University of Alberta

Molecular Dynamics Simulations of Polyethylenimine Mediated Nucleic Acid Complexation with Implications for Non-viral Gene Delivery

by

Chongbo Sun

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Mechanical Engineering

Spring 2013
Edmonton, Alberta

To my parents and my wife, for your love and support.

Abstract

Gene therapy is a promising therapeutic technique that involves delivering nucleic acids into cells. Polycations have evolved into a major category of gene carriers. Polyethylenimine (PEI) is one of the most effective polycationic carriers. Furthermore, modifying PEI with certain lipophilic moieties was found to greatly improve its performance. Despite the great potential of PEI-based carriers, the understanding of complexation of nucleic acids with PEIs is still lacking at the atomistic level. In addition, the mechanisms for the beneficial effects of lipid modification remain unclear and to be probed.

In this dissertation, a series of molecular dynamics simulations were performed to investigate the PEI/nucleic acids complexation. We started our simulations from single DNA interacting with single PEI and investigated eight 600 Da PEIs with four different architectures and at two protonation ratios. We found that for these low molecular weight PEIs, compared with the protonation state, the degree of branching has a smaller effect on binding. We then increased the size of the system to incorporate one DNA with multiple PEIs and increased the molecular weight of the PEIs to 2 kDa. Unlike in the case of 600 Da PEIs, the simulations revealed distinct binding modes of branched and linear PEIs to DNA, demonstrating that the molecular weight of PEI is an important factor in PEI/DNA complexation. Following this simulation, complexation/aggregation of DNA molecules medicated by PEIs was studied by simulating multiple DNA molecules with excessive PEIs. We found that native PEIs condense DNA through two mechanisms - polyion bridging and electrostatic screening of the DNA charges. The effects of lipid substitution on polycation mediated nucleic acids aggregation was then explored by adopting lipid-modified PEIs in the simulations of multiple DNAs and siRNAs complexa-

tion. The lipid moieties were found to associate significantly with one another, which provides another mechanism of aggregating nucleic acids and stabilizing the formed polyplexes. The effects of lipid length and substitution level on the formed polyplexes were also investigated. This dissertation will advance the understanding of PEI/nucleic acids polyplexes at atomistic level. Moreover, the methodology adopted suggests a framework for systematically evaluating polycationic carriers using molecular simulations.

Acknowledgements

First and foremost, my sincere gratitude to my supervisor, Dr. Tian Tang, for her guidance, support, care and patience, without which this dissertation would not be possible. Looking back the past five years, I feel so lucky to have the opportunity conducting research with Dr. Tang. The academic training I gained under her supervision will be a lifelong asset for me.

My sincere thanks to our collaborator, my supervisor committee member, Dr. Hasan Uludağ, for his insight, help, support and valuable input in my publications. It is my honour to work with Dr. Uludağ and to gain the insight from the experimental perspective.

Many thanks to Dr. Javier Cuervo for his help in validating the PEI force field parameters using quantum mechanics calculation.

Special thanks to the WestGrid supporting staff and department IT staff for their effective and timely technical support.

Thanks to the developers of CHARMM force field, NAMD package and VMD program, for providing such powerful computational tools.

I am also grateful to my friends at University of Alberta for their companionship and support.

Contents

List of Tables

List of Figures

List of Abbreviations

DNA	DeoxyriboNucleic Acid
RNA	RiboNucleic Acid
siRNA	small interfering RNA
RNAi	RNA interference
PEI	Polyethylenimine
PAMAM	Polyamidoamine
PLL	poly-L-lysine
N	Nitrogen
O	Oxygen
OA	Oleic Acid
CA	Caprylic Acid
LA	Linoleic Acid
HMW	High Molecular Weight
LMW	Low Molecular Weight
MD	Molecular Dynamics
CG	Coarse Grained
COM	Center Of Mass
CHARMM	Chemistry at HARvard Molecular Mechanics

Chapter 1
Introduction [1]

1.1 Background

Gene therapy is an emerging therapeutic technique which involves delivering genetic materials into cells to treat diseases including cancers, hereditary diseases, viral infections, and for immunization purpose [1, 2]. Gene therapy normally uses carrier molecules such as viruses, liposomes and polycations to deliver nucleic acids into cells [2, 3]. Viruses, yet the most common and efficient delivery carriers, are greatly limited in their general use due to the safety concerns [4]. Synthetic polymers, as alternatives to viral carriers, can condense and form nanoparticles with nucleic acids to facilitate the gene delivery, with the advantages of less toxicity, low cost, easiness to produce and versatility for different applications [1, 2]. Polyethylenimine (PEI), as one of the most effective gene delivery polymers, has been intensely studied since 1995 [5, 6]. PEI and PEI-based gene carriers bear potentials to become practical carriers in future clinical usage [7]. Despite the great potential of PEI and PEI-based gene carriers, a detailed understanding of the interaction of nucleic acids with PEIs at atomistic level is still lacking. To better reveal the role of PEI as gene carrier and help design more effective PEI-based carriers, it is of significance to understand the binding and complexation of nucleic acids with PEIs, and interaction between the complexes formed and cell membrane at atomistic level.

Molecular dynamics (MD) simulation has been a useful tool in studying biological systems. It can provide images at atomic resolution, predict the properties of nanoscale systems, and evaluate strategies for designing new systems [8--10]. Recent advances in computer hardware and software as well as new methodologies

[1] Sections 1.1 - 1.3 of this chapter are adapted with permission from: C. Sun, T. Tang, Journal of Adhesion Science and Technology, 2012, DOI:10.1080/01694243.2012.693830, in press.

have further strengthened the power of computer simulations. In this dissertation, we employ all-atom MD to study the complexation of nucleic acids with native and lipid-modified PEIs.

1.2 Challenges in Designing Synthetic Cationic Carriers

The lack of effective carriers is a major impediment for successful applications of gene therapy. To design effective carriers, it is of significance to review the challenges that the carriers will encounter on the delivery pathway. A successful gene delivery involves the following steps [2]. The first step is that the carriers complex with and condense the nucleic acids and protect them from extracellular physical and biochemical degradation before reaching target cells; the next step is the complexes passing through the cell membrane; after the complexes enter the cell, they need to escape from the endosomes and release the nucleic acids from the carriers; the final step is migration across the nuclear membrane and entry into the nucleus in the case of plasmid DNA (pDNA), and targeting the cytosol in the case of small interfering RNA (siRNA). The pathway for successful delivery of DNA is illustrated in Figure 1.1. To successfully deliver nucleic acids into the cell nucleus, gene carriers must overcome a series of obstacles, both extracellular and intracellular. Two excellent detailed reviews can be found in the works of Pouton *et al.* [11] and Pack *et al.* [12]. Here we provide a concise summary.

1.2.1 Extracellular challenges

Gene carriers face a series of barriers in transporting genes to the membrane of target cells. These include binding and condensing nucleic acids and maintaining the complex in solution, remaining stable and surviving in the blood stream, and binding to the target cells [12].

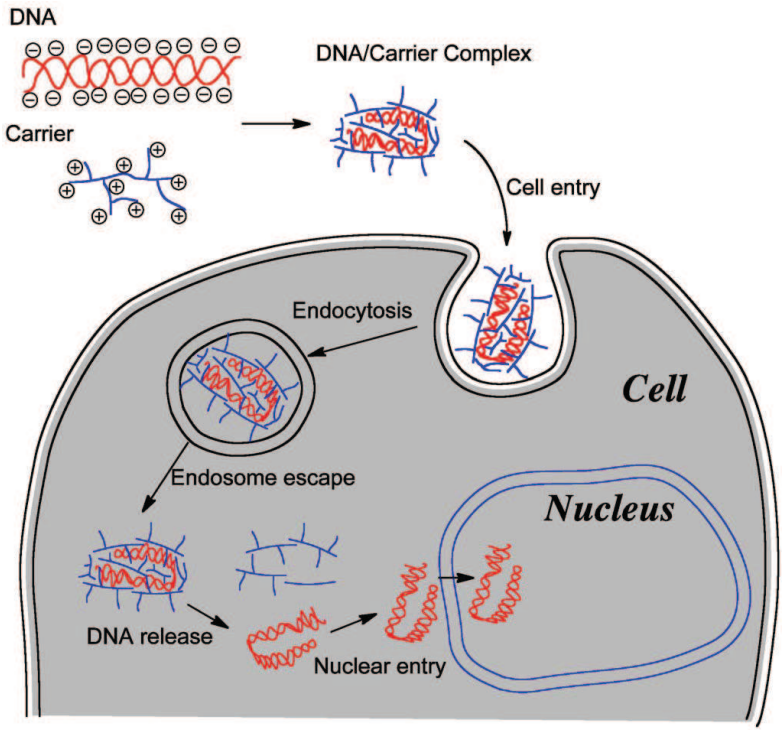

Figure 1.1: Pathway for DNA delivery using carrier.

Nucleic acids condensing and packaging

Unprotected nucleic acids can be degraded by nucleolytic enzymes within minutes, hence the nucleic acids need to be protected from enzymes [12]. Cationic polymers can bind to and condense nucleic acids into small, compact nanoparticles through electrostatic interactions between the negative phosphate groups along the DNA backbone and positive charges on the carriers [13]. The nanoparticles formed are typically toroidal or spherical structures with diameters ranging from tens to several hundreds nanometers [14, 15]. The size and structure of polycationic carriers are known to affect nucleic acids binding and condensation [16, 17]. But the mechanism of the size and structure effect is still not clear. It should be noted that strong binding between carriers and nucleic acids does not necessarily correlate with high gene delivery efficiency, the reason being that tight binding might hinder transcription [12]. An efficient polymer needs to have sufficient binding strength to complex with nucleic acids for protection and have the ability to release the nucleic acids under certain conditions.

Cell uptake and cell targeting

The nanoparticles formed need to be uptaken by the cells in order to function. The uptake depends on the size of the nanoparticle and the charge ratio of polymer to nucleic acids [18]. For some applications like in cancer treatments, where the goal is to kill specific cancer cells, it is required that the therapeutic gene be delivered to a specific type of cells. Primitive polymers normally are not capable of targeting specific cells but may be chemically modified with targeting ligands that can allow cell targeting and even improve cell uptake [19]. The targeting modifications have a broad range of parameters including the conjugation chemistry, the length of spacer between the ligand and the polymers, the ligand-receptor binding strength, and the number of ligands attached to each polymer [12]. These parameters affect the delivery efficiency in a complex way and require careful optimization.

1.2.2 Intracellular challenges

After being internalized into cells, gene carriers have to overcome a new set of intracellular obstacles.

Endolysosomal escape

The carrier/nucleic acids complexes generally enter cells through endocytosis [20], and in the endocytic pathway, the complexes are trafficked to acidic vesicles filled with degradative enzymes [12]. The vesicles at the first stage are termed early endosomes, from which the internalized complexes may be transported back to the membrane and out of the cell by exocytosis [12]. More generally, the complexes stay in the cell and are then trafficked into vesicles termed late endosomes which rapidly acidify to pH 5 [12]. Complexes can subsequently be trafficked into lysosomes, where pH drops to $\sim$4.5 and various degradative enzymes exist [12]. The complexes must escape from these acidic vesicles before being degraded. Certain materials, including PEI and polyamidoamine, are believed to escape the endosomes through the ‘proton-spong’ mechanism [1, 5]. The mechanism hypothesizes that the polymers undergoing large changes in protonation cause influx of protons and counterions into the endocytic vesicles and eventually burst the vesicles [1, 5].

Unpackaging

The complexes can protect nucleic acids from enzymatic degradation, on the other hand the complexes also prevent binding of the nucleic acids with the proteins for gene expression. Polymers must, therefore, release the nucleic acids at the end of the delivery process. It has been found that increased gene expression can be realized by reducing the polymer/nucleic acids binding strength through reducing the number of positive charges [22], or decreasing the polymer molecular weight [16]. To balance the desired strong binding strength at early stage of the delivery process and weaker binding at the late stage, the carriers are ideally to be designed to release the genes by responding to the environmental changes at the action sites.

1.3 Previous Experiments and Simulations

To overcome the above mentioned challenges, it is necessary to understand the complexation of nucleic acids with polycations. Experimental and simulation efforts have been spent to characterize the nucleic acids/PEI complexes. Some of these efforts related to this dissertation are summarized below.

1.3.1 Experimental characterizations

Most experimental attempts to date have focused on studying the transfection efficacy and how it depends on the size and chemical modifications of the PEI [13, 19, 23]. There have also been a few experimental works on the biophysical characterization of PEI/nucleic acids complexes. Choosakoonkriang *et al.* studied the complexes formed by DNA and PEIs of different molecular weights and structures using infrared spectroscopy, scanning calorimetry and isothermal titration calorimetry [24]. They found that DNA remained in the B conformation in all cases; PEIs destabilized the complexes at lower PEI nitrogen to DNA phosphate ratio, but not at higher ratios; no direct correlation was found between the size or zeta potential of the complexes and the molecular weight of the PEI. They also found that the transfection efficiency strongly depends on the ratio of the PEI to DNA and the molecular weight of the PEI. Specifically, 2 kDa PEI did not transfect the cells

and only positively charged complexes transfected the cells. Utsuno and Uludag recently studied the thermodynamics of PEI/DNA complexation in different solutions and at different pH using isothermal titration calorimetry [25]. They found that at low pH PEI had a greater tendency to complex with DNA [25]. They also found that PEI had two modes of binding to DNA -- binding to the DNA grooves and binding to the DNA phosphate backbone.

There also exist several experimental works studying the interaction of DNA with polycations similar to PEI. Prevette *et al.* investigated the binding of plasmid DNA with a series of poly(glycoamidoamine)s (PGAAs) using isothermal titration calorimetry and infrared spectroscopy [26]. They concluded that the binding mechanism was likely a combination of electrostatics and hydrogen bonding, namely that the long-range electrostatic interaction initiated the attraction and the hydroxyl groups in the carbohydrate comonomer further enhanced the association through hydrogen bonding to the DNA base pairs. Lipid modification was found to improve the delivery efficiency of nucleic acids [27]. Patel and Anchordoquy investigated the role of ligand hydrophobicity in DNA condensation by studying the thermodynamics of three different polyamines (cobalt hexamine, spermine and lipospermine) binding to plasmid DNA [28]. The work concluded that there was no significant contribution from hydrophobicity to spermine-DNA binding, while a larger binding affinity was found for lipospermine-DNA binding due to the higher hydrophobicity of lipospermine. The work also postulated that the steric barrier introduced by the acyl chains in lipospermine hampered DNA to be packaged into a highly condensed state.

The above experimental efforts provide information on binding of DNA with polycations and shed light on the design of effective PEI-based gene carriers. Simulations, on the other hand, can provide microscopic view at DNA complexation with polycations and quantify the interplay among various parameters in designing gene carriers.

1.3.2 Coarse grained simulations

Coarse grained (CG) method is an important simulation technique at mesoscopic scale, where, instead of using one particle for each atom as in all atom simulations, several to tens of atoms are coarse grained as a bead. In a CG simulation, water is often treated implicitly. Under these simplifications, often CG method makes it possible to simulate models unmanageable by atomistic models within the current computational capacity. Indeed, CG method bridges atomistic (microscopic) and continuum (macroscopic) levels and can yield some qualitative insight when studying biological systems.

CG models have been employed to study polycation induced DNA condensation. As the atomic representation is lacking in the CG models, the CG simulations are not specific for PEIs and nucleic acids, but rather for general oppositely charged polyelectrolyte molecules. Stevens performed coarse grained molecular dynamics (CG-MD) simulations to investigate the condensation of semiflexible polyelectrolytes in the presence of di-, tri-, and tetra-valent counterions [29]. It was found that a single polyelectrolyte can be condensed into toroidal and rodlike structures by the trivalent and tetravalent counterions, while no condensates form or stay stable for divalent counterions. Winkler *et al.* performed simulations of two flexible, oppositely charged polyelectrolyte chains and found glasslike condensed structures at large Bjerrum lengths [30]. Hayashi *et al.*, in a series of CG Monte Carlo simulations, have studied the complexation of multiple polyanions and polycations with varying chain length, charge density, charge ratio, and salt concentration [31--33]. These simulations demonstrated that the net charge and charge density of the complexes formed were minimized to reduce the electrostatic interaction and the number of the complexes formed was maximized due to entropic effect. Dias *et al.* used CG Monte Carlo simulations to study the polyanion compaction by shorter polycations [34]. These simulations showed that increasing the number of polycations or the number of charges per polycation can lead to greater collapse of the polyanion. Ziebarth *et al.*, in an effort to simulate DNA condensation by PEI-poly(ethylene glycol) diblock copolymers [35], performed CG-MD simulations for the complexation of linear polyanion with block copolymers consisting of cationic and neutral hy-

drophilic blocks. The simulations showed that increasing the cationic block length in the copolymer can result in greater condensation of the polyanion, and the morphology of the complexes formed is dependent on both the size and architecture of the polyelectrolytes.

CG modeling has also shown promise in simulating the interaction between DNA/carrier complex and biological membrane. Voulgarakis *et al.* performed CG Monte Carlo simulations of dendrimer/DNA complexes interacting with a cell membrane [36]. Their results indicated that there is a limited domain in the parameter space where the complex can arrive intact and attach to the membrane long enough for cellular entry to occur. The lifetime of the complex on the membrane depends exponentially on the molecular weight of the polymers, and hence the molecular weight is critical in determining whether a specific delivery system will succeed.

The above mentioned CG-MD and CG Monte Carlo simulations have provided insightful understanding of how chain length, charge density, charge ratio and counterion concentration can influence the polycation/polyanion complexation. However, the oversimplified structures of nucleic acids and polycation cannot yield a quantitative characterization of the complexation. On the other hand, simulations with explicit representation of all atoms permit the investigation of PEI mediated nucleic acid delivery with atomistic details.

1.3.3 Atomistic simulations

At the atomistic scale, Ziebarth and Wang [37] studied the complexation between DNA and linear PEIs using all atom molecular dynamics (MD) simulations, and also made a comparison with complexation between DNA and poly-L-lysine. They found that DNA remained in the B form upon complexation with PEI; the charged amine groups mainly interacted with the DNA phosphate groups; PEI can better neutralize the charge of DNA compared with poly-L-lysine. To the best of our knowledge, this was the only atomic simulation work for DNA/PEI complexes.

In the past five years, there have been several other full atomic simulation works studying similar systems such as polyamidoamine (PAMAM) with DNA or siRNA. Ouyang and coworkers performed MD simulations to study the struc-

ture, dynamics and energetics of siRNA complexation with 6 different polycations including 2 PAMAM dendrimers, 2 dendritic poly-L-lysines and 2 linear poly-L-lysines [38, 39]. They found that all polycations could bind to the siRNA at a low polycation to siRNA charge ratio, while only a fraction of polycations could bind to the siRNA at a high charge ratio. Mills *et al.* carried out MD simulations of DNA--PAMAM dendrimer interaction [40], the calculated free energy as a function of the separation between the DNA and the PAMAM dendrimer agreed very well with the single-molecule pulling experiments. Pavan and coworkers, in a series of papers [41--44], reported MD simulation results of interactions between nucleic acids and dendrimers including PAMAM, UV-degradable dendrons and triazine dendrimers . Their studies demonstrated that the dendrimer flexibility and its ability to reorganize its structure to interact with DNA are important to the binding affinity.

The above MD works all focused on polycations interacting with a single nucleic acid segment at the atomistic scale. And there has not been an atomic simulation work studying the polycation induced nucleic acids condensation. Dai *et al.* however recently performed all atomic MD simulations to study the multivalent-ion mediated attraction between DNA molecules [45]. The simulation confirmed the experimentally observed polyamine-induced DNA attraction, which was explained by the formation of ion bridges between the two DNA molecules. The interaction potential was found to be more attractive for polyamine with larger valence and higher charge density.

In the case of PEI, whether and how the architecture and the protonation state of the PEI may affect the binding structure and energy, how multiple PEIs bind to DNA at different PEI to DNA ratios, how the PEIs mediate the DNA aggregation, and how the complexes formed interact with lipid membrane are still not clear at atomic level.

1.4 Objectives and Organization of Dissertation

The ultimate objective of simulating polycationic gene carriers is to test and screen the designed carriers, thus reducing the amount of experiments required in the carrier development cycle, and hence accelerating the development pace and reducing the development cost. As a first step, this dissertation is devoted to providing a framework of simulating the complexation of nucleic acids with polycationic gene carriers. We took PEI as a representative polycation and performed a series of all-atom MD simulations in ascending complexity on PEI/nucleic acids polyplexes. These simulations consist of studies on the interaction of single DNA with single PEI molecules, that of single DNA with multiple PEI molecules, PEI mediated complexation of multiple DNA molecules, effects of lipid modification on PEI mediated DNA complexation, and effects of lipid modification on siRNA complexation. In analyzing the data, we focused on the structural parameters of the nucleic acid/PEI polyplexes and tried to correlate the structural data with their functions. The remaining of the dissertation is organized as follows.

Chapter 2 presents the simulations of a DNA dodecamer binding with a 600 Da PEI that can have four different degrees of branching and two different protonation states. The simulations revealed the effects of PEI branching and protonation state on the binding pattern of DNA with PEI. Chapter 3 presents the simulations of a 3-fold dodecamer DNA binding with branched and linear 2 kDa PEIs. Instead of simulating the interaction between a single DNA and a single PEI, multiple PEIs were simulated to complex with the DNA at two PEI/DNA N/P ratios. The simulations shed light on the effects of PEI architecture and N/P ratio on the complexation of DNA with 2 kDa PEIs. Chapter 4 presents the simulations of PEI mediated complexation of multiple DNA molecules, which aim to elucidate the mechanisms of polycation induced DNA complexation. In Chapter 5, we adopted a lipid-modified PEI based on 600 Da native PEI and investigated the effect of lipid modification on the structure of the PEI/DNA polyplexes. In Chapter 6 the complexation of siRNA is simulated, which is facilitated by native and four types of lipid-modified 2 kDa PEIs used in experiments. The effects of different lipid modification, in terms of

the length and substitution level of the lipid, on the siRNA complexation were investigated. Simulations and results from these five chapters have been published or submitted for publication. The structures of the published/submitted works are maintained in these chapters, i.e., each chapter has its own introduction, method, results, conclusion and bibliography sections. An overall conclusion of this dissertation and future prospects in simulating the roles of polycationic gene carriers are given in Chapter 7.

Bibliography

[1] F. Ledley, *Hum. Gene Ther.* **1995**, *6*, 1129--1144.

[2] D. Luo, W. Saltzman, *Nat. Biotechnol.* **2000**, *18*, 33--37.

[3] R. Mulligan, *Science* **1993**, *260*, 926--932.

[4] S. Lehrman, *Nature* **1999**, *401*, 517--518.

[5] O. Boussif, F. Lezoualc'h, M. A. Zanta, M. D. Mergny, D. Scherman, B. Demeneix, J. P. Behr, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 7297--7301.

[6] W. Godbey, K. Wu, A. Mikos, *J. Control. Release* **1999**, *60*, 149--160.

[7] D. Jere, H. L. Jiang, R. Arote, Y. K. Kim, Y. J. Choi, M. H. Cho, T. Akaike, C. S. Chot, *Expert Opin. Drug Del.* **2009**, *6*, 827--834.

[8] T. Cheatham, P. Kollman, *Annu. Rev. Phys. Chem.* **2000**, *51*, 435--471.

[9] T. Cheatham, M. Young, *Biopolymers* **2000**, *56*, 232--256.

[10] D. Lu, A. Aksimentiev, A. Shih, E. Cruz-Chu, P. Freddolino, A. Arkhipov, K. Schulten, *Phys. Biol.* **2006**, *3*, S40--S53.

[11] C. Pouton, L. Seymour, *Adv. Drug Delivery Rev.* **2001**, *46*, 187--203.

[12] D. Pack, A. Hoffman, S. Pun, P. Stayton, *Nat. Rev. Drug Discovery* **2005**, *4*, 581--593.

[13] R. Kircheis, L. Wightman, E. Wagner, *Adv. Drug Delivery Rev.* **2001**, *53*, 341--358.

[14] E. Wagner, M. Cotten, R. Foisner, M. Birnstiel, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 4255.

[15] H. Hansma, R. Golan, W. Hsieh, C. Lollo, P. Mullen-Ley, D. Kwoh, *Nucleic Acids Res.* **1998**, *26*, 2481.

[16] D. Schaffer, N. Fidelman, N. Dan, D. Lauffenburger, *Biotechnol. Bioeng.* **2000**, *67*, 598--606.

[17] M. Wolfert, P. Dash, O. Nazarova, D. Oupicky, L. Seymour, S. Smart, J. Strohalm, K. Ulbrich, *Bioconjugate Chem.* **1999**, *10*, 993--1004.

[18] D. Schaffer, D. Lauffenburger, *J. Biol. Chem.* **1998**, *273*, 28004.

[19] D. Schaffert, E. Wagner, *Gene Ther.* **2008**, *15*, 1131--1138.

[20] M. Marsh, *Endocytosis*, M. Marsh (Ed.), Oxford University Press, **2001**.

[21] J. Behr, *Chimia* **1997**, *51*, 34--36.

[22] P. Erbacher, A. Roche, M. Monsigny, P. Midoux, *Biochim. Biophys. Acta* **1997**, *1324*, 27--36.

[23] W. Godbey, K. Wu, A. Mikos, *J. Biomed. Mater. Res.* **1999**, *45*, 268--275.

[24] S. Choosakoonkriang, B. Lobo, G. Koe, J. Koe, C. Middaugh, *J. Pharm. Sci.* **2003**, *92*, 1710--1722.

[25] K. Utsuno, H. Uludag, *Biophys. J.* **2010**, *99*, 201--207.

[26] L. E. Prevette, T. E. Kodger, T. M. Reineke, M. L. Lynch, *Langmuir* **2007**, *23*, 9773--9784.

[27] A. Alshamsan, A. Haddadi, V. Incani, J. Samuel, A. Lavasanifar, H. Uludag, *Mol. Pharm.* **2009**, *6*, 121--133.

[28] M. Patel, T. Anchordoquy, *Biophys. J.* **2005**, *88*, 2089--2103.

[29] M. Stevens, *Biophys. J.* **2001**, *80*, 130--139.

[30] R. Winkler, M. Steinhauser, P. Reineker, *Phys. Rev. E* **2002**, *66*.

[31] Y. Hayashi, M. Ullner, P. Linse, *J. Chem. Phys.* **2002**, *116*, 6836--6845.

[32] Y. Hayashi, M. Ullner, P. Linse, *J. Phys. Chem. B* **2003**, *107*, 8198--8207.

[33] Y. Hayashi, M. Ullner, P. Linse, *J. Phys. Chem. B* **2004**, *108*, 15266--15277.

[34] R. Dias, A. Pais, M. Miguel, B. Lindman, *J. Chem. Phys.* **2003**, *119*, 8150--8157.

[35] J. Ziebarth, Y. Wang, *J. Phys. Chem. B* **2010**, *114*, 6225--6232.

[36] N. K. Voulgarakis, K. O. Rasmussen, P. M. Welch, *J. Chem. Phys.* **2009**, *130*.

[37] J. Ziebarth, Y. Wang, *Biophys. J.* **2009**, *97*, 1971--1983.

[38] D. Ouyang, H. Zhang, D.-P. Herten, H. S. Parekh, S. C. Smith, *J. Phys. Chem. B* **2010**, *114*, 9220--9230.

[39] D. Ouyang, H. Zhang, H. S. Parekh, S. C. Smith, *J. Phys. Chem. B* **2010**, *114*, 9231--9237.

[40] M. Mills, B. Orr, M. M. B. Holl, I. Andricioaei, *Biophys. J.* **2010**, *98*, 834--842.

[41] G. M. Pavan, A. Danani, S. Pricl, D. K. Smith, *J. Am. Chem. Soc.* **2009**, *131*, 9686--9694.

[42] G. M. Pavan, L. Albertazzi, A. Danani, *J. Phys. Chem. B* **2010**, *114*, 2667--2675.

[43] G. M. Pavan, M. A. Kostiainen, A. Danani, *J. Phys. Chem. B* **2010**, *114*, 5686--5693.

[44] G. M. Pavan, M. A. Mintzer, E. E. Simanek, O. M. Merkel, T. Kissel, A. Danani, *Biomacromolecules* **2010**, *11*, 721--730.

[45] L. Dai, Y. Mu, L. Nordenskiöld, J. R. C. van der Maarel, *Phys. Rev. Lett.* **2008**, *100*, 118301.

Chapter 2
Molecular Dynamics Simulations of DNA/PEI Complexes: Effect of PEI Branching and Protonation State [1]

2.1 Introduction

Complexes formed by DNA and synthetic polymers are of great research interest due to their application in gene therapy, which involves delivering genetic materials into cells for therapeutic purposes [1, 2]. This approach offers tremendous hope for patients with cancer, hereditary disease, viral infection, and will potentially change the vaccination technology to an advanced new level [1, 2]. Gene therapy uses carrier molecules such as viruses, synthetic polymers and carbon nanotubes as vehicles to deliver nucleic acids into cell [2--4]. Viruses are yet the most common and efficient delivery carriers. However, their toxicity and immunogenicity greatly limit their general use [5]. Synthetic polymers are an alternative to viral carriers with the advantages of less toxicity, low cost, easiness to produce and versatility for different applications [1, 2]. Polyethylenimine (PEI) is one of the most effective synthetic polymers to deliver nucleic acids into cells through endocytosis [6, 7]. PEI can condense nucleic acids and form nanoparticles via electrostatic interaction between negatively charged nucleic acid phosphate groups and positively charged PEI amine groups. The nanoparticles thus formed can facilitate cellular uptake of the nucleic acids and protect the nucleic acids from degradation during the delivery path. The efficacy of the PEI as a gene delivery vector, however, has been found to depend on the structure and molecular weight of the PEI used [8, 9]. High-molecular-weight (HMW) PEIs (e.g., 25 kDa) can yield a high transfection effi-

[1] A version of this chapter has been published. Reprinted with permission from: C. Sun, T. Tang, H. Uludağ, J. Cuervo, Biophysical Journal, 2011, 100 (11), p2754-2763. Copyright 2011 Elsevier.

ciency but also display high cytotoxicity, while low-molecular-weight (LMW) PEIs (e.g., 600 to 2000 Da) have low cytotoxicity but give poor transfection efficiencies. Cross-linked and grafted LMW PEIs, however, can overcome the high cytotoxicity of HMW PEIs and low transfection efficiency of naked LMW PEIs [10]. While the field is progressing fast in making better PEI-based gene delivery systems, a detailed understanding of the structure and properties of the nucleic acids/PEI complexes is still lacking. It is critical to elucidate the interaction of DNA molecules with carriers at the atomistic level to understand the role of carrier molecules and design more effective DNA/polymer complexes.

To date, experimental studies have focused on studying transfection efficacy using PEIs of different sizes and with different chemical modifications [8, 11, 12]. Utsuno and Uludag [13] recently performed a thermodynamics analysis of PEI/DNA complexes in different solutions and at different pH using isothermal titration calorimetry, and they found that PEI at low pH had a greater tendency to form a complex with DNA. They also concluded that PEI had two modes of binding to DNA, one involves PEI binding to the DNA groove and the other involves external binding of PEI to the DNA phosphate backbone.

On the theoretical front, Ziebarth *et al.* [14] performed all atomic molecular dynamics (MD) simulations of DNA/PEI complexes, where they focused on the formed structures and PEI's ability to neutralize DNA, and made a comparison with the DNA complexed with the poly-L-lysine carrier. To the best of our knowledge, that was the only all atomic simulation work for DNA/PEI complexes. The PEIs employed in Ziebarth's work are in linear form, while branched PEIs are also widely used as a gene delivery vector [9].

The protonability of PEI has been credited for its success as a gene delivery vector [6]. Compared with other polymers such as poly-L-lysine, PEI has a high buffer capacity over a broad range of pH values [6, 15]. It is found that PEI has a much higher protonation ratio of amine groups at low pH than at high pH [16]. Experimentally it has been also found that low pH environment can facilitate transfection [17], presumably affecting the protonation state of PEI. Hence, it is of relevance to study the interaction between DNA and PEIs with different protonation

ratios. In Ziebarth's work [14], two protonation ratios (100% and 50%) were investigated. As reported in most experimental works, the protonation ratio under physiological conditions ranged from 10% to 50% [16, 18--21]. In this work we adopted two protonation ratios (23% and 46%) that are relevant to the gene delivery environment. Commercial PEIs have a large structural dispersity in term of branching. To elucidate whether PEIs with different architectures bind differently to DNA, we have also studied the effect of PEI branching on the complex binding. Specially, we performed all atom molecular dynamics simulations with explicit water and counterions to study the structures formed by the DNA duplex $\mathrm{d(CGCGAATTCGCG)_2}$ and 8 different PEIs. These PEIs have similar molecular weights of $\sim$570 Da, but have 4 different degrees of branching and 2 protonation states of amine groups. Such LMW PEIs are used in the simulations not only because the size of the system that can be simulated in MD is rather limited, but also because LMW PEI based gene delivery vectors have shown increasing promise in practical applications [10, 22]. Through this work, we revealed the effect of degree of branching and protonation states on DNA/PEI binding.

2.2 Methods

2.2.1 Initial structures

The DNA simulated was a Drew-Dickerson dodecamer composed of 24 nucleotides, d(CGCGAATTCGCG)$_2$, which carries a total charge of -22 (fully de-protonated) in physiological solution. The initial structure of this dodecamer was built to be a canonical B form using AMBER NAB tool [23]. Four structures of PEI with different degree of branching were adopted in this work, as shown in Figure 2.1. All four structures consist of 13 amine groups and have a similar molecular weight at around 570 Da. To differentiate the four structures, we introduce the following terminology: 'Purely-Linear (PL)' PEI has 13 amine groups connected in a chain without any branching, representing a linear PEI structure; 'Semi-Linear (SL)' PEI has three short chains, each containing one amine group, distributed nearly uniformly along the primary chain (we refer to the longest chain in the PEI structure as the primary

chain), representing a nearly linear or lightly branched PEI structure; 'Moderately-Branched (MB)' PEI has a short branch with one amine group and a longer branch with two amine groups on the primary chain, representing a moderately branched structure; 'Hyper-Branched (HB)' PEI has four branches each containing one or two amine groups connected to the middle four nitrogens on the primary chain, representing a hyper branched PEI structure. In the remaining part of this chapter, each structure will be referred to by two capital letters, e.g., HB for the hyper-branched structure.

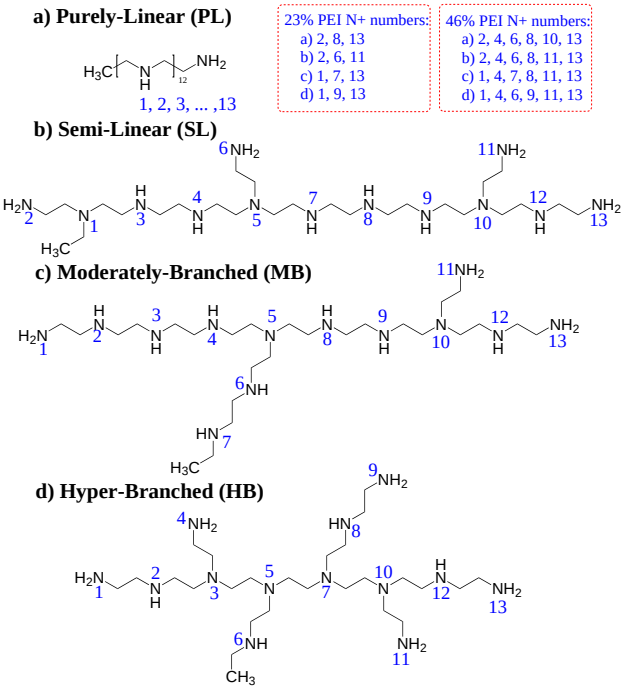

Figure 2.1: Molecular structures, nitrogen numbering (indexed by blue numbers near nitrogens), and protonation sites (specified in red dashed squares) for four PEIs with similar molecular weight but different degree of branching. (a) Purely-Linear structure, (b) Semi-Linear structure, (c) Moderately-Branched structure, (d) Hyper-Branched structure.

There has not been a conclusive value for the protonation ratio of PEI amine groups under physiological conditions. The protonation ratio reported in most experimental works ranged from 10% to 50% [16, 18--21]. Ziebarth *et al.* recently performed a Monte Carlo simulation of linear PEI, where the protonation ratio of PEI amine groups was reported to be approximately 55% under physiological condition with a nearly alternating arrangement of protonated and unprotonated

amines [24]. In their thermodynamics study of 600 Da PEI binding to DNA, Utsuno and Uludag [13] found that 47% PEI amine groups were protonated at pH 6 while 21% were protonated at pH 8. In this work, we chose two protonation ratios close to these values, namely that, one protonation state with 3 out of the total 13 amine groups protonated and the other one with 6 amine groups protonated. We assigned the protonation sites on the primary and secondary amines as they are more nucleophilic (e.g., with higher pK_a) than the tertiary amines [16]. In addition, we assigned the protonation sites as uniformly as possible and separated the neighboring protonation sites as far as possible to minimize thermodynamic interactions between the protonated amines. The uniform distribution of the protonation sites has also been confirmed theoretically [24]. The PEI nitrogen index and protonation sites are illustrated in Figure 2.1. Since the two protonation states correspond to approximately 23% and 46% protonated amines, respectively, in the following sections as well as in tables and figures, we will simply refer to them as '23% systems' (or '23%') and '46% systems' (or '46%').

Separate MD simulations were first carried out for each individual PEI with explicit water and counterions, and the final configurations of these simulations were adopted as the initial configurations for PEIs in the complex formation. Details about the MD simulations are described in Section 2.2.3.

2.2.2 Force field for PEI

CHARMM 27 force field was used for all the molecules in our simulations. However, the residues for PEI do not originally exist in the CHARMM force field. They were devised by adopting parameters from analogous residues available in the CHARMM force field following the CHARMM General Force Field methodology [25]. A comparison was made to Ref. 14, where AMBER force field was used, by repeating a simulation with the same simulation procedure. Similar results were obtained, demonstrating the similarity of these two force fields in describing the DNA/PEI systems. We have further validated the torsional parameters for PEIs by *ab initio* calculations and by repeating two simulations using a different set of torsional parameters reported in Ref. 26. Details about the development and validation

of the force field for PEIs are given in the Appendix A.

2.2.3 Molecular dynamics simulations

The MD simulations were performed using MD package NAMD [27] with the CHARMM 27 force field [28, 29]. TIP3P water model [30], periodic boundary condition, and full electrostatics with particle-mesh Ewald method [31] were used for all MD simulations. A cutoff of 12 Å was used for van der Waals interactions and electrostatics pairwise calculations. All bonds containing hydrogen atoms were constrained (SHAKE algorithm [32]) during all the runs, which allowed us to use a time step of 2 fs.

The configuration of each PEI residue was minimized in NAMD using the devised force field. All the residues for constructing each PEI were then manipulated and glued using VMD [33] and minimized using NAMD to generate a starting configuration for each PEI. This starting configuration was then solvated into a water box with a solvation shell of 16 Å in thickness, and adequate number of Cl- ions were added to the water box to neutralize the system. The system was minimized for 5000 steps to remove bad contact and then gradually heated from 0 K to 300 K in 20 ps. The heated system was equilibrated for 6 ns at 300 K and 1 bar. The final configuration of each PEI was used as the starting configuration for corresponding PEI in the complex simulations.

In constructing the initial system for each of the eight DNA/PEI complexes, the DNA and PEI were first separated by 30 Å and then solvated in a water box with a solvation shell of 18 Å in thickness. Ions (3 Cl- for 23% systems, 6 Cl- for 46% systems and 22 Na+ for all systems) were then added to the water box by randomly replacing the same number of water molecules. During the simulations, the systems that consisted of DNA, PEI, ions and water were first minimized for 2000 steps with the solute atoms fixed, and then 2000 steps with the solute atoms harmonically restrained, followed by 1000 steps of unrestrained minimization. The systems were then heated to 300 K in 20 ps with 10 $\mathrm{kcal}/(\mathrm{mol}\times\mathrm{\AA}^2)$ harmonic restraint on the non-hydrogen solute atoms. The restraint was kept on for another 3 ns at 300 K and 1 bar to allow the counterions to relax around the DNA and PEI. The restraint was

then removed and NPT ensemble simulation was performed for 60 ns for the 23% systems and 40 ns for the 46% systems. The simulation time for the 23% systems was longer than that for the 46% systems because we found it took more time for the complexes in the 23% systems to equilibrate and stabilize, and we will discuss this in the next section. Trajectories were saved every 1000 steps. VMD [33] was used for visualization and trajectories analysis.

2.3 Results

In this section, we demonstrate our simulation results on the flexibility of the PEI, the formation of complexes from 8 different PEI molecules, the deformation of the molecules during the complex formation process, and how the PEIs bind to the DNA at atomic level. Through these results, we discuss how the molecular structure of the PEI and its protonation ratio affect its binding with DNA.

2.3.1 PEI flexibility

Figure 2.2 shows the radii of gyration, R_g, of the eight PEIs in the single PEI simulations over the 6 ns simulation time. It can be seen that HB is the most compact one among the four structures with the smallest R_g, which remains almost constant during the entire simulation. And the degree of ionization does not affect the R_g of the HB PEI. This is explainable as the atoms in the highly branched structure are distributed closer to its center of mass. SL and MB PEIs have similar R_g, which fluctuate more than that of HB PEI, demonstrating that SL and MB PEIs are more flexible than the HB PEI. The R_g of PL fluctuates most among the four PEIs. This is expected as its linear chain configuration makes it the most flexible structure. Intuitively, one may expect that the 46% PEIs should in general have larger R_g than the 23% PEIs as they have a higher charge density and presumably possess a more extended structure. While this is true for the SL and MB structures, our results show that HB PEI has similar R_g at 23% and 46% protonation ratios. This can be attributed to the dendritic structure of the HB PEI, which has resulted in mechanically stiff molecule. Even though the electrostatic repulsion at 46% is larger, it is

not sufficient to cause a clear increase in R_g. In addition, the 23%-PL appears to have a large R_g than the 46%-PL. This may be caused by configuration sampling, as the flexible PL PEI can adopt many equilibrium configurations which may not be sufficiently sampled during the 6 ns MD run.

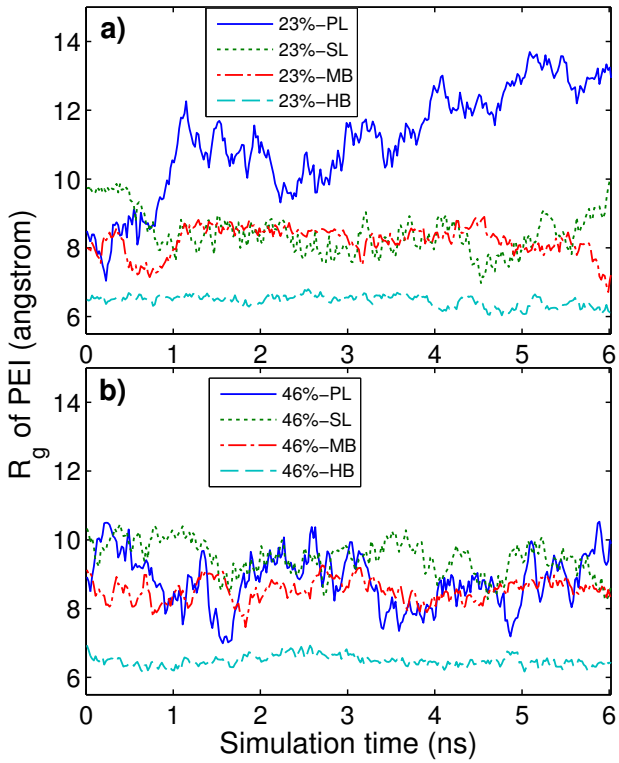

Figure 2.2: Radii of gyration of PEIs in the single PEI simulations. (a) 23% PEIs, (b) 46% PEIs.

The PEI structure after 6 ns of equilibration was used to form complex with DNA. Because of the fluctuation seen in Figure 2.2, the initial PEI configuration for complexation would be different if it was taken at a different time during this equilibration period. However, we do not think the initial configurations of PEIs in the complex simulations would affect the general results reported in this chapter. In fact, we have performed simulations for the 23%-PL system with different initial PEI structures, and similar binding results were obtained. Furthermore, we repeated the simulation for 'System 50%-PEI(50)' described in Ref. 14 and obtained similar results in binding structure, ion distribution and radial distribution function of PEI nitrogens around DNA (see Appendix A). The initial PEI structures in these two

works were very unlikely to be same as the simulations were run separately and with different force fields.

2.3.2 Complex formation

Figure 2.3 shows the configurations of the 8 complexes at the last stage of the simulations. The PL, SL and MB PEIs mainly interact with one strand of DNA and have a significant part of PEI aligning with the DNA backbone. The HB PEIs tend to stay in the DNA major groove and interact with both strands of the DNA. In all the simulations, the DNA preserved its B form with distinguishable minor and major grooves. The Watson-Crick DNA base pairs at the middle of the DNAs remain intact, however, in 5 out of the 8 cases (46%-SL, 23%-MB, 46%-MB, 23%-HB, and 46%-HB), one terminal base pair at one end or two terminal base pairs at two ends of the DNA are broken, as can be seen from Figure 2.3(d-h). The broken bases can in turn attach to the PEIs, as shown in Figure 2.3(d, g). This however does not have a significant effect on the overall binding pattern, as will be discussed in the next subsection.

Figure 2.4 shows the center of mass (COM) distances between the DNA and the PEIs during the complexation process, with time zeroed at the moment the restraints were removed from the solutes. The COM distances all start from 30 Å as the COM of the PEIs were separated by 30 Å from the DNA COM at the beginning of the simulations. For the 23% systems, the COM distances decreases to a series of plateaus after 20 ns, indicating the formation of DNA/PEI complexes. We further define the bound state as a state where a significant part of the PEI is in close contact with the DNA, i.e., there is at most a monolayer of water molecules between the PEI and the DNA. By visually checking the complex structures, all 23% PEIs were found to bind to the DNA within 20 ns. Compared with the 23% protonated PEIs, the 46% protonated PEIs are faster in moving toward the DNA except the 46%-SL PEI. By visually checking the complex structures, we find all 46% PEIs bind to the DNA within 7 ns. This is expected as the electrostatic force, the main driving force for binding, is larger in the 46% systems than that in the 23% systems. In addition, the curves for the 23% systems fluctuate more than their 46% counterparts,

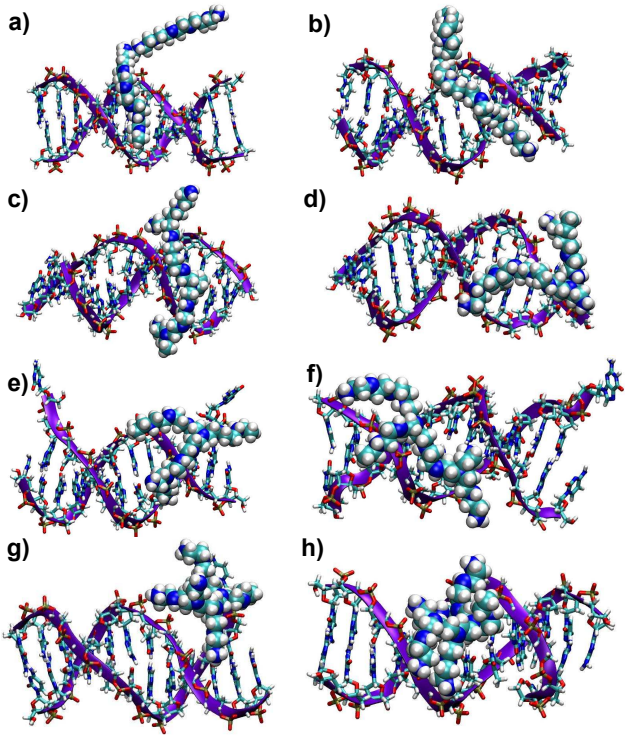

Figure 2.3: Snapshots for each complex at the last stage of the simulations: (a) 23%-PL, (b) 46%-PL, (c) 23%-SL, (d) 46%-SL,(e) 23%-MB, (f) 46%-MB,(g) 23%-HB, (h) 46%-HB.

which is an indication that the complexes formed in the 23% systems to be less stable. Note that shorter COM distance on these plots does not necessarily indicate tighter binding. This is because the PEIs in our simulations are short molecules compared with the DNA, and their locations along the DNA axis can greatly affect the COM distances. This is clear from Figure 2.3(b) and Figure 2.3(d), where the horizontal location of the 46%-PL is much closer to the DNA COM than the 46%-SL. This explains the much larger DNA-PEI COM distance for the 46%-SL shown in Figure 2.4(b).

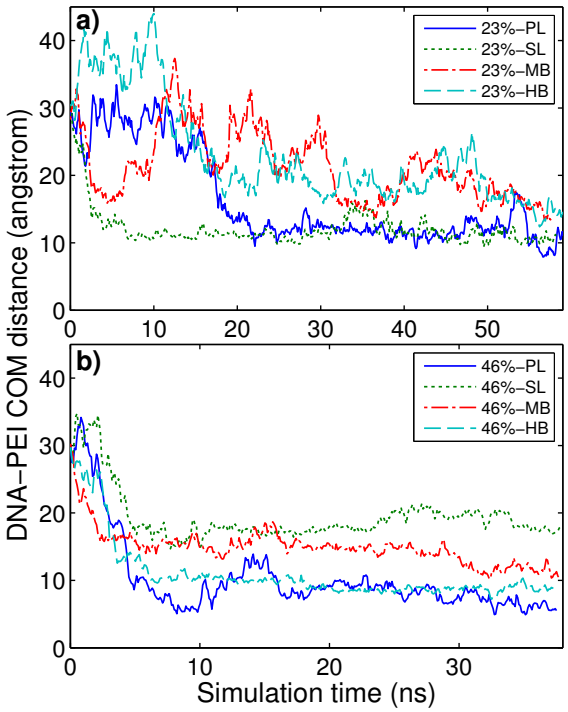

Figure 2.4: Center of mass distance between the DNA and each PEI as a function of simulation time. Time is zeroed at the moment when the restraints were removed from the solutes. (a) 23% systems, (b) 46% systems.

2.3.3 Binding pattern

We then examine how the PEIs bind to the DNA at atomic level. As have been shown in the MD works of Korolev *et al.* [34, 35] on polyamines including spermine, spermidine, putrescine and diaminopropane, the amine groups mainly interact with DNA phosphate groups but can also interact with other electronegative

atoms in the DNA grooves. In this work, we study where and how the PEIs bind to the electronegative atoms of DNA (oxygens and nitrogens), and address the stability of such binding.

Table 2.1 summarizes the average number of PEI nitrogens in close contact (within 4 Å) with DNA electronegative N/O(nitrogens/oxygens) atoms, averaged over the last 20 ns of the simulations. It can be seen that PEI nitrogens in the 46% systems are more probable to be in close contact with the DNA. Except the 46%-MB PEI, the average number of PEI nitrogens within 4 Å of DNA N/O in the 46% systems is more than twice of their counterparts in the 23% systems. Moreover, the PEI predominantly interacts with the DNA backbone oxygens although it also interacts with the DNA base N/O. Note, the summation of numbers in the 'backbone O' and 'Base N/O' columns usually is higher than the number in the 'All N/O' column. This is because some PEI nitrogens can be simultaneously in close contact with the DNA backbone and the base N/O, while such nitrogens were only counted once when calculating the number of PEI nitrogens in close contact with all DNA N/O. We further distinguished the PEI nitrogens that interact with O1P & O2P, O3' & O5', and O4' in the DNA backbone oxygens in Table 2.1. It can be found that for the 46% systems, the PEI nitrogens are much more probable to interact with O1P & O2P atoms than with O3' & O5'. The PEI nitrogens in the 23% systems tend to be almost equally probable to interact with O1P & O2P atoms and with O3' & O5' atoms. Except the 23%-PL PEI, the PEI nitrogens are very unlikely to be in close contact with O4' atoms on the DNA sugar rings, and some nitrogens can interact with multiple backbone oxygens simultaneously.

Figure 2.5 shows the percentage of time during which the individual PEI nitrogens are in close contact (within 4 Å) with the DNA electronegative atoms in the last 20 ns of the simulations. 100% means a nitrogen is within 4 Å of at least one DNA electronegative atoms at all time during the last 20 ns of the simulations, and 0% means a nitrogen is not within 4 Å of any DNA electronegative atoms at all during the last 20 ns of the simulations. Nitrogens in protonated amine groups are marked with '+' for the 23% systems and '*' for the 46% systems in Figure 2.5. Several observations can be made from the figure: First, nitrogens in protonated

Table 2.1: Average number of PEI nitrogens within 4 Å of the DNA electronegative atoms (oxygens and nitrogens) in the last 20 ns simulations.

Systems	All	Backbone	Backbone O			Base
	N/O	O	O1P & O2P	O3' & O5'	O4'	N/O
23%-PL	2.25	2.08	1.00	0.83	0.79	0.71
46%-PL	6.19	6.10	6.09	0.72	0	0.09
23%-SL	2.26	2.10	1.71	1.20	0.25	0.29
46%-SL	5.07	4.72	4.39	0.99	0.28	0.47
23%-MB	2.34	1.96	1.46	1.00	0.23	0.46
46%-MB	3.14	2.97	2.94	0.31	0.02	0.22
23%-HB	2.68	1.82	1.56	0.76	0.02	0.98
46%-HB	5.51	5.12	5.11	0.64	0	0.40

amine groups are generally more probable to be in close contact with the DNA. In fact, out of the 23 nitrogens that are in close contact with the DNA for over 50% of the time only 6 are not protonated. Secondly, nitrogens in the 46% systems are generally more probable to be in close contact to the DNA than their counterparts in the 23% systems. For 23% PL, SL, MB, and HB PEIs, the average percentages of time of PEI nitrogens in close contact with DNA are respectively 17%, 17%, 18%, and 21%. In contrast, the corresponding percentages for the 46% systems are 48%, 39%, 24%, and 42%, respectively. Thirdly, for 46%-PL and 46%-SL, unprotonated nitrogens sandwiched by two protonation sites have higher probability to be in close contact with DNA. Specifically, all the 3 unprotonated nitrogens that are within 4 Å of the DNA for over 50 % of the time are located between two protonated nitrogens. Such an observation is not so clear for 46%-MB and 46%-HB because of their branched structure, nitrogens with neighboring indices may not be located next to each other. Nor is this seen in the 23% systems, since the few protonated nitrogens are located too far apart to strongly affect the unprotonated nitrogens in between.

Let us examine Figure 2.5 together with Figure 2.1 to further explore how the location of the nitrogens might affect their contact with the DNA. For 46%-PL PEI, the PEI nitrogens in the middle of the polymer chain are more probable to be in close contact with the DNA than those at the two ends. While for 23%-PL PEI, those PEI nitrogens at the two ends are more probable to be in close contact

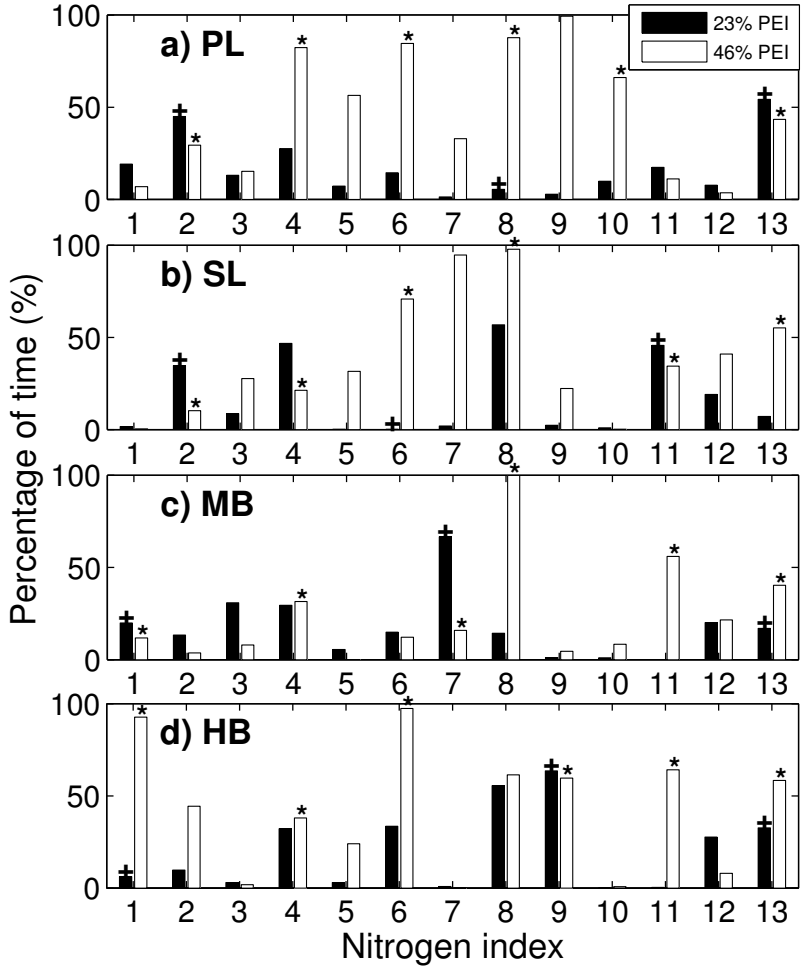

Figure 2.5: Percentage of time during which each individual PEI nitrogen was in close contact (within 4 Å) with any DNA electronegative atoms during the last 20 ns of the simulations. Nitrogen numbering is the same as that in Figure 1. Nitrogens in protonated amine groups are marked with '+' for the 23% systems and '*' for the 46% systems. (a) PL, (b) SL, (c) MB, (d) HB.

with the DNA than those in the middle. The same phenomenon was also observed for SL PEI. A possible explanation to such behavior is that for the 23% PEIs, the electrostatic interaction is not strong enough to cause a large part of the PL or SL chain to be in close contact with the DNA. Having the two ends in close contact with the DNA allows the majority (2 out of 3) of the charges to bind, while giving some flexibility to the middle part of the PEI molecule. In the 46% systems, however, the electrostatic interaction is sufficiently large to cause the majority of the nitrogens in the 46% PEIs, which are located in the middle, to be in close contact with the DNA, leaving the end nitrogens with more fluctuation. Based on this observation, we can make the following conjecture. If a LMW PL or SL PEI is to form a complex with a DNA, at a high protonation ratio such as 46%, the complex might be more stable for longer PEI because of its low percentage of end nitrogens. At a low protonation ratio such as 23%, shorter PEIs might form tighter complex with DNA as higher percentage of end nitrogens are available. The above phenomenon becomes less pronounced as the degree of branching is increased to MB, and disappears for HB, since all the protonation sites are located at the branch ends.

As the PEI nitrogens mainly interact with the DNA backbone oxygens, we plot the radial distribution function (RDF) of the PEI nitrogens around the DNA backbone oxygens in Figure 2.6. Figure 2.6(a) and Figure 2.6(b) are respectively the RDF plots for all PEI nitrogens and for protonated PEI nitrogens around the DNA backbone oxygens in the 23% systems. Figure 2.6(c) and Figure 2.6(d) are the same RDF plots for the 46% systems. These RDF plots were generated based on trajectories of the last 20 ns simulations. In all cases, a step distance of 0.2 Å was used and the curves were normalized by the total number of PEI nitrogens--13. For almost all the RDF curves, there are two predominant peaks, one at around 3 Å and the other at 5 Å from the oxygens. The first peak corresponds to the expected distance for direct contact between the PEI amine groups and the DNA oxygens through hydrogen bonding. The second peak corresponds to the distance for indirect interaction such as hydrogen bonding mediated by one water molecule. For some RDF curves, there exists a less distinct third peak at around 7 Å. This third peak corresponds weak indirect interactions such as hydrogen bonding mediated by two or more wa-

ter molecules. For the 23% systems, it can be seen clearly from Figure 2.6(a) that the second peaks dominate over the first peaks, while in Figure 2.6(b), the first peaks are more pronounced than the second peak. This indicates the protonated PEI nitrogens are more likely to be in direct contact with the DNA, while among all nitrogens, most are in indirect contact with the DNA. For the RDF of the 46% systems shown in Figure 2.6(c) and Figure 2.6(d), it can be seen the profiles of the first peaks from the two figures are almost identical. This indicates that the majority of the PEI nitrogens in direct contact with the DNA are from the protonated amine groups.

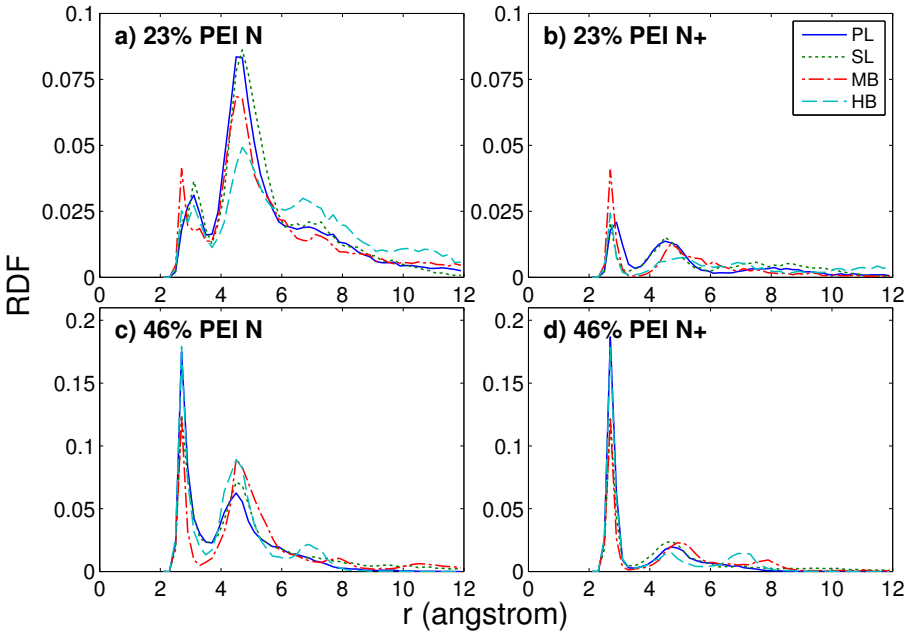

Figure 2.6: Radial distribution function of the PEI nitrogens around the DNA backbone oxygens based on the last 20 ns trajectory of the simulations. (a) 23% all PEI nitrogens, (b) 23% protonated PEI nitrogens, (c) 46% all PEI nitrogens, (d) 46% protonated PEI nitrogens.

To quantify the numbers of PEI nitrogens involved in each peaks of the RDF, we plotted the cumulative number of PEI nitrogens around the DNA backbone oxygens in Figure 2.7. From Figure 2.7(a) for all PEI nitrogens in the 23% systems, it can be seen that approximately 2 PEI nitrogens are within 4 Å of the DNA backbone oxygens for all the four PEI structures, which corresponds to the first peak in Figure 2.6(a). There are about 6 PEI nitrogens at 4--6 Å from the DNA backbone

oxygens. These PEI nitrogens account for approximately half of the total number of PEI nitrogens and make the second peaks in Figure 2.6(a) dominant. Therefore, for the 23% systems, the majority of the PEI nitrogens are in indirect interaction with the DNA at 4--6 Å. By examining the cumulative number in Figure 2.7(b), it can be seen that for the 23% systems, there is about 1 protonated PEI nitrogen within 4 Å of DNA backbone oxygens and about 1 protonated PEI nitrogen at 4--6 Å from the DNA backbone oxygens. Note that in most of the cases shown in Figure 2.7(b), the cumulative numbers sharply increase from zero to a plateau at 2.3 Å. This indicates that the direct contact between protonated amine groups and the DNA is strong hydrogen bonding. For the 46% systems, the fact that the majority of the PEI nitrogens in direct contact with the DNA are from the protonated amine groups can be further confirmed from the cumulative number curves in Figure 2.7(c,d). Specifically, the cumulative number of all PEI nitrogens within 3 Å of the DNA backbone oxygens are very close to that for the protonated PEI nitrogens. About 3 out of 6 protonated PEI nitrogens are in direct contact with DNA, while most unprotonated PEI nitrogens are in indirect interaction with DNA.

To further demonstrate the stability of the formed complexes, we have plotted the RDF and cumulative number curves based on trajectories within different time windows in the simulations (Figures A.26--A.33 in Appendix A). For the 23% systems the figures show that even after 49 ns of simulation, the curves are still evolving with time, and the order of the curves corresponding to different PEI structures do not remain the same at all time. This indicates that the complexes formed in the 23% systems are not stable, which is consistent with the fact that the majority of the nitrogens bind to DNA through indirect interactions. Compared with the 23% systems, the RDF and cumulative number curves for the 46% systems demonstrate more stability (i.e., less variations among different simulation windows). Moreover, the curves corresponding to different PEI structures are closer to one another compared with the 23% systems. In fact, after 40 ns of simulations, these curves essentially overlap with one another. This indicates that the degree of branching has vanishingly small effect on the binding at the protonation ration of 46%.

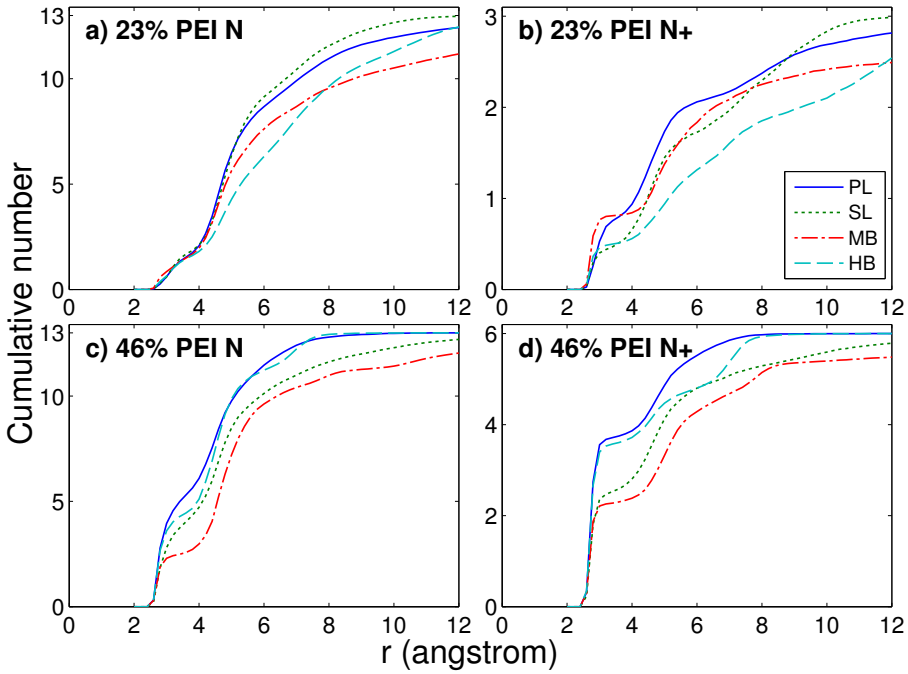

Figure 2.7: Cumulative number of the PEI nitrogens around the DNA backbone oxygens based on the last 20 ns trajectory of the simulations. (a) 23% all PEI nitrogens, (b) 23% protonated PEI nitrogens, (c) 46% all PEI nitrogens, (d) 46% protonated PEI nitrogens.

2.4 Implications and Limitations

2.4.1 Implications

This is the first MD study to investigate the effect of degree of branching and protonation state on PEI binding of DNA. The results shed light on detailed binding mechanism(s) of PEI to DNA, and will help to better understand and design PEI-based gene carriers. A clear outcome from the current study is the beneficial effect of higher PEI protonation state on DNA binding, given by the (i) shorter complex formation time, and (ii) more intimate contact of PEI nitrogens with DNA at the higher protonation state. Changing the pH of a PEI solution is the practical way to enhance the protonation state of PEI, where pH changes from 6.0 to 9.0 was observed to change the fraction of protonated amines from 47% to 13% in our hands [13]. Consistent with the MD results here, a better DNA binding was observed when PEI interaction to DNA was investigated at low pH [13, 17], where the PEI molecule becomes highly protonated without significantly affecting the charge of the DNA. These experimental studies were conducted with $\sim$600 Da branched PEI [13] or $\sim$25,000 Da linear PEI [17], but the role of protonation on DNA binding should be independent of the size and the architecture of the PEI molecule employed. The initial binding constant K_1 (estimated after fitting the titration heat with a single set of identified sites model) was found to be enhanced at lower medium pH for such an interaction [13]. A stronger binding is likely to result in better ability of PEI to deliver extracellular DNA molecules into cells, resulting in better gene expression [17]. Tailoring for a stronger binding is also beneficial when one considers the use of such complexes *in vivo*, where highly bound complexes were shown to be more resilient against degradation [36]. Although the predominant PEI-DNA interaction is expected to be between the electronegative oxygen atoms on the DNA backbone and protonated PEI nitrogens, our simulations also predicted interactions with the DNA base oxygens and nitrogens, implying DNA groove binding of the PEI. This was experimentally shown to be the case in our hands [13] as well as in independent studies [37].

It is known that both linear and branched forms of PEI can complex with DNA

and form particles suitable for cell uptake and gene expression. Experimentally, independent labs comparing DNA binding with linear vs branched PEI have reported stronger DNA binding by the branched PEI [38--41]. The functional consequence of the stronger binding could be better gene expression, due to increased cellular uptake and/or better protection against degradation, but less stable complexes (i.e., complexes formed with linear PEI) may result in better gene expression under some conditions since less stable complexes are also more prone to free the DNA inside the cells and make it available for transcription [42]. In our simulations, a clear trend is not observed for how the binding of the protonated amines with the DNA backbone oxygens is affected by the degree of branching. It is yet to be investigated whether this remains to be the case for PEIs with higher molecular weight. The PEI molecules chosen for the present study had similar molecular sizes, and it is well known that DNA binding is significantly influenced by the size of the PEI as well as its architecture [42]. This issue will be the subject of our future studies to better understand the role of architecture in combination with the molecular size.

2.4.2 Limitations

The PEI molecules simulated in this work are small molecules with low molecular weight. Experimentally what has been demonstrated to be most effective in gene delivery are PEIs with higher molecular weight (e.g., $\sim$25 kDa). However, simulating such large molecules using MD is not practical even with the state-of-art computation capacity. Our results with low molecular weight PEIs are still expected to shed light onto binding in the DNA/PEI complexes, as we believe the binding mechanism at atomistic level is the same for all molecular weight PEIs. In addition, the high toxicity of larger PEI molecules has limited their usage in practical situations, while recent success in delivering nucleic acids with modified low molecular weight PEIs [10, 22] has encouraged us to study the DNA interactions with LMW PEIs.

The current work has focused on single PEI binding with single DNA molecule. When more than one PEI and DNA molecules are present, multiple PEIs can bind to a single DNA segment, and a single PEI can bridge multiple DNA molecules. There

can also be interactions between multiple complexes formed. These are interesting, yet practically related problems to be investigated in our future work.

Finally, it is known that counterion release during the complex formation can play an important role in binding. Counterion release is clearly observed in our simulations on longer PEIs (see Appendix A). However, for smaller PEIs with fewer charges, no distinct correlation has been found between binding and counterion release. Whether increasing salt concentration, i.e., adding more ions in the simulation, will change the scenario needs further exploration.

2.5 Conclusions

In this work, we performed all atomic molecular dynamics simulations of a DNA duplex $d(CGCGAATTCGCG)_2$ with PEIs of 4 different architectural structures and 2 protonation ratios. Our results shed insight on how the degree of branching and protonation state of the PEI will affect the binding to DNA. We found that: (i) The PEIs primarily bind to the DNA backbone through the formation of hydrogen bonding with the backbone oxygens. (ii) The 46% protonated PEIs bind to the DNA mainly through direct hydrogen bonding, while for the 23% protonated PEIs, indirect interaction mediated by water molecules plays an important role in binding. This results in less stable complex formation for the 23% protonated PEIs. Such findings are also consistent with the experimental results where more stable binding is found at low pH [13], since higher protonation ratio is expected at lower pH values. (iii) At 23% protonation ratio, the RDF and cumulative number of PEI nitrogens around DNA backbone oxygens show some difference between the different PEI structures, but we do not observe a systematic trend for such a difference, and the less stable complexation also leads to fluctuations in the behavior of these curves. At 46% protonation ratio, the effect of PEI structure essentially diminishes. In general, our results show that for the LMW PEI structures investigated here, the degree of branching has a smaller influence on the DNA binding than the protonation ratio of the polymers.

Bibliography

[1] F. Ledley, *Hum. Gene Ther.* **1995**, *6*, 1129--1144.

[2] D. Luo, W. Saltzman, *Nat. Biotechnol.* **2000**, *18*, 33--37.

[3] R. Mulligan, *Science* **1993**, *260*, 926--932.

[4] D. Pantarotto, R. Singh, D. McCarthy, M. Erhardt, J. Briand, M. Prato, K. Kostarelos, A. Bianco, *Angew. Chem., Int. Ed.* **2004**, *43*, 5242--5246.

[5] S. Lehrman, *Nature* **1999**, *401*, 517--518.

[6] O. Boussif, F. Lezoualc'h, M. A. Zanta, M. D. Mergny, D. Scherman, B. Demeneix, J. P. Behr, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 7297--7301.

[7] W. Godbey, K. Wu, A. Mikos, *J. Control. Release* **1999**, *60*, 149--160.

[8] W. Godbey, K. Wu, A. Mikos, *J. Biomed. Mater. Res.* **1999**, *45*, 268--275.

[9] L. Wightman, R. Kircheis, V. Rossler, S. Carotta, R. Ruzicka, M. Kursa, E. Wagner, *J. Gene Med.* **2001**, *3*, 362--372.

[10] D. Jere, H. L. Jiang, R. Arote, Y. K. Kim, Y. J. Choi, M. H. Cho, T. Akaike, C. S. Chot, *Expert Opin. Drug Del.* **2009**, *6*, 827--834.

[11] R. Kircheis, L. Wightman, E. Wagner, *Adv. Drug Delivery Rev.* **2001**, *53*, 341--358.

[12] D. Schaffert, E. Wagner, *Gene Ther.* **2008**, *15*, 1131--1138.

[13] K. Utsuno, H. Uludag, *Biophys. J.* **2010**, *99*, 201--207.

[14] J. Ziebarth, Y. Wang, *Biophys. J.* **2009**, *97*, 1971--1983.

[15] M. X. Tang, F. C. Szoka, *Gene Ther.* **1997**, *4*, 823--832.

[16] J. Suh, H. Paik, B. Hwang, *Bioorg. Chem.* **1994**, *22*, 318--327.

[17] Y. Fukumoto, Y. Obata, K. Ishibashi, N. Tamura, I. Kikuchi, K. Aoyama, Y. Hattori, K. Tsuda, Y. Nakayama, N. Yamaguchi, *Cytotechnology* **2010**, *62*, 73--82.

[18] J. Nagaya, M. Homma, A. Tanioka, A. Minakata, *Biophys. Chem.* **1996**, *60*, 45--51.

[19] A. von Harpe, H. Petersen, Y. Li, T. Kissel, *J. Control. Release* **2000**, *69*, 309--322.

[20] G. Koper, R. van Duijvenbode, D. Stam, U. Steuerle, A. Borkovec, *Macromolecules* **2003**, *36*, 2500--2507.

[21] S. Choosakoonkriang, B. Lobo, G. Koe, J. Koe, C. Middaugh, *J. Pharm. Sci.* **2003**, *92*, 1710--1722.

[22] A. Alshamsan, A. Haddadi, V. Incani, J. Samuel, A. Lavasanifar, H. Uludag, *Mol. Pharm.* **2009**, *6*, 121--133.

[23] D. Case, T. Darden, I. T.E. Cheatham, C. Simmerling, J. Wang, R. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. Merz, B.Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvary, K.F.Wong, F. Paesani, J. Vanicek, X. Wu, S. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. Mathews, M. Seetin, C. Sagui, V. Babin, P. Kollman, *AMBER 10*, University of California, San Francisco, **2008**.

[24] J. D. Ziebarth, Y. Wang, *Biomacromolecules* **2010**, *11*, 29--38.

[25] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. MacKerell, Jr., *J. Comput. Chem.* **2010**, *31*, 671--690.

[26] H. Dong, J. Hyun, C. Durham, R. Wheeler, *Polymer* **2001**, *42*, 7809--7817.

[27] J. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. Skeel, L. Kale, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781--1802.

[28] B. Brooks, R. Bruccoleri, D. Olafson, D. States, S. Swaminathan, M. Karplus, *J. Comput. Chem.* **1983**, *4*, 187--217.

[29] A. MacKerel Jr., C. Brooks III, L. Nilsson, B. Roux, Y. Won, M. Karplus, *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*, Vol. 1 of *The Encyclopedia of Computational Chemistry*, P. v. R. Schleyer et al. (Ed.), John Wiley & Sons: Chichester, **1998**, pp. 271--277.

[30] W. Jorgensen, *J. Am. Chem. Soc.* **1981**, *103*, 335--340.

[31] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089--10092.

[32] J. Ryckaert, G. Ciccotti, H. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327--341.

[33] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33--38.

[34] N. Korolev, A. Lyubartsev, A. Laaksonen, L. Nordenskiold, *Biophys. J.* **2002**, *82*, 2860--2875.

[35] N. Korolev, A. Lyubartsev, A. Laaksonen, L. Nordenskiold, *Eur. Biophys. J.* **2004**, *33*, 671--682.

[36] P. Mullen, C. Lollo, Q. Phan, A. Amini, M. Banaszczyk, J. Fabrycki, D. Wu, A. Carlo, P. Pezzoli, C. Coffin, D. Carlo, *Biochim. Biophys. Acta, Gen. Subj.* **2000**, *1523*, 103--110.

[37] Y. Zhou, Y. Li, *Spectrochim. Acta A* **2004**, *60*, 377--384.

[38] F. P. Seib, A. T. Jones, R. Duncan, *J. Control. Release* **2007**, *117*, 291--300.

[39] J. Intra, A. K. Salem, *J. Control. Release* **2008**, *130*, 129--138.

[40] L. D. Hahn, H. Kong, D. J. Mooney, *Macromol. Biosci.* **2010**, *10*, 1210--1215.

[41] A. Kwok, S. L. Hart, *Nanomed.* **2011**, *7*, 210--219.

[42] M. Bertschinger, G. Backliwal, A. Schertenleib, M. Jordan, D. L. Hacker, F. M. Wurm, *J. Control. Release* **2006**, *116*, 96--104.

Chapter 3
Molecular Dynamics Simulations for Complexation of DNA with 2 kDa PEI Reveal Profound Effect of PEI Architecture on Complexation [1]

3.1 Introduction

Complexation between DNA and synthetic polycations has drawn great interest due to the applications of synthetic polycations as gene carriers [1, 2]. Among the polycations, polyethylenimine (PEI) is one of the most effective synthetic molecules serving as gene carriers [3, 4]. It can condense nucleic acids into nanoparticles, which can facilitate the cellular uptake of nucleic acids and protect the nucleic acids from degradation during the delivery process. It was found that the efficacy of PEI as a gene carrier depends on the structure and molecular weight of the PEI [5, 6]. High molecular weight (HMW) PEIs (i.e., $>$25 kDa) are more efficient in DNA delivery but also display high cytotoxicity. On the contrary, low molecular weight (LMW) PEIs (e.g., 1--5 kDa) display low cytotoxicity but are also less efficient. Modifying LMW PEIs, for example, through lipid substitution [7, 8] or disulfide cross-linking [9], can overcome the limitations of DNA delivery efficiency. PEI of 2 kDa, in particular, is a good platform for such modifications, and some modified 2 kDa PEIs have been proved to be as effective as or even more effective than 25 kDa PEIs for gene transfection [10, 11]. It is therefore of great interest to investigate the binding of LMW PEIs to DNA in order to elucidate their roles in the delivery process.

[1] A version of this chapter has been published. Reprinted with permission from: C. Sun, T. Tang, H. Uludağ, Journal of Physical Chemistry B, 2012, 116 (8), p2405-2413. Copyright 2012 American Chemical Society.

Commercial PEIs have a large structural diversity and are usually categorized into two basic forms, linear and branched. Linear PEIs (lP) are composed of almost all secondary amines while branched PEIs (bP) consist of primary, secondary, and tertiary amines. Both lP and bP have been adopted in gene delivery and transfection studies [6, 12--15]. Experiments on transfection using lP and bP showed that the structural difference could affect the transfection efficiency and sometimes to a significant degree [6, 12--15]. However, there has not been a clear conclusion as to whether lP or bP is more effective as gene carriers. Experimental efforts have also been made to elucidate the relationship between transfection and PEI/DNA complexation. Itaka *et al.* investigated the intracelluar trafficking and DNA release of lP and bP formed polyplexes [13]. They found that the bP/DNA polyplexes were more stable and the DNA could be kept in a condensed state even after 24 hours, while lP/DNA polyplexes could be quickly decondensated and yield a considerably higher and faster gene expression. Their atomic force microscopy results also revealed more effective condensation of DNA by bP than by lP, supporting the restricted release of DNA from bP/DNA polyplexes. Dai *et al.* recently studied the complexation, decondensation, transfection efficiency, and cellular uptake of lP/DNA and bP/DNA polyplexes at different N/P ratios [15]. Their results further confirmed the higher capacity of bP in condensing DNA and the better capacity of lP in releasing DNA from the polyplexes. Despite these experimental findings, the underlying mechanism of the structure-function relationship for PEI-based carriers remains to be probed at the atomistic level.

To understand the role of carrier molecules and to design more effective PEI-based gene carriers, it is crucial to gain a detailed understanding of the complexation of DNA with PEIs at atomistic level. With the fast growing computational capacity, simulating the complexation of nucleic acids and polycations in all-atom representation is becoming feasible. For example, Ziebarth *et al.* [16] simulated the complexation of DNA with linear PEIs (900 Da and 1700 Da) and compared it with the complexation between DNA and poly-L-lysine. They found that DNA remained in the B form in the DNA/PEI complex, the charged PEI amines mainly interacted with the DNA phosphate groups, and PEI had a higher capability in neu-

tralizing DNA than poly-L-lysine. Pavan *et al.* reported a series of molecular dynamics (MD) simulations on the complexation of DNA with dendrimers including PAMAM, UV-degradable dendrons and triazine dendrimers [17--20]. The simulation results showed that the flexibility of dendrimers and their ability to reorganize their structures to comply with DNA structure were important for binding affinity. The complexation of DNA with 600 Da PEIs of different architecture and protonation state was explored in a recent work [21]. It was found that the protonation state of the PEI greatly affected the PEI's complexation with the DNA. In particular, the binding for 46% protonated PEIs was achieved mainly through direct interaction between the protonated amines on the PEI and the electronegative oxygens on the DNA backbone. For the 23% protonated PEIs, however, indirect interaction mediated by water molecules played an important role in binding. Four PEI architectures were simulated with increasing degree of branching, but no strong influence was found on the complexation of these LMW PEIs with the DNA. The above studies have demonstrated the power of atomistic simulations in investigating complexation of nucleic acids with polycations and revealed important details that are not readily accessible by experimental techniques.

In this work, we performed a series of large scale all-atom MD simulations to study the complexation of DNA with 2 kDa PEIs. How the PEI's architecture might influence its complexation with the DNA is a main focus of this study. While previous MD results [21] demonstrated insignificant effects of branching for LMW (600 Da) PEIs, our simulations below for two 2 kDa PEI molecules with different architectures (representing a lP and a bP, respectively) show that the scenario is considerably different. Experimentally, the existence of free PEIs at high N/P ratios has been shown to contribute to cellular uptake and transfection [15, 22]. This underlines the importance of incorporating different PEI/DNA ratios in the simulations. In our simulations for both lP and bP, we have used two PEI/DNA and hence two N/P ratios. For transfection purposes, the DNA/PEI complexes are usually prepared without salt or with 154 mM NaCl to mimic physiological osmolarity. For both the lP and bP, we performed simulations at both zero and 154 mM salt concentrations. Our results reveal the different binding characteristics of lP and bP

in binding to DNA and the effect of salt concentration on the complexation.

3.2 Methods

3.2.1 Simulated systems and procedure

The DNA simulated is a 3-fold Drew-Dickerson dodecamer d((CGCGAATTCGCG)$_3$) composed of 72 nucleotides carrying a total charge of -70 in the fully deprotonated state. The initial structure of the DNA was built to be a canonical B form using the AMBER NAB tool [23]. Two types of PEIs in branched and linear forms were simulated, each consisting of 43 amine groups with a molecular mass of 1874 Da. The chemical structures and protonation sites of the two PEIs are shown in Figure 3.1. Twenty amine groups were chosen to be protonated for each PEI type, corresponding to a protonation ratio of 47% on experimentally determined value at pH = 6 [24]. The protonation sites were assigned to only the primary and secondary amines and were arranged as uniformly as possible to minimize thermodynamic interactions among the protonated amines. An MD simulation was first performed for each PEI with explicit water and 20 Cl- counterions. The structure of each PEI at the end of the simulation was adopted as the initial configuration for PEIs in the complex formation simulations.

Eight systems were simulated to study the complexation of the DNA with multiple PEIs, four of which contain one DNA and four PEIs (DNA/PEI number ratio = 1/4) and the other four contain one DNA and eight PEIs (DNA/PEI number ratio = 1/8). They correspond to N/P ratios (ratio of the total number of N atoms on PEIs to the number of DNA phosphates) of $\sim$2.5 and $\sim$5, respectively. At each DNA/PEI number ratio, the two PEI architectures and two salt concentrations (0 and 154 mM) were simulated. The information on the eight systems, together with the two systems involving individual PEIs is summarized in Table 3.1. In the remaining part of this chapter, each system will be referred by its name in the first column of Table 3.1. The systems with 154 mM salt are designated with '-S' in their names to be distinguished from the systems with zero salt. In systems D-4bP, D-4bP-S, D-4lP and D-4lP-S, the DNA/PEI charge ratio is 7/8; in systems D-8bP, D-8bP-S,

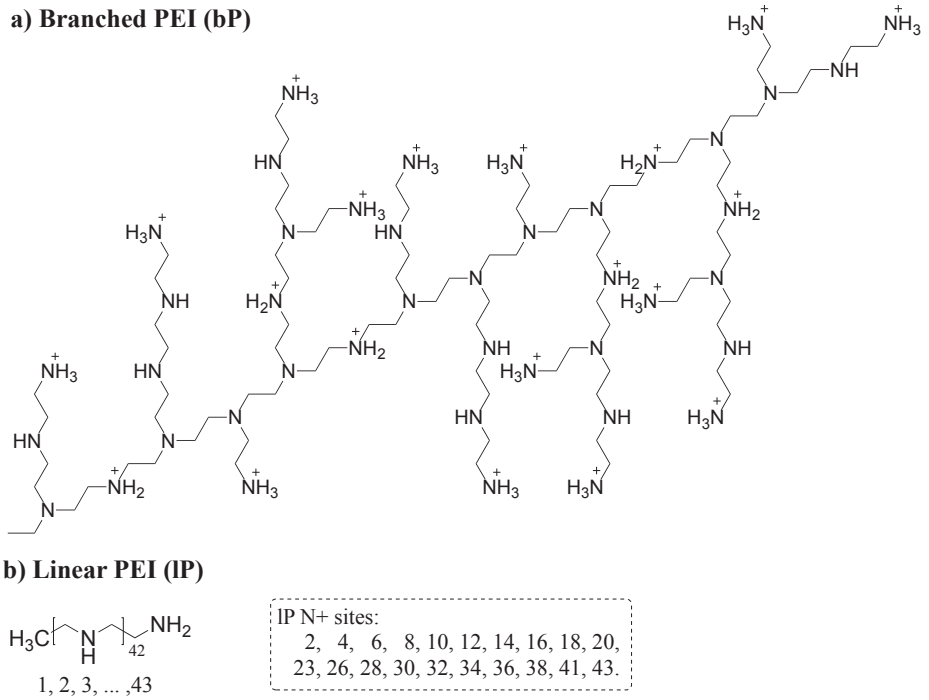

Figure 3.1: Molecular structure and protonation sites of the PEIs studied: (a) bP, (b) lP.

D-8lP and D-8lP-S, the DNA/PEI charge ratio is 7/16. An overall cationic charge is chosen for the DNA/PEI complexes since that better represents the charge of complexes used for experimental purposes. For the zero salt simulations, only neutralizing Cl- ions were added to account for the difference between DNA and PEI charges. At the salt concentration of 154 mM, additional Na+ and Cl- ions of equal amount were added to the solution, and only these additional ions were counted in the calculation of salt concentration. In constructing the initial configurations for each of the eight systems involving complex formation, the principal axes of the PEIs were initially aligned parallel to the DNA axis, and the center of mass (COM) of each PEI was positioned at 25 Å away from the DNA axis. Detailed arrangement of the initial configurations are illustrated in Figure 3.2.

3.2.2 Simulation details

CHARMM 27 force field [25, 26] was used for all molecules except for the PEIs as the force field for PEI is not available in CHARMM. A CHARMM format force

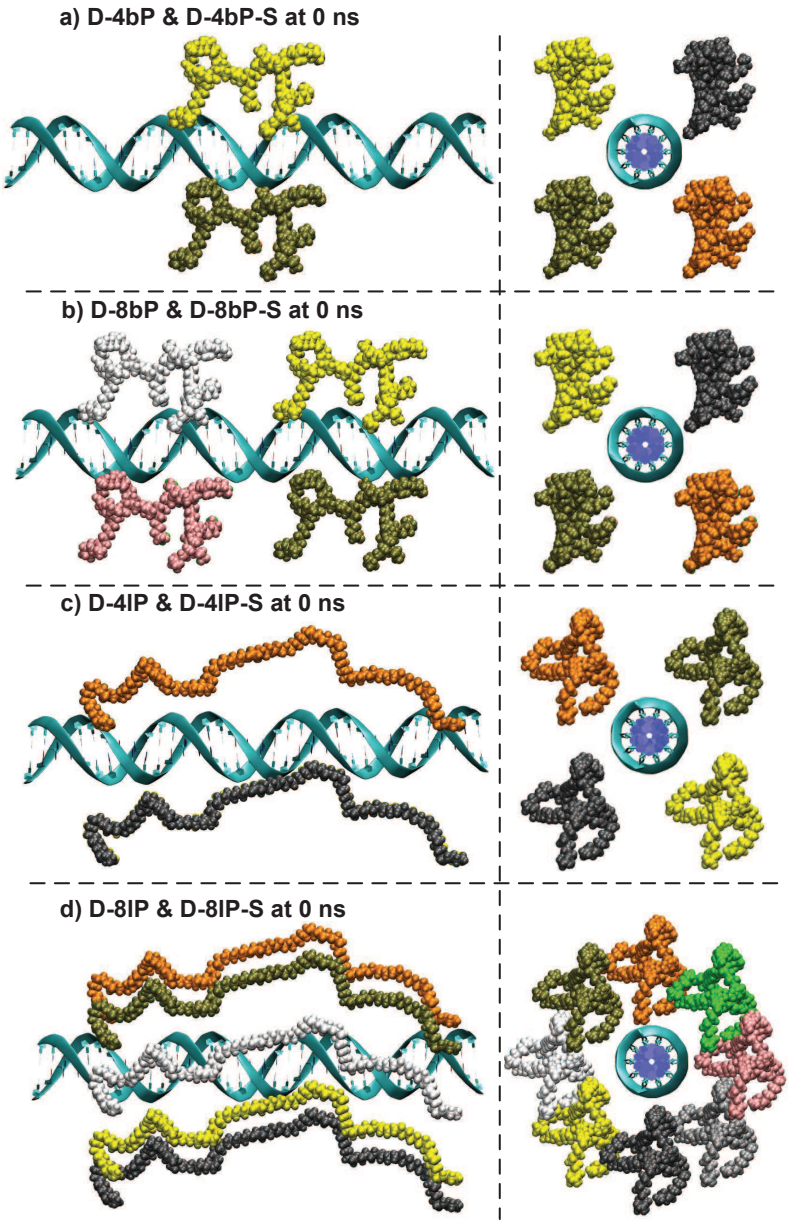

Figure 3.2: Initial configurations of the systems: (a) D-4bP & D-4bP-S, (b) D-8bP & D-8bP-S, (c) D-4lP & D-4lP-S, (d) D-8lP & D-8lP-S. Left panel -- side view, right panel -- axis view. Different PEIs are represented in different colors; water and ions are removed for clarity. Note that because all the PEIs in each model have identical initial configurations, when viewed from a particular direction, some of the PEI molecules may be covered by other PEIs and thus are not visible in certain subfigures.

Table 3.1: Information of the 10 systems simulated.

System name	No. of DNA/PEI	N/P ratio	DNA/PEI charges	No. of atoms	Simulation box size ($Å^3$)	Ion (mM)	Time (ns) restr.+free
bP	0/1	N/A	0/20	37160	$58 \times 69 \times 92$	0	0.2+50
D-4bP	1/4	2.5	70/80	131789	$88 \times 93 \times 158$	0	4+200
D-4bP-S	1/4	2.5	70/80	131297	$88 \times 93 \times 157$	154	50+200
D-8bP	1/8	5	70/160	130321	$88 \times 93 \times 157$	0	4+200
D-8bP-S	1/8	5	70/160	129841	$88 \times 93 \times 157$	154	50+200
lP	0/1	N/A	0/20	215834	$123 \times 150 \times 117$	0	0.2+100
D-4lP	1/4	2.5	70/80	131666	$88 \times 93 \times 158$	0	4+200
D-4lP-S	1/4	2.5	70/80	131174	$88 \times 93 \times 157$	154	50+200
D-8lP	1/8	5	70/160	130039	$88 \times 93 \times 157$	0	4+200
D-8lP-S	1/8	5	70/160	129559	$88 \times 93 \times 157$	154	50+200

field was devised for the PEIs based on the CHARMM General Force Field [27], which has been validated through *ab initio* calculations, a study on sensitivity of MD results to torsional parameters, and comparison with previous works [21]. All simulations were performed using the MD package NAMD [28]. TIP3P water model [29], periodic boundary condition, full electrostatics with particle-mesh Ewald method [30], cutoff distance 10 Å for van der Waals interactions and electrostatics pairwise calculations, SHAKE algorithm [31] to constrain all bonds containing hydrogens, and a time step of 2 fs were used for all of the simulations.

For each system, the DNA and PEIs were solvated into a water box, the size of which is large enough so that the solutes are at least 36 Å away from their nearest periodic images in each direction. Cl- ions to neutralize the system and NaCl salt ions for the systems in 154 mM salt concentration were then added to the water box by randomly replacing equivalent amount of water molecules using VMD [32]. During each simulation, the system was first minimized for 5000 steps, then heated from 0 K to 300 K in 20 ps with 10 $\mathrm{kcal}/(\mathrm{mol}\times Å^2)$ harmonic restraint on the nonhydrogen atoms of the solute. The restraint was kept on for a specific period (0.2 ns for bP and lP, 4 ns for the four complex formation simulations with zero salt, and 50 ns for the four complex formation simulations with 154 mM salt) at 300 K and 1 bar to relax the ions around the solutes. The restraint was kept longer for systems with larger amounts of ions to allow them to relax. The restraint was then removed and NPT ensemble simulation was performed for 50 ns for bP, 100 ns for lP, and 200 ns for complex formation simulations. A total length of 1966.6 ns trajectory

was generated from the 10 simulations. VMD [32] was used for visualization and trajectories analysis.

3.3 Results and Discussions

3.3.1 PEI flexibility

Figure 3.3 shows the radii of gyration, R_g, of the bP and lP in the single PEI simulations. It can be seen that bP has a much smaller R_g ($\sim$ 12 Å) than lP. This is expected as the atoms in the branched PEI are distributed closer to its center of mass in a dendritic manner. In addition, R_g of bP remains almost constant during the entire 50 ns of the simulation, which implies that the bP molecule undergoes very little deformation in the simulation. This is a result of both the inflexible dendritic structure of bP and the fact that each PEI N+ in bP is closely surrounded by several other PEI N+, and hence any large configurational change from the equilibrated structure will introduce a significant energy penalty. On the contrary, R_g of lP fluctuates significantly during the 100 ns simulation, demonstrating the high flexibility of lP.

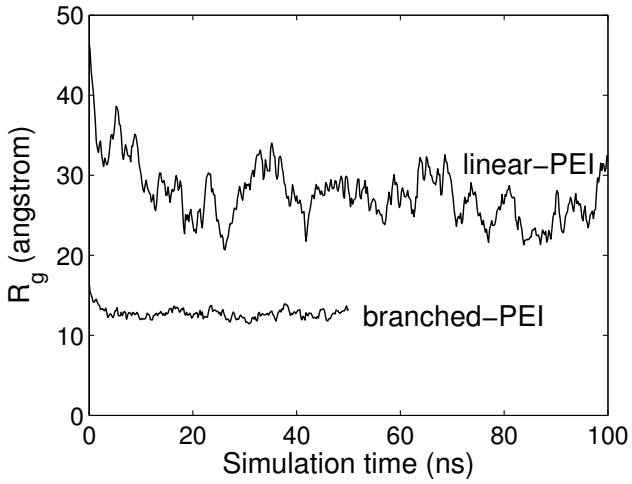

Figure 3.3: Radius of gyration of the PEIs as a function of simulation time.

The flexibility of polycations is known to play important roles in their binding with DNAs and RNAs [17, 18, 20]. Through MD simulations, Pavan *et al.* showed that the flexibility of dendrimers and their ability to reorganize their structure for

interactions with siRNA significantly affect the binding affinity [18]. They found that rigid dendrimers can reorganize their peripheral groups to generate a larger number of contacts to the nucleic acid and display higher affinity than flexible dendrimers [20]. As will be seen in the subsections below, by studying the dynamics of the PEI/DNA complexation, we also observed distinct binding configurations for linear and branched PEIs, which mainly resulted from the difference in flexibility.

3.3.2 Dynamics of the complex formation

By visually examining the simulation trajectories, we found that for the 4 systems with one DNA and four PEIs, all the PEIs move toward the DNA swiftly during the initial several ns after the restraint is removed and each PEI established significant contacts with the DNAs within 20 ns. For the 4 systems with one DNA and eight PEIs, the speed of the PEIs moving toward the DNA was slightly slower. In systems D-8bP and D-8bP-S, all PEIs established significant contacts with the DNA within 20 ns, while in systems D-8lP and D-8lP-S, some PEIs did not bind to the DNA even at the end of the simulation. Figure 3.4 shows the configurations of the eight systems at the final stage of the simulations. It can be seen that for systems D-4bP, D-4bP-S, D-8bP, D-8bP-S, D-4lP and D-4lP-S, all the PEIs bind to the DNA with a significant part of the molecules complying with the DNA. However, the scenario for systems D-8lP and D-8lP-S was different. In D-8lP, two PEIs have only a small fraction of the molecules in contact with the DNA, and in D-8lP-S two PEIs are completely separated from the DNA.

To quantify the dynamics of the interaction of PEIs with the DNA, we plotted the binding state of individual PEIs to the DNA in terms of the number of Ns from each PEI in close contact with the DNA (i.e., within 4 Å of any N/O atoms of the DNA) as a function of simulation time. Four angstroms was chosen as the cutoff distance because this is the distance within which a PEI amine group can form a direct hydrogen bond with a DNA [21]. Figure 3.5 summarizes the results for the four systems with 4 PEIs and Figure 3.6 for the four systems with 8 PEIs. Figures 3.5 and 3.6 also provide the numbers of Ns from all the PEIs that are in close contact with the DNA. Each subfigure in Figures 3.5 and 3.6 contains two curves

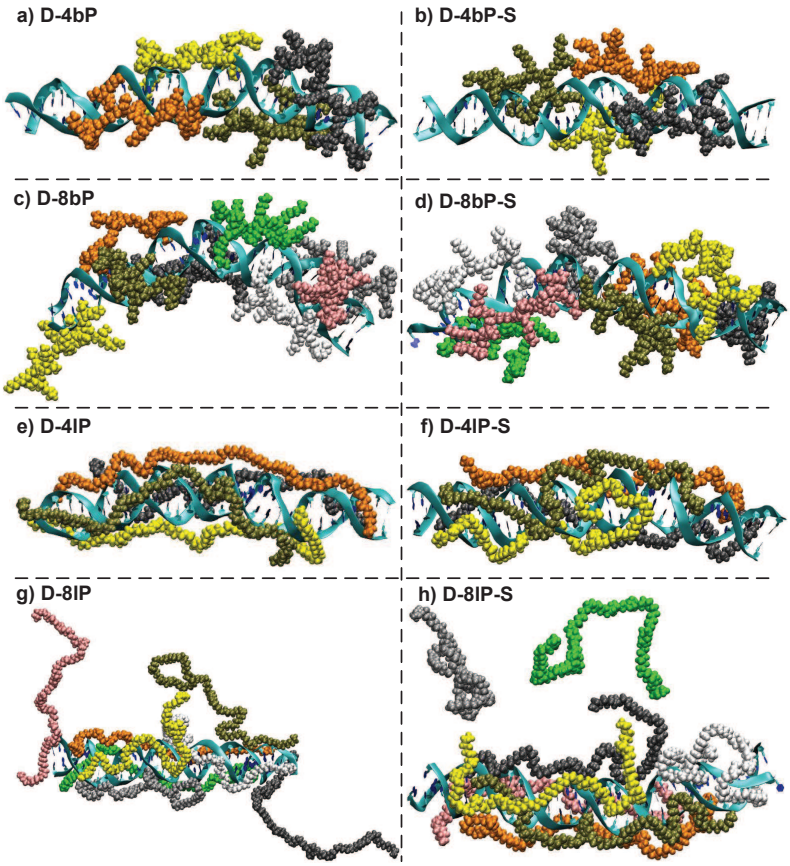

Figure 3.4: Snapshots of the systems at the final stage of the simulations: (a) D-4bP, (b) D-4bP-S, (c) D-8bP, (d) D-8bP-S, (e) D-4lP, (f) D-4lP-S, (g) D-8lP, (h) D-8lP-S. Different PEIs are represented in different colors; water and ions are removed for clarity.

associated with different salt concentrations. In Figure 3.5, each curve in the top 8 subfigures corresponds to one of the 4 PEIs in a particular system. Each curve in the bottom 2 subfigures describes the total number of Ns of all the 4 PEIs in close contact with the DNA in a particular system. Similarly, in Figure 3.6, each curve in the top 16 subfigures corresponds to one of the 8 PEIs in a particular system, and each curve in the bottom 2 subfigures corresponds to all the 8 PEIs in a particular system.

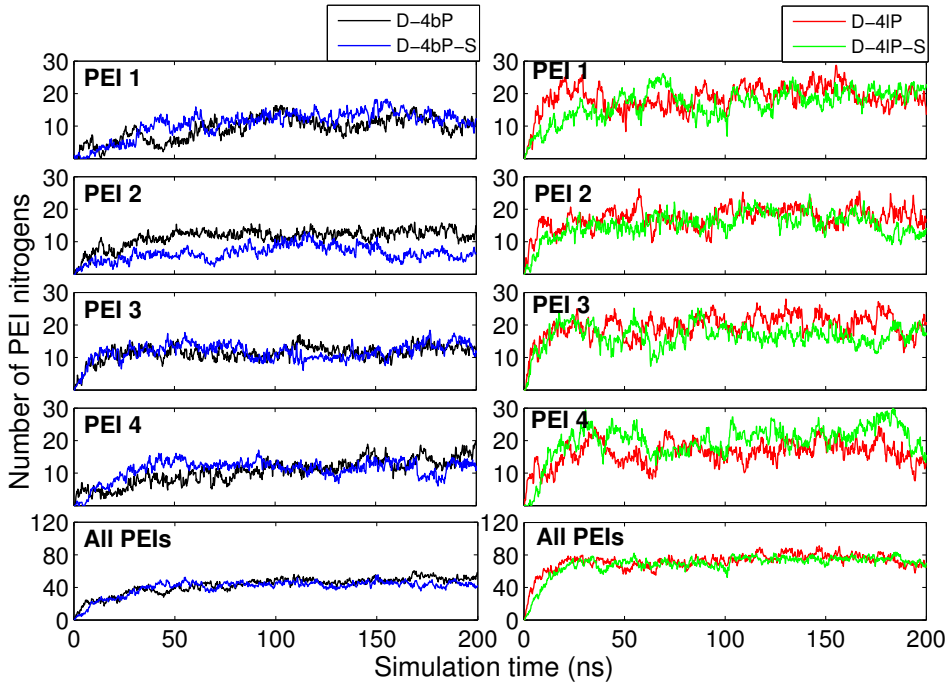

Figure 3.5: Number of nitrogens for each PEI and all PEIs within 4 Å of any N/O atom of the DNA as a function of time for systems D-4bP, D-4bP-S, D-4lP and D-4lP-S.

The first observation from Figures 3.5 and 3.6 is that the two curves in each subfigure have a very similar trend, demonstrating that the ion concentration plays negligible role in affecting the dynamics of PEIs' binding with the DNA. Secondly, at the DNA/PEI charge ratio of 7/8 (Figure 3.5), all the PEIs in each system move toward the DNAs quickly in the first 20 ns reflected by the rapid increase in the number of PEI Ns in close contact with the DNA. The curves for all PEIs stabilize at $\sim$50 ns for bP and at $\sim$25 ns for lP, demonstrating the faster kinetics of lP in binding with the DNA. Each lP has about 18 Ns in close contact with the DNA,

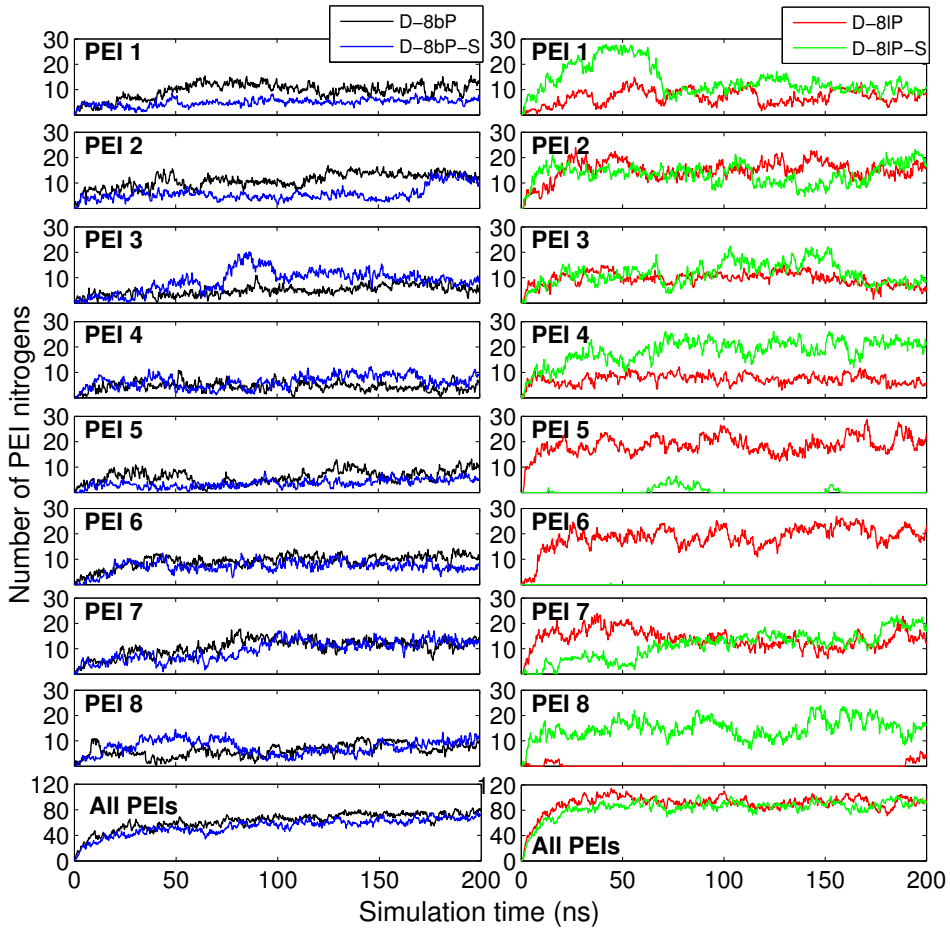

Figure 3.6: Number of nitrogens for each PEI and all PEIs within 4 Å of any N/O atom of the DNA as a function of time for systems D-8bP, D-8bP-S, D-8lP and D-8lP-S.

which is 50% more than that for bP (∼12). This can be explained by the higher flexibility of lP which allows it to comply more easily with the DNA. In contrast, the rigid dendritic bP can only have part of its molecules facing the DNA to form close contact, and the rest stays away from the DNA so as to avoid significant configurational changes (see Figure 3.4(a-d)). In addition, the curves for each lP fluctuate more than that for the bPs. This can again be attributed to the higher flexibility of lP, which make its contacts with the DNA change more frequently.

At the DNA/PEI charge ratio of 7/16, from the final configurations shown in Figure 3.4(c-d, g-h) the DNAs seem to be saturated with their surfaces fully covered by PEIs. This situation is reflected in Figure 3.6 as we see a competition among the PEIs for binding to the DNAs. Specifically, for systems D-8bP and D-8bP-S, all the 8 PEIs bind to the DNA through the entire simulation, but each PEI has fewer Ns (∼9) in close contact with the DNA compared with systems D-4bP and D-4bP-S (∼12). For systems D-8lP and D-8lP-S, the competition is more intense, and some PEIs lose contact with the DNA during the simulation or do not have close contact with the DNA at all from the beginning. For example, PEI 6 in D-8lP-S does not make any close contact with the DNA during the entire simulation. PEI 8 in D-8lP and PEI 5 in D-8lP-S only have very few Ns in close contact with the DNAs for short periods. The large fluctuation for each PEI in systems D-8lP and D-8lP-S also reflects the intense competition for binding. Although there are more bP molecules bound with DNA than lPs, the number of all PEI Ns in close contact with the DNA is still larger for the lPs. In particular, this number is ∼70 for D-8bP and D-8bP-S (equilibrated after 75 ns of the simulations) and ∼90 for D-8lP and D-8lP-S (equilibrated quickly after 25 ns of the simulations). This can be explained by the fact that the flexible lPs can better conformally coat the DNA, resulting in more intimate contact.

Table 3.2 summarizes the average numbers of the PEI Ns in close contact with the DNAs during the last 50 ns of the simulations. It can be seen at both DNA/PEI charge ratios, lPs have significantly more Ns in close contact with the DNAs than bPs, and when the charge ratio changes from 7/8 to 7/16, in all systems the PEIs have 20--30 more Ns to establish close contact with the DNAs. At the charge ratio

Table 3.2: Average numbers of the PEI nitrogens in close contact with the DNAs during the last 50 ns of the simulations.

System	D-4bP	D-4bP-S	D-8bP	D-8bP-S
Number	50.9	44.3	72.7	65.6
System	D-4lP	D-4lP-S	D-8lP	D-8lP-S
Number	74.6	74.4	92.7	89.7

of 7/8, all PEIs bind to the DNA, and the complex formed by the DNA and 4 PEIs is positively charged with a net charge of +10. Despite this overall positive charge, when excess PEIs are present, they continue to bind to the DNA, and the complexes formed by the DNA and the 8 bPs or the 6 lPs represent significantly positively charged particles. The amine groups on the PEIs interact with the DNA N/O through direct or indirect hydrogen bonding [21], and it is this local interaction that facilitates the continuing binding of the PEIs to the DNA. The stoppage of binding in the case of 8 lPs is unlikely driven by the positive charge of the complex since all 8 bPs bind to the DNA. Rather, it is driven by the fact that the entire DNA surface has been covered by the 6 lPs, prohibiting the local interaction of the other 2 lPs with the DNA. At zero and 154 mM salt concentrations, the numbers for D-4lP and D-4lP-S differ only by 0.2 Ns, and the difference between D-8lP and D-8lP-S is only 3 Ns. The less than 4% relative difference demonstrates that the salt ions have negligible effects on the binding of lPs to the DNA. This difference between D-4bP and D-4bP-S is 6.6 Ns and that between D-8bP and D-8bP-S is 7.1 Ns. This difference is relatively small (10-15% relative difference), but it also indicates that the salt ions have a stronger influence on the binding of bPs to the DNA.

The contribution of the PEI flexibility to binding can be further confirmed by examining the radius of gyration of the PEI molecules after the binding. R_g of each PEI in the complexation simulations is plotted in Figures B.1 and B.2 in the Appendix B. In all cases, the bP maintains an almost constant R_g of $\sim$12 Å, nearly identical to the R_g value before the binding (see Figure 3.3). This implied that a bP undergoes little deformation as it binds to the DNA. On the contrary, smaller fluctuations in R_g was observed for the lP molecules after the binding, compared with the fluctuations before the binding (see Figure 3.3). This indicated that the

lPs have conformed themselves to the DNA, lost some degrees of freedom and become less flexible. In addition, in systems D-8lP and D-8lP-S, the R_g values of the unbound PEIs (PEI 6 in D-8lP-S during the entire simulation; and PEI 8 in D-8lP and PEI 5 in D-8lP-S during most time of the simulations) fluctuated more than the bound PEIs, further confirming the ability of lPs to conform to the DNA upon binding.

Macromolecular association in solution can cause water molecules adhered to the surface of macromolecules to be released into the bulk solution. This process is entropically favorable since the water molecules on the macromolecular surface are less mobile [33]. Hence, counting the number of water molecules released from the macromolecules can give us an idea of the strength of interaction in terms of entropic gain from water release and changes in solution accessible surface area upon macromolecular binding. Figure 3.7 shows the number of water molecules in the hydration shell of the DNA or PEIs (within 3 Å from the molecules) as a function of simulation time. As the complexes form, the number of water molecules decreased concurrently, i.e., the water molecules were released from the macromolecules. It can be seen that lPs displaced more water molecules than bPs, with $\sim$500 released in D-4lP/D-4lP-S and $\sim$750 released in D-8lP/D-8lP-S, compared to $\sim$350 in D-4bP/D-4bP-S and $\sim$500 in D-8bP/D-8bP-S. Moreover, water release for lPs start to level off within 40 ns while it took them $\sim$100 ns to level off for the bPs. This again demonstrated the faster kinetics of DNA's complexation with lPs than with bPs due to the higher flexibility of the lP. Comparing Figure 3.7 with the subfigures of all PEIs in Figures 3.5 and 3.6, we found that the number of released water molecules correlates well with the number of PEI Ns in close contact with the DNA: with more PEI Ns coming into close contact with the DNA, more water molecules were released. At zero and 154 mM salt concentrations, the numbers of water molecules released for two systems with the same number and species of molecules are very close, confirming again that the 154 mM salt does not affect on the complexation of PEIs with the DNA to a significant degree.

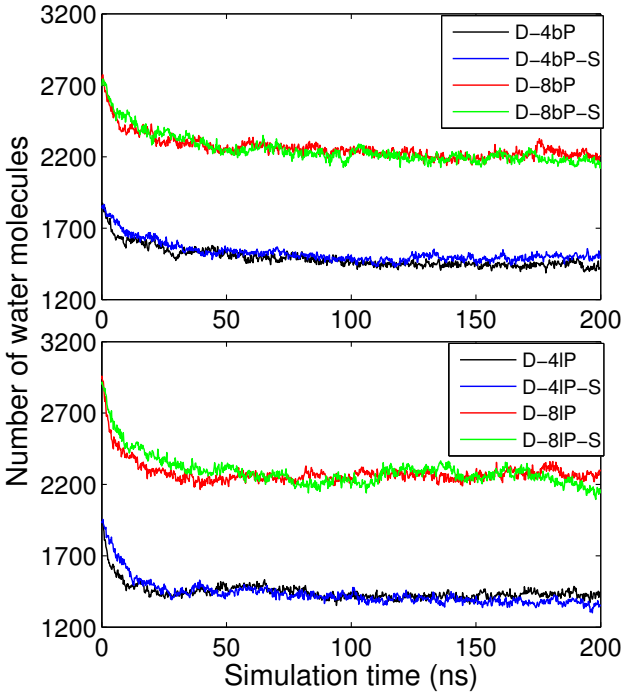

Figure 3.7: Number of water molecules in the hydration shell (within 3 Å of the DNA or PEIs) as a function of simulation time.

3.3.3 Charge neutralization

The simulation results clearly show that PEIs can bind with the DNA and form an overall positively charged particle to overneutralize the DNAs. In our previous MD work of 600 Da PEI mediated DNA aggregation [34], we have demonstrated that the neutralization of the DNA charges by PEIs plays an important role in PEI mediated DNA aggregation [34]. When the PEI/DNA charge ratio is above 1/1, a DNA aggregate can be formed, and when the charge ratio is reduced to 1/4, the DNA aggregate becomes unstable and eventually breaks [34]. To investigate how 2 kDa PEIs neutralize the DNA charges, we plotted the cumulative distributions, with respect to the DNA C1' atoms, of protonated PEI Ns, Na+ ions, Cl- ions, and the net charge of PEI and ions, averaged over the last 50 ns of the simulations (Figure 3.8). In each subfigure, the straight dashed black line indicates the total charge of the DNA (-70), and the blue solid curve indicates the total charge of PEI and ions within a given distance to their nearest DNA C1' atoms. At the intersection of the black line and blue curve, the DNA charges are 100% neutralized by the PEIs

and ions. For all the 8 systems, at larger distances beyond the intersection of the black line and blue curve, the PEI and ion charges exceed the DNA charges to some extent, and the DNA is overneutralized at such distances. Comparing the subfigures on the left column and the ones on the right, we found that charge neutralization (the solid blue line) in systems with 154 mM salt has a very similar characteristic as their counterparts with zero salt concentration. For systems D-4bP and D-4bP-S, the DNA charges are 100% neutralized at a distance of $\sim$13 Å from their C1' atoms, and the DNA is slightly overneutralized beyond this distance. However, at the same charge ratio for linear PEIs in systems D-4lP and D-4lP-S, the four PEIs neutralize the DNA at a much shorter distance of $\sim$9 Å from the DNA C1' atoms. This can be attributed to the high flexibility of the lPs that can comply more easily with the DNA, resulting in shorter separation of the PEI N+ from the DNA. This is also consistent with our previous finding that lP has more number of Ns in close contact with the DNA.

At the DNA/PEI charge ratio of 7/16, the DNAs are all 100% neutralized at a distance of $\sim$8 Å from their C1' atoms similar to the scenario for D-4lP and D-4lP-S, but the DNAs are significantly overneutralized beyond this distance, and the over-neutralization can reach a maximum of $\sim$20%. The difference between bP and lP lies where the overneutralization reaches its maximum. For systems D-8bP and D-8bP-S, the maximum is located at $\sim$20 Å, while for systems D-8lP and D-8lP-S, the maximum is located at a much shorter distance of $\sim$12 Å. Comparing the PEI N+ distribution in systems D-8bP and D-8bP-S with it in systems D-8lP and D-8lP-S (green dashed curves in Figure 3.8 (e,f,g,h)), we found that in the cases of D-8bP and D-8bP-S $\sim$85 PEI N+ are within 10 Å of the DNA C1' atoms and $\sim$140 PEI N+ are within 20 Å of the DNA C1' atoms, while the two numbers for D-8lP and D-8lP-S are $\sim$90 and $\sim$110, respectively. Clearly the bP charges are located further away from the DNA. This once again can be explained by the higher flexibility of lP and the resulting more intimate binding structure compared with bP. If we define the charge center of a PEI as $\frac{\sum_{i=1}^{N} q_i r_i}{\sum_{i=1}^{N} q_i}$, where q_i is the charge of atom i, r_i is its location, and the summation is over all the N atoms in the PEI, then in binding with a DNA, the bPs have their charge centers located further from the DNA axis

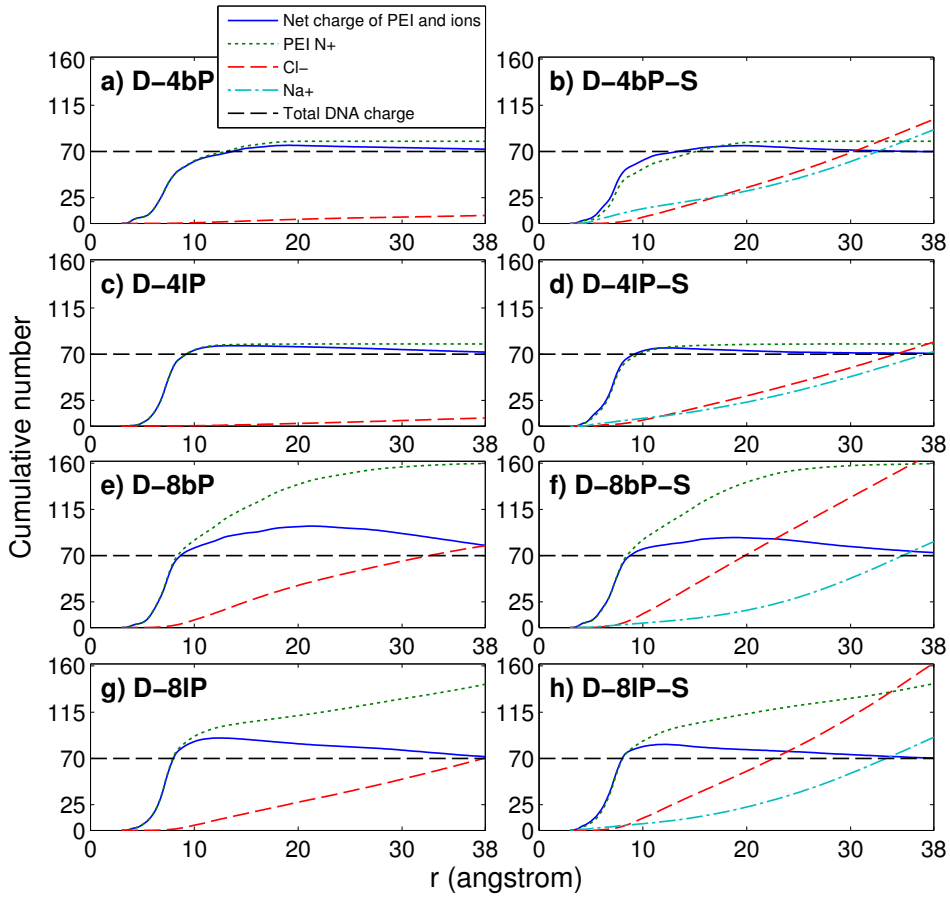

Figure 3.8: Cumulative numbers of protonated PEI nitrogens, Na+, Cl-, and net charge of PEI/Na+/Cl- as a function of the distance from any C1' DNA atom, averaged over the last 50 ns of each simulation. The total charge of the DNA in each system is plotted by a straight dashed black lines as reference. (a) D-4bP, (b) D-4bP-S, (c) D-4lP, (d) D-4lP-S, (e) D-8bP, (f) D-8bP-S, (g) D-8lP, (h) D-8lP-S.

compared with lPs.

On the basis of the analyses of Sections 3.3.2 and 3.3.3, we revealed distinct modes of bP and lP binding with DNAs. A bP tends to have a part of the molecule in close contact with the DNA and leaves the remaining part outward. In the scenario of one DNA segment with multiple bPs, the bPs resemble beads adhering on the surface of the DNA, with their overall charge center located further from the DNA axis compared with that of lPs. The lPs tend to adhere on the DNA surface like cords; because of the high flexibility, the overall charge center of lPs are closer to the DNA axis than the that of bPs. The different binding mode of bP and lP with DNA can affect how the PEIs contribute to DNA aggregation, i.e., when they mediate the condensation of plasmid DNAs. The lP molecules bind very tightly to one DNA segment, leaving little room for interaction with other DNA segments. In addition, for the same amount of molecules, lPs provide better surface coverage of the DNA, the net result being that fewer lPs can bind to the DNA segment and serve as polyion bridges in DNA aggregation. On the contrary, the outward bP moiety might facilitate the PEIs to attract other DNA segments to form more stable DNA/PEI nanoparticles. The different binding modes of lP and bP elucidated here provide a mechanistic explanation to the experimental finding that bP forms more stable nanoparticles with DNA, while lP has a better capacity to unload DNA inside cells [13, 15].

In our previous study of the complexation between single 600 Da PEI and DNA [21], we did not find that the architecture of PEIs had a profound effect on DNA binding, in that four PEIs of different degrees of branching had a similar number of Ns in close contact with DNA. Specifically, at the same protonation ratio of 46%, a linear 600 Da PEI has an average number of 6.2 Ns in close contact with the DNA N/O, while a highly branched 600 Da PEI has an average number of 5.5 Ns in close contact with the DNA N/O. So, the difference is about 0.7 Ns per PEI. If we have 12 600 Da PEIs complexing with the DNA (total molecular weight similar to four PEIs in this study), then the difference will be about 8.4 Ns, and for 24 600 Da PEIs complexing with the DNA (total molecular weight similar to eight PEIs in this study), the difference will be about 16.8 Ns. However, the difference observed

for the 2 kDa linear and branched PEIs is much larger, being 20-30 Ns in the case of 4 PEIs complexing with the DNA. In the case of 8 PEIs complexing with the DNA, all the branched PEIs bind to the DNA, while only 6 out of the 8 linear PEIs bind to the DNA. Nevertheless, the number of Ns in close contact with the DNA in the case of linear PEI still exceeds the case of branched PEI by more than 20. Clearly, the effect of PEI architecture manifests differently for different PEI sizes. At 600 Da, the branched PEIs have short branches; hence, the steric hindrance that each branch experiences in binding with DNA is rather small, and as a result, the branched 600 Da PEI binds to DNA in a similar way as a linear 600 Da PEI. The 2 kDa branched PEIs have more and longer branches so that the dendritic nature of PEI gives greater steric hindrance for amines in binding to DNA, which can explain the distinctly different binding mode it displays compared with that of linear PEIs.

The influence of the molecular weight of PEIs is also reflected in the neutralizing distance for the DNA. In excess of PEIs, both 2 kDa bP and lP fully neutralize the DNA at a distance of $\sim$8 Å from the DNA C1' atoms, which is significantly shorter compared with $\sim$12 Å at which the 600 Da branched PEIs fully neutralize the DNA [34]. This implies that 2 kDa PEIs might form nanoparticles with higher DNA density compared with 600 Da PEIs, which can further facilitate the membrane transfer and better protect DNA from degradation on the delivery route.

3.4 Conclusions

We performed a series of all-atom MD simulations to study the complexation of DNA with 2 kDa branched and linear PEIs. The results revealed the distinct modes of bP and lP in binding to DNA. The bPs bound to DNA like beads adhering to the DNA surface, with little deformation upon binding. The lPs were very flexible and bound to DNA like cords conforming to the DNA surface. The tighter binding of lPs with DNA results in the overall charge center of the lPs being located closer to the DNA axis. In particular, at a PEI/DNA charge ratio of close to 1 (8/7), bP and lP fully neutralized the DNA at $\sim$13 Å and $\sim$9 Å from the DNA C1' atoms, respectively. The tighter binding of lPs further causes more water to be displaced

and potentially leads to energetically more stable DNA-lPs complexes. On the other hand, the lPs provides better surface coverage of the DNA, which limits the number of lPs that can complex with the DNA and the interaction of lPs with multiple DNAs. This can have a negative effect on DNA aggregation needed for cell uptake.

Compared with the results for 2 kDa PEIs, previous work on 600 Da PEIs did not show such significant dependence on PEI architecture. This demonstrates that molecular weight of PEI is an important factor in DNA/PEI complexation. Further evidence for this exists in the fact that in excess of PEIs at a PEI/DNA charge ratio of 16/7, both bP and lP fully neutralized the DNA at a distance of $\sim$8 Å from the DNA C1' atoms, which is a significantly shorter distance compared with $\sim$12 Å in the case of excessive 600 Da PEIs. Finally, our simulations in both physiological and zero-salt conditions showed that the presence of salt had a small effect on DNA/PEI complexation, with a slightly larger influence on the bP molecules.

Bibliography

[1] F. Ledley, *Hum. Gene Therapy* **1995**, *6*, 1129--1144.

[2] D. Luo, W. Saltzman, *Nature Biotechnol.* **2000**, *18*, 33--37.

[3] O. Boussif, F. Lezoualc'h, M. A. Zanta, M. D. Mergny, D. Scherman, B. Demeneix, J. P. Behr, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 7297--7301.

[4] W. Godbey, K. Wu, A. Mikos, *J. Controlled Release* **1999**, *60*, 149--160.

[5] W. Godbey, K. Wu, A. Mikos, *J. Biomed. Mater. Res.* **1999**, *45*, 268--275.

[6] L. Wightman, R. Kircheis, V. Rossler, S. Carotta, R. Ruzicka, M. Kursa, E. Wagner, *J. Gene Med.* **2001**, *3*, 362--372.

[7] K. C. R. Bahadur, B. Landry, H. M. Aliabadi, A. Lavasanifar, H. Uludag, *Acta Biomater.* **2011**, *7*, 2209--2217.

[8] H. M. Aliabadi, B. Landry, R. K. Bahadur, A. Neamnark, O. Suwantong, H. Uludag, *Macromol. Biosci.* **2011**, *11*, 662--672.

[9] D. Jere, H. L. Jiang, R. Arote, Y. K. Kim, Y. J. Choi, M. H. Cho, T. Akaike, C. S. Chot, *Expert Opin. Drug Delivery* **2009**, *6*, 827--834.

[10] A. Neamnark, O. Suwantong, R. K. C. Bahadur, C. Y. M. Hsu, P. Supaphol, H. Uludag, *Mol. Pharm.* **2009**, *6*, 1798--1815.

[11] C. Y. M. Hsu, M. Hendzel, H. Uludag, *J. Gene Med.* **2011**, *13*, 46--59.

[12] J. Wiseman, C. Goddard, D. McLelland, W. Colledge, *Gene Therapy* **2003**, *10*, 1654--1662.

[13] K. Itaka, A. Harada, Y. Yamasaki, K. Nakamura, H. Kawaguchi, K. Kataoka, *J. Gene Med.* **2004**, *6*, 76--84.

[14] J. Intra, A. K. Salem, *J. Controlled Release* **2008**, *130*, 129--138.

[15] Z. Dai, T. Gjetting, M. A. Mattebjerg, C. Wu, T. L. Andresen, *Biomaterials* **2011**, *32*, 8626--8634.

[16] J. Ziebarth, Y. Wang, *Biophys. J.* **2009**, *97*, 1971--1983.

[17] G. M. Pavan, A. Danani, S. Pricl, D. K. Smith, *J. Am. Chem. Soc.* **2009**, *131*, 9686--9694.

[18] G. M. Pavan, L. Albertazzi, A. Danani, *J. Phys. Chem. B* **2010**, *114*, 2667--2675.

[19] G. M. Pavan, M. A. Kostiainen, A. Danani, *J. Phys. Chem. B* **2010**, *114*, 5686--5693.

[20] G. M. Pavan, M. A. Mintzer, E. E. Simanek, O. M. Merkel, T. Kissel, A. Danani, *Biomacromolecules* **2010**, *11*, 721--730.

[21] C. Sun, T. Tang, H. Uludag, J. E. Cuervo, *Biophys. J.* **2011**, *100*, 2754--2763.

[22] S. Boeckle, K. von Gersdorff, S. van der Piepen, C. Culmsee, E. Wagner, M. Ogris, *J. Gene Med.* **2004**, *6*, 1102--1111.

[23] D. Case, T. Darden, I. T.E. Cheatham, C. Simmerling, J. Wang, R. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. Merz, B.Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvary, K.F.Wong, F. Paesani, J. Vanicek, X. Wu, S. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. Mathews, M. Seetin, C. Sagui, V. Babin, P. Kollman, *AMBER 10*, University of California, San Francisco, **2008**.

[24] K. Utsuno, H. Uludag, *Biophys. J.* **2010**, *99*, 201--207.

[25] B. Brooks, R. Bruccoleri, D. Olafson, D. States, S. Swaminathan, M. Karplus, *J. Comput. Chem.* **1983**, *4*, 187--217.

[26] A. MacKerel Jr., C. Brooks III, L. Nilsson, B. Roux, Y. Won, M. Karplus, *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*, Vol. 1 of *The Encyclopedia of Computational Chemistry*, P. v. R. Schleyer et al. (Ed.), John Wiley & Sons: Chichester, **1998**, pp. 271--277.

[27] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. MacKerell, Jr., *J. Comput. Chem.* **2010**, *31*, 671--690.

[28] J. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. Skeel, L. Kale, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781--1802.

[29] W. Jorgensen, *J. Am. Chem. Soc.* **1981**, *103*, 335--340.

[30] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089--10092.

[31] J. Ryckaert, G. Ciccotti, H. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327--341.

[32] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33--38.

[33] J. Dunitz, *Science* **1994**, *264*, 670.

[34] C. Sun, T. Tang, H. Uludag, *Biomacromolecules* **2011**, *12*, 3698--3707.

Chapter 4
Molecular Dynamics Simulations of PEI Mediated DNA Aggregation [1]

4.1 Introduction

DNA condensation has been studied for many years due to its fundamental biological importance, such as the tight packaging of DNA in the chromatin structure and its regulation [1]. More recently, great interest in nonviral gene delivery for therapeutic purposes has stimulated development of systems that can condense DNA and package it suitable for cellular uptake [2]. It was experimentally found that multivalent ions can induce DNA condensation while monovalent or divalent ions lack this capability. Certain synthetic cationic polymers such as polyethylenimine (PEI) can condense nucleic acids into nanoparticulate aggregates and have been employed as effective gene carriers [3, 4]. The compactness and stability of the formed nanoparticulate aggregates were found to be relevant to the delivery efficacy; more stable polymer formulations were correlated with better uptake into cells and, ultimately, better gene expression [5, 6]. The ratio of polymer (e.g., PEI) to DNA in the aggregate formation is known to be critical for transfection [7, 8]; excess PEI gives the aggregates an overall positive charge for increased interaction with membranes and contributes favorably to cell modifications. Several experimental tools are available for the study of aggregate physical features and overall stability [9]. However, little is known about the structures of DNA/polymer complexes at the atomistic level due to limitations of the experimental tools. The role of polymeric cations in maintaining the aggregates stability as well as molecular kinetics in such aggregates remains to be probed.

[1] A version of this chapter has been published. Reprinted with permission from: C. Sun, T. Tang, H. Uludağ, Biomacromolecules, 2011, 12 (10), p3698-3707. Copyright 2011 American Chemical Society.

DNA is a highly charged polyelectrolyte with a charge density of -2e/3.4 Å. Mean-field theory such as Poisson-Boltzmann equation always predicts a repulsive interaction between like-charged polyelectrolytes and thus cannot explain the multivalent ion induced DNA condensation. Coarse grained simulations of polyelectrolytes normally treat several or tens of atoms on the polyelectrolytes as a unit and water as a continuous dielectric medium. This method has been employed in the past decade to study oppositely charged polyelectrolyte interactions [10--16] and had some success in predicting how chain length, charge density, charge ratio, and ion concentration affect the complexation of polycation and polyanion. However, coarse grained simulations neglect the fine details of the molecules, especially the water structure around the binding sites, and thus are only appropriate for interactions over distances exceeding the atomic scale. In the case of polymer mediated DNA aggregation, the surfaces of the DNA segments can be as close as a few angstroms [17]. In such situations, water molecules play a crucial role in arranging their structure (polarity) to mediate the strong electrostatic interaction and to form hydrogen bonds; thus, an atomistic description is necessary in order to understand polycation induced DNA aggregation. Recent experiments have also demonstrated the crucial role of atomic topologies in strong polyelectrolyte interactions. For example, it was found that double-stranded RNA resists condensation by trivalent counterions, which can otherwise condense double-stranded DNA [18], and divalent counterions can condense triple-stranded DNA but lack the capability to condense double-stranded DNA [19]. These findings further underline the requirement to incorporate an all-atom representation in simulating polycation mediated nucleic acids aggregation and condensation. Atomic simulations have recently been employed to study the complexation of single nucleic acid molecule with polycations [20--27], where the structure, dynamics, and energetics of the nucleic acid/polycation complexes were investigated, but studies investigating polycation mediated aggregation of multiple DNAs are yet missing.

In this work, we performed a series of large scale all-atom molecular dynamics (MD) simulations to study the PEI mediated DNA aggregation. Specifically, we studied the mechanism and dynamics of PEI induced DNA aggregation, how close

the DNA segments are in the DNA/PEI polyplex, and how the compactness of the DNAs is affected by the DNA/PEI charge ratio. The PEI studied in this work is a 568 Da low molecular weight (LMW) PEI. Such LMW PEIs provide an optimal system size that can be simulated by all-atom MD and are increasingly employed as gene delivery carriers due to suitable compatibility with cellular systems [7, 28, 29]. To our knowledge, this study is the first atomistic simulations on polycation mediated DNA aggregation involving multiple DNA molecules. The results help to elucidate on the mechanism of PEI mediated DNA aggregation at the atomic resolution and, moreover, to understand DNA aggregation and condensation involving other polycations of similar characteristics.

4.2 Methods

The DNA simulated in this work was a Drew-Dickerson dodecamer d(CGCGAATT CGCG)$_2$ composed of 24 nucleotides, which carries a total charge of -22 in the fully deprotonated state. The PEI simulated was a branched PEI consisting of 13 amine groups with a molecular weight of 568 Da. The chemical structure and protonation sites of the PEI are shown in Figure 4.1. We chose to have 6 amine groups protonated (46% protonated) to be consistent with the protonation ratio of 47% for 600 Da PEI from our recent study [30]. The protonation sites were assigned to both the primary and secondary amines and were distributed as uniformly as possible to minimize thermodynamic interactions between the protonated amines [31]. An MD simulation was first performed for 6 ns for PEI surrounded by explicit water and counterions, and the configuration of the PEI at the end of the simulation was adopted as the initial configurations for PEIs in the complex formation simulations [31]. Seven separate systems were simulated in this study, and their information is summarized in Table 4.1. Each system consists of a certain number of DNA(s), PEIs, ions, and water molecules. Details of the simulated systems and the explanations on their designations are given below.

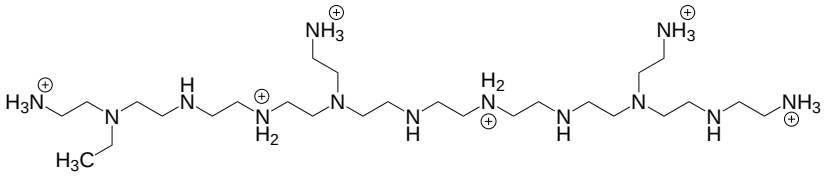

Figure 4.1: Molecular structure and protonation sites of the PEI studied.

Table 4.1: Information on the seven different systems simulated in this study.

System name	No. of DNA/PEI	Charge ratio DNA/PEI	No. of atoms	Simulation box size ($Å^3$)	No. of Na+/Cl-	Time (ns)
D-4P	1/4	22/24	43244	$74 \times 79 \times 74$	0/2	100
D-8P	1/8	22/48	60882	$88 \times 93 \times 73$	0/26	100
2D-8P-50	2/8	44/48	77149	$122 \times 78 \times 79$	0/4	100
2D-8P	2/8	44/48	65965	$105 \times 78 \times 79$	0/4	100
2D-2P	2/2	44/12	64423	$104 \times 78 \times 79$	36/4	200
4D-16P	4/16	88/96	97007	$107 \times 112 \times 79$	0/8	130
4D-28P	4/28	88/168	175910	$117 \times 122 \times 120$	0/80	200

4.2.1 Simulated systems and procedure

The first two systems shown in Table 4.1, namely D-4P and D-8P, each contains a single DNA and multiple PEIs, 4 PEIs in D-4P and 8 PEIs in D-8P. For each system, the principal axes of the PEIs were initially aligned parallel to the DNA axis, and the center of mass (COM) of each PEI was positioned at 25 Å away from the DNA COM. Each system was simulated for 100 ns, and in both cases, four PEIs were attached to the DNA at the late stage of the simulations [32]. The structure of the D-4P system at 15 ns was then used to construct the systems that involve multiple DNAs to study the DNA aggregation. The D-4P and D-8P systems are also discussed in Section 4.3.2 to address the ability of PEI to neutralize DNA.

To study the PEI-mediated aggregation of multiple DNA molecules, we first simulated two systems, named 2D-8P-50 and 2D-8P (Table 4.1), each containing two DNAs and eight PEIs. The initial configurations of these two systems were constructed by solvating two identical D-4P complexes in the simulation box, as shown in Figure 4.2(a) and Figure 4.2(b), respectively. The axes of the DNAs in the two complexes were aligned to be parallel, and the COM of the two complexes were initially separated by 50 Å in 2D-8P-50 and by 32 Å in 2D-8P. The distinction

between these two systems is marked by the ‘‘50’’ in the name of system 2D-8P-50, which refers to the initial separation of 50 Å between the two D-4P complexes. The reason for using two different COM distances was to examine the likelihood of aggregation at these separations.

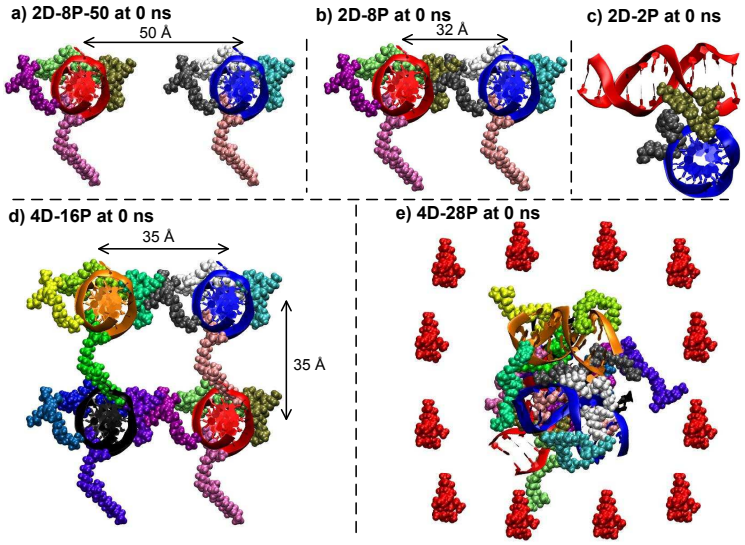

Figure 4.2: Initial configurations of the systems: (a) 2D-8P-50, (b) 2D-8P, (c) 2D-2P, (d) 4D-16P, (e) 4D-28P. Different PEIs are represented in different colors (except in (e) where the extra 12 PEIs are in red); water and ions are removed for clarity.

Another simulation was conducted for a system that consists of two DNAs and two PEIs, indicated by 2D-2P in Table 4.1. While the 2D-8P-50 and 2D-8P systems both have a PEI/DNA charge ratio of $\sim$1/1, the 2D-2P system has a PEI/DNA charge ratio that is much smaller than one ($\sim$1/4). Hence, simulation on this system allows us to investigate the effect of electrostatic screening by PEI on DNA aggregation. The initial configuration of 2D-2P is based on the configuration of 2D-8P at 100 ns of its simulation. Specifically, as will be discussed later, we found that two out of the eight PEIs in 2D-8P are bridging the two DNAs at the end of the simulation. We kept these two PEIs and removed the other six PEIs. 36 Na+ ions were assigned on the locations of the protonated nitrogens on the removed 6 PEIs. The initial configuration of the 2D-2P system is illustrated in Figure 4.2(c).

To further study the formation of aggregate in the case of multiple DNA segments, we performed a simulation on a system with 4 DNAs and 16 PEIs, named

4D-16P in Table 4.1. In constructing the initial configuration for 4D-16P, we followed a similar procedure to that of 2D-8P. In particular, we adopted four D-4P complexes and arranged them on the four corners of a square, as shown in Figure 4.2(d). The axes of the four DNAs were aligned to be parallel, and the COM of each D-4P complex was separated from the COM of its neighboring complex by 35 Å.

To investigate the effect of excess PEIs on the DNA aggregation, 12 PEIs were added to the 4D-16P system at the simulation time of 100 ns, and the new system is referred to as 4D-28P. The added 12 PEIs surrounded the 4D-16P complex in a circular fashion, with 8 PEIs located at 42 Å from the COM of 4D-16P and 4 PEIs located at 50 Å from the COM of 4D-16P. The initial configuration of 4D-28P is shown in Figure 4.2(e).

4.2.2 Simulation details

A CHARMM format force field was devised for PEI based on the CHARMM General Force Field [33], and CHARMM 27 force field [34, 35] was used for all other molecules. The force field parameters for PEI have been carefully validated through *ab initio* calculations, a study on sensitivity of MD results to torsional parameters, and comparison with previous works [31]. The simulations were performed using MD package NAMD [36]. TIP3P water model [37], periodic boundary condition, and full electrostatics with particle-mesh Ewald method [38] were used for all the simulations. A cutoff of 12 Å was used for van der Waals interactions and electrostatics pairwise calculations. All bonds containing hydrogen atoms were constrained using the SHAKE algorithm [39] during all the simulations, and a time step of 2 fs was used.

For each system described in Section 4.2.1, the DNA(s) and PEIs were solvated into a water box, the size of which was set to be large enough to make sure the DNA(s) and PEIs are at least 36 Å away from their nearest periodic images in each direction. Ions (numbers summarized in Table 4.1) were then added to the water box by randomly replacing the same number of water molecules using VMD [40]. During each simulation, the system was first minimized for 2000 steps

with DNA(s)/PEIs fixed, and then 2000 steps with DNA(s)/PEIs non-hydrogen atoms harmonically restrained, followed by 1000 steps of unrestrained minimization. The system was then gradually heated from 0 to 300 K in 20 ps with 10 $\mathrm{kcal}/(\mathrm{mol}\times\mathrm{\AA}^2)$ harmonic restraint on the DNA(s)/PEIs non-hydrogen atoms. The restraint was kept on for another 4 ns at 300 K and 1 bar to relax the ions around the solutes. The restraint was then removed, and NPT ensemble simulation was performed for 100-200 ns (simulation time for each system shown in Table 4.1). VMD [40] was used for visualization and trajectories analysis.

4.2.3 Definitions and acronyms

To facilitate the discussion of the simulation results, we introduced the following definitions and acronyms in analyzing the simulation trajectories.

In systems 2D-8P and 4D-16P, each DNA is labeled with a capital letter (A, B, C, or D), and the four PEIs initially associated with the DNA in a D-4P complex (see Section 4.2.1) are labeled with the same capital letter plus a number. For example, ‘‘A1, A2, A3, A4’’ stand for the four PEIs initially associated with DNA molecule A. In system 4D-28P, we keep the same labels for the 16 PEIs from 4D-16P and label the extra 12 PEIs by ‘‘E1--E12’’. The acronyms for the DNAs and PEIs in each system are summarized in Table 4.2.

To describe the binding state of PEI to DNA, a PEI N is said to be ‘‘in close contact with the DNA’’ if it is within 4 Å of any N/O atoms of the DNA. We chose 4 Å because this is the distance within which the PEI amine groups can form direct hydrogen bond with the DNA [31]. A PEI is said to be ‘‘bound’’ to a DNA molecule if it has at least one N in close contact with this DNA. If a PEI is ‘‘bound’’ to two or more DNA molecules simultaneously, this PEI is said to be bridging or forming a polyion bridge between the DNAs.

To investigate the DNA-DNA spacing in the aggregates, we defined the ‘‘shortest distance’’ and ‘‘root-mean-square (RMS) distance’’ between two DNA molecules. We first represent each DNA as a series of points each being the COM of a Watson-Crick DNA base pair (see Figure C.2 in Appendix C). For each dodecamer studied in this work, there are 12 such points, and connecting neigh-

Table 4.2: Acronyms of the DNAs and PEIs in each system.

System	DNAs	PEIs
2D-8P	A, B	A1-A4, B1-B4
2D-2P	A, B	A1, B4
4D-16P	A, B, C, D	A1-A4, B1-B4, C1-C4, D1-D4
4D-28P	A, B, C, D	A1-A4, B1-B4, C1-C4, D1-D4, E1-E12

boring points results in 11 segments. For a pair of segments from two different DNAs, we can calculate their shortest distance, and there are 121 such distances ($d_i, i = 1 \ldots 121$) between all pairs of segments from the two DNA dodecamers. We defined the minimum of these 121 distances as the ''shortest distance'' d_{shortest} and the root-mean-square of these 121 distances as the ''RMS distance'' d_{RMS} ($d_{\mathrm{RMS}} = \sqrt{(\Sigma_{i=1}^{121} d_i^2)/121}$). ''Shortest distance'' is a parameter to quantify the closet approach of two DNA molecules, whereas ''RMS distance'' is a parameter to characterize the closeness of two DNA molecules as two entities.

4.3 Results and Discussions

For the system 2D-8P-50, the two D-4P complexes stayed separate during the 100 ns simulation. Figure 4.3(a) shows the configuration of 2D-8P-50 at the end of the simulation, where the two DNAs are separated at a COM distance of $\sim$42 Å. No PEI molecule simultaneously binds to both DNAs. We attribute the lack of aggregation in this case to the fact that the two D-4P complexes were initially separated by a relatively large distance; namely, each complex is 50 Å from its closest complex and 72 Å from its second closest complex as a periodic image. Because each D-4P complex carries a total charge of +2, an overall repulsive electrostatic interaction between the two complexes is expected. A sufficiently close approach might be necessary to allow the attractive interaction between the positively and negatively charged parts of the complexes to form an aggregate. Although sufficient diffusion of the macromolecular complexes could accomplish this, the 100 ns simulation time was relatively short, so that the diffusion alone apparently did not bring the two complexes close enough to form an aggregate in the 2D-8P-50 system. To obtain an aggregate within a practical simulation time, we brought the

D-4P complexes at shorter separations of 32 Å for the 2D-8P system and 35 Å for the 4D-16P system at the beginning of the simulations. In both cases, aggregation happened shortly after the simulation started, and the complexes were never separated again thereafter (see movies in online Supporting Information available at ‘‘http://dx.doi.org/10.1021/bm2009476”). The overall repulsive electrostatic interaction was therefore not an impediment for aggregate formation. In the following subsections, structural analysis were conducted to examine the mechanism of aggregation and to characterize the formed aggregates.

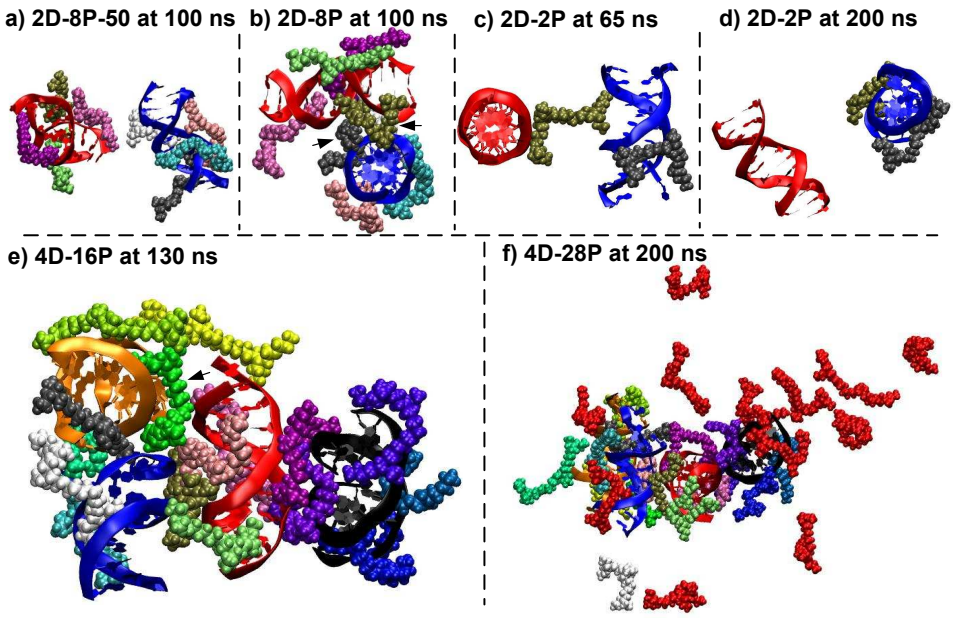

Figure 4.3: Snapshots of the systems in the simulations: (a) 2D-8P-50 at 100 ns, (b) 2D-8P at 100 ns, (c) 2D-2P at 65 ns, (d) 2D-2P at 200 ns, (e) 4D-16P at 130 ns, and (f) 4D-28P at 200 ns. Different PEIs are represented in different colors (except in (f) where the added 12 PEIs are in red); water and ions are removed for clarity.

4.3.1 Mechanism of aggregation

Figure 4.3(b) shows the conformation of the system 2D-8P at the last stage of the 100 ns simulation. It can be seen that all eight PEIs are attached to the DNAs with significant fraction of each PEI in contact with the DNAs. Two PEIs (indicated by the two black arrows in Figure 4.3(b)) bind to the two DNAs concurrently, bridging them so that they are closer to each other compared with the initial configuration.

Figure 4.3(e) is the snapshot of the 4D-16P system at the end of its simulation. As in the 2D-8P system, all PEIs are attached to the DNAs, and about half of the PEIs participate in bridging the DNA molecules. In particular, one PEI (indicated by the black arrow in Figure 4.3(e)) binds to three DNAs simultaneously.

To quantify the ability of PEI to bridge the DNA molecules, we plotted the binding state of individual PEIs to each DNA in terms of the number of Ns from each PEI in close contact with each DNA (see Section 4.2.3 for definition of ‘‘close contact’’), as shown in Figure 4.4 for the 2D-8P system and in Figure 4.5 for the 4D-16P system. In Figure 4.4, each subfigure corresponds to one of the 8 PEIs in the 2D-8P system and contains two curves, each of which describes the number of Ns of this PEI in close contact with a particular DNA (see Section 4.2.3 for the acronyms of the molecules in each system). Similarly, the 16 subfigures in Figure 4.5 correspond to the 16 PEIs in the 4D-16P system, and the four curves in each subfigure describe the binding state of a PEI with the four DNAs. In Figure 4.4, at the beginning of the simulation, PEI A1--A4 and DNA A constitute one D-4P complex, and PEI B1--B4 and DNA B constitute the other D-4P complex. Except for B1 and B2, each PEI has at least brief periods during which it forms a polyion bridge between the two DNAs (see Section 4.2.3 for definition of ‘‘polyion bridge’’). The bridges are transient; they form and break during the simulations. For example, the bridge formed by PEI A1 breaks for several times at around 25, 40, 70, and 85 ns. Notably, PEIs A1 and B4 contribute to the aggregation of the two DNAs significantly, each having more than 1 Ns in close contact with each DNA for longer than 50% of the entire simulation time. At around 60 ns, PEI A1 has as many as 4 Ns in close contact with each DNA. The same happens to PEI B4 at around 50 ns. Figure 4.5 demonstrates the polyion bridging in the 4D-16P system. Out of the 16 PEIs, 8 PEIs (A1, A3, B1, B4, C1, C3, C4, and D4) participate in bridging two or three DNAs for longer than 50% of the simulation time. In fact, the fraction of PEIs that contribute to bridging has increased from 25% in 2D-8P to 50% in 4D-16P. DNAs A, B, and C are mutually bridged (A and B bridged by PEIs A3, B4, C3; A and C bridged by PEIs A1, A2, C3, C4; B and C bridged by PEIs B3, C1, C3), while DNA D is only bridged with DNA B by PEIs B1 and D4.

Noticeably, PEI C3 is bridging three DNAs (A, B, and C) from ∼30 ns until the end of the simulation.

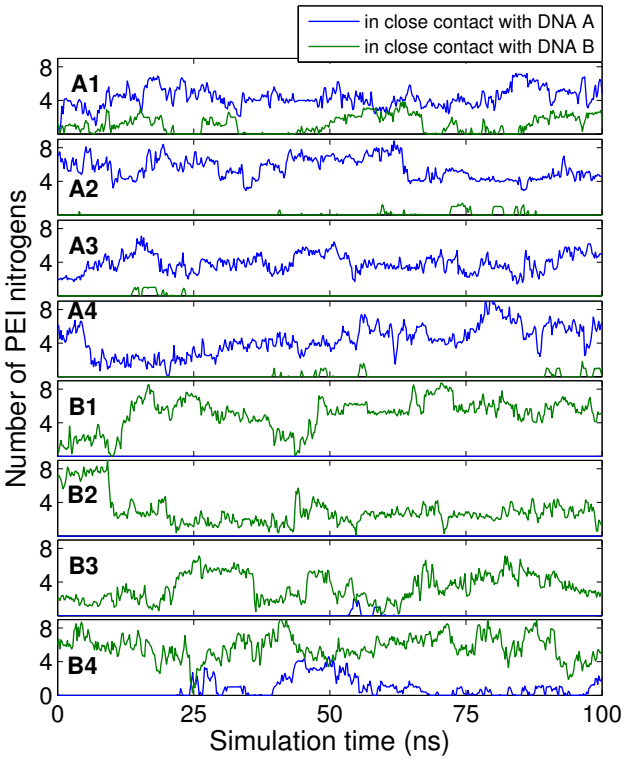

Figure 4.4: Number of nitrogens for each PEI within 4 Å of any N/O atom of each DNA as a function of time for the 2D-8P system which contains 2 DNAs and 8 PEIs. The 8 subfigures correspond to the 8 PEIs in the system; the 2 curves in each subfigure correspond to the 2 DNAs in the system.

The above results clearly demonstrate the presence of bridging PEIs when a DNA aggregate is formed. The bridging PEIs are not ‘‘locked” in bound state (see Section 4.2.3 for definition of ‘‘bound’’) and undergo reversible binding. The polyion bridging is likely one of the key mechanisms causing the DNA aggregation. Another possible mechanism can be the electrostatic screening of DNA charges by PEIs at a short distance from the DNA surface so that the strong repulsive electrostatic interaction between DNAs at close separation is weakened (or shielded). To verify this, we performed the simulation of a system with a lower PEI to DNA charge ratio. The objective is to test if aggregation can remain with fewer PEIs, i.e., with less electrostatic screening. One way to do so is to start a new simulation as we did for 2D-8P, i.e., separate two DNA/PEI complexes first and check if they

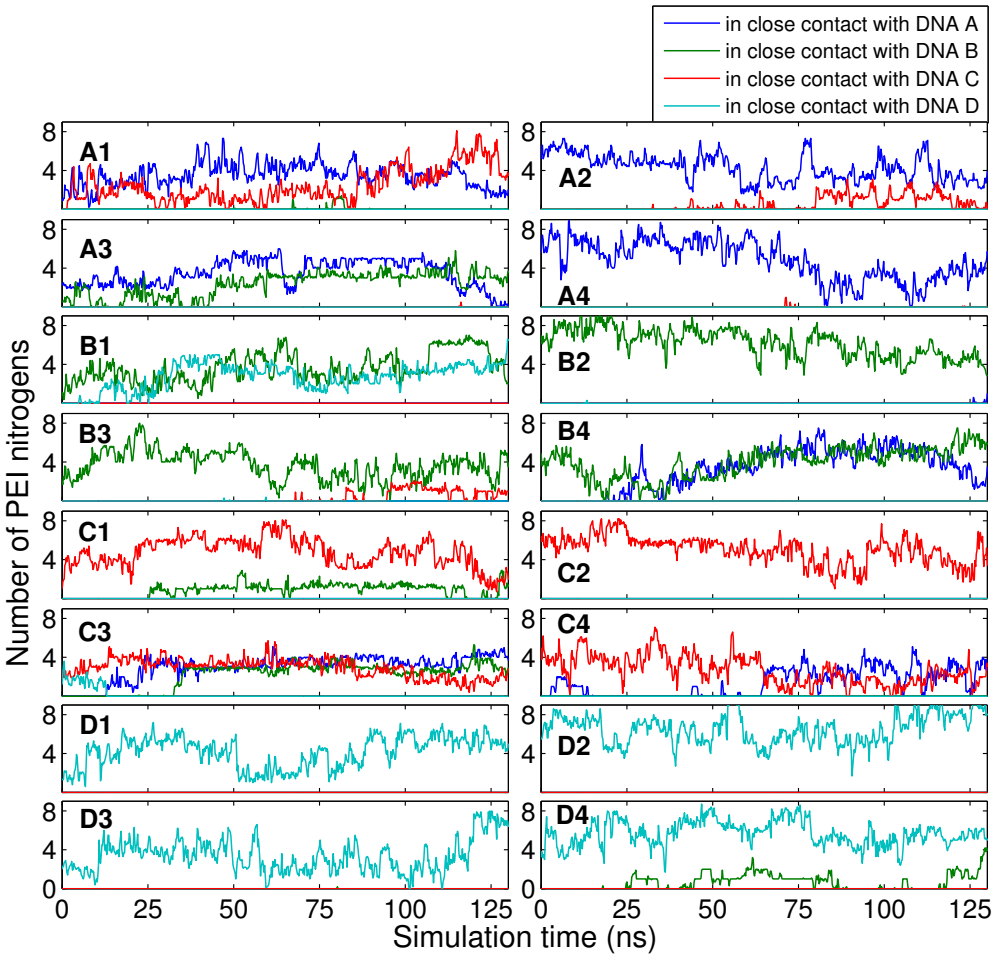

Figure 4.5: Number of nitrogens for each PEI within 4 Å of any N/O atom of each DNA as a function of time for the 4D-16P system which contains 4 DNAs and 16 PEIs. The 16 subfigures correspond to the 16 PEIs in the system; the 4 curves in each subfigure correspond to the 4 DNAs in the system.

aggregate; however, one can encounter the same problem as we experienced in simulating 2D-8P-50. That is, the DNA aggregation could happen, while it might not be observed within the simulation time limited by current computational capability. As an alternative, we simulated the system 2D-2P with two DNAs and two PEIs following the procedure described in Section 4.2.1; i.e., we made use of the final configuration from the 2D-8P simulation, kept the two PEIs bridging the two DNAs (PEIs A1 and B4) and replaced the other 6 PEIs with Na+ ions. If the aggregate becomes looser or breaks, then the role of electrostatic screening will be verified. By visually checking the configurational change during the simulation, we observed loosening of the aggregate during the simulation (Figure 4.3(c)) and ultimate breakup of the aggregate at around 150 ns as seen in Figure 4.3(d) (also see movie in online Supporting Information available at ‘‘http://dx.doi.org/10.1021/bm2009476”). Figure 4.6 plots the number of Ns of each PEI in close contact with each DNA for 2D-2P. It can be seen that the polyion bridge lasts for 150 ns before it breaks and is not restored after the breakage. Clearly, the electrostatic screening of the PEI molecules also plays an important role in maintaining the DNAs in an aggregated form.

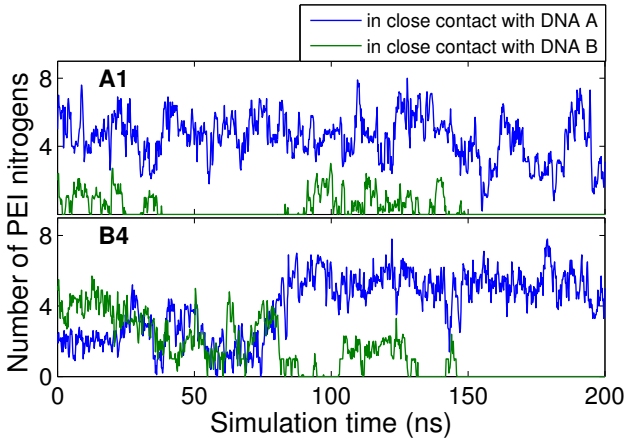

Figure 4.6: Number of nitrogens for each PEI within 4 Å of any N/O atom of each DNA as a function of time for the 2D-2P system which contains 2 DNAs and 2 PEIs. The 2 subfigures correspond to the 2 PEIs in the system; the 2 curves in each subfigure correspond to the 2 DNAs in the system.

In experiments, the DNA/PEI complex is typically prepared in excess of PEIs. To examine DNA aggregation in excess of PEIs, we performed the simulation of

4D-28P by adding 12 extra PEIs to the 4D-16P system at 100 ns and setting this time to be zero for the 4D-28P simulation. Figure 4.7 shows the number of Ns from each PEI in close contact with the DNAs in the 200 ns simulation. It can be seen that the polyion bridging between DNAs follows a similar characteristic as in 4D-16P. Eight out of the 28 PEIs (A1, A3, A4, B1, B4, C1, C3, and D4) participate in bridging two or three DNAs for longer than 50% of the simulation time. PEIs A3 and C3 are bridging three DNAs for most of the simulation time. Five out of the added 12 PEIs (E1, E4, E8, E11, and E12) bind with the DNAs for significantly long periods; however, they mainly bind with one DNA with only very short time windows to bridge two DNAs. For example, PEI E8 is only bound to DNA C over most time of the simulation while bridging DNAs A and C for several ns at $\sim$170 ns. Noticeably, some of the original 16 PEIs were ‘‘replaced’’ by the added PEIs in that they unbound from the DNAs while allowing the newly added PEIs to form the DNA binding. For example, PEI D3 was replaced during the time window of 50-180 ns (also, see movie in online Supporting Information available at ‘‘http://dx.doi.org/10.1021/bm2009476”). Seven of the added PEIs (E2, E3, E5, E6, E7, E9, and E10) make no or negligible contact with the aggregate during the simulation. By studying the binding state of each PEI to each DNA in the 4D-28P system, we found that $\sim$18 PEIs were bound with the DNAs at the late stage of the simulation (see Figure C.1 in Appendix C). As the 18 PEIs carry a total charge of +108 and the 4 DNAs carry a total charge of -88, the formed DNA/PEI aggregate is even more positively charged than the aggregate obtained in the 4D-16P system which carries a net charge of +8. This is consistent with the experimental measurements of the ζ-potential of DNA/PEI complexes; it is well established that the gradual addition of PEI molecules results in a progressive increase in ζ-potential, ultimately reaching $> +30$ mV in the presence of excess PEI and indicating cationic nature of the final aggregates [7].

4.3.2 Charge neutralization

As demonstrated in Section 4.3.1, electrostatic screening plays an crucial role in DNA aggregation. To investigate how PEIs neutralize the DNA charges, we plotted

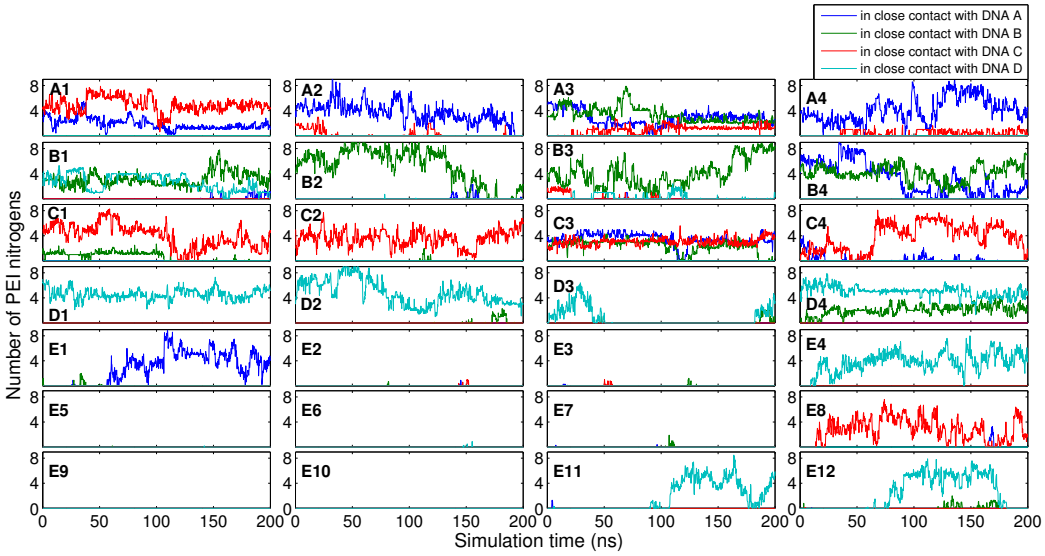

Figure 4.7: Number of nitrogens for each PEI within 4 Å of any N/O atom of each DNA as a function of time for the 4D-28P system which contains 4 DNAs and 28 PEIs. The 28 subfigures correspond to the 28 PEIs in the system; the 4 curves in each subfigure correspond to the 4 DNAs in the system.

the cumulative distributions, with respect to the DNA C1' atoms, of protonated PEI Ns, Na+/Cl- ions, and the net charge of PEI and ions, averaged over the last 40 ns of the simulations (Figure 4.8). The C1' atoms are on the sugar rings of the DNAs, located inside the DNA helix at a distance of $\sim$5 Å from the surface of DNA defined by the phosphorus atoms (see Figure C.2 in Appendix C). In systems D-4P, 2D-8P, and 4D-16P, the DNA/PEI charge ratio is approximately 1/1; in systems D-8P and 4D-28P, the DNA/PEI charge ratio is approximately 1/2; and in system 2D-2P, the DNA/PEI charge ratio is approximately 4/1. In each subfigure of Figure 4.8, the straight dashed black line indicates the total charge all the DNAs in the system carry, and the blue solid curve is the total charge of PEI and ions within given distance to their nearest DNA C1' atoms. If the black line and blue curve intersect, the DNA charges are 100% neutralized by the PEI and ions at the distance where they intersect. At larger distances, the PEI and ions charges could exceed the DNA charges, and the DNA would be ''overneutralized'' at such distances. This is the case for the five systems in Figure 4.8, except the 2D-2P system. It can be seen, for these five systems, with DNA/PEI charge ratio of 1/1 or 1/2, the curves for the net charge of PEI and ions have a similar characteristic. That is, the DNA(s) are 50%

neutralized at a distance of ∼7 Å and 100% neutralized at a distance of ∼12 Å from their C1' atoms; the DNA(s) are slightly ''overneutralized'' beyond the distance of ∼12 Å from their C1' atoms and the ''overneutralization'' maximize at ∼15 Å. For systems D-4P, 2D-8P, and 4D-16P, all the PEI charges are located within 15 Å from the DNA C1' atoms, supported by the observation that the cumulative number of PEI N+ is constant beyond 15 Å. For systems D-8P and 4D-28P, more PEI charges still accumulate gradually beyond 15 Å, which however are largely neutralized by Cl-. The scenario for 2D-2P is very different from the other 5 cases. The DNAs are not fully neutralized even at a distance of 25 Å from their C1' atoms; all the 12 charges from the two PEIs are within 8 Å from the DNA C1' atoms and the DNAs are only about 50% neutralized at a distance of 12 Å from their C1' atoms. This demonstrates that PEI is much more capable of neutralizing the DNA at a short distance from the DNA surface than monovalent ions. We have also plotted the cumulative net charge of PEI/Na+/Cl- based on three time windows at the late stage of the simulations as an evidence of convergence of the simulation trajectories (see Figure C.4 in Appendix C).

4.3.3 DNA-DNA spacing in the aggregate

The DNA-DNA spacing reflects the compactness of DNA molecules in the aggregate and has been a great interest of experimental studies [1, 17, 41, 42]. The DNA-DNA spacing from our MD trajectories was analyzed, and Figures 4.9 to 4.12 show the distance between any two DNA molecules as a function of simulation time for the 2D-8P, 2D-2P, 4D-16P, and 4D-28P systems, respectively (See Section 4.2.3 for the definitions of ''shortest distance'' and ''RMS distance'').

In Figure 4.9 for the 2D-8P system, it can be seen the two curves decrease at the beginning of the simulation, indicating the approach of the two DNAs. The curves flatten after 40 ns with the average values of ∼23 Å for the d_{shortest} and ∼28 Å for the d_{RMS}. In Figure 4.10 for the 2D-2P system, the two curves remain relatively stable without significant fluctuation from 0 to 150 ns when the two DNAs are still bridged by one or two PEIs. At about 150 ns when the polyion bridge breaks, the two curve increase dramatically, indicating the separation of the two DNAs. In

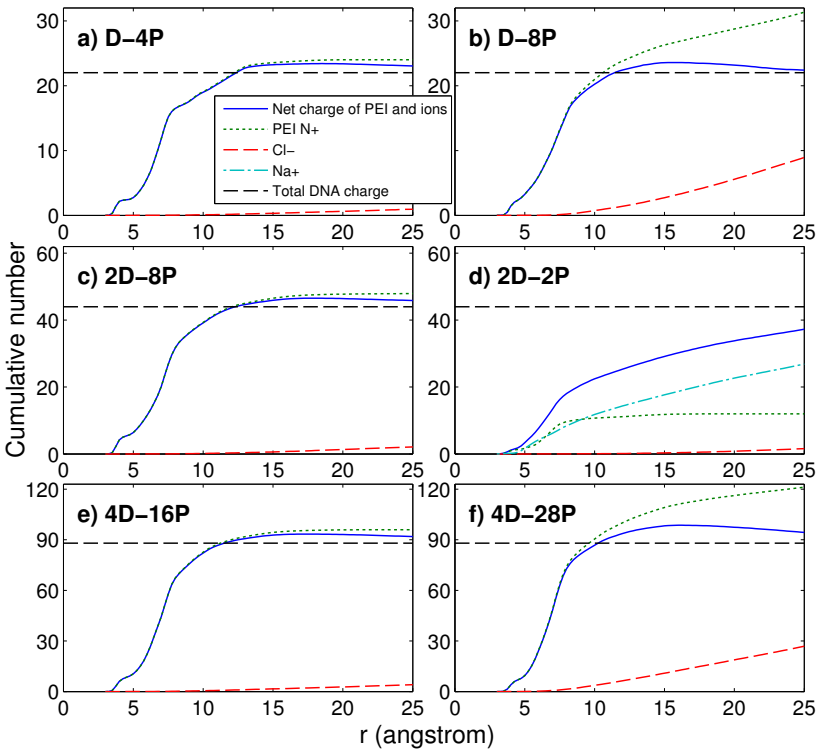

Figure 4.8: Cumulative numbers of protonated PEI nitrogens, Na+, Cl-, and net charge of PEI/Na+/Cl- as a function of the distance from any C1' DNA atom (averaged over the last 40 ns of each simulation). The total charge of all the DNAs in each system is plotted by a straight dashed black lines as reference. (a) D-4P, (b) D-8P, (c) 2D-8P, (d) 2D-2P, (e) 4D-16P, (f) 4D-28P.

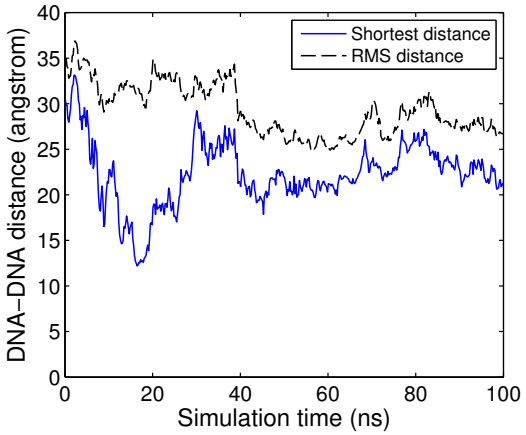

Figure 4.9: DNA-DNA distance in the 2D-8P system which contains two DNAs: (a) shortest distance; (b) root mean square distance (see texts for definition of these distances).

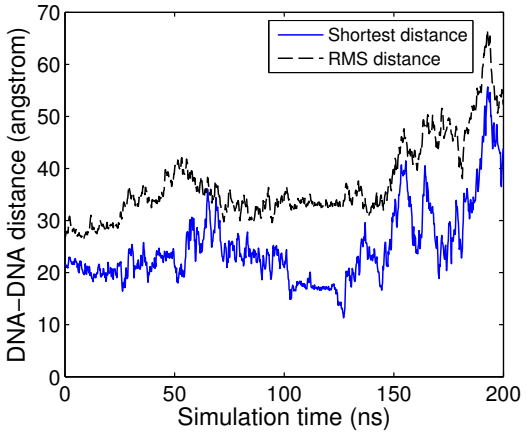

Figure 4.10: DNA-DNA distance in the 2D-2P system which contains two DNAs: (a) shortest distance; (b) root mean square distance (see texts for definition of these distances).

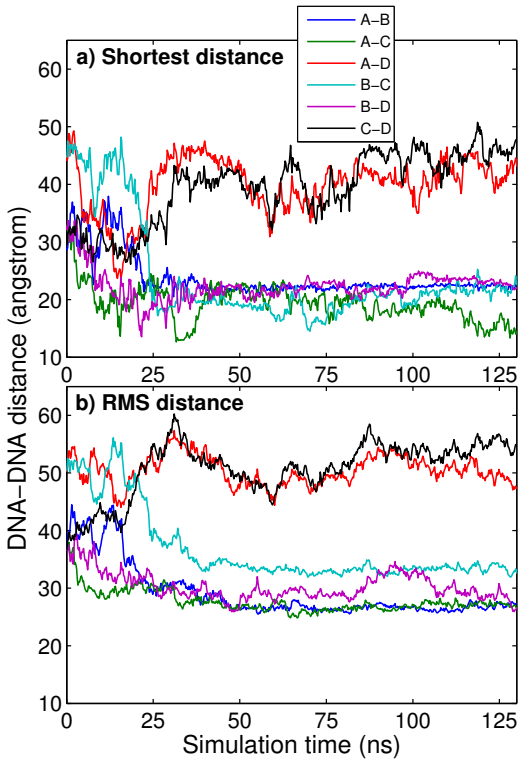

Figure 4.11: DNA-DNA distances in the 4D-16P system which contains four DNAs (A, B, C, and D): (a) shortest distance; (b) root mean square distance (see texts for definition of these distances).

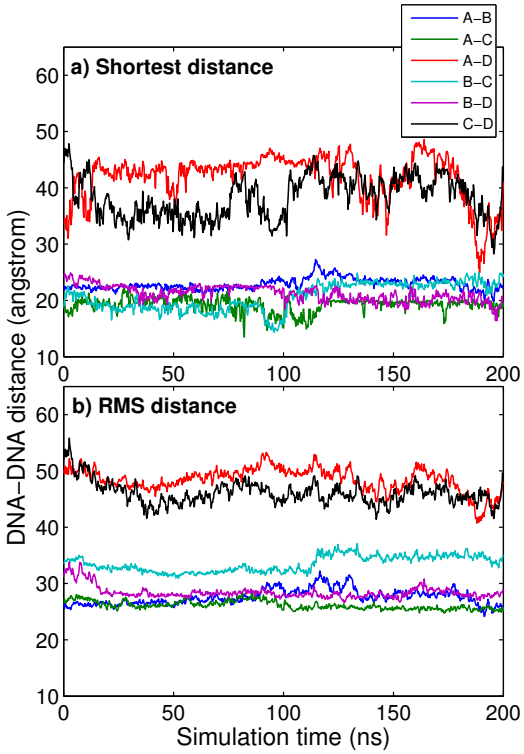

Figure 4.12: DNA-DNA distances in the 4D-28P system which contains four DNAs (A, B, C, and D): (a) shortest distance; (b) root mean square distance (see texts for definition of these distances).

Figure 4.11 for 4D-16P, we observe a similar characteristic as seen in Figure 4.9 for the pairs of DNAs bridged by PEIs (A--B, A--C, B--C, B--D), while, for the pairs of DNAs not bridged by PEIs (A--D, C--D), the d_{RMS} and d_{shortest} values are much larger than 23 Å and 28 Å. In Figure 4.12 for the 4D-28P system, the curves for the DNA pairs bridged by PEIs (A--B, A--C, B--C, B--D) remain almost constants during the entire 200 ns simulation, with fluctuations at a similar magnitude as their counterparts in the 4D-16P system after 50 ns. The DNAs pairs not bridged in the 4D-16P system (A--D, C--D) stayed separate during the simulation of the 4D-28P system. This demonstrates that the added PEIs, although can bind to the DNAs and even replace the previously attached PEIs as shown earlier, do not affect the DNA-DNA spacing in the aggregate.

During the simulations, d_{shortest} of two DNAs can be less than 20 Å, such as the DNA pair in the 2D-8P system at around 18 ns ($d_{\mathrm{shortest}} \sim 12$ Å) and the DNA pair A-C in the 4D-16P system at around 30 and 120 ns ($d_{\mathrm{shortest}} \sim 15$ Å). This is unexpected as the diameter of a DNA molecule is about 20 Å. By further examining the simulation trajectories, it was revealed that the two DNAs adopted an L or T shape arrangement with one end of a DNA nearly perpendicular to the other DNA (see Figure C.3 in Appendix C). In such configurations, d_{shortest} does not reflect the shortest interduplex distance of DNA segments in compact plasmid DNAs, which is the distance between the axes of two parallel DNA segments. Table 4.3 summarizes average d_{shortest} and d_{RMS} during the last 40 ns of the simulations for the bridged DNA pairs in the 2D-8P, 4D-16P and 4D-28P systems. Average d_{shortest} for the 9 pairs of DNAs is 21.4 Å, and average d_{RMS} is 29.0 Å. Because of the special configurations mentioned above for the short DNAs simulated in this work, the actual average DNA-DNA spacing in plasmid DNA/PEI complex should be larger than the average d_{shortest} (21.4 Å). As d_{RMS} in our calculations characterizes the average distance between two short DNAs that may not have parallel axes, the average DNA-DNA spacing in plasmid DNA/PEI complex should be smaller than the average d_{RMS} (29.0 Å) obtained here. Hence, we believe that the average DNA-DNA spacing in plasmid DNA/PEI complex should be between 21.4 Å and 29.0 Å. The DNA interduplex distance in bacteriophages is $\sim$27 Å [41] and that for DNAs

Table 4.3: DNA-DNA distance averaged over the last 40 ns of the simulations (Å).

System	2D-8P	4D-16P				4D-28P				Average
DNA-DNA	A-B	A-B	A-C	B-C	B-D	A-B	A-C	B-C	B-D	
d_{shortest}	23.2	22.3	17.1	21.4	23.1	22.7	19.4	23.1	20.0	21.4
d_{RMS}	27.7	26.7	27.1	33.3	29.9	27.3	25.5	34.7	28.4	29.0

wrapped around the histone core of nucleosomes was also found to be $\sim$27 Å [1]. Our simulation results demonstrate that the DNAs in the DNA/PEI complex are as compact as those in bacteriophages and nucleosomes. The spacing between DNAs condensed by 35 kDa poly-L-arginine was reported to be $\sim$28 Å in a recent experimental study [17], within the range of spacing obtained from our simulations. We have also plotted the radii of gyration of the DNAs in each aggregates (see Figure C.5 in Appendix C), which generally follow a similar trend as ‘‘RMS distance’’ shown in Figures 4.9 to 4.12.

4.3.4 Implications

On the theoretical front of investigating DNA aggregation, Savelyev and Papoian studied the inter-DNA interaction in NaCl and KCl solution using all-atom MD simulations and generated the repulsive interaction potential profiles of two parallel DNA oligomers [43]. Dai *et al.* performed a series of all-atom MD simulations to study DNA attraction mediated by multivalent ions including putrescine (2+), spermidine (3+), spermine (4+), and cobalt hexamine (3+), the inter-DNA interaction potential profiles were calculated, and the dynamics of the complexes was investigated [44]. The above two works, however, only studied the inter-DNA interaction mediated by monovalent and multivalent ions. To the best of our knowledge, the work presented here is the first all-atom MD simulation of cationic polymer mediated DNA aggregation involving multiple DNA molecules.

Our simulations revealed dynamics of the PEI mediated DNA aggregation, which is unaccessible through experiments. It is likely that the obtained results will be applicable to other polymeric carriers apart from PEI. During the simulations of DNA-PEI aggregation, bridging PEIs were observed connecting multiple DNA molecules even though they were initially confined to a single DNA molecule. The

number and binding strength of such bridging PEIs are likely to dictate the stability of aggregates. The bridging PEIs, however, were found to be highly dynamic and the binding to multiple DNA molecules were reversible. Only a select number of PEIs participated in DNA bridging and some remained exclusively committed to DNA molecules that they were originally bound to (within the limitation of simulation times). A critical PEI/DNA ratio was needed for stable aggregation; whereas 4 PEI molecules/DNA (corresponding to mass ratio of $\sim$0.31) gave stable complexes, 1 PEI molecule/DNA (corresponding to mass ratio of $\sim$0.08) was found insufficient to maintain DNA aggregation. Under experimental conditions, we previously reported a polymer/DNA mass ratios of $\sim$0.4 for complete DNA binding irrespective of the molecular weight of PEI [45]. Since almost complete DNA binding by carriers is a prerequisite for DNA aggregation, the simulations results were consistent with the experimentally investigated DNA aggregation. It was worthwhile to note that aggregation successfully occurred even though there was a net positive charge for individual PEI/DNA complexes (each D-4P complex carries a net charge of +2). Local attractive forces have compensated for the overall repulsive force associated with like-charged entities. An overall positive charge is typically observed in PEI/DNA complexes prepared with excess PEI [7], and in this regard, our simulations concurred with the experimental observations.

A significant observation derived from the current studies was the ability of free PEI molecules to replace PEI molecules already bound to DNA molecules in an aggregate. This process will have implications for preparation of DNA complexes to be used for transfection, as well as for dissociation of DNA complexes essential for functional transfection. The fact that excess DNA-binding molecules might displace already bound molecules might be utilized to better engineer DNA complexes prepared by step-by-step addition of transfection complexes [46], where complexes are prepared by sequential addition of constituent molecules. For example, A-B-C-D complexes were prepared by complexing DNA (A) with polymeric polycations (B), followed by the addition of lipophilic substituents (C) and cell-binding moieties (D). Our results support the experimental observation that such sequential addition of DNA-binding molecules can lead to stable incorporation of

individual constituents added at later stages of assembly. Optimizing the complex properties in this approach were mostly experimentally driven, and the current study provide an alternative means to further fine-tune the bound state of the complexes. It might be possible to compare the relative binding affinity of substituents *in silico* beforehand and determine which molecules might be more stably incorporated into the complexes. The final state of the specific molecules in the complexes might be better understood in this way. For example, MD simulations could be used to determine if cell-binding moieties are embedded in a complex or retained on the surface readily accessible for target binding.

Macromolecular displacement of DNA-binding molecules in aggregates might be especially critical for complex stability. Interaction with naturally occurring macromolecules in the extracellular matrix (e.g., heparan sulfate) has undesirable consequence on transfection [47, 48], which leads to inhibition of DNA uptake. MD simulations can help identify DNA carriers resistant to such displacements and enhance the stability of DNA aggregates that is needed for cellular uptake. Displacement with intracellular macromolecules, on the other hand, is essential to release the DNA in free form for efficient transcription [49, 50]. Specific proteins capable of interacting with DNA complexes were identified [51, 52], whose binding to complexes were shown to facilitate nuclear uptake and ultimately transgene expression. The atomistic MD simulations might shed important insight into this process, providing a better means to predict the complex stability in the presence of these intracellular molecules. Simulating displacement of carrier molecules with known DNA-binding molecules such as histones will reveal information about DNA disassembly inside the cells.

4.3.5 Limitations and future studies

Even though the simulations reported here are the-state-of-the-art in terms of model size and simulation time, the current computational limitations restricted us to focusing on only one species of PEI, where the protonation state, architecture, and molecular weight were fixed. Such factors can potentially affect PEI mediated DNA aggregation and will be part of our future work. We already know that PEIs

with lower degree of protonation display lower binding affinity to DNA (i.e., binding is less stable in MD simulations) [31]. The linear form of the PEI also behaves in a similar fashion as compared to the branched PEIs simulated here [31]. The level of protonation employed in this study was more realistic of PEI's state under physiological conditions (i.e., pH between 6 and 7) [30], and branched PEI is more commonly used for DNA delivery. On the basis of our past simulations for 23% protonated PEIs [31], which mimics PEI's state at pH 8 [30], several conjectures can be made on DNA aggregation by the less protonated PEIs. First, due to their lower binding affinity to DNA [31], the aggregates mediated by less protonated PEIs would be less stable and the PEI exchange among the DNA molecules might be more common. Because a considerable fraction of Ns in 23% PEIs contribute to binding with DNA through indirect hydrogen bonding mediated by water molecules [31], we expect the 23% PEIs to neutralize the DNA at a larger distance from the DNA C1' atoms compared with the 46% PEIs; i.e., the capability of 23% PEIs to neutralize the DNA would be weaker. This might cause the formed aggregate in excess PEIs to be less positively charged. The looser binding of 23% PEI to DNA may also increase the DNA-DNA spacing in the aggregate. These conjectures will be tested via additional simulations.

Another limitation is the size of DNA aggregates studies here, where the largest aggregate was composed of 4 DNA molecules leading to a size of $\sim$10 nm with excess PEIs. The aggregate size formed with the PEI/DNA complexes are larger in reality. For example, $\sim$100 nm aggregates are routinely reported for the 25 kDa branched PEI, and we recently reported aggregates as large as 500-700 nm for the PEIs with MW of 0.6-2.0 kDa [45]. Therefore, a larger numbers of DNA complexes will be needed to realistically simulate DNA aggregation employed for cell transfections. Alternatively, longer chain DNA molecules might be needed to achieve more realistic aggregation. MD simulations can help understand what determines the aggregate size obtained under experimental conditions and, more importantly, what makes the aggregation stop before an exuberant aggregate is formed consuming all of the DNA and PEI molecules. Since size is important in the transfection efficiency of the DNA aggregates [53, 54], better control of transfection could be

achieved as a result of such simulations.

4.4 Conclusions

We performed a series of all-atom MD simulations to study PEI mediated DNA aggregation. The results clearly demonstrate that PEIs contribute to DNA aggregation through two mechanisms: (i) forming polyion bridges between DNA segments and (ii) screening the negative DNA charges at a short distance from the surface of DNA molecules. As a consequence of the latter mechanism, the PEI/DNA charge ratio needs to be above certain value in order to maintain a stable aggregation. Compared with monovalent ions, PEIs are shown to be more capable of neutralizing the DNAs at close distance and provide full neutralization at $\sim$12 Å from the DNA C1' atoms, when the PEI/DNA charge ratio is above 1. The DNA-DNA spacing in the DNA/PEI aggregates were between 21.4 Å and 29.0 Å. Excess PEIs were capable of binding to the already positively charged aggregate and further increase its charge. They can also replace the PEIs previously bound to the DNAs in the aggregate. The binding of excess PEIs, however, does not change the DNA-DNA spacing.

Bibliography

[1] K. Luger, A. Mader, R. Richmond, D. Sargent, T. Richmond, *Nature* **1997**, *389*, 251--260.

[2] D. Luo, W. Saltzman, *Nat. Biotechnol.* **2000**, *18*, 33--37.

[3] O. Boussif, F. Lezoualc'h, M. A. Zanta, M. D. Mergny, D. Scherman, B. Demeneix, J. P. Behr, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 7297--7301.

[4] W. Godbey, K. Wu, A. Mikos, *J. Control. Release* **1999**, *60*, 149--160.

[5] A. Kwok, S. L. Hart, *Nanomedicine* **2011**, *7*, 210--219.

[6] M. Abbasi, H. Uludag, V. Incani, C. Y. M. Hsu, A. Jeffery, *Biomacromolecules* **2008**, *9*, 1618--1630.

[7] K. C. R. Bahadur, B. Landry, H. M. Aliabadi, A. Lavasanifar, H. Uludag, *Acta Biomater.* **2011**, *7*, 2209--2217.

[8] M. Bertschinger, A. Schertenleib, J. Cevey, D. L. Hacker, F. M. Wurm, *Mol. Biotechnol.* **2008**, *40*, 136--143.

[9] M. Bertschinger, G. Backliwal, A. Schertenleib, M. Jordan, D. L. Hacker, F. M. Wurm, *J. Control. Release* **2006**, *116*, 96--104.

[10] M. Stevens, *Biophys. J.* **2001**, *80*, 130--139.

[11] R. Winkler, M. Steinhauser, P. Reineker, *Phys. Rev. E* **2002**, *66*.

[12] Y. Hayashi, M. Ullner, P. Linse, *J. Chem. Phys.* **2002**, *116*, 6836--6845.

[13] Y. Hayashi, M. Ullner, P. Linse, *J. Phys. Chem. B* **2003**, *107*, 8198--8207.

[14] Y. Hayashi, M. Ullner, P. Linse, *J. Phys. Chem. B* **2004**, *108*, 15266--15277.

[15] R. Dias, A. Pais, M. Miguel, B. Lindman, *J. Chem. Phys.* **2003**, *119*, 8150--8157.

[16] J. Ziebarth, Y. Wang, *J. Phys. Chem. B* **2010**, *114*, 6225--6232.

[17] J. DeRouchey, V. A. Parsegian, D. C. Rau, *Biophys. J.* **2010**, *99*, 2608--2615.

[18] L. Li, S. A. Pabit, S. P. Meisburger, L. Pollack, *Phys. Rev. Lett.* **2011**, *106*.

[19] X. Qiu, V. A. Parsegian, D. C. Rau, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 21482--21486.

[20] J. Ziebarth, Y. Wang, *Biophys. J.* **2009**, *97*, 1971--1983.

[21] G. M. Pavan, A. Danani, S. Pricl, D. K. Smith, *J. Am. Chem. Soc.* **2009**, *131*, 9686--9694.

[22] G. M. Pavan, M. A. Kostiainen, A. Danani, *J. Phys. Chem. B* **2010**, *114*, 5686--5693.

[23] G. M. Pavan, L. Albertazzi, A. Danani, *J. Phys. Chem. B* **2010**, *114*, 2667--2675.

[24] G. M. Pavan, M. A. Mintzer, E. E. Simanek, O. M. Merkel, T. Kissel, A. Danani, *Biomacromolecules* **2010**, *11*, 721--730.

[25] D. Ouyang, H. Zhang, D.-P. Herten, H. S. Parekh, S. C. Smith, *J. Phys. Chem. B* **2010**, *114*, 9220--9230.

[26] D. Ouyang, H. Zhang, H. S. Parekh, S. C. Smith, *J. Phys. Chem. B* **2010**, *114*, 9231--9237.

[27] D. Ouyang, H. Zhang, H. S. Parekh, S. C. Smith, *Biophys. Chem.* **2011**, *158*, 126 -- 133.

[28] D. Jere, H. L. Jiang, R. Arote, Y. K. Kim, Y. J. Choi, M. H. Cho, T. Akaike, C. S. Chot, *Expert Opin. Drug Del.* **2009**, *6*, 827--834.

[29] H. M. Aliabadi, B. Landry, R. K. Bahadur, A. Neamnark, O. Suwantong, H. Uludag, *Macromol. Biosci.* **2011**, *11*, 662--672.

[30] K. Utsuno, H. Uludag, *Biophys. J.* **2010**, *99*, 201--207.

[31] C. Sun, T. Tang, H. Uludag, J. E. Cuervo, *Biophys. J.* **2011**, *100*, 2754--2763.

[32] C. Sun, T. Tang, *J. Adhes. Sci. Technol.* **2012**, DOI:10.1080/01694243.2012.693830, In press.

[33] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. MacKerell, Jr., *J. Comput. Chem.* **2010**, *31*, 671--690.

[34] B. Brooks, R. Bruccoleri, D. Olafson, D. States, S. Swaminathan, M. Karplus, *J. Comput. Chem.* **1983**, *4*, 187--217.

[35] A. MacKerel Jr., C. Brooks III, L. Nilsson, B. Roux, Y. Won, M. Karplus, *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*, Vol. 1 of *The Encyclopedia of Computational Chemistry*, P. v. R. Schleyer et al. (Ed.), John Wiley & Sons: Chichester, **1998**, pp. 271--277.

[36] J. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. Skeel, L. Kale, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781--1802.

[37] W. Jorgensen, *J. Am. Chem. Soc.* **1981**, *103*, 335--340.

[38] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089--10092.

[39] J. Ryckaert, G. Ciccotti, H. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327--341.

[40] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33--38.

[41] M. Cerritelli, N. Cheng, A. Rosenberg, C. McPherson, F. Booy, A. Steven, *Cell* **1997**, *91*, 271--280.

[42] X. Qiu, D. C. Rau, V. A. Parsegian, L. T. Fang, C. M. Knobler, W. M. Gelbart, *Phys. Rev. Lett.* **2011**, *106*, 028102.

[43] A. Savelyev, G. A. Papoian, *J. Am. Chem. Soc.* **2007**, *129*, 6060--6061.

[44] L. Dai, Y. Mu, L. Nordenskiöld, J. R. C. van der Maarel, *Phys. Rev. Lett.* **2008**, *100*, 118301.

[45] R. K. C. Bahadur, H. Uludag, *J. Biomat. Sci. Poly. Ed.* **2011**, *22*, 873--892.

[46] K. Kostarelos, A. Miller, *Chem. Soc. Rev.* **2005**, *34*, 970--994.

[47] M. S. Shim, X. Wang, R. Ragan, Y. J. Kwon, *Microsc. Res. Tech.* **2010**, *73*, 845--856.

[48] S. Danielsen, S. Strand, C. Davies, B. Stokke, *Biochim. Biophys. Acta* **2005**, *1721*, 44--54.

[49] M. Thibault, S. Nimesh, M. Lavertu, M. D. Buschmann, *Mol. Ther.* **2010**, *18*, 1787--1795.

[50] H. H. Chen, Y.-P. Ho, X. Jiang, H.-Q. Mao, T.-H. Wang, K. W. Leong, *Mol. Ther.* **2008**, *16*, 324--332.

[51] F. M. Munkonge, V. Amin, S. C. Hyde, A.-M. Green, I. A. Pringle, D. R. Gill, J. W. S. Smith, R. P. Hooley, S. Xenariou, M. A. Ward, N. Leeds, K.-Y. Leung, M. Chan, E. Hillery, D. M. Geddes, U. Griesenbach, E. H. Postel, D. A. Dean, M. J. Dunn, E. W. F. W. Alton, *J. Biol. Chem.* **2009**, *284*, 26978--26987.

[52] M. Kaouass, R. Beaulieu, D. Balicki, *J. Control. Release* **2006**, *113*, 245--254.

[53] K. Kunath, A. von Harpe, D. Fischer, T. Kissel, *J. Control. Release* **2003**, *88*, 159--172.

[54] Y. Higuchi, S. Kawakami, S. Fumoto, F. Yamashita, M. Hashida, *Biol. Pharm. Bull.* **2006**, *29*, 1521--1523.

Chapter 5
Probing the Effects of Lipid Substitution on Polycation Mediated DNA Aggregation: A Molecular Dynamics Simulations Study [1]

5.1 Introduction

Gene delivery has been extensively studied in the past two decades as a means to treat diseases associated with defective gene expression [1, 2]. Although there have been as many as 1786 clinical trials to date [3], the lack of safe and efficient gene delivery carriers is still a major impediment for the successful application of such treatment. Polycations, such as polyethylenimine (PEI) [4, 5], are an important category of nonviral carriers since they are effective and do not arouse the safety concerns associated with viral carriers [6, 7]. Moreover, compared with viral carriers, polycationic carriers have the advantage of being readily engineered with other functional groups, making it possible to tailor their properties for different applications. Experimentally, it has been found that modifying polycations with lipophilic and hydrophobic moieties can enhance the performance of polycation-based gene delivery carriers [8]. Khalil *et al.* [9] and Pham *et al.* [10] investigated the cellular interaction and transfection efficiency of lipid modified peptides, and found that the lipid modification yielded more stable polyplexes and led to higher cellular uptake. Hydrophobic modification of chitosan was also found to facilitate DNA condensation by forming stable polyplexes with DNA and to enhance gene delivery with improved cell entry [11--13]. Lipid modification of poly-L-lysine was found to

[1] A version of this chapter has been published. Reprinted with permission from: C. Sun, T. Tang, H. Uludağ, Biomacromolecules, 2012, 13 (9), p2982-2988. Copyright 2012 American Chemical Society.

greatly enhance the DNA delivery efficiency due to increased cellular uptake and better protection from DNA degradation [14--17]. Neamnark *et al.* studied the delivery and transfection efficiency of 2 kDa PEIs modified with caprylic, myristic, palmitic, stearic, oleic and linoleic acids [18]. They found that as little as one lipid substitution per PEI (with linoleic acid substituent) could transform the ineffective 2 kDa native PEI into effective carriers comparable with highly effective 25 kDa PEI. Bahadur *et al.* investigated the efficiency of 0.6, 1.2, and 2 kDa PEIs modified by palmitic acid, and found that the lipid substitution led to a higher zeta potential of the formed polyplexes, increased cell uptake of the DNA, and enhanced transgene expression [19]. Despite the experimental evidence, the molecular mechanism contributing to the beneficial effects of such modification in gene delivery is not clear, and remains to be investigated.

Two recent works have attempted to address the effect of lipids on the aggregation of DNA. Patel and Anchordoquy experimentally investigated the spermine and lipospermine induced DNA condensation [20]. They found that while lipospermines gave higher DNA binding affinity due to their higher hydrophobicity, they lacked the capacity to condense the DNA into compact toroidal structures. The steric hindrance introduced by the acyl chain in lipospermine was postulated to preclude packaging of DNA into compact dimensions. Posocco *et al.*, using mesoscopic coarse-grained simulations, studied the binding of cholesterol-modified dendrimers to DNAs [21]. It was shown that the cholesterol modified dendrimers could form self-assembly through the interaction among the hydrophobic units (cholesterol), which was a reason they could condense DNA more effectively compared with native dendrimers. The influence of lipids on peptide aggregation was investigated in the recent work by Hung *et al.* [22] and Todorova *et al.* [23] via computer simulations. They found that in the absence of lipids, peptides manifested higher flexibility and aggregated through interactions among the aromatic cores. In the presence of lipids, the head lipid groups more favorably interacted with the hydrophilic regions on the peptides while the lipid tails mainly interacted with the hydrophobic regions. Such interactions interfered with the interactions among the aromatic cores and prohibited the aggregation of peptides. Clearly, lipids can have

different roles that may contribute positively or negatively to the aggregation; this demands a careful examination on DNA aggregation mediated by lipid modified polycations.

Computer simulations, especially all-atom molecular dynamics (MD) simulations, have proven to a powerful tool in studying interaction of polycations with nucleic acids, viruses and other drug molecules [24--35]. For example, besides the works mentioned above, all-atom MD simulations has been used to evaluate the ability of different copolymers to incorporate lipophilic drugs into micelles, which yielded results in good comparison with experimental data [35]. In this work, in order to elucidate the role of lipid substitution in polycation mediated DNA aggregation and condensation, we performed a series of large scale all-atom MD simulations. PEI was considered as the representative polycation and oleic acid (C18, 1) as the representative lipid substituent in this study. Our study determined the location of the lipid moieties in the formed polyplexes, and shed light on the effects of lipid substitution on DNA binding and aggregation.

5.2 Methods

5.2.1 Simulated Systems and Procedure

The DNA simulated in this work was a Drew-Dickerson dodecamer d(CGCGAATT CGCG)$_2$ carrying a total charge of -22. The initial structure of the DNA was built to be a canonical B form using AMBER NAB tool [36]. The lmPEI simulated is a 831 Da branched PEI consisting of 13 amine groups with a single oleic acid (OA) lipid tail grafted on a primary amine. The chemical structure and protonation sites of the lmPEI are shown in Figure 5.1. A total of six primary or secondary amines were chosen to be protonated corresponding to a protonation ratio of 46%. We chose the 46% protonation ratio to be consistent with the protonation ratio of 47% for 600 Da PEI at pH $=$ 6 from our recent study [37]. The protonation sites were distributed as uniformly as possible to minimize thermodynamic interactions between the protonated amines [32]. The initial structure of the lmPEI was built in VMD [38] and then energetically minimized in NAMD [39]. Five separate systems

were simulated in this study, and their information is summarized in Table 5.1. To get an equilibrated configuration of the lmPEI, system lmPEI consisting of a single lmPEI molecule with explicit water and counterions was first simulated for 6 ns. The configuration of the lmPEI at the end of the simulation was adopted as the initial configuration of the lmPEIs in the following simulations. To study the interaction of one DNA with one lmPEI, and to determine the location of the cationic and lipophilic moieties of the lmPEI relative to the DNA, system 1D-1P containing one DNA and one lmPEI was simulated for 50 ns. The initial configuration of system 1D-1P is shown in Figure 5.2(a), where the principal axes of the lmPEI were aligned parallel to the DNA axis and its center of mass (COM) was positioned at 25 Å from the DNA COM. To investigate the binding of multiple lmPEIs to a DNA, we adopted the final configuration of 1D-1P and added three more lmPEIs to form the new system 1D-4P. The added 3 lmPEIs in 1D-4P were again aligned parallel to the DNA axis with their COM at 25 Å away from the DNA COM (Figure 5.2(b)). The 1D-4P system was simulated for 100 ns. Four identical equilibrated 1D-4P complexes were then used to construct the system 4D-16P to study lmPEI mediated DNA aggregation. The four 1D-4P complexes were arranged on the four corners of a square, as shown in Figure 5.2(c). The axes of the four DNAs were aligned to be parallel and the COM of each 1D-4P complex was separated from the COM of its neighboring complex by 35 Å. The 4D-16P system was simulated for 100 ns. To investigate the effect of excess lmPEIs on the DNA aggregation, 12 lmPEIs were added to the 4D-16P system at the end of the 100 ns simulation, and the new system is referred to as 4D-28P. The added 12 lmPEIs surrounded the 4D-16P polyplex in a circular fashion located at 42 Å from the COM of 4D-16P system as shown in Figure 5.2(d). The 4D-28P system was simulated for 200 ns. It should be pointed out that there are many ways of specifying the initial configurations for systems 4D-16P and 4D-28P. One particular reason the present initial configurations are chosen is that the same initial configurations were used in a previous work on native PEI mediated DNA aggregation [33]. Having the same initial setting allows us to best address the influence of lipid in the DNA aggregation. To facilitate the data presentation, in systems 4D-16P and 4D-28P, each DNA is labeled with a cap-

Table 5.1: Information of the five systems simulated in this study.

System name	No. of DNA/lmPEI	Charge ratio DNA/lmPEI	No. of atoms	Simulation box size ($Å^3$)	No. of Na+/Cl-	Time (ns)
lmPEI	0/1	0/6	12856	$64 \times 48 \times 41$	0/6	6
1D-1P	1/1	22/6	34210	$74 \times 69 \times 66$	22/6	50
1D-4P	1/4	22/24	65285	$81 \times 92 \times 86$	0/2	100
4D-16P	4/16	88/96	96278	$104 \times 105 \times 87$	0/8	100
4D-28P	4/28	88/168	163034	$117 \times 117 \times 117$	0/80	200

ital letter (A, B, C or D), and each lmPEI is labeled with a number (1-16 in 4D-16P; 17-28 for the additional 12 lmPEIs in 4D-28P).

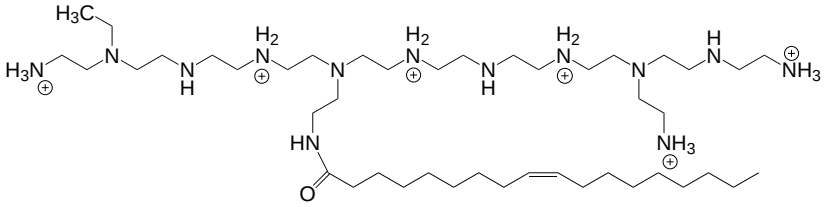

Figure 5.1: Molecular structures and protonation sites of the lipid modified PEI studied in this work.

5.2.2 Simulation Details

A CHARMM format force field was developed and validated [32] for PEI based on the CHARMM General Force Field [40], and CHARMM 27 force field [41, 42] was used for all other molecules. All simulations were performed using NAMD [39]. A time step of 2 fs, TIP3P water model [43], periodic boundary condition, full electrostatics with particle-mesh Ewald method [44], cutoff of 12 Å for van der Waals interactions and electrostatics pairwise calculations, and the SHAKE algorithm [45] were used for all the simulations. During each simulation, the system was first minimized for 5000 steps. The system was then heated from 0 to 300 K in 20 ps with 10 $\mathrm{kcal}/(\mathrm{mol}\times Å^2)$ harmonic restraint on the non-hydrogen atoms of the DNAs and lmPEIs. The simulation was continued for 4 ns at 300 K and 1 bar with the restraint to have the ions relax around the solutes. We then removed the restraint and NPT ensemble simulation was performed for the period of time indicated in Table 5.1 for each system. The length of the simulations was shown to be sufficient to generate

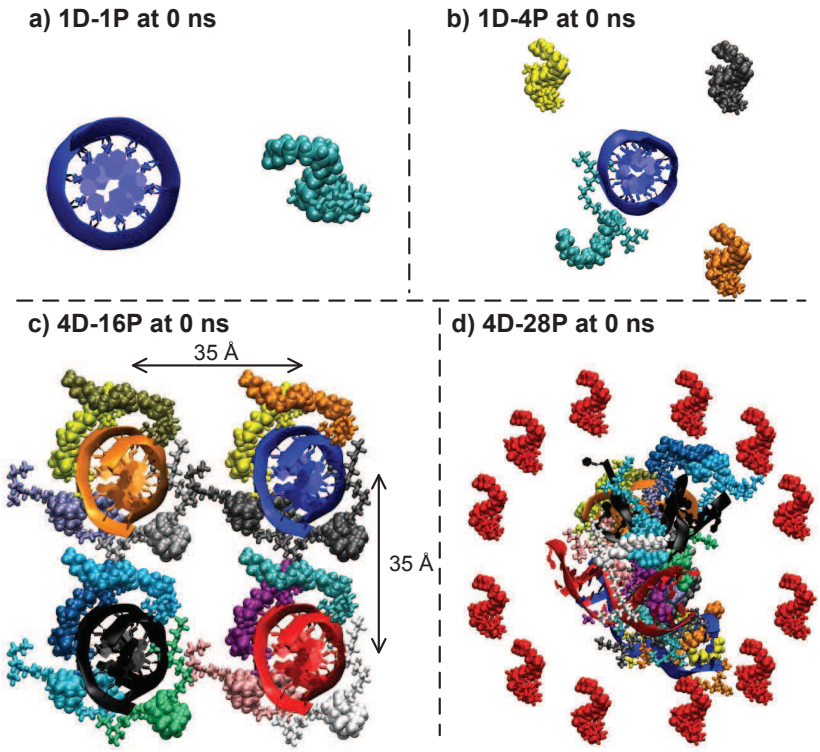

Figure 5.2: Snapshots of the initial configurations: (a) 1D-1P at 0 ns, (b) 1D-4P at 0 ns, (c) 4D-16P at 0 ns, (d) 4D-28P at 0 ns. Different lmPEIs are represented by different colors (except in (d) where the extra 12 lmPEIs are in red); the OA moieties on the lmPEIs are represented by spheres; water and counterions are removed for clarity.

dynamic equilibrium, an evidence of which is given in Appendix D. VMD [38] was used for visualization and trajectories analysis.

5.3 Results and Discussions

5.3.1 Location of Lipophilic Moieties and Lipid Association

Although the size and charge of the carrier/DNA polyplexes are routinely assessed in an experimental setting, the structural details especially the location of the lipophilic moieties in the lipid-modified polycation/DNA polyplexes have not been investigated [8]. Figure 5.3 shows the final configurations for systems 1D-1P, 1D-4P, 4D-16P, and 4D-28P, where the lipid moieties on the lmPEIs are represented by spheres. In system 1D-1P (Figure 5.3(a)), the cationic moiety of the lmPEI conforms to the DNA while the lipid tail remains on the outside with no obvious interactions with the DNA molecule. In system 1D-4P (Figure 5.3(b)), the four lipid tails still stay outside of the complex with three of them being associated with one another. In system 4D-16P (Figure 5.3(c)), a DNA aggregate is formed and a large lipid association involving multiple lmPEIs is formed in the middle of the four DNA molecules. The remaining lipids also stay associated with one another on the periphery of the DNA aggregate. In system 4D-28P (Figure 5.3(d)), the DNA aggregate and the large lipid association continue to exist. In addition, some of the added lmPEIs are attached to the outer surface of the formed DNA aggregate.

To quantify the location of cationic and lipophilic moieties of the lmPEIs relative to the DNA, in Figure 5.4 we plotted the cumulative percentage of the lmPEI nitrogens and lipid carbons for system 1D-4P as a function of distance from any DNA C1' atom, averaged over the last 40 ns of the simulation. The C1' atoms are on the sugar rings of the DNAs, located inside the DNA helix at a distance of $\sim$5 Å from the surface of DNA defined by the phosphorus atoms. In Figure 5.4, taking the lmPEI nitrogens for example, the cumulative percentage at a given distance r is the percentage of all the lmPEI nitrogens within a distance r from any DNA C1' atoms. The curve for the lmPEI nitrogens rises quickly from 0% at 4 Å to $\sim$90% at 9 Å, demonstrating that most lmPEI nitrogens stay between 4 and 9 Å from the

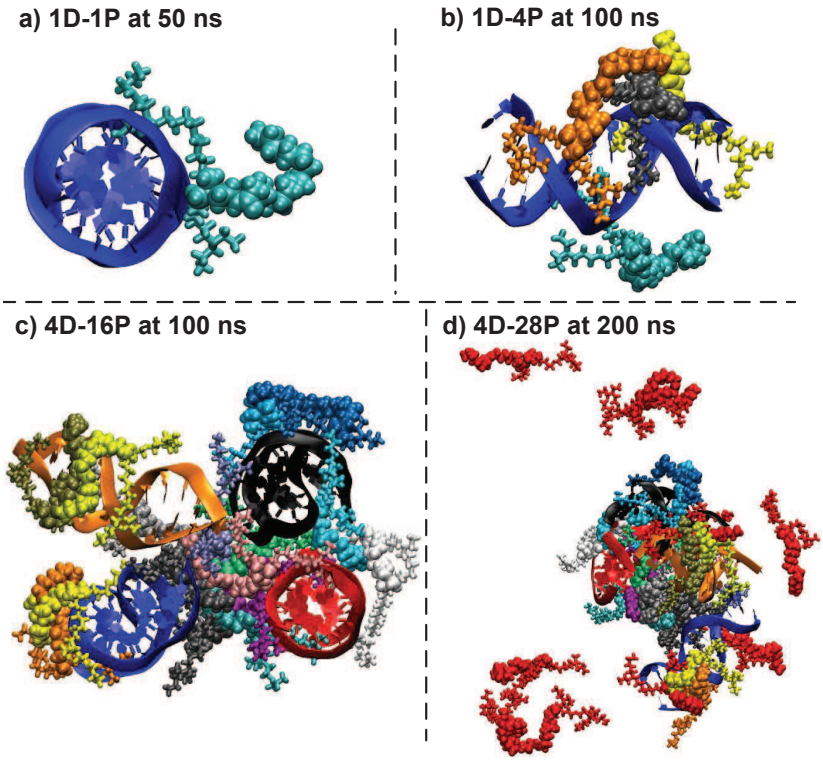

Figure 5.3: Snapshots of the final configurations: (a) 1D-1P at 50 ns, (b) 1D-4P at 100 ns, (c) 4D-16P at 100 ns, (d) 4D-28P at 200 ns. Different lmPEIs are represented by different colors (except in (d) where the extra 12 lmPEIs are in red); the OA moieties on the lmPEIs are represented by spheres; water and counterions are removed for clarity.

DNA C1' atoms. The curve for the lipid carbons only reaches ∼30% at 10 Å, indicating that only ∼30% of the lipid carbons are within 10 Å of the DNA C1' atoms. These curves clearly show that, as observed visually in Figure 5.3(b), the cationic moieties bind closely to the DNA while the lipid substituents tend to stay away from the DNA instead of being located inside the DNA grooves.

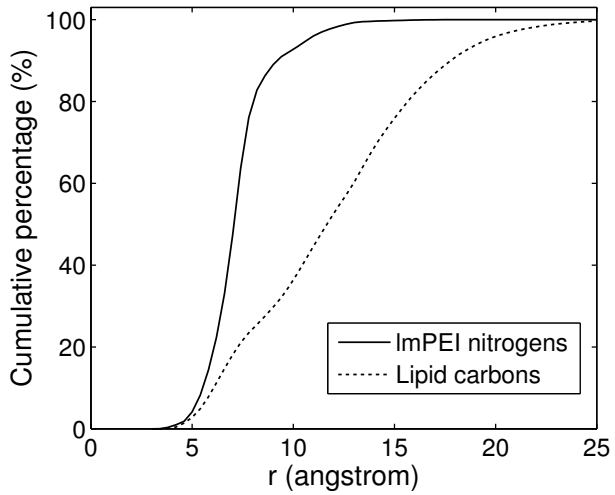

Figure 5.4: Cumulative percentage of the lmPEI nitrogens and lipid carbons as functions of the distance from any DNA C1' atom in system 1D-4P.

As seen in Figure 5.3(c), the lipid moieties that stay on the periphery of the lmPEI/DNA complex become associated with one another when multiple lmPEI/DNA complexes are placed together. This can play a significant role in aggregating the DNAs. To quantify the association among the lipid tails in the aggregate, in Figure 5.5 we tabulated, between each pair of lipid tails, the number of pairs of lipid carbons that are closer than 5 Å apart, averaged over the last 40 ns of the simulations. The calculations were performed for both systems 4D-16P and 4D-28P. The numbers on the top and right of each subfigure are the lmPEI indices. Each pair of lmPEIs in a system can form a pair of lipid tails. This results in 120 pairs of lipid tails in systems 4D-16P and 378 pairs in 4D-28P, corresponding to 120 cells in Figure 5.5(a) and 378 cells in Figure 5.5(b), respectively. Each lipid tail has 18 carbons; thus, between a pair of lipid tails there are 324 pairs of carbons. Among these 324 pairs of carbons, the number of pairs within 5 Å is counted and given in the cell corresponding to this pair of lipid tails. The cell is left empty where no

carbon pairs are found to be within 5 Å. For example, number 11 on the top left cell of Figure 5.5(a) means that out of the 324 pairs of carbons between lmPEI 1 and lmPEI 2 in system 4D-16P, 11 pairs are separated by 5 Å or less. We choose 5 Å as the criterion because this is the closet carbon-carbon distance within which the free energy for the association of two alkane molecules is negative indicating that their associaltion is energetically favorable [46]. If one or more pairs of lipid carbons between two lipid tails is closer than 5 Å apart, the two lipid tails are said to be associated. In Figure 5.5(a), consider the rows 1, 2, 6-10 and columns 2, 6-10, 13, all the cells formed by these rows and columns have nonzero numbers, which are marked with red squares. This indicates that lmPEIs 1, 2, 6-10 and 13 are mutually associated and they form a large association involving 8 lmPEIs in system 4D-16P. It can be seen in Figure 5.3(c) that this association stays in the middle of the polyplex. Since the 8 lmPEIs bind to different DNAs, the lipid association contributes to holding the DNAs in an aggregated form. Four other lipid associations each involving only two lmPEIs also exist in the polyplex (illustrated by the squares of green, blue, orange and olive colors in Figure 5.5(a)). Overall, each lmPEI in system 4D-16P is associated with at least one other lmPEIs through lipids, demonstrating the significance of lipid association in the polyplex.

In system 4D-28P, the 12 extra lmPEIs are indexed by numbers 17-28, as shown in Figure 5.5(b). Comparing the columns 2-16 of Figure 5.5(b) with Figure 5.5(a), it can be seen that all the colored squares stay in the same location. This means that, after adding 12 more lmPEIs, the lipid associations formed in system 4D-16P preserve in system 4D-28P. In addition, lmPEI 27 joins the lipid association between lmPEIs 3 and 4 (3 cells marked by green squares in Figure 5.5(b)); lmPEI 20 joins the lipid association between lmPEIs 11 and 12 (3 cells marked by orange squares in Figure 5.5(b)); lmPEIs 19, 21, 23, and 24 form a new association involving four lmPEIs (6 cells marked by violet squares in Figure 5.5(b)); lmPEIs 17 and 18 are associated (cyan square in Figure 5.5(b)); as well lmPEIs 22 and 26 are associated (yellow square in Figure 5.5(b)). Only two lmPEIs, PEIs 25 and 28, are not associated with any other lmPEIs. It should be noted that not all the extra lmPEIs are bound to the DNA aggregate. In particular, by examining the binding state of each

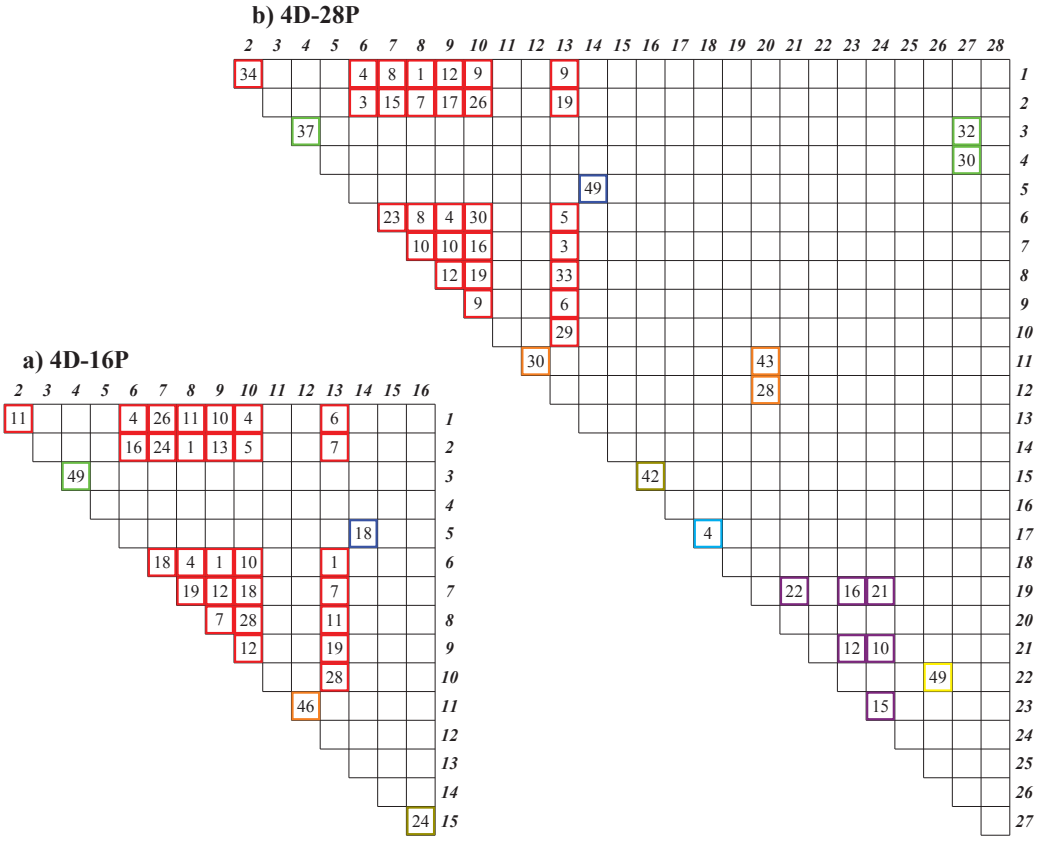

Figure 5.5: Number of pairs of lipid carbons that are closer than 5 Å apart between each pair of lipid tails in (a) 4D-16P, and (b) 4D-28P. The numbers on the top and right of each subfigure are lmPEI indices. Only the non-zero numbers are shown and marked with colored squares. The lmPEIs involved in the same association are marked by the same color.

lmPEI with the DNAs (Figure D.3 in Appendix D), we found that, during the last 40 ns of the simulation, lmPEIs 17, 20, 22, and 27 directly bind to the DNAs for more than 50% of the time. lmPEIs 18 and 26 bind directly to the DNAs for short periods of time; also they attach to the polyplex through lipid association with lmPEIs 17 and 22. lmPEIs 19, 21, 23, and 24 do not bind to any DNAs and exist in an associated form in the solution. lmPEIs 25 and 28 neither bind to the DNAs nor are associated with any other lmPEIs.

5.3.2 Polyion Bridging and DNA Charge Neutralization

Two main mechanisms have been identified in native PEI mediated DNA aggregation [33]: polyion bridging (i.e., a polycation binding with multiple DNA segments simultaneously; see detailed definition in Appendix D) and DNA charge neutralization. Not surprisingly, we found that polyion bridging also plays an important role in lmPEI mediated DNA aggregation. Specially, five lmPEIs participated in bridging two or three DNAs for longer than 50% of the simulation time in both systems 4D-16P and 4D-28P (see Figures D.2 and D.3 in Appendix D). However, the intensity of polyion bridging appears to be slightly weaker than that in native PEI mediated DNA aggregation, as in both the system containing 4 DNAs with 16 native PEIs and the system containing 4 DNAs with 28 native PEIs, eight PEIs participated in bridging DNAs for longer than 50% of the simulation time [33]. We attribute this to the steric disturbance on the polyion bridging arising from the lipid tails on the lmPEIs.

To investigate how lmPEIs neutralize the DNA charges and whether the lipid modification introduces any effect on the charge neutralization, we plotted the cumulative distributions, with respect to the DNA C1' atoms, of protonated PEI nitrogens, Cl- ions, and the net charge of PEI and ions, averaged over the last 40 ns of the simulations (Figure 5.6). The results for lmPEIs (left column) are compared with those for native PEIs [33] (right column). In each subfigure of Figure 5.6, the straight dashed black line indicates the total charge all the DNAs in the system carry, and the blue solid curve is the total charge of PEI and ions within a given distance to their nearest DNA C1' atoms. At the distance where black line and blue curve

intersect, the DNA charges are 100% neutralized by the PEI and ions. At larger distances, the PEI and ions charges exceed the DNA charges, and the DNA(s) are 'overneutralized'. It can be seen that the PEIs in all the 6 systems demonstrate similar characteristics in neutralizing the DNA(s). Quantitatively, the distance at which the lmPEIs 100% neutralize the DNA(s) is shorter than the distance at which the native PEIs 100% neutralize the DNA(s). Specifically, for system 1D-4P, such distance is $\sim$10 Å for lmPEIs and $\sim$12 Å for native PEIs. For system 4D-28P, such distance is $\sim$8 Å for lmPEIs and $\sim$10 Å for native PEIs. The distance at which the 'overneutralization' maximizes is approximately the same for all the 6 system, being $\sim$15 Å. However, in excess of PEIs, the degree of overneutralization is higher for lmPEIs, which can be seen by comparing the peak values of the blue solid curves in Figure 5.6(e) and (f). The overneutralization of the DNAs can also be quantified using the number of PEI molecules bound to the DNAs during the simulation. Here we say a lmPEI is bound to DNA if it has one or more nitrogens within 4 Å of any DNA N/O atoms (see Appendix D for details). Figure 5.7 shows this number as a function of simulation time for the system of 4D-28P with lmPEIs. On average, 19.7 lmPEIs are directly bound with the DNAs during the last 40 ns of the simulation for system 4D-28P. These 19.7 lmPEIs carry a positive charge of 118 and the DNAs carry a negative charge of -88. Besides the lmPEIs directly bound to the DNAs, there are some lmPEIs attached to the polyplex through lipid association with the lmPEIs directly bound to the DNAs. Thus, the resulting polyplex carries a positive charge higher than $+30$. To compare, in the system of 4 DNAs with 28 native PEIs [33], on average 18.2 PEIs were bound to the DNAs during the last 40 ns of the simulation. Overall, our results show that the PEI's capability in neutralizing DNA is slightly enhanced by the lipid substitution.

5.3.3 Water Release during the Aggregation Process

Macromolecular association in aqueous environment is normally accompanied by the release of water molecules previously adhering to the surfaces of the macromolecules. Since the water molecules on the macromolecular surfaces are less mobile, such release is an entropically favorable process with a free energy reduction of

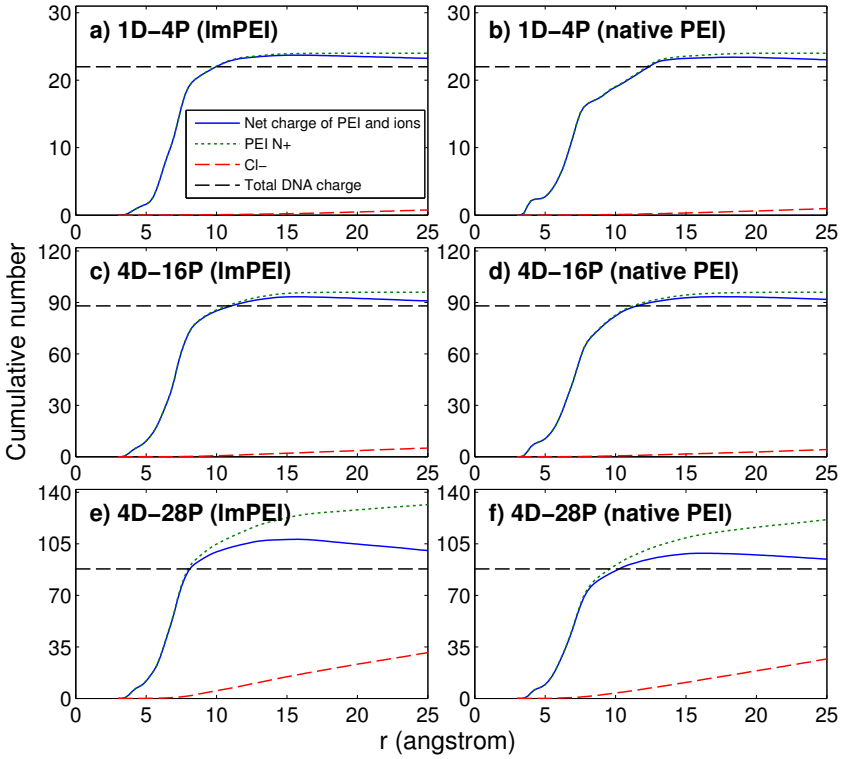

Figure 5.6: Cumulative numbers of PEI charges, Cl-, and net charge of PEI/Cl- as a function of the distance from any C1' DNA atom (averaged over the last 40 ns of each simulation). The total charge of all the DNAs in each system is plotted by a straight dashed black lines as reference. (a) D-4P (lmPEI), (b) D-4P (native PEI), (c) 4D-16P (lmPEI), (d) 4D-16P (native PEI), (e) 4D-28P (lmPEI), (f) 4D-28P (native PEI).

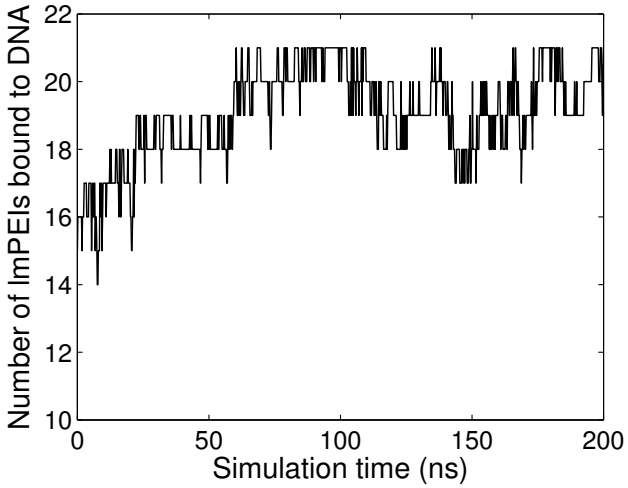

Figure 5.7: Number of lmPEI molecules bound to DNA during the simulation of system 4D-28P.

Table 5.2: Number of water molecules released from the solutes during the aggregation process.

System	4D-16P (lmPEI)	4D-16P (native PEI)	4D-28P (lmPEI)	4D-28P (native PEI)
# of waters	1003	794	1345	844

up to 2 kcal/mol at 300 K [47]. Therefore, counting the number of water molecules released during the macromolecule association allows us to assess the entropic gain from water release upon macromolecular binding. Table 5.2 summarizes the number of water molecules released from the hydration shell of the solutes (within 3 Å from the molecules) during the aggregation process for systems 4D-16P and 4D-28P and their counterparts involving native PEIs [33]. Detailed information on the calculation is given in Appendix D. It can be seen that the numbers of released water molecules are much higher for the lmPEI systems than for the native PEI systems. System 4D-16P has $\sim$20% more water molecules released, and system 4D-28P has as high as $\sim$60% more water molecules released compared to its native PEI counterpart. There are two reasons for the much larger amount of water release in the presence of lmPEIs. First, there are a greater number of lmPEI molecules ($>$19.7) in the lmPEI/DNA polyplex than the number of native PEIs ($\sim$18.2) in the native PEI/DNA polyplex. Second, the lmPEIs in the lmPEI/DNA polyplex are significantly associated, resulting in more water release.

5.3.4 Discussion

From our simulation results, several effects of the lipid substitution on DNA aggregation can be identified. First, compared with the polyplex formed by DNA and native PEIs [33], the existence of hydrophobic moieties on the periphery of the lmPEI/DNA polyplex can present the hydrophobic groups more effectively for interaction with cell membranes and other hydrophobic biological entities on the delivery path. The external location of lipids is expected to facilitate the internalization of the DNAs through cell membranes, supporting the experimentally observed higher cellular uptake of lmPEI/DNA polyplexes compared with native PEI/DNA polyplexes [19]. The peripheral lipids can also enhance the interfacial interaction

among lmPEI/DNA polyplexes and drive their growth into larger polyplexes, which is confirmed in our experimental observation that the size of lmPEI/DNA polyplex became larger after 30 min (unpublished results). Moreover, the lipids on the periphery of the polyplex can presumably reduce the accessibility of degrading nucleases to the DNA molecules and hence protect the DNAs from degradation. Second, the lmPEI/DNA polyplex formed in our simulation has demonstrated enhanced stability compared with native PEI/DNA polyplex [33]. A strong evidence of this is that when we added 12 extra PEIs to the polyplex formed by 4 DNAs and 16 native PEIs, we found that some of the original 16 PEIs were ‘‘replaced’’ by the added PEIs in that they unbound from the DNAs while allowing the newly added PEIs to bind to the DNAs [33]. Interestingly, this did not happen to our current 4D-28P system after adding 12 lmPEIs (see Figure D.3 in Appendix D). All the original 16 lmPEIs bound firmly to the DNAs during the entire 200 ns simulation. We attribute the increased stability to the intensive linkage formed among the lmPEIs through lipid association which we observed in systems 4D-16P and 4D-28P. In fact, despite the steric hindrance associated with the presence of the lipid tails, the radius of gyration of the four DNAs aggregated by lmPEIs is found to be nearly identical to that of the four DNAs aggregated by native PEIs (see Figure D.4 in Appendix D). The lipid association has compensated for the steric hindrance as well as the electrostatic repulsion between the likely charged lmPEIs and allowed the formation of a network in which the lmPEIs collectively aggregate the DNAs and all the DNAs are mutually connected. In contrast, native PEIs work individually in aggregating the DNAs and only a fraction of the DNAs in the aggregate are mutually connected by the native PEIs [33], resulting in polyplexes with an overall lower stability. Another support for the enhanced stability of the lmPEI/DNA polyplex is the significantly larger amount of water molecules released from the lmPEI/DNA polyplex compared with the native PEI/DNA polyplex. Because water release is associated with entropy gain and free energy reduction, more water release contributes favorably to increase the stability of the formed polyplex. The low stability of polyplexes formed by low molecular weight (LMW) native PEIs might be a major reason for the low cellular uptake of these polyplexes. Modifying LMW PEIs

with lipid substitution could overcome this drawback of native LMW PEIs while taking the advantage of the low toxicity associated with LMW PEIs. Finally, in excess of PEIs, the polyplex formed by lmPEIs and DNAs is more positively charged compared with that formed by native PEIs and DNAs. This is consistent with the experimental finding that lmPEI aggregated polyplexes have a higher ζ potential than native PEI aggregated polyplexes [19, 48].

In an experimental investigation of the dissociation of different polylexes formed by 2 kDa native PEI and 23 lmPEIs with different types and amount of lipid substitutions, it was found that 7 of the lmPEIs formed polyplexes that were more difficult to dissociate compared with native PEI formed polyplexes, while the other 16 lmPEIs formed polyplexes that were easier to dissociate [18]. This suggests that the lipid modification might weaken the stability of the polyplex in some cases. Our simulation results have shown that the degree of polyion bridging is slightly weaker for lmPEI mediated DNA aggregation. The weakened polyion bridging by lipid substitution could make the polyplex easier to dissociate while the network formed among the lmPEIs from lipid association could provide more resistance to the polyplex dissociation. For the systems studied in this work, the weakened stability from less intensive polyion bridging is more than compensated by the enhanced stability due to lipid association, and the lmPEI/DNA polyplexes manifest higher stability. However, for lmPEIs modified with different types and amount of lipid substitutions, the significance of these two effects might be reversed. The delicate balance between these two effects provides an explanation for the experimentally observed different dissociation results for lmPEIs with different lipid substitution [18].

5.4 Conclusions

When lmPEIs bind with a DNA, the cationic moieties of the lmPEIs form close contact with the DNA whereas the lipid moieties stay at the periphery. Compared with native PEIs, which aggregate DNAs through polyion bridging and charge neutralization, lmPEIs mediate DNA aggregation through an additional mechanism: asso-

ciation among lipid tails of different lmPEIs. The lipid association is significant and it further stabilizes the lmPEI/DNA polyplex. However, the lipid substitution weakens the polyion bridging and this might have an opposite effect on the stability enhanced by lipid association. The peripheral location of the lipid moieties attached to the lmPEI/DNA polyplex increases the hydrophobicity of the formed polyplex and contributes favorably to the interaction of the polyplex with cell membrane.

Bibliography

[1] R. Mulligan, *Science* **1993**, *260*, 926--932.

[2] D. Schaffert, E. Wagner, *Gene Therapy* **2008**, *15*, 1131--1138.

[3] http://www.wiley.co.uk/genmed/clinical/.

[4] O. Boussif, F. Lezoualc'h, M. A. Zanta, M. D. Mergny, D. Scherman, B. Demeneix, J. P. Behr, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 7297--7301.

[5] W. Godbey, K. Wu, A. Mikos, *J. Controlled Release* **1999**, *60*, 149--160.

[6] F. Ledley, *Hum. Gene Therapy* **1995**, *6*, 1129--1144.

[7] D. Luo, W. Saltzman, *Nature Biotechnol.* **2000**, *18*, 33--37.

[8] V. Incani, A. Lavasanifar, H. Uludag, *Soft Matter* **2010**, *6*, 2124--2138.

[9] I. Khalil, S. Futaki, M. Niwa, Y. Baba, N. Kaji, H. Kamiya, H. Harashima, *Gene Ther.* **2004**, *11*, 636--644.

[10] W. Pham, M. Kircher, R. Weissleder, C. Tung, *ChemBioChem* **2004**, *5*, 1148--1151.

[11] W. Liu, X. Zhang, S. Sun, G. Sun, K. De Yao, D. Liang, G. Guo, J. Zhang, *Bioconjugate Chem.* **2003**, *14*, 782--789.

[12] S. Chae, S. Son, M. Lee, M. Jang, J. Nah, *J. Control. Release* **2005**, *109*, 330--344, 12th International Symposium on Recent Advances in Drug Delivery Systems, Salt Lake City, UT, FEB 21-24, 2005.

[13] F. Hu, M. Zhao, H. Yuan, J. You, Y. Du, S. Zeng, *Int. J. Pharm.* **2006**, *315*, 158--166.

[14] M. Abbasi, H. Uludag, V. Incani, C. Olson, X. Lin, B. A. Clements, D. Rutkowski, A. Ghahary, M. Weinfeld, *Biomacromolecules* **2007**, *8*, 1059--1063.

[15] M. Abbasi, H. Uludag, V. Incani, C. Y. M. Hsu, A. Jeffery, *Biomacromolecules* **2008**, *9*, 1618--1630.

[16] B. A. Clements, V. Incani, C. Kucharski, A. Lavasanifar, B. Ritchie, H. Uludag, *Biomaterials* **2007**, *28*, 4693--4704.

[17] V. Incani, E. Tunis, B. A. Clements, C. Olson, C. Kucharski, A. Lavasanifar, H. Uludag, *J. Biomed. Mat. Research A* **2007**, *81A*, 493--504.

[18] A. Neamnark, O. Suwantong, R. K. C. Bahadur, C. Y. M. Hsu, P. Supaphol, H. Uludag, *Mol. Pharm.* **2009**, *6*, 1798--1815.

[19] K. C. R. Bahadur, B. Landry, H. M. Aliabadi, A. Lavasanifar, H. Uludag, *Acta Biomater.* **2011**, *7*, 2209--2217.

[20] M. Patel, T. Anchordoquy, *Biophys. J.* **2005**, *88*, 2089--2103.

[21] P. Posocco, S. Pricl, S. Jones, A. Barnard, D. K. Smith, *Chem. Sci.* **2010**, *1*, 393--404.

[22] A. Hung, I. Yarovsky, *J. Mol. Graphics & Modelling* **2011**, *29*, 597--607.

[23] N. Todorova, A. Hung, I. Yarovsky, *J. Phys. Chem. B* **2010**, *114*, 7974--7982.

[24] J. Ziebarth, Y. Wang, *Biophys. J.* **2009**, *97*, 1971--1983.

[25] G. M. Pavan, A. Danani, S. Pricl, D. K. Smith, *J. Am. Chem. Soc.* **2009**, *131*, 9686--9694.

[26] G. M. Pavan, M. A. Kostiainen, A. Danani, *J. Phys. Chem. B* **2010**, *114*, 5686--5693.

[27] G. M. Pavan, L. Albertazzi, A. Danani, *J. Phys. Chem. B* **2010**, *114*, 2667--2675.

[28] G. M. Pavan, M. A. Mintzer, E. E. Simanek, O. M. Merkel, T. Kissel, A. Danani, *Biomacromolecules* **2010**, *11*, 721--730.

[29] D. Ouyang, H. Zhang, D.-P. Herten, H. S. Parekh, S. C. Smith, *J. Phys. Chem. B* **2010**, *114*, 9220--9230.

[30] D. Ouyang, H. Zhang, H. S. Parekh, S. C. Smith, *J. Phys. Chem. B* **2010**, *114*, 9231--9237.

[31] G. Doni, M. A. Kostiainen, A. Danani, G. M. Pavan, *Nano Lett.* **2011**, *11*, 723--728.

[32] C. Sun, T. Tang, H. Uludag, J. E. Cuervo, *Biophys. J.* **2011**, *100*, 2754--2763.

[33] C. Sun, T. Tang, H. Uludag, *Biomacromolecules* **2011**, *12*, 3698--3707.

[34] C. Sun, T. Tang, H. Uludag, *J. Phys. Chem. B* **2012**, *116*, 2405--2413.

[35] A. O. Kasimova, G. M. Pavan, A. Danani, K. Mondon, A. Cristiani, L. Scapozza, R. Gurny, M. Moeller, *J. Phys. Chem. B* **2012**, *116*, 4338--4345.

[36] D. Case, T. Darden, I. T.E. Cheatham, C. Simmerling, J. Wang, R. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. Merz, B.Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvary, K.F.Wong, F. Paesani, J. Vanicek, X. Wu, S. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. Mathews, M. Seetin, C. Sagui, V. Babin, P. Kollman, *AMBER 10*, University of California, San Francisco, **2008**.

[37] K. Utsuno, H. Uludag, *Biophys. J.* **2010**, *99*, 201--207.

[38] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33--38.

[39] J. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. Skeel, L. Kale, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781--1802.

[40] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. MacKerell, Jr., *J. Comput. Chem.* **2010**, *31*, 671--690.

[41] B. Brooks, R. Bruccoleri, D. Olafson, D. States, S. Swaminathan, M. Karplus, *J. Comput. Chem.* **1983**, *4*, 187--217.

[42] A. MacKerel Jr., C. Brooks III, L. Nilsson, B. Roux, Y. Won, M. Karplus, *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*, Vol. 1 of *The Encyclopedia of Computational Chemistry*, P. v. R. Schleyer et al. (Ed.), John Wiley & Sons: Chichester, **1998**, pp. 271--277.

[43] W. Jorgensen, *J. Am. Chem. Soc.* **1981**, *103*, 335--340.

[44] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089--10092.

[45] J. Ryckaert, G. Ciccotti, H. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327--341.

[46] A. S. Thomas, A. H. Elcock, *J. Phys. Chem. Lett.* **2011**, *2*, 19--24.

[47] J. Dunitz, *Science* **1994**, *264*, 670.

[48] H. M. Aliabadi, B. Landry, R. K. Bahadur, A. Neamnark, O. Suwantong, H. Uludag, *Macromol. Biosci.* **2011**, *11*, 662--672.

Chapter 6
A Molecular Dynamics Simulation Study on the Effect of Lipid Substitution on Polyethylenimine Mediated siRNA Complexation [1]

6.1 Introduction

RNA interference (RNAi) via small interfering RNA (siRNA) is a promising therapeutic strategy which have attracted intense attention during the past decade [1, 2]. Protective carriers are needed to deliver siRNA to the target site (cytosol) since siRNA is rapidly degraded by nucleases in the biological environment in its native form. siRNA itself cannot permeate cell membrane on its own, given the anionic charge of the cell membrane and the siRNA itself. Delivering siRNA into cells in an effective way is a major impediment for its successful therapeutic applications. Polycationic carriers, such as polyethylenimine (PEI) [3, 4], have evolved into a major approach for siRNA delivery with the advantage of being readily modified with other functional groups, making it possible to tailor their properties for different applications [5]. In addition, polycationic carriers do not arouse the safety concerns associated with viral carriers [6, 7]. High molecular weight (HMW, $\sim$25 kDa) PEIs is one class of effective carriers for siRNA delivery and often considered as 'gold standard' in non-viral gene delivery, however, the high toxicity and limited biodegradability prohibit their clinical use. Low molecular weight (LMW, $<$2 kDa) PEIs display acceptable toxicity but cannot effectively deliver siRNA into cells. Modifying polycations with lipophilic and hydrophobic moieties was found to improve the performance of polycation-based gene delivery carriers [5, 8]. Mod-

[1] A version of this chapter has been accepted for publication. Reprinted from: C. Sun, T. Tang, H. Uludağ, Biomaterials, 2013, DOI:10.1016/j.biomaterials.2013.01.011, in press.

ifying poly-L-lysine (PLL) with a lipid, for example, greatly enhanced the siRNA delivery efficiency with increased cellular uptake and better protection from siRNA degradation [9, 10]. Cationic carriers containing cholesterol were found to improve siRNA delivery through an enhanced interaction with cell membrane [11]. The impact of different lipid substitution on the assembly and delivery of siRNA by 2 kDa PEI was investigated, and lipid substitution was found to significantly increase the cellular uptake and lead to effective gene knockdown with minimal cytotoxicities [12]. It was also found that not all the lipid modifications were beneficial and the performance of modified PEIs depended on the nature of the substituted lipids and the level of substitution [12]. Despite the experimental evidence for the beneficial effects of the lipid modifications, the molecular mechanism behind the beneficial effects is not clear, and remains to be probed.

Computer simulations especially molecular dynamics (MD) simulations are playing an increasingly important role in studying nucleic acid complexation with carriers. Recent MD simulation studies provided insight into the self-assembly process and structure-binding relationship of siRNA with dendrimers [13--20]. They demonstrated the validity of using MD simulations to study the interaction between siRNA and supramolecular carriers. For example, the radius of gyration of the dendrimer calculated from the simulation was consistent with the SAXS measured value [20]. However, these simulation studies only investigated the interaction of carriers with a single siRNA molecule, while in practical systems, the carriers are often interacting with multiple siRNA molecules to condense the siRNAs into nanoparticles. Simulations involving multiple siRNA molecules will provide a more realistic insight on the role of carriers in siRNA complexation process.

In order to elucidate the mechanisms of PEI mediated siRNA complexation and condensation, we performed a series of large scale all-atom MD simulations with four siRNA molecules. We specially focused on the role of lipid substitution in siRNA binding and condensation. A branched 2 kDa PEI was adopted as the native PEI and four PEIs modified with caprylic acid (CA) and linoleic acid (LA) at two substitution levels were adopted as the lipid modified PEIs. These conjugates were chosen due to the availability of extensive experimental data on their siRNA

delivery capabilities. Our study aims to shed light on the structure-function relationship by analyzing the structures of PEI/siRNA polyplexes and correlating them with experimental data.

6.2 Methods

6.2.1 Simulated Systems and Procedure

The siRNA simulated in this study has the following sequence: sense: 5'-CAGAAA GCUUAGUACCAAATT-3', antisense: 5'-UUUGUACUAAGCUUUCUGTC-3', which was used extensively to silence P-glycoprotein [9, 10, 12]. It is composed of 42 nucleotides carrying a total charge of -40 in the fully de-protonated state. The initial structure of the siRNA was built to be a canonical A form using the AMBER NAB tool [21]. The native PEI simulated is a branched PEI which consists of 43 amine groups (primary, secondary and tertiary) and has a molecular mass of 1874 Da [22]. Four lipid-modified PEIs were adopted and referred to as: P_1CA, PEI with one caprylic acid (CA); P_3CA, PEI with three CAs; P_1LA, PEI with one linoleic acid (LA); and P_3LA, PEI with three LAs. These levels of substitution are in the practical range where functional differences were observed from the native PEI in siRNA delivery [12]. To facilitate the discussion herein, we generally refer to both the native PEI and lipid modified PEIs as PEIs. The chemical structures and protonation sites of the five PEIs are shown in Figure 6.1. For the native PEI, 20 amine groups (marked by '+' and $\#1,\ \#2,\ \#3$ in Figure 6.1) were chosen to be protonated, corresponding to a protonation ratio of 47% as recently found experimentally at pH = 6 [23]. The protonation sites were assigned to only the primary and secondary amines, and were arranged as uniformly as possible to minimize thermodynamic interactions between the protonated amines. For P_1CA and P_1LA, the corresponding lipid is grafted on the native PEI at site $\#1$ as shown in Figure 6.1. Similarly, for P_3CA and P_3LA, three lipids are grafted on the native PEI at sites $\#1,\ \#2,\ \#3$ as shown in Figure 6.1. The initial structures of the PEIs were built in VMD [24] and then energetically minimized in NAMD [25] to obtain the initial PEI structures for the subsequent MD simultations. Five initial MD

Table 6.1: Information of the five single PEI systems simulated in this study.

System name	Charge of the PEI	No. of atoms	Simulation box size (Å^3)	No. of Na+/Cl-	Simulation time restrained+free (ns)
P2k	20	37160	$58 \times 69 \times 92$	0/20	$0.2 + 50$
P_1CA	19	20096	$58 \times 58 \times 58$	0/19	$0.2 + 40$
P_3CA	17	20123	$58 \times 58 \times 58$	0/17	$0.2 + 40$
P_1LA	19	20215	$58 \times 58 \times 58$	0/19	$0.2 + 40$
P_3LA	17	20042	$58 \times 58 \times 58$	0/17	$0.2 + 40$

simulations were first performed for the PEIs, each of which contained one PEI with explicit water and a number of Cl- ions to neutralize the systems (Table 6.1). The structure of each PEI at the end of the simulation was adopted as the initial configuration for PEIs in the simulations of PEI mediated siRNA complexation.

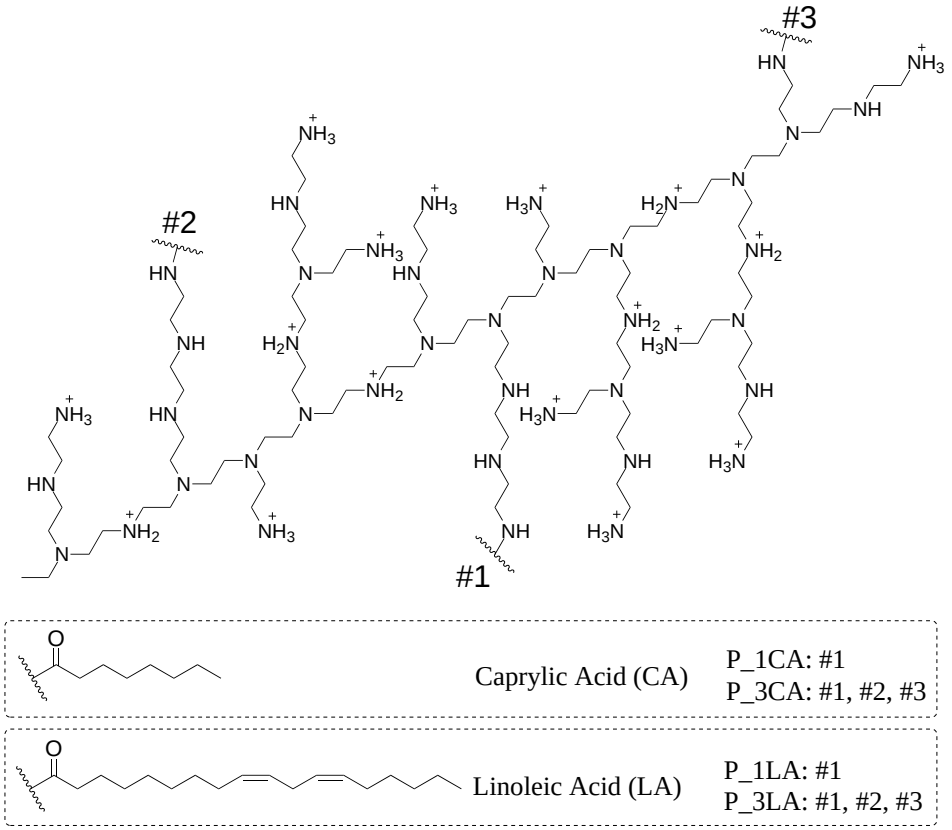

Figure 6.1: Molecular structure, protonation sites and lipid substitution sites of the five PEIs studied.

Five systems were then simulated to study the PEI mediated siRNA complexation, each of which contains 4 siRNAs and 18 PEIs corresponding to a PEI/siRNA N/P charge ratio of $\sim$2. For each system, appropriate amount of Na+ and Cl- ions

Table 6.2: Information of the five PEI/siRNA systems simulated in this study.

System name	Lipid no./type	Charge ratio siRNA/PEI	No. of atoms	Simulation box size ($Å^3$)	No. of Na+/Cl-	Time (ns) restr.+free
4R-18P	none	160/360	157456	$115 \times 115 \times 115$	136/336	$10 + 200$
4R-18P_1CA	1 CA	160/342	157210	$115 \times 115 \times 115$	136/318	$10 + 200$
4R-18P_3CA	3 CA	160/306	156817	$115 \times 115 \times 115$	136/282	$10 + 200$
4R-18P_1LA	1 LA	160/342	156787	$115 \times 115 \times 115$	136/318	$10 + 200$
4R-18P_3LA	3 LA	160/306	156496	$115 \times 115 \times 115$	136/282	$10 + 200$

were added to simulate the salt concentration of 154 mM at physiological levels. Detailed information of the five systems is summarized in Table 6.2. In this work, each system will be referred to by its name in the first column of Table 6.2. In constructing the initial configurations for each of the five systems, the axes of the four siRNAs were aligned to be parallel to one another and positioned on the four corners of a square with 35 Å side length. The principal axes of the PEIs were initially aligned parallel to the siRNA axes, and the center of mass (COM) of each PEI was positioned at 25 Å away from the axis of its neighboring siRNA(s). Detailed arrangement of the initial configurations is illustrated in Figure 6.2.

6.2.2 Simulation Details

A CHARMM format force field was developed and validated [26] for PEI based on the CHARMM General Force Field [27], and CHARMM 27 force field [28, 29] was used for all other molecules. All simulations were performed using the MD package NAMD [25]. TIP3P water model [30], periodic boundary condition, full electrostatics with particle-mesh Ewald method [31], cutoff distance 10 Å for van der Waals interactions and electrostatics pairwise calculations, SHAKE algorithm [32] to constrain all bonds containing hydrogens, and a time step of 2 fs were used for all the simulations.

For each system, the PEI/siRNA molecule(s) were first solvated into a cubic water box. Ions were then added into the water box by randomly replacing equivalent amount of water molecules using VMD [24]. During each simulation, the system was first minimized for 5000 steps, then heated from 0 K to 300 K in 20 ps with 10 $\mathrm{kcal}/(\mathrm{mol}\times Å^2)$ harmonic restraint on the non-hydrogen atoms of the solute. The restraint was kept on for 200 ps for the single PEI systems or 10 ns for the

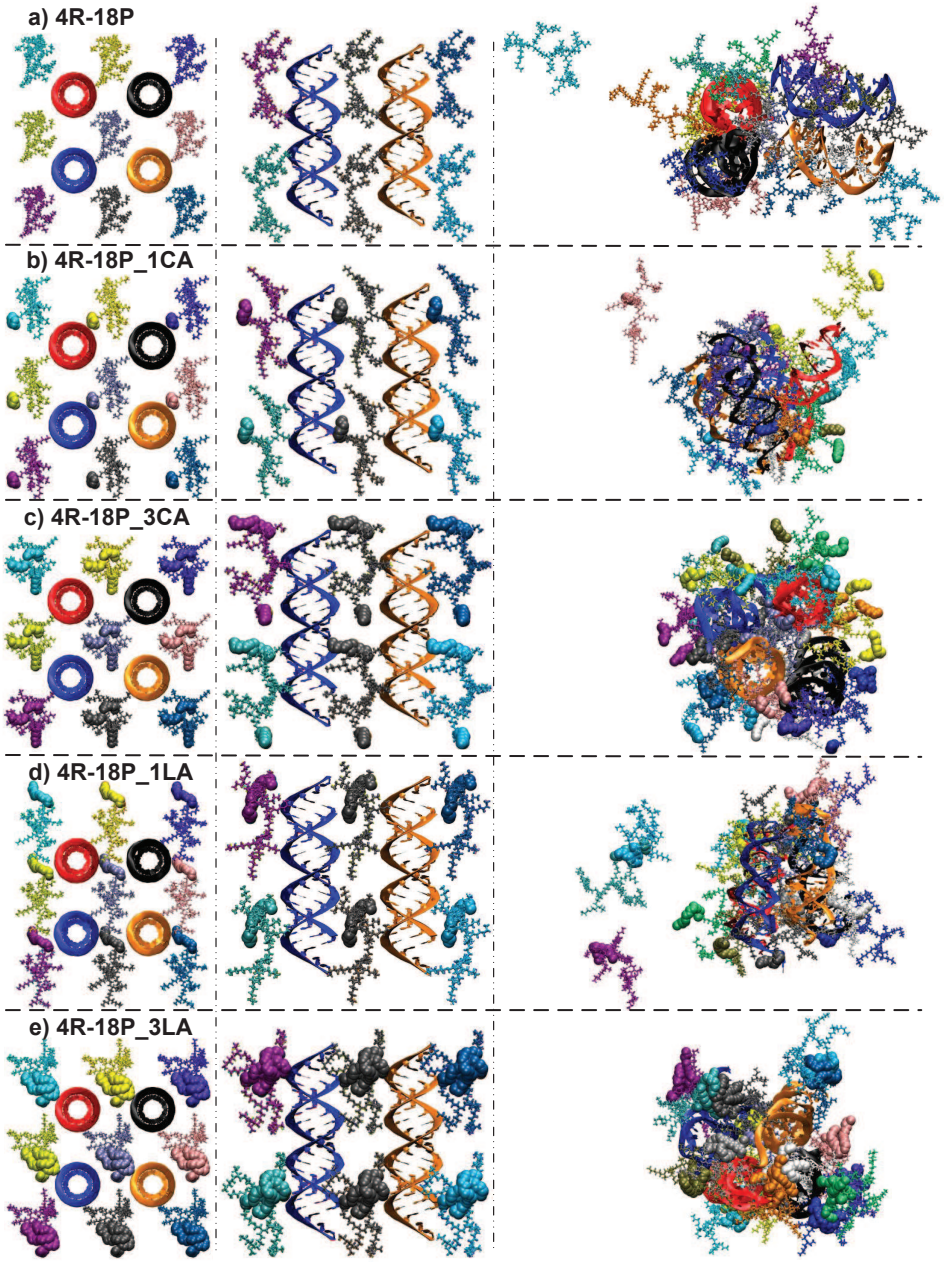

Figure 6.2: Initial (left panel -- axis view, central panel -- side view) and final configurations (right panel) of the five PEI/siRNA systems: (a) 4R-18P, (b) 4R-18P_1CA, (c) 4R-18P_3CA, (d) 4R-18P_1LA, (e) 4R-18P_3LA. Different PEIs and siRNAs are represented in different colors; the lipid moieties on the PEIs are represented by spheres; water and ions are removed for clarity.

PEI/siRNA complex systems at 300 K and 1 bar to relax the ions around the solutes. The restraint was then removed and NPT ensemble simulation was performed for 40-50 ns for the single PEI systems or 200 ns for the PEI/siRNA complex systems. VMD [24] was used for visualization and trajectories analysis.

6.3 Results and Discussion

6.3.1 Dynamics of siRNA Complexation

Figure 6.2 shows snapshots of the initial and final configurations for the five polyplexes, where the lipid moieties on the PEIs in Figure 6.2(b-d) are represented by spheres. For all five systems, the four siRNAs are centrally sequestered in the polyplexes formed with the PEIs at the end of the simulations. In systems 4R-18P, 4R-18P_1CA and 4R-18P_1LA, several PEIs are not attached to the formed polyplexes and exist freely in solution, unlike the polyplexes formed with PEIs containing higher lipid content, P_3CA and P_3LA.

To investigate the dynamics of PEIs binding during the complexation process, we plotted the numbers of PEIs bound to siRNAs as a function of time (Figure 6.3). Here we define a PEI to be bound to siRNAs if it has at least one N within 4 Å of any siRNA N/O atoms. For all five systems, the numbers of PEIs bound to the siRNAs rise quickly to $\sim$16 during the first 10 ns of the simulations. The curves display significant fluctuation from 10 ns to 50 ns, after which the fluctuations diminish to some degree. At the late stage of the simulations, the bound PEIs stabilize at around 16, 16, 18, 15, 16 for systems 4R-18P, 4R-18P_1CA, 4R-18P_3CA, 4R-18P_1LA and 4R-18P_3LA, respectively. Since a PEI carries a positive charge of 17-20 and the four siRNAs carry a negative charge of -160, all the five polyplexes formed are positively charged, which is consistent with the experimental observations [12]. Comparison between data in Figure 6.3 and the final configurations of the polyplexes suggest that some lipid modified PEI molecules involved in a polyplex are not directly bound to the siRNAs. These PEIs in fact attach to the polyplex through lipid association with other PEIs, which is an important mechanism for siRNA complexation by lipid modified PEIs and will be discussed in detail later.

Table 6.3: Number of water molecules released from the solute during the complexation process.

System	4R-18P	4R-18P_1CA	4R-18P_3CA	4R-18P_1LA	4R-18P_3LA
no. of waters	1577	1588	1910	1861	2054

Macromolecular association in aqueous environments usually results in the release of water molecules previously adhering to surfaces of the macromolecules. In our work, monitoring the number of the water molecules on the PEIs/siRNAs surface allows us to monitor the complexation process and to gauge whether dynamic equilibrium has been reached. Since the water molecules on the macromolecular surfaces are less mobile than those in the free bulk state, such release is an entropically favorable process with a free energy reduction of up to 2 kcal/mol at 300 K [33]. Therefore, the number of water molecules released during the macromolecule association allows us to assess the entropic gain upon macromolecular binding. Figure 6.4 shows the numbers of water molecules in the hydration shell of the PEIs/siRNAs (within 3 Å from the molecules) as a function of simulation time. For all five systems, the number of water molecules in the hydration shell drop quickly during the first 50 ns, indicating that as the polyplexes form, a large number of water molecules were displaced from the PEI and siRNA surfaces into the bulk during this period. The curves start to level off after 100 ns, indicating dynamic equilibriums have been reached. Table 6.3 summarizes the number of water molecules released from the hydration shell of the solutes at the end of the simulation time. Detailed information on this calculation is given in Appendix E. The released water molecules are higher for the systems with lipid-modified PEIs than for system 4R-18P with native PEIs. With more lipid substitution on PEIs, more water molecules are released: 4R-18P_1CA (1588) vs. 4R-18P_3CA (1910) and 4R-18P_1LA (1861) vs. 4R-18P_3LA (2054). The longer alkyl chain in P_1LA and P_3LA leads to more water molecules being released, as compared with the systems with the same level of CA substitution. The lipids are expected to associate in aqueous solution to reduce their solvent accessible surface. The larger amount of water release for systems with lipid modified PEIs is likely to reflect water released from the lipid association. We will address the lipid association later in this work.

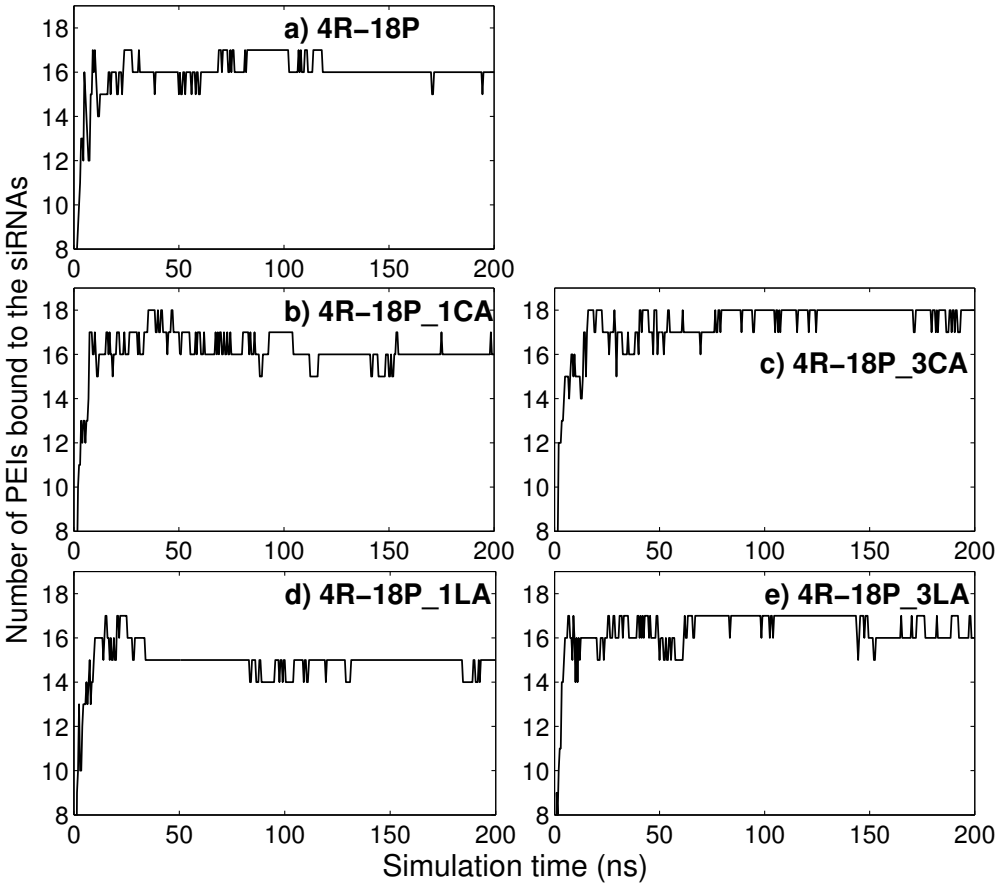

Figure 6.3: Number of PEI molecules bound to the siRNAs as a function of simulation time. (a) 4R-18P, (b) 4R-18P_1CA, (c) 4R-18P_3CA, (d) 4R-18P_1LA, (e) 4R-18P_3LA.

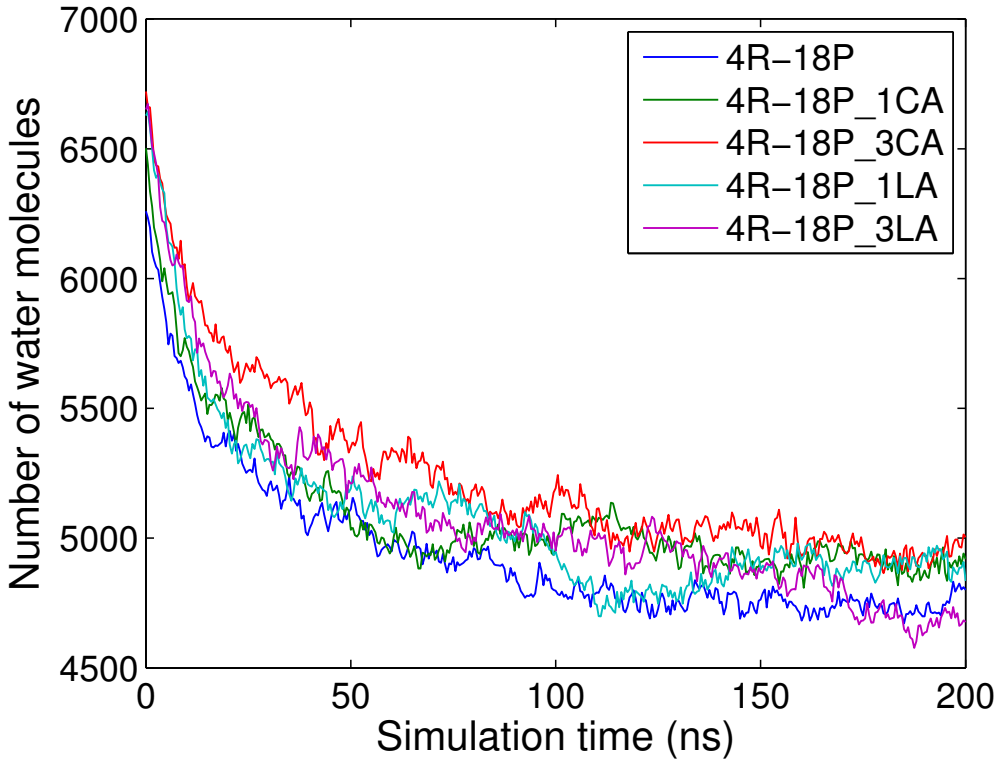

Figure 6.4: Number of water molecules in the hydration shell (within 3 Å of the siRNAs and PEIs) as a function of simulation time.

6.3.2 Polyion Bridging and siRNA Charge Neutralization

We previously identified two main mechanisms for native PEI mediated DNA complexation [34]: polyion bridging (i.e., a polycation spanning across multiple DNA segments simultaneously; see detailed definition in Appendix E) and DNA charge neutralization. As expected, we found that polyion bridging also plays an important role in PEI mediated siRNA complexation. Specially, for systems 4R-18P, 4R-18P_1CA, 4R-18P_3CA, 4R-18P_1LA and 4R-18P_3LA, 7, 7, 8, 9, 8 PEIs, respectively, participated in bridging two or three siRNAs for longer than 50% of the simulation time (see Figures E.1-E.5 in Appendix E). These close numbers indicate that lipid substitution does not affect the polyion bridging behavior of the PEIs.

To investigate how the five different PEIs neutralize the siRNA charges, we plotted the cumulative distributions, with respect to the siRNA C1' atoms, of PEI N+, Na+, Cl- and the total charge of PEI/salt ions, averaged over the last 80 ns of the simulations (Figure 6.5). The C1' atoms are on the sugar rings of the siRNAs, located inside the siRNA helix at a distance of $\sim$5 Å from the surface of siRNA defined by the phosphorus atoms. In each subfigure of Figure 6.5, the dashed black line indicates the -160 charge of the four siRNAs, and the blue solid curve is the total charge of PEI and salt ions within given distance to their nearest siRNA C1' atoms. The four siRNA are 100% neutralized by the PEIs and salt ions at the distance where the black dashed line and blue curve intersect. It can be seen that the curves for the five systems share a similar characteristic. For all the systems, the PEIs/ions neutralize the siRNA at a distance of $\sim$8 Å from the siRNA C1' atoms. Within 8 Å, it can be seen that the distribution of Cl- and Na+ ions is minimal and almost identical, and the curves for the net charge of PEI/ions basically overlaps with the curves for the PEI charge. Therefore, we conclude that the PEIs contribute dominantly in neutralizing the siRNA within the 8 Å of the siRNA C1' atoms. Being capable of neutralizing the siRNAs at such a short distance is a major mechanism for PEI mediated siRNA complexation. At distances beyond the intersection of the black line and the blue curve, the PEI and ion charges exceed the siRNA charges, *i.e.*, the siRNAs are over-neutralized at such distances. The five systems have a similar extent of maximum over-neutralization of $\sim$25 charges at $\sim$18 Å

from the siRNA C1' atoms. The similar neutralizing characteristic for the five PEIs demonstrates that the lipid substitution on PEIs does not affect their capability of neutralizing the siRNA molecules. In Figure E.6 of Appendix E, we plotted the charge neutralization curves based on four different time windows at the late stage of the simulations as an evidence for convergence of the simulation trajectories.

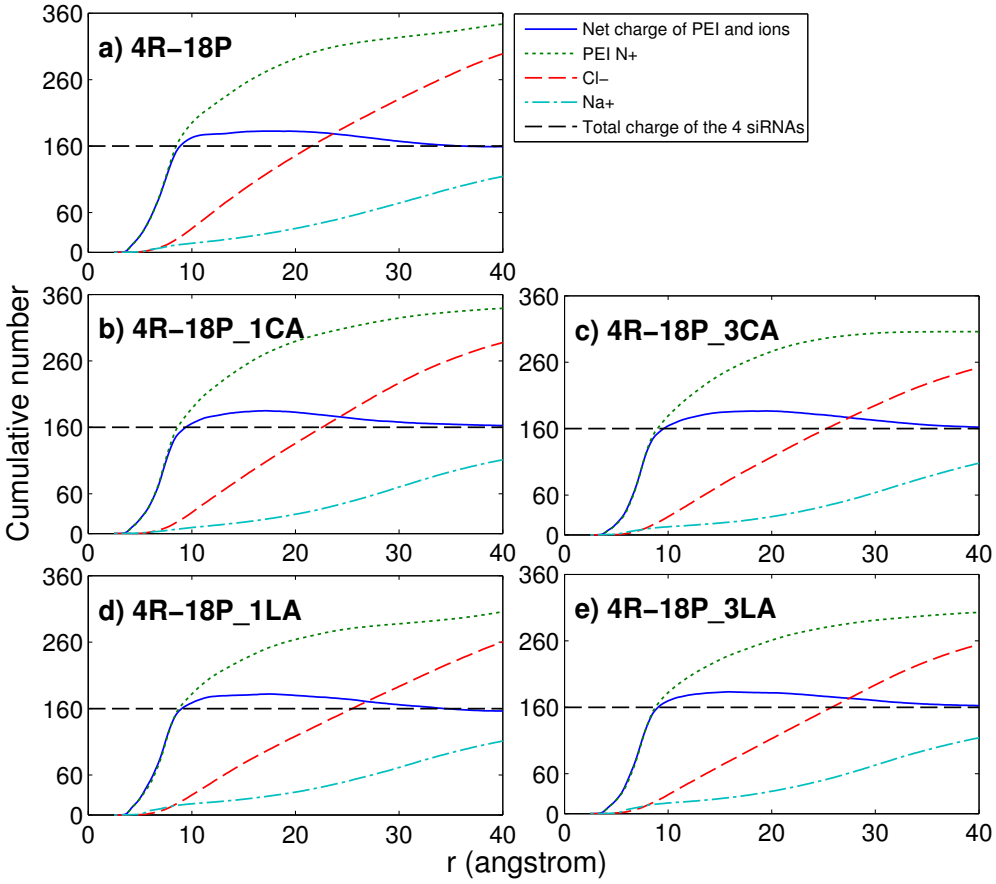

Figure 6.5: Cumulative numbers of protonated PEI Ns, Na+, Cl-, and net charge of PEI/Na+/Cl- as a function of the distance from any RNA C1' atom (averaged over the last 80 ns of each simulation). The total charge of the four siRNAs is plotted by a straight dashed black lines as reference in each subfigure. (a) 4R-18P, (b) 4R-18P_1CA, (c) 4R-18P_3CA, (d) 4R-18P_1LA, (e) 4R-18P_3LA.

6.3.3 Lipid Association

Visual examination of the final configurations in Figure 6.2(b-e) shows that some lipid tails from different PEIs are associated at the end of the simulation. The much larger amount of water release with lipid modified PEIs was also indicative

of significant association among the lipids. To quantify the lipid association among different PEIs, in Figure 6.6, we tabulated the number of pairs of lipid carbons that are closer than 5 Å between each pair of PEIs, averaged over the last 40 ns of the simulations. The numbers on the top and right of each subfigure are the PEI indices. The 18 PEIs in each system results in 153 possible pairing of PEIs, corresponding to 153 cells in each subfigure of Figure 6.6. A CA lipid has 8 carbons so that one pair of P_1CA would have 64 pairing of carbons. Similarly, for 4R-18P_3CA, 4R-18P_-1LA and 4R-18P_3LA, all possible pairing of carbons would be 576, 324 and 2916, respectively. Among these possible carbon pairs, the number of pairs within 5 Å are counted and given in the cell corresponding to this pair of PEIs. The cells are left empty where no carbon pair was found to be within 5 Å. For examples, number 9 on the top left of Figure 6.6(a) means that out of the 64 pairs of carbons between PEI 1 and PEI 3 in system 4D-18R_1CA, 9 pairs are separated by 5 Å or less; number 169 on the bottom of Figure 6.6(d) means that out of the 2916 pairs of carbons between PEI 15 and PEI 16 in system 4D-18R_3LA, 169 pairs are separated by 5 Å or less. We choose 5 Å as the criterion because this is the closest carbon-carbon distance within which the free energy for the association of two alkane molecules is negative, indicating that their association is energetically favorable [35]. The situation of at least one pair of carbon being closer than 5 Å apart is considered to represent linked PEI molecules through lipid association. Only two pairs of PEIs are linked in system 4D-18R_1CA (Figure 6.6(a)). For system 4D-18R_3CA with increased level of substitution, the intensity of lipid association is dramatically increased, with 11 pairs of PEIs linked (Figure 6.6(b)). Between each of these 11 pairs of PEIs, 5.7 pairs of lipid carbons (on average) are closer than 5 Å apart. Systems 4D-18R_-1LA and 4D-18R_3LA have 6 and 8 pairs of PEIs linked, respectively, with the average numbers of lipid carbon pairs between each pair of linked PEIs being 35 and 107. There are three PEIs mutually linked in systems 4D-18R_1LA and 4D-18R_3LA, which are marked by gray cells in Figure 6.6(c, d). When the level of lipid substitution increases, the number of linked PEIs through lipid association do not increase as dramatically for LA substituted PEIs as that for CA substituted PEIs, however, the average number of lipid carbon pairs between each pair of linked PEIs

increases by approximately three times for the longer lipid. For systems 4D-18R_-1CA and 4D-18R_1LA with one lipid substitution per PEI, with longer lipid chain, the intensity of lipid association increased dramatically in terms of both number of linked PEIs and average number of lipid carbon pairs between each pair of linked PEIs. For systems 4D-18R_3CA and 4D-18R_3LA with three lipid substitution per PEI, with longer lipid chain, the number of linked PEIs actually decrease from 11 to 8, however, the average number of lipid carbon pairs between each pair of linked PEIs is almost 19 times as high for 4D-18R_3LA.

To investigate how the lipid associations evolve during the simulations, we plotted the number of lipid carbon pairs that are closer than 5 Å apart between certain pairs of PEIs as a function of simulation time in Figure 6.7. Only those pairs of PEIs that have lipid association during the last 40 ns as marked in Figure 6.6 are considered. For 4D-18R_1CA (Figure 6.7(a)), the curves undergo rapid fluctuation between 0 and 20, indicating that the lipid associations between the two pairs of PEIs are highly unstable. Increasing the lipid substitution level from 1 CA to 3 CA per PEI does not change the unstable nature of the lipid associations for the 11 pairs of PEIs in 4D-18R_3CA (Figure 6.7(b)). The associations fluctuate rapidly between 0 and 40, for example, the lipid association between PEIs 11 and 12 breaks for $\sim$20 times during the 200 ns simulation time. For 4D-18R_1LA with long LA lipid (Figure 6.7(c)), both the magnitude of fluctuation and the frequency of breakage of lipid associations are reduced. Five lipid associations (PEIs 1-5, 6-8, 6-14, 8-14, 11-15) out of the six still break for several times during the simulations. Unlike for CA, increasing the LA substitution level stabilizes the lipid association (Figure 6.7(d)), evidenced by that the eight lipid associations never break apart once they are formed in the simulations.

For PEIs with 3 CA or 3 LA substituents, the three lipids within one PEI could associate and this can make the PEI molecule more rigid. Table 6.4 tabulates the lipid association status among the three lipids on individual PEIs in systems 4D-18R_3CA and 4D-18R_3LA. If none of the three lipids on a PEI is associated, it is marked with ‘N’; if the three lipids are mutually associated, it is marked with ‘A’; otherwise, the three lipids are partially associated and marked with ‘P’. For all

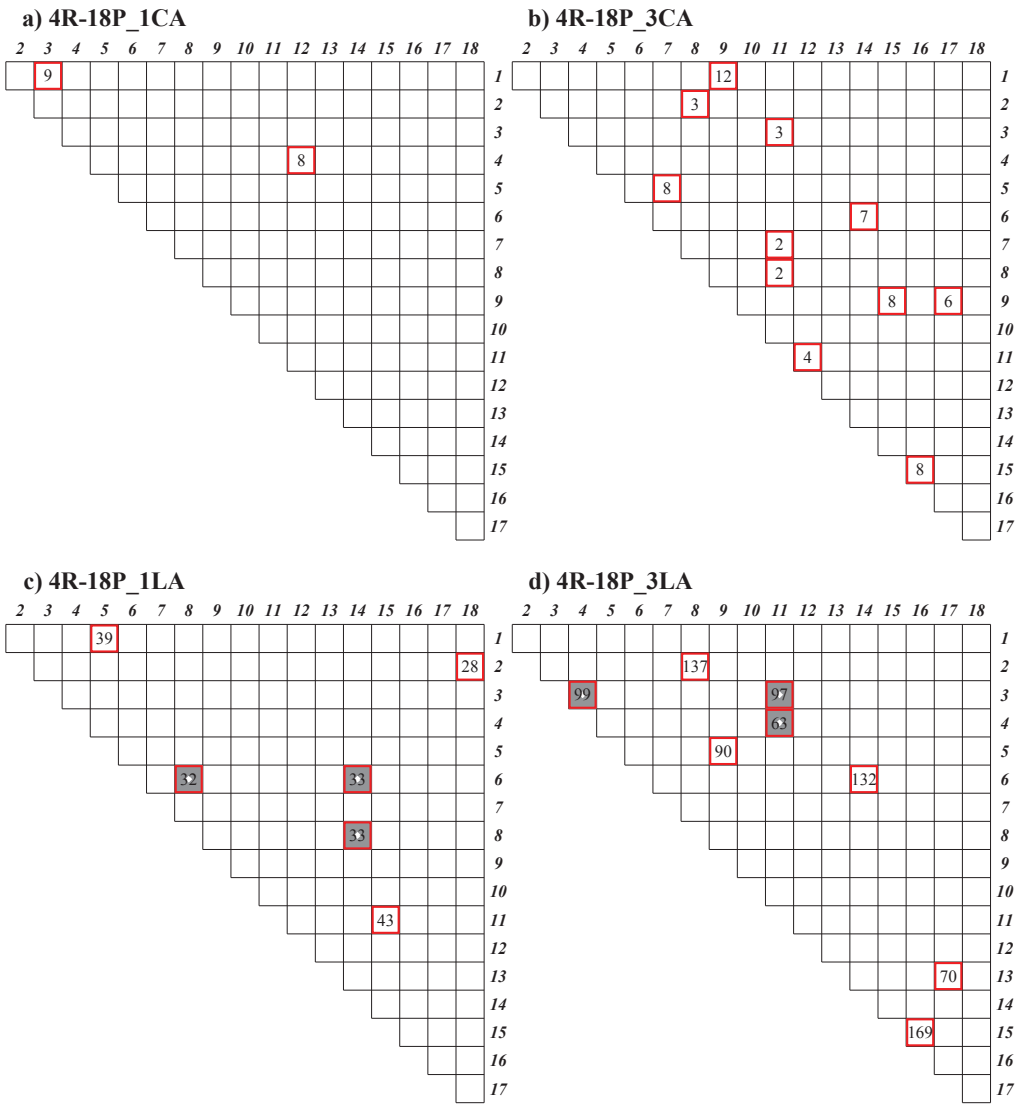

Figure 6.6: Number of pairs of lipid carbons that are closer than 5 Å apart between each pair of PEIs, averaged over the last 40 ns of the simulations in (a) 4R-18P_-1CA, (b) 4R-18P_3CA, (c) 4R-18P_1LA, (d) 4R-18P_3LA. The numbers on the top and right of each subfigure are PEI indices. Only the non-zero numbers are shown and marked with red squares. The PEIs involved in the same association are marked by gray cells in (c, d).

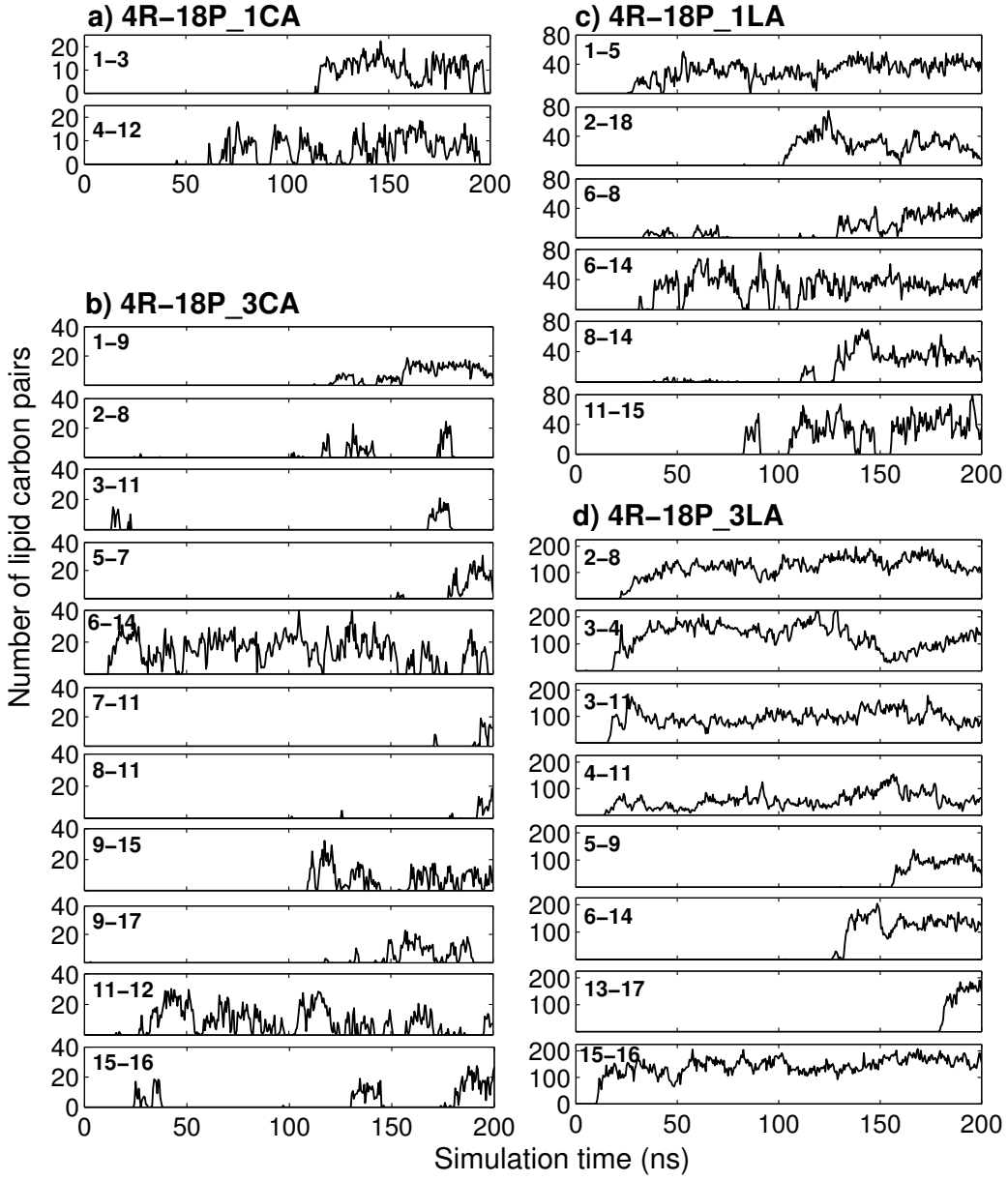

Figure 6.7: Dynamics of lipid association: Number of pairs of lipid carbons that are closer than 5 Å between certain pairs of PEIs as a function of simulation time in (a) 4R-18P_1CA, (b) 4R-18P_3CA, (c) 4R-18P_1LA, (d) 4R-18P_3LA. The two numbers in each subfigure are the indices of the pair of PEIs.

Table 6.4: Lipid association among the three lipids on individual PEIs in systems 4R-18P_3CA and 4R-18P_3LA. (N--none; P--partially; A--all.)

PEI index	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
4R-18P_3CA	N	P	N	P	P	P	P	N	N	P	N	P	A	P	N	P	P	A
4R-18P_3LA	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A

the 18 PEIs in 4D-18R_3LA, the three lipids on each PEI are mutually associated, while for 4D-18R_3CA only two PEIs have their three lipids mutually associated. Ten PEIs in 4D-18R_3CA have the three lipids partially associated and six PEIs have no self lipid association. The universal lipid association on individual P_3LAs in system 4D-18R_3LA could cause the P_3LAs possess more rigidity and hence manifest a mere profound steric effect when binding to siRNA. Such effect can be seen from the compactness of the siRNAs as demonstrated below.

6.3.4 Compactness of the siRNAs

To gauge the compactness of the siRNA molecules in the polyplexes and their stability, we plotted the radii of gyration R_g of the four siRNA molecules as a function of simulation time in each system (Figure 6.8). For a given polyplex, a higher value of R_g is an indicator of more loosely arranged siRNA molecules in the polyplex. For all the five systems, the R_g curves follow a generally decreasing trend during the first 50 ns of the simulations, indicating the formation of the siRNA polyplexes while the siRNA molecules are moving closer. The R_g values fluctuate from 50 ns to 200 ns. During the last 80 ns of the simulations, the average R_g of the four siRNAs are 31.0 Å, 29.1 Å, 27.9 Å, 27.0 Å and 28.0 Å, and the variance of the R_g are 0.19 Å, 0.08 Å, 0.05 Å, 0.05 Å and 0.11 Å for systems 4R-18P, 4R-18P_1CA, 4R-18P_3CA, 4R-18P_1LA and 4R-18P_3LA, respectively. The 4R-18P has the most loose and least stable siRNA structure among the five polyplexes indicated by the largest R_g and largest variance. This indicates that the lipid modification compacts and stabilizes the siRNA polyplexes. The siRNAs in system 4R-18P_1CA have a larger and more fluctuating R_g compared to the siRNAs in 4R-18P_3CA. This is expected as the more lipid substitution per PEI, the more profoundly the effects of the lipid substitution manifest. However, the siRNAs in system 4R-18P_3LA have a larger R_g compared to the siRNAs in 4R-18P_1LA with less lipid substitution. It

seems therefore that higher lipid substitution level does not necessarily lead to more compact siRNA structure. This can be explained by the fact that the association of the three LA lipids on individual PEIs generates a significant steric hindrance for condensing the siRNAs.

6.3.5 Location of Cationic and Lipophilic Moieties of PEIs

Although the size and charge of carrier/siRNA polyplexes are routinely assessed in experiments, the structural details especially the location of the lipophilic moieties in the polyplexes have not been investigated experimentally [8]. To assess the location of cationic and lipophilic moieties of the PEIs relative to the siRNAs, in Figure 6.9 we plotted the cumulative percentage (left panel) and radius distribution function (right panel) of the PEI Ns and lipid carbons as a function of distance from any siRNA N/O atom, averaged over the last 80 ns of the simulation. Taking the PEI Ns for example, the cumulative percentage at a given distance r is the percentage of all the PEI Ns within a distance r from any siRNA N/O atoms. For all the five systems, the cumulative percentage curves for the PEI Ns rise quickly from 0% at 2.5 Å to $\sim$40% at 5 Å, and to $\sim$80% at 15 Å. The curves for lipid carbons rise from 3 Å (0%) following a similar trend as the curves for PEI Ns, but the lipid carbon curves generally right shift by 1--2 Å of the PEI Ns curves. This clearly shows that the lipophilic moieties are located further away from the siRNAs than the cationic moieties. The corresponding subfigures in the right panel of Figure 6.9 show the detailed distribution of the PEI Ns and lipid carbons with respect to the siRNA N/O atoms. The five systems have a similar pattern of PEI N distribution. There are two predominant peaks: one at $\sim$2.5 Å and one at $\sim$4.5 Å. The first peak corresponds to the expected distance for direct contact between the PEI amine groups and the siRNA N/O atoms through hydrogen bonding. The second peak corresponds to the distance for indirect interactions, such as hydrogen bonding mediated by one water molecule. Each of the two peaks corresponds to $\sim$20% of the total PEI Ns. The similar pattern of PEI N distribution among the five systems demonstrates that lipid modification does not seem to affect the interaction between the siRNAs and the cationic moieties of PEIs, which is also consistent with the similar neutralization

capability the five different PEIs demonstrate. The lipid carbons have one predominant peak at ∼4 Å, and this peak accounts for ∼40% of the total lipid carbons for systems 4R-18P_1CA and 4R-18P_1LA, ∼35% carbons for system 4R-18P_3CA and ∼25% carbons for system 4R-18P_3LA. The less amount of lipid carbons at the vicinity of siRNAs in systems 4R-18P_3CA and 4R-18P_3LA is probably due to the lipid association among the three lipids on individual PEIs as discussed in the previous subsection, which makes the lipid tails more rigid and hence harder to comply with the siRNAs. For systems 4R-18Pi_1LA and 4R-18P_3LA with long LA lipids, there is a second peak for lipid carbons at ∼8 Å, which is more pronounced for system 4R-18P_3LA. This can be explained by the fact that the long LA lipids are more probable to have their lipid carbons located further from the siRNAs due to their physical length. The lipid association among the three lipids on individual PEIs for system 4R-18P_3LA contributes to its more pronounced second peak because such association makes the lipid moieties harder to comply with the siRNAs as discussed above.

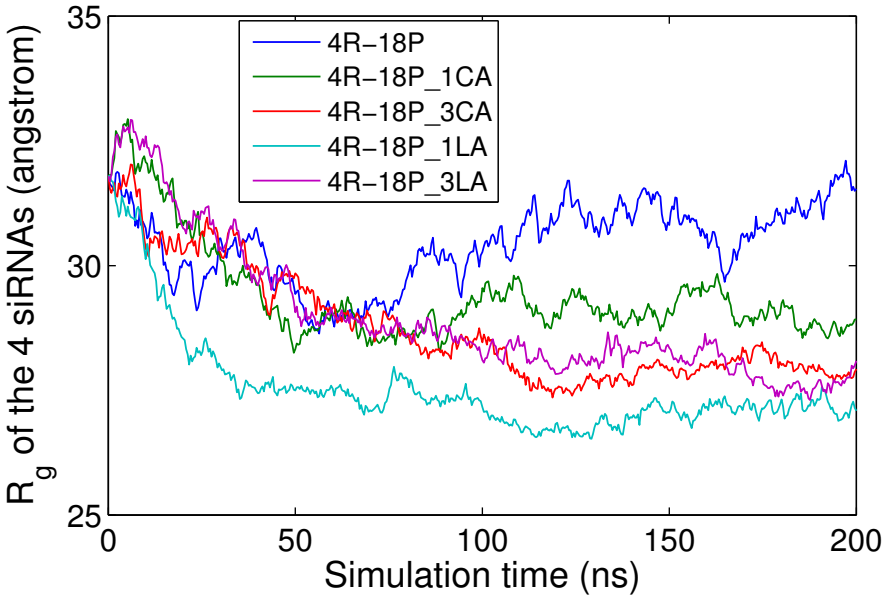

Figure 6.8: Radius of gyration of the four siRNAs in each system as a function of simulation time.

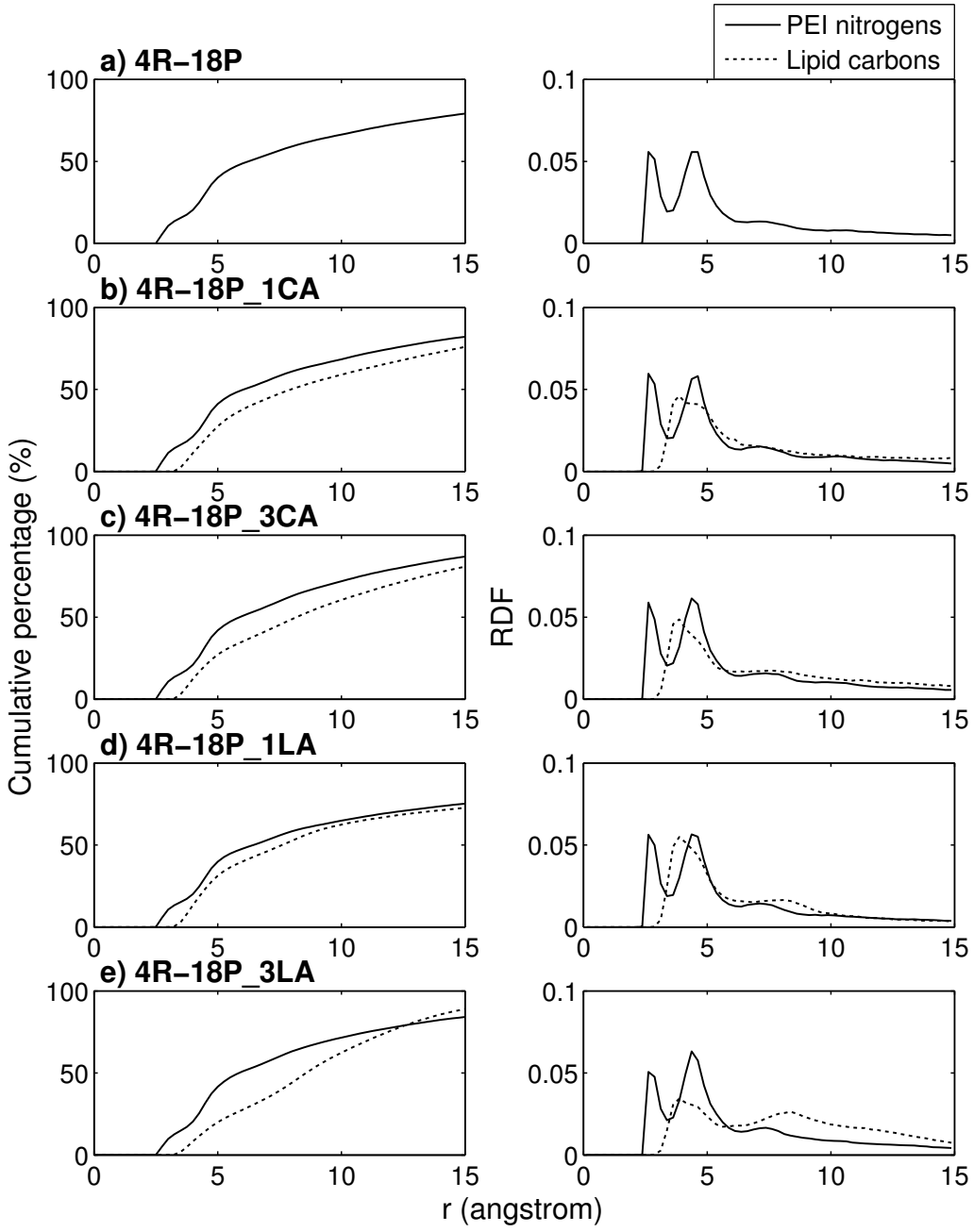

Figure 6.9: Cumulative percentage (left panel) and radial distribution function (right panel) of the PEI Ns and lipid carbons as functions of the distance from the siRNA N/O atom in (a) 4R-18P, (b) 4R-18P_1CA, (c) 4R-18P_3CA, (d) 4R-18P_-1LA, (e) 4R-18P_3LA.

6.3.6 Implications

Molecular details of siRNA polyplexes derived from MD simulations are expected to be correlated to molecular features of siRNA complexes, as well as biologically relevant performance, such as cellular uptake, intracellular trafficking and functional silencing. Although one ultimately wishes to seek correlations with functional silencing effects, it is premature to undertake this at this stage since polyplex dissociation needs to take place for silencing and details of polyplex dissociation and interactions with endogenous solutes (such as the siRNA target mRNA) have not been attempted before with MD simulations. However, experimental studies on cellular uptake of polyplexes have been reported by our group prior to the MD simulations. Cellular uptake is a relatively simple, single-step process that should be more amenable for correlations with MD simulations (as compared to silencing).

The cellular uptake of siRNA polyplexes formed with CA- and LA-substituted PEIs is summarized in Figure 6.10 [12, 36, 37]. Our experimental observations are derived from 3 different cell lines and we employed polymer:siRNA weight ratios of 2:1 or 8:1 in these studies (corresponding to (10-14):1 and (40-56):1 molar ratios, respectively, depending on the level and nature of lipid substitution). The simulations in this study were obtained by using a molar ratio of 4.5:1. First clear observation in all experimental uptake studies was the improved cellular delivery of siRNA with lipid substitutions as compared to native PEI. More stable nature of the polyplexes formed by lipid-substituted PEIs (given by R_g in this study) is confirmatory of general observations on the correlation between the complex stability and cellular uptake [38, 39]. The presence of lipid moiety, which enhances the compatibility of polyplexes with lipid membrane, cannot be ruled out in this context. The siRNA uptake with CA-substituted PEIs is consistent in all three cell types employed, kidney tubule MDCK cells, breast cancer MDA-435 cells and breast cancer MDA-231 cells [12, 36, 37]; the uptake was correlated to CA substitution level under all conditions. MD simulation indicated low degree of lipid interactions with CA at low substitution and improving interactions with increasing substitution levels, resulting in higher stability of the polyplex (smaller fluctuation in R_g). Hence, increased stability of polyplexes with increasing CA levels revealed with MD sim-

ulations was consistent with the experimental uptake results. The siRNA uptake with the LA-substituted PEIs, on the other hand, was more complicated. At low substitution levels (e.g., 1 LA per PEI), LA substitution was more effective than CA for facilitating siRNA uptake. The MD simulations also indicated better interaction among LA lipids at such low substitution levels. Higher LA substitutions (e.g., 3 LAs per PEI) did not always lead to higher uptake in experimental studies, supportive of MD simulation results that indicated low level of substitutions to be sufficient to induce lipid-lipid interactions in the case of this longer lipid. The rigid PEI structure obtained with higher level of LA substitution could inhibit further siRNA binding, an observation noted on PAMAM dendrimers by Paven *et al.* [15]. The R_g values obtained for LA substituted PEI was not indicative of a correlation between the substitution level and the polyplex stability. The overall cellular uptake studies with this substituent also did not indicate a clear correlation between the uptake and substitution level. In that sense, the lack of strong correlations between LA substitution level and cellular uptake of siRNA was reflective of the effect of LA on calculated R_g values in this study.

One experimental observation not reproduced in this study is the higher charge (ζ-potential) of the polyplexes formed with lipid-substituted PEIs, as compared to polyplexes with native PEI. Better assembly of the polyplexes was evident with higher lipid substitution (see Figure 6.2), which would have led to higher charges if all PEI molecules were equally charged. A limited numbers of PEIs were used in our simulations whereas the experimental studies usually employ higher PEI:siRNA ratios (as indicated above with the molar ratios). It is possible that more lipid-substituted PEIs could assemble to the polyplex with 4 siRNA and MD simulations might reveal such an effect when polyplexes are simulated with higher PEI:siRNA ratios. Secondly, the charges on PEI are considered fixed in this study, whereas a dynamic protonation state might change the overall charge of the polyplexes.

The peripheral lipids on the polyplexes can better protect the siRNAs from degrading nucleases, and also facilitate the interaction of the polyplexes with cell membranes and other hydrophobic biological entities on the delivery path. This lipid distribution and the stable lipid association are expected to be beneficial for in-

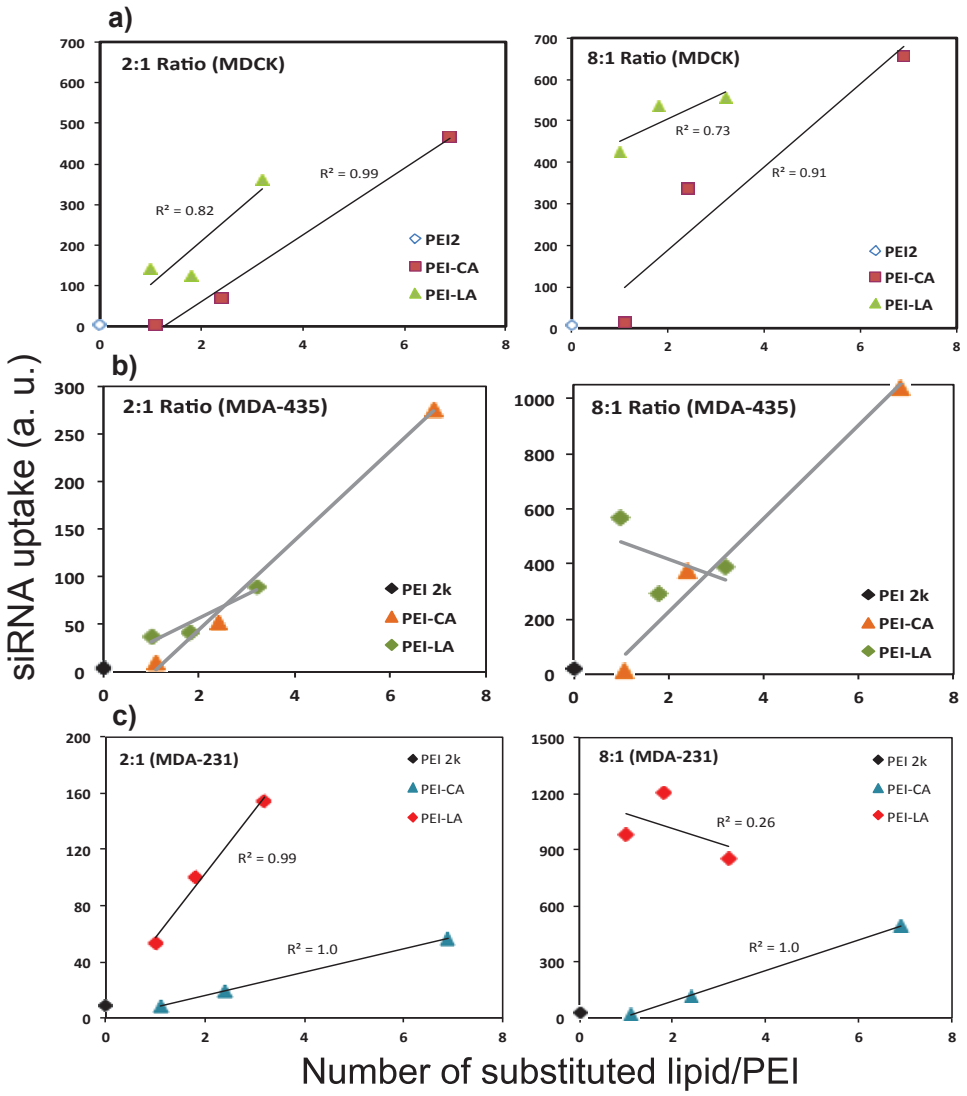

Figure 6.10: The siRNA delivery into (a) MDCK, (b) MDA-435 and (c) MDA-231 cells with native and CA and LA substituted PEIs. The uptake was quantitated with FAM-labeled siRNAs and expressed in arbitrary fluorescence units. Experimental details on siRNA for MDCK, MDA-435 and MDA-231 cells can be found in References 12, 36, 37.

ternalization of the polyplexes through cell membranes, supporting the experimentally observed higher cellular uptake obtained with lipid modified carriers [9, 12]. On the other hand, the peripheral location of lipids also provides an opportunity for siRNA polyplexes towards an undesirable aggregation state via facilitated hydrophobic interactions. It will be useful to design systems where the lipid moieties are exposed to periphery only after contact with cell surfaces, so that cell penetration is enhanced while aggregation in solution is suppressed. Also, the stable lipid association is undesired for the siRNA unloading at the target site. These opposing effects should be considered in rational design of lipid-modified polymer based carriers.

Our simulations can be adopted for investigating the complexation of siRNA molecules with other polycationic carriers. Through interpreting the structural properties of a siRNA polylex formed with a designed polycation, we can evaluate the siRNA complexation capability of this polycation from MD simulation. Future MD simulations can also be conducted to assess the performance of the polycation in releasing the delivered siRNA, thus helping to better interpret the experimental results on silencing efficiency and eventually helping screen candidate design schemes.

6.4 Conclusions

We performed a series of all-atom MD simulations to study siRNA complexation mediated by native and lipid-modified PEIs. We found that the lipid modification does not affect PEI's capability to neutralize the siRNAs. All five PEIs used in the simulations can completely neutralize the siRNAs at a distance of $\sim$8 Å from the siRNA C1' atoms. Polyion bridging plays an important role in siRNA complexation, which is not affected by the substituted lipids. The lipophilic moieties are located further away from the siRNAs compared to the cationic moieties. The lipid associations between short lipids (CA) form and break frequently for one and three CA substituted PEI. The lipid associations between long lipids (LA) are more stable, where the lipid associations never break once they form during the simulation for three LA substituted PEI. The results also revealed that siRNA structures medi-

ated by lipid modified PEIs are more compact and stable. For PEIs modified with short lipids (CA), increasing the lipid substitution level from one to three lipids per PEI makes the effects of lipid modification manifest more dramatically, resulting in more compact and stable siRNA structure. For PEIs modified with long lipids (LA), increasing the lipid substitution from one to three lipids per PEI does not change the amount of PEI linkage via lipid association much, and it has a reverse effect on compacting siRNA structure due to increased steric hindrance brought by the lipid association among the three lipids on individual PEIs.

Bibliography

[1] S. Elbashir, J. Harborth, W. Lendeckel, A. Yalcin, K. Weber, T. Tuschl, *Nature* **2001**, *411*, 494--498.

[2] D. Castanotto, J. J. Rossi, *Nature* **2009**, *457*, 426--433.

[3] O. Boussif, F. Lezoualc'h, M. A. Zanta, M. D. Mergny, D. Scherman, B. Demeneix, J. P. Behr, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 7297--7301.

[4] W. Godbey, K. Wu, A. Mikos, *J. Control. Release* **1999**, *60*, 149--160.

[5] H. M. Aliabadi, B. Landry, C. Sun, T. Tang, H. Uludag, *Biomaterials* **2012**, *33*, 2546--2569.

[6] F. Ledley, *Hum. Gene Ther.* **1995**, *6*, 1129--1144.

[7] D. Luo, W. Saltzman, *Nat. Biotechnol.* **2000**, *18*, 33--37.

[8] V. Incani, A. Lavasanifar, H. Uludag, *Soft Matter* **2010**, *6*, 2124--2138.

[9] M. Abbasi, H. M. Aliabadi, E. H. Moase, A. Lavasanifar, K. Kaur, R. Lai, C. Doillon, H. Uludag, *Pharm. Res.* **2011**, *28*, 2516--2529.

[10] M. Abbasi, A. Lavasanifar, L. G. Berthiaume, M. Weinfeld, H. Uludag, *Cancer* **2010**, *116*, 5544--5554.

[11] W. J. Kim, C.-W. Chang, M. Lee, S. W. Kim, *J. Control. Release* **2007**, *118*, 357--363.

[12] H. M. Aliabadi, B. Landry, R. K. Bahadur, A. Neamnark, O. Suwantong, H. Uludag, *Macromol. Biosci.* **2011**, *11*, 662--672.

[13] D. Ouyang, H. Zhang, D.-P. Herten, H. S. Parekh, S. C. Smith, *J. Phys. Chem. B* **2010**, *114*, 9220--9230.

[14] D. Ouyang, H. Zhang, H. S. Parekh, S. C. Smith, *J. Phys. Chem. B* **2010**, *114*, 9231--9237.

[15] G. M. Pavan, L. Albertazzi, A. Danani, *J. Phys. Chem. B* **2010**, *114*, 2667--2675.

[16] G. M. Pavan, P. Posocco, A. Tagliabue, M. Maly, A. Malek, A. Danani, E. Ragg, C. V. Catapano, S. Pricl, *Chem --Eur J* **2010**, *16*, 7781--7795.

[17] G. M. Pavan, M. A. Mintzer, E. E. Simanek, O. M. Merkel, T. Kissel, A. Danani, *Biomacromolecules* **2010**, *11*, 721--730.

[18] G. M. Pavan, M. A. Kostiainen, A. Danani, *J. Phys. Chem. B* **2010**, *114*, 5686--5693.

[19] V. Vasumathi, P. K. Maiti, *Macromolecules* **2010**, *43*, 8264--8274.

[20] L. B. Jensen, K. Mortensen, G. M. Pavan, M. R. Kasimova, D. K. Jensen, V. Gadzhyeva, H. M. Nielsen, C. Foged, *Biomacromolecules* **2010**, *11*, 3571--3577.

[21] D. Case, T. Darden, I. T.E. Cheatham, C. Simmerling, J. Wang, R. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. Merz, B.Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvary, K.F.Wong, F. Paesani, J. Vanicek, X. Wu, S. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. Mathews, M. Seetin, C. Sagui, V. Babin, P. Kollman, *AMBER 10*, University of California, San Francisco, **2008**.

[22] C. Sun, T. Tang, H. Uludag, *J. Phys. Chem. B* **2012**, *116*, 2405--2413.

[23] K. Utsuno, H. Uludag, *Biophys. J.* **2010**, *99*, 201--207.

[24] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33--38.

[25] J. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. Skeel, L. Kale, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781--1802.

[26] C. Sun, T. Tang, H. Uludag, J. E. Cuervo, *Biophys. J.* **2011**, *100*, 2754--2763.

[27] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. MacKerell, Jr., *J. Comput. Chem.* **2010**, *31*, 671--690.

[28] B. Brooks, R. Bruccoleri, D. Olafson, D. States, S. Swaminathan, M. Karplus, *J. Comput. Chem.* **1983**, *4*, 187--217.

[29] A. MacKerel Jr., C. Brooks III, L. Nilsson, B. Roux, Y. Won, M. Karplus, *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*, Vol. 1 of *The Encyclopedia of Computational Chemistry*, P. v. R. Schleyer et al. (Ed.), John Wiley & Sons: Chichester, **1998**, pp. 271--277.

[30] W. Jorgensen, *J. Am. Chem. Soc.* **1981**, *103*, 335--340.

[31] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089--10092.

[32] J. Ryckaert, G. Ciccotti, H. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327--341.

[33] J. Dunitz, *Science* **1994**, *264*, 670.

[34] C. Sun, T. Tang, H. Uludag, *Biomacromolecules* **2011**, *12*, 3698--3707.

[35] A. S. Thomas, A. H. Elcock, *J. Phys. Chem. Lett.* **2011**, *2*, 19--24.

[36] H. M. Aliabadi, B. Landry, P. Mandipoor, C. Y. M. Hsu, H. Uludag, *Eur. J. Pharmaceutics and Biopharmaceutics* **2012**, *81*, 33--42.

[37] H. M. Aliabadi, B. Landry, P. Mandipoor, H. Uludag, *Mol. Pharm.* **2011**, *8*, 1821--1830.

[38] C. L. Waite, S. M. Sparks, K. E. Uhrich, C. M. Roth, *BMC Biotechnol.* **2009**, *9*.

[39] T. Suma, K. Miyata, T. Ishii, S. Uchida, H. Uchida, K. Itaka, N. Nishiyama, K. Kataoka, *Biomaterials* **2012**, *33*, 2770--2779.

Chapter 7
Conclusions and Future Prospects

7.1 Conclusions

In this dissertation, we performed a series of all-atom MD simulations for the complexation of nucleic acids with native and lipid-modified PEIs. The simulation results generally agreed well with experimental data and moreover, explored some structural information of the polyplexes which cannot be accessed through experiments.

We started our simulations from the simplest case -- a single DNA interacting with a single PEI and the PEI has the lowest molecular weight (600 Da) among all the commercially available PEIs. Eight 600 Da PEIs with four different architectures and at two protonation ratios (23% and 46%) were adopted. We found that the PEIs primarily bind to the DNA backbone through the formation of hydrogen bonding with the backbone oxygens. The 46% protonated PEIs bind to the DNA more tightly and form more stable complexes compared with 23% protonated PEIs. Compared with the protonation state, the degree of branching has a smaller effect on binding, which essentially diminishes at the protonation ratio of 46%.

As molecular weight is an important parameter in affecting PEI's performance as gene carrier, we then chose two 2 kDa PEIs with distinct architectures (linear and branched with 14 primary amine groups) to check if the PEI architecture affects the complexation at the 2 kDa molecular weight. Because the DNA is normally mixed with excessive PEIs in experimental settings, we increased the complexity of the systems by introducing multiple PEI molecules in the simulations. The numbers of PEI molecules simulated correspond to two PEI/DNA N/P ratios. Also, the effect of ion concentration was investigated by simulating each system at two ion concentrations -- 0 and 154 mM. The simulations revealed distinct binding modes of branched and linear 2 kDa PEIs to DNA, with branched PEIs adhering to the

DNA surface like beads and linear PEIs adhering to the DNA surface like cords. This demonstrates that molecular weight of PEI is an important factor in PEI/DNA complexation. Further evidence for this exists in the fact that in excess of PEIs at a PEI/DNA charge ratio of 16/7, both branched and linear PEIs fully neutralize the DNA at a distance of $\sim$8 Å from the DNA C1' atoms, which is a significantly shorter distance compared with $\sim$12 Å in the case of excessive 600 Da PEIs. For both branched and linear PEIs, the addition of 154 mM salt ions was found to have only a small effect on PEI/DNA complexation compared to salt-free conditions.

As the first step on the delivery pathway, PEIs need to condense nucleic acids into nanoparticles. Hence, increasing the system's complexity from one nucleic acid molecule to multiple nucleic acid molecules is of significance to simulating a more realistic situation and to exploring the mechanisms of PEI mediated nucleic acid aggregation. In Chapter 4, we performed a series of MD simulations for 600 Da branched PEI mediated DNA aggregation at various PEI/DNA ratios. We found that PEIs condense DNA through two mechanisms -- polyion bridging and electrostatic screening of the DNA charges. At PEI/DNA charge ratio $>$1, PEIs can completely neutralize DNAs at a short distance ($\sim$12 Å from the C1' atoms), and this distance was found to be insensitive to the exact value of the charge ratio. When excess PEIs were added to a formed PEI/DNA polyplex, they were found to bind to the aggregate and increase its cationic charge. Part of the added PEIs also replaced the PEIs previously bound to the aggregate. The excess PEIs, however, do not change the spacing of the DNAs in the polyplexes.

To probe the effects of lipid substitution on PEI mediated DNA aggregation, following the methodology established in Chapter 4, we replaced the native 600 Da PEIs with lipid-modified PEIs at the substitution level of one OA substitution per PEI. We found that the lipid moieties associate significantly with one another, which serves as an additional mechanism of aggregating the DNAs and stabilizing the formed polyplex. In addition, some lipid moieties stay at the periphery of the polyplex and increase the hydrophobicity of the formed polyplex. The enhanced stability and hydrophobicity might contribute to the better cellular uptake of polyplexes formed with lipid-modified PEIs.

To further explore the effects of lipid length and substitution level on the formed polyplexes, in Chapter 6, we adopted siRNA and five PEI molecules according to those used in experiments. The five PEIs include one native 2 kDa branched PEI and its four analogues with two species of lipid substituent at two substitution levels. We found the lipid modification does not affect PEI's capability of neutralizing the siRNA charge, neither does it affect the polyion bridging. Similarly as observed in Chapter 5, significant linkages among the lipid modified PEIs via association of lipid moieties were observed and this results in more stable and compact PEI/siRNA polyplexes. The lipid associations between short lipids form and break frequently while the lipid associations between long lipids are more stable. For PEIs modified with short lipids, increasing the lipid substitution level results in more compact and stable siRNA structure. For PEIs modified with long lipids, increasing the lipid substitution does not change the amount of PEI linkage via lipid association, and has a reverse effect on compacting siRNA structure due to increased steric hindrance brought by the lipid association on individual PEIs.

The results presented in this dissertation will advance the understanding of the complexation of nucleic acids with native and lipid-modified PEIs at atomistic level and shed light on the structure-function relationship of PEI-based carriers. Some of the conclusions can be applied to other polycationic carriers. More importantly, the methodology and analyzing techniques used in this dissertation can serve as a framework for simulating other polycationic carriers, thus helping design more effective polycationic carriers.

7.2 Future Prospects

Up to date, development of functional gene carriers has been mostly driven by experimental efforts, but we believe that using computational simulations to characterize the complexation of nucleic acids and carriers will make a significant impact and ultimately revolutionize the carrier development process. Although the current computational capability has limited the atomistic simulations to systems of relatively small size ($<$1,000,000 atoms) and short time scale ($<$1000 ns) compared

with those in practical conditions, the length and time scale accessible through simulations are rapidly increasing. In the next decade or so, we can anticipate atomistic simulations of many practical aspects of carrier based gene delivery to be realized. For instances, the following large scale simulations can be performed in helping design advanced carriers:

7.2.1 Structures and dynamics of nucleic acids/carrier complexes of practical size

The size of the systems simulated in this dissertation is limited by the current computational capacity and is only in the order of 10 nm, which is much smaller compared to the nanoparticles formed in experimental settings (typically in the order of 100 nm). In the future, with the ever advancing computational hardware/software, simulating nucleic acids/carrier systems of practical sizes in atomic resolution can be realized. This will allow us to directly compare structures and dynamics of complexation revealed in simulations with experimental data. Some important aspects of the complexation currently not feasible with atomic simulations could be explored through the simulations of practical sized systems. For example, we can predict at what size the complexes will stop growing with different carriers.

7.2.2 Interaction and dynamics of nucleic acids/carrier complexes with cell membranes, endosomal membranes and endogenous molecules

The nucleic acids/carrier complexes will interact with cell membranes and possibly encounter other biological entities including endosomal membranes and endogenous molecules on the delivery pathway. These interactions are paramount for the delivery process and could significantly affect the performance of a carrier, hence, simulating the interaction of complexes with cell membranes, endosomal membranes and endogenous is of great relevance.

7.2.3 Mechanisms and dynamics of nucleic acids releasing from nucleic acids/carrier complexes

At the functioning sites, the carriers need to unload the nucleic acids from the complexes, which is a crucial step for successful gene deliveries. The molecular mechanisms and dynamics of the nucleic acids releasing from the complexes can be elucidated from atomistic simulations. Moreover, we can use simulations to evaluate the nucleic acids releasing capability of different carriers and to investigate the effects of different function groups of the carriers on nucleic acids releasing.

From the perspective of MD force field development, with continually advancing force field, the simulation results will be more accurate and provide us with more reliable information. For example, with the next generation polarizable force field, verification of the proton sponge effect [1] of PEI carrier can be realized through MD simulations.

Bibliography

[1] J. Behr, *Chimia* **1997**, *51*, 34--36.

Appendix A
Supporting Information for Chapter 2

A.1 Details about developing the force field for PEI

The force field for PEI were developed based on the CHARMM General Force Field (CGenFF) [1]. The force field for the building blocks of PEI are available in CGenFF, which are "RESI EAMM, $CH_3-CH_2-NH_3$", "RESI DMAM, $CH_3-NH-CH_3$", and "RESI TMAM, $N-(CH_3)_3$", corresponding to the primary, secondary and tertiary amines, respectively. Based on these three residues all van der Walls parameters, most bonded parameters, and partial charges for each atom were determined. The remaining angle and torsion parameters were adopted from existing parameters for analogous atom groups in CGenFF. It has been argued, in the CGenFF paper [1], that this methodology can be advantageous over bonded parameters parameterized from quantum mechanics calculations as the existing parameters have been further tuned and validated against experiments after the initial quantum mechanics calculations.

A.2 Complex formation using CHARMM force field (this work) *VS.* using AMBER force field (Ref. 2)

To examine how the complex formation can be affected by using different force fields, we made a comparison to Ref. 2, where AMBER force field was used, by performing a simulation of 'System 50%-PEI(20)' defined in Ref. 2 with the same simulation procedure. Very close results were obtained, demonstrating the similarity of these two force fields in describing the DNA/PEI systems.

The simulation parameters reported in Ref. 2 were adopted as much as possible in our simulation:

1. A 20 monomers purely linear PEI (see definition of ‘‘purely linear’’ in the main texts) with 10 monomers (index 2, 4, 6, 7, 12, 13, 14, 16, 17 and 20) protonated;

2. 27 Na+, 15 Cl- counterions added in the water box to neutralize the system;

3. 50 Å initial separation between the DNA and PEI centers of mass;

4. SHAKE algorithm, 2 fs time step;

5. 10 Å cutoff for van der Waals and direct electrostatic calculation;

6. Particle mesh Ewald method;

7. 20 ps of heating to 300K after the minimization;

8. 1.2 ns NPT simulation with restraints on DNA and PEI after heating.

The differences between the two simulations exist in the following aspects:

1. Force field: CHARMM (Our simulation) vs. Amber (Ref. 1);

2. Initial configurations of PEI and initial relative position of DNA and PEI;

3. Water box size: $84 \times 93 \times 92$ Å^3 (our simulation) vs. $95 \times 100 \times 80$ Å^3 (Ref. 2);

4. Simulation time: we used a much longer simulation time (40 ns and 60 ns) since it appears that it took longer for our system to equilibrate.

A.2.1 Results

Figure A.1 plots the center of mass (COM) distance between the DNA and the PEI as a function of simulation time. The COM distance fluctuates around a constant for several ns at the beginning of the simulation and again at about 8-12 ns, but overall the COM distance decreases as the PEI approaches the DNA. The COM distance becomes stable at about 20 ns of the simulation when the complex has been formed with a significant part of the PEI in close contact with the DNA. The same characteristics was found in Ref. 2 (Shown in Figure A.2), however the rates

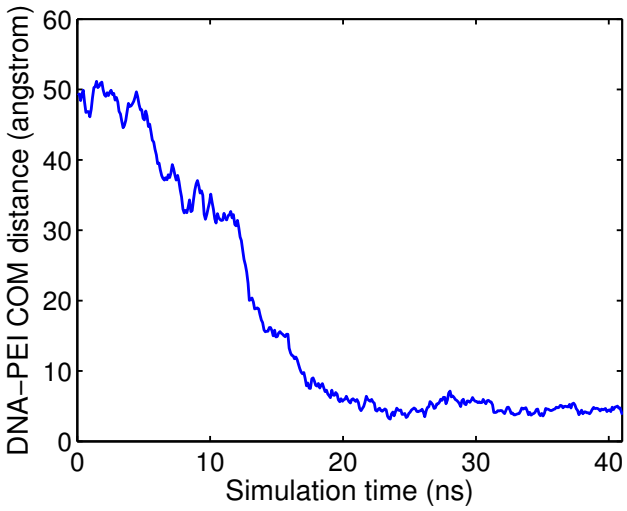

Figure A.1: Center of mass distance between the DNA and the PEI as a function of simulation time. Time is zeroed at the moment when the restraints were removed from the DNA and the PEI.

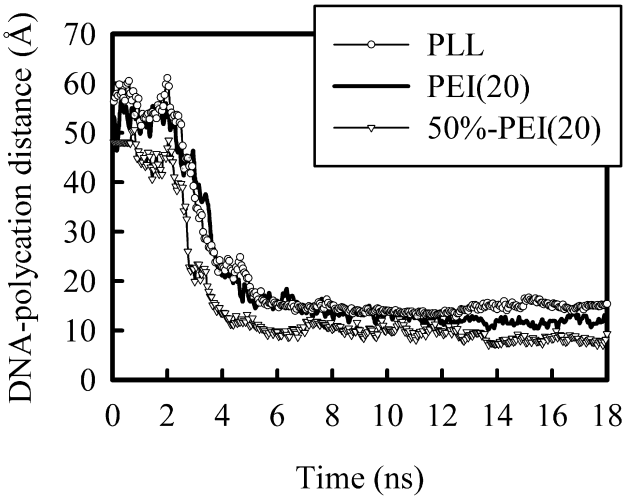

FIGURE 1 Plot of the distance between the centers of mass of DNA and the polycation chains as a function of simulation time for the first three systems, PEI(20), PLL(20), and 50%-PEI(20). Time zero corresponds to the moment when restraints on the chains were removed.

Figure A.2: Figure 1 from Ref. 2 [3].

of the complex formation in the two simulations are different. In our simulation the COM distance takes about 20 ns to stabilize, while in Ziebarth's simulation it only took about 5 ns. This could be due to the different force field used, but more probably, it may be due to the difference in the initial position of the PEI relative to the DNA and hence the positions of the image molecules (from periodic boundary condition).

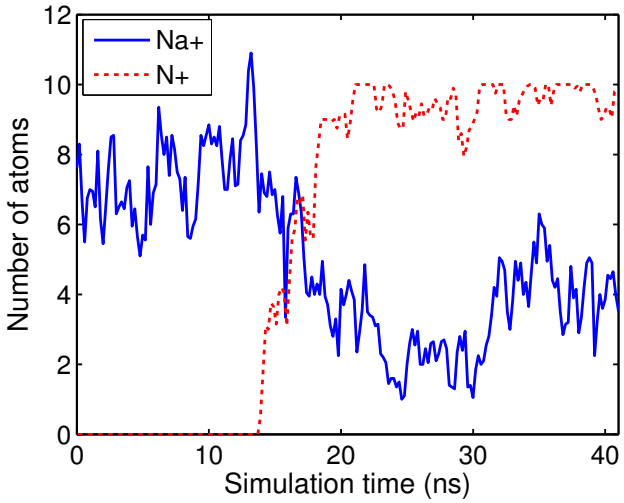

Figure A.3: The number of Na+ ions and protonated amine nitrogens within 10 Å of any C1' DNA atom as a function of time.

Figure A.3 plots the number of Na+ ions and protonated amine nitrogens within 10 Å of any C1' DNA atom as a function of time. As the PEI approaches the DNA, the number of Na+ decreases from a value of about 7 to about 3, indicating the release of Na+ around DNA is due to its association with the PEI. This curve was not plotted for the same system (DNA and a purely-linear 50% protonated PEI with 20 amine groups) in Ref. 2, however, the same phenomenon was found for system I (DNA and a 100% protonated PEI with 20 amine groups) in FIGURE 3 of Ref. 2 (Shown in Figure A.4).

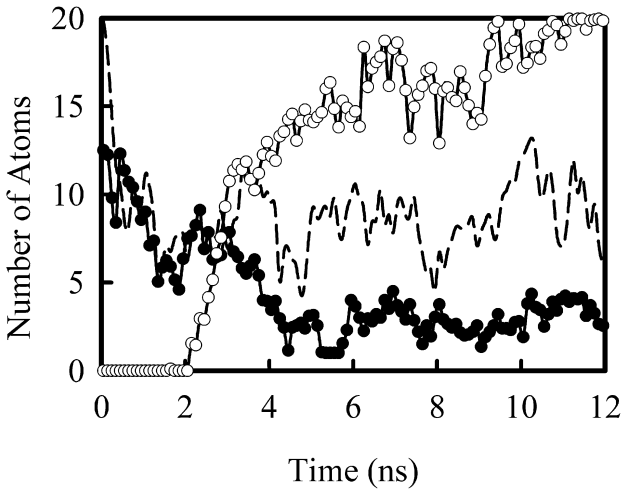

FIGURE 3 Plot of the number of Na^+ ions (*solid circles*) and protonated amine nitrogens (*open circles*) for system I (PEI(20)) within 10 Å of any C1′ DNA atom as a function of time for system I (PEI(20)). The dashed line shows the number of Na^+ ions for system VI as a reference. The number of Na^+ ion around DNA helix is reduced as PEI chain approaches the DNA helix.

Figure A.4: Figure 3 from Ref. 2 [3].

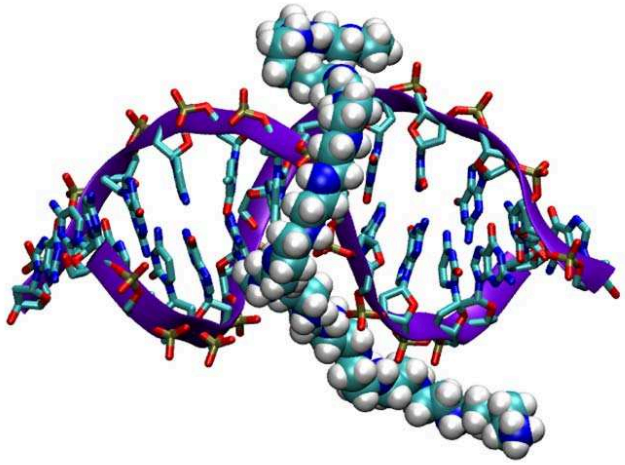

Figure A.5: A snapshot of the complex at the final stage of MD.

Figure A.5 is a snapshot of the complex at the final stage of the molecular dynamics (MD) simulation. The PEI is mainly in contact with only one strand of the DNA with a significant section of the PEI comply to the backbone of one DNA strand. The snapshot looks similar to the snapshot for the same system shown in FIGURE 6(f) of Ref. 2 (not shown here).

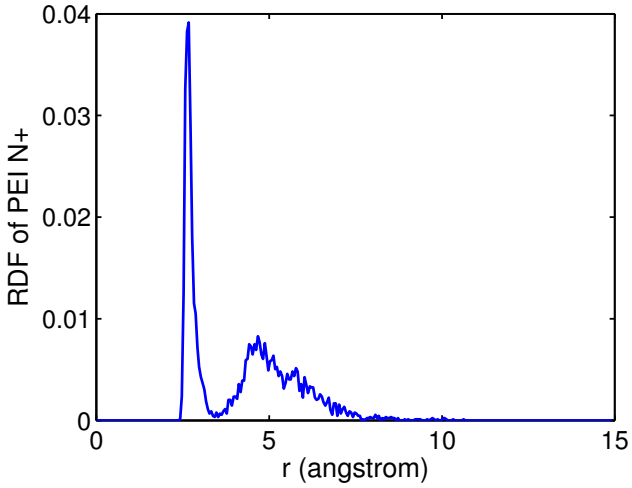

Figure A.6: Radial distribution functions (RDF) of Nitrogen atoms in protonated amine groups around the O1P and O2P DNA atoms.

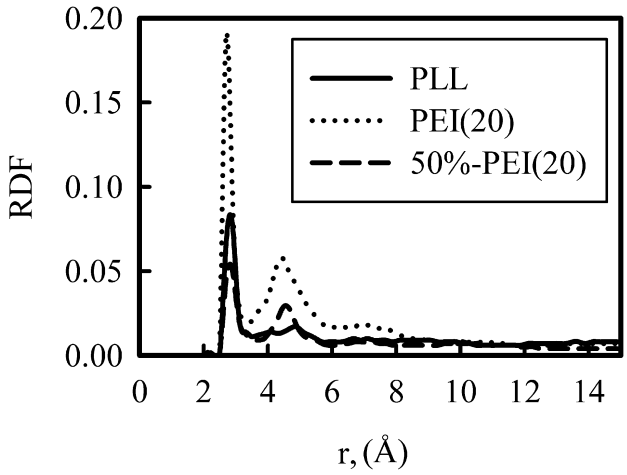

FIGURE 7 Radial distribution functions of polycation amine groups shown in figure legends around the O1P and O2P DNA atoms. In the case of the 50%-PEI(20) simulation, only charged amine groups are included.

Figure A.7: Figure 7 from Ref. 2 [3].

Figure A.6 is the radial distribution functions (RDF) of Nitrogen atoms in protonated amine groups around the O1P and O2P DNA atoms, where distance step 0.05 Å was used in generating the figure. The RDF curve has two peaks, one at about 3 Å and the other one at about 5 Å. The first peak corresponds to direct contact between the amine groups and the DNA O1P, O2P atoms, the second peak corresponds to secondary interaction such as water-mediated hydrogen bonding. The RDF curve resembles that in FIGURE 7 of Ref. 2 (Shown in Figure A.7).

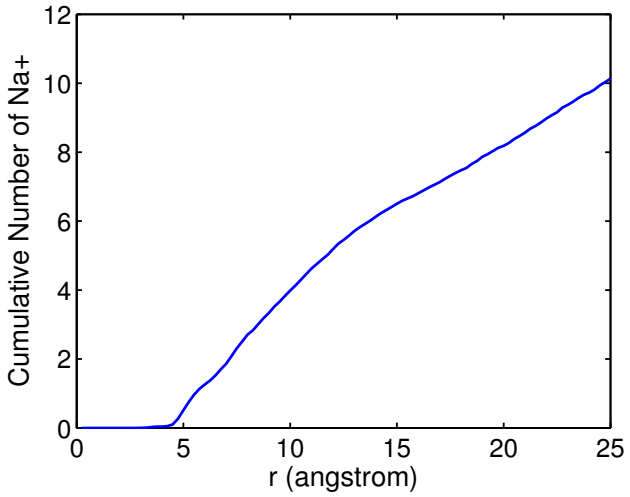

Figure A.8: Cumulative number of sodium ions as a function of the distance from any C1' DNA atom.

Figure A.8 shows the average cumulative number of sodium ions as a function of the distance from any C1' DNA atom during the last 6 ns of the simulation. The closet Na+ exists at about 5 Å from the C1' atoms. From 5 Å to 25 Å, the number of Na+ around the DNA gradually increases and approximately 10 Na+ are within 25 Å. The curve is very similar to the curve for 50%-PEI(20) in FIGURE 8 of Ref. 2 (Shown as dotted line in Figure A.9).

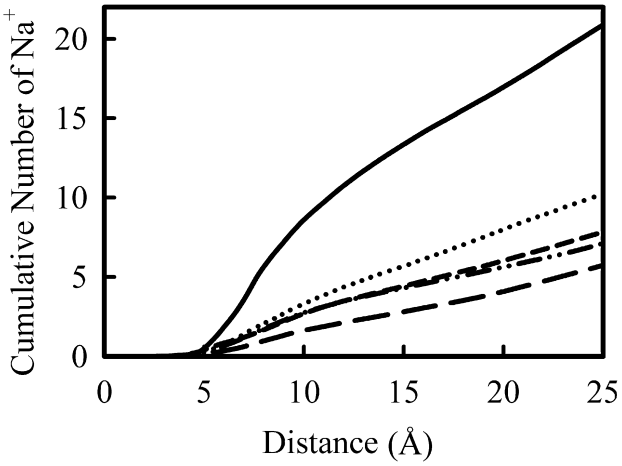

FIGURE 8 Cumulative number of sodium ions as a function of the distance from any C1′ DNA atom for each simulation. From the top line down, the Na^+ (*solid line*), 50%-PEI(20) (*dotted line*), PEI(20) (*short dashes*), PLL (*dots-dashes*), and 50%-PEI(40) (*long dashes*) systems are shown.

Figure A.9: Figure 8 from Ref. 2 [3].

A.3 Torsional parameters validation

In order to verify the correctness of the torsional parameters used for PEIs in the MD simulations, we have calculated the torsional potential energy surface (PES) as a function of certain representative dihedral angles at *ab initio* and molecular mechanics (MM) levels. Five compound models were used to calculated the PES: neutral N-Ethylethylenediamine (2-MI), secondary amine protonated N-Ethylethylenediamine (2-MI-P-A), primary amine protonated N-Ethylethylenediamine (2-MI-P-B), neutral branched trimethylimine (3-TMI) and primary amine protonated branched trimethylimine (3-TMI-P). The structure and atom type information of these five models is shown in Figure A.10. In conjunction, the five models encompass all 13 possible combinations of non-hydrogen atom types in a dihedral term for the PEIs studied in this work. The studied dihedrals are summarize in Table A.1. The *ab initio* quantum mechanical (QM) calculations were carried out using Gaussian 09 [4] at MP2/6-31+G* level. The MM calculations were performed using package CHARMM 33b2 [5] with our devised force field for PEI. For each dihedrals, 25 dihedral angles were calculated from -180° to 180° at an interval of 15°.

Figures A.11 to A.23 show the comparison of the PES calculated using QM and MM for the 13 dihedrals, respectively. A good agreement is observed in between the QM-PES and the MM-PES. In particular the overall shape, location of the maxima and minima and most of the relative energies of the QM model are reproduced by the MM calculation. For some dihedrals (e.g., NZ0-CX0-CX2-NZ2 shown in Figure A.20), evident discrepancies exist, however the overall behavior is reproduced. The results obtained support the notion that the parameters used in the MD simulation make a model that is a good description of the intra-molecular interactions in the PEIs.

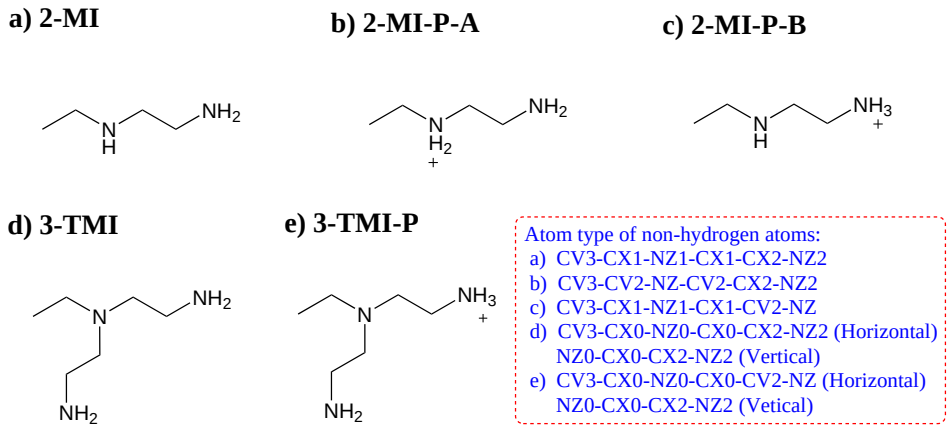

Figure A.10: Molecular structures of the five compound models. (a) neutral N-Ethylethylenediamine (2-MI), (b) secondary amine protonated N-Ethylethylenediamine (2-MI-P-A), (c) primary amine protonated N-Ethylethylenediamine (2-MI-P-B), (d) neutral branched trimethylimine (3-TMI), (e) primary amine protonated branched trimethylimine (3-TMI-P). Atom types of the atoms in each compound are specified in the red dashed square: for (a), (b), (c), the list corresponds to the atom types of the non-hydrogen atoms in the chain from left to right; for (c), (d), the first list corresponds the atom types of the non-hydrogen atoms in the horizontal chain from left to right, and the second list corresponds to the non-hydrogen atoms in the vertical branch from top to bottom.

Table A.1: The 13 dihedrals calculated

Dihedral	Model compound	QM/MM profiles
CV3-CX1-NZ1-CX1	2-MI	Figure A.11
CX1-NZ1-CX1-CX2	2-MI	Figure A.12
NZ1-CX1-CX2-NZ2	2-MI	Figure A.13
CV3-CV2-NZ-CV2	2-MI-A	Figure A.14
CV2-NZ-CV2-CX2	2-MI-A	Figure A.15
NZ-CV2-CX2-NZ2	2-MI-A	Figure A.16
CX1-NZ1-CX1-CV2	2-MI-B	Figure A.17
NZ1-CX1-CV2-NZ	2-MI-B	Figure A.18
CV3-CX0-NZ0-CX0	3-TMI	Figure A.19
NZ0-CX0-CX2-NZ2	3-TMI	Figure A.20
CX0-NZ0-CX0-CV2	3-TMI-P	Figure A.21
NZ0-CX0-CV2-NZ	3-TMI-P	Figure A.22
CX0-NZ0-CX0-CX2	3-TMI-P	Figure A.23

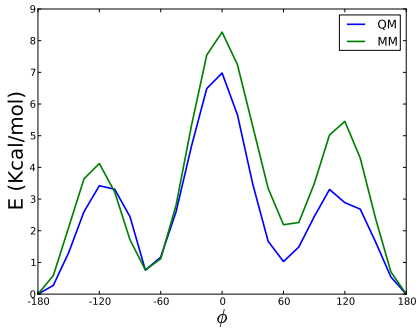

Figure A.11: Potential energy profiles of dihedral CV3-CX1-NZ1-CX1 calculated using QM and MM.

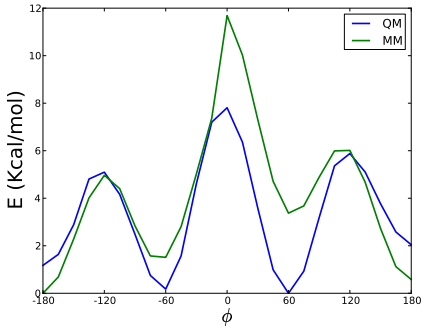

Figure A.12: Potential energy profiles of dihedral CX1-NZ1-CX1-CX2 calculated using QM and MM.

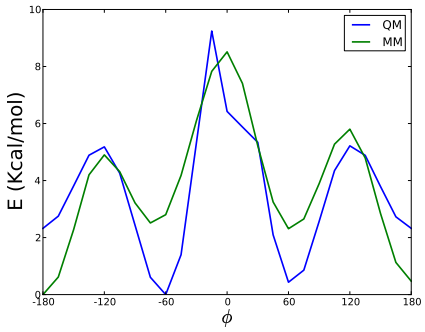

Figure A.13: Potential energy profiles of dihedral NZ1-CX1-CX2-NZ2 calculated using QM and MM.

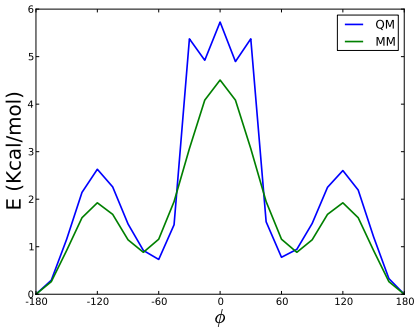

Figure A.14: Potential energy profiles of dihedral CV3-CV2-NZ-CV2 calculated using QM and MM.

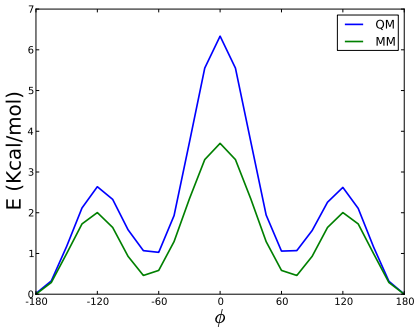

Figure A.15: Potential energy profiles of dihedral CV2-NZ-CV2-CX2 calculated using QM and MM.

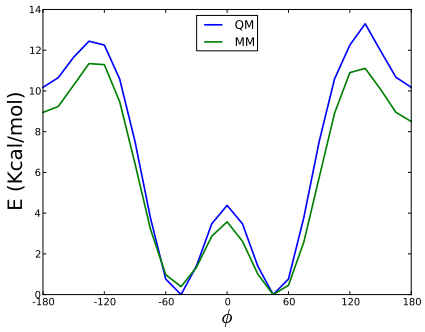

Figure A.16: Potential energy profiles of dihedral NZ-CV2-CX2-NZ2 calculated using QM and MM.

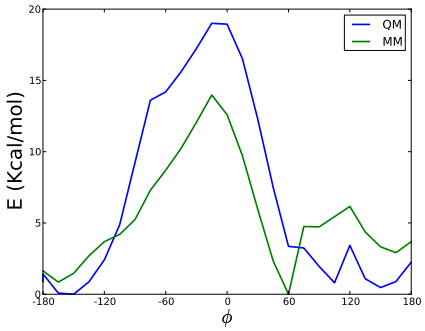

Figure A.17: Potential energy profiles of dihedral CX1-NZ1-CX1-CV2 calculated using QM and MM.

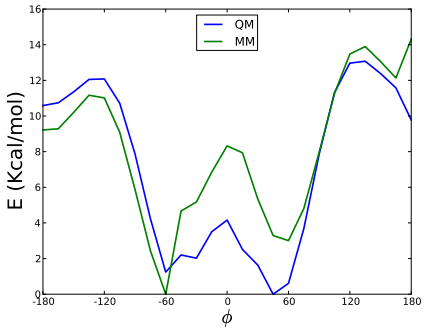

Figure A.18: Potential energy profiles of dihedral NZ1-CX1-CV2-NZ calculated using QM and MM.

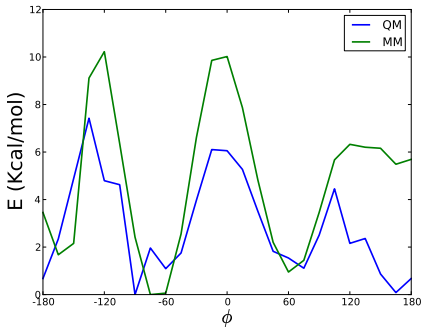

Figure A.19: Potential energy profiles of dihedral CV3-CX0-NZ0-CX0 calculated using QM and MM.

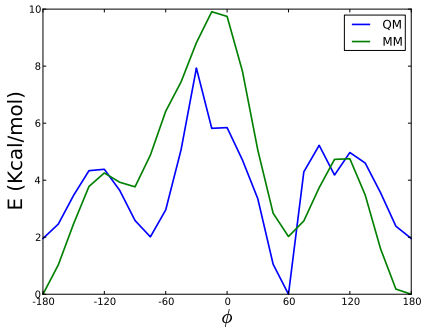

Figure A.20: Potential energy profiles of dihedral NZ0-CX0-CX2-NZ2 calculated using QM and MM.

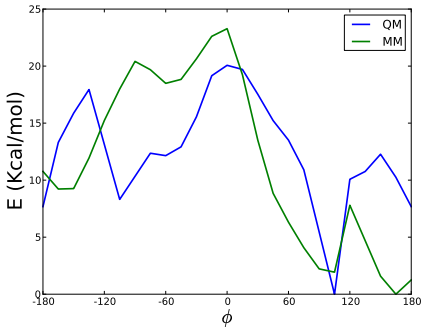

Figure A.21: Potential energy profiles of dihedral CX0-NZ0-CX0-CV2 calculated using QM and MM.

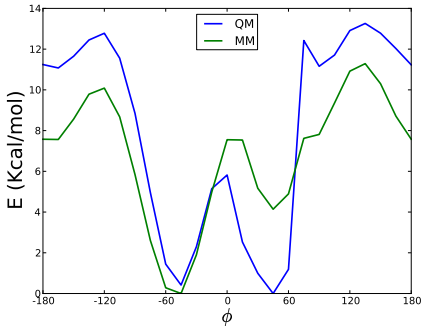

Figure A.22: Potential energy profiles of dihedral NZ0-CX0-CV2-NZ calculated using QM and MM.

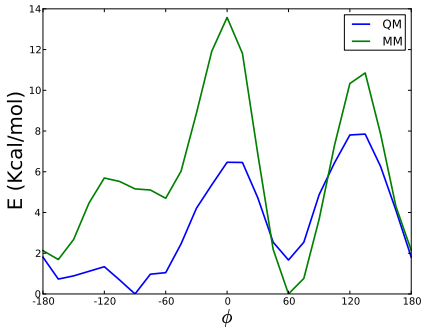

Figure A.23: Potential energy profiles of dihedral CX0-NZ0-CX0-CX2 calculated using QM and MM.

Table A.2: Torsional parameters of the 13 dihedrals. k_ϕ in kcal/mol, and δ in degree.

Dihedral	CGeenFF			Dong's work [6]		
	k_ϕ	n	δ	k_ϕ	n	δ
CV3-CX1-NZ1-CX1	1.26	3	0	1.0	3	0
CX1-NZ1-CX1-CX2	1.26	3	0	1.0	3	0
NZ1-CX1-CX2-NZ2	0.15	3	0	0.6	3	0
CV3-CV2-NZ-CV2	0.10	3	0	1.0	3	0
CV2-NZ-CV2-CX2	0.10	3	0	1.0	3	0
NZ-CV2-CX2-NZ2	0.15	3	0	0.6	3	0
CX1-NZ1-CX1-CV2	1.26	3	0	1.0	3	0
NZ1-CX1-CV2-NZ	0.15	3	0	0.6	3	0
CV3-CX0-NZ0-CX0	1.26	3	0	1.0	3	0
NZ0-CX0-CX2-NZ2	0.15	3	0	0.6	3	0
CX0-NZ0-CX0-CV2	1.26	3	0	1.0	3	0
NZ0-CX0-CV2-NZ	0.15	3	0	0.6	3	0
CX0-NZ0-CX0-CX2	1.26	3	0	1.0	3	0

A.4 The sensitivity of DNA/PEI binding pattern to the torsional parameters of the PEIs

We have also examined how sensitive our results are to variations in the torsional parameters. To do this, we replaced our 13 dihedral parameters completely with that used in Dong's MD simulations of PEI [6] and repeated two of the eight MD simulations for DNA/PEI complex formation (23%-PL and 46%-HB). The torsional parameters for PEI used in Dong's simulation were obtained by ''fitting an energy profile from a density functional calculation of for dimethylethylenediamine (DMEDA) into the dihedral angle torsion functional form'' [6], and the simulations yielded good results in comparison with experimental data. Table A.2 summaries the torsional parameters used in our work and in Ref. 6, where the torsional energy function is in the form of $E_{\text{torsion}} = k_\phi(1 + \cos(n\phi - \delta))$.

Figures A.24 and A.25 show the RDF and cumulative number of the PEI nitrogens around the DNA backbone oxygens based on the last 20 ns trajectory of the simulations for the 23%-PL system and the 46%-HB system, respectively. It can be observed from Figures A.24 and A.25 that, using a different set of torsional parameters which were derived from density function calculations and have been validated

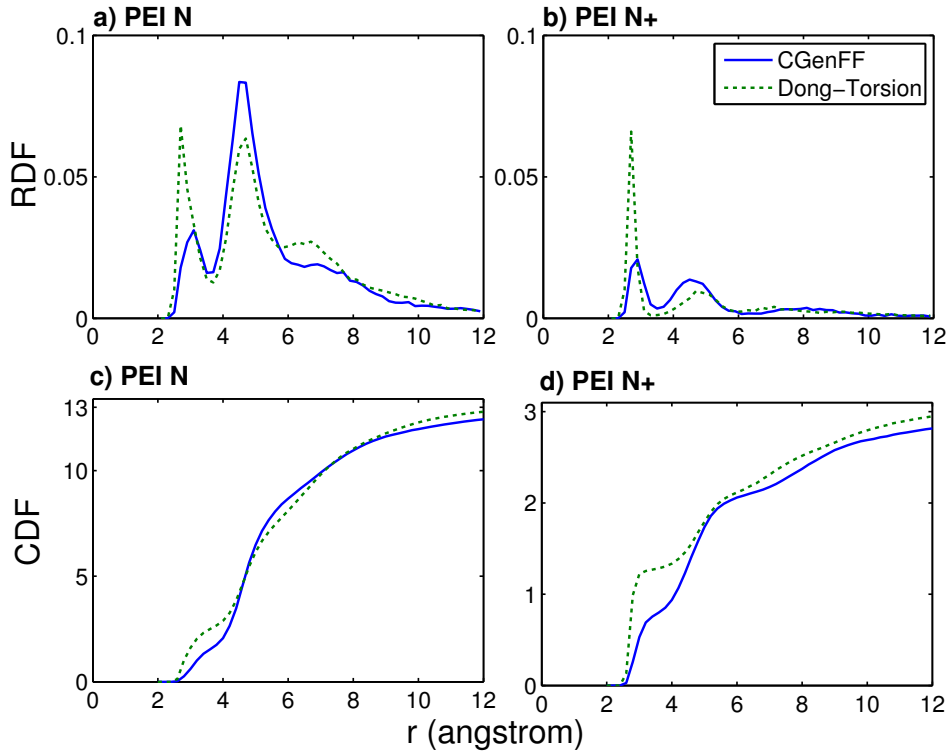

Figure A.24: Radial distribution function (RDF) and cumulative number (CDF) of the PEI nitrogens around the DNA backbone oxygens based on the last 20 ns trajectory of the simulations for 23%-PL system. (a) RDF of all PEI nitrogens, (b) RDF of protonated PEI nitrogens, (c) CDF of all PEI nitrogens, (d) CDF of protonated PEI nitrogens.

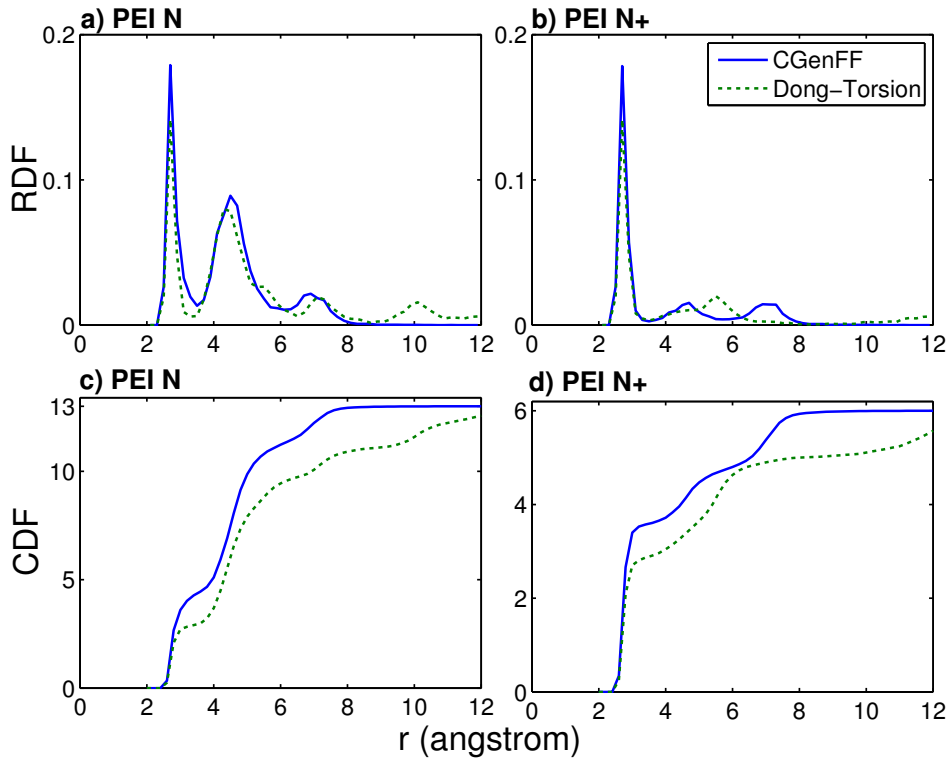

Figure A.25: Radial distribution function (RDF) and cumulative number (CDF) of the PEI nitrogens around the DNA backbone oxygens based on the last 20 ns trajectory of the simulations for 46%-HB system. (a) RDF of all PEI nitrogens, (b) RDF of protonated PEI nitrogens, (c) CDF of all PEI nitrogens, (d) CDF of protonated PEI nitrogens.

against experimental data, very similar results to our previous ones were obtained. Specially, the discrepancies associated with the two sets of torsional parameters in Figures A.24 and A.25 are comparable with the discrepancies among different simulation time windows using a single set of parameters as shown in Appendix A.5 of this document (Figures A.26 to A.33). For example, the discrepancies in Figure A.24(a) are comparable with the discrepancies in Figure A.26 among different simulation windows. None of the new results obtained using Dong's torsional parameters changes the conclusions we made in the manuscript.

Based on the calculations in Appendices A.3 and A.4, we believe that, for the focus of our study here which is the binding of DNA with PEIs, the force field we used is quantitatively meaningful. The force field for PEI might need to be further calibrated and validated if the objective is to study the conformation of PEIs in solution or crystal PEIs. However to study its binding to DNA, using the CGenFF principle to generate the force field parameters is a valid approach.

A.5 Radial distribution function (RDF) and cumulative number curves within different time windows in the simulations

Figure A.26 and Figure A.27 are respectively the RDF plots for all PEI nitrogens and for the protonated PEI nitrogens around the DNA backbone oxygens in the 23% systems. Figure A.28 and Figure A.29 are the same RDF plots for the 46% systems. Figures A.30 to A.33 are the corresponding cumulative number plots for 23% systems and 46% systems, respectively. These RDF and cumulative number plots were generated based on trajectories within different time windows in the simulations.

Figures A.26, A.27, A.30 and A.31 show that even after 49 ns of simulation, the curves are still evolving with time, and the order of the curves corresponding to different PEI structures do not maintain the same at all time. This indicates that the complexes formed in the 23% systems are not stable, which is consistent with the fact that the majority of the nitrogens bind to DNA through indirect interactions.

Compared with the 23% systems, the RDF and cumulative number curves for the 46% systems in Figures A.28, A.29, A.32 and A.33 demonstrate more stability (i.e., less variations among different simulation windows). Moreover, the curves corresponding to different PEI structures are closer to one another compared with the 23% systems. In fact, after 40 ns of simulations, these curves essentially overlap with one another. This indicates that at the protonation ration of 46%, the degree of branching has vanishingly small effect on the binding.

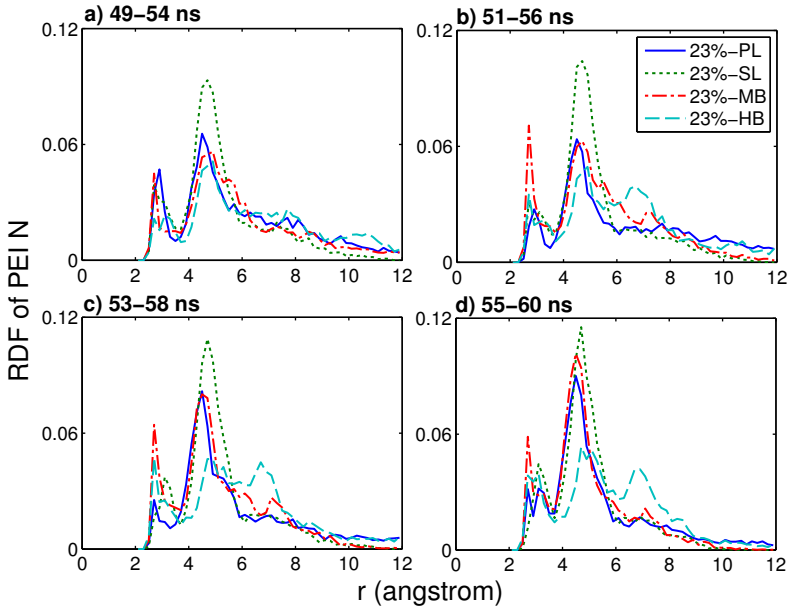

Figure A.26: Radial distribution functions of all PEI nitrogens around the DNA backbone oxygens for the 23% systems, plotted for different simulation time windows. (a) 49--54 ns, (b) 51--56 ns, (c) 53--58 ns, (d) 55--60 ns.

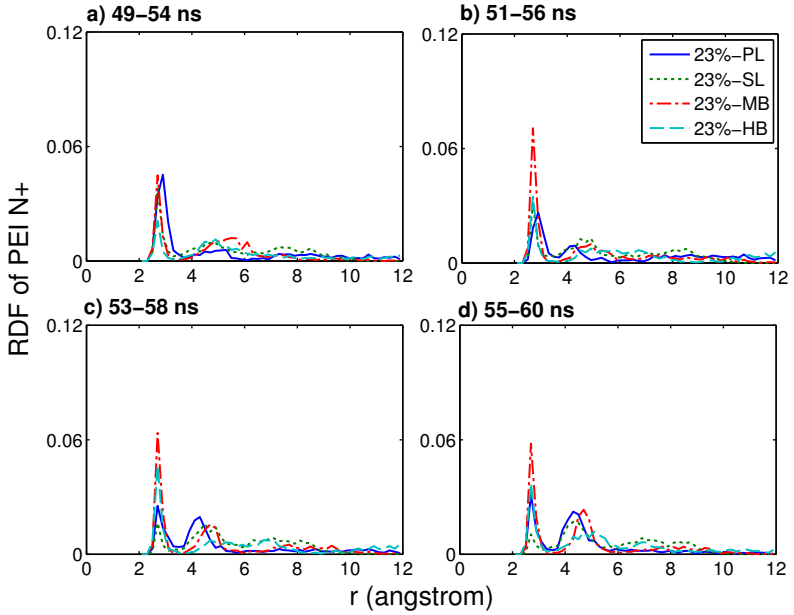

Figure A.27: Radial distribution functions of the PEI protonated nitrogens around the DNA backbone oxygens for the 23% systems, plotted for different simulation time windows. (a) 49--54 ns, (b) 51--56 ns, (c) 53--58 ns, (d) 55--60 ns.

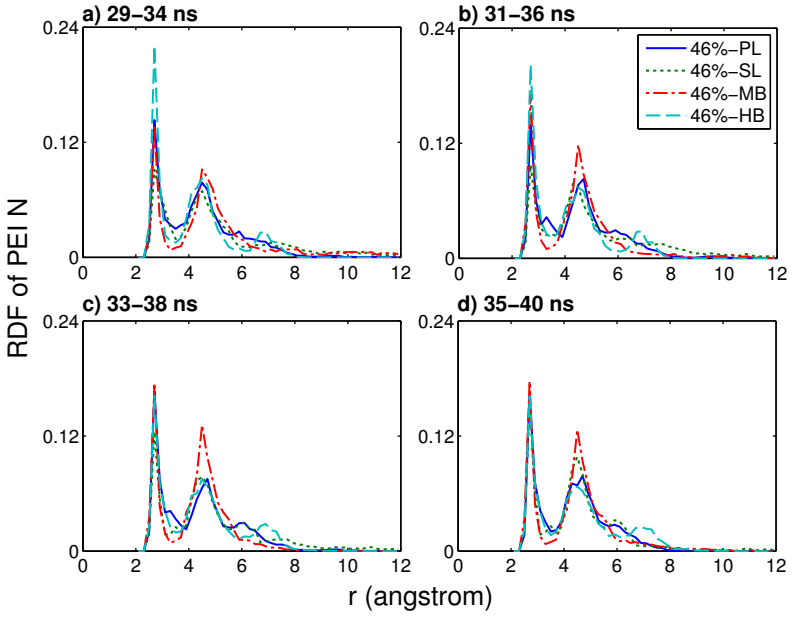

Figure A.28: Radial distribution functions of all PEI nitrogens around the DNA backbone oxygens for the 46% systems, plotted for different simulation time windows. (a) 29--34 ns, (b) 31--36 ns, (c) 33--38 ns, (d) 35--40 ns.

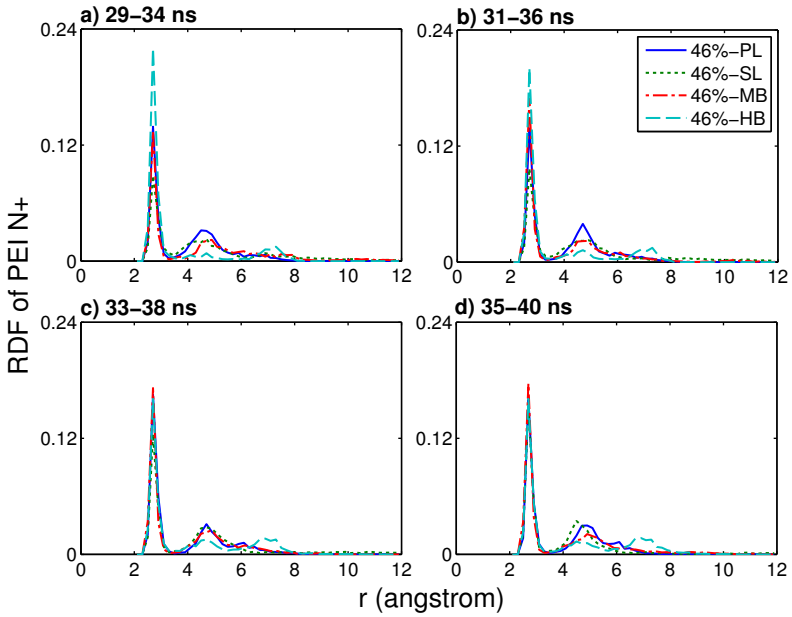

Figure A.29: Radial distribution function of the PEI protonated nitrogens around the DNA backbone oxygens for the 46% systems, plotted for different simulation time windows. (a) 29--34 ns, (b) 31--36 ns, (c) 33--38 ns, (d) 35--40 ns.

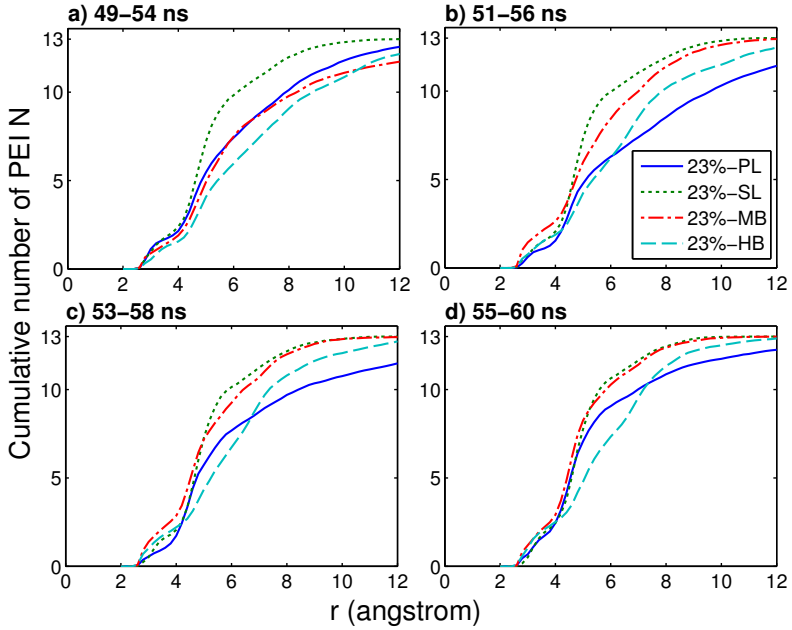

Figure A.30: Cumulative number of the PEI nitrogens around the DNA backbone oxygens for the 23% systems, plotted for different simulation time windows. (a) 49--54 ns, (b) 51--56 ns, (c) 53--58 ns, (d) 55--60 ns.

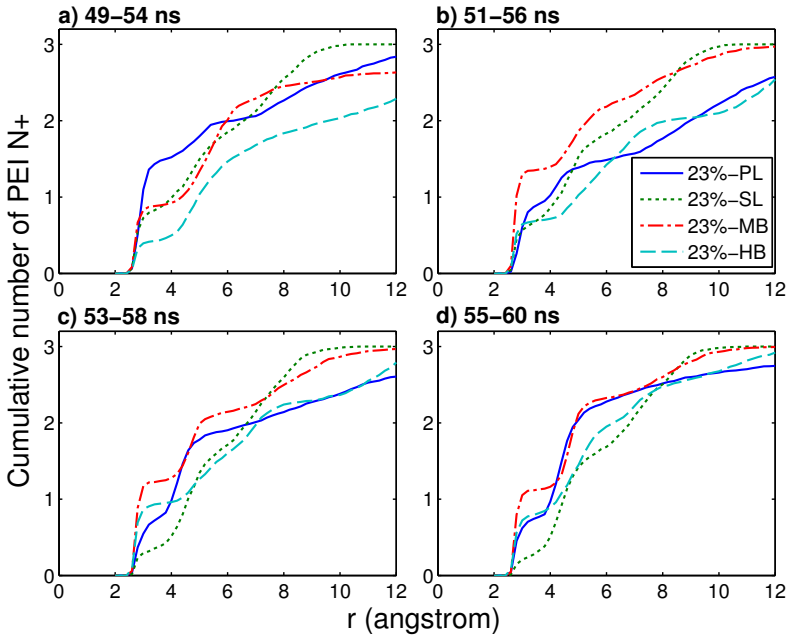

Figure A.31: Cumulative number of the PEI protonated nitrogens around the DNA backbone oxygens for the 23% systems, plotted for different simulation time windows. (a) 49--54 ns, (b) 51--56 ns, (c) 53--58 ns, (d) 55--60 ns.

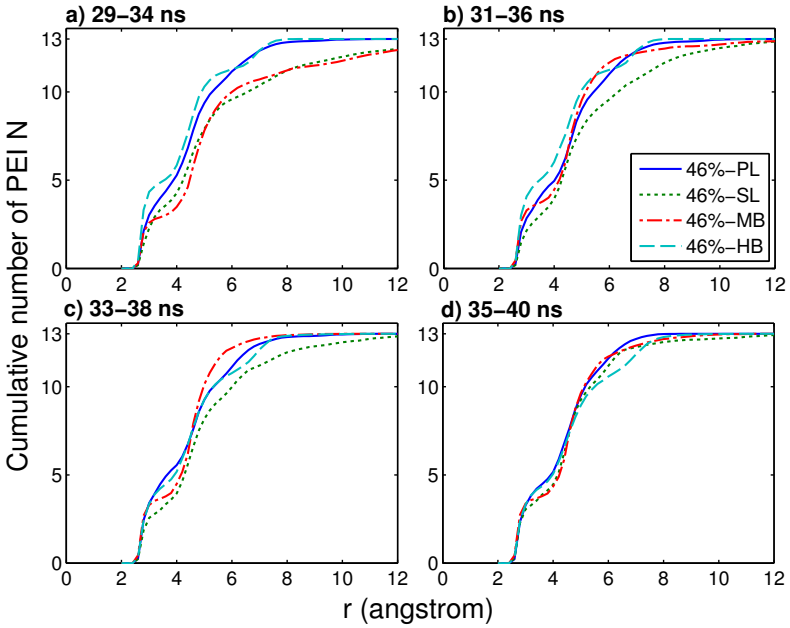

Figure A.32: Cumulative number of the PEI nitrogens around the DNA backbone oxygens for the 46% systems, plotted for different simulation time windows. (a) 29--34 ns, (b) 31--36 ns, (c) 33--38 ns, (d) 35--40 ns.

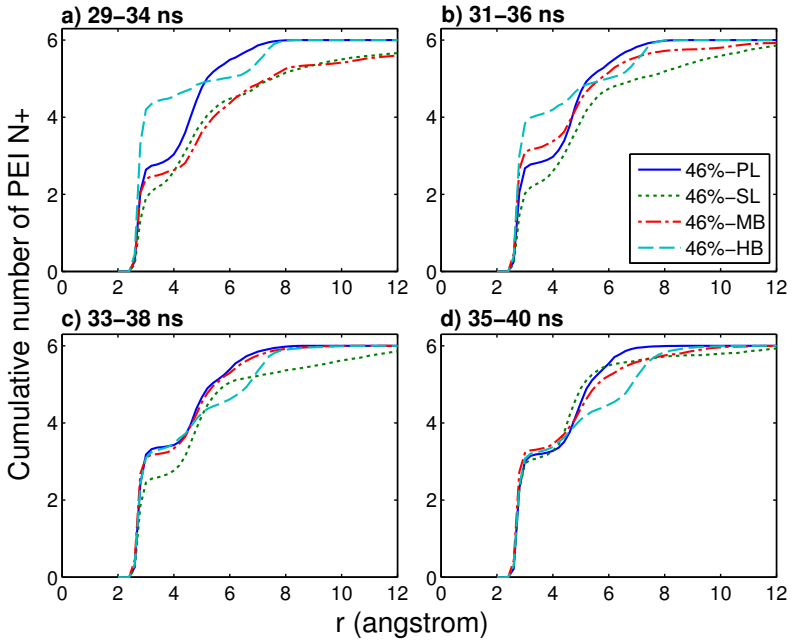

Figure A.33: Cumulative number of the PEI protonated nitrogens around the DNA backbone oxygens for the 46% systems, plotted for different simulation time windows. (a) 29--34 ns, (b) 31--36 ns, (c) 33--38 ns, (d) 35--40 ns.

Bibliography

[1] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. MacKerell, Jr., *J. Comput. Chem.* **2010**, *31*, 671--690.

[2] J. Ziebarth, Y. Wang, *Biophys. J.* **2009**, *97*, 1971--1983.

[3] Reprinted from Biophysical Journal, Vol 97, Jesse Ziebarth and Yongmei Wang, Molecular Dynamics Simulations of DNA-Polycation Complex Formation, Pages No. 1971, Copyright (2009), with permission from Elsevier.

[4] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, *Gaussian 09 Revision A.1*, Gaussian Inc. Wallingford CT 2009.

[5] B. R. Brooks, C. L. Brooks, III, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch,

L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, *J. Comput. Chem.* **2009**, *30*, 1545--1614.

[6] H. Dong, J. Hyun, C. Durham, R. Wheeler, *Polymer* **2001**, *42*, 7809--7817.

Appendix B
Supporting Information for Chapter 3

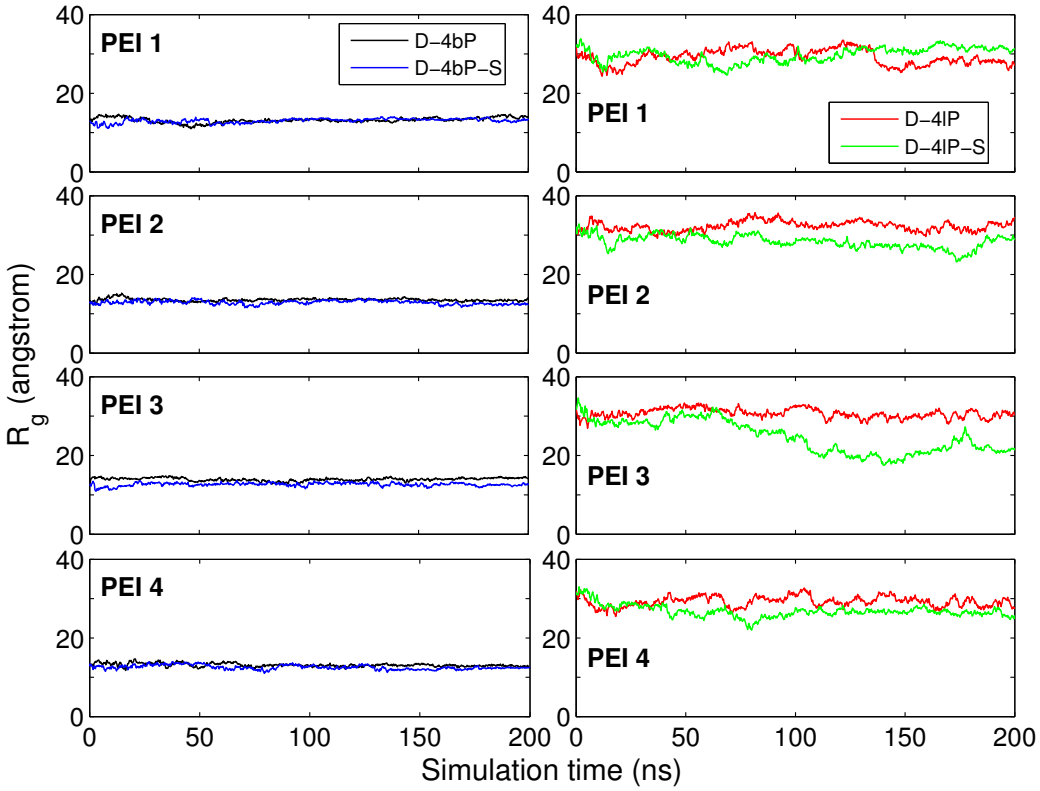

Figure B.1: Radius of gyration of each PEI as a function of simulation time in systems D-4bP, D-4bP-S, D-4lP and D-4lP-S.

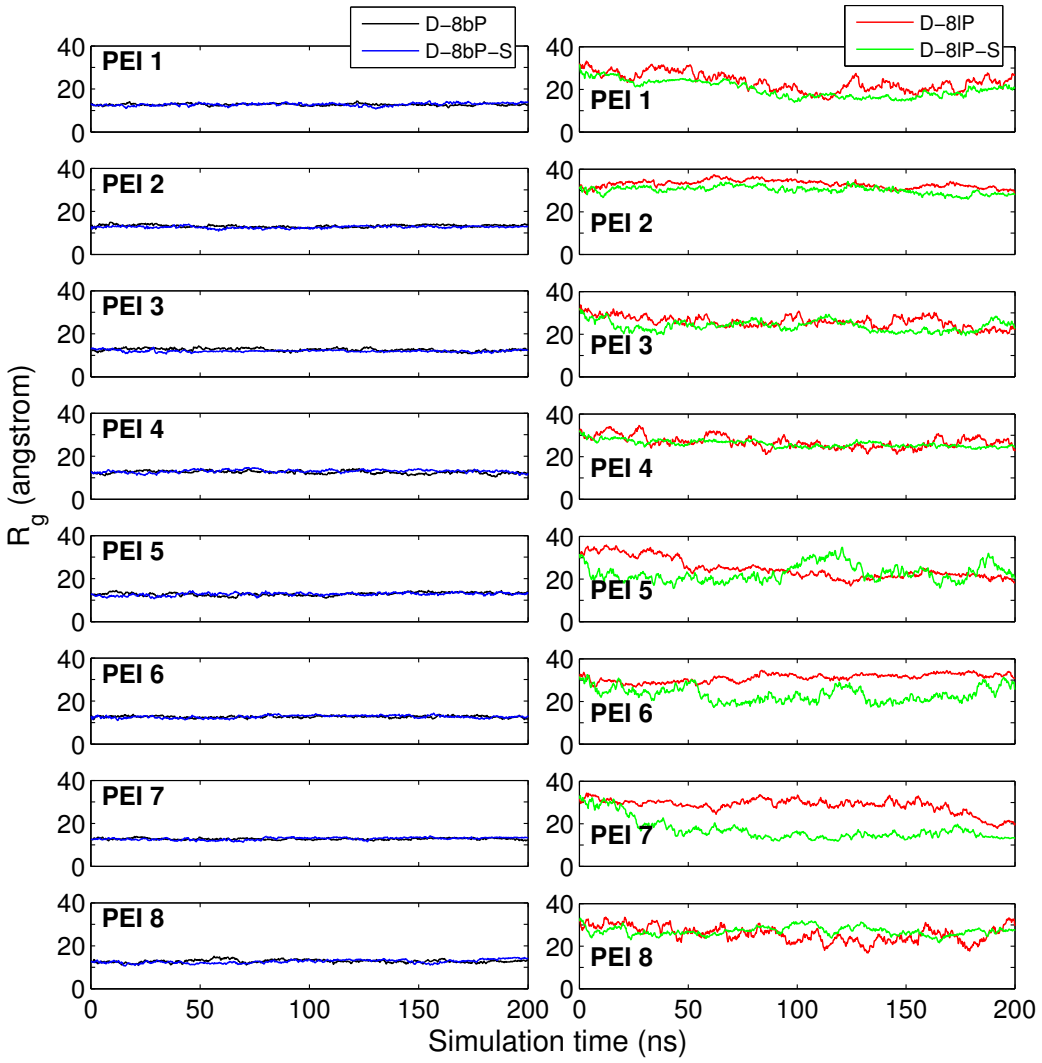

Figure B.2: Radius of gyration of each PEI as a function of simulation time in systems D-8bP, D-8bP-S, D-8lP and D-8lP-S.

Appendix C
Supporting Information for Chapter 4

C.1 Number of PEIs bound to DNA in system 4D-28P

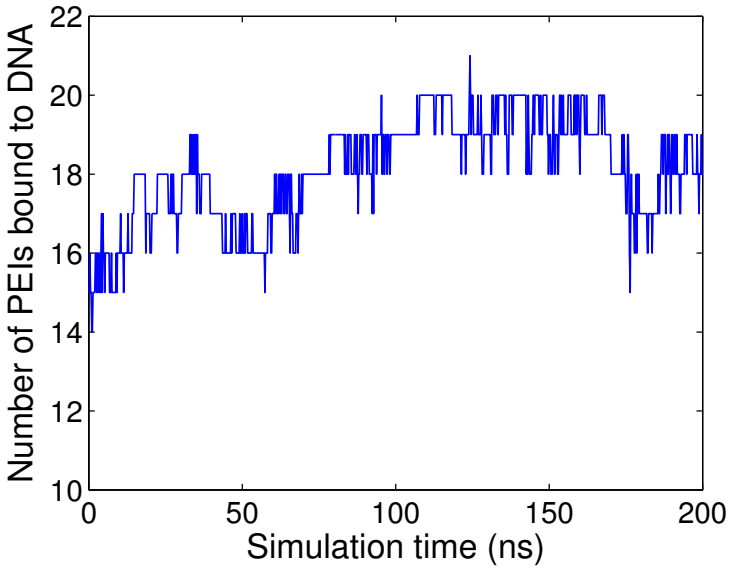

Figure C.1: Number of PEI molecules bound to DNA during the simulation for the 4D-28P system. The definition for 'bound' is that at least 1 out of the 13 nitrogens on a PEI are within 4 Å of any DNA N/O atoms.

C.2 Structure of DNA dodecamer and special configurations of two DNAs

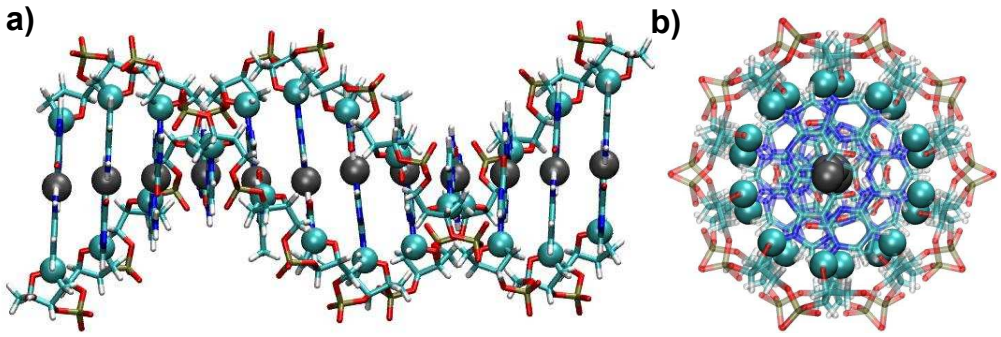

Figure C.2: DNA dodecamer in canonical B form. C1' atoms are represented in small blue spheres; the centers of mass of each Watson-Crick base pair are represented in large grey spheres. (a) Side view, (b) Axis view.

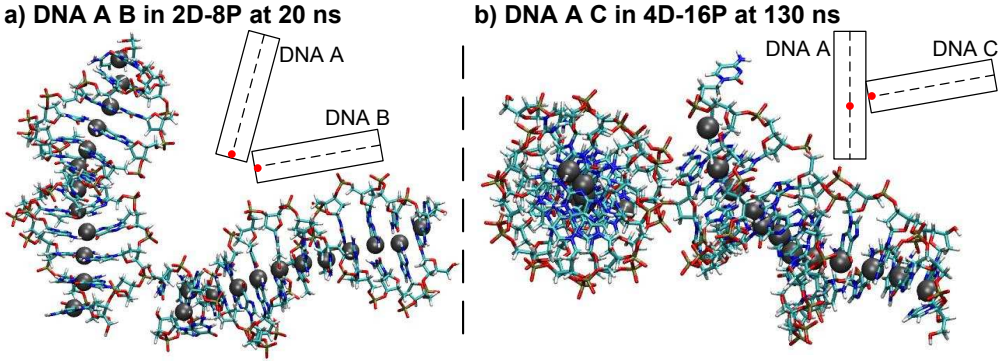

Figure C.3: Configurations of two DNAs for which d_{shortest} can not truly reflect the shortest inter-duplex distances. (a) DNA A B in 2D-8P at 20 ns, (b) DNA A C in 4D-16P at 130 ns. The centers of mass of each Watson-Crick base pair are represented in grey spheres. The two schematic drawings at the top-right corners of each subfigure illustrate the closet points for these two configurations.

C.3 Evidence of convergence of the simulation trajectories

Figure C.4 shows the charge neutralization curves (cumulative net charge of PEI/Na+/Cl- as a function of the distance from any C1' DNA atom) based on three time windows (last 10 ns, 2nd last 10 ns and 3rd last 10ns) at the late stage of the simulations. It can be seen that the curves almost overlap with one another, demonstrating the convergence of the simulation trajectories.

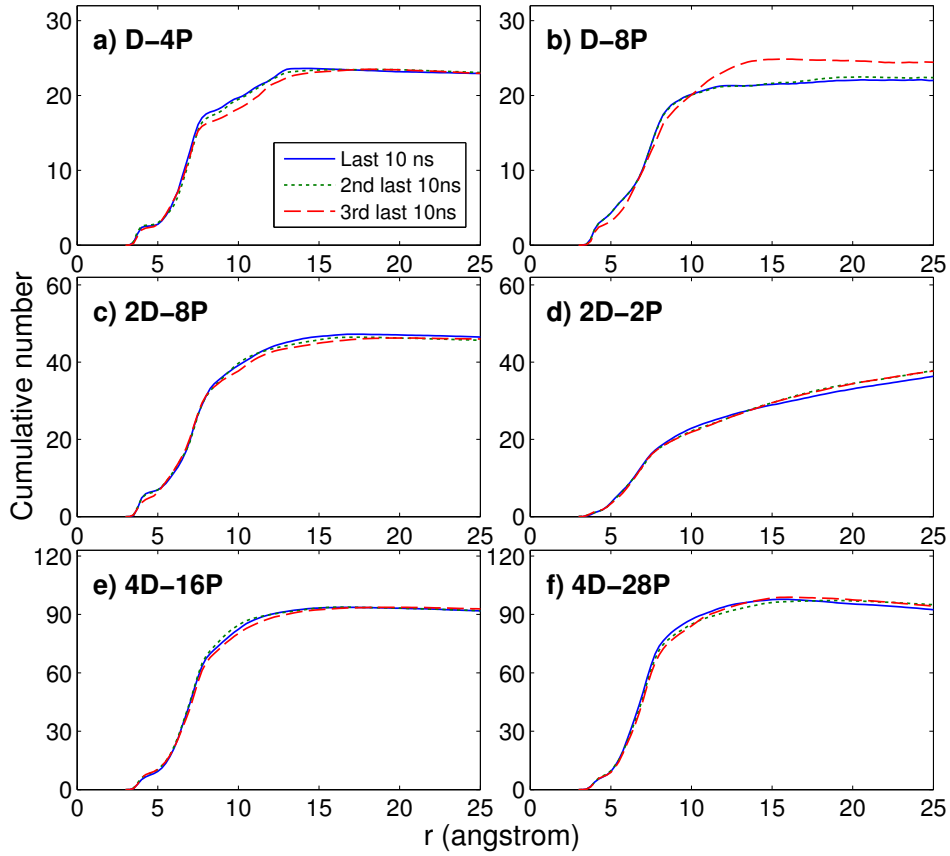

Figure C.4: Cumulative numbers of net charge of PEI/Na+/Cl- as a function of the distance from any C1' DNA atom based on three time windows (last, 2nd last and 3rd last 10 ns) at the late stage of the simulation. (a) D-4P, (b) D-8P, (c) 2D-8P, (d) 2D-2P, (e) 4D-16P, (f) 4D-28P.

C.4 Radii of gyration of the DNA aggregates

Figure C.5 shows the radii of gyration R_g of the DNA aggregates (2 DNAs in the 2D-8P and 2D-2P systems, and 4 DNAs in the 4D-16P and 4D-28P systems) as a function of simulation time, not accounting for the PEIs, in each system. For a given aggregate, a higher value of R_g is an indicator of more loosely arranged DNA molecules in the aggregate. For the 2D-8P and 4D-16P systems, the R_g curves decrease gradually in the beginning 50 ns, indicating the DNAs become more intimate. They then fluctuate for the rest of the simulation. As the DNAs in these two systems are relatively close at the beginning of the simulations, R_g only undergoes a small decrease. For the 2D-2P system, R_g at 0 ns is equal to that for 2D-8P at 100 ns as the initial configuration for the two DNAs is adopted from 2D-8P at 100 ns. The R_g curve increases slightly during the beginning 50 ns, indicating the two DNAs become less intimate. It fluctuates during the period of 50--150 ns, but at around 150 ns, R_g sharply increases from $\sim$20 Å to $\sim$25 Å within a time window of about 10 ns, corresponding to the separation of the two DNAs. The R_g curve continues to increase with strong fluctuation thereafter, reflecting the diffusion of independent DNAs. For the 4D-28P system, R_g at 0 ns is equal to that for the 4D-16P system at 100 ns as the initial configuration for the four DNAs is adopted from 4D-16P at 100 ns. It can the seen that, over the entire 200 ns simulation time, the R_g curve remains almost constant at around 25 Å, which again demonstrates that the compactness of the DNAs in the aggregate was barely affected by the excessive PEIs. Generally, these curves follow a similar trend as 'RMS distance' in Figures 9 to 12 in the main text.

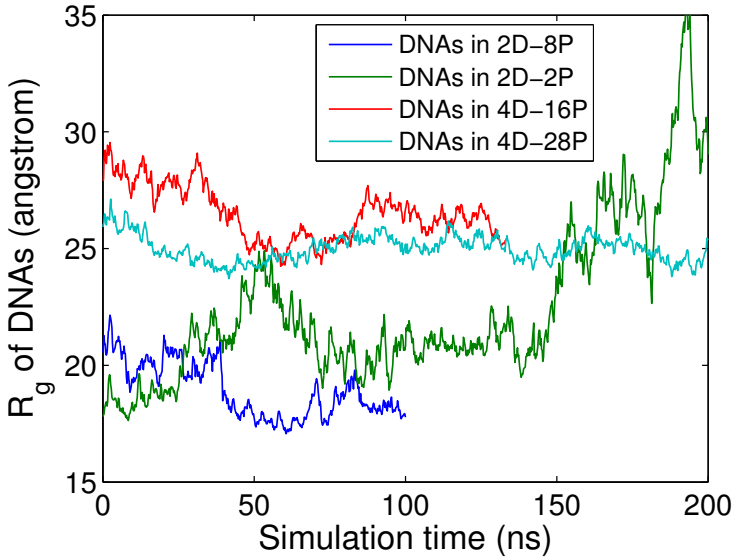

Figure C.5: Radius of gyration of all the DNAs in a system as a function of simulation time. (a) 2D-8P, (b) 2D-2P, (c) 4D-16P, (d) 4D-28P.

Appendix D
Supporting Information for Chapter 5

D.1 Evidence of convergence of the simulation trajectories

Figure D.1 shows the charge neutralization curves (cumulative net charge of PEI and ions as a function of the distance from any DNA C1' atom) based on four time windows (last 10 ns, 2nd last 10 ns, 3rd last 10 ns and 4th last 10 ns) at the late stage of the simulations. It can be seen that the curves almost overlap with one another, demonstrating the convergence of the simulation trajectories.

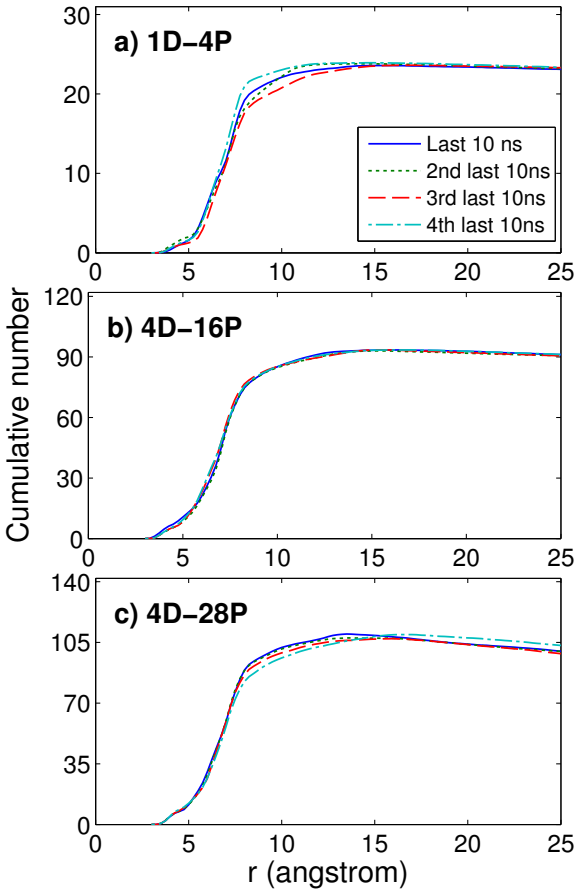

Figure D.1: Cumulative numbers of net charge of PEI and ions as a function of the distance from any DNA C1' atom based on four time windows (last 10 ns, 2nd last 10 ns, 3rd last 10 ns and 4th last 10 ns) at the late stage of the simulation. (a) D-4P, (b) 4D-16P, (c) 4D-28P.

D.2 Binding of individual lmPEIs to each DNA

A PEI N is said to be 'in close contact with the DNA' if it falls within 4 Å of any N/O atoms of the DNA. We chose 4 Å because this is the distance within which the PEI amine groups can form direct hydrogen bond with the DNA [1]. A PEI is said to be 'bound' to a DNA molecule if it has one or more Ns in close contact with this DNA. If a PEI is 'bound' to two or more DNA molecules simultaneously, we say that this PEI form a polyion bridge between the DNAs. To quantify the ability of lmPEI to bridge the DNA molecules, we plotted the binding state of individual lmPEIs to each DNA in terms of the number of Ns from each lmPEI in close contact with each DNA, as shown in Figure D.2 for the 4D-16P system and in Figure D.3 for the 4D-28P system. In Figure D.2, each subfigure corresponds to one of the 16 lmPEIs in the 4D-16P system, and it contains 4 curves each of which describes the number of Ns of this lmPEI in close contact with a particular DNA. Similarly, the 28 subfigures in Figure D.3 correspond to the 28 lmPEIs in the 4D-28P system and the 4 curves in each subfigure describe the binding state of a lmPEI with the four DNAs. In Figure D.2, out of the 16 lmPEIs, five lmPEIs (2, 6, 8, 10, and 13) participate in bridging two or three DNAs for longer than 50% of the simulation time. In Figure D.3, there are also five lmPEIs (2, 6, 10, 13 and 18) participating in bridging two or three DNAs for longer than 50% of the simulation time, and lmPEI 10 bridges DNAs A, C and D during most time of the simulation. Six out of the added 12 lmPEIs (17, 18, 20, 22, 26 and 27) bind with at least one DNA for significantly long periods. All the original 16 lmPEIs keep binding to the DNAs and none of them was 'replaced' by the added lmPEIs.

D.3 Calculation of water release

In order to calculate the number of water molecules released during the aggregation process, we counted the number of water molecules within 3 Å of the solutes in each system as summarized in Table D.1. To determine the number of water molecules released for a particular system, we subtract the number of water molecules within 3 Å of the polyplex from the total number of water molecules within 3 Å of the

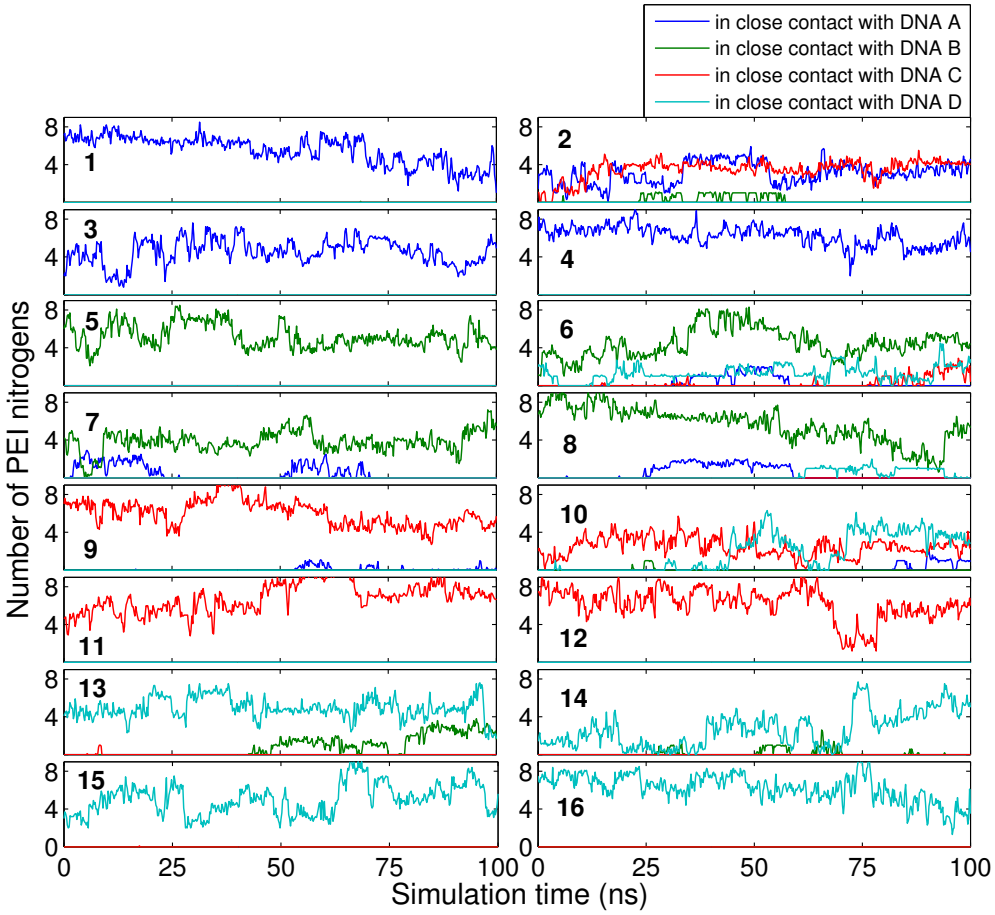

Figure D.2: Number of nitrogens for each lmPEI that are within 4 Å of any N/O atom of each DNA as a function of the simulation time for the 4D-16P system.

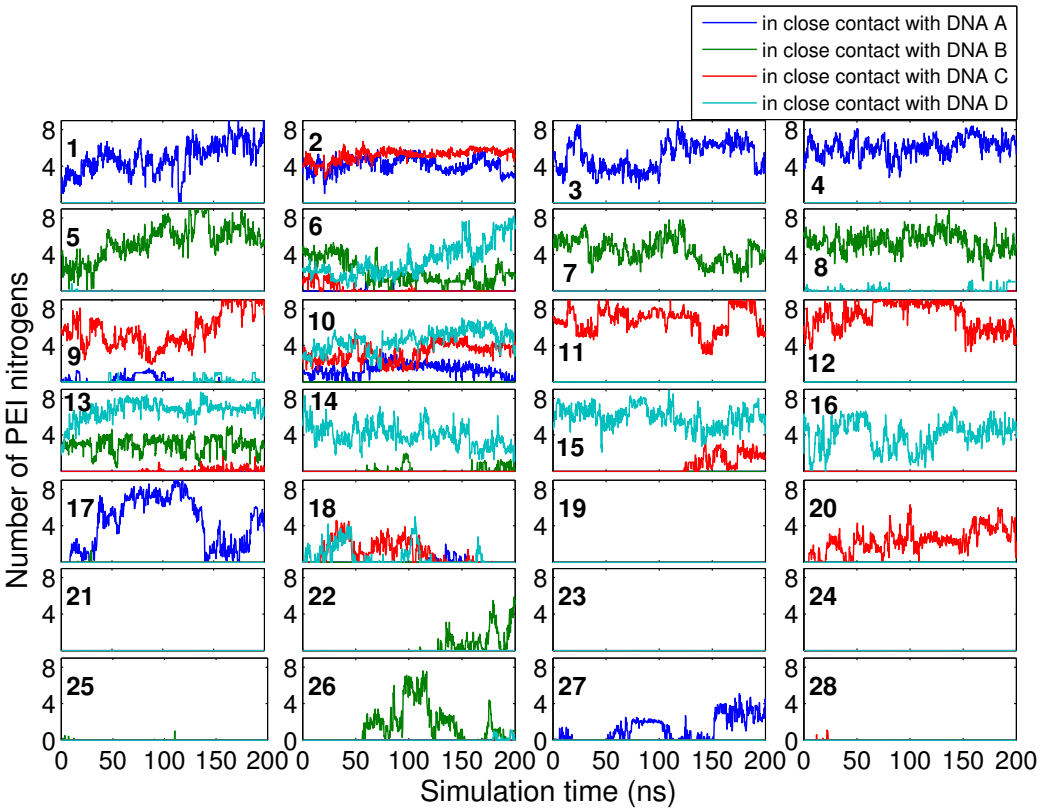

Figure D.3: Number of nitrogens for each lmPEI that are within 4 Å of any N/O atom of each DNA as a function of the simulation time for the 4D-28P system.

Table D.1: Number of water molecules within 3 Å of the solute at the late stage of each system

System	Time window/entire time	No. of waters
DNA	last 5 ns / 20 ns	330.8
lmPEI	last 2 ns / 6 ns	99.4
native PEI	last 2 ns / 6 ns	77.8
4D-16P (lmPEI)	last 40 ns / 100 ns	1911.1
4D-16P (native PEI)	last 40 ns / 130 ns	1773.6
4D-28P (lmPEI)	last 40 ns / 200 ns	2700.9
lmPEI 19,21,23,24 in 4D-28P (lmPEI)	last 40 ns / 200 ns	337.0
4D-28P (native PEI)	last 40 ns / 200 ns	2657.5

individual molecules when they are separated. For example, for system 4D-16P with lmPEIs, the number of water molecules is calculated as '$330.8 \times 4(DNAs) + 99.4 \times 16(lmPEIs) - 1911.1 = 1002.5$'. For system 4D-28P with lmPEIs, as lmPEIs 19, 21, 23, 24 are associated in the solution and not bound to the polyplex, we subtract the number of released water molecules resulting from the association of these four lmPEIs from the number of released water molecules from the whole system.

D.4 Radii of gyration of the DNAs in systems 4D-16P and 4D-28P and comparison with their counterparts involving native PEIs

Figure D.4 shows the radii of gyration R_g of the four DNAs as a function of simulation time in systems 4D-16P (lmPEI), 4D-16P (native PEIs), 4D-28P (lmPEI) and 4D-28P (native PEIs). It can be seen that R_g of the four DNAs aggregated by lmPEIs in 4D-16P (average R_g over last 40 ns $= 25.7$ Å) is nearly identical to that of the four DNAs aggregated by 16 native PEIs (average R_g over last 40 ns $= 26.3$ Å). So is R_g of the four DNAs in 4D-28P with lmPEIs (average R_g over last 40 ns $= 24.6$ Å) compared with R_g of the four DNAs aggregated by 28 native PEIs (average R_g over last 40 ns $= 24.9$ Å).

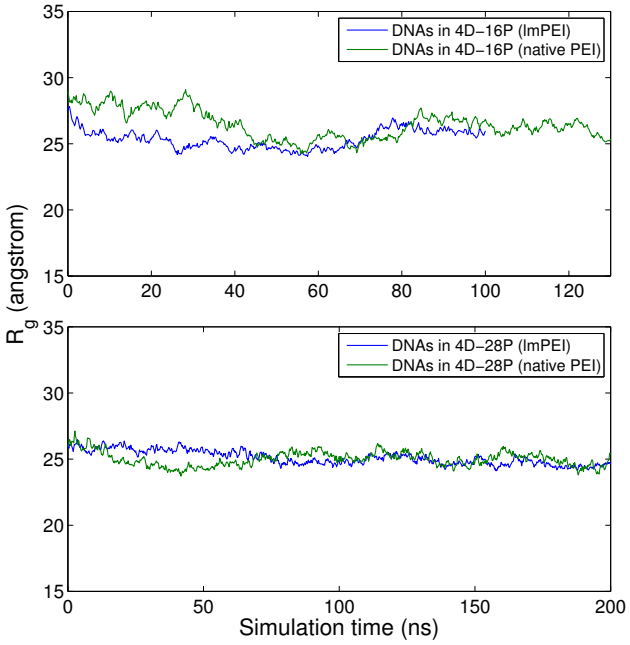

Figure D.4: Radius of gyration of the four DNAs in each system as a function of simulation time.

D.5 DNA-DNA spacing

DNA-DNA spacing is an important parameter to gauge how compact the DNAs are condensed. To investigate the DNA-DNA spacing in the aggregates, we have defined the 'shortest distance' and 'root mean square (RMS) distance' between two DNA molecules in Ref. 2. We first represent each DNA as a series of points each being the COM of a Watson-Crick DNA base pair. For each dodecamer studied in this work, there are 12 such points, and connecting neighboring points results in 11 segments. For a pair of segments from two different DNAs, we can calculate their shortest distance, and there are 121 such distances ($d_i, i = 1 \ldots 121$) between all pairs of segments from the two DNA dodecamers. We defined the minimum of these 121 distances as the 'shortest distance' d_{shortest} and the root mean square of these 121 distances as the 'RMS distance' d_{RMS} ($d_{\mathrm{RMS}} = \sqrt{(\Sigma_{i=1}^{121} d_i^2)/121}$). Figures D.5 and D.6 plot d_{shortest} and d_{RMS} for systems 4D-16P and 4D-28P, respectively. The average values of d_{shortest} and d_{RMS} for each pair of DNAs over the last 40 ns of the simulations are summarized in Table D.2. Among the 6 pairs, the aggregation of three of them (A-C, B-D, C-D) involves direct bridging by the

same lmPEIs, while the other three (A-B, A-D, B-C) are aggregated only through the lipid association among different lmPEIs. Overall, the directly bridged DNA pairs have smaller d_{shortest} and d_{RMS}. In fact, for the (A-C, B-D, C-D) pairs, the average values are 18.5 Å for d_{shortest} and 30.1 Å for d_{RMS}, which are not significantly different from the respective values of 21.4 Å and 29.0 Å for native PEI mediated DNA aggregation. Due to steric effect, DNA pairs brought together by lipid association (A-B, A-D, B-C) show much larger d_{shortest} and d_{RMS} (except for A-B in 4D-16P system which has slightly smaller d_{RMS} than that of C-D).

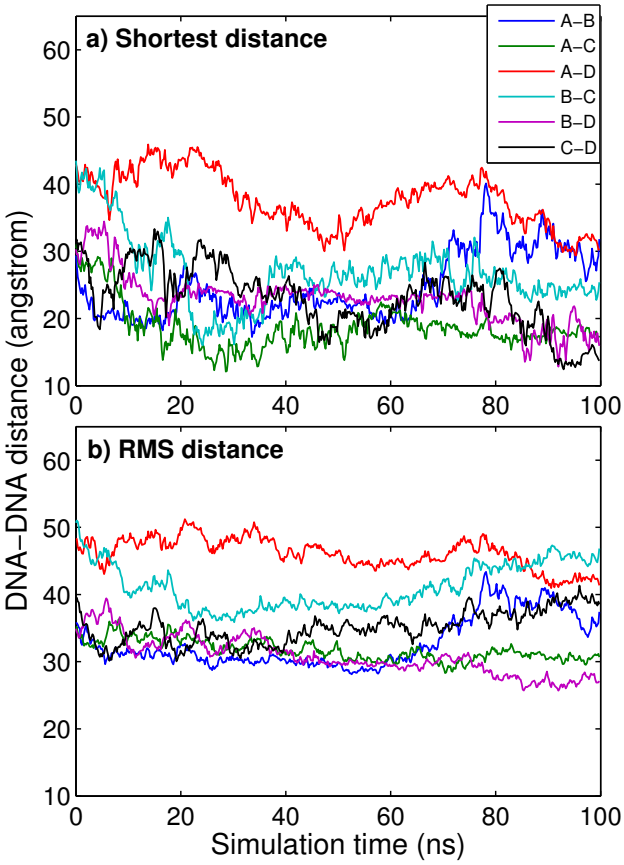

Figure D.5: Distance between DNAs for the 4D-16P system. (a) Shortest distance. (b) Root mean square distance.

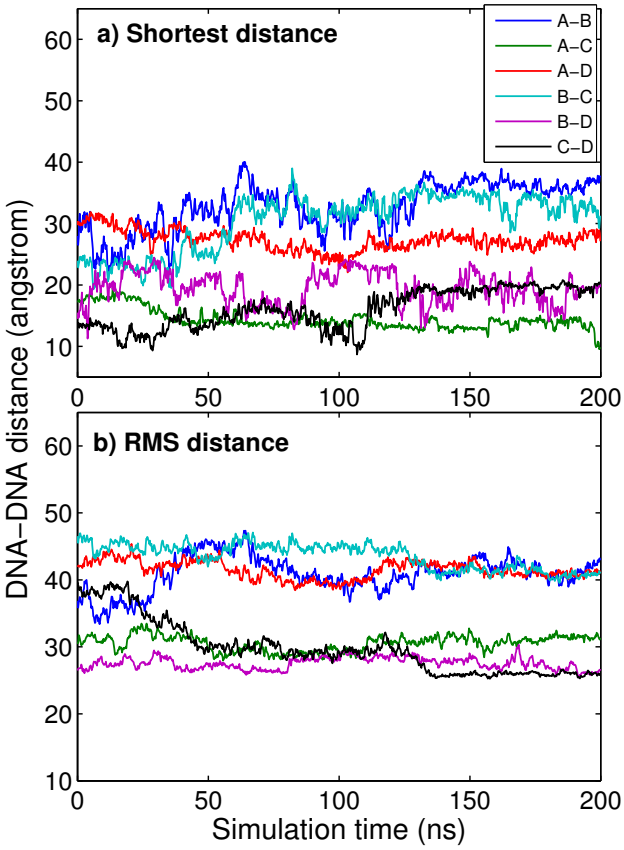

Figure D.6: Distance between DNAs for the 4D-28P system. (a) Shortest distance. (b) Root mean square distance.

Table D.2: DNA-DNA distance (Å) averaged over the last 40 ns of the simulations.

System	4D-16P						4D-28P					
DNA-DNA	A-B	A-C	A-D	B-C	B-D	C-D	A-B	A-C	A-D	B-C	B-D	C-D
d_{shortest}	29.0	18.2	36.5	26.0	20.3	20.8	36.1	13.6	27.2	32.8	18.5	19.7
d_{RMS}	36.3	30.7	44.7	43.3	28.4	37.0	41.5	31.2	41.0	41.1	27.1	26.0

Bibliography

[1] C. Sun, T. Tang, H. Uludag, J. E. Cuervo, *Biophys. J.* **2011**, *100*, 2754--2763.

[2] C. Sun, T. Tang, H. Uludag, *Biomacromolecules* **2011**, *12*, 3698--3707.

Appendix E
Supporting Information for Chapter 6

E.1 Calculation of water release

In order to calculate the number of water molecules released during the aggregation process, we counted the number of water molecules within 3 Å of the solutes in each system as summarized in Table E.1. To determine the number of water molecules released for a particular system, we subtract the number of water molecules within 3 Å of the polyplex from the total number of water molecules within 3 Å of the individual molecules when they are separated. For example, for system 4R-18P, the number of released water molecules is calculated as '555.8×4 (4 siRNAs) $+227.6 \times 18$ (18 P2ks) $-4743.1 = 1577$'.

E.2 Binding of individual PEIs to each siRNA

We define 'a PEI N is in close contact with the siRNA' if it falls within 4 Å of any N/O atoms of the siRNA. A PEI is defined to be 'bound' to a siRNA molecule if it has one or more Ns in close contact with this siRNA. If a PEI is 'bound' to two or more siRNA molecules simultaneously, we say that this PEI form a polyion bridge between the siRNAs. To quantify the polyion bridging the PEIs form among siRNA molecules, we plotted the binding state of individual PEIs to each siRNA in terms of the number of Ns from each PEI in close contact with each siRNA, as shown in Figures E.1 to E.5 for systems 4R-18P, 4R-18P-1CA, 4R-18P-3CA, 4R-18P-1LA and 4R-18P-3LA, respectively. Each subfigure in Figures E.1 to E.5 corresponds to one of the 18 PEIs in a particular system, and contains 4 curves each of which describes the number of Ns of this PEI in close contact with a particular siRNA. For system 4R-18P shown in Figure E.1, out of the 18 PEIs, seven PEIs (1, 2, 3,

Table E.1: Number of water molecules within 3 Å of the solute at the late stage of each system

System	Time window / Entire simulation time	No. of waters
siRNA	last 40 ns / 100 ns	555.8
P2k	last 20 ns / 50 ns	227.6
P_1CA	last 20 ns / 40 ns	237.5
P_3CA	last 20 ns / 40 ns	258.3
P_1LA	last 20 ns / 40 ns	252.3
P_3LA	last 20 ns / 40 ns	254.6
4R-18P	last 50 ns / 200 ns	4743.1
4R-18P_1CA	last 50 ns / 200 ns	4909.2
4R-18P_3CA	last 50 ns / 200 ns	4961.8
4R-18P_1LA	last 50 ns / 200 ns	4903.9
4R-18P_3LA	last 50 ns / 200 ns	4752.2

4, 5, 6 and 8) participate in bridging two or three siRNAs for longer than 50% of the simulation time, where PEI 2 is bridging the four siRNA for most time of the simulation. For systems 4R-18P_1CA, 4R-18P_3CA, 4R-18P_1LA and 4R-18P_-3LA, there are respectively 7, 8, 9 and 8 PEIs participating in bridging two or three siRNAs for longer than 50% of the simulation time.

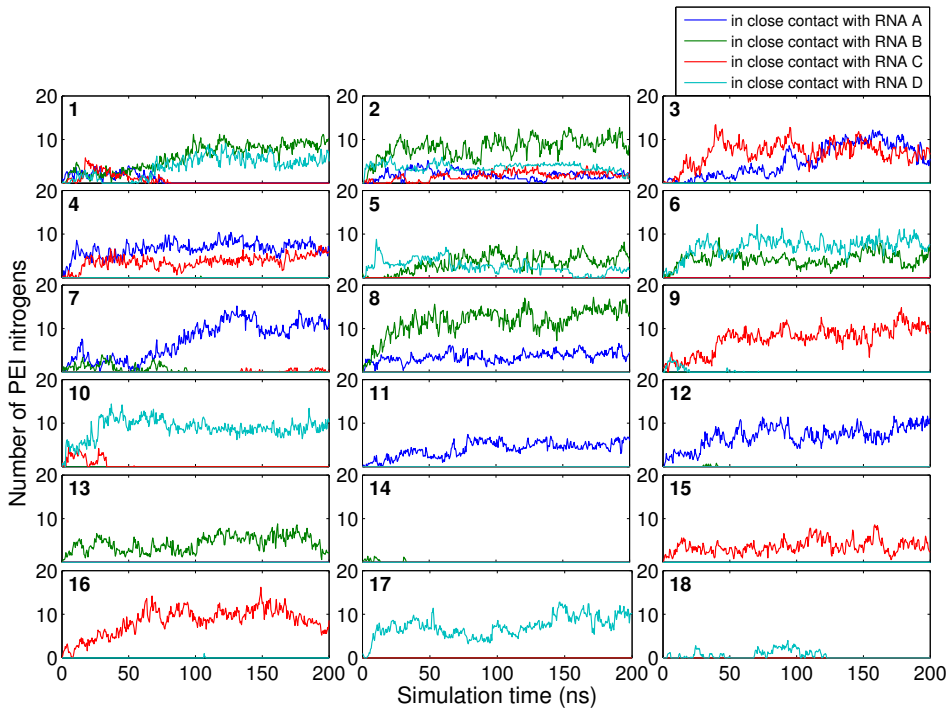

Figure E.1: Number of nitrogens for each PEI that are within 4 Å of any N/O atom of each siRNAs as a function of simulation time for system 4R-18P.

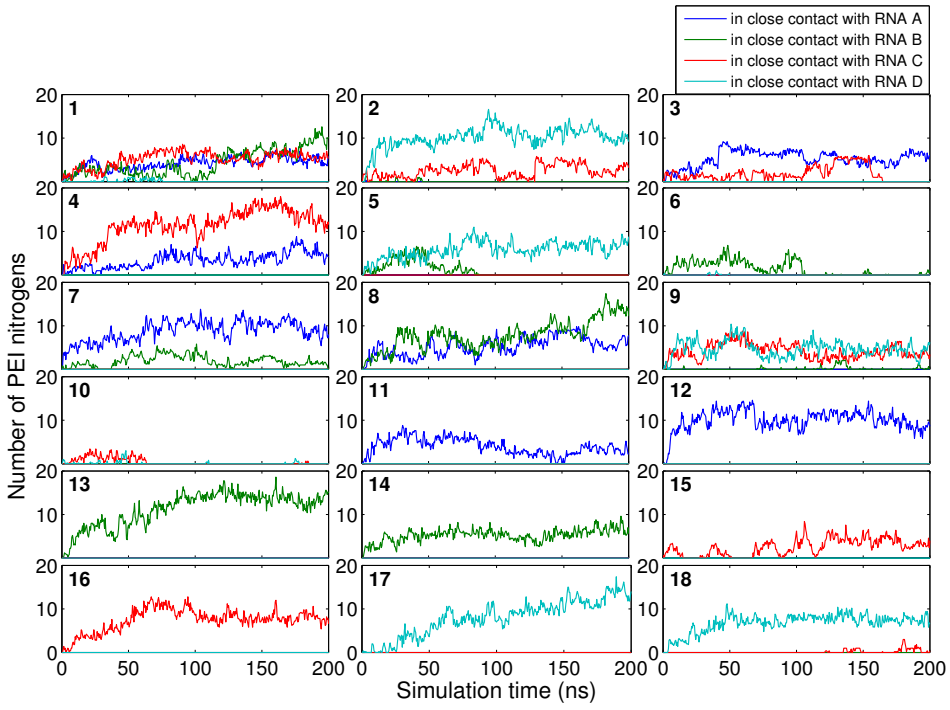

Figure E.2: Number of nitrogens for each PEI that are within 4 Å of any N/O atom of each siRNAs as a function of simulation time for system 4R-18P_1CA.

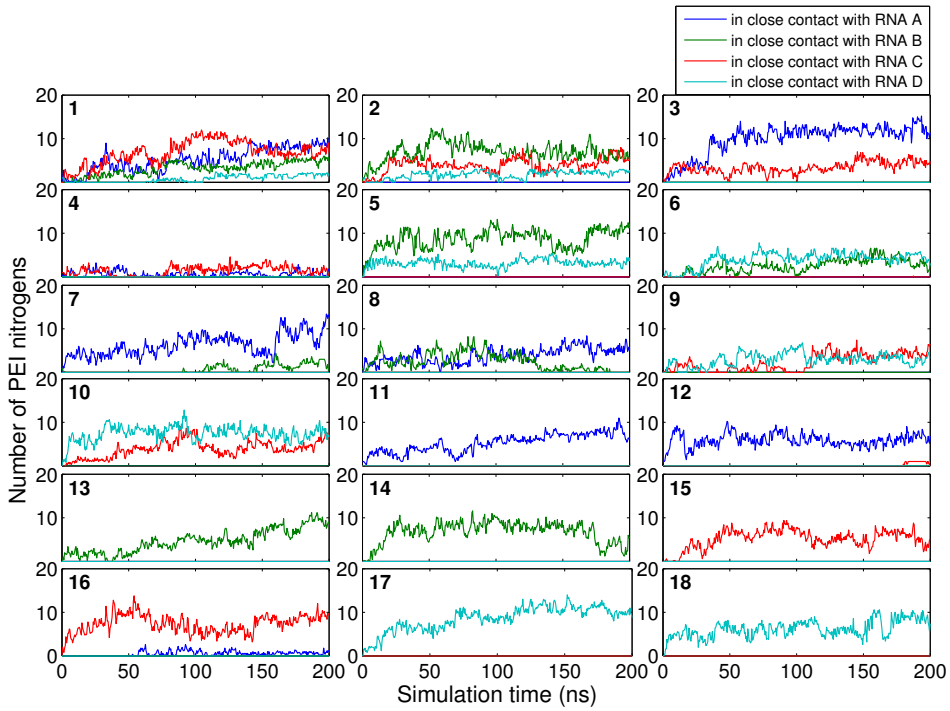

Figure E.3: Number of nitrogens for each PEI that are within 4 Å of any N/O atom of each siRNAs as a function of simulation time for system 4R-18P_3CA.

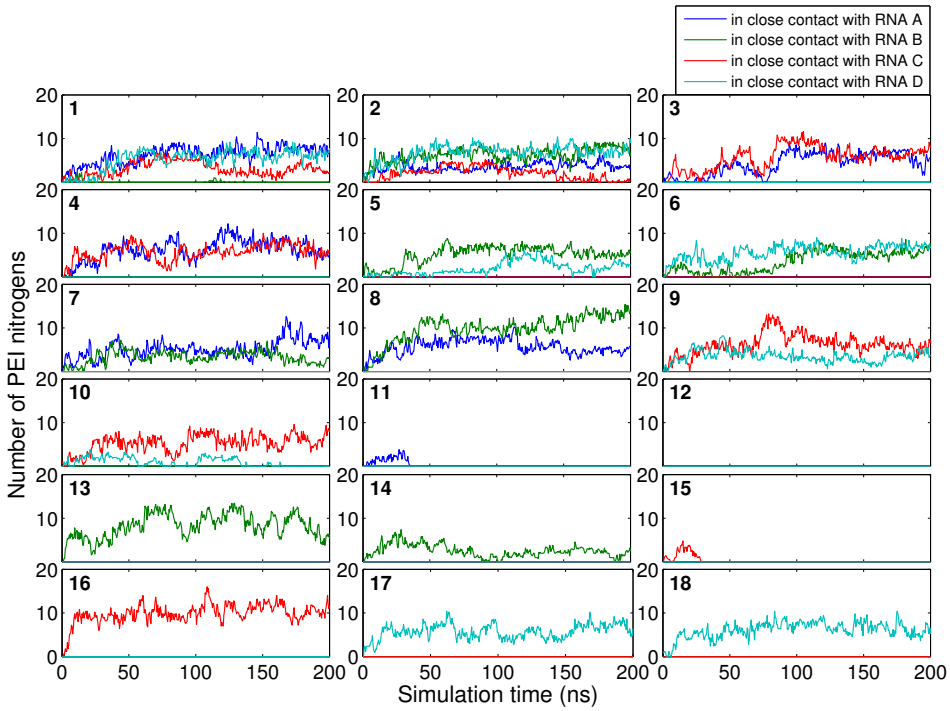

Figure E.4: Number of nitrogens for each PEI that are within 4 Å of any N/O atom of each siRNAs as a function of simulation time for system 4R-18P_1LA.

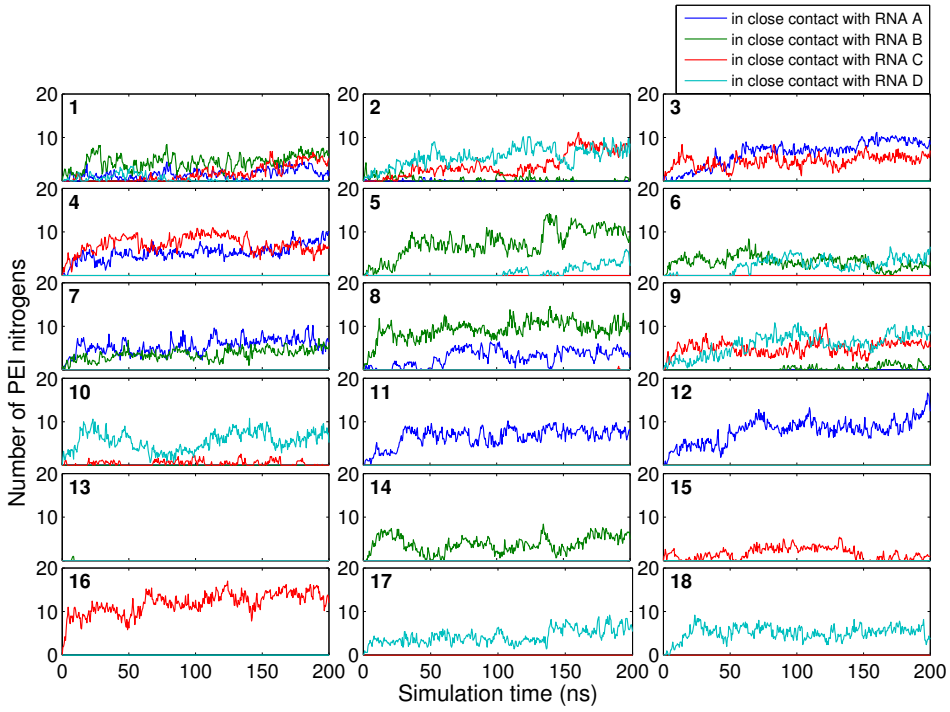

Figure E.5: Number of nitrogens for each PEI that are within 4 Å of any N/O atom of each siRNAs as a function of simulation time for system 4R-18P_3LA.

E.3 Evidence of convergence of the simulation trajectories

Figure E.6 shows the charge neutralization curves (cumulative net charge of PEI and ions as a function of the distance from any siRNA C1' atom) based on four time windows (last 20 ns, 2nd last 20 ns, 3rd last 20 ns and 4th last 20 ns) at the late stage of the simulations. It can be seen that the curves almost overlap with one another, demonstrating the convergence of the simulation trajectories.

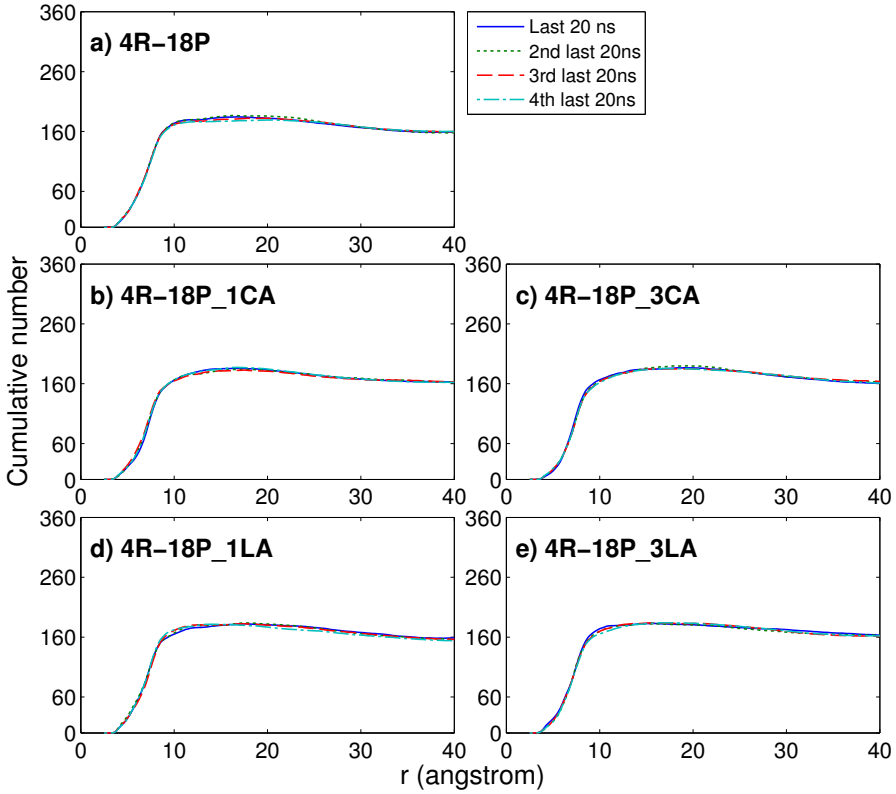

Figure E.6: Cumulative numbers of net charge of PEI and ions as a function of the distance from any siRNA C1' atom based on four time windows (last 20 ns, 2nd last 20 ns, 3rd last 20 ns and 4th last 20 ns) during the last 80 ns of the simulations. (a) 4R-18P, (b) 4R-18P_1CA, (c) 4R-18P_3CA, (d) 4R-18P_1LA, (e) 4R-18P_3LA.

Appendix F
Additional Details of the Simulations

In constructing the molecular structure of the PEIs, we split the PEI into a series of building blocks each of which corresponds to a newly created residue in the developed force field rtf file. Take the 2 kDa branched PEI in Chapter 3 (the most complicated native PEI structure simulated in this dissertation) for example, as shown in Figure F.1, we split the PEI into 16 residues along one chain. Because that some residues are identical, only eight distinct residues are required to build this PEI. This approach provides us a great flexibility to reuse the residues in building other PEIs. For example, in building the lipid-modified PEIs in Chapter 6, we only needed to create four new residues with lipid moieties based on residues 4 and 14 shown in Figure F.1 and reuse the other residues.

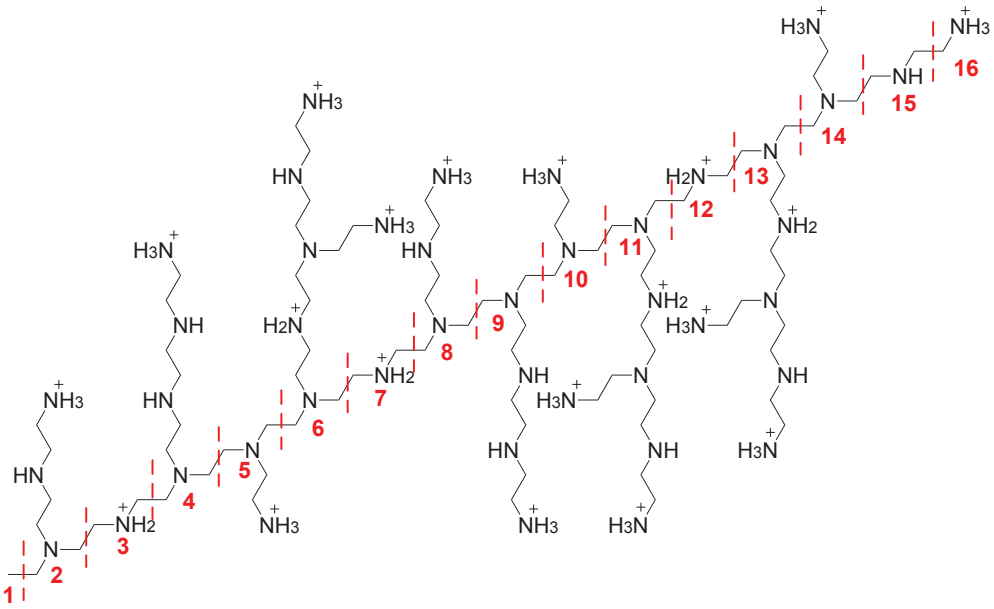

Figure F.1: 2 kDa branched PEI simulated in Chapter 3. The PEI is decided into 16 residues by the red dashed lines along one chain. Residues 2 and 8 are identical, so are residues 3, 7 and 12; residues 4 and 9; residues 5, 10 and 14; residues 6, 11 and 13.

The force field parameters for native PEIs were adopted mainly based on residues ‘‘EAMM”, ‘‘DMAM”, ‘‘TMAM” in the the CHARMM force field [1] following the CHARMM General Force Field methodology [2]. The force field parameters for the lipid moieties on the lipid-modified PEIs simulated in Chapters 5 and 6 were adopted mainly based on residue ‘‘OLEO (oleic acid)” in the CHARMM force field [1].

The production runs of all the simulations were performed with Langevin dynamics implemented in NAMD as a thermostat to keep the temperature constant at 300 K [3]. The barostat used in all the simulations is a modified Nosé-Hoover method in which Langevin dynamics is used to control fluctuations [4, 5] as implemented in NAMD [3].

Below is a sample NAMD configuration file used in simulating system 4R-18P in Chapter 6:

```
#############################################
## JOB DESCRIPTION                         ##
#############################################
# heating of RNA in a Water Box
set fs             0
#############################################
## ADJUSTABLE PARAMETERS                   ##
#############################################
set X               119.9
set Y               119.9
set Z               119.9
set CX               0
set CY               0
set CZ               0

set PX            120
set PY            120
set PZ            120
```

```
structure          4rna-18bp.psf
coordinates        4rna-18bp.pdb

set consfileName   4rna-18bp.pdb
set temperature    300
set outputname     equil/4rna-18bp_$fs

firsttimestep      $fs

#####################################################
## SIMULATION PARAMETERS                           ##
#####################################################

# Input
paraTypeCharmm      on
parameters          par_all27_prot_na_pei_lipid.prm
temperature         0

# Force-Field Parameters
exclude             scaled1-4
1-4scaling          1.0
cutoff              10
switching           on
switchdist          8.5
pairlistdist        12

# Integrator Parameters
timestep            2.0
rigidBonds          all
nonbondedFreq       1
```

```
fullElectFrequency  2
stepspercycle       10

# Constant Temperature Control
langevin            on
langevinDamping     10
langevinTemp        $temperature
langevinHydrogen    off

# Periodic Boundary Conditions
cellBasisVector1     $X    0.   0.
cellBasisVector2     0.    $Y   0.
cellBasisVector3     0.    0    $Z
cellOrigin           $CX   $CY  $CZ

wrapAll             on
wrapWater           on
wrapNearest         off

margin                1

# PME (for full-system periodic electrostatics)
PME                 yes
PMEGridSizeX        $PX
PMEGridSizeY        $PY
PMEGridSizeZ        $PZ

# Constant Pressure Control (variable volume)
useGroupPressure      yes
useFlexibleCell       no
useConstantArea       no
```

```
langevinPiston        on
langevinPistonTarget  1.01325
langevinPistonPeriod  200.
langevinPistonDecay   100.
langevinPistonTemp    $temperature

# Output
outputName          $outputname
binaryoutput        off
restartfreq         5000
dcdfreq             5000
xstFreq             5000
outputEnergies      5000
outputPressure      5000

#####################################################
## EXTRA PARAMETERS                                ##
#####################################################

fixedAtoms          on
fixedAtomsForces    on
fixedAtomsFile      $consfileName
fixedAtomsCol       B

# Harmonic restraints
constraints         on
consref             $consfileName
conskfile           $consfileName
conskcol            B
```

```
#############################################
## EXECUTION SCRIPT                        ##
#############################################

minimize            0

# Turn off until later
langevinPiston      off

# Minimize with atoms fixed
minimize            2000

# Minimize all atoms
fixedAtoms          off
constraintScaling   10
minimize            2000

constraintScaling   0
minimize            1000

# heat with restrained
constraintScaling     10.0
set temp              30
while { $temp <= 300 } {
langevinTemp          $temp
run                   1000
set  temp             [expr $temp + 30]}

langevinPiston        on
run                   5000000
```

```
constraintScaling     0
run                   100000000
```

Bibliography

[1] A. MacKerel Jr., C. Brooks III, L. Nilsson, B. Roux, Y. Won, M. Karplus, *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*, Vol. 1 of *The Encyclopedia of Computational Chemistry*, P. v. R. Schleyer et al. (Ed.), John Wiley & Sons: Chichester, **1998**, pp. 271--277.

[2] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. MacKerell, Jr., *J. Comput. Chem.* **2010**, *31*, 671--690.

[3] J. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. Skeel, L. Kale, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781--1802.

[4] G. Martyna, D. Tobias, M. Klein, *J. Chem. Phys.* **1994**, *101*, 4177--4189.

[5] S. Feller, Y. Zhang, R. Pastor, B. Brooks, *J. Chem. Phys.* **1995**, *103*, 4613--4621.