BABEŞ-BOLYAI UNIVERSITY
FACULTY OF PHYSICS

SUMMARY OF THE PHD THESIS

Modeling and simulations of complex biomolecular systems

Author:
Andrada-Elena AILENEI
(căs. TERTECI-POPESCU)

Supervisor:
Prof. Dr. Titus Adrian BEU

Cluj-Napoca
2023

Abstract

Polyethyleneimine (PEI) is one of the most promising non-viral vectors for gene delivery due to its high efficiency in forming complexes with DNA. This thesis focuses on investigating PEI at the all-atom (AA) and coarse-grained (CG) resolutions.

We developed an AA CHARMM force field (FF) for linear and branched PEI, based on ab initio calculations on model polymers. As a novelty with respect to previous studies, we optimized consistently the partial atomic charges and the entire set of bonded parameters. The AA FF was validated by molecular dynamics (MD) simulations of PEIs of various molecular weights and protonation fractions. The resulting statical (gyration radii, end-to-end distances) and dynamical (diffusion coefficients) properties are in line with previous computational and experimental studies. We performed MD simulations of DNA-PEI complex formation, which determined the most favorable attachment conformations.

To develop a CG Martini FF for linear PEI, we first identified entire AA residues as CG beads, and selected the non-bonded parameters from standard Martini types. The bonded parameters were derived based on applying the Boltzmann Inversion technique on probability distributions of bond lengths, angles, and dihedrals resulted from the AA simulations. The CG simulations of PEI revealed a very good agreement with the AA quantities and experimental results, validating the developed CG FF.

The developed AA and CG FFs for PEI can be used, jointly with the CHARMM/ Martini FFs for DNA, in extensive simulations of DNA-PEI condensation, which are expected to provide meaningful knowledge for the realization of efficient gene delivery protocols.

Keywords

atomistic force fields, coarse-grained force fields, cationic polymers, molecular dynamics, polyethyleneimine, non-viral vectors, gene delivery.

Contents

1

Gene delivery

1.1 General aspects

The discovery of deoxyribonucleic acid (DNA) in the 1950s marked the beginning of modern molecular biology research. This field gained great popularity and interest due to providing prospects for treating some hereditary diseases, which were believed to be incurable up to this point. The concept of delivering nucleic acids into cells via a carrier was introduced in the 1970s, laying the foundations of a novel research field.

Gene delivery vectors are grouped into two major categories – viral (virus capsids) and non-viral (cationic polymers, lipids, inorganic nanoparticles), each presenting key advantages and limitations. The delivery vector is required to have a good safety profile (low toxicity), high transfection efficiency, good stability, favorable transgene expression, easy production, and cost-effectiveness.

Viral vectors feature natural tropism to infect cells and high transfer efficiency (both in vivo and ex vivo), however presenting major issues, such as low packaging capacity for the genetic material and high immunogenicity and toxicity [1]. Nevertheless, it is noteworthy that most of the currently approved gene therapy drugs are based on viral vectors.

Non-viral vectors offer low cytotoxicity and immunogenicity, easy manufacturing, and cost-effectiveness, but fall behind viral carriers by low transfection efficiency [2]. Owing to their straightforward chemistry, low production cost, and low cytotoxicity, *cationic polymers* are currently the most widely used non-viral vectors for gene delivery protocols. Cationic polymers condense the genetic payload into *polyplexes* via electrostatic interactions between the polymer's protonated amino groups (positively charged) and the phosphate groups of DNA/RNA (negatively charged). The most commonly employed non-viral carriers are polyethyleneimine, poly(amidoamine), poly-L-lysine, chitosan, and cationic dendrimers [3].

Polyethyleneimine (PEI) is a highly attractive polymer for gene delivery, its monomers being composed of an amine group (unprotonated – NH, or protonated – NH_2^+) and two aliphatic spacers (CH_2). PEI can have different geometries, namely linear (*L*PEI), and branched (*B*PEI). The distribution of primary, secondary, and tertiary amines in PEI is linked to its considerable buffering capacity, which relates to the ability of PEI to escape from the endosome.

The limitations of experimental measurements at the atomic scale create a significant knowledge gap regarding some of the properties of gene delivery complexes, and their formation and condensation mechanisms, which need to be addressed for practical and safe protocols to be realized. As a means to overcome such limitations, and learn optimal conditions for efficient gene delivery approaches, deterministic simulation methods, such as molecular dynamics (MD), are most commonly employed.

1.2 Thesis outline

The first chapter conveys the general aspects regarding gene delivery, clinical trials, types of employed vectors, and current limitations. The second chapter covers theoretical aspects on molecular dynamics including propagators, temperature and pressure coupling, and electrostatics, along with a description of molecular mechanics force fields. The third and fourth chapters present the developed force fields and original results of the present thesis, being grouped into two categories, namely the ones involving the all-atom representation (third chapter) and those covering the coarse-grained scale (fourth chapter). The thesis concludes with a final chapter sumarizing the central results of the study and future research prospects.

2

Theoretical background

2.1 Molecular dynamics integrators

Computer simulations are an effective means to study molecular systems at the microscopic scale, providing data that is not readily accessible in experiments [4; 5]. Molecular dynamics is a *deterministic* class of simulations, which produces the trajectory of particles (atoms, molecules) by solving Newton's equations of motion numerically. In the case of classical MD, the forces acting on the system particles are modeled by molecular mechanics (MM) force fields (FFs), which describe bonded and non-bonded interactions.

The central principle in classical molecular dynamics is to divide the simulation time into discrete time steps, calculate the forces acting on the atoms of the system, and update their positions and velocities by solving Newton's equations of motion:

$$m_i \frac{d^2 \boldsymbol{r}_i}{dt^2} = \boldsymbol{F}_i, \quad i = 1, 2, ..., N, \tag{2.1}$$

where m_i is the mass of particle i. $\boldsymbol{F}_i$ is the resulting force acting upon particle i, produced by external and interatomic fields.

At each time step, an MD integrator, such as the *Velocity Verlet* algorithm [6; 7], is employed to propagate the velocities and positions to the next time step. The Velocity Verlet method is a two-step algorithm, which uses synchronized positions and velocities for the integration of the equations of motion. In the predictor step, the velocity is advanced at mid interval and the position – over an entire time step:

$$\begin{aligned} \boldsymbol{v}\left(t + \frac{\delta t}{2}\right) &= \boldsymbol{v}(t) + \dot{\boldsymbol{v}}\frac{\delta t}{2} + O(\delta t^2) \\ \boldsymbol{r}(t + \delta t) &= \boldsymbol{r}(t) + \boldsymbol{v}\left(t + \frac{\delta t}{2}\right)\delta t + O(\delta t^3) \end{aligned} \tag{2.2}$$

In the corrector step, the acceleration and the velocity are propagated at the end of a full time interval:

$$\begin{aligned} \boldsymbol{a}(t + \delta t) &= \frac{\boldsymbol{F}\left(\boldsymbol{r}(t + \delta t)\right)}{m} \\ \boldsymbol{v}(t + \delta t) &= \boldsymbol{v}\left(t + \frac{\delta t}{2}\right) + \boldsymbol{a}(t + \delta t)\frac{\delta t}{2} \end{aligned} \tag{2.3}$$

The velocity Verlet algorithm is time-reversible and properly conserves the total energy of the system. It is an $O(\delta t^3)$ algorithm and has small storage requirements [8]. Velocity Verlet is the most commonly employed integration method in modern MD simulation programs.

2.2 Temperature and pressure coupling

In many applications, a thermostat can be employed to keep the temperature at a constant value, since, generally, one needs to calculate quantities from the NVT or NPT ensembles (constant number of particles – N, constant volume – V or pressure – P, respectively, and constant temperature – T) [9; 10]. Many biomolecular systems revealed different behaviors/properties at different temperatures (for example protein folding/unfolding [11]), so the issue of temperature coupling becomes particularly important. The *Berendsen* algorithm [12] implies that the system is coupled to a heat bath of temperature T_0, and, by rescaling the velocities of the particles, attempts to correct the deviations of the system temperature, T, from T_0.

Since the experiments are usually performed under constant pressure, MD simulations generally require employing a barostat to produce meaningful results, which can be readily compared to the experimental data. The *Parrinello-Rahman* algorithm [13; 14] ensures a constant value of the pressure during the simulation by adjusting the coordinates of the particles and the simulation box vectors. Notably, this algorithm generates trajectories consistent with an NPT ensemble [10].

2.3 Molecular mechanics force fields

In the field of computational chemistry and molecular dynamics, a force field (FF) is a model specifying the functional form of the potential energy from which the actual forces acting on the atoms are derived. The potential incorporates terms for bonded-, and non-bonded interactions, as described in the following.

In most of the presently employed FF models, the non-bonded interactions are comprised of repulsion and dispersion terms present in the *van der Waals* potential and electrostatic interactions ($U_{Coulomb}$). The $12-6$ *Lennard-Jones (LJ) potential* is the most commonly employed pair potential for van der Waals interactions, being expressed as [15]:

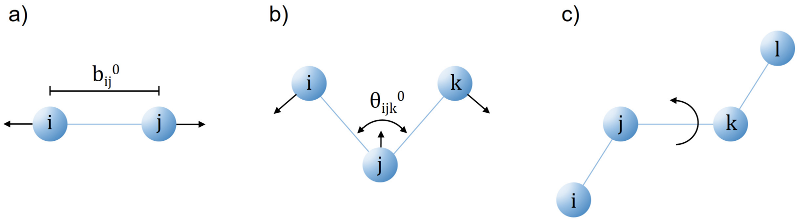


Figure 2.1: Bonded interactions. After Abraham et al. [10]

$$U_{LJ}(r_{ij}) = \begin{cases} 4\varepsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] & , r_{ij} < r_{cut} \\ 0 & , r_{ij} \geqslant r_{cut} \end{cases} \tag{2.4}$$

where r_{ij} is the inter-distance between atoms i and j, σ is the zero-potential distance, and ε is the depth of the potential well. r_{cut} is a cut-off distance, beyond which the LJ interaction is considered negligible.

For the pair of atoms i and j, the $(\sigma/r_{ij})^{12}$ term determines repulsion at small r_{ij}, on account of overlap of electronic orbitals. The second term in equation 2.4, namely $-(\sigma/r_{ij})^6$, characterizes attraction at large r_{ij} values (corresponding to London dispersion forces) due to dipole-dipole interactions.

The *Coulomb potential* defines the *electrostatic* interaction between two particles characterized by charges q_i and q_j, and positions $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$, respectively:

$$U_{Coulomb}(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}} \tag{2.5}$$

where $r_{ij} = |\boldsymbol{r}_i - \boldsymbol{r}_j|$ denotes the distance between atoms i and j, ε_0 is the vacuum permittivity and ε_r is the relative permittivity.

Bonded interactions represent the stretching of bonds, bending of angles, and torsion of the dihedral angles (Figure 2.1 a, b, and c, respectively). *Bond stretching* (Figure 2.1a) can be represented by:

$$u_b(b) = \frac{1}{2}K_b(b - b_0)^2 \tag{2.6}$$

This potential reproduces the Morse potential minimum very well, however, the breaking of bonds and atom separation are not accounted for in this approximation [16].

Angle bending (Figure 2.1b) can also be modeled by a harmonic potential having the form of Hooke's law [11]:

$$u_\theta(\theta) = \frac{1}{2}K_\theta(\theta - \theta_0)^2 \tag{2.7}$$

where θ_0 is the reference angle, and K_θ is the angle force constant.

A proper dihedral angle (Figure 2.1c) is defined as the angle ϕ between the ijk and jkl planes, with the convention that a null angle represents the *cis* geometric isomer, and a 180° dihedral illustrates the *trans* configuration [10]. The potential for the torsion of dihedral angles is defined as:

$$u_\phi(\phi) = \frac{1}{2}K_\phi[1 + cos(n\phi - \phi_s)] \tag{2.8}$$

where n is the dihedral multiplicity, and ϕ_s is the phase shift.

Improper dihedral angles are used to model "out-of-plane bending" in planar parts of molecules (such as rings), and the usually considered potential is harmonic [10]:

$$u_\xi(\xi) = \frac{1}{2}K_\xi(\xi - \xi_0)^2 \tag{2.9}$$

ξ_0 is the reference angle, and K_ξ - the force constant for the out-of-plane bending [16].

Since the motion of an atom as part of a molecule is strongly coupled to the motions of other atoms, a potential term is sometimes useful to account for the effect of coupled bond stretching and angle bending motions. To this end, the Urey-Bradley potential is employed to model the interaction between the non-bonded atoms 1 and 3 as a consequence of the coupled stretch and bending motions [16]:

$$u_{UB}(r_{1,3}) = \frac{1}{2}K_{UB}(r_{1,3} - r_{1,3;0})^2 \tag{2.10}$$

where $r_{1,3}$ represents the distance between atoms 1 and 3, and $r_{1,3;0}$ – the respective equilibrium distance.

3

All-atom modeling and simulation of polyethyleneimine

3.1 Introduction

The field of gene delivery received increasing popularity in the last few decades, and, naturally, molecular dynamics (MD) simulations have been employed in this context to gain knowledge on the efficient design of delivery vectors. However, due to the very challenging nature of force field (FF) development, very few computational studies employed parameters specifically optimized for the simulated compounds, relying on general-purpose FFs. Specifically, the FFs employed in previous studies (for example the ones used in Ref. 17, 18, 19 etc) were only partially optimized for PEI, having either the partial atomic charges, or the dihedral terms adjusted.

Our group previously developed a FF for PEI based on –C–C–N– residues [20], which was characterized by a consistent parametrization of all bonded parameters and atomic charges. In this work, defining symmetric residues (–C–N–C–), we parametrized FFs for two PEI architectures, namely linear-, and branched polymers. The high quality of the resulting force field was demonstrated by performing AA simulations involving PEIs of various sizes and protonation fractions (f_P), comparing their structural and dynamical results with experimental data, and simulating the formation of DNA-PEI complexes.

Note: The nomenclature of the polymers contains a letter denoting the architecture: L - for linear, or B - for branched PEIs. The number of monomers forming (a) the chain in the case of LPEI, or (b) each of the three branches in the case of BPEI is also indicated. Finally, to highlight the protonation patterns, either "-p0", or "-un" is appended to the polymer names, "-p0" indicating non-protonated PEIs, and "-un" standing for uniformly protonated chains with 1-in-n protonated units.

3.2 Parametrization of a CHARMM force field for linear polyethyleneimine

Aiming to develop a simple, yet versatile force field for polyethyleneimine, that could be used in conjunction with the CHARMM FF for other compounds of interest (such

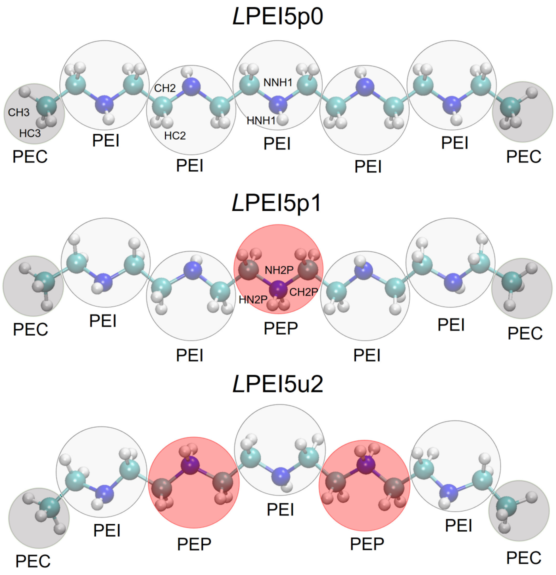


Figure 3.1: *L*PEI model polymers employed in the parametrization of the AA force field for linear chains, evidencing the defined residue- and atom types.

as DNA), we chose to define as few atom-, and residue types as possible. As seen in Figure 3.1, we developed the FF for *L*PEI on three model polymers, by firstly mapping whole functional groups into single *symmetrical* residues. Only integer charges (either $0e$ or $+1e$) were considered for the defined residues. Specifically, we defined the following residue types: *PEI*: $CH_2{-}NH{-}CH_2$ – non-protonated residue; *PEP*: $CH_2{-}NH_2{}^+{-}CH_2$ – protonated residue; and CH_3 – ending methyl residue.

To adequately characterize the connectivity between adjacent atoms in the chain, we defined nine atom types: *CH2* and *HC2*: – C and H atoms in the PEI residue; *NNH1* and *HNH1*: – N and H atoms in the NH group of the PEI residue; *CH2P*: – C atom in the PEP residue; *NH2P* and *HN2P*: – N and H atoms in the $NH_2{}^+$ group of the PEP residue; and *CH3* and *HC3*: – C and H atoms in the PEC residue.

3.2.1 Lennard-Jones parameters

The form of the LJ potential reads:

$$U_{LJ} = \sum_{atoms\ i,j} \varepsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min,ij}}{r_{ij}} \right)^{6} \right] \tag{3.1}$$

where ε_{ij} is the potential well depth, and $R_{min,ij}$ is the potential minimum position.

The Lennard-Jones parameters, specifically, the ε and R_{min} values for each atom type in the new *L*PEI model were assigned based on the most suitable correspondent in CHARMM36. The CH2 and CH2P atoms were identified as aliphatic carbons for CH_2 groups, adjacent to neutral and positive amino groups, respectively, while CH3 – as aliphatic C for CH_3 groups. The NHN1 and NH2P atoms were modeled as neutral dimethylamine- ($CH_3{-}NH{-}CH_3$) and secondary NH_2^+ nitrogens, respectively. For the hydrogens, we actually chose the suitable types based on their connectivity to a heavy atom: for example, HC2 was identified as aliphatic H for CH_2 groups, while HNH1 was most accurately described by an H atom from a dimethylamine group.

3.2.2 Partial atomic charges

Having in the view that one of the objectives of this work was to develop a coarse-grained force field for PEI, we chose to constrain the charges of individual residues (to become CG beads) to integer values.

The Coulomb potential depends on the partial atomic charges, and reads:

$$U_{Coulomb} = \sum_{atoms\ i,j} \frac{q_i q_j}{\epsilon_0 r_{ij}} \tag{3.2}$$

where q_i and q_j are the charges of atoms i and j, respectively, ϵ_0 represents the dielectric constant (multiplied by 4π in SI units), and r_{ij} is the i–j interdistance.

To parametrize the partial atomic charges, the force field ToolKit (ffTK) [21] is based on the central idea that the modeled molecules need to behave realistically in solvated environments, so the water molecules are required to have reasonable initial orientations relative to the considered compounds. To this end, the hydrogens in PEI were classified as hydrogen bond *donors*, while the nitrogens in the non-protonated residues – as *acceptors*. For each atom categorized either as donor, or as acceptor, a complex was constructed, which consisted of the *L*PEI model and a H_2O molecule placed in a representative relative position. For every *L*PEI-H_2O system, the relative distance, and rotation angle of the water molecule were optimized quantum mechanically, at HF/6-31G(d) level, while constraints were imposed on the other degrees of freedom.

In accordance with the standard CHARMM convention, all aliphatic hydrogen atoms were attributed the charge $+0.09e$, and were no longer included in the optimization process. We imposed neutrality to the non-protonated *L*PEI5p0 model, $+1e$ charge to *L*PEI5p1, and $+2e$ to *L*PEI5u2.

3.2.3 Bond and angle parameters

The bond- and angle potential has the following harmonic form:

$$U_{bond} + U_{angle} = \sum_{bonds} K_b(b - b_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 \tag{3.3}$$

where K_b and K_θ are the bond- and angle force constants, while b_0 and θ_0 are the respective equilibrium values.

For the optimization of the bond- and angle parameters, ffTK relies on iteratively adjusting the MM energy of distortion in such a way that it reproduces the QM counterpart as accurately as possible. The local landscape of the potential energy surface (PES) can be described by Hessian matrices, which can be computed through vibrational calculations performed with Gaussian09 [22]. Therefore, ffTK employs Hessian matrices to characterize the QM energy change determined by small deviations of the internal coordinates (ICs) from the equilibrium geometry. By contrast, the MM energy of distortion is readily given by total energies computed with the trial force constants and equilibrium values.

The MM-QM deviations are already small after the first optimization step, decreasing even more by the end of the optimization process, as also shown by the decrease in the objective function (of roughly 61%). Indeed, the very low values of the maximum MM-QM differences, namely $\sim 2 \times 10^{-2}$ Å for the bonds, and $\sim 2^\circ$ for the angles, confirm the high quality of the parameters. In the original paper of the CHARMM FF release [23], Vanommeslaeghe et al. indicated that the differences with respect to the QM data are required to be smaller than 3×10^{-2} Å, and 3° for the equilibrium bonds and angles, correspondingly.

The bonded parameters were averaged over their values in the three central residues (i.e. units 2, 3, and 4) of the model pentamers, to neglect any unwanted marginal effects of the PEC ends. Regarding the effect of protonation, we found a weaker C–N bond in the protonated PEP residue (CH2P–NH2P), than its correspondent in the PEI residue (CH2–NNH1), demonstrated by a 20% decrease in the force constant.

An initial inspection of the angle stiffness values emphasized that K_θ is higher in the vicinity, or inside protonated residues. The stiffer angles found in the presence of protonated sites are not linked to the constituting bonds, the latter actually being weaker inside the PEP residue.

3.2.4 Dihedral parameters

The dihedral potential depends on the force constant K_ϕ, the multiplicity n, and the phase shift δ, and is expressed as:

$$U_{dihedrals} = \sum_{dihedrals} K_{\phi}[1 + cos(n\phi - \delta)] \tag{3.4}$$

The approach to characterize the potential energy surface by Hessian matrices proves suitable for the bond- and angle optimization, since it accurately describes the energy landscape in the vicinity of minima. By contrast, for the dihedral angles, the treatment of high-energy regions, which is also of importance, can not be modeled by harmonic potentials. Accordingly, ffTK employs QM torsion scans of coordinates, for every dihedral angle along the model backbone, used to perform the fitting of the MM profiles. Specifically, for each scan, a series of QM optimizations of the geometry was compiled at MP2/6-31G(d) level, keeping the considered dihedral constrained, while allowing the remainder of the molecule to completely relax. Regarding the sequence of geometry optimizations, the dihedrals were typically distorted in steps of 10° in the symmetric interval of -90° and +90°.

The dihedral parametrization implied iteratively minimizing the deviations of the MM energy profile from the QM data until an acceptable fit (and a very small root-mean-square error) was achieved, which meant performing hundreds of optimization steps based on the downhill-, and simulated annealing techniques. The quality of the dihedral optimization is warranted by the match between the MM and QM torsion energies, and by the comprehensive set of model molecules, which includes the shortest chains that could accommodate the protonation patterns required for the present investigations.

Comparing our dihedral contributions with the ones reported in the literature (for example the ones in Ref. 24), we found significantly lower MM-QM deviations, which indicate a higher quality of the parameters.

3.3 Parametrization of a CHARMM force field for branched polyethyleneimine

We introduced two new residue types for the FF parametrization of branched PEIs, one for the branch connector, and the second for NH_2 ends, since it appears that experimental *B*PEI exhibits NH_2 endings more favorably: *PEY*: $N(CH_2)_3$ – branching connector residue, and *PEN*: $NH_2{-}CH_2$ – ending amine residue.

Figure 3.2, presents the two model polymers chosen for parametrizing *B*PEI molecules: *B*PEI1p0, used for deriving the branch connector (PEY) parameters, and *L*PEI1p0-*N*, employed for the PEN ending. In the following, the letter "-*N*" appended to the name

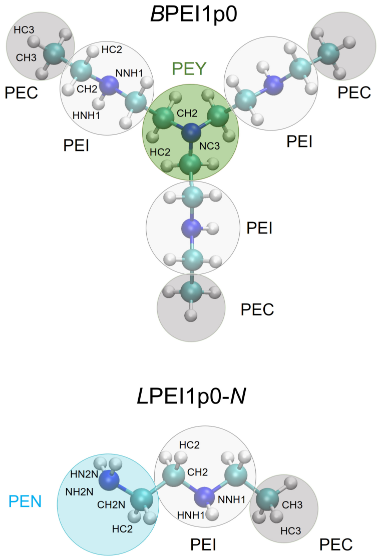


Figure 3.2: *B*PEI1p0 and *L*PEI1p0-*N* model molecules employed in the parametrization of the FF for branched chains, evidencing the defined residue- and atom-types.

of the molecules represents their ending in NH_2 groups, as opposed to the case of PEIs ending in CH_3 methyl groups, which bear no distinctive symbol.

We defined five additional atom types, to account for the new connectivities: *NC3*: – N atom in the tertiary amine; *CH2*: – C atom in the tertiary amine; *NH2N*: – N atom in the terminal PEN residue; *CH2N*: – C atom in the terminal PEN residue; and *HN2N*: – H atom bonded to NH2N.

To ensure consistency with our previously parametrized FF for *L*PEI, we kept the already optimized parameters unchanged, and only optimized the ones involving the new residue types, PEY and PEN. Also for reasons of consistency, we followed the workflow implemented in ffTK [21] to derive the parameters for *B*PEI.

As per ffTK convention, we adopted the LJ parameters by analogy with CHARMM36. The atomic charges were optimized according to QM water interaction profiles of *B*PEI and single H_2O molecules (at HF/6-31G(d) level). The bond- and angle parameters were derived from the fitting of MM distortion energy to the QM target profile. Lastly, the dihedral parameters were extracted based on torsion energy scans, by fitting the MM potential energy surface (PES) to its QM counterpart (performed at MP2/6-31G(d) level).

3.3.1 Lennard-Jones parameters

The NC3 atom type in the branch connector was modeled as neutral trimethylamine nitrogen, and the ε and R_{min} values were adopted from the NG301 type defined in CHARMM36. For the CH2 and HC2 atom types in the PEY residue, we adopted the same LJ parameters as for the equivalents from the PEI residue, however, optimizing new partial charges. The NH2N and HN2N types from the PEN ending residue were modeled as neutral methylamine nitrogen and hydrogen, respectively, being identified with the NG321 and HGPAM2 types in CHARMM36. CH2N was modeled as the CH2 atom type, maintaining the ε and R_{min} values of the CG321 standard type.

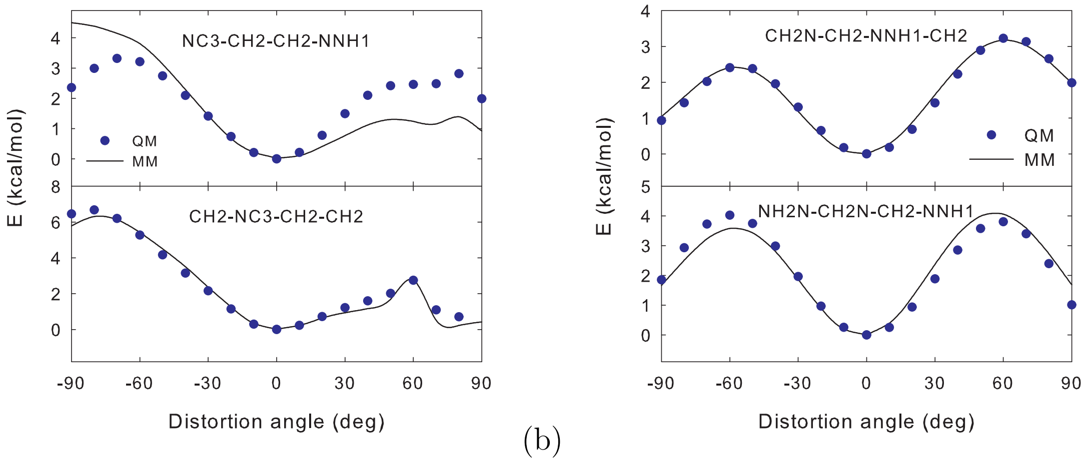
3.3.2 Partial atomic charges

The partial atomic charges were determined from QM interaction profiles for complexes formed of either *B*PEI1p0 or *L*PEI1p0, and water molecules located in typical positions relative to the model molecule. To facilitate representative positions and orientations of the water molecules, the water-accessible atoms were categorized based on their hydrogen-bonding abilities into two major classes: *acceptors* – the nitrogen atoms, and *donors* – the hydrogen atoms in the model compounds. For each water-accessible atom, a polymer–(single-H_2O-molecule) complex was assembled, and the distance and rotation angle relative to the model molecule were optimized quantum mechanically, while keeping the other degrees of freedom fixed. We imposed neutrality for the *B*PEI1p0 and *L*PEI1p0 model polymers, thereby, just the charges for the CH2 and NC3 atom types from PEY were included in the optimization process, along with those for CH2N, NH2N and HN2N from PEN.

3.3.3 Bond and angle parameters

Since the Hessian matrices reflect the curvature of the PES, ffTK employs them to describe energy variations produced by minor distortions about the equilibrium configuration. Explicitly, for each internal coordinate, a minor distortion is produced, and the resulting QM energy variation is computed using the Hessian matrix. The MM energy change, on the other hand, is calculated simply by employing the trial bond and angle parameters and computing the difference between total energies characterizing the distorted and undistorted configurations. The difference between the MM and QM energy is minimized iteratively until acceptable agreement is achieved.

The differences between the MM and QM optimized bonds and, respectively, angles in the initial and final optimization steps decrease considerably during the adjustment



(a) (b)

Figure 3.3: Torsion energy profiles for the dihedrals formed by the backbone atoms of (a) *B*PEI1p0 and (b) *L*PEI1p0-*N*.

procedure, which is reflected by a ∼ 60% drop in the objective function value. The respective differences at the end of the optimization process amount to $\sim 1.1 \times 10^{-2}$ Å for bonds, and, respectively ∼ 2.3° for angles, demonstrating the very good match of the MM and QM data. The MM-QM differences are below 3×10^{-2} Å and 3°, respectively for bonds and angles, as recommended by the authors of the original CHARMM FF [23], indicating a reasonable agreement.

As a comparison with earlier studies, our force constant for the C–N bond in the branch connector residue (namely CH2–NC3) is by ∼ 2.4% lower than the value reported by Mintis et al. [19], who adopted the bonded parameters for *B*PEI from the General Amber force field [25]. Notably, their bonded parameters were not *specifically* parametrized for *B*PEI, and were assigned by similarity with comparable structures.

3.3.4 Dihedral parameters

Each backbone dihedral was distorted from the equilibrium geometry, and QM torsion scans were performed, which were used as target data for the fitting of the MM profiles. The scans comprised a series of geometry optimizations (at MP2/6-31G(d) level), the model compound being allowed to relax freely while the dihedral of interest was constrained. The multiplicities were assigned integer values (1, 2, or 3) based on local symmetry considerations, and the phases were allowed to take only the values 0° or 180°, to conform to the CHARMM prescription.

The QM and MM torsion energy profiles for the backbone dihedrals involving the tertiary nitrogen, NC3, and, alternatively, the CH2N and NH2N atoms in the NH_2

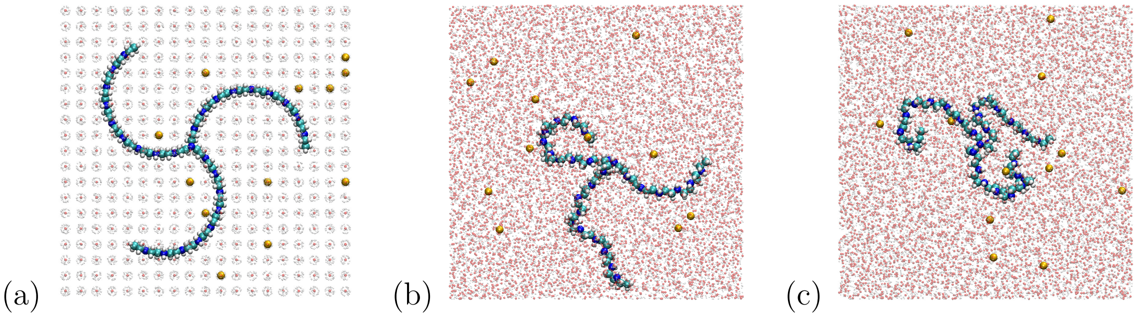


Figure 3.4: Snapshots of the solvated *B*PEI9u2 system: (a) initial configuration, (b) after equilibration, and (c) at the end of a 10 ns run.

ending residue, are illustrated in Figure 3.3 (a) and (b), respectively. The correspondence between the MM and QM torsion energy profiles for NC3–CH2–CH2–NNH1 (Figure 3.3a, top panel) is very good around the central minimum. However, the differences at high distortions, are a consequence of the connectivity of NC3 with three branches, which mixes this dihedral with other internal coordinates, and results in a more restricted motion of the tertiary nitrogen. For the rest of the dihedrals, there is a very good agreement between the MM and QM data.

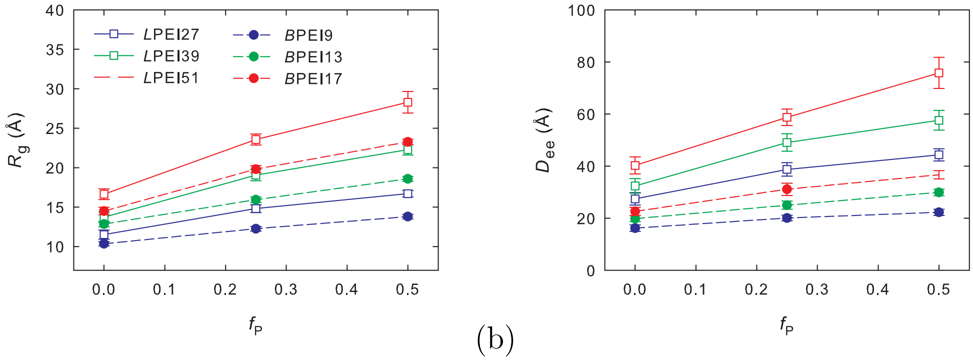
Our parametrization of the dihedral angles stands out from previous works [17; 19] by the very good match between the MM and QM torsion profiles, while still considering a single multiplicity for each dihedral.

3.4 All-atom simulations of linear/branched polyethyleneimine

3.4.1 Simulation methodology

The developed force field was employed in MD simulations of solvated linear- and branched PEIs. We considered four uniform protonation fractions (f_P), namely: 0 (non-protonated), and $1/n$ (1-in-n protonated units), with n taking the values 2, 3, and 4. The polymer sizes were chosen so that all chains perfectly accommodate all considered protonation patterns. As a general rule, the chains were composed of $12N+3$ monomers, enclosed by ending methyl groups (CH_3 units). We thus considered three chain lengths, namely *L*PEIs composed of 27, 39 and 51 monomers, and *B*PEIs with equal branches and molecular weights (M_W) as close as possible to their linear equivalents.

We solvated the systems in TIP3P water [26] with the molecules oriented randomly and positioned on a regular grid, with a view to rigorously control the density (1 g/cm^3), and ensure the randomness required for each initial configuration to achieve a proper statistical ensemble. We considered coiled and “triskele” arrangements as initial



(a) (b)

Figure 3.5: Protonation dependence of the radius of gyration (a) and end-to-end distance (b) for the branched PEIs (dashed lines), along with the profiles for linear PEIs (continuous lines).

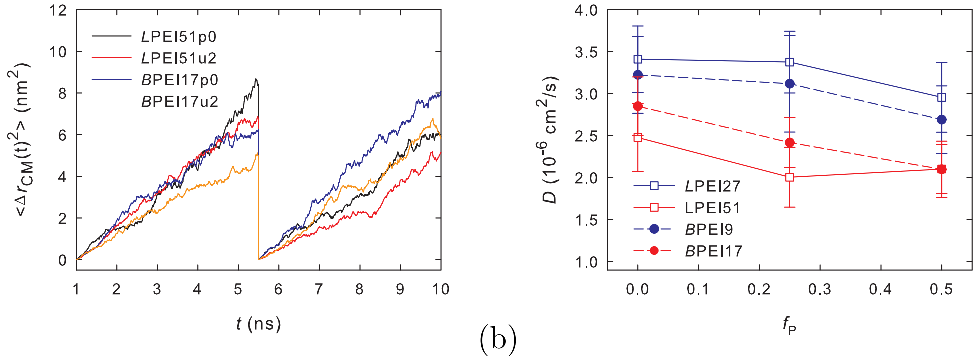
conformations for *L*- and *B*PEI, and inserted a number of Cl^- ions, equal to the number of protonated sites, in random positions to neutralize the systems. Figures 3.4 a, b, and c illustrate the initial configuration, evidencing the grid water and neutralizing counterions, the configuration after equilibration, and, respectively, a snapshot of the final conformation, for selected systems involving *B*PEI9u2.

For the actual analysis, the first nanosecond of each simulation was discarded to further ensure properly equilibrated configurations, and neglect the initial transitory behavior. The statistical errors were calculated after performing the ensemble averages, as standard deviations of the time-dependent values from the respective time-averaged quantities.

We used as MD integrator the velocity Verlet algorithm implemented in NAMD, and the “leap-frog” method in Gromacs, with a 2 fs time step. A 12 Å cutoff was used for the short-ranged interactions, and periodic boundary conditions (PBC) were applied in all cartesian directions. The electrostatics was treated with the Particle-Mesh-Ewald (PME) method using a mesh spacing of 1 Å. The temperature was fixed at 310 K, while the pressure was maintained at 1 atm.

3.4.2 Radius of gyration. End-to-end distance

The structural features of the simulated linear- and branched PEIs were investigated by means of radii of gyration (R_g), and end-to-end distances (D_{ee}). In polymer science, the radius of gyration is used to describe the spatial extents of the polymers, and is calculated as:



(a) (b)

Figure 3.6: (a) Time dependence of the mean squared displacement of the molecule's center of mass relative to a reference position, and (b) protonation dependence of the diffusion coefficient, for the non-protonated and alternatively protonated LPEI51 and BPEI17.

$$R_g = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_i - r_{CM})^2} \tag{3.5}$$

where N is the total number of polymer atoms, r_i is the position of atom i, and r_{CM} is the position of the center of mass (CM) of the polymer.

For *linear* PEIs, the end-to-end distance was simply calculated as the average distance between the CMs of the ending methyl groups (PEC–PEC distance). For the branched polymers, on the other hand, D_{ee} was computed as the average distance between the CMs of the three pairs of PEC residues correspondingly ending each of the three branches.

Figure 3.5 illustrates the dependences of the ensemble- and time-averaged R_g and D_{ee} of L- and BPEIs on the protonation fraction. A steady increase with the protonation fraction is evident both for R_g, and D_{ee}, owed to the electrostatic repulsion between (an increasing number of) positively charged (protonated) units. In addition, a shift to higher values can be observed for the profiles corresponding to heavier PEIs. The slopes of the R_g, and D_{ee} profiles appear to also increase with the molecule size, with the highest slope featured by PEI51. We found that BPEIs consistently exhibit lower values than LPEIs both in terms of R_g and D_{ee}, for all molecular weights and protonations considered.

3.4.3 Diffusion coefficient

With a view to analyzing the dynamical properties of the polymers, we calculated the diffusion coefficient employing Einstein's formula:

$$D = \lim_{t\to\infty} \frac{1}{6t} \left\langle \Delta\mathbf{r}_{\mathrm{CM}}(t)^2 \right\rangle \tag{3.6}$$

where $\langle \Delta\mathbf{r}_{\mathrm{CM}}(t)^2 \rangle$ is the mean squared displacement (MSD) of the CM of the molecule relative to a reference position, averaged over time and ensemble.

The time dependence of the MSD is shown in Figure 3.6 (a), comparatively for *L*PEI51 and *B*PEI17. As a validation, on both considered threads, the MSD exhibits a quasi-linear increase in time, while D fluctuates around a stable value. Higher slopes in the MSDs are evidenced by the non-protonated PEIs, indicating their increased mobilities, as compared to alternatively protonated polymers. The average slopes appear to increase with the decrease in the polymer's size, which is an intuitive result, since more compact objects, that also have lower molecular weights, present enhanced mobility.

As a general finding, the diffusion coefficient (Figure 3.6b) shows a downward trend both with increasing protonation, and with M_W. Evidently, the increase in protonation determines the expansion of the polymers, resulting in the decrease of mobility. The *B*PEIs with the highest M_W exhibit higher diffusion coefficients than the *L*PEI equivalents, which is in line with the lower radii of gyration found for the branched configurations. Conversely, for the PEIs with the lowest mass, the diffusion coefficient presents higher values for *L*PEIs, which can be linked to the increased rigidity of *B*PEI9 (owed to its short branches).

3.5 All-atom simulations of DNA/(linear-/branched-) polyethyleneimine

We employed the newly developed force field for *L*- and *B*PEI, along with the CHARMM parameters for DNA from Ref. 27, in MD simulations of a Drew-Dickerson dodecamer (DDD), d[CGCGAATTCGCG]$_2$, interacting with either linear, or branched PEIs. The polymers were arranged circularly around the DNA helix at 20 Å from its axis. The simulated DNA-PEI systems were neutralized with an appropriate number of Cl$^-$ counterions, and were immersed in rectangular boxes filled with TIP3P water. The simulations were carried out at physiological temperature and pressure conditions, namely, the temperature was set to 310 K, and the pressure was fixed at a value of 1 bar.

For the systems composed of DNA and 4 *B*PEI5u2 polymers, Figure 3.7 illustrates top and side views of the initial configuration, along with snapshots after $\sim$ 30 ns. Visually inspecting the simulations, we found that *B*PEIs present a higher affinity in attaching their branches to the minor groove of DNA, while for *L*PEIs, structured

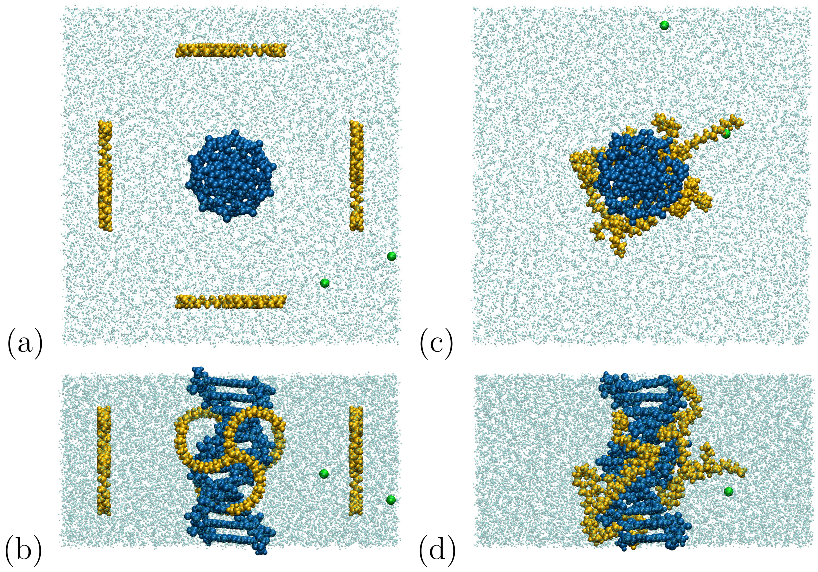


Figure 3.7: Initial- (a, b) and intermediate configuration (c, d) of the system composed of DNA and 4 *B*PEI9u2 molecules – top and side views (upper/lower row).

arrangements relative to DNA's helix are less evident.

To provide a picture on the variation of free energy in dependence on a reaction coordinate, for example the interdistance between atoms, we calculated the potential of mean force (PMF):

$$\mathrm{PMF} = -k_{\mathrm{B}}T \log P_{\text{P-N}} + const. \tag{3.7}$$

where k_{B} is Boltzmann's constant, T is the temperature, and $P_{\text{P-N}}$ is the normalized radial probability distribution of the distances between the phosphates (P) of DNA and nitrogens (N) in PEI. The value of the constant is chosen in such a way that the minimum of the PMF corresponds to zero.

The PMF profiles for the DNA–*L*PEI9u2/*B*PEI5u2 systems present a major minimum at ~ 4 Å, revealing similar P-N distances in the attached states. We found that the studied linear polymer is able to form complexes with DNA at intermediate P-N distances (larger than the average distance at which the PMF exhibits the major minimum), while the branched variant appears to form DNA-PEI polyplexes (slightly) more favorably.

4

Coarse-grained modeling and simulations of polyethyleneimine

4.1 Introduction

The atomistic computational studies mentioned in the previous chapters presented major limitations in the case of complex biomolecular systems, such as those composed of DNAs and delivery vectors, since simulating a massive number of atoms (millions) becomes a laborious task, even for high-performance computers. Given that the interest in efficient gene delivery protocols is continuously rising, a method that allows for much larger systems, and also longer time spans, to be simulated efficiently, was necessary. Without notable loss of accuracy, coarse-grained (CG) methods rely on mapping multiple atoms into a single interaction site, and thus the system size is significantly reduced. The lower computational costs facilitate the achievement of considerably increased simulation times, as the reduced number of interaction sites implies stability over larger time steps. Martini [28–30] is a widely popular CG force field model, used as a standard modeling choice in the present study, which is based on mapping, on average, four atoms into a single CG *bead*.

4.2 Development of a MARTINI force field for linear polyethyleneimine

We mapped entire *L*PEI residues into CG interaction sites, as seen in Figure 4.1, and identified the positions of the beads with the centers of mass (CM) of the respective *L*PEI monomers: *PEI*: non-protonated bead mapping the $CH_2{-}NH{-}CH_2$ monomer; *PEP*: protonated bead mapping the $CH_2{-}NH_2{}^+{-}CH_2$ monomer; *PEC*: chain ending bead mapping the CH_3 (methyl) group.

The Martini potential energy function has the following form:

$$\begin{aligned}U_{bonded} =& \frac{1}{2}\sum_{bonds} K_b(b-b_0)^2 + \frac{1}{2}\sum_{angles} K_\theta(\theta-\theta_0)^2 + \\ & \sum_{dihedrals} K_\psi[1+cos(n\psi-\psi_0)]\end{aligned} \tag{4.1}$$

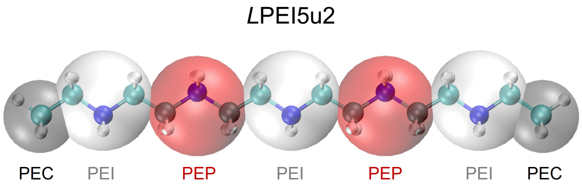


Figure 4.1: Coarse-grained mapping of *L*PEI5u2, evidencing the defined beads.

$$U_{non-bonded} = \sum_{beads\ i,j} \left\{ \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} + 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] \right\} \tag{4.2}$$

where the bonded interactions depend on the force constants, K_b, K_θ and K_ψ, and on the equilibrium bonds and angles, b_0 and θ_0. The dihedral potential is dependent on the multiplicity, n, and the reference phase, ψ_0. The non-bonded terms include electrostatic- and Lennard-Jones (LJ) interactions, the latter being characterized by the potential well depth, ε_{ij}, and the zero-potential distance, σ_{ij}.

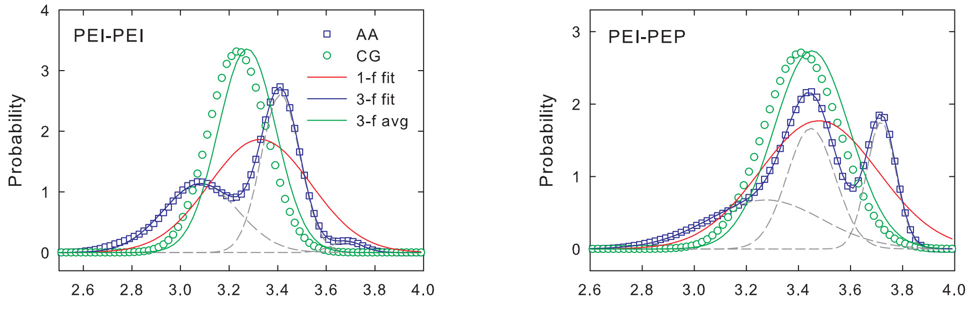
From the AA simulations, we computed probability distributions for bonds, angles, and dihedrals formed by the CMs of neighboring residues. For the CG bonded terms, using single- and multi-function fits, we extracted the CG force constants, equilibrium values (bonds, angles), multiplicities, and phases (the last two for the dihedral terms) by Boltzmann inversion [31] on the reference AA distributions.

The non-bonded terms were assigned the parameters for standard types in Martini, by performing trial CG simulations for various combinations, and verifying the agreement between the CG radii of gyration (R_g) and end-to-end distances (D_{ee}) and their AA equivalents. The match between the CG and AA data was quantized by calculating the root-mean-square deviations (RMSD) of the respective quantities.

4.2.1 Bonded parameters via Boltzmann inversion of all-atom distributions

The CG parameters were extracted using the Boltzmann inversion method, directly from the AA reference distributions. The probability distributions extracted from AA simulations present multiple peaks, which is undesirable for directly employing the Boltzmann inversion procedure. For the PEI–PEI bond depicted in Figure 4.2 (a), the probability distribution is obviously determined by two distinct Gaussian functions, while the PEI–PEP distribution in Figure 4.2 (b) can be constructed from three independent functions, from which neither can be considered negligible.

We thus considered three-Gaussian functions (3-f):



(a) (b)

Figure 4.2: Probability distributions for the (a) PEI-PEI, and (b) PEI-PEP distances. The AA-, and CG simulation results (SNda-Qd-3-3-P model) are represented with blue squares and green circles. The 1 Gaussian fit, 3-f fit, and 3-f average are depicted with red, blue, and green lines, respectively.

$$P = \sum_{i=1}^{3} A_i e^{-U_i/k_B T}, \quad U_{i,b} = \frac{1}{2} K_{b,i}(b - b_{0,i})^2, \quad U_{i,\theta} = \frac{1}{2} K_{\theta,i}(\theta - \theta_{0,i})^2 \tag{4.3}$$

The 3-f individual peaks are represented in Figure 4.2 with dashed grey lines, their distinctive maximum positions and peak heights clearly confirming the need for a multi-peak fitting function.

Even by considering 3-f functions to fit the AA probability distributions, the issue of single force constants for each bond- and angle type (specified by the employed analytical potential functions) in conventional MD codes persists. We considered the inverse of the *unique* force constant to be equal to the weighted average of the inverses of individual K_b values (from the 3-f fits), as follows:

$$\frac{1}{K_b} = \sum_{i=1}^{3} \frac{w_i}{K_{b,i}}, \quad \frac{1}{K_\theta} = \sum_{i=1}^{3} \frac{w_i}{K_{\theta,i}} \tag{4.4}$$

while the equilibrium values for respective distances and angles were straightforwardly calculated as the weighted averages of the individual values:

$$b = \sum_{i=1}^{3} w_i b_{0,i}, \quad \theta = \sum_{i=1}^{3} w_i \theta_{0,i}; \quad w_i = \frac{S_i}{\sum_{i=1}^{3} S_i} \tag{4.5}$$

The AA probability distributions for the dihedrals (ψ) formed by four consecutive beads revealed qualitatively different outlines than the case of bonds and angles. A new scheme was devised based on the form of the AA profiles, and on the fact that widely used MD programs are able to model torsions with multiple functions. Specifically, we used four functions (4-f) to fit the AA probability distributions:

$$P_\psi = Ae^{-U_\psi/k_BT}, \quad U_\psi = \sum_{i=1}^{4} K_{\psi,i}\left[1 + cos(n_i\psi - \psi_{0,i})\right] \tag{4.6}$$

where the multiplicity n_i takes the values 1, 2, 3, and 4, while the phase $\psi_{0,i}$ is either 0°, or 180°.

4.3 Coarse-grained simulations of linear polyethyleneimine

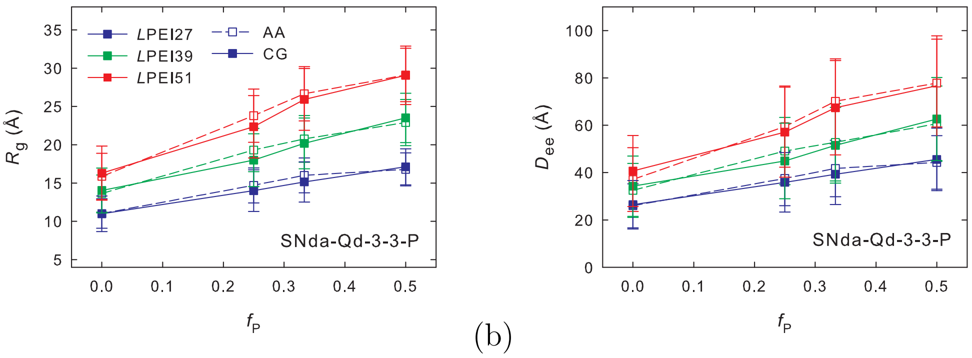
We performed coarse-grained MD simulations of *L*PEI chains solvated in *standard* or alternatively, *polarizable* Martini water. The sizes and protonations of the employed polymers match the ones considered in the AA study of *L*PEI, namely: the chains comprised 27-, 39-, and 51 monomers, and the protonation fractions (f_P) were equal to 0, 1/4, 1/3, and 1/2. The solvated *L*PEI systems were neutralized by adding a number of Cl^- ions (modeled by Qa particles) equal to the number of PEP beads.

The electrostatics was treated with the reaction field method, employing a dielectric constant (ε_r) of 2.5 in the case of solvating the systems in polarizable water, or, alternatively an increased value of 15 in conjunction with non-polarizable water. We imposed the same temperature and pressure conditions as the ones used for the AA simulations, namely 310 K, and 1 bar. To this end, we used the velocity rescaling [32] thermostat with a 1 ps coupling constant, and the Parinello-Rahman barostat [13; 14] with a 12 Å coupling constant.

4.3.1 Selecting suitable Martini types

In order to develop an additive CG FF compatible with Martini, the bead types need to be identified with standard types in Martini so as to be assigned compatible non-bonded parameters. To this end, we selected various combinations from the standard interaction sites for our PEI, PEP, and PEC beads. Concretely, for the PEI bead, we considered three standard nonpolar types, specifically Nda, Nd, and N0 (as well as their “S” and “T” variants), and one polar type, P2, suggested by Mahajan and Tang [33]. The protonated PEP bead was modeled as a charged particle ($+1e$) of Qda, Qd, or Q0 type (along with their “S” and “T” options), which also indicates the hydrogen donor nature of this bead. For comparison, we also considered in our test simulations, for the PEI–PEP combination, the pair of standard types selected by Wei and Lujten [18], namely SNda–SQd, and the one of Mahajan and Tang, specifically P2–Qd.

In performing the bulk CG simulations, we considered the bonded parameters extracted using 3-*f* functions for the bonds, both schemes (1-*f* and 3-*f*) for the angles,



(a) (b)

Figure 4.3: Radius of gyration (a) and end-to-end distance (b) for *L*PEI using polarizable water and the SNda-Qd-3-3-P FF model. AA (CG) values are represented with empty (filled) squares and are connected with dashed (continuous) lines.

and 4-f fits for the dihedrals. In the following sections, we considered the “3-3” bonded model (meaning the modeling with 3-f funtions for bonds and angles) since it presents, in fact, the lowest RMSD with respect to the atomistic values of R_g and D_{ee}, validating this scheme for simulations employing polarizable water.

4.3.2 Optimal force field model with polarizable/standard Martini water

Our findings suggest that the PEI particle needs to be modeled as a small neutral Martini bead-type, which is consistent with its definition as a bead mapping only three heavy atoms, in contrast to the standard 4-to-1 mapping. The agreement of the AA and CG structural quantities can be readily assessed from Figure 4.3, which display the R_g and D_{ee} values for the SNda-Qd-3-3-P model (SNda - standard type selected for the PEI bead; Qd - bead type selected for the PEP bead).

By calculating the root-mean-squared deviation between the AA and CG values for R_g and D_{ee}, we determined that the most suitable PEI–PEP combination for polarizable water to be SNda–Qd. Notably, for the model used in conjunction with standard water, it is recommended that the bead-type combination be SN0–SQ0, and that the polymers be weakly protonated (meaning the protonation fraction is smaller than 0.25).

We computed the diffusion coefficients for the studied CG polymers using Einstein’s formula (also employed in the AA analysis) and the averaged values are presented in Figure 4.4 along with the ones obtained from atomistic simulations. The diffusion coefficient is known to be a numerically sensitive quantity, so we obtained only a semi-quantitative agreement between the AA and CG data. Nevertheless, the overall decreasing trend for higher protonations is preserved. In addition, we found

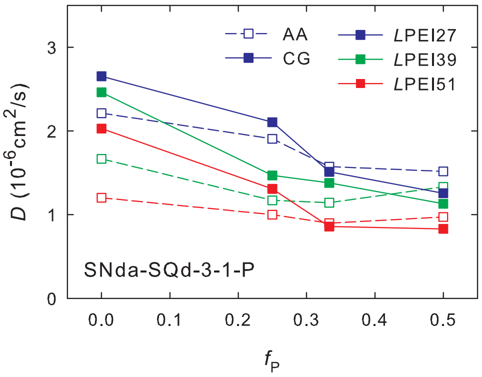


Figure 4.4: Diffusion coefficient for *L*PEI using polarizable water and the SNda-SQd-3-1-P FF model. AA (CG) values are represented with empty (filled) squares and are connected with dashed (continuous) lines.

that the CG diffusion coefficients for weakly protonated chains show an overestimation tendency, while for the strongly protonated polymers, the CG values appear to be lower than the AA equivalents.

4.4 Coarse-grained simulations of DNA and polyethyleneimine

We employed the new FF in conjunction with the Martini FF for nucleic acids in CG simulations of DNA-PEI polyplex formation. We used the newly developed SNda-SQd-3-1-P FF for PEI, and the Martini FF for DNA developed by Uusitalo et al. [30], in conjunction with the polarizable water FF of Yesylevskyy et al. [34].

The systems comprised a (single or double) Drew-Dickerson dodecamer (DDD, with the nucleotide sequence d[CGCGAATTCGCG]$_2$) and a selected number of *L*PEI15 chains (4 or 8, respectively, having either 1/4, or 1/2 protonation fractions) positioned circularly around the DNA molecule at a distance of 25 Å from its axis. The systems were enclosed in cubic boxes, and subjected to periodic boundary conditions in all directions. The DNA molecule was constrained along the z-axis, and we added a suitable number of Na$^+$ or Cl$^-$ ions to attain system neutrality.

The initial- and intermediate configurations for a system composed of a single DDD and four *L*PEI15u2 polymers are depicted in Figure 4.5. The uniform attachment of *L*PEI to DNA's minor and major grooves is evident in Figure 4.5 (d).

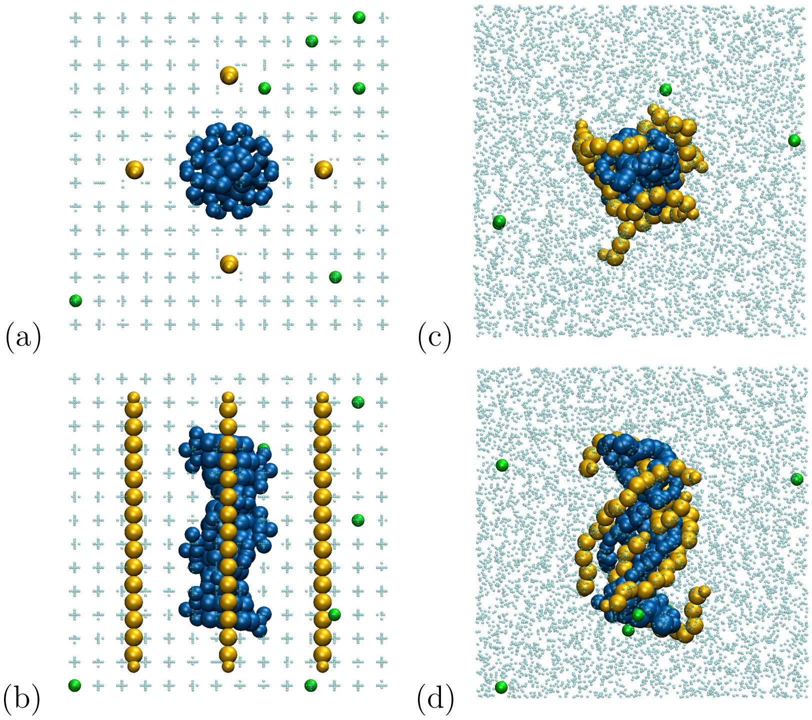


Figure 4.5: Initial- (a, b) and intermediate configuration (c, d) of the coarse-grained system composed of DNA and 4 *L*PEI15u2 molecules – top and side views (upper/lower row).

4.4.1 Potential of mean-force

To quantify the formation of DNA-*L*PEI complexes, we employed the PMF between the phosphate groups of DNA (mapped into BB1 beads) and the protonated PEP sites of *L*PEI:

$$\text{PMF} = -k_{\text{B}}T \log P_{\text{BB1-PEP}} + const. \tag{4.7}$$

where $P_{\text{BB1-PEP}}$ is the normalized probability distribution of the radial distances between BB1 and PEP.

The PMFs for the BB1–PEP interaction for the systems composed of a double DDD and either 8 *L*PEI15u4, or 8 *L*PEI15u2 molecules reveal the main minimum is obtained at ~ 4.6 Å, which is higher than the PMF minimum obtained from the AA simulations. This discrepancy can be linked to the coarser representation in the case of CG modeling, along with the differences in the sizes of the simulated molecules.

5

Conclusions

With a view to studying the formation and condensation of complexes involving DNA and PEI, which is known as an efficient gene delivery vector, and given the absence of reliable force fields for PEI in the literature, the present thesis focused on deriving high-quality parameters for this particular cationic polymer.

By following a bottom-up approach, we developed an all-atom FF based on high-quality ab initio data, which was employed in molecular dynamics simulations of solvated PEI polymers. Subsequently, by extracting probability distributions for bonds, angles, and dihedrals formed by adjacent residues (mapped as beads), we parametrized a coarse-grained FF for PEI, which was employed in CG simulations of solvated PEI- and DNA-PEI systems.

Modeling PEI at the atomistic scale involved deriving all the force constants and equilibrium values for bonds, angles, and dihedrals formed by consecutive atoms, as well as the partial atomic charges and the LJ parameters, needed to characterize the intramolecular forces acting up in molecules of arbitrary sizes and protonation patterns. The AA force field that we developed complies with the CHARMM standard, which includes parameters for a multitude of biological molecules.

The high quality of the present parametrization stems from consistently deriving the partial atomic charges, along with the whole set of bonded parameters, which is in itself a novelty compared to earlier studies. The comprehensive FF validation involved performing extensive AA molecular dynamics simulations, and comparing the results with experimental data. Notably, we found gyration radii, end-to-end distances, and diffusion coefficients in very good agreement with experimental counterparts for PEIs of similar molecular weights and protonation fractions.

To put the developed AA FF to the test, we carried out simulations of DNA-PEI complex formation, which showed the most probable attachment distances in terms of PEI branching. In particular, branched PEI appears to exhibit a higher attachment affinity to the minor grove of DNA, and a more efficient complex formation than linear chains.

The new AA FF for PEI [35; 36], is suitable for simulations involving PEI polymers of arbitrary branching, protonation, and molecular weight, representing an indispensable tool in future computational studies, which are expected to provide a complete characterization of PEI. The new parameters can be also employed in conjunction with a wide variety of molecules already included in the CHARMM force field, opening the way to a broad range of studies, which are not restricted necessarily to the field of gene delivery.

By mapping the atomistic residues into coarse-grained beads, we extracted probability distributions for bond lengths, angles, and dihedrals formed by consecutive particles from the AA simulations, and developed a CG FF for linear PEI[37; 38]. We selected for our beads suitable types from the Martini standard interaction sites, so the new CG FF conforms with Martini, which is one of the most extensively used CG models.

The high quality of our CG model for PEI is manifestly evidenced by the very good match between the CG and AA structural properties, and consequently, by the agreement with experimental measurements. The FF was employed in trial simulations of DNA-PEI polyplex formation, providing qualitative results, in line with experimental observations.

The derived CG parameters can be used, jointly with the Martini FF for DNA, in extensive simulations of DNA-PEI condensation, which requires simulating systems of millions of atoms on the microsecond time scale to achieve relevant phenomenological results. Such simulations could provide meaningful knowledge in terms of the ratio between DNA phosphates and PEI protonated sites, salinity, pH, and temperature of the solutions, of great importance for the development of efficient and non-toxic gene delivery protocols.

List of publications

Thesis-related publications

1. **A.-E. Terteci-Popescu**, T. A. Beu, “Branched Polyethyleneimine: CHARMM Force Field and Molecular Dynamics Simulations”, Journal of Computational Chemistry, 43 (31), 2072-2083, DOI: 10.1002/jcc.27005 (2022); IF: 3.672; AIS: 0.865.

2. T. A. Beu, **A.-E. Ailenei**, R.-I. Costinaş, “Martini Force Field for Protonated Polyethyleneimine”, Journal of Computational Chemistry, 41 (4), 349-361, DOI: 10.1002/jcc.26110 (2020); IF: 3.376; AIS: 0.923.

3. T. A. Beu, **A.-E. Ailenei**, A. Farcaş, “Atomistic and coarse-grained modeling of polyethyleneimine”, Chemical Physics Letters, 714, 94-98, DOI: 10.1016/j.cplett.2018.10.071 (2019); IF: 2.029; AIS: 0.367.

4. T. A. Beu, **A.-E. Ailenei**, A. Farcaş, “CHARMM force field for protonated polyethyleneimine”, Journal of Computational Chemistry, 39 (31), 2564-2575, DOI: 10.1002/jcc.25637 (2018); IF: 3.194; AIS: 1.141.

   TOTAL AIS: 3.296

Other publications

1. **A.-E. Ailenei**, T. A. Beu, “Ion transport through gated carbon nanotubes: Molecular dynamics simulations using polarizable water”, Journal of Molecular Structure, 1245, 131022, DOI: 10.1016/j.molstruc.2021.131022 (2021); IF: 3.196.

Conference contributions

1. **A.-E. Ailenei**, R.-I. Costinaş, T. A. Beu, “Polyethyleneimine: Coarse-grained modeling and simulation”, 12th International Conference on Processes in Isotopes and Molecules (PIM 2019), 25-27 September 2019, Cluj-Napoca, Romania, poster presentation.

2. **A.-E. Ailenei**, R.-I. Costinaş, T. A. Beu, “Coarse-grained simulations of solvated polyethyleneimine”, TIM 19 Physics Conference, 29-31 May 2019, Timişoara, Romania, poster presentation.

References

[1] C. A. M. Clévio Nóbrega, Liliana Mendonça, *A Handbook of Gene and Cell Therapy* (Springer, 2020), 1st ed.

[2] R. Narain, ed., *Polymers and Nanomaterials for Gene Therapy* (Elsevier, Woodhead Publishing, Cambridge, 2016).

[3] F. Alexis, E. Pridgen, L. K. Molnar, and O. C. Farokhzad, Molecular Pharmaceutics **5**, 505 (2008).

[4] M. Allen and D. Tildesley, *Computer Simulation of Liquids*, Oxford Science Publ (Clarendon Press, 1989).

[5] M. P. Allen, *Introduction to molecular dynamics simulation* (2004), URL `https://udel.edu/~arthij/MD.pdf`.

[6] R. Hockney and J. Eastwood, *Computer Simulation Using Particles* (CRC Press., 1988), 1st ed.

[7] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, The Journal of Chemical Physics **76**, 637 (1982).

[8] T. A. Beu, *Basic molecular dynamics* (2021), URL `http://www.phys.ubbcluj.ro/~titus.beu/MDsquad/chapters/C3%20Basic%20MD.pdf`.

[9] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, SoftwareX **1-2**, 19 (2015), ISSN 2352-7110, URL `http://www.sciencedirect.com/science/article/pii/S2352711015000059`.

[10] M. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and the GROMACS development team (2016), URL `www.gromacs.org`.

[11] A. Leach, *Molecular Modelling: Principles and Applications* (Pearson Education, 2001).

[12] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, The Journal of Chemical Physics **81**, 3684 (1984).

[13] M. Parrinello and A. Rahman, Journal of Applied Physics **52**, 7182 (1981).

[14] S. Nosé and M. Klein, Molecular Physics **50**, 1055 (1983).

[15] D. C. Rapaport, *The Art of Molecular Dynamics Simulation* (Cambridge University Press, 2004), 2nd ed.

[16] R. J. Sadus, *Molecular Simulation of Fluids: Theory, Algorithms and Object-orientation*, v. 1 (Elsevier, 1999).

[17] C. Sun, T. Tang, H. Uludağ, and J. E. Cuervo, Biophysical J. **100**, 2754 (2011).

[18] Z. Wei and E. Luijten, J. Chem. Phys. **143**, 243146 (2015).

[19] D. G. Mintis, T. S. Alexiou, and V. G. Mavrantzas, The Journal of Physical

Chemistry B **124**, 6154 (2020).

[20] T. A. Beu and A. Farcaş, J. Comput. Chem. **38**, 2335 (2017).

[21] C. G. Mayne, J. Saam, K. Schulten, E. Tajkhorshid, and J. C. Gumbart, J. Comput. Chem. **34**, 2757 (2013).

[22] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, and M. A. R. et. al., *Gaussian 09 Revision A.01*, gaussian Inc. Wallingford CT 2009.

[23] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, et al., J. Comput. Chem. **31**, 671 (2010).

[24] S. Basu, R. M. Venable, B. Rice, E. Ogharandukun, J. B. Klauda, R. W. Pastor, and P. L. Chandran, Macromolecular Chemistry and Physics **220**, 1800423 (2019).

[25] J. Wang, P. Cieplak, and P. A. Kollman, J. Comput. Chem. **25**, 1049 (2000).

[26] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).

[27] K. Hart, N. Foloppe, C. M. Baker, E. J. Denning, L. Nilsson, and A. D. MacKerell, Journal of Chemical Theory and Computation **8**, 348 (2011).

[28] S. J. Marrink, A. H. de Vries, and A. E. Mark, The Journal of Physical Chemistry B **108**, 750 (2004).

[29] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, The Journal of Physical Chemistry B **111**, 7812 (2007).

[30] J. J. Uusitalo, H. I. Ingólfsson, P. Akhshi, D. P. Tieleman, and S. J. Marrink, Journal of Chemical Theory and Computation **11**, 3932 (2015), pMID: 26574472.

[31] D. Reith, M. Pütz, and F. Müller-Plathe, Journal of Computational Chemistry **24**, 1624 (2003).

[32] G. Bussi, D. Donadio, and M. Parrinello, The Journal of Chemical Physics **126**, 014101 (2007).

[33] S. Mahajan and T. Tang, Journal of Computational Chemistry **40**, 607 (2019).

[34] S. O. Yesylevskyy, L. V. Schäfer, D. Sengupta, and S. J. Marrink, PLoS Comput Biol **6**, e1000810 (2010).

[35] T. A. Beu, A.-E. Ailenei, and A. Farcaş, Journal of Computational Chemistry **39**, 2564 (2018).

[36] A.-E. Terteci-Popescu and T. A. Beu, Journal of Computational Chemistry **43**, 2072 (2022).

[37] T. A. Beu, A.-E. Ailenei, and A. Farcaş, Chemical Physics Letters **714**, 94 (2019).

[38] T. A. Beu, A.-E. Ailenei, and R.-I. Costinaş, Journal of Computational Chemistry **41**, 349 (2020).