University of Alberta

Potential of Mean Force of Polyethylenimine Mediated DNA Attraction

by

Sampada Bagai

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

Master of Science

Department of Mechanical Engineering

©Sampada Bagai

Spring 2013
Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

Dedicated to my family, my teachers and my soul mate

Abstract

The aggregation of DNA molecules induced by cationic polymers is of importance to applications in gene delivery. In this work, we performed a series of umbrella sampling molecular dynamics simulations to calculate the potential of mean force (PMF) between two DNA molecules in the presence of polyethylenimine (PEI) molecules using the weighted histogram analysis method. The distance between the centers of mass of the two DNAs was chosen as the reaction coordinate, and the location and depth of the global minimum in the PMF curve were used to gauge the compactness and stability of the formed aggregate. The effects of the PEI to DNA charge ratio (N/P charge ratio), protonation state of the PEI and lipid modification of PEI were investigated. Compared with small multivalent ions, PEIs give rise to stronger DNA attraction. The DNA aggregation was found to be more favorable at higher N/P charge ratios and higher PEI protonation ratios, with the depth of the PMF well more strongly influenced by the N/P charge ratio. The lipid substitution in the PEI molecules promotes the formation of more stable aggregates in case of long lipids but not in the case of short lipids.

Acknowledgements

The duration of my Masters study in University of Alberta has been the most valuable experience of my life. I would like to gratefully acknowledge the enthusiastic supervision of my supervisor, Dr. Tian Tang, for her continuous guidance and inspiration. The weekly meetings with her gave me a much needed analysis of the progress made in my research. I particularly remember the meetings in the beginning of my Masters in which she took time to motivate me, which gave me much needed boost to take the leap from undergraduate studies to research. Apart from educating me on different aspects of my research, she helped me develop the professional qualities to be a good researcher. Her dedication towards work and her students has always inspired me to be as dedicated as her and be a better person.

I would also like to thank Dr Rong Long and Dr. Hasan Uludag for reviewing this work. I would also like to acknowledge the financial support provided by National Science and Engineering Research Council of Canada, Alberta Innovates Technology Futures, and Canada Foundation for Innovation.

I am grateful to all the Mechanical Engineering department office staff, IT personnel, Compute Canada and the high performance computing facility at the National Institute for Nanotechnology, Edmonton, Canada for all the technical and computational support.

I am also thankful to Chongbo Sun for his continuous guidance and support during my research and degree program. He was always available to answer questions on almost any topic related to my research. I would also like to take this opportunity to thank my colleagues Cuiying, Waqas, Chongbo, Tamran, Roya and Mohammad for providing a pleasant work environment.

I am extremely lucky to have Abhishek Saxena as my partner. Even being far apart, he has always been by my side. Without his constant support and affection, my studies and stay in Canada would have been extremely difficult. I am also grateful to all my friends Prashant, Olena, Stella, Dilpreet, Sparshi and Manu for a wonderful time.

Last but definitely not the least; my biggest gratitude is towards my parents, Mr. Vijay Kumar Bagai and Mrs. Vandana Bagai. They have always trusted me and supported every decision of mine, even sending me far away from them for my studies, without doubting my capability. Their constant hardwork, sacrifice and prayers have made me what I am today. I would also like to thank my younger brother Nikher for being a friend and giving me much needed breaks by his loving and notorious activities. I would like to thank my grandparents for being proud of me and giving me their blessings.

Contents

List of Tables

List of Figures

Nomenclature

Write names of symbols assumed on your own

Symbols

$\boldsymbol{q}$	Coordinates
$\boldsymbol{p}$	Momentum
H	Hamiltonian
K	Kinetic energy
U	Potential energy
N	Number of particles
k_B	Boltzmann constant
T	Temperature
h	Planck's constant
β	$1/k_bt$
F	Helmholtz free energy
V	Volume
E	Internal energy
S	Entropy
M	Mass
a	Acceleration
r	Distance between two particles
$\acute{\varepsilon}, \sigma$	Lennard Jones parameters
$\emptyset$	Electrostatic potential
ε	Dielectric constant
d	Charge on the particle

W	Potential of mean force
x	Reaction coordinate
P	Probability distribution function
U_{bias}	Biasing potential
Δ	Initial distance between centers of mass
k	Force constant
N_i	Total number of counts in a simulation
n	Number of counts in a bin of a particular simulation
$\mathbb{Z}$	Partition function
δ	Delta function
c	Biasing factor
m	Total number of bins
w	Weighting factor
s	Total number of simulations
var	Variance
$<>$	Average
$\overline{p}$	Probability of a bin
λ	Lagrange multiplier
$\overline{F}$	Change in free energy
$\mathring{A}$	Angstrom
Cl	Chlorine
Na	Sodium

Superscript

0	Unbiased state

Subscript

i	Simulation number
0	Unbiased system
j	Bin number

Abbreviations

DNA	Deoxyribonucleic Acid
PEI	Polyethylenimine
µm	Micrometer
ATP	Adenosine Triphosphate
RNA	Ribonucleic Acid
nm	Nanometer
HMW	High Molecular Weight
LMW	Low Molecular Weight
N/P	PEI To DNA
Da	Dalton
kDa	Kilo Dalton
PMF	Potential Of Mean Force
MD	Molecular Dynamics
WHAM	Weighted Histogram Analysis Method
PBC	Periodic Boundary Conditions
PME	Particle Mesh Ewald
FEP	Free Energy Perturbation

MC	Monte Carlo
MH	Multiple Histogram
MM-PBSA	Molecular Mechanics/Poisson Boltzmann Surface Area
LIE	Linear Interaction Energy
COM	Centre Of Mass
US	Umbrella Sampling
LA	Linoleic Acid
OA	Oleic Acid
CA	Caprylic Acid
pdb	Protein Data Bank
psf	Protein Structural File
TIP3P	Transferable Intermolecular Potential 3P
ps	Picoseconds
kcal	Kilo Calorie
PLL	Poly-L-Lysine
siRNA	Small-Interfering RNA

Chapter 1

Introduction

This work is motivated by the considerable interest in DNA condensation and its role in gene therapy. Gene therapy requires gene delivery vectors which can efficiently transfer the gene inside the cell. This chapter starts with an introduction to gene therapy and DNA condensation. The first section explains how DNA is packed inside a cell and how DNA condensation occurs *in-vitro*. This is followed by a review of the theoretical studies on the mechanisms governing DNA condensation in the presence of cationic ions in a controlled environment and also on the polycation mediated DNA aggregation. This chapter also introduces a cationic polymer, polyethylenimine, which has been identified as an effective gene carrier. Lastly, the chapter provides an outline of the thesis.

1.1 Introduction to DNA Condensation

The phenomenon of transforming the DNA molecule from an extended state into a compact and more ordered state is known as DNA condensation [1]. DNA molecule is the source of all the genetic information in cells and is generally a large molecule, for example the T4 phage DNA has ~160,000 base pairs occupying a length of ~54 μm [1]. This molecule has to fit in a capsid of about

100 nm in diameter; which translates into a linear compression of 540 times [1]. An Adenosine Triphosphate (ATP) driven process translocates and condenses DNA into a highly compact state, almost like a crystalline structure [2, 3].

DNA condensation has been of interest to researchers in various fields as it explains how a great amount of genetic information is packed inside highly dense and compact capsid or nuclei [1]. DNA condensation is also the pre-requisite for gene therapy [4], therefore, the condensation of DNA into compact forms is directly correlated to the efficacy of gene delivery. Gene therapy offers a potential treatment for cancer and other hereditary diseases by raising vaccination to a higher level [5]. It is based on insertion of genes into cells to eliminate the root cause of disease. The addition of such a foreign gene allows the cells to work in the same way as they were meant to do before they were initially defected. The discovery that gene could be transferred within the nucleic acids and the foreign genes can permanently become a part of the genome has provided a base to scientists for the invention of therapeutic remedies [6]. In order to transfer the gene inside the cell genome, a carrier is required. It was found that viruses can act as carriers for foreign genes by infecting the cells, thereby, transferring the gene inside the cell; but the viruses as gene carriers can be unsafe for human. Other non-viral gene carriers, including polymers, were then developed as a substitute to the viruses.

In nature, a large DNA is condensed and compacted to fit inside the nucleus; this condensation takes place due to many factors including chromosomal domains, polyamines, histone proteins and the space competition with RNA and proteins [7, 8, 9]. The study of DNA condensation began with the study of the structure and morphology of DNA inside a capsid [2, 10]. This was facilitated first by X-Ray diffraction and then by electron microscopy, and the initial studies suggested that DNA was packed in a toroidal structure inside a phage head [11]. Richards *et al.* [12] suggested that the condensed DNA was wound in a uniform curvature. Earnshaw *et al.* [12] proposed two models for the toroidal structure of DNA condensates, namely "ball of string" and "co-axial spool". In the "ball of string" model, the DNA is wrapped around a spherical ball with any radius and any order. The co-axial spool on the other hand was an ordered model, wherein the DNA was supposed to be wound around a common axis in concentric circles. The latter is the most commonly supported model [10]. Cerretelli *et al.* [2], in their experiment using cryo-electromagnetic imaging, showed that there can be multiple concentric DNA wounds inside a capsid.

Recently, the factors affecting DNA condensation have been of interest in various studies due to the significance of DNA condensation in gene delivery. The observation that DNA condensation can occur in the presence of some condensing agents, like cationic ions, has been confirmed in various experimental settings [13, 14]. Gosule *et al.* [13] demonstrated that adding spermidine in a test tube with low ionic strength (0.001 M) aqueous buffer could cause spontaneous DNA

condensation. In an experiment by Wilson *et al*. [14], it was shown that DNA was condensed from its extended coil form upon the addition of trivalent or tetravalent cations. Widom *et al.* [15] in their experiment, proved the capability of cobalt hexamine to induce DNA condensation.

Not all types of cationic ions have the ability to condense DNA in aqueous solution [16]. It was shown nearly three decades ago that multivalent cations are capable of condensing DNA in aqueous solutions while monovalent and divalent ions lack this ability [17, 18]. Raspaud *et al.* [19] demonstrated the aggregation of DNA by multivalent ion (spermidine) in four different states of DNA. Chattoraj *et al.* [20], in their experiment using electron microscope, observed that the formation of round and compact DNA molecules was caused due to the presence of spermidine molecule. Cationic polymers have also proven to be successful in the condensation of DNA [21, 22]. Hansma *et al.* [23], observed the formation of toroids and short rods with a contour size of 300 nm using atomic force microscope when polylysine (PLL) was added to a solution with DNA molecules. Different polycations have been identified as potential gene carriers and condensing agents by careful study of DNA-polycation complex formation.

1.2 Theoretical studies on DNA Condensation

Some theoretical work has been done in identifying the impact of thermodynamic and kinetic factors on DNA aggregation [24], but there is still a lot to discover

[25, 26, 27, 28]. Various thermodynamic energies contribute to the energy changes involved in DNA condensation, which include free energy change due to DNA bending, entropy changes due to mixing of the condensates, energy change due to coulombic forces and energy changes due to hydration of the molecules [1].

Bending free energy is directly proportional to the length of the DNA and inversely proportional to the square of the radius of curvature [1]. The bending of DNA molecule is essential for the formation of toroids and hence, it is necessary to overcome this bending rigidity. One way to do so is to use condensing agents which are highly mobile and have a tendency to induce bending [1]. For instance, it was observed that the flexibility of DNA increases in the presence of monovalent and divalent cations at specific concentrations [29].

Entropy of Mixing: The addition of condensing agents leads to the displacement of solute particles and separation of DNA from solvent [1]. Due to this movement of solute particles, entropy is lost due to and entropy changes vary throughout the condensation process.

Coulombic Interactions: Coulombic interactions are one of the most important contributors towards the condensation process. In order for the condensation to take place, the repulsion among the negative charges on the phosphate backbone of DNA molecules needs to be shielded, which is generally done by introducing

cations [14]. Apart from the repulsion, coulombic interactions also contribute to ion fluctuations [25, 30], competition for cation binding, undercharging [1, 31] as well as overcharging [19, 25, 32] of the DNA condensates.

Hydration Forces: These forces arise as a result of the change in the distribution of water molecules around the DNA due to the interactions between the condensing agent and DNA. These forces are generally favorable towards condensation since the release of water molecules upon the binding of the condensation with the DNA is entropically favorable [1, 33].

There have been several theoretical studies on the mechanisms governing cation-induced DNA condensation and aggregation. Wilson and Bloomfield [14] stated that charge neutralization was a prerequisite for the DNA condensation process; and for condensation to occur, it was necessary that at least 90% of the charge was neutralized. Oosawa [34] proved that the addition of counterions reduced the repulsion between like-charged polyelectrolytes. It was done by calculating the repulsive forces between the two rod shaped macro ions in terms of the charge density on the rods. The study shows that the fluctuations in the surface charge density of the polyelectrolytes (with counterions bound to them) are related to the distance between the chains; producing a long range attractive force which is similar in nature to van der Waal forces but depends inversely on the square of the separation distance. Manning [35] proposed that counterions condense on a polyelectrolyte, such that the linear charge density of the polyelectrolyte gets

lowered to a certain value. He defined a parameter for the polyelectrolyte $\xi=l_B/b$ (l_B is Bjerrum length and b is distance between two neighboring charges on the backbone of the polyelectrolyte); ξ should be greater than 1 for counterions to condense on the polyelectrolyte's backbone. According to Manning [35], the charge density of the DNA in a solution is a function of the dielectric constant because of which the effect of charge neutralization caused by a specific cationic carrier on a DNA molecule might change, depending on the solution. As an example of this, we can state the work of *Wilson et al.* [14], who suggested that although divalent molecules are incapable of condensing DNA molecules in water, they can cause condensation in a solution with 50% methanol. By decreasing the dielectric constant of the solution, more cations can associate with DNA in the solution. Since it was predicted that mono- and di-valent cations are not successful in condensing the DNA and reducing the repulsion, most of the studies on DNA condensation have suggested the use of condensates with a valency of 3+ or more to neutralize DNA phosphate's negative charge [20]. While many of these works [36] demonstrated that electrostatic screening of DNA's charges by counterions could reduce the repulsion between DNAs at close separation, these studies were not able to fully explain DNA condensation.

In particular, electrostatic screening can only reduce the distance between the segments on the DNA, whereas DNA in a condensed form involves lateral contact of the segments [37]. This led to the study of other mechanisms that are not based on electrostatic effects. In the simulation work by Sun *et al*. [38], where the

aggregation of two and four short DNA molecules in the presence of PEI molecules was investigated, it was found that in addition to the electrostatic screening, *polyion bridging* (a polyion simultaneously binds to multiple DNA molecules) is an important mechanism responsible for polycation-medicated DNA aggregation [38]. The polyion bridging is a reversible temporary bond between the two DNAs that brings them closer to each other as compared to their initial separation [38]. Other mechanisms proposed to explain DNA condensation by cations also exist, such as the formation of Wigner crystal by counterions [27], zipper motif model [39] and the creation of long range attractive hydration forces because of the counterions [33]. Shklovskii [27] provided an explanation for rod like polyelectrolyte condensation in the presence of counterions and later, extended it to DNA condensation. According to their theory in a DNA molecule, the phosphate groups carrying the negative charges, are located along the two spirals and are separated by a wide groove. The counterions tend to approach the phosphate groups, thereby forming a one dimensional Wigner crystal along the spiral. A large energy per ion is released when the crystalline spots on two such DNAs touch each other [27]. The *zipper motif model* extended the interaction theory for helical molecules to explain the phenomenon of DNA condensation [39]. Specifically Kornyshev *et al.* [39] proposed that the counterions arrange themselves in the grooves between the negatively charged phosphate groups, hence forming stripes of negative and positive charges. Two DNAs align such that the oppositely charged stripes come in close contact with each other (forming a zipper), thereby creating an electrostatic attraction between the two DNA

molecules [39]. Theories based on the long range attractive hydration forces suggest that the reconfiguration of water molecules between two DNA molecules in the presence of counterions results in the attraction of the two DNA molecules. Multivalent counterions upon adsorption cause re-distribution of the water molecules, thereby, making the DNAs come together and form a hexagonal array [33].

1.3 Introduction to Polyethylenimine (PEI)

One of the cationic polymers, PEI is able to form toroidal nanoscale aggregates with DNA molecules, [40, 41] which can subsequently facilitate *in-vivo* as well as *in-vitro* delivery of oligonucleotides and plasmid DNAs [42]. The micro- and nano-scale aggregates help in the cellular uptake of the DNA and protect them from degrading in the process of gene delivery [41]. PEI was first identified by Boussif *et al.* [42] as an effective gene carrier. Apart from its non-pharmaceutical usage like water purification and shampoo manufacturing, PEI has proven to be relatively safe for internal use in humans. [43]. Compared with other candidate synthetic gene carriers, PEI has a high cationic charge density due to its structure in which every third heavy atom is a protonable amino nitrogen [42]. This gives PEI a high buffering capacity at different pH values. PEI is highly water soluble because of the presence of repeating ethylamine units. Also, PEI has the ability to transfect a wide variety of cells [44, 45], making it a potential non-viral gene delivery carrier that shows great promise in clinical settings [45]. The steps

involved in the transfer of an external gene to the cell using PEI include *binding and uptake, endosomolysis* and *nuclear entry* (Fig. 1.1) [46].

Binding and Uptake: The first step involves binding of the positively charged DNA-PEI polyplexes to the cell surface that is negatively charged. This can be generally done in two ways: *passive targeting* or *active targeting*. Active targeting relies on the recognition of specific ligands by the cell while passive targeting makes use of the enhanced permeability and retention effect due to which the polyplexes have higher tendency to pile-up around targeted cells as compared to normal cells [47]. Generally, active targeting has proven to be more successful in transferring polyplexes to different types of cell surfaces due to their receptor media uptake [48, 49, 50]. The uptake of the polyplexes is generally governed by endocytosis [51], but the exact nature of this process and the effect of various factors on the endocytosis are still not completely known. For example, Kopatz *et al.* [52] used fluorescent labeled linear PEIs to observe the uptake process of DNA cation polyplexes and suggested a phagocytosis mediated by actin filament.

Endosomolysis: This is the most critical stage, as it can lead to the destruction of the DNA by an acidic endonuclease present in lysosomes [53]. Various studies have shown that PEI has been successful in alleviating this endo-lysosomal effect [54, 55]. Behr [54] postulated "proton sponge hypothesis", which explained the role of PEI in escaping this effect. He proposed that at physiological pH, PEI has a certain number of protonable nitrogens; when it enters the endosomes, the pH

value reduces, thereby increasing the number of protonable nitrogens [54]. This generates a charge gradient causing a high concentration of chloride ions to be released. This leads to high influx of water or hydrogen ions that make the endosome to swell, burst and release DNA molecule. Thus, the polyplexes are saved from lysosomal degradation [54]. Sonawane *et al*. [55] also reported the higher efficiency of PEI over polylysine during endosomolysis.

Nuclear Entry: Once the DNA has been de-complexed from the PEI, plasmid DNA has to enter the nucleus, which is a very crucial step because of three main reasons, namely: large size of DNA molecule [56], short half-life of DNA [57] and the nucleus double membrane barrier [58]. The DNA can enter the nucleus either through nucleus pore or during mitosis when the nuclear envelope breaks [59].

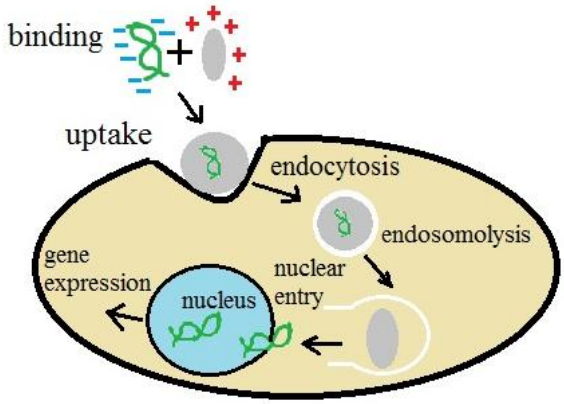


Fig. 1.1 Picture depicting the steps involved in transfer of an external gene.

Many experiments have been conducted to investigate DNA condensation and aggregation facilitated by PEI. Akinc *et al.* [60] tested the efficiency of PEIs to

transfer DNA molecules and concluded that the transfection efficiency is due to their ability to avoid acidic lysosomes. A dynamic light scattering study showed that PEI was able to aggregate small DNAs (150 base pairs), giving it an advantage over smaller cations that form aggregates only with long DNAs (at least 400 base pairs) [61]. Studies have shown that the size of the gene delivery vector is an important factor in determining the efficiency of the uptake into the cells [62, 63]. In general, it has been found that the size and stability of the nanoparticles of DNA-PEI polyplexes significantly affect the delivery and transfection process, with more compact and stable particles showing better cellular penetration and uptake [64, 65]. Dunlap *et al*. [40] compared the DNA condensates formed by PEI and lipospermine and found that the condensates formed by PEI were more compact than those formed by lipospermine.

PEIs have a wide range of molecular weights, protonation ratios and degrees of branching, and these properties affect the transfection efficiency of PEIs [40, 66]. In terms of *molecular weight*, high molecular weight (HMW) PEIs (usually 25 kDa or higher) in general have shown to yield a higher transfection efficiency compared with low molecular weight (LMW) PEIs, but HMW PEIs are also associated with high cytotoxicity that restricts their clinical usage [67, 41]. Yu *et al.* [68] in their experiment reported that on increasing the weight of a LMW PEI, the gene expression increases. Kunath *et al*. [69] compared the transfection efficiency and cytotoxicity of a LMW PEI (5.4 kDa) and a HMW PEI (25 kDa). They observed that while the transfection efficiency of HMW PEI was higher

than of LMW PEI at low N/P ratio, the difference in the zeta potential of the formed polyplexes was comparable at high N/P ratio. They also suggested that the cytotoxicity of HMW PEI was one magnitude higher than that of LMW PEI [69].

The *protonation ratio* of the PEI is the percentage of nitrogens which are protonated, and it is generally regulated by the pH value of the environment. The role of protonation ratio of PEI was investigated by Utsuno and Uludag [66], who suggested that PEI bound with DNA more rapidly in a low pH value environment when the PEI possesses a higher protonation ratio. In their simulation based study, Sun *et al.* [70] compared two protonation states of PEI, 23% and 46% protonated. They concluded that higher protonation state PEI was able to form stronger and more stable complexes with DNA. They observed that protonated nitrogen in 46% protonated PEIs interacted mainly with the electronegative oxygen at DNA backbone; whereas in the case of 23% protonated PEI, indirect interaction by water molecules played an important role [70].

The PEI molecule contains different structures based on the *degree of branching*. A branched PEI is synthesized by the catalysis of aziridine which leads to its ring opening [71] whereas linear PEI can be formed either by polymerization of aziridine at low temperature or polymerization of oxazoline [72]. Dai *et al.* [73] studied the effect of PEI structure on the cellular uptake and transfection efficiency by examining polyplexes formed by DNA with linear as well as branched PEIs. They found that branched PEI was more effective in causing DNA

condensation and resulted in better cellular uptake than its linear counterpart. On the other hand, many studies have also shown that linear PEIs have better transfection efficiency and higher cell viability as compared to branched PEIs [74]. Wightman *et al.* [75] in their experiment compared the transfection efficiency and the aggregation capability of linear and branched PEI in a salt containing buffer, both weighing 22 kDa. They observed that upon the addition of salt, the linear PEI/DNA complexes are able to aggregate much more rapidly as compared to the branched PEI/DNA complexes, which contributes to the higher efficiency of linear PEIs.

The *PEI to DNA charge ratio* (N/P charge ratio) is another important factor affecting DNA condensation [76]. It was observed that high N/P charge ratio created a net positive charge on the DNA-PEI polyplexes, improved cellular uptake and extended DNA retention in the nucleus [77]. Xie *et al.* [78] in their experiment to show the effect of PEI/DNA complexation on gene expression observed that at higher N/P ratio, the diameter of the complexes formed was smaller. Although increasing the N/P charge ratio increases the transfection efficiency, it increases the cytotoxicity as well [79].

Although PEI is capable of forming stable and compact DNA condensates, as stated earlier, it has been observed that LMW PEI are not as effective as HMW; and HMW on the other hand are more toxic. It has been found that substituting LMW PEIs with lipids increase their efficiency, making the condensates more

stable [80, 81]. Incani *et al.* [82] observed that lipid substitution of poly-lysine resulted in improved efficiency in delivering DNA and also protected DNA from damaging. Bahadur *et al.* [83] substituted different molecular weight PEIs with palmitic acid and observed that the lipid substitution increases the zeta potential of the condensates. Chen *et al.* [84] in their experiment compared native 25 kDa PEI to PEI substituted with lauric acid and thioctic acid, and found that substitution with acid increases the transfection efficiency. Neanmark *et al.* [85] in their experiment observed that even one lipid substitution on 2kDa PEI makes it more efficient than native 25kDa PEI. Sun *et al.* [86] used molecular dynamics simulations to study the role of lipid substitution on PEI mediated DNA condensation. They observed that lipids tend to associate with each other, thereby stabilizing the DNA aggregate [86]. Posocco *et al*. [87] also observed that self-assembly of the lipids is an important mechanism favoring the stability of the polyplexes. Hsu *et al.* [88] explained that the reason of increased cellular uptake of lipid modified PEIs is the difference in the transfection pathways used by them and that used by native PEIs. Although the lipid substitution of PEIs increase the stability of the formed polyplexes, Patel *et al.* [89] observed that the steric hindrances of the lipids do not allow them to form compact toroidal DNA structures. Therefore, the role of lipids in affecting PEI-based DNA carriers is still unclear and needs further investigation.

1.4 Outline of the thesis

Despite the great interest in polycation mediated DNA condensation and aggregation, the quantification of DNA-DNA attraction in the presence of polycations, to the best of our knowledge, is still absent. Calculating Potential of Mean Force (PMF) of DNA attraction mediated by polycations such as PEI, allows us to examine the size and stability of the formed DNA-polycation aggregates and understand how they can be affected by various material/structure parameters involved in the system. These issues are of great importance in understanding the role of polycations in DNA aggregation and ultimately in polycation-based gene delivery. Recently, the development of tools like MD simulation techniques and advancement in the computational capacity has taken this research to a higher level by providing atomic details of the interactions. In this thesis, we utilize the MD simulations to quantify the spontaneity of the aggregation process. The remainder of the thesis is organized as follows:

Chapter 2: Preliminaries

This chapter provides a brief introduction to the MD methods. It also describes in detail the calculation of free energies to assess the spontaneity of the aggregation process [90]. This will include statistical mechanics explanation of calculation of free energies and different methods to evaluate free energy from MD simulation. Specifically, we will explain in detail the calculation of PMF [90] through weighted histogram analysis method [91] and umbrella sampling simulations [92].

Chapter 3: Simulation Details

This chapter gives the simulated models of the molecules (DNA, native and lipid-substituted PEI) and details of the MD simulation procedure and parameters used.

Chapter 4: Results for Native PEI mediated DNA Attraction

This chapter provides the results for the native PEI mediated DNA attraction. A total of five systems were simulated, with four of them containing 46% protonated PEIs and the last containing 23% protonated PEIs. Each system contains two DNAs but may have a different number of PEIs. The PMF for each system is calculated and the results are used to quantify and compare the effect of N/P charge ratio and protonation ratio of PEI on the attraction process. The results are also compared with interaction potential of smaller multivalent cations mediated DNA attraction.

Chapter 5: Results for Lipid-Modified PEI-Mediated DNA Aggregation

In this chapter, we studied the effect of lipid substitution on PEI mediated DNA attraction. We simulated three systems with different types of lipid substitution i.e. oleic acid, linoleic acid and caprylic acid. PMF for each of these systems was calculated and compared among each other and to that of the native PEI.

Chapter 6: Conclusions and Future Work

This chapter summarizes the results from all the calculations. It also identifies some potential areas in which this study can be extended.

Chapter 2

Preliminaries

In this chapter, an outline of the basic principles of statistical mechanics and molecular dynamics will be provided. The need for free energy calculation and different methods that are used to calculate the free energy differences will be introduced, followed by a detailed description of the Weighted Histogram Analysis Method (WHAM).

2.1 Statistical Mechanics

Statistical mechanics relates the properties of atoms and molecules observed at microscopic level to the macroscopic properties of materials. The microscopic state of a system is described by a complete set of coordinates $\boldsymbol{q}$ and momentum $\boldsymbol{p}$. The total energy of the system is represented by Hamiltonian (H($\boldsymbol{p}$, $\boldsymbol{q}$)) which is a function of $\boldsymbol{p}$ and $\boldsymbol{q}$. The Hamiltonian is the sum of the kinetic energy (as a function of momentum $\boldsymbol{p}$) and the potential energy (as a function of the coordinates $\boldsymbol{q}$) and is given by:

$$H(\boldsymbol{p}, \boldsymbol{q}) = K(\boldsymbol{p}) + U(\boldsymbol{q}) \tag{2.1}$$

where $K(\boldsymbol{p})$ is the kinetic energy while $U(\boldsymbol{q})$ is the potential energy of the system. A phase space is a hypothetical space with $\boldsymbol{p}$ and $\boldsymbol{q}$ being its axis. A state of a system that contains N particles can be described by p_1, $p_2 \ldots p_{3\mathrm{N}}$ and q_1, $q_2 \ldots q_{3N}$ and is represented by a point in the phase space (which contains information of all the N particles). For each of the N particles, there will be 3 degrees of freedom for coordinates and 3 degrees of freedom for the momentum; hence the phase space will have a dimension of *6*N.

One of the most important concepts in statistical mechanics is the ensemble. An ensemble is a theoretical model consisting of all the states of a system, considered all at once such that each of this could correspond to a real state in which the system exists [93]. For example, particular states of an isolated system which have a constant Hamiltonian at energy *E* form a microcanonical ensemble [94]. There are different types of ensembles with different macroscopic constraints. A canonical ensemble (or N*VT* ensemble) represents a closed system in which the states are at constant volume and temperatures [95]. Other important ensembles are the grand canonical ensemble (or μVT ensemble) in which the chemical potential, volume and temperature remain constant, and Isothermal-Isobaric ensemble (or N*pT* ensemble) in which the N particles are at constant pressure *p* and temperature *T* [96].

In a canonical ensemble [97], the probability of finding a system in a volume element of $d\boldsymbol{p}^{3N} d\boldsymbol{q}^{3N}$ in the phase space is given by the Boltzmann distribution:

$$P(\boldsymbol{p}^{3N}, \boldsymbol{q}^{3N}) d\boldsymbol{p}^{3N} d\boldsymbol{q}^{3N} = \frac{1}{\mathrm{Z} h^{3N} N!} e^{[-\beta H(\boldsymbol{p}^{3N}, \boldsymbol{q}^{3N})]}\, d\boldsymbol{p}^{3N} d\boldsymbol{q}^{3N}, \tag{2.2}$$

where $\beta = 1/k_B T$, k_B is the Boltzmann constant, T is the temperature, h is Planck's constant and Z is the canonical partition function given by:

$$\mathrm{Z} = \frac{1}{h^{3N} N!} \int e^{[-\beta H(\boldsymbol{p}^{3N}, \boldsymbol{q}^{3N})]}\, d\boldsymbol{p}^{3N} d\boldsymbol{q}^{3N}. \tag{2.3}$$

The partition function is a summation or integral over all states. Every ensemble has a particular partition function which can be used to find the macroscopic properties. As an example, the macroscopic Helmholtz free energy F can be calculated from canonical partition function as:

$$F = -k_B T \ln \mathrm{Z}. \tag{2.4}$$

From this Helmholtz free energy, we can calculate all other thermodynamic quantities like entropy, internal energy, specific heat etc. Any macroscopic property (say A) can also be evaluated by an ensemble average of a microscopic property (say a):

$$A = \frac{1}{Q h^{3N} N!} \int a(\boldsymbol{p}^{3N}, \boldsymbol{q}^{3N}) e^{[-\beta H(\boldsymbol{p}^{3N}, \boldsymbol{q}^{3N})]}\, d\boldsymbol{p}^{3N} d\boldsymbol{q}^{3N} \tag{2.5}$$

An important goal of computational chemistry and biology is to provide detailed molecular models to study the movements of molecules, which can't be obtained by experiments. The calculation of free energies is important in meeting the above goal as it helps us to understand the kinetics of many processes like phase transformations, folding of proteins, conformational changes in DNA etc. Free energy calculations can also be used to do a comparative study of the stability of states [98, 99]. Let us take an example in which we want to test the stability of DNA after a base flipping process. If the system has temperature *T*, volume *V* and a fixed number of particles N, then the Helmholtz free energy $F = E - TS$ is minimum at equilibrium; where *E* is the internal energy and *S* is the entropy. We can compare the stability by calculating the Helmholtz free energy for initial and final configurations [100]. If the final configuration has a lower value of Helmholtz free energy, it is more stable and the change is acceptable. The spontaneity of a process involving binding of different molecules can be gauged by calculating the binding free energy. Although computationally demanding, calculation of binding free energies has been shown to be useful in various fields including medicine and biotechnology [90].

Helmholtz free energy or other thermodynamic quantities can't be directly calculated from MD simulations as they depend on the volume occupied in the phase space [100]. This can be seen from (2.6), which gives the expression for calculation of Helmholtz free energy in a NVT ensemble.

$$F = -k_B T \ln\left(\frac{\int d\boldsymbol{p}^N d\boldsymbol{q}^N e^{-H\left(\boldsymbol{p}^N, \boldsymbol{q}^N\right)/k_B T}}{h^{3N} N!}\right), \tag{2.6}$$

where h is the Planck constant. The expression inside the natural log involves a phase space integration which practically can't be determined from computer simulations [100]. Hence, there is a need to devise methods to determine the free energy. For many decades, different methods have been proposed to solve this problem. These methods are explained in detail in section 2.3.

2.2 Molecular Dynamics Simulations

Molecular dynamics simulations help us to study the behavior of a system at molecular level [101]. According to the ergodicity hypothesis, time average over a sufficiently long period of time is equivalent to ensemble average over a sufficiently large number of microstates, using which we can relate quantities calculated from MD simulations to thermodynamic properties determined from ensemble based calculations. MD uses classical mechanics to study the motion of an atom. Each atom in an MD simulation is considered to be a sphere whose center carries a point charge. The trajectory of the atoms is obtained by integrating Newton's equations of motion

$$\boldsymbol{F}_k = M_k \boldsymbol{a}_k, \tag{2.7}$$

with

$$\boldsymbol{F}_k = -\nabla_k U(\boldsymbol{q}), \tag{2.8}$$

where $\boldsymbol{F}_k$ is the total force acting on atom k moving with acceleration $\boldsymbol{a}_k$, M is the particle's rest mass and $U(\boldsymbol{q})$ is the potential energy [102]. Integration of the complex Newton's equations is computationally demanding and several algorithms have been used to provide approximate solutions. These methods include Verlet integration [103], Beeman's algorithm [104], constraint algorithm [105, 106, 107], Symplectic integrator [108], Particle Mesh Ewald Method [109] etc. An efficient algorithm should be time reversible and reduce the amount of force calculations.

Molecular dynamics considers two main types of interactions: bonded and non-bonded. Bonded interactions are the ones which are responsible for formation and stability of one molecule like chemical bonds. In simulations these interactions are defined along with their angle, torsion, length etc. Non-bonded interactions take into account the interactions between two molecules or distant atoms which can interact with each other. Calculating these non-bonded interactions is computationally demanding and many techniques and algorithms have been formulated to approximate them with accuracy [109, 110]. The total energy in MD simulation will be due to kinetic energy and potential energy. The kinetic energy changes occur due to movements of the atoms and potential energy changes occurs due to the bonded and non-bonded interactions.

The two major non-bonded interactions are van der Waals and Electrostatic forces; and how MD treats these interactions will be explained below. Van der Waals are caused by correlations in the fluctuating polarizations of nearby particles. They are weak, temporary and short ranged interactions as compared to chemical bonds [111]. It should also be mentioned that when these atoms or molecules get very close, they experience a repulsive force which can be attributed to the fact that the structure of molecules is getting deformed. Lennard-Jones potential provides the following expression taking into consideration both the attractive and repulsive part of these interactions [112] :

$$U_{LJ} = 4\acute{\varepsilon}\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right], \tag{2.9}$$

where $\acute{\varepsilon}$ and σ are Lennard-Jones parameters depending on the interacting particles and r is the distance between two particles. The power of 12 indicates that at very small distances the potential will become highly positive which suggests repulsion at very short distances. As stated earlier, calculating non-bonded interactions is highly computationally demanding and hence a cutoff radius is used in MD simulations so that as the distance between the molecules increases above a certain value the interactions due to van der Waals forces is neglected.

Apart from van der Waals forces, electrostatic forces between two charged particles also play an important role. In MD simulations, each atom is assigned a

net charge depending on its type, bond etc. Unlike Lennard-Jones potential, electrostatic potential energy is long ranged and not short lived. The equation governing electrostatic potential energy is:

$$\emptyset_{ij} = \frac{1}{4\pi\varepsilon}\frac{d_i d_j}{r}, \tag{2.10}$$

where ε is the dielectric constant, r is the distance between two particles and d_i, d_j are the charges on the particles i and j respectively. Being pair wise interactions, the electrostatic potential energy calculations are also computationally demanding and cutoff is necessary but their long range characteristics present a need for special algorithms. Ewald summation [113] and particle mesh ewald (PME) method [109] are among the methods used to tackle the long range electrostatic interactions.

A molecular system can be very large but computing interactions of a large system is not possible, hence the system is divided into small sub systems to make the computation feasible. But a limitation of such a system would be the treatment of molecules on the surfaces which will face different forces as compared to other molecules inside the system. In order to overcome these limitations periodic boundary conditions (PBC) are used in MD simulations [114]. This can be explained by Fig. 2.1, in which the yellow box is the system being simulated and neighboring boxes are replica of this box which helps to nullify the effect of

surface properties. Although a 2 dimensional figure is provided in this work, an infinite array of such boxes is used in all three dimensions in actual MD simulations.

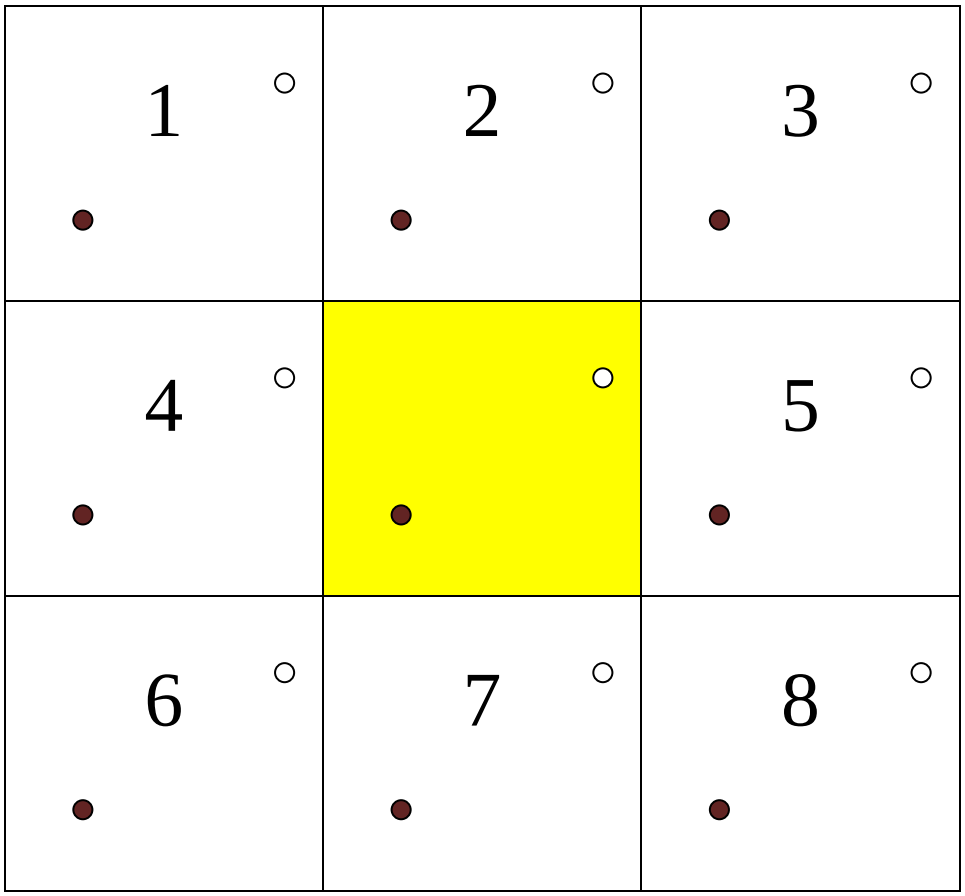


Fig. 2.1 Periodic boundary conditions representation

2.3 Weighted Histogram Analysis Method (WHAM)

The most common methods used to calculate the free energy difference are broadly divided into three categories: *thermodynamic integration*, *perturbation techniques* and *probability based methods*. These methods don't exactly calculate

the free energy of one system but they calculate the free energy difference between the present system and a reference system [100]. The present system is the system under consideration, whereas the reference system can be an ideal system e.g. ideal gas or liquid. The theoretical formulation for the calculation of free energy differences can be dated back to 1935, when Kirkwood [115] introduced the concept of order parameters to calculate free energy differences in two thermodynamic states of liquids. Kirkwood also introduced the concept of perturbation technique for the calculation of equation of liquid state [116]. Some of the earlier works included calculation of hydration energies by Mruzik *et al.* [117] and the calculation of excess free energy for water molecules by Mezei [118]. Most of these studies were based on the Thermodynamic Integration (TI) method. The derivation of TI method is pretty straightforward. For applying TI method a coupling parameter is defined which can attain a value between (0, 1). The reference state or completely decoupled state i.e. in which there are no interactions between molecules are defined by coupling parameter equal to 0 whereas the fully coupled state or final state will be defined by coupling parameter equal to 1. In most of the cases the mass of the system remains constant; hence, the kinetic energy of the system doesn't change [100] and the free energy difference of the system only depends upon the potential energy part of the Hamiltonian. The Helmholtz free energy is then differentiated w.r.t. the coupling parameter and the free energy difference between two states is obtained by summation over discreet values of coupling parameters [100]. The computational power at that time was not developed enough to apply these

techniques to practical chemistry and biology problems. In 1938, Landau applied perturbation technique to thermodynamic calculations; but it was only in 1954 when the perturbation technique was used to calculate free energies and to derive the equation of states for non polar gases at high temperature [97]. This technique later came to be known as Free Energy Perturbation (FEP) technique [119]. The FEP method was introduced by Zwanzig in 1954 for finding the high temperature equation of state for argon and nitrogen [119]. FEP method has been used for many years for systems with different complexities. From calculating the free energy of solute dissolved in water to really complex problems of protein interactions, this method has been successfully applied to various fields of computational chemistry and biology [120]. The basic concept of FEP is to find the free energy difference between an initial and a final state. The initial state is called the unperturbed state and the final state is expressed in terms of perturbation [121]. An example of unperturbed case can be an unbound ligand and enzyme which reaches to a final state of bound complex. Zwanzig [119] in his formulation of FEP chose a rigid sphere fluid as an unperturbed system while taking Lenard Jones potential to be the perturbation. The approach lined by Zwanzig [119] to find the free energy difference between two states involved expressing Hamiltonian of the final state as the sum of perturbation function and Hamiltonian of the initial state. The free energy difference is then expressed as an average over the perturbation function. As a special case of perturbation technique, Widom formulated potential distribution theorem which calculated the excess chemical potential of a system upon addition of a test particle [122]. With

the increasing complexity of problems, e.g. simulation of complex protein, nucleic, ligand molecules, scientists started switching to these perturbation techniques. To mention a few, Tembe *et al.* [123] used FEP to study ligand–receptor assemblies; and in 1985, Jorgensen *et al.* [120] estimated the excess solvation free energy of methanol and ethanol using FEP. With the advancement in simulation techniques, such as Monte Carlo (MC) and molecular dynamics, the importance of calculating the free energy differences was realized in various chemical and biological problems. Since MC was devised earlier, most of the initial work was done in MC. There were many attempts to formulate new methods to find free energy differences with minimum statistical error. One such method was introduced by Bennet [124], called the acceptance ratio method or maximum overlap method; he showed that the statistical error can be reduced if the overlap between the two states is increased.

Although a lot of work was done using these techniques, it was realized that these methods have a disadvantage of slow convergence due to which the results obtained from these were different from experimental results [121]. As the size and complexity of simulation systems were increasing, there was a need to devise new methods which had better sampling and hence, were more accurate. This need drew the attention of researchers towards determining the potential of mean force (PMF) [90]. Potential of mean force $W(x)$ was first introduced by Kirkwood [115] which is calculated along a certain reaction coordinate x and is defined using average distribution function $< P(x) >$ as $W(x) = - k_B T \ln <$

$P(x) >$. The derivative of PMF gives the mean force on one particle due to all the configurations of all the other particles [125]. In other words it is the potential which gives the average force acting on some degree of freedom. PMF curve represents a free energy map along this degree of freedom or reaction coordinate. The reaction coordinate can be any property of a system to which a parameter has been assigned. Examples for the reaction coordinate include distance between two atoms, angle between two molecules etc. Calculation of PMF from a direct molecular dynamics simulation was not practical because of the presence of large energy barriers which might not provide sufficient sampling along the reaction coordinate x [126]. To overcome this difficulty, umbrella sampling simulations were introduced [92]. In umbrella sampling, an additional biasing potential is applied to the reaction coordinate to enhance the sampling in a small region and thus, help in crossing the energy barrier. Such a potential is centered at different values of reaction coordinate which are close to each other. Thus, a series of MD simulations were needed to give sufficient sampling for the entire range of reaction coordinate. The results obtained from all the sampling windows were then unbiased and used to calculate the unbiased probability and the associated PMF [126]. Many methods were developed to calculate the PMF from the results obtained by umbrella sampling simulations.

One such method was multiple histogram (MH) method formulated by Ferrenberg et al. [127] as an extension to the single histogram method [128] formulated by the same authors, in which the maximum and minimum value of a desired

quantity was calculated through a single simulation. MH method allows one to study the behavior of a desired quantity of the system in a given time frame. This method requires performing many simulations to reduce the error. An advantage of MH method is that it helps in identifying the need of additional simulations to give the most accurate result. Kumar *et al.* [91] introduced the weighted histogram analysis method (WHAM) which was an extension of the MH method [127]. This method is based on the maximum overlap method given by Bennett [124], in which a weighting factor is assigned such that it minimizes the statistical error while estimating the unbiased distribution function from the simulation data. The advantages of WHAM method over other methods are that it utilizes all the information from umbrella sampling simulation which reduces the statistical errors, it allows multiple overlaps in probability distribution to obtain better sampling and accurate result, and it can be applied to multidimensional cases as well [91]. Various other methods have been proposed to perform free energy calculations, such as the molecular mechanics/Poisson Boltzmann surface area (MM-PBSA) method and the linear interaction energy (LIE) method. In this thesis, we explain the detailed theory for the WHAM method and apply this method to evaluate PEI mediated DNA attraction. Although the derivation below is done for NVT ensemble the derivation for NPT ensemble is similar and the final WHAM equations apply to both ensembles.

Let us have *s* computer simulations and in each simulation the system is subjected to a biasing potential $U_{bias,i}(\boldsymbol{q})$, where $\boldsymbol{q}$ is a vector for the position of all atoms

and *i* represents the simulation number (*i*=1,2,3……*s*). Hence the potential energy of a biased system, $U_i(\boldsymbol{q})$, will be given as:

$$U_i(\boldsymbol{q}) = U_0(\boldsymbol{q}) + U_{bias,i}(\boldsymbol{q}) \tag{2.11}$$

where $U_0(\boldsymbol{q})$ gives the potential energy of the unbiased system. Now, we introduce a reaction coordinate (e.g. distance, angle etc) x. It should be noted that the reaction coordinate is chosen such that $U_{bias,i}$ will depend on $\boldsymbol{q}$ through x. For example, in this study, the two DNA molecules are harmonically restrained at their centers of mass and the distance between these centers of mass is the reaction coordinate. In this case, the expression governing $U_{bias,i}$ is:

$$U_{bias,i}(\boldsymbol{q}) = 2 * \frac{1}{2} k \left(\frac{x-\alpha_i}{2}\right)^2 \tag{2.12}$$

where α_i is the distance between the centers of mass at which the biasing potential is zero and x gives the value of the reaction coordinate at a particular time. By varying α_i, the reaction coordinate obtained from the simulation can cover a desired range for which the PMF is needed. Umbrella sampling (US) simulations are done through the entire range of x. US technique is used along with MD simulations in order to obtain good sampling density in a series of narrow windows along the reaction coordinate [129]. As explained earlier, a biasing potential is added which increases the potential barrier in all the regions except for the region where x is close to α_i [92]. A number of US simulations are

performed and the results obtained from these simulations are used to construct histograms by dividing the entire range of x into bins. Let us define N_i as the total number of counts for the i^{th} simulation and n_{ij} as the number of counts in the j^{th} bin of the i^{th} simulation. Then the probability of j^{th} bin in the i^{th} simulation (P_{ij}) is given by:

$$P_{ij} = \frac{n_{ij}}{N_i} \tag{2.13}$$

On the other hand, P_{ij} can be expressed in terms of phase space integration as:

$$P_{ij} = \frac{\int e^{-\beta U_i(\boldsymbol{q})}\delta(x-x_\mathrm{j})d\boldsymbol{q}}{\int e^{-\beta U_i(\boldsymbol{q})}d\boldsymbol{q}} = \frac{\int e^{-\beta U_0(\boldsymbol{q})}e^{-\beta\, U_{bias,i}(\boldsymbol{q})}\,\delta(\mathrm{x}-\mathrm{x_j})d\boldsymbol{q}}{\int e^{-\beta U_i(\boldsymbol{q})}d\boldsymbol{q}} =$$

$$\frac{e^{-\beta\, U_{bias,i}(x_j)}\int e^{-\beta U_0(\boldsymbol{q})}\,\delta(\mathrm{x}-\mathrm{x_j})d\boldsymbol{q}}{\int e^{-\beta U_i(\boldsymbol{q})}d\boldsymbol{q}} = \frac{\mathrm{Z}_0}{\mathrm{Z_i}}e^{-\beta\, U_{bias,i}(x_j)}\,P_j^0 \tag{2.14}$$

where x_j is the value of the reaction coordinate at the center of the j^{th} bin and $\mathrm{Z}_0 = \int e^{-\beta U_0(\boldsymbol{q})}\,d\boldsymbol{q}$ gives the configurational partition function for NVT ensemble. $\mathrm{Z_i} = \int e^{-\beta U_i(\boldsymbol{q})}\,d\boldsymbol{q}$ is the configurational partition function for the i^{th} simulation where the biasing potential $U_{bias,i}$ is applied. $P_j^0 = \frac{\int e^{-\beta U_0(\boldsymbol{q})}\delta(x-x_\mathrm{j})d\boldsymbol{q}}{\int e^{-\beta U_0(\boldsymbol{q})}d\boldsymbol{q}}$ is the probability of j^{th}bin in the unbiased system. The function $\delta(x - x_\mathrm{j})$ is the delta function which means that this function will always give a value of zero except when $x = x_\mathrm{j}$. It is known that PMF is given by $W = k_B T \ln P$ and the free energy of the system in canonical ensemble is $F = -k_B T \ln Z$. Therefore,

$$F_0 - F_i = k_B T \ln \frac{Z_i}{Z_0} = k_B T \ln \left[\frac{P_j^0}{P_{ij}} e^{-\beta\, U_{bias,i}(x_j)} \right] = W_{ij} - W_j^0 - U_{bias,i}(x_j) \quad (2.15)$$

where F_0 is the free energy of the unbiased system, F_i is free energy of system in the i^{th} simulation, W_j^0 is the PMF of the unbiased system in bin j and W_{ij} is the PMF in the j^{th} bin of the i^{th} simulation. Now from (2.15), we can observe that the left hand side of the equation is independent of the bin j; whereas on the right hand side all the terms are functions of j. The difference is, however, expected to be independent of j.

Theoretically, P_j^0 calculated from P_{ij} should be the same for all biasing potential $U_{bias,i}$. However, this is not the case practically due to statistical errors. Let us introduce $\Omega_{ij} = \frac{Z_i}{Z_0} e^{\beta\, U_{bias,i}(x_j)} P_{ij}$ which gives unbiased probability of the j^{th} bin calculated from the i^{th} simulation. Averaging Ω_{ij} from all s simulations should give the actual unbiased probability distribution P_j^0, i.e.,

$$< \Omega_{ij} > = P_j^0 \quad (2.16)$$

Let us introduce some variables for the ease of calculation. Let f_i be the normalizing constant such that $P_{ij} = f_i c_{ij} P_j^0$, where $c_{ij} = e^{-\beta U_{bias,i}(x_j)}$ is the biasing factor. Clearly from (2.14) $f_i = \frac{Z_0}{Z_i}$. In addition, normalization of

probability distribution requires that $\sum_1^m P_{ij} = 1$, where m is the total number of bins, hence, $f_i = \frac{1}{\sum_1^m e^{-\beta U_{bias,i}(x_j)} P_j^0}$. Thus, we get,

$$\Omega_{ij} = \frac{n_{ij}}{N_i c_{ij} f_i}. \tag{2.17}$$

To perform WHAM, we calculate the unbiased probability as a weighted sum of Ω_{ij} from all *s* simulations.

$$P_j^0 = \sum_{i=1}^s w_i \Omega_{ij}. \tag{2.18}$$

We have to find the value of the weighting factor w_i, with the constraint $\sum_{i=1}^s w_i = 1$. Its value is determined such that the variance of P_j^0 given by (2.18) can be minimized. The variance of P_j^0 is

$$var\left(P_j^0\right) = < \left(P_j^0 - < P_j^0 >\right)^2 >. \tag{2.19}$$

From (2.18) and (2.19),

$$var\left(P_j^0\right) = < \left(\sum_{i=1}^s w_i \Omega_{ij} - < \sum_{i=1}^s w_i \Omega_{ij} >\right)^2 >$$

$$= < \left(\sum_{i=1}^s w_i (\Omega_{ij} - < \Omega_{ij} >)\right)^2 > \tag{2.20}$$

Let $\Omega_{ij} - < \Omega_{ij} > = \alpha_{ij}$, then

$$var\left(P_j^0\right) = < \left(\sum_{i=1}^{s} w_i \alpha_{ij}\right)^2 > \quad (2.21)$$

$$= < \sum_{i=1}^{s} {w_i}^2 {\alpha_{ij}}^2 + \sum_{k \neq l=1}^{s} w_k w_l \alpha_{kj} \alpha_{lj} > \quad (2.22)$$

$$= \sum_{i=1}^{s} {w_i}^2 < {\alpha_{ij}}^2 > + \sum_{k \neq l=1}^{s} w_k w_l < \alpha_{kj} \alpha_{lj} > \quad (2.23)$$

Now $< {\alpha_{ij}}^2 > = var(\Omega_{ij})$. Assuming that simulations k and l are independent which implies that $< \alpha_{kj} \alpha_{lj} > = 0$, it can be concluded that

$$var\left(P_j^0\right) = \sum_{i=1}^{s} {w_i}^2 var(\Omega_{ij}). \quad (2.24)$$

Using (2.16) & (2.17), we get

$$var\left(P_j^0\right) = \sum_{i=1}^{s} \frac{w_i^2}{{(N_i f_i C_{ij})}^2} \, var\left(n_{ij}\right). \quad (2.25)$$

Next, we need to find $var(n_{ij})$. If there are N_i independent samples in a simulation *i*, then the probability of finding *n* samples in a bin is given by the binomial distribution as $C_N^n = \bar{p}^n (1 - \bar{p})^{N-n}$, where $\bar{p}$ is the probability of the bin. When *N* is large, mean of *n* will be equal to the variance of *n* and will be given by:

$$var\left(n_{ij}\right) = \langle n_{ij} \rangle = \langle \Omega_{ij} \, N_i c_{ij} f_i \rangle = N_i f_i c_{ij} P_j^0. \quad (2.26)$$

Substituting (2.26) into (2.25),

$$var\left(P_j^0\right) = \sum_{i=1}^{s} \frac{w_i^2}{(N_i f_i c_{ij})} P_j^0. \quad (2.27)$$

We already stated that $\sum_{i=1}^{s} w_i = 1$ as a constraint. To minimize $var\left(P_j^0\right)$ with this constraint define:

$$Q(w_i) = \sum_{i=1}^{s} \frac{w_i^2}{(N_i f_i c_{ij})} P_j^0 - \lambda \sum_{i=1}^{s} w_i, \quad (2.28)$$

where λ is the Lagrange multiplier. Differentiating $Q(w_i)$ with respect to w_i and setting the results to 0 gives

$$\frac{\partial Q}{\partial w_l} = \frac{2w_l}{(N_l f_l c_{lj})} P_j^0 - \lambda = 0, \quad (2.29)$$

$$\text{i.e., } w_l = \frac{(N_l f_l c_{lj})}{2P_j^0} \lambda. \quad (2.30)$$

Using the constraint $\sum_{i=1}^{s} w_i = 1$ on (2.30), we get:

$$\sum_{l=1}^{s} w_l = \frac{\lambda \sum_{l=1}^{s} (N_l f_l c_{lj})}{2P_j^0} = 1, \quad (2.31)$$

Therefore,

$$\lambda = \frac{2P_j^0}{\sum_{l=1}^{S}(N_l f_l c_{lj})} \tag{2.32}$$

From here we get the value of w_i (using (2.30)) as:

$$w_i = \frac{N_i f_i c_{ij}}{\sum_{l=1}^{S} N_l f_l c_{lj}} \tag{2.33}$$

Substituting it in (2.17), we get

$$P_j^0 = \frac{\sum_{i=1}^{S} n_{ij}}{\sum_{i=1}^{S} N_i f_i c_{ij}}, \tag{2.34}$$

or

$$P_j^0 = \frac{\sum_{i=1}^{S} n_{ij}}{\sum_{i=1}^{S} N_i e^{-\beta U_{bias,i}(x_j)} f_i} \tag{2.35}$$

where $f_i = \frac{1}{\sum_{j=1}^{m} e^{-\beta U_{bias,i}(x_j)} P_j^0}$ as explained earlier. These equations for P_j^0 (2.35) and f_i show a set of coupled equations. There are s equations for P_j^0 and m equations for f_i, which should be solved iteratively to self consistency. As explained in the beginning of this section these methods can be used for calculating the free energy difference ($\Delta\bar{F}_i$) between two states. In such cases to determine the free energy, we assume the free energy or partition function of one of the states to be arbitrary. f_i can be used to calculate the change in free energy $\Delta\bar{F}_i$ for each simulation as,

$$\Delta\bar{F}_i = -\frac{1}{\beta}\ln\left\{\sum_{j=1}^{m} P_j^0 e^{-\beta\, U_{bias,i}(x_j)}\right\} \tag{2.36}$$

The calculations start with assuming an arbitrary initial value of f_i and using that value to calculate P_j^0. The value of P_j^0, hence calculated is used to obtain another value for f_i and these iterations continue until self consistency is reached.

Chapter 3

Simulation Details

This chapter explains the models of DNA and PEI that were simulated in this work. It also provides the details of the procedure used to perform MD simulations.

3.1 DNA-native PEI systems simulated

The DNA dodecamer molecule used in this study was a Dickerson-Drew B-DNA [130]. The DNA molecule had 24 nucleotides with the sequence of $d(CGCGAATTCGCG)^2$. The DNA molecule carried a total charge of -22 in its fully deprotonated state. As explained in chapter 1, PEI can have different properties including molecular weight, protonation ratio and degree of branching. The PEI (Fig. 3.1) simulated in this work has thirteen amine groups with a molecular weight of 586 Da [70]. It has been shown by earlier MD simulations that the degree of branching has a negligible effect for this LMW PEI [70]. Two protonation ratios are used: 46% and 23%. The former corresponds to 6 out of the 13 amine groups being protonated while the latter corresponds to 3 out of the 13 amine groups being protonated (Fig. 3.1). The charges are distributed nearly uniformly along the structure of the PEI molecule. These protonation ratios are

close to the results obtained by Utsuno *et al*., [66] where 600 Da PEI was found to have 47% protonated amines at pH 6 and 21% protonated amines at pH 8.

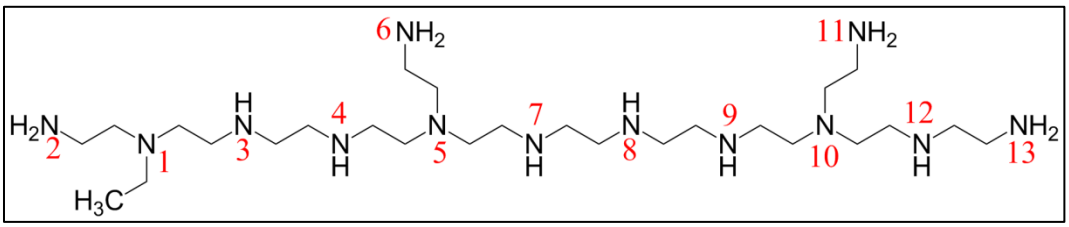


Fig. 3.1 Molecular structure of the PEI simulated. The nitrogen numbers are given in red. For 46% protonated PEI nitrogens 2, 4, 6, 8, 11 and 13 are protonated whereas for 23% protonated PEI nitrogens 2, 6 and 11 are protonated.

In order to study the native PEI mediated DNA aggregation, a total of five systems were simulated. Four of these systems comprised of two DNA molecules and 46% protonated PEIs. The fifth system contained 23% protonated PEIs along with two DNA molecules. The four systems with 46% protonated PEIs contain 8, 6, 4 and 2 PEIs respectively, and are in turn referred to as systems 2D-8P, 2D-6P, 2D-4P and 2D-2P throughout the thesis. These four systems are simulated and compared in order to address the effect of N/P charge ratio on the PMF. The fifth simulated system consists of two DNAs and eight 23% protonated PEIs, and is referred to as system 2D-8P(23%). Since system 2D-8P(23%) has the same N/P charge ratio as system 2D-4P, their comparison allows us to study the effect of PEI's protonation state on the PMF.

The initial configurations of the systems described above are shown in Fig. 3.2, with systems 2D-8P and 2D-8P(23%) having the same initial configuration. For

each system, the two DNAs at the beginning of the simulation had their principle axes aligned parallel to each other as well as to the axes of the PEIs. In specifying the initial configuration, we have considered the two important mechanisms governing PEI mediated DNA aggregation, which are polyion bridging and charge neutralization. In particular, we placed two PEIs in between the two DNAs so that they can form polyion bridges. The two PEIs were placed such that their centers of mass (COMs) were equidistant from the COMs of the two DNAs. The rest of the PEIs were placed around the DNA in order to capture the effect of charge neutralization. They were placed such that their COMs were at a distance of 20 Å from the COM of the nearest DNA molecule. Each system was placed in a water box large enough to make sure that each solute molecule was at a distance of at least 16 Å from the edge of the water box, i.e. the distance between the outermost PEI and the closest PEI in the periodic boxes is 32 Å. The distance between the DNA molecule and the closest DNA in the periodic boxes is therefore 72 Å. To maintain neutralization of the simulation box, certain amounts of sodium or chlorine ions were added to the water box by replacing equal number of water molecules. For example, in 2D-8P system, the total negative charge from DNA molecules was -44 and the eight PEIs carrying a charge of +6 each was able to neutralize the negative charges of DNA while creating a charge of +4 in the simulation box. This charge was neutralized by adding 4 chlorine ions in the box, which is done by randomly replacing 4 water molecules. It should be noted that these monovalent counterions were just added to make the water box

neutral and didn't actually facilitate DNA aggregation. The details of the five simulated systems are tabulated in Table 3.1.

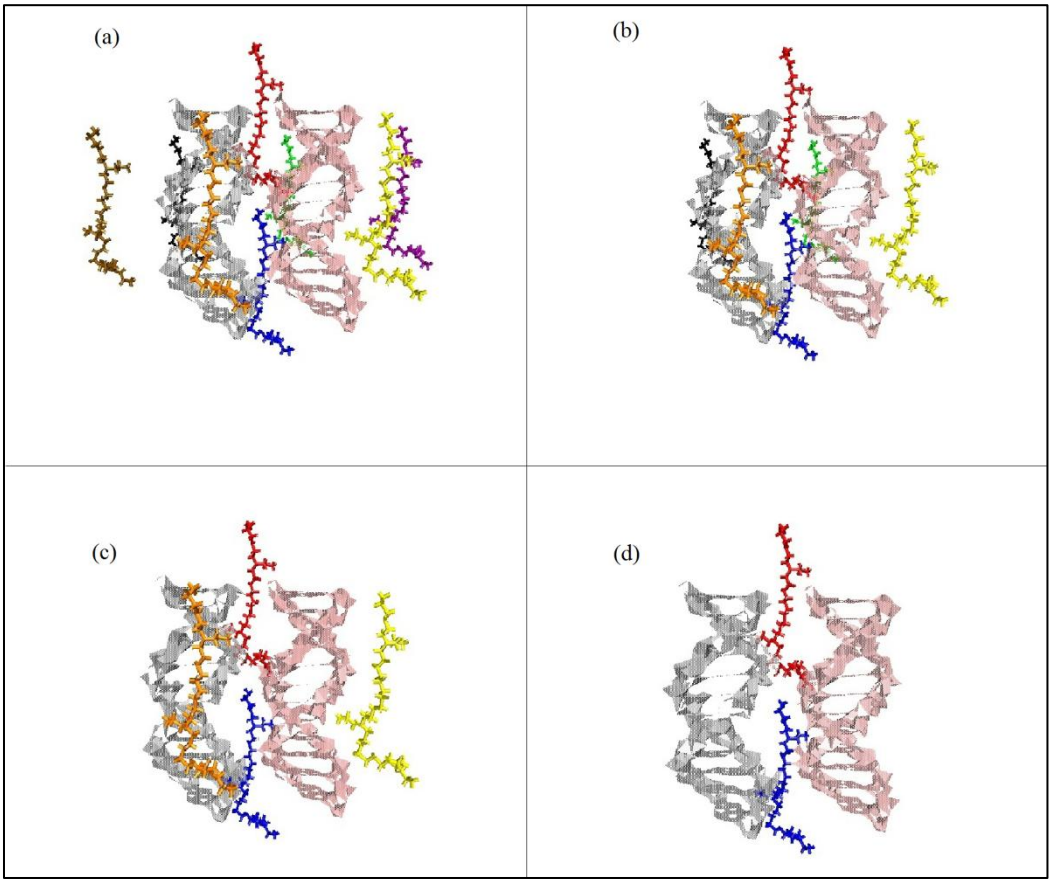


Fig. 3.2 Initial configurations (at 0 ns) of the simulated systems each containing 2 DNAs and a different number of PEIs: (a) 2D-8P and 2D-8P(23%), (b) 2D-6P, (c) 2D-4P and (d) 2D-2P.

3.2 DNA-lipid modified PEI systems simulated

In order to study the effect of lipid substitution on PEI mediated DNA aggregation, we substituted the 46% protonated 586 Da PEI (described in section

3.1) with three different lipids namely linoleic acid (LA), oleic acid (OA) and caprylic acid (CA). Fig. 3.3 gives the chemical structure of the lipids.

Table 3.1 Details of the simulated systems

Name of the system	2D-8P	2D-6P	2D-4P	2D-2P	2D-8P(23%)
No. of PEI	8	6	4	2	8
Protonation ratio of PEI	46%	46%	46%	46%	23%
N/P charge ratio	48/44=1.09	36/44=0.81	24/44=0.54	12/44=0.27	24/44=0.54
Type and No. of ions	4 Cl^-	8 Na^+	20 Na^+	32 Na^+	20 Na^+

OA and LA are the longer lipids (with 18 carbon atoms) and CA is the shorter lipid (with 8 carbon atoms). In addition, OA and LA have the same number of carbon atoms and a similar structure except the presence of an extra double bond in LA. The lipids are grafted to the nitrogen number 6 in Fig. 3.1. The lipid substituted PEI structure is made such that the protonation ratio of PEI remains 46%. In particular, the charges are re-distributed along the PEI such that 6 out of 13 nitrogens (nitrogen number: 2, 4, 7, 9, 11 and 13 in Fig. 3.1) are protonated. Three systems are simulated, namely, 2D-8P(OA), 2D-8P(LA) and 2D-8P(CA), with each of them containing 2 DNA and 8 lipid-substituted PEI molecules. In order to perform fair comparison among these three systems and with system 2D-8P, the initial configurations of the DNA and the lipid-modified PEI are kept the same as in Fig. 3.2a. Again, each system was placed in a water box large such that the distance between the outermost PEI and the closest PEI in the periodic boxes

is 32 Å. The excess positive charge due to the PEIs was neutralized by adding 4 chlorine ions in the box, which is done by randomly replacing 4 water molecules.

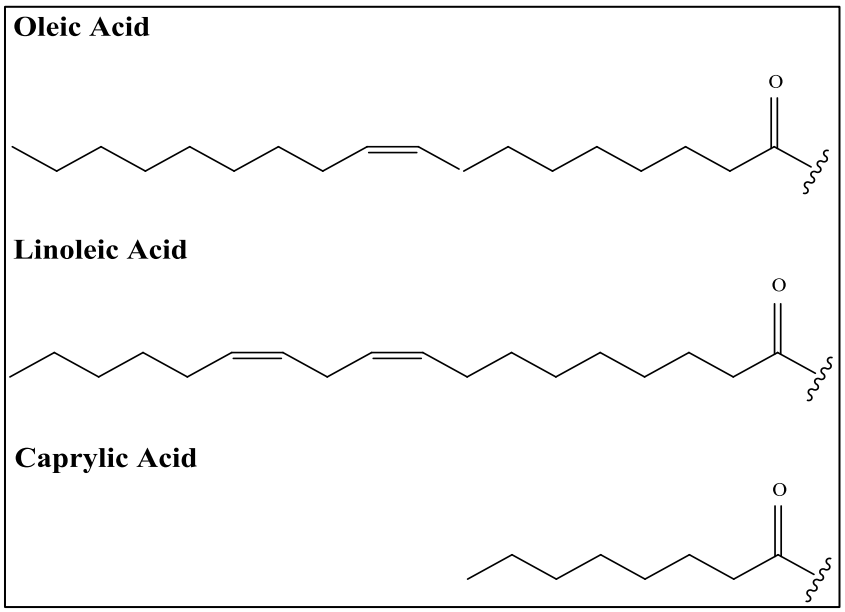


Fig. 3.3 Chemical structures of oleic acid, linoleic acid and caprylic acid which are substituted on nitrogen number 6 marked in red on the PEI in Figure 3.1.

3.3 Simulation details

The MD simulations performed in this work were run using NAMD [131]. NAMD is a parallel computing based MD code designed for the simulation of biomolecular systems and was developed by the Parallel Programming Laboratory and the Theoretical and Computational Biophysics group at the University of Illinois at Urbana Champaign. NAMD uses object oriented structure and prioritized message driven execution techniques of Charm++ [131]. A simulation box is divided into equal sized cells which are just larger than the

assigned cut off radius. Each cell can interact with its entire neighbouring cell. The interactions are calculated not only within the cells but also with the neighbouring cells. A major advantage of using NAMD is that these cells can be simulated on different processors independently which shortens the processing time [132]. A typical NAMD simulation requires four input files: (1) Protein Data Bank (pdb) file containing information about the coordinates for all the atoms, (2) Protein structural file (psf) which includes all the structural information, (3) A parameter file which includes information about the force field including bond length, strength etc., and (4) An NAMD configuration file which governs the physical conditions of the simulation like temperature, pressure, cut-off distances etc.

A force field consists of all the information related to the bonded and non-bonded interactions. In this work, we used CHARMM general force field [133] for the PEIs and CHARMM 27 force field [134] for the rest of the molecules. TIP3P (transferable intermolecular potential 3P) model [135] was used for treating water molecules in which each of the three atoms is assigned a partial charge as well as a Lennard-Jones potential parameter [136]. PME method [109] with periodic boundary condition was used for calculating electrostatic interaction. Calculation of long-ranged electrostatic interactions is a major time-consuming part of the simulation; hence, approximations are made to calculate them. Two of the most common algorithms used for calculating long-ranged electrostatic interactions are PME method and standard cut-off method [110]. In a simulation based study by

Rog [137], they compared the artefacts of both the methods on charged and uncharged systems. For the charged system, they concluded that the cut-off method generated artefacts related to the order and mobility of the charged particles making the results very different as compared to experimental results. On the other hand, results obtained by using PME method for charged systems were compliant to the experimental data, hence making PME a superior method for long-ranged electrostatic calculations in MD simulations. The PME method has been most widely employed for MD simulations of large and complex systems. A time step of 2 fs was used for all the simulations. SHAKE algorithm [138] was used to constrain the hydrogen bonds and the non-bonded interactions were cut off at 12 Å. VMD [139] was used to visualize the systems and analyze the trajectories obtained from the MD simulations.

Before simulating the system, the PEI and DNA molecules were separately pre-equilibrated in water. For each simulation, the system was first minimized for 5000 steps in order to eliminate bad contact. This includes 2000 steps of minimization with constrained solute molecules followed by 2000 steps of minimization with harmonically restrained solutes and 1000 steps of unrestrained minimization. Afterwards, the system was gradually heated from 0 to 300 K in 20 ps. A harmonic restraint of 10 kcal/mol/$Å^2$ was then applied on the non-hydrogen atoms of the solutes at 300 K and 1 bar pressure, and dynamics was run for 4 ns to allow the ions and water molecules around the DNAs and PEIs to relax. After removal of the restraint, the system was then simulated in an isothermal-isobaric

ensemble condition at 300 K temperature and 1 bar pressure for 20 ns. The trajectory of each simulation was saved at every 1000 steps and the data from the last 10 ns of the 20 ns simulation were used for analysis. For the lipid-modified PEI-mediated DNA aggregation, the systems were simulated for 30 ns, keeping all the other parameters the same. This is because in the presence of the lipid, the molecules are larger and the attainment of dynamic equilibrium requires more time. It was made sure that the number of data points for both native and lipid substituted PEIs should be the same for accurate comparison, hence, in both the cases the data obtained from the last 10 ns was used to perform the PMF calculations.

3.4 PMF calculation

As mentioned earlier in order to perform umbrella sampling, we chose the distance between the COMs of the two DNAs as the reaction coordinate *x* and umbrella sampling windows were created at 1 Å interval along the reaction coordinate. A total of 29 windows were created with the value of the reaction coordinate varying from 22 Å to 50 Å, which provided sufficient sampling for the entire range of the reaction coordinate. For each window *i* (*i = 1, 2 …29*), the COM of each DNA was subjected to a harmonic restraint with a biasing potential in the form of $k\left((x-\alpha_i)/2\right)^2/2$, where the force constant k = 2.0 kcal/mol/Å^2 and α_i is the equilibrium separation associated with the i^{th} biasing potential. The choice of the value of k is based on previous studies involving binding of

molecules with similar size to our simulated molecules [140, 18]. For this thesis, very large amount of simulations were performed with total run time of $[(5\ systems * 29\ windows * 20\ ns) + (3\ systems * 29\ windows * 30\ ns)] = 5.5\ \mu s$. Following WHAM, the trajectories collected from all the simulations were used to calculate the PMF. Specifically, the entire range of the reaction coordinate was divided into bins of size 0.25 Å. Such a size was determined by a convergence test in which the dependence of the PMF on the bin size was found, the results of which are presented in chapter 4. A histogram was then constructed for the number of counts in each bin vs. the reaction coordinate. Once the histogram was created, the unbiased probability *P* as a function of the reaction coordinate and the free energy shift $\Delta\bar{F}_i$ for each umbrella sampling simulation were calculated using the WHAM [91] equations which were derived in section 2.2.1 and are again mentioned below:

$$P_j^0 = \frac{\sum_{i=1}^{s} n_i(x_j)}{\sum_{i=1}^{s} N_i e^{-\beta U_{bias,i}(x_j)} e^{-\beta \Delta \bar{F}_i}} \tag{3.1}$$

$$\Delta\bar{F}_i = -\frac{1}{\beta}\ln\left\{\sum_{j=1}^{m} P_j^0 e^{-\beta U_{bias,i}(x_j)}\right\} \tag{3.2}$$

$$U_{bias,i}(x_j) = 2 * \frac{1}{2} k \left(\frac{x_j - \alpha_i}{2}\right)^2 \tag{3.3}$$

In these equations, $s = 29$, is the total number of umbrella sampling simulations, N_i gives the total number of counts in simulation i and $n_i(x_j)$ is the number of counts in simulation i that fall into the bin centered at x_j. The biasing potential $U_{bias,i}(x_j)$ as given in (3.3) is twice the harmonic potential applied on each DNA

since both DNAs are subjected to the harmonic restraint. A MATLAB code (given in appendix A-1) was developed to solve the above equations iteratively to a satisfactory convergence to arrive at the values of $P(x)$ and $\Delta\bar{F}_i$, after which the PMF $W(x)$ was calculated using

$$W(x) = -k_B T \ln P(x) \tag{3.4}$$

In order to confirm that converged results are obtained, we used two conditions in the MATLAB code (appendix A1). The first condition is setting the maximum number of iterations = 100000, so that the loop doesn't run infinitely. This condition was applied in order to make sure that the code does not have any error; this number of iterations was found to be sufficiently more than those required for convergence of results. The second condition is to set the maximum error tolerance = 0.0001 between two consecutive values of $\Delta\bar{F}_i$. The datum of the PMF was adjusted such that at sufficiently large value of x, the PMF became zero.

Chapter 4

Results for native PEI mediated DNA attraction

In this chapter, we present the PMF obtained from the MD simulations of native PEI mediated DNA attraction. We first demonstrate the basic characteristics of the PMF curve using the example of the 2D-8P system. A comparison is then made among the systems with different N/P charge ratios and PEI protonation states, which allows us to address the stability and the compactness of the aggregate under different conditions.

4.1 PMF for system 2D-8P

As mentioned in the previous chapter, 29 simulations were run for every system. Each simulation has a different initial COM separation distance between the two DNAs described by the harmonic potential applied at the two COMs. Because of the interactions between the two DNAs, they don't remain parallel and the COM separation does not remain at the initial value; instead it varies with the simulation time. Fig. 4.1 represents one such curve in which the initial distance between the two COMs was 25 Å. When the solutes are released at 0 ns, large fluctuations are found in the first 5 ns, which is due to the imbalance of forces acting on the DNAs: repulsion between the DNAs, attraction introduced by the PEIs and the

spring force from the harmonic potential applied at COMs. As the simulation continues dynamic equilibration of these forces is gradually achieved, leading to decreased amplitude of fluctuations. It can be observed from Fig 4.1 that the COM distance has a decreasing trend indicating that the equilibrium separation between the two DNAs should be smaller than 25 Å.

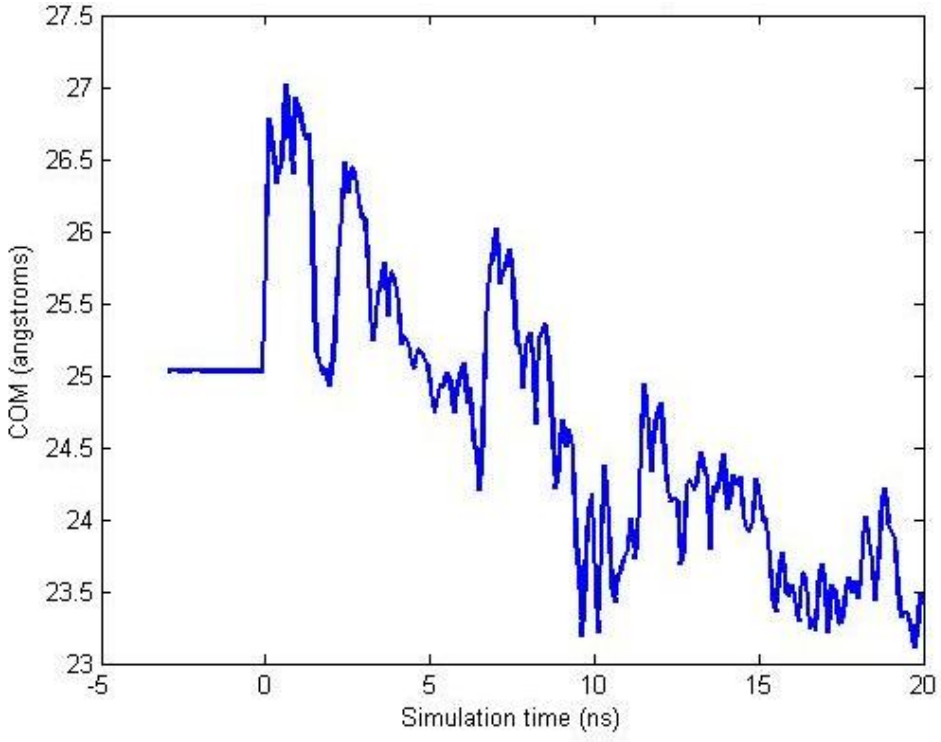


Fig. 4.1 Variation of COM distance between the two DNAs in the presence of eight native PEIs when the initial distance was 25 Å. Graphs for all other initial COM distances are given in appendix 1.

From the total simulation time of 20 ns, the data obtained from last 10 ns was used to perform the PMF calculation. Using the data obtained from all 29 simulations, a histogram was constructed for the entire range of reaction coordinates (x) (Fig. 4.2). As mentioned in chapter 3 in order to construct the

histogram, the entire range of the reaction coordinate was divided into bins of size 0.25 Å. The PMF was then calculated by iteratively solving equations (3.1) and (3.2).

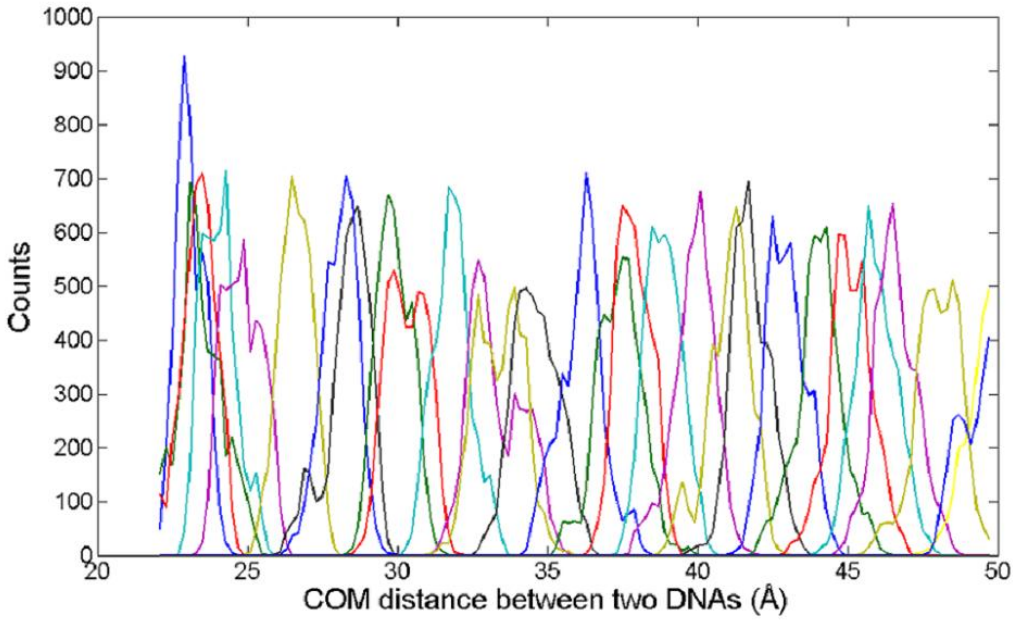


Fig. 4.2 Histogram for the 2D-8P system showing the number of counts for the entire range of the reaction coordinates. A total of 29 simulations were performed to generate the histogram. Different curves in the figure corresponds to simulations with different biasing potential $U_{bias,i}(x)$ (i = 1, 2, …, 29).

The PMF for system 2D-8P is plotted along the reaction coordinate as shown in Fig. 4.3. As the COM distance between the two DNAs (x) increases, the PMF curve first takes the form of a branch with decreasing slope which is characteristic of the repulsive interactions between the two negatively charged DNA molecules. The curve then reaches its global minimum at 23.6 Å which is the most important

point in the curve as it provides information about the compactness and stability of the DNA aggregate. Beyond the global minimum, there is an overall increasing trend in the PMF till 35 Å, which represents the attraction between the two DNA molecules due to the presence of the PEI molecules. After 35 Å, the curve fluctuates without overall increase in the PMF, indicating negligible interaction beyond this separation. It can be observed from Fig. 4.3 that the global minimum of the PMF curve is -6.7 kcal/mol and is located at a separation of x = 23.6 Å. This result is in accordance to the structure obtained from the simulation study of Sun *et al*., [38] where the shortest distance between two DNAs aggregated by eight 600 Da 46% protonated PEIs was found to be 23.2 Å.

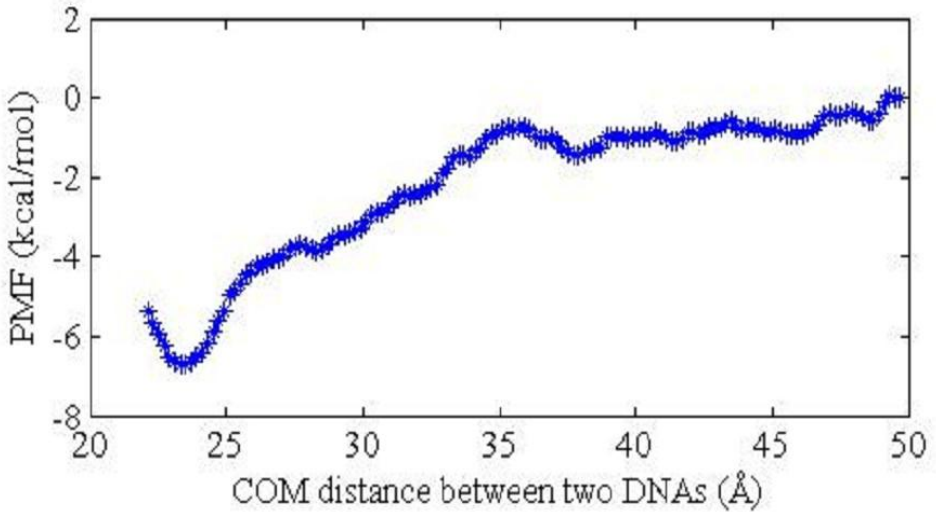


Fig. 4.3 PMF vs. DNA COM distance for system 2D-8P

The size and the stability of the DNA aggregates are crucial for gene delivery and transfection [141] and are important criteria for determining the extent of the DNA aggregation. The PMF curve has characteristics which allow us to examine these two factors. The location of the global minimum in the PMF curve is the COM distance at which the interaction between the two DNAs is neither repulsive

nor attractive. That is, it is the equilibrium separation between the two DNAs and is an indicator of the compactness of the aggregate. In addition, the depth of the PMF well, i.e., the absolute value of the PMF at its global minimum, is an indicator of the stability of the aggregate. A larger depth of the PMF well corresponds to larger amount of energy needed to dissociate the aggregate and hence, higher stability.

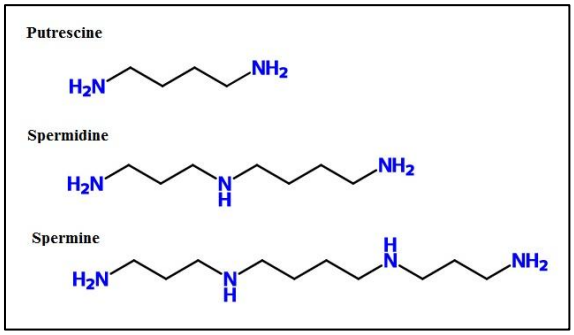


Fig. 4.4 Molecular structure of putrescine, spermidine and spermine molecules used by *Dai et al.* [18].

In a study, Dai *et al*. [18] performed MD simulations to investigate the attraction between DNA molecules in the presence of multivalent ions and calculated the interaction potential between two parallel 10-base pair DNAs due to these multivalent ions. Three types of ions were used in their simulations: putrescine (divalent), spermidine (trivalent) and spermine (tetravalent) (Fig. 4.4). Several observations can be made on the characteristics of multivalent ion and polyion medicated DNA aggregations from the comparison between their work and the work presented in this thesis. First, the minimum of the interaction potential for putrescine, spermidine and spermine occurs respectively at x = 23.8 Å, 22.8 Å and

23.7 Å, with the depth of the potential well valued at 2 k_BT (1.18 kcal/mol), 6 k_BT (3.5 kcal/mol) and 9 k_BT (5.3 kcal/mol), respectively [18]. Comparing those with the depth of the PMF (6.7 kcal/mol) for PEI (which has a valence of +6) mediated DNA attraction, a general trend of stronger DNA-DNA attraction with cations of higher valence is observed. It can also imply that more stable DNA aggregates are formed by cations of higher valence. It should be pointed out that in the work of Dai et al. [18], the DNA charges are just neutralized by the multivalent ions (except for the case of spermidine where there is a net charge of +2 neutralized by Cl- ions), while in the current simulation, the two DNAs and eight PEIs carry a net charge of +4, i.e., the DNAs are over neutralized by the PEIs. Despite this, stronger DNA-DNA attraction is found for PEI mediated DNA aggregation. This is due to the PEI's ability to interact locally with the DNA and bring the DNAs together via polyion bridging [38, 70]. The polyion bridges formed by PEIs are much stronger than the bridges formed by small multivalent ions, contributing to larger attraction between the DNAs.

Secondly, from the data in the current work and in the work by Dai *et.al.*, [18] a clear relation between the valence of the ions and the equilibrium separation between the two DNAs cannot be identified. In particular, spermidine, despite having a smaller valence than spermine, gives rise to a smaller equilibrium separation between the two DNAs [18], and the equilibrium separations associated with spermine and PEI are approximately the same. It is interesting to note that experimentally different conclusions also exist as to whether polycations

can form more compact DNA aggregates compared with small multivalent ions. In the experiments by Kasyanenko and Afanasieva [143], the DNA aggregation in the presence of different trivalent ions (Fe3+, La3+, [Co(NH3)6]3+, spermidine ions) and cationic PLL molecules was compared. They observed that the increase in valence of the cations led to more reduction in the DNA volume [143]. On the other hand, a different experiment reported that for the same cationic charge, smaller ions formed more compact and stable DNA condensates [18]. A possible explanation for this is that while polycations have higher valence and can aggregate the DNAs through stronger polyion bridging [38], their large size and the resultant steric effect can prevent them from forming more compact DNA aggregates compared with those formed through small cations. Finally, comparing Fig. 4.3 with the curve in [18] relating the interaction potential to the interduplex distance, Fig. 4.3 is much less smoother and shows many local minima along the curve. This is not due to any numerical issue, because the good amount of data obtained from a large number of simulations (29 for each system as compared to 4 simulations for each system in [18]) with umbrella sampling windows densely distributed along the reaction coordinate resulted in a well overlapped histogram (see Fig. 4.2). Instead, this is due to the fact that compared with small multivalent ions, PEI is a more flexible molecule and can adopt more configurations in the MD simulation. Throughout this thesis, only the global minimum in the PMF and the associated DNA separation value will be used to gauge the strength of the DNA attraction.

It should be pointed out that polyions such as PEIs are highly charged slow diffusing molecules and hence it is important to make sure that they attain equilibrium configuration in order to obtain accurate results. In order to confirm this; we did a convergence test where we chose different 10 ns - time slots from our simulation time and calculated the PMF for each of them. Table 4.1 compares the PMF well results for the 2D-8P system and shows that the results are converging concluding the attainment of equilibrium. Finally the bin size (0.25Å) used to generate the histogram, was also chosen based on a convergence test where the bin size was gradually reduced until the depth of the PMF became insensitive to the change in bin size. This is shown in Table 4.2.

Table 4.1 Comparison between the depths of PMF well obtained from different 10 ns - time slots from our simulation time for 2D-8P system. The depth of PMF well reflects convergence of results.

Time Slot (ns)	Depth of PMF well (kcal/mol)
7.5-17.5	7.9
8.0-18.0	7.5
8.5-18.5	7.1
9.0-19.0	6.8
9.5-19.5	6.7
10.0-20.0	6.7

Table 4.2 Comparison between the depths of PMF well obtained for different bin width for 2D-8P system. The depth of PMF well reflects convergence of results.

Bin Width	Depth of PMF (kcal/mol)
0.7	6.19
0.65	6.35
0.6	6.22
0.55	6.16
0.5	6.64
0.45	6.67
0.4	6.76
0.35	6.74
0.3	6.72
0.25	6.7
0.2	6.7

4.2 Effect of N/P charge ratio

As mentioned in chapter 1, various studies in the literature have shown that the N/P charge ratio plays an important role in polycation mediated DNA aggregation [76, 66, 79]. The effect of N/P charge ratio is studied in this work by comparing the PMF for systems 2D-8P, 2D-6P, 2D-4P and 2D-2P with N/P charge ratio varying from ~1 to ~0.25. For each system, 29 simulations were run and the resulting data was used for calculating the PMF. The PMF curves for systems 2D-6P, 2D-4P and 2D-2P are respectively shown in Fig. 4.5(a)-(c). For the 2D-6P system, the PMF curve has a global minimum at 25.6 Å at which the PMF value is -4.04 kcal/mol. For the 2D-4P system, the minimum is obtained at 26.1 Å with a PMF value of -2.92 kcal/mol. For the 2D-2P system, the minimum is obtained at 28.9 Å which corresponds to a PMF value of -1.96 kcal/mol. Although it can be

observed from Fig. 4.4 that there are many local minima for all these three systems, the value of the deepest minimum is used.

Fig. 4.6 plots the equilibrium separation as well as the depth of PMF well at that separation for all the four systems. It is clear from Fig. 4.6 that as the N/P charge ratio increases, the equilibrium separation of the DNAs decreases and the depth of the PMF well increases, both varying nearly linearly with the N/P charge ratio. This suggests that systems with higher N/P charge ratio are able to form not only more compact but also stronger DNA aggregate. The enhanced stability with increasing N/P charge ratio is also reflected through the smoothness of the PMF curve. PMF curves given in Figs. 4.3 and 4.5 show an increase in fluctuation, i.e., an increase in the number of local minima along the reaction coordinate, as the number of PEIs reduces. This implies that at a lower N/P charge ratio, there is an increase in the probability of finding the DNAs at these local minima, which are shallower and associated with lower stability. In fact, in an earlier work involving the simulation of two DNAs aggregated by two PEIs [38], after 200 ns simulation, the two DNAs were found to be separated by a distance of 55 Å, and there were no PEIs simultaneously connecting the two DNAs. The shallow well in the PMF as shown in Fig. 4.5(c) together with the existence of multiple local energy minima beyond 35 Å have contributed to the inability of forming sufficiently stable DNA aggregates that can sustain much longer simulations. The results here are consistent with experimental observation that higher N/P charge ratio facilitates the DNA condensation process and improves the uptake and retention

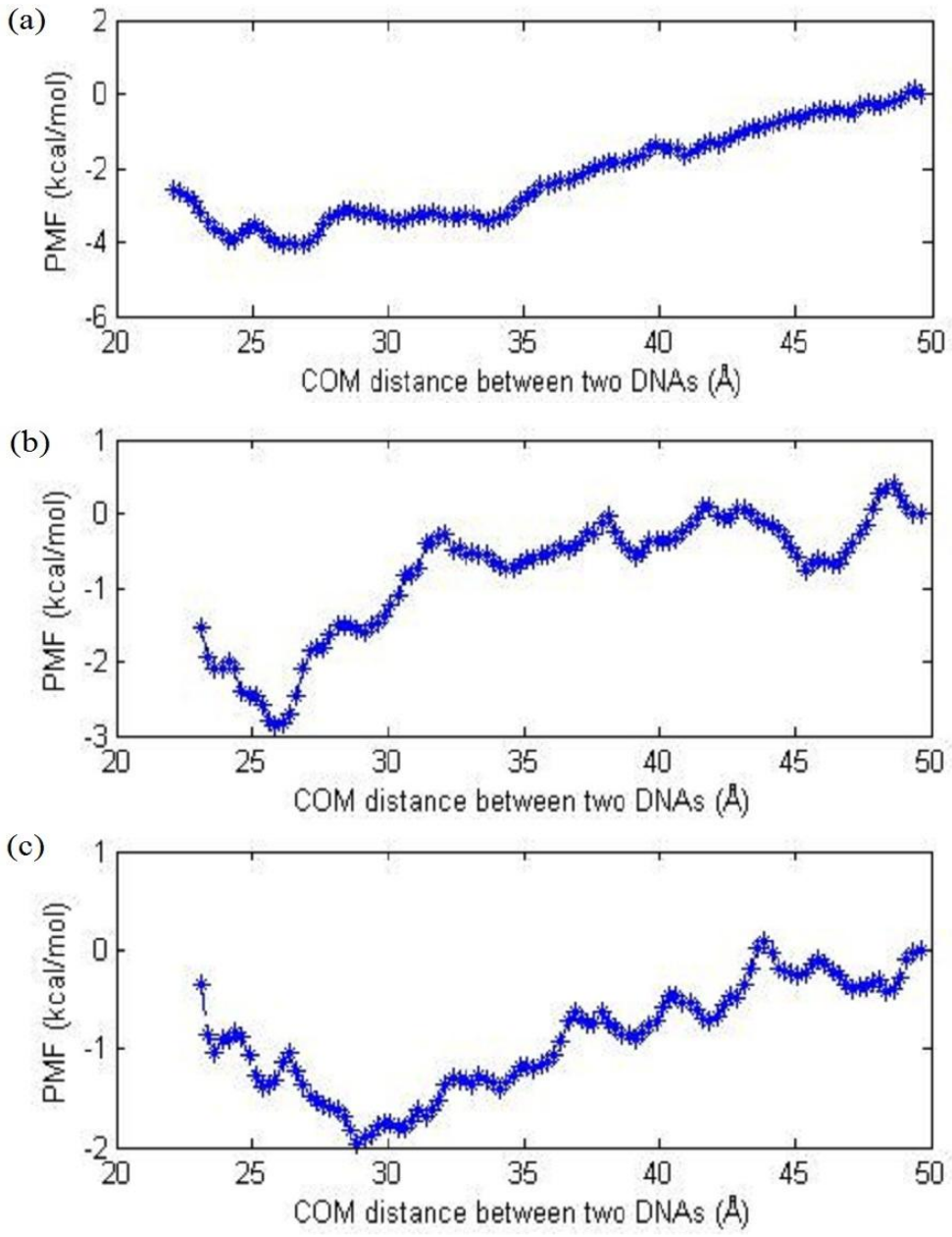


Fig. 4.5 PMF vs. DNA COM distance for systems (a) 2D-6P, (b) 2D-4P and (c) 2D-2P.

of DNA in the cell and nucleus [144, 79]. For example, Guo *et al.* [144] studied the effect of N/P charge ratio on the transfection efficiency of polyplexes formed by DNA with PLL and dioeoylphosphatidylethanolamine/cholesteryl hemisuccinate liposomes. They found that at a constant lipid/DNA ratio, an

increase in the N/P charge ratio (from 0 to 5) rapidly increased the transfection efficiency, and suggested that this result could be applied to PEI as well [144].

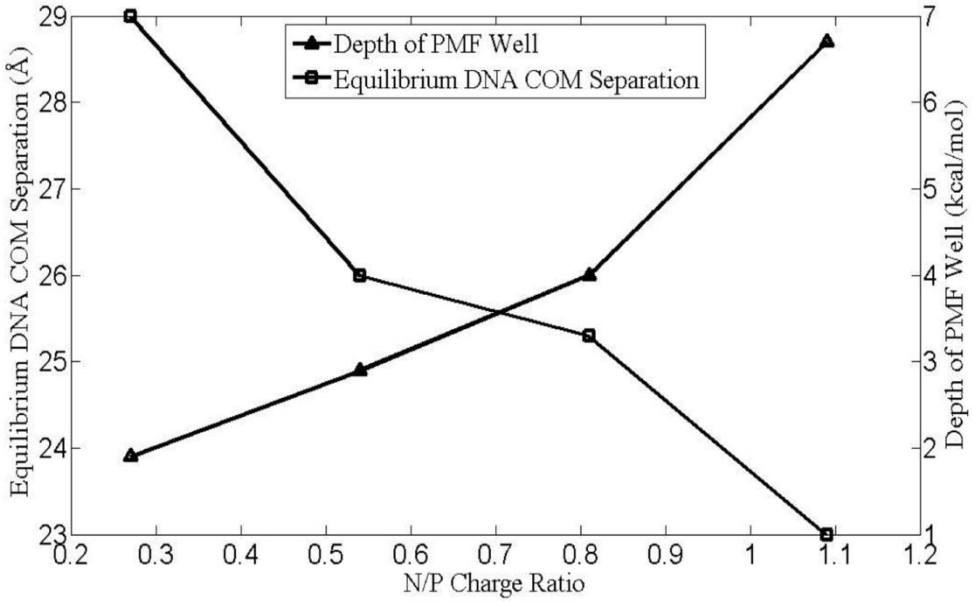


Fig. 4.6 Equilibrium separation between the DNAs and depth of the PMF well plotted against the N/P charge ratio for 46% protonated PEI mediated DNA attraction.

4.3 Effect of PEI protonation ratio

Apart from N/P charge ratio, protonation ratio of the PEI is also a crucial factor in the process of DNA aggregation [41, 70, 66]. In general, it has been observed experimentally that increasing protonation ratio promotes the stability of the aggregation [70, 66]. To study the effect of protonation ratio, we calculated the PMF for system 2D-8P(23%), where 3 out of the 13 nitrogens on the PEI are protonated. The obtained PMF curve is shown in Fig. 4.7. Compared with the PMF (Fig. 4.3) for the 2D-8P system that contains the same number of PEIs but at

a higher protonation ratio (46%), Fig. 4.7 shows a much shallower and wider PMF well with a more or less flat region that ranges from 30 to 35 Å along the reaction coordinate. The global minimum is located at 34.2 Å, signifying 48.7% larger COM separation than that of system 2D-8P. The depth of the PMF well is 2.5 kcal/mol, more than 62% smaller than that of the 2D-8P system. These data demonstrate the formation of significantly smaller and stronger DNA aggregate by PEIs with higher protonation ratio.

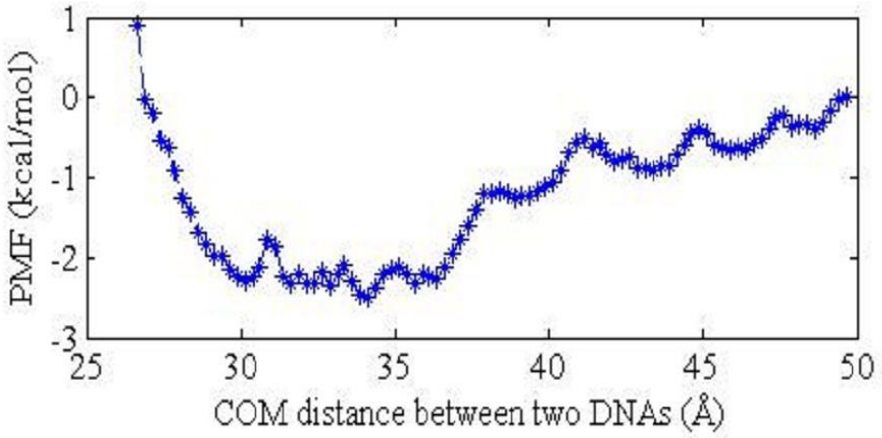


Fig. 4.7 PMF vs. DNA COM distance for 2D-8P(23%) system.

Polycations can aggregate DNAs not only via charge neutralization but also via polyion bridging. Therefore, the total charge that the polycations carry (or the N/P charge ratio) may not be the only factor that affects the properties of the DNA aggregate. To further explore this, we compare in Table 4.3 the PMF calculated for the three systems: 2D-8P, 2D-8P(23%) and 2D-4P. 2D-8P and 2D-8P(23%) contain the same number of PEI molecules, but have different N/P charge ratio due to their different protonation ratios. 2D-8P(23%) and 2D-4P differ in both PEI number and protonation ratio, but have the same overall PEI charges and hence the same N/P charge ratio.

Table 4.3 Comparison of PMF in systems 2D-8P, 2D-8P(23%) and 2D-4P.

System	No. of PEI	Protonation ratio of PEI	N/P charge ratio	Location of global minimum in PMF (Å)	Depth of PMF well (kcal/mol)
2D-8P	8	46%	1.09	23.0	6.7
2D-4P	4	46%	0.54	26.1	2.9
2D-8P (23%)	8	23%	0.54	34.2	2.5

Clearly among the three systems, 2D-8P, the one with the highest protonation ratio and N/P charge ratio, forms the most compact and stable DNA aggregate. The more interesting comparison is between 2D-8P(23%) and 2D-4P. The location of the global minimum in the PMF curve is at 26.1 Å for 2D-4P which is 8.10 Å less than the 34.2 Å obtained for 2D-8P(23%). In addition, the depth of the PMF well for 2D-4P is slightly larger than that for 2D-8P(23%). Therefore, it is clear that at the same N/P charge ratio, 46% protonated PEIs form more stable and much more compact DNA aggregate than 23% protonated PEIs. We can also observe from Table 4.3 that the effect of changing the protonation ratio on the compactness of the aggregate is more prominent than changing the N/P charge ratio. Specifically, changing the N/P charge ratio from 0.54 in 2D-4P to 1.09 in 2D-8P reduces the equilibrium separation by only 11.8%, whereas increasing the protonation ratio, at the same N/P charge ratio, from 23% in 2D-8P(23%) to 46% in 2D-4P reduces it by 23.6%. On the other hand, the depth of the PMF well is more strongly influenced by the N/P charge ratio. In particular, it is increased by

16% on doubling the protonation ratio (from 2D-8P(23%) to 2D-4P) and by 131% on doubling the N/P charge ratio (from 2D-4P to 2D-8P). It is known that one disadvantage of increasing the N/P charge ratio is the increase in cytotoxicity [79]. These observations suggest that by properly selecting the PEI protonation ratio and the N/P charge ratio, a good compromise can be found between the compactness of the DNA aggregate, its stability and cytotoxicity.

Chapter 5

Results for lipid-modified PEI-mediated DNA attraction

In this chapter, we present the PMFs obtained from the MD simulations of lipid-modified PEI-mediated DNA attraction. The results obtained from the systems with three different lipid modifications i.e. oleic acid (OA), linoleic acid (LA) and caprylic acid (CA), are compared. The results are also compared with the PMF obtained for native PEI mediated DNA aggregation, which allows us to assess the effect of different lipids on the stability and compactness of the aggregate.

5.1 PMF for lipid-modified PEI-mediated DNA attraction

As shown in Chapter 3 of Fig. 3.3, among the three lipids studied here, OA and LA are the longer lipids (with 18 carbon atoms) and CA is the shorter lipid (with 8 carbon atoms). OA and LA have the same number of carbon atoms and a similar structure except the presence of an extra double bond in LA. Again 29 simulations were run for every system. Each simulation has a different initial COM separation distance between the two DNAs, which corresponds to the springs in the harmonic restraint being un-stretched. The variation of the COM separation with simulation time, which is caused by the DNA interactions, can be obtained from

the simulation. Fig. 5.1 represents one such curve for which the PEIs were substituted by OA and the initial distance between the COM of the two DNAs was 30 Å. From the data obtained from all 29 simulations, a histogram (Fig 5.2) was constructed on the same grounds as done in the previous chapter.

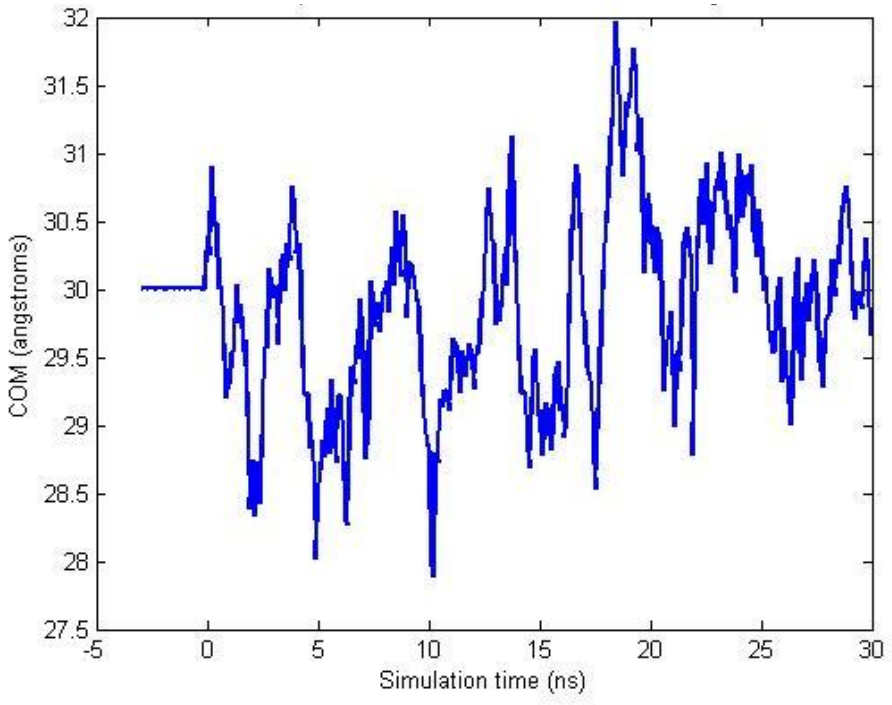


Fig. 5.1 Variation of COM distance between the two DNAs in the presence of eight OA substituted PEIs when the initial DNA COM distance was 30 Å. Graphs for all other initial COM distances are given in appendix 1.

The PMF for system 2D-8P(OA) is plotted along the reaction coordinate as shown in Fig. 5.3. As the COM distance between the two DNAs (x) increases, the curve demonstrates qualitative similarity to the PMF of system 2D-8P. Initially, the PMF curve takes the form of a branch with decreasing slope which is characteristic of the repulsive interactions between the two negatively charged DNA molecules.

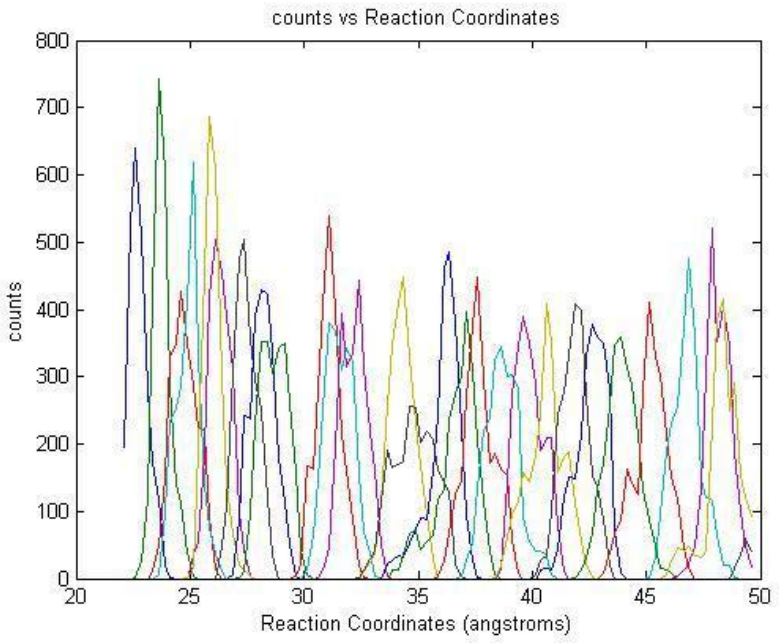


Fig. 5.2 Histogram for the 2D-8P(OA) system showing the number of counts for the entire range of the reaction coordinates. A total of 29 simulations were performed to generate the histogram. Different curves in the figure corresponds to simulations with different biasing potential $V(x)$ (i = 1, 2, …, 29).

The curve then reaches its global minimum at 25.6 Å, beyond which there is an overall increasing trend in the PMF till 44 Å. After 44 Å, the curve fluctuates without overall increase in the PMF, indicating negligible interaction beyond this separation. It can be observed from Fig. 5.3 that the global minimum of the PMF curve is -7.8 kcal/mol and is located at x = 25.6 Å.

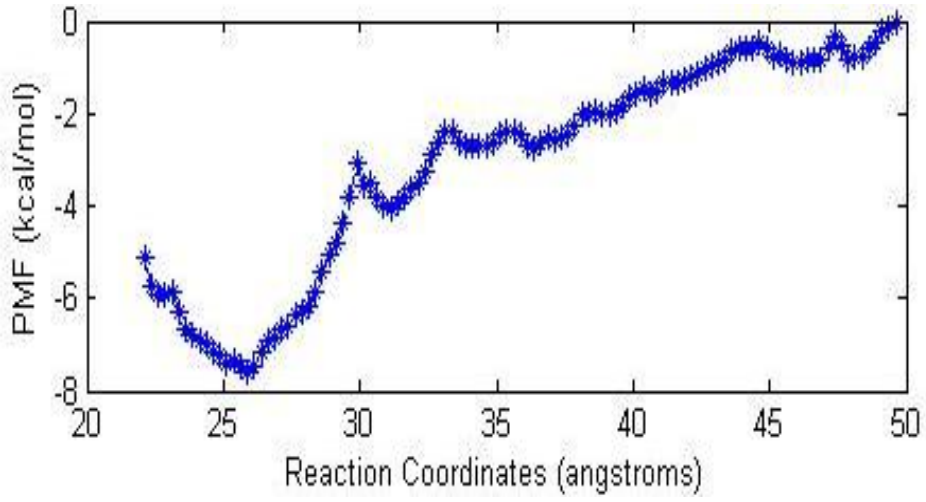


Fig. 5.3 PMF vs. DNA COM distance for system 2D-8P(OA)

The PMF for system 2D-8P(LA) is plotted along the reaction coordinate as shown in Fig. 5.4. The curve demonstrates similar characteristics as that for system 2D-8P(OA) with similar amplitude of fluctuation. On the other hand, the curve for 2D-8P(LA) appears to level off at a larger separation (~ 48 Å compared with ~ 44 Å for 2D-8P(OA)). The curve reaches its global minimum at 24.9 Å with a value of -8 kcal/mol.

The PMF for system 2D-8P(CA) is plotted along the reaction coordinate as shown in Fig. 5.5. The curve reaches its global minimum at 25 Å. After 43 Å, the curve fluctuates without overall increase in the PMF. The global minimum of the PMF curve obtained at the global minimum is -5.6 kcal/mol.

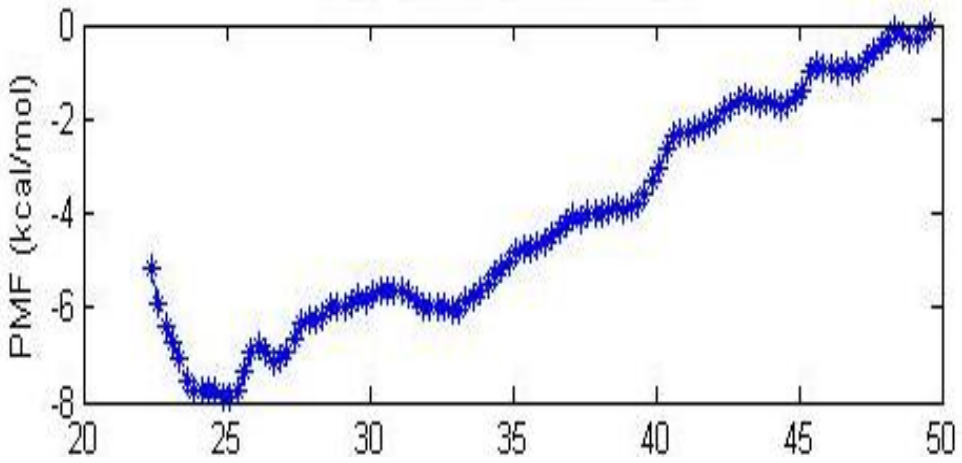


Fig. 5.4 PMF vs. DNA COM distance for system 2D-8P(LA)

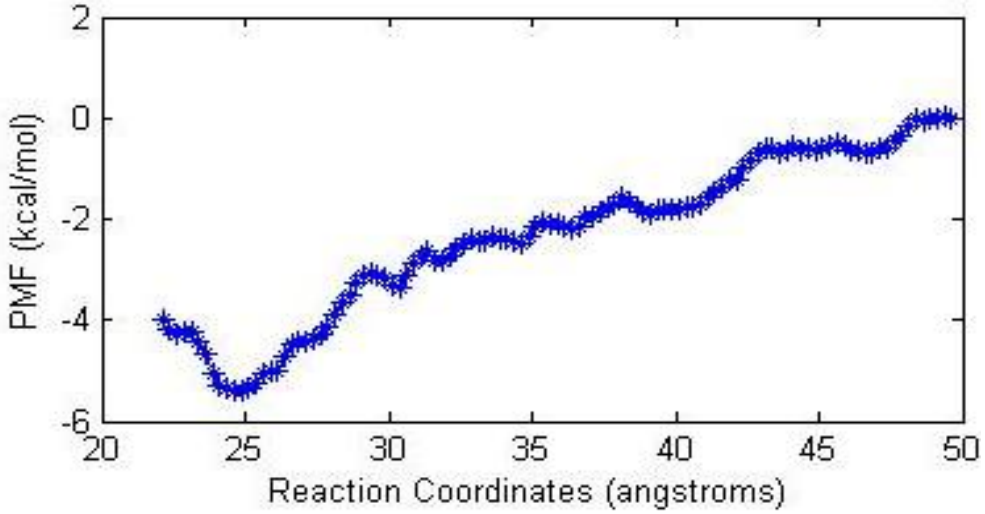


Fig. 5.5 PMF vs. DNA COM distance for system 2D-8P(CA)

5.2 Effect of Lipid Substitution

It has been reported in literature that substituting PEIs with lipids increase their efficiency in gene delivery, which might be due to the increase in the stability of the condensates [80, 81]. In order to verify this, the effect of lipid substitution is studied in this work by comparing the PMF for systems 2D-8P(OA), 2D-8P(LA),

2D-8P(CA) and 2D-8P without lipid. The location of the global minimum in each curve and the depth of the PMF well are given in Table. 5.1.

Table 5.1 Equilibrium separation between the DNAs and depth of the PMF well for lipid substituted and native PEI mediated DNA attraction.

System	Location of global minimum in PMF (Å)	Depth of PMF well (kcal/mol)
2D-8P(OA)	25.6	7.8
2D-8P(LA)	25.0	8.0
2D-8P (CA)	24.9	5.6
2D-8P	23.7	6.7

It can be clearly observed from Table 5.1 that the location of global minimum in the PMF increases on substituting PEI with lipids, 1.2 Å to 2 Å depending on the lipid, suggesting an increase in size of the DNA aggregates. This observation can be attributed to the increase in the size of the PEI molecules due to the lipid substitution and the steric hindrances caused by the lipids which resist the formation of compact structures [89]. It can also be observed by comparing the all four PMF curves (Fig. 5.3 - Fig 5.5 and Fig. 4.3), in case of lipid-modified PEI-mediated DNA attraction there is an increase in fluctuation in the curves, i.e., an increase in the number of local minima along the reaction coordinate. The fluctuation is particularly larger in 2D-8P(OA) and 2D-8P(LA) systems. This may also be due to the fact that the PEIs substituted by lipids, especially by long lipids, are of larger size and can accommodate more configurations than the native PEIs.

As compared to native PEI, for the longer lipids (LA and OA) the depth of PMF curve has significantly increased by 16%-20%; whereas for short lipid (CA) it has significantly decreased by 17%. This indicates that despite the larger size, the DNA aggregate formed by OA & LA substituted PEIs are more stable than native PEI aggregated DNAs. On the other hand, the DNA aggregate mediated by CA substituted PEI is not only larger, but also less stable. A possible explanation to this can be attributed to the mechanisms governing PEI mediated DNA aggregation. Since the charge on each PEI is the same for all four systems, the effect of charge neutralization is the same for all 4 systems. Another common mechanism responsible for DNA aggregation in the presence of the four types of PEIs is polyion bridging. When the PEI is substituted with a lipid, third mechanism known as lipid association (Fig. 5.6) also plays an important role in DNA aggregation [86]. The simulation based study by Sun *et al.* [86] showed that although lipid substitution promotes DNA aggregation through lipid association, it hinders polyion bridging. Hence, the stability of the DNA aggregation is dependent on both mechanisms and also on the steric hindrances caused by lipid substitution [87]. For large lipids like OA and LA, the lipid association is significant [145], which may overcome the adverse effect on DNA aggregation caused by steric hindrances and less formation of polyion bridging, and lead to a more stable aggregate when compared to the native PEI. On the other hand, a small lipid like CA has comparatively lesser extent of lipid association which may not be able to compensate for the decrease in attraction due to lesser polyion bridging and increased steric hindrances. In an experimental study by Neamnark

et al. [85] , the effect of different lipid substitutions on the cellular uptake and transfection efficiency of PEI mediated gene delivery was studied and CA was found to be ineffective in both cellular uptake and transfection as compared to other lipids.

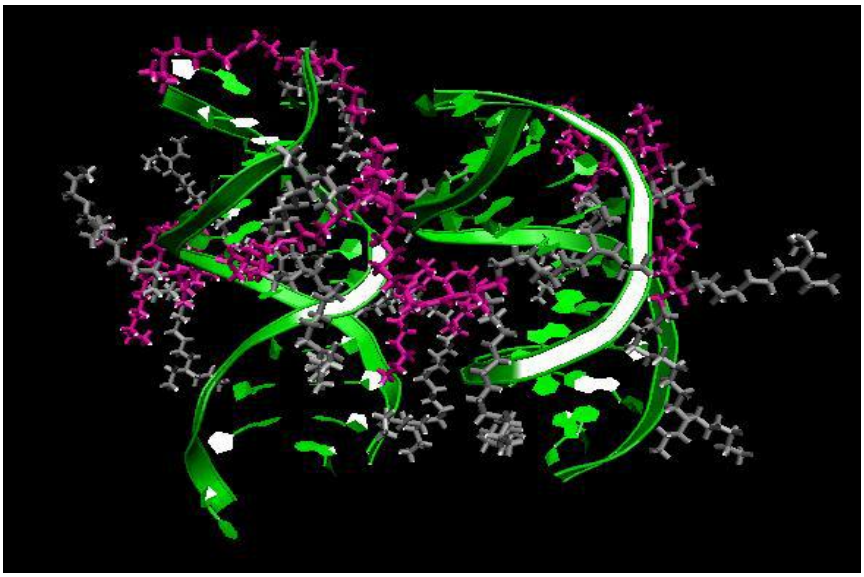


Fig. 5.6 A screenshot of the DNA aggregate depicting lipid association in LA-modified PEI-mediated DNA aggregation. The green molecules represent the two DNA molecules, the grey molecules represent the PEI molecules and the magenta segments in the PEI molecules represent the lipid chains. It can be observed that lipid chains prefer to stay associated during the simulations

Comparing systems 2D-8P(OA) with 2D-8P(LA), their depths of PMF well are similar, while the aggregate formed in 2D-8P(LA) is smaller compared to 2D-8P(OA). Experimentally, Neanmark *et al*. [85] observed that as compared to OA and CA, LA substituted PEIs were more effective in cellular uptake and showed higher transfection efficiency during gene delivery. Hochgraf *et al*. [146]

suggested that LA had a tendency to lower the crystallinity, i.e., cause disorder, of the cell membrane. Roach *et al.* [147] studied the effect of different fatty acids on membrane properties and found that LA could reduce the membrane's phase transition temperature. Both of these effects might increase the rate of motion of the molecules that have to move across the membrane and hence promote cellular uptake [85]. Despite these experimental studies, it was not clear whether there was a fundamental difference between the DNA aggregates formed by OA substituted PEIs and those formed by LA substituted PEIs. It is encouraging to see from our simulations that LA substituted PEIs were able to form more compact DNA aggregates compared with OA, although the molecular mechanism behind this observation requires further investigation.

Recently, there was a simulation work performed by Sun *et al.* [145], where the effect of different lipid substitutions was studied on PEI mediated siRNA (small-interfering RNA) complexation. In particular, they simulated LA and CA substituted PEI mediated siRNA aggregation without applying any restraints on the molecules. Each simulated system involves 4 siRNAs and 18 lipid modified PEIs, and for each system (a particular type of substitution at a given substitution level), one unrestrained simulation was performed for a long time (200 ns). Compared with their work, our results show a few similarities but also some differences. Specifically, they observed that as the length of the lipid increases (from CA to LA), siRNA aggregate became more compact and stable. On the other hand, it was reported that the siRNA aggregates mediated by lipid modified

PEIs, whether the substitution was by LA or CA, were more compact and stable than those by native PEIs. Whereas in the present work, CA substituted PEIs was not able to form more compact or stable aggregate, and all aggregates formed by lipid modified PEIs are larger than that formed by native PEIs. Several factors can contribute to such difference. Firstly, the native PEIs simulated in this work are around 600 Da while the native PEIs in [145] are around 2kDa. The larger PEIs in [145] may help form stronger polyion bridging and lipid association. Secondly, the aggregates formed in [145] involve more molecules (4 siRNAs and 18 PEIs). Although compared with native PEIs, CA substituted PEIs cannot form more stable aggregate for two DNAs, when they collectively aggregate more DNAs and form a network where all the DNAs are mutually connected, the effect of lipid association can become more pronounced, and hence, increase the stability of the aggregate more significantly. It should be noted that the stability of the aggregate in [145] was assessed by examining the fluctuation in the radius of gyration of the siRNA molecules, which was not as quantitative as the depth of PMF used in the present work. However, defining PMF for more than two DNAs or RNAs is challenging since it is not clear what would be the best choice for the reaction coordinate.

Conclusions and Future Work

6.1 Conclusions

The aggregation of DNA molecules induced by cationic polymers is of importance to applications in gene delivery. In this work, umbrella sampling MD simulations were performed to study the interaction of two DNAs in the presence of PEI molecules. Because modifying PEI with lipids increases its efficiency in DNA delivery and reduces the cytotoxicity, simulations were performed for systems involving both native PEIs and PEIs modified with three different lipids, i.e. linoleic acid, oleic acid and caprylic acid. The PMF of interaction between the two DNAs was calculated using WHAM with the reaction coordinate being the COM distance between the two DNAs. The calculations allow us to quantitatively assess PEI mediated DNA attraction and address the effect of N/P charge ratio, PEI's protonation state and lipid substitution. To the best of our knowledge this is the first work to perform quantitative analysis of polycation (specifically PEI) mediated DNA aggregation using potential of mean force calculations.

Compared with small multivalent ions, PEIs give rise to stronger DNA attraction even with an N/P charge ratio at which the DNAs are over neutralized. As the N/P charge ratio and/or the protonation ratio of the PEI increases the DNA aggregate becomes more compact and stable. The effect of changing the protonation ratio on

the compactness of the aggregate is more prominent than changing the N/P charge ratio, while the depth of the PMF well is more strongly influenced by the N/P charge ratio. Comparing the depth of PMF well it was found that of the three lipid substitutions studied here, linoleic acid produced the most stable aggregates. On the other hand, it was observed that substitution with shorter lipid, like caprylic acid, was not effective in aggregating the DNA molecules. Compared to native PEIs, substitution with longer lipids (linoleic acid and oleic acid) showed larger but significantly more stable aggregates. Although in this study we have used PEI as a representative polycation, the results obtained here may be applicable to other polycations mediated DNA aggregation.

6.2 Limitations and Future Work

The simulation systems used in this study are different from experiments in few ways. Firstly, in this study, the interaction of two DNA molecules was under focus; whereas in an experimental system there would be many DNAs interacting with one another. Secondly, the size of the DNA molecules and PEI molecules in experimental setup are much larger than what is considered in the simulation models. Also the N/P charge ratio studied in this study was very small as compared to the practical purposes for which N/P charge ratio is greater than 10. For the case of native PEI-mediated DNA attraction, we have compared two different parameters; an interesting study would be to analyze the effect of various other factors (e.g. size of PEI, degree of branching etc) and also check for cases

with N/P ratio>1 to study the effect of excess PEIs. For the case of lipid-modified PEI-mediated DNA attraction, one interesting study would be to study in detail why linoleic acid substitution has different results as compared to oleic acid with such a small difference in their molecular structure. An addition to this can be to use high molecular weight PEIs, which can show more distinguishable results than small PEIs.

References

[1] V. A. Bloomfield, *Biopolymers,* vol. 44, pp. 269-282, 1998.

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

[3] M. Morita, M. Tasaka and H. Fujisawa, *Virology,* vol. 193, pp. 748-752, 1993.

[4] V. Vijayanathan, T. Thomas, A. Shirahata and T. Thomas, *Biochem,* vol. 40, pp. 13644-13651, 2001.

[5] F. D. Ledley, *Hum. Gene Ther,* vol. 6, pp. 1129-1144, 1995.

[6] J. Wolff and J. Lederberg, *Hum. Gene Ther.,* vol. 5, pp. 469-480, 1994.

[7] F. Mühlbacher, H. Schiessel and C. Holm, *Phys. Rev. E.,* vol. 74, p. 031919, 2006.

[8] R. Dame, *Mol. Microbiol.,* vol. 56, pp. 858-870, 2005.

[9] A. Travers and G. Muskhelishvili, *Curr. Opin. Genet. Dev.,* vol. 15, pp. 507-514, 2005.

[10] W. Earnshaw, J. King, S. Harrison and F. Eiserling, *Cell,* vol. 14, pp. 559-568, 1978.

[11] S. Klimenko, T. Tikchonenko and V. Andreev, *J. Mol. Biol.,* vol. 23, pp. 523-533, 1967.

[12] K. Richards, R. Williams and R. Calendar, *J. Mol. Biol.,* vol. 78, pp. 255-259, 1973.

[13] L. Gosule and J. Schellman, *Nature,* vol. 259, pp. 333-335, 1976.

[14] R. W. Wilson and V. A. Bloomfield, *Biochem.,* vol. 18, pp. 2192-2196, 1979.

[15] J. Widom and R. Baldwin, *J. Mol. Biol.,* vol. 1980, pp. 431-453, 1980.

[16] J. A. Schellman and N. Parthasarathy, *J. Mol. Biol.,* vol. 175, pp. 313-329, 1984.

[17] O. V. Zribi, H. Kyung, R. Golestanian, T. B. Liverpool and G. C. L. Wong, *Phys. Rev. E,*

vol. 73, pp. 031911-031920, 2006.

[18] L. Dai, Y. Mu, L. Nordenskiold and J. R. C. v. d. Maarel, *Phys. Rev. Lett.,* vol. 100, pp. 118301-118304, 2008.

[19] E. Raspaud, I. Chaperon, A. Leforestier and F. Livolant, *Biophys. J.,* vol. 77, pp. 1547-1555, 1999.

[20] D. K. Chattoraj, L. C. Gosule and J. A. Schellman, *J. Mol. Biol,* vol. 1971, pp. 327-337, 1978.

[21] A. C. Toma, M. Frutos, F. Livolant and E. Raspaud, *Biomacromolecules,* vol. 10, pp. 2129-2134, 2009.

[22] E. Wagner, M. Cotten, R. Foisner and M. Birnstiel, *P. Natl. Acad. Sci. USA,* vol. 10, pp. 4255-4259, 1991.

[23] H. G. Hansma, R. Golan, W. Hsieh, C. P. Lollo, P. Mullen-Ley and D. Kwoh, *Nucleic Acids Res.,* vol. 26, pp. 2481-2487, 1998.

[24] T. Sakaue and K. Yoshikawa, *J. Chem. Phys.,* vol. 117, pp. 6323-6330, 2002.

[25] T. T. Nguyen and B. I. Shklovskii, *Phys. Rev. E.,* vol. 65, pp. 031409-1-031409-7, 2002.

[26] I. Rouzina and V. A. Bloomfield, *Biophys. Chem.,* vol. 64, pp. 139-155, 1997.

[27] B. Shklovskii, *Phys. Rev. Lett.,* vol. 82, pp. 3268-3271, 1999.

[28] M. J. Stevens, *Phys. Rev. Lett.,* vol. 82, pp. 101-104, 1999.

[29] P. J. Hagerman, *Annu. Rev. Biophys. Chem.,* vol. 17, pp. 265-286, 1988.

[30] M. Olvera de la Cruz, L. Belloni, M. Delsanti, J. P. Dalbiez, O. Spalla and M. Drifford, *J. Chem. Phys.,* vol. 103, pp. 5781-5791, 1995.

[31] W. M. Gelbart, R. F. Bruinsma, P. A. Pincus and V. A. Parsegian, *Phys. Today,* vol. 53, pp. 38-44, 2000.

[32] E. M. Mateescu, C. Jeppensen and P. Pincus, *Europhys. Lett.,* vol. 46, pp. 493-498, 1999.

[33] D. Rau and V. Parsegian, *Biophys. J.,* vol. 61, pp. 246-259, 1992.

[34] F. Oosawa, *Biopolymers,* vol. 6, pp. 1633-1647, 1968.

[35] G. S. Manning, *Q. Rev. Biophys.,* vol. 11, pp. 179-246, 1978.

[36] C.-M. Wu, C.-Y. Chen, S.-Y. Lin and H.-L. Chen, *React. Funct. Poly.,* vol. 71, pp. 266-271, 2011.

[37] P.-E. Sottas, E. Larquet, A. Stasiak and J. Dubochet, *Biophys. J.,* vol. 77, pp. 1858-1870, 1999.

[38] C. Sun, T. Tang and H. Uludağ, *Biomacromolecules,* vol. 12, p. 3698–3707, 2011.

[39] A. Kornyshev and S. Leikin, *Phys. Rev. Lett.,* vol. 82, pp. 4138-4141, 1999.

[40] D. D. Dunlap, A. Maggi, M. R. Soria and L. Monaco, *Nucleic Acids Res.,* vol. 25, pp. 3095-3101, 1997.

[41] W. Godbey, K.K.Wu and A. Mikos, *J. Biomed. Mater. Res.,* vol. 45, p. 268–275, 1999.

[42] O. Boussif, F. Lezoualch, A. Z. M., D. M. M. and D. Scherman, *Proc. Natl. Acad. Sci.,* vol. 92, p. 7297–7301, 1995.

[43] C. Braun, J. Vetro, D. Tomalia, G. Koe, J. Koe and C. Middaugh, *J. Pharm. Sci.,* vol. 94, pp. 423-436, 2005.

[44] O. Boussif, M. Zanta and J.-P. Behr, *Gene Ther.,* vol. 3, pp. 1074-1080, 1996.

[45] H. Dong, L. Ding, F. Yan, H. Ji and H. Ju, *Biomaterials,* vol. 32, pp. 3875-3882, 2011.

[46] W. Godbey, K. Wu and A. Mikos, *J. Control. Release,* vol. 60, pp. 149-160, 1999.

[47] V. P. Torchilin, Drug Delivery, Berlin Heidelberg: Springer , 2010, pp. 3-53.

[48] S. Grosse, Y. Aron, I. Honore, G. Thevenot, C. Danel, A. Roche, M. Monsigny and I. Fajac, *J. Gene. Med.,* vol. 6, pp. 345-356, 2004.

[49] P. Erbacher, J. Remy and J. Behr, *Gene Ther.,* vol. 6, pp. 138-145, 1999.

[50] R. Kircheis, W. L., A. Schreiber, B. Robitza, V. Rossler, M. Kursa and E. Wagner, *Gene Ther.,* vol. 8, pp. 28-40, 2001.

[51] M. Marsh and H. McMahon, *Science,* vol. 285, pp. 215-220, 1999.

[52] I. Kopatz, J. Remy and J. Behr, *J. Gene. Med.,* vol. 6, pp. 769-776, 2004.

[53] C. Evans and R. Aguilera, *Gene,* vol. 322, pp. 1-15, 2003.

[54] J. Behr, *Bioconjug. Chem.,* vol. 5, pp. 382-389, 1995.

[55] N. Sonawane, F. Szoka and A. Verkman, *J. Biol. Chem.,* vol. 278, pp. 44826-44831, 2003.

[56] G. Lukacs, P. Haggie, O. Seksek, D. Lechardeur, N. Freedman and A. Verkman, *J. Biol. Chem,* vol. 275, pp. 1625-1629, 2000.

[57] H. Pollard, G. Toumaniantz, J. Amos, H. Avet-Loiseau, G. Guihard, J. Behr and D. Escande, *J. Gene. Med.,* vol. 3, pp. 153-164, 2001.

[58] B. Talcott and M. Moore, *Trends Cell. Biol.,* vol. 9, pp. 312-318, 1999.

[59] S. Brunner, T. Sauer, S. Carotta, M. Cotten, M. Saltik and E. Wagner, *Gene Ther.,* vol. 7, pp. 401-407, 2000.

[60] A. Akinc, M. Thomas, A. Klibanov and R. Langer, *J. Gene. Med.,* vol. 7, pp. 657-663, 2005.

[61] H. Thomas, H.-T. Nelly, C. Marc and R. Didier, *Colloid. Surface.,* vol. 163, pp. 71-80, 2000.

[62] R. Mahato, L. Smith and A. Rolland, Advances in Genetics, New York: Advances in Genetics, 1999.

[63] M. Molas, J. C. Bartrons and J. C. Perales, *Biochimica et Biophysica Acta ,* vol. 1572, pp. 37-44, 2002.

[64] Y. Ren, X. Jiang, P. Deng and H. Mao, *Biomacromolecules ,* vol. 11, pp. 3432-3439, 2010.

[65] S. Prabha, W. Zhou, J. Panyam and V. Labhasetwar, *Int. J. Pharm.,* vol. 244, no. 105-115, 2002.

[66] K. Utsuno and H. Uludag., *Biophys. J.,* vol. 97, pp. 1971-1983, 2010.

[67] D. Fischer, T. Bieber, Y. Li, H.-P. Elsässer and T. Kissel, *Pharmaceut. Res.,* vol. 16, pp. 1273-1279, 1999.

[68] J. Yu, J. Quan, J. Huang, J. Nah and C. Cho, *J. Mater. Sci.,* vol. 20, pp. 2501-2510, 2009.

[69] K. Kunath, A. Von Harpe, D. Fischer, H. Petersen, U. Bickel, K. Voigt and T. Kissel, *J. Control. Release,* vol. 89, pp. 113-125, 2003.

[70] C. Sun, T. Tang, H. Uludağ and J. Cuervo, *Biophys. J.,* vol. 100, pp. 2754-2763, 2011.

[71] A. Kichler, J. Behr and P. Erbacher, Nonviral Vectors for Gene Therapy, San Diego: Academic Press, 1999.

[72] B. Brissault, A. Kichler, C. Guis, C. Leborgne, O. Dalos and H. Cheradame, *Bioconjug. Chem.,* vol. 14, pp. 581-587, 2003.

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

[74] A. Bragonzi, A. Boletta, A. Biffi, A. Muggia, G. Sersale, S. Cheng, C. Bordignon, B. Assael and M. Conese, *Gene Ther.,* vol. 6, pp. 1995-2004, 1999.

[75] L. Wightman, R. Kircheis, V. Rössler, S. Garotta, R. Ruzicka, M. Kursa and E. Wagner, *J. Gene Med.,* vol. 3, pp. 362-372, 2001.

[76] K. Minagawa, Y. Matsuzawa, K. Yoshikawa, M. Matsumoto and M. Doi, *FEBS J.,* vol. 295, pp. 67-69, 1991.

[77] Y. Oh, D. Suh, J. Kim, H. Choi, K. Shin and J. Ko, *Gene Ther.,* vol. 9, pp. 1627-1632, 2002.

[78] Q. Xie, G. Xinyong, C. Xianjin and W. Yayu, *Cytotechnology,* vol. Forthcoming, pp. 1-9, 2012.

[79] Q. Q. Zhao, J. L. Chen, T. F. Lv, C. X. He, G. P. Tang, W. Q. Liang, Y. Tabata and J. Q. Gao, *Biol. Pharm. Bull.,* vol. 32, pp. 706-710, 2009.

[80] V. Incani, A. Lavasanifar and H. Uludağ, *Soft Matter,* vol. 6, pp. 2124-2138, 2010.

[81] Z. Liu, Z. Zhang, C. Zhou and Y. Jiao, *Prog. Polym. Sci.,* vol. 35, pp. 1144-1162, 2010.

[82] V. Incani, X. Lin, A. Lavasanifar and H. Uludağ, *ACS Appl. Mater. Interfaces,* vol. 1, pp. 841-848, 2009.

[83] K. C. R. Bahadur, B. Landry, H. M. Aliabadi, A. Lavasanifar and H. Uludag, *Acta.*

Biomater., vol. 7, pp. 2209-2217, 2011.

[84] X. Chen, Z. Yuan, X. Yi, R. Zhuo and F. Li, *Nanotechnology,* vol. 23, p. 415602, 2012.

[85] A. Neamnark, O. Suwantong, R. K. C. Bahadur, C. Y. M. Hsu, P. Supaphol and H. Uludag, *Mol. Pharmaceut.,* vol. 6, pp. 1798-1815, 2009.

[86] C. Sun, T. Tang and H. Uludağ, *Biomacromolecules,* vol. 13, pp. 2982-2988, 2012.

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

[88] C. Hsu and H. Uludaĝ, *Biomater.,* vol. 33, pp. 7834-7848, 2012.

[89] M. Patel and T. Anchordoquy, *Biophys. J.,* vol. 88, pp. 2089-2103, 2005.

[90] T. Batuğ, P.-C. Chen, S. Patra and S. Kuyucak, *J. Chem. Phys.,* vol. 128, p. 155104, 2008.

[91] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman and J. M. Rosenberg, *J. Comp. Chem,* vol. 13, p. 1011–1021, 1992.

[92] G. Torrie and J. Valleau, *J. Comp. Chem.,* vol. 23, pp. 187-199, 1977.

[93] C. Kittel and H. Kroemer, Thermal Physics, San Francisco : W.H. Freeman and Company, 1980.

[94] D. A. Beard and H. Qian, Cambridge: Cambridge University Press, 2008.

[95] L. Landau and E. Lifshitz, Statistical Physics, Oxford: Butterworth-Heinemann, 1996.

[96] J. Walecka, Introduction to Statistical Mechanics, Singapore: World Scientific, 2011.

[97] L. D. Landauĭ, Statistical Physics, 1938.

[98] M. Shirts, D. Mobley and J. Chodera, *Annu. Rep. Comput. Chem.,* vol. 3, pp. 41-59, 2007.

[99] S. Park, F. Khalili-Araghi, E. Tajkhorshid and K. Schulten, *J. Chem. Phys,* vol. 119, pp. 3559-3566, 2003.

[100] D. Frenkel and B. Smit, Understanding Molecular Simulation From Algorithm to Application, San Diego: Academic Press, 1996.

[101] C. Alemán and S. Galembeck, *Chem. Phys.,* vol. 232, pp. 151-159, 1998.

[102] B. J. Alder and T. Wainwright, *J. Chem. Phys. ,* vol. 31, pp. 459-466, 1959.

[103] L. Verlet, *Phys. Rev.,* vol. 159, pp. 98-103, 1967.

[104] D. Beeman, *J. Comp. Phys.,* vol. 20, pp. 130-139, 1976.

[105] H. Abe, W. Braun, T. Noguti and N. Gō, *Comp. Chem.,* vol. 8, pp. 239-247, 1984.

[106] A. Jain, N. Vaidehi and G. Rodriguez, *J. Comp. Phys.,* vol. 106, pp. 258-268, 1993.

[107] T. Noguti and N. Gō, *J. Phys. Soc. Japan,* vol. 52, pp. 3685-3690, 1983.

[108] R. Ruth, *Nuclear Sci.,* vol. 30, pp. 2669-2671, 1983.

[109] T. Darden, D. York and L. Pedersen, *J. Chem. Phys.,* vol. 98, pp. 10089-10092, 1993.

[110] O. De Souza, *J. Biomol. Struct. Dyn,* vol. 16, p. 1205, 1999.

[111] J. Israelachvili, Intermolecular and Surface Forces, Academic Press, 1985-2004.

[112] J. E. Lennard-Jones, *Proc. R. Soc. Lond.,* vol. 106, pp. 463-477, 1924.

[113] P. Ewald, *Ann. Phys.,* vol. 369, pp. 253-287, 1921.

[114] M. Allen and D. Tildesley, Computer Simulation of Liquids, Oxford University Press, 1987.

[115] J. G. Kirkwood, *J. Chem. Phys.,* vol. 3, pp. 300-313, 1935.

[116] J. G. Kirkwood, Theory of Liquids, Spring Branch: Routledge, 1968.

[117] M. R. Mruzik, F. F. Abraham, D. E. Schreiber and G. M. Pound, *J. Chem. Phy,* vol. 64, p. 481–491, 1975.

[118] M. Mezei, *Mol. Phys.,* vol. 47, p. 1307–1315, 1982.

[119] R. W. Zwanzig, *J. Chem. Phys.,* vol. 22, pp. 1420-1426, 1954.

[120] W. L. Jorgensen and C. Ravimohan, *J. Chem. Phys.,* vol. 83, p. 3050–3054, 1985.

[121] C. Chipot and A. Pohorille, Free Energy Calculations: Theory and Applications in Chemistry and Biology, New York: Springer, 2007.

[122] B. Widom, *J. Chem. Phy,* vol. 39, p. 2808–2812, 1963.

[123] B. L. Tembe and J. A. McCammon, *Comput. Chem.,* vol. 8, pp. 281-283, 1984.

[124] C. H. Bennett, *J. Chem. Phy,* vol. 22, p. 245–268, 1976.

[125] A. R. Leach, Molecular Modelling: Principles and Applications, Harlow: Addison Wesley Longman, 1998.

[126] B. Roux, *Comp. Phys. Commun.,* vol. 91, pp. 275-282, 1995.

[127] A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.,* vol. 63, p. 1195–1198, 1989.

[128] A. Ferrenberg and R. Swendsen, *Phys. Rev. Lett.,* vol. 61, pp. 2635-2638, 1988.

[129] G. Patey and J. Valleau, *J. Chem. Phys.,* vol. 63, pp. 2334-2339, 1975.

[130] V. Tereshko, G. Minasov and M. Egli, *J. Am. Chem. Soc.,* vol. 121, pp. 470-471, 1999.

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

[132] J. Phillips, G. Zheng, S. Kumar and L. Kale, "NAMD: Biomolecular Simulation on Thousands of Processors," in *SC 2002*, 2002.

[133] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes and I. M. A. Vorobyov, *J. Comput. Chem.,* vol. 31, pp. 671-690, 2010.

[134] A. MacKerel Jr., C. Brooks III, L. Nilsson, B. Roux and Y. Won, The Encyclopedia of Computational Chemistry, vol. 1, Chichester: John Wiley & Sons, 1998, pp. 271-277.

[135] W. L. Jorgensen, *J. Am. Chem. Soc.,* vol. 103, pp. 335-340, 1981.

[136] A. D. J. MacKerell, D. Bashford, R. L. Bellott, R. L. J. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo and D. T. Nguyen, *J. Phys. Chem.,* vol.

102, pp. 3586-3616, 1998.

[137] T. Rog, *Acta. Biochim. Pol.,* vol. 50, p. 789, 2003.

[138] J. Ryckaert, G. Ciccotti and H. Berendsen, *J. Comput. Phys.,* vol. 23, pp. 327-341, 1977.

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

[140] S. Khalid, P. Bond, J. Holyoake, R. Hawtin and M. Sansom, *J. Roy. Soc. Interface,* vol. 5, pp. 241-250, 2008.

[141] M. Abbas, H. Uludag, V. Incani, C. Hsu and A. Jeffery, *Biomacromolecules,* vol. 9, pp. 1618-1630, 2008.

[142] C. K. Young, H. S. Jae, H. K. Young, K. Seong-Chun, J. L. Sang, R. K. Seung, W. K. Dong, P. Seon-Mee, J. Y. Sei, L. Sang-Jeon, X. De-Gang, X. Wen-Xie and W. K. Ki, *J. Korean Med. Sci. ,* vol. 22, pp. 48-56, 2007.

[143] N. Kasyanenko and D. Afanasieva, vol. 16, Netherlands: Springer, 2008, pp. 29-38.

[144] W. Guo and R. Lee, *Bioscience Rep.,* vol. 20, pp. 419-432, 2000.

[145] C. Sun, T. Tang and H. Uludag, *Biomaterials,* p. in press, 2013.

[146] E. Hochgraf, S. Mokady and U. Cogan, *J. Nutr.,* vol. 127, pp. 681-686, 1997.

[147] C. Roach, S. E. Feller, J. A. Ward, S. R. Shaikh, M. Zerouga and W. Stillwell, *Biochem.,* vol. 43, pp. 6344-6351, 2004.

Appendix

A-1 Matlab Code for Performing WHAM Calculations

```
clear all
clc

n_windows=textread('input.txt','Nsimulation
%d',1,'headerlines',0);

%spring constant
k=0;
x_0=0;
for m=1:n_windows
[filen(m) x_0(m) k(m)] = textread('input.txt','%s %f
%f',1,'headerlines',5+m);
end
 filen = char(filen);

 bin_first=textread('input.txt','minbin %f',1,'headerlines',1);
 bin_last=textread('input.txt','maxbin %f',1,'headerlines',2);
 bin_width=textread('input.txt','binwidth %f',1,'headerlines',3);
 first_frame=textread('input.txt','start_frame
%d',1,'headerlines',4);
 last_frame=textread('input.txt','end_frame
%d',1,'headerlines',5);

 %making histogram

 %making bin

 no_bins=(bin_last-bin_first)/bin_width;
 bin_value=[];
 for m=1:no_bins
     bin_value(m)= ((bin_first+((m-
1)*bin_width)+bin_first+((m)*bin_width))/2);
 end

 for m=1:no_bins
bin_edges(m) = bin_first + (m-1)*bin_width;
end

 %entering data
 for m=1:n_windows
     frame = [];
     reac = [];
 [frame reac]=textread(filen(m,:),'%f %f');
 n_frames = size (frame,1);

 use_frame=[];
```

```
 use_reac=[];
 co=1;
 for n=1:n_frames
    if ((frame(n)<=last_frame)&&(frame(n)>=first_frame))
    use_frame(co,:) = frame(n);
    use_reac(co,:) = reac(n);
    co=co+1;
 end
 end

 %counting


n=histc(use_reac,bin_edges);

counts1(:,m)= n(:,1);
 end

 index = 1;
 for i=1:no_bins
     flag=1;
     for m=1:n_windows
       if (counts1(i,m)~=0)
           flag=0;
       end
     end
     if(flag==0)
         counts(index,:) = counts1(i,:);
         bins(index,:)=bin_value(i);
         index=index+1;
     end
 end

 n_bins = index-1;
%K_b(kcal/molK) and T (K)
k_B = 0.001982923700;
T = 300;

%define maximum number of iterations and tolerance value
max_iterations = 100000;
error_tol = 0.0000001;

i=0;
error = 1;

%find total counts in every simulation
totalcounts = sum(counts);

%initialize r_i r_oldi
for c= 1:n_windows
    r_i(c)=2;
    r_oldi(c)=0;
end
```

```
while ((i<max_iterations)&&(error>error_tol))

    %copy all the r_i into r_oldi
    for c=1:n_windows
        r_oldi(c)=r_i(c);
    end

    i=i+1;
    %find numerators for every bin
    for b = 1:n_bins
        %find numerator for the given bin (sum counts over all
windows)
        numerator = sum(counts(b,:));
        denominator = 0;
        %next step is to find the denominator
        for m = 1:n_windows
            bias = 0.5*k(m)*((bins(b)-x_0(m))^2);
            denominator = denominator +
totalcounts(m)*(1/r_i(m))*exp((0-bias)/(k_B*T));
        end
        prob(b)=numerator/denominator;
    end
    for m=1:n_windows
       summation= 0;
       for b=1:n_bins
           bias = 0.5*k(m)*((bins(b)-x_0(m))^2);
           summation = summation + prob(b)*exp((0-bias)/(k_B*T));
       end
       r_i(m) = summation;
    end

    %find error
    for m=1:n_windows
        temperror(m)=abs(r_i(m)-r_oldi(m));
        end
    error = min(temperror);

end
r_i;
sum=0;
for b=1:n_bins
sum=sum+prob(b);
end
sum;
con=1/sum;
su=0;
%adjust probability and find pmf
for b=1:n_bins
n_pro(b)=con*prob(b);
su=su+n_pro(b);
pmf(b)= 0-(k_B*T*log(n_pro(b)));
end
su;
pmf;
n_pro
```

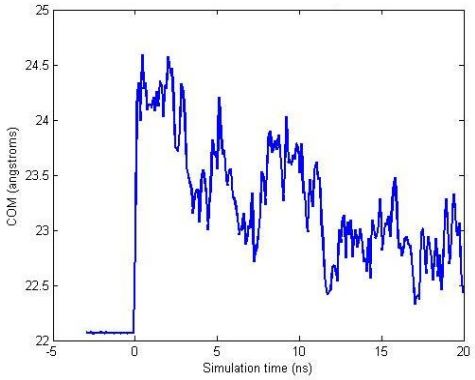
```
%adjust ratio & find free enrgy difference
for m=1:n_windows
    ad_ri(m)=con*r_i(m);
    free_diff(m)=0-(k_B*T*log(ad_ri(m)));
end
ad_ri
free_diff
%shift pmf
for b=1:n_bins
    ad_pmf(b)=pmf(b)-pmf(n_bins);
end

plot (bins, ad_pmf,'--*')
xlabel('COM distance between two DNAs (Å)')
ylabel('PMF (kcal/mol)')


figure
plot (bins, counts)
xlabel('COM distance between two DNAs (Å)')
ylabel('Counts')
```

Sample Input File

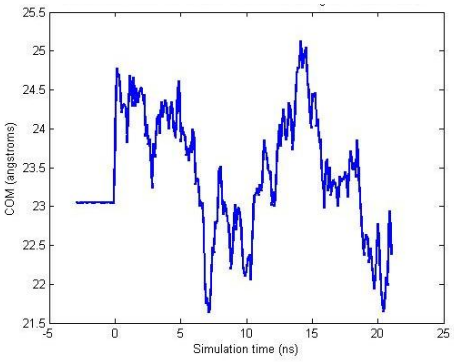
```
Nsimulation 18
minbin 22
maxbin 40
binwidth 0.25
start_frame 5386
end_frame 10386
DPD_22   22.0 1.0
DPD_23   23.0 1.0
DPD_24   24.0 1.0
DPD_25   25.0 1.0
DPD_26   26.0 1.0
DPD_27   27.0 1.0
DPD_28   28.0 1.0
DPD_29   29.0 1.0
DPD_30   30.0 1.0
DPD_31   31.0 1.0
DPD_32   32.0 1.0
DPD_33   33.0 1.0
DPD_34   34.0 1.0
DPD_35   35.0 1.0
DPD_36   36.0 1.0
DPD_37   37.0 1.0
DPD_39   39.0 1.0
DPD_38   38.0 1.0
DPD_40   40.0 1.0
```

A-2 COM distance vs. simulation time graphs

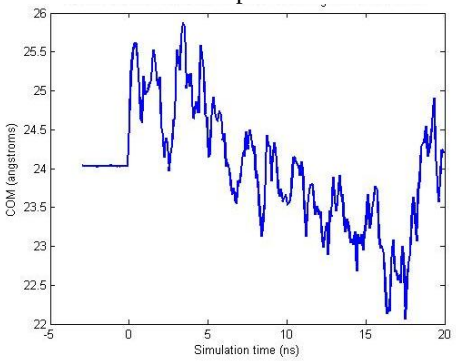
System 2D-8P



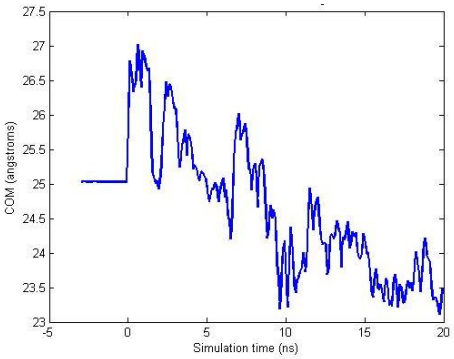
COM initial separation = 22 Å



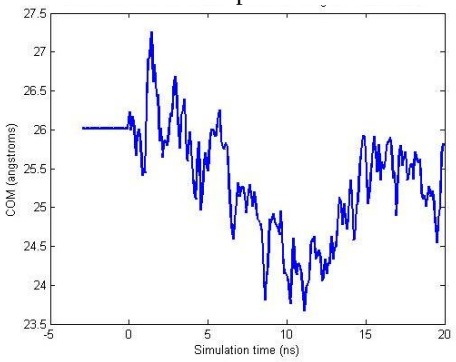
COM initial separation = 23 Å



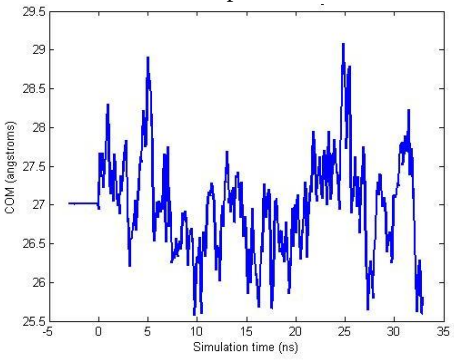
COM initial separation = 24 Å



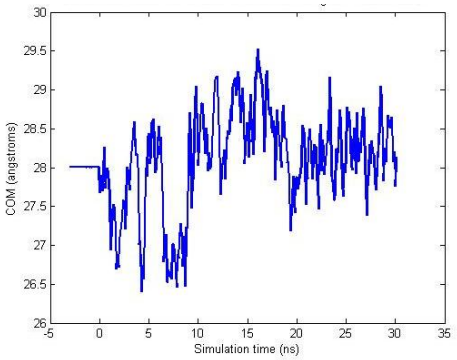
COM initial separation = 25 Å



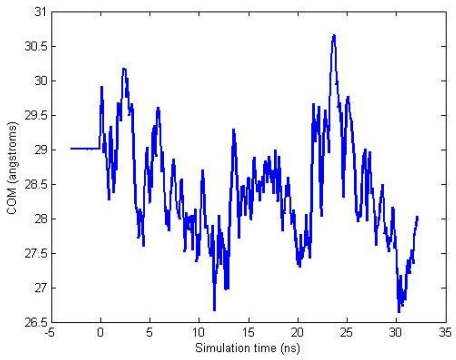
COM initial separation = 26 Å



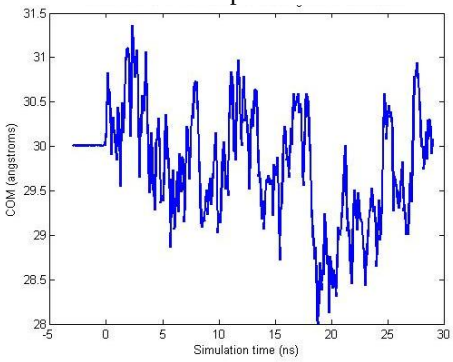
COM initial separation = 27 Å



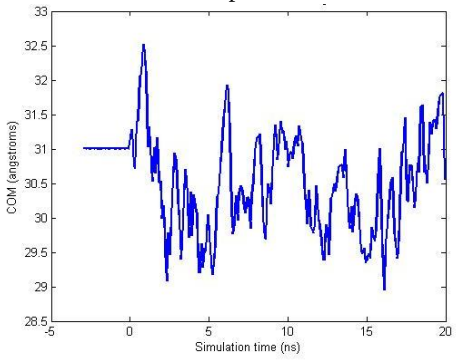
COM initial separation = 28 Å



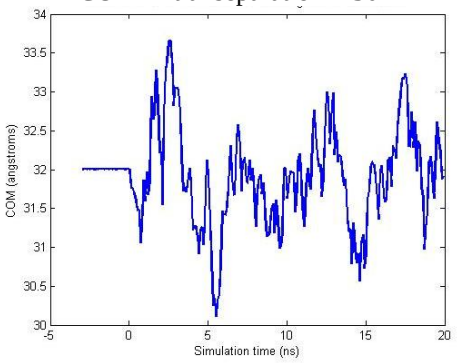
COM initial separation = 29 Å



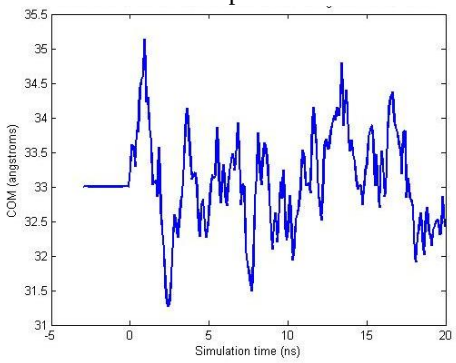
COM initial separation = 30 Å



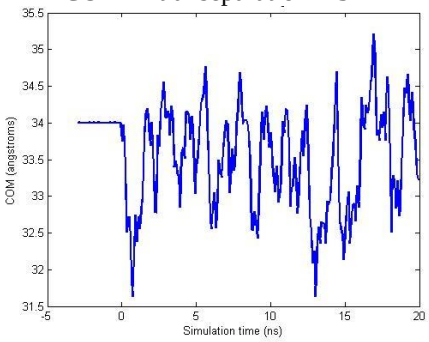
COM initial separation = 31 Å



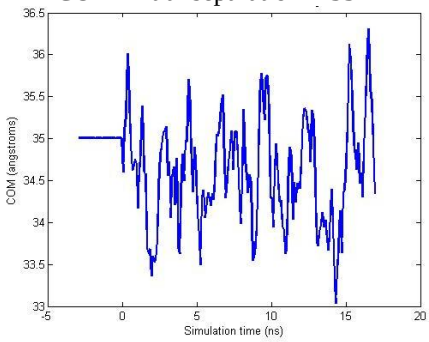
COM initial separation = 32 Å



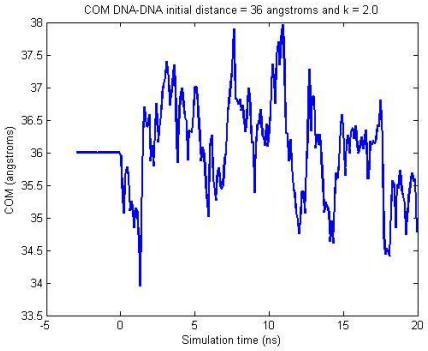
COM initial separation = 33 Å



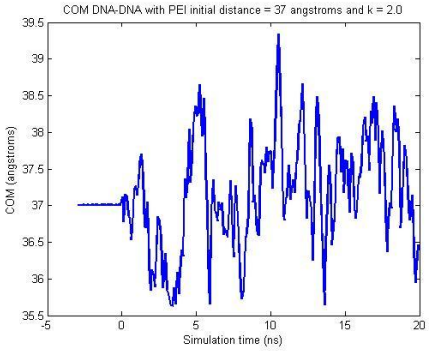
COM initial separation = 34 Å



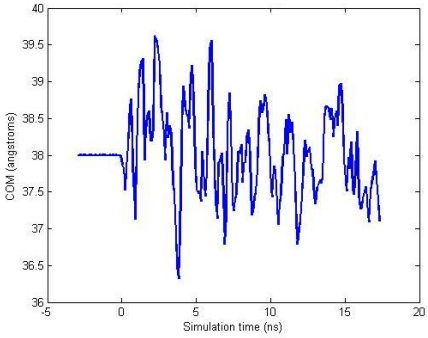
COM initial separation = 35 Å



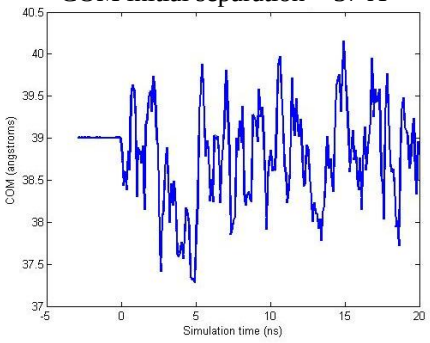
COM initial separation = 36 Å



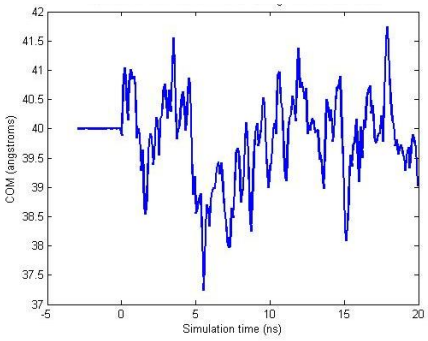
COM initial separation = 37 Å



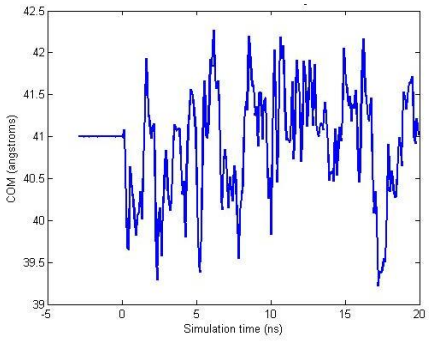
COM initial separation = 38 Å



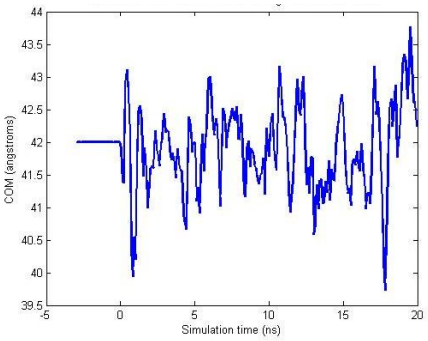
COM initial separation = 39 Å



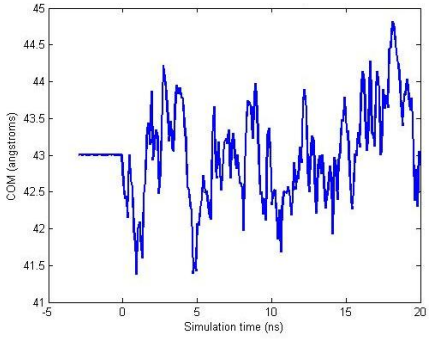
COM initial separation = 40 Å



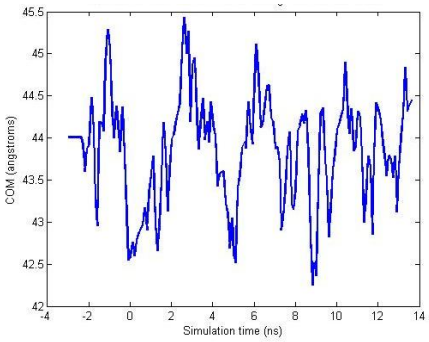
COM initial separation = 41 Å



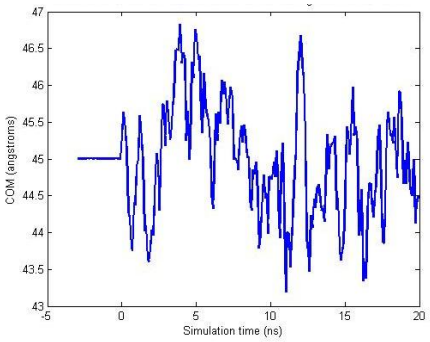
COM initial separation = 42 Å



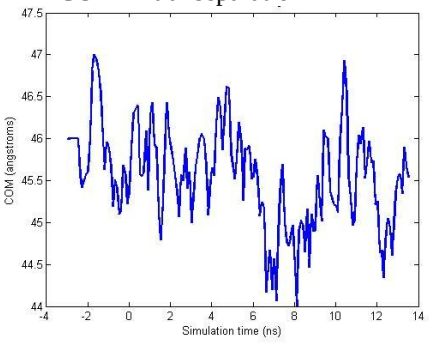
COM initial separation = 43 Å



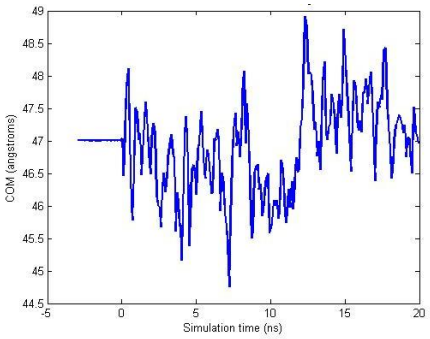
COM initial separation = 44 Å



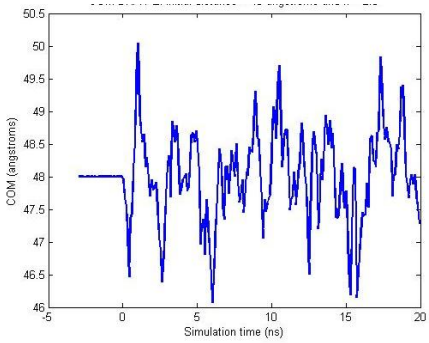
COM initial separation = 45 Å



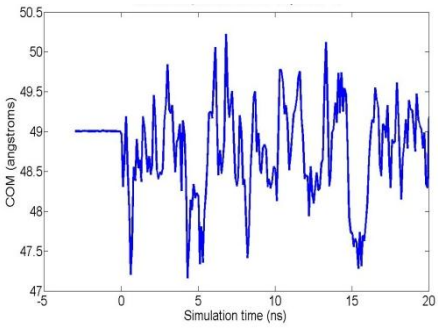
COM initial separation = 46 Å



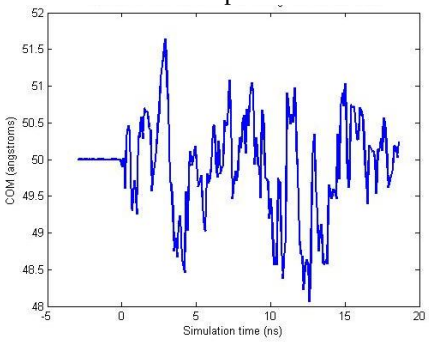
COM initial separation = 47 Å



COM initial separation =48 Å

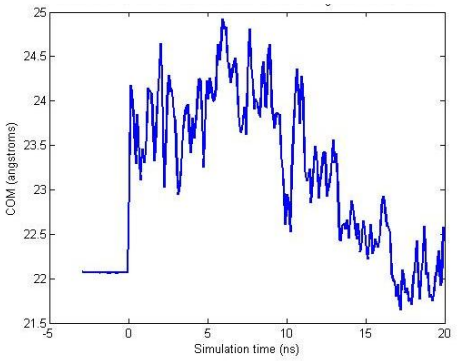


COM initial separation = 49 Å

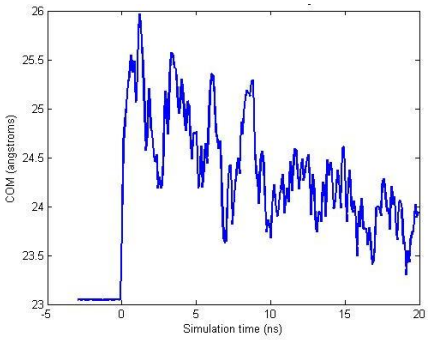


COM initial separation = 50 Å

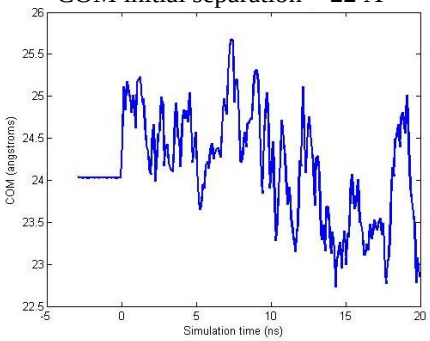
System 2D-6P



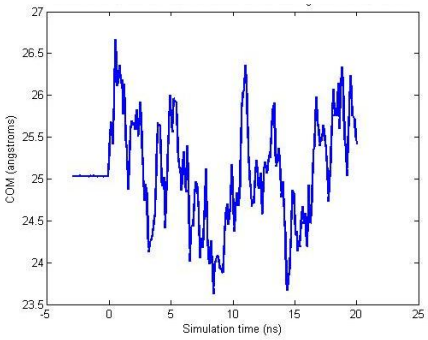
COM initial separation = 22 Å



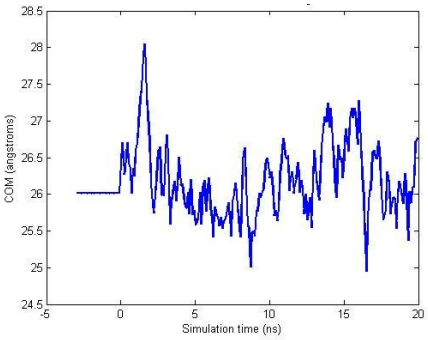
COM initial separation = 23 Å



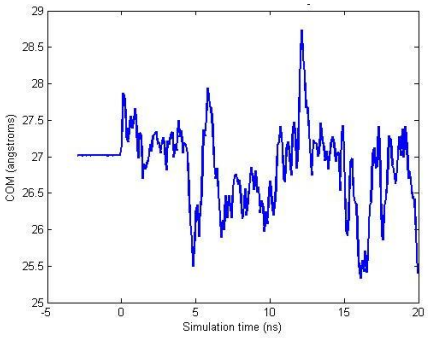
COM initial separation = 24 Å



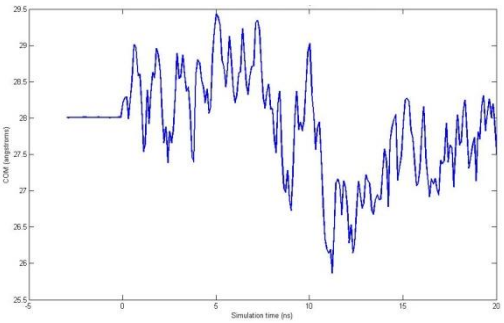
COM initial separation = 25 Å



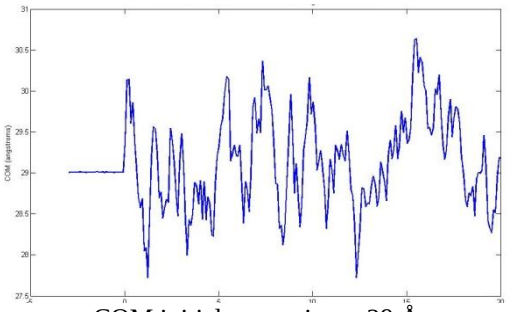
COM initial separation = 26 Å



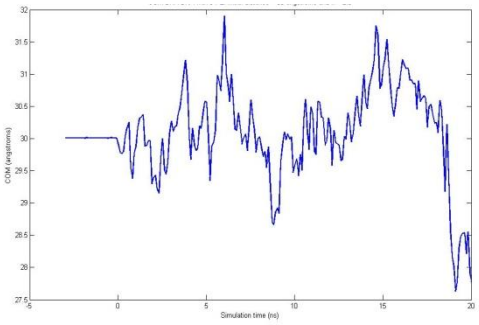
COM initial separation = 27 Å



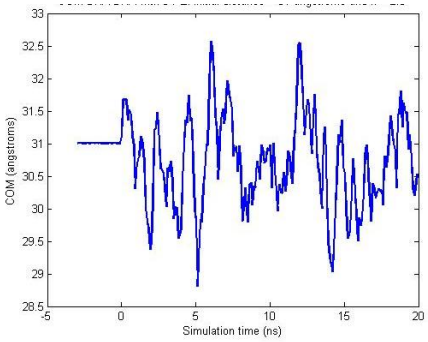
COM initial separation = 28 Å



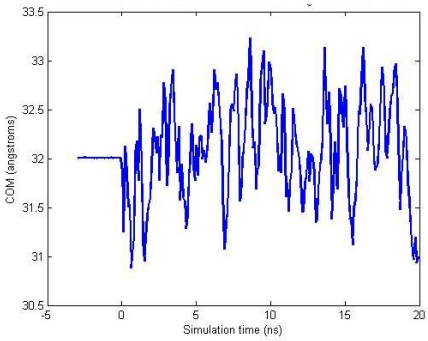
COM initial separation = 29 Å



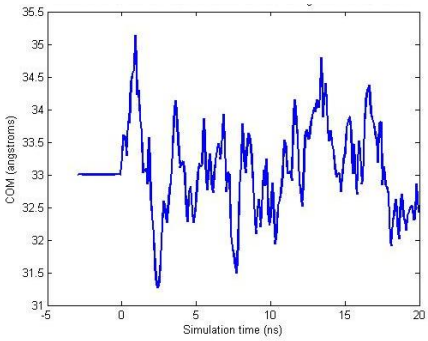
COM initial separation = 30 Å



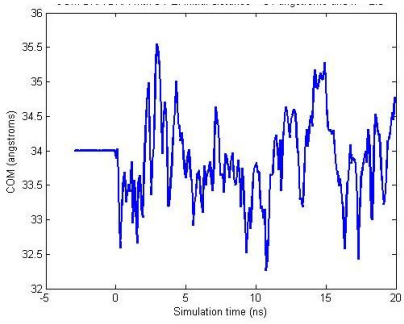
COM initial separation = 31 Å



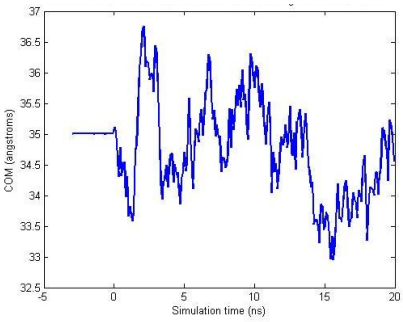
COM initial separation = 32 Å



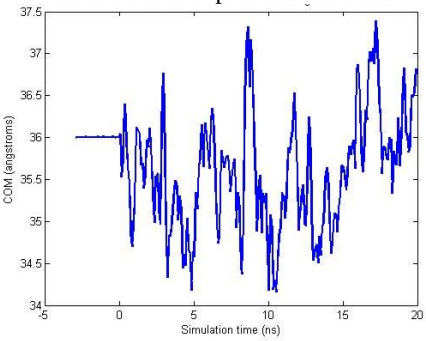
COM initial separation = 33 Å



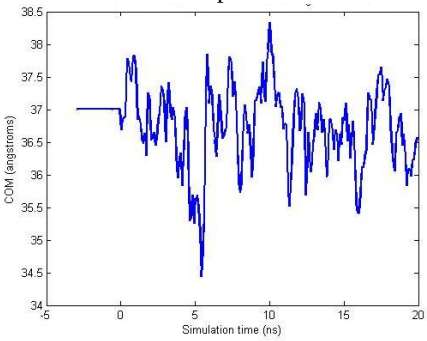
COM initial separation = 34 Å



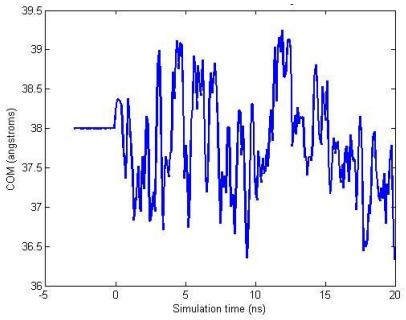
COM initial separation = 35 Å



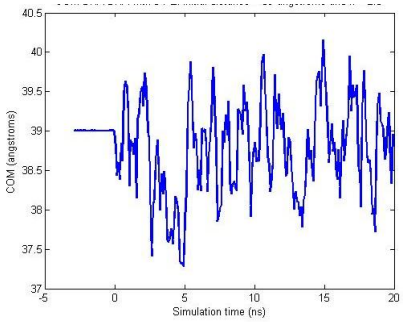
COM initial separation = 36 Å



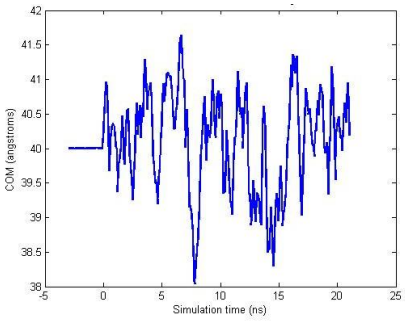
COM initial separation = 37 Å



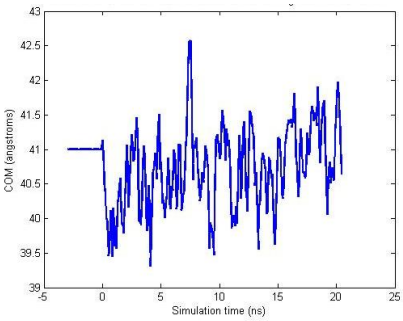
COM initial separation = 38 Å



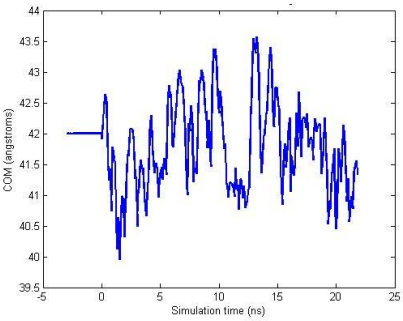
COM initial separation = 39 Å



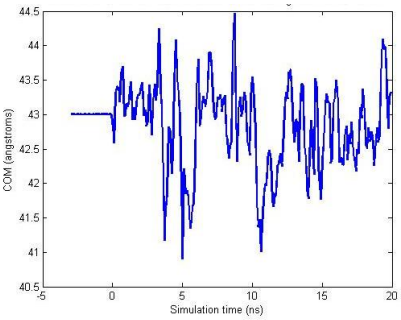
COM initial separation = 40 Å



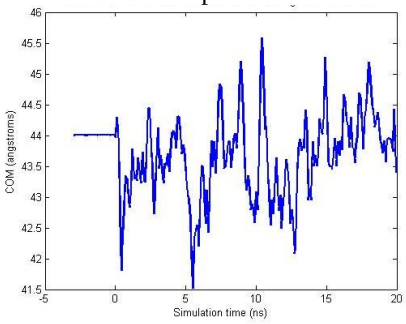
COM initial separation = 41 Å



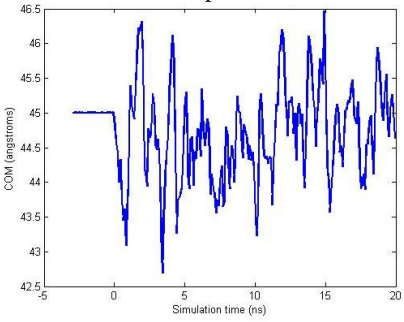
COM initial separation = 42 Å



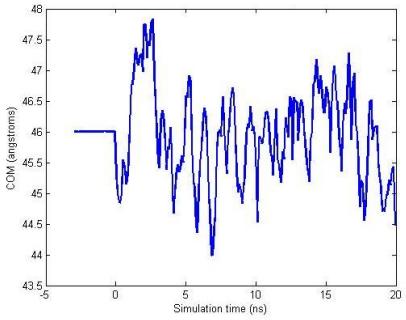
COM initial separation = 43 Å



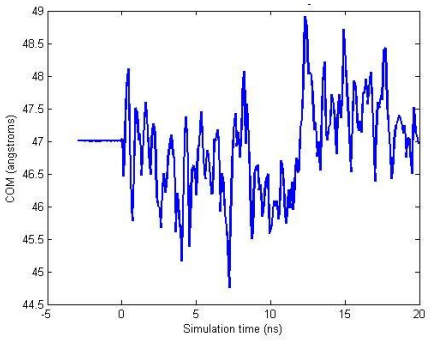
COM initial separation = 44 Å



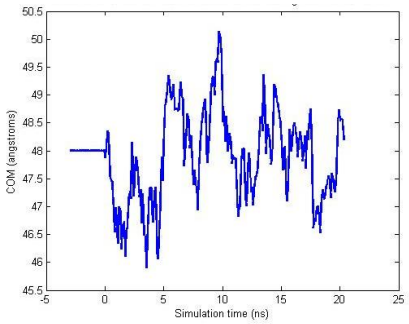
COM initial separation = 45 Å



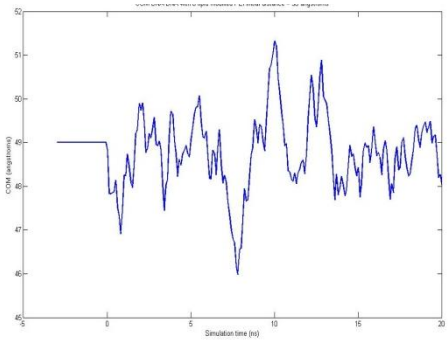
COM initial separation = 46 Å



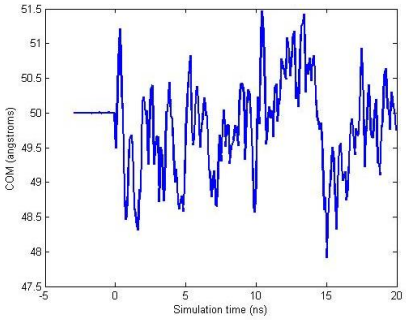
COM initial separation = 47 Å



COM initial separation =48 Å

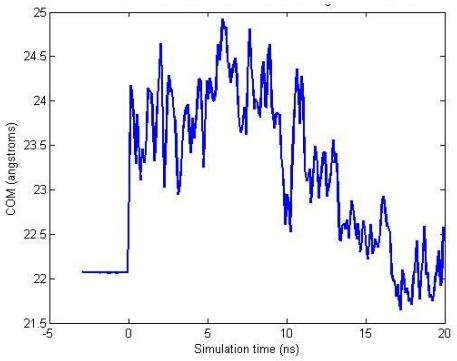


COM initial separation = 49 Å

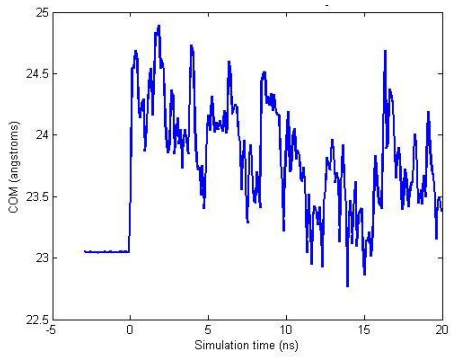


COM initial separation = 50 Å

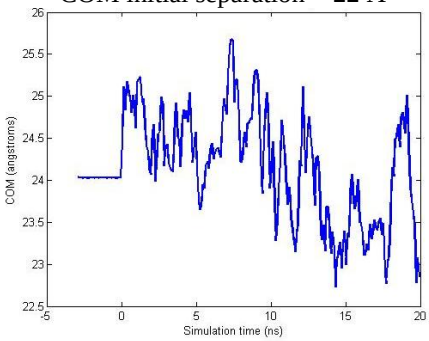
System 2D-4P



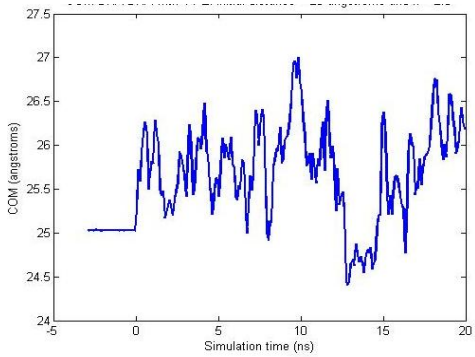
COM initial separation = 22 Å



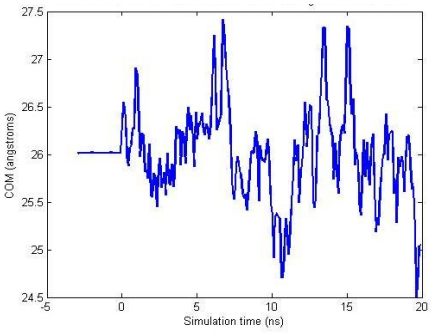
COM initial separation = 23 Å



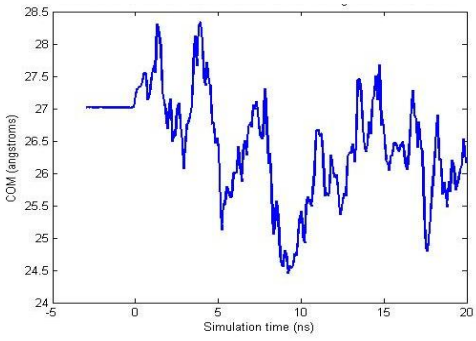
COM initial separation = 24 Å



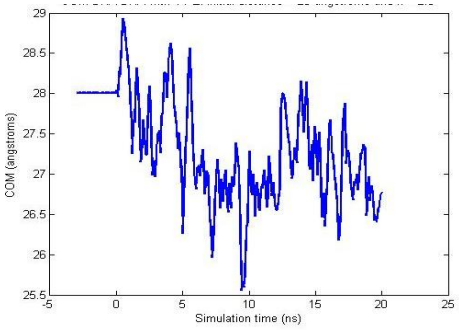
COM initial separation = 25 Å



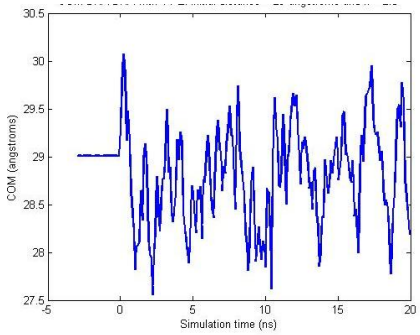
COM initial separation = 26 Å



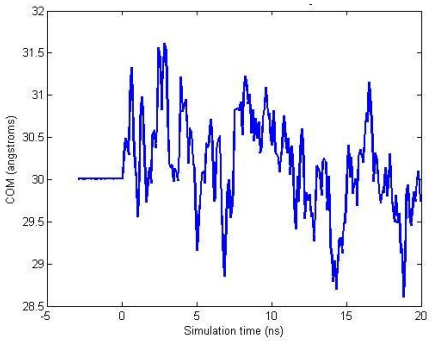
COM initial separation = 27 Å



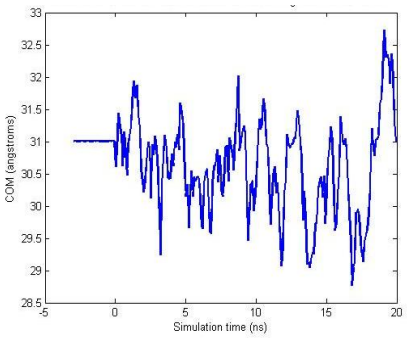
COM initial separation = 28 Å



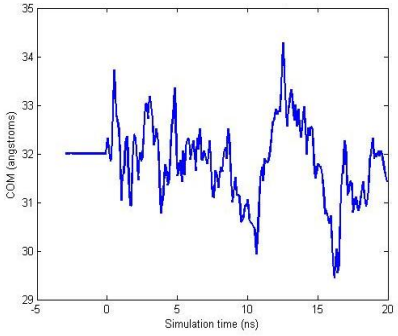
COM initial separation = 29 Å



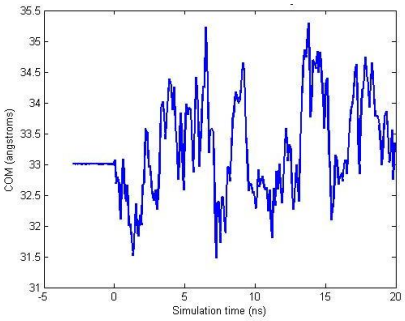
COM initial separation = 30 Å



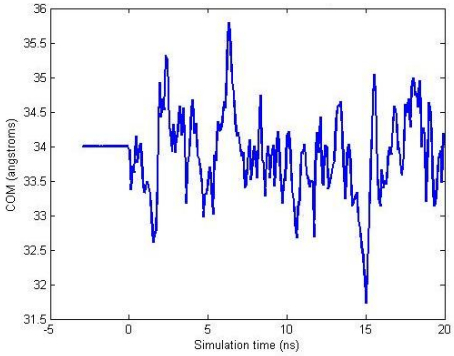
COM initial separation = 31 Å



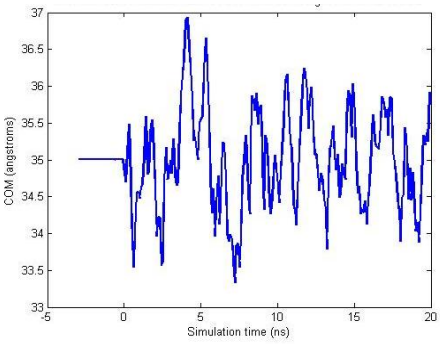
COM initial separation = 32 Å



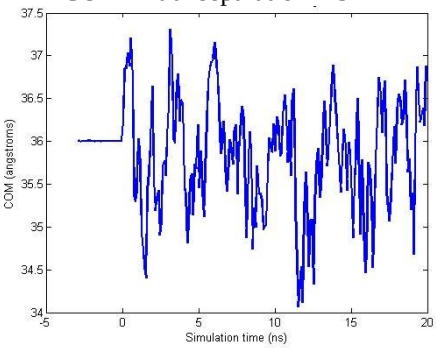
COM initial separation = 33 Å



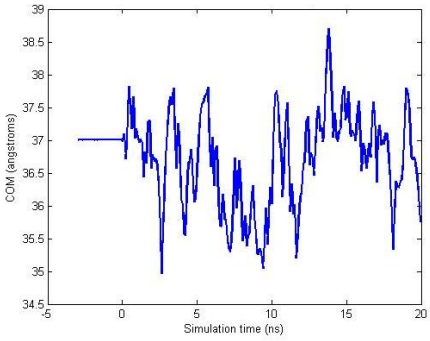
COM initial separation = 34 Å



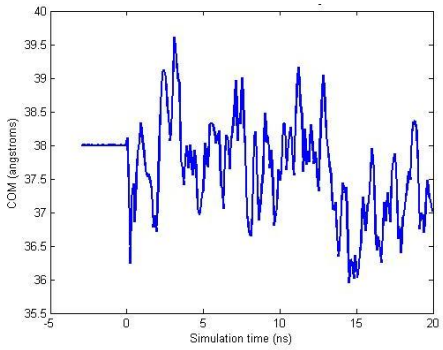
COM initial separation = 35 Å



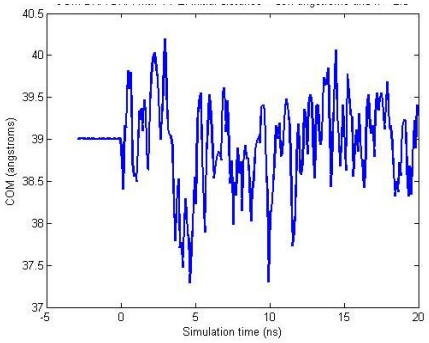
COM initial separation = 36 Å



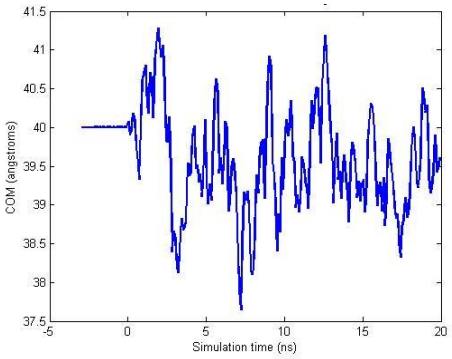
COM initial separation = 37 Å



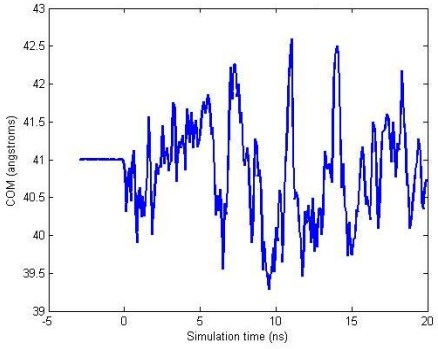
COM initial separation = 38 Å



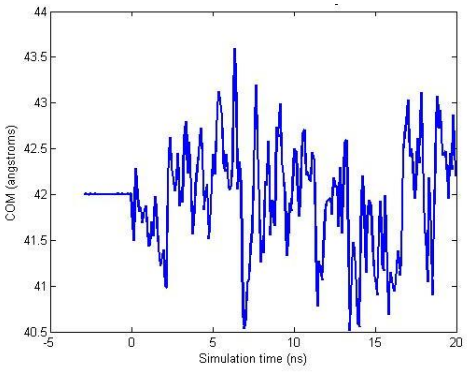
COM initial separation = 39 Å



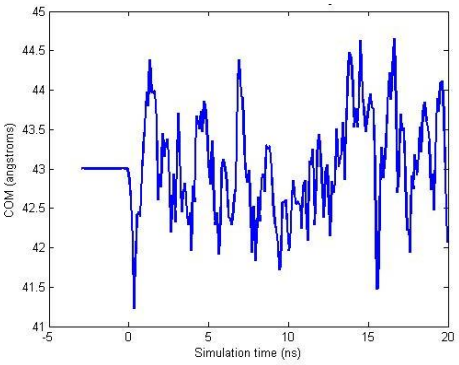
COM initial separation = 40 Å



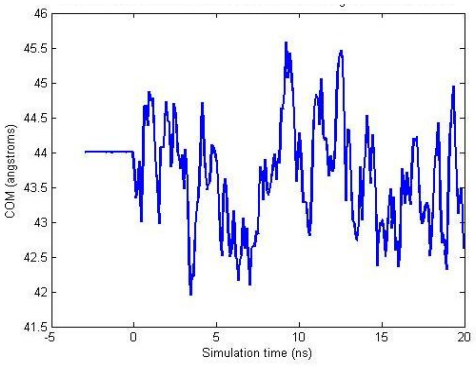
COM initial separation = 41 Å



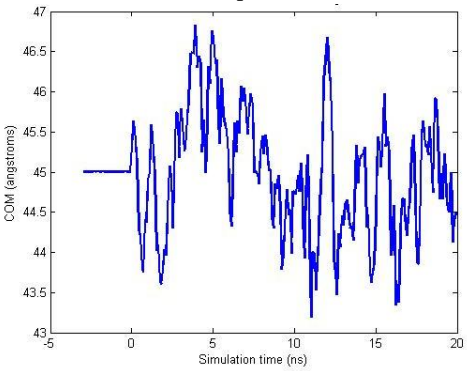
COM initial separation = 42 Å



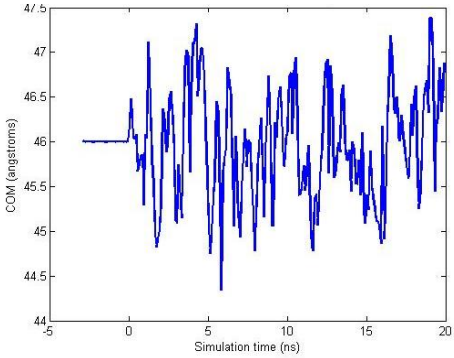
COM initial separation = 43 Å



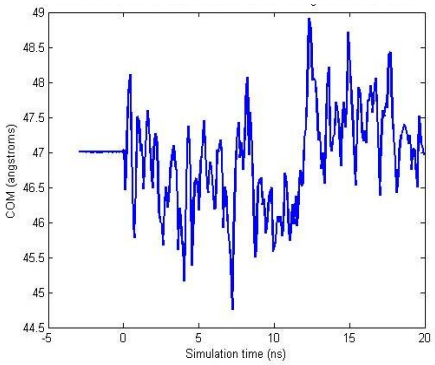
COM initial separation = 44 Å



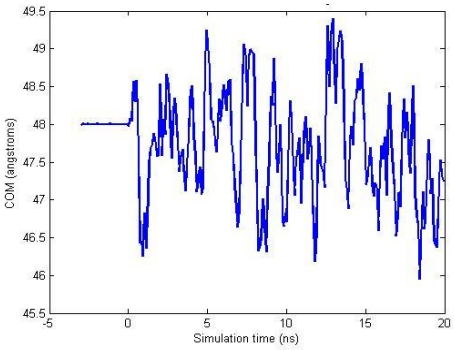
COM initial separation = 45 Å



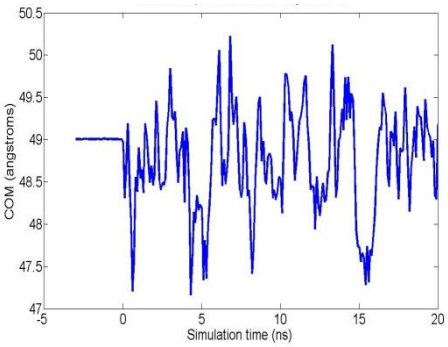
COM initial separation = 46 Å



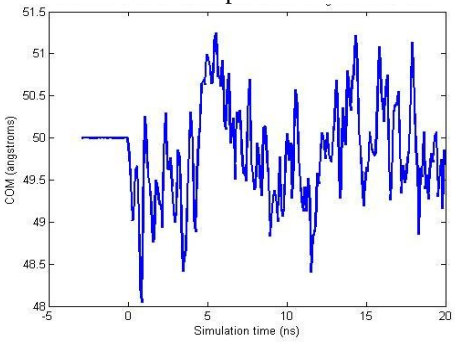
COM initial separation = 47 Å



COM initial separation =48 Å

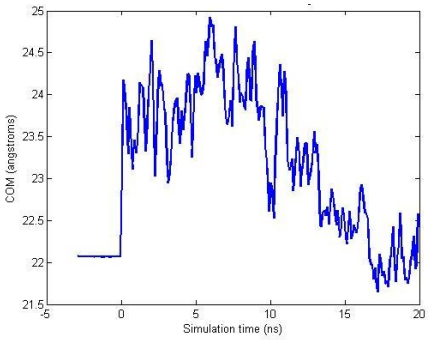


COM initial separation = 49 Å

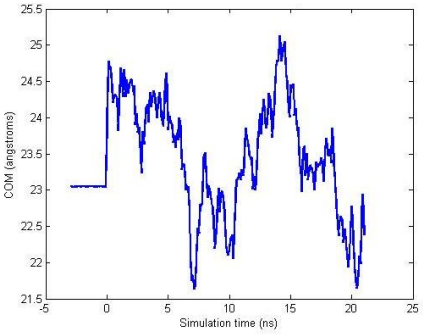


COM initial separation = 50 Å

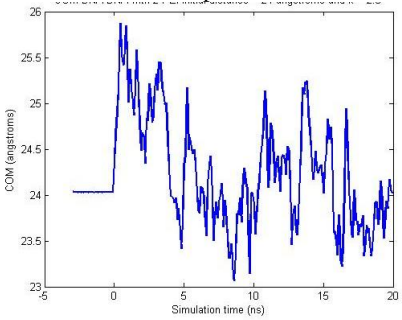
System 2D-2P



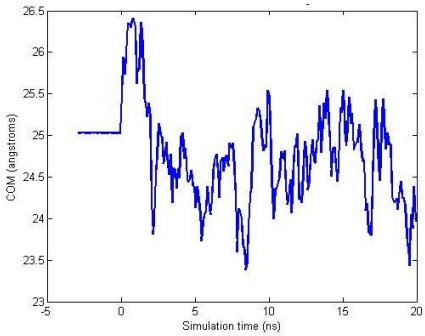
COM initial separation = 22 Å



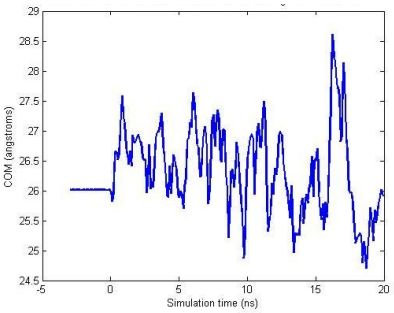
COM initial separation = 23 Å



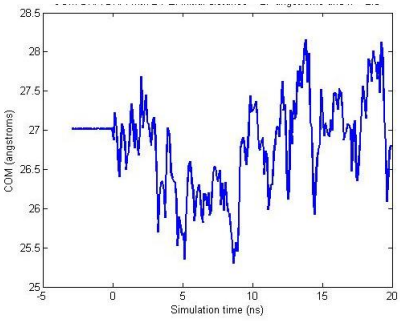
COM initial separation = 24 Å



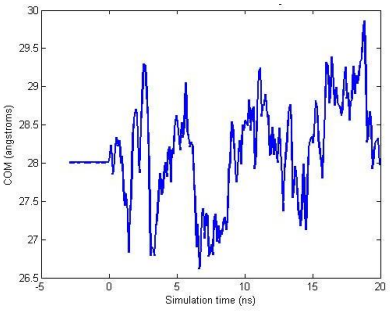
COM initial separation = 25 Å



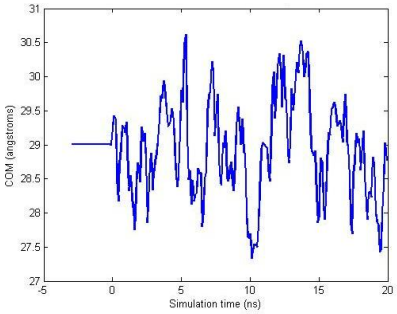
COM initial separation = 26 Å



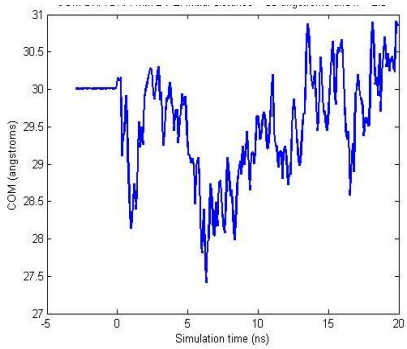
COM initial separation = 27 Å



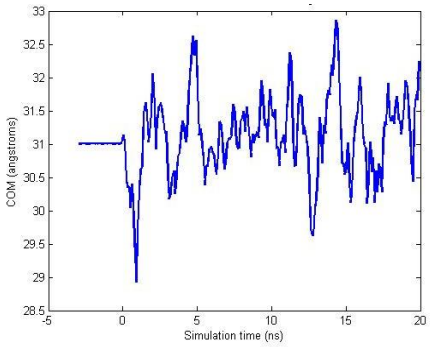
COM initial separation = 28 Å



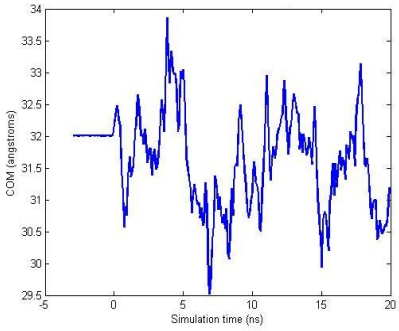
COM initial separation = 29 Å



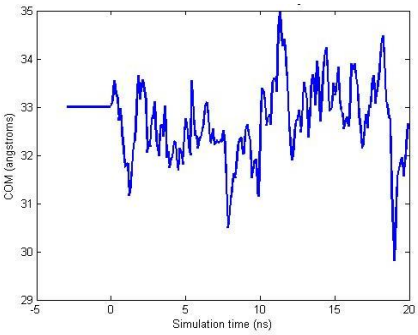
COM initial separation = 30 Å



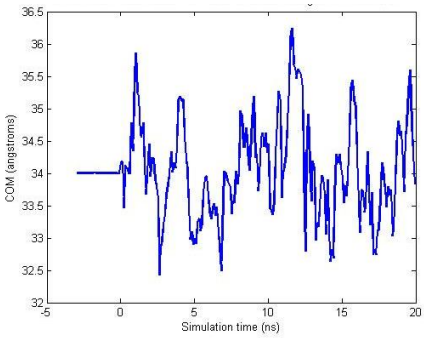
COM initial separation = 31 Å



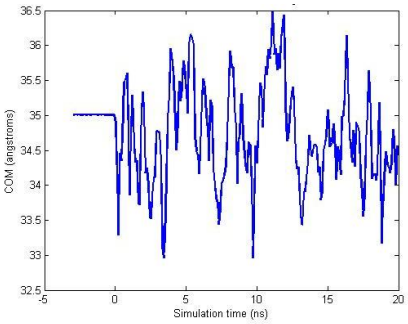
COM initial separation = 32 Å



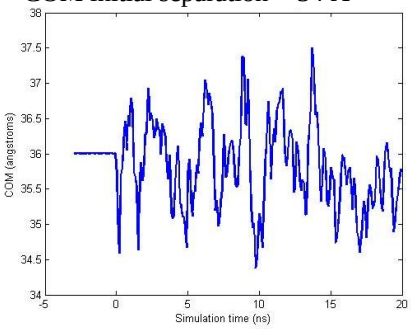
COM initial separation = 33 Å



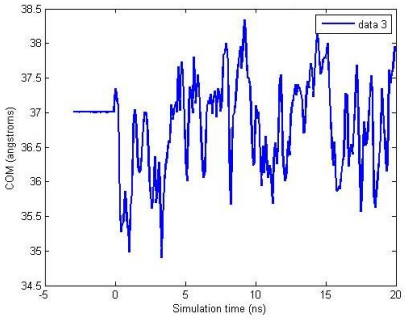
COM initial separation = 34 Å



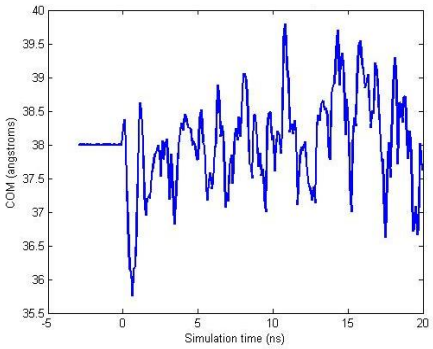
COM initial separation = 35 Å



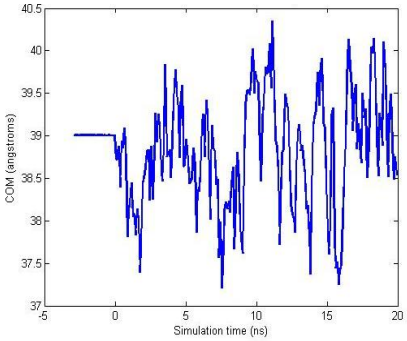
COM initial separation = 36 Å



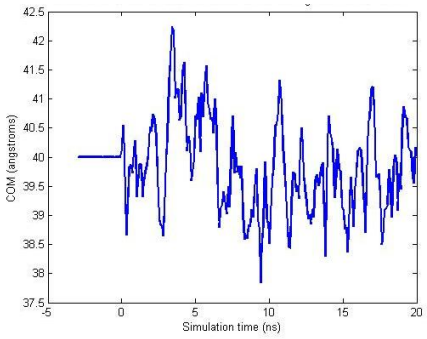
COM initial separation = 37 Å



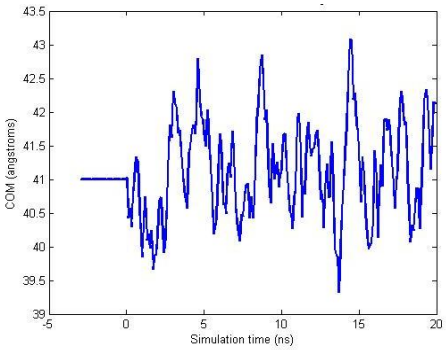
COM initial separation = 38 Å



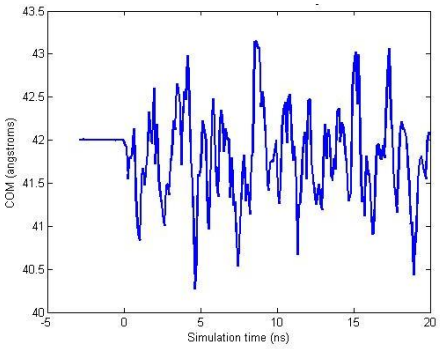
COM initial separation = 39 Å



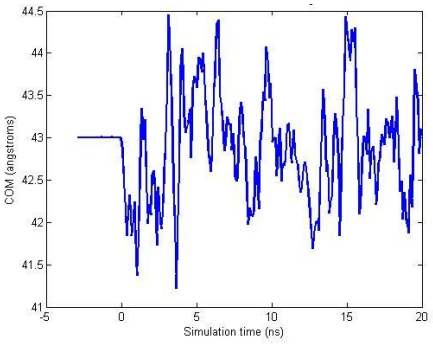
COM initial separation = 40 Å



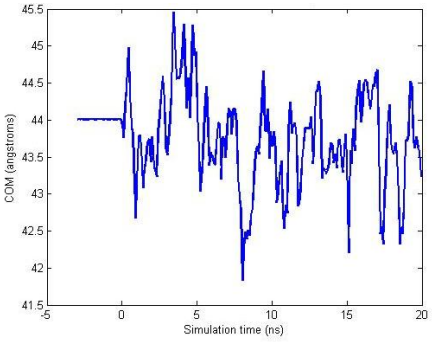
COM initial separation = 41 Å



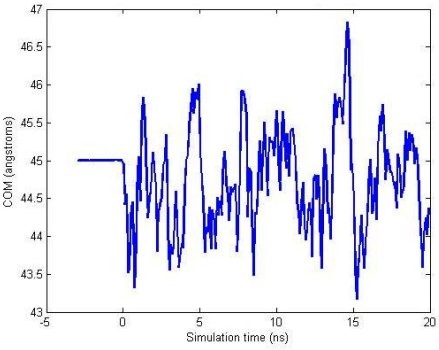
COM initial separation = 42 Å



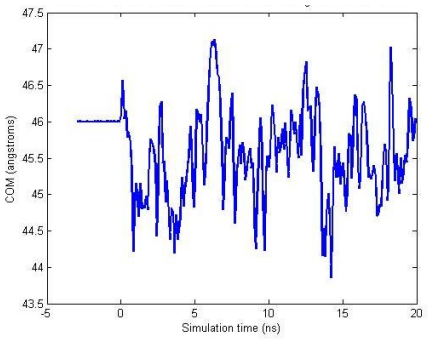
COM initial separation = 43 Å



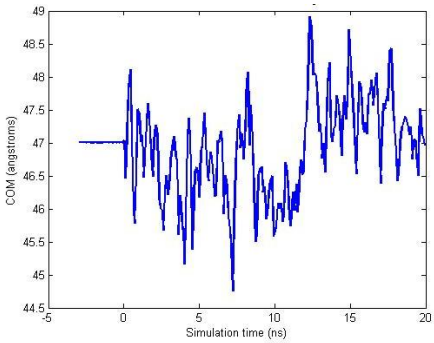
COM initial separation = 44 Å



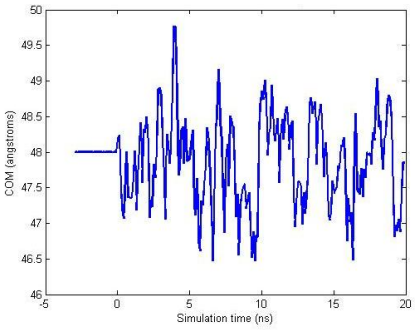
COM initial separation = 45 Å



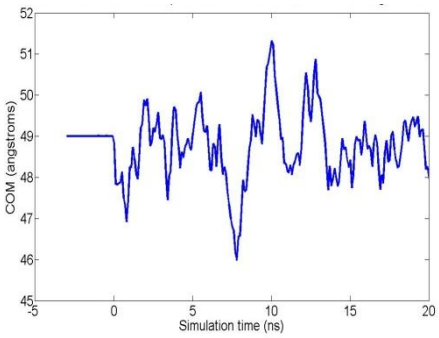
COM initial separation = 46 Å



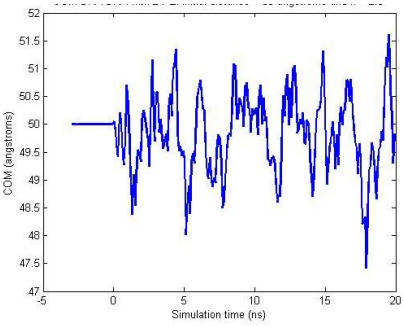
COM initial separation = 47 Å



COM initial separation =48 Å

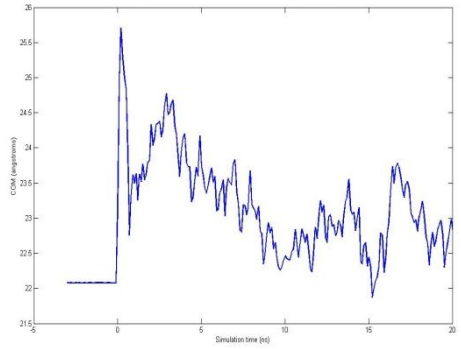


COM initial separation = 49 Å

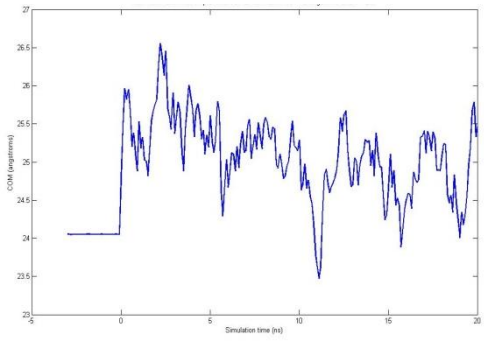


COM initial separation = 50 Å

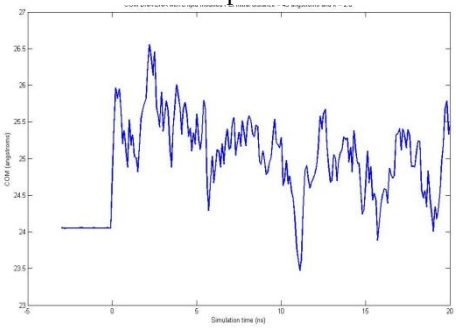
System 2D-8P(23%)



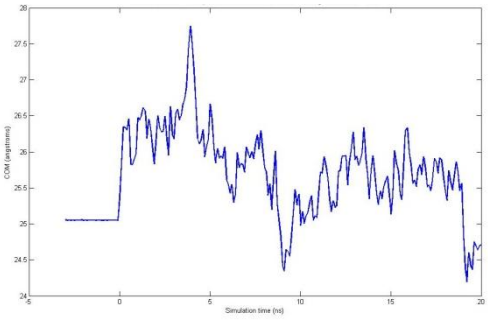
COM initial separation = 22 Å



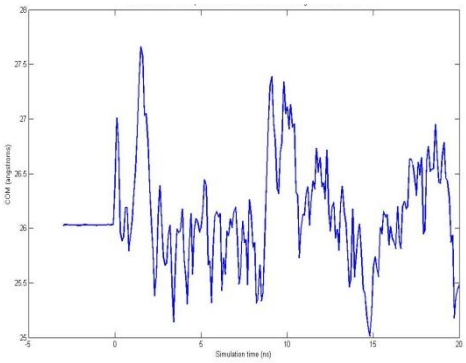
COM initial separation = 23 Å



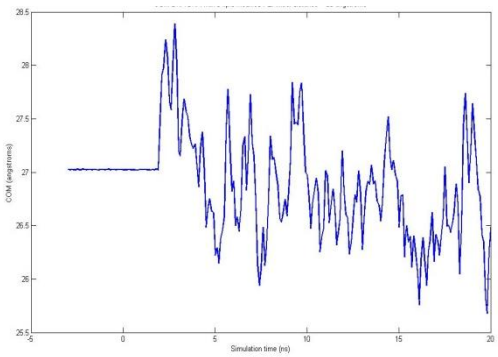
COM initial separation = 24 Å



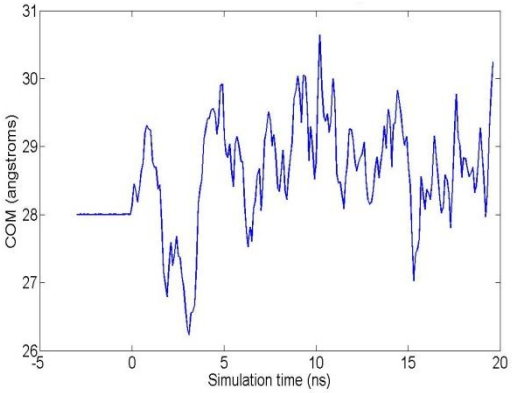
COM initial separation = 25 Å



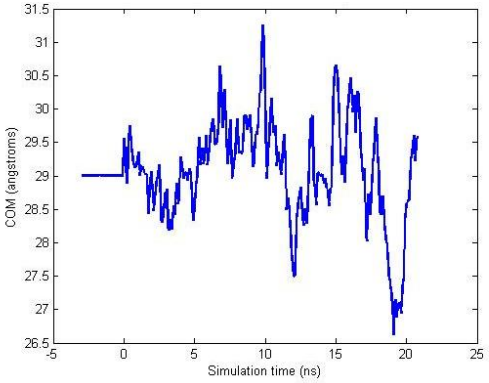
COM initial separation = 26 Å



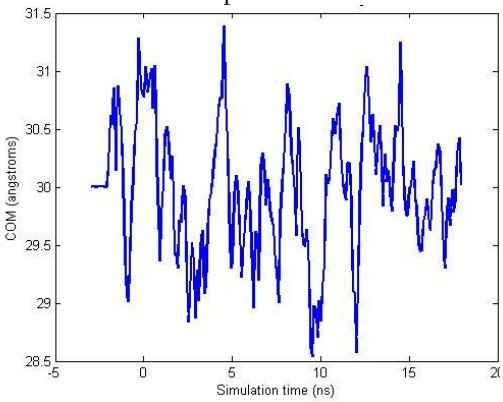
COM initial separation = 27 Å



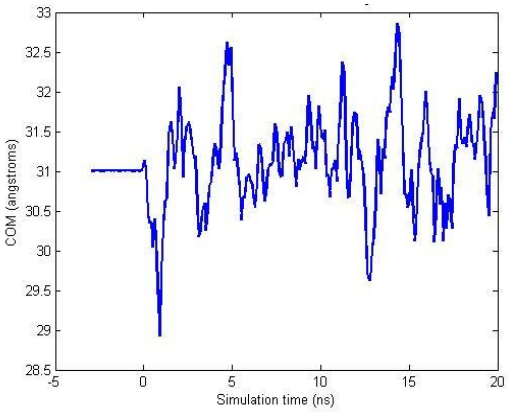
COM initial separation = 28 Å



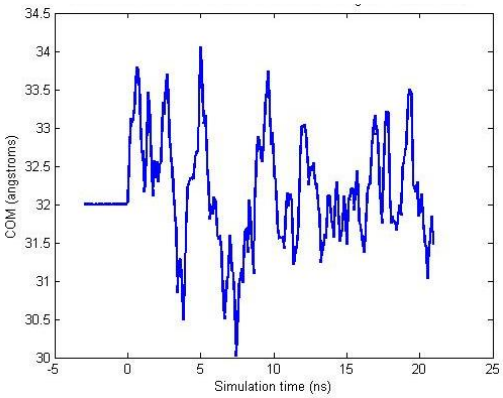
COM initial separation = 29 Å



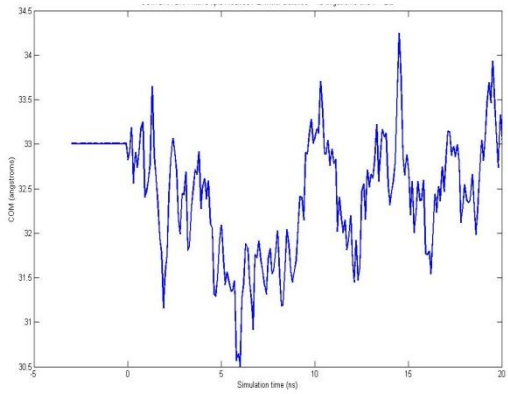
COM initial separation = 30 Å



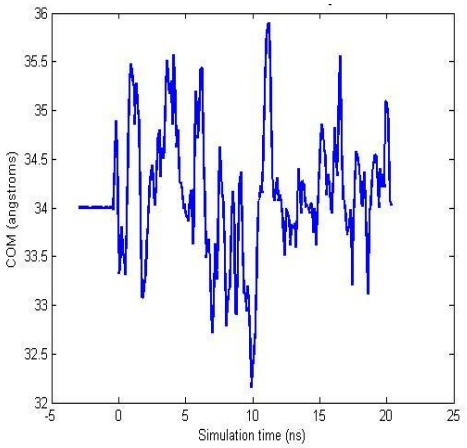
COM initial separation = 31 Å



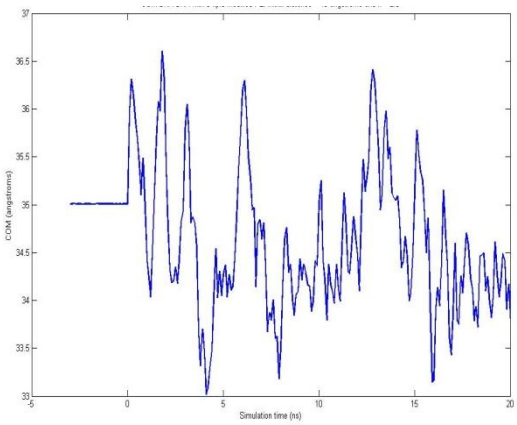
COM initial separation = 32 Å



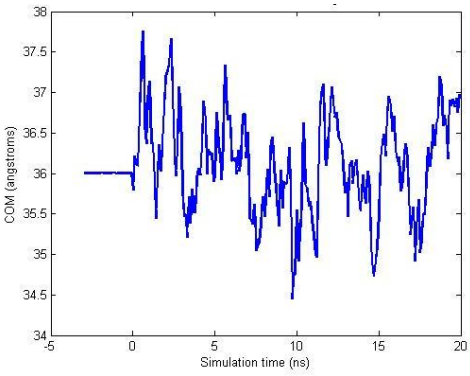
COM initial separation = 33 Å



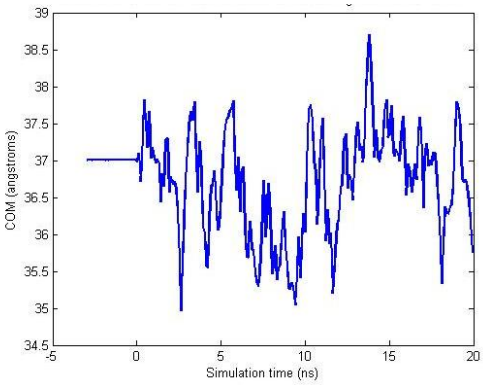
COM initial separation = 34 Å



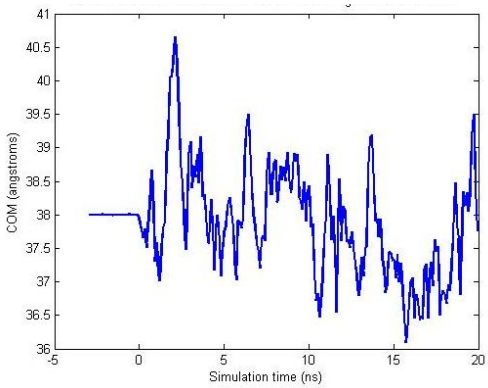
COM initial separation = 35 Å



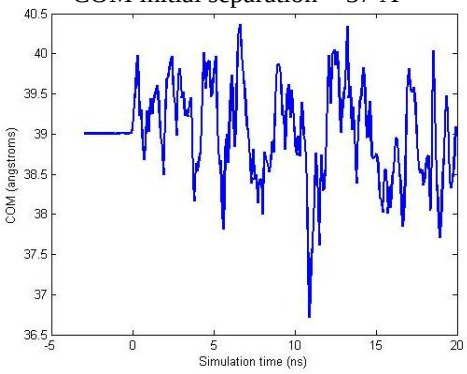
COM initial separation = 36 Å



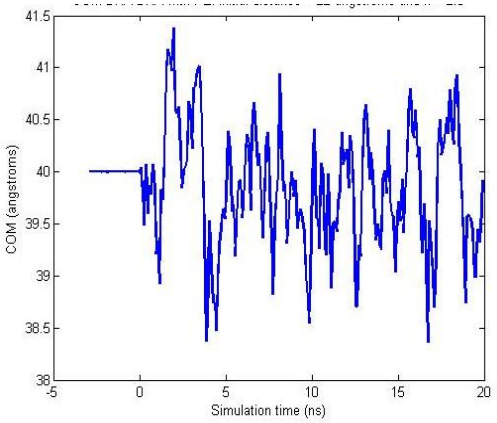
COM initial separation = 37 Å



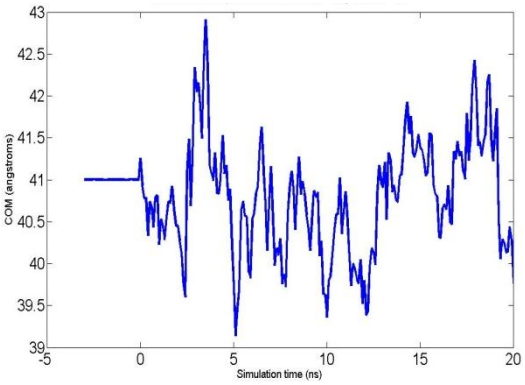
COM initial separation = 38 Å



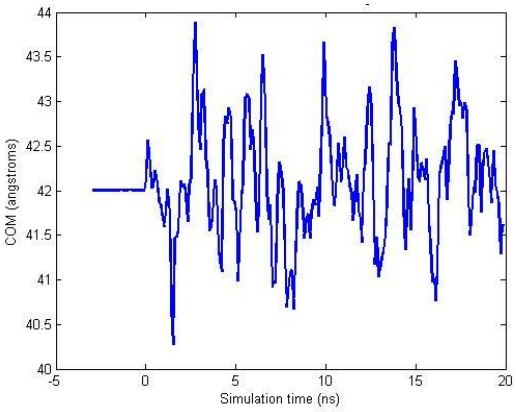
COM initial separation = 39 Å



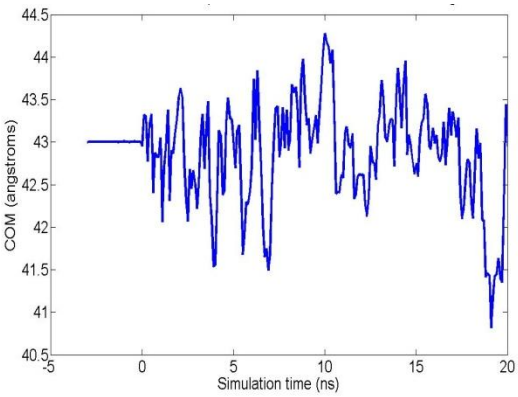
COM initial separation = 40 Å



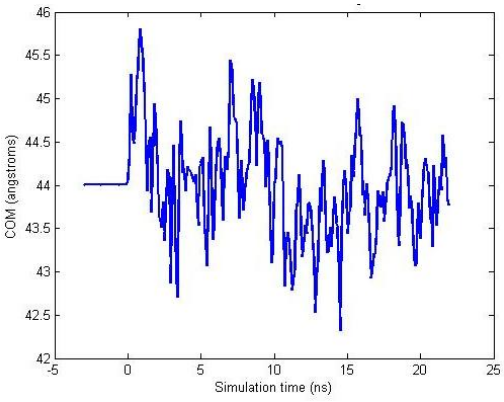
COM initial separation = 41 Å



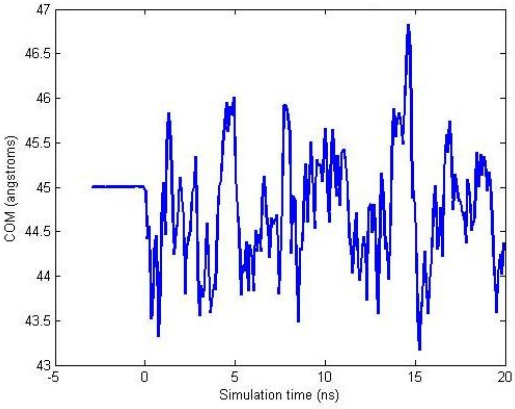
COM initial separation = 42 Å



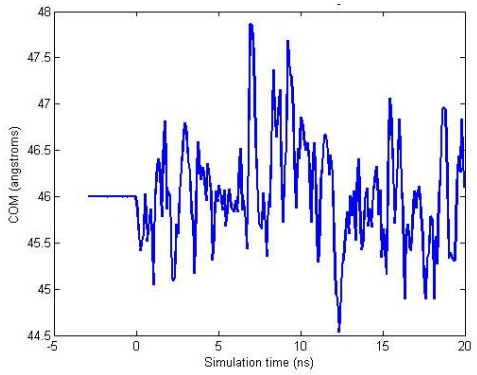
COM initial separation = 43 Å



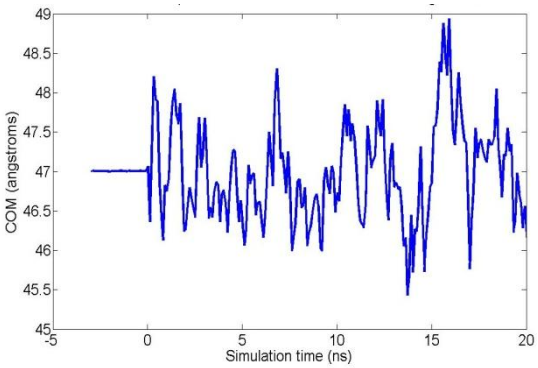
COM initial separation = 44 Å



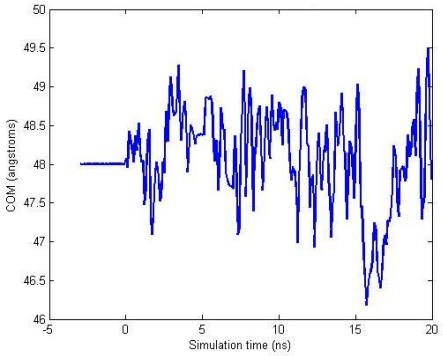
COM initial separation = 45 Å



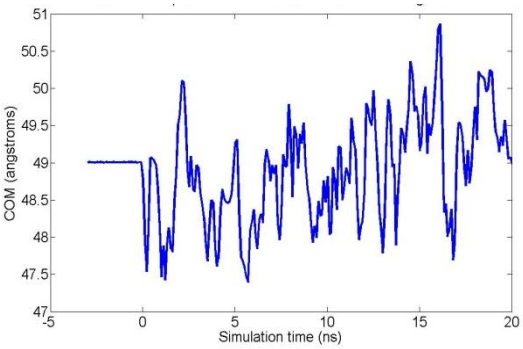
COM initial separation = 46 Å



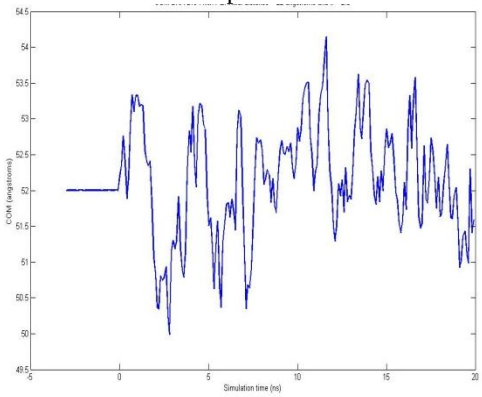
COM initial separation = 47 Å



COM initial separation =48 Å

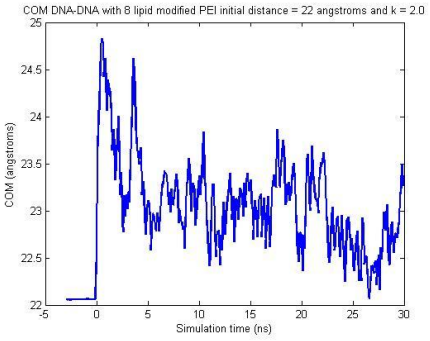


COM initial separation = 49 Å

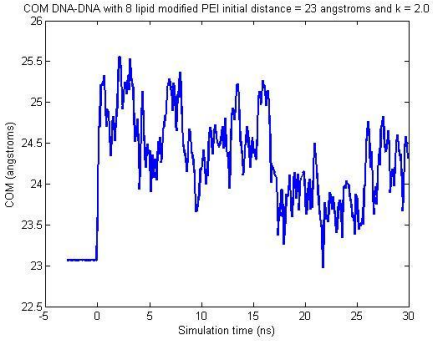


COM initial separation = 50 Å

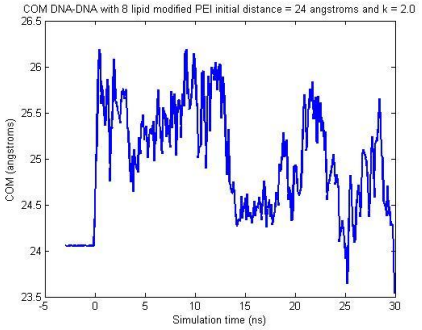
System 2D-8P(OA)



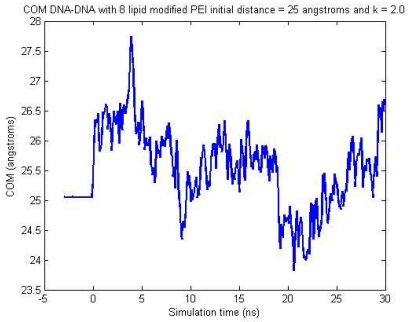
COM initial separation = 22 Å



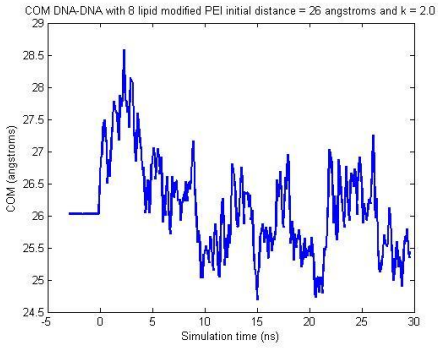
COM initial separation = 23 Å



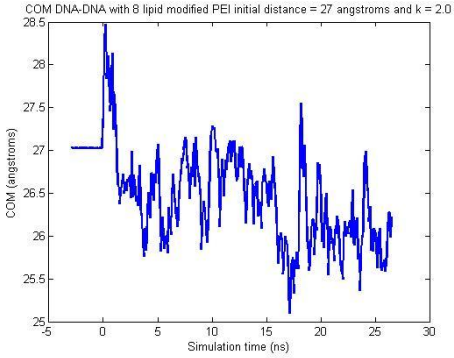
COM initial separation = 24 Å



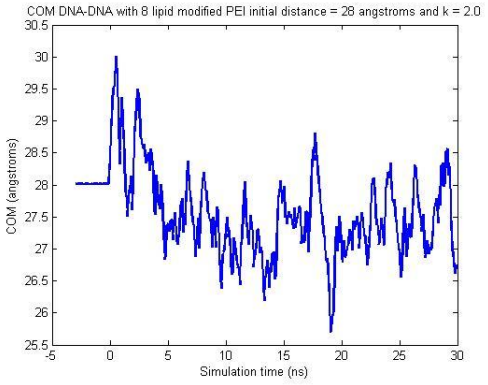
COM initial separation = 25 Å



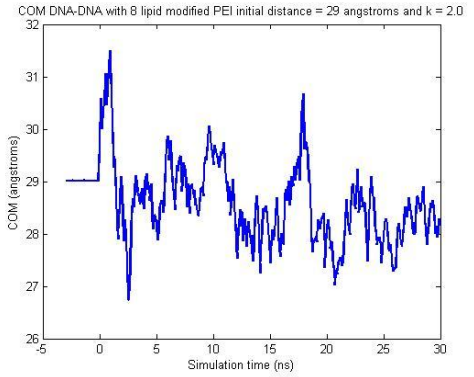
COM initial separation = 26 Å



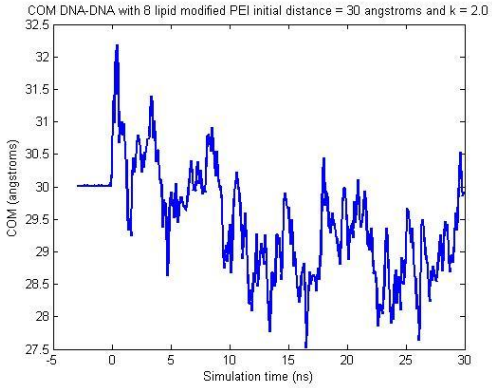
COM initial separation = 27 Å



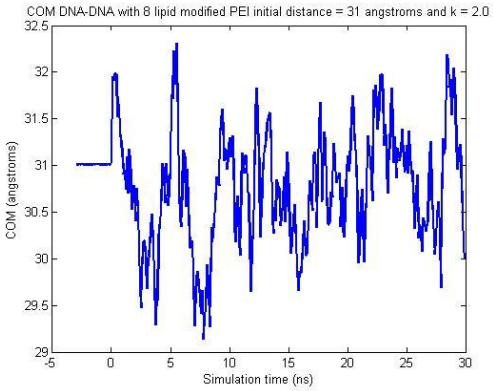
COM initial separation = 28 Å



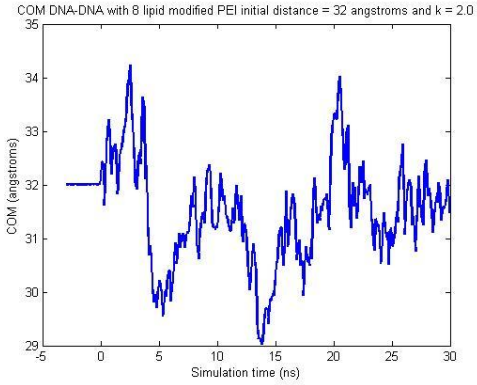
COM initial separation = 29 Å



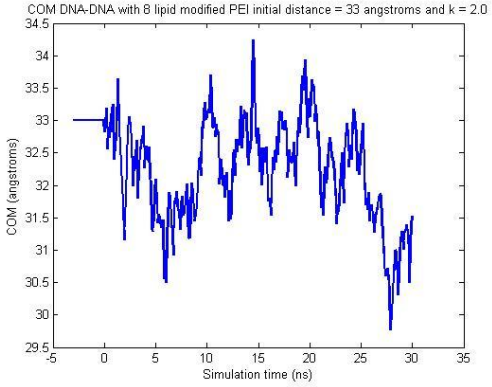
COM initial separation = 30 Å



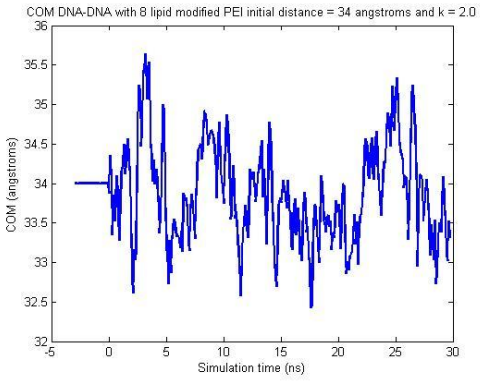
COM initial separation = 31 Å



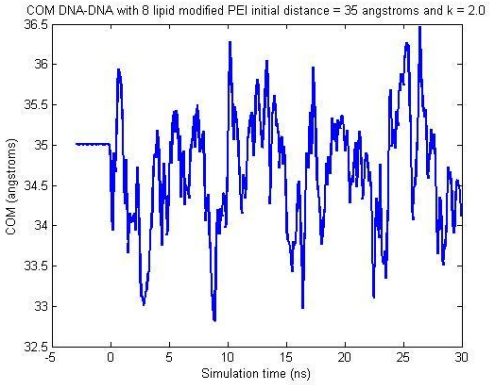
COM initial separation = 32 Å



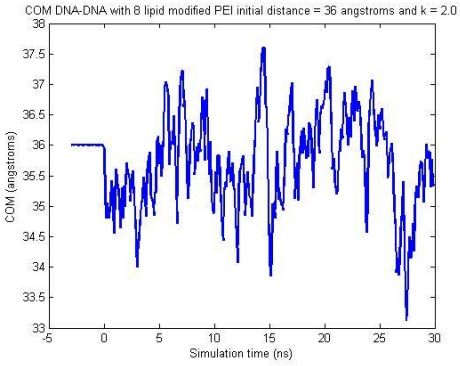
COM initial separation = 33 Å



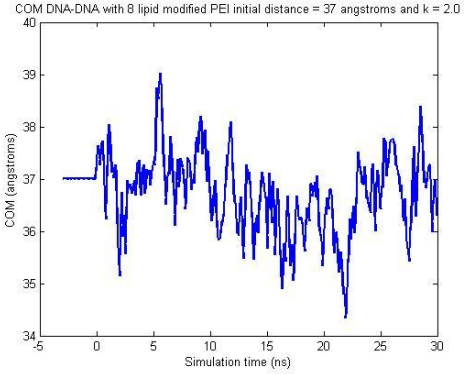
COM initial separation = 34 Å



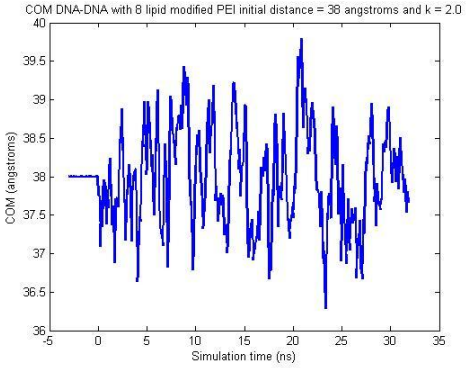
COM initial separation = 35 Å



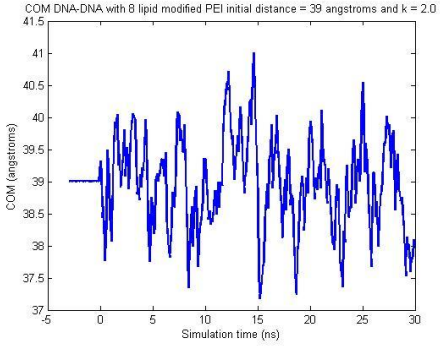
COM initial separation = 36 Å



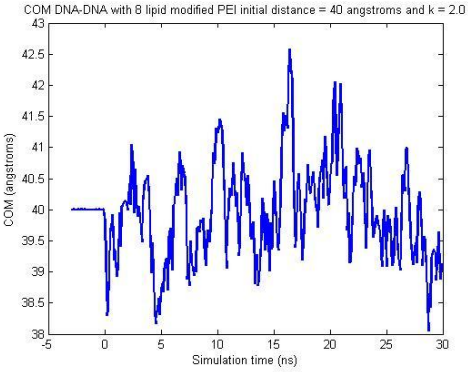
COM initial separation = 37 Å



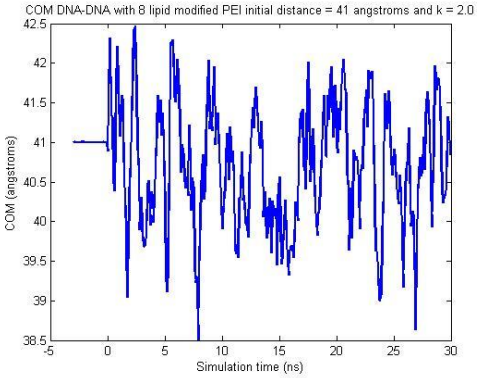
COM initial separation = 38 Å



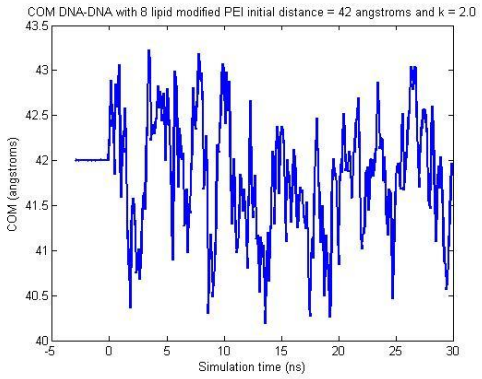
COM initial separation = 39 Å



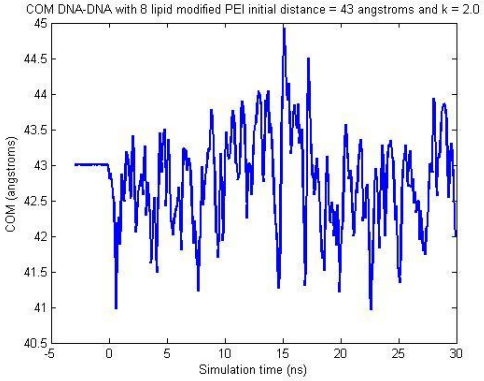
COM initial separation = 40 Å



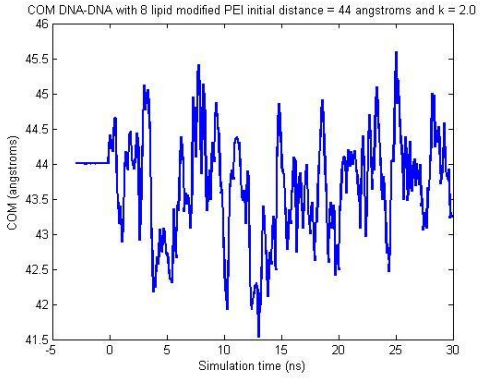
COM initial separation = 41 Å



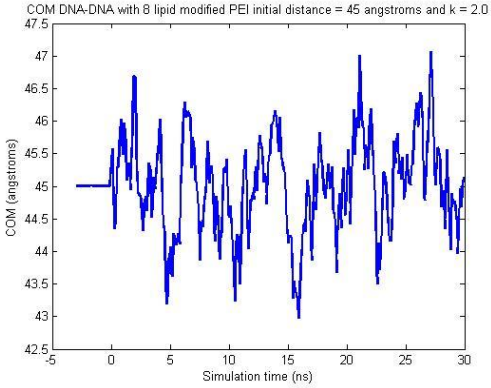
COM initial separation = 42 Å



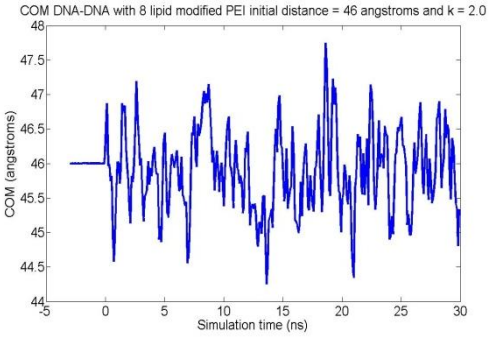
COM initial separation = 43 Å



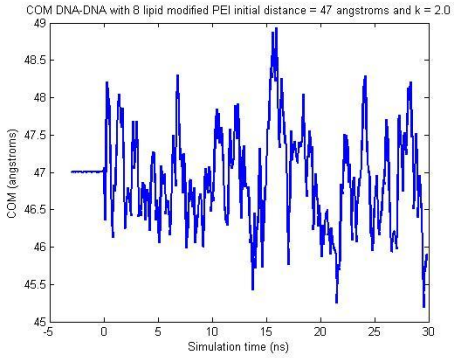
COM initial separation = 44 Å



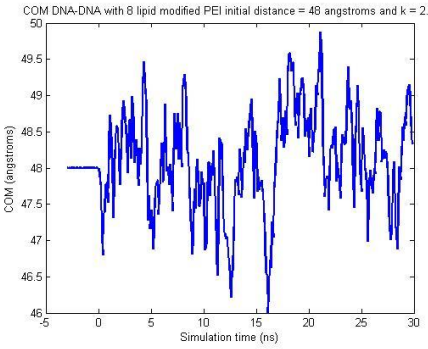
COM initial separation = 45 Å



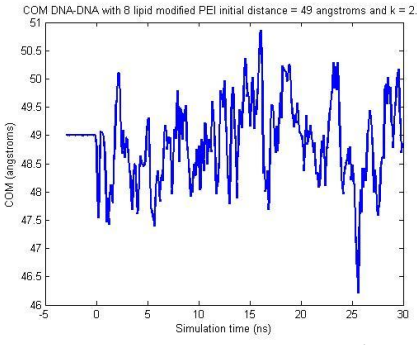
COM initial separation = 46 Å



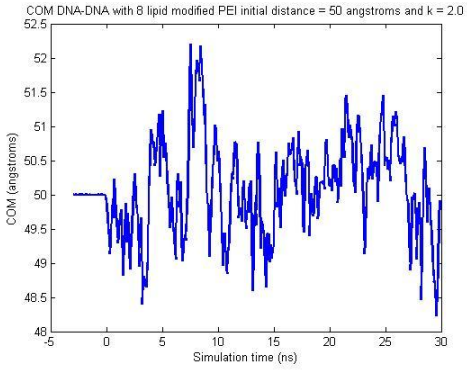
COM initial separation = 47 Å



COM initial separation =48 Å

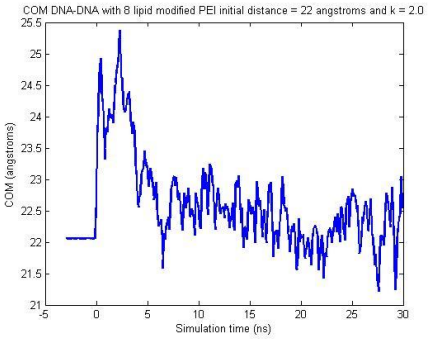


COM initial separation = 49 Å

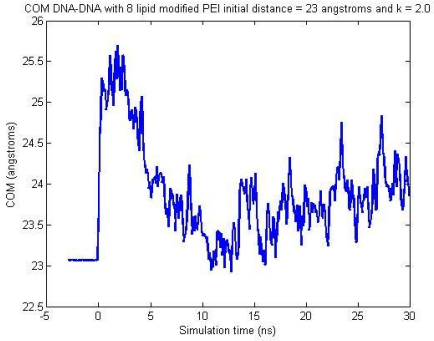


COM initial separation = 50 Å

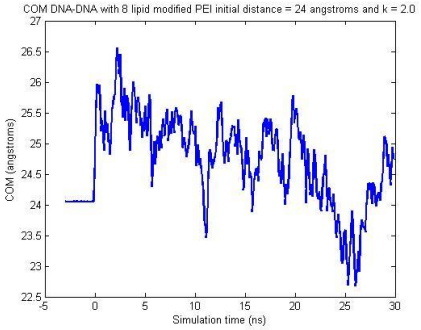
System 2D-8P(LA)



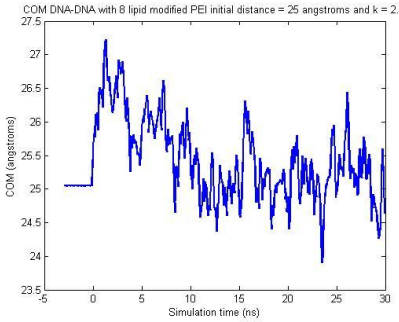
COM initial separation = 22 Å



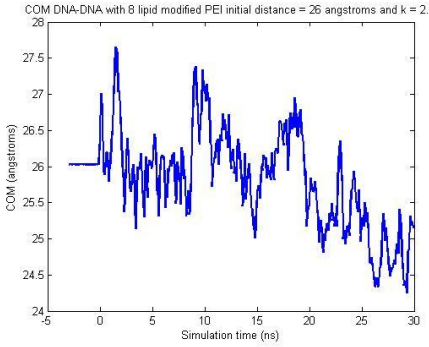
COM initial separation = 23 Å



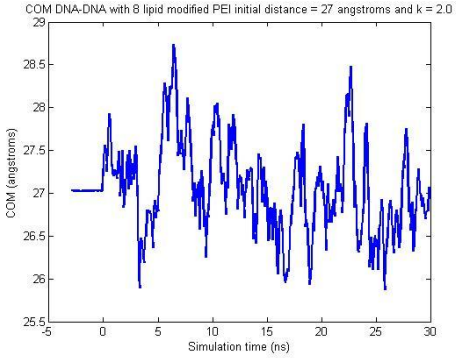
COM initial separation = 24 Å



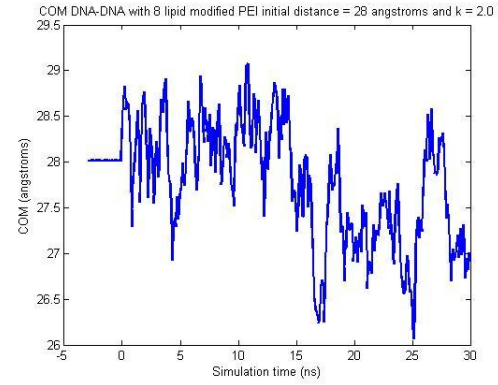
COM initial separation = 25 Å



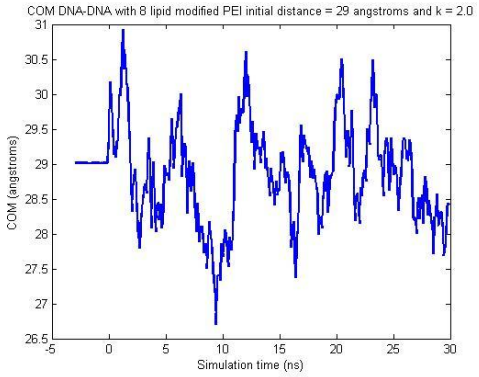
COM initial separation = 26 Å



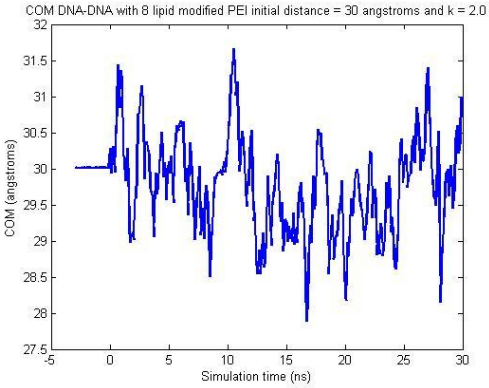
COM initial separation = 27 Å



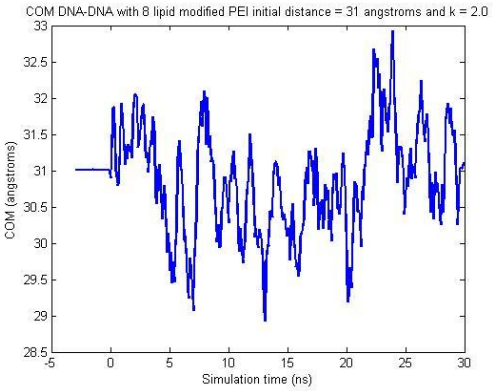
COM initial separation = 28 Å



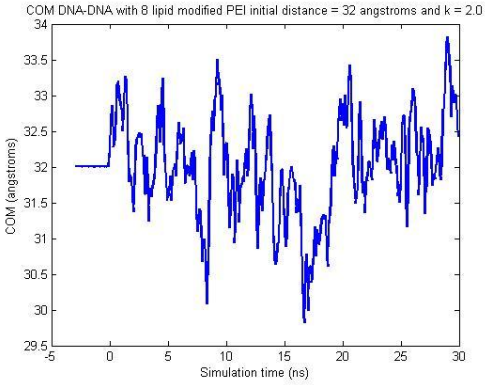
COM initial separation = 29 Å



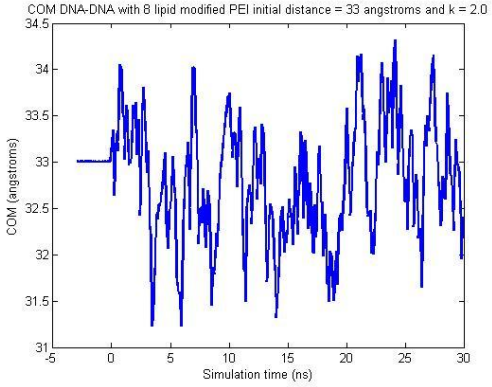
COM initial separation = 30 Å



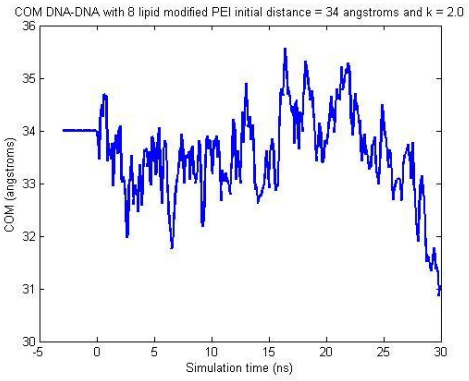
COM initial separation = 31 Å



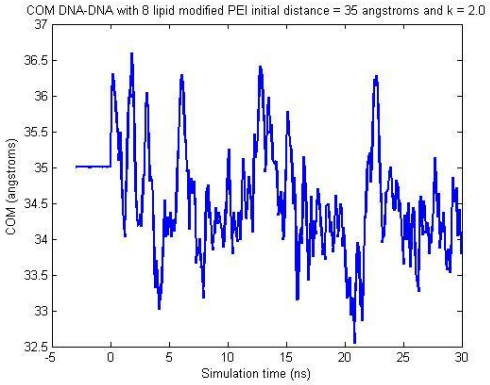
COM initial separation = 32 Å



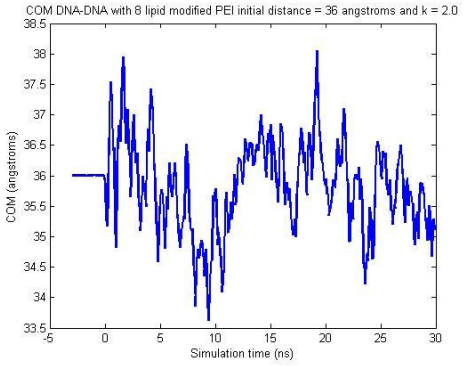
COM initial separation = 33 Å



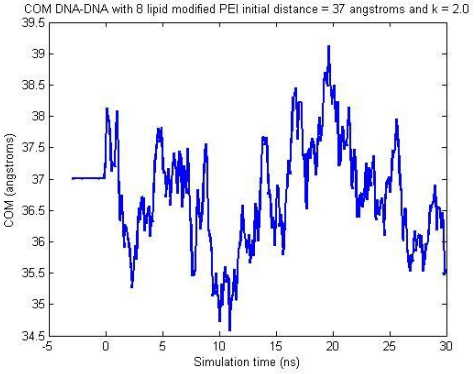
COM initial separation = 34 Å



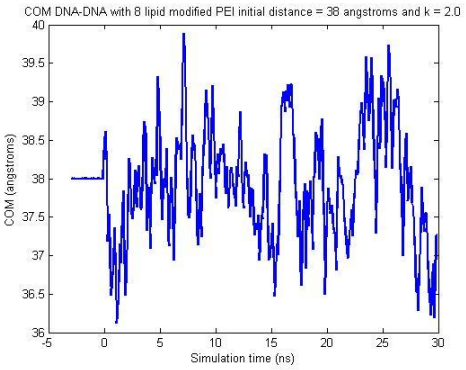
COM initial separation = 35 Å



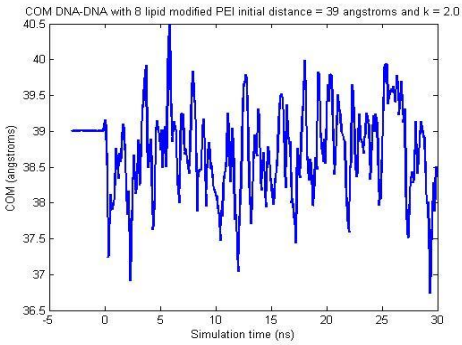
COM initial separation = 36 Å



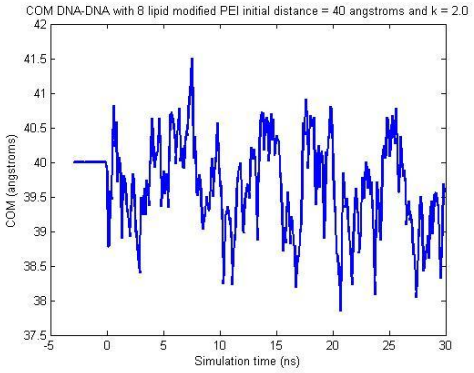
COM initial separation = 37 Å



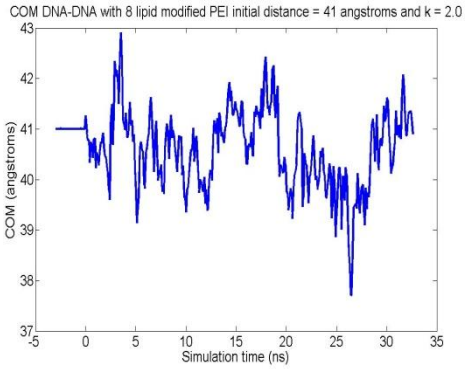
COM initial separation = 38 Å



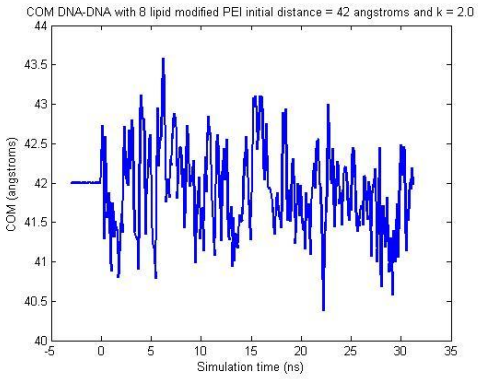
COM initial separation = 39 Å



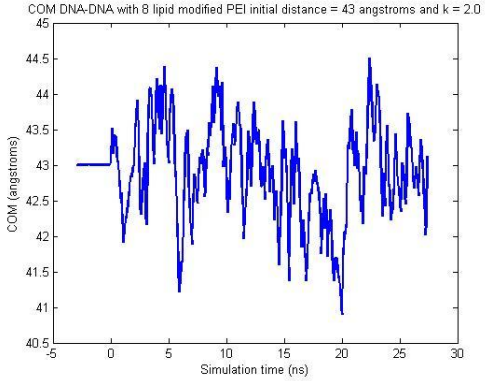
COM initial separation = 40 Å



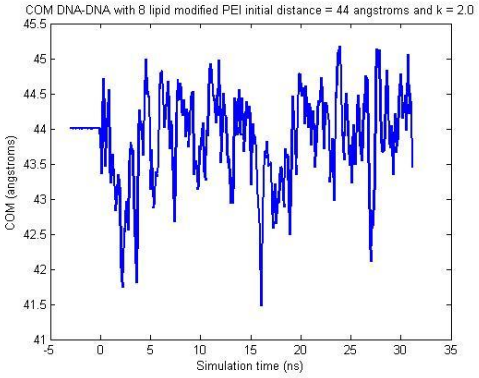
COM initial separation = 41 Å



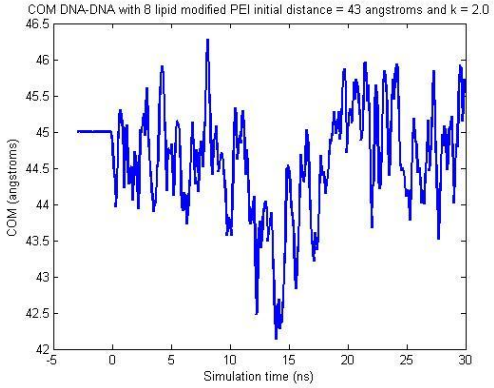
COM initial separation = 42 Å



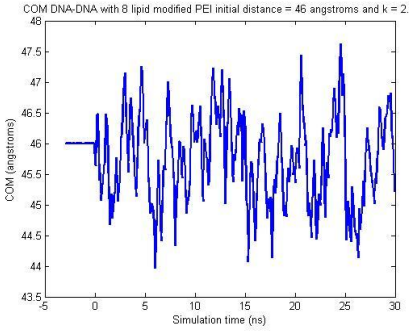
COM initial separation = 43 Å



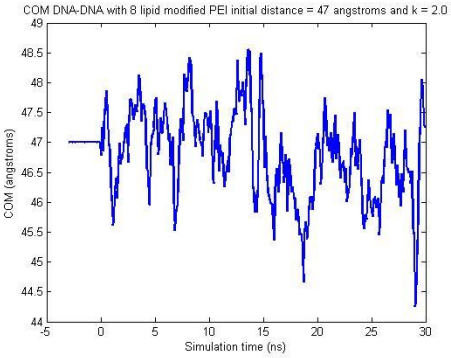
COM initial separation = 44 Å



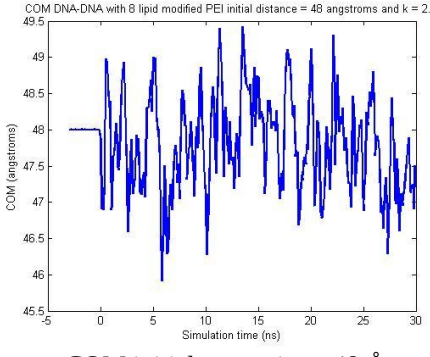
COM initial separation = 45 Å



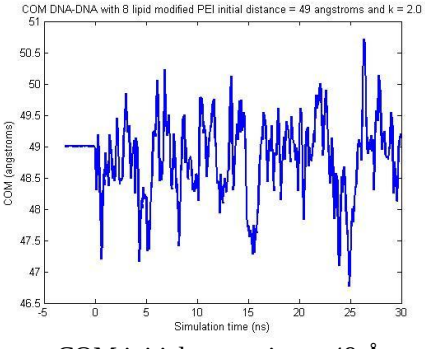
COM initial separation = 46 Å



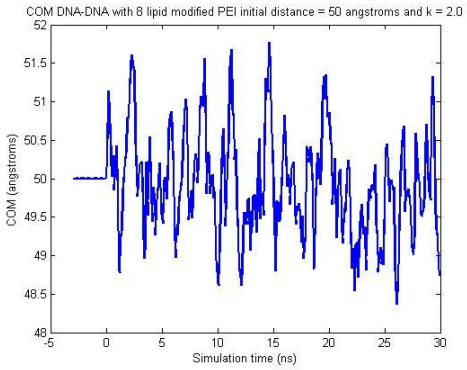
COM initial separation = 47 Å



COM initial separation =48 Å

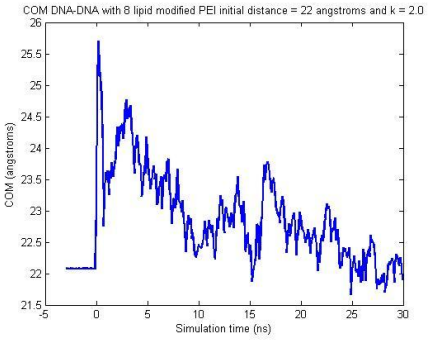


COM initial separation = 49 Å

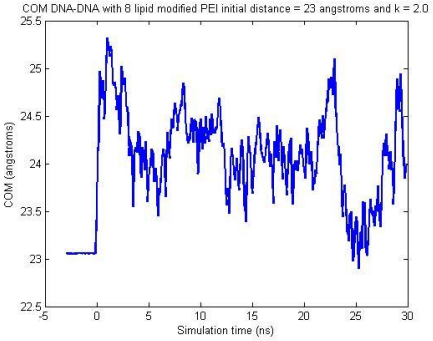


COM initial separation = 50 Å

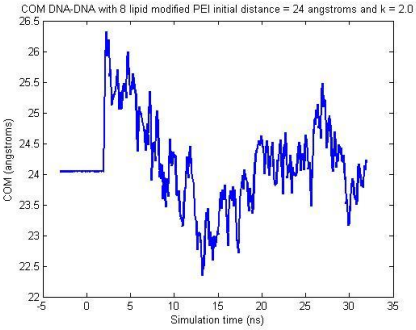
System 2D-8P(CA)



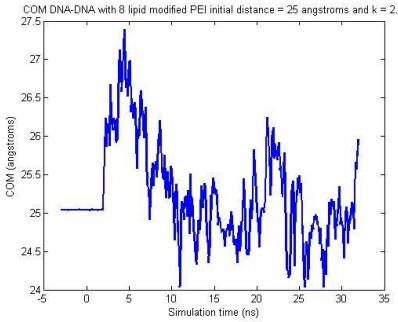
COM initial separation = 22 Å



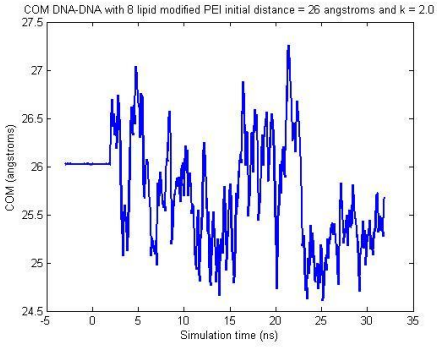
COM initial separation = 23 Å



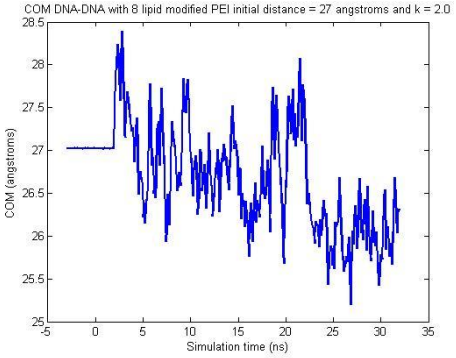
COM initial separation = 24 Å



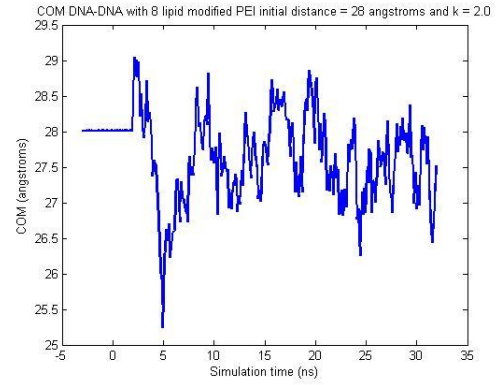
COM initial separation = 25 Å



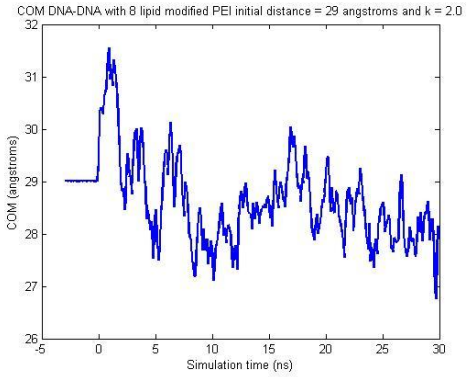
COM initial separation = 26 Å



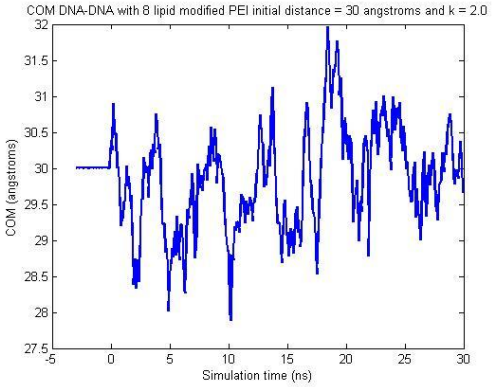
COM initial separation = 27 Å



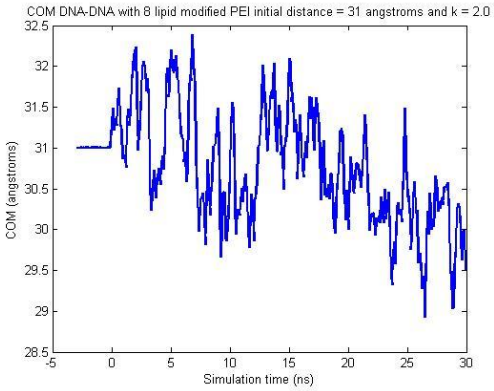
COM initial separation = 28 Å



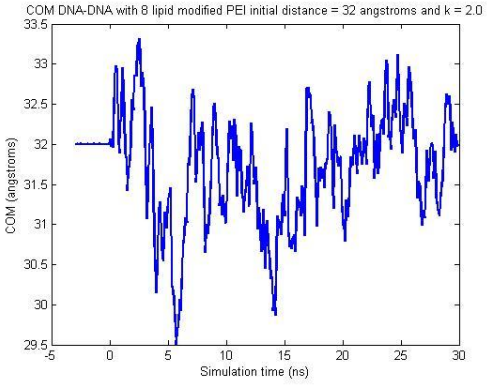
COM initial separation = 29 Å



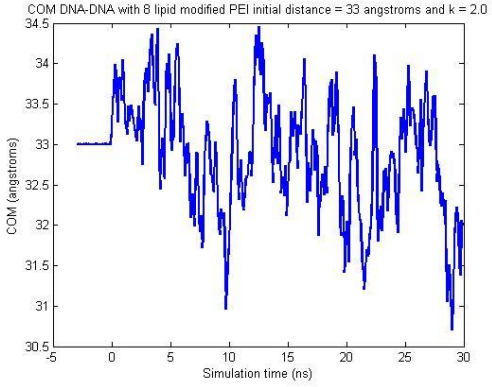
COM initial separation = 30 Å



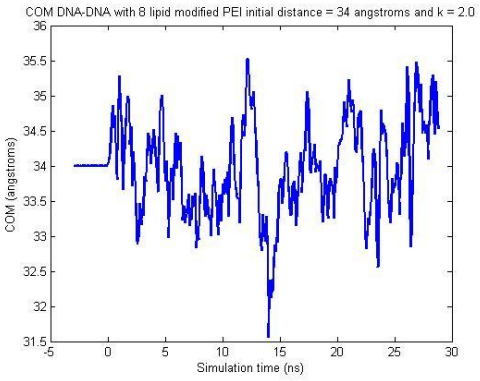
COM initial separation = 31 Å



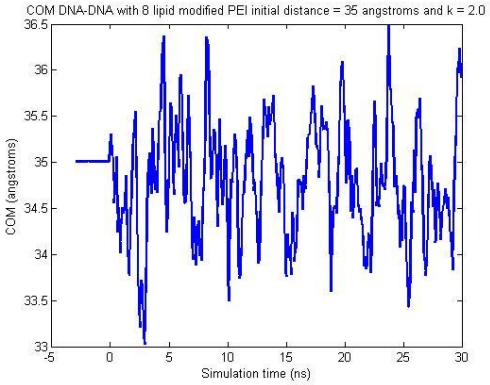
COM initial separation = 32 Å



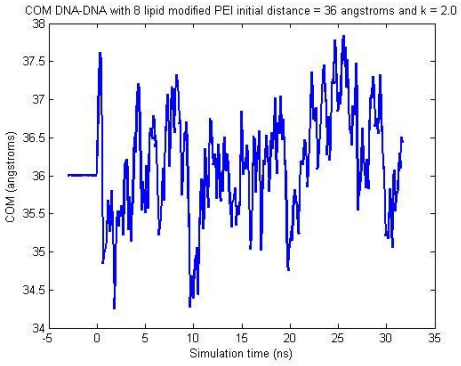
COM initial separation = 33 Å



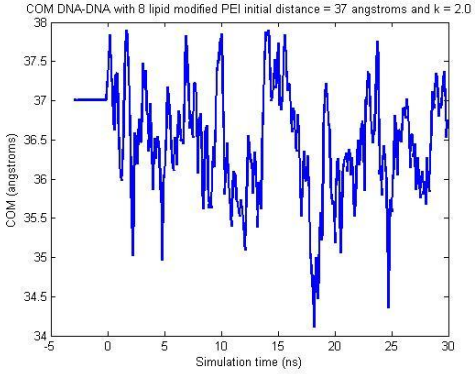
COM initial separation = 34 Å



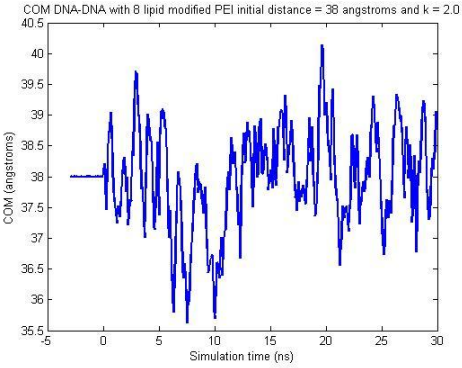
COM initial separation = 35 Å



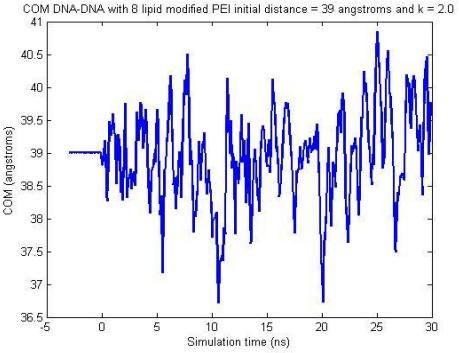
COM initial separation = 36 Å



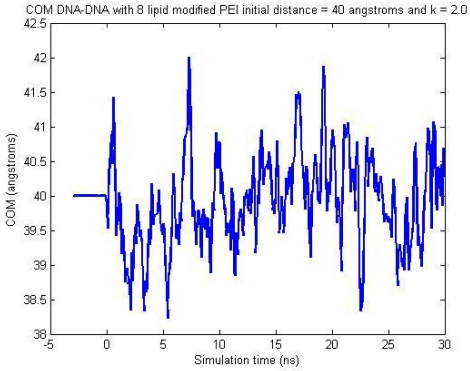
COM initial separation = 37 Å



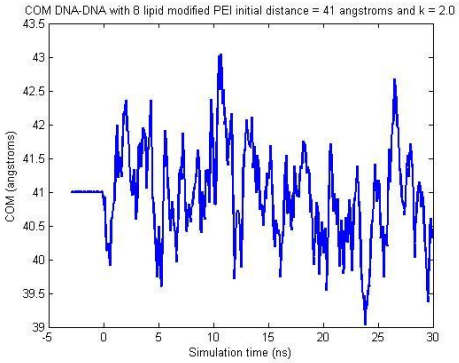
COM initial separation = 38 Å



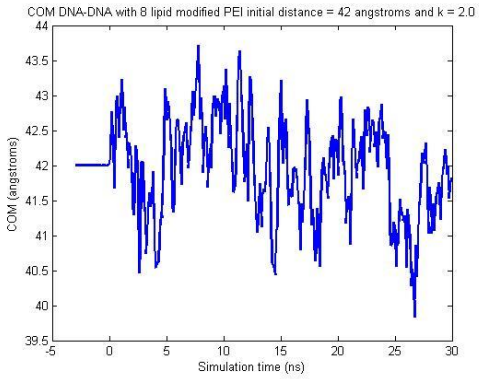
COM initial separation = 39 Å



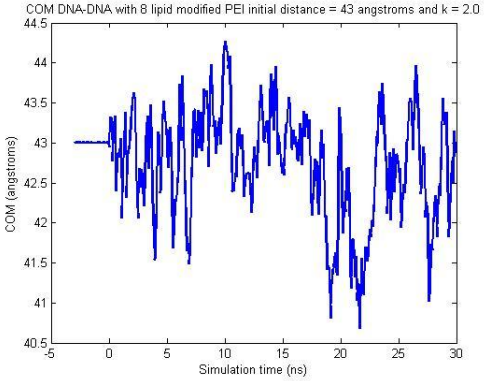
COM initial separation = 40 Å



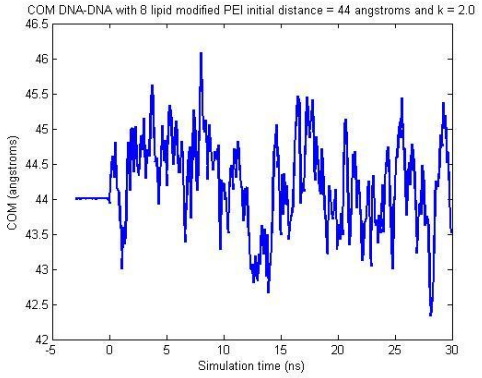
COM initial separation = 41 Å



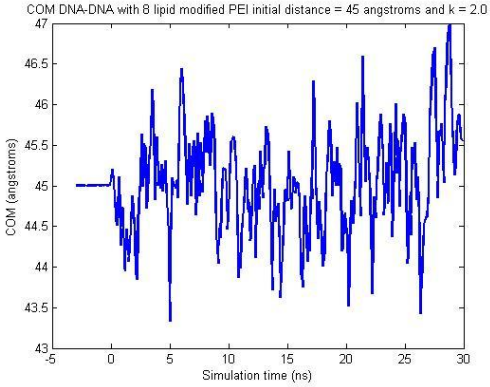
COM initial separation = 42 Å



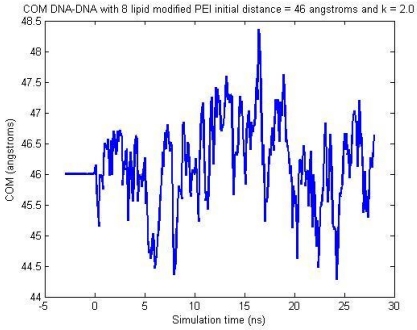
COM initial separation = 43 Å



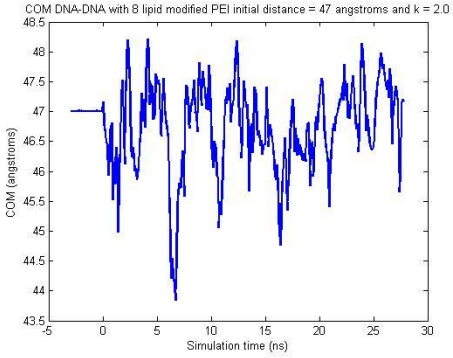
COM initial separation = 44 Å



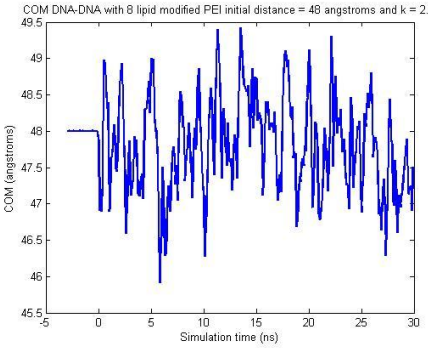
COM initial separation = 45 Å



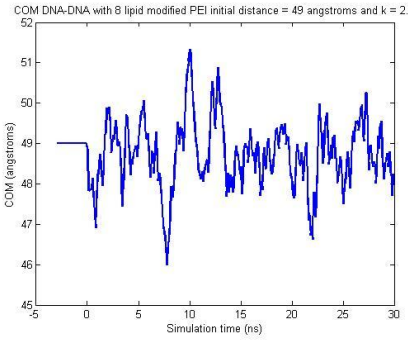
COM initial separation = 46 Å



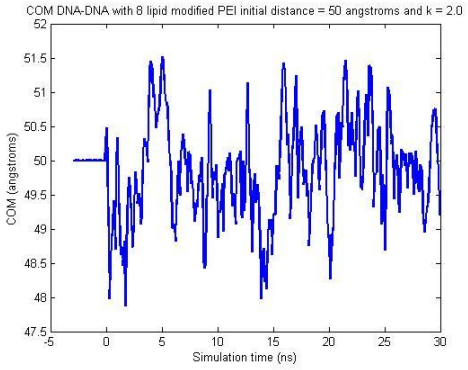
COM initial separation = 47 Å



COM initial separation =48 Å

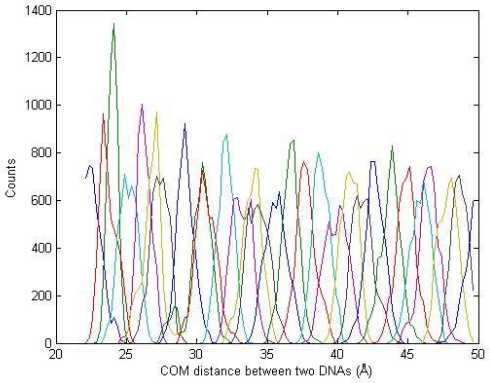


COM initial separation = 49 Å

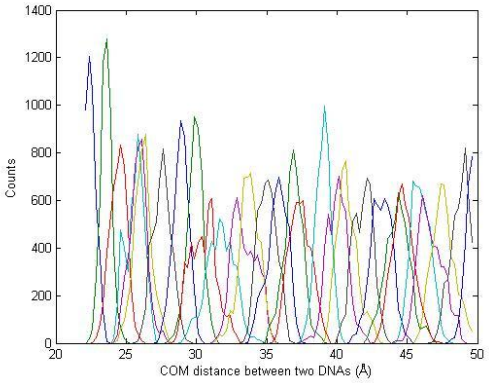


COM initial separation = 50 Å

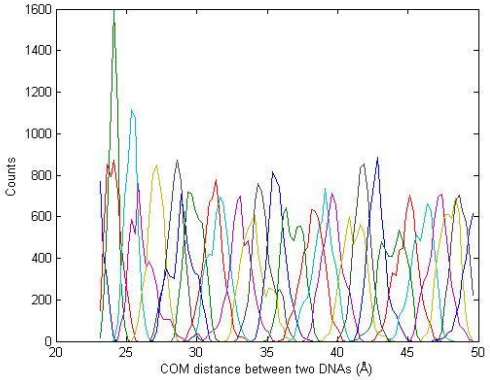
A-3 Histogram Charts



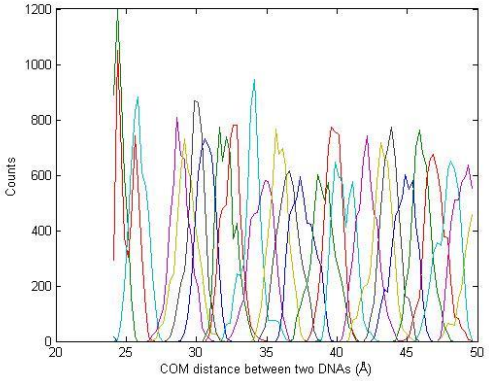
Histogram for 2D-6P



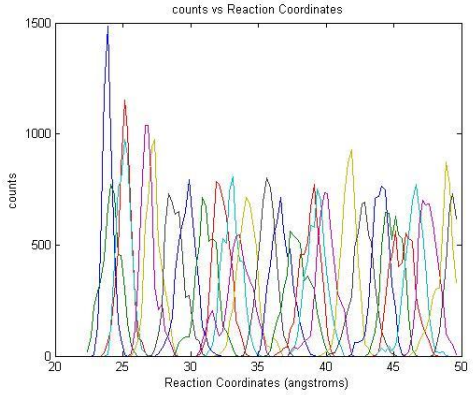
Histogram for 2D-4P



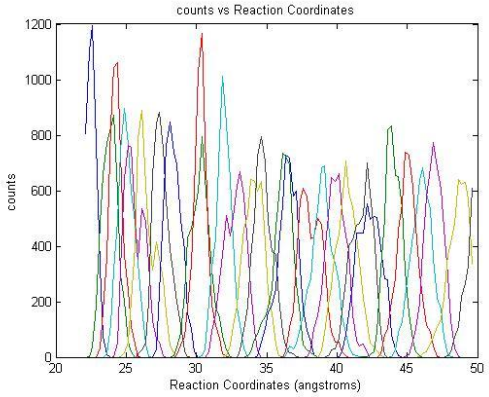
Histogram for 2D-2P



Histogram for 2D-8P(23%)



Histogram for 2D-8P(LA)



Histogram for 2D-8P(CA)