

# MARTINI Coarse-Grained Model for Polyethylenimine

*Subhamoy Mahajan and Tian Tang\**

Department of Mechanical Engineering, University of Alberta, Edmonton, AB, Canada

## ABSTRACT

As a polycation with diverse applications in biomedical and environmental engineering, polyethylenimine (PEI) can be synthesized with varying degrees of branching, polymerization and can exist in different protonation states. There have been some interests in molecular modeling of PEI at all-atom or coarse-grained (CG) levels, but present CG models are limited to linear PEIs. Here we present the methodology to systematically categorize bond lengths, bond angles and dihedral angles, which allows us to model branched PEIs. The CG model was developed under the Martini scheme based on eight ~600 Da PEIs, with four different degree of branching at two different protonation states. Comparison of the CG model with all-atom simulations shows good agreement for both local (distributions for bonded interactions) and global (end-to-end distance, radius of gyration) properties, with and without salt. Compatibility of the PEI model with other CG bio-molecules developed under the Martini scheme will allow for large-scale simulations of many PEI-enabled processes.

## 1. INTRODUCTION

Polyethylenimine (PEI) is a polymer of aziridine, which can be synthesized as linear [1] or branched [2] structures with varying degrees of polymerization. Because of its polycationic nature in acidic environment, PEI has diverse applications in filtration and has been used to floc humic acid [3], remove toxic Chromium (VI) [4], [5] and separate multivalent salt ions from water [6]. PEI-modified mesoporous molecular sieve MCM-41, also referred to as CO<sub>2</sub> “molecular basket” [7]–[11], has been used to adsorb CO<sub>2</sub> from N<sub>2</sub> [7], [12] and flue gas [8], [11]. Similar PEI-modified molecular baskets have been developed for the adsorption of NO<sub>x</sub> [8] and H<sub>2</sub>S [13]. Removal of H<sub>2</sub>S has further use in purifying hydrocarbons for fuel cells [14]. PEI is also a useful polyelectrolyte in biofuel cells [15], [16] and tandem solar cells [17], and has been used in methanol cells to promote electrocatalysis and stability [18]. Sensing of biomolecules, such as glucose [19] and proteins [20], [21], using PEI in nanocomposites has also been demonstrated.

Among the numerous applications of PEI, its usage as a gene delivery carrier is of particular note. While viral carriers are very effective in cell targeting and internalization, their use is limited by elevated levels of cytotoxicity. PEI has been shown to be one of the most effective non-viral gene delivery carriers, owing to its large amine density and high buffering capacity [22]–[25]. Delivery using PEI also allows for great versatility, made possible by adjusting the PEI’s molecular weight (MW) [26], [27], degree of branching [28], [29], method of complexation [22] with the genetic material, and modifications made to the complexes [30]–[32]. An excellent review on the properties of PEI and its contribution to gene delivery can be found in Lungwitz et al. [33]

Many experimental studies have been done to evaluate the performance of PEI in gene delivery [33], [34], and to explore methods that can lead to balanced efficacy and cytotoxicity. These

experiments are supplemented by theoretical investigations that have appeared in recent years, mostly through all-atom (AA) MD simulations. AA force field parameters have been developed for linear PEIs using general AMBER force field by Ziebarth et al. [35], and for PEIs with different degrees of branching using CHARMM General Force Field (CGenFF) [36] by Sun et al. [37]. CGenFF was also used to develop a model for linear PEI grafted with polyethylene glycol by Wei et al. [38], by parameterizing dihedral angle parameters against ab initio calculations. Recently, Beu et al. [39] used CGenFF to obtain all the parameters for linear PEIs by comparing with ab initio calculations. These AA force fields have been used to study different aspects of PEI-based gene delivery. For instance, the effect of PEI structure (degrees of branching) on DNA-PEI interaction was studied for low [35], [37] and high MW [40] PEIs. DNA aggregation by PEI was studied with [41] and without [42], [43] lipid modifications on the PEI. Antimicrobial properties of PEI were studied using MD simulations of PEI and lipid bilayers [44], [45]. While providing molecular details on the interaction and functionality of the PEI during gene delivery, the AA MD simulations are incapable of matching the length and time scales in the real delivery process, where the complexes formed by PEI and genetic material are ~100 nm and larger in size and the entire delivery can take hours [34].

On a completely different scale, Yang and May [46] used the continuum level Poisson-Boltzmann theory for linear polycations to study the proton sponge effect, a mechanism hypothesized for PEI-mediated gene delivery. The same process was investigated by Freeman et al. [47] using a system of kinetic equations to predict time-dependent functions such as volume and osmotic pressure of the endosome. While these simulations could match the length and time scales, they are limited by significantly simplified molecular structure and very approximate local interactions.

Coarse-grained (CG) MD simulations sets a middle ground between the costly AA simulations and the approximate continuum level theories. Coarse-graining an atomic structure involves representing a few atoms by a single bead, thereby decreasing the total number of atoms (and dimension of the phase space). The intra- and intermolecular interactions are captured by careful parameterization of the CG model, which can produce quite accurate results via a reliable CG force field where a lower dimensional potential energy surface (PES) describes the AA PES [48]. There are various methods to coarse-grain an AA model, one of which is the bead-spring model where the bond lengths are modeled using a harmonic spring, whereas the bond angles and torsions are not considered. As one of the simplest CG models, the bead-spring model has been used to study the protonation state of linear PEI in various pH through a Monte Carlo simulation [49], and the interactions of DNA with lipid modified PEI using a MD simulation [50]. Jorge et al. [51] studied the complexation of DNA with linear PEIs using a CG Monte Carlo simulation, where the polyions were approximated using harmonic bond length and bond angle potentials [51]. An increasingly popular CG method is the Martini scheme developed by Marrink et al. [52], which provides a 1000-fold speedup with accuracy comparable to AA simulations without the requirement of parametrization for every new simulation [53]. In addition, new CG models developed under the Martini scheme are compatible with previously developed Martini CG models which can be adopted directly to allow for simulations of complex multi-component systems. Using Martini scheme, Wei and Luijten [38] developed a CG model for polyethylene glycol (PEG) grafted PEI (PEG-g-PEI) with various grafting density. While to our knowledge this is the first and only Martini CG model for PEI in the literature, the focus of this work was on the complexation of PEG-g-PEI with small interfering RNA and not on extensive parameterization of the CG PEI.

Because of this their CG model is only applicable to linear PEI with a single and high protonation ratio of 82%.

In this work, we aim at performing a general coarse-graining of PEI by considering a series of PEI molecules that have different degrees of branching and protonation ratios. It is known that PEIs can have vastly different branching structures and protonation states, which can lead to their different functionalities. Our extensive parameterization can therefore greatly benefit future CG simulations involving PEIs. Like in any CG modeling work, to be able to validate the CG models against the AA ones, the molecules we parameterized are relatively small (~600 Da). However, they form building blocks for larger PEIs and can be easily used and extended to different structures. The direct application of low MW PEI cannot be ruled out, for instance, ~600 Da PEI grafted with  $\beta$ -cyclodextrin has also been reported to be quite efficient in *in vivo* gene expression [54]. The compatibility of CG models within Martini will provide the opportunity for integrating the CG PEI models developed here with other well-established CG models for lipids [52], [55], proteins [56], [57], nucleic acids [58], polymers [59], [60], solvents [55], [61], [62] and ions [55], to study PEI-related biomedical systems with suitable length and time scales.

## 2. MODEL

Martini [48] follows a “bottom-up” approach to model bonded interactions and a “top-down” approach for non-bonded interactions. First, the atomistic structures are coarse-grained by mapping multiple atoms into a CG bead, which typically contains 4 heavy atoms. A bead type [52], [55], [56], [58], [60], [61], [63] is then assigned to each CG bead, based on finding a molecular analog[55] for the AA structure within the bead and comparing free energies values

from experiments and simulations. The sizes of CG beads under Martini are categorized into being “standard”, “small” and “tiny”. A “standard” bead [52] is usually created with 4:1 mapping (four heavy atoms to one CG bead), while a “small” bead [55] is typically used to model ring-like structures, with 3:1 or 2:1 mapping. Uncommonly used, “tiny” beads [58] were introduced to model base stacking in DNA, with 3:1 or 2:1 mapping. It should be noted that the resolution of mapping for each bead is not absolute; higher or lower or even fractional mapping ratios can be used [52]. The assigned bead type and size determine the parameters for non-bonded interactions. The bonded parameters are then determined by conducting CG and AA simulations and comparing the probability distributions of bond lengths, bond angles and dihedrals angles from the two sets of the simulations. The parameters are tuned by trial and error till a good match is obtained between the CG and AA simulations.

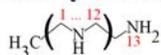
Below, we follow these basic principles to develop a CG model for four different PEI structures with two different protonation ratios (**Figure 1**) [37]. These four structures were named in Sun et al. [37] as purely linear (PL), semi-linear (SL), moderately branched (MB) and hyperbranched (HB), based on their degree of branching. In the context of gene delivery, the two protonation ratios, 23% and 46%, represent the protonation states of branched PEIs under physiological (pH=8) and late endosomal (pH=6) conditions, respectively [64]. The same protonation ratios are applied to all four structures to allow for a comparative study. The nitrogen atoms which get protonated at the two protonation ratios are indicated in **Figure 1** using a table. Henceforth, 23% and 46% protonated SL PEI will be referred to as SL23 and SL46 respectively, and the other PEIs are named in the same manner.

## All-Atom

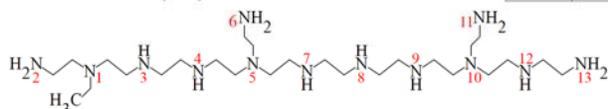
| PEI | 23%    | 46%           |
|-----|--------|---------------|
| PL  | 2,8,13 | 2,4,6,8,10,13 |
| SL  | 2,6,11 | 2,4,6,8,11,13 |
| MB  | 1,7,13 | 1,4,7,8,11,13 |
| HB  | 1,9,13 | 1,4,6,9,11,13 |

## Coarse-Grained

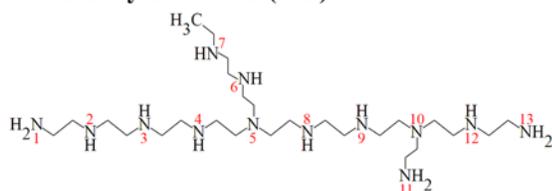
### Purely-Linear (PL)



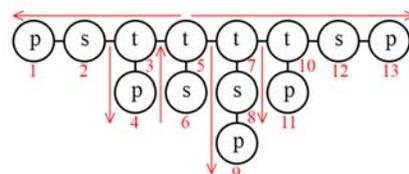
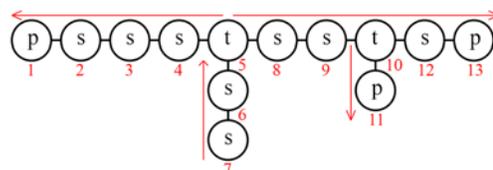
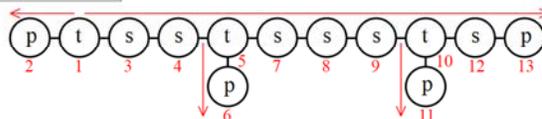
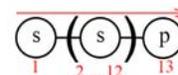
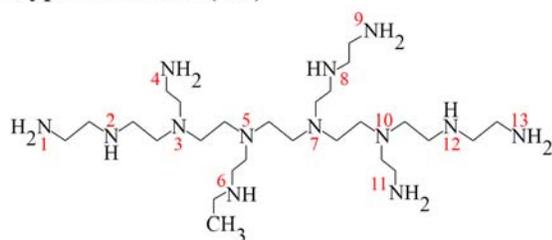
### Semi-Linear (SL)



### Moderately-Branched (MB)



### Hyper-Branched (HB)



**Figure 1.** Left Panel: AA structures of the PEIs studied in this work. Right Panel: corresponding CG structures. Nitrogens on the left and CG beads on the right are numbered in a consistent way. The inserted table shows the indices of the protonated nitrogens (AA) or beads (CG) at the two different protonation ratios (23% and 46%). Details on the labeling of the CG beads ('p', 's', etc.) are given in **Section 2.1**. The arrows in the right panel are used to indicate the direction from C terminal to N terminal inside the beads.

## 2.1. CG Mapping

The eight AA PEI structures are shown in **Figure 1** (left panel) each consisting of 13 C-C-N groups. To coarse-grain these structures, the center of mass of each C-C-N group is mapped to a

single CG bead (including the hydrogens). The beads that contain primary, secondary or tertiary nitrogens are labeled as ‘*p*’, ‘*s*’ and ‘*t*’ respectively. If a bead is charged, the letter ‘*q*’ is appended to its original label. For instance, a bead that contains a protonated primary nitrogen would be labeled as ‘*pq*’. This results in six different bead labels ‘*tq*’, ‘*t*’, ‘*sq*’, ‘*s*’, ‘*pq*’ and ‘*p*’. The PEI structures considered here only have uncharged tertiary nitrogen, so ‘*tq*’ beads are absent. AA structure of each bead mapped in this way has a C and an N terminal, which makes it asymmetrical. To specify the orientation of atoms inside the beads, red arrows in **Figure 1** (right panel) are used to indicate the direction from the C terminal to the N terminal. This information is critical in classifying bond lengths, bond angles and dihedral angles, which will be discussed in **Section 2.3**. The mapping scheme followed by Wei and Luijten [38] differs from ours in several aspects. Firstly, we chose to map each C-C-N group into a CG bead whereas Wei and Luijten [38] chose to map the C-N-C groups [38]. As a result, the two beads at the end of their linear PEI were asymmetric, with one containing four heavy atoms and the other containing two heavy atoms. Using a mapping scheme that produces asymmetric end beads for a symmetric polymer (linear PEI as considered in [38]) is undesirable [65]. Secondly, our mapping scheme is the same for all beads (C-C-N), whereas Wei and Luijten [38] used four different types of mapping, two for end beads, one for the central linear beads (C-N-C) and another for the tertiary beads.

## 2.2. Non-bonded Interactions and Selection of Bead Type

The van der Waals interaction between two beads is described by the 12-6 Lennard-Jones (LJ) potential

$$V_{LJ}(r_{ij}) = 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (1)$$

where,  $r_{ij}$  is the distance between two beads  $i$  and  $j$ , and the LJ parameters  $\sigma_{ij}$  and  $\varepsilon_{ij}$  are determined by the bead type [55]. In addition, the charged beads interact with each other by the Columbic potential

$$V_{el}(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}} \quad (2)$$

where  $q_i$  and  $q_j$  are the charges of beads  $i$  and  $j$  respectively,  $\varepsilon_0$  is the permittivity of free space, and  $\varepsilon_r$  is the relative dielectric constant of the solvent. Like the exclusion of 1-4 pair interactions in AA simulations, 1-2 pair interactions are neglected in CG simulations.

**Table 1.** Hydration free energy ( $\Delta G^{hyd}$ ) and partitioning free energy between water and Hexadecane ( $\Delta G_{HW}^{part}$ ), Chloroform ( $\Delta G_{CW}^{part}$ ), or Octanol ( $\Delta G_{OW}^{part}$ ). Comparison is made between ethylamine (experimental value) and Martini CG beads (tabulated data reported in the literature except for C<sub>5</sub> and C<sub>2</sub> in PW). The partitioning free energy is the same for polarizable water (PW) and non-polarizable water (W) [61].

| beads                | $\Delta G^{hyd} \left( \frac{kJ}{mol} \right)$ |       | $\Delta G_{HW}^{part} \left( \frac{kJ}{mol} \right)$ | $\Delta G_{CW}^{part} \left( \frac{kJ}{mol} \right)$ | $\Delta G_{OW}^{part} \left( \frac{kJ}{mol} \right)$ |
|----------------------|--|-------|--|--|--|
|                      | PW[61]   | W[55] | W/PW[55], [61]                                       |  |  |
| P <sub>5</sub>       | -27.7  | -25   | -28  | -18  | -10  |
| P <sub>4</sub>       | -21.6  | -18   | -23  | -14  | -9   |
| P <sub>3</sub>       | -21.6  | -18   | -21  | -10  | -7   |
| P <sub>2</sub>       | -15.6  | -14   | -17  | -2   | -2   |
| P <sub>1</sub>       | -15.6  | -14   | -11  | -2   | -1   |
| N <sub>da</sub>      | -10.9  | -9    | -7   | 0  | 3  |
| N <sub>d</sub>       | -10.9  | -9    | -7   | 0  | 3  |
| N <sub>a</sub>       | -10.9  | -9    | -7   | 0  | 3  |
| N <sub>0</sub>       | -5.2   | -2    | -2   | 6  | 5  |
| C <sub>5</sub>       | N/A  | 1     | 5  | 10   | 6  |
| C <sub>4</sub>       | 2.9  | 5     | 9  | 13   | 9  |
| C <sub>3</sub>       | 2.9  | 5     | 13   | 13   | 14   |
| C <sub>2</sub>       | N/A  | 10    | 16   | 15   | 16   |
| C <sub>1</sub>       | 9.7  | 14    | 18   | 18   | 17   |
| Ethylamine[66], [67] | -19  |       | -18  | 2  | 1  |

The “standard” CG beads can be polar (P), non-polar (N), apolar (C), or charged (Q) [55]. These bead types are further subdivided to capture different chemical natures of the beads. Polar and apolar bead types are each subdivided into five levels from 1 to 5 depending on their polarity (low to high), while non-polar and charged beads are subdivided based on their hydrogen-bonding capacity (d = donor, a = acceptor, da = both, 0 = none) [55]. In this work, all the uncharged beads have the same molecular analog, ethylamine. To choose the most suitable type for them, free energy of hydration and partitioning free energy between water and other organic solvents (hexadecane, chloroform and octanol) are evaluated for different Martini beads [55] and compared with the corresponding experimental values for ethylamine [66], [67] (**Table 1**). Examining all the data, P<sub>2</sub> is found to be the best candidate for the uncharged beads. For the charged beads, the type Q<sub>d</sub> is chosen because of their capability to serve as hydrogen bond donor. Different from our work, Wei and Luijten [38] used “small” beads based on the rationale that Martini mapping is usually 4:1, and the 3:1 mapping used in their work (and ours) is finer. However, the most important difference between “standard” and “small” beads is that “small” beads were introduced to model ring structures where the range and strength of the LJ interaction are smaller compared with “standard” beads [55]. This treatment will allow the “small” beads to form densely packed ring structure which are absent in PEI. Such ring-forming behaviors were in fact observed in Lee et al. [60] for low MW PEG when “small” beads were assigned in the CG model. To avoid the unrealistic formation of rings, we used “standard” size for our CG beads. Furthermore, our mapping scheme allows us to use only two types of beads for all possible PEI structures, while the four different mappings in Wei and Luijten [38] require five bead types. It should also be noted that free energy comparison, such as shown in **Table 1**, was not presented in Wei and Luijten [38] and it is not clear why the specific bead types were chosen.

### 2.3. Bonded Interactions

Bond lengths (eq 3) are modeled using a harmonic potential  $V_b$ , where  $r_{ij}$  is the distance between two bonded beads  $i$  and  $j$ ,  $K_b$  is the force constant, and  $r_{eq}$  is the equilibrium bond length. Bond angles (eq 4) are modelled using a cosine type harmonic potential  $V_a$ , where  $\theta_{ijk}$  is the angle between consecutive beads  $i$ ,  $j$  and  $k$ ,  $K_a$  is the force constant and  $\theta_{eq}$  is the equilibrium bond angle.

$$V_b(r_{ij}) = \frac{1}{2}K_b(r_{ij} - r_{eq})^2 \quad (3)$$

$$V_a(\theta_{ijk}) = \frac{1}{2}K_a \left( \cos(\theta_{ijk}) - \cos(\theta_{eq}) \right)^2 \quad (4)$$

The dihedral angles (eq 5) are modeled using a sum of periodic potentials ( $V_d$ ), where  $\varphi_{ijkl}$  is the dihedral angle between consecutive beads  $i$ ,  $j$ ,  $k$  and  $l$ .  $K_{d,w}$  is the force constant for the  $w^{\text{th}}$  periodic potential,  $n_w$  determines the number of peaks in this potential, and  $\varphi_{eq,w}$  specifies the location of those peaks. The parameter  $n_w$  in eq 5 is an integer, which depends on  $w$  as shown in eq 6, where  $[\cdot]$  is the greatest integer function.

$$V_d(\varphi_{ijkl}) = \sum_{w=1}^W K_{d,w} [1 + \cos(n_w \varphi_{ijkl} - \varphi_{eq,w})] \quad (5)$$

$$n_w = \left\lfloor \frac{w+1}{2} \right\rfloor \quad (6)$$

The parameters in eq 3-6 are determined by first making an initial guess for their values, followed by tuning of these values by trial and error. To explain the procedure, we will use the bond lengths as an example. First, AA simulations are performed for all the eight AA structures, from which we obtain the AA trajectories for the PEIs. These AA trajectories are then converted

to CG trajectories, using the mapping scheme discussed in **Section 2.1**. From the CG trajectories, the probability distributions for the bond lengths can be calculated. Recognizing that many of the bonds in the eight structures are identical in nature (explained in **Section 2.4**), the probability distributions of these identical bonds are averaged to obtain a single, averaged, probability distribution for each distinct bond length. These averaged probability distributions are obtained directly from the AA simulations (without running any CG simulations), and therefore are referred to as the *averaged AA distributions*. The initial guess for the CG bond length parameters in eq 3 can now be obtained by applying a direct Boltzmann inversion (DBI)[68] on the averaged AA distributions. This method was used by Seo et al. [57] to improve dihedral angle parameters for Martini proteins, but was found to be an effective way to generate the initial guess for all the bonded parameters in this work. Specifically, from the averaged AA distribution  $P_b^{AA}$  for a bond length we can calculate the potential energy  $U_b^{AA}$  for this bond by  $U_b^{AA} = -k_B T \ln(P_b^{AA})$ , where  $k_B$  is the Boltzmann's constant and  $T$  is the absolute temperature.  $U_b^{AA}$  is then curve fitted with the bond length potential (eq 3) to obtain the initial guess for the bond length parameters  $K_b$  and  $r_{eq}$ .

The initial guess for bond angle and dihedral parameters can be obtained in the same way, which provides a complete set of initial parameters that allow for CG simulations. The CG simulations will generate new probability distributions for bond lengths, angles, and dihedrals, which will be referred to as the *CG distributions*. It should be noted that CG distributions are obtained for individual structures while the averaged AA distributions are obtained by averaging distributions from all structures. The CG simulations are run one structure after another starting from SL46 followed by MB46, HB46, PL46, SL23, MB23, HB23, and finally PL23. The CG distributions are compared with the corresponding averaged AA distributions, and depending on the difference, the CG parameters are updated until a good match is obtained between the two sets of distributions.

Bonded parameters determined from previously simulated structures are used directly if the same bonded interactions appear again in a structure simulated later. For example, the parameters for bond length ‘ $t$ - $s$ ’ are determined first from SL46 and then directly used for the same bond in MB46, HB46, SL23, MB23 and HB23. CG distributions for the ‘ $t$ - $s$ ’ bond obtained from simulating these structures (MB46, HB46, SL23, MB23, HB23) are checked to make sure that they also match well with the averaged AA distribution (see **Supporting Information** for details). As new bond lengths, angles and dihedrals appear in a new structure, their parameters are obtained as mentioned above. This is continued until all the bonded parameters for the eight PEI structures are found.

Adjusting the bonded parameters to match the averaged AA and CG distributions is an iterative process. For the bond lengths and angles, the parameters in eqs 3 and 4 control the location of the extrema ( $r_{eq}, \theta_{eq}$ ) and the width of the distributions ( $K_b, K_a$ ). They can therefore be systematically varied to achieve better matching. Such a trial and error method is not trivial for the dihedral angle potentials (eq 5). For most dihedral angles the CG distribution matches the averaged AA distribution very well using the initial guess of parameters from DBI and eight periodic functions ( $W=8$  in eq 5), with a few requiring more than eight. However, for some dihedrals good matching between CG and averaged AA distributions cannot be obtained with DBI alone. For those dihedrals, a method similar to the iterative Boltzmann inversion [69] (IBI) is used. Consider a specific dihedral for which we would like to determine the parameters in eq 5. In the  $m^{\text{th}}$  iteration of IBI ( $m \geq 1$ ), a CG simulation is performed based on the current bonded parameters, and a CG distribution  $(P_d^{\text{CG}})_m$  is obtained for this dihedral. Its corresponding potential energy,  $(U_d^{\text{CG}})_m$ , is calculated using  $(U_d^{\text{CG}})_m = -k_B T \ln(P_d^{\text{CG}})_m$ . The difference between  $(U_d^{\text{CG}})_m$  and the potential energy evaluated from the averaged AA distribution,  $U_d^{\text{AA}} = -k_B T \ln P_d^{\text{AA}}$ , is computed and

$(\delta U_d^{\text{CG}})_m = (U_d^{\text{CG}})_m - U_d^{\text{AA}}$  is curve fitted to a function  $(\delta V_d)_m$  in the form of eq 5. The new potential for the dihedral angle is then obtained using  $(V_d)_{m+1} = (V_d)_m + \eta(\delta V_d)_m$ , where  $\eta$  is a positive real number less than one.

## 2.4. Nomenclature for Bonded Interactions

To group similar bonds, angles and dihedrals (for calculating averaged AA and CG distributions) we systematically name them as follows. Consider first the bond lengths in the eight structures. The bond formed between beads  $i$  and  $j$  will be referred to as bond  $i$ - $j$ , where  $i$  and  $j$  are the indices for the CG beads shown in **Figure 1**. Intuitively one may think that bonds  $i$ - $j$  and  $j$ - $i$  are identical, but this is not necessarily the case for the CG beads. For instance, bonds 5-7 and 9-10 in SL46 look similar in the CG structure, however their AA structures show a difference because the tertiary nitrogen is near the center for bond 5-7 while it is near the edge for bond 9-10. Therefore, bonds 5-7 and 9-10 are treated differently based on the order of their beads with respect to the red arrow in **Figure 1**. Along the direction of the red arrow, which points from the C terminal to the N terminal, bond 5-7 is ‘ $t$ - $s$ ’ while bond 9-10 is ‘ $s$ - $t$ ’. If we look at bonds 4-5 and 5-8 in MB23 (note: carbons of bead 5 are present in the side chain), along the direction of the red arrows, they are both ‘ $t$ - $s$ ’ and hence identical. Many bond lengths in the eight structures can then be identified to be the same. For example, bonds 1-3, 5-7 and 10-12 in SL46 and SL23, 10-12 in MB46, 4-5, 5-8 and 10-12 in MB23, as well as 2-3, 7-8 and 10-12 in HB46 and HB23 are all ‘ $t$ - $s$ ’; while bonds 9-10 in SL46, 4-5 and 9-10 in SL23, 5-6 and 9-10 in MB46 and MB23, and 5-6 in HB23 are ‘ $s$ - $t$ ’. The bond length distributions of these identical bonds, based on CG trajectories mapped from AA simulations, are averaged to obtain the averaged AA distribution as mentioned earlier.

Similarly, when examining the bond angles, denoted as angle  $i-j-k$  where  $i, j$  and  $k$  are the indices for the CG beads in **Figure 1**, we will follow the direction of the red arrows (from C to N terminals). For example, bond angles 1-3-4 and 5-7-8 in SL46 both belong to ' $t-s-sq$ ', whereas bond angles 8-9-10 in SL46 and MB46, and 5-6-7 in MB46 and MB23 all belong to ' $sq-s-t$ '. Special care should be taken for beads that contain a tertiary nitrogen because each of them has three associated bond angles, two of which can be arranged from C to N terminals (referred to as *normal bond angles*), while the third can only be arranged from N to N terminals (referred to as an *N-type bond angle*). For example, in SL46 bond angle 6-5-7 is an N-type bond angle while angles 4-5-7 and 4-5-6 are normal bond angles. Since the red arrows in **Figure 1** point the direction from C to N terminals, we can identify N-type bond angles as the angles where the direction of the red arrows changes. By definition, the bead in the middle of an N-type bond angle must be either ' $tq$ ' or ' $t$ '. To denote the N-type bond angles, the letter 'N' is prefixed to the label; for example, bond angle 6-5-7 in SL46 can be either ' $Npq-t-s$ ' or ' $Ns-t-pq$ '. N-type angles are symmetrical because they are from one N terminal to another, and the terminal beads are connected to same tertiary bead, making both ' $Npq-t-s$ ' and ' $Ns-t-pq$ ' equivalent.

A similar analysis can be done for dihedral angles, denoted as dihedral  $i-j-k-l$ . Like bond angles, a dihedral angle is called an N-type dihedral if it can only be arranged from N to N terminals. An N-type dihedral must contain a ' $t$ ' or ' $tq$ ' bead, and this bead may be located at either the second or the third position. To keep the naming simple, we arrange the beads such that the first three beads form an N-type bond angle and the last three beads form a normal bond angle. In such a representation, the second bead is always a tertiary bead. For instance, dihedral 6-5-7-8 in SL46 belongs to ' $Npq-t-s-sq$ ' and dihedral 7-6-5-8 in MB46 belongs to ' $Nsq-t-s-sq$ '. Even though N-type dihedral angles are from N to N terminal they are not symmetrical like N-type bond angles.

Dihedral angle ' $Ni-j-k-l$ ' is not the same as ' $Nl-k-j-i$ ' because the former has ' $Ni-j-k$ ' as an N-type bond angle while the latter has ' $Nl-k-j$ ' as an N-type bond angle.

The parameters for all bond lengths, bond angles and dihedral angles present in the eight PEI structures were obtained after many trial-and-error iterations and are reported in **Tables S1-S3** in **Supporting Information**.

## 2.5. AA simulations

AA simulations were run on the eight different PEI structures separately. Each system consisted one PEI in TIP3P water [70], with a proper number of  $Cl^-$  ions added to neutralize the system i.e., the salt concentration was zero. To study the effect of salts, additional AA simulations were run for all the PEI structures in 150 mM *KCl* salt solution mimicking the physiological conditions [31]. Force field parameters for the AA models were adopted from Sun et al. [37], which was developed based on the CGenFF [36]. The initial structure for the PEI in each AA simulation was solvated in a rectangular box with a margin of 2 nm from all sides. Each system was first subjected to steepest descent energy minimization. This was followed by constrained NVT simulation at 300 K and then constrained NPT simulation at 300 K and 1 bar, where all the bonds in the PEI and water were constrained using the LINCS algorithm [71]. The temperature was maintained using a velocity-rescaling thermostat [72] with a time constant of 0.1 ps, and the pressure was maintained using Berendsen barostat [73] with a time constant of 2 ps and compressibility of  $4.5 \times 10^{-5} \text{ bar}^{-1}$ . Finally, all constraints were removed except the bonds involving hydrogen atoms, and NPT simulation was run using Parrinello-Rahman barostat [74] and velocity-rescaling thermostat [72] with the same time constants and compressibility. The NPT simulation was performed for 100 ns and the last 50 ns was used for analysis. In all simulations, periodic boundary condition was applied

in all directions and equations of motion were integrated using the leapfrog algorithm with a time step of 2 fs. Short-range non-bonded interactions were cut off at 1.2 nm, while long-range electrostatic interactions were handled using particle-mesh Ewald [75] (PME). A dispersion correction was introduced to adjust both pressure and energy to compensate for the cutoff of non-bonded interactions.

## 2.6. CG simulation

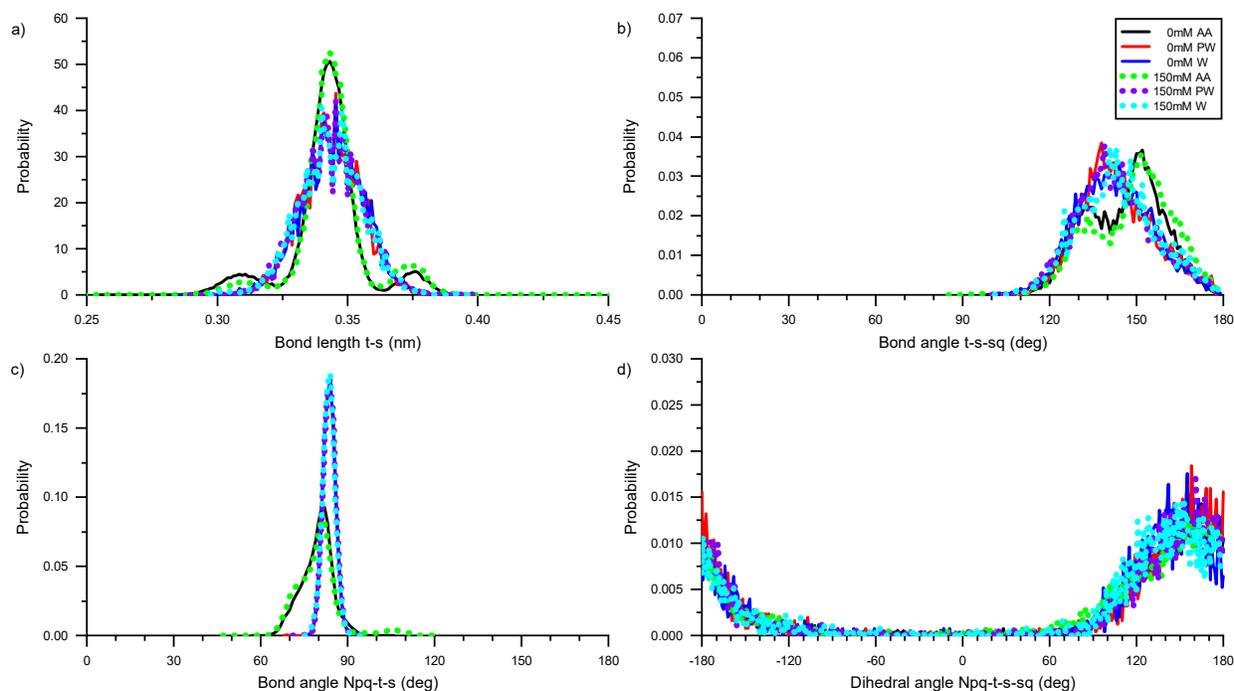
CG Simulations corresponding to AA simulations were conducted in polarizable Martini water [61] (PW), which were used to determine the bonded parameters for the CG model. Using these bonded parameters, further CG simulations were conducted in non-polarizable Martini water [52] (W). Since simulations with non-polarizable water are much faster but less accurate, this comparison allows us to assess the feasibility of employing non-polarizable water for PEI-related CG simulations. Initial structures of the PEIs in the CG simulations were obtained by mapping the equilibrated configurations at the end of the AA simulations using the mapping scheme specified in **Section 2.1**. All CG simulations were performed at the same temperature and pressure as in the AA simulations, following the same steps with the exclusion of a constrained NVT simulation after energy minimization. Since the low-amplitude high-frequency fluctuations are removed from the PES in the CG models [48], no additional constrained NVT simulation was necessary prior to the NPT simulation. The LJ potential was cut off at 1.2 nm by smoothly shifting it to zero between 0.9 and 1.2 nm. The electrostatic forces in the long range were handled using PME [75], and in the short range were cut off at 1.4 nm and smoothly shifted between 0.0 and 1.4 nm. Periodic boundary condition was applied in all the directions. The dielectric constant was taken to be 2.5 in PW [61] and 15 in W [52]. The same thermostat and barostat used in the AA simulations, with identical

coupling parameters, were used except that a time constant of 0.5 ps was used in the Berendsen thermostat for both constrained and unconstrained NPT simulations in CG-W. The smoother PES allowed a larger time step of 10 fs to be used in all CG simulations. Simulation parameters for systems with 150 mM *KCl* are identical to those used for the 0 mM *KCl* systems. It is worth pointing out that dynamics in CG simulations is faster compared to AA simulations due to the smoother PES. The speedup factor can be calculated by comparing the self-diffusion coefficients of individual water molecules in AA and CG simulations. Because each CG water bead in Martini consists of four water molecules, the standard time scaling factor is taken to be four [52]. Unless otherwise specified, time in this work for all CG simulations is the effective time, scaled according to this speedup factor. For all systems, the unconstrained NPT simulation was run for an effective time of 100 ns and the last 50 ns were used for analysis. All the MD simulations, AA and CG, were performed using GROMACS software package version 4.6.5. [71]

## **2.7. Potential of mean force calculation**

Potential of mean force (PMF) was calculated between DNA and PEI using umbrella sampling, taking the distance between their centers of mass (com) as the reaction coordinate. A Drew-Dickerson dodecamer DNA  $d(\text{CGCGAATTCGCG})_2$  with 24 nucleotides in its fully deprotonated state and the SL46 PEI were used for the PMF calculation. For AA simulations, CHARMM27 force field [76] was used for the DNA, parameters developed and validated by Sun et al. [37] based on CGenFF [36] were used for the PEI, and TIP3P [70] model was used for the water molecules. For CG simulations, CG parameters developed in this work were used for the PEI, and the Martini DNA [58] and polarizable water model [61] were adopted without modification.

In the AA simulations, the DNA and PEI were initially placed 1.5 nm apart (com-com distance) along the z-axis, with their principal axes aligned with the x-axis. The center of mass of the PEI was then pulled along the z-axis, while restraining the DNA, to generate a set of configurations serving as initial configurations in the umbrella sample simulations. These configurations correspond to DNA-PEI center of mass distance between 1.2 nm and 5.4 nm, with a step of 0.1 nm. Each umbrella sampling simulation was performed for 20 ns where the DNA-PEI center of mass distance was restrained with a harmonic spring of force constant  $1000 \text{ kJmol}^{-1}\text{nm}^{-2}$ . The PMF was then calculated from all of the umbrella sampling simulations using the weighted histogram analysis method [77]. The CG simulations followed the same procedure except that the initial configurations were generated at every 0.05 nm, and the harmonic spring had a force constant of  $2000 \text{ kJmol}^{-1}\text{nm}^{-2}$ . Each umbrella sampling simulation was performed for 100 ns. All simulations were done in a cuboidal box of  $6 \times 4 \times 12 \text{ nm}^3$  containing water and 150 mM KCl salt, at the temperature of 300 K and pressure of 1 bar.



**Figure 2.** Representative averaged AA distributions (AA) and CG distributions for SL46: (a) Bond length “ $t-s$ ”, (b) Bond angle “ $t-s-sq$ ”, (c) Bond angle “ $Npq-t-s$ ” (same as “ $Ns-t-pq$ ”), and (d) Dihedral angle “ $Npq-t-s-sq$ ”. Each subfigure contains six curves that correspond to AA simulations with 0 mM and 150 mM *KCl*, CG-PW with 0 mM and 150 mM *KCl*, and CG-W with 0 mM and 150 mM *KCl*.

### 3. RESULTS AND DISCUSSION

#### 3.1. Bonded parameters

For all the eight structures, there are 99 different bonded distributions and the quality of matching for a few of them is shown in **Figure 2**. As described earlier, the bonded parameters were first developed for PEI structures in PW with 0 mM *KCl* salt, but then applied to CG simulations in 150 mM *KCl* (PW) and in W (0 mM and 150 mM *KCl*). Therefore, each subfigure contains six curves: two averaged AA distributions obtained from AA (0 mM and 150 mM *KCl*)

simulations, two CG distributions from CG-PW (0 mM and 150 mM *KCl*) simulations and two from CG-W (0 mM and 150 mM *KCl*) simulations. Systems with 0 mM and 150 mM salt concentrations showed negligible difference in the averaged AA distributions. This is well captured by our CG distributions, which remain unaltered irrespective of the water model or salt concentration used.

**Figure 2a** represents the distributions of bond length '*t-s*', where the CG distributions under all conditions (in PW or W, with or without salt) are able to capture the location of the peak (most probable bond length) and the width of the averaged AA distribution. The distributions of two bond angles '*t-s-sq*' and '*Npq-t-s*' are shown in **Figure 2b** and **Figure 2c** respectively. Clearly, the CG distributions of '*t-s-sq*' overlap well with the averaged AA distribution, whereas there are noticeable differences between the CG and averaged AA distributions for bond angle '*Npq-t-s*'. Specifically, the location of the peak, which dictates the most probable bond angle, is slightly larger in the CG distributions (84° as compared to 81° in the averaged AA distributions). This implies that the distance between the terminal beads ('*pq*' and '*s*') is slightly larger in the CG simulations. We calculated time average of this distance from the CG simulation to be 0.457 nm, which is smaller than the equilibrium distance (0.527 nm) in the LJ potential (eq 1). Therefore, the terminal beads experience a repulsive force. As a result, further decreasing the equilibrium bond angle ( $\theta_{eq}$  in eq 4) did not improve the quality of the matching. Additional adjustment of the force constant ( $K_a$ ) was performed but also found to be ineffective: lowering  $K_a$  causes the most probably bond angle to further increase, while using a higher  $K_a$  makes the probability distribution very sharp. Similar observations were made for most N-type bond angles. The difficulty of finding a perfect matching for an N-type bond angle may be associated with the bulky structure near the tertiary bead, which is simultaneously attached to three beads. It is challenging to accommodate

the steric hindrance of these beads and the non-bonded interactions among all of them. Nevertheless, the CG distributions still match the averaged AA distributions reasonably well and similar degree of discrepancy has been deemed acceptable [58]. The dihedral angle “*Npq-t-s-sq*” in **Figure 2d** is modeled perfectly, and it is interesting to notice that the N-type dihedral angle does not suffer from the same discrepancy as seen in **Figure 2c**. This is because the dihedral angle involves four beads which experience less steric hindrance, and the distance between terminal beads is well above the equilibrium distance in the LJ potential.

The comparisons between averaged AA and CG distributions for all bonded interactions are given in the **Supporting Information**. A vast majority of the averaged AA distributions for bond lengths and bond angles contain a single peak; these distributions are excellently matched by our CG distributions. A few averaged AA distributions for bond lengths and bond angles show multiple peaks, which is not an artifact of averaging multiple bond lengths/bond angles. Wei and Luijten [38] also observed such distributions, even though they used a different mapping scheme. In those cases, our CG distribution accurately captures the average bond length/angle, and the shape of the distribution matches reasonably well. For dihedral angles, the CG distributions matched almost perfectly with the averaged AA distributions except for a few dihedrals in the HB PEIs, primarily due to steric hindrance in those highly branched structures. Even then, the CG distributions have properly captured the location of the peaks, and only overestimated or underestimated certain peak heights. The averaged AA distributions were obtained from the last 50 ns of the 100 ns simulations. As a check of equilibrium, the averaged AA distributions were also evaluated over the periods of 50-75 ns and 75-100 ns (see Figure S6 in Supporting Information). The distributions over different time windows almost overlap, confirming the attainment of equilibrium.

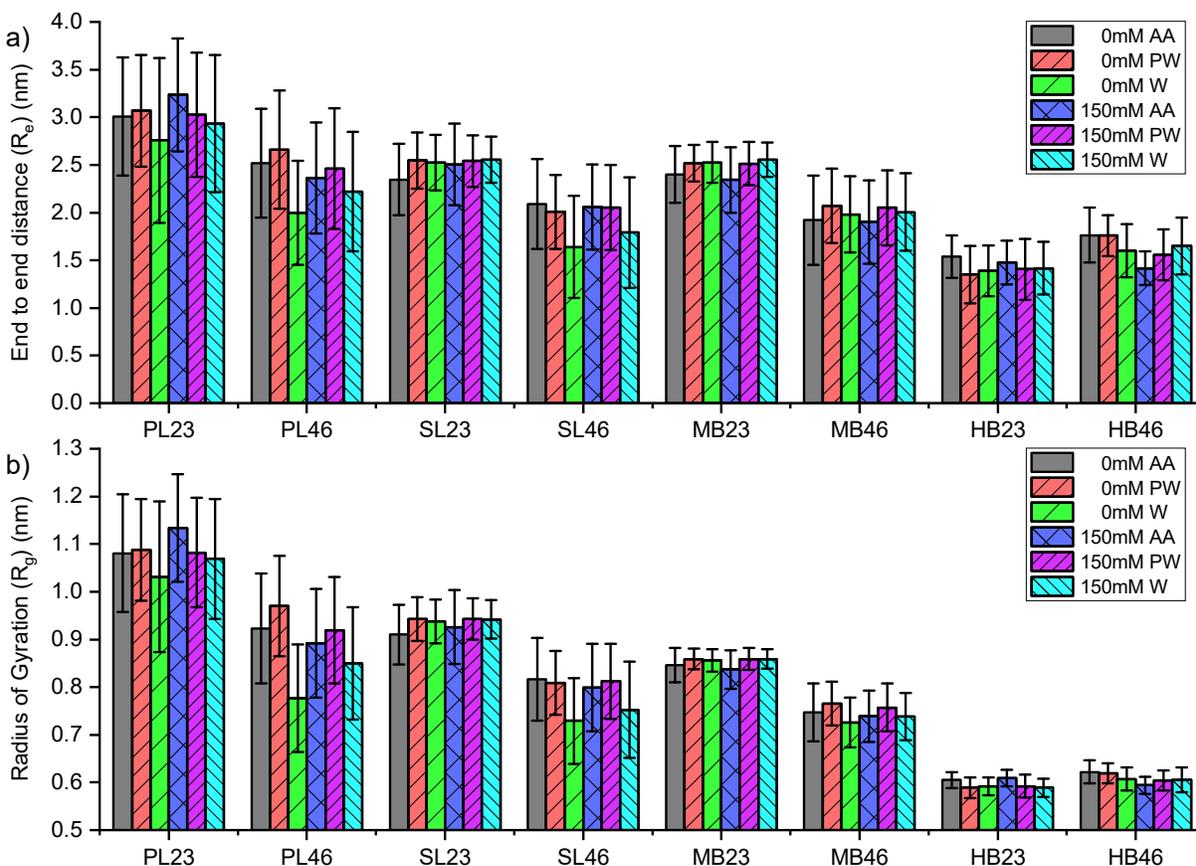
### 3.2. Global properties

To further validate the CG model, two global properties of the polymers, namely the end-to-end distance ( $R_e$ ) and radius of gyration ( $R_g$ ), are calculated from the CG simulations and compared with the results from AA simulations (using the mapped trajectories). For the branched structures (SL, MB and HB),  $R_e$  was taken to be the end-to-end distance of the longest linear chain (bead 1 to bead 13 in **Figure 1**). Average value of  $R_e$  and  $R_g$  are represented in **Figure 3**, along with their standard deviations. Each subfigure contains eight sets of data, one for each PEI molecule. For each set, there are six columns that correspond to results from AA, CG-PW and CG-W simulations, with either 0 mM or 150 mM *KCl*.

Considering the standard deviations, a good overlap is observed between the AA and CG results in all systems. The standard deviation tends to be higher in the PL and SL structures, which is expected since the PL and SL structures are more flexible and undergo greater fluctuations. For a given structure at a given protonation ratio, even though the CG parameters were obtained from the 0 mM *KCl* system in PW, the addition of salt did not have a strong influence on the values of  $R_e$  and  $R_g$ , nor did it affect the comparison with its AA counterparts. In fact, the difference between the result in 0 mM AA simulation and that in 150 mM AA simulation is also small, indicating that the screening effect of salt on such small molecules is negligible, although it is known to strongly affect the conformation of macromolecules [78].

At a given protonation ratio, the AA simulations predict  $R_e$  and  $R_g$  to decrease with the increase in degree of branching (PL > SL  $\geq$  MB > HB). Overall the same trend is predicted by the CG simulations, except for PL46 and SL46 simulated in W. These two systems also have the largest discrepancy when compared with the corresponding AA simulations. Clearly, all systems studied

in the CG-PW simulations have produced results that match better with the AA simulations. In the following section, we further discuss the quality of the water model (PW vs. W).

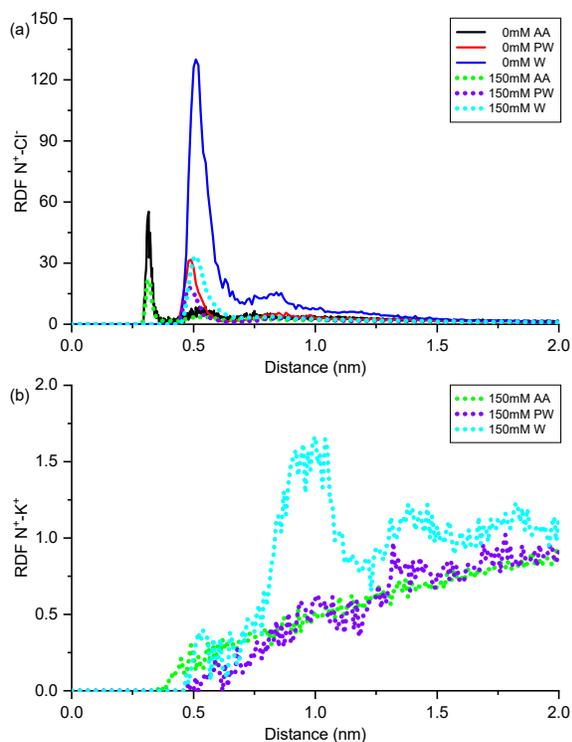


**Figure 3.** Comparison of end-to-end distance (a), and radius of gyration (b) between AA, CG-PW and CG-W simulations with 0 mM and 150 mM *KCl*. The error bars represent the standard deviations of the quantities.

### 3.3. Quality of the water model

The Martini PW model includes partial charges to model the dipole moment of water [61], and is thus superior to the Martini W model in predicting ion-ion and ion-water interactions. In **Figure 3**, it has been observed that the W model is less accurate than the PW model, especially in linear

structures with higher protonation ratios. However, there is still reasonable overlap between the results in W and those from AA simulations, and it captures the trend of  $R_e$  and  $R_g$  accurately except for PL46. In addition, the W water model is computationally more efficient than PW, as each PW has two additional particles with charges [61]. It is therefore of interest to further examine the quality of the W model. To do so, we plot two radial distribution functions (RDFs) in **Figure 4** for system SL46. The RDFs for the other structures share qualitative similarity and hence are not shown here.



**Figure 4.** Radial distribution functions for (a) chloride ions around protonated nitrogens; and (b) potassium ions around protonated nitrogens. All data are for SL46 in AA, CG-PW and CG-W simulations with 0 mM and 150 mM KCl.

**Figure 4a** shows the RDF of  $Cl^-$  around  $N^+$  in both 0 mM and 150 mM  $KCl$  solutions. RDFs from the AA simulation (with and without salt) both show a prominent first peak at 0.315 nm

which corresponds to the first co-ordination shell, and a secondary peak at 0.528 nm which corresponds to  $N^+$  and  $Cl^-$  separated by a single layer of water. The system with 150 mM  $KCl$  has a higher number density of  $Cl^-$  (six times the density in 0 mM  $KCl$ ), and thus has a lower peak height since the RDF is normalized with respect to the bulk density. In case of CG simulations, the RDF was calculated between charged CG beads (' $sq$ ' or ' $pq$ ') and CG beads representing chloride ions. Compared with the AA results, the peak at 0.315 nm is absent in all CG simulations because each CG chloride ion is internally solvated with four water molecules [55]. The first peak of the RDF in CG-PW is at 0.492 nm and that for CG-W is at 0.510 nm, irrespective of the salt concentration. These values in fact match very well with the second peak in the RDFs from AA simulations. Like the AA simulations, for both water models (PW or W), the height of the peaks is lower for the system containing 150 mM  $KCl$ .

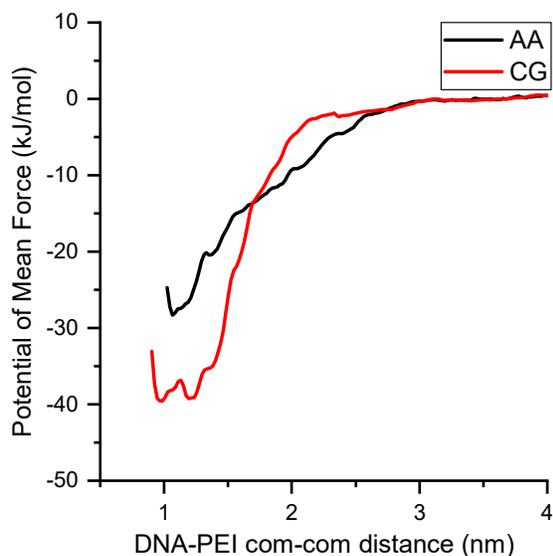
More interestingly, at a given salt concentration (0 mM or 150 mM), the peak height is considerably larger in CG-W, which signifies that  $Cl^-$  are on average closer to  $N^+$  in CG-W. **Figure 4b** represents the RDF of  $K^+$  around  $N^+$  in the simulations with 150 mM  $KCl$ . The RDF for CG-PW closely follows the AA simulation, whereas the RDF of CG-W has a significant amount of  $K^+$  accumulation near  $N^+$  between 0.8 and 1.2 nm. This can be explained by the attraction of  $K^+$  towards  $Cl^-$  which is accumulated near  $N^+$  (as seen in **Figure 4a**). Clearly, the W model overestimates ion-ion interactions, which was also reported by Wei and Luijten [38].

From the discussions above, it can be concluded that while simulating the CG PEI in W gives reasonable predictions for the global properties of the PEIs, it lacks the capability to properly capture the distribution of ions around a charged polymer [61]. The over-predicted electrostatic interactions in W can in turn influence the binding of the PEIs with other charged molecules such as nucleic acids in gene delivery. Even interactions with neutral molecules can be influenced if

they have local charges. Therefore, despite of its computational efficiency, the W model should be used with caution in CG PEI simulations.

### 3.4. DNA-PEI interaction

To assess the compatibility of the developed CG PEI model with existing Martini DNA [58], PMF calculations were performed between a SL46 PEI and a DNA at AA and CG levels (see **Section 2.7** for details). As seen in **Figure 5**, the PMF generated from AA simulations shows an initial decrease corresponding to repulsion between the two molecules at close proximity. The PMF reaches the global minimum of -29 kJ/mol at 1.1 nm, beyond which it increases with the DNA-PEI center of mass distance, representing an attractive force. A stationary value is reached at 2.6 nm where the interaction between the DNA and PEI diminishes. The PMF obtained from the CG simulations shows similar behavior. It reaches a global minimum of -39 kJ/mol at 1.1 nm and becomes stationary after 2.4 nm. The location of the global minimum is captured accurately, as well as the distance at which the PMF plateaus. The depth of PMF is larger from the CG simulations corresponding to stronger DNA-PEI interaction, but the overall matching is well within the accepted range of Martini [55], [56], [58], [63]. The results indicate that when our CG PEI model is used to simulate interaction with Martini DNA, no additional parameterization is needed.



**Figure 5:** Potential of mean force between DNA and PEI using the distance between their centers of mass as the reaction co-ordinate.

### 3.5. Application of the CG PEI model

The CG model developed here is based on small PEIs with MW of ~600 Da. Since we considered 8 molecules with different structures (degrees of branching) and protonation states, the model has included most of the key building blocks present in a typical PEI, in particular protonated and non-protonated primary and secondary amines, and tertiary amines. PEIs with higher MW such as 25kDa [79] have shown to be very effective in gene delivery. Bond lengths, bond angles and dihedrals developed in this work are likely to appear again if such larger PEIs are to be investigated. The parameters determined here can therefore be directly used, to reduce the number of new bonded interactions that need to be parametrized. In fact, higher MW PEIs constructed with repeating units of PL, SL and MB PEIs studied here can be generated without introducing any additional bonded parameters. This will allow us to model large-scale polymer solutions with

different polydispersity index. Our model can also be used as a starting point to model modified PEIs such as PEG-g-PEI, Poly(2-methyl-2-oxazoline), Poly(2-ethyl-2-oxazoline), Poly(2-propyl-2-oxazoline) and Ethoxylated PEI, with additional determination of the parameters associated with the junctions and side groups.

The selection of bead types and hence non-bonded parameters were determined by comparing various free energies with experimental values, and they are directly applicable to other PEI molecules with different MW and degree of branching, in a wide range of solvents. The CG PEI model can also be used in conjunction with other Martini models, e.g. DNAs [58], proteins [56], [63], lipids [55], etc., to study their interaction and usage in applications such as gene delivery. In such cases, additional potential of mean force calculations should be performed to validate the interaction between the PEI and other Martini models.

While there are CG models for branched polymers such as PAMAM [80] and branched polyethylene [81], they have been limited to very specific structures. Even for a molecule like PEI with a wide range of branching, existing CG models are limited to linear chains [38], [50], [51]. To our knowledge our methodology to model branched polymers is one of the first, and the same approach can be adopted to coarse-grain other branched polymers.

## **4. Conclusion**

We have developed a coarse-grained model of PEI, under the Martini scheme, that is applicable to PEIs with different degrees of branching, molecular weights and protonation ratios. Based on various solvation and partitioning free energies, two types of beads were selected to model the non-bonded interactions, and they can be used to model any PEI structures. Extensive

parameterization was done for the bonded interactions, with probability distributions that compare well with all-atom simulations. Our coarse-grained model accurately captures the behaviors of PEI in water without salt, as well as in aqueous solution at physiological salt concentration relevant to biological processes. Compared with the Martini polarizable water, the non-polarizable water model predicts poorer but still acceptable structural parameters ( $R_e$ ,  $R_g$ ) for the PEIs. However, it significantly overestimates the ion-ion interactions, leading to accumulation of ions around the protonated amines which is much higher than results from all-atom simulations. The model developed here can be used in conjunction with other Martini models can be further expanded to model higher MW PEI and modified PEIs.

## ASSOCIATED CONTENT

### **Supporting Information**

Discussion on the number of bonded parameters; CG parameters for bond length, bond angle and dihedral angles; comparison between averaged AA distributions and their corresponding CG distributions.

## AUTHOR INFORMATION

### **Corresponding Author**

\*Email: [tian.tang@ualberta.ca](mailto:tian.tang@ualberta.ca) Phone: +17804925476

### **Author Contributions**

S. M. and T. T. conceived the PEI mapping scheme and simulations procedures. S.M. did all the simulations and parameterizations. Both authors analyzed simulations results and wrote the paper.

## Funding Sources

Natural Science and Engineering Research Council (NSERC)

MITACS Globalink Graduate Fellowship

## Notes

The authors declares no competing financial interest.

## ACKNOWLEDGEMENTS

We acknowledge the computing resources and technical support from Western Canada Research Grid (WestGrid). Financial support from the Natural Science and Engineering Research Council (NSERC) of Canada and MITCAS is gratefully acknowledged.

## References

- [1] T. Saegusa, H. Ikeda, H. Fujii, *Macromolecules*, **1972**, *5*, 108–108.
- [2] T. D. Perrine, W. R. Landis, *J. Polym. Sci. Part A-1*, **1967**, *5*, 1993–2003.
- [3] G. Harold T., J. K. Edzwald, *Environ. Sci. Technol.*, **1979**, *13*, 299–305.
- [4] I. Korus, K. Loska, *Desalination*, **2009**, *247*, 390–395.
- [5] R. Zhao, X. Li, B. Sun, Y. Li, Y. Li, R. Yang, C. Wang, *J. Mater. Chem. A*, **2017**, *5*, 1133–1144.
- [6] Y. Lv, H. Yang, H. Liang, L. Wan, Z. Xu, *J. Memb. Sci.*, **2015**, *476*, 50–58.
- [7] X. Xu, C. Song, J. M. Andr, B. G. Miller, A. W. Scaroni, *Microporous Mesoporous Mater.*, **2003**, *62*, 29–45.
- [8] X. Xu, C. S. T, B. G. Miller, A. W. Scaroni, *Fuel Process. Technol.*, **2005**, *86*, 1457–1472.
- [9] X. Wang, V. Schwartz, J. C. Clark, X. Ma, S. H. Overbury, X. Xu, C. Song, *J. Phys. Chem. C*, **2009**, *113*, 7260–7268.

- [10] D. Wang, X. Ma, C. Sentorun-shalaby, C. Song, *Ind. Eng. Chem. Res.*, **2012**, *51*, 3048–3057.
- [11] X. Xu, C. Song, B. G. Miller, A. W. Scaroni, *Ind. Eng. Chem. Res.*, **2005**, *44*, 8113–8119.
- [12] X. Xu, C. Song, J. M. Andresen, B. G. Miller, A. W. Scaroni, *Energy & Fuels*, **2002**, *16*, 1463–1469.
- [13] X. Wang, X. Ma, X. Xu, L. Sun, C. Song, *Top. Catal.*, **2008**, *49*, 108–117.
- [14] X. Xu, I. Novochinskii, C. Song, *Energy & Fuels*, **2005**, *19*, 2214–2215.
- [15] D. Ivnitski, B. Branch, P. Atanassov, C. Apblett, *Electrochem. commun.*, **2006**, *8*, 1204–1210.
- [16] M. Das, L. Barbora, P. Das, P. Goswami, *Biosens. Bioelectron.*, **2014**, *59*, 184–191.
- [17] Y. Zhou, C. Fuentes-Hernandez, J. Won Shim, T. M. Khan, B. Kippelen, *Energy Environ. Sci.*, **2012**, *5*, 9827–9832.
- [18] Y. Cheng, C. Xu, P. Kang, S. Ping, *Appl. Catal. B Environ.*, **2014**, *158*, 140–149.
- [19] C. Shan, H. Yang, J. Song, D. Han, A. Ivaska, L. Niu, *Anal. Chem.*, **2009**, *81*, 2378–2382.
- [20] Q. Ju, D. Tu, R. Li, H. Zhu, J. Chen, Z. Chen, M. Huang, X. Chen, *J. Am. Chem. Soc.*, **2012**, *134*, 1323–1330.
- [21] L. Ma, N. Sun, J. Zhang, C. Tu, X. Cao, D. Duan, A. Diao, S. Man, *Nanoscale*, **2017**, *9*, 17699–17703.
- [22] O. Boussif, F. Lezoualc'h, M. A. Zanta, M. D. Mergny, D. Scherman, B. Demeneix, J. P. Behr, *Proc. Natl. Acad. Sci. U. S. A.*, **1995**, *92*, 7297–301.
- [23] H. K. Shete, R. H. Prabhu, V. B. Patravale, *J. Nanosci. Nanotechnol.*, **2014**, *14*, 460–74.
- [24] A. Akinc, M. Thomas, A. M. Klibanov, R. Langer, *J. Gene Med.*, **2005**, *7*, 657–663.
- [25] A. Kichler, *J. Gene Med.*, **2001**, *3*, 135–144.

- [26] W. T. Godbey, K. K. Wu, A. G. Mikos, *J. Biomed. Mater. Res.*, **1998**, *45*, 268–275.
- [27] K. Kunath, A. Von Harpe, D. Fischer, H. Petersen, *J. Control. Release*, **2003**, *89*, 113–125.
- [28] M. X. Tang, F. C. Szoka, *Gene Ther.*, **1997**, *4*, 823–832.
- [29] L. Wightman, R. Kircheis, S. Carotta, R. Ruzicka, M. Kursa, *J. Gene Med.*, **2001**, *3*, 362–372.
- [30] R. Kircheis, L. Wightman, E. Wagner, *Adv. Drug Deliv. Rev.*, **2001**, *53*, 341–358.
- [31] S. Mishra, P. Webster, M. E. Davis, *Eur. J. Cell Biol.*, **2004**, *83*, 97–111.
- [32] M. Ogris, S. Brunner, S. Schu, R. Kircheis, E. Wagner, *Gene Ther.*, **1999**, *6*, 595–605.
- [33] U. Lungwitz, M. Breunig, T. Blunk, A. Göpferich, *Eur. J. Pharm. Biopharm.*, **2005**, *60*, 247–266.
- [34] Z. U. Rehman, D. Hoekstra, I. S. Zuhorn, *ACS Nano*, **2013**, *7*, 3767–3777.
- [35] J. Ziebarth, Y. Wang, *Biophys. J.*, **2009**, *97*, 1971–1983.
- [36] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, D. M. J. A., *J. Comput. Chem.*, **2010**, *31*, 671–690.
- [37] C. Sun, T. Tang, H. Uludağ, J. E. Cuervo, *Biophys. J.*, **2011**, *100*, 2754–2763.
- [38] Z. Wei, E. Luijten, *J. Chem. Phys.*, **2015**, *143*, 243146.
- [39] T. A. Beu, A. Farcaş, *J. Comput. Chem.*, **2017**, *38*, 2335–2348.
- [40] C. Sun, T. Tang, H. Uludağ, *J. Phys. Chem. B*, **2012**, *116*, 2405–2413.
- [41] S. Bagai, C. Sun, T. Tang, *J. Phys. Chem. B*, **2014**, *118*, 7070–7076.
- [42] C. Sun, T. Tang, H. Uludağ, *Biomacromolecules*, **2011**, *12*, 3698–3707.
- [43] S. Bagai, C. Sun, T. Tang, *J. Phys. Chem. B*, **2013**, *117*, 49–56.
- [44] A. Y. Kostritskii, D. A. Kondinskaia, A. M. Nesterenko, A. A. Gurtovenko, *Langmuir*, **2016**, *32*, 10402–10414.

- [45] S. John, M. Hussain, U. Turabe, C. Dhand, M. Venkatesh, E. Tze, L. Goh, S. Harini, C. Eugene, R. Rui, S. Ramakrishna, S. Sunder, R. W. Beuerman, C. Shekhar, N. Kumar, X. Jun, R. Lakshminarayanan, *Acta Biomater.*, **2016**, *37*, 155–164.
- [46] S. Yang, S. May, *J. Chem. Phys.*, **2008**, *129*, 185105.
- [47] E. C. Freeman, L. M. Weiland, W. S. Meng, *J. Biomater. Sci. Polym. Ed.*, **2013**, *24*, 398–416.
- [48] W. G. Noid, *J. Chem. Phys.*, **2013**, *139*, 090901.
- [49] J. D. Ziebarth, Y. Wang, *Biomacromolecules*, **2010**, *11*, 29–38.
- [50] J. Ziebarth, Y. Wang, *J. Phys. Chem. B*, **2010**, *114*, 6225–6232.
- [51] A. F. Jorge, R. S. Dias, A. A. C. C. Pais, *Biomacromolecules*, **2012**, *13*, 3151–3161.
- [52] S. J. Marrink, A. H. de Vries, A. E. Mark, *J. Phys. Chem. B*, **2004**, *108*, 750–760.
- [53] S. J. Marrink, D. P. Tieleman, *Chem. Soc. Rev.*, **2013**, *42*, 6801–6822.
- [54] G. P. Tang, H. Y. Guo, F. Alexis, X. Wang, S. Zeng, *J. Gene Med.*, **2006**, *8*, 736–744.
- [55] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, A. H. De Vries, *J. Phys. Chem. B*, **2007**, *111*, 7812–7824.
- [56] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, S. J. Marrink, *J. Chem. Theory Comput.*, **2008**, *4*, 819–834.
- [57] M. Seo, S. Rauscher, R. Pomès, D. P. Tieleman, *J. Chem. Theory Comput.*, **2012**, *8*, 1774–1785.
- [58] J. J. Uusitalo, H. I. Ingólfsson, P. Akhshi, D. P. Tieleman, S. J. Marrink, *J. Chem. Theory Comput.*, **2015**, *11*, 3932–3945.
- [59] G. Rossi, L. Monticelli, S. R. Puisto, I. Vattulainen, T. Ala-Nissila, *Soft Matter*, **2011**, *7*, 698–708.

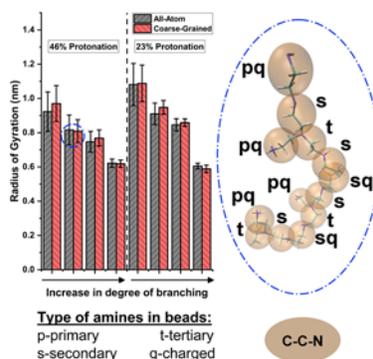
- [60] H. Lee, A. H. De Vries, S. J. Marrink, R. W. Pastor, *J. Phys. Chem. B*, **2009**, *113*, 13186–13194.
- [61] S. O. Yesylevskyy, L. V. Schäfer, D. Sengupta, S. J. Marrink, *PLoS Comput. Biol.*, **2010**, *6*, 1–17.
- [62] J. Michalowsky, L. V. Schäfer, C. Holm, J. Smiatek, *J. Chem. Phys.*, **2017**, *146*, 054501.
- [63] D. H. De Jong, G. Singh, W. F. D. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schäfer, X. Periole, D. P. Tieleman, S. J. Marrink, *J. Chem. Theory Comput.*, **2013**, *9*, 687–697.
- [64] K. Utsuno, H. Uludağ, *Biophys. J.*, **2010**, *99*, 201–207.
- [65] Martini, Martini tutorial: Polymers.
- [66] E. M. Duffy, W. L. Jorgensen, *J. Am. Chem. Soc.*, **2000**, *122*, 2878–2888.
- [67] M. H. Abraham, J. A. Platts, A. Hersey, A. J. Jeo, R. W. Taft, *J. Pharm. Sci.*, **1999**, *88*, 670–679.
- [68] W. Tschöp, K. Kremer, J. Batoulis, T. Bürger, O. Hahn, *Acta Polym.*, **1998**, *49*, 61–74.
- [69] F. Mu, D. Reith, M. Pu, *J. Comput. Chem.*, **2003**, *24*, 1624–1636.
- [70] W. L. Jorgensen, *J. Am. Chem. Soc.*, **1981**, *103*, 335–340.
- [71] B. Hess, C. Kutzner, D. Van Der Spoel, E. Lindahl, *J. Chem. Theory Comput.*, **2008**, *4*, 435–447.
- [72] G. Bussi, D. Donadio, M. Parrinello, *J. Chem. Phys.*, **2007**, *126*, 014101.
- [73] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *J. Chem. Phys.*, **1984**, *81*, 3684–3690.
- [74] A. Parrinello, M. Rahman, *J. Appl. Phys.*, **1981**, *52*, 7182–7190.
- [75] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.*, **1993**, *98*, 10089–10092.
- [76] A. D. MacKerell, N. Banavali, N. Foloppe, *Biopolymers*, **2000**, *56*, 257–265.

- [77] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman, *J. Comput. Chem.*, **1992**, *13*, 1011–1021.
- [78] A. V. V Dobrynin, M. Rubinstein, *Prog. Polym. Sci.*, **2005**, *30*, 1049–1118.
- [79] B. Abdallah, A. Hassan, C. Benoist, D. Goula, J. P. P. Behr, B. A. A. Demeneix, *Hum. Gene Ther.*, **1996**, *7*, 1947–1954.
- [80] H. Lee, R. G. Larson, *J. Phys. Chem. B*, **2006**, *110*, 18204–18211.
- [81] E. Panizon, D. Bochicchio, L. Monticelli, G. Rossi, *J. Phys. Chem. B*, **2015**, *119*, 8209–8216.

# Table of Contents use only

## MARTINI Coarse-Grained Model for Polyethylenimine

*Subhamoy Mahajan and Tian Tang\**



**Summary:** A Martini coarse-grained (CG) model for polyethylenimine was developed based on eight structures with four different degrees of branching and two different protonation states. Extensive parametrization was performed for 99 bonded interactions, which are capable of describing both linear and branched polymer architectures. The model was validated by comparing structural properties predicted by the CG model with those predicted by all-atom models, in absence and presence of salt (KCl).