UNIVERSITATEA “BABEŞ-BOLYAI” CLUJ-NAPOCA

FACULTATEA DE FIZICĂ

SPECIALIZAREA BIOFIZICĂ ȘI FIZICĂ MEDICALĂ

LUCRARE DE DISERTAȚIE

Coordonator științific

Prof. Dr. Titus Adrian Beu

Absolvent

Alex-Ovidiu Mircea

2023

UNIVERSITATEA “BABEŞ-BOLYAI” CLUJ-NAPOCA

FACULTATEA DE FIZICĂ

SPECIALIZAREA BIOFIZICĂ ȘI FIZICĂ MEDICALĂ

LUCRARE DE DISERTAŢIE

PARAMETRIZATION OF MOLECULAR MECHANICS FORCE FIELD FOR GENE DELIVERY VECTORS

Coordonator științific

Prof. Dr. Titus Adrian Beu

Absolvent

Alex-Ovidiu Mircea

2023

Abstract

The aim of this work is to present a new force field for a polyethylenimine – polyethylene glycol (PEI-PEG) connector, as well as observations on the way the PEI-PEG copolymer behaves in aqueous solution, both in the present and the absence of DNA. A brief introduction to gene therapy reveals motivations for the present work. A mathematical description of the machinery behind simulations and parametrizations is presented. Next, the work presents the various parameters obtained for the connector, discussing obtained values in relation to two recently developed relevant force fields. The evolution of the solvated PEI-PEG copolymer is analyzed, followed by the discussion of DNA-copolymer interactions. The connector force field is observed to be of quality as there is no unphysical behavior detected, a sign of good interplay between the three custom parameter sets required by the copolymer. This agreement is further evidenced by the complexation pattern observed, displaying both the natural hydrophilicity of polyethylene glycol and the strength and stability of polyethylenimine – DNA electrostatic interactions.

Table of contents

INTRODUCTION

In this work, a custom force field for polyethylenimine – polyethylene glycol is presented, along with observations regarding copolymer behavior and polyplex formation. Motivated by recent relevant developments in terms of quality parametrizations, the missing elements for a unified PEI-PEG force field were identified and parametrized. Then, to observe the way the components of the unified force field interacted with one another, a series of molecular dynamics simulations were run, providing valuable information on copolymer behavior and polyplex formation. Three chapters comprise the structure of the presented work.

In the first chapter, a brief introduction on gene therapy and molecular dynamics is given. Some relevant notions and considerations are discussed, and the general simplified mechanism of gene delivery is presented. Polyethylenimine and polyethylene glycol are introduced in the context of non-viral gene delivery vectors. The importance of molecular dynamics is also briefly touched upon, then a few objectives for the work are set.

In the second chapter, the theoretical framework is presented, beginning with a mathematical description of molecular dynamics. Several important notions are introduced, including the explicit form of the potential felt by atoms within a simulation, the periodic boundary conditions, the particle mesh Ewald method and the Langevin equation, Then, the Force Field Toolkit (ffTK) parametrization workflow is introduced, which was used in the actual parametrization process, followed by a general system to be simulated and analyzed.

The third chapter includes the results of the work, starting with a step-by-step discussion of all the obtained parameters according to the ffTK procedure. Next, a set of solvated copolymer simulations is analyzed, and a few quantities of interest are presented, namely the radius of gyration and the diffusion coefficient. Another set of simulations, this time of three copolymer molecules surrounding the Drew-Dickerson dodecamer, is also presented, analyzed and discussed.

A conclusions section summarizes the findings regarding parametrization and the simulation sets. An outlook is given, outlining several directions for further research starting from the present results.

CHAPTER 1

GENE THERAPY AND MOLECULAR DYNAMICS

1.1. Introduction to gene therapy.

The goal of gene therapy may be viewed as the bringing about of therapeutic effects via intentional genetic changes brought to cells [1]. Specifically, this field is concerned with delivering certain genetic material into cells to treat diseases, be they acquired or inherited. This interesting concept started to materialize in the 1960s with a series of studies on the viability of the permanent presence of foreign DNA into mammalian cells [2], such that it would integrate with the already-present genetic material in a way which would preserve functionality. The field came into existence in the early 1970s as papers began to be published referencing gene therapy [2], although at the same time some authors expressed caution when dealing with human gene research [3].

The research in the field of gene therapy quickly accelerated in the 1990s. Between 1990 and 2015, the clinical trials which were ongoing, approved and waiting to begin or have been finished summed to more than 2300 [4]. Examples of pathologies which receive significant attention are cancers, hematological malignancies, neurological disorders, etc. By 2019, roughly half of all the clinical trials focused on cancers, while hematological malignancies made up about 20% [4].

In general, for proper delivery of genetic material, a suitable delivery vector is required. A vector acts as a mode of transport and as a protector of genetic material until it may be safely released. The general philosophy is as follows: the encapsulated DNA approaches a target cell, is absorbed inside the cell, then becomes free of the encapsulating vector and enters the nucleus, where it may interact with the already-present genetic material. One may arrive at the following questions, among others, when discussing gene therapy vectors: “What materials should the vector be made from? What are the properties of the vector? Is it compatible with physiological conditions? How well does it protect the genetic material held within? How well does it deliver the genetic material?” and so on.

Delivery vectors can be divided into two main groups: viral and non-viral. Viral vector methods make use of and manipulate virus reproduction mechanisms in order to deliver genetic material within target cells. Within clinical trials which are either ongoing or have been completed, viral vectors make up most of the clinical study delivery methods, as can be seen in table 1.1. To date, about 53% of the total number of clinical trials worldwide have used only the four most popular choices: adenoviruses, retroviruses, plasmid DNA and lentiviruses.

Of the non-viral methods, polyplexes represent a promising delivery method. A polyplex is simply a complex formed between a polymer and genetic material [5]. In general, employed polymers tend to be polycations (positively charged polymers) as they offer greater stability due to electrostatic interactions between themselves and genetic material. For some time, one significant disadvantage of polyplex methods was the inability of polyplexes to properly facilitate the escaping of genetic material into the cytoplasm before being attacked by enzymes. However, after Akinc et al explored the capacity of polyethylenimine for high transfection efficiency through the lens of the proton sponge effect [6], the interest for this delivery method increased significantly [7].

Vector type	Number of clinical trials
Adenovirus	573
Retrovirus	538
Plasmid DNA	483
Lentivirus	364
Adeno-associated virus	350
Vaccinia virus	139
Lipofection	126
Poxvirus	110
Herpes simplex virus	105
Other / Not known	897
Total	3685

Table 1.1. Most prevalent vector types worldwide in clinical studies on gene therapy. Except for plasmid DNA, all explicit entries are viral vectors. After [8].

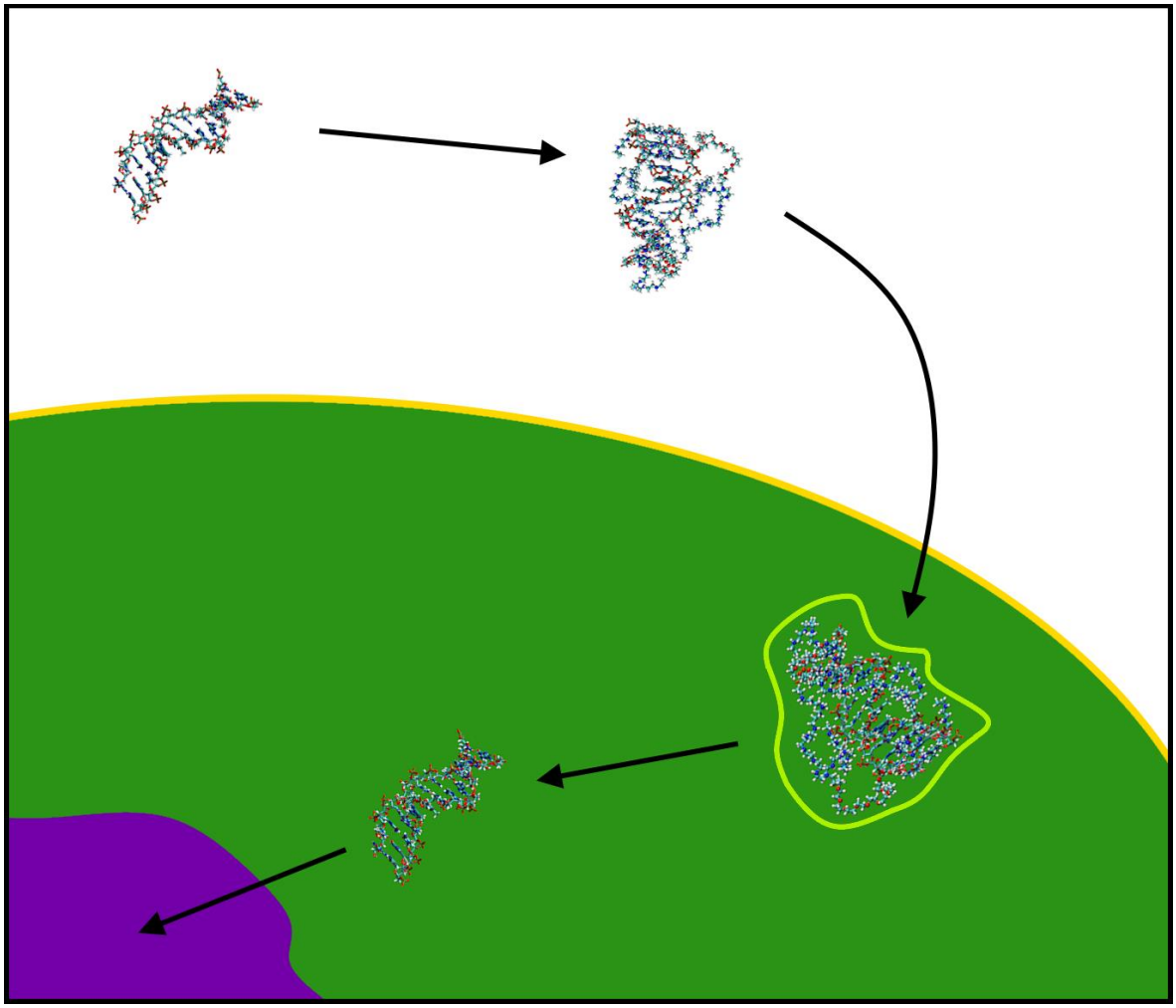

Fig. 1.1. Gene therapy vector delivering its material to a cell. After [9].

Figure 1.1 shows a simplified general schematic of the polyplex delivery mechanism method. The genetic material is approached by a polycation and forms a stable complex with it. The complex then reaches a target cell and is assimilated via an endosome. Via the proton sponge effect, an intake of charges takes place until the pressure is too much for the endosome to resist, resulting in the escape of the genetic material into the cytoplasm (green), to be delivered to the nucleus (purple).

With the proton sponge hypothesis in mind, one can quickly realize that polymers which allow for significant protonation are suitable both for complexation and for endosomal escape. It should then come as no surprise that polyethylenimine (PEI) has been studied as a non-viral delivery vector for both DNA and RNA in detail [10]. However, one major drawback is that

polyethylenimine tends to be toxic to transfected cells in virtue of its protonated amine groups, although research is still ongoing to comprehensively understand the exact mechanism [11].

To reduce the toxicity of this otherwise valuable polycation, several modifications have been proposed. A popular modification is using polyethylenimine in conjunction with polyethylene glycol (PEG), another widely studied and used material. Polyethylene glycol is useful for increasing the stability of compounds it attaches to, as well as increasing solubility in water and reducing observed toxicity [12][13]. The process of attaching polyethylene glycol to a given compound or structure is called PEGylation. The combination of high transfection efficiency, polyplex stability, solubility, and reduced cellular toxicity makes this conjugated material a promising choice for successful gene delivery.

1.2. Computational approach: molecular dynamics.

However, an important downside when analyzing complexation in traditional experiments is the very small size of both genetic material and delivery vectors. Although some investigation methods do exist which have excellent resolution, such as atomic force microscopy, the dynamic nature of complexation requires other means to better understand the process.

A complementary set of investigation tools is usually provided via computational approaches. Highly accurate results can be provided by quantum mechanical simulations of systems of interest, but these have a major drawback in the form of the sheer computational effort. To address this issue, one may instead employ the methodology of molecular dynamics and produce simulations of much larger scale molecular systems with reduced computational effort, while also retaining much of the accuracy of quantum mechanical methods. Crucial to any molecular dynamics simulation are force fields, which describe the potential produced by configurations of predetermined atom types. In essence, force fields are determined by a series of parameters, which determine the characteristics of resulting potentials.

Though not a perfect method by any means, molecular dynamics brought significant contributions in several fields. It was developed and began to be used as early as the 1950s [14]

shortly after the computer was created. In 1975, the first protein folding molecular dynamics simulation was published in Nature [15]. It has been an important tool in understanding biological phenomena at the molecular scale ever since. Molecular dynamics simulations also revealed the way molecular motors operate at an atomistic level [16]. When analyzing energetic neutron/ion collisions on solids or solid surfaces, molecular dynamics provides the standard analysis tool at the atomic level [17]. In 2020, molecular dynamics have been used on a massive scale in order to successfully investigate the spike protein of SARS-CoV-2, which plays a key role in cellular infection [18].

The method of molecular dynamics is therefore useful for a range of important applications. The focus of this work lies with biomolecular systems simulations. Recalling the polyplex delivery method illustrated previously, there are several steps which can be investigated. The initial complexation is a good starting point towards a comprehensive computational analysis of the entire delivery phenomenon. Understanding how DNA forms a complex with polyethylenimine – polyethylene glycol in the first place is fundamental for later studies on its absorption into cells and genetic material escape from the endosome, as well as how the complex itself unwraps.

These ideas provide a powerful motivation for studying DNA-PEI-PEG complex formation. To this end, there has been recent work in producing quality, custom force fields for both polymers, which are crucial to be able to simulate realistic behavior at the atomistic level. However, the junction connecting the two polymers did not receive such a treatment. The need for parametrizing a full force field for this vital component thus arises. An overall description of the mechanism used to calculate the evolution of a simulated system is given in the next chapter, along with the parametrization workflow used for finding a full force field for the connector.

In this work, the following three objectives are followed. First, based on recently produced force fields for polyethylenimine and polyethylene glycol, a full parametrization for the junction must be done. It would be most useful not to modify already known parameters to avoid redundant work. Second, using a unified force field, the PEI-PEG copolymer will be investigated via a series of simulations. The final objective is to analyze DNA-PEI-PEG complex formation using the unified copolymer force field and describe the phenomenon using relevant quantities.

CHAPTER 2

THEORY AND METHOD

2.1. Mathematical description of molecular dynamics.

The software employed for all MD simulations is Nanoscale Molecular Dynamics (NAMD) [19]. The way it works is by applying Newton's second law of motion on every atom:

$$\overline{F} = m \cdot \overline{a}, \tag{1}$$

where $\overline{F}$ is the force acting upon an atom, m is its mass and $\overline{a}$ is the acceleration imposed upon the atom. The force itself is calculated as the gradient of a potential U:

$$\overline{F} = -\overline{\nabla U}, \tag{2}$$

where $\overline{\nabla U}$ is the gradient of the potential felt by an atom. Thus, it is possible to calculate the acceleration imposed upon an atom by knowing the potential due to all other atoms present in the system. Bear in mind, however, that this means U is a function of all atom positions. To calculate the positions and velocities of all atoms at a certain time, NAMD uses the velocity Verlet algorithm [19]. It relates the positions and velocities at a time t to those at time $t + \Delta t$ The following scheme [20] illustrates the mathematical steps:

$$\text{Compute } \overline{x(t+\Delta t)} = \overline{x(t)} + \overline{v(t)} + \frac{\overline{a(t)} \cdot (\Delta t)^2}{2} \tag{3.1}$$

$$\text{Compute } \overline{a(t+\Delta t)} \text{ using the updated potential } U(t+\Delta t) \tag{3.2}$$

$$\text{Compute } \overline{v(t+\Delta t)} = \overline{v(t)} + \frac{\overline{a(t)} + \overline{a(t+\Delta t)}}{2} \tag{3.3}$$

where $\overline{x(t)}$ means the position, $\overline{v(t)}$ is the velocity and $\overline{a(t)}$ is the acceleration of an atom at time t while Δt is the time step.

Thus, the atom positions are calculated from a Taylor-like expansion around time t, the accelerations are computed from the two equations involving the forces $\overline{F}$ acting upon the atoms (eq. 1 and 2) and the velocities are obtained via another expression reminiscent of a Taylor expansion. However, the "acceleration-like" term in equation (3.3) may in fact be viewed as form of "average" acceleration over the interval $[t,\ t + \Delta t]$. The assumption of Δt being small is made. Then, due to the continuous nature of the time evolution of physical systems, it can be assumed that the average acceleration on that interval is approximately the arithmetic mean of the acceleration at the two interval endpoints. Finally, an interesting observation is that the acceleration at any point in time depends only on the positions of atoms, since the potential U itself has the same dependence.

This is more easily seen by looking at the form of the employed potential [21]:

$$U = U_{nonbonded} + U_{bonded} \tag{4.1}$$

$$U_{nonbonded} = \sum_{atoms\ i,j} \frac{q_i \cdot q_j}{\varepsilon_{eff} \cdot r_{ij}} + \sum_{atoms\ i,j} \varepsilon_{i,j} \cdot \left[\left(\frac{R_{ij}}{r_{ij}} \right)^{12} - 2 \cdot \left(\frac{R_{ij}}{r_{ij}} \right)^{6} \right] \tag{4.2}$$

$$U_{bonded} = \sum_{bonds\ i} k_{b_i} \cdot \left(b_i - b_{0_i} \right)^2 + \sum_{angles\ i} k_{\theta_i} \cdot \left(\theta_i - \theta_{0_i} \right)^2 +$$

$$\sum_{dihedrals\ i} k_{\varphi_i} \cdot (1 - \cos(n_i \cdot \varphi_i - \delta_i)) + \sum_{impropers\ i} k_{\psi_i} \cdot \left(\psi_i - \psi_{0_i} \right)^2 \tag{4.3}$$

The form of the full potential U may seem complicated and difficult to understand initially, but its additive nature allows for easy interpretation of all of its contributions. Equation (4.1) shows that the overall potential is composed of so-called nonbonded and bonded potentials. Nonbonded contributions model the long-range interactions between atoms. Equation (4.2) reveals two types of nonbonded contributions. The first kind is due to electrostatic interaction between two atoms i and j with charges q_i and q_j at a distance r_{ij} in a medium of effective dielectric constant ε_{eff}. The second is due to dispersive forces, modeled via a Lennard-Jones (LJ) potential. Again, $r_{ij}$is the distance between two atoms i and j. R_{ij} is the potential minimum and ε_{ij} is the potential well depth for atoms i and j. These last two values are calculated according to the Lorentz-Berthelot mixing rules, as showing in equations (5.1) and (5.2):

$$R_{ij} = \frac{R_i + R_j}{2} \tag{5.1}$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_i \cdot \varepsilon_j} \tag{5.2}$$

The values $(R_i,\ \varepsilon_i)$ and $\left(R_j,\ \varepsilon_j\right)$ are specific to atoms i and j respectively. Atoms of the same element may or may not have the same $(R,\ \varepsilon)$ parameters.

The bonded potential relevant to this work has 4 types of contributions, as illustrated by equation (4.3). All four have the form of a force constant k multiplied by a dependence of internal coordinates. Three of them – bonds, angles and improper dihedrals contributions – have a harmonic form while the dihedrals contribution has a periodic form. The first sum in equation (4.3) models bond stretching and squishing. The second sum is the potential energy due to angle bending. The third sum is the energy due to torsions about some bond shared by two angles. The presence of the fourth sum is due to the addition of "improper dihedrals", which serve the purpose of keeping a molecule (or a region of it) planar. In equation (4.3), all values with a subscript 0 indicate equilibrium values. b refers to bonds, θ refers to angles, φ refers to dihedrals and ψ refers to improper dihedrals. Particularly for dihedral terms, n_i means multiplicity, which describes the number of minima modeled, and can have the values 1, 2, 3, 4 or 6. Finally, $\delta_i$means phase shift, which can be 0 or 180 degrees.

The simulations are run in so-called "periodic boundary conditions" (PBC). Periodic boundary conditions provide a way to emulate an infinitely large medium. This allows for calculations of quantities such as radius of gyration, diffusion coefficient, etc. without the issue of results being influenced by the presence of walls. Figure 2.1 illustrates a simpler 2-dimensional example of PBC employed in a simulation. The region of interest is represented by the shaded box, while the other 8 are periodic copies of it and are not normally shown. The 3-dimensional equivalent would consist of 27 periodic copies of the simulated box to properly surround the simulated volume. Employing PBC allows particles to travel to other regions of the box not just by crossing it directly, but also by crossing a *boundary*. In short, when one particle would fully exit the box, a copy of it would enter from the "other side". The original particle is now part of one of the periodic copies and an image of it entered the volume of interest. In this way, effects which arise due to constraints such as walls are reduced without having the simulated system be surrounded by a vacuum.

However, a different challenge arises due to PBC being employed in simulations. The nature of the nonbonded potential would imply that, at long distances, the periodic copies would have an unwanted influence on the simulated system. For example, an atom could interact via electrostatic interactions with images of itself, which is undesirable. To avoid this issue, *the minimum image criterion* is employed. This principle states that a particle should interact with at most one image of every other particle present in the simulation, whether it is a real particle or a periodic copy. This can be easily employed via a cut-off distance, beyond which interactions with other particles are ignored. Figure 2.2 shows how a particle may interact with any other particle in its cut-off distance, even though some may be periodic images.

The bonded interactions are not affected by this new restraint. However, the nonbonded potential suffers a cut-off due to its long-range nature. The LJ contribution is not as strongly affected due to its fast convergence to 0, but still requires a smooth, switching function at the cut-off distance to avoid a sudden jump. A separate switching distance parameter is thus required to implement this correction. The bigger issue arises with the effect on calculating the electrostatic potential. Its dependence of $\frac{1}{r}$ means a much slower convergence to 0, meaning larger errors introduced due to the cut-off. To solve this, the particle mesh Ewald (PME) [24][25] method is employed.

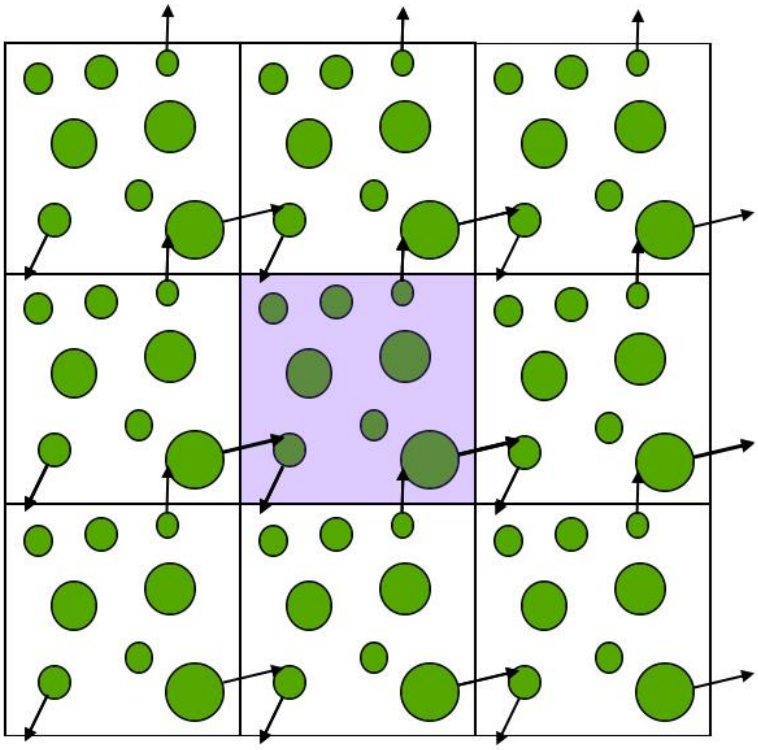

Fig. 2.1. 2D example of periodic boundary conditions. The color-shaded square is the simulation box. After Gkeka [22].

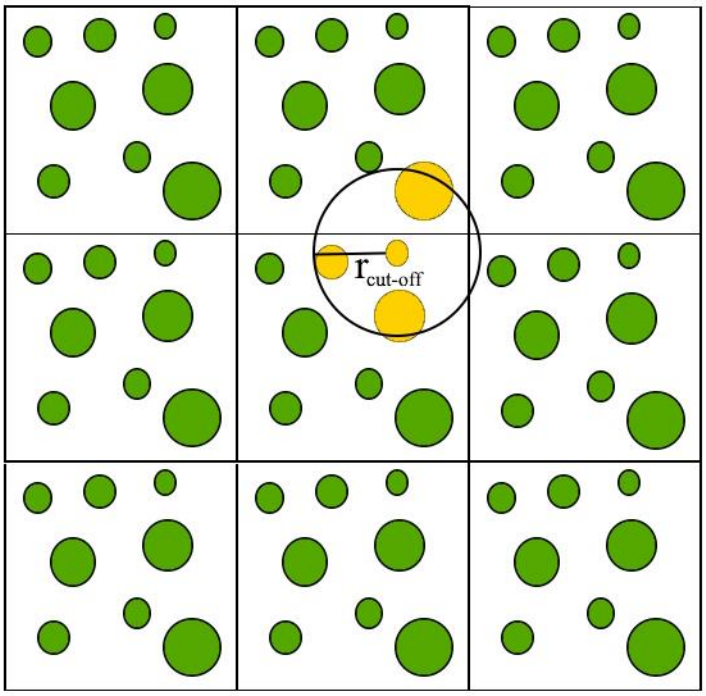

Fig. 2.2. Minimum image criterion employed via an appropriate cut-off distance, which is denoted by $r_{cut\text{-}off.}$ The central square is the simulation box. After [23].

In the PME method, the electrostatic potential is decomposed into two parts: a short-range contribution and a long-range one. The short-range part deals with electrostatic interactions an atom is subjected to within the cut-off distance. The long-range contribution is evaluated by considering its Fourier transform. The decomposition is done such that both computations converge quickly. Then, taking advantage of the Fast-Fourier Transform algorithm by evaluating the charge density on a grid, the computational complexity is overall reduced from $O(N^2)$ to $O(N \cdot \log(N))$, a significant gain in speed while also having the advantage of including initially discarded interactions.

Temperature control was done using the Langevin thermostat. An interesting consequence of using Langevin dynamics is the intentional introduction of random fluctuations, as well as a friction-like influence [26]. One form of the Langevin equation used reads:

$$\frac{d^2}{dt^2}\left(\overline{x(t)}\right) = \frac{-\overline{\nabla U}}{m} - \gamma \cdot \frac{d}{dt}\left(\overline{x(t)}\right) + \sqrt{\frac{2 \cdot \gamma \cdot k_B \cdot T}{m}} \cdot \overline{R(t)} \tag{6.1}$$

$\overline{R(t)}$ is a Gaussian process (an extension of multivariate Gaussian distributions) with mean 0, is delta-correlated to itself and is stationary (which may be understood as non-changing in time). γ is an important parameter, called the damping coefficient. It is measured in $(ps)^{-1}$, its value is a real positive number, and it effectively controls the influence of the friction-like term and of the random thermal fluctuations. Care should be taken when setting the value of this parameter. Too high of a value will overdamp the system, making it move mostly under the influence of friction and thermal fluctuations. In the limit of high γ, the effect of the deterministic potential U diminishes:

$$\frac{d^2}{dt^2}\left(\overline{x(t)}\right) = -\gamma \cdot \frac{d}{dt}\left(\overline{x(t)}\right) + \sqrt{\frac{2 \cdot \gamma \cdot k_B \cdot T}{m}} \cdot \overline{R(t)} \tag{6.2}$$

which resembles the equation of motion for Brownian motion [27], and for even larger γ:

$$\frac{d^2}{dt^2}\left(\overline{x(t)}\right) = -\gamma \cdot \frac{d}{dt}\left(\overline{x(t)}\right) \tag{6.3}$$

which is just motion under the influence of a friction-like force. Too low of a value and the two terms linked to the damping coefficient will not have a significant impact on the system's dynamics. In the limit of small γ, the Langevin equation approaches:

$$\frac{d^2}{dt^2}\left(\overline{x(t)}\right) = \frac{-\overline{\nabla U}}{m} \tag{6.4}$$

which is the same as combining equations (1) and (2). Therefore, it becomes apparent that the value of γ is in fact crucial for determining in which type of dynamics one is interested. With increasing γ, the system dynamics are governed either by the deterministic potential U, by friction and thermal fluctuations, by friction alone or a combination of the potential, friction, and thermal fluctuations. It should be noted that the velocity Verlet algorithm is affected only when it comes to acceleration computation, which is done by using the Langevin equation instead.

2.2. Force field parametrization workflow.

Some preparation is required for properly simulating systems of interest. One crucial step is ensuring there are parameter sets adequate for the investigated systems. Particularly, in this work, parameters for DNA, polyethylenimine (PEI), polyethylene glycol (PEG) and a PEI-PEG connector are required. A substantial amount of the required information did exist, albeit from various sources. The force field for the DNA strand is due to Foloppe et al [28] and MacKerrel et al [29]. The parameters related to polyethylenimine have been developed by Beu et al [30][31]. The used water model is TIP3P, with a force field due to Jorgensen et al [32]. Parameters for polyethylene glycol have been developed in a recent study by Beu et al which has not been published yet. The only parameters which are not yet readily available are specific to the connector, being crucial for properly modeling the dynamics of the PEI-PEG bonding site.

To address this issue, the VMD plugin Force Field Toolkit (ffTK) [33] was employed. Essential to this work, ffTK allows the user to fully construct and refine force fields. Note that LJ parameters are assigned by analogy between custom and predefined CHARMM atom types. Looking back at equations (4.2) and (4.3), this means that all the remaining parameters – force constants, equilibrium values, etc. – for the connector must be calculated and refined. Improper dihedral terms are not present since the connector does not have to be planar.

The following steps represent the ffTK workflow:

- Generate the initial structure to be parametrized. (7.1)
- Assign **LJ parameters** by analogy with standard CHARMM types. (7.2)
- Compute the quantum mechanical **relaxed structure** of the initial structure. (7.3)
- Calculate quantum mechanical water interaction data. (7.4)
- Obtain and refine **partial charges** on defined atom types. (7.5)
- Calculate quantum mechanical vibrational data (generate Hessian matrix). (7.6)
- Obtain and refine **bonds and angles parameters** for defined bond and angle types. (7.7)
- Calculate quantum mechanical torsion potential energy data. (7.8)
- Obtain and refine **dihedral parameters** for defined dihedral types. (7.9)

This procedure can prove to be tedious, especially when refining parameters. It can also become time-consuming, again when refining parameters, but also when computing quantum mechanical data, particularly torsion potential energy data. A crucial consequence of ffTK's workflow is that operations are sensitive to the thoroughness and carefulness with which previous

data have been obtained. Errors during any step of the workflow may propagate with highly undesirable effects. For instance, an incorrectly optimized molecular structure will result in poor water interaction data. This, in turn, will result in an abnormal or nonrepresentative partial charges distribution and so on. Lastly, there are limitations on the size of the structure which ffTK can properly parametrize. This gave rise to a fundamental problem regarding the PEI-PEG connector structure, which will be discussed in the next section.

Still, ffTK allows for comprehensively and accurately generating representative parameters without any prior molecule-specific data required. This means that an experienced user can create a useful, quality force field for any feasible molecular structure. In the context of powerful computational resources at affordable prices, ffTK is an excellent tool for systematically formulating accurate descriptions for a myriad of structures in parallel by many researchers, which accelerates the increase of simulation accuracy.

2.3. System preparation.

Bridging the chapters regarding theoretical considerations and results, the focus of this section will be system preparation, both for the parametrization procedure and for the simulated system. When considering the geometry of the PEI-PEG connector, an initial survey of commercially available PEI-g-PEG (polyethylene glycol grafted on polyethylenimine) revealed an important challenge. The copolymers' sheer complexity made the task of proper parametrization almost impossible. Since ffTK has been designed for small molecules [33], a good approach for larger systems is usually *divide et impera*. This works particularly well for polymers due to the nature of their structure. Yet, the repeating unit itself for the commercially available PEI-g-PEG would have required additional parametrizations of smaller substructures. The complexity of this task made it infeasible to parametrize the PEI-PEG connection site as was; this is the problem alluded to in the previous section. What is more, one of the motivations of this work was to use available force fields for PEI [30] [31] and PEG in conjunction with a novel connector parameter set in the MD simulations.

Overcoming this challenge was done by reducing the copolymer complexity. Inspired by the work of Wei et al [34], it was considered that the PEG side chain would attach directly to one of the nitrogen atoms of the PEI main chain. Another reason for making this simplification was skepticism over the regularity of the PEI-g-PEG found during the market survey. The repeating unit themselves presented large complexity, which would have to be upheld along the entire length of the copolymer chain. Considering the large lengths of the chains being apparently produced, it was asserted that it is not very probable for the exact structure to repeat itself without any modifications whatsoever. The simplified structure was accepted as an approximation for this work's purpose and may serve as a starting point for future studies,

In molecular dynamics, parameters are not defined between individual groups of atoms but between groups of *atom types*. Atom types represent, in simple terms, ways to classify atoms in a

molecule based on local topology, as illustrated in the following examples. Polar hydrogens are of a different type than aliphatic hydrogen atoms. Depending on how many hydrogen atoms are bound to a particular aliphatic carbon atom, four distinct hydrogen atom types may be required. A hydroxyl group oxygen atom requires a different atom type than ether oxygen atoms, and so on. One purpose of this classification is to define parameters for groups of atoms instead of individual ones, which significantly speeds up and simplifies the parametrization process. It can also reduce the computational effort during simulations, since some quantities (e.g., products of atom type partial charge pairs, LJ parameters for heterogenous atom types) can be computed once initially and then simply accessed later. Bonds between atom types define bond types, angles between two neighboring bond types define angle types, and dihedrals involving two neighboring angle types define dihedral types. Finally, a molecule's structure is subdivided into residues, which model functional groups of a few heavy atoms and any hydrogen atoms bound to them.

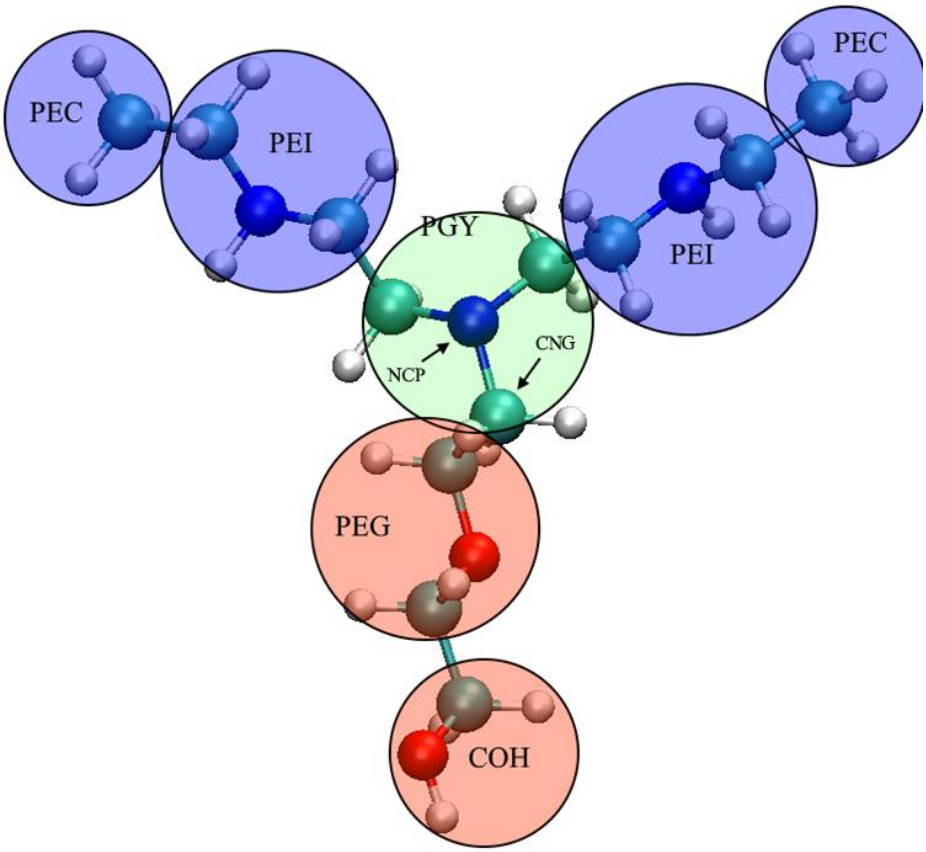

Fig. 2.3. Model geometry for the parametrization of the proposed PEI-PEG connector.

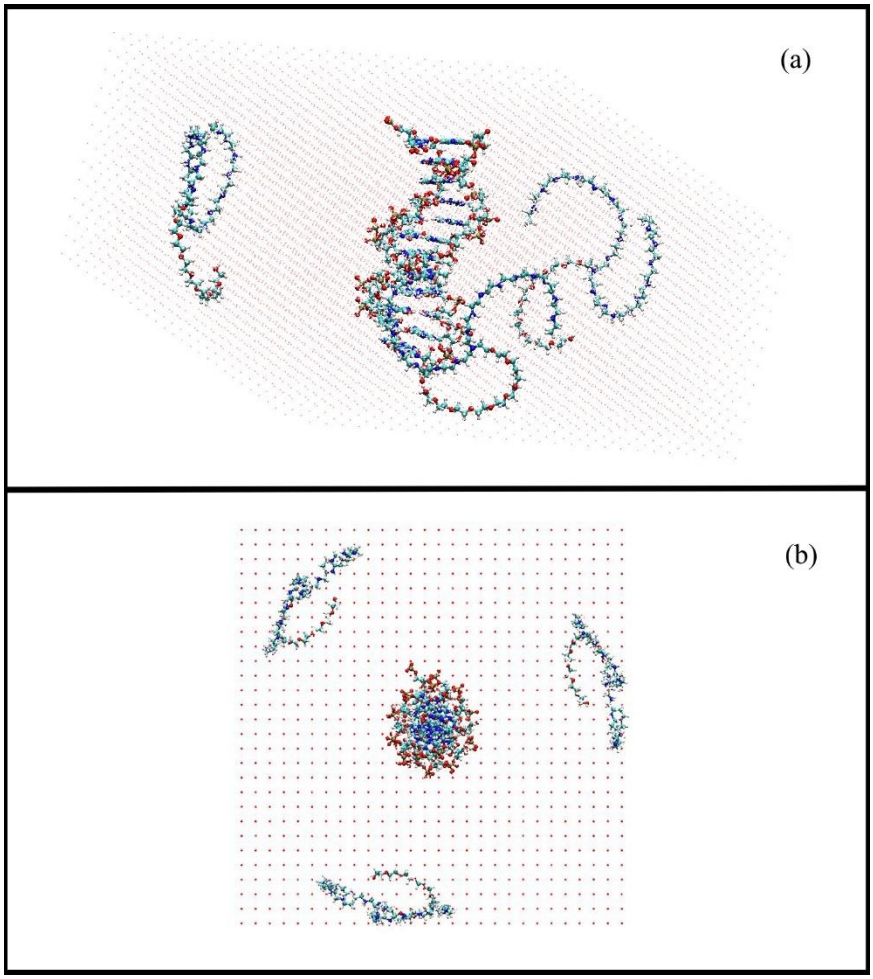

Fig. 2.4 The DNA-PEI-PEG system's initial configuration (a) from the side and (b) along the z axis. The central structure is the Drew-Dickerson dodecamer [35]. The three large molecules around the dodecamer are the PEI-PEG copolymers. Red dots represent water molecules.

Figure 2.3 shows the model system used for parametrizing the desired connector. The residue PGY is a new residue introduced in this work and was parametrized according to the ffTK workflow. All other residues are already defined in previous works. PGY contains 4 atom types. NCP is a new atom type introduced in this work. It is the nitrogen atom anchoring a PEG residue on the polyethylenimine chain. The other new atom type is CNG, a carbon atom type providing the direct link between NCP and polyethylene glycol atoms. The two carbon atoms connecting NCP to two PEI residues are of type CH2 [31]. The six hydrogen atoms are of type HC2 [31].

This choice for connector architecture kept in line with the motivation of using previously created force fields and use analogy in defining new structures. Beu et al defined a residue named PEY [30], which served as the branching point for polyethylenimine chains. It is mandatory to have the illustrated extra structures attached to PGY: the residue represents the site where polyethylene glycol attaches itself to polyethylenimine. Only the new bond, angle, and dihedral types within PGY, and the partial charges on NCP and CNG are parametrized. All other

parameters are set to the values in the already produced force fields for polyethylenimine [30][31] and polyethylene glycol and kept constant. This forces the connector force field to adapt to the pre-existing parameters.

Preparing the simulated system was less challenging. Three large PEI-PEG copolymer molecules were placed around a Drew-Dickerson dodecamer [35]. Each individual copolymer was composed of two PEI arms and a PEG arm. Each arm stretched for 9 repeating units, ended by a methyl group for PEI arms and by a (CH_2)-OH for the PEG arms. The PEI arms are composed of two residues, the five neutral polyethylenimine PEI residue and the four protonated polyethylenimine PEP residue alternating along the arm length. The system was solvated using a script containing Python routines developed by T. A. Beu [36]. When checking the net overall charge, it was found that no additional ions were needed for electrical neutralization. Figure 2.4 (a) and (b) shows the solvated DNA-PEI-PEG initial configuration.

As stated previously, NAMD was used to run all the simulations in this work. For visualization of analyzed systems, for extraction of data, and for processing it, Visual Molecular Dynamics (VMD) [37], a series of built-in plugins, custom TCL [38] and custom Python [39] scripts were employed. For the creation of graphs, the software SciDAVis [40] was used. The software Gaussian09 [41], a powerful software for quantum mechanics computation, was requested by ffTK to obtain all the reference quantum mechanical data needed for proper parametrization. This finishes system preparation and the theoretical chapter of this work.

CHAPTER 3

RESULTS AND DISCUSSION

3.1. Parametrization of CHARMM force field for linear polyethylenimine – linear polyethylene glycol connector.

As anticipated, a custom force field for the PGY residue in the given form must be created before any relevant MD simulations can be done. Additionally, the structure must be parametrized without bringing any modifications to pre-existing employed force fields. While this results in a reduction of complexity for the task at hand, it also brings a constraint. The new parameters are pushed towards values which accommodate the reference quantum mechanical data *and* the already-defined force fields. NCP plays a similar role in PGY as NC3 does in PEY. It should then be no surprise that it was identified with standard CHARMM atom type NG301 in accordance with local topology. Similarly, CNG is topologically the same as the CHARMM CG321 atom type. The Lennard-Jones parameters (R, ε) were then immediately assigned. Afterwards, at the MP2/6-31G(d) level of theory, a simple geometry optimization of the model configuration was done to relax it and obtain the optimized geometry of the connector.

3.1.1. Partial charges optimization.

Differences in the force fields begin with partial charges calculation. As per ffTK procedure, an initial series of QM calculations done in Gaussian09 were required to gather water interaction target data. ffTK provides an automatic interaction site detection feature which classified oxygen and nitrogen atoms as acceptors and hydrogen atoms as donors. Afterwards, it generated a series of Gaussian09 input files, each regarding a geometry optimization of one water

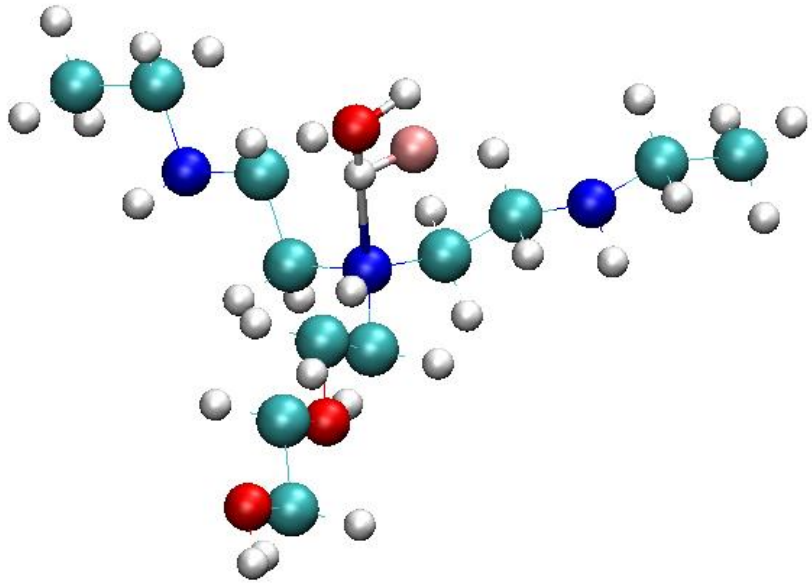

Fig. 3.1. A water molecule placed such that it interacts primarily with the connector nitrogen atom NCP. The pink particle is an auxiliary atom used for placing the water molecule and plays no role in the calculations.

molecule and the copolymer structure. One such input was written for each interaction site found by ffTK, exemplified in figure 3.1. Additionally, a water molecule and two copolymer energy calculation input files were generated. All inputs were, by default, ran at the HF/6-31G(d) level of theory, except for one MP2/6-31G(d) copolymer energy computation.

The resulting data was used as input for partial charge optimization. The only atoms which were involved at this step were the three carbon atoms and the nitrogen atom in PGY. Everything else was excluded in virtue of either having already been parametrized or, in the case of PGY hydrogen atoms, receiving a charge of +0.09e by CHARMM standard. Two conditions were imposed. First, the entire residue should have a net charge of zero as it does not represent an ionic structure. Second, the CH2 atoms within PGY should have the same charge in virtue of local symmetry.

Table 3.1 shows the calculated partial charges on PGY atoms in comparison with the ones on PEY [30]. Two tactics were employed in charge optimization: using the PEY distribution as a starting point and starting with all partial charges set to zero. This was done to check the behavior

of the optimization process, particularly to observe any potential differences between the two starting points. Several individual runs were done for both strategies, spanning a total of around two hundred optimizations. The partial charges rapidly converged towards the presented set. No difference was found between the two strategies or across runs in terms of the final partial charge distribution, indicating a single, strongly attracting minimum.

Though the values themselves do significantly differ, the general behavior remains the same for both PGY and PEY. Carbon atoms are weakly charged, although PGY carbons tend to be more negatively charged than their branched polyethylenimine counterparts. Despite a difference of almost 38% in final value, NCP still retains a strong negative charge, as expected when considering element electronegativity. No further adjustments were required, and the set was accepted as optimized.

PGY		PEY [30]	
Atom type	Partial charge (e)	Analogous atom type	Partial charge (e)
HC2	+0.09	HC2	+0.09
CH2	+0.02	CH2	+0.085
CNG	-0.086	CH2	+0.085
NCP	-0.494	NC3	-0.795

Table 3.1. Optimized PGY partial charges in comparison with PEY charge distribution.

3.1.2. Bonds and angles optimization.

The next step involves computing and optimizing parameters describing bond and angle types. To this end, Gaussian09 was employed to do a frequency calculation at the MP2/6-31G(d) for the model structure. This was done in order to construct the Hessian matrix, useful for describing the quantum mechanical potential energy surface (PES) with respect to internal coordinates. The generated PES is then used as target data for parametrization of bond and angle types within ffTK. Starting with a set of initial values, a molecular mechanical PES may be computed, then the difference between the QM-PES and MM-PES is observed. Variations of the initial set are done in order to search for a better agreement between the two potential energy

landscapes. The set implying the best agreement is kept and used for new initial values. The process is repeated until convergence to a stable set is achieved.

Again, as during the previous step, everything that was already parametrized was excluded from the optimization procedure. This left four new bond types and ten new angle types, each with its own force constant and equilibrium value. Two strategies were employed: starting from an initial guess (courtesy of ffTK's built-in function) and starting from the equivalent PEY parameters. A few individual runs were done for both strategies, and each run was continued until convergence to a set was reached. Both strategies resulted in very close final values, but the guess strategy took significantly longer to converge. This was somewhat expected given the structural similarity of PGY and PEY. Table 3.2 shows the set of values upon which the process converged.

The PEY force field provides once more a natural comparison. In terms of bond types, the observed relative differences are not greater than 6.5% in the case of force constants, while equilibrium bond lengths differ by less than 1.2%. Bond types NCP CH2 and NCP CNG have somewhat larger force constants and are close in size while the other two present slightly lower values than their polyethylenimine equivalents. This indicates that the developed connector favors the main chain's structural integrity to some extent, and at the same time, it slightly disfavors PEG binding onto the PEI chain. The equilibrium bond length differences are very small, on the order of 10^{-2} Å. This was again somewhat expected given the structural similarities of the two connectors. Still, it is worth noting that the presence of a bound PEG residue causes a distortion in PGY's geometry in the form of an elongated NCP CNG bond compared to NCP CH2, a detail revealed via using the new CNG atom type instead of another CH2 type.

Regarding angle types, there is a preference for medium to low strength force constants, which indicates bending flexibility, and for average to relatively sharp equilibrium angles, between 105 and 115 degrees. The stiffest angle is CNG CPG OPG; compared to typical values, its force constant is quite large while its equilibrium angle is relatively small. Together with the medium-sized k_θ and θ_0 of NCP CNG CPG, this indicates a tendency of the connector to not allow much bending of the bound polyethylene glycol along the CNG CPG bond. In contrast, the angle type CH2 NCP CNG allows for greater flexibility in virtue of having the smallest force constant of all used angle types, while also having the largest equilibrium value. If one also considers the values for CH2 NCP CH2 and CH2 CH2 NCP, two important observations arise. PGY allows the main PEI chain to bend flexibly at the polyethylene glycol binding site. It also allows the PEG

chain to effectively "lean" towards the main chain while trying to keep its first residue from "leaning" towards the binding site itself. This behavior differs from PEY, which is significantly more resistant to bending in virtue of its larger force constants. To give a few examples, for the angle type CH2 NCP CNG, the force constant shows a drop of 54% compared to CH2 NC3 CH2. Similarly, it drops by 47% for CH2 NCP CH2 and by 33% for CH2 CH2 NCP in comparison with their equivalent PEY types.

Summing up, substituting a polyethylenimine chain at a branching point for polyethylene glycol results in four notable effects:

1) A distortion in the geometry of the connector,
2) Significantly greater flexibility of the main chain at the binding site,
3) A local preference for polyethylene glycol to bend onto polyethylenimine,
4) A local tendency to block polyethylene glycol from bending around the connector.

Bond/ angle type definition	k_b (kcal/mol/Å^2)/ k_θ (kcal/mol/rad^2)	b_0 (Å)/ θ_0 (degrees)
NCP CH2	335.412	1.423
NCP CNG	338.548	1.430
CNG CPG	290.779	1.516
CNG HC2	333.148	1.098
CH2 CH2 NCP	45.740	109.476
CH2 NCP CH2	41.880	122.934
CH2 NCP CNG	36.175	124.114
HC2 CH2 NCP	53.455	113.071
NCP CNG HC2	47.942	111.759
NCP CNG CPG	57.804	109.129
CNG CPG HC2	49.921	112.102
CNG CPG OPG	72.504	105.433
HC2 CNG HC2	49.771	106.932
HC2 CNG CPG	47.359	107.471

Table 3.2. Optimized values for the new bond and angle types. CPG and OPG are aliphatic carbon and ether oxygen atom types respectively, from the unpublished PEG force field.

3.1.3. Dihedral angles optimization.

The final step in the parametrization procedure is finding the set of dihedral force constants, multiplicities and phase shifts for the new dihedral angles. Due to the periodic nature of their potential energy, a series of relevant backbone dihedrals had to be individually scanned. A scan is simply a series of geometry optimizations which lock a certain dihedral angle in a fixed configuration and allowing the rest of the molecule to relax. This was done in Gaussian09 at the MP2/6-31G(d) level of theory, going from -90 to 90 degrees in steps of 15 degrees for each input file generated by ffTK.

The result was compiled by ffTK into a QM PES of the dihedral angles. This can be reproduced to some extent by a MM PES generated by a set of dihedral angle types. The better the PES agreement, the better the set – so long as its behavior is not unphysical. A dihedral type may receive several multiplicities, but it is generally better to have as few multiplicities as possible to reduce the computational effort during simulations. Deciding the number of employed multiplicities for a given of dihedral is ultimately a matter of know-how as there is no clearcut way, other than observing unphysical behavior, to decide whether a small increase in accuracy is worth the increased computational effort.

Table 3.3 illustrates the final set of dihedral parameters for the new dihedral types. Again, two strategies were followed: guessing initial sets and borrowing equivalent PEY parameters. Effectively, the best agreement between target data and optimized parameters was reached with the second strategy. Better MM PES and QM PES agreement was reached with the first strategy, but the corresponding sets were observed to produce unphysical behavior, which is highly undesirable, and were thus rejected.

The presented set has the advantage of one multiplicity per dihedral, reducing the computational effort while remaining in fair agreement with the target data, as may be seen in figure 3.2. Just like in the previous steps, only new types may be changed; everything else must be excluded from the process. This imposed condition is considered the cause of the few disagreements between the MM and QM energies, as several degrees of freedom are not available in the parametrization of PGY. Nevertheless, the parametrization procedure concludes with this set of dihedral parameters being accepted, allowing copolymer simulations to be run.

Dihedral angle type definition	k_φ (kcal/mol)	n	δ (degrees)
CH2 CH2 NCP CH2	0.9450	3	0
NCP CNG CPG HC2	2.7060	1	0
NCP CNG CPG OPG	1.4450	3	0
CNG CPG OPG CPG	0.9620	3	0
CH2 NCP CNG HC2	0.6830	3	0
HC2 CH2 NCP CNG	2.1230	1	180
NNH1 CH2 CH2 NCP	1.9420	2	0
CH2 NCP CH2 HC2	2.3080	1	180
CH2 NCP CNG CPG	0.8050	2	0
HC2 CH2 CH2 NCP	0.5720	1	180
CH2 CH2 NCP CNG	1.4330	3	0
HC2 CNG CPG HC2	0.1610	1	180
HC2 CNG CPG OPG	0.3260	3	180

Table 3.3. Dihedral angle final parameter set. NNH1 is an unprotonated nitrogen atom type present in the PEI force field [31].

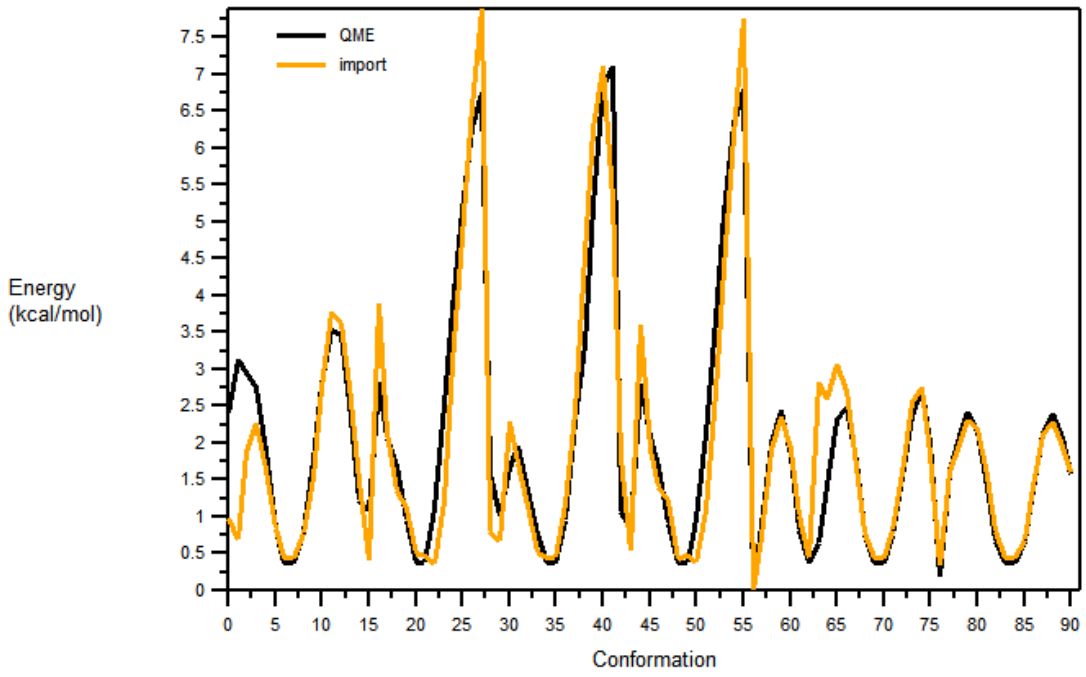

Fig. 3.2 QM energy (concatenated scans) vs MM energy produced by the optimized parameters.

3.2. Molecular dynamics simulations. PEI-PEG and DNA-PEI-PEG solvated systems.

3.2.1. Model structure relaxation.

Before doing larger simulations, it is useful to analyze the effect that the combined force field has on geometry via molecule relaxation. This is done easily and quickly in NAMD and provides an initial validation step for the calculated parameters.

The full set was observed to produce mostly small deviations from the relaxed geometry of the model structure optimized earlier. This can be readily seen in figure 3.3. This is an initial confirmation of the quality of the combined force fields for polyethylenimine, polyethylene glycol and the newly developed connector. In particular, the largest backbone difference is observed with NCP, slightly pulling out of the molecule's plane as it tries to locally accommodate two different behaviors. Overall, no unphysical behavior was noticed during the relaxation.

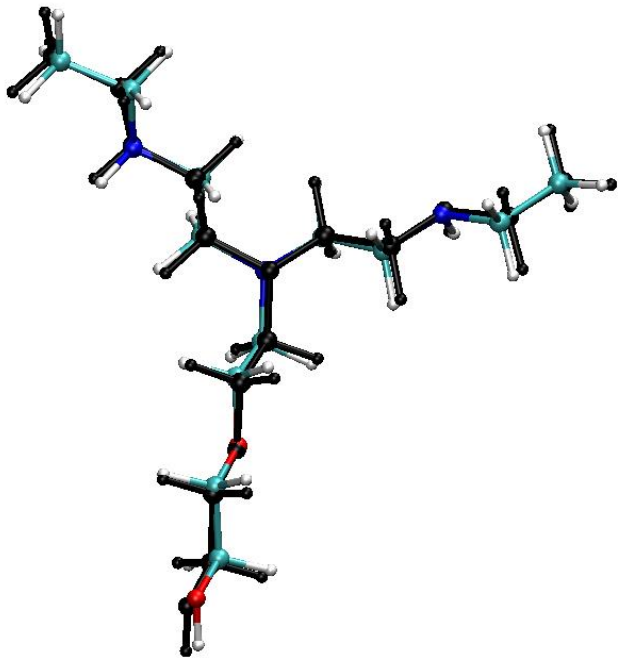

Fig. 3.3. QM (black) vs MM (colored) relaxed model structure.

3.2.2. PEI-PEG simulations.

To characterize the PEI-PEG copolymer model used in the polyplex simulations, a set of 20 separate, solvated PEI-PEG simulations was run. The simulated system consisted of one copolymer molecule surrounded by a cubic water box of side length 60 Å. A few water molecules were replaced by chloride ions to bring the net charge of the system to zero for a total of 20403 atoms. The addition of 8 chloride ions was necessary due to the charged PEP residues contained within the copolymer. Failure to properly neutralize a simulated system may result in long term errors due to improperly used PME, which requires a net neutral periodic cell (i.e., a neutral system).

Each simulation was run for 2510000 steps, each step representing 2 femtoseconds. An initial minimization process of 10000 steps is done before each simulation to help relax the system. The following 2500000 are the actual simulation time. The cumulated system evolution time is therefore a grand total of 100 nanoseconds, which was deemed sufficient to extract a few quantities of interest, namely radius of gyration and diffusion coefficient.

Periodic boundary conditions were employed, as was the Langevin thermostat for temperature control, at a temperature of 300 K and a damping constant of 1 ps^{-1}. The related Langevin piston was used to keep the pressure around 1.01325 bar. Particle Mesh Ewald was used for calculating electrostatic interactions, with a cut-off distance of 12 Å and a switching distance of 10 Å.

Five nanoseconds per simulation provides enough time for the system to reach an equilibrated phase (around the first 1-2 nanoseconds) and provide adequate data for investigation. However, after around five nanoseconds, the system evolution produces data which is not of use for extracting the aforementioned quantities of interest. Running shorter simulations also allows faster relevant data gathering, allowing an accelerated convergence of the quantities of interest, especially for the notoriously slowly converging diffusion coefficient.

3.2.2.1. Copolymer radius of gyration.

The radius of gyration R_{gyr} of a system is a measure of the distribution its mass elements around its center of mass (CM). Numerically, one calculated R_{gyr} as the square root of the average atom – CM distance squared, weighed by atom masses:

$$R_{gyr} = \sqrt{\frac{\sum_i m_i*(r_i - r_{CM})^2}{M}} \quad (8)$$

where M is the total mass of the system and r_{CM} is the CM of the configuration. By averaging the radius of gyration across all simulations, one finds the ensemble-average value, or $< R_{gyr} >$, which is more representative in general.

Figure 3.4 shows the average over the 20 simulations of the radius of gyration's time evolution. An initial drop in the calculated value may be observed mainly within the first nanosecond: this is a practical example of the equilibration phase. After around 2 nanoseconds, $< R_{gyr} >$ fluctuates around some value, which was found to be ~12.029 Å with a standard deviation σ = 0.244 Å, two orders of magnitude smaller. Indeed, this can be seen as the fact that, after 2 ns, the time evolution of the radius of gyration is a tendency to stay within a small distance of the calculated value.

Beu et al found that branched polyethylenimine of similar configuration with the currently analyzed copolymer in terms of size and protonation displays different radii of gyration depending on protonation [30]. The results obtained here seem to agree more with a protonation of 0.25 rather than 0.5, that is, with 1 in 4 PEI residues being protonated rather than 1 in 2. Indeed, though the polyethylenimine main chain does have almost the same number of positive residues and neutral ones, the polyethylene glycol side chain has no protonated residues. This reduced the protonated residue fraction to around 3 in 10, which is perfectly in line with the observed behavior.

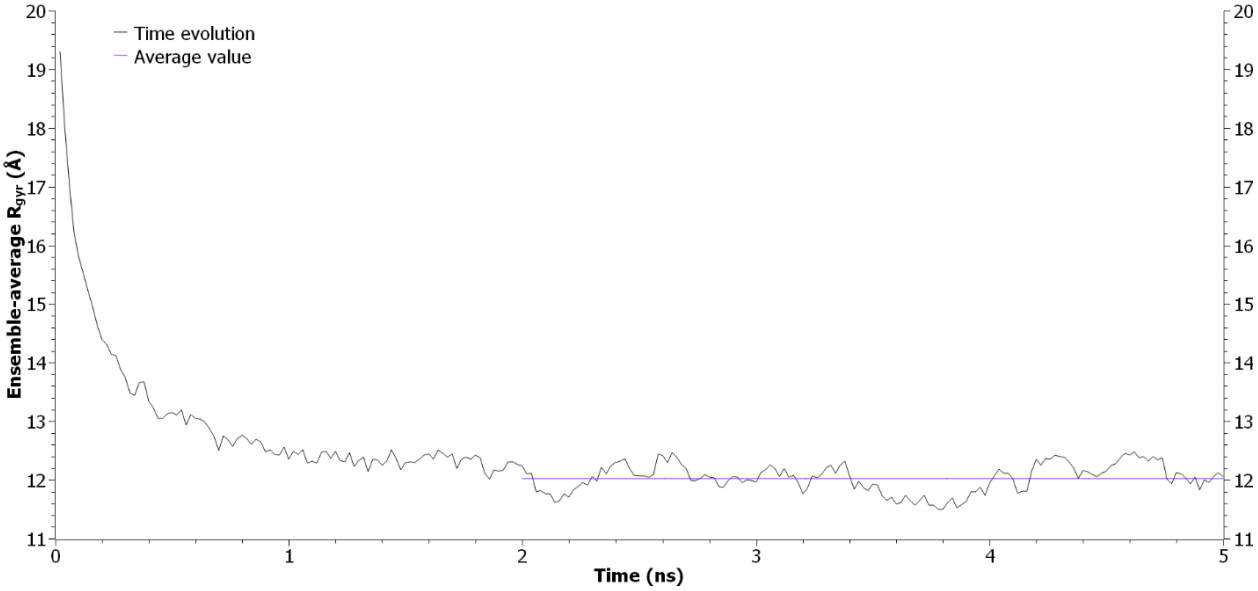

Fig. 3.4. Ensemble-averaged radius of gyration evolution.

3.2.2.2. Diffusion coefficient.

Next, the focus shifts to obtaining the diffusion coefficient. This is done by using the Einstein relation with ensemble-averaged mean-square-displacement (MSD) of the copolymer:

$$D = \frac{1}{6} * \lim_{t \to \infty} \frac{MSD(t)}{t} \tag{9}$$

The MSD is a measure of the displacement of the entire copolymer at some time, usually with respect to the initial configuration. It is known for a rather slow convergence in the sense of the large fluctuations in the MSD. This in turn is a natural consequence of the many interactions which take place, especially considering that Langevin dynamics specifically contains a term inducing randomness due to thermal agitation. Still, this usually requires a simulation set which provides sufficient data for the random fluctuations to cancel out and reduce the uncertainty in the diffusion coefficient.

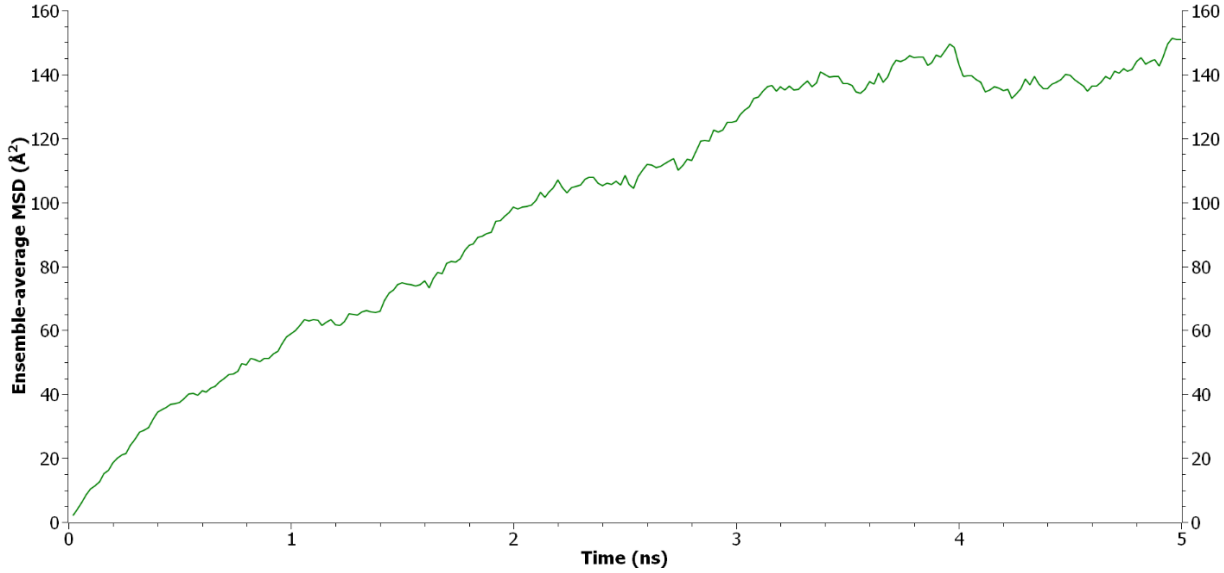

Fig. 3.5. Ensemble-averaged MSD time evolution.

Figure 3.5 shows the ensemble-averaged time evolution of the mean-square-displacement of the copolymer. After roughly 1 ns, the MSD becomes quasilinear until just before 4 ns, at which point decorrelation happens: the MSD starts fluctuating around a value instead of increasing. This shows why 5 ns per simulation were sufficient in this work, allowing for faster sampling in the interval of interest. Per the Einstein formula, the diffusion coefficient may be calculated from the slope of the MSD. Choosing the right time period is again a matter of know-how. Generally, one should look for a quasilinear behavior, after equilibration, but before decorrelation. Otherwise, unwanted influence can significantly affect the results.

By linear regression ($R^2 > 0.99$) the value $D = (3.2 \pm 0.4) * 10^{-6} \frac{cm^2}{s}$ was found on the [1ns, 3.8 ns] interval. Typically, errors tend to reach orders of magnitude even as high as the one for the diffusion coefficient itself. The relatively low error obtained is due to the large number of simulations in the set, which helps reduce the statistical fluctuations. As with the radius of gyration, the findings of the diffusion coefficient are more in agreement with the behavior of a 1 in 4 PEI protonation branched polyethylenimine of similar size [30]. With these findings in mind, an interesting observation arises. When isolated in a neutral water solution, a relatively small polyethylenimine chain of 1 in 2 protonation with an attached polyethylene glycol side chain (with approximately equal mass distribution per branch) seems to behave similarly to a branched polyethylenimine molecule of similar architecture, but with half the protonation.

3.2.3. DNA-PEI-PEG simulations.

Having characterized the behavior of the employed copolymer structure, the focus moves to the final stage of the study: the solvated DNA-PEI-PEG system, or DPP for brevity. For analyzing the complexation of DPP, another set of 20 simulations was run. In this set, systems were allowed to evolve for 10 nanoseconds after an initial minimization process, totaling once more 200 nanoseconds of simulation time. As shown already in chapter 2, figure 2.4., the initial configuration of the system exhibits three copolymer molecules placed at 30 Å from the dodecamer. Due to the inherent charge of the dodecamer, the system already had a net charge of zero, so there was no need to add ions of any kind for neutralization.

In order to reduce the simulated volume and thus the computational effort, the three arms of all copolymer molecules were rotated to form spiral configurations. Although improbable, it is not an impossible configuration, and the dynamics of the system will modify them to more relaxed, more probable geometries anyway. A similar argument applies to the initial lattice arrangement of water molecules. The DNA strand was subject to restraining the z coordinates of its atoms to their initial values: this allows the simulation of an infinitely long DNA chain. All in all, this makes the simulation volume 90 Å x 90 Å x 40 Å, fitting 30103 atoms within. Except for the number of steps per simulation, the box size and the DNA z coordinate restriction, all simulation conditions for this set are identical to those used for the solvated copolymer set.

Seeking to investigate and understand polyplex formation, two quantities were deemed of interest: the distribution of atoms around the DNA strand and the potential of mean force. The first one is a simple way to numerically describe how the polycations approach the dodecamer in time. The second quantity describes the free energy variation as a function of the way one kind of atom is distributed radially around another. In the process of calculating the potential of mean force, one also calculates the radial distribution function, a useful quantity for describing the surrounding particle configuration of a solvated molecule. For example, it can be used to find both the first solvation shell and the number of particles which, on average, are in the shell [42].

3.2.3.1 Atom distribution around DNA.

Beginning with the distribution of atoms around the DNA strand, figure 3.7 (a) shows the ensemble-average (i.e., average over all runs) within 5 Å of the dodecamer, which shall be named counting cut-off. After an extremely brief initial transient state, the water reaches a maximum of 1616 atoms before it begins to drop at around 0.5 nanoseconds, at which time atoms belonging to polyethylenimine arms have already started to enter the investigated volume. The drop in water and rise in polyethylenimine continues steadily until around 3 nanoseconds, at which point both tendencies begin to slow down. By 6 nanoseconds, water atom outflow and inflow are close enough to produce little changes over the [6 ns, 10 ns] interval. Polyethylenimine atoms display a similar behavior as their number barely increases in the same time period. Polyethylene glycol atoms, on the other hand, are almost completely absent, barely reaching 1% of the total by the 9 nanoseconds mark.

A similar behavior can be observed in figure 3.7 (b) with some qualitative differences. With the counting cut-off at 15 Å, the polycation atoms enter the much larger target volume earlier, but also reach steady state sooner. Another interesting difference is that polyethylene glycol atoms represent a significantly larger proportion than before. At the same time, the polyethylenimine atoms constitution with respect to the total number drops even though more than 90% are within the counting cut-off distance from the dodecamer. Evidently, in a larger volume, more copolymer atoms can fit in, but the same is true for water.

The natural step was to compare the steady state behavior at a few counting cut-off distances, precisely what is depicted in figure 3.7 (c). By analyzing the ensemble-average evolution of the three quantities, it was noticed that by 6 nanoseconds, the simulations tend to enter steady state (within a reasonable deviation) no matter the counting cut-off. As expected, the water content around DNA grows with increasing cut-off to the detriment of copolymer atoms. Even so, polyethylene glycol displays an interesting increase up to 15 Å before starting to decrease as well. Figure 3.7 (d) reveals more about the competitive behavior of the copolymer atoms. What it shows is that, as anticipated, the majority of polyethylenimine atoms tend to stay very close to the DNA strand while polyethylene glycol atoms keep a steady distance from the dodecamer.

To sum up these findings, by simply binning into three categories the different atoms within a certain counting cut-off distance of DNA across a few distances, it was possible to numerically describe the competitive behavior of polyethylenimine atoms in the presence of DNA. More than half reside *within* 5 Å of any DNA atom and, in doing so, removes some of the water initially solvating the dodecamer. Polyethylene glycol, on the other hand, prefers the water-rich region at larger distances, illustrated by the fact that, in steady state, roughly 76.5% of all polyethylene glycol atoms in the system are within 5 Å to 15 Å away from DNA atoms. These findings are presented in table 3.4. Figure 3.8 shows snapshots from a few simulations at various times, illustrating typical DPP configurations during complexation.

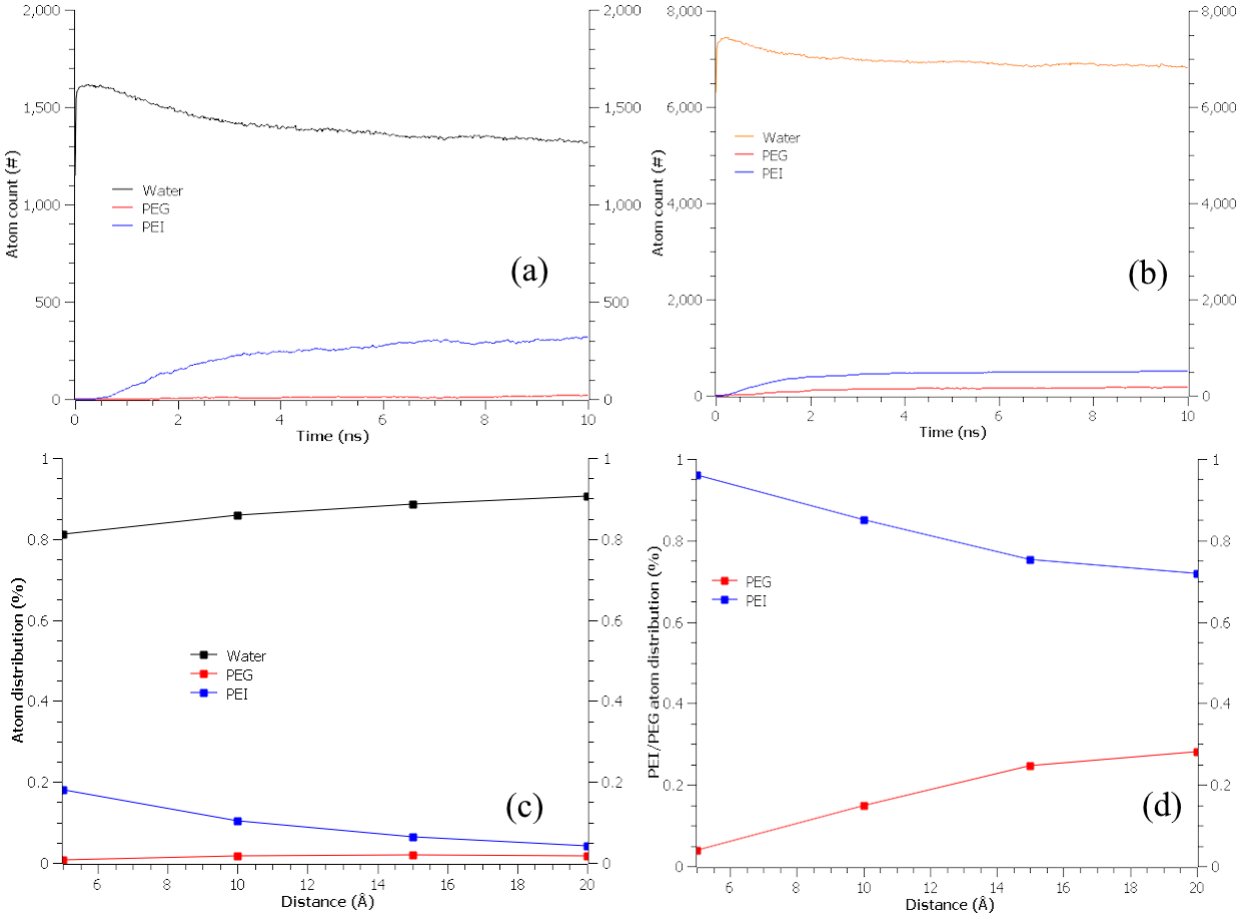

Fig. 3.6. Top: atom distribution around DNA in time at 5 Å (a) and 10 Å (b).

Bottom: normalized steady-state atom distribution; all residue types (c) and copolymer (d).

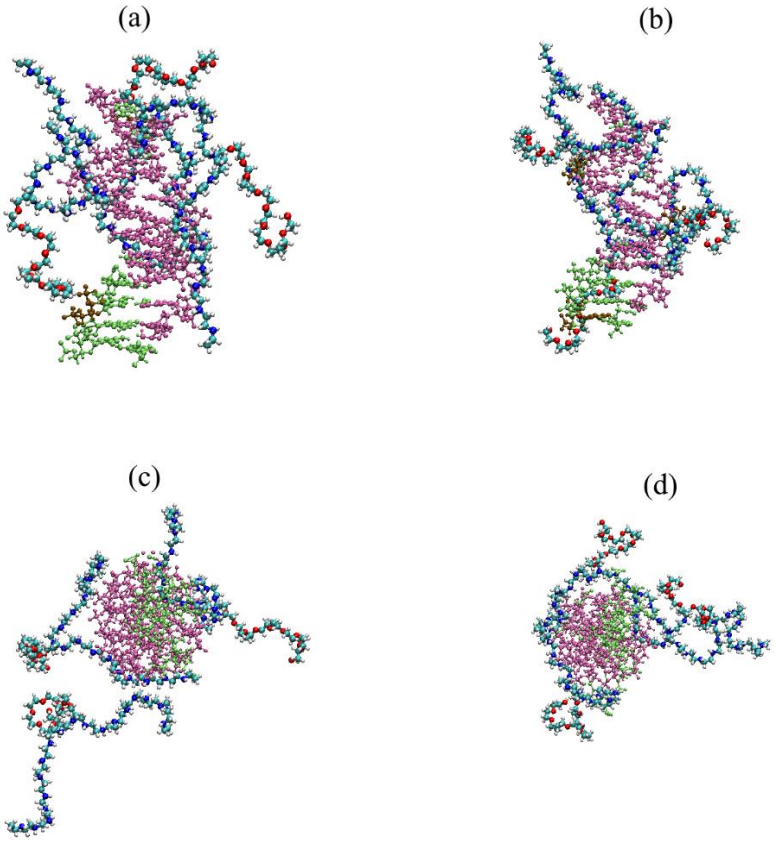

Fig. 3.7. Various snapshots from four simulations. Water atoms not shown. DNA has been colored as follows: brown – near PEG; purple – near PEP or PEI; green – near neither.

Quantity	Atom kind			Distance
	Water	Polyethylenimine	Polyethylene glycol	
Steady state atom count	1316	317	17	5 Å
	3765	475	97	10 Å
	6857	501	177	15 Å
	10460	509	200	20 Å
Polymer atom percentage	N/A	96.1 %	3.9 %	5 Å
		85 %	15 %	10 Å
		75.2 %	24.8 %	15 Å
		72 %	28 %	20 Å

Table 3.4. Atom distribution data. More than 78% of polyethylene glycol atoms lie between 5 Å and 15 Å, while more than 90% of polyethylenimine atoms are within 10 Å of the DNA strand.

A computational study by Ziebarth et al [43] on PEI/DNA and PEI/siRNA complexes showed that the amount of water around the nucleic chains decreases as polyethylenimine chains approach and form polyplexes. In their study, they discussed the release of water as polyethylenimine chains bind to DNA and siRNA. Though smaller in scope and simulated system, the presented results agree qualitatively. While discussing the distribution of atoms around the DNA strand, it was established that the number of atoms close to it decreases as the employed copolymers approach the dodecamer. Via comparison of internal and external results, it can then be argued that the presence of the polyethylene glycol side chain does not hinder polyethylenimine's capacity to expel water from the vicinity of the DNA. A future direction for research would be to run simulations of the same system, but with the presence of positive and negative ions, and observe ion expulsion patterns.

Ziebarth et al also showed that, once bound, the polyethylenimine chains do not change their positions significantly within the polyplex [43]. This is in agreement with the visually observed behavior in this work, especially when considering the first copolymer which manages to bind to DNA during a simulation. However, some rearrangement was sporadically observed to allow unbound PEI and PEP residues to approach the dodecamer.

Finally, due to the natural charge distribution of DNA, the polyethylenimine chains tended to approach and bind directly onto the phosphates in the backbone. However, though the minor groove would be theoretically a more attractive site in virtue of its higher charge density, the copolymers tend to simply approach the dodecamer and bind to whichever region is closest to them. This sometimes includes polyethylenimine chains following the minor grooves, but it also includes the same chains binding to DNA in a way that crosses over the major groove, turning against the nucleic chain. This behavior is reproduced in the work of Ziebarth et al [43] by linear polyethylenimine, further indicating that, when it comes to complexation, the employed copolymer behaves akin to a linear polyethylenimine chain of similar protonation. In other words, polyethylene glycol does not directly participate in forming the DPP complex, at least on a small scale.

3.2.3.2. Radial pair distribution function. Potential of mean force.

To better understand the polyplex formation, the radial pair distribution function (RPDF) and potential of mean force (PMF) are employed. The RPDF requires two categories of atoms, for example x and y, and provides data regarding the probability of finding an atom of kind y at some distance around an atom of kind x. In this work, two such RPDFs have been calculated. As determined previously, polyethylenimine atoms tend to stay very close to the dodecamer. However, since there are both protonated (PEP) and neutral (PEI) residues in the two polyethylenimine arms, it quickly becomes apparent that the two nitrogen atoms within may have different distributions.

For investigation, the RPDFs for DNA P – PEP N and DNA P – PEI N were computed, then used for calculating the respective potentials of mean function, which are illustrated in figure 3.9. The potential of mean force is a quantity proportional to negative the logarithm of the normalized RPDF:

$$PMF(x,y) = -k_B T * \ln\left(RPDF(x,y)\right) \tag{10}$$

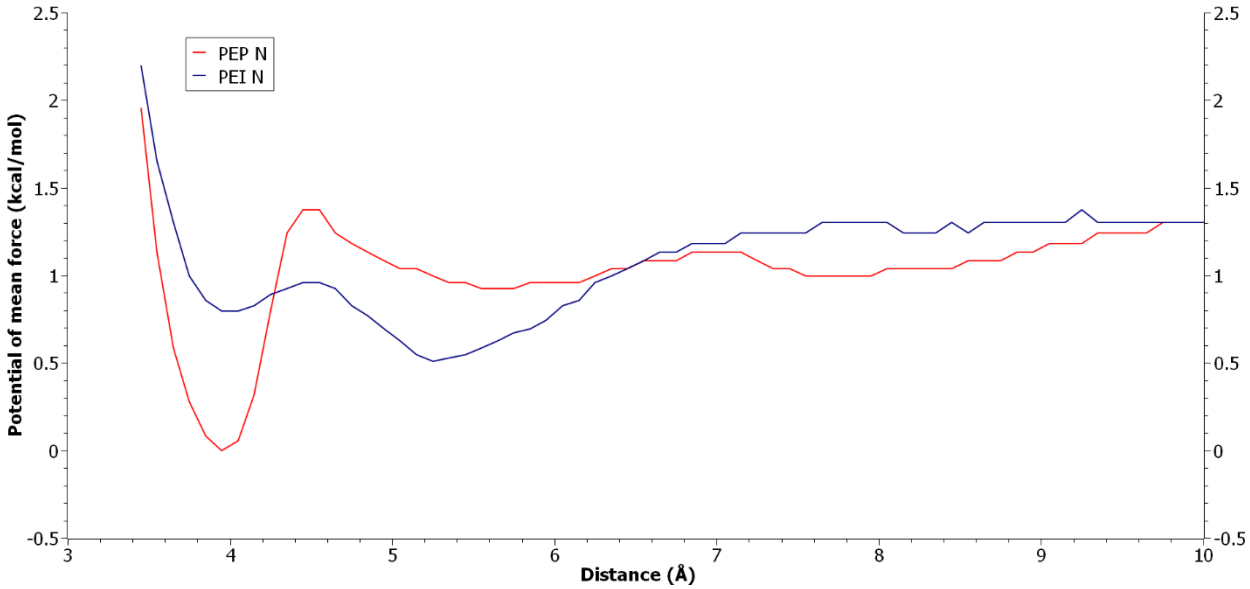

Figure 3.8. The potential of mean force for phosphorous – PEP N and phosphorous – PEI N.

The potential of mean force can be used to derive a force by using equation (2). This is the mean force which two particles, one from category x and one from y, would experience at some distance (hence the name *potential of mean force*).

Particularly, x is the set of DNA phosphorous atoms. For the PEP N curve, y was the set of all protonated nitrogen atoms in the polycation. For the PEI N curve, y was the complementary set of unprotonated nitrogen copolymer atoms. A slight modification was made in that both curves were shifted by the value which makes the PEP N curve have a global minimum of zero.

In the case of PEP nitrogen, what may be observed is a deep, relatively sharp minimum at 3.95 Å, the same value reported by Beu et al [30]. This indicates a strong preference for protonated N to stay at roughly that distance away from DNA phosphorus atoms. The sharpness of the minimum indicates a stable configuration once complexation has been achieved, providing an explanation for the observed lack of mobility that polyethylenimine chains have once bound to the dodecamer. Away from the minimum, there is some weak fluctuation as the PMF slowly increases with larger distances.

In the case of unprotonated nitrogen, the region in which PEP N is stably bound does not confer the same stability after complexation. There is a minimum visible at around 4 Å, but quite weak and easy to escape. The more favorable area, as per the PEI N curve, is an area at around 5.25 Å. This indicates that polyethylenimine's unprotonated nitrogen atoms are more favorably arranged at a distance quite larger compared to protonated sites. This is of course to be expected, as unprotonated nitrogen atoms do not have the same affinity for the DNA phosphate groups as protonated ones do. Nevertheless, it is in agreement with the observation regarding the polyethylenimine atom distribution around DNA.

CONCLUSIONS

A molecular mechanics force field for a (polyethylenimine) – (polyethylene glycol) copolymer connector has been developed and tested by various means. Initially, a full parametrization was required, which implied suitably choosing Lennard-Jones potential parameters for key new atom types, finding the optimized geometry of the connector, obtaining a partial charge distribution across the entire connector, calculating equilibrium values and force constants for new bond and angle types, then finally developing a parameter set for novel dihedrals in and around the connector. Periodically, comparisons were made between the parameters of the connector and a homogeneous branched polyethylenimine connector.

A series of simulations were run in order to establish the validity of the newly obtained parameter set. In parallel, the simulations helped reveal the interplay between the force field and two additional related ones developed recently. Lots of simulations were run to observe solvated polyethylenimine – polyethylene glycol copolymer behavior. The behavior was numerically investigated by calculating the radius of gyration and the diffusion coefficient. Comparisons were made between presented data and recently published findings for branched polyethylenimine of similar size. Finally, DNA-PEI-PEG complexation was analyzed to determine the behavior of the copolymer in the presence of the Drew-Dickerson dodecamer. Visual investigation determined that polyethylene glycol does not hinder the main polyethylenimine chain's ability to quickly and stably attach itself to the nucleic chain. The competitive behavior of polyethylenimine was analyzed by considering the atom distribution around DNA. The hydrophilicity of polyethylene glycol was also evidenced as most of its atoms tended to keep a reasonable distance from the dodecamer, in a more water-rich region. Finally, the potential of mean force was employed to analyze differences between protonated and unprotonated polyethylenimine nitrogen atoms during complexation, with results closely matching existing ones.

This work serves as a good starting point for future studies on the topic of DNA-PEI-PEG complexation. Directions for improvement include larger simulated systems, protonation fraction variation, addition of anions and cations to better reproduce physiological conditions, developing a coarse-grained force field, all culminating in large-scale simulations to analyze the effects of larger polyethylenimine – polyethylene glycol copolymers on much longer DNA strands.

BIBLIOGRAPHY

[1] Richard C. Mulligan, *The Basic Science of Gene Therapy*, Science, **260**, 926-932, 1993

[2] Friedmann T. *A brief history of gene therapy*. Nat Genet., **2** (2), 93-8, 1992

[3] Maurice S. Fox, John W. Littlefield, *Reservations Concerning Gene Therapy*, Science, **173**, 195-195, 1971.

[4] Hanna E, Rémuzat C, Auquier P, Toumi M., *Gene therapies development: slow progress and promising prospect*. J Mark Access Health Policy. **5(1)**, 1265293, 2017

[5] Elena Junquera, Emilio Aicart, *Recent progress in gene therapy to deliver nucleic acids with multivalent cationic vectors*, Advances in Colloid and Interface Science, **233**, 161-175, 2016.

[6] Akinc, A., Thomas, M., Klibanov, A.M., Langer, R., *Exploring polyethylenimine-mediated DNA transfection and the proton sponge hypothesis*, J. Gene Med., **7**, 657-663, 2005.

[7] Salameh JW, Zhou L, Ward SM, Santa Chalarca CF, Emrick T, Figueiredo ML., *Polymer-mediated gene therapy: Recent advances and merging of delivery techniques*, Wiley Interdiscip Rev Nanomed Nanobiotechnol.**12(2),** e1598, 2020

[8] https://a873679.fmphost.com/fmi/webd/GTCT (accessed: 02.06.2023)

[9] https://www.genome.gov/news/news-release/Gene-therapy-shows-promise-for-treating-Niemann-Pick-disease-type-C1 (accessed: 19.05.2023)

[10] Nimesh Surendra, *Polyethylenimine as a Promising Vector for Targeted siRNA Delivery*, Current Clinical Pharmacology, **7(2)**, 121 – 130, 2012

[11] Khansarizadeh M, Mokhtarzadeh A, Rashedinia M, et al., *Identification of possible cytotoxicity mechanism of polyethylenimine by proteomics analysis*, Human & Experimental Toxicology, **35(4),** 377-387, 2016

[12] Tang GP, Zeng JM, Gao SJ, Ma YX, Shi L, Li Y, Too HP, Wang S., *Polyethylene glycol modified polyethylenimine for improved CNS gene transfer: effects of PEGylation extent*, Biomaterials. **24(13),** 2351-62, 2003

[13] Anh K. Lam, Erika L. Moen, Jennifer Pusavat, Cassandra L. Wouters, Hannah Panlilio, Maya J. Ferrell, Matthew B. Houck, Daniel T. Glatzhofer, Charles V. Rice, *PEGylation of Polyethylenimine Lowers Acute Toxicity while Retaining Anti-Biofilm and β-Lactam Potentiation Properties against Antibiotic-Resistant Pathogens*, ACS Omega, **5 (40)**, 26262-26270, 2020

[14] Ciccotti, G., Dellago, C., Ferrario, M. et al., *Molecular simulations: past, present, and future (a Topical Issue in EPJB)*. Eur. Phys. J. B**, 95**, 3 (2022).

[15] Levitt M, Warshel A, *Computer simulation of protein folding*. Nature. **253** (5494): 694–698, 1975

[16] Palma, C.-A.; Björk, J.; Rao, F.; Kühne, D.; Klappenberger, F.; Barth, J.V., *Topological Dynamics in Supramolecular Rotors*. Nano Letters. **148** (8): 4461–4468, 2014

[17] Smith, R., *Atomic & ion collisions in solids and at surfaces: theory, simulation and applications*., Cambridge University Press, Cambridge, UK, 1997

[18] Casalino L, Dommer AC, Gaieb Z, et al., *AI-driven multiscale simulations illuminate mechanisms of SARS-CoV-2 spike dynamics*. The International Journal of High Performance Computing Applications., **35(5)**, 432-451, 2021

[19] https://www.ks.uiuc.edu/Research/namd/2.9/ug/node5.html (accessed: 21.05.2023)

[20] William C. Swope, Hans C. Andersen, Peter H. Berens, Kent R. Wilson; *A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters*. J. Chem. Phys., **76 (1)**, 637–649, 1982

[21] MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FT, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M., *All-atom*

empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B.,**102(18),** 3586-3616, 1998

[22] Gkeka, Paraskevi., *Molecular dynamics studies of peptide-membrane interactions: insights from coarse-grained models*, 2010.

[23] https://www.compchems.com/molecular-dynamics-periodic-boundary-conditions-pbc/#minimum-image-convention-and-cut-off-radius (accessed: 21:05.2022)

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

[25] https://courses.physics.illinois.edu/phys466/sp2013/projects/2002/team1/Ewald3.htm (accessed: 22.05.2023)

[26] Schlick, Tamar (2002). *Molecular Modeling and Simulation*. Springer. p. 480

[27] Langevin, P., *Sur la théorie du mouvement brownien [On the Theory of Brownian Motion]*. C. R. Acad. Sci. Paris. **146**: 530–533, 1908

[28] Foloppe, N., MacKerell, Jr., A.D. *All-Atom Empirical Force Field for Nucleic Acids: 1) Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data.*, J. Comput. Chem, **21**, 86-104, 2000

[29] MacKerell, Jr., A.D. and Banavali, N. *All-Atom Empirical Force Field for Nucleic Acids: 2) Application to Molecular Dynamics Simulations of DNA and RNA in Solution.*, J. Comput. Chem, 2000, **21**, 105-120.

[30] Terteci-Popescu, A.-E., Beu, T. A., *Branched polyethyleneimine: CHARMM force field and molecular dynamics simulations*, J. Comput. Chem., **43 (31)**, 2072, 2022

[31] Beu, T. A., Ailenei, A.-E., Farcaş, A., *CHARMM Force Field for Protonated Polyethyleneimine*, J. Comput. Chem., **39**, 2564– 2575, 2018

[32] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML, *Comparison of simple potential functions for simulating liquid water.* The Journal of Chemical Physics. **79** (2): 926–935, 1983

[33] C.G. Mayne, J. Saam, K. Schulten, E. Tajkhorshid, J.C. Gumbart., *Rapid parameterization of small molecules using the Force Field Toolkit*, J. Comput. Chem. 2013, 34, 2757-2770.

[34] Wei Z, Luijten E., *Systematic coarse-grained modeling of complexation between small interfering RNA and polycations*. J Chem Phys., **143(24),** 243146, 2015

[35] https://www.rcsb.org/structure/4c64 (accessed: 15.05.2023)

[36] https://phys.ubbcluj.ro/~titus.beu/MDsquad/libraries.html (accessed: 16.05.2023)

[37] https://www.ks.uiuc.edu/Research/vmd/ (accessed: 22.05.2023)

[38] https://www.tcl.tk/about/language.html (accessed: 26.05.2023)

[39] https://www.python.org/ (accessed: 26.05.2023)

[40] https://scidavis.sourceforge.net/ (accessed: 22.05.2023)

[41] Gaussian 09, Revision E.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2013.

[42]https://chem.libretexts.org/Bookshelves/Biological_Chemistry/Concepts_in_Biophysical_Chemistry_(Tokmakoff)/01%3A_Water_and_Aqueous_Solutions/01%3A_Fluids/1.02%3A_Radial_Distribution_Function (accessed: 29.05.2023)

[43] Ziebarth JD, Kennetz DR, Walker NJ, Wang Y., *Structural Comparisons of PEI/DNA and PEI/siRNA Complexes Revealed with Molecular Dynamics Simulations*. J Phys Chem B., **121(8)** 1941-1952,

Anexa 2

DECLARAȚIE PE PROPRIE RĂSPUNDERE

Subsemnatul, Mircea Alex-Ovidiu, declar că Lucrarea de disertație pe care o voi prezenta în cadrul examenului de finalizare a studiilor la Facultatea de Fizică, din cadrul Universității Babeș-Bolyai, în sesiunea iulie 2023 , sub îndrumarea Prof. dr. Beu Titus Adrian, reprezintă o operă personală. Menționez că nu am plagiat o altă lucrare publicată, prezentată public sau un fișier postat pe Internet. Pentru realizarea lucrării am folosit exclusiv bibliografia prezentată și nu am ascuns nici o altă sursă bibliografică sau fișier electronic pe care să le fi folosit la redactarea lucrării.

Prezenta declarație este parte a lucrării și se anexează la aceasta.

Data, Nume,

26.06.2023 Mircea Alex-Ovidiu

Semnătură