Cardiff University - School of Chemistry

Polymer-drug delivery. Combining computer simulations and experimental techniques.

Thesis deposited posthumously for the degree of
Doctor of Philosophy

on behalf of
Glib Meleshko (1055862)

2015

DECLARATION

We, the candidate's supervisors, Dr Alison Paul and Dr James A Platts, confirm that the work presented in this thesis is the result of the independent work/investigation of Dr Glib Meleshko, except where otherwise stated. Other sources are acknowledged by explicit references.

Subsequent to submission and assessment of the thesis for the posthumous award of the PhD, but prior to submission of the final thesis to the library, minor changes have been made to the thesis by the supervisors. These were minor corrections for presentation and clarification purposes, and are as requested by the external examiners. Any additions to the thesis have been clearly indicated in the text. The opinions expressed remain those of the candidate throughout.

Dr Alison Paul Dr James A Platts

Abstract

The conformation that polymer-drug conjugates form in solution have a significant effect on properties that are important for designing of drug delivery systems. For N-(2-hydroxypropyl)methacrylamide (HPMA) copolymer conjugates it is known that aggregation number, size and shape affect the rate at which a drug is enzymatically cleaved from the polymer backbone. Investigation of conformational properties could lead to creation of polymeric systems with the ability to keep drug levels at biologically active and safe concentrations for desired period of time. This project is focused on establishing the required modeling methodologies to describe the solution configurations of HPMA based conjugates. Therefore structural properties such as size, shape and density distribution of a range of HPMA copolymers have been investigated. The suitability of atomistic force fields has been assessed against rotational barriers and relative conformational energies obtained from *ab initio* and DFT data for a monomer and dimer of HPMA.

Following this, the AMBER99 parameter set was chosen for all molecular dynamics simulation. Radius of gyration (R_g), radial distribution function (RDF), shape, and density profiles of particular atom types were calculated for a range of HPMA homopolymers sizes from 4 to 200 repeat units (2 to 35 kDa). Results were interpreted in the context of Flory's mean field approach, and compared with data obtained from small angle neutron scattering (SANS) experiments. Results of this study were used for investigation of HPMA conjugates with drug mimics. A range of linear amines (aminohexane(C6), aminooctane(C8), aminododecane (C12)), hydroxyl and fluoro terminated linear amines as well as aromatic aminoanthracene (ANC), aminocrysene (AC) and aminoanthraquinone (ANQ), bound to the polymeric carrier via a tetrapeptide linker (glycine-phenylalanine-leucine-glycine) (GFLG) (Mw ~ 30 kDa) were selected as model objects for study of the effect of drug type and loading on HPMA copolymer conformation.

Using obtained results for further investigation we have progressed to more complex systems of mixed polymer conjugates containing drug-mimic parts with similar parameters of hydrophobicity, but different in terms of flexibility for drug-mimic chain. In order to provide corresponding comparison we have selected systems of Adamantane (Ad)/ANC, Ad/ANQ, methyl-Adamantane (AdMe)/C12, hydroxyl-Adamantane (AdOH)/C10OH and Ad/C10 conjugate

mixtures and investigated the effect of changing ratio of drug-mimic parts for these systems. Analysis of SANS data revealed how conformation can be affected by the drug mimic's intrinsic volume variation, and allowed us to get closer to finding answers for questions that can increase effective use of polymer-drug therapeutics. SANS experimental scattering curves were compared with theoretical curves, predicted from molecular dynamics (MD) simulations. Parameters such as size and shape fitted to SANS data were compared with relevant simulated structures. Based on results of previous studies as well as additional polymer synthesis and characterization process we were able to develop reliable all-atom (AA) and coarse-gain (CG) computer models for simulation of HPMA polymer. Required tools and software were developed. Various methods were used to increase performance efficiency of Molecular Dynamics calculations. Among them highest practical impact has a domain decomposition parallelization strategy.

Acknowledgements

Many people have contributed towards the completion of this thesis and I would like offer my thanks to all of those who have taken the time to give me support over the years.

Firstly, I would like to express my gratitude to my supervisors Dr. Alison Paul and Dr. James Platts for giving me the chance to undertake this research project. I wish to thank both of them for selecting my project for Soft Matter Research Group research and all the effort they have invested in it within this time. I am extremely grateful for their advices, patience, enthusiasm, and encouragement that guided me and helped to finish this project. As well I would like to thank my mentor Dr. Willock for his patience and all help he have given me during this project.

I would like to thank the instrument scientists at ISIS and the ILL facilities; special thanks go to Dr. Richard Heenan and Dr. Ralf Schweins who have helped a lot in understanding SANS characterization method and application of this knowledge for studied systems.

I thank Dr. Maria J. Vicent for her advices and help on each stage of this project.

I wish to thank the Soft Matter Group and Theoretical and Computational Chemistry Group teams for all advices and help in the lab over the years; especially to Ibrahim, Jamie, Zaineb, Ziri, Abi and other students and staff members.

I would also like to mention Dr Rob Jenkins, Dr. Luca Monticelli and Dr. Peter Griffiths for all the guidance and help they have provided.

Finally, special thanks go to my family for their support and in particular to my grandparents to whom I dedicate this thesis. They are the one who should be really getting this degree.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AA all-atom

Ab Amino biphenyl

Ab-OH 4-amino-4'-hydroxybiphenyl

Ac Amino chrysene

An Aniline

Anc Amino anthracene

Anq Amino anthraquinone

Ap 1-amino-2-propanol

C10 1-aminodecance

C10-OH 1-aminodecan-1-ol

C12 1-aminododecane

C12-OH 1-aminododecan-1-ol

C14 1-aminotetradecane

C16 1-aminohexadecane

C18 1-aminooctadecane

C6 1-aminohexane

C6-OH 1-aminohexan-1-ol

C8 1-aminooctane

C8-OH 1-aminooctan-1-ol

CG coarse-grained

CPU central processing unit

D2O Deuterium Oxide

DFT Density functional theory

DLS Dynamic light scattering

DMSO Dimethyl sulphoxide

DNA Deoxyribonucleic acid

DOF degrees of freedom

Dox Doxorubicin

EPR Enhance permeability and retention effect

GAL Glactosamine

GFLG Glycine-phenylalanine-leucine-glycine

HF Hartree–Fock

HPMA N-(2-hydroxypropylmethacrylamide)

ILL Institue Laue-Langevien

LJ Lennard-Jones

LDA Local Density Approximation

MTD Maximum tolerated dose

MD Molecular Dynamics

NMR Nuclear Magnetic resonance

ØNp Para nitrophenol

PEG Poly ethylene glycol

PGA (α-L-glutamic acid)

PGG-PTX poly-γ-glutamyl-glutamate paclitaxel

PLL Poly(L-lysine)

Rg Radius of gyration

RDF radial distribution function

SALS Small-angle light scattering

SANS Small-angle neutron scattering

SEM Scanning electron microscopy

QM/MM quantum mechanics/molecular mechanics

Chapter 1: Introduction

1.1. Polymers for therapeutics

1.1.1. Challenges of cancer treatment

First documented in ancient Egypt, in 1500 BC [1], cancer still causes more than 7 million deaths worldwide every year [2]. Of course, methods of cancer treatment are far ahead of the cauterization (a method to destroy tissue with a hot instrument called "the fire drill") and nowadays effective treatment can achieve full recession in 17-75% of cancer cases [3]. However cancer remains one of the most common causes of death and is predicted to move ahead of heart diseases in the near future [4]. Cancer can be treated by surgery, chemotherapy, radiation therapy, immunotherapy, monoclonal antibody therapy or other methods. The choice of therapy depends upon the location and grade of the tumor and the stage of the disease, as well as the general health state of the patient. These methods have shown successful clinical results, but the price and the risk of negative effects with currently used therapeutics are reasons for development of new experimental cancer treatments.

Among them are polymer therapy methods, developed in the second half of the 20th century, which can be considered as a key to finding a safer cure for different types of cancer [5]. These promising forecasts are based on distinct advantages, polymer-drug therapy have over conventional chemotherapy. These include:

- passive tumor targeting owing to the enhanced permeability and retention (EPR) effect, a phenomenon that arises from the hyperpermeability of angiogenic tumor blood vessels [6]

- the potential to bypass mechanisms of drug resistance after cellular uptake by the endocytic route [7,8]

- the ability to target desired sites in the human body and localize drug action where is required [8-12]

- the ability to keep drug levels at biologically active and safe concentrations for desired period of time [7-10]

In general, these features define principles for drug-delivery system design and application. The systematic study and development of such systems allowed introduction of more effective methods for cancer treatment, using known drugs but with decreased side-effects and enhanced activity.

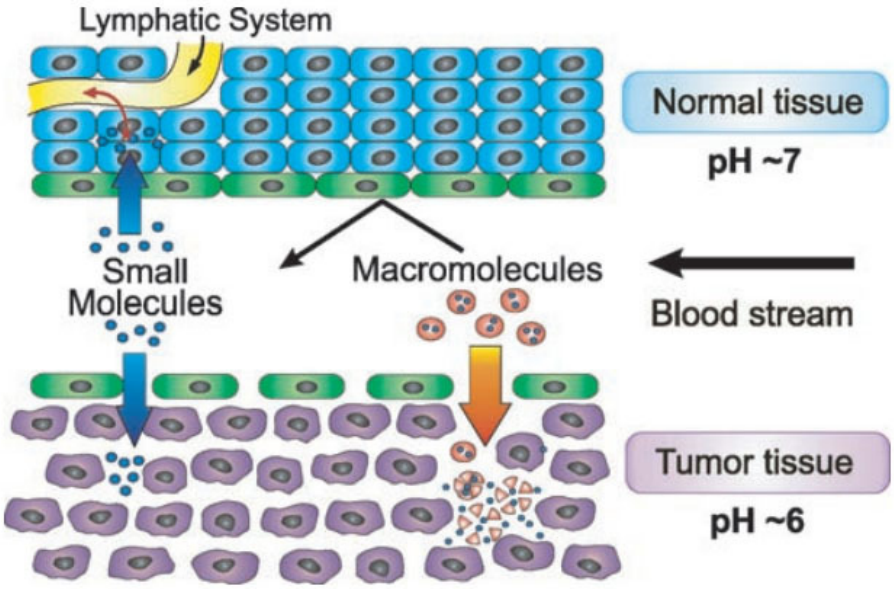


Figure 1.1. Schematic representation EPR effect for normal and tumor tissue [12].

In general, the EPR effect (Figure 1.1) can be described as passive targeting mechanism based on ability of low-molecular-weight compounds to diffuse into normal and tumour tissue through the cell layer of blood capillaries. Macromolecules, however, cannot pass through the capillary walls of healthy tissue. In contrast to the blood capillaries in most normal tissues, the endothelial layer of the capillaries in the tumour tissue is deformed so that macromolecules can pass, leading to passive targeting of tumour sites. The EPR effect is observed for macromolecules with molecular weights greater than 20 kDa [5]. It has been shown that the EPR effect is molecular weight dependent, and that higher molecular weight conjugates have greater passive targeting [13]. However, in order for the polymer to be excreted by the kidneys, the molecular weight must be below 45 kDa [14].

Apart from passive targeting, introducing polymer to a drug-delivery system provides the potential for active intercellular and extracellular targeting by various mechanisms, such as: pH and enzyme-responsible side-chain groups, optional targeting groups, or a combination of them [13].

Such selective action can be applied for direct improvement of biological activity for selected drugs and demands further investigation.

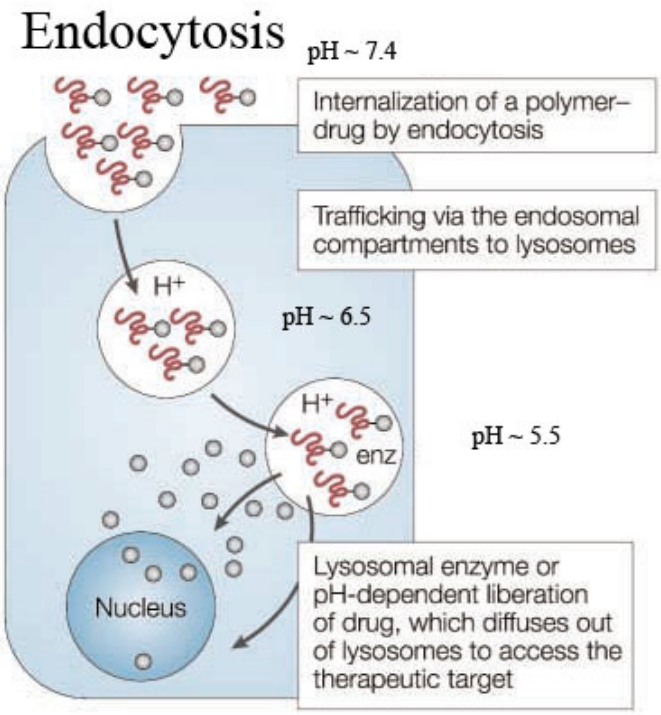


Figure 1.2: Schematic representation of the endocytosis process [4].

One mechanism of intercellular targeting often applied for polymer-drug delivery is endocytosis, a term first used by De Duve [15, 16]. Initiated by a receptor interaction, an adsorptive interaction with the cell wall, or via passive fluid phase, takes place in cell (Figure 1.2). The endosomal compartment will then traffic the conjugated drug to a primary lysosome containing proteolytic enzymes, which are naturally designed to degrade proteins [17]. Finally, the drug is released by enzyme interaction or break of pH-dependent linker. During this process pH will be changed from the 7.4 in the blood stream to the 6.5 in the primary endosome and 7.5 in the secondary endosome. Based on these principles, further application of this mechanism for polymer-drug delivery as cancer targeting lysosomotropic agents was proposed by Ringsdorf [18]. Later, experiments with ^{14}C labelled polymers injected into rats muscle tumors demonstrated that the tumor tissue had polymer uptake with a factor of 6 higher than the normal muscle tissue [19].

1.1.2. Polymer-drug delivery systems development

There are various systems related to the term “polymer therapeutics”. It is used for polymeric drugs, polymer–drug conjugates, polymer–protein conjugates, polymeric micelles to which drug is covalently bound, and multicomponent systems. (Figure 1.3) All are based on water-soluble polymers, either bioactive or as an inert functional part of a drug-delivery system for improved drug, protein or gene delivery. During the past two decades, an effective biological rationale has emerged for the design of each of the subclasses of polymer therapeutic based on progress in polymer, analytical, computational and soft-matter chemistry. New systems have been introduced and carried to different stages of clinical trials. Results of such trials show benefits of introducing of polymer therapeutics for cancer treatment and motivate research groups to further investigation. Polymeric materials for such systems can be extracted from plants, animals and seaweed, particularly polyanions and polysulphates. Or, like DIVEMA (pyran copolymer) and COPAXONE (random copolymer of L-alanine, L-lysine, L-glutamic acid and L-tyrosine), they can be a product of chemical synthesis [16].

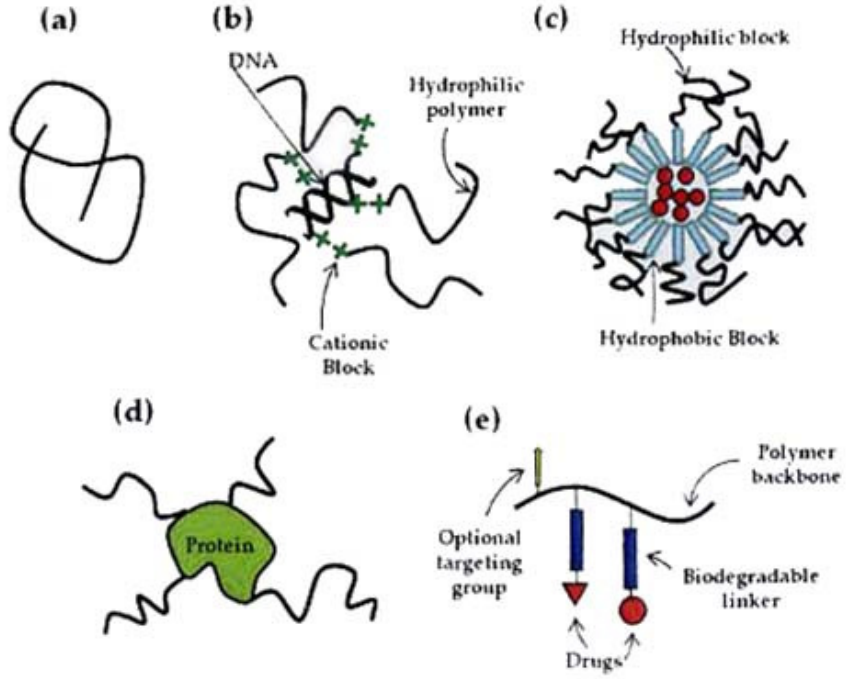


Figure 1.3. Different polymer therapeutics: (a) polymeric drugs, (b) polyplexels, (c) polymeric micells, (d) polymer-protein conjugates, (e) polymer-drug conjugates [5].

Polymer–protein and polymer–drug conjugates share many common features, but the biological principles for their design is quite different. Peptide-, protein- and antibody-based drugs were first used for therapy but had a range of limitations, crucial for clinical application. These often include a short plasma half-life, poor stability and immunogenicity. Such limitations created a need for future method development. In the 1970s research by Davis Abuchowski and colleagues introduced the so-called PEGylation technique, designed to increase protein solubility and stability, and to reduce protein immunogenicity [17]. The result was a need for less frequent dosing caused by prolonged plasma half-life. These method is still being used and its clinical value is now well established.

At the same time another concept was introduced and described in the works of De Duve and Ringsdorf [16-18]. It was shown that covalent conjugation of a low molecular weight drug to a polymer backbone would alter drug pharmacokinetics at a cellular level by restricting uptake to the endocytic route. These opened a new area for research and allowed use of new polymers for drug-delivery. As well as new perspectives, new challenges arose, resulting in several important requirements for both drug and carrier. Whereas the linker binding the drug to the polymer chain must be stable when in circulation to ensure the EPR effect, at some point the drug must be released from the polymer backbone. As previously mentioned, this can be accomplished in two main ways: using a linker sensitive to the pH change of endocytosis or using an enzymatically degradable linker sensitive to the enzymes encountered within the lysosome [14, 15].

In general, a polymer for drug delivery, and any of its metabolic products, should be:

- non toxic
- water-soluble
- biodegradable
- well characterized
- protected from inactivation by plasma enzyme
- macromolecules with a molecular weight in a range of 20-40 kDa
- able to bind to the desired drugs
- easily synthesized

Regarding these principles, several polymer carriers have been developed over the past 40-50 years. Many of them were selected for usage in drug-delivery systems: poly(α-L-glutamic acid) (PGA) [6], polyethylene glucol (PEG) [7], poly(L-lysine) (PLL) [8]), poly(vinyl- pyrrolidone) (PVP) [20], poly(ethyleneimine) (PEI) [20], linear polyamidoamines, alternating cyclocopolymer of divinyl ether and maleic anhydride (DIVEMA) [20] and N-(2-hydroxypropyl)methacrylamide (HPMA) [9]. Several natural polymers have been successfully transferred to the clinic: (dextran (α-1,6 polyglucose), dextrin (α-1,4 poly- glucose), hyaluronic acid and chitosans [20].

However, despite all advantages and huge potential for usage in biochemistry, polymer therapies have some unsolved problems and unanswered practical questions. Among them are the undefined structure of some co-polymers, and details of polymer interaction with the cell and drug-release mechanism [4]. Another aspect should be pointed out: it is crucial to understand solution behavior of such polymeric systems and how it is linked to a conformation and drug-loading.

1.1.3. Polymer-drug conjugates in current use

There are various systems available for application in polymer-drug delivery. These include (Figure 1.4):

i) Polymer–drug conjugate + low molecular weight drug or other type of therapy eg PGA paciltaxel + Carboplatin [20, 21]
ii) Polymer–drug conjugate (HPMA-GFLG-Dox, HPMA-GFLG-Gal-Dox [9])
iii) Polymer–drug conjugate + polymer–drug conjugate (HPMA-doxorubicin/mesochlorin [22])
iv) Single polymeric carrier carrying a combination of drugs (HPMA-GFLG-Dox-AGM, PEG-Epi-NO, CPT-PEG(-LHRH)-BH3 [22-25])

Based on different strategies, these systems can enhance drug performance by introducing new features by selective action and complex release mechanisms. The presence of two or more therapeutic agents on a single polymeric chain opens new therapeutic possibilities but at the same time poses new challenges.

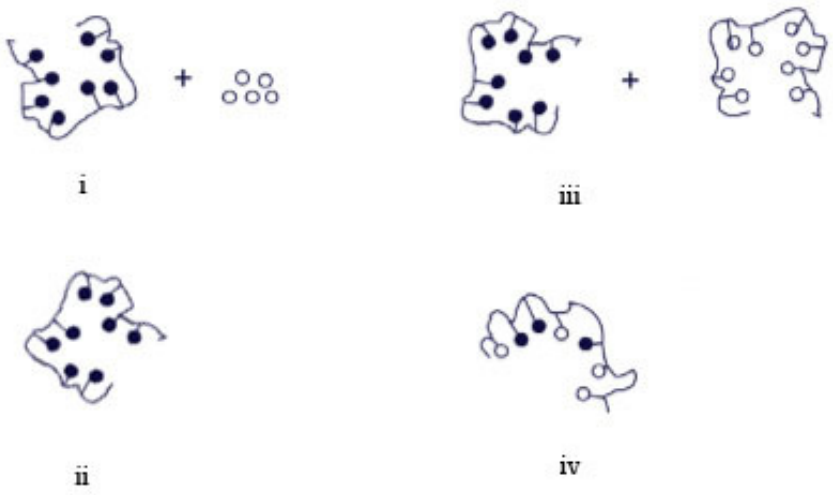


Figure 1.4. Different types of polymer-drug delivery systems

Among the most promising candidates for polymeric carrier part for drug-delivery systems are HPMA polymer and its conjugates. It was shown that HPMA is both suitable as an model object for investigation and can be used as a component for medical treatment. They fulfill all requirements for drug-delivery system: they are highly hydrophilic, non-immunogenic, non-toxic, have a binding sites for drugs to be attached and reside well in the blood circulation. HPMA homopolymer was originally developed by Kopecek and colleagues as a plasma expander. Collaborative research with Duncan and colleagues in the early 1980s produced two HPMA copolymer–doxorubicin conjugates that subsequently progressed into I/II clinical phases [24].

Using doxorubicin as a model drug, significant efforts at the preclinical level have also been invested in improving the passive targeting properties of the HPMA copolymers, as well as in enhancing drug delivery to tumor cells by active targeting [23]. As HPMA copolymers have already shown activity in chemotherapy refractory breast-cancer patients and this polymer has proven clinical safety [9]. Various HPMA-conjugates were studied within the Soft Matter Research Group so synthetic and characterization techniques are well known and successfully applied [19-20]. Conjugates are water soluble and can be obtained from parent a polymer by

substitution reaction relatively easily. For these reasons HPMA was chosen as the polymeric carrier for our investigation.

1.1.4 Importance of linker, loading, conformation.

The conformations that polymer-drug conjugates may form in solution have a significant effect on properties that are important for designing of drug delivery systems. In practice, these depend on drug type, loading and mixture rates. For example, it was demonstrated that for selected HPMA copolymer conjugates (Figure 1.5) the effect of these factors can be different for each system. Thus, for aminopropanol (Ap) increase of drug-loading from 5 to 10 mol % affects size, but not the overall shape of the molecule, keeping it spherical. For HPMA-*co*-MA-GFLG-*para*-nitrophenol (ONp) the shape of the molecule adopted in solution will vary from sphere to a rod for 5 and 10 mol% loading respectively. At the same time, for both 5 and 10 mol % dodecylamine (C12) conjugates scattering results suggest rod-like structure with change of length as loading increased (Figure 1.6)

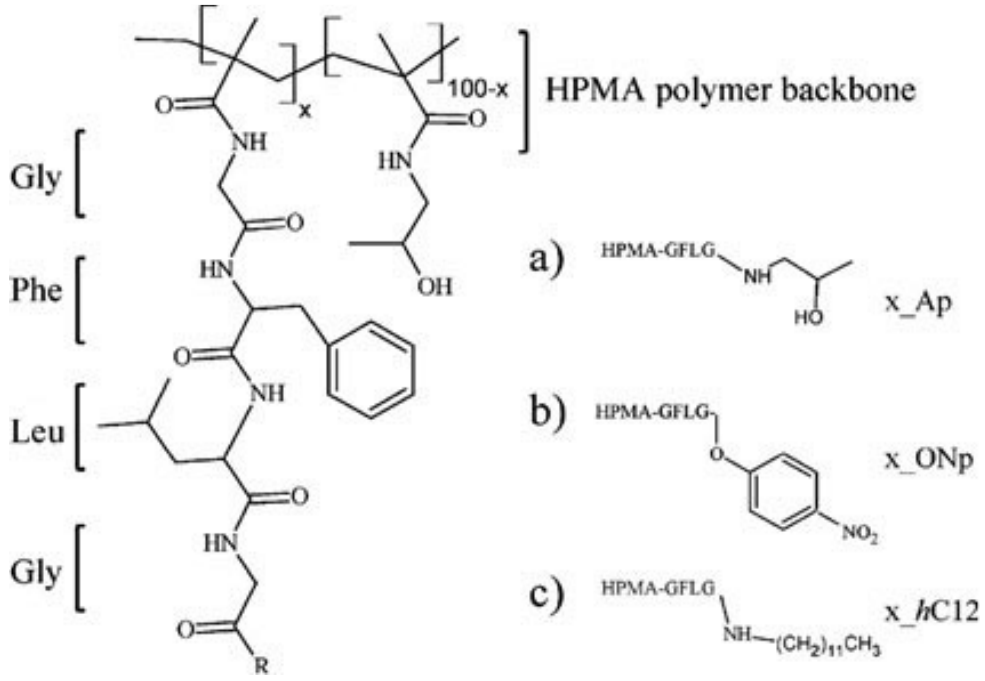


Figure 1.5: Structure of the HPMA-*co*-MA-GFLG copolymer where x defines the mole percent of GFLG-R side chain loading and R is the bound drug mimic: a) aminopropanol, x_Ap; b) *para*-nitrophenol, x_ONp, c) dodecylamine, x_*h*C12.

Further investigations suggested that as the size of the hydrocarbon increases, the conjugate conformation will change from a fully flexible Gaussian coil structure (Ap) to a more rigid chain structure (hexylamine-decylamine (C6 – C10)) and finally to a rigid rod structure for larger alkanes (C12 – C18). Changes made in the hydrophobicity of the substituent demonstrated conclusively that hydrophobicity rather than chain structure is responsible for the conformation changes observed. Where the conjugates have been modified with aromatic substituents, the results follow the same trend: as the substituent increases in size, the flexibility reduces from a worm-like chain for aniline (An) to a rod for aminoanthracene (Anc) [24].

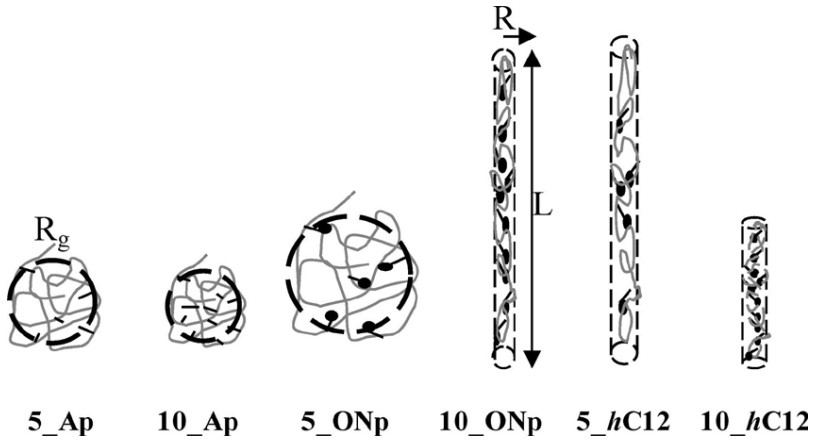


Figure 1.6.: Schematic representation of the best-fit morphologies concluded from the SANS analysis. [24]

In addition, it was shown that for HPMA copolymers conjugate aggregation as well as the size and shape adopted by the molecule in solution will affect the rate at which the drug is enzymatically cleaved via GFLG linker from the polymer backbone [21, 25]. As demonstrated by Ulbrich, a series of copolymer HPMA-GLFG conjugates containing 6.9, 3.3 and 1.5 mol % of p-nitroaniline (NAp) formed aggregates with association numbers of 5, 2, and 1 respectively [25]. The rate at which free NAp was released was measured using UV spectroscopy. It was found that the overall amount and speed of NAp released was lower for conjugates with higher association numbers i.e. higher loading (Figure 1-7 a). Aggregate formation and polymers coiled structure was suggested to act as shield from the active site of the enzyme. (Figure 1-7 b).

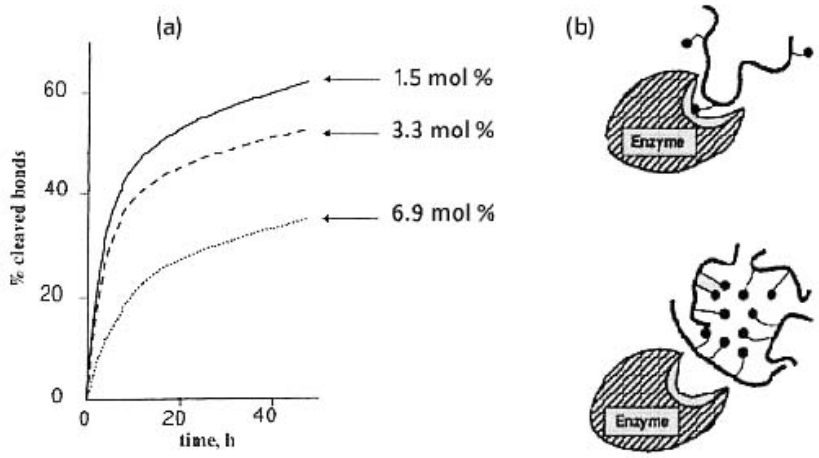


Figure 1.7. Influence of structural properties on biological behavior: (a) observed release kinetics for conjugates containing different mol % of GLFG-Nap side chains, (b) influence of the type of aggregates formed in solution on enzyme activity [25].

It is also known that clinical performance related to parameters, such as maximum tolerated dose (MTD), may be linked to the conformation of a polymer-drug conjugate.[19] For example, Paul et el showed that for clinically tested antitumor therapeutics: PK1(FCE28068) and PK2(FCE28069) (Figure 1.8) differ in sizes of conjugates adopted in solution. Presence of Galactosamine (Gal) as a targeting group leads PK2 to a less compact structure (R_g = 10.5 nm) than PK1 (R_g = 7.8 nm), which corresponds to equivalent decreasing of the MTD for PK2 comparing to PK1 with a factor of two for almost at the same level of the drug content (Figure 1.8. a). AGM-Dox study also showed conformational difference for HPMA conjugates containing different mixed drugs [26].

These studies suggest that drug type and drug loading, as well as presence of side-groups, have an influence on solution conformation, drug-release mechanism and potentially to clinical performance, and hence lead to understanding of drug-delivery systems design. More in depth and systematic studies will provide the basis of design rules for next generation of conjugates, establishing a mechanism for a controlled delivery and release of drug or combination of drugs.

compound	side chain content (mol %)	Dox[f] content (wt %)	Gal content (mol %)	Dox release (% at 24 h)	MTD (mg/m^{-2} Dox-equiv)	R_a[d] (nm ± 0.5)
			Control Conjugates[a]			
HPMA copolymer–GFLG–Ap	5	n/a	n/a	n/a	n/a	ND
HPMA copolymer–GFLG–Ap	10	n/a	n/a	n/a	n/a	ND
HPMA copolymer–GFLG–Gal–Ap	10	n/a	≈2	n/a	n/a	4.5
			Conjugates in Clinical Trials[b]			
HPMA copolymer–GFLG–Dox (FCE28068)	5	8.5	n/a	60[e]	320	7.8
HPMA copolymer–GFLG–Gal–Dox (FCE28069)	10	7.5	1.5–2.5	60[e]	160	10.5
			Pharmaceutical Formulations[a,c]			
HPMA copolymer–GFLG–Dox (lyophilized FCE28068)	5	8.5	n/a	n/a	320	7.8
HPMA copolymer–GFLG–Gal–Dox (lyophilized FCE28069)	10	7.5	1.5–2.5	n/a	160	9.0

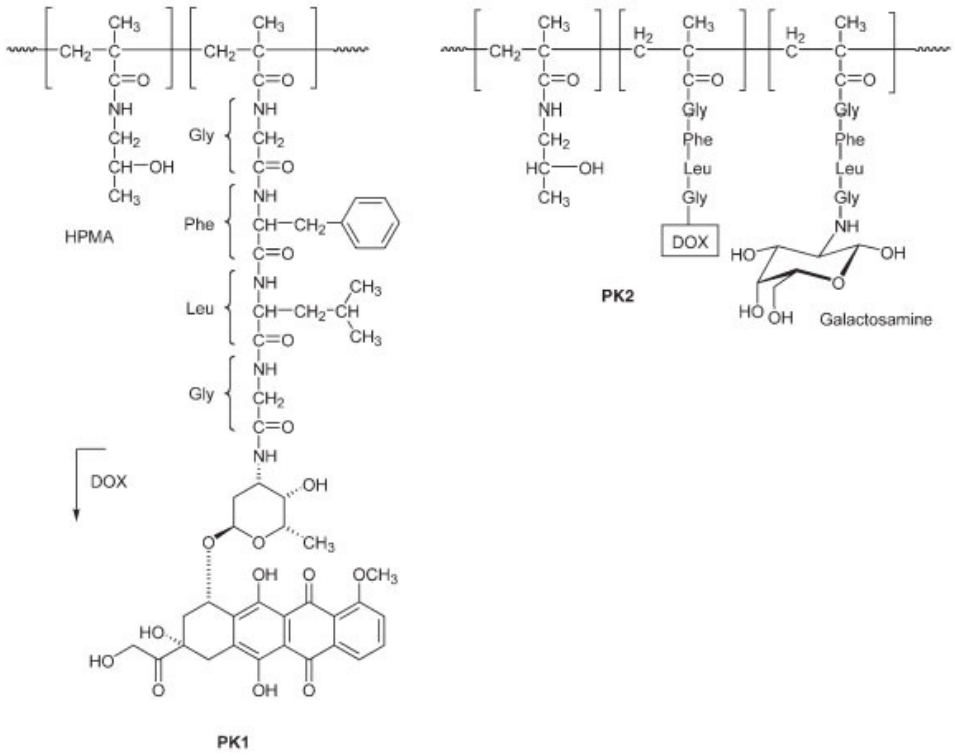


Figure 1.8. a. Results of clinical trials for seleched polymer-drug therapeutics; b. Chemical structure of the polymeric antitumor therapeutics: PK1 (left) and PK2 (right).

1.1.5 Solvent conjugate interactions

In theory, structural parameters of a single polymer chain depends on the solvent [30]. For a good solvent the chain is more expanded while for a bad solvent the chain is coiled. In the limit of a very bad solvent the polymer chain collapses to form a hard sphere, while in good solvent the chain swells in order to maximize the number of polymer-solvent interactions [31]. For this case the radius of gyration, R_g, is approximated using Flory's mean field approach which yields a scaling for the radius of gyration of:

$$R_g = R_0 * N^{\upsilon} \tag{1.1}$$

where R_0 is the radius of gyration of the monomer unit, N is the number of repeated units and ν is a solvent factor. A good solvent will give $\upsilon \sim 2/3$ while for bad solvent the tight spherical conformation leads to $\upsilon \sim 1/3$. Therefore, polymer in good solvent has a larger size and behaves like a fractal object. In bad solvent it behaves like a solid sphere. In the so-called θ solvent, $\upsilon \cong 1/2$, which is the result of simple random walk and the chain behaves as if it were an ideal chain. However, R_g provides only general information on conformation size and structure. Separate axes analysis for such systems can contribute more detailed characteristics of polymer-drug conjugate, such as shape, relative sizes, sphericity, etc. The inclusion of pendant linker groups greatly complicates the behavior due to the complex molecular structure and associated intra-molecular interaction caused by varying hydrophobicity of each part of the conjugate molecule. Advanced experimental techniques are often applied to determine conjugates structure and study solvent behavior of such systems. These techniques include viscosity, Dynamic light scattering (DLS) and Small Angle Neutron Scattering (SANS) measurement followed by data analysis [37].

1.2 Experimental characterization of polymer solution conformation

1.2.1 Small Angle Neutron Scattering (SANS)

1.2.1.1. General information and application

Small Angle Neutron Scattering (SANS) is a neutron scattering technique that provides accurate data for the study of materials on the nanometre to micrometre length scales. The experiment consists of a well collimated beam of neutrons being passed through a sample and detectors to

count the number of neutrons scattered as a function of angle and neutron wavelength. This data can then be used to extract information about the shape, size, arrangement, and interactions of the components of the sample. SANS studies date from the early works of A. Guinier in 1938 [29], and later publications by O. Kratky, W. Beeman, P. Schmidt, V. Luzzati where the theoretical basis and experimental techniques were described. In the 1960s, the method became increasingly important as it allowed one to get low-resolution structural information on the overall shape and internal structure of the studied system in the absence of crystals, making water-soluble polymers accessible for measurement and characterization [30].

A breakthrough in SANS experiments came in the 1970s with development of new experiments using new powerful neutron source beams and advanced detectors as well as result analysis and model fitting evolution. Further progress in this field was introduced by application of a contrast variation technique from available deuteration methods. Nowadays more than twenty neutron facilities are in operation worldwide including Grenoble (France), Oxfordshire (UK), Brookhaven (US), Risø (Denmark), Studsvik (Sweeden), and Julich (Germany) [31].

The main challenge of SANS as a structural method yet remained the same; to extract information about the three-dimensional structure of the system from the one-dimensional experimental data. Until method development, only overall particle parameters (e.g. volume, radius of gyration) of the macromolecules were directly determined from the experimental data, whereas three-dimensional models were limited to simple geometrical shapes (e.g. ellipsoids, cylinders, etc.). Additional characterization methods such as electron microscopy were often used to obtain necessary parameters for building such models. In the 1980s, progress in structural methods led to the possibility of structural conclusions on studied systems based on trial-and-error models. Use of computers for such procedures allowed increased efficiency and allowed easier access to these methods. The result is a gradual expansion of studies in many directions.

Currently SANS is used in a wide range of scientific fields, but finds particular use in studies of soft condensed matter, such as: molecular self-assembly and interactions in a range of systems, polymer solutions, gels and blends, colloids, micelles, and microemulsions as well biophysics (e.g. lipids and lipid-protein complexes), biology (e.g. solution structures of proteins) and hard

condensed matter (e.g. superconductors and magnetic materials) [37]. Specifically important for biological research it was previously shown that SANS characterization can be carried out under physiologically relevant conditions (body temperature, biological pH, ionic strength, and at a range of conjugate concentrations up to 50 mg mL^{-1}). Besides, SANS is non-invasive, and does not rely on the use of chemical modification, for example. Another significant advantage of SANS is the possibility to obtain both overall and local structural information for macromolecules [36].

1.2.1.1. SANS data analysis and fitting

In order to obtain the required information on a system's structure or properties, a fitting process is required as part of SANS characterization. The analysis requires application of least-squares methods, and the basic principles of linear and non-linear least squares methods are summarized with emphasis on applications in the analysis of small-angle scattering data. These include indirect Fourier transformation, square-root deconvolution, size distribution determinations, and modelling. For reasons of mathematical convenience, the scattering pattern is normally described by intensity (I) as a function of the amplitude of the scattering vector or momentum transfer, Q:

$$Q = \frac{4\pi \sin\theta}{\lambda} \tag{1.2}$$

where λ is the wavelength of the incident radiation, and θ is half the angle between the incident and scattered radiation. In the SANS experiment, λ is fixed and θ is small (typically <3°), so I(Q) versus Q is essentially the intensity as a function of scattering angle. Although I(Q) versus Q is related to the shape of the macromolecule in solution, the profile is not intuitively informative, and to interpret a scattering profile in terms of a structure, it is useful to Fourier transform the scattering profile to obtain the interatomic distance distribution function, P(R), of the scattering particle.

In experimental conditions I(Q), from a polymer solution of volume fraction φ_p is given by Equation 1.3, where ΔF is the difference in scattering length density, F, between the scattering body (of volume V_p) and the solvent, and B_{inc} is the background incoherent scattering.

$$I(Q) = \varphi_p V_p (\Delta\rho)^2 P(Q, R) + B_{inc} \quad (1.3)$$

The form factor, P(Q, R), describes the size and shape, and form factors are known for a range of particle morphologies. These can be combined with appropriate contrast and scaling terms to model the scattering data.

1.2.2 Previous studies of polymer-drug delivery systems in Cardiff Soft Matter Research Group

Previously, the importance of polymer-conjugate conformation in solution for therapeutic outcome as well as application of SANS techniques to study the solution conformation of HPMA conjugates was investigated in work of Soft Matter Research Group of Cardiff School of Chemistry [11]. Focused on polymer mediated drug-delivery systems and physicochemical characterization of macromolecules in solution, various systems were studied and results of these research can be implemented for future investigations. It was shown that, for a range of HPMA copolymer conjugates, both drug type and drug loading have an influence on solution conformation, providing insight into overall drug release and release kinetics observed for HMPA-GFLG-doxorubicin/aminoglutethimide conjugates [22, 23].

Intensive investigation of polymer conformation and morphology linked to drug-type and loading was carried out to explore the solution behavior of HPMA-conjugates and uncover the influence of side-chain on its properties. A wide range of alkane, aromatic, fluorinated and hydroxyl-terminated conjugates were synthesized and characterized by NMR and SANS techniques. This led to important information on structure-property relationships for HPMA-conjugates, introduced some principles of polymer-drug design and established methodology for characterization process. Importance and effective usage of contrast variation experiment as well as application of SANS method was demonstrated [22, 23].

Results of this research were published in a number of articles and demonstrated the importance of HPMA-copolymers conformation in solution for their therapeutic outcome [11, 22, 23]. Despite progress to date, and many benefits demonstrated, SANS experiments are expensive, time consuming and difficult to perform, data interpretation can be a long and complex process. The

possibility of introducing molecular modelling approach to virtually “synthesize” new conjugates by screening their structure and solution behavior is therefore a highly desirable objective.

1.3 All-atom modelling

1.3.1 Motivation and general strategy for computer modelling

An all-atom (AA) model of a polymer contains all the atoms which are present in the real polymer. Their interactions are modelled as realistically as possible, using parameters for simulated systems that are compared with results of experiment or between various simulation conditions. AA simulations are powerful tools for investigation of various biological systems for a number of reasons. For example, experimental determination of chosen parameters for such systems is often time and money consuming process, that can be enhanced by introducing results of molecular dynamics (MD) simulation [38]. In biophysics and molecular biology, the method is frequently applied for ligand docking, simulations of lipid bilayers, homology modeling and *ab initio* prediction of protein structure by simulating folding of the polymer chain from random coil. MD is also frequently used to refine three-dimensional structures of proteins and other macromolecules based on experimental constraints from crystallography or spectroscopy [32].

Design of MD simulations should account for the available computational power and properties of system, chosen as object of research. Simulation size (number of particles, simulation box sizes), timestep and total time duration are important parameters to be taken in account for MD strategy and methodology development. However, the simulations should be long enough to be relevant to the time scales of the natural processes being studied. Parallel algorithms, such as domain decomposition and multiple CPU usage can increase effective simulation time and are often used as long with other methods, such as SHAKE [39] and other timestep-related algorithms. These and other methods will be discussed in more detail in Chapter 2.

In general, use of these methods are often based of decreasing number of parameters calculated, such as number of particles in the system or timestep for simulation process, allowing to obtain required information while maintaining desired accuracy. For simulating molecules in a solvent, as we wish to do, a choice should be made between explicit solvent and implicit solvent. Explicit solvent particles (such as the TIP3P, SPC/E and SPC-f water models) must be calculated

expensively by the force field, while implicit solvents use a mean-field approach [58]. Using an explicit solvent is computationally expensive, requiring inclusion of roughly ten times more particles in the simulation. However, implicit solvent models are not always able to reproduce specific system properties, such as molecule-solvent interactions. These may lead to critical errors in system structure or require longer simulation times for equilibration. (Figure 1.9) Such limitations should be taken in account while performing simulations in solute conditions.

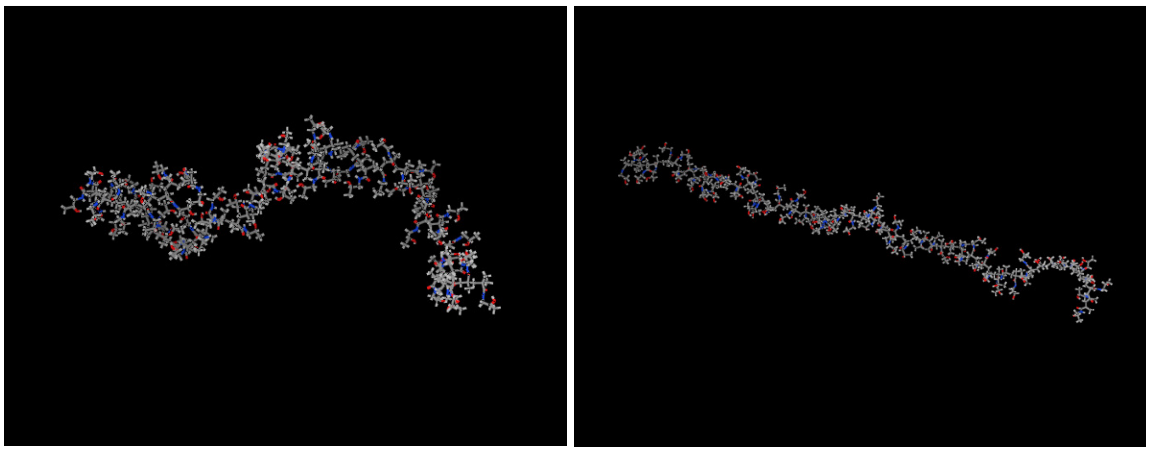


Figure 1.9. Results of MD simulation for HPMA-64 polymer molecule. Left - implicit solvent model (Rg = 16 Å), right - explicit water solvent model (Rg = 33 Å). Time = 1 ns, T = 310K

Another important parameter that has a great influence on simulation process is sizes of simulated system. In MD, the simulation box size must be large enough to avoid boundary condition artefacts. Boundary conditions are often treated by choosing fixed values at the edges or by employing periodic boundary conditions in which one side of the simulation loops back to the opposite side, mimicking intramolecular forces.

A molecular dynamics simulation requires the definition of a potential function, or a description of the terms by which the particles in the simulation will interact usually referred to as a force field. Interatomic potentials are mathematical functions used to describe the potential energy of a statistical mechanical model formed by a system of particles Potentials and their form can be defined at many levels of physical accuracy and depends on available recourses as well as needed

level of details for simulation process. Different force fields are designed for different purposes. These include less detailed "united-atom" interatomic potentials and more accurate AA- interatomic potentials representation. Among them AMBER[40] (widely used for proteins and DNA) CHARMM[41] (widely used for both small molecules and macromolecules) OPLS[42] (developed and optimized for liquid simulations of organic molecules) and MMFF[43] (developed at Merck, for a broad range of molecules) along with many others.

1.3.2 Application for computer simulations to synthetic polymers studies

The increasing number of papers on polymer-conjugate simulations in recent years indicates interest in this area. Growth of computational facilities available to implement techniques such as molecular mechanics, molecular dynamics, mesoscale and statistical mechanics has led to studies of various properties, such as solubility, sizes and shape of obtained structures, polymer chain collapse time, distribution of individual fractions and density distribution. Some studies, of polymer solvation and modelling of pH-sensitive drug release [44] discuss simulation results in comparison with experimental data. Correlations of calculated parameters with experimental observations suggest that computer modelling techniques can be used for accurate prediction of polymer behaviour, and that computational methods have potential use in screening polymeric systems prior to time consuming synthesis and experimental evaluation.

Despite the increase of the interest in applying coarse-graining (CG) techniques (*vide infra*) for polymeric systems, all-atom models provide a detailed overview of intra- and intermolecular interactions, which can be used for subsequent mapping and adjustment of beads in CG models. Nevertheless, for short simulations or for systems of relatively small-size, AA simulations are often carried out. Both explicit and implicit water models can be used for solvent effects as well as vacuum simulations in selected systems. In addition, some of the specific parameters that are based on atoms' geometry can be obtained from AA simulations more accurately than CG methods. However, reverse mapping tools are intensively developed so the geometry resulting from CG simulations can be afterwards transferred to AA representation [46].

Mixed QM-MM modelling methods are also becoming more affordable and accessible for research purpose for highest level of details and can already handle hundreds of atoms [57]. The QM-MM

approach was introduced in the 1976 paper of Warshel and Levitt [45]. These methods open a wide range of applications for AA simulations to polymers of different nature to obtain desired parameters of model systems and such predict properties. In general, this method is based on separate handling of different groups of atoms at different levels of detail. (Figure 1.10). Most parameters can be obtained by highly-accurate quantum chemistry methods for selected atoms while the rest of the system interactions and dynamics will be carried via MD approach. The benefits of such a scheme are efficiency and the ability to explore both timescales and level of details beyond limits of other techniques. However currently QM-MM methods can be implemented only for a limited selection of systems and are not able to provide a realistic models of large-scale polymer molecules in solution. Future development of this approach might allow the study of such systems which will open a new possibilities for research and modeling.

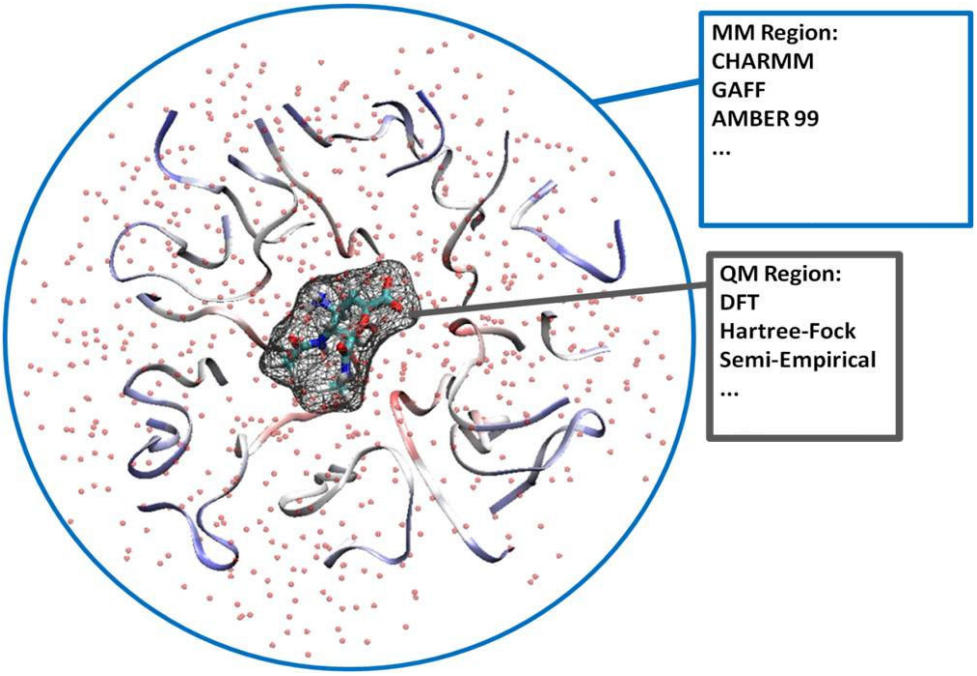


Figure 1.10. Schematic representation of QM-MM method

A wide variety of polymers in a wide range of sizes and structure have been studied involving AA MD simulations for different research purposes. In recent decades, various techniques to handle

the problems of different size and time-scales have been developed. These techniques can be used for research porpoises in different problems like: role of hydrogen bonds in protein and polymer stability, analysis of polymer chain flexibility, polymer chain collapse time, distribution of individual fractions within the molecule and density distribution. AA computer simulations analysis can help in answering many questions regarding the time scale and the nature of process occurs in various systems. Such studies are aimed at understanding the coupling between the molecule and solvent dynamics, solution behavior and its likely influence on biological processes.

1.4 Coarse-grained modelling

1.4.1 Motivation and general strategy

Computational modeling can be beneficial for understanding the dynamical and structural properties of molecular systems. However, AA MD simulations of relatively large systems (thousands to tens of thousands of atoms in solvent) such as polymers are computationally expensive and inefficient. Therefore, to minimize computational costs and time, coarse-grained (CG) models can be applied to reduce the amount of particles in the system, which leads to increase of simulation efficiency. By grouping atoms together, coarse graining reduces the number of degrees of freedom and, therefore, the computational cost. Moreover, due to the removal of high-frequency motions, such as the vibrational movements of hydrogen atoms, larger steps can be made in MD simulations, allowing longer simulations (Figure 1.11). Examples of such CG MD approaches have been applied to a wide range of molecular events, such as investigating protein folding and elucidating protein–protein interactions. A large diversity of coarse graining approaches is available; they range from qualitative, solvent-free models via more realistic models with explicit but simplified water to models including chemical specificity. CG simulation methods for polymer systems include Monte Carlo methods, dissipative particle dynamics, and Brownian dynamics as well as conventional MD.

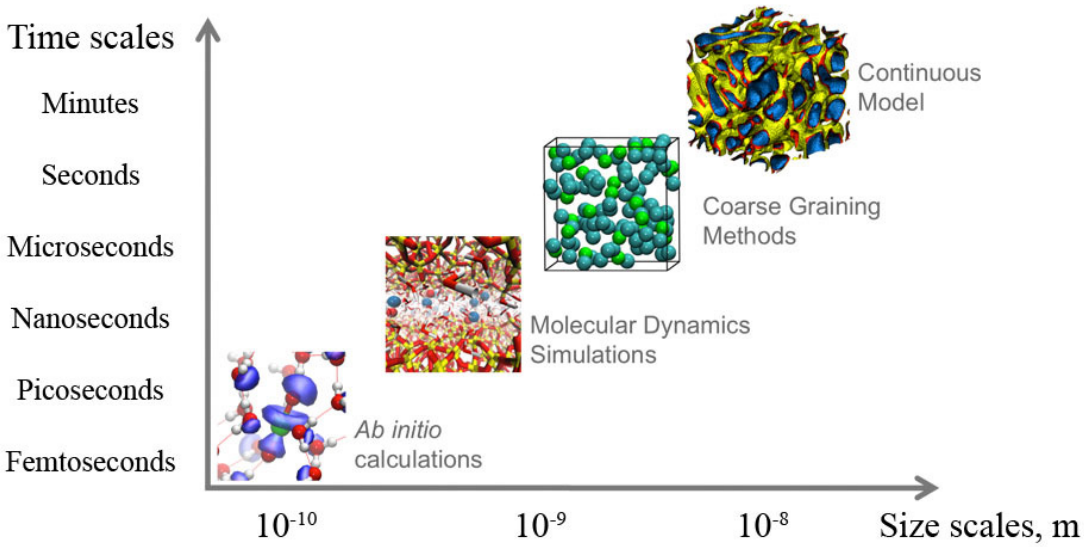


Figure 1.11. Schematic representation of multistate simulations.

1.4.2 CG models of synthetic polymers

An atomistic model of a polymer contains all the atoms which are present in the real polymer. At the other end of the detail scale are coarse-grained and lattice models. They can provide information on molecules structure from representation of atomic groups by beads or other objects. At the same time a list of requirements applied for conversion procedures and often reverse CG to AA conversion is required (Figure 1.12.). How effectively a CG model performs mainly depends on the chosen procedure: i) the model resolution (particles per CG bead), ii) the mapping procedure (reproducing the properties of groups of atoms into those of selected beads), iii) the potential energy function entering the CG Hamiltonian, and iv) the experimental and/or AA simulation properties against which the CG model is calibrated. CG models of polymer chains have improved our understanding of the general features of polymer structure and dynamics. Generic freely jointed, bead-spring and self-avoiding model chains are still the subject of computational and theoretical investigation. In general, two different strategies are commonly followed to define the effective interactions acting between the CG sites. In structure-based CG models, the CG

interaction parameters are tuned to reproduce accurately structural features of the system (typically, radial distribution functions) derived from all-atom MD simulations.

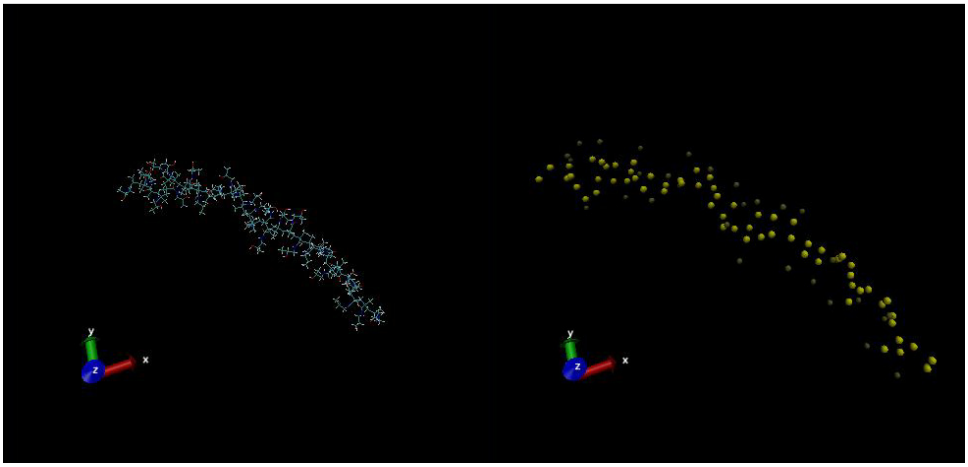


Figure 1.12. Example of AA and CG representation of polymer chain.

Compared to structure-based CG models, thermodynamics-based models rely on different coarse-graining strategy. In the thermodynamics-based approach, interaction parameters are chosen in such a way as to reproduce selected thermodynamic properties of the system, and only a few structural properties. Therefore thermodynamics-based approaches have been often used for modeling biological environments. The model could be later combined with a structure-based CG description of hydrophilic groups, resulting in a CG model of amphiphilic molecules such as diblock copolymers (poly(ethyleneoxide)-poly(ethylethylene)) [32].

Different methodologies have been used in the development of CG models for polymers. One of the main advantage for introducing CG modelling in polymer studies is the ability to expand available molecule sizes for exploration as well as extend simulation time to polymer chain-solution equilibrium state. Most of the development of CG models has so far been focused on typical benchmark chains like polycarbonates, polystyrene, polyamide, etc [32]. However, CG modelling was successfully applied in various polymer studies and more systems are now being adopted for CG simulations. These include poly(γ-glutamyl-glutamate) paclitaxel conjugates, such as PGG-PTX [32], PGG-PTX-PEG-npRGD [47], polyethylene glycol-based polymer conjugates

[48-49], Poly(ethylene oxide)−Poly(propylene oxide)− Poly(ethylene oxide) (PEO−PPO−PEO) Block Copolymers [48], and amphipols (APols) [50].

1.4.3 MARTINI coarse-grained modelling method

An example of thermodynamics-based CG models is the MARTINI force field, which was originally developed for lipids and then extended to proteins, fullerenes and carbohydrates and later extended for use with polymers, bi-layers and other systems [51-55]. In this approach, interaction parameters are determined by reproducing densities and free energies of partitioning. With respect to other methods, this offers a number of advantages. Parameterization of nonbonded interactions is based on experimental data, and thus does not rely entirely on AA results and its possible errors. Non-bonded interactions are modelled by simple Lennard-Jones and Coulomb functions, therefore, the parameters developed for new building blocks can be made compatible with existing MARTINI ones. However, introducing new groups requires parameterization of the developed model against known parameters. For example, parametrization of the bonded interactions for the copolymer chain can be performed using distance, angle, and dihedral distributions obtained from the atomistic simulations of the homopolymer chains.

The MARTINI model is based on a four-to-one mapping, *i.e.* on average four heavy atoms are represented by a single interaction centre. The four-to-one mapping is often chosen as an optimum between computational efficiency on the one hand and chemical representability on the other hand. Mapping of water is consistent with this choice, as four water molecules are mapped to a CG water bead. Ions are represented by a single CG bead, which represents both the ion and its first hydration shell. Ring-like molecules are mapped with a higher resolution of up to two non-hydrogen atoms to one MARTINI particle. (Figure 1.13)

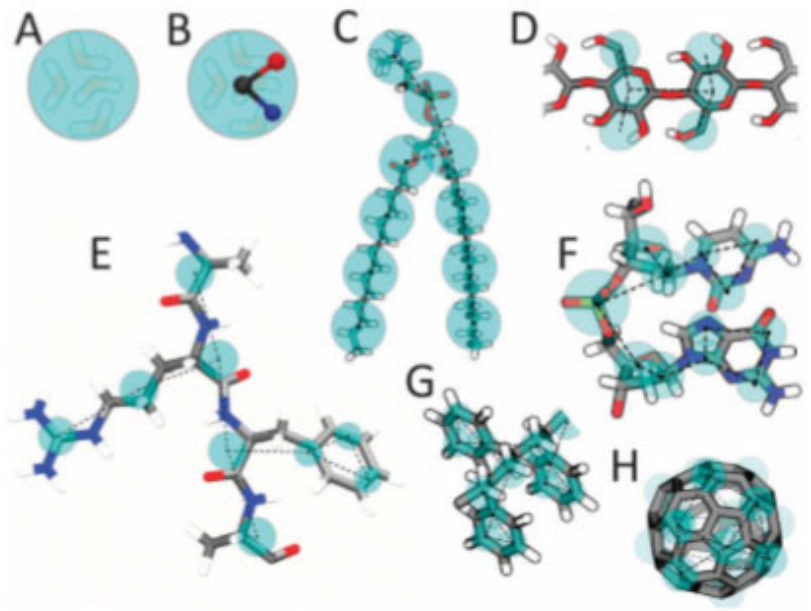


Figure 1.13. MARTINI mapping examples A) Standard water particle representing four water molecules. (B) Polarizable water molecule with embedded charges. (C) DMPC lipid. (D) Polysaccharide fragment. (E) Peptide. (F) DNA fragment. (G) Polystyrene fragment. (H) Fullerene.

Based on the chemical nature of relative structure, the CG beads are assigned a specific particle type with polar character. The MARTINI model has four main types of particle: polar (P), non-polar (N), apolar (C), and charged (Q). Within each type, several sub-types are presented and can be described either by a letter denoting the hydrogen-bonding capabilities (d = donor, a = acceptor, da = both, 0 = none) or by a number indicating the degree of polarity (from 1 = low polarity to 5 = high polarity), giving a total of 18 particle types or 'building blocks'. Non-bonded interactions are described by a Lennard-Jones (LJ) 12-6 potential. The strength of the interaction, determined by the value of the LJ well-depth ε_{ij}, depends on the interacting particle types i and j. The value of ε_{ij} ranges from 5.6 kJ mol^{-1} for interactions between strongly polar groups to 2.0 kJ mol^{-1}, mimicking the hydrophobic effect. The effective size of the particles is governed by the LJ parameter s = 0.47 nm for all normal particle types. For the special class of particles in ring-like molecules, slightly reduced parameters are defined to model ring–ring interactions: s = 0.43 nm,

and ε_{ij} is scaled to 75% of the standard value. Detailed information on MARTINI mapping logic and parameter selection and validation can be found in Chapter 3.

The MARTINI model has been thoroughly validated against experimental and all-atom simulation data on diffusivities, hydrodynamics, end-to-end-distances, bond lengths, angles, and dihedrals for a range of synthetic polymers [51-58]. Besides, MARTINI has been parameterized for a large number of chemical building blocks, which offers the possibility to build-up CG models of a wide range of molecules. MARTINI in many cases is comparable in accuracy to atomistic models, particularly in thermodynamics, but also has a number of limitations such as lack of electronic polarizability in the standard force field and inability to obtain parameters, such as particular atom distribution profile or information on some of the important interactions. For example, cation–π interactions or the strong electrostatically driven interactions between aromatic molecules are difficult to include in a coarse-grained force field to reproduce both the strength of these interactions and the geometries resulting from these interactions. However, the wide use of MARTINI and extensive testing in the original papers clearly shows the degree of agreement with experiments and atomistic simulations and allows an assessment of whether MARTINI is accurate enough for a particular application, such as polymer conjugates structure modelling and solvent behavior simulation

Currently, there is a growing number of basic polymer systems for which parameters have been derived, including polyethyleneglycol (PEG), polystyrene, triblock copolymers polyethyleneoxide–polypropyleneoxide– polyethyleneoxide (PEO–PPO–PEO), polyurea, Nafion ionomers, polyester coatings composed of two dicarboxylic acids and a diol, esterified to neopentyl glycol monomers and crosslinked by hexa(methoxymethyl)melamine, polymer nanofibres composed of nylon-6 (polycaprolactam), PAMAM dendrimers, PEG-conjugated PAMAM dendrimers and amphipols [56]. At the same time HPMA copolymers haven't been investigated using MARTINI-based models. This lack is one of the main motivations for this project as arising need in such research is now clear.

Clarification of project aims [added by supervisors]

1.5 AIMS

The goal of the research was to develop molecular modelling techniques to predict the conformation of complex substituted polymer structures in solution, focusing on polymer-drug conjugates as a target system. There were several discrete aims contributing to this:

1) The synthesis and characterization of the solution behaviour of HPMA-GFLG based polymer-drug mimic conjugates, using mixed drug-mimics to discern the cumulative effect on solution morphology adopted.

2) The development of appropriate molecular dynamics approaches to study large molecules in solution, including identification of the appropriate parameter sets, solvent models and analytical methods for data interpretation.

3) The comparison of experimental (particularly small-angle neutron scattering data) and modelling data in order to validate the modelling approach to determining solution conformation.

4) The translation of these tools to coarse-grained modelling of conjugates as a predicitive tool for conformation of novel conjugates.

1.5 References

[1] S.I. Hajdu, M.J. Thun, L.M. Hannan, A. Jemal, Cancer, 2011, 117 (5): 1097–102

[2] "Cancer Fact sheet N°297". World Health Organization. February 2014. Retrieved 10 June 2014.

[3] Cancer Facts & Figures 2015. Atlanta: American Cancer Society; 2015.

[4] The Power of Prevention Chronic disease . . . the public health challenge of the 21st century National Center for Chronic Disease Prevention and Health Promotion, 2009

[4] R. Duncan, Nat Rev Drug Discov. 2003 May; 2(5):347-60.

[5] L. W. Seymour, K. Ulbrich, P. S. Steyger, M. Brerenton, V. Subr, J. Strohalm, R. Duncan, Br. J. Cancer 1994, 70, 636–641

[6] T. Minko, P. Kopeckova, V. Poz- harov, J. Kopecek, J. Controlled Release 1998, 54, 223–233.

[7] R. B. Greenwald, Y. H. Chloe, McGuire, C. D. Conover, Adv. Drug Delivery Rev. 2003, 55, 217–250.

[8] D. Hreczuk-Hirst, D. Chicco, L. German, R. Duncan, Int. J. Pharm. 2001, 230, 57–66.

[9] R. Duncan. Polymeric Drug Delivery Systems (Ed.: G. S. Kwon), Marcel Dekker, New York, 2005, pp. 1–92.

[10] L. Varticovski, Z. R. Lu, K. Mitchell, I. De Aos, J. Kopecek, J. Controlled Release 2001, 74, 275–281;

[11] M. J. Vicent, S. Manzanaro, J. A. De la Fuente, R. Duncan, J. Drug Targeting 2004, 12, 503–515.

[12] R. Haag and F. Kratz. Angew. Chem. Int. Ed. 2006, 45, 1198–1215

[13] Y. Noguchi, J. Wu, R. Duncan, J. Strohalm, K. Ulbrich, T. Akaike and H. Maeda. Japanese J. Cancer Res, 1998, 89, 307-314.

[14] L. W. Seymour, R. Duncan, J. Strohalm and J. Kopecek. J. Biomed. Mater. Res., 1987, 21, 1341-1358

[15] Y. Chau, DSpace@MIT: Massachusetts Institute of Technology, 2005. Web. 12 Dec. 2012.

[16] C. De Duve. Scientific American, 1963, 208, 64-72.

[17] Y. Luo, G. D. Prestwich, Curr Cancer Drug Targets. 2002 Sep; 2(3):209-26.

[18] H. Ringsdorf, J. Polymer Sci. Part C-Polymer Symposium, 1975, 135-153.

[19] V. Hofmann, H. Ringsdorf and Schaumloffel. Unpublished Work.
[20] A. Abuchowski, T. Van Es, N.C. Palczuk, F.F. Davis, J. Biol. Chem., 1977, 252, 3578-3581.
[21] L.Gros, H.Ringsdorf, H.Schupp, Angew.Chem.1981,93, 311– 331
[22] A. Paul, A., M. J. Vicent, R. Duncan, Biomacromolecules 2007, 8(5), pp. 1573-1579.
[23] A. Paul, Biomacromolecules 2010, 11(8), pp. 1978-1982.
[24] F. Kratz, U. Beyer and M. T. Schutte, Critical Reviews in Therapeutic Drug Carrier Systems, 1999, 16, 245-288.
[25] V. Giménez et al. J. Controlled Release 2012 159(2), pp. 290-301.
[26] J. D. Byrne, T. Betancourt, L. Adv Drug Deliv Rev. 2008; 60(15):1615-26
[27] P. A. Vasey, S. B. Kaye, R. Morrison, C. Twelves, P. Wilson, R. Duncan, A. H. Thomson, L. S. Murray, T. E. Hilditch, T. Murray, S. Burtles, D. Fraier, E. Frigerio, J. Cassidy, Clin. Cancer Res. 1999, 5, 83–94.
[28] K. Ulbrich, C. Konak, Z. Tuzar, J. Kopecek. Makromolekulare Chemie- Macromolecular Chemistry and Physics, 1987, 188, 1261-1272.
[29] F. Greco, M. J. Vicent, N. A. Penning, R. I. Nicholson, R, Duncan J Drug Target. 2005; 13(8-9):459-70.
[30] P. G. De Gennes, Scaling Concepts in Polymer Physics, Cornell University Press, Ithaca and London, 1979
[31] P. J. Flory, Principles of Polymer Chemistry, Cornell University Press, Ithaca, 1953;
[32] L. X. Peng, A. Ivetac, A. S. Chaudhari, S. Van, G. Zhao, L. Yu, S. B. Howell, J. A. McCammon, D. A. Gough Biopolymers 2010; 93 (11):936-951
[33] A. Guinier, C.R. Hebd, Séances Acad. Sci. 1937, 204, 1115.
[34] L. A. Feigin, D. I. Svergun. Structure Analysis by Small-Angle X-Ray and Neutron Scattering. Springer Science+Business Media, LLC
[35] I.S. Anderson et al. Neutron Imaging and Applications, Neutron Scattering Applications and Techniques, Springer Science Business Media, LLC 2009
[36] D. I. Svergun, M. H. J. Koch, 2003 Rep. Progr. Phys. 66 (10): 1735–1782
[37] G. J. M. Koper, Experimental and molecular dynamics characterization of dense microemulsion systems: morphology, conductivity and SAXS, Soft Matter, 2014; 10(43):8685 - 8697

[38] Y. Li, T. Hou, Current Medicinal Chemistry, 2010 Vol. 17, No. 36
[39] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, J. Comput. Physics 1977 23 (3): 327–341.
[40] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman J. Am. Chem. Soc. 1995 117: 5179–5197
[41] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus J Comp Chem 1983, 4 (2): 187–217.
[42] W. L. Jorgensen, J. Tirado-Rives J. Am. Chem. Soc. 1988 110 (6): 1657–1666.
[43] T. A. Halgren, J. Comput. Chem.; 1996; 490-519
[44] Z. Luo, J. Jiang J. Controlled Release 162 (2012) 185–193
[45] A. Warshel, M. Levitt, J. Mol. Biol. 1976; 103 (2): 227–49.
[46] F. Miller-Plathe, CHEMPHYSCHEM 2002, 3, 754 - 769
[47] L. X. Peng, S. K. Das, L. Yu, S. B. Howell, D. A. Gough J Mol Model (2011) 17:2973–2987
[48] S. Nawaz, P. Carbone J. Phys. Chem. B 2014, 118, 1648−1659
[49] H. Maeda, T. Sawa, T. Konno, J. Controlled Release, vol. 74, no. 1–3, pp. 47–61, 2001.
[50] J. D. Perlmutter, W. J. Drasler II, W. Xie, J. Gao, J.-L. Popot, J. N. Sachs, Langmuir 2011, 27, 10523–10537
[51] S.J. Marrink, H.J. Risselada, S. Yefimov, D.P. Tieleman, A.H. de Vries J Phys Chem B, 111:7812-7824, 2007.
[52] S.J. Marrink, M. Fuhrmans, H.J. Risselada, X. Periole in "Coarse graining of condensed phase and biomolecular systems", G. Voth ed., CRC press, Chapter 2, 2008.
[53] A. Catte, J.C. Patterson, D. Bashtovyy, M.K. Jones, F. Gu, L. Li, A. Rampioni, D. Sengupta, T. Vuorela, P. Niemela, M. Karttunen, S.J. Marrink, I. Vattulainen, J.P. Segrest Biophys. J., 2008, 94:2306-2319
[54] M. Velinova, D. Sengupta, A. Tadjer, S.J. Marrink. Langmuir, 2011, 27:14071–14077
[55] F.J. van Eerden, D.H. de Jong, A.H de Vries, T.A. Wassenaar, S.J. Marrink. BBA Biomembranes, 2015
[56] H.I. Ingólfsson, C.A. Lopez, J.J. Uusitalo, D.H. de Jong, S. Gopal, X. Periole, S.J. Marrink. WIREs Comput. Mol. Sci., 4:225–248, 2014
[57] H. M. Senn, W. Thiel Angew Chem Int Ed Engl. 2009;48(7):1198-229
[58] B. Xia, V. Tsui, D. A. Case, H. J. Dyson, P. E. Wright, J. Biol. NMR,22, 317-331 (2002).

Chapter 2: Methods

2.1. Computer simulation methods

2.1.1. Hartree–Fock method

Hartree–Fock (HF) is a method of approximation for the determination of the wave function and energy of a many-body system in a stationary state, widely used to solve the time-independent Schrödinger equation for a multi-electron atom or molecule [1]. In order to do so the HF method introduces a series of approximations: the Born–Oppenheimer approximation assuming that the full molecular wave function is a function of the coordinates of each of the nuclei position, in addition to those of the electrons [2]. Second, relativistic effects are neglected. The momentum operator is assumed to be completely non-relativistic. The variational solution in HF is assumed to be a linear combination of a finite number of basis functions, where the finite basis set is assumed to be approximately complete. As well each energy eigenfunction is assumed to be describable by the single Slater determinant or by a single permanent of N spin-orbitals in the case of bosons. Finally, the mean field approximations are implied. Effects arising from electron correlation are completely neglected for the electrons of opposite spin, but are taken into account for electrons of parallel spin.

The Hartree–Fock method is sometimes called the self-consistent field method (SCF) following the fact that the equations in HF are solved using a nonlinear method such as iteration due to nonlinearities introduced by the method's approximation. In deriving the Hartree equation as an approximate solution of the Schrödinger equation, Hartree required the final field as computed from the charge distribution to be "self-consistent" with the assumed initial field. Thus, self-consistency is a compulsory requirement of the solution. Another basic approximation is known as so-called the Fock matrix. The Fock matrix is an approximation of the Hamiltonian operator which includes the effects of electron-electron repulsion as an average and does not take in count the electron correlation energy (2.1.). The Fock matrix is defined by the Fock operator:

$$\hat{F}(i) = \hat{h}(i) + \sum_{j=1}^{n/2} [2\hat{J}_j(i) - \hat{K}_j(i)] \qquad (2.1)$$

where:

$\hat{F}(i)$ is the Fock operator for the *i*-th electron in the system,

$\hat{h}(i)$ is the one-electron hamiltonian for the *i*-th electron,

n is the number of electrons and $\frac{n}{2}$ is the number of occupied orbitals in the closed-shell system,

$\hat{J}_j(i)$ is the Coulomb operator, defining the repulsive force between the j-th and i-th electrons in the system,

$\hat{K}_j(i)$ is the exchange operator, defining the quantum effect produced by exchanging two electrons.

Now Finding the Hartree–Fock wave functions is equivalent to solving the following eigenfunction equation:

$$\hat{F}(1)\phi_i(1) = \epsilon_i\phi_i(1) \tag{2.2}$$

where $\phi_i\ (1)$ are a set of one-electron wave functions, called the Hartree–Fock molecular orbitals. Typically, in Hartree–Fock calculations, the one-electron wave functions are approximated by a linear combination of atomic orbitals (basis set) also called Slater-type orbitals [3]. Furthermore, in the interests of saving computation time it is common for the basis sets to be composed of a linear combination of one or more Gaussian-type orbitals (GTOs), rather than Slater-type orbitals [4]. Nowadays, there are hundreds of basis sets composed GTOs. The smallest of these are called minimal basis sets, and are composed of the minimum number of basis functions required to represent all of the electrons on each atom. The largest of these can contain hundreds of basis functions on each atom which makes them relatively more accurate but sometimes not affordable for particular system [5].

Minimal basis sets are fixed and are unable to adjust to different molecular environments. So in order to obtain more accurate results it is common to represent valence orbitals by more than one

basis function, called split-valence. They arise from the work of John Pople and typically are represented as: X-YZG, where X represents the number of primitive Gaussians comprising each core atomic orbital basis function, and Y and Z indicate that the valence orbitals are composed of two basis functions each, the first one composed of a linear combination of Y primitive Gaussian functions, the other composed of a linear combination of Z primitive Gaussian functions [6].

Of the five simplifications outlined above, the final one is typically the most important. Neglecting electron correlation can lead to large deviations from experimental results. A number of approaches also address the fact that HF methods neglect electron correlation are called post-Hartree–Fock methods. For example, Møller–Plesset perturbation theory, described below, treats correlation as a perturbation of the Fock operator [7]. Others expand the true multi-electron wave function in terms of a linear combination of Slater determinants, such as multi-configurational self-consistent field, configuration interaction, quadratic configuration interaction, and complete active space SCF (CASSCF). Still others (such as variational quantum Monte Carlo) modify the Hartree–Fock wave function by multiplying it by a correlation function. Another alternative to Hartree–Fock calculations is density functional theory, which treats both exchange and correlation energies using calculations that are a hybrid of the two methods.

2.1.2 Density Functional Theory methods

Density functional theory (DFT) is a quantum chemistry method, widely used in computer modelling to investigate the electronic structure of various many-body systems, such as atoms, molecules, and condensed phases. The method is based on calculation of total electronic density distribution as a means to calculate the electronic energy. In DFT the properties of a many-electron system can be determined by using functionals which describe energy as a function of spatially dependent electron density. Modern DFT is made possible by the existence of two theorems that were introduced and proven by Hohenberg and Kohn in 1964 [8, 9].

Theorem I: For any system of interacting particles in an external potential $V_{ext}(r)$, the density is uniquely determined. These means the external potential is a unique functional of the density.

Theorem II: A universal functional for the energy E[n] can be defined in terms of the density. The exact ground state is the global minimum value of this functional. The correct density that minimizes the energy is then the ground state density.

As any other first-principle method, DFT uses the many body Schrödinger equation and is solved using various approximations such as Born-Oppenheimer approximation, local-density approximation (LDA) or Generalized gradient approximations (GGA) depending on the chosen method. Nuclei generate a static external potential V_{ext}. Therefore, a stationary electronic state is then described by a wavefunction satisfying the many-electron time-independent The Schrödinger equation of electron motion (2.3):

$$\hat{H}\Psi = \left[\hat{T} + \hat{V} + \hat{U}\right]\Psi = \left[\sum_{i}^{N}\left(-\frac{\hbar^2}{2m_i}\nabla_i^2\right) + \sum_{i}^{N} V(\vec{r}_i) + \sum_{i<j}^{N} U(\vec{r}_i, \vec{r}_j)\right]\Psi = E\Psi$$

(2.3)

where, for the N-electron system, $\hat{H}$ is the Hamiltonian, E is the total energy, $\hat{T}$ is the kinetic energy, $\hat{V}$ is the potential energy from the external field due to positively charged nuclei, $\hat{U}$ is the electron-electron interaction energy, E is the total energy, Ψ is the wave function of the system, $V_{ext}(r_i)$ is the potential of the nuclei, ∇^2 is the second derivative of position, ħ is Planck's constant over 2π and m_e is the mass of an electron. In DFT the particle density $n(r_i)$ for a normalized wavefunction Ψ is given by:

$$n(\vec{r}) = N\int \mathrm{d}^3 r_2 \cdots \int \mathrm{d}^3 r_N \Psi^*(\vec{r}, \vec{r}_2, \ldots, \vec{r}_N)\Psi(\vec{r}, \vec{r}_2, \ldots, \vec{r}_N).$$

(2.4)

Therefore for a given ground-state density $n_0(r_i)$ it is possible, in principle, to calculate the corresponding ground-state wavefunction Ψ_0:

$$\Psi_0 = \Psi(n_0)$$

(2.5)

Consequently the expectation value of an observable $\hat{O}$ in a ground state is a functional of n_0 :

$$O[n_0] = \left\langle \Psi[n_0] \middle| \hat{O} \middle| \Psi[n_0] \right\rangle .$$

(2.6)

In the particular case of ground-state energy:

$$E_0 = E[n_0] = \left\langle \Psi[n_0] \middle| \hat{T} + \hat{V} + \hat{U} \middle| \Psi[n_0] \right\rangle$$

(2.7)

where the contribution of the external potential V_{ext} can be written explicitly for the ground-state density n_0 :

$$[V_{ext}[n]]_0 = \int n_0(r) V_{ext}(r) d^3r$$

(2.8)

In the general case:

$$V_{ext}[n] = \int n(r) V_{ext}(r) d^3r$$

(2.9)

while the functionals $T[n]$ and $U[n]$ are called universal functionals, $V_{ext}[n]$ depends on the particular system of interest and is called a non-universal functional. So we can now describe the system energy as:

$$E[n] = T[n] + U[n] + \int n(r) V_{ext}(r) d^3r$$

(2.10)

By expressing the electron density in terms of single particle wavefunctions ψ_i that reproduce the density $n(r)$ of the original many-body system we get an eigenvalue equation, which is the typical

representation of the Kohn–Sham equations named after Walter Kohn and Lu Jeu Sham, who introduced the concept at the University of California, San Diego in 1965 [9]:

$$n(r) \triangleq n_S(r) = \sum_i^N |\psi_i(r)^2|$$

(2.11)

Then the kinetic energy can be written as:

$$T_S[n(r)] = -\frac{\hbar^2}{2m_e}\sum_i \int \psi_i(r)\nabla^2\psi_i(r)d^3r$$

(2.12)

The potential energy can be further divided into coulombic electron-electron interactions and exchange-correlation energy:

$$E_S[n(r)] = \frac{e^2}{2}\iint \frac{n(r)n(r')}{|r-r'|}d^3r d^3r' + E_{XC}[n(r)]$$

(2.13)

where the second term denotes the so-called Hartree term describing the electron-electron Coulomb repulsion and E_{XC} is the exchange-correlation energy. The Local Density Approximation (LDA) defines the exchange-correlation energy that depends only on the density at the coordinate where the functional is evaluated:

$$E_{XC}^{LDA}[n] = \int \varepsilon_{XC}(n)n(r)d^3r$$

(2.14)

where ε_{XC} is the exchange-correlation energy per unit volume of a uniform electron gas. The local spin-density approximation (LSDA) is a straightforward generalization of the LDA to include electron spin:

$$E_{XC}^{LSDA}\left[n_\uparrow, n_\downarrow\right] = \int \varepsilon_{XC}(n_\uparrow, n_\downarrow)n(r)d^3r$$

(2.15)

The Generalised Gradient Approximation (GGA) take into account the gradient of the density at the same coordinate:

$$E_{XC}^{GGA}\left[n_{\uparrow}, n_{\downarrow}\right] = \int \varepsilon_{XC}(n_{\uparrow}, n_{\downarrow}, \nabla n_{\uparrow}, \nabla n_{\downarrow}) n(r) d^3r \tag{2.16}$$

DFT can be considered as one of the most popular and versatile methods available in condensed-matter physics and computational chemistry due to comparatively large systems it is able to evaluate within the range of quantum chemistry methods as well as relatively low cost of corresponding calculations compared to post-Hartree Fock methods. In many applications, hybrid DFT methods are used even more widely. They are based on a use of a combination of the HF exchange energy with LSDA exchange energy with a specified amount of the exact exchange mixture. One of the most widely used DFT methods is B3LYP (Becke, three-parameter, Lee-Yang-Parr) [11]. This method operates a combination of exact exchange from Hartree-Fock theory with LSDA and Lee-Yang-Parr, LYP correlation energy.

$$E_{XC}^{B3LYP} = E_{XC}^{LDA} + a_0\left(E_X^{HF} - E_X^{LDA}\right) + a_X\left(E_X^{GGA} - E_X^{LDA}\right) + a_C\left(E_C^{GGA} - E_C^{LDA}\right) \tag{2.17}$$

where $a_0 = 0.2$, $a_X = 0.72$, and $a_C = 0.81$, E_X^{GGA} and E_C^{GGA} are generalized gradient approximations: the Becke 88 exchange functional and the correlation functional of Lee, Yang and Parr, and E_C^{LDA} is the local-density approximation to the correlation functional.

2.1.3 *Ab initio* methods

In contrast to "simple" quantum chemistry methods, *ab initio* methods can be used to converge to the exact solution. Being more expensive in the computation terms than DFT, *ab initio* calculations provide more accurate data, treating the Schrödinger equation as an eigenvalue equation of the electronic molecular Hamiltonian, with a discrete set of solutions. There are several common classes of *ab initio* electronic structure methods: Hartree–Fock methods, Post-Hartree–Fock methods and Multi-reference methods. Among them one of the most popular are Møller–Plesset perturbation theory (MPn) [7], Configuration interaction (CI) [12] and Coupled cluster (CC) [13]

methods which belongs to Post-Hartree–Fock class. They start with a Hartree–Fock calculation and subsequently correct for electron-electron repulsion, referred to also as electronic correlation.

In Møller–Plesset perturbation theory the basic idea is that the difference between the Fock operator and the exact Hamiltonian can be considered as a perturbation:

$$\hat{\mathbf{H}} = \hat{F} + \hat{V}$$

(2.18)

Corrections can be made to any order of the energy and the wavefunction (MP2, MP3, MP4, MP5):

$$E = E_{HF} + E^{(1)} + E^{(2)} + E^{(3)} + \cdots$$

$$\Psi = \Psi_{HF} + \Psi^{(1)} + \Psi^{(2)} + \Psi^{(3)} + \cdots$$

(2.19)

One of the most popular methods is the lowest level of correction, MP2.

$$E_{corr}^{MP2} = \sum_{ijkl} \frac{2(ik|jl) - (il|jk)}{(\varepsilon_i + \varepsilon_j + \varepsilon_k + \varepsilon_l)(ik|jl)}$$

(2.20)

MP2 is relatively fast for investigating systems that consist of tens and even hundreds of particles. A disadvantage is that it is not variational, so the correlation energy can be overestimated. In practice MP2 must be used with a reasonable basis set (6-31G* or better). Subsequent MP-levels MP3, MP4 and MP5 are more complicated and much more time-consuming.

2.1.4. Molecular mechanics methods

Molecular mechanics (MM) uses classical Newtonian mechanics to model molecular systems. The potential energy of all systems in molecular mechanics is calculated using force fields. In contrast to QM, electronic motion is neglected and energy is calculated as a function of nuclear positions

only. This approach is used to perform calculations on systems containing large number of atoms, where quantum methods can't be applied.In general, MM can be described using "balls on springs" model, which is based on following statements:

- Each atom (or group of atoms) has its own type and is represented by a single particle which is treated due to this type.
- Each particle is assigned a radius (typically the van der Waals radius), polarizability, and charge.
- Bonded interactions are treated as "springs" with an equilibrium distance equal to the experimental or calculated bond length

A wide range of problems can be solved by implementing MM methods to evaluate parameters of investigated system. For example, search of the global minimum of the system can be performed using force field for the system's energy. There are several force fields available for use in MM, among them: AMBER [14], CHARMM [15], GROMOS [16], OPLS [17], UFF [18], MMFF [19], MARTINI [20] and others. Selection of the most suitable force field should be made based on the system features as every force field is designed for use in particular systems.

2.1.5. Force field selection

In the context of molecular modelling, a force field is the set of parameters used to describe the potential energy of a system of particles as a function of position of these particles. Force fields can be used both for all-atom and coarse-grained simulations. "All-atom" force fields provide parameters for every atom type in a system while the "coarse-grained" force fields, which are frequently used in long-time simulations of polymers and proteins, provide a more crude representations for increased computational efficiency treating a group of atoms as a single bead. A typical force field includes both covalent bonding and nonbonded or "non-covalent" terms that describe long-range electrostatic and van der Waals forces. The specific form of the terms depends on the chosen force field, but a general form for the total energy in an additive force field is

$$E = E_{covalent} + E_{noncovalent} \tag{2.21}$$

where the components of the covalent and noncovalent contributions are given by the following summations:

$$E_{covalent} = E_{bond} + E_{angle} + E_{dihedral}$$
$$E_{noncovalent} = E_{electrostatic} + E_{van\ der\ Waals} \quad (2.22)$$

Generally the bond and angle terms are defined as harmonic potentials centered near to equilibrium bond-lengths derived either from experimental data or theoretical calculations of electronic structure performed with more accurate benchmark methods. Parameter sets may also include "improper" dihedral terms, which function as correction factors for out-of-plane deviations (for example, they can be used to keep planar molecules such as benzene or cyclopropane in appropriate planar state).

The non-bonded interactions are typically modeled using a "6–12 Lennard-Jones potential", which is based on idea of rapid fall of attractive van der Waals forces with distance as r^{-6} and repulsive forces rise as r^{-12}, where r represents the distance between two atoms. Generally, a cutoff radius is used to speed up the calculation so that atom pairs whose distances are greater than the cutoff value will not nave non-bonded interaction. It is shown on Figure 2.1, where ε is the depth of the potential well, σ is the finite distance at which the inter-particle potential is zero, r is the distance between the particles and r_m is the distance at which the potential reaches its minimum value. The electrostatic terms do not fall off rapidly with distance, so their basic functional form is the Coulomb potential, which falls off as r^{-1}. A cutoff radius similar to that used for the van der Waals terms may be implemented, but in order to obtain more accurate simulation results more sophisticated and computationally intensive methods such as particle mesh Ewald (PME) and the multipole algorithm may be used.

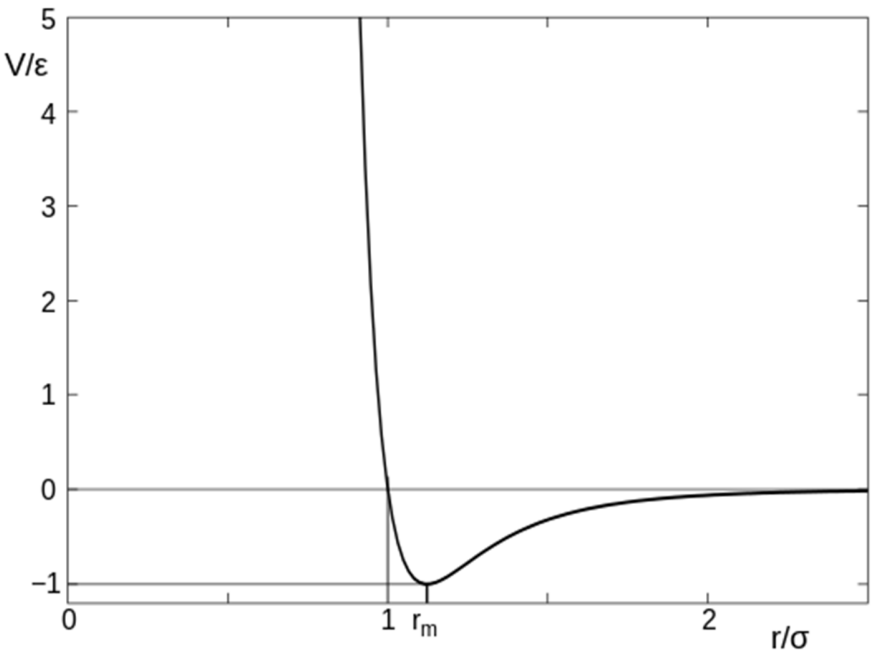


Figure 2.1. Typical shape of 12-6 Lennard-Jones potential in MM

As stated previously, different force fields are designed for different purposes. For example, MM2 was developed by primarily for conformational analysis of small organic molecules [48]. It is designed to reproduce the equilibrium geometry of such molecules as precisely as possible. Another set, called CFF was developed by , Lifson and coworkers as a general method for unifying studies of energies, structures and vibration of general molecules and molecular crystals [49]. AMBER, CHARMM and GROMOS have been developed primarily for molecular dynamics of biomolecules, although they can be also applied to perform energy minimization. The term "AMBER force field" generally refers to the functional form used by the family of AMBER force fields originally developed by Peter Kollman's group at the University of California, San Francisco [14]. The functional form of the AMBER force field is quite common, and can be written as:

$$V(r^N) = \sum_{\text{bonds}} k_b(l - l_0)^2 + \sum_{\text{angles}} k_a(\theta - \theta_0)^2$$

$$+\sum_{\text{torsions}}\frac{1}{2}V_n[1+\cos(n\omega-\gamma)]$$

$$+\sum_{j=1}^{N-1}\sum_{i=j+1}^{N}\left\{\epsilon_{i,j}\left[\left(\frac{r_{0ij}}{r_{ij}}\right)^{12}-2\left(\frac{r_{0ij}}{r_{ij}}\right)^{6}\right]+\frac{q_iq_j}{4\pi\epsilon_0 r_{ij}}\right\}$$

(2.23)

OPLS (Optimized Potentials for Liquid Simulations) was developed by Jorgensen for treating interatomic forces in simulations of organic molecules and biopolymers [17]. Its functional form is similar to AMBER and can be written as:

$$V(r^N)=\sum_{\text{bonds}}K_r(r-r_0)^2+\sum_{\text{angles}}k_\theta(\theta-\theta_0)^2+$$

$$\frac{V_1}{2}[1+\cos(\phi-\phi_0)]+\frac{V_2}{2}[1-\cos 2(\phi-\phi_0)]+\frac{V_3}{2}[1+\cos 3(\phi-\phi_0)]+\frac{V_4}{2}[1-\cos 4(\phi-\phi_0)]$$

$$+\sum_{i>j}f_{ij}\left(\frac{A_{ij}}{r_{ij}^{12}}-\frac{C_{ij}}{r_{ij}^{6}}+\frac{q_iq_je^2}{4\pi\epsilon_0 r_{ij}}\right)$$

(2.24)

Several sets of OPLS parameters have been published. Among them OPLS-ua (united atom), which includes hydrogen atoms next to carbon implicitly in the carbon parameters, and can be used to save simulation time and OPLS-aa (all atom) includes every atom explicitly. A distinctive feature of the OPLS parameters is that they were optimized to fit experimental properties of liquids, such as density and heat of vaporization, in addition to fitting gas-phase torsional profiles.

2.1.6. Molecular dynamics simulations

Molecular dynamics is a computer simulation method where the trajectories of atoms and molecules are determined by numerically solving the Newton's equations of motion (Eq. 2.25) for a system of interacting particles. Interaction forces between the particles and potential energy in MD simulations are typically defined by molecular mechanics force fields. The method was

originally developed within theoretical physics and was later implemented for computer calculation, and is applied today in chemical physics, materials science and biochemistry. First introduced by Alder and Wainwright in the late 1950's for study of the interactions between hard spheres, MD was continuously developed during these years [21]. The next major advance was in 1964, when Rahman carried out the first simulation using a realistic potential for liquid argon [22]. Finally, the first molecular dynamics simulation of a molecular system was done by Rahman and Stillinger in their simulation of liquid water in 1974 [23].

$$F_i = m_i * a_i = m_i * \frac{dv_i}{dt} = m * \frac{d^2 r_i}{dt^2}$$
$$F_i = -\nabla_i E$$

(2.25)

The first protein simulation was carried in 1977 on bovine pancreatic trypsin inhibitor (BPTI) [24]. Since then, the number of simulation techniques has greatly expanded; there exist now many specialized techniques for particular problems, including mixed quantum mechanical - classical simulations (QM/MM), coarse-grain MD simulations for complex biological systems, etc. There are several reasons for such an intensive development: one is the ability to evaluate behaviour of the systems that was not reachable within quantum chemistry techniques. Another is that results of MD simulations can be combined with experiments such as X-ray crystallography, SANS and NMR structure determination to obtain details on morphological properties of chosen systems. Despite these advantages, running long MD simulations for large structures (tens of thousands atoms or more) can be complicated and generate cumulative errors in numerical integration that must be minimized with proper selection of algorithms and parameter sets.

Molecular dynamics simulations generate information about studied system at the microscopic level, which includes atomic positions and velocities. The conversion of parameters to macroscopic observables such as pressure, energy, heat capacities, etc., requires the use of statistical mechanics which provides mathematical expressions that relate macroscopic properties to the distribution and motion of the atoms and molecules of the N-body system. For a system with N atoms in a volume Ω, we can define its internal energy: E ≡ K +U, where K is the kinetic energy,

$$K \equiv \sum_{i=1}^{N} \frac{1}{2} m_i [x_i(t)]^2$$

(2.26)

and U is the potential energy,

$$U = U(x^{3N}(t))$$

(2.27)

where $x^{3N}(t)$ denotes the collective of 3D coordinates $x_1(t)$, $x_2(t)$, . . . , $x_N(t)$. During a process of MD simulation, the evaluation of the potential as a function of the internal coordinates of atoms can be considered as the most computationally intensive task, and the most expensive to evaluate is the non-bonded or non-covalent contributions to a total energy. This computational cost can be reduced by employing electrostatics methods such as Particle Mesh Ewald [25], P3M [26] or other spherical cutoff techniques.

Another factor that impacts total CPU time required by a simulation is the size of the integration timestep chosen for the simulation. It must be small enough to avoid discretization errors, which can occur if less than the fastest vibrational frequency in the evaluated system. Too small timestep will also have an influence on simulation speed so it tends to be as large as possible for particular simulation conditions. Typical timesteps for classical all-atom MD simulations are usually selected within the range of 1-2 femtoseconds, which can be extended by using algorithms such as SHAKE [27], which fix the vibrational movement of the fastest hydrogen atoms. Total simulation time should be long enough to be relevant to the timescales of the processes being studied: specifically, the simulated time should be enough for the system to reach its equilibrium state. For example, recent publications about the dynamics of proteins and DNA use data from simulations of nanoseconds (10^{-9} s) to microseconds (10^{-6} s). To obtain these simulations, several CPU-days to CPU-years are needed. Parallel algorithms, such as Domain Decomposition allow the calculations to be distributed among CPUs, allowing to massively increase simulation speed.

An important condition for running MD simulation is that the simulation box size must be large enough to avoid boundary condition artifacts, such as self-interaction. Boundary conditions are often treated by choosing fixed values at the edges or by implementing periodic boundary conditions (PBC) in which simulation box is repeated in all directions containing the same amount of particles with the same coordinates to mimic bulk phase conditions. The minimum-image convention is a common form of PBC particle tracking in which each individual particle interacts with the closest image of the remaining particles in the system.

2.1.6.1. Microcanonical ensemble (NVE)

In the microcanonical, or NVE ensemble, the system is isolated from changes in moles (N), volume (V) and energy (E). It corresponds to an adiabatic process with no heat exchange. A microcanonical molecular dynamics trajectory may be seen as an exchange of potential and kinetic energy, with total energy being conserved. For a system of N particles with coordinates $\boldsymbol{X}$ and velocities $\boldsymbol{V}$, the following pair of first order differential equations may be written in Newton's notation as

$$F(X) = -\nabla U(X) = M\dot{V}(t)$$
$$V(t) = \dot{X}(t).$$

(2.28)

The potential energy function $U(X)$ of the system is a function of the particle coordinates $\boldsymbol{X}$. The first equation comes from Newton's laws; the force $\boldsymbol{F}$ acting on each particle in the system can be calculated as the negative gradient of $U(X)$. For every timestep, each particle's position $\boldsymbol{X}$ and velocity $\boldsymbol{V}$ may be integrated with a symplectic method such as Verlet. The time evolution of $\boldsymbol{X}$ and $\boldsymbol{V}$ is called a trajectory. Given the initial positions (e.g. from theoretical knowledge) and velocities (e.g. randomized Gaussian), we can calculate all future (or past) positions and velocities.

One frequent source of confusion is the meaning of temperature in MD. Commonly we have experience with macroscopic temperatures, which involve a huge number of particles. But temperature is a statistical quantity. If there is a large enough number of atoms, statistical

temperature can be estimated from the instantaneous temperature, which is found by equating the kinetic energy of the system to $nk_B T/2$, where n is the number of degrees of freedom of the system. A temperature-related phenomenon arises due to the small number of atoms that are used in MD simulations. For example, consider simulating the growth of a copper film starting with a substrate containing 500 atoms and a deposition energy of 100 eV. In the real world, the 100 eV from the deposited atom would rapidly be transported through and shared among a large number of atoms (10^{10} or more) with no big change in temperature. When there are only 500 atoms, however, the substrate is almost immediately vaporized by the deposition. Something similar happens in biophysical simulations. The temperature of the system in NVE is naturally raised when macromolecules such as proteins undergo exothermic conformational changes and binding.

2.1.6.2. Canonical ensemble (NVT)

In the canonical ensemble, moles (N), volume (V) and temperature (T) are conserved. It is also sometimes called constant temperature molecular dynamics (CTMD). In NVT, the energy of endothermic and exothermic processes is exchanged with a thermostat. A variety of thermostat methods are available to add and remove energy from the boundaries of an MD system in a more or less realistic way, approximating the canonical ensemble. Popular techniques to control temperature include velocity rescaling, the Nosé-Hoover thermostat, Nosé-Hoover chains, the Berendsen thermostat, the Andersen thermostat and Langevin dynamics. Note that the Berendsen thermostat might introduce the “flying ice cube effect”, which leads to unphysical translations and rotations of the simulated system. It is not trivial to obtain a canonical distribution of conformations and velocities using these algorithms. How this depends on system size, thermostat choice, thermostat parameters, time step and integrator is the subject of many articles in the field.

2.1.6.3. Isothermal–isobaric (NPT) ensemble

In the isothermal–isobaric ensemble, moles (N), pressure (P) and temperature (T) are conserved. In addition to a thermostat, a barostat is needed. It corresponds most closely to laboratory conditions with a flask open to ambient temperature and pressure. In the simulation of biological membranes, isotropic pressure control is not appropriate. For lipid bilayers, pressure control occurs under constant membrane area (NPAT) or constant surface tension "gamma" (NPγT).

2.1.6.4. Generalized ensembles

There are several integration algorithm available for performing MD simulations. All the integration algorithms assume the positions, velocities and accelerations can be approximated by a Taylor series expansion:

$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \cdots$$
$$v(t+\delta t) = v(t) + a(t)\delta t + \frac{1}{2}b(t)\delta t^2 + \cdots$$
$$a(t+\delta t) = a(t) + b(t)\delta t + \cdots \quad (2.29)$$

Where r is the position, v is the velocity, a is the acceleration. To derive the Verlet algorithm one can write

$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2$$
$$r(t-\delta t) = r(t) - v(t)\delta t + \frac{1}{2}a(t)\delta t^2 \quad (2.30)$$

Summing these two equations, we obtain

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + a(t)\delta t^2 \quad (2.31)$$

The Verlet algorithm uses positions and accelerations of particles at time *t* and the positions from time *t-dt* to calculate particle's coordinates at time *t+dt.* There is no explicit velocity in Verlet algorithm. Among the advantages of this method is the modest storage requirements and general simplicity. The disadvantage is that the use of the algorithm has only moderate precision. In the Leap-frog algorithm algorithm, the velocities are first calculated at time *t+1/2dt*; these are used to calculate the positions, *r*, at time *t+dt*.

$$r(t+\delta t) = r(t) + v(t+\frac{1}{2}\delta t)\delta t$$

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

(2.32)

In this way, the velocities *leap* over the positions, then the positions *leap* over the velocities. The advantage of this algorithm is that the velocities are explicitly calculated, however, the disadvantage is that they are not calculated at the same time as the positions. The velocities at time *t* can be found from the relationship:

$$v(t)=\tfrac{1}{2}[v\left(t-\tfrac{1}{2}\delta t\right)+v\left(t+\tfrac{1}{2}\delta t\right)]$$

(2.33)

Another integration algorithm is the Velocity Verlet algorithm. This algorithm yields positions, velocities and accelerations at time *t*.

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

$$v(t+\delta t)=v(t)+\frac{1}{2}[a(t)+a(t+\delta t)]\delta t$$

(2.34)

The replica exchange MD (REMD) method is a generalized ensemble, also called parallel tempering, that was originally created to deal with the slow dynamics of disordered spin systems. REMD tries to overcome the multiple-minima problem by exchanging the temperature of non-interacting replicas of the system running at several temperatures.

2.1.6.5. Water Models

There are different types of water models that can be used within the MD simulation. Water models can represent solvent explicitly or implicitly, mimicking solvation effects with mean field terms. In most explicit water models the water molecule is modelled as a three-centre flexible or rigid-body model. To mimic solvent effect implicitly the general class of reaction field methods are used. They are based on the idea that around each molecule there is a 'cavity' within which the Couloumb interactions are treated explicitly. Outside of this sphere the medium is assumed to have a uniform dielectric constant. First it was introduced by Barker and Watts in 1973 [28]. The effective pairwise potential in R-field is:

$$U_{AB} = q_A q_B \left[\frac{1}{r_{AB}} + \frac{(\varepsilon_{RF} - 1) r_{AB}^2}{(2\varepsilon_{RF} + 1) r_c^3} \right]$$

(2.35)

where r_c is the cut-off radius. The reaction field in the center of the cavity is given by :

$$E_{RF} = \frac{2(\varepsilon_{RF} - 1)}{2\varepsilon_{RF} + 1} \frac{\vec{M}}{r_c^3}$$

(2.36)

where $\vec{M} = \sum \mu_i$ is the total dipole moment of all the molecules in the cavity. The polarizable continuum model (PCM) is a particular case of R-field method and is often used in quantum and MD simulations. There are two types of PCM that were introduced so far: dielectric PCM (D-PCM) and conductor-like PCM (C-PCM) which treat the continuum as a polarizable dielectric or conductor-like. The molecular free energy of solvation is computed as the sum of three terms:

$$G_{solvation} = G_{elec} + G_{d-r} + G_{cav}$$

(2.37)

where G_{elec} is electrostatic, G_{d-r} is dispersion-repulsion and G_{cav} is cavitation free energy.

2.1.7. Coarse-graining techniques

In order to minimize computational costs as well as accessing sizes of a “real” polymer and time scales, the number of calculated particles were reduced using coarse-grained (CG) techniques. Given the selected research object – HPMA polymer conjugates, as well as previous polymer studies that involve application of CG models, the MARTINI force field was selected for the CG parameterization. This has had success in application to proteins and various polymers by experimental validation of their structural properties, reasonable correlation with AA simulation results and the reproduction of polymers thermodynamics [20]. The MARTINI force field dictates that a group of roughly 4–5 atoms are represented as an interaction centre, or bead. These beads interact through a set of short-ranged Lennard-Jones potentials to reproduce characteristic

properties resulting from AA simulations. Charged groups interact via a Coulombic energy function, and bonded potentials are used to describe the chemical connectivity of the beads.

2.1.8. Software and methods

Initial structures of HPMA monomer and dimer were generated using the polymer builder function in MOE 2009.10 [29] and DFT-optimized with B3LYP [30, 31] and 6-31G(d) basis set using Gaussian03 [32]. *Ab initio* and DFT calculations were performed using Turbomole 5.10 [33] and Gaussian03. Relative conformational energies were tested against MP2/def2-TZVP and MP2/aug-cc-pVTZ data, while rotational energy barriers were tested against B3LYP/6-31G(d) and BP86/6-31G(d) [34-37] data. Rigid potential energy scans, in which all geometrical parameters other than the one being scanned are fixed at their equilibrium geometry, were employed in the latter for better comparison with force field results. *Ab initio* data was generated in the gas phase only, whereas the effect of aqueous solvation on DFT data employed the polarizable continuum model (PCM) [38]. Force field tests were carried out using AMBER99 [14], MMFF94 [19] and OPLS-aa [17] parameter sets in both gas-phase and reaction field model of aqueous solvation.

Structures of larger HPMA oligomers and polymers were built using the Polymer Builder module of MOE. All polymers were constructed as a racemic mix of S- and R- monomers with random distribution of chirality. Atomic partial charges were assigned through MOE using parameters from AMBER99 force field, and built polymers were energy minimized with AMBER99 force field. MD simulations were carried out in DL-POLY 4 [39], starting from minimized structures. AMBER99 parameters were converted for DL-POLY using an in-house utility [40]. The conversion required was validated by comparing results of bond stretch, angle bending, dihedral, electrostatic and total energies calculated by DL-POLY with AMBER and MOE packages.

Simulations in DL-POLY were performed using periodic boundary conditions in cubic boxes with sizes of 20x20x20 Å for smallest HPMA oligomers to a 600x600x600 Å for HPMA-265. Starting structures were equilibrated using zero-temperature minimization algorithm running MD simulation at 10 K for 0.2 ns. Both minimization and further MD simulation for implicit solvent models were run using isothermal-isochoric (NVT) Nose-Hoover ensemble with 0.5 ps relaxation constant [41]. For explicit solvents minimization were carried in isothermal-isobaric (NPT)

ensemble with both thermostat and barostat relaxation constants equal to 0.5 ps. After equilibration temperature was scaled to 310 K and MD simulation run with a timestep of 2 fs and all bonds constrained using the SHAKE algorithm [42]. Domain decomposition based on the link cell algorithm of Hockney and Eastwood was chosen as a parallelization technique for MD calculations [26]. Simulations were performed using 8 to 64 processors in a) vacuum, b) a reaction field representation and c) explicit water [43, 44]. Trajectory snapshots were saved every 1000 timesteps for later reprocessing. Each cycle of the MD simulations was run for 0.2 - 2 ns and this step was repeated until equilibrium in R_g and ellipsoid semi-axis parameters was reached.

2.1.9. Development of programming tools

MD simulations were analysed using the in-house "analyse_hist" tool [40], generating radius of gyration and radial distribution functions. In addition we have developed shape analysis based on moments of inertia. Here the moments of inertia matrix is calculated for each frame of the saved trajectory and diagonalised to give eigenvalues and three mutually perpendicular eigenvectors. These eigenvectors are then used as an axis system centred on the centre of mass of the polymer. The extent of the polymer in each of the three directions is found by calculating the furthest atom from the centre of mass in each direction. These distances are used to define an ellipsoid:

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1 \tag{2.38}$$

where a, b and c are the furthest atom distances in the x, y and z directions respectively.

To produce a "density profile" we consider the volume occupied by the polymer to consist of a series of concentric ellipsoidal shells. The shell to which an atom belongs is identified by putting the atom co-ordinates relative to the centre of mass in the inertial axis system into the left hand side of the equation for the ellipsoid. The fractional result can be used to assign the atom to a particular shell. By carrying out this process on all atoms we arrive at a density profile and, provided the shape of the polymer has reached equilibrium, averaging over frames of the trajectory can improve the statistical significance of the profiles produced. Such profiles are used to discuss

structural aspects of the polymer system, for example if a core region exists in which one particular atom type is concentrated. All developed scripts and parts of code are provided in Appendix.

2.2 Experimental methods and materials

2.2.1 Materials

2.2.1.1. Poly N-(2-hydroxypropyl)methacrylamide parent polymer

N-(2-Hydroxypropyl)methacrylamide (HPMA) copolymer precursors containing two different loadings of ortho-nitrophenol (ØNp) terminated Gly-Phe-Leu-Gly peptidyl side chains (Figure 2.2) were purchased from Polymer Laboratories Ltd. U.K. The degree of side chain loading was found by UV spectroscopy (ε270 nm = 9500 L/mol/cm) to be 2.6 and 7.8 mol %. These values have been used throughout this work whenever side-chain concentrations have been needed. However, by convention, polymers are referred to hereafter by the original feedstock ratios of 5 and 10 mol%. Molecular weights for the 5 and 10 mol % copolymers were 45770 g/mol and 38978 g/mol respectively with polydispersities of 1.74 and 1.46 Mw/Mn.

2.1.1.2. Linear amines

1-aminohexane (C6), 1-aminooctane (C8), 1-aminodecance (C10), 1-aminododecane (C12) and 1-aminotetradecane (C14) were purchased from Sigma Aldrich and used as received (> 99.5 % purity).

2.1.1.3. Aromatic amines

2-aminoanthracene (Anc, 99 %), 2-aminoanthraquinone (Anq, 97 %) and 6-aminochrysene (Ac, 97%) were purchased from Sigma Aldrich and used as received (> 99.5 % purity).

2.1.1.4. Amino-alcohols

1-Aminohexan-1-ol (C6-OH), 1-aminooctan-1-ol (C8-OH) and 1-aminodecan-1-ol (C10-OH) were purchased from Tokyo Chemical Industry. 1-aminododecan-1-ol (C12-OH) was purchased from Toronto Research Chemicals Inc. All were used as received (> 99.5 % purity)

2.1.1.5. Adamantane derivatives

1 – Adamantine (Ad), 3-Aminoadamantan-1-ol (AdOH), 3- methyl - 1 – aminoadamantane ($AdCH_3$) were purchased from Fluorochem. All were used as received (> 99.5 % purity)

2.1.1.6. Other reagents

1-Amino-2-propanol (Ap) and triethylamine were purchased from Sigma Aldrich and kept over molecular sieves. Dialysis tubing was benzoylated with a Mw cut-off of 2000 g/mol. All other solvents were purchased from Sigma Aldrich and used as received. Any water used was of Millipore Grade.

2.2.1. Methods

2.2.1. Aminolysis reaction

The nitrophenol group of the HPMA parent polymers above were substituted with a library of different amines using an aminolysis reaction described below (Figure 2.2) based on that utilised by Vicent et al. [45].

2.2.1.1. Aminolysis reaction

5 or 10 mol % HPMA copolymer precursor (1 eq. in respect to ØNp) was dissolved in a minimal volume of anhydrous dimethyl sulphoxide (DMSO) transferred into a round-bottom flask and flushed with nitrogen. The flask was placed in the oil bath and set to stir at 45°C with magnetic stirrer until polymer was completely dissolved. 1.1 eq of chosen amine and 1 eq. of triethylamine were added to the reaction mixture, and stirred for three to five hours under nitrogen atmosphere at the same temperature with constant spectroscopic monitoring of evolution of free nitrophenol (λmax = 430nm). After reaction was complete (level of nitrophenol constant) the mixture was quenched with 3 eq. of 2-aminopropanol (1.028 uL) and left to stand for 1 hour. The obtained mixture was then precipitated into a 45 ml of vigorously stirred mixture of diethyl ether:acetone (8:1), the suspension filtered, the solid dissolved in minimal water and purified by dialysis against H2O (MW cut-off 2000 g*mol-1) for 2-3 days (6-8 water changes). Purified polymer was then freeze dried to give the conjugate as an off-white solid. Typical yields were 60 % based on polymer

starting weight. UV/VIS and 1H NMR in D_2O were used to confirm complete substitution of the ONp groups.

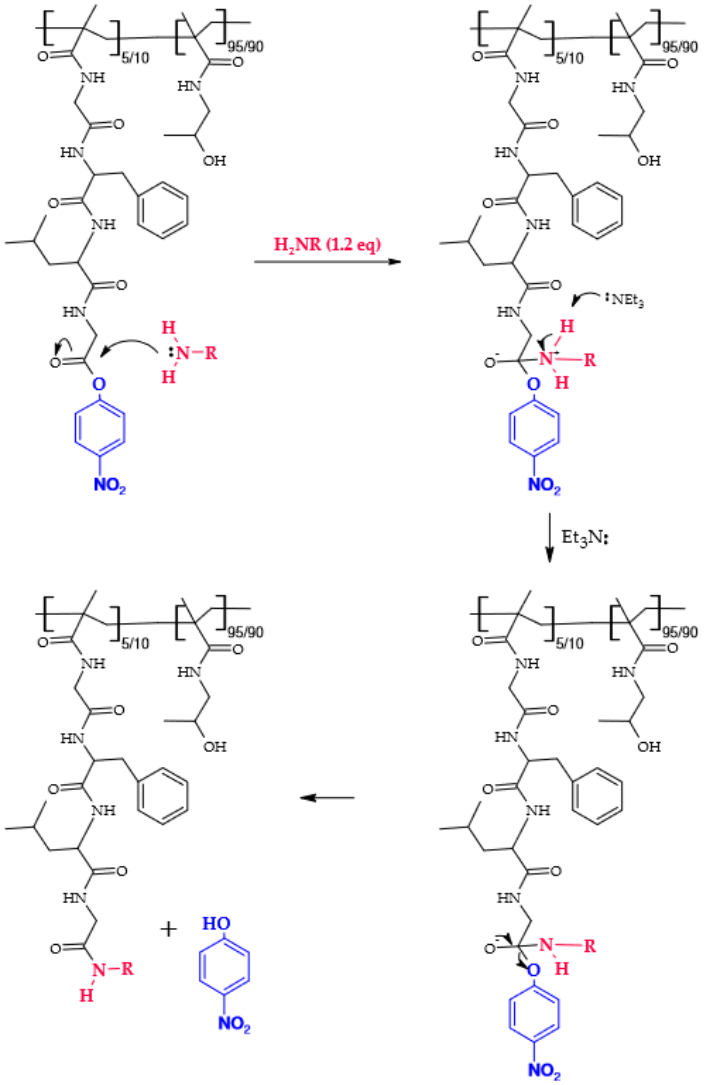


Figure 2.2: Reaction mechanism for the aminolysis reaction used to modify the parent copolymer HPMA.

2.2.1.2. Ultra-Violet spectroscopy (UV-Vis)

HPMA aminolysis reactions were followed by UV/Vis spectroscopy, using a Jasco V-570 dual beam spectrometer measuring between 200 and 500 nm. Samples were diluted in 99.9% pure DMSO, the spectrum plotted and peak intensities measured using Jasco Spectra Manager. An example data set for the reaction between 10 mol % HPMA starting polymer and dodecylamine is

shown in Figure 2.3. After the addition of dodecylamine, upon the occurrence of aminolysis, the bound nitrophenol peak visible for the parent polymer at $\lambda = 270$ nm is lost and replaced by a peak for unbound nitrophenol at $\lambda = 430$ nm. As the extinction coefficient of the bound nitrophenol on the parent polymer is known ($\varepsilon(270\text{ nm}) = 9500$ L/mol/cm), the concentration of nitrophenol being substituted could be measured and therefore the amount of amine content inversely known.

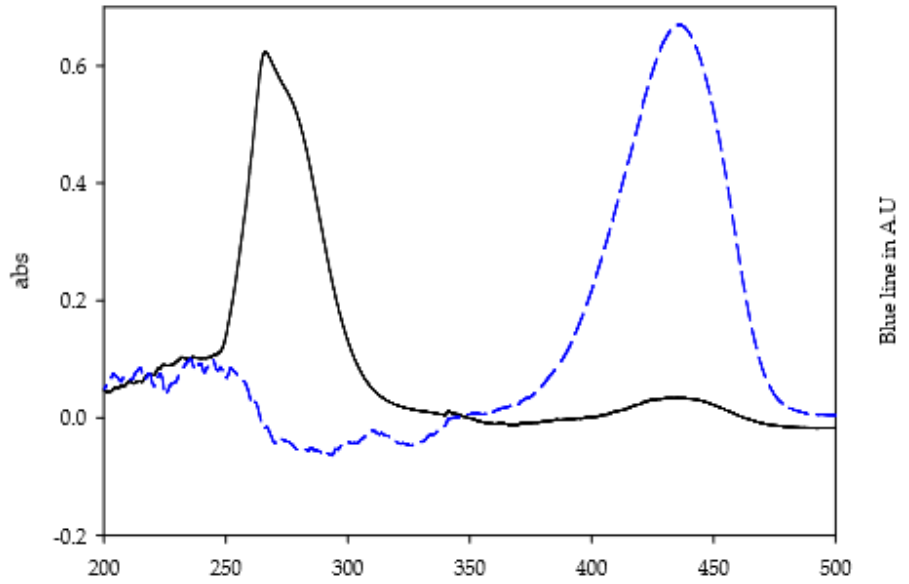


Figure 2.3: UV spectra showing the 10 mol % HPMA parent polymer before (solid line) and after (dashed line) the addition of dodecylamine. (C James, thesis, 2011, Cardiff University).

2.2.1.3. Nuclear Magnetic Resonance (NMR)

Nuclear magnetic resonance (NMR) was performed on a Bruker DRX-500 MHz spectrometer using a solvent suppression pulse program to reduce the signal from any water present ($\delta = 4.8$) [46]. Samples were prepared in D_2O at concentration of 5 mg/ml. For conjugates modified with a linear amine, the rise of a peak relating to the amine terminal $-CH_3$ group ($\delta = 0.79$) follows the substitution process. For conjugates modified with an aromatic group, peaks relating to the aromatic amines NH ($\delta = 7.4 - 7.8$). NMR spectra were predicted and plotted using g-NMR software package [47].

2.2.2.4. Small-angle neurton scattering (SANS)

2.2.2.4.1. SANS instrumentation

In a SANS experiment, the number of neutrons scattered as a function of the scattering vector, Q, is measured, where Q depends on the scattering angle and neutron wavelength (Appendix A 1). For this project, facilities of the D11 and LOQ diffractometers, located at the ILL and ISIS respectively, were used. D11 uses a fixed neutron wavelength and multiple detector distances to obtain a possible Q range between 0.00031 and 1 $Å^{-1}$. At LOQ a white beam of neutrons (of wavelength range from 2.2 to 10 Å) is used in combination with time-of-flight methods to produce a Q range between 0.006 and 0.24 $Å^{-1}$.

2.2.2.4.2. SANS data analysis

The intensity of neutrons scattered is determined by the number of particles (n_p), their volume (V_p), the particles' scattering contrast with solvent ($\Delta\rho$), their shape (P(Q)), interparticle interactions (S(Q)) and a contribution from any background incoherent scattering (B_{inc})

$$I(Q) = \varphi_p V_p (\Delta\rho)^2 P(Q, R) + B_{inc} \tag{3.36}$$

The first three terms Equation 3.36 are Q-independent, and refer to the absolute intensity of scattering. The FISH software used in this project groups these terms into one parameter, which "scales" the fit intensity to match the actual scattering data. Fitted parameters can be validated by comparing the reported scale factor, Sfit,, to one calculated, S_{calc} obtained from Equation 3.37.

$$S_{calc} = V\rho\ \varphi(\Delta\rho^2) \tag{3.37}$$

The particle volume, V_p, is calculated from the best fit parameters (e.g. rod length and radius of a cylinder). The volume fraction, ϕ, and scattering contrast, $\Delta\rho$, are obtained by using the known sample experimental parameters.

2.2.2.4.3. FISH modelling

FISH refers to the data fitting program developed by R. Heenan [51], used to obtain fitted parameters for known models from results of SANS results. FISH fitting procedure is based on the use of an iterative, linear least-squares process to fit known models containing the equations describing the scattering expected for different form factors, structure factors, contrast steps, polydispersities and background scattering to the experimental data. Parameters can be fixed manually and the program used to optimise the fit of the remaining parameters. All equations have been reproduced from the FISH manual.

2.2.2.4.4. Used models

HPMA conjugates have been studied previously and their scattering described as Gaussian coils, cylinders, and Kholodenko worm-like chains. These models were used for fitting procedures. Further evidence for a particular conjugate shape can be found by plotting the data in the Kratky representation of $I(Q)*Q^2$ versus Q.

2.2.2.4.5. SANS sample preparation

HPMA-conjugates were dissolved at concentrations between 0.5 and 2 wt% in D_2O (pH 5.5, 0.1M + 0.02 phosphate buffer, PBS) and placed in quartz cells (Hellma) with a pathlength of 1 or 2 mm. Samples were thermostatted to 27-37°C. Typical measuring times were between 45 and 90 minutes per sample. Sample solution scattering was normalised for the scattering and transmission of the solvent and cell, and then placed on an absolute intensity scale by reference to a standard (D_2O). Ap 100% was used as a standard sample for benchmark, as described in previous studies [50].

2.3. References

[1] D. R. Hartree The calculation of Atomic Structures. New York: Wiley & Sons, 1957.

[2] M. Born, R. Oppenheimer, Ann. Physik, 1927, 84, p. 457.

[3] J. C. Slater Phys. Rev. 1930, 36: 57.

[4] S. F. Boys Proc. Royal Soc A 1950, 200 (1063): 542–554.

[5] E. Davidson, D. Feller, 1986, Chem. Rev. 86 (4): 681–696.

[6] R. Ditchfield, W.J. Hehre, J. A. Pople, J. Chem. Phys. 1971, 54 (2): 724–728.

[7] C. Møller, M. S. Plesset, Phys. Rev. 1934, 46 (7): 618–622.

[8] P. Hohenberg, W. Kohn, Phys. Rev., vol. 136, p. B864, 1964.

[9] W. Kohn, L. J. Sham, Phys. Rev., vol. 140, p. A1133, 1965.

[10] W. Kohn, L. J. Sham, Phys. Rev., vol. 140, p. A1133, 1965.

[11] C. Lee, W. Yang, R. G. Parr Phys. Rev. B 1988, 37 (2): 785–789.

[12] C. D. Sherrill, H. F. Schaefer III in Löwdin, Per-Olov. Advances in Quantum Chemistry 34. San Diego: Academic Press. 1999, pp. 143–269.

[13] H. G. Kümmel in Bishop, R. F.; Brandes, T.; Gernoth, K. A.; Walet, N. R.; Xian, Y. Recent progress in many-body theories Proceedings of the 11th international conference. Singapore: World Scientific Publishing, 2002. pp. 334–348.

[14] W.D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, J. Am. Chem. Soc. 1995 117: 5179–5197.

[15] B.R. Brooks, R.E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus J Comp Chem 1983 4 (2): 187–217.

[16] W. F. van Gunsteren, H. J. C. Berendsen, Groningen Molecular Simulation (GROMOS) Library Manual, BIOMOS b.v., Groningen, 1987.

[17] W. L. Jorgensen, J. Tirado-Rives J. Am. Chem. Soc. 1988, 110 (6): 1657–1666.

[18] A.K. Rappe, C.J. Casewit, K.S. Colwell, W.A. Goddard III, W.M. Skiff, J.Am. Chem. Soc. 1992, 114, 10024–10035

[19] T. A. Halgren, J. Comp. Chem.; 1996; 490-519.

[20] S. J. Marrink, H. J. Risselada, S. Yefimov; D. P. Tieleman, A. H. de Vries, J Phys Chem B 2007, 111 (27): 7812–7824.

[21] B. J. Alder, T. E. Wainwright J. Chem. Phys. 1959 31 (2): 459.

[22] A. Rahman, Phys. Rev. 1964, 136: A405–A411.

[23] F. H. Stillinger and A. Rahman, J. Chem. Phys. 1974, 61, 4973-4980.

[24] J.A. McCammon, in "Report of the 1976 Workshop, Models for Protein Dynamics," H.J.C. Berendsen, Ed., Centre Europeen de Calcul Atomique et Moleculaire, Universite de Paris IX, France, 1977 pp. 137-152.

[25] T. Darden, L. Perera, L. Li, L. Pedersen Structure 1999, 7, R55–R60

[26] R. W. Hockney, J. W. Eastwood Computer Simulation Using Particles; IOP: London, 1988.

[27] A. Jain, N Vaidehi, G. Rodriguez J Comput. Phys. 1993 106 (2): 258–268.

[28] J.A. Barker, R.O. Watts Mol. Phys 1973 26 (3): 789–792.

[29] Molecular Operating Environment (MOE), 2013.08; Chemical Computing Group, Montreal, Canada, 2013.

[30] A. D. Becke Phys. Rev. A 1988, 38 (6): 3098–3100.

[31] C. Lee, W. Yang, R. G. Parr, Phys. Rev. B 1988, 37 (2): 785–789.

[32] Gaussian 03, Revision C.02, M. J. Frisch, et al, Gaussian, Inc., Wallingford CT, 2004.

[33] TURBOMOLE V5.1 2010, TURBOMOLE GmbH, since 2007.

[34] F. Haase and R. Ahlrichs, J. Comput. Chem., 1993, 14, 907.

[35] A. Schafer, C. Huber and R. Ahlrichs, J. Chem. Phys., 1994, 100, 5829.

[36] A. D. Becke, Phys. Rev. A, 1988, 38, 3098.

[37] J. P. Perdew, Phys. Rev. B: Condens. Matter Mater. Phys., 1986, 33, 8822.

[38] M. Cossi, N. Rega, G. Scalmani, V. Barone J. Comput. Chem. 2003, 24(6): 669-681.

[39] I. T. Todorov, W. Smith, K. Trachenko and M. T. Dove, J. Mater. Chem., 2006, 16, 1911.

[40] D. J. Willock, personal communication http:// theory.chem.cf.ac.uk/~dave/soft_intro.html.

[41] P. Cieplak and P. A. Kollman, J. Comput. Chem., 2000, 21, 1049.

[42] W. Smith and T. R. Forester, Comput. Phys. Commun., 1994, 79, 63.

[43] K. Sharp and B. Honig, Annu. Rev. Biophys. Biophys. Chem., 1990, 19, 301.

[44] J. A. Barker and R. O. Watts, Mol. Phys., 1973, 26, 789.

[45] M. J. Vicent, F. Greco, R. I. Nicholson, A. Paul, P. C. Griffiths and R. Duncan, Angew. Chem. Int. Ed. 2005, 44, 4061-4066.

[46] Y. Kasuya, Z. R. Lu, P. Kopecková, T. Minko, S. E. Tabibi and J. Kopecek. J. Controlled Release, 2001, 74, 203-211.

[47] A. L. Parrill, gNMR version 3, J. Chem. Inf. Comput. Sci., 1996, 36 (1), pp 153–153

[48] N. L. Allinger, J. Am. Chem. Soc., 1977, 99, pp 8127–8134.

[49] S. Lifson and A. Warshel, J. Chem. Phys. 1968, 49, 5116.

[50] A. Paul, M. J. Vicent and R. Duncan, Biomacromolecules 2007, 8, 1573-1579.

[51] R. K. Heenan, FISH manual1, Rutherford Appleton Laboratory, Didcot, U.K.

Chapter 3: HPMA conjugates synthesis and characterisation

3.1. Introduction

It is known for polymer-drug conjugates that solution behaviour derives from the polymer backbone properties as well as drug solubility, size and loading [1]. Exploring the influence of drug and carrier combinations by systematically changing the amount and properties of the drug part can yield useful information for controlling polymer-drug conformation, and thereby biological activity and clinical performance. Thus, various experimental and theoretical researches were held during recent years [2-5]. Among them, experimental investigation on mono-substituted HPMA polymer conjugates [1, 4] demonstrate the ability to change polymer-drug conjugate sizes and morphology changing the solubility and ratio of drug-mimic part. Results of this investigation showed both the degree of loading and the structure of the substituent have an extensive influence on the conformation of the conjugate in solution (Figure 3.1).

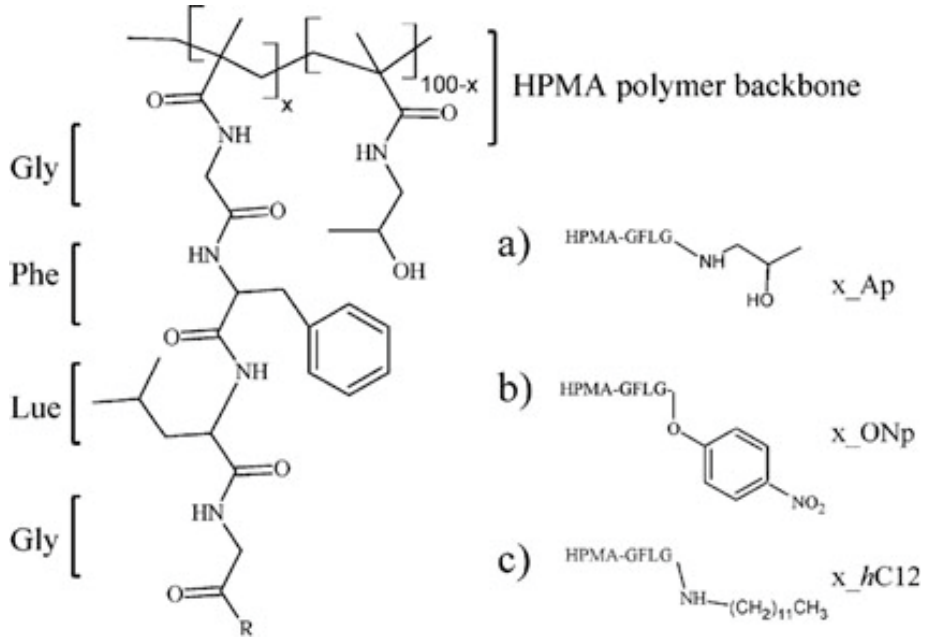


Figure 3.1: Structure of the HPMA-*co*-MA-GFLG copolymer where x defines the mole percent of GFLG-R side chain loading and R is the bound drug mimic: a) aminopropanol, x_Ap; b) *para*-nitrophenol, x_ONp, c) dodecylamine, x_*h*C12. [4]

For selected loadings of parent polymer conjugate, modified by the addition of linear substituents, it has been shown that as the size of the hydrocarbon increases the conjugate conformation will change from a fully flexible Gaussian coil structure (aminopropanol (Ap)) to a more rigid chain structure (hexylamine-decylamine (C6 – C10)) and finally to a rigid rod structure for larger alkanes (C12 – C18). Changes made in the hydrophobicity of the substituent demonstrated conclusively that hydrophobicity rather than chain structure is responsible for the conformation changes observed. Where the conjugates have been modified with aromatic substituents, the results follow the same trend: as the substituent increases in size, the flexibility reduces from a worm-like chain for aniline (An) to a rod for aminoanthracene (Anc) [1].

Based on these results we were able to expand this research area, exploring combinations of drug-mimics. Besides addressing the problem of general ability of producing desired structures, this also allowed us to investigate the influence of substituent's structure and size on properties and solution behaviour of polymer-drug conjugates. In order to provide the required data for further analysis, we have chosen systems with highly-soluble and low- or non-soluble drug parts; aromatic and non-aromatic substituents; and rigid/flexible alkanes mixtures in selected ratios of 1:9, 1:3 and 1:1. Therefore selected systems were first synthesised and purified and then characterised using SANS experiments and data analysis. Effective influence of ratio changes on the solution behaviour were tracked by changes in SANS curves and fitted parameters, compared with mono-substituted polymer-conjugates and analysed for further conclusions.

3.2. Synthesis and purification

3.2.1 Conjugates with mono substituents

HPMA-co-MA-GFLG-ONp (HPMA) parent polymer was described above in Chapter 2.2.1 was selected as precursor for all of the synthesized polymers in order to obtain comparable experimental results. The nitrophenol group of the HPMA parent polymers were substituted with selected amines using an aminolysis reaction (shown in Figure 3.1). Both HPMA parent copolymer precursor with 5 mol% or 10 mol% loading of ONp were dissolved [1.2 mol equiv. agents in respect to nitrophenol groups] [1] in a minimal volume (5-10 ml) of anhydrous dimethyl sulphoxide (DMSO) in round-bottom flask and the system flushed with nitrogen.

The flask was heated to 45°C to ensure complete dissolution. 1.1 mol eq of chosen amine (RNH2) and 1 eq of triethylamine were added to the reaction mixture. Then solution was stirred for three to five hours under nitrogen atmosphere at 45°C, monitoring evolution of free nitrophenol concentration as well as substituent concentration (if possible) with UV/VIS. After reaction was completed (level of nitrophenol remains constant), the reaction was quenched by addition of 3 eq of 2-aminopropanol for 1 h. UV/VIS and 1H NMR in D_2O were used to confirm complete substitution of the ONp groups with amine of choice. Results of characterization are provided in Supplementary Information. The conjugate was precipitated by dropwise addition of reaction mixture to a 450 ml of constantly stirred mixture of acetone:diethyl ether (8:1). The crude product obtained by filtration was dissolved in a minimal volume of distilled water and purified by dialysis against H2O (MW cut-off 2000 g*mol-1) for 2-3 days (6-8 water changes per day). Purified polymer was then freeze dried and stored in the freezer.

3.2.2. Conjugates with mixed substituents

Synthesis of mixed substituents is similar to that for mono-substitution. The difference is simply the need to perform synthesis in two stages in order to control the mixture ratio for substituted polymer. The first stage is reaction with substituent that was selected as a minor mixture component. Then, after reaction with first substituent is complete (progress can be evaluated by UV-Vis spectra), second substituent of choice is added.

For the first stage, the selected first amine of choice (R_1NH_2) was added in reaction mixture with respect to required ratio (1 mol eq for 100%). Then 1 eq of triethylamine was added, and the solution was set to stir until reaction is considered to be complete. Progress was monitored by UV-VIS spectroscopy, sampling reaction every 30 min. After the first reaction is completed, 3 eq of the second substituent were added to the mixture and stirred for 1 h. UV/VIS and 1H NMR in D_2O were used to confirm complete substitution of the ONp groups with amine(s) of choice. Results of this conformation are provided in Supplementary information. The conjugate was precipitated by dropwise addition of reaction mixture to a 450 ml of constantly stirred mixture of acetone:diethyl ether (8:1). The crude product obtained by filtration was dissolved in a minimal volume of distilled water and purified by dialysis against H2O (MW cut-off 2000 g*mol-1) for 2-3 days (6-8 water changes per day). Purified polymer was then freeze dried and stored in the freezer.

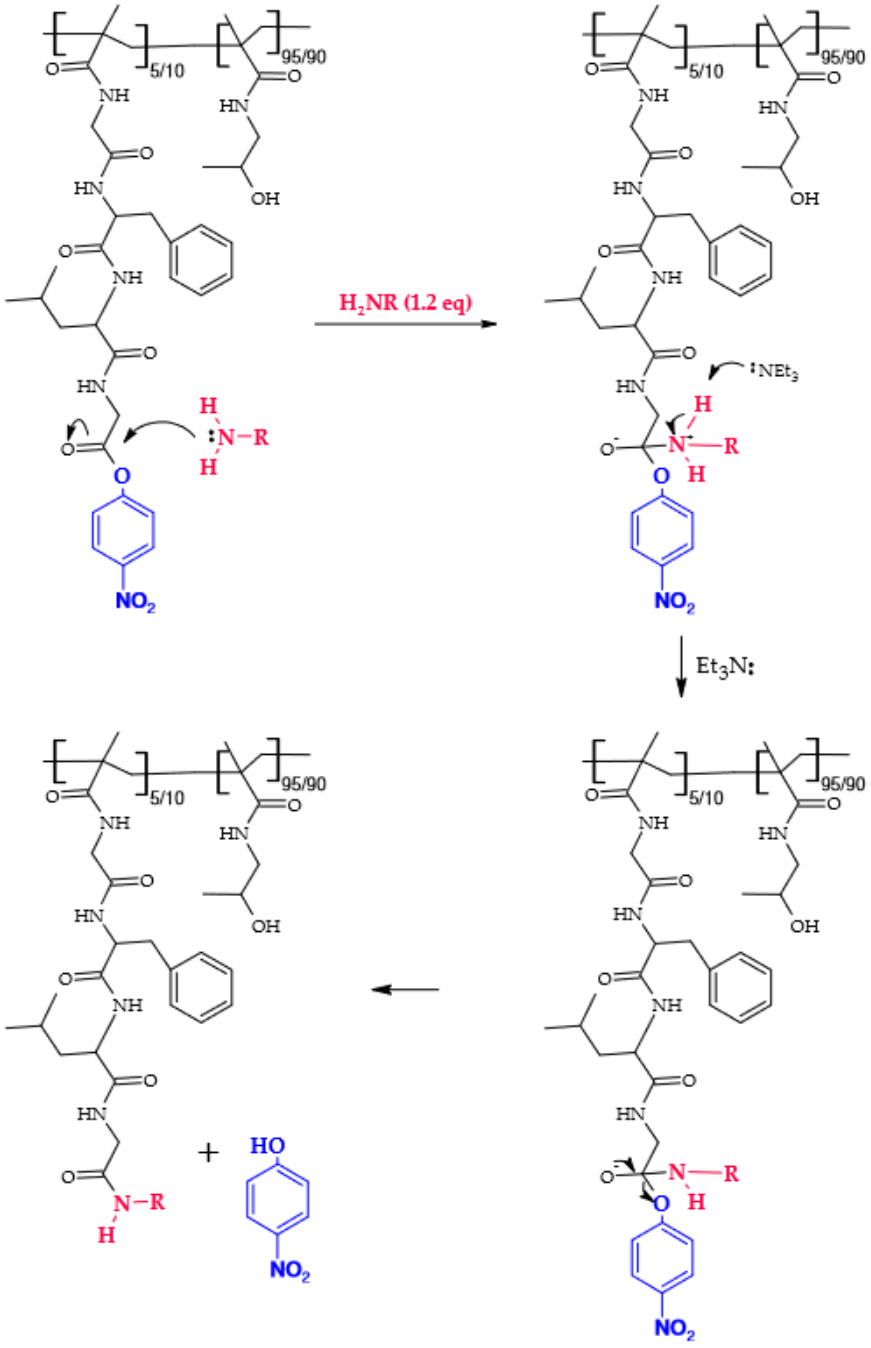


Figure 3.2: The aminolysis reaction mechanism of parent HPMA copolymer modification. [3]

3.3. Characterisation

3.3.1. UV-Vis and NMR data analysis

Aminolysis reactions were monitored using UV/Vis spectroscopy, using a Jasco V-570 dual beam spectrometer measuring between 200 and 600 nm. Samples were diluted in DMSO, the spectrum plotted and peak intensities measured using Jasco Spectra Manager. An example data set for the reaction between 10 mol% HPMA starting polymer and ANC is shown in Figure 3.3. Start of sampling (Figure 3.3 blue line) was made before aminoanthracene (ANC) was added to reaction mixture. After ANC was added Sample #1 was taken (Figure 3.3 orange line). Rising peak at λ = 451 nm shows presence of free ANC which decreases after ANC substitution of *o*-nitrophenol (oNp) group. After the Sample #4, since no major difference was noticed between spectroscopy results, reaction can be considered as fully completed.

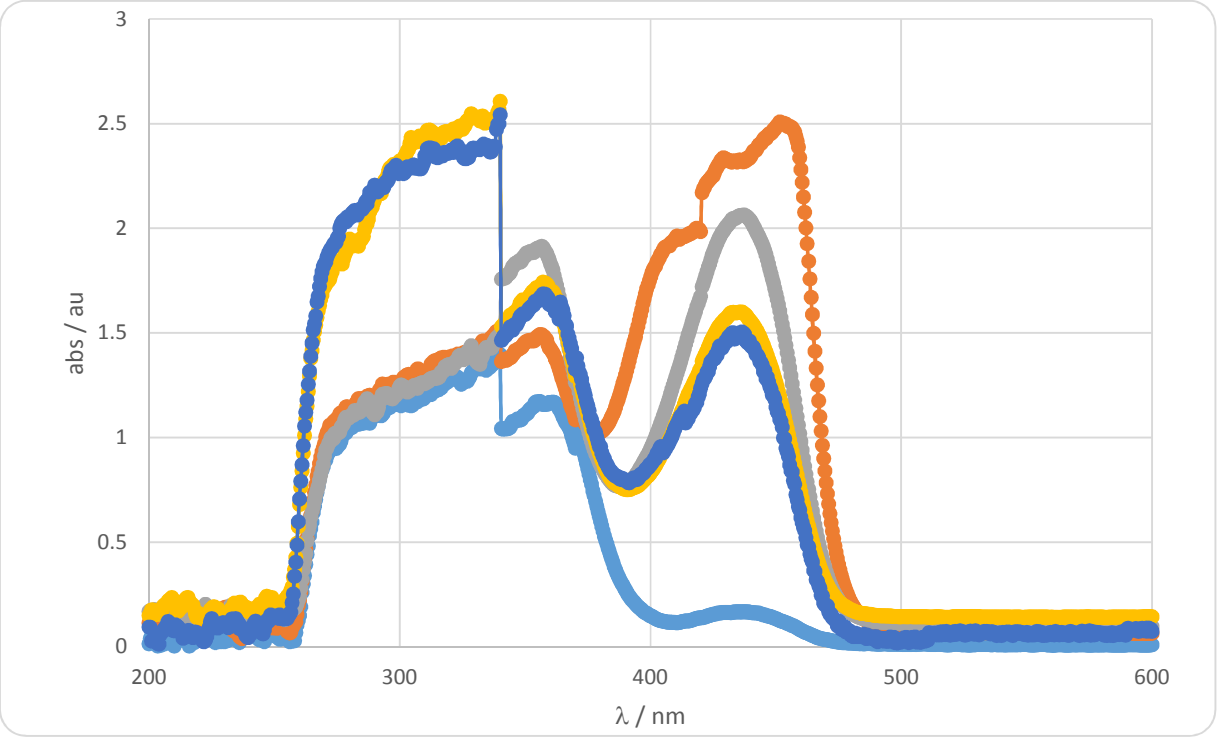
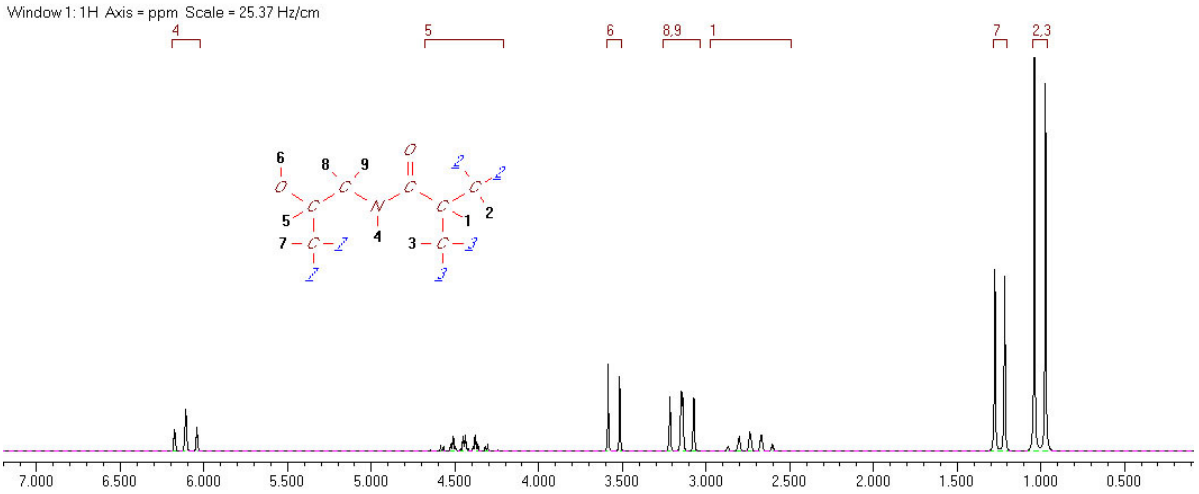


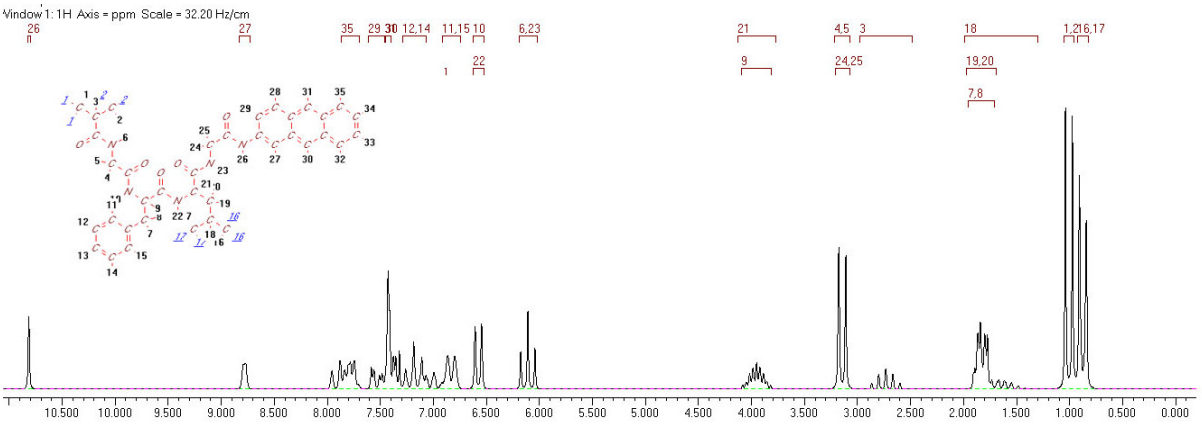
Figure 3.3 UV spectra of aminolysis reaction progress of HPMA parent polymer with ANC

As well as monitoring reaction progress with UV/Vis spectroscopy, NMR was used to confirm polymer's structure. Example spectra are shown below in Figure 3.4 (predicted) and 3.5

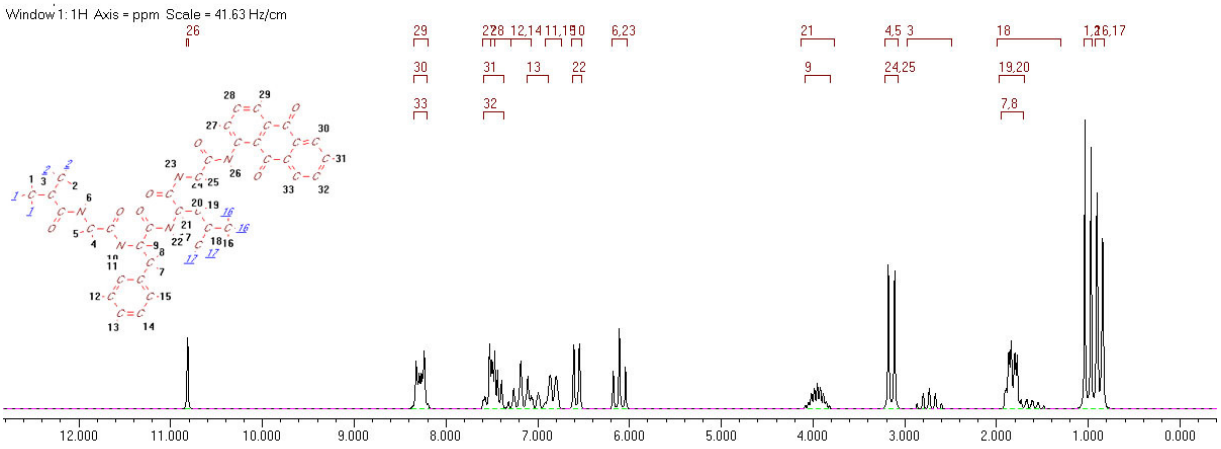
(experimental). NMR spectra were analysed and compared with predicted peak shifts and ratios in order to complete the characterization process for synthesized polymers.



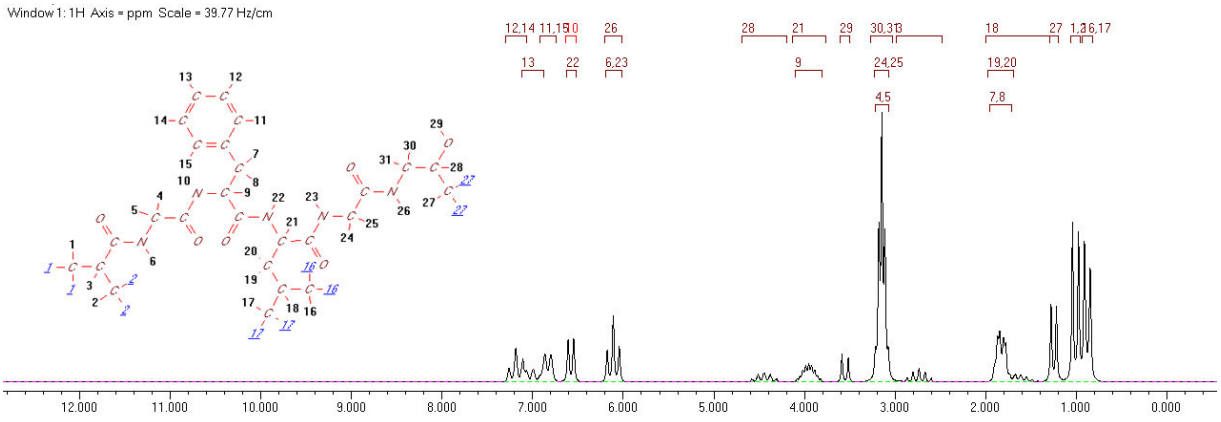
(a)



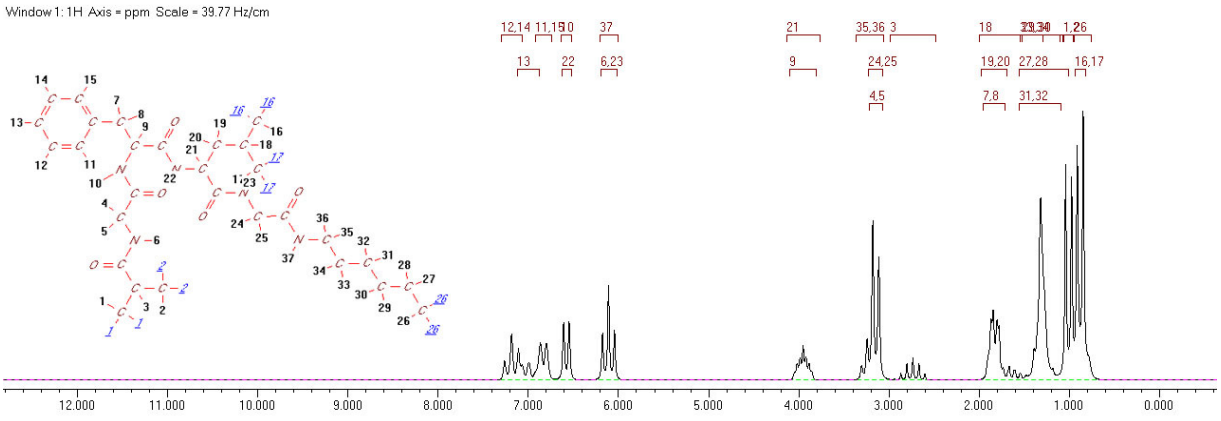
(b)


Window 1: 1H Axis = ppm Scale = 41.63 Hz/cm
12.000
11.000
10.000
9.000
8.000
7.000
6.000
5.000
4.000
3.000
2.000
1.000
0.000

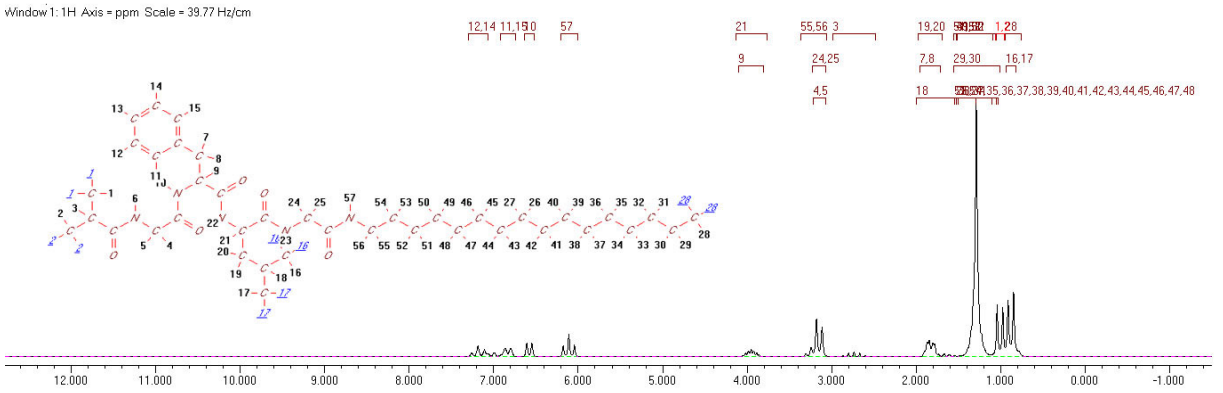
(c)


Window 1: 1H Axis = ppm Scale = 39.77 Hz/cm
12.000
11.000
10.000
9.000
8.000
7.000
6.000
5.000
4.000
3.000
2.000
1.000
0.000

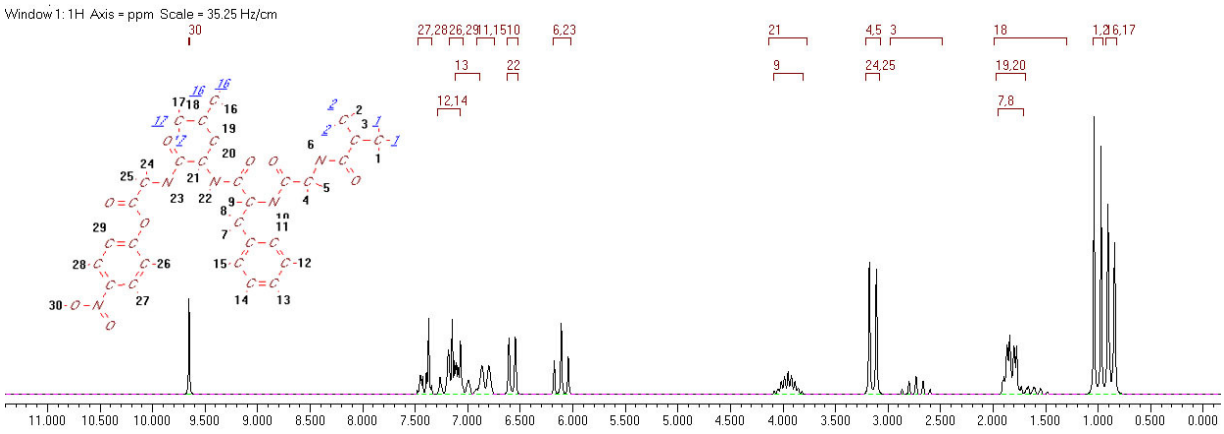
(d)


Window 1: 1H Axis = ppm Scale = 39.77 Hz/cm
12.000
11.000
10.000
9.000
8.000
7.000
6.000
5.000
4.000
3.000
2.000
1.000
0.000

(e)



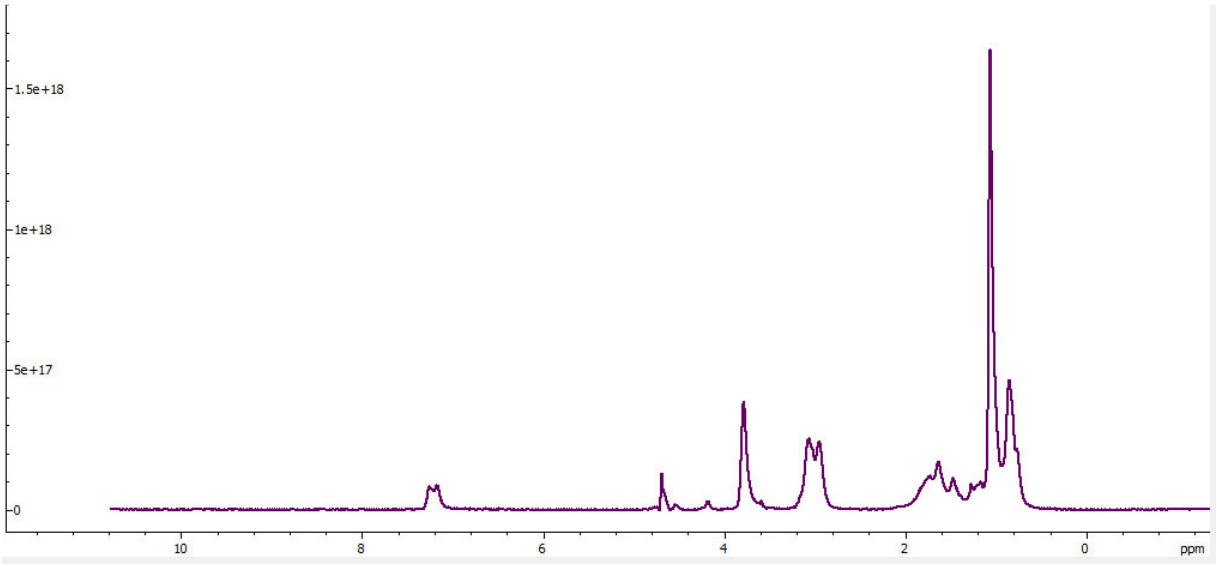
(f)



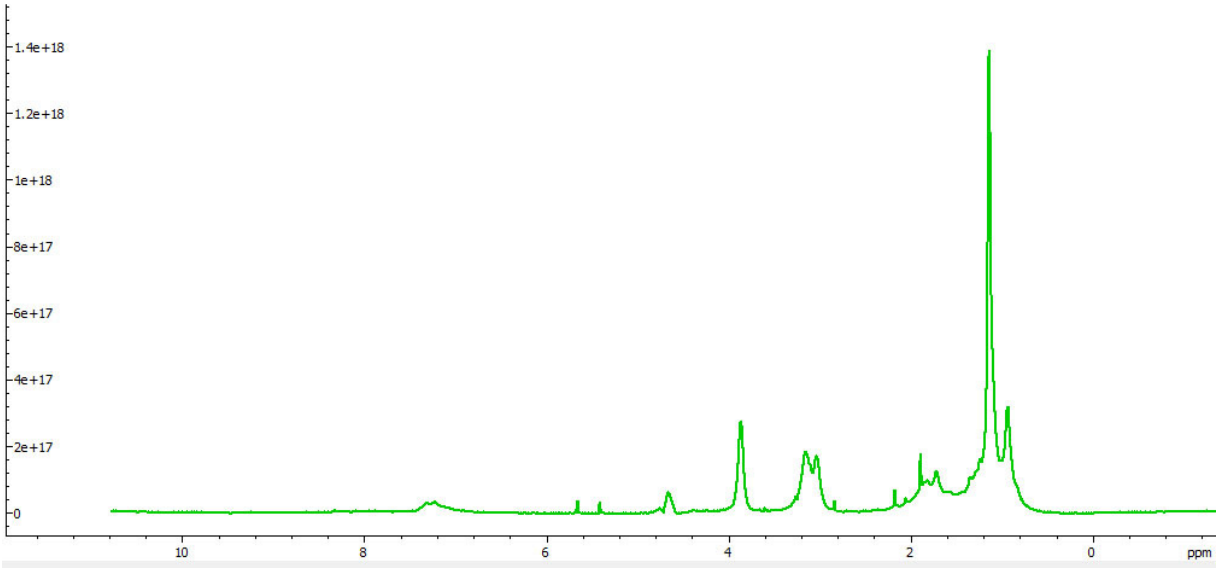
(g)

Figure 3.4 Predicted NMR spectra for a) HPMA, b) HPMA-ANC, c) HPMA-ANQ, d) HPMA-Ap, e) HPMA-C6, f) HPMA-C14 g) HPMA-ONp

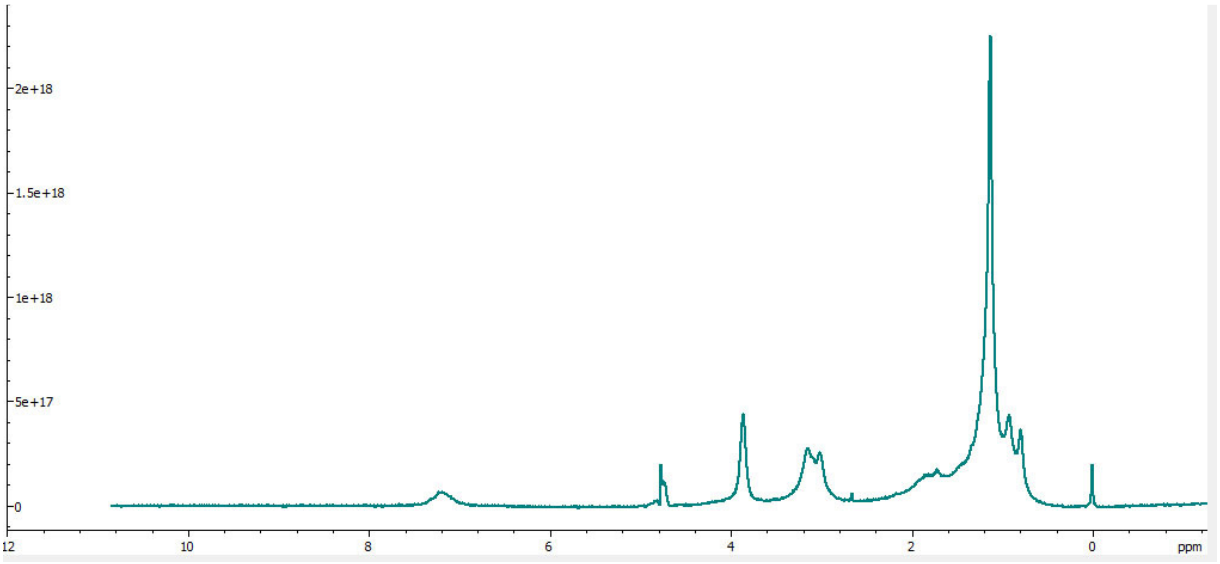
For mono-substituted polymers it was possible to provide conclusive results to show substitution of ONp group to amine of choice by comparing peaks ratio at selected intensities. In the case of substitution with aromatic drug-mimic parts (ANC, ANQ) new peaks at 10.6-10.8 ppm appeared (H atom on carbonyl) as well as ratio of peaks at 7.5-8.3 ppm increased due to contribution of hydrogens bonded with aromatic carbons.


1.5e+18
1e+18
5e+17
0
10
8
6
4
2
0
ppm

(a)


1.4e+18
1.2e+18
1e+18
8e+17
6e+17
4e+17
2e+17
0
10
8
6
4
2
0
ppm

(b)



(c)

Figure 3.5 Experimental NMR spectra for a) HPMA-ANC, b) HPMA-ONp, c) HPMA-C14 polymer conjugates

The same characterization methods were used for mixed conjugates as for mono-substituted polymer molecules. In order to characterize selected polymer conjugates we compared the ratio of protons in side-chain of polymer conjugate, for example signals from hydrogens bonded to aromatic carbons (~7.2 ppm) to signals from hydrogens, bonded to carbon with two carboxamide group (4.187 ppm). For 100% HPMA-ANC conjugates, the selected ratio is 7, for ANQ = 6, and for polymer-Alkane conjugates = 5/2. Applying these methods to experimental results from NMR data analysis we were able to charecterize mixed conjugates for future study, using advanced techniques such as SANS data analysis. However, for C6/C14 mixture it is not possible to obtain conclusive results due to a minor contribution of side-chain groups in the resulting spectra and inability to track changes in alkane intensities area. The robustness of the experimental methodology demonstrated with a wide range of compounds [1] suggests that the products are of the desired composition.

3.2.2. SANS data analysis

Validating obtained polymers as desired products, we were able to proceed to structural characterisation. In order to obtain more information on the sizes and properties of the selected HPMA-conjugates, SANS technique was applied. This also provided us with the ability to carry out the direct comparison between experimental data and computer modelling results (discussed in Chapter 5). As well as allowing to fit direct dimensions for solubilised polymers, SANS was successfully used for characterization of polymer-drug conjugates [2-5], due to its ability to be carried out under normal physiological conditions (body temperature, biological pH, ionic strength of 0.1M (NaCl), conjugate concentration of 50 mg/ml used in clinical research) without destroying or interacting with target molecules. Characteristics SANS plots for each family of conjugates are given below.

3.2.2.1. Mixed linear substituents

C6 and C14 were chosen as model substituents due to noticeable difference in sizes of polymer molecules. Series of 1:9, 1:3 and 1:1 mixtures were synthesized, characterised and studied using SANS techniques. For 100% C14 and 100% C6 substituted conjugates, a difference of a factor of two in length of solubilised molecules was used for a comparison. As shown in Table 3.1, gradual increase of C14 forces the system to adopt a more compact shape with decreasing length of the polymer molecule but retaining similar radius for fitted cylinder. This shows the influence of changing ratio of drug-parts and can be explained by difference in solubility parameters for alkane substituents of different alkane chain length.

3.2.2.2. Mixed aromatic substituents

In order to investigate the contribution of aromatic substituents on overall shape and sizes of polymer conjugates, the same series of mixed systems were developed for ANQ/ANC series of polymers. For ANC and ANQ 100% substituted copolymers, shape and size difference is also well described and can be explained by difference in hydrophobicity for these aromatic molecules. Thus, 100% ANQ conjugate can be described as a cylinder with length of 200±10Å and radius of 40±5Å and 100% ANC described as a cylinder with length of 118±10Å and radius of 37±5Å. Nevertheless, the scattering curves obtained from SANS experiment are similar for these two

systems and demonstrate rod-like shape on Kratky representation. However, mixed systems with these two aromatic substituents follow a similar trend, expanding with increasing ANQ presence in the mixture.

In fact, polymer molecules with mixed aromatic substituents demonstrate another interesting feature – adding of another aromatic substituent changes affects the size of polymer molecule, expanding its length. For example, for ANQ/ANC 25/75% system the best fitted parameters are 425±10Å length and 53±5Å radius, which is twice as long as obtained for pure ANQ and four times longer than 100% ANC fitted cylinder. These data can be explained by self-repulsion of these two parts within the polymer chain.

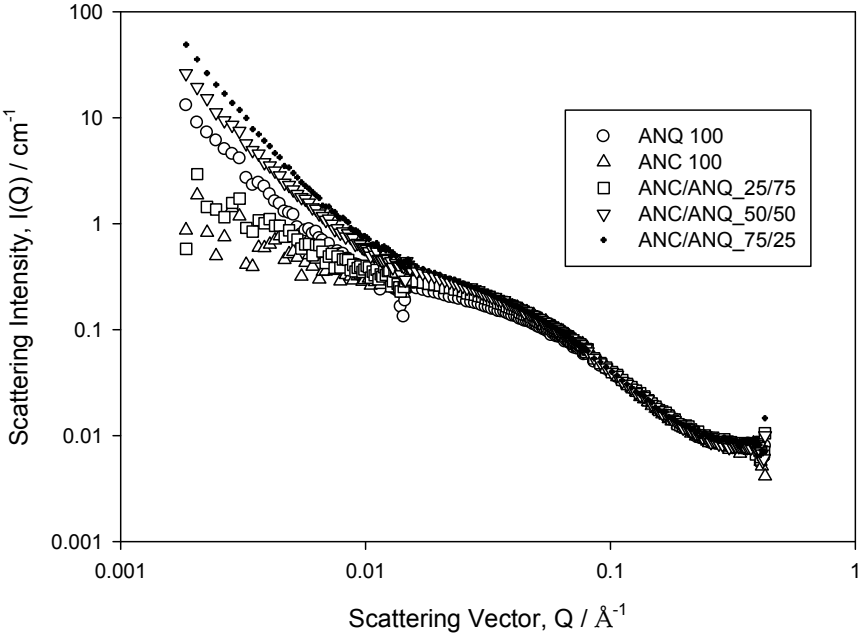
3.2.2.3. Mixed linear/aromatic substituents

For ANQ/C14 conjugates, results of SANS data analysis are different due to contribution of aromatic groups. However, transition from pure ANQ to pure C14 conjugates is noticeable and can be tracked both on Kratky and raw scattering data plots. For pure ANQ conjugates we can observe rod-like structure while for C14 substituted polymers results show spherical morphology. Increasing amounts of C14 substituent in the mixture gradually progresses from rod-like to spherical structures, which can be extremely useful both for research purpose as well as clinical application of produced drug-polymers. Being able to track changes and evaluate contribution of increasing hydrophobic/hydrophilic groups, we decided to progress to further investigation, comparing similar substituents in terms of chemical make-up but different in structure.

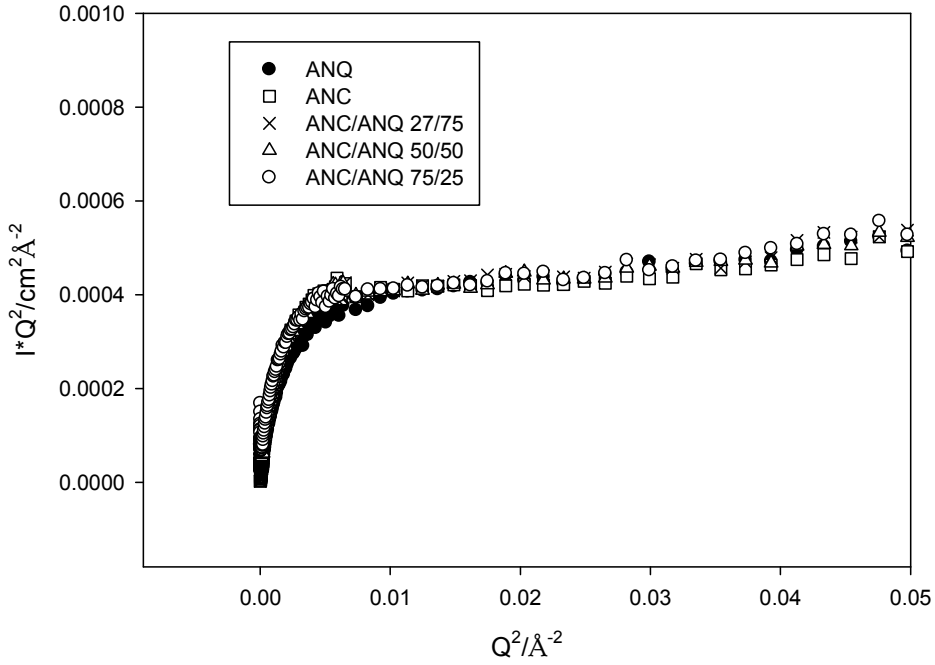
3.2.2.4. Mixed adamantine substituents

In order to obtain required information we have selected C10 and Ad substituens which have same properties and molecular formula but are different in structure and flexibility. For C10 substituent, the flexible alkane chain differs greatly from the rigid bonds in Ad. This defines difference both in enclosure volume and can have an effect on overall structure behavior. This difference was investigated using results of comparison between mixed systems of Ad/ANC, Ad/ANQ and Ad/C10 conjugates which are presented in Figure 3.6.

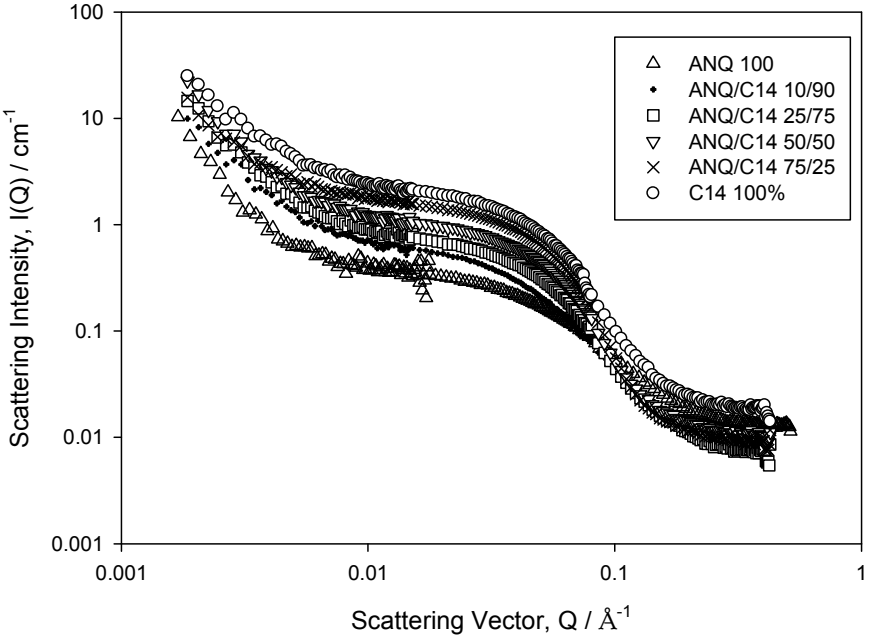
ANC/ANQ



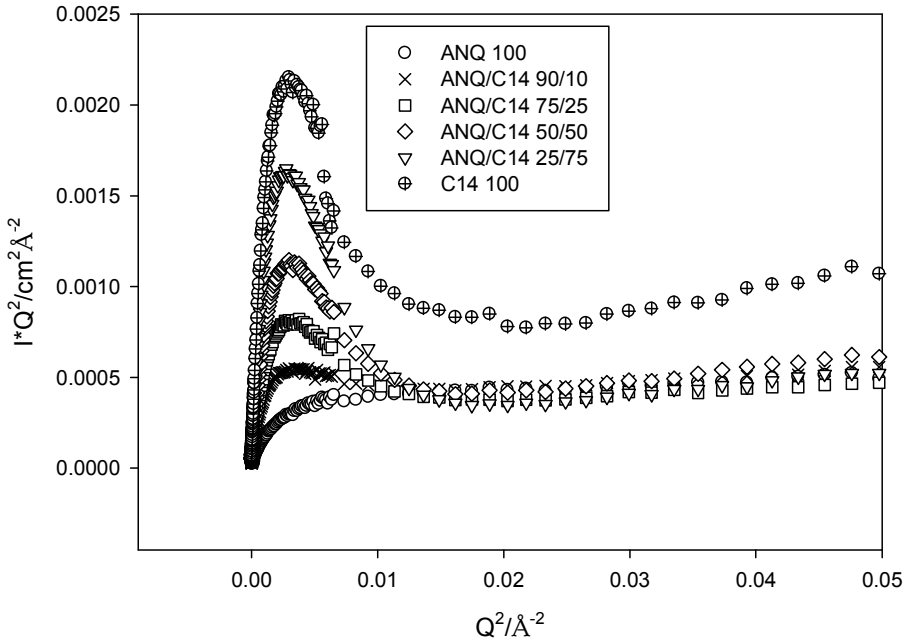
ANC/ANQ



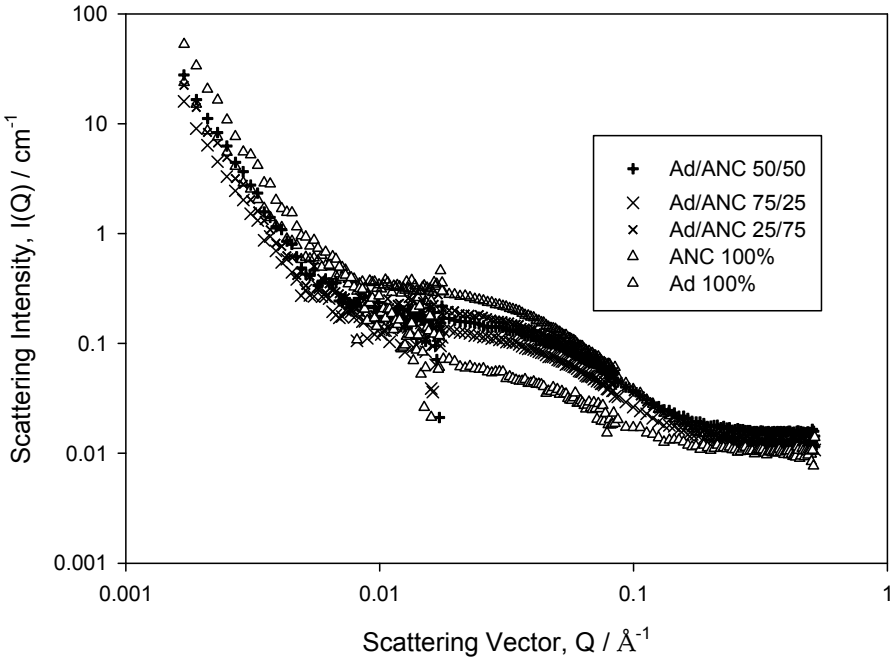
ANQ/C14



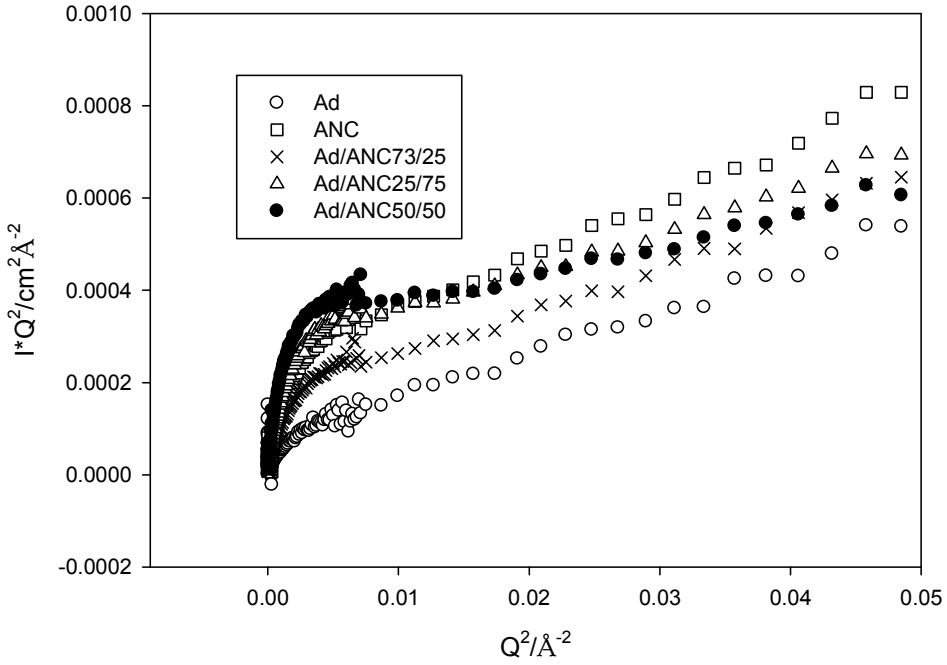
ANQ/C14



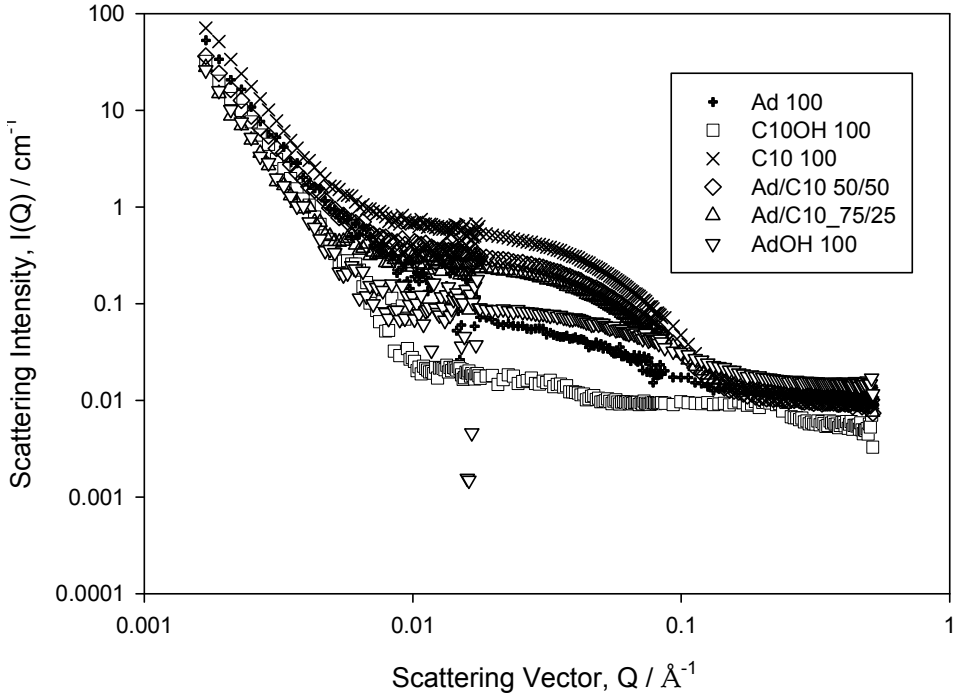
Ad/ANC


Scattering Intensity, I(Q) / cm-1
Scattering Vector, Q / Å-1
Ad/ANC 50/50
Ad/ANC 75/25
Ad/ANC 25/75
ANC 100%
Ad 100%

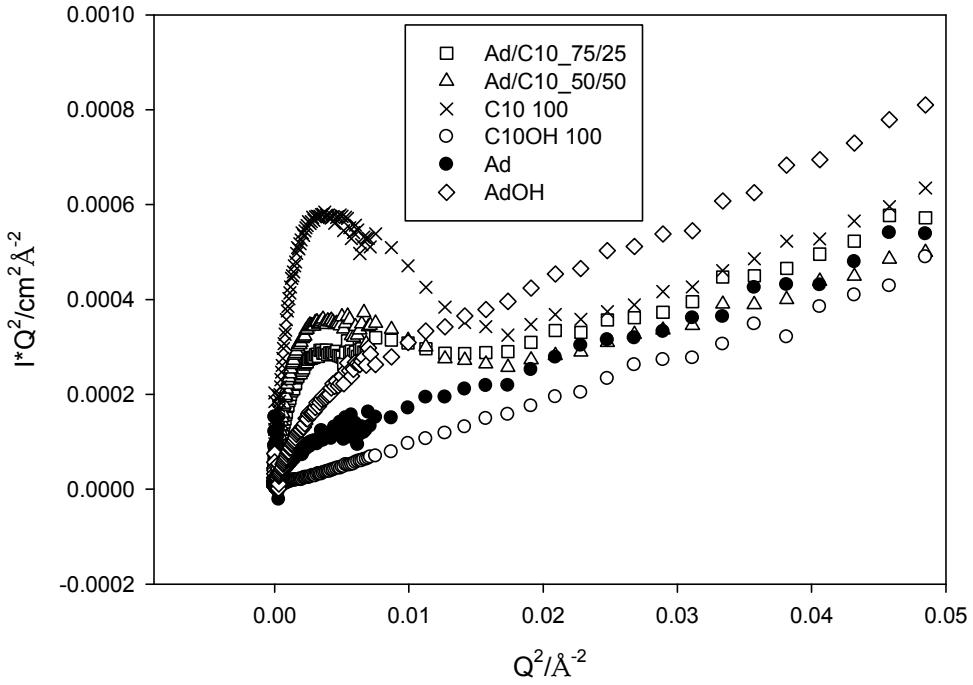
Ad/ANC


I*Q2/cm2Å-2
Q2/Å-2
Ad
ANC
Ad/ANC73/25
Ad/ANC25/75
Ad/ANC50/50

Ad/C10


Ad 100
C10OH 100
C10 100
Ad/C10 50/50
Ad/C10_75/25
AdOH 100
Scattering Intensity, I(Q) / cm-1
Scattering Vector, Q / Å-1
100
10
1
0.1
0.01
0.001
0.0001
0.001
0.01
0.1
1

Ad/C10


Ad/C10_75/25
Ad/C10_50/50
C10 100
C10OH 100
Ad
AdOH
I*Q2/cm2Å-2
Q2/Å-2
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
-0.0002
0.00
0.01
0.02
0.03
0.04
0.05

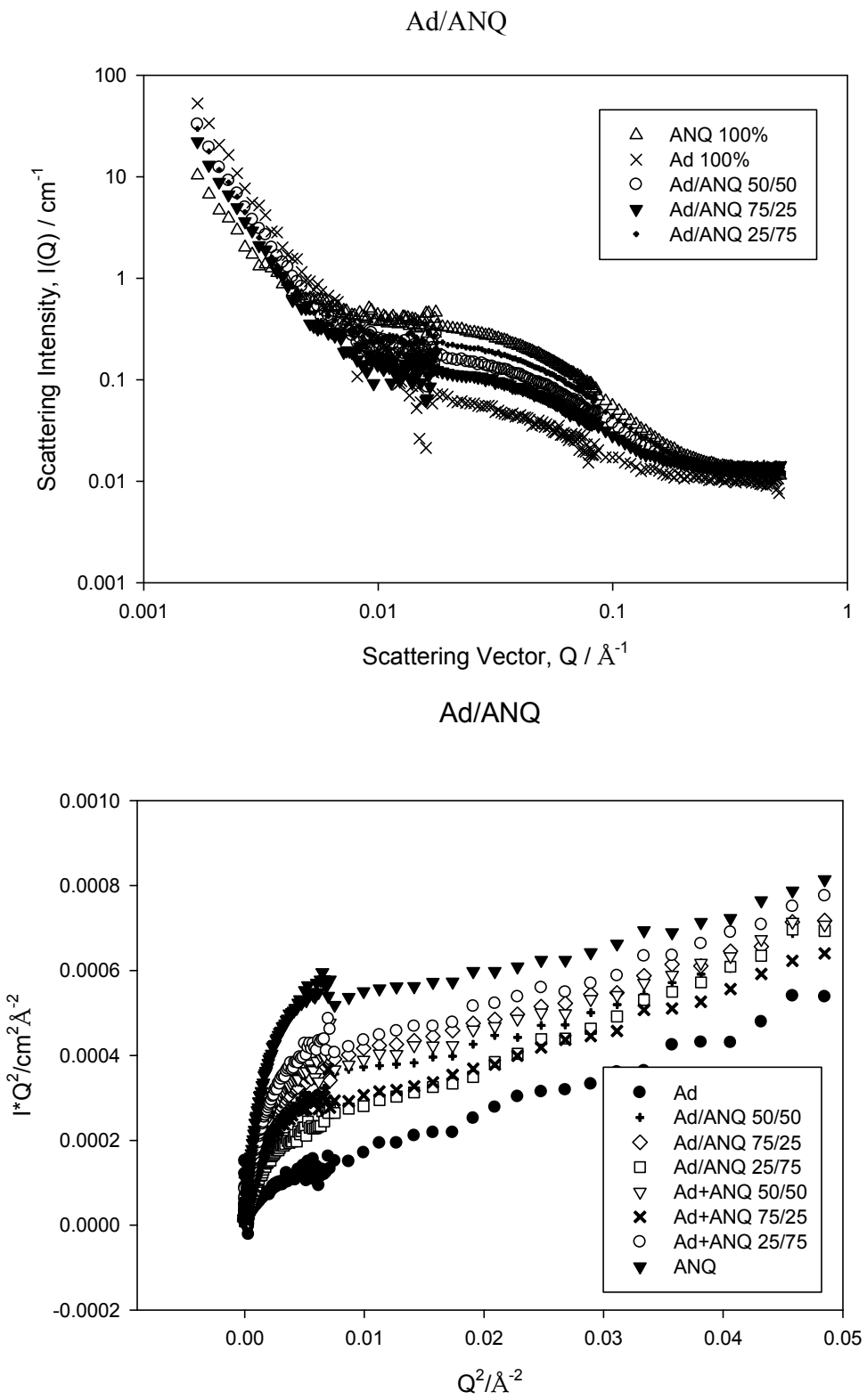


Figure 3.6 Scattering curves for ANC/ANQ ANQ/C14, Ad/ANC, Ad/C10 and Ad/ANQ polymer conjugates

3.3.2.2. SANS data analysis

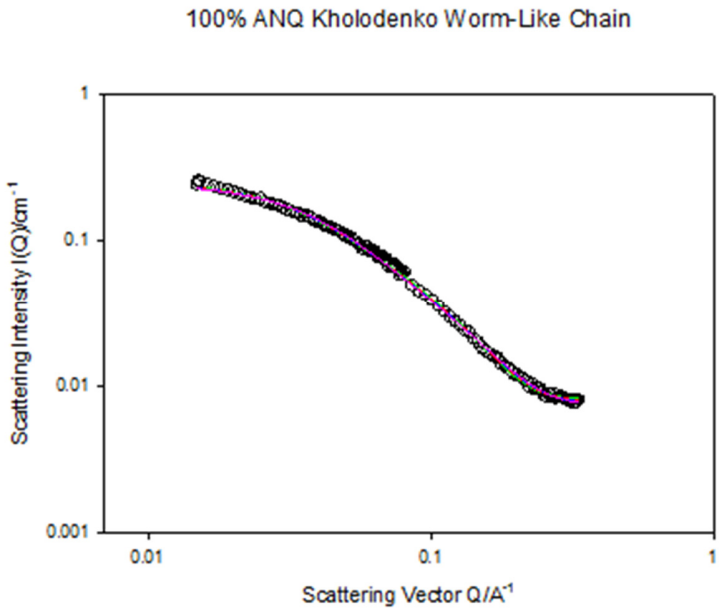
The data fitting program FISH was used to fit the whole of the scattering data to form factors for various shapes. Due to similarity of selected systems of mixed conjugates with mono-substituted HPMA-copolymers that was described in previous studies, the same data fitting technique was used for SANS data analysis [1]. Kholodenko worm-like chain model as an expansion of the Gaussian coil was used for fitting procedures as it was shown previously to be effective representation for HPMA-copolymers. Further evidence for a particular conjugate shape can be found by plotting the data in the Kratky representation of $I(Q)*Q^2$ versus Q. In this plot, scattering data from a coil, rod and worm-like chain have distinctive forms.

Experimental data was used for further analysis and fitted to model shapes. Results of best fits for cylinder model are provided in Table 3.1. For selected series of conjugates gradual changes in sizes can be observed. For example, C6/C14 mixture retains similar radius as well as increased length with increasing amount of C6, which correlates with results for mono-substituted C6 and C14 conjugates. However, in some cases changes in parameters do not follow the same trend and no correlation found for mixture ratio and sizes of obtained structures. Despite that, SANS data analysis is able to provide information on changes of structure and morphology of polymer-drug conjugates with mixed substituents. Based on scattering curves for series of HPMA-copolymers it was shown that properties of mixed conjugates are defined both by types of drug-mimics and ratio of selected substituents. This allows us to use available data for known conjugates to predict possible parameters for mixed substituents, as well as to change the ratio in mixture to obtain desired structure with chosen sizes and preferred shape which links to biological performance.

Conjugates and ratio of substituents	Rg, Å ±5Å	Length, Å ±10Å	Radius, Å ±5Å
C6 100%	80	171	27
C14 100%	88	89	30
C6/C14 50/50%	97	146	27
ANC 100%	37	118	17
ANQ 100%	40	200	20
ANQ/ANC 25/75%	53	425	16
ANQ/ANC 50/50%	47.5	203.5	16
ANQ/ANC 75/25%	56	145	15
C14/ANQ 10/90%	55	146	22
C14/ANQ 25/75%	61	175	27
C14/ANQ 50/50%	69.5	108	29
C14/ANQ 75/25%	68.5	93	29.5
ANQ/ANC 10/90%	49	165	17
Ap 100%	33	100	14
C14/C6 10/90%	72	204	23
C14/C6 75/25%	67	101	30
C14/C6 50/50%	97	146	27
C14/C6 25/75%	63	115	25.5
ANQ+ANC 75/25%	44	190	16
C14 +ANQ 10/90%	50	128	21
C14 +ANQ 50/50%	63	97	27
C14 +ANQ 75/25%	75	92	29
C14/C6 75/25%	81	100	29
ANQ+ANC 50/50%	68	170	16

Table 3.1. SANS fitting results* for selected HPMA-conjugates both for mono-substituted polymers and mixtures of drug-mimic parts. X/Y refers to poly-substituted polymer conjugate, X+Y refers to mixture of mono-substituted polymer-conjugates. [Note from AP: Table includes parameters from the 'best fit' to each model, but is not intended indicate which model gives the most appropriate fit to the data].

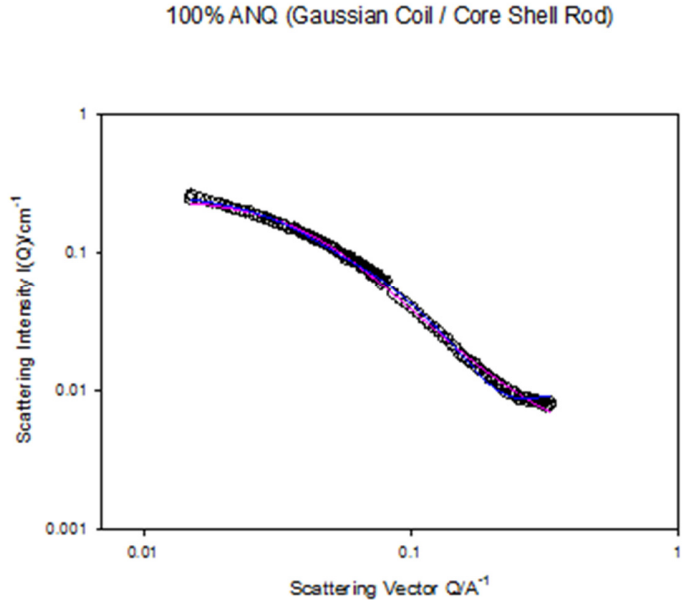
A



Fitted parameters

R_{ax} / Å (± 5 Å)	n	*l* / Å (± 10 Å)	n*l*	swse
10	1	135	135	805
10	2	78	156	786
9	3	58	176	816
9	4	48	194	852

B



Gaussian coil fit (pink line)

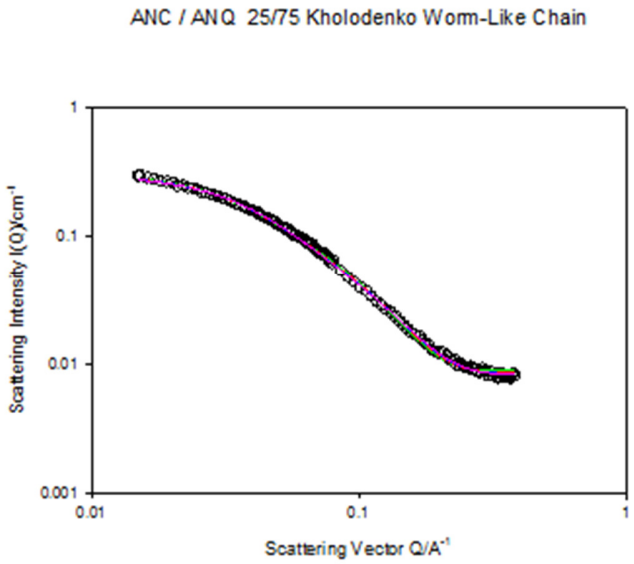
Core/shell rod fit (blue line)

Fitted parameters

Gaussian Coil: Rg = 36 Å, SWSE = 1063

Core Shell Rod: L= 135 Å, R= 14.68Å, SWSE=1237

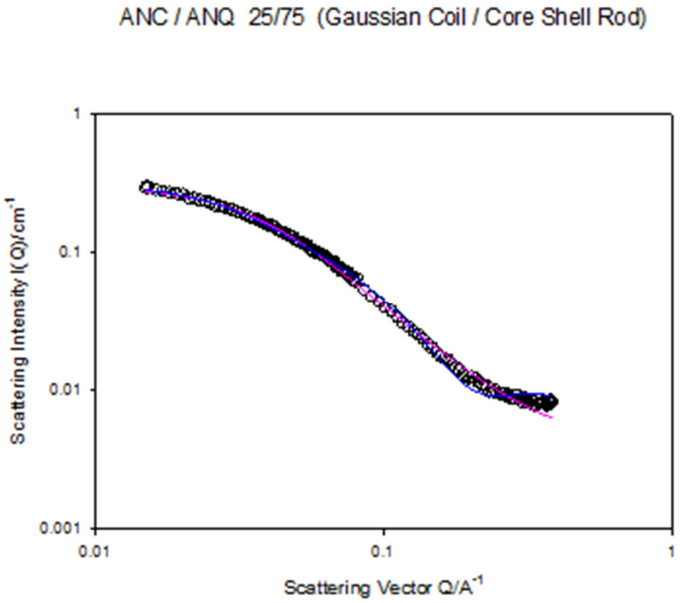
C



Fitted parameters

R_{ax} / Å (± 5 Å)	n	*l* / Å (± 10 Å)	n*l*	swse
11	1	143	143	639
11	2	81	163	493
10	3	61	184	442
10	4	51	204	429

D



Gaussian coil fit (pink line)

Core/shell rod fit (blue line)

Fitted Parameters

Gaussian Coil: Rg = 39 Å,

SWSE = 914

Core Shell Rod: L= 139 Å,

R= 16 Å, SWSE=1585

Figure 3.7. Representation of best fits in FISH software for 100% ANQ and 27/75 ANC/ANQ mixed conjugates.

3.4. Conclusions

It is important to synthesize desired polymer-drug conjugates in controllable reactions, not only to develop novel therapeutic agents but also to provide reliable data against which modelling methods can be tested. Having a wide database of SANS experimental data and fitting for mono-substituted HPMA conjugates, we have started from reproducing previous results, and then proceeded to more complicated synthesis and SANS experiments on conjugates with mixed substituents.

Our aim was to control solution behaviour of polymer-drug conjugates using information obtained from SANS data. Combining mixtures of various drug mimics, we have developed characterisation methodology as well as investigated influence of changing ratio of hydrophilic/hydrophobic drug mimics for mixed polymer conjugates. Thus, for systems containing aromatic/alkane substituent mixtures we were able to gradually change the overall structure and morphology of HPMA-polymer conjugates from rod-like for aromatic substituents to a sphere-like for predominantly alkane mixtures.

Further investigation of mixed polymer conjugates contained drug-mimics with similar parameters of hydrophobicity, but different in terms of flexibility. In order to provide corresponding comparison we have selected systems of Ad/ANC, Ad/ANQ and Ad/C10 conjugate mixtures and investigated an effect of changing ratio of drug-mimic parts for these systems.

This allowed us to demonstrate the possibility to obtain desired polymer-drug molecules with variable ratio of selected drugs *in situ* controllable reactions, allowing us to proceed to the next research aims, *i.e.* computer modeling of desired structures which have various advantages among experimental variation approach but requires preliminary validation using known results.

3.5. References

[1] C. James, PhD thesis, Cardiff University, (2011)

[2] P. A. Vasey, S. B. Kaye, R. Morrison, C. Twelves, P. Wilson, R. Duncan, A. H. Thomson, L. S. Murray, T. E. Hilditch, T. Murray, S. Burtles, D. Fraier, E. Frigerio and J. Cassidy. Clin. Cancer Res., 1999, 5, 83-94.

[3] R. Duncan, L. W. Seymour and K. B. O'Hare. J. Controlled Release, 1991, 19, 331- 346.

[4]. A. Paul, M. J. Vicent and R. Duncan. Biomacromolecules, 2007, 8, 1573-1579.

[5] M. J. Vicent, F. Greco, R. I. Nicholson, A. Paul, P. C. Griffiths and R. Duncan. Angew. Chem. Int. Ed., 2005, 44, 4061-4066.

[6] R. K. Heenan, FISH manual, Rutherford Appleton Laboratory, Didcot, U.K

Chapter 4: All-atom modelling of HPMA

4.1. Introduction

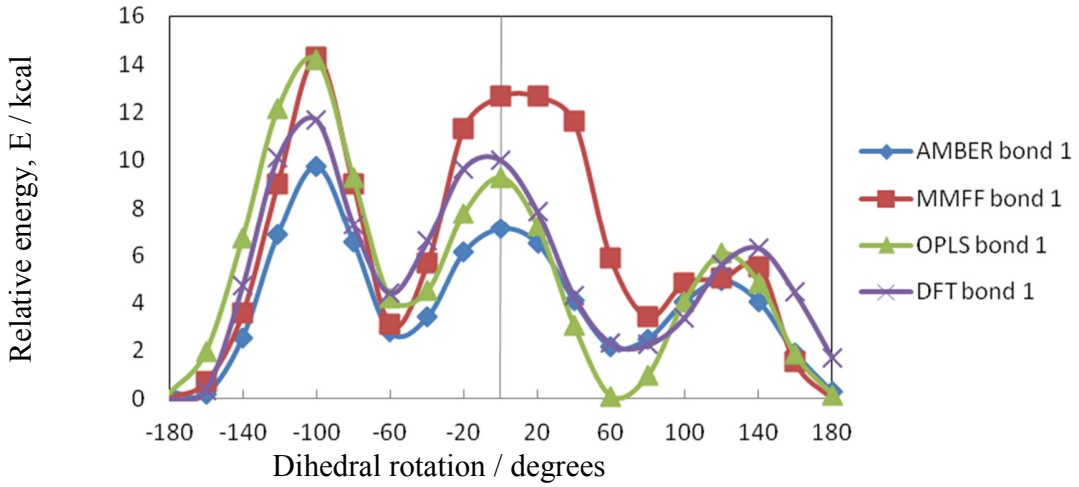
Experimentally, detailed studies of solution behaviour are time consuming and often complex, requiring synthesis and characterisation for numerous compounds to establish structure-property relationships using methods such as light, X-ray or neutron scattering, spectroscopic studies and viscosity measurements. Whilst a high level of insight into solution sub-structure can be obtained [2-4], these studies are undeniably expensive and time consuming. Development of modelling methodology techniques to accurately simulate solution conformation of polymer-drug systems would be a valuable tool for preliminary studies of unknown combinations of polymer carriers and drugs, as well as for more detailed investigation of experimentally characterised structures. The challenges of carrying out such calculations for polymer solutions, and in particular for the complex multi-component polymeric systems of interest in drug delivery are significant. AA simulation of systems of this size raises demands on code optimization and computational computational resources and requires substantial CPU time. As well as force field models, finding appropriate solvent models and development of specific software is an essential part of this research. Methods and tools developed can be used for polymer–drug studies, neutron scattering simulations and have wider application in various theoretical and experimental fields.

In this work, HPMA homopolymers with a range of molecular weight (MW) from 2 kDa to 35 kDa were built and studied via MD simulations. The results of MD simulation in explicit, R-field and distance dependent dielectric solvent models were interpreted using Flory mean field approach. As well as conventional parameters of interest, such as the radius of gyration R_g, the molecular volume, chain collapse time, radial distribution function (RDF) and direction dependent density profiles were investigated. These data were used to develop a model to describe the full range of shape variations with ellipsoidal parameters, which can be compared to results obtained from small-angle neutron scattering (SANS) experiments. After tools and methods validation, provided with this study and further implementation of new atom types presented in conjugates, we were able to procced to HPMA copolymer studies. Results of that investigation is presented in Chapter 5.

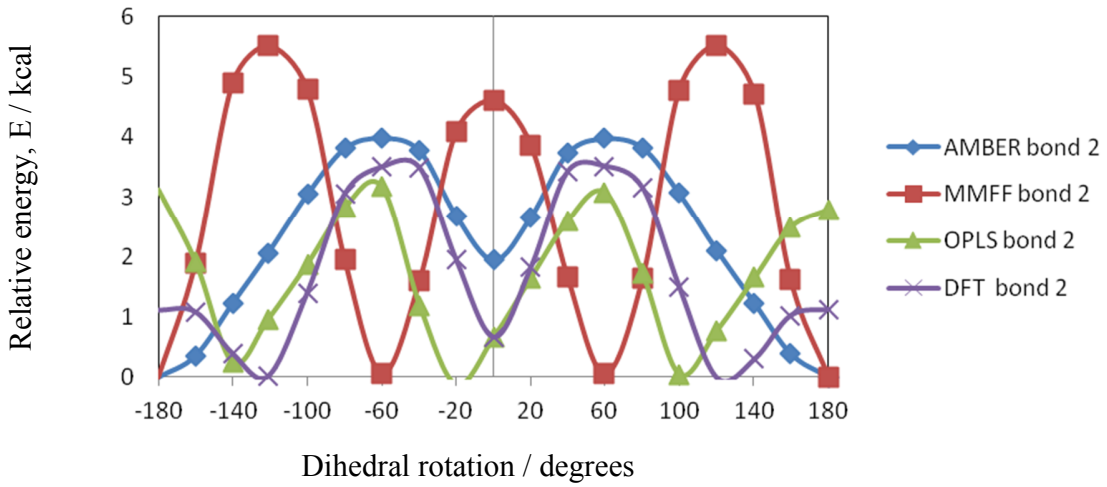
4.2. Force-field validation

Three force fields available within MOE have parameters for all atoms in HPMA, namely MMFF94x, AMBER99 and OPLS-AA. Comparisons of each of these against DFT and *ab initio* methods were used to determine which parameter set was most suitable for our purpose. Potential energies as a function of dihedral rotation for each of two flexible C-C bonds were calculated: Figure 4.1 shows the results of such calculations with various methods. This shows that OPLS-AA and AMBER99 have similar energy profiles to that generated by DFT, whereas MMFF94 shows different number and position of minima and markedly different energy barriers to the DFT results.

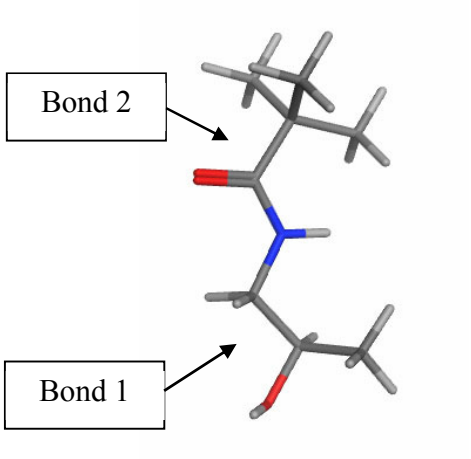
A systematic conformational search of HPMA dimer then was performed using OPLS-aa, AMBER99 and MMFF94 parameter sets with rejection energy of Van der Waals contacts set to 10 kcal/mol. Relative energies of thirteen obtained conformations were calculated using DFT methods were compared with forcefield results, as shown in Figure 4.2. *Ab initio* calculations results were used as a benchmark and show good correlation with BP86 PCM results: mean square deviation in conformational energies for BP86 PCM and MP2/def2-TZVP were found as 0.89 kcal and for BP86 PCM and MP2/aug-cc-pVTZ 1.01 kcal.



(a)



(b)



(c)

Figure 4.1 Dihedral energies plots obtained from AMBER99, OPLS-AA, MMFF94 and DFT (B3LYP) calculations for a) C-C bond 1, b) C-C bond; c) – representation of both bonds.

As shown in Figure 4.2, relative energies obtained by BP86, AMBER99 and MMFF94x are similar, whereas OPLS-aa results in significantly different energy ordering of conformations, mean square deviations from results of BP86 PCM calculations for the following force fields in kilocalories are Bond 1: 1.92, 2.62, 1.62 and Bond 2: 1.11, 3.12, 1.22 for AMBER99, MMFF94 and OPLS-aa respectively. Considering the results of these tests, AMBER99 was chosen as the most appropriate force field parameter set for further investigation of HPMA copolymers.

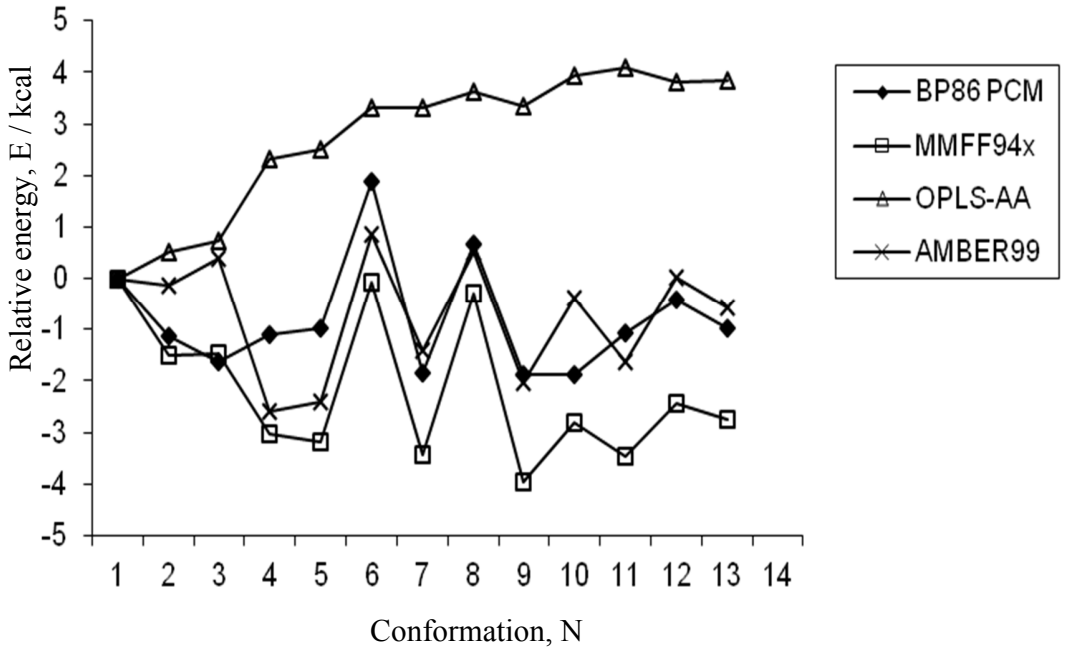


Figure 4.2. Results of total energies calculations for DFT and *ab-initio* methods

4.3. Solvent model selection

HPMA molecules with sizes from 4 to 265 monomers were built with random distribution of R-HPMA and S-HPMA fragments in racemic mixture. In order to select a solvent model suitable for polymeric HPMA, R_g obtained for explicit water and two implicit solvent models were compared using a Flory mean field approach (Figure 4.3.). It is well known that the size of a polymer chain depends on the solvent. In a "good" solvent the chain expands to allow polymer-solvent contacts, whereas in a "bad" solvent the chain segments stay close to each other. From the Flory approximation $R_g = R_0 * N^{\nu}$, where R_0 is R_g of the monomer unit, N is the number of repeated units and ν is a solvent factor. For good solvent, $\nu \approx 3 / 5$, while for bad solvent, $\nu \approx 1 / 3$ [5, 6].

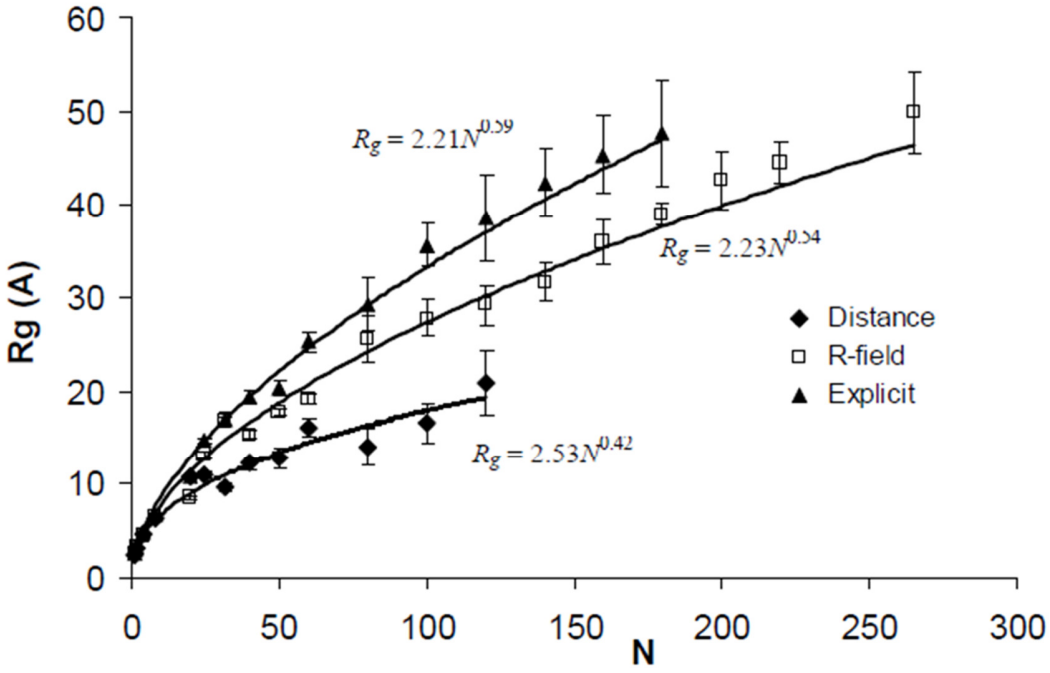


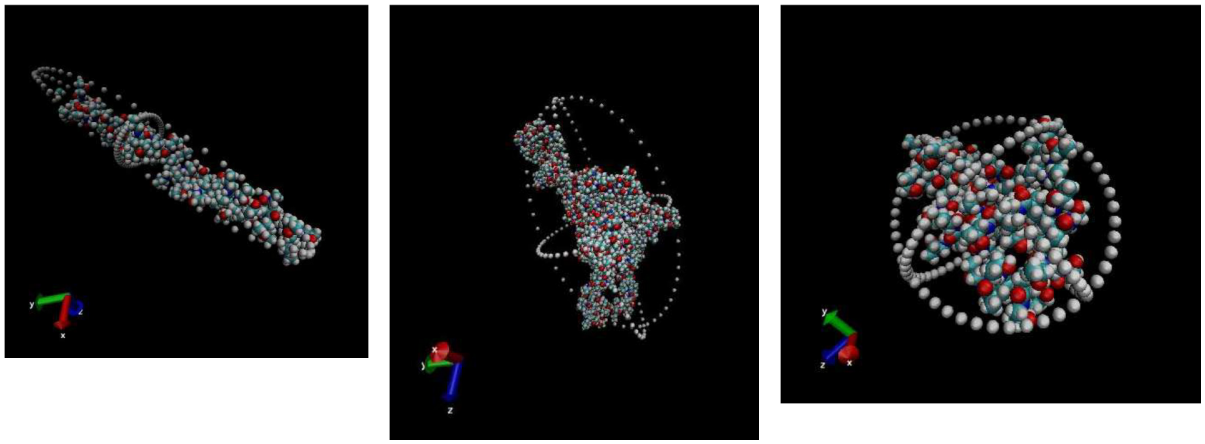
Figure 4.3. Comparison of different solvent models for a range of HPMA-copolymers. Lines are calculated fits to a power law, parameters of fits are shown above.

This theory was implemented for HPMA polymers (Figure 4.3) and the obtained data show striking differences between solvent models. It is evident that R_g resulting from explicit water and R-field solvation are in reasonable agreement, whereas the distance-dependent dielectric solvation model leads to much smaller values of R_g. Adding lines of best fit with the same form as Flory approximation, *i.e.* power law, results in fitted value of ν is close to the ideal value of 0.60 for "good" solvents, with 0.59 for explicit solvent and 0.54 for R-field, and identical R_0 values of 2.2 Å. Extrapolating the power law best fit line to N = 265 allows comparison with experimental data of SANS experiment, which yields R*g* = 75±3 Å[7]. Values of R_g obtained were 66±4 Å for explicit water and 49±4 Å for R-field, and hence both models can be used for further investigation of "real" polymers.

4.4. Data analysis

4.4.1. General concept

MD simulations were analysed using our own C code which allows for the analysis of trajectories for geometric data relevant to the system being studied. Data generated in this case includes R_g and radial distribution functions (RDF). We also examined the shape of polymers by calculating enclosing ellipsoids. This was done based on the moment of inertia matrix, calculated for each frame of the saved trajectory. This matrix was diagonalised to give eigenvalues and mutually perpendicular eigenvectors centred on the centre of mass. The eigenvectors correspond to the axes of rotation for the polymer as a whole, the axis corresponding to the smallest eigenvalue is the axis requiring the lowest torque to rotate the polymer as a rigid body. For a rod-like molecule, this naturally identifies the direction in which the polymer is elongated, *i.e.* the rod direction, in a way that is independent of the polymer orientation in space. The extent of the polymer in each of the three directions was then found using the polymer centre of mass as the origin and taking the component of each atom's position vector along the eigenvector directions.



a) b) c)

Figure 4.4. Representation of ellipsoid characterization approach for evaluation in time for MD trajectories of a) "rod-like", b) "disc-like", and c) spherical conformations for structures of different HPMA-conjugates sizes 100-120 monomer units.

The component furthest from the centre of mass in each direction was then used to define the semi-principal axes {a, b, c} of the enclosing ellipsoid. The ratio of these axes can then be used to define

molecule's shape. For example, "rod-like" or described as prolate spheroids with ratio of axes Rx ≅ Rz < Ry in a ratio of 1:2.5-4 (Figure 4.4.a). "Disk-like" molecules can be represented as oblate spheroids with Rx ≅ Rz > Ry (Figure 4.4.b) and for spheres axes will be nearly equal Rx ≅ Rz ≅ Ry (Figure 4.4.c). This approach is complementary to the use of the R_g to describe polymer shape [8]. Indeed, for polymers with uniformly distributed mass the eigenvectors of the moment of inertia matrix and those of the R_g tensor will have the same directions [9,10]. This has lead to the assumption that the two approaches are interchangeable [11].The main difference is the inclusion of the atomic masses in the moment of inertia matrix, which means that the heavier atoms have a greater influence on the matrix elements, whereas R_g tensor did not originally contain the atomic masses and simply defines the geonmetric shape of a polymer based on the co-ordinates of the monomer units [9]. Results for structures simulated using R-field solvent model are presented in Table 4.1 and suggest no systematic influence of tacticity of the initial linear polymer structure on results of MD simulations and corresponding parameters.

Monomer	Tacticity	R_g (Å)	STD	Average R_g (Å)	Rx (Å)	Ry (Å)	Rz (Å)	Ratio	Vol (Å^3)
HPMA-20	isotactic	8.7	0.15	8.5 ± 0.2	11.2	13.6	11.9	1.2	7557.3
	atactic	8.3	0.19		10.5	13.8	11.8	1.2	7111.9
	syndiotactic	8.4	0.22		9.1	14.0	12.2	1.3	6543.7
HPMA-40	isotactic	15.4	0.07	15.0 ± 0.4	12.9	16.0	12.9	1.2	11167.2
	atactic	14.6	0.06		11.7	21.2	13.8	1.7	14237.5
	syndiotactic	15.2	0.10		11.8	18.0	14.5	1.4	12950.6
HPMA-60	isotactic	19.6	0.10	19.1 ± 0.5	15.6	20.6	16.5	1.3	22232.0
	atactic	18.4	0.12		13.1	24.2	19.0	1.5	25150.2
	syndiotactic	19.4	0.05		14.9	20.4	15.4	1.3	19661.9
HPMA-80	isotactic	25.7	0.19	25.0 ± 2.5	15.0	31.8	15.1	2.1	30059.0
	atactic	27.0	0.14		13.9	26.1	22.7	1.4	34431.9
	syndiotactic	22.4	0.06		17.1	20.1	17.7	1.2	25509.4
HPMA-100	isotactic	26.3	0.07	27.8 ± 2	17.0	29.8	18.6	1.7	39479.5
	atactic	29.4	0.13		17.4	30.6	18.7	1.7	41683.1
	syndiotactic	27.7	0.11		16.8	30.8	24.8	1.5	53596.5
HPMA-120	isotactic	28.2	0.08	29.1 ± 2	19.4	25.1	19.6	1.3	39957.1
	atactic	27.7	0.13		17.5	35.9	19.8	1.9	51905.8
	syndiotactic	31.3	0.09		21.3	28.1	23.0	1.3	57615.1
HPMA-140	isotactic	30.3	0.06	31.6 ± 2	21.3	29.9	28.8	1.2	76734.6
	atactic	31.7	0.12		19.2	33.0	25.8	1.5	68398.4
	syndiotactic	33.0	0.09		21.6	31.2	23.7	1.4	67130.3

Cont///

Monomer	Tacticity	R_g (Å)	STD	Average R_g (Å)	Rx (Å)	Ry (Å)	Rz (Å)	Ratio	Vol (Å^3)
HPMA-160	isotactic	37.0	0.06	35.0 ± 2	17.2	34.4	29.7	1.5	73236.5
	atactic	32.7	0.16		18.2	35.3	25.7	1.6	69182.5
	syndiotactic	35.3	0.10		29.5	37.1	18.5	1.5	84810.4
HPMA-180	isotactic	39.6	0.05	38.9 ± 1	23.9	31.9	27.2	1.2	86891.8
	atactic	38.7	0.19		19.7	57.7	24.7	2.6	117248.5
	syndiotactic	38.3	0.07		20.3	33.9	34.4	1.2	98862.1
HPMA-200	isotactic	40.7	0.12	42.5 ± 3	22.3	45.3	26.3	1.9	111624.0
	atactic	45.7	0.19		22.4	37.1	29.5	1.4	102741.3
	syndiotactic	41.3	0.17		25.6	36.3	30.8	1.3	120233.9
HPMA-220	isotactic	44.7	0.13	44.0 ± 2	23.3	39.7	28.3	1.5	109798.9
	atactic	43.6	0.10		24.2	38.4	28.1	1.5	109463.8
	syndiotactic	42.3	0.06		25.3	42.3	25.7	1.7	115258.0
HPMA-265	isotactic	45.5	0.05	49.0 ± 4	31.3	42.9	29.5	1.4	165614.0
	atactic	47.4	0.10		35.7	50.3	50.1	1.2	376274.3
	syndiotactic	42.1	0.07		25.2	58.9	40.0	1.8	247750.8

Table 4.1. Results of HPMA copolymers MD simulations. Obtained parameters are averaged over 200 000 timesteps over several simulations for each build parameters and calculated for equilibrated structures. STD – standard deviation, Rx,Ry,Rz – distances to a furthest atom in appropriate direction, where Ry – always largest and ratio is calculated from Ry/((Rx+Rz)/2)

4.4.2. Ellipsoid-based characterisation

As well as common methods of trajectory analysis, results were interpreted using an ellipsoid model which can provide more detailed characterization of polymer shape. This is achieved due to the separate axis analysis which makes it possible to obtain a unified density distribution profile for both spherical and non-spherical molecules. The ratio of distances to a furthest atom in the appropriate direction can be used to describe a wide range of molecules of different shape. Linear unfolded polymers can be described as a prolate spheroids with ratio of Rx ≅ Rz < Ry of 1:2.5-4 (Figure 4.5a); disc shape molecules can be presented as a oblate spheroids with Rx ≅ Rz > Ry (Figure 4.5b) or a tri-axial ellipsoid with Rx>Rz>Ry, where Ry is largest semi-axis of ellipsoid. Spherical shape can be assigned to cases when distances are equal or approximately equal for all directions (Figure 4.5).

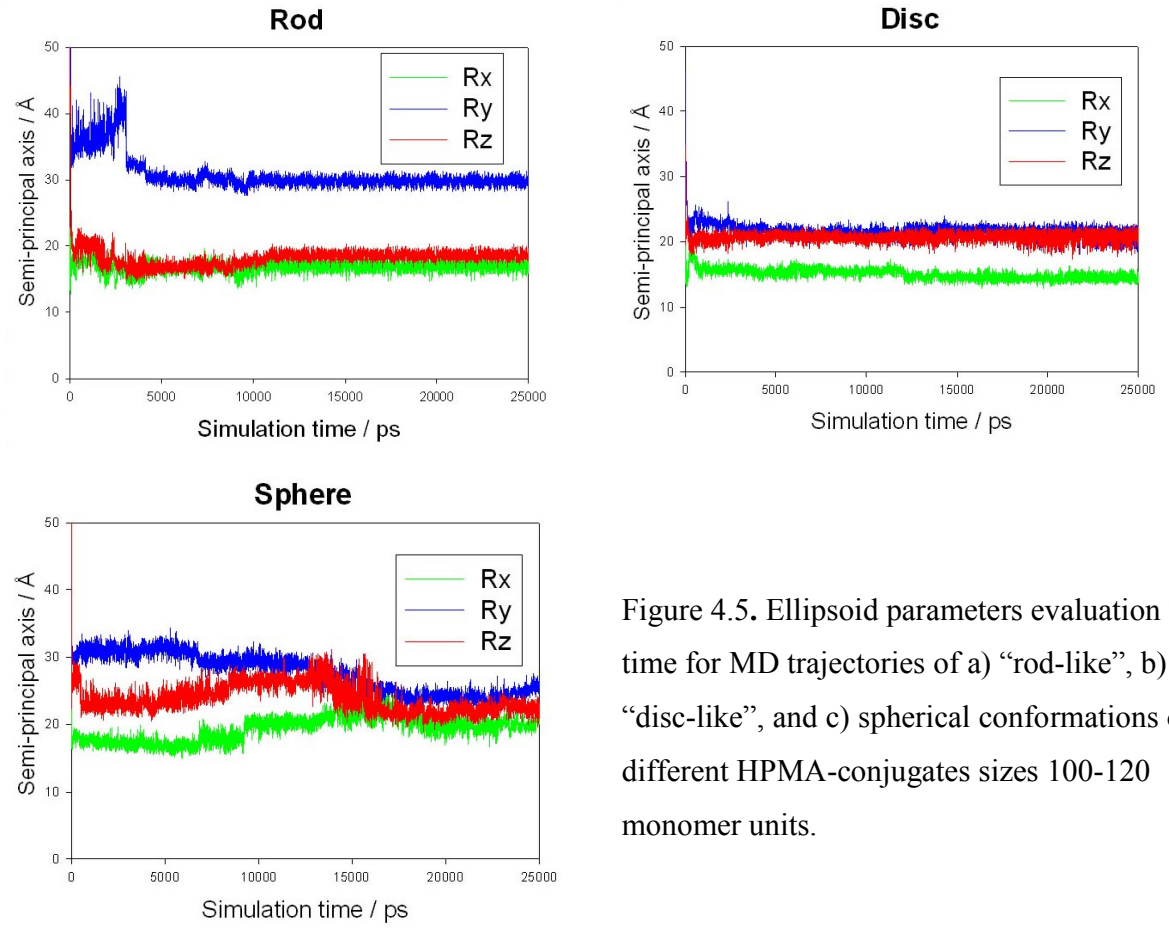


Figure 4.5**.** Ellipsoid parameters evaluation in time for MD trajectories of a) "rod-like", b) "disc-like", and c) spherical conformations of different HPMA-conjugates sizes 100-120 monomer units.

It was found that results of analysis for polymer chain collapse show that ellipsoid model is able to screen equilibration time more accurate and provide detailed analysis of changes in molecular shape and size then conventional R_g analysis (Figure 4.6.). While R_g time dependent plot flattens after 200 ps, separate axis analysis shows shape changes from 0 to 9 ns.

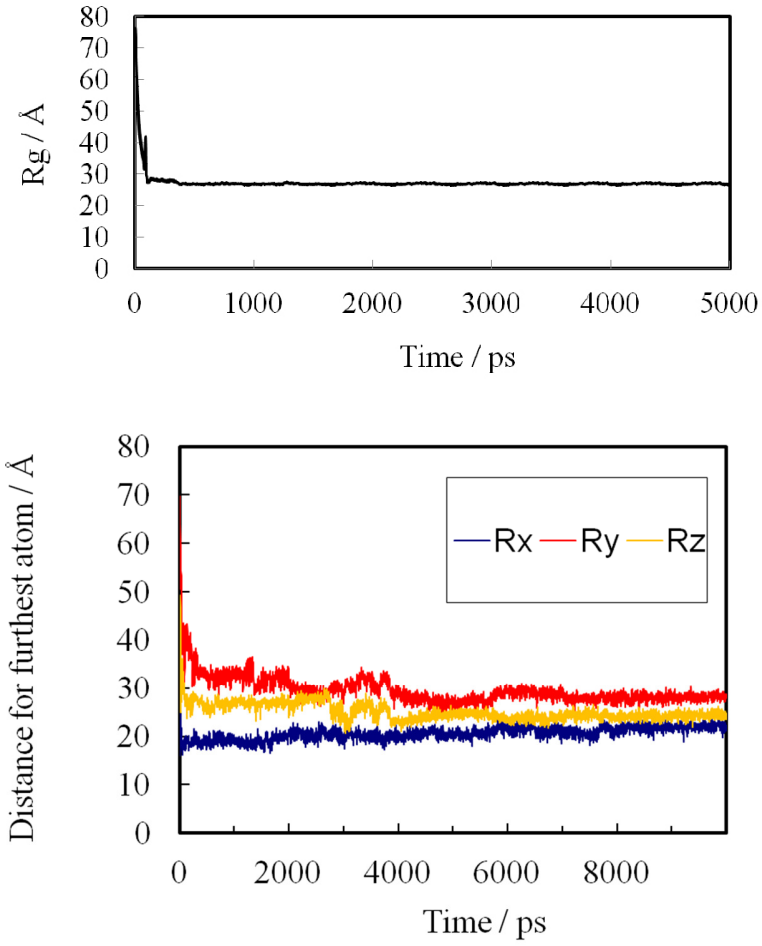
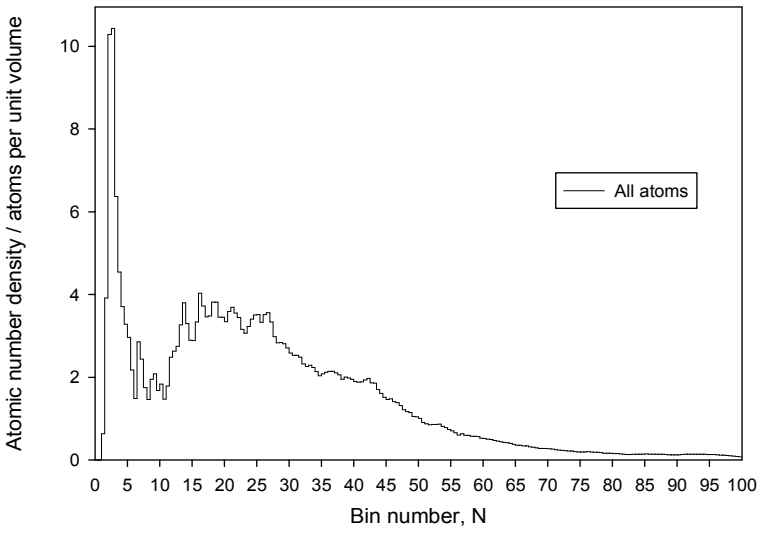


Figure 4.6: Comparison of separate axis analysis with R_g results for the same run of HPMA-120 homopolymer a) R_g evaluation in time 0-5 ns b) distances to the furthest atom in axis directions for 0-10 ns.

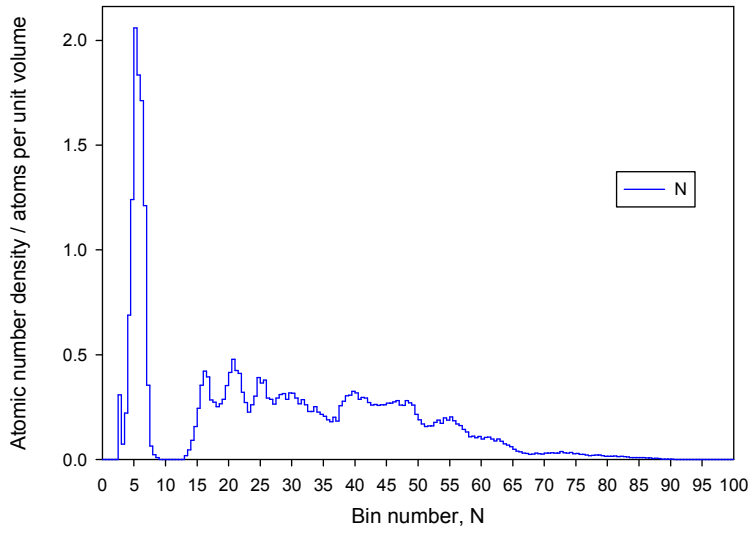
4.4.3. Density distribution

In addition to the overall shape of the polymer we wish to describe the distribution of atom types within the ellipsoid. This is an important aspect of polymers for drug delivery applications since positioning of the drug compound deep within the volume of the polymer will restrict accessibility and so a change of conformation in response to environment can be important in the delivery process. To probe this, we produce a density profile for a chosen atomic species through the volume occupied by the polymer. To do this we consider a series of ellipsoidal shells within the volume. The shell to which an atom belongs is identified by putting the atom co-ordinates relative to the centre of mass in the inertial axis system into the equation for the ellipsoid. The fractional result can be used to assign the atom to a particular shell. For practical purposes we use 100 shells with atoms binned into the nearest appropriate shell so that a histogram can be produced to show the distribution of any atom type through the volume occupied by the polymer. By carrying out this process on all atoms we arrive at a density profile and, provided the shape of the polymer has reached equilibrium, averaging over frames of the trajectory improves the statistical significance of the histograms produced. Such histograms display structural aspects of the polymer, for example whether a core region exists in which one particular atom type is concentrated.

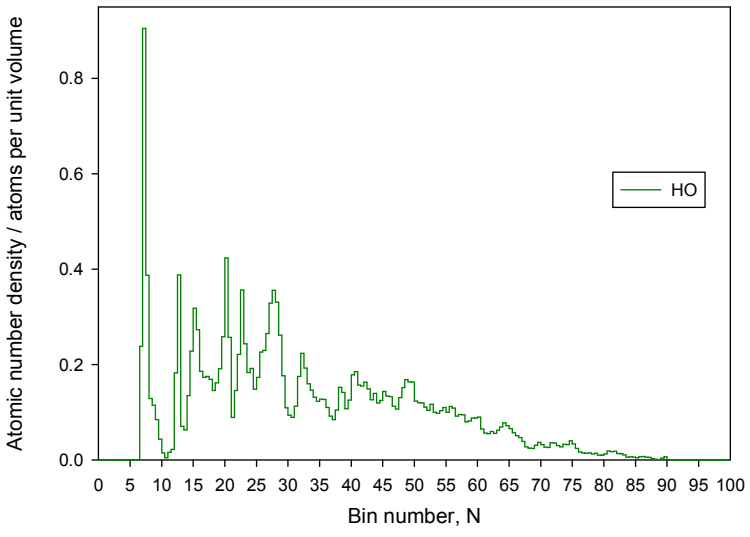
Absolute distances from the molecular centre can be obtained by scaling to the value of the largest axis. Results can therefore be interpreted as an atom density distribution from the center of mass, or center of scattering length, in all directions within the ellipsoid. Density profiles for HPMA-265, shown in Figure 4.7 exhibit a main peak in density distribution 0-3 Å from the center of the mass, which can be explained by relatively smaller sizes of bin volume, and is followed by drop in the number of atoms per volume and a series of peaks in 9-27 Å. As well 87.7% of the overall mass of the polymer is allocated in half of the ellipsoid space from the center of the mass (0-30Å).


Atomic number density / atoms per unit volume
Bin number, N
All atoms
0
2
4
6
8
10
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100

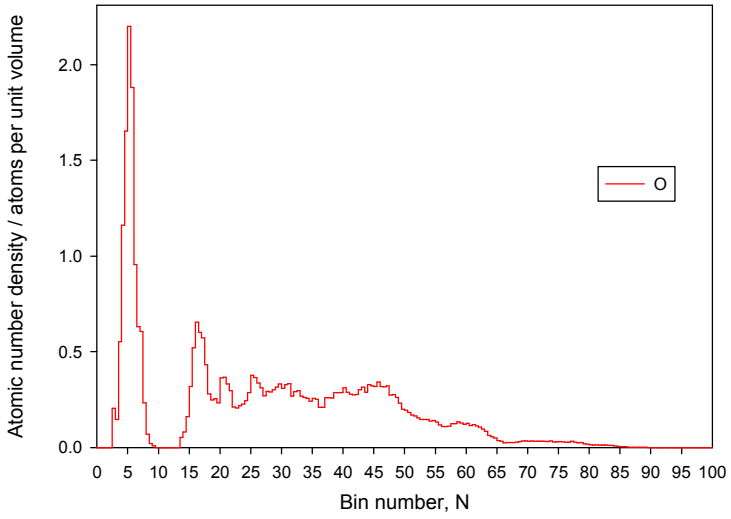
(a)


Atomic number density / atoms per unit volume
Bin number, N
N
0.0
0.5
1.0
1.5
2.0
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100

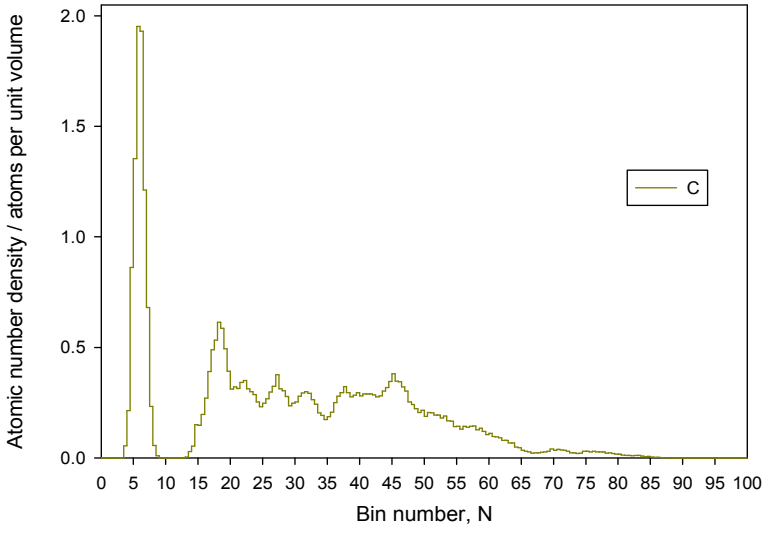
(b)


Atomic number density / atoms per unit volume
0.0
0.2
0.4
0.6
0.8
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Bin number, N
HO

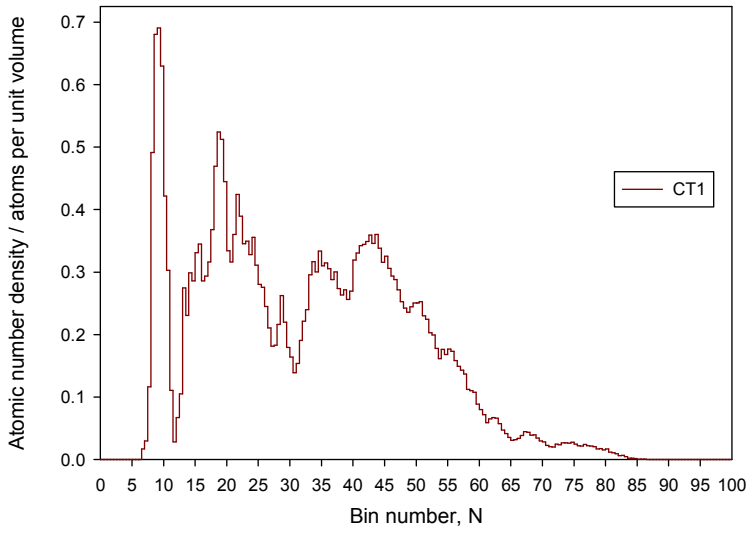
(c)


Atomic number density / atoms per unit volume
0.0
0.5
1.0
1.5
2.0
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Bin number, N
O

(d)



(e)

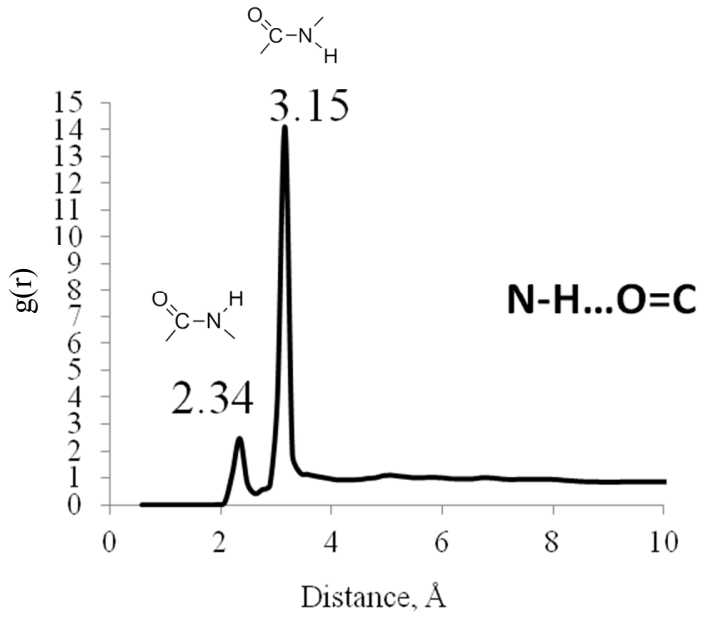


(f)

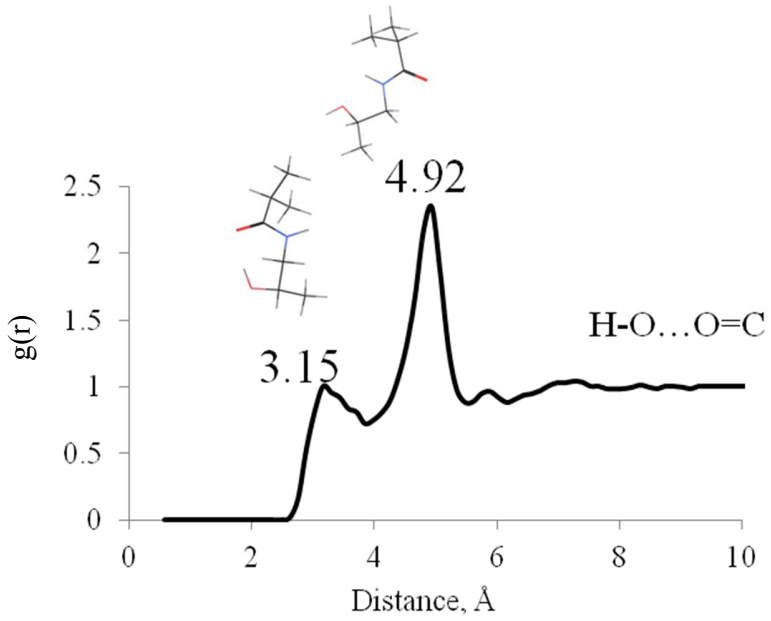
Figure 4.7: Density distribution profiles averaged from MD trajectory of equlibrated HPMA-265 polymer fitted in ellipsoid with dimensions 31.3 Å, 42.9 Å and 29.5 Å for a) all atoms, b) Nitrogen atoms, c) Hydrogen in hydroxylic group, d) Oxygen in carbonyl group, e) Carbon in carbonyl group, f) Carbon in polymer backbone.

4.4.4. Radial distribution function

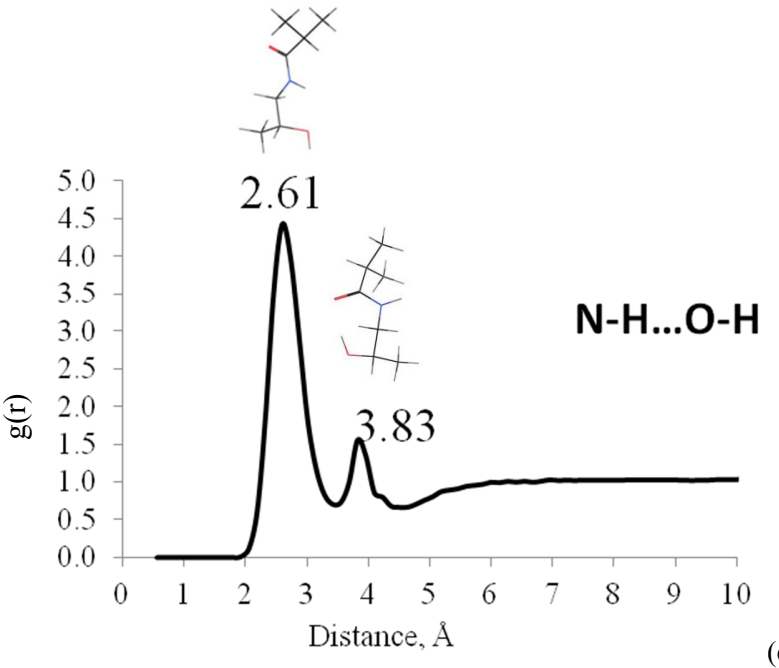
Radial distribution functions for selected atoms were calculated from MD trajectories of HPMA-265. Results of the N-H...O=C distance distribution are shown in Figure 4.8.a, and indicate the presence of both *cis*- and *trans*- isomers of the peptide group in the folded polymer, with a ratio of trans to cis isomers of 5.5:1. The side group of HPMA can also take on two distinct conformations, an internal hydrogen bond between the alcohol and carbonyl groups gives a "coil" like conformation while in the alternative extended form the alcohol and amine groups are in relatively close proximity. For the H-O...O=C and N-H...O-H radial distribution functions (Figure 4.8.b,c) the results indicate that the extended form of the HPMA monomer is preferred within the polymer chain with a ratio of extended conformation to coiled of 2.5:1. This corresponds to the relative calculated energies for the monomer unit which gave a difference between conformations of 2.2 kcal/mol (AMBER99). Results for the N-H...O=C radial distribution function (Figure 4.8.d) also suggests the presence of intermonomer bonding for 1/4 of total number of monomer units. In general, results of RDF calculations indicate the presence of hydrogen bonding both within monomer units and between different monomers within the polymer chain.



(a)



(b)



(c)

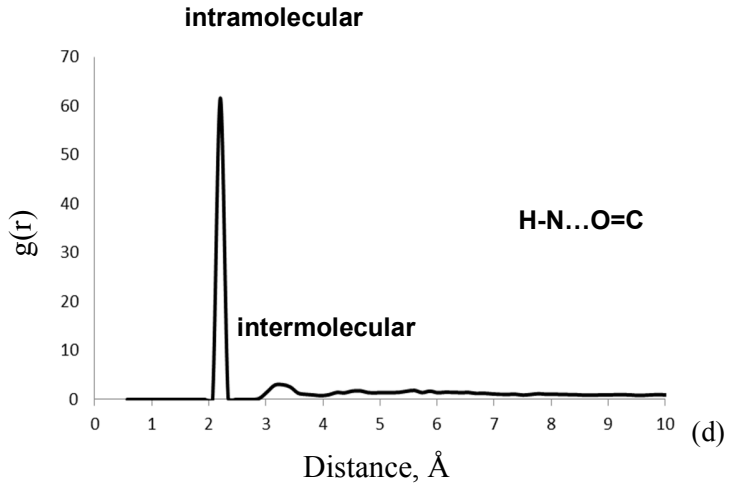


Figure 4.8**.** Radial distribution function of a) N-H...O=C b) H-O…O=C c) N-H…O-H and d) H-N…O=C as above bonds obtained from MD trajectories and averaged over 200,000 conformations of equilibrated structure of HPMA-265.

4.5. Discussion

In order to select a solvent model suitable for polymeric HPMA, R_g was obtained for explicit water and two implicit solvent models; reaction field (R-field) and distance-dependent dielectric. While the explicit water model would be expected to give the most accurate model of the solvated polymer, the large system sizes required for the higher molecular weight polymers become prohibitively expensive in terms of computer time. The R-field and distance-dependent dielectric solvent models are implicit and so involve only minor computational overheads compared to simulation of the HPMA polymer alone.

Figure 4.3 shows R_g data for HPMA polymers from MD simulation in different solvent models, and demonstrates striking differences between these models. It is evident that R_g resulting from explicit water simulation is the largest for a given value of N, while that from the R-field model is significantly smaller and the distance-dependent dielectric solvation model leads to much smaller values of R_g. Lines of best fit with the same power law form as the Flory approximation results in fitted values of ν of 0.63 for explicit solvent and 0.58 for R-field. These values are close to the ideal of 0.60 for "good" solvents. However, the distance dependent dielectric constant approach to solvent simulation yields ν=0.40. This indicates that simulations using a distance dependent dielectric constant representation of the water solvent would lead to a much more compact HPMA

conformation than using an explicit water model, whereas the R-field approach gives rather more expanded polymer that is in better agreement the explicit water results. Best fit values of R_0 are 1.88, 1.92 and 2.76 Å for explicit water, R-field and distance dependent dielectric models, respectively. These compare with an average over all conformations of HPMA monomer of 2.50 Å, indicating that in realistic solvent models each repeat unit is more compact in the polymer.

For N = 265 experimental SANS data is available which gives the value of R_g = 70±5 Å. For the explicit water calculations computer resources limit us to a maximum chain length of N = 180, corresponding to *ca.* 72,000 water molecules. However, extrapolating the power law best fit line yields an estimated value of 62 Å in explicit water. The R-field results for N = 265 give R_g averaged from five independent simulations as 49±4 Å, while extrapolation of the power-law fit gives 48 Å. In contrast, the distanc edependent dielectric model extrapolates to just 26 Å. Thus, it is clear that the implicit solvent models considered in this work give rise to significantly more compact structures than does explicit solvation, and that the latter is in markedly better agreement with experiment than the former. However, simulation of the full N = 265 polymer in explicit solvent is beyond our current computational resources. An alternative to all-atom MD simulation is to use coarse-grained methods, which will be discussed in Chapter 6.

Results for oligomers and polymers of various sizes, simulated using R-field solvent model, are presented in Table 4.1 and suggest no systematic influence of tacticity of the initial linear polymer structure on results of MD simulations and corresponding parameters. It was found that overall shape of the molecule changes with polymer size from approximately spherical for smaller molecules to more elongated for larger structures. Data for the final entry in Table 4.1. for HMPA-265 also indicate that the atactic form is rather more spherical than either iso- or syndiotactic forms, as evidenced by a smaller ratio of long: short axis. Moreover, the value of separate axis analysis is clear in this data, since structures with quite different overall shape may have similar R_g values.

4.6. Conclusions

MD calculations were performed for model HPMA conjugates and methods refined to obtain reproducible results. After a number of tests AMBER99 force field was proven to be suitable for HPMA modelling. An explicit solvent model shows R_g close to the SANS experimental result, but

long simulations of larger polymers requires excessive computing resources. The reaction field implicit solvent model showed good results in both efficiency and accuracy so it was chosen for further investigation. A distance-dependent dielectric model of solvation was found not to be accurate for MD simulations, as it gives rise to too small radius of gyration, with the polymer chain forming a hard sphere. Along with running MD simulations, shape analysis tools were developed to determine the overall shape of the polymers in solution. Distribution functions were obtained based on an ellipse model, using i) overall atom distribution; ii) distribution of atom type; iii) specific labelled atoms (e.g. locate drugs or linkers). Adapting this approach for calculating scattering length density profiles will allow us to predict SANS data and compare with that obtained from experiments. Obtained data and established methodology can be now expanded to be applied for investigation of hetero-polymers, such as HPMA-GFLG-copolymers with mono or mixed drug parts conjugated.

4.7. References

[1] Polymer Therapeutics I: Polymers as Drugs, Conjugates and Gene Delivery Systems, ed. R. Duncan, H. Ringsdorf and R. Satchi-Fainaro, Springer, New York, 2006, vol. 192, pp. 1–8.

[2] M. Laradji, H. Guo and M. Zuckermann, Phys. Rev. B: Condens. Matter Mater. Phys., 1994, 49, 3199

[3] L. Huynh, J. Grant, J. C. Leroux, P. Delmas and C. Allen, Pharm. Res., 2008, 25, 147

[4] R. Chang and A. Yethiraj, J. Chem. Phys., 2011, 114, 7688

[5] R. Duncan Nat Rev Drug Discov. 2003, 2:347-60.

[6] R. Chang and A. Yethiraj, J. Chem. Phys., 2011, 114, 7688.

[7] Paul, A., Vicent, M. J. and Duncan, R. Biomacromolecules 2007, 8, 1573-1579.

[8] K. Solc and W. H. Stockmayer, J. Chem. Phys., 1971, 54, 2756;

[9] D. N. Theodorou and U. W. Suter, Macromolecules, 1985, 18, 1206.

[10] E. J. Rawdon, J. C. Kern, M. Piatek, P. Plunkett, A. Stasiak and K. C. Millett, Macromolecules, 2008, 41, 8281.

[11] N. Rawat and P. Biswas, J. Phys. Chem. B, 2012, 116, 6326.

Chapter 5: All-atom modelling of conjugates

5.1. Introduction

The conformations that polymer-drug conjugates may form in solution have a significant effect on properties that are important for designing of drug delivery systems, and can lead to the ability to keep drug levels at biologically active and safe concentrations for desired period of time [1]. It is also known that conformations of polymers formed in solutions depend on type and amount of side chains [2]. Based on Chapters 3 and 4, we are in a position to model polymer conjugates with various drug-mimics and compare computer modelling with experimental data. Methodology established in Chapter 4 was used to prepare and perform simulations, with additional atom types and parameters for drug-mimics introduced and tested. We have chosen a range of both hydrophilic and hydrophobic side-chains, based on previous studies of conjugates carried out in the Soft Matter research group. Results of experimental data analysis were compared with parameters predicted by computer simulation. Based on this comparison, we were able to make conclusions on the adequacy of chosen models as well as develop tools that can be useful for further investigations.

5.2. Previous studies

This project builds on the experimental studies of structural dependence for HPMA conjugates on side chain type and loading reported in Chapter 3. Trying to replicate experimental results, we have also investigated changes in hydrophobicity for both linear and aromatic systems with the same drug mimics and degree of loading as in experiments. Polymer conjugates were modelled using random distribution of side chains within the molecule. Following Chapter 4, we adopted analysis tools to compare shape and size parameters rather than simply calculating R_g. Counting the fact that shapes in SANS models are represented as a combination of geometrical forms, such as sphere, rod, coil, etc. we have used ellipsoid-based characterization tools. This allowed us to compare experimental results of SANS fitting with computer simulation results, obtained from trajectory analysis. This comparison will be shown and discussed later in this chapter.

In general, the results from previous structure and morphology study were used to help correlate modelling to real measurements, so that a suitable methodology for MD simulations can be

developed. Per contra, some of the modelling results suggest that experimental results can be clarified with theoretical predictions, based on these results. Although there are differences in data obtained from SANS experiment fitting and MD simulations, it is clear that introducing modelling techniques to drug discovery projects can be beneficial for refining existing theoretical models as well as better understanding of experimental data, and ultimately will allow us to establish intelligent approach in drug-polymer conjugates design.

5.3. Polymer-drug conjugates modelling

It is known that combining molecular modelling techniques with experimental methods data analysis can provide reliable model for polymer-drug conjugate characterization [4]. Proven to be a reliable tool for morphology investigation [5], molecular modelling of polymer behavior is demanding for computer resources as well as complicated in establishing and validating of developed model. Among the challenges are the size of molecules and simulated systems (tens of thousands atoms of polymer as well as hundreds of thousands solvent molecules), necessity to parameterize custom atom and bond types and a need to choose or introduce relevant parameters for comparison. Despite these difficulties, increasing amounts of research of polymers involve computer simulations and more publications on polymer-drug conjugates modeling were introduced recently [6].

One of the main reasons for such an intensive interest in involving computational chemistry methods for medical studies are effective decrease of investigation time – while being highly accurate, experimental methods often involve use of rare and expensive equipment. Another concern is low effectiveness for overall process of drug discovery, which requires improvement; on average, about 35% of discovery projects succeed in delivering experimental drugs suitable for clinical testing. The project stages of target identification and screening, hit-to-lead, lead optimization, and preclinical candidate selection have individual success rates ranging from 69 to 85% [7]. This leads to a conclusion that ability to predict chosen properties before time and money consuming process of testing and trials can be a major advantage.

Some recent publications focus on polymer-drug characterization using multiscale modelling tools along with experimental data techniques. [6, 10, 11]. PEG and PGG are mainly chosen as polymer carriers in such research, which shows correlation between theoretical and experimental

parameters. Proving the concept of computer simulations being able to predict specific properties, these projects also reveal trends for future drug discovery process. In such research, where there is a need to work with massive database of combinations, it is crucial to have an instrument that can provide preliminary data analysis to define pathways for further investigation. For example, reference [6] explored how the shape and size of poly-γ-glutamyl-glutamate paclitaxel (PGG-PTX) can be controlled by varying hydrophobic and hydrophilic loadings. In order to do so, all-atom MD simulations of PGG-PTX were run and then continued as coarse-grained (CG) simulation. Results show that a PGG-PTX molecule has a strong tendency to form coil shapes, regardless of the PTX loading fraction and spatial PTX arrangement, although globular shape was found for certain conditions. This is valuable information for drug-delivery purposes, suggesting PGG-PTX may confer a long circulation half-life and result in significant accumulation of the drug at the tumor site. As well as obtaining useful data for comparison and investigations, all-atom MD simulations allowed mapping and collection of other information on coarse-grained models of simulated polymer conjugates. Following these steps we have used obtained data to progress to CG modelling which will be discussed in the next chapter (Chapter 6).

5.3.1. MD simulations of HPMA conjugates

Following Chapter 4, the AMBER99 parameter set was chosen for all molecular dynamics simulation. Reaction field (R-field) was chosen as a suitable solvent model based from solvent model comparison. Added atom types were tested to match AMBER99 force field parameters. This approach was used for investigation of HPMA conjugates with drug mimics. A range of linear amines (aminohexane(C6), aminooctane(C8), aminododecane (C12)), hydroxyl and fluoro terminated linear amines as well as aromatic aminoanthracene (ANC), aminocrysene(AC) and aminoanthraquinone (ANQ), bound to the polymeric carrier via a tetrapeptide linker glycine-phenylalanine-leucine-glycine (GFLG), which is cleaved by lysosomal cathepsins at low pH and was identified as a suitable spacer, were selected as model objects for study of the effect of drug type and loading on HPMA copolymer conformation. For 5% loading, 6 co-polymer-GFLG-R fragments were randomly incorporated into the HPMA molecule, and for 10% loading 17 fragments were randomly distributed within the polymer chain along with a random mixture of 86 R- and S- optical isomers of HPMA fragments. SANS experimental scattering curves were compared with theoretical curves, predicted from MD simulations. Parameters such as size and

shape fitted to SANS data were compared with relevant simulated structures. Results of MD simulation in R-field solvent model are presented in Table 5.1. Table 5.2 contains relative results for SANS best fitted models. Numbers in conjugate names correspond to loading percentage.

Ellipsoids						
Conjugate	Rg, Å	a, Å	b, Å	c, Å	Volume ($\times 10^{-19}$ cm^3)	Aspect ratio
HPMA-ANC-10	22.80	25.39	47.29	23.42	1.138	1.91
HPMA-ANQ-10	21.43	23.64	45.56	28.99	1.320	1.73
HPMA-AC-10	21.13	19.82	71.60	32.80	1.252	2.80
HPMA-AP-10	20.18	24.75	35.74	28.79	1.066	1.33
HPMA-C12-5	20.27	24.56	37.85	30.74	1.196	1.37
HPMA-C12-10	21.04	26.49	36.49	33.45	1.353	1.22
HPMA-C6-5	19.49	22.79	41.11	31.25	1.226	1.52
HPMA-C6-10	21.28	24.49	35.06	29.26	1.052	1.30
HPMA-C6-F-10	23.24	19.16	49.12	32.36	1.275	1.91
HPMA-C8-F-10	20.15	22.00	36.45	30.34	1.018	1.39
HPMA-C6-OH-10	22.15	24.62	41.94	25.96	1.122	1.66
HPMA-C8-OH-10	25.16	22.5	43.8	29.87	1.232	1.67

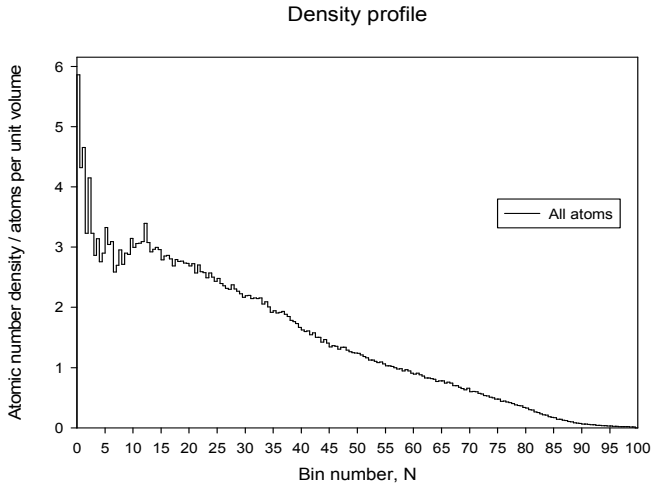
Table 5.1: MD simulation results for chosen conjugates, where a,b,c – enclosing ellipsoid semi-axis and aspect ratio is main axis to averaged remaining semi-axis ratio.

	Sphere	Cylinder		Volume ($\times 10^{-19}$ cm^3)	Aspect ratio
Conjugate	Radius, Å	Length, Å	Radius, Å		
HPMA-ANC-10	-	105±5	25	2.060	2.10
HPMA-ANQ-10	45	-	-	3.815	1
HPMA-AC-10		130±5	25	2.551	2.60
HPMA-AP-10	-	-	-	-	-
HPMA-C12-5	-	385±5	25	7.556	7.70
HPMA-C12-10	-	145±5	32	4.662	2.26
HPMA-C6-5	89	-	-	29.515	1
HPMA-C6-10	-	155±5	23	2.575	3.36
HPMA-C6-F-10	-	140±5	15	0.989	4.69
HPMA-C8-F-10	-	160±5	25	3.140	3.20
HPMA-C6-OH-10	-	135±5	16	1.085	4.22
HPMA-C8-OH-10	-	140±5	15	0.989	4.69

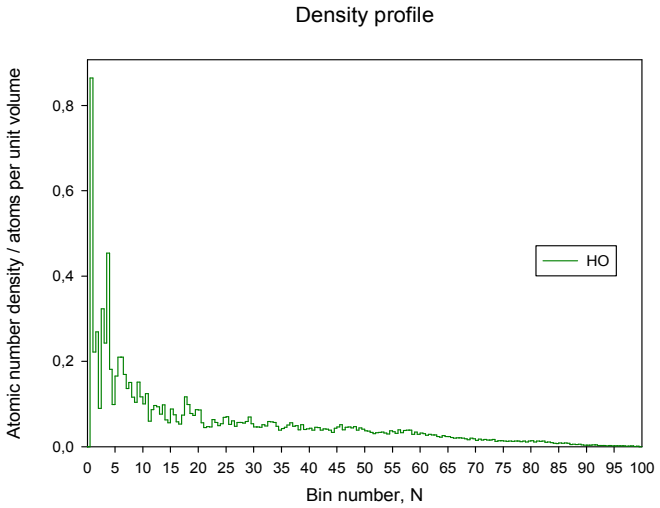
Table 5.2: SANS fitting results for chosen conjugates, where Length and Radius are relevant length and radius of fitted cylinders or spheres, aspect ratio is length to radius ratio. [12]

Our ellipsoid analysis tool allowed us not only to unify experimental data with MD simulation results by comparing shapes of obtained molecules, but also to analyze morphology. Both for all and particular atom types, density distribution within the structure is valuable information for morphology investigation, as shown in figure 5.1. Ultimately this will allow screening of drug distribution or, for example, solvent molecules' density profile which can be compared with experimental data. These data can help to explain changes in solvent behavior of polymer-drug conjugates based on variation of drug part and loading ratio.

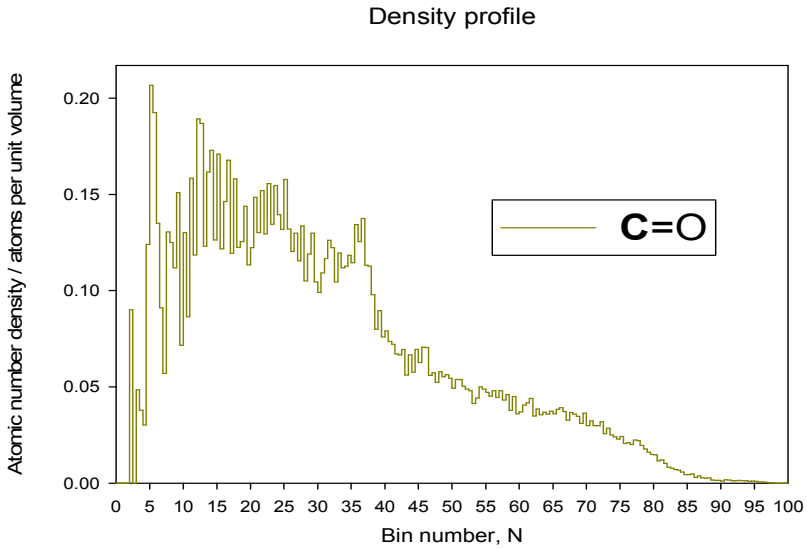
a)



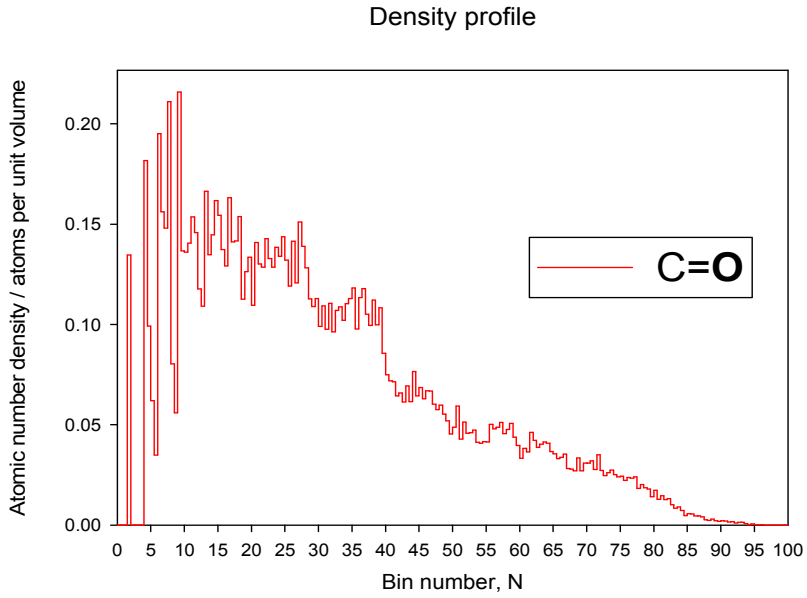
b)



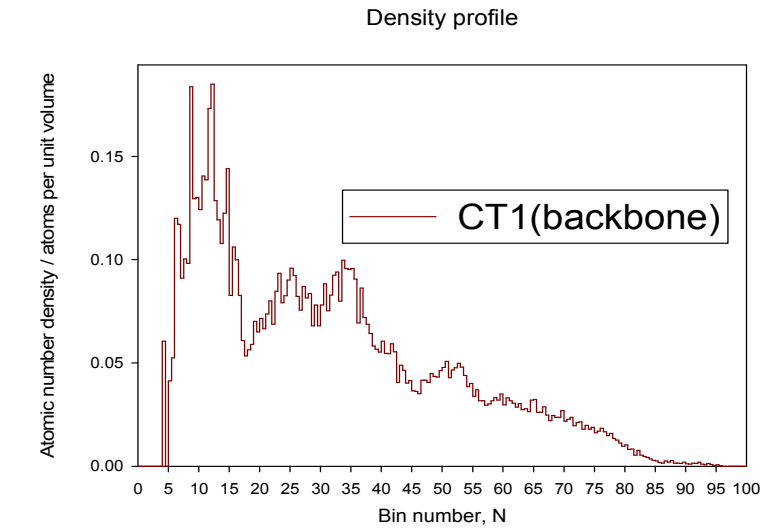
c)

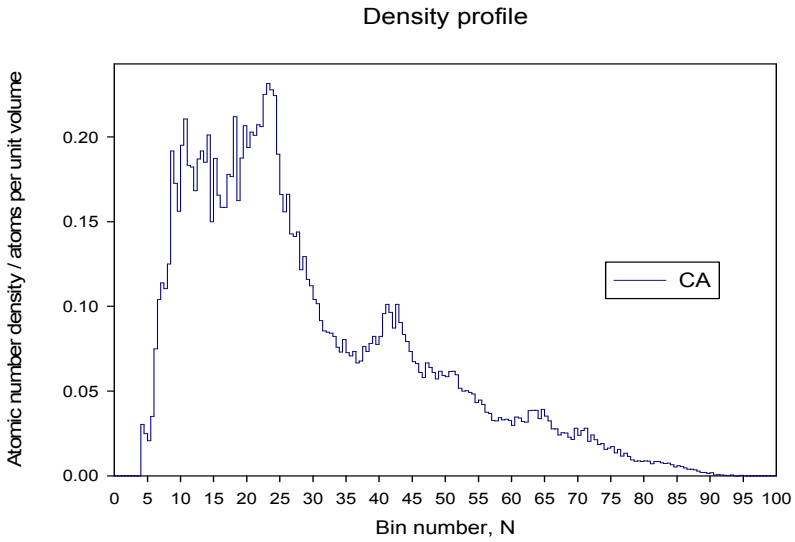

Density profile
Atomic number density / atoms per unit volume
0.00
0.05
0.10
0.15
0.20
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Bin number, N
C=O

d)


Density profile
Atomic number density / atoms per unit volume
0.00
0.05
0.10
0.15
0.20
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Bin number, N
C=O

e)





(f)

Figure 5.1: Density distribution profiles averaged from MD trajectory of HPMA-GFLG-ANC for a) all atoms, (b) Oxygen in hydroxyl, (c) Oxygen in carbonyl, (d) Carbon in carbonyl, (e) backbone C atom, (f) aromatic Carbon (drug-mimic atoms)

5.3.2. Result analysis

Results presented in Table 5.1 and 5.2 represent data obtained by SANS and MD simulation for replicated structures after equilibration. All conjugates were fitted to cylinder model and can be also described as elongated ellipsoids. To match these shapes we have applied ellipsoid base model and obtained a range of ellipsoids from sphere-like to rod-like. For linear amine (C6 and C12) conjugates, both for 5% and 10% loading we find that MD simulation suggests oblate spheroid shapes for equilibrated structures ($a \sim b > c$). Experimental data shows more rod like shape for corresponding conjugates. We have also noticed differences in volumes for obtained structures – for both C6 and C12 conjugates with 5% side chain loading these differences are more striking, but even for 10% loading the predicted volume is several times larger than experimental results. While having volumes calculated as 1.05-1.35 $\times$ 10^{-19} cm^3 for MD simulated structures, SANS results suggest 2.5 – 29.5 $\times$ 10^{-19} cm^3 for linear C6 and C12 amines. Meanwhile, aspect ratios for ellipsoid/cylinder dimensions are also quite different, reaching a maximum of 7.70 for HPMA-C8-F-10 from experimental data fitting with corresponding 1.196 result from modelling results [12]. For HPMA conjugates with Hydroxyl and Fluoro-terminated amines, MD generally predicts similar oblate and prolate spheroids as for linear amine substituents, though predicted volumes mostly match experimental data (except HPMA-C8-F-10, where SANS fitting suggests 3.140 $\times$ 10^{-19} cm^3) slightly overestimating values of SANS fitting results.

For both ANC and AC, simulation follows experimental data reasonably well, showing similar aspect ratio (1.91 and 2.80 from MD; 2.10 and 2.60 from SANS for ANC and AC respectively) but different volumes (1.138 $\times$ 10^{-19} cm^3 and 1.252 $\times$ 10^{-19} cm^3 against 2.060 $\times$ 10^{-19} cm^3 and 2.551$\times$ 10^{-19} cm^3). In general, both MD and SANS data analysis suggest a shape of elongated ellipsoid for ANC and AC substituents, but for HPMA-ANQ-10 experimental results demonstrate sphere-like structure, where computer simulation predict cylindrical shape. However, dimensions of “real” molecules are underestimated by computer simulations with a factor of 2-4 by difference in volume of obtained structures.

There are various factors that can contribute to this major difference in volumes between computer simulated and suggested by best fits of SANS data. First, simulations were carried with a single polymer conjugate molecule and there are evidence of aggregation of HPMAconjugate molecules in solvent which will be discussed later in this Chapter. Secondly, SANS suggested structures are

able to describe variation of shapes for evaluated molecules but from ellipsoid model which is used for computer simulation data analysis – obtained volumes represent enclosed ellipsoid volume for a molecule. So, for example, the same molecule would be interpreted with a different models and results for volume would be different. Third reason is accuracy, there are limits on how precise data obtained from experiment, including both instrumental and chosen error, but from computer simulations obtained dimensions are extremely precise. Finally, selected solvent model can require additional changes in order to prevent polymer molecules for rapid collapse or overfolding. Explicit solvent simulations are suggested as alternative source of data generation.

However, MD simulations provided valuable data and suggested important information for both experimental and modelling future studies. For example, modelling can provide information on changes in volumes within the simulation and suggest structures which have similar volume to SANS data. Such comparison is provided for HPMA-ANC-10 simulated and experimentally obtained structures (Figure 5.2). For HPMA-C6-5 conjugate it is shown that single molecule cannot represent SANS provided data which suggest an evidence of aggregation or error in analysis of experimental data (Figure 5.3).

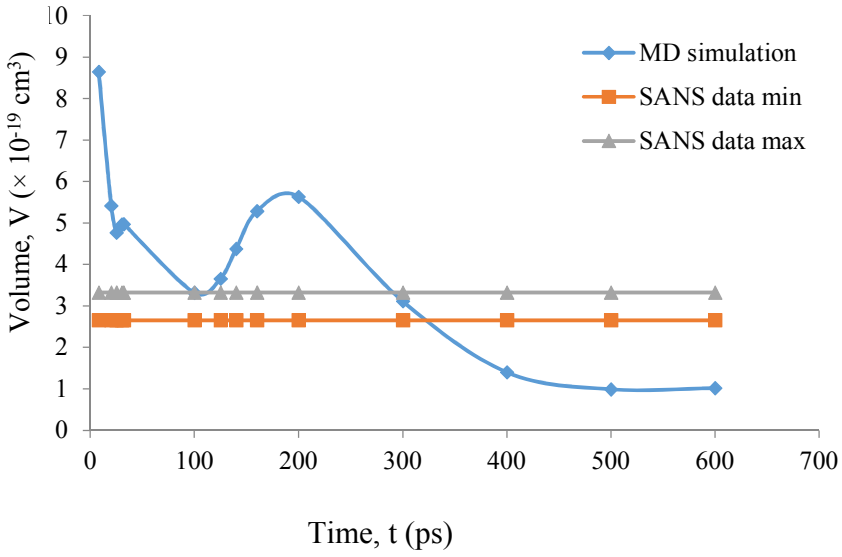


Figure 5.2: Volume comparison between SANS and MD simulation data for HPMA-ANC-10

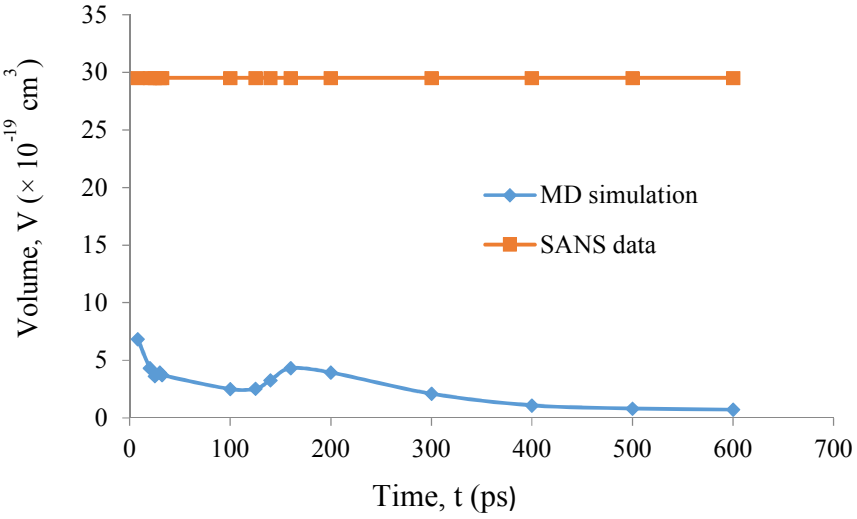


Figure 5.3: Volume comparison between SANS and MD simulation data for HPMA-C6-5

Density profiles analysis provided ability to screen distributions of various atoms within the molecule. In general, atoms tend to localize near to the center of mass. Approximately, 50-55% of atoms for all obtained results for HPMA-copolymers were allocated within a 25% of the ellipsoids volume, which is surrounding the center. At the same time hydroxyl atoms show even more core-shell like behavior, allocating ~ 90% of this atom type within the nearest 10% to the center of mass. This can be explained by relatively strong hydrogen bonds that are formed and might be investigated further with explicit solvent model MD simulations. For both atoms from carbonyl group we can observe similar behavior with most atoms distributed evenly within the 50% of ellipsoid volume.

One of the most desirable features of this method is the ability to screen distribution of drugs or drug-mimic parts within the molecule, crucial information for investigating drug-polymer properties. Being able to obtain information on whether atoms are mainly shielded from the solvent or localized on the surface of the molecule is a useful tool for polymer-drug characterization and modelling results analysis. From results of MD simulation, shown on Figure 5.1f for CA atom type, which corresponds to aromatic carbon atoms in ANC group, we can observe some evidence for core-shell like structure, with 60% of atoms allocated in 30% of ellipsoid space around center. This is followed by a drop of density, demonstrating similar behavior to other atoms. This corresponds to low solubility of aromatic groups and follows available experimental data.

5.5. Conclusions

Computational simulation of polymer-drug conjugates has a practical application in obtaining details on morphological information of polymer structure, and can be combined with experimental techniques to provide a view of solution behaviour. This allows design of new materials for drug delivery, characterization of known biological agents and investigation of structure dependence on copolymer types and drug loading in conjugate systems. Results for aromatic drug conjugates showed better agreement with experiment than linear amines, and suggest evidence of aggregation for some structures. Designed tools can be used for various MD simulation result analysis, including particular atom or group of atoms position screening, density profile and sizes in time evaluation. This leads to deeper understanding of obtained modelling data for particular HPMA-polymers study, as well as any other MD simulation data analysis in various fields. However, more simulations are required for full investigation, which includes explicit solvent simulations, expanding database of simulated structures, including mixed conjugates and obtaining comparison for other available SANS data.

5.6. References

[1] R. Duncan, Y.-N Sat Ann. Oncol. 1998, 9 (Suppl.2): 39.

[2] P. C. Griffiths, A. Paul, B. Apostolovic, H.-A. Klok, E. de Luca, S. M. King, R. K. Heenan, J. Controlled Release 2011, 153, 173-179

[3] G. Meleshko, J. Kulhavy, A. Paul, D. J. Willock and J. A. Platts, RSC Advances, 2013

[4] J Comput Aided Mol Des. 2012 January; 26(1): 15–26.

[5] N. Koga, R. Tatsumi-Koga, G. Liu, R. Xiao, T. B. Acton, G. T. Montelione, D. Baker Nature. 2012, 491: 222-7.

[6] L. X. Peng, A. Ivetac, A. S. Chaudhari, S. Van, G. Zhao, L. Yu, S. B. Howell, J. A. McCammon, D. A. Gough Biopolymers 2010, 93, 936-951.

[7] S. M. Paul, D. S. Mytelka, C. T. Dunwiddie, C. C. Persinger, B. H. Munos, S. R. Lindborg, A. L. Schacht Nat Rev Drug Discov 2010, 9:203–214.

[8] J. A. McCammon, B. R. Gelin, M. Karplus Nature 1977 267:585–590.

[9] J. J. Wendoloski, Z. R. Wasserman, F. R. Salemme, J Comput Aided Mol Des 1987, 1:313–322

[10] L. X. Peng, A. Ivetac, A. S. Chaudhari, S. Van, G. Zhao, L. Yu, S. B. Howell, J. A. McCammon, D. A. Gough Biopolymers 2010, 93, 936-951.

[11] L. X. Peng, S. K. Das, L. Yu et al J. Mol. Model 2011, 17 2973-2987.

[12] C. James PhD Thesis Cardiff University 2011.

Chapter 6: Coarse-grained modelling of HPMA

6.1. Introduction

AA simulations of HPMA and its conjugates were carried in various solvent models obtain methodology for HPMA-polymer simulation and prove concept of calculation reliability. However, there are some limitations for AA-simulation, especially time limitation – some processes can occur on a time scale that is too long to be studied by atomistic simulations. These include the dynamics of full-scale biopolymers in explicit solvent or, for example, self-assembly of biological materials. Despite the variety of optimization techniques and methods suitable for evaluating molecular systems of various size, AA MD simulations of a relatively large systems can require enormous CPU power which makes it unreasonable to perform.

Coarse-grained (CG) molecular modeling allows simulations to be run on length and time scales that are 2–3 orders of magnitude larger compared to AA ones, by treating evaluated systems as separate "beads" consisting of multiple atoms, rather than individual atoms. This has a massive impact on CPU requirements, contributed by decreasing number of particles for simulation, especially explicit solvent [1]. For example, MARTINI force field is known to accelerate the dynamics of water by approximately a factor of 4 [8]. Longer simulation times can be reached for each system by use of larger time steps for stable simulation systems. For example, typical time steps used are tens of femtoseconds for MD, and >100 femtoseconds for dissipative particle dynamics (DPD), compared with 1–4 femtoseconds employed in AA MD. Another factor that makes CG models faster is the reduced number of degrees of freedom (DOF) which leads to smothering of the potential energy surface and reducing friction between particles [12]. These factors make CG simulations a powerful tool for biological simulations.

Even though setting up CG model and validating selected parameters for simulations need additional time and require specific tools to be developed, such advantages as increase in simulation speed and longer simulation time available makes them beneficial for research in many areas of molecular simulation. However AA modelling is still required for examining molecules in detail, as well as establishing initial models and can be combined with CG, as well as QM/MM data to provide fully multiscale modelling of objects of interest.

6.2. Model and methods

6.2.1. Force field selection

Following previous reports [1-7], we have chosen the MARTINI force-field as a suitable approach for CG molecular modeling, as it was successfully applied for numerous biological simulations including protein and polymer simulations. In order to perform simulations we have established initial mapping of interaction centers or beads, then compared result analysis for AA and GC simulations of relatively small HPMA fragments (8-32 monomer units). Comparing such parameters as radius of gyration (R_g) as well as bond distance and angle distribution between interaction centers, we were able to set initial parameters for MD simulations of HPMA polymers using MARTINI model approach. Following this comparison, as well as interaction matrix and beads type assigning logic from building blocks' thermodynamical properties [8], we have defined possible variations for beads in HPMA-homopolymer and compared results of CG simulations for chosen parameter sets with AA simulation results.

6.2.2 Beads mapping

6.2.2.1. Interaction sites types and mapping principles

The MARTINI model is based on a four-to-one mapping, i.e., on average four heavy atoms are represented by a single interaction centre. Four main types of interaction sites are presented in force field: polar (P), nonpolar (N), apolar (C), and charged (Q). Each bead type has a number of subtypes which are either distinguished by a letter indicating the hydrogen-bonding capabilities (d = donor, a = acceptor, da = both, 0 = none), or by a number that corresponds to the degree of polarity (from 1, which corresponds to low polarity, to 5, high polarity). In order to provide initial mapping based on AA structures, bead centres were allocated in centres of mass for corresponding group of atoms that make up a bead.

$$R_j = \frac{1}{M}\sum_{i=1}^{N} m_i \, r_{ij} \qquad (5.1)$$

where *i* is an atom index and *j* is an index for *x,y,z* and *M* is the total molecular mass of atoms included in atomic group representing bead.

6.2.2.2. Mapping results

In general, a single HPMA monomer unit consists of 24 atoms (11 heavy atoms) and can be represented as three beads: an apolar group which represents backbone (B), a polar bead mimicking carbonyl amide group (N) and a polar bead for terminating hydroxypropyl group (C), as shown in Figure 6.1. These groups correspond to selection of available beads: B can be represented by SC1-SC3 (typical apolar beads used for alkane fragment descriptions with 3 carbon atoms) interaction center, N is interpreted as SP4-SP5 (polar groups corresponding to physical parameters, fitted from similar molecules like acetic acid, methylformamide, etc.) bead type and C can be defined by a wide range of types, both "normal" and "small" or "cycle" P1-P5 and SP1-SP5 (beads selection and mapping will be discussed further in these chapter). To obtain the best representation of HPMA polymer behavior we have provided a series of validation procedures which will be discussed later in this Chapter. An example of MARTINI mapping for a separate HPMA monomer unit is presented in Figure 6.1.

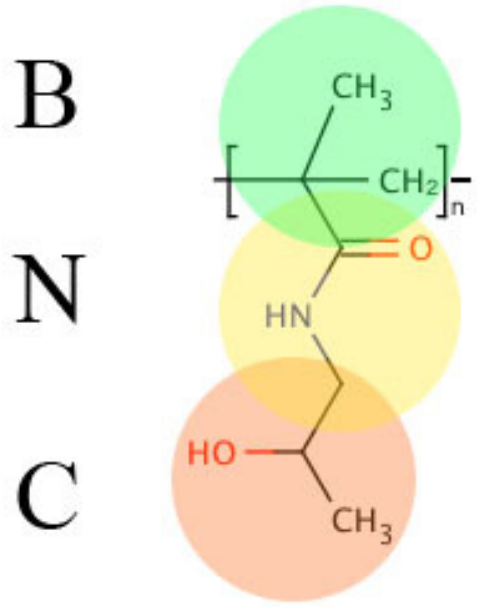


Figure 6.1. General principle for MARTINI CG mapping for a separate HPMA monomer unit

6.2.2.3. Conversion validation

Converting tools developed were validated using build in check for R_g value and center of mass position for AA molecule and CG converted structure. For all converted structures values match.

6.2.2.4. Parameters selection

In MARTINI force field, bond lengths are defined by a harmonic potential and angles are typically defined using a cosine harmonic potential. Bonded interactions can be described by the following set of potential energy functions interacting between bonded sites i, j, k with equilibrium distance d_b and angle φ_a

$$V_b = \frac{1}{2}K_b(d_{ij} - d_b)^2 \quad (2)$$

$$V_a = \frac{1}{2}K_a[\cos(\varphi_{ijk}) - \cos(\varphi_a)]^2 \quad (3)$$

where K_b is the harmonic force constant and K_a is the angle force constant.

The bonded potential V_b is used for chemically bonded sites and the angle potential V_a represents chain stiffness. For most cases [8] K_a is weaker, inducing flexibility of the molecule at the coarse-grained level. The improper dihedral angle potential as well as proper dihedrals can be included for more accurate polymer backbone behaviour, but would also had an impact on calculation time and for this level of detail was considered to be unnecessary at this stage of research. The parameters of the bonded interactions for CG HPMA model were tuned so as to reproduce the distributions of bonds and angles obtained from AA simulations. Trajectories of HPMA polymers obtained from simulations were used to provide benchmark data. We tracked the trajectories of the center of mass of each bead type and analysed the distributions of the CG bonded interactions afterwards, based on previous mapping. In this way we refined our CG parametrization of bonded interactions using a trial-and-error procedure to achieve the best agreement between the CG

distributions and the target AA distributions. HPMA-8 was chosen as suitable benchmark model for adjusting both bonded and non-bonded interactions.

MARTINI CG beads were originally parameterized to reproduce the free energy of transfer between water and oil for a set of 18 selected chemical compounds. All particle pairs *i* and *j* at distance r_{ij} interact via a Lennard-Jones (LJ) potential:

$$V_{\text{Lennard-Jones}}(r_{ij}) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \qquad (5.4)$$

The strength of the interaction, determined by the value of the well depth ε_{ij}, depends on the interacting particle types. Values range from 5.6 kJ/mol for interactions between strongly polar groups to 2.0 kJ/mol for interactions between polar and apolar groups mimicking the hydrophobic effect. The effective size of the beads can be adjusted by defining the LJ parameter σ. All normal size particles will have σ = 0.47 nm. For groups referred as small reduced LJ parameters were used: σ = 0.43 nm and ε_{ij} is scaled to 75% of the standard value.

CG simulations were performed with the GROMACS simulation package, version 4.6.5 [9]. Initial structures were generated from AA simulations, applying mapping principles described above. Scripts to generate topologies of CG structures from AA simulations were developed and written using awk programming language.

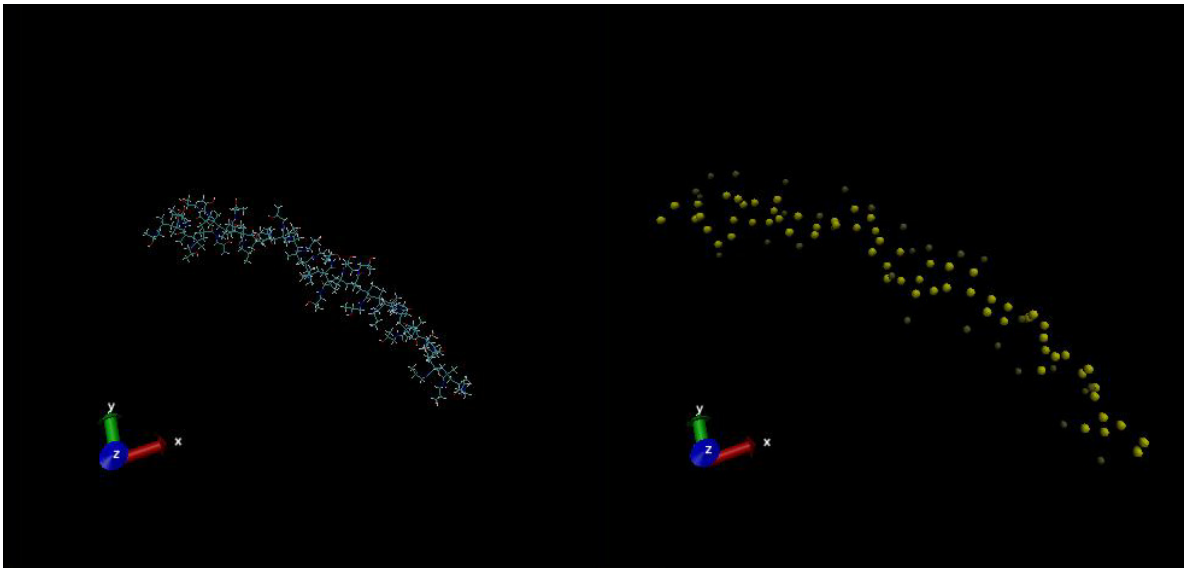


Figure 6.2. Example of beads mapping for HPMA-32 polymer chain.

The polymers were solvated in a box containing SPC [11] water beads to mimic explicit solvent, adjusting the number of solvent particles and box size to match water density of 1 g/ml. Periodic boundary conditions were applied in all directions. All CG simulations were run in similar manner as AA simulations – preliminary optimization were followed by long-run simulation. The pressure was set to 1 bar and temperature to 310 K and the NPT ensemble was used for optimization stage and NVT ensemble for MD run. A time step of 40 fs was chosen for optimization stage and 25 fs for MD simulation. Optimization stage was carried for 200000 steps and simulations were set to run 80000000 steps and continued for multiple repeats until reaching equilibrium state, which can be justified by changes in sizes, R_g and other calculated parameters' deviation from the average within the selected timescale. For the range of HPMA polymers from HPMA-8 to HPMA-64 being simulated, total simulation time varies 200 ns to 0.2 ms. For each structure 3-4 independent runs were performed with different starting trajectory.

The LJ potential falls steadily from a distance of rshift = 0.8 nm to rcut = 1.2 nm. To mimic the effect of a distance-dependent screening in chosen algorithm the electrostatic potential is shifted from rshift = 0.0 nm to rcut. The neighbor list for all non-bonded interactions was updated every 10 steps using a neighbor list cutoff equal to rcut. Simulation parameters used are as follows: temperature was kept constant at 310 K using the Berendsen temperature coupling algorithm [10] with a time constant of 1 ps. Semi-isotropic pressure coupling was applied using the Berendsen algorithm with a pressure of 1 bar independently in the plane of the membrane and perpendicular to the membrane. A time constant of 4.0 ps and a compressibility of 2×10^{-5} bar^{-1} was used. Following MARTINI force field principles [7, 8], the mass of the CG beads is unified as 72 amu for all beads, except for small beads, for which the mass is set to 45 amu. This approximation was made for a reasons of computational efficiency - using this setup allows us to run stable MD simulations with selected time step and lower CPU time.

6.3. Data analysis

R_g was used to describe the sizes of the polymer chain in equilibrated systems. This can be directly compared with previous AA MD results as the same equations were used in analysis tools [12] described in chapter 3. R_g for CG MD simulation analysis was calculated with the g_gyrate GROMACS utility [13]. Tracking of particular bonds and angles during simulation time, were carried using the program *g_bond* to calculate the distribution of the bond length in time and *g_angle* to calculate the distribution of angles [13]. Both CG and converted AA trajectories were analysed with the same set of tools. Converting scripts were developed and written using awk.

6.4. Results

6.4.1. Parameter selection

In order to establish the primary parameterization, HPMA-8 was used as a model for both bond and angle distribution. Trajectory from equilibrated AA MD simulation was used as a benchmark. Results of CG simulations for BB, NC, BN bonds as well as BBB, BBN and BNC angles were compared with corresponding data for centers of mass of selected atom groups. After fitting, some parameter sets were found to be satisfactory for the established model and remained fixed afterwards. With this "trial and error" method we were able to obtain a set of parameters for further research. Results of this fitting presented in Table 6.1. By changing the equilibrium angle or the equilibrium bond distance we were able to tune the position of peaks. At the same time, altering the force constant effectively changed distribution peaks in shape, *i.e.* the probability of obtaining average value near selected equilibrium constant. Appropriate plots for BB, NC, BN bonds distances and BBB, BBN and BNC angles are presented in Figure 6.3.

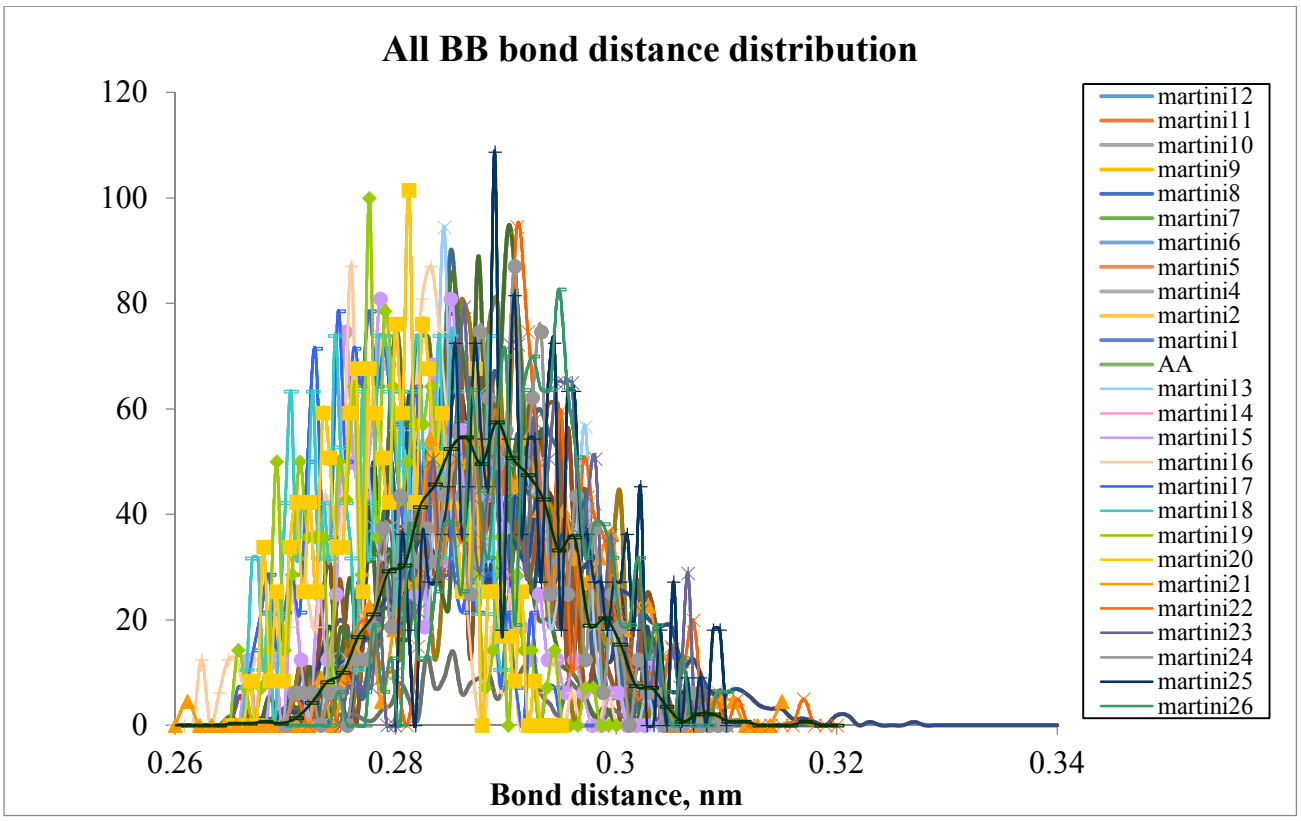
Set name	Beads selection	Bond	Equilibrium distance, nm	Force constant Kb, (kJ mol−1 nm−2)	Angle	Equilibrium value, deg	Force constant Ka, (kJ mol−1)	Distribution results				Rg, nm
								Bond	Average bond distance, nm	Angle	Average angle, deg	
AA								**BB**	0.28	**BBN**	107	0.651
								BN	0.25	**BNC**	120	
								NC	0.245	**BBB**	145.5	
martini1	SC1-SP4-P1	**BB**	0.28	3000	**BBN**	128	12	**BB**	0.3	**BBN**	111.5	0.612
		BN	0.24	4000	**BNC**	150	35	**BN**	0.28	**BNC**	123.5	
		NC	0.248	2000	**BBB**	108	250	**NC**	0.24	**BBB**	139	
martini2	SC1-SP4-P1	**BB**	0.28	3000	**BBN**	128	12	**BB**	0.29	**BBN**	112	0.589
		BN	0.24	4000	**BNC**	150	35	**BN**	0.26	**BNC**	127.5	
		NC	0.248	2000	**BBB**	108	250	**NC**	0.255	**BBB**	141.5	
martini3	SC1-SP4-P1	**BB**	0.28	3000	**BBN**	128	12	**BB**	0.285	**BBN**	112	0.613
		BN	0.24	4000	**BNC**	150	35	**BN**	0.255	**BNC**	127.5	
		NC	0.248	2000	**BBB**	108	250	**NC**	0.25	**BBB**	141.5	
martini4	SC1-SP4-P2	**BB**	0.28	3000	**BBN**	128	12	**BB**	0.29	**BBN**	110	0.624
		BN	0.24	4000	**BNC**	150	35	**BN**	0.2585	**BNC**	127.5	
		NC	0.248	2000	**BBB**	108	250	**NC**	0.25	**BBB**	143.5	
martini5	SC1-SP4-P3	**BB**	0.28	3000	**BBN**	128	12	**BB**	0.2875	**BBN**	110.5	0.621
		BN	0.24	4000	**BNC**	150	35	**BN**	0.2575	**BNC**	127	
		NC	0.248	2000	**BBB**	108	250	**NC**	0.25	**BBB**	145.5	
martini6	SC1-SP4-P4	**BB**	0.28	3000	**BBN**	128	12	**BB**	0.295	**BBN**	109	0.632
		BN	0.24	4000	**BNC**	150	35	**BN**	0.2525	**BNC**	129.5	
		NC	0.248	2000	**BBB**	108	250	**NC**	0.245	**BBB**	142	

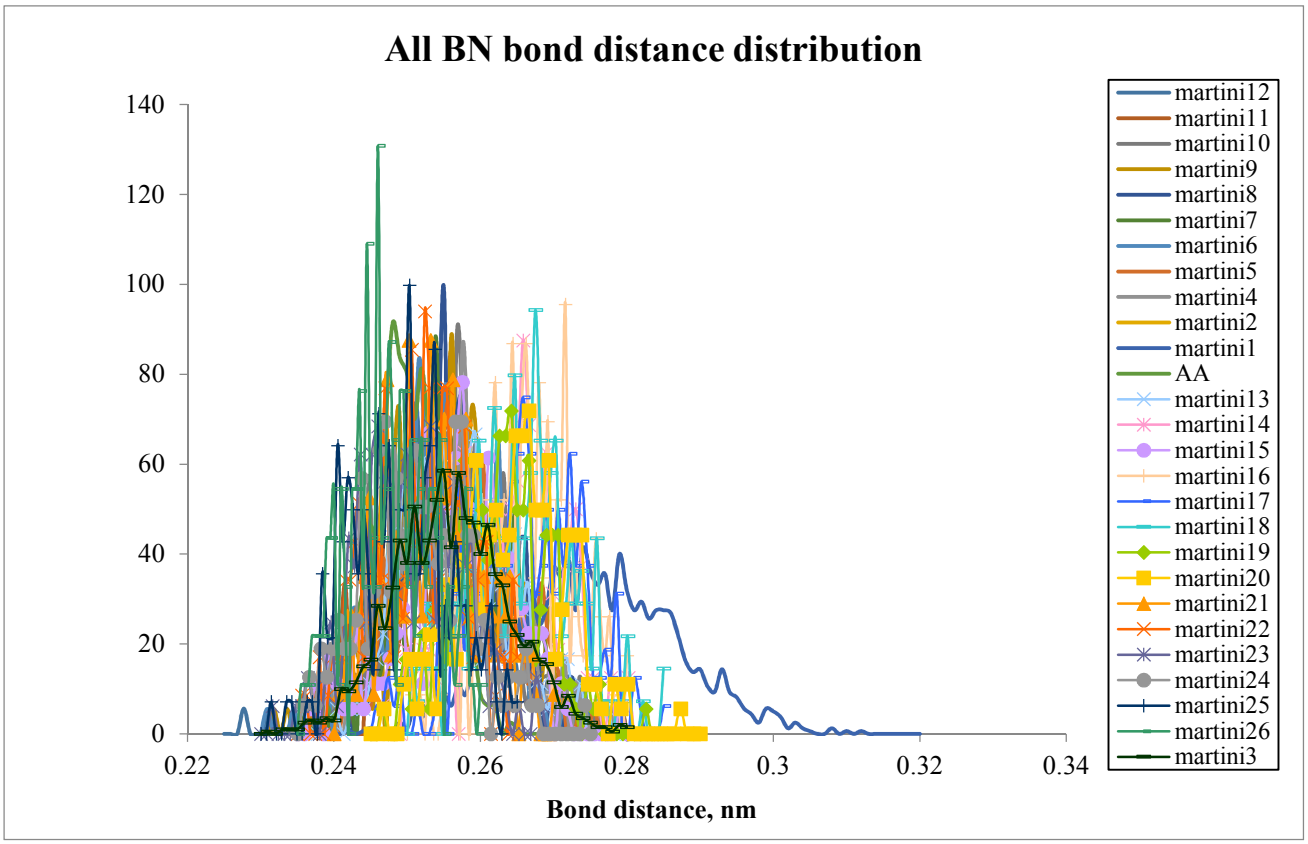
martini11	SC1-SP4-P5	**BB**	0.28	3000	**BBN**	128	12	**BB**	0.29	**BBN**	109.5	0.657
		BN	0.24	4000	**BNC**	150	35	**BN**	0.26	**BNC**	127.5	
		NC	0.248	2000	**BBB**	108	250	**NC**	0.245	**BBB**	145	
martini7	SC1-SP5-P1	**BB**	0.28	3000	**BBN**	128	12	**BB**	0.2875	**BBN**	109.5	0.631
		BN	0.24	4000	**BNC**	150	35	**BN**	0.2575	**BNC**	128	
		NC	0.248	2000	**BBB**	108	250	**NC**	0.2475	**BBB**	146.5	
martini8	SC1-SP5-P2	**BB**	0.28	3000	**BBN**	128	12	**BB**	0.29	**BBN**	110	0.633
		BN	0.24	4000	**BNC**	150	35	**BN**	0.255	**BNC**	131	
		NC	0.248	2000	**BBB**	108	250	**NC**	0.25	**BBB**	144	
martini9	SC1-SP5-P3	**BB**	0.28	3000	**BBN**	128	12	**BB**	0.29	**BBN**	106	0.636
		BN	0.24	4000	**BNC**	150	35	**BN**	0.2525	**BNC**	127.5	
		NC	0.248	2000	**BBB**	108	250	**NC**	0.2475	**BBB**	145.5	
martini10	SC1-SP5-P4	**BB**	0.28	3000	**BBN**	128	12	**BB**	0.295	**BBN**	109.5	0.619
		BN	0.24	4000	**BNC**	150	35	**BN**	0.26	**BNC**	127.5	
		NC	0.248	2000	**BBB**	108	250	**NC**	0.245	**BBB**	142.5	
martini12	SC1-SP5-P5	**BB**	0.28	3000	**BBN**	128	12	**BB**	0.2925	**BBN**	109.5	0.627
		BN	0.24	4000	**BNC**	150	35	**BN**	0.2475	**BNC**	127.5	
		NC	0.248	2000	**BBB**	108	275	**NC**	0.2475	**BBB**	145	
martini13	SC1-SP4-P2	**BB**	0.28	5000	**BBN**	128	12	**BB**	0.29	**BBN**	108	0.639
		BN	0.24	4000	**BNC**	150	35	**BN**	0.2575	**BNC**	132	
		NC	0.248	2750	**BBB**	108	25	**NC**	0.25	**BBB**	147	
martini14	SC1-SP4-P3	**BB**	0.275	5000	**BBN**	128	12	**BB**	0.28	**BBN**	127	0.625
		BN	0.245	4000	**BNC**	150	35	**BN**	0.265	**BNC**	130	
		NC	0.248	7500	**BBB**	130	25	**NC**	0.2375	**BBB**	148.5	
martini15	SC1-SP5-P1	**BB**	0.275	5000	**BBN**	128	12	**BB**	0.285	**BBN**	128.5	0.653
		BN	0.245	4000	**BNC**	150	35	**BN**	0.2575	**BNC**	129.5	
		NC	0.235	7500	**BBB**	130	25	**NC**	0.23	**BBB**	140	

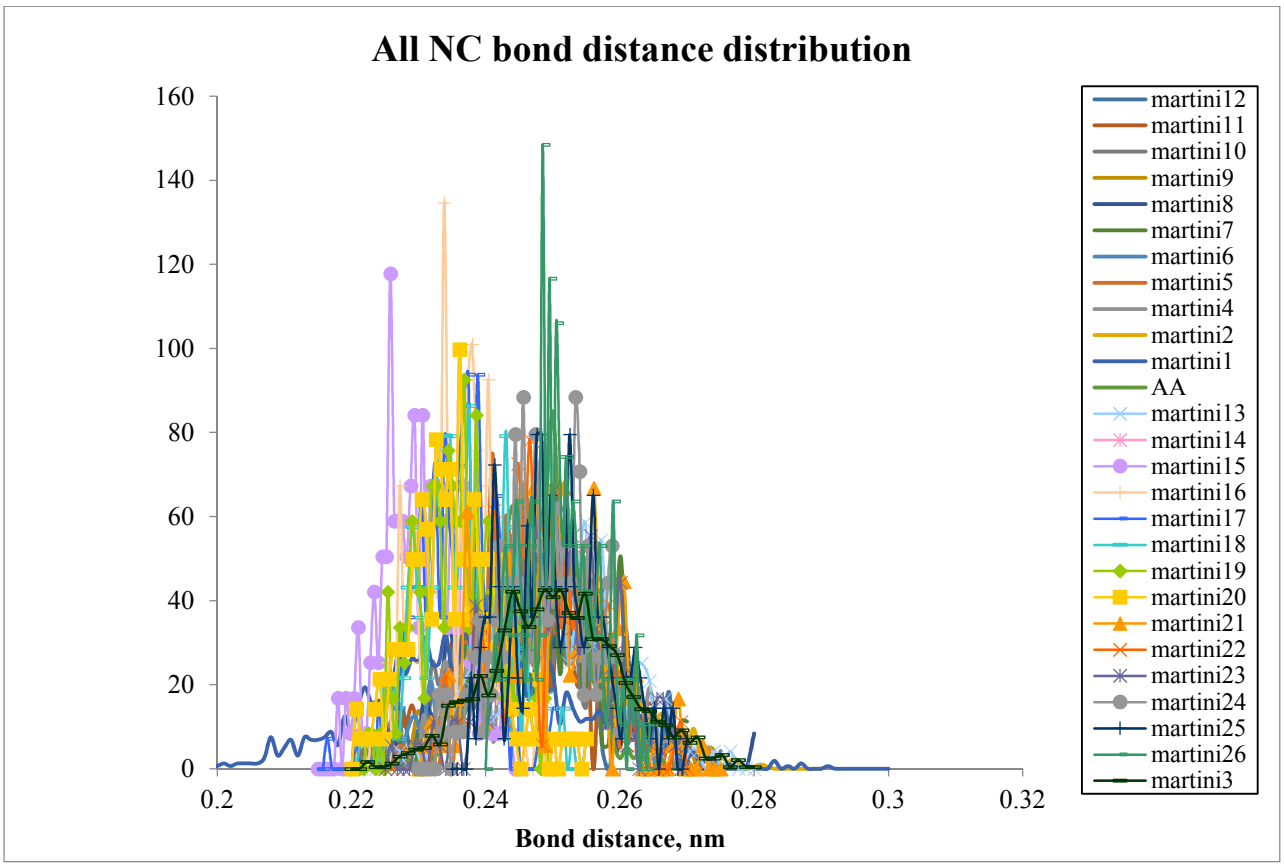
martini16	SC1-SP5-P2	**BB**	0.275	5000	**BBN**	128	12	**BB**	0.28	**BBN**	122.5	0.623
		BN	0.245	4000	**BNC**	150	35	**BN**	0.265	**BNC**	133	
		NC	0.235	7500	**BBB**	130	25	**NC**	0.235	**BBB**	146	
martini17	SC1-SP5-P3	**BB**	0.275	5000	**BBN**	128	12	**BB**	0.2825	**BBN**	126	0.629
		BN	0.245	4000	**BNC**	150	35	**BN**	0.265	**BNC**	132.5	
		NC	0.235	7500	**BBB**	130	25	**NC**	0.2325	**BBB**	150	
martini18	SC1-SP4-P3	**BB**	0.275	5000	**BBN**	128	12	**BB**	0.2775	**BBN**	130	0.633
		BN	0.245	4000	**BNC**	150	35	**BN**	0.2675	**BNC**	128.5	
		NC	0.235	7500	**BBB**	130	25	**NC**	0.2375	**BBB**	148	
martini19	SC1-SP4-P1	**BB**	0.275	5000	**BBN**	128	12	**BB**	0.2825	**BBN**	124.5	0.631
		BN	0.245	4000	**BNC**	150	35	**BN**	0.265	**BNC**	130.5	
		NC	0.235	7500	**BBB**	130	25	**NC**	0.235	**BBB**	146	
martini20	SC1-SP4-P3	**BB**	0.275	5000	**BBN**	128	12	**BB**	0.28	**BBN**	127	0.646
		BN	0.245	4000	**BNC**	150	35	**BN**	0.2675	**BNC**	130	
		NC	0.235	7500	**BBB**	130	25	**NC**	0.2375	**BBB**	148.5	
martini21	SC1-SP4-P1	**BB**	0.282	3500	**BBN**	128	12	**BB**	0.2875	**BBN**	110	0.622
		BN	0.248	3000	**BNC**	155	35	**BN**	0.255	**BNC**	130.5	
		NC	0.238	4000	**BBB**	108	280	**NC**	0.2525	**BBB**	146	
martini22	SC1-SP5-P2	**BB**	0.285	3500	**BBN**	128	12	**BB**	0.295	**BBN**	109.5	0.628
		BN	0.248	3000	**BNC**	160	35	**BN**	0.25	**BNC**	128.5	
		NC	0.236	4000	**BBB**	130	285	**NC**	0.2475	**BBB**	145.5	
martini23	SC1-SP5-P2	**BB**	0.287	3000	**BBN**	128	12	**BB**	0.2925	**BBN**	108	0.621
		BN	0.248	4000	**BNC**	155	40	**BN**	0.25	**BNC**	128	
		NC	0.234	4000	**BBB**	130	290	**NC**	0.2475	**BBB**	147.5	
martini24	SC1-SP5-P2	**BB**	0.282	3500	**BBN**	128	12	**BB**	0.29	**BBN**	106.5	0.62
		BN	0.248	4500	**BNC**	150	40	**BN**	0.255	**BNC**	126.5	
		NC	0.238	5000	**BBB**	130	295	**NC**	0.245	**BBB**	144.5	

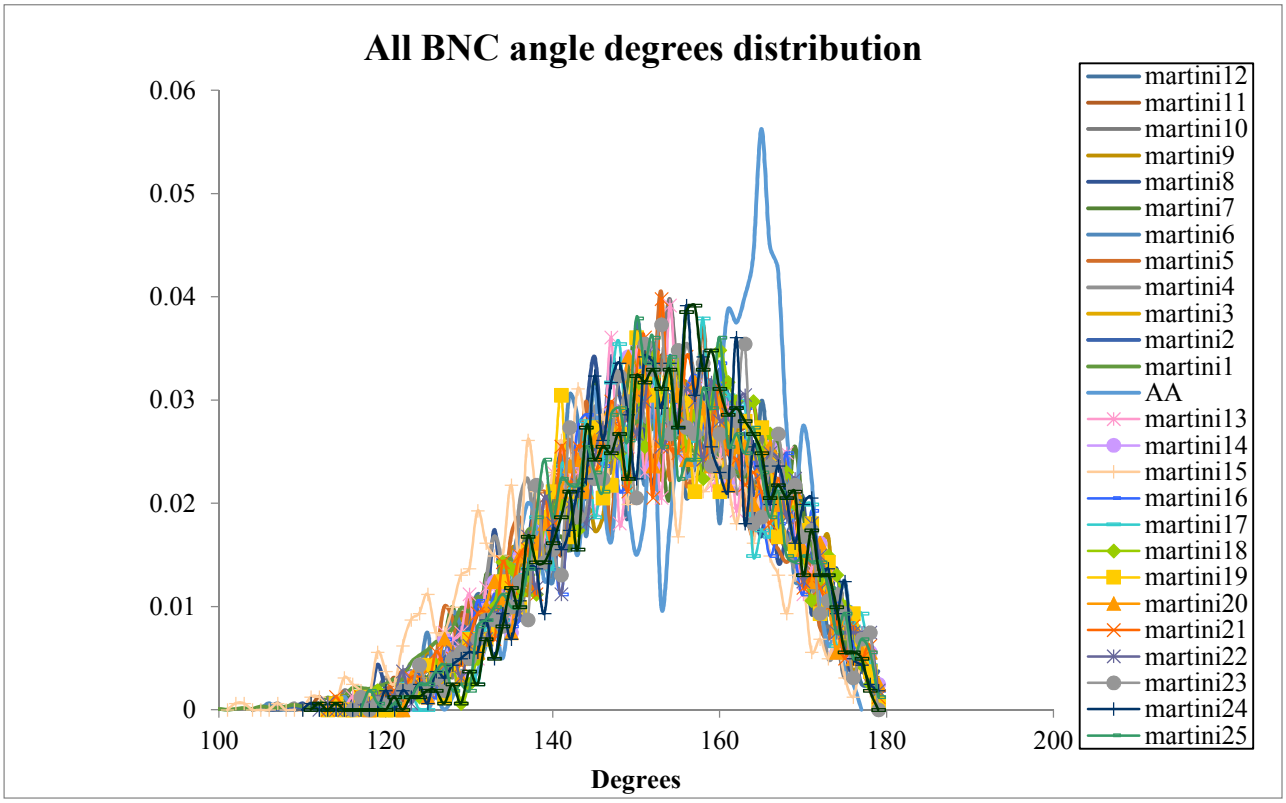
martini25	SC1-SP5-P2	**BB**	0.287	5000	**BBN**	128	12	**BB**	0.295	**BBN**	107.5	0.612
		BN	0.248	3500	**BNC**	150	45	**BN**	0.2475	**BNC**	129	
		NC	0.234	6000	**BBB**	130	300	**NC**	0.2525	**BBB**	148	
martini26	SC1-SP5-P2	**BB**	0.287	5500	**BBN**	108	12	**BB**	0.29	**BBN**	109	0.611
		BN	0.248	3500	**BNC**	160	45	**BN**	0.2475	**BNC**	132	
		NC	0.234	7500	**BBB**	128	305	**NC**	0.2525	**BBB**	145	

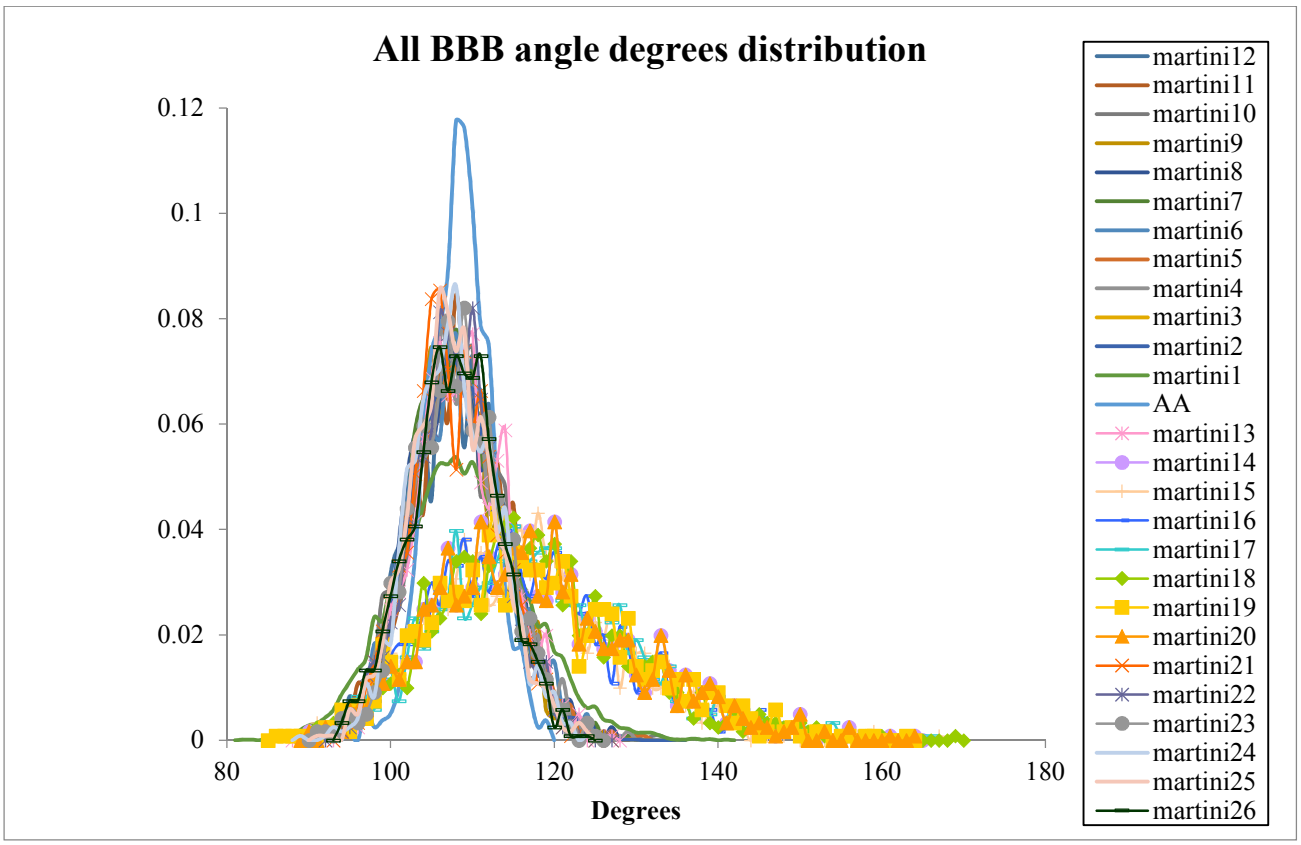
Table 6.1. Results of HPMA-8 simulation with various parameter set selection


All BB bond distance distribution
120
100
80
60
40
20
0
0.26
0.28
0.3
0.32
0.34
Bond distance, nm
martini12
martini11
martini10
martini9
martini8
martini7
martini6
martini5
martini4
martini2
martini1
AA
martini13
martini14
martini15
martini16
martini17
martini18
martini19
martini20
martini21
martini22
martini23
martini24
martini25
martini26


All BN bond distance distribution
140
120
100
80
60
40
20
0
0.22
0.24
0.26
0.28
0.3
0.32
0.34
Bond distance, nm
martini12
martini11
martini10
martini9
martini8
martini7
martini6
martini5
martini4
martini2
martini1
AA
martini13
martini14
martini15
martini16
martini17
martini18
martini19
martini20
martini21
martini22
martini23
martini24
martini25
martini26
martini3


All NC bond distance distribution
160
140
120
100
80
60
40
20
0
0.2
0.22
0.24
0.26
0.28
0.3
0.32
Bond distance, nm
martini12
martini11
martini10
martini9
martini8
martini7
martini6
martini5
martini4
martini2
martini1
AA
martini13
martini14
martini15
martini16
martini17
martini18
martini19
martini20
martini21
martini22
martini23
martini24
martini25
martini26
martini3


All BNC angle degrees distribution
0.06
0.05
0.04
0.03
0.02
0.01
0
100
120
140
160
180
200
Degrees
martini12
martini11
martini10
martini9
martini8
martini7
martini6
martini5
martini4
martini3
martini2
martini1
AA
martini13
martini14
martini15
martini16
martini17
martini18
martini19
martini20
martini21
martini22
martini23
martini24
martini25



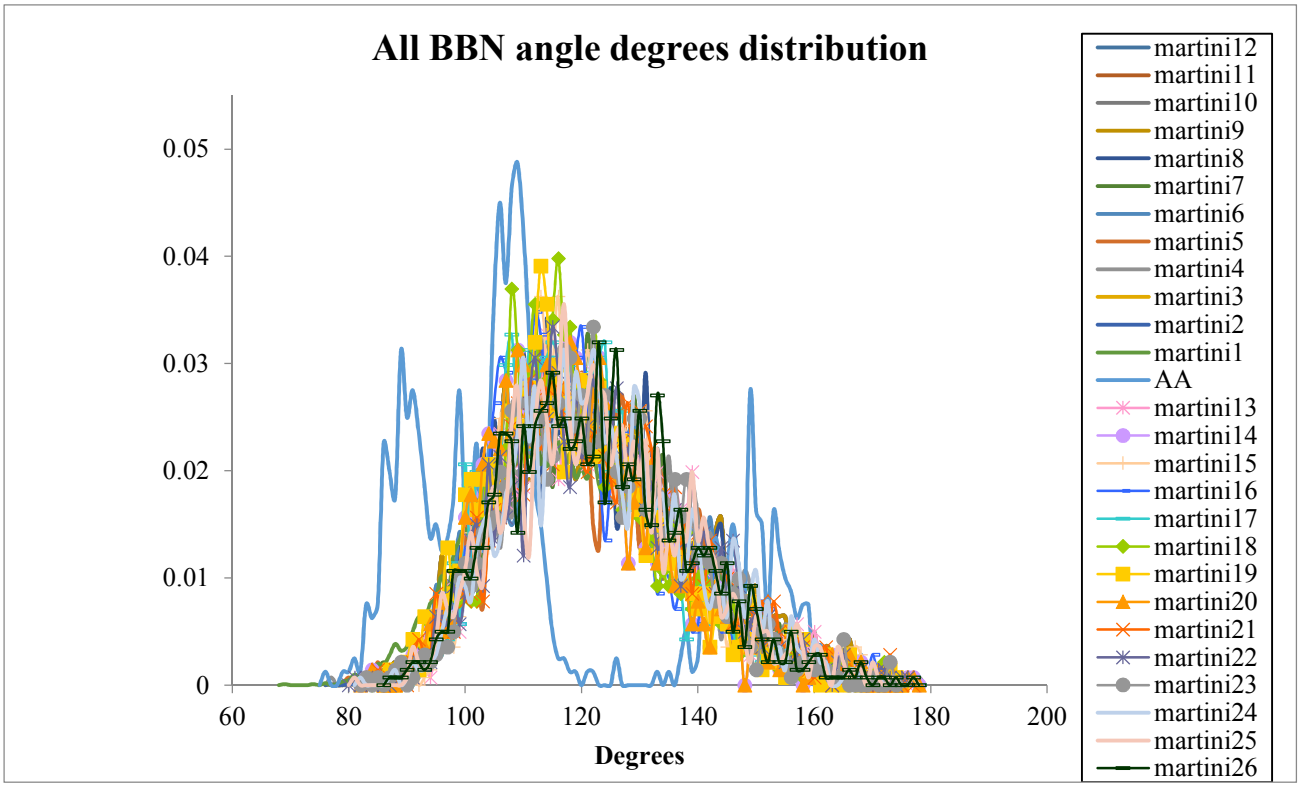


Figure 6.3. Distribution of BB, NC, BN bonds distances and BBB, BBN and BNC angles degrees obtained from AA and CG simulations of HPMA-8.

6.4.2. Radius of gyration

Results of explicit water AA and CG simulations R_g are presented in Figure 6.4

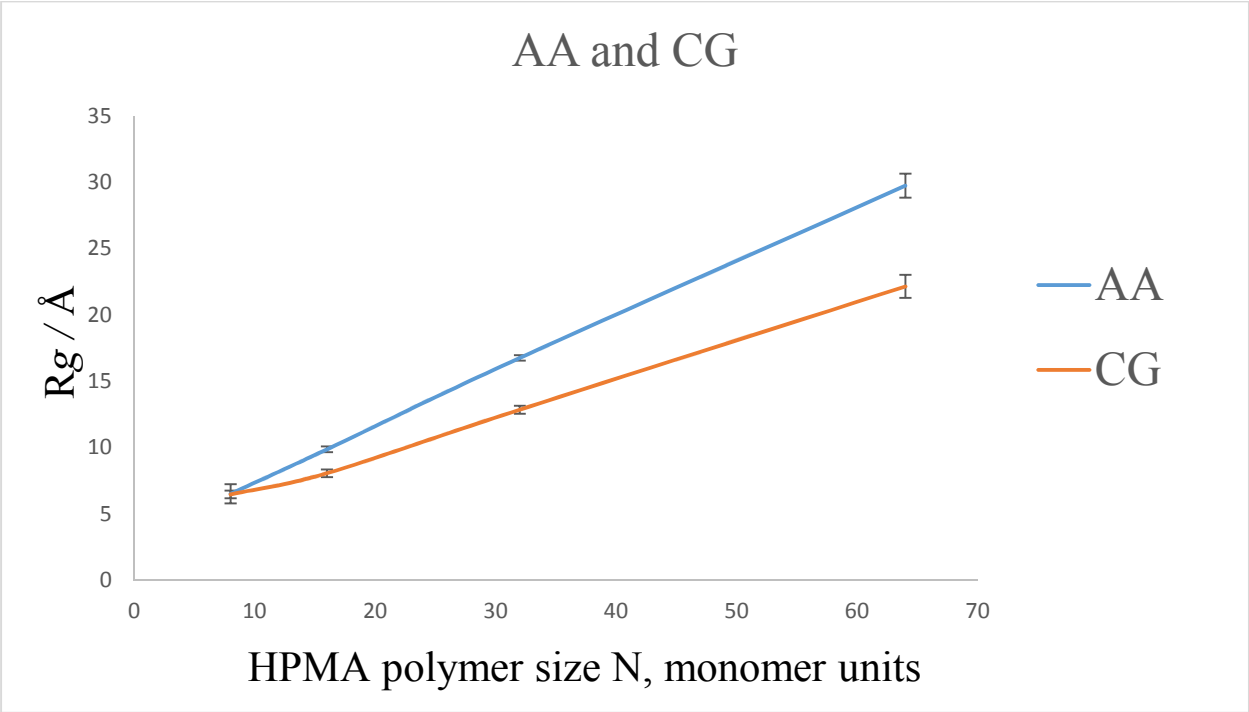


Figure 6.4: AA and CG explicit water simulations comparison

There was a minor change between the average R_g values of the different parameter sets for smaller polymers (8-64 monomer units). However for larger polymer chains changing the bead type and polarity had a massive influence on the size of the equilibrated structures. Results for HPMA-265 show how parameter variation influences R_g of polymer chains are presented in Table 6.2, and can be compared with experimentally obtained value for this structure (R_g = 75±3 Å[14]). The key parameter for polymer chain stiffness is considered to be BBB angle, which defines backbone flexibility, so other parameters were fixed.

Selected beads type	BBB parameter settings		Rg, Å
	Angle, degrees	Force constant	
SC3-SP5-P1	130	25	45.3
SC3-SP5-P1	130	50	52.1
SC3-SP5-P3	130	25	27.9
SC3-SP5-P3	130	50	41.8
SC3-SP5-SP1	130	25	56.4
SC3-SP5-SP3	130	50	66.5
SC3-SP5-SP1	130	25	69.7
SC3-SP5-SP3	130	50	81.3

Table 6.2. R_g results for various MARTINI parameter sets

It is clear that increasing force constant of backbone angle makes polymer chain less flexible, and so will prevent it from colliding and effectively increases R_g. Another way to control polymer chain stiffness is to introduce torsion rotation parameters. However, these will increase calculation time as well as affect on other parameters and require additional time for configuration and adjusting. As well as changing parameters, varying of bead types have an effect on a polymer sizes due to both changing polarity and beads effective sizes. Replacing “normal” bead with “small” as well as increasing polarity of terminal bead affected structure, increasing R_g of equilibrated molecule and meaning that the polymer structure remains more rod-like.

At the end of extensive paramaterisation, including comparison of CG with AA bond length and angle distributions as well as estimates of R_g for different parameter sets, the final set of MARTINI parameters that we propose for CG description of HPMA polymers are reported in Table 6.3.

Bond	**Equilibrium distance** / nm	**K_b /** kJ mol^{-1} nm^{-2}	**Angle**	**Equilibrium value / °**	**K_a /** kJ mol^{-1} $°^{-1}$
BB	0.275	5000	**BBN**	128	12
BN	0.245	4000	**BNC**	150	35
NC	0.235	7500	**BBB**	130	50

Table 6.3.: All bonded interaction parameters of our final CG model of HPMA polymer.

6.5. Conclusions

Due to various advantages of CG MD simulations for investigation of polymer behaviour in explicit water, it was important to establish initial mapping and parameterization scheme, based on previous AA MD simulation results as well as experimental data available. We were able to develop the necessary tools and perform fitting process for all bonded parameters and introduced preferable beads types for HPMA polymer GC simulations with MARTINI force field. In this chapter, development and validation of the established model was described, as well as results of GC MD simulations for HPMA polymers of different sizes with various parameters. These results will ultimately allow researchers to develop models for HPMA-conjugates that are applicable to full scale, long time simulations, and hence allow to obtain information on solution behavior and properties of desired polymeric drugs with faster rate and lower CPU cost.

6.6. References

[1] L.Peng, A. Ivetac, A. Chaudhari, S. Van, G. Zhao, L. Yu, S. Howell, J. A. McCammon, D. Gough Biopolymers 2010, 93, 936-951

[2] J. Perlmutter, W. Drasler, W. Xie, J. Gao, Jean-Luc Popot, J. Sachs, Langmuir 2011, 27, 10523–10537

[3] G. Rossi, P. F. J. Fuchs, J. Barnoud and L. Monticelli, J. Phys. Chem. B 2012, 116, 14353−14362

[4] R. Baron, A.H. de Vries, P.H. Huenenberger, W.F. van Gunsteren 2006 J. Phys. Chem. B 110:8464-8473.

[5] D. Reith, M. Putz, F. J. Muller-Plathe, Comput Chem 2003, 24, 1624–1636.

[6] S. J. Marrink, J. H. Rissalada, S. Yefimov, D. P. Tieleman, A. H. de Vries, J. Phys. Chem. B 2007, 111, 7812–7824.

[7] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tileman, S. J. Marrink, J Chem Theor. Comput 2008, 4, 819–834.

[8] S. J. Marrink, J. H. Rissalada, S. Yefimov, D. P. Tieleman, A. H. de Vries, J. Phys. Chem. B 2007, 111, 7812–7824.

[9] D. van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen, J. Comput. Chem. 2005, 26, 1701-1718.

[10] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, J. Chem. Phys. 1984, 81, 3684–3690.

[11] J. Hermans, H. J. C. Berendsen, W. F. van Gunsteren, J. P. M. Postma, Biopolymers 1984, 23, 1513−1518.

[12] E. Lindahl, B. Hess, D. van der Spoel, J. Mol. Mod. 7 (2001) 306-317.

[13] B. Hess, D. van der Spoel, E. Lindahl. GROMACS Manual VERSION 4.6.5, 2013.

[14] P. C. Griffiths, A. Paul, B. Apostolovic, H.-A. Klok, E. de Luca, S. M. King, R. K. Heenan, J. Controlled Release, 2011, 153, 173–179

Chapter 7: Overall conclusions and further work

7.1. Conclusions

This project was dedicated to developing and combining various methods and techniques in order to investigate novel drug-delivery systems, particularly N-(2-hydroxypropyl) methacrylamide (HPMA) conjugates. Prior to this study it was known that there is a strong link between a polymer-drug conjugate properties, such as drug type and loading, and conformation adopted in solution. Structure and morphology of the adopted conformation will define behaviour and biological activity of such systems. Despite detailed experimental work on this problem, there was a need for more in-depth study of these relationships, and ultimately to provide a way to control drug behaviour on its early development stages prior to clinical studies.

The key aim of this project was to combine experimental and computer modelling approaches in order to obtain more detailed information on studied polymer-drug delivery systems and provide reliable methodology for future study. Therefore, structural properties such as size, shape and density distribution of a range of HPMA copolymers have been investigated. The suitability of atomistic force fields has been assessed against rotational barriers and relative conformational energies obtained from *ab initio* and density functional theory (DFT) data for a monomer and dimer of HPMA.

Following this, the AMBER99 parameter set was chosen for all molecular dynamics simulation. Radius of gyration (R_g), radial distribution function (RDF), shape, and density profiles of particular atom types were calculated for a range of HPMA homopolymers sizes from 4 to 200 repeat units (2 to 35 kDa). Results are interpreted in the context of Flory's mean field approach, and compared with data obtained from small angle neutron scattering (SANS) experiments. Results for a set of solvent models and simulation conditions showed suitability of selected force field in combination with selected explicit and implicit solvent models.

This methods was then used for investigation of HPMA conjugates with drug mimics. A range of linear amines (aminohexane (C6), aminooctane (C8), aminododecane (C12)), hydroxyl and fluoro terminated linear amines as well as aromatic aminoanthracene (ANC), aminocrysene (AC) and aminoanthraquinone (ANQ), bound to the polymeric carrier via a tetrapeptide linker (glycine-phenylalanine-leucine-glycine) (GFLG) (Mw ~ 30 kDa) were selected as model objects for study of the effect of drug type and loading on HPMA copolymer conformation. SANS experimental

scattering curves were compared with theoretical curves, predicted from MD simulations. Parameters such as size and shape fitted to SANS data were compared with relevant simulated structures.

Finally, results obtained from AA simulations as well as available experimental data were used to build and adjust a coarse-grained (CG) model for HPMA polymers. Non-bonded and bonded parameters were provided from comparison against AA data, and bead mapping was developed for HPMA homopolymer chains. Applying this model allowed efficient simulation of polymer chains and provided a powerful tool for further investigations.

Summarizing the results of this project, it is important to mention its practical impact. New mixed HPMA polymer-drug conjugate systems were introduced and characterized. Combining mixtures of various drug mimics, we have developed characterisation methodology as well as investigated influence of changing ratio of hydrophilic/hydrophobic drug mimics. Thus, for systems containing aromatic/alkane substituent mixtures we were able to gradually change the overall structure and morphology of HPMA-polymer conjugates from rod-like for dominantly aromatic substituents to a sphere-like for predominantly alkane mixtures.

Building on these results, we have progressed to more complex systems of mixed polymer conjugates containing drug-mimic parts with similar parameters of hydrophobicity, but different in terms of flexibility for drug-mimic chain. In order to provide corresponding comparison we have selected systems of Adamantane (Ad)/ANC, Ad/ANQ, methyl-Adamantane (AdMe)/C12, hydroxyl-Adamantane (AdOH)/C10OH and Ad/C10 conjugate mixtures and investigated the effect of varying ratio of drug-mimics for these systems. Results of SANS data analysis revealed how conformation can be affected by the drug-part's intrinsic volume variation, and allowed us to get closer to finding answers for questions that can increase effective use of polymer-drug therapeutics.

In summary, it was demonstrated that computational simulation of polymer-drug conjugates has a practical application in obtaining details on morphological information of polymer structure, and can be combined with experimental techniques to provide a view of solution behaviour. This will allow design of new materials for drug delivery, characterization of known biological agents and investigation of structure dependence on copolymer types and drug loading in conjugate systems.

7.2 Future work

Further understanding of the processes and principles behind polymer-drug behaviour in solution is required for future progress. These can be achieved by more in-depth studies of HPMA-based systems using the tools developed herein. Force field potentials, bead mapping and other information obtained for CG model can be now applied for the next stages of research, alongside AA results, to provide a multi-scale tool for investigation of specific sites in polymer-drug molecules.

In general, preliminary studies were made for mixed systems; these can be continued with more intense study to uncover new information on polymer-drug solution behaviour. New systems can be designed and studied using developed methods in order to obtain information on contribution of various factors on solution behaviour, such as hydrophobicity and intrinsic volume of the drug part or variation of drug mimic ratio for mixed systems. This would only become possible after preliminary study and proof of concept which have been demonstrated in this project.

It is obvious that systematic approach is required for further studies. Having a wide database of polymer-drug conjugates, structure analysis combined with a way to predict properties of as yet non-existing systems is a key for intelligent experimental design and can increase efficiency of such studies. Application of these advanced approaches that combine experimental data analysis with computer modelling techniques should ultimately provide a reliable control mechanism for drug-conjugate behaviour in the body.

Appendix Software developed

Mol2 to car format convertor

```
BEGIN { printf ("!BIOSYM archive 3\nPBC=OFF\ncar file generated from g98
co-ords\n")
        printf ("!DATE Mon Jul  3 20:49:24 1995\n")

      maxatoms=1000
      for (iatom=0; iatom<=maxatoms; iatom++)
        {
          num_neigh[iatom]=0
        }
      natom=0
      found_struct = 0
      found_bond   = 0
      }

{
#$1 ~ /TRIPOS/ && $2 ~/ATOM/ { found_struct=1 }

#  if( found_struct == 1 && (NF == 9 || NF == 10) )
   if( NF == 9 || NF == 10 )
   {
              foundff=0
              natom++
              fftype[natom]="UNKNOWN"
              if ($6 ~ /C.3/)
                {
                   fftype[natom] = "CT"
                   foundff=1
                }
              else if ($6 ~ /C.2/)
                {
                   fftype[natom] = "C"
                   foundff=1
                }
              else if ($6 ~ /H/)
                {
                   fftype[natom] = "HC"
                   foundff=1
                }
              else if ($6 ~ /O.2/)
                {
                   fftype[natom] = "O"
                   foundff=1
                }
              else if ($6 ~ /O.3/)
                {
                   fftype[natom] = "OH"
                   foundff=1
                }
              else if ($6 ~ /N.am/)
                {
                   fftype[natom] = "N"
```

```
                    foundff=1
                  }

                if ($2 ~ /A/)
                     label[natom] = "H"
              else
                     label[natom] = $2

                x[natom] = $3
                y[natom] = $4
                z[natom] = $5
                q[natom] = $9

                if  (foundff == 0)
                  printf("PROBLEM : Atom ff type %s not known\n", $6)
  }


#$1 ~ /TRIPOS/ && $1 ~/BOND/ { found_struct=0
#                             found_bond=1 }
#found_bond == 1 && NF == 4  {
  if (NF == 4)
  {
#               printf("Found bond info %s %s %s %s\n", $1, $2, $3, $4)
                num_neigh[$2] = num_neigh[$2]+1
                neigh[$2,num_neigh[$2]]=$3

                num_neigh[$3] = num_neigh[$3]+1
                neigh[$3,num_neigh[$3]]=$2
   }
}

END {
          for (iatom=1; iatom <= natom; iatom++)
           {
             for (ineigh=1; ineigh<= num_neigh[iatom]; ineigh++)
               {
                 neigh_index = neigh[iatom,ineigh]

                 for (jneigh=1; jneigh<=num_neigh[neigh_index]; jneigh++)
                   {

                   if (fftype[iatom] == "HC")
                   {
                    if ( label[neigh[iatom,ineigh]] ~ "O")
                       fftype[iatom] = "HO"
                    else if ( label[neigh[iatom,ineigh]] ~ "N")
                         fftype[iatom] = "H"
                    else if ( fftype[neigh[iatom,ineigh]] ~ "CT")
                    {
#                     printf(".......%d %s is a neighbour of this
neighbour!\n",
#                                neigh[neigh_index,jneigh],
fftype[neigh[neigh_index,jneigh]]);
                       if (fftype[neigh[neigh_index,jneigh]] ~ "O" ||
fftype[neigh[neigh_index,jneigh]] ~ "N" )
                          fftype[iatom] = "H1"
```

```
                    }
                  }
                }
              }

           if (label[iatom] ~ "H1" || label[iatom] ~ "H2" || label[iatom]
~ "H3" )
              label[iatom] = substr(label[iatom],1,1)

/* Change fftype according to H tests before here */
            printf("%-5s %14.9f %14.9f %14.9f GROU 1      %3s     %2s
%6.3f\n" ,
                  label[iatom], x[iatom], y[iatom], z[iatom],
fftype[iatom], label[iatom], q[iatom])
          }
         printf ("end\nend\n")
    }
```

Force field population script for MARTINI HPMA model

```
BEGIN {
jatom=0
beadcount=1
maxbeads=0
i1=1
i2=4
nmonomer=16
#natoms is the number of CG beads times 3
#printf("natom= %d\n", natom)
printf("[BB]\n")
 for (jatom = 0; jatom < nmonomer-1; jatom++) {
printf("%d     %d\n", i1, i2)
i1=i1+3
i2=i2+3
 }
}
END {
}
```

```
BEGIN {
jatom=0
beadcount=1
maxbeads=0
i1=1
i2=4
i3=7
nmonomer=16
#natoms is the number of CG beads times 3
#printf("natom= %d\n", natom)
printf("[BBB]\n")
 for (jatom = 0; jatom < nmonomer-2; jatom++) {
printf("  %d      %d     %d\n", i1, i2, i3)
i1=i1+3
i2=i2+3
i3=i3+3
 }
}
END {
}
```

```
BEGIN {
jatom=0
beadcount=1
maxbeads=0
i1=1
i2=4
i3=5
nmonomer=16
#natoms is the number of CG beads times 3
#printf("natom= %d\n", natom)
printf("[BBN]\n")
 for (jatom = 0; jatom < nmonomer-2; jatom++) {
printf("  %d      %d     %d\n", i1, i2, i3)
i1=i1+3
```

```
i2=i2+3
i3=i3+3
 }
}
END {
}


BEGIN {
jatom=0
beadcount=1
maxbeads=0
i1=1
i2=2
nmonomer=16
#natoms is the number of CG beads times 3
#printf("natom= %d\n", natom)
printf("[BN]\n")
 for (jatom = 0; jatom < nmonomer; jatom++) {
printf("%d     %d\n", i1, i2)
i1=i1+3
i2=i2+3
 }
}
END {
}


BEGIN {
jatom=0
beadcount=1
maxbeads=0
i1=1
i2=2
i3=3
nmonomer=16
#natoms is the number of CG beads times 3
#printf("natom= %d\n", natom)
printf("[BNC]\n")
 for (jatom = 0; jatom < nmonomer-1; jatom++) {
printf("  %d      %d     %d\n", i1, i2, i3)
i1=i1+3
i2=i2+3
i3=i3+3
 }
}
END {
}


BEGIN {
jatom=0
beadcount=1
maxbeads=0
i1=2
i2=3
nmonomer=16
#natoms is the number of CG beads times 3
```

```
#printf("natom= %d\n", natom)
printf("[NC]\n")
 for (jatom = 0; jatom < nmonomer; jatom++) {
printf("%d     %d\n", i1, i2)
i1=i1+3
i2=i2+3
 }
}
END {
}
```

RDF calculation for bonds and angles

```
#!/bin/bash

/usr/local/gromacs/bin/g_bond -s traj -f traj -n index_CN.ndx -d bonds_CN
/usr/local/gromacs/bin/g_bond -s traj -f traj -n index_NC.ndx -d bonds_NC
/usr/local/gromacs/bin/g_bond -s traj -f traj -n index_BB.ndx -d bonds_BB
```

```
#!/bin/bash

/usr/local/gromacs/bin/g_angle  -f traj.xtc -type angle -n index_CNC -od
angle_CNC
/usr/local/gromacs/bin/g_angle  -f traj.xtc -type angle -n index_BBB -od
angle_BBB
/usr/local/gromacs/bin/g_angle  -f traj.xtc -type angle -n index_BBN -od
angle_BBN
```

AA to CG mapping tool

```
BEGIN{
      template="%4s  %5i %4s %3s  %4i    %8.3f%8.3f%8.3f%6.2f%6.2f    \n"
      atom_count_SC1=0
      atom_count_SP5=0
      atom_count_P1=0

      i=0
         amass=0
         xc_o_m = 0
         yc_o_m = 0
         zc_o_m = 0
         txc_o_m = 0
         tyc_o_m = 0
         tzc_o_m = 0
         xc_o_m_cg = 0
         yc_o_m_cg = 0
         zc_o_m_cg = 0
         txc_o_m_cg = 0
         tyc_o_m_cg = 0
         tzc_o_m_cg = 0
         bead = 1
         beadcount=1
       groupIdentifierCounter=0
       firstMonomersAtom=1
     maxatoms=1000
       for (iatom=0; iatom<=maxatoms; iatom++)
         {
           num_neigh[iatom]=0
              natom=0
              found_struct = 0
              found_bond   = 0
         }
#/**  init()  **/
       }

 function init() {
                  groupIdentifierCounter=0
               bmass_SC1 = 0
               bmass_SP5= 0
               bmass_P1= 0
               xb_SC1 = 0
               yb_SC1 = 0
               zb_SC1 = 0
               xb_SP5 = 0
               yb_SP5 = 0
               zb_SP5 = 0
               xb_P1 = 0
               yb_P1 = 0
               zb_P1 = 0
               t_xb_SC1 = 0
               t_yb_SC1 = 0
               t_zb_SC1 = 0
               t_xb_SP5 = 0
               t_yb_SP5 = 0
               t_zb_SP5 = 0
               t_xb_P1 = 0
```

```
              t_yb_P1 = 0
              t_zb_P1 = 0

 }



{
      if($10 == "BACKBONE") {
                 i++
      }
      if(i == 3) {
            execute()
            init()
            i=1

                    }

    if( NF == 9 || NF == 10 ) {
              foundff=0
              natom++
              fftype[natom]="UNKNOWN"
              btype[natom]="UNKNOWN"
                  if ($6 ~ /H/ || $2 ~ /A/ ) {
                   mass[natom]=1.007
                   amass=amass+mass[natom]
                  } else if ($6 ~ /C/) {
                   mass[natom]=12.011
                   amass=amass+mass[natom]
                         groupIdentifierCounter++
                  } else if ($6 ~ /N/) {
                   mass[natom]=14.007
                   amass=amass+mass[natom]
                  } else if ($6 ~ /O/) {
                   mass[natom]=15.999
                   amass=amass+mass[natom]
                  } else if ($2 ~ /A/) {
                   mass[natom]=1.007
                   amass=amass+mass[natom]
                }
                      if(groupIdentifierCounter < 4) {
                           btype[natom] = "SC1"
                      } else if(groupIdentifierCounter < 5) {
                           btype[natom] = "SP5"
                      } else {
                           btype[natom] = "P1"
                      }
                  x[natom] = $3
              y[natom] = $4
              z[natom] = $5
           }

  }
#/****start of rgyr****/
 function center_aa()
 {
   for (iatom = 0; iatom <= natom; iatom++)
```

```
          {
             xc_o_m = xc_o_m + x[iatom] * mass[iatom]
             yc_o_m = yc_o_m + y[iatom] * mass[iatom]
             zc_o_m = zc_o_m + z[iatom] * mass[iatom]
          }
txc_o_m = xc_o_m / amass
tyc_o_m = yc_o_m / amass
tzc_o_m = zc_o_m / amass

 }
  function calcrgyr_aa()
  {
   for (iatom = 0; iatom <= natom; iatom++)
          {
            dist_aa_atom = mass[iatom] * ((x[iatom] - txc_o_m) ^ 2 +
(y[iatom] - tyc_o_m) ^ 2 + (z[iatom] - tzc_o_m) ^ 2)
            dist_aa = dist_aa + dist_aa_atom
          }
   rgyr_aa_2 = dist_aa/amass
   rgyr_aa = sqrt(rgyr_aa_2)
  }
#/*****end of the rgyr****/
function execute() {
          for (iatom = firstMonomersAtom; iatom <= natom; iatom++)
          {
                   if (btype[iatom] ~ /SC1/)
                    {
                             atom_count_SC1++
                       bmass_SC1 = bmass_SC1 + mass[iatom]
                       xb_SC1 = xb_SC1 + x[iatom] * mass[iatom]
                       yb_SC1 = yb_SC1 + y[iatom] * mass[iatom]
                       zb_SC1 = zb_SC1 + z[iatom] * mass[iatom]
                    }
                   else if (btype[iatom] ~ /SP5/)
                    {
                             atom_count_SP5++
                       bmass_SP5= bmass_SP5 + mass[iatom]
                       xb_SP5 = xb_SP5 + x[iatom] * mass[iatom]
                       yb_SP5 = yb_SP5 + y[iatom] * mass[iatom]
                       zb_SP5 = zb_SP5 + z[iatom] * mass[iatom]
                    }
                   else if (btype[iatom] ~ /P1/)
                    {
                             atom_count_P1++
                       bmass_P1= bmass_P1 + mass[iatom]
                       xb_P1 = xb_P1 + x[iatom] * mass[iatom]
                       yb_P1 = yb_P1 + y[iatom] * mass[iatom]
                       zb_P1 = zb_P1 + z[iatom] * mass[iatom]
                    }

          }
 firstMonomersAtom = natom + 1
#/** printf("......%d is a mass of SC1\n", bmass_SC1) **/
#/** printf("......%d is a mass of SP5\n", bmass_SP5) **/
#/**  printf("......%d is a mass of P1\n", bmass_P1) **/
```

```
t_xb_SC1 = xb_SC1 / (bmass_SC1 * 10)
t_yb_SC1 = yb_SC1 / (bmass_SC1 * 10)
t_zb_SC1 = zb_SC1 / (bmass_SC1 * 10)

t_xb_SP5 = xb_SP5 / (bmass_SP5 * 10)
t_yb_SP5 = yb_SP5 / (bmass_SP5 * 10)
t_zb_SP5 = zb_SP5 / (bmass_SP5 * 10)

t_xb_P1 = xb_P1 / (bmass_P1 * 10)
t_yb_P1 = yb_P1 / (bmass_P1 * 10)
t_zb_P1 = zb_P1 / (bmass_P1 * 10)

             xc_o_m_cg = xc_o_m_cg + t_xb_SC1  * bmass_SC1
             yc_o_m_cg = yc_o_m_cg + t_yb_SC1  * bmass_SC1
             zc_o_m_cg = zc_o_m_cg + t_zb_SC1  * bmass_SC1

             xc_o_m_cg = xc_o_m_cg + t_xb_SP5  * bmass_SP5
             yc_o_m_cg = yc_o_m_cg + t_yb_SP5  * bmass_SP5
             zc_o_m_cg = zc_o_m_cg + t_zb_SP5  * bmass_SP5

             xc_o_m_cg = xc_o_m_cg + t_xb_P1  * bmass_P1
             yc_o_m_cg = yc_o_m_cg + t_yb_P1  * bmass_P1
             zc_o_m_cg = zc_o_m_cg + t_zb_P1  * bmass_P1

#/* Change fftype according to H tests before here */
            printf("%5i%4s%5s%5i%8.3f%8.3f%8.3f%8.4f%8.4f%8.4f\n" ,
                  bead,"HPMA", "SC1", beadcount, t_xb_SC1, t_yb_SC1,
t_zb_SC1, "0.00", "0.00", "0.00")
                        beadcount = beadcount + 1
                        cgmass=cgmass+bmass_SC1
                               atom_count_SC1=0
            printf("%5i%4s%5s%5i%8.3f%8.3f%8.3f%8.4f%8.4f%8.4f\n" ,
                  bead,"HPMA", "SP4", beadcount, t_xb_SP5, t_yb_SP5,
t_zb_SP5, "0.00", "0.00", "0.00")
                        beadcount = beadcount + 1
                        cgmass=cgmass+bmass_SP5
                               atom_count_SP5=0
            printf("%5i%4s%5s%5i%8.3f%8.3f%8.3f%8.4f%8.4f%8.4f\n" ,
                  bead,"HPMA", "P2", beadcount, t_xb_P1, t_yb_P1,
t_zb_P1, "0.00", "0.00", "0.00")
                        beadcount = beadcount + 1
                        cgmass=cgmass+bmass_P1
                               atom_count_P1=0
bead=bead+1
}

END {
        center_aa()
        calcrgyr_aa()
      execute()
        init()

txc_o_m_cg = xc_o_m_cg / cgmass
tyc_o_m_cg = yc_o_m_cg / cgmass
tzc_o_m_cg = zc_o_m_cg / cgmass
```

```
      printf ("10.00 10.00 10.00\n")
#/**      printf("Total mass from aa = %5f\n", amass) **/
#/**      printf("Total mass from cg = %5f\n", cgmass) **/
#/**      printf("Center of mass from aa = O1 ( %5f %5f %5f )\n",
txc_o_m, tyc_o_m, tzc_o_m) **/
#/**      printf("Center of mass from cg = O2 ( %5f %5f %5f )\n",
txc_o_m_cg, tyc_o_m_cg, tzc_o_m_cg) **/
#/**      printf("Rgyr from aa = %5f\n", rgyr_aa) **/

}
```

Analyse_hist in-host tool modifications

```
#include <stdio.h>
#include <stdlib.h>
#include <errno.h>
#include <math.h>
#include <limits.h>
#include "maxima.h"
#include "data.h"

/* routine to assign atomic scattering factor according to element type
label */

double atomic_bscat_list( char *element )
{

int iloop;

for (iloop = 0; iloop < NUM_ELEMENTS; iloop++)
   {
      if ( *element ==  period_table[iloop].elem[0]
        && *(element+1) == period_table[iloop].elem[1] ) return
period_table[iloop].bscat ;
   }

return -100.0;

}
```

Build ellipsoid

```
#include <stdio.h>
#include <math.h>

#include "maxima.h"
#include "ewald.h"
#include "structures.h"

void build_ellipse(atom *p_ellipse_atoms, int num_ellipse_atoms,
                   double aaa, double bbb, double ccc,
                   double *p_axis1, double *p_axis2, double *p_centre)
{
atom *p_atom;
int iii, iatom;
double xxx[3], yyy[3], centre[3];
double theta, dtheta;

   printf("\nbuilding ellipse using %d atoms\n", num_ellipse_atoms);
   printf("aaa= %10.6f bbb= %10.6f ccc=%10.6f\n",aaa,bbb,ccc);
   printf("axis 1: %10.6f  %10.6f  %10.6f \n", *p_axis1, *(p_axis1+1),
*(p_axis1+2));
   printf("axis 2: %10.6f  %10.6f  %10.6f \n", *p_axis2, *(p_axis2+1),
*(p_axis2+2));

   p_atom = p_ellipse_atoms;

   for (iatom=0; iatom < num_ellipse_atoms; iatom++)
     {
        sprintf(p_atom->group,"DRAW");
        sprintf(p_atom->group_no,"1");
        sprintf(p_atom->pot,"dr");
        sprintf(p_atom->elem,"H");
        p_atom->mass = 1.0;
        p_atom->part_chge=0.0;

        p_atom++;
     }

   xxx[0]= *p_axis1;
   xxx[1]= *(p_axis1+1);
   xxx[2]= *(p_axis1+2);

   yyy[0]= *p_axis2;
   yyy[1]= *(p_axis2+1);
   yyy[2]= *(p_axis2+2);

   centre[0] = *p_centre;
   centre[1] = *(p_centre+1);
   centre[2] = *(p_centre+2);

   printf("Have xxx %10.6f %10.6f %10.6f\n", xxx[0], xxx[1], xxx[2]);
   printf("Have yyy %10.6f %10.6f %10.6f\n", yyy[0], yyy[1], yyy[2]);
   printf("Have centre %10.6f %10.6f %10.6f\n", centre[0], centre[1],
centre[2]);
   printf("Have aaa %10.6f bbb %10.6f ccc %10.6f\n", aaa, bbb, ccc);

   p_atom = p_ellipse_atoms;
```

n

```
   theta= 0.0;
   dtheta= two_pi/NUM_ELLIPSE_DOTS;
   for (iii=0; iii< NUM_ELLIPSE_DOTS; iii++)
     {
       sprintf(p_atom->label,"%s%d","H",iii+1);

       theta += dtheta;
       p_atom->x=   xxx[0] * aaa * ccc * cos(theta)
                  + yyy[0] * bbb * ccc * sin(theta);

       p_atom->y=   xxx[1] * aaa * ccc * cos(theta)
                  + yyy[1] * bbb * ccc * sin(theta);

       p_atom->z=   xxx[2] * aaa * ccc * cos(theta)
                  + yyy[2] * bbb * ccc * sin(theta);

       p_atom->x += centre[0];
       p_atom->y += centre[1];
       p_atom->z += centre[2];

       p_atom++;
     }

return;
}
```

Build line

```
#include <stdio.h>
#include <math.h>

#include "maxima.h"
#include "ewald.h"
#include "structures.h"

void build_line(atom *p_line_atoms, int num_line_atoms,
                double aaa, double *p_line, double *p_centre)
{
atom *p_atom;
int iii, iatom;
double xxx[3], centre[3];
double dx, scale;

   printf("\nbuilding line using %d atoms\n", num_line_atoms);
   printf("aaa= %10.6f \n",aaa);
   printf("direction 1: %10.6f  %10.6f  %10.6f \n", *p_line);

   p_atom = p_line_atoms;

   for (iatom=0; iatom < num_line_atoms; iatom++)
     {
        sprintf(p_atom->group,"DRAW");
        sprintf(p_atom->group_no,"1");
        sprintf(p_atom->pot,"dr");
        sprintf(p_atom->elem,"H");
        p_atom->part_chge=0.0;

        p_atom++;
     }

   xxx[0]= *p_line;
   xxx[1]= *(p_line+1);
   xxx[2]= *(p_line+2);

   centre[0] = *p_centre;
   centre[1] = *(p_centre+1);
   centre[2] = *(p_centre+2);

   p_atom = p_line_atoms;
   dx = aaa/NUM_LINE_DOTS;
   scale=0;
   for (iii=0; iii< NUM_LINE_DOTS; iii++)
     {
       sprintf(p_atom->label,"%s%d","H",iii+1);

       p_atom->x=   xxx[0] * scale;
       p_atom->y=   xxx[1] * scale;
       p_atom->z=   xxx[2] * scale;

       p_atom->x += centre[0];
       p_atom->y += centre[1];
       p_atom->z += centre[2];

       p_atom++;
```

```
        scale += dx;
      }

return;
}
```

Moments of inertia based on mass

```
#include <stdio.h>
#include "maxima.h"
#include "structures.h"

/* protype list for this routine */
void move_molecule(atom *p_molecule, int num_atoms, double *move_vec);

void cube_roots(double *p_coeffs, double *p_roots);

/************************/
/**** DEBUG Routines ****/

void centre_of_mass(double *p_c_of_m, double *p_total_mass, atom
*p_molecule,
                    int num_atoms, int which_mol );

/**** DEBUG Routines ****/
/************************/
/*-------------------------------------------------------------------------
----*/


/* work out the moment of interia matrix for the molecule */

void moments_of_inertia(atom *p_molecule, int num_atoms, double
*p_c_of_m,
                            double *p_m_of_inertia, double *p_eigenvals )
{
  int icomp,this_atom;
  atom *p_atom;

  double atomic_mass;
  double *p_comp0, *p_comp1, *p_comp2, *p_comp3, *p_comp4, *p_comp5, r2;
  double vec[3];

  double cubic_coeffs[3];

/************************/
/**** DEBUG Variables ****/

  double c_of_m_dum[3];
  double totm_dum;

/**** DEBUG Variables ****/
/************************/
/*-------------------------------------------------------------------------
-----*/

/****** Zero moment of inertia matrix *******/

   p_comp0= p_m_of_inertia;
   p_comp1= p_m_of_inertia+1;
   p_comp2= p_m_of_inertia+2;
   p_comp3= p_m_of_inertia+3;
   p_comp4= p_m_of_inertia+4;
   p_comp5= p_m_of_inertia+5;
```

```
   *p_comp0 =0.0;
   *p_comp1 =0.0;
   *p_comp2 =0.0;
   *p_comp3 =0.0;
   *p_comp4 =0.0;
   *p_comp5 =0.0;

/****** move molecule to centre of mass *****/
/*  printf("Moving molecule to centre of mass at %10.6f %10.6f %10.6f\n",
                                                 *p_c_of_m, *(p_c_of_m+1),
*(p_c_of_m+2)); */
  vec[0]= - *p_c_of_m;
  vec[1]= - *(p_c_of_m+1);
  vec[2]= - *(p_c_of_m+2);
  move_molecule(p_molecule, num_atoms, &vec[0]);

/*****************************************/
/****** debug ****************************/

  centre_of_mass(&c_of_m_dum[0], &totm_dum, p_molecule, num_atoms, -1 );
  printf("Centre of mass for shifted molecule:  %10.6f %10.6f %10.6f\n",
                           c_of_m_dum[0], c_of_m_dum[1], c_of_m_dum[2]);

/****** debug ****************************/
/*****************************************/

/****** Calculate M_of_I matrix *************/

   p_atom= p_molecule;
   for (this_atom=0; this_atom <= num_atoms; this_atom++)
     {
       atomic_mass = p_atom->mass;
/*       printf("Atomic mass of %s (elem >>%s<< = %10.6f ", p_atom-
>label,
                                                                  p_atom-
>elem, atomic_mass);     */
/*       printf("position: %10.6f %10.6f %10.6f\n", p_atom->x, p_atom->y,
p_atom->z);            */

       r2= p_atom->x * p_atom->x + p_atom->y * p_atom->y + p_atom->z *
p_atom->z;

/*************************************************************************
******/
/*** Store as a triangular array since the moment of inertia matrix is
********/
/*** symmetric so you can use the elements :
********/
/***
********/
/***              0  1  2
********/
/***                 3  4
********/
/***                    5
********/
```

```
/*************************************************************************
******/

       (*p_comp0) += atomic_mass * (r2 - p_atom->x * p_atom->x);
/* 0 */

       (*p_comp1) -= atomic_mass * p_atom->x * p_atom->y;
/* 1 */

       (*p_comp2) -= atomic_mass * p_atom->x * p_atom->z;
/* 2 */

       (*p_comp3) += atomic_mass * (r2 - p_atom->y * p_atom->y);
/* 3 */

       (*p_comp4) -= atomic_mass * p_atom->y * p_atom->z;
/* 4 */

       (*p_comp5) += atomic_mass * (r2 - p_atom->z * p_atom->z );
/* 5 */

       p_atom++;
     }

/****** move molecule back from centre of mass *****/

   move_molecule(p_molecule, num_atoms, p_c_of_m);

/****** Work out eigen vectors of matrix, easy in this case as it is 3x3
so *******/
/****** use standard form from Handbook
*******/
/****** Assumes cubic term coefficient is 1
*******/
/****** So cubic_coeffs[2]= sq coeff, [1]= linear and [0]= constant term
*******/

/************************/
/****** DEBUG DEBUG ******/

/*  *p_comp0=  1.0; */
/*  *p_comp1=  1.0; */
/*  *p_comp2=  2.0; */
/*  *p_comp3=  1.0; */
/*  *p_comp4= -1.0; */
/*  *p_comp5=  1.0; */

/****** DEBUG DEBUG ******/
/************************/

  printf("DEBUG>> Matrix formed in moments_of_inertia.c : \n");
  printf("%10.6f %10.6f %10.6f\n", *p_comp0, *p_comp1, *p_comp2);
  printf("%10.6f %10.6f %10.6f\n", *p_comp1, *p_comp3, *p_comp4);
  printf("%10.6f %10.6f %10.6f\n", *p_comp2, *p_comp4, *p_comp5);

  cubic_coeffs[2] = -*p_comp0 - *p_comp3 - *p_comp5;
```

```
  cubic_coeffs[1] = -*p_comp4 * *p_comp4 - *p_comp1 * *p_comp1
                                         - *p_comp2 * *p_comp2;

  cubic_coeffs[1]+= *p_comp0 * *p_comp3 + *p_comp0 * *p_comp5
                                        + *p_comp3 * *p_comp5;

  cubic_coeffs[0] =      -*p_comp0 * *p_comp3 * *p_comp5
                   -2.0* *p_comp1 * *p_comp2 * *p_comp4
                      +  *p_comp0 * *p_comp4 * *p_comp4
                      +  *p_comp3 * *p_comp2 * *p_comp2
                      +  *p_comp5 * *p_comp1 * *p_comp1;

/*  printf("\nGives cubic coeffs:  %10.6f %10.6f %10.6f\n",
                 cubic_coeffs[0], cubic_coeffs[1], cubic_coeffs[2]);
*/
/**** Test roots ***/
/*   cubic_coeffs[0]=-56.0; */
/*   cubic_coeffs[1]=-22.0; */
/*   cubic_coeffs[2]=  5.0; */
/*******************/
  cube_roots(&cubic_coeffs[0], p_eigenvals);

/*  printf("cubic equation coefficients 0= %10.6f 1=%10.6f 2=%10.6f\n",
                                  cubic_coeffs[0], cubic_coeffs[1],
cubic_coeffs[2]); */

/*  printf("eigenvalues of moment of inertia tensor are : %10.6f %10.6f
%10.6f\n",
                                  *p_eigenvals, *(p_eigenvals+1),
*(p_eigenvals+2));   */

return;
}
```

Moments of inertia based on bscat

```
#include <stdio.h>
#include "maxima.h"
#include "structures.h"

/* protype list for this routine */
void move_molecule(atom *p_molecule, int num_atoms, double *move_vec);

void cube_roots(double *p_coeffs, double *p_roots);

/************************/
/**** DEBUG Routines ****/

void centre_of_bscat(double *p_c_of_b, double *p_total_mass, atom
*p_molecule,
                     int num_atoms, int which_mol );

/**** DEBUG Routines ****/
/************************/
/*-------------------------------------------------------------------------
----*/

/* work out the moment of interia matrix for the molecule based on SANs
bscat parameters */

void moments_of_inertia_bscat(atom *p_molecule, int num_atoms, double
*p_c_of_b,
                              double *p_m_of_inertia_b, double
*p_eigenvals_bscat )
{
  int icomp,this_atom;
  atom *p_atom;

  double bfact;
  double *p_comp0, *p_comp1, *p_comp2, *p_comp3, *p_comp4, *p_comp5, r2;
  double vec[3];

  double cubic_coeffs[3];

/************************/
/**** DEBUG Variables ****/

  double c_of_b_dum[3];
  double totb_dum;

/**** DEBUG Variables ****/
/************************/
/*-------------------------------------------------------------------------
-----*/

/****** Zero moment of inertia matrix *******/

   p_comp0= p_m_of_inertia_b;
   p_comp1= p_m_of_inertia_b+1;
   p_comp2= p_m_of_inertia_b+2;
   p_comp3= p_m_of_inertia_b+3;
   p_comp4= p_m_of_inertia_b+4;
```

```
   p_comp5= p_m_of_inertia_b+5;

   *p_comp0 =0.0;
   *p_comp1 =0.0;
   *p_comp2 =0.0;
   *p_comp3 =0.0;
   *p_comp4 =0.0;
   *p_comp5 =0.0;

/****** move molecule to centre of mass *****/
/*  printf("Moving molecule to centre of mass at %10.6f %10.6f %10.6f\n",
                                                  *p_c_of_b, *(p_c_of_b+1),
*(p_c_of_b+2)); */
  vec[0]= - *p_c_of_b;
  vec[1]= - *(p_c_of_b+1);
  vec[2]= - *(p_c_of_b+2);
  move_molecule(p_molecule, num_atoms, &vec[0]);

/*****************************************/
/****** debug ****************************/

  centre_of_bscat(&c_of_b_dum[0], &totb_dum, p_molecule, num_atoms, -1 );
  printf("Centre of bscat for shifted molecule:  %10.6f %10.6f %10.6f\n",
                           c_of_b_dum[0], c_of_b_dum[1], c_of_b_dum[2]);

/****** debug ****************************/
/*****************************************/

/****** Calculate M_of_I matrix *************/

   p_atom= p_molecule;
   for (this_atom=0; this_atom <= num_atoms; this_atom++)
     {
       bfact = p_atom->bscat;
/*       printf("Atomic mass of %s (elem >>%s<< = %10.6f ", p_atom-
>label,
                                                                 p_atom-
>elem, atomic_mass);     */
/*       printf("position: %10.6f %10.6f %10.6f\n", p_atom->x, p_atom->y,
p_atom->z);            */

       r2= p_atom->x * p_atom->x + p_atom->y * p_atom->y + p_atom->z *
p_atom->z;

/*********************************************************************
******/
/*** Store as a triangular array since the moment of inertia matrix is
********/
/*** symmetric so you can use the elements :
********/
/***
********/
/***             0  1  2
********/
/***                3  4
********/
```

```
/***                    5
********/
/*************************************************************************
******/

       (*p_comp0) += bfact * (r2 - p_atom->x * p_atom->x);           /* 0
*/

       (*p_comp1) -= bfact * p_atom->x * p_atom->y;                 /* 1
*/

       (*p_comp2) -= bfact * p_atom->x * p_atom->z;                 /* 2
*/

       (*p_comp3) += bfact * (r2 - p_atom->y * p_atom->y);          /* 3
*/

       (*p_comp4) -= bfact * p_atom->y * p_atom->z;                 /* 4
*/

       (*p_comp5) += bfact * (r2 - p_atom->z * p_atom->z );         /* 5
*/

       p_atom++;
     }

/****** move molecule back from centre of mass *****/

   move_molecule(p_molecule, num_atoms, p_c_of_b);

/****** Work out eigen vectors of matrix, easy in this case as it is 3x3
so *******/
/****** use standard form from Handbook
*******/
/****** Assumes cubic term coefficient is 1
*******/
/****** So cubic_coeffs[2]= sq coeff, [1]= linear and [0]= constant term
*******/

/*************************/
/****** DEBUG DEBUG ******/

/*  *p_comp0=  1.0; */
/*  *p_comp1=  1.0; */
/*  *p_comp2=  2.0; */
/*  *p_comp3=  1.0; */
/*  *p_comp4= -1.0; */
/*  *p_comp5=  1.0; */

/****** DEBUG DEBUG ******/
/*************************/

  printf("DEBUG>> Matrix formed in moments_of_inertia_bscat.c : \n");
  printf("%10.6f %10.6f %10.6f\n", *p_comp0, *p_comp1, *p_comp2);
  printf("%10.6f %10.6f %10.6f\n", *p_comp1, *p_comp3, *p_comp4);
  printf("%10.6f %10.6f %10.6f\n", *p_comp2, *p_comp4, *p_comp5);
```

```
  cubic_coeffs[2] = -*p_comp0 - *p_comp3 - *p_comp5;

  cubic_coeffs[1] = -*p_comp4 * *p_comp4 - *p_comp1 * *p_comp1
                                         - *p_comp2 * *p_comp2;

  cubic_coeffs[1]+= *p_comp0 * *p_comp3 + *p_comp0 * *p_comp5
                                        + *p_comp3 * *p_comp5;

  cubic_coeffs[0] =      -*p_comp0 * *p_comp3 * *p_comp5
                    -2.0* *p_comp1 * *p_comp2 * *p_comp4
                       +  *p_comp0 * *p_comp4 * *p_comp4
                       +  *p_comp3 * *p_comp2 * *p_comp2
                       +  *p_comp5 * *p_comp1 * *p_comp1;

  printf("\nGives cubic coeffs:  %10.6f %10.6f %10.6f\n",
                  cubic_coeffs[0], cubic_coeffs[1], cubic_coeffs[2]);
/**** Test roots ***/
/*   cubic_coeffs[0]=-56.0; */
/*   cubic_coeffs[1]=-22.0; */
/*   cubic_coeffs[2]=  5.0; */
/*******************/
  cube_roots(&cubic_coeffs[0], p_eigenvals_bscat);

/*  printf("cubic equation coefficients 0= %10.6f 1=%10.6f 2=%10.6f\n",
                                  cubic_coeffs[0], cubic_coeffs[1],
cubic_coeffs[2]); */

 printf("eigenvalues of moment of inertia tensor are : %10.6f %10.6f
%10.6f\n",

*p_eigenvals_bscat,

*(p_eigenvals_bscat+1),

*(p_eigenvals_bscat+2));

return;
}
```

y

Rg calculator

```
#include <stdio.h>
#include <math.h>
#include "maxima.h"
#include "structures.h"

/* protype list for this routine */

void min_image( double *x, double *y, double *z);

/*-------------------------------------------------------------------------
----*/

void radius_gyration(atom *p_molecule, int num_atoms, double *p_c_of_m,
                     double total_mass, double *p_rgyr )
{
  int icomp, iatom;
  atom *p_atom;

  double vec[3], dist2;

  double cubic_coeffs[3];

/****** Zero rgyr value *******/

   *p_rgyr=0.0;

   p_atom=p_molecule;
   for (iatom=0; iatom<=num_atoms; iatom++)
     {
/*** get atom to centre of mass vector ***/
       vec[0] = p_atom->x - *p_c_of_m;
       vec[1] = p_atom->y - *(p_c_of_m+1);
       vec[2] = p_atom->z - *(p_c_of_m+2);

       min_image( &vec[0], &vec[1], &vec[2]);

       dist2= vec[0]*vec[0]+vec[1]*vec[1]+vec[2]*vec[2];

       *p_rgyr += p_atom->mass * dist2;

       p_atom++;
     }

   *p_rgyr = sqrt(*p_rgyr / total_mass);

   printf("Make this molecule's radius of gyration %10.6f Angstroms\n",
*p_rgyr);

return;
}
```

[note added by supervisors]

The following publication arising from this work has been published in the peer-reviewed scientific literature:

All-atom molecular dynamics simulation of HPMA polymers.

Glib Meleshko, Jiri Kulhavy, Alison Paul, David J Willock, James A Platts

RSC Advances, 2014, 4 (14), 7003 – 7012