in water are shown in Fig. 30, 31, 32 & 33, respectively. The water molecules have not been presented in the images for better visualisation of the polymer chain and its interaction with polymer. The PAA chain is shown as stick model. In Fig. 30 the PAA chain can be seen as a stretched conformation in the first snapshot i.e. at 0 ns but folds shortly after. PAA starts to interact with Na in the region of the strongly bound sodium ion represented by a violet ball. Visualisation can be seen in the following pages.

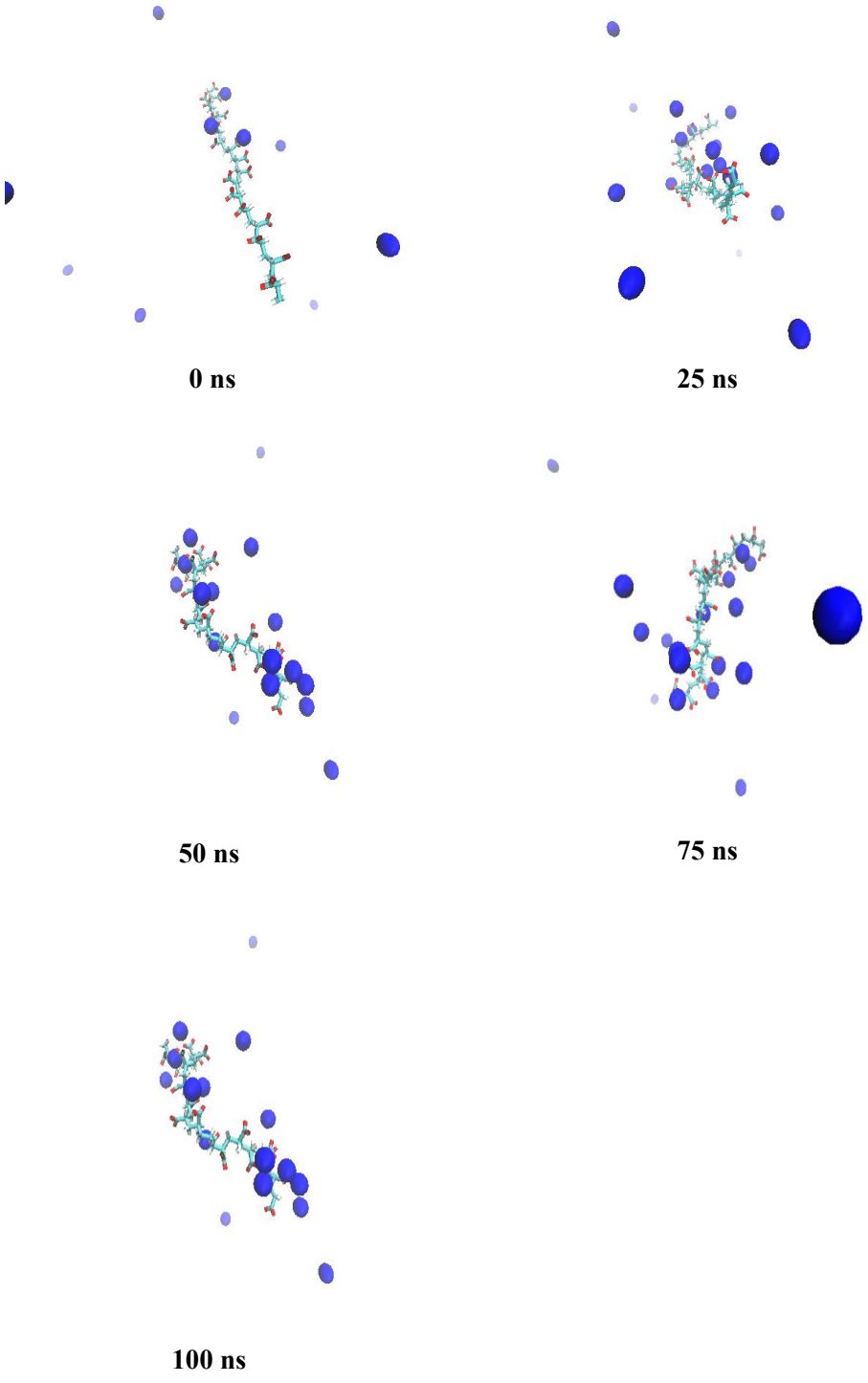


Figure 30 Snapshots taken from MD trajectory of 20-mer model of fully deprotonated PAA with Na ions and water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (100 ns). PAA is shown as sticks with carbon (cyan) and oxygen (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: sodium ions (violet).

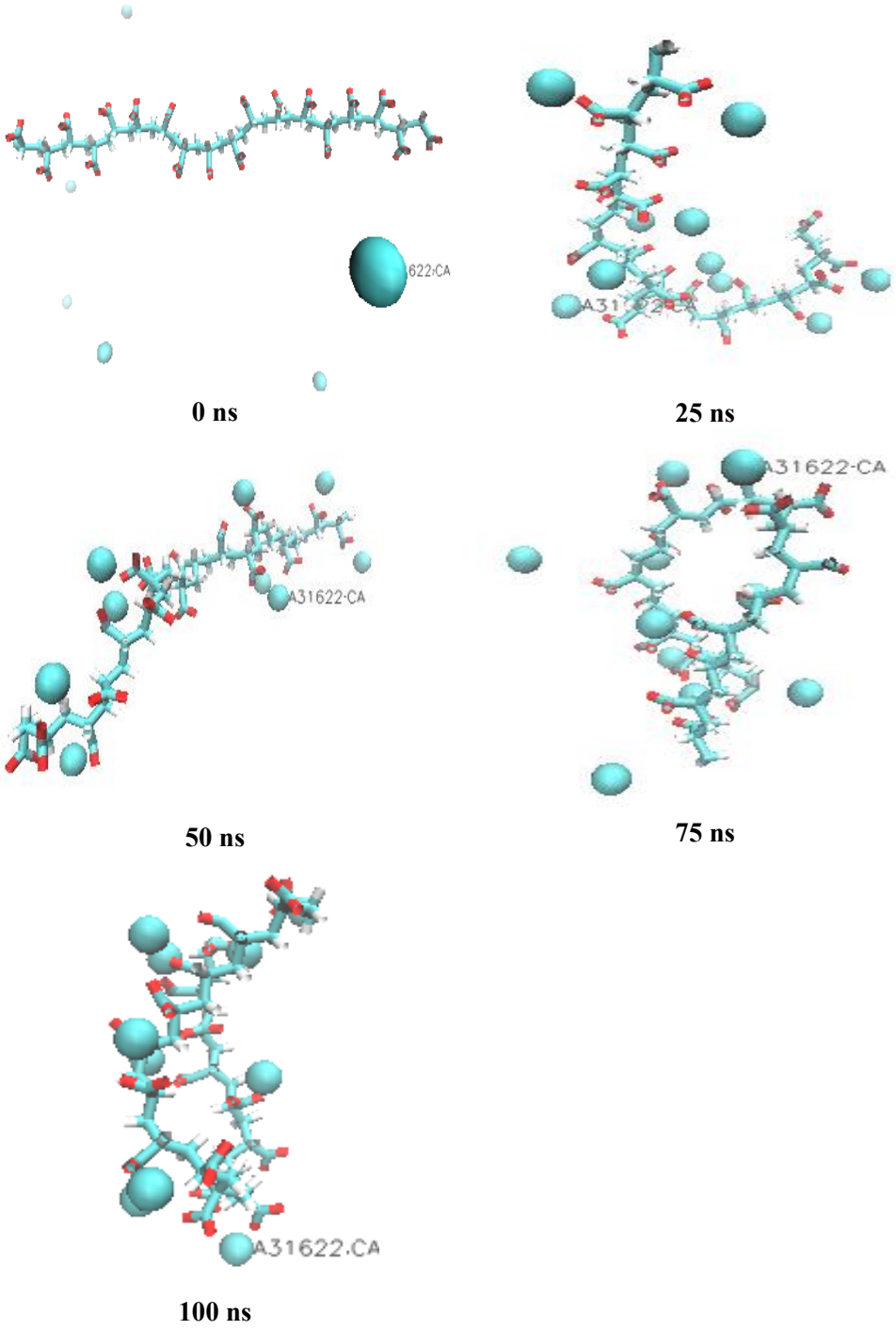


Figure 31 Snapshots taken from MD trajectory of 20-mer model of fully deprotonated PAA with Ca2+ ions and water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (100 ns). PAA is shown as sticks with carbon (cyan) and oxygen (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: calcium ions (cyan)

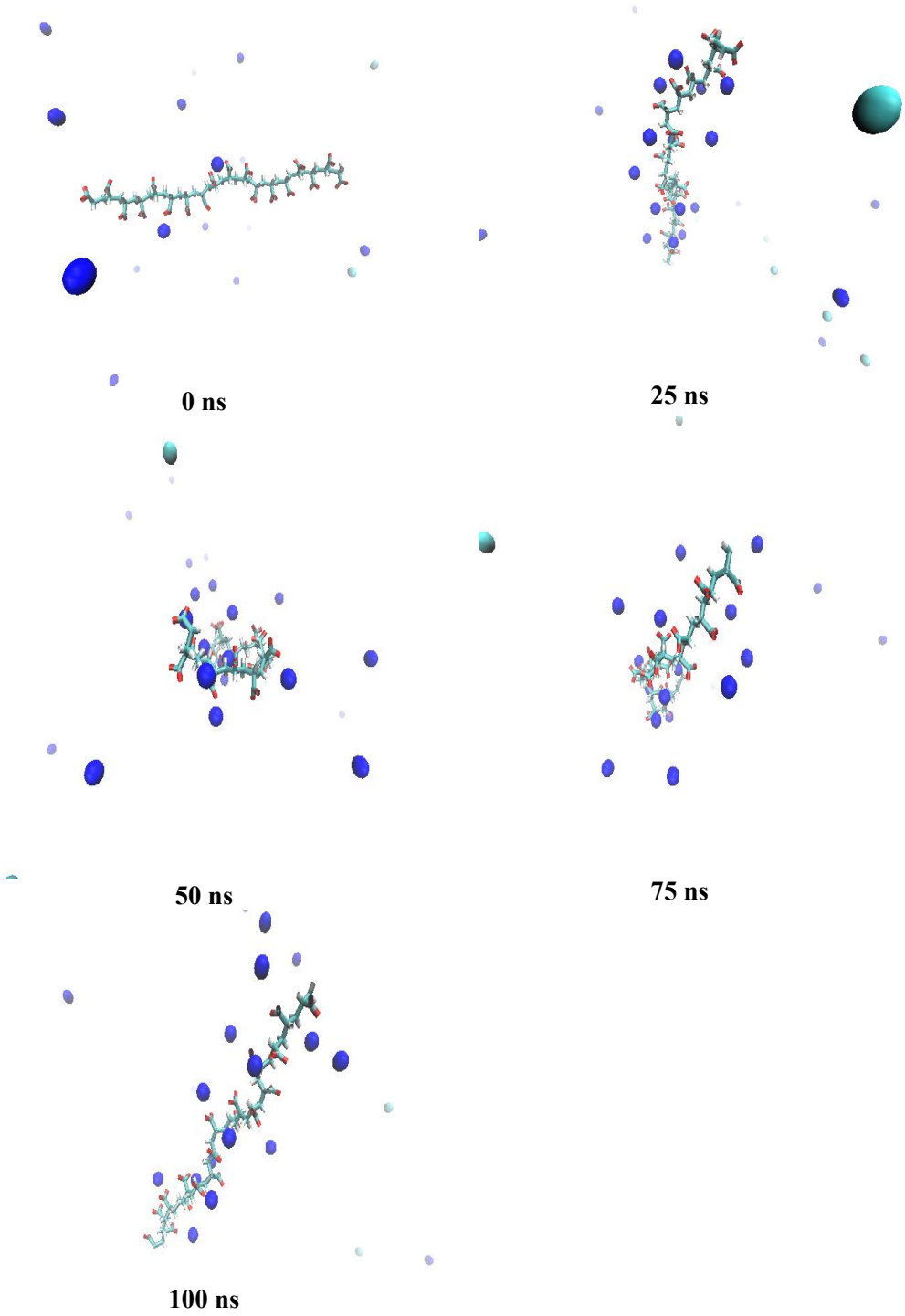


Figure 32 Snapshots taken from MD trajectory of 20-mer model of fully deprotonated PAA with NaCl ions and water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (100 ns). PAA is shown as sticks with carbon (cyan) and oxygen (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: sodium ions (violet) and chlorine ions (teal).

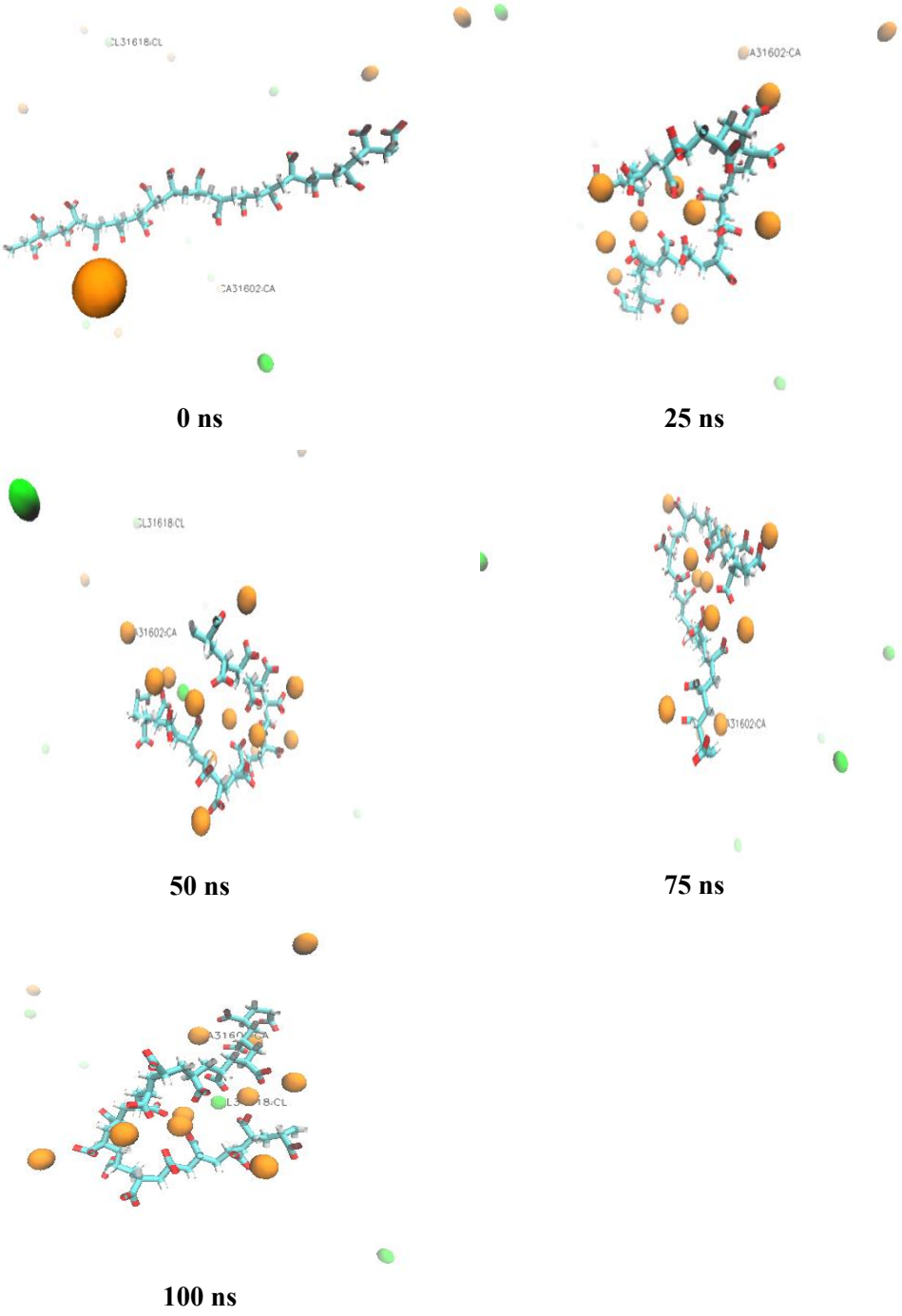


Figure 33 Snapshots taken from MD trajectory of 20-mer model of fully deprotonated PAA with $CaCl_2$ ions and water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (100 ns). PAA is shown as sticks with carbon (cyan) and oxygen (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: calcium ions (orange) and chlorine ions (green)

5.1.2 Quality assurance

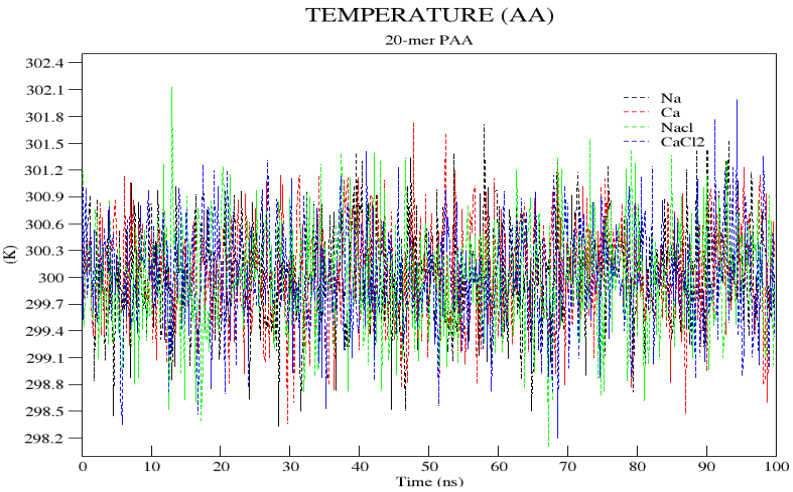
First, we check if the system is well equilibrated. As shown in Fig 34a, 34b, the average temperature and pressure, calculated as a time average from all 3 single independent runs, fluctuate steadily. The average temperature and pressure of all 3 independent production runs is presented in the Table 6. All the systems present an average temperature and pressure of 300.0 ± 0.04 K and 2.2 ± 0.39 bar respectively. The average energies (see Fig. 34c, 34d, 34e) of our system are constant, the kinetic energy is $2.31 \times 10^{5} \pm 0.1 \times 10^{-2}$ kJ/mol, the potential energy is -1.49×10^{6} kJ/mol except the system with calcium and chlorine ions showing a higher level of potential energy -1.51×10^{6} which is well inside the limits of to be considered as a stable system and the total energy is -1.259×10^{6}. The system with $CaCl_2$ shows a higher value of total energy, 1.27×10^{6}. The values are constant over the whole trajectories, this indicates a well equilibrated and stable system.

(The '±' intervals are provided to approximate the result for the time interval at which the results are visible. Even though the results were recorded at small time steps of the simulation, it is computationally expensive to record results for every fs/ns of simulation hence, the approximation is made to present the exact value. This was done by observing the values at the preceding and following time steps at which the graphs present the results under observation and the approximation was done through the slope of the graph).

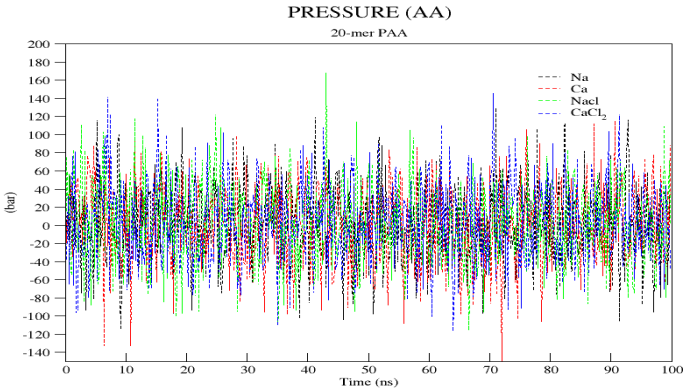
Table 6 System properties calculated as a time average from all 3 single independent production runs

Property	***Average value of over 3 independent production runs***			
	Na	***Ca***	***NaCl***	***$CaCl_2$***
Temperature (K)	300.0224	300.0431	300.0119	300.0256
Pressure (bar)	2.235045	0.302917	1.013022	2.33349
Potential Energy (KJ/mol)	-1.49E+06	-1.50E+06	-1.49E+06	-1.51E+06
Kinetic Energy (KJ/mol)	2.37E+05	2.37E+05	2.37E+05	2.37E+05
Total Energy (KJ/mol)	-1253242	-1259549	-1257884	-1273530

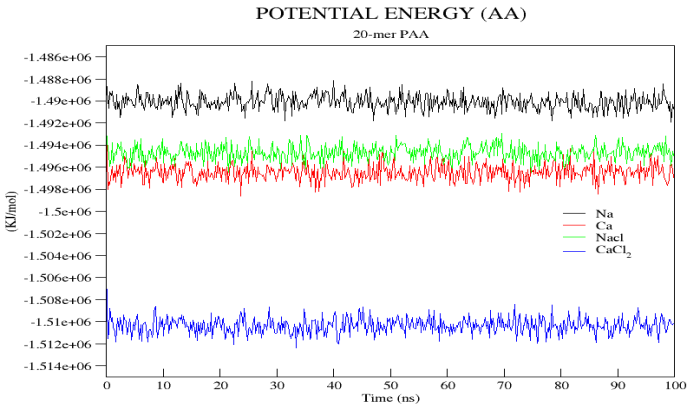
a)


TEMPERATURE (AA)
20-mer PAA
Na
Ca
Nacl
CaCl2
(K)
Time (ns)

b)


PRESSURE (AA)
20-mer PAA
Na
Ca
Nacl
CaCl2
(bar)
Time (ns)

c)


POTENTIAL ENERGY (AA)
20-mer PAA
Na
Ca
Nacl
CaCl2
(KJ/mol)
Time (ns)

d)

e)

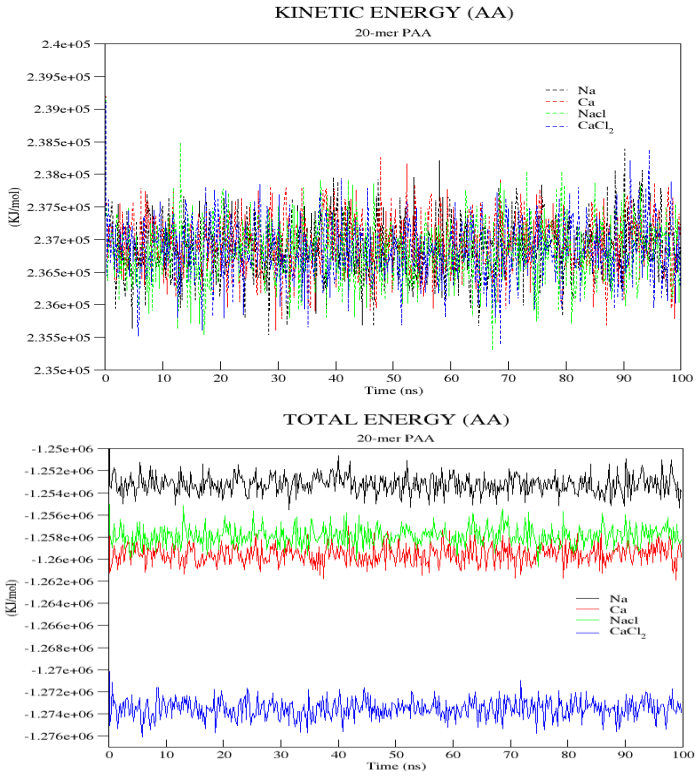


Figure 34 Average energy profiles of 3 independent AA production runs (100 ns) for 20-mer model of fully deprotonated PAA with Na, Ca, NaCl, and $CaCl_2$ ions and water

5.1.3 Radius of Gyration

The radius of gyration R_g of the fully deprotonated PAA is shown in Fig. 35. The time averaged R_g for the four systems is varies between 0.9859 nm and 1.1058 nm. This is slightly lower than the R_g value of PGA because the chain backbone is smaller (2 atoms per monomer instead of 3 atoms).

As we can see from Fig. 35 the radius of gyration stays stable for the production runs for the system containing Na and NaCl until 26 ns of simulation time. PAA shows a sharp conformational change at around 40 ns time stamp of the simulation with NaCl ions undergoing folding and returns to stretched conformation at 60 ns and shows a constant change in conformation till the end of the simulation. For the system containing Ca2+ ions, the system goes from stretched to a folded conformation in the very beginning pf the simulation and keeps folding for the next 50 ns where a sharp drop in R_g can be seen. PAA stays in the same conformation for 20 ns and at 66 ns PAA completely folds hence R_g value decreases substantially due to the bridging effect between the ends of the monomer which leads to this significant drop of R_g value.

Table 7 Average Radius of gyration values calculated over 3 production runs

System	***Average (nm)***
Na	1.1058977
Ca	0.9859738
NaCl	1.0734410
$CaCl_2$	0.8549989

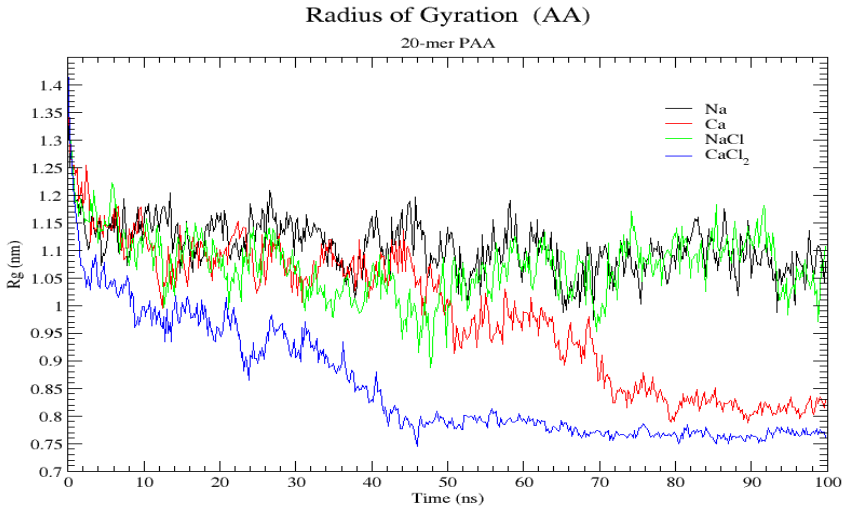


Figure 35 Average curve of the radius of gyration Rg over time of 100 ns for all 4 systems over 3 independent production runs of PAA (fully deprotonated) AA simulation.

Whereas in the system containing $CaCl_2$ ions a sharp conformational change can be seen at the very beginning of the simulation similar to the system with Ca2+ ions within 5 nanoseconds. As seen in Fig 35, PAA with $CaCl_2$ ions shows a constant decrease in R_g for the first 46 nanoseconds of the simulation time resulting as the PAA continuously keeps folding further as seen in Fig. 35. The bridging effect in PAA can be seen at 48 nanoseconds where PAA folds almost completely and tends to a value of $R_g = 0$. PAA remains in the folded conformation for the remainder of the simulation time. For the systems containing Na, and NaCl ions the value of R_g remains nearly the same for the time-period of 22 ns after which a decrease in the value can be seen for NaCl. After short decrease, the R_g value increases to its final value after 50 ns of MD simulation. Fig. 35 shows the conformational change of PAA, for $CaCl_2$ the R_g value decreases substantially due to the bridging effect between the ends of the monomer which leads to this significant drop of R_g value. For $CaCl_2$, the R_g value starts to decrease within the initial 10ns of the simulation and goes on to decrease till around 47 ns reaching its final value and stays stable till the end of the simulation. The average R_g value fluctuates around 1.1 nm until the end of the simulation.

5.1.4 End–to–End distance

Table 8 Average end-to-end distance values calculated over 3 independent production runs

System	*Average (nm)*
Na	2.8027660
Ca	2.2487405
NaCl	2.7691287
$CaCl_2$	1.2491734

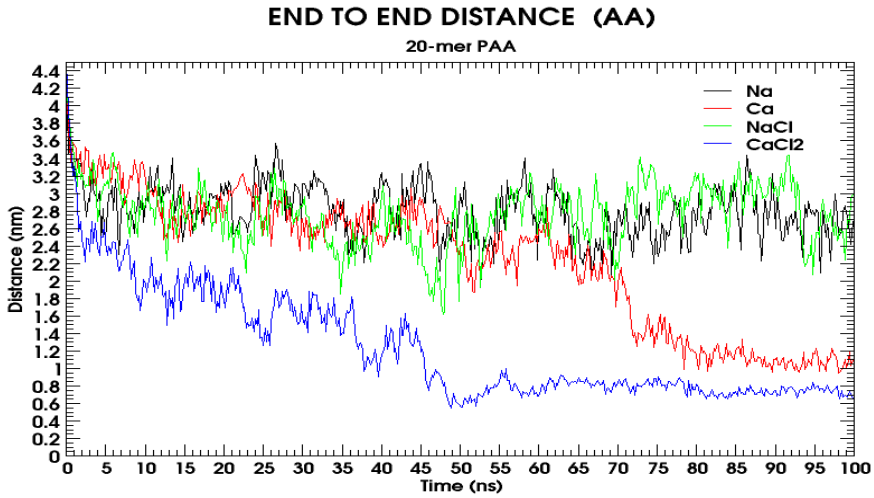


Figure 36 Average curve of the end-to-end distance R over time of 100 ns for all 4 systems over 3 independent production runs of PAA (fully deprotonated) AA simulation.

The time averaged *R* for our simulated 20-mer model of PAA is presented in the Table 8, which is in between the previously reported values for *R*.

The change of the distance from head to tail group of PAA behaves similar to the change of the radius of gyration. For the first 43 ns there is no major difference in the average curve for Na, Ca2+ and NaCl systems production runs, as they fluctuate around 2.8 nm. While the average end-to-end distance for $CaCl_2$ steadily decreases over a period of first 50 ns and then stabilizes at 1.2 nm until the end of the simulation, a sudden decrease in the system containing Ca2+ ions occurs at 71 ns as PAA folds because of that the distribution of the end-to-end distance shown in Figure 19. Snapshots taken from MD trajectory for run_2. PAA is shown as sticks with carbon (cyan) and oxygen (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: sodium ions (violet) and chlorine ions (orange)Root-mean-square dev36 broadens, and two local minima appear at 2 nm and 1.4 nm. For $CaCl_2$ system only one minima is observed at 0.6 nm. As PAA relaxes back to a more stretched conformation in the system containing NaCl ions at 47 ns after another 10 ns the end-to-end distance reaches its former value.

5.1.5 Root-mean-square deviation

Table 9 Average RMSD values calculated over 3 independent production runs

System	*Average (nm)*
Na	0.6067087
Ca	0.7686634
NaCl	0.6352509
$CaCl_2$	0.9496047

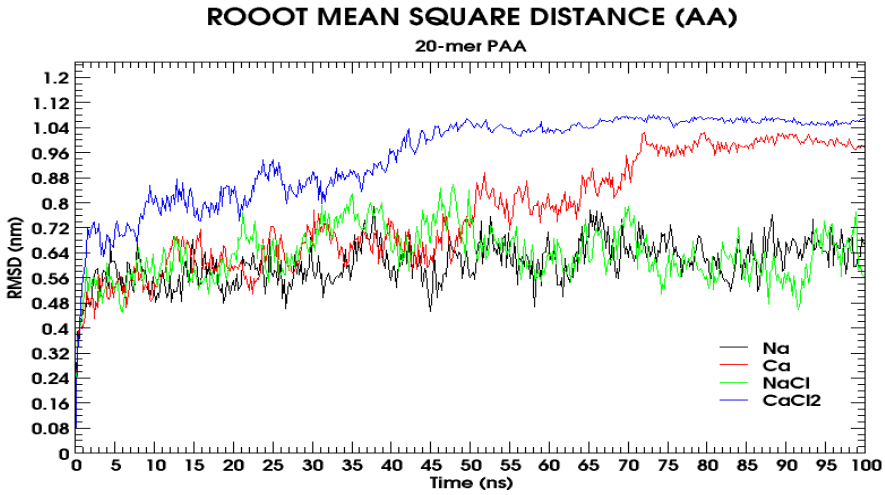


Figure 37 Average curve of the backbone RMSD over time of 100 ns for all 4 systems over 3 independent production runs of PAA (fully deprotonated) AA simulation

The backbone rmsd of PAA with sodium and sodium chloride ions and water fluctuates around an average value of 0.64 ± 0.004 nm (see Fig. 37). System containing calcium and calcium chloride ions show drastic increase in rmsd. PAA shows a steady increase in RMSD continuously through the 100 ns for $CaCl_2$ system reaching a stable value at around 65 ns. In the presence of calcium ion, again PAA shows a steady increase in rmsd and has a sharp increase in RMSD value at 45 ns stabling at a value of 0.76 nm at 72 ns until the end of the simulation. This is in agreement with the results of R_g and R.

5.1.6 Solvent accessible surface area

The result of solvent accessible surface area is in complete agreement with the previously calculated radius of gyration and end to end distance (see Fig. 35, 36). Due to the drastic bending and folding of PAA at 70 ns with calcium ions the SASA value has a sharp decrease (see Fig. 38). In presence of calcium chloride ions, the value R_g and R consistently decreased over the simulation period of 100 ns. This effect can be clearly seen by the value of SASA for calcium chloride system where the value has a consistent decrease from the initial value of 18.96 nm to15.49 nm at the end of 100 ns.

Table 10 Average SASA values calculated over 3 independent production runs

System	*Average (nm^2)*
Na	18.5317199
Ca	17.8639613
NaCl	18.3685279
$CaCl_2$	16.8614613

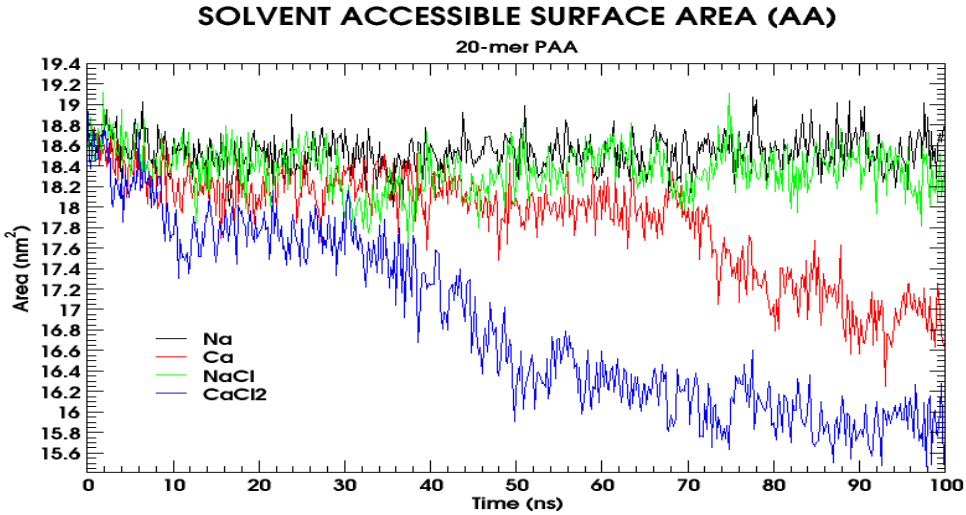


Figure 38 Average curve of the SASA over time of 100 ns for all 4 systems over 3 independent production runs of PAA (fully deprotonated) AA simulation

5.1.7 Hydrogen Bonds

Again, the conformational change occurring in the system containing Ca2+ and $CaCl_2$ ions can be observed in Fig. 39 as a decrease after 4 ns of simulation time. As the ends of the chain tend to fold in attendance of calcium ions, the number of free carboxylate ions decreases and thus the number of hydrogen bonds. The number of hydrogen bonds (N_{hb}) for Na and NaCl fluctuate around 116 during the whole simulation, whereas N_{hb} reaches a local minimum of around 102 during for NaCl.

Table 11 Average hydrogen bonds calculated over 3 independent production runs

System	*Average*
Na	116.4746667
Ca	93.8000000
NaCl	115.1119999

$CaCl_2$	85.7066668

The average N_{hb}, calculated as time average, is 114. As we can see from the RDF between carboxylate oxygens and water hydrogen atoms (see), every oxygen atom is surrounded by nearly 3 hydrogen atoms for Na and NaCl system and 2 hydrogen atoms for Ca and $Cacl_2$ system, so there should be around 80 hydrogen bonds for systems containing calcium ions.

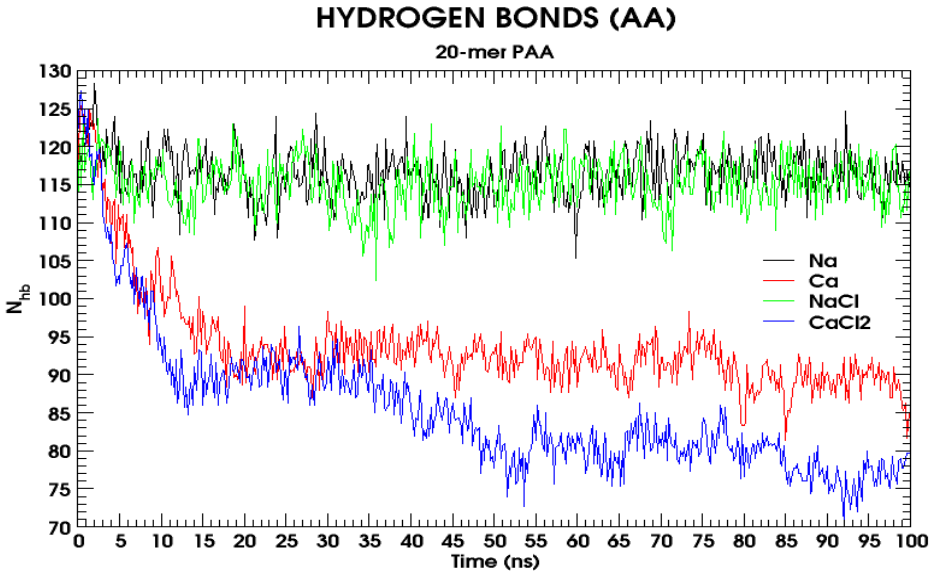


Figure 39 Average curve of the change of hydrogen bonds over time of 100 ns for all 4 systems over 3 independent production runs of PAA (fully deprotonated) AA simulation

5.1.8 Radial Distribution Functions

5.1.8.1 RDF between the centre of mass of the PAA residues and the oxygen atoms of water molecules

The sharp peak at 0.26 ± 0.002 nm in indicates a well-structured solvation shell with ordered distribution of water molecules around the polymer chain.

This in not in good agreement with the results of Sulatha *et al.*[30] They observed a first peak at 0.37 nm for FF-1 and 0.35 nm for FF-2. Biermann *et al.*[25] reported a first peak at 0.35 nm. As our system has provided accurate results in agreement with the literature for all the other properties under investigation, the system setup cannot be doubted for this disagreement. Further, investigation is needed to look into the possible reasons for his mismatch of results and will be taken up in future studies.

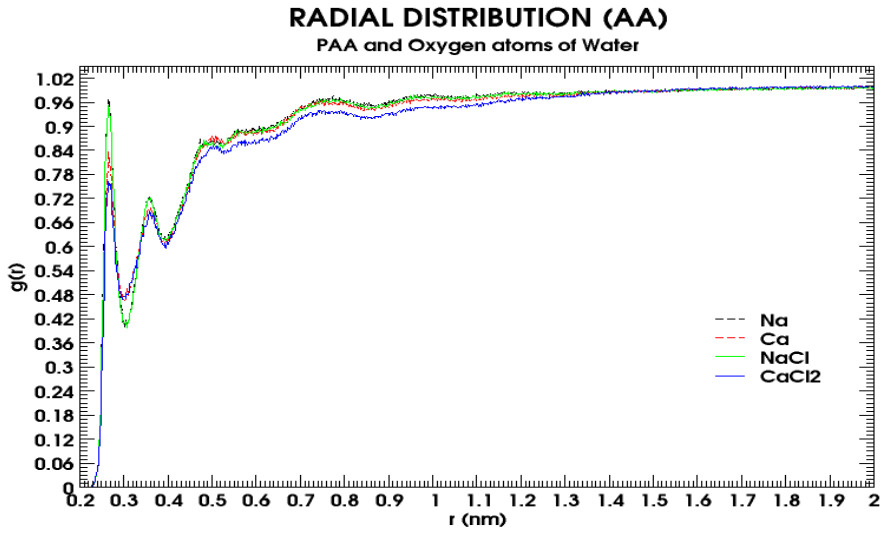


Figure 40 RDF between PAA (center of mass) and water oxygens (OW) for all for all 4 systems over 3 independent production runs of PAA (fully deprotonated) AA simulation

5.1.8.2 RDF between oxygen atoms of carboxylate group of PAA and the hydrogen atoms of water molecules

The RDF between the carboxylate oxygen atoms and water hydrogen atoms in show two peaks at 0.168 ± 0.002 nm and 0.316 ± 0.002 nm. This is similar to the hydrogen bonding distance in bulk water. which is in agreement with the Sulatha *et al.*[30]

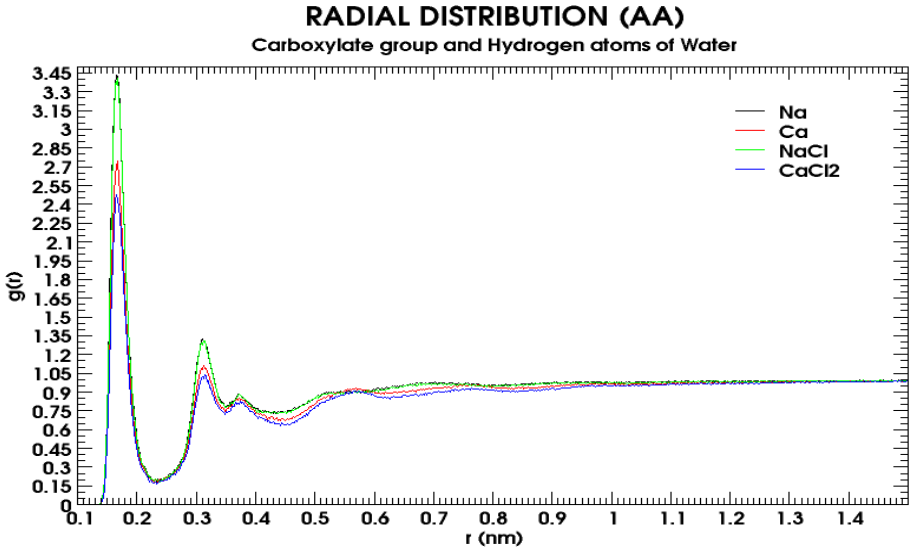


Figure 41 RDF between carboxylate oxygens of PAA and water hydrogens (HW) for all for all 4 systems over 3 independent production runs of PAA (fully deprotonated) AA simulation.

5.1.8.3 RDF between oxygen atoms of carboxylate group of PAA and the oxygen atoms of water molecules

The RDF between of the oxygen atoms of the carboxylate group with respect to water hydrogens shows a strong peak at 0.26 ± 0.002 nm (see). A shoulder peak at 0.464 ± 0.002 nm follows. Again, the results of Sulatha *et al.*[30] match ours. There RDFs

show two 2 peaks at 0.28 and 0.48 nm, respectively 0.25 and 0.45 nm based on the simulation with FF-1 and FF-2.

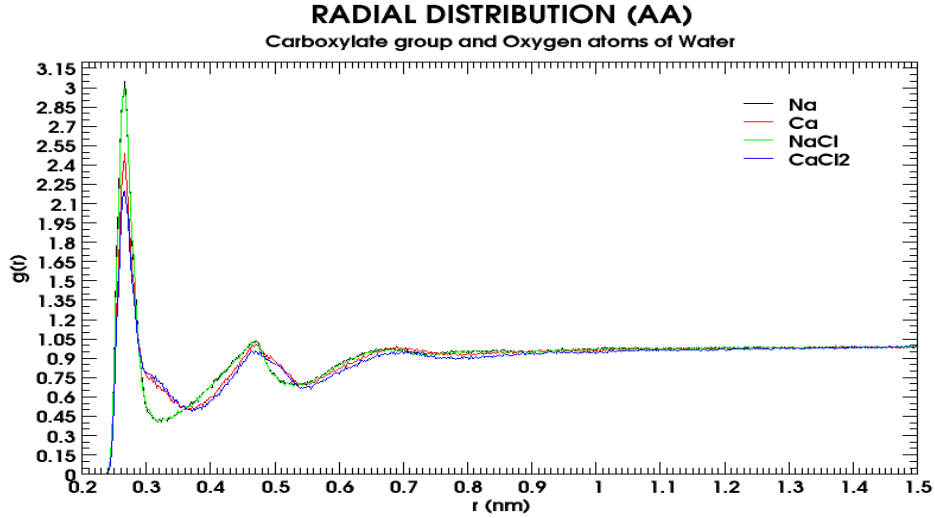


Figure 42 RDF between carboxylate oxygens of PAA and water oxygens (OW) for all for all 4 systems over 3 independent production runs of PAA (fully deprotonated) AA simulation

5.1.8.4 RDF between oxygen atoms of carboxylate group of PAA and the ions

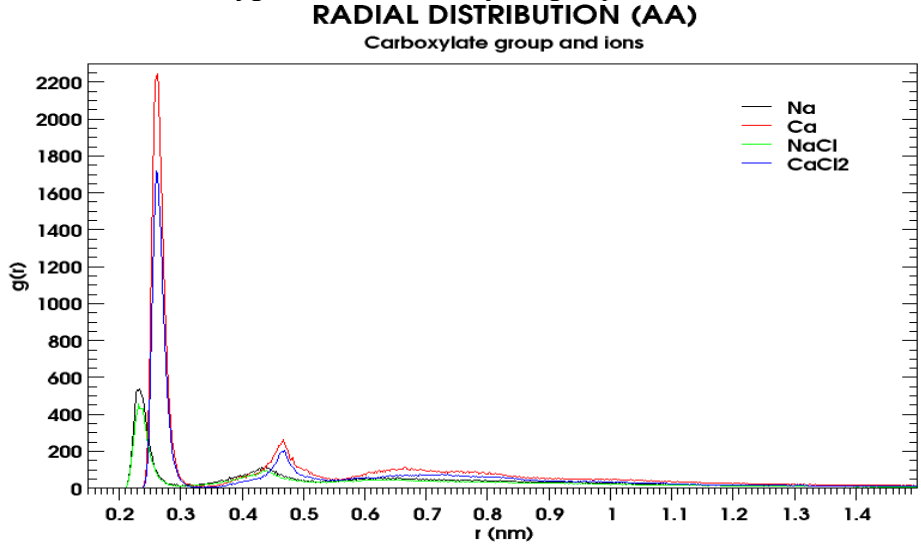


Figure 43 RDF between carboxylate oxygens of PAA and ions for all for all 4 systems over 3 independent production runs of PAA (fully deprotonated) AA simulation

As shown by Fig. 43, the RDF between the carboxylate oxygens and sodium counterions exhibits a strong first peak at 0.232 ± 0.002 nm, suggesting a preferred condensation of sodium ions on to the chain backbone of PAA, for calcium ions and carboxylate oxygens exhibits a strong first peak at 0.26 ± 0.003 nm and a weaker shoulder peak at 0.47 ± 0.002 nm. Sulatha *et al.*[30] discovered a similar behaviour for simulations with FF-2 parameter set. Their RDFs show a first peak at ~0.21 nm. The higher value of RDF for the system containing calcium counterion clearly shows better affinity of PAA towards the divalent counterions.

5.2 ANALYSIS OF TRAJECTORIES FROM COARSE-GRAINED MD SIMULATIONS OF FULLY DEPROTONATED 20-MER MODEL OF POLYACRYLIC ACID (PAA)

PAA (fully deprotonated) is simulated in a system containing polarizable water model from MARTINI and 4 different ions individually (see Table 5). Polymer model was generated using the mapping scheme described in section 3.2.1 using MARTINI scheme for modelling of polymer.

The four generated boxes each include one PAA model short-chain model. Periodic boundary conditions were applied in all directions. For PAA simulations, cubic boxes with 9.8 nm edge length were used for the short-chain models. The new refined polarizable water model from MARTINI was used to solvate the system. In order to neutralize the system charge of the polymer, water molecules were replaced by ions according to the number of monomers in the polymer chain (see Table 5). In order to account for the ionic strength of the synthetic tap water (15°dH) used in experiments additional water molecules were exchanged by sodium and chloride ions. The setup for each simulated system is summarized in Table.4.

All simulations were performed on the RWTH Compute Cluster using GROMACS 2019 simulation package[116]. All systems were first energy minimized using steepest descent algorithm with 500,000 integration steps or until the maximum force on any atom in the system did not exceed a value of 10.0 kJ/mol/nm. For neighbor searching, Verlet cutoff-scheme was used having short range Van der Waals cut-off of 1.0 with periodic boundary conditions. For electrostatic forces, Particle-mesh Ewald method of order 8 was used having short range cut-off of 1.2 keeping relative dielectric constant 2.5 and relative dielectric constant of the reaction field equal to 1. For Van der waals forces, twin range cut-offs with neighbor list cut-off and VdW cut-off are used. For the treatment of long-range electrostatic interaction and long-range dispersion corrections for energy and pressure was applied. Grid dimensions are controlled with Fourier spacing of 0.15. The relative strength of the Ewald-shifted direct potential at rcoulomb is given by 'ewald-rtol'. For doing PME for VdW-interactions, ewald-rtol-lj is used to control the relative strength of the dispersion potential at rvdw. Both the values are kept default 1e-5 and 1e-3 respectively.

After heating, the systems were equilibrated for 50 ns each at 300 K using the velocity rescale thermostat (coupling time 1 ps) according to Bussi *et al.*[117] in NVT ensemble and at 1 bar using Berendsen barostat[118, 119] (coupling time 12 ps and compressibility $3 \cdot 10^{-4}$ bar^{-1}) in NPT ensemble. During the equilibration, position restraints on every polymer bead were activated. Three independent production runs, 500 ns each, were performed for every system in NPT ensemble using Parrinello-Rahman extended-ensemble pressure coupling and Verlet cutoff scheme with a cutoff distance of $r_{cut} = 1.2$ nm for the Lennard-Jones interactions according to a recent publication[120]. The Particle Mesh Ewald (PME) method[112] with a grid spacing of 0.15 nm in agreement with the parameters employed during the parameterization of the refined polarizable water model[105]. Consistently, a time step of 20 fs was used. The neighbor list was updated every 10 steps using the Verlet neighbor search (VNS) algorithm[120]. Bond lengths were held constant by the LINCS algorithm[113, 121] and the background permittivity was set to $\varepsilon_{bg} = 2.5$ as recommended when using the polarizable water model[106, 120]. The final box sizes are shown in Table.4.

A chain length of 20 monomer of PAA (fully deprotonated) is simulated in a system containing water and different quantity of monovalent and divalent ions (sodium, calcium, sodium chloride and calcium chloride) (see Table 12). The system is equilibrated at 300k and 1 bar of pressure. The system is simulated for 500 ns with a time step of 20 femtoseconds. System configuration for production run is provided in appendix 9.15-9.2.18.

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

The simulation results are validated with the results obtained from AA simulations, literature and experimental results.

Table 12 Number of ions in Na & Ca (salt free) and NaCl & $CaCl_2$ (excess salt) systems of 20-mer PAA

Salt	***Number of ions***	
	Na/Ca	***Cl***
Na	20	-
NaCl	27	7
Ca	10	-
$CaCl_2$	17	14

5.2.1 Visualization of PAA interaction with Na, Ca, NaCl & $CaCl_2$ using VMD

Some representative snapshots of the MD simulations of the PAA MARTINI models are shown in Fig.44-47 for short 20-mer chain model with different ions. The snapshots were taken after 500 ns of MD simulation. The water molecules have not been presented in the images for better visualisation of the polymer chain and its interaction with polymer. The PAA chain is shown as stick model. Snapshots from each system can be seen on the following pages.

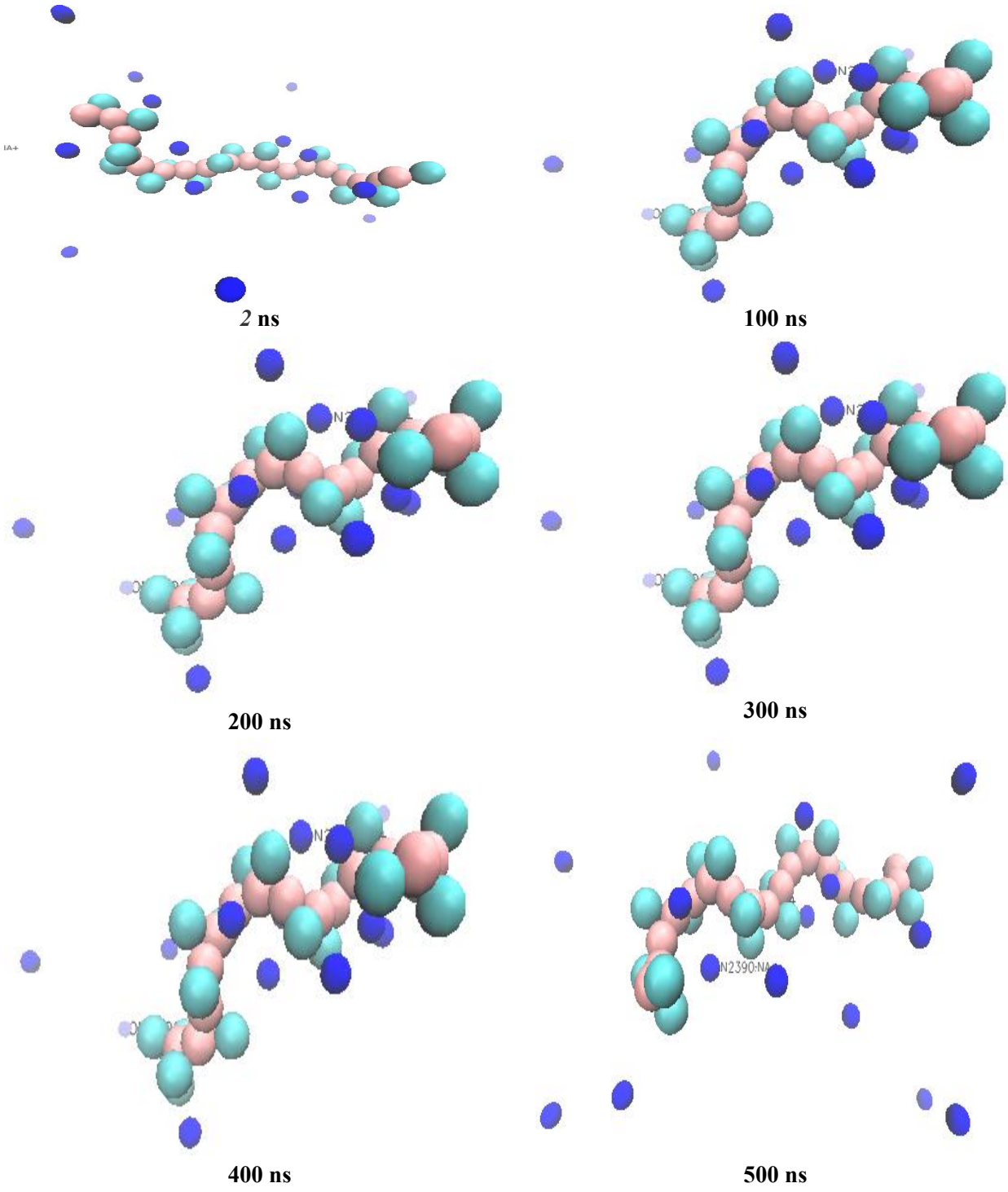


Figure 44 Snapshots taken from MD trajectory of 20-mer MARTINI model of fully deprotonated PAA with Na ions and polarized water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (500 ns). PAA is shown as beads with SC1 (pink) and SQa (cyan). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: sodium ions (violet).

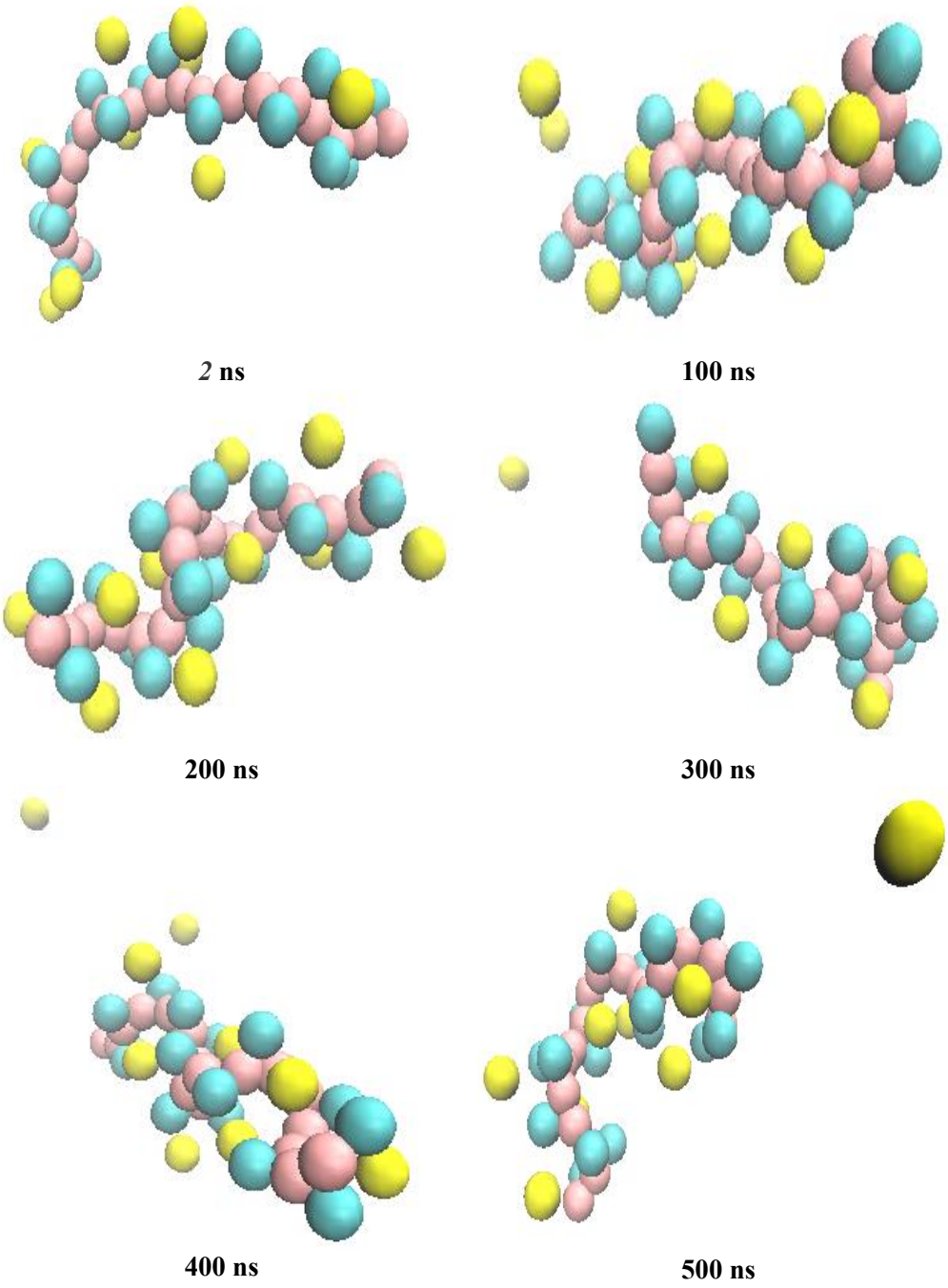


Figure 45 Snapshots taken from MD trajectory of 20-mer MARTINI model of fully deprotonated PAA with Ca2+ ions and polarizable water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (500 ns). PAA is shown as beads with SC1 (pink) and SQa (cyan). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: Calcium ions (yellow).

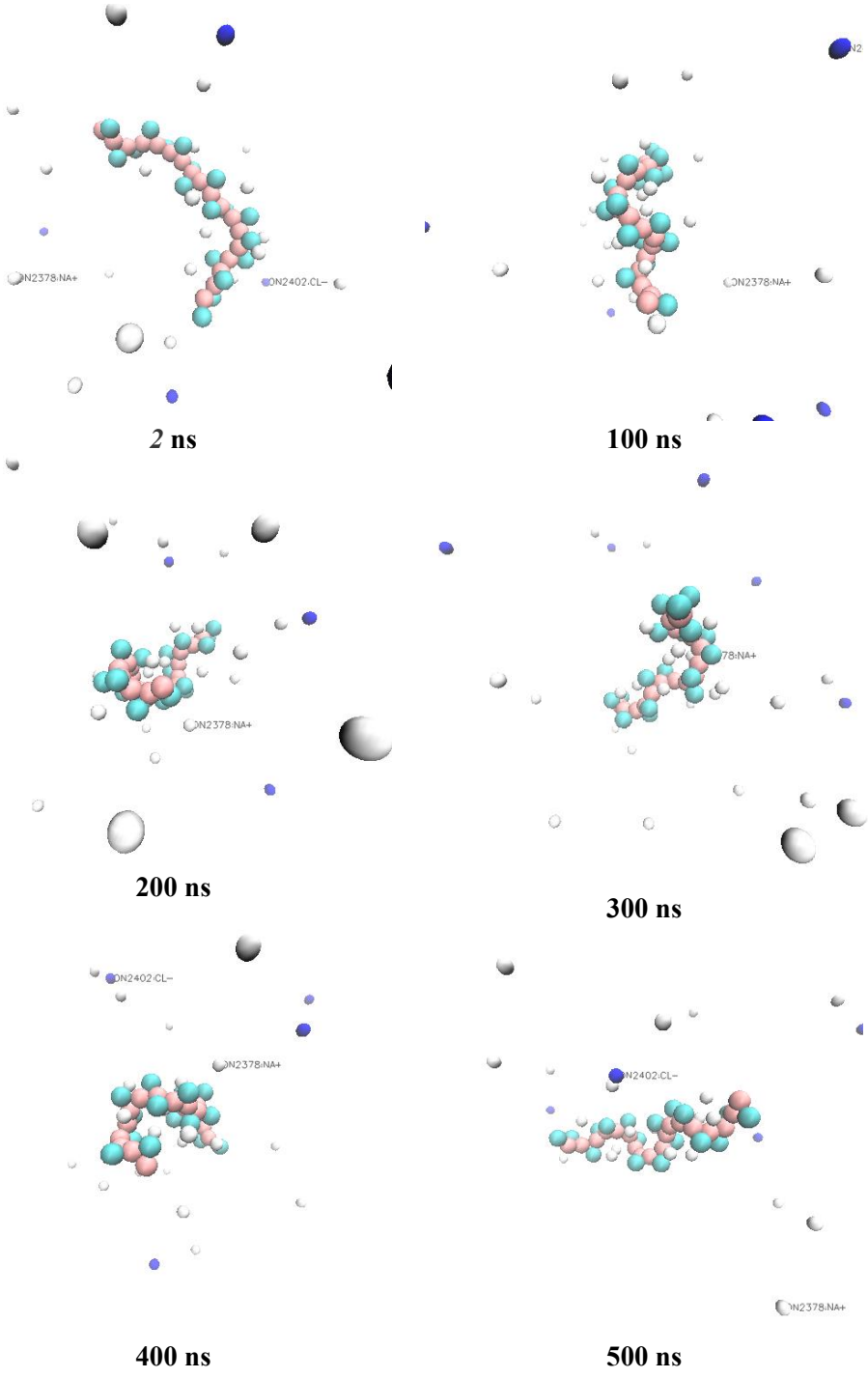


Figure 46 Snapshots taken from MD trajectory of 20-mer MARTINI model of fully deprotonated PAA with NaCl ions and polarizable water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (500 ns). PAA is shown as beads with SC1 (pink) and SQa (cyan). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: sodium ions (silver) and chlorine (indigo)

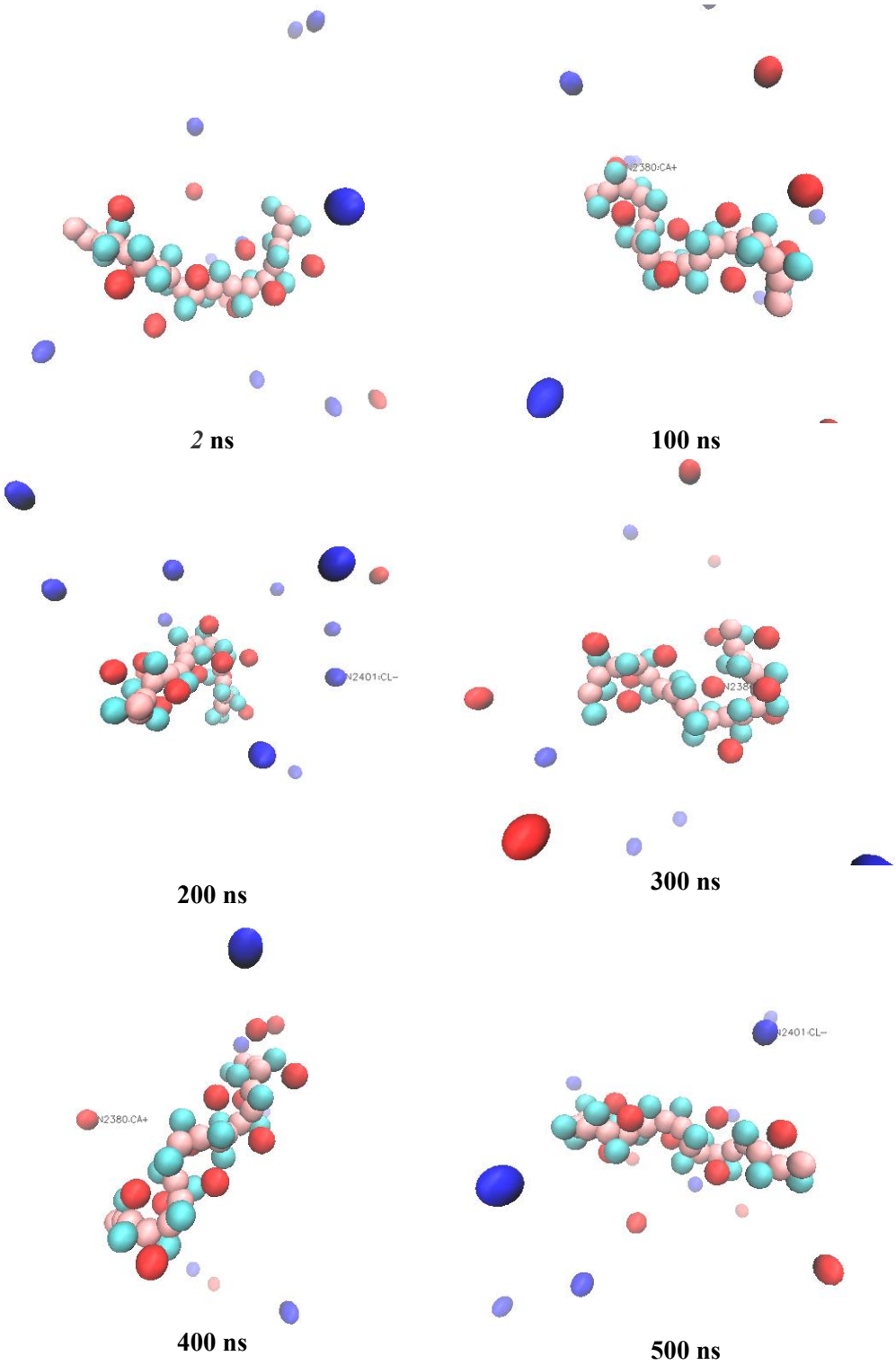


Figure 47 Snapshots taken from MD trajectory of 20-mer MARTINI model of fully deprotonated PAA with $CaCl_2$ ions and water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (500 ns). PAA is shown as beads with SC1 (pink) and SQa (cyan). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: calcium ions (red) and chlorine ions (violet).

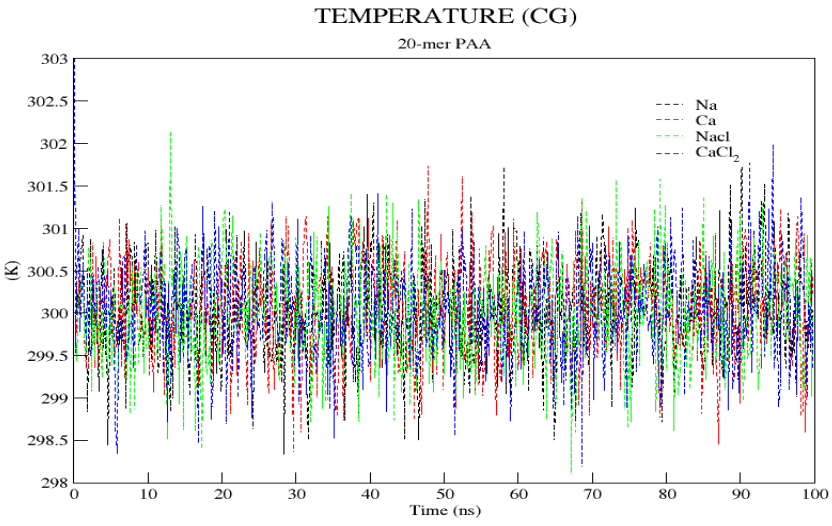
5.2.2 Quality assurance

First, we check again if the system is well equilibrated. As shown in Fig 48a, 48b, the average temperature and pressure, calculated as a time average from all 3 single independent runs for each system, fluctuate steadily. The average temperature and pressure of all 3 independent production runs are presented in the Table 13. All the systems present an average temperature and pressure of 300.0 ±0.1K and 0.3 ±1.34 bar respectively. The average energies (see Fig. 48c, 48d, 48e) of our system are constant, the kinetic energy is $2.08x10^4 \pm 0.5x10^{-3}$ kJ/mol, the potential energy is $-7.49x10^4 \pm 0.6x10^{-3}$ kJ/mol except the system with calcium chloride ions showing a higher level of potential energy $-8.30x10^4$ KJ/mol which is well inside the limits of to be considered as a stable system and the total energy is $-1.273x10^6 \pm 0.02$ KJ/mol. The values are constant over the whole trajectories, this indicates a well equilibrated and stable system.

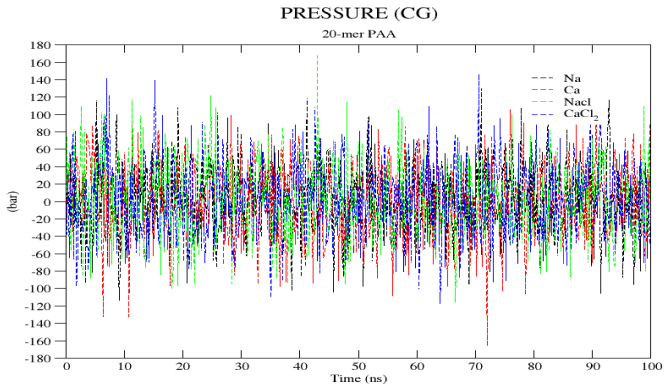
Table 13 System properties calculated as a time average from all 3 single independent runs

Property	Average value of over 3 independent production runs			
	Na	**Ca**	**NaCl**	**$CaCl_2$**
Temperature (K)	300.1699	300.032	300.0345	300.0255865
Pressure (bar)	-2.08983	1.371777	1.0130221	1.178518
Potential Energy (KJ/mol)	-7.72E+04	-7.91E+04	-7.85E+04	-8.30E+04
Kinetic Energy (KJ/mol)	2.09E+04	2.09E+04	2.08E+04	2.08E+04
Total Energy (KJ/mol)	-5.63E+04	-5.82E+04	-6.22E+04	-1.27E+06

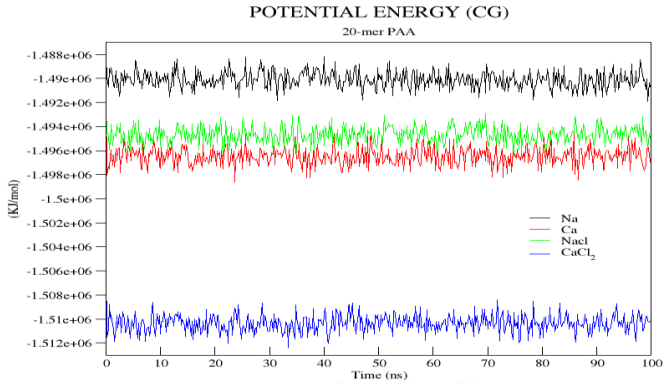
a)



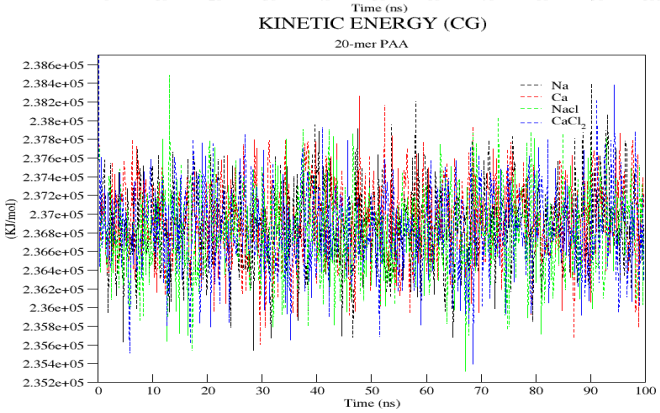
b)



c)



d)



e)

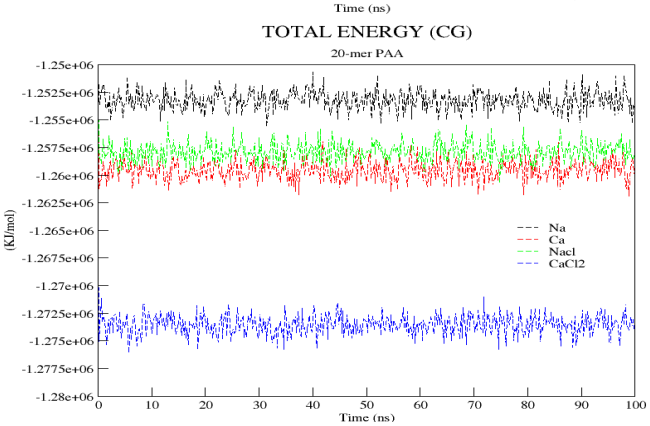


Figure 48 Average energy profiles of 3 independent AA production runs (500 ns) for 20-mer model of fully deprotonated PAA with Na, Ca, NaCl, and $CaCl_2$ ions and water

5.2.3 Radius of Gyration

Table 14 Average Radius of gyration values calculated over 3 production runs

System	*Average (nm)*
Na	0.9979
Ca	0.9383
NaCl	0.9974
$CaCl_2$	0.9352

The radius of gyration R_g of the fully deprotonated PAA CG models is shown in Fig. 49. The time averaged R_g from all 3 single independent runs for each of the four systems varies between 0.93 nm and 0.99 nm in salt free and excess salt conditions. This is lower than the R_g value of PAA obtained from the all atomistic model because the chain backbone is smaller (2 atoms per monomer instead of 3 atoms).

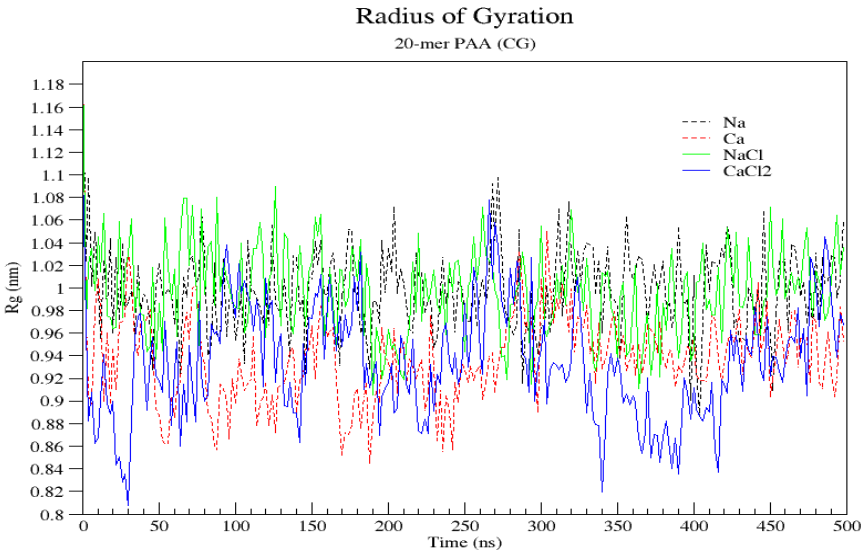


Figure 49 Average curve of the radius of gyration Rg over time of 500 ns for all 4 systems over 3 independent production runs of PAA (fully deprotonated) CG simulation

PAA MARTINI model in salt free system containing sodium ions shows similar conformational changes throughout the first 100 ns having an average R_g value of 1.058 nm which is similar to AA results. Similarly, in the system containing calcium, sodium chloride and calcium chloride ions the average R_g value for first 100 ns correspondence to 0.996, 1.045 and 0.932 nm respectively. These values are in close proximity to AA models. In the remaining 400 ns PAA does not go through significant conformational changes in the presence of Na and NaCl ions but rapidly goes from stretched to bended and folded conformation throughout the remaining simulation. In systems with divalent ions (Ca and $CaCl_2$) bending and folding similar to AA model can be seen in first 100 ns and a constant change in conformation fluctuating between stretched and folded can be very well seen from the Fig. 49 till the end of the simulation. PAA reaches is minimum value of 0.807 nm at

24nm and values closer to it at 330 ns in presence on $CaCl_2$ ions. Hence justifying the results obtained from AA models and the results obtained Tomida *et al.*[122] and Karppi *et al.*[123] showcasing the absorption and conformation effects on polyelectrolytes by divalent ions.

5.2.4 End–to–End distance

The time averaged *R* for our simulated 20-mer MARTINI model of PAA is presented in the Table 15, which is in close proximation with the previously reported values from AA model for PAA short chain model. The values of R from the PAA MARTINI for first 100 ns for system containing sodium, sodium chloride, calcium and calcium chloride are 2.685, 2.408, 2.706, 2.313 nm respectively.

The end-to-end distance for the salt free system in acceptable range in comparison to the values obtained from AA models. Whereas the values for excess salt systems are higher than the values obtained from the AA models. This can be explained by the mapping method used in this study.

Table 15 Average end-to-end distance values calculated over 3 independent production runs

System	Average (nm)
Na	2.6894
Ca	2.4159
NaCl	2.7062
$CaCl_2$	2.3070

As the carboxylate residues are mapped as a single bead, the bond formation between the ions and hydrogen atoms of water with the carboxylate group is weak hence the chain remains in a more stretched conformation.

The change of the distance from head to tail bead of PAA behaves similar to the change of the radius of gyration. Systems containing sodium ions do not present major fluctuation in the value of *R*. Systems with divalent ions (Ca2+ and $CaCl_2$) show drastic variations in the conformation of the chain.

At 25 ns a sudden drop in the value of *R* for $CaCl_2$ can be seen which is in agreement with the radius of gyration. PAA going from folded to stretched conformation in the system containing only calcium ions can be seen at the same time interval. A similar trend, where the system containing Ca2+ and $CaCl_2$ behaving in complete opposite trend i.e. folding and stretching, can be seen throughout the simulation till the end of 500 ns.

Though, as observed in the AA model PAA systems containing divalent ions had a uniform decrease in the value of *R* during 100 ns which our CG models fails to depict and does not present the bridging effect seen in AA simulation at 50 ns (see Fig 36). But our coarse grain model can accurately predict the average value for 100 ns simulation.

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

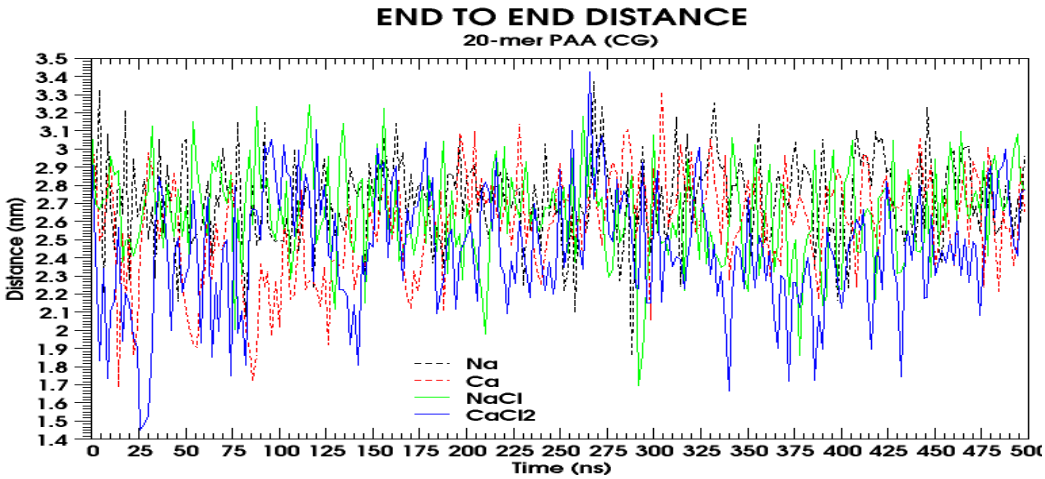


Figure 50 Average curve of the end-to-end distance R over time of 500 ns for all 4 systems over 3 independent production runs of PAA (fully deprotonated) CG simulation.

5.2.5 Root-mean-square deviation

Table 16 Average RMSD values calculated over 3 independent production runs

System	Average (nm)
Na	0.5020
Ca	0.5302
NaCl	0.4896
$CaCl_2$	0.4984

The backbone rmsd of PAA MARTINI with sodium and sodium chloride ions and water fluctuates around an average value of 0.50 ± 0.3 nm (see Fig. 51) while systems with calcium and calcium chloride show rmsd value fluctuating between 0.53±1.5 nm. Salt free system with divalent ions show a higher rmsd during 500 ns of simulation compared to excess salt divalent ion system.

The average value of rmsd obtained from our CG model for the systems containing Na, Ca2+, NaCl and $CaCl_2$ ions over first 100 ns of simulation are 0.503, 0.530, 0.490, 0.497 nm respectively. This is in complete disagreement with our AA simulation results of PAA where $CaCl_2$ system showcased a constant increase in rmsd value during 100 ns run and a similar pattern was seen in salt free Ca2+ system but the values of Ca2+ remained less than the $CaCl_2$ system throughout the 100 ns. The average values predicted by our CG model are far smaller than the values obtained from AA models for the systems containing divalent ions. Though the rmsd results are in agreement with the radius of gyration and end-to-end distance results obtained from CG simulation of PAA.

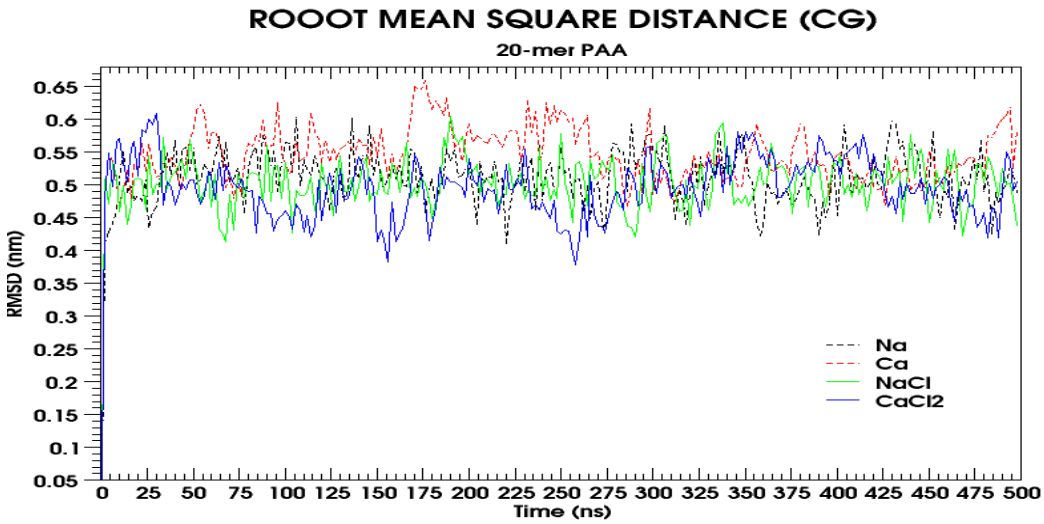


Figure 51 Average curve of the backbone RMSD over time of 500 ns for all 4 systems over 3 independent production runs of PAA (fully deprotonated) CG simulation

5.2.6 Solvent accessible surface area

Table 17 Average SASA values calculated over 3 independent production runs

System	Average (nm^2)
Na	17.3151
Ca	16.8738
NaCl	17.2857
$CaCl_2$	16.9084

The result of solvent accessible surface area is in complete agreement with the previously calculated radius of gyration and end to end distance (see Fig. 49, 51). Due to the drastic bending and folding of PAA at 25 ns with calcium ions the value SASA has a sharp decrease (see Fig. 52). In presence of calcium chloride ions, the value R_g and R consistently fluctuated between there maximum 17.3 nm and minimum 16.2 nm value over the simulation period of 500 ns. This effect can be clearly seen by the value of SASA for calcium chloride system where the value has a consistent decrease from the initial value of 18.96 nm to15.49 nm at the end of 100 ns.

The average values obtained from for the first 100 ns simulation of our model for the system containing Na, Ca, NaCl and $CaCl_2$ ions are 17.315, 16.869, 17.290, 16.90 nm^2. These values are in very close proximation with the values obtained from the AA simulation. (see Table 10)

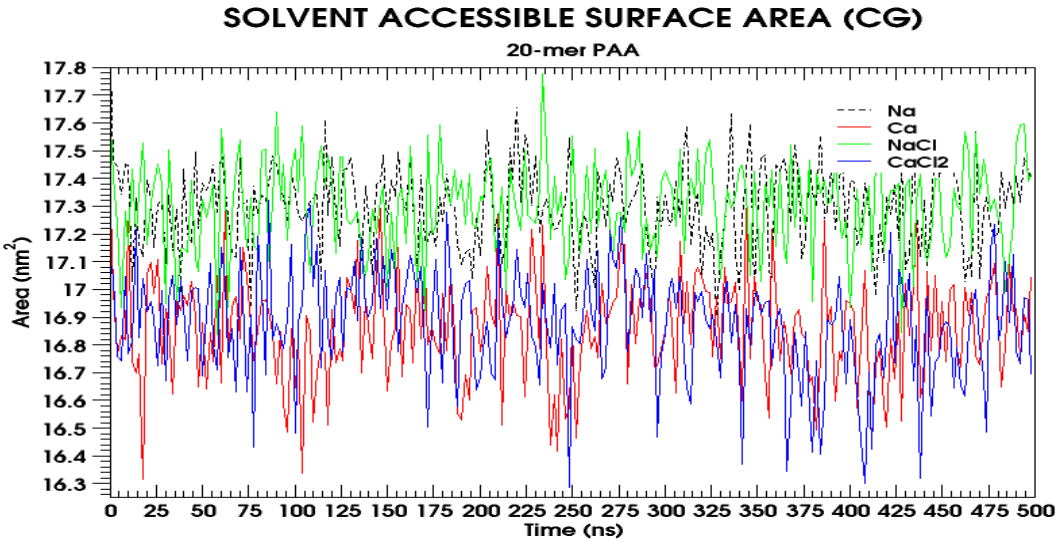


Figure 52 Average curve of the SASA over time of 500 ns for all 4 systems over 3 independent production runs of PAA (fully deprotonated) AA simulation

5.2.7 Radial Distribution Functions

5.2.7.1 RDF between the centre of mass of PAA beads and polarized water beads

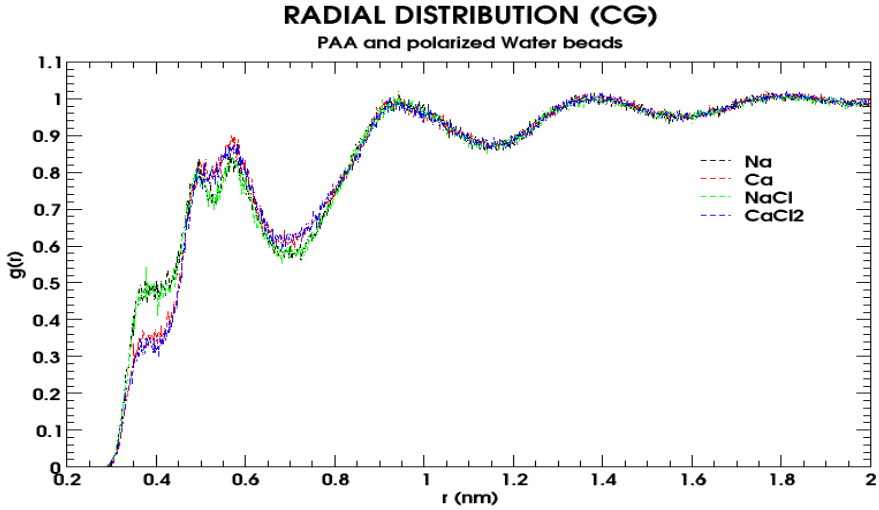


Figure 52 RDF between PAA (center of mass) and polarized water beads (PW) for all for all 4 systems averaged over 3 independent production runs of PAA (fully deprotonated) CG simulation

The sharp peak at 0.584± 0.002 nm in indicates a well-structured solvation shell with ordered distribution of polarized water beads around the polymer chain. This value is bigger than the value obtained in AA simulation. This result can be corrected by adjusting the dielectric constant for each individual simulation.

5.2.7.2 RDF between SQa beads and 'W' of Polarized water bead

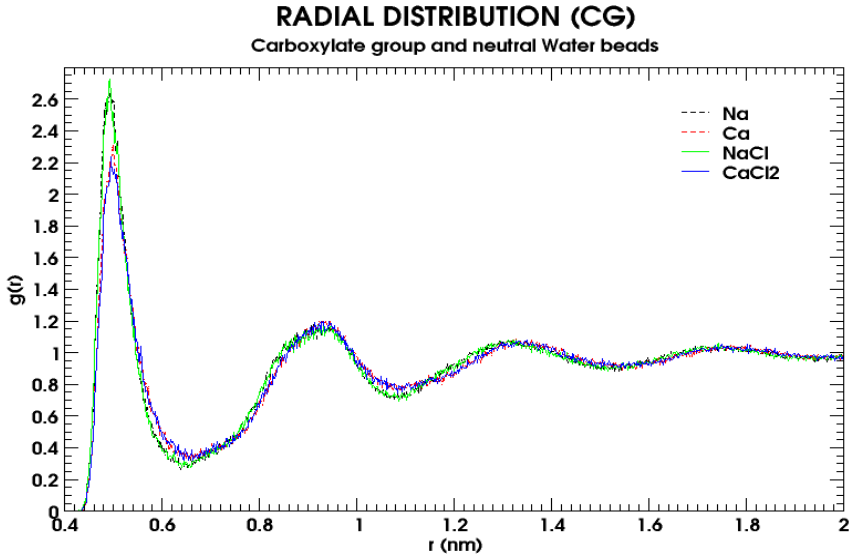


Figure 53 RDF between Carboxylate group (SQa) and LJ interaction site 'W' for all for all 4 systems averaged over 3 independent production runs of PAA (fully deprotonated) CG simulation

The RDF between the SQa beads (carboxylate group) with respect to neutral water beads shows a strong peak at 0.49 ± 0.002 nm (see Fig. 53). A shoulder peak at 0.94 ± 0.002 nm follows

5.2.7.3 RDF between SQa beads and the Polarized water beads

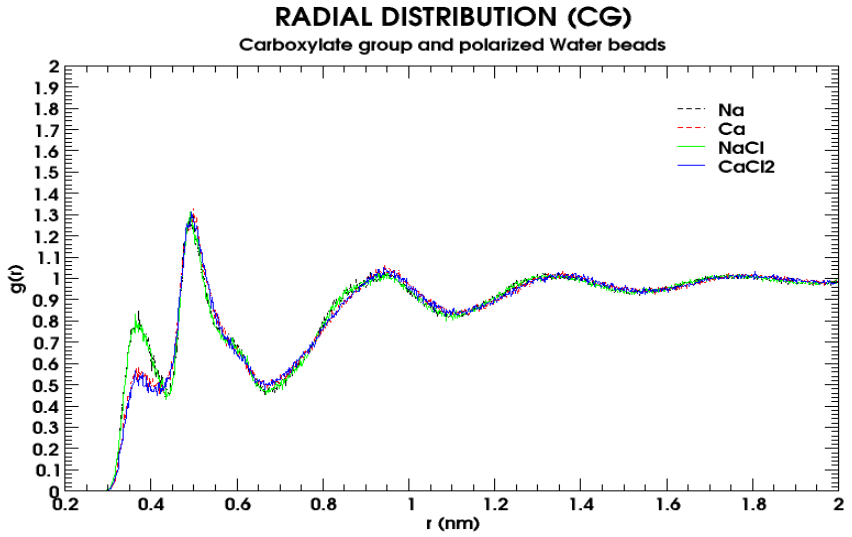


Figure 54 RDF between Carboxylate group (SQa) and PW for all for all 4 systems averaged over 3 independent production runs of PAA (fully deprotonated) CG simulation

The RDF between the SQa (carboxylate group) beads and polarized water beads shown in in show two peaks at 0.36 ± 0.002 nm and 0.49 ± 0.002 nm for sodium and sodium chloride systems and at 0.37 ± 0.002 and 0.49 ± 0.002 for systems containing calcium and calcium chloride ions.

5.2.7.4 RDF between SC1 beads and the Polarized water beads

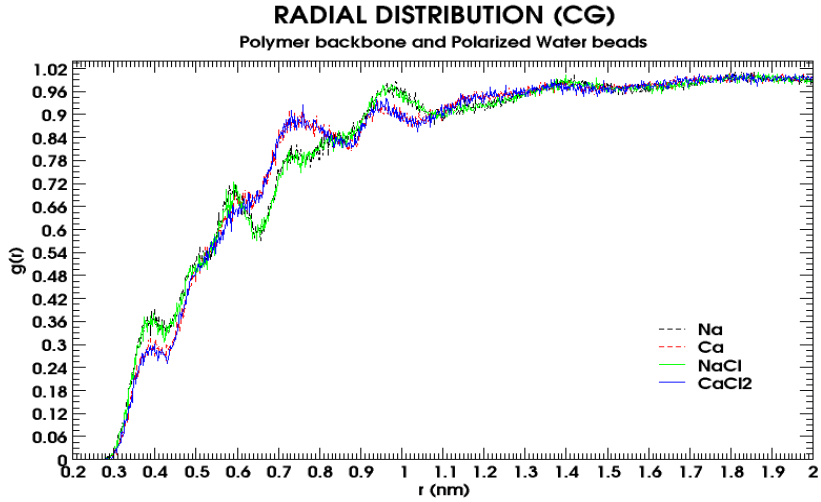


Figure 55 RDF between aliphatic backbone (SC1) and PW for all for all 4 systems averaged over 3 independent production runs of PAA (fully deprotonated) CG simulation

5.2.7.5 RDF between SC1 and 'W' of Polarized water bead

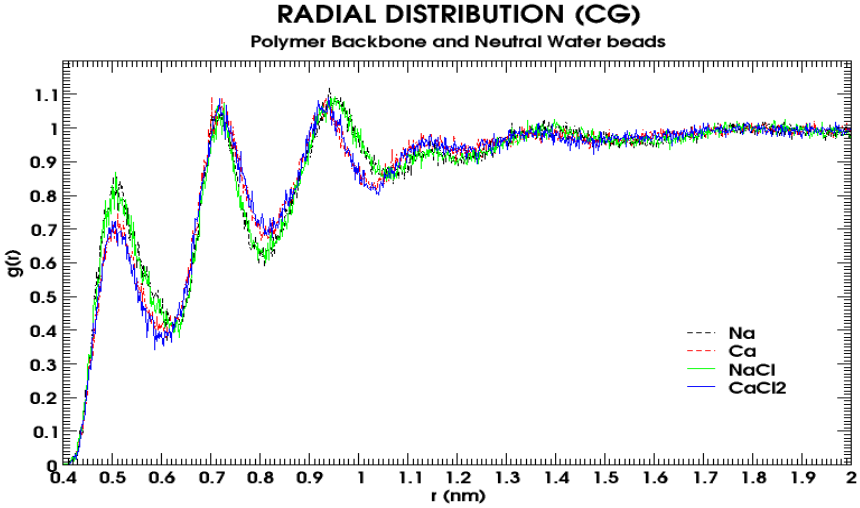


Figure 56 RDF between aliphatic backbone (SC1) and 'W' of polarized water bead for all for all 4 systems averaged over 3 independent production runs of PAA (fully deprotonated) CG simulation

5.2.7.6 RDF between SQa beads and Ions

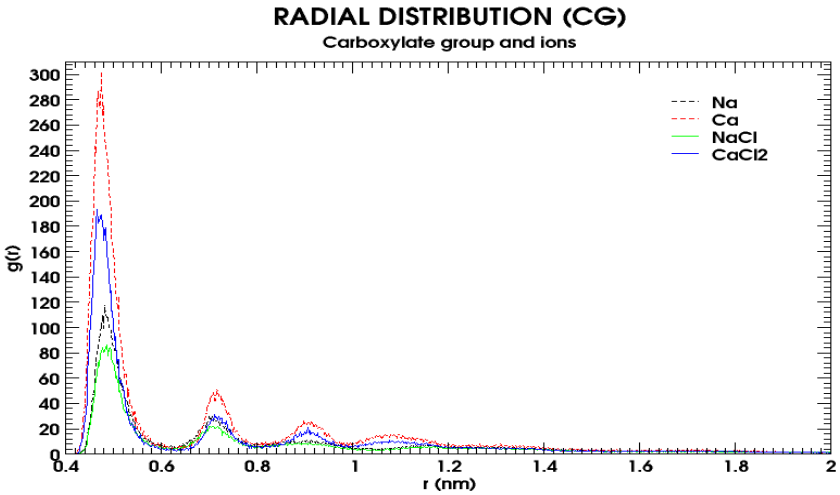


Figure 57 RDF between Carboxylate group (SQa) and Ions for all for all 4 systems averaged over 3 independent production runs of PAA (fully deprotonated) CG simulation

As shown by Fig.57, the RDF between the SQa (carboxylate group) and monovalent and divalent counterions exhibits a strong first peak at 0.47 ± 0.004 nm, suggesting a preferred condensation of calcium ions on to the chain backbone of PAA. A weaker shoulder peak at can be seen at 0.71 ± 0.004 nm.

5.2.7.7 RDF between SC1 beads and Ions

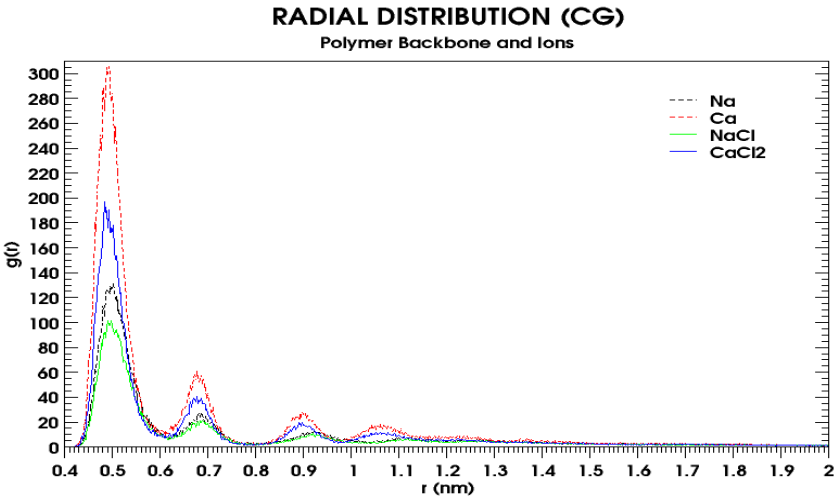


Figure 58 RDF between aliphatic backbone (SC1) and Ions for all for all 4 systems averaged over 3 independent production runs of PAA (fully deprotonated) CG simulation

As shown by Fig.58, the RDF between the SC1 (aliphatic backbone) and monovalent and divalent counterions exhibits a strong first peak at 0.48 ± 0.004 nm, suggesting a preferred condensation of calcium ions on to the chain backbone of PAA. A weaker shoulder peak at can be seen at 0.67 ± 0.004 nm.

5.3 ANALYSIS OF TRAJECTORIES FROM ALL-ATOMISTIC MD SIMULATIONS OF FULLY DEPROTONATED 62-MER MODEL OF POLYACRYLIC ACID (PAA)

The four generated boxes each include one PAA long-chain model with periodic boundary conditions applied in all directions. For PAA simulations, cubic boxes with 14.4 nm edge length were used for the long-chain models. The new refined SPC/E water model was used to solvate the system. In order to neutralize the system charge of the polymer, water molecules were replaced by ions according to the number of monomers in the polymer chain (see Table 18). In order to account for the ionic strength of the synthetic tap water (15°dH) used in experiments additional water molecules were exchanged by sodium and chloride ions. The setup for each simulated system is summarized in Table 18.

Table 18 Number of ions in Na & Ca (salt free) and NaCl & $Cacl_2$ (excess salt) systems of 62-mer PAA

Salt	***Number of ions***	
	Na/Ca	***Cl***
Na	62	-
NaCl	81	19
Ca	31	-
$CaCl_2$	50	38

All simulations were performed on the RWTH Compute Cluster using GROMACS 2019 simulation package[116]. All systems were first energy minimized using steepest descent algorithm with 50,000 integration steps or until the maximum force on any atom in the system did not exceed a value of 1000.0 kJ/mol/nm. For neighbor searching, Verlet cutoff-scheme was used having short range Van der Waals cut-off of 1.0 with periodic boundary conditions. For electrostatic forces, Particle-mesh Ewald method of order 8 was used having short range cut-off of 1.2 keeping relative dielectric constant and relative dielectric constant of the reaction field equal to 1. For Van der waals forces twin range cut-offs with neighbor list cut-off *rlist* and VdW cut-off *rvdw* are used. For the treatment of long-range electrostatic interaction and long-range dispersion, corrections for energy and pressure was applied. Grid dimensions are controlled with Fourier spacing of 0.15. The relative strength of the Ewald-shifted direct potential at rcoulomb is given by 'ewald-rtol' in the mdp. For doing PME for VdW-interactions, ewald-rtol-lj is used to control the relative strength of the dispersion potential at rvdw. Both the values are kept default 1e-5 and 1e-3 respectively.

After heating, the systems were equilibrated for 140 ps each at 300 K using Velocity rescale thermostat along with the velocity Verlet algorithm[124, 125] integrator in NVT ensemble and at 1 bar using Berendsen barostat[118, 119](coupling time 2 ps and compressibility 4.5×10^{-5} bar^{-1}) in NPT ensemble. During the equilibration, position restraints on every backbone atom of the PAA model and on every polymer bond were activated. Three independent production runs, 100 ns each, were performed for every system in NPT ensemble using a Verlet cutoff scheme with a cutoff distance of r_{cut} = 1.2 nm for the Lennard-Jones interactions according to a recent publication[120]. Consistently, a time step of 2 fs was used. The neighbor list was updated every 10 fs using the Verlet neighbor search (VNS) algorithm[120]. Bond lengths were held constant by the LINCS algorithm[113, 121] and Parrinello-Rahman extended-ensemble pressure coupling was used. The final box sizes are shown in Table.4.

A chain length of 62 monomer of PAA (fully deprotonated) is simulated in a system containing water and different quantity of monovalent and divalent ions (sodium, calcium, sodium chloride and calcium chloride) to simulate salt free and excess salt solutions (see

Table 18). The system is equilibrated at 300k and 1 bar of pressure. The system is simulated for 100 ns with a time step of 2 femtoseconds. System configuration for production run is provided in appendix 9.10-9.13. The simulation results are validated with the results from literature and experimental results.

5.3.1 Visualization of 62-mer PAA interaction with Na, Ca, NaCl & $CaCl_2$ using VMD

Some representative snapshots of the MD simulations of the simulated system consisting of PAA 62-mer all atomistic model of fully deprotonated PAA with Na, Ca2+, NaCl and $CaCl_2$ ions and water are shown in Fig. 59, 60, 61, 62, respectively. The PAA chain is shown as stick model. The PAA chain is stretched conformation in the first snapshot but folds shortly after. The snapshots are available on the following pages.

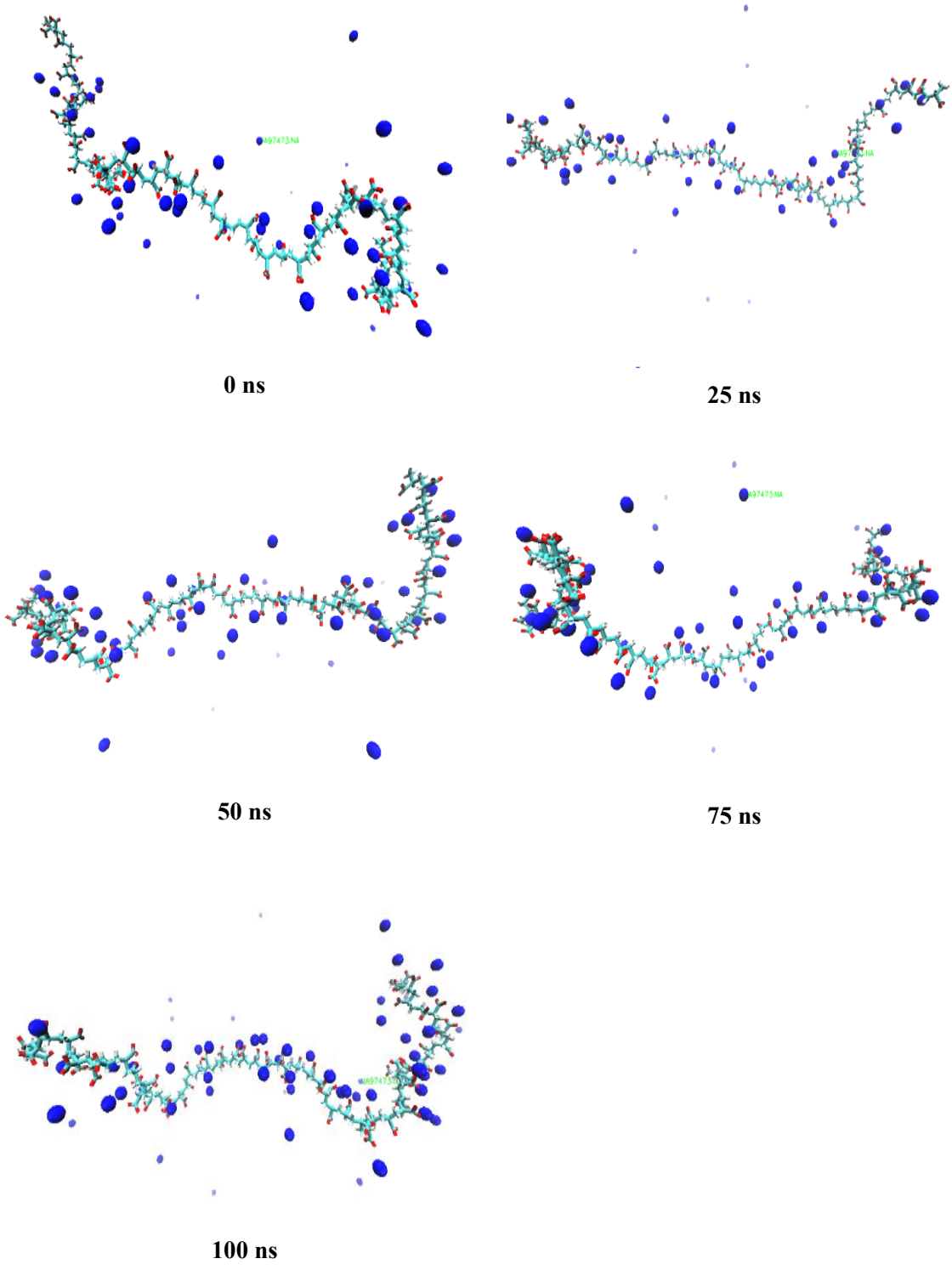


Figure 59 Snapshots taken from MD trajectory of 62-mer model of fully deprotonated PAA with Na ions and water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (100 ns). PAA is shown as sticks with carbon (cyan) and oxygen (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: sodium ions (violet).

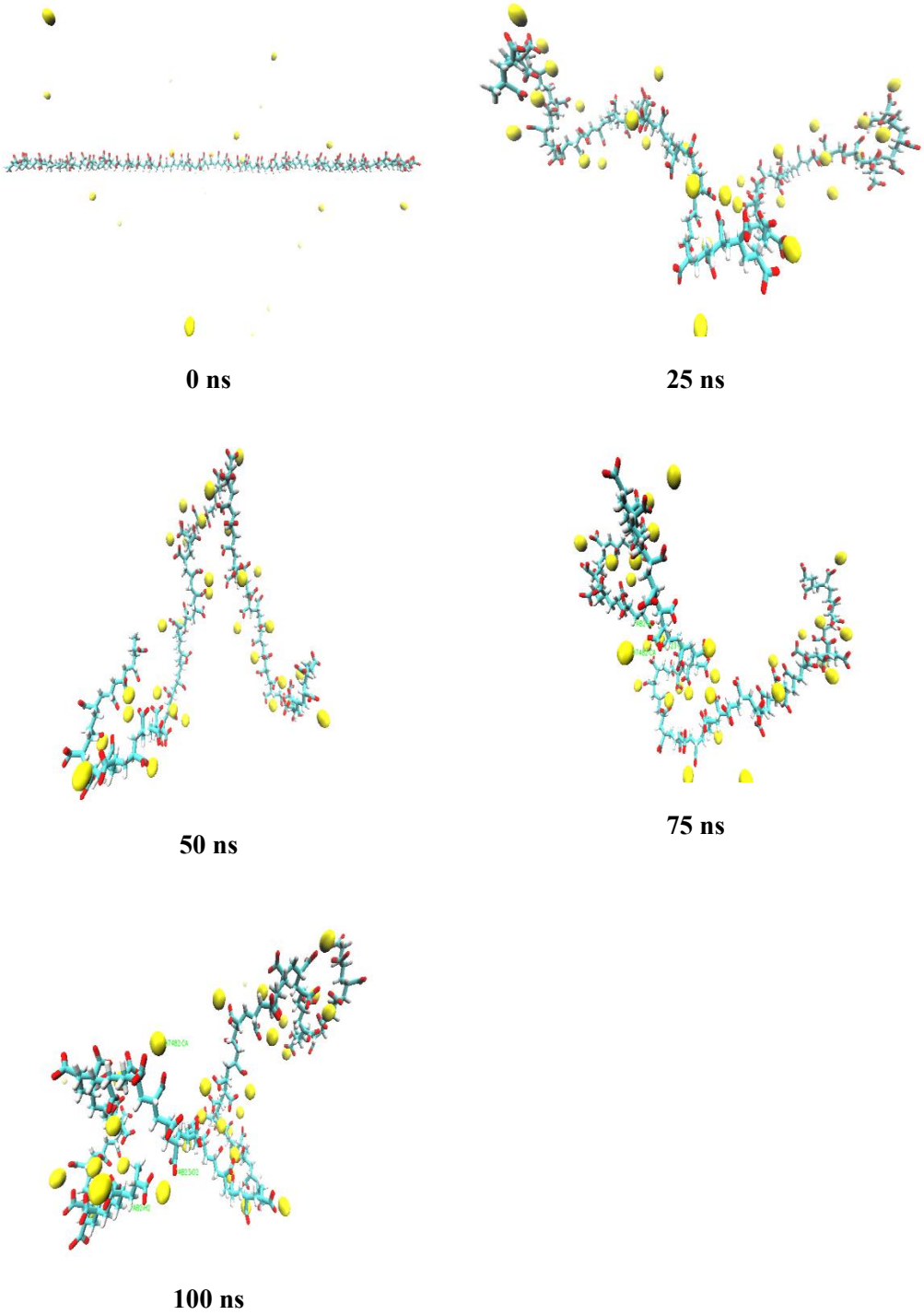


Figure 60 Snapshots taken from MD trajectory of 62-mer model of fully deprotonated PAA with Ca ions and water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (100 ns). PAA is shown as sticks with carbon (cyan) and oxygen (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: Calcium ions (orange).

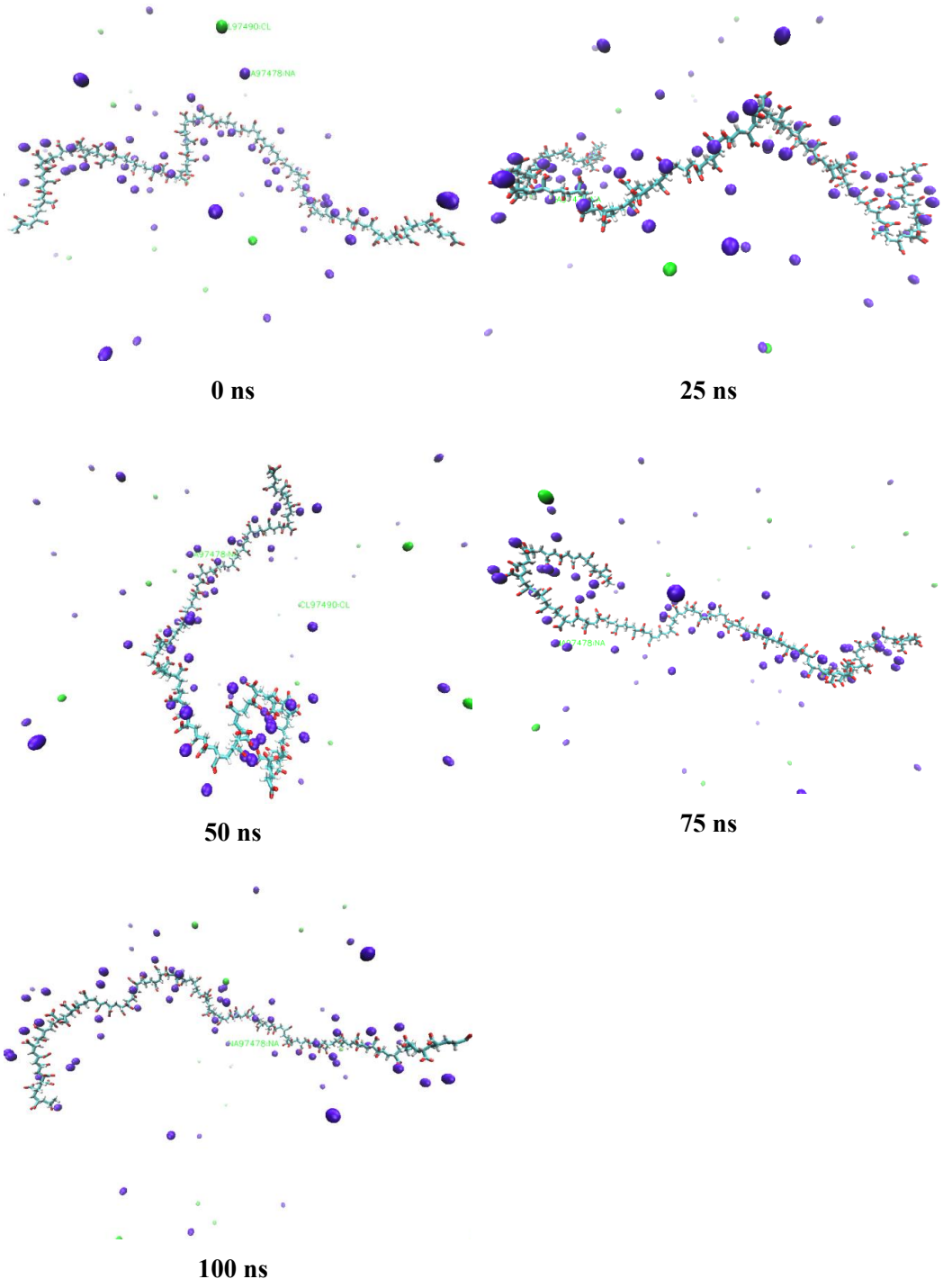


Figure 61 Snapshots taken from MD trajectory of 62-mer model of fully deprotonated PAA with NaCl ions and water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (100 ns). PAA is shown as sticks with carbon (cyan) and oxygen (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: sodium ions (violet) and chlorine ions (green).

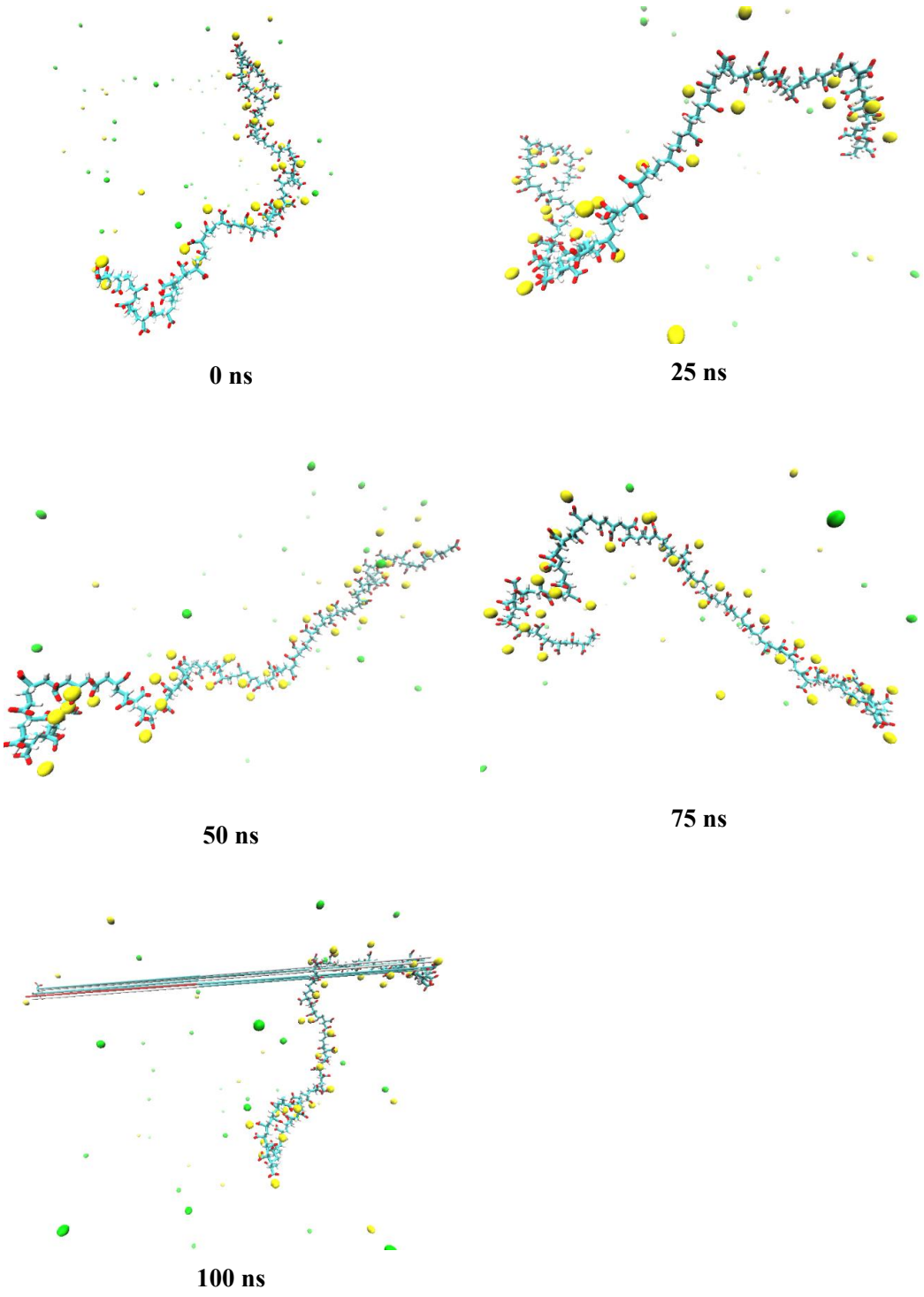


Figure 62 Snapshots taken from MD trajectory of 62-mer model of fully deprotonated PAA with NaCl ions and water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (100 ns). PAA is shown as sticks with carbon (cyan) and oxygen (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: calcium ions (orange) and chlorine ions (green).

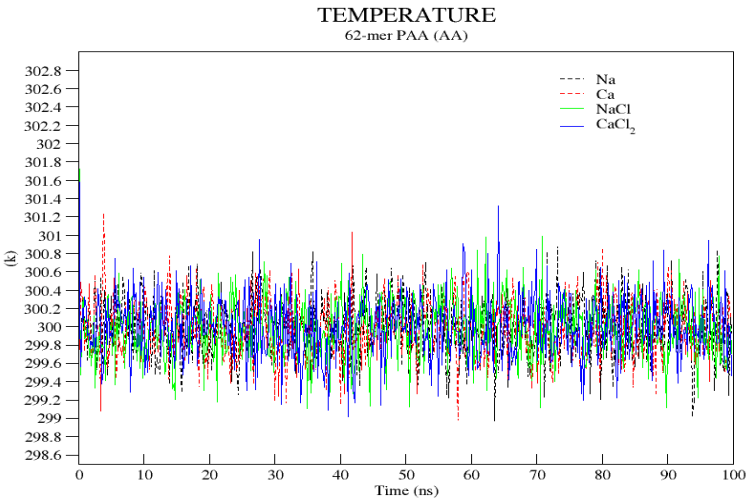
5.3.2 Quality assurance

First, we check again if the system is well equilibrated. As shown in Fig 63a, 63b, the average temperature and pressure, calculated as a time average from all 3 single independent runs, fluctuate steadily throughout the 100 ns simulation. The average temperature and pressure of all 3 independent production runs is presented in the Table 19. All the systems present an average temperature and pressure of 300.0 ±0.01 K and 1.71 ±1.32 bar respectively. The average energies (see Fig. 63c, 63d, 63e) of our system are constant, the kinetic energy is $7.30x10^5$ kJ/mol, the potential energy is $-4.62x10^6 \pm 0.7x10^{-5}$ kJ/mol except the system with calcium and chlorine ions showing a higher level of potential energy which is stable through the 100 ns simulation hence the system is considered stable and the total energy is $-1.259x10^6$. The system with $CaCl_2$ shows a higher value of total energy, $-3.89x10^6 \pm 0.7x10^{-5}$. The values are constant over the whole trajectories, this indicates a well equilibrated and stable system.

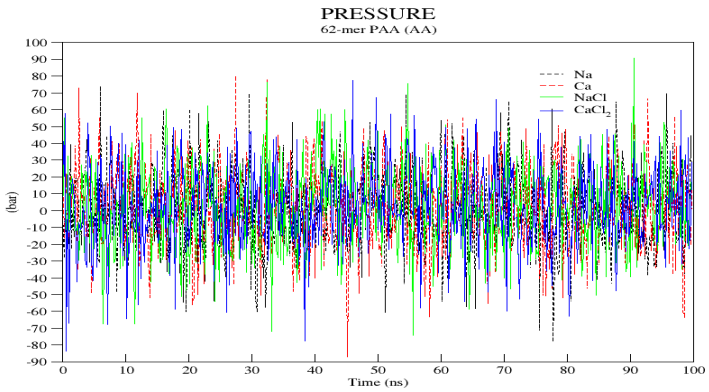
Table 19 System properties calculated as a time average from all 3 single independent runs

Property	Average value of over 3 independent production runs			
	Na	***Ca***	***NaCl***	***$CaCl_2$***
Temperature (K)	300.00	300.00	299.99	300.01
Pressure (bar)	0.71	1.35	2.58	2.20
Potential Energy (KJ/mol)	-4.60E+06	-4.62E+06	-4.61E+06	-4.66E+06
Kinetic Energy (KJ/mol)	7.30E+05	7.30E+05	7.30E+05	7.30E+05
Total Energy (KJ/mol)	-3.87E+06	-3.89E+06	-3.88E+06	-3.93E+06

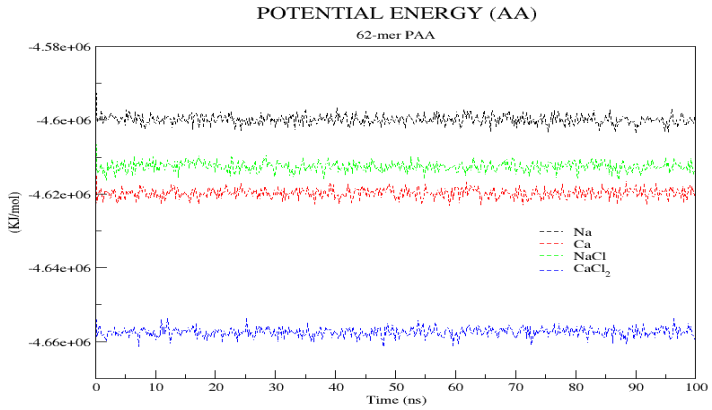
a)



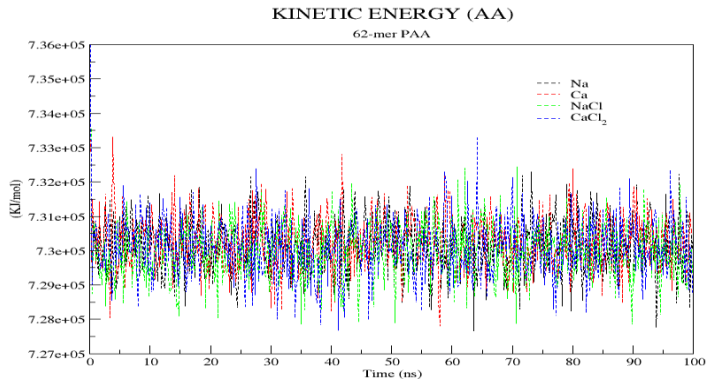
b)



c)



d)



e)

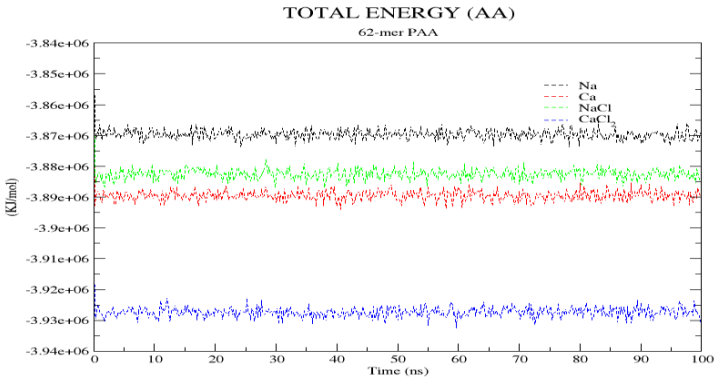


Figure 63 Average energy profiles of 3 independent AA production runs (100 ns) for 62-mer model of fully deprotonated PAA with Na, Ca, NaCl, and $CaCl_2$ ions and water

5.3.3 Radius of Gyration

Table 20 Average value of Radius of gyration calculated over 3 production runs

System	Average (nm)
Na	2.8044
Ca	2.1615
NaCl	2.8324
$CaCl_2$	2.4129

The time averaged radius of gyration Rg of the fully deprotonated polyacrylic acid is shown in Table 20. This is bigger than the Rg value of PAA 20-mer as the chain size of the long chain PAA is longer than the short chain PAA. The simulation results show us long chain PAA goes under the conformational change from stretched to bended for all the systems in first 30 ns. The radius of gyration for all systems consistently keeps decreasing for the rest of the simulation till the end of 100 ns simulation time.

When compared with system containing 20-mer PAA chain and sodium ions, the short chain does not show major conformational change and fluctuate around its average value whereas in the system containing long chain PAA shortens in the first 19 nanoseconds and fluctuate around the average value for the next 81 nanoseconds.

A similar behaviour is observed (see Fig 64) in the system with calcium ions where the radius of gyration drastically reduces in first 35 nanoseconds and fluctuate around average value of radius of gyration for the system till the end of the simulation.

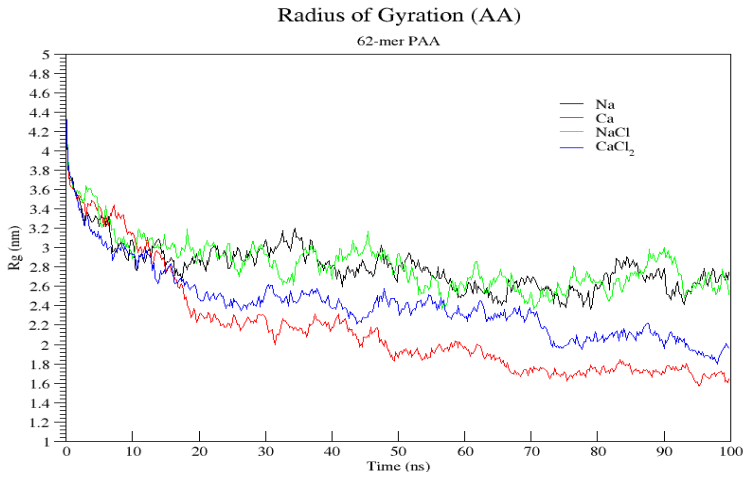


Figure 64 Average curve of the radius of gyration Rg over time of 100 ns for all 4 systems over 3 independent production runs of PAA (fully deprotonated) AA simulation.

The long polymer chain containing Ca2+ and $Cacl_2$ ions shows a similar behaviour compare two short chain PAA but the polymer chain shows sharp change in the conformation which can be seen through the reduced value of radius of gyration in first 16 nanoseconds. The long-chain polymer continues to decrease as the simulation goes ahead till it reaches 100

nanosecond a similar behaviour was seen in the short chain where the radius of gyration consistently decreased over the period of 100 ns.

The difference between the long and the short chain polymer can be observed by the change of conformation of the long-chain polymer in the presence of calcium and calcium chloride ions from Fig. 64. It can be observed the radius of gyration of the long-chain polymer containing calcium chloride ions stays bigger i.e. PAA tends to remain in stretched conformation than the long chain PAA simulated with calcium ions where as in the short polymer chain it was seen the radius of gyration for the short chain PAA system containing calcium chloride was far lower than the system containing calcium ions this this is due to excess chlorine ions bonding with the polymer chain leaving less residues for calcium and to bond with the polymer hence, less effect of divalent ions on the polymer chain length can be seen in the presence of chlorine atoms.

5.3.4 End–to–End distance

Table 21 Average end-to-end distance values calculated over 3 independent production runs

System	***Average (nm)***
Na	7.1021
Ca	4.6455
$NaCl$	7.1577
$CaCl_2$	5.0183

Due to the folding of PAA in the first 5 ns, the end-to-end distance enhance drastically from 5.5 to up to 9.5 ± 0.04 nm (see Fig 65), which is opposite to the short chain system where the value of R decreased in the first 5 ns. The value of R for systems containing Na and NaCl ions fluctuates throughout the 100 ns simulation but no major change in value is seen which is in agreement with the obtained value of R_g. The polymer system containing only calcium ions drop in the value of *R* after 2 ns with an initial increase of the value from 5.5 to 7.7 nm. For the system containing calcium chloride ions value of *R* increases to 8.9 nm at 10 ns and then goes on to decrease for the rest of the simulation till it reaches its final value of 2.68 nm. The system containing calcium chloride ions show increase in the value of *R* in the first 5 nanoseconds reaching the value of 8.7 nm and then goes on decreasing for the rest of the simulation till 75 nanoseconds where a sharp drop from 5.05 nm to 2.64 nm can be seen in Fig. 65. For the remaining 25 nanoseconds the value of R constantly fluctuate till it reaches its final value of 3.36 nm.

A similar trend was also seen in the short chain PAA, in the long-chain PAA it can be seen from the Fig. 65 the value of *R* are similar for the system containing calcium and calcium chloride ions for a 10 to 60 ns whereas in the short chain polymer it was seen the value of *R* was consistently different between the two systems and the value for all the system containing only calcium ions remained to be higher than the system containing calcium and chlorine ions thought the 100 ns simulation time. Again, the results are in complete agreement with the value of the radius of gyration of the system.

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

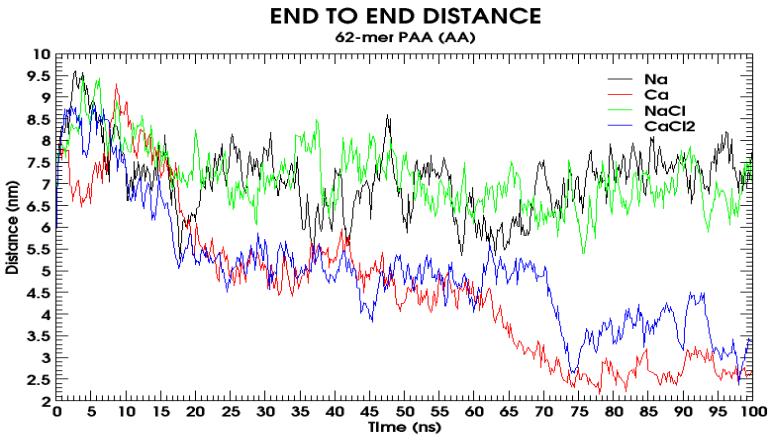


Figure 65 Average curve of the end-to-end distance R over time of 100 ns for all 4 systems over 3 independent production runs of PAA (fully deprotonated) AA simulation

5.3.5 Root-mean-square deviation

Table 22 Average RMSD values calculated over 3 independent production runs

System	Average (nm)
Na	2.1231
Ca	2.7144
NaCl	2.0927
$CaCl_2$	2.6467

The backbone RMSD of long chain PAA in presence of monovalent/divalent ions and water fluctuates around an average value presented in Table 21. From Fig. 66 it can be seen the RMSD value is on a constant rise since the beginning of the simulation for all the 4 systems. All 4 system show a similar growth in the RMSD value till 17 ns from where the system containing only calcium and calcium chloride ions have significantly higher growth in the RMSD value till the end of 100 ns simulation time. This is due to the higher affinity of PAA towards the divalent counterions resulting in major conformational changes and structure of PAA. The final values at the end of simulation for 100 ns for system containing Na, Ca2+, NaCl and $CaCl_2$ ions are 2.12, 3.29, 2.27 and 3.32 nm respectively which are far higher than the initial value of 0.49×10^{-3} nm. These values are in agreement with the values of R and R_g.

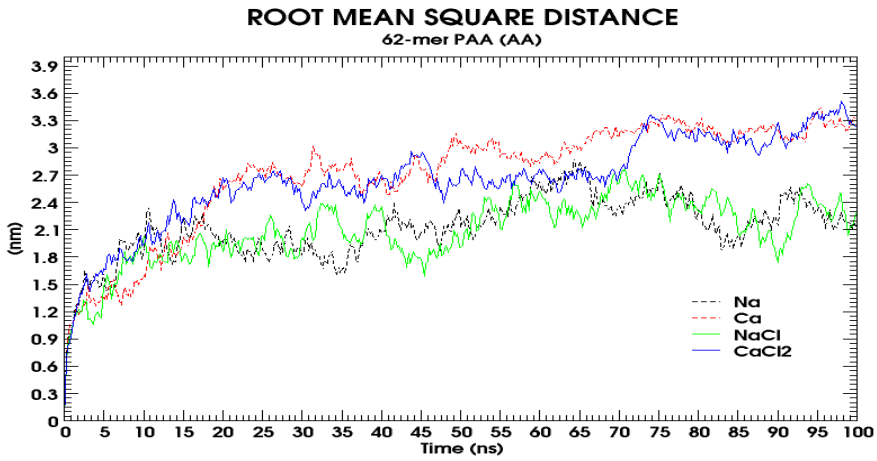


Figure 66 Average curve of the backbone RMSD over time of 100 ns for all 4 systems over 3 independent production runs of PAA (fully deprotonated) AA simulation

5.3.6 Solvent accessible surface area

Table 23 Average SASA values calculated over 3 independent production runs

System	Average (nm^2)
Na	53.8838
Ca	50.5183
NaCl	53.9347
$CaCl_2$	50.4003

The 62-mer AA model of PAA has a time average SASA for all four systems presented in the Table 23. This is almost thrice the value of SASA of the 20-mer model of polyacrylic acid (see Table 10). Due to high degree of folding of PAA in the presence of divalent ions as observed from the results of R and R_g the value of SASA are well explainable and in agreement with our previous results.

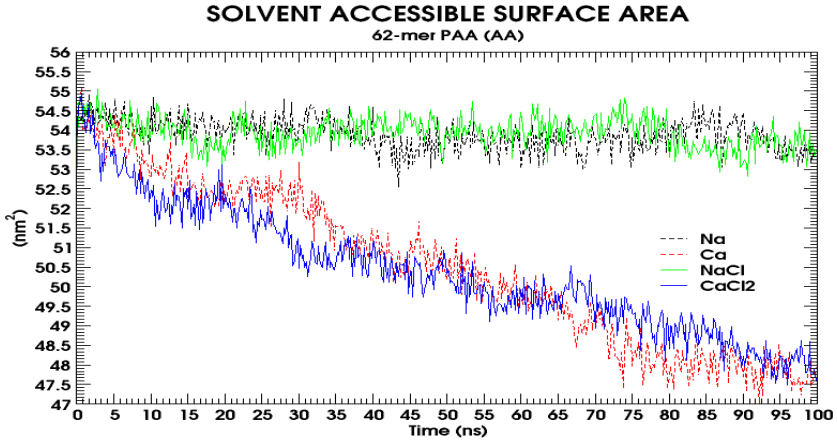


Figure 67 Average SASA values calculated over 3 independent production runs

5.3.7 Hydrogen bonds

Again, the conformational change happening during the early stages of the simulation in the system containing calcium and calcium chloride ions in are clearly visible in the Fig. 68 as a decrease in the value of hydrogen bonds after 1 ns simulation time. As the ends of the chain tend to fold in presence of calcium ions, the number of free carboxylate ions decreases and thus the number of hydrogen bonds. The number of hydrogen bonds (N_{hb}) for each system are presented in Table 24, The value of N_{hb} reaches a local minimum of around 256 in calcium chloride excess salt simulation which represents a highly folded structure of PAA with calcium ions trapped in it.

As we can see from the RDF between carboxylate oxygens and water hydrogen atoms in system containing sodium chloride ions (see Fig. 70) every oxygen atom is surrounded by nearly 3 hydrogen atoms, so there should be around 372 hydrogen bonds.

Table 24 Average hydrogen bonds calculated over 3 independent production runs

System	Average
Na	354.0107
Ca	281.0820
NaCl	353.2300
$CaCl_2$	266.1547

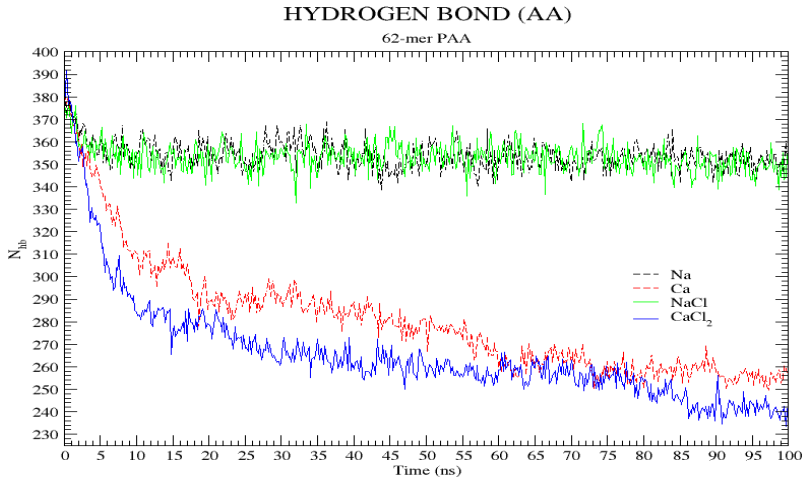


Figure 68 Average curve of the change of hydrogen bonds over time of 100 ns for all 4 systems over 3 independent production runs of PAA (fully deprotonated) AA simulation

5.3.8 Radial Distribution Functions

5.3.8.1 RDF between the centre of mass of the Polyacrylic acid residues and the oxygen atoms of water molecules

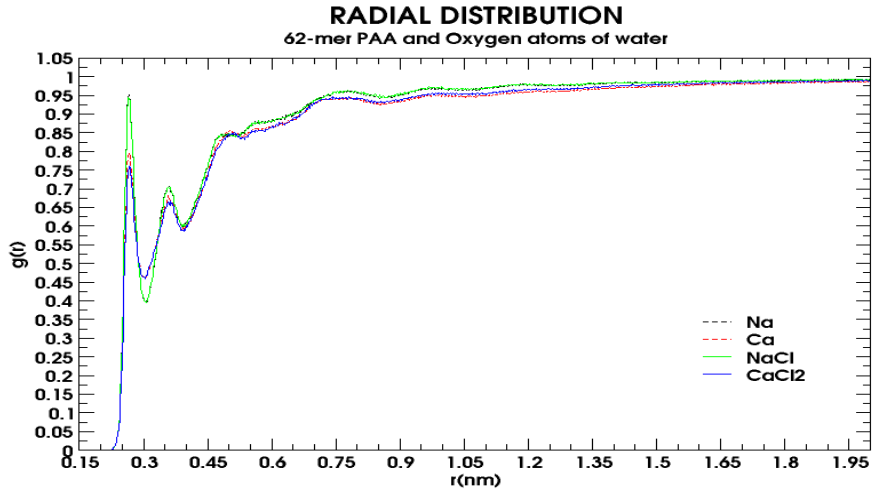


Figure 69 RDF between PAA (center of mass) and water oxygens (OW) for all for all 4 systems over 3 independent production runs of PAA (fully deprotonated) AA simulation

The sharp peak at 0.26 ± 0.002 nm in indicates a well-structured solvation shell with ordered distribution of water molecules around the polymer chain. This is similar to the results observed for short chain PAA.

5.3.8.2 RDF between oxygen atoms of carboxylate group of PAA and the hydrogen atoms of water molecules

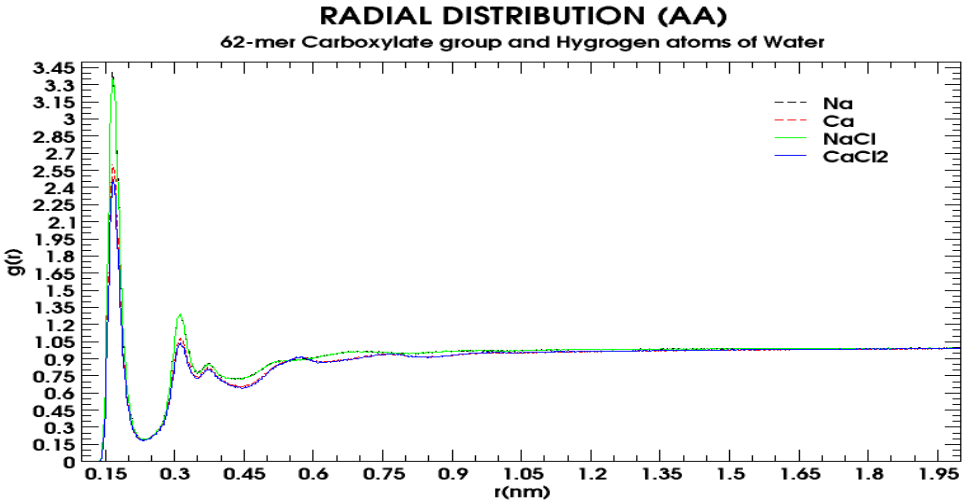


Figure 70 RDF between carboxylate oxygens of PAA and hydrogen atoms of water (HW) for all for all 4 systems over 3 independent production runs of PAA (fully deprotonated) AA simulation

The RDF between the carboxylate oxygen atoms and water oxygen atoms in show two peaks at 0.166 ± 0.002 nm and 0.312 ± 0.002 nm, which is much shorter than the observed distance in bulk water for short chain PAA.

5.3.8.3 RDF between oxygen atoms of carboxylate group of PAA and the oxygen atoms of water molecules

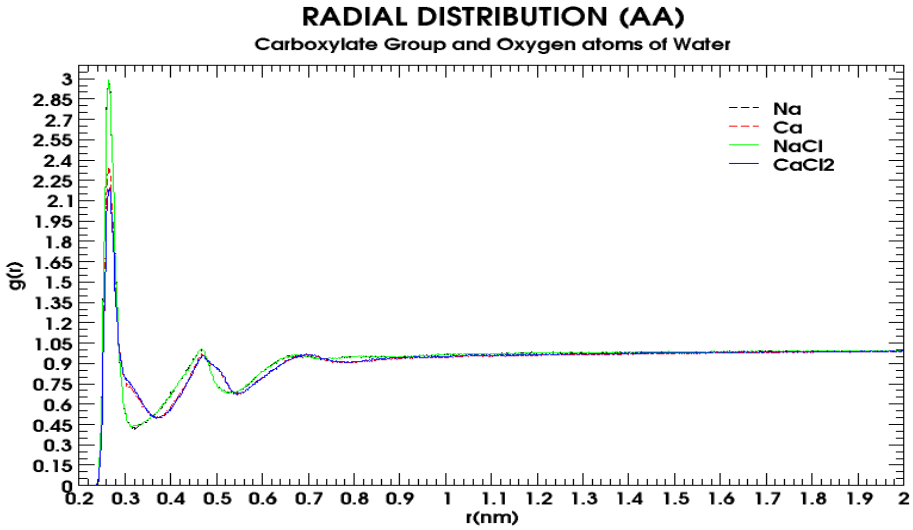


Figure 71 RDF between carboxylate oxygens of PAA and water oxygens (OW) for all for all 4 systems over 3 independent production runs of PAA (fully deprotonated) AA simulation

The RDF between the carboxylate oxygen atoms and water oxygen atoms in show two peaks at 0.266 ± 0.002 nm and 0.466 ± 0.002 nm. Again, the results of short chain PAA matches the results for long chain PAA. For short chain, RDFs show two 2 peaks at 0.28 and 0.48 nm, respectively and 0.25 and 0.45 nm based on the simulation with FF-1 and FF-2.

5.3.8.4 RDF between oxygen atoms of carboxylate group of PAA and the ions

As shown by Fig.72, the RDF between the carboxylate oxygens and monovalent and divalent counterions exhibits a strong first peak at 0.234 ± 0.002 nm for sodium and sodium chloride and at 0.260 ± 0.002 nm for calcium and calcium chloride ions, suggesting a

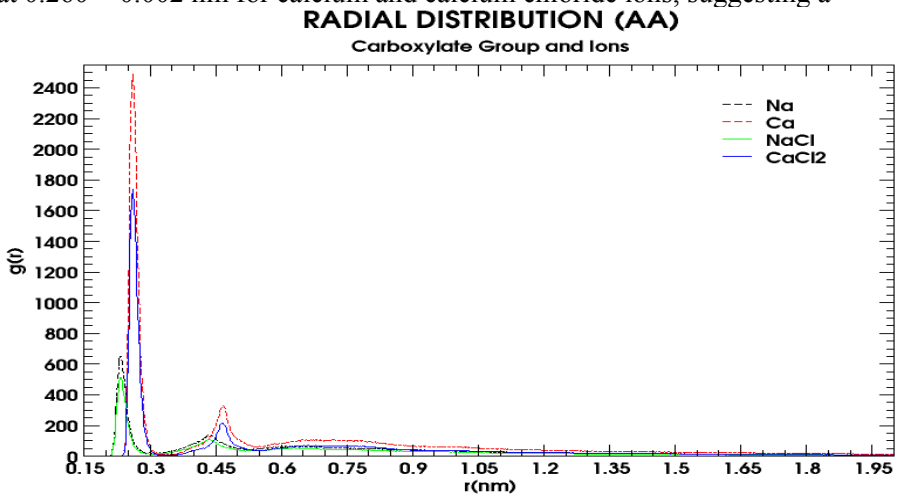


Figure 72 RDF between carboxylate oxygens of PAA and ions for all for all 4 systems over 3 independent production runs of PAA (fully deprotonated) AA simulation

preferred condensation of divalent ions on to the chain backbone of PAA and a weaker shoulder peak at 0.434 ± 0.002 nm for sodium and sodium chloride ions and at

0.464 ± 0.002 nm for calcium and calcium chloride ions. Sulatha *et al.*[30] discovered a similar behaviour for simulations with FF-2 parameter set. Their RDFs show a first peak at ~0.21 nm

5.4 Analysis of trajectories from Coarse-grained MD simulations of fully deprotonated 62-mer model of Polyacrylic acid (PAA)

PAA (fully deprotonated) MARTINI model is simulated in a system containing polarizable water model from MARTINI and 4 different ions individually (see Table 18). Polymer model was generated using the mapping scheme described in section 3.2.1 using MARTINI scheme for modelling of polymer. The systems were equilibrated at 300 K and 1 bar pressure for 50 ns. The simulation was carried out for 500 ns with a time step of 20 fs. Detailed system configurations are provided in appendix 9.15 – 9.18.

The four generated boxes each include one PAA MARTINI long-chain model with periodic boundary conditions applied in all directions. For PAA simulations, cubic boxes with 14.4 nm edge length were used for the long-chain models. The new polarized water model model was used to solvate the system. In order to neutralize the system charge of the polymer, water molecules were replaced by ions according to the number of monomers in the polymer chain (see Table 18). To account for the ionic strength of the synthetic tap water (15°dH) used in experiments additional water molecules were exchanged by sodium, calcium and chloride ions depending upon the salt used in the simulation setup. The setup for each simulated system is summarized in Table 5.

All simulations were performed on the RWTH Compute Cluster using GROMACS 2019 simulation package[116]. All systems were first energy minimized using steepest descent algorithm with 500,000 integration steps or until the maximum force on any atom in the system did not exceed a value of 10.0 kJ/mol/nm. For neighbor searching, Verlet cutoff-scheme was used having short range Van der Waals cut-off of 1.0 with periodic boundary conditions. For electrostatic forces, Particle-mesh Ewald method of order 8 was used having short range cut-off of 1.2 keeping relative dielectric constant and relative dielectric constant of the reaction field equal to 2.5 and 1 respectively. For Van der waals forces, twin range cut-offs with neighbor list cut-off and VdW cut-off are used. For the treatment of long-range electrostatic interaction and long-range dispersion, corrections for energy and pressure was applied. Grid dimensions are controlled with Fourier spacing of 0.15. The relative strength of the Ewald-shifted direct potential at rcoulomb is given by 'ewald-rtol'. For doing PME for VdW-interactions, ewald-rtol-lj is used to control the relative strength of the dispersion potential at rvdw. Both the values are kept default 1e-5 and 1e-3 respectively.

After heating, the systems were equilibrated for 50 ns each at 300 K using the velocity rescale thermostat (coupling time 1 ps) according to Bussi *et al.*[117] in NVT ensemble and at 1 bar using Berendsen barostat[118, 119] (coupling time 12 ps and compressibility $3 \cdot 10^{-4}$ bar^{-1}) in NPT ensemble using verlet velocity integrator. During the equilibration, position restraints on every backbone bead of the PAA model and on every polymer bead were activated. Three independent production runs, 500 ns each, were performed for every system in NPT ensemble using a Verlet cutoff scheme with a cutoff distance of $r_{cut} = 1.2$ nm for the Lennard-Jones interactions according to a recent publication[120]. Consistently, a time step of 2 fs was used. The neighbor list was updated every 10 fs using the Verlet neighbor search (VNS) algorithm[120]. Bond lengths were held constant by the LINCS algorithm[113, 121] and Parrinello-Rahman extended-ensemble pressure coupling was used. The background permittivity was set to $\varepsilon_{bg} = 2.5$ as recommended when using the polarizable water model[106, 120]The final box sizes are shown in Table.4.

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

5.4.1 Visualization of PAA interaction with Na, Ca, NaCl & $CaCl_2$ using VMD

Snapshots of the simulated system consisting of PAA MARTINI 62-mer model of fully deprotonated PAA consisting of Na, Ca, NaCl and $CaCl_2$ ions and polarized water are shown in Fig. 73, 74, 75 & 76, respectively. The PAA chain is shown as stick model. The PAA chain is stretched conformation in the first snapshot.

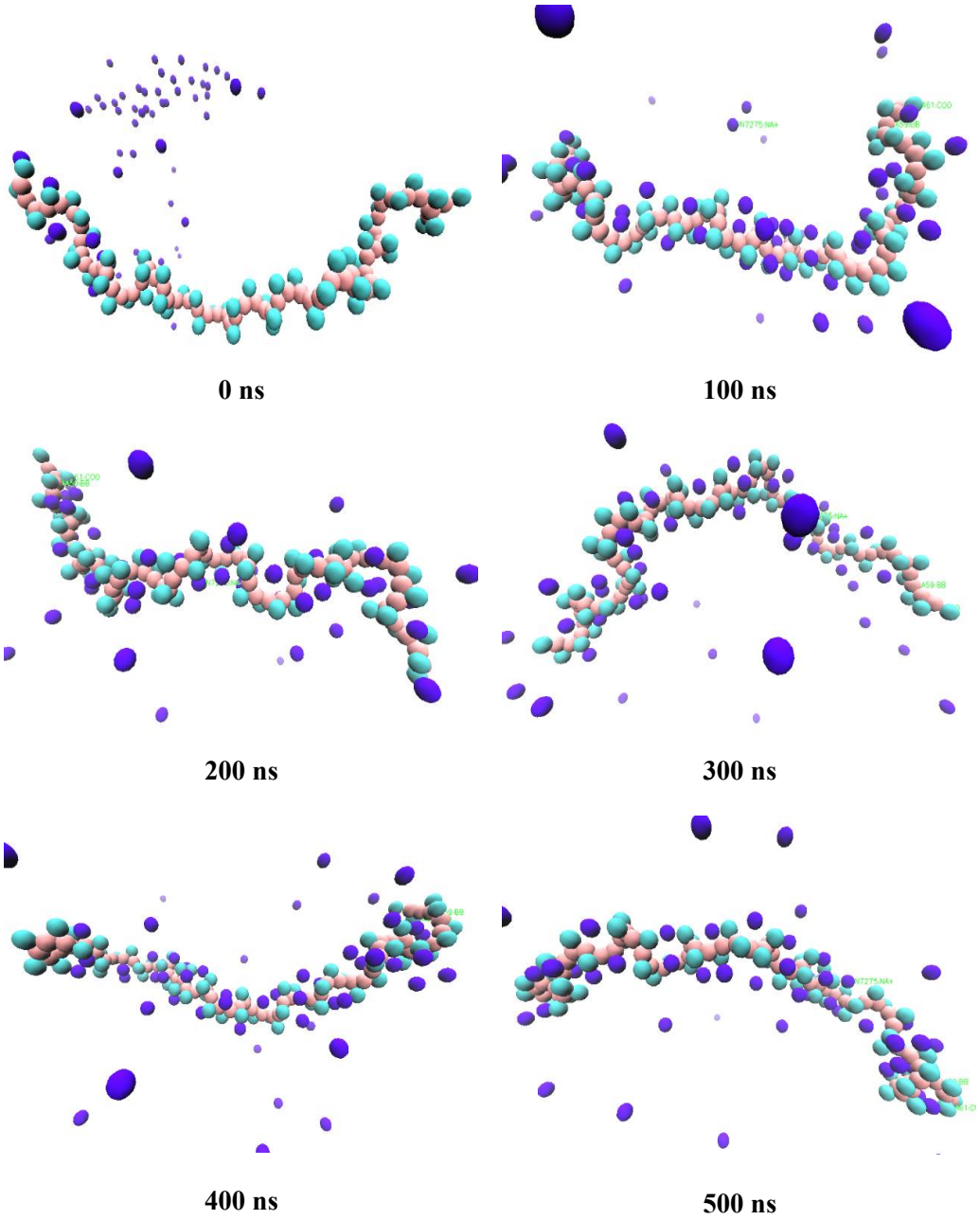


Figure 73 Snapshots taken from CG MD trajectory of 62-mer model of fully deprotonated PAA with Na ions and polarized water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (500 ns). PAA is shown as beads with SC1 (pink) and SQa (cyan). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: sodium ions (violet).

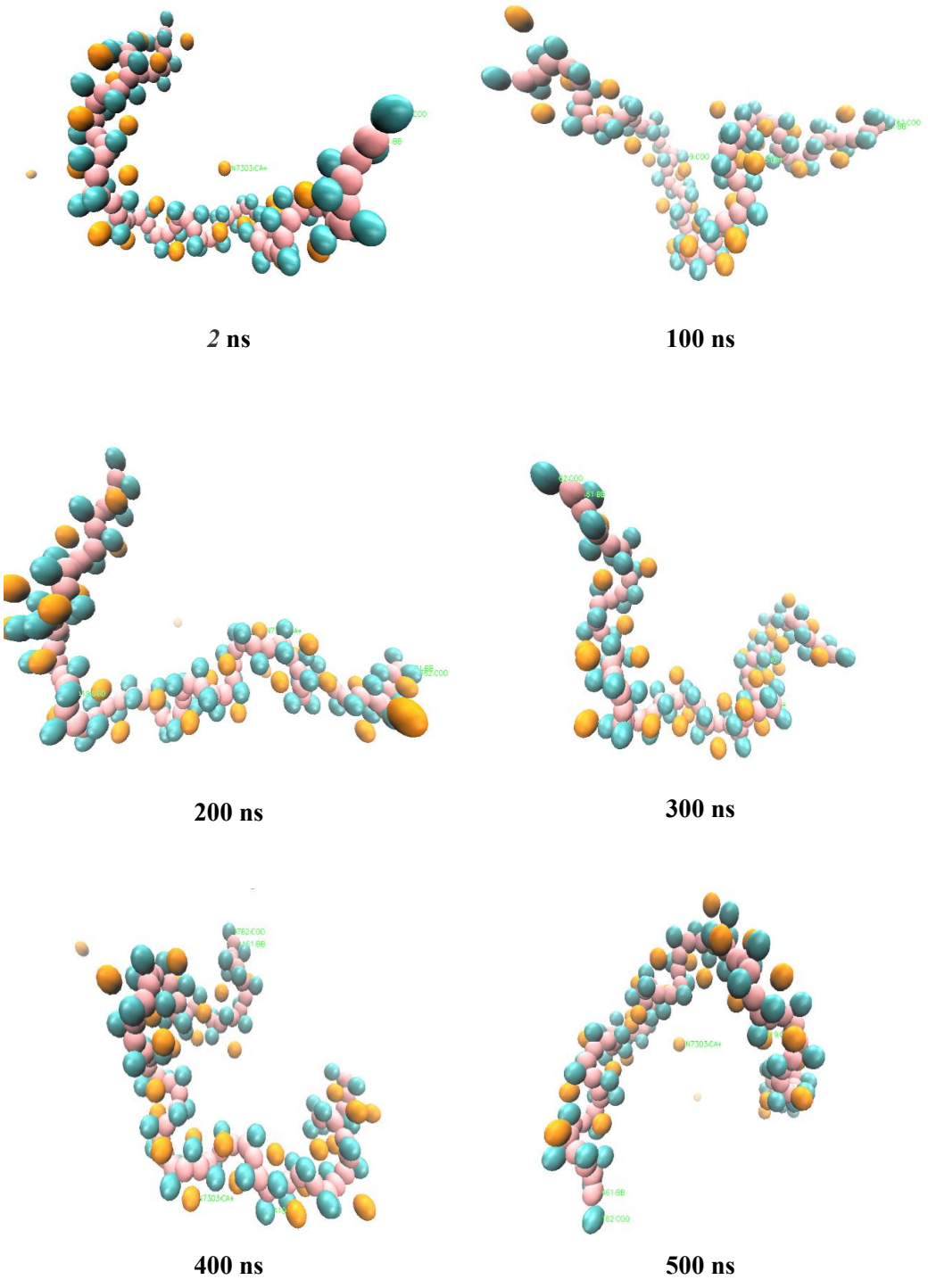


Figure 74 Snapshots taken from MD trajectory of 62-mer MARTINI model of fully deprotonated PAA with Ca2+ ions and polarizable water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (500 ns). PAA is shown as beads with SC1 (pink) and SQa (cyan). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: Calcium ions (orange).

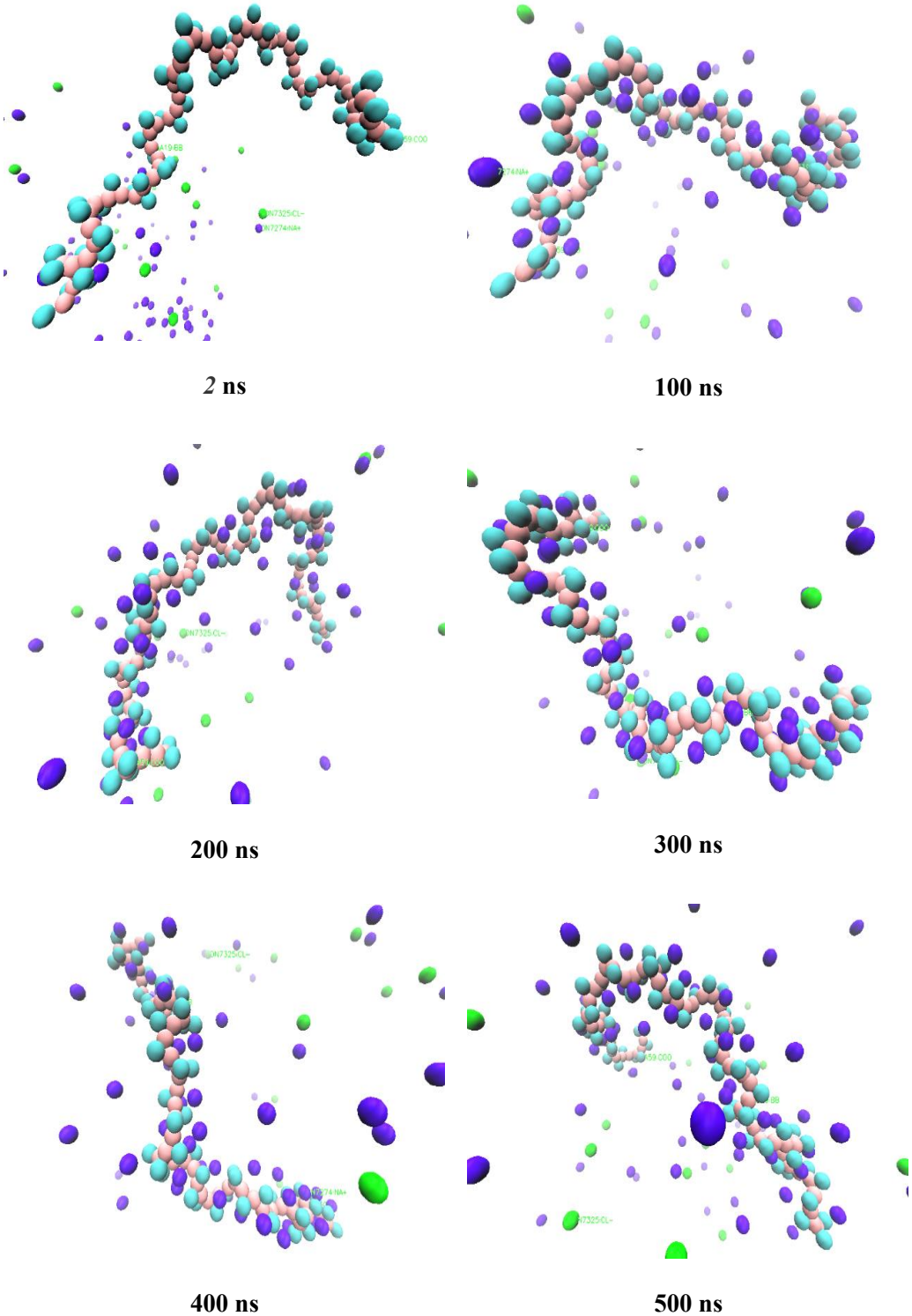


Figure 75 Snapshots taken from MD trajectory of 62-mer MARTINI model of fully deprotonated PAA with NaCl ions and polarizable water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (500 ns). PAA is shown as beads with SC1 (pink) and SQa (cyan). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: sodium ions (violet) and chlorine (green)

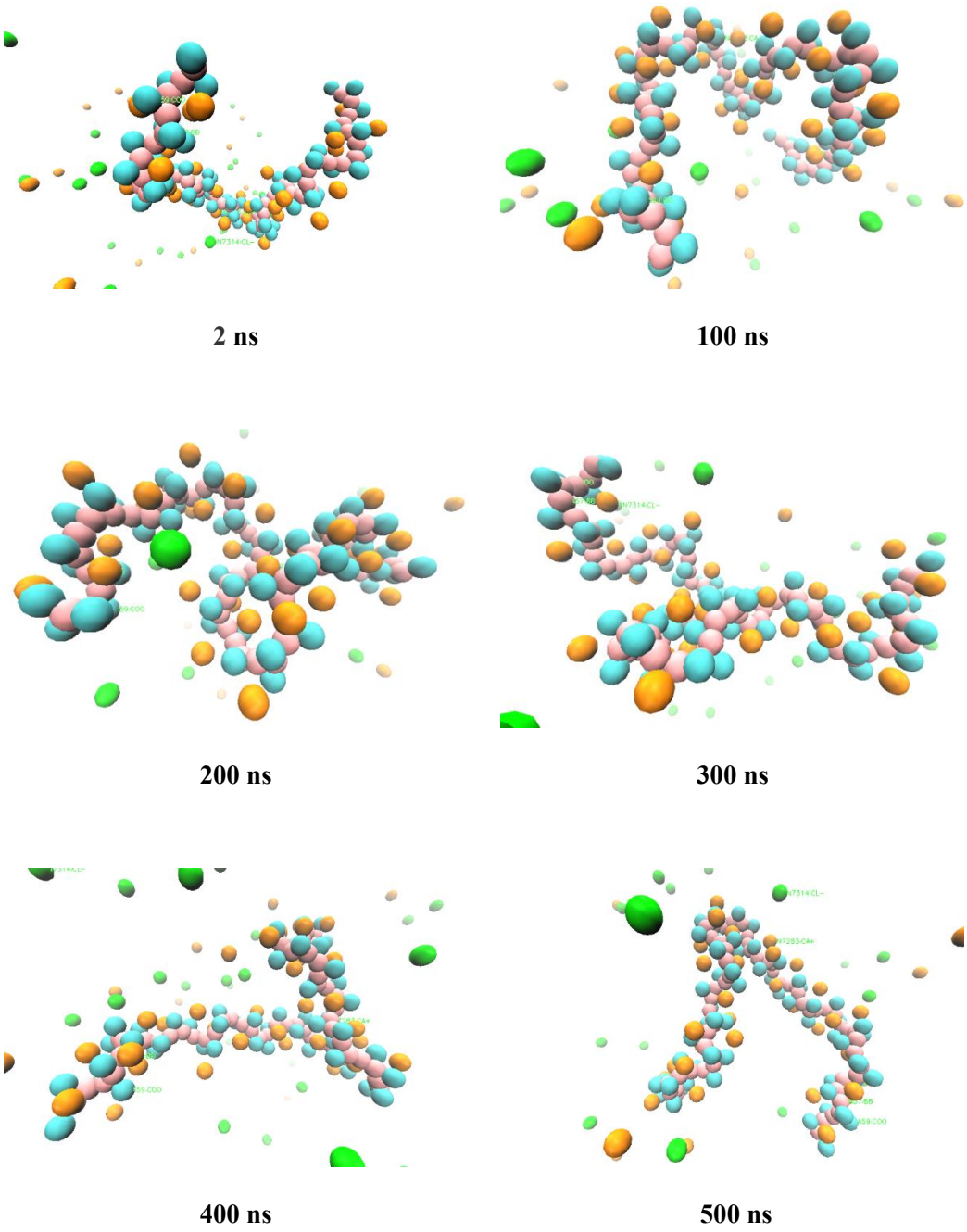


Figure 76 Snapshots taken from MD trajectory of 62-mer MARTINI model of fully deprotonated PAA with $CaCl_2$ ions and polarized water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (500 ns). PAA is shown as beads with SQa (cyan) and SC1 (pink). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: calcium ions (orange) and chlorine ions (green).

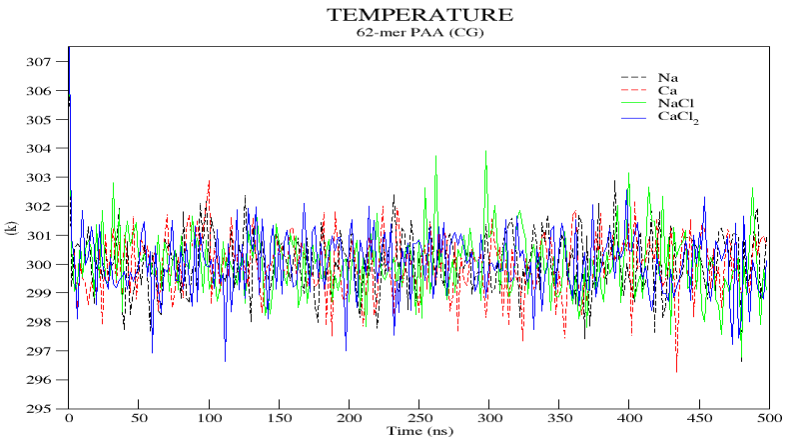
5.4.2 Quality assurance

Fist we check again if the system is well equilibrated. As shown in Fig. 77a, 77b the average temperature and pressure, calculated as a time average from all 3 single independent runs, fluctuate steadily. The average temperature and pressure of all four systems calculated over 3 independent production runs is presented in Table 24. The average energies (see Fig. 77c, 77d, 77e) of our system are constant. The values are constant over the whole trajectories, this indicates a well equilibrated and stable system.

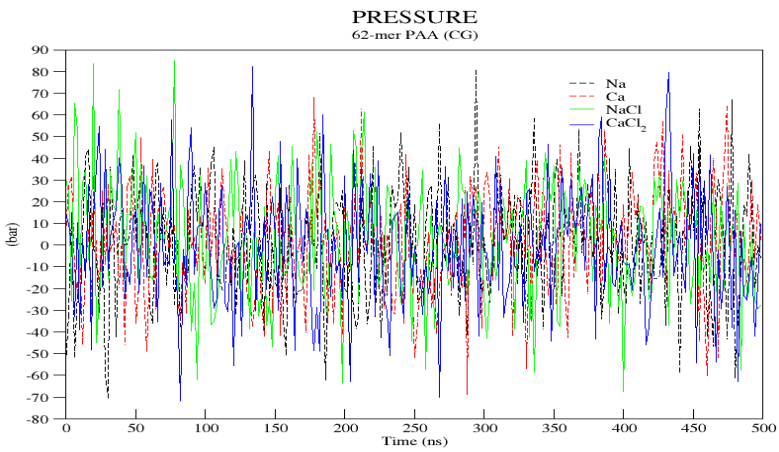
Table 25 System properties calculated as a time average from all 3 single independent runs

Property	**Average value of over 3 independent production runs**			
	Na	***Ca***	***NaCl***	***$CaCl_2$***
Temperature (K)	300.1053	300.0196	300.0535	300.0339
Pressure (bar)	0.2250	1.6807	2.8474	-0.5022
Potential Energy (KJ/mol)	-2.36E+05	-2.42E+05	-2.40E+05	-2.54E+05
Kinetic Energy (KJ/mol)	6.36E+04	6.38E+04	6.34E+04	6.35E+04
Total Energy (KJ/mol)	-1.72E+05	-1.78E+05	-1.77E+05	-1.91E+05

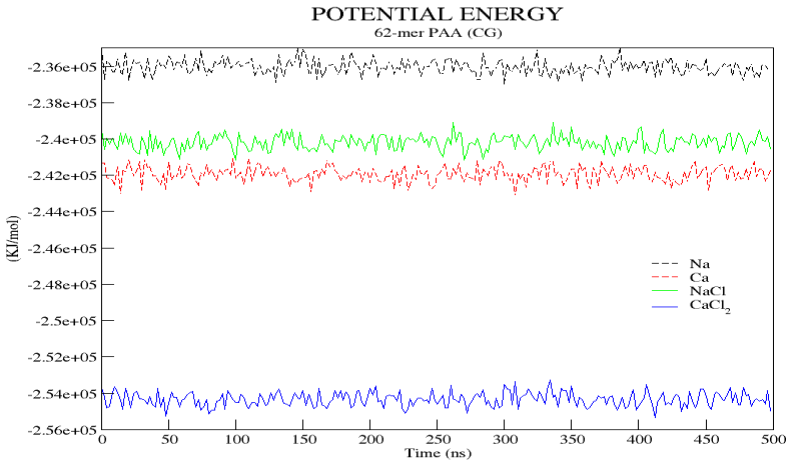
a)



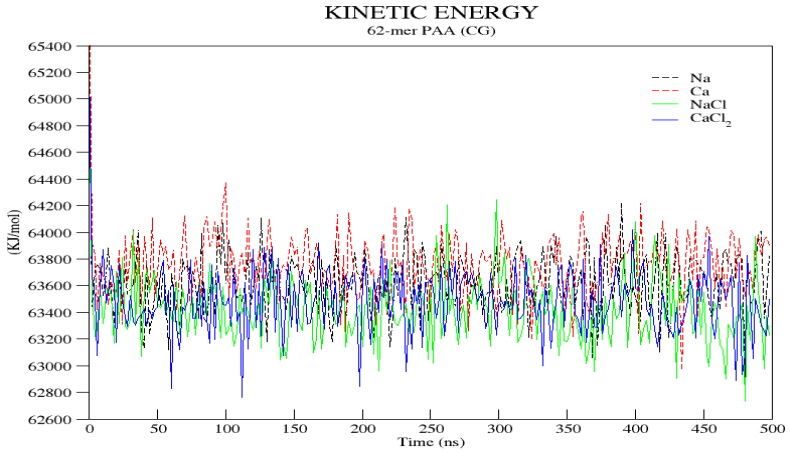
b)



c)



d)



e)

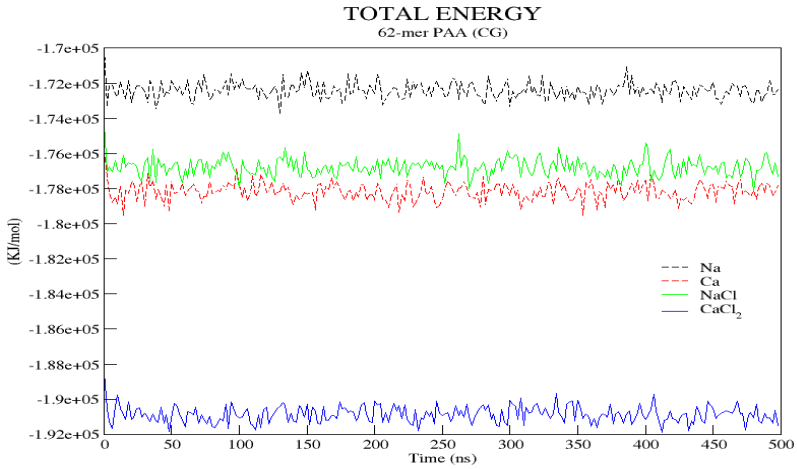


Figure 77 Average energy profiles of 3 independent CG production runs (500 ns) for 62-mer model of fully deprotonated PAA MARTINI with Na, Ca, NaCl, and $CaCl_2$ ions and water

5.4.3 Radius of Gyration

Table 26 Average Radius of gyration values calculated over 3 production runs

System	Average (nm)
Na	2.5814
Ca	2.0595
NaCl	2.4916
$CaCl_2$	1.9850

The radius of gyration R_g of the fully deprotonated coarse-grained polyacrylic acid is shown in Fig. 78. The time averaged R_g are shown in Table 26. This is slightly smaller than the R_g value obtained from the AA simulations because the chain backbone is smaller (2 atoms per monomer instead of 3 atoms). Compared to the average curve of long chain PAA obtained from AA simulation for 100 ns shown in Fig 64 the average value of R_g obtained from our CG models for the systems containing Na, Ca, NaCl, and $CaCl_2$ for first 100 ns are 2.59, 2.07, 2.45, 2.19 nm which are in good proximity with results obtained from AA. Although the trajectories generated from our CG model does not correspond to the trajectories obtained from AA model for 100 ns simulations. For the 62-mer MARTINI model of PAA the deviation from the results published by Reith *et al.*[28] increases.

For the systems with sodium and sodium chloride ions, R_g does not show any major deviation from the average value of R_g for the corresponding systems through 500 ns of simulation and fluctuates around its average value of 2.5 and 2.49 nm. A major drop in the R_g value can be observed in the first 28 nanosecond from 2.77 nm to 1.71 nm in the presence of only calcium ions after which the system tends to fluctuate around the average R_g of the system with maxima 2.42 nm at 350 ns and minima of 1.61 at 456 nm till the end of the simulation. After a short increase, the R_g value decreases to its final value after 20 ns of MD simulation for calcium ion system. Similarly, R_g for the system containing Calcium chloride ions show a sharp increase at 6ns reaching a maximum value of 2.5 nm. PAA further shows a sharp drop in the value of R_g throughout the system till the end of the simulation which can be seen in between 196 ns (1.62 nm) and at the minimum value of 1.55 nm at 462 ns till it reaches its final value of 1.76 nm.

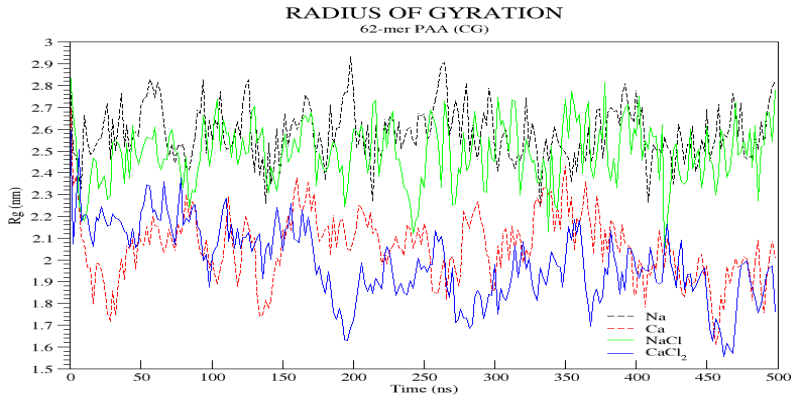


Figure 78 Average change of the radius of gyration over time for all 3 independent production runs of PAA (fully deprotonated) for 4 different systems of PAA (fully deprotonated) CG simulation

5.4.4 End–to–End distance

Table 27 Average end-to-end distance values calculated over 3 independent production runs

System	Average (nm)
Na	5.3433
Ca	4.5886
NaCl	5.0538
$CaCl_2$	4.4732

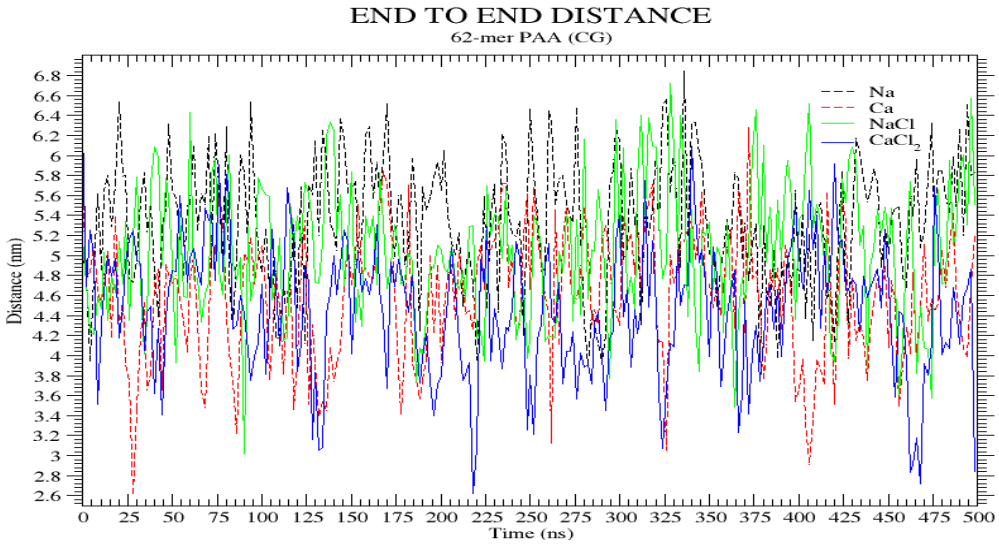


Figure 79 Average curve of the end-to-end distance R over time of 500 ns for all 4 systems over 3 independent production runs of PAA (fully deprotonated) CG simulation

The change of end-to-end distance R for long chain PAA during the simulation is similar to the progress of the radius of gyration in Fig. 78. Whenever PAA interacts strongly with the sodium ions and starts to fold, the end-to-end distance shortens. As the interaction between PAA and calcium ions exceed the electrostatic repulsion forces between the residues of PAA chain, PAA folds and the R value decreases (see Fig. 79). The average value of R for all the systems are shown in Table 27.

The average value of R for the first 100 ns of the simulation in the presence of Na, Ca2+, NaCl and $CaCl_2$ are 5.38, 4.50, 4.97, 4.74 nm respectively which are far smaller for the sodium and sodium chloride systems. This implies a slightly more stretched conformation of PAA in the presence of Na and NaCl ions, as the R value for pure PAA is only 3.12 ± 0.52 nm. This confirms the weak interactions between PAA beads and sodium ions compared to calcium and calcium chloride ions.

5.4.5 Root-mean-square deviation

Table 28 Average end-to-end distance values calculated over 3 independent production runs

System	Average (nm)
Na	1.3336
Ca	1.3791
NaCl	1.2978
$CaCl_2$	1.3626

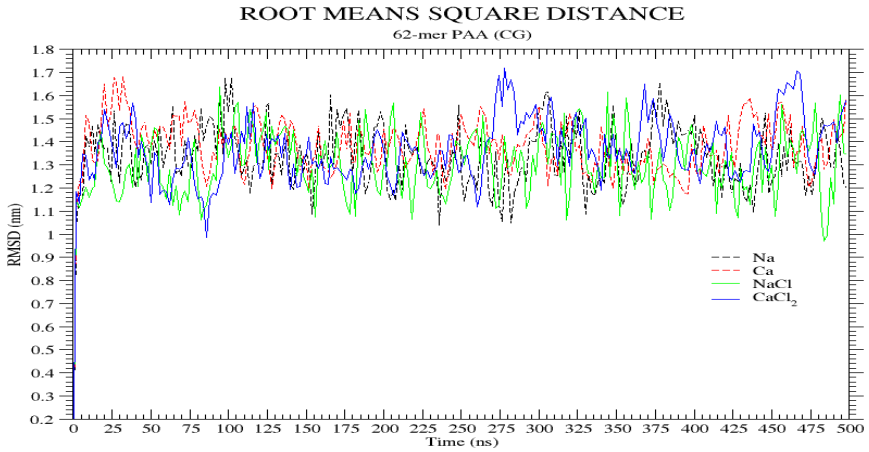


Figure 80 Average RMSD values calculated over 3 independent production runs

The backbone RMSD of long chain PAA MARTINI model within all the systems along with polarized water fluctuates around an average value presented in Table 28 till the end of the simulation. No sharp changes on the value of RMSD is seen during 500 ns of simulation. The values for first 100 ns for system containing Na, Ca2+, NaCl and $CaCl_2$ ions are 1.35, 1.41, 1.23 and 1.27 nm respectively which are far lower than the values obtained from all-atomistic simulations.

5.4.6 Solvent accessible surface area

Table 29 Average SASA values calculated over 3 independent production runs

System	Average (nm^2)
Na	52.2494
Ca	50.6153
NaCl	52.2743
$CaCl_2$	50.4675

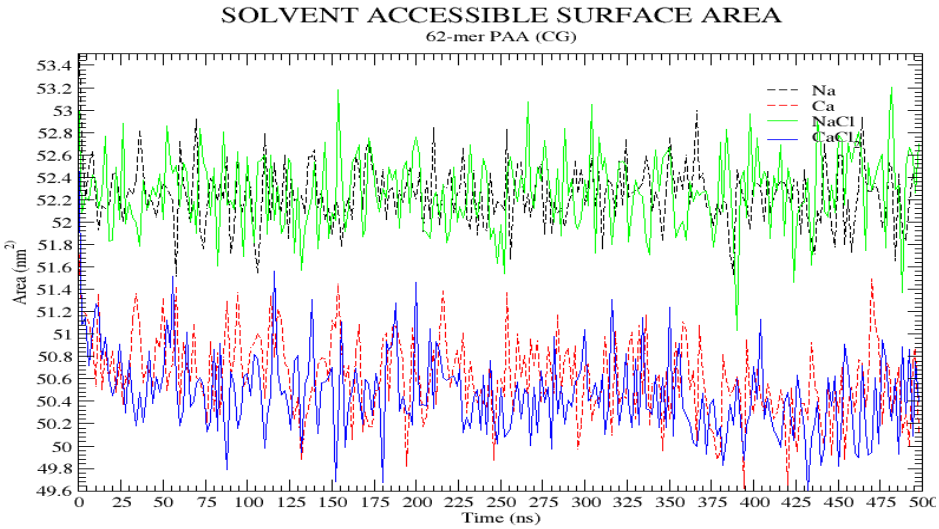


Figure 81 Average SASA values calculated over 3 independent production runs

The 62-mer MARTINI model of PAA has a time average SASA for all four systems presented in the Table 29. This is almost same as the value of SASA of the 62-mer AA model of polyacrylic acid (see Table 23).

5.4.7 Radial Distribution Functions

5.4.7.1 RDF between the centre of mass of PAA beads and polarized water beads

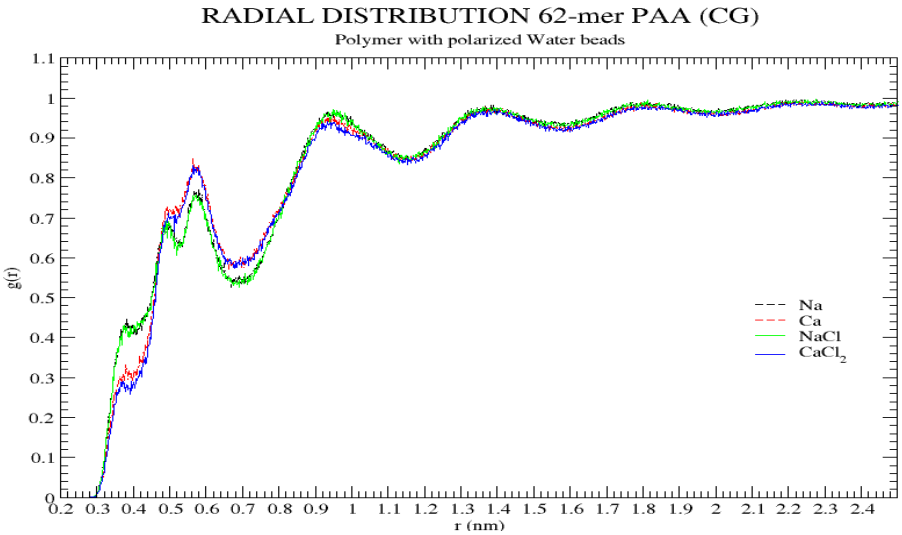


Figure 82 RDF between PAA (centre of mass) and polarized water beads (PW) for all for all 4 systems averaged over 3 independent production runs of PAA (fully deprotonated) CG simulation

The sharp peak at 0.57 ±0.002 nm in indicates a well-structured solvation shell with ordered distribution of polarized water beads around the polymer chain. This value is bigger than the value obtained in AA simulation. This result can be corrected by adjusting the dielectric constant for each individual simulation.

5.4.7.2 RDF between SQa beads and 'W' of polarized water bead

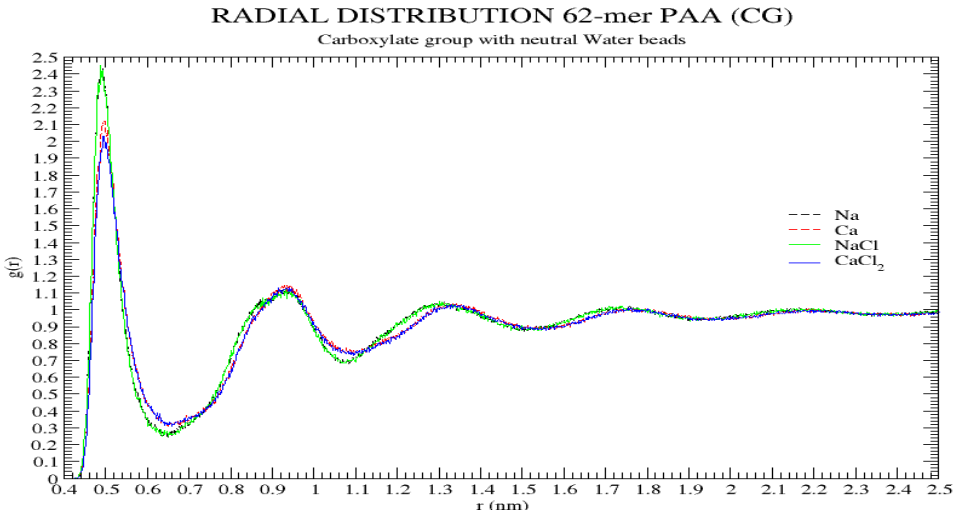


Figure 83 RDF between SQa beads and LJ interaction site 'W' for all for all 4 systems averaged over 3 independent production runs of PAA (fully deprotonated) CG simulation

The RDF between the SQa beads (carboxylate group) with respect to neutral water beads shows a strong peak at 0.48 ± 0.002 nm (see Fig. 83). A shoulder peak at 0.93 ± 0.002 nm follows.

5.4.7.3 RDF between SQa beads and the Polarized water beads

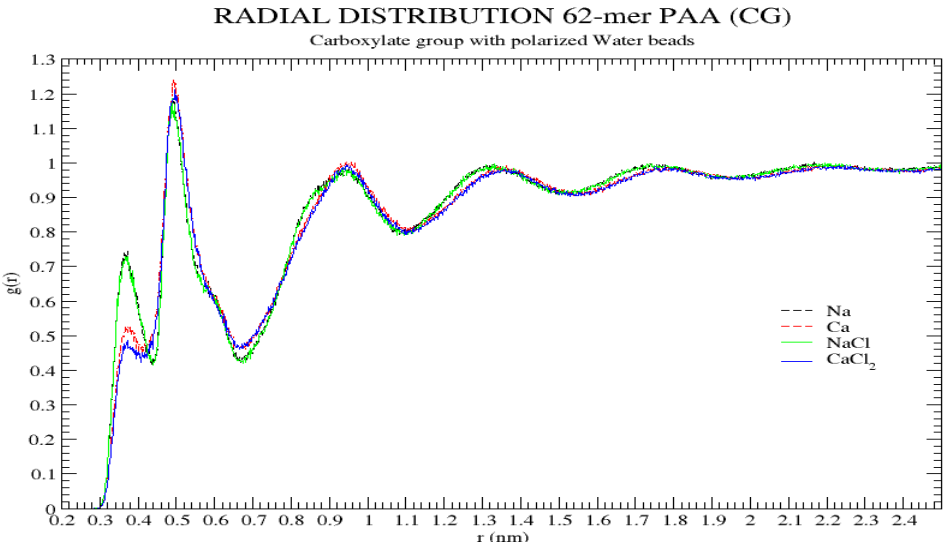


Figure 84 RDF between SQa beads and PW for all for all 4 systems averaged over 3 independent production runs of PAA (fully deprotonated) CG simulation

The RDF between the SQa (carboxylate group) beads and polarized water beads shown in in show a small first peaks at 0.36 ± 0.002 nm and an major peak at 0.48 ± 0.002 nm for sodium and sodium chloride systems. Similarly, at 0.37 ± 0.002 and 0.496 ± 0.002 for systems containing calcium and calcium chloride ions.

5.4.7.4 RDF between SC1 beads and the Polarized water beads

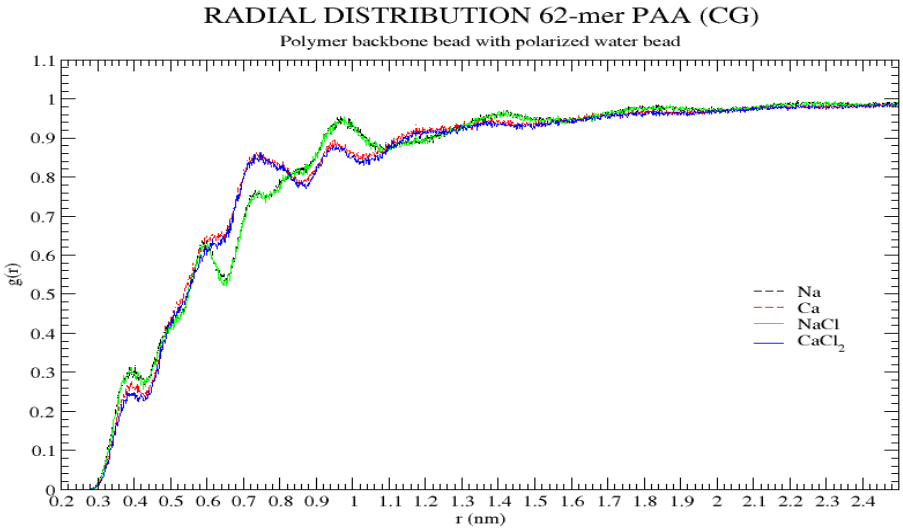


Figure 85 RDF between SC1 beads and PW for all for all 4 systems averaged over 3 independent production runs of PAA (fully deprotonated) CG simulation

The RDFs between the aliphatic backbone atoms mapped by using SC1 bead and polarized water bead show a constant growth in the RDF value. First peak is observed at 0.38 ± 0.002.

5.4.7.5 RDF between SC1 and 'W' of polarized water bead

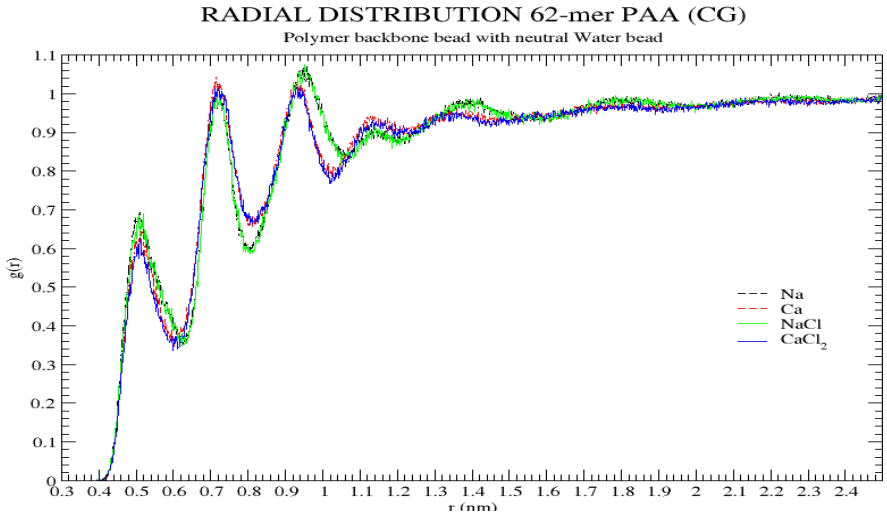


Figure 86 RDF between SC1 beads and 'W' of polarized water bead for all for all 4 systems averaged over 3 independent production runs of PAA (fully deprotonated) CG simulation

The RDFs between the aliphatic backbone atoms mapped by using SC1 bead and neutral water bead presents a first peak at 0.51 ± 0.002 nm having value 0.694 followed by a major peak at 0.72 ± 0.002 nm having a value of 1.023 and a second major peak at 0. 0.958 nm having value corresponding to 1.071.

5.4.7.6 RDF between SQa beads and Ions

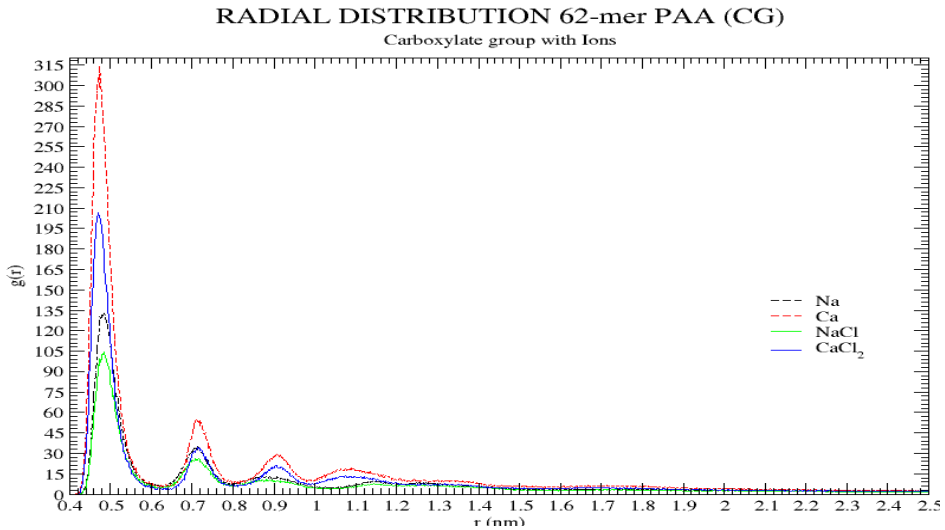


Figure 87 RDF between SQa beads and Ions for all for all 4 systems averaged over 3 independent production runs of PAA (fully deprotonated) CG simulation

As shown by Fig.87, the RDF between the SQa (carboxylate group) and monovalent and divalent counterions exhibits a strong first peak at 0.484 ± 0.002 nm, suggesting a preferred condensation of calcium ions on to the chain backbone of PAA. A weaker shoulder peak at can be observed at 0.706 ± 0.004 nm.

5.4.7.7 RDF between SC1 beads and Ions

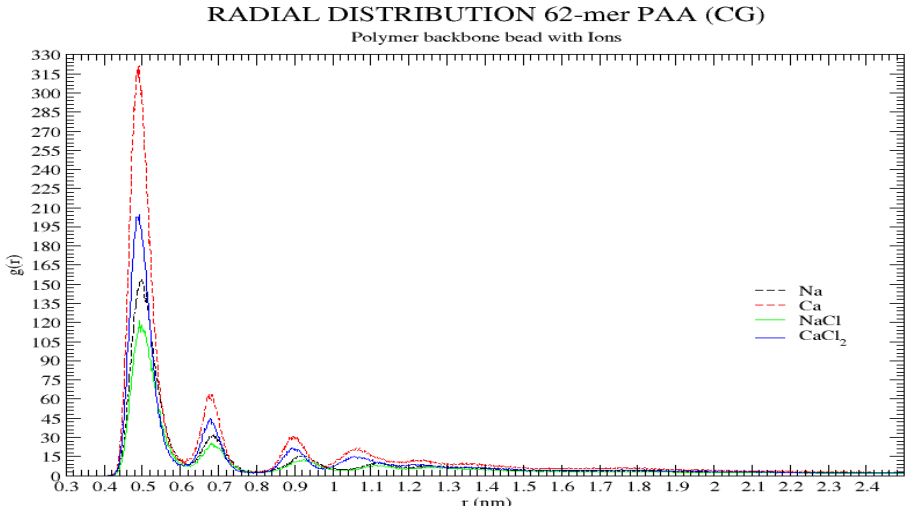


Figure 88 RDF between SC1 beads and Ions for all for all 4 systems averaged over 3 independent production runs of PAA (fully deprotonated) CG simulation

As shown by Fig.88, the RDF between the SC1 (aliphatic backbone) and monovalent and divalent counterions exhibits a strong first peak at 0.492 ± 0.004 nm, suggesting a preferred condensation of calcium ions on to the chain backbone of PAA. A weaker shoulder peak at can be seen at 0.678 ± 0.004 nm

5.5 ANALYSIS OF TRAJECTORIES FROM ALL-ATOMISTIC MD SIMULATIONS OF FULLY DEPROTONATED 20-MER MODEL OF GAMMA-L-POLYGLUTAMIC ACID (Γ-L-PGA)

The four generated boxes each include one γ-L-PGA short-chain model with periodic boundary conditions applied in all directions. For γ-L-PGA simulations, cubic boxes with 12 nm edge length were used for the short-chain models. The new refined SPC/E water model was used to solvate the system. In order to neutralize the system charge of the polymer, water molecules were replaced by ions according to the number of monomers in the polymer chain (see Table 30). In order to account for the ionic strength of the synthetic tap water (15°dH) used in experiments additional water molecules were exchanged by sodium and chloride ions. The setup for each simulated system is summarized in Table.4.

All simulations were performed on the RWTH Compute Cluster using GROMACS 2019 simulation package[116]. All systems were first energy minimized using steepest descent algorithm with 50,000 integration steps or until the maximum force on any atom in the system did not exceed a value of 1000.0 kJ/mol/nm. For neighbor searching, Verlet cutoff-scheme was used having short range Van der Waals cut-off of 1.0 with periodic boundary conditions. For electrostatic forces, Particle-mesh Ewald method of order 8 was used having short range cut-off of 1.2 keeping relative dielectric constant and relative dielectric constant of reaction field equal to 1. For Van der waals forces, twin range cut-offs with neighbor list cut-off and VdW cut-off are used. For the treatment of long-range electrostatic interaction and long-range dispersion corrections for energy and pressure was applied. Grid dimensions are controlled with Fourier spacing of 0.15. The relative strength of the Ewald-shifted direct potential at rcoulomb is given by 'ewald-rtol'. For doing PME for VdW-interactions, ewald-rtol-lj is used to control the relative strength of the dispersion potential at rvdw. Both the values are kept default 1e-5 and 1e-3 respectively.

Table 30 Number of ions in Na & Ca (salt free) and NaCl & Cacl2 (excess salt) systems of 20-mer γ-l-PGA

Salt	***Number of ions***	
	Na/Ca	***Cl***
Na	*20*	-
NaCl	*31*	*11*
Ca	*10*	-
$CaCl_2$	*21*	*22*

After heating, the systems were equilibrated for 140 ps each at 300 K using the velocity rescale thermostat (coupling time 1 ps) according to Bussi *et al.*[117] in NVT ensemble and at 1 bar using Berendsen barostat[118, 119] (coupling time 2 ps and compressibility 4.5×10^{-5} bar^{-1}) in NPT ensemble using verlet velocity integrator. During the equilibration, position restraints on every backbone atom of the PAA model and on every polymer bond were activated. Three independent production runs, 100 ns each, were performed for every system in NPT ensemble using a Verlet cutoff scheme with a cutoff distance of $r_{cut} = 1.2$ nm for the Lennard-Jones interactions according to a recent publication[120]. Consistently, a time step of 2 fs was used. The neighbor list was updated every 10 steps using the Verlet neighbor search (VNS) algorithm[120]. Bond lengths were held constant by the LINCS algorithm[113, 121] and Parrinello-Rahman extended-ensemble pressure coupling was used.

A chain length of 20 monomer of γ-L-PGA (fully deprotonated) is simulated in a system containing water and different quantity of monovalent and divalent ions (sodium, calcium,

sodium chloride and calcium chloride) (see Table 30). The system is equilibrated at 300k and 1 bar of pressure. The system is simulated for 100 ns with a time step of 2 femtoseconds. System configuration for production run is provided in appendix 9.10-9.13. Since no previous literature is available on γ-L-PGA, the simulation results will be verified from the experimental results being conducted by Dr. rer. nat Carlos Gonzalez Lopez, RWTH Aachen university.

5.5.1 Visualization of PAA interaction with Na, Ca, NaCl & $CaCl_2$ using VMD

Snapshots of the simulated system consisting of γ-L-PGA, 20-mer model of fully deprotonated PAA with Na, Ca, NaCl and $CaCl_2$ ions and water are shown in Fig. 89. 90, 91 & 92 respectively. The γ-L-PGA chain is shown as stick model. The snapshots of the systems are available on the following pages

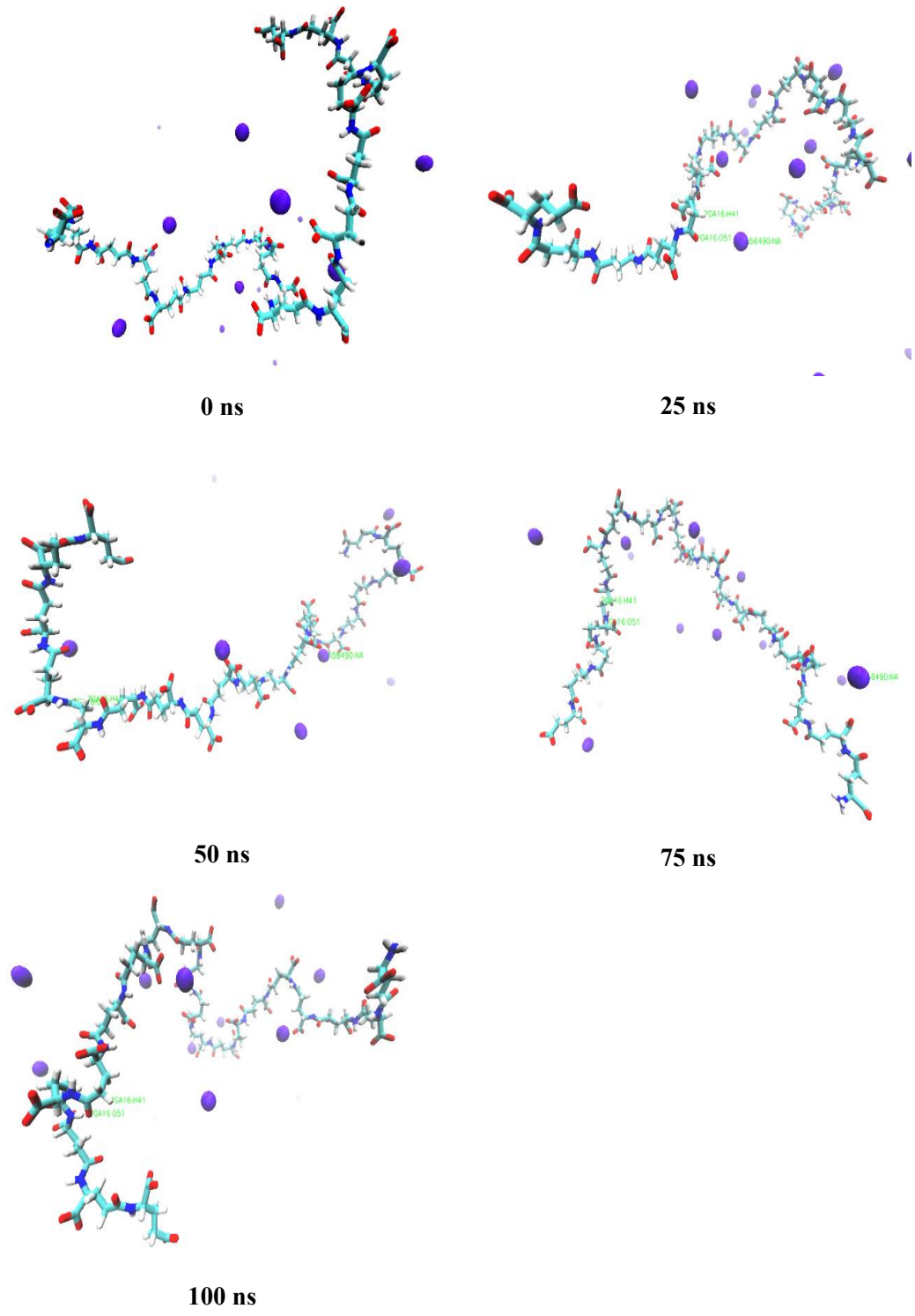


Figure 89 Snapshots taken from MD trajectory of 20-mer model of fully deprotonated γ-l-PGA with Na ions and water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (100 ns). γ-l-PGA is shown as sticks with carbon (cyan) and oxygen (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: sodium ions (violet).

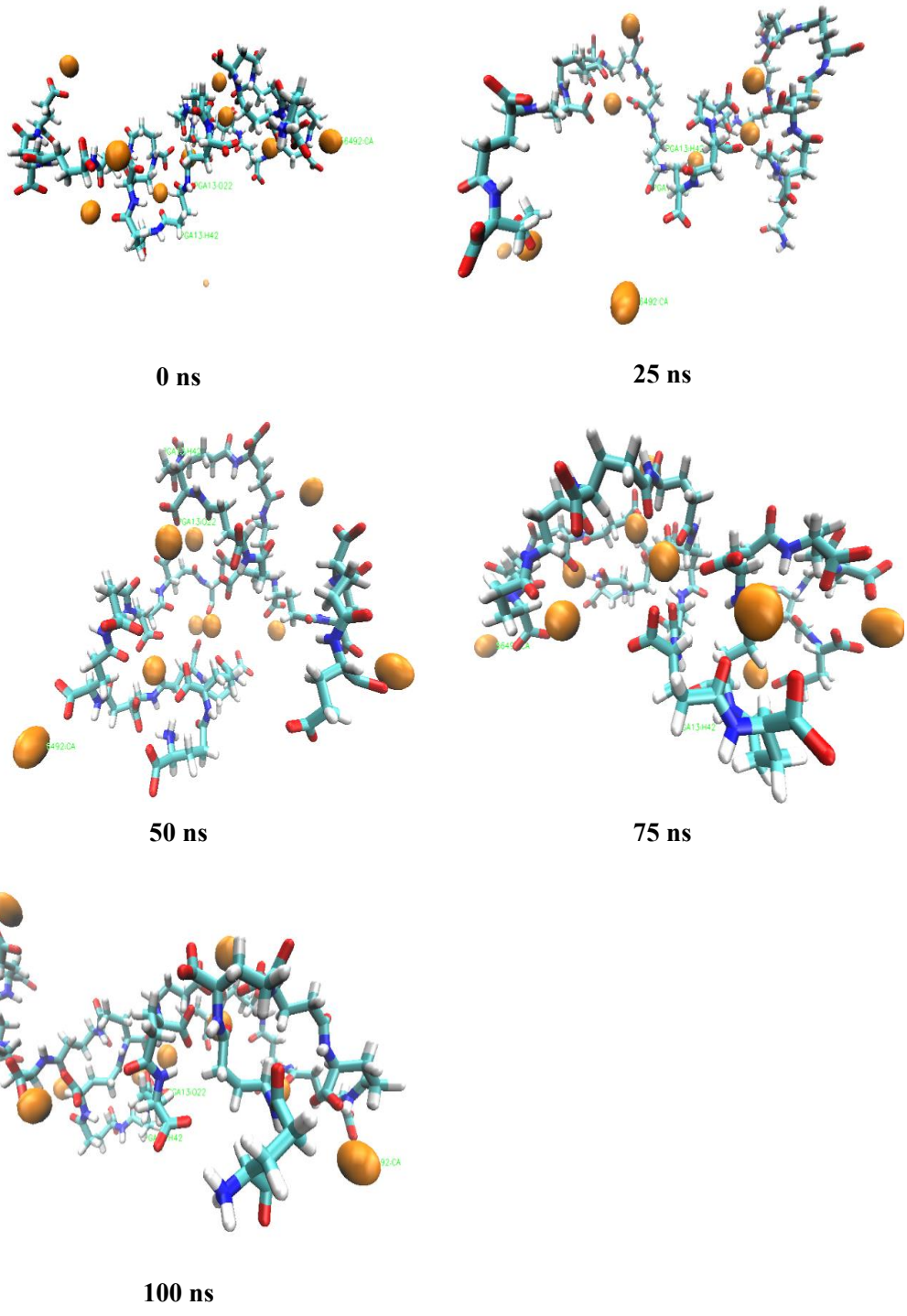


Figure 90 Snapshots taken from MD trajectory of 20-mer model of fully deprotonated γ-l-PGA with Ca2+ ions and water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (100 ns). γ-l-PGA is shown as sticks with carbon (cyan) and oxygen (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: calcium ions (orange)

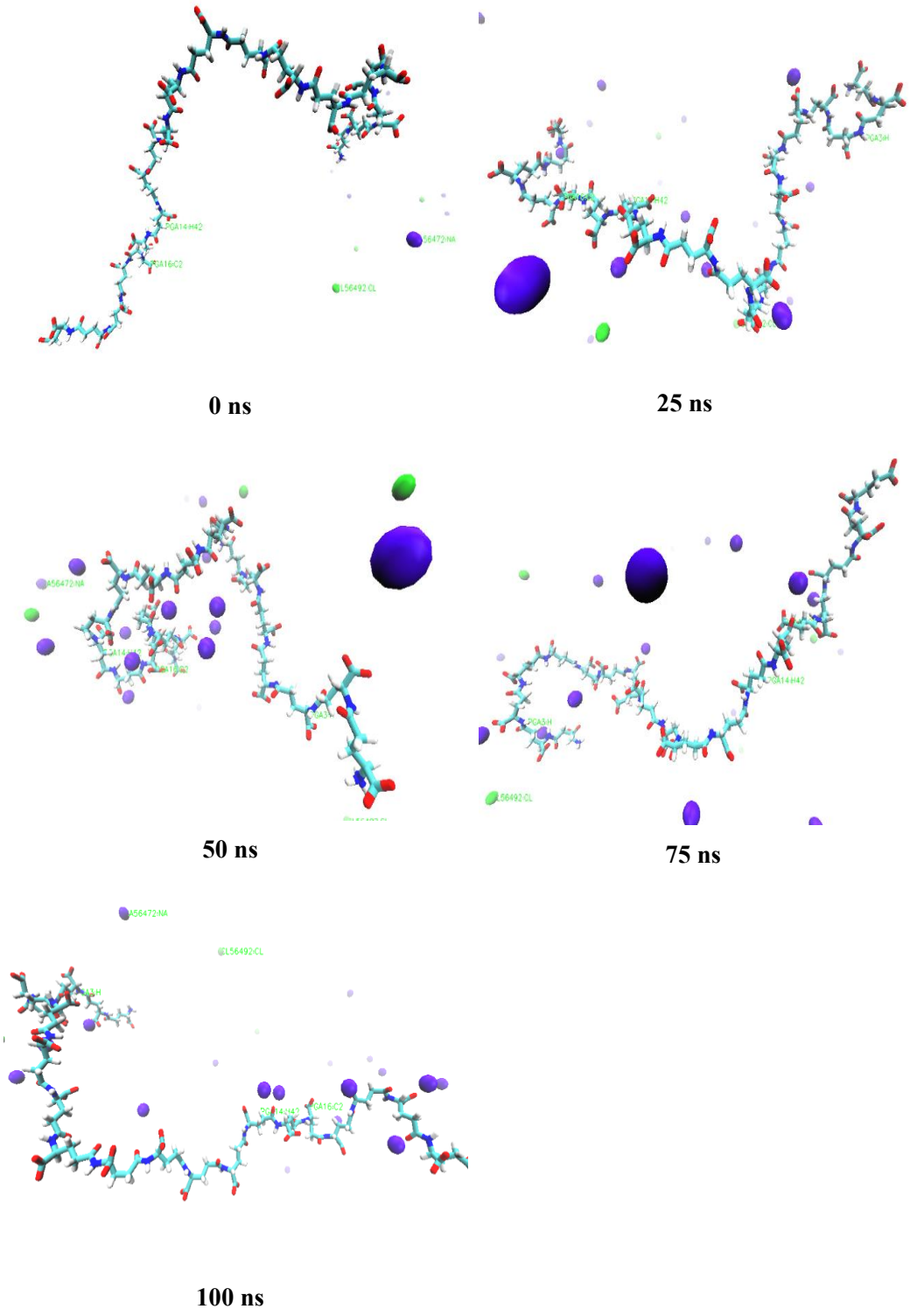


Figure 91 Snapshots taken from MD trajectory of 20-mer model of fully deprotonated γ-l-PGA with NaCl ions and water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (100 ns). γ-l-PGA is shown as sticks with carbon (cyan) and oxygen (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: sodium ions (violet) and chlorine ions (green).

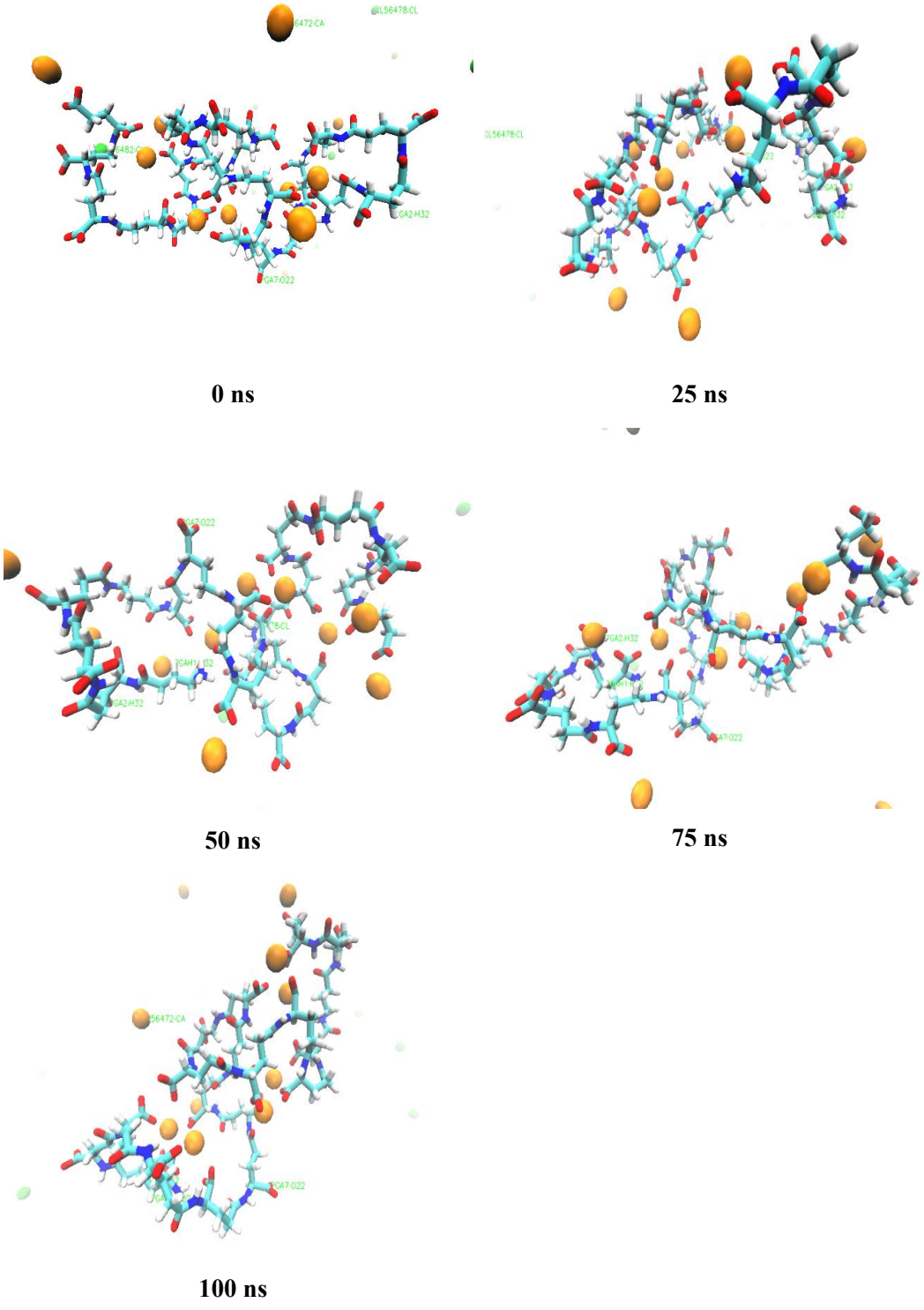


Figure 92 Snapshots taken from MD trajectory of 20-mer model of fully deprotonated γ-l-PGA with $CaCl_2$ ions and water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (100 ns). γ-l-PGA is shown as sticks with carbon (cyan) and oxygen (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: calcium ions (orange) and chlorine ions (green).

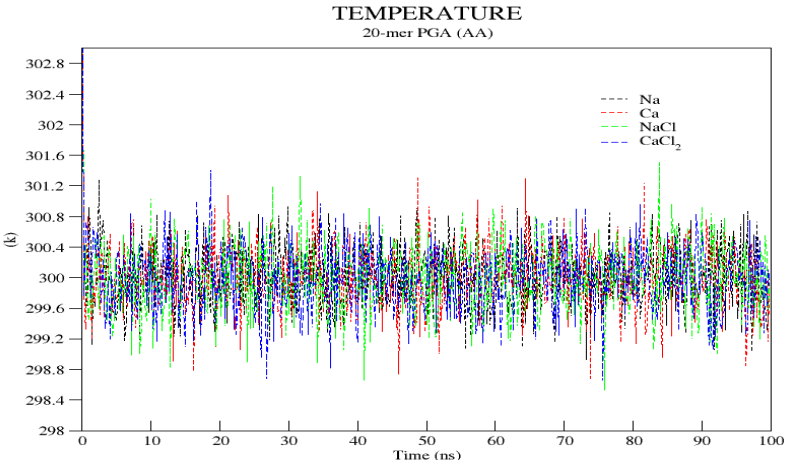
5.5.2 Quality assurance

Table 31 System properties calculated as a time average from all 3 single independent runs

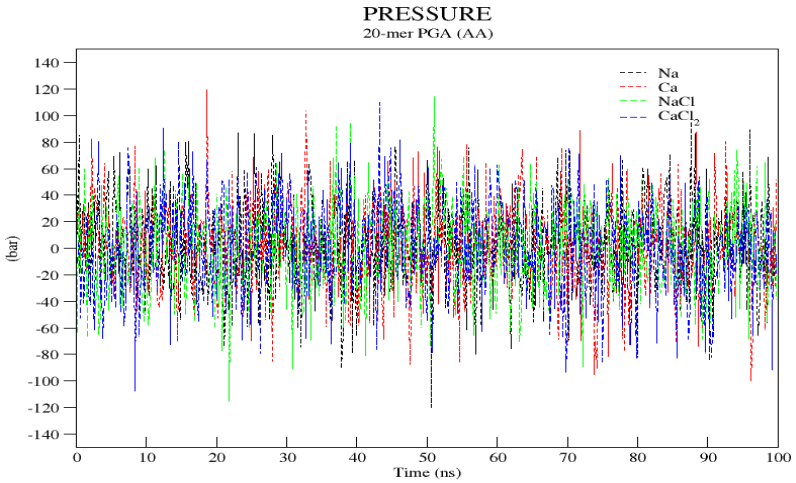
Property	***Average value of over 3 independent production runs***			
	Na	***Ca***	***NaCl***	***$CaCl_2$***
Temperature (K)	300.03	300.00	300.01	299.99
Pressure (bar)	2.31	2.06	0.31	0.50
Potential Energy (KJ/mol)	-2.65E+06	-2.66E+06	-2.66E+06	-2.68E+06
Kinetic Energy (KJ/mol)	4.23E+05	4.23E+05	4.23E+05	4.23E+05
Total Energy (KJ/mol)	-2.23E+06	-2.23E+06	-2.24E+06	-2.26E+06

Fist we check if the system is well equilibrated. As shown in Fig 93a,93b the average temperature and pressure, calculated as a time average from all 3 single independent runs, fluctuate steadily. The average temperature and pressure of all 3 independent production runs is 300.0 ± 0.1 K, respectively 1.30 ± 1 bar. The average energies (see Fig. 93c, 93d, 93e) of our systems are constant, the kinetic energy is 4.23×10^5 kJ/mol, the potential energy is $-2.66 \times 10^6 \pm 1.41 \times 10^4$kJ/mol and the total energy is $-2.24 \times 10^6 \pm 1.41 \times 10^4$ kJ/mol. The values are constant over the whole trajectories, this indicates a well equilibrated and stable system.

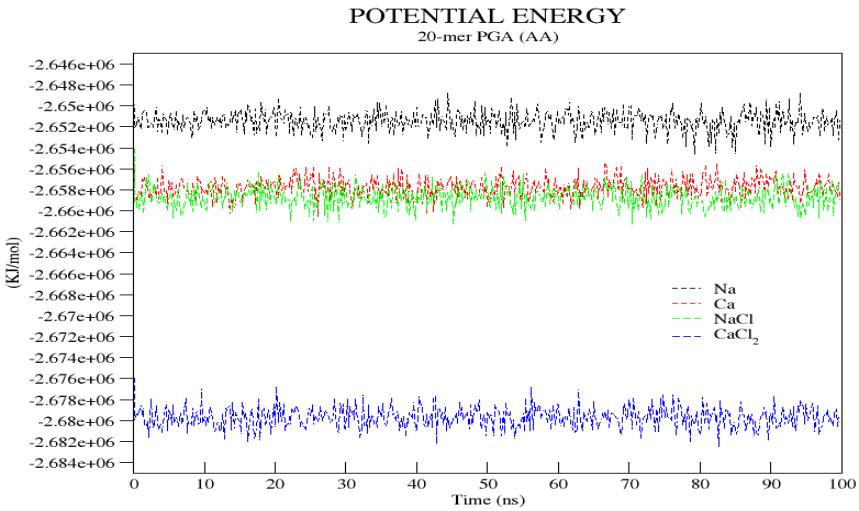
a)



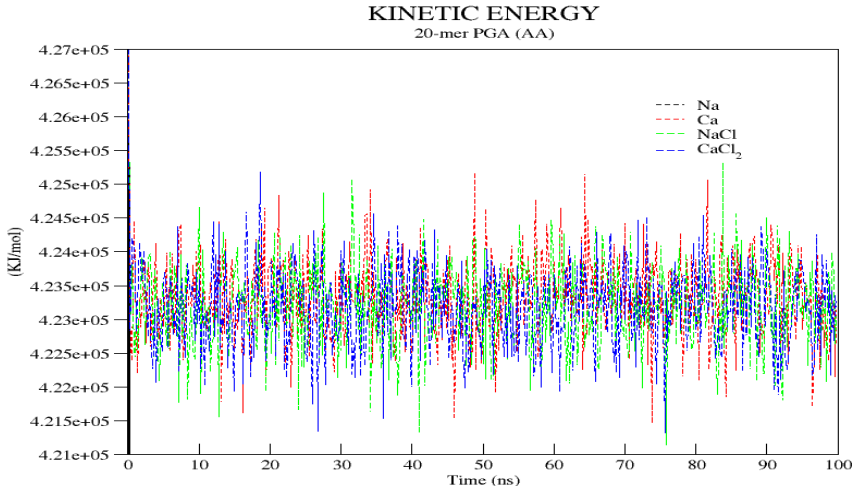
b)


PRESSURE
20-mer PGA (AA)
Na
Ca
NaCl
CaCl2
(bar)
Time (ns)

c)


POTENTIAL ENERGY
20-mer PGA (AA)
Na
Ca
NaCl
CaCl2
(KJ/mol)
Time (ns)

d)


KINETIC ENERGY
20-mer PGA (AA)
Na
Ca
NaCl
CaCl2
(KJ/mol)
Time (ns)

e)

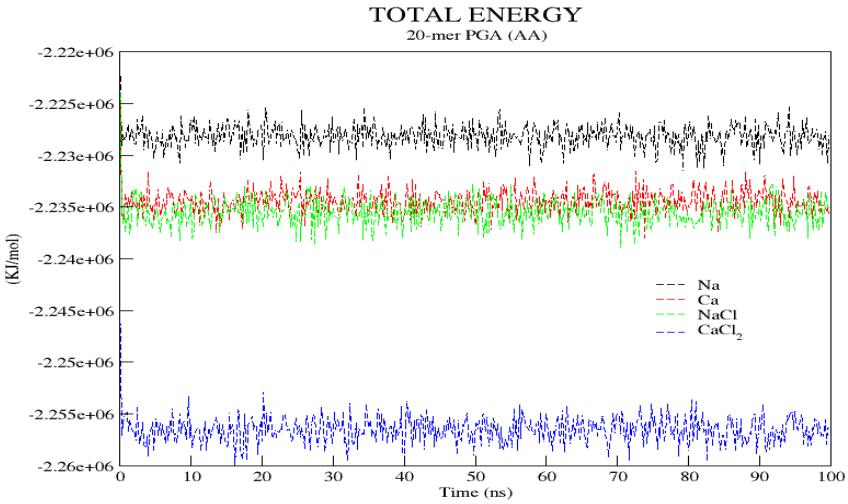


Figure 93 Average energy profiles of 3 independent AA production runs (100 ns) for 20-mer model of fully deprotonated γ-l-PGA with Na, Ca2+, NaCl, and $CaCl_2$ ions and water

5.5.3 Radius of Gyration

Table 32 Average Radius of gyration values calculated over 3 production runs

System	***Average (nm)***
Na	2.10
Ca	1.35
NaCl	2.00
$CaCl_2$	1.09

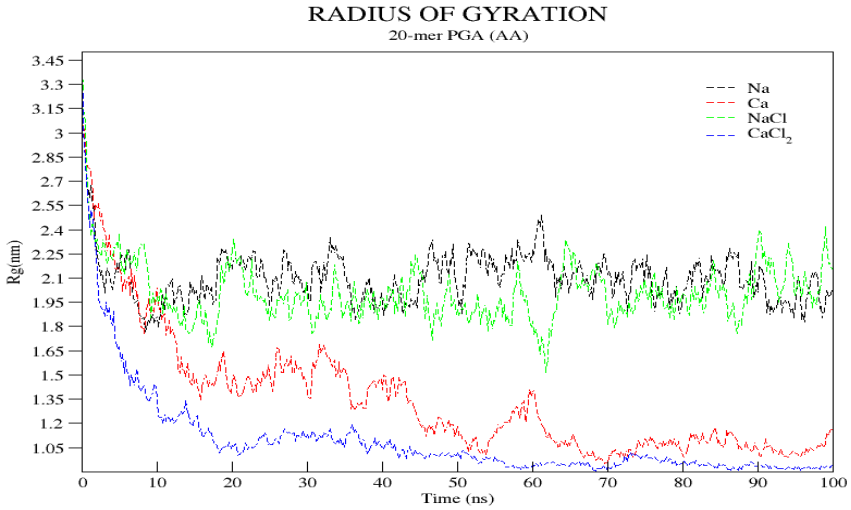


Figure 94 Average change of the radius of gyration over time for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems of γ-l-PGA (fully deprotonated) AA simulation

The average radius of gyration R_g of the fully deprotonated poly-γ-l-glutamic acid is shown in Fig 94. The time averaged value of R_g for all the simulated systems is presented in Table 32. This is slightly bigger than the R_g value of PAA because the chain backbone is longer (3 atoms per monomer instead of 2 atoms). Compared to the average curve of PAA in Fig 35 for the salt free system containing only sodium ions there is major drop in the Rg value in the first 10 nanosecond from 3.56 nm to 1.85 nm. A similar sharp drop in R_g was also observed in PAA. Similarly, for the excess salt system containing sodium chloride ions, a sharp drop in R_g is seen in the first 20 ns to a value of 1.38 nm similar to PAA. After short increase, the Rg value decreases to its final value after 20 ns of MD simulation (see Fig. 94) and fluctuates around the average R_g value till the end of the simulation. The average R_g value of γ-l-PGA fluctuates a lot because of the strong interaction with the sodium ions leading to structural changes of the PAA chain and inclusion of sodium ions but no major conformational changes are observed.

The average radii of gyration for γ-l-PGA in presence of calcium and calcium chloride ions are shown in Fig. 94. As the interaction between the carboxylate ions of γ-l-PGA and the calcium ions is preferred a sharp drop in R_g for both systems can be seen. The Rg value of γ-l-PGA decreases steadily during the first 20 ns and 52 ns simulation time for calcium and calcium chloride system respectively due to the strong chain folding. After this time the radius of gyration for γ-l-PGA remains steady. The time averaged Rg value of γ-l-PGA is shown in Table 32. A bigger difference was expected as γ-l-PGA has 297 atoms, which is nearly twice the number of atoms of PAA that has only 162 atoms. It can be concluded that the folding of PAA is hindered due to the strong repulsion of negatively charged carboxylate groups in close distance. As the γ-l-PGA chain has 3 instead of 2 backbone atoms, the repulsive forces are weaker, and the chain is able to fold stronger in the presence of divalent ions.

5.5.4 End–to–End distance

Table 33 Average end-to-end distance values calculated over 3 independent production runs

System	***Average (nm)***
Na	5.06
Ca	2.48
NaCl	4.80
$CaCl_2$	1.84

Again, the *R* value fluctuates a lot but there is no big change occurring after 20 ns simulation time for all the systems. This is in agreement with the value of R_g. The analysis of the polymer end-to-end distances in Fig.95, due to the folding of γ-l-PGA in the first 20 ns in the systems containing calcium chloride ions, the end-to-end distance shortens drastically from 6.5 nm to 1.63 nm on average, which is nearly the same as for the 20-mer model of PAA (see Fig. 36). In the system containing calcium ions γ-l-PGA shows a sharp drop in R from 6.74 nm to 3.07 nm in first 14 ns and steadily folds till the end of the simulation till it reaches its final value 1.81 nm. This sharp drop in value of R was not observed for PAA where the polymer chain steadily folded for 70 ns. For the systems consisting of sodium and sodium chloride ions the *R* value drops sharply for 10 ns showing sudden folding of the polymer chain as the chain interacts with sodium ions. The value of R increases in the 7

nanoseconds i.e. the chain attains a stretched conformation and fluctuates a lot but there is no big conformational change happening after 20 ns simulation time showing no big difference between PAA and γ-l-PGA in the presence of Na and NaCl ions. Since the γ-l-PGA backbone is longer than the PAA backbone, a bigger distance between the end groups was expected for γ-l-PGA. However, the analysis of *R* confirms the stronger folding behaviour of γ-l-PGA due to weaker electrostatic repulsion.

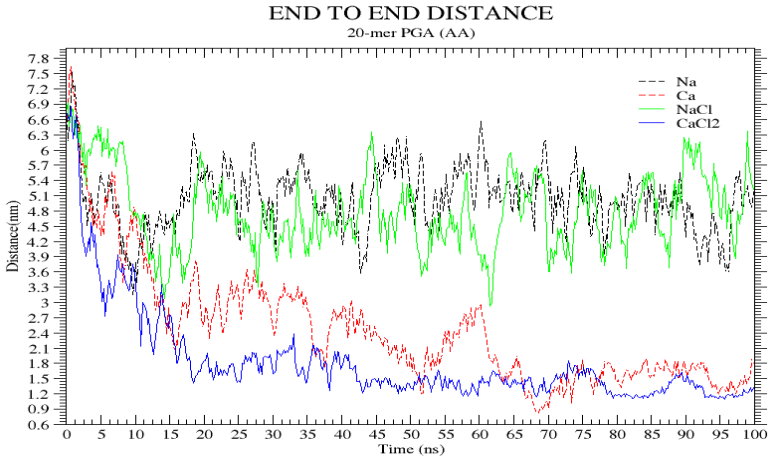


Figure 95 Average curve of the end-to-end distance R over time of 100 ns for all 4 systems over 3 independent production runs of γ-l-PGA (fully deprotonated) AA simulation.

5.5.5 Root-mean-square deviation

The backbone RMSD of γ-l-PGA with sodium and sodium chloride ions and water fluctuates around an average value of 1.62 ±0.05 nm (see Fig. 96). All the systems show drastic increase in RMSD for 17 ns and fluctuates around the average rmsd value till the end of simulation until it reaches its final value of 1.68, 2.4, 1.61, 2.6 nm for systems containing Na, NaCl, and $CaCl_2$ ions respectively. γ-l-PGA shows an increase in RMSD continuously through the 56 ns simulation time with calcium chloride ions later reaching a stable value of

Table 34 Average RMSD values calculated over 3 independent production runs

System	***Average (nm)***
Na	1.5707
Ca	2.2584
NaCl	1.6696
$CaCl_2$	2.5336

2.68 ±0.05 nm. In the presence of calcium ion γ-l-PGA has a major increase in RMSD value for 15 ns and continuously increases till the end of the simulation. The values of rmsd are in agreement with the results of R_g and R.

The comparison of the average backbone RMSD of PAA and γ-l-PGA can be observed from Fig 37 and 96. Nevertheless, the increase is very small indicating that γ-l-PGA remains stably folded.

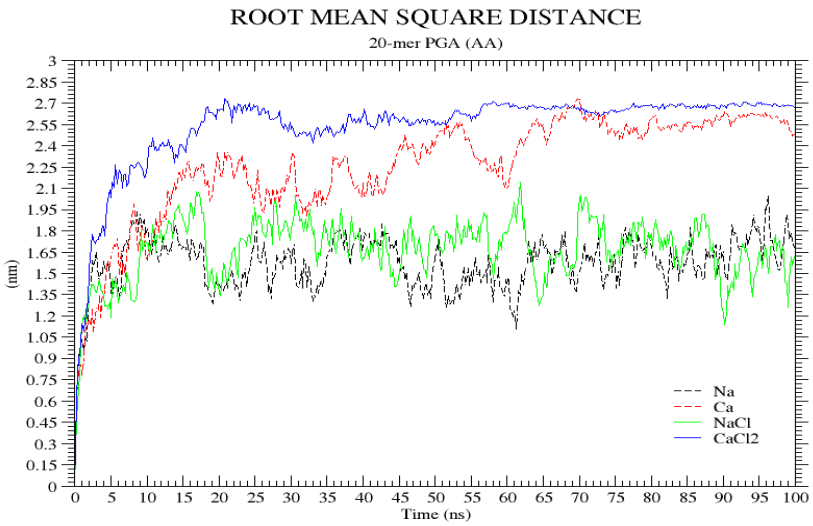


Figure 96 Average curve of the backbone RMSD over time of 100 ns for all 4 systems over 3 independent production runs of γ-l-PGA (fully deprotonated) AA simulation

5.5.6 Solvent accessible surface area

The result of solvent accessible surface area is in complete agreement with the previously calculated radius of gyration and end to end distance (see Fig. 94, 95). Due to the drastic bending and folding of γ-l-PGA for 100 ns with calcium ions SASA steadily decrease (see Fig. 97). Not surprisingly, the *SASA* of PAA is far below the value for γ-L-PGA (see Fig. 38) due to the different structure and backbone length of γ-L-PGA. In presence of sodium and sodium chloride ions the value R_g and R remained consistent over the simulation period of 100 ns. This effect can be clearly seen by the value of SASA for Na and NaCl systems where the value stays consistent after an initial drop in the value. The average curves of the solvent accessible surface area for γ-l-PGA in presence of various ions are shown in Fig. 97. The time averaged SASA for γ-l-PGA are presented in Table 35. Again, the difference is between PAA and γ-l-PGA is caused by the size of the γ-l-PGA chain.

Table 35 Average SASA values calculated over 3 independent production runs

System	***Average (nm²)***
Na	36.6614
Ca	29.9056
NaCl	36.3592
$CaCl_2$	27.9185

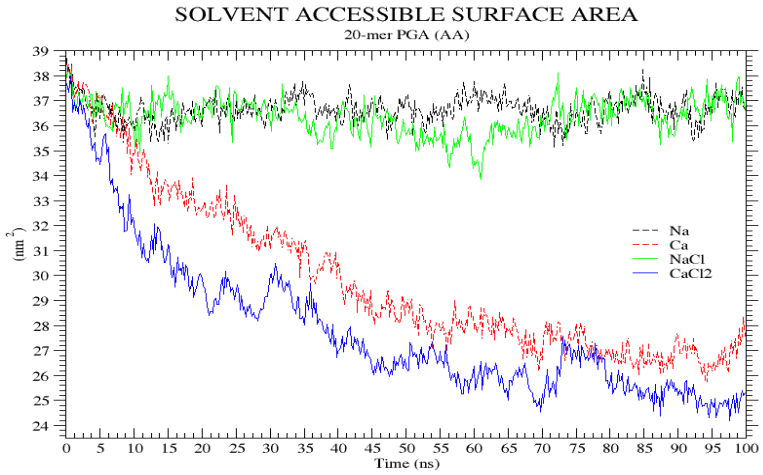


Figure 97 Average curve of the SASA over time of 100 ns for all 4 systems over 3 independent production runs of γ-l-PGA (fully deprotonated) AA simulation

The comparison with Fig.38 indicates decrease in the SASA due to the presence of divalent ions in the simulation system. Although both polymers are in contact with ions during the simulation, which should theoretically lower the SASA, our model correctly depicts it and shows higher interaction between calcium ions in salt free and excess salt systems. As shown before, the polymers tend have a more folded conformation when divalent ions are present, therefore the SASA decreases. Not surprisingly, the SASA of PAA is far below the value for γ-L-PGA (see Fig. 37) as a result of the different structure and backbone length of γ-L-PGA.

5.5.7 Hydrogen bonds

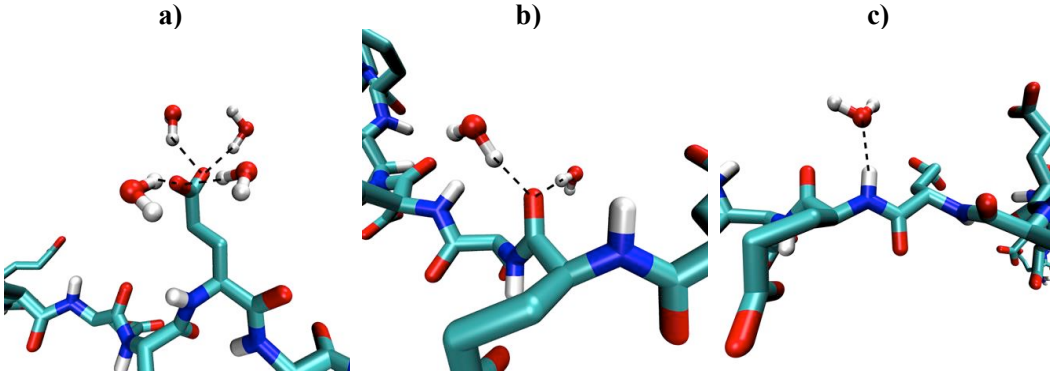


Figure 98 3 types of hydrogen bonds between γ-l-PGA and water: (a) between carboxylate oxygens and water hydrogens, (b) between carbonyl oxygens and water hydrogens and (c) between water oxygens and amino hydrogens

The hydration behaviour of γ-l-PGA with the attendance of ions is shown in Fig.99. For γ-l-PGA simulation with only water and ions, average hydrogen bonds formed are shown in Table 36. The number of hydrogen bonds between water and divalent ions decreases due to the folding of the polymer chain and contraption of divalent ions leaving smaller number of free carboxylate ions thus the smaller number of hydrogen bonds.

Table 36 Average hydrogen bonds calculated over 3 independent production runs

System	Average
Na	180.1767
Ca	137.7300
NaCl	179.3593
$CaCl_2$	131.6600

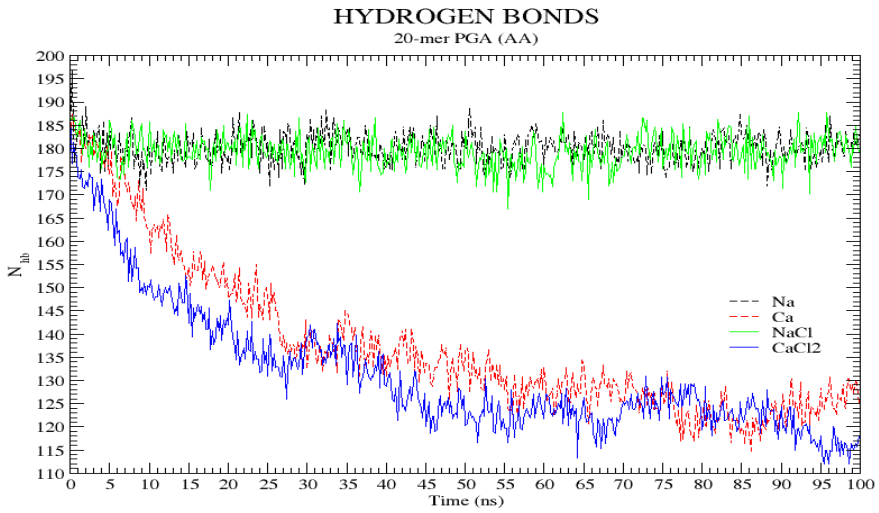


Figure 99 Average curve of the change of hydrogen bonds over time of 100 ns for all 4 systems over 3 independent production runs of γ-l-PGA (fully deprotonated) AA simulation

In total hydrogen bonds formed by γ-l-PGA (see Fig. 99) with the surrounding water molecules, divided on hydrogen bonds between carboxylate oxygen atoms and water hydrogen atoms, hydrogen bonds between carbonyl oxygen atoms and water hydrogen atoms and hydrogen bonds between water oxygen atoms and amino hydrogen atoms. As expected, the number of hydrogen bonds between γ-l-PGA and water decreases in presence of divalent ions. In water these groups would be able to from 3 hydrogen bonds per carboxylate.

The average N_{hb}, calculated as time average (see Table 36). As we can see from the RDF between carboxylate oxygens and water hydrogen atoms , every oxygen atom is surrounded by nearly 2 hydrogen atoms for Na and NaCl system and 1.5 hydrogen atoms for Ca and $CaCl_2$ systems.

5.5.8 Radial Distribution Functions

5.5.8.1 RDF between the centre of mass of γ-l-PGA residues and the Hydrogen atoms of water molecules

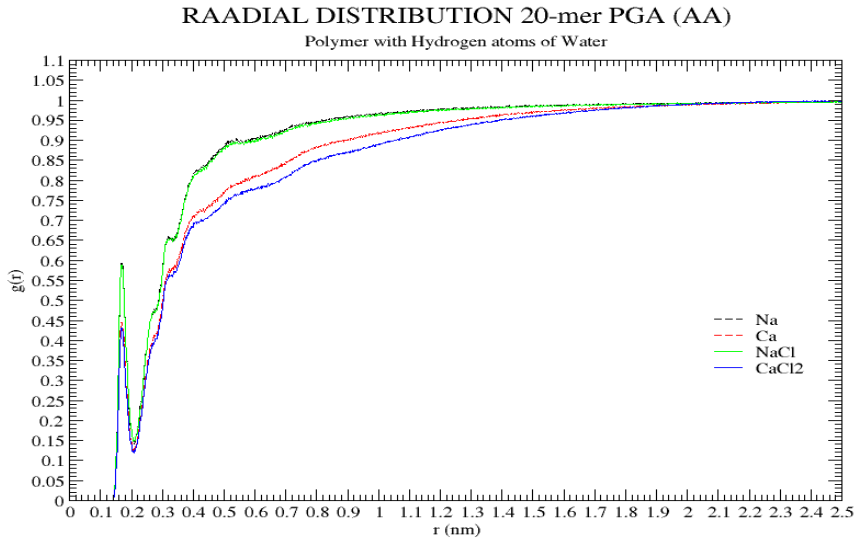


Figure 100 Average RDF between γ-l-PGA (centre of mass) and water hydrogen (HW) for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

The RDF between the centre of mass of γ-l-PGA atoms and hydrogen atoms of water molecules shown in Fig.100 show a peak at 0.322 ± 0.002 nm.

5.5.8.2 RDF between centre of mass of the γ-l-PGA residues and the Oxygen atoms of water molecules

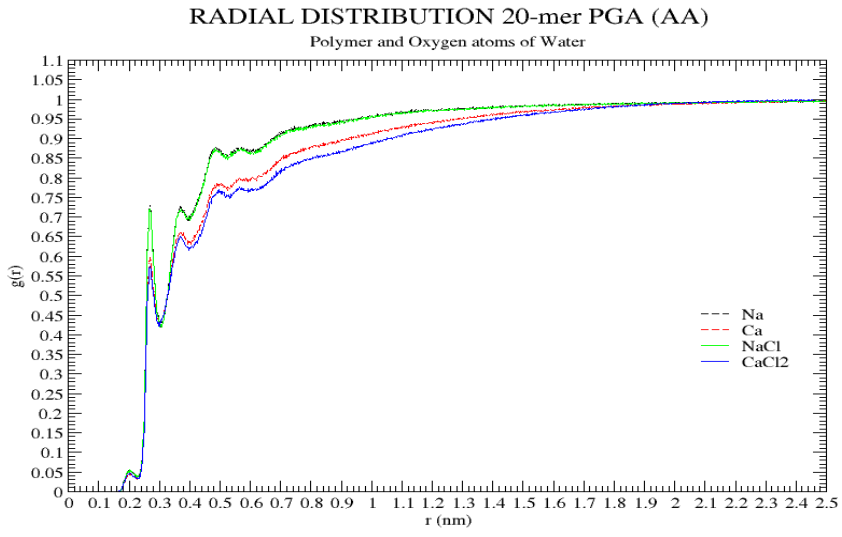


Figure 101 Average RDF between γ-l-PGA (centre of mass) and water oxygen (OW) for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation.

The sharp peak at 0.268 ± 0.002 nm in Fig. 101 indicates a well-structured solvation shell with ordered distribution of water molecules around the polymer chain.

5.5.8.3 RDF between centre of mass of the γ-l-PGA residues and Ions

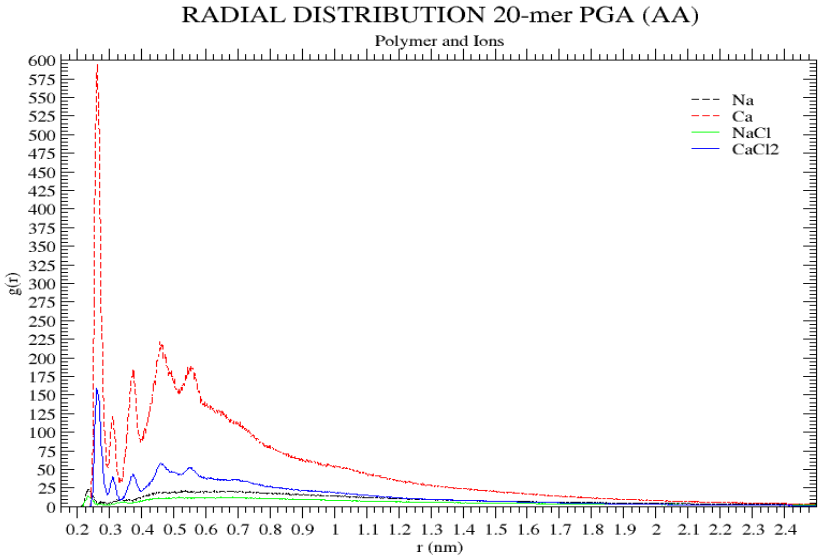


Figure 102 Average RDF between γ-l-PGA (centre of mass) and ions for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation.

The RDF between the centre of mass of γ-l-PGA atoms and ions is shown in Fig.102 shows ion condensation around the polymer at 0.262 ± 0.002 nm for calcium and calcium chloride systems. For, sodium and sodium chloride system peak can be seen at 0.236 ± 0.002 nm.

5.5.8.4 RDF between carboxylate group and Hydrogen atoms of Water

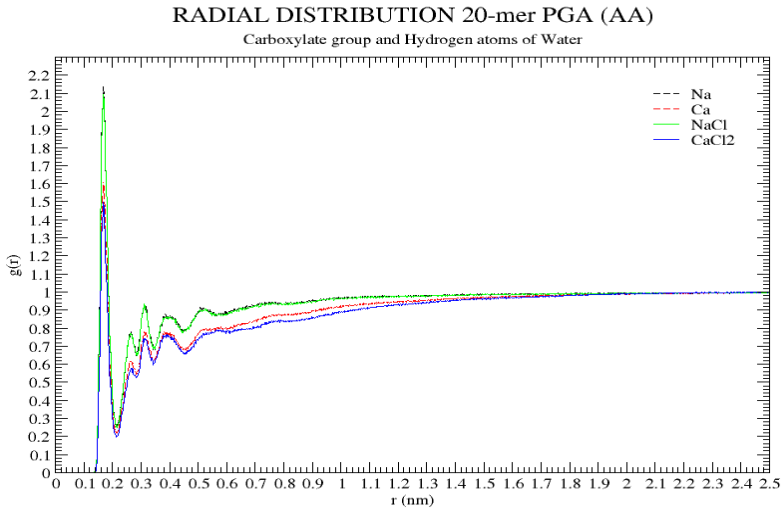


Figure 103 Average RDF between oxygen atoms of carboxylate and hydrogen atoms (HW) of water for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation.

The RDF between the carboxylate oxygen atoms and water hydrogen atoms in Fig.103 shows main peaks at 0.170 ± 0.002 nm followed by peaks at 0.269 ± 0.002 and 0.312 ± 0.002 nm.

5.5.8.5 RDF between carboxylate group and Oxygen atoms of Water

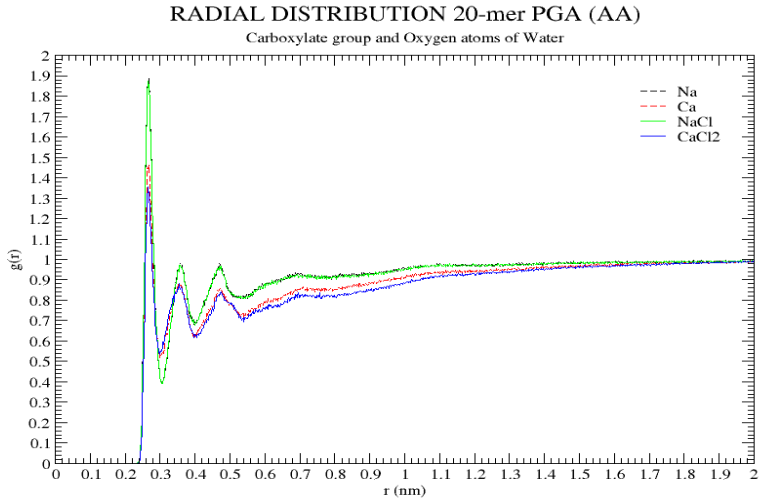


Figure 104 Average RDF between oxygen atoms of carboxylate and oxygen atoms of water (OW) for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation.

The RDF between of the oxygen atoms of the carboxylate group with respect to water hydrogens shows a strong peak at 0.266 ± 0.002 nm and followed by a shoulder peak at 0.358 ± 0.002 nm follows.

5.5.8.6 RDF between carboxylate group and Ions

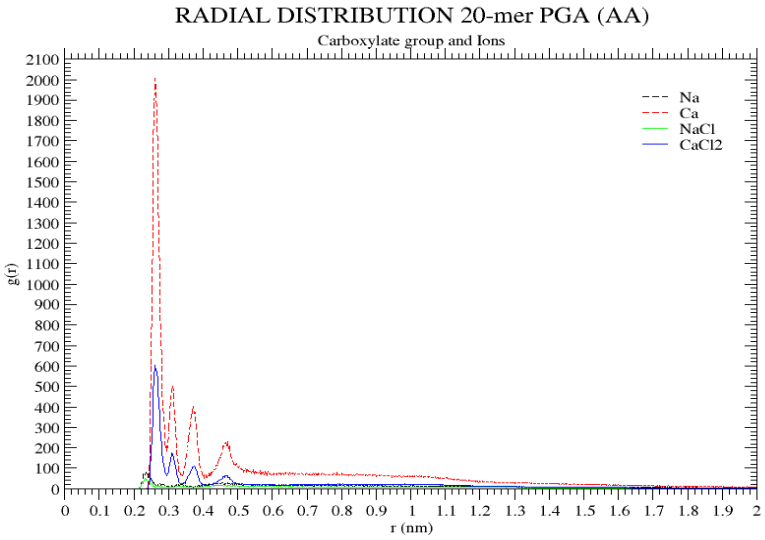


Figure 105 Average RDF between oxygen atoms of carboxylate and ions for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation.

As shown by Fig.105, the RDF between the carboxylate oxygens and sodium counterions exhibits a strong first peak at 0.232 ± 0.002 nm, suggesting a preferred condensation of sodium ions on to the chain backbone of PAA. For calcium ions and carboxylate oxygens exhibits a strong first peak at 0.262 ± 0.003 nm and a weaker shoulder peak at 0.312 ± 0.002 nm.

5.5.8.7 RDF between carbonyl group and Hydrogen atoms of Water

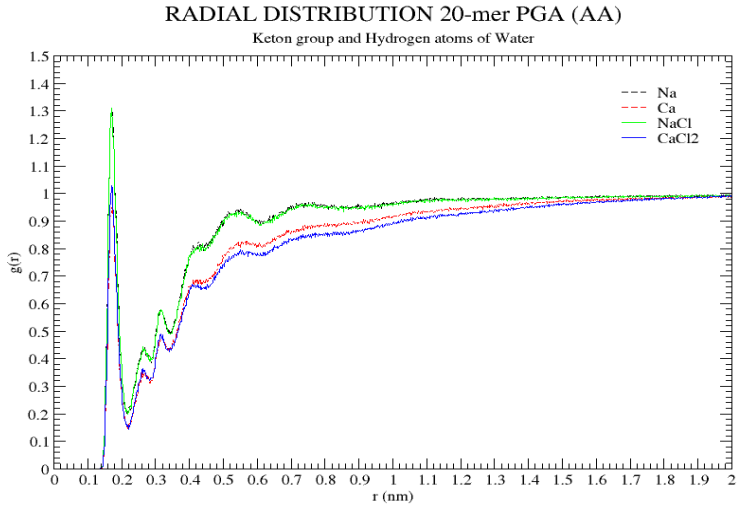


Figure 106 Average RDF between oxygen atoms of ketone and water hydrogen atoms (HW) for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation.

The RDF between the carbonyl group atoms and water hydrogen atoms in Fig. 106 shows the major peak at 0.170 ± 0.002 nm followed by minor peaks at 0.264 ± 0.002 and 0.316 ± 0.002 nm.

5.5.8.8 RDF between carbonyl group and Oxygen atoms of Water

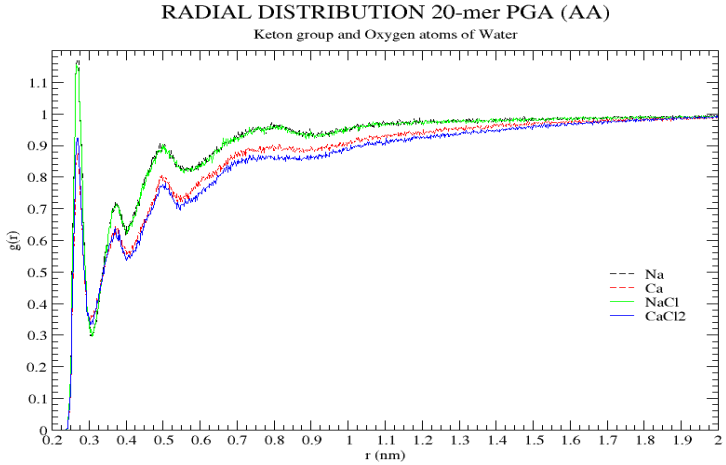


Figure 107 Average RDF between oxygen atoms of ketone and water oxygen atoms (OW) for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation.

The RDF between of the carbonyl group with respect to oxygen atoms of water shows a strong peak at 0.266 ± 0.002 nm (see Fig.107). A shoulder peak at 0.374 ± 0.002 nm and 0.5 ± 0.002 follows.

5.5.8.9 RDF between carbonyl group and Ions

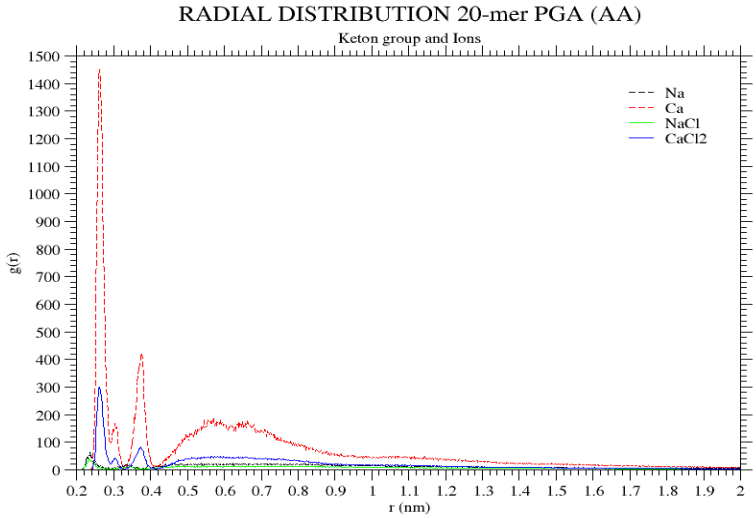


Figure 108 Average RDF between oxygen atoms of ketone and ions for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation.

As shown by Fig108, the RDF between carbonyl group and monovalent and divalent counterions exhibits a strong first peak at 0.262 ± 0.002 nm and a minor peak at 0.376 ±0.002 , suggesting a preferred condensation of calcium ions on to the chain backbone of PAA. For sodium ions and sodium chloride, exhibits a strong first peak at 0.232 ± 0.003 nm.

5.5.8.10 RDF between Amino group and Hydrogen atoms of Water

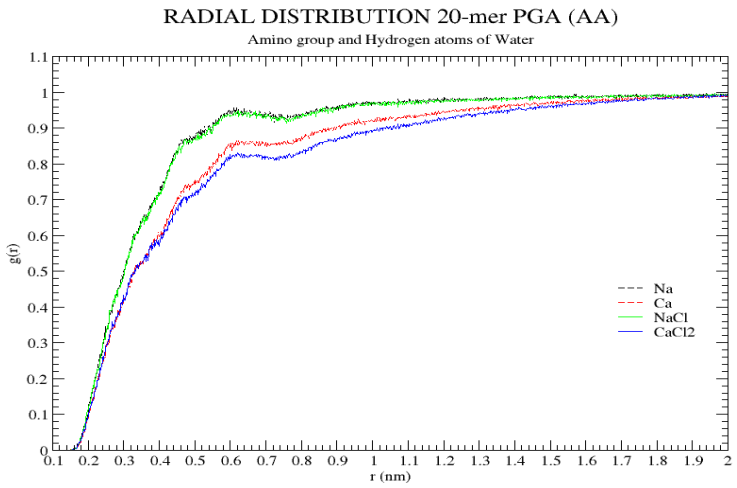


Figure 109 Average RDF between amino group and water hydrogen (OW) for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

The RDF between the amino group atoms and water hydrogen atoms in Fig.109 shows a linear growth in the distribution function till it gets converged into a stable value of 1.

5.5.8.11 RDF between Amino group and Oxygen atoms of Water

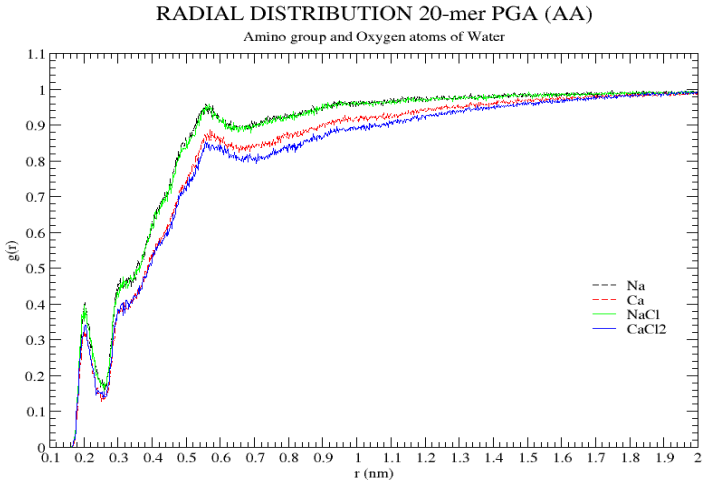


Figure 110 Average RDF between amino group and water oxygen (OW) for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

The RDF between the amino group with respect to oxygen atoms of water shows a strong peak at 0.397 ± 0.002 nm (see Fig 110).

5.5.8.12 RDF between Amino group and Ions

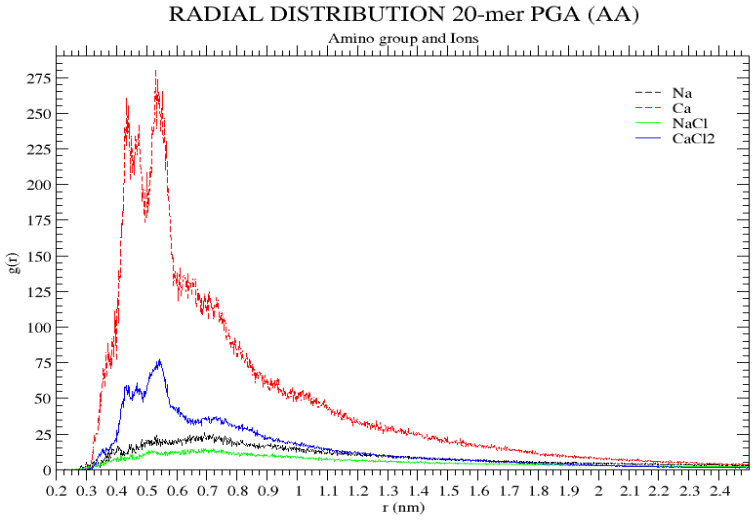


Figure 111 Average RDF between amino group and ions for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

5.6 ANALYSIS OF TRAJECTORIES FROM COARSE GRAIN MD SIMULATIONS OF FULLY DEPROTONATED 20-MER MODEL OF GAMMA-L-POLYGLUTAMIC ACID (Γ-L-PGA)

γ-l-PGA (fully deprotonated) MARTINI 20-mer model is simulated in a system containing polarizable water model from MARTINI and 4 different ions individually (see Table 30). Polymer model was generated using the mapping scheme described in section 3.2.2 using MARTINI scheme for modelling of polymer.

The four generated boxes each include one γ-l-PGA short-chain model. Periodic boundary conditions were applied in all directions. For γ-l-PGA simulations, cubic boxes with 12 nm edge length were used for the short-chain models. The new refined polarizable water model from MARTINI was used to solvate the system. In order to neutralize the system charge of the polymer, water molecules were replaced by ions according to the number of monomers in the polymer chain (see Table 30). In order to account for the ionic strength of the synthetic tap water (15°dH) used in experiments additional water molecules were exchanged by sodium and chloride ions. The setup for each simulated system is summarized in Table.30.

All simulations were performed on the RWTH Compute Cluster using GROMACS 2019 simulation package[116]. All systems were first energy minimized using steepest descent algorithm with 500,000 integration steps or until the maximum force on any atom in the system did not exceed a value of 10.0 kJ/mol/nm. For neighbor searching, Verlet cutoff-scheme was used having short range Van de Waals cut-off of 1.0 with periodic boundary conditions. For electrostatic forces, Particle-mesh Ewald method of order 8 was used having short range cut-off of 1.2 keeping relative dielectric constant 2.5 and relative dielectric constant of reaction field equal to 1. For Van der waals forces, twin range cut-offs with neighbor list cut-off and VdW cut-off are used. For the treatment of long-range electrostatic interaction and long-range dispersion corrections for energy and pressure was applied. Grid dimensions are controlled with flourier spacing of 0.15. The relative strength of the Ewald-shifted direct potential at rcoulomb is given by 'ewald-rtol'. For doing PME for VdW-interactions, ewald-rtol-lj is used to control the relative strength of the dispersion potential at rvdw. Both the values are kept default 1e-5 and 1e-3 respectively.

After heating, the systems were equilibrated for 50 ns each at 300 K using the velocity rescale thermostat (coupling time 1 ps) according to Bussi *et al.*[117] in NVT ensemble and at 1 bar using Berendsen barostat[118, 119] (coupling time 12 ps and compressibility $3 \cdot 10^{-4}$ bar^{-1}) in NPT ensemble using velocity verlet integrator. During the equilibration, position restraints on every polymer bead were activated. Three independent production runs, 500 ns each, were performed for every system in NPT ensemble using Parrinello-Rahman extended-ensemble pressure coupling and Verlet cutoff scheme with a cutoff distance of $r_{cut} = 1.2$ nm for the Lennard-Jones interactions according to a recent publication[120]. The Particle Mesh Ewald (PME) method[112] with a grid spacing of 0.15 nm in agreement with the parameters employed during the parameterization of the refined polarizable water model[105]. Consistently, a time step of 20 fs was used. The neighbor list was updated every 10 steps using the Verlet neighbor search (VNS) algorithm[120]. Bond lengths were held constant by the LINCS algorithm[113, 121] and the background permittivity was set to $\varepsilon_{bg} = 2.5$ as recommended when using the polarizable water model[106, 120]. The final box sizes are shown in Table.4.

A chain length of 20 monomer of γ-l-PGA (fully deprotonated) is simulated in a system containing water and different quantity of monovalent and divalent ions (sodium, calcium, sodium chloride and calcium chloride) (see Table 30). The system is equilibrated at 300k and 1 bar of pressure. The system is simulated for 500 ns with a time step of 20

femtoseconds. System configuration for production run is provided in appendix 9.15-9.18. The simulation results are validated with the results from all atomistic simulation

5.6.1 Visualization of gamma-l-polyglutamic acid (γ-l-PGA) interaction with Na, Ca, NaCl & $CaCl_2$ using VMD

Some representative snapshots of the MD simulations of the γ-l-PGA MARTINI models are shown in Fig.112-115 for short 20-mer CG chain model with different ions. The snapshots were taken after 500 ns of MD simulation. The polarized water beads have not been presented in the images for better visualisation of the polymer chain and its interaction with ions. The γ-l-PGA chain is shown as stick model. Snapshots from each system can be seen on the following pages.

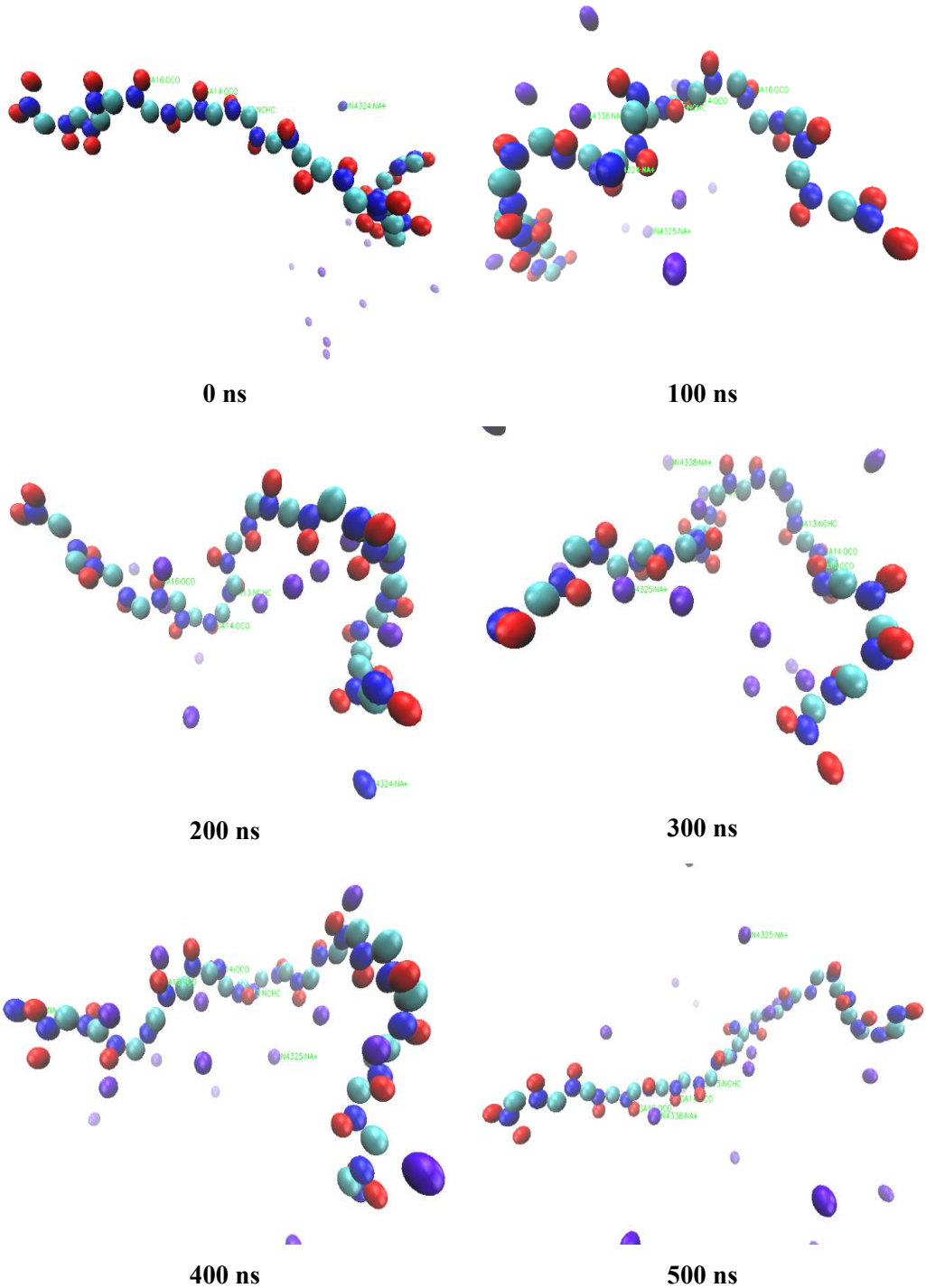


Figure 112 Snapshots taken from MD trajectory of 20-mer MARTINI model of fully deprotonated γ-l-PGA with Na ions and polarized water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (500 ns). γ-l-PGA is shown as beads with SNa (cyan), SQd/SNd (blue) and SQa (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: sodium ions (violet).

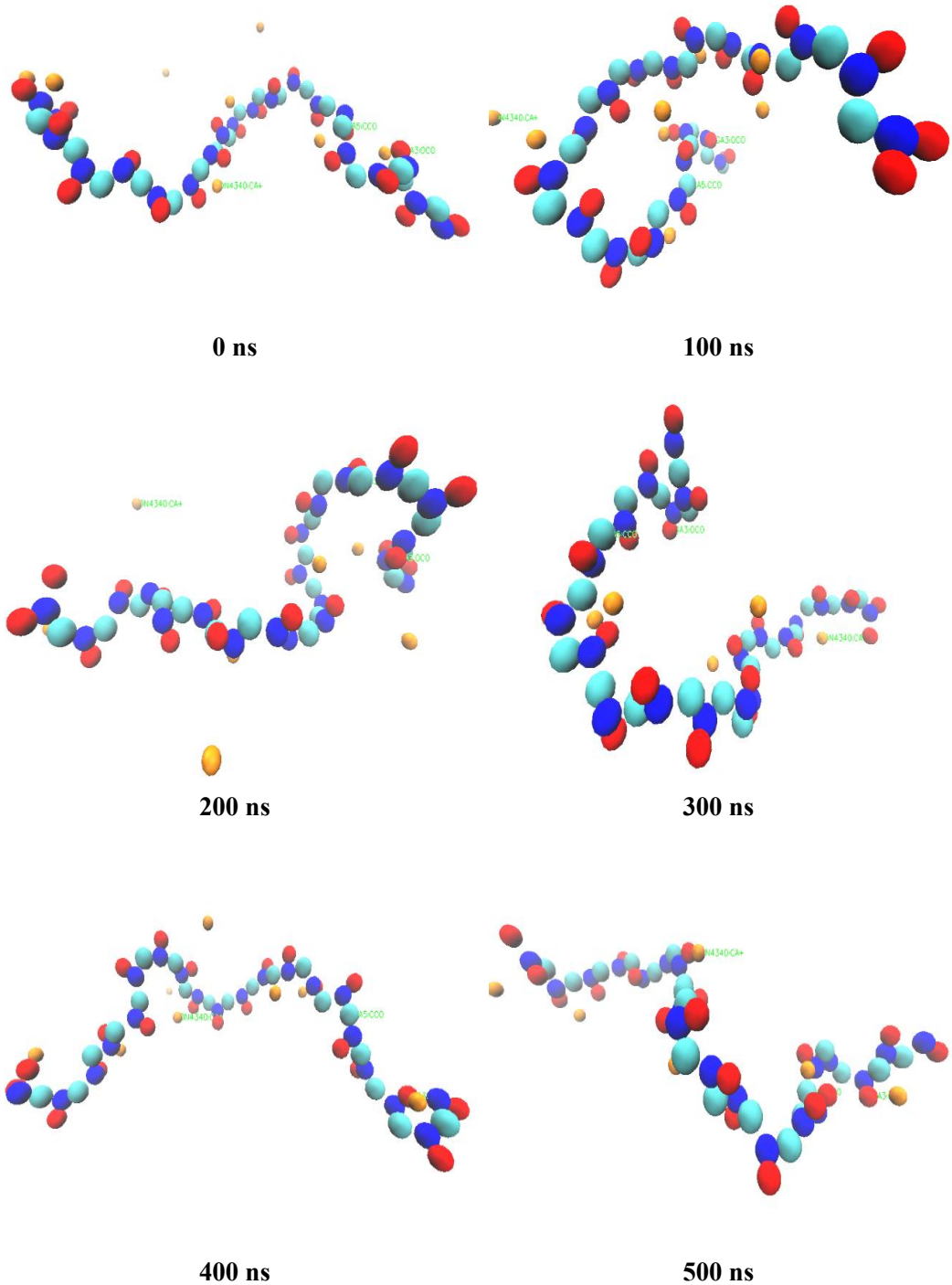


Figure 113 Snapshots taken from MD trajectory of 20-mer MARTINI model of fully deprotonated γ-l-PGA with Ca2+ ions and polarized water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (500 ns). γ-l-PGA is shown as beads with SNa (cyan), SQd/SNd (blue) and SQa (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: calcium ions (orange)

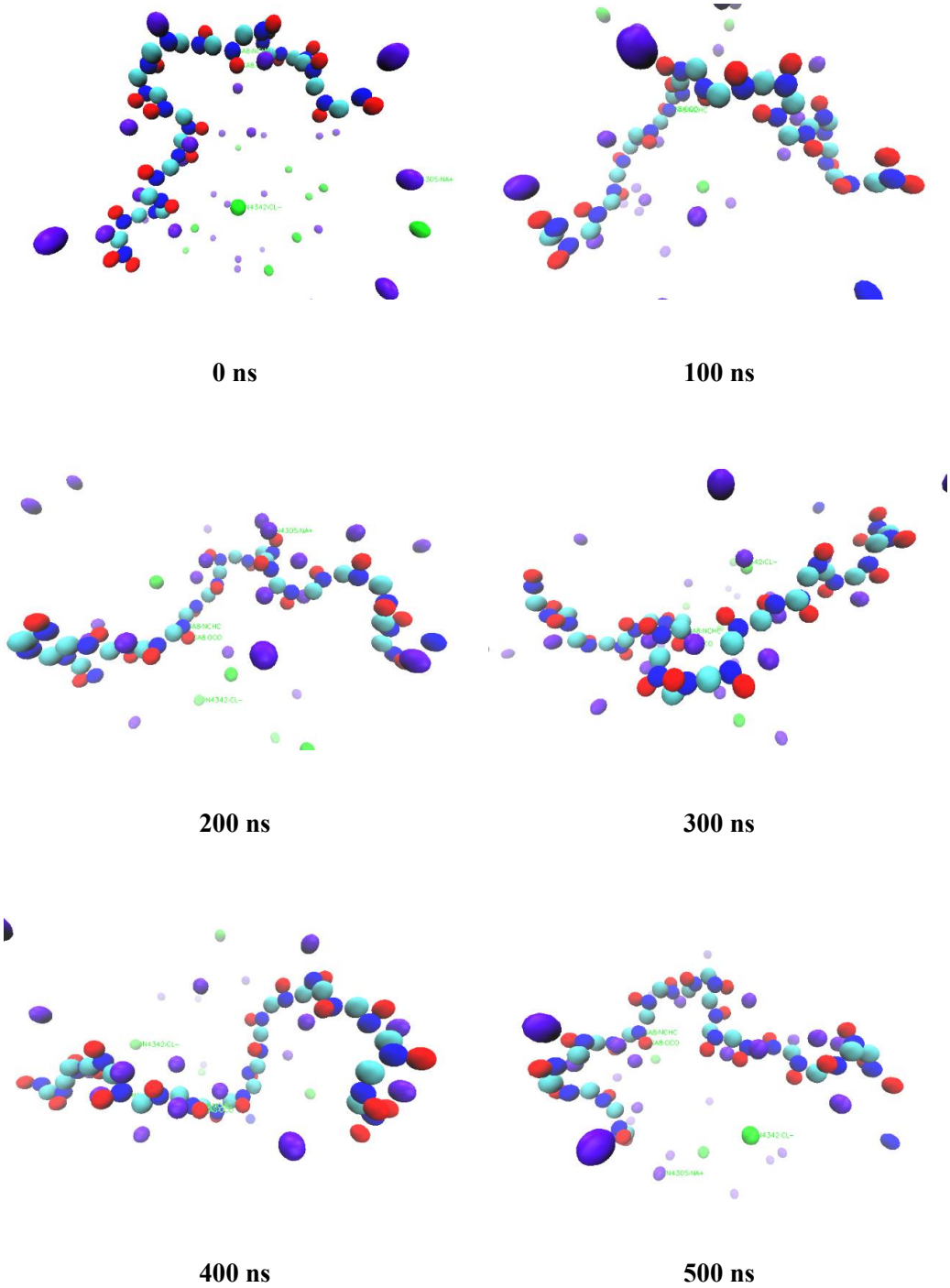


Figure 114 Snapshots taken from MD trajectory of 20-mer MARTINI model of fully deprotonated γ-l-PGA with NaCl ions and polarized water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (500 ns). γ-l-PGA is shown as beads with SNa (cyan), SQd/SNd (blue) and SQa (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: sodium ions (violet) and chlorine ions (green).

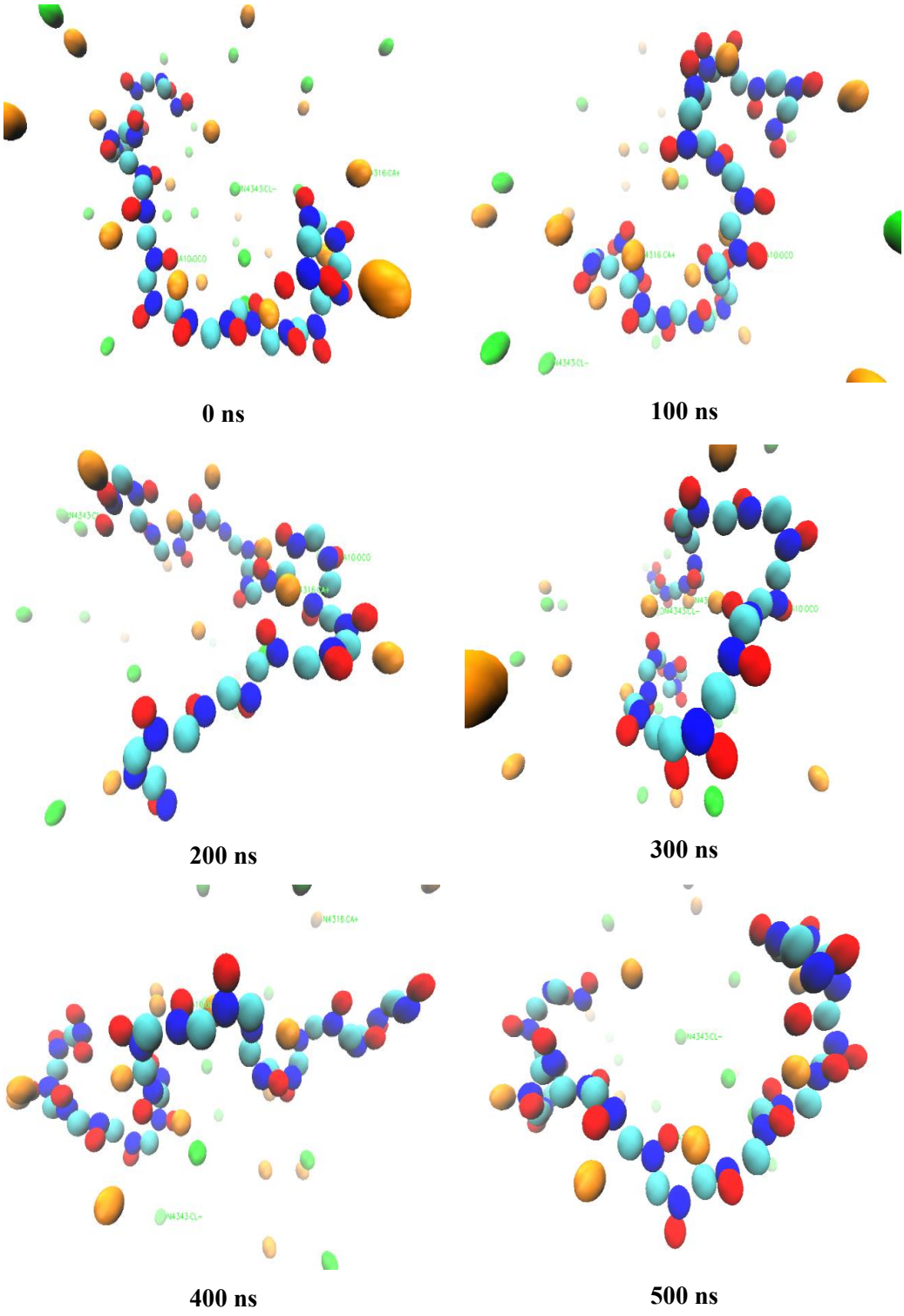


Figure 115 Snapshots taken from MD trajectory of 20-mer MARTINI model of fully deprotonated γ-l-PGA with $CaCl_2$ ions and polarized water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (500 ns). γ-l-PGA is shown as beads with SNa (cyan), SQd/SNd (blue) and SQa (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: calcium ions (orange) and chlorine ions (green).

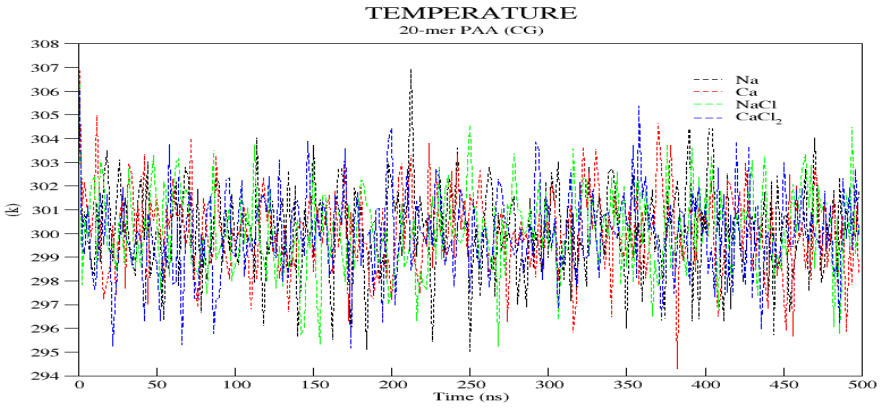
5.6.2 Quality assurance

Fist we check if the system is well equilibrated. As shown in Fig.116a,116b the average temperature and pressure, calculated as a time average from all 3 single independent runs, fluctuate steadily. The average temperature and pressure of all 3 independent production runs is shown in Table 37. Similarly, the average energies (see Fig.116c, 116d, 116e) of our systems are constant (see Table.37). The values are constant over the whole trajectories (500 ns), this indicates a well equilibrated and stable system.

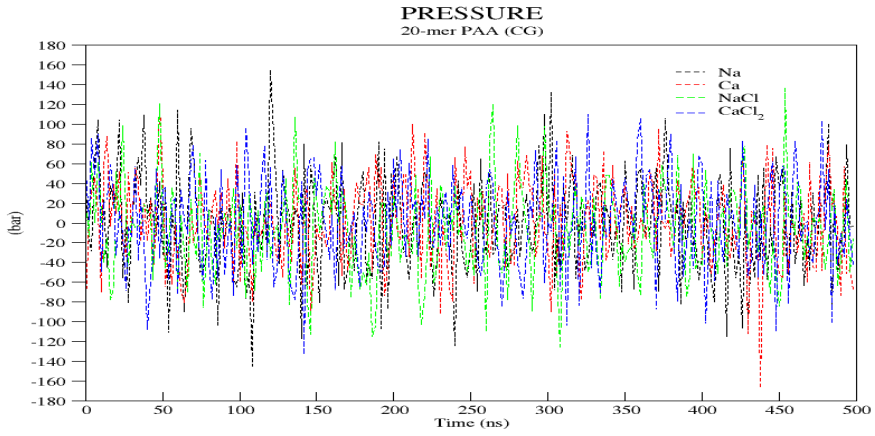
Table 37 System properties calculated as a time average from all 3 single independent runs

Property	***Average value of over 3 independent production runs***			
	Na	***Ca***	***NaCl***	***$CaCl_2$***
Temperature (K)	300.0882	300.1699	300.1015	300.0732
Pressure (bar)	2.4025	4.7417	1.4713	-0.2338
Potential Energy (KJ/mol)	-1.34E+05	-1.36E+05	-1.36E+05	-1.43E+05
Kinetic Energy (KJ/mol)	3.78E+04	3.79E+04	3.77E+04	3.77E+04
Total Energy (KJ/mol)	-9.57E+04	-9.77E+04	-9.83E+04	-1.05E+05

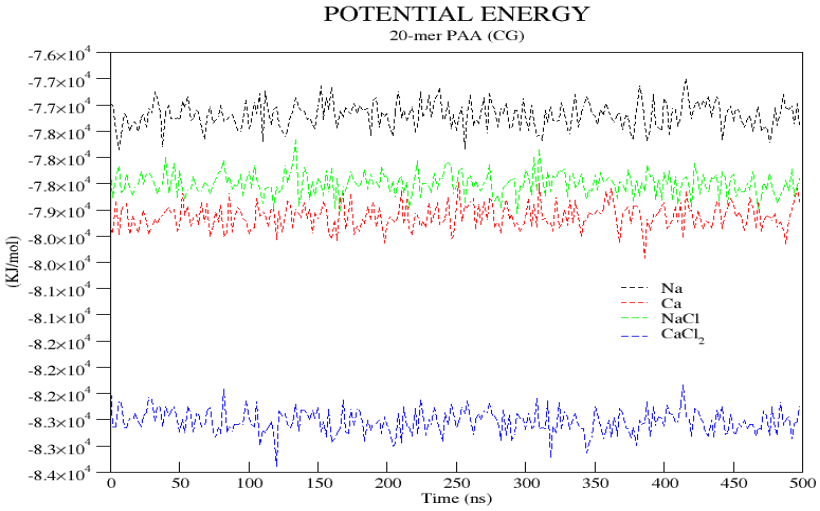
a)



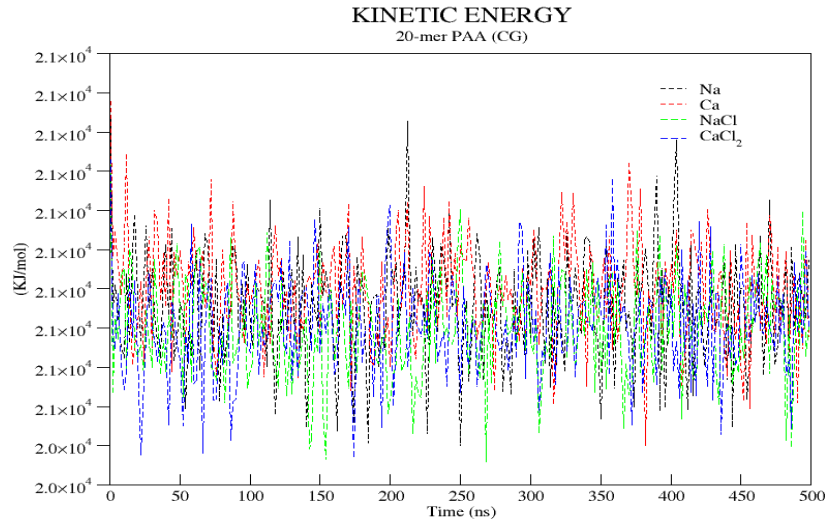
b)



c)



d)



e)

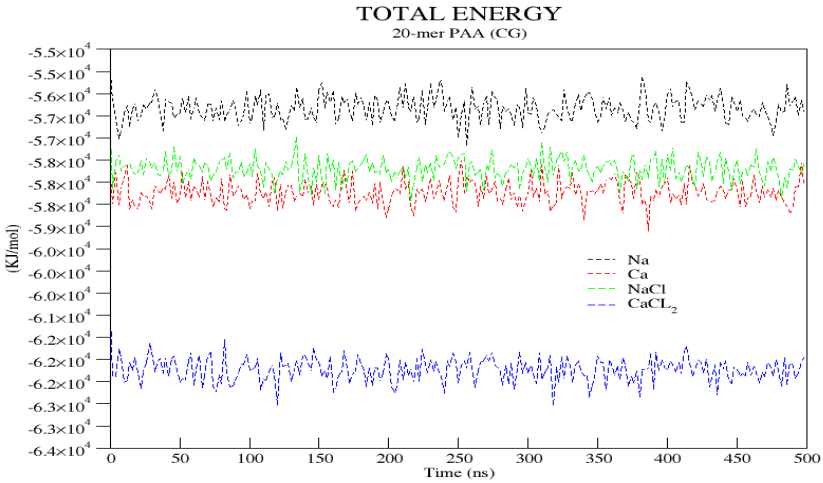


Figure 116 Average energy profiles of 3 independent CG production runs (500 ns) for 20-mer model of fully deprotonated γ-l-PGA with Na, Ca, NaCl, and $CaCl_2$ ions and water

5.6.3 Radius of Gyration

Table 38 Average Radius of gyration values calculated over 3 production runs

System	Average (nm)
Na	2.0650
Ca	1.8793
NaCl	2.0008
$CaCl_2$	1.8318

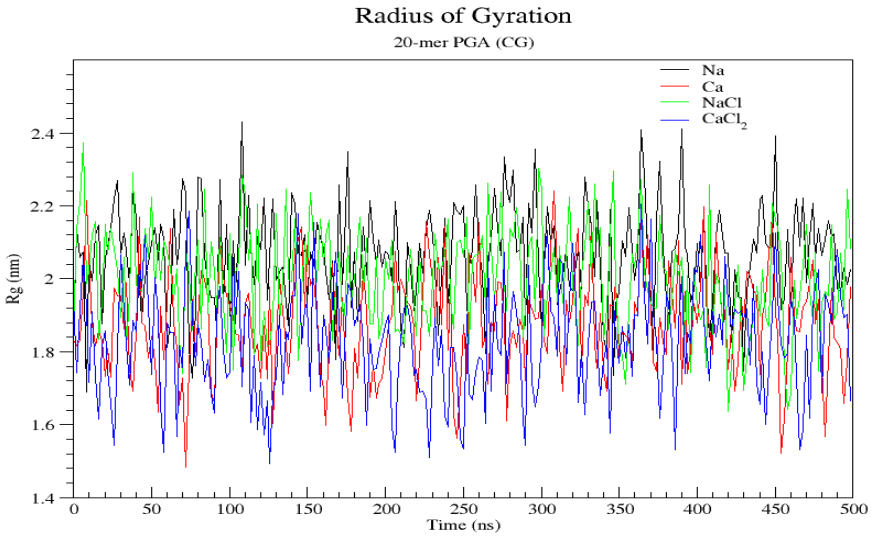


Figure 117 Average change of the radius of gyration over time for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems of γ-l-PGA (fully deprotonated) CG simulation

The average radius of gyration R_g of the fully deprotonated poly-γ-l-glutamic acid CG model in presence of monovalent and divalent ions is shown in Fig 117. The time averaged R_g is presented in Table 38.

For the first 100 ns of simulation the average R_g values of the system containing sodium, calcium, sodium chloride, and calcium chloride ions are 1.0, 0.94, 1.01, 0.93 nm respectively. This is smaller than the R_g value of γ-l-PGA obtained from AA simulation because the chain backbone of CG model is longer (due to three-to-one mapping). Compared with γ-l-PGA AA model, the radius of gyration for the γ-l-PGA MARTINI model exhibits more fluctuations, indicating a more flexible structure.

Radius of gyration for the small-chain γ-L-PGA MARTINI model is within the error limits of the result obtained with all-atomistic simulations as shown in Fig.117. The comparison between the small-chain PAA and γ-l-PGA MARTINI models shows again the influence of the different backbone length, the average radius of gyration for γ-L-PGA is nearly twice as large as for PAA long-chain MARTINI model.

5.6.4 End–to–End distance

Table 39 Average end-to-end values calculated over 3 production runs

System	*Average (nm)*
Na	4.0250
Ca	3.7745
NaCl	3.9680
$CaCl_2$	3.6887

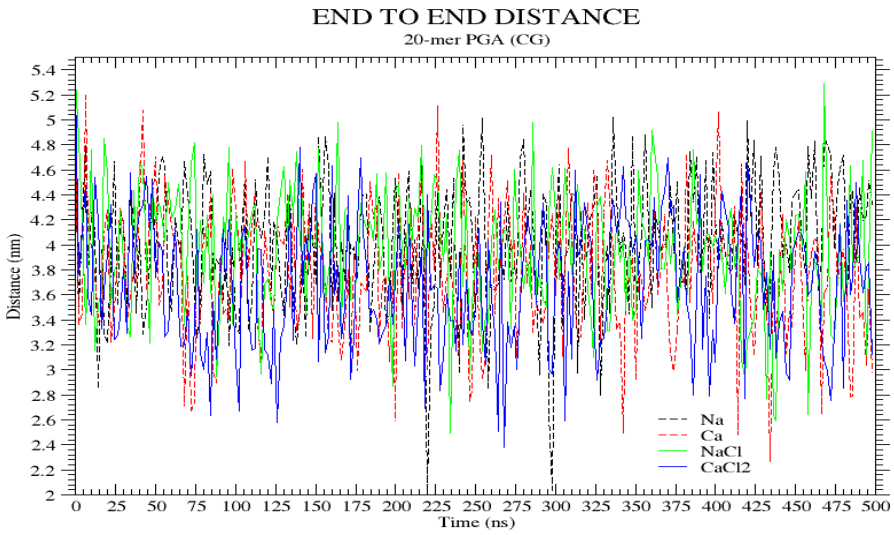


Figure 118 Average curve of the end-to-end distance R over time of 500 ns for all 4 systems over 3 independent production runs of γ-l-PGA (fully deprotonated) CG simulation.

Like the radius of gyration, end-to-end distance for the γ-L-PGA MARTINI model constantly fluctuates during the simulation. The average distance from head to tail group is shown in Table 39.

The average value of R for first 100 ns for the systems containing sodium, calcium, sodium chloride and calcium chloride systems are 3.96, 3.86, 4.04, 3.74 which are not in agreement with the γ-L-PGA all-atomistic simulations (see Fig.95). According to the higher flexibility of the polymer chain, the end-to-end distance for the γ-L-PGA MARTINI model fluctuates strongly. A comparison with all-atomistic simulations reveals a more coiled polymer structure for the MARTINI model (see Fig.95).

5.6.5 Root-mean-square deviation

The backbone RMSD of γ-l-PGA with ions and water is presented in Table 40 (see Fig.119). The average rmsd value for first 100 ns for the systems containing sodium, calcium, sodium chloride and calcium chloride are 0.87, 0.89, 0.86, and 0.91 respectively. These values are lower than the observed values from the γ-l-PGA all atomistic simulations (see Table.34).

Although the results in agreement with the values of R_g and R obtained from the CG simulations of γ-l-PGA.

Table 40 Average RMSD values calculated over 3 independent production runs

System	*Average (nm)*
Na	0.8768
Ca	0.8920
NaCl	0.8795
$CaCl_2$	0.9261

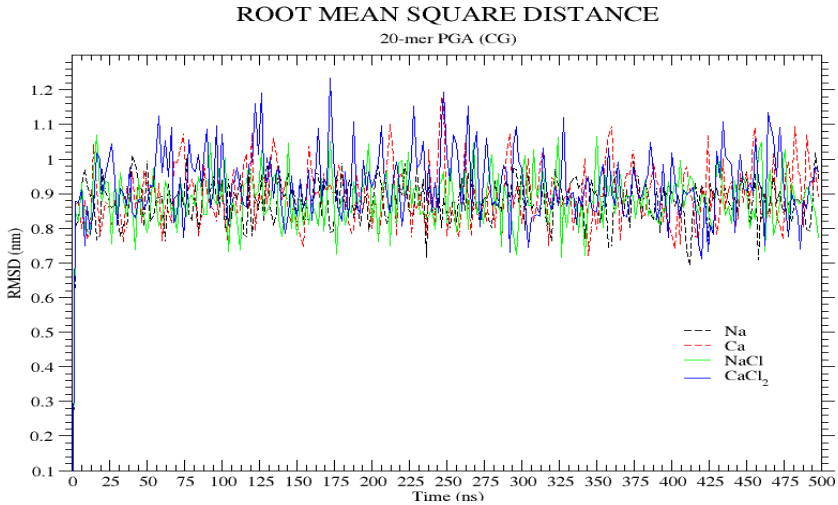


Figure 119 Average curve of the backbone RMSD over time of 500 ns for all 4 systems over 3 independent production runs of γ-l-PGA (fully deprotonated) CG simulation

5.6.6 Solvent accessible surface area

Table 41 Average SASA values calculated over 3 independent production runs

System	Average (nm^2)
Na	30.8153
Ca	30.6438
NaCl	30.7611
$CaCl_2$	30.6177

The average *SASA* calculated for all the salt free and excess salt systems for the fully deprotonated γ-L-PGA MARTINI model is shown in Table 41. No comparison with the literature is available.

The average value of rmsd for first 100 ns of simulation for the system are 30.84, 30.64, 30.80, 30.65 nm but the result is in reasonable agreement with the results from all-atomistic simulations. Differences between the MARTINI and all-atomistic simulations are expected due to the different probe radius.

Compared with the results from all-atomistic simulations, the deviation of solvent accessible surface area grows further for the short-chain MARTINI models of γ-L-PGA.

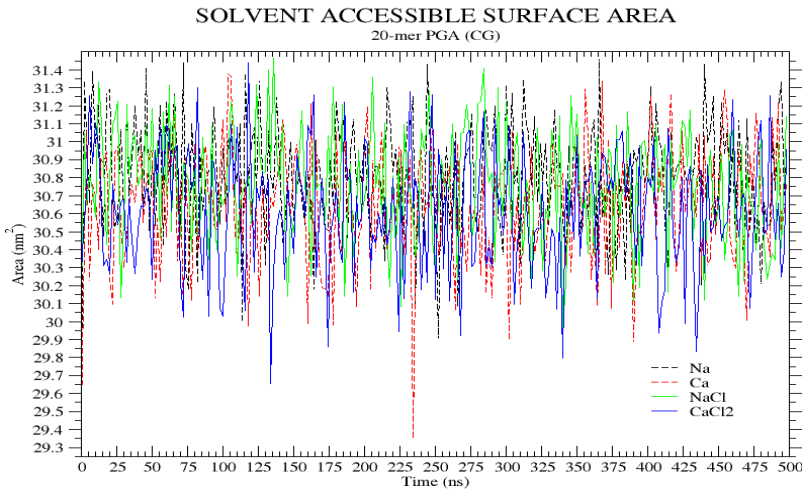


Figure 120 Average curve of the SASA over time of 500 ns for all 4 systems over 3 independent production runs of γ-l-PGA (fully deprotonated) CG simulation

5.6.7 Radial Distribution Functions

5.6.7.1 RDF between the centre of mass of the γ-l-PGA beads and the polarized water

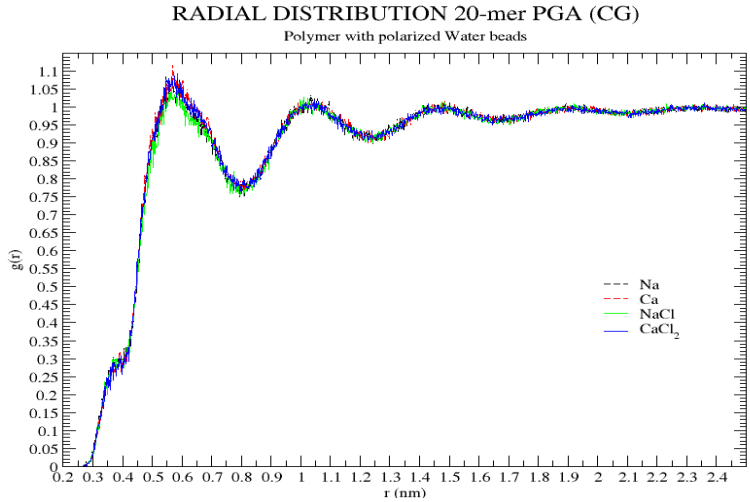


Figure 121 Average RDF between γ-l-PGA (centre of mass) and polarized water for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

The sharp peak at 0.570 ± 0.002 nm in indicates a well-structured solvation shell with ordered distribution of polarized water beads around the polymer chain. This value is bigger than the value obtained in AA simulation. This result can be corrected by adjusting the dielectric constant for each individual simulation.

5.6.7.2 RDF between the centre of mass of the γ-l-PGA beads and the neutral water bead

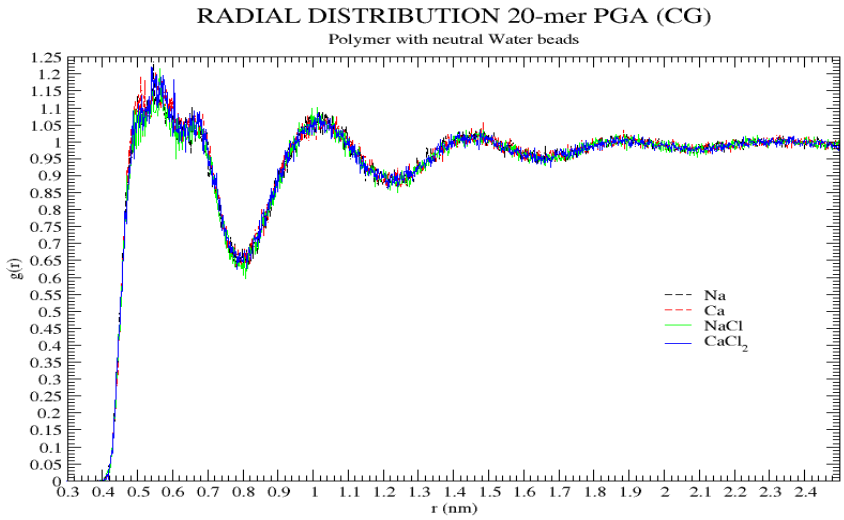


Figure 122 Average RDF between γ-l-PGA (centre of mass) and neutral water bead (W) for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

The RDF between centre of mass of γ-l-PGA and neutral water beads of polarized water shows us a major peak at 0.552 ± 0.002 nm shown in Fig.122.

5.6.7.3 RDF between SQa beads and the polarized water

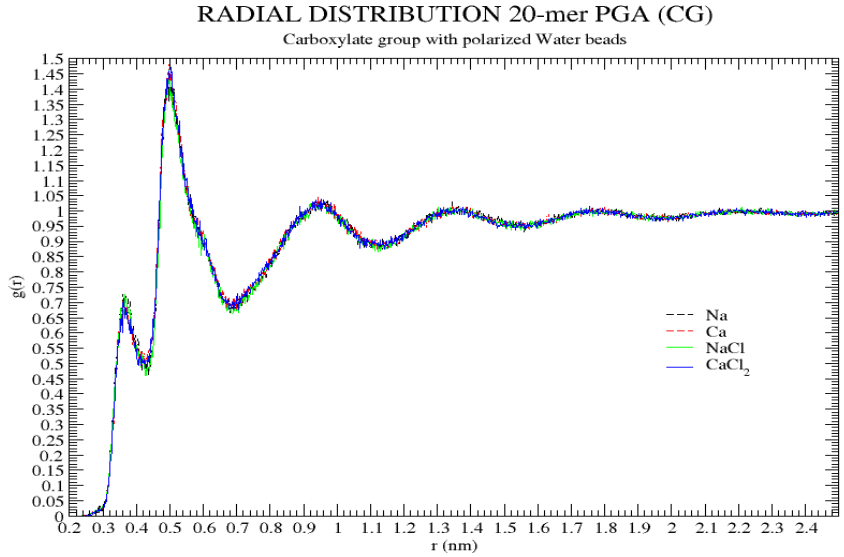


Figure 123 Average RDF between carboxylate group (SQa) and polarized water for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

The RDF between the SQa beads (carboxylate group) with respect to polarized water shows a strong peak at 0.496 ± 0.002 nm (see Fig.123). A shoulder peak can be observed before the major peak at 0.358 ± 0.002 nm follows.

5.6.7.4 RDF between SNa beads and the polarized water

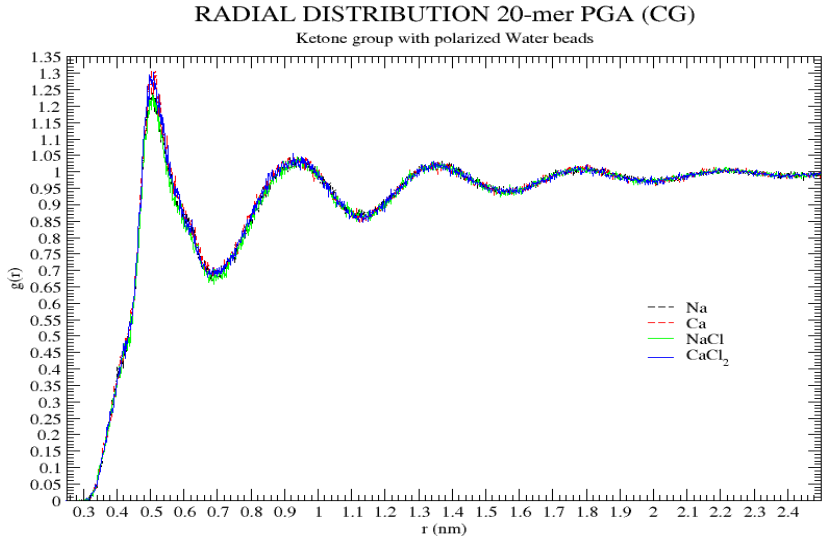


Figure 124 Average RDF between carbonyl group (SNa) and polarized water for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

The RDF between the SNa beads (carbonyl group) with respect to polarized water shows a strong peak at 0.500 ± 0.002 nm (see Fig.124). A shoulder peak can be observed following the major peak at 0.944 ± 0.002 nm.

5.6.7.5 RDF between SQd beads and polarized water

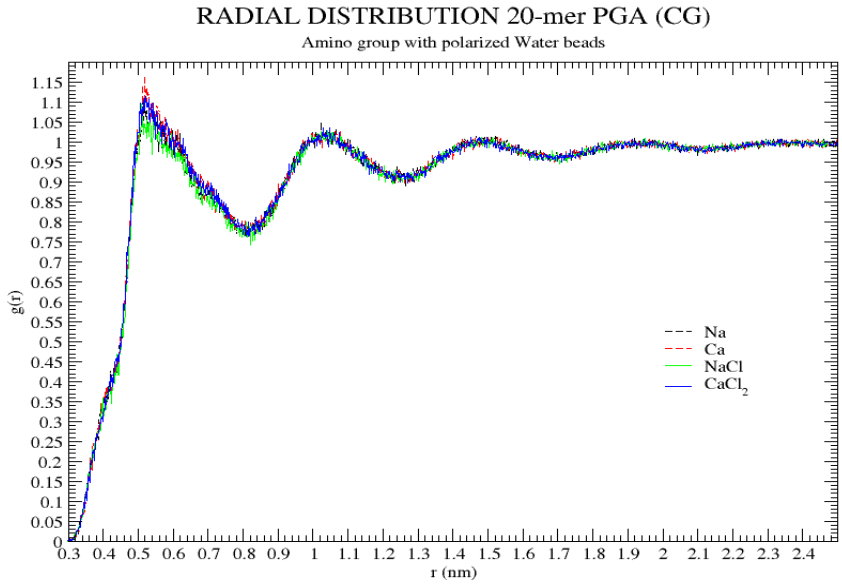


Figure 125 Average RDF between amino group (SQd) and polarized water for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

The RDF between the SQd beads (amino group) with respect to polarized water shows a strong peak at 0.520 ± 0.002 nm (see Fig.125). A shoulder peak can be observed following the major peak at 1.028 ± 0.002 nm.

5.6.7.6 RDF between SQa beads and 'W' of polarized water

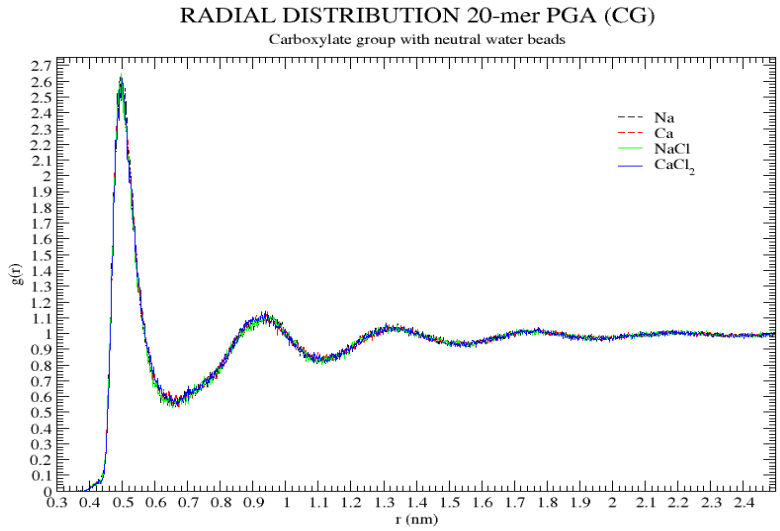


Figure 126 Average RDF between carboxylate group (SQa) and neutral water bead 'W' of polarized water for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

The RDF between the SQa beads (carboxylate group) with respect to neutral water beads of polarized water shows a strong peak at 0.496 ± 0.002 nm (see Fig.126). A shoulder peak at 0.934 ± 0.002 nm follows.

5.6.7.7 RDF between SNa beads and "W' of polarized water

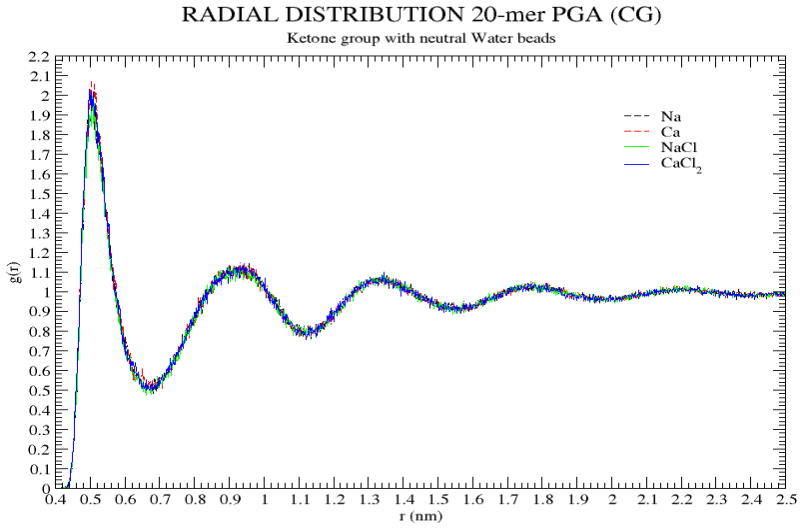


Figure 127 Average RDF between carbonyl group (SNa) and neutral bead (W) of polarized water for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

The RDF between the SNa beads (carbonyl group) with respect to neutral water beads of polarized water shows a strong peak at 0.502 ± 0.002 nm (see Fig.127). A shoulder peak at 0.944 ± 0.002 nm follows.

5.6.7.8 RDF between SQd beads and 'W' of polarized water

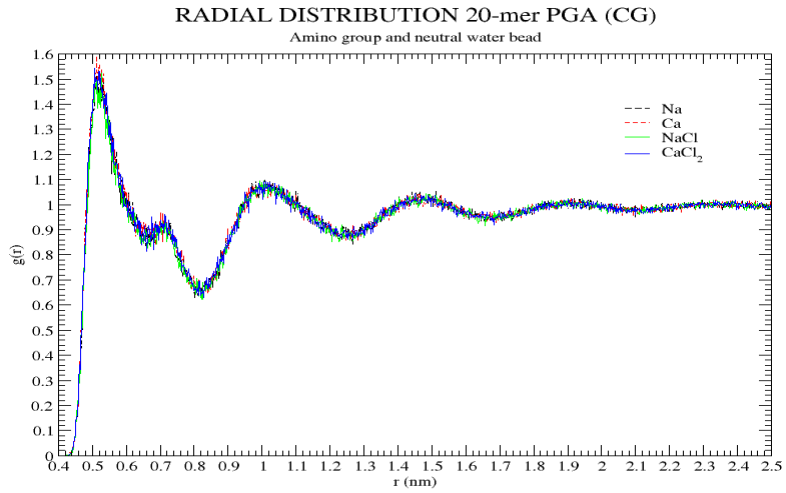


Figure 128 Average RDF between amino group (SQd) and neutral water bead 'W' of polarized water for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

The RDF between the SQd beads (amino group) with respect to neutral water beads of polarized water shows a strong peak at 0.512 ± 0.002 nm (see Fig.128). A shoulder peak at 0.716 ± 0.002 nm follows.

5.6.7.9 RDF between SQa beads and Ions

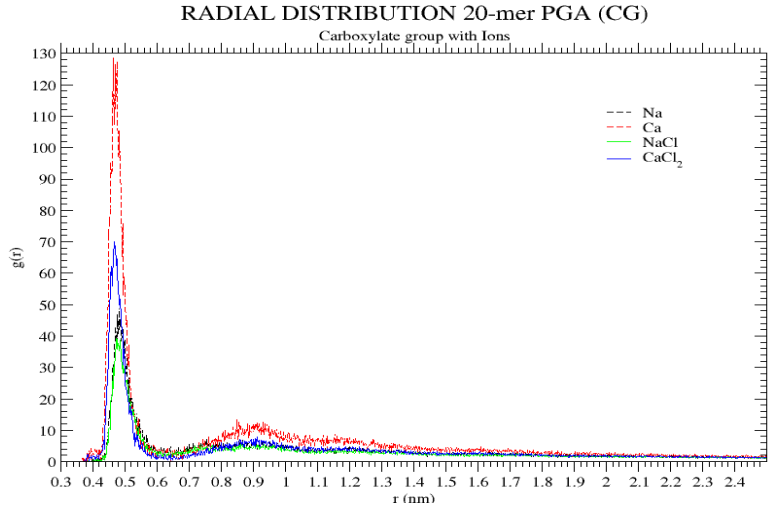


Figure 129 Average RDF between carboxylate group (SQa) and ions for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

As shown by Fig.129, the RDF between the SQa (carboxylate group) and monovalent and divalent counterions exhibits a strong first peak at 0.464 ± 0.002 nm.

5.6.7.10 RDF between SNa beads and Ions

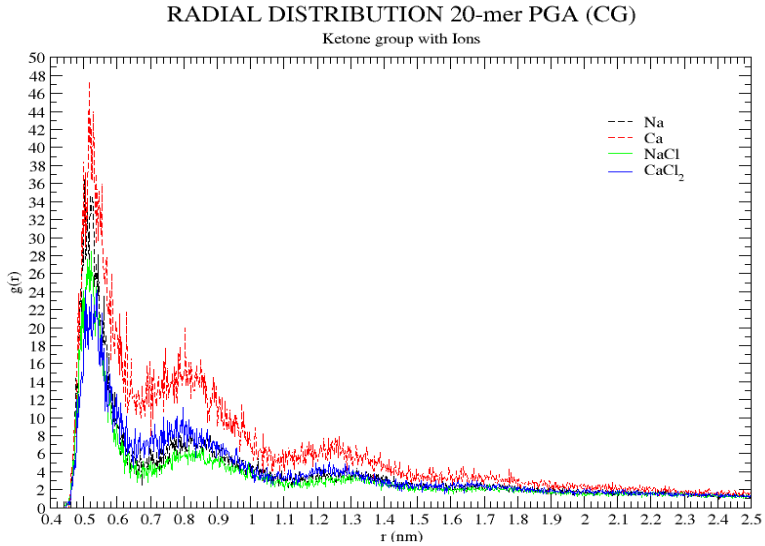


Figure 130 Average RDF between carbonyl group (SNa) and ions for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

As shown by Fig.130, the RDF between the SNa (carbonyl group) and monovalent and divalent counterions exhibits a strong first peak at 0.474 ± 0.002 nm, suggesting a preferred condensation of calcium ions on to the chain backbone of γ-l-PGA.

5.6.7.11 RDF between SQd beads and Ions

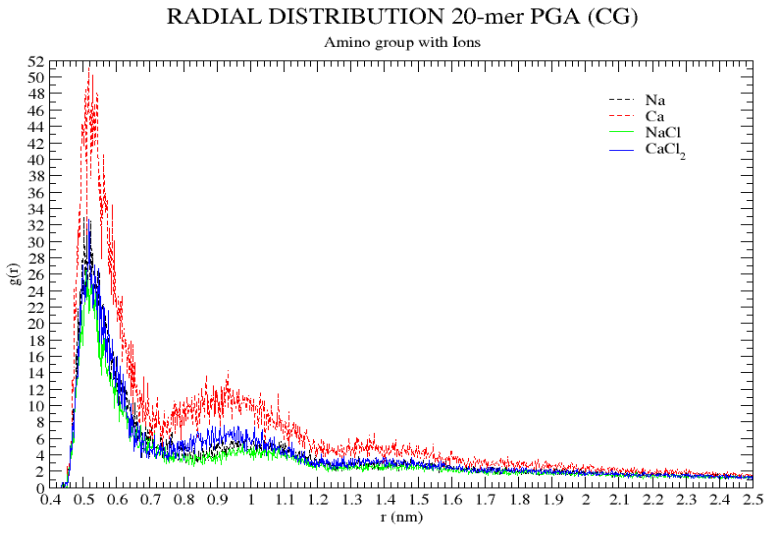


Figure 131 Average RDF between amino group (SQd) and ions of polarized water for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

As shown by Fig.131, the RDF between the SQd (amino group) and monovalent and divalent counterions exhibits a strong first peak at 0.516 ± 0.002 nm, suggesting a preferred condensation of calcium ions on to the amino group.

5.7 Analysis of trajectories from All-Atomistic MD simulations of fully deprotonated 62-mer model of gamma-L-polyglutamic acid (γ-L-PGA)

Table 42 Number of ions in Na & Ca (salt free) and NaCl & Cacl2 (excess salt) systems of 62-mer γ-l-PGA

Salt	*Number of ions*	
	Na/Ca	*Cl*
Na	*62*	-
NaCl	*95*	*33*
Ca	*31*	-
$CaCl_2$	*64*	*66*

Simulations under process

5.8 Analysis of trajectories from Coarse Grained MD simulations of fully deprotonated 62-mer model of gamma-L-polyglutamic acid (γ-L-PGA)

γ-l-PGA (fully deprotonated) MARTINI 62-mer model is simulated in a system containing polarizable water model from MARTINI and 4 different ions individually (see Table 42). Polymer model was generated using the mapping scheme described in section 3.2.2 using MARTINI scheme for modelling of polymer.

The four generated boxes each include one γ-l-PGA long-chain model. Periodic boundary conditions were applied in all directions. For γ-l-PGA simulations, cubic boxes with 18 nm edge length were used for the short-chain models. The new refined polarizable water model from MARTINI was used to solvate the system. In order to neutralize the system charge of the polymer, water molecules were replaced by ions according to the number of monomers in the polymer chain (see Table 42). In order to account for the ionic strength of the synthetic tap water (15°dH) used in experiments additional water molecules were exchanged by sodium and chloride ions. The setup for each simulated system is summarized in Fig. 42.

All simulations were performed on the RWTH Compute Cluster using GROMACS 2019 simulation package[116]. All systems were first energy minimized using steepest descent algorithm with 500,000 integration steps or until the maximum force on any atom in the system did not exceed a value of 10.0 kJ/mol/nm. For neighbor searching, Verlet cutoff-scheme was used having short range Van der Waals cut-off of 1.0 with periodic boundary conditions. For electrostatic forces, Particle-mesh Ewald method of order 8 was used having short range cut-off of 1.2 keeping relative dielectric constant 2.5 and relative dielectric constant of reaction field equal to 1. For Van der waals forces, twin range cut-offs with neighbor list cut-off and VdW cut-off are used. For the treatment of long-range electrostatic interaction and long-range dispersion corrections for energy and pressure was applied. Grid dimensions are controlled with flourier spacing of 0.15. The relative strength of the Ewald-shifted direct potential at rcoulomb is given by 'ewald-rtol'. For doing PME for VdW-interactions, ewald-rtol-lj is used to control the relative strength of the dispersion potential at rvdw. Both the values are kept default 1e-5 and 1e-3 respectively.

After heating, the systems were equilibrated for 50 ns each at 300 K using the velocity rescale thermostat (coupling time 1 ps) according to Bussi *et al.*[117] in NVT ensemble and at

1 bar using Berendsen barostat[118, 119] (coupling time 12 ps and compressibility $3 \cdot 10^{-4}$ bar^{-1}) in NPT ensemble using velocity verlet integrator. During the equilibration, position restraints on every polymer bead were activated. Three independent production runs, 500 ns each, were performed for every system in NPT ensemble using Parrinello-Rahman extended-ensemble pressure coupling and Verlet cutoff scheme with a cutoff distance of $r_{cut} = 1.2$ nm for the Lennard-Jones interactions according to a recent publication[120]. The Particle Mesh Ewald (PME) method[112] with a grid spacing of 0.15 nm in agreement with the parameters employed during the parameterization of the refined polarizable water model[105]. Consistently, a time step of 20 fs was used. The neighbor list was updated every 10 steps using the Verlet neighbor search (VNS) algorithm[120]. Bond lengths were held constant by the LINCS algorithm[113, 121] and the background permittivity was set to $\varepsilon_{bg} = 2.5$ as recommended when using the polarizable water model[106, 120]. The final box sizes are shown in Table.4.

A chain length of 62 monomer of γ-l-PGA (fully deprotonated) is simulated in a system containing water and different quantity of monovalent and divalent ions (sodium, calcium, sodium chloride and calcium chloride) (see Table 42). The system is equilibrated at 300k and 1 bar of pressure. The system is simulated for 500 ns with a time step of 20 femtoseconds. System configuration for production run is provided in appendix 9.15-9.18. The simulation results can be only validated through AA simulations which are currently under simulation. Hence, the results of these simulations could not verify. They do provide an insight into the functioning and capability of MARTINI for simulating long chain γ-l-PGA. The validity of the results will be done in further work.

5.8.1 Visualization of gamma-l-polyglutamic acid (γ-l-PGA) interaction with Na, Ca, NaCl & $CaCl_2$ using VMD

Some representative snapshots of the MD simulations of the γ-l-PGA long chain MARTINI models are shown in Fig.132-135 with different ions. The snapshots were taken after 500 ns of MD simulation. The water molecules have not been presented in the images for better visualisation of the polymer chain and its interaction with ions. The γ-l-PGA chain is shown as stick model. Snapshots from each system can be seen on the following pages.

In Fig. 133, Inter chain bridging can be clearly visualized in the snapshot taken at 500ns. Polymer bridging flocculation is a widespread but poorly understood phenomenon, with little quantitative data available. Its strength and mechanism can be investigated by neutron diffraction and X-ray diffraction experiments. For strong polyelectrolytes, the polymer charge is sufficient to firmly attract some counterions to stay very near the charge sites, as can indeed be reflected by the high counterion concentration predicted by the Poisson−Boltzmann law. It is assumed that the condensed counterions are much less mobile than the free ions due to stronger attraction with the monomer charge, thereby leading to a decreased effective charge density of the chain. If multivalent counterions are added, they can replace the condensed monovalent counterions[126] and change the sign of the local charge near the site, which will attract the site without condensed ions. A bridging effect can therefore take place, which can further contract the macromolecules. The bridging concept has been used to account for the phase separation of polyelectrolytes with the addition of multivalent counterions.[127]

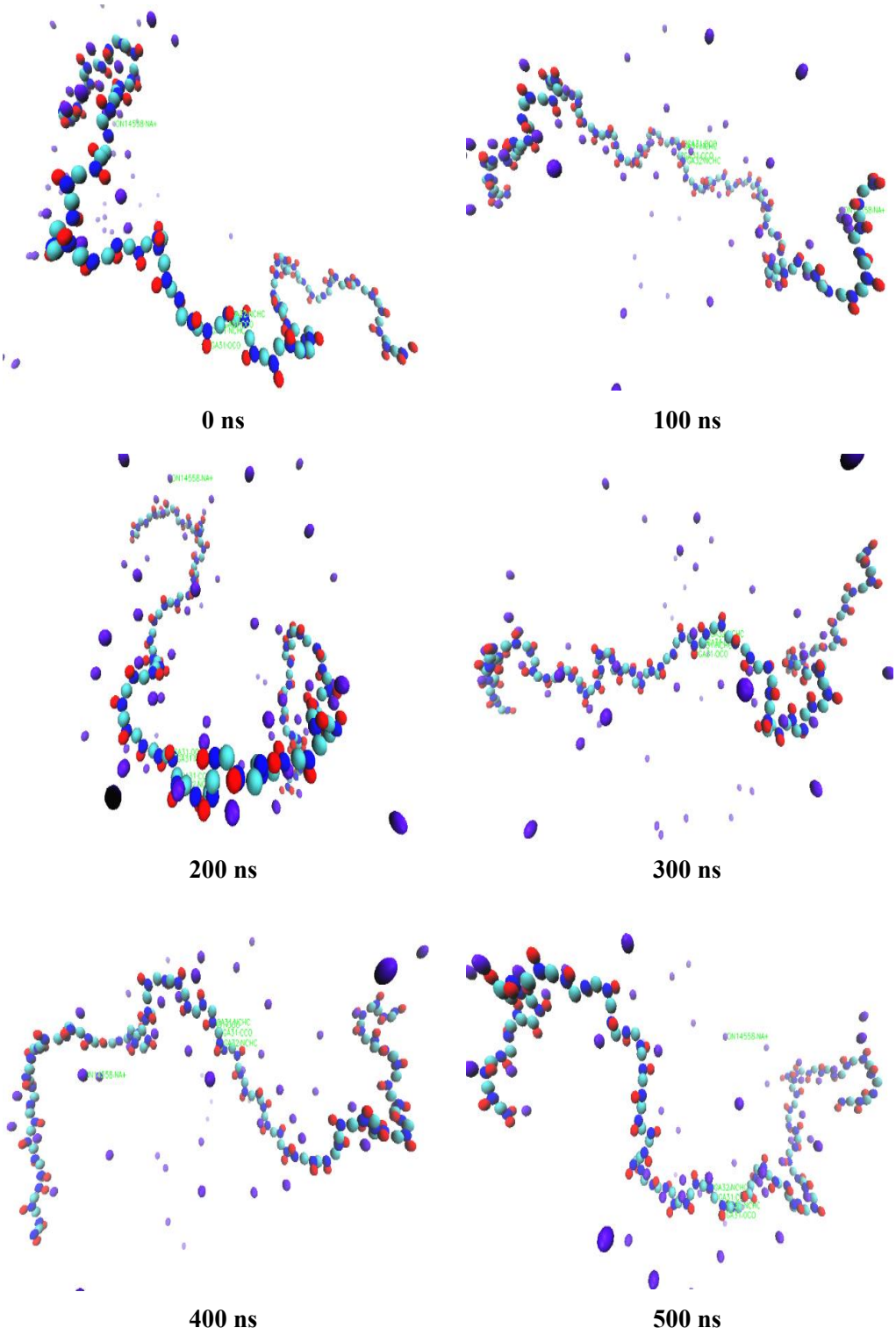


Figure 132 Snapshots taken from MD trajectory of 62-mer MARTINI model of fully deprotonated γ-l-PGA with Na ions and polarized water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (500 ns). γ-l-PGA is shown as beads with SNa (cyan), SQd/SNd (blue) and SQa (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: sodium ions (violet).

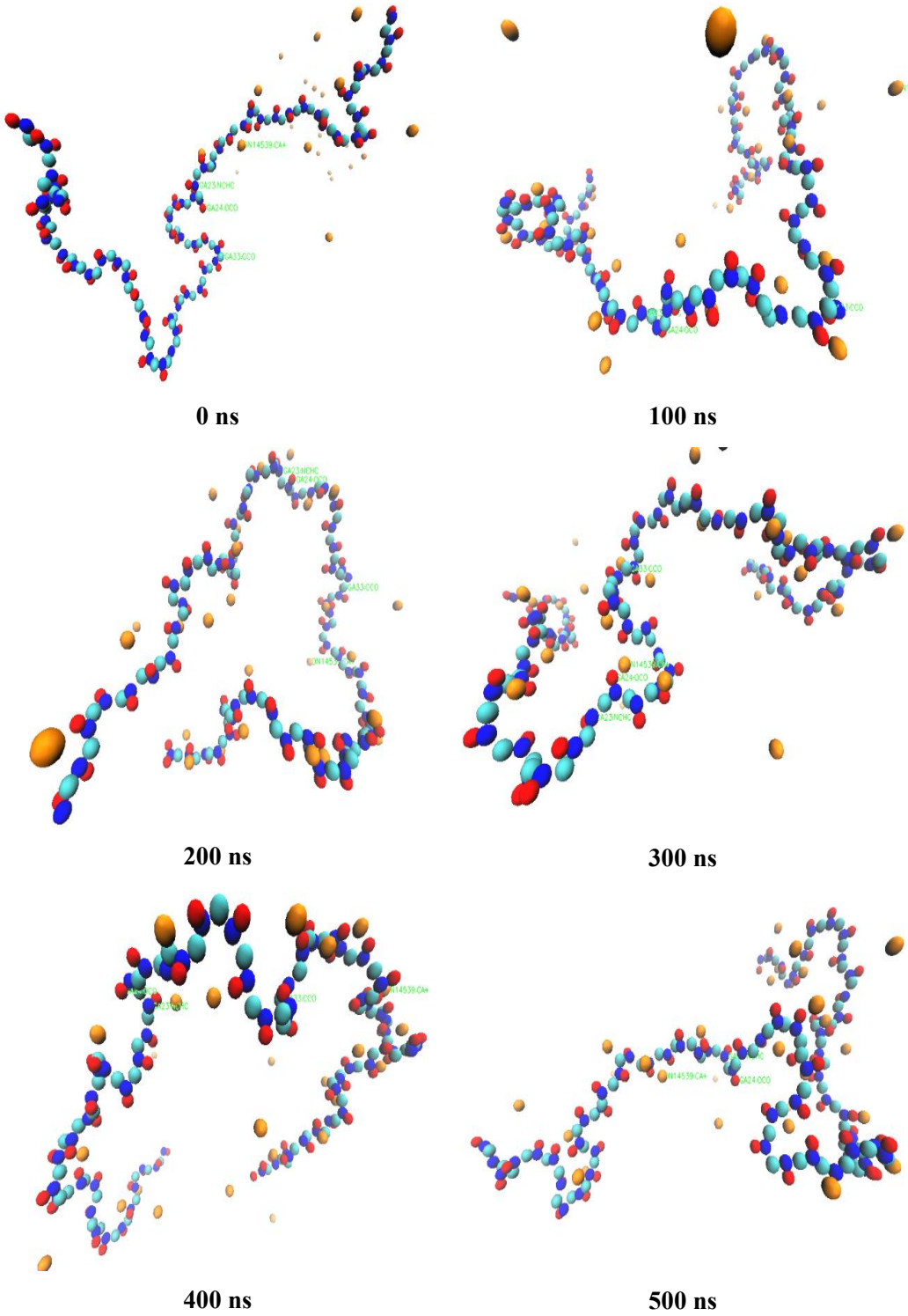


Figure 133 Snapshots taken from MD trajectory of 62-mer MARTINI model of fully deprotonated γ-l-PGA with Ca2+ ions and polarized water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (500 ns). γ-l-PGA is shown as beads with SNa (cyan), SQd/SNd (blue) and SQa (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: calcium ions (orange)

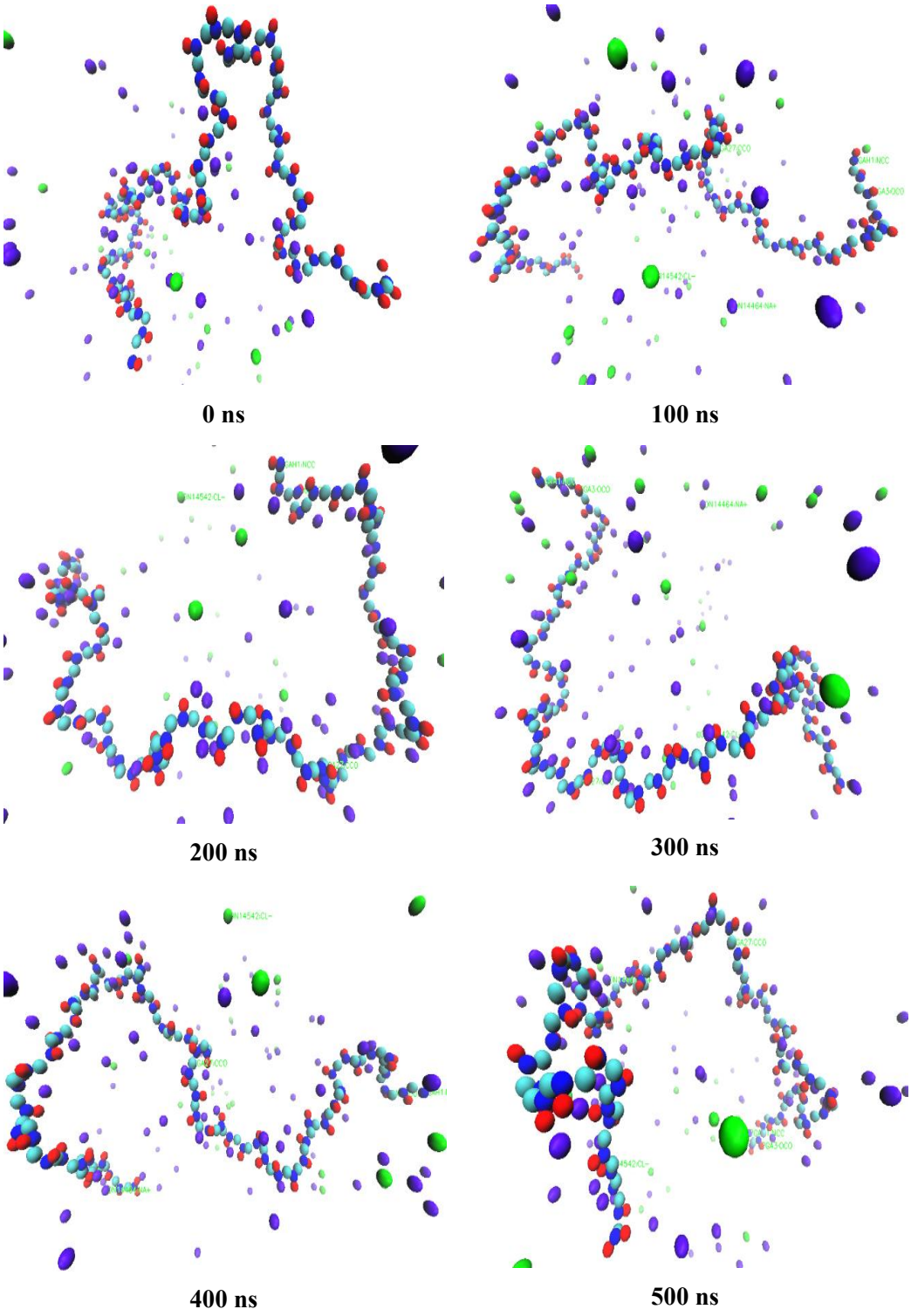


Figure 134 Snapshots taken from MD trajectory of 62-mer MARTINI model of fully deprotonated γ-l-PGA with NaCl ions and polarized water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (500 ns). γ-l-PGA is shown as beads with SNa (cyan), SQd/SNd (blue) and SQa (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: sodium ions (violet) and chlorine ions (green).

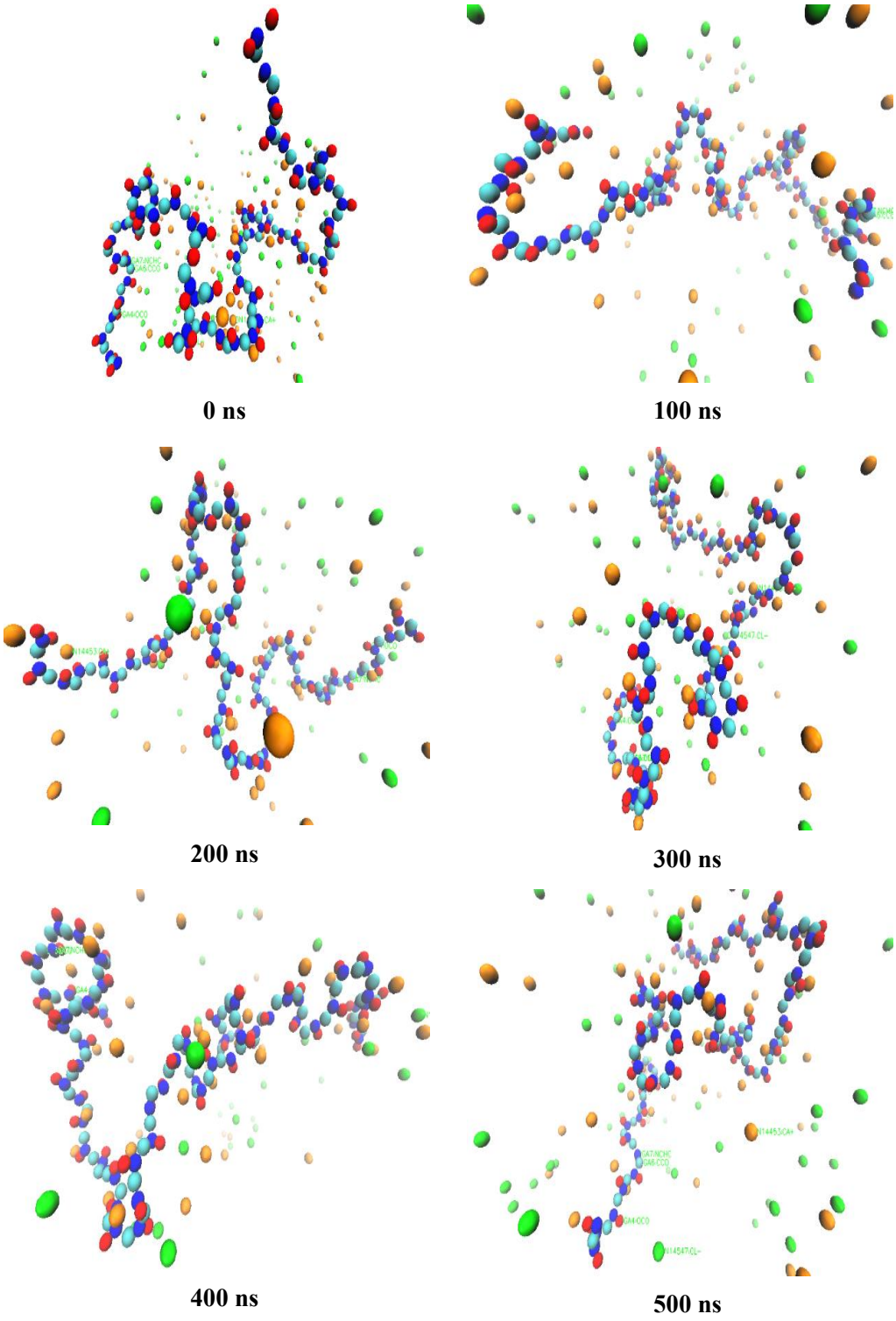


Figure 135 Snapshots taken from MD trajectory of 62-mer MARTINI model of fully deprotonated γ-l-PGA with $CaCl_2$ ions and polarized water (not shown here) for run_1 from initial structure (0 ns) to last snapshot (500 ns). γ-l-PGA is shown as beads with SNa (cyan), SQd/SNd (blue) and SQa (red). Aliphatic hydrogens are not shown for clarity. Ions are represented by balls: calcium ions (orange) and chlorine ions (green).

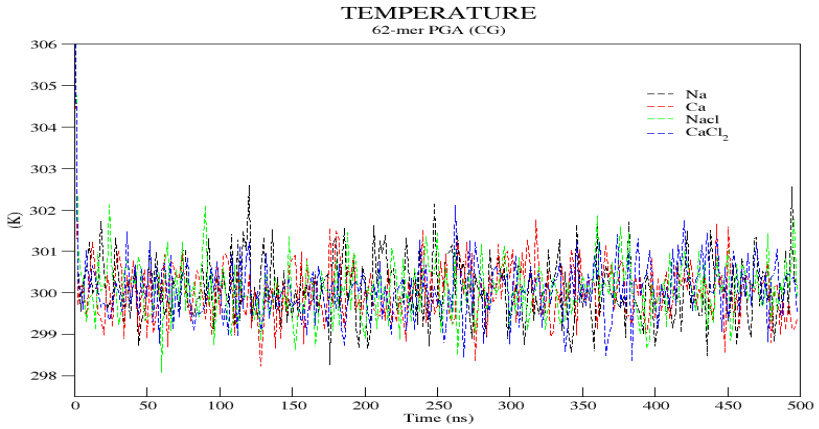
5.8.2 Quality assurance

Fist we have to check if the system is well equilibrated. As shown in Fig 136a,136b the average temperature and pressure, calculated as a time average from all 3 single independent runs, fluctuate steadily. The average temperature and pressure of all 3 independent production runs is shown in Table 43. Similarly, the average energies (see Fig. 136c, 136d, 136e) of our systems are constant (see Table 43). The values are constant over the whole trajectories (500 ns), this indicates a well equilibrated.

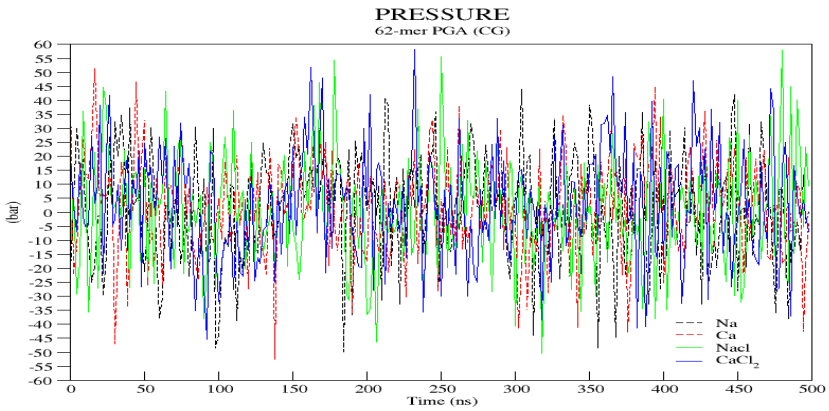
Table 43 System properties calculated as a time average from all 3 single independent runs

Property	***Average value of over 3 independent production runs***			
	Na	***Ca***	***NaCl***	***$CaCl_2$***
Temperature (K)	300.1240	300.0627	300.0875	300.0959
Pressure (bar)	1.3413	2.2910	0.4168	3.0031
Potential Energy (KJ/mol)	-4.46E+05	-4.53E+05	-4.53E+05	-4.74E+05
Kinetic Energy (KJ/mol)	1.27E+05	1.27E+05	1.27E+05	1.27E+05
Total Energy (KJ/mol)	-3.19E+05	-3.25E+05	-3.27E+05	-3.47E+05

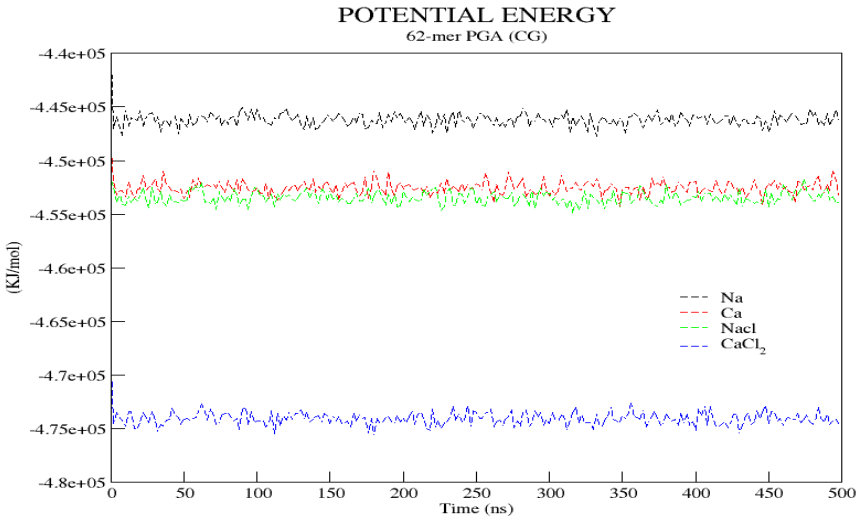
a)



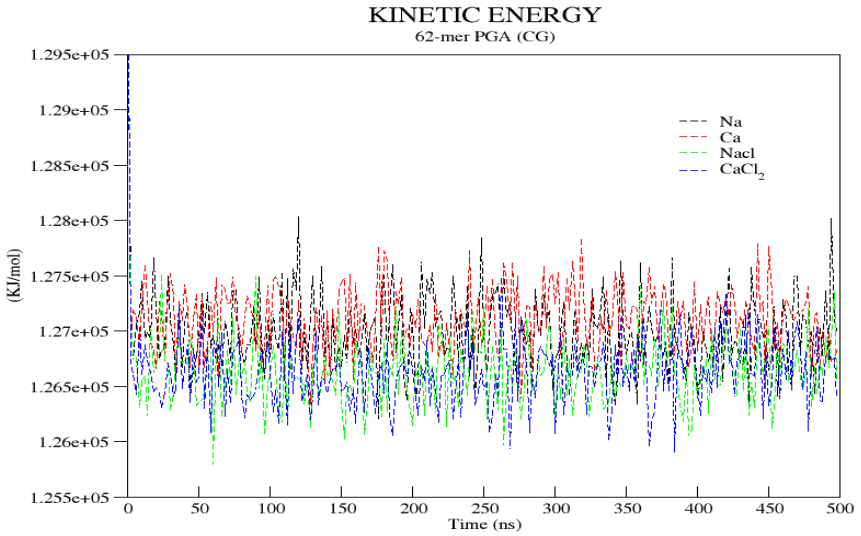
b)



c)



d)



e)

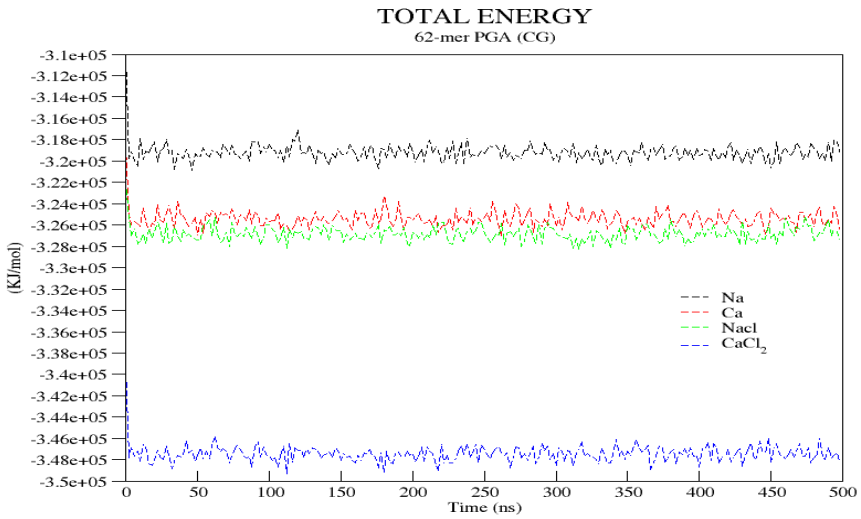


Figure 136 Average energy profiles of 3 independent CG production runs (500 ns) for 62-mer model of fully deprotonated γ-l-PGA with Na, Ca, NaCl, and $CaCl_2$ ions and water

5.8.3 Radius of Gyration

Table 44 Average change of the radius of gyration over time for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems of γ-l-PGA (fully deprotonated) CG simulation

System	Average (nm)
Na	5.0579
Ca	4.1421
NaCl	4.8016
$CaCl_2$	3.9483

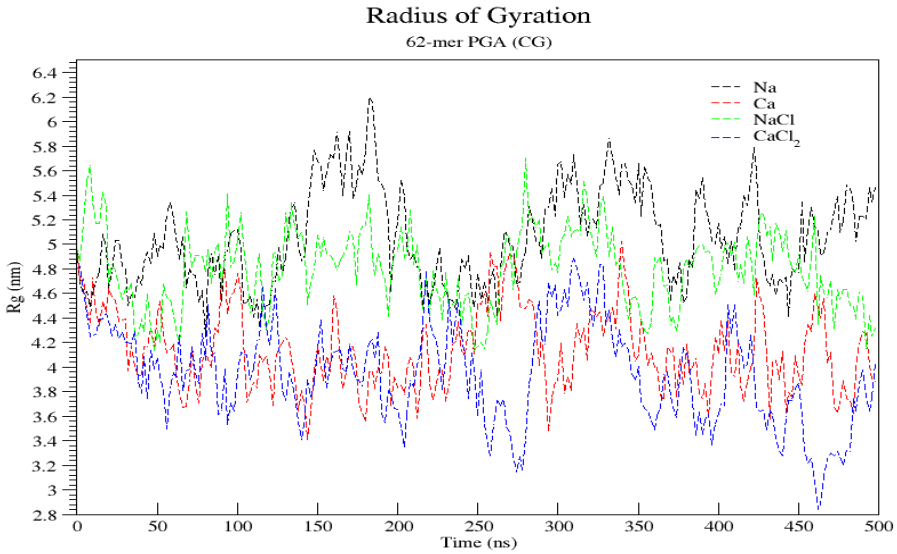


Figure 137 Average change of the radius of gyration over time for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems of γ-l-PGA (fully deprotonated) CG simulation

The average radius of gyration R_g of the fully deprotonated 62-mer poly-γ-l-glutamic acid CG model in presence of monovalent and divalent ions is shown in Fig. 137. The time averaged R_g is presented in Table 44. For the first 100 ns of simulation the average R_g values of the system containing sodium, calcium, sodium chloride, and calcium chloride ions are 4.83, 4.29, 4.78, 4.09 nm respectively. Due to unavailability of all atomistic results no comparison cannot be drawn, and the validity of the results cannot be checked.

5.8.4 End–to–End distance

Like the radius of gyration, the end-to-end distance for the 62-mer γ-L-PGA MARTINI model is fluctuates during the simulation. The average distance from head to tail group is shown in Table 45. The average value of R for first 100 ns for the systems containing sodium, calcium, sodium chloride and calcium chloride systems are 5.80, 5.59, 5.94, 5.39 nm. Due to unavailability of all atomistic results no comparison can't be drawn, and the validity of the results cannot be checked.

Table 45 Average end-to-end values calculated over 3 production runs

System	Average (nm)
Na	5.7326
Ca	5.8674
NaCl	5.8798
$CaCl_2$	5.5760

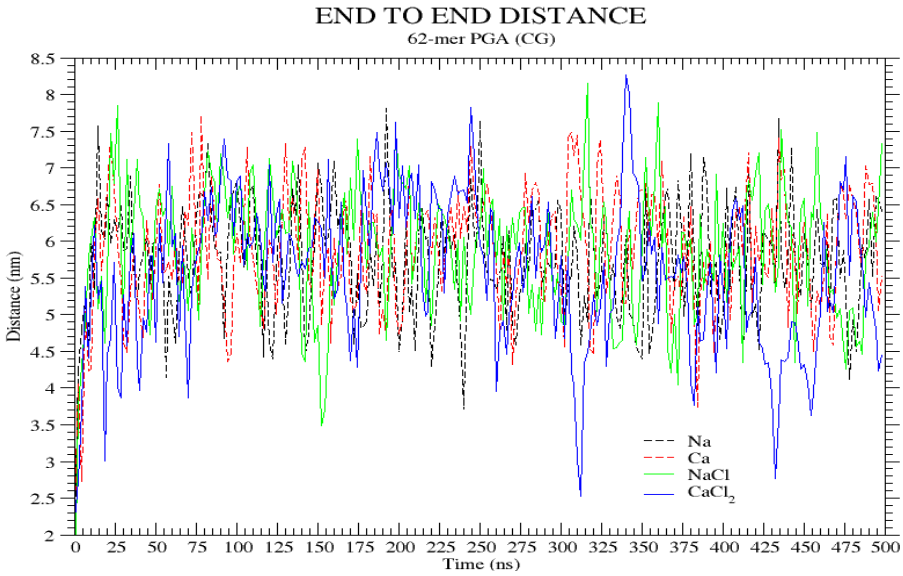


Figure 138 Average curve of the end-to-end distance R over time of 500 ns for all 4 systems over 3 independent production runs of γ-l-PGA (fully deprotonated) CG simulation

5.8.5 Root-mean-square deviation

Table 46 Average RMSD values calculated over 3 independent production runs

System	Average (nm)
Na	2.3883
Ca	2.6136
NaCl	2.3564
$CaCl_2$	2.4847

The backbone RMSD of γ-l-PGA with ions and water is presented in Table 46 (see Fig. 139). The average rmsd value for first 100 ns for the systems containing sodium, calcium, sodium chloride and calcium chloride are 2.38, 2.48, 2.27, 2.31 nm respectively. The results in agreement with the values of R_g and R obtained from the CG simulations of γ-l-PGA.

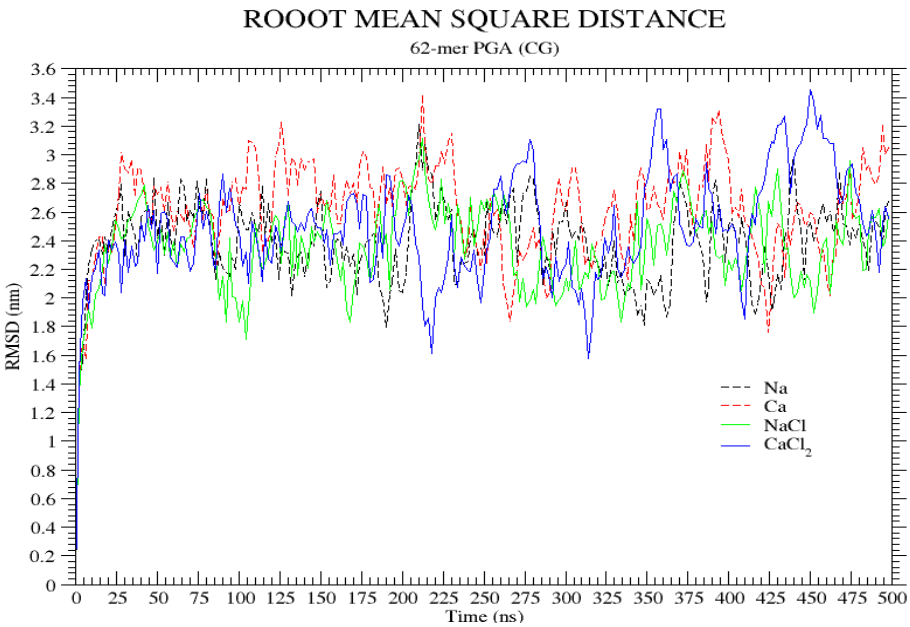


Figure 139 Average curve of the backbone RMSD over time of 500 ns for all 4 systems over 3 independent production runs of γ-l-PGA (fully deprotonated) CG simulation

5.8.6 Solvent accessible surface area

The average *SASA* calculated for all the salt free and excess salt systems for the fully deprotonated γ-L-PGA MARTINI model is shown in Table 47. No comparison with the literature is available. The average value of rmsd for first 100 ns of simulation for the system are 94.44, 93.95, 94.36, 93.7 nm^2

Table 47 Average SASA values calculated over 3 independent production runs

System	Average (nm^2)
Na	94.4641
Ca	93.9732
NaCl	94.4124
$CaCl_2$	93.7677

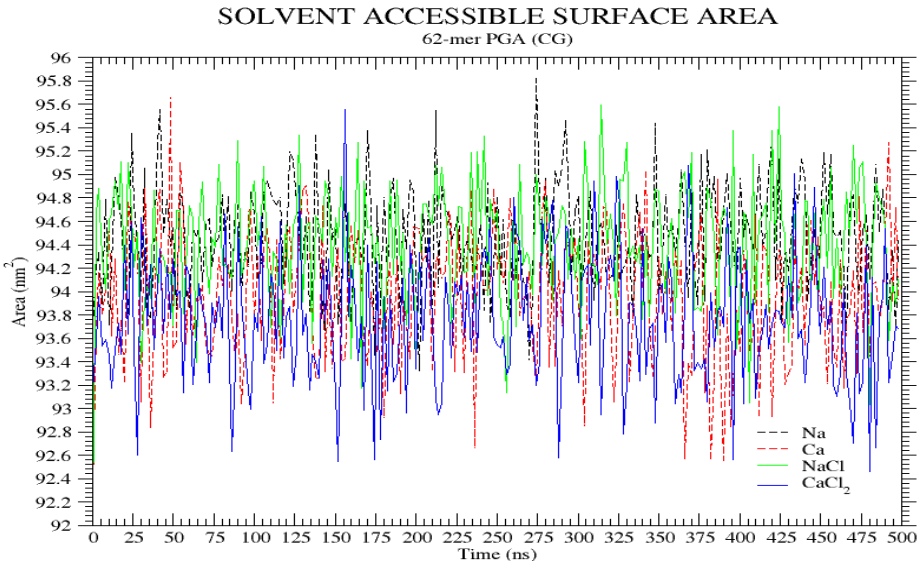


Figure 140 Average curve of the SASA over time of 500 ns for all 4 systems over 3 independent production runs of γ-l-PGA (fully deprotonated) CG simulation

5.8.7 Radial Distribution Functions

5.8.7.1 RDF between the centre of mass of the γ-l-PGA beads and the polarized water

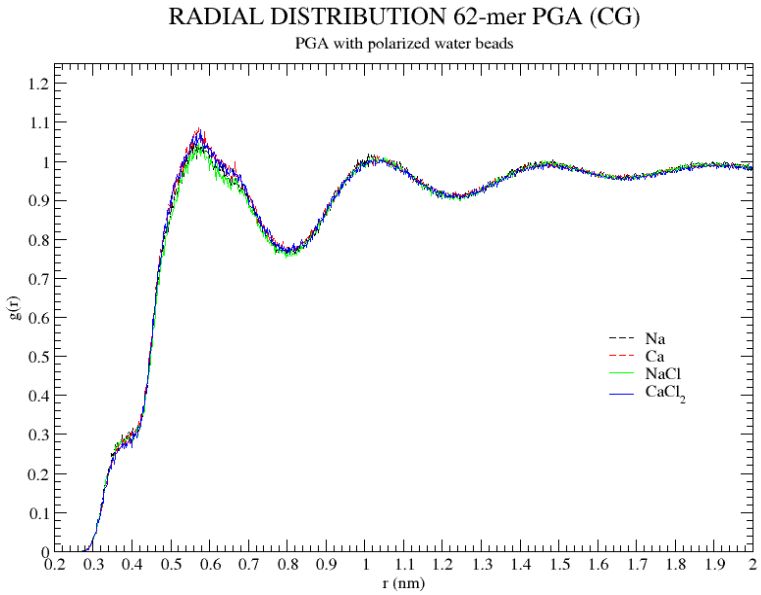


Figure 141 Average RDF between γ-l-PGA (centre of mass) and polarized water for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

The sharp peak at 0.572 ± 0.002 nm in indicates a well-structured solvation shell with ordered distribution of polarized water beads around the polymer chain.

5.8.7.2 RDF between the centre of mass of the γ-l-PGA beads and the neutral water bead

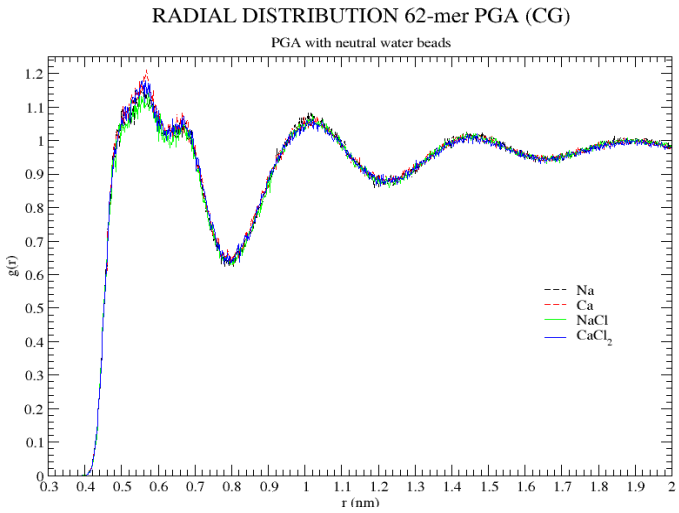


Figure 142 Average RDF between γ-l-PGA (centre of mass) and neutral water bead (W) for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

The RDF between centre of mass of γ-l-PGA and neutral water beads of polarized water shows us a major peak at 0.568 ± 0.002 nm shown in Fig. 142. A small shoulder peak can also be observed at 0.666 nm.

5.8.7.3 RDF between SQa beads and the polarized water

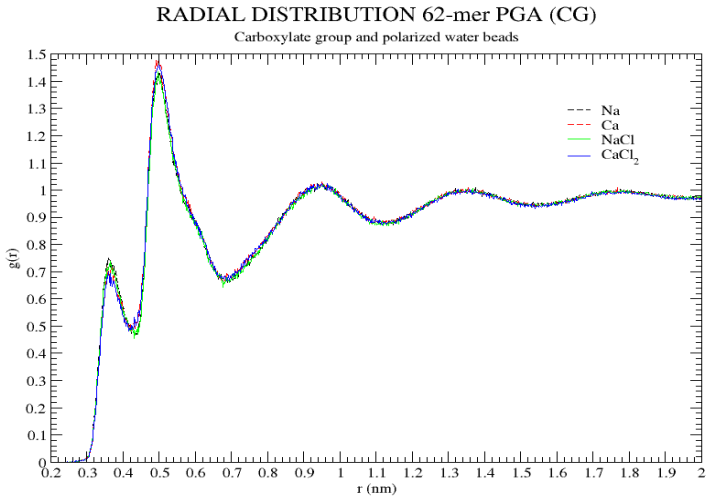


Figure 143 Average RDF between carboxylate group (SQa) and polarized water for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

The RDF between the SQa beads (carboxylate group) with respect to polarized water shows a strong peak at 0.496 ± 0.002 nm (see Fig. 143). A shoulder peak can be observed before the major peak at 0.356 ± 0.002 nm follows.

5.8.7.4 RDF between SNa beads and the polarized water

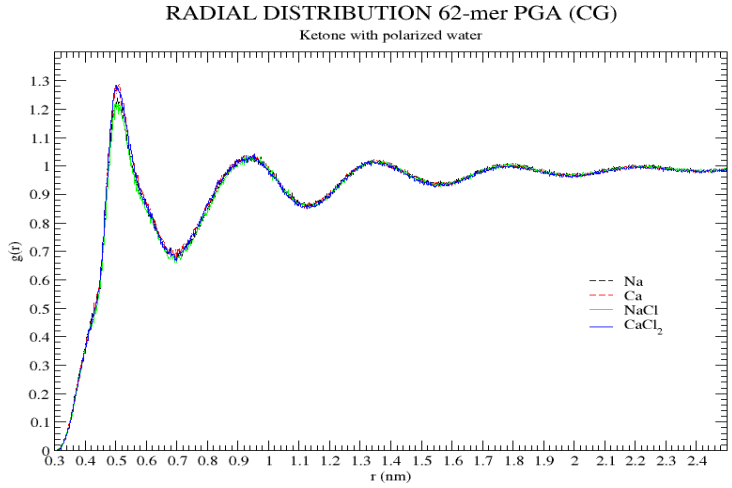


Figure 144 Average RDF between carbonyl group (SNa) and polarized water for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

The RDF between the SNa beads (carbonyl group) with respect to polarized water shows a strong peak at 0.500 ± 0.002 nm (see Fig. 144). A shoulder peak can be observed following the major peak at 0.950 ± 0.002 nm

5.8.7.5 RDF between SQd beads and polarized water

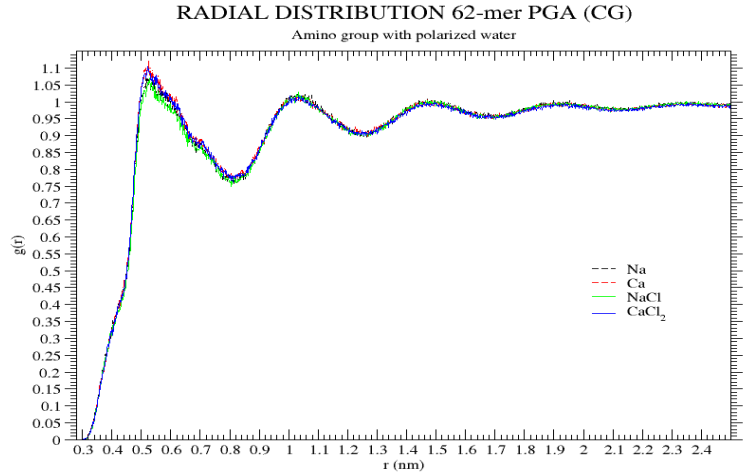


Figure 145 Average RDF between amino group (SQd) and polarized water for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

The RDF between the SQd beads (amino group) with respect to polarized water shows a strong peak at 0.514 ± 0.002 nm (see Fig. 145). A shoulder peak can be observed following the major peak at 1.036 ± 0.002 nm.

5.8.7.6 RDF between SQa beads and 'W' of polarized water

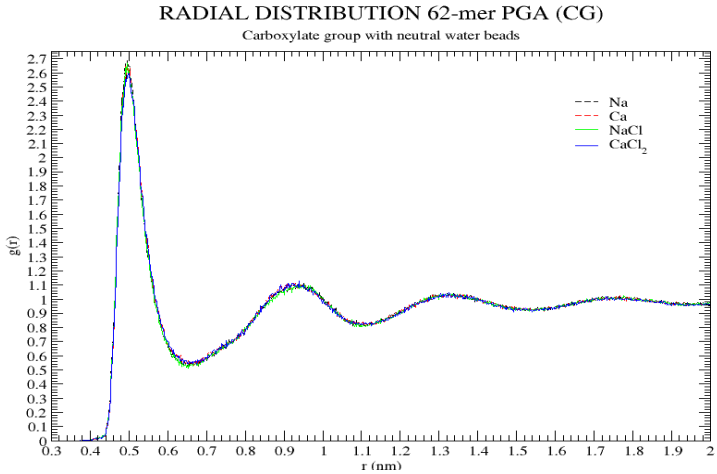


Figure 146 Average RDF between carboxylate group (SQa) and neutral water bead 'W' of polarized water for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

The RDF between the SQa beads (carboxylate group) with respect to neutral water beads of polarized water shows a strong peak at 0.496 ± 0.002 nm (see Fig. 126). A shoulder peak at 0.938 ± 0.002 nm follows

5.8.7.7 RDF between SNa beads and 'W' of polarized water

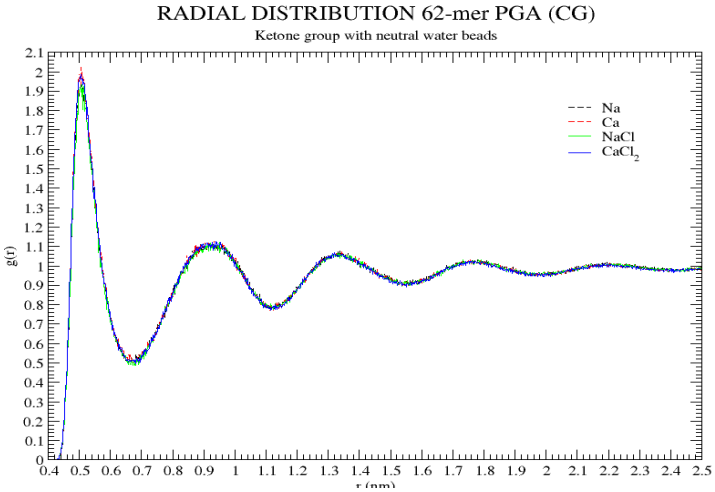


Figure 147 Average RDF between carbonyl group (SNa) and neutral bead (W) of polarized water for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

The RDF between the SNa beads (carbonyl group) with respect to neutral water beads of polarized water shows a strong peak at 0.500 ± 0.002 nm (see Fig. 127). A shoulder peak at 0.960 ± 0.002 nm follows

5.8.7.8 RDF between SQd beads and 'W' of polarized water

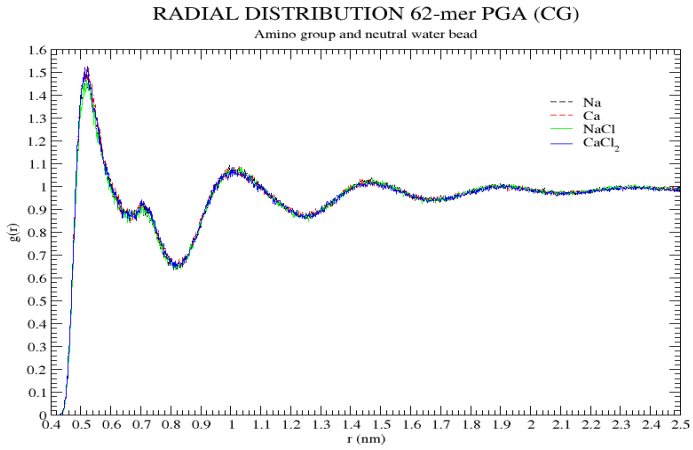


Figure 148 Average RDF between amino group (SQd) and neutral water bead 'W' of polarized water for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

The RDF between the SQd beads (amino group) with respect to neutral water beads of polarized water shows a strong peak at 0.512 ± 0.002 nm (see Fig. 128). A shoulder peak at 0.716 ± 0.002 nm follows.

5.8.7.9 RDF between SQa beads and Ions

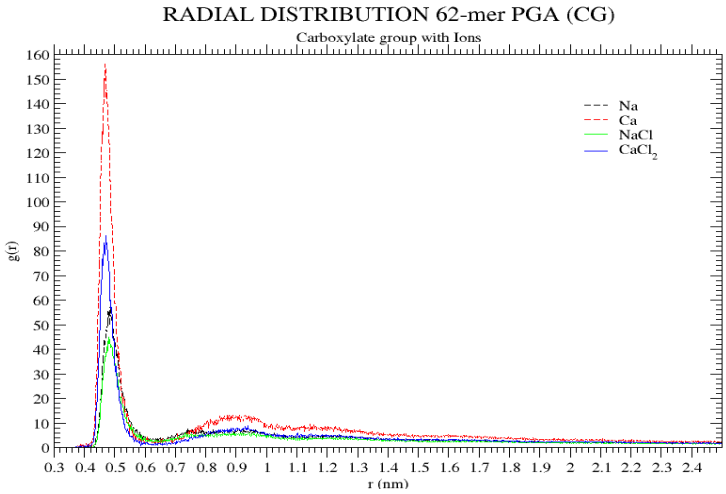


Figure 149 Average RDF between carboxylate group (SQa) and ions for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

As shown by Fig. 149, the RDF between the SQa (carboxylate group) and monovalent and divalent counterions exhibits a strong first peak at 0.468 ± 0.002 nm.

5.8.7.10 RDF between SNa beads and Ions

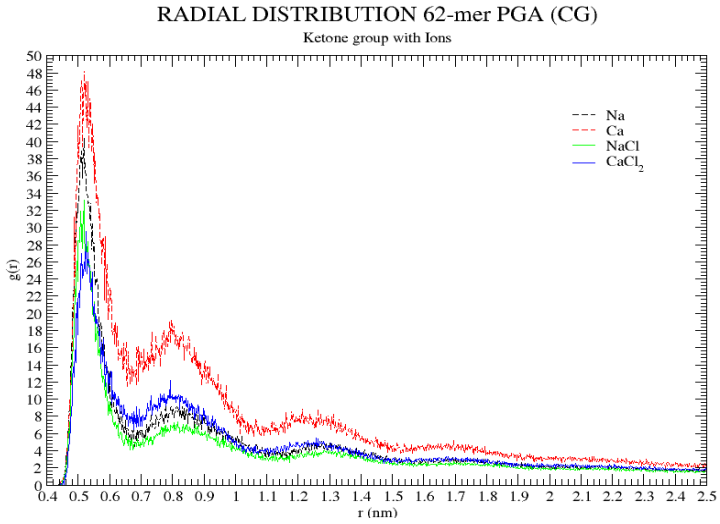


Figure 150 Average RDF between carbonyl group (SNa) and ions for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

As shown by Fig.150, the RDF between the SNa (carbonyl group) and monovalent and divalent counterions exhibits a strong first peak at 0.520 ± 0.002 nm, suggesting a preferred condensation of calcium ions on to the chain backbone of γ-l-PGA.

5.8.7.11 RDF between SQd beads and Ions

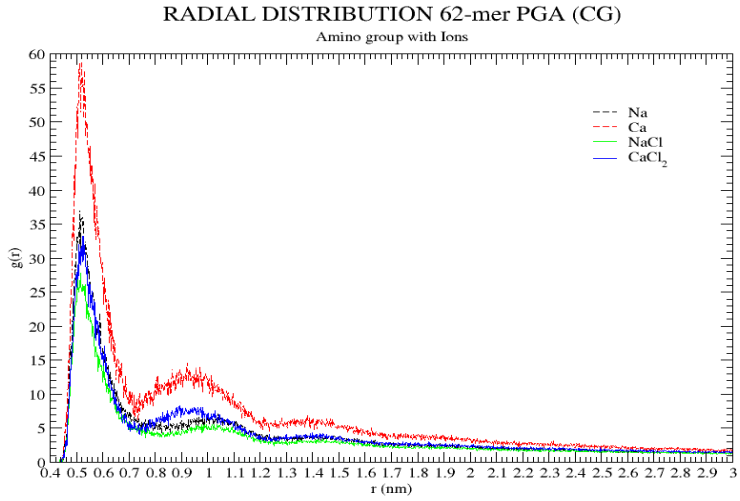


Figure 151 Average RDF between amino group (SQd) and ions of polarized water for all 3 independent production runs of γ-l-PGA (fully deprotonated) for 4 different systems simulation

As shown by Fig.151, the RDF between the SQd (amino group) and monovalent and divalent counterions exhibits a strong first peak at 0.520 ± 0.002 nm, suggesting a preferred condensation of calcium ions on to the amino group.

6 Discussion

In previous chapters all the characteristic properties of PAA and γ-l-PGA were presented and elaborated using all-atomistic and coarse-grained molecular dynamics simulation methods. In this chapter, first a comparison is drawn between the structural behaviour of PAA and γ-l-PGA. Since only all-atomistic simulation results for small chain for both the polymers are available, conclusion of the characteristic behaviour of only small chain polymers can be drawn.

Secondly, a comparison and discussion on the validation of our coarse-grained model is presented by comparing the previously presented results for short and long chain PAA and short chain γ-l-PGA. Again, due to the unavailability of the γ-l-PGA long chain all atomistic results, we do not consider this case.

6.1 Comparison of structural behavior of PAA and γ-l-PGA

For comparison of the structural properties and behaviour of PAA and γ-l-PGA the polymer, we need to look at the basic characteristic properties of these two polymers in salt free and excess salt conditions. From Fig. 151 it can be clearly observed the influence of divalent ions on the conformation of PAA and γ-l-PGA is far larger than that of monovalent ions. Both molecules show shrinkage in chain over the 100 ns simulation time in the presence of Calcium ions and furthermore in the presence of calcium chloride ions.

γ-l-PGA shows extreme folding of chain, showing a shrinkage of 78.5% whereas PAA shows 48% shrinkage in the presence of calcium chloride ions. In the presence of sodium chloride ions γ-l-PGA and PAA shows 39.4% and 24% decrease in size respectively. This can be explained due to the weaker electrostatic forces chain repulsion due to the different length of the polymer backbone chain and longer side chain.

The conformational change behaviour as observed from our systems are very similar for both the polymers, showing entrapment of ions with the folded chain leading to sharp decrease in chain length and remaining constant after a period of time.

Table 48 Standard deviation of radius of gyration from average values obtained from 3 independent runs for all atomistic systems of 20-mer PAA and γ-l-PGA

	PAA				PGA			
	Na	Ca	NaCl	$CaCl_2$	Na	Ca	NaCl	$CaCl_2$
Std Deviation over 100 ns (nm)	0.0477	0.1303	0.0591	0.1082	0.1628	0.3874	0.1900	0.3028
Std Deviation over 70-100 ns (nm)	0.0391	0.1607	0.0449	0.0885	0.1171	0.2954	0.1335	0.1530

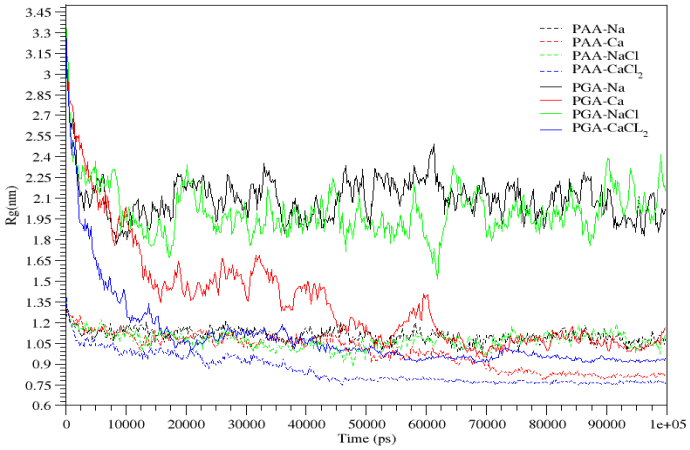


Figure 151 Time average radius of gyration of 20-mer PAA and γ-l-PGA from all atomistic simulation.

As the end-to-end distance is another parameter for measuring the ability of the polymer to change the conformation of molecules as the interaction with ions proceeds. Modifications in the end-to-end distance before and after MD can be related to the conformational changes caused by the interactions. From our all atomistic simulation (see Fig. 152) a change of 80.44% in γ-l-PGA and 85% in PAA in presence of calcium chloride and 22.7% in γ-l-PGA and 35% in PAA in presence of sodium chloride was observed.

Table 49 Standard deviation of end to end distance from average values obtained from 3 independent runs for all atomistic systems of 20-mer PAA and γ-l-PGA

	PAA				PGA			
	Na	Ca	NaCl	$CaCl_2$	Na	Ca	NaCl	$CaCl_2$
Std Deviation over 100 ns (nm)	0.3074	0.7594	0.3540	0.6323	0.6296	1.2015	0.7592	0.9602
Std Deviation over 70-100 ns (nm)	0.2587	1.0561	0.3215	0.4937	0.5382	0.9642	0.6655	0.5424

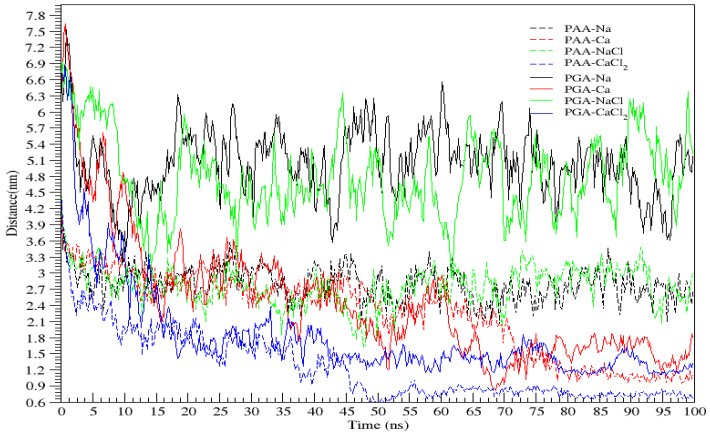


Figure 152 Time average end-to-end distance of 20-mer PAA and γ-l-PGA from all atomistic simulation

RMSD calculation is usually used to measure the similarity of 3-dimensional structure and to analyse the structural stability of polymer. As RMSD is only a measure for a global backbone deviation, the backbone RMSD of PAA with sodium and sodium chloride ions shows a deviation in the early stages of the simulation (within 15 ns) reaching a value of 0.681 and 0.691 nm whereas γ-l-PGA show a larger flexibility during the simulation reaching a value of 1.936 nm and 2.074 nm showing a 99.9% and 99.97% deviation from the initial value of PAA and γ-l-PGA in presence of sodium and sodium chloride ions respectively. After reaching a stable conformation i.e. the ions are bonded with the backbone PAA backbone becomes stiff hence the RMSD value reaches a stable value and shows minor fluctuation around the average RMSD value. The backbone of γ-l-PGA shows more flexibility compared to PAA as the fluctuations of RMSD value around the average value are higher compared to PAA during the rest of the simulation. This can be observed from the standard deviation of the RMSD value from the average RMSD value (see Table 50).

Table 50 Standard deviation of RMSD from average values obtained from 3 independent runs for all atomistic systems of 20-mer PAA and γ-l-PGA

	PAA				**PGA**			
	Na	Ca	NaCl	$CaCl_2$	Na	Ca	NaCl	$CaCl_2$
Std Deviation over 100 ns (nm)	0.0718	0.1739	0.0855	0.1399	0.1890	0.3775	0.2209	0.2987
Std Deviation over 70-100 ns (nm)	0.0580	0.2180	0.0677	0.1128	0.1455	0.3076	0.1732	0.1404

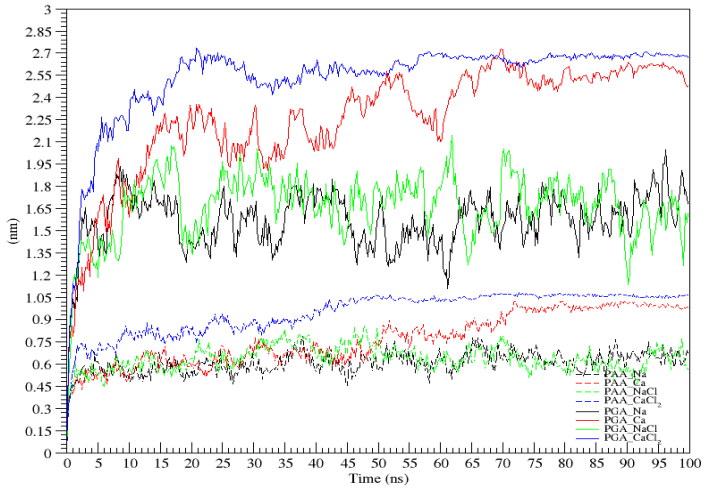


Figure 153 Time average RMSD of 20-mer PAA and γ-l-PGA from all atomistic simulation

Similarly, in the presence of calcium and calcium chloride ions PAA and γ-l-PGA shows a sharp increase in the RMSD in the early stages of the simulation, showing a flexible backbone structure. For PAA, the flexibility of the backbone was observed for 70 ns as the RMSD value increased from the initial value of 0.000492 nm to 1.024 nm in presence of calcium ions and 0.000508 to 1.068 nm in the presence of calcium chloride ions within 50 ns simulation time after which the backbone shows a stiffer characteristics and shows minor fluctuations around the average RMSD value for both the system. γ-l-PGA shows a similar behaviour in the presence of calcium ions though the increase in RMSD value for γ-l-PGA in calcium chloride system shows a flexible backbone for initial 20 ns reaching a value of 2.73 nm showing a deviation of 99.98% from the initial structure. In the presence of calcium ions γ-l-PGA shows high flexibility in the backbone for 70 ns, similar to PAA, reaching a maximum value of 2.731 nm. In the last 30 ns of the simulation all the systems reach a stable conformation and the structure possess a stiff backbone. This is represented by the standard deviation of backbone chain from the mean RMSD value of all the systems in Table 50.

The solvent accessible surface area (SASA) is used as a parameter in characterization of polymer folds and in describing solvation of macromolecules like PAA and γ-l-PGA. The results are in complete agreement with the radii of gyration and end to end distance values. Not surprisingly, the SASA of PAA is far below the value of γ-l-PGA (see Fig. 51) as a result of different structure and backbone length of γ-l-PGA.

From Table 51, it can be clearly observed γ-l-PGA shows a higher affinity towards the divalent ions in comparison to PAA as the rate of decrease of γ-l-PGA is much higher that the PAA in divalent ion systems. This is due to the presence of additional functional groups in the backbone chain which allows formation of more bonds. Since the backbone of γ-l-PGA shows higher flexibility than PAA, γ-l-PGA is able to easily fold and entrap ions within the folded conformation hence the lower SASA results.

Since γ-l-PGA works as a hydrogen bond donor and acceptor at the same time, the total number of hydrogen bonds formed is lot bigger than the corresponding number of hydrogen bonds of PAA. γ-l-PGA has one more carboxylate ion in the C-terminus therefore it is able form more hydrogen bonds. Likewise, the first peak in RDF between the carboxylate oxygens of γ-l-PGA and water hydrogens is slightly higher. Thereby γ-l-PGA coordinates more water molecules in the first solvation shell than PAA.

Table 51 Standard deviation of SASA from average values obtained from 3 independent runs for all atomistic systems of 20-mer PAA and γ-l-PGA

	PAA				PGA			
	Na	Ca	NaCl	$CaCl_2$	Na	Ca	NaCl	$CaCl_2$
Std Deviation over 100 ns (nm)	0.1769	0.5206	0.2322	0.8737	0.5453	3.2485	0.7248	0.8737
Std Deviation over 70-100 ns (nm)	0.1903	0.7624	0.2021	0.9337	0.5833	2.9901	0.6542	0.9337

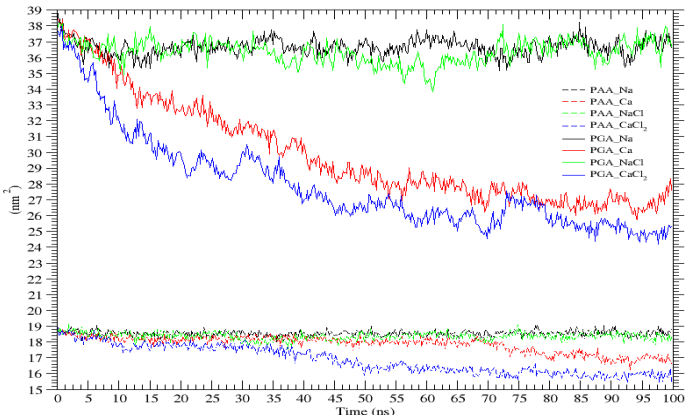


Figure 154 Time average SASA of 20-mer PAA and γ-l-PGA from all atomistic simulation

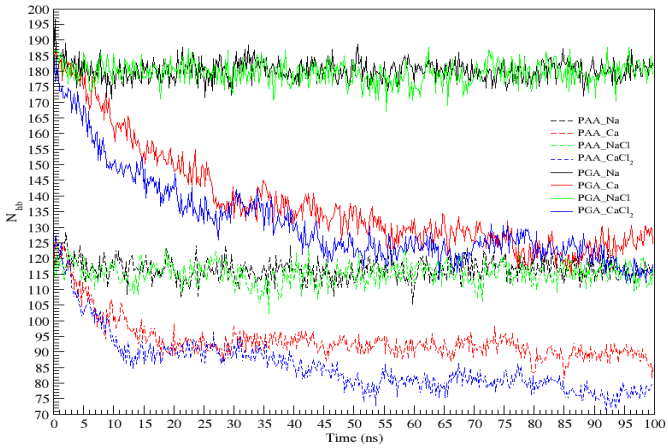


Figure 155 Time average number of hydrogen bonds of 20-mer PAA and γ-l-PGA from all atomistic simulation

The interaction of polyelectrolytes with water and counterions is another property that must be investigated in order to present the characteristic difference of PAA and γ-l-PGA. For this purpose, radial distribution functions of water and sodium & calcium ions around different models are calculated.

The radial distribution function (RDF) between two kind of particles A and B describes the frequency with a particle of type B can be found at a distance *r* from a particle type A, based on this frequency that two particles of an ideal solution are in this distance. Therefore, it can provide a better understanding of the hydration shell and counterion interaction.

Regarding the water interaction, no major difference between PAA and γ-l-PGA models can be observed. Since both PAA and γ-l-PGA have 20 deprotonated carboxylate groups, they have very similar, well-structured solvation shell. The peak in the RDF between the center of mass of the polymer and the oxygen atoms of water (see Fig 156a) is not only different in height, which equals the number of oxygen atoms in the first solvation shell, but also in shape. The RDF´s of oxygen water around the centre of mass of each polymer repeating unit show a comparatively small first peak for both PAA and γ-l-PGA (see Fig. 156a) implying a strong screening of the centre of mass due to the highly charged environment. The RDF´s of water oxygen around carboxylate groups display a prominent first peak. Every carboxylate group is surrounded by approximately 3 water oxygen (see Fig 156b). Moreover, several smaller peaks follow the first peak, which implies a well-structured solvation shell with ordered distribution of water molecules around the polymer chain. The comparison of RDF´s of sodium ions around the carboxylate group in Fig. 156c shows a higher peak for PAA model compared to γ-l-PGA. This suggests that PAA is more strongly interacting with sodium and calcium counterions that γ-l-PGA.

For both long-chain polymer models similar distribution functions are obtained. This shows that the interaction of PAA and γ-l-PGA with water and sodium and calcium counterions are independent from chain length.

As calcium ions play a crucial role in the interaction between polymer and surfactant, the influence of calcium ions on polymer structure must be determined. Subsequently, this will allow a distinction between effects caused specifically by the presence of surfactants molecules and other unspecific influences caused by surrounding ions.

From the results obtained from all atomistic simulations and presented in previous chapter structural changes are visible for both PAA and γ-l-PGA. The calcium ions bind tightly to the carboxylate groups of the polymer. The electrostatic repulsion between the negatively charged groups is screened more efficiently when using divalent ions, resulting in a stronger coiling behaviour.

The RDF´s of calcium and sodium ions around the carboxylate group pf PAA and γ-l-PGA short chain all atomistic model is shown in Fig. 156c confirm the visual impression. The sodium peak is negligible compared to the calcium peak.

Furthermore, the RDF´s suggest that water molecules are blocked from the polymer in the presence of calcium ions as the peak height decreases drastically compared to Fig. 156a. Additionally, the interaction of PAA with calcium ions is slightly stronger compared to γ-l-PGA.

As already seen in Fig. 156, the strong shielding of the electrostatic repulsion caused by the presence of the calcium ion leads to the major structural changes in both polymer model. The size of PAA and γ-l-PGA model decreases significantly regardless of the chain length. Additionally, the stronger coiling behaviour reduces the solvent accessible surface. Other simulation studies have shown, that "like-charge attraction"[31, 36] of carboxylate ions can occur via bridging calcium ions. This effect is caused by stronger binding affinity of calcium ions to the carboxylate groups compared to sodium ions, which is verified by the RDF´s an from Raman spectroscopy[33].

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

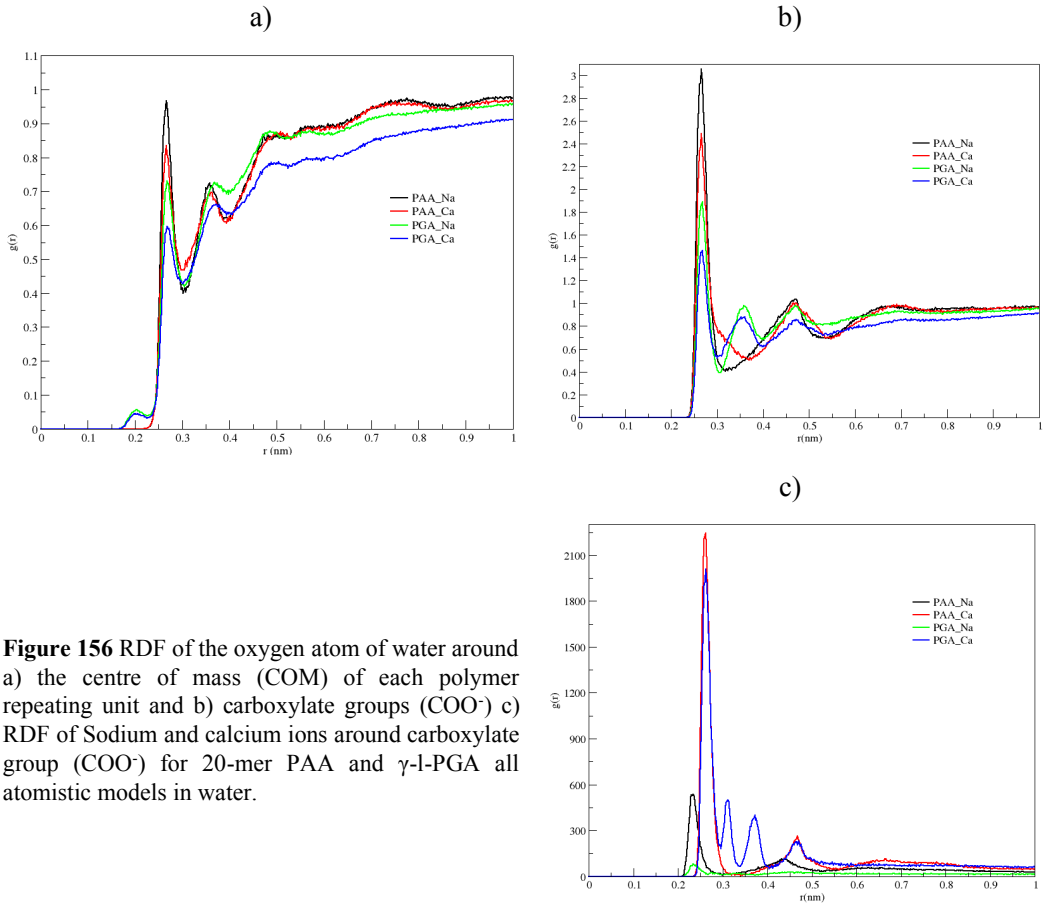


Figure 156 RDF of the oxygen atom of water around a) the centre of mass (COM) of each polymer repeating unit and b) carboxylate groups (COO^-) c) RDF of Sodium and calcium ions around carboxylate group (COO^-) for 20-mer PAA and γ-l-PGA all atomistic models in water.

6.2 Comparison of Coarse-grained model with All atomistic Model

As mentioned in a previous chapter, the radius of gyration is commonly used for proteins as a measure of its compactness, but it is also useful for polymers and polyelectrolytes as it gives a sense of the size of a polymer coil.

The radius of gyration for the 20-mer MARTINI model of PAA with sodium chloride ions is stable over the course of time with an average value of 0.09 ± 0.04 nm (see Fig. 157). As seen in Table.52, this result is in good agreement with other simulation studies[25, 26, 28, 30] and previously performed all-atomistic simulations. Compared with PAA, the radius of gyration for the γ-L-PGA MARTINI model exhibits more fluctuations, indicating a more flexible structure. Due to the longer polymer backbone the average value for the radius of gyration is 2.00 ± 0.14 nm, which is slightly larger than value calculated based on all-atomistic simulations (see Table 53).

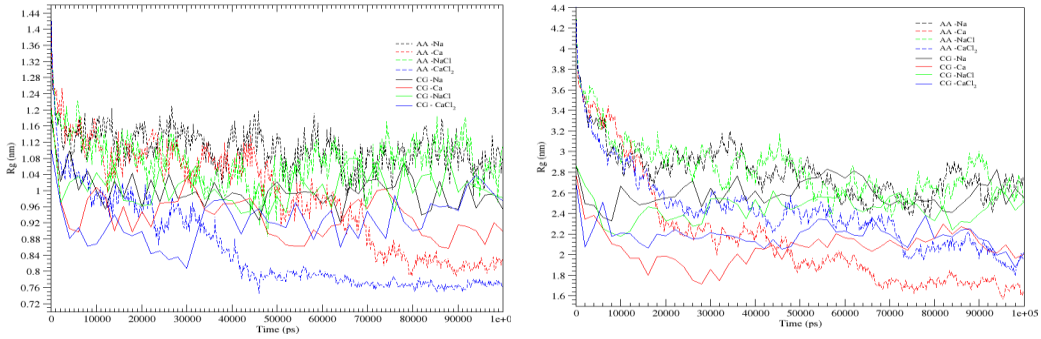
For the 62-mer MARTINI model of PAA the deviation from the results published by Reith *et al.*[28] increases. As Reith *et al.*[28] followed the first presented mapping approach, where one repeating unit was mapped to one CG bead, the results from all-atomistic simulations are more trustworthy. A comparison shows that CG and all-atomistic simulations are in perfect agreement. The radius of gyration for the long-chain γ-L-PGA MARTINI model is within the error limits of the result obtained with all-atomistic simulations as shown in Table 53.

Table 52 Radius of gyration Rg and end-to-end distance R for the MARTINI model of fully deprotonated 20-mer PAA calculated in previous chapter. The results for the PAA MARTINI model are compared with simulation results from AA MD simulations.

		20-mer		62-mer	
		R_g (nm)	R (nm)	R_g (nm)	R (nm)
MARTINI		0.9 ± 0.04	2.7 ± 0.3	2.49 ± 0.13	5.05 ± 0.7
AA MD		1.07 ± 0.07	2.7 ± 0.4	2.83 ± 0.27	7.1 ± 1.3
Biermann *et al.*[25]		1.27*	3.6		
Reith *et al.*[28]		1.06#		2.3#	
Sulatha *et al.*[30]	FF-1	0.994	2.1		
	FF-2	0.997	2.8		
Katiyar *et al.*[26]		1.16 ± 0.01			

**for 23-mer PAA*

#extrapolated



a) b)

Figure 157 Comparison of time average radius of gyration from all atomistic and coarse-grained simulations over 100 ns of a) 20-mer PAA b) 62-mer PAA

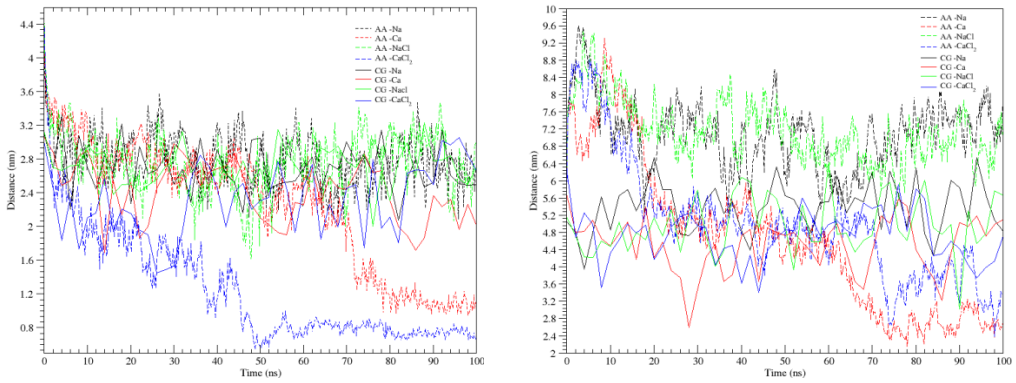
The comparison between the long-chain PAA and γ-L-PGA MARTINI models shows again the influence of the different backbone length, the average radius of gyration for γ-L-PGA is nearly twice as large as for PAA long-chain MARTINI model.

The end-to-end distance *R,* defined as the distance between the first (head group) and last (tail group) monomer of a polymer, is another characteristic of polymer coiling behavior.

Like the radius of gyration, the end-to-end distance for the PAA MARTINI model is stable during the simulation. The average distance from head to tail group is 2.7 ± 0.3 nm, which is in good agreement with the results obtained from the FF-2 parameter set published by Sulatha *et al.*[30] and from all-atomistic simulations (see Table.53). According to the higher flexibility of the polymer chain, the end-to-end distance for the γ-L-PGA short chain MARTINI model fluctuates strongly around an average value of 3.9 ± 0.6 nm. A comparison with all-atomistic simulations reveals a slightly less coiled polymer structure for the MARTINI model (see Table.53).

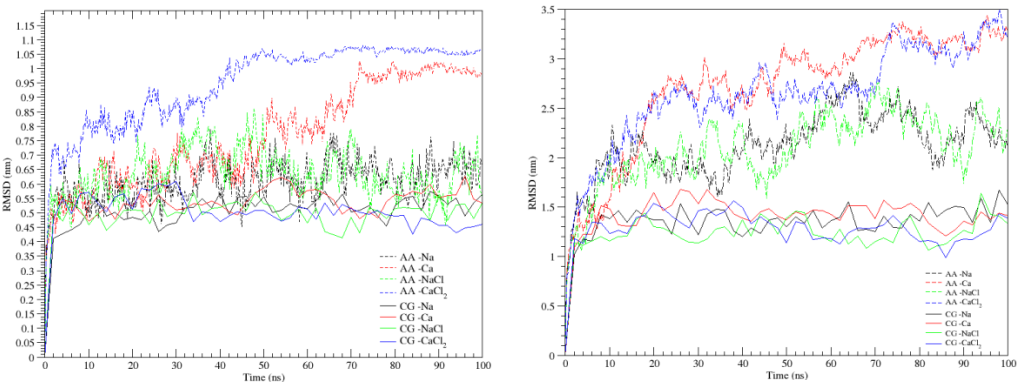
Again, the long-chain MARTINI model of PAA shows perfect agreement with the all-atomistic simulation regarding the end-to-end distance while the deviation for γ-L-PGA MARTINI models is slightly larger but still within the error limits of the all-atomistic simulation model (see Table.53).

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66



a) b)

Figure 158 Comparison of time average end-to-end distance from all atomistic and coarse-grained simulations over 100 ns of a) 20-mer PAA b) 62-mer PAA



a) b)

Figure 159 Comparison of time average RMSD from all atomistic and coarse-grained simulation over 100 ns of a) 20-mer PAA b) 62-mer PAA

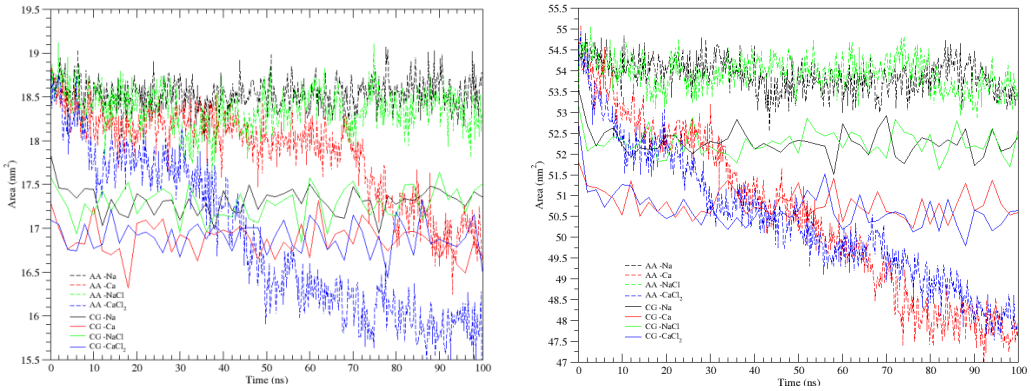
Table 53 Comparison of radii of gyration R_g, end-to-end distances R and *SASA* for the MARTINI models of PAA and γ-L-PGA with the respective results calculated based on all-atomistic MD (AA MD) simulations.

			R_g (nm)	R (nm)	*SASA* (nm^2)
PAA	20-mer	MARTINI	0.99 ± 0.04	2.7 ± 0.3	17.3 ± 0.2
		AA MD	1.07 ± 0.07	2.7 ± 0.4	18.4 ± 0.2
	62-mer	MARTINI	2.49 ± 0.13	5.1 ± 0.7	52.2 ± 0.5
		AA MD	2.83 ± 0.27	7.2 ± 1.3	53.9 ± 1.0
γ-L-PGA	20-mer	MARTINI	2.00 ± 0.14	3.9 ± 0.6	30.8 ± 0.5
		AA MD	2.00 ± 0.16	4.8 ± 0.5	36.3 ± 0.9
	62-mer	MARTINI	4.8 ± 0.38	5.9 ± 0.9	94.4 ± 0.8
		AA MD	*Results not available*		

SASA was first describe by Lee and Richards[115] as a sphere, representing a solvent molecule, rolling over the van der waals surface of a polymer and by tracing the centre of sphere. As of Gromacs version 2019, the SASA of a molecule is calculated based on the double cubic

lattice method as proposed by Eisenhaber *et al.*[128]. Here, a probe radius of 0.235 nm is chosen to represent the rolling sphere according to the increased radii of a MARTINI water bead.

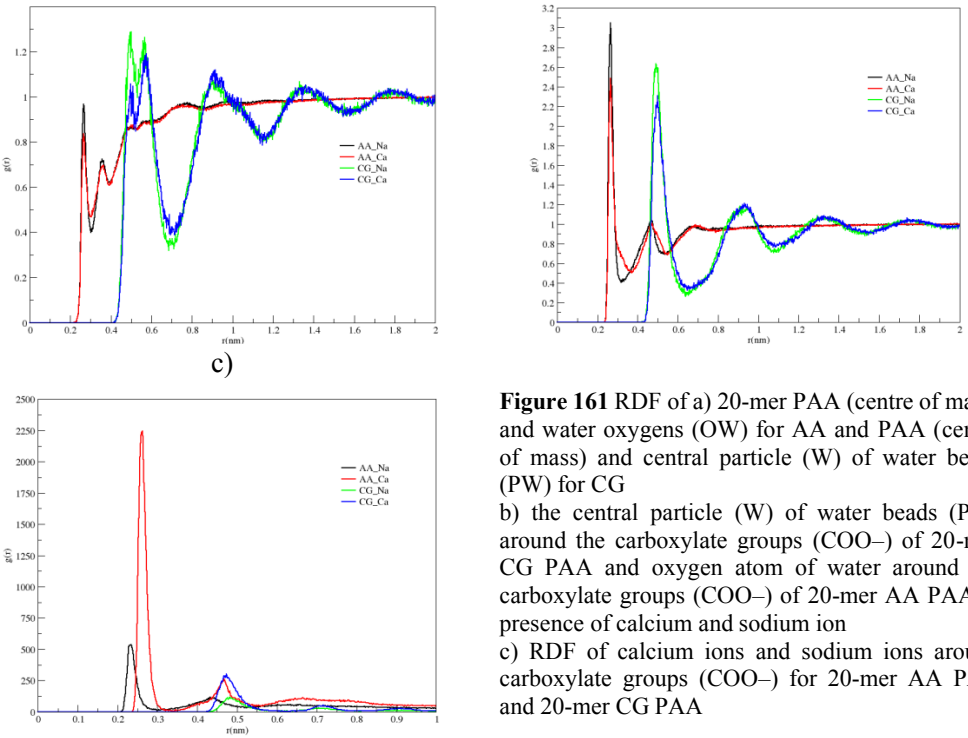
The average SASA calculated for the fully deprotonated PAA and γ-l-PGA MARTINI and all atomistic models has been presented in the results section. Difference between MARTINI and all atomistic simulations are expected due to the different probe radius.



a) b)

Figure 160 Comparison of time average SASA from all atomistic and coarse-grained simulation over 100 ns of **a)** 20-mer PAA **b)** 62-mer PAA

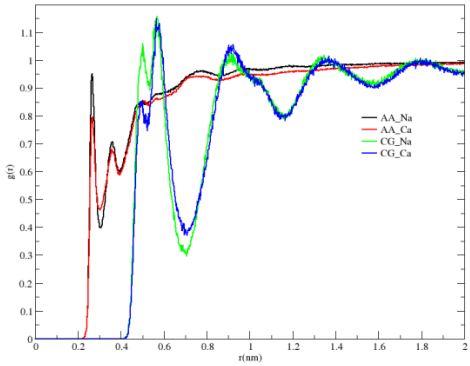
a) b)



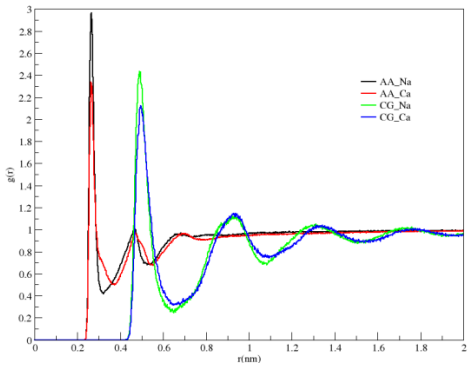
c)

Figure 161 RDF of a) 20-mer PAA (centre of mass) and water oxygens (OW) for AA and PAA (centre of mass) and central particle (W) of water beads (PW) for CG
b) the central particle (W) of water beads (PW) around the carboxylate groups (COO–) of 20-mer CG PAA and oxygen atom of water around the carboxylate groups (COO–) of 20-mer AA PAA in presence of calcium and sodium ion
c) RDF of calcium ions and sodium ions around carboxylate groups (COO–) for 20-mer AA PAA and 20-mer CG PAA

a)



b)



c)

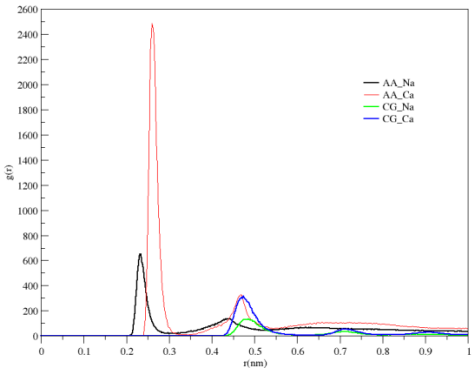
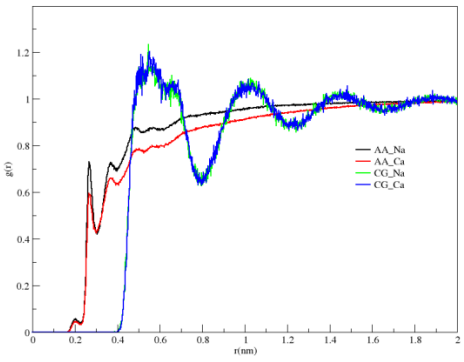
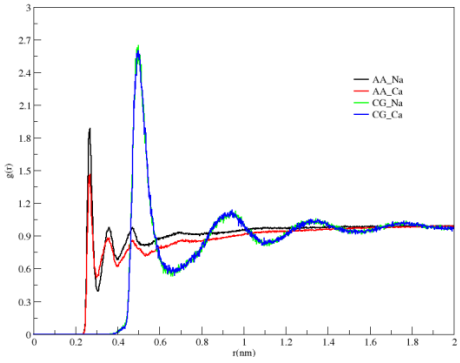


Figure 162 RDF of a) 62-mer PAA (centre of mass) and water oxygens (OW) for AA and PAA (centre of mass) and central particle (W) of water beads (PW) for CG
b) the central particle (W) of water beads (PW) around the carboxylate groups (COO^-) of 62-mer CG PAA and oxygen atom of water around the carboxylate groups (COO^-) of 62-mer AA PAA in presence of calcium and sodium ion
c) RDF of calcium ions and sodium ions around carboxylate groups (COO^-) for 62-mer AA PAA and 62-mer CG PAA

a)



b)



c)

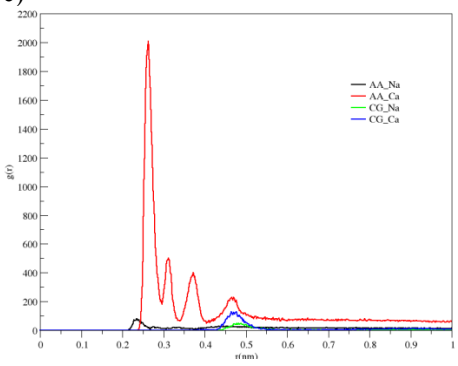


Figure 163 RDF of a) 20-mer γ-l-PGA (centre of mass) and water oxygens (OW) from AA and γ-l-PGA (centre of mass) and central particle (W) of water beads (PW) from CG
b) the central particle (W) of water beads (PW) around the carboxylate groups (COO^-) of 20-mer CG γ-l-PGA and oxygen atom of water around the carboxylate groups (COO^-) of 20-mer AA γ-l-PGA in presence of calcium and sodium ion
c) RDF of calcium ions and sodium ions around carboxylate groups (COO^-) from 20-mer AA γ-l-PGA and 20-mer CG γ-l-PGA

The calculated RDF´s for all-atomistic and coarse-grained simulations of PAA and γ-l-PGA of chain length 20 and 62 are compared and can be seen in Fig. 161, 162 & 163. In the first figures (see Fig. 161a, 162a, 163a) a comparison of hydration behaviour of our polymers from AA and CG systems is shown. The RDF´s of water beads around the centre of mass of each polymer repeating unit show a comparatively higher first peak for both PAA and γ-l-PGA implying a weak screening of the centre of mass. The RDF´s of water beads around the carboxylate groups (SQa/Qa beads) display a prominent first peak. For PAA, the coarse-grained RDF´s presents a lower number of water beads surrounding the SQa/Qa (see Fig. 161b, 162b). Whereas for γ-l-PGA, the CG model over predicts this value (See Fig. 163b). Although the CG RDF´s show a well solvated system for both the polymers as they are present several smaller peaks following the prominent peak implying ordered distribution of water molecules around polymer chain. These results are in qualitative agreement with previously performed all-atomistic simulation.

The coarse-grained models are unable to depict the RDF´s in agreement with all atomistic results but show a similar graph representation. The offset in the AA and CG graphs can be explained due to a larger probe distance used in CG models as the bead sizes are bigger than the actual size of the atoms. Another parameter contributing to this offset is the charge associated with the neutral bead depicting the oxygen atom of water molecules in polarized water viz. zero. whereas in all atomistic simulations the charge given to the oxygen atom is -0.8476. By representing a water molecule with a bead, the internal structure of the water molecule is lost i.e. the information carried by the RDF's of H-O and H-H. To add this information in a coarse-grained model pressure correction can be used to perfectly match the radial distribution function and hence perfectly reproduce the compressibility, which determines density fluctuations, would be accurately reproduced. For pressure correction, iterative Boltzmann inversion can be used to approximate the RDF's compared to all-atomistic system. To obtain a coarse-grained model that can reproduce the all-atom pressure, one can also extrapolate between pressure corrected coarse-grained SPC/E model.

The comparison of the RDF´s of sodium and calcium ions (Qd beads) around the SQa/Qa beads in Fig. 161c, 162c, 163c shows much lower peaks and distribution for PAA MARTINI and γ-l-PGA MARTINI models. From the Fig. 161c, 162c, 163c, it can be concluded that while simulating the CG polymer in polarized water gives reasonable predictions for the global properties of the polymer, it lacks the capability to properly capture the distribution of ions around a polymer. The under-predicted electrostatic interactions in polarized water can in turn influence the binding of the polymer with other charged molecules such as nucleic acids in gene delivery. Even interactions with neutral molecules can be influenced if they have local charges. Therefore, despite of its computational efficiency, the polarized water model should be used with caution in CG polymer simulations.

These significant deviations reflect the fact that the coarse-grained models approximate a system with significantly more degrees of freedom, more spatially varying and effectively anisotropic interactions.

7 Summary and Outlook

The aim of this research report was to parameterize polyacrylic acid (PAA) and polyglutamic acid (PGA), which is a biodegradable alternative for PAA. In order to investigate the polymer-ion interaction using molecular dynamics (MD) simulations. The parameterization of polyacrylic acid and polyglutamic acid (γ-L-PGA) according to AMBER99SB force field with Antechamber and ACPYPE was performed successfully. The obtained RESP charges seemed reasonable and the analysis of the MD simulation of PAA & γ-L-PGA with water with addition of sodium, calcium, sodium chorine and calcium chloride ions was in a good agreement with the literature[25, 28, 30]. The RDF analysis showed that sodium ions tend to condensate on the polymer backbone resulting in strong artificial chain folding and enclosing of sodium ions although this is unlikely to happen as reported by Raman spectroscopy[44]. On the other hand, calcium ions showed stronger condensation on the polymer backbone. To solve this problem, different force fields like GROMOS, OPLS or CHARMM should be tested. However, parameterization is difficult, as some of them require semi-empirical fitting of data.

Furthermore, this report focused on a second polymer, polyglutamic acid (PGA). Poly-γ-glutamic acid (γ-L-PGA) was also the subject of investigation in this report as it as an amino acid and thus its topology is defined in AMBER99SB. The analysis of the simulations revealed a similar hydration behaviour of PAA and γ-L-PGA. However, γ-L-PGA showed a stronger folding behaviour compared to PAA, which had a rather stretched conformation due to the strong chain rigidity. The strong folding of γ-L-PGA is caused by the weaker electrostatic chain repulsion due to the different length of the polymer backbone and the longer side chains. Furthermore, the RDF analysis showed a weaker interaction between γ-L-PGA and the surrounding sodium ions. This seems to be reasonable but unfortunately there are no results reported in literature and requires experimental results to validate these findings. For comparison γ-L-PGA should also be simulated within different force fields.

The studied impact of different counterions on the structure of PAA and γ-L-PGA shows clearly the divalent counterions reduce the chain dimensions compared to the monovalent ones, promoting interchain association at high concentration via electrostatic bridging which should lead to viscosity reduction.

Future simulations will focus on the interaction between polymer and ions through all atomistic simulations (currently under simulation) in order to understand the interaction mechanisms and consequently the influence of ion on polymers and therefore influence of polymer on the washing performance of enzymes in laundry detergents.

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

8 Conclusion

In conclusion, we have developed an all atomistic and coarse-grained model to describe the structure of polyelectrolyte solution allowing for systematic variation of chain length, tacticity, and concentration which can be linked to the parent atomistic model. The model uses an explicit solvent (water) and the polyelectrolytes and counterions were modelled explicitly. The CG model is basically based on the effective pair potential taken from the all-atomistic MD simulations. CG bonded potentials were calculated and obtained separately from atomistic sampling of random walk of a single polyelectrolyte chain in vacuum. CG models make available the possibility to simulate longer-chain and for longer period, in comparison with MD simulation.

Comparison of CG and AA models show that the resulting CG models reproduces the atomistic ally sampled local conformation of PAA and PGA chains in solution. The CG model was able to reproduce the positions of radial distribution functions (RDF) curve between polyelectrolytes chain and ions, although the heights of the peak are underestimated.

Upon studying the counterion distribution around the PAA and PGA chain through CG models. The measure was not represented accurately by direct comparison with the explicit solvent atomistic model. The CG model does not replicate the local structural characteristics of the electrolyte in the atomistic explicit solvent system and demonstrates an underprediction of the counterion density. Because of the lofty negative charge density of the polyelectrolyte chain, dielectric properties of hydration water differ from bulk water as well as water molecules engaged in facilitating ionic interactions at the pair level. In previous, alternative developments of CG implicit solvent models, the illustration of counterion distributions around the polyelectrolyte has been attained with coarse graining procedures that iteratively improve a set of effective potentials until counterion distribution functions obtained from detailed atomistic simulations of polyelectrolyte solutions are reproduced. With the dielectric constant used in this study peaks in the RDF curve underestimates the heights of the peaks. Because of the strong electrostatic interactions between the polyelectrolyte charges and its counterions, the local concentration of ions close to polyelectrolyte chains are much higher than at infinite dilution of ions where the pair potential was developed. By fine tuning the dielectric permittivity established on local concentration of counterions around PAA and PGA chains, the CG RDF curve can be considerably enhanced in order to replicate the counterion structure around the chain from AA simulations. Unbonded potentials also needs to be fitted into the models which can be developed from calculating the pair PMF's between two small molecules or ions in water representing corresponding CG beads.

This study shows that the interactions between the counterion and the polyelectrolytes chain analyzed by CG model are too weak matched to AA models. Impacts of adding salt on the shift of radius of gyration of PAA chains with various chain length were also studied. The divergence of CG model and atomistic simulation on R_g values demonstrate the non-transferability of our CG model upon salt conditions, which can be corrected by deriving new potentials at each different salt concentration. Due to unavailability of results from the all atomistic simulation of PGA, no comments can be made on the validation of our CG PGA models. AA models for PGA are currently under simulation and will be presented in the next study. The model requires further development and has to be tested in comparison with experimental systems of the same degree of polymerization and also under conditions of finite concentrations and varied ionic strengths before being passed as a valid models to predict similar results as of all atomistic MD simulations.

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

9 Appendix

9.1 GENERATION OF POLYMER MODELS USING POLYBUILD

To generate a model for fully deprotonated PAA with a chain length of 20, Polybuild must be executed by entering

python polypoid –p pab–syndio –l 20 –f mm2 --smilesonly

–p	name of the polymer	*–f*	force field (always mm2)
–l	length of polymer chain	*--smilesonly*	only *.smiles* file as output

Polybuild creates a file called *pab-syndio.smiles* that contains the polymer structure in smiles code. The full library entry for deprotonated PAA can be found in. This chapter explains how to add new polymer molecules to the Polybuild library.

9.2 HOW TO MODIFY (ADD A NEW MONOMER TO) LIBPOLYMERS.XML

If we want to add a new monomer to the polybuild library, we have used this structure (example for fully deprotonated polyacrylic acid):

```
<polymer>
  <!--
  Polymer: Polyacrylic Acid deprotonated
  Force fields: mm2
  -->
  <name>pab-syndio</name>
  <structure>syndio</structure>
  <scheme>A B</scheme>
  <head>
        <smi>[H]</smi>
  </head>
  <r_unit>
        <monomer>
              <name>A</name>
              <mass>71</mass>
              <smi>C([H])([H])C([H])(C(=O)[O-])</smi>
        </monomer>
        <monomer>
              <name>B</name>
              <mass>71</mass>
```

```
                <smi>C([H])([H])C(C(=O)[O-])([H])</smi>
            </monomer>
        </r_unit>
        <tail>
            <smi>[H]</smi>
        </tail>
        <forcefields>
            <forcefield>
                <ffname>mm2</ffname>
                <at_head>5</at_head>
                <at_unit>
                    <monomer>
                        <name>A</name>
                        <at_mon>1 5 5 1 5 3 7 47</at_mon>
                        <name>B</name>
                        <at_mon>1 5 5 1 3 7 47 5 </at_mon>
                    </monomer>
                </at_unit>
                <at_tail>5</at_tail>
            </forcefield>
        </forcefields>
    </polymer>
```

In order to define a new polymer, we have added a new entry starting with <*polymer*>. This is followed by a short definition of the polymer and its <*name*>. In polybuild we have the option to create isotactic, syndiotactic or atactic polymers, which can be set by the entry <*structure*>.

To get a syndiotactic structure we need to define the <*scheme*> with A and B. This means that the polymer has two different building blocks that alternate strictly. Of course, this can be more than two building blocks. Polybuild has a built-in function to create polymers with different percentages of the monomers and it can use different statistical methods (equiprobable or Bernoulli) to distribute the monomers within the polymer.

```
    <polymer>
        <!--
        Polymer: Polyacrylic Acid deprotonated
        Force fields: mm2
        -->
        <name>pab-syndio</name>
```

<structure>syndio</structure>

<scheme>A B</scheme>

The second part defines the polymer structure in smiles code. Every polymer consists of a head group (<*head*>), repeating units (<*r_unit*>) and a tail group (<*tail*>). For different building blocks, different monomers must be defined within section <*r_unit*>. Polybuild can create polymers with different length but it can create polymers with a specific molecular weight as well. Because of this, defining a <*mass*> for every monomer is necessary.

<head>

<smi>[H]</smi>

</head>

<r_unit>

<monomer>

<name>A</name>

<mass>71</mass>

<smi>C([H])([H])C([H])(C(=O)[O-])</smi>

</monomer>

<monomer>

<name>B</name>

<mass>71</mass>

<smi>C([H])([H])C(C(=O)[O-])([H])</smi>

</monomer>

</r_unit>

<tail>

<smi>[H]</smi>

</tail>

However, polybuild will not only generate a *.smiles* file but also an *.mm2* file with the polymer topology. To create a topology, polybuild needs a force field (for the added polymers it's always mm2) and the definition of the used atom types within the chosen force field. The atom types of head group (<*at_head*>), the different monomers (<*at_mon*>) and tail group (<*at_tail*>) are defined in the last part. Each atom needs a specific number in dependence of its chemical function within the polymer.

<forcefields>

<forcefield>

<ffname>mm2</ffname>

<at_head>5</at_head>

<at_unit>

<monomer>

```
                    <name>A</name>
                    <at_mon>1 5 5 1 5 3 7 47</at_mon>
                    <name>B</name>
                    <at_mon>1 5 5 1 3 7 47 5 </at_mon>
                </monomer>
            </at_unit>
            <at_tail>5</at_tail>
        </forcefield>
    </forcefields>
</polymer>
```

There are 2 files and one webpage in

/media/storage/b/data/m.koenig/Melanie_home/software/polybuild/Atom_Types

With help of these it is possible to determine the atom type number of every atom. For example, the atom types of the carboxylic acid group (-COOH) in PAA: the carbon in this group is an sp^2-hybridized carbonyl atom, so it gets number 3; one of the oxygen gets number 7 because it is a carbonyl oxygen (=O); the other oxygen, which has one bonded hydrogen, is considered as an alcohol, so it gets number 6 and the hydrogen gets 21.

9.3 RENAMING RESIDUES

```
import shutil

gehtlosbei = 31

shutil.copyfile('PAB_atactic.pdb','PAB_atactic_copy.pdb')

polymer = open('PAB_atactic_copy.pdb', 'r+')
lines = polymer.readlines()

list = []
i = 0
start='1H'

#search for 1H to define blocks
for line in lines:
        if line.find(start) > -1:
                list.append(i)
        i = i + 1
print list

#rename first residue (headgroup)
k = 0
y = 0
y = list[(k+1)] - list[k]
if y == 11:
        for i in range(0,10):
                lines[list[k]+i] = lines[list[k]+i].replace('UNK','PAH')

        lines[list[k]] = lines[list[k]].replace('1H','HH',1)
        lines[list[k]+1] = lines[list[k]+1].replace(' C',' C',1)
        lines[list[k]+2] = lines[list[k]+2].replace('2H','H1',1)
        lines[list[k]+3] = lines[list[k]+3].replace('3H','H2',1)
        lines[list[k]+4] = lines[list[k]+4].replace(' C','C1',1)
        lines[list[k]+5] = lines[list[k]+5].replace(' H','H3',1)
```

```
        lines[list[k]+6] = lines[list[k]+6].replace(' C','C2',1)
        lines[list[k]+7] = lines[list[k]+7].replace('O','O',1)
        lines[list[k]+8] = lines[list[k]+8].replace(' O','O1',1)
        lines[list[k]+9] = lines[list[k]+9].replace('H','H',1)

if y == 10:
        for i in range(0,9):
                lines[list[k]+i] = lines[list[k]+i].replace('UNK','PBH')

        lines[list[k]] = lines[list[k]].replace('1H','HH',1)
        lines[list[k]+1] = lines[list[k]+1].replace(' C',' C',1)
        lines[list[k]+2] = lines[list[k]+2].replace('2H','H1',1)
        lines[list[k]+3] = lines[list[k]+3].replace('3H','H2',1)
        lines[list[k]+4] = lines[list[k]+4].replace(' C','C1',1)
        lines[list[k]+5] = lines[list[k]+5].replace(' H','H3',1)
        lines[list[k]+6] = lines[list[k]+6].replace(' C','C2',1)
        lines[list[k]+7] = lines[list[k]+7].replace(' O','O1',1)
        lines[list[k]+8] = lines[list[k]+8].replace(' O','O2',1)

#rename following residues (PAA or PAB) in dependence of block length
k = 1
y = 0
while k < len(list)-2:

        y = list[(k+1)] - list[k]
        if y == 8:
                for i in range(-1,7):
                        lines[list[k]+i] = lines[list[k]+i].replace('UNK','PAB')
                        if k < 9:
                        lines[list[k]+i] = lines[list[k]+i].replace(' 1 ', ' ' + str(k+1) + ' '  )
                        else:
                        lines[list[k]+i] = lines[list[k]+i].replace(' 1 ', str(k+1) + ' ')
                if k % 2:
                        lines[list[k]-1] = lines[list[k]-1].replace('C','C',1)
                        lines[list[k]] = lines[list[k]].replace('1H','H1',1)
```

```
                lines[list[k]+1] = lines[list[k]+1].replace('2H','H2',1)
                lines[list[k]+2] = lines[list[k]+2].replace(' C','C1',1)
                lines[list[k]+3] = lines[list[k]+3].replace(' C','C2',1)
                lines[list[k]+4] = lines[list[k]+4].replace(' O','O1',1)
                lines[list[k]+5] = lines[list[k]+5].replace(' O','O2',1)
                lines[list[k]+6] = lines[list[k]+6].replace(' H','H3',1)

            else:
                lines[list[k]-1] = lines[list[k]-1].replace('C','C',1)
                lines[list[k]] = lines[list[k]].replace('1H','H1',1)
                lines[list[k]+1] = lines[list[k]+1].replace('2H','H2',1)
                lines[list[k]+2] = lines[list[k]+2].replace(' C','C1',1)
                lines[list[k]+3] = lines[list[k]+3].replace(' H','H3',1)
                lines[list[k]+4] = lines[list[k]+4].replace(' C','C2',1)
                lines[list[k]+5] = lines[list[k]+5].replace(' O','O1',1)
                lines[list[k]+6] = lines[list[k]+6].replace(' O','O2',1)


        elif y == 9:
            for i in range(-1,8):
                lines[list[k]+i] = lines[list[k]+i].replace('UNK','PAA')
            if k < 9:
                lines[list[k]+i] = lines[list[k]+i].replace(' 1 ', ' ' + str(k+1) + ' '  )
            else:
                lines[list[k]+i] = lines[list[k]+i].replace(' 1 ', str(k+1) + ' ')

            if k % 2:
                lines[list[k]-1] = lines[list[k]-1].replace('C','C',1)
                lines[list[k]] = lines[list[k]].replace('1H','H1',1)
                lines[list[k]+1] = lines[list[k]+1].replace('2H','H2',1)
                lines[list[k]+2] = lines[list[k]+2].replace(' C','C1',1)
                lines[list[k]+3] = lines[list[k]+3].replace(' C','C2',1)
                lines[list[k]+4] = lines[list[k]+4].replace(' O',' O',1)
                lines[list[k]+5] = lines[list[k]+5].replace(' O','O1',1)
                lines[list[k]+6] = lines[list[k]+6].replace(' H','H3',1)
```

```
                lines[list[k]+7] = lines[list[k]+7].replace('1H','H4',1)

        else:
                lines[list[k]-1] = lines[list[k]-1].replace('C','C',1)
                lines[list[k]] = lines[list[k]].replace('1H','H1',1)
                lines[list[k]+1] = lines[list[k]+1].replace('2H','H2',1)
                lines[list[k]+2] = lines[list[k]+2].replace(' C','C1',1)
                lines[list[k]+3] = lines[list[k]+3].replace(' H','H3',1)
                lines[list[k]+4] = lines[list[k]+4].replace(' C','C2',1)
                lines[list[k]+5] = lines[list[k]+5].replace(' O',' O',1)
                lines[list[k]+6] = lines[list[k]+6].replace(' O','O1',1)
                lines[list[k]+7] = lines[list[k]+7].replace('1H','H4',1)

    k = k + 1


#rename first residue (tailgroup)
k = len(list)-2
y = list[(k+1)] - list[k]
if y == 7:
        for i in range(-1,9):
                lines[list[k]+i] = lines[list[k]+i].replace('UNK','PAT')
                lines[list[k]+i] = lines[list[k]+i].replace(' 1 ','20 ')

        lines[list[k]-1] = lines[list[k]-1].replace('C','C',1)
        lines[list[k]] = lines[list[k]].replace('1H','H1',1)
        lines[list[k]+1] = lines[list[k]+1].replace('2H','H2',1)
        lines[list[k]+2] = lines[list[k]+2].replace(' C','C1',1)
        lines[list[k]+3] = lines[list[k]+3].replace(' C','C2',1)
        lines[list[k]+4] = lines[list[k]+4].replace('O','O',1)
        lines[list[k]+5] = lines[list[k]+5].replace(' O','O1',1)
        lines[list[k]+6] = lines[list[k]+6].replace(' H ',' H ',1)
        lines[list[k]+7] = lines[list[k]+7].replace('1H','H3',1)
        lines[list[k]+8] = lines[list[k]+8].replace('2H','H4',1)
```

```
if y == 6:
        for i in range(-1,8):
                lines[list[k]+i] = lines[list[k]+i].replace('UNK','PBT')
                lines[list[k]+i] = lines[list[k]+i].replace(' 1 ','20 ')

        lines[list[k]-1] = lines[list[k]-1].replace('C','C',1)
        lines[list[k]] = lines[list[k]].replace('1H','H1',1)
        lines[list[k]+1] = lines[list[k]+1].replace('2H','H2',1)
        lines[list[k]+2] = lines[list[k]+2].replace(' C','C1',1)
        lines[list[k]+3] = lines[list[k]+3].replace(' C','C2',1)
        lines[list[k]+4] = lines[list[k]+4].replace(' O','O1',1)
        lines[list[k]+5] = lines[list[k]+5].replace(' O','O2',1)
        lines[list[k]+6] = lines[list[k]+6].replace('1H','H3',1)
        lines[list[k]+7] = lines[list[k]+7].replace('2H','H4',1)

f = open('PAB_atactic_renamed.pdb', 'w')

for line in lines:
        f.write(line)

polymer.close
f.close
```

9.4 GROMACS TOPOLOGY OF FULLY DEPROTONATED PAA FOR AMBER99SB FORCE FIELD

Deprotonated repeating unit

```
[ PAB ]
 [ atoms ]
   C     CT          -0.007678     1
   H1    HC           -0.022983     2
   H2    HC           -0.022983     3
   C1    CT          -0.183658     4
   C2     C           0.833072     5
   O1    O2           -0.806779     6
   O2    O2           -0.806779     7
   H3    HC            0.017789     8
 [ bonds ]
   C          -C1
   C    H1
   C    H2
```

```
 C    C1
 C1   H3
 C1   C2
 C1   +C
 C2   O1
 C2   O2
```

Deprotonated head group

```
[ PBH ]
 [ atoms ]
   HH    HC          0.040521     1
   C     CT         -0.249123     2
   H1    HC          0.040521     3
   H2    HC          0.040521     4
   C1    CT         -0.060095     5
   C2     C          0.794538     6
   O1    O2         -0.803602     7
   O2    O2         -0.803602     8
   H3    HC          0.000322     9
 [ bonds ]
   HH     C
    C    H1
    C    H2
    C    C1
   C1    H3
   C1    C2
   C2    O1
   C2    O2
   C1    +C
```

Deprotonated tail group

```
[ PBT ]
 [ atoms ]
   C     CT         -0.209424     1
   H1    HC          0.022221     2
   H2    HC          0.022221     3
   C1    CT          0.155759     4
   C2     C          0.821426     5
   O1    O2         -0.815820     6
   O2    O2         -0.815820     7
   H3    HC         -0.090282     8
   H4    HC         -0.090282     9
 [ bonds ]
    C         -C1
    C    H1
    C    H2
    C    C1
   C1    H3
   C1    C2
   C1    H4
```

C2 O1
C2 O2

9.5 ANALYSIS PROTOCOL FOR POLYACRYLIC ACID (PAA)

The analysis protocol for fully deprotonated PAA is given below.

```
#!/bin/bash
folder=$PWD
cd $folder

#########################

for ((i=1; i<=3; i++ ))
    do

cd run_$i

#################### Analysis

###temperature

echo 12 0 | g_energy -s md_0_1.tpr -f ener.edr -o temperature.xvg
###pressure

echo 14 0 | g_energy -s md_0_1.tpr -f ener.edr -o pressure.xvg

###energies

echo 11 0 | g_energy -s md_0_1.tpr -f ener.edr -o total_energie.xvg

echo 9 0 | g_energy -s md_0_1.tpr -f ener.edr -o potential.xvg

echo 10 0 | g_energy -s md_0_1.tpr -f ener.edr -o kinetic.xvg

###radius of gyration
```

```
echo 1 | gmx gyrate -s md_0_1.tpr -f traj_comp.xtc -o gyrate.xvg
```

###Radial Distribution function (RDF)

#first create new index entrys

```
gmx make_ndx -f npt.gro << EOF
a C2
a OW
a HW1 | a HW2
a O1 | a O2
q
EOF
```

#calculate RDF between different groups

```
echo 1 17  | gmx rdf  -s md_0_1.tpr -f traj_comp.xtc -o rdf_PAA_OW.xvg -n index.ndx –rdf res_com
echo 19 17  | gmx rdf  -s md_0_1.tpr -f traj_comp.xtc -o rdf_COO_OW.xvg -n index.ndx
echo 19 18  | gmx rdf  -s md_0_1.tpr -f traj_comp.xtc -o rdf_COO_HW.xvg -n index.ndx
echo 19 14  | gmx rdf  -s md_0_1.tpr -f traj_comp.xtc -o rdf_COO_Na.xvg -n index.ndx
```

#####Hydrogens bonds

```
echo 1 | gmx hbond -s md_0_1.tpr -f traj_comp.xtc –num hbond.xvg –tu ns
```

###Surface Accesible Area

```
echo 1 | gmx sasa -s md_0_1.tpr -f traj_comp.xtc -o sasa.xvg
```

###end-to-end distance

```
echo 2 4 |gmx mindist -s md_0_1.tpr -f traj_comp.xtc -od mindist.xvg
```

```
    cd ..
done
```

9.6 ANALYSIS PROTOCOL FOR POLY-Γ-GLUTAMIC ACID (Γ-PGA)

The analysis protocol for α-PGA is based on the previous developed protocol for PAA. As only the types of hydrogen bonds differ, this part of the protocol had to be adjusted:

```
#####Hydrogens bonds

#first create new index entrys

gmx make_ndx -f npt.gro << EOF
a OW
a HW1 | a HW2
a OE1 | a OE2 | a OC1 | a OC2
r NGLU
r CGLU
a O
a N | a H | a H1 | a H2 | a H3
q
EOF

#calculate hydrogen bonds between different groups

echo 1 12 | gmx hbond -s md_0_1.tpr -f traj_comp.xtc -num hbonds.xvg -tu ns

echo 21 12 | gmx hbond -s md_0_1.tpr -f traj_comp.xtc -num hbonds_COO.xvg -n
index.ndx -tu ns

echo 24 12 | gmx hbond -s md_0_1.tpr -f traj_comp.xtc -num hbonds_O.xvg -n
index.ndx -tu ns

echo 25 12 | gmx hbond -s md_0_1.tpr -f traj_comp.xtc -num hbonds_N.xvg -n
index.ndx -tu ns
```

9.7 GROMACS WORKFLOW

Set path to the GROMACS library (contains the topology of the deprotonated PAA)

export GMXLIB=/media/storage/b/data/m.koenig/Melanie_home/
data/PAA_models/top2itp/top

##System preparation.
#Topology generation for the 20-mer model of deprotonated PAA within AMBER99SB force field and with SPC/E water model

echo 2 6 | pdb2gmx –f paa_renamed.pdb –o 20PAA.gro

#Generation of a simulation box with V=945.2 nm^3 around PAA

gmx editconf –f 20PAA.gro –o 20PAA_box.gro –c –box 9.814 9.814 9.814 –bt cubic

#Solvation of the box with water

gmx solvate –cp 20PAA_box.gro –cs spc216.gro –o 20PAA_solv.gro –p topol.top

#Replacing the water with a suitable amount of ions

gmx grompp –f ions.mdp –c 20PAA_solv.gro –p topol.top –o ions.tpr

echo 6 | gmx genion –s ions.tpr –o 20PAA_solv_ions.gro –p topol.top –np 26 –nn 6

#Energy minimization

gmx grompp –f minim.mdp –c 20PAA_solv_ions.gro –p topol.top –o em.tpr

gmx mdrun –v –deffnm em

#Analysis of the potential energy curve

echo 9 0 | gmx energy –f em.edr –o potential.xvg

##System equilibration

#Creating index file with two groups

```
gmx make_ndx –f 20PAA_solv_ions.gro << EOF
name 1 Polymer
name 12 others
q
EOF
```

#NVT equilibration
gmx grompp –f nvt.mdp –c em.gro –p topol.top –o nvt.tpr –n index.ndx

gmx mdrun –v –deffnm nvt

#Analysis of temperature progression

echo 14 0 | gmx energy –f nvt.edr –o temp.xvg

#NPT equilibration

gmx grompp –f npt.mdp –c nvt.gro –t nvt.cpt –p topol.top –o npt.tpr –n index.ndx

gmx mdrun –v –deffnm npt

#Analysis of pressure and density progression

echo 15 0 | gmx energy –f npt.edr –o pressure.xvg

echo 21 0 | gmx energy –f npt.edr –o density.xvg

##Production run

gmx grompp –f md.mdp –c npt.gro –t npt.cpt –p topol.top –o md_0_1.tpr –n index.ndx

gmx mdrun –deffnm md_0_1

#When running on RWTH cluster (in parallel)

gmx_mpi mdrun –deffnm md_0_1

9.8 .ITP FILE OF 20-MER PAA COARSE GRAINED MODEL

[moleculetype]

;molname	nrexcl
PAA	2

[atoms]

;nr	type	resnr	residu	atom	cgnr	charge	mass
1	SQa	1	PAH	COO	1	-1	44.0098
2	SC1	1	PAH	BB	2	0	48.5901
3	SQa	2	PAA	COO	3	-1	44.0098
4	SC1	2	PAA	BB	4	0	27.0458
5	SQa	3	PAA	COO	5	-1	44.0098
6	SC1	3	PAA	BB	6	0	27.0458
7	SQa	4	PAA	COO	7	-1	44.0098
8	SC1	4	PAA	BB	8	0	27.0458
9	SQa	5	PAA	COO	9	-1	44.0098
10	SC1	5	PAA	BB	10	0	27.0458
11	SQa	6	PAA	COO	11	-1	44.0098
12	SC1	6	PAA	BB	12	0	27.0458
13	SQa	7	PAA	COO	13	-1	44.0098
14	SC1	7	PAA	BB	14	0	27.0458
15	SQa	8	PAA	COO	15	-1	44.0098
16	SC1	8	PAA	BB	16	0	27.0458
17	SQa	9	PAA	COO	17	-1	44.0098
18	SC1	9	PAA	BB	18	0	27.0458
19	SQa	10	PAA	COO	19	-1	44.0098
20	SC1	10	PAA	BB	20	0	27.0458
21	SQa	11	PAA	COO	21	-1	44.0098
22	SC1	11	PAA	BB	22	0	27.0458
23	SQa	12	PAA	COO	23	-1	44.0098
24	SC1	12	PAA	BB	24	0	27.0458
25	SQa	13	PAA	COO	25	-1	44.0098
26	SC1	13	PAA	BB	26	0	27.0458

27	SQa	14	PAA	COO	27	-1	44.0098
28	SC1	14	PAA	BB	28	0	27.0458
29	SQa	15	PAA	COO	29	-1	44.0098
30	SC1	15	PAA	BB	30	0	27.0458
31	SQa	16	PAA	COO	31	-1	44.0098
32	SC1	16	PAA	BB	32	0	27.0458
33	SQa	17	PAA	COO	33	-1	44.0098
34	SC1	17	PAA	BB	34	0	27.0458
35	SQa	18	PAA	COO	35	-1	44.0098
36	SC1	18	PAA	BB	36	0	27.0458
37	SQa	19	PAA	COO	37	-1	44.0098
38	SC1	19	PAA	BB	38	0	23.0422
39	SQa	20	PAT	COO	39	-1	44.0098

[bonds]

;i	j	funct	length	force.c.
1	2	1	0.298	46000
2	3	1	0.273	35000
3	4	1	0.273	35000
4	5	1	0.273	35000
5	6	1	0.273	35000
6	7	1	0.273	35000
7	8	1	0.273	35000
8	9	1	0.273	35000
9	10	1	0.273	35000
10	11	1	0.273	35000
11	12	1	0.273	35000
12	13	1	0.273	35000
13	14	1	0.273	35000
14	15	1	0.273	35000
15	16	1	0.273	35000
16	17	1	0.273	35000
17	18	1	0.273	35000
18	19	1	0.273	35000
19	20	1	0.273	35000

20	21	1	0.273	35000	
21	22	1	0.273	35000	
22	23	1	0.273	35000	
23	24	1	0.273	35000	
24	25	1	0.273	35000	
25	26	1	0.273	35000	
26	27	1	0.273	35000	
27	28	1	0.273	35000	
28	29	1	0.273	35000	
29	30	1	0.273	35000	
30	31	1	0.273	35000	
31	32	1	0.273	35000	
32	33	1	0.273	35000	
33	34	1	0.273	35000	
34	35	1	0.273	35000	
35	36	1	0.273	35000	
36	37	1	0.273	35000	
37	38	1	0.273	35000	
38	39	1	0.298	46000	

[angles]

;i	j	k	funct	angle	force.c.
1	2	3	1	133	210
2	3	4	1	43	250
3	4	5	1	142	70
4	5	6	1	43	250
5	6	7	1	142	70
6	7	8	1	43	250
7	8	9	1	142	70
8	9	10	1	43	250
9	10	11	1	142	70
10	11	12	1	43	250
11	12	13	1	142	70
12	13	14	1	43	250
13	14	15	1	142	70

14	15	16	1	43	250
15	16	17	1	142	70
16	17	18	1	43	250
17	18	19	1	142	70
18	19	20	1	43	250
19	20	21	1	142	70
20	21	22	1	43	250
21	22	23	1	142	70
22	23	24	1	43	250
23	24	25	1	142	70
24	25	26	1	43	250
25	26	27	1	142	70
26	27	28	1	43	250
27	28	29	1	142	70
28	29	30	1	43	250
29	30	31	1	142	70
30	31	32	1	43	250
31	32	33	1	142	70
32	33	34	1	43	250
33	34	35	1	142	70
34	35	36	1	43	250
35	36	37	1	142	70
36	37	38	1	43	250
37	38	39	1	133	210

9.9 .ITP FILE OF 62-MER PAA COARSE GRAINED MODEL

[moleculetype]

;molname	nrexcl
PAA	2

[atoms]

;nr	type	resnr	residu	atom	cgnr	charge	mass
1	SQa	1	PAH	COO	1	-1	44.0098
2	SC1	1	PAH	BB	2	0	48.5901
3	SQa	2	PAA	COO	3	-1	44.0098
4	SC1	2	PAA	BB	4	0	27.0458
5	SQa	3	PAA	COO	5	-1	44.0098
6	SC1	3	PAA	BB	6	0	27.0458
7	SQa	4	PAA	COO	7	-1	44.0098
8	SC1	4	PAA	BB	8	0	27.0458
9	SQa	5	PAA	COO	9	-1	44.0098
10	SC1	5	PAA	BB	10	0	27.0458
11	SQa	6	PAA	COO	11	-1	44.0098
12	SC1	6	PAA	BB	12	0	27.0458
13	SQa	7	PAA	COO	13	-1	44.0098
14	SC1	7	PAA	BB	14	0	27.0458
15	SQa	8	PAA	COO	15	-1	44.0098
16	SC1	8	PAA	BB	16	0	27.0458
17	SQa	9	PAA	COO	17	-1	44.0098
18	SC1	9	PAA	BB	18	0	27.0458
19	SQa	10	PAA	COO	19	-1	44.0098
20	SC1	10	PAA	BB	20	0	27.0458
21	SQa	11	PAA	COO	21	-1	44.0098
22	SC1	11	PAA	BB	22	0	27.0458
23	SQa	12	PAA	COO	23	-1	44.0098
24	SC1	12	PAA	BB	24	0	27.0458
25	SQa	13	PAA	COO	25	-1	44.0098
26	SC1	13	PAA	BB	26	0	27.0458
27	SQa	14	PAA	COO	27	-1	44.0098
28	SC1	14	PAA	BB	28	0	27.0458

29	SQa	15	PAA	COO	29	-1	44.0098
30	SC1	15	PAA	BB	30	0	27.0458
31	SQa	16	PAA	COO	31	-1	44.0098
32	SC1	16	PAA	BB	32	0	27.0458
33	SQa	17	PAA	COO	33	-1	44.0098
34	SC1	17	PAA	BB	34	0	27.0458
35	SQa	18	PAA	COO	35	-1	44.0098
36	SC1	18	PAA	BB	36	0	27.0458
37	SQa	19	PAA	COO	37	-1	44.0098
38	SC1	19	PAA	BB	38	0	27.0458
39	SQa	20	PAA	COO	39	-1	44.0098
40	SC1	20	PAA	BB	40	0	27.0458
41	SQa	21	PAA	COO	41	-1	44.0098
42	SC1	21	PAA	BB	42	0	27.0458
43	SQa	22	PAA	COO	43	-1	44.0098
44	SC1	22	PAA	BB	44	0	27.0458
45	SQa	23	PAA	COO	45	-1	44.0098
46	SC1	23	PAA	BB	46	0	27.0458
47	SQa	24	PAA	COO	47	-1	44.0098
48	SC1	24	PAA	BB	48	0	27.0458
49	SQa	25	PAA	COO	49	-1	44.0098
50	SC1	25	PAA	BB	50	0	27.0458
51	SQa	26	PAA	COO	51	-1	44.0098
52	SC1	26	PAA	BB	52	0	27.0458
53	SQa	27	PAA	COO	53	-1	44.0098
54	SC1	27	PAA	BB	54	0	27.0458
55	SQa	28	PAA	COO	55	-1	44.0098
56	SC1	28	PAA	BB	56	0	27.0458
57	SQa	29	PAA	COO	57	-1	44.0098
58	SC1	29	PAA	BB	58	0	27.0458
59	SQa	30	PAA	COO	59	-1	44.0098
60	SC1	30	PAA	BB	60	0	27.0458
61	SQa	31	PAA	COO	61	-1	44.0098
62	SC1	31	PAA	BB	62	0	27.0458
63	SQa	32	PAA	COO	63	-1	44.0098

64	SC1	32	PAA	BB	64	0	27.0458
65	SQa	33	PAA	COO	65	-1	44.0098
66	SC1	33	PAA	BB	66	0	27.0458
67	SQa	34	PAA	COO	67	-1	44.0098
68	SC1	34	PAA	BB	68	0	27.0458
69	SQa	35	PAA	COO	69	-1	44.0098
70	SC1	35	PAA	BB	70	0	27.0458
71	SQa	36	PAA	COO	71	-1	44.0098
72	SC1	36	PAA	BB	72	0	27.0458
73	SQa	37	PAA	COO	73	-1	44.0098
74	SC1	37	PAA	BB	74	0	27.0458
75	SQa	38	PAA	COO	75	-1	44.0098
76	SC1	38	PAA	BB	76	0	27.0458
77	SQa	39	PAA	COO	77	-1	44.0098
78	SC1	39	PAA	BB	78	0	27.0458
79	SQa	40	PAA	COO	79	-1	44.0098
80	SC1	40	PAA	BB	80	0	27.0458
81	SQa	41	PAA	COO	81	-1	44.0098
82	SC1	41	PAA	BB	82	0	27.0458
83	SQa	42	PAA	COO	83	-1	44.0098
84	SC1	42	PAA	BB	84	0	27.0458
85	SQa	43	PAA	COO	85	-1	44.0098
86	SC1	43	PAA	BB	86	0	27.0458
87	SQa	44	PAA	COO	87	-1	44.0098
88	SC1	44	PAA	BB	88	0	27.0458
89	SQa	45	PAA	COO	89	-1	44.0098
90	SC1	45	PAA	BB	90	0	27.0458
91	SQa	46	PAA	COO	91	-1	44.0098
92	SC1	46	PAA	BB	92	0	27.0458
93	SQa	47	PAA	COO	93	-1	44.0098
94	SC1	47	PAA	BB	94	0	27.0458
95	SQa	48	PAA	COO	95	-1	44.0098
96	SC1	48	PAA	BB	96	0	27.0458
97	SQa	49	PAA	COO	97	-1	44.0098
98	SC1	49	PAA	BB	98	0	27.0458

99	SQa	50	PAA	COO	99	-1	44.0098
100	SC1	50	PAA	BB	100	0	27.0458
101	SQa	51	PAA	COO	101	-1	44.0098
102	SC1	51	PAA	BB	102	0	27.0458
103	SQa	52	PAA	COO	103	-1	44.0098
104	SC1	52	PAA	BB	104	0	27.0458
105	SQa	53	PAA	COO	105	-1	44.0098
106	SC1	53	PAA	BB	106	0	27.0458
107	SQa	54	PAA	COO	107	-1	44.0098
108	SC1	54	PAA	BB	108	0	27.0458
109	SQa	55	PAA	COO	109	-1	44.0098
110	SC1	55	PAA	BB	110	0	27.0458
111	SQa	56	PAA	COO	111	-1	44.0098
112	SC1	56	PAA	BB	112	0	27.0458
113	SQa	57	PAA	COO	113	-1	44.0098
114	SC1	57	PAA	BB	114	0	27.0458
115	SQa	58	PAA	COO	115	-1	44.0098
116	SC1	58	PAA	BB	116	0	27.0458
117	SQa	59	PAA	COO	117	-1	44.0098
118	SC1	59	PAA	BB	118	0	27.0458
119	SQa	60	PAA	COO	119	-1	44.0098
120	SC1	60	PAA	BB	120	0	27.0458
121	SQa	61	PAA	COO	121	-1	44.0098
122	SC1	61	PAA	BB	122	0	34.5633
123	SQa	62	PAT	COO	123	-1	44.0098

[bonds]

;i	j	funct	length	force.c.
1	2	1	0.298	46000
2	3	1	0.273	35000
3	4	1	0.273	35000
4	5	1	0.273	35000
5	6	1	0.273	35000
6	7	1	0.273	35000
7	8	1	0.273	35000

8	9	1	0.273	35000
9	10	1	0.273	35000
10	11	1	0.273	35000
11	12	1	0.273	35000
12	13	1	0.273	35000
13	14	1	0.273	35000
14	15	1	0.273	35000
15	16	1	0.273	35000
16	17	1	0.273	35000
17	18	1	0.273	35000
18	19	1	0.273	35000
19	20	1	0.273	35000
20	21	1	0.273	35000
21	22	1	0.273	35000
22	23	1	0.273	35000
23	24	1	0.273	35000
24	25	1	0.273	35000
25	26	1	0.273	35000
26	27	1	0.273	35000
27	28	1	0.273	35000
28	29	1	0.273	35000
29	30	1	0.273	35000
30	31	1	0.273	35000
31	32	1	0.273	35000
32	33	1	0.273	35000
33	34	1	0.273	35000
34	35	1	0.273	35000
35	36	1	0.273	35000
36	37	1	0.273	35000
37	38	1	0.273	35000
38	39	1	0.273	35000
39	40	1	0.273	35000
40	41	1	0.273	35000
41	42	1	0.273	35000
42	43	1	0.273	35000

43	44	1	0.273	35000
44	45	1	0.273	35000
45	46	1	0.273	35000
46	47	1	0.273	35000
47	48	1	0.273	35000
48	49	1	0.273	35000
49	50	1	0.273	35000
50	51	1	0.273	35000
51	52	1	0.273	35000
52	53	1	0.273	35000
53	54	1	0.273	35000
54	55	1	0.273	35000
55	56	1	0.273	35000
56	57	1	0.273	35000
57	58	1	0.273	35000
58	59	1	0.273	35000
59	60	1	0.273	35000
60	61	1	0.273	35000
61	62	1	0.273	35000
62	63	1	0.273	35000
63	64	1	0.273	35000
64	65	1	0.273	35000
65	66	1	0.273	35000
66	67	1	0.273	35000
67	68	1	0.273	35000
68	69	1	0.273	35000
69	70	1	0.273	35000
70	71	1	0.273	35000
71	72	1	0.273	35000
72	73	1	0.273	35000
73	74	1	0.273	35000
74	75	1	0.273	35000
75	76	1	0.273	35000
76	77	1	0.273	35000
77	78	1	0.273	35000

78	79	1	0.273	35000
79	80	1	0.273	35000
80	81	1	0.273	35000
81	82	1	0.273	35000
82	83	1	0.273	35000
83	84	1	0.273	35000
84	85	1	0.273	35000
85	86	1	0.273	35000
86	87	1	0.273	35000
87	88	1	0.273	35000
88	89	1	0.273	35000
89	90	1	0.273	35000
90	91	1	0.273	35000
91	92	1	0.273	35000
92	93	1	0.273	35000
93	94	1	0.273	35000
94	95	1	0.273	35000
95	96	1	0.273	35000
96	97	1	0.273	35000
97	98	1	0.273	35000
98	99	1	0.273	35000
99	100	1	0.273	35000
100	101	1	0.273	35000
101	102	1	0.273	35000
102	103	1	0.273	35000
103	104	1	0.273	35000
104	105	1	0.273	35000
105	106	1	0.273	35000
106	107	1	0.273	35000
107	108	1	0.273	35000
108	109	1	0.273	35000
109	110	1	0.273	35000
110	111	1	0.273	35000
111	112	1	0.273	35000
112	113	1	0.273	35000

113	114	1	0.273	35000
114	115	1	0.273	35000
115	116	1	0.273	35000
116	117	1	0.273	35000
117	118	1	0.273	35000
118	119	1	0.273	35000
119	120	1	0.273	35000
120	121	1	0.273	35000
121	122	1	0.273	35000
122	123	1	0.298	46000

[angles]

;i	j	k	funct	angle	force.c.
1	2	3	1	133	210
2	3	4	1	43	250
3	4	5	1	142	70
4	5	6	1	43	250
5	6	7	1	142	70
6	7	8	1	43	250
7	8	9	1	142	70
8	9	10	1	43	250
9	10	11	1	142	70
10	11	12	1	43	250
11	12	13	1	142	70
12	13	14	1	43	250
13	14	15	1	142	70
14	15	16	1	43	250
15	16	17	1	142	70
16	17	18	1	43	250
17	18	19	1	142	70
18	19	20	1	43	250
19	20	21	1	142	70
20	21	22	1	43	250
21	22	23	1	142	70
22	23	24	1	43	250

23	24	25	1	142	70
24	25	26	1	43	250
25	26	27	1	142	70
26	27	28	1	43	250
27	28	29	1	142	70
28	29	30	1	43	250
29	30	31	1	142	70
30	31	32	1	43	250
31	32	33	1	142	70
32	33	34	1	43	250
33	34	35	1	142	70
34	35	36	1	43	250
35	36	37	1	142	70
36	37	38	1	43	250
37	38	39	1	142	70
38	39	40	1	43	250
39	40	41	1	142	70
40	41	42	1	43	250
41	42	43	1	142	70
42	43	44	1	43	250
43	44	45	1	142	70
44	45	46	1	43	250
45	46	47	1	142	70
46	47	48	1	43	250
47	48	49	1	142	70
48	49	50	1	43	250
49	50	51	1	142	70
50	51	52	1	43	250
51	52	53	1	142	70
52	53	54	1	43	250
53	54	55	1	142	70
54	55	56	1	43	250
55	56	57	1	142	70
56	57	58	1	43	250
57	58	59	1	142	70

58	59	60	1	43	250
59	60	61	1	142	70
60	61	62	1	43	250
61	62	63	1	142	70
62	63	64	1	43	250
63	64	65	1	142	70
64	65	66	1	43	250
65	66	67	1	142	70
66	67	68	1	43	250
67	68	69	1	142	70
68	69	70	1	43	250
69	70	71	1	142	70
70	71	72	1	43	250
71	72	73	1	142	70
72	73	74	1	43	250
73	74	75	1	142	70
74	75	76	1	43	250
75	76	77	1	142	70
76	77	78	1	43	250
77	78	79	1	142	70
78	79	80	1	43	250
79	80	81	1	142	70
80	81	82	1	43	250
81	82	83	1	142	70
82	83	84	1	43	250
83	84	85	1	142	70
84	85	86	1	43	250
85	86	87	1	142	70
86	87	88	1	43	250
87	88	89	1	142	70
88	89	90	1	43	250
89	90	91	1	142	70
90	91	92	1	43	250
91	92	93	1	142	70
92	93	94	1	43	250

93	94	95	1	142	70
94	95	96	1	43	250
95	96	97	1	142	70
96	97	98	1	43	250
97	98	99	1	142	70
98	99	100	1	43	250
99	100	101	1	142	70
100	101	102	1	43	250
101	102	103	1	142	70
102	103	104	1	43	250
103	104	105	1	142	70
104	105	106	1	43	250
105	106	107	1	142	70
106	107	108	1	43	250
107	108	109	1	142	70
108	109	110	1	43	250
109	110	111	1	142	70
110	111	112	1	43	250
111	112	113	1	142	70
112	113	114	1	43	250
113	114	115	1	142	70
114	115	116	1	43	250
115	116	117	1	142	70
116	117	118	1	43	250
117	118	119	1	142	70
118	119	120	1	43	250
119	120	121	1	142	70
120	121	122	1	43	250
121	122	123	1	133	210

9.10 IONS.MDP FILE

; ions.mdp - used as input into grompp to generate ions.tpr

; Parameters describing what to do, when to stop and what to save

integrator = steep ; Algorithm (steep = steepest descent minimization)

emtol = 1000.0 ; Stop minimization when the maximum force < 1000.0 kJ/mol/nm

```
emstep      = 0.01        ; Energy step size
nsteps      = 50000       ; Maximum number of (minimization) steps to perform
;
; Parameters describing how to find the neighbors of each atom and how to calculate the interactions
nstlist     = 10          ; Frequency to update the neighbour list and long-range forces
cutoff-scheme   = Verlet
ns_type     = grid        ; Method to determine neighbour list (simple, grid)
coulombtype     = PME     ; Treatment of long-range electrostatic interactions
rcoulomb    = 1.0         ; Short-range electrostatic cut-off
rvdw        = 1.0         ; Short-range Van der Waals cut-off
pbc         = xyz         ; Periodic Boundary Conditions (yes/no)
```

9.11 ALL ATOMISTIC ENERGY MINIMIZATION.MDP FILE

```
; minim.mdp - used as input into grompp to generate em.tpr
;
; Run parameters
Integrator  = steep             ; Algorithm (steep = steepest descent minimization)
tinit = 0
dt = 0.002
init-step = 0
comm-mode = Linear
nstcomm = 100
comm-grps = system
emtol   = 1000.0  ; Stop minimization when the maximum force < 1000.0 kJ/mol/nm
emstep = 0.002          ; Energy step size
nsteps   = 50000                ; Maximum number of (minimization) steps to perform
;
; Output control
Nstxout = 500           ; save coordinates every 1 ps
Nstvout = 500           ; save velocities every 1 ps
nstenergy  = 500                ; save energies every 1 ps
nstxtcout = 500                 ; co-ordinates to xtc
nstlog  = 500                   ; update log file every 1.0 ps
nstfout  = 500
nstcalcenergy = 100     ; calculate energies
nstcheckpoint = 1000
xtc-precision = 1000
xtc_grps = system
;
; LANGEVIN DYNAMICS OPTIONS
; Friction coefficient (amu/ps) and random seed
bd-fric = 0
ld-seed  = -1
;
; Max number of iterations in relax_shells
niter  = 20
;
```

```
; NEIGHBOR SEARCHING
cutoff-scheme  = Verlet
ns-type = grid                  ; Method to determine neighbor list (simple, grid)
nstlist  = 10               ; Frequency to update the neighbor list and long-range forces
rlist   = 1
rvdw    = 1.0                   ; Short-range Van der Waals cut-off
verlet-buffer-tolerance  = 0.005
pbc = xyz                       ; Periodic Boundary Conditions (yes/no)
periodic_molecules  = no
;
; ELECTROSTATICS
coulombtype     = PME           ; Treatment of long-range electrostatic interactions
pme-order  = 8
rcoulomb = 1.2                  ; Short-range electrostatic cut-off
rcoulomb-switch  = 0
epsilon-r  = 1                  ; 2.5 (with polarizable water)
epsilon-rf  = 1
;
; VAN DER WAALS
vdwtype  = Cut-off
vdw-modifier  = Potential-shift-verlet
;
; CUT-OFF LENGHTS
rvdw-switch   = 0
;
; Ewald
DispCorr = EnerPres ;Apply long range dispersion corrections for Energy and Pressure
Table-extension  = 1
energygrp-Table  =
fourierspacing = 0.15
fourier-nx  = 0
fourier-ny  = 0
fourier-nz   = 0
ewald-rtol  = 1e-05
ewald-rtol-lj  = 1e-3
```

```
ewald-geometry  = 3d
lj-pme-comb-rule = Geometric
epsilon-surface  = 0
optimize_fft  = yes
;
; OPTIONS FOR WEAK COUPLING ALGORITHMS
; Temperature coupling
tcoupl  = No
tc-grps   =
tau-t   =
ref-t   =
pcoupl  = No
pcoupltype   = Isotropic
tau-p  = 1
ref-p   =
refcoord_scaling   = No
andersen_seed  = 815131
;
; GENERATE VELOCITIES FOR STARTUP RUN
gen_vel  = no
gen-temp = 300
gen_seed   = -1                    ;173529
;
; OPTIONS FOR BONDS
constraints   = none
constraint-algorithm  = LINCS
continuation  = no
Shake-SOR   = no
shake-tol  = 0.0001
lincs-order = 8
lincs-iter  = 1
lincs-warnangle = 30
morse = no
```

9.12 ALL ATOMISTIC NVT.MDP FILE

```
Title = Thermorun
define  = -DPOSRES  ; position restrain the protein
;
; Run parameters
Integrator = md-vv              ; velocity verlet
nsteps   = 500000               ; 2 * 500000 = 1 ns
dt = 0.002                      ; 2 fs
init_step   = 0
energygrps   = Polymer Ion SOL
;
; Output control
nstxout                 = 500                   ; save coordinates every 1.0 ps
nstvout                 = 500                   ; save velocities every 1.0 ps
nstenergy               = 500                   ; save energies every 1.0 ps
nstxtcout               = 500           ; co-ordinates to xtc
nstlog                  = 500                   ; update log file every 1.0 ps
nstfout                 = 500
nstcalcenergy           = 500           ; calculate energies
nstcheckpoint           = 1000
xtc-precision           = 1000
xtc_grps                = system
;
; Neighbor searching
cutoff-scheme       = Verlet
ns_type                         = grid                  ; search neighboring grid cells
nstlist                 = 10                    ; 20 fs, largely irrelevant with Verlet

rlist           = 1.0           ; in nm
;
; Electrostatics
coulombtype             = PME           ; Particle Mesh Ewald for long-range electrostatics
pme-order               = 8                     ; 4 is cubic interpolation
rcoulomb-switch     = 0
rcoulomb            = 1.2                               ; short-range electrostatic cutoff (in
nm); With Verlet rcoulomb!=rvdw is not supported ; For PME rcouloumb>rvdw is possible
```

```
epsilon_r          = 1.0
;
; Vanderwaals
vdwtype            = Cut-off
rvdw-switch        = 0
rvdw               = 1.0            ; short-range van der Waals cutoff (in nm)
DispCorr           = EnerPres               ; account for cut-off vdW scheme
;
; Ewald
fourierspacing     = 0.15
fourier_nx         = 0
fourier_ny         = 0
fourier_nz         = 0
ewald_rtol         = 1e-5
ewald-rtol-lj      = 1e-3
ewald_geometry     = 3d
lj-pme-comb-rule   = Geometric
epsilon_surface    = 0
optimize_fft       = yes
;
; Bond parameters
constraints              = all-bonds    ; all bonds (even heavy atom-H bonds) constrained
constraint_algorithm     = LINCS        ; holonomic constraints
continuation             = no           ; first dynamics run
lincs_iter               = 1            ; accuracy of LINCS
lincs_order              = 8            ; also related to accuracy
lincs_warnangle          = 30
shake_tol                = 0.0001
morse                    = no
;
; Temperature coupling is on
tcoupl                   = V-rescale    ; modified Berendsen thermostat
tc-grps                  = Polymer Ion SOL   ; two coupling groups - more accurate
nsttcouple               = 1                          ; frequency for coupling temperature
tau_t                    = 0.1 0.1 0.1       ; time constant, in ps
```

```
ref_t                    = 300 300 300       ; reference temperature, one for each group, in K
;
; Pressure coupling is off
pcoupl                   = no                ; no pressure coupling in NVT
;
; Periodic boundary conditions
pbc                      = xyz               ; 3-D PBC
;
; Velocity generation
gen_vel                  = yes               ; assign velocities from Maxwell distribution
gen_temp                 = 300               ; temperature for Maxwell distribution
gen_seed                 = -1                ; generate a random seed
```

9.13 ALL ATOMISTIC NPT.MDP FILE

```
title               = Thermorun
define              = -DPOSRES  ; position restrain the protein
;
; Run parameters
integrator          = md-vv             ; velocity verlet
nsteps              = 500000            ; 2 * 500000 = 1 ns
dt                  = 0.002             ; 2 fs
init_step           = 0
energygrps          = Polymer Ion SOL
;
; Output control
nstxout                     = 500       ; save coordinates every 1 ps
nstvout                     = 500       ; save velocities every 1 ps
nstenergy                   = 500       ; save energies every 1 ps
nstxtcout                   = 500       ; co-ordinates to xtc
nstlog                      = 500       ; update log file every 1.0 ps
nstfout                     = 500
nstcalcenergy               = 500       ; calculate energies
nstcheckpoint               = 1000
xtc-precision               = 1000
xtc_grps                    = system
;
; Neighbor searching
cutoff-scheme               = Verlet
ns_type                     = grid      ; search neighboring grid cells
nstlist                     = 10        ; 20 fs, largely irrelevant with Verlet scheme
rlist                       = 1.0       ; in nm
;
; Electrostatics
coulombtype                 = PME       ; Particle Mesh Ewald for long-range electrostatics
pme-order                   = 8         ; 4 is cubic interpolation
rcoulomb                    = 1.2       ; short-range electrostatic cutoff (in nm)
```

```
rcoulomb-switch         = 0
epsilon_r               = 1
;
; VANDERWAALS
vdwtype                 = Cut-off
rvdw-switch             = 0
rvdw                    = 1.0           ; short-range van der Waals cutoff (in nm)
DispCorr                = EnerPres      ; account for cut-off vdW scheme
;
; EWALD
fourierspacing          = 0.15          ;grid spacing for FFT (0.15 is
reccommended according to manual)
fourier_nx              = 0
fourier_ny              = 0
fourier_nz              = 0
ewald_rtol              = 1e-5          ;relative strength of Ewald-shifted
direct potential at rcoulomb
ewald-rtol-lj           = 1e-3
ewald_geometry          = 3d
lj-pme-comb-rule        = Geometric
epsilon_surface         = 0
optimize_fft            = yes
;
; Bond parameters
constraints             = all-bonds     ; all bonds (even heavy atom-H bonds)
constrained
continuation            = yes           ; Restarting after NVT
constraint_algorithm    = lINCS         ; holonomic constraints
lincs_iter              = 1             ; accuracy of LINCS
lincs_order             = 8             ; also related to accuracy
lincs_warnangle         = 30
shake_tol               = 0.0001
morse                   = no
;
; Temperature coupling is on
```

```
tcoupl                  = V-rescale             ; modified Berendsen thermostat
tc-grps                 = Polymer Ion SOL       ; two coupling groups - more accurate
nsttcouple              = 1                     ; frequency for coupling temperature
tau_t                   = 0.1 0.1 0.1           ; time constant, in ps. For production with NH, use 2 ps
ref_t                   = 300 300 300           ; reference temperature, one for each group, in K
;
; Pressure coupling is on
pcoupl                  = Berendsen             ; Pressure coupling on in NPT
pcoupltype              = isotropic             ; uniform scaling of box vectors
tau_p                   = 2.0                   ; time constant, in ps
ref_p                   = 1.0                   ; reference pressure, in bar
compressibility         = 4.5e-5                ; isothermal compressibility of water, bar^-1
refcoord_scaling        = com
;
; Periodic boundary conditions
pbc                     = xyz                   ; 3-D PBC
;
;
; Velocity generation
gen-vel                 = no                    ; Velocity generation is off
```

9.14 ALL ATOMISTIC .MDP FILE

```
title                    = 20-mer of fully deprotonated PAA with Na at 15dH
; Run parameters
integrator               = md                    ; leap frog
nsteps                   = 50000000              ; 2 * 50000000 = 100 ns
dt                       = 0.002                 ; 2 fs
; Output control
nstxout                  = 100000                ; save coordinates every 200 ps
nstvout                  = 100000                ; save velocities every 200 ps
nstenergy                = 100000                ; save energies every 200 ps
nstlog                   = 100000                ; update log file every 200 ps
nstxout-compressed       = 100000                ; save compressed coordinates every 200 ps
                                                 ;
nstxout-compressed replaces nstxtcout
compressed-x-grps        = System                ; replaces xtc-grps
energygrps               = Polymer Ion SOL
;
; Bond parameters
continuation             = no                    ; Restarting after NPT
constraint_algorithm     = lincs                 ; holonomic constraints
constraints              = all-bonds             ; all bonds (even heavy atom-H bonds) constrained
lincs_iter               = 1                     ; accuracy of LINCS
lincs_order              = 4                     ; also related to accuracy
;
; Neighborsearching
cutoff-scheme            = Verlet
ns_type                  = grid                  ; search neighboring grid cells
nstlist                  = 10                    ; 20 fs, largely irrelevant with Verlet scheme
rcoulomb                 = 1.2                   ; short-range electrostatic cutoff (in nm)
rvdw                     = 1.0                   ; short-range van der Waals cutoff (in nm)
;
```

```
; Electrostatics
coulombtype             = PME           ; Particle Mesh Ewald for long-range electrostatics
pme_order               = 4             ; cubic interpolation
fourierspacing          = 0.15          ; grid spacing for FFT
;
; Temperature coupling is on
tcoupl                  = V-rescale     ; modified Berendsen thermostat
tc-grps                 = Polymer Ion SOL   ; two coupling groups - more accurate
nsttcouple              = 1             ; frequency for coupling temperature
tau_t                   = 0.1 0.1 0.1   ; time constant, in ps. For production with NH, use 2 ps
ref_t                   = 300 300 300   ; reference temperature, one for each group, in K
;
; Pressure coupling is on
pcoupl                  = Parrinello-Rahman     ; Pressure coupling on in NPT
pcoupltype              = isotropic             ; uniform scaling of box vectors
tau_p                   = 2.0                   ; time constant, in ps
ref_p                   = 1.0                   ; reference pressure, in bar
compressibility         = 4.5e-5                ; isothermal compressibility of water, bar^-1
;
; Periodic boundary conditions
pbc                     = xyz           ; 3-D PBC
;
; Dispersion correction
DispCorr                = EnerPres      ; account for cut-off vdW scheme
;
; Velocity generation
gen_vel                 = yes           ; Velocity generation is off
gen-seed                = -1            ; pseudo random seed
```

9.15 COARSE GRAINED ENERGY MINIMISATION.MDP FILE

```
; minim.mdp - used as input into grompp to generate em.tpr
;
; Run parameters
integrator              = steep             ; Algorithm (steep = steepest descent minimization)
tinit                   = 0
dt                      = 0.002
init-step               = 0
comm-mode               = Linear
nstcomm                 = 100
comm-grps               = System
emtol                   = 10                ; Stop minimization when the maximum force < 10.0 kJ/mol/nm
emstep                  = 0.002       ; Energy step size
nsteps                  = 500000            ; Maximum number of (minimization) steps to perform
;
; Output control
nstxout             = 500           ; save coordinates every 1 ps
nstvout             = 500           ; save velocities every 1 ps
nstenergy           = 500           ; save energies every 1 ps
nstxtcout     = 500     ; co-ordinates to xtc
nstlog              = 500           ; update log file every 1.0 ps
nstfout       = 500
nstcalcenergy = 100     ; calculate energies
nstcheckpoint = 1000
xtc-precision = 1000
xtc_grps      = System
;
; LANGEVIN DYNAMICS OPTIONS
; Friction coefficient (amu/ps) and random seed
bd-fric                 = 0
ld-seed                 = -1
;
; Max number of iterations in relax_shells
```

```
niter                   = 20
;
; NEIGHBOR SEARCHING
cutoff-scheme                   = Verlet
ns-type                 = grid          ; Method to determine neighbor list
(simple, grid)
nstlist                 = 10            ; Frequency to update the neighbor list and
long-range forces
rlist                   = 1
rvdw                            = 1.0           ; Short-range Van der Waals
cut-off
verlet-buffer-tolerance  = 0.005
pbc                             = xyz           ; Periodic Boundary
Conditions (yes/no)
periodic_molecules      = no
;
; ELECTROSTATICS
coulombtype                     = PME                   ; Treatment of long-
range electrostatic interactions
pme-order               = 8
rcoulomb                        = 1.2           ; Short-range electrostatic cut-
off
rcoulomb-switch         = 0
epsilon-r               = 2.5                   ; 2.5 (with polarizable water)
epsilon-rf              = 1
;
; VAN DER WAALS
vdwtype                 = Cut-off
vdw-modifier            = Potential-shift-verlet
;
; CUT-OFF LENGHTS
rvdw-switch             = 0
;
; Ewald
DispCorr                = EnerPres              ;Apply long range dispersion corrections for
Energy and Pressure
Table-extension         = 1
```

energygrp-**Table** =
fourierspacing = 0.15
fourier-nx = 0
fourier-ny = 0
fourier-nz = 0
ewald-rtol = 1e-05
ewald-rtol-lj = 1e-3
ewald-geometry = 3d
lj-pme-comb-rule = Geometric
epsilon-surface = 0
optimize_fft = yes
;
; OPTIONS FOR WEAK COUPLING ALGORITHMS
; Temperature coupling
tcoupl = No
tc-grps =
tau-t =
ref-t =
pcoupl = No
pcoupltype = Isotropic
tau-p = 4.0
ref-p =
refcoord_scaling = No
andersen_seed = 815131
;
; GENERATE VELOCITIES FOR STARTUP RUN
gen_vel = no
gen-temp = 300
gen_seed = -1 ;173529
;
; OPTIONS FOR BONDS
constraints = none
constraint-algorithm = LINCS
continuation = no
Shake-SOR = no

```
shake-tol              = 0.0001
lincs-order            = 8
lincs-iter             = 1
lincs-warnangle        = 90
morse                  = no
```

9.16 COARSE GRAINED NVT.MDP FILE

```
title               = Thermorun
define              = -DPOSRES  ; position restrain the protein
;
; Run parameters
integrator              = md-vv                 ; velocity verlet
nsteps                  = 25000000                  ; 2 * 25000000 = 50 ns
dt                          = 0.002                         ; 2 fs
init_step           = 0
energygrps          = System
;
; Output control
nstxout                 = 500                   ; save coordinates every 1.0 ps
nstvout                 = 500                   ; save velocities every 1.0 ps
nstenergy               = 500                   ; save energies every 1.0 ps
nstxtcout           = 500           ; co-ordinates to xtc
nstlog                  = 500                   ; update log file every 1.0 ps
nstfout             = 500
nstcalcenergy       = 500           ; calculate energies
nstcheckpoint       = 1000
xtc-precision       = 1000
xtc_grps            = System
;
; Neighbor searching
cutoff-scheme       = Verlet
ns_type                     = grid                  ; search neighboring grid cells
nstlist                 = 10                    ; 20 fs, largely irrelevant with Verlet
rlist               = 1.0           ; in nm
verlet-buffer-tolerance  = 0.005
```

```
;
; Electrostatics
coulombtype          = PME          ; Particle Mesh Ewald for long-range electrostatics
pme-order            = 8            ; 4 is cubic interpolation
rcoulomb-switch      = 0
rcoulomb             = 1.2          ; short-range electrostatic cutoff (in nm); With Verlet rcoulomb!=rvdw is not supported ; For PME rcouloumb>rvdw is possible
epsilon_r            = 2.5
;
; Vanderwaals
vdwtype              = Cut-off
vdw-modifier         = Potential-shift-verlet
rvdw-switch          = 0
rvdw                 = 1.0          ; short-range van der Waals cutoff (in nm)
DispCorr             = EnerPres     ; account for cut-off vdW scheme
;
; Ewald
fourierspacing       = 0.15
fourier_nx           = 0
fourier_ny           = 0
fourier_nz           = 0
ewald_rtol           = 1e-5
ewald-rtol-lj        = 1e-3
ewald_geometry       = 3d
lj-pme-comb-rule     = Geometric
epsilon_surface      = 0
optimize_fft         = yes
;
; Bond parameters
constraints          = none         ; all bonds (even heavy atom-H bonds) constrained
constraint_algorithm = LINCS        ; holonomic constraints
continuation         = no           ; first dynamics run
lincs_iter           = 1            ; accuracy of LINCS
lincs_order          = 8            ; also related to accuracy
```

```
lincs_warnangle         = 90
shake_tol               = 0.0001
morse                   = no
;
; Temperature coupling is on
tcoupl                  = V-rescale       ; modified Berendsen thermostat
tc-grps                 = System                  ; two coupling groups - more accurate
nsttcouple              = 1                       ; frequency for coupling temperature
tau_t                   = 0.1                ; time constant, in ps
ref_t                   = 300                ; reference temperature, one for each group, in
K
;
; Pressure coupling is off
pcoupl                  = no                 ; no pressure coupling in NVT
;
; Periodic boundary conditions
pbc                     = xyz                     ; 3-D PBC
;
; Velocity generation
gen_vel                 = yes                ; assign velocities from Maxwell
distribution
gen_temp                = 300                ; temperature for Maxwell distribution
gen_seed                = -1                 ; generate a random seed
```

9.17 COARSE GRAINED NPT.MDP FILE

```
title                   = Barostat
define                  = -DPOSRES  ; position restrain the protein
;
; Run parameters
integrator              = md-vv                   ; verlet velocity
nsteps                  = 25000000          ; 2 * 25000000 = 50 ns
dt                      = 0.002                   ; 2 fs
init_step               = 0
energygrps              = System
;
; Output control
```

```
nstxout                         = 500           ; save coordinates every 1 ps
nstvout                         = 500           ; save velocities every 1 ps
nstenergy                       = 500           ; save energies every 1 ps
nstxtcout                       = 500           ; co-ordinates to xtc
nstlog                          = 500           ; update log file every 1.0 ps
nstfout                         = 500
nstcalcenergy                   = 500           ; calculate energies
nstcheckpoint                   = 1000
xtc-precision                   = 1000
xtc_grps                        = System
;
; Neighbor searching
cutoff-scheme                   = Verlet
ns_type                         = grid          ; search neighboring
grid cells
nstlist                         = 10            ; 20 fs, largely irrelevant with Verlet
scheme
rlist                   = 1.0           ; in nm
verlet-buffer-tolerance  = 0.005
;
; Electrostatics
coulombtype                     = PME           ; Particle Mesh Ewald for
long-range electrostatics
pme-order                       = 8             ; 4 is cubic interpolation
rcoulomb                = 1.2           ; short-range electrostatic cutoff (in nm)
rcoulomb-switch                 = 0
epsilon_r               = 2.5
;
; VANDERWAALS
vdwtype                 = Cut-off
vdw-modifier            = Potential-shift-verlet
rvdw-switch                     = 0
rvdw                    = 1.0           ; short-range van der Waals cutoff (in nm)
DispCorr                        = EnerPres      ; account for cut-off vdW scheme
;
; EWALD
```

```
fourierspacing          = 0.15          ;grid spacing for FFT (0.15 is
reccommended according to manual)
fourier_nx              = 0
fourier_ny              = 0
fourier_nz              = 0
ewald_rtol              = 1e-5          ;relative strength of Ewald-shifted
direct potential at rcoulomb
ewald-rtol-lj           = 1e-3
ewald_geometry          = 3d
lj-pme-comb-rule        = Geometric
epsilon_surface         = 0
optimize_fft            = yes
;
; Bond parameters
constraints             = none          ; all bonds (even heavy atom-H bonds)
constrained
continuation            = yes           ; Restarting after NVT
constraint_algorithm    = lINCS         ; holonomic constraints
lincs_iter              = 1             ; accuracy of LINCS
lincs_order             = 8             ; also related to accuracy
lincs_warnangle         = 90
shake_tol               = 0.0001
morse                   = no
;
; Temperature coupling is on
tcoupl                  = V-rescale     ; modified Berendsen
thermostat
tc-grps                 = System        ; two coupling
groups - more accurate
nsttcouple              = 1             ; frequency for
coupling temperature
tau_t                   = 0.1           ; time constant, in ps.
For production with NH, use 2 ps
ref_t                   = 300           ; reference temperature,
one for each group, in K
;
; Pressure coupling is on
pcoupl                  = Berendsen     ; Pressure coupling on in NPT
```

```
pcoupltype              = isotropic             ; uniform scaling of box vectors
tau_p                   = 12.0                  ; time constant, in ps
ref_p                   = 1.0                   ; reference pressure, in bar
compressibility         = 3e-4          ; isothermal compressibility of water, bar^-1
refcoord_scaling        = com
;
; Periodic boundary conditions
pbc                                     = xyz                                   ; 3-D
PBC
;
;
; Velocity generation
gen-vel                         = no                    ; Velocity generation is off
```

9.18 MARTINI .MDP FILE

```
; Run parameters
integrator              = md         ; leap-frog integrator
nsteps                  = 25000000    ; 20 * 2500000 = 500 ns
dt                      = 0.02       ; 20 fs
define                  = -DDISRES    ; distance restraints of the protein
disre                   = simple     ; Enable Distance Restraints
disre_fc                = 1000

; Output control
nstxout                 = 25000      ; save coordinates every 10 ps
nstvout                 = 25000      ; save velocities every 10 ps
nstenergy               = 25000      ; save energies every 10 ps
nstlog                  = 25000      ; update log file every 10 ps
nstxout-compressed      = 50000      ; save comp. coord. every 100 ps
compressed-x-grps       = System     ; replaces xtc-grps
energygrps              = Protein  others
; Bond parameters
continuation            = no
constraints             = none
constraint_algorithm    = Lincs
lincs_order             = 4
lincs_warnangle         = 90
; Neighborsearching
cutoff-scheme           = Verlet
nstlist                 = 10
ns_type                 = grid
verlet-buffer-tolerance = 0.005
coulombtype             = PME
pme_order               = 4          ; cubic interpolation
rcoulomb                = 1.1
epsilon_r               = 2.5        ; for polarizable water
epsilon_rf              = 0
vdw_type                = Cut-off
vdw-modifier            = Potential-shift-verlet
```

```
rvdw                = 1.1
fourierspacing      = 0.12      ; grid spacing for FFT
; Temperature coupling is on
tcoupl              = V-rescale   ; modified Berendsen thermostat
tc-grps             = Protein  others ; two coupling groups
tau_t               = 1.0    1.0  ; time constant, in ps
ref_t               = 300    300  ; reference temperature
; Pressure coupling is on
Pcoupl              = parrinello-rahman
Pcoupltype          = isotropic
tau_p               = 12.0      ;parrinello-rahman
compressibility     = 3e-4
ref_p               = 1.0
; Periodic boundary conditions
pbc                 = xyz       ; 3-D PBC
; Dispersion correction
DispCorr            = EnerPres    ; account for cut-off vdW scheme
; Velocity generation
gen_vel     = yes               ; Velocity generation is on
```

9.19 PROTEIN DISTANCE RESTRAINTS

```
#include "Protein.itp"

;include Distance restraints between calcium ions and surrounding residues
#ifdef DISRES
[ distance_restraints ]
; ai  aj  type index type' low  up1  up2 fac
705   3   1     0     1    0.0  0.2  0.3 2.0
705 104   1     1     1    0.0  0.2  0.3 3.0
705 192   1     2     1    0.0  0.2  0.3 2.0
705 203   1     3     1    0.0  0.2  0.3 1.0
706 418   1     4     1    0.0  0.2  0.5 1.0
706 423   1     5     1    0.0  0.2  0.5 1.0
706 481   1     6     1    0.0  0.2  0.5 3.0
706 485   1     7     1    0.0  0.2  0.5 3.0
#endif
```

10 References

1. Broze, G., *Handbook of Detergents: Properties*. CRC Press: 1999.
2. Madkour, T. M.; Barakat, A. M., Computer simulation of polymers. *Computational and Theoretical Polymer Science* **1997,** *7* (1), 35-46.
3. Gartner III, T. E.; Jayaraman, A., Modeling and simulations of polymers: A Roadmap. *Macromolecules* **2019,** *52* (3), 755-786.
4. Darre, L.; Iglesias-Fernandez, J.; Kohlmeyer, A.; Wacklin, H.; Domene, C., Molecular Dynamics Simulations and Neutron Reflectivity as an Effective Approach to Characterize Biological Membranes and Related Macromolecular Assemblies. *J. Chem. Theory Comput.* **2015,** *11* (10), 4875.
5. Fenter, P.; Kerisit, S.; Raiteri, P.; Gale, J. D., Is the Calcite-Water Interface Understood? Direct Comparisons of Molecular Dynamics Simulations with Specular X-Ray Reflectivity Data. *J. Phys. Chem. C* **2013,** *117* (10), 5028.
6. Schneidman-Duhovny, D.; Hammel, M.; Sali, A., Foxs: A Web Server for Rapid Computation and Fitting of Saxs Profiles. *Nucleic Acids Res.* **2010,** *38*, W540.
7. Kucerka, N.; Holland, B. W.; Gray, C. G.; Tomberli, B.; Katsaras, J., Scattering Density Profile Model of Popg Bilayers as Determined by Molecular Dynamics Simulations and Small-Angle Neutron and X-Ray Scattering Experiments. *J. Phys. Chem. B* **2012,** *116* (1), 232.
8. Chen, P. C.; Hub, J. S., Validating Solution Ensembles from Molecular Dynamics Simulation by Wide-Angle X-Ray Scattering Data. *Biophys. J.* **2014,** *107* (2), 435.
9. Knight, C. J.; Hub, J. S., Waxsis: A Web Server for the Calculation of Saxs/Waxs Curves Based on Explicit-Solvent Molecular Dynamics. *Nucleic Acids Res.* **2015,** *43* (W1), W225.
10. Jankowski, E.; Marsh, H. S.; Jayaraman, A., Computationally Linking Molecular Features of Conjugated Polymers and Fullerene Derivatives to Bulk Heterojunction Morphology. *Macromolecules* **2013,** *46* (14), 5775.
11. Miller, N. C.; Cho, E.; Junk, M. J. N.; Gysel, R.; Risko, C.; Kim, D.; Sweetnam, S.; Miller, C. E.; Richter, L. J.; Kline, R. J.; Heeney, M.; McCulloch, I.; Amassian, A.; Acevedo-Feliz, D.; Knox, C.; Hansen, M. R.; Dudenko, D.; Chmelka, B. F.; Toney, M. F.; Bredas, J. L.; McGehee, M. D., Use of X-Ray Diffraction, Molecular Simulations, and Spectroscopy to Determine the Molecular Packing in a Polymer-Fullerene Bimolecular Crystal. *Adv. Mater.* **2012,** *24* (45), 6071.
12. Doi, M.; Edwards, S. F., *The Theory of Polymer Dynamics*. 1988.
13. Gartner, T. E.; Jayaraman, A., Modeling and Simulations of Polymers: A Roadmap. *Macromolecules* **2019,** *52* (3), 755-786.
14. Jobic, H.; Theodorou, D. N., Quasi-Elastic Neutron Scattering and Molecular Dynamics Simulation as Complementary Techniques for Studying Diffusion in Zeolites. *Microporous Mesoporous Mater.* **2007,** *102* (1–3), 21.
15. Middleton, L. R.; Tarver, J. D.; Cordaro, J.; Tyagi, M.; Soles, C. L.; Frischknecht, A. L.; Winey, K. I., Heterogeneous Chain Dynamics and Aggregate Lifetimes in Precise Acid-Containing Polyethylenes: Experiments and Simulations. *Macromolecules* **2016,** *49* (23), 9176.
16. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. **2004,** *25* (9), 1157-1174.
17. Chenoweth, K.; Cheung, S.; van Duin, A. C. T.; Goddard, W. A.; Kober, E. M., Simulations on the Thermal Decomposition of a Poly(dimethylsiloxane) Polymer

Using the ReaxFF Reactive Force Field. *Journal of the American Chemical Society* **2005,** *127* (19), 7192-7202.
18. Borodin, O.; Smith, G. D.; Douglas, R., Force Field Development and MD Simulations of Poly(ethylene oxide)/LiBF4 Polymer Electrolytes. *The Journal of Physical Chemistry B* **2003,** *107* (28), 6824-6837.
19. Sun, H.; Ren, P.; Fried, J. R., The COMPASS force field: parameterization and validation for phosphazenes. *Computational and Theoretical Polymer Science* **1998,** *8* (1), 229-246.
20. Rossi, G.; Monticelli, L.; Puisto, S. R.; Vattulainen, I.; Ala-Nissila, T., Coarse-graining polymers with the MARTINI force-field: polystyrene as a benchmark case. *Soft Matter* **2011,** *7* (2), 698-708.
21. Reith, D.; Pütz, M.; Müller-Plathe, F., Deriving effective mesoscale potentials from atomistic simulations. **2003,** *24* (13), 1624-1636.
22. Elvers, B.; Hawkins, S.; Russey, W., *Ullmann's encyclopedia of industrial chemistry*. Wiley Online Library: 1989.
23. Adamczyk, Z.; Bratek, A.; Jachimska, B.; Jasinski, T.; Warszynski, P., Structure of poly (acrylic acid) in electrolyte solutions determined from simulations and viscosity measurements. *The Journal of Physical Chemistry B* **2006,** *110* (45), 22426-22435.
24. Biermann, O., *Molecular Dynamics Simulation Study of Polyelectrolyte Adsorption on Cellulose Surfaces*. 2002.
25. Biermann, O.; Haedicke, E.; Koltzenburg, S.; Seufert, M.; Mueller-Plathe, F., Hydration of polyelectrolytes studied by molecular dynamics simulation. *arXiv preprint cond-mat/0101115* **2001**.
26. Katiyar, R. S.; Jha, P. K., Phase behavior of aqueous polyacrylic acid solutions using atomistic molecular dynamics simulations of model oligomers. *Polymer* **2017,** *114*, 266-276.
27. Laguecir, A.; Ulrich, S.; Labille, J.; Fatin-Rouge, N.; Stoll, S.; Buffle, J., Size and pH effect on electrical and conformational behavior of poly (acrylic acid): Simulation and experiment. *European polymer journal* **2006,** *42* (5), 1135-1144.
28. Reith, D.; Müller, B.; Müller-Plathe, F.; Wiegand, S., How does the chain extension of poly (acrylic acid) scale in aqueous solution? A combined study with light scattering and computer simulation. *The Journal of Chemical Physics* **2002,** *116* (20), 9100-9106.
29. Schweins, R.; Hollmann, J.; Huber, K., Dilute solution behaviour of sodium polyacrylate chains in aqueous NaCl solutions. *Polymer* **2003,** *44* (23), 7131-7141.
30. Sulatha, M. S.; Natarajan, U., Origin of the Difference in Structural Behavior of Poly(acrylic acid) and Poly(methacrylic acid) in Aqueous Solution Discerned by Explicit-Solvent Explicit-Ion MD Simulations. *Industrial & Engineering Chemistry Research* **2011,** *50* (21), 11785-11796.
31. Bulo, R. E.; Donadio, D.; Laio, A.; Molnar, F.; Rieger, J.; Parrinello, M., “Site Binding” of Ca2+ Ions to Polyacrylates in Water: A Molecular Dynamics Study of Coiling and Aggregation. *Macromolecules* **2007,** *40* (9), 3437-3442.
32. Huber, K., Calcium-induced shrinking of polyacrylate chains in aqueous solution. *The Journal of Physical Chemistry* **1993,** *97* (38), 9825-9830.
33. Koda, S.; Nomura, H.; Nagasawa, M., Raman spectroscopic studies on the interaction between divalent counterion and polyion. *Biophysical chemistry* **1983,** *18* (4), 361-367.

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

34. Lages, S.; Lindner, P.; Sinha, P.; Kiriy, A.; Stamm, M.; Huber, K., Formation of Ca2+-induced intermediate necklace structures of polyacrylate chains. *Macromolecules* **2009,** *42* (12), 4288-4299.
35. Lages, S.; Michels, R.; Huber, K., Coil-Collapse and Coil-Aggregation due to the Interaction of Cu2+ and Ca2+ Ions with Anionic Polyacrylate Chains in Dilute Solution. *Macromolecules* **2010,** *43* (6), 3027-3035.
36. Molnar, F.; Rieger, J., "Like-charge attraction" between anionic polyelectrolytes: molecular dynamics simulations. *Langmuir* **2005,** *21* (2), 786-789.
37. Schweins, R.; Goerigk, G.; Huber, K., Shrinking of anionic polyacrylate coils induced by Ca 2+, Sr 2+ and Ba 2+: A combined light scattering and ASAXS study. *The European Physical Journal E: Soft Matter and Biological Physics* **2006,** *21* (2), 99-110.
38. Schweins, R.; Huber, K., Collapse of sodium polyacrylate chains in calcium salt solutions. *The European Physical Journal E* **2001,** *5* (1), 117-126.
39. Schweins, R.; Lindner, P.; Huber, K., Calcium induced shrinking of NaPA chains: A SANS investigation of single chain behavior. *Macromolecules* **2003,** *36* (25), 9564-9573.
40. Tong, K.; Song, X.; Sun, S.; Xu, Y.; Yu, J., Molecular dynamics study of linear and comb-like polyelectrolytes in aqueous solution: effect of Ca2+ ions. *Molecular Physics* **2014,** *112* (16), 2176-2183.
41. Lages, S.; Goerigk, G. n.; Huber, K., SAXS and ASAXS on dilute sodium polyacrylate chains decorated with lead ions. *Macromolecules* **2013,** *46* (9), 3570-3580.
42. Ezhova, A.; Huber, K., Specific Interactions of Ag+ Ions with Anionic Polyacrylate Chains in Dilute Solution. *Macromolecules* **2014,** *47* (22), 8002-8011.
43. Goerigk, G.; Schweins, R.; Huber, K.; Ballauff, M., The distribution of Sr2+ counterions around polyacrylate chains analyzed by anomalous small-angle X-ray scattering. *EPL (Europhysics Letters)* **2004,** *66* (3), 331.
44. Koda, S.; Nomura, H.; Nagasawa, M., Raman spectroscopic studies on the interaction between counterion and polyion. *Biophysical chemistry* **1982,** *15* (1), 65-72.
45. Tsukida, N.; Muranaka, H.; Ide, M.; Maeda, Y.; Kitano, H., Effect of Neutralization of Poly (acrylic acid) on the Structure of Water Examined by Raman Spectroscopy. *The Journal of Physical Chemistry B* **1997,** *101* (34), 6676-6679.
46. Vao-soongnern, V.; Merat, K.; Horpibulsuk, S., Interaction of the calcium ion with poly (acrylic acid) as investigated by a combination of molecular dynamics simulation and X-ray absorption spectroscopy. *Journal of Polymer Research* **2016,** *23* (1), 7.
47. Reith, D.; Meyer, H.; Müller-Plathe, F., Mapping atomistic to coarse-grained polymer models using automatic simplex optimization to fit structural properties. *Macromolecules* **2001,** *34* (7), 2335-2345.
48. Tjipangandjara, K. F.; Somasundaran, P., Effects of the conformation of polyacrylic acid on the dispersion-flocculation of alumina and kaolinite fines. *Advanced Powder Technology* **1992,** *3* (2), 119-127.
49. Bajaj, I.; Singhal, R., Poly (glutamic acid) – An emerging biopolymer of commercial interest. *Bioresource Technology* **2011,** *102* (10), 5551-5561.
50. Khalil, I. R.; Burns, A. T. H.; Radecka, I.; Kowalczuk, M.; Khalaf, T.; Adamus, G.; Johnston, B.; Khechara, M. P., Bacterial-Derived Polymer Poly-y-Glutamic Acid (y-PGA)-Based Micro/Nanoparticles as a Delivery System for Antimicrobials and Other Biomedical Applications. **2017,** *18* (2), 313.

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

51. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **1996,** *14* (1), 33-38.
52. Ashiuchi, M., Microbial production and chemical transformation of poly-γ-glutamate. *Microbial Biotechnology* **2013,** *6* (6), 664-674.
53. Ogunleye, A.; Bhat, A.; Irorere, V. U.; Hill, D.; Williams, C.; Radecka, I., Poly-γ-glutamic acid: production, properties and applications. *Microbiology* **2015,** *161* (1), 1-17.
54. Bajaj, I.; Singhal, R., Poly (glutamic acid)–an emerging biopolymer of commercial interest. *Bioresource technology* **2011,** *102* (10), 5551-5561.
55. Schaffer, J.; Woodhams, R., Polyelectrolyte builders as detergent phosphate replacements. *Industrial & Engineering Chemistry Product Research and Development* **1977,** *16* (1), 3-11.
56. Yangxin, Y.; Jin, Z.; Bayly, A. E., Development of surfactants and builders in detergent formulations. *Chinese Journal of Chemical Engineering* **2008,** *16* (4), 517-527.
57. Marchand, G.; Soetens, J.-C.; Jacquemin, D.; Bopp, P. A., Effect of the cation model on the equilibrium structure of poly-L-glutamate in aqueous sodium chloride solution. *The Journal of chemical physics* **2015,** *143* (22), 224505.
58. Satoh, M.; Fujii, Y.; Kato, F.; Komiyama, J., Solvent‐and salt‐induced coil‐helix transition of alkali metal salts of poly (L‐glutamic acid) in aqueous organic solvents. *Biopolymers* **1991,** *31* (1), 1-10.
59. Satoh, M.; Komiyama, J.; Iijima, T., Salt effects on the helix‐coil transition of sodium poly (L‐glutamate) in aqueous solution. *Biopolymers* **1982,** *21* (9), 1927-1931.
60. Tiffany, M. L.; Krimm, S., New chain conformations of poly (glutamic acid) and polylysine. *Biopolymers* **1968,** *6* (9), 1379-1382.
61. Mikhonin, A. V.; Myshakina, N. S.; Bykov, S. V.; Asher, S. A., UV resonance Raman determination of polyproline II, extended 2.51-helix, and β-sheet ψ angle energy landscape in poly-L-lysine and poly-L-glutamic acid. *Journal of the American Chemical Society* **2005,** *127* (21), 7712-7720.
62. Tiffany, M. L.; Krimm, S., Circular dichroism of the “random” polypeptide chain. *Biopolymers* **1969,** *8* (3), 347-359.
63. Xiong, K.; Ma, L.; Asher, S. A., Conformation of poly-L-glutamate is independent of ionic strength. *Biophysical chemistry* **2012,** *162*, 1-5.
64. Fedorov, M. V.; Goodman, J. M.; Schumm, S., The effect of sodium chloride on poly-l-glutamate conformation. *Chemical Communications* **2009,** (8), 896-898.
65. Fedorov, M. V.; Goodman, J. M.; Schumm, S., To switch or not to switch: the effects of potassium and sodium ions on α-poly-L-glutamate conformations in aqueous solutions. *Journal of the American Chemical Society* **2009,** *131* (31), 10854-10856.
66. Zanuy, D.; Alemán, C., Poly (γ-glutamic acid) in aqueous solution: molecular dynamics simulations of 10-and 20-residue chains at different temperatures. *Biomacromolecules* **2001,** *2* (3), 651-657.
67. Zanuy, D.; Alemán, C.; Muñoz-Guerra, S., On the helical conformation of un-ionized poly (γ-D-glutamic acid). *International journal of biological macromolecules* **1998,** *23* (3), 175-184.
68. Holzwarth, G.; Doty, P., The ultraviolet circular dichroism of polypeptides1. *Journal of the American Chemical Society* **1965,** *87* (2), 218-228.
69. Jacobson, A., Salt effects on the conformation of α‐poly‐L‐glutamic acid. *Biopolymers* **1964,** *2* (3), 237-244.

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

70. Wada, A., Helix-coil transformation and titration curve of poly-L-glutamic acid. *Molecular Physics* **1960,** *3* (5), 409-416.
71. Krimm, S.; Mark, J.; Tiffany, M. L., Influence of counterions on the helical conformations of charged polypeptide chains. **1969**.
72. Satoh, M.; Komiyama, J.; Iljima, T., Dependence on ionic strength of charge-induced coil-helix transition of sodium poly (l-glutamate) in aqueous solution. *Biophysical chemistry* **1981,** *14* (4), 347-355.
73. Zanuy, D.; Aleman, C.; Munoz-Guerra, S., On the helical conformation of un-ionized poly(gamma-D-glutamic acid). *Int J Biol Macromol* **1998,** *23* (3), 175-84.
74. Ogasawara, N.; Kasahara, K.; Iwai, R.; Takahashi, T., Unfolding of α-helical 20-residue poly-glutamic acid analyzed by multiple runs of canonical molecular dynamics simulations. *PeerJ* **2018,** *6*, e4769.
75. A.J. Barnes, W. J. O.-T., J. Yarwood, Molecular Liquids: Dynamics and Interactions. D. Reidel Publishing Comapny.
76. Giovanni Ciccotti, R. K., Alessandro Sergi, Handbook of Materials Modeling. Yip, S. E., Ed. Springer Publishing.
77. He, X.; Luo, L.-S., A priori derivation of the lattice Boltzmann equation. *Physical Review E* **1997,** *55* (6), R6333-R6336.
78. Higgins, J. S.; Benoit, H. C., *Polymers and neutron scattering*. Clarendon Press: United Kingdom, 1994.
79. Meller, J., Molecular dynamics. *eLS* **2001**.
80. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* **2006,** *65* (3), 712-725.
81. Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; De Vries, A. H., The MARTINI force field: coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B* **2007,** *111* (27), 7812-7824.
82. Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J., The MARTINI coarse-grained force field: extension to proteins. *Journal of chemical theory and computation* **2008,** *4* (5), 819-834.
83. Periole, X.; Cavalli, M.; Marrink, S.-J.; Ceruso, M. A., Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition. *Journal of Chemical Theory and Computation* **2009,** *5* (9), 2531-2543.
84. Wassenaar, T. A.; Ingólfsson, H. I.; Böckmann, R. A.; Tieleman, D. P.; Marrink, S. J., Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations. *Journal of chemical theory and computation* **2015,** *11* (5), 2144-2155.
85. Periole, X.; Marrink, S.-J., The Martini coarse-grained force field. *Biomolecular Simulations: Methods and Protocols* **2013**, 533-565.
86. Marrink, S. J.; De Vries, A. H.; Mark, A. E., Coarse grained model for semiquantitative lipid simulations. *The Journal of Physical Chemistry B* **2004,** *108* (2), 750-760.
87. Katchalsky, A.; Gillis, J., Theory of the potentiometric titration of polymeric acids. *Recueil des Travaux Chimiques des Pays-Bas* **1949,** *68* (9), 879-897.
88. Po, H. N.; Senozan, N. M., The Henderson-Hasselbalch Equation: Its History and Limitations. *Journal of Chemical Education* **2001,** *78* (11), 1499.

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66

89. Krieger, E.; Koraimann, G.; Vriend, G., Increasing the precision of comparative models with YASARA NOVA−a self‐parameterizing force field. *Proteins: Structure, Function, and Bioinformatics* **2002,** *47* (3), 393-402.
90. Frisch, M.; Trucks, G.; Schlegel, H. B.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. e., Gaussian 09. Gaussian, Inc. Wallingford, CT: 2009.
91. Becke, A. D., Density‐functional thermochemistry. III. The role of exact exchange. *The Journal of chemical physics* **1993,** *98* (7), 5648-5652.
92. Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical review B* **1988,** *37* (2), 785.
93. Vosko, S.; Wilk, L.; Nusair, M., Accurate spin-dependent electron liquid correlation energies for local spin density functional calculations: a critical analysis. *Can. J. Phys* **1989,** *58*, 1200-1211.
94. Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A., Self‐consistent molecular orbital methods. XX. A basis set for correlated wave functions. *The Journal of Chemical Physics* **1980,** *72* (1), 650-654.
95. Hariharan, P. C.; Pople, J. A., The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretica chimica acta* **1973,** *28* (3), 213-222.
96. Hehre, W. J.; Ditchfield, R.; Pople, J. A., Self—consistent molecular orbital methods. XII. Further extensions of Gaussian—type basis sets for use in molecular orbital studies of organic molecules. *The Journal of chemical physics* **1972,** *56* (5), 2257-2261.
97. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Antechamber: an accessory software package for molecular mechanical calculations. *J. Am. Chem. Soc* **2001,** *222*, U403.
98. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of molecular graphics and modelling* **2006,** *25* (2), 247-260.
99. Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A., A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry* **1993,** *97* (40), 10269-10280.
100. Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A., Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *Journal of Computational Chemistry* **1995,** *16* (11), 1357-1377.
101. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollmann, P. A., Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *Journal of the American Chemical Society* **1993,** *115* (21), 9620-9631.
102. Sousa da Silva, A. W.; Vranken, W. F., ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Research Notes* **2012,** *5* (1), 1-8.
103. Gromacs. http://www.gromacs.org/.
104. Vögele, M.; Holm, C.; Smiatek, J., Coarse-grained simulations of polyelectrolyte complexes: MARTINI models for poly (styrene sulfonate) and poly (diallyldimethylammonium). *The Journal of chemical physics* **2015,** *143* (24), 243151.

105. Michalowsky, J.; Schäfer, L. V.; Holm, C.; Smiatek, J., A refined polarizable water model for the coarse-grained MARTINI force field with long-range electrostatic interactions. *The Journal of Chemical Physics* **2017,** *146* (5), 054501.
106. Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J., Polarizable water model for the coarse-grained MARTINI force field. *PLoS Comput Biol* **2010,** *6* (6), e1000810.
107. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J., GROMACS: fast, flexible, and free. *Journal of computational chemistry* **2005,** *26* (16), 1701-1718.
108. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **1995,** *91* (1–3), 43-56.
109. Berendsen, H.; Grigera, J.; Straatsma, T., The missing term in effective pair potentials. *Journal of Physical Chemistry* **1987,** *91* (24), 6269-6271.
110. Berendsen, H. J.; Postma, J. v.; van Gunsteren, W. F.; DiNola, A.; Haak, J., Molecular dynamics with coupling to an external bath. *The Journal of chemical physics* **1984,** *81* (8), 3684-3690.
111. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* **1981,** *52* (12), 7182-7190.
112. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *The Journal of chemical physics* **1995,** *103* (19), 8577-8593.
113. Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G., LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry* **1997,** *18* (12), 1463-1472.
114. Frisch, M.; Trucks, G.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G., Gaussian 09, revision a. 02, gaussian. *Inc., Wallingford, CT* **2009,** *200*.
115. Lee, B.; Richards, F. M., The interpretation of protein structures: estimation of static accessibility. *Journal of molecular biology* **1971,** *55* (3), 379-IN4.
116. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1*, 19-25.
117. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *The Journal of chemical physics* **2007,** *126* (1), 014101.
118. Hess, B., Force fields, thermo- and barostats.
119. Rühle, V. J. J. C., Pressure coupling/barostats. **2008**.
120. De Jong, D. H.; Baoukina, S.; Ingólfsson, H. I.; Marrink, S. J., Martini straight: Boosting performance using a shorter cutoff and GPUs. *Computer Physics Communications* **2016,** *199*, 1-7.
121. Hess, B., P-LINCS: A parallel linear constraint solver for molecular simulation. *Journal of Chemical Theory and Computation* **2008,** *4* (1), 116-122.
122. Tomida, T.; Tomida, M.; Nishihara, Y.; Nakabayashi, I.; Okazaki, T.; Masuda, S. J. P., Properties of polyacryloylacetone for adsorption of divalent metal ions. **1990,** *31* (1), 102-105.
123. Karppi, J.; Åkerman, S.; Åkerman, K.; Sundell, A.; Penttilä, I. J. J. o. p. r., Adsorption of metal cations from aqueous solutions onto the pH responsive poly (vinylidene fluoride grafted poly (acrylic acid)(PVDF-PAA) membrane. **2010,** *17* (1), 71.

124. Pastor, R. W.; Brooks, B. R.; Szabo, A. J. M. P., An analysis of the accuracy of Langevin and molecular dynamics algorithms. **1988,** *65* (6), 1409-1419.
125. Martys, N. S.; Mountain, R. D. J. P. R. E., Velocity Verlet algorithm for dissipative-particle-dynamics-based models of suspensions. **1999,** *59* (3), 3733.
126. Manning, G. S., *J. Chem. Phys.* **1969,** *51*, 924.
127. Olvera de la Cruz, M.; Belloni, L.; Delsanti, M.; Dalbiez, J. P.; Spalla, O.; Drifford, M., *J. Chem. Phys.* **1995,** *103*, 5781.
128. Eisenhaber, F.; Lijnzaad, P.; Argos, P.; Sander, C.; Scharf, M., The double cubic lattice method: efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *Journal of Computational Chemistry* **1995,** *16* (3), 273-284.

System Administrator 26 Oct 2020 at 13:07:37 18:7e:b9:0a:6c:66