

A Coarse-Grained Simulation Framework to Study Polyethylenimine-DNA Nanoparticles
in Gene Delivery

by

Subhamoy Mahajan

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Mechanical Engineering

University of Alberta

© Subhamoy Mahajan, 2022

Abstract

Cellular DNAs contain genetic information vital for the functioning of the cells. Corruption of this information, or genetic disorders, can lead to various diseases. A potential treatment is to deliver exogenous DNAs with the correct genetic information to malignant cells to achieve a therapeutic response. However, DNAs are prone to degradation and are not efficient in overcoming cellular barriers, thus requiring specialized gene carriers. Among the non-viral gene carriers, the polymer polyethylenimine (PEI) has shown potential. Adding PEI to DNA forms nanoparticles (NPs) that protect DNA from degradation and help it overcome cellular barriers such as cellular uptake, endosomal escape, and nuclear trafficking. The efficacy of gene delivery depends on the properties of PEIs and NPs but their relationship is not well understood. Current experimental studies are limited because molecules inside cells cannot be observed with infinite precision, whereas current molecular simulations have not modeled large systems relevant for gene delivery.

This dissertation studies various steps of PEI-DNA gene delivery using large-scale coarse-grained (CG) molecular dynamics simulations. Three main studies have been performed. The first study includes the development of a CG forcefield for PEIs that capture its diverse molecular properties (degree of branching, molecular weight, and protonation ratio) and interaction with DNA. In the second study, the molecular aggregation mechanism behind PEI-DNA NP formation was explored using CG simulations with a large number of PEIs and DNAs at different N/P ratios. The aggregation was found to be dependent on the diffusion of PEIs, DNAs, and NPs and their electrostatic interactions. The N/P ratio was found to be an important control parameter for electrostatic interactions that can alter NP properties such as size, size distribution, shape, and rate of NP growth. Furthermore, for a high N/P ratio, a two-step addition of PEIs was found to make the NPs smaller and more spherical, which has the potential to increase the efficacy of cellular

uptake. The third study performed large-scale CG simulations to determine the effects of endosomal acidification on PEI-DNA NPs, an inevitable step in gene delivery. Simulations of endosomal acidification revealed that NP undergoes structural changes. NPs prepared at low N/P ratio underwent further aggregation, whereas at high N/P ratio they dissociated. Dissociation of NPs increased the osmotic pressure and reduced the NP's size that respectively help endosomal escape and nuclear trafficking. These findings support the observation of the strong efficacy of gene delivery at a high N/P ratio. The structural changes in the NP during dissociation were explained using a free energy landscape of PEIs, which revealed dissociation to be driven by repulsion between PEIs bound to the same DNA pair and repulsion between DNAs. These observations suggest a PEI with moderate molecular weight and degree of branching can increase NP dissociation and thereby the efficacy of gene delivery.

To assist the comparison of molecular simulations with experimental fluorescence microscopy used to study gene delivery, a new *in-silico* fluorescence microscopy technique was developed. The new technique converted molecular simulation trajectories into images that are comparable to the images obtained from experimental fluorescence microscopy. The cross-validation of *in-silico* images, experimental images, and molecular simulations bridged their analysis and generated new information such as determining the occurrence of NP dissociation in experimental images that were not originally reported. Comparison of *in-silico* images and molecular simulations can also determine equivalence of properties for future comparison between experiments and simulations. Furthermore, the comparison can be used to assess and develop image analysis algorithms for experimental images.

Overall, this dissertation developed a framework for performing and analyzing large-scale CG simulations of different steps in PEI-DNA gene delivery and its comparison with experiments.

Preface

All chapters in this thesis are my original work. Some chapters have been submitted or published in peer-reviewed journals and preprint as described below. Therefore, some chapters have their own introduction, methods, results, discussions, and references, and their supporting information is added in the Appendix. The abbreviations used in the thesis are available in the list of abbreviations and are defined in chapters after its first use. Certain symbols and terminology have been modified from the published articles to maintain consistency throughout the thesis. Acknowledgements specific to the chapter are provided before the chapter's references.

A version of **Chapter 3** has been published as (1) Mahajan S., and Tang T., Martini Coarse-Grained Model for Polyethylenimine. *J. Comp. Chem* **2019**, 40, 607-618, (2) Mahajan S., and Tang T., Erratum: "Martini Coarse-Grained Model for Polyethylenimine" [J. Comput. Chem. 2019, 40, 607-618 DOI:10.1002/jcc.25747]. *J. Comp. Chem* **2020**, 41, 1730-1734, and (3) Mahajan S., and Tang T., Comment on "Martini force field for protonated polyethyleneimine", *J. Comp. Chem* **2021**, 42, 261-263. As the first author, I am responsible for designing, performing, and analyzing simulations, generating figures, and writing and revising the manuscript. As the supervisor Dr. Tian Tang contributed to analyzing simulation results and revising the manuscript. Furthermore, Dr. Tian Tang was responsible for formulating the research question for manuscript (1), acquiring financial support, and arranging computational resources.

A version of **Chapter 4** has been published as Mahajan S., and Tang T., Polyethylenimine-DNA Ratio Strongly Affects Their Nanoparticle Formation: A Large-Scale Coarse-Grained Molecular Dynamics Study, *J. Phys. Chem. B* **2019**, 123, 9629-9640. As the first author, I was responsible for designing, performing, and analyzing simulations, conceptualizing the research question, generating figures, and writing and revising the manuscript. As the supervisor Dr. Tian

Tang contributed to analyzing simulation results, verifying the accuracy of aggregation mechanism, revising the manuscript, acquiring financial support and arranging computational resources.

Chapter 5 has been submitted for publication as Mahajan S., Tang T., Polyethylenimine-DNA Nanoparticles Under Endosomal Acidification and Implication to Gene Delivery. As the first author, I was responsible for designing, performing, and analyzing simulations, conceptualizing the research question, generating figures, writing and revising the manuscript, and writing the open-source computer code. As the supervisor Dr. Tian Tang contributed to analyzing simulation results, verifying the accuracy of dissociation mechanism, revising the manuscript, acquiring financial support, and arranging computational resources.

Chapter 6 has been submitted for publication as Mahajan S., Tang T., Meeting Experiments at the Diffraction Barrier with In-silico Fluorescence Microscopy and has been published as a preprint Mahajan S., Tang T., Meeting Experiments at the Diffraction Barrier: an In-silico Widefield Fluorescence Microscopy, *bioRxiv* doi:10.1101/2021.03.02.433395. As the first author, I was responsible for designing, performing, and analyzing simulations, conceptualizing the research question, generating figures, writing and revising the manuscript, and writing the open source computer code. As the supervisor Dr. Tian Tang contributed to analyzing simulation results, developing the color mixing algorithm, revising the manuscript, acquiring financial support, and arranging computational resources.

Chapter 7 includes revised versions of Mahajan S. *et al. J. Comput. Chem.* **2019** 40, 607-618, Mahajan S. *et al. J. Phys. Chem. B* **2019**, 123, 9629-9640, and the two submitted manuscripts discussed in **Chapter 5, 6**. Additionally, some future perspectives are discussed that have not been published elsewhere.

Acknowledgments

I had the privilege of working on an exciting Ph.D. research project that has expanded my thinking and curiosity. First, I would like to thank my supervisor, Dr. Tian Tang, for her expert advice and for training me to be an independent researcher. I am also grateful for her support and excitement in exploring research areas that were new to both of us.

I would like to acknowledge the financial support from Natural Sciences and Engineering Research Council of Canada (NSERC) through my supervisor that has made my Ph.D. program possible. I am grateful for the recognition of my Ph.D. research from MITACS Globalink Graduate Fellowship, Sadler Graduate Scholarship, Alberta Graduate Excellence Scholarship, and RR Gilpin Memorial Scholarship.

I am indebted to my early mentors, my high school physics teacher S. Remankantha Kurup who instilled in me the love of science and logical reasoning, and Dr. Puneet Patra and Dr. Baidurya Bhattacharya for introducing me to molecular dynamics and statistical mechanics. Without them, I would not think of pursuing graduate studies.

My Ph.D. program would be very difficult without the support of my colleagues and friends, Dr. Wenyuan Sun, Rajorshi Paul, Dr. Yousef Nademi, Zhen Yang, Dr. Deniz Meneksedag-Erol, Dr. Xiaoyu Sun, and Dr. Sajjad Kaviani.

I would also like to thank my parents Priya Bala Mahajan and Parikshit Mahajan, and my older brother Tanmoy Mahajan for supporting my interest in Engineering and Mathematics. Special thanks to Natasha Ting for her love that has sustained me through the critical years of my Ph.D. and the pandemic, and for her encouragement that has allowed me to push through my limits.

Table of Contents

1. Introduction	1
1.1. Characteristics of PEI and DNA	3
1.2. PEI-based gene delivery pathway	4
1.3. Nanoparticle formation.....	7
1.4. Cellular attachment and uptake	9
1.5. Endosomal escape	11
1.6. Nuclear trafficking	12
1.7. Theoretical studies.....	13
1.8. Motivation and objective.....	17
1.9. Thesis outline	18
1.10. References.....	19
2. Simulation Methodology	31
2.1. Equations of motion	32
2.1.1. NVE ensemble.....	34
2.1.2. NVT ensemble.....	34
2.1.3. NPT ensemble	36
2.2. Forcefields	38
2.2.1. Bonded interactions	38
2.2.2. Non-bonded interactions	39
2.2.3. Modifications to non-bonded interactions.....	40
2.2.4. AA and CG forcefields	42
2.3. Radial distribution function.....	44
2.4. Potential of mean force.....	45
2.5. References	46
3. Coarse-Grained Model for Polyethylenimine	49
3.1. Introduction	49
3.2. Model.....	53
3.2.1. CG mapping.....	55
3.2.2. Non-bonded interactions and selection of bead type.....	56

3.2.3.	Bonded interactions	58
3.2.4.	Nomenclature for bonded interactions	61
3.2.5.	AA simulation methodology	63
3.2.6.	CG simulation methodology.....	64
3.2.7.	Potential of mean force calculation	66
3.3.	Results and discussions	67
3.3.1.	Bonded parameters	67
3.3.2.	Global properties	70
3.3.3.	Quality of the water model	72
3.3.4.	DNA-PEI interaction	74
3.3.5.	Applications of the CG-PEI model.....	75
3.4.	Conclusion.....	76
3.5.	Acknowledgements	77
3.6.	References	77
4.	Polyethylenimine-DNA Nanoparticle Formation.....	85
4.1.	Introduction	85
4.2.	Methods.....	88
4.3.	Results	91
4.3.1.	Mode of aggregation and roles of PEI.....	91
4.3.2.	NP charge and size	98
4.3.3.	Rate of NP growth.....	104
4.3.4.	Influence of salt concentration	106
4.4.	Discussions.....	108
4.5.	Conclusions	113
4.6.	Acknowledgements	114
4.7.	References	114
5.	Endosomal Acidification of PEI-DNA Nanoparticles	120
5.1.	Introduction	120
5.2.	Results	123
5.2.1.	Acidic resizing.....	123
5.2.2.	Acidification of NPs prepared with $\alpha = 2$	126

5.2.3.	Acidification of NPs prepared with $\alpha = 10$	129
5.2.3.1.	Kinetics of bridging PEIs	131
5.2.3.2.	NP structural changes	134
5.2.3.3.	NP swelling and shrinkage	138
5.3.	Discussions	140
5.4.	Conclusions	144
5.5.	Methods	145
5.5.1.	Simulations	145
5.5.1.1.	Aggregation simulations	146
5.5.1.2.	Acidification simulations	148
5.5.2.	Free energy landscape	149
5.5.3.	<i>In-silico</i> fluorescence microscopy	150
5.5.4.	Transition diagram	151
5.5.5.	Radius of gyration	153
5.6.	Data and code availability	154
5.7.	Acknowledgements	154
5.8.	References	154
6.	<i>In-silico</i> Fluorescence Microscopy	160
6.1.	Introduction	160
6.2.	Results	164
6.2.1.	Setup of the <i>in-silico</i> microscope	164
6.2.2.	Generating <i>in-silico</i> monochrome image	167
6.2.3.	Generating <i>in-silico</i> microscopy image and video	170
6.2.4.	Applications	173
6.3.	Discussions	179
6.4.	Conclusions	182
6.5.	Data and code availability	183
6.6.	Acknowledgements	183
6.7.	References	184
7.	Conclusion and Future Perspective	190
7.1.	Overall conclusions	190

7.2. Future perspectives.....	194
7.2.1. Extension of the PEI forcefield	194
7.2.2. Further explorations on nanoparticle formation	195
7.2.3. Further explorations on endosomal acidification	195
7.2.4. Future of CG-MD simulations of PEI-DNA gene delivery.....	196
7.3. References	196
Unified Bibliography.....	198
Appendix A: Supporting Information for Chapter 3	237
Appendix B: Supporting Information for Chapter 4	259
Appendix C: Supporting Information for Chapter 5	275
Appendix D: Supporting Information for Chapter 6	286

List of Tables

Table 3.1: Hydration and partitioning free energy of CG beads.....	57
Table 4.1: Details of simulated systems.....	90
Table 5.1: Details of aggregation simulations.....	147
Table 5.2: Details of acidification simulations.....	149

List of Figures

Figure 1.1: Representative molecular structures of branched and linear PEI.	4
Figure 1.2: Cellular barriers of non-viral gene delivery.....	6
Figure 1.3: Cellular uptake pathways.....	10
Figure 3.1 AA and CG structures of PEI.	54
Figure 3.2: Validation of bonded parameters.....	69
Figure 3.3: Validation of global properties.	71
Figure 3.4: Comparison of water models.	73
Figure 3.5: Validating compatibility of PEI and DNA forcefield.....	75
Figure 4.1: All-atom (black) and coarse-grained (blue) structures of PEI used in this study.	89
Figure 4.2: Largest NP formed at the end of 4 μ s simulation.	92
Figure 4.3: Number of PEIs in different roles, plotted against simulation time.	93
Figure 4.4: Snapshots of a single PEI binding to a DNA.....	95
Figure 4.5: Steady-state NP size distributions (black) and average charge of NPs as a function of size (red).	99
Figure 4.6: Characteristics of charge and size of NPs.....	101
Figure 4.7: Size of NP and rate of NP growth.	105
Figure 4.8: Influence of salt concentration on NP aggregation.....	108
Figure 5.1: Schematic representation of the average number of DNAs in a NP vs. PEI/DNA number ratio (α) under physiological (grey) and endosomal (black) pH.	121
Figure 5.2: Properties of aggregated NPs in physiological and endosomal pH.....	124
Figure 5.3: Properties of NPs under endosomal acidification for $\alpha = 2$	127
Figure 5.4: Properties of NPs under endosomal acidification for $\alpha = 10$	131
Figure 5.5: PEI roles and their transitions during endosomal acidification simulations for $\alpha = 10$	133
Figure 5.6: Principal transition diagram of NPs in acidification simulations for $\alpha = 10$	137
Figure 5.7: Radius of gyration of NPs in acidification simulations for $\alpha = 10$	139
Figure 5.8: Structures of PEI at physiological and endosomal pH.	146
Figure 5.9: Generation of principal transition diagram.....	152
Figure 6.1: Framework of <i>in-silico</i> microscopy.....	163

Figure 6.2: Schematic of the <i>in-silico</i> microscope.....	165
Figure 6.3: <i>In-silico</i> microscopy images generated with different parameters.	169
Figure 6.4: Demonstration of color mixing.	171
Figure 6.5: Images with different color combinations.	172
Figure 6.6: Gaining new information by comparing <i>in-silico</i> images, experimental images, and detailed MS data.....	174
Figure 6.7: Bridging data obtained from MS and <i>in-silico</i> images.....	176
Figure 6.8: Single-particle tracking using <i>in-silico</i> microscopy on MS of DMPC lipid bilayer	179

List of Symbols

Latin Alphabets

a, a_1, a_2	A real number; Chapter 2, Appendix D
$a_{\mathcal{N}}$	Variance of Gaussian noise; Appendix D
$a_{\mathcal{P}}$	Effectiveness of Poisson noise ($a_{\mathcal{P}} \in [0,1]$); Appendix D
A_{macro}	Macroscopic property of A ; Chapter 2
A_{micro}	Microscopic property of A ; Chapter 2
b	Bridging PEI; Chapter 4, 5, Appendix B, C
(B_{ℓ}, B_m, B_n)	Dimensions of the simulation box at a given time; Appendix D
$(B_{\ell}^*, B_m^*, B_n^*)$	Dimensions of the largest (across time) simulation box; Appendix D
\mathbf{c}	Tensor containing edge vectors of a triclinic simulation box (dimensions 3×3); Chapter 2
C	A constant; Chapter 2, Appendix D
${}^a C_b$	Number of ways of choosing b number from a ; Appendix A
d_i	Minimum distance between PEI and i^{th} DNA; Chapter 5
D	Diffusion coefficient; Appendix B
$\text{erf}(\cdot)$	Error function; Chapter 2
E	Total energy; Chapter 2
f	Free PEI; Chapter 4, 5, Appendix B, C
f_s	FWHM scaling factor; Chapter 6, Appendix D
F	Free energy; Chapter 5
\mathcal{F}_O	Object focal plane; Chapter 6
\mathcal{F}_I	Image focal plane; Chapter 6
$g^k(\cdot)$	k^{th} constraint equation; Chapter 2
$h(\mathbf{r})$	Reaction coordinate associated with position \mathbf{r} ; Chapter 2
\mathcal{H}	Hamiltonian of a system; Chapter 2
\mathbb{H}	Heaviside step function; Chapter 5, 6
HSV	Hue, saturation, value coordinate system; Chapter 6, Appendix D

H_a	Hue of color a , where $a \in \{r: \text{red}, g: \text{green}, b: \text{blue}, o: \text{orange}, y: \text{yellow}\}$; Appendix D
$\{H_j, V_j\}$	A set of saturated colors; Appendix D
(H_j, S_j, V_j)	Hue, saturation, and value of j^{th} color; Chapter 6, Appendix D
$(H_{mix}, S_{mix}, V_{mix})$	Hue, saturation, and value of mixed color; Chapter 6, Appendix D
i	An index; Chapter 2, 3, 5, Appendix B Square root of -1 ; Chapter 6, Appendix D
I	Resultant fluorophore intensity; Chapter 6, Appendix D
I_0	Maximum PSF intensity; Chapter 6, Appendix D
I_1	Maximum PSF intensity set as 1; Chapter 6, Appendix D
I_c	Noise corrupted image; Appendix D
I_0^i	Prediction of optimal I_0 after i iterations; Appendix D
I_{DNA}	Monochrome intensity of DNA; Chapter 6
I_{PEI}	Monochrome intensity of PEI; Chapter 6
I_{thres}	Threshold intensity for Manders' coefficients; Chapter 6
j	An index; Chapter 2, 3, 6, Appendix B, D
J_0	Zeroth order Bessel function of the first kind; Chapter 6, Appendix D
k	An index; Chapter 2, 3, 5, Appendix C
k	Wavenumber of emitted light; Appendix D
k'	Wavenumber of emitted light scaled with f_s ; Chapter 6, Appendix D
k_B	Boltzmann constant; Chapter 2, 3, 5, Appendix B
\mathcal{K}	Kinetic energy; Chapter 2
K_{ang}	Force constant of bond angle potential; Chapter 2, 3, Appendix A
K_{bond}	Force constant of bond length potential; Chapter 2, 3, Appendix A
K_{dih}	Force constant of dihedral angle potential; Chapter 2
$K_{dih.w}$	Force constant of w^{th} periodic potential; Chapter 3, Appendix A
$\tilde{\mathcal{K}}$	Kinetic energy sampled from a probability distribution; Chapter 2
l	An index; Chapter 2, 3, Appendix D
ℓmn	Right-handed coordinate system in the object; Chapter 6, Appendix D
$\ell' m' n'$	Right-handed coordinate system in the image; Chapter 6, Appendix D

(ℓ_j, m_j, n_j)	Position of i^{th} fluorophore in ℓmn coordinates; Chapter 6, Appendix D
L	Largest element in \mathbf{c} ; Chapter 2
\mathcal{L}	<i>In-silico</i> microscope; Chapter 6
L_j	Luminance of color with index j ; Appendix D
m	An index; Chapter 3
m_i	Mass of the particle associated with i^{th} generalized coordinate; Chapter 2 Mass of the i^{th} bead (particle) in a NP; Chapter 5
M	Magnification of <i>in-silico</i> microscope; Chapter 6, Appendix D
\mathbf{M}	Mass tensor (dimensions $3N \times 3N$); Chapter 2
\mathcal{MS}	Molecular specimen; Chapter 6, Appendix D
$MSD(t)$	Lateral mean squared displacement; Appendix D
M_1, M_2	Manders' colocalization coefficient; Chapter 6
M_n	Number average molecular weight; Chapter 1
M_w	Weight average molecular weight; Chapter 1
n	An integer; Chapter 5, 6, Appendix B, C, D
n	Optical axis; Chapter 6
n_b	Average number of bridging PEI between a pair of bridged DNAs; Chapter 4, 5, Appendix B, C
n_{dih}	An integer parameter in the dihedral angle potential; Chapter 2
$n_{dih,w}$	An integer parameter in w^{th} periodic potential; Chapter 3, Appendix A
n_I	Coordinate of image focal plane; Chapter 6
n_{nl}	Number of timesteps neighbor list is updated; Chapter 2
n_O	Coordinate of object focal plane; Chapter 6, Appendix D
N, N_1, N_2	Number of particles; Chapter 2
N	Number of fluorophore particles; Chapter 6, Appendix D
\mathcal{N}	Gaussian noise; Appendix D
NA	Numerical aperture of <i>in-silico</i> microscope; Chapter 6, Appendix D
$N_{a/d}$	Number of circles denoting aggregation/dissociation in a transition diagram; Appendix C
$N_{a/d,early}$	$N_{a/d}$ that occur before 50% of the simulation time; Appendix C

$N_{a/d,late}$	$N_{a/d}$ that occur before 50% of the simulation time; Appendix C
N_b	Number of bridging PEI; Chapter 4, 5, Appendix B, C
N_{beads}	Number of beads in a NP; Chapter 5, Appendix B, D
N_c	Number of constraints; Chapter 2
N_{dof}	Number of degrees of freedom; Chapter 2
N_{ex}	Number of circles denoting NPs exchanging DNAs in a transition diagram; Appendix C
N_f	Number of free PEI; Chapter 4, 5, Appendix B, C
N_{ion}	Number of ions (potassium or chloride); Chapter 4
N_p	Number of peripheral PEI; Chapter 4, 5, Appendix B, C
$N_{phos}(n)$	Number of phosphate beads in close contact with n protonated amine beads; Appendix B
$N_{s,early}$	$N_{structure}$ that occur before 50% of the simulation time; Appendix C
$N_{s,late}$	$N_{structure}$ that occur after 50% of the simulation time; Appendix C
$N_{structure}$	Number of NP structures in a transition diagram; Appendix C
$N_{transition}$	Number of transitions in one direction within a transition diagram; Appendix C
N_{DNA}	Number of DNAs; Chapter 4, 5, Appendix B, C
N_{NP}	Number of NPs; Chapter 5
N_{PEI}	Number of PEIs; Chapter 4, 5, Appendix B
OPD	Optical path difference; Appendix D
p	Peripheral PEI; Chapter 4, 5, Appendix B, C
\mathbf{p}	Momentum vector of N particles (dimensions $3N$); Chapter 2
$\mathcal{p}(\cdot)$	Probability density of a microstate; Chapter 2
\mathbf{p}_i	Momentum vector of i^{th} particle (dimensions 3); Chapter 2
p_i	i^{th} generalized momentum; Chapter 2
P	Thermodynamic pressure; Chapter 2
\mathcal{P}	Poisson noise; Appendix D
$PMF(\cdot)$	Potential of mean force; Chapter 2
$PSF(\cdot, \cdot, \cdot)$	Gandy point spread function; Chapter 6

P_v	Scalar virial pressure; Chapter 2
\mathbf{P}_v	Virial pressure tensor (dimensions 3×3); Chapter 2
P_q	Probability of finding a PEI at reaction coordinate q ; Chapter 5
P_{bond}^{AA}	Probability distribution of both lengths averaged across all bonds in AA simulations; Chapter 3
$(P_{dih}^{CG})_m$	Probability distribution of dihedral angle (sum of periodic potentials) obtained from CG simulations in the m^{th} iteration of iterative Boltzmann inversion; Chapter 3
$(P_{\ell'}, P_{m'}, P_{n'})$	Dimensions of the box over which PSF is calculated; Chapter 6, Appendix D
PSF_{GL}	Gibson and Lanni point spread function; Appendix D
PSF_{MG}	Modified Gandy point spread function; Appendix D
q	Reaction coordinate for free energy; Chapter 5, Appendix C
q_i	Charge associated with particle with generalized position r_i ; Chapter 2, 3
Q_{DNA}	Charge of a DNA; Chapter 4, 5, Appendix B
Q_{NP}	Charge of a NP; Chapter 4, 5, Appendix B
Q_{PEI}	Charge of a PEI; Chapter 4, 5, Appendix B
Q_{NP}^{pred}	Predicted charge of a NP; Chapter 4, 5, Appendix B
r	Radial distance; Chapter 2
\mathbf{r}	Position vector of N particles (dimensions $3N$); Chapter 2
\tilde{r}	Normalized radius; Appendix D
r'	Distance of point image from the optical axis; Chapter 6, Appendix D
r_b	Distance from which non-bonded interactions are modified; Chapter 2
r_{bind}	Binding distance between PEI and DNA; Chapter 5
$r_{bond,eq}$	Equilibrium bond length; Chapter 2, 3, Appendix A
r_c	Cut-off distance for non-bonded potential; Chapter 2
r_i	i^{th} generalized momentum; Chapter 2
\mathbf{r}_i	Position vector of i^{th} particle (dimensions 3); Chapter 2, Appendix D
$r_{i,com}$	Distance between i^{th} bead (particle) in a NP and center of mass of the NP; Chapter 5

r_{ij}	Distance between particles i and j ; Chapter 2, 3, Appendix B, D
r_{nl}	Distance for maintaining neighbor list; Chapter 2
\mathbb{R}_{NP}	Rate of NP growth; Chapter 4, Appendix B
r^k	k^{th} constrained distance for a particle pair; Chapter 2
$RDF(\cdot)$	Radial distribution function; Chapter 2
RGB	Red, green, blue coordinate system; Appendix D
R_e	End-to-end distance; Chapter 3, Appendix A
R_g	Radius of gyration; Chapter 3, 5, Appendix A, C
R_{hyd}	Hydrodynamic radius; Appendix B, D
s	Simulation box scaling factor; Chapter 2
s_{NP}	Size of a NP; Number of DNAs in a NP; Chapter 4, 5, Appendix B, C
$S(\cdot)$	Structure factor; Chapter 2
t	Simulation time; Chapter 2, 3, 4, 5, 6, Appendix B, D
t_0	Reference time for calculating MSD; Appendix D
t_{acid}	Total time to acidify all PEIs; Chapter 5
t_{sim}	Total simulation time; Chapter 5, Appendix C
t_g	Thickness of coverslip in non-design condition; Appendix D
t_g^*	Thickness of coverslip in design condition; Appendix D
t_{oil}	Thickness of immersion oil in non-design condition; Appendix D
t_{oil}^*	Thickness of immersion oil in design condition; Appendix D
t_s	Distance of object (in the specimen) from the coverslip; Appendix D
t_{s0}	Distance of object focal plane from the coverslip; Appendix D
T	Thermodynamic temperature; Chapter 2, 3, 5, Appendix B
T_k	Kinetic temperature; Chapter 2
U_{bond}^{AA}	Potential energy associated with P_{bond}^{AA} ; Chapter 3
U_{dih}^{AA}	Potential energy associated with P_{dih}^{AA} ; Chapter 3
$(U_{dih}^{CG})_m$	Potential energy associated with $(P_{dih}^{CG})_m$; Chapter 3
V	Thermodynamic volume; Chapter 2
$\mathcal{V}(\cdot)$	Scalar potential energy; Chapter 2

$\mathcal{V}_a(\cdot)$	Potential for a , where a is <i>ang</i> for bond angle, <i>bond</i> for bond length, <i>dih</i> for dihedral angle, <i>dih</i> s for sum of periodic potentials, <i>e</i> for electrostatic (Coulombic), and <i>LJ</i> for Lennard-Jones; Chapter 2, 3
$(V_{dih})_m$	Parameterized potential energy after m^{th} iteration of iterative Boltzmann inversion; Chapter 3
w	An index; Chapter 3
W	Total number of periodic in dihedral angle potential; Chapter 3
\mathcal{W}	Wiener noise; Chapter 2
x	Reaction coordinate for potential of mean force; Chapter 2
xyz	Right-handed coordinate system; Chapter 2, 3, 5, 6, Appendix C
(x_d, y_d, z_d)	Position of the detector; Appendix D
Greek Symbols	
α	PEI/DNA number ratio; Chapter 4, 5, Appendix B, C
α_c	α at which DNAs are just neutralized by PEIs; Chapter 5
α'	Number ratio of bound PEIs and DNAs; Chapter 5, Appendix B
α_c^p	α_c at physiological pH; Chapter 5
α_c^e	α_c at endosomal pH; Chapter 5
β	Scalar isothermal compressibility; Chapter 2
$\boldsymbol{\beta}$	Isothermal compressibility tensor (dimensions 3×3); Chapter 2
γ	An integer; Chapter 2
$\Gamma(\cdot)$	Gamma function; Chapter 2
$\delta(\cdot)$	Dirac's delta function; Chapter 2, 6
$(\delta U_{dih}^{\text{CG}})_m$	Difference between potential energy obtained from CG and AA simulations at m^{th} iteration of iterative Boltzmann inversion; Chapter 3
$(\delta V_{dih})_m$	$(\delta U_{dih}^{\text{CG}})_m$ curve fitted to the potential form of V_{dih} ; Chapter 3
ΔG^{hyd}	Hydration free energy; Chapter 3
$\Delta G_{AW}^{\text{part}}$	Partitioning free energy to move a solute from A to W, where A can be H (hexadecane), C (Chloroform), O (Octanol) and W is water; Chapter 3
Δr	Small change in radial distance; Chapter 2
Δt	Integration time step or simulation time step; Chapter 2

Δz	Defocus of the <i>in-silico</i> microscope; Appendix D
$\Delta \mathcal{K}$	Change in the kinetic energy in consecutive timesteps; Chapter 2
$(\Delta \ell', \Delta m', \Delta n')$	Voxel dimensions for PSF calculation; Chapter 6, Appendix D
$\epsilon(\cdot)$	Energy of a microstate; Chapter 2
ϵ_0	Permittivity of vacuum; Chapter 2, 3
ϵ_{ij}	Lennard-Jones energy parameter between particles <i>i</i> and <i>j</i> ; Chapter 2, 3
ϵ_r	Relative permittivity of the medium; Chapter 2, 3
η	Viscosity of a fluid; Appendix B
θ	An angle; Chapter 6, Appendix D
$\theta_{ang,eq}$	Equilibrium bond angle; Chapter 2, 3, Appendix A
θ_{ijk}	Bond angle associated with particles <i>i</i> , <i>j</i> , and <i>k</i> ; Chapter 2, 3
ϑ	Positive real number less than 1; Chapter 3
κ^2	Relative shape anisotropy; Appendix B
λ	Wavelength of emitted light in vacuum; Chapter 6
μ	Refractive index of virtual immersion oil; Chapter 6, Appendix D
μ_g	Refractive index of coverslip in non-design condition; Appendix D
μ_g^*	Refractive index of coverslip in design condition; Appendix D
μ_{oil}	Refractive index of immersion oil in non-design condition; Appendix D
μ_{oil}^*	Refractive index of immersion oil in design condition; Appendix D
μ_s	Refractive index of specimen; Appendix D
ξ	Fourier space frequency; Chapter 2
ρ	Global number density; Chapter 2
ρ_{lmn}	Number density of particles in <i>lmn</i> ; Chapter 6
$\varrho(\cdot)$	Pair distribution function; Chapter 2
σ_{ij}	Lennard-Jones distance parameter between particles <i>i</i> and <i>j</i> ; Chapter 2, 3
τ_T	Time constant for a thermostat; Chapter 2
τ_P	Time constant for a barostat; Chapter 2
$\phi_{dih,eq}$	Equilibrium phase of dihedral angle potential; Chapter 2
$\phi_{dih,eq,w}$	Equilibrium phase of w^{th} periodic potential; Chapter 3, Appendix A

ϕ_{ijkl}	Dihedral angle associated with particles $i, j, k,$ and l ; Chapter 2, 3
χ_i	Eigen values of gyration tensor. $i \in \{1,2,3\}$; Appendix B
ψ	Maximum half-angle in virtual immersion oil; Chapter 6, Appendix D
Ω	Inertia of the non-rigid boundary; Chapter 2
Operators	
$\arg(\cdot)$	Phase of a complex number; Chapter 6, Appendix D
\dot{a}	First order time derivative; Chapter 2
\ddot{a}	Second order time derivative; Chapter 2
$ \mathbf{a} $	Magnitude of a vector; Appendix D
A^T	Transpose of a tensor; Chapter 2
$\det(\cdot)$	Determinant of a tensor; Chapter 2
$\max(\cdot)$	Maximum of all arguments; Chapter 5, Appendix D
$\max_n(\cdot)$	n^{th} largest number after sorting all arguments; Chapter 6, Appendix D
$\min(\cdot)$	Minimum of all arguments; Chapter 6, Appendix D
$Tr(\cdot)$	Trace of a tensor; Chapter 2
\in	Belongs to; Chapter 2, Appendix D
\leftarrow or \rightarrow	Transition; Chapter 5, Appendix B, C
$*$	Convolution; Chapter 6, Appendix D
\forall	For all; Appendix D
∇_r	Gradient with respect to vector r ; Chapter 2
$\langle \cdot \rangle$	Average; Chapter 2, 4, 5, Appendix B, C, D
$\langle \cdot \rangle_{ss}$	Averaged across steady state time interval; Chapter 2, 5
$[\cdot]$	Greatest integer function; Chapter 3
\sum	Sum; Chapter 2, 3, 6, Appendix B, D
\prod	Product; Chapter 2

List of Abbreviations

AA	All-atom
ATP	Adenosine triphosphate
bp	Base pair
CD	Circular dichroism
CG	Coarse-grain
CGenFF	CHARMM General Force Field
CHARMM	Chemistry at Harvard Macromolecular Mechanics
CME	Clathrin-mediated endocytosis
CMY	Cyan-magenta-yellow
COM	Center of mass
COSM	Computational optical-sectioning microscopy
CvME	Caveolae-mediated endocytosis
DBI	Direct Boltzmann inversion
DNA	Deoxyribonucleic acid
DMPC	Dimyristoyl phosphatidylcholine
FEM	Finite element method
FFT	Fast Fourier transform
FRET	Förster resonance energy transfer
FWHM	Full-width-at-half-maximum
GL	Gibson and Lanni
HB	Hyperbranched
HIV	Human immunodeficiency virus
HMW	High molecular weight
HSV	Hue-saturation-value
HSL	Hue-saturation-luminance
IBI	Iterative Boltzmann inversion
JACoP	Just Another Colocalization Plugin

LMW	Low molecular weight PEI
mRNA	Messenger RNA
miRNA	Micro RNA
LINCS	Linear Constraint Solver
LMW	Low molecular weight
LJ	Lennard-Jones
MB	Moderately branched
MC	Monte Carlo
MCM	Mesoporous molecular sieve
MD	Molecular dynamics
MG	Modified Gandy
MS	Molecular simulation
MSD	Mean-squared displacement
MW	Molecular weight
NA	Nucleic acids
NP	Nanoparticle
NPA	Number-Pressure-Enthalpy
NPC	Nuclear pore complex
NLS	Nuclear localization sequence
NMR	Nuclear magnetic resonance
NPT	Number-Pressure-Temperature
NTF	Nuclear transport factor
NVE	Number-Volume-Energy
NVT	Number-Volume-Temperature
DNA	Deoxyribonucleic acid
OSM	Optical sectioning microscopy
OPD	Optical path difference
OCV	Orange-cyan-violet
PAMAM	Poly(amidoamine)

PBC	Periodic boundary condition
PEG	Polyethylene glycol
PEG-g-PEI	PEG grafted PEI
PEI	Polyethylenimine
PES	Potential energy surface
PME	Particle-mesh Ewald
PMF	Potential of mean force
PL	Purely linear
PSF	Point spread function
PW	Polarizable Martini water
RDF	Radial distribution function
RGB	Red-green-blue
RISC	RNA induced silencing complex
RNA	Ribonucleic acid
siRNA	Short interfering RNA
SCE	Smoluchowski coagulation equation
SIM	Structure illuminated microscopy
SL	Semi-linear
SOFI	Super-resolution optical fluctuation imaging
STED	Stimulated emission depletion
TIP3P	Transferable intermolecular potential 3 points
VMD	Visual molecular dynamics
W	Non-polarizable Martini water
WHAM	Weighted-histogram analysis method
YCM	Yellow-cyan-magenta

Chapter 1

Introduction

DNAs present in cells contain genetic information for producing proteins, which are responsible for most cellular functions.[1,2] Alterations to the genetic information can make cells malignant because toxic proteins are produced or essential proteins are not produced.[3,4] Gene therapy can potentially treat these genetic disorders and other diseases by gene addition to produce essential proteins, and gene silencing to eradicate toxic proteins.[5] Compared to unsuccessful gene therapies in the late 1970s,[6] the recent progress has been significant with over ten gene addition and silencing therapies approved for human use to treat neuromuscular disease, inherited blindness, cancer, blood disorders, *etc.*[7–10] It is also promising to note that over two thousand other therapies are undergoing trials to treat human immunodeficiency virus (HIV), hepatitis B, diabetes, *etc.*[10,11]

Gene addition can involve the delivery of exogenous DNAs such as plasmid DNA containing several genes or oligonucleotides containing one gene.[12,13] Within the nucleus, exogenous DNAs transcribe their genetic information onto a messenger RNA (mRNA), which in turn travels to the cytoplasm for protein production.[14] Consequently, delivery of exogenous mRNA into the cytoplasm can also achieve gene addition.[15] In gene silencing, short-interfering RNA (siRNA) or microRNA (miRNA) delivered to the cytoplasm forms an RNA-induced silencing complex (RISC) which cleaves mRNA in two.[16] This essentially stops gene expression, i.e., the production of a protein by the mRNA.

Gene delivery of nucleic acids (NAs; DNAs or RNAs) is necessary for gene therapies. In certain tissues and cells, gene delivery of naked NAs can be achieved by injection,[17] gene gun,[18] and electroporation.[19] With these techniques, the efficiency of gene expression is low[20] because NAs are prone to degradation and are not effective in entering cells.[21,22] A positive therapeutic response can be achieved by administering large volumes of NAs and multiple doses,[20,23] which has led to at least six therapies (Kynamro, Spinraza, Exondys 51, Neovasculgen, Macugen, and Vyondys 53) approved for human use with some restrictions due to their toxicity.[7,9,24,25]

For prolonged stability of NAs and their targeted systemic delivery, specialized gene carriers are required which can be viral or non-viral.[26] In the case of viral carriers, viruses such as recombinant adenovirus (a type of common cold virus) are used for their ability to invade cells and deliver genetic material.[27] The main drawbacks of a viral carrier are the possibility of a patient experiencing an adverse immune response[27] and developing immunity against the viral carrier.[28] Despite these limitations, seven viral carrier-based gene therapies (Zolgensma, Gendicine, Rexin-G, Glybera, Strimvelis, Luxturna, and Zynteglo) have recently been approved for human use with some restrictions.[7,9,24,25] Non-viral carriers are synthetically manufactured molecules whose properties can be modified to increase their efficacy of gene delivery and decrease the risk of an adverse immune response.[29] Furthermore, compared to viral carriers, they are easier and cheaper to manufacture.[29] However, current non-viral carriers are not as efficacious as viral ones, with several treatments undergoing clinical trials and none approved for human use.[10,26]

Among non-viral carriers of DNA, the polymer polyethylenimine (PEI) is promising for its high efficacy based on a mechanism not followed by viruses.[26,30] A thorough understanding of

the mechanism would allow the efficient design of PEI and potentially other polymeric non-viral carriers. The following sections, provide a detailed literature review of experimental (**Section 1.1-1.6**) and theoretical studies (**Section 1.7**) of PEI-based gene delivery, motivations and objectives of the thesis (**Section 1.8**), and a thesis outline (**Section 1.9**).

1.1. Characteristics of PEI and DNA

PEI is a polymer with repeating units of ethylamine, -C-C-N-, that can be synthesized with various molecular weights (MWs) and degrees of branching (**Figure 1.1**).[31,32] The MW of synthesized PEI follows a distribution, which is quantified by a number average (M_n) or a weight average molecular weight (M_w).[33] The heterogeneity of polymer's MW is quantified using the polydispersity index defined as the M_w/M_n . [33] Commercially available PEI can have M_n and polydispersity index in the range 0.6-800 kDa and 1.4-13 respectively.[34–36] The average degree of branching of a synthesized PEI is quantified by the molar ratio of primary, secondary, and tertiary amines in PEI solution.[37,38] To the best of our knowledge no experimental technique exists that can determine the distribution of degree of branching of PEIs in a solution. Due to the abundance of amine groups, PEI acts as a buffer over a wide range of pH.[37] At the physiological pH of ~7.4, PEI is positively charged and its protonation ratio has been reported to be 10-20%,[39] ~21-33%,[34] and 55-67%. [40] The difference in these results likely arises from the difference in the molar ratio of primary, secondary and tertiary amines, i.e., the average degree of branching. For example, Suh *et al.*[39] observed a protonation ratio of 10-20% for PEIs with 25% primary amines, and Tang and Szoka[40] observed a protonation ratio of 55-67% for PEIs with 50% primary amines.

DNAs contain two backbone strands that run anti-parallel to each other, where each strand is composed of negatively charged phosphate and deoxyribose sugar.[2] A base is attached to each

deoxyribose sugar, which can be adenine (A), guanine (G), thymine (T), or cytosine (C).[2] The ordering of these bases stores the genetic information in the DNA.[2] The base in each strand interacts with its complementary base in the other strand, i.e., the bases exist in pairs (base pairs) of A and T, and G and C.[2] The base pairs (bp) interact using hydrogen bonding, which gives DNAs get their characteristic double-helical structure with a major and minor groove.[2] Plasmid DNAs used for gene delivery have supercoiled circular (each strand forms a closed-loop) structures that are composed of several thousand base pairs,[26,41,42] whereas oligonucleotides are small, rigid, and have linear structures with 10-30 bp.[43]

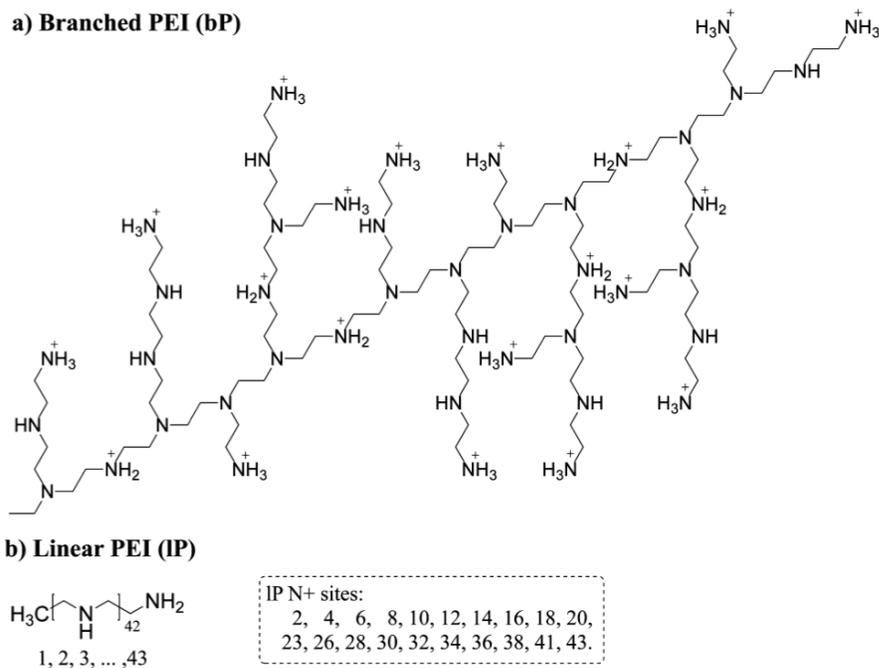


Figure 1.1: Representative molecular structures of branched and linear PEI. The box denotes the location of protonated nitrogen in linear PEI. [Reprinted with permission from Sun *et al.*[44] Copyright © 2012 American Chemical Society]

1.2. PEI-based gene delivery pathway

PEI-based gene delivery begins with the preparation of PEI-DNA nanoparticles (NPs) (Section 1.3). During *in vitro* gene delivery, a solution of NPs is added to the cells,[12] while in

in vivo gene delivery the NP solution can be injected into veins (intravenous injection[45]) and spine (cerebrospinal fluid injection),[46] or instilled into lungs.[47] The gene delivery process can be observed using fluorescence microscopy techniques, such as widefield, optical-sectioning,[48–50] and super-resolution[51–55]. Cellular components, PEIs, and DNAs can be marked with fluorophores that absorb and emit light of a specific range of wavelengths. [56] The emitted light, i.e., *fluorescence* is tracked visually or by image processing to provide information on the dynamics of NPs within the cells[57] with a spatial resolution of 20-200 nm and temporal resolution of milli-seconds to seconds.[54,58,59] Additionally, fluorescence microscopy provides information on various barriers experienced by NPs during gene delivery.[57]

In vivo delivery faces barriers even before they reach their target cells, i.e., *extracellular barrier*, whereafter both *in vivo* and *in vitro* deliveries experience the same *intracellular barriers* (**Figure 1.2**).[26] The first intracellular barrier is the attachment of NPs onto the cell surface and their subsequent entry into the cell known as *cellular uptake* (**Section 1.4**).[60] Uptake of NPs mainly occurs through endocytosis, where they are engulfed by the cell membrane and trapped in endosomes.[61] Endosomes are continuously acidified, making the environment hostile for DNAs.[62] Therefore, endosomal escape of NPs into the cytoplasm is required to preserve the DNAs, thus being the second barrier of gene delivery (**Section 1.5**).[63] Since DNAs can only transcribe their genetic information in the nucleus,[14] the final barrier is the nuclear trafficking of DNAs or NPs (**Section 1.6**).[64]

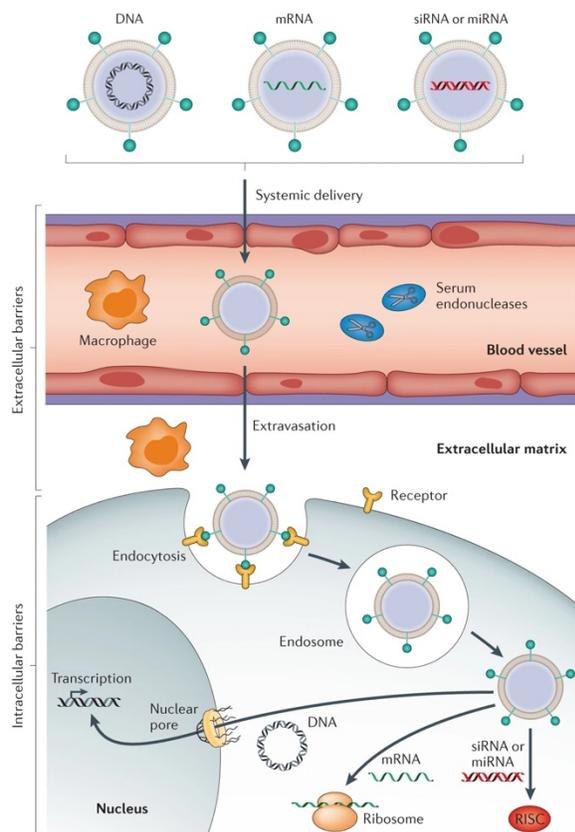


Figure 1.2: Cellular barriers of non-viral gene delivery. *In vivo* delivery experiences both extracellular and intracellular barriers, whereas *in vitro* delivery only experiences intracellular barriers.[26] Extracellular barriers include degradation of PEI-DNA NPs by macrophage and serum endonuclease, and intracellular barriers include cellular uptake through endocytosis, endosomal escape, and nuclear trafficking of DNAs[26] (siRNA and mRNA delivery is not discussed). [Reproduced with permission from Yin *et al.*[26] Copyright © 2014 Macmillan Publishers Limited.]

The efficacy of gene delivery is intricately tied to the properties of PEIs and NPs, which have not been fully elucidated. Here, details regarding the efficacy of the entire gene delivery process are provided, and details regarding each intracellular barrier are provided in the subsequent sections. High N/P ratio (ratio of PEI amines to DNA phosphates) of 5 to 10 have been reported to have the highest efficacy, although the exact number varied between different cell lines.[21,65,66] The presence of free PEIs after the formation of NPs at a high N/P ratio appears

to be crucial (**Section 1.3**). [67,68] Studies have reported a reduction in the efficacy when the free PEIs are removed, and an increase in the efficacy when free PEIs are added with a time delay (0 to 4 hours).[67,68] M_n of PEI associated with high efficacy is less clear, and both moderate[69] (12 kDa) and high[36] (70 kDa) values have been proposed to have high efficacy. Since PEIs are polydisperse, the high efficacy in these studies can be a result of any MW in their distribution. Only two studies have fractionated PEI (25 kDa) into groups of different MW and identified 4-10 kDa to be the most ideal.[70,71] Conflicting efficacy results have been reported for linear and branched PEI depending on the cell line.[72,73] Comparison between PEIs with different average degrees of branching is limited to only Krämer *et al.*[74], where an average degree of branching of 52% was found to be the most efficacious in two different cell lines.

Cytotoxicity of PEI has been a recent concern, as it was reported to correlate with successful gene delivery.[75] Specifically, it has been reported to increase with the concentration of the PEI or N/P ratio,[76,77] MW,[67,69,70] and degree of branching.[78] As a result, several studies have demonstrated that high efficacy and low cytotoxicity can be achieved using a linear or low MW PEI.[77,78]

1.3. Nanoparticle formation

The addition of PEIs neutralizes and condenses plasmid DNAs into compact structures that can further aggregate,[40,42,79] whereas for oligonucleotides condensation is not observed due to its rigidity.[80,81] The properties of PEIs control NP properties, which can in turn influence the efficacy of DNAs or NPs crossing different intracellular barriers(discussed later in **Section 1.4-1.6**).[73]

The charge of the NPs is quantified by their average zeta potential, which is mainly controlled by the N/P ratio.[73,81–83] At physiological pH, average zeta potential is negative for

N/P ratio < 2 and positive for N/P ratio > 4 . [73,81–83] The N/P ratio associated with zero average zeta potential is expected to vary with the protonation ratio of PEI at a given pH, which can, in turn, vary with its MW and degree of branching (**Section 1.1**). [34,39,40] At a high N/P ratio, some PEIs are unable to bind with DNA because NPs develop a positive zeta potential, and these PEIs can be found free in the solution. [68,84]

The shape of NPs has been studied for different N/P ratios. For PEI-plasmid DNA NPs, oblong, folded-loops, toroids, and circular structures are obtained at a low N/P ratio < 1 (0.08 charge equivalent), [42] while toroids and spheres are observed at a high N/P ratio > 8 . [40,42,79] In the literature, shapes of PEI-oligonucleotide NPs have only been reported for a high N/P ratio, where network structure and spheres are observed. [81] In general, the shape of NPs are visually assessed, [40,42,79,81] which has limited its characterization with detailed PEI properties.

The size of NPs is quantified by their hydrodynamic radius and is affected by the N/P ratio, ionic strength of the solution, and the MW and degree of branching of PEI. The mean hydrodynamic radius is observed to be the largest for moderate N/P ratio between 2 and 4, where the mean zeta potential is close to zero. [73,81,82,85] Increasing the ionic strength of the solution (higher salt concentration), increased the mean hydrodynamic radius, and the increase in hydrodynamic radius for NPs prepared with linear PEI was more pronounced than the ones prepared with branched PEI. [72] In the literature, the effect of degree of branching has only been compared between a linear and branched PEI (multiple degrees of branching are not compared). [42,85] In addition, the results in these studies varied between cell lines. [42,85] The distribution of hydrodynamic radius is also affected by the N/P ratio. For plasmid DNA, the distribution is bimodal for a low N/P ratio of 2, whereas it is unimodal for N/P ratio > 3 . [40,79]

Studies that reported NP size distribution for oligonucleotides are limited, with one study reporting unimodal distribution for N/P = 2.2.[83]

1.4. Cellular attachment and uptake

The surface of the cell membrane contains several lipids and proteins,[86] and their interaction with NPs, for cellular attachment, can be both specific and non-specific.[60,87] Non-specific cellular attachment can be enhanced by preparing positively charged NPs that interact with the negatively charged proteoglycans (a type of protein) on the cell membrane,[60,88] and large NPs that sedimentation onto the cell surface.[89] The presence of free PEIs also affect cell association, and has been reported to both increase[67] and decrease[68] the cell association of NPs. To target specific cell receptors (specific interaction), specialized ligands such as transferrin and folic acid can be used.[87,90] *In vivo* delivery has additional constraints. For example, if NPs are injected intravenously, highly charged NPs can interact with non-target cells (such as macrophages **Figure 1.2**) thereby reducing its efficacy.[26] Furthermore, interaction with non-target molecules can increase the NP's radius through aggregation, which can then block blood vesicles.[91] Once NPs are attached, their cellular uptake occurs through three endocytic pathways: macropinocytosis forming macropinosomes of diameter $> 1 \mu\text{m}$; clathrin-mediated endocytosis (CME) forming early-endosomes of diameter $\sim 90 \text{ nm}$; and Caveolae-mediated endocytosis (CvME) forming caveolar vesicle of diameter $\sim 60 \text{ nm}$ (**Figure 1.3**).[92,93] All pathways internalize NPs in vesicles that are continuously acidified, and the eventual fusion with lysosomes and acidic degradation is only avoided in CvME. [92]

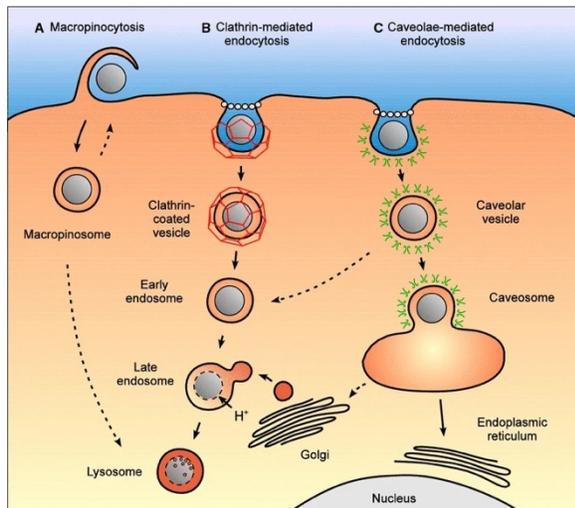


Figure 1.3: Cellular uptake pathways. (a) micropinocytosis, (b) clathrin-mediated endocytosis, and (c) caveolae-mediated endocytosis. [92] [Reproduced with permission from Hillaireau and Couvreur [92] Copyright © 2009 Birkhäuser Verlag, Basel/Switzerland]

Cellular attachment and uptake of PEI-based NPs are strongly influenced by the N/P ratio. Oh *et al.*[21] demonstrated that cellular uptake is possible for negatively-charged NPs formed at a N/P ratio of 1, and the uptake increases with the rise in N/P ratio i.e. with increasing NP charge. In the literature, cellular uptake of PEI-based NPs is believed to occur through CME, which traps the NPs in endosomes.[60] However, Rejman *et al.*[61] observed that both CME and CvME take place, with the efficacy of uptake dependent on the NP size.[61] Specifically, they found that CME can uptake NPs with a diameter of 50, 100, and 200 nm, and the uptake of 50 nm NPs was higher than the other two.[61] They also reported CvME to only uptake NPs with a diameter of more than 500 nm, which is inconsistent with other works that reported caveolar vesicle to be 50-100 nm in diameter.[93–95] In the literature, the effect of NP shape on cellular uptake has been discussed for several types of NP but not for PEI-based ones.[96,97]

1.5. Endosomal escape

Endosomes are continuously acidified by proton pumps at the rate of ~200 protons per second,[98] reducing their pH from the physiological (~7.4) to late-endosomal (~5.5).[99] At such acidic pH, DNAs are prone to degradation.[62] PEI's high efficacy is largely associated with its ability to escape endosomes, and protect the DNAs from degradation until their escape.[30,62] The escape mechanism is believed to be a consequence of its buffering capacity over a wide range of pH,[40] and is referred to as the “proton sponge” hypothesis.[12,30,63] According to the hypothesis, buffering of endosomal pH results in an influx of chloride ions and water molecules, which leads to endosomal swelling and eventually its burst.[100] Several experiments support the hypothesis. For example, Kichler *et al.*[101] found that suppression of proton pumps decreases the efficacy of gene delivery, and Sonawane *et al.*[100] observed an increase in chloride ions inside the endosome and endosomal swelling. Subsequently, several mechanistic studies were performed that shed more light on the endosomal escape mechanism. Vermeulen *et al.*[102] demonstrated that the minimum number of PEI required to cause endosomal burst is proportional to the endosomal volume, i.e., NPs must be prepared with a minimum N/P ratio. Consistent with this observation, Boeckle *et al.*[68] observed enhanced endosomal escape due to accumulation of free PEIs in the endosomes when they were additionally administered 1 or 4 hours after administering NPs. Comparing PEI of different MWs, Breunig *et al.*[103] found inhibition of proton pumps to impact high MW PEIs more significantly. That is, the “proton sponge” character of PEI might increase with its MW. Both linear and branched PEI have been observed to induce endosomal escape,[104] but their relative strength is not well explored in the literature. Small modifications to the “proton sponge” hypothesis have been suggested. Rehman *et al.*[43] proposed endosomes

to rupture locally instead of a global burst, and Vermeulen *et al.*[102] suggested that the endosomal membrane can become permeable to small molecules due to increased stress on the membrane.

Several studies have opposed the “proton sponge” hypothesis. Choosakoonkriang *et al.*[73] questioned the buffering capacity of PEI at endosomal pH. However, for similar data, Benjaminsen *et al.*[105] argued the buffering capacity of PEI to be sufficient. Yue *et al.*[67] observed only partial reduction in PEI’s efficacy when proton pumps were inhibited. Using this result, they claimed the “proton sponge” hypothesis to be prevalent but not dominant. Other possible mechanisms of endosomal escape include PEI-induced endosomal membrane damage.[106,107] Klemm *et al.*[106] demonstrated that PEI’s membrane damaging capacity increases with its concentration, but not the protonation state. Additionally, Bieber *et al.*[107] reported that both free PEI and PEI in the NP can cause membrane damage.

1.6. Nuclear trafficking

The final hurdle of gene delivery is for the DNAs to reach the nucleus. The genetic material in the nucleus is protected by nuclear membranes, which contain several pores called the nuclear pore complex (NPC) through which all transport of materials (such as proteins and mRNA) take place.[108] The NPC is filled with filamentous proteins called FG nucleoporin that regulate the transport of materials based on their physical size.[64,109] The mode of transport across the NPC is determined using spherical gold particles, which revealed the transport to be passive (i.e. diffusion), active (regulated by FG nucleoporin), and inactive for diameter < 9, 9-39, and > 39 nm respectively.[64,110,111] Active transport is a multi-step process, where specific short amino acid sequences called nuclear localization sequences (NLSs) selectively bind with a nuclear transport factor (NTF), which in turn binds to FG nucleoporin for nuclear entry.[109] Materials of any size can enter the nucleus during cell division when the protective nuclear membrane is absent.[60]

The exact mechanism of nuclear entry is not fully understood for PEI-based gene delivery. Some have suggested cell division to be a necessary step,[112] while others have observed nuclear entry in non-dividing cells.[22,113] For the latter case, Andersen *et al.*[113] reported NPs to exclusively bind with the NPC for their nuclear transport, although exact interactions with NLSs and NTFs were not determined.

There is also no consensus regarding the state of NPs after their endosomal escape. For example, Rehman *et al.*[43] observed oligonucleotides to dissociate from the NP inside the endosome, and Godbey *et al.*[62] found undissociated NPs inside the nucleus. Several studies have suggested the dissociation of DNAs from NPs to be essential for gene delivery and proposed the mechanism to be interexchange reactions with cytosolic molecules.[104] To this end, several studies measured the stability of NPs in the presence of anionic molecules such as heparin. Low stability of NPs was observed when prepared with PEI of low MW (stability of 2 kDa ~ 25 kDa < 750 kDa[114,115]), degree of branching[114,116], and N/P ratio.[116] While low stability of NPs is required for easy dissociation and transcription of genetic information, it should not be too low because it can lead to degradation of DNAs in the endosome.

1.7. Theoretical studies

Several all-atom (AA) and coarse-grained (CG) forcefields have been developed to study PEI. AA forcefields for linear PEIs with different protonation ratios have been developed by Ziebarth *et al.*[117], Beu *et al.*[118,119], and Wei and Luijten[120], and for PEI with different degrees of branching and protonation ratio by Sun *et al.*[121] CG forcefields developed by Wei and Luijten[120], Ziebarth *et al.*[122] and Jorge *et al.*[123] are limited to only linear PEI. These forcefields can characterize PEI's properties with its precise MW and structure using molecular dynamics (MD) or Monte Carlo (MC) simulations, while the same is not yet possible with

experimental techniques. While direct cross-validation with experiments is not yet possible in most cases because of the unknown structure and protonation ratio of synthesized PEIs (**Section 1.1**), some qualitative comparisons with experiments are available. For example, the radius of gyration is observed to increase with MW,[118,119] and protonation ratio (due to internal electrostatic repulsions),[118,119,124] and decrease with degree of branching;[44,121] diffusion coefficient increases with MW and protonation ratio[118,119]. Additionally, from a CG-MC simulation, protonation ratio of 2.2 kDa linear PEIs is reported to be ~55% at physiological pH, which compares well with 55-67% for 25 kDa branched PEI in experiments.[40]

PEI-DNA binding has been studied with AA-MD and CG-MC simulations. The primary driving forces for PEI-DNA binding are electrostatic interactions and hydrogen bonds between PEI's nitrogen and DNA's oxygen atoms.[117,121,125] For 0.6 kDa PEI and 12 bp DNA, Sun *et al.*[121] observed PEIs with different degrees of branching to bind similarly with the DNA backbone, where their interaction is weaker for the low protonation ratio of PEI. When MW of PEI is increased to 2 kDa, a distinct binding pattern is visible: linear PEIs wrap around the DNA conforming to its surface, whereas branched PEIs bind to it like beads.[44] In contrast, 0.9 kDa linear PEI binds along the DNA's major groove in Kondinskaia *et al.*[125], and along the major and minor groove in Ziebarth *et al.*[117] For sufficient N/P, PEIs binding to a DNA can neutralize its charge. The distance from the DNA's surface at which its charge is completely neutralized provides important characteristics of PEI-DNA binding strength. The neutralizing distance has been reported to be 0.8-1.3 nm for PEI with a high protonation ratio, i.e., its protonation at endolysosomal pH.[44,117,126] Specifically, the neutralization distance decreased with the increase in N/P ratio,[44] decrease in degree of branching,[44] and was not strongly affected by the MW (comparing 0.6 kDa[44] and 2 kDa[126]). More importantly, the neutralization distance

for PEI is much smaller than that observed for monovalent salts, which allows PEI to aggregate DNAs.[117]

Studies regarding PEI and DNA aggregation are not well explored in the literature. Sun *et al.*[126] studied the aggregation for a small system of PEIs and DNAs containing at most four 12 bp DNAs and twenty-eight 0.6 kDa PEIs. The aggregation was promoted by PEI, by forming polyion bridges and screening DNA's negative charge.[126] Therefore, a minimum number of PEIs, i.e., a minimum N/P ratio is required to begin the aggregation process.[126] Although the DNAs were initialized parallel to each other, some DNA pairs rearranged to form T and L shapes.[126] On a large scale, these binding modes can dictate the shape of the NP. At a high N/P ratio, NPs are positively charged and PEIs bound to the DNA are easily exchanged with free PEIs in the solution without disrupting the NP.[126] The potential of mean force (further discussed in **Chapter 2**) between two aggregating DNAs highlights a stronger role of the N/P ratio.[127] For instance, an increase in the N/P ratio makes NPs more compact and stable.[127] Computational studies exploring large-scale aggregation of PEIs and DNAs are absent in the literature.

Interactions of PEI and model cellular membranes have been studied using AA-MD in the context of PEI's antimicrobial properties,[128,129] and these results can explain their membrane destabilization properties in gene delivery. Kostritskii *et al.*[129] found linear PEI-membrane binding to be electrostatically driven, where the binding is strengthened with an increase in PEI's protonation ratio. The PEI-membrane binding causes lipid segregation, which is the first step towards a pore formation in membranes.[129] In a similar study, John *et al.*[128] explored the membrane destabilization capability of linear and branched PEI. They found branched PEIs to form pores in zwitterionic and anionic membranes at both physiological and endosomal pH; whereas linear PEIs form small pores in zwitterionic and anionic membranes at physiological pH

and only transient-unstable pores in anionic membranes at endosomal pH.[128] These pore formations by free PEIs can likely aid both cellular uptake and endosomal escape. Large-scale studies of NP-membrane interactions have been performed in the literature[130] but not for PEI-DNA NPs.

Other than the membrane interactions discussed above, studies on endosomal escape remain focused on validating the “proton sponge” hypothesis using continuum-level theories.[131,132] Yang and May[131] used self-consistent field theory and Poisson-Boltzmann equations to model PEI-DNA interactions with decreasing endosomal pH. The free PEIs increase the osmotic pressure, whereas PEIs bound to a DNA decrease it.[131] That is, for endosomal escape with the “proton sponge” hypothesis, a large concentration of free PEIs is needed. Freeman *et al.*[132] used a system of differential equations to model the transport of ions, protons, and water through the endosomal membrane and protonation of a dendritic (highly-branched) polymer. Their results are consistent with the “proton sponge” hypothesis.[132] Additionally, they found that the polymer’s ability to buffer the endosome is essential for endosomal escape.[132] The rate of protonation of the polymer and its swelling (due to protonation) did not significantly affect the osmotic pressure and thereby endosomal escape.[132]

To the best of our knowledge, no theoretical study has been performed to study the nuclear trafficking of PEI-based NPs. However, Zhou *et al.*[133] modeled the entire PEI-based gene delivery process using a kinetic model, which includes nuclear trafficking. The kinetic model was fitted to experimental data with branched PEIs of $M_n = 1.8$ and 25 kDa. [133] Analysis of rate constants revealed 25 kDa PEI to have a much higher rate of endosomal escape than 1.8 kDa, and the opposite is true for nuclear trafficking.[133] It is worth noting that kinetic models do not capture the complex nature of different steps in gene delivery.

1.8. Motivation and objective

The efficacy of PEI-based gene delivery is closely linked to its molecular properties such as MW, degree of branching, and protonation ratio, as well as PEI-DNA nanoparticle preparation procedure such as the choice of N/P ratio, ionic strength of the solution, *etc.* Furthermore, the efficacy of overcoming each intracellular barrier was dependent on the properties of PEIs and NPs. Overcoming the first barrier, cellular uptake, depends on the charge and size of NPs. For the second carrier, endosomal escape, PEI's buffering and membrane destabilization capacity are important. Finally, nuclear trafficking depends on the NP's size and stability. Determining the optimal properties of PEI through experiments is difficult because synthesized PEIs are polydisperse and have an unknown distribution of branched structures. Furthermore, various steps of gene delivery are assessed using fluorescence microscopy which can have a spatial resolution of 20-200 nm and temporal resolution of milli-seconds to seconds.[54,58,59] Therefore, any interactions or processes that occur below the resolution limit are not detected.

On the other hand, theoretical approaches are not limited by such complexities and can be used to study various steps of gene delivery with PEIs with a specific MW, degree of branching and protonation ratio, and at fine spatio-temporal resolution. AA simulations using standard computing clusters have a spatial resolution of picometers and a temporal resolution of femtoseconds but are limited to length scales of ~ 10 nm and timescales of ~ 100 ns. Therefore, they are ideal for studying PEI-DNA interactions but cannot be used to study various steps in gene delivery which occurs in length scales ~ 1 μm to time-scales of minutes.[43] Continuum-level theories can study much larger length- and time-scales, but such simulations would lack essential molecular details, such as grooves in a DNA, and forces arising from ion-ion correlations (for example in the Poisson-Boltzmann method[134]). For simulating various steps of gene delivery,

CG simulations provide a good balance between accuracy, molecular detail, and computational speed-up, and can study macroscopic length- and time-scales.[135] However, the literature lacks CG forcefields for PEI that capture various degrees of branching, protonation ratios, and MWs.

This thesis will extend the theoretical understanding of PEI-based gene delivery using CG-MD simulations. Current AA-MD simulations of PEI-DNA NPs are limited to a small number of molecules with at most four 12 bp DNAs and twenty-eight 0.6 kDa PEIs.[126] Most properties of NPs that are relevant for gene delivery can only be observed for NPs formed with a large number of molecules. CG simulations can study the formation of such large NPs and investigate how their properties change with properties of PEI and NP preparation methodology. Furthermore, the simulations enable the exploration of changes in NP properties when subject to different intracellular conditions, such as endosomal acidification. Overall, the thesis will focus on the following objectives: (1) develop a CG forcefield for PEI that captures a wide range of branched structures, MWs, and protonation ratios, i.e., create a framework to simulate various steps of gene delivery; (2) understand the formation of PEI-DNA NPs, which would internalize in endosomes; (3) understand the effects of endosomal acidification on PEI-DNA NPs; and (4) develop analysis tools to compare simulations results with experiments.

1.9. Thesis outline

Chapter 2 details the simulation methodology, which includes details of the MD method, AA and CG forcefields, radial distribution function, and potential of mean force.

Chapter 3 describes the development of CG-forcefield for PEI with different degrees of branching and protonation ratios in non-polarizable or polarizable water. The structural properties of PEI and its interaction with the existing CG-DNA forcefield are validated against AA simulations.

CG-MD simulations are performed in **Chapter 4** to study PEI-DNA NP formation at physiological pH for different N/P ratios and KCl salt concentrations. The influence of N/P ratio and salt concentration on NP properties such as size, charge, rate of aggregation, and shape is analyzed. An aggregation mechanism is proposed based on electrostatics and diffusion, and a methodology is proposed to control the size and shape of the NP.

Chapter 5 studies the role of endosomal acidification on PEI-DNA NPs and its consequences on endosomal escape and nuclear trafficking. CG-MD simulations are performed to study NP aggregation at physiological and endosomal pH, as well as acidification of NPs from physiological to endosomal pH at instant and slow ($1 \text{ H}^+/\text{ns}$) acidification rates. Acidification-driven structural changes of NPs are explained using a novel transition diagram and free energy landscape of PEIs. The simulation results are used to propose design criteria for PEI to enhance endosomal escape and nuclear trafficking.

In **Chapter 6**, a novel *in-silico* (virtual) fluorescence microscopy is described for molecular simulations that can bridge the gap between experiments and simulations. The chapter details how to use the novel technique to compare results between molecular simulations and experimental fluorescence microscopy, and how to generate new information from these comparisons.

Chapter 7 provides the overall conclusion of the thesis and potential future work.

1.10. References

- [1] Alberts B., Johnson A., Lewis J., Morgan D., Raff M., Roberts K., et al. Proteins. *In Molecular Biology of The Cell* 6th ed. *Garland Science* **2015**, 109–72.
- [2] Alberts B., Johnson A., Lewis J., Morgan D., Raff M., Roberts K., et al. DNA, Chromosome, and Genomics. *In Molecular Biology of the Cell* 6th ed. *Garland Science* **2015**, 173–236.
- [3] Gilmore I.R., Fox S.P., Hollins A.J., Sohail M., Akhtar S. The Design and Exogenous Delivery of siRNA for Post-transcriptional Gene Silencing. *J. Drug Target.* **2004**, 12, 315–

40.

- [4] Scollay R. Gene Therapy: A Brief Overview of the Past, Present, and Future. *Ann. N. Y. Acad. Sci.* **2001**, 935, 26–30.
- [5] Bulcha J.T., Wang Y., Ma H., Tai P.W.L., Gao G. Viral vector platforms within the gene therapy landscape. *Signal Transduct. Target. Ther.* **2021**, 6, 53.
- [6] Terheggen H.G., Lowenthal A., Lavinha F., Colombo J.P., Rogers S. Unsuccessful Trial of Gene Replacement in Arginase Deficiency. *Eur. J. Pediatr.* **1975**, 119, 1–3.
- [7] Ramasamy T., Munusamy S., Ruttala H.B., Kim J.O. Smart Nanocarriers for the Delivery of Nucleic Acid-Based Therapeutics: A Comprehensive Review. *Biotechnol. J.* **2021**, 16, e1900408.
- [8] Evans C.H. The Vicissitudes of Gene Therapy. *Bone Jt. Res.* **2019**, 8, 469–71.
- [9] Bulaklak K., Gersbach C.A. The Once and Future Gene Therapy. *Nat. Commun.* **2020**, 11, 5820.
- [10] Ginn S.L., Amaya A.K., Alexander I.E., Edelstein M., Abedi M.R. Gene Therapy Clinical Trials Worldwide to 2017: An Update. *J. Gene Med.* **2018**, 20, e3015.
- [11] Hardee C.L., Arévalo-Soliz L.M., Hornstein B.D., Zechiedrich L. Advances in Non-Viral DNA Vectors for Gene Therapy. *Genes (Basel)*. **2017**, 8, 65.
- [12] Boussif O., Lezoualc'h F., Zanta M.A., Mergny M.D., Scherman D., Demeneix B., et al. A Versatile Vector for Gene and Oligonucleotide Transfer into Cells in Culture and In Vivo: Polyethylenimine. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, 92, 7297–301.
- [13] Liu J., Song L., Liu S., Jiang Q., Liu Q., Li N., et al. A DNA-Based Nanocarrier for Efficient Gene Delivery and Combined Cancer Therapy. *Nano Lett.* **2018**, 18, 3328–34.
- [14] Alberts B., Johnson A., Lewis J., Morgan D., Raff M., Roberts K., et al. How Cells Read the Genome: From DNA to Protein. In *Molecular Biology of The Cell* 6th ed. *Garland Science* **2015**, 299–368.
- [15] Zhang X., Men K., Zhang Y., Zhang R., Yang L., Duan X. Local and Systemic Delivery of mRNA Encoding Survivin-T34A by Lipoplex for Efficient Colon Cancer Gene Therapy. *Int. J. Nanomedicine* **2019**, 14, 2733–51.
- [16] Ameres S.L., Martinez J., Schroeder R. Molecular Basis for Target RNA Recognition and Cleavage by Human RISC. *Cell* **2007**, 130, 101–12.
- [17] Wolff J.A., Malone R.W., Williams P., Chong W., Acsadi G., Jani A., et al. Direct Gene Transfer into Mouse Muscle In Vivo. *Science* **1990**, 247, 1465–8.

- [18] Yang N.S., Burkholder J., Roberts B., Martinell B., McCabe D. In Vivo and In Vitro Gene Transfer to Mammalian Somatic Cells by Particle Bombardment. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, 87, 9568–72.
- [19] Rols M.-P., Delteil C., Golzio M., Dumond P., Cros S., Teissie J. In Vivo Electrically Mediated Protein and Gene Transfer in Murine Melanoma. *Nat. Biotechnol.* **1998**, 16, 168–71.
- [20] Li S., Huang L. Nonviral Gene Therapy: Promises and Challenges. *Gene Ther.* **2000**, 7, 31–4.
- [21] Oh Y.-K., Suh D., Kim J.M., Choi H.-G., Shin K., Ko J.J. Polyethylenimine-Mediated Cellular Uptake, Nucleus Trafficking and Expression of Cytokine Plasmid DNA. *Gene Ther.* **2002**, 9, 1627–32.
- [22] Pollard H., Remy J.S., Loussouarn G., Demolombe S., Behr J.P., Escande D. Polyethylenimine but not Cationic Lipids Promotes Transgene Delivery to the Nucleus in Mammalian Cells. *J. Biol. Chem.* **1998**, 273, 7507–11.
- [23] Liu F., Song Y., Liu D. Hydrodynamics-Based Transfection in Animals by Systemic Administration of Plasmid DNA. *Gene Ther.* **1999**, 6, 1258–66.
- [24] Wang F., Qin Z., Lu H., He S., Luo J., Jin C., et al. Clinical Translation of Gene Medicine. *J. Gene Med.* **2019**, 21, e3108.
- [25] Shahryari A., Jazi M.S., Mohammadi S., Nikoo H.R., Nazari Z., Hosseini E.S., et al. Development and Clinical Translation of Approved Gene Therapy Products for Genetic Disorders. *Front. Genet.* **2019**, 10, 868.
- [26] Yin H., Kanasty R.L., Eltoukhy A.A., Vegas A.J., Dorkin J.R., Anderson D.G. Non-Viral Vectors for Gene-Based Therapy. *Nat. Rev. Genet.* **2014**, 15, 541–55.
- [27] Thomas C.E., Ehrhardt A., Kay M.A. Progress and Problems with the Use of Viral Vectors for Gene Therapy. *Nat. Rev. Genet.* **2003**, 4, 346–58.
- [28] Roberts D.M., Nanda A., Havenga M.J.E., Abbink P., Lynch D.M., Ewald B.A., et al. Hexon-Chimaeric Adenovirus Serotype 5 Vectors Circumvent Pre-Existing Anti-Vector Immunity. *Nature* **2006**, 441, 239–43.
- [29] Pack D.W., Hoffman A.S., Pun S., Stayton P.S. Design and Development of Polymers for Gene Delivery. *Nat. Rev. Drug Discov.* **2005**, 4, 581–93.
- [30] Behr J.-P. The Proton Sponge: A Trick to Enter Cells the Viruses did not Exploit. *Chimia (Aarau).* **1997**, 51, 34–6.

- [31] Dick C.R., Ham G.E. Characterization of Polyethylenimine. *J. Macromol. Sci. Part A-Chem.* **1970**, 4, 1301–14.
- [32] Saegusa T., Ikeda H., Fujii H. Crystalline Polyethylenimine. *Macromolecules* **1972**, 5, 108–108.
- [33] Young R.J., Lovell P.A. Concepts and Nomenclature. *In* Introduction to Polymers 3rd ed. **2011**, 3–14.
- [34] Utsuno K., Uludağ H. Thermodynamics of Polyethylenimine-DNA Binding and DNA Condensation. *Biophys. J.* **2010**, 99, 201–7.
- [35] Kircheis R., Kichler A., Wallner G., Kursu M., Ogris M., Felzmann T., et al. Coupling of Cell-Binding Ligands to Polyethylenimine for Targeted Gene Delivery. *Gene Ther.* **1997**, 4, 409–18.
- [36] Godbey W.T., Wu K.K., Mikos A.G. Size Matters: Molecular Weight Affects the Efficiency of Poly(ethylenimine) as a Gene Delivery Vehicle. *J. Biomed. Mater. Res.* **1999**, 45, 268–75.
- [37] von Harpe A., Petersen H., Li Y., Kissel T. Characterization of Commercially Available and Synthesized Polyethylenimines for Gene Delivery. *J. Control. Release* **2000**, 69, 309–22.
- [38] Mattice W.L., Porter L.J. Molecular Weight Averages and ¹³C NMR Intensities Provide Evidence for Branching in Proanthocyanidin Polymers. *Phytochemistry* **1984**, 23, 1309–11.
- [39] Suh J., Paik H.J., Hwang B.K. Ionization of Poly(ethylenimine) and Poly(allylamine) at Various pH's. *Bioorg. Chem.* **1994**, 22, 318–27.
- [40] Tang M.X., Szoka F.C. The Influence of Polymer Structure on the Interactions of Cationic Polymers with DNA and Morphology of the Resulting Complexes. *Gene Ther.* **1997**, 4, 823–32.
- [41] Kreiss P., Cameron B., Rangara R., Mailhe P., Aguerre-Charriol O., Airiau M., et al. Plasmid DNA Size Does Not Affect the Physicochemical Properties of Lipoplexes but Modulates Gene Transfer Efficiency. *Nucleic Acids Res.* **1999**, 27, 3792–8.
- [42] Dunlap D.D., Maggi A., Soria M.R., Monaco L. Nanoscopic Structure of DNA Condensed for Gene Delivery. *Nucleic Acids Res.* **1997**, 25, 3095–101.
- [43] Rehman Z.U., Hoekstra D., Zuhorn I.S. Mechanism of Polyplex- and Lipoplex-Mediated Delivery of Nucleic Acids: Real-Time Visualization of Transient Membrane Destabilization Without Endosomal Lysis. *ACS Nano* **2013**, 7, 3767–77.

- [44] Sun C., Tang T., Uludağ H. Molecular Dynamics Simulations for Complexation of DNA with 2 kDa PEI Reveal Profound Effect of PEI Architecture on Complexation. *J. Phys. Chem. B* **2012**, 116, 2405–13.
- [45] Chemin I., Moradpour D., Wieland S., Offensperger W.-B., Walter E., Behr J.-P., et al. Liver-Directed Gene Transfer: A Linear Polyethylenimine Derivative Mediates Highly Efficient DNA Delivery to Primary Hepatocytes In Vitro and In Vivo. *J. Viral Hepat.* **1998**, 5, 369–75.
- [46] Goula D., Remy J.S., Erbacher P., Wasowicz M., Levi G., Abdallah B., et al. Size, Diffusibility and Transfection Performance of Linear PEI/DNA Complexes in the Mouse Central Nervous System. *Gene Ther.* **1998**, 5, 712–7.
- [47] Gautam A., Densmore C.L., Golunski E., Xu B., Waldrep J.C. Transgene Expression in Mouse Airway Epithelium by Aerosol Gene Therapy With PEI-DNA Complexes. *Mol. Ther.* **2001**, 3, 551–6.
- [48] Webb R.H. Confocal Optical Microscopy. *Reports Prog. Phys.* **1996**, 59, 427–71.
- [49] Huisken J., Swoger J., del Bene F., Wittbrodt J., Stelzer E.H.K. Optical Sectioning Deep Inside Live Embryos by Selective Plane Illumination Microscopy. *Science* **2004**, 305, 1007–9.
- [50] Denk W., Strickler J.H., Webb W.W. Two-Photon Laser Scanning Fluorescence Microscopy. *Science* **1990**, 248, 73–6.
- [51] Gustafsson M.G.L. Surpassing the Lateral Resolution Limit by a Factor of Two Using Structured Illumination Microscopy. *J. Microsc.* **2000**, 198, 82–7.
- [52] Hell S.W., Wichmann J. Breaking the Diffraction Resolution Limit by Stimulated Emission: Stimulated-Emission-Depletion Fluorescence Microscopy. *Opt. Lett.* **1994**, 19, 780–2.
- [53] Rust M.J., Bates M., Zhuang X. Sub-Diffraction-Limit Imaging by Stochastic Optical Reconstruction Microscopy (STORM). *Nat. Methods* **2006**, 3, 793–6.
- [54] Xu J., Tehrani K.F., Kner P. Multicolor 3D Super-Resolution Imaging by Quantum Dot Stochastic Optical Reconstruction Microscopy. *ACS Nano* **2015**, 9, 2917–25.
- [55] Liu W., Toussaint Jr. K.C., Okoro C., Zhu D., Chen Y., Kuang C., et al. Breaking the Axial Diffraction Limit: A Guide to Axial Super-Resolution Fluorescence Microscopy. *Laser Photon. Rev.* **2018**, 12, 1700333.
- [56] Webb D.J., Brown C.M. Epi-Fluorescence Microscopy. *In Cell Imaging Techniques. Methods in Molecular Biology (Methods and Protocols)*, vol 931, 2nd ed. *Humana Press*

- 2012, 29–59.
- [57] Suh J., Wirtz D., Hanes J. Efficient Active Transport of Gene Nanocarriers to the Cell Nucleus. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, 100, 3878–82.
- [58] Mondal P.P. Temporal Resolution in Fluorescence Imaging. *Front. Mol. Biosci.* **2014**, 1, 11.
- [59] Abbe E. Beiträge zur Theorie des Mikroskops und der Mikroskopischen Wahrnehmung. *Arch. Für Mikroskopische Anat.* **1873**, 9, 413–68.
- [60] Khalil I.A., Kogure K., Akita H., Harashima H. Uptake Pathways and Subsequent Intracellular Trafficking in Nonviral Gene Delivery. *Pharmacol. Rev.* **2006**, 58, 32–45.
- [61] Rejman J., Bragonzi A., Conese M. Role of Clathrin- and Caveolae-Mediated Endocytosis in Gene Transfer Mediated by Lipo- and Polyplexes. *Mol. Ther.* **2005**, 12, 468–74.
- [62] Godbey W.T., Barry M.A., Saggau P., Wu K.K., Mikos A.G. Poly(Ethylenimine)-Mediated Transfection: A New Paradigm for Gene Delivery. *J. Biomed. Mater. Res.* **2000**, 51, 321–8.
- [63] Shete H.K., Prabhu R.H., Patravale V.B. Endosomal Escape: A Bottleneck in Intracellular Delivery. *J. Nanosci. Nanotechnol.* **2014**, 14, 460–74.
- [64] van der Aa M.A.E.M., Mastrobattista E., Oosting R.S., Hennink W.E., Koning G.A., Crommelin D.J.A. The Nuclear Pore Complex: The Gateway to Successful Nonviral Gene Delivery. *Pharm. Res.* **2006**, 23, 447–59.
- [65] Bahadur K.C.R., Landry B., Aliabadi H.M., Lavasanifar A., Uludağ H. Lipid Substitution on Low Molecular Weight (0.6-2.0 kDa) Polyethylenimine Leads to a Higher Zeta Potential of Plasmid DNA and Enhances Transgene Expression. *Acta Biomater.* **2011**, 7, 2209–17.
- [66] Abdallah B., Hassan A., Benoist C., Goula D., Behr J.P., Demeneix B.A. A Powerful Nonviral Vector for In Vivo Gene Transfer into the Adult Mammalian Brain: Polyethylenimine. *Hum. Gene Ther.* **1996**, 7, 1947–54.
- [67] Yue Y., Jin F., Deng R., Cai J., Dai Z., Lin M.C.M., et al. Revisit Complexation Between DNA and Polyethylenimine - Effect of Length of Free Polycationic Chains on Gene Transfection. *J. Control. Release* **2011**, 152, 143–51.
- [68] Boeckle S., von Gersdorff K., van der Pijpen S., Culmsee C., Wagner E., Ogris M. Purification of Polyethylenimine Polyplexes Highlights the Role of Free Polycations in Gene Transfer. *J. Gene Med.* **2004**, 6, 1102–11.
- [69] Fischer D., Bieber T., Li Y., Elsässer H.-P., Kissel T. A Novel Non-viral Vector for DNA

- Delivery Based on Low Molecular Weight, Branched Polyethylenimine: Effect of Molecular Weight on Transfection Efficiency and Cytotoxicity. *Pharm. Res.* **1999**, 16, 1273–9.
- [70] Werth S., Urban-Klein B., Dai L., Höbel S., Grzelinski M., Bakowsky U., et al. A Low Molecular Weight Fraction of Polyethylenimine (PEI) Displays Increased Transfection Efficiency of DNA and siRNA in Fresh or Lyophilized Complexes. *J. Control. Release* **2006**, 112, 257–70.
- [71] Bieber T., Elsässer H.P. Preparation of a Low Molecular Weight Polyethylenimine for Efficient Cell Transfection. *Biotechniques* **2001**, 30, 74–7, 80–1.
- [72] Wightman L., Kircheis R., Rössler V., Carotta S., Ruzicka R., Kursa M., et al. Different Behavior of Branched and Linear Polyethylenimine for Gene Delivery In Vitro and In Vivo. *J. Gene Med.* **2001**, 3, 362–72.
- [73] Choosakoonkriang S., Lobo B.A., Koe G.S., Koe J.G., Middaugh C.R. Biophysical Characterization of PEI/DNA Complexes. *J. Pharm. Sci.* **2003**, 92, 1710–22.
- [74] Krämer M., Stumbé J.-F., Grimm G., Kaufmann B., Krüger U., Weber M., et al. Dendritic Polyamines: Simple Access to New Materials With Defined Treelike Structures for Application in Nonviral Gene Delivery. *ChemBioChem* **2004**, 5, 1081–7.
- [75] Godbey W.T., Wu K.K., Mikos A.G. Poly(Ethyleneimine)-Mediated Gene Delivery Affects Endothelial Cell Function and Viability. *Biomaterials* **2001**, 22, 471–80.
- [76] Putnam D., Gentry C.A., Pack D.W., Langer R. Polymer-Based Gene Delivery With Low Cytotoxicity by a Unique Balance of Side-Chain Termini. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, 98, 1200–5.
- [77] Morimoto K., Nishikawa M., Kawakami S., Nakano T., Hattori Y., Fumoto S., et al. Molecular Weight-Dependent Gene Transfection Activity of Unmodified and Galactosylated Polyethyleneimine on Hepatoma Cells and Mouse Liver. *Mol. Ther.* **2003**, 7, 254–61.
- [78] Kafil V., Omid Y. Cytotoxic Impacts of Linear and Branched Polyethylenimine Nanostructures in A431 Cells. *Bioimpacts* **2011**, 1, 23–30.
- [79] Erbacher P., Bettinger T., Belguise-Valladier P., Zou S., Coll J.L., Behr J.P., et al. Transfection and Physical Properties of Various Saccharide, Poly(Ethylene Glycol), and Antibody-Derivatized Polyethylenimines (PEI). *J. Gene Med.* **1999**, 1, 210–22.
- [80] Saminathan M., Antony T., Shirahata A., Sigal L.H., Thomas T., Thomas T.J. Ionic and

- Structural Specificity Effects of Natural and Synthetic Polyamines on the Aggregation and Resolubilization of Single-, Double-, and Triple-Stranded DNA. *Biochemistry* **1999**, 38, 3821–30.
- [81] dos Santos A.L.G., Bochot A., Tsapis N., Artzner F., Bejjani R.A., Thillaye-Goldenberg B., et al. Oligonucleotide–Polyethylenimine Complexes Targeting Retinal Cells: Structural Analysis and Application to Anti-TGFbeta-2 Therapy. *Pharm. Res.* **2006**, 23, 770–81.
- [82] D’Andrea C., Pezzoli D., Malloggi C., Candeo A., Capelli G., Bassi A., et al. The Study of Polyplex Formation and Stability by Time-Resolved Fluorescence Spectroscopy of SYBR Green I-Stained DNA. *Photochem. Photobiol. Sci.* **2014**, 13, 1680–9.
- [83] Hou S., Ziebacz N., Wieczorek S.A., Kalwarczyk E., Sashuk V., Kalwarczyk T., et al. Formation and Structure of PEI/DNA Complexes: Quantitative Analysis. *Soft Matter* **2011**, 7, 6967–72.
- [84] Ko Y.T., Bickel U., Huang J. Polyethylenimine/Oligonucleotide Polyplexes Investigated by Fluorescence Resonance Energy Transfer and Fluorescence Anisotropy. *Oligonucleotides* **2011**, 21, 109–14.
- [85] Kunath K., von Harpe A., Fischer D., Petersen H., Bickel U., Voigt K., et al. Low-Molecular-Weight Polyethylenimine as a Non-Viral Vector for DNA Delivery: Comparison of Physicochemical Properties , Transfection Efficiency and In Vivo Distribution With High-Molecular-Weight Polyethylenimine. *J. Control. Release* **2003**, 89, 113–25.
- [86] Alberts B., Johnson A., Lewis J., Morgan D., Raff M., Roberts K., et al. Membrane Structure. *In Molecular Biology of The Cell* 6th ed. *Garland Science* **2015**, 565–96.
- [87] Paulos C.M., Reddy J.A., Leamon C.P., Turk M.J., Low P.S. Ligand Binding and Kinetics of Folate Receptor Recycling In Vivo: Impact on Receptor-Mediated Drug Delivery. *Mol. Pharmacol.* **2004**, 66, 1406–14.
- [88] Kopatz I., Remy J.-S., Behr J.-P. A Model for Non-Viral Gene Delivery: Through Syndecan Adhesion Molecules and Powered by Actin. *J. Gene Med.* **2004**, 6, 769–76.
- [89] Ogris M., Steinlein P., Kursa M., Mechtler K., Kircheis R., Wagner E. The Size of DNA/Transferrin-PEI Complexes Is an Important Factor for Gene Expression in Cultured Cells. *Gene Ther.* **1998**, 5, 1425–33.
- [90] Daniels T.R., Bernabeu E., Rodríguez J.A., Patel S., Kozman M., Chiappetta D.A., et al. The Transferrin Receptor and the Targeted Delivery of Therapeutic Agents Against Cancer. *Biochim. Biophys. Acta* **2012**, 1820, 291–317.

- [91] Ogris M., Brunner S., Schüller S., Kircheis R., Wagner E. PEGylated DNA/Transferrin–PEI Complexes: Reduced Interaction With Blood Components, Extended Circulation in Blood and Potential for Systemic Gene Delivery. *Gene Ther.* **1999**, 6, 595–605.
- [92] Hillaireau H., Couvreur P. Nanocarriers' Entry Into the Cell: Relevance to Drug Delivery. *Cell. Mol. Life Sci.* **2009**, 66, 2873–96.
- [93] Conner S.D., Schmid S.L. Regulated Portals of Entry Into the Cell. *Nature* **2003**, 422, 37–44.
- [94] Bareford L.M., Swaan P.W. Endocytic Mechanisms for Targeted Drug Delivery. *Adv. Drug Deliv. Rev.* **2007**, 59, 748–58.
- [95] Lisanti M.P., Scherer P.E., Vidugiriene J., Tang Z., Hermanowski-Vosatka A., Tu Y.-H., et al. Characterization of Caveolin-rich Membrane Domains Isolated from an Endothelial-rich Source: Implications for Human Disease. *J. Cell Biol.* **1994**, 126, 111–26.
- [96] Chithrani B.D., Ghazani A.A., Chan W.C.W. Determining the Size and Shape Dependence of Gold Nanoparticle Uptake into Mammalian Cells. *Nano Lett.* **2006**, 6, 662–8.
- [97] Kettler K., Veltman K., van de Meent D., van Wezel A., Hendriks A.J. Cellular Uptake of Nanoparticles as Determined by Particle Properties, Experimental Conditions, and Cell Type. *Environ. Toxicol. Chem.* **2014**, 33, 481–92.
- [98] Wang C., Wang Y., Li Y., Bodemann B., Zhao T., Ma X., et al. A Nanobuffer Reporter Library for Fine-Scale Imaging and Perturbation of Endocytic Organelles. *Nat. Commun.* **2015**, 6, 8524.
- [99] Asokan A., Cho M.J. Exploitation of Intracellular pH Gradients in the Cellular Delivery of Macromolecules. *J. Pharm. Sci.* **2002**, 91, 903–13.
- [100] Sonawane N.D., Szoka F.C., Verkman A.S. Chloride Accumulation and Swelling in Endosomes Enhances DNA Transfer by Polyamine-DNA Polyplexes. *J. Biol. Chem.* **2003**, 278, 44826–31.
- [101] Kichler A., Leborgne C., Coeytaux E., Danos O. Polyethylenimine-Mediated Gene Delivery: A Mechanistic Study. *J. Gene Med.* **2001**, 3, 135–44.
- [102] Vermeulen L.M.P., Brans T., Samal S.K., Dubruel P., Demeester J., De Smedt S.C., et al. Endosomal Size and Membrane Leakiness Influence Proton Sponge-Based Rupture of Endosomal Vesicles. *ACS Nano* **2018**, 12, 2332–45.
- [103] Breunig M., Lungwitz U., Liebl R., Fontanari C., Klar J., Kurtz A., et al. Gene Delivery With Low Molecular Weight Linear Polyethylenimines. *J. Gene Med.* **2005**, 7, 1287–98.

- [104] Itaka K., Harada A., Yamasaki Y., Nakamura K., Kawaguchi H., Kataoka K. In Situ Single Cell Observation by Fluorescence Resonance Energy Transfer Reveals Fast Intra-Cytoplasmic Delivery and Easy Release of Plasmid DNA Complexed With Linear Polyethylenimine. *J. Gene Med.* **2004**, 6, 76–84.
- [105] Benjaminsen R. V, Matthebjerg M.A., Henriksen J.R., Moghimi S.M., Andresen T.L. The Possible “ Proton Sponge ” Effect of Polyethylenimine (PEI) Does Not Include Change in Lysosomal pH. *Mol. Ther.* **2013**, 21, 149–57.
- [106] Klemm A.R., Young D., Lloyd J.B. Effects of Polyethyleneimine on Endocytosis and Lysosome Stability. *Biochem. Pharmacol.* **1998**, 56, 41–6.
- [107] Bieber T., Meissner W., Kostin S., Niemann A., Elsasser H.-P. Intracellular Route and Transcriptional Competence of Polyethylenimine-DNA Complexes. *J. Control. Release* **2002**, 82, 441–54.
- [108] Wentz S.R. Gatekeepers of the Nucleus. *Science* **2000**, 288, 1374–7.
- [109] Strambio-De-Castillia C., Niepel M., Rout M.P. The Nuclear Pore Complex: Bridging Nuclear Transport and Gene Regulation. *Nat. Rev. Mol. Cell Biol.* **2010**, 11, 490–501.
- [110] Panté N., Kann M. Nuclear Pore Complex Is Able to Transport Macromolecules with Diameters of ~39 nm. *Mol. Biol. Cell* **2002**, 13, 425–34.
- [111] Paine P.L., Moore L.C., Horowitz S.B. Nuclear Envelope Permeability. *Nature* **1975**, 254, 109–14.
- [112] Brunner S., Sauer T., Carotta S., Cotten M., Saltik M., Wagner E. Cell Cycle Dependence of Gene Transfer by Lipoplex , Polyplex and Recombinant Adenovirus. *Gene Ther.* **2000**, 7, 401–7.
- [113] Andersen H., Parhamifar L., Hunter A.C., Shahin V., Moghimi S.M. AFM Visualization of Sub-50 nm Polyplex Disposition to the Nuclear Pore Complex Without Compromising the Integrity of the Nuclear Envelope. *J. Control. Release* **2016**, 244, 24–9.
- [114] Bertschinger M., Backliwal G., Schertenleib A., Jordan M., Hacker D.L., Wurm F.M. Disassembly of Polyethylenimine-DNA Particles In Vitro: Implications for Polyethylenimine-Mediated DNA Delivery. *J. Control. Release* **2006**, 116, 96–104.
- [115] Neamark A., Suwantong O., Bahadur R.K.C., Hsu C.Y.M., Supaphol P., Uludağ H. Aliphatic Lipid Substitution on 2 kDa Polyethylenimine Improves Plasmid Delivery and Transgene Expression. *Mol. Pharm.* **2009**, 6, 1798–815.
- [116] Kwok A., Hart S.L. Comparative Structural and Functional Studies of Nanoparticle

- Formulations for DNA and siRNA Delivery. *Nanomedicine* **2011**, 7, 210–9.
- [117] Ziebarth J., Wang Y. Molecular Dynamics Simulations of DNA-Polycation Complex Formation. *Biophys. J.* **2009**, 97, 1971–83.
- [118] Beu T.A., Farcaş A. CHARMM Force Field and Molecular Dynamics Simulations of Protonated Polyethylenimine. *J. Comput. Chem.* **2017**, 38, 2335–48.
- [119] Beu T.A., Ailenei A.-E., Farcaş A. CHARMM Force Field for Protonated Polyethyleneimine. *J. Comput. Chem.* **2018**, 39, 2564–75.
- [120] Wei Z., Luijten E. Systematic Coarse-Grained Modeling of Complexation Between Small Interfering RNA and Polycations. *J. Chem. Phys.* **2015**, 143, 243146.
- [121] Sun C., Tang T., Uludağ H., Cuervo J.E. Molecular Dynamics Simulations of DNA/PEI Complexes: Effect of PEI Branching and Protonation State. *Biophys. J.* **2011**, 100, 2754–63.
- [122] Ziebarth J.D., Wang Y. Understanding the Protonation Behavior of Linear Polyethylenimine in Solutions Through Monte Carlo Simulations. *Biomacromolecules* **2010**, 11, 29–38.
- [123] Jorge A.F., Dias R.S., Pais A.A.C.C. Enhanced Condensation and Facilitated Release of DNA Using Mixed Cationic Agents: A Combined Experimental and Monte Carlo Study. *Biomacromolecules* **2012**, 13, 3151–61.
- [124] Choudhury C.K., Roy S. Structural and Dynamical Properties of Polyethylenimine in Explicit Water at Different Protonation States : A Molecular Dynamics Study. *Soft Matter* **2013**, 9, 2269–81.
- [125] Kondinskaia D.A., Kostritskii A.Y., Nesterenko A.M., Antipina A.Y., Gurtovenko A.A. Atomic-Scale Molecular Dynamics Simulations of DNA-Polycation Complexes: Two Distinct Binding Patterns. *J. Phys. Chem. B* **2016**, 120, 6546–54.
- [126] Sun C., Tang T., Uludağ H. Molecular Dynamics Simulations of PEI Mediated DNA Aggregation. *Biomacromolecules* **2011**, 12, 3698–707.
- [127] Bagai S., Sun C., Tang T. Potential of Mean Force of Polyethylenimine-Mediated DNA Attraction. *J. Phys. Chem. B* **2013**, 117, 49–56.
- [128] Fox S.J., Fazil M.H.U.T., Dhand C., Venkatesh M., Goh E.T.L., Harini S., et al. Insight Into Membrane Selectivity of Linear and Branched Polyethylenimines and Their Potential as Biocides for Advanced Wound Dressings. *Acta Biomater.* **2016**, 37, 155–64.
- [129] Kostritskii A.Y., Kondinskaia D.A., Nesterenko A.M., Gurtovenko A.A. Adsorption of

- Synthetic Cationic Polymers on Model Phospholipid Membranes: Insight from Atomic-Scale Molecular Dynamics Simulations. *Langmuir* **2016**, 32, 10402–14.
- [130] Zhang S., Gao H., Bao G. Physical Principles of Nanoparticle Cellular Endocytosis. *ACS Nano* **2015**, 9, 8655–71.
- [131] Yang S., May S. Release of Cationic Polymer-DNA Complexes From the Endosome: A Theoretical Investigation of the Proton Sponge Hypothesis. *J. Chem. Phys.* **2008**, 129, 185105.
- [132] Freeman E.C., Weiland L.M., Meng W.S. Modeling the Proton Sponge Hypothesis: Examining Proton Sponge Effectiveness for Enhancing Intracellular Gene Delivery Through Multiscale Modeling. *J. Biomater. Sci. Polym. Ed.* **2013**, 24, 398–416.
- [133] Zhou J., Yockman J.W., Kim S.W., Kern S.E. Intracellular Kinetics of Non-Viral Gene Delivery Using Polyethylenimine Carriers. *Pharm. Res.* **2007**, 24, 1079–87.
- [134] Grønbech-Jensen N., Mashl R.J., Bruinsma R.F., Gelbart W.M. Counterion-Induced Attraction between Rigid Polyelectrolytes. *Phys. Rev. Lett.* **1997**, 78, 2477–80.
- [135] Noid W.G. Perspective: Coarse-Grained Models for Biomolecular Systems. *J. Chem. Phys.* **2013**, 139, 090901.

Chapter 2

Simulation Methodology

The physical laws used to describe a system is dependent on its length- and time-scale.[1] In principle, the physical phenomena in macroscopic scales can be described using quantum or classical mechanics at the microscopic scales.[1] The connection between these scales is achieved using statistical techniques referred to as *statistical physics* or *statistical mechanics*. [2] For example, the macroscopic state can be described using thermodynamic variables such as pressure (P), absolute temperature (T), and volume (V), whereas the classical microscopic state using positions (\mathbf{r}) and momenta (\mathbf{p}) of all particles in the system.

For a given macroscopic state, such as constant P , V , and T , the system of particles can exist in many microscopic states. The collection of microscopic systems that share the same macroscopic state are referred to as an *ensemble*. [2] A macroscopic property A_{macro} can then be determined from the microscopic scales as an average over the ensemble using $A_{macro} = \int A_{micro}(\mathbf{r}, \mathbf{p}) \rho(\mathbf{r}, \mathbf{p}) d\mathbf{r} d\mathbf{p} = \langle A_{micro} \rangle$, where $\rho(\mathbf{r}, \mathbf{p})$ is the probability density, and $\langle \cdot \rangle$ denotes average. [2,3] For a state (\mathbf{r}, \mathbf{p}) with energy $\epsilon(\mathbf{r}, \mathbf{p})$, the probability density is proportional to the Boltzmann factor $e^{-\epsilon(\mathbf{r}, \mathbf{p})/k_B T}$, where k_B is the Boltzmann constant. [4] For example, $\epsilon = \mathcal{K} + \mathcal{V}$ for constant N, V, T and $\epsilon = \mathcal{K} + \mathcal{V} + PV$ for constant N, P, A , where \mathcal{K} is the kinetic energy, \mathcal{V} is the potential energy and A is enthalpy. For complex systems, the integrals must be solved numerically, which gets increasingly demanding with the number of particles. There are mainly two types of simulations that indirectly evaluates these integrals, Monte Carlo (MC) and molecular dynamics (MD).

MC simulations uses probability theories to sample states (\mathbf{r}, \mathbf{p}) from a given ensemble and A_{micro} is averaged over the sampled states to determine A_{macro} . [5] Since the Boltzmann distribution is exponential in nature, sampling of high energy states are not necessary. In MD simulations the states (\mathbf{r}, \mathbf{p}) are evolved in time using equations of motion (Section 2.1). [6] Based on the *ergodic hypothesis*, time evolution of a state (\mathbf{r}, \mathbf{p}) visits all possible states in the ensemble for a sufficiently long simulation time. [3] That is, time averages in MD simulations are approximately equal to ensemble averages. [3] However, if the simulation time is low or the system of particles are trapped in a potential well, ergodicity of MD simulations is not guaranteed. [5] In such cases, advanced sampling techniques are required for MD.

In this chapter we briefly discuss the equations of motion (Section 2.1) and forcefields used in MD simulations (Section 2.2). Additionally, theory behind radial distribution function (Section 2.3) and potential of mean force (Section 2.4) evaluated from MD simulations are discussed.

2.1. Equations of motion

Molecular dynamics is simulated using the Hamiltonian equations of motion (Eq 2.1), where r_i and p_i are the i^{th} generalized coordinate and momentum respectively, \mathbf{r} and \mathbf{p} are the corresponding vectors of all particles, \mathcal{H} is the Hamiltonian, and $(\dot{})$ represents time derivative. [7] In most MD software packages, the Cartesian coordinate system is used, i.e., r_i and p_i are xyz components of a particle's position and momentum respectively. For a system of N particles, \mathbf{r} and \mathbf{p} are vectors of size $3N$. [8]

$$\dot{r}_i = \frac{\partial \mathcal{H}(\mathbf{r}, \mathbf{p})}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial \mathcal{H}(\mathbf{r}, \mathbf{p})}{\partial r_i} \quad 2.1$$

The particles in the system are placed in a simulation box with the most common being a cuboid. To predict bulk macroscopic properties using MD simulations, a periodic boundary

condition (PBC) is used, which assumes infinite identical copies of the simulation box in some or all xyz directions.[9] Another interpretation of PBC is that when a particle crosses a periodic boundary and leaves the simulation box, it reenters the simulation box from an opposite periodic boundary with the same velocity.[9] Consequently, as the particle moves across the periodic boundaries, the angular momentum of the system is not conserved.[10] However, the angular momentum does not drift significantly because multiple particles cross the periodic boundaries in opposite directions. For calculating the potential energy, the interactions between particles are only evaluated for the *nearest image*.[11]

The equations of motion are solved numerically for known initial conditions for \mathbf{r} and \mathbf{p} . Typically, the momentum of the center of mass (COM) and angular momentum of the system about the COM is initialized to zero. In this thesis, the leap-frog algorithm is used to solve the equations of motion using Eq 2.2 and 2.3, where t is the simulation time, Δt is the integration time step (simulation time step), $\nabla_{\mathbf{r}}$ is the gradient operator with respect to \mathbf{r} , and \mathbf{M} is the mass tensor (a diagonal tensor).[6]

$$\mathbf{p}\left(t + \frac{\Delta t}{2}\right) = \mathbf{p}\left(t - \frac{\Delta t}{2}\right) - \Delta t \nabla_{\mathbf{r}} \mathcal{V} \quad 2.2$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \Delta t \mathbf{M}^{-1} \mathbf{p}\left(t + \frac{\Delta t}{2}\right) \quad 2.3$$

The Leap-frog algorithm is time-reversible, conserves energy and momentum, and requires (computational) memory to store $6N$ real numbers associated with \mathbf{r} and \mathbf{p} (storage of acceleration is not need).[6] The positions calculated in Leap-frog is identical to that of Verlet[12] and velocity Verlet algorithms[13], which are also time-reversible, and conserves energy and momentum. However, Verlet[12] and velocity Verlet algorithms[13] both require memory for $9N$ variables associated with position, momentum, and acceleration.[6] The momentum evaluated

using velocity Verlet algorithm[13] is more accurate than that evaluated from leap-frog[6] and Verlet,[12] but is computationally more expensive.[14]

MD simulations can include constraints on the positions, such as $g^k \equiv |\mathbf{r}_i(t) - \mathbf{r}_j(t)| - r^k = 0$, where k runs from 1 to N_c , N_c is the number of constraints, \mathbf{r}_i is the positions of i^{th} particle, and r^k is the constrained distance. The constraint equations can be solved using the linear constraint solver (LINCS), where constraints are applied in two steps after integrating the equations of motion.[15]

The Hamiltonian of the system depends on the ensemble simulated. In **Section 2.1.1-2.1.3**, NVE, NVT, and NPT ensembles are discussed that are relevant to biomolecular simulations. These ensembles are named after the constant macroscopic properties, which can be number of particles N , volume V , energy E , temperature T and pressure P .

2.1.1. NVE ensemble

NVE ensemble (microcanonical ensemble) models a closed system, i.e., the system enclosed in a rigid boundary that does not exchange mass or energy with its surroundings.[3] If the forces acting of each particle can be expressed as the gradient of a scalar potential energy \mathcal{V} that does not depend on \mathbf{p} , then the Hamiltonian is given by Eq 2.4.[7] Since there is no external forces acting on the system of particles, \mathcal{V} is due to interparticle interactions (further discussed in **Section 2.2**) and the total energy E is conserved. Furthermore, the kinetic energy is given by $\mathcal{K} = \mathbf{p}^T \mathbf{M}^{-1} \mathbf{p} / 2$, where \mathbf{M} is the diagonal mass tensor.

$$\mathcal{H} = \mathcal{K}(\mathbf{p}) + \mathcal{V}(\mathbf{r}) = E \tag{2.4}$$

2.1.2. NVT ensemble

NVT ensemble (canonical ensemble) models an open system with rigid boundaries that freely exchanges energy with a large heat reservoir. In thermal equilibrium, the temperatures of

the open system and the heat reservoir are equal. In statistical mechanics, the relation between momenta \mathbf{p} and the temperature T is given by the Maxwell-Boltzmann distribution in Eq 2.5, where m_i is the mass of the particle associated with r_i , and \prod denotes product of a sequence.[3] The generalized momenta are variable separable in Eq 2.5, which indicates each p_i follows a normal distribution with mean of 0 and variance of $m_i k_B T$, i.e., $\langle p_i^2 \rangle = m_i k_B T$. Therefore, the average kinetic energy of the system is $\langle \mathcal{K} \rangle = N_{dof} k_B T / 2$, where $N_{dof} = 3N - N_c$ is the number of degrees of freedom and N_c is the number of constraints.[16]

$$p(\mathbf{p}) = \left(\prod_{i=1}^{3N} \frac{1}{\sqrt{2\pi m_i k_B T}} \right) \exp \left(- \sum_i \frac{p_i^2}{2m_i k_B T} \right) \quad 2.5$$

In MD simulations, the algorithm used to moderate the temperature of a system is known as a *thermostat*. There are a number of well-known thermostats, based on different principles: velocity-rescaling,[17] Nosé-Hoover[18,19], Berendsen,[20] Andersen,[21] etc. Nosé-Hoover thermostat[18,19] modifies the Hamiltonian to explicitly sample NVT ensemble, which is computationally demanding. In Andersen thermostat,[21] particles undergo a stochastic collision based on a Poisson process and the particles' momentum after the collisions is sampled from a normal distribution. Since the sampled momentum can be opposite in direction to the particle's momentum, it can lead to unstable MD simulations. In velocity-rescaling thermostat, this instability is avoided by scaling all generalized momenta by a factor such that the probability distribution of kinetic energy given by Eq 2.6 is reproduced. Eq 2.6 is obtained from the Maxwell-Boltzmann distribution by a change of variables and $\Gamma(\cdot)$ denotes the gamma function. To ensure the kinetic energy of the system \mathcal{K} follows Eq 2.6, a random kinetic energy $\tilde{\mathcal{K}}$ is sampled from the probability distribution in Eq 2.6 and every generalized momentum p_i is scaled by a factor

$(\tilde{\mathcal{K}}/\mathcal{K})^{1/2}$. If the initial p_i 's follow a normal distribution centered at zero, scaling p_i 's keeps the distribution normal and only changes its variance which is proportional to the temperature.[17]

$$p(\mathcal{K}) = \frac{\mathcal{K}^{\frac{3N}{2}-1}}{\Gamma\left(\frac{3N}{2}\right) (k_B T)^{3N/2}} \exp\left(-\frac{\mathcal{K}}{k_B T}\right) \quad 2.6$$

Abruptly changing the momenta \mathbf{p} by a large scaling factor can lead to fast fluctuations in the particle's momentum, which is undesirable.[17] To solve this issue, \mathcal{K} is increased using a stochastic process shown in Eq 2.7, where \mathcal{W} is a Wiener noise, and τ_T is a time constant that controls the rate at which \mathcal{K} approaches $N_{dof}k_B T/2$.[17] One drawback of velocity-rescaling thermostat is that \mathbf{p} is not explicitly sampled from Maxwell-Boltzmann distribution. Berendsen thermostat,[20] follows a temperature control protocol that is similar to Eq 2.7 but does not contain the stochastic noise term. As a result, \mathcal{K} approaches $N_{dof}k_B T/2$ but does not follow the distribution shown in Eq 2.6.[20]

$$\Delta\mathcal{K} = \left(\frac{N_{dof}k_B T}{2} - \mathcal{K}\right) \frac{\Delta t}{\tau_T} + 2(\mathcal{W}(t + \Delta t) - \mathcal{W}(t)) \sqrt{\frac{k_B T \mathcal{K}}{2\tau_T}} \quad 2.7$$

2.1.3. NPT ensemble

NPT ensemble models an open system with non-rigid boundaries in contact with a large heat reservoir. The surroundings freely exchange energy with the system and performs work against the non-rigid boundary. In MD simulations, NPT ensembles are modelled using a thermostat to control the temperature and a *barostat* to control the pressure. There are several barostats in the literature such as, Parrinello-Rahman,[22] Andersen,[21] and Berendsen.[20] In Parrinello-Rahman barostat,[22] a NPA ensemble is modelled where A refers to the enthalpy. Thereafter, the temperature is corrected using velocity-rescaling thermostat to obtain the NPT

ensemble. Parrinello-Rahman barostat models NPA by changing the Hamiltonian to include the energy associated with the non-rigid boundary. The equations of motion associated with this modified Hamiltonian is given in Eq 2.8 and 2.9, where \mathbf{c} is a tensor containing edge vectors of a triclinic simulation box, \mathbf{P}_v is the virial pressure tensor given by Eq 2.10, P is the required thermodynamic pressure, $V = \det(\mathbf{c})$ is the volume of the system, and $\mathbf{\Omega}$ is a tensor that measures the inertia of the non-rigid boundary (Eq 2.11).[22] In Eq 2.11, $\boldsymbol{\beta}$ is the isothermal compressibility tensor, τ_p is the relaxation time constant for the barostat, and L is the largest element in the tensor \mathbf{c} . [14]

$$\ddot{\mathbf{r}} = -\mathbf{M}^{-1}(\nabla_{\mathbf{r}}\mathcal{V}(\mathbf{r}) + 2\mathbf{p}) \quad 2.8$$

$$\ddot{\mathbf{c}} = V\mathbf{\Omega}^{-1}\mathbf{c}^{-1}(\mathbf{P}_v - P) \quad 2.9$$

$$\mathbf{P}_v = \frac{1}{V} \left(\sum_{i=1}^N \frac{\mathbf{p}_i \mathbf{p}_i^T}{m_i} - \nabla_{\mathbf{r}_i} \mathcal{V}(\mathbf{r}_i) \mathbf{r}_i^T \right) \quad 2.10$$

$$\mathbf{\Omega}^{-1} = \frac{4\pi^2 \boldsymbol{\beta}}{2\tau_p^2 L} \quad 2.11$$

The simulations in the thesis only uses cuboidal or cubic simulation box, where \mathbf{c} , $\mathbf{\Omega}^{-1}$, and $\boldsymbol{\beta}$ are diagonal tensors.[14] Furthermore, the dimensions of the simulation box are scaled by the same value in all directions, which is achieved by replacing \mathbf{P}_v with a scalar virial pressure $P_v = Tr(\mathbf{P}_v)/3$ and using $\beta_{ii} = \beta$, where $Tr(\cdot)$ is the trace of a tensor.[14] Over the simulation time, P_v can fluctuate between positive and negative values with its average being equal to the required thermodynamic pressure P , which is always positive.[22] The main drawback for Parrinello-Rahman barostat, is that the algorithm becomes unstable if the initial conditions are far away from the NPT ensemble.[14] Therefore, other approximate barostats such as Berendsen[20]

is used to bring the MD system close to the NPT ensemble before implementing Parrinello-Rahman.[14]

Berendsen barostat scales the dimensions of a cuboidal simulation box with a constant s to slowly increase P_v to the required equilibrium pressure P using Eq 2.12.[20] Berendsen barostat is similar to velocity-rescaling thermostat but does not contain random noise, and thus P_v does not follow the correct distribution associated with NPT ensemble.[20]

$$s = 1 - \frac{\beta \Delta t}{3\tau_p} (P_v - P) \quad 2.12$$

2.2. Forcefields

A set of functional forms used to describe the interaction between particles and its associated parameters are referred to as the *forcefield*. [23] Interactions are broadly classified into bonded and non-bonded interactions.[23] Bonded interactions are defined for a set of particles that are expected to stay in close proximity due to the presence of covalent bonds. In contrast, non-bonded interactions apply to all pair of atoms, unless specified otherwise.

2.2.1. Bonded interactions

Several potentials have been defined in the literature for bonded interactions.[23] Here, the discussion is limited to one example where bond length, bond angle, and dihedral angle are each described by a single potential. The bond length potential is usually modeled using a harmonic potential as shown in Eq 2.13, where r_{ij} is the distance between particles i and j , K_{bond} is the force constant, and $r_{bond,eq}$ is equilibrium bond length.[23] For bond lengths that are constrained, only the parameter $r_{bond,eq}$ is required.[23]

$$\mathcal{V}_{bond}(r_{ij}) = \frac{1}{2} K_{bond} (r_{ij} - r_{bond,eq})^2 \quad 2.13$$

A bond angle θ_{ijk} can be defined for three particles with indices i , j , and k , where the angle is formed at the particle j . [23] The bond angle potential can be defined using a cosine-harmonic potential as shown in Eq 2.14, where K_{ang} is the force constant and $\theta_{ang,eq}$ is the equilibrium bond angle. [23]

$$\mathcal{V}_{ang}(\theta_{ijk}) = \frac{1}{2} K_{ang} (\cos \theta_{ijk} - \cos \theta_{ang,eq})^2 \quad 2.14$$

For four particles i , j , k , and l , the angle between the planes containing particles i , j , and k , and particles j , k , and l is defined as the dihedral angle ϕ_{ijkl} . The dihedral angle potential is given by the periodic potential shown in Eq 2.15, where K_{dih} is a force constant, n_{dih} is an integer, $2\pi/n_{dih}$ is the period, and $\phi_{dih,eq}$ is the phase. [23]

$$\mathcal{V}_{dih}(\phi_{ijkl}) = K_{dih} (1 + \cos(n_{dih}\phi_{ijkl} - \phi_{dih,eq})) \quad 2.15$$

2.2.2. Non-bonded Interactions

Non-bonded interactions are evaluated using a pairwise potential $\mathcal{V}(r_{ij})$, i.e., using 2-body interactions. The electrostatic interactions are modeled using the Coulombic potential given by Eq 2.16, where q_i is the charge of particle i , ϵ_0 is permittivity of vacuum, and ϵ_r is the relative permittivity of the medium.

$$\mathcal{V}_e(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \quad 2.16$$

The van der Waals interactions are usually modeled with a Lennard-Jones (LJ) potential as shown in Eq 2.17, where σ_{ij} is the interparticle distance at which potential is zero, and $-\epsilon_{ij}$ is the minimum interparticle potential energy. [24] The interparticle distance at which the force is zero is given by $2^{1/6}\sigma_{ij}$, which is referred to as the *van der Waal's diameter*. In Eq 2.17, the first term in

the bracket models short-range repulsive interaction, while the second term is the *dispersion energy* (London, Debye or Keesom energy) that models long-range attraction.[25,26]

$$\mathcal{V}_{LJ}(r_{ij}) = 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad 2.17$$

In general, the non-bonded interactions take the form $\mathcal{V}(r_{ij}) = Cr_{ij}^{-\gamma}$, where γ is an integer greater than 0. For example, γ is 2 for hydrogen bonding, 4 for interaction of a point charge and non-polar molecule (or freely rotating dipole), etc.[27]

Evaluation of non-bonded interactions is the most computationally demanding step in MD simulations for multiple reasons. First, every particle in the simulation can interact through non-bonded interactions, making the computational complexity to be of the order N^2 , N being the number of particles in the system.[28] Second, when $\gamma < 3$, the potential decays slowly and the simulation would require a large box[28] or evaluation of interactions with multiple periodic images. Finally, evaluation of potential includes several multiplication and division operations, which are significantly more expensive than addition and subtraction.[9] For these reasons non-bonded interactions are modified for computational efficiency.

2.2.3. Modifications to non-bonded interactions

A cutoff radius r_c is chosen and the interactions below and above the cutoff are referred to as *short-range* and *long-range* interactions respectively.[28] A *neighbor list* is maintained for short-range interactions, which stores a list of particle pairs that are within a radial distance of r_{nl} , and the list is updated in every n_{nl} timesteps.[28] Typically $r_{nl} \geq r_c$ because the particles outside the cut-off radius can go inside and vice-versa between neighbor list updates.[28] Interparticle distance are only necessary for particles in the neighbor list in between neighbor list updates.[28] This improves computational efficiency by decreasing the number of distance evaluations.[28]

Each particle only interacts with a small number of particles (in the neighbor list) $\ll N$ via short-range interactions, making the computational complexity of evaluating these interactions of the order N . Simply truncating the potential at r_c can make the potential and its force discontinuous and non-differentiable, which does not conserve energy in NVE ensemble.[28] Furthermore, discontinuous forces lead to inaccurate predictions of scalar virial pressure in NPT ensemble. Different strategies are used to cut off LJ and Coulombic potentials and handle their long-range interactions.

LJ potential is modified either by shifting or scaling the potential with a switching function for $r_{ij} \in (r_b, r_c)$, r_b is the smallest radial distance at which the potential is modified. The function is chosen such that the resultant potential and force are continuous and differentiable.[23,28] Long-range LJ potential is not explicitly modified, rather its effective contribution to the total energy and scalar virial pressure is evaluated.[23] Since r_{ij}^{-12} decays to zero much faster than r_{ij}^{-6} , the energy and pressure contributions are only determined for the r_{ij}^{-6} term.[23] As a result, it is referred to as the *dispersion correction*. [23] For a system of particles interacting with $V(r_{ij}) = 4\varepsilon_{ij}(\sigma_{ij}/r_{ij})^6$, dispersion correction is calculated analytically for $V(r_{ij}) = r_{ij}^{-6}$ and multiplied with the dispersion constant averaged over all particle pairs $\langle 4\varepsilon_{ij}\sigma_{ij}^6 \rangle$. [23] Since the parameters ε_{ij} and σ_{ij} does not change in time, the evaluation of dispersion correction has a constant computational complexity. This making the computational complexity of evaluating short- and long-range LJ potential to be of the order of N .

Coulombic potential decays to zero much slower than LJ potential, therefore their interaction needs special attention. Here, two modification schemes are discussed, the reaction-field[29] and particle-mesh Ewald (PME).[30] In the reaction-field approximation, each particle

observes a continuum with a fixed dielectric constant beyond the cutoff radius r_c .^[29] Then, the interparticle potential energy is given by the sum of short-range Coulombic interaction and an effective interaction with the continuum.^[29] Furthermore, the potential energy at the cutoff r_c is set to zero.^[23] This makes the computational complexity of evaluating Coulombic potential with reaction-field to be of the order N . Under PME, the Coulombic interactions are separated into short- (first term) and long-range (second term) interactions using the error function $\text{erf}(\cdot)$ as shown in Eq 2.18.^[31] A real constant a is chosen such that the short-range interaction is approximately zero at r_c .^[31] The long-range interactions are calculated in the Fourier space, where it decays quickly.^[30] To take advantage of Fast Fourier Transform (FFT) algorithm, charges are assigned to regular grids and the potential in Fourier space is interpolated using cardinal B -splines.^[30] PME is more accurate than reaction-field approach but has a computational complexity of the order $N \log(N)$.^[30]

$$v_e(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} (1 - \text{erf}(aq_{ij})) + \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \text{erf}(aq_{ij}) \quad 2.18$$

2.2.4. AA and CG forcefield

In AA MD simulations, equations of motion of atoms are solved, whereas in CG MD simulations they are solved for *beads* which represents multiple atoms.^[32] For both simulations, forcefields can be derived using top-down and/or bottom-up approaches.^[32,33] That is, AA forcefield can be derived from quantum mechanical simulations (bottom-up) and experimental data (top-down), whereas CG forcefield can be derived from AA simulations and experimental data.^[32] The functional form of AA and CG forcefield can be similar, but their parameters are significantly different. For example, mass of beads would be higher than atoms, CG length

parameters (such as $q_{bond,eq}$ and σ_{ij}) would be larger than AA, and CG force constants (such as K_{bond} , K_{ang} , and K_{dih}) would be smaller than AA.

Simulation time steps in CG simulations can be larger than AA simulations because of their smaller force constants.[32] Since multiple atoms are represented as a bead, CG simulation time is scaled up by a factor, which may not be the same for each bead.[32] Typically, the effective scaling factor for time is obtained by the ratio between CG and AA diffusion coefficients of the solvent.[34] Due to scaling of time, higher timestep, and lower number of particles in CG simulations, their computational efficiency is at least four orders higher than AA-simulations. Therefore, CG simulations can simulate much larger length- and time-scales than AA simulations.

Atom pairs that are very close to each other due to bonded interactions can make the simulations unstable due to large repulsive force from non-bonded interactions. In AA simulations, non-bonded interactions are excluded for 1-2, 1-3, and 1-4 particle interactions, where particles 1-2, 2-3, and 3-4 are connected via covalent bonds and particles 1-3 and 1-4 are not directly connected.[23] That is, the AA forcefield is derived such that the bonded interactions also account for the close range non-bonded interactions that are excluded.[23] Similarly, in CG forcefield, non-bonded interactions are typically excluded for 1-2 particle interactions.[34]

To perform AA and CG simulations, good initial conditions (\mathbf{r}, \mathbf{p}) are required, lack of which might lead to unstable simulations or unrealistic predications. In AA simulations, initial momenta are generated from Maxwell-Boltzmann distribution, and momentum of center of mass is set to zero (and also in regular intervals). Furthermore, initial positions \mathbf{r} are modified using energy minimization protocol such as steepest descent.[14] NVE and NVT simulations can be performed after the energy minimization protocol. For AA NPT simulations using Parrinello-Rahman barostat, initialization is done in three steps to bring the system close to the desired

condition. First, energy is minimized using steepest-descent, followed by generation of NVT ensemble with velocity-rescaling, and finally approximate NPT ensemble is generated using Berendsen barostat and velocity-rescaling thermostat. The second step is typically neglected in CG NPT simulations.[34] Both NVT and NPT simulations require a period of equilibration before statistics can be collected for analysis.

2.3. Radial distribution function

Among fluids, one of the most important statistical measure is the spatial distribution of atoms at equilibrium, which contains information of the local structure and its fluctuations.[35] One such measure is the *pair distribution function* (Eq 2.19), which is the probability of finding two particles at positions \mathbf{r}_1 and \mathbf{r}_2 respectively.[35] In Eq 2.19, $\delta(\cdot)$ is Dirac's delta function and C is the number of permutations of choosing particles 1 and 2. For NVT ensemble, variables \mathbf{r} and \mathbf{p} can be separated and \mathbf{p} does not contribute to the pair distribution.[35]

$$\varrho(\mathbf{r}_1, \mathbf{r}_2) = C \frac{\int \delta(\mathbf{r}'_1 - \mathbf{r}_1) \delta(\mathbf{r}'_2 - \mathbf{r}_2) e^{-\frac{\epsilon(\mathbf{r}', \mathbf{p}')}{k_B T}} d\mathbf{r}' d\mathbf{p}'}{\int e^{-\frac{\epsilon(\mathbf{r}', \mathbf{p}')}{k_B T}} d\mathbf{r}' d\mathbf{p}'} \quad 2.19$$

The pair distribution function is only dependent on the distance between the particle pair r , i.e., $\varrho(\mathbf{r}_1, \mathbf{r}_2) \equiv \varrho(r)/4\pi r^2 V$. [35] $\varrho(\mathbf{r}_1, \mathbf{r}_2)$ has dimensions of N^2/V^2 and $\varrho(r)$ has dimensions of N^2/r , where V is the volume and r is the radial distance. Using $\varrho(r)$, a *radial distribution function* (RDF) can be defined as shown in Eq 2.20, which describes the number density of the fluid at a distance of r from a particle relative to the global number density ρ . [35] For simulations, the $RDF(r)$ is calculated by dividing r into bins of equal width Δr , and determining the number density in the spherical shell of volume $4\pi r^2 \Delta r$ relative to the global number density of the system.

$$RDF(r) = \frac{\varrho(r)}{4\pi r^2 V \rho^2} \quad 2.20$$

Several important properties can be evaluated from $RDF(r)$. For example, the location of peaks in RDF between solute and solvent particles specifies the location of the solvation shells (first, second, third, etc.), and the integration of $4\pi\rho r^2 RDF(r)$ between two valleys specifies the number of particles in the solvation shell. Furthermore, the Fourier transform of $RDF(r)$ can be related to the structure factor, which can then be validated with X-ray scattering experiments.[35] For a isotropic system, the structure facto is given by Eq 2.21. [35]

$$S(\xi) = 4\pi\rho \int_0^\infty \frac{\sin \xi r}{\xi r} (RDF(r) - 1)r^2 dr \quad 2.21$$

2.4. Potential of mean force

For NVT ensemble, the potential of mean force $PMF(\mathbf{x})$ is given by Eq 2.22, where the function $h(\mathbf{r})$ transforms the positions of particles which are constrained to a reaction coordinate \mathbf{x} .[36] It is readily seen that gradient of $PMF(\mathbf{x})$ with respect to \mathbf{x} yields an ensemble average of the force $-\nabla_{\mathbf{x}}\mathcal{V}$ at given reaction coordinate \mathbf{x} .

$$PMF(\mathbf{x}) = -k_B T \ln \left[\frac{\int \delta(\mathbf{x} - h(\mathbf{r})) e^{-\frac{\mathcal{V}(\mathbf{r})}{k_B T}} d\mathbf{r}}{\int e^{-\frac{\mathcal{V}(\mathbf{r})}{k_B T}} d\mathbf{r}} \right] \quad 2.22$$

Evaluation of potential of mean force is challenging because MD simulations mostly sample the states (\mathbf{r}) associated with low $PMF(\mathbf{x})$. Advanced sampling techniques are required to accurately calculate $PMF(\mathbf{x})$, such as the umbrella sampling.[37] In umbrella sampling,[37] the reaction coordinate is separated into multiple bins and a MD simulation performed for each one.[37] At each bin, a harmonic biasing potential is added centered at its mid-point to ensure good sampling in the MD simulations.[37] Thereafter, the weighted-histogram analysis method

(WHAM) is employed to determine $PMF(\mathbf{x})$ of the unbiased system from the states sampled from the biased MD simulations.[38]

Potential of mean force can be used to validate the CG forcefield against an AA forcefield.[39] If the parameters for CG forcefield is determined accurately, the difference between AA and CG potential of mean force would be minimal.

2.5. References

- [1] Huang K. A Macroscopic View of Matter. *In* Introduction to Statistical Physics 2nd ed. *CRC Press* **2010**, 1–14.
- [2] Tuckerman M.E. Theoretical Foundations of Classical Statistical Mechanics. *In* Statistical Mechanics: Theory and Molecular Simulation. *Oxford University Press* **2010**, 53–73.
- [3] Huang K. The Statistical Approach. *In* Introduction to Statistical Physics 2nd ed. *CRC Press* **2010**, 65–82.
- [4] Landau L.D., Lifshitz E.M., Pitaevskii L.P. The Fundamental Principles of Statistical Physics. *In* Course of Theoretical Physics Vol. 5, Statistical Physics, 2nd ed. *Pergamon Press* **1969**, 1–32.
- [5] Tuckerman M.E. Monte Carlo. *In* Statistical Mechanics: Theory and Molecular Simulation. *Oxford University Press* **2010**, 277–311.
- [6] Allen M.P., Tildesley D.J. Molecular Dynamics. *In* Computer Simulation of Liquids. *Clarendon Press* **1989**, 71–109.
- [7] Goldstein H. The Hamilton Equations of Motion. *In* Classical Mechanics 2nd ed. *Addison-Wesley* **1980**, 339–77.
- [8] Lindahl E., Hess B., van der Spoel D. GROMACS 3.0 : A Package for Molecular Simulation and Trajectory Analysis. *J. Mol. Model.* **2001**, 7, 306–17.
- [9] Allen M.P., Tildesley D.J. Introduction. *In* Computer Simulation of Liquids. *Clarendon Press* **1989**, 1–32.
- [10] Hoover W.G. Historical Development and Scope. *In* Molecular Dynamics. Lecture Notes in Physics, vol 258. *Springer-Verlag* **1986**, 1–41.
- [11] Abraham M.J., van der Spoel D., Lindahl E., Hess B. Introduction. *In* GROMACS User Manual Version 5.1.4. **2016**, 1–6.

- [12] Verlet L. Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, 159, 98–103.
- [13] Swope W.C., Andersen H.C., Berens P.H., Wilson K.R. A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. *J. Chem. Phys.* **1982**, 76, 637–49.
- [14] Abraham M.J., van der Spoel D., Lindahl E., Hess B. Algorithms. *In* GROMACS User Manual Version 5.1.4. **2016**, 11–66.
- [15] Hess B., Bekker H., Berendsen H.J.C., Fraaije J.G.E.M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, 18, 1463–72.
- [16] Allen M.P., Tildesley D.J. Statistical Mechanics. *In* Computer Simulation of Liquids. *Clarendon Press* **1989**, 33–70.
- [17] Bussi G., Donadio D., Parrinello M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, 126, 014101.
- [18] Nosé S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, 52, 255–68.
- [19] Hoover W.G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, 31, 1695–7.
- [20] Berendsen H.J.C., Postma J.P.M., van Gunsteren W.F., DiNola A., Haak J.R. Molecular Dynamics With Coupling to an External Bath. *J. Chem. Phys.* **1984**, 81, 3684–90.
- [21] Andersen H.C. Molecular Dynamics Simulations at Constant Pressure and/or Temperature. *J. Chem. Phys.* **1980**, 72, 2384–93.
- [22] Parrinello M., Rahman A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, 52, 7182–90.
- [23] Abraham M.J., van der Spoel D., Lindahl E., Hess B. Interaction Functions and Forcefields. *In* GROMACS User Manual Version 5.1.4. **2016**, 67–118.
- [24] Lennard-Jones J.E. The Equation of State of Gases and Critical Phenomena. *Physica* **1937**, 4, 941–56.
- [25] Israelachvili J.N. Interactions Involving Polar Molecules. *In* Intermolecular and Surface Forces 3rd ed. *Academic Press* **2011**, 71–90.
- [26] Israelachvili J.N. Van der Waals Forces. *In* Intermolecular and Surface Forces 3rd ed. *Academic Press* **2011**, 107–32.
- [27] Israelachvili J.N. Thermodynamics and Statistical Aspects of Intermolecular Forces. *In*

- Intermolecular and Surface Forces 3rd ed. *Academic Press* **2011**, 23–52.
- [28] Allen M.P., Tildesley D.J. Some Tricks of the Trade. *In* Computer Simulation of Liquids. *Clarendon Press* **1989**, 140–81.
- [29] Barker J.A., Watts R.O. Monte Carlo Studies of the Dielectric Properties of Water-Like Models. *Mol. Phys.* **1973**, 26, 789–92.
- [30] Essmann U., Perera L., Berkowitz M.L., Darden T., Lee H., Pedersen L.G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, 103, 8577–93.
- [31] Tuckerman M.E. Evaluation of Energies and Forces. *In* Statistical Mechanics: Theory and Molecular Simulation. *Oxford University Press* **2010**, 652–62.
- [32] Noid W.G. Perspective: Coarse-Grained Models for Biomolecular Systems. *J. Chem. Phys.* **2013**, 139, 090901.
- [33] Beu T.A., Ailenei A.-E., Farcaş A. CHARMM Force Field for Protonated Polyethyleneimine. *J. Comput. Chem.* **2018**, 39, 2564–75.
- [34] Marrink S.J., de Vries A.H., Mark A.E. Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B* **2004**, 108, 750–60.
- [35] Tuckerman M.E. The Canonical Ensemble. *In* Statistical Mechanics: Theory and Molecular Simulation. *Oxford University Press* **2010**, 133–213.
- [36] Kirkwood J.G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, 3, 300–13.
- [37] Torrie G.M., Valleau J.P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, 23, 187–99.
- [38] Kumar S., Rosenberg J.M., Bouzida D., Swendsen R.H., Kollman P.A. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, 13, 1011–21.
- [39] Marrink S.J., Risselada H.J., Yefimov S., Tieleman D.P., de Vries A.H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, 111, 7812–24.

Chapter 3

Coarse-Grained Model for Polyethylenimine*

*Version of this chapter's sections has been published in: Mahajan S., and Tang T., Martini Coarse-Grained Model for Polyethylenimine. *J. Comp. Chem* **2019**, 40, 607-618, with permission from *Wiley Periodicals, Inc.* Copyright © 2018.

Mahajan S., and Tang T., Erratum: “Martini Coarse-Grained Model for Polyethylenimine” [*J. Comput. Chem.* 2019, 40, 607-618 DOI:10.1002/jcc.25747]. *J. Comp. Chem* **2020**, 41, 1730-1734, with permission from *Wiley Periodicals, Inc.* Copyright © 2020.

Mahajan S., and Tang T., Comment on “Martini force field for protonated polyethyleneimine”, *J. Comp. Chem* **2021**, 42, 261-263, with permission from *Wiley Periodicals, Inc.* Copyright © 2020.

3.1. Introduction

Polyethylenimine (PEI) is a polymer of aziridine, which can be synthesized as linear[1] or branched[2] structures with varying degrees of polymerization. Because of its polycationic nature in acidic environment, PEI has diverse applications in filtration and has been used to floc humic acid[3], remove toxic Chromium (VI)[4,5] and separate multivalent salt ions from water[6]. PEI-modified mesoporous molecular sieve MCM-41, also referred to as CO₂ “molecular basket”[7–11], has been used to adsorb CO₂ from N₂[7,12] and flue gas[8,11]. Similar PEI-modified molecular baskets have been developed for the adsorption of NO_x[8] and H₂S[13]. Removal of H₂S has further use in purifying hydrocarbons for fuel cells.[14] PEI is also a useful polyelectrolyte

in biofuel cells[15,16] and tandem solar cells[17], and has been used in methanol cells to promote electrocatalysis and stability[18]. Sensing of biomolecules, such as glucose[19] and proteins[20,21], using PEI in nanocomposites has also been demonstrated.

Among the numerous applications of PEI, its usage as a gene delivery carrier is of note. While viral carriers are very effective in cell targeting and internalization, their use is limited by their high immunogenicity.[22] PEI has been shown to be one of the most effective non-viral gene delivery carriers, owing to its large amine density and high buffering capacity.[23–26] Delivery using PEI also allows for great versatility, made possible by adjusting the PEI's molecular weight (MW)[27,28], degree of branching[29,30], method of complexation[23] with the genetic material, and modifications made to the complexes[31–33]. An excellent review on the properties of PEI and its contribution to gene delivery can be found in Lungwitz *et al.*[34]

Many experimental studies have been done to evaluate the performance of PEI in gene delivery[34,35], and to explore methods that can lead to balanced efficacy and cytotoxicity.[36,37] These experiments are supplemented by theoretical investigations that have appeared in recent years, mostly through all-atom (AA) molecular dynamics (MD) simulations. AA force field parameters have been developed for linear PEIs using general AMBER force field by Ziebarth *et al.*[38], and for PEIs with different degrees of branching using CHARMM General Force Field (CGenFF)[39] by Sun *et al.*[40] CGenFF was also used to develop a model for linear PEI grafted with polyethylene glycol by Wei *et al.*[41], by parameterizing dihedral angle parameters against ab initio calculations. Recently, Beu *et al.*[42] used CGenFF to obtain all the parameters for linear PEIs by comparing with ab initio calculations. These AA force fields have been used to study different aspects of PEI-based gene delivery. For instance, the effect of PEI structure (degrees of branching) on DNA-PEI interaction was studied for low[38,40] and high MW[43] PEIs. DNA

aggregation by PEI was studied with[44] and without[45,46] lipid modifications on the PEI. Antimicrobial properties of PEI were studied using MD simulations of PEI and lipid bilayers[47,48]. While providing molecular details on the interaction and functionality of the PEI during gene delivery, the AA MD simulations are incapable of matching the length and time scales in the real delivery process, where the complexes formed by PEI and genetic material are greater than 100 nm in size and the entire delivery can take hours[35].

On a completely different scale, Yang and May[49] used the continuum level Poisson-Boltzmann theory for linear polycations to study the proton sponge effect, a mechanism hypothesized for PEI-mediated gene delivery. The same process was investigated by Freeman *et al.*[50] using a system of kinetic equations to predict time-dependent functions such as volume and osmotic pressure of the endosome. While these simulations could match the length and time scales, they are limited by significantly simplified molecular structure and very approximate local interactions.

Coarse-grained (CG) MD simulations sets a middle ground between the costly AA simulations and the approximate continuum level theories. Coarse-graining an atomic structure involves representing a few atoms by a single bead, thereby decreasing the total number of atoms (and dimension of the phase space). The intra- and intermolecular interactions are captured by careful parameterization of the CG model, which can produce accurate results via a reliable CG force field where a lower dimensional potential energy surface (PES) describes the AA PES.[51] There are various methods to coarse-grain an AA model, one of which is the bead-spring model where the bond lengths are modeled using a harmonic spring, whereas the bond angles and torsions are not considered. As one of the simplest CG models, the bead-spring model has been used to study the protonation state of linear PEI in various pH through a Monte Carlo (MC) simulation[52],

and the interactions of DNA with lipid modified PEI using a MD simulation[53]. Jorge *et al.*[54] studied the complexation of DNA with linear PEIs using a CG-MC simulation, where the polyions were approximated using harmonic bond length and bond angle potentials. An increasingly popular CG method is the Martini scheme developed by Marrink *et al.*[55], which provides a 1000-fold speedup with accuracy comparable to AA simulations without the requirement of parametrization for every new simulation[56]. In addition, new CG models developed under the Martini scheme are compatible with previously developed Martini CG models which can be adopted directly to allow for simulations of complex multi-component systems. Using Martini scheme, Wei and Luijten[41] developed a CG model for polyethylene glycol (PEG) grafted PEI (PEG-g-PEI) with various grafting density. While to our knowledge this is the first and only Martini CG model for PEI in the literature, the focus of this work was on the complexation of PEG-g-PEI with small interfering RNA and not on extensive parameterization of the CG PEI. Because of this their CG model is only applicable to linear PEI with a single and high protonation ratio of 82%.

In this work, we aim at performing a general coarse-graining of PEI by considering a series of PEI molecules that have different degrees of branching and protonation ratios. It is known that PEIs can have vastly different branching structures and protonation states[1,29,57,58], which can lead to their different functionalities. Our extensive parameterization can therefore greatly benefit future CG simulations involving PEIs. Like in any CG modeling work, to be able to validate the CG models against the AA ones, the molecules we parameterized are relatively small (~600 Da). However, they form building blocks for larger PEIs and can be easily used and extended to different structures. The direct application of low MW PEI cannot be ruled out, for instance, ~600 Da PEI grafted with β -cyclodextrin has also been reported to be quite efficient in *in vivo* gene

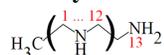
expression.[59] The compatibility of CG models within Martini will provide the opportunity for integrating the CG PEI models developed here with other well-established CG models for lipids[55,60], proteins[61,62], nucleic acids[63], polymers[64,65], solvents[60,66,67] and ions[60], to study PEI-related biomedical systems with suitable length and time scales.

3.2. Model

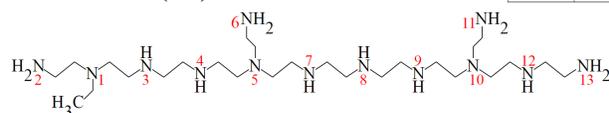
Martini[51] follows a “bottom-up” approach to model bonded interactions and a “top-down” approach for non-bonded interactions. First, the atomistic structures are coarse-grained by mapping multiple atoms into a CG bead, which typically contains 4 heavy atoms. A bead type[55,60,61,63,65,66,68] is then assigned to each CG bead, based on finding a molecular analog[60] for the AA structure within the bead and comparing free energies values from experiments and simulations. The sizes of CG beads under Martini are categorized into being “standard”, “small” and “tiny”. A “standard” bead[55] is usually created with 4:1 mapping (four heavy atoms to one CG bead), while a “small” bead[60] is typically used to model ring-like structures, with 3:1 or 2:1 mapping. Uncommonly used, “tiny” beads[63] were introduced to model base stacking in DNA, with 3:1 or 2:1 mapping. It should be noted that the resolution of mapping for each bead is not absolute; higher or lower or even fractional mapping ratios can be used.[55] The assigned bead type and size determine the parameters for non-bonded interactions. The bonded parameters are then determined by comparing the probability distributions of bond lengths, bond angles and dihedrals angles from AA and CG simulations. The parameters are tuned by trial and error till a good match is obtained between the CG and AA simulations.

All-Atom

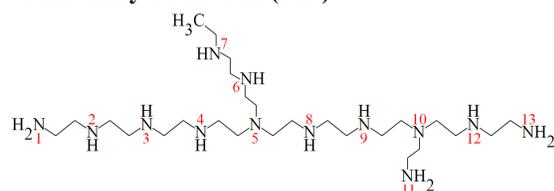
Purely-Linear (PL)



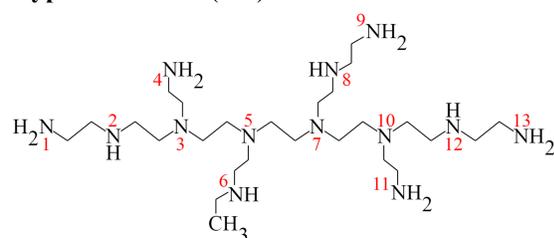
Semi-Linear (SL)



Moderately-Branched (MB)



Hyper-Branched (HB)



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

Coarse-Grained

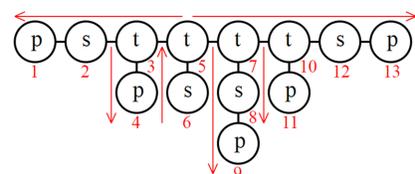
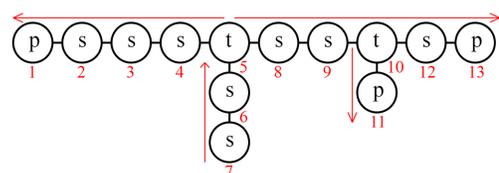
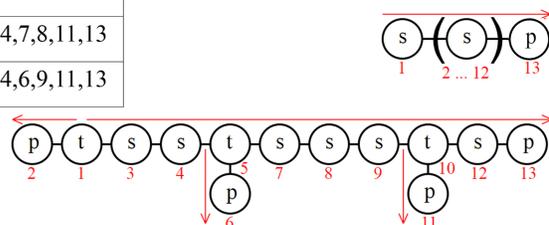


Figure 3.1: AA and CG structures of PEI. Left Panel: AA structures of the PEIs studied in this work. Right Panel: corresponding CG structures. Nitrogens on the left panel and CG beads on the right panel are numbered consistently. The inserted table shows the indices of the protonated nitrogens (AA) or beads (CG) at the two different protonation ratios (23% and 46%). Details on the labeling of the CG beads ('p', 's', etc.) are given in **Section 3.2.1**. The arrows in the right panel are used to indicate the direction from C terminal to N terminal inside the beads.

Below, we follow these basic principles to develop a CG model for four different PEI structures with two different protonation ratios (**Figure 3.1**). These four structures were named in Sun *et al.*[40] as purely linear (PL), semi-linear (SL), moderately branched (MB) and hyperbranched (HB), based on their degree of branching. In the context of gene delivery, the two protonation ratios, 23% and 46%, represent the protonation states of branched PEIs under physiological (pH = 8) and late endosomal (pH = 6) conditions, respectively.[69] The same

protonation ratios are applied to all four structures to allow for a comparative study. The nitrogen atoms which get protonated at the two protonation ratios are indicated in **Figure 3.1** using a table. The protonation sites were chosen by placing them as far apart from each other as possible.[40] Since pKa of amines follows the order of primary > secondary > tertiary, only primary and secondary amines are protonated.[40] Henceforth, 23% and 46% protonated SL PEI will be referred to as SL23 and SL46 respectively, and the other PEIs are named in the same manner.

3.2.1. CG mapping

The eight AA PEI structures are shown in **Figure 3.1** (left panel), each consist of 13 C-C-N groups. To coarse-grain these structures, the center of mass (COM) of each C-C-N group is mapped to a single CG bead (including the hydrogens). The beads that contain primary, secondary or tertiary nitrogen are labeled as ‘*p*’, ‘*s*’ and ‘*t*’ respectively. If a bead is charged, the letter ‘*q*’ is appended to its original label. For instance, a bead that contains a protonated primary nitrogen would be labeled as ‘*pq*’. This results in six different bead labels ‘*tq*’, ‘*t*’, ‘*sq*’, ‘*s*’, ‘*pq*’ and ‘*p*’. The PEI structures considered here only have uncharged tertiary nitrogen, so ‘*tq*’ beads are absent. AA structure of each bead mapped in this way has a C and an N terminal, which makes it asymmetrical. To specify the orientation of atoms inside the beads, red arrows in **Figure 3.1** (right panel) are used to indicate the direction from the C terminal to the N terminal. This information is critical in classifying bond lengths, bond angles and dihedral angles, which will be discussed in **Section 3.2.3**. The mapping scheme followed by Wei and Luijten[41] differs from ours in several aspects. Firstly, we chose to map each C-C-N group into a CG bead whereas Wei and Luijten[41] chose to map the C-N-C groups. As a result, the two beads at the end of their linear PEI were asymmetric, with one containing four heavy atoms and the other containing two heavy atoms. Using a mapping scheme that produces asymmetric end beads for a symmetric polymer (linear PEI

as considered in [41]) is undesirable.[70] Secondly, our mapping scheme is the same for all beads (C-C-N), whereas Wei and Luijten[41] used four different types of mapping, two for end beads, one for the central linear beads (C-N-C) and another for the tertiary beads.

3.2.2. Non-bonded interactions and selection of bead type

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

$$\mathcal{V}_{LJ}(r_{ij}) = 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad 3.1$$

where, r_{ij} is the distance between two beads i and j , and the LJ parameters σ_{ij} and ε_{ij} are determined by the bead type.[60] In addition, the charged beads interact with each other by the Columbic potential

$$\mathcal{V}_e(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}} \quad 3.2$$

Where q_i and q_j are the charges of beads i and j respectively, ε_0 is the permittivity of free space, and ε_r is the relative dielectric constant of the solvent. Like the exclusion of 1-4 pair interactions in AA simulations, 1-2 pair interactions are neglected in CG simulations.

The “standard” CG beads can be polar (P), non-polar (N), apolar (C), or charged (Q).[60] These bead types are further subdivided to capture different chemical natures of the beads. Polar and apolar bead types are each subdivided into five levels from 1 to 5 depending on their polarity (low to high), while non-polar and charged beads are subdivided based on their hydrogen-bonding capacity (d = donor, a = acceptor, da = both, 0 = none).[60] In this work, all the uncharged beads have the same molecular analog, ethylamine. To choose the most suitable bead type, free energy of hydration and partitioning free energy to move a solute from organic solvents (hexadecane,

chloroform and octanol) to water are evaluated for different Martini beads[60], and compared with the corresponding experimental values for ethylamine[71,72] (**Table 3.1**). Examining all the data, P₁ and P₂ are found to be the best candidates for the unchanged beads. For the charged beads, the type Q_d is chosen because of their capability to serve as hydrogen bond donor.

Table 3.1: Hydration and partitioning free energy of CG beads. Hydration free energy (ΔG^{hyd}) and partitioning free energy to move a solute from hexadecane to water (ΔG_{HW}^{part}), from chloroform to water (ΔG_{CW}^{part}), or from octanol to water (ΔG_{OW}^{part}). Comparison is made between ethylamine[71,72] (experimental value) and Martini CG beads[60,66] (tabulated data reported in the literature except for C5 and C2 in PW). The partitioning free energy is the same for polarizable water (PW) and non-polarizable water (W).[66]

beads	$\Delta G^{hyd} \left(\frac{\text{kJ}}{\text{mol}} \right)$		$\Delta G_{HW}^{part} \left(\frac{\text{kJ}}{\text{mol}} \right)$	$\Delta G_{CW}^{part} \left(\frac{\text{kJ}}{\text{mol}} \right)$	$\Delta G_{OW}^{part} \left(\frac{\text{kJ}}{\text{mol}} \right)$
	PW[66]	W[60]	W[60] considered applicable to PW[66]		
P ₅	-24.5	-25	-28	-18	-10
P ₄	-18.7	-18	-23	-14	-9
P ₃	-18.7	-18	-21	-10	-7
P ₂	-13.5	-14	-17	-2	-2
P ₁	-13.5	-14	-11	-2	-1
N _{da}	-8.3	-9	-7	0	3
N _d	-8.3	-9	-7	0	3
N _a	-8.3	-9	-7	0	3
N ₀	-3.2	-2	-2	6	5
C ₅	N/A	1	5	10	6
C ₄	4.0	5	9	13	9
C ₃	4.0	5	13	13	14
C ₂	N/A	10	16	15	16
C ₁	10.6	14	18	18	17
Ethylamine[71,72]	-19		-9	-2	-1

Different from our work, Wei and Luijten[41] used “small” beads based on the rationale that Martini mapping is usually 4:1, and the 3:1 mapping used in their work (and ours) is finer. However, the most important difference between “standard” and “small” beads is that “small” beads were introduced to model ring structures where the range and strength of the LJ interaction are smaller compared with “standard” beads.[60] This treatment allows the “small” beads to form ring structures that are absent in PEI. Such ring-forming behaviors were in fact observed in Lee *et al.*[65] for low MW PEG when “small” beads were assigned in the CG model. To avoid such unrealistic ring formation, we used the “standard” size CG beads. Furthermore, our mapping scheme allows us to use only two types of beads for all possible PEI structures, while the four different mappings in Wei and Luijten[41] require five bead types. It should also be noted that free energy comparison, such as shown in **Table 3.1**, was not presented in Wei and Luijten[41] and it is not clear why the specific bead types were chosen.

3.2.3. Bonded interactions

Bond lengths (Eq 3.3) are modeled using a harmonic potential, \mathcal{V}_{bond} where r_{ij} is the distance between two bonded beads i and j , K_{bond} is the force constant, and $r_{bond,eq}$ is the equilibrium bond length. Bond angles (Eq 3.4) are modelled using a cosine type harmonic potential \mathcal{V}_{ang} , where θ_{ijk} is the angle between consecutive beads i , j and k , K_{ang} is the force constant and $\theta_{ang,eq}$ is the equilibrium bond angle.

$$\mathcal{V}_{bond}(r_{ij}) = \frac{1}{2}K_{bond}(r_{ij} - r_{bond,eq})^2 \quad 3.3$$

$$\mathcal{V}_{ang}(\theta_{ijk}) = \frac{1}{2}K_{ang}(\cos \theta_{ijk} - \cos \theta_{ang,eq})^2 \quad 3.4$$

The dihedral angles (Eq 3.5) are modeled using a sum of periodic potentials (\mathcal{V}_{dih}), where ϕ_{ijkl} is the dihedral angle between consecutive beads i , j , k and l . $K_{dih,w}$ is the force constant for

the w^{th} periodic potential, $n_{dih,w}$ determines the number of peaks in this potential, and $\phi_{dih,eq,w}$ specifies the location of those peaks. The parameter $n_{dih,w}$ in Eq 3.5 is an integer, which depends on w as shown in Eq 3.6, where $\lceil \cdot \rceil$ is the greatest integer function.

$$\mathcal{V}_{dih}(\varphi_{ijkl}) = \sum_{w=1}^W K_{dih,w} [1 + \cos(n_{dih,w} \varphi_{ijkl} - \phi_{dih,eq,w})] \quad 3.5$$

$$n_{dih,w} = \left\lceil \frac{w+1}{2} \right\rceil \quad 3.6$$

The parameters in Eq 3.3-3.6 are determined by first making an initial guess for their values, followed by tuning of these values by trial and error. To explain the procedure, we will use the bond lengths as an example. First, AA simulations are performed for all the eight AA structures, from which we obtain the AA trajectories for the PEIs. These AA trajectories are then converted to CG trajectories, using the mapping scheme discussed in **Section 3.2.1**. From the mapped CG trajectories, the probability distributions for the bond lengths can be calculated. Recognizing that many of the bonds in the eight structures are identical in nature (explained in **Section 3.2.4**), the probability distributions of these identical bonds are averaged to obtain a single, averaged, probability distribution for each distinct bond length. These averaged probability distributions are obtained directly from the AA simulations (without running any CG simulations), and therefore are referred to as the *averaged AA distributions*. The initial guess for the CG bond length parameters in Eq 3.3 can now be obtained by applying a direct Boltzmann inversion (DBI)[73] on the averaged AA distributions. This method was used by Seo *et al.*[62] to improve dihedral angle parameters for Martini proteins, but was found to be an effective way to generate the initial guess for all the bonded parameters in this work. Specifically, from the averaged AA distribution P_{bond}^{AA} for a bond length we can calculate the potential energy U_{bond}^{AA} for this bond by $U_{bond}^{\text{AA}} =$

$-k_B T \ln(P_{bond}^{AA})$, where k_B is the Boltzmann's constant and T is the absolute temperature. U_{bond}^{AA} is then curve fitted with the bond length potential (Eq 3.3) to obtain the initial guess for the bond length parameters K_{bond} and $r_{bond,eq}$.

The initial guess for bond angle and dihedral parameters can be obtained in the same way, which provides a complete set of initial parameters that allow for CG simulations. The CG simulations will generate new probability distributions for bond lengths, angles, and dihedrals, which will be referred to as the *CG distributions*. It should be noted that CG distributions are obtained for individual structures while the averaged AA distributions are obtained by averaging distributions from the eight PEI structures. The CG simulations are run one structure after another starting from SL46 followed by MB46, HB46, PL46, SL23, MB23, HB23, and finally PL23. The CG distributions are compared with the corresponding averaged AA distributions, and depending on the difference, the CG parameters are updated until a good match is obtained between the two sets of distributions. Bonded parameters determined from previously simulated structures are used directly if the same bonded interactions appear again in a structure simulated later. For example, the parameters for bond length 't-s' are determined first from SL46 and then directly used for the same bond in MB46, HB46, SL23, MB23 and HB23. CG distributions for the 't-s' bond obtained from simulating these structures (MB46, HB46, SL23, MB23, HB23) are checked to make sure that they also match well with the averaged AA distribution (see **Appendix A** for details). As new bond lengths, angles and dihedrals appear in a new structure, their parameters are obtained as mentioned above. This is continued until all the bonded parameters for the eight PEI structures are found.

Adjusting the bonded parameters to match the averaged AA and CG distributions is an iterative process. For the bond lengths and angles, the parameters in Eqs 3.3 and 3.4 control the

location of the extrema ($r_{bond,eq}$, $\theta_{ang,eq}$) and the width of the distributions (K_{bond} , K_{ang}). They can therefore be systematically varied to achieve better matching. Such a trial-and-error method is not trivial for the dihedral angle potentials (Eq 3.5). For most dihedral angles the CG distribution matches the averaged AA distribution very well using the initial guess of parameters from DBI and eight periodic functions ($W = 8$ in Eq 3.5), with a few requiring more than eight. However, for some dihedrals good matching between CG and averaged AA distributions cannot be obtained with DBI alone. For those dihedrals, a method similar to the iterative Boltzmann inversion[74] (IBI) is used. Consider a specific dihedral for which we would like to determine the parameters in Eq 3.5. In the m^{th} iteration of IBI ($m \geq 1$), a CG simulation is performed based on the current bonded parameters, and a CG distribution ($P_{dih_s}^{\text{CG}})_m$ is obtained for this dihedral. Its corresponding potential energy, ($U_{dih_s}^{\text{CG}})_m$, is calculated using ($U_{dih_s}^{\text{CG}})_m = -k_B T \ln(P_{dih_s}^{\text{CG}})_m$. The difference between ($U_{dih_s}^{\text{CG}})_m$ and the potential energy evaluated from the averaged AA distribution, $U_{dih_s}^{\text{AA}} = -k_B T \ln P_{dih_s}^{\text{AA}}$, is computed and ($\delta U_{dih_s}^{\text{CG}})_m = (U_{dih_s}^{\text{CG}})_m - U_{dih_s}^{\text{AA}}$ is curve fitted to a function (δV_{dih_s}) $_m$ in the form of Eq 3.5. The new potential for the dihedral angle is then obtained using (V_{dih_s}) $_{m+1} = (V_{dih_s})_m + \vartheta(\delta V_{dih_s})_m$, where ϑ is a positive real number less than one.

3.2.4. Nomenclature for bonded interactions

To group similar bonds, angles and dihedrals (for calculating averaged AA and CG distributions) we systematically name them as follows. First, consider the bond lengths in the eight structures. The bond formed between beads i and j will be referred to as bond i - j , where i and j are the indices for the CG beads shown in **Figure 3.1**. Intuitively one may think that bonds i - j and j - i are identical, but this is not necessarily the case for the CG beads. For instance, bonds 5-7 and 9-10 in SL46 look similar in the CG structure, however their AA structures show a difference

because the tertiary nitrogen is near the center for bond 5-7 while it is near the edge for bond 9-10. Therefore, bonds 5-7 and 9-10 are treated differently based on the order of their beads with respect to the red arrow in **Figure 3.1**. Along the direction of the red arrow, which points from the C terminal to the N terminal, bond 5-7 is ‘*t-s*’ while bond 9-10 is ‘*s-t*’. If we look at bonds 4-5 and 5-8 in MB23 (note: carbons of bead 5 are present in the side chain), along the direction of the red arrows, they are both ‘*t-s*’ and hence identical. Many bond lengths in the eight structures can then be identified to be the same. For example, bonds 1-3, 5-7 and 10-12 in SL46 and SL23, 10-12 in MB46, 4-5, 5-8 and 10-12 in MB23, as well as 2-3, 7-8 and 10-12 in HB46 and HB23 are all ‘*t-s*’; while bonds 9-10 in SL46, 4-5 and 9-10 in SL23, 5-6 and 9-10 in MB46 and MB23, and 5-6 in HB23 are ‘*s-t*’. The bond length distributions of these identical bonds, based on CG trajectories mapped from AA simulations, are averaged to obtain the averaged AA distribution as mentioned earlier.

Similarly, when examining the bond angles, denoted as angle *i-j-k*, the direction of the red arrows (from C to N terminals) are followed. For example, bond angles 1-3-4 and 5-7-8 in SL46 both belong to ‘*t-s-sq*’, whereas bond angles 8-9-10 in SL46 and MB46, and 5-6-7 in MB46 and MB23 all belong to ‘*sq-s-t*’. Special care should be taken for beads that contain a tertiary nitrogen because each of them has three associated bond angles, two of which can be arranged from C to N terminals (referred to as *normal bond angles*), while the third can only be arranged from N to N terminals (referred to as an *N-type bond angle*). For example, in SL46 bond angle 6-5-7 is an N-type bond angle while angles 4-5-7 and 4-5-6 are normal bond angles. Since the red arrows in **Figure 3.1** point from C to N terminals, we can identify the N-type bond angles as the angles where the direction of the red arrow changes. By definition, the bead in the middle of an N-type bond angle must be either ‘*tq*’ or ‘*t*’. To denote the N-type bond angles, the letter ‘N’ is prefixed

to the label; for example, bond angle 6-5-7 in SL46 can be either ‘ $Npq-t-s$ ’ or ‘ $Ns-t-pq$ ’. N-type angles are symmetrical because they are from one N terminal to another, and the terminal beads are connected to same tertiary bead, making both ‘ $Npq-t-s$ ’ and ‘ $Ns-t-pq$ ’ equivalent.

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

The parameters for all bond lengths, bond angles and dihedral angles present in the eight PEI structures were obtained after many trial-and-error iterations and are reported in **Tables A1-A3** in **Appendix A**.

3.2.5. AA simulation methodology

AA simulations were run on the eight different PEI structures separately. Each system consisted one PEI in TIP3P water[75], with a proper number of Cl^- ions added to neutralize the system i.e., the salt concentration was zero. To study the effect of salts, additional AA simulations were run for all the PEI structures in 150 mM KCl salt solution mimicking the physiological conditions.[32] Force field parameters for the AA models were adopted from Sun *et al.*[40], which was developed based on the CGenFF[39]. The initial structure for the PEI in each AA simulation

was solvated in a rectangular box with a margin of 2 nm from all sides. Each system was first subjected to steepest descent energy minimization. This was followed by constrained NVT simulation at 300 K and then constrained NPT simulation at 300 K and 1 bar, where all the bonds in the PEI and water were constrained using the LINCS algorithm.[76] The temperature was maintained using a velocity-rescaling thermostat[77] with a time constant of 0.1 ps, and the pressure was maintained using Berendsen barostat[78] with a time constant of 2 ps and compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$. Finally, all constraints were removed except the bonds involving hydrogen atoms, and NPT simulation was run using Parrinello-Rahman barostat[79] and velocity-rescaling thermostat[77] with the same time constants and compressibility. The NPT simulation was performed for 100 ns and the last 50 ns was used for analysis. In all simulations, periodic boundary condition was applied in all directions and equations of motion were integrated using the leapfrog algorithm with a time step of 2 fs. Short-range non-bonded interactions were cut off at 1.2 nm, while long-range electrostatic interactions were handled using particle-mesh Ewald[80,81] (PME). A dispersion correction was introduced to adjust both pressure and energy to compensate for the cutoff of non-bonded interactions.

3.2.6. CG simulation methodology

CG Simulations corresponding to AA simulations were conducted with P₁-Q_d and P₂-Q_d beads in polarizable Martini water[66] (PW) and non-polarizable water[55] (W). The bonded parameters for the CG simulations were determined from P₁-Q_d PEIs in PW and 0 mM KCl. Since simulations with non-polarizable water are much faster but less accurate, this comparison allows us to assess the feasibility of employing non-polarizable water for PEI-related CG simulations. Initial structures of the PEIs in the CG simulations were obtained by mapping the equilibrated configurations at the end of the AA simulations using the mapping scheme specified in **Section**

3.2.1. All CG simulations were performed at the same temperature and pressure as in the AA simulations, following the same steps with the exclusion of a constrained NVT simulation after energy minimization. Since the low-amplitude high-frequency fluctuations are removed from the PES in the CG models[51], no additional constrained NVT simulation was necessary prior to the NPT simulation. The LJ potential was cut off at 1.2 nm by smoothly shifting it to zero between 0.9 and 1.2 nm. The electrostatic forces in the long range were handled using PME[80,81], and in the short range were cut off at 1.4 nm and smoothly shifted between 0.0 and 1.4 nm. Periodic boundary condition was applied in all the directions. The dielectric constant was taken to be 2.5 in PW[66] and 15 in W[55]. The same thermostat and barostat used in the AA simulations, with identical coupling parameters, were used except that a time constant of 0.5 ps was used in the Berendsen thermostat for both constrained and unconstrained NPT simulations in CG simulations with W. The smoother PES allowed a larger time step of 10 fs to be used in all CG simulations. Simulation parameters for systems with 150 mM KCl are identical to those used for the 0 mM KCl systems. It is worth pointing out that dynamics in CG simulations is faster compared to AA simulations due to the smoother PES. The speedup factor can be calculated by comparing the self-diffusion coefficients of individual water molecules in AA and CG simulations. Because each CG water bead in Martini consists of four water molecules, the standard time scaling factor is taken to be four.[55] Unless otherwise specified, time (except for time step and time constant) in this work for all CG simulations is the effective time, scaled according to this speedup factor. For all systems, the unconstrained NPT simulation was run for an effective time of 100 ns and the last 50 ns were used for analysis. All the MD simulations, AA and CG, were performed using Gromacs software package version 4.6.5.[76]

3.2.7. Potential of mean force calculation

Potential of mean force (PMF) was calculated between DNA and PEI using umbrella sampling, taking the distance between their centers of mass (COM) as the reaction coordinate. A Drew-Dickerson dodecamer DNA d(CGCGAATTCGCG)₂ with 24 nucleotides (12 base pairs) in its fully deprotonated state and the SL46 PEI were used for the PMF calculation. For AA simulations, CHARMM27 force field[82] was used for the DNA, parameters developed and validated by Sun *et al.*[40] based on CGenFF[39] were used for the PEI, and TIP3P[75] model was used for the water molecules. For CG simulations, CG parameters developed in this work were used for the PEI (P₁-Q_d and P₂-Q_d), and the Martini DNA[63] and polarizable water model[66] were adopted without modification.

In the AA simulations, the DNA and PEI were initially placed 1.5 nm apart (COM-COM distance) along the *z*-axis, with their principal axes aligned with the *x*-axis. The COM of the PEI was then pulled along the *z*-axis, while restraining the backbone carbon and oxygen atoms in DNA with a spring of force constant 1000 kJmol⁻¹nm⁻², to generate a set of configurations serving as initial configurations in the umbrella sample simulations. These configurations correspond to DNA-PEI COM-COM distance between 1.2 nm and 5.4 nm, with a step of 0.1 nm. Each umbrella sampling simulation was performed for 20 ns where the DNA-PEI COM-COM distance was restrained with a harmonic spring of force constant 1000 kJmol⁻¹nm⁻². The PMF was then calculated from all of the umbrella sampling simulations using the weighted histogram analysis method.[83] The CG simulations followed the same procedure except that the initial configurations were generated at every 0.05 nm, and the harmonic spring had a force constant of 2000 kJmol⁻¹nm⁻². Each umbrella sampling simulation was performed for 200 ns. All simulations were done in a

cuboidal box with initial dimensions of $6 \times 4 \times 12 \text{ nm}^3$ containing water and 150 mM KCl salt, at the temperature of 300 K and pressure of 1 bar.

All PMF were calculated using Gromacs 5.1.4,[84] and its average and standard deviation was calculated from 1000 umbrella sampling trajectories generated using Bayesian bootstrap.[85] The time window used to calculate AA- and CG-PMF are 10-20 and 50-200 ns respectively.

3.3. Results and discussions

3.3.1. Bonded parameters

For all the eight structures, there are 99 different bonded distributions and the quality of matching for a few of them is shown in **Figure 3.2**. As described earlier, the bonded parameters were first developed for P_1 - Q_d PEI structures in PW (PW- P_1) with 0 mM KCl salt. Then, other CG simulations were performed with the same bonded parameters for P_2 - Q_d PEIs, salt concentration of 150 mM KCl, and W. Therefore, each subfigure contains ten curves: two averaged AA distributions at 0 and 150 mM KCl, and two CG distributions at 0 and 150 mM KCl for PW- P_1 , PW- P_2 , W- P_1 , and W- P_2 . Systems with 0 and 150 mM salt concentrations showed negligible difference in the averaged AA distributions. This is well captured by our CG distributions, which remain unaltered irrespective of the water model, salt concentration or bead type (P_1 or P_2) used. No difference is observed for P_1 and P_2 beads because the non-bonded interactions P_1 - P_1 , P_1 - Q_d , P_1 -PW, and P_1 -W are equivalent to P_2 - P_2 , P_2 - Q_d , P_2 -PW, and P_2 -W respectively. That is, P_1 - Q_d and P_2 - Q_d models are equivalent if the system only contains PEI and water.[60]

Figure 3.2a represents the distributions of bond length ' t - s ', where the CG distributions from different simulations (in PW or W, with or without salt, P_2 or P_1) can capture the location of peak (most probable bond length) and width of the averaged AA distribution. The distributions of two bond angles ' t - s - sq ' and ' Npq - t - s ' are shown in **Figure 3.2b** and **Figure 3.2c** respectively.

The CG distributions of ‘*t-s-sq*’ overlap well with the averaged AA distribution, whereas there are noticeable differences between the CG and averaged AA distributions for bond angle ‘*Npq-t-s*’. Specifically, the location of the peak, which dictates the most probable bond angle, is slightly larger in the CG distributions (84° as compared to 81° in the averaged AA distributions). This implies that the distance between the terminal beads (‘*pq*’ and ‘*s*’) is slightly larger in the CG simulations. We calculated time average of this distance from the CG simulation to be 0.457 nm, which is smaller than the equilibrium distance (0.527 nm) in the LJ potential (Eq 3.1). Therefore, the terminal beads experience a repulsive force. As a result, further decreasing the equilibrium bond angle ($\theta_{ang,eq}$ in Eq 3.4) did not improve the quality of the matching. Additional adjustment of the force constant (K_{ang}) was performed but also found to be ineffective: lowering K_{ang} causes the most probably bond angle to further increase, while using a higher K_{ang} makes the probability distribution very sharp. Similar observations were made for most N-type bond angles. The difficulty of finding a perfect matching for an N-type bond angle may be associated with the bulky structure near the tertiary bead, which is simultaneously attached to three beads. It is challenging to accommodate the steric hindrance of these beads and the non-bonded interactions among all of them. Nevertheless, the CG distributions still match the averaged AA distributions reasonably well and similar degree of discrepancy has been deemed acceptable.[63] The dihedral angle “*Npq-t-s-sq*” in **Figure 3.2d** is modeled perfectly, and it is interesting to notice that the N-type dihedral angle does not suffer from the same discrepancy as seen in **Figure 3.2c**. This is because the dihedral angle involves four beads which experience less steric hindrance, and the distance between terminal beads is well above the equilibrium distance in the LJ potential.

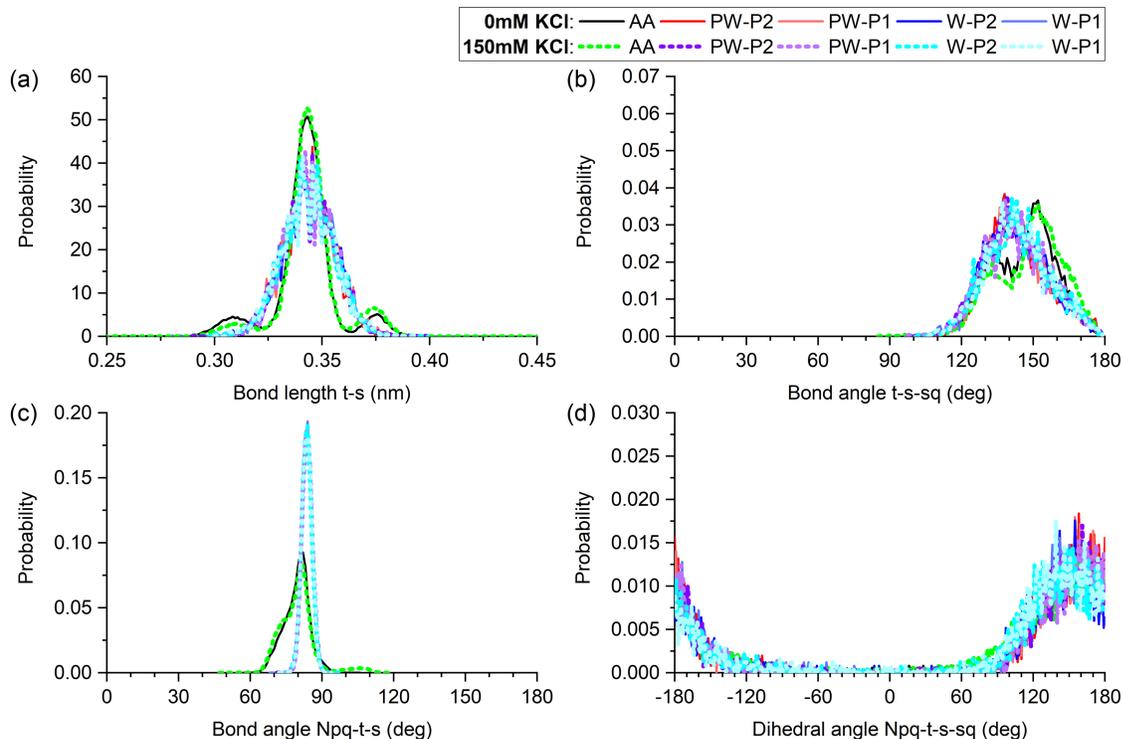


Figure 3.2: Validation of bonded parameters. Representative averaged AA distributions (AA) and CG distributions for SL46: (a) Bond length “ $t-s$ ”, (b) Bond angle “ $t-s-sq$ ”, (c) Bond angle “ $Npq-t-s$ ” (same as “ $Ns-t-pq$ ”), and (d) Dihedral angle “ $Npq-t-s-sq$ ”. Each subfigure contains ten curves that correspond to AA simulations with 0 and 150 mM KCl, CG simulations with 0 and 150 mM KCl with P_1 - Q_d PEI in PW (PW-P1), P_2 - Q_d PEI in PW (PW-P2), P_1 - Q_d PEI in W (W-P1), and P_2 - Q_d PEI in W (W-P2).

The comparisons between averaged AA and CG distributions for all bonded interactions are given in the **Appendix A**. A vast majority of the averaged AA distributions for bond lengths and bond angles contain a single peak; these distributions are excellently matched by our CG distributions. A few averaged AA distributions for bond lengths and bond angles show multiple peaks, which is not an artifact of averaging multiple bond lengths/bond angles. Wei and Luijten[41] also observed such distributions, even though they used a different mapping scheme. In those cases, our CG distribution accurately captures the average bond length/angle, and the shape of the distribution matches reasonably well. For dihedral angles, the CG distributions matched almost

perfectly with the averaged AA distributions except for a few dihedrals in the HB PEIs, primarily due to steric hindrance in those highly branched structures. Even then, the CG distributions have properly captured the location of the peaks, and only overestimated or underestimated certain peak heights. The averaged AA distributions were obtained from the last 50 ns of the 100 ns simulations. As a check of equilibrium, the averaged AA distributions were also evaluated over the periods of 50-75 ns and 75-100 ns (see **Figure A6** in **Appendix A**). The distributions over different time windows almost overlap, confirming the attainment of equilibrium.

3.3.2. Global properties

To further validate the CG model, two global properties of the polymers, namely the end-to-end distance (R_e) and radius of gyration (R_g), are calculated from the CG simulations and compared with the results from AA simulations (using the mapped trajectories). For the branched structures (SL, MB and HB), R_e was taken to be the end-to-end distance of the longest linear chain (bead 1 to bead 13 in **Figure 3.1**). Average value of R_e and R_g are represented in **Figure 3.3**, along with their standard deviations. Each subfigure contains eight sets of data, one for each PEI molecule. For each set, there are ten columns that correspond to results from AA, PW-P1, PW-P2, W-P1, and W-P2 simulations, with either 0 or 150 mM KCl.

The average and standard deviation of R_e and R_g is comparable for P₁ and P₂ because both have an identical forcefield. Considering the standard deviations, a good overlap is observed between the AA and CG results in all systems. The standard deviation tends to be higher in the PL and SL structures, which is expected since the PL and SL structures are more flexible and undergo greater fluctuations. Even though the CG parameters were obtained from the PW-P1 at 0 mM KCl, the addition of salt did not have a strong influence on the values of R_e and R_g , nor did it affect the comparison with its AA counterparts. In fact, the difference between the result in 0 and 150 mM

AA simulation is small, indicating that the screening effect of salt on such small molecules is negligible, although it is known to strongly affect the conformation of macromolecules.[86]

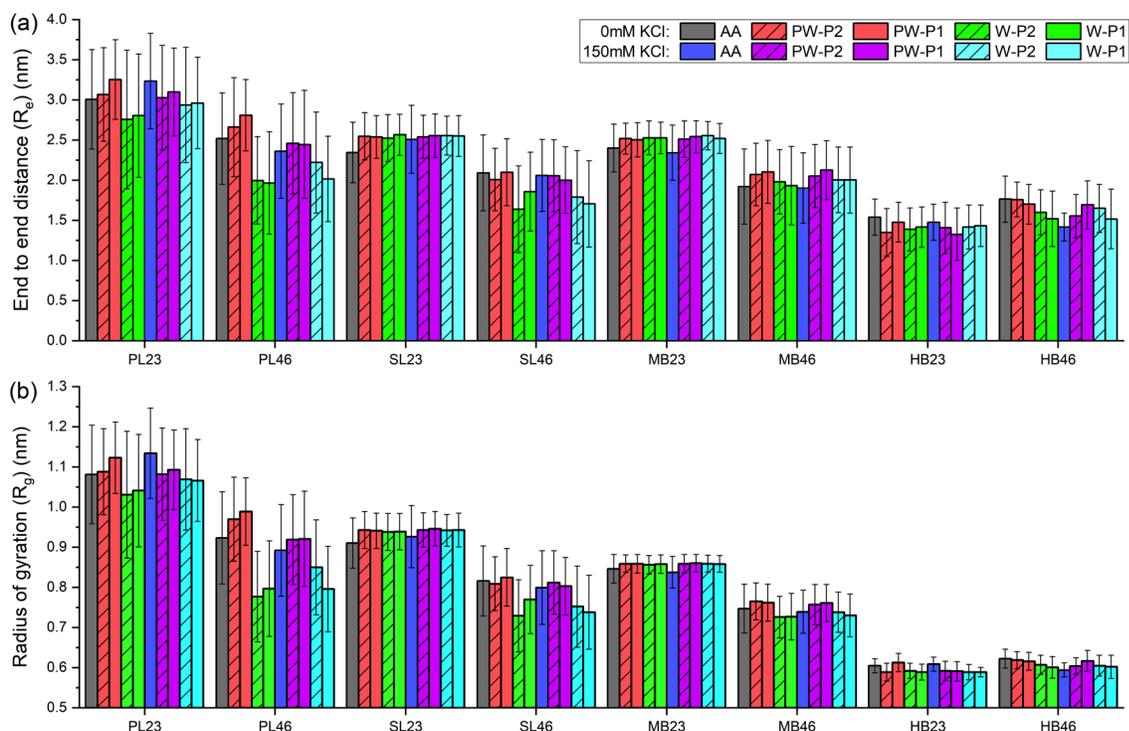


Figure 3.3: Validation of global properties. Comparison of (a) end-to-end distance, and (b) radius of gyration between AA, PW-P1, PW-P2, W-P1, and W-P2 simulations with 0 mM and 150 mM KCl. The error bars represent the standard deviations of the quantities.

At a given protonation ratio, the AA simulations predict R_e and R_g to decrease with the increase in degree of branching ($PL > SL \geq MB > HB$). Overall, the same trend is predicted by the CG simulations, except for PL46 and SL46 simulated in W. These two systems also have the largest discrepancy when compared with the corresponding AA simulations. Clearly, all systems studied in the CG-PW simulations have produced results that match better with the AA simulations. In the following section, we further discuss the quality of the water model (PW vs. W).

3.3.3. Quality of the water model

The Martini PW model includes partial charges to model the dipole moment of water[66], and is thus superior to the Martini W model in predicting ion-ion and ion-water interactions. In **Figure 3.3**, it has been observed that the W model is less accurate than the PW model, especially in linear structures with higher protonation ratios. However, there is still reasonable overlap between the results in W and those from AA simulations, and it captures the trend of R_e and R_g accurately except for PL46. In addition, the W water model is computationally more efficient than PW, as each PW has two additional particles with charges.[66] It is therefore of interest to further examine the quality of the W model. To do so, we plot two radial distribution functions (RDFs) in **Figure 3.4** for system SL46. The RDFs for the other structures share qualitative similarity and hence are not shown here.

Figure 3.4a shows the RDF of Cl^- around N^+ in both 0 mM and 150 mM KCl solutions. RDFs from the AA simulation (with and without salt) both show a prominent first peak at 0.315 nm which corresponds to the first co-ordination shell, and a secondary peak at 0.528 nm which corresponds to Cl^- and N^+ separated by a single layer of water. The system with 150 mM KCl has a higher number density of Cl^- (six times the density in 0 mM KCl) and has a lower peak height because the RDF is normalized with respect to the bulk density. In case of CG simulations, the RDF was calculated between charged PEI beads ('*sq*' or '*pq*') and CG beads representing chloride ions. The RDFs for PEIs modeled with P_1 beads are very similar to the ones modeled with P_2 beads because they have equivalent forcefield when considering only PEI and water. Compared with the AA results, the peak at 0.315 nm is absent in all CG simulations because each CG chloride ion is internally solvated with four water molecules[60]. The first peak of the RDF in PW- P_1 and PW- P_2 is at 0.492 nm and that for W- P_1 and W- P_2 is at 0.510 nm, irrespective of the salt concentration.

These values in fact match very well with the second peak in the AA RDFs. Like the AA simulations, for both water models (PW or W), the height of the peaks is lower for the system containing 150 mM KCl.

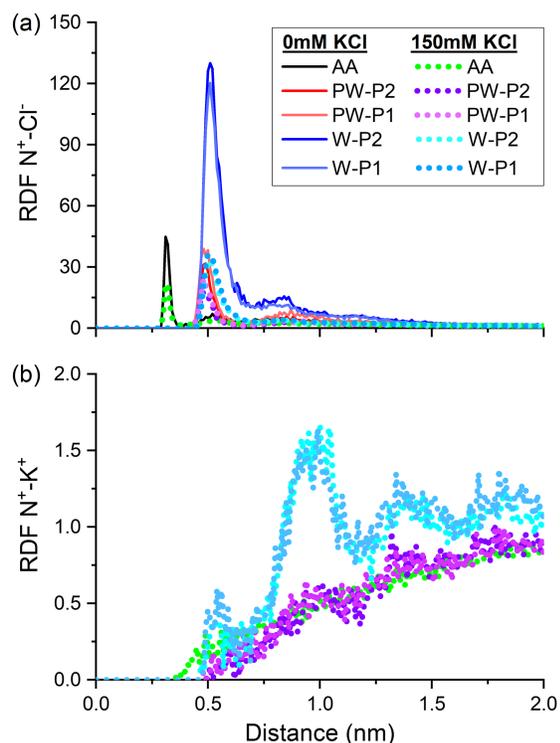


Figure 3.4: Comparison of water models. Radial distribution functions for (a) chloride ions around protonated nitrogens in 0 and 150 mM KCl; and (b) potassium ions around protonated nitrogens in 150 mM KCl. All data are for SL46 in AA, PW-P1, PW-P2, W-P1, and W-P2 simulations.

More interestingly, at a given salt concentration (0 or 150 mM), the peak height is considerably larger in CG-W, which signifies that Cl⁻ are on average closer to N⁺ in CG-W. **Figure 3.4b** represents the RDF of K⁺ around N⁺ in the simulations with 150 mM KCl. The RDF for CG-PW closely follows the AA simulation, whereas the RDF of CG-W has a significant amount of K⁺ accumulation near N⁺ between 0.8 and 1.2 nm. This can be explained by the attraction of K⁺ towards Cl⁻ which is accumulated near N⁺ (as seen in **Figure 3.4a**). Clearly, the W model overestimates ion-ion interactions, which was also reported by Wei and Luijten [41].

From the discussions above, it can be concluded that while simulating the CG PEI in W gives reasonable predictions for the global properties of the PEIs, it poorly predicts the distribution of ions around a charged polymer[66]. The over-predicted electrostatic interactions in W can in turn influence the binding of the PEIs with other charged molecules such as nucleic acids in gene delivery. Therefore, the W model should be used with caution in CG PEI simulations.

3.3.4. DNA-PEI interaction

To assess the compatibility of the developed CG PEI model with existing Martini DNA[63], PMF calculations were performed between a SL46 PEI and 12 base pair DNA at AA and CG levels (see **Section 3.2.7** for details). In **Figure 3.5**, the PMF generated from AA simulations shows an initial decrease corresponding to repulsion between the two molecules at close proximity. The PMF reaches the global minimum of -33 kJ/mol at 1.06 nm, beyond which it increases with the DNA-PEI COM-COM distance, representing an attractive force. A stationary value is reached at 3 nm where the interaction between the DNA and PEI is negligible. The PMF obtained from the CG simulations shows similar behavior: PW-P1 reaches a global minimum of -36 kJ/mol at 0.98 nm and becomes stationary after 3 nm; PW-P2 reaches a global minimum of -32 kJ/mol at 0.95 nm and becomes stationary after 3 nm. The location of the global minimum, well depth, and the shape of AA-PMF is accurately captured by both PW-P1 and PW-P2 and their difference with AA-PMF is well within the accepted range of Martini.[60,61,63,68] These results indicate that when our CG PEI models does not require additional parameterization to simulate interactions with the Martini DNA.[63] Overall, PW-P1 is better than PW-P2 at modeling AA-PMF.

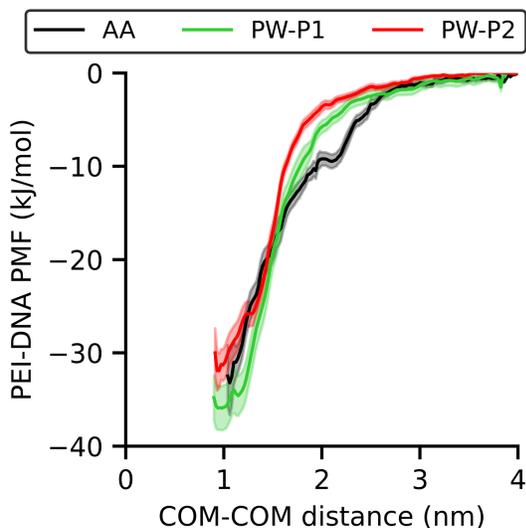


Figure 3.5: Validating compatibility of PEI and DNA forcefield. Potential of mean force between DNA and PEI using the distance between their COM as the reaction co-ordinate. Solid line represents the average over 1000 Bayesian bootstrap trajectories[85] and the corresponding shaded region represent the standard deviation.

3.3.5. Applications of the CG-PEI model

The CG model developed here is based on small PEIs with MW of ~600 Da. Since we considered 8 molecules with different structures (degrees of branching) and protonation states, the model has included most of the key building blocks present in a typical PEI, in particular protonated and non-protonated primary and secondary amines, and tertiary amines. PEIs with higher MW such as 25kDa[87] have shown to be very effective in gene delivery. Bond lengths, bond angles and dihedrals developed in this work are likely to appear again if such larger PEIs are to be investigated. The parameters determined here can therefore be directly used, to reduce the number of new bonded interactions that need to be parametrized. In fact, higher MW PEIs constructed with repeating units of PL, SL and MB PEIs studied here can be generated without introducing any additional bonded parameters. This will allow us to model large-scale polymer solutions with different polydispersity index. Our model can also be used as a starting point to

model modified PEIs such as PEG-g-PEI, Poly(2-methyl-2-oxazoline), Poly(2-ethyl-2-oxazoline), Poly(2-propyl-2-oxazoline) and Ethoxylated PEI, with additional determination of the parameters associated with the junctions and side groups.

The selection of bead types and hence non-bonded parameters were determined by comparing various free energies with experimental values, and they are directly applicable to other PEI molecules with different MW and degree of branching, in a wide range of solvents. The CG PEI model can also be used in conjunction with other Martini models, e.g. DNAs[63], proteins[61,68], lipids[60], etc., to study their interaction and usage in applications such as gene delivery. In such cases, additional potential of mean force calculations should be performed to validate the interaction between the PEI and other Martini models.

While there are CG models for branched polymers such as PAMAM[88] and branched polyethylene[89], they have been limited to very specific structures. Even for a molecule like PEI with a wide range of branching, existing CG models are limited to linear chains[41,53,54]. To our knowledge our methodology to model branched polymers is one of the first, and the same approach can be adopted to coarse-grain other branched polymers.

3.4. Conclusion

We have developed a coarse-grained model of PEI, under the Martini scheme, that is applicable to PEIs with different degrees of branching, molecular weights, and protonation ratios. Based on various solvation and partitioning free energies, two types of beads were selected to model the non-bonded interactions, and they can be used to model any PEI structures. Extensive parameterization was done for the bonded interactions, with probability distributions that compare well with all-atom simulations. Our coarse-grained model accurately captures the behaviors of PEI in water without salt, as well as in aqueous solution at physiological salt concentration relevant to

biological processes. Compared with the Martini polarizable water, the non-polarizable water model predicts poorer but still acceptable structural parameters (R_e , R_g) for the PEIs. However, it significantly overestimates the ion-ion interactions, leading to accumulation of ions around the protonated amines which is much higher than results from all-atom simulations. The model developed here can be used in conjunction with other Martini models and can be further expanded to model higher MW PEI and modified PEIs.

3.5. Acknowledgements

We acknowledge the computing resources and technical support from Western Canada Research Grid (WestGrid). Financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) and MITACS Globalink Graduate Fellowship is gratefully acknowledged.

3.6. References

- [1] Saegusa T., Ikeda H., Fujii H. Crystalline Polyethylenimine. *Macromolecules* **1972**, *5*, 108–108.
- [2] Perrine T.D., Landis W.R. Analysis of Polyethylenimine by Spectrophotometry of its Copper Chelate. *J. Polym. Sci. Part A-1 Polym. Chem.* **1967**, *5*, 1993–2003.
- [3] Glaser H.T., Edzwald J.K. Coagulation and Direct Filtration of Humic Substances with Polyethylenimine. *Environ. Sci. Technol.* **1979**, *13*, 299–305.
- [4] Korus I., Loska K. Removal of Cr(III) and Cr(VI) Ions From Aqueous Solutions by Means of Polyelectrolyte-Enhanced Ultrafiltration. *Desalination* **2009**, *247*, 390–5.
- [5] Zhao R., Li X., Sun B., Li Y., Li Y., Yang R., et al. Branched Polyethylenimine Grafted Electrospun Polyacrylonitrile Fiber Membrane: A Novel and Effective Adsorbent for Cr(VI) Remediation in Wastewater. *J. Mater. Chem. A* **2017**, *5*, 1133–44.
- [6] Lv Y., Yang H.-C., Liang H.-Q., Wan L.-S., Xu Z.-K. Nanofiltration Membranes via Co-Deposition of Polydopamine/Polyethylenimine Followed by Cross-Linking. *J. Memb. Sci.* **2015**, *476*, 50–8.

- [7] Xu X., Song C., Andrésen J.M., Miller B.G., Scaroni A.W. Preparation and Characterization of Novel CO₂ “Molecular Basket” Adsorbents Based on Polymer-Modified Mesoporous Molecular Sieve MCM-41. *Microporous Mesoporous Mater.* **2003**, 62, 29–45.
- [8] Xu X., Song C., Miller B.G., Scaroni A.W. Adsorption Separation of Carbon Dioxide From Flue Gas of Natural Gas-Fired Boiler by a Novel Nanoporous “Molecular Basket” Adsorbent. *Fuel Process. Technol.* **2005**, 86, 1457–72.
- [9] Wang X., Schwartz V., Clark J.C., Ma X., Overbury S.H., Xu X., et al. Infrared Study of CO₂ Sorption Over “Molecular Basket” Sorbent Consisting of Polyethylenimine-Modified Mesoporous Molecular Sieve. *J. Phys. Chem. C* **2009**, 113, 7260–8.
- [10] Wang D., Ma X., Sentorun-Shalaby C., Song C. Development of Carbon-Based “Molecular Basket” Sorbent for CO₂ Capture. *Ind. Eng. Chem. Res.* **2012**, 51, 3048–57.
- [11] Xu X., Song C., Miller B.G., Scaroni A.W. Influence of Moisture on CO₂ Separation from Gas Mixture by a Nanoporous Adsorbent Based on Polyethylenimine-Modified Molecular Sieve MCM-41. *Ind. Eng. Chem. Res.* **2005**, 44, 8113–9.
- [12] Xu X., Song C., Andresen J.M., Miller B.G., Scaroni A.W. Novel Polyethylenimine-Modified Mesoporous Molecular Sieve of MCM-41 Type as High-Capacity Adsorbent for CO₂ Capture. *Energy Fuels* **2002**, 16, 1463–9.
- [13] Wang X., Ma X., Xu X., Sun L., Song C. Mesoporous-Molecular-Sieve-Supported Polymer Sorbents for Removing H₂S from Hydrogen Gas Streams. *Top. Catal.* **2008**, 49, 108–17.
- [14] Xu X., Novochinskii I., Song C. Low-Temperature Removal of H₂S by Nanoporous Composite of Polymer - Mesoporous Molecular Sieve MCM-41 as Adsorbent for Fuel Cell Applications. *Energy Fuels* **2005**, 19, 2214–5.
- [15] Ivnitski D., Branch B., Atanassov P., Apblett C. Glucose Oxidase Anode for Biofuel Cell Based on Direct Electron Transfer. *Electrochem. Commun.* **2006**, 8, 1204–10.
- [16] Das M., Barbora L., Das P., Goswami P. Biofuel Cell for Generating Power From Methanol Substrate Using Alcohol Oxidase Bioanode and Air-Breathed Laccase Biocathode. *Biosens. Bioelectron.* **2014**, 59, 184–91.
- [17] Zhou Y., Fuentes-Hernandez C., Won Shim J., Khan T.M., Kippelen B. High Performance Polymeric Charge Recombination Layer for Organic Tandem Solar Cells. *Energy Environ. Sci.* **2012**, 5, 9827–32.
- [18] Cheng Y., Xu C., Shen P.K., Jiang S.P. Effect of Nitrogen-Containing Functionalization on the Electrocatalytic Activity of PtRu Nanoparticles Supported on Carbon Nanotubes for

- Direct Methanol Fuel Cells. *Appl. Catal. B Environ.* **2014**, 158, 140–9.
- [19] Shan C., Yang H., Song J., Han D., Ivaska A., Niu L. Direct Electrochemistry of Glucose Oxidase and Biosensing for Glucose Based on Graphene. *Anal. Chem.* **2009**, 81, 2378–82.
- [20] Ju Q., Tu D., Liu Y., Li R., Zhu H., Chen J., et al. Amine-Functionalized Lanthanide-Doped KGdF₄ Nanocrystals as Potential Optical/Magnetic Multimodal Bioprobes. *J. Am. Chem. Soc.* **2012**, 134, 1323–30.
- [21] Ma L., Sun N., Zhang J., Tu C., Cao X., Duan D., et al. Polyethylenimine-Coated Fe₃O₄ Nanoparticles Effectively Quench Fluorescent DNA, Which Can Be Developed as a Novel Platform for Protein Detection. *Nanoscale* **2017**, 9, 17699–703.
- [22] Pack D.W., Hoffman A.S., Pun S., Stayton P.S. Design and Development of Polymers for Gene Delivery. *Nat. Rev. Drug Discov.* **2005**, 4, 581–93.
- [23] Boussif O., Lezoualc'h F., Zanta M.A., Mergny M.D., Scherman D., Demeneix B., et al. A Versatile Vector for Gene and Oligonucleotide Transfer into Cells in Culture and In Vivo: Polyethylenimine. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, 92, 7297–301.
- [24] Shete H.K., Prabhu R.H., Patravale V.B. Endosomal Escape: A Bottleneck in Intracellular Delivery. *J. Nanosci. Nanotechnol.* **2014**, 14, 460–74.
- [25] Akinc A., Thomas M., Klibanov A.M., Langer R. Exploring Polyethylenimine-Mediated DNA Transfection and the Proton Sponge Hypothesis. *J. Gene Med.* **2005**, 7, 657–63.
- [26] Kichler A., Leborgne C., Coeytaux E., Danos O. Polyethylenimine-Mediated Gene Delivery: A Mechanistic Study. *J. Gene Med.* **2001**, 3, 135–44.
- [27] Godbey W.T., Wu K.K., Mikos A.G. Size Matters: Molecular Weight Affects the Efficiency of Poly(ethylenimine) as a Gene Delivery Vehicle. *J. Biomed. Mater. Res.* **1999**, 45, 268–75.
- [28] Kunath K., von Harpe A., Fischer D., Petersen H., Bickel U., Voigt K., et al. Low-Molecular-Weight Polyethylenimine as a Non-Viral Vector for DNA Delivery: Comparison of Physicochemical Properties, Transfection Efficiency and In Vivo Distribution With High-Molecular-Weight Polyethylenimine. *J. Control. Release* **2003**, 89, 113–25.
- [29] Tang M.X., Szoka F.C. The Influence of Polymer Structure on the Interactions of Cationic Polymers with DNA and Morphology of the Resulting Complexes. *Gene Ther.* **1997**, 4, 823–32.
- [30] Wightman L., Kircheis R., Rössler V., Carotta S., Ruzicka R., Kurska M., et al. Different Behavior of Branched and Linear Polyethylenimine for Gene Delivery In Vitro and In Vivo.

- J. Gene Med.* **2001**, 3, 362–72.
- [31] Kircheis R., Wightman L., Wagner E. Design and Gene Delivery Activity of Modified Polyethylenimines. *Adv. Drug Deliv. Rev.* **2001**, 53, 341–58.
- [32] Mishra S., Webster P., Davis M.E. PEGylation Significantly Affects Cellular Uptake and Intracellular Trafficking of Non-Viral Gene Delivery Particles. *Eur. J. Cell Biol.* **2004**, 83, 97–111.
- [33] Ogris M., Brunner S., Schüller S., Kircheis R., Wagner E. PEGylated DNA/Transferrin–PEI Complexes: Reduced Interaction With Blood Components, Extended Circulation in Blood and Potential for Systemic Gene Delivery. *Gene Ther.* **1999**, 6, 595–605.
- [34] Lungwitz U., Breunig M., Blunk T., Göpferich A. Polyethylenimine-Based Non-Viral Gene Delivery Systems. *Eur. J. Pharm. Biopharm.* **2005**, 60, 247–66.
- [35] Rehman Z.U., Hoekstra D., Zuhorn I.S. Mechanism of Polyplex- and Lipoplex-Mediated Delivery of Nucleic Acids: Real-Time Visualization of Transient Membrane Destabilization Without Endosomal Lysis. *ACS Nano* **2013**, 7, 3767–77.
- [36] Morimoto K., Nishikawa M., Kawakami S., Nakano T., Hattori Y., Fumoto S., et al. Molecular Weight-Dependent Gene Transfection Activity of Unmodified and Galactosylated Polyethyleneimine on Hepatoma Cells and Mouse Liver. *Mol. Ther.* **2003**, 7, 254–61.
- [37] Kafil V., Omidi Y. Cytotoxic Impacts of Linear and Branched Polyethylenimine Nanostructures in A431 Cells. *Bioimpacts* **2011**, 1, 23–30.
- [38] Ziebarth J., Wang Y. Molecular Dynamics Simulations of DNA-Polycation Complex Formation. *Biophys. J.* **2009**, 97, 1971–83.
- [39] Vanommeslaeghe K., Hatcher E., Acharya C., Kundu S., Zhong S., Shim J., et al. CHARMM General Force Field (CGenFF): A Force Field for Drug-Like Molecules Compatible With the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, 31, 671–90.
- [40] Sun C., Tang T., Uludağ H., Cuervo J.E. Molecular Dynamics Simulations of DNA/PEI Complexes: Effect of PEI Branching and Protonation State. *Biophys. J.* **2011**, 100, 2754–63.
- [41] Wei Z., Luijten E. Systematic Coarse-Grained Modeling of Complexation Between Small Interfering RNA and Polycations. *J. Chem. Phys.* **2015**, 143, 243146.
- [42] Beu T.A., Farcaş A. CHARMM Force Field and Molecular Dynamics Simulations of

- Protonated Polyethylenimine. *J. Comput. Chem.* **2017**, 38, 2335–48.
- [43] Sun C., Tang T., Uludağ H. Molecular Dynamics Simulations for Complexation of DNA with 2 kDa PEI Reveal Profound Effect of PEI Architecture on Complexation. *J. Phys. Chem. B* **2012**, 116, 2405–13.
- [44] Bagai S., Sun C., Tang T. Lipid-Modified Polyethylenimine-Mediated DNA Attraction Evaluated by Molecular Dynamics Simulations. *J. Phys. Chem. B* **2014**, 118, 7070–6.
- [45] Sun C., Tang T., Uludağ H. Molecular Dynamics Simulations of PEI Mediated DNA Aggregation. *Biomacromolecules* **2011**, 12, 3698–707.
- [46] Bagai S., Sun C., Tang T. Potential of Mean Force of Polyethylenimine-Mediated DNA Attraction. *J. Phys. Chem. B* **2013**, 117, 49–56.
- [47] Kostritskii A.Y., Kondinskaia D.A., Nesterenko A.M., Gurtovenko A.A. Adsorption of Synthetic Cationic Polymers on Model Phospholipid Membranes: Insight from Atomic-Scale Molecular Dynamics Simulations. *Langmuir* **2016**, 32, 10402–14.
- [48] Fox S.J., Fazil M.H.U.T., Dhand C., Venkatesh M., Goh E.T.L., Harini S., et al. Insight Into Membrane Selectivity of Linear and Branched Polyethylenimines and Their Potential as Biocides for Advanced Wound Dressings. *Acta Biomater.* **2016**, 37, 155–64.
- [49] Yang S., May S. Release of Cationic Polymer-DNA Complexes From the Endosome: A Theoretical Investigation of the Proton Sponge Hypothesis. *J. Chem. Phys.* **2008**, 129, 185105.
- [50] Freeman E.C., Weiland L.M., Meng W.S. Modeling the Proton Sponge Hypothesis: Examining Proton Sponge Effectiveness for Enhancing Intracellular Gene Delivery Through Multiscale Modeling. *J. Biomater. Sci. Polym. Ed.* **2013**, 24, 398–416.
- [51] Noid W.G. Perspective: Coarse-Grained Models for Biomolecular Systems. *J. Chem. Phys.* **2013**, 139, 090901.
- [52] Ziebarth J.D., Wang Y. Understanding the Protonation Behavior of Linear Polyethylenimine in Solutions Through Monte Carlo Simulations. *Biomacromolecules* **2010**, 11, 29–38.
- [53] Ziebarth J., Wang Y. Coarse-Grained Molecular Dynamics Simulations of DNA Condensation by Block Copolymer and Formation of Core - Corona Structures. *J. Phys. Chem. B* **2010**, 114, 6225–32.
- [54] Jorge A.F., Dias R.S., Pais A.A.C.C. Enhanced Condensation and Facilitated Release of DNA Using Mixed Cationic Agents: A Combined Experimental and Monte Carlo Study.

- Biomacromolecules* **2012**, 13, 3151–61.
- [55] Marrink S.J., de Vries A.H., Mark A.E. Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B* **2004**, 108, 750–60.
- [56] Marrink S.J., Tieleman D.P. Perspective on the Martini model. *Chem. Soc. Rev.* **2013**, 42, 6801–22.
- [57] Dick C.R., Ham G.E. Characterization of Polyethylenimine. *J. Macromol. Sci. Part A-Chem.* **1970**, 4, 1301–14.
- [58] Suh J., Paik H.J., Hwang B.K. Ionization of Poly(ethylenimine) and Poly(allylamine) at Various pH's. *Bioorg. Chem.* **1994**, 22, 318–27.
- [59] Tang G.P., Guo H.Y., Alexis F., Wang X., Zeng S., Lim T.M., et al. Low Molecular Weight Polyethylenimines Linked by β -Cyclodextrin for Gene Transfer Into the Nervous System. *J. Gene Med.* **2006**, 8, 736–44.
- [60] Marrink S.J., Risselada H.J., Yefimov S., Tieleman D.P., de Vries A.H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, 111, 7812–24.
- [61] Monticelli L., Kandasamy S.K., Periole X., Larson R.G., Tieleman D.P., Marrink S.J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, 4, 819–34.
- [62] Seo M., Rauscher S., Pomès R., Tieleman D.P. Improving Internal Peptide Dynamics in the Coarse-Grained MARTINI Model: Toward Large-Scale Simulations of Amyloid- And Elastin-Like Peptides. *J. Chem. Theory Comput.* **2012**, 8, 1774–85.
- [63] Uusitalo J.J., Ingólfsson H.I., Akhshi P., Tieleman D.P., Marrink S.J. Martini Coarse-Grained Force Field: Extension to DNA. *J. Chem. Theory Comput.* **2015**, 11, 3932–45.
- [64] Rossi G., Monticelli L., Puisto S.R., Vattulainen I., Ala-Nissila T. Coarse-Graining Polymers With the MARTINI Force-Field: Polystyrene as a Benchmark Case. *Soft Matter* **2011**, 7, 698–708.
- [65] Lee H., de Vries A.H., Marrink S.J., Pastor R.W. A Coarse-Grained Model for Polyethylene Oxide and Polyethylene Glycol: Conformation and Hydrodynamics. *J. Phys. Chem. B* **2009**, 113, 13186–94.
- [66] Yesylevskyy S.O., Schäfer L. V., Sengupta D., Marrink S.J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comput. Biol.* **2010**, 6, e1000810.
- [67] Michalowsky J., Schäfer L. V., Holm C., Smiatek J. A Refined Polarizable Water Model

- for the Coarse-Grained MARTINI Force Field With Long-Range Electrostatic Interactions. *J. Chem. Phys.* **2017**, 146, 054501.
- [68] de Jong D.H., Singh G., Bennett W.F.D., Arnarez C., Wassenaar T.A., Schäfer L. V., et al. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **2013**, 9, 687–97.
- [69] Utsuno K., Uludağ H. Thermodynamics of Polyethylenimine-DNA Binding and DNA Condensation. *Biophys. J.* **2010**, 99, 201–7.
- [70] Martini tutorial: Polymers, <http://cgmartini.nl>, **2018**.
- [71] Abraham M.H., Platts J.A., Hersey A., Jeo A.J., Taft R.W. Correlation and Estimation of Gas–Chloroform and Water–Chloroform Partition Coefficients by a Linear Free Energy Relationship Method. *J. Pharm. Sci.* **1999**, 88, 670–9.
- [72] Duffy E.M., Jorgensen W.L. Prediction of Properties From Simulations: Free Energies of Solvation in Hexadecane, Octanol, and Water. *J. Am. Chem. Soc.* **2000**, 122, 2878–88.
- [73] Tschöp W., Kremer K., Batoulis J., Bürger T., Hahn O. Simulation of Polymer Melts. I. Coarse-Graining Procedure for Polycarbonates. *Acta Polym.* **1998**, 49, 61–74.
- [74] Reith D., Pütz M., Müller-Plathe F. Deriving Effective Mesoscale Potentials From Atomistic Simulations. *J. Comput. Chem.* **2003**, 24, 1624–36.
- [75] Jorgensen W.L. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water. *J. Am. Chem. Soc.* **1981**, 103, 335–40.
- [76] Hess B., Kutzner C., van der Spoel D., Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, 4, 435–47.
- [77] Bussi G., Donadio D., Parrinello M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, 126, 014101.
- [78] Berendsen H.J.C., Postma J.P.M., van Gunsteren W.F., DiNola A., Haak J.R. Molecular Dynamics With Coupling to an External Bath. *J. Chem. Phys.* **1984**, 81, 3684–90.
- [79] Parrinello M., Rahman A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, 52, 7182–90.
- [80] Darden T., York D., Pedersen L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, 98, 10089–92.
- [81] Essmann U., Perera L., Berkowitz M.L., Darden T., Lee H., Pedersen L.G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, 103, 8577–93.

- [82] MacKerell Jr A.D., Banavali N., Foloppe N. Development and Current Status of the CHARMM Force Field for Nucleic Acids. *Biopolymers* **2000**, 56, 257–65.
- [83] Kumar S., Rosenberg J.M., Bouzida D., Swendsen R.H., Kollman P.A. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, 13, 1011–21.
- [84] Abraham M.J., Murtola T., Schulz R., Páll S., Smith J.C., Hess B., et al. GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism From Laptops to Supercomputers. *SoftwareX* **2015**, 1–2, 19–25.
- [85] Hub J.S., de Groot B.L., van der Spoel D. g_wham-A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **2010**, 6, 3713–20.
- [86] Dobrynin A. V, Rubinstein M. Theory of Polyelectrolytes in Solutions and at Surfaces. *Prog. Polym. Sci.* **2005**, 30, 1049–118.
- [87] Abdallah B., Hassan A., Benoist C., Goula D., Behr J.P., Demeneix B.A. A Powerful Nonviral Vector for In Vivo Gene Transfer into the Adult Mammalian Brain: Polyethylenimine. *Hum. Gene Ther.* **1996**, 7, 1947–54.
- [88] Lee H., Larson R.G. Molecular Dynamics Simulations of PAMAM Dendrimer-Induced Pore Formation in DPPC Bilayers With a Coarse-Grained Model. *J. Phys. Chem. B* **2006**, 110, 18204–11.
- [89] Panizon E., Bochicchio D., Monticelli L., Rossi G. MARTINI Coarse-Grained Models of Polyethylene and Polypropylene. *J. Phys. Chem. B* **2015**, 119, 8209–16.

Chapter 4

Polyethylenimine-DNA Nanoparticle Formation*

*A version of this chapter's sections has been adapted with permission from Mahajan S., and Tang T., Polyethylenimine-DNA Ratio Strongly Affects Their Nanoparticle Formation: A Large-Scale Coarse-Grained Molecular Dynamics Study, *J. Phys. Chem. B* **2019**, 123, 9629-9640. Copyright © 2019 *American Chemical Society*.

4.1. Introduction

Over the last few decades, gene delivery has been gaining attention as a means to treat genetic defects.[1,2] In this process, genetic materials such as DNAs are “delivered” to the cell to induce a therapeutic response against genetic defects, which can be hereditary or arise from harsh environmental conditions. One challenge in DNA delivery is to find an effective delivery system because DNAs by themselves are ineffective in cell internalization and are prone to degradation.[3] Gene delivery was first achieved using viral vectors because viruses are capable of transferring their genetic materials into a cell to reproduce.[4] Although viral vectors are modified so that they cannot replicate, they may still result in immune response in the host and therefore have significant safety concerns. As a result, non-viral delivery has been investigated due to the low cost, ease of synthesis, and relatively lower safety concerns. The efficacy of non-viral delivery, however, is not yet at par with viral delivery.

There have been numerous studies on non-viral vectors and among them, polyethylenimine (PEI) has shown a lot of potential.[3,5–9] PEIs are comprised of repeating amine groups, which can form a wide variety of structures from completely linear to highly dendritic. They are highly

soluble in water, and due to the presence of many nitrogen atoms, they act as a buffer over a large range of pH.[1,10] At physiological pH, PEIs are weak bases and positively charged, with ~21% of their nitrogen being protonated.[11] Meanwhile, DNAs are strong acids and negatively charged due to the presence of phosphates on their backbone. Adding PEIs to DNAs results in charge neutralization and condensation of DNAs, and aggregation of these condensed DNAs to form nanoparticles (NPs).[12] Properties of the NPs such as hydrodynamic radius, surface charge and shape depend on the structure and molecular weight of PEI being used, molecular modifications made to PEIs such as the grafting of aliphatic chains, as well as the N/P ratio (ratio of the total number of amines in PEIs to that of phosphates in DNAs). These NP properties play a significant role in the various steps of a gene delivery process.

Take for example the step of NP internalization, i.e., cellular uptake. A successful cellular uptake starts with the adsorption of the NP onto the plasma membrane. The overall negatively charged nature of the plasma membrane promotes the adsorption of NPs with positive surface charge.[13,14] Adsorption of negatively charged NPs has also been observed due to the presence of cationic sites on the plasma membrane, but the adsorption is limited.[15] Neutral NPs are difficult to be adsorbed[16], but in intravenous delivery they have the advantage of avoiding non-specific interactions with other biomolecules such as proteins before reaching the target cells[17,18]. Upon adsorption, NPs are internalized and it has been shown that PEI-based gene delivery primarily uses clathrin-mediated endocytosis.[19] This endocytic pathway encapsulates NPs in endosomes with a diameter ranging from 100-150 nm[19,20], thereby placing a limit on the size of the NPs that can be internalized. It is therefore important to be able to control the size and charge of the NPs, in order to improve their cellular uptake.

Extensive studies have been carried out to modulate the size (in terms of hydrodynamic radius) and charge (in terms of zeta potential) of DNA-PEI NPs by changing the N/P ratio.[21,22] Zeta potential is calculated to quantify the surface charge of the NP, which is negative for low N/P ratio and positive for high N/P ratio. It is well accepted that for intermediate N/P ratios where the zeta potential is close to zero, the hydrodynamic radius of the NPs is several folds larger than other N/P ratios.[21] The hydrodynamic radius is not a single value but follows certain distributions.[10,12,23] Identifying these distributions can be important because a successful gene delivery might be enabled by a range of hydrodynamic radius and not the mean value.[24,25] Time variation of the distributions has not been addressed, which can be important in understanding the NP formation process and how we may control it. The NPs are known to form various shapes such as spheres, toroids, rods, and folded loops[10,12,23], which might be crucial for cellular uptake[24,26–28], but their effect on gene delivery is not well explored.

In recent years, many molecular dynamics (MD) simulations have been performed to generate atomistic images on the aggregation of DNAs and PEIs. Dynamics of DNA-PEI binding and their binding free energy were studied for PEIs with different molecular weight, degrees of branching and protonation ratios using all-atom (AA) MD simulations.[29–33] Binding of large polycations to single DNA (or RNA) was also studied with coarse-grained (CG) MD simulations.[34–36] Aggregation of multiple DNAs was studied using AA MD; these simulations were small in size and the largest system contained only four short DNAs each having 12 base pairs.[37,38] At a larger length scale, CG simulations were used to study the aggregation of DNAs, modeled as linear polyelectrolytes, by mono- and multi-valent ions.[39–48] Similar CG simulations are not available for DNA aggregation mediated by polycations such as PEI. Due to their size and structure, polymers can aggregate DNAs in ways different from mono- and multi-

valent ions. Moreover, modeling DNAs as linear polyelectrolytes[39–48] is highly approximate, which lacks the double helical structure and major/minor grooves. These structural features are important as PEIs have been observed to bind with DNA along major and minor grooves.[29] Such details cannot be captured by mean-field theories such as Poisson-Boltzmann. Together with the size limitation in AA simulations, there is a need to study PEI-induced DNA aggregation and NP formation using large-scale CG simulations with models comprised of accurate molecular structures.

In this work, large-scale CG-MD simulations are performed to investigate the aggregation of a large number of DNA and PEI molecules. Since N/P ratio plays a significant role in determining the properties of NPs, we focus on its influence on aggregation dynamics, and NP size, shape and charge. This is done by simulating 5 systems with different PEI/DNA number ratios (hence different N/P ratios) and at two different salt concentrations. Similar to the experiments of Ogris *et al.*[49], the systems are prepared by keeping the number of DNAs fixed while changing the number of PEIs. The rest of the paper is organized as follows. Preparation of the simulated systems and details of the simulations are provided in **Section 4.2**. Results are presented in **Section 4.3**. In **Section 4.3.1-4.3.3** we focus on systems with only neutralizing salt ions, and address NP shape (**Section 4.3.1**), charge and size (**Section 4.3.2**) and the rate of NP growth (**Section 4.3.3**). Results in the presence of 150 mM salt are presented and compared in **Section 4.3.4**. Further discussion of the results, their comparison with experiments and practical implications are given in **Section 4.4**.

4.2. Methods

The DNA used in the simulations is a Drew-Dickerson dodecamer $d(\text{CGCGAATTCGCG})_2$ in its canonical B form with a charge (Q_{DNA}) of -22 . At the CG level, it is modeled using the stiff

elastic network in the Martini force field developed by Uusitalo *et al.*[50] The PEI simulated is a 586 Da branched PEI with its structure and protonation sites shown in **Figure 4.1**. [29] Three amines are protonated ($Q_{PEI} = 3$) according to the $\sim 21\%$ protonation ratio at physiological pH.[11,29] The CG PEI model was developed in our previous work within the Martini framework and was shown to be compatible with Martini DNA.[51] Specifically, each CG bead contains C-C-N atoms with the accompanying hydrogens, and the mapping scheme is displayed as blue enclosures in **Figure 4.1**. The PEI beads are classified as ‘*t*’, ‘*s*’, ‘*p*’ and ‘*pq*’, which respectively denote beads containing tertiary, secondary, primary and protonated primary nitrogens.[51]

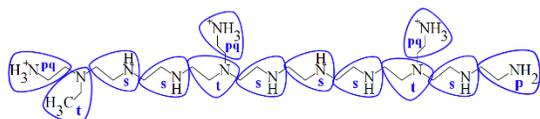


Figure 4.1: All-atom (black) and coarse-grained (blue) structures of PEI used in this study. The blue enclosures represent the CG mapping scheme. The CG beads were classified based on the type of nitrogen within the bead, i.e., tertiary (*t*), secondary (*s*), primary (*p*), and protonated primary (*pq*).[51]

Five systems are simulated each containing the same number of DNAs ($N_{DNA} = 27$) but a different number of PEIs (N_{PEI}), corresponding to the PEI/DNA number ratio (α) of 2, 4, 6, 8 and 10 (**Table 4.1**). In each system, the principal axes of the 27 DNAs were aligned and arranged in a $3 \times 3 \times 3$ array, which was then placed in a cubic simulation box of side length 25 nm. The distance between the centers of mass of two closest DNAs was 6 nm. The PEIs were randomly placed within the box, and the solutes were solvated with polarizable Martini water[52]. An appropriate amount of Martini K^+ or Cl^- ions[53] were introduced to neutralize each system (see **Table 4.1**). Systems with $\alpha = 2-6$ contained excess DNAs and required cationic (K^+) ions to neutralize the systems, whereas the systems with $\alpha = 8-10$ had excess PEIs and required anionic (Cl^-) ions. For each system in **Table 4.1**, an additional system was simulated by adding 150 mM KCl and

correspondingly removing an appropriate amount of water molecules. Comparison of the two systems at the same α allows us to address the effect of salt.

Table 4.1: Details of simulated systems. N_{DNA} , N_{PEI} , N_{ion} are respectively the total numbers of DNAs, PEIs and ions. The N/P ratio is calculated using $13\alpha/22$, where 13 is the number of nitrogen atoms in each PEI and 22 is the number of phosphorus atoms in each DNA. Corresponding systems with 150 mM KCl had the same values for N_{DNA} and N_{PEI} , but an additional 150mM KCl in each system.

α	N_{DNA}	N_{PEI}	N_{ion}	N/P ratio
2	27	54	432 (K ⁺)	1.18
4	27	108	270 (K ⁺)	2.36
6	27	162	108 (K ⁺)	3.55
8	27	216	54 (Cl ⁻)	4.73
10	27	270	216 (Cl ⁻)	5.91

All simulations were performed using the Gromacs 5 package.[54] Each system was first energy minimized using steepest-descent, followed by a constrained MD in NPT ensemble for 1 ns, and an unconstrained MD in NPT ensemble for 4 μ s. Since the potential energy surface of the Martini force field is smooth[55], NVT equilibration after energy minimization was not required. In the constrained MD simulations, all the bonds of PEI and the backbone bonds of DNA were constrained using LINCS[56]. The temperature was maintained at 300 K using velocity rescaling thermostat[57] with a time constant of 0.1 ps. The pressure was maintained at 1 bar using Berendsen thermostat[58] with a time constant of 3 ps and compressibility of 3×10^{-4} bar⁻¹. Unconstrained NPT simulations were conducted at the same temperature and pressure, using the same parameters for the thermostat. However, the pressure was maintained using Parrinello-Rahman barostat[59] with a time constant of 5 ps and compressibility of 4.5×10^{-4} bar⁻¹. For all

simulations, a neighbour-list was maintained within a radius of 1.1 nm using the Verlet scheme[60], and was updated every 40 time steps. Electrostatic interactions were calculated using a relative dielectric constant of 2.5 and were cut off at 1.1 nm using the reaction field scheme[61]. Similarly, van der Waals interactions were cut-off at 1.1 nm using the potential-shift-Verlet scheme. Periodic boundary condition was applied in all directions, and leapfrog integrator was used with a time step of 5 fs. Unless otherwise specified, all the simulation time reported below are scaled by a factor of 4, whereas the simulation parameters (time constants, time step) mentioned above are unscaled.[53]

4.3. Results

4.3.1. Mode of aggregation and roles of PEI

At the end of the simulation, several DNA-PEI NPs are formed in each system. **Figure 4.2** shows the largest NP for different α . The size of the NP (s_{NP}) is quantified by the number of DNAs in the NP. Upon visual inspection, in **Figure 4.2a** where $\alpha = 2$, the aggregation is very limited. In **Figure 4.2b, c** where $\alpha = 4$ and 6, the DNAs in the NP tend to connect to one another from end-to-end, leading to a loose linear structure resembling a string. On the contrary, in **Figure 4.2d, e** where $\alpha = 8$ and 10, the DNAs in the NP have established some side-by-side connections forming more compact branched structures, representing a different mode of aggregation. By calculating the relative shape anisotropy (see **Appendix B.1**), the branched structures are closer to being spherical compared with the linear structures. Clearly, the mode of aggregation is dependent on the number of PEIs present in the system and how they bind to the DNA. To facilitate the discussion, we identify different roles of PEI in the dynamics of DNA aggregation. A PEI is defined to be bound to a DNA if the minimum distance between any one bead in the PEI and any one bead in the DNA is less than 0.53 nm, which is the equilibrium distance for the Lennard-Jones interaction potential between DNA and PEI beads (see **Appendix B.2** for further discussion on

DNA-PEI binding). A PEI is categorized as *free* if it is not bound to any DNA, *peripheral* if it is bound to only one DNA, and *bridging* if it is simultaneously bound to two or more DNAs to form a polycation bridge.

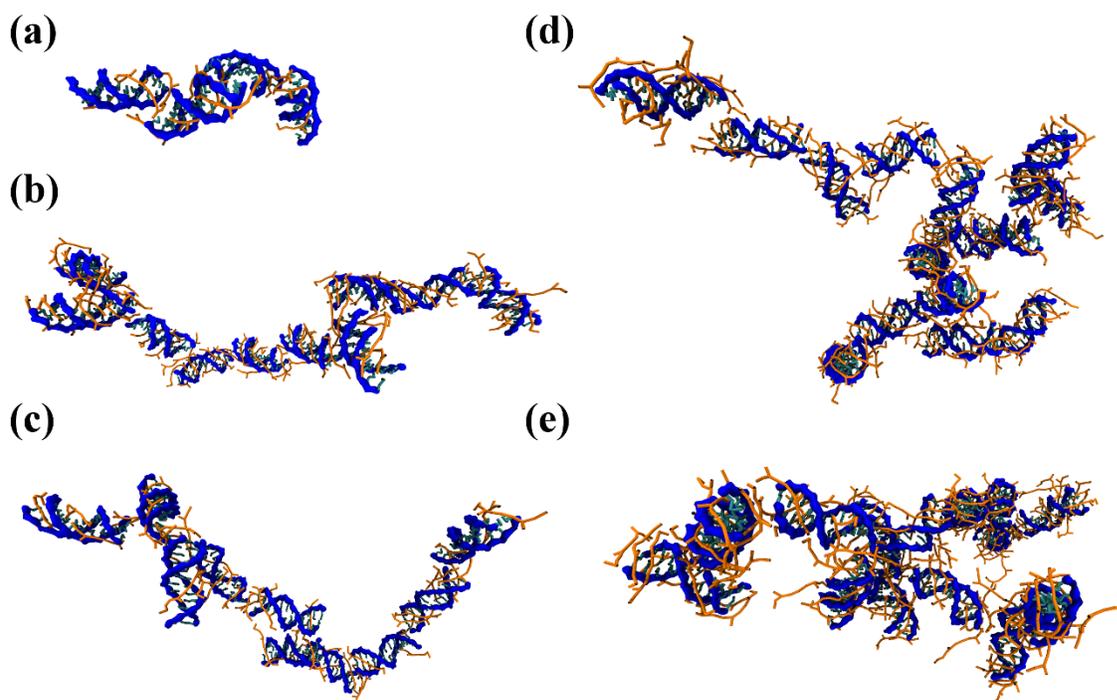


Figure 4.2: Largest NP formed at the end of 4 μ s simulation. (a) $\alpha = 2$, $s_{NP} = 4$; (b) $\alpha = 4$, $s_{NP} = 10$; (c) $\alpha = 6$, $s_{NP} = 10$; (d) $\alpha = 8$, $s_{NP} = 15$ and (e) $\alpha = 10$, $s_{NP} = 14$. DNAs are represented with blue backbone and cyan base pairs, and PEIs are represented in orange.

Figure 4.3a-c shows the numbers of free (N_f), peripheral (N_p) and bridging (N_b) PEIs as functions of simulation time, for different α 's. For all systems, PEIs starts as free (**Figure 4.3a**), but the number quickly reduces as the PEIs become peripheral (**Figure 4.3b**) or bridging (**Figure 4.3c**). At any time, N_b is significantly lower than N_p , indicating that the majority of free PEIs convert to peripheral PEIs. Furthermore, by examining the number of conversions among free, peripheral and bridging PEIs, we find that direct conversion from free to bridging PEI is rare. That is, bridging PEI is formed dominantly from peripheral PEI, and peripheral PEI from free PEI (see **Appendix B.3** for more details).

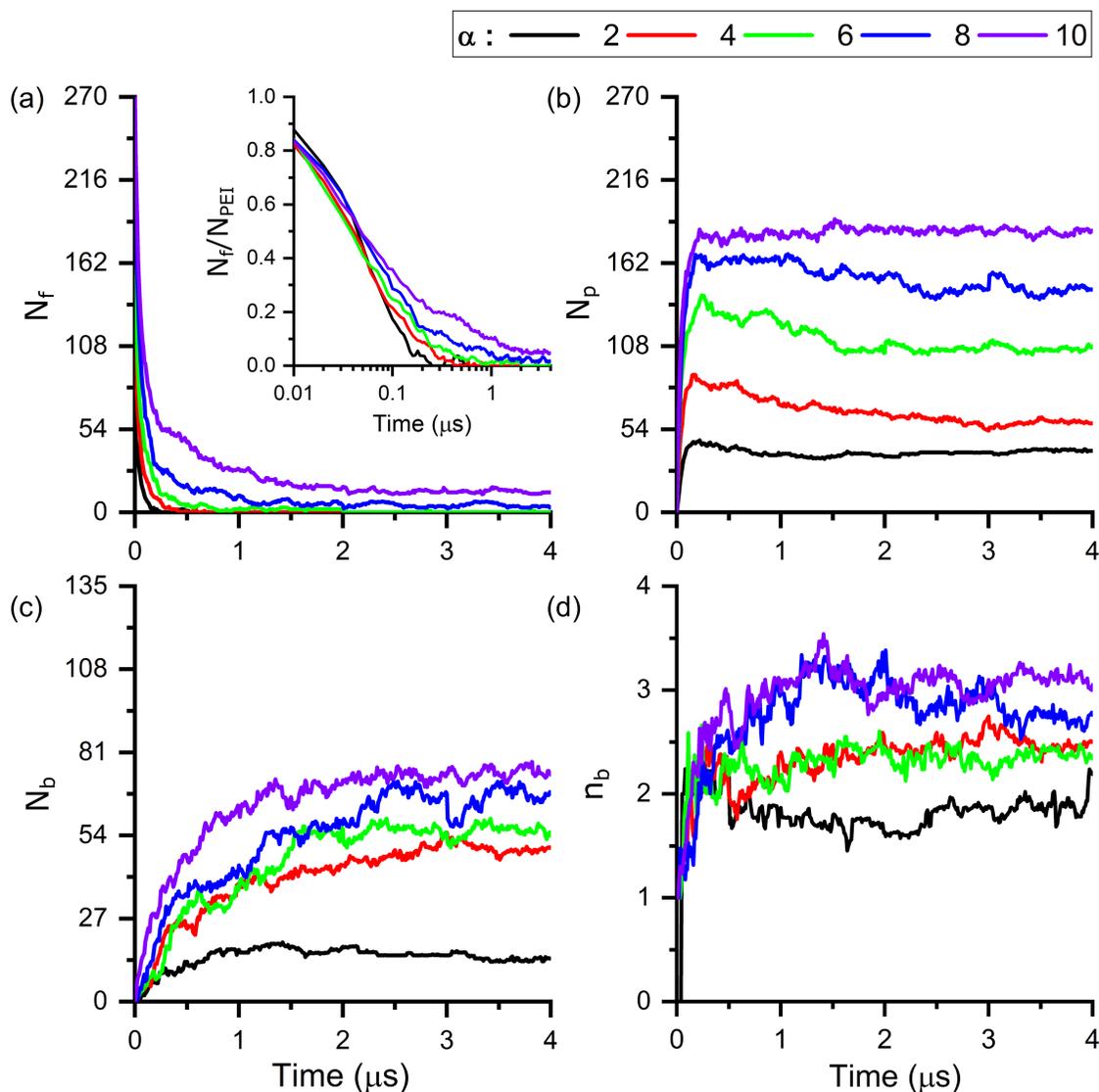


Figure 4.3: Number of PEIs in different roles, plotted against simulation time. (a) free (N_f), (b) peripheral (N_p), (c) bridging (N_b), and (d) average number of bridging PEI between a pair of bridged DNAs (n_b).

In **Figure 4.3a**, free PEIs shows an exponential-like decay, where the rate of decay (negative of the slope) decreases with time for all α . For systems with excess DNA ($\alpha = 2-6$) N_f decreases to zero before 500 ns, whereas for systems with excess PEI ($\alpha = 8-10$) N_f remains non-zero even at the end of the simulations. The non-zero N_f suggests that in excess of PEIs there exists a maximum number of PEIs that can bind to a DNA. To estimate this maximum, we calculate the

average number of PEIs bound to a DNA, $(N_p + N_b)/N_{DNA}$, over the last 2 μ s. The average is 7.89 for $\alpha = 8$ and 9.48 for $\alpha = 10$, which implies that the maximum number of PEIs that binds to a DNA increases with α . Given that the number of DNA is the same for all systems, the results suggest that the fashion in which the PEIs bind to the DNA may vary with α .

In **Figure 4.3b**, N_p increases and reaches its maximum value at ~ 250 ns for all α . However, its behavior beyond 250 ns is different. For systems with excess DNA ($\alpha = 2-6$), N_p slowly decreases, which is due to the conversion of peripheral PEIs to bridging ones. At the same time no new peripheral PEI is formed due to the exhaustion of free PEIs. For systems with excess PEI ($\alpha = 8-10$), N_p remains almost constant because there is an abundance of free PEIs. As peripheral PEIs convert to bridging ones, it is likely that some of the DNA phosphates previously “covered” by the peripheral PEIs are now exposed, allowing free PEIs to bind. Therefore, N_p remains constant due to a dynamic balance between the conversions from free to peripheral PEIs and from peripheral to bridging PEIs.

In **Figure 4.3c**, bridging PEIs increases monotonically with time, and the final equilibrium value of N_b is larger for higher α . A larger number of bridging PEIs could imply a greater number of DNAs being aggregated together; it could also imply a larger number of bridges being formed between a pair of aggregated DNAs. To investigate this, we plot the average number of bridging PEIs between a pair of bridged DNAs (n_b) in **Figure 4.3d**. Details on the calculation of n_b can be found in **Appendix B.4**. The 5 n_b curves appear to form 3 groups: $\alpha = 2$ corresponds to the lowest value of n_b , $\alpha = 4$ and 6 correspond to an intermediate n_b , and the highest value of n_b occurs for $\alpha = 8$ and 10. Interestingly these 3 groups correlate with the observations made earlier on the mode of aggregation, where the aggregation was limited for $\alpha = 2$, end-to-end for $\alpha = 4$ and 6 and side-by-side for $\alpha = 8$ and 10. More discussions are available in **Appendix B.4**.

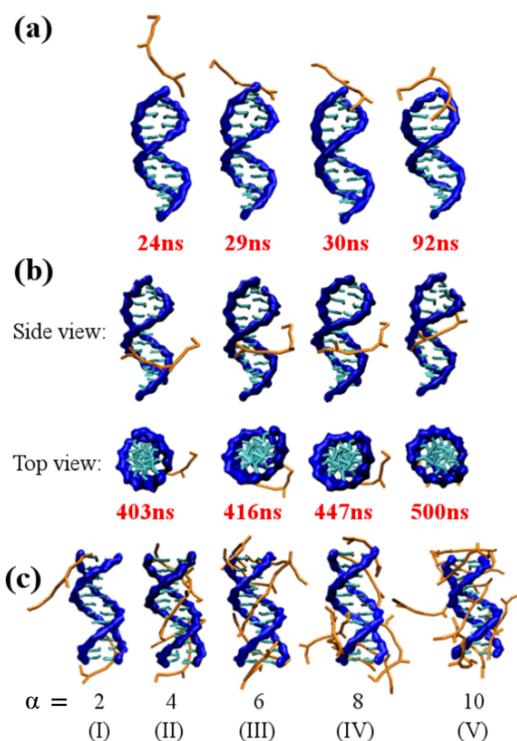


Figure 4.4: Snapshots of a single PEI binding to a DNA. (a) from the terminal and (b) at the center. (c) Snapshots of PEIs bound to a DNA for different PEI/DNA ratios at 500 ns. DNAs are represented with blue backbone and cyan base pairs, and PEIs are represented in orange.

Having understood the different roles PEIs play and the conversion between these roles during the aggregation process, we now discuss why different modes of aggregations are found under different PEI/DNA ratio. The aggregation starts by PEIs binding to the DNAs (conversion from free to peripheral PEIs), followed by bridging of different DNAs (conversion from peripheral to bridging PEIs). As discussed earlier, different PEI/DNA ratios can lead to different fashions in which PEIs bind to a DNA. To explore this, we first look at single PEI binding to a DNA and report two distinct ways in which a PEI approaches a DNA. **Figure 4.4a** shows snapshots, taken at different time as indicated, of a single PEI binding to a DNA at one of its terminals. Since the PEI in this study has three protonated beads, it can make three primary points of contact with the DNA. As shown in **Figure 4.4a**, the PEI first makes a one-point contact (24 ns), and then slowly

moves along the backbone of the DNA to form a two-point contact (29 and 30 ns). Some PEIs further align themselves with the backbone or bend over the terminal (the PEI shown at 92 ns) to form a three-point contact, although two-point contact is dominant for PEIs binding at the DNA terminals.

Figure 4.4b shows snapshots, also taken at different time as indicated, of a single PEI binding to a DNA at its center. Similarly, the PEI first makes a one-point contact (403 ns), gradually moves closer to the DNA backbone (416 and 447 ns) forming a two-point contact and can eventually form a three-point contact completely aligned with the backbone (500 ns). It should be noted that the snapshots in **Figure 4.4a, b** are selected from our simulation trajectory to clearly show the binding process; they are not an indication that binding at a DNA's center occurs later than binding at one of the terminals.

Figure 4.4c shows one DNA from each system ($\alpha = 2-10$) at 500 ns which has had PEIs bound to it but has not aggregated with other DNAs yet. At low α values (**Figure 4.4c-I**), the number of bound PEIs are very few, resulting in limited aggregation as seen in **Figure 4.2a**. At moderate α values (**Figure 4.4c-II, III**), there are several peripheral PEIs and they bind to the DNA at different locations. PEIs at the terminals mostly form two-point contact leaving one additional binding site available, whereas PEIs at the center mostly form three-point contact making it difficult to interact with another DNA to form a bridge. If this DNA were to aggregate with another DNA, it would be facilitated by the PEIs at the terminals but not those at the center. This forces the DNAs to undergo end-to-end aggregation forming loose linear NP as seen in **Figure 4.2b, c**. For systems with excess PEI, i.e., large α (**Figure 4.4c-IV, V**), the competition among the PEIs to bind with the DNA prevents PEIs binding at the center from complete alignment with the backbone. This misalignment reduces three-point contacts made by center-binding PEIs,

allowing them to bind with other DNAs through a lateral side-by-side aggregation, and forming more compact branched NPs as seen in **Figure 4.2d-e**.

The PEI/DNA ratio α represents how crowded the PEIs are around the DNAs. Since each DNA has 22 deprotonated phosphates and each PEI has 3 protonated amines, on average each DNA can be saturated with 7.33 PEIs. The average number of PEIs bound to a DNA, $(N_p + N_b)/N_{DNA}$, in the last 2 μ s is 2, 4, 6, 7.89 and 9.48 for $\alpha = 2, 4, 6, 8$ and 10 respectively. The first three numbers (< 7.33) represent under-saturation of the DNAs by the PEIs, while the latter two (> 7.33) represent over-saturation. The data shows not only increased PEI crowding with the increase in α , but also the extra accommodation of PEIs when α is large. From **Figure 4.4a, b** it is clear that bound PEIs takes time (~ 100 ns) to conform to the DNA structure, while free PEIs continue to bind with the DNA. In the inset of **Figure 4.3a**, we plot the ratio between the numbers of free PEI and the total number of PEI (N_f/N_{PEI}) as a function of time using a semi-log plot. For the first 100 ns N_f/N_{PEI} is similar for all α , implying that the initial rate of decay of N_f is proportional to N_{PEI} . In other words, the rate of PEI binding to DNA increases with α and more PEIs can bind to a DNA within this time window for higher α . DNAs for $\alpha = 2, 4, 6$ are under-saturated and the PEIs still have time to adjust their configurations and become aligned to the DNAs at the center. Whereas the extra accommodation of PEIs made by the DNAs at $\alpha = 8$ and 10 are traded with misalignment of center-binding PEIs. The misalignment is expected to be larger for $\alpha = 10$, as on average each DNA in $\alpha = 10$ accommodates 1.6 more PEIs than in $\alpha = 8$. Overall, our results above demonstrate that the PEI/DNA ratio influences the initial rate of PEI binding to individual DNAs and subsequent alignment of PEI to the DNA backbones. The alignment, in turn, affects the mode of aggregation of multiple DNAs, leading to different shapes of the NPs. It should be noted that DNAs can also be over-saturated if a deprotonated phosphate bead simultaneously interacts

with multiple protonated amine beads. This is in fact observed in our simulations. While the number of phosphate beads interacting with three amine beads is negligible, a fraction of phosphate beads can interact with two amine beads at the same time, and the number increases with α (see **Appendix B.5** for further discussions).

4.3.2. NP charge and size

Figure 4.5 shows the size distribution of the NPs (black) and the average charge of the NPs at a given size (red). Data used to generate the plots are collected from the last 0.5 μs of the simulations, which is regarded as steady state (see **Appendix B.6** for details). The horizontal axis of each subplot is the size of the NP (s_{NP}) quantified by the number of DNAs in the NP. The time-averaged number of NPs having size s_{NP} are plotted on the left axis representing the size distribution. Meanwhile, the sum of charges from all NPs with size s_{NP} is calculated. Dividing this total charge by the number of NPs at the same s_{NP} gives the average NP charge (Q_{NP}) and it is plotted on the right axis. If for a given s_{NP} no NPs are found, Q_{NP} is set to zero (shown as the red dashed line in **Figure 4.5**). When $\alpha = 4$ and for $s_{NP} = 15$, Q_{NP} is non-zero while the average number of NPs appears to be zero, which seems contradictory. We point out that the average number of NPs is in fact not zero but a very small number (0.025), which represents unstable NPs formed for a short period of time during the last 0.5 μs . Similar unstable NPs are observed for $\alpha = 2$ at $s_{NP} = 5$ and $\alpha = 6$ at $s_{NP} = 13$ and 17. No unstable NPs are observed for systems with excess PEIs. The statistical significance of the average number and charge of NPs is ensured by the large time window of 0.5 μs .

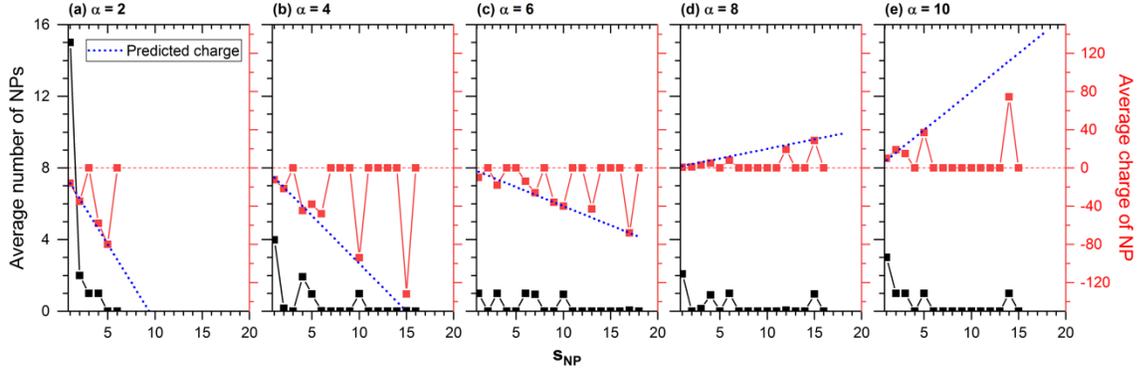


Figure 4.5: Steady-state NP size distributions (black) and average charge of NPs as a function of size (red). $\alpha = 2, 4, 6, 8,$ and 10 for (a-e) respectively. The predicted charge (blue) is calculated using Eq. 4.1.

Considering only the non-zero Q_{NP} , it is interesting to see that its magnitude increases almost linearly with s_{NP} for all α . Consulting the initial slopes (before 250 ns) of **Figure 4.3b, c** it is clear that the rate of formation of peripheral PEIs is much faster than that of bridging PEIs. As a result, for any α the DNAs in the system are likely to have a similar number of peripheral PEIs before they begin to aggregate. Therefore, it is reasonable to hypothesize that if these DNAs aggregate into a NP, its charge would be proportional to the number of DNAs in it. As a very rough approximation, we assume all the PEIs in the system are bound to DNAs before they begin to aggregate. Under this assumption, the NP charge can be predicted (Q_{NP}^{pred}) as a function of s_{NP} and α using Eq. 4.1, where Q_{DNA} , N_{DNA} , Q_{PEI} and N_{PEI} are the charge and number of a single DNA and a single PEI respectively. In **Figure 4.5**, Q_{NP}^{pred} is plotted (blue dots) alongside the average NP charge from the simulations and are found to be in good agreement with each other. This suggests that the assumption is quite reasonable, which is because most of the PEIs in the systems have become bound to the DNAs before they begin to aggregate. The remaining small number of PEIs that bind to the NPs later in the simulation only results in small deviation from the predicted charge. The deviation of Q_{NP} from Q_{NP}^{pred} is negligible for $\alpha = 2-8$ because the number

of free PEIs (N_f , see **Figure 4.3a**) is eventually zero (for $\alpha = 2-6$) or close to zero ($\alpha = 8$). The discrepancy, however, is considerable for $\alpha = 10$ because N_f (= 13 at steady state) is not negligible. The excessive PEIs in $\alpha = 10$ cannot all be accommodated by the DNAs and therefore the charge of NPs is lower than our prediction. The discrepancy is particularly large for large NPs (see $s_{NP} = 14$ in **Figure 4.5e**) because although some PEIs bind to the NPs later in the simulation, they tend to bind to smaller NPs due to their lower positive charge.

$$Q_{NP}^{pred} = s_{NP}(N_{DNA}Q_{DNA} + N_{PEI}Q_{PEI})/N_{DNA} = s_{NP}(Q_{DNA} + \alpha Q_{PEI}) \quad 4.1$$

Since Q_{NP} at the steady state is proportional to s_{NP} , the quantity Q_{NP}/s_{NP} represents the degree of charge neutralization of DNAs in the NP. While Q_{NP} in **Figure 4.5** is calculated at the steady state, it is of interest to study how the degree of neutralization changes during the aggregation process. In **Figure 4.6a**, the average $\langle Q_{NP}/s_{NP} \rangle$ is shown for different α , where the average is performed over different time windows and additionally for each time window, it is averaged over different s_{NP} . The prediction Q_{NP}^{pred}/s_{NP} from Eq. 4.1 is also shown for comparison. For $\alpha = 2$ and 4 the prediction (Eq. 4.1) could accurately estimate $\langle Q_{NP}/s_{NP} \rangle$ for all time windows, whereas for $\alpha = 6-10$ the prediction is accurate only for $t > 1 \mu\text{s}$. The lack of agreement for $t < 1 \mu\text{s}$ is because not all the PEIs are bound to DNAs during this time window. Although our approximation of NP charge (Eq. 4.1) is simple, its prediction is accurate for $t > 1 \mu\text{s}$ (see **Appendix B.7** for an extended, time-dependent model for $Q_{NP}^{pred}(t)$). In the following, to ease the discussion on the aggregation process we will use Q_{NP}^{pred} instead of Q_{NP} which does not change with simulation time.

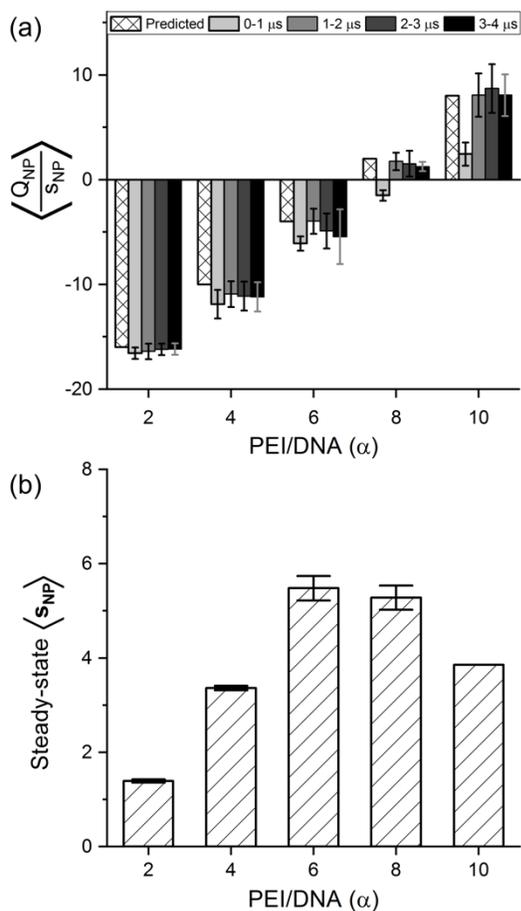


Figure 4.6: Characteristics of charge and size of NPs. (a) Average charge associated with one DNA in the NPs, representing the degree of neutralization. (b) Steady-state average NP size.

Q_{NP}^{pred}/s_{NP} takes the values of $-16, -10, -4, 2,$ and 8 for $\alpha = 2, 4, 6, 8$ and 10 respectively.

This implies that on average, the NPs in systems with $\alpha = 2-6$ are negative (under-neutralized), and the NPs in systems with $\alpha = 8-10$ are positive (over-neutralized). NPs with the same sign of net charge would have long-range electrostatic repulsion and the strength of repulsion would depend on the product of their net charges. Since the charge of a NP is approximately proportional to its size (Eq. 4.1), we define an intrinsic property, $(Q_{NP}^{pred}/s_{NP})^2$, as the *specific repulsion*. The specific repulsion directly depends on the degree of neutralization of DNAs in a NP and approximates the strength of repulsion between two NPs with $s_{NP} = 1$. It only depends on α of

the system, and we classify it as strong for $\alpha = 2$, moderate for $\alpha = 4$ and 10, and weak for $\alpha = 6$ and 8. Despite the long-range repulsion NPs do aggregate, which can only be achieved through local electrostatic attraction, i.e., the attraction between protonated amines in one NP and deprotonated phosphates in another NP. Therefore, aggregation of DNAs is hindered by long-range electrostatic repulsion but promoted by the diffusion of NPs which brings two NPs close enough so that they may aggregate via local electrostatic attraction.

To understand the effects of diffusion we will borrow some results from the Smoluchowski coagulation equation[62] (SCE) for spherical Brownian particles. SCE states that the aggregation of two particles is faster if the two particles have a larger ratio between their radii. For example, aggregation is slow for two large particles due to their low diffusion coefficient, and for two small particles due to their small radius of sphere of action which prevents their collision. On the other hand, it is easier for a small particle which has a higher diffusion coefficient to collide and aggregate with a large particle which has a larger radius of sphere of action. It should be noted that strictly speaking SCE is not directly applicable for our systems because our aggregating particles (DNAs) resemble cylinders and not spheres. However, the qualitative results mentioned above help us better understand the size distribution observed in **Figure 4.5**.

Regardless of the value of α , the NP size distribution in **Figure 4.5** has a peak located at $s_{NP} = 1$ representing unaggregated DNAs. Excluding $s_{NP} = 1$, three distinct shapes are observed for the size distribution akin to the specific repulsion defined above. For $\alpha = 2$, we hardly notice any aggregation of DNAs (only small-NPs with $s_{NP} \leq 3$ are formed), and the distribution is unimodal with the peak at $s_{NP} = 2$. This is because of the strong specific repulsion, and the lack of PEIs to form bridges (see **Figure 4.3c, d**). For $\alpha = 4$ and 10, the distribution appears to be bimodal, with peaks at $s_{NP} = 4$ and 10 for $\alpha = 4$, and at $s_{NP} = 5$ and 14 for $\alpha = 10$. The two peaks represent

the formation of moderate ($4 \leq s_{NP} \leq 6$) and large NPs ($s_{NP} \geq 7$), respectively. At these α 's, the specific repulsion is moderate, which at the early stage of the aggregation process allows the aggregation of small NPs to form a few moderate NPs. Afterwards, a small NP would aggregate faster with a moderate NP than another small NP. As a result, a moderate NP would grow in size to form a large NP while new moderate NPs are formed from the aggregation of small NPs. This leads to the bimodal distributions seen in **Figure 4.5b, e**. The two peaks remain separated because the aggregation of a moderate NP and a large NP is unlikely: they both diffuse slowly and have a large net charge (proportional to s_{NP}), which gives rise to large long-range electrostatic repulsion (Eq. 4.1). For $\alpha = 6$ and 8 the distribution is multimodal which resembles a uniform distribution. In these cases, the specific repulsion is weak which enhances the aggregation of small NPs at the early stage of the aggregation process. Consequently, many moderately-sized NPs are formed, with a broader size distribution.

$$\langle s_{NP} \rangle = \frac{\sum_{s_{NP}=1}^{s_{NP}=27} s_{NP} N_{NP}(s_{NP}; t)}{\sum_{s_{NP}=1}^{s_{NP}=27} N_{NP}(s_{NP}; t)} = \frac{N_{DNA}}{\sum_{s_{NP}=1}^{s_{NP}=27} N_{NP}(s_{NP}; t)} \quad 4.2$$

Figure 4.6b shows the average NP size ($\langle s_{NP} \rangle$) at the steady state. $\langle s_{NP} \rangle$ at any time is calculated from Eq. 4.2, where $N_{NP}(s_{NP}; t)$ is the number of NPs at time t that has size s_{NP} . So $\langle s_{NP} \rangle$ is essentially obtained by dividing N_{DNA} by the total number of NPs. Data in **Figure 4.6b** are obtained from the steady state stage of the simulation (last 0.5 μ s), and the error bars arise from the standard deviation in the total number of NPs. The steady state $\langle s_{NP} \rangle$ is lowest for $\alpha = 2$ (strong specific repulsion), moderate for $\alpha = 4$ and 10 (moderate specific repulsion) and highest for $\alpha = 6$ and 8 (weak specific repulsion). Therefore, the specific repulsion not only controls the characteristic of the size distribution but also the steady state $\langle s_{NP} \rangle$.

4.3.3. Rate of NP growth

Figure 4.7a shows the average NP size ($\langle s_{NP} \rangle$) as a function of time for all α , where $\langle s_{NP} \rangle$ was calculated using Eq. 4.2. Upon visual inspection, we notice that $\langle s_{NP} \rangle$ exhibits linear trend over certain time windows. For simplicity, three time windows 0-2, 2-2.7 and 2.7-3.5 μs were chosen over which $\langle s_{NP} \rangle$ was approximated by linear functions. Essentially each curve $\langle s_{NP} \rangle$ in **Figure 4.7a** is fitted with a piecewise linear function which is continuous at 2 and 2.7 μs . Two additional constraints were imposed in the curve fitting. Firstly, $\langle s_{NP} \rangle$ is set to be 1 at $t = 0$, representing initial unaggregated DNAs. Secondly, $\langle s_{NP} \rangle$ at $t = 3.5 \mu\text{s}$ is set to be the steady state $\langle s_{NP} \rangle$ value shown in **Figure 4.6b**. The piecewise linear functions are then determined by least squares fitting and shown as dashed lines in **Figure 4.7a**. The rates of NP growth (r_{NP}), defined as the slopes of the piecewise linear functions, are shown in **Figure 4.7b** for different α and time windows.

The first observation made from **Figure 4.7b** is that the initial rate (0-2 μs) increases as α increases from 2 to 6. This is expected because the number of peripheral PEIs increases with α (see **Figure 4.3b**), which leads to higher probability of converting to bridging PEIs that aggregates DNAs together. The same increase in r_{NP} is observed as α increases from 8 to 10. However, both $\alpha = 8$ and 10 have lower r_{NP} than $\alpha = 6$, and there is a sharp decrease in r_{NP} from $\alpha = 6$ to $\alpha = 8$. At first glance, this is counter-intuitive, especially since the specific repulsion is similar for $\alpha = 6$ and 8. This sharp decrease most likely arises from the difference in the mode of aggregation between $\alpha = 2-6$ and $\alpha = 8-10$. Since $\alpha = 8-10$ leads to branched NPs which are more compact, these NPs are expected to have a smaller radius of sphere of action than $\alpha = 2-6$ for the same s_{NP} ($s_{NP} > 1$). This reduces the chances of collisions between NPs for $\alpha = 8-10$ (branched) when compared to $\alpha = 6$ (linear) and therefore a sharp decrease in r_{NP} is observed between $\alpha = 6$ and 8.

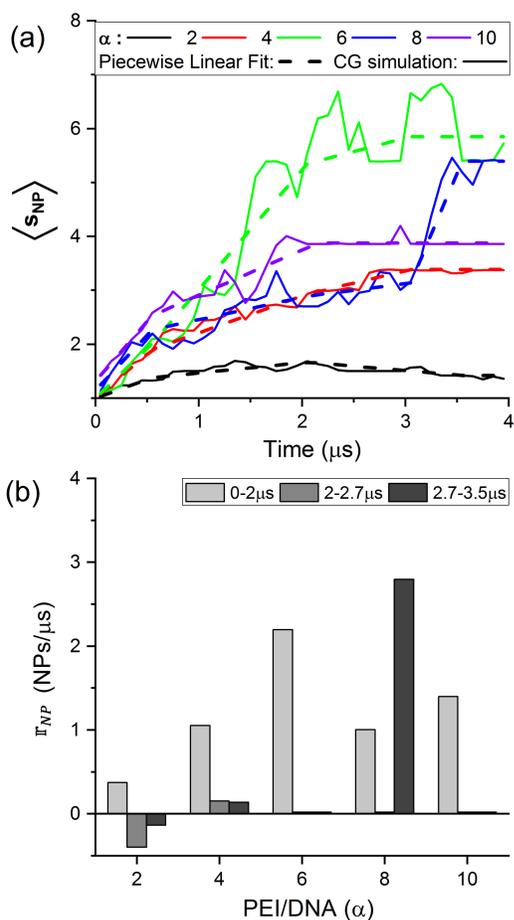


Figure 4.7: Size of NP and rate of NP growth. (a) Average NP size as a function of time. (b) Rate of NP growth calculated as the slope of the curves in (a) in the specified time ranges.

Another observation from **Figure 4.7b** is that except for $\alpha = 8$, r_{NP} decreases with time, i.e., r_{NP} is highest for 0-2 μs followed by 2-2.7 μs , and lowest for 2.7-3.5 μs . This reduction primarily occurs for two reasons. Firstly, as NPs increase in size, their charge increases proportionately (see Eq. 4.1) and as a result, electrostatic repulsion between them increases which reduces r_{NP} . Secondly, as NPs aggregate the total number of NPs in the system decreases which reduces the probability of collision.[62] Although according to SCE, as the size of the largest NP increases, the size ratio between the largest and smallest NPs would increase, which could potentially increase r_{NP} , this effect is secondary in our simulations. Reduction of r_{NP} also occurs

due to other reasons which are specific to different systems. For $\alpha = 2$, the average number of bridging PEIs between a pair of bridged DNAs (n_b) is low (see **Figure 4.3d**), making the aggregate unstable. This instability even makes r_{NP} negative for $t > 2 \mu\text{s}$. For $\alpha = 4$ and 6, the reduction in r_{NP} also occurs due to the instability of loose linear NPs. This can be seen from the relatively low average number of NPs (~ 0) for large NPs in **Figure 4.5**. Another evidence for the instability is the fluctuations in $\langle s_{NP} \rangle$ for $\alpha = 6$ (**Figure 4.7a**). For $\alpha = 10$ as the peripheral PEIs accumulate and surround each DNA it becomes more difficult for the phosphates to be accessible to amines in another NP, which reduces local electrostatic attractions and hence reduces r_{NP} to zero.

The system with $\alpha = 8$ did not show a monotonic relation between r_{NP} and time. Instead, r_{NP} is highest for 2.7-3.5 μs , followed by 0-2 μs and is zero for 2-2.7 μs . Since the specific repulsion is weak for both $\alpha = 8$ and 6 while such a behavior was not observed for $\alpha = 6$, it is most likely associated with the different mode of aggregation. One possible explanation for $r_{NP} = 0$ within 2-2.7 μs is that during this period the compact NPs hardly collide with each other. Instead, they diffuse close to each other and begin to aggregate from 2.7 μs . Unlike $\alpha = 8$, we do not observe $r_{NP} > 0$ for $\alpha = 10$ later in the simulation, which can be due to the greater long-range electrostatic repulsion and the lack of exposed phosphates. The PEI/DNA ratio, therefore, influences the rate of NP growth not only by the long-range electrostatic repulsion but also by the mode of aggregation.

4.3.4. Influence of salt concentration

The aggregation in the presence of 150 mM KCl was qualitatively similar to that in the absence of salts (0 mM KCl), with some minor quantitative differences. To avoid repetition, we will only present key results for $\alpha = 10$, while the rest of the results can be found in **Appendix B.8**. In **Figure 4.8a**, we plot N_f , N_p and N_b as a function of time for 0 and 150 mM KCl systems. N_f

with 150 mM salt is higher for most of the simulation time, implying that the rate of DNA-PEI binding decreases with increase in salt. This can be explained by the competition between PEI and K^+ in binding with the DNA and screening of DNA-PEI attractions. However, in the last 1 μ s of the simulation N_f is similar for both salt concentrations. This is expected because PEIs form more stable binding with DNAs as compared to K^+ . Consistent with the results for N_f , N_p is lower at higher salt concentration. The number of bridging PEIs N_b is practically identical for both salt concentrations, that is, the small changes in N_p do not affect N_b (this is consistent with the discussion in **Appendix B.4**). **Figure 4.8b** shows the average number of bridging PEIs between a pair of bridged DNAs (n_b) as a function of time, which is also hardly affected by salt concentration.

Figure 4.8c plots the NP size distribution (black) and average NP charge as a function of s_{NP} (red) for both 0 and 150 mM KCl. When 150 mM KCl is added, the shape of the NP size distribution does not change and remains bimodal. However, the size of the largest NP increases from $s_{NP} = 14$ to $s_{NP} = 19$. The presence of 150 mM KCl has introduced a screening effect on the electrostatic repulsion between smaller NPs, leading to their aggregation into larger NPs.[7,63] The NP charge in presence of 150 mM KCl also increases linearly with s_{NP} similar to the case of 0 mM KCl. Predicted charge from Eq. 4.1 is shown as the dotted blue line and is in good agreement with simulation data when $s_{NP} < 6$ (similar to $\alpha = 10$ for 0 mM KCl). In **Figure 4.8d**, the average NP size $\langle s_{NP} \rangle$ is plotted as a function of time, which shows similar characteristics for both 0 and 150 mM KCl. The initial slope of $\langle s_{NP} \rangle$ is clearly higher at higher salt concentration, representing a higher rate of NP growth. In fact, $\langle s_{NP} \rangle$ is larger for 150 mM KCl during the entire time of the simulations. Both the increased rate of NP growth and size in the presence of 150 mM KCl can be explained by the screening of long-range electrostatic repulsions between NPs by salt.

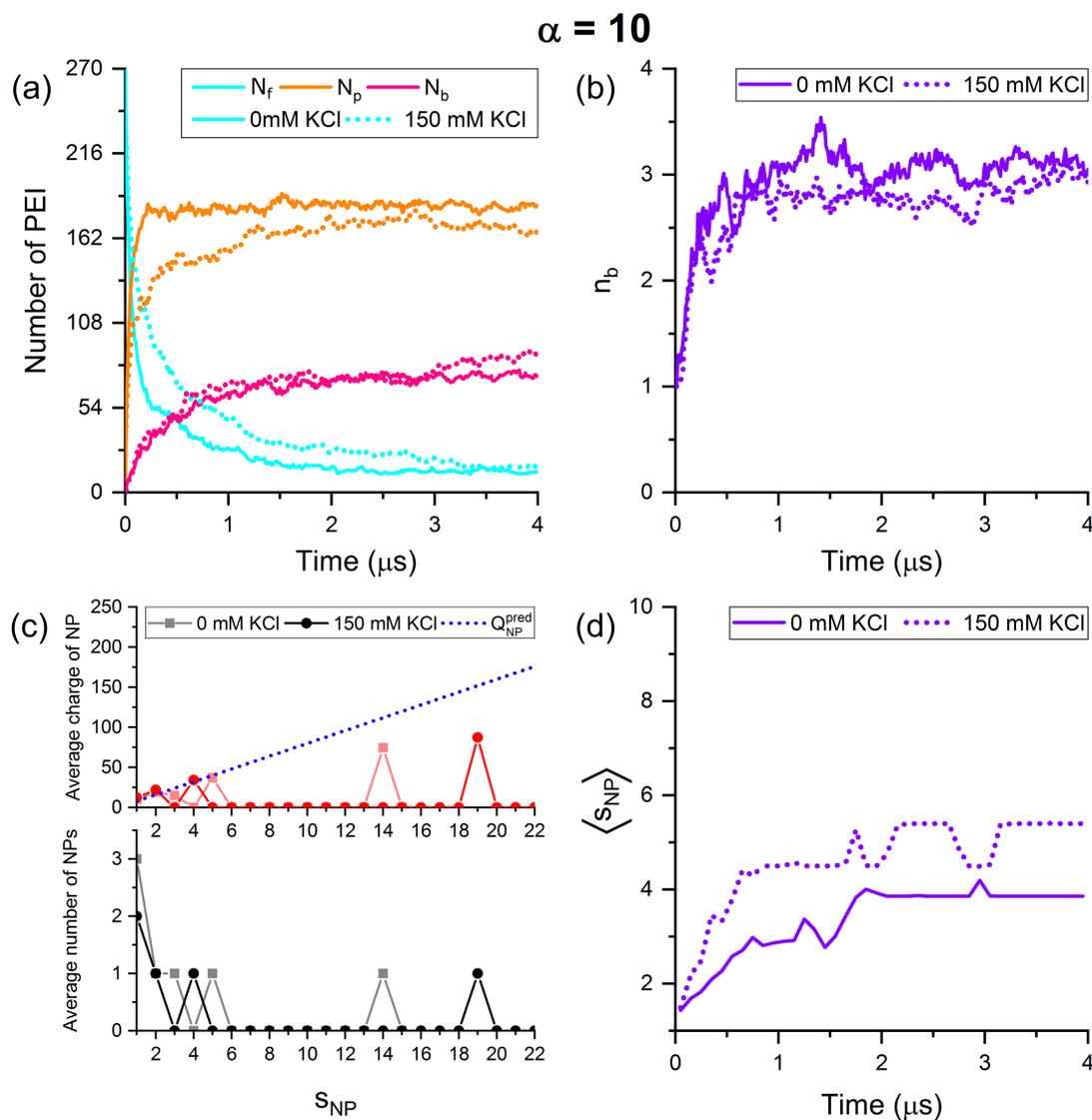


Figure 4.8: Influence of salt concentration on NP aggregation. (a) number of PEIs in different roles; (b) average number of bridging PEIs between a pair of bridged DNAs (n_b); (c) steady state size distribution (black) and average NP charge vs size (red); (d) average NP size as a function of time. The results shown here are only for $\alpha = 10$, and other results can be found in **Appendix B.8**.

4.4. Discussions

Charge of DNA-PEI NPs can be quantified in experiments using the zeta potential. The zeta potential is negative for low N/P ratio and positive for sufficiently large N/P ratio, although the exact N/P ratio to cause zero zeta potential varies for different PEIs and DNAs. Previous work

has reported zero zeta potential at N/P ratio ~ 5 for 600 Da PEI[64], ~ 2.3 for 2 kDa PEI[21], 2.7 to 3.84 for 25 kDa PEI[21,23,64], ~ 3 for 750 kDa PEI[21] and ~ 2.2 for 800 kDa[17] PEI. In our simulations of 586 Da PEIs, the NPs are negative for $N/P < 3.55$ and positive for $N/P > 4.73$ (**Table 4.1**). The location of zero zeta potential can be approximated using linear interpolation, which turns out to be at $N/P \sim 4.4$ (See **Appendix B.9** for details). These results agree well with the experimental findings.

In experiments, the size of NPs is expressed in terms of hydrodynamic radius measured by dynamic light scattering.[21,65] Although the NP size distributions we presented earlier in **Figure 4.5** are in terms of number of DNAs in the NP, calculation of hydrodynamic radius results in distributions that are qualitatively similar (see **Appendix B.10** for details). For 2, 25 and 750 kDa PEIs, Choosakoonkriang *et al.*[21] found the mean hydrodynamic radius to be largest for intermediate N/P ratios at which the zeta potential is nearly zero. This is consistent with our results, where the NP size is largest for $\alpha = 6$ and 8, which has the lowest negative and positive charge respectively. Erbacher *et al.*[23] studied NP size distribution at different N/P ratios using 25 kDa PEI. They observed bimodal size distribution at $N/P = 2$, consistent with the observation made for $\alpha = 4$ ($N/P = 2.36$) in our simulation. Dunlap *et al.*[12] reported several peaks in the NP size distribution for a DNA system over-neutralized by 25 kDa PEI at 150mM NaCl, which is similar to the system with $\alpha = 10$ in our simulation. Aggregation of semi-flexible polyelectrolytes by multivalent salts studied using CG simulations showed similar shapes of the size distribution.[44] Specifically, they observed limited aggregation for highly under-neutralized systems similar to $\alpha = 2$, bimodal distribution for lightly under-neutralized or over-neutralized systems similar to $\alpha = 4$ and 10, and uniform distribution for near-neutral systems similar to $\alpha = 6$ and 8. It should be

noted that these distributions were found when the Bjerrum length is about twice the diameter of the beads in the polyelectrolyte, while the results may be different for other Bjerrum lengths.

On the other hand, there are some differences between experimental studies and our results on the NP size distribution. Unimodal size distribution was observed by Erbacher *et al.*[23] for $N/P = 3$ and 5, where the zeta potential was nearly zero. Our simulation results predict a more or less uniform distribution at these N/P ratios. Since the PEIs in our work have drastically different molecular weight (586 Da as compared to 25 kDa in Erbacher *et al.*[23]), and different size distributions were observed for 22 and 25 kDa PEIs[12], we can expect the NP size distribution to be dependent on the molecular weight of PEI.

Choosakoonkriang *et al.*[21] found DNAs in NPs to have different circular dichroism (CD) spectra for systems with excess PEIs and excess DNAs. They proposed the changes in CD spectra to be caused by changes in base-base interaction of a DNA. Interestingly, we observed different modes of aggregation due to crowding and misalignment of PEIs to the DNA backbone. The crowding of PEIs in the major and minor grooves might alter the base-base interactions in a DNA and in turn the CD spectra.

At low N/P ratio < 1 , plasmid DNA-22 kDa PEI NPs exhibited different shapes such as oblong, folded-loops, toroids and circular structures at 15 mM NaCl, and linear structures with limited branches at 150 mM KCl.[12] At high N/P ratio, rounded globular NPs are observed.[12] These results agree well with this work, where linear structures with limited branches are observed at $\alpha < 8$ (**Appendix B**) and branched spherical-like structures at $\alpha \geq 8$ for 0 and 150 mM KCl. Shape of PEI-oligonucleotide NPs have only been reported for high N/P ratios, where network or spherical structures are observed.[66] These structures are similar to the branched NPs observed at $\alpha \geq 8$.

The 12 base-pair DNA studied here is comparable to oligonucleotides and short-interfering RNAs and micro RNAs which can contain 15-25 base-pairs.[67,68] Since the primary forces of aggregation is long-range electrostatic repulsion and diffusion, the RNA aggregation mechanism is expected to be similar. Aggregation of large DNAs (such as plasmid DNAs) and HMW PEIs is expected to follow a similar mechanism, where aggregation is controlled by diffusion, long-range electrostatic repulsion, and local electrostatic attraction. However, due to large size of molecules, their diffusion coefficient is expected to be lower and long-range electrostatic repulsions and local electrostatic attraction higher.

The different modes of aggregation and the mechanisms behind can allow us to propose a strategy for controlling the hydrodynamic radius and shape of a NP. While an efficient gene delivery would require compaction and aggregation of multiple NAs, NPs with a very large hydrodynamic radius are difficult for cellular uptake.[24] Also, spherical NPs are expected to have higher cellular uptake,[24] so there is a need to generate NPs that have an optimal number of NAs, are sphere-like, compact and stable. A numerical experiment was done to achieve this. In particular, for $\alpha = 4$, a large linear NP was formed as shown in **Figure 4.2b**. Linear NPs larger than this ($S_{NP} > 10$) was unstable, as pointed out in **Section 4.3.2**. At the end of the 4 μ s simulation, 162 PEIs were added to change α of the system to 10. To make the system electroneutral K^+ ions in the $\alpha = 4$ system were removed and an appropriate amount of Cl^- ions were added. NPT simulation was run on this new system for another 3.5 μ s. Comparing the largest NP before (**Figure 4.2b**) and after (**Figure B10**) the simulation, we observe that the NP retains its size (in terms of number of DNAs) but becomes more compact (with reduced hydrodynamic radius). The DNAs in the NP are brought closer by the newly added PEIs through the formation of side-by-side contacts. This implies that compact and stable NPs can be formed using a two-step addition of PEIs. The first step involves

the addition of PEIs so that the DNAs are moderately under-neutralized. Aggregation is expected to occur in an end-to-end fashion forming linear NPs, but the number of DNAs in the aggregate will not be too large due to their instability. In the second step, a large number of PEIs are introduced to bypass the N/P ratio associated with weak specific repulsion to prevent further aggregation. The previously formed linear NPs are then condensed into stable NPs that are more compact and closer to being spherical, with the help of excess PEIs (see **Appendix B.11**). Since the largest NP formed at $4 \mu\text{s}$ for $\alpha = 4$ has smaller s_{NP} than $\alpha = 10$, the NP created using the two-step process would have a lower hydrodynamic radius than the one formed using a single step process.

Finally, we comment on the limitation of this work. As with all MD simulations, the length and time scales of our simulated systems are still small compared with experiments. We have used a low molecular weight (LMW) PEI, 586 Da, while 25 kDa PEI is the most commonly studied PEI in experiments. It should be noted that there are experiments which explored the efficacy of LMW PEIs (600-800 Da).[5,6,69] As well, there are interests in introducing functional modification (e.g., lipid substitution) to LMW PEIs (0.6-2 kDa) to achieve a balance between efficacy and toxicity.[9] To be able to observe the binding and aggregation process within the timeframe permitted by MD simulations, we have used DNA and PEI concentrations that are larger than those in experiments. However, the local concentration of DNAs and PEIs near a NP can be much greater than their concentration in the bulk solution, and we believe it is this local concentration that affects the binding most. Setting a higher concentration in simulation essentially accelerates the diffusion process of the DNAs and PEIs before their binding and aggregation, which would have taken more time to occur in experiments. We also point out that compared with similar works in the literature,[35] our systems are considerably larger, and the investigation on multi-NP formation

allowed us to obtain size distribution from the simulations which has never been attempted before. The good agreement of our results with experiments has shown great potential of using simulations as a predictive tool to assist in the design of delivery systems. One way of extending the length of the simulations is to use larger time steps, e.g., 10 fs instead of 5 fs used in this work. While simulations of Martini PEI[51] and Martini DNA[50] separately in polarizable water[52] were found to be numerically stable for a time step of 10 fs, our initial attempt of using 10 fs for simulating Martini PEI and DNA together in polarizable water failed due to numerical instability. Since large Martini simulations of PEI and DNA together had not been performed before, the reason behind the numerical instability is not yet clear and remains to be explored.

4.5. Conclusions

Large-scale coarse-grained molecular dynamics simulations are performed to study gene delivery nanoparticles formed from the aggregation of DNAs and PEIs. The effects of PEI to DNA number ratio, or equivalently the N/P ratio, are investigated on the shape, charge and size of the nanoparticles. At very low N/P ratio, the aggregation of DNAs is limited, and as the N/P ratio increases the nanoparticles change from a loose linear structure to a compact branched structure, which is closer to a sphere. Such a transition in the mode of aggregation is caused by the different alignments of PEIs with DNA backbones prior to aggregation, which dictates their ability to serve as bridging PEIs. Except for very large nanoparticles at high N/P ratios, the charge of a nanoparticle is proportional to the number of DNAs in it. Their ratio allows for the definition of an intrinsic property called specific repulsion, which controls the characteristics of the steady state size distribution of NPs: unimodal for strong specific repulsion, bimodal for moderate specific repulsion, and more or less uniform for weak specific repulsion. The rate of nanoparticle growth is affected by the N/P ratio through the long-range electrostatic repulsions and mode of aggregation.

Presence of salt does not have any qualitative influence on the formation and characteristics of the nanoparticles. Quantitative, adding salt causes an increase in the rate of nanoparticle growth and their steady state size, due to the screening of electrostatic repulsion between nanoparticles brought by the ions.

4.6. Acknowledgements

We acknowledge the computing resources and technical support from Western Canada Research Grid (WestGrid). Financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) and MITACS Globalink Graduate Fellowship is gratefully acknowledged.

4.7. References

- [1] Boussif O., Lezoualc'h F., Zanta M.A., Mergny M.D., Scherman D., Demeneix B., et al. A Versatile Vector for Gene and Oligonucleotide Transfer into Cells in Culture and In Vivo: Polyethylenimine. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, 92, 7297–301.
- [2] Huang P., Zhao J., Wei C., Hou X., Chen P., Tan Y., et al. Erythrocyte Membrane Based Cationic Polymer-mcDNA Complexes as an Efficient Gene Delivery System. *Biomater. Sci.* **2017**, 5, 120–7.
- [3] Godbey W.T., Barry M.A., Saggau P., Wu K.K., Mikos A.G. Poly(Ethylenimine)-Mediated Transfection: A New Paradigm for Gene Delivery. *J. Biomed. Mater. Res.* **2000**, 51, 321–8.
- [4] Aposhian V.H., Barr S.M., Keck K. Experimental Gene Delivery Systems for Mammalian Cells-The Polyoma Pseudovirus System. *Adv. Enzyme Regul.* **1977**, 16, 275–88.
- [5] Godbey W.T., Wu K.K., Mikos A.G. Size Matters: Molecular Weight Affects the Efficiency of Poly(ethylenimine) as a Gene Delivery Vehicle. *J. Biomed. Mater. Res.* **1999**, 45, 268–75.
- [6] Forrest M.L., Koerber J.T., Pack D.W. A Degradable Polyethylenimine Derivative with Low Toxicity for Highly Efficient Gene Delivery. *Bioconjug. Chem.* **2003**, 14, 934–40.
- [7] Wightman L., Kircheis R., Rössler V., Carotta S., Ruzicka R., Kursu M., et al. Different

- Behavior of Branched and Linear Polyethylenimine for Gene Delivery In Vitro and In Vivo. *J. Gene Med.* **2001**, 3, 362–72.
- [8] Boeckle S., von Gersdorff K., van der Pijpen S., Culmsee C., Wagner E., Ogris M. Purification of Polyethylenimine Polyplexes Highlights the Role of Free Polycations in Gene Transfer. *J. Gene Med.* **2004**, 6, 1102–11.
- [9] Bahadur K.C.R., Landry B., Aliabadi H.M., Lavasanifar A., Uludağ H. Lipid Substitution on Low Molecular Weight (0.6-2.0 kDa) Polyethylenimine Leads to a Higher Zeta Potential of Plasmid DNA and Enhances Transgene Expression. *Acta Biomater.* **2011**, 7, 2209–17.
- [10] Tang M.X., Szoka F.C. The Influence of Polymer Structure on the Interactions of Cationic Polymers with DNA and Morphology of the Resulting Complexes. *Gene Ther.* **1997**, 4, 823–32.
- [11] Utsuno K., Uludağ H. Thermodynamics of Polyethylenimine-DNA Binding and DNA Condensation. *Biophys. J.* **2010**, 99, 201–7.
- [12] Dunlap D.D., Maggi A., Soria M.R., Monaco L. Nanoscopic Structure of DNA Condensed for Gene Delivery. *Nucleic Acids Res.* **1997**, 25, 3095–101.
- [13] Godbey W.T., Wu K.K., Mikos A.G. Tracking the Intracellular Path of Poly(ethylenimine)/DNA Complexes for Gene Delivery. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, 96, 5177–81.
- [14] Bieber T., Meissner W., Kostin S., Niemann A., Elsasser H.-P. Intracellular Route and Transcriptional Competence of Polyethylenimine-DNA Complexes. *J. Control. Release* **2002**, 82, 441–54.
- [15] Verma A., Stellacci F. Effect of Surface Properties on Nanoparticle–Cell Interactions. *Small* **2010**, 6, 12–21.
- [16] Villanueva A., Cañete M., Roca A.G., Calero M., Veintemillas-Verdaguer S., Serna C.J., et al. The Influence of Surface Functionalization on the Enhanced Internalization of Magnetic Nanoparticles in Cancer Cells. *Nanotechnology* **2009**, 20, 115103.
- [17] Ogris M., Brunner S., Schüller S., Kircheis R., Wagner E. PEGylated DNA/Transferrin–PEI Complexes: Reduced Interaction With Blood Components, Extended Circulation in Blood and Potential for Systemic Gene Delivery. *Gene Ther.* **1999**, 6, 595–605.
- [18] Li S., Rizzo M.A., Bhattacharya S., Huang L. Characterization of Cationic Lipid–Protamine–DNA (LPD) Complexes for Intravenous Gene Delivery. *Gene Ther.* **1998**, 5, 930–7.

- [19] Khalil I.A., Kogure K., Akita H., Harashima H. Uptake Pathways and Subsequent Intracellular Trafficking in Nonviral Gene Delivery. *Pharmacol. Rev.* **2006**, 58, 32–45.
- [20] Hillaireau H., Couvreur P. Nanocarriers' Entry Into the Cell: Relevance to Drug Delivery. *Cell. Mol. Life Sci.* **2009**, 66, 2873–96.
- [21] Choosakoonkriang S., Lobo B.A., Koe G.S., Koe J.G., Middaugh C.R. Biophysical Characterization of PEI/DNA Complexes. *J. Pharm. Sci.* **2003**, 92, 1710–22.
- [22] D'Andrea C., Pezzoli D., Malloggi C., Candeo A., Capelli G., Bassi A., et al. The Study of Polyplex Formation and Stability by Time-Resolved Fluorescence Spectroscopy of SYBR Green I-Stained DNA. *Photochem. Photobiol. Sci.* **2014**, 13, 1680–9.
- [23] Erbacher P., Bettinger T., Belguise-Valladier P., Zou S., Coll J.L., Behr J.P., et al. Transfection and Physical Properties of Various Saccharide, Poly(Ethylene Glycol), and Antibody-Derivatized Polyethylenimines (PEI). *J. Gene Med.* **1999**, 1, 210–22.
- [24] Chithrani B.D., Ghazani A.A., Chan W.C.W. Determining the Size and Shape Dependence of Gold Nanoparticle Uptake into Mammalian Cells. *Nano Lett.* **2006**, 6, 662–8.
- [25] Chithrani B.D., Chan W.C.W. Elucidating the Mechanism of Cellular Uptake and Removal of Protein-Coated Gold Nanoparticles of Different Sizes and Shapes. *Nano Lett.* **2007**, 7, 1542–50.
- [26] Zhang S., Gao H., Bao G. Physical Principles of Nanoparticle Cellular Endocytosis. *ACS Nano* **2015**, 9, 8655–71.
- [27] Ding H.M., Ma Y.Q. Theoretical and Computational Investigations of Nanoparticle-Biomembrane Interactions in Cellular Delivery. *Small* **2015**, 11, 1055–71.
- [28] Kettler K., Veltman K., van de Meent D., van Wezel A., Hendriks A.J. Cellular Uptake of Nanoparticles as Determined by Particle Properties, Experimental Conditions, and Cell Type. *Environ. Toxicol. Chem.* **2014**, 33, 481–92.
- [29] Sun C., Tang T., Uludağ H., Cuervo J.E. Molecular Dynamics Simulations of DNA/PEI Complexes: Effect of PEI Branching and Protonation State. *Biophys. J.* **2011**, 100, 2754–63.
- [30] Sun C., Tang T. Study on the Role of Polyethylenimine as Gene Delivery Carrier Using Molecular Dynamics Simulations. *J. Adhes. Sci. Technol.* **2014**, 28, 399–416.
- [31] Sun C., Tang T., Uludağ H. Molecular Dynamics Simulations for Complexation of DNA with 2 kDa PEI Reveal Profound Effect of PEI Architecture on Complexation. *J. Phys. Chem. B* **2012**, 116, 2405–13.

- [32] Kondinskaia D.A., Kostritskii A.Y., Nesterenko A.M., Antipina A.Y., Gurtovenko A.A. Atomic-Scale Molecular Dynamics Simulations of DNA-Polycation Complexes: Two Distinct Binding Patterns. *J. Phys. Chem. B* **2016**, 120, 6546–54.
- [33] Ziebarth J., Wang Y. Molecular Dynamics Simulations of DNA-Polycation Complex Formation. *Biophys. J.* **2009**, 97, 1971–83.
- [34] Ziebarth J., Wang Y. Coarse-Grained Molecular Dynamics Simulations of DNA Condensation by Block Copolymer and Formation of Core - Corona Structures. *J. Phys. Chem. B* **2010**, 114, 6225–32.
- [35] Wei Z., Luijten E. Systematic Coarse-Grained Modeling of Complexation Between Small Interfering RNA and Polycations. *J. Chem. Phys.* **2015**, 143, 243146.
- [36] Zhan B., Shi K., Dong Z., Lv W., Zhao S., Han X., et al. Coarse-Grained Simulation of Polycation/DNA-Like Complexes: Role of Neutral Block. *Mol. Pharm.* **2015**, 12, 2834–44.
- [37] Sun C., Tang T., Uludağ H. Molecular Dynamics Simulations of PEI Mediated DNA Aggregation. *Biomacromolecules* **2011**, 12, 3698–707.
- [38] Bagai S., Sun C., Tang T. Potential of Mean Force of Polyethylenimine-Mediated DNA Attraction. *J. Phys. Chem. B* **2013**, 117, 49–56.
- [39] Tom A.M., Rajesh R., Vemparala S. Aggregation Dynamics of Rigid Polyelectrolytes. *J. Chem. Phys.* **2016**, 144, 034904.
- [40] Lee K.-C., Borukhov I., Gelbart W.M., Liu A.J., Stevens M.J. Effect of Mono- and Multivalent Salts on Angle-Dependent Attractions Between Charged Rods. *Phys. Rev. Lett.* **2004**, 93, 128101.
- [41] Stevens M.J. Bundle Binding in Polyelectrolyte Solutions. *Phys. Rev. Lett.* **1999**, 82, 101–4.
- [42] Grønbech-Jensen N., Mashl R.J., Bruinsma R.F., Gelbart W.M. Counterion-Induced Attraction between Rigid Polyelectrolytes. *Phys. Rev. Lett.* **1997**, 78, 2477–80.
- [43] Fazli H., Golestanian R. Aggregation Kinetics of Stiff Polyelectrolytes in the Presence of Multivalent Salt. *Phys. Rev. E* **2007**, 76, 041801.
- [44] Sayar M., Holm C. Equilibrium Polyelectrolyte Bundles With Different Multivalent Counterion Concentrations. *Phys. Rev. E* **2010**, 82, 031901.
- [45] Sayar M., Holm C. Finite-Size Polyelectrolyte Bundles at Thermodynamic Equilibrium. *Europhys. Lett.* **2007**, 77, 16001.
- [46] Diehl A., Carmona H.A., Levin Y. Counterion Correlations and Attraction Between Like-

- Charged Macromolecules. *Phys. Rev. E* **2001**, 64, 011804.
- [47] Fazli H., Mohammadinejad S., Golestanian R. Salt-Induced Aggregation of Stiff Polyelectrolytes. *J. Phys. Condens. Matter* **2009**, 21, 424111.
- [48] Tom A.M., Rajesh R., Vemparala S. Aggregation of Flexible Polyelectrolytes: Phase Diagram and Dynamics. *J. Chem. Phys.* **2017**, 147, 144903.
- [49] Ogris M., Steinlein P., Kursa M., Mechtler K., Kircheis R., Wagner E. The Size of DNA/Transferrin-PEI Complexes Is an Important Factor for Gene Expression in Cultured Cells. *Gene Ther.* **1998**, 5, 1425–33.
- [50] Uusitalo J.J., Ingólfsson H.I., Akhshi P., Tieleman D.P., Marrink S.J. Martini Coarse-Grained Force Field: Extension to DNA. *J. Chem. Theory Comput.* **2015**, 11, 3932–45.
- [51] Mahajan S., Tang T. Martini Coarse-Grained Model for Polyethylenimine. *J. Comput. Chem.* **2019**, 40, 607–18.
- [52] Yesylevskyy S.O., Schäfer L. V., Sengupta D., Marrink S.J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comput. Biol.* **2010**, 6, e1000810.
- [53] Marrink S.J., de Vries A.H., Mark A.E. Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B* **2004**, 108, 750–60.
- [54] Abraham M.J., Murtola T., Schulz R., Páll S., Smith J.C., Hess B., et al. GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism From Laptops to Supercomputers. *SoftwareX* **2015**, 1–2, 19–25.
- [55] Marrink S.J., Tieleman D.P. Perspective on the Martini model. *Chem. Soc. Rev.* **2013**, 42, 6801–22.
- [56] Hess B., Bekker H., Berendsen H.J.C., Fraaije J.G.E.M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, 18, 1463–72.
- [57] Bussi G., Donadio D., Parrinello M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, 126, 014101.
- [58] Berendsen H.J.C., Postma J.P.M., van Gunsteren W.F., DiNola A., Haak J.R. Molecular Dynamics With Coupling to an External Bath. *J. Chem. Phys.* **1984**, 81, 3684–90.
- [59] Parrinello M., Rahman A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, 52, 7182–90.
- [60] Verlet L. Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, 159, 98–103.
- [61] Tironi I.G., Sperb R., Smith P.E., van Gunsteren W.F. A Generalized Reaction Field

- Method for Molecular Dynamics Simulations. *J. Chem. Phys.* **1995**, 102, 5451–9.
- [62] von Smoluchowski M. Versuch Einer Mathematischen Theorie der Koagulationskinetik kolloider Lösungen. *Zeitschrift f. Physik. Chemie.* **1917**, 92, 129–68.
- [63] Goula D., Remy J.S., Erbacher P., Wasowicz M., Levi G., Abdallah B., et al. Size, Diffusibility and Transfection Performance of Linear PEI/DNA Complexes in the Mouse Central Nervous System. *Gene Ther.* **1998**, 5, 712–7.
- [64] Tang G.P., Guo H.Y., Alexis F., Wang X., Zeng S., Lim T.M., et al. Low Molecular Weight Polyethylenimines Linked by β -Cyclodextrin for Gene Transfer Into the Nervous System. *J. Gene Med.* **2006**, 8, 736–44.
- [65] Putnam D., Gentry C.A., Pack D.W., Langer R. Polymer-Based Gene Delivery With Low Cytotoxicity by a Unique Balance of Side-Chain Termini. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, 98, 1200–5.
- [66] dos Santos A.L.G., Bochot A., Tsapis N., Artzner F., Bejjani R.A., Thillaye-Goldenberg B., et al. Oligonucleotide–Polyethylenimine Complexes Targeting Retinal Cells: Structural Analysis and Application to Anti-TGFbeta-2 Therapy. *Pharm. Res.* **2006**, 23, 770–81.
- [67] Wojnilowicz M., Glab A., Bertucci A., Caruso F., Cavaliere F. Super-resolution Imaging of Proton Sponge-Triggered Rupture of Endosomes and Cytosolic Release of Small Interfering RNA. *ACS Nano* **2019**, 13, 187–202.
- [68] Rehman Z.U., Hoekstra D., Zuhorn I.S. Mechanism of Polyplex- and Lipoplex-Mediated Delivery of Nucleic Acids: Real-Time Visualization of Transient Membrane Destabilization Without Endosomal Lysis. *ACS Nano* **2013**, 7, 3767–77.
- [69] Godbey W.T., Wu K.K., Mikos A.G. Poly(Ethylenimine)-Mediated Gene Delivery Affects Endothelial Cell Function and Viability. *Biomaterials* **2001**, 22, 471–80.

Chapter 5

Endosomal Acidification of PEI-DNA Nanoparticles

5.1 Introduction

Gene therapy treats diseases by introducing genetic material such as DNA into malignant cells.[1] For a successful delivery, DNAs have to overcome several barriers such as cell association, cellular uptake, endosomal escape, and nuclear trafficking.[2–4] Specialized carriers are required to overcome these barriers since DNAs, by themselves, are susceptible to enzyme degradation along the delivery pathway.[5,6] Viral carriers such as adenovirus are often used because their cell-proliferating molecular machinery has been honed by evolution, making them highly efficient.[7] This is, however, accompanied by the risk of immunogenicity which needs to be managed.[8] One solution is to use synthetically manufactured non-viral carriers, whose efficacy can be tweaked by changing its molecular structure, mass, chemical composition, etc.[2] Polyethylenimine (PEI) has stood out as one of the most studied non-viral carriers due to its high efficacy and versatility.[9]

PEI-based DNA delivery follows an endocytic pathway, where the PEI-DNA nanoparticles (NPs) are trapped in early endosomes after cellular uptake.[10] Early endosomes are continuously acidified by proton pumps (H^+ -ATPase), leaving the DNAs vulnerable to acidic digestion.[11,12] PEI's high efficacy arises from its ability to protect DNAs by absorbing protons[3] and escape from endosomes with the DNAs.[13] The endosomal escape mechanism is not fully understood but is widely accredited to the “proton sponge” hypothesis.[13,14] Under this hypothesis, PEIs act as a buffer to absorb protons (like a sponge), which results in more protons

pumped into the endosome. Consequently, there is an electrostatics-driven entry of counterions[15] (such as chloride) and osmotic entry of water.[13,14] Such influxes cause endosomal swelling, and if substantial, it can lead to endosomal burst or localized rupture which release all the NPs into the cytoplasm.[13,16,17] Other proposed endosomal escape mechanisms are associated with PEI's membrane damaging properties.[10,18]

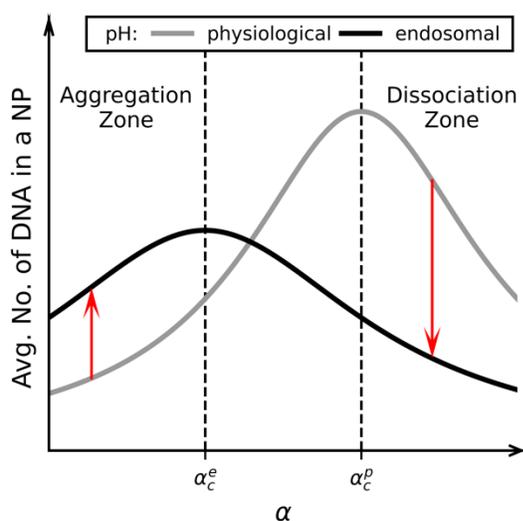


Figure 5.1: Schematic representation of the average number of DNAs in a NP vs. PEI/DNA number ratio (α) under physiological (grey) and endosomal (black) pH. Dashed vertical lines mark the critical α values at which the average number of DNAs in a NP is maximized: (left) under endosomal pH; (right) under physiological pH.

Theoretical studies of endosomal acidification have been limited, with only two modeling works[19,20] investigating the “proton sponge” hypothesis. Additionally, the experimental work by Vermeulen *et al.*[16] proposed an endosomal burst criterion based on its size. As endosomal acidification is expected to alter the morphology of endosomes (swell, rupture, burst, etc.), the same is possible for NPs. For example, upon endosomal acidification, decondensation of plasmid DNAs[17,21] and dissociation of oligonucleotides from NPs[17] have been reported based on fluorescence microscopy measurements. The detailed mechanism involved in such processes is

unclear due to the spatio-temporal resolution limit of fluorescence microscopy.[22,23] Furthermore, the location where genetic materials are released from the NPs is not well agreed upon and has been proposed to occur in the endosome[17], in the cytoplasm[21], or not occur at all[6,24] (i.e., NPs stay intact in the nucleus). This warrants a theoretical study exploring the effects of endosomal acidification on NP properties.

Further motivation comes from **Chapter 4**[25] where PEI-DNA aggregation under physiological pH was simulated for different PEI/DNA number ratios α (proportional to the N/P ratio which is the ratio of PEI nitrogen to DNA phosphate). The results hinted that endosomal acidification might alter NP properties depending on the α value at which the NPs were prepared. In particular, aggregation of NPs was promoted by diffusion and hindered by the *specific repulsion* (electrostatic repulsion between two NPs, where the net charge of each NP is normalized by the number of DNAs in it), which scales quadratically with α (**Chapter 4**[25]). For low and high α , the specific repulsion between NPs was high leading to the formation of small NPs, while large NPs were formed at moderate α . At a critical α (α_c), where the DNAs were just neutralized by the PEIs, the specific repulsion was the lowest and NP aggregation (quantified by the average number of DNAs in the NPs) the highest; as shown schematically in **Figure 5.1** (grey curve). We hypothesize that similar phenomenon exists at endosomal pH (**Figure 5.1**, black curve), and that α_c at endosomal pH (α_c^e) is less than that at physiological pH (α_c^p) because PEIs are more protonated at endosomal pH and hence require smaller α to neutralize the DNAs. According to this hypothesis, NPs prepared at physiological pH with $\alpha < \alpha_c^e$ would aggregate further when placed in the endosomal environment (upward arrow; **Figure 5.1**), and conversely would dissociate when prepared with $\alpha > \alpha_c^p$ (downward arrow; **Figure 5.1**). The present work explores

the validity of this hypothesis, which we refer to as the *acidic resizing* of NPs, and its implications to gene delivery.

5.2 Results

5.2.1 Acidic resizing

Four aggregation simulations (see **Section 5.5.1** for simulation details) were performed, at physiological or endosomal pH, and with $\alpha = 2$ or 10. $\alpha = 2$ is less than α_c^e , and $\alpha = 10$ is greater than α_c^p (**Table 5.1**). A PEI-DNA pair was considered bound if their minimum distance was below 0.528 nm, the van der Waals diameter between their beads. A NP was defined by a set of PEIs and DNAs bound together (directly or indirectly through other PEIs and DNAs), and the number of DNAs bound together (directly or indirectly through other PEIs and DNAs), and the number of DNAs present in the NP was used to quantify the NP size s_{NP} . The average NP size $\langle s_{NP} \rangle$ at any time was obtained by averaging s_{NP} across all NPs, while the steady state average size $\langle s_{NP} \rangle_{ss}$, (**Figure 5.2a**) was obtained from the last 1 μ s of each simulation (see total simulation time in **Table 5.1**). As shown in **Figure 5.2a**, for $\alpha = 2$, $\langle s_{NP} \rangle_{ss}$ is smaller at physiological pH than at endosomal pH, while the result is opposite for $\alpha = 10$. The results support the acidic resizing of NPs proposed in **Figure 5.1**. Other results of the aggregation simulations are presented in **Appendix C.1**.

To explain the acidic resizing, the free energy landscape of PEI was calculated at the steady state (**Figure 5.2b, c**) along a reaction coordinate q (**Section 0**) by setting its lowest value to zero. q is in the range of: (i) 0-1 when PEI is *free*, i.e., not bound to any DNA, (ii) 1-2.2 when PEI is on the periphery of a NP or *peripheral*, i.e., bound to only one DNA, (iii) 2.2-3.4 when PEI is *2-bridging*, i.e., bridging two DNAs, (iv) 3.4-4.6 when PEI is *3-bridging*, and (v) 4.6-5.8 when PEI is *4-bridging*. The term *bridging PEI* collectively refers to all n -bridging PEIs, where n is an integer greater than 1.

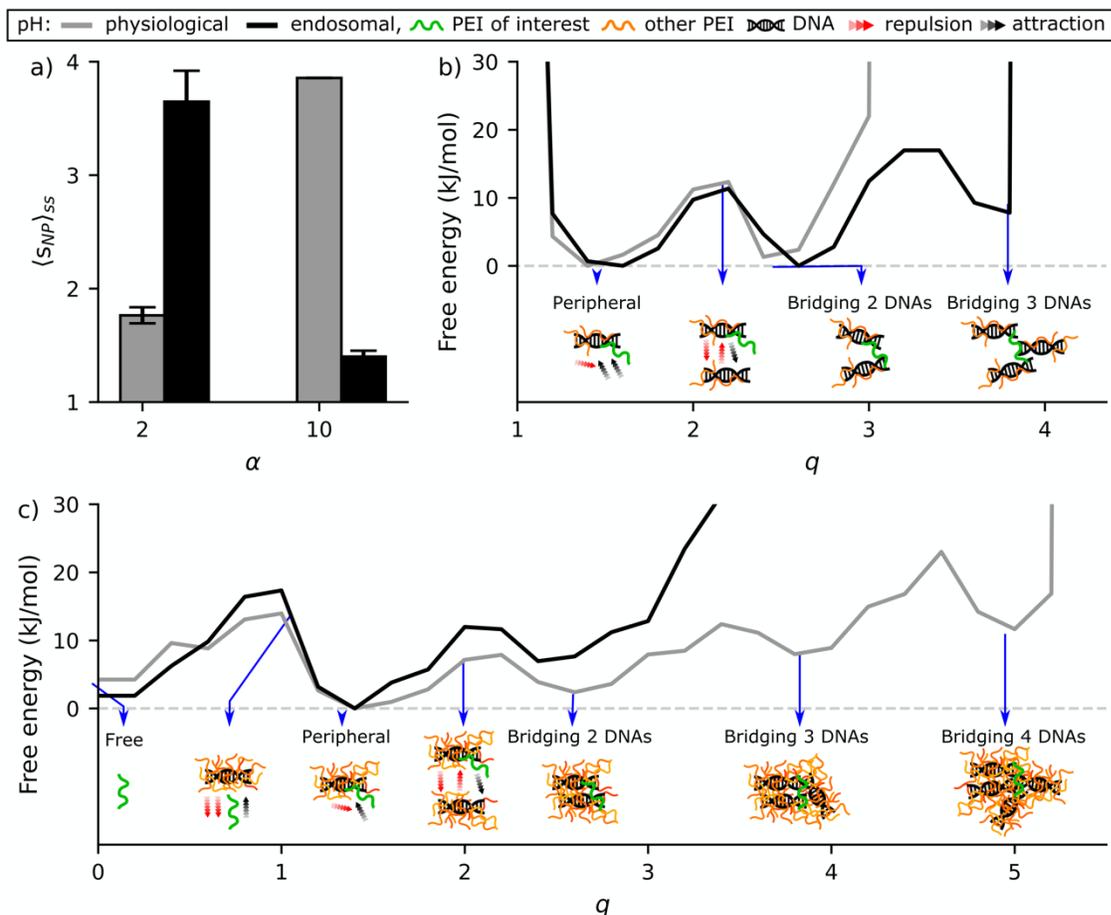


Figure 5.2: Properties of aggregated NPs in physiological and endosomal pH. (a) Steady state average NP size $\langle s_{NP} \rangle_{ss}$ for aggregation simulations at physiological and endosomal pH. Free energy landscape of PEI at physiological and endosomal pH along reaction coordinate q (see Methods **Section 0**) for (b) $\alpha = 2$, and (c) $\alpha = 10$ at the steady state. Last 1 μ s of each simulation is considered as the steady state (**Table 5.1**).

The free energy landscapes in **Figure 5.2b, c** contain local minima representing stable equilibrium states, such as $q = 0, 1.5, 2.5, 3.8,$ and 5 for free, peripheral, 2-bridging, 3-bridging, and 4-bridging PEIs respectively. Two consecutive stable states are separated by a maximum, representing an unstable equilibrium state. Analogous to a reversible chemical reaction, the rate of transition from one stable state to another is regulated by the energy barrier in between. The forward energy barrier associated with increasing q arises due to two repulsive forces. The first is

the repulsive force between a PEI and those bound to the DNA it is approaching. This is referred to as *PEI-PEI repulsion* which increases with the number of PEIs (i.e. α) in the system and the degree of protonation. The second is the repulsive force between the DNAs bound to a PEI and the DNA the PEI is approaching. This repulsion is referred to as *DNA-DNA repulsion* and is proportional to the specific repulsion(**Chapter 4**[25]) and the number DNAs bound to the PEI (which is n for an n -bridging PEI). The backward energy barrier associated with decreasing q arises due to the attraction between a PEI and the DNA it is approaching, which is referred to as *PEI-DNA attraction* and increases with the degree of protonation of PEI. Conformational changes in the NP such as relative position of DNAs and PEIs can also modulate these attractive and repulsive forces.

For $\alpha = 2$ and physiological pH, the energy barriers at $q = 1.1$ and 3 are very high, denoting the absence of free and 3-bridging PEIs. As pH is decreased from physiological to endosomal condition, PEI-PEI repulsion and PEI-DNA attraction are strengthened due to the increase in PEI charge, and DNA-DNA repulsion is weakened due to the decrease in specific repulsion (DNAs are more neutralized). The strengthening of PEI-PEI repulsion is not significant at such low α . As a result, the energy barrier at $q = 1.1$ remains high, preventing the appearance of free PEIs. Meanwhile, the increased PEI-DNA attraction and reduced DNA-DNA repulsion have lowered the energy barrier at $q = 3$, allowing the formation of stable 3-bridging PEIs and leading to further aggregation of NPs.

For $\alpha = 10$ and physiological pH, there is a large energy barrier at $q = 5.2$, denoting the absence of stable 5-bridging PEIs. Energy barrier near $q = 1$ is not very high and free PEIs are present. Decreasing the pH to endosomal level strengthens PEI-PEI repulsion and PEI-DNA attraction. The DNA-DNA repulsion is also strengthened due to the increase in specific repulsion

(DNAs are more over-neutralized). At this large α , the increase of PEI-PEI and DNA-DNA repulsions are significant, causing a large energy barrier to appear at $q = 3.1$. Consequently, the free energy minima at $q = 3.8$ and 5 , present for physiological pH, have disappeared for endosomal pH. In addition, while the free energy landscape remains similar for q between 0 and 1.5 , at endosomal pH the free energy for $q > 1.5$ is higher than that in physiological pH. All these observations support the conclusion that if the NPs are prepared with $\alpha = 10$, upon endosomal acidification, they tend to dissociate into smaller NPs, with PEIs bridging a smaller number of DNAs and preferring to be peripheral. The acidification simulations allow us to have a closer and more direct look at the resizing of NPs during the acidification process.

5.2.2 Acidification of NPs prepared with $\alpha = 2$

Endosomal acidification was simulated for two acidification rates, instant and slow ($1 \text{ H}^+/\text{ns}$). The initial configuration of acidification simulations was taken from the last timestep of the aggregation simulation performed under physiological pH for the same α . The charge of PEIs was increased to model endosomal acidification (see **Section 5.5.1.2** for details). To facilitate the discussions below, the acidification simulations are referred to as *slow* or *instant acidification simulations*, whereas the aggregation simulations are simply referred to as *physiological* or *endosomal pH simulations*.

Figure 5.3a, b shows the *in-silico* fluorescence microscopy images (**Section 5.5.3**) of the NPs during acidification, where red and green represent fluorescence of DNA and PEI beads respectively. The color between red and green is produced based on the ratio of DNA and PEI fluorescence, as shown in the colormap. Bright colors indicate particles are in-focus, whereas diffused colors indicate they are out-of-focus. Visually examining **Figure 5.3a, b**, larger NPs are observed after $8 \mu\text{s}$. Consistent with **Section 5.2.1**, no free PEI (pure green fluorescence) is

observed in the solution. Quantitatively, the average NP size $\langle s_{NP} \rangle$ (Figure 5.3c) demonstrates similar results. Further aggregation of NPs is observed at both acidification rates as indicated by an increase in $\langle s_{NP} \rangle$ with time. Steady state average size $\langle s_{NP} \rangle_{ss}$ for slow acidification simulation is comparable to that in endosomal pH simulation (shown in black) and higher than that in instant acidification simulation. That is, $\langle s_{NP} \rangle_{ss}$ increases with the decrease in acidification rate.

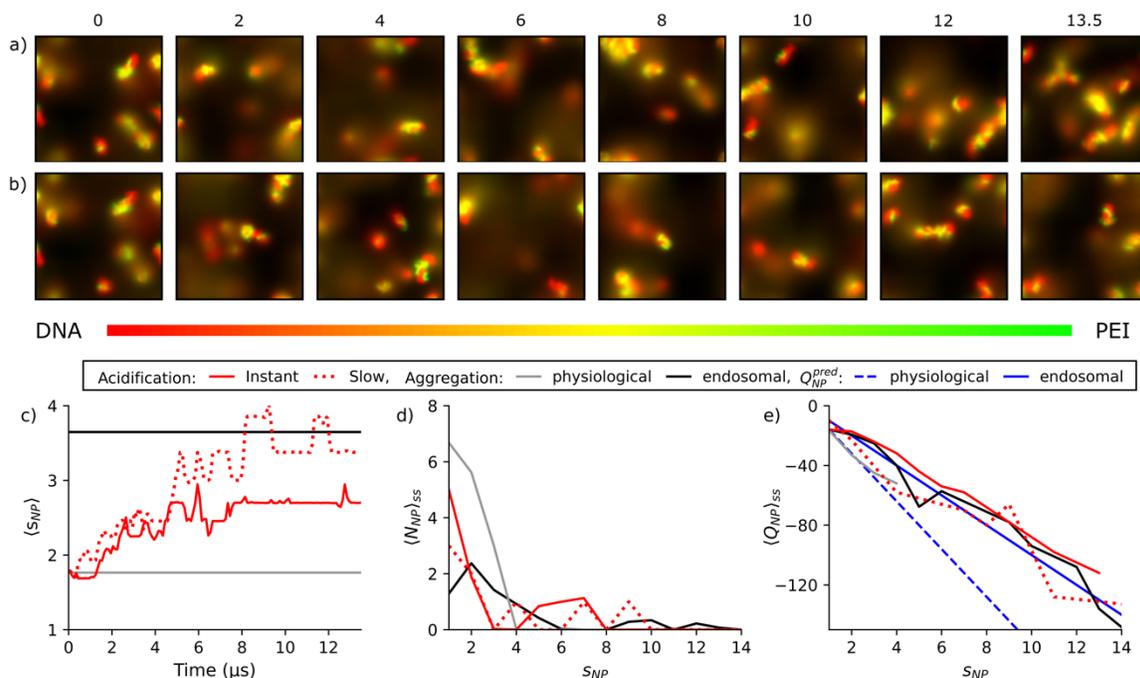


Figure 5.3: Properties of NPs under endosomal acidification for $\alpha = 2$. In-silico microscopy images of (a) instant and (b) slow acidifications, where DNA and PEI emit red and green colors respectively. The colocalization of DNA and PEI produces color between red and green, as shown in the colormap. Simulation time is reported above each image in microseconds. (c) Time evolution of average NP size $\langle s_{NP} \rangle$ for slow and instant acidification simulations, average performed across all NPs. (d) The steady state number of NPs $\langle N_{NP} \rangle_{ss}$ vs. size s_{NP} representing NP size distribution, average performed over time during the steady state. (e) The steady state charge of NPs $\langle Q_{NP} \rangle_{ss}$ vs. s_{NP} , averaged across NPs with the same size s_{NP} and over time during the steady state. The predicted NP charge Q_{NP}^{pred} is calculated from steady state of aggregation simulations using Eq. 5.1. The last 1 μ s of each simulation is considered as the steady state (Table 5.1, Table 5.2).

The steady state number of NPs with size s_{NP} , $\langle N_{NP} \rangle_{ss}$, is plotted against s_{NP} in **Figure 5.3d**, which describes the steady state size distribution of the NPs. $\langle N_{NP} \rangle_{ss}$ exhibits similar characteristics for both acidification rates, with slow acidification forming larger NPs and instant acidification having more unaggregated DNAs. $\langle N_{NP} \rangle_{ss}$ from endosomal pH simulation is significantly different, forming larger NPs containing as many as 12 DNAs. According to the Smoluchowski coagulation theory, the likelihood of two NPs aggregating together increases with the difference in their size (**Chapter 4**[25,26]). This makes $\langle N_{NP} \rangle_{ss}$ highly dependent on the initial size distribution of NPs in the simulation. Endosomal pH simulation begins with only unaggregated DNAs, which first aggregates to form a few small NPs ($s_{NP} = 2$ or 3). These NPs preferentially bind with unaggregated DNAs ($s_{NP} = 1$) to grow into moderate NPs ($4 \leq s_{NP} \leq 9$). The moderate NPs then preferentially bind with unaggregated DNAs and grow into large NPs ($s_{NP} \geq 10$). The probability of small NPs aggregating with similarly sized small NPs to form stable moderate NPs is low and as a result, in endosomal pH simulation, the NPs tend to be either large or small. The initial size distribution in the acidification simulations is $\langle N_{NP} \rangle_{ss}$ from the physiological pH simulation (**Figure 5.3d**), with several small NPs and unaggregated DNAs. Upon instant acidification, all the PEIs are protonated at the beginning of the simulation, which decreases the specific repulsions of NPs all at once. As a result, all small NPs compete to bind with unaggregated DNAs and grows simultaneously into moderate NPs. Due to the concurrent competition, it is difficult for moderate NPs to bind with unaggregated DNA in order to grow into large NPs. In slow acidification simulation, the PEIs are protonated sequentially, and therefore the specific repulsion of some NPs decreases earlier. These NPs therefore have the advantage of being able to grow first into larger NPs amid the existing competition. Therefore, the NPs resized during

slow acidification are larger than those in instant acidification but smaller than those in endosomal pH simulation.

The steady state charge of NPs with size s_{NP} , $\langle Q_{NP} \rangle_{ss}$, is plotted against s_{NP} in **Figure 5.3e**. The NP charges at physiological and endosomal pH are predicted (Q_{NP}^{pred}) using Eq 5.1 and plotted in **Figure 5.3e**, where Q_{DNA} is the charge of a DNA, Q_{PEI} is the charge of a PEI, and α' is the ratio of the total number of bound PEIs and DNAs (**Chapter 4**[25]). This prediction assumes that all DNAs are bound to an equal number of PEIs. All $\langle Q_{NP} \rangle_{ss}$ decreases with s_{NP} in an almost linear fashion, which agrees with Eq 5.1. Therefore, each DNA is bound to a similar number of PEIs before and after acidification. Since free PEIs are absent before and after acidification (**Figure 5.1b**; **Figure 5.3a, b**), it can be inferred that the number of PEIs bound to each DNA remains constant during acidification.

$$Q_{NP}^{pred} = s_{NP}(Q_{DNA} + \alpha'Q_{PEI}) \quad 5.1$$

The above results confirm that NPs prepared at low α further aggregate upon endosomal acidification, and that the acidic resizing hypothesis is valid regardless of the acidification rate. Of more interest is the examination on the acidification of NPs prepared at high α , which according to the hypothesis, can be subjected to dissociation.

5.2.3 Acidification of NPs prepared with $\alpha = 10$

In-silico fluorescence microscopy images generated for slow and instant acidification of NPs prepared with $\alpha = 10$ (**Figure 5.4a, b** respectively) show that green fluorescence in the solution increases with time, indicating that bound PEIs (peripheral or bridging) become free upon acidification. This is further supported by the change in color of NPs from yellow to orange, corresponding to reduced PEI fluorescence in the NPs. Along with the increase in free PEIs, NPs are observed to become smaller, suggesting NP dissociation. Many free PEIs are observed at 0.2

μs in **Figure 5.4a**, whereas similar density of free PEIs is observed in **Figure 5.4b** at a later time (1 μs). That is, the release of free PEI from the NPs is faster upon instant acidification than during slow acidification. Similarly, NP dissociation starts at 0.2 μs in **Figure 5.4a** and later (0.8 μs) in **Figure 5.4b**. After 3 μs , the difference between the two acidification rates appears minimal. These observations are quantified by $\langle s_{NP} \rangle$ in **Figure 5.4c**, which again shows that NP dissociation occurs for both acidification rates and is faster under instant acidification. The steady state NP size, $\langle s_{NP} \rangle_{SS}$, is comparable for acidification and endosomal pH simulations. Interestingly, the onset of dissociation in slow acidification simulation, when $\langle s_{NP} \rangle$ starts decreasing, is comparable to the total acidification time of 0.81 μs (**Table 5.2**). In other words, NP dissociation does not occur until all the PEIs in the system are acidified.

The steady state NP size distribution $\langle N_{NP} \rangle_{SS}$ shown in **Figure 5.4d**, is similar for acidification and endosomal pH simulations, with slow acidification forming a lower number of unaggregated DNAs ($s_{NP} = 1$). That is, instant acidification is more efficient in dissociating NPs than slow acidification. This can be explained by a sudden increase in PEI-PEI and DNA-DNA repulsions when the instant acidification is applied, which makes the NPs unstable and prone to dissociation. In contrast, during slow acidification newly protonated PEIs can adjust their conformations and attach to the NPs via local attractions with the DNAs, before other PEIs are protonated. Such conformational changes help reduce repulsive forces within the NPs and delay NP dissociation. Similar to **Figure 5.3**, the steady state NP charges $\langle Q_{NP} \rangle_{SS}$ in **Figure 5.4e** scale linearly with s_{NP} and are well predicted by Eq 5.1. This implies that after acidification, each DNA in the system is bound to a similar number of PEIs. Since bound PEIs are lost during acidification (**Figure 5.4a, b**), the result suggests that each DNA loses a similar number of PEIs to the solution.

Understanding how PEIs, especially the bridging ones, are lost is crucial because it controls the structural and geometrical changes of NPs, and hence if and how the NPs dissociate.

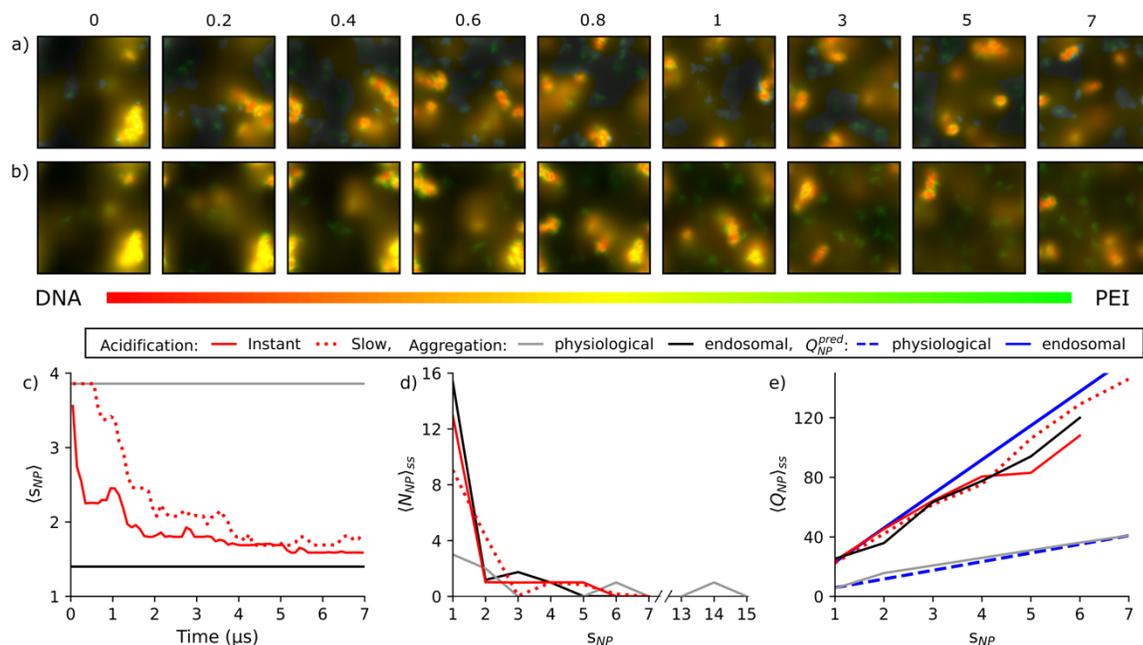


Figure 5.4: Properties of NPs under endosomal acidification for $\alpha = 10$. In-silico microscopy images of (a) instant and (b) slow acidifications, where DNA and PEI emit red and green colors respectively. The colocalization of DNA and PEI produces color between red and green, as shown in the colormap. Simulation time is reported above each image in microseconds. (c) Time evolution of average NP size $\langle s_{NP} \rangle$ for slow and instant acidification simulations, average performed across all NPs. (d) The steady state number of NPs $\langle N_{NP} \rangle_{SS}$ vs. size s_{NP} representing NP size distribution, average performed over time during the steady state. (e) The steady state charge of NPs $\langle Q_{NP} \rangle_{SS}$ vs. s_{NP} , averaged across NPs with the same size s_{NP} and over time during the steady state. The predicted NP charge Q_{NP}^{pred} is calculated from steady state of aggregation simulations using Eq. 5.1. The last 1 μ s of each simulation is considered as the steady state (**Table 5.1, Table 5.2**).

5.2.3.1 Kinetics of bridging PEIs

The numbers of free (N_f/N_{DNA}), peripheral (N_p/N_{DNA}), and bridging (N_b/N_{DNA}) PEIs per DNA, and the average number of bridging PEIs between a pair of bridged DNAs (n_b) are

calculated in **Figure 5.5a** as a function of time for both acidification rates (similar analysis for $\alpha = 2$ in **Appendix C.2**). Corresponding steady state averages from endosomal pH simulation are shown for reference using horizontal lines. The general trend demonstrates an exponential-like decay for N_b/N_{DNA} and n_b , and a corresponding logistic-like growth in N_f/N_{DNA} . N_p/N_{DNA} remains steady with small fluctuations. The initial change in N_f/N_{DNA} , N_p/N_{DNA} , N_b/N_{DNA} and n_b is faster for instant acidification than for slow acidification, and the difference caused by acidification rate is negligible after 2 μ s.

The decay of N_b/N_{DNA} and n_b signifies the loss of bridging PEIs due to acidification, likely caused by increased repulsion between PEIs within the same NP. **Figure 5.5a** shows the stabilization of N_b/N_{DNA} when n_b in the acidification simulations reach the steady state average in the endosomal pH simulation (at ~ 3 μ s). Because N_b is the product of n_b and the total number of bridged DNA pairs, the results indicate that the number of bridged DNA pairs and the number of bridging PEIs for each DNA pair reach equilibrium at the same time. In turn, this suggests that the primary PEI-PEI repulsion is between bridging PEIs bound to the same DNA pair. The steady state value of N_b/N_{DNA} in the acidification simulations is higher than that in the endosomal pH simulation, which is caused by the higher number of bridged DNA pairs in the acidification simulations. Consistently, $\langle s_{NP} \rangle_{SS}$ is higher in the acidification simulations (**Figure 5.4c**). During acidification, the free energy landscape of PEIs transitions from the grey curve towards the black curve in **Figure 5.2c**. Correspondingly, the energy minimum at $q = 2.5$ (2-bridging PEI) is increased relative to the one at $q = 1.5$ (peripheral PEI), facilitating the conversion of PEIs from bridging to peripheral. However, as N_b/N_{DNA} and n_b decrease the PEI-PEI and DNA-DNA repulsions reduce, which hinders further increasing of the energy minimum at $q = 2.5$ and release of the bridging PEIs.

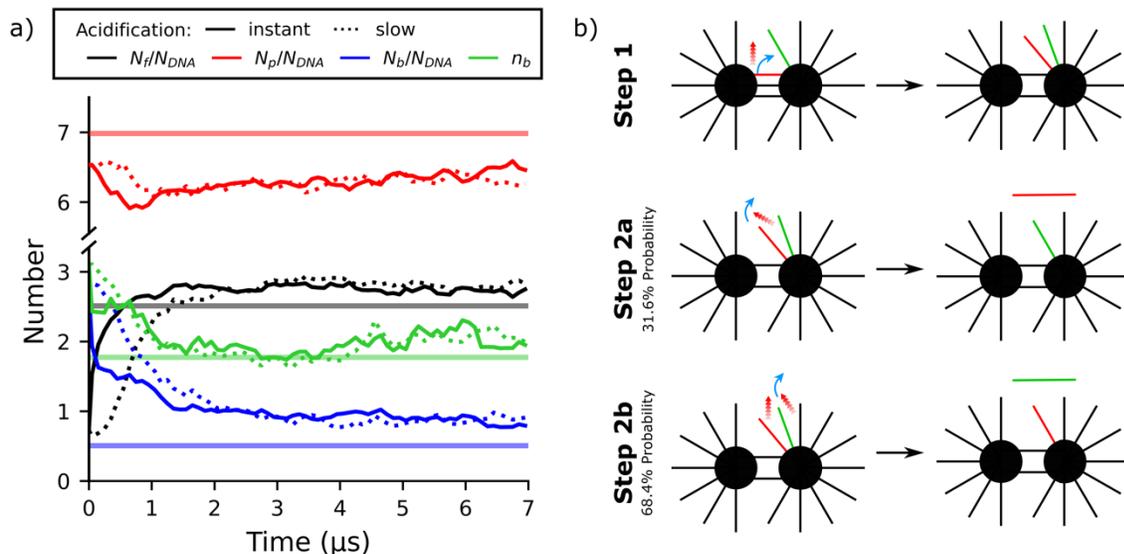


Figure 5.5: PEI roles and their transitions during endosomal acidification simulations for $\alpha = 10$. (a) Number of free (black), peripheral (red), and bridging (blue) PEIs per DNA, and the average number of bridging PEIs between a pair of bridged DNAs (green). The horizontal lines represent the steady state averages from endosomal pH simulations. (b) Schematic of transition from bridging to free PEIs. A bridging PEI that is eventually lost is shown in red and a peripheral PEI close to it is shown in green. The blue arrow indicates the transition, and the red arrowheads represent repulsion.

Detailed analysis shows that bridging PEIs convert to peripheral ones and peripheral PEIs convert to free ones (see **Appendix C.3**), while maintaining a dynamic balance of N_p/N_{DNA} . At the steady state of the physiological pH simulation, all peripheral PEIs conform to the DNAs so that their repulsions with other bound PEIs are minimized. Upon acidification, as a bridging PEI converts to peripheral, its conformation change tends to increase its repulsion with other bound PEIs, driving it or another peripheral PEI to become free. The increase of N_f/N_{DNA} over time is driven by such transitions. The free energy landscape in **Figure 5.2c** provides further evidence for such increase. Since acidification reduces the free energy minimum at $q = 0$ (free PEI) relative to the one at $q = 1.5$ (peripheral PEI), peripheral PEIs tend to become free and N_f/N_{DNA} would increase.

Figure 5.5b provides a graphical illustration for the typical process of losing a bridging PEI, where DNAs are shown as filled circles and PEIs as straight lines (**Figure 5.5b**). The bridging PEI lost is shown in red, and a nearby peripheral PEI is shown in green. Step 1 shows the transition of the PEI from bridging to peripheral. Two possible scenarios are presented in Step 2a and Step 2b for the peripheral to free PEI transition, along with their respective probabilities. In Step 2a, the original bridging PEI becomes free, whereas in Step 2b the neighboring peripheral PEI is freed by PEI-PEI repulsion.

5.2.3.2 NP structural changes

Structural changes in the NP are illustrated using a *transition diagram*, where each NP structure is represented with a network. The transition diagram can illustrate NP aggregation, dissociation, internal restructuring, as well as exchange of DNAs among multiple NPs (see **Section 5.5.4** for details). Since the simulations are dynamic, the transition diagrams are complex (**Appendix C.4** for $\alpha = 2$ and 10). Further simplifications are made by removing the loops corresponding to cyclic transitions so that only principal or “net” transitions are retained. This simplified diagram is referred to as the *principal transition diagram* (see **Section 5.5.4** for details).

Figure 5.6 reports the principal transition diagrams for the two largest NPs during instant (**Figure 5.6a, b**) and slow (**Figure 5.6c, d**) acidifications. Diagrams for other NPs at $\alpha = 10$, and $\alpha = 2$ are shown in **Section S5**. A black arrow in the diagram indicates the transition between two NP structures, along with the *time of the last transition* (due to the removal of cyclic transitions in between). Each NP structure is comprised of nodes (filled circles) representing the DNAs and edges (lines or curves) between two nodes representing the bridging PEIs. The thickness of each edge is proportional to the number of PEIs bridging the DNA pair, averaged over the period of existence for the NP structure. An edge is crossed out with a red line if the DNA pair becomes

unbridged after a transition, which is referred to as *bridge scission*. An unbridged DNA pair is connected with a blue edge if they become bridged after a transition, which is referred to as *bridge formation*. In the diagram the nodes are arranged to reduce overlapping edges, and the distance between two DNA nodes is not related to their physical separation in the NP. The relative positions of nodes are kept fixed in the same transition diagram to easily identify the DNAs at different simulation time. For example, the DNA marked using red arrows in **Figure 5.6a** is the same DNA at different simulation time. A blue circled plus symbol is used to denote aggregation of two or more NPs, while a red circled minus symbol is used to denote dissociation.

The observed structural changes are similar for both acidification rates and NP sizes. Several bridge scissions occur over time, turning a branched NP structure at 0 μs with many mutually bridged DNAs (especially in **Figure 5.6a, c**) into a few linear ones where the DNAs are almost sequentially connected. Almost all bridge scissions are observed for thin edges (except for 1315 ns in **Figure 5.6a**, 52.6 ns in **Figure 5.6b**, and 1472 and 4107 ns in **Figure 5.6d**) with a lower number of bridging PEIs. Generally, the thickness of an edge decreases with time due to the decrease in n_b (**Figure 5.5a**), making it more prone to bridge scission. Two observations can be made for the location of the bridge scissions. First, DNAs that are bridged with several others in a crowded setting are more likely to lose their bridging due to increased DNA-DNA repulsion. The DNA marked with the red arrows in **Figure 5.6a** is bridged with six other DNAs at the beginning of the simulation, making it the most crowded DNA in the NP. Five of these six bridges experience scission at 7.2, 61.2, 105.8 (two bridges), and 1746.4 ns respectively. The only surviving bridge is with the DNA marked by green arrows, and this pair had the thickest edge among the six at the beginning of the simulation. The same is true for the heavily crowded DNA marked by red arrows in **Figure 5.6c**, which loses all its bridges except the one with the DNA highlighted by green

arrows. Second, in NPs having linear structures with minimal mutual association amongst the DNAs, bridge scissions tend to occur near the terminals. A DNA located at the center of such a NP experiences almost equal DNA-DNA repulsion from the two directions, whereas for a DNA near the terminal the repulsive forces would be unbalanced leading to higher probability of bridge scission. For example, the loss of bridged DNA pairs occurs near the terminals at 1315, 1746.4 and 3914.2 ns in **Figure 5.6a**, and at 3001 and 3744.8 ns in **Figure 5.6c**.

Sufficient loss of bridged DNA pairs leads to NP dissociation. In **Figure 5.6a**, the first dissociation occurs at 105.8 ns near the crowded DNA marked with the red arrow. This dissociation forms three linearly structured NPs with $s_{NP} = 6, 5$ and 3 , which further dissociate near their terminals at 1315, 1746.4 and 3914.2 ns respectively, with larger NPs dissociating earlier in time. Similar observations are made in **Figure 5.6b-d**. Terminal DNAs in linearly structured NPs are more prone to dissociation for larger NPs because of they experience stronger unbalanced repulsive force from other DNAs in the same NP. The NP with $s_{NP} = 5$ in **Figure 5.6a**, created after the dissociation at 1315 ns, undergoes internal restructuring with three bridge formations and two bridge scissions. As a result, a branched NP structure with mutually bridged DNAs is formed, but such restructuring does not seem to be a dominant mode of transition. The final dissociated NPs have similar structures regardless of the acidification rate.

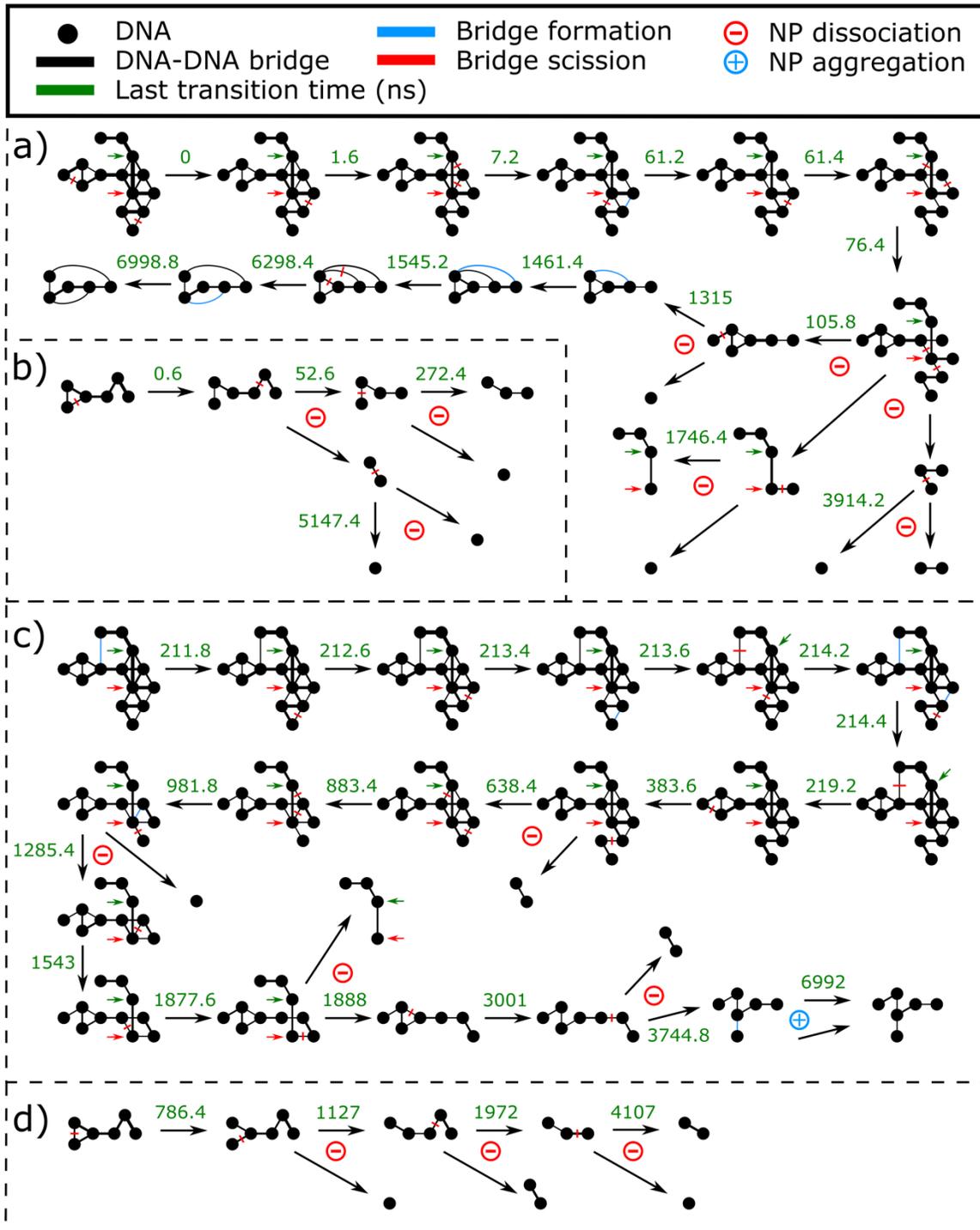


Figure 5.6: Principal transition diagram of NPs in acidification simulations for $\alpha = 10$. (a) Largest and (b) second-largest NP in the instant acidification simulation. (c) Largest and (d) second-largest NP in the slow acidification simulation.

The greatest difference found in NP dissociation upon slow and instant acidifications is the onset of the principal transition. Under instant acidification, the first principal transition occurs at 0 ns (**Figure 5.6a**), whereas under slow acidification it occurs at 211.8 ns (**Figure 5.6c**). Prior to 211.8 ns, several transition loops are observed (see **Appendix C.4**) suggesting that the loss of bridged DNA pairs is more reversible under slow acidification than that under instant acidification. Such reversibility can also be seen in **Figure 5.6c**: a bridged DNA pair is formed at 211.8 ns, lost at 214.2 ns, reformed at 214.4 ns, and lost again at 219.2 ns. This higher reversibility delays NP dissociation under the slow acidification condition.

5.2.3.3 NP swelling and shrinkage

The radius of gyration averaged across all NPs, $\langle R_g \rangle$, (see **Section 5.5.5**) is shown in **Figure 5.7a** for the two acidification simulations. The steady state average radii of gyration evaluated from physiological and endosomal pH simulations are shown using light-grey and dark-grey horizontal lines respectively. The overall decreasing trend in $\langle R_g \rangle$ is consistent with $\langle S_{NP} \rangle$ (**Figure 5.4c**) and therefore mainly due to dissociation of NPs. When NPs are not dissociating, their radius of gyration (R_g) can increase (corresponding to NP swelling), decrease (corresponding to NP shrinkage), or remain steady. To explore this, R_g for the largest and second-largest NP, before their dissociation, is shown in **Figure 5.7b, c** respectively. The light-grey horizontal line in each subfigure represents the NP's steady state average R_g from the physiological pH simulation. Since the time at which the NPs dissociate is different in slow and instant acidifications, time on the horizontal axis of **Figure 5.7b, c** is reported as the percentage of the total lifetime of the undissociated NPs.

In **Figure 5.7b**, an overall increase in R_g is observed for both acidification simulations, i.e., NP swells. This swelling can be explained by the bridge scissions that occur in the NPs (0-105.8

ns in **Figure 5.6a**; 0-638.4 ns in **Figure 5.6c**) and subsequent increase in DNA-DNA distance. The snapshots of the largest NP are shown in **Figure 5.7b** as insets: grey for $t = 0$, black for just before dissociation under instant acidification, and red for just before dissociation under slow acidification. Swelling of the NPs is visually detected, where “arms” containing DNA and PEI beads extend outwards from the NP’s center after their acidification.

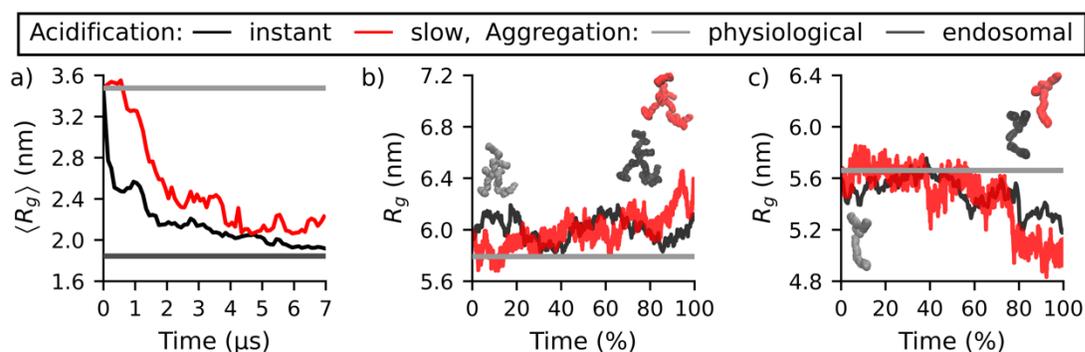


Figure 5.7: Radius of gyration of NPs in acidification simulations for $\alpha = 10$. (a) Average radius of gyration $\langle R_g \rangle$ as a function of time. Radius of gyration of the (b) largest NP with $s_{NP} = 14$, and (c) second-largest NP with $s_{NP} = 6$ before dissociation. In (b, c), snapshots of the NPs are shown in grey at $t = 0$, and in black (instant acidification) or red (slow acidification) at the onset of dissociation.

In **Figure 5.7c**, R_g decreases over time for both acidification simulations, i.e., NP shrinks. Similar to **Figure 5.7b**, the snapshots of the second-largest NP are shown in grey at $t = 0$, and black (instant acidification) or red (slow acidification) at the onset of dissociation. Different from the largest NP shown in **Figure 5.7b**, the second-largest NP has a more linear, rod-like structure. Under instant acidification, the NP shrinkage is produced by the bending of the rod-like structure, whereas under slow acidification it occurs due to a DNA moving to the center of the NP after a bridge scission (786.4 ns in **Figure 5.6d**; $t = 70\%$). NP swelling and shrinkage are therefore dependent on the initial NP geometry, structure and the internal restructuring that take place upon acidification.

5.3 Discussions

Endosomal acidification is found to cause a resizing of NPs, which strongly depends on the N/P ratio at which they are prepared. At physiological pH, NPs prepared with a low N/P ratio (1.18; $\alpha = 2$) has a small average size (in terms of the number of DNAs present in the NPs) and no free PEIs are found in the solution (**Section 5.2.1**). Upon endosomal acidification, the average NP size increases (**Figure 5.3c**) and the number of free PEIs remains zero (**Section 5.2.1**). Increase in NP size indicates further aggregation of small NPs into larger ones. On the contrary, under physiological pH NPs prepared with a high N/P ratio (5.91; $\alpha = 10$) has a larger average size with free PEIs in the solution (**Section 5.2.1**). Upon endosomal acidification, the average NP size decreases (**Figure 5.4c**) and the number of free PEIs in the solution increases (**Figure 5.5a**), suggesting the dissociation of large NPs into smaller ones. The term “dissociation” here may not imply complete separation of DNAs from the PEIs as it was used in some literature [21,27], although some PEIs are freed during the process. Similar to our results, Rehman *et al.*[17] observed dissociation of oligonucleotides from NPs within the endosome of HeLa cells, where the NPs were prepared at N/P ratio of 5 using 22 kDa linear PEI and 20-base pair oligonucleotides. This is comparable with our study using N/P ratio of 5.91 and 12-base pair DNAs, although the 586 Da PEI in this work is much smaller.

Acidic resizing of NPs has strong implications in the endosomal escape and nuclear trafficking of DNAs. Several mechanisms have been proposed in the literature for PEI-facilitated endosomal escape. First, based on the “proton sponge” hypothesis,[13,15] the osmotic pressure of the endosome can increase with the number of PEIs in it, due to the entry of water and counterions caused by PEIs’ protonation. Second, endosomal acidification of free PEIs can also increase the osmotic pressure, even without the entry of water and counterions, based on the theoretical study

of Yang and May[19]. This is supported by several experimental studies, such as in CT26 and 293T cells, where the presence of free PEIs in the endosome increased the efficacy of gene delivery.[28,29] Third, oncotic pressure or colloidal osmotic pressure in the endosome can increase with a rise in the number of free PEIs. This effect has been reported for non-viral carriers such as lactosylated poly(ethylene glycol)[30] in Hepatoma cells but not for PEIs. Increase in osmotic pressure can lead to endosomal swelling and its subsequent rupture or burst. Finally, PEIs can damage the endosomal membrane and the membrane damaging capability correlates positively with the number of PEIs in the endosome.[10,18] Results in the present work demonstrate that the ability for PEIs to induce endosomal escape depends on the N/P ratio at which the NPs are prepared. When prepared with a high N/P ratio, the number of PEIs is large, free PEIs are present and increase upon endosomal acidification. Furthermore, NP dissociation increases the number of NPs and thereby increases the colloidal osmotic pressure. These effects all lead to high osmotic pressure and membrane damage that can facilitate endosomal escape. In contrast, when the NPs are prepared with a low N/P ratio, the number of PEIs is small, and no free PEIs are available even after endosomal acidification. In fact, endosomal acidification results in further aggregation of NPs which can indeed reduce the colloidal osmotic pressure. Consequently, such NPs are ineffective in causing osmotic pressure increase and membrane damage. Consistent with our findings, low transfection efficacy has been observed in lung of mice[31] and Raw264 cells[5] for N/P ratio of 1, as compared to high transfection efficacy in COS-7 and CHO-K1 cells for N/P ratio of 6.[32]

After the endosomal escape, NPs are released into the cytoplasm, where bound PEIs protect the DNAs from degradation[3]. As the pH rises to cytosolic pH (6.9-7.4),[33] PEIs could get deprotonated which in turn can affect PEI-DNA interactions and sizes of the NPs. For example, dissociated NPs could aggregate again due to reduced PEI-PEI repulsion, which meanwhile could

be hindered by the more voluminous condition in the cytoplasm that promotes NP dispersion.[17] From experiments, there has not been consensus regarding whether complete dissociation of DNAs from the PEIs is necessary for successful transfection. Some reported that DNAs and PEIs never fully separated in cells such as EA.hy 926, COS-7, and CFPAC-1,[6,24] while others argued that complete separation was vital for gene delivery[17,21,27] in cells such as 293T and HeLa. Simulations results in this work suggest that regardless of the N/P ratio at which the NPs are prepared, complete separation of PEIs from DNAs is unlikely without the presence of molecules that has more binding affinity to DNA than PEI. It has been reported that nuclear trafficking of NPs can occur passively through diffusion if their diameter is less than 9 nm, or actively via short amino acid sequences and nuclear transport proteins if their diameter is between 9 and 39 nm;[34–36] NPs with diameter more than 39 nm cannot be trafficked.[34,35] Since the diameter of NPs prepared for gene delivery is on the order of 100 nm,[17,32] their dissociation into smaller NPs appears essential, which has been observed in our simulations for NPs prepared with high N/P ratio. Endosomal escape of NPs prepared with low N/P ratio is inferior, and DNAs could be digested in the endo-lysosomal compartment. However, if they are able to successfully escape from the endosome, nuclear trafficking may not pose further barriers because the NPs would likely dissociate as the pH rises. On the other hand, these NPs might be subjected to enzyme degradation in the cytoplasm as our *in-silico* microscopy reveals that DNAs are exposed at low N/P ratio (red colored regions in **Figure 5.3a**). Pollard et al.[6] reported poor transfection efficacy of NPs prepared with N/P ratio of 4.2 (1 charge equivalent) when injected into the cytoplasm of COS-7 cells, which could be due to the degradation of DNAs.

The rate of endosomal acidification can affect the acidic resizing of NPs which depends on the number of proton pumps in the endosome and the available chemical energy (adenosine-

triphosphate; ATP). The actual endosomal acidification rate is expected to be on the order of 10^{-7} H^+ /ns,[37] but such a low rate is not accessible by simulations due to the current computational limit. Two acidification rates, instant and slow (1 H^+ /ns), are simulated in this work to qualitatively assess its effect. The results suggest a limited influence of acidification rate on the overall dissociation of NPs prepared with high N/P ratio, and thereby on their endosomal escape and nuclear trafficking. For the NPs prepared with a low N/P ratio, slow acidification produces larger NPs. It is therefore reasonable to suspect that at an acidification rate of 10^{-7} H^+ /ns, average NP size can be even larger, further reducing the probability of endosomal escape.

To increase the efficacy of PEI-based gene delivery with a high N/P ratio, it is desirable to enhance the NP dissociation upon endosomal acidification, which would translate to higher efficacy of endosomal escape and nuclear trafficking. There is a remarkable consistency in the dissociation mechanisms at both acidification rates such as, loss of bridging PEIs due to repulsion from PEIs bound to the same DNA, bridge scission between DNA pairs with a low number of bridging PEIs, NP dissociation near crowded DNAs due to DNA-DNA repulsion, etc. We can exploit these mechanisms to design PEIs that promote NP dissociation during acidification. Specifically, the following strategies are proposed for NP preparation: reducing DNA-DNA separation, promoting DNA crowding (increasing mutual bridging of DNAs), and forming NPs with a moderate number of bridging PEIs. We suspect a PEI with moderate molecular weight (MW) and degree of branching is an effective candidate. If the MW of the PEI is too low, it might hinder the formation of large NPs that can later dissociate, whereas if the MW is too high the DNAs might be far apart causing difficulties in NP dissociation due to low DNA-DNA repulsion. High transfection efficacy has been reported for PEI with moderate[38] (12 kDa better than 1616 kDa in K5 Cells) and high[39] (70 kDa better than 10 kDa and 2 kDa in EA.hy 926 cells) MW. However,

these results are reported for polydisperse PEIs and are not informative of the optimal MW of PEIs. Werth *et al.*[40] fractionated commercially available polydisperse 25 kDa and found moderate MW PEIs (4-10 kDa) had the highest transfection efficacy in SKOV-3 cells, with undetectable transfection efficacy of PEIs above 10 kDa and below 2 kDa. Bieber *et al.*[41] fractionated the same PEI and found those with MW in the range of 0.5-10 kDa had the highest transfection efficacy in PaTu 8902 cells. These results are consistent and align with the hypothesis that moderate MW PEIs are likely more efficacious. If a PEI has a dendritic structure, it can crowd several DNAs, but the number of bridging PEIs would be low and hence their repulsion, making NP dissociation difficult. While linear PEIs could create a large number of bridging PEIs, bridge scission upon endosomal acidification could be inadequate. Comparison of transfection efficacy between linear and branched PEIs has been reported to differ between cell lines and between *in-vitro* and *in-vivo*. [32,42] To the best of our knowledge, the only study that systematically explored the effect of the degree of branching of the PEI (measured using ^{13}C NMR spectroscopy) on the transfection efficacy was conducted by Krämer *et al.*[43] Consistent with our supposition, the transfection efficacy was found to be highest for the degree of branching of 58% (among 0, 58, 72, and 100%) for NIH/3T3 and COS-7 cells. Further investigation on the combined effect of degree of branching and MW on transfection efficacy in different cell lines is needed to ascertain the optimal PEI properties for gene delivery, by performing simulations with high MW PEI with different degrees of branching.

5.4 Conclusions

Large-scale coarse-grained simulations are performed to study the size and structural changes of PEI-DNA NPs when subjected to endosomal acidification. The results reveal an acidic resizing of NPs that is highly dependent on the N/P ratio at which they are prepared. NPs prepared

with a low N/P ratio at physiological pH aggregate further upon acidification, whereas those prepared with a high N/P ratio dissociate. The extent of NP aggregation is sensitive to the rate of acidification, with more aggregation occurring under slow acidification ($1 \text{ H}^+/\text{ns}$) as compared to instant acidification. However, acidification rate has limited influence on the overall dissociation of NPs prepared with high N/P ratio, and the main consequence of a slower acidification is a delay in the onset of NP dissociation.

During endosomal acidification of NPs prepared at high N/P ratio, some PEIs bound to DNAs are released into the solution due to the repulsion between PEIs bridging the same DNA pair. Reduction in the number of bridging PEIs leads to bridge scission, i.e., disconnection, between some DNA pairs, making the NP more prone to dissociation. Bridge scission is more likely to occur near DNAs that are crowded (bridged to several other DNAs), and near the terminals of a NP with a linear structure. Principal transition diagrams are created to illustrate the structural changes of the NPs during acidification, and free energy landscapes of the PEIs facilitate the understanding of the forces (PEI-PEI repulsion, DNA-DNA repulsion, PEI-DNA attraction) driving those structural changes.

5.5 Methods

5.5.1 Simulations

Martini coarse-grained (CG) molecular dynamics (MD) simulations were performed for both aggregation and acidification. Specifics for the two types of simulations can be found in **Section 5.5.1.1** and **Section 5.5.1.2** respectively, with common details described below.

The DNA simulated was the Drew-Dickerson dodecamer $d(\text{CGCGAATTCGCG})_2$ in its canonical B form, which was modeled using the stiff Martini DNA forcefield.[45] In both physiological and endosomal pH, the DNAs were completely deprotonated and the charge on each

DNA $Q_{DNA} = -22$. A 586 Da semi-linear PEI with a small degree of branching was used,[44] each carrying a charge of $Q_{PEI} = 3$ at physiological (referred to as PEI3) and $Q_{PEI} = 6$ at endosomal pH (referred to as PEI6).[44,46] Their structures are shown in **Figure 5.8**. The protonated and unprotonated beads were modeled with Q_d and P_1 bead types[47] respectively, based on Martini models developed in **Chapter 3**.[48,49] The solvent was modeled using the polarizable Martini water[50] with 150 mM of KCl. All simulations were performed in Gromacs 5.[51]

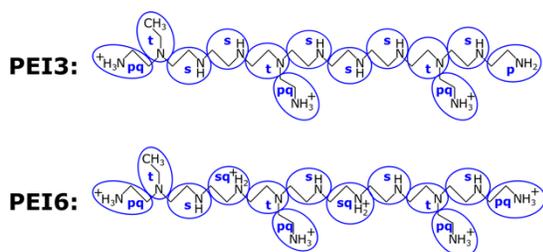


Figure 5.8: Structures of PEI at physiological and endosomal pH (PEI3 is SL23 in **Chapter 3**, and PEI6 is SL46 in **Chapter 3**). All-atom (AA) structures are shown in black,[44] whereas the CG structures and mapping scheme are shown in blue. The text in blue represents the type of nitrogen contained in the CG bead, tertiary (t), secondary (s), protonated secondary (sq), primary (p), and protonated primary (pq).

5.5.1.1 Aggregation simulations

Four systems with 27 DNAs were simulated, each at a specific pH (physiological or endosomal) and PEI/DNA number ratio ($\alpha = 2$ or 10). The details of these simulated systems are summarized in **Table 5.1**. The simulations were performed following **Chapter 4**[25], but with a different bead type for unprotonated PEI beads for improved accuracy (**Chapter 3**[49]). All the DNAs were aligned and placed in a 3x3x3 array at the center of a 25 nm cubic box. The closest distance between the centers of mass of DNAs was 6 nm. All the PEIs were placed randomly inside the box, followed by the addition of Martini polarizable water.[50] Appropriate amounts of Martini

potassium and chloride ions[47] were added to first neutralize the system and then reach the salt concentration of 150 mM.

Table 5.1: Details of aggregation simulations. PEI/DNA number ratio (α), N/P ratio, pH, number of DNAs (N_{DNA}) and PEIs (N_{PEI}), charge of DNA (Q_{DNA}) and PEI (Q_{PEI}), critical α (α_c) at which the total charge of PEIs and DNAs is zero, and total simulations time (t_{sim}). The critical α is calculated using the equation $\alpha_c = -Q_{DNA}/Q_{PEI}$.

System	α	N/P ratio	pH	N_{DNA}	N_{PEI}	Q_{DNA}	Q_{PEI}	α_c	t_{sim} (μ s)
P2	2	1.18	Physiological	27	54	-22	3	7.33	9.5
P10	10	5.91	Physiological	27	270	-22	3	7.33	8
E2	2	1.18	Endosomal	27	54	-22	6	3.66	12
E10	10	5.91	Endosomal	27	270	-22	6	3.66	6

The initial configuration was first energy-minimized using steepest-descent, followed by a constrained NPT simulation for 1 ns and an unconstrained NPT simulation of different times (see t_{sim} in **Table 5.1**). The length of unconstrained NPT simulations varied based on the time required for a system to reach its steady state, which was defined by observing a stable value for the average NP size $\langle S_{NP} \rangle$ for 1 μ s. For constrained NPT simulations, all the bonds in PEI, DNA backbone, and polarizable Martini water were constrained using LINCS. Van der Waals interaction was cut off at 1.1 nm using the potential-shift-Verlet scheme. Long-range electrostatic interaction was handled by the reaction-field scheme[52] while short-ranged electrostatics was modeled using Coulombic interaction cut off at 1.1 nm. The relative dielectric constant of the solvent and the reaction-field were set to be 2.5 and ∞ respectively.[50] A neighbor list was maintained for the cutoff radius of 1.1 nm using the Verlet scheme[53], which was updated every 20 steps. The pressure was maintained at 1 bar using Berendsen barostat with a time constant of 3 ps and

compressibility of $3 \times 10^{-4} \text{ bar}^{-1}$. The temperature was maintained at 300 K using a velocity-rescaling thermostat with a time constant of 0.1 ps. Initial velocities were generated from the Maxwell distribution for 300 K. A leap-frog integrator was used with a timestep of 5 fs, while applying a periodic boundary condition in all directions. All simulation time reported in this work was scaled up by a factor of four (excluding time constants and timestep), a standard factor used to match the diffusion coefficient of water in AA and CG simulations. Unconstrained NPT simulations were similar to the constrained ones, with three differences: (i) initial velocities of the CG beads were assigned from the last step of constrained simulations, (ii) Parrinello-Rahman barostat was used to maintain the pressure at 1 bar, using a time constant of 5 ps and compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$, and (iii) constraints were only applied to bonds present in the Martini polarizable water.

5.5.1.2 Acidification simulations

The configuration of beads in the last timestep of an aggregation simulation ($\alpha = 2$ or 10) at physiological pH, i.e., system P2 or P10, was used as the initial configuration for the corresponding acidification simulation. Endosomal acidification was modeled by increasing the protonation of the PEIs from PEI3 to PEI6. This was achieved by modifying three CG beads in PEI3 from P_1 to Q_d (**Figure 5.8**). Subsequently, an appropriate number of water molecules was replaced, at random, with chloride ions to maintain charge neutrality. Two acidification rates were studied, namely instant and slow. In instant acidification, all PEIs was changed from PEI3 to PEI6 at the beginning of the simulation. In slow acidification, one PEI was changed from PEI3 to PEI6 every 3 ns, i.e., the acidification rate is $1 \text{ H}^+/\text{ns}$. The PEIs were protonated in the order of their molecule ID. Since the PEIs were placed randomly at the beginning of the aggregation simulation,

the acidification of PEIs was, in essence, random. The details of these simulations are summarized in **Table 5.2**.

Unlike the aggregation simulations, only unconstrained NPT simulations were performed, which followed the same settings of unconstrained NPT aggregation simulations (see **Section 5.5.1.1**). The total simulation time reported in **Table 5.2** is only for the acidification simulation and does not include the aggregation simulation time. The total simulation time was different for $\alpha = 2$ and 10, allowing each system to reach steady state where $\langle s_{NP} \rangle$ was not changing significantly for 1 μs .

Table 5.2: Details of acidification simulations. PEI/DNA number ratio (α), type of acidification, number of DNAs (N_{DNA}) and PEIs (N_{PEI}), acidification rate, total time to acidify all PEIs (t_{acid}), and total simulation time (t_{sim}).

System	α	Acidification	N_{DNA}	N_{PEI}	Acidification rate	t_{acid} (μs)	t_{sim} (μs)
I2	2	Instant	27	54	∞	0	13.5
I10	10	Instant	27	270	∞	0	7
S2	2	Slow	27	54	1 H ⁺ /ns	0.162	13.5
S10	10	Slow	27	270	1 H ⁺ /ns	0.810	7

5.5.2 Free energy landscape

The free energy of PEI (F) was calculated using Eq 5.2, where k_B is the Boltzmann's constant, T is the absolute temperature, and P_q is the probability of finding a PEI with reaction coordinate q at the steady state. The reaction coordinate q for a PEI was defined by Eq 5.3, where N_{DNA} is the total number of DNAs in the system, d_i is the minimum distance between the PEI and the i^{th} DNA, r_{bind} is a distance that defines binding between PEI and DNA, and \mathbb{H} is the Heaviside step function. When a PEI is bound with the i^{th} DNA, $d_i < r_{bind}$ and $\mathbb{H}(r_{bind} - d_i)$ is 1. In this case, r_{bind}/d_i is in the range of 1-1.2, where the upper limit of 1.2 arises due to repulsive van der Waals forces. In

contrast, if a PEI is not bound to the i^{th} DNA, $d_i > r_{bind}$, $\mathbb{H}(d_i - r_{bind})$ is 1 and r_{bind}/d_i is in the range 0-1. The first term in Eq 5.3 calculates r_{bind}/d_i for the closest unbound DNA using the max function, and the value ranges between 0 and 1. In the second term in Eq 5.3, r_{bind}/d_i is summed over all bound DNAs, and the increase in q is in the range of 1-1.2 for every bound DNA. For a free PEI, q only contains the contribution from the closest unbound DNA, and ranges from 0 to 1. The range of q is 1-2.2 when a PEI is bound to one DNA, since the two terms in Eq 5.3 contribute values of 0-1 and 1-1.2 respectively. Similarly, the range of q is between k to $1 + 1.2k$ when PEI is bound to k DNAs ($k > 0$). Theoretically there can be some overlap between the ranges of q for a PEI bound to $k - 1$ and k DNAs. However, the overlap observed is minimal. Therefore, in our results the range of q is taken as $1 + 1.2(k - 1)$ to $1 + 1.2k$ when a PEI is bound to k DNAs ($k > 0$; see **Appendix C.6**). Consequently, q positively correlates with the number of DNAs the PEI is bound with, hence this continuous variable is a good candidate to be used as the reaction coordinator in the energy landscape. The binding distance r_{bind} was taken as 0.528 nm, the van der Waals' diameter between PEI and DNA beads. The reaction coordinate q was calculated for all PEIs in the system and the range of q was divided into bins of width 0.2 to calculate the probability P_q required for the free energy calculation in Eq 5.2. F provides a quantitative measure for the likelihood of PEIs to form simultaneous contact with multiple DNAs.

$$F(q) = -k_B T \ln (P_q) \quad 5.2$$

$$q = \max\left(\frac{r_{bind}}{d_i} \mathbb{H}(d_i - r_{bind})\right) + \sum_{i=0}^{N_{DNA}} \frac{r_{bind}}{d_i} \mathbb{H}(r_{bind} - d_i) \quad 5.3$$

5.5.3 *In-silico* fluorescence microscopy

In-silico fluorescence microscopy images were generated using an open source tool developed in **Chapter 6**.^[54,55] Briefly, the Gandy point-spread-function^[56] (PSF) was used to

model a virtual-microscope with the numerical aperture of 1.3, the refractive index of immersion oil and specimen equal to 1.51, and full-width-at-half-maximum (FWHM) scaling factor (**Chapter 6**[54]) of 400. The PSF was calculated over a volume of $15 \times 15 \times 25 \text{ nm}^3$ and voxel dimensions of $0.1 \times 0.1 \times 0.2 \text{ nm}^3$, where the third dimension was the optical axis, which is taken to be the z-axis. All DNA and PEI particles emitted light of 670 nm (modeling Cy5) and 518 nm (modeling FITC) respectively. For $\alpha = 2$, the maximum intensities of light emission were 0.1 and 0.5 respectively for DNA and PEI particles; the corresponding values were 0.1 and 0.15 for $\alpha = 10$. All images were generated by applying the periodic boundary condition in all directions and time-averaging over 50 timesteps, i.e., an effective exposure time of 10 ns. Red and green colors were assigned to DNAs and PEIs respectively, and the colors were mixed based on the color mixing scheme developed in **Chapter 6**. [54]

5.5.4 Transition diagram

As shown in **Figure 5.9a**, a NP structure is graphically presented by a network consisting of nodes (filled circles) that represent DNAs, and edges (lines) connecting the nodes represent the presence of bridging PEI(s) between them. A unique ID was assigned to each NP structure (e.g., U_1, U_2 , etc.). For purposes of explaining the method, nodes corresponding to the same DNA in different NP structures were shown using the same color. While generating the transition diagram (and principal transition diagram) two NP structures were considered to be identical, hence having the same ID, if they had the same nodes and edges.

A transition diagram was created from the IDs of multiple NP structures and arrows indicating the conversions between them. **Figure 5.9b-d** (left) shows some examples, where the number accompanying each arrow stores the time of the transition. A transition diagram illustrates (i) internal restructuring of a NP if one unique ID transitions to another, (ii) NP aggregation if

multiple IDs transition into one, (iii) NP dissociation if one ID transitions into multiple, and (iv) exchange of DNAs among several NPs if multiple IDs transition to multiple but different ones. For NP aggregation, arrows were drawn from the aggregating NPs to a ‘+’ symbol, and a separate arrow was drawn from the ‘+’ symbol to the aggregated NP (see **Figure 5.9c, d**). Similarly, for NP dissociation and exchange of DNAs among NPs, ‘-’ (see **Figure 5.9c, d**) and ‘×’ (not shown) symbols were used respectively.

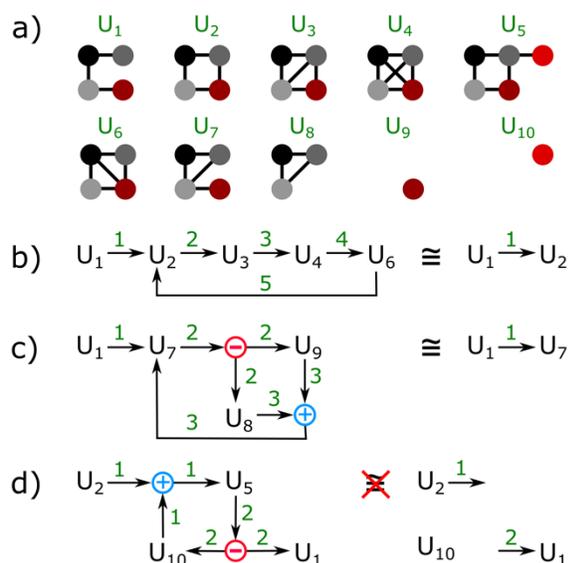


Figure 5.9: Generation of principal transition diagram. (a) Network representations of NP structures, where nodes are DNAs and edges denote the presence of bridging PEI between a pair of DNAs. Nodes corresponding to the same DNA in different NP structures are shown using the same color. (b), (c) Generation of a principal transition diagram (right) from a complete transition diagram (left) by removing reversible transitions and transition loops. (d) An example where loops cannot be removed from the complete transition diagram (left). Naively removing the loop leads to an unphysical transition diagram (right). In (b)-(d) each number over a directed edge is the time, in nanoseconds, at which the transition occurs.

A principal transition diagram contains the initial and final (at the last timestep) NP structures, as well as transitions between them representing the “net” changes. In some cases, transitions produced IDs already present in the transition diagram. This led to the appearance of a

transition loop which was removed to obtain the “net” transitions. For example, the “net” transition was U_1 to U_2 in **Figure 5.9b**, and U_1 to U_7 in **Figure 5.9c**, represented by the principal transition diagrams on the right side of these subfigures. Such loop removal was carried out at each time step when new transitions were added to the transition diagram. Care was taken for the cases where no transition path was available after the removal to reach the NP structure at the present time step. For example, in **Figure 5.9d**, if the loop $U_{10} \rightarrow '+' \rightarrow U_5 \rightarrow '-' \rightarrow U_{10}$ were removed while keeping the initial (U_2 and U_{10}) and current IDs (U_1 and U_{10}), there would be no transition path from U_2 to U_1 (**Figure 5.9d**, right), and therefore the loop was kept in the principal transition diagram.

After generating the principal transition diagram, the IDs were replaced with the corresponding NP structures to generate **Figure 5.6**. For each NP structure, the average number of bridging PEIs between each DNA pair was calculated and represented by the thickness of the edge connecting the DNA nodes. The average was performed over the time range when the NP structure was present in the principal transition diagram. The codes for generating the principal and complete transition diagrams have been made available to the public.[57]

5.5.5 Radius of gyration

The radius of gyration R_g was calculated using Eq 5.4, where $r_{i,com}$ is the distance of the i^{th} bead in the NP from its center of mass, m_i is the mass of the i^{th} bead, and N_{beads} is the total number of beads in the NP. For a NP that crossed a periodic boundary, R_g was calculated after making its structure whole using an open source tool.[57] The average R_g reported in **Figure 5.7** was calculated by root-mean-square averaging.

$$R_g = \left(\frac{1}{N_{beads}} \sum_{i=0}^{N_{beads}} m_i r_{i,com}^2 \right)^{1/2} \quad 5.4$$

5.6 Data and code availability

<https://github.com/subhamoymahajan/NPanalysis> version v.1.3.

<https://github.com/subhamoymahajan/in-silico-microscopy> version v.1.2.2.

5.7 Acknowledgments

We acknowledge the computing resources and technical support from Western Canada Research Grid (WestGrid). Financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC), Sadler Graduate Scholarship, Alberta Graduate Excellence Scholarship, R. R. Gilpin Memorial Scholarship, and MITACS Globalink Graduate Fellowship is gratefully acknowledged.

5.8 References

- [1] Verma I.M., Somia N. Gene Therapy - Promises, Problems and Prospects. *Nature* **1997**, 389, 239–42.
- [2] Lungwitz U., Breunig M., Blunk T., Göpferich A. Polyethylenimine-Based Non-Viral Gene Delivery Systems. *Eur. J. Pharm. Biopharm.* **2005**, 60, 247–66.
- [3] Godbey W.T., Barry M.A., Saggau P., Wu K.K., Mikos A.G. Poly(Ethylenimine)-Mediated Transfection: A New Paradigm for Gene Delivery. *J. Biomed. Mater. Res.* **2000**, 51, 321–8.
- [4] Khalil I.A., Kogure K., Akita H., Harashima H. Uptake Pathways and Subsequent Intracellular Trafficking in Nonviral Gene Delivery. *Pharmacol. Rev.* **2006**, 58, 32–45.
- [5] Oh Y.-K., Suh D., Kim J.M., Choi H.-G., Shin K., Ko J.J. Polyethylenimine-Mediated Cellular Uptake, Nucleus Trafficking and Expression of Cytokine Plasmid DNA. *Gene Ther.* **2002**, 9, 1627–32.
- [6] Pollard H., Remy J.S., Loussouarn G., Demolombe S., Behr J.P., Escande D. Polyethylenimine but not Cationic Lipids Promotes Transgene Delivery to the Nucleus in Mammalian Cells. *J. Biol. Chem.* **1998**, 273, 7507–11.
- [7] Kay M.A., Glorioso J.C., Naldini L.M. Viral Vectors for Gene Therapy: The Art of Turning Infectious Agents Into Vehicles of Therapeutics. *Nat. Med.* **2001**, 7, 33–40.

- [8] Glover D.J., Lipps H.J., Jans D.A. Towards Safe, Non-Viral Therapeutic Gene Expression in Humans. *Nat. Rev. Genet.* **2005**, 6, 299–310.
- [9] Yin H., Kanasty R.L., Eltoukhy A.A., Vegas A.J., Dorkin J.R., Anderson D.G. Non-Viral Vectors for Gene-Based Therapy. *Nat. Rev. Genet.* **2014**, 15, 541–55.
- [10] Bieber T., Meissner W., Kostin S., Niemann A., Elsasser H.-P. Intracellular Route and Transcriptional Competence of Polyethylenimine-DNA Complexes. *J. Control. Release* **2002**, 82, 441–54.
- [11] Godbey W.T., Wu K.K., Mikos A.G. Poly(Ethylenimine)-Mediated Gene Delivery Affects Endothelial Cell Function and Viability. *Biomaterials* **2001**, 22, 471–80.
- [12] Akinc A., Thomas M., Klibanov A.M., Langer R. Exploring Polyethylenimine-Mediated DNA Transfection and the Proton Sponge Hypothesis. *J. Gene Med.* **2005**, 7, 657–63.
- [13] Behr J.-P. The Proton Sponge: A Trick to Enter Cells the Viruses did not Exploit. *Chimia (Aarau)*. **1997**, 51, 34–6.
- [14] Boussif O., Lezoualc'h F., Zanta M.A., Mergny M.D., Scherman D., Demeneix B., et al. A Versatile Vector for Gene and Oligonucleotide Transfer into Cells in Culture and In Vivo: Polyethylenimine. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, 92, 7297–301.
- [15] Sonawane N.D., Szoka F.C., Verkman A.S. Chloride Accumulation and Swelling in Endosomes Enhances DNA Transfer by Polyamine-DNA Polyplexes. *J. Biol. Chem.* **2003**, 278, 44826–31.
- [16] Vermeulen L.M.P., Brans T., Samal S.K., Dubruel P., Demeester J., De Smedt S.C., et al. Endosomal Size and Membrane Leakiness Influence Proton Sponge-Based Rupture of Endosomal Vesicles. *ACS Nano* **2018**, 12, 2332–45.
- [17] Rehman Z.U., Hoekstra D., Zuhorn I.S. Mechanism of Polyplex- and Lipoplex-Mediated Delivery of Nucleic Acids: Real-Time Visualization of Transient Membrane Destabilization Without Endosomal Lysis. *ACS Nano* **2013**, 7, 3767–77.
- [18] Klemm A.R., Young D., Lloyd J.B. Effects of Polyethyleneimine on Endocytosis and Lysosome Stability. *Biochem. Pharmacol.* **1998**, 56, 41–6.
- [19] Yang S., May S. Release of Cationic Polymer-DNA Complexes From the Endosome: A Theoretical Investigation of the Proton Sponge Hypothesis. *J. Chem. Phys.* **2008**, 129, 185105.
- [20] Freeman E.C., Weiland L.M., Meng W.S. Modeling the Proton Sponge Hypothesis:

- Examining Proton Sponge Effectiveness for Enhancing Intracellular Gene Delivery Through Multiscale Modeling. *J. Biomater. Sci. Polym. Ed.* **2013**, 24, 398–416.
- [21] Itaka K., Harada A., Yamasaki Y., Nakamura K., Kawaguchi H., Kataoka K. In Situ Single Cell Observation by Fluorescence Resonance Energy Transfer Reveals Fast Intra-Cytoplasmic Delivery and Easy Release of Plasmid DNA Complexed With Linear Polyethylenimine. *J. Gene Med.* **2004**, 6, 76–84.
- [22] Abbe E. Beiträge zur Theorie des Mikroskops und der Mikroskopischen Wahrnehmung. *Arch. Für Mikroskopische Anat.* **1873**, 9, 413–68.
- [23] Huang B., Bates M., Zhuang X. Super-Resolution Fluorescence Microscopy. *Annu. Rev. Biochem.* **2009**, 78, 993–1016.
- [24] Godbey W.T., Wu K.K., Mikos A.G. Tracking the Intracellular Path of Poly(ethylenimine)/DNA Complexes for Gene Delivery. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, 96, 5177–81.
- [25] Mahajan S., Tang T. Polyethylenimine-DNA Ratio Strongly Affects Their Nanoparticle Formation: A Large-Scale Coarse-Grained Molecular Dynamics Study. *J. Phys. Chem. B* **2019**, 123, 9629–40.
- [26] von Smoluchowski M. Versuch Einer Mathematischen Theorie der Koagulationskinetik kolloider Lösungen. *Zeitschrift f. Physik. Chemie.* **1917**, 92, 129–68.
- [27] Clamme J.P., Azoulay J., Mély Y. Monitoring of the Formation and Dissociation of Polyethylenimine/DNA Complexes by Two Photon Fluorescence Correlation Spectroscopy. *Biophys. J.* **2003**, 84, 1960–8.
- [28] Boeckle S., von Gersdorff K., van der Pijpen S., Culmsee C., Wagner E., Ogris M. Purification of Polyethylenimine Polyplexes Highlights the Role of Free Polycations in Gene Transfer. *J. Gene Med.* **2004**, 6, 1102–11.
- [29] Yue Y., Jin F., Deng R., Cai J., Dai Z., Lin M.C.M., et al. Revisit Complexation Between DNA and Polyethylenimine - Effect of Length of Free Polycationic Chains on Gene Transfection. *J. Control. Release* **2011**, 152, 143–51.
- [30] Oishi M., Nagasaki Y., Itaka K., Nishiyama N., Kataoka K. Lactosylated Poly(ethylene Glycol)-siRNA Conjugate through Acid-Labile β -Thiopropionate Linkage to Construct pH-Sensitive Polyion Complex Micelles Achieving Enhanced Gene Silencing in Hepatoma Cells. *J. Am. Chem. Soc.* **2005**, 127, 1624–5.

- [31] Rudolph C., Lausier J., Naundorf S., Müller R.H., Rosenecker J. In Vivo Gene Delivery to the Lung Using Polyethylenimine and Fractured Polyamidoamine Dendrimers. *J. Gene Med.* **2000**, 2, 269–78.
- [32] Choosakoonkriang S., Lobo B.A., Koe G.S., Koe J.G., Middaugh C.R. Biophysical Characterization of PEI/DNA Complexes. *J. Pharm. Sci.* **2003**, 92, 1710–22.
- [33] Roos A., Boron W.F. Intracellular pH. *Physiol. Rev.* **1981**, 61, 296–434.
- [34] van der Aa M.A.E.M., Mastrobattista E., Oosting R.S., Hennink W.E., Koning G.A., Crommelin D.J.A. The Nuclear Pore Complex: The Gateway to Successful Nonviral Gene Delivery. *Pharm. Res.* **2006**, 23, 447–59.
- [35] Panté N., Kann M. Nuclear Pore Complex Is Able to Transport Macromolecules with Diameters of ~39 nm. *Mol. Biol. Cell* **2002**, 13, 425–34.
- [36] Paine P.L., Moore L.C., Horowitz S.B. Nuclear Envelope Permeability. *Nature* **1975**, 254, 109–14.
- [37] Wang C., Wang Y., Li Y., Bodemann B., Zhao T., Ma X., et al. A Nanobuffer Reporter Library for Fine-Scale Imaging and Perturbation of Endocytic Organelles. *Nat. Commun.* **2015**, 6, 8524.
- [38] Fischer D., Bieber T., Li Y., Elsässer H.-P., Kissel T. A Novel Non-viral Vector for DNA Delivery Based on Low Molecular Weight, Branched Polyethylenimine: Effect of Molecular Weight on Transfection Efficiency and Cytotoxicity. *Pharm. Res.* **1999**, 16, 1273–9.
- [39] Godbey W.T., Wu K.K., Mikos A.G. Size Matters: Molecular Weight Affects the Efficiency of Poly(ethylenimine) as a Gene Delivery Vehicle. *J. Biomed. Mater. Res.* **1999**, 45, 268–75.
- [40] Werth S., Urban-Klein B., Dai L., Höbel S., Grzelinski M., Bakowsky U., et al. A Low Molecular Weight Fraction of Polyethylenimine (PEI) Displays Increased Transfection Efficiency of DNA and siRNA in Fresh or Lyophilized Complexes. *J. Control. Release* **2006**, 112, 257–70.
- [41] Bieber T., Elsässer H.P. Preparation of a Low Molecular Weight Polyethylenimine for Efficient Cell Transfection. *Biotechniques* **2001**, 30, 74–7, 80–1.
- [42] Wightman L., Kircheis R., Rössler V., Carotta S., Ruzicka R., Kursa M., et al. Different Behavior of Branched and Linear Polyethylenimine for Gene Delivery In Vitro and In Vivo.

- J. Gene Med.* **2001**, 3, 362–72.
- [43] Krämer M., Stumbé J.-F., Grimm G., Kaufmann B., Krüger U., Weber M., et al. Dendritic Polyamines: Simple Access to New Materials With Defined Treelike Structures for Application in Nonviral Gene Delivery. *ChemBioChem* **2004**, 5, 1081–7.
- [44] Sun C., Tang T., Uludağ H., Cuervo J.E. Molecular Dynamics Simulations of DNA/PEI Complexes: Effect of PEI Branching and Protonation State. *Biophys. J.* **2011**, 100, 2754–63.
- [45] Uusitalo J.J., Ingólfsson H.I., Akhshi P., Tieleman D.P., Marrink S.J. Martini Coarse-Grained Force Field: Extension to DNA. *J. Chem. Theory Comput.* **2015**, 11, 3932–45.
- [46] Utsuno K., Uludağ H. Thermodynamics of Polyethylenimine-DNA Binding and DNA Condensation. *Biophys. J.* **2010**, 99, 201–7.
- [47] Marrink S.J., Risselada H.J., Yefimov S., Tieleman D.P., de Vries A.H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, 111, 7812–24.
- [48] Mahajan S., Tang T. Martini Coarse-Grained Model for Polyethylenimine. *J. Comput. Chem.* **2019**, 40, 607–18.
- [49] Mahajan S., Tang T. Erratum : “Martini Coarse-Grained Model for Polyethylenimine” [J. Comput. Chem. 2019, 40, 607-618, DOI:10.1002/jcc.25747]. *J. Comput. Chem.* **2020**, 41, 1730–4.
- [50] Yesylevskyy S.O., Schäfer L. V., Sengupta D., Marrink S.J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comput. Biol.* **2010**, 6, e1000810.
- [51] Abraham M.J., Murtola T., Schulz R., Páll S., Smith J.C., Hess B., et al. GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism From Laptops to Supercomputers. *SoftwareX* **2015**, 1–2, 19–25.
- [52] Tironi I.G., Sperb R., Smith P.E., van Gunsteren W.F. A Generalized Reaction Field Method for Molecular Dynamics Simulations. *J. Chem. Phys.* **1995**, 102, 5451–9.
- [53] Verlet L. Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, 159, 98–103.
- [54] Mahajan S., Tang T. Meeting experiments at the diffraction barrier: an in-silico widefield fluorescence microscopy. *BioRxiv 10.1101/2021.03.02.433395 (Version 1)* **2021**.
- [55] Mahajan S. In-silico-microscopy v.1.2.2, <https://github.com/subhamoymahajan/in-silico->

microscopy, **2021**.

[56] Gandy R.O. Out-of-Focus Diffraction Patterns for Microscope Objectives. *Proc. Phys. Soc. B* **1954**, 67, 825–31.

[57] Mahajan S. NPanalysis v.1.3, <https://github.com/subhamoymahajan/NPanalysis>, **2021**.

Chapter 6

In-silico Fluorescence Microscopy

* A version of this chapter's sections has been published in a pre-print. Adapted from Subhamoy Mahajan, and Tian Tang, Meeting Experiments at the Diffraction Barrier: An *In-silico* Widefield Fluorescence Microscopy, *bioRxiv* **2021**, DOI: 10.1101/2021.03.02.433395. Copyright © 2021 Subhamoy Mahajan, Tian Tang, CC-BY-NC-ND 4.0 International license.

6.1. Introduction

Microscopy has enabled the exploration of tissues, cells, and its components,[1–3] structure and properties of biochemicals,[4–6] surface properties of materials,[7,8] and advances in many other fields. Fluorescence microscopy accounting for more than eighty percent of all microscopy images,[9] has enabled qualitative and quantitative analysis of live-cell processes.[10] Such analysis includes, but is not limited to the fluorophores' colocalization,[11–13] intensity,[14–16] count,[12,14] diameter,[12] area,[17] volume,[11] shape,[14] mean-squared displacement (MSD),[12] position,[11,12,18] diffusion,[12,15] speed,[11] *etc.*, which form a key understanding of various biological processes. Excellent reviews on different fluorescence microscopy techniques and their applications can be found in Lichtman and Conchello,[19] Stephens and Allan,[20] and Huang *et al.*[1]

The intrinsic limitation of fluorescence microscopy arises due to diffraction,[21] which is quantified by the effective point-spread-function (PSF) of a microscope. Standard fluorescence microscopy, such as widefield, experiences the resolution limit (i.e., diffraction barrier) of ~200 nm in the lateral direction when imaging cells because only visible spectra can be used to avoid

photodamage to the cells.[10,21] The images also suffer from further reduction in resolution due to the detection of out-of-focus fluorescence light. Like widefield, optical sectioning microscopy (OSM) techniques such as confocal,[22] light-sheet,[23] two-photon,[24] *etc.* are also diffraction-limited, but they improve the resolution of images by reducing the amount of out-of-focus fluorescence. Computational optical-sectioning microscopy (COSM) can be used, on a stack of 2D images obtained from different object focal planes (i.e., an equivalent 3D image), to further reduce the amount of out-of-focus fluorescence.[25,26] This is achieved by applying theoretical or experimental knowledge of the microscope's PSF and certain deconvolution algorithms.[26,27] COSM can also be used directly on 2D images and those obtained from widefield microscopy. Even with advances in OSM and COSM observing fine details in most cellular organelles seemed impossible until a few decades ago. Innovations in the field of super-resolution microscopy[28–32]ⁱ have broken the traditional diffraction barrier to achieve resolution as low as 20 nm. However, atomistic resolution on the order of angstroms still remains out of reach in fluorescence microscopy.

On the other hand, molecular simulations (MS) can probe biochemical systems with molecular,[33] sub-molecular,[34,35] or atomic[36] resolutions. Here for the first time, we present a framework for performing *in-silico* (i.e., virtual) fluorescence microscopy on MS and demonstrate its wide and versatile applications. It is critical to mention that in the literature the term “computational microscopy” refers to either a collection of MS not at all linked to microscopy,[37,38] or computational techniques applied on microscopy images to improve their quality and therefore not related to MS.[39,40] The framework proposed here is also distinct from existing studies in the literature that model microscopes, which focused on developing different PSFs in order to more accurately extract information from microscopy images.[41–44] Since the

development of COSM,[25,26] synthetic microscopy images have been generated by convoluting position density of objects placed in 2D or 3D space (considered “ground truth”) with a PSF to test analysis tools in experimental microscopy.[25,45–48] As such, the “ground truth” used are not realistic and do not possess any predictive power. Much more information can be gained by bridging the worlds of physics-based MS and experimental fluorescence microscopy, which to the best of our knowledge has not been investigated before.

The *in-silico* fluorescence microscopy created in this work can achieve functionalities similar to experimental fluorescence microscopies. As an example, widefield fluorescence microscopy follows simple optics, where the entire specimen is illuminated, imaging fluorescence from both in- and out-of-focus. Fluorophores with different emission peaks can be detected using one or multiple cameras.[49] While both color and monochrome cameras are available, monochrome cameras are often used because of their higher sensitivity and resolution.[49] Artificial colors can be digitally assigned to the monochrome images and superimposed to produce a colored microscopy image.[50] Images (monochrome or colored) taken at different times can be combined to form a microscopy video, to examine the temporal variations in the locations of different fluorophores and their colocalization. The optics of OSM is similar to widefield, where a thin optical section around the object focal plane is illuminated instead of the entire specimen and/or most of the out-of-focus fluorescence is blocked using an aperture. Similarly, the optics of some super-resolution microscopies such as stimulated emission depletion[29] (STED) microscopy resembles that of widefield but with a lower effective emission wavelength leading to higher resolution.[51]

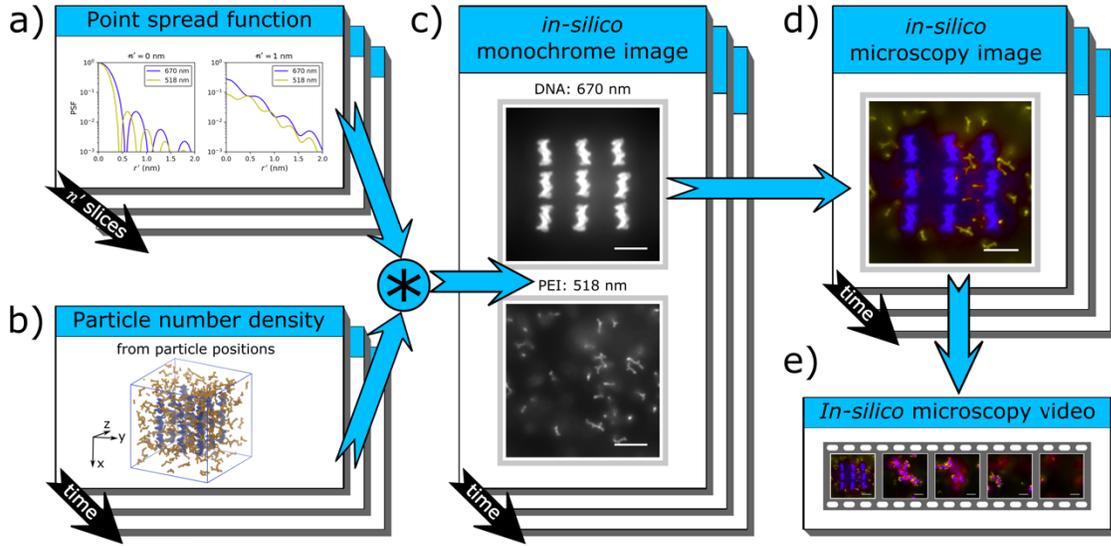


Figure 6.1: Framework of *in-silico* microscopy. (a) PSF with $I_0 = 1$, $\psi = 59.4^\circ$, $\mu = 1.51$, $P_{\ell'} = P_{m'} = P_{n'} = 25$ nm, $\Delta \ell' = \Delta m' = 0.1$ nm, $\Delta n' = 0.05$ nm, and $f_s = 530$ (Eq 6.1). In this work, PSF is calculated for $\lambda = 461$ (not shown), 518 and 670 nm (corresponding to emission peaks of DAPI, FITC, and Cy5), at different n' (black arrow). (b) A polyethylenimine (PEI)-DNA aggregation simulation[34] (**Appendix D.1.2**) is used as an MS specimen. The box represents the initial configuration of the MS, in the xyz coordinate system shown. DNAs are shown in blue and PEIs in orange. Particles in DNA and PEI molecules are assigned two different fluorophore types and their number density ($\rho_{\ell mn}$) is calculated based on their positions, at different simulation times (black arrow). (c) *In-silico* monochrome images for DNA (top) and PEI (bottom) are obtained (at different simulation times; black arrow) using the convolution (*) between PSF specified in (a) and $\rho_{\ell mn}$ (Eq 6.2). In the images shown, $t = 0$, n is taken to be the z -axis and the object focal plane is at $n_o = 12$ nm. DNA and PEI particles emit light with $(\lambda, I_0) = (670$ nm, 0.13) and (518 nm, 0.27) respectively. Bright-white and dim-diffused-white colors represent in-focus and out-of-focus fluorescence respectively. (d) An *in-silico* microscopy image is generated by assigning indigo hue to the top figure in (c) and yellow hue to the bottom figure in (c), and colors are mixed in the hue-saturation-value space using Eq 6.3-6.5. (e) Microscopy images generated at different simulation times can be combined into a microscopy video. Scale bars in (c-e), 5 nm.

Analogous principles are followed in the *in-silico* fluorescence microscopy with the additional advantage of tunable resolution and amount of out-of-focus fluorescence. Detailed

particle positions from an MS and PSF are used to generate fluorescence intensities, turning an MS “specimen” into *in-silico* monochrome images/videos, which are then superimposed with different hues to form colored *in-silico* microscopy images/videos (**Figure 6.1**). MS and experimental microscopy calculate physical properties using different principles, but they can now be directly compared through the *in-silico* images/videos. Since precise positions of particles are known through MS (the ground truth), a direct link between the position/motion of particles (**Figure 6.1b**) and *in-silico* microscopy images/videos (**Figure 6.1c-e**) can be established. This will not only allow cross-validation between experiments and MS, but also aid in the understanding of subcellular processes and mechanisms by combining knowledge from MS and experiments which may cover different length and time scales. Three-dimensional MS trajectories, although containing a large amount of quantitative information, are tedious (if not difficult) to view and analyze on a two-dimensional screen. The *in-silico* fluorescence microscopy presented here aims to provide a novel easy-to-use open-source visualization toolbox, which allows researchers to observe more by reducing the quantitative details.

6.2. Results

6.2.1. Setup of the *in-silico* microscope

A linear and lateral shift-invariant *in-silico* microscope (\mathcal{L}) is setup to observe an MS specimen (\mathcal{MS}) with an arbitrarily chosen right-handed rectangular coordinate system ℓmn , where ℓm forms the lateral plane and n is the optical axis (**Figure 6.2**). The microscope is focused on the object focal plane (\mathcal{F}_o in **Figure 6.2**) where $n = n_o$. Selected microscopy images generated with different n -axis and n_o are shown in **Figure 6.3a**. For a given ℓmn , images taken at different n_o provide insight on the 3D structure of the \mathcal{MS} .

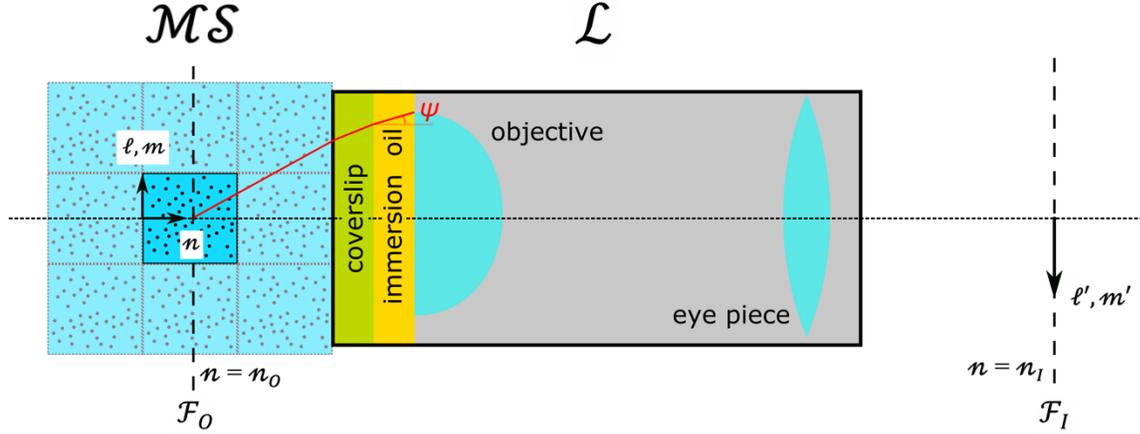


Figure 6.2: Schematic of the *in-silico* microscope. \mathcal{MS} is the MS specimen being viewed under the *in-silico* microscope \mathcal{L} . The central box in \mathcal{MS} is the original MS system, and the adjacent boxes with equal dimensions to the original MS system are its periodic images, if periodic boundary condition is applied. \mathcal{L} consists of a virtual cover slip, immersion oil, objective lens, and eyepiece. $\ell m n$ is a right-handed coordinate system of the MS, where n is the optical axis. \mathcal{F}_O and \mathcal{F}_I are the object and image focal planes respectively. The object focal plane is located at $n = n_o$ and image focal plane at $n = n_i$.

The image of \mathcal{MS} is produced in the image focal plane $n = n_i$ (\mathcal{F}_I in **Figure 6.2**), magnifying the \mathcal{MS} coordinates by $-M$; i.e., in-focus fluorophore particle with coordinates (ℓ_j, m_j, n_o) produces a focal spot at $(-M\ell_j, -Mm_j, n_i)$. An image coordinate system $\ell' m'$ is introduced which scales the ℓm coordinates by $-1/M$, such that the image coordinates of the focal spot $(-M\ell_j, -Mm_j, n_i)$ are given by $(\ell', m') = \left(\frac{-M\ell_j}{-M}, \frac{-Mm_j}{-M} \right) = (\ell_j, m_j)$. Fluorophore particles in \mathcal{MS} , both in- and out-of-focus, each generates an intensity profile around its own focal spot, which is characterized by the PSF. For the *in-silico* microscope with an ideal aberration-free high-magnification objective lens, the PSF is modeled using Eq 6.1.[42]

$$PSF(r', n') \equiv PSF(\ell', m', n') = I_0 \left| \frac{3}{2(1 - \cos^{3/2} \psi)} \int_0^\psi e^{-i\ell' \mu n' \cos \theta} J_0(\ell' \mu r' \sin \theta) \sin \theta \cos^{1/2} \theta d\theta \right|^2 \quad 6.1$$

Eq 6.1 describes the intensity produced at a point (ℓ', m') in \mathcal{F}_I by a fluorophore particle located at $(0, 0, n_o - n')$, a distance of n' away from \mathcal{F}_O . $r' = \sqrt{(\ell')^2 + (m')^2}$, i is the unit imaginary number, and J_0 is zeroth order Bessel function of the first kind. The wavenumber $k' = 2\pi f_s/\lambda$, where λ is the wavelength of the emitted light in vacuum and f_s a scaling factor introduced to tune the FWHM. I_0 is the maximum PSF intensity, $\psi = \sin^{-1}(NA/\mu)$ the maximum half-angle in the virtual immersion oil (**Figure 6.2**), NA the numerical aperture of the virtual objective lens, and μ the refractive index of the virtual immersion oil. The factor $3/2(1 - \cos^{3/2} \psi)$ is a normalization constant to ensure the maximum of PSF is I_0 for $n' = 0$. [42] In this form of the PSF, Gandy [42] assumed that the microscope is in design condition and the effective refractive index of the specimen is comparable to the immersion oil, which makes the PSF depth-invariant. Because of this, the location of the object plane (n_o) does not explicitly appear in Eq 6.1. The location of the image focal plane (n_i) also does not appear in Eq 6.1 because for a microscope, n_i solely depends on the design parameters (focal length of objective, eye piece, thickness of coverslip and immersion oil, tube length, *etc.*) and its effect is felt through the magnification M . For example, magnification in a simplified microscope with only an objective lens is given by the ratio between the distances of the image and object from the objective lens. Since the microscope is linear and lateral shift-invariant, the PSF defined for a fluorophore particle at $(0, 0, n_o - n')$ can be used to calculate the contribution of particles located elsewhere by a simple shift operation. It is worth noting that the *in-silico* microscope presented in this work is not limited to the PSF given in Eq 6.1. Other PSF available in the literature, such as those proposed by Gibson and Lanni, [41] Hell *et al.*, [44] can also be used. More discussions on this are given in **Appendix D.2**.

For computational efficiency, $PSF(\ell', m', n')$ is predetermined with $I_0 = 1$ at grid points within a cuboidal box that has a dimension of $(P_{\ell'}, P_{m'}, P_{n'})$ and constant grid spacing of $\Delta\ell'$, $\Delta m'$ and $\Delta n'$. Typical PSF curves are shown in **Figure 6.1a**. Increasing f_s will increase h' , which is equivalent to decreasing λ , compressing the PSF along r' axis (**Figure 6.1a**) and reducing the “spread” of the fluorescence intensity. This effectively decreases the FWHM making the *in-silico* microscopy images sharper (**Figure 6.3b**) and can be used to model super-resolution microscopy (See **Appendix D.3**). Increasing I_0 elongates the PSF along the vertical axis, causing the intensity of some local maxima in the PSF (**Figure 6.1a**) to exceed the minimum detection threshold of human vision. This makes the *in-silico* microscopy images brighter while increasing the radial distance over which each fluorophore particle contributes to the resultant image (**Figure 6.3c**). A concise guide on how to choose f_s and I_0 is available in **Appendix D.4**. $P_{\ell'}/2$ and $P_{m'}/2$ are respectively the maximum lateral distances in directions ℓ' and m' over which the fluorescence of a particle located at $(0, 0, n_0 - n')$ is calculated. In general, $P_{\ell'}$ and $P_{m'}$ should be large enough such that the PSF decays to zero within the box of dimension $(P_{\ell'}, P_{m'})$. $P_{n'}/2$ is the maximum distance of a fluorophore particle from \mathcal{F}_0 for which its fluorescence contribution is calculated, i.e., $P_{n'}$ is the thickness of the excited specimen around \mathcal{F}_0 . Therefore, decreasing $P_{n'}$ decreases the amount of out-of-focus fluorescence (**Figure 6.3d**) and can be used to model OSM (**Appendix D.3**).

6.2.2. Generating *in-silico* monochrome image

Particles in an MS are assigned to different fluorophore types, each emitting light at a specific wavelength λ . For each fluorophore type, the resultant fluorophore intensity I detected at \mathcal{F}_I , when the object focal plane is at n_0 , is calculated as the convolution between PSF (given n -axis, n_0 , ψ , f_s , λ , and I_0) and particle number density $\rho_{\ell mn} = \sum_{j=1}^N \delta(\ell - \ell_j, m - m_j, n - n_j)$

using Eq 6.2. The coordinates (ℓ_j, m_j, n_o) specify the position of the j^{th} fluorophore particle in the MS, N is the number of fluorophore particles in the MS, and δ is the Dirac delta function. The convolution operator is responsible for the lateral shift-operation on the PSF based on the position of each fluorophore particle. Therefore, convolution of the PSF with $\rho_{\ell mn}$ provides the resultant diffraction pattern produced by all fluorophores on the camera.

$$I(\ell', m', n_o) = PSF(\ell', m', n_o) * \rho_{\ell mn} = \sum_{j=1}^N PSF(\ell' - \ell_j, m' - m_j, n_o - n_j) \quad 6.2$$

Similar to PSF , for computational efficiency I is predetermined with $I_0 = 1$ at discrete points where PSF was evaluated. I values calculated from $I_0 = 1$ are hereafter denoted by I_1 . To generate images, I_1 is scaled with the actual chosen I_0 value and any intensity above 1 is treated as 1; i.e., $I = \min\{I_0 I_1, 1\}$. When I is rendered as an image for a fluorophore type, it is referred to as the *in-silico* monochrome image. Periodic boundary condition (PBC) can be applied while calculating I . The number of periodic images that contribute to I depends on the dimension of the box $(P_{\ell'}, P_{m'}, P_{n'})$ used to predetermine the PSF (**Appendix D.1.1**). Because the size of the MS can change over the course of the simulation, a white image frame larger than the MS is created and the monochrome image is scaled with respect to the white image frame before being placed at its center (**Appendix D.1.1**). This allows the comparison of images generated at different simulation times. An example of the 3D distribution of fluorophore particles and the corresponding *in-silico* monochrome images are shown in **Figure 6.1b-c**. The white image frame is highlighted in **Figure 6.1c-e** by adding a grey background.

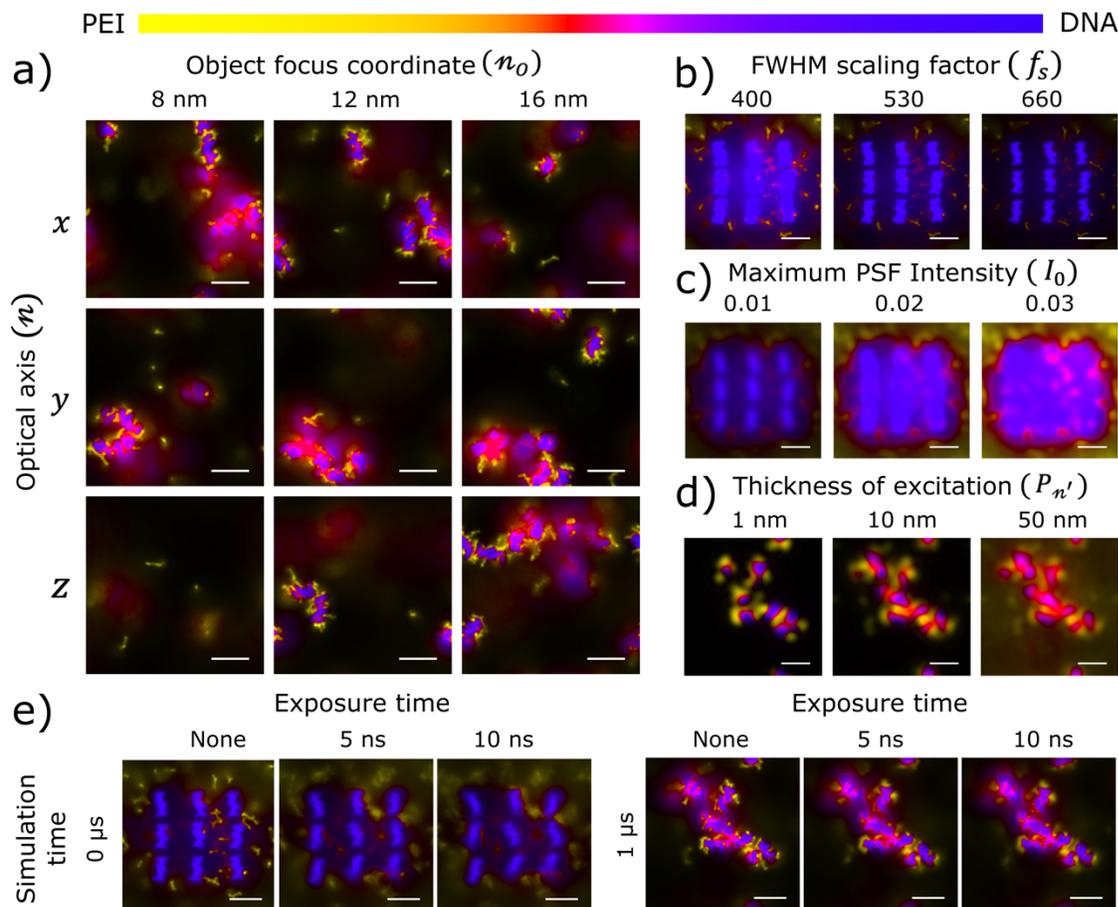


Figure 6.3: *In-silico* microscopy images generated with different parameters. MS on PEI-DNA aggregation[34] (Appendix D.1.2) is used as the specimen. Unless otherwise specified, the PSF is modeled with $\psi = 59.4^\circ$, $\mu = 1.51$, n is z-axis, $n_o = 12$ nm, $\Delta\ell' = \Delta m' = 0.1$ nm, $\Delta n' = 0.05$ nm, $P_{\ell'} = P_{m'} = P_{n'} = 25$ nm, $f_s = 530$, $(\lambda, I_0) = (670$ nm, 0.13) for DNA and (518 nm, 0.27) for PEI; DNA and PEI particles are assigned indigo and yellow hues respectively (colocalization color bar on the top); and no time-averaging is performed. (a) Images with different n and n_o at $t = 3$ μ s. (b) Images with different f_s , $t = 0$ μ s and $I_0 = 0.2$ for all particles. (c) Images with different I_0 at $t = 0$ μ s, $f_s = 130$. (d) Images with different $P_{n'}$ at $t = 1$ μ s, $f_s = 130$. I_0 for DNA and PEI are (0.04, 0.12) (left), (0.01, 0.03) (middle), and (0.008, 0.02) (right). (e) Images with different exposure time at $t = 0$ and 1 μ s. Scale bars, 5 nm.

6.2.3. Generating *In-silico* microscopy image and video

The final *in-silico* microscopy image is generated by selecting a color for each monochrome image and superimposing them. The colors are mixed in the hue-saturation-value (HSV) space. Each fluorophore type is assigned a hue, saturation of 1, and value equal to $I = \min\{I_0 I_1, 1\}$. The hue, saturation, and value of mixed color are given by Eq (6.3-6.5), where (H_j, V_j) are the hue and value of the j^{th} color, $\arg()$ returns the phase of a complex number, and $\max_n(V_j)$ represents the n^{th} largest V_j after sorting V_j of the colors being mixed (**Appendix D.5**). For example, if the colors being mixed have values 0.2, 0.5 and 0.5, then $\max_1(V_j) = \max_2(V_j) = 0.5$ and $\max_3(V_j) = 0.2$.

$$H_{mix} = \arg\left(\sum_{j=1}^N V_j e^{iH_j}\right) \quad 6.3$$

$$V_{mix} = \max_1(V_j) \quad 6.4$$

$$S_{mix} = 1 - \frac{\max_3(V_j)}{\max_1(V_j)} \quad 6.5$$

For two-color mixing the third largest V_j is zero, resulting in a fully saturated color (**Figure 6.4a**). When the third largest V_j is non-zero, it represents the mixing of three or more colors, and the mixed color is desaturated. A graphical representation of four-color mixing is shown in **Figure 6.4b**. A concise guide for choosing hues is provided in **Appendix D.6**. A typical *in-silico* microscopy image generated from a two-color mixture of indigo (assigned to **Figure 6.1c**, top) and yellow (assigned to **Figure 6.1c**, bottom) hues is shown in **Figure 6.1d**.

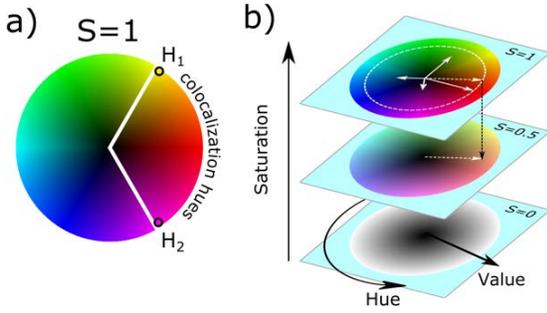


Figure 6.4: Demonstration of color mixing. (a) Two-color mixing always results in a fully saturated color. When hues H_1 and H_2 are chosen for two fluorophore types, all possible mixed colors (for different V_1 and V_2) is shown using the minor sector of the circle. (b) Demonstration of four-color mixing. The hue, saturation, and value are represented by the azimuthal angle, vertical distance and the radial distance respectively. Colors associated with all hue-value combinations are shown at three saturation levels 0, 0.5, and 1. The four colors being mixed have hues of 0° , 90° , 200° and 300° , and values of 0.8, 0.6, 0.4 and 0.3 respectively, which are shown using solid-white arrows. The hue and value of the mixed color are calculated using Eq 6.3 and 6.4 based on the sum of the four complex numbers $V_j e^{iH_j}$. The resultant complex number $V_{mix} e^{iH_{mix}}$ is shown by the dashed-white arrow in the $S = 1$ plane. The mixed color has a value of 0.8 and hue of 19.5° . The saturation of mixed color is 0.5 (Eq 6.5). The drop in saturation to the $S = 0.5$ plane is shown by the dashed-black arrow.

Existing color mixing techniques often use the RGB (red-green-blue) or CMY (cyan-magenta-yellow) color space. At most three fluorophore types can be superimposed in these methods and they can only be associated with the primary (in RGB) or secondary colors (in CMY). In the HSL (hue-saturation-luminance) color mixing scheme developed by Demanols and Davoust,[52] I for one fluorophore type can be associated with any fully saturated hue. However, this method cannot mix more than two hues because it does not follow the associative law; consequently, mixing more than two colors is order dependent. In contrast, the new color mixing scheme presented here is superior to previous methods because an arbitrary number of fully saturated hues can be mixed. This allows great flexibility in choosing hues for different

fluorophores, such as choosing color-safe colocalization hues for color-blind readers (**Appendix D.6**). Choice of non-standard colors has the added benefit of producing stronger color contrast in *in-silico* (**Figure 6.5**) and experimental (**Appendix D.7**) images. Even if the resultant V_{mix} is the same for different color combinations, the contrast in images can be different because the relative luminance[53] (brightness) is not the same for all hues. For example, relative luminance[53] is highest for yellow and lowest for blue, with yellow having ~ 10 times the relative luminance[53] of blue at the saturation of 1. For further discussion on color contrast and relative luminance,[53] see **Appendix D.8**.

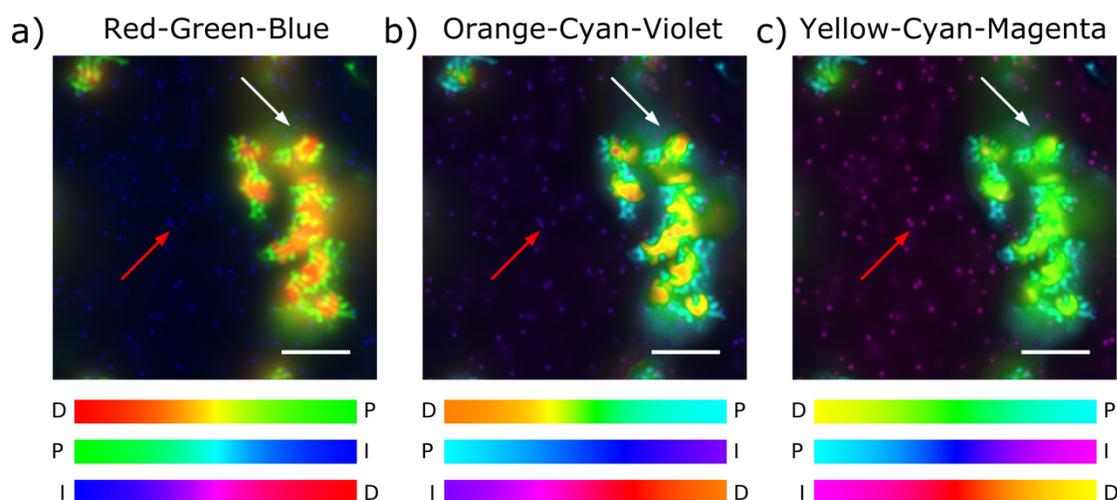


Figure 6.5: Images with different color combinations. (a) red-green-blue, (b) orange-cyan-violet, and (c) yellow-cyan-magenta (colocalization color bars below each subfigure; D: DNA, P: PEI, I: ions). MS on PEI-DNA aggregation[34] (**Appendix D.1.2**) at $t = 3 \mu\text{s}$ is used as the specimen. PSF is modeled with $\psi = 59.4^\circ$, $\mu = 1.51$, n is x -axis, $n_o = 4 \text{ nm}$, $\Delta\ell' = \Delta m' = 0.1 \text{ nm}$, $\Delta n' = 0.05 \text{ nm}$, $P_{\ell'} = P_{m'} = P_{n'} = 25 \text{ nm}$, $f_s = 530$, $(\lambda, I_o) = (670 \text{ nm}, 0.13)$ for DNA, $(518 \text{ nm}, 0.27)$ for PEI and $(461 \text{ nm}, 0.4)$ for ions. Visibility for ions over black (red arrow) and non-black (white arrow) backgrounds is dependent on the color combinations. Ion visibility for color combination of D-P-I follows yellow-cyan-magenta > orange-cyan-violet > red-green-blue over black background, and orange-cyan-violet > yellow-cyan-magenta > red-green-blue over non-black background. Overall orange-cyan-violet combination performs best among the three.

Time-averaged *in-silico* microscopy images can be generated by superimposing time-averaged *in-silico* monochrome images. The time over which average is performed represents an effective exposure time (**Figure 6.3e**, and **Appendix D.9**). As fluorophore particles move, a time-averaged image captures the motion blur arising from the particle's motion. When the particle's diffusion coefficient is high so is the motion blur and vice versa. Finally, multiple images generated at different simulation times, with or without time averaging, can be combined to create an *in-silico* microscopy video (**Figure 6.1e**).

6.2.4. Applications

In-silico microscopy is operated on MS data using physics-based PSF. Therefore, similar to experimental images, *in-silico* images should be treated as data[54] and both qualitative and quantitative analysis can be performed. More importantly, since MS has predictive power, MS data associated with *in-silico* images serves as the ground truth, enabling cross-validation, attainment of new knowledge, examination of existing analysis tools and creation of new analysis tools, to name a few. The wide and versatile application of the *in-silico* microscopy is demonstrated using the following three examples.

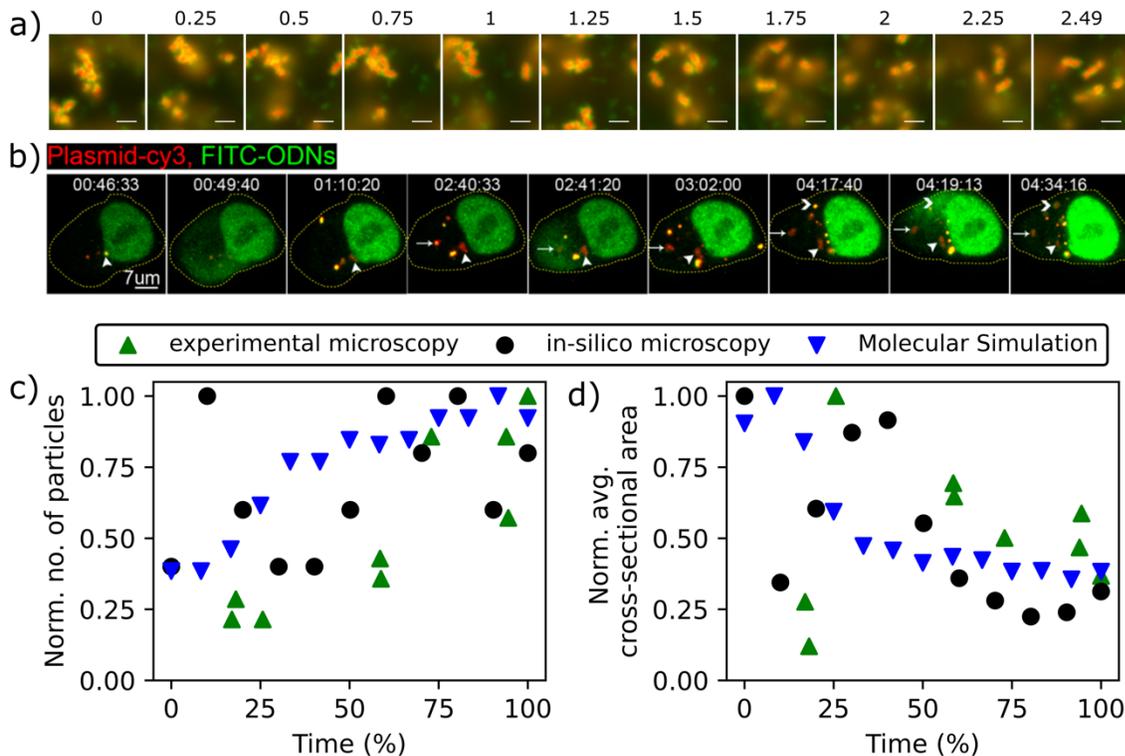


Figure 6.6: Gaining new information by comparing *in-silico* images, experimental images, and detailed MS data. (a) In-silico microscopy images from endosomal acidification simulation of PEI-DNA nanoparticles (Appendix D.1.2). PSF is modeled with $\psi = 59.4^\circ$, $\mu = 1.51$, n is z-axis, $n_o = 0$ nm, $\Delta\ell' = \Delta m' = 0.1$ nm, $\Delta n' = 0.05$ nm, $P_{\ell'} = P_{m'} = P_{n'} = 25$ nm, $f_s = 260$, $(\lambda, I_o) = (670$ nm, 0.04) for DNA and (518 nm, 0.06) for PEI; DNA and PEI particles are assigned red and green hues respectively. Above each image the simulation time is specified in microseconds, with endosomal acidification marking the start of the simulation. Scale bar, 5 nm. (b) Experimental fluorescence microscopy images by Rehman *et al.*[2] of PEI-mediate delivery of plasmid DNA (red) and oligonucleotides (green). The time is in hours:minutes:second format, and scale bars are absent in the original publication. [Reprinted with permission from Rehman *et al.*[2] Copyright © 2013 American Chemical Society]. (c) Number of particles and (d) cross-sectional area of nanoparticles, normalized by the corresponding maximum value from *in-silico* images, experimental images and MS data. Threshold of 0.588 and 0.196 was used for *in-silico* and experimental images respectively to calculate number and cross-sectional area. Cross-sectional area corresponding to MS was calculated from the square of the hydrodynamic radius (Appendix D.1.4).

First, new information can be gained by comparing data from MS, *in-silico* images and experimental images. In **Figure 6.6a** *in-silico* microscopy images are generated for an endosomal acidification simulation on PEI-DNA nanoparticles (see **Appendix D.1.2**). In **Figure 6.6b** PEI-based delivery of plasmid DNA (red) and oligonucleotides (green) inside HeLa cells is shown, which also undergoes endosomal acidification.[2] Qualitative similarity between **Figure 6.6a** and **Figure 6.6b** (appearance of smaller nanoparticles as time increases) can be established by visual inspection. Quantitatively, normalized number of particles (**Figure 6.6c**) and normalized average cross-sectional area of nanoparticles (**Figure 6.6d**) are calculated (**Appendix D.1.3, D.1.4**). In **Figure 6.6c**, a general increasing trend is observed for both experimental microscopy and MS. The trend from *in-silico* microscopy is less clear due to the large fluctuations caused by particles moving in and out of the focal plane of interest. Similar fluctuation is also observed in experimental microscopy at ~90% of the total time. In **Figure 6.6d**, a general decreasing trend of average cross-sectional area is observed for *in-silico* microscopy, experimental microscopy, and MS. A transient increase at short time (10-40% of the total time) is also observed for *in-silico* and experimental microscopy images. Clearly, qualitative and quantitative agreement between MS and experimental microscopy is achieved. In addition, detailed MS data has provided direct evidence for nanoparticle dissociation under endosomal acidification, which means dissociation of plasmid DNA from the nanoparticles likely occurred in Rehman *et al.*[2] although it was not reported in the original article. The comparison also suggests that an overall decrease in average cross-sectional area is a good indicator for nanoparticle dissociation, which can be used as a new analysis tool for experimental microscopy.

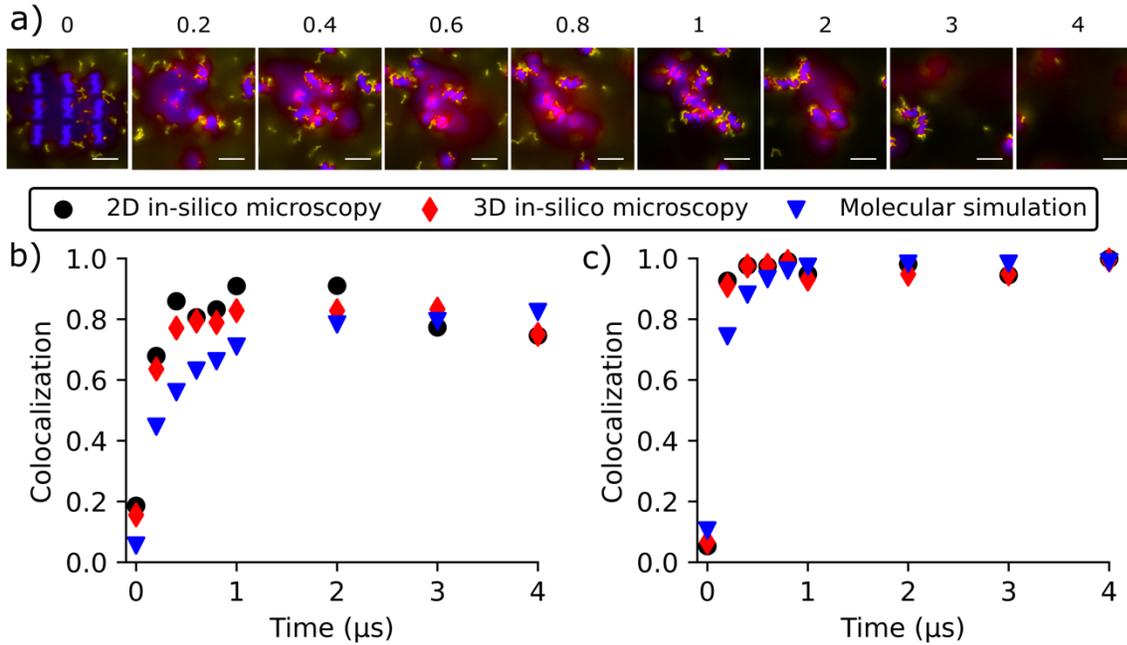


Figure 6.7: Bridging data obtained from MS and *in-silico* images. (a) *In-silico* microscopy images of PEI-DNA aggregation simulation[34] (**Appendix D.1.2**) at different time for visual colocalization analysis. The PSF is modeled with $\psi = 59.4^\circ$, $\mu = 1.51$, n is z-axis, $\Delta\ell' = \Delta m' = 0.1$ nm, $\Delta n' = 0.05$ nm, $P_{\ell'} = P_{m'} = P_{n'} = 25$ nm, $f_s = 530$, $(\lambda, I_0) = (670$ nm, 0.13) for DNA and (518 nm, 0.27) for PEI; DNA and PEI particles are assigned indigo and yellow hues respectively. 2D images (shown) were generated using $n_o = 12$ nm, and 3D images (not shown) were generated with $\Delta n_o = 1$ nm. The time, in microseconds, is shown above each *in-silico* image. Scale bar, 5 nm. Colocalization analysis was performed using Manders' coefficients[13] with JACoP plugin[55] in Fiji ImageJ[56]: (b) fraction of PEI colocalized with DNA (Eq 6.6), and (c) fraction of DNA colocalized with PEI (Eq 6.7) were calculated using a threshold of 0.2685 for both PEI and DNA. The MS values in (b) and (c) correspond to the fraction of PEI beads having at least one DNA bead within 0.81 nm, and fraction of DNA beads having at least one PEI bead within 0.85 nm.

Second, *in-silico* microscopy can bridge the analyses of data from MS and experimental microscopy, despite their distinct data acquisition techniques. **Figure 6.7a** shows the *in-silico* microscopy images of a PEI-DNA aggregation simulation at different time. The DNAs and PEIs begin unaggregated at 0 μ s, and aggregate over time. Complexation of DNA and PEI can be

observed as the color of DNAs and PEIs change from indigo and yellow to magenta. Most of the DNAs are complexed at 0.2 μs . Meanwhile, free PEIs not complexed with DNAs are observed to decrease. The lowest number of free PEI is observed at 1 μs , whereafter it does not change significantly. Quantitative data can be obtained using a dedicated colocalization analysis, such as Manders' coefficients[13] M_1 (**Figure 6.7b**) and M_2 (**Figure 6.7c**) given by Eq 6.6 and 6.7 respectively. I_{PEI} and I_{DNA} are monochrome intensities of PEI and DNA respectively at a given pixel, I_{thres} is a threshold intensity, $\mathbb{H}(\cdot)$ is the Heaviside step function, and summations are performed over all pixels.[55] Physically, M_1 is the fraction of total PEI fluorescence that is colocalized with DNAs having fluorescence above the threshold I_{thres} . Similarly, M_2 is the fraction total DNA fluorescence that is colocalized with PEIs having fluorescence above the threshold I_{thres} . The Manders' coefficients[13] are calculated from a single 2D image, as well as from multiple 2D images generated using $\Delta n_o = 1$ nm (i.e., 3D *in-silico* microscopy in **Figure 6.7b-c**).

$$M_1 = \frac{\sum I_{PEI} \mathbb{H}(I_{DNA} - I_{thres})}{\sum I_{PEI}} \quad 6.6$$

$$M_2 = \frac{\sum I_{DNA} \mathbb{H}(I_{PEI} - I_{thres})}{\sum I_{DNA}} \quad 6.7$$

Using detailed MS data, the fraction of PEI beads that have at least one DNA bead within 0.81 nm shows excellent agreement with M_1 (**Figure 6.7b**). The cutoff 0.81 nm is the sum between the first minimum in the PSF of PEI and the distance at which the PSF of DNA is equal to the threshold I_{thres} for $n' = 0$. Similarly, the fraction of DNA beads that have at least one PEI bead within 0.85 nm is equivalent to M_2 (**Figure 6.7c**), where 0.85 nm is the sum between the first minimum in the DNA PSF and the distance at which the PEI PSF is equal to the threshold I_{thres} for $n' = 0$ (**Appendix D.1.4**). **Figure 6.7b-c** shows that the equivalency of properties calculated

from MS and from *in-silico*, and hence experimental, microscopy images which suggests a way to bridge data analyses in these two completely different fields. The example here also demonstrates that it is feasible to estimate 3D contacts of molecules (and likely other properties) by visual inspection and quantitative analysis of 2D microscopy images (**Figure 6.7a**).

Third, the *in-silico* microscopy can be used as a platform to assess image analysis tools. In **Figure 6.8**, single-particle tracking is performed on *in-silico* images of a dimyristoyl phosphatidylcholine (DMPC) lipid bilayer MS[57] using TrackMate[58] in Fiji ImageJ[56] (**Appendix D.1.3**). The phosphorous atom in each DMPC molecule was selected to emit fluorescence (**Figure 6.8a**). TrackMate[58] detects several “tracks” of particles (“spots”), which are shown with different colors in **Figure 6.8b**. As the particles move across the periodic boundary in the MS, they are considered to be different by TrackMate[58] and therefore produce new “tracks”. The lateral mean-squared-displacement (MSD) calculated from “tracks” with more than 20 consecutive particle positions is shown in **Figure 6.8c** (**Appendix D.1.3**). Due to PBC, the “tracks” are short and the sampling is poor beyond 1 ns. Such sampling issues are not observed when lateral MSD is calculated directly from MS (with *msd* function in Gromacs) using positions of only phosphorous atoms and PBC (**Figure 6.8c**). The lateral MSD determined from MS and *in-silico* images are significantly different even at short time (< 0.4 ns). This difference is not caused by PBC because the sampling is sufficient at short time; rather it is likely due to incorrect detection of particles (for example arrows in **Figure 6.8a**) and “tracks”. Researchers therefore should exercise caution when directly applying particle tracking to experimental microscopy images in order to evaluate the mobility of the particles.

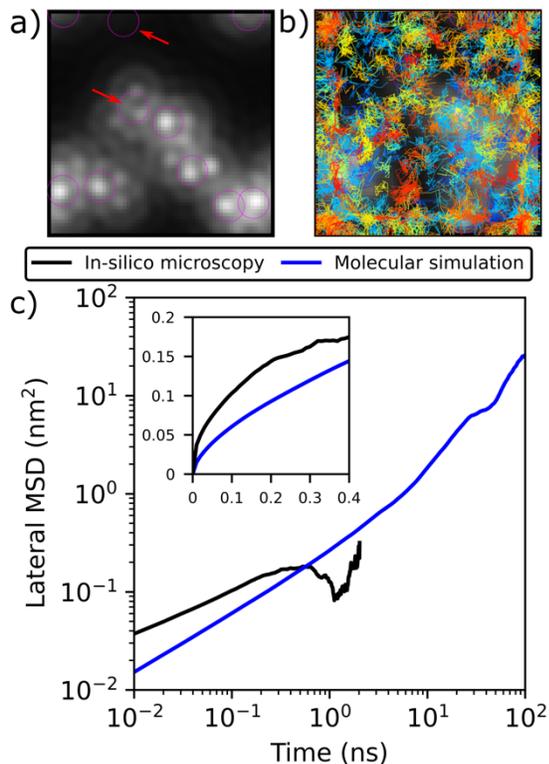


Figure 6.8: Single-particle tracking using *in-silico* microscopy on MS of DMPC lipid bilayer[57] (Appendix D.1.2). (a) *In-silico* microscopy image at time 0 μ s. The PSF is modeled with $\psi = 59.4^\circ$, $\mu = 1.51$, n is z -axis, $n_o = 3$ nm, $\Delta\ell' = \Delta m' = 0.1$ nm, $\Delta n' = 0.2$ nm, $P_{\ell'} = P_{m'} = 25$ nm, $P_{n'} = 2$ nm, $f_s = 530$, $(\lambda, I_o) = (670$ nm, 0.7) was used for phosphorous atoms in Dimyristoyl phosphatidylcholine (DMPC). Particles detected by TrackMate[58] plugin in Fiji ImageJ[56] is shown with purple circles. (b) Particle's tracks recognized by TrackMate.[58] A unique color is assigned to each track. (c) The lateral (in the ℓm plane) positions of phosphorous atoms as functions of time detected by TrackMate[58] were used to calculate the lateral mean squared displacement (MSD) and compared with the corresponding value from MS.

6.3. Discussions

A novel *in-silico* fluorescence microscopy is presented as an open-source toolbox (*in-silico-microscopy*, v1.2.2), which can work with different optical axis, object focal plane, exposure time, and color combinations; and generate images and videos with the desired resolution, contrast, brightness, and amount out-of-focus fluorescence. Images that resemble experimental widefield, OSM and super-resolution microscopy images can be generated by changing parameter such as

the amount of out-of-focus fluorescence and FWHM scaling factor (see **Appendix D.3** for more details). Properties calculated from different experimental microscopies will differ and corresponding *in-silico* microscopies can be used to quantify the difference. Other fluorescence microscopies can also be modeled by changing the PSF,[44,59,60] which is allowed by the modular nature of the toolbox. Moreover, PSF obtained from experiments can be implemented to model non-ideal objective lens with aberrations.[26] It is recognized that some fluorescence microscopies such as super-resolution optical fluctuation imaging[47] (SOFI), structure illuminated microscopy[61,62] (SIM), Förster resonance energy transfer[63,64] (FRET) microscopy, *etc.* involve mechanisms other than diffraction of light and cannot be modelled by simply changing the PSF. Nevertheless, our powerful toolbox has laid the foundation to generate *in-silico* fluorescence microscopy images that can be compared to those and other techniques such as X-ray microscopy (can be modelled with a PSF[65]), which would greatly enhance cross-validation and integration between simulations and experiments.

Generally, any image analysis developed for experimental microscopy is also applicable to *in-silico* microscopy, making the list of applications long. Three major categories of applications are presented in this work, namely generating new information by cross-comparison, determining equivalence of properties calculated from MS and experiments, and assessing image analysis tools or estimating errors using MS-based ground truth. For the former two, similar experimental and MS systems are required, which is possible in many fields of research such as gene delivery (shown in Results), biological motors,[66] role of lipid in membrane organization,[67] chromosomal dynamics,[68] *etc.* Error estimation for imaging analysis is a common practice in the literature where randomly placed objects are usually used as the ground truth.[45–48] The *in-silico* microscopy enables the use of MS data as the ground truth, which are physics-based, dynamic,

and more realistic than synthetic data. For better comparison with experiments, *in-silico* images can be generated with and without noise (**Appendix D.10**), before and after deconvolution (**Appendix D.11**), further demonstrating the versatility of the toolbox. Care should be taken while using existing image analysis tools, as these algorithms do not employ PBC which may be present in some MS. The existence of PBC in the MS does not affect analysis such as colocalization, fluorescence intensity and deconvolution (**Appendix D.11**), while it might affect the calculation of count, radius, area, volume and shape, if the fluorescence of the particles is divided across the periodic boundary (**Figure D.1**). In the toolbox v1.2.2 provided, count, area, and volume can be evaluated with the presence of PBC, while existing image analysis algorithms do not yet have this capability.

Software such as visual molecular dynamics (VMD) can visualize 3D structures, where the particles' colors are linearly blended with the background color according to the depth of the particles. As a result, it is difficult to visually estimate distances between particles. Furthermore, opacity of the particles limits the view of the particles behind, and increasing transparency worsens the depth perception. Typically, proximity of particles at a specific time is determined by rotating/translating the MS system and zooming into the area of interest, which is a time-consuming process. The toolbox presented here can be used as a standalone visualization tool for MS, where complex 3D data is condensed into 2D images. Key features from the plane of interest (object focal plane) is stored in high resolution, while the information away from the plane of interest is stored in low resolution. Proximity between two or more types of particles can be visualized using their colocalized hues, which is superior to depth perception in a software like VMD. For example, in **Figure 6.7** it is demonstrated that 3D contacts can be predicted with visual inspection of 2D images. Other plausible applications include the analysis of morphological

changes in molecules, aggregation or dissociation of molecules, multi-phase diffusion, *etc.* While *in-silico* images are corrupted by PSF, the amount of corruption can be controlled using the FWHM scaling factor or by changing the functional form of PSF, which does not need to model a microscope when used as a visualization tool. The visualization of colocalized fluorophores is improved using the newly developed color mixing scheme for both *in-silico* (**Figure 6.5**) and experimental (**Appendix D.7**) images, representing an added strength of the standalone visualization tool.

Although the toolbox is developed for MS such as molecular dynamics, microscopy images can be generated for other non-molecular simulations such as the finite element method (FEM), by treating the nodes in an FEM mesh as particles (the FEM nodes data must be converted to “gro” coordinate file format to be directly usable by the v1.2.2 of the toolbox). For continuum-level models such as Poisson-Boltzmann, where discrete position coordinates are unavailable, the general methodology demonstrated in this work can still be applied to create microscopy images by the convolution of PSF and particle densities in continuous form.

In the future, the *in-silico* fluorescence microscopy framework can be advanced to include the consideration of spectrum emitted by a light source, spectrum absorbed and emitted by fluorophores, probabilities of fluorophore excitation and emission, blinking rate of fluorophores, photobleaching, *etc.*

6.4. Conclusions

A novel open-source toolbox for performing *in-silico* (virtual) fluorescence microscopy on molecular simulations is presented. The toolbox has the ability to generate *in-silico* microscopy images that models different experimental fluorescence microscopy images. This brings the seemingly remote fields of microscopy and simulations together by generating new knowledge

from direct cross-validation, determining equivalence of properties extracted from MS and experimental images, assessing and developing algorithms for experimental image analysis, *etc.* It also paves the path for other *in-silico* microscopy techniques applied to molecular and non-molecular simulations. The work also reports the development of a new color mixing scheme, which allows the visualization of multi-fluorophore colocalization with arbitrary color assignment to the fluorophores. We expect this to be beneficial for *in-silico* microscopy, experimental fluorescence microscopy, and standalone visualization tools. We further hope this new open-source toolbox would spread the joy of creating and observing beautiful and powerful images, to theoreticians and experimentalists alike.

6.5. Data and code availability

The data that support the findings of this study are openly available in the University of Alberta Libraries Dataverse network at <https://doi.org/10.7939/DVN/F3JKZH>, Version 2.0. The open-source codes used in this work along with tutorials is maintained on GitHub, <https://github.com/subhamoymahajan/in-silico-microscopy>. This article is based on version v1.2.2.

6.6. Acknowledgements

We acknowledge the computing resources and technical support from Western Canada Research Grid (WestGrid). Financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC), Sadler Graduate Scholarship, Alberta Graduate Excellence Scholarship, R. R. Gilpin Memorial Scholarship, and Mitacs Globalink Graduate Fellowship is gratefully acknowledged.

6.7. References

- [1] Huang B., Bates M., Zhuang X. Super-Resolution Fluorescence Microscopy. *Annu. Rev. Biochem.* **2009**, 78, 993–1016.
- [2] Rehman Z.U., Hoekstra D., Zuhorn I.S. Mechanism of Polyplex- and Lipoplex-Mediated Delivery of Nucleic Acids: Real-Time Visualization of Transient Membrane Destabilization Without Endosomal Lysis. *ACS Nano* **2013**, 7, 3767–77.
- [3] Frederiksen R.S., Alarcon-Llado E., Krogstrup P., Bojarskaite L., Buch-Månson N., Bolinsson J., et al. Nanowire-Aperture Probe: Local Enhanced Fluorescence Detection for the Investigation of Live Cells at the Nanoscale. *ACS Photonics* **2016**, 3, 1208–16.
- [4] Erickson H.P. Size and Shape of Protein Molecules at the Nanometer Level Determined by Sedimentation, Gel Filtration, and Electron Microscopy. *Biol. Proc. Online* **2009**, 11, 32–51.
- [5] Jeynes J.C.G., Geraki K., Jeynes C., Zhaohong M., Bettiol A.A., Latorre E., et al. Nanoscale Properties of Human Telomeres Measured with a Dual Purpose X-ray Fluorescence and Super Resolution Microscopy Gold Nanoparticle Probe. *ACS Nano* **2017**, 11, 12632–40.
- [6] Ahmadi Y., Nord A.L., Wilson A.J., Hütter C., Schroeder F., Beeby M., et al. The Brownian and Flow-Driven Rotational Dynamics of a Multicomponent DNA Origami-Based Rotor. *Small* **2020**, 16, 2001855.
- [7] Melitz W., Shen J., Kummel A.C., Lee S. Kelvin Probe Force Microscopy and Its Application. *Surf. Sci. Rep.* **2011**, 66, 1–27.
- [8] Giessibl F.J. Advances in Atomic Force Microscopy. *Rev. Modern Phys.* **2003**, 75, 949–83.
- [9] Huang B., Babcock H., Zhuang X. Breaking the Diffraction Barrier: Super-Resolution Imaging of Cells. *Cell* **2010**, 143, 1047–58.
- [10] Taylor D.L., Wang Y.-L. Fluorescently Labelled Molecules as Probes of the Structure and Function of Living Cells. *Nature* **1980**, 284, 405–10.
- [11] Valm A.M., Cohen S., Legant W.R., Melunis J., Hershberg U., Wait E., et al. Applying Systems-Level Spectral Imaging and Analysis to Reveal the Organelle Interactome. *Nature* **2017**, 546, 162–7.
- [12] Adhikari S., Moscatelli J., Smith E.M., Banerjee C., Puchner E.M. Single-Molecule Localization Microscopy and Tracking With Red-Shifted States of Conventional BODIPY

- Conjugates in Living Cells. *Nat. Commun.* **2019**, 10, 3400.
- [13] Manders E.M.M., Verbeek F.J., Aten J.A. Measurement of Co-localization of Objects in Dual-Color Confocal Images. *J. Microsc.* **1993**, 169, 375–82.
- [14] Qamar S., Wang G., Randle S.J., Ruggeri F.S., Varela J.A., Lin J.Q., et al. FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation- π Interactions. *Cell* **2018**, 173, 720–34.
- [15] Capoulade J., Wachsmuth M., Hufnagel L., Knop M. Quantitative Fluorescence Imaging of Protein Diffusion and Interaction in Living Cells. *Nat. Biotechnol.* **2011**, 29, 835–9.
- [16] Chong S., Dugast-Darzacq C., Liu Z., Dong P., Dailey G.M., Cattoglio C., et al. Imaging Dynamic and Selective Low-Complexity Domain Interactions That Control Gene Transcription. *Science* **2018**, 361, eaar2555.
- [17] Mauthe M., Orhon I., Rocchi C., Zhou X., Luhr M., Hijlkema K.-J., et al. Chloroquine Inhibits Autophagic Flux by Decreasing Autophagosome-Lysosome Fusion. *Autophagy* **2018**, 14, 1435–55.
- [18] Xu Z., Song W., Crozier K.B. Direct Particle Tracking Observation and Brownian Dynamics Simulations of a Single Nanoparticle Optically Trapped by a Plasmonic Nanoaperture. *ACS Photonics* **2018**, 5, 2850–9.
- [19] Lichtman J.W., Conchello J. Fluorescence Microscopy. *Nat. Methods* **2005**, 2, 910–9.
- [20] Stephens D.J., Allan V.J. Light Microscopy Techniques for Live Cell Imaging. *Science* **2003**, 300, 82–6.
- [21] Abbe E. Beiträge zur Theorie des Mikroskops und der Mikroskopischen Wahrnehmung. *Arch. Für Mikroskopische Anat.* **1873**, 9, 413–68.
- [22] Webb R.H. Confocal Optical Microscopy. *Reports Prog. Phys.* **1996**, 59, 427–71.
- [23] Huisken J., Swoger J., del Bene F., Wittbrodt J., Stelzer E.H.K. Optical Sectioning Deep Inside Live Embryos by Selective Plane Illumination Microscopy. *Science* **2004**, 305, 1007–9.
- [24] Denk W., Strickler J.H., Webb W.W. Two-Photon Laser Scanning Fluorescence Microscopy. *Science* **1990**, 248, 73–6.
- [25] Preza C., Ollinger J.M., McNally J.G., Thomas Jr L.J. Point-Spread Sensitivity Analysis for Computational Optical-Sectioning Microscopy. *Micron Microsc. Acta* **1992**, 23, 501–13.
- [26] McNally J.G., Preza C., Conchello J.-A., Thomas L.J. Artifacts in Computational Optical-Sectioning Microscopy. *J. Opt. Soc. Am. A* **1994**, 11, 1056–67.

- [27] Sarder P., Nehorai A. Deconvolution Methods for 3-D Fluorescence Microscopy Images. *IEEE Signal Process. Mag.* **2006**, 23, 32–45.
- [28] Gustafsson M.G.L. Surpassing the Lateral Resolution Limit by a Factor of Two Using Structured Illumination Microscopy. *J. Microsc.* **2000**, 198, 82–7.
- [29] Hell S.W., Wichmann J. Breaking the Diffraction Resolution Limit by Stimulated Emission: Stimulated-Emission-Depletion Fluorescence Microscopy. *Opt. Lett.* **1994**, 19, 780–2.
- [30] Rust M.J., Bates M., Zhuang X. Sub-Diffraction-Limit Imaging by Stochastic Optical Reconstruction Microscopy (STORM). *Nat. Methods* **2006**, 3, 793–6.
- [31] Xu J., Tehrani K.F., Kner P. Multicolor 3D Super-Resolution Imaging by Quantum Dot Stochastic Optical Reconstruction Microscopy. *ACS Nano* **2015**, 9, 2917–25.
- [32] Liu W., Toussaint Jr. K.C., Okoro C., Zhu D., Chen Y., Kuang C., et al. Breaking the Axial Diffraction Limit: A Guide to Axial Super-Resolution Fluorescence Microscopy. *Laser Photon. Rev.* **2018**, 12, 1700333.
- [33] Yuan H., Huang C., Li J., Lykotrafitis G., Zhang S. One-Particle-Thick, Solvent-Free, Coarse-Grained Model for Biological and Biomimetic Fluid Membranes. *Phys. Rev. E* **2010**, 82, 011905.
- [34] Mahajan S., Tang T. Polyethylenimine-DNA Ratio Strongly Affects Their Nanoparticle Formation: A Large-Scale Coarse-Grained Molecular Dynamics Study. *J. Phys. Chem. B* **2019**, 123, 9629–40.
- [35] Uusitalo J.J., Ingólfsson H.I., Akhshi P., Tieleman D.P., Marrink S.J. Martini Coarse-Grained Force Field: Extension to DNA. *J. Chem. Theory Comput.* **2015**, 11, 3932–45.
- [36] Jung J., Nishima W., Daniels M., Bascom G., Kobayashi C., Adedoyin A., et al. Scaling molecular dynamics beyond 100,000 processor cores for large-scale biophysical simulations. *J. Comput. Chem.* **2019**, 40, 1919–30.
- [37] Ingólfsson H.I., Arnarez C., Periole X., Marrink S.J. Computational ‘Microscopy’ of Cellular Membranes. *J. Cell Sci.* **2016**, 129, 257–68.
- [38] Dror R.O., Dirks R.M., Grossman J.P., Xu H., Shaw D.E. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annu. Rev. Biophys.* **2012**, 41, 429–52.
- [39] Navruz I., Coskun A.F., Wong J., Mohammad S., Tseng D., Nagi R., et al. Smart-Phone Based Computational Microscopy Using Multi-Frame Contact Imaging on a Fiber-Optic Array. *Lab Chip* **2013**, 13, 4015–23.
- [40] Aidukas T., Eckert R., Harvey A.R., Waller L., Konda P.C. Low-Cost, Sub-Micron

- Resolution, Wide-Field Computational Microscopy Using Opensource Hardware. *Sci. Rep.* **2019**, *9*, 7457.
- [41] Gibson S.F., Lanni F. Experimental Test of an Analytical Model of Aberration in an Oil-Immersion Objective Lens Used in Three-Dimensional Light Microscopy. *J. Opt. Soc. Am. A* **1992**, *9*, 154–66.
- [42] Gandy R.O. Out-of-Focus Diffraction Patterns for Microscope Objectives. *Proc. Phys. Soc. B* **1954**, *67*, 825–31.
- [43] Richards B., Wolf E. Electromagnetic Diffraction in Optical Systems, II. Structure of the Image Field in an Aplanatic System. *Proc. R. Soc. London A* **1959**, *253*, 358–79.
- [44] Hell S., Reiner G., Cremer C., Stelzer E.H.K. Aberrations in Confocal Fluorescence Microscopy Induced by Mismatches in Refractive Index. *J. Microsc.* **1993**, *169*, 391–405.
- [45] Sehayek S., Gidi Y., Glembockyte V., Brandaõ H.B., François P., Cosa G., et al. A High-Throughput Image Correlation Method for Rapid Analysis of Fluorophore Photoblinking and Photobleaching Rates. *ACS Nano* **2019**, *13*, 11955–66.
- [46] Lehmussola A., Ruusuvoori P., Selinummi J., Huttunen H., Yli-Harja O. Computational Framework for Simulating Fluorescence Microscope Images With Cell Populations. *IEEE Trans. Med. Imaging* **2007**, *26*, 1010–6.
- [47] Girsault A., Lukes T., Sharipov A., Geissbuehler S., Leutenegger M., Vandenberg W., et al. SOFI Simulation Tool: A Software Package for Simulating and Testing Super-Resolution Optical Fluctuation Imaging. *PLoS One* **2016**, *11*, e0161602.
- [48] Venkataramani V., Herrmannsdörfer F., Heilemann M., Kuner T. SuReSim: Simulating Localization Microscopy Experiments From Ground Truth Models. *Nat. Methods* **2016**, *13*, 319–21.
- [49] Webb D.J., Brown C.M. Epi-Fluorescence Microscopy. In *Cell Imaging Techniques. Methods in Molecular Biology (Methods and Protocols)*, vol 931, 2nd ed. *Humana Press* **2012**, 29–59.
- [50] Xia J., Kim S.H.H., Macmillan S., Truant R. Practical Three Color Live Cell Imaging by Widefield Microscopy. *Biol. Proc. Online* **2006**, *8*, 63–8.
- [51] Hell S.W. Toward Fluorescence Nanoscopy. *Nat. Biotechnol.* **2003**, *21*, 1347–55.
- [52] Demandolx D., Davoust J. Multicolour Analysis and Local Image Correlation in Confocal Microscopy. *J. Microsc.* **1997**, *185*, 21–36.
- [53] Anderson M., Motta R., Chandrasekar S., Stokes M. Proposal for a Standard Default Color

- Space for the Internet—sRGB. *Color Imaging Conf.* **1996**, 238–45.
- [54] Cromeey D.W. Digital Images Are Data: And Should Be Treated as Such. *In* Cell Imaging Techniques. Methods in Molecular Biology (Methods and Protocols), vol 931, 2nd ed. *Humana Press* **2013**, 1–27.
- [55] Bolte S., Cordelières F.P. A Guided Tour Into Subcellular Colocalization Analysis in Light Microscopy. *J. Microsc.* **2006**, 224, 213–32.
- [56] Schindelin J., Arganda-Carreras I., Frise E., Kaynig V., Longair M., Pietzsch T., et al. Fiji: An Open-Source Platform for Biological-Image Analysis. *Nat. Methods* **2012**, 9, 676–82.
- [57] Miettinen M.S. Molecular Dynamics Simulation Trajectory of a Fully Hydrated DMPC Lipid Bilayer [Data set], <http://doi.org/10.5281/zenodo.51635>, **2013**.
- [58] Tinevez J.Y., Perry N., Schindelin J., Hoopes G.M., Reynolds G.D., Laplantine E., et al. Trackmate: An Open and Extensible Platform for Single-Particle Tracking. *Methods* **2017**, 115, 80–90.
- [59] Sergeeva E.A., Katichey A.R., Kirillin M.Y. Two-Photon Fluorescence Microscopy Signal Formation in Highly Scattering Media: Theoretical and Numerical Simulation. *Quantum Electron.* **2010**, 40, 1053–61.
- [60] Preza C., Snyder D.L., Conchello J.-A. Theoretical Development and Experimental Evaluation of Imaging Models for Differential-Interference-Contrast Microscopy. *J. Opt. Soc. Am. A* **1999**, 16, 2185–99.
- [61] Preza C. Simulating Structured-Illumination Microscopy in the Presence of Spherical Aberrations. *Proc. SPIE 7904, Three-Dimensional Multidimens. Microsc. Image Acquis. Process. XVIII*, **2011**, 79040D.
- [62] Shterman D., Gjonaj B., Bartal G. Experimental Demonstration of Multi Moiré Structured Illumination Microscopy. *ACS Photonics* **2018**, 5, 1898–902.
- [63] Uster P.S., Pagano R.E. Resonance Energy Transfer Microscopy: Observations of Membrane-Bound Fluorescent Probes in Model Membranes and in Living Cells. *J. Cell Biol.* **1986**, 103, 1221–34.
- [64] Corry B., Hurst A.C., Pal P., Nomura T., Rigby P., Martinac B. An Improved Open-Channel Structure of MscL Determined From FRET Confocal Microscopy and Simulation. *J. Gen. Physiol.* **2010**, 136, 483–94.
- [65] Selin M., Fogelqvist E., Holmberg A., Guttman P., Vogt U., Hertz H.M. 3D Simulation of the Image Formation in Soft X-Ray Microscopes. *Opt. Express* **2014**, 22, 30756–68.

- [66] Noji H., Yasuda R., Yoshida M., Kinosita Jr K. Direct Observation of the Rotation of F1-ATPase. *Nature* **1997**, 386, 299–302.
- [67] Veatch S.L., Keller S.L. Separation of Liquid Phases in Giant Vesicles of Ternary Mixtures of Phospholipids and Cholesterol. *Biophys. J.* **2003**, 85, 3074–83.
- [68] Sage D., Neumann F.R., Hediger F., Gasser S.M., Unser M. Automatic Tracking of Individual Fluorescence Particles: Application to the Study of Chromosome Dynamics. *IEEE Trans. Image Process.* **2005**, 14, 1372–83.

Chapter 7

Conclusion and Future Perspective*

*A version of this chapter's sections has been published in: Subhamoy Mahajan and Tian Tang, Martini Coarse-Grained Model for Polyethylenimine. *J. Comp. Chem* **2019**, 40, 607-618, with permission from *Wiley Periodicals, Inc.* Copyright © 2018,

Adapted with permission from Mahajan S., and Tang T., Polyethylenimine-DNA Ratio Strongly Affects Their Nanoparticle Formation: A Large-Scale Coarse-Grained Molecular Dynamics Study, *J. Phys. Chem. B* **2019**, 123, 9629-9640. Copyright © 2019 *American Chemical Society*, Subhamoy Mahajan and Tian Tang, Meeting Experiments at the Diffraction Barrier: An *In-silico* Widefield Fluorescence Microscopy, *bioRxiv* 2021, DOI:10.1101/2021.03.02.433395. Copyright © 2021 Subhamoy Mahajan and Tian Tang, CC-BY-NC-ND 4.0 International license.

7.1. Overall conclusions

An extensive literature review on experimental and theoretical studies of PEI-DNA gene delivery is provided in **Chapter 1**, which identified several knowledge gaps. Experiments have found properties of PEIs and PEI-DNA NPs to dictate the efficacy of gene delivery, but there is no consensus on the desirable properties that maximize the efficacy. Understanding of molecular mechanisms of gene delivery remains restricted due to the resolution limit of experimental techniques. While MD simulations can shed light on the mechanisms, existing simulations are performed at small length- and time-scales, which cannot be used to study large NPs relevant for

gene delivery. This dissertation created a framework for studying PEI-DNA gene delivery using large-scale CG-MD simulations. Particularly, the simulations focused on the formation of PEI-DNA NPs and structural changes in the NPs during the endosomal acidification step of gene delivery.

Chapter 3 developed the first CG-forcefield for PEI that can model PEI with different degrees of branching, molecular weight, and protonation ratios. Using eight 600 Da PEIs, ~46% of all relevant bonded interactions of PEI were parameterized, which can be used to model most branched structures with any molecular weight. This is a significant improvement over existing forcefields for linear PEI,[1,2] which typically contained less than ~16% of all relevant bonded interactions. The CG-forcefield consistently predicted the structural properties of PEI for different degrees of branching and protonation ratios, as well as in different water models and salt concentrations. Furthermore, the forcefield was compatible with the existing CG-DNA forcefield without any additional parameterization, allowing accurate simulation of NPs formed by PEIs and DNAs.

The 600 Da semi-linear PEI with 23% protonation ratio and 12 bp DNA were used to study NP formation at physiological pH in **Chapter 4**. The simulations contained 27 DNAs and as many as 270 PEIs, which were much larger than existing MD simulations[3] containing at most 4 DNAs and 28 PEIs. Since NP properties were strongly affected by the N/P ratio in experiments, the simulations were performed at five N/P ratios (1.2 to 6). Aggregation of PEIs and DNAs occurred in two distinct timescales: at low timescale, a similar number of PEIs bound to each DNA forming PEI-DNA complexes, and at large timescale, bound PEIs mediated the aggregation of PEI-DNA complexes. The shape of the NPs was linear at a low N/P ratio (< 4.7) due to end-to-end aggregation and branched at a high N/P ratio (> 4.7) due to side-by-side aggregation. The charge

of NPs was equal to the product of the number of DNAs in the NP and the charge of each PEI-DNA complex, which varied linearly with the N/P ratio. The steady-state average size (defined as the number of DNAs in the NP), and size distribution depended on an intrinsic electrostatic repulsion between NPs referred to as the specific repulsion, which was quadratic in the N/P ratio. At low and high N/P ratios, the specific repulsion was high leading to small average size and unimodal size distribution, whereas at moderate N/P ratios the specific repulsion was low leading to large average size and bimodal size distribution. The initial rate of NP growth in general increased with the N/P ratio, but a sharp decrease was observed when the mode of aggregation changed from end-to-end to side-by-side. The results on the shape, average charge, average size, and size distribution of NPs were in qualitative agreement with experiments, while other results such as the number of bridging or peripheral PEI, rate of NP growth, mode of aggregation, etc. provided additional insights that could not be extracted from experiments. Based on the aggregation mechanism, a two-step addition of PEIs was proposed as a strategy to control the shape and average size of the NPs and thereby the efficacy of cellular uptake.

The aggregation mechanism in **Chapter 4** suggests that the largest NPs are produced at a critical N/P ratio where the charge of the PEI-DNA complex is close to zero, i.e., with low specific repulsion. Since the charge of PEI is dependent on the pH, the charge of PEI-DNA complex and thereby the critical N/P ratio can be altered by the pH. Therefore, the aggregation mechanism suggests that endosomal acidification can alter the average size of NPs, which can either increase (further aggregation of NPs) if NPs are prepared at a low N/P ratio or decrease (dissociation of NPs) if prepared at a high N/P ratio. Motivated by this idea, **Chapter 5** explored the effects of endosomal acidification on the size and structure of NPs prepared at physiological pH. Indeed, NP aggregation simulations at physiological and endosomal pH, as well as NP acidification

simulations at an instant and slow acidification in **Chapter 5** confirmed this hypothesis. It is likely that the strong efficacy reported in experiments at a high N/P ratio is associated with the dissociation of NPs, which can enhance endosomal escape by increasing osmotic pressure and nuclear trafficking.

The dissociation mechanism was studied in **Chapter 5** using a novel principal transition diagram that demonstrated the NP's structural changes during dissociation, as well as a free energy diagram of PEIs that explained the forces behind the structural changes. Acidification of PEI primarily increased the repulsion between PEIs bridging the same DNA pair leading to weakening or even scission of the bridge. Bridge scission was more likely if the number of PEIs bridging a DNA pair was low. Furthermore, bridge scission was likely to occur near crowded DNAs experiencing strong DNA-DNA repulsion, and at the terminal of linear NPs where DNA-DNA repulsions were unbalanced. These mechanistic understanding suggested that a PEI with moderate molecular weight and degree of branching can enhance NP dissociation and thereby the efficacy of gene delivery.

Finally, in **Chapter 6**, a novel *in-silico* fluorescence microscopy technique was reported that allows direct comparison between molecular simulations and fluorescence microscopy experiments. The particle trajectories in molecular simulations are converted using a point-spread-function to model how the simulated particles would appear under a physical microscope. Monochrome *in-silico* images were first generated, and multiple monochrome *in-silico* images can be combined to produce a colored *in-silico* image. In this regard, a new color mixing scheme was derived in the HSV color space. Furthermore, the images can be corrupted by Gaussian and Poisson noise to model noise in physical microscopes and averaged over time to model the effects of exposure time. Since the *in-silico* microscopy models several aspects of an experimental

microscope, their images are equivalent, i.e., any property that can be calculated from experimental images can also be determined from *in-silico* images. Three broad applications *in-silico* microscopy were discussed in the chapter: generation of new information from cross-validation, determining equivalence of properties, and assessing/developing algorithms for microscopy image analysis. Further applications include the use of *in-silico* fluorescence microscopy as a visualization tool. For example, in **Chapter 5**, *in-silico* microscopy was used as a visualization tool to qualitatively analyze endosomal acidification, which correlated well with quantitative analysis of detailed MD data. Experimental fluorescence microscopy of gene delivery produces rich spatio-temporal data that contains several hidden dynamic properties, and *in-silico* microscopy would be an indispensable tool to extract, analyze and understand these properties.

7.2. Future perspectives

7.2.1. Extension of the PEI forcefield

The modeling methodology described in **Chapter 3** can be followed to extend the PEI forcefield, capturing all possible branched structures and protonation ratio relevant for gene delivery. Coarse-graining of a set of PEI structures would be required that contains all possible CG-bonded interactions that do not appear in the eight 600 Da PEIs studied. The PEI forcefield can be extended to model PEIs modified with other functional groups. For example, forcefield development of PEG-g-PEI would require additional parameterization of PEG chains and bonded parameters near the graft, whereas reparameterization of bonded interactions between PEI beads would not be required. It is worth noting that forcefield for linear PEG-g-PEI is available in the literature,[2] but not for branched ones. In general, the developed PEI forcefield would be compatible with other Martini forcefields, such as lipids, proteins, siRNA, etc., that are parameterized using the Martini methodology. However, extensive validations for nonbonded

interactions between PEI and the molecule would be required. If the AA- and CG-PMF profiles are dissimilar, parameterization of PEI's non-bonded interaction would be necessary.

7.2.2. Further explorations on nanoparticle formation

In **Chapters 4** and **5**, NP formation is studied using a 12 bp DNA and 600 Da semi-linear PEI. Additional aggregation mechanisms may be present when PEIs with higher molecular weight and different degrees of branching are studied. Furthermore, DNAs with a large number of base-pairs can involve condensation, which is not captured with 12 bp DNAs. Aggregation of such large molecules may involve analysis similar to **Chapter 4**, such as tracking the number of DNAs in each NP, number of PEIs with different roles, mode of aggregation, etc. However, additional indicators would be required to track the conformational state of large PEI and DNA molecules, as they aggregate, i.e., parameters such as radius of gyration of individual molecules. Sun *et al.*[4,5] observed PEI-DNA binding to be affected by the degree of branching only for high molecular weight PEI. Such an effect can be examined for large-scale PEI-DNA aggregation involving high molecular weight PEIs. Since PEIs used in experiments are inherently polydisperse, understanding their effect on NP aggregation would be relevant.

7.2.3. Further explorations on endosomal acidification

The effects of endosomal acidification can be dependent on the PEI's molecular weight, degree of branching, and polydispersity. The discussion of driving forces in **Chapter 5**, such as PEI-PEI repulsion, DNA-DNA repulsion, and PEI-DNA attraction should remain relevant when NPs are prepared with PEIs of different properties. However, the forces may not be sufficient for NP dissociation, in which case it would lead to NP swelling or shrinkage as discussed in **Chapter 5**. NP swelling is more likely for plasmid DNAs that have been observed to decondense in the cytoplasm.[6] **Chapter 5** assumes a fixed protonation state of PEI at physiological and endosomal

pH, and PEIs are protonated at random. However, the protonation state of PEI and its acidification are expected to be highly dynamic within an actual endosome. Accurate modeling of such processes would require advanced simulation techniques such as MC-MD,[7] constant pH MD,[8] etc. Since the entry of water in the endosome increase the osmotic pressure, an increase in the pressure of the MD simulation may be relevant to NP aggregation/dissociation.

7.2.4. Future of CG-MD simulations of PEI-DNA gene delivery

The results in **Chapters 4 and 5** demonstrate good qualitative agreement with the corresponding experiments. This shows the tremendous potential of CG-MD simulations in determining the structure-function relationship of PEI in the context of gene delivery. Prior to this dissertation, there were no large-scale CG-MD simulations in the literature that studied PEI-DNA gene delivery, and **Chapter 3-5** provides a framework for all future studies. Some AA-MD simulations have explored the interaction of PEI with model membranes.[9,10] These studies can be extended at the CG scale to study cellular uptake and endosomal escape of NPs with more realistic membranes. Since membranes have been extensively modeled in Martini,[11–13] simulation of NP-membrane interactions would only require the compatibility of PEI with existing Martini lipid and protein (if the membrane contains proteins) forcefield. MD simulations of nuclear trafficking are absent in the literature and experimental studies are limited. CG simulations of nuclear trafficking would require a detailed 3D structure and forcefield associated with the nuclear pore complex and its compatibility with our CG-PEI forcefield. Unlike PEI-membrane interactions, the study of nuclear trafficking would require more efforts in extending the CG forcefield developed in this work.

7.3. Reference

[1] Ziebarth J, Wang Y. Coarse-Grained Molecular Dynamics Simulations of DNA

- Condensation by Block Copolymer and Formation of Core - Corona Structures. *J Phys Chem B* 2010;114:6225–32.
- [2] Wei Z, Luijten E. Systematic coarse-grained modeling of complexation between small interfering RNA and polycations. *J Chem Phys* 2015;143:243146.
- [3] Sun C, Tang T, Uludağ H. Molecular dynamics simulations of PEI mediated DNA aggregation. *Biomacromolecules* 2011;12:3698–707.
- [4] Sun C, Tang T, Uludağ H, Cuervo JE. Molecular dynamics simulations of DNA/PEI complexes: Effect of PEI branching and protonation state. *Biophys J* 2011;100:2754–63.
- [5] Sun C, Tang T, Uludağ H. Molecular dynamics simulations for complexation of DNA with 2 kDa PEI reveal profound effect of PEI architecture on complexation. *J Phys Chem B* 2012;116:2405–13.
- [6] Rehman ZU, Hoekstra D, Zuhorn IS. Mechanism of polyplex- and lipoplex-mediated delivery of nucleic acids: Real-time visualization of transient membrane destabilization without endosomal lysis. *ACS Nano* 2013;7:3767–77.
- [7] Ziebarth JD, Wang Y. Understanding the protonation behavior of linear polyethylenimine in solutions through Monte Carlo simulations. *Biomacromolecules* 2010;11:29–38.
- [8] Lee MS, Salsbury Jr. FR, Brooks III CL. Constant-pH Molecular Dynamics Using Continuous Titration Coordinates. *Proteins Struct Funct Genet* 2004;56:738–52.
- [9] John S, Hussain M, Turabe U, Dhand C, Venkatesh M, Tze E, et al. Insight into membrane selectivity of linear and branched polyethylenimines and their potential as biocides for advanced wound dressings. *Acta Biomater* 2016;37:155–64.
- [10] Kostritskii AY, Kondinskaia DA, Nesterenko AM, Gurtovenko AA. Adsorption of Synthetic Cationic Polymers on Model Phospholipid Membranes: Insight from Atomic-Scale Molecular Dynamics Simulations. *Langmuir* 2016;32:10402–14.
- [11] Louhivuori M, Risselada HJ, van der Giessen E, Marrink SJ. Release of content through mechano-sensitive gates in pressurized liposomes. *Proc Natl Acad Sci* 2010;107:19856–60.
- [12] Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, De Vries AH. The MARTINI force field: Coarse grained model for biomolecular simulations. *J Phys Chem B* 2007;111:7812–24.
- [13] Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ. The MARTINI coarse-grained force field: Extension to proteins. *J Chem Theory Comput* 2008;4:819–34.

Unified Bibliography

- [1.1] Alberts B., Johnson A., Lewis J., Morgan D., Raff M., Roberts K., *et al.* Proteins. *In* Molecular Biology of The Cell 6th ed. *Garland Science* **2015**, 109–72.
- [1.2] Alberts B., Johnson A., Lewis J., Morgan D., Raff M., Roberts K., *et al.* DNA, Chromosome, and Genomics. *In* Molecular Biology of the Cell 6th ed. *Garland Science* **2015**, 173–236.
- [1.3] Gilmore I.R., Fox S.P., Hollins A.J., Sohail M., Akhtar S. The Design and Exogenous Delivery of siRNA for Post-transcriptional Gene Silencing. *J. Drug Target.* **2004**, 12, 315–40.
- [1.4] Scollay R. Gene Therapy: A Brief Overview of the Past, Present, and Future. *Ann. N. Y. Acad. Sci.* **2001**, 935, 26–30.
- [1.5] Bulcha J.T., Wang Y., Ma H., Tai P.W.L., Gao G. Viral vector platforms within the gene therapy landscape. *Signal Transduct. Target. Ther.* **2021**, 6, 53.
- [1.6] Terheggen H.G., Lowenthal A., Lavinha F., Colombo J.P., Rogers S. Unsuccessful Trial of Gene Replacement in Arginase Deficiency. *Eur. J. Pediatr.* **1975**, 119, 1–3.
- [1.7] Ramasamy T., Munusamy S., Ruttala H.B., Kim J.O. Smart Nanocarriers for the Delivery of Nucleic Acid-Based Therapeutics: A Comprehensive Review. *Biotechnol. J.* **2021**, 16, e1900408.
- [1.8] Evans C.H. The Vicissitudes of Gene Therapy. *Bone Jt. Res.* **2019**, 8, 469–71.
- [1.9] Bulaklak K., Gersbach C.A. The Once and Future Gene Therapy. *Nat. Commun.* **2020**, 11, 5820.
- [1.10] Ginn S.L., Amaya A.K., Alexander I.E., Edelstein M., Abedi M.R. Gene Therapy Clinical Trials Worldwide to 2017: An Update. *J. Gene Med.* **2018**, 20, e3015.
- [1.11] Hardee C.L., Arévalo-Soliz L.M., Hornstein B.D., Zechiedrich L. Advances in Non-Viral DNA Vectors for Gene Therapy. *Genes (Basel).* **2017**, 8, 65.
- [1.12] Boussif O., Lezoualc'h F., Zanta M.A., Mergny M.D., Scherman D., Demeneix B., *et al.* A Versatile Vector for Gene and Oligonucleotide Transfer into Cells in Culture and In Vivo: Polyethylenimine. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, 92, 7297–301.
- [1.13] Liu J., Song L., Liu S., Jiang Q., Liu Q., Li N., *et al.* A DNA-Based Nanocarrier for Efficient Gene Delivery and Combined Cancer Therapy. *Nano Lett.* **2018**, 18, 3328–34.

- [1.14] Alberts B., Johnson A., Lewis J., Morgan D., Raff M., Roberts K., *et al.* How Cells Read the Genome: From DNA to Protein. *In* Molecular Biology of The Cell 6th ed. *Garland Science* **2015**, 299–368.
- [1.15] Zhang X., Men K., Zhang Y., Zhang R., Yang L., Duan X. Local and Systemic Delivery of mRNA Encoding Survivin-T34A by Lipoplex for Efficient Colon Cancer Gene Therapy. *Int. J. Nanomedicine* **2019**, 14, 2733–51.
- [1.16] Ameres S.L., Martinez J., Schroeder R. Molecular Basis for Target RNA Recognition and Cleavage by Human RISC. *Cell* **2007**, 130, 101–12.
- [1.17] Wolff J.A., Malone R.W., Williams P., Chong W., Acsadi G., Jani A., *et al.* Direct Gene Transfer into Mouse Muscle In Vivo. *Science* **1990**, 247, 1465–8.
- [1.18] Yang N.S., Burkholder J., Roberts B., Martinell B., McCabe D. In Vivo and In Vitro Gene Transfer to Mammalian Somatic Cells by Particle Bombardment. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, 87, 9568–72.
- [1.19] Rols M.-P., Delteil C., Golzio M., Dumond P., Cros S., Teissie J. In Vivo Electrically Mediated Protein and Gene Transfer in Murine Melanoma. *Nat. Biotechnol.* **1998**, 16, 168–71.
- [1.20] Li S., Huang L. Nonviral Gene Therapy: Promises and Challenges. *Gene Ther.* **2000**, 7, 31–4.
- [1.21] Oh Y.-K., Suh D., Kim J.M., Choi H.-G., Shin K., Ko J.J. Polyethylenimine-Mediated Cellular Uptake, Nucleus Trafficking and Expression of Cytokine Plasmid DNA. *Gene Ther.* **2002**, 9, 1627–32.
- [1.22] Pollard H., Remy J.S., Loussouarn G., Demolombe S., Behr J.P., Escande D. Polyethylenimine but not Cationic Lipids Promotes Transgene Delivery to the Nucleus in Mammalian Cells. *J. Biol. Chem.* **1998**, 273, 7507–11.
- [1.23] Liu F., Song Y., Liu D. Hydrodynamics-Based Transfection in Animals by Systemic Administration of Plasmid DNA. *Gene Ther.* **1999**, 6, 1258–66.
- [1.24] Wang F., Qin Z., Lu H., He S., Luo J., Jin C., *et al.* Clinical Translation of Gene Medicine. *J. Gene Med.* **2019**, 21, e3108.
- [1.25] Shahryari A., Jazi M.S., Mohammadi S., Nikoo H.R., Nazari Z., Hosseini E.S., *et al.* Development and Clinical Translation of Approved Gene Therapy Products for Genetic Disorders. *Front. Genet.* **2019**, 10, 868.

- [1.26] Yin H., Kanasty R.L., Eltoukhy A.A., Vegas A.J., Dorkin J.R., Anderson D.G. Non-Viral Vectors for Gene-Based Therapy. *Nat. Rev. Genet.* **2014**, 15, 541–55.
- [1.27] Thomas C.E., Ehrhardt A., Kay M.A. Progress and Problems with the Use of Viral Vectors for Gene Therapy. *Nat. Rev. Genet.* **2003**, 4, 346–58.
- [1.28] Roberts D.M., Nanda A., Havenga M.J.E., Abbink P., Lynch D.M., Ewald B.A., *et al.* Hexon-Chimaeric Adenovirus Serotype 5 Vectors Circumvent Pre-Existing Anti-Vector Immunity. *Nature* **2006**, 441, 239–43.
- [1.29] Pack D.W., Hoffman A.S., Pun S., Stayton P.S. Design and Development of Polymers for Gene Delivery. *Nat. Rev. Drug Discov.* **2005**, 4, 581–93.
- [1.30] Behr J.-P. The Proton Sponge: A Trick to Enter Cells the Viruses did not Exploit. *Chimia (Aarau)*. **1997**, 51, 34–6.
- [1.31] Dick C.R., Ham G.E. Characterization of Polyethylenimine. *J. Macromol. Sci. Part A-Chem.* **1970**, 4, 1301–14.
- [1.32] Saegusa T., Ikeda H., Fujii H. Crystalline Polyethylenimine. *Macromolecules* **1972**, 5, 108–108.
- [1.33] Young R.J., Lovell P.A. Concepts and Nomenclature. *In* Introduction to Polymers 3rd ed. **2011**, 3–14.
- [1.34] Utsuno K., Uludağ H. Thermodynamics of Polyethylenimine-DNA Binding and DNA Condensation. *Biophys. J.* **2010**, 99, 201–7.
- [1.35] Kircheis R., Kichler A., Wallner G., Kursu M., Ogris M., Felzmann T., *et al.* Coupling of Cell-Binding Ligands to Polyethylenimine for Targeted Gene Delivery. *Gene Ther.* **1997**, 4, 409–18.
- [1.36] Godbey W.T., Wu K.K., Mikos A.G. Size Matters : Molecular Weight Affects the Efficiency of Poly(ethylenimine) as a Gene Delivery Vehicle. *J. Biomed. Mater. Res.* **1999**, 45, 268–75.
- [1.37] von Harpe A., Petersen H., Li Y., Kissel T. Characterization of Commercially Available and Synthesized Polyethylenimines for Gene Delivery. *J. Control. Release* **2000**, 69, 309–22.
- [1.38] Mattice W.L., Porter L.J. Molecular Weight Averages and ¹³C NMR Intensities Provide Evidence for Branching in Proanthocyanidin Polymers. *Phytochemistry* **1984**, 23, 1309–11.

- [1.39] Suh J., Paik H.J., Hwang B.K. Ionization of Poly(ethylenimine) and Poly(allylamine) at Various pH's. *Bioorg. Chem.* **1994**, 22, 318–27.
- [1.40] Tang M.X., Szoka F.C. The Influence of Polymer Structure on the Interactions of Cationic Polymers with DNA and Morphology of the Resulting Complexes. *Gene Ther.* **1997**, 4, 823–32.
- [1.41] Sun C., Tang T., Uludağ H. Molecular Dynamics Simulations for Complexation of DNA with 2 kDa PEI Reveal Profound Effect of PEI Architecture on Complexation. *J. Phys. Chem. B* **2012**, 116, 2405–13.
- [1.42] Kreiss P., Cameron B., Rangara R., Mailhe P., Aguerre-Charriol O., Airiau M., *et al.* Plasmid DNA Size Does Not Affect the Physicochemical Properties of Lipoplexes but Modulates Gene Transfer Efficiency. *Nucleic Acids Res.* **1999**, 27, 3792–8.
- [1.43] Dunlap D.D., Maggi A., Soria M.R., Monaco L. Nanoscopic Structure of DNA Condensed for Gene Delivery. *Nucleic Acids Res.* **1997**, 25, 3095–101.
- [1.44] Rehman Z.U., Hoekstra D., Zuhorn I.S. Mechanism of Polyplex- and Lipoplex-Mediated Delivery of Nucleic Acids: Real-Time Visualization of Transient Membrane Destabilization Without Endosomal Lysis. *ACS Nano* **2013**, 7, 3767–77.
- [1.45] Chemin I., Moradpour D., Wieland S., Offensperger W.-B., Walter E., Behr J.-P., *et al.* Liver-Directed Gene Transfer: A Linear Polyethylenimine Derivative Mediates Highly Efficient DNA Delivery to Primary Hepatocytes In Vitro and In Vivo. *J. Viral Hepat.* **1998**, 5, 369–75.
- [1.46] Goula D., Remy J.S., Erbacher P., Wasowicz M., Levi G., Abdallah B., *et al.* Size, Diffusibility and Transfection Performance of Linear PEI/DNA Complexes in the Mouse Central Nervous System. *Gene Ther.* **1998**, 5, 712–7.
- [1.47] Gautam A., Densmore C.L., Golunski E., Xu B., Waldrep J.C. Transgene Expression in Mouse Airway Epithelium by Aerosol Gene Therapy With PEI-DNA Complexes. *Mol. Ther.* **2001**, 3, 551–6.
- [1.48] Webb R.H. Confocal Optical Microscopy. *Reports Prog. Phys.* **1996**, 59, 427–71.
- [1.49] Huisken J., Swoger J., del Bene F., Wittbrodt J., Stelzer E.H.K. Optical Sectioning Deep Inside Live Embryos by Selective Plane Illumination Microscopy. *Science* **2004**, 305, 1007–9.
- [1.50] Denk W., Strickler J.H., Webb W.W. Two-Photon Laser Scanning Fluorescence

- Microscopy. *Science* **1990**, 248, 73–6.
- [1.51] Gustafsson M.G.L. Surpassing the Lateral Resolution Limit by a Factor of Two Using Structured Illumination Microscopy. *J. Microsc.* **2000**, 198, 82–7.
- [1.52] Hell S.W., Wichmann J. Breaking the Diffraction Resolution Limit by Stimulated Emission: Stimulated-Emission-Depletion Fluorescence Microscopy. *Opt. Lett.* **1994**, 19, 780–2.
- [1.53] Rust M.J., Bates M., Zhuang X. Sub-Diffraction-Limit Imaging by Stochastic Optical Reconstruction Microscopy (STORM). *Nat. Methods* **2006**, 3, 793–6.
- [1.54] Xu J., Tehrani K.F., Kner P. Multicolor 3D Super-Resolution Imaging by Quantum Dot Stochastic Optical Reconstruction Microscopy. *ACS Nano* **2015**, 9, 2917–25.
- [1.55] Liu W., Toussaint Jr. K.C., Okoro C., Zhu D., Chen Y., Kuang C., *et al.* Breaking the Axial Diffraction Limit: A Guide to Axial Super-Resolution Fluorescence Microscopy. *Laser Photon. Rev.* **2018**, 12, 1700333.
- [1.56] Webb D.J., Brown C.M. Epi-Fluorescence Microscopy. *In Cell Imaging Techniques. Methods in Molecular Biology (Methods and Protocols)*, vol 931, 2nd ed. *Humana Press* **2012**, 29–59.
- [1.57] Suh J., Wirtz D., Hanes J. Efficient Active Transport of Gene Nanocarriers to the Cell Nucleus. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, 100, 3878–82.
- [1.58] Mondal P.P. Temporal Resolution in Fluorescence Imaging. *Front. Mol. Biosci.* **2014**, 1, 11.
- [1.59] Abbe E. Beiträge zur Theorie des Mikroskops und der Mikroskopischen Wahrnehmung. *Arch. Für Mikroskopische Anat.* **1873**, 9, 413–68.
- [1.60] Khalil I.A., Kogure K., Akita H., Harashima H. Uptake Pathways and Subsequent Intracellular Trafficking in Nonviral Gene Delivery. *Pharmacol. Rev.* **2006**, 58, 32–45.
- [1.61] Rejman J., Bragonzi A., Conese M. Role of Clathrin- and Caveolae-Mediated Endocytosis in Gene Transfer Mediated by Lipo- and Polyplexes. *Mol. Ther.* **2005**, 12, 468–74.
- [1.62] Godbey W.T., Barry M.A., Saggau P., Wu K.K., Mikos A.G. Poly(Ethylenimine)-Mediated Transfection: A New Paradigm for Gene Delivery. *J. Biomed. Mater. Res.* **2000**, 51, 321–8.
- [1.63] Shete H.K., Prabhu R.H., Patravale V.B. Endosomal Escape: A Bottleneck in Intracellular

- Delivery. *J. Nanosci. Nanotechnol.* **2014**, 14, 460–74.
- [1.64] van der Aa M.A.E.M., Mastrobattista E., Oosting R.S., Hennink W.E., Koning G.A., Crommelin D.J.A. The Nuclear Pore Complex: The Gateway to Successful Nonviral Gene Delivery. *Pharm. Res.* **2006**, 23, 447–59.
- [1.65] Bahadur K.C.R., Landry B., Aliabadi H.M., Lavasanifar A., Uludağ H. Lipid Substitution on Low Molecular Weight (0.6-2.0 kDa) Polyethylenimine Leads to a Higher Zeta Potential of Plasmid DNA and Enhances Transgene Expression. *Acta Biomater.* **2011**, 7, 2209–17.
- [1.66] Abdallah B., Hassan A., Benoist C., Goula D., Behr J.P., Demeneix B.A. A Powerful Nonviral Vector for In Vivo Gene Transfer into the Adult Mammalian Brain: Polyethylenimine. *Hum. Gene Ther.* **1996**, 7, 1947–54.
- [1.67] Yue Y., Jin F., Deng R., Cai J., Dai Z., Lin M.C.M., *et al.* Revisit Complexation Between DNA and Polyethylenimine - Effect of Length of Free Polycationic Chains on Gene Transfection. *J. Control. Release* **2011**, 152, 143–51.
- [1.68] Boeckle S., von Gersdorff K., van der Pijpen S., Culmsee C., Wagner E., Ogris M. Purification of Polyethylenimine Polyplexes Highlights the Role of Free Polycations in Gene Transfer. *J. Gene Med.* **2004**, 6, 1102–11.
- [1.69] Fischer D., Bieber T., Li Y., Elsässer H.-P., Kissel T. A Novel Non-viral Vector for DNA Delivery Based on Low Molecular Weight, Branched Polyethylenimine: Effect of Molecular Weight on Transfection Efficiency and Cytotoxicity. *Pharm. Res.* **1999**, 16, 1273–9.
- [1.70] Werth S., Urban-Klein B., Dai L., Höbel S., Grzelinski M., Bakowsky U., *et al.* A Low Molecular Weight Fraction of Polyethylenimine (PEI) Displays Increased Transfection Efficiency of DNA and siRNA in Fresh or Lyophilized Complexes. *J. Control. Release* **2006**, 112, 257–70.
- [1.71] Bieber T., Elsässer H.P. Preparation of a Low Molecular Weight Polyethylenimine for Efficient Cell Transfection. *Biotechniques* **2001**, 30, 74–7, 80–1.
- [1.72] Wightman L., Kircheis R., Rössler V., Carotta S., Ruzicka R., Kurska M., *et al.* Different Behavior of Branched and Linear Polyethylenimine for Gene Delivery In Vitro and In Vivo. *J. Gene Med.* **2001**, 3, 362–72.
- [1.73] Choosakoonkriang S., Lobo B.A., Koe G.S., Koe J.G., Middaugh C.R. Biophysical

- Characterization of PEI/DNA Complexes. *J. Pharm. Sci.* **2003**, 92, 1710–22.
- [1.74] Krämer M., Stumbé J.-F., Grimm G., Kaufmann B., Krüger U., Weber M., *et al.* Dendritic Polyamines: Simple Access to New Materials With Defined Treelike Structures for Application in Nonviral Gene Delivery. *ChemBioChem* **2004**, 5, 1081–7.
- [1.75] Godbey W.T., Wu K.K., Mikos A.G. Poly(Ethylenimine)-Mediated Gene Delivery Affects Endothelial Cell Function and Viability. *Biomaterials* **2001**, 22, 471–80.
- [1.76] Putnam D., Gentry C.A., Pack D.W., Langer R. Polymer-Based Gene Delivery With Low Cytotoxicity by a Unique Balance of Side-Chain Termini. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, 98, 1200–5.
- [1.77] Morimoto K., Nishikawa M., Kawakami S., Nakano T., Hattori Y., Fumoto S., *et al.* Molecular Weight-Dependent Gene Transfection Activity of Unmodified and Galactosylated Polyethyleneimine on Hepatoma Cells and Mouse Liver. *Mol. Ther.* **2003**, 7, 254–61.
- [1.78] Kafil V., Omidi Y. Cytotoxic Impacts of Linear and Branched Polyethyleneimine Nanostructures in A431 Cells. *Bioimpacts* **2011**, 1, 23–30.
- [1.79] Erbacher P., Bettinger T., Belguise-Valladier P., Zou S., Coll J.L., Behr J.P., *et al.* Transfection and Physical Properties of Various Saccharide, Poly(Ethylene Glycol), and Antibody-Derivatized Polyethylenimines (PEI). *J. Gene Med.* **1999**, 1, 210–22.
- [1.80] Saminathan M., Antony T., Shirahata A., Sigal L.H., Thomas T., Thomas T.J. Ionic and Structural Specificity Effects of Natural and Synthetic Polyamines on the Aggregation and Resolubilization of Single-, Double-, and Triple-Stranded DNA. *Biochemistry* **1999**, 38, 3821–30.
- [1.81] dos Santos A.L.G., Bochot A., Tsapis N., Artzner F., Bejjani R.A., Thillaye-Goldenberg B., *et al.* Oligonucleotide–Polyethyleneimine Complexes Targeting Retinal Cells: Structural Analysis and Application to Anti-TGFbeta-2 Therapy. *Pharm. Res.* **2006**, 23, 770–81.
- [1.82] D’Andrea C., Pezzoli D., Malloggi C., Candeo A., Capelli G., Bassi A., *et al.* The Study of Polyplex Formation and Stability by Time-Resolved Fluorescence Spectroscopy of SYBR Green I-Stained DNA. *Photochem. Photobiol. Sci.* **2014**, 13, 1680–9.
- [1.83] Hou S., Ziebacz N., Wiczorek S.A., Kalwarczyk E., Sashuk V., Kalwarczyk T., *et al.* Formation and Structure of PEI/DNA Complexes: Quantitative Analysis. *Soft Matter*

- 2011**, 7, 6967–72.
- [1.84] Ko Y.T., Bickel U., Huang J. Polyethylenimine/Oligonucleotide Polyplexes Investigated by Fluorescence Resonance Energy Transfer and Fluorescence Anisotropy. *Oligonucleotides* **2011**, 21, 109–14.
- [1.85] Kunath K., von Harpe A., Fischer D., Petersen H., Bickel U., Voigt K., *et al.* Low-Molecular-Weight Polyethylenimine as a Non-Viral Vector for DNA Delivery: Comparison of Physicochemical Properties , Transfection Efficiency and In Vivo Distribution With High-Molecular-Weight Polyethylenimine. *J. Control. Release* **2003**, 89, 113–25.
- [1.86] Alberts B., Johnson A., Lewis J., Morgan D., Raff M., Roberts K., *et al.* Membrane Structure. *In* Molecular Biology of The Cell 6th ed. *Garland Science* **2015**, 565–96.
- [1.87] Paulos C.M., Reddy J.A., Leamon C.P., Turk M.J., Low P.S. Ligand Binding and Kinetics of Folate Receptor Recycling In Vivo: Impact on Receptor-Mediated Drug Delivery. *Mol. Pharmacol.* **2004**, 66, 1406–14.
- [1.88] Kopatz I., Remy J.-S., Behr J.-P. A Model for Non-Viral Gene Delivery: Through Syndecan Adhesion Molecules and Powered by Actin. *J. Gene Med.* **2004**, 6, 769–76.
- [1.89] Ogris M., Steinlein P., Kursa M., Mechtler K., Kircheis R., Wagner E. The Size of DNA/Transferrin-PEI Complexes Is an Important Factor for Gene Expression in Cultured Cells. *Gene Ther.* **1998**, 5, 1425–33.
- [1.90] Daniels T.R., Bernabeu E., Rodríguez J.A., Patel S., Kozman M., Chiappetta D.A., *et al.* The Transferrin Receptor and the Targeted Delivery of Therapeutic Agents Against Cancer. *Biochim. Biophys. Acta* **2012**, 1820, 291–317.
- [1.91] Ogris M., Brunner S., Schüller S., Kircheis R., Wagner E. PEGylated DNA/Transferrin–PEI Complexes: Reduced Interaction With Blood Components, Extended Circulation in Blood and Potential for Systemic Gene Delivery. *Gene Ther.* **1999**, 6, 595–605.
- [1.92] Hillaireau H., Couvreur P. Nanocarriers’ Entry Into the Cell: Relevance to Drug Delivery. *Cell. Mol. Life Sci.* **2009**, 66, 2873–96.
- [1.93] Conner S.D., Schmid S.L. Regulated Portals of Entry Into the Cell. *Nature* **2003**, 422, 37–44.
- [1.94] Bareford L.M., Swaan P.W. Endocytic Mechanisms for Targeted Drug Delivery. *Adv. Drug Deliv. Rev.* **2007**, 59, 748–58.

- [1.95] Lisanti M.P., Scherer P.E., Vidugiriene J., Tang Z., Hermanowski-Vosatka A., Tu Y.-H., *et al.* Characterization of Caveolin-rich Membrane Domains Isolated from an Endothelial-rich Source: Implications for Human Disease. *J. Cell Biol.* **1994**, 126, 111–26.
- [1.96] Chithrani B.D., Ghazani A.A., Chan W.C.W. Determining the Size and Shape Dependence of Gold Nanoparticle Uptake into Mammalian Cells. *Nano Lett.* **2006**, 6, 662–8.
- [1.97] Kettler K., Veltman K., van de Meent D., van Wezel A., Hendriks A.J. Cellular Uptake of Nanoparticles as Determined by Particle Properties, Experimental Conditions, and Cell Type. *Environ. Toxicol. Chem.* **2014**, 33, 481–92.
- [1.98] Wang C., Wang Y., Li Y., Bodemann B., Zhao T., Ma X., *et al.* A Nanobuffer Reporter Library for Fine-Scale Imaging and Perturbation of Endocytic Organelles. *Nat. Commun.* **2015**, 6, 8524.
- [1.99] Asokan A., Cho M.J. Exploitation of Intracellular pH Gradients in the Cellular Delivery of Macromolecules. *J. Pharm. Sci.* **2002**, 91, 903–13.
- [1.100] Sonawane N.D., Szoka F.C., Verkman A.S. Chloride Accumulation and Swelling in Endosomes Enhances DNA Transfer by Polyamine-DNA Polyplexes. *J. Biol. Chem.* **2003**, 278, 44826–31.
- [1.101] Kichler A., Leborgne C., Coeytaux E., Danos O. Polyethylenimine-Mediated Gene Delivery: A Mechanistic Study. *J. Gene Med.* **2001**, 3, 135–44.
- [1.102] Vermeulen L.M.P., Brans T., Samal S.K., Dubruel P., Demeester J., De Smedt S.C., *et al.* Endosomal Size and Membrane Leakiness Influence Proton Sponge-Based Rupture of Endosomal Vesicles. *ACS Nano* **2018**, 12, 2332–45.
- [1.103] Breunig M., Lungwitz U., Liebl R., Fontanari C., Klar J., Kurtz A., *et al.* Gene Delivery With Low Molecular Weight Linear Polyethylenimines. *J. Gene Med.* **2005**, 7, 1287–98.
- [1.104] Itaka K., Harada A., Yamasaki Y., Nakamura K., Kawaguchi H., Kataoka K. In Situ Single Cell Observation by Fluorescence Resonance Energy Transfer Reveals Fast Intracytoplasmic Delivery and Easy Release of Plasmid DNA Complexed With Linear Polyethylenimine. *J. Gene Med.* **2004**, 6, 76–84.
- [1.105] Benjaminsen R. V, Matthebjerg M.A., Henriksen J.R., Moghimi S.M., Andresen T.L. The Possible “ Proton Sponge ” Effect of Polyethylenimine (PEI) Does Not Include Change

- in Lysosomal pH. *Mol. Ther.* **2013**, 21, 149–57.
- [1.106] Klemm A.R., Young D., Lloyd J.B. Effects of Polyethyleneimine on Endocytosis and Lysosome Stability. *Biochem. Pharmacol.* **1998**, 56, 41–6.
- [1.107] Bieber T., Meissner W., Kostin S., Niemann A., Elsasser H.-P. Intracellular Route and Transcriptional Competence of Polyethylenimine-DNA Complexes. *J. Control. Release* **2002**, 82, 441–54.
- [1.108] Wentz S.R. Gatekeepers of the Nucleus. *Science* **2000**, 288, 1374–7.
- [1.109] Strambio-De-Castillia C., Niepel M., Rout M.P. The Nuclear Pore Complex: Bridging Nuclear Transport and Gene Regulation. *Nat. Rev. Mol. Cell Biol.* **2010**, 11, 490–501.
- [1.110] Panté N., Kann M. Nuclear Pore Complex Is Able to Transport Macromolecules with Diameters of ~39 nm. *Mol. Biol. Cell* **2002**, 13, 425–34.
- [1.111] Paine P.L., Moore L.C., Horowitz S.B. Nuclear Envelope Permeability. *Nature* **1975**, 254, 109–14.
- [1.112] Brunner S., Sauer T., Carotta S., Cotten M., Saltik M., Wagner E. Cell Cycle Dependence of Gene Transfer by Lipoplex, Polyplex and Recombinant Adenovirus. *Gene Ther.* **2000**, 7, 401–7.
- [1.113] Andersen H., Parhamifar L., Hunter A.C., Shahin V., Moghimi S.M. AFM Visualization of Sub-50 nm Polyplex Disposition to the Nuclear Pore Complex Without Compromising the Integrity of the Nuclear Envelope. *J. Control. Release* **2016**, 244, 24–9.
- [1.114] Bertschinger M., Backliwal G., Schertenleib A., Jordan M., Hacker D.L., Wurm F.M. Disassembly of Polyethylenimine-DNA Particles In Vitro: Implications for Polyethylenimine-Mediated DNA Delivery. *J. Control. Release* **2006**, 116, 96–104.
- [1.115] Neamark A., Suwantong O., Bahadur R.K.C., Hsu C.Y.M., Supaphol P., Uludağ H. Aliphatic Lipid Substitution on 2 kDa Polyethylenimine Improves Plasmid Delivery and Transgene Expression. *Mol. Pharm.* **2009**, 6, 1798–815.
- [1.116] Kwok A., Hart S.L. Comparative Structural and Functional Studies of Nanoparticle Formulations for DNA and siRNA Delivery. *Nanomedicine* **2011**, 7, 210–9.
- [1.117] Ziebarth J., Wang Y. Molecular Dynamics Simulations of DNA-Polycation Complex Formation. *Biophys. J.* **2009**, 97, 1971–83.
- [1.118] Beu T.A., Farcaş A. CHARMM Force Field and Molecular Dynamics Simulations of Protonated Polyethylenimine. *J. Comput. Chem.* **2017**, 38, 2335–48.

- [1.119] Beu T.A., Ailenei A.-E., Farcaş A. CHARMM Force Field for Protonated Polyethyleneimine. *J. Comput. Chem.* **2018**, 39, 2564–75.
- [1.120] Wei Z., Luijten E. Systematic Coarse-Grained Modeling of Complexation Between Small Interfering RNA and Polycations. *J. Chem. Phys.* **2015**, 143, 243146.
- [1.121] Sun C., Tang T., Uludağ H., Cuervo J.E. Molecular Dynamics Simulations of DNA/PEI Complexes: Effect of PEI Branching and Protonation State. *Biophys. J.* **2011**, 100, 2754–63.
- [1.122] Ziebarth J.D., Wang Y. Understanding the Protonation Behavior of Linear Polyethylenimine in Solutions Through Monte Carlo Simulations. *Biomacromolecules* **2010**, 11, 29–38.
- [1.123] Jorge A.F., Dias R.S., Pais A.A.C.C. Enhanced Condensation and Facilitated Release of DNA Using Mixed Cationic Agents: A Combined Experimental and Monte Carlo Study. *Biomacromolecules* **2012**, 13, 3151–61.
- [1.124] Choudhury C.K., Roy S. Structural and Dynamical Properties of Polyethylenimine in Explicit Water at Different Protonation States : A Molecular Dynamics Study. *Soft Matter* **2013**, 9, 2269–81.
- [1.125] Kondinskaia D.A., Kostritskii A.Y., Nesterenko A.M., Antipina A.Y., Gurtovenko A.A. Atomic-Scale Molecular Dynamics Simulations of DNA-Polycation Complexes: Two Distinct Binding Patterns. *J. Phys. Chem. B* **2016**, 120, 6546–54.
- [1.126] Sun C., Tang T., Uludağ H. Molecular Dynamics Simulations of PEI Mediated DNA Aggregation. *Biomacromolecules* **2011**, 12, 3698–707.
- [1.127] Bagai S., Sun C., Tang T. Potential of Mean Force of Polyethylenimine-Mediated DNA Attraction. *J. Phys. Chem. B* **2013**, 117, 49–56.
- [1.128] Fox S.J., Fazil M.H.U.T., Dhand C., Venkatesh M., Goh E.T.L., Harini S., *et al.* Insight Into Membrane Selectivity of Linear and Branched Polyethylenimines and Their Potential as Biocides for Advanced Wound Dressings. *Acta Biomater.* **2016**, 37, 155–64.
- [1.129] Kostritskii A.Y., Kondinskaia D.A., Nesterenko A.M., Gurtovenko A.A. Adsorption of Synthetic Cationic Polymers on Model Phospholipid Membranes: Insight from Atomic-Scale Molecular Dynamics Simulations. *Langmuir* **2016**, 32, 10402–14.
- [1.130] Zhang S., Gao H., Bao G. Physical Principles of Nanoparticle Cellular Endocytosis. *ACS Nano* **2015**, 9, 8655–71.

- [1.131] Yang S., May S. Release of Cationic Polymer-DNA Complexes From the Endosome: A Theoretical Investigation of the Proton Sponge Hypothesis. *J. Chem. Phys.* **2008**, 129, 185105.
- [1.132] Freeman E.C., Weiland L.M., Meng W.S. Modeling the Proton Sponge Hypothesis: Examining Proton Sponge Effectiveness for Enhancing Intracellular Gene Delivery Through Multiscale Modeling. *J. Biomater. Sci. Polym. Ed.* **2013**, 24, 398–416.
- [1.133] Zhou J., Yockman J.W., Kim S.W., Kern S.E. Intracellular Kinetics of Non-Viral Gene Delivery Using Polyethylenimine Carriers. *Pharm. Res.* **2007**, 24, 1079–87.
- [1.134] Grønbech-Jensen N., Mashl R.J., Bruinsma R.F., Gelbart W.M. Counterion-Induced Attraction between Rigid Polyelectrolytes. *Phys. Rev. Lett.* **1997**, 78, 2477–80.
- [1.135] Noid W.G. Perspective: Coarse-Grained Models for Biomolecular Systems. *J. Chem. Phys.* **2013**, 139, 090901.
- [2.1] Huang K. A Macroscopic View of Matter. *In* Introduction to Statistical Physics 2nd ed. *CRC Press* **2010**, 1–14.
- [2.2] Tuckerman M.E. Theoretical Foundations of Classical Statistical Mechanics. *In* Statistical Mechanics: Theory and Molecular Simulation. *Oxford University Press* **2010**, 53–73.
- [2.3] Huang K. The Statistical Approach. *In* Introduction to Statistical Physics 2nd ed. *CRC Press* **2010**, 65–82.
- [2.4] Landau L.D., Lifshitz E.M., Pitaevskii L.P. The Fundamental Principles of Statistical Physics. *In* Course of Theoretical Physics Vol. 5, Statistical Physics, 2nd ed. *Pergamon Press* **1969**, 1–32.
- [2.5] Tuckerman M.E. Monte Carlo. *In* Statistical Mechanics: Theory and Molecular Simulation. *Oxford University Press* **2010**, 277–311.
- [2.6] Allen M.P., Tildesley D.J. Molecular Dynamics. *In* Computer Simulation of Liquids. *Clarendon Press* **1989**, 71–109.
- [2.7] Goldstein H. The Hamilton Equations of Motion. *In* Classical Mechanics 2nd ed. *Addison-Wesley* **1980**, 339–77.
- [2.8] Lindahl E., Hess B., van der Spoel D. GROMACS 3.0: A Package for Molecular Simulation and Trajectory Analysis. *J. Mol. Model.* **2001**, 7, 306–17.

- [2.9] Allen M.P., Tildesley D.J. Introduction. *In Computer Simulation of Liquids. Clarendon Press* **1989**, 1–32.
- [2.10] Hoover W.G. Historical Development and Scope. *In Molecular Dynamics. Lecture Notes in Physics*, vol 258. *Springer-Verlag* **1986**, 1–41.
- [2.11] Abraham M.J., van der Spoel D., Lindahl E., Hess B. Introduction. *In GROMACS User Manual Version 5.1.4.* **2016**, 1–6.
- [2.12] Verlet L. Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, 159, 98–103.
- [2.13] Swope W.C., Andersen H.C., Berens P.H., Wilson K.R. A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. *J. Chem. Phys.* **1982**, 76, 637–49.
- [2.14] Abraham M.J., van der Spoel D., Lindahl E., Hess B. Algorithms. *In GROMACS User Manual Version 5.1.4.* **2016**, 11–66.
- [2.15] Hess B., Bekker H., Berendsen H.J.C., Fraaije J.G.E.M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, 18, 1463–72.
- [2.16] Allen M.P., Tildesley D.J. Statistical Mechanics. *In Computer Simulation of Liquids. Clarendon Press* **1989**, 33–70.
- [2.17] Bussi G., Donadio D., Parrinello M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, 126, 014101.
- [2.18] Nosé S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, 52, 255–68.
- [2.19] Hoover W.G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, 31, 1695–7.
- [2.20] Berendsen H.J.C., Postma J.P.M., van Gunsteren W.F., DiNola A., Haak J.R. Molecular Dynamics With Coupling to an External Bath. *J. Chem. Phys.* **1984**, 81, 3684–90.
- [2.21] Andersen H.C. Molecular Dynamics Simulations at Constant Pressure and/or Temperature. *J. Chem. Phys.* **1980**, 72, 2384–93.
- [2.22] Parrinello M., Rahman A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, 52, 7182–90.
- [2.23] Abraham M.J., van der Spoel D., Lindahl E., Hess B. Interaction Functions and Forcefields. *In GROMACS User Manual Version 5.1.4.* **2016**, 67–118.

- [2.24] Lennard-Jones J.E. The Equation of State of Gases and Critical Phenomena. *Physica* **1937**, 4, 941–56.
- [2.25] Israelachvili J.N. Interactions Involving Polar Molecules. *In* Intermolecular and Surface Forces 3rd ed. *Academic Press* **2011**, 71–90.
- [2.26] Israelachvili J.N. Van der Waals Forces. *In* Intermolecular and Surface Forces 3rd ed. *Academic Press* **2011**, 107–32.
- [2.27] Israelachvili J.N. Thermodynamics and Statistical Aspects of Intermolecular Forces. *In* Intermolecular and Surface Forces 3rd ed. *Academic Press* **2011**, 23–52.
- [2.28] Allen M.P., Tildesley D.J. Some Tricks of the Trade. *In* Computer Simulation of Liquids. *Clarendon Press* **1989**, 140–81.
- [2.29] Barker J.A., Watts R.O. Monte Carlo Studies of the Dielectric Properties of Water-Like Models. *Mol. Phys.* **1973**, 26, 789–92.
- [2.30] Essmann U., Perera L., Berkowitz M.L., Darden T., Lee H., Pedersen L.G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, 103, 8577–93.
- [2.31] Tuckerman M.E. Evaluation of Energies and Forces. *In* Statistical Mechanics: Theory and Molecular Simulation. *Oxford University Press* **2010**, 652–62.
- [2.32] Noid W.G. Perspective: Coarse-Grained Models for Biomolecular Systems. *J. Chem. Phys.* **2013**, 139, 090901.
- [2.33] Beu T.A., Ailenei A.-E., Farcaş A. CHARMM Force Field for Protonated Polyethyleneimine. *J. Comput. Chem.* **2018**, 39, 2564–75.
- [2.34] Marrink S.J., de Vries A.H., Mark A.E. Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B* **2004**, 108, 750–60.
- [2.35] Tuckerman M.E. The Canonical Ensemble. *In* Statistical Mechanics: Theory and Molecular Simulation. *Oxford University Press* **2010**, 133–213.
- [2.36] Kirkwood J.G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, 3, 300–13.
- [2.37] Torrie G.M., Valleau J.P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, 23, 187–99.
- [2.38] Kumar S., Rosenberg J.M., Bouzida D., Swendsen R.H., Kollman P.A. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, 13, 1011–21.
- [2.39] Marrink S.J., Risselada H.J., Yefimov S., Tieleman D.P., de Vries A.H. The MARTINI

Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, 111, 7812–24.

- [3.1] Saegusa T., Ikeda H., Fujii H. Crystalline Polyethylenimine. *Macromolecules* **1972**, 5, 108–108.
- [3.2] Perrine T.D., Landis W.R. Analysis of Polyethylenimine by Spectrophotometry of its Copper Chelate. *J. Polym. Sci. Part A-1 Polym. Chem.* **1967**, 5, 1993–2003.
- [3.3] Glaser H.T., Edzwald J.K. Coagulation and Direct Filtration of Humic Substances with Polyethylenimine. *Environ. Sci. Technol.* **1979**, 13, 299–305.
- [3.4] Korus I., Loska K. Removal of Cr(III) and Cr(VI) Ions From Aqueous Solutions by Means of Polyelectrolyte-Enhanced Ultrafiltration. *Desalination* **2009**, 247, 390–5.
- [3.5] Zhao R., Li X., Sun B., Li Y., Li Y., Yang R., et al. Branched Polyethylenimine Grafted Electrospun Polyacrylonitrile Fiber Membrane: A Novel and Effective Adsorbent for Cr(VI) Remediation in Wastewater. *J. Mater. Chem. A* **2017**, 5, 1133–44.
- [3.6] Lv Y., Yang H.-C., Liang H.-Q., Wan L.-S., Xu Z.-K. Nanofiltration Membranes via Co-Deposition of Polydopamine/Polyethylenimine Followed by Cross-Linking. *J. Memb. Sci.* **2015**, 476, 50–8.
- [3.7] Xu X., Song C., Andrése J.M., Miller B.G., Scaroni A.W. Preparation and Characterization of Novel CO₂ “Molecular Basket” Adsorbents Based on Polymer-Modified Mesoporous Molecular Sieve MCM-41. *Microporous Mesoporous Mater.* **2003**, 62, 29–45.
- [3.8] Xu X., Song C., Miller B.G., Scaroni A.W. Adsorption Separation of Carbon Dioxide From Flue Gas of Natural Gas-Fired Boiler by a Novel Nanoporous “Molecular Basket” Adsorbent. *Fuel Process. Technol.* **2005**, 86, 1457–72.
- [3.9] Wang X., Schwartz V., Clark J.C., Ma X., Overbury S.H., Xu X., et al. Infrared Study of CO₂ Sorption Over “Molecular Basket” Sorbent Consisting of Polyethylenimine-Modified Mesoporous Molecular Sieve. *J. Phys. Chem. C* **2009**, 113, 7260–8.
- [3.10] Wang D., Ma X., Sentorun-Shalaby C., Song C. Development of Carbon-Based “Molecular Basket” Sorbent for CO₂ Capture. *Ind. Eng. Chem. Res.* **2012**, 51, 3048–57.
- [3.11] Xu X., Song C., Miller B.G., Scaroni A.W. Influence of Moisture on CO₂ Separation from Gas Mixture by a Nanoporous Adsorbent Based on Polyethylenimine-Modified

- Molecular Sieve MCM-41. *Ind. Eng. Chem. Res.* **2005**, 44, 8113–9.
- [3.12] Xu X., Song C., Andresen J.M., Miller B.G., Scaroni A.W. Novel Polyethylenimine-Modified Mesoporous Molecular Sieve of MCM-41 Type as High-Capacity Adsorbent for CO₂ Capture. *Energy Fuels* **2002**, 16, 1463–9.
- [3.13] Wang X., Ma X., Xu X., Sun L., Song C. Mesoporous-Molecular-Sieve-Supported Polymer Sorbents for Removing H₂S from Hydrogen Gas Streams. *Top. Catal.* **2008**, 49, 108–17.
- [3.14] Xu X., Novochinskii I., Song C. Low-Temperature Removal of H₂S by Nanoporous Composite of Polymer - Mesoporous Molecular Sieve MCM-41 as Adsorbent for Fuel Cell Applications. *Energy Fuels* **2005**, 19, 2214–5.
- [3.15] Ivnitski D., Branch B., Atanassov P., Apblett C. Glucose Oxidase Anode for Biofuel Cell Based on Direct Electron Transfer. *Electrochem. Commun.* **2006**, 8, 1204–10.
- [3.16] Das M., Barbora L., Das P., Goswami P. Biofuel Cell for Generating Power From Methanol Substrate Using Alcohol Oxidase Bioanode and Air-Breathed Laccase Biocathode. *Biosens. Bioelectron.* **2014**, 59, 184–91.
- [3.17] Zhou Y., Fuentes-Hernandez C., Won Shim J., Khan T.M., Kippelen B. High Performance Polymeric Charge Recombination Layer for Organic Tandem Solar Cells. *Energy Environ. Sci.* **2012**, 5, 9827–32.
- [3.18] Cheng Y., Xu C., Shen P.K., Jiang S.P. Effect of Nitrogen-Containing Functionalization on the Electrocatalytic Activity of PtRu Nanoparticles Supported on Carbon Nanotubes for Direct Methanol Fuel Cells. *Appl. Catal. B Environ.* **2014**, 158, 140–9.
- [3.19] Shan C., Yang H., Song J., Han D., Ivaska A., Niu L. Direct Electrochemistry of Glucose Oxidase and Biosensing for Glucose Based on Graphene. *Anal. Chem.* **2009**, 81, 2378–82.
- [3.20] Ju Q., Tu D., Liu Y., Li R., Zhu H., Chen J., et al. Amine-Functionalized Lanthanide-Doped KGdF₄ Nanocrystals as Potential Optical/Magnetic Multimodal Bioprobes. *J. Am. Chem. Soc.* **2012**, 134, 1323–30.
- [3.21] Ma L., Sun N., Zhang J., Tu C., Cao X., Duan D., et al. Polyethylenimine-Coated Fe₃O₄ Nanoparticles Effectively Quench Fluorescent DNA, Which Can Be Developed as a Novel Platform for Protein Detection. *Nanoscale* **2017**, 9, 17699–703.
- [3.22] Pack D.W., Hoffman A.S., Pun S., Stayton P.S. Design and Development of Polymers

- for Gene Delivery. *Nat. Rev. Drug Discov.* **2005**, 4, 581–93.
- [3.23] Boussif O., Lezoualc'h F., Zanta M.A., Mergny M.D., Scherman D., Demeneix B., et al. A Versatile Vector for Gene and Oligonucleotide Transfer into Cells in Culture and In Vivo: Polyethylenimine. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, 92, 7297–301.
- [3.24] Shete H.K., Prabhu R.H., Patravale V.B. Endosomal Escape: A Bottleneck in Intracellular Delivery. *J. Nanosci. Nanotechnol.* **2014**, 14, 460–74.
- [3.25] Akinc A., Thomas M., Klibanov A.M., Langer R. Exploring Polyethylenimine-Mediated DNA Transfection and the Proton Sponge Hypothesis. *J. Gene Med.* **2005**, 7, 657–63.
- [3.26] Kichler A., Leborgne C., Coeytaux E., Danos O. Polyethylenimine-Mediated Gene Delivery: A Mechanistic Study. *J. Gene Med.* **2001**, 3, 135–44.
- [3.27] Godbey W.T., Wu K.K., Mikos A.G. Size Matters: Molecular Weight Affects the Efficiency of Poly(ethylenimine) as a Gene Delivery Vehicle. *J. Biomed. Mater. Res.* **1999**, 45, 268–75.
- [3.28] Kunath K., von Harpe A., Fischer D., Petersen H., Bickel U., Voigt K., et al. Low-Molecular-Weight Polyethylenimine as a Non-Viral Vector for DNA Delivery: Comparison of Physicochemical Properties, Transfection Efficiency and In Vivo Distribution With High-Molecular-Weight Polyethylenimine. *J. Control. Release* **2003**, 89, 113–25.
- [3.29] Tang M.X., Szoka F.C. The Influence of Polymer Structure on the Interactions of Cationic Polymers with DNA and Morphology of the Resulting Complexes. *Gene Ther.* **1997**, 4, 823–32.
- [3.30] Wightman L., Kircheis R., Rössler V., Carotta S., Ruzicka R., Kursa M., et al. Different Behavior of Branched and Linear Polyethylenimine for Gene Delivery In Vitro and In Vivo. *J. Gene Med.* **2001**, 3, 362–72.
- [3.31] Kircheis R., Wightman L., Wagner E. Design and Gene Delivery Activity of Modified Polyethylenimines. *Adv. Drug Deliv. Rev.* **2001**, 53, 341–58.
- [3.32] Mishra S., Webster P., Davis M.E. PEGylation Significantly Affects Cellular Uptake and Intracellular Trafficking of Non-Viral Gene Delivery Particles. *Eur. J. Cell Biol.* **2004**, 83, 97–111.
- [3.33] Ogris M., Brunner S., Schüller S., Kircheis R., Wagner E. PEGylated DNA/Transferrin–PEI Complexes: Reduced Interaction With Blood Components, Extended Circulation in

- Blood and Potential for Systemic Gene Delivery. *Gene Ther.* **1999**, 6, 595–605.
- [3.34] Lungwitz U., Breunig M., Blunk T., Göpferich A. Polyethylenimine-Based Non-Viral Gene Delivery Systems. *Eur. J. Pharm. Biopharm.* **2005**, 60, 247–66.
- [3.35] Rehman Z.U., Hoekstra D., Zuhorn I.S. Mechanism of Polyplex- and Lipoplex-Mediated Delivery of Nucleic Acids: Real-Time Visualization of Transient Membrane Destabilization Without Endosomal Lysis. *ACS Nano* **2013**, 7, 3767–77.
- [3.36] Morimoto K., Nishikawa M., Kawakami S., Nakano T., Hattori Y., Fumoto S., et al. Molecular Weight-Dependent Gene Transfection Activity of Unmodified and Galactosylated Polyethyleneimine on Hepatoma Cells and Mouse Liver. *Mol. Ther.* **2003**, 7, 254–61.
- [3.37] Kafil V., Omid Y. Cytotoxic Impacts of Linear and Branched Polyethylenimine Nanostructures in A431 Cells. *Bioimpacts* **2011**, 1, 23–30.
- [3.38] Ziebarth J., Wang Y. Molecular Dynamics Simulations of DNA-Polycation Complex Formation. *Biophys. J.* **2009**, 97, 1971–83.
- [3.39] Vanommeslaeghe K., Hatcher E., Acharya C., Kundu S., Zhong S., Shim J., et al. CHARMM General Force Field (CGenFF): A Force Field for Drug-Like Molecules Compatible With the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, 31, 671–90.
- [3.40] Sun C., Tang T., Uludağ H., Cuervo J.E. Molecular Dynamics Simulations of DNA/PEI Complexes: Effect of PEI Branching and Protonation State. *Biophys. J.* **2011**, 100, 2754–63.
- [3.41] Wei Z., Luijten E. Systematic Coarse-Grained Modeling of Complexation Between Small Interfering RNA and Polycations. *J. Chem. Phys.* **2015**, 143, 243146.
- [3.42] Beu T.A., Farçaş A. CHARMM Force Field and Molecular Dynamics Simulations of Protonated Polyethylenimine. *J. Comput. Chem.* **2017**, 38, 2335–48.
- [3.43] Sun C., Tang T., Uludağ H. Molecular Dynamics Simulations for Complexation of DNA with 2 kDa PEI Reveal Profound Effect of PEI Architecture on Complexation. *J. Phys. Chem. B* **2012**, 116, 2405–13.
- [3.44] Bagai S., Sun C., Tang T. Lipid-Modified Polyethylenimine-Mediated DNA Attraction Evaluated by Molecular Dynamics Simulations. *J. Phys. Chem. B* **2014**, 118, 7070–6.
- [3.45] Sun C., Tang T., Uludağ H. Molecular Dynamics Simulations of PEI Mediated DNA

- Aggregation. *Biomacromolecules* **2011**, 12, 3698–707.
- [3.46] Bagai S., Sun C., Tang T. Potential of Mean Force of Polyethylenimine-Mediated DNA Attraction. *J. Phys. Chem. B* **2013**, 117, 49–56.
- [3.47] Kostrikskii A.Y., Kondinskaia D.A., Nesterenko A.M., Gurtovenko A.A. Adsorption of Synthetic Cationic Polymers on Model Phospholipid Membranes: Insight from Atomic-Scale Molecular Dynamics Simulations. *Langmuir* **2016**, 32, 10402–14.
- [3.48] Fox S.J., Fazil M.H.U.T., Dhand C., Venkatesh M., Goh E.T.L., Harini S., et al. Insight Into Membrane Selectivity of Linear and Branched Polyethylenimines and Their Potential as Biocides for Advanced Wound Dressings. *Acta Biomater.* **2016**, 37, 155–64.
- [3.49] Yang S., May S. Release of Cationic Polymer-DNA Complexes From the Endosome: A Theoretical Investigation of the Proton Sponge Hypothesis. *J. Chem. Phys.* **2008**, 129, 185105.
- [3.50] Freeman E.C., Weiland L.M., Meng W.S. Modeling the Proton Sponge Hypothesis: Examining Proton Sponge Effectiveness for Enhancing Intracellular Gene Delivery Through Multiscale Modeling. *J. Biomater. Sci. Polym. Ed.* **2013**, 24, 398–416.
- [3.51] Noid W.G. Perspective: Coarse-Grained Models for Biomolecular Systems. *J. Chem. Phys.* **2013**, 139, 090901.
- [3.52] Ziebarth J.D., Wang Y. Understanding the Protonation Behavior of Linear Polyethylenimine in Solutions Through Monte Carlo Simulations. *Biomacromolecules* **2010**, 11, 29–38.
- [3.53] Ziebarth J., Wang Y. Coarse-Grained Molecular Dynamics Simulations of DNA Condensation by Block Copolymer and Formation of Core - Corona Structures. *J. Phys. Chem. B* **2010**, 114, 6225–32.
- [3.54] Jorge A.F., Dias R.S., Pais A.A.C.C. Enhanced Condensation and Facilitated Release of DNA Using Mixed Cationic Agents: A Combined Experimental and Monte Carlo Study. *Biomacromolecules* **2012**, 13, 3151–61.
- [3.55] Marrink S.J., de Vries A.H., Mark A.E. Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B* **2004**, 108, 750–60.
- [3.56] Marrink S.J., Tieleman D.P. Perspective on the Martini model. *Chem. Soc. Rev.* **2013**, 42, 6801–22.
- [3.57] Dick C.R., Ham G.E. Characterization of Polyethylenimine. *J. Macromol. Sci. Part A-*

- Chem.* **1970**, 4, 1301–14.
- [3.58] Suh J., Paik H.J., Hwang B.K. Ionization of Poly(ethylenimine) and Poly(allylamine) at Various pH's. *Bioorg. Chem.* **1994**, 22, 318–27.
- [3.59] Tang G.P., Guo H.Y., Alexis F., Wang X., Zeng S., Lim T.M., et al. Low Molecular Weight Polyethylenimines Linked by β -Cyclodextrin for Gene Transfer Into the Nervous System. *J. Gene Med.* **2006**, 8, 736–44.
- [3.60] Marrink S.J., Risselada H.J., Yefimov S., Tieleman D.P., de Vries A.H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, 111, 7812–24.
- [3.61] Monticelli L., Kandasamy S.K., Periole X., Larson R.G., Tieleman D.P., Marrink S.J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, 4, 819–34.
- [3.62] Seo M., Rauscher S., Pomès R., Tieleman D.P. Improving Internal Peptide Dynamics in the Coarse-Grained MARTINI Model: Toward Large-Scale Simulations of Amyloid-And Elastin-Like Peptides. *J. Chem. Theory Comput.* **2012**, 8, 1774–85.
- [3.63] Uusitalo J.J., Ingólfsson H.I., Akhshi P., Tieleman D.P., Marrink S.J. Martini Coarse-Grained Force Field: Extension to DNA. *J. Chem. Theory Comput.* **2015**, 11, 3932–45.
- [3.64] Rossi G., Monticelli L., Puisto S.R., Vattulainen I., Ala-Nissila T. Coarse-Graining Polymers With the MARTINI Force-Field: Polystyrene as a Benchmark Case. *Soft Matter* **2011**, 7, 698–708.
- [3.65] Lee H., de Vries A.H., Marrink S.J., Pastor R.W. A Coarse-Grained Model for Polyethylene Oxide and Polyethylene Glycol: Conformation and Hydrodynamics. *J. Phys. Chem. B* **2009**, 113, 13186–94.
- [3.66] Yesylevskyy S.O., Schäfer L. V., Sengupta D., Marrink S.J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comput. Biol.* **2010**, 6, e1000810.
- [3.67] Michalowsky J., Schäfer L. V., Holm C., Smiatek J. A Refined Polarizable Water Model for the Coarse-Grained MARTINI Force Field With Long-Range Electrostatic Interactions. *J. Chem. Phys.* **2017**, 146, 054501.
- [3.68] de Jong D.H., Singh G., Bennett W.F.D., Arnarez C., Wassenaar T.A., Schäfer L. V., et al. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **2013**, 9, 687–97.

- [3.69] Utsuno K., Uludağ H. Thermodynamics of Polyethylenimine-DNA Binding and DNA Condensation. *Biophys. J.* **2010**, 99, 201–7.
- [3.70] Martini tutorial: Polymers, <http://cgmartini.nl>, **2018**.
- [3.71] Abraham M.H., Platts J.A., Hersey A., Jeo A.J., Taft R.W. Correlation and Estimation of Gas–Chloroform and Water–Chloroform Partition Coefficients by a Linear Free Energy Relationship Method. *J. Pharm. Sci.* **1999**, 88, 670–9.
- [3.72] Duffy E.M., Jorgensen W.L. Prediction of Properties From Simulations: Free Energies of Solvation in Hexadecane, Octanol, and Water. *J. Am. Chem. Soc.* **2000**, 122, 2878–88.
- [3.73] Tschöp W., Kremer K., Batoulis J., Bürger T., Hahn O. Simulation of Polymer Melts. I. Coarse-Graining Procedure for Polycarbonates. *Acta Polym.* **1998**, 49, 61–74.
- [3.74] Reith D., Pütz M., Müller-Plathe F. Deriving Effective Mesoscale Potentials From Atomistic Simulations. *J. Comput. Chem.* **2003**, 24, 1624–36.
- [3.75] Jorgensen W.L. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water. *J. Am. Chem. Soc.* **1981**, 103, 335–40.
- [3.76] Hess B., Kutzner C., van der Spoel D., Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, 4, 435–47.
- [3.77] Bussi G., Donadio D., Parrinello M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, 126, 014101.
- [3.78] Berendsen H.J.C., Postma J.P.M., van Gunsteren W.F., DiNola A., Haak J.R. Molecular Dynamics With Coupling to an External Bath. *J. Chem. Phys.* **1984**, 81, 3684–90.
- [3.79] Parrinello M., Rahman A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, 52, 7182–90.
- [3.80] Darden T., York D., Pedersen L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, 98, 10089–92.
- [3.81] Essmann U., Perera L., Berkowitz M.L., Darden T., Lee H., Pedersen L.G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, 103, 8577–93.
- [3.82] MacKerell Jr A.D., Banavali N., Foloppe N. Development and Current Status of the CHARMM Force Field for Nucleic Acids. *Biopolymers* **2000**, 56, 257–65.
- [3.83] Kumar S., Rosenberg J.M., Bouzida D., Swendsen R.H., Kollman P.A. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The

- Method. *J. Comput. Chem.* **1992**, 13, 1011–21.
- [3.84] Abraham M.J., Murtola T., Schulz R., Páll S., Smith J.C., Hess B., et al. GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism From Laptops to Supercomputers. *SoftwareX* **2015**, 1–2, 19–25.
- [3.85] Hub J.S., de Groot B.L., van der Spoel D. g_wham-A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **2010**, 6, 3713–20.
- [3.86] Dobrynin A. V, Rubinstein M. Theory of Polyelectrolytes in Solutions and at Surfaces. *Prog. Polym. Sci.* **2005**, 30, 1049–118.
- [3.87] Abdallah B., Hassan A., Benoist C., Goula D., Behr J.P., Demeneix B.A. A Powerful Nonviral Vector for In Vivo Gene Transfer into the Adult Mammalian Brain: Polyethylenimine. *Hum. Gene Ther.* **1996**, 7, 1947–54.
- [3.88] Lee H., Larson R.G. Molecular Dynamics Simulations of PAMAM Dendrimer-Induced Pore Formation in DPPC Bilayers With a Coarse-Grained Model. *J. Phys. Chem. B* **2006**, 110, 18204–11.
- [3.89] Panizon E., Bochicchio D., Monticelli L., Rossi G. MARTINI Coarse-Grained Models of Polyethylene and Polypropylene. *J. Phys. Chem. B* **2015**, 119, 8209–16.
- [4.1] Boussif O., Lezoualc'h F., Zanta M.A., Mergny M.D., Scherman D., Demeneix B., et al. A Versatile Vector for Gene and Oligonucleotide Transfer into Cells in Culture and In Vivo: Polyethylenimine. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, 92, 7297–301.
- [4.2] Huang P., Zhao J., Wei C., Hou X., Chen P., Tan Y., et al. Erythrocyte Membrane Based Cationic Polymer-mcDNA Complexes as an Efficient Gene Delivery System. *Biomater. Sci.* **2017**, 5, 120–7.
- [4.3] Godbey W.T., Barry M.A., Saggau P., Wu K.K., Mikos A.G. Poly(Ethylenimine)-Mediated Transfection: A New Paradigm for Gene Delivery. *J. Biomed. Mater. Res.* **2000**, 51, 321–8.
- [4.4] Aposhian V.H., Barr S.M., Keck K. Experimental Gene Delivery Systems for Mammalian Cells-The Polyoma Pseudovirus System. *Adv. Enzyme Regul.* **1977**, 16, 275–88.
- [4.5] Godbey W.T., Wu K.K., Mikos A.G. Size Matters: Molecular Weight Affects the

- Efficiency of Poly(ethylenimine) as a Gene Delivery Vehicle. *J. Biomed. Mater. Res.* **1999**, 45, 268–75.
- [4.6] Forrest M.L., Koerber J.T., Pack D.W. A Degradable Polyethylenimine Derivative with Low Toxicity for Highly Efficient Gene Delivery. *Bioconjug. Chem.* **2003**, 14, 934–40.
- [4.7] Wightman L., Kircheis R., Rössler V., Carotta S., Ruzicka R., Kurska M., et al. Different Behavior of Branched and Linear Polyethylenimine for Gene Delivery In Vitro and In Vivo. *J. Gene Med.* **2001**, 3, 362–72.
- [4.8] Boeckle S., von Gersdorff K., van der Pijpen S., Culmsee C., Wagner E., Ogris M. Purification of Polyethylenimine Polyplexes Highlights the Role of Free Polycations in Gene Transfer. *J. Gene Med.* **2004**, 6, 1102–11.
- [4.9] Bahadur K.C.R., Landry B., Aliabadi H.M., Lavasanifar A., Uludağ H. Lipid Substitution on Low Molecular Weight (0.6-2.0 kDa) Polyethylenimine Leads to a Higher Zeta Potential of Plasmid DNA and Enhances Transgene Expression. *Acta Biomater.* **2011**, 7, 2209–17.
- [4.10] Tang M.X., Szoka F.C. The Influence of Polymer Structure on the Interactions of Cationic Polymers with DNA and Morphology of the Resulting Complexes. *Gene Ther.* **1997**, 4, 823–32.
- [4.11] Utsuno K., Uludağ H. Thermodynamics of Polyethylenimine-DNA Binding and DNA Condensation. *Biophys. J.* **2010**, 99, 201–7.
- [4.12] Dunlap D.D., Maggi A., Soria M.R., Monaco L. Nanoscopic Structure of DNA Condensed for Gene Delivery. *Nucleic Acids Res.* **1997**, 25, 3095–101.
- [4.13] Godbey W.T., Wu K.K., Mikos A.G. Tracking the Intracellular Path of Poly(ethylenimine)/DNA Complexes for Gene Delivery. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, 96, 5177–81.
- [4.14] Bieber T., Meissner W., Kostin S., Niemann A., Elsasser H.-P. Intracellular Route and Transcriptional Competence of Polyethylenimine-DNA Complexes. *J. Control. Release* **2002**, 82, 441–54.
- [4.15] Verma A., Stellacci F. Effect of Surface Properties on Nanoparticle–Cell Interactions. *Small* **2010**, 6, 12–21.
- [4.16] Villanueva A., Cañete M., Roca A.G., Calero M., Veintemillas-Verdaguer S., Serna C.J., et al. The Influence of Surface Functionalization on the Enhanced Internalization of

- Magnetic Nanoparticles in Cancer Cells. *Nanotechnology* **2009**, 20, 115103.
- [4.17] Ogris M., Brunner S., Schüller S., Kircheis R., Wagner E. PEGylated DNA/Transferrin–PEI Complexes: Reduced Interaction With Blood Components, Extended Circulation in Blood and Potential for Systemic Gene Delivery. *Gene Ther.* **1999**, 6, 595–605.
- [4.18] Li S., Rizzo M.A., Bhattacharya S., Huang L. Characterization of Cationic Lipid–Protamine–DNA (LPD) Complexes for Intravenous Gene Delivery. *Gene Ther.* **1998**, 5, 930–7.
- [4.19] Khalil I.A., Kogure K., Akita H., Harashima H. Uptake Pathways and Subsequent Intracellular Trafficking in Nonviral Gene Delivery. *Pharmacol. Rev.* **2006**, 58, 32–45.
- [4.20] Hillaireau H., Couvreur P. Nanocarriers' Entry Into the Cell: Relevance to Drug Delivery. *Cell. Mol. Life Sci.* **2009**, 66, 2873–96.
- [4.21] Choosakoonkriang S., Lobo B.A., Koe G.S., Koe J.G., Middaugh C.R. Biophysical Characterization of PEI/DNA Complexes. *J. Pharm. Sci.* **2003**, 92, 1710–22.
- [4.22] D'Andrea C., Pezzoli D., Malloggi C., Candeo A., Capelli G., Bassi A., et al. The Study of Polyplex Formation and Stability by Time-Resolved Fluorescence Spectroscopy of SYBR Green I-Stained DNA. *Photochem. Photobiol. Sci.* **2014**, 13, 1680–9.
- [4.23] Erbacher P., Bettinger T., Belguise-Valladier P., Zou S., Coll J.L., Behr J.P., et al. Transfection and Physical Properties of Various Saccharide, Poly(Ethylene Glycol), and Antibody-Derivatized Polyethylenimines (PEI). *J. Gene Med.* **1999**, 1, 210–22.
- [4.24] Chithrani B.D., Ghazani A.A., Chan W.C.W. Determining the Size and Shape Dependence of Gold Nanoparticle Uptake into Mammalian Cells. *Nano Lett.* **2006**, 6, 662–8.
- [4.25] Chithrani B.D., Chan W.C.W. Elucidating the Mechanism of Cellular Uptake and Removal of Protein-Coated Gold Nanoparticles of Different Sizes and Shapes. *Nano Lett.* **2007**, 7, 1542–50.
- [4.26] Zhang S., Gao H., Bao G. Physical Principles of Nanoparticle Cellular Endocytosis. *ACS Nano* **2015**, 9, 8655–71.
- [4.27] Ding H.M., Ma Y.Q. Theoretical and Computational Investigations of Nanoparticle–Biomembrane Interactions in Cellular Delivery. *Small* **2015**, 11, 1055–71.
- [4.28] Kettler K., Veltman K., van de Meent D., van Wezel A., Hendriks A.J. Cellular Uptake of Nanoparticles as Determined by Particle Properties, Experimental Conditions, and Cell

- Type. *Environ. Toxicol. Chem.* **2014**, 33, 481–92.
- [4.29] Sun C., Tang T., Uludağ H., Cuervo J.E. Molecular Dynamics Simulations of DNA/PEI Complexes: Effect of PEI Branching and Protonation State. *Biophys. J.* **2011**, 100, 2754–63.
- [4.30] Sun C., Tang T. Study on the Role of Polyethylenimine as Gene Delivery Carrier Using Molecular Dynamics Simulations. *J. Adhes. Sci. Technol.* **2014**, 28, 399–416.
- [4.31] Sun C., Tang T., Uludağ H. Molecular Dynamics Simulations for Complexation of DNA with 2 kDa PEI Reveal Profound Effect of PEI Architecture on Complexation. *J. Phys. Chem. B* **2012**, 116, 2405–13.
- [4.32] Kondinskaia D.A., Kostritskii A.Y., Nesterenko A.M., Antipina A.Y., Gurtovenko A.A. Atomic-Scale Molecular Dynamics Simulations of DNA-Polycation Complexes: Two Distinct Binding Patterns. *J. Phys. Chem. B* **2016**, 120, 6546–54.
- [4.33] Ziebarth J., Wang Y. Molecular Dynamics Simulations of DNA-Polycation Complex Formation. *Biophys. J.* **2009**, 97, 1971–83.
- [4.34] Ziebarth J., Wang Y. Coarse-Grained Molecular Dynamics Simulations of DNA Condensation by Block Copolymer and Formation of Core - Corona Structures. *J. Phys. Chem. B* **2010**, 114, 6225–32.
- [4.35] Wei Z., Luijten E. Systematic Coarse-Grained Modeling of Complexation Between Small Interfering RNA and Polycations. *J. Chem. Phys.* **2015**, 143, 243146.
- [4.36] Zhan B., Shi K., Dong Z., Lv W., Zhao S., Han X., et al. Coarse-Grained Simulation of Polycation/DNA-Like Complexes: Role of Neutral Block. *Mol. Pharm.* **2015**, 12, 2834–44.
- [4.37] Sun C., Tang T., Uludağ H. Molecular Dynamics Simulations of PEI Mediated DNA Aggregation. *Biomacromolecules* **2011**, 12, 3698–707.
- [4.38] Bagai S., Sun C., Tang T. Potential of Mean Force of Polyethylenimine-Mediated DNA Attraction. *J. Phys. Chem. B* **2013**, 117, 49–56.
- [4.39] Tom A.M., Rajesh R., Vemparala S. Aggregation Dynamics of Rigid Polyelectrolytes. *J. Chem. Phys.* **2016**, 144, 034904.
- [4.40] Lee K.-C., Borukhov I., Gelbart W.M., Liu A.J., Stevens M.J. Effect of Mono- and Multivalent Salts on Angle-Dependent Attractions Between Charged Rods. *Phys. Rev. Lett.* **2004**, 93, 128101.

- [4.41] Stevens M.J. Bundle Binding in Polyelectrolyte Solutions. *Phys. Rev. Lett.* **1999**, 82, 101–4.
- [4.42] Grønbech-Jensen N., Mashl R.J., Bruinsma R.F., Gelbart W.M. Counterion-Induced Attraction between Rigid Polyelectrolytes. *Phys. Rev. Lett.* **1997**, 78, 2477–80.
- [4.43] Fazli H., Golestanian R. Aggregation Kinetics of Stiff Polyelectrolytes in the Presence of Multivalent Salt. *Phys. Rev. E* **2007**, 76, 041801.
- [4.44] Sayar M., Holm C. Equilibrium Polyelectrolyte Bundles With Different Multivalent Counterion Concentrations. *Phys. Rev. E* **2010**, 82, 031901.
- [4.45] Sayar M., Holm C. Finite-Size Polyelectrolyte Bundles at Thermodynamic Equilibrium. *Europhys. Lett.* **2007**, 77, 16001.
- [4.46] Diehl A., Carmona H.A., Levin Y. Counterion Correlations and Attraction Between Like-Charged Macromolecules. *Phys. Rev. E* **2001**, 64, 011804.
- [4.47] Fazli H., Mohammadinejad S., Golestanian R. Salt-Induced Aggregation of Stiff Polyelectrolytes. *J. Phys. Condens. Matter* **2009**, 21, 424111.
- [4.48] Tom A.M., Rajesh R., Vemparala S. Aggregation of Flexible Polyelectrolytes: Phase Diagram and Dynamics. *J. Chem. Phys.* **2017**, 147, 144903.
- [4.49] Ogris M., Steinlein P., Kursá M., Mechtler K., Kircheis R., Wagner E. The Size of DNA/Transferrin-PEI Complexes Is an Important Factor for Gene Expression in Cultured Cells. *Gene Ther.* **1998**, 5, 1425–33.
- [4.50] Uusitalo J.J., Ingólfsson H.I., Akhshi P., Tieleman D.P., Marrink S.J. Martini Coarse-Grained Force Field: Extension to DNA. *J. Chem. Theory Comput.* **2015**, 11, 3932–45.
- [4.51] Mahajan S., Tang T. Martini Coarse-Grained Model for Polyethylenimine. *J. Comput. Chem.* **2019**, 40, 607–18.
- [4.52] Yesylevskyy S.O., Schäfer L. V., Sengupta D., Marrink S.J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comput. Biol.* **2010**, 6, e1000810.
- [4.53] Marrink S.J., de Vries A.H., Mark A.E. Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B* **2004**, 108, 750–60.
- [4.54] Abraham M.J., Murtola T., Schulz R., Páll S., Smith J.C., Hess B., et al. GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism From Laptops to Supercomputers. *SoftwareX* **2015**, 1–2, 19–25.
- [4.55] Marrink S.J., Tieleman D.P. Perspective on the Martini model. *Chem. Soc. Rev.* **2013**, 42,

6801–22.

- [4.56] Hess B., Bekker H., Berendsen H.J.C., Fraaije J.G.E.M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, 18, 1463–72.
- [4.57] Bussi G., Donadio D., Parrinello M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, 126, 014101.
- [4.58] Berendsen H.J.C., Postma J.P.M., van Gunsteren W.F., DiNola A., Haak J.R. Molecular Dynamics With Coupling to an External Bath. *J. Chem. Phys.* **1984**, 81, 3684–90.
- [4.59] Parrinello M., Rahman A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, 52, 7182–90.
- [4.60] Verlet L. Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, 159, 98–103.
- [4.61] Tironi I.G., Sperb R., Smith P.E., van Gunsteren W.F. A Generalized Reaction Field Method for Molecular Dynamics Simulations. *J. Chem. Phys.* **1995**, 102, 5451–9.
- [4.62] von Smoluchowski M. Versuch Einer Mathematischen Theorie der Koagulationskinetik kolloider Lösungen. *Zeitschrift f. Physik. Chemie.* **1917**, 92, 129–68.
- [4.63] Goula D., Remy J.S., Erbacher P., Wasowicz M., Levi G., Abdallah B., et al. Size, Diffusibility and Transfection Performance of Linear PEI/DNA Complexes in the Mouse Central Nervous System. *Gene Ther.* **1998**, 5, 712–7.
- [4.64] Tang G.P., Guo H.Y., Alexis F., Wang X., Zeng S., Lim T.M., et al. Low Molecular Weight Polyethylenimines Linked by β -Cyclodextrin for Gene Transfer Into the Nervous System. *J. Gene Med.* **2006**, 8, 736–44.
- [4.65] Putnam D., Gentry C.A., Pack D.W., Langer R. Polymer-Based Gene Delivery With Low Cytotoxicity by a Unique Balance of Side-Chain Termini. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, 98, 1200–5.
- [4.66] dos Santos A.L.G., Bochot A., Tsapis N., Artzner F., Bejjani R.A., Thillaye-Goldenberg B., et al. Oligonucleotide–Polyethylenimine Complexes Targeting Retinal Cells: Structural Analysis and Application to Anti-TGF β 2 Therapy. *Pharm. Res.* **2006**, 23, 770–81.
- [4.67] Wojnilowicz M., Glab A., Bertucci A., Caruso F., Cavalieri F. Super-resolution Imaging of Proton Sponge-Triggered Rupture of Endosomes and Cytosolic Release of Small Interfering RNA. *ACS Nano* **2019**, 13, 187–202.

- [4.68] Rehman Z.U., Hoekstra D., Zuhorn I.S. Mechanism of Polyplex- and Lipoplex-Mediated Delivery of Nucleic Acids: Real-Time Visualization of Transient Membrane Destabilization Without Endosomal Lysis. *ACS Nano* **2013**, 7, 3767–77.
- [4.69] Godbey W.T., Wu K.K., Mikos A.G. Poly(Ethylenimine)-Mediated Gene Delivery Affects Endothelial Cell Function and Viability. *Biomaterials* **2001**, 22, 471–80.
- [5.1] Verma I.M., Somia N. Gene Therapy - Promises, Problems and Prospects. *Nature* **1997**, 389, 239–42.
- [5.2] Lungwitz U., Breunig M., Blunk T., Göpferich A. Polyethylenimine-Based Non-Viral Gene Delivery Systems. *Eur. J. Pharm. Biopharm.* **2005**, 60, 247–66.
- [5.3] Godbey W.T., Barry M.A., Saggau P., Wu K.K., Mikos A.G. Poly(Ethylenimine)-Mediated Transfection: A New Paradigm for Gene Delivery. *J. Biomed. Mater. Res.* **2000**, 51, 321–8.
- [5.4] Khalil I.A., Kogure K., Akita H., Harashima H. Uptake Pathways and Subsequent Intracellular Trafficking in Nonviral Gene Delivery. *Pharmacol. Rev.* **2006**, 58, 32–45.
- [5.5] Oh Y.-K., Suh D., Kim J.M., Choi H.-G., Shin K., Ko J.J. Polyethylenimine-Mediated Cellular Uptake, Nucleus Trafficking and Expression of Cytokine Plasmid DNA. *Gene Ther.* **2002**, 9, 1627–32.
- [5.6] Pollard H., Remy J.S., Loussouarn G., Demolombe S., Behr J.P., Escande D. Polyethylenimine but not Cationic Lipids Promotes Transgene Delivery to the Nucleus in Mammalian Cells. *J. Biol. Chem.* **1998**, 273, 7507–11.
- [5.7] Kay M.A., Glorioso J.C., Naldini L.M. Viral Vectors for Gene Therapy: The Art of Turning Infectious Agents Into Vehicles of Therapeutics. *Nat. Med.* **2001**, 7, 33–40.
- [5.8] Glover D.J., Lipps H.J., Jans D.A. Towards Safe, Non-Viral Therapeutic Gene Expression in Humans. *Nat. Rev. Genet.* **2005**, 6, 299–310.
- [5.9] Yin H., Kanasty R.L., Eltoukhy A.A., Vegas A.J., Dorkin J.R., Anderson D.G. Non-Viral Vectors for Gene-Based Therapy. *Nat. Rev. Genet.* **2014**, 15, 541–55.
- [5.10] Bieber T., Meissner W., Kostin S., Niemann A., Elsasser H.-P. Intracellular Route and Transcriptional Competence of Polyethylenimine-DNA Complexes. *J. Control. Release* **2002**, 82, 441–54.
- [5.11] Godbey W.T., Wu K.K., Mikos A.G. Poly(Ethylenimine)-Mediated Gene Delivery

- Affects Endothelial Cell Function and Viability. *Biomaterials* **2001**, 22, 471–80.
- [5.12] Akinc A., Thomas M., Klivanov A.M., Langer R. Exploring Polyethylenimine-Mediated DNA Transfection and the Proton Sponge Hypothesis. *J. Gene Med.* **2005**, 7, 657–63.
- [5.13] Behr J.-P. The Proton Sponge: A Trick to Enter Cells the Viruses did not Exploit. *Chimia (Aarau)*. **1997**, 51, 34–6.
- [5.14] Boussif O., Lezoualc'h F., Zanta M.A., Mergny M.D., Scherman D., Demeneix B., et al. A Versatile Vector for Gene and Oligonucleotide Transfer into Cells in Culture and In Vivo: Polyethylenimine. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, 92, 7297–301.
- [5.15] Sonawane N.D., Szoka F.C., Verkman A.S. Chloride Accumulation and Swelling in Endosomes Enhances DNA Transfer by Polyamine-DNA Polyplexes. *J. Biol. Chem.* **2003**, 278, 44826–31.
- [5.16] Vermeulen L.M.P., Brans T., Samal S.K., Dubruel P., Demeester J., De Smedt S.C., et al. Endosomal Size and Membrane Leakiness Influence Proton Sponge-Based Rupture of Endosomal Vesicles. *ACS Nano* **2018**, 12, 2332–45.
- [5.17] Rehman Z.U., Hoekstra D., Zuhorn I.S. Mechanism of Polyplex- and Lipoplex-Mediated Delivery of Nucleic Acids: Real-Time Visualization of Transient Membrane Destabilization Without Endosomal Lysis. *ACS Nano* **2013**, 7, 3767–77.
- [5.18] Klemm A.R., Young D., Lloyd J.B. Effects of Polyethyleneimine on Endocytosis and Lysosome Stability. *Biochem. Pharmacol.* **1998**, 56, 41–6.
- [5.19] Yang S., May S. Release of Cationic Polymer-DNA Complexes From the Endosome: A Theoretical Investigation of the Proton Sponge Hypothesis. *J. Chem. Phys.* **2008**, 129, 185105.
- [5.20] Freeman E.C., Weiland L.M., Meng W.S. Modeling the Proton Sponge Hypothesis: Examining Proton Sponge Effectiveness for Enhancing Intracellular Gene Delivery Through Multiscale Modeling. *J. Biomater. Sci. Polym. Ed.* **2013**, 24, 398–416.
- [5.21] Itaka K., Harada A., Yamasaki Y., Nakamura K., Kawaguchi H., Kataoka K. In Situ Single Cell Observation by Fluorescence Resonance Energy Transfer Reveals Fast Intracytoplasmic Delivery and Easy Release of Plasmid DNA Complexed With Linear Polyethylenimine. *J. Gene Med.* **2004**, 6, 76–84.
- [5.22] Abbe E. Beiträge zur Theorie des Mikroskops und der Mikroskopischen Wahrnehmung. *Arch. Für Mikroskopische Anat.* **1873**, 9, 413–68.

- [5.23] Huang B., Bates M., Zhuang X. Super-Resolution Fluorescence Microscopy. *Annu. Rev. Biochem.* **2009**, 78, 993–1016.
- [5.24] Godbey W.T., Wu K.K., Mikos A.G. Tracking the Intracellular Path of Poly(ethylenimine)/DNA Complexes for Gene Delivery. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, 96, 5177–81.
- [5.25] Mahajan S., Tang T. Polyethylenimine-DNA Ratio Strongly Affects Their Nanoparticle Formation: A Large-Scale Coarse-Grained Molecular Dynamics Study. *J. Phys. Chem. B* **2019**, 123, 9629–40.
- [5.26] von Smoluchowski M. Versuch Einer Mathematischen Theorie der Koagulationskinetik kolloider Lösungen. *Zeitschrift f. Physik. Chemie.* **1917**, 92, 129–68.
- [5.27] Clamme J.P., Azoulay J., Mély Y. Monitoring of the Formation and Dissociation of Polyethylenimine/DNA Complexes by Two Photon Fluorescence Correlation Spectroscopy. *Biophys. J.* **2003**, 84, 1960–8.
- [5.28] Boeckle S., von Gersdorff K., van der Pipen S., Culmsee C., Wagner E., Ogris M. Purification of Polyethylenimine Polyplexes Highlights the Role of Free Polycations in Gene Transfer. *J. Gene Med.* **2004**, 6, 1102–11.
- [5.29] Yue Y., Jin F., Deng R., Cai J., Dai Z., Lin M.C.M., et al. Revisit Complexation Between DNA and Polyethylenimine - Effect of Length of Free Polycationic Chains on Gene Transfection. *J. Control. Release* **2011**, 152, 143–51.
- [5.30] Oishi M., Nagasaki Y., Itaka K., Nishiyama N., Kataoka K. Lactosylated Poly(ethylene Glycol)-siRNA Conjugate through Acid-Labile β -Thiopropionate Linkage to Construct pH-Sensitive Polyion Complex Micelles Achieving Enhanced Gene Silencing in Hepatoma Cells. *J. Am. Chem. Soc.* **2005**, 127, 1624–5.
- [5.31] Rudolph C., Lausier J., Naundorf S., Müller R.H., Rosenecker J. In Vivo Gene Delivery to the Lung Using Polyethylenimine and Fractured Polyamidoamine Dendrimers. *J. Gene Med.* **2000**, 2, 269–78.
- [5.32] Choosakoonkriang S., Lobo B.A., Koe G.S., Koe J.G., Middaugh C.R. Biophysical Characterization of PEI/DNA Complexes. *J. Pharm. Sci.* **2003**, 92, 1710–22.
- [5.33] Roos A., Boron W.F. Intracellular pH. *Physiol. Rev.* **1981**, 61, 296–434.
- [5.34] van der Aa M.A.E.M., Mastrobattista E., Oosting R.S., Hennink W.E., Koning G.A., Crommelin D.J.A. The Nuclear Pore Complex: The Gateway to Successful Nonviral

- Gene Delivery. *Pharm. Res.* **2006**, *23*, 447–59.
- [5.35] Panté N., Kann M. Nuclear Pore Complex Is Able to Transport Macromolecules with Diameters of ~39 nm. *Mol. Biol. Cell* **2002**, *13*, 425–34.
- [5.36] Paine P.L., Moore L.C., Horowitz S.B. Nuclear Envelope Permeability. *Nature* **1975**, *254*, 109–14.
- [5.37] Wang C., Wang Y., Li Y., Bodemann B., Zhao T., Ma X., et al. A Nanobuffer Reporter Library for Fine-Scale Imaging and Perturbation of Endocytic Organelles. *Nat. Commun.* **2015**, *6*, 8524.
- [5.38] Fischer D., Bieber T., Li Y., Elsässer H.-P., Kissel T. A Novel Non-viral Vector for DNA Delivery Based on Low Molecular Weight, Branched Polyethylenimine: Effect of Molecular Weight on Transfection Efficiency and Cytotoxicity. *Pharm. Res.* **1999**, *16*, 1273–9.
- [5.39] Godbey W.T., Wu K.K., Mikos A.G. Size Matters: Molecular Weight Affects the Efficiency of Poly(ethylenimine) as a Gene Delivery Vehicle. *J. Biomed. Mater. Res.* **1999**, *45*, 268–75.
- [5.40] Werth S., Urban-Klein B., Dai L., Höbel S., Grzelinski M., Bakowsky U., et al. A Low Molecular Weight Fraction of Polyethylenimine (PEI) Displays Increased Transfection Efficiency of DNA and siRNA in Fresh or Lyophilized Complexes. *J. Control. Release* **2006**, *112*, 257–70.
- [5.41] Bieber T., Elsässer H.P. Preparation of a Low Molecular Weight Polyethylenimine for Efficient Cell Transfection. *Biotechniques* **2001**, *30*, 74–7, 80–1.
- [5.42] Wightman L., Kircheis R., Rössler V., Carotta S., Ruzicka R., Kursa M., et al. Different Behavior of Branched and Linear Polyethylenimine for Gene Delivery In Vitro and In Vivo. *J. Gene Med.* **2001**, *3*, 362–72.
- [5.43] Krämer M., Stumbé J.-F., Grimm G., Kaufmann B., Krüger U., Weber M., et al. Dendritic Polyamines: Simple Access to New Materials With Defined Treelike Structures for Application in Nonviral Gene Delivery. *ChemBioChem* **2004**, *5*, 1081–7.
- [5.44] Sun C., Tang T., Uludağ H., Cuervo J.E. Molecular Dynamics Simulations of DNA/PEI Complexes: Effect of PEI Branching and Protonation State. *Biophys. J.* **2011**, *100*, 2754–63.
- [5.45] Uusitalo J.J., Ingólfsson H.I., Akhshi P., Tieleman D.P., Marrink S.J. Martini Coarse-

- Grained Force Field: Extension to DNA. *J. Chem. Theory Comput.* **2015**, 11, 3932–45.
- [5.46] Utsuno K., Uludağ H. Thermodynamics of Polyethylenimine-DNA Binding and DNA Condensation. *Biophys. J.* **2010**, 99, 201–7.
- [5.47] Marrink S.J., Risselada H.J., Yefimov S., Tieleman D.P., de Vries A.H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, 111, 7812–24.
- [5.48] Mahajan S., Tang T. Martini Coarse-Grained Model for Polyethylenimine. *J. Comput. Chem.* **2019**, 40, 607–18.
- [5.49] Mahajan S., Tang T. Erratum : “Martini Coarse-Grained Model for Polyethylenimine” [J. Comput. Chem. 2019, 40, 607-618, DOI:10.1002/jcc.25747]. *J. Comput. Chem.* **2020**, 41, 1730–4.
- [5.50] Yesylevskyy S.O., Schäfer L. V., Sengupta D., Marrink S.J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comput. Biol.* **2010**, 6, e1000810.
- [5.51] Abraham M.J., Murtola T., Schulz R., Páll S., Smith J.C., Hess B., et al. GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism From Laptops to Supercomputers. *SoftwareX* **2015**, 1–2, 19–25.
- [5.52] Tironi I.G., Sperb R., Smith P.E., van Gunsteren W.F. A Generalized Reaction Field Method for Molecular Dynamics Simulations. *J. Chem. Phys.* **1995**, 102, 5451–9.
- [5.53] Verlet L. Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, 159, 98–103.
- [5.54] Mahajan S., Tang T. Meeting experiments at the diffraction barrier: an in-silico widefield fluorescence microscopy. *BioRxiv 10.1101/2021.03.02.433395 (Version 1)* **2021**.
- [5.55] Mahajan S. In-silico-microscopy v.1.2.2, <https://github.com/subhamoymahajan/in-silico-microscopy>, **2021**.
- [5.56] Gandy R.O. Out-of-Focus Diffraction Patterns for Microscope Objectives. *Proc. Phys. Soc. B* **1954**, 67, 825–31.
- [5.57] Mahajan S. NPanalysis v.1.3, <https://github.com/subhamoymahajan/NPanalysis>, **2021**.
- [6.1] Huang B., Bates M., Zhuang X. Super-Resolution Fluorescence Microscopy. *Annu. Rev. Biochem.* **2009**, 78, 993–1016.
- [6.2] Rehman Z.U., Hoekstra D., Zuhorn I.S. Mechanism of Polyplex- and Lipoplex-Mediated

- Delivery of Nucleic Acids: Real-Time Visualization of Transient Membrane Destabilization Without Endosomal Lysis. *ACS Nano* **2013**, 7, 3767–77.
- [6.3] Frederiksen R.S., Alarcon-Llado E., Krogstrup P., Bojarskaite L., Buch-Månson N., Bolinsson J., et al. Nanowire-Aperture Probe: Local Enhanced Fluorescence Detection for the Investigation of Live Cells at the Nanoscale. *ACS Photonics* **2016**, 3, 1208–16.
- [6.4] Erickson H.P. Size and Shape of Protein Molecules at the Nanometer Level Determined by Sedimentation, Gel Filtration, and Electron Microscopy. *Biol. Proc. Online* **2009**, 11, 32–51.
- [6.5] Jeynes J.C.G., Geraki K., Jeynes C., Zhaohong M., Bettiol A.A., Latorre E., et al. Nanoscale Properties of Human Telomeres Measured with a Dual Purpose X-ray Fluorescence and Super Resolution Microscopy Gold Nanoparticle Probe. *ACS Nano* **2017**, 11, 12632–40.
- [6.6] Ahmadi Y., Nord A.L., Wilson A.J., Hütter C., Schroeder F., Beeby M., et al. The Brownian and Flow-Driven Rotational Dynamics of a Multicomponent DNA Origami-Based Rotor. *Small* **2020**, 16, 2001855.
- [6.7] Melitz W., Shen J., Kummel A.C., Lee S. Kelvin Probe Force Microscopy and Its Application. *Surf. Sci. Rep.* **2011**, 66, 1–27.
- [6.8] Giessibl F.J. Advances in Atomic Force Microscopy. *Rev. Modern Phys.* **2003**, 75, 949–83.
- [6.9] Huang B., Babcock H., Zhuang X. Breaking the Diffraction Barrier: Super-Resolution Imaging of Cells. *Cell* **2010**, 143, 1047–58.
- [6.10] Taylor D.L., Wang Y.-L. Fluorescently Labelled Molecules as Probes of the Structure and Function of Living Cells. *Nature* **1980**, 284, 405–10.
- [6.11] Valm A.M., Cohen S., Legant W.R., Melunis J., Hershberg U., Wait E., et al. Applying Systems-Level Spectral Imaging and Analysis to Reveal the Organelle Interactome. *Nature* **2017**, 546, 162–7.
- [6.12] Adhikari S., Moscatelli J., Smith E.M., Banerjee C., Puchner E.M. Single-Molecule Localization Microscopy and Tracking With Red-Shifted States of Conventional BODIPY Conjugates in Living Cells. *Nat. Commun.* **2019**, 10, 3400.
- [6.13] Manders E.M.M., Verbeek F.J., Aten J.A. Measurement of Co-localization of Objects in Dual-Color Confocal Images. *J. Microsc.* **1993**, 169, 375–82.

- [6.14] Qamar S., Wang G., Randle S.J., Ruggeri F.S., Varela J.A., Lin J.Q., et al. FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation- π Interactions. *Cell* **2018**, 173, 720–34.
- [6.15] Capoulade J., Wachsmuth M., Hufnagel L., Knop M. Quantitative Fluorescence Imaging of Protein Diffusion and Interaction in Living Cells. *Nat. Biotechnol.* **2011**, 29, 835–9.
- [6.16] Chong S., Dugast-Darzacq C., Liu Z., Dong P., Dailey G.M., Cattoglio C., et al. Imaging Dynamic and Selective Low-Complexity Domain Interactions That Control Gene Transcription. *Science* **2018**, 361, eaar2555.
- [6.17] Mauthe M., Orhon I., Rocchi C., Zhou X., Luhr M., Hijlkema K.-J., et al. Chloroquine Inhibits Autophagic Flux by Decreasing Autophagosome-Lysosome Fusion. *Autophagy* **2018**, 14, 1435–55.
- [6.18] Xu Z., Song W., Crozier K.B. Direct Particle Tracking Observation and Brownian Dynamics Simulations of a Single Nanoparticle Optically Trapped by a Plasmonic Nanoaperture. *ACS Photonics* **2018**, 5, 2850–9.
- [6.19] Lichtman J.W., Conchello J. Fluorescence Microscopy. *Nat. Methods* **2005**, 2, 910–9.
- [6.20] Stephens D.J., Allan V.J. Light Microscopy Techniques for Live Cell Imaging. *Science* **2003**, 300, 82–6.
- [6.21] Abbe E. Beiträge zur Theorie des Mikroskops und der Mikroskopischen Wahrnehmung. *Arch. Für Mikroskopische Anat.* **1873**, 9, 413–68.
- [6.22] Webb R.H. Confocal Optical Microscopy. *Reports Prog. Phys.* **1996**, 59, 427–71.
- [6.23] Huisken J., Swoger J., del Bene F., Wittbrodt J., Stelzer E.H.K. Optical Sectioning Deep Inside Live Embryos by Selective Plane Illumination Microscopy. *Science* **2004**, 305, 1007–9.
- [6.24] Denk W., Strickler J.H., Webb W.W. Two-Photon Laser Scanning Fluorescence Microscopy. *Science* **1990**, 248, 73–6.
- [6.25] Preza C., Ollinger J.M., McNally J.G., Thomas Jr L.J. Point-Spread Sensitivity Analysis for Computational Optical-Sectioning Microscopy. *Micron Microsc. Acta* **1992**, 23, 501–13.
- [6.26] McNally J.G., Preza C., Conchello J.-A., Thomas L.J. Artifacts in Computational Optical-Sectioning Microscopy. *J. Opt. Soc. Am. A* **1994**, 11, 1056–67.
- [6.27] Sarder P., Nehorai A. Deconvolution Methods for 3-D Fluorescence Microscopy Images.

- IEEE Signal Process. Mag.* **2006**, 23, 32–45.
- [6.28] Gustafsson M.G.L. Surpassing the Lateral Resolution Limit by a Factor of Two Using Structured Illumination Microscopy. *J. Microsc.* **2000**, 198, 82–7.
- [6.29] Hell S.W., Wichmann J. Breaking the Diffraction Resolution Limit by Stimulated Emission: Stimulated-Emission-Depletion Fluorescence Microscopy. *Opt. Lett.* **1994**, 19, 780–2.
- [6.30] Rust M.J., Bates M., Zhuang X. Sub-Diffraction-Limit Imaging by Stochastic Optical Reconstruction Microscopy (STORM). *Nat. Methods* **2006**, 3, 793–6.
- [6.31] Xu J., Tehrani K.F., Kner P. Multicolor 3D Super-Resolution Imaging by Quantum Dot Stochastic Optical Reconstruction Microscopy. *ACS Nano* **2015**, 9, 2917–25.
- [6.32] Liu W., Toussaint Jr. K.C., Okoro C., Zhu D., Chen Y., Kuang C., et al. Breaking the Axial Diffraction Limit: A Guide to Axial Super-Resolution Fluorescence Microscopy. *Laser Photon. Rev.* **2018**, 12, 1700333.
- [6.33] Yuan H., Huang C., Li J., Lykotrafitis G., Zhang S. One-Particle-Thick, Solvent-Free, Coarse-Grained Model for Biological and Biomimetic Fluid Membranes. *Phys. Rev. E* **2010**, 82, 011905.
- [6.34] Mahajan S., Tang T. Polyethylenimine-DNA Ratio Strongly Affects Their Nanoparticle Formation: A Large-Scale Coarse-Grained Molecular Dynamics Study. *J. Phys. Chem. B* **2019**, 123, 9629–40.
- [6.35] Uusitalo J.J., Ingólfsson H.I., Akhshi P., Tieleman D.P., Marrink S.J. Martini Coarse-Grained Force Field: Extension to DNA. *J. Chem. Theory Comput.* **2015**, 11, 3932–45.
- [6.36] Jung J., Nishima W., Daniels M., Bascom G., Kobayashi C., Adedoyin A., et al. Scaling molecular dynamics beyond 100,000 processor cores for large-scale biophysical simulations. *J. Comput. Chem.* **2019**, 40, 1919–30.
- [6.37] Ingólfsson H.I., Arnarez C., Periole X., Marrink S.J. Computational ‘Microscopy’ of Cellular Membranes. *J. Cell Sci.* **2016**, 129, 257–68.
- [6.38] Dror R.O., Dirks R.M., Grossman J.P., Xu H., Shaw D.E. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annu. Rev. Biophys.* **2012**, 41, 429–52.
- [6.39] Navruz I., Coskun A.F., Wong J., Mohammad S., Tseng D., Nagi R., et al. Smart-Phone Based Computational Microscopy Using Multi-Frame Contact Imaging on a Fiber-Optic

- Array. *Lab Chip* **2013**, 13, 4015–23.
- [6.40] Aidukas T., Eckert R., Harvey A.R., Waller L., Konda P.C. Low-Cost, Sub-Micron Resolution, Wide-Field Computational Microscopy Using Opensource Hardware. *Sci. Rep.* **2019**, 9, 7457.
- [6.41] Gibson S.F., Lanni F. Experimental Test of an Analytical Model of Aberration in an Oil-Immersion Objective Lens Used in Three-Dimensional Light Microscopy. *J. Opt. Soc. Am. A* **1992**, 9, 154–66.
- [6.42] Gandy R.O. Out-of-Focus Diffraction Patterns for Microscope Objectives. *Proc. Phys. Soc. B* **1954**, 67, 825–31.
- [6.43] Richards B., Wolf E. Electromagnetic Diffraction in Optical Systems, II. Structure of the Image Field in an Aplanatic System. *Proc. R. Soc. London A* **1959**, 253, 358–79.
- [6.44] Hell S., Reiner G., Cremer C., Stelzer E.H.K. Aberrations in Confocal Fluorescence Microscopy Induced by Mismatches in Refractive Index. *J. Microsc.* **1993**, 169, 391–405.
- [6.45] Sehayek S., Gidi Y., Glembockyte V., Brandaõ H.B., François P., Cosa G., et al. A High-Throughput Image Correlation Method for Rapid Analysis of Fluorophore Photoblinking and Photobleaching Rates. *ACS Nano* **2019**, 13, 11955–66.
- [6.46] Lehmußola A., Ruusu vuori P., Selinummi J., Huttunen H., Yli-Harja O. Computational Framework for Simulating Fluorescence Microscope Images With Cell Populations. *IEEE Trans. Med. Imaging* **2007**, 26, 1010–6.
- [6.47] Girsault A., Lukes T., Sharipov A., Geissbuehler S., Leutenegger M., Vandenberg W., et al. SOFI Simulation Tool: A Software Package for Simulating and Testing Super-Resolution Optical Fluctuation Imaging. *PLoS One* **2016**, 11, e0161602.
- [6.48] Venkataramani V., Herrmannsdörfer F., Heilemann M., Kuner T. SuReSim: Simulating Localization Microscopy Experiments From Ground Truth Models. *Nat. Methods* **2016**, 13, 319–21.
- [6.49] Webb D.J., Brown C.M. Epi-Fluorescence Microscopy. *In Cell Imaging Techniques. Methods in Molecular Biology (Methods and Protocols)*, vol 931, 2nd ed. *Humana Press* **2012**, 29–59.
- [6.50] Xia J., Kim S.H.H., Macmillan S., Truant R. Practical Three Color Live Cell Imaging by Widefield Microscopy. *Biol. Proc. Online* **2006**, 8, 63–8.
- [6.51] Hell S.W. Toward Fluorescence Nanoscopy. *Nat. Biotechnol.* **2003**, 21, 1347–55.

- [6.52] Demandolx D., Davoust J. Multicolour Analysis and Local Image Correlation in Confocal Microscopy. *J. Microsc.* **1997**, 185, 21–36.
- [6.53] Anderson M., Motta R., Chandrasekar S., Stokes M. Proposal for a Standard Default Color Space for the Internet—sRGB. *Color Imaging Conf.* **1996**, 238–45.
- [6.54] Cromey D.W. Digital Images Are Data: And Should Be Treated as Such. *In* Cell Imaging Techniques. Methods in Molecular Biology (Methods and Protocols), vol 931, 2nd ed. *Humana Press* **2013**, 1–27.
- [6.55] Bolte S., Cordelières F.P. A Guided Tour Into Subcellular Colocalization Analysis in Light Microscopy. *J. Microsc.* **2006**, 224, 213–32.
- [6.56] Schindelin J., Arganda-Carreras I., Frise E., Kaynig V., Longair M., Pietzsch T., et al. Fiji: An Open-Source Platform for Biological-Image Analysis. *Nat. Methods* **2012**, 9, 676–82.
- [6.57] Miettinen M.S. Molecular Dynamics Simulation Trajectory of a Fully Hydrated DMPC Lipid Bilayer [Data set], <http://doi.org/10.5281/zenodo.51635>, **2013**.
- [6.58] Tinevez J.Y., Perry N., Schindelin J., Hoopes G.M., Reynolds G.D., Laplantine E., et al. Trackmate: An Open and Extensible Platform for Single-Particle Tracking. *Methods* **2017**, 115, 80–90.
- [6.59] Sergeeva E.A., Katichev A.R., Kirillin M.Y. Two-Photon Fluorescence Microscopy Signal Formation in Highly Scattering Media: Theoretical and Numerical Simulation. *Quantum Electron.* **2010**, 40, 1053–61.
- [6.60] Preza C., Snyder D.L., Conchello J.-A. Theoretical Development and Experimental Evaluation of Imaging Models for Differential-Interference-Contrast Microscopy. *J. Opt. Soc. Am. A* **1999**, 16, 2185–99.
- [6.61] Preza C. Simulating Structured-Illumination Microscopy in the Presence of Spherical Aberrations. *Proc. SPIE 7904, Three-Dimensional Multidimens. Microsc. Image Acquis. Process. XVIII*, **2011**, 79040D.
- [6.62] Shterman D., Gjonaj B., Bartal G. Experimental Demonstration of Multi Moiré Structured Illumination Microscopy. *ACS Photonics* **2018**, 5, 1898–902.
- [6.63] Uster P.S., Pagano R.E. Resonance Energy Transfer Microscopy: Observations of Membrane-Bound Fluorescent Probes in Model Membranes and in Living Cells. *J. Cell Biol.* **1986**, 103, 1221–34.

- [6.64] Corry B., Hurst A.C., Pal P., Nomura T., Rigby P., Martinac B. An Improved Open-Channel Structure of MscL Determined From FRET Confocal Microscopy and Simulation. *J. Gen. Physiol.* **2010**, 136, 483–94.
- [6.65] Selin M., Fogelqvist E., Holmberg A., Guttmann P., Vogt U., Hertz H.M. 3D Simulation of the Image Formation in Soft X-Ray Microscopes. *Opt. Express* **2014**, 22, 30756–68.
- [6.66] Noji H., Yasuda R., Yoshida M., Kinosita Jr K. Direct Observation of the Rotation of F1-ATPase. *Nature* **1997**, 386, 299–302.
- [6.67] Veatch S.L., Keller S.L. Separation of Liquid Phases in Giant Vesicles of Ternary Mixtures of Phospholipids and Cholesterol. *Biophys. J.* **2003**, 85, 3074–83.
- [6.68] Sage D., Neumann F.R., Hediger F., Gasser S.M., Unser M. Automatic Tracking of Individual Fluorescence Particles: Application to the Study of Chromosome Dynamics. *IEEE Trans. Image Process.* **2005**, 14, 1372–83.
- [7.1] Ziebarth J, Wang Y. Coarse-Grained Molecular Dynamics Simulations of DNA Condensation by Block Copolymer and Formation of Core - Corona Structures. *J Phys Chem B* 2010;114:6225–32.
- [7.2] Wei Z, Luijten E. Systematic coarse-grained modeling of complexation between small interfering RNA and polycations. *J Chem Phys* 2015;143:243146.
- [7.3] Sun C, Tang T, Uludağ H. Molecular dynamics simulations of PEI mediated DNA aggregation. *Biomacromolecules* 2011;12:3698–707.
- [7.4] Sun C, Tang T, Uludağ H, Cuervo JE. Molecular dynamics simulations of DNA/PEI complexes: Effect of PEI branching and protonation state. *Biophys J* 2011;100:2754–63.
- [7.5] Sun C, Tang T, Uludağ H. Molecular dynamics simulations for complexation of DNA with 2 kDa PEI reveal profound effect of PEI architecture on complexation. *J Phys Chem B* 2012;116:2405–13.
- [7.6] Rehman ZU, Hoekstra D, Zuhorn IS. Mechanism of polyplex- and lipoplex-mediated delivery of nucleic acids: Real-time visualization of transient membrane destabilization without endosomal lysis. *ACS Nano* 2013;7:3767–77.
- [7.7] Ziebarth JD, Wang Y. Understanding the protonation behavior of linear polyethylenimine in solutions through Monte Carlo simulations. *Biomacromolecules* 2010;11:29–38.
- [7.8] Lee MS, Salsbury Jr. FR, Brooks III CL. Constant-pH Molecular Dynamics Using

- Continuous Titration Coordinates. *Proteins Struct Funct Genet* 2004;56:738–52.
- [7.9] John S, Hussain M, Turabe U, Dhand C, Venkatesh M, Tze E, et al. Insight into membrane selectivity of linear and branched polyethylenimines and their potential as biocides for advanced wound dressings. *Acta Biomater* 2016;37:155–64.
- [7.10] Kostritskii AY, Kondinskaia DA, Nesterenko AM, Gurtovenko AA. Adsorption of Synthetic Cationic Polymers on Model Phospholipid Membranes: Insight from Atomic-Scale Molecular Dynamics Simulations. *Langmuir* 2016;32:10402–14.
- [7.11] Louhivuori M, Risselada HJ, van der Giessen E, Marrink SJ. Release of content through mechano-sensitive gates in pressurized liposomes. *Proc Natl Acad Sci* 2010;107:19856–60.
- [7.12] Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, De Vries AH. The MARTINI force field: Coarse grained model for biomolecular simulations. *J Phys Chem B* 2007;111:7812–24.
- [7.13] Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ. The MARTINI coarse-grained force field: Extension to proteins. *J Chem Theory Comput* 2008;4:819–34.

Appendix A

Supporting Information for Chapter 3*

*A version of this chapter's sections has been published. Adapted from Subhamoy Mahajan and Tian Tang, Martini Coarse-Grained Model for Polyethylenimine. *J. Comp. Chem* **2019**, 40, 607-618, with permission from *Wiley Periodicals, Inc.* Copyright © 2018.

Based on our mapping scheme, any PEI structure will consist of six different beads 'tq', 't', 'sq', 's', 'pq' and 'p'. According to the nomenclature discussed in **Section 3.2.4**, a primary bead can only appear at the end of the name for a bond length, a normal bond angle or a normal dihedral angle. Consequently, the number of different bond lengths '*i-j*' is $4 \times 6 = 24$, the number of different normal bond angles '*i-j-k*' is $4 \times 4 \times 6 = 96$, and the number of different normal dihedral angles '*i-j-k-l*' is $4 \times 4 \times 4 \times 6 = 384$. The N-type bond angles '*Ni-j-k*' is symmetrical, with '*j*' being a tertiary bead but no restrictions on beads '*i*' and '*k*'. When $i \neq k$, the number of ways of choosing two different beads *i* and *k* without a specific order is 6C_2 . For $i = k$, the number of combinations is trivially 6. So, the total number of N-type bond angles is $2 \times ({}^6C_2 + 6) = 42$, where the factor of 2 accounts for '*j*' being 'tq' or 't'. The N-type dihedral angle '*Ni-j-k-l*' is not symmetric, and the bead '*j*' is always tertiary, and the bead *k* cannot be primary. So, the total possible N-type dihedral angles is $6 \times 2 \times 4 \times 6 = 288$. In summary, a general PEI structure can have at most 24 bond lengths, 138 bond angles and 672 dihedral angles.

However, the nitrogens in PEI get protonated based on their pKa values, and after one nitrogen gets protonated the pKa values of other nitrogens change. This makes many of the combinations mentioned above unrealistic and only a fraction of the combinations exists in

practice. For example, in the context of gene delivery, PEIs are typically protonated at a ratio that is less than 50%.[1] In addition, the primary and secondary nitrogens usually get protonated first, which subsequently suppresses the protonation of tertiary nitrogens. Finally, it is unlikely that two adjacent nitrogens will both get protonated. With these constraints, it can be shown that the number of bond lengths, bond angles and dihedral angles reduces drastically to 12, 49 and 152 respectively. In our model, we have parameterized 12 bond lengths, 31 bond angles and 55 dihedral angles. This accounts for 100% of the bond lengths, 63% of the bond angles and 36% of the dihedral angles, a significant portion of the possible bonded parameters.

The bond length, bond angle and dihedral angle parameters are shown in **Table A.1-A3** respectively, whereas their corresponding distributions are shown in **Figure A.1-A3**.

Table A.1: Bond length parameters, $r_{bond,eq}$ (nm) and K_{bond} (kJ/nm²/mol) are defined in Eq. 3.3.

<i>bond</i>	K_{bond}	$r_{bond,eq}$	<i>bond</i>	K_{bond}	$r_{bond,eq}$	<i>bond</i>	K_{bond}	$r_{bond,eq}$
<i>t-t</i>	20000	0.335	<i>t-sq</i>	6000	0.286	<i>t-s</i>	20000	0.330
<i>t-pq</i>	7000	0.296	<i>t-p</i>	3000	0.303	<i>sq-t</i>	6000	0.342
<i>sq-s</i>	5000	0.357	<i>s-t</i>	20000	0.360	<i>s-sq</i>	5000	0.357
<i>s-s</i>	10000	0.364	<i>s-pq</i>	5000	0.360	<i>s-p</i>	12000	0.376

Table A.2: Bond angle parameters, $\theta_{ang,eq}$ (deg) and K_{ang} (kJ/mol) are defined in Eq. 3.4.

<i>ang</i>	K_{ang}	$\theta_{ang,eq}$	<i>ang</i>	K_{ang}	$\theta_{ang,eq}$	<i>ang</i>	K_{ang}	$\theta_{ang,eq}$
<i>t-t-t</i>	100	132	<i>t-t-s</i>	50	157	<i>t-t-pq</i>	100	128
<i>t-t-p</i>	50	145	<i>t-sq-s</i>	20	134	<i>t-s-sq</i>	70	151
<i>t-s-s</i>	500	152	<i>t-s-pq</i>	200	180	<i>t-s-p</i>	200	180
<i>sq-t-t</i>	20	91	<i>sq-t-s</i>	50	125	<i>sq-t-pq</i>	200	120

<i>sq-s-t</i>	20	159	<i>sq-s-sq</i>	20	121	<i>sq-s-s</i>	50	141
<i>s-t-t</i>	200	138	<i>s-t-sq</i>	200	133	<i>s-t-s</i>	100	154
<i>s-t-pq</i>	50	145	<i>s-t-p</i>	30	180	<i>s-sq-t</i>	20	141
<i>s-sq-s</i>	20	168	<i>s-s-t</i>	500	180	<i>s-s-sq</i>	200	180
<i>s-s-s</i>	500	180	<i>s-s-pq</i>	200	180	<i>Nt-t-t</i>	1700	79
<i>Nt-t-s/Ns-t-t</i>	1000	68	<i>Nsq-t-sq</i>	100	87	<i>Ns-t-s</i>	700	88
<i>Ns-t-pq/Npq-t-s</i>	2000	80	<i>Ns-t-p/Np-t-s</i>	100	50			

Table A.3: Dihedral angle parameters, $K_{dih,w}$ (kJ/mol), $n_{dih,w}$ and $\varphi_{dih,eq,w}$ (deg) are defined in Eq. 3.5.

$\varphi_{dih,eq,w}$	$K_{dih,w}$	$n_{dih,w}$	$\varphi_{dih,eq,w}$	$K_{dih,w}$	$n_{dih,w}$	$\varphi_{dih,eq,w}$	$K_{dih,w}$	$n_{dih,w}$
<i>t-t-t-s</i>			<i>t-t-t-pq</i>			<i>t-t-t-p</i>		
180	0.7611	1	180	5.005	1	180	0.575	1
90	1.5556	1	-90	1.058	1	90	3.14347	1
180	1.18076	2	180	2.794	2	0	3.5126	2
-90	1.59562	2	-90	1.124	2	90	1.37915	2
0	0.85896	3	180	1.061	3	180	0.59154	3
90	1.10652	3	-90	0.6954	3	-90	1.0108	3
0	0.35776	4	180	0.1284	4	180	1.5454	4
90	2.429	4	-90	0.2789	4	90	0.56349	4
0	0.28804	5						
90	1.25512	5						
0	0.37902	6						
90	0.51536	6						

<i>t-t-s-pq</i>			<i>t-sq-s-t</i>			<i>t-sq-s-s</i>		
180	2.273	1	180	0.1635	1	0	0.1789	1
-90	0.1678	1	90	0.05581	1	90	0.6414	1
180	0.323	2	0	0.4016	2	0	0.04005	2
-90	0.146	2	90	0.107	2	90	0.1861	2
0	0.1812	3	0	0.1962	3	0	0.00052	3
90	0.04275	3	-90	0.2463	3	90	0.287	3
0	0.1086	4	0	0.3595	4	0	0.1764	4
90	0.0268	4	90	0.03185	4	90	0.3615	4
<i>t-s-sq-t</i>			<i>t-s-sq-s</i>			<i>t-s-s-t</i>		
180	0.5667	1	180	0.03045	1	180	0.2794	1
-90	0.2793	1	-90	0.09518	1	-90	0.4257	1
0	0.13	2	0	0.834	2	180	0.4685	2
-90	0.6036	2	-90	0.2247	2	90	0.2221	2
180	0.00712	3	0	0.05384	3	180	0.01302	3
-90	0.02012	3	90	0.03155	3	90	0.02205	3
0	0.04612	4	180	0.08784	4	180	0.05282	4
-90	0.05911	4	-90	0.05518	4	-90	0.1746	4
<i>t-s-s-s</i>			<i>sq-t-t-t</i>			<i>sq-t-t-pq</i>		
0	0.9592	1	180	2.624	1	180	3.967	1
-90	0.3031	1	90	2.487	1	-90	4.394	1
0	0.07864	2	180	0.3305	2	0	0.1698	2
-90	0.3132	2	-90	1.375	2	-90	0.8341	2
180	0.00904	3	0	1.151	3	180	0.8689	3

-90	0.1049	3	90	0.6597	3	90	0.918	3
180	0.08386	4	0	0.9143	4	180	0.5322	4
-90	0.03749	4	90	0.9613	4	90	0.1127	4
<i>sq-t-t-sq</i>			<i>sq-t-t-s</i>			<i>sq-s-t-sq</i>		
180	4.466	1	180	0.6084	1	180	0.275	1
-90	0.8974	1	90	0.5301	1	-90	0.117	1
0	0.3155	2	0	0.8167	2	0	0.2408	2
90	0.449	2	90	0.2646	2	-90	0.05002	2
0	0.9992	3	0	0.2556	3	0	0.2305	3
90	0.3155	3	90	0.491	3	-90	0.09074	3
180	0.1448	4	180	0.07	4	0	0.2816	4
-90	0.00843	4	90	0.8255	4	-90	0.04807	4
			180	0.3319	5			
			-90	0.4926	5			
<i>sq-s-t-s</i>			<i>sq-s-t-pq</i>			<i>sq-s-sq-s</i>		
180	0.2602	1	180	0.1808	1	180	0.9085	1
90	0.43	1	-90	0.4601	1	90	0.1517	1
0	0.3921	2	0	0.05521	2	0	0.08763	2
90	0.0492	2	90	0.1435	2	90	0.1437	2
0	0.08434	3	0	0.508	3	0	0.08808	3
90	0.346	3	-90	0.1165	3	-90	0.09503	3
180	0.2742	4	180	0.08161	4	0	0.04727	4
90	0.1182	4	90	0.03025	4	-90	0.03526	4
<i>sq-s-s-s</i>			<i>sq-s-s-pq</i>			<i>s-t-t-t</i>		

180	0.578	1	180	0.2917	1	180	6.07634	1
90	0.5483	1	-90	0.08782	1	-90	4.39625	1
0	0.02419	2	180	0.1641	2	180	1.1485	2
90	0.06867	2	-90	0.09495	2	-90	0.12722	2
0	0.07933	3	0	0.04617	3	180	2.13296	3
-90	0.04552	3	90	0.01113	3	90	1.08089	3
180	0.00144	4	180	0.01469	4	180	0.72975	4
-90	0.0156	4	-90	0.07517	4	-90	0.70235	4
<i>s-t-t-s</i>			<i>s-t-t-p</i>			<i>s-t-sq-s</i>		
0	4.34	1	180	4.771	1	180	0.5226	1
90	3.11083	1	-90	2.489	1	-90	0.589	1
0	0.16994	2	180	1.36	2	0	0.3605	2
90	2.3226	2	90	0.06439	2	-90	0.4849	2
0	0.55404	3	180	1.346	3	0	0.1488	3
90	0.9776	3	90	0.9888	3	90	0.01395	3
0	0.66166	4	180	0.5621	4	180	0.2104	4
90	0.39966	4	90	0.05767	4	90	0.03304	4
<i>s-t-s-s</i>			<i>s-t-s-pq</i>			<i>s-t-s-p</i>		
180	4.958	1	180	2.7435	1	180	3.269	1
-90	0.4269	1	-90	0.14	1	-90	2.217	1
180	0.7339	2	0	0.0732	2	180	0.6888	2
-90	0.05046	2	-90	0.3342	2	-90	0.6913	2
0	0.3602	3	0	0.0365	3	0	0.09744	3
90	0.612	3	-90	0.2216	3	-90	0.8553	3

180	0.0559	4	180	0.0375	4	180	0.2292	4
90	0.3238	4	90	0.05039	4	-90	0.3393	4
<i>s-sq-t-s</i>			<i>s-sq-t-pq</i>			<i>s-sq-s-t</i>		
0	0.0455	1	180	0.4716	1	0	0.08534	1
-90	0.01309	1	-90	0.166	1	90	0.1719	1
0	0	2	0	0.3157	2	0	0.4751	2
90	0.07727	2	90	0.1404	2	90	0.5068	2
0	0.4454	3	180	0.1711	3	0	0.2635	3
-90	0.1521	3	90	0.512	3	-90	0.1141	3
0	0.03718	4	0	0.07675	4	0	0.03519	4
-90	0.05089	4	90	0.0176	4	-90	0.1021	4
<i>s-sq-s-sq</i>			<i>s-sq-s-s</i>			<i>s-s-t-s</i>		
0	0.01675	1	180	0.5089	1	0	1.058	1
90	0.00732	1	90	0.1826	1	90	0.3069	1
0	0.2463	2	180	0.2	2	0	0.08945	2
90	0.1003	2	-90	0.116	2	-90	0.153	2
0	0.06287	3	0	0.1657	3	0	0.152	3
90	0.07754	3	-90	0.1259	3	-90	0.1812	3
180	0.00352	4	180	0.05827	4	180	0.08197	4
-90	0.01123	4	90	0.00657	4	-90	0.1495	4
<i>s-s-t-pq</i>			<i>s-s-t-p</i>			<i>s-s-sq-s</i>		
180	0.8354	1	180	1.481	1	180	1.206	1
-90	0.03209	1	-90	2.452	1	90	0.8237	1
180	0.2513	2	180	0.6452	2	0	0.1526	2

90	0.1863	2	-90	0.1292	2	-90	0.3173	2
0	0.1615	3	0	0.01608	3	180	0.00028	3
-90	0.09505	3	90	0.193	3	90	0.06832	3
0	0.1572	4	180	0.02013	4	0	0.156	4
-90	0.2764	4	-90	0.05453	4	90	0.07099	4
<i>s-s-s-t</i>			<i>s-s-s-sq</i>			<i>s-s-s-s</i>		
0	0.4347	1	0	0.4184	1	0	0.07194	1
90	0.7823	1	-90	0.175	1	-90	0.03906	1
0	0.2341	2	0	0.1899	2	0	0.1423	2
90	0.3782	2	-90	0.391	2	-90	0.07014	2
0	0.01605	3	0	0.04892	3	0	0.07076	3
90	0.2112	3	-90	0.1252	3	-90	0.03901	3
0	0.00614	4	0	0.05003	4	180	0.02093	4
90	0.1048	4	-90	0.09302	4	90	0.07606	4
<i>s-s-s-pq</i>			<i>Nt-t-t-t</i>			<i>Nt-t-t-s</i>		
180	0.1214	1	0	1.516	1	180	0.19626	1
90	0.2415	1	90	0.1462	1	-90	0.81532	1
180	0.1527	2	180	0.1592	2	180	0.03962	2
90	0.1075	2	-90	1.787	2	90	1.38598	2
180	0.1189	3	0	1.35744	3	180	1.15752	3
90	0.1045	3	-90	2.6552	3	-90	1.5939	3
180	0.02334	4	180	0.73934	4	180	0.34302	4
-90	0.04644	4	90	1.357	4	-90	0.12256	4
			180	0.84582	5	0	1.25506	5

			90	0.481	5	90	0.65496	5
			0	0.71588	6	180	0.60978	6
			-90	0.74564	6	-90	0.26408	6
<i>Nt-t-t-pq</i>			<i>Nt-t-t-p</i>			<i>Nt-t-s-pq</i>		
0	1.277	1	0	4.54	1	0	2.454	1
90	4.679	1	90	2.542	1	90	1.146	1
0	2.832	2	180	1.438	2	180	0.3954	2
-90	1.727	2	-90	0.2625	2	-90	0.2008	2
180	1.267	3	0	1.607	3	180	0.08704	3
-90	1.045	3	-90	1.164	3	90	0.3218	3
180	0.05346	4	180	1.233	4	180	0.306	4
90	0.4889	4	90	0.3317	4	-90	0.01641	4
<i>Nsq-t-sq-s</i>			<i>Ns-t-t-s</i>			<i>Ns-t-t-pq</i>		
0	1.972	1	0	1.0906	1	0	4.116	1
-90	0.8696	1	-90	2.9302	1	90	3.842	1
0	1.055	2	0	2.61728	2	180	0.2675	2
-90	0.6807	2	-90	1.73518	2	-90	4.374	2
180	0.09864	3	180	2.8734	3	180	1.809	3
-90	0.04226	3	-90	0.47268	3	90	2.065	3
180	0.3563	4	0	1.63992	4	0	1.188	4
-90	0.3498	4	90	0.20031	4	90	0.0058	4
<i>Ns-t-t-p</i>			<i>Ns-t-s-s</i>			<i>Npq-t-s-sq</i>		
0	1.7	1	0	3.456	1	0	4.543	1
-90	0.50968	1	90	0.8971	1	-90	1.85	1

0	4.214	2	180	1.431	2	180	0.9629	2
90	0.04158	2	-90	0.3018	2	90	1.244	2
180	0.00686	3	0	0.2482	3	180	0.1907	3
90	1.12014	3	-90	0.2842	3	-90	0.2465	3
180	0.35091	4	180	0.215	4	0	0.04138	4
90	0.36246	4	-90	0.08677	4	-90	0.1814	4
<i>Npq-t-s-s</i>			<i>Npq-t-s-pq</i>			<i>Npq-t-s-p</i>		
0	4.438	1	0	2.922	1	0	3.892	1
-90	1.372	1	90	1.656	1	90	1.224	1
180	1.278	2	180	0.4585	2	180	1.51	2
90	0.6364	2	-90	0.3462	2	-90	0.03169	2
180	0.4575	3	180	0.07853	3	0	0.8277	3
-90	2.015	3	90	0.3462	3	-90	0.00439	3
0	0.4068	4	180	0.1724	4	180	0.2952	4
90	1.244	4	90	0.1287	4	90	0.3258	4
<i>Np-t-s-pq</i>								
0	2.678	1						
-90	0.2122	1						
180	0.2772	2						
-90	0.4798	2						
180	0.04036	3						
-90	0.00104	3						
180	0.2256	4						
90	0.0028	4						

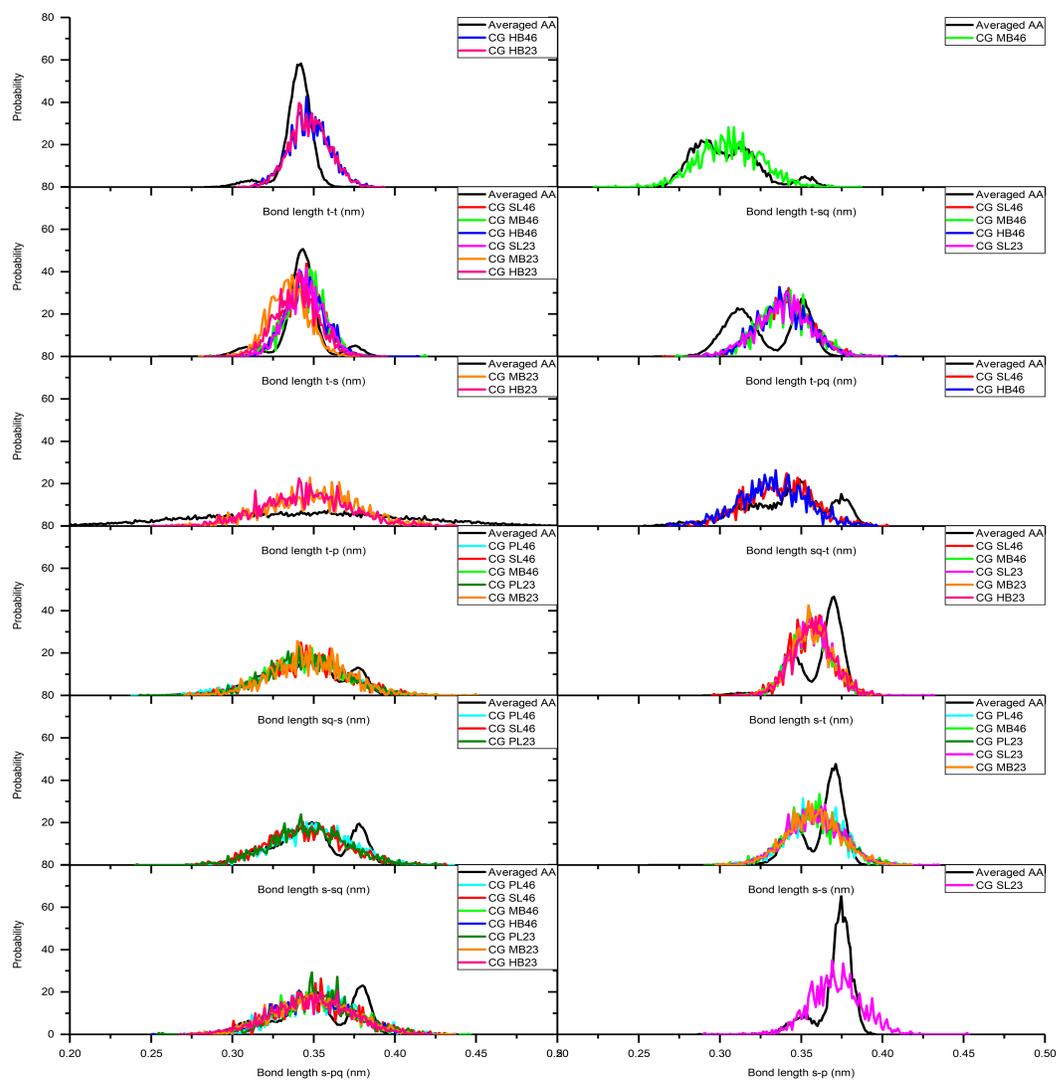
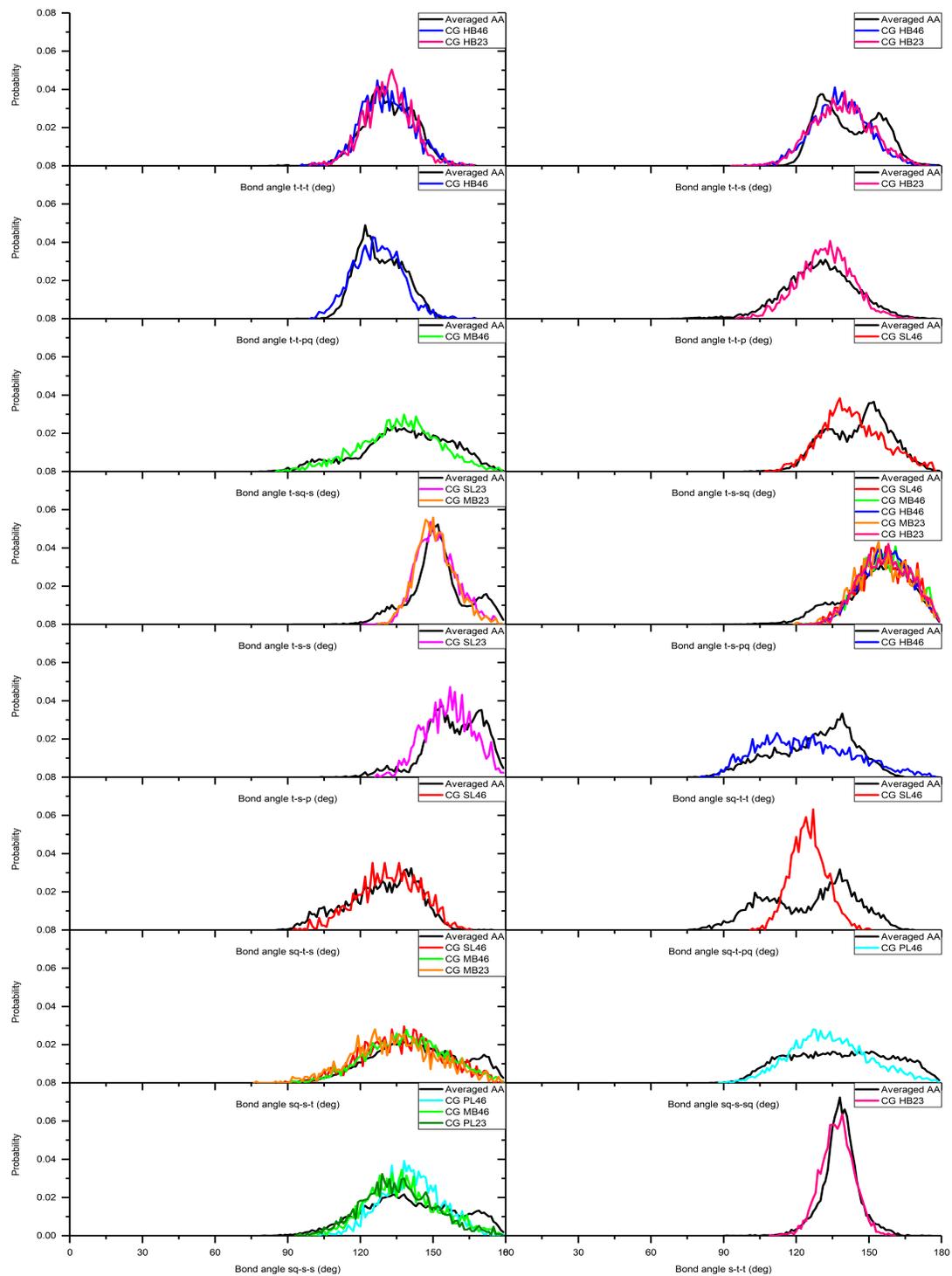


Figure A.1: Averaged AA distributions (black) and CG distributions for bond lengths: PL46 (cyan), SL46 (red), MB46 (green), HB46 (blue), PL23 (olive), SL23 (magenta), MB23 (orange), and HB23 (pink).



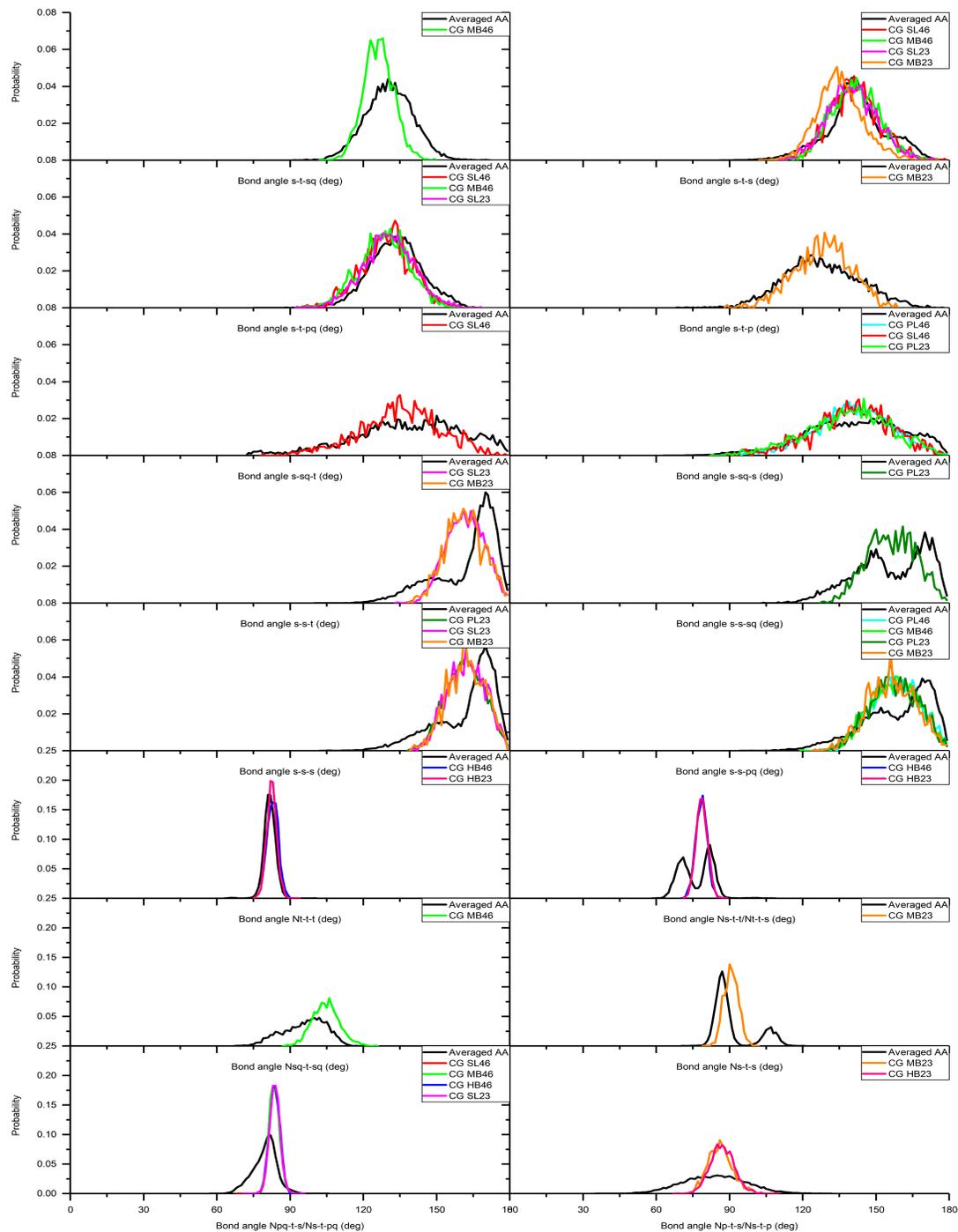
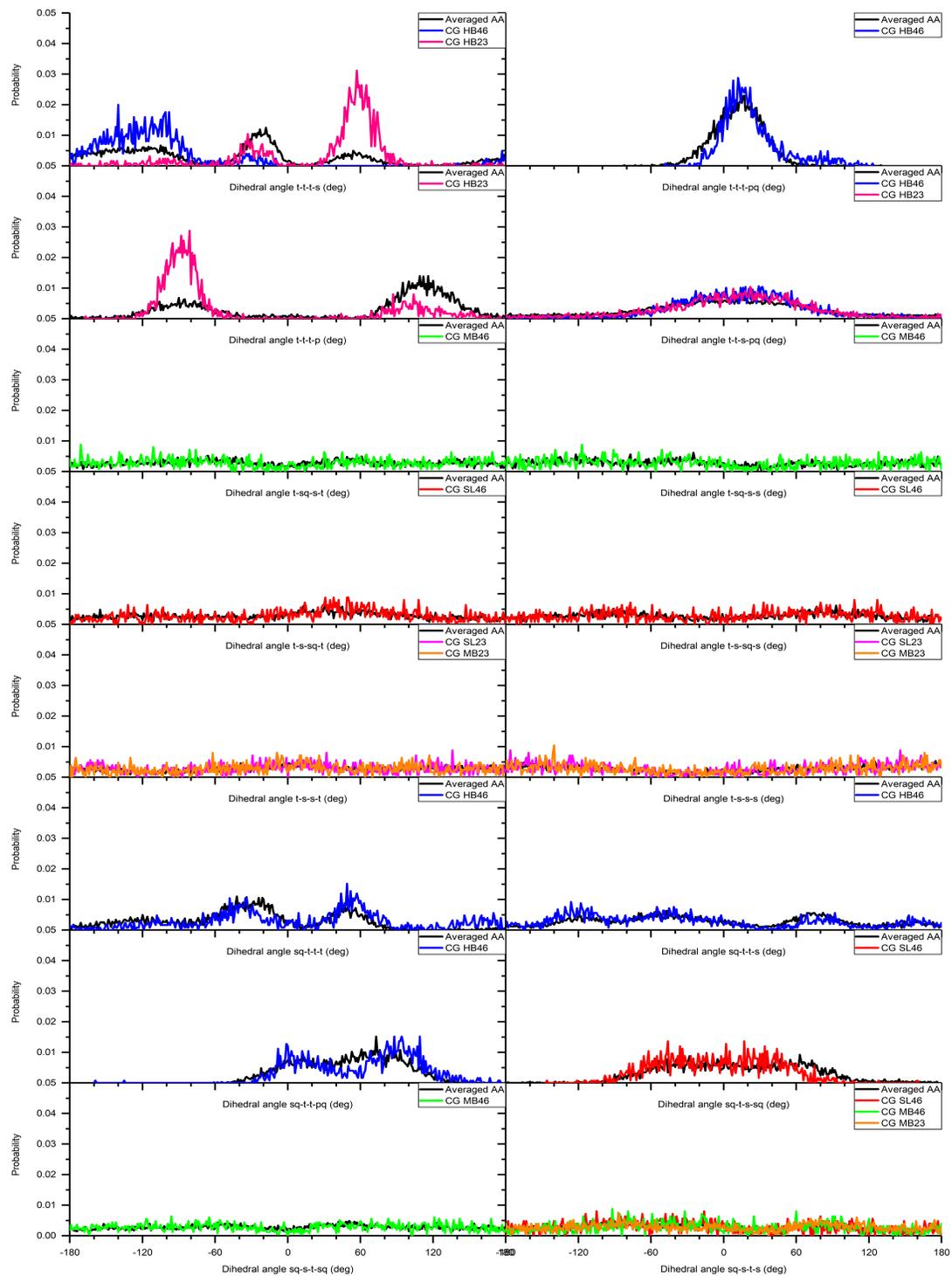
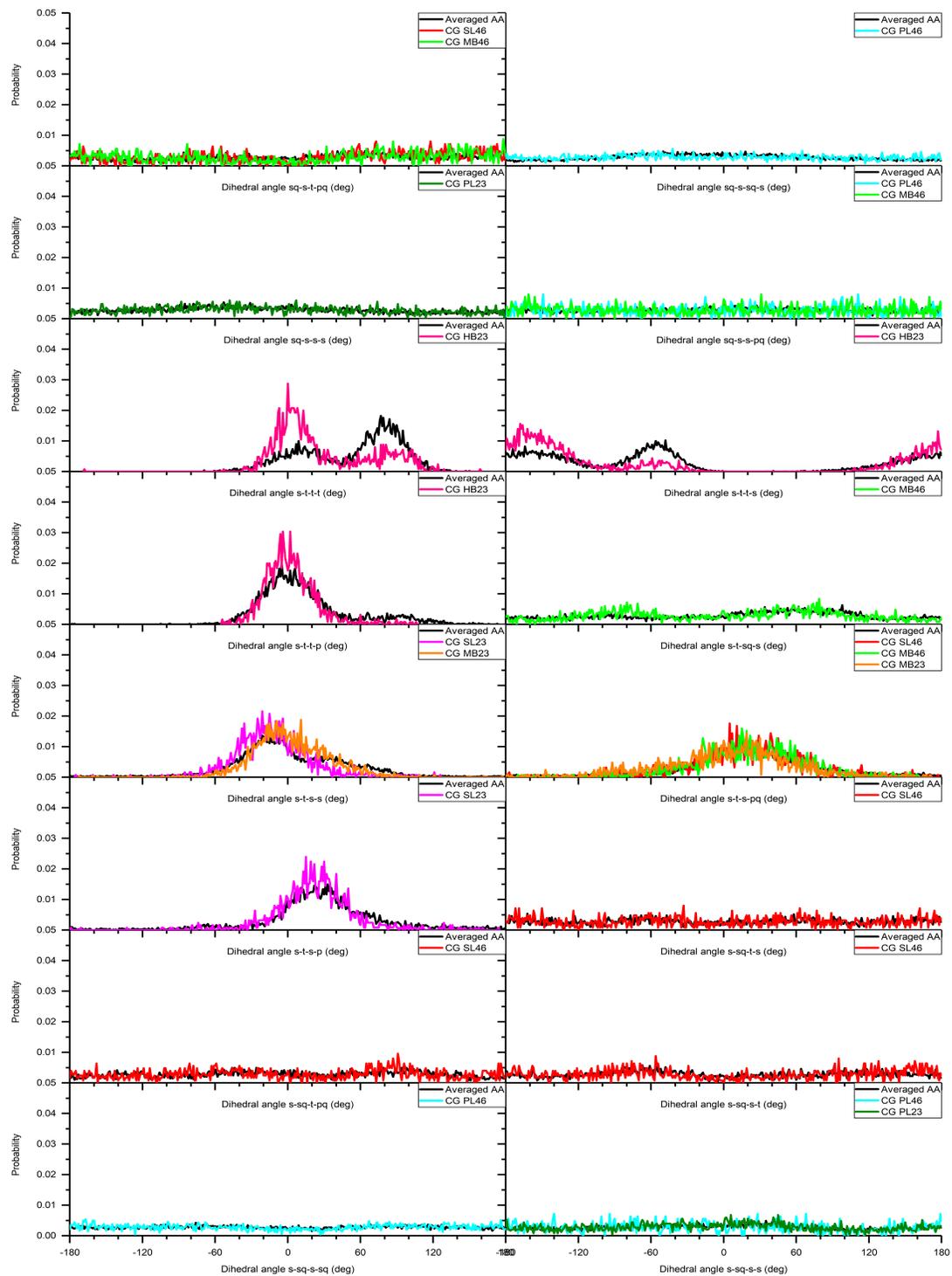


Figure A.2: Averaged AA distributions (black) and CG distributions for bond angles: PL46 (cyan), SL46 (red), MB46 (green), HB46 (blue), PL23 (olive), SL23 (magenta), MB23 (orange), and HB23 (pink).





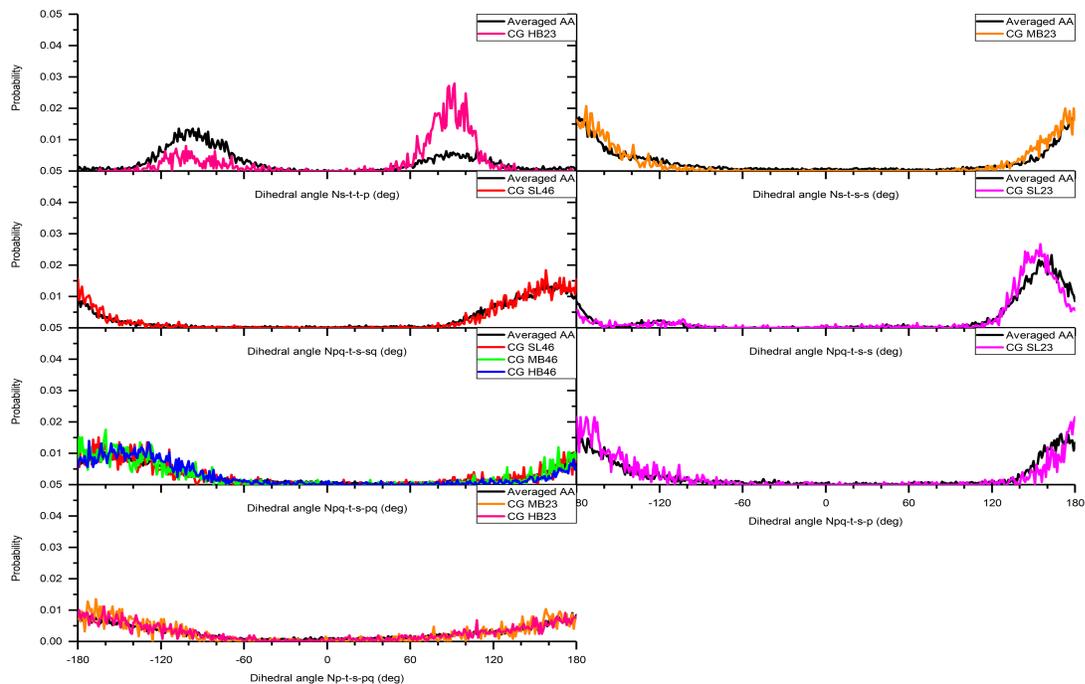


Figure A.3: Averaged AA distributions (black) and CG distributions for dihedral angles: PL46 (cyan), SL46 (red), MB46 (green), HB46 (blue), PL23 (olive), SL23 (magenta), MB23 (orange), and HB23 (pink).

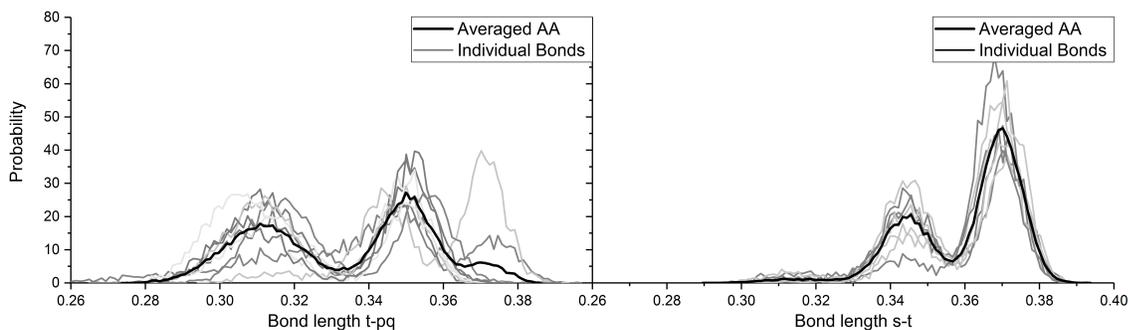


Figure A.4: Bond length distributions obtained from AA simulations are plotted, where multiple shades of grey is used to represent different AA distributions obtained from individual bonds and black represents the averaged AA distribution. The left panel is for bond length ' $t-pq$ ' and the right panel for ' $s-t$ '.

It can be seen that the average AA distributions for some bond lengths and bond angles have more than one prominent peak. In **Figure A.4** left panel, we plot the AA distributions for the

'*t-pq*' bond obtained from individual bonds (bond 1-2, 5-6, and 10-11 in SL46 and SL23, 10-11 in MB46, and 3-4 and 10-11 in HB46) in grey, and the averaged AA distribution in black. Similar plot is generated in **Figure A.4** the right panel for the '*s-t*' bond. Clearly, multiple peaks are present even for the AA distributions from individual bonds, so it is not a result of averaging the AA distributions.

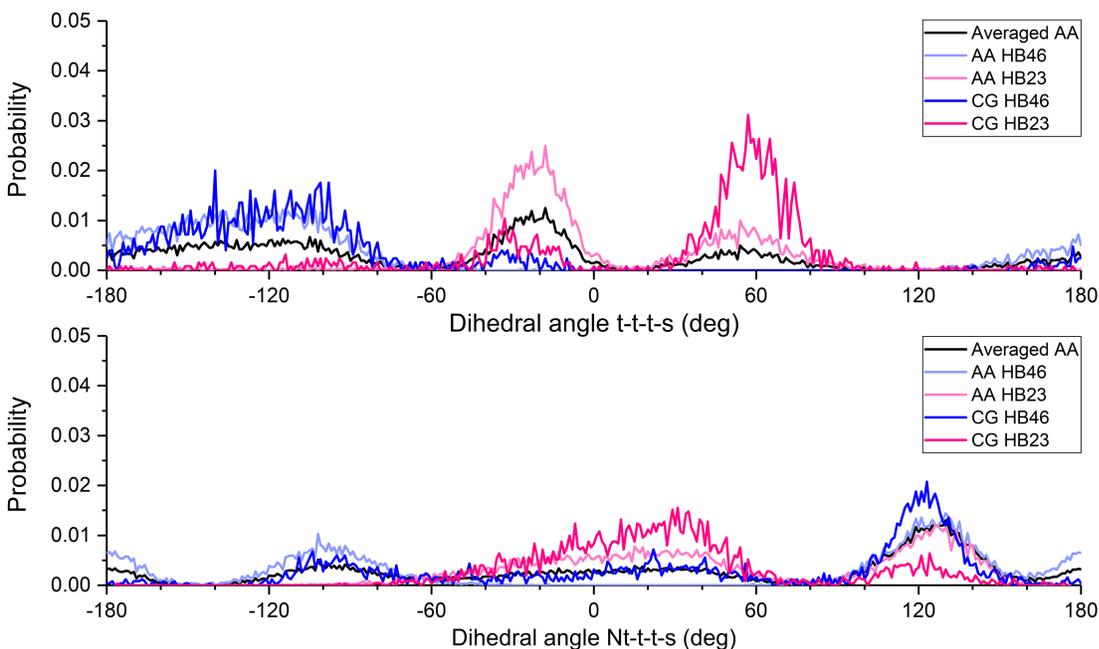


Figure A.5: Dihedral angle distributions for '*t-t-t-s*' (top) and '*Ns-t-t-s*' (bottom). For each dihedral angle five distributions were plotted namely, the averaged AA distribution (black), AA distribution only averaged over HB46 (light blue), AA distribution only averaged over HB23 (light pink), CG distribution averaged over HB46 (blue) and CG distribution averaged over HB23 (pink).

Dihedral angles '*t-t-t-s*' and '*Ns-t-t-s*' are especially interesting as the CG distributions in HB46 and HB23 do not match very well the averaged AA distributions. To check this in more detail, we computed the AA distribution of dihedral angle '*t-t-t-s*' in HB46 by performing an average over the all '*t-t-t-s*' angles present in HB46 (dihedral 5-7-10-12), but not over the dihedrals present in other structures. The result is presented in **Figure A.5** (top panel) as AA HB46. The

same is done over the ‘*t-t-t-s*’ angles present in HB23 (dihedral 5-7-10-12), shown in **Figure A.5** (top panel) as AA HB23. This subfigure also contains two curves for the corresponding CG distributions, and the averaged AA distribution. **Figure A.5** (bottom panel) shows a similar comparison for dihedral angle ‘*Ns-t-t-s*’ (dihedral 8-7-10-12 in both HB23 and HB46). For both dihedrals ‘*t-t-t-s*’, and ‘*Ns-t-t-s*’ we observe that CG distributions matches better with AA distribution than averaged AA distribution. Although the dihedral angles parameters were obtained to match the averaged AA distributions, the AA distributions of a specific structure can still be properly captured by the local structural restraints in the CG model.

Reference AA distributions were calculated for a few bonded distributions using trajectories from 50-75 ns and 75-10 ns (**Figure A.6**). Curves from different time windows almost overlap, confirming the attainment of equilibrium in the 100 ns AA simulations.

Additional comparison is made between our CG model and the AA force field recently developed by Beu *et al.*[2]. In Beu *et al.*[2] linear PEIs with 14, 26 and 50 monomers were simulated at three different protonation ratios: 50%, 33% and 25%. The exact location of the protonated nitrogens, however, were not specified. We created CG PEIs of the same length and protonated them according to two protonation ratios 50% and 25% (see **Figure A.7** for details), but we emphasize that there can be discrepancies between our work and the work of Beu *et al.* [2] on the sites of protonation. Another notable difference is that each linear PEI in Beu *et al.* [2] consists of monomeric units CCN but has an additional NH₂ group, which is not present in our model. **Figure A.8** shows the comparison of radius of gyration (R_g) and end-to-end distance (R_e) between Beu *et al.* [2] (AA) and our model (CG). Considering the relatively large standard deviation in the CG values, which is expected, the results are quite comparable. The difference is more significant in 25% protonated PEIs, suggesting that the discrepancies are likely caused by

differences in protonation sites, since there are more ways of assigning the protonated nitrogens in the case of 25% protonation ratio.

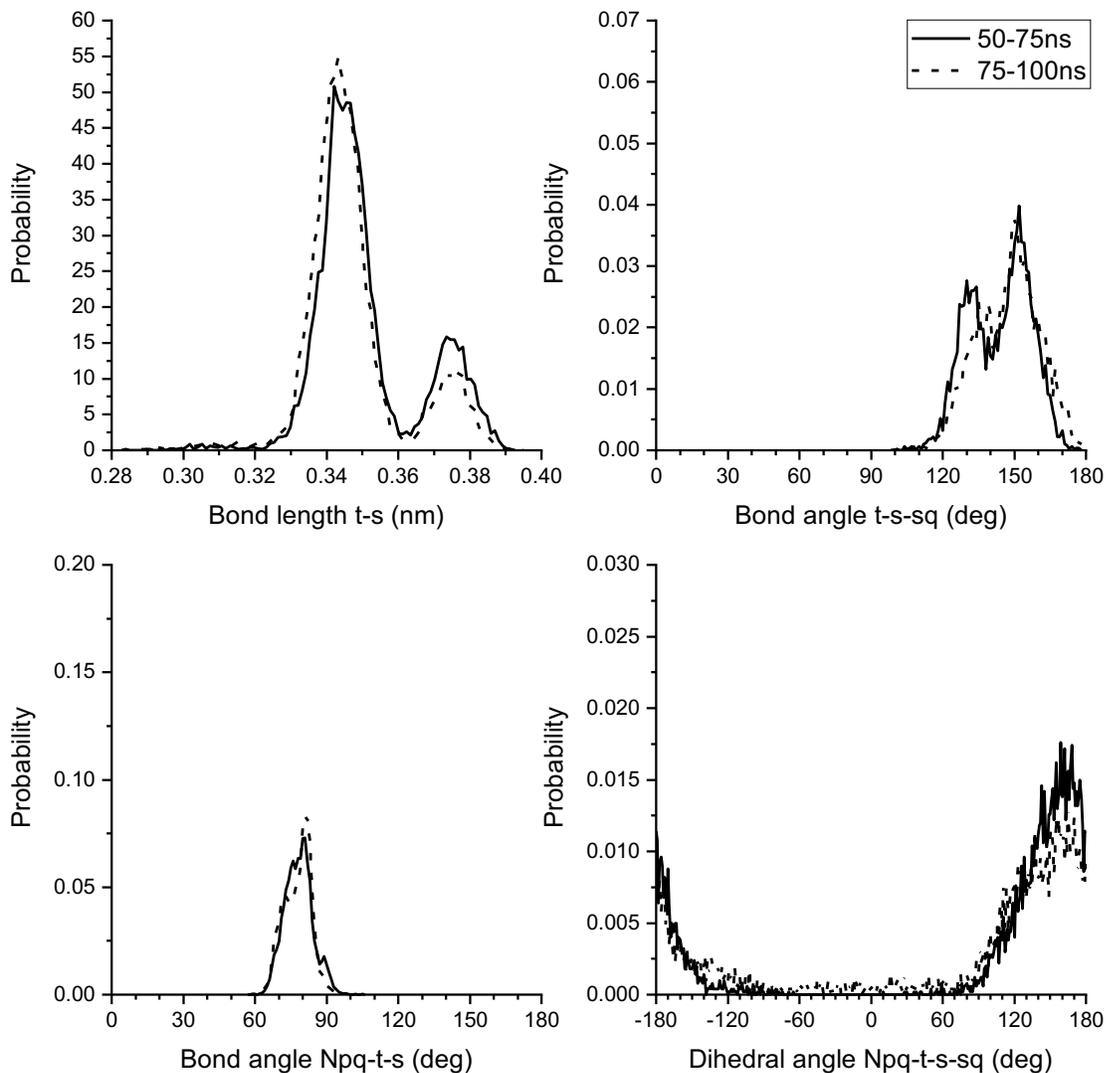
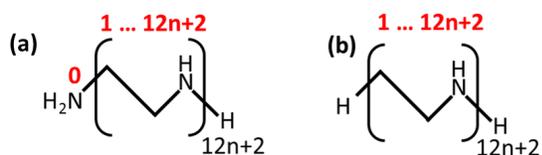


Figure A.6: Bond length distribution of ‘*t-s*’ (top left), bond angle distributions of ‘*t-s-sq*’ (top right) and ‘*Npq-t-s*’ (bottom left), and dihedral angle distribution of ‘*Npq-t-s-sq*’ (bottom right), sampled from different time windows of the AA simulations.



Protonation ratio	No. of monomers	Site of protonation
50%	14	2,4,6,8,10,12,14
	26	2,4,6,8,10,12,14,16,18,20,22,24,26
	50	2,4,6,8,10,12,14,16,18,20,22,24,26,28,30,32,34,36,38,40,42,44,46,48,50
25%	14	2,8,14
	26	1,6,11,16,21,26
	50	2,6,10,14,18,22,26,30,34,38,42,46,50

Figure A.7: (a) structure of linear PEI simulated in Beu *et al.* [2]; and (b) linear PEI created here to mimic that structure, based on which CG models were generated and simulated. The indices of protonated nitrogens in our work are given in the figure for the two protonation ratios.

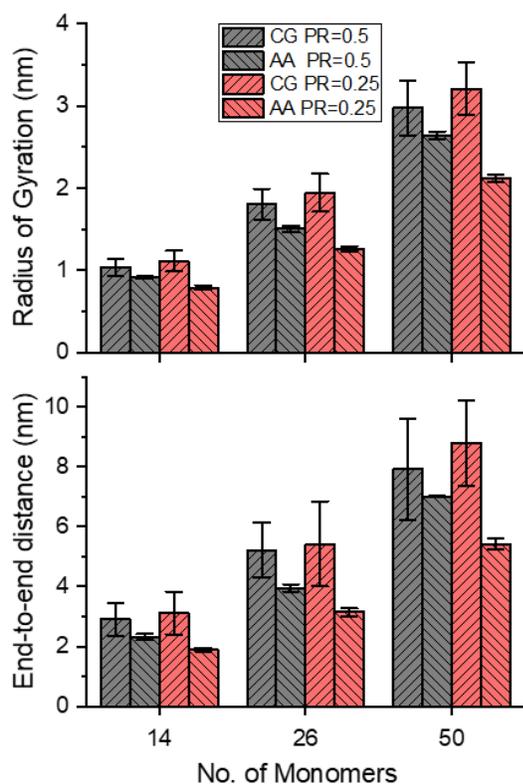


Figure A.8: Comparison of radius of gyration (top) and end-to-end distance (bottom) for linear PEIs as modeled by Beu *et al.* [2] at AA level and by our CG model.

References

- [1] Utsuno K., Uludağ H. Thermodynamics of Polyethylenimine-DNA Binding and DNA Condensation. *Biophys. J.* **2010**, 99, 201–7.
- [2] Beu T.A., Farcaş A. CHARMM Force Field and Molecular Dynamics Simulations of Protonated Polyethylenimine. *J. Comput. Chem.* **2017**, 38, 2335–48.

Appendix B

Supporting Information for Chapter 4*

*A version of this chapter's sections has been published. Adapted with permission from Subhamoy Mahajan and Tian Tang, Polyethylenimine-DNA Ratio Strongly Affects Their Nanoparticle Formation: A Large-Scale Coarse-Grained Molecular Dynamics Study. *J. Phys. Chem. B* **2019**, 123(45), 9629-9640. Copyright © 2019 American Chemical Society.

B.1. Relative shape anisotropy of nanoparticles

The shapes of NPs as seen in **Figure 5.2**, are significantly different for $\alpha = 2-6$ and $\alpha = 8-10$. To quantify the difference in shape, the relative shape anisotropy (κ^2) is calculated using Eq. S1, where χ_1 , χ_2 , and χ_3 are the three eigenvalues of the gyration tensor of the NP. κ^2 would be 0 if all the beads in the NP are spherically symmetrical about its center of mass and 1 if the beads lie on a straight line.[1] κ^2 is found to be 0.61, 0.53, 0.69, 0.30 and 0.35, respectively, for the largest NP at $\alpha = 2-10$. Therefore, the largest NP formed for $\alpha = 8-10$ are more compact and closer to being spherical than the ones formed for $\alpha = 2-6$.

$$\kappa^2 = \frac{3}{2} \frac{\chi_1^2 + \chi_2^2 + \chi_3^2}{(\chi_1 + \chi_2 + \chi_3)^2} - \frac{1}{2} \quad \text{B.1}$$

B.2. PEI-DNA binding

A PEI can bind to a DNA via Coulombic and/or Lennard-Jones interactions. That is, the binding can occur between uncharged beads. This can be seen by plotting the minimum distance between a DNA-PEI pair calculated using (i) all the beads and (ii) only the charged beads. In

Figure B.1, the minimum distance calculated using these two methods is plotted for the 23rd DNA and 75th PEI in system $\alpha = 10$. A horizontal dashed line is added to mark the cutoff distance for DNA-PEI binding (0.53 nm). From 0 to 1.25 μs the minimum distances calculated using both methods are similar, which implies that the initial attraction between DNA and PEI is dominated by long-range electrostatic interactions. After 1.25 μs , the minimum distance calculated using (ii) is above the cutoff, while the minimum distance calculated using (i) remains below the cutoff. This shows that a PEI and DNA can remain bound to each other through Lennar-Jones interactions assisted by long-range electrostatic interactions.

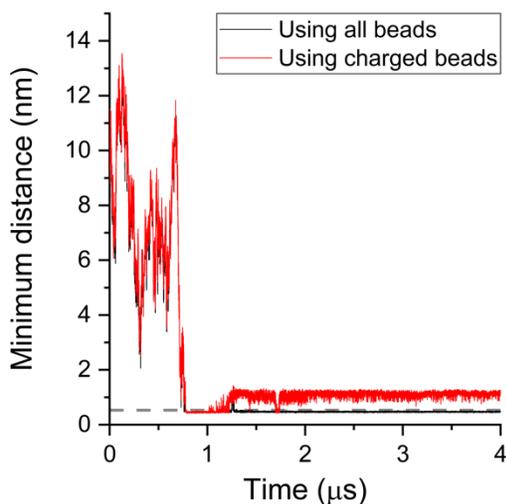


Figure B.1: Minimum distance between the 23rd DNA and 75th PEI for $\alpha = 10$, calculated using all the beads (black), and only charged beads (red).

B.3. Conversion of PEI roles

Over the simulation time of 4 μs , the role of each PEI (free/peripheral/bridging) was determined for every 0.2 ns. For each PEI, its role was determined between two consecutive time steps (0.2 ns apart) to track the conversion. **Table C1** reports these total number of conversions between free and peripheral ($f \rightarrow p, p \rightarrow f$), between peripheral and bridging ($p \rightarrow b, b \rightarrow p$), and

between free and bridging ($f \rightarrow b$, $b \rightarrow f$) PEIs. Clearly, the conversions between free and bridging PEI is 2-4 orders of magnitude lower than other conversions. This indicates bridging PEI is primarily formed from peripheral PEI and peripheral PEI is primarily formed from free PEI.

Table B.1: Total number of conversions between free (f), peripheral (p) and bridging (b) PEIs over the 0-4 μ s of the simulation.

α	$f \rightarrow p$	$p \rightarrow f$	$p \rightarrow b$	$b \rightarrow p$	$f \rightarrow b$	$b \rightarrow f$
2	266	218	2529	2515	1	1
4	747	653	15718	15669	4	2
6	1537	1394	26008	25960	10	5
8	3962	3771	34302	34233	15	18
10	5086	4854	37530	37453	35	41

B.4. Average number of bridging PEIs between a pair of bridged DNAs (n_b)

To calculate n_b we first find the number of bridging PEIs and number of DNA pairs that are bridged together. They are evaluated every 0.2 ns and averaged over a range of 10 ns. We then evaluate n_b as the ratio of the total number of bridging PEIs and the total number of bridged DNA pairs. The result is given in **Figure 5.3d** as a function of time.

For the five PEI/DNA ratios (α) studied n_b forms three distinct groups, one for $\alpha = 2$, one for $\alpha = 4$ and 6, and one for $\alpha = 8$ and 10. This can be explained as follows. When PEIs bind with a DNA, they neutralize some of its phosphates and screen their electric field. Additional PEIs binding to the same DNA would likely bind near phosphates which are not screened. Moreover, bound PEIs would repel each other, so to reduce the potential energy they would be as far apart as possible. Therefore, it is reasonable to expect bound PEIs to be equally-spaced around a DNA,

which is supported by our observations in **Figure 5.4c**. Aggregating DNAs are in contact through a small portion of their surface area which we refer to as the *bridging region*. Therefore, only a fraction of PEIs bound to a DNA can become bridging PEIs, and they are located in the bridging region. Moreover, the number of bridging PEIs between each pair of bridged DNAs has to be a whole number. As a result when α increases, n_b cannot increase continuously, but rather in a step-wise fashion. For our systems, we observe the increase in n_b between $\alpha = 2$ and 4, and between $\alpha = 6$ and 8.

The value of n_b can also increase if there is an increase in the size of the bridging region. For example, DNAs aggregating side-by-side would have a larger bridging region than DNAs aggregating end-to-end. Therefore, the higher value of n_b for $\alpha = 8$ and 10 compared to $\alpha = 4$ and 6 is probably due to a combination of more bound PEIs and increased bridging region.

B.5. Additional mechanism of DNA over-saturation

Clearly from **Figure 5.4c**, over-saturation of DNAs for systems with excess PEIs mainly occurs through misalignment of center-binding PEIs, which accommodates more protonated amine beads in the NP than the number of phosphate beads. Another mechanism for DNA over-saturation is through the simultaneous binding of one phosphate bead with more than one protonated amine beads. **Figure B.2** plots the number of phosphate beads $N_{phos}(n)$, that are in close contact with n protonated amine beads. The number of phosphate beads without any close contact with protonated amine beads, $N_{phos}(0)$, decreases with increasing α , which is expected as the number of bound PEIs increases with α . Consequently, the number of phosphate beads bound with only one protonated amine bead, $N_{phos}(1)$, increases with α . There is also a fraction of phosphate beads, up to 20%, in close contact with two protonated amine beads, which increases

monotonously with α aiding in the over-saturation of DNAs. Simultaneous contact of a phosphate bead with 3 protonated amine beads is rare.

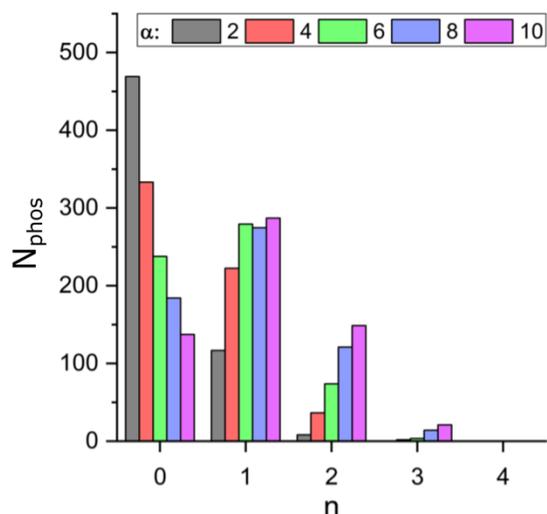


Figure B.2: Number of deprotonated phosphate beads $N_{phos}(n)$, that are in close contact with n protonated amine beads.

B.6. Determination of steady-state

First, NP size distribution was evaluated at every 100 ns, by calculating the number of NPs for each s_{NP} at every 0.2 ns and averaging it over a time range of 100 ns. The NP size distribution and NP charge as a function of size were considered to be at the steady-state if the NP size distribution did not change significantly for over 1 μ s. Every system was observed to reach steady-state at 3.5 μ s.

B.7. Time-dependent nanoparticle charge

The charge of NPs can be predicted using an extended time-dependent model. Since each DNA is likely to be neutralized by a similar number of peripheral PEIs before they aggregate, charge of NPs can be predicted using Eq. B.2. In this equation, each DNA is assumed to have the

same number of bound PEI $\alpha' = (N_{PEI} - N_f(t; \alpha))/N_{DNA}$, where $N_{PEI} - N_f(t; \alpha)$ is total number of bound PEIs.

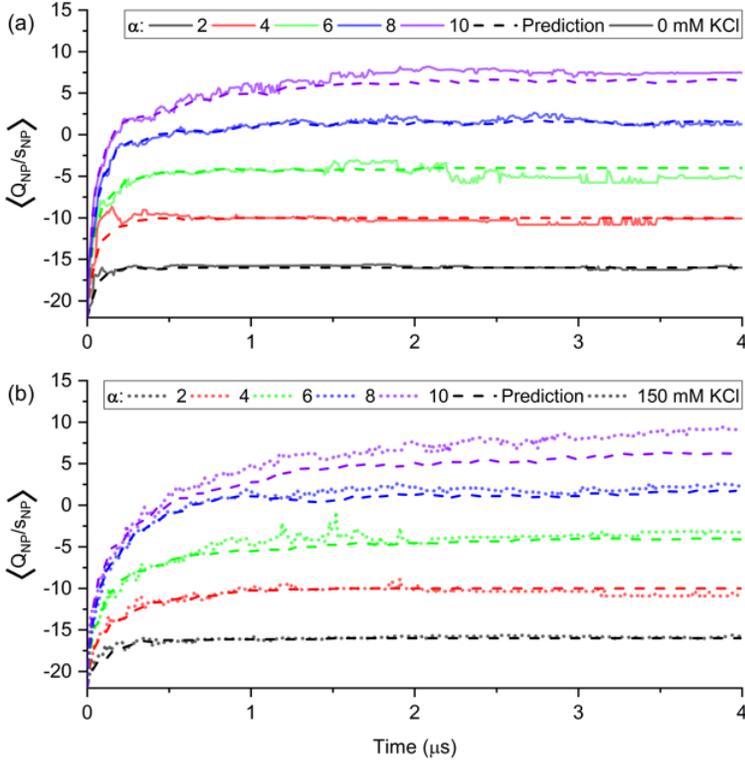


Figure B.3: $\langle Q_{NP}/S_{NP} \rangle$ as a function of time for systems with (a) 0 mM KCl, (b) 150 mM KCl. Q_{NP}^{pred}/S_{NP} is calculated using Eq. S2 and is added for comparison.

In **Figure B.3**, we plot $\langle Q_{NP}/S_{NP} \rangle$ as a function of time for different α , where the average is performed over all NPs. As explained in the main text, this quantity is a measure of the degree of neutralization of the DNAs in a NP. Data are presented for both 0 and 150 mM KCl (subfigure **a** and **b** respectively). The prediction Q_{NP}^{pred}/S_{NP} from Eq. S2 is added for comparison. The prediction is accurate for the entire simulation time and for both 0 and 150 mM KCl.

$$\begin{aligned}
Q_{NP}^{pred}(t; \alpha) &= s_{NP} \left(Q_{DNA} + \frac{Q_{PEI} (N_{PEI} - N_f(t; \alpha))}{N_{DNA}} \right) \\
&= s_{NP} (Q_{DNA} + Q_{PEI} \alpha')
\end{aligned}
\tag{B.2}$$

B.8. Aggregation in presence of 150 mM KCl

There are many similarities between the aggregation of PEIs and DNAs in the presence of 0 mM KCl and 150 mM KCl. The key similarities and differences have been reported in the main text (“Influence of salt concentration”) using $\alpha = 10$. Here we discuss the rest of the results. Similar to systems with 0 mM KCl, the difference in the mode of aggregation is observed in systems with 150 mM KCl. However, the transition from end-to-end aggregation to lateral side-by-side aggregation occurs between $\alpha = 4$ and 6. Specifically, in **Figure B.4**, we find limited aggregation for $\alpha = 2$, end-to-end aggregation for $\alpha = 4$ and side-by-side aggregation for $\alpha = 6-10$.

In **Figure B.5**, we plot N_f , N_p , N_b , and n_b as functions of simulation time for both 0 and 150 mM KCl. For $\alpha = 2, 4$, and 8, N_f , N_b , N_p , and n_b at 150 mM KCl shows similar characteristics to 0 mM KCl as discussed in the main texts, whereas some differences are observed for $\alpha = 6$. Specifically, we observe N_p to be lower (**Figure B.5b**) and N_b (**Figure B.5c**) to be higher at higher salt concentration during the entire simulation. Similar to 0 mM KCl n_b for the five systems form three groups in **Figure B.5d** corresponding to lowest, intermediate and highest n_b values respectively. However, while at 0 mM KCl $\alpha = 6$ belongs to the group of intermediate n_b , with the presence of 150 mM KCl $\alpha = 6$ belongs to the group with the highest n_b . All these different characteristics for $\alpha = 6$ are associated with the different mode of aggregation observed at 150 mM KCl. At 0 mM KCl N_b increases monotonically with α but at 150 mM KCl N_b for $\alpha = 6$ is greater than $\alpha = 8$. This can be explained by similar n_b values for $\alpha = 6$ and 8, and higher number of DNAs being aggregated for $\alpha = 6$ (to be discussed next).

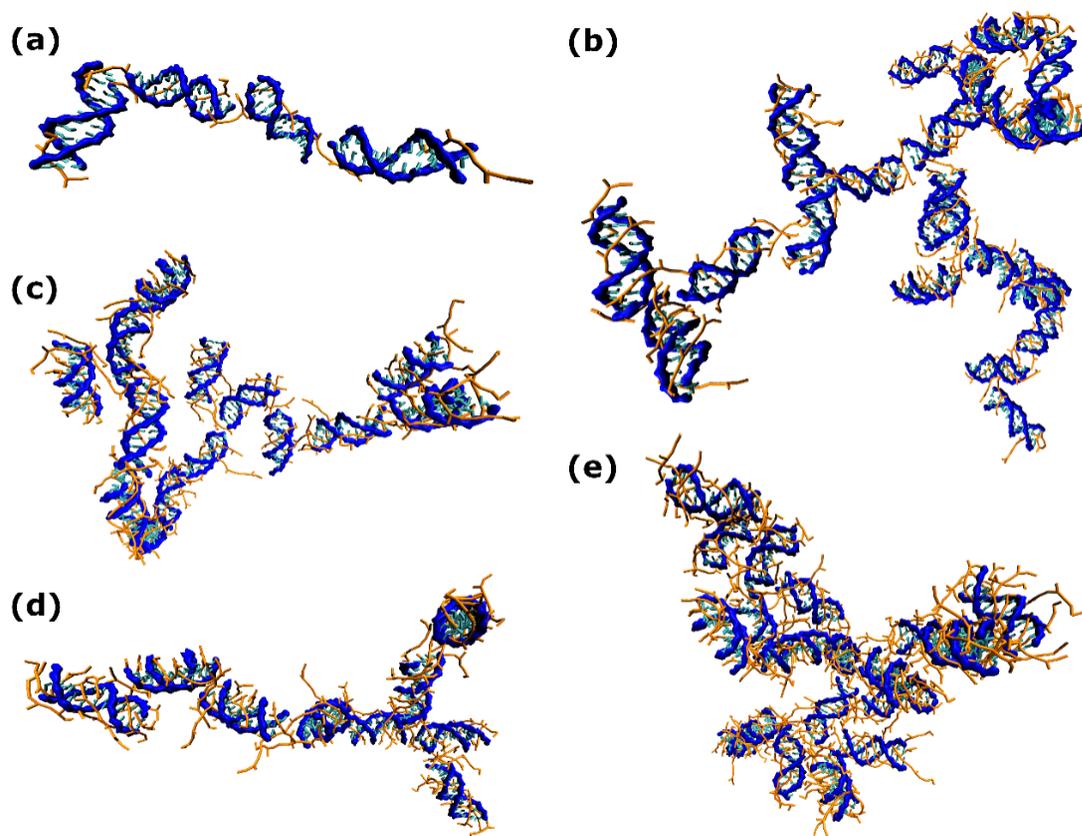


Figure B.4: Largest NP formed at the end of 4 μ s simulation for systems with 150 mM KCl. (a) $\alpha = 2$, $s_{NP} = 4$; (b) $\alpha = 4$, $s_{NP} = 24$; (c) $\alpha = 6$, $s_{NP} = 14$; (d) $\alpha = 8$, $s_{NP} = 9$ and (e) $\alpha = 10$, $s_{NP} = 19$. DNAs are represented with blue backbone and cyan base pairs, and PEIs in orange.

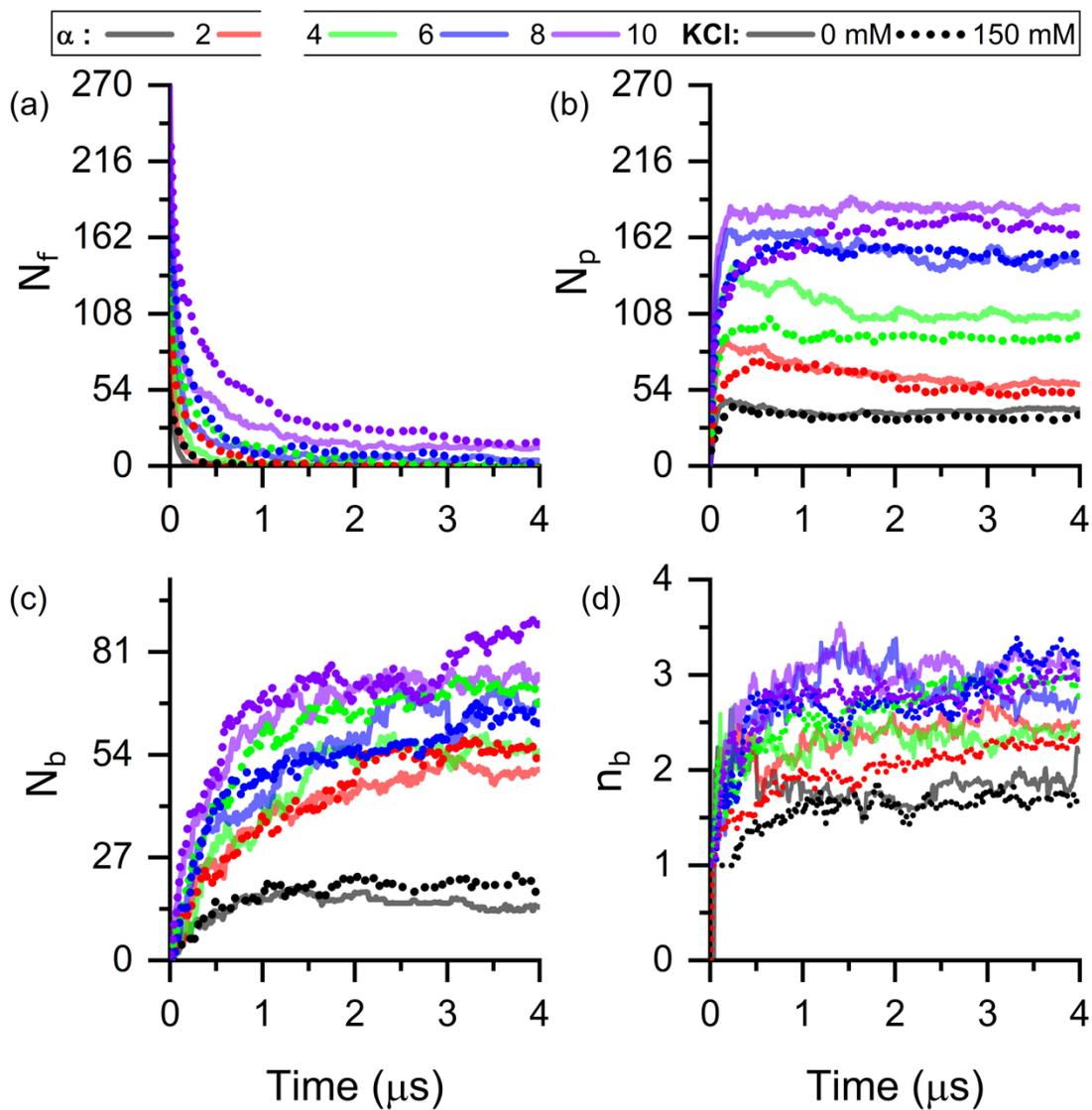


Figure B.5: Number of PEI in different roles, plotted against simulation time: (a) number of free PEIs (N_f) (b) number of peripheral PEIs (N_p) and (c) number of bridging PEIs (N_b). (d) Average number of bridging PEI between a pair of bridged DNAs (n_b).

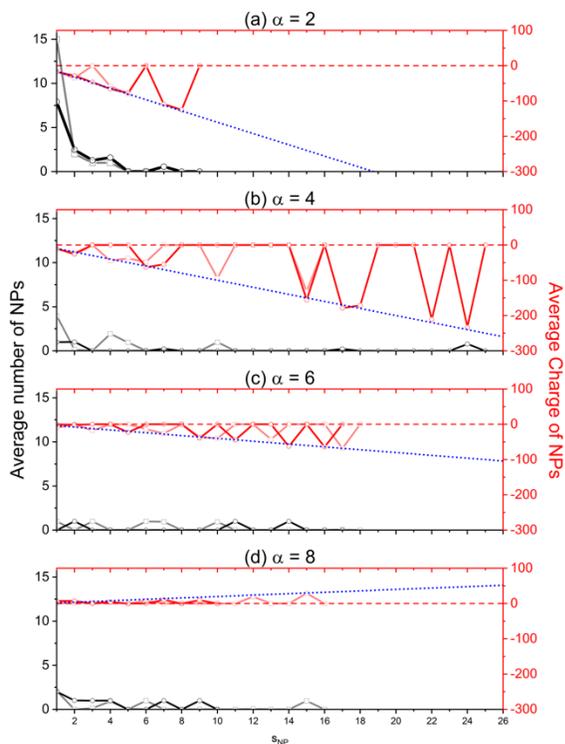


Figure B.6: NP size (black) and charge (red) distributions for $\alpha = 2-8$. The predicted charge (blue) is calculated using Eq. 1 (main text).

Figure B.6 plots the NP size distribution (black) and average NP charge as a function of its size. Data shown are obtained from the average over 3.5 to 4 μs of the simulations for both 0 and 150 mM KCl. **Figure B.7a** shows the mean NP size $\langle s_{NP} \rangle$ as a function of time, which for 150 mM KCl is fitted with continuous piecewise linear functions in **Figure B.7b**. **Figure B.7c** shows the steady-state NP size $\langle s_{NP} \rangle$ for the systems with 150 mM KCl, again obtained by average over 3.5 to 4 μs of the simulations. The slopes of the piecewise linear fit in **Figure B.7b** were plotted in **Figure B.7d** as a measure of the rate of NP growth (r_{NP}). These plots show that the characteristics of aggregation are similar for 0 and 150 mM KCl. Besides the quantitative difference discussed in the main texts (“Influence of salt concentration”), the biggest difference observed under the two salt concentrations is for $\alpha = 8$. Firstly, the largest NP formed for $\alpha = 8$ is smaller at 150 mM KCl than at 0 mM KCl. Secondly, while under 0 mM KCl the steady-state

$\langle s_{NP} \rangle$ is similar for $\alpha = 6$ and 8, with 150 mM KCl the value is noticeably smaller for $\alpha = 8$. Finally, for $\alpha = 8$ under 0 mM KCl, r_{NP} was observed to first decrease from the 0-2 μ s time window to zero during the 2-2.7 μ s time window, and then increase during the 2.7-3.5 μ s time window. Such an increase is not seen under 150 mM KCl. Because $\alpha = 8$ corresponds to weak specific repulsion, the screening from the ions may not be strong enough to enhance aggregation. However, it is unclear why for this α the aggregation appears to be weakened by the addition of salt. It is possible that due to the small specific repulsion and reduced probability of NP collision, the NPs may take longer to further grow, but this would need additional investigations.

Increase in NP size is observed for all α except $\alpha = 8$. The increase in NP size is largest for $\alpha = 4$ which can be explained as follows. The presence of salt is expected to promote NP growth by screening the NP-NP repulsions. Since the specific repulsions are weak for $\alpha = 6$ and 8, a significant increase in NP growth can be expected for $\alpha = 2, 4$ and 10. However, for $\alpha = 2$ NP growth is hindered due to the lack of PEIs in the system. For $\alpha = 10$, NP growth is also hindered as the peripheral PEIs accumulate and surround the DNAs, making the phosphate beads inaccessible for local NP-NP attraction. The linear nature of the NPs at $\alpha = 4$ further promotes NP growth as discussed in the main text, and hence the increase in NP size is most significant for $\alpha = 4$.

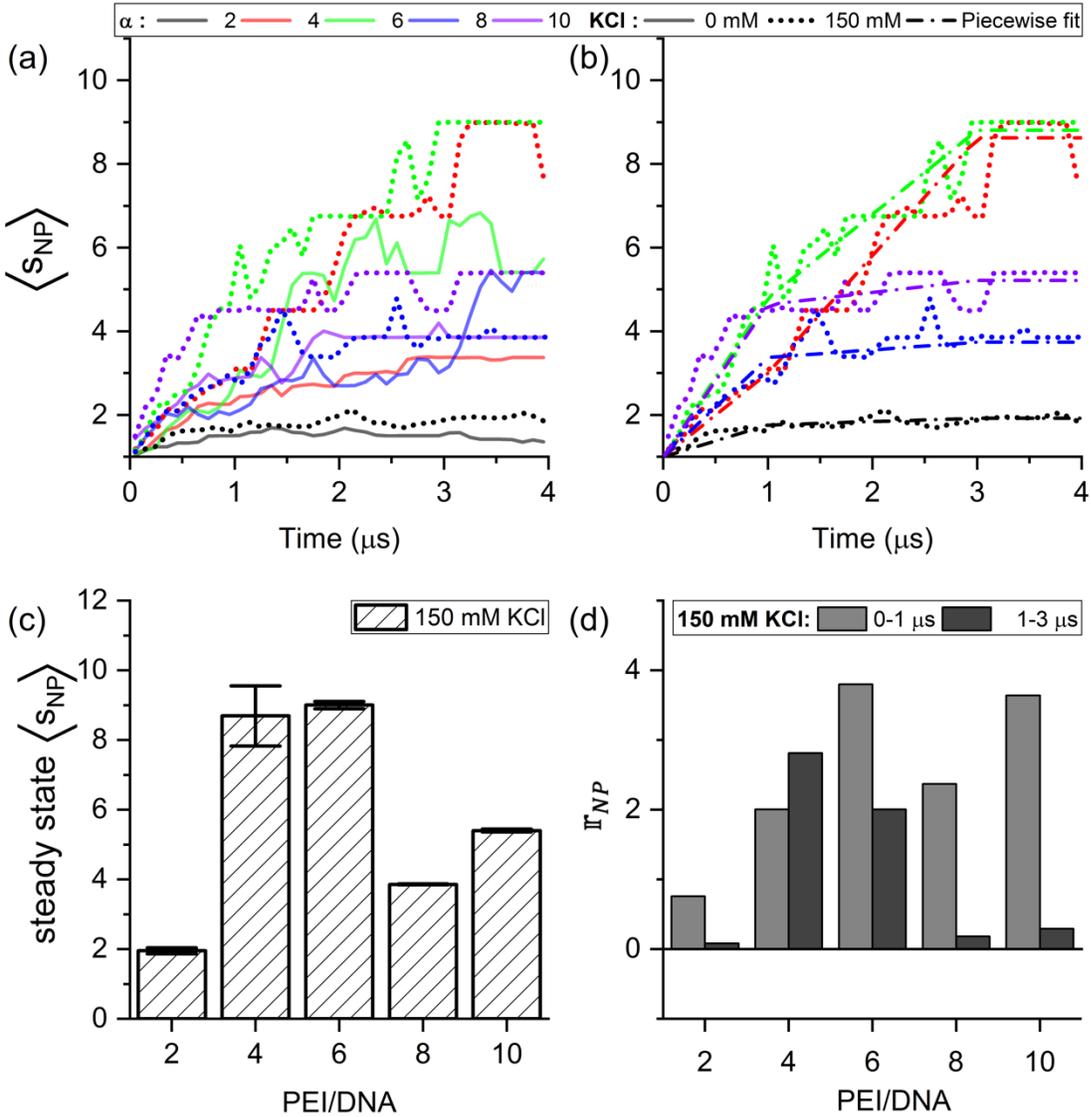


Figure B.7: (a) Comparison of $\langle s_{NP} \rangle$ as a function of time between 0 and 150 mM KCl (b) Continuous piecewise linear function fit of $\langle s_{NP} \rangle$ as a function of time for 150 mM KCl (c) Steady-state $\langle s_{NP} \rangle$ for 150 mM KCl. (d) rate of NP growth for 150 mM KCl.

B.9. Zeta potential

In experiments, charge of NPs is measured using their mean zeta potential. In **Figure B.8**, we calculate the average NP charge at steady-state for different N/P ratios to make comparisons with experimental mean zeta potential. The average NP charge is calculated for 0 and 150 mM KCl using two approaches: (1) by considering only bound PEIs and (2) also by considering both

bound PEIs and ions. From linear interpolation, the location of zero average NP charge is found to be at $N/P \sim 4.4$ for all the curves in **Figure B.8**. The shape of the curves also resembles the shape of mean zeta potential vs. N/P ratio reported in most experiments (see reference 17, 21, 23, 64 in the main article).

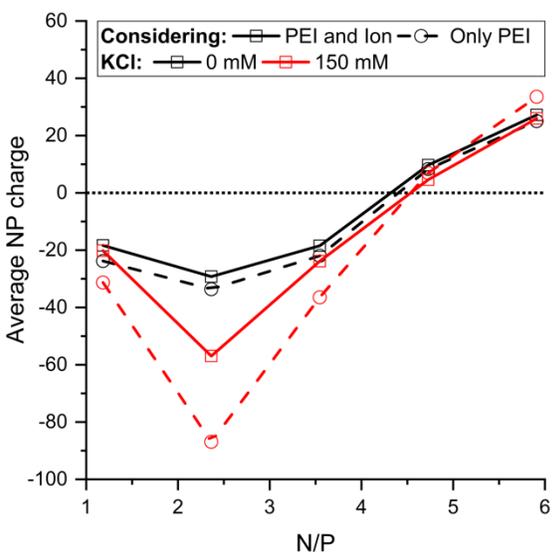


Figure B.8: Average NP charge calculated as a function of N/P ratio.

B.10. Distribution of hydrodynamic radius

The hydrodynamic radius (R_{hyd}) of a specific NP is calculated using the Kirkwood[2] definition shown in Eq. B.3, where r_{ij} is the distance between two beads i and j in the NP, and N_{beads} is the total number of beads in the NP.

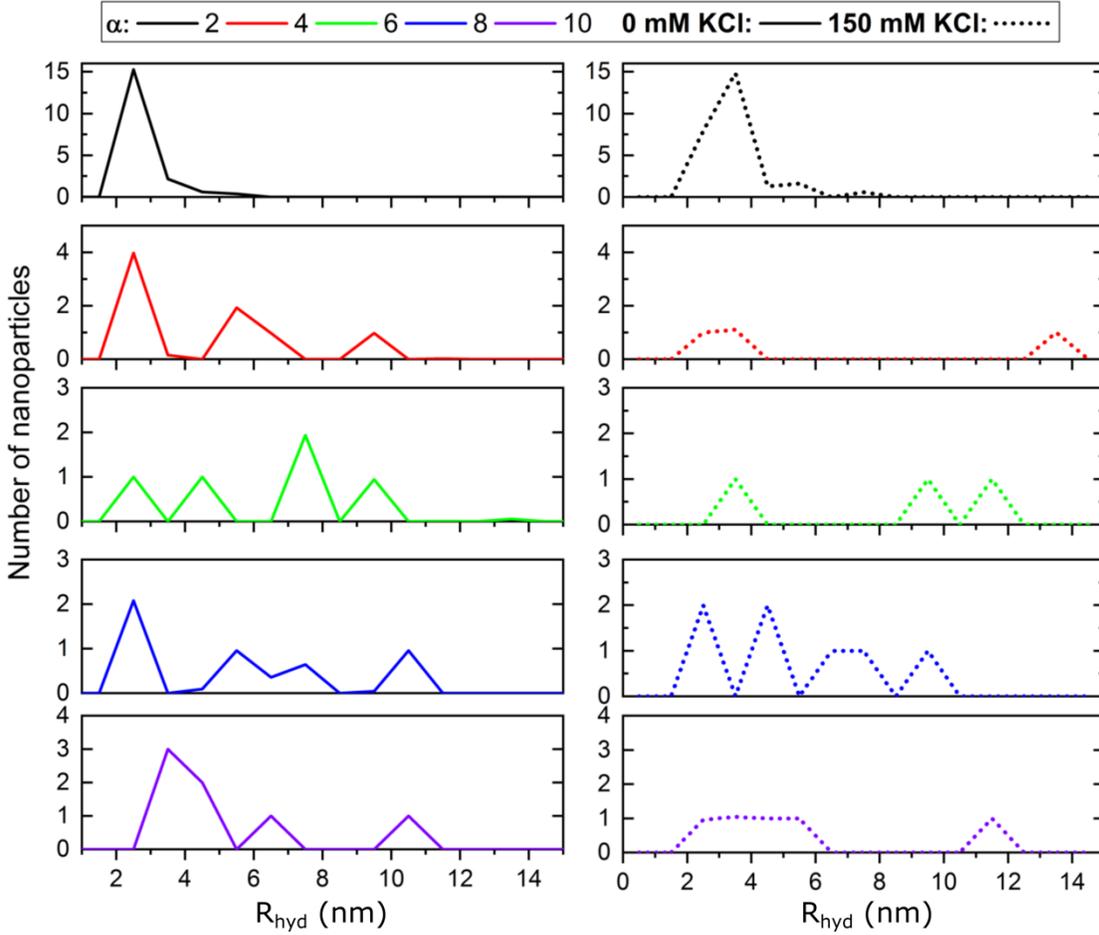


Figure B.9: Distribution of hydrodynamic radius (R_{hyd}) for 0 and 150 mM KCl at different α .

This definition is used instead of the Stokes radius ($k_B T / 6\pi\eta D$) because the latter would be subjected to errors associated with the evaluation of diffusion coefficient D of macromolecules and viscosity η . Eq. B.3 is usually used for polymers where the beads are connected via covalent bonds. In our case, each NP is composed of multiple molecules that can potentially leave the NP and join another one. In the steady-state of the simulations, most NPs are stable and the DNAs and PEIs within do not change with time. For each of these NPs, $\sum_{i<j} 1/r_{ij}$ is calculated at each time and a time average is performed to evaluate R_{hyd} using Eq. B.3, where N_{beads} is number of beads in the NP. A few NPs have lost or gained molecules during the steady-state stage of the simulations. For each of these NPs, the average is performed only over the time in which they are present. The

collected R_{hyd} data are used to generate a histogram (**Figure B.9**). If a NP is only present for a fraction of the time, its histogram value is scaled by the same fraction. The first peak in each distribution corresponds to unaggregated DNAs. Neglecting this peak, the R_{hyd} distribution of the NPs has a similar shape as the size distribution shown in **Figure 5.5** (black curve), i.e., unimodal for $\alpha = 2$, bimodal for $\alpha = 4$ and 10, and multi-modal (closer to uniform) for $\alpha = 6$ and 8.

$$\frac{1}{R_{hyd}} = \frac{1}{N_{beads}^2} \left\langle \sum_{i < j} \frac{1}{r_{ij}} \right\rangle \quad \text{B.3}$$

B.11. Distribution of hydrodynamic radius

Figure B.10 shows the largest NP formed at $\alpha = 10$ through a two-step PEI addition process. A branched NP is formed which has smaller size (less number of DNAs) than the largest NP formed at $\alpha = 10$ using a one-step PEI addition. The shape of these NPs is quantified using the relative shape anisotropy (κ^2 as shown in Eq. S1). The κ^2 averaged over last 0.5 μs of the simulation is found to be 0.32 ± 0.02 and 0.24 ± 0.02 for the NP formed using one- and two-step process respectively. This shows the largest NP obtained using a two-step process is also more compact and closer to being spherical than the one obtained using a one-step process.

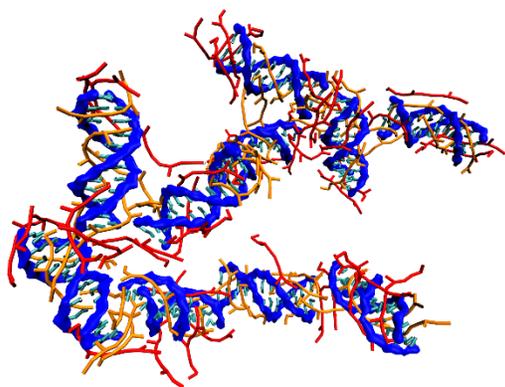


Figure B.10: Largest NP formed after 3.5 μs for $\alpha = 10$ achieved by adding 162 PEIs to $\alpha = 4$.

References

- [1] Theodorou D.N., Suter U.W. Shape of Unperturbed Linear Polymers: Polypropylene. *Macromolecules* **1985**, 18, 1206–14.
- [2] Cloizeaux J. Des, Jannink G. Standard Continuous Model and Perturbation Calculations. *In* Polymers in Solution Their Modelling and Structure. *Oxford University Press* **1990**, 354–430.

Chapter C

Supporting Information for Chapter 5*

C.1. Additional results for physiological and endosomal pH simulations

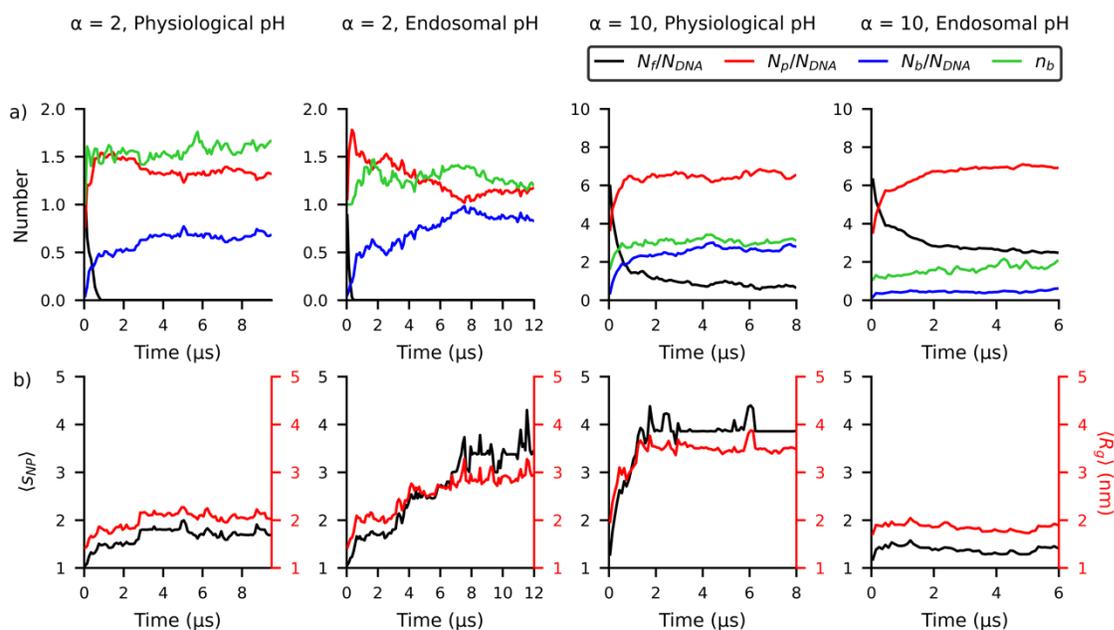


Figure C.1: a) Number of free (N_f/N_{DNA}), peripheral (N_p/N_{DNA}), and bridging PEI (N_b/N_{DNA}) normalized by the total number of DNAs, and average number of bridging PEIs between a pair of bridged DNAs (n_b). b) NP size ($\langle S_{NP} \rangle$; left axis) and radius of gyration ($\langle R_g \rangle$; right axis) averaged across all NPs. The figure titles provided are for (a) and (b), and the legend is for (a). The maximum y-axis in (a) is equal to the α of the system.

C.2. Kinetics of bridging PEIs for $\alpha = 2$

In **Figure C.2**, the general trend shows N_f/N_{DNA} is zero, N_p/N_{DNA} decreases, N_b/N_{DNA} increases, and n_b remains steady. The main driving force for this change is the conversion of

peripheral to bridging PEIs due to the reduced energy barrier at $q = 3$ after acidification (**Figure 5.2b**).

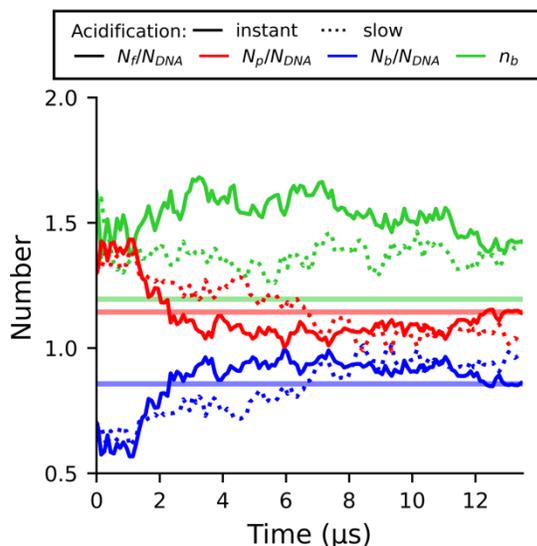


Figure C.2: Number of free (N_f/N_{DNA}), peripheral (N_p/N_{DNA}), and bridging PEI (N_b/N_{DNA}) normalized by the total number of DNAs, and average number of bridging PEI between a pair of bridged DNAs (n_b). The horizontal lines represent the steady-state value obtained from the endosomal pH simulation at $\alpha = 2$.

C.3. Transition between PEIs

The transition of free (f), peripheral (p), and bridging (b) PEIs are calculated between consecutive timesteps (0.2 ns apart). The total number transitions are normalized by the corresponding simulation time t_{sim} in microseconds and by α to account for the different number of PEIs in the systems (**Table C.1**). While **Table C.1** presents data for both $\alpha = 2$ and 10, discussions below are focused on the results for $\alpha = 10$ to support **Chapter 5**.

Table C.1 shows that during the acidification the normalized transitions between free and bridging PEI is very low. This indicates that bridging PEIs convert to peripheral ones and peripheral PEIs convert to free ones, while maintaining a dynamic balance of N_p/N_{DNA} .

The reversibility of PEI transitions can be quantified by the ratio between normalized transitions in one direction and the net normalized transitions. For bridging to peripheral PEI transitions, the ratio $\frac{p \leftarrow b}{p \leftarrow b - p \rightarrow b}$ is ~ 376 and ~ 326 for slow and instant acidification respectively. Similarly, for peripheral to free PEI transitions, the ratio is ~ 53 and ~ 43 for slow and instant acidification respectively. This reaffirms the observations made from **Figure 5.6**, that slow acidification is more reversible than instant acidification.

Table C.1: Number of transitions between PEIs normalized by the total simulation time (t_{sim}) in microseconds and by the PEI/DNA number ratio (α) at which the NPs were prepared. f , p , and b denote free, peripheral, and bridging PEIs respectively, and the arrows denote the direction of transition.

α	System	t_{sim} (μ s)	$f \rightarrow p$	$f \leftarrow p$	$p \rightarrow b$	$p \leftarrow b$	$f \rightarrow b$	$f \leftarrow b$
2	Physiological	9.5	18.05	15.58	213.16	213.32	0.11	0
2	Endosomal	12	8.42	6.50	415.62	414.71	0	0
2	Slow acidification	13.5	0	0	657.41	657.15	0	0
2	Instant acidification	13.5	0	0	605.30	605.11	0	0
10	Physiological	8	157.74	154.84	925.00	923.96	1.74	1.85
10	Endosomal	6	71.30	68.57	168.13	167.90	0.10	0.08
10	Slow acidification	7	40.36	41.14	281.44	282.19	0.04	0.06
10	Instant acidification	7	33.76	34.56	250.57	251.34	0.03	0.04

The transition of PEI states calculated from the first and last timesteps is shown in **Table C.2**. This can be used to calculate the probability that a PEI changes its state. For slow acidification, there are 18 $b \rightarrow f$, 40 $b \rightarrow p$ and 18 $b \rightarrow b$ transitions. Among the initially bridging PEIs that

have changed their role, the percentage of conversion to peripheral PEIs is $\frac{b \rightarrow p}{(b \rightarrow p) + (b \rightarrow f)} = 69\%$.

The corresponding percentage for instant acidification is 67.8%, and the average of the two is 68.4%. This number is given in **Figure 5.5b** (Step 2b) of the main text. Consequently, $100\% - 68.4\% = 31.6\%$ is the average probability of bridging PEIs converting to free ones (Step 2a; **Figure 5.5b**).

Table C.2: Total number of transitions between PEIs by only considering the first and last timestep. α is PEI/DNA number ratio, f , p , and b denote free, peripheral and bridging PEIs respectively, and the arrows denote the direction of transition.

α	System	$f \rightarrow f$	$f \rightarrow p$	$f \rightarrow b$	$p \rightarrow f$	$p \rightarrow p$	$p \rightarrow b$	$b \rightarrow f$	$b \rightarrow p$	$b \rightarrow b$
2	Physiological	0	31	18	0	5	0	0	0	0
2	Endosomal	0	27	19	0	5	3	0	0	0
2	Slow acidification	0	0	0	0	24	12	0	5	13
2	Instant acidification	0	0	0	0	26	10	0	5	13
10	Physiological	16	162	63	2	13	12	0	1	1
10	Endosomal	67	153	14	2	31	2	0	1	0
10	Slow acidification	10	8	0	47	123	6	18	40	18
10	Instant acidification	13	5	0	43	129	4	19	40	17

C.4. Transition diagram

The transition diagrams of the entire acidification process are shown in **Figure C.3** (left panel) for $\alpha = 2$, and in **Figure C.4** (left panel) for $\alpha = 10$. In these diagrams, detailed internal

structures of NPs (as displayed in **Figure 5.6**) are replaced by large circles which are colored blue-red depending on their last occurrence time in the simulation. Small circles denoting aggregation/dissociation are colored light-grey, and the ones denoting multiple NPs exchanging DNAs are colored dark-grey. Arrows connecting the circles (large or small) represent transitions and its thickness is given by $1 + \log((N_{transition} + 1)/2)$, where $N_{transition}$ is the number of transitions in one direction. As explained in the main text, loops corresponding to cyclic transitions are removed to retain only principal or “net” transitions in the principal transition diagram. In **Figure C.3** and **Figure C.4**, large circles and transitions are colored green if they are present in the principal transition diagram. The relative position of circles has no physical significance. For clarity, the transitions that are absent in the principal transition diagram are removed in the right panel of **Figure C.3** and **Figure C.4**. The following discussion of the transition diagram is based on their properties summarized in **Table C.3**.

The transition diagrams during instant and slow acidifications for $\alpha = 2$ are shown in **Figure C.3a, b** respectively. The number of NP structures in the transition diagram ($N_{structure}$) are similar for both acidification rates (**Table C.3**). The number of small circles denoting aggregation/dissociation ($N_{a/d}$) is comparable to $N_{structure}$ for both acidification rates, indicating a large fraction of the NPs undergo aggregation/dissociation. By counting the $N_{a/d}$ that last occurred before ($N_{a/d,early}$) and after ($N_{a/d,late}$) 50% of the total simulation time, we find the reversibility arises later in the simulation. This likely reflects the instability of large NPs due to low number of PEIs. $N_{a/d}$ in instant acidification is higher than that in slow acidification, which indicates NP aggregation/dissociation in instant acidification are more reversible.

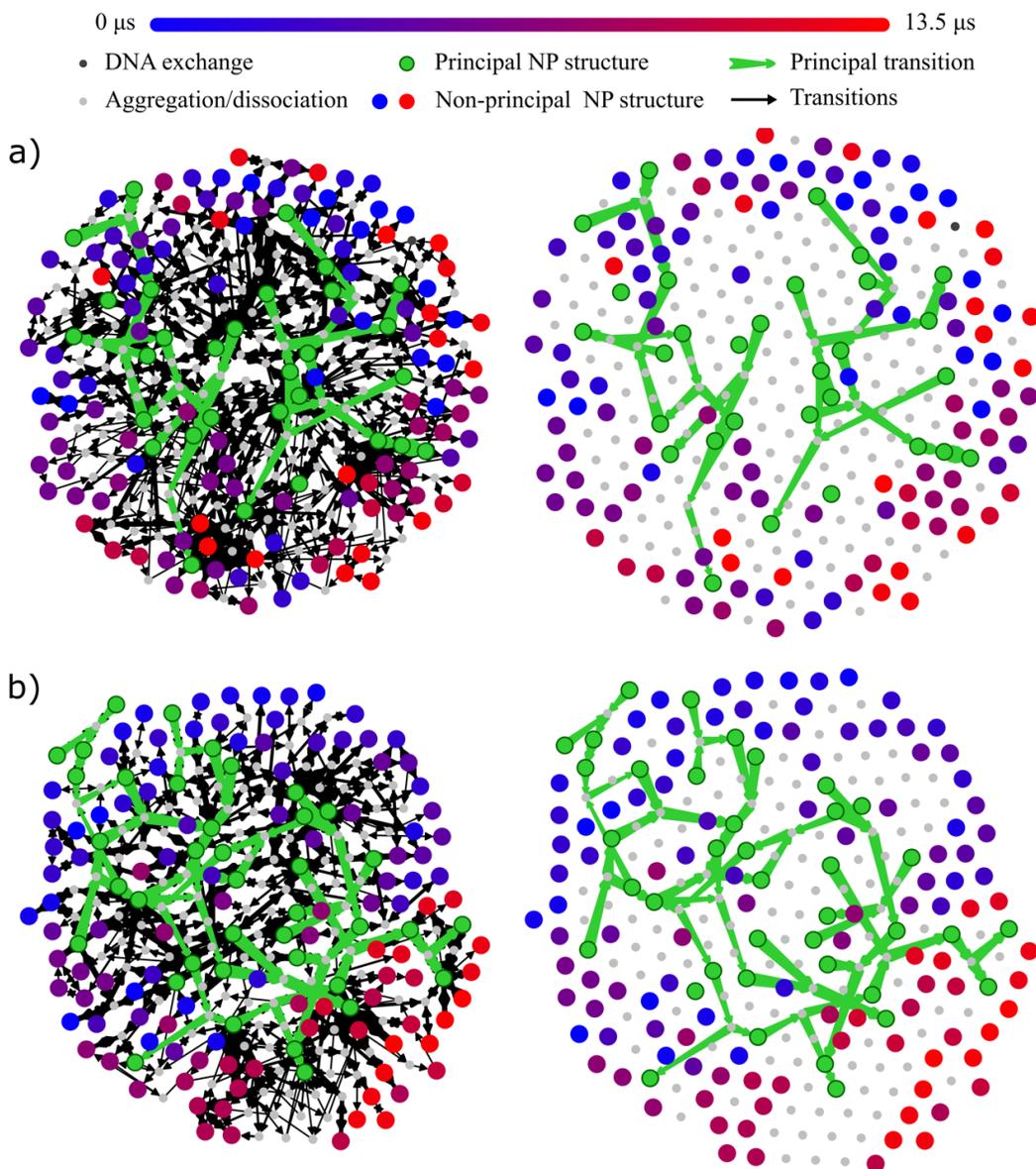


Figure C.3: Transition diagram (left panel) for (a) instant acidification, and (b) slow acidification of NPs prepared with $\alpha = 2$. The arrows between the nodes represent NP transition and its thickness is given by $1 + \log((N_{transitions} + 1)/2)$, where $N_{transition}$ is the number of transitions in one direction. The non-principal transitions are removed in the right panel.

The transition diagrams during instant and slow acidifications for $\alpha = 10$ are shown in **Figure C.4a, b** respectively. $N_{structure}$ in slow acidification is higher than that in instant acidification, whereas $N_{a/d}$ and N_{ex} (number of small circles denoting multiple NPs exchanging

DNAs) are similar. This demonstrates that, unlike $\alpha = 2$ where reversibility is mainly caused by aggregation/dissociation, the higher reversibility during slow acidification for $\alpha = 10$ is due to internal restructuring. Further investigating the $N_{structure}$ that last occurred before ($N_{s,early}$) and after ($N_{s,late}$) 50% of the total simulation time reveals that the reversibility is mainly observed at the beginning of the simulation.

Table C.3: Properties of the transition diagram: number of NP structures ($N_{structure}$), $N_{structure}$ that last occurred before ($N_{s,early}$) and after ($N_{s,late}$) 50% of the total simulation time, number of small circles denoting aggregation/dissociation ($N_{a/d}$), $N_{a/d}$ that last occurred before ($N_{a/d,early}$) and after ($N_{a/d,late}$) 50% of the total simulation time, and number of small circles denoting multiple NPs exchanging DNAs (N_{ex}). The numbers within the parenthesis are the corresponding values obtained from the principal transition diagram in the right panel.

Transition Diagram	$N_{structure}$	$N_{s,early}$	$N_{s,late}$	$N_{a/d}$	$N_{a/d,early}$	$N_{a/d,late}$	N_{ex}
$\alpha = 2$, Slow acidification	150 (39)	91 (22)	59 (17)	120 (18)	47 (6)	73 (12)	0 (0)
$\alpha = 2$, Instant acidification	151 (30)	84 (9)	67 (21)	143 (10)	55 (4)	88 (6)	1 (0)
$\alpha = 10$, Slow acidification	223 (42)	158 (20)	65 (22)	78 (11)	41 (5)	37 (6)	2 (0)
$\alpha = 10$, Instant acidification	131 (37)	70 (16)	61 (21)	75 (9)	37 (2)	38 (7)	1 (0)

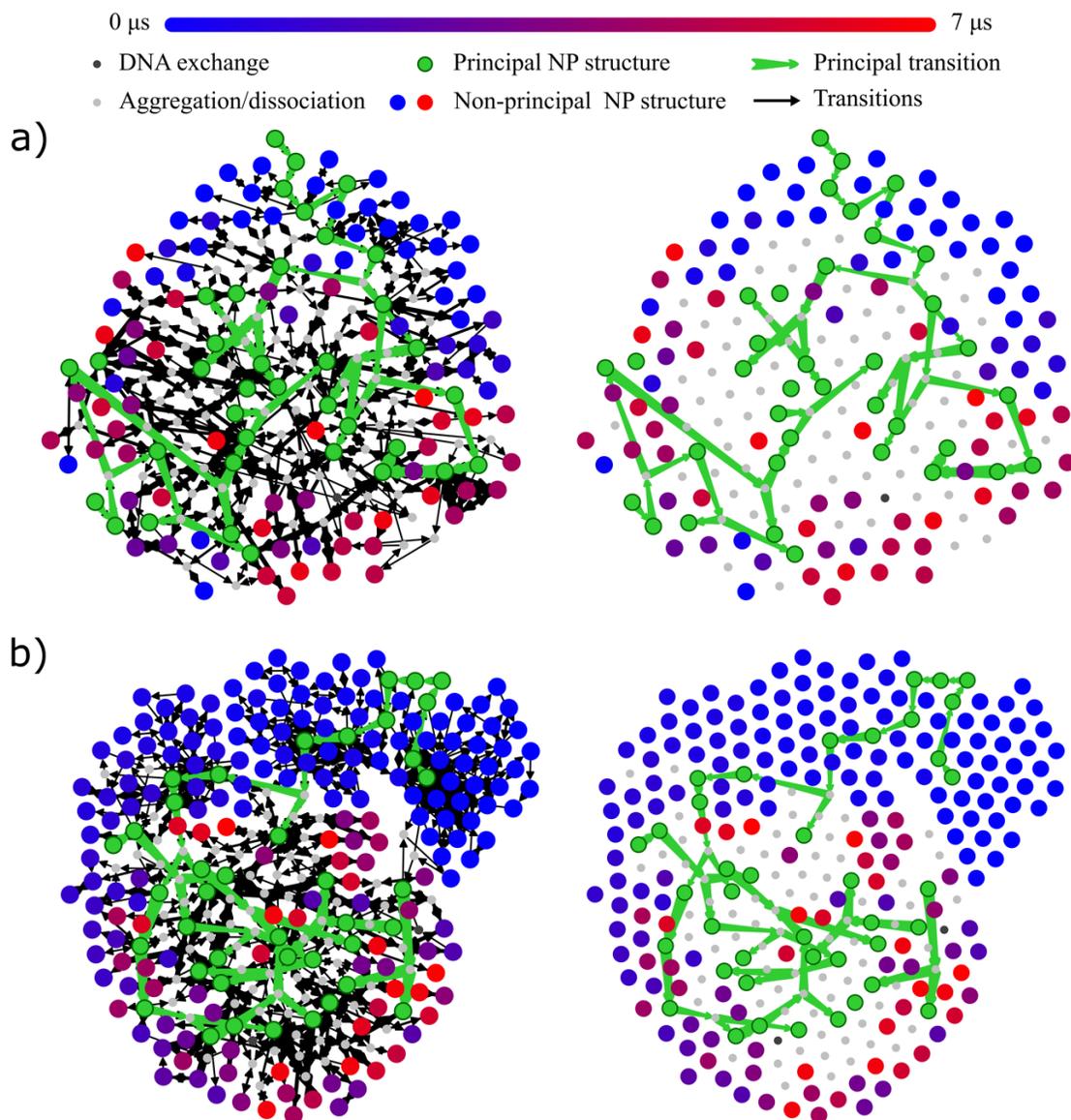


Figure C.4: Transition diagram (left panel) for (a) instant acidification, and (b) slow acidification of NPs prepared with $\alpha = 10$. The arrows between the nodes represent NP transition and its thickness is given by $1 + \log((N_{transitions} + 1)/2)$, where $N_{transition}$ is the number of transitions in one direction. The non-principal transitions are removed in the right panel.

C.5. Principal transition diagram

Acidification simulations for $\alpha = 10$ begins with seven NPs. The principal transition diagram for the two largest NPs were shown in the main text (**Figure 5.6**), and the diagrams for the remaining NPs, named NP3-NP7, are shown in **Figure C.5**.

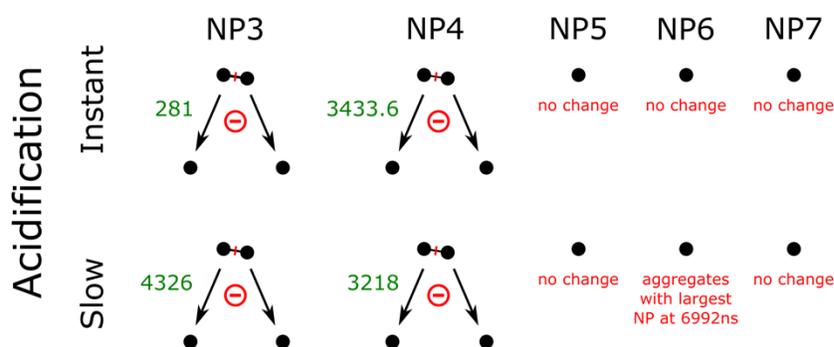


Figure C.5: Principal transition diagrams of NPs at $\alpha = 10$. The black circles denote DNAs, where the relative position of circles have no physical significance. The black lines represent bridging PEIs, and its thickness is proportional to number of PEIs bridging the DNA pair. The red line represents bridge scission, and the circled ‘-’ represents NP dissociation. The last transition time, in nanoseconds, is shown in green.

The principal transition diagrams for $\alpha = 2$ are shown in **Figure C.6**, where **Figure C.6a-d** pertains to instant acidification and **Figure C.6e-g** to slow acidification. NPs that do not undergo any change in the principal transition diagram are not shown. In **Figure C.6**, the transitions are similar for both acidification rates and all NP sizes. Most transitions are NP aggregations, which tend to occur near the terminal of linearly structured NPs. In **Figure C.6a**, e loops are formed in the principal transition diagram that cannot be removed. This is analogous to a catalytic chemical reaction, with the DNA marked by the red arrow acting as a catalyst to help NPs transition from one structure to another (3925 ns in **Figure C.6a**, 5381.4 ns in **Figure C.6e**).

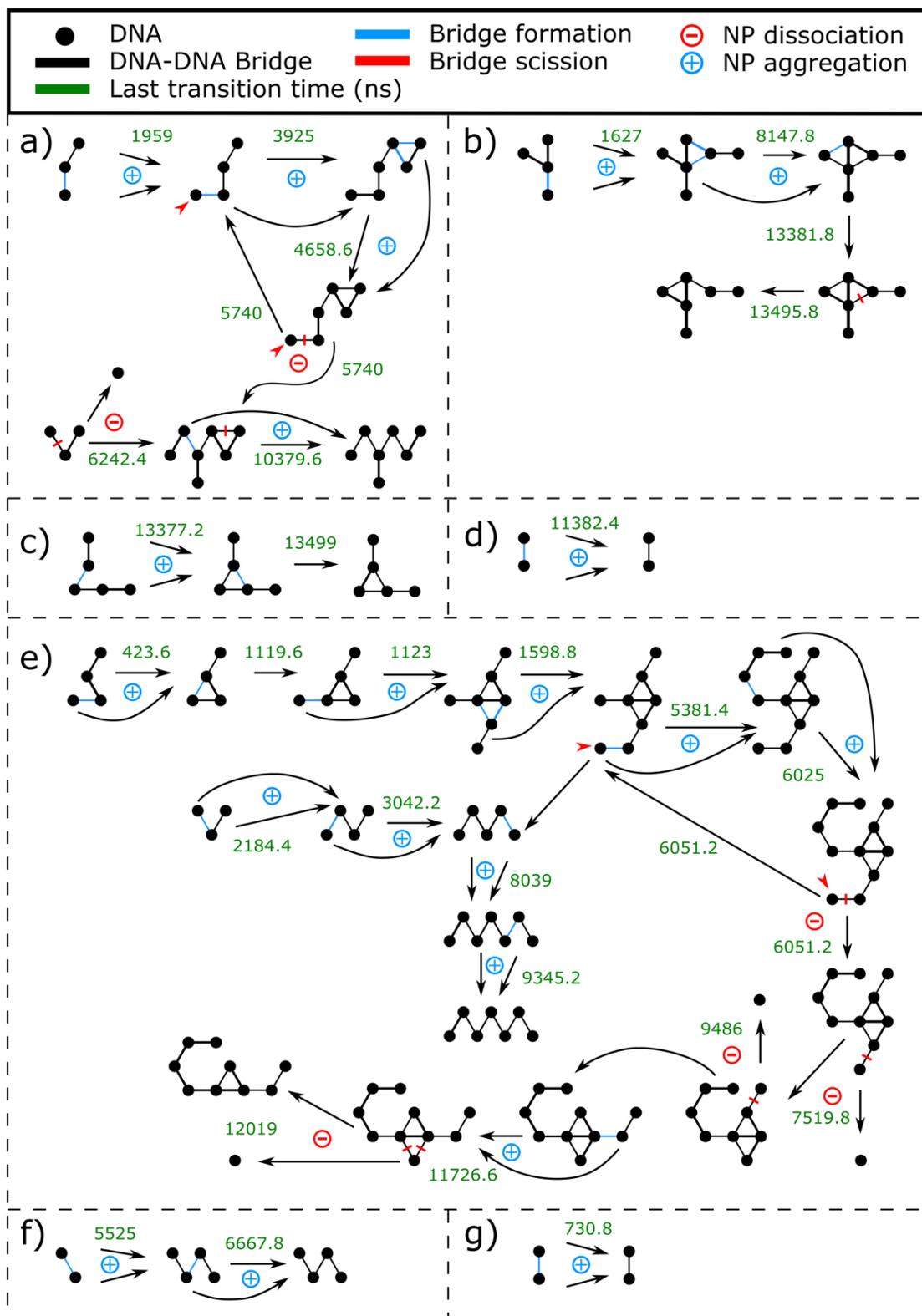


Figure C.6: Principal transition diagram for $\alpha = 2$ during (a-d) instant and (e-g) slow acidifications.

Relative position of circles (DNAs) has no physical significance.

C.6. Relation between reaction coordinate q and PEI state

The probability of the reaction coordinate q is plotted in **Figure C.7** for acidification simulations with $\alpha = 2$ and 10, and for different PEI states (free, peripheral, and n -bridging). For $\alpha = 2$, the curves are absent for free PEIs at both acidification rates, and 4-bridging PEIs at slow acidification due to the absence of such PEIs in the simulation. For $\alpha = 10$, the curves for free PEIs are identical for both acidification rates. Dashed vertical lines represent $q = 1 + 1.2(k - 1)$, where k is a positive integer. In the main text, these q values are used to differentiate PEI states in the free energy diagram (**Figure 5.2b** and **c**). **Figure C.7** shows limited overlap of curves corresponding to different PEI states near the vertical lines, i.e., these q values are appropriate to differentiate PEIs.

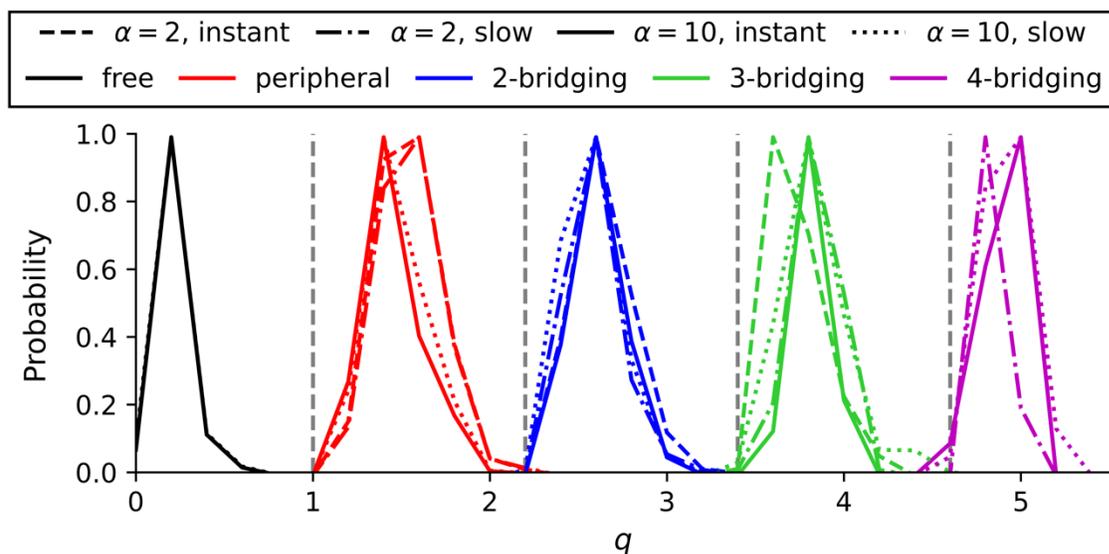


Figure C.7: Probability of the reaction coordinate q for different acidification simulations and PEI state.

Chapter D

Supporting Information for Chapter 6*

* A version of this chapter's sections has been published in a pre-print. Adapted from Subhamoy Mahajan, and Tian Tang, Meeting Experiments at the Diffraction Barrier: An *In-silico* Widefield Fluorescence Microscopy, *bioRxiv* **2021**, DOI: 10.1101/2021.03.02.433395. Copyright © 2021 Subhamoy Mahajan and Tian Tang, CC-BY-NC-ND 4.0 International license.

D.1. Methods

D.1.1. Generating *in-silico* microscopy images and videos

The *in-silico* monochrome images were rendered using *matplotlib*[1] *imshow* with a grey colormap. A 2D cross-sectional view depicting the use of PBC and white image frame is shown in **Figure D.1**. The number of periodic images of fluorophores that contributes to $I(\ell', m')$ depends on the dimensions $(P_{\ell'}, P_{m'}, P_{n'})$ specified for the predetermination of PSF. However, the range of (ℓ', m') coordinates correspond to the original MS specimen (center box in \mathcal{MS} , **Figure 6.2**). For example, if an MS specimen is a cube with side length of 100 nm and $P_{\ell'}, P_{m'}, P_{n'} = 300$ nm, I will be calculated for the image coordinates $0 \leq \ell', m' \leq 100$ nm, while particles (and their periodic images) located at $\ell \in [\ell' - 150, \ell' + 150]$, $m \in [m' - 150, m' + 150]$ and $n \in [n' - 150, n' + 150]$ can all contribute to I at (ℓ', m') . In each direction ℓ or m , the dimension of the white image frame is greater than or equal to the largest MS specimen during the entire trajectory. For example, if an MS simulation produces two MS specimen with dimensions of (100, 200, 300) and (200, 100, 300) nm in the ℓmn directions, the white image frame is no smaller than 200×200 nm².

For generating colored microscopy images, all mixed HSV colors were converted to RGB colors based on Smith[2] before rendering each *in-silico* microscopy image. The final *in-silico* microscopy images were produced using *imshow* in *matplotlib*[1] v3.1.3. Videos were created in .mov format with ‘mp4v’ codec using VideoWriter class from OpenCV-python v.3.4.4 (<https://libraries.io/pypi/opencv-python/3.4.4.19>).

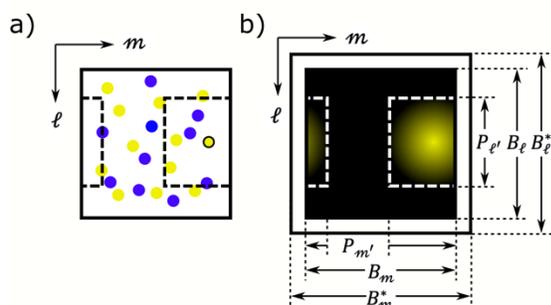


Figure D.1: Handling of periodic boundary condition and white frame. a) MS box when seen along the optical axis n . The ℓm axes are shown for reference. The yellow and indigo circles represent particles of two different fluorophore types. The PSF for the yellow circle with black outline is calculated over a 3D cuboidal box centered around it with dimension $(P_{\ell'}, P_{m'}, P_{n'})$. The 2D cross-sectional view of the box with dimension $(P_{\ell'}, P_{m'})$ is shown by black dashed lines, which is split into two parts due to PBC. b) The image is generated from the fluorescence of the yellow particle with black outline shown in (a). The largest MS box in the trajectory is represented by the white frame with dimensions B_{ℓ}^* and B_m^* . The MS box for the current time is represented by the black background with dimensions B_{ℓ} and B_m and placed at the center of the white image frame.

D.1.2. Molecular simulations

The PEI-DNA aggregation simulation used in this work was a Martini coarse-grained molecular dynamics simulation[3] performed using the Gromacs 5 package.[4] The system contained 27 DNAs, 270 PEIs and 150 mM KCl. In the initial configuration, the DNAs were arranged in a 3x3x3 cubic lattice inside a cubic box of side 25 nm, with 6 nm as the closest distance between the centers of mass of two DNAs. The PEIs and ions were placed randomly in the cubic

box. The configuration of the system after energy minimization and constrained simulation of 1 ns in NPT ensemble is shown in **Figure 6.1b**. Thereafter, an unconstrained NPT simulation was run for 4 μ s.

The last configuration of the PEI-DNA aggregation simulation was used as the initial configuration for the endosomal acidification simulation. To simulate acidification, every 0.4 ns the protonation ratio of a randomly selected PEI was changed from 23% to 46%,^[5] followed by adding an appropriate number of chloride ions to keep the system electroneutral. After all PEIs were acidified (in 108 ns), unconstrained NPT simulation was run for an additional 2.5 μ s.

The DMPC lipid bilayer simulation dataset produced by Miettinen^[6] (Copyright © 2013 Miettinen, Creative Commons Attribution 4.0 International License) was used to calculate mean-squared-displacement. The system consisted of 128 DMPC and 5097 SPC water molecules and was run for 110 ns in unconstrained NPT ensemble using the Gromacs 3^[7] package.

D.1.3. Analysis of *in-silico* microscopy images

To determine the number and cross-sectional area of particles from an *in-silico* image, a binary image was produced using a threshold intensity; any pixel with intensity above the threshold was considered to be part of a particle. If two pixels were vertically or horizontally adjacent (while considering PBC) to each other and both have intensity above the threshold, they were considered to be part of the same particle; pixels diagonal to each other were not considered to be adjacent. Pixels belonging to the same particle were grouped together. The number of such groups was defined as the number of particles, and the cross-sectional area of a particle was calculated from the product between the number of pixels in the particle and the area of a pixel ($\Delta l' \Delta m'$). For experimental images, the number of particles and cross-sectional area were calculated using

“Analyze Particles” feature in Fiji ImageJ.[8] The number of particles and cross-sectional area were then normalized by their corresponding maximum values.

Single-particle tracking was performed using TrackMate v6.0.2 in Fiji ImageJ.[8,9] Voxel dimensions were calibrated to the grid used for PSF calculation, i.e., $(\Delta\ell', \Delta m', \Delta n') = (0.1, 0.1, 0.2)$ nm. Time interval was set as 1 ns (based on the lipid simulation[6]). The entire image was used for analysis, i.e., crop settings was set to default. Particles were detected using Laplacian of Gaussian detector with an “estimated blob diameter” of 9.0 and threshold diameter of 6.0. All detected particles were selected for tracking using the “Simple LAP” tracker. Maximum linking distance, gap-closing distance, and gap-closing frame gap were set to 5.0, 5.0 and 2 respectively. Lateral MSD was calculated using $MSD(t) = \langle |\mathbf{r}_j(t + t_0) - \mathbf{r}_j(t_0)|^2 \rangle$, where \mathbf{r}_j was the 2D position vector of j^{th} particle, and the average $\langle \cdot \rangle$ was performed over different “tracks” and different reference time t_0 .

D.1.4. Analysis of MS data

The number of nanoparticles in a MS was calculated based on **Chapter 4**. [3] PEIs and DNAs were considered to be bound if their minimum distance was less than 0.528 nm. A PEI-DNA nanoparticle is a collection of all PEIs and DNAs that are directly or indirectly bound to each other. At each time, the number of nanoparticles was averaged over the 10 ns after this time. Hydrodynamic radius of a nanoparticle was calculated using $R_{hyd} = N_{beads}^2 / \langle \sum_{j < l} 1/r_{jl} \rangle$, where N_{beads} was the number of beads in a nanoparticle (including those from both PEIs and DNAs), r_{jl} was the distance between beads j and l , and $\langle \cdot \rangle$ represented ensemble average.[10] Cross-sectional area of a nanoparticles was approximated by R_{hyd}^2 averaged over 10 ns.

The fraction of PEI beads that had at least one DNA bead within a cutoff distance was determined using the *mindist* function in Gromacs. The total number of contacts between PEIs and

DNA was calculated using the *group* option where all DNAs were specified as the first group. That is, contacts between a PEI bead and multiple DNA beads were treated as one. Then, this number of contacts was divided by the total number of PEI beads to obtain the fraction of PEI beads in contact with at least one DNA bead. Similar approach was taken to determine the fraction of DNA beads in contact with at least one PEI bead.

D.2. Point spread function for *in-silico* microscopy

To use a point spread function (PSF) for *in-silico* microscopy, it should be a function of coordinates (ℓ', m', n') , where $\ell' m'$ are the scaled image coordinates numerically equal to object coordinates ℓm , and $n' = n_o - n_j$ is distance of an object from the object focal plane. The PSF proposed by Gibson and Lanni[11] (Eq D.1, D.2) describes the diffraction pattern produced by an object located along the optical axis as a function of the detector position (x_d, y_d, z_d) in image coordinates. In Eq D.1, the subscript GL represents Gibson and Lanni, C is a constant, $r = \sin \theta / \sin \psi = (\mu \sin \theta) / \text{NA}$ is the normalized radius in a medium with maximum half angle ψ and refractive index μ , NA is the numerical aperture, M is the magnification of the microscope (typically $\gg \text{NA}$), k is the wave vector in vacuum, and OPD is the difference in the optical paths followed by the light in non-design and design conditions.

The OPD is given by Eq D.2, where μ_a and t_a are respectively the refractive index and thickness of medium $a \in \{oil, oil^*, g, g^*\}$; where *oil* and *g* are respectively immersion oil and coverslip, and the superscript ‘*’ denotes the corresponding value under design condition. t_s and μ_s denote the distance of the object from the coverslip and the effective refractive index of the specimen. Δz is the defocus of the microscope arrangement and is given by $\frac{\Delta z}{\mu_{oil}} = \frac{t_s}{\mu_s} + \frac{t_g}{\mu_g} + \frac{t_{oil}}{\mu_{oil}} -$

$\frac{t_g^*}{\mu_g^*} - \frac{t_{oil}^*}{\mu_{oil}^*}$. It is also the distance by which the stage has to be moved (by changing t_{oil}) to achieve

the best geometric focus.[11]

$$PSF_{GL}(x_d, y_d, z_d) = \left| \frac{C}{z_d} \int_0^1 J_0 \left(\frac{kNA(x_d^2 + y_d^2)^{1/2} r}{M} \right) e^{i k \text{OPD}} r dr \right|^2 \quad \text{D.1}$$

$$\begin{aligned} \text{OPD} = \Delta z & \sqrt{\mu_{oil}^2 - NA^2 r^2} + t_s \left\{ \sqrt{\mu_s^2 - NA^2 r^2} - \frac{\mu_{oil}}{\mu_s} \sqrt{\mu_{oil}^2 - NA^2 r^2} \right\} \\ & + t_g \left\{ \sqrt{\mu_g^2 - NA^2 r^2} - \frac{\mu_{oil}}{\mu_g} \sqrt{\mu_{oil}^2 - NA^2 r^2} \right\} \\ & - t_{g^*} \left\{ \sqrt{\mu_{g^*}^2 - NA^2 r^2} - \frac{\mu_{oil}}{\mu_{g^*}} \sqrt{\mu_{oil}^2 - NA^2 r^2} \right\} \\ & - t_{oil^*} \left\{ \sqrt{\mu_{oil^*}^2 - NA^2 r^2} - \frac{\mu_{oil}}{\mu_{oil^*}} \sqrt{\mu_{oil}^2 - NA^2 r^2} \right\} \end{aligned} \quad \text{D.2}$$

To use the Gibson and Lanni[11] PSF in *in-silico* microscopy it should be transformed into a function of (ℓ', m', n') . Since the objective lens is lateral-shift invariant, having the detector at (x_d, y_d) and the object at $(0,0)$ is equivalent to having the detector at $(0,0)$ and the object at

$(-x_d, -y_d)$, i.e., $(\ell', m') = \left(\frac{x_d}{M}, \frac{y_d}{M} \right)$ and $r' = \sqrt{(\ell')^2 + (m')^2} = \frac{\sqrt{x_d^2 + y_d^2}}{M}$. Let the microscope in

non-design condition be focused on a plane located at a distance of t_{s0} below the coverslip, which corresponds to the object focal plane n_o . Focusing the microscope at t_{s0} implies that the best geometric focus $\Delta z = 0$ is achieved by adjusting t_{oil} . The distance of an out-of-focus object from the coverslip, t_s , is the sum of the distance between the object focal plane and the coverslip (t_{s0}) and the distance of the object from the object focal plane ($-n'$), i.e., $t_s = t_{s0} - n'$. The defocus produced by this out-of-focus object is given by $\Delta z = \frac{(t_s - t_{s0})\mu_{oil}}{\mu_s} = -\frac{n'\mu_{oil}}{\mu_s}$. In Eq D.1, the

maximum value of the PSF is given by $(C/2z_d)^2$, which occurs for $\text{OPD} = 0$ and $r' = 0$. To maintain consistency with the PSF definition in **Chapter 6**, $(C/2z_d)^2$ is replaced with I_0 .

Furthermore, to control the resolution of the *in-silico* images the full-width-at-half maximum (FWHM) scaling factor f_s is introduced with $\ell' = f_s \ell$. Using the aforementioned substitutions, the PSF in Eq D.1, D.2 can be rewritten as Eq D.3, D.4. This form of the PSF, if denoted by $PSF_{GL}(\ell', m', n'; t_{sO})$, can be used in the *in-silico* microscopy to generate images using the convolution $I(\ell', m', n_o) = PSF_{GL}(\ell', m', n_o; t_{sO}) * \rho_{\ell mn}$.

$$PSF_{GL}(\ell', m', n'; t_{sO}) = 4I_0 \left| \int_0^1 J_0(\ell' NA r' r) e^{i k' \text{OPD}} r dr \right|^2 \quad \text{D.3}$$

$$\begin{aligned} \text{OPD} = & -n' \sqrt{\mu_s^2 - \text{NA}^2 r^2} + t_{sO} \left\{ \sqrt{\mu_s^2 - \text{NA}^2 r^2} - \frac{\mu_{oil}}{\mu_s} \sqrt{\mu_{oil}^2 - \text{NA}^2 r^2} \right\} \\ & + t_g \left\{ \sqrt{\mu_g^2 - \text{NA}^2 r^2} - \frac{\mu_{oil}}{\mu_g} \sqrt{\mu_{oil}^2 - \text{NA}^2 r^2} \right\} \\ & - t_{g^*} \left\{ \sqrt{\mu_{g^*}^2 - \text{NA}^2 r^2} - \frac{\mu_{oil}}{\mu_{g^*}} \sqrt{\mu_{oil}^2 - \text{NA}^2 r^2} \right\} \\ & - t_{oil^*} \left\{ \sqrt{\mu_{oil^*}^2 - \text{NA}^2 r^2} - \frac{\mu_{oil}}{\mu_{oil^*}} \sqrt{\mu_{oil}^2 - \text{NA}^2 r^2} \right\} \end{aligned} \quad \text{D.4}$$

To demonstrate the similarity between the PSF by Gandy[12] and Gibson and Lanni[11], PSF by Gandy[12] is transformed using $r = \sin \theta / \sin \psi$ and $\mu_{oil} \sin \psi = \text{NA}$ in Eq D.5 (Note $\mu_s = \mu_{oil}$ in Gandy[12]). Comparing the PSF by Gibson and Lanni[11] (Eq D.3) and Gandy[12] (Eq D.5), we note that their OPD is the same if the *in-silico* microscope operates in design condition ($\mu_{oil} = \mu_{oil^*}$, $\mu_g = \mu_{g^*}$, $t_g = t_{g^*}$) and $\mu_s = \mu_{oil}$. Moreover, PSF by Gibson and Lanni[11] does not have the apodization factor $\cos^{-1/2}(\theta) = \left[1 - \left(\frac{\text{NA} r}{\mu_{oil}} \right)^2 \right]^{-1/4}$, which is also pointed out by Haerberlé.[13] Adding the apodization factor, the new PSF given by Eq. D.6 (with OPD given by Eq D.4) is referred to as the modified Gandy PSF (indicated by subscript MG).

$$PSF(\ell', m', n') = I_0 \left| \frac{3 \sin^2 \psi}{2(1 - \cos^{3/2} \psi)} \int_0^1 e^{-i k' n' [\mu_{oil}^2 - \text{NA}^2 r^2]^{1/2}} J_0(\ell' r' \text{NA} r) r \left[1 - \left(\frac{\text{NA} r}{\mu_{oil}} \right)^2 \right]^{-1/4} dr \right|^2 \quad \text{D.5}$$

$$PSF_{MG}(\ell', m', n'; \tau_{s0}) = I_0 \left| \frac{3 \sin^2 \psi}{2(1 - \cos^{3/2} \psi)} \int_0^1 J_0(\ell' r' NA r') e^{i \ell' OPD r'} \left[1 - \left(\frac{NA r'}{\mu_{oil}} \right)^2 \right]^{-1/4} dr' \right|^2 \quad D.6$$

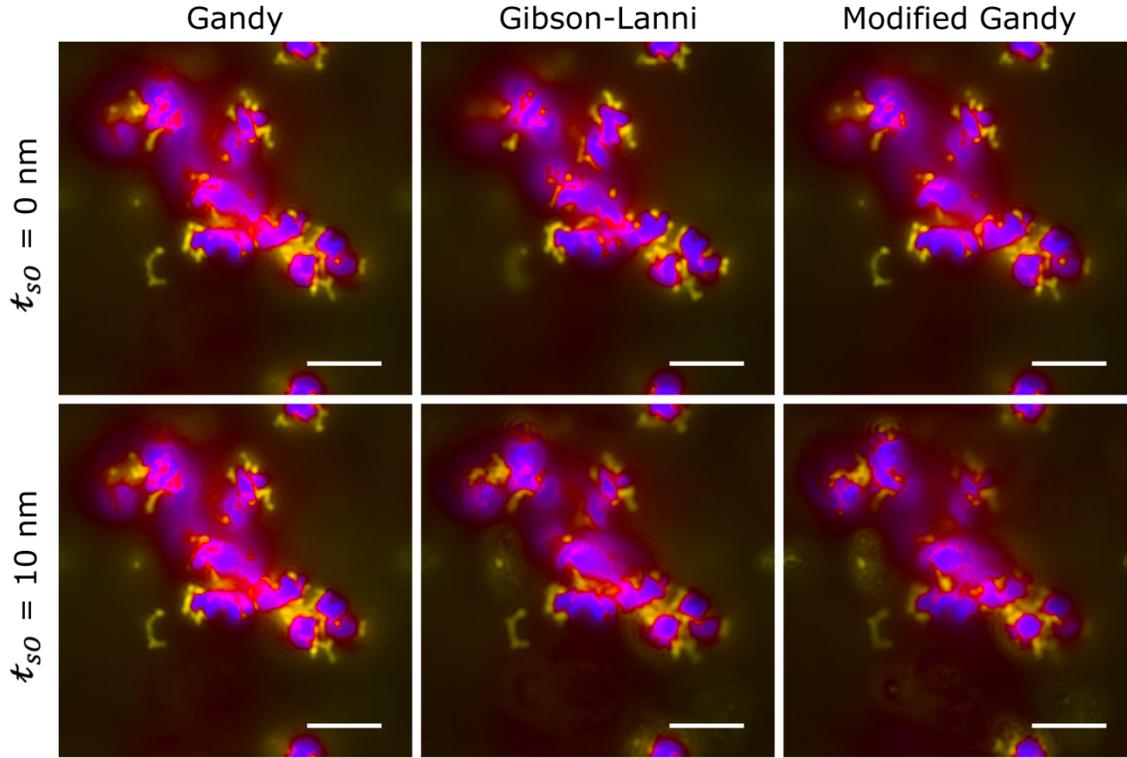


Figure D.2: *In-silico* microscopy images generated with different PSF models. Molecular simulation (MS) on PEI-DNA aggregation[3] (Section D.1.2) is used as the specimen. PSF is modeled with $NA = 1.3$, $\mu = \mu_{oil} = \mu_{oil}^* = 1.51$ ($\psi = 59.4$), $\tau_{oil}^* = 300$ nm, $\mu_g = \mu_g^* = 1.522$, $\tau_g = \tau_g^* = 320$ nm, $\mu_s = 1.33$, $f_s = 530$, n is z-axis, $n_o = 12$ nm, $\Delta \ell' = \Delta m' = 0.1$ nm, $\Delta n' = 0.4$ nm, $P_{\ell'} = P_{m'} = P_{n'} = 25$ nm, $(\lambda, I_0) = (670$ nm, 0.13) for DNA and (518 nm, 0.27) for PEI. The *in-silico* microscope is focused at the distance of $\tau_{s0} = 0$ and 10 nm below the coverslip. Scale bar, 5 nm.

Comparison of *in-silico* microscopy images generated with Gandy,[12] Gibson and Lanni,[11] and modified Gandy PSFs is shown in **Figure D.2** using the polyethylenimine (PEI)-DNA aggregation simulation[3] (Section D.1.2). The *in-silico* microscope is in design condition ($\tau_{a^*} = \tau_a$, $\mu_{a^*} = \mu_a$) and focused at a distance of $\tau_{s0} = 0$ and 10 nm below the coverslip. At $\tau_{s0} = 0$ nm, the Gandy[12] PSF considers the apodization factor but not the refractive index mismatch

($\mu_s = \mu_{oil}$), the Gibson and Lanni[11] PSF considers the refractive index mismatch ($\mu_s \neq \mu_{oil}$) but not the apodization factor, whereas the modified Gandy PSF considers both. The difference between the images is minimal. At $t_{s0} = 10$ nm, the Gandy[12] PSF produces an image identical to that for $t_{s0} = 0$ nm because it is depth-invariant, whereas noticeable changes are observed in the images produced from the Gibson and Lanni[11] and modified Gandy PSFs. For these two PSFs, the background appears to be brighter for the depth of 10 nm, which is due to spherical aberrations arising from refractive index mismatch. These images are also less sharp due to increased spherical aberration at higher depth t_{s0} . Other PSFs such as those by Hell *et al.*,[14] and Richards and Wolf[15] can be used by transforming the PSF to a function of ($l', m', n'; t_{s0}$) but it is not shown in this work.

D.3. Modes of *in-silico* microscopy

To demonstrate different modes of the *in-silico* microscopy, a synthetic model cell is constructed with a nucleus and transfected with PEI-DNA nanoparticles using the following steps. First, the all-atom structure of a nuclear DNA created by Sun *et al.*[16] was used to create a Martini DNA[17] using the stiff forcefield. Second, a model nucleus was created by placing multiple Martini DNAs randomly in a cubic box of side 55 nm using Gromacs[4] *insert-molecules* function with a van der Waal radius of 1 nm. Martini DNAs with at least one bead located outside a sphere (sharing the center with the box) of diameter 55 nm is removed. Third, the model cell was constructed by placing the model nucleus in a cubic box of side 170 nm centered at (100, 100, 100) nm. Two-hundred replicas of an aggregated PEI-DNA nanoparticle from a previous work[3] were added at into the model cell with randomized orientations and locations using Gromacs[4] *insert-molecules*. A total of 5400 DNAs and 54000 PEIs were added into the model cell to form the transfected PEI-DNA nanoparticles.

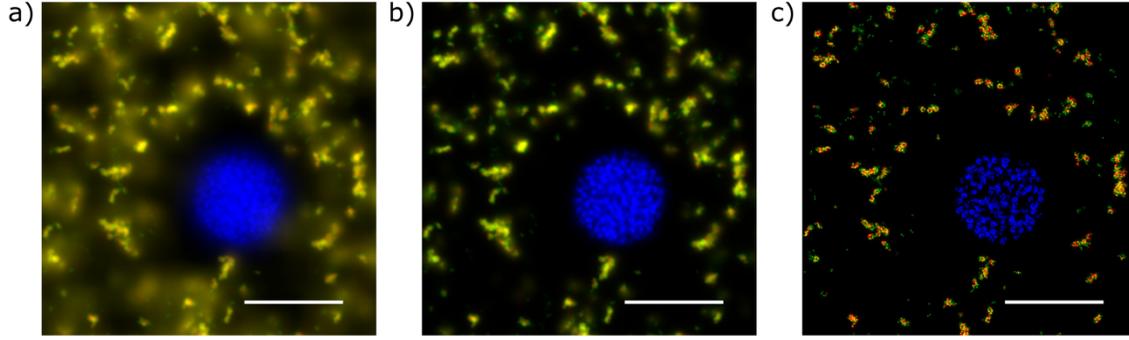


Figure D.3: *In-silico* images of a model cell transfected with PEI-DNA nanoparticles. The Gandy[12] PSF is calculated with $\psi = 59.4^\circ$, $\mu = 1.51$, n is z , $n_o = 85$ nm, $\Delta\ell' = \Delta m' = \Delta n' = 0.1$ nm, and $P_{\ell'} = P_{m'} = 25$ nm. The phosphate particles of nuclear DNA, phosphate particles in DNA of the nanoparticles, and amine particles in PEI emit light of wavelength 461, 670, and 518 nm respectively, and are assigned blue, red, and green hues respectively. (a) Widefield image is generated by using $f_s = 130$, $P_{n'} = 50$ nm, and $(I_0, \lambda) = (461 \text{ nm}, 0.04)$, $(518 \text{ nm}, 0.06)$, and $(670 \text{ nm}, 0.05)$. (b) OSM image is generated using $f_s = 130$, $P_{n'} = 10$ nm, and $(I_0, \lambda) = (461 \text{ nm}, 0.1)$, $(518 \text{ nm}, 0.09)$, and $(670 \text{ nm}, 0.06)$. (c) super-resolution microscopy is generated using $f_s = 520$, $P_{n'} = 2$ nm, and $(I_0, \lambda) = (461 \text{ nm}, 4.0)$, $(518 \text{ nm}, 2.0)$, and $(670 \text{ nm}, 2.0)$. Scale bars, 50 nm.

In-silico microscopy images for this model cell is generated in widefield (**Figure D.3a**), optical sectioning microscopy (OSM; **Figure D.3b**), and super-resolution (**Figure D.3c**) modes. The thickness of excitation $P_{n'}$ for the widefield mode should be such that the *in-silico* image does not change when $P_{n'}$ is increased. For wavelengths 461, 518 and 670 nm, and FWHM scaling factor $f_s = 130$, such a $P_{n'}$ value was determined to be 50 nm (**Figure D.3a**). To generate an *in-silico* image in OSM mode, FWHM scaling factor f_s is kept the same and the thickness of excitation $P_{n'}$ is reduced. For our model cell $P_{n'} = 10$ nm was found to produce an *in-silico* image (**Figure D.3b**) that resembles an experimental OSM image. Since deconvolution algorithms aim to reduce out-of-focus fluorescence, *in-silico* images in OSM mode also resemble experimental widefield microscopy image after deconvolution. For example, deconvoluted experimental widefield microscopy images of Schaffer *et al.*[18] are visually similar to **Figure D.3b**. To

generate an *in-silico* image in super-resolution mode, FWHM scaling factor f_s is increased and the thickness of excitation $P_{n'}$ is reduced. For our model cell $f_s = 520$ and $P_{n'} = 2$ nm was found to produce an *in-silico* image (**Figure D.3c**) that resembles an experimental super-resolution image.

D.4. Choosing maximum fluorescence intensity and FWHM scaling factor

The optimal combination of I_0 and f_s can be determined based on the following guiding principles: (1) All fluorophore particles of interest in focus should be visible. (2) Different particles of the same fluorophore type, therefore emitting the same color, should be distinguishable if and only if they are not bound to each other. (3) Colocalized hues (i.e., mixture of assigned hues) should be visible if and only if fluorophore particles of different types are bound to each other. (4) The user should take into consideration the desired level of resolution. It is suggested that the combination of I_0 and f_s be determined from a simulation time at which the configuration of the simulated system is best known. For example, for the MS on PEI-DNA aggregation[3] (**Section D.1.2**), the initial unaggregated configuration can be used to guide the choice of I_0 and f_s . In this case the basic principles translate to: (1) Both DNA and PEI should be visible. (2) Individual DNA molecules should be distinguishable because they are unaggregated. (3) Little to no colocalization of DNA and PEI should be observed because DNAs and PEIs are not yet bound to each other. (4) Individual particles in DNA and PEI molecules should not be identifiable in order to make better comparison with experiments. *In silico* microscopy images for the initial configuration are presented in **Figure D.4** for 25 combinations of I_0 and f_s . Based on the guiding principles, the best combinations are found to be $I_0 = 0.2$ and $f_s = 530$, and $I_0 = 0.3$ and $f_s = 660$.

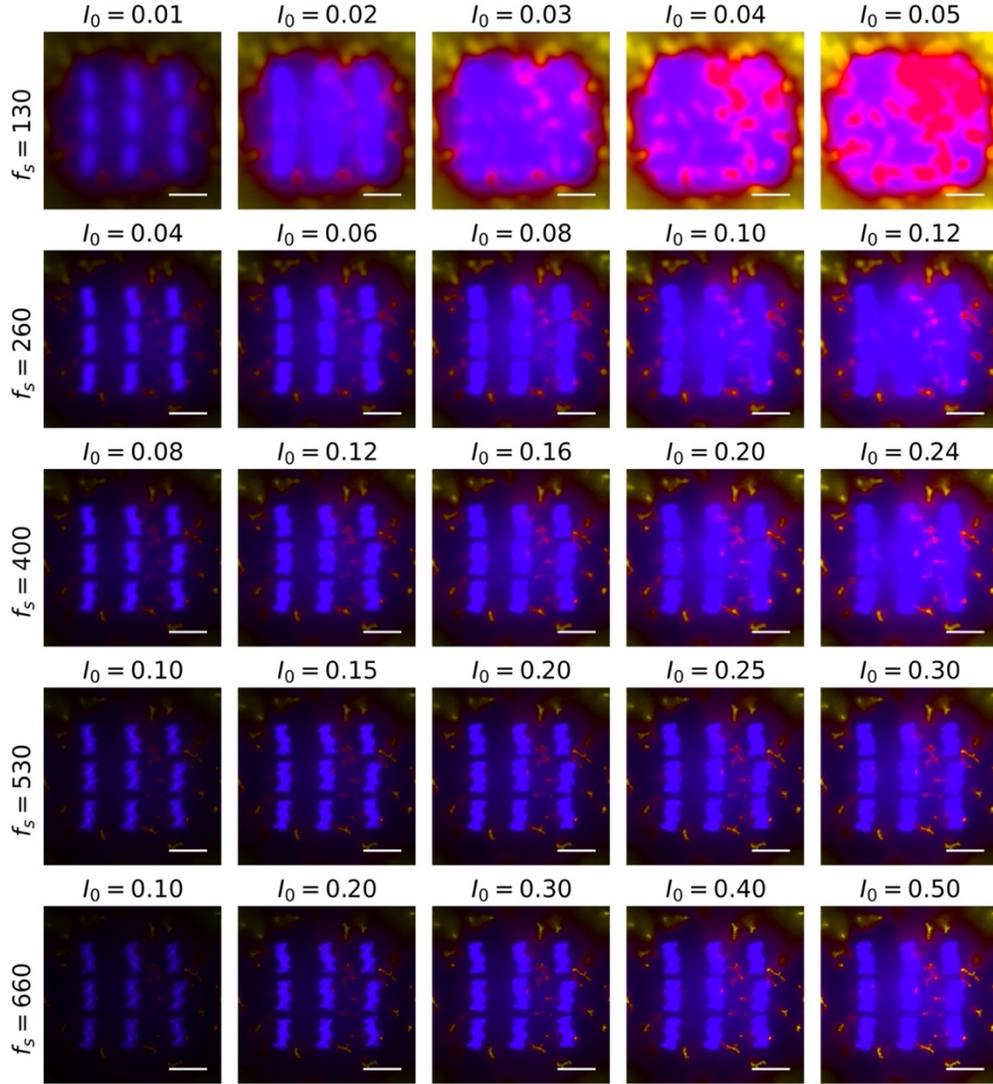


Figure D.4: *In-silico* microscopy images for the initial configuration of a PEI-DNA aggregation MS, using different combinations of maximum fluorescence intensity I_0 and FWHM scaling factor f_s . I_0 and f_s are kept the same for all fluorophore types. The Gandy[12] PSF is modelled with $\psi = 59.4^\circ$, $\mu = 1.51$, n is z-axis, $n_o = 12$ nm, $\Delta\ell' = \Delta m' = 0.1$ nm, $\Delta n' = 0.05$ nm, $P_{\ell'} = P_{m'} = P_{n'} = 25$ nm, $\lambda = 670$ nm for DNA and 518 nm for PEI; DNA and PEI particles are assigned indigo and yellow hues respectively; and no time-averaging is performed.

Depending on the MS, it may be necessary to choose different I_0 for different fluorophore types. For each fluorophore type (i.e., a specific wavelength of emitted light), a histogram of I_1 (I evaluated with $I_0 = 1$) at a reference simulation time can help guide the adjustment of I_0 . Such

histograms are plotted in **Figure D.5a** for DNA (670 nm) and PEI (518 nm) particles in the initial configuration of the MS on PEI-DNA aggregation[3]. Here $f_s = 530$ is used, and both histograms are normalized so that the maximum count is one. The peak observed at low I_1 represents the fluorescence from particles out of focus, and the tail of the histogram at larger I_1 represents fluorescence from particles in focus. To improve I_0 iteratively, let I_0^i be the value of I_0 at the i^{th} iteration with $I_0^0 = 1$. Since all $I_0^i I_1 > 1$ will be replaced with 1 while generating images, I_0^i must be such that $I_0^i I_1$ of the in-focus particles is close to 1 (bright) and that of the out-of-focus particles is close to 0. So, practically one would first determine the maximum I_1 at which the histogram becomes zero, denoted by $\max I_1$, and then propose $I_0^1 = 1/\max I_1$. $\max I_1$ for the two fluorophore types in **Figure D.5a** are marked by the dashed lines with full opacity. I_0^1 is determined to be 0.08 and 0.27 for DNA and PEI particles respectively (12.5 and 3.7 on the horizontal axis of **Figure D.5a**). The *in-silico* microscopy image with these I_0^1 values is shown in **Figure D.5b**. Further adjustments of I_0 can be made depending on the quality of the images. For instance, in **Figure D.5b** while the fluorescence of PEI is appropriate the brightness of the DNAs can be improved. We can also arrive at this conclusion by looking at the histogram in **Figure D.5a**. Since the histogram tail for DNA is very broad, scaling I_1 by a small factor of 0.08 leads to low intensity for in-focus DNA particles near the beginning of the histogram tail (~ 3 on the horizontal axis of **Figure D.5a**). Further adjustments to I_0 can be done recursively using $I_0^{i+1} = I_0^i/(1 - I_0^i)$. Following the guiding principles previously described, I_0 for DNA is increased over four iterations to $I_0^5 = 0.13$. $1/I_0^i$ for each iteration is shown with decreasing opacity in **Figure D.5a**. The corresponding *in-silico* microscopy images are shown in **Figure D.5c-f**.

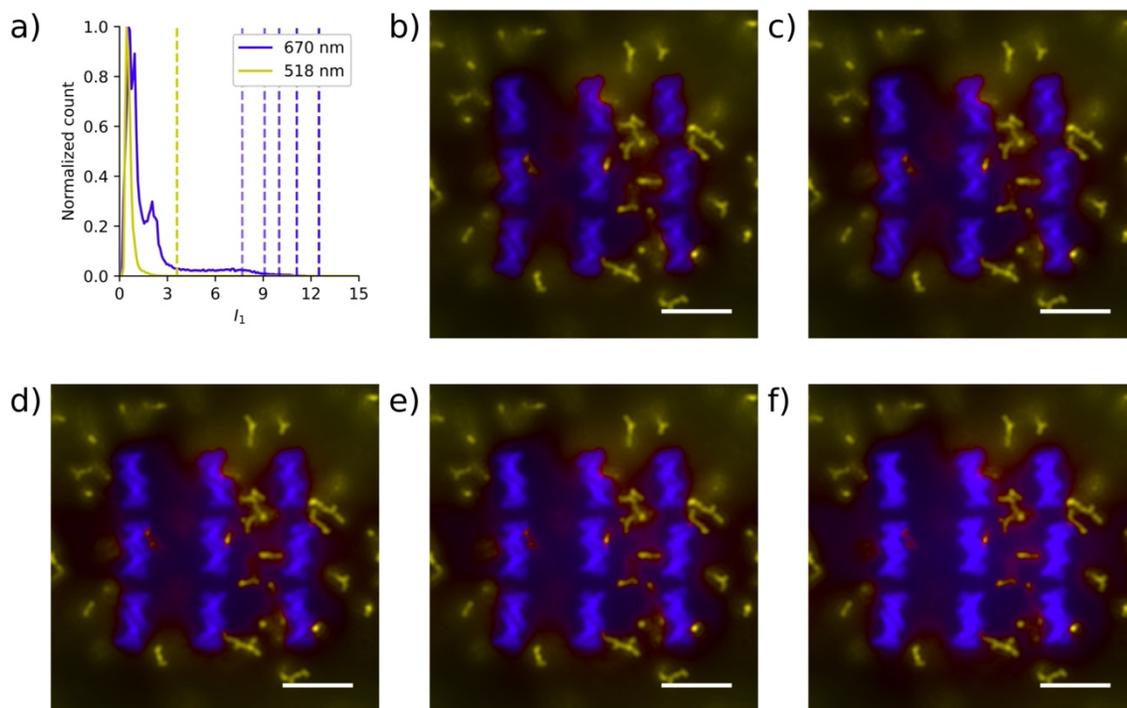


Figure D.5: Adjusting I_0 iteratively for a given f_s . a) Histogram of I_1 (monochrome intensity evaluated with $I_0 = 1$) for DNA-PEI aggregation MS at time $t = 0$. The curves are normalized along the vertical axis so that the maximum count is 1. The Gandy[12] PSF is modelled with $\psi = 59.4^\circ$, $\mu = 1.51$, $n = z$, $n_0 = 12$ nm, $\Delta\ell' = \Delta m' = 0.1$ nm, $\Delta n' = 0.05$ nm, $P_{\ell'} = P_{m'} = P_{n'} = 25$ nm, $f_s = 530$, $I_0 = 1$, $\lambda = 670$ nm for DNA and 518 nm for PEI. For each wavelength, a dashed line with 100% opacity is drawn at $\max I_1$ where the histogram decays to zero. $\max I_1 = 1/0.27$ for 518 nm and $1/0.08$ for 670 nm. The indigo dashed lines with less than 100% opacity are drawn at $1/0.09$, $1/0.10$, $1/0.11$ and $1/0.13$ from right to left, whose reciprocals correspond to I_0^i , I_0 after the first, second, third and fourth iterations, respectively. b-f) *In-silico* microscopy images generated with the parameters described in (a), and I_0 of (DNA, PEI) as (0.08, 0.27), (0.09, 0.27), (0.10, 0.27), (0.11, 0.27), and (0.13, 0.27) for (b)-(f) respectively. DNA and PEI particles are assigned indigo and yellow hues respectively; and no time-averaging is performed. Scale bar for (b)-(f), 5 nm.

D.5. Development of color mixing scheme

Hue-saturation-value (HSV) color space was developed to model an artist's way of producing colors.[2] However, to the best of our knowledge, mixing of colors has not been defined in the HSV space. Here we develop a color mixing scheme using HSV, in order to meet our requirements to observe colocalization.

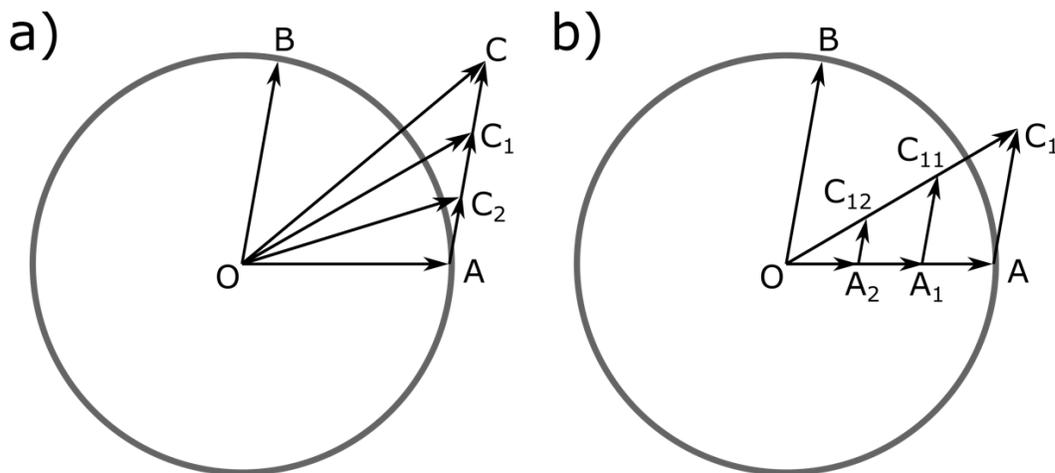


Figure D.6: Demonstrating vector addition in HSV. a) The circle centered at O has a unit radius. All points within the circle corresponds to a color with saturation of one. The polar angle represents the hue and radial distance from the center represents the value. The points A and B lie on the circumference. The vectors OA and OB represents two colors with the same value (=1) and saturation (=1), but different hues (the polar angle). Addition of the vectors OA and OB (or equivalently AC) produces OC. Decreasing the value of OB (i.e., AC) produces vectors AC₁, AC₂, etc. Addition of OA to AC₁ and AC₂ produces vectors OC₁ and OC₂; reducing the value of OB makes the resultant vector (and hue) closer to OA. Hues between OC and OB can be obtained by fixing the value of OB while decreasing the value of OA. b) Addition of vectors OA and AC₁ produces OC₁. Scaling vectors OA and AC₁ by the same factor yields vectors OA₁ and A₁C₁₁, OA₂ and A₂C₁₂, etc. The resultant vectors OC₁, OC₁₁, and OC₁₂ all have same hues but different values.

Let (H_{mix}, V_{mix}) be the hue and value produced from mixing a set of saturated colors $\{(H_1, V_1), (H_2, V_2), \dots, (H_n, V_n)\}$ denoted as $\{H_j, V_j\}$, where $H_j \in [0, 360^\circ)$ and $V_j \in [0, 1]$. From artistic intuition, combination of red and yellow with different proportion should yield different

shades of orange (yellowish-orange, orangish-red, etc.), and not hues such as blue and green. In the color wheel, this points to the minor sector connecting the two hues being mixed (**Figure D.7b**). This artistic intuition of hue mixing can be modeled by treating each element in $\{H_j, V_j\}$ as a two-dimensional vector or complex number ($V_j e^{iH_j}$) and adding them (**Figure D.6a**). The hue of the mixed colors is the angle evaluated from the positive x -axis counter-clockwise to the resultant vector, or the argument of the resultant complex number,

$$H_{mix} = \arg \left(\sum_{j=1}^N V_j e^{iH_j} \right) \quad \text{D.7}$$

The resultant value, however, cannot be simply set to the magnitude of the resultant vector or the complex number because the magnitude can be more than one, while a value cannot.[2] Here we introduce a new mapping that will create the vector of $V_{mix} e^{iH_{mix}}$ from the sum, $\sum_{j=1}^N V_j e^{iH_j}$. To define V_{mix} , we recognize that $\sum_{j=1}^N a_1 V_j e^{iH_j} = a_1 \sum_{j=1}^N V_j e^{iH_j}$. That is, if the vector for each color is scaled by a non-negative real number a_1 , the hue of the resultant vector stays the same while the resultant value (V_{mix}) is scaled by a_1 (**Figure D.6b**). We wish to define V_{mix} so that all physically valid scaling of the colors ($V_j e^{iH_j}$) to be mixed will produce values that are not more than 1. Physically valid scaling is specified by the range of a_1 . Since the values of individual colors after the scaling should be less than or equal to one,

$$\begin{aligned} a_1 V_j \leq 1 \quad \forall j &\Rightarrow a_1 \max_1(V_j) \leq 1 \\ \Rightarrow a_1 &\leq \frac{1}{\max_1(V_j)} \Rightarrow a_1 \in \left[0, \frac{1}{\max_1(V_j)} \right] \end{aligned}$$

where, $\max_n(V_j)$ returns the n^{th} largest from sorted values (V_j) of colors being mixed.

V_{mix} is defined such that for the entire range of a_1 , $a_1 V_{mix} \leq 1$. Conveniently, we set

$$\max(a_1)V_{mix} = 1$$

$$V_{mix} = \max_1(V_j) \quad \text{D.8}$$

The following considerations are taken when defining the saturation of mixed colors S_{mix} . Similar to value, saturation cannot be more than one.[2] It is desirable for two-color mixtures to be fully saturated ($S_{mix} = 1$), and for mixtures involving more than two colors to be desaturated ($S_{mix} < 1$). We propose the following expression for S_{mix} ,

$$S_{mix} = 1 - a_2 \max_3(V_j)$$

where a_2 is a positive real number. When $\max_3(V_j) = 0$, it represents a two-color mixture and therefore $S_{mix} = 1$. For $\max_3(V_j) > 0$, it represents mixture of more than two colors and $S_{mix} < 1$. If after sorting the three largest values of V_j are equal, i.e., $\max_3(V_j) = \max_2(V_j) = \max_1(V_j)$, $\max_3(V_j)$ achieves its largest possible value and S_{mix} is smallest. Conveniently, we set this lowest $S_{mix} = 0$, i.e.,

$$0 = 1 - a_2 \max_1(V_j) \Rightarrow a_2 = \frac{1}{\max_1(V_j)}$$

In turn,

$$S_{mix} = 1 - \frac{\max_3(V_j)}{\max_1(V_j)} \quad \text{D.9}$$

It is interesting to note that expressions of value (V) and saturation (S) provided by Smith[2] to describe the conversion from RGB to HSV are very similar to the expressions of S_{mix} and V_{mix} in Eq D.9 and D.10.

$$V = \max(R, G, B)$$

$$S = 1 - \frac{\min(R, G, B)}{\max(R, G, B)}$$

Here, R , G , and B refer to the intensities of red, green and blue light respectively.

D.6. Choosing hues for fluorophore types

D.6.1. Choosing hues for two fluorophore types

Adding hues 180° apart results in colors along the line joining them rather than a sector of the color wheel (**Figure D.7a**). Such hue pairs should be avoided, to ensure a wide range of colocalization hues are available.

D.6.2. Choosing hues for three fluorophore types

Three fluorophore types can mutually mix, in the form of three pairs. Mixing every pair of fluorophore type would produce a range of hues on the minor sector of the color wheel corresponding to the hue pair. In order to clearly identify the colocalization of hues, care should be taken to avoid overlap between the three minor sectors. For example, if the three fluorophore types are assigned red ($H_r = 0^\circ$), green ($H_g = 120^\circ$), and blue ($H_b = 240^\circ$) hues, the minor sectors will not overlap (**Figure D.7b**). On the contrary, if the three fluorophore types are assigned red, orange ($H_o = 30^\circ$) and yellow ($H_y = 60^\circ$) hues, the minor sectors overlap (**Figure D.7c**). In such a case, a mixed hue of $H_{mix} = 40^\circ$ can represent colocalization of orange-yellow or colocalization of red-yellow, and it would not be possible to provide a unique physical interpretation of the mixed hue.

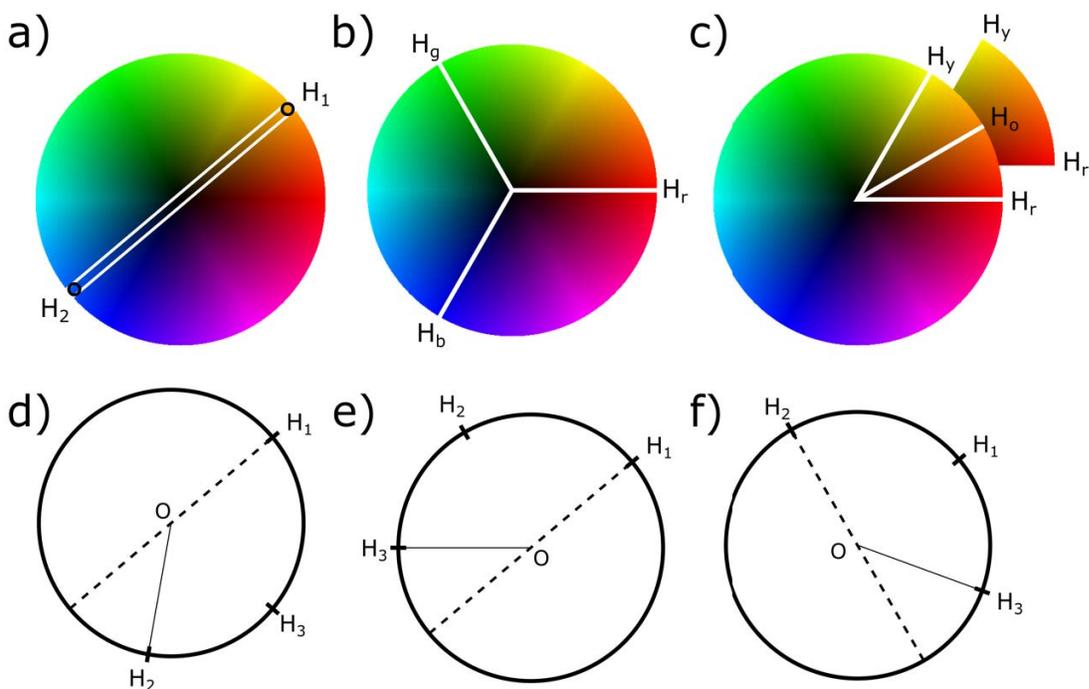


Figure D.7: Choosing hues to represent colocalization. a) Colocalization of fluorophores with hues H_1 and H_2 180° apart produces hues along the line connecting them. b) If three fluorophore types are assigned red (H_r), green (H_g) and blue (H_b) hues, the minor sectors between each hue pairs do not overlap. c) If three fluorophore types are assigned red, orange (H_o), and yellow (H_y) hues, the minor sectors between the hue pairs overlap. For three hues $0 \leq H_1 < H_2 < H_3 < 360^\circ$ d) if $H_2 > H_1 + 180^\circ$ the minor sectors overlap, e) if $H_3 < H_1 + 180^\circ$ the minor sectors overlap, f) If $H_3 > 180^\circ + H_2$ the minor sectors overlap.

We can determine a set of mathematical conditions for the assigned hues so that the minor sectors between hue pairs do not overlap, i.e., a unique physical interpretation exists for each mixed hue. Let H_1 , H_2 , and H_3 be the hues for the three fluorophore types such that $0 \leq H_1 < H_2 < H_3 < 360^\circ$. If $H_2 > H_1 + 180^\circ$, then H_3 will lie on the minor sector of $H_1 - H_2$, i.e., the minor sectors will overlap (**Figure D.7d**). To avoid this and the situation that $H_2 = H_1 + 180^\circ$, the following condition is imposed

$$H_2 < H_1 + 180^\circ \quad \text{D.10}$$

If $H_1 \geq 180^\circ$, H_1 , H_2 , and H_3 will be in the same minor sector, therefore

$$H_1 < 180^\circ \quad \text{D.11}$$

If $H_3 < H_1 + 180^\circ$, H_2 will lie on the minor sector of $H_1 - H_3$ (**Figure D.7e**). Similarly, if $H_3 > H_2 + 180^\circ$, H_1 will lie on the minor sector of $H_2 - H_3$ (**Figure D.7f**). To avoid such situations along with imposing $H_3 < 360^\circ$, H_3 should satisfy

$$H_1 + 180^\circ < H_3 < \min(H_2 + 180^\circ, 360^\circ) \quad \text{D.12}$$

D.6.3. Choosing hues for more than three fluorophore types

Using more than three fluorophore types, the minor sectors of different hue pairs would always overlap because the three non-overlapping sectors would span the entire color wheel. Therefore, it is not possible to provide a unique physical interpretation for all color colocalizations. This is not a limitation of the present color mixing scheme, but rather stems from the limitation of a standard human eye, which can only perceive three dimensional colors (H, S, V are three dimensions, similarly R, G, B). If it becomes necessary to use more than three fluorophore types to represent particles in an MS, we recommend that one of the overlapping pairs of hues be assigned to particles that do not colocalize in the MS, so that colocalization of particles can be uniquely inferred from the hue of mixed colors. In the case where this is impossible, multiple *in-silico* microscopy images can be generated for different triad of fluorophores.

D.6.4. Considerations for color-blind readers

Since color-blindness reduces the dimensionality of perceivable colors, use of more than two fluorophore types is not ideal, unless certain pairs of fluorophores never colocalize. When using two fluorophore types, the difference in their hues should be close, but not equal, to 180° in order to increase the color variation as much as possible. It would also be desirable if the chosen hues are perceived in the same way by readers with and without color-blindness. For example, yellow and blue are perceived in the same way by trichromats (no color blindness), protanopes

and deuteranopes, whereas red-pink and cyan are perceived in the same way by trichromats and tritanopes.[19] In **Chapter 6**, yellow-blue combination was not used because their hues are 180° apart. Instead, yellow-indigo (hues 60°-255°) combination was used, where indigo is a named color that is closest to blue.

D.7. Implementing color-mixing scheme on experimental images

The color-mixing scheme described in this work can also be implemented on existing experimental images. First, the channels of an RGB image are split to obtain three (or two) monochrome images. Then hues can be assigned to each monochrome image and colors can be mixed using Eq. 6.3-6.5. To demonstrate this, fluorescence microscopy images of Simpson *et al.*[20] were obtained from Image Data Resource repository (<https://idr.openmicroscopy.org>). **Figure D.8a** represents the original images from the repository showing HeLa Kyoto cells, where colors were mixed in RGB scheme. Three monochrome images were extracted corresponding the red, green, and blue channels. Then violet, cyan and orange hues were assigned to the monochrome images and mixed using Eq 6.3-6.5 to obtain **Figure D.8b**. **Figure D.8c** is obtained using Fiji ImageJ[8] by adding the monochrome images in yellow, cyan and magenta channels. From the region of interest, it is clear that higher color contrast can be produced using the color mixing scheme described in this work.

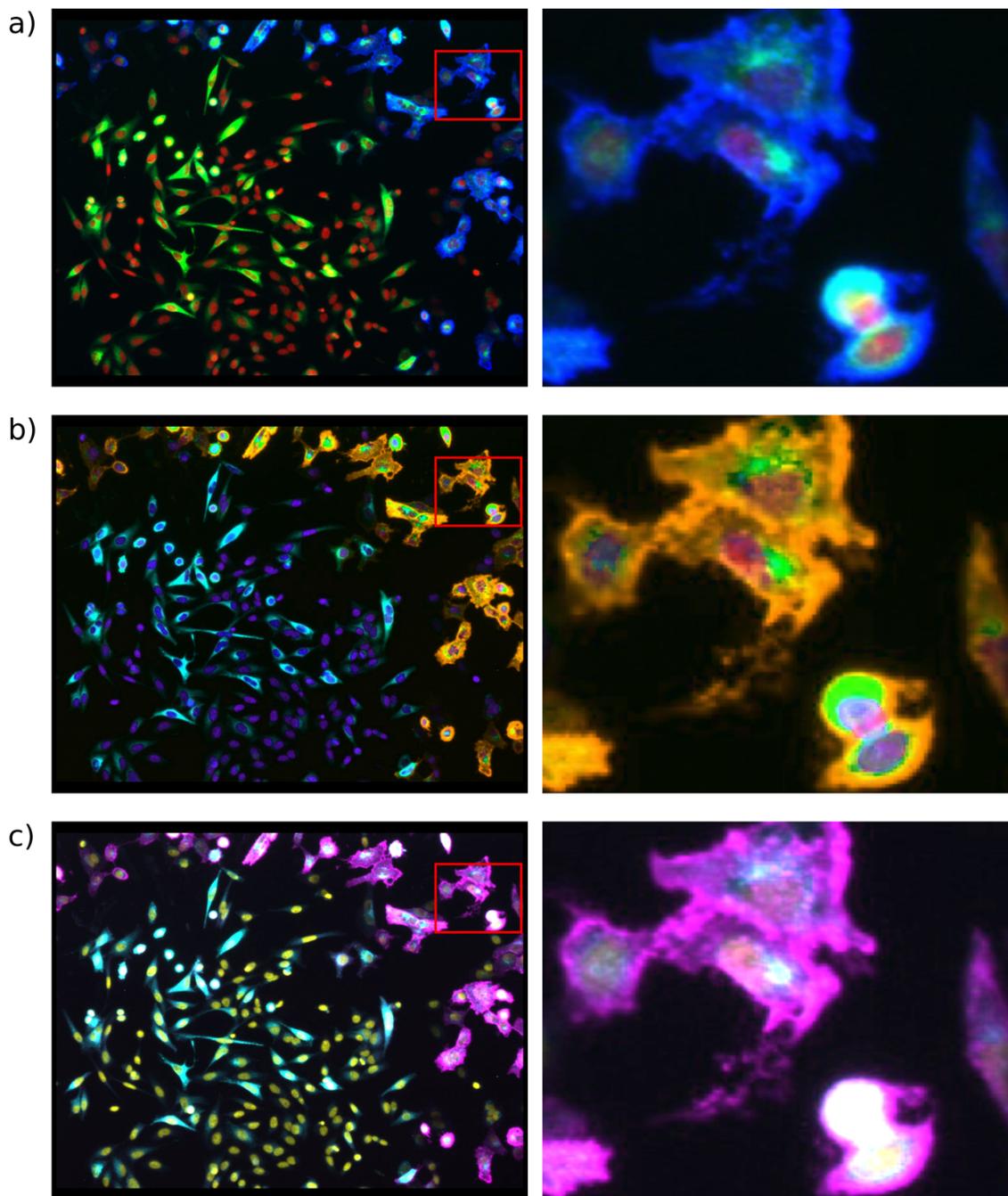


Figure D.8: (a) HeLa cells, where dapi-stained nucleus, expressed CFP-tsO45, and stained cell surface tsO45 are shown in red, green and blue respectively (left).[20] Copyright Simpson *et al.*[20] licensed under CC BY 4.0; Image ID: 171499, Well ID: 61941, Plate ID: 408, “idr0009-simpson-secretion/ScreenA/0001-03—2005-08-01/F5”. The image was modified by adding a red rectangle denoting the region of interest, which is shown in the right panel. (b) Intensities from red, green,

and blue channels in (a) were used to mix colors in violet, cyan and orange using Eq. 6.3-6.5 (left). The region of interest is shown on the right. (c) The red, green, and blue channels in (a) were changed to yellow, cyan and magenta channels using Fiji ImageJ[8] (left). The region of interest is shown on the right.

D.8. Luminance of hues and its effects on color contrast

The visibility of fluorescence is affected by the contrast between the colors of the fluorophore (foreground) and its immediate surroundings (background), especially when the fluorophore occupies a small number of pixels. Quantitatively, a color contrast ratio[21,22] can be defined using $(L_1 + 0.05)/(L_2 + 0.05)$ where L_1 and L_2 are respectively the relative luminance[23] of foreground and background colors, and the number 0.05 arises from typical viewing flare.[24] To achieve a high contrast, this ratio should be much greater than or much smaller than 1. Relative luminance for different pure hues ($V, S = 1$) is shown in **Figure D.9**, where the polar angle represents hue and the radial distance represents relative luminance. Increase in value V or decrease in saturation S increases the relative luminance of a color.

For example, the visibility of ions (fluorophores with very small size) is examined in **Figure 6.3e** by using different color combinations for DNA-PEI-Ion. On a black background as indicated by the red arrows, the visibility follows the trend of yellow-cyan-magenta (YCM) > orange-cyan-violet (OCV) > red-green-blue (RGB). This is because the relative luminance of magenta > violet > blue (**Figure D.9**) while the black background has a relative luminance of 0. The colors of DNA and PEI have negligible effect here.

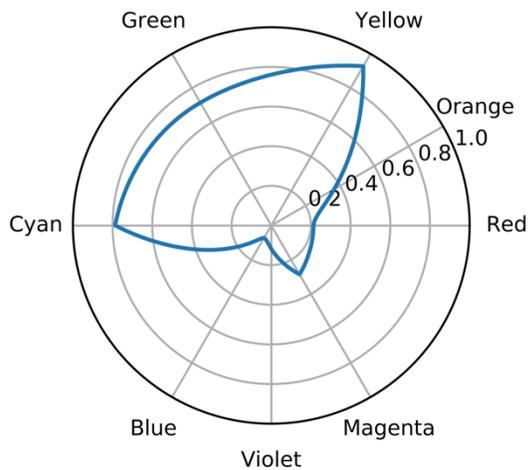


Figure D.9: Relative luminance of pure hues. The polar angle represents the hue, and the radial distance represents the relative luminance. The polar angle is marked for red (0°), orange (30°), yellow (60°), green (120°), cyan (180°), blue (240°), violet (270°), and magenta (300°). The relative luminance is shown for hues with saturation and value of one.

The visibility of ions highlighted by white arrows in **Figure 6.3e** follows a different trend, $OCV > YCM > RGB$. To understand this better, in **Figure D.10** we show a zoomed-in view of the ion (left), along with the relative luminance (middle) and hue (right) of pixels as a function of the distance from the central pixel. Clearly the visibility of ion in RGB is low due to poor color contrast, and higher in YCM and OCV due to better color contrast. Though the color contrast is similar for YCM and OCV, visibility is higher in OCV due to the larger number of pixels around the center with similar luminance, which makes the ions appear larger.

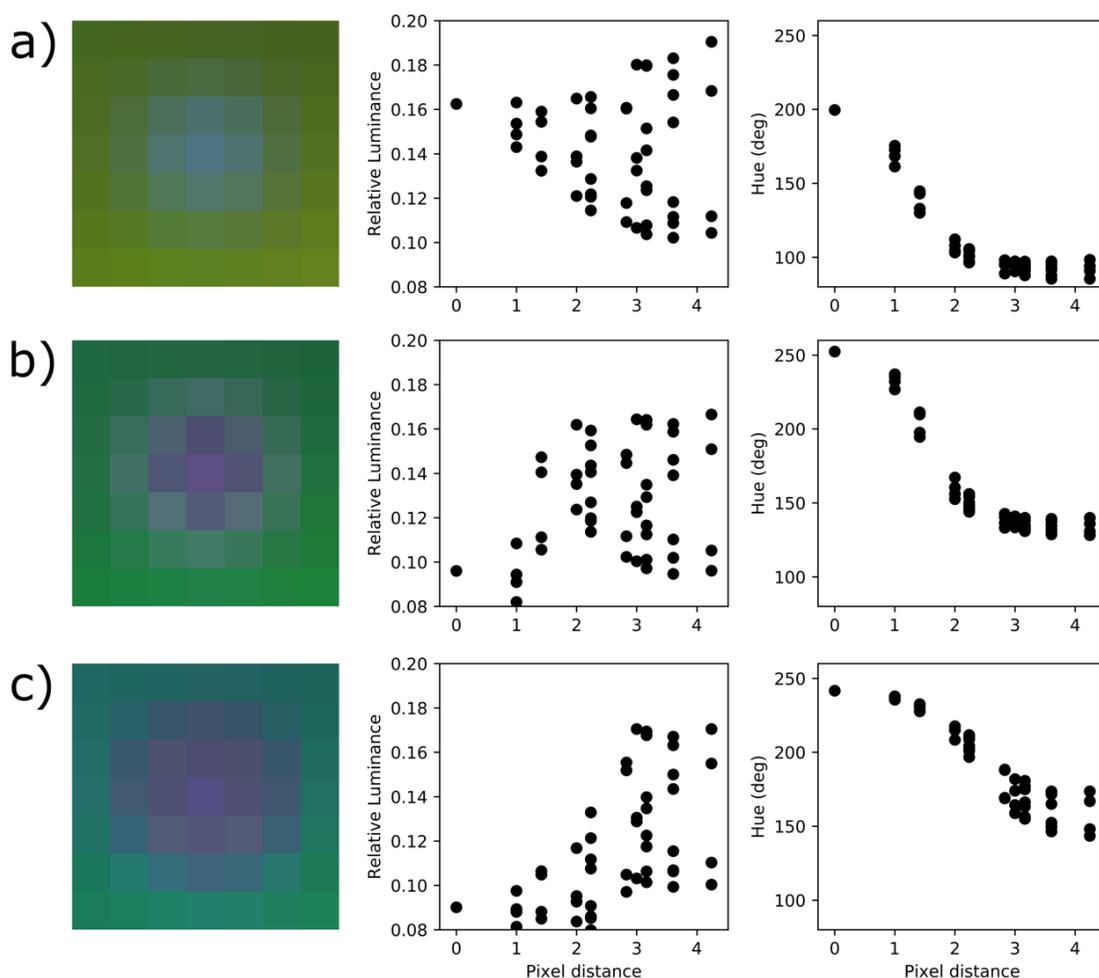


Figure D.10: Characteristics of ion visibility. The visibility characteristics of an ion shown by the white arrow in **Figure 6.3e** for color combinations a) red-green-blue, b) yellow-cyan-magenta, and c) orange-cyan-violet. For each color combination the image on the left is a zoomed *in-silico* microscopy image centered at the ion of interest. The figure in the middle shows the relative luminance as a function of pixel distance from the central pixel. The figure at the right shows the hue of the pixels as a function of pixel distance from the central pixel. Each pixel corresponds to a 0.1 nm by 0.1 nm square (same as $\Delta\ell'$, $\Delta m'$).

The difference in visibility can also be explained from **Figure D.9**. In these cases, the background color is not black but a saturated hue arising from the colocalization of PEI and DNA with trace amounts of ions; the foreground color is a saturated hue arising from the colocalization

of PEI and ion with trace amounts of DNA. This can be confirmed from the hue data in **Figure D.10** (right) as the pixel moves away from the center, changing from foreground to background. Therefore, we examine the variation in relative luminance in **Figure D.9** from a hue halfway between PEI and ion to a hue halfway between PEI and DNA. For RGB, the hue is 180° at the midpoint between PEI (120°) and ion (240°), and 60° at the midpoint between PEI and DNA (0°). In the hue range of 180° to 60° , the variation of relative luminance in **Figure D.9** is small (also see **Figure D.10a** (middle) where the relative luminance of pixels varies from 0.16 in the foreground to 0.16 ± 0.04 in the background), and the contrast ratio is expected to be close to 1. Consequently, the visibility of ion is low. For YCM, the variation of relative luminance in **Figure D.9** from hue 240° (midpoint hue between PEI and ion) to 120° (midpoint hue between PEI and DNA) is large, and thereby the color contrast and ion visibility is high. For OCV, the hue range of 225° (midpoint hue between PEI and ion) to 105° (midpoint hue between PEI and DNA) overlaps significantly with that of YCM. Therefore, the color contrast is expected to be similar. On the other hand, the hue range from PEI to ion is smaller in OCV (180° to 270°) than in YCM (180° to 300°). This results in a more gradual change of hue from foreground to background (see **Figure D.10b, c** (right)) when OCV is used, and hence more pixels around the center having similar relative luminance which is smaller than that of the background (see **Figure D.10b, c** (middle)). Consequently, the ion appears bigger.

D.9. Time-integrated and time-averaged images

In experimental microscopy, increasing exposure time increases the number of photons sensed by detectors. Within the framework of *in-silico* microscopy, the number of photons detected is modelled by integrating the fluorescence intensities over a certain period of time (the equivalent “exposure time”). Numerical integration further translates to a summation. Since all resultant

monochrome fluorescence intensities I are proportional to I_0 (see Eq 6.1 and 6.2), when intensities from multiple images (at different time steps) are added, the overall intensity I increases. As a result, the optimal I_0 determined based on histograms without any time integration/summation (**Figure D.5**) may no longer be optimal for time-integrated images. On the other hand, time-averaged intensity is not expected to change the magnitude of the overall intensity significantly, and optimal I_0 determined without any time averaging (**Figure D.5**) can still be used. Therefore, time-averaged image is preferred over time-integrated image, which closely models the experiments with a certain exposure time.

D.10. Adding Poisson-Gaussian noise

Noise in an imaging system can be modelled using a Poisson noise (\mathcal{P}) and an additive zero-mean Gaussian noise (\mathcal{N}). The corrupted image I_c for a noise free image I is given by Eq D.13, where I has values between 0 and 1, $a_p \in [0,1]$ is the effectiveness of the Poisson noise, and a_N is the variance of the Gaussian noise. Using $a_p = 1$ would add a Poisson noise similar to Lanza *et al.*,[25] and $a_p = 0$ would add no Poisson noise. The factor 255 is multiplied to I to convert it into an unsigned 8-bit integer. After determining the Poisson random variable $\mathcal{P}(255I)$ it is divided by 255 to convert the unsigned 8-bit integer to a real number between 0 and 1. In **Figure D.11**, corrupted images are produced with $a_p \in \{0,0.5,1\}$ and $a_N \in \{0,0.1\}$.

$$I_c = I(1 - a_p) + \frac{a_p \mathcal{P}(255I)}{255} + \mathcal{N}(0, a_N) \quad \text{D.13}$$

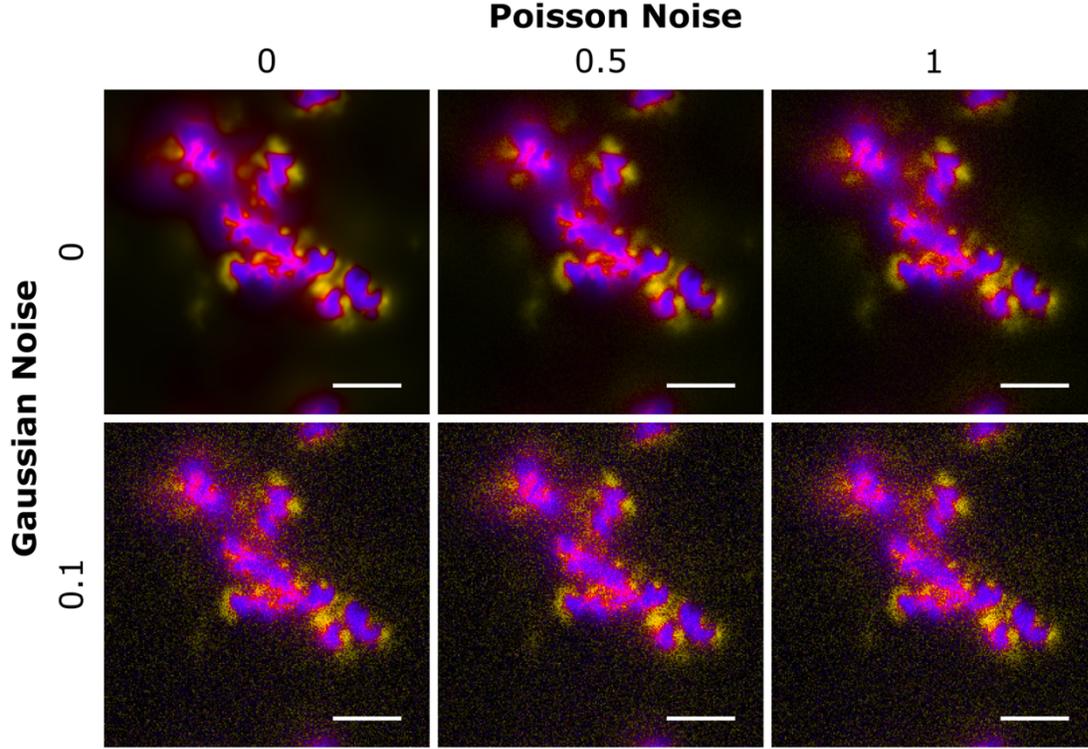


Figure D.11: Noise corrupted images with Poisson noise parameter $a_p = 0, 0.5$ and 1 , and Gaussian noise parameter $a_N = 0$ and 0.1 . The Gandy[12] PSF is modeled with $t = 1 \mu\text{s}$, $\psi = 59.4^\circ$, $\mu = 1.51$, n is z-axis, $n_o = 12 \text{ nm}$, $\Delta\ell' = \Delta m' = 0.1 \text{ nm}$, $\Delta n' = 0.05 \text{ nm}$, $P_{\ell'} = P_{m'} = P_{n'} = 25 \text{ nm}$, $f_s = 530$, $(\lambda, I_0) = (670 \text{ nm}, 0.13)$ for DNA and $(518 \text{ nm}, 0.27)$ for PEI; DNA and PEI particles are assigned indigo and yellow hues respectively. Scale bar, 5 nm .

D.11. Deconvolution of *in-silico* microscopy images

To demonstrate the use of existing deconvolution algorithms on *in-silico* microscopy images, first 3D *in-silico* images were calculated using $\Delta n_o = 1 \text{ nm}$. Then deconvolution was performed using 150 iterations of Richardson-Lucy algorithm[26,27] with DeconvolutionLab2[28] in Fiji ImageJ.[8] To determine the efficiency of the deconvolution algorithm, 3D *in-silico* images of the same system in OSM mode (produced by reducing the thickness of excitation $P_{n'}$ to 1 nm) were used as ground truth. From **Figure D.12**, it is clear that the accuracy of Richardson-Lucy algorithm is high, and errors are minimal.

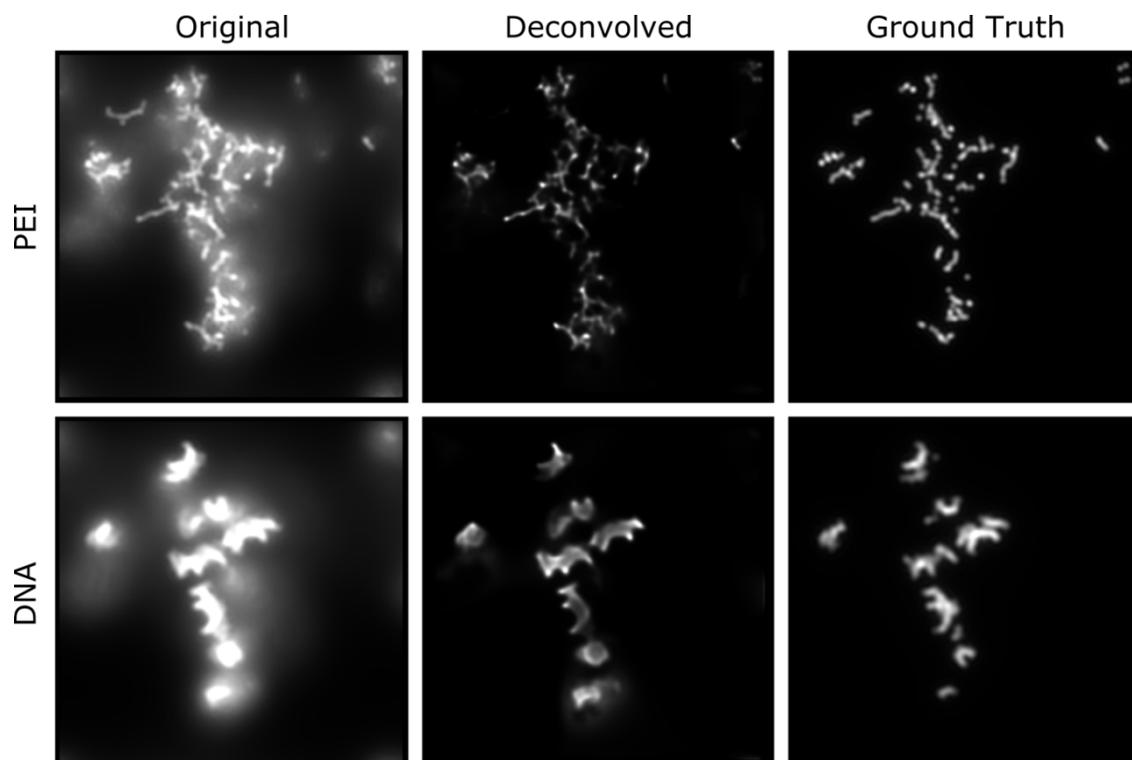


Figure D.12: Original 3D *in-silico* microscopy image is calculated using Gandy[12] PSF, $t = 2 \mu\text{s}$, $\psi = 59.4^\circ$, $n = z$, $\Delta n_o = 0.1 \text{ nm}$, $\Delta l' = \Delta m' = 0.1 \text{ nm}$, $\Delta n' = 0.05 \text{ nm}$, $P_{l'} = P_{m'} = P_{n'} = 25 \text{ nm}$, $f_s = 530$, $(\lambda, I_0) = (670 \text{ nm}, 0.13)$ for DNA and $(518 \text{ nm}, 0.27)$ for PEI. Deconvoluted 3D *in-silico* image is obtained using 150 iterations of Richardson-Lucy algorithm[26,27] with DeconvolutionLab2[28] in Fiji ImageJ[8]. The ground truth is the original 3D *in-silico* image calculated with $P_{n'} = 1 \text{ nm}$ (keeping other parameters the same). All the images are shown for $n_o = 7.8 \text{ nm}$; 3D images are not shown.

References

- [1] Hunter J.D. Matplotlib : A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, 9, 90–5.
- [2] Smith A.R. Color Gamut Transforma Pairs. *Proc. 5th Annu. Conf. Comput. Graph. Interact. Tech.* **1978**, 12–9.
- [3] Mahajan S., Tang T. Polyethylenimine-DNA Ratio Strongly Affects Their Nanoparticle Formation: A Large-Scale Coarse-Grained Molecular Dynamics Study. *J. Phys. Chem. B* **2019**, 123, 9629–40.
- [4] Abraham M.J., Murtola T., Schulz R., Páll S., Smith J.C., Hess B., et al. GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism From Laptops to

- Supercomputers. *SoftwareX* **2015**, 1–2, 19–25.
- [5] Mahajan S., Tang T. Martini Coarse-Grained Model for Polyethylenimine. *J. Comput. Chem.* **2019**, 40, 607–18.
- [6] Miettinen M.S. Molecular Dynamics Simulation Trajectory of a Fully Hydrated DMPC Lipid Bilayer [Data set], <http://doi.org/10.5281/zenodo.51635>, **2013**.
- [7] Lindahl E., Hess B., van der Spoel D. GROMACS 3.0 : A Package for Molecular Simulation and Trajectory Analysis. *J. Mol. Model.* **2001**, 7, 306–17.
- [8] Schindelin J., Arganda-Carreras I., Frise E., Kaynig V., Longair M., Pietzsch T., et al. Fiji: An Open-Source Platform for Biological-Image Analysis. *Nat. Methods* **2012**, 9, 676–82.
- [9] Tinevez J.Y., Perry N., Schindelin J., Hoopes G.M., Reynolds G.D., Laplantine E., et al. Trackmate: An Open and Extensible Platform for Single-Particle Tracking. *Methods* **2017**, 115, 80–90.
- [10] Cloizeaux J. Des, Jannink G. Standard Continuous Model and Perturbation Calculations. In *Polymers in Solution Their Modelling and Structure*. New York, NY: Oxford University Press **1991**, 354–430.
- [11] Gibson S.F., Lanni F. Experimental Test of an Analytical Model of Aberration in an Oil-Immersion Objective Lens Used in Three-Dimensional Light Microscopy. *J. Opt. Soc. Am. A* **1992**, 9, 154–66.
- [12] Gandy R.O. Out-of-Focus Diffraction Patterns for Microscope Objectives. *Proc. Phys. Soc. B* **1954**, 67, 825–31.
- [13] Haeberlé O. Focusing of Light Through a Stratified Medium: A Practical Approach for Computing Microscope Point Spread Functions. Part I: Conventional Microscopy. *Opt. Commun.* **2003**, 216, 55–63.
- [14] Hell S., Reiner G., Cremer C., Stelzer E.H.K. Aberrations in Confocal Fluorescence Microscopy Induced by Mismatches in Refractive Index. *J. Microsc.* **1993**, 169, 391–405.
- [15] Richards B., Wolf E. Electromagnetic Diffraction in Optical Systems, II. Structure of the Image Field in an Aplanatic System. *Proc. R. Soc. London A* **1959**, 253, 358–79.
- [16] Sun C., Tang T., Uludağ H., Cuervo J.E. Molecular Dynamics Simulations of DNA/PEI Complexes: Effect of PEI Branching and Protonation State. *Biophys. J.* **2011**, 100, 2754–63.
- [17] Uusitalo J.J., Ingólfsson H.I., Akhshi P., Tieleman D.P., Marrink S.J. Martini Coarse-Grained Force Field: Extension to DNA. *J. Chem. Theory Comput.* **2015**, 11, 3932–45.

- [18] Schaffer D. V., Fidelman N.A., Dan N., Lauffenburger D.A. Vector Unpacking as a Potential Barrier for Receptor-Mediated Polyplex Gene Delivery. *Biotechnol. Bioeng.* **2000**, 67, 598–606.
- [19] Viénot F., Brettel H., Ott L., M'Barek A.B., Mollon J.D. What Do Colour-Blind People See? *Nature* **1995**, 376, 127–8.
- [20] Simpson J.C., Joggerst B., Laketa V., Verissimo F., Cetin C., Erfle H., et al. Genome-Wide RNAi Screening Identifies Human Proteins With a Regulatory Function in the Early Secretory Pathway. *Nat. Cell Biol.* **2012**, 14, 764–74.
- [21] Ergonomic requirements for office work with visual display terminals (VDTs) — Part 3: Visual display requirements – Amendment 1, ISO 9241-3:1992/AMD 1:2000. *Geneva, Switzerland: ISO* **1992**.
- [22] American National Standard for Human Factors Engineering of Visual Display Terminal Workstations, Section 6, 17-20, ANSI/HFS 100-1988. *New York, United States* **1988**.
- [23] Anderson M., Motta R., Chandrasekar S., Stokes M. Proposal for a Standard Default Color Space for the Internet—sRGB. *Color Imaging Conf.* **1996**, 238–45.
- [24] Colour Measurement and Management in Multimedia Systems and Equipment - Part 2.1: Default Colour Space - sRGB, IEC/4WD 61966-2-1. *Geneva, Switzerland: IEC* **1998**.
- [25] Lanza A., Morigi S., Sgallari F., Wen Y.W. Image Restoration With Poisson-Gaussian Mixed Noise. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* **2014**, 2, 12–24.
- [26] Richardson W.H. Bayesian-Based Iterative Method of Image Restoration. *J. Opt. Soc. Am.* **1972**, 62, 55–9.
- [27] Lucy L.B. An Iterative Technique for the Rectification of Observed Distributions. *Astron. J.* **1974**, 79, 745–54.
- [28] Sage D., Donati L., Soulez F., Fortun D., Schmit G., Seitz A., et al. DeconvolutionLab2: An Open-Source Software for Deconvolution Microscopy. *Methods* **2017**, 115, 28–41.