NORTHWESTERN UNIVERSITY

# Molecular Dynamics Study of Charged Nanomaterials: Electrostatics and Self-assembly

A THESIS

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Applied Physics

By

Yaohua Li

EVANSTON, ILLINOIS

January 2021

This dissertation is dedicated to those who fought to save lives during the Covid-19 global pandemic.

# ABSTRACT

Molecular Dynamics Study of Charged Nanomaterials: Electrostatics and Self-assembly

Yaohua Li

Although electrostatics interactions in fluids have been studied for many decades, new results in this field are still challenge our classical understanding of electrolytes. The combination of electrostatics and self-assembly yields many interesting yet challenging problems that are of fundamental scientific interest and show promise for industrial applications. In this dissertation, I introduce our work on the topic of charged nanomaterials in aqueous salt solutions and how electrostatics play a role in different systems. The work summarized here is an attempt to develop methods to correctly model nanoscale charged systems in both low and high salt environments, and tackle the problem of simulating large systems with MD simulations by using coarse-graining techniques. First, we study nanoparticles immersed in concentrated monovalent salt (>0.5 mol/L[1]) using multi-scale molecular dynamics (MD) simulations involving atomic resolution and coarse-grained representations with implicit solvent. We find a surprising attractive to repulsive and then attractive re-entrant behavior as a function of salt concentration that cannot be predicted by previous theories, and propose a rational explanation. Next, we explore the interaction

[1]the unit mol/L is also written as M for short

of cylindrical interfaces in NaCl soluitons to find the screening length of charged cylinders and compare with the prediction of Poisson-Boltzmann equation. We also find a depletion attraction between cylinders at high monovalent salt concentrations. We compare the results of MD simulations to mean-field theories as well as liquid state theory that incorporates ion correlations, and we show that the ion short-range correlations significantly impact the interactions between cylinders in concentrated monovalent salt solutions. Finally, we look into the complex biological system of bacterial microcompartment (MCP) assembly. Using all atom (AA, explicit water and ion) and coarse-grained (CG, implicit ion) MD simulations, combined with thermodynamics analysis, we find that electrostatic interactions (hydrogen bonds and charge distributions) play an important role in the self-assembly of native propanediol utilization (Pdu) MCPs. Combining AA and CG MD simulations, we predict various polyhedral and extended assembly shapes, and we predict what kinds of mutations lead to the success or failure of MCP assembly. The simulation and theoretical predictions match with the experimental observation of our collaborators and with published experiments.

Thesis Advisor: Prof. Monica Olvera de la Cruz

# Table of Contents

## Acknowledgements

I would like to express my deepest gratitude to many people who have offered great support throughout my doctoral studies. First of all, I must thank my advisor Prof. Monica Olvera de la Cruz for granting me the opportunity to work in the Olvera group, and her continuous support and guidance to my research career. Although we are a big group of over 20 people, Prof. Olvera de la Cruz devotes a great amount of time and attention to working with us and shows care to every single person in the group. The insightful ideas she pointed out during our discussion gave me a good direction of problem to study. Next I would love to thank Prof. Pulak Dutta and Prof. John Marko for being my committee member, and giving very helpful suggestions during my prospectus and beyond. I remember running into Prof. Marko and Prof. Dutta around campus and chatting briefly, about science and life. I wish I had more chances to do so.

During my research, many colleges have given important help and suggestions. I specially thank Dr. Martin Girard and Dr. Jaime A. Millan for teaching me MD simulations, Dr. Meng Shen, Dr. Felipe Jimenez, Dr. Baofu Qiao and Dr. Kyle Hoffmann for very helpful discussions and teaching me all-atom MD techniques, Dr. Trung D. Nguyen for a lot of help in programming and science. Also, my work and life would be far less joyful and fruitful without the companion and support of many colleges in the Olvera group and other friends at Northwestern. Special thanks to Curt Waltman and Ali Ehlen for

improving the English of this document.

I would like to thank my family, including my parents and my beloved wife in China. Despite the huge physical distance between us, we always stay connected and they have unconditionally supported me with love and care. The year 2020 has been a difficult year for everyone, I would like to express special thanks to those who risk their lives fighting the virus so that many of us can be safe and healthy.

This work would not have been possible without the support of Sherman Fairchild Foundation, Department of Energy Award DE-FG02-08ER46539 and Center for Computation & Theory of Soft Materials (CCTSM) funding support.

## List of Abbreviations

AA All-atom

bp base pair (DNA)

BMC Bacterial Microcompartment

BMV Bacterial Microcompartment Vertex

CG Coarse-grained

DLVO Derjaguin-Landau-Vervey-Overbeek

GFP Green Florescent Protein IBI Iterative Boltzmann Inversion

LJ Lenard-Jones

LS Liquid State

MCP microcompartment

MC Monte Carlo

MD Molecular Dynamics

NP Nanoparticle

NPT Constant Particle number, Pressure and Temperature

NVT Constant Particle number, Volume and Temperature

PB Poisson-Boltzmann

Pdu 1,2-propanediol utilization

PDB Protein Data Bank

PM Primitive Model

PMF Potential of Mean Force

CHAPTER 1

# Introduction

In aqueous environments, from living matter to micro or nanoscale materials, charged interfaces are ubiquitous. Most naturally occurring charged interfaces carry charge due to the ubiquity of weak acids and bases in biology. For example, each monomer of Deoxyribonucleic acid (DNA) carries 1 electron charge located on its phosphate group [**1**], resulting in a linear charge density of 5.88 e/nm. These charged interfaces have a non-trivial impact on the physical and chemical processes happening near them, such as the aggregation or solvation of DNA and proteins. Understanding the behavior of charged particles such as nanoparticles and ions near solid-liquid interfaces is of great interest both for fundamental science and industrial applications such as electrodes and capacitors [**2, 3**], and environmental applications such as extracting chemicals using surface adsorption [**4**]. To describe these processes using a theoretical framework, one has to properly take into account the role of ions, including the counterions and additional salt in the solution. Until recently, the most widely used theories for electrolytes were mean-field theories based on the Poisson-Boltzmann (PB) equation[**5, 6**], which for monovalent salt writes:

$$\nabla^2\psi = 2c_0 e \sinh(\frac{e\psi(\boldsymbol{r})}{k_B T}) \tag{1.1}$$

Where $\phi$ is the electrostatic potential, $e$ being the absolute charge of an electron, $c_0$ is the bulk ion concentration; or its linearized version $\nabla^2\phi(\boldsymbol{r}) = \kappa^2\phi(\boldsymbol{r})$ where $\kappa^2 = \frac{4\pi\beta}{\varepsilon}\sum_s q_s^2 c_s$, $q_s$ and $c_s$ are the charge and concentration for ion species s respectively. The solvent (water) is treated as a continuum with one parameter, the dielectric constant $\varepsilon_r$. The well known DLVO (named after Boris Derjaguin and Lev Landau, Evert Verwey, and Theodoor Overbeek) theory is based on the PB equation. The Gouy-Chapman theory applies PB equation to uniformly charged interfaces and introduces the concept of diffused double layers. The ion hard core can be partially accounted for by applying the fact that ions cannot get a closer distance to the wall than the ion radius, which is called modified Gouy-Chapman theory[**6, 5**]. These mean-field methods achieved some success in the regime of low surface charge and low ion concentration[**7, 8**] and even high charge and low ion concentration [**9**]. But as the concentration increases, these theories inevitably break down, because ions have a finite size and interact with the interface and each other via excluded volume interactions[**10**], which are not taken into account in mean-field theories. Based on the salt concentration, this thesis can be divided into 2 parts. Part I: interactions between nano-scale interfaces at high salt concentration, where ion correlations are important and explicit-ion models or methods are required. This part includes Chapter 2 and 3. Part II studies the interactions and self-assembly of proteins at physiological conditions (about 100 to 300 millimolar monovalent salts[**11, 12**] and 300 K temperature), where the geometry and multi-component nature of biological systems add complexity to the problem. This part corresponds to Chapter 4

### 1.1. Interaction of Interfaces at High Salt

The Primitive Mode (PM), which depicts salt ions as hard spheres, gets one step further from mean-field approaches for spherically symmetric ions by including simplistic ion correlations[**13, 14, 15**]. The unrestricted (does not limit the size and charge of cation and anion to be equal) PM has 2 free parameters, the size of cations and anions. One of the difficulties of studying electrolytes with explicit ion is that the Coulomb interaction is a long-range interaction that cannot be simply summed with a cutoff. MD and Monte Carlo (MC) simulations of bulk systems usually take advantage of periodic boundary conditions and particle-mesh Ewald summation[**16**]. A lot of progress has been made towards a better understanding of electrolytes in the bulk or near interfaces by the PM, either used with MD or MC simulations, or integral equations such as Ostern-Zernike equation[**17**]. Just to list a few examples, hyper-netted chain theory was used to qualitatively predict the electrophoretic mobility of DNA[**17**]. Hypernetted chain calculations show that the electrostatic screening length for multivalent salts increases at high salt concentration, and such effects can be enhanced by the shape of the multivalent ions[**18**]. The interaction between charged rods in salt solutions can be calculated by solving the Ostern-Zernike equation with HNC/MSA closure self-consistently up to 1M of monovalent salt, and the result agrees well with MC simulations [**19**]. These results show that with the exception of detailed properties of solvent structure, such as the hydration shell around ions, integral equation methods can reach very similar results compared to MC or MD simulations. However, although quantitatively more accurate, all-atom (AA) MD simulation is rarely compared to the results made by PM and there is little knowledge of what error PM has, and to what extent it should be compared with experiments.

While physical quantities such as ionic activity and osmotic coefficients can be fitted to experimental data by tuning the hard-sphere sizes [**15**], it is conceptually straightforward to see that this method does not extend well to the very high concentration regime where ions are crowded. The physical meaning of the hard-sphere radius in PM usually corresponds to the hydrated ion radius. However, it is not hard to imagine that as the concentration increases, the structure of the hydration shell changes, as the ion-ion and ion-water correlation[**20, 21, 22**] change depending on the temperature and ion concentration. The pair correlation function predicted by the PM largely deviates from classical AA MD simulations [**23**]. At sufficiently high ion concentrations, the mean distance between ion decreases, and lots of the thermodynamic observable such as the osmotic pressure depends on the pair correlation function. Therefore, the PM cannot provide a satisfactory description of concentrated salt solutions.

During my doctoral work, I seek an improved theoretical or simulational framework to describe the interaction between nanoscale charged objects in aqueous salt solutions, and determine when the available tools such as the PB equation, the PM, and other tools can be used as a reasonably good approximation. In particular, I study two problems for this aim:

(1) The interaction of charged/uncharged nanoparticles (NP) in NaCl solution

(2) The interaction of cylindrical charged surfaces in NaCl solution

In problem (1), I study nanoparticles with an attempt to explain recent experiments of like-charge attraction of DNA coated nanoparticles in monovalent and multivalent salt solutions[**24**]. In this study, using an empirical salt potential that preserves accurate

ion-ion pair correlations with implicit solvent, collaborators and I found 3 regimes of re-entrant behavior for the interaction between two like-charged NPs as the surface charge density increases, see Fig. 2.1. This work was published in [**21**] and described in chapter 2. Problem (2) is not only an extension of problem (1) to cylindrical geometry, but also facilitates the assessment of longer-range ion screening behavior because the screened electrostatic potential from charged cylinders decays more slowly than spheres. Also, liquid state theory is applied to this problem so that we can compare its results with MD simulations and separate the terms of steric interaction from electrostatic contributions. Cylinders are of particular interest because they are extensive in one dimension and curved in another, and we find that concentration of salt for the different interactions regimes found in problem (1) differs due to the change of geometry. Cylinders are also of practical interest because they are oftentimes used as a model for rigid polymers like DNA and nanostructures such as peptide amphiphiles. This work is in preparation for journal publication.

### 1.2. Self-assembly of Bacterial Microcompartments

Part II is a study of a multi-component biological system, in collaboration with Prof. Danielle Tullman-Ercek group in the Chemical and Biological Engineering department of Northwestern. Bacterial Microcompartments (MCP), which are made of protein shells and inner scaffolds (cargo), have gained increasing attention in recent years. They belong to the broader class of membraneless organelle, which breaks the traditional perception that all space inside the cell is divided by membranes[**25**]. These protein organelles are

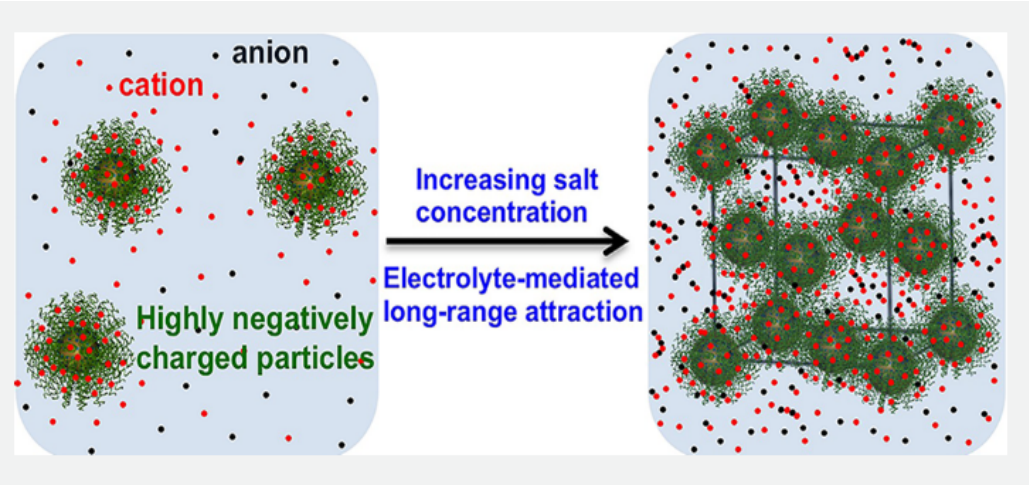



Figure 1.1. Small angle X-ray scattering (SAXS) experiments show that like-charged DNA coated NPs that are repulsive at low salt concentration experience long-range attraction at high salt concentrations, leading to crystallization and glass-like behavior. Reprinted from Kewalramani et al. ACS central science 2016[**24**]

found in various bacterial phyla and are postulated to help many of these organisms survive in hostile environments such as the gut of their hosts[**26**]. There are three essential components (shell proteins) that form the MCP, the hexagonal protein BMC-H, pentagonal protein BMC-P, and pseudo-hexagonal protein BMC-T (see Fig.1.2). These proteins can assemble into icosahedral or polyhedral compartments[**27**] or cylindrical tubes[**28, 27**] depending on the protein expression. AA and coarse-grained MD simulations were used to study the formation of a small part of MCP shell[**29**]. A 3-component model (hexamer, pentamer, and cargo) with spontaneous curvature was proposed and predicted the assembly of closed MCPs[**30**]. However, the transition of MCPs to cylinders was not predicted and the role of quasi-hexagonal BMC-T was not studied. Collaborators and I utilize AA and Coarse-grained (CG) MD simulation as well as thermodynamics to understand the process of polygonal proteins assembling into closed or extended structures, which sheds

light on how to control the morphology for engineering applications. We compare theory predictions with *in vivo* experiments and published results of MCPs and get qualitative or semi-quantitative agreement.

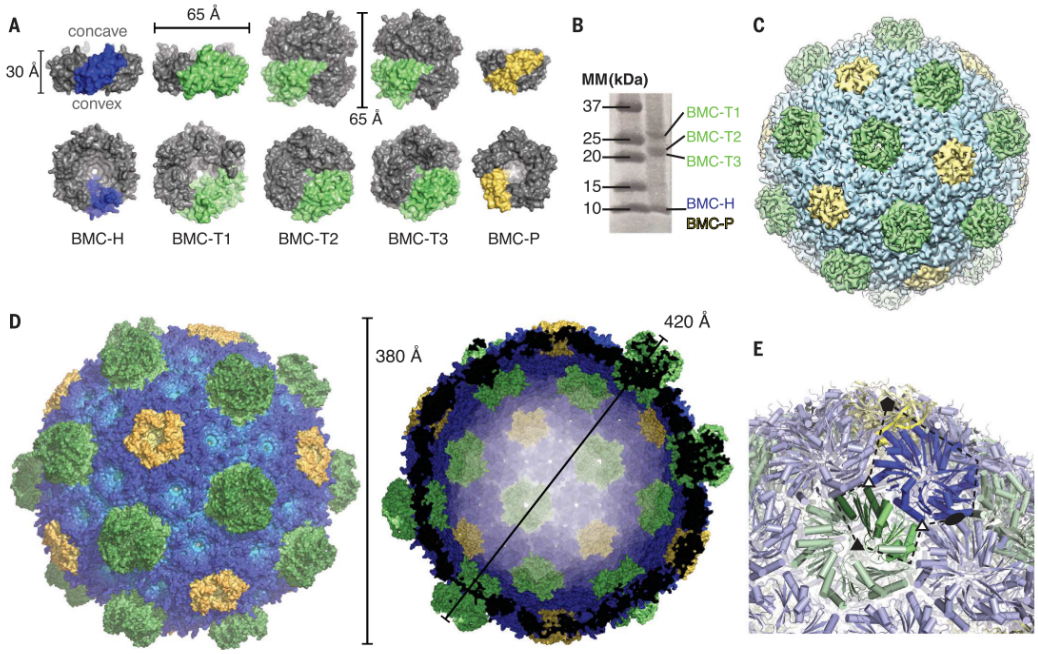



Figure 1.2. A small MCP reconstructed from cryo-electron microscopy (cryo-EM). (A) shows the 3 major component of shell proteins: hexagonal BMC-H, pseudo-hexagonal BMC-T (further divided into 3 types), and pentagonal BMC-P. 3 types of shell proteins tile into icosahedral shell that usually enclose enzyme as cargo (not shown in this example). Reprinted from M. Sutter et al. Science 2017 [**31**].

CHAPTER 2

# Strong Attractions and Repulsions by Monovalent Salt

Authors: Yaohua Li, Martin Girard, Meng Shen, Jaime Andres Millan and Monica Olvera de la Cruz

Author contributions: M.O.d.l.C. designed research; Y.L., M.G., and M.S. performed simulations; Y.L., M.G., M.S., J.A.M., and M.O.d.l.C. analyzed results; M.O.d.l.C. supervised research; and Y.L., M.G., M.S., J.A.M., and M.O.d.l.C. wrote the paper.
This work was published in [**32**].

## 2.1. Chapter Abstract

Controlling interactions between proteins and nanoparticles in electrolyte solutions is crucial for advancing biological sciences and biotechnology. The assembly of charged nanoparticles (NPs) and proteins in aqueous solutions can be directed by modifying the salt concentration. High concentrations of monovalent salt can induce the solubilization or crystallization of NPs and proteins. By using a multiscale coarse-grained MD approach, we show that, due to ionic correlations in the electrolyte, NPs pairs at high monovalent salt concentrations interact via remarkably strong long-range attractions or repulsions, which can be split into three regimes depending on the surface charge densities of the NPs. NPs with zero to low surface charge densities interact via a long-range attraction that is stronger and has a similar range to the depletion attraction induced by polymers with radii of gyrations comparable to the NP diameter. On the other hand, moderately

charged NPs with smooth surfaces as well as DNA-functionalized NPs with no possibility of hybridization between them interact via a strong repulsion of range and strength larger than the repulsion predicted by models that neglect ionic correlations, including the Derjaguin-Landau-Vervey-Overbeek (DLVO) model. Interactions between strongly charged NPs ($> 2e/nm^2$), including both smooth and DNA-functionalized NPs, show an attractive potential well at intermediate-to-high salt concentrations, which demonstrates that electrolytes can induce aggregation of strongly charged NPs. Our work provides an improved understanding of the role of ionic correlations in NP assembly and design rules to utilize the salting-out process to crystallize NPs.

## 2.2. Background

The interactions between charged nanoparticles in aqueous salt solutions bear an important role in biological science, soft-matter physics, and nanotechnology [**33, 34, 35, 36**]. High concentration salt has been used to manipulate interactions between particles or macromolecules. Specifically, proteins and nanoparticles (NPs) crystallize upon the addition of high concentration monovalent or multivalent salt, which is a phenomenon commonly referred to as "salting out" [**37, 38, 39**]. Recent experimental work (see Fig.1.1) demonstrates the attractive interaction between gold nanoparticles grafted with negatively charged DNA that has no possibility of forming hydrogen bonds, and therefore the interaction must be related to salt-induced interactions. The range of such a force, which extends to several nms, is much larger than the ion sizes, and therefore cannot be understood by condensed multivalent salts or ion bridging[**40, 41, 42, 43**]. A lot of our

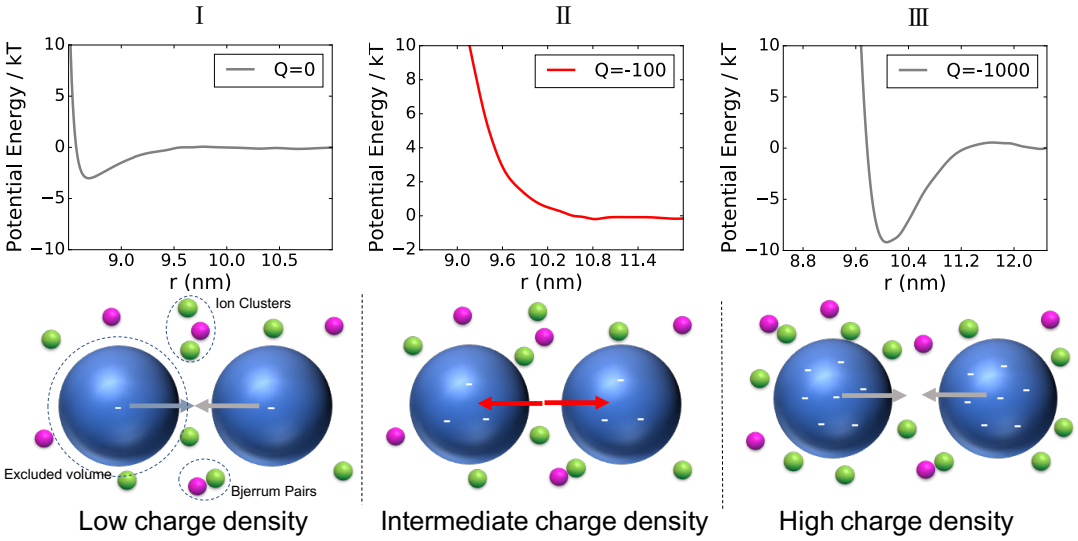



Figure 2.1. Three regimes of NP-NP interactions found in the simulations. In regime I, where the NPs have zero or low charge densities, the interaction is attractive due to depletion interactions mediated by the clustering of ions. In regime II, where the charge is sufficiently large, we find a repulsion that is stronger and longer range than the double-layer force predicted by the DLVO theory. In regime III, where the NPs have very high charge densities resulting in strong counterion condensation on the NP surfaces, the potential shows a strong short-range repulsion and a deep long-range attractive well. van der Waals interactions between NPs are not included in our model. Reprinted from Li et al. Ref. [**32**]

understandings about interactions in salt solutions rely on the DLVO (named after Boris Derjaguin and Lev Landau, Evert Verwey, and Theodoor Overbeek) theory and primitive model (PM) MD simulations. The DLVO theory, which explains the interaction by two terms: 1) screened Coulomb potential that provides repulsion, which keeps particles from aggregating at low concentration, and 2) dispersion attraction, which explains the aggregation upon increasing particle concentration. While this theory has the merit of a simple physical picture, it fails to explain the phenomenon that occurs at high salt concentrations or with multivalent salts. This is because it adopts a mean-field approach

for the electrostatic interactions, namely linearized PB equation:

$$\nabla^2\phi(\boldsymbol{r}) = \kappa^2\phi(\boldsymbol{r}) \tag{2.1}$$

This equation is also the base of Debye-Hückel theory for salt solutions. To derive this equation there are two assumptions, assuming the potential of mean force of ions $w(r) = q\phi(r)$, which neglect ion correlations; and linearization, which requires that the field cannot be large at any point in the solution. The first assumption breaks down at high salt concentration, where ions are frequently colliding, and the latter breaks down when there are strongly charged objects. On the other hand, the explicit salt primitive model (PM) sees ions as hard spherical particles with Coulomb potentials. It takes into account the discrete nature of ions, and therefore explains the ion correlations to some extent. Some previous theory and simulation work using the PM has been summarized in Chapter 1. However, when one tries to perform simulation at high salt concentrations, one has the dilemma of ion hydration sizes, as analyzed in the introduction1. The common practice is to take the hydration shell diameter of ions as the diameter of hard spheres. But as one increases the concentration, these spheres start overlapping well before the experimental saturating concentration. Therefore the primitive model does not describe high salts well either. To overcome the difficulties of these existing methods, a way to treat ion correlations in a consistent manner is needed.

There are several ways to coarse-grain an ionic solvent into an implicit-solvent, explicit ion model, including the Iterative Boltzmann Inversion (IBI), relative entropy method, and machine learning methods [**44, 45, 46**]. And the ion potential calculated by these methods is usually solved by Monte Carlo (MC) or MD simulations. The IBI is a method

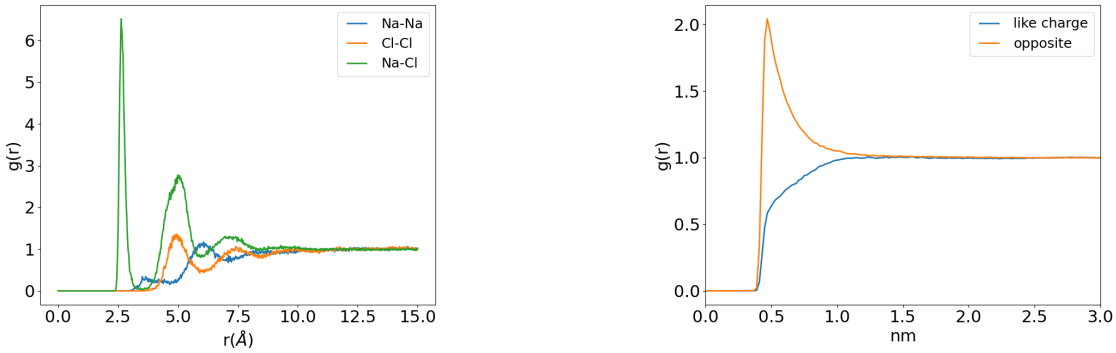



Figure 2.2. RDF of Primitive Model monovalent salt and Atomistic RDF. the PM gives rise to simple ion-ion RDF, where as AA RDF has much more features in the short range of 1nm

to calculate empirical potential from known correlation functions, whether calculated from full atom simulations or experiments. It has a clear physical meaning from statistical mechanics and is easy to implement. The relative entropy is slightly more general, minimizing the relative entropy $S_{rel} = \sum p_T \ln(p_T/p_M)$ optimizes a system to reproduce specific target system[**44**]. Machine learning methods have gained increasing popularity in recent years. While it has been shown to reproduce *ab initio* MD simulations efficiently, it is hard to relate to classical physical concepts such as pair additive potentials. For the problem of salt ion interactions, it is shown here that IBI calculated potential suffice to reproduce key physical quantities of interest, such as pair distribution functions of ions and bulk osmotic pressure. Therefore, for the sake of simplicity, I used the IBI method to derive ion-ion pair potentials from atomistically calculated pair correlation functions. The results are shown in Figure 2.5. Then, the ion-ion pair potentials are applied to two systems: simple smooth nanospheres and DNA grafted nanoparticles. From these studies, we learned that the interactions of like-charged nanoparticles have 3 distinct interaction

regimes (See Fig. 2.1, which differs drastically from the anticipated screened repulsion with monotonic screening length as a function of salt concentrations. Similar effects have been reported in primitive model simulations, but the effect here is much stronger. This suggests that we should look at the interactions of charged nanoparticles from a new, non-mean-field perspective.

Molecular Dynamics has been frequently applied to ionic solution systems [**47, 48**]. Given a Hamiltonian, MD solves the trajectory of particles by calculating the force and numerically integrating the equations of motion[**49**]. It is generally accepted that in the dilute regime (<100 mM) of monovalent salt concentration, the electrostatic interactions can be described to a reasonable accuracy using mean-field theory such as nonlinear or linearized PB equation. Simplified discrete models such as the Primitive Model (PM), which depict ions as hard or soft spheres [**50, 51, 52, 53**] can be applied to a higher concentration of several hundred milimolar (mM), lots of biochemical phenomena happen at those salt concentrations, where the steric effects of the ions comes into play. However, recent theoretical work[**18**] and our own simulations [**32**] reveal that mean-field descriptions or even the Primitive Model, which includes some simplistic steric effects, leads to unsatisfactory results because of their failure to account for ion correlations. The pair correlation function predicted by PM 2.2 has a single peak and exponential tail, whereas All-Atomistic [1] MD simulations and X-ray scattering experiments demonstrate a sophisticated g(r) with a much more significant peak and oscillatory tail. This is because the PM, together with a lot of implicit solvent CG MD simulations, only model ions as hard

[1]The terms All-Atom, Atomistic, and full atom are used interchangably, all meaning the same thing

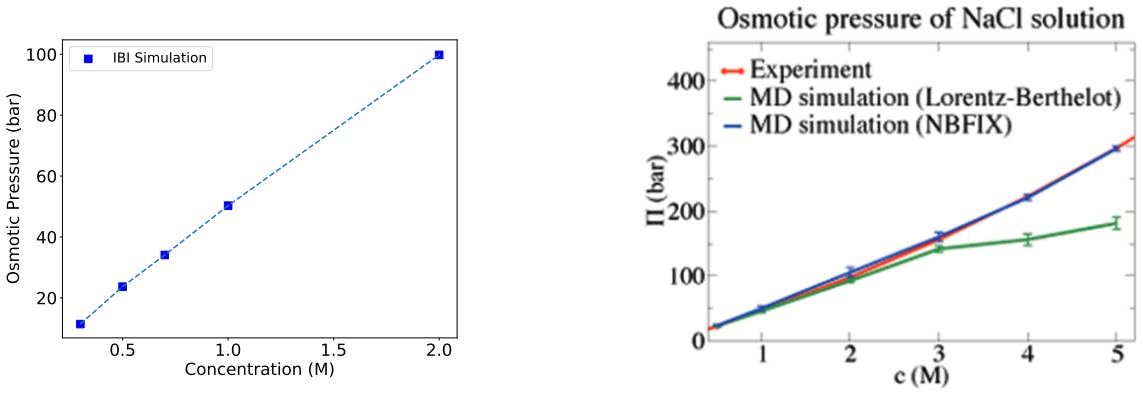



Figure 2.3. (a) Osmotic pressure calculated from IBI NaCl potentials (this work). (b) Osmotic pressure from previous published atomistic MD simulations and comparison with experiments. Reprinted from Luo et al. JCP 2010 [**57**].

spheres, and neglect detailed interaction of the solvent (water). Yet in more detailed studies, water is known to form a hydration layer around ions or any interface in contact with water [**54**]. Bearing in mind that implicit solvent model is an efficient way to simplify theory or computer simulations, because most of the time molecular details about water is irrelavant to the physical quantity of interest, we seek to formulate a method to describe concentrated electrolyte using implicit solvent while preserving the ion-ion correlations. Here we investigate the interaction of spherical nanoparticles in NaCl solution, which is very commonly found in biological and chemical labs as well as in nature, both *in vivo* and *ex vivo* [**55, 56, 35, 40**]. The problem of nano rods in salt solution will be investigated in chapter 3, where the difference in geometry leads to results that are similar qualitatively but different in several ways.

## 2.3. Methods

We first studied NaCl solution by All-Atomistic (explicit-water) simulations. A periodic box with 0.3M-1M water is simulated in a constant temperature, constant pressure canonical ensemble (NPT). We use the SPC/E model for ions and Tip3P water model. A cubic box with 5nm side length is initialized, so the box contain 3575 to 4131 depending on the salt concentration. For small concentrations, the box size is increased so there are at least 30 ion pairs in the box. The ion-ion pair correlation function (g(r)) is calculated directly by counting number of atoms within radius r from the simulation trajectory using VMD [**58**]. Also calculated is the osmotic pressure of the NaCl solution(Fig. 2.3). The relative dielectric constant, or dielectric permitivity $\epsilon_r(c)$(c is the salt concentration, see Fig. 2.4) of water is calculated at different NaCl concentrations by a statistical average of water dipole moments [**34**]. Note that the tip3p water model consistantly overestimate the dielectric constant compared to experimental measurements, wile the SPC/E model underestimate it. But both models give qualitatively similar trends. This $\epsilon_r(c)$ calculation would prove useful in the coarse-grain model when we replace the water with a continuum. We derive the short-range interaction between CG ions following the Iterative Boltzmann inversion (IBI) scheme described in [**45**] (see fig. 2.6 (d)). The short-range part of the effective potential between ions are represented by a table. An initial guess potential is needed, in our case we use the Boltzmann inversion of AA RDF $g_AA(r)$, and perform CG MD simulation of pure NaCl. The RDF between 3 combination of NaCl ions are calculated from the CG MD trajectory, and modified iteratively by equation

$$U_{i+1} = U_i(r) - k_B T log(\frac{g_i(r)}{g_{AA}(r)}) \tag{2.2}$$

$g_{AA}(r)$ stands for the RDF calculated in all atomistic simulations. These short-ranged represent the physical force of the hydration shell and repulsion from ion electron cloud. They are applied in addition to a long-range Coulomb force with an aforementioned effective dielectric constant calculated from AA simulations, so that the g(r) match with AA consistently.

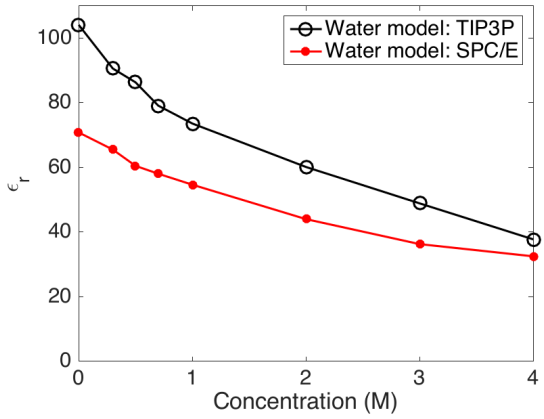



Figure 2.4. Dielectric constant value vs. NaCl concentration using the TIP3P water model (open circles) and SPC/E water model (filled circles). Reprinted from the SI appendix of [**32**]

With the CG salt potential and dielectric constant calculated from atomistic simulations, the interactions of two charged or neutral NPs are studied. Two NPs are placed symmetrically in a tetragonal simulaiton box, as shown in Fig. 2.6. The total force on each of the two NPs are measured at 60 different distances, each distance for $10^3$ times at intervals of 200 timesteps (450 ps). The force data points are fitted to a spline curve to reduce statistical noise, and then integrated to give the potential of mean force between two NPs. Each simulation is performed 3 times with different randomly generated initial states, and the results are the average of all 3 runs. The resulting mean force are shown in 2.6 (a), and the potentials integrated from the force curves, plus a hard core repulsion

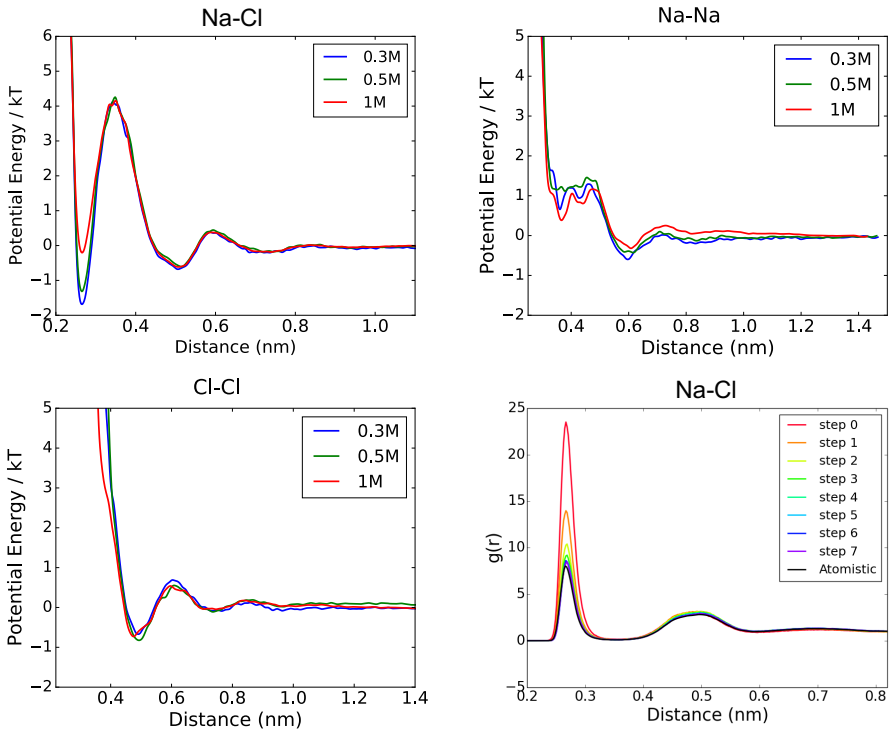



Figure 2.5. (A)-(C)Empirical potentials for NaCl ions calculated from IBI. (D) The Na-Cl radial distribution function of different iteration steps during the IBI calculaiton. Reprinted from[**32**]

at 8.5 nm are shown in (b) and (c).

## 2.4. Results and Discussions

Firstly, for neutral NPs in NaCl, as the salt concentration increases, an attractive interaction occurs. This attraction is negligible below 0.3M salt concentration, and increases nonlinearly above some threshold between 0.3 and 0.5 M (Fig. 2.6). Assuming that the attractive force originates from the depletion interaction[**59, 60**] of correlated ions, ion cluster analysis is preformed using the code by Dr. Trung Dac Nguyen [**61**]

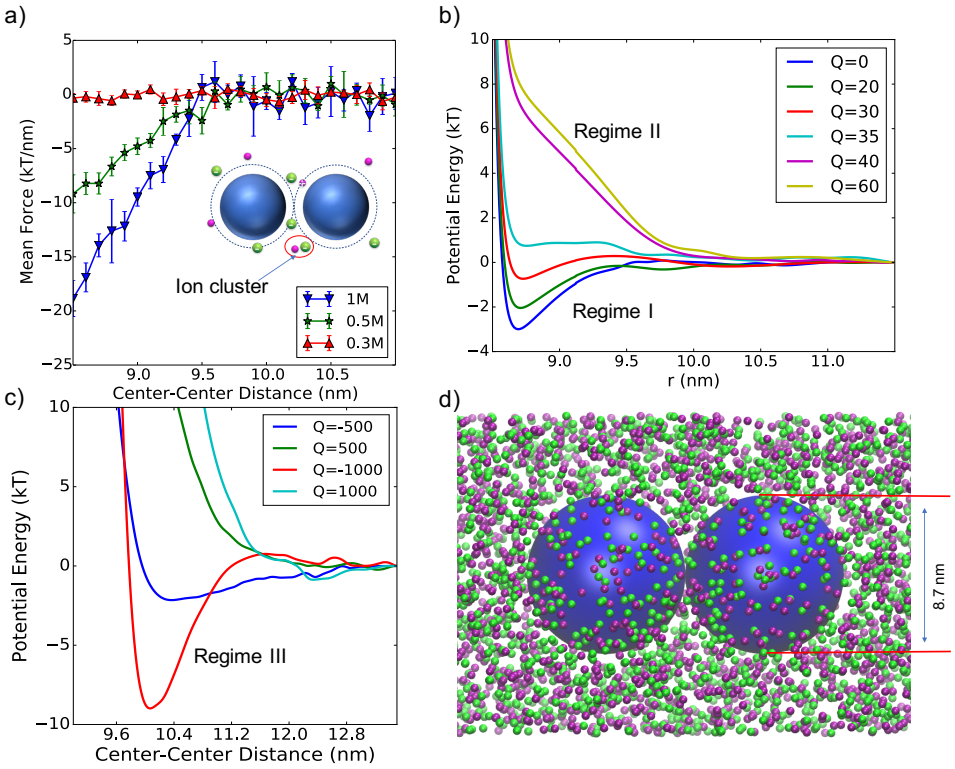



Figure 2.6. Interactions between NPs in NaCl solutions. (a) Average force between two electroneutral NPs in solutions with different NaCl concentrations. (b and c) NP Interaction potentials in 0.5 M NaCl. NPs have effective excluded volume diameter of 8.7 nm. The charge Q are in units of e. As the NP surface charge density increases, the interaction changes from attractive to repulsive. For very highly charged NPs, a deep, wide attractive potential well is found for negatively charged NPs. (d) Snapshot of MD simulation probing the interaction between two NPs immersed in 0.5 M NaCl salt. Small purple and green beads represent sodium and chloride ions, respectively. Large blue spheres represent the NPs. Reprinted from [**32**]

(Fig. 2.7). The analysis shows that at high salt concentration with our IBI salt potential model, the population of larger clusters is increased. These ion clusters, within their finite lifetime, behave as short chains and lead to depletion force between NPs [**59**]. The range of attractive interaction ( 1nm) also matches with the length scale of dimers and trimers,

which accounts for the majority of ion clusters.

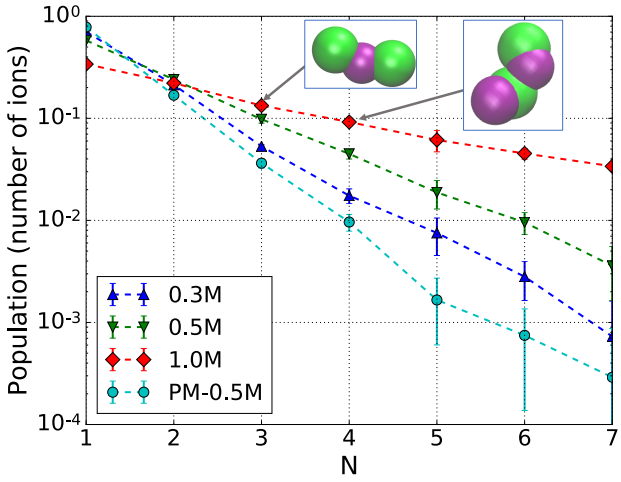



Figure 2.7. Cluster-size distribution in NaCl solutions. N denotes the number of ions in a cluster regardless of ion type, as a measure of cluster size. Dashed lines are guide to the eyes. Simulation results for the Primitive Model show a much lower number of clusters of ions. It is interesting to note that the population of even numered clusters are not significantly larger than odd numbers, which are not charge neutral. This indicates short-range interactions other than electrostatic interaction play an important role in the clustering behavior. Examples of ion timers (n = 3) and tetramers (n = 4) found in simulation are shown in the inset. Reprinted from [**32**]

For charged NPs in NaCl, more interesting types of interactions can occur. For weakly charged NPs, depletion attractions still dominate the interaction except for a week repulsion tail at large distances. Above some threshold charge density, which is dependent on the salt concentration, the interaction rapidly transition to a strongly repulsive potential (see Fig.2.6 (b)). In this regime, the range and strength of repulsion exceed the values predicted by the DLVO theory (see a comparison in Fig. 2.9), even with the correction of decreased water dielectric constant as the concentration of salt increases (Fig. 2.4) [**62**].

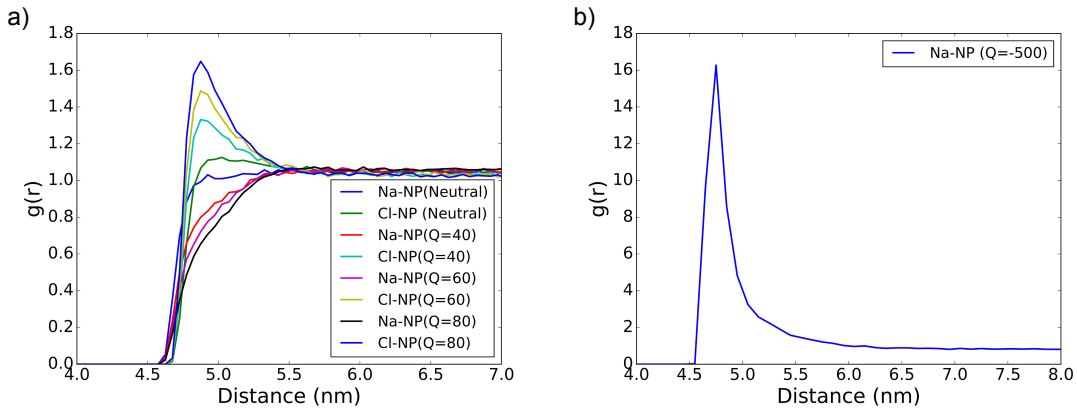



Figure 2.8. Ion distribution around NPs of different charge. The distribution of highly charged NP (Q=-500 e) is shown separatedly in (b)

This happens due to two mechanisms. First, the formation of ion clusters reduces the number of free ions in the solution, leading to a weaker screening of like-charge repulsions. Second, for strongly charged NPs, the depletion of ions in the middle gap breaks down because the electrostatic potential from NPs overwhelms thermo energy and the association of ion clusters.

We are interested in predicting the self-assembly of these charged NPs in salt solution. However, in the PMF calculation shown above, only 2 NPs are simulated, corresponding to the dilute limit of NPs. In order to test the effectiveness of the calculated potential of mean force in many-body circumstances, we introduce a 3 rd NP in our simulation so that three NPs form the vertices of an equilateral triangle. The total force and corresponding potential along one edge is shown in Fig. 2.10. At low surface charge density of NP, similar behavior of the effective force is maintained in the presence of multibody effect. However, at high surface charge density, the mean force calculated from 3 bodies is not linear-additive of 2-body calculations.

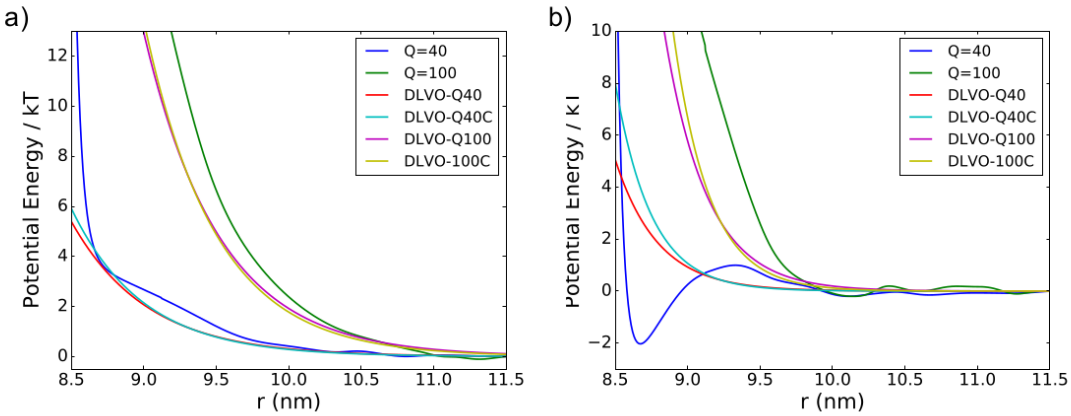



Figure 2.9. Comparison of simulated effective potentials between two NPs at salt concentrations of (A) 0.3 M and (B) 1 M with DLVO theory, and DLVO theory with corrected dielectric constants that depends on the ion concentration (denoted in the labels by DLVO-Q40C and DLVO-Q100C). The numbers following Q indicate the total charge on each NP, in the unit of elementary charges. All of the curves are plotted without van der Waals attraction. At intermediate concencentrations (such as 0.3M), DLVO theory gives a reasonable match to simulations, slightly underestimating the repulsion. At high concentrations, the DLVO theory deviates from simulation results for both low-charge and high-charge NPs. Reprinted from [**32**] SI appendix

Polymer grafted surfaces can carry much more charge than smooth surfaces and demonstrate interesting behavior of polymer as well as its interaction with ions[**63, 64, 35, 65**]. Interestingly, NPs with extremely high charge density (>2 $e/nm^2$) can also have a long-range attraction (Fig. 2.6 (c)). This surface charge density can be achieved by Densely grafted DNA-Au NPs, which is one order of magnitude larger than the regime discussed above for smooth NPs. This re-entrant behavior can be understood by realizing that at this charge density, the counterions condense into a tight layer around the NP (Fig. 2.6). Therefore the NP and shell of counterion can be renormalized into a larger sphere with smaller total charge. The asymmetry for positive and negative NP probably

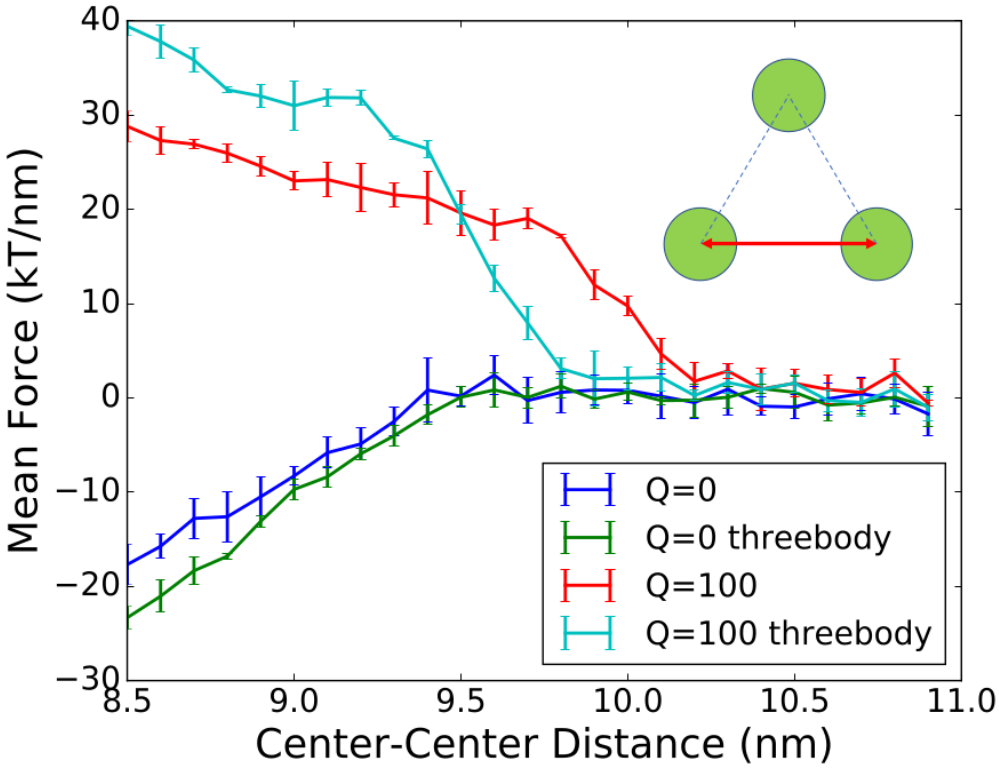



Figure 2.10. Simulated mean force along lateral direction from 3 NPs forming the vertices of an equilateral triangle (as shown in the inset), compared with two NP calculations (blue and red curves). For neutral NPs, the lateral force in three body occasion is roughly 1.5 times the force between pairs of NP, which is the value one would get assuming the ion-induced force is linearly additive. For NPs with a large charge (in this case 100 elementary charges), the mean force in 3-body simulations is stronger at short distances but then decrease faster.

originate from the different core size and hydration structure of sodium and chloride ions. One can understand this transition using a simple lattice model for the ion condensation. Consider there are $m$ counterions condensed onto the spherical surface, the Helmholtz

free energy of condensed ion is:

$$F = U_e + U_{ion} - k_B T \ln \Omega \tag{2.3}$$

$U_e = k_B T Z l_B / R$ is the electrostatic energy between ions and the NP. $\Omega = [\binom{N}{m}]$ is the possible numeber of configurations for the lattice model. The exact number of $m$ can be determined by minimizing the free energy with respect to $m$. We know that $m \sim Z$ where $Z$ is the valence of the NP because of electroneutrality condition. $\binom{N}{m} = \frac{N!}{m!(N-m)!}$. Using Stirlin's approximation,

$$F_{condense} = U_e + U_{ion} - [N \ln N - m \ln m - (N-m) \ln(N-m)] \tag{2.4}$$

the free energy difference by condensation of an ion from the solution is

$$\Delta F = F_{condense} - F_{solution} \tag{2.5}$$

We make a coarse estimate by assumming the total number of sites $N = \frac{4\pi R^2}{\pi a^2}$, a being the ion radii. Notice that the first term $U_e$ is the same for both types of ions (the difference in ion dimeter is much smaller compared to the NP radius). We do not have an expression for $F_{solution}$ since the ions experience long-range Coulomb potential, but if we assume that the solution free energy for the cations and anions do not differ significantly, and take the radius of ion to be 0.175 nm for Na and 0.21 nm for Cl, the free energy of condensation for Na is about 500 $k_B T$ lower than Cl. Therefore, in this case ion condensation happens for sodium but not for chloride.

To compare the smooth NP results with a more realistic model, we computed the potential interaction between non-complementary DNA-grafted NPs [**24, 35**] with a multiscale MD model that includes the DNA chains explicitly [**66**]. The interaction potential fitted from simulation data are shown in Fig. 2.11 A and B. Two grafting densities of DNA chains (indicated in the figure legend) are studied. An exact comparison of charge density to the smooth-sphere model is not meaningful because the charge is spread along the DNA strands (Fig. 2.11 D). Therefore, the ion distribution is mediated by the conformation of the DNA chains as well as electrostatic potentials originating from the chains, which extends for a few nanometers. Both the geometry and charge density on the NP surfaces can have a nontrivial impact on the potential of mean force. Nevertheless, a potential minimum is discovered in DNA-grafted NPs at intermediate distance ($\sim$34 nm from NP center) which is also reproduced in 8.7-nm smooth NP simulations at similar charge density (2.2 e/$nm^2$) (Fig. 2.6 C). In the case of higher DNA-grafting density, the increased charge density leads to a stronger repulsion, overcoming the potential minimum. Interestingly, the screening lengths fitted from explicit chain simulations demonstrate a decrease followed by an increase as the salt concentration increases, as reported in [**67, 68**]. We estimate the minimum of the screening length to be found at around 500 mM for NaCl in this system, but a similar trend is not found in the smooth sphere model because the exponential tail is below statistical noise. By analyzing the results in the smooth-sphere and DNA-functionalized NPs models, we find that in concentrated monovalent salts, ion correlations can lead to attractions not only between neutral or weakly charged NPs but also between strongly charged NPs at proper charge densities and salt concentrations. The problem of increased screening length at high monovalent salt concentration is a

highly debated subject [**69, 68, 70**], we continue the discussion in Chapter 3 by looking at the MD simulation results of both interface geometries, cylinder and sphere.

## 2.5. Conclusions and Outlook

By studying the interaction of NPs with empirical salt potential derived from AA MD simulations, we show that at high monovalent salt concentration, ion correlations are not negligible and leads to attractions and repulsions that are not expected by mean-field theories. Therefore, it is necessary to use proper CG methods to capture correct ion-ion correlations. The interplay between NP charge and correlated ions leads to interesting interaction behaviors that can be summarized into three regimes, as shown in Fig. 2.1. This method to calculate ion interaction using IBI and concentration-dependent dielectric constant is easy to impliment and captures important correlation effects at high salt concentration. It can be applied to CG simulation of multivalent ions where the association of ions is important but more detailed atomistic simulaitons are not affordable. The same scheme can be applied to the effective interaction between ions and other interfaces, the case of cylindrical interface is discussed in 3.

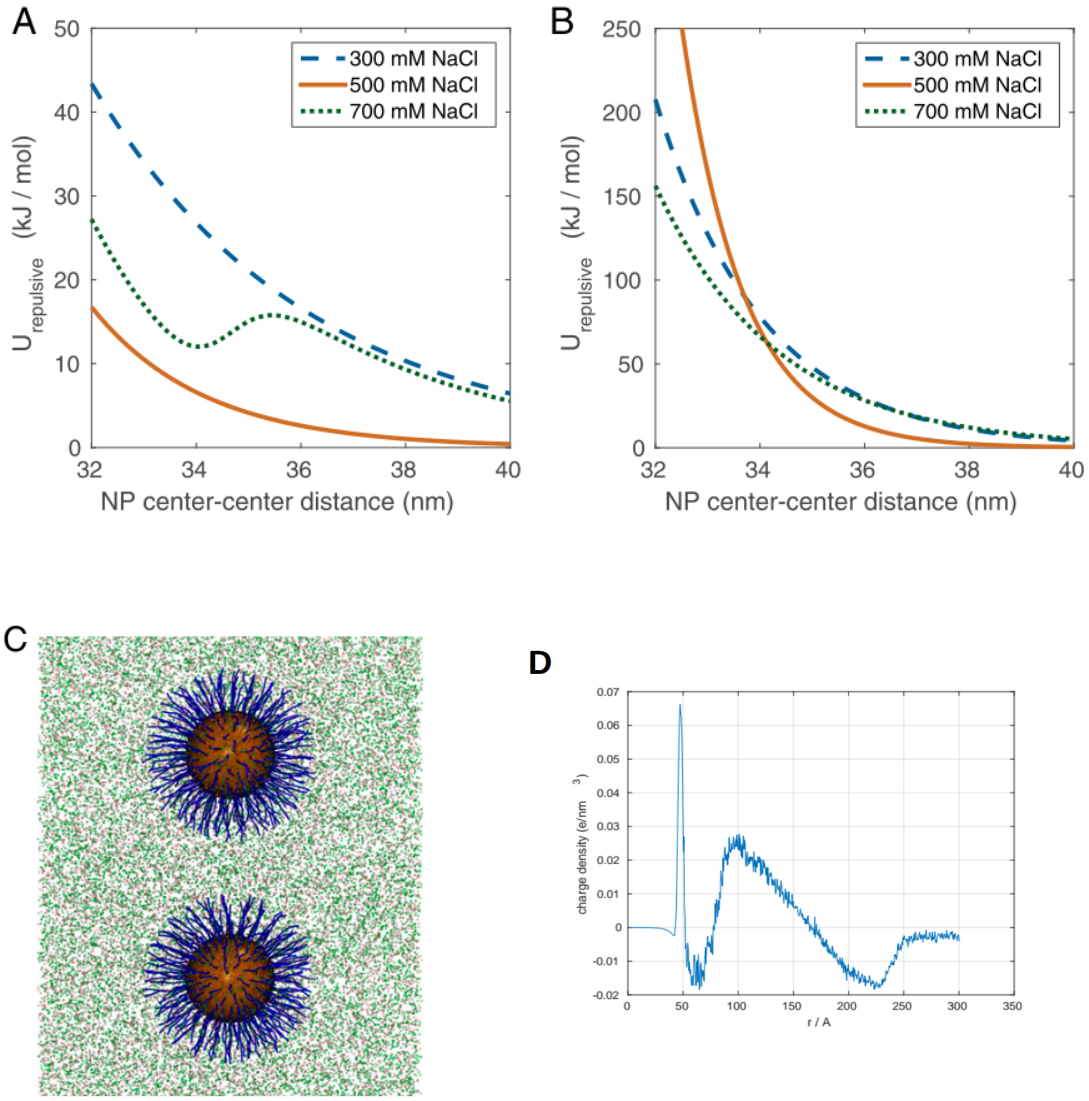



Figure 2.11. Potential of mean force between two DNA-coated NPs with (A) DNA surface density of 0.2 strands/nm$^2$ (total bare charge Q = -6,480 e) and (B) DNA surface density of 0.325 strands/nm$^2$ (Q = -110,580 e) for salt concentrations between 300 and 700 mM. The DNA strands consist of 20 base pairs (bp) of dsDNA and 6 bp of ssDNA. All curves are obtained from fits. In both cases the effective Debye length is nearly equal for 300 and 700 mM, going through a minimum at 500 mM. Onset of DNA overlap regime is around 35.0 nm. (C) Simulation snapshot for DNA-grafted gold NP interactions in NaCl solution. Ten percent of the actual ions are shown for visual clarity; water is treated as a continuous medium with a dielectric constant calculated from all atomistic simulations. The simulation box is larger than shown in the figure due to cropping. (A-C) image courtesy of Martin Girard. (D) The radial distribution of net charge around NPs.

CHAPTER 3

# Cylindrical Interfaces in Monovalent Salt Solutions

Authors: Yaohua Li, Felipe Jimenez-Angeles, Monica Olvera de la Cruz

Contributions: Y.L. performed the MD simulations, F.J.A. performed liquid state theory calculations. Y.L., F.J.A, M.O.d.l.C analyzed the data.

This work is in preparation for journal publication.

## 3.1. Chapter Abstract

The interaction between cylindrical rods is of interest in soft matter physics because they are often used to approximate the interaction between rigid linear molecules such as DNA or other nanostructures. Cylinders of various charge density in concentrated salt solution are studied using Molecular Dynamics (MD) and liquid state theory with HNC/MSA closure. We investigate two models for the ions in the solution, referred to as the restricted primitive model and a coarse-grain model built from all-atom MD simulations using the iterative Boltzmann inversion (IBI). Both models show that cylinders experience attractive interactions at or above concentrations of 2 mol/L (2M), regardless of their charge density. MD simulation and the liquid state calculations of the Primitive Model (PM) agree with these simulations up to 2M. Beyond 2M, PM shows a stronger attractive well compared to IBI due to the different ways the two models describe hydration shells. We show that at concentrations larger than 3M for NaCl solutions, more

detailed CG models such as IBI is required to accurately describe the behavior of NaCl ions.

## 3.2. Background

The curvature of solid-liquid interfaces plays a crucial role in their interaction with ions and subsequently, the interactions between these interfaces. Surfaces of cylindrical geometry have a unique combination of two properties: extensive in one dimension and curved in another dimension[**71**]. Charged cylinders have been frequently used as a simplified model for DNA as well as other nanostructures such as carbon nanotubes. Numerous studies have been done on the interactions of cylindrical objects in salt solutions [**72, 73**]. Most of the mean-field approaches work well in the weak electric field, low concentration regime, or with only counterions (no added salts). Monte Carlo or MD studies using the primitive model (PM), which depict the ions as hard spheres, have met some success in predicting correlation related effects in the high concentration (>100 mM) regime [**47**], but real ion correlations are more complex than the hard-sphere model in PM salts[**40, 74, 21, 18**]. Recently, Smith et al. discovered an anomalous increase in electrostatic screening length by measuring the force between cylindrical and salt activity coefficient[**67, 68**], and has raised considerable attention[**75, 76, 69**]. Recent all-atom simulation work found a much smaller increase in the screening length in bulk monovalent salt solution [**69**]. It might be helpful to study the screening length near cylindrical interfaces to figure out the origin of this discrepancy between experiments and theories or simulations.

Attractive interactions have been found experimentally for rods in both monovalent and

multivalent salts. While the attraction for cylinders in multivalent salt solutions has been successfully explained by the correlation effects of condensed counterions [**40, 73**], the aggregation of like-charged rods in monovalent salt is not well understood. There have been numerous studies on the interactions of spherical interfaces (colloids) and planes [**53, 77, 78**], in contrast, the interaction of cylinders in very high concentration salt is seldom studied. We aim to use MD simulaitons and the LS theory to explain phenomena such as salting out of rod-like molecules at the regime of very high ionic strength. Since the PB method has been shown to work well in the low ion concentration, low electric charge regime, we review the results of the PB equation as a starting point, and then look at the correlation effects neglected by such methods at high concentration.

Efforts to improve solutions based on the Poisson Boltzmann equation was pursued in the 1970s [**79**]. The PB equation predict that the ion correlations in solution decay in an exponential manner, with a characteristic screening parameter given by

$$\kappa_D^2 = \frac{\beta}{\varepsilon_r \varepsilon_0} \sum n_j q_j^2 \tag{3.1}$$

An asymptotic solution of the nonlinear PB equation in cylindrical and spherical geometry was derived in [**8**] series expansion form:

$$\begin{aligned} \psi = & \frac{e^{-X}}{X^{j/2}}(A_{00} + \frac{A_{10}}{X} + \frac{A_{20}}{X^2} + ...) \\ & + \frac{e^{-3X}}{X^{3/2}}(A_{01} + \frac{A_{11}}{X} + \frac{A_{21}}{X^2} + ...) \\ & + ... \end{aligned} \tag{3.2}$$

where $X$ is the dimensionless length $\kappa r$, and $A_{ij}$ are constants with the dimension of electrostatic potential. The coefficients $A_{ij}$ are determined by the boundary condition at the cylindrical interface, either by the surface charge density or the surface potentials. When $\kappa r >> 1$, the higher-order terms decay exponentially faster than lower-order terms(second rows and afterward) or by a higher power law (first row after the first term). For this solution to converge, $\kappa$ which scales as $\sqrt{c}$ has to be big enough. However, when $c$ is large, ions are close to one another and interact with steric effects, and the assumption that ion correlation is neglected cannot hold. This means that the PB equation, even the nonlinear form, has limited applicability. For appropriate ionic strength and at large distances, the series solution reduces to the cylindrical solution of linearized PB equation $A\frac{e^{-X}}{X^{j/2}}$.

Ion short-range correlations such as excluded volume interactions have been shown to play an important role in interactions in salt solutions, especially at high salt concentrations[**17, 21**]. Correlations caused by the finite excluded volumes of ions are not included in PB-based mean-field methods. Jimanez-Angeles et al. [**19**] study the interaction between two charged cylinders in 1M NaCl solution incorporating the PM ion correlations by means of the liquid state theory, and found that large size monovalent salt ions can introduce attractive force between cylinders. Here we present a study of cylinders from low to high monovalent salt concentrations (30 mM to 3 M), using a combination of multi-scale MD simulations and liquid state theory. We further explore whether the attractive interactions

found in [**19**] can be extrapolated to other concentrations and compare it to MD simulation using the CG potential derived from AA MD simulations as described in Chapter 2.

## 3.3. Methods

### 3.3.1. MD simulation Methods

We perform MD simulations with periodic boundary conditions (PBC) in all 3 directions, modeling an infinitely long cylinder. The box dimension perpendicular to the axial direction is taken large enough so that the ion density goes to the bulk value before reaching PBC box boundary. Two kinds of rod models are considered: rods with large spheres splayed along the axial direction (referred to as smooth rods, see Fig. 3.1) and cylindrical walls made of beads of 0.6 nm diameter (called rough rods). In real charged polymers like DNA, the charge is not uniformly distributed. Here we investigate the common features of interactions of rod-like molecules in high salt concentrations in general without considering sub-nanometer chemical details. These rods are immersed in a solution of explicit ions, implicit water with concentration-dependent dielectric constant calculated from atomistic MD simulations [**21**]. The interaction potentials between these ions are calculated from IBI so that they reproduce the pair correlation function from AAMD simulations. Since most thermodynamic quantities of our interest, such as osmotic pressure are largely decided by pair correlation functions [**80**](see Fig. 2.3), we expect this model to represent more accurate behavior of electrolyte compared to simplified models such as the primitive model [**13**]. In this work, we include a wider range of concentration than in Ref. [**21**] (0.03M-3M). The NaCl ions below 0.1M are described by the restricted

primitive model with a diameter of 0.47 nm, since IBI at this concentration is costly and unnecessary because ions have lower chances of being close to each other. More details of this model have been described in chapter 2. Single cylinders immersed in NaCl ions

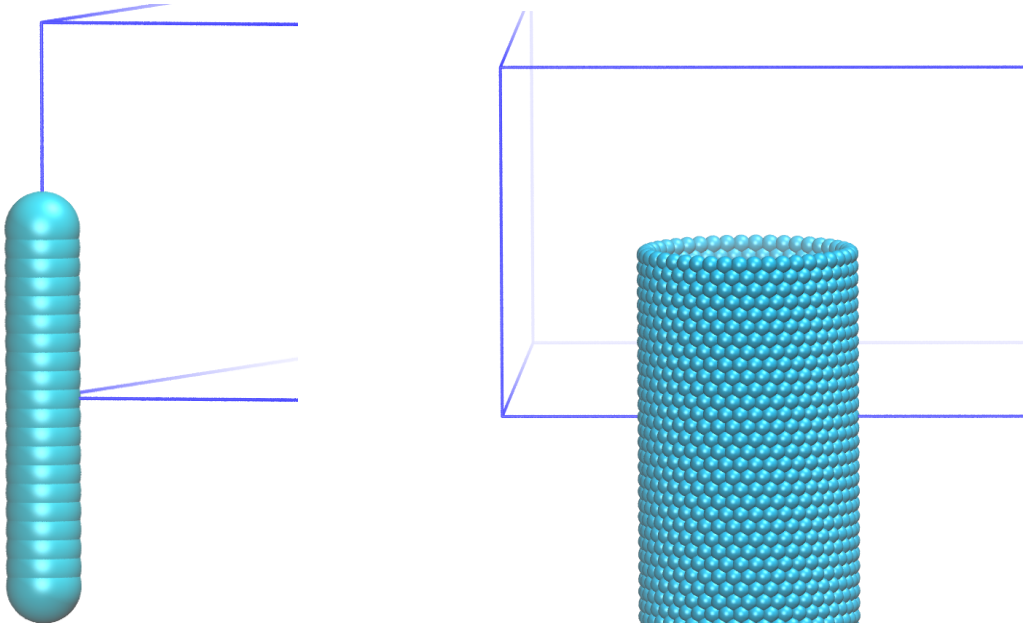


Figure 3.1. The two model cylinders studied in this work. The blue frame indicates the size of the periodic MD simulation box. (a) The smooth rod model (b) The rough rod model

are simulated for $1 \times 10^7$ MD steps with a reduced timestep of 0.01, and ion distribution is calculated using the equilibrium trajectory after $1 \times 10^6$ timesteps. The potential of mean force (PMF, $W(r)$) between the cylinders are calculated by fixing 2 cylinders at distance $r$, and recording the average force from all particles acting on each cylinder. At each distance, MD simulations of at least $4 \times 10^6$ steps are performed, and the total force is recorded at an interval of 200 steps to get a time average at the end. The mean force is numerically integrated to get the PMF.

### 3.3.2. Liquid State Theory

The liquid state theory method used in this work is the same as Ref. [**19**], and it is briefly summarize here. Consider 2 cylinders (species $\alpha$), the total correlation function can be related to direct correlation functions by the Ornstein-Zernike intergral equation:

$$h_{ij}(\boldsymbol{r_{12}}) = c_{ij}(\boldsymbol{r_{12}}) + \sum_{m=1}^{k} \int \rho_m h_{\alpha m}(\boldsymbol{r_{13}}) c_{mj}(\boldsymbol{r_{23}}) d\boldsymbol{r_3} \tag{3.3}$$

To solve Eq. 3.3, another relation between the direct correlation and total correlation is needed, namely the closure. For ions and the cylinder, the HNC closure is used:

$$log[g_{\alpha j}(\boldsymbol{r_{12}})] = -\beta u_{\alpha j}(\boldsymbol{r_{12}}) - c_{\alpha j}(\boldsymbol{r_{12}}) \tag{3.4}$$

and for ion-ion correlations, the mean spherical approximation (MSA) is applied:

$$c_{mj}(\boldsymbol{r_{23}}) = -\beta u_{mj}(\boldsymbol{r_{23}}) \tag{3.5}$$

for $|\boldsymbol{r_{23}}| > a$, $a$ is the ion diameter. For $|\boldsymbol{r_2 3}| < a$, the Ornstein-Zernike equations for a k-component homogeneous fluid is solved. The pair distribution function $g(r) \equiv h(r) + 1$. Therefore we have:

$$g_{\alpha j}(\boldsymbol{r_3}) = exp[-\beta u_{\alpha j}(\boldsymbol{r_3}) + \sum_{m=1}^{k} \int \rho_m h_{\alpha m}(\boldsymbol{r_4}) c_{mj}(\boldsymbol{r_{34}}) d\boldsymbol{r_3}] \tag{3.6}$$

Where we have simplified the notations $r_{12} \rightarrow r_3$, $r_{13} \rightarrow r_4$ and $r_{23} \rightarrow r_{34}$. To solve the integral equation set the space is discretized into a mesh in elliptic coordinates[**19**].

Inserting the electrostatic potentials

$$u_{ij}(r_{34}) = \begin{cases} \frac{z_i z_j e^2}{\epsilon r_{34}} & r_{34} > a \\ \infty & r_{34} > a \end{cases} \tag{3.7}$$

$$u_{\alpha j}(\boldsymbol{r_3}) = \begin{cases} \frac{4\pi z_j e R\sigma}{\epsilon}[ln(r_{13}) + ln(r_{23})] & r_{13} > R_a \text{and} r_{23} > R_a \\ \infty & r_{13} < R_a \text{or} r_{23} < R_a \end{cases} \tag{3.8}$$

and assuming an initial electrostatic potential for each mesh point, one can solve the integral equation for $g_{\alpha j}$ and $c$ self-consistently until the solution converges. For the mean force between the cylinders, the Born-Green-Yvon is used [**81, 19**].

$$F_{\beta\gamma}(\tau) = -\frac{du_{\beta\gamma}(\tau)}{d\tau} - \sum_{j=1}^{k} \rho_j \int g^{(3)}_{\beta\gamma j}(\boldsymbol{r_1, r_2, r_3}; \tau) \times \frac{du_{\gamma j}(\boldsymbol{r_{23}})}{dr_{23}} \cos\theta dv_3 \tag{3.9}$$

where $g^{(3)}_{\beta\gamma j}(\boldsymbol{r_1, r_2, r_3}; \tau)$ is the three-particle correlation function between two fix cylinders placed at a distance $\tau$ and a third particle (the ion). The total mean force can be further decomposed into:

$$F_{\beta\gamma}(r) = F_E(r) + F_C(r) \tag{3.10}$$

Where the subscript E stands for electrostatic contribution and C stands for the entropic contribution (collision of particles). And the expressions for $F_E$ and $F_C$ are: [**19**]

(3.11)

$$F_E(r) = -\frac{du_{\beta\gamma}(r)}{dr} - \sum_{j=+,-} \rho_j \int g_{\beta\gamma j}(\boldsymbol{r_{23}; r}) \times \frac{du_{\gamma j}(\boldsymbol{r_{23}})}{dr_2 3} \cos\theta dV F_C = k_B T \int \rho_0(\theta) \cos(\theta) dA$$

### 3.4. Results and Discussion

Since we are interested in the interaction between rods, which is highly related to the osmotic pressure of the electrolyte solution, we calculate the osmotic pressure of a simulation box with only NaCl ions and implicit water, the results are shown in Fig. 2.3. The osmotic pressure is very close to the values given by either AA MD simulations or experimental data.

We then study a single, infinitely long rod in NaCl solution. We will use the leading term in Eq. 3.2 to fit our data and discuss the consequences of ion-ion short-range correlations that are omitted in the PB equation. The radial distributions of coions and counterions around the cylinder from 0.3 M to 1 M are plotted in Fig. 3.2. The results for 0.03 M is shown in Fig. 3.10 of the Appendix. Below 1 M of salt concentration, the distribution function of ions is monotonic after the first peak. We can achieve a reasonably good fit with the first term of Eq. 3.2. Interestingly, the exponential constant $\kappa$ from fitting the coion distributions and counterion distributions are different, which is related to the steric effect felt by cations and anions. This is another effect not predicted by mean-field theory (Section 3.6 Fig. 3.10).

At concentrations of 2 M and above, the ion distributions are no longer monotonic after the first peak, see Fig. 3.3. The strong short-range correlation of ions causes over-screening [**75**]. This phenomenon arises as long as excluded volume interaction is included in the physical model of ions. For example, we use liquid state (LS) theory with HNC/MSA closure and obtain a very similar distribution (dashed lines in Fig. 3.3). This over-screening behavior is also called charge inversion or charge reversal[**82, 19**]. Gonzalez et al. and others have discussed the transition from monotonic decay to oscillatory

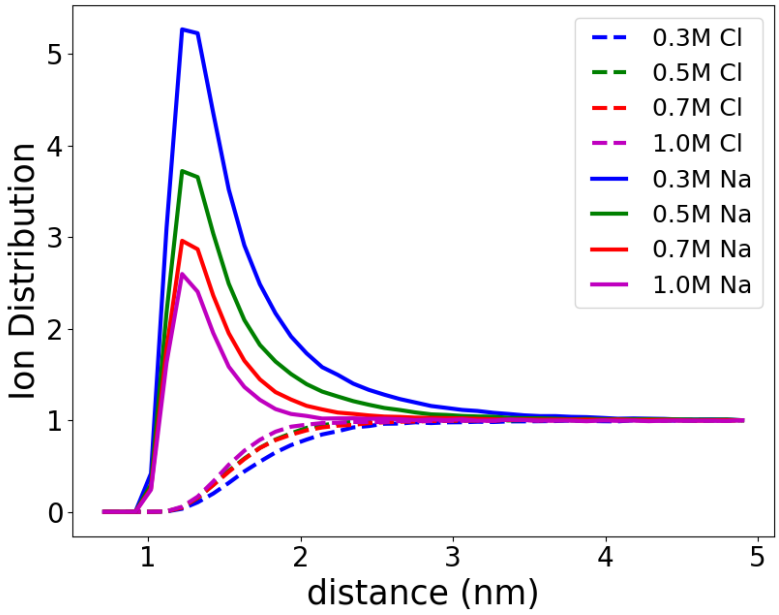



Figure 3.2. The radial distribution of monovalent salt around the charged cylinder. The X axis is the distance from the cylinder axis, Y axis is the normalized ion concentration.

distributions [**75, 18**] using dressed ion theories. This happens when the screening parameter $\kappa$ is no longer a real number, but a pair of complex conjugates. The resulting form of electrostatic potential is

$$\phi \sim \frac{1}{2\pi\varepsilon_r} \times \frac{e^{-\kappa_R r}}{r} \cos(\kappa_I r - \theta) \tag{3.12}$$

However, a fitting using this function form does not in general result in good conversion, indicating that detailed short-range correlations in our model modify the shape of ion distributions compared to the prediction of dressed ion theories.

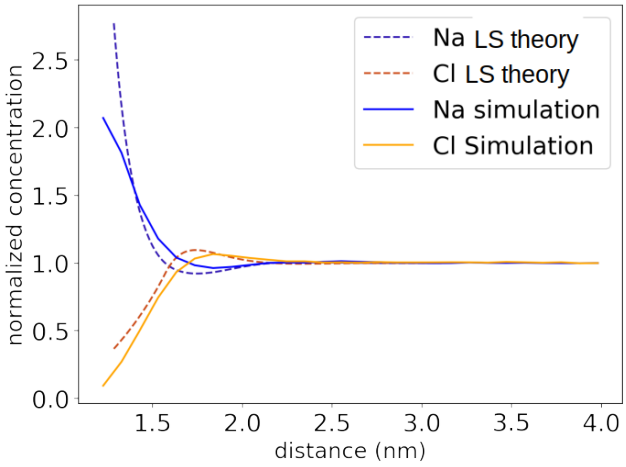



Figure 3.3. The distribution of 2M NaCl around charged cylinder. The X axis is the distance from the cylinder axis.

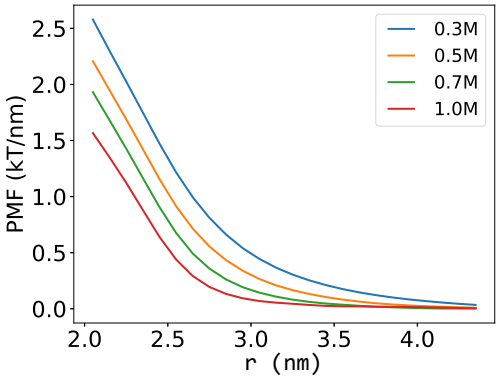
### 3.4.1. Potential of Meanforce Between 2 Cylinders

Next, we study the interaction between rods in NaCl solutions by calculating the potential of mean force (PMF) of two charged or neutral cylinders. Since at sufficiently low concentration, the ion distribution agrees well with PB equation, we only calculate the PMF for rods in $> 0.3M$ concentration.

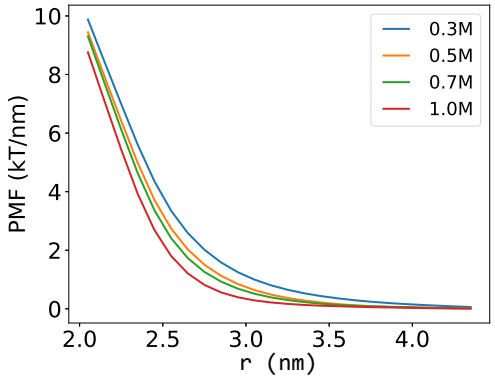
First, we consider two rods with a diameter of 2nm, which is roughly the thickness of a double-stranded DNA. Using the "smooth rod" model, we study the charge density of 3 e/nm and 6 e/nm. The latter is approximately the linear charge density of DNA. The interaction of DNAs in a salt concentration below 1 M is purely repulsive (see Fig. 3.4a). To our knowledge, there is no closed-form expression for the interaction between two cylinders in salt from PB or other theories. However, since the ion distribution in the dilute limit follows a dominating term that decays as $e^{-\kappa r}/r^{1/2}$, we hypothesize that

the interaction at large distances ($> 0.5$ nm surface distance) also follow the same functional form. Discarding the data points with distance $< 2.5$nm which are dominated by short-range interactions such as excluded volumes, a good fit ($R^2 > 0.999$, see Fig. 3.12b in Section 3.6) can be obtained using the first term of Eq. 3.2. At the concentration of 1 M, the data points at mid-range slightly deviate from the fitted curve, but still with $R^2 > 0.99$. A screening parameter $\kappa$ can be determined from the fitting and compared against the screening parameter $\kappa_D$ predicted by Debye-Huckel theory. At concentrations below 0.7 M, Debye-Huckel predicts a screening length reasonably close to the simulation value. For a pair of cylindrical interfaces, we do not find a significantly increased screening length as discovered by [**67**]. Recent large-scale MD simulations [**69**] also report much less increase in the electrostatic screening at high salt concentration. When the concentration further increases, the PMF takes another functional form, so a natural extension to the screening parameter $\kappa_D$ needs to be defined with care. Note that the $\kappa$ fitted from simulation data depends on the rod surface charge density, which is in line with the breakdown of DH theory at high surface charges.

Above 2M of concentration, the PMF is not monotonic. For 2M, we find an small attractive well of 0.1 $k_BT/nm$ (see Fig.3.6 a), which is non-trivial for a long segment of DNA (tens of nm). Note that the yellow shaded area indicate the cumulative error assuming that the PMD must go to 0 at the cutoff distance of 6 nm. Since the PMF is calculated by numerical integration of the mean force, the error is a random walk in $r$ and PMF starting from the right. Therefore, although the error region appears to be big enough to draw a smooth curve without attraction in Fig. 3.6 a, such a curve (a random walk going




(a) The PMF between two infinite-length rods with 2nm diameter, 3 e/nm in NaCl solutions of different concentrations. Errorbars are comparable to line thickness and not shown. $r$ is the distance between the 2 axes of the cylinders, the same for other PMF plots below.




(b) The PMF between two infinite-length rods with 2nm diameter, 6 e/nm in NaCl solutions of different concentrations. $r$ is the distance between the 2 axes of the cylinders, the same for other PMF plots below.

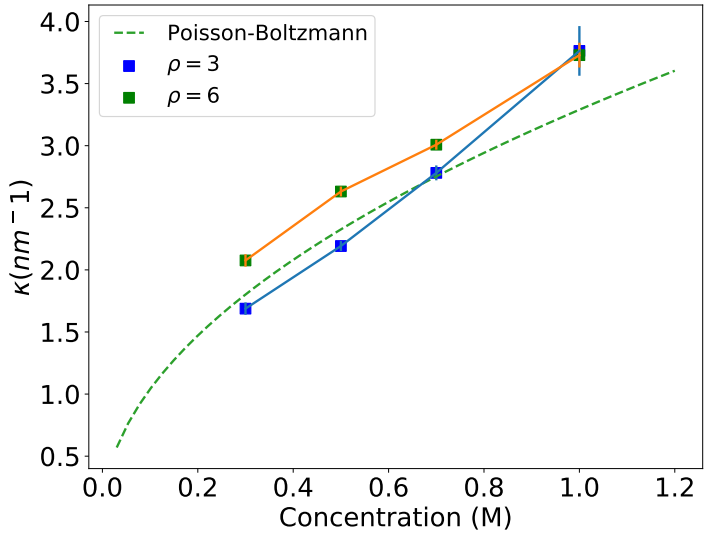
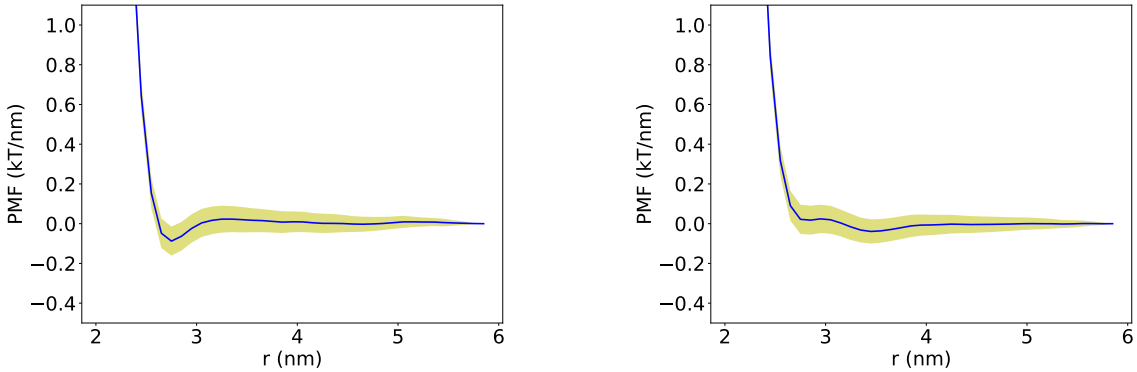



Figure 3.5. Screening parameter $\kappa$ determined from fitting the PMF to 3.2

in one direction in many consecutive steps) is highly unlikely. A 2D time-averaged ion density profile confirms that this attractive interaction arises from the depletion of ions

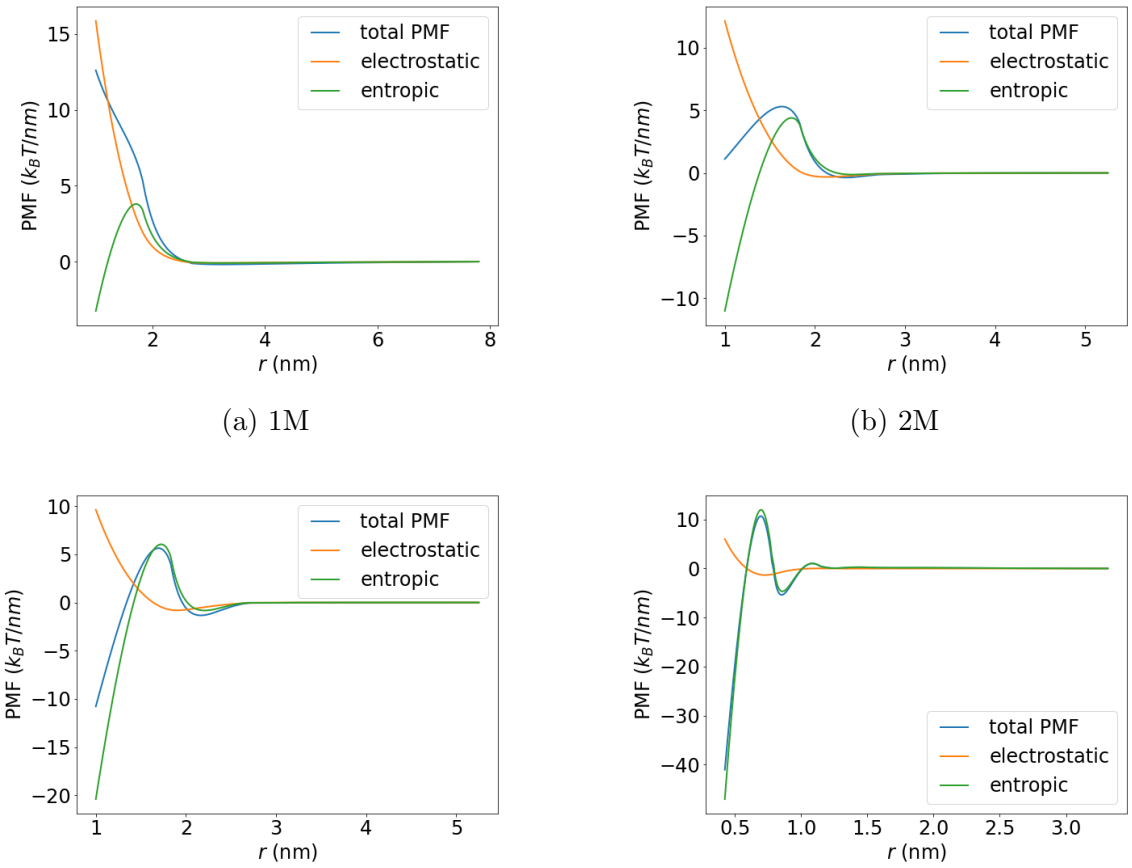



(a) 2M (b) 3M

Figure 3.6. The PMF between two infinite-length rods with 2nm diameter in NaCl solutions of different concentrations. Errorbars are comparable to line thickness and not shown.

(see Appendix Fig. 3.13). This attraction indicates a complete breakdown of PB based theories at high concentration, as they always predict monotonic interactions.

The mean force between the 2 rods can be calculated using the Born-Green-Yvon equation and the MSA-HNC closure [**19**]. This calculation also predicts similar trend: monotonic repulsion below 1M and oscillatory interactions at high concentration, see Fig. 3.7. The LS theory results agree with MD simulaiton results using the PM (Fig. 3.8). It is important to note that in the LS theory, the ions are modeled as perfectly hard spheres with the size of ions together with the hydration water molecules. At high concentrations, the hydration shells approach close-packing and cause strong oscillation in the interaction. MD simulations with the IBI potentials, however, have softer short-range interactions at the hydration shell, and thus predict a smaller attractive well.




(a) 1M

(b) 2M

(c) 3M

(d) 5M

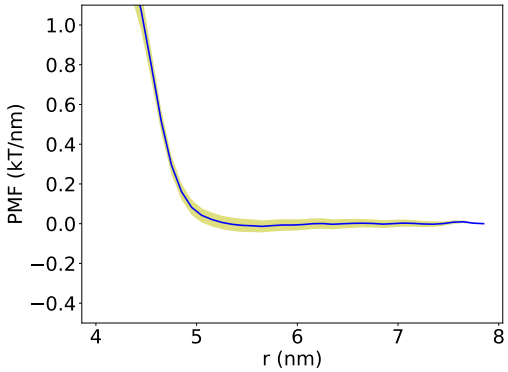
Figure 3.7. The PMF between two cylinders of 4 nm diameter calculated from liquid state theory using the MSA-HNC closure. Note that the plot range for 5M has been increased because the interaction is much stronger

The mean force calculated with HNC/MSA closure (Fig. 3.7) is decomposed into electrostatic and entropic contributions Eq. 3.10. Interestingly, this decomposition reveals that steric effects (the green curve in Fig. 3.7) have a major contribution to the rod-rod interactions in magnitude. However, it is important to bear in mind that the collision of ions is heavily coupled to their electrostatic interactions. Without the electrostatic interactions, the ions would not be in such a distribution. In Eq. 3.10, the direct steric

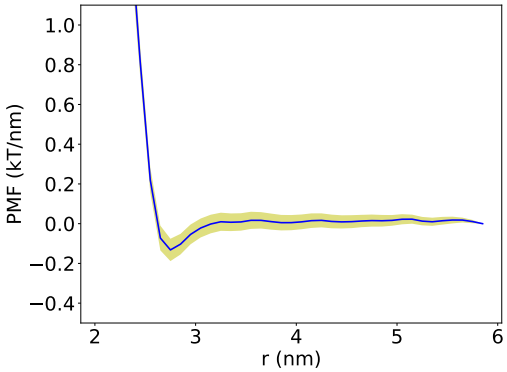
interaction between the rods is not included. Therefore, as the distance between rods decrease to below 2 nm, which is the rod diameter, the entropic interaction become strongly attractive due to the osmotic pressure of ions from both sides. In reality, the rods are not allowed to go below the diameter because of their excluded volume. From this analysis, we can see that the collision of ions under Coulomb interactions, rather than the direct Coulomb interaction between charged rods and ions, account for the majority of rod-rod interaction at high salt concentrations.

To properly compare with the results from LS theory, MD simulation of the PM is also performed with the same charge density and rod diameter. The results of PM agree well with the LS calculations (Fig. 3.8), again confirming that the MSA/HNC closure captures the correlation effects to a good accuracy. Both the LS and MD simulation of RPM deviate from the IBI MD simulaions at 3M of salt concentration. The major difference between the two is that the IBI potential has a soft potential barrier of $\sim 1k_BT$ that mimics the hydration shell of ions, which breaks apart when the mean ion-ion distance is comparable to hydration shell sizes. From this comparison, we infer that at high salt concentration, even a simple 1:1 electrolyte can demonstrate more sophisticated behavior which is not described by the PM.

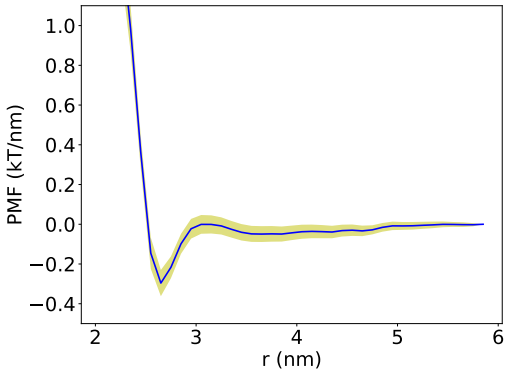
It is worth noticing that compared to the case of spherical interfaces (see 2), the onset concentration of attractive interaction is higher, for both spherical models (smooth and explicit chain) and both cylinder models (smooth and rough rods). A possible reason is that the unscreened electrostatic potential decays in a different functional form (logarithmic) for cylinders compared to spheres (1/r). Correspondingly, the depletion force for uncharged hard spheres follows other forms (linear plus cubic for spherical interface [**59**]




(a) 1M




(b) 2M




(c) 3M

Figure 3.8. PMF calculated from MD simulation of RPM with ion diameter 0.425 nm, rod charge density 6e / nm. 5M concentration is not simulated due to computation limit.

and $x^{3/2}$ for cylindrical interfaces using the relation that depletion potential is proportional to the overlap volume, $x = 2R - r$), leading to the difference in electrostatics-entropy competition for different geometries. It would be interesting to also compare to the result of planar interfaces at high salt concentration.

Figure 3.11 in the appendix shows us that within half a nm, even the strongly charged DNA is over-screened by concentrated electrolytes. The DNA, together with the condensed counterions can be seen as a weakly charged rod, which is attracted to another DNA by depletion attraction.

To study the effect of different curvature, we also study cylinders of 8 nm diameters. The latter can be a simplified model for larger nanorods such as carbon nanotubes and peptide amphiphiles. The thicker rods follow qualitatively similar behavior compared to the 2 nm rods. At 1M, no detectable attraction is found. The screening length is 0.38 $\pm$0.01 nm. At 2M salt concentration, even highly charged cylinders can exhibit depletion-type attractions since the surface charge is strongly screened by the electrolyte (Fig. 3.9). Interestingly, at 2M concentration, compared to an uncharged rod, a charged rod has a more pronounced attractive well. This can be rationalized by the ion distribution and net charge distribution Fig. 3.11. At such high salt concentrations, even strongly charged rods are completely screened within 1 nm from the surface. The charged rod and tightly bound layer of ions can be seen as one object, and attractions are caused by depletion effects of ions and ion clusters.

## 3.5. Conclusions

In this paper we study the interaction of rod-shaped macromolecules or their rod-like assemblies immersed in monovalent salt solutions using explicit ion MD simulations and liquid state theories. We reproduce the previously known result that at high salt concentrations, the ion distributions transition from monotonic to oscillatory functions. By fitting the PMF between the rods, we extract a screening parameter, which systematically deviates from PB predictions to a higher value up to 1M. We do not find the anomalous increase reported in Ref. [**68**]. Ref. [**69**] calculate the PMF between ions using large-scale AA MD simulation, and also fails to find a long-ranged exponential decay. Our simulations indicate that this disagreement is not explained by the boundary effect of cylindrical interfaces. As pointed out in Ref. [**69**], the disagreement in MD simulations and experiments might indicate that the experimentally observed long-range repulsion is not an equilibrium bulk property.

At above 2M of monovalent salt, the PMF between two neutral or like-charged rods can be attractive due to the depletion effect of ions. For 2M, we find a small attractive well of 0.1 $k_BT/nm$ (see Fig.3.6 a), which is non-trivial for a long segment of DNA (tens of nm). At this salt concentration, even strongly charged rods can be sufficiently screened, and have a strong attractive interaction. MD and LS theory of the PM predicts stronger depletion force compared to the IBI model because the hard-sphere model overestimates the steric effects. The interaction for rods of different diameters from 2nm and 8nm follows qualitatively similar trends. For large rods at high salt concentrations, the strong charge on cylindrical surfaces enhance salt-induced attraction rather than compromise it. These results explain the aggregation of highly charged rod-like molecules such as DNA

in concentrated salt solutions, which are contrary to the expectation of the PB equation.

## 3.6. Appendix of Chapter 3

The exponential parameter $\kappa$ is obtained by fitting the coion and counterion distribution function around the rod, respectively, leaving out the first few points which are dominated by short-range interactions. In the dilute regime ($< 0.1$M), $\kappa$ match reasonably well with PB predictions (Eq. 3.1).

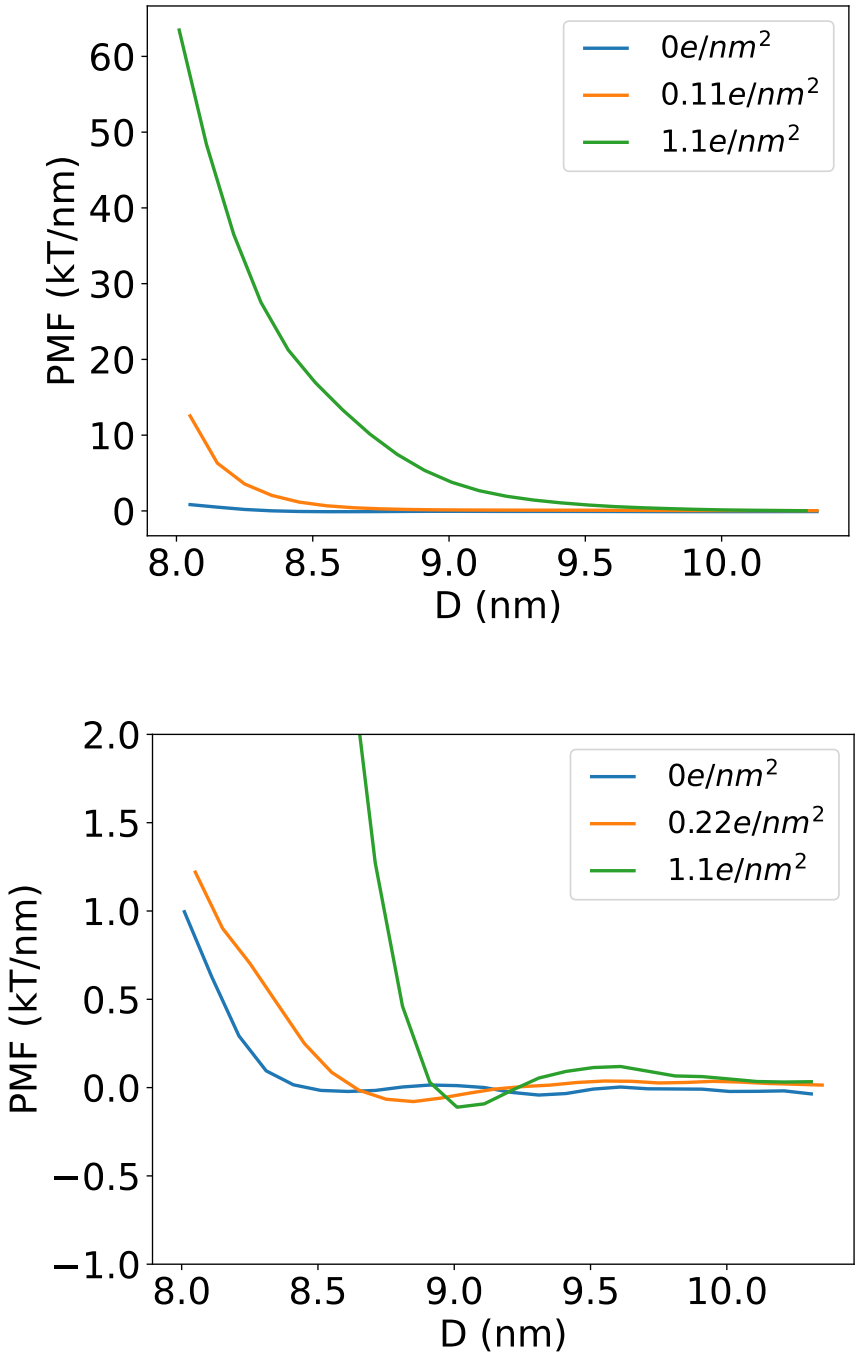



Figure 3.9. The PMF for 7 nm cylinders of different surface charge density in (a) 1M NaCl and (b) 2M NaCl. The relative error is comparable to line thickness.

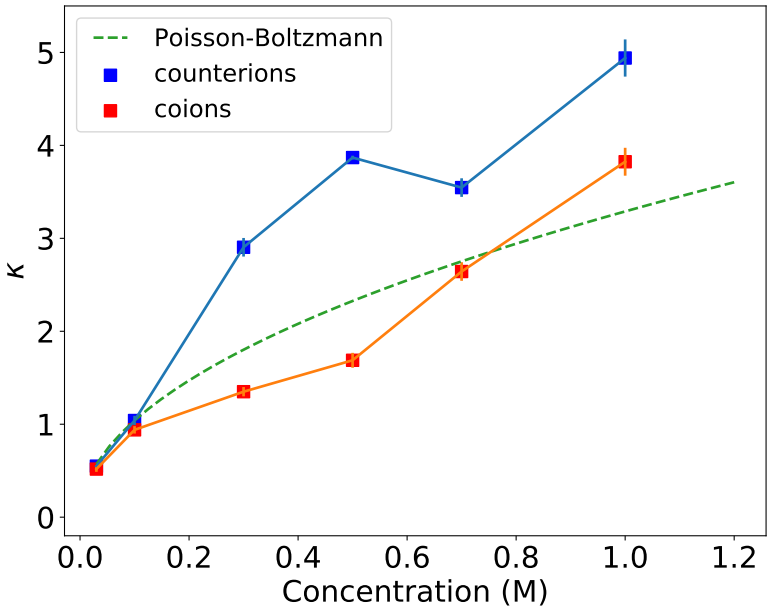



Figure 3.10. Screening parameter $\kappa$ as a function of NaCl concentration. As concentration increases, $\kappa$ deviate from the PB value, although the ion distribution functions can still be fitted to the first term of PB solutions. Interestingly, the $\kappa$ fitted from coion and counterion distributions differ from one another at concentrations above 0.3M, suggesting that the screening behavior of ions are affected by short-range interactions in a different manner.

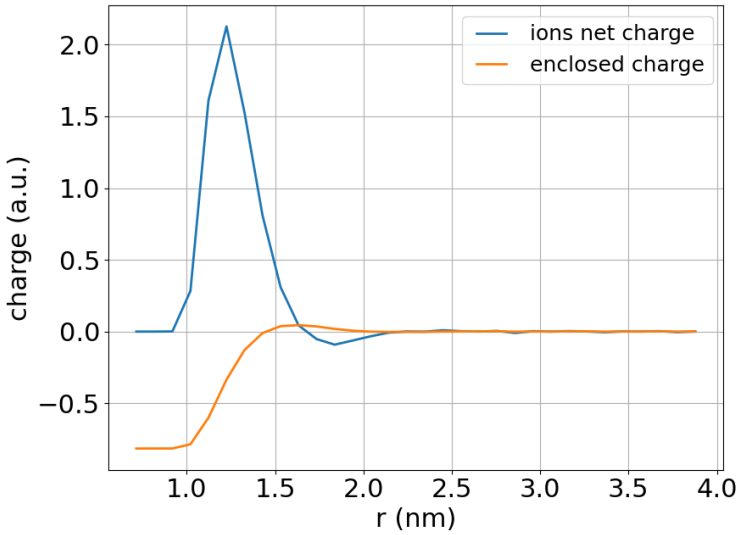
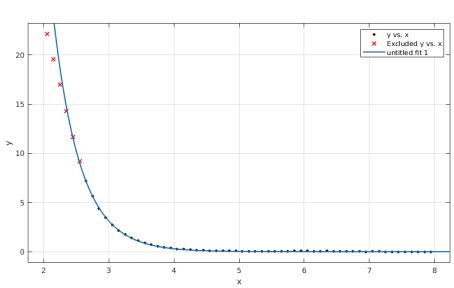
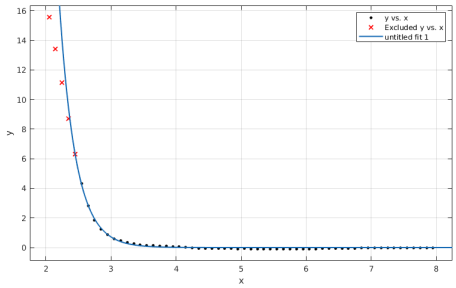



Figure 3.11. Blue line: Net charge and of 2M NaCl around charged cylinder. Yellow line: cumulative charge including the DNA with in radius r.




(a) fitting PMF for 2 DNAs in 0.3M NaCl solution




(b) fitting PMF for 2 DNAs in 1M NaCl solution

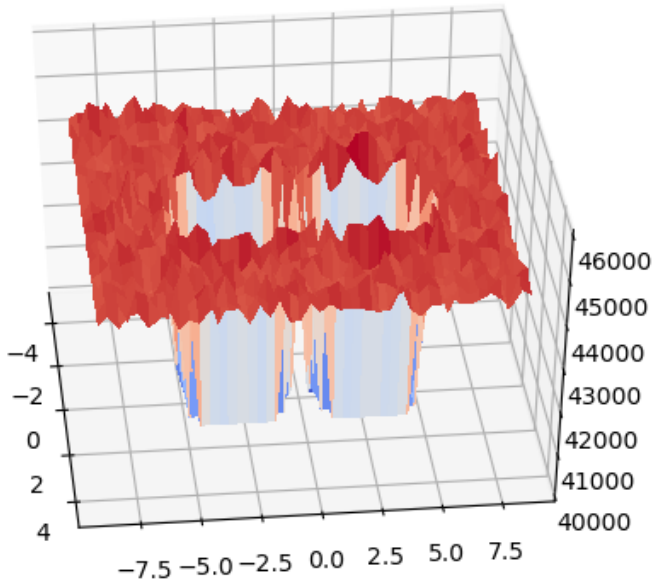



Figure 3.13. 2D density map of ion distributions when two cylinders are at a distance of 5 nm or surface-surface distance of 1nm. The gap region between two cylinders (light yellow) has a lower ion concentration than the bulk, therefore the osmotic pressure of the bulk ion pushes two cylinders to the center.

CHAPTER 4

# MD Simulation and Thermodynamics of Bacterial Microcompartments

Authors: Yaohua Li*, Nolan W. Kennedy*, Siyu Li, Carolyn E. Mills, Danielle Tullman-Ercek and Monica Olvera de la Cruz

Contributions: Y. Li constructed the CG model and performed the MD simulations, S. Li constructed the thermodynamic analysis. N. W. Kennedy and C. E. Mills performed the experiments, C.E.M analyzed and interpreted computational and experimental findings. Y.L, N.W.K, S.L., C.E.M., M.O.d.l.C. and D.T.E wrote the paper. M. O. d.l. C and D.T.E. supervised the research.

This work was submitted to ACS Central Science in December 2020.

## 4.1. Chapter Abstract

Bacterial microcompartments compartmentalize the enzymes that aid the chemical and energy production in many bacterial species. These protein organelles are found in various bacterial phyla and are postulated to help many of these organisms survive in hostile environments such as the gut of their hosts. Metabolic engineers are interested in repurposing these organelles for non-native functions. Here, we use computational, theoretical and experimental approaches to determine mechanisms that effectively control microcompartment self-assembly. As a model system, we find via multiscale modeling and mutagenesis studies, the interactions responsible for the binding of hexamer-forming

proteins propanediol utilization bacterial microcompartments from Salmonella and establish conditions that form various morphologies. We determine how the changes in the microcompartment hexamer protein preferred angles and interaction strengths can modify the assembled morphologies including the naturally occurring polyhedral microcompartment shape, as well as other extended shapes or quasi-closed shapes. We demonstrate experimentally that such altered strengths and angles are achieved via amino acid mutations. A thermodynamic model that agrees with the coarse-grained simulations provides guidelines to design microcompartments. These findings yield insight in controlled protein assembly and provide principles for assembling microcompartments for biochemical or energy applications as nanoreactors.

## 4.2. Background

Compartmentalization of cell components enables a variety of functions and ensures that biochemical processes happen without interfering with one another. Proteins play a crucial rule in compartmentalization. Examples of closed compartments formed by protein components include viral capsids, which protect enclosed nucleic acids via robust assembly rules [**83, 84**], and bacterial microcompartments (MCPs), which aid the breakdown of chemicals and energy production that allow bacteria to thrive in various environments [**26**]. MCPs, which are polyhedral structures capable of controlling the transport of specific molecules, are found in a large variety of bacterial species [**85, 86, 87**]. Due to their native ability to enhance chemical transformations within cells, synthetic biologists are working to repurpose MCPs for non-native, industrially relevant purposes [**86, 88**]. Therefore, determining and understanding assembly mechanisms of MCPs is a relevant to

industry and to life and physical sciences problem. The envelopes of MCPs are assembled from proteins, much like some viral capsids. In contrast to some capsids, however, these shells are often irregular in shape, contain multiple types of proteins, and enclose specific enzymes as cargo [**89**]. Three types of shell protein components form the envelope of 1,2-propanediol utilization (Pdu) MCPs, ethanolamine utilization (Eut) MCPs and many other MCPs [**90**]. These three MCP shell components are composed entirely of two different protein domains, the bacterial microcompartment (BMC) domain (which comprises the hexameric protein and trimeric proteins BMC-T) and the bacterial microcompartment vertex (BMV) domain (which comprises the pentameric proteins), see Fig.1.2. Six BMC-domain monomers form the flat, six-sided, hexameric structures (BMC-H); three tandem BMC-domain dimers form the flat, six-sided, pseudohexameric trimers (BMC-T), and five BMV-domain monomers form the flat, five-sided, pentameric structures (BMC-P). When hexagonal shape components (BMC-H and BMC-T) assemble with pentagonal components (BMC-P) into closed shells, the topological constraints of the Euler's polyhedron formula are obeyed [**91**]. This formula states that 12 pentagonal components (BMC-P) are needed to form a topologically closed shell, which explains the ubiquitous icosahedral symmetry of spherical shells [**92**] and their buckling into icosahedral shells [**93**]. However, the closed shells formed by many MCP systems, including the Pdu MCP system, have complex closed or extended geometries [**28, 94, 95**] in which shell protein components are present at different ratios. Polyhedral MCPs are predicted in multicomponent closed shells using elasticity theory [**96, 97**] , yet the conditions, including protein size, shape and interactions, to achieve specific structures are unknown. In certain conditions these MCP proteins assemble to produce extended cylindrical structures [**28**]. A

recent study demonstrated that open "icosahedral cages" can be formed in the absence of BMC-P proteins [**31**]. These cages contain open spaces that enable BMC-P binding such that Euler's polyhedron formula is still obeyed when the shell closes. These studies have shed light on some basic aspects of MCP shell structure. The rules which govern MCP assembly into various morphologies, such as the nature of attractive that bind the component proteins and what factors control the assembly shape, are not well-understood.

As an experimental model system we study the PduA BMC-H protein of Salmonella enterica serovar Typhimurium LT2 [**98**]. By computationally studying specific interactions between PduA hexamers, we find the conditions for assembly into closed polyhedral or extended shapes and examine a specific mutation site on these hexamers that affects the assembly of MCPs into different morphologies. We use all-atom (AA) MD simulations to determine the interaction energy and equilibrium bending and twisting angles of native and mutated hexamers and predict what mutations can lead to self-assembly of MCPs or extended shapes. Experimental results with mutated PduA hexamers qualitatively corroborate the simulation predictions. Our study reveals that electrostatic and hydrogen bonding interactions between the arginine and valine residues on the edge of PduA are key to the self-assembly of PduA hexamers and MCP formation, as the majority of the mutations to this residue negatively impact MCP formation [**99**]. We use the structural information and interaction energy from AA MD simulations to construct a coarse-grained (CG) model for PduA self-assembly. Moreover, we determine what assembled geometries are accessible given different stoichiometric ratio of the 3 major MCP component proteins. Finally, we use this CG framework in combination with a theoretical thermodynamic model to explore how modulating BMC-H/BMC-H and BMC-H/BMC-P interactions can

permit access to an array of assembled morphologies. Simulation and theoretical findings agree with in vivo experiments of MCP encapsulating green florescent proteins (GFP). These findings will help guide future studies that seek to repurpose MCPs with specific morphologies and shed light onto the rules governing MCP assembly in general. This study also shows that electrostatics is crucial for MCP assembly, removing charge interactions change the assembly behavior of BMC-H proteins and result in the formation of flat sheets.

## 4.3. Molecular Dynamics Simulations

### 4.3.1. All Atom MD simulaitons

We first investigate the interaction between two native BMC-H subunits using AA MD simulations of the PduA crystal structure from S. enterica LT2 (PDB id: 3ngk) [**100**]. The two PduA hexamers bind into a highly stable dimer (see Fig. 4.1 A) as a result of complementary shape and hydrogen bonding (see Table 4.1), in which the arginine at the 79 th position (ARG79) binds to the backbone carbonyl oxygen of VAL25. The side chain of ARG79 fits in the pocket formed by the residues ASN29, VAL25, HIS81 and LYS26 (Fig. 4.1A inset). This result corresponds to previous studies indicating ARG79 plays a role in inter-hexamer binding [**99, 101, 31, 102**]. To better quantify the relative orientation of the two hexamers, we calculated the angle between the norm of the hexagonal planes (Fig. 4.1 B). The angle is projected onto the y-z plane and x-z plane to decompose into a bending angle $\theta_b = 10 \pm 4^\circ$ (error bars are standard deviations) and a twisting angle $\theta_t = 5 \pm 3^\circ$ at equilibrium (Fig. 4.1 C), with a mean total angle $\theta = 11 \pm 4^\circ$. (A histogram of the bending and twisting angles are shown in SI Fig

4.7). This calculation suggests that in solution PduA hexamer complex mostly adopt a near coplanar configuration, corresponding to the flat facets observed in polyhedral MCP structures [**31**]; there is another orientation with the hexamers at a 30° angle relative to each other [**31**]. The latter bent configuration is likely due to collective interactions among other proteins and scaffold, as shown in our coarse-grained model described below, and therefore, it is not observed in simulations of two proteins.

We study the potential of mean force (PMF) of two PduA hexamers using umbrella sampling MD simulations (Fig. 4.1 D). The binding energy ($\varepsilon_{hh,AA}$) is estimated to be 11 ± 2 kCal/mol. This binding energy falls in the range of reported values of hydrogen bonds [**103**], further indicating that hydrogen bonds are a major contributor to PduA binding. The calculated binding energy and bending and twisting angles from atomistic simulations provide semi-quantitative guidance for building larger scale models.

### 4.3.2. Coarse-grained MD simulations

Since AA simulations cannot include many proteins and cover the time scales required to assemble the proteins into different MCPs morphologies, we use CG modeling to study the assembly of MCPs at larger time and length scales. CG simulations have been used to follow the assembly of simplified protein models into MCPs or viral capsids [**30, 108**]. MD simulations of one shell protein predict the stable formation of small icosahedral capsids [**109**]. CG models have also elucidated the initial faceting process [**29**] and the formation of a complete MCP [**30**]. However, these models have not investigated how interactions between multiple shell components results in various MCP morphologies such as cylinders, and the molecular origin of spontaneous curvature has not been explained.

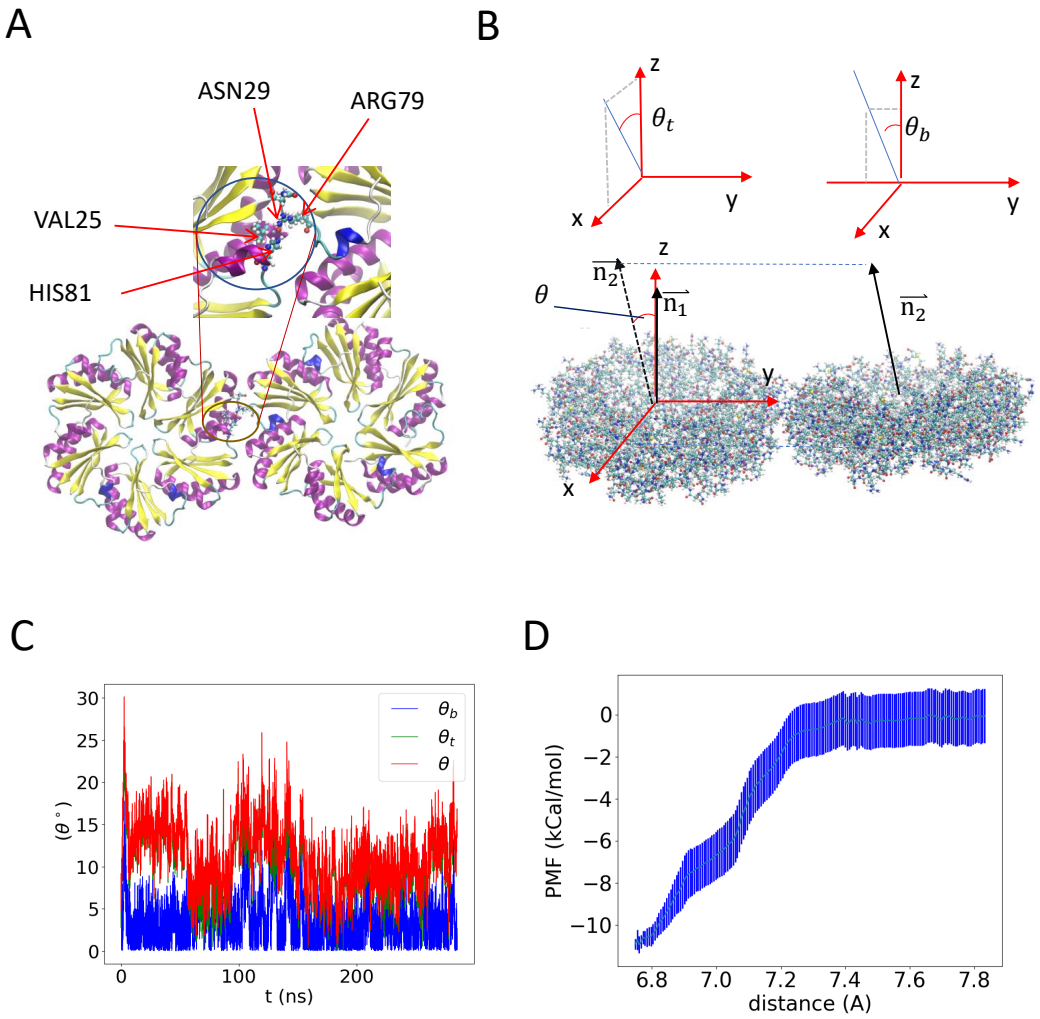



Figure 4.1. All atom (AA) molecular dynamics (MD) simulations results of a pair of PduA hexamers. (A) The equilibrium configuration. The arginine residue on each edge sticks out to form hydrogen bond with the backbone of valine in the neighboring PduA hexamer. The two arginine residues are highlighted by increased bead size. Water and ions are hidden for visual clarity. (B) Illustration of the reference frame and angle between the two PduA hexamers. The angle is decomposed into twisting angle $\theta_t$ and bending angle $\theta_b$ by projecting the normal of the neighboring protein ($\boldsymbol{n_2}$) onto the XZ plane and the YZ plane, respectively (see SI Fig. 4.7 for a different representation) (C) The MD simulation values of $\theta_b$ and $\theta_t$. (D) Potential of mean force between two PduA hexamers shows an attractive binding energy of about 11 kCal/mol.

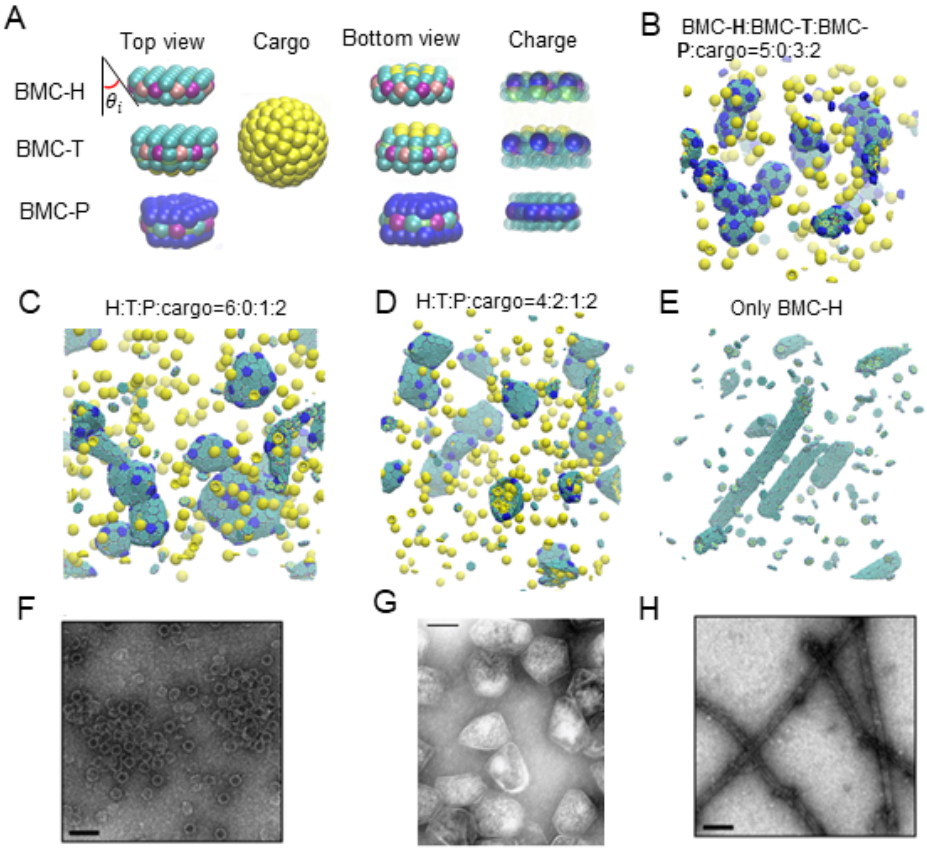



Figure 4.2. Coarse-grained (CG) model and MD simulations results. (A) Illustration of the CG model. The sides of BMC-H and BMC-T are inclined at an angle 25 according to their pdb structures (PDB id 3ngk [**100**] and 4fay [**104**]). The green beads interact via excluded volume (changed to blue color for pentamers), the purple and pink beads are short-range attractive sites representing the arginine hydrogen bonds. The sphere of yellow beads is the cargo and the yellow beads on the BMC-H and BMC-T proteins are the residues that bind to the cargo. To show the negative (blue) and positive (green) charged sites clearly, all the non-charged sites are shown as semi-transparent and in smaller size on the right column. (B) (E) Snapshots of CG simulations (the hexamers, pentamers and cargo are in green, blue and yellow, respectively): a red dot is marked on the center of BMC-T to distinguish them from BMC-H. (B) A system without BMC-T, with a ratio BMC-H:BMC-P:cargo=5:3:2 forms T=3 icosahedral shells resembling in vitro electron micrographs of compartments from Haliangium ochraceum (F) which form triangulation number[**105**] T=9 shells. (C) (D) Increasing the number of BMC-H or adding BMC-T can enable assembly into polyhedral shapes that resemble the shape of purified MCPs shown in (F). (E) BMC-H proteins alone form cylinders, reproducing observations of in vitro BMC-H assembly shown in (H). (G) is reprinted with permission from C. Fan, et al. PNAS 2010, 107, 7509-7514 [**106**], (F) and (H) are reprinted with permission from A. R. Hagen, et al. Nano Letters 2018, 18, 7030-7037 [**107**].

We hypothesize that the stoichiometric ratio of different components and the interaction strength between them control the morphology of assembly products [**90, 28, 106**]. We construct a CG model using the approximate shapes and charge distributions of the aforementioned PduA structure (PDB: 3ngk [**100**] ) for the BMC-H model, the PduB homolog from Lactobacillus reuteri (PDB: 4fay [**101**] ) for the BMC-T model, and the PduN homolog GrpN from Rhodospirillium rubrum (PDB: 4i7a [**110**] ) for the BMC-P model. We validate this model by comparing assembled morphologies against known experimental observations of various MCP and cylinder shapes [**90, 28, 106**]. The components of our CG model are shown in Fig. 4.2 A. In the CG simulations the charges of the BMC-H, BMC-T and BMC-P proteins are computed from the AA structures (see SI Fig. 4.8), the mechanical properties such as the spontaneous curvature of BMC-H assemblies result from the inclined edge shape; that is, the inclination angle $\theta_i \sim 25^\circ$ obtained from the PDB (see CG simulations results using other inclination angles in Figs 4.9 in SI). We first explore the various geometries with different stoichiometric ratio of BMC-H and BMC-P. We find icosahedral shells assembled with BMC-H : BMC-P = 5:3, which resemble the shape of in vitro MCPs from Haliangium ochraceum [**28, 111**] (Fig. 4.2 B). When we reduce the content of BMC-P to BMC-H: BMC-P = 6:1, the addition of hexameric proteins allows the shell to assemble into the irregular shapes of Pdu MCPs reported in the literature [**90, 106, 112**], illustrating how varying shell protein content can control morphology of MCP structures formed. When a small portion of BMC-H is replaced by BMC-T, similar structures are found (Figs. 4.2 C and D). However, when the content of BMC-T further increases, the asphericity increases (see SI Fig. 4.11). The impact of BMC-T on the assembly can be understood by the difference in bending angle and

rigidity of the BMC-H and BMC-T (a bending angle analysis is given in the SI). We note that the BMC-H : BMC-P ratio used in these simulations is lower than that typically observed for Pdu MCPs in vivo and the MCPs simulated are about 20–40 nm in diameter, compared to typical reported Pdu MCP diameters of 40–600 nm [**27, 95, 112, 113**]. These smaller MCPs equilibrate faster, allowing more efficient exploration of parameter space while still qualitatively reproducing the characteristic polyhedral shapes of MCPs observed in experiments [**28, 106, 114**]. Simulations with BMC-H : BMC-P ratios of 8:1 produced bigger, more aspherical shells (SI Fig. 4.12).

Long cylinders are obtained when only BMC-H proteins are in the assembly, as shown in Fig. 4.2 E. The shape is consistent with both in vivo and ex vivo experiments [**115, 116**], in which overexpressing PduA proteins resulted in the assembly of long cylindrical tubes, and in vitro experiments in which the BMC-H from H. ochraceum spontaneously assembled into tubes in the absence of other shell proteins . The ability to reproduce vastly different shapes (cylinders and polyhedra) suggests that this simple CG model captures the key interactions of Pdu MCP proteins.

Once our CG model has been validated by comparing the results with experimentally observed morphologies at different stoichiometry, we explore how the docking angle of BMC-H proteins impacts MCP assembly. AA MD simulations reveal (Table 4.2) that the docking angle of BMC-Hcan be tuned by mutating the arginine at the binding site. To this end, we use a combination of AA MD simulations, mutation experiments, and CG modeling to determine the role of hexamer-hexamer interactions on BMC-H assembly. AA MD simulations, the PduA conservation score (SI Fig. 4.13) and prior work in the field [**117, 99, 115, 111**] indicate that ARG79 is a key residue for controlling

PduA interactions. Thus, we hypothesized that mutations to the arginine residue at the 79 th position (ARG79) of PduA could impact PduA, and, subsequently, MCP assembly. We perform AA simulations of the hexamer-hexamer interface on fourteen PduA mutants. Of these, eleven mutants retain some degree of hexamer-hexamer interfacial contact for the duration of the simulations, and their bending and twisting angles are shown in Fig. 4.3 A-C. Three mutants result in hexamers that entirely dissociate from each other, completely eliminating interfacial contacts (Table 4.2 and Fig. 4.14 in the SI provide further details). The hexamer-hexamer dissociation arises when the arginine is replaced by a negatively charged residue or hydrophilic residue with low isoelectric point (aspartic acid, glutamic acid and glutamine), indicating that Coulombic interactions play an important role in stabilizing the hexamer-hexamer interface in PduA. The alpha-helix of the opposite hexamer has a negative partial charge, which is thought to complement the native, positively charged arginine residue [**102**]. Therefore, mutating the arginine to the opposite charge result in total dissociation of the PduA.

## 4.4. Mutagenesis Experiments Comparison with Simulations of Hexamers Assembly

We turned to an experimental approach to assess the assembly of PduA mutants and validate the above described AA MD simulations. When PduA is overexpressed in E. coli cells it naturally assembles into long cylinders that span the length of the cell body. The assembled protein structures form bundles within the cell cytoplasm, preventing cell division and leading to extended chains of cells [**95**] (SI Fig. 4.15). This allows for cell

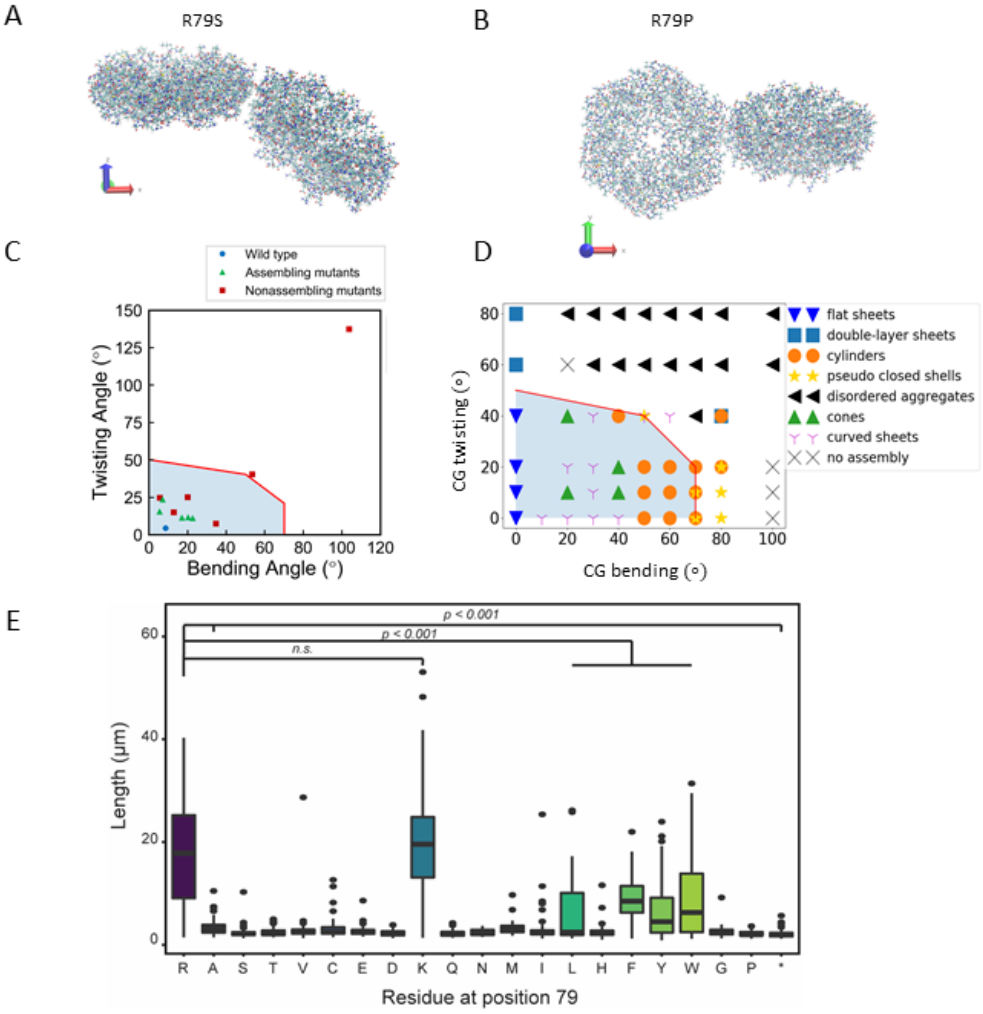



Figure 4.3. PduA mutant orientation studies. An example of a mutation with large bending angle is shown in (A) and large twisting angle in (B). (C) The twisting and bending angle calculated for 11 R79 mutation from AA simulations (corresponding to the data in Table 4.2). The PMF for 3 mutations and WT are calculated and shown in Table 4.2 . The red line indicates the critical angle to form extended structures as predicted by CG simulations. Triangle and square markers indicate whether the mutants self assemble in experiments (E and SI Fig. 4.15) (D) The morphologies formed in CG simulation of 256 PduA proteins with given bending and twisting angles. The shaded area below the solid red lines in (C) and (D) indicate the system can form extended structures, which we predict to cause long chains of cells to form in experiments. At bending angles of 70 and 80, the pseudo closed shells (yellow stars) resemble quasi-isosahedra of T=1 in that each hexamer has 5 neighbors (SI Figure 4.16). (E) Distribution of cell length populations for each PduA variant. R79R (WT) and R79K were significantly longer than other PduA variants ($p < 0.001$, t-test).

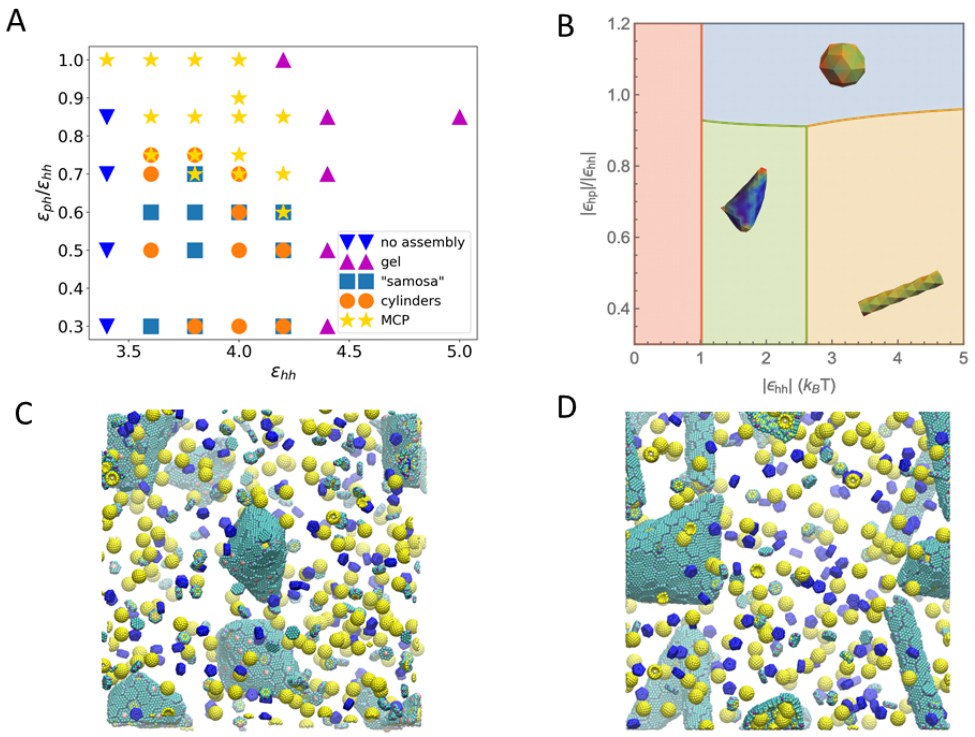



Figure 4.4. (A) Phase diagram of assembled shapes from CG MD simulations. The axes are the parameters defined in Eq.1. The simulation box is constructed from a unit of 4 BMC-H, 2 BMC-T, 2 enzymes, and 1 BMC-P, replicated 5 times in x, y and z directions. (B) Phase diagram from thermodynamics analysis. When the hexamers binding energy is weak, there is no assembly (red region). With stronger hexamer-hexamer interaction ($\varepsilon_{hh}$), MCP, cylindrical and samosa shaped shells are formed, corresponding to the blue, yellow and green regions, respectively. (C) and (D) are snapshots of CG simulations (the hexamers, pentamers and cargo are in green, blue and yellow, respectively). (C) The "samosa" shape ($\varepsilon_{hh} = 3.8, \varepsilon_{ph} = 1.9$) is a quasi-closed surface without BMC-P proteins. They are different from MCPs in that they have no pentamers, and the vertices are sharp cones with a hole or defect at the tip. An example of a 4-fold defect is shown in the center. (D) Co-existence of cylinders and "samosas" at ($\varepsilon_{hh} = 3.8, \varepsilon_{ph} = 2.28$).

length to be used as a proxy for protein assembly and allows us to rapidly assess the assembly state of different PduA mutants. We hypothesized that mutants which formed stable dimers and had limited bending and twisting angles in AA MD simulations would

be more likely to lead to the linked cell phenotype, indicative of protein self-assembly into cylinders. Indeed, the two PduA variants predicted to form hydrogen bonds (the native ARG79 and the R79K mutant, where "R79K" indicates that the arginine (R) at 79th position in the protein sequence is mutated to a lysine (K)), have among the lowest bending and twisting angles in simulations (Fig. 4.3C, Table 4.2), and overexpression of both proteins leads to highly-elongated cells (Fig. 4.3E, SI Fig. 4.15) compared to the negative control mutants (R79A and R79*, where "*" indicates a stop codon) ($p < 0.0001$). We also verified that the expression level of various PduA mutants did not correlate strongly with cell length, indicating that differences in assembly state are not likely to be due to differences in expression. These results indicate successful assembly of WT PduA and PduA-R79K into cylinders. Surprisingly, we found that substitution of ARG79 with large hydrophobic residues (R79L) or aromatic residues (R79F, R79Y, R79W) also enabled some amount of assembly, as indicated by the presence of linked cells during overexpression (Fig. 4.3 E, SI Fig. 4.15), although to a lesser degree than the native ARG79 ($p < 0.001$). This assembly is predicted by the low bending and twisting angles calculated for these mutants from our simulations (Fig. 4.3 C). Free energy calculations for the binding energy of select mutants are also provided (R79K, R79W and R79P), but these do not correlate as well with cell length as bending and twisting angles—while hydrophobic residues tryptophan (R79W) and proline (R79P) have similar binding energies, experiments indicate that they do not exhibit similar self-assembly behavior. Overall, two residues that form hydrogen bonds (R79R and R79K) and aromatic residues (R79W, R79F and R79Y) both confer small bending and twisting angles between hexamers. Together, these results imply that bending and twisting angles as

determined by our AA MD simulations can provide insight into the assembly of PduA. The combination of experimental and AA simulation results described above suggests that PduA assembly is largely influenced by the inclination angle and twisting angle between hexamers. To understand what assembled morphologies may be accessed by these different PduA mutants, we performed CG MD simulations on PduA alone (no other proteins and no cargo) using the BMC-H model described above, and documented how preferred inclination and twisting angle between hexamers impacts the morphology of PduA-only structures formed (Fig. 4.3 D). From the inclined geometry assumed in the CG model, the resulting bending angle between two hexamer planes is approximately (SI Fig. 4.9A). The CG model predicts that at small , the mutated PduA assemble into extended structures (flat/curve sheet, cylinder/cone) with low curvature, whereas at large , the hexamers form structureless aggregates or non-extended pseudo-closed shells with high curvature (See SI Fig. 4.16 for snapshots of assembled morphologies). The flat or curved sheets are expected to roll into cylinders or cones in the presence of cargo and cell confinements. The predicted critical angles for self-assembly into extended structures from CG simulations are given by the red line in Figs. 4.3 C and D. We note that several non-assembling mutants have bending and twisting angles (calculated by AA simulations) below this line. This quantitative discrepancy is because AA simulations have detailed interatomic interactions omitted in the CG model that overestimate the CG critical bending and twisting angle. That is, agreement with the in vivo experiments is found if we take into account this shift by moving red line to the lower left. Taken together, the AA and CG simulations indicate that pure PduA hexamers that form interfaces with small bending and twisting angles can assemble into extended structures. Mutations with big

angles, despite having comparable attractive interaction strength, tend to form pseudo closed shells, unstructured aggregates and stacked layers (see Fig. 4.16 in SI).

## 4.5. Multicomponent Microcompartment Assembly: Thermodynamic Model Comparison with Coarse-Grained Simulations and Experiments

Extending our exploration of how shell protein interactions impact MCP assembly, we next used CG MD simulations and thermodynamic modeling to explore how hexamer-hexamer and pentamer-hexamer interaction strengths impact assembled morphologies. We hypothesized that modifying these interaction strength ratios would allow us to tune the morphologies of MCPs formed, enabling downstream engineering efforts. The thermodynamic model is constructed by numerically minimizing a free energy that includes protein-protein binding, protein-cargo interaction, elastic penalty of shell bending and chemical potentials (see Materials and Methods), using parameters consistent with the CG model described in the previous section. Based on the hexamer-hexamer binding energies calculated in our R79 mutation simulations, we specified a range of accessible BMC-H–BMC-H interaction ($\varepsilon_{hh}$) and BMC-H–BMC-P ($\varepsilon_{ph}$) interaction energies. We used a protein ratio of BMC-H:BMC-T:BMC-P:cargo = 4:2:1:2 for CG MD simulations (Fig. 4.4 A). For simplicity, the interaction parameters for BMC-T/BMC-T and BMC-T/BMC-H were set equal to the BMC-H/BMC-H interaction energy, because PduB trimers have conserved arginines in positions analogous to the arginines on PduA hexamers, and these are thought to be critical for binding [**31**] (in the SI Fig. 4.11 D, we discuss the effect of changing these values). The data shown in Fig. 4.4 A are for a BMC-H and BMC-T inclination angle of 25°; however, data taken for other inclination

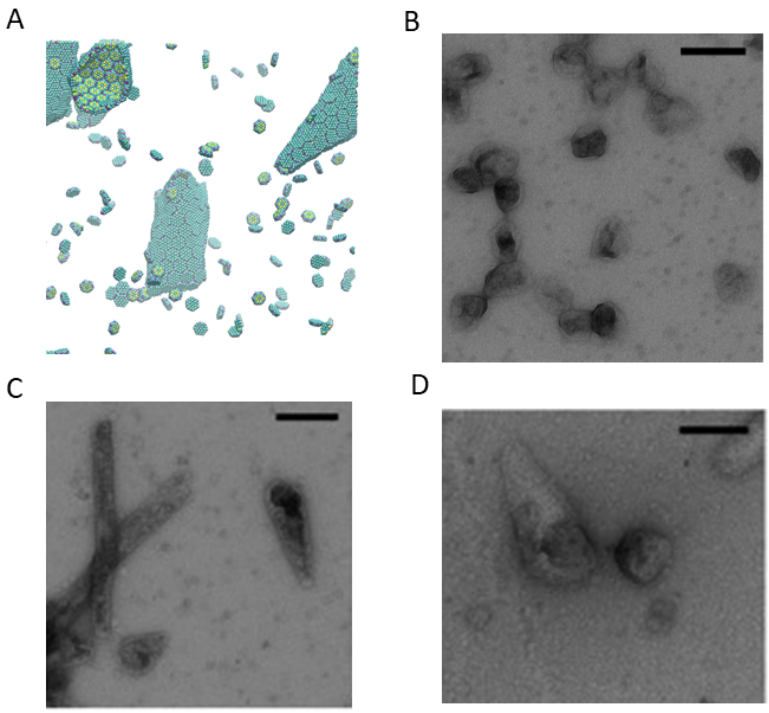



Figure 4.5. Additional morphologies formed in the CG model and comparisons to morphologies observed by TEM in experiments. (A) With a slightly smaller tilting angle, a system of only BMC-H can form cones similar to the ones observed in (C) and (D). (B) native MCP shells have similar shape compared to simulated MCPs with BMC-H, BMC-T, and BMC-P in Fig. 4.2 C and D. When BMC-P has weaker interactions($\varepsilon_{hh} = 3.8, \epsilon_{ph} = 2.28$) as shown in Fig. 4.4 D the proteins form coexisting cylinders and "samosas", similar to those observed in (C) and (D). (B) (D) are reprinted with permission from T. M. Nichols, et al. Biochemical Eng. J. 2020, 156, 107496 [**27**]

angles produces qualitatively similar morphologies and trends (SI Fig. 4.9), suggesting that the observed trends hold for a variety of inclination angles. The different morphologies predicted by the CG simulation and by the thermodynamic theory as a function of

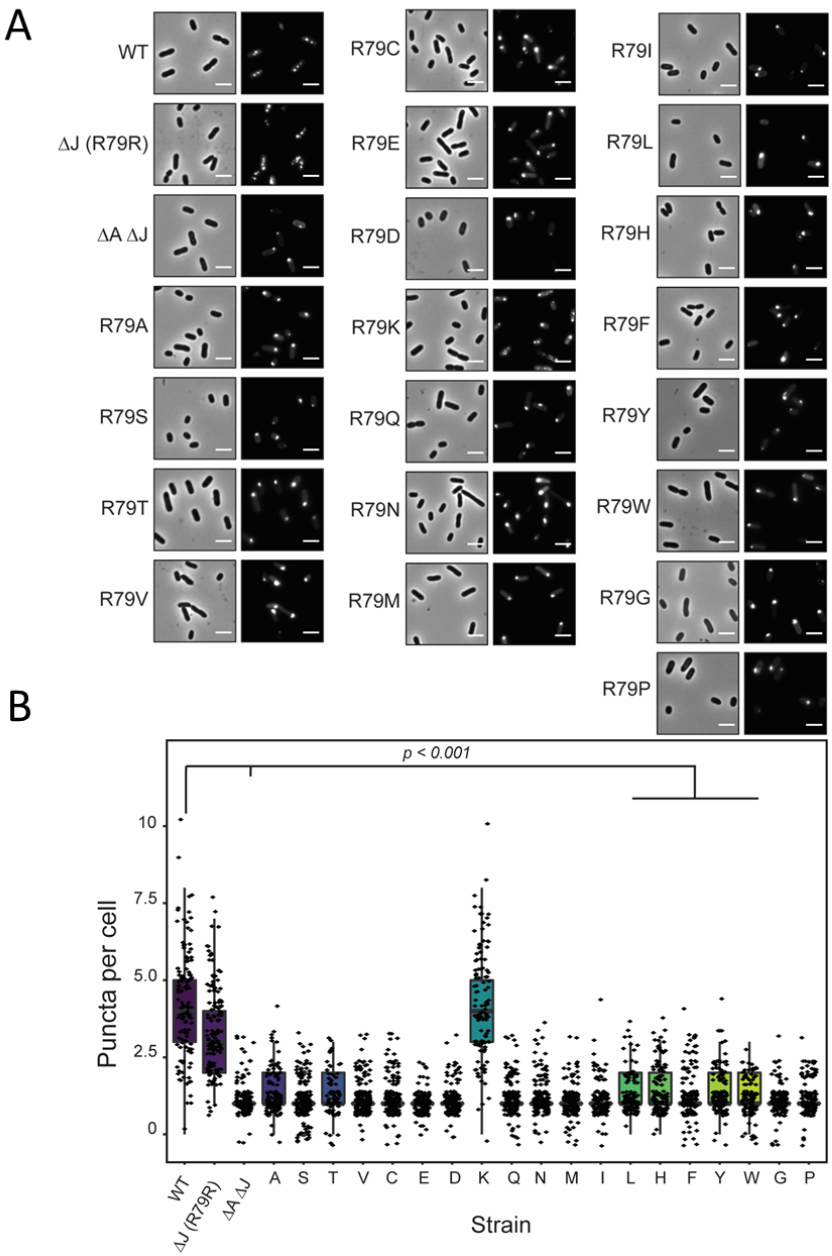



Figure 4.6. (A) GFP encapsulation assay for PduA variant strains. Phase contrast microscopy (left) and fluorescence (right) micrographs of modified Salmonella strains. GFP-containing MCPs appear as bright puncta in the cell cytoplasm (see WT image), while malformed MCPs appear as polar bodies (see $\Delta A \Delta J$ image). Scale bar = 3 µm. (B) Population distribution of puncta (MCPs) per cell for each strain. ΔJ and Δ A::R79K ΔJ had significantly more puncta per cell than other variant strains and the negative control $\Delta A \Delta J$ strain ($p < 0.001$, t-test).

pentamer-hexamer and hexamer-hexamer interactions are shown Figs. 4.4 A and B, respectively. The phase diagram constructed using CG MD simulations shows that closed MCP structures form when $\varepsilon_{hh}$ and is between 3.6 and 4.2 and $\varepsilon_{ph}/\varepsilon_{hh} > 0.7$. This range of energy values corresponds to a hexamer-hexamer attraction energy between 4 and 6 kCal/mol, obtained using Eq.4.1, which does not include the repulsion caused by screened electrostatics (adding this screened Coulomb energy contribution, the overall attractive energy is between 3.6 and 5.6 kCal/mol). When hexamer-hexamer interaction strength is increased beyond this MCP-forming range ( $> 4.4$), the proteins bind irreversibly and assemble into gel-like structures. When hexamer-hexamer attractions are weakened ($< 3.4$), no assembled morphologies are observed. When is small compared to ($< 0.7$), assembled structures do not contain pentamers. These structures include cylinders and structures resembling "samosas" with four-fold coordinated defective holes at the corners (Fig. 4.4 C). We note that these samosa-like structures are also observed when inextensible sheets are folded to form quasi-closed surface, for example in macroscopic inelastic membranes [**118**]. The thermodynamic theory (described in Materials and Methods) predicts changes in morphology with changing pentamer-hexamer ($\epsilon_{hp}$) and hexamer-hexamer ($\epsilon_{hh}$) interaction energies. By comparing the formation energy of each tested morphology, we obtain the phase diagram shown in Fig. 4.4 B. In the region of $\epsilon_{hh} < 1.0 k_B T$ (in red), the nucleation barrier is large, and, as a result, no shells could be formed. At a moderate value of $\epsilon_{hh}$, the shells can assemble into multiple shapes depending on the value of $\epsilon_{hp}$. When $\epsilon_{ph}$ is sufficiently high, the shell assembles into a small closed shell (MCP) (blue region). In the lower pentamer-hexamer interaction region, hexamers aggregate without pentamers, and in such case, cargo is required to provide the curvature

to close the shell. The competition between cargo-hexamer interaction and the bending energy decide whether a cylindrical shape or a "samosa" shape is assembled. While the cylindrical shells can encapsulate dense cargo at a considerable bending energy cost, the "samosa" shell can encapsulate loose cargo at a smaller bending energy penalty. These results are in qualitative agreement with the CG simulations in Fig. 4.4 A, except that the CG simulations for large values produce gels of connected sheets trapped in a local minimum. The lack of quantitative agreement can be explained by the fact that the thermodynamic model assumes equilibrium and constant reservoir concentrations during the assembly. Many of the morphologies observed in our CG simulations (Fig. 4.4 ) have been observed in the literature for different MCP systems assembled under various conditions (Fig. 4.5) [**106, 115, 110, 116**], supporting our hypothesis that shell protein interaction strength plays an important role in determining MCP morphology. To provide experimental support for the observed effects of changing hexamer-hexamer binding energy on MCP assembly in our models, we used a green fluorescent protein (GFP) encapsulation assay to probe MCP formation in strains expressing the PduA mutants characterized above. In this assay, GFP is fused to an N-terminal signal sequence that is sufficient for targeting GFP to the lumen of MCPs 14, 41. This results in bright, fluorescent puncta in the cell cytoplasm if MCPs are present (Fig. 4.6 A). However, if MCPs do not form properly, fluorescence will be observed instead at one or both poles of the cells; these are termed polar bodies (Fig. 4.6A). In this way, it is possible to determine if a strain is capable of forming MCPs. To this end, we constructed 19 mutant strains of Salmonella enterica serovar Typhimurium LT2 to test the effect of altering PduA assembly on overall MCP assembly. Alterations were made to the pdu operon as shown in Fig 4.6B. Because PduJ

is capable of overcoming loss of PduA function [**119**], we have knocked out the pduJ open reading frame from these strains. Thus, the WT and $\Delta J$ strains serve as positive controls, as these both contain functional, WT PduA. Note that fluorescent puncta are visible throughout the cytoplasm of these strains, as expected when MCPs form (Fig. 4.6A). The $\Delta A \Delta J$ strain, in which both the pduA and pduJ open reading frames have been knocked out, serves as a negative control. Only polar bodies are observed in this strain, indicating improper MCP assembly (Fig. 4.6A). In the experimental strains, a point mutant of PduA is encoded at the pduA locus (Fig. 4.6A) and puncta were counted in each cell to determine the effect of these mutants on MCP assembly. Notably, only strains containing PduA mutants capable of forming hydrogen bonds (R79R and R79K) are also capable of forming MCPs (Fig. 4.6 B). These strains had significantly more puncta than all other mutant strains ($p < 0.001$) (Fig. 4.6B). This includes strains that were demonstrated to have lower, but detectable, levels of assembly (R79L, R79F, R79Y, R79W) (Fig. 4.3E). This is likely due to the fact that these strains are not able to form hydrogen bonds with the adjacent hexamer and therefore have lower binding stabilities. To confirm these results, we attempted to purify MCPs from a number of the PduA point mutant strains which did not show MCP assembly in the GFP encapsulation assay. However, we were unable to purify MCPs from these strains, as evidenced by the lack of the standard MCP banding pattern by sodium dodecyl sulfate gel electrophoresis (SDS-PAGE). Together, these results demonstrate that only PduA variants capable of forming hydrogen bonds enable MCP assembly. This validates the results shown in Fig. 4.4 A which indicates that reduction of $\varepsilon_{hh}$ disrupts MCP assembly and leads to no assembly.

## 4.6. Discussion and Conclusion

Combined multi-scale MD simulations and in vivo mutation experiments show that hydrogen bonding and Coulomb interactions of arginine residues on the hexagon edges (ARG79) of PduA play a crucial role in native PduA self-assembly. In our experiments, mutations of ARG79 to most other residues negatively affect the assembly of Pdu proteins into tubes, cones or polyhedral compartments. From the fourteen mutations tested by All Atom (AA) MD simulations we found that those that have smaller bending and twisting angles are shown experimentally to be more likely to assemble into extended structures. We determine the shape of these extended structures (tubes or cones) and confirm that small twisting and bending angles promote their formation by performing coarse-grained (CG) MD simulations. The CG model also determines the conditions to form polyhedral and icosahedral shells. Bacterial microcompartments (BMC) contain hexameric (BMC-H), pseudohexameric trimers (BMC-T), and pentameric (BMC-P) structures, which are considered in our CG model. Our CG model includes twisting and bending by considering the shape, the position of the charges and specific short-range interaction (including hydrogen bonding and hydrophobic interaction) of Pdu proteins in the form of BMC-H, BMC-T and BMC-P structures. In agreement with our experimentally observed trends, the CG model predicts that small bending and twisting angles facilitate PduA self-assembly into extended cylinders. However, the agreement is only qualitative because the CG bending angle used in our CG simulations is overestimated; that is, tends to be larger than the equilibrium bending angle obtained in the AA simulations, which include more detailed interactions. According to our CG MD simulations, BMC-H proteins that maintain proper bending and twisting angles can form MCPs when BMC-P proteins and

BMC-T proteins are present and BMC-P interactions with BMC-H are strong enough. For BMC-P/BMC-H mixtures that can form MCPs, the number ratio of BMC-H to BMC-P determines the size and asphericity of assembled shells, with a higher ratio of BMC-H leading to larger, more elongated shells. We found that increasing the content of BMC-T increases the asphericity of assembled structures (SI Fig. 4.11) because BMC-T prefers smaller bending angles. In some assemblies, pentameric units are stuck in 6-coordinated sites, which is against the expectations of elasticity theory, since pentameric dislocations should repel one another [**93**]. This is likely due to local kinetic traps that are not overcome by annealing (see methods). Other morphologies including "samosas" and cylinders arise in our CG simulations and theoretical arguments by modifying the hexamer and pentamer interactions via mutations. These morphologies are robust in simulations using other CG bending angles (see SI Figs. 4.9 and 4.10) and are also observed using a thermodynamics model that includes interactions, bending energies and concentrations of the components via chemical potentials. Our experiments demonstrate that reducing the hexamer-hexamer interaction strength abolishes MCP assembly in GFP encapsulation assays, as predicted by our CG MD simulations and theory. Admittedly, our CG model has many parameters that are not completely determined. For example the physiological condition in the cell is more complex than Debye-Huckel theory can describe, and the salt concentration can range from 100 to 300 mM [**11, 12**]. And there are multivalent salts and other marcomolecules. If the electrostatic interaction is modified by changing the Bjerrium length $l_B$, the resulting assembly shape changes slightly 4.17. At the limit when $l_B = 0$, the BMC-H assemble into flat structures (Figure not shown). We have tested the

robustness of our result to different bending angles and different bending energy by exploring the phase diagram of these variables. Our main conclusions are robust to changes in these quantities. Our work explains the necessary conditions for BMC domain proteins to assemble into closed MCPs, as well as cylinders, and suggests that it is possible to systematically modify the shape of the resulting assembled structures.

## 4.7. Materials and Methods

(for experimental methods please refer to the methods and supplementary information of the publication.)

**All atom (AA) molecular dynamics (MD) simulations.** The PduA hexamers (PDB id: 3ngk) in the AA MD simulations are downloaded from the PDB and solvated in water containing 100 mM NaCl and 4 mM MgCl2. Afterward, the system undergoes a short constant volume, temperature (NVT) equilibration of 100 ps with the backbone restrained. Then, restraints are released for one PduA hexamer, while the other still has backbone restraint, and constant pressure, temperature (NPT) production simulations of 100 ns and umbrella sampling simulations are performed. For native PduA, an NPT extended simulation of 200 ns is preformed, the PduA-PduA angle calculated are similar to 100 ns simulations. More details for AA MD simulations are in the SI. Coarse-grained (CG) MD simulations. The BMC-H, BMC-T, and BMC-P models include short-range attractive interactions modeled by a Lennard-Jones-Gauss potential on a pair of attractive sites on the side of CG proteins (Fig. 4.2A), given by

$$U_{LJG}^{(ij)} = \varepsilon_{ij}[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^{6} - 0.8\exp(\frac{-(r-r_0)^2}{2\sigma^2})] \tag{4.1}$$

where $\sigma = 1.0$ nm is the bead size, $r_0$ is 1.2 nm roughly corresponding to the length of arginine side chain and i j refers to the three species, p for pentamer (BMC-P), h for hexamer (BMC-H) and t for trimer (BMC-T), $\varepsilon_{ij}$ is the binding energy between species i and j, where we choose the hexamer-hexamer interaction based on the AAMD simulation ($\varepsilon_{ij} = \varepsilon_{ij,AA} + \mu$, $\mu$ is the CG chemical potential). The coefficient 0.8 is chosen so that the CG LJG potential resembles the AA PMF. To reduce the enormous parameter space, the value of $\varepsilon_{ij}$ is set equal for the BMC-H and BMC-T proteins. The consequence of this assumption is discussed in the SI. is unknown and left as a free parameter. In addition to the hydrogen bonding, the proteins' charges (Fig. 4.2 A right column) interact via a screened Coulomb potential with Debye length of 0.91 nm. The resulting total hexamer-hexamer interaction from two sites is then about -7$k_BT$ (or -4.4 kCal/mol) per edge. The spherical cargo is a generic model for enzymes encapsulated in MCPs, they are attracted to BMC-H and BMC-T proteins at six sites by a standard 12-6 Lennard-Jones (LJ) potential with $\varepsilon = 1.8$, resulting in a total cargo/shell protein binding energy of $\sim 7.5k_BT$. The energy scale for this LJ potential was selected such that the cargo can bind to BMC-H and BMC-T, but the MCP structures are not heavily deformed by the strength of the cargo/shell protein attraction. Note that the hexamer-hexamer interaction energy in the CG model is smaller than that calculated from AA simulations because the protein concentration in CG simulations is higher (See SI appendix for details of MD simulations).

**PduA self-assembly assay**. The PduA self-assembly assay was carried out as previously described (see SI "Extended Materials and Methods"). Once cultures were grown to saturation, strains were imaged using phase contrast microscopy. All images were

adjusted equally for brightness and contrast and were cropped to an area of 500 x 500 pixels. Measurements of cell length were done using the segmented line tool in ImageJ as described previously [**95**].

**Thermodynamic theory**. In this section, we describe the thermodynamic model that explains the morphologies observed in CG MD simulations. Consider $N_s$ protein monomers (corresponding to BMC-H hexamers in the MD simulations) and $N_c$ enzyme cargo initially placed in the simulation box. When the system reaches equilibrium, the monomers are distributed either as free monomers ($N_0$) or as subunits of an oligomer ($N_q$) that contains q monomers. The total number of monomers is conserved by the relation $N_s = N_0 + qN_q$. Similarly, the free ($N_f$) and encapsidated ($N_{ce}$) cargo is subject to the conservation constraint . The free energy thus can be written as a sum of ideal free energy of each component and the interaction energy arising from shell bending energy, monomer-monomer binding energy and the cargo-monomer interaction:

$$\begin{aligned}\frac{G}{k_BT} &= N_0\left(\ln \rho_0 v_0 - 1\right) + U_0 + N_q\left(\ln \rho_q v_0 - 1\right) + U_q/k_BT \\ &+ N_{cf}(\ln \rho_{cf} - 1) + N_{ce}\left(\ln \rho_{ce} v_0\right) - 1) + U_c/k_BT \\ &+ qN_q(\epsilon_{bend} + \epsilon_{nbond}) + N_q N_{ce}\epsilon_c/k_BT\end{aligned} \tag{4.2}$$

where $v$ denotes the standard volumetric cell size $a^3$, with $a$ the approximated monomer size,$\epsilon_{bend}, \epsilon_{bond}$, and $\epsilon_c$ represent the bending energy, binding energy per monomer and cargo-monomer interaction, respectively. Note that we have assumed the total internal energy for monomers ($U_0 + U_q$) and cargo ($U_c$) are constant. Minimizing the free energy

with respect to the number of free monomers we obtain the equilibrium condition:

$$k_B T \ln(\rho_q v_0) = q k_B T \ln(\rho_0 v_0) + N_{ce} k_B T \ln(\rho_{cf} v_0) - [q(\epsilon_{bend} + \epsilon_{nbond}) + N_{ce}\epsilon_c], \tag{4.3}$$

and the law of mass action for the particle density 42

$$\rho_q = c_0 e^{-\frac{\Delta G}{k_B T}} \tag{4.4}$$

where $\Delta G = q\epsilon - q\mu_0 - N_{ce}\mu_c$ is the formation energy with the excess energy given by $\epsilon = \epsilon_{bend} + \epsilon_{bond} + N_{ce}\epsilon_c/q$ and the monomer and cargo chemical potentials are $\mu_0 = k_B T ln(\rho_0 v_0)$ and $\mu_c = k_B T ln(\rho_{cf} v_0)$, respectively.

To investigate the formation of the shell, we incorporate the line tension 43, 44so that

$$\Delta G = q\epsilon - q\mu_0 - \sigma\epsilon_{hh}\sqrt{q(Q-q)} \tag{4.5}$$

where $\mu = \mu_0 + \frac{N_{ce}}{q}\mu_c$ and $\sigma = \sqrt{\frac{4\pi A}{Q}}$, $Q$ is the monomer number of a complete shell and is the area factor of hexamer. Since the main contribution to the bending energy comes from the block geometry, it is reasonable to set the corresponding rigidity ($k_b$) as a piece-wise function, i.e., when no geometry overlaps and is finite when overlap occurs. In such a case, only the corners of samosa shape and cylinder shape assemblies are subject to the bending penalty, with the local curvature defined by the cargo size ($r_c$). Thus, we

write the formation energy for each morphology as

$$\Delta G_{\mathrm{MCP}} = 6(q - \frac{22q}{Q} + 20\frac{q}{Q}\frac{\epsilon_{hp}}{\epsilon_{hh}})(-\epsilon_{hh}) + \sigma\epsilon_{hh}\sqrt{q(Q-q)} - q\mu_0 \tag{4.6}$$

$$\Delta G_{\mathrm{cylinder}} = qk_b(1 - cos(\theta_s - \theta_0)) + (6q - 4\pi\frac{r_c}{a})(-\epsilon_{hh})$$

$$+ \frac{Aa}{4r_c}q(-\epsilon_c) - q\mu_0 \tag{4.7}$$

$$\Delta G_{\mathrm{samosa}} = k_b 4\frac{6q}{Q}(1 - \cos(\theta_s - \theta_0) + 6q(1 - \frac{4}{Q})(-\epsilon_{hh}) + \sigma\epsilon_{hh}\sqrt{q(Q-q)}$$

$$+ 4\frac{6q}{Q}(-\epsilon_c) - \frac{6q}{Q}\mu_{cf} - q\mu_0, \tag{4.8}$$

$$\Delta G_{\mathrm{flat}} = (6q - 2\sqrt{\pi qA})(-\epsilon_{hh}) - q\mu_s, \tag{4.9}$$

where $\theta_s$ is the angle between normal vectors of connected subunits and $\theta_0$ is the spontaneous angle determined by the intrinsic properties of the protein. (See SI Table S5 for the list of thermodynamic parameters).

## 4.8. Appendix for Chapter 4

### 4.8.1. Extended Computational Methods

Atomistic molecular dynamics simulation methods. All atom molecular dynamics simulations were performed using the GROMACS simulation package (version 2016.3)[**120, 121**],using Amber ff99SB-ILDN force field [**122**]. A tetragonal box of $24 \times 16\mathrm{nm}^3$ was used for constant volume, constant temperature (NVT) simulations of 2 PduA hexamers, with periodic boundary conditions. The proteins were equilibrated for 100 ps in a constant volume, constant temperature (NVT) simulation with the backbones restrained with a force constant of 1000 kJ/mol/nm$^2$. In this work all simulations were performed at the

temperature of 300 K unless otherwise noted. Then we released the restraints of one hexamer such that it was free to move or rotate to explore all possible configurations, while still ensuring that all atoms associated with protein molecules were at least 5 nm away from box edges. We then performed 100 ns constant pressure and temperature (NPT) simulations to obtain the angle between the norm of two hexamer planes. An additional 100 ns simulation NPT simulation was performed for the native PduA (pdb id: 3ngk). The angles plotted in Fig. 4.1 are defined by the norm of hexamer planes defined by the carbon on 3 of the 6 arginines at 79th position. To calculate the angle, the PduA hexamer whose backbone was restrained was taken as a reference frame and the unit vector of the other PduA was projected onto this cartesian frame to obtain bending and twisting angles. A summary of all the calculated angles is shown in Table 4.1. Mutations to the PduA structures were performed using the PyMol software. To probe the PduA hexamer binding energy, steered simulations were employed to generate approximately 18 snapshots along the reaction coordinate (the center of mass distance between the two proteins) at an interval of about 1 . 10 ns constant pressure (NPT) umbrella sampling simulations were then performed for each snapshot using a spring constant of 2000 $kJ/mol/nm^2$. Weight histogram analysis was used to calculate the potential of mean force between the two proteins. The error is estimated by the bootstrapping tool implemented in g_wham [**123**], with 100 bootstraps. The potential of mean force is shown in Figure 1 A of main text. In the sampling of PduA hexamer binding, we took advantage of the fact that arginine 79 is always the binding location, and the natural docking orientation of the two proteins falls into a narrow angular range, allowing us to have decent sampling using 10 ns umbrella windows. The umbrella sampling simulations are repeated for native PduA

and results are similar. For PduA R79 mutants, this specific short range interaction is lost, and many different docking orientations thus become available. Therefore the PMF for these mutants are not calculated. Correlation of BMC-H dimerization and hydrogen bonding. To investigate the nature of the interaction between BMC-H proteins, we used the H-bond calculator in VMD [**58**], with a cutoff distance of 3.3 and cutoff angle of 30, to count the number of hydrogen bonds in umbrella sampling simulations. The number of H-bonds correlates well with the attachment state of the two PduA hexamers. We also calculated the H-bonds on the 79th residue for hydrophilic residues, summarized in Table 4.1 in the main text. Only the native PduA and the PduA R79K mutant are able to form hydrogen bonds between the 79th residue with the main chain carbonyl oxygen of VAL 25. The lysine at position 26 in PduA (LYS26) has also been shown to form H-bonds . However, in our simulations, we find that this residue only forms H-bonds when the H-bond cutoff is increased to 3.5 , and even then these H-bonds only have an occupancy of 11%.

## 4.9. Supplementary Figures and Tables

Table 4.1. Add caption

| Residue | Residue 1 | Residue 2 | Average |
|---|---|---|---|
| native | ARG79-Side VAL25-Main 67.16% | ARG79-Side VAL25-Main 63.93% | 65.50% |
| R79K | LYS79-Side VAL25-Main 41.54% | LYS79-Side VAL25-Main 31.34% | 36.40% |
| R79C | 0 | 0 | 0.00% |
| R79N | ASN79-Side ALA28-Main 8.21% | ASN79-Side LYS26-Main 3.98% | 6.10% |
| R79S | SER79-Side LYS26-Main 47.51% | SER79-Side LYS26-Main 27.36% | 37.45% |
| R79T | THR79-Side ASN29-Side 4.23% | THR79-Side LYS26-Main 4.73% | 4.48% |
| R79W | TRP79-Side ASN29-Side 2.24% | 0 | 1.12% |
| R79Y | TYR79-Side VAL25-Main 1.74% | ASN29-Side TYR79-Side 6.97% | 4.36% |

Table 4.2. Add caption

| Mutation | Hydrophobicity or charge | Bending/deg | Twisting/deg | Assemble cylinder | Binding energy (kCal/mol) |
|---|---|---|---|---|---|
| native | positive charge | 8.56 | 4.51 | Y | |
| R79K | positive charge | 7.01 | 23.57 | Y | |
| R79D | negative charge | detach | detach | N | N/A |
| R79E | negative charge | detach | detach | N | N/A |
| R79Q | hydrophilic (5.65) | detach | detach | N | N/A |
| R79N | hydrophilic | 5.59 | 24.76 | N | N/A |
| R79S | hydrophilic | 53.44 | 40.43 | N | N/A |
| R79T | hydrophilic | 20.05 | 25.15 | N | N/A |
| R79Y | hydrophilic | 17.04 | 11.39 | Y | N/A |
| R79C | hydrophilic | 12.8 | 15.08 | N | N/A |
| R79A | hydrophobic | 34.66 | 7.52 | N | N/A |
| R79P | hydrophobic | 103.76 | 137.52 | N | |
| R79L | hydrophobic | 20.26 | 11.66 | Y | N/A |
| R79F | hydrophobic | 5.49 | 15.54 | Y | N/A |
| R79W | hydrophobic | 22.59 | 11.12 | Y | |

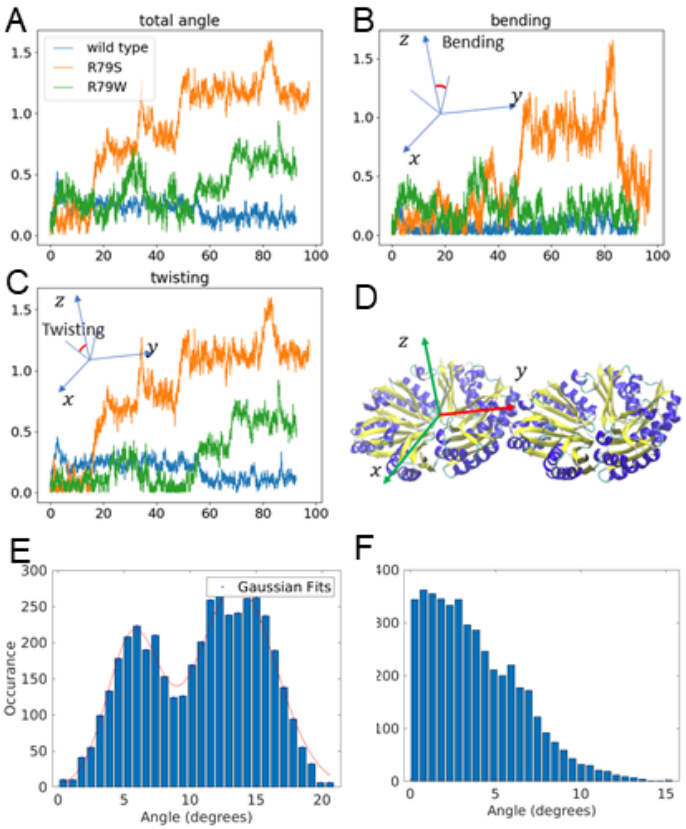



Figure 4.7. Orientation analysis of two PduA proteins in all atom simulations, comparing 3 charasteristic types of protein: the native PduA has low overall angles and the highest stability; 2 R79W mutation hexamers have intermediate stability and have slightly larger twisting angles; R79S mutation is least stable and do not self assemble according to our experiments. (A) The total angles of native PduA and two mutants, R79S and R79W. A Cartesian coordinate frame is established on the fixed BMC-H protein as shown in (D), and the surface normal of the other BMC-H is projected onto yz and xz plane to obtain the bending and twisting angles, respectively. The distribution of native PduA bending angle (E) and twisting angles (F).

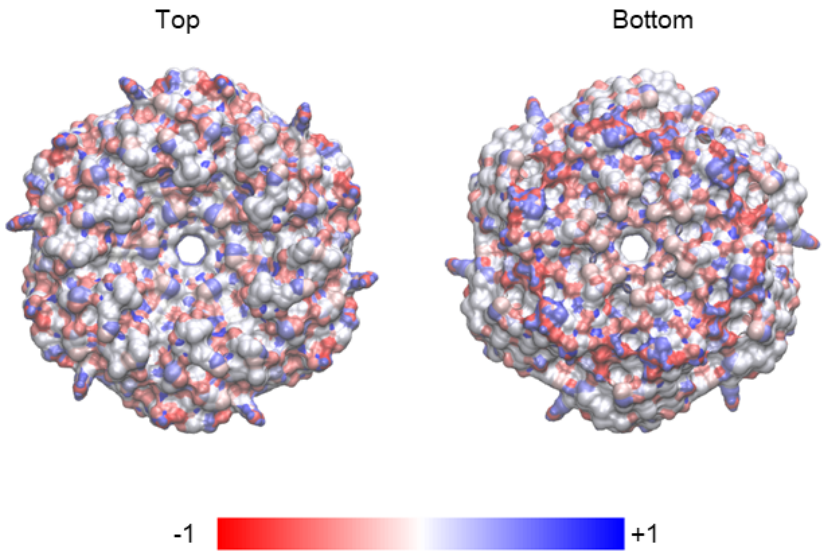



Figure 4.8. The charge distribution of PduA (pdb id: 3ngk) calculated by PDB2PQR server [**124, 125**], visualized using VMD [**58**].

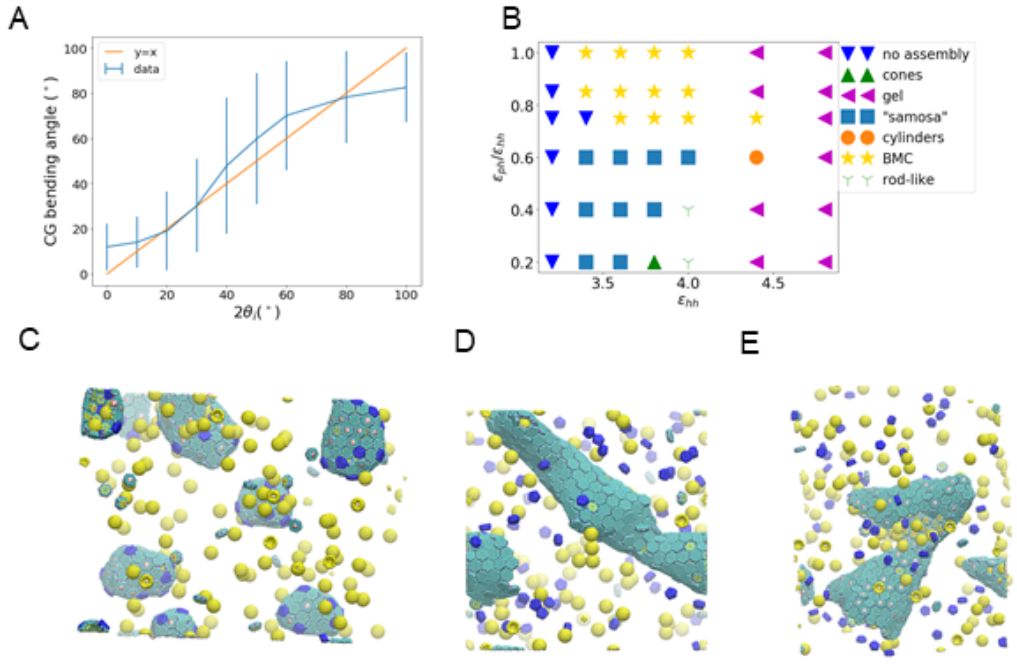



Figure 4.9. Analysis of coarse-grained (CG) bending angle. (A) CG MD simulations using different shows that the CG bending angle is $\sim 2\theta_i$. The number ratios are BMC-H : BMC-T : BMC-P : cargo=4:2:1:2 (B) Phase diagram calculated using $\theta_i = 15°$ corresponding to a cg bending angle of 30°. (C), (D) and (D) are snapshots of CG simulations where the hexamers, pentamers and cargo are in green, blue and yellow, respectively: (C) MCP observed at $\varepsilon_{hh} = 3.8, \varepsilon_{ph} = 3.8$, (D) Rod-like structure observed at $\varepsilon_{hh} = 4.0, \varepsilon_{ph} = 0.8$, and (E) Cones observed at $\varepsilon_{hh} = 3.8, \varepsilon_{ph} = 0.76$.

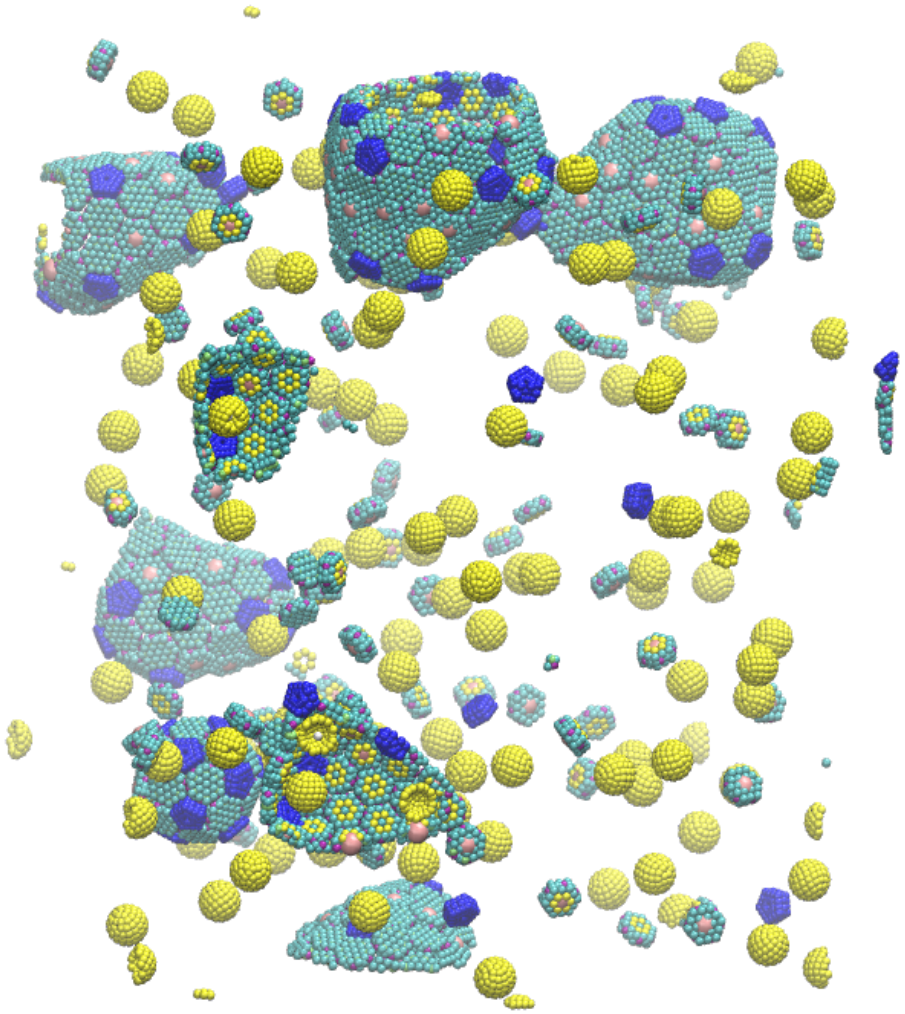


Figure 4.10. Snapshots of MCPs observed in CG MD simulations using the inclination angle $\theta_i$ CG bending angle of 10 (hexamers, pentamers and cargo in green, blue and yellow, respectively). The MCPs are qualitatively similar across different $\theta_i$, suggesting that the shape of MCPs formed does not depend sensitively on $\theta_i$, but is primarily determined by stoichiometric ratios.

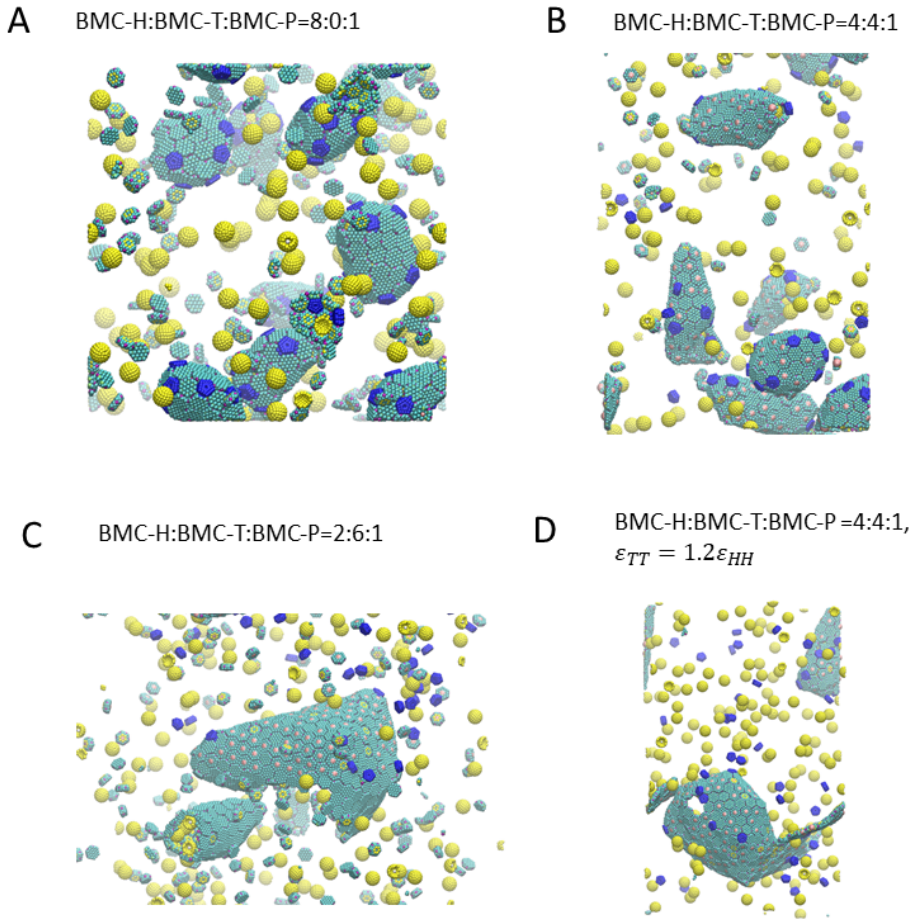



Figure 4.11. The role of BMC-T proteins in determining MCP shape. (A) to (D) are snapshots of CG MD simulations where the hexamers, pentamers and cargo are in green, blue and yellow, respectively. (A) (C) An increasing number of BMC-T proteins relative to BMC-H proteins causes the assembly to shift from higher symmetry shells in (A) to more aspherical polyhedral shells in (C). The calculated asphericity are: (A) $0.15 \pm 0.09$, (B) $0.2 \pm 0.2$ and (C) 0.2 (uncertainty nor available because there are two few data points). (D) Increasing the BMC-T/BMC-T interaction relative to the BMC-H/BMC-H interaction by a factor 1.2 causes BMC-T binding to be favored, having a similar effect as increasing the relative number of BMC-T proteins.

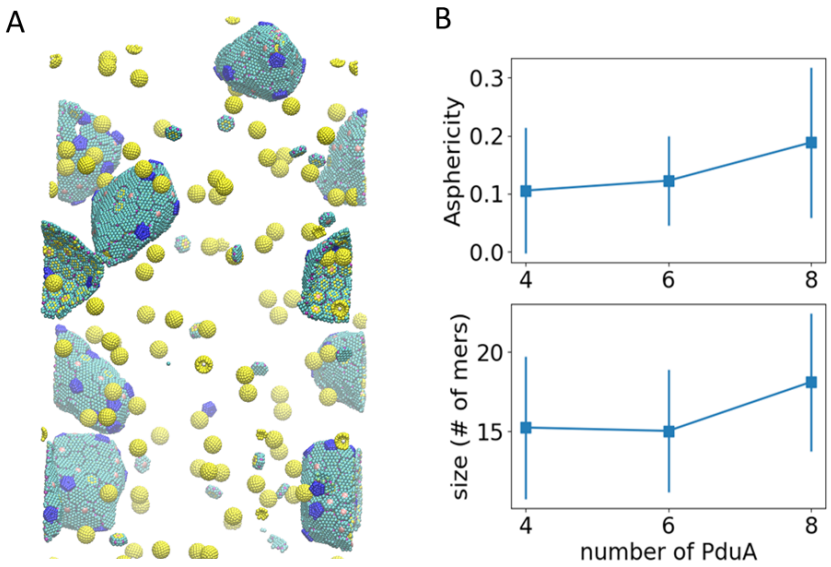



Figure 4.12. The change of MCP shapes for different BMC-H ratios. (A) The final CG simulation snapshot with (hexamers, pentamers and cargo in green, blue and yellow, respectively). (B) Asphericity and radius of gyration for varying $n_B MC - H$ in a repeating cell. This cell is replicated 5 times in x, y and z direction. The error bar represents standard deviation of all observed MCPs in the equilibrium configuration.

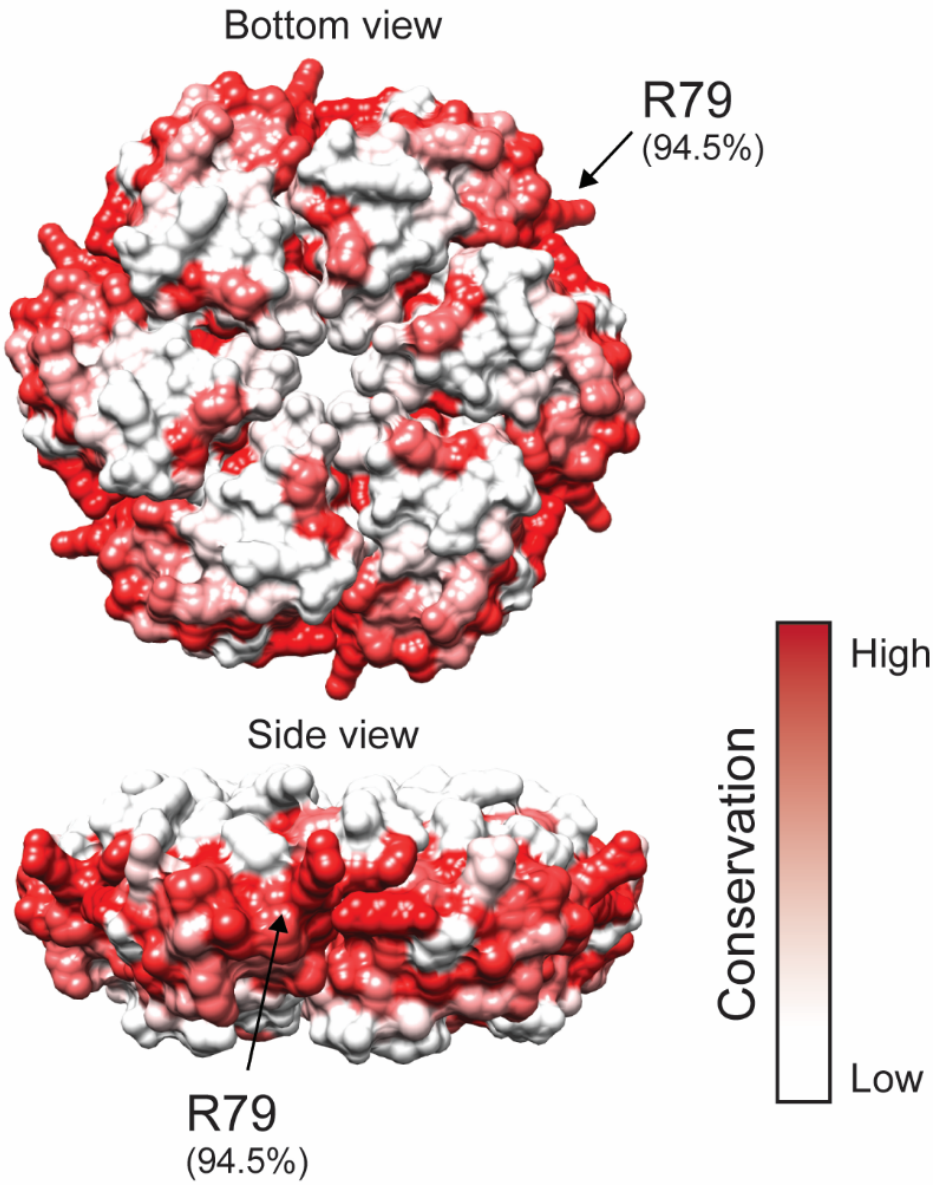



Figure 4.13. PduA conservation scores as determined by a multiple sequence alignment of 192 PduA homologs. Percent conservation scores overlaid is on the PduA structure, with ARG79. Red = high percent conservation, white = low percent conservation. Image Courtesy of Nolan Kennedy.

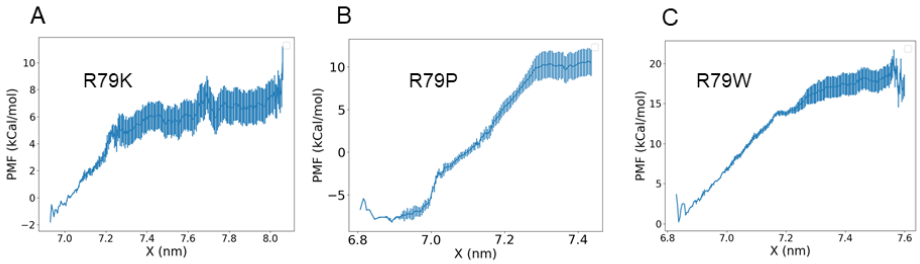



Figure 4.14. PMF of 3 selected mutations, R79K, R79P and R79W. R79K forms a hydrogen bond with the main chain VAL25 on the other hexamer and has similar binding energy to WT PduA. R79P and R79W both interact with neighboring PduA via hydrophobic interactions, but the hexamer-hexamer bending and twisting angles of R79P are much larger than R79W.

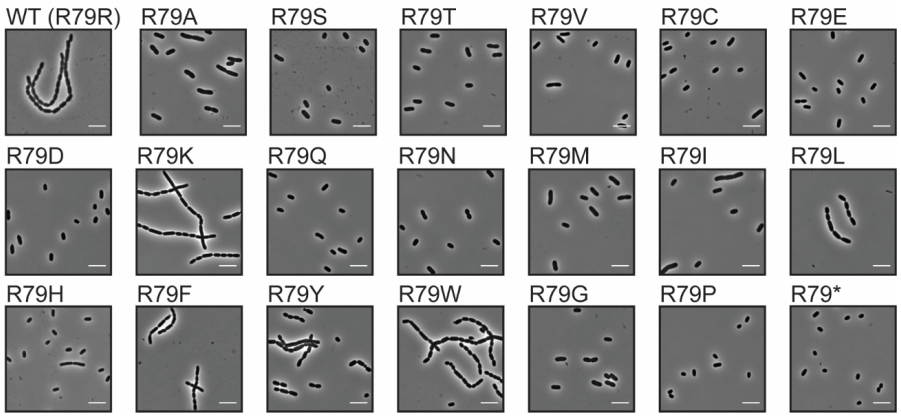



Figure 4.15. PduA self-assembly assay. Phase contrast microscopy images of cells overexpressing PduA variants (scale bar = 5 µm).

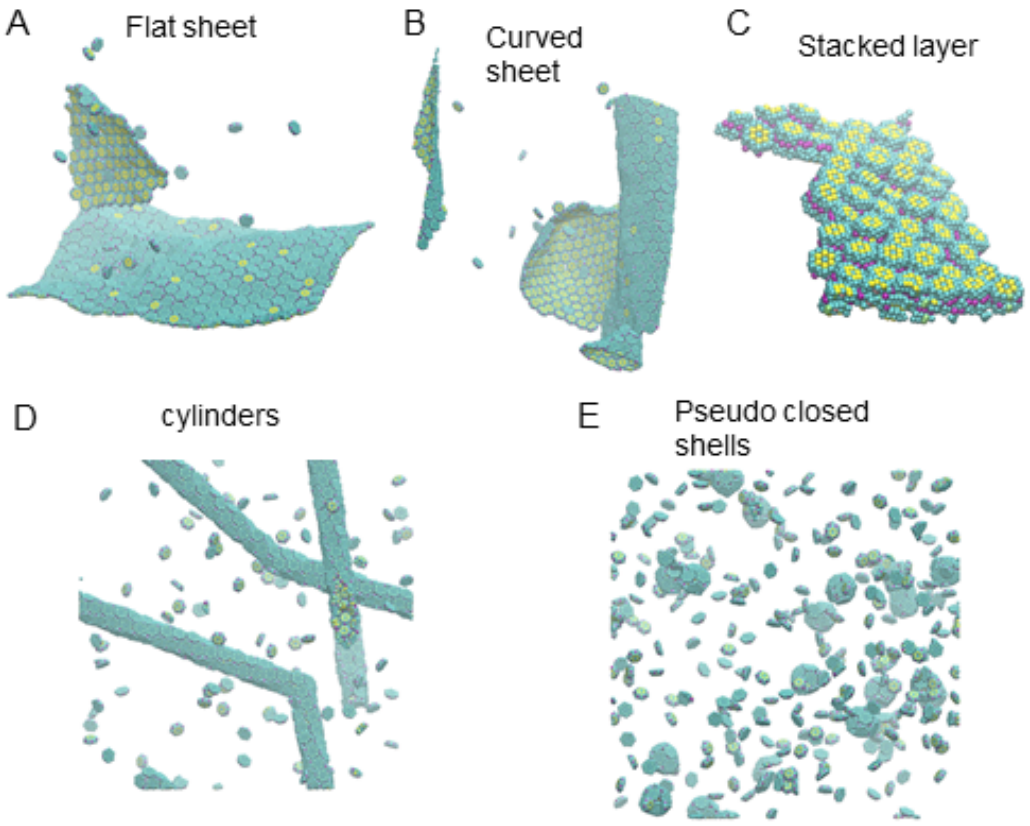



Figure 4.16. Examples of the morphologies discovered in the phase diagram of main text Fig. 4.3D. (A) Flat sheet observed at . (B) Curved sheet observed at $\theta_i = 5, \theta_t = 0$. (C) Stacked layer aggregate at $\theta_i = 0, \theta_t = 80$. (D) Cylinders at $\theta_i = 25, \theta_t = 10$. (E) Pseudo closed shells observed at $\theta_i = 40, \theta_t = 0$.

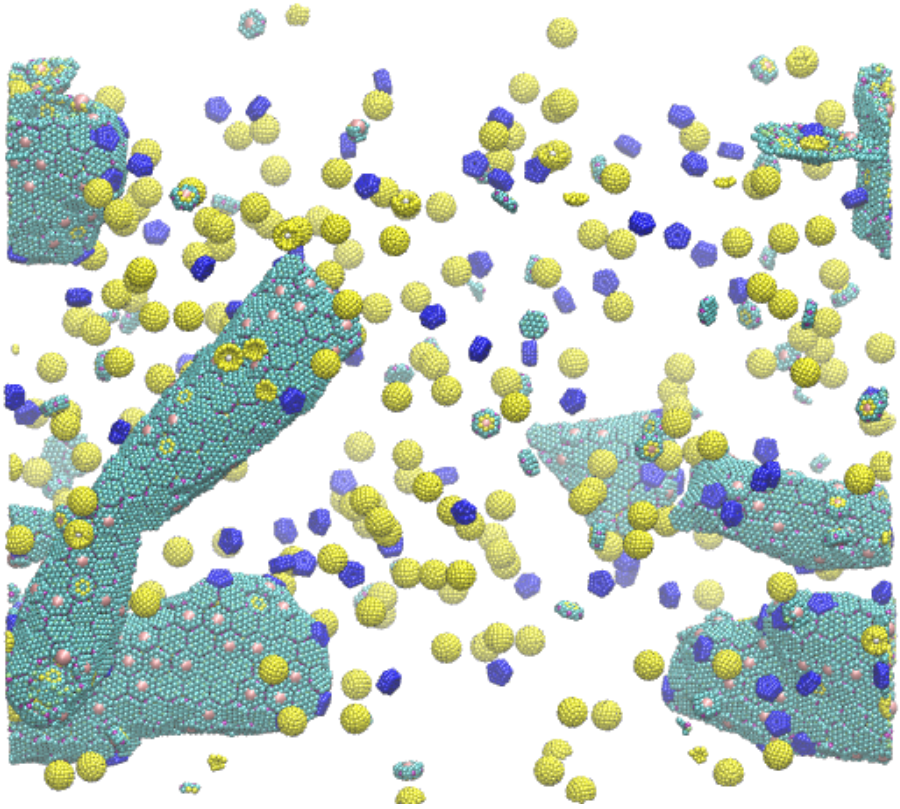


Figure 4.17. MCP morphologies formed when $l_B$ = 0.3nm. Number ratio BMC-H:BMC-T:BMC-P:cargo=4:2:1:2, all interaction parameters are the same as in Fig/4.2 except for $l_B$.

CHAPTER 5

# Summary and Outlook

In previous chapters, I have shown studies of charged nanomaterials (including NP, DNA, and proteins) in different salt environments. We find that ion correlations play an important role in the interaction of spheres and cylinders at high concentrations of monovalent salt, leading to depletion type attractions and oscillatory ion distributions. We find a slight increase in screening length compared to that predicted by the Poisson-Boltzmann (PB) equation for both geometries. We compare the results of the primitive model (PM) and the iteravive Boltzmann inversion (IBI) potentials for ions at moderate to high monovalent salt concentrations (0.5 M-2 M), and found that the two models agree. At higher salt concentrations, the PM overestimates the depletion attraction.

The potential of mean force (PMF) for spheres and cylinders are different in several ways. The spheres experience attractive interaction above 0.5 M. As the concentration increases, the electrostatic repulsion overcomes the depletion attraction. Eventually, at a very high concentration, counterion condensation leads to a re-entrant attraction. For cylinders, the critical concentration for attraction is between 1 and 2 M. Above this concentration, the cylinders are always weakly attractive for the all charge densities (that are accessible by ion dissociation or chemically modifying the cylindrical surface) that we tested. There appears to be a nonmonotonic trend in the screening length as a function of salt concentration, as discovered in MD simulations of DNA grafted NPs ([**21**] and Fig.

2.11), but this trend is not universal for other interfaces, such as cylinders. In conclusion, better CG potentials that contain realistic ion correlations such as IBI are required to accurately capture nanomaterial interactions in concentrated monovalent salt solutions.

Some physics and chemical details are not included in the sphere and the cylinder models. For example, the solvent is treated as a uniform continuum with a dielectric constant. All-atom simulations have demonstrated that polar solvent near solid-liquid interfaces cannot be described by a uniform, scalar dielectric constant [**3, 126, 127, 128**]. In our model, the complex solvation effect of water molecules is partially included via the IBI method by ensuring that the ion pair correlation function is the same as in a simulation with explicit water, but the long-range effect from dielectric mismatch is not included. CG MD simulations that include the effect of dielectric mismatch in the Olvera group have shown that for a monovalent, symmetric electrolyte, the dielectric mismatch has a minor impact on charge distributions [**2**]. These results support our simplicifation for the smooth sphere and cylinder. Intuitively, for ions with either size or valence asymmetry, the force from boundary polarization is different for different ion species. Therefore, dielectric effects are likely to play an important role in environments containing multivalent salt or salt with a large size-asymmetry. Polarization forces should be carefully taken into account when applying the findings in this thesis to systems with large dielectric contrast (such as metals) or highly asymmetric ions. For spheres and cylinders grafted with charged polymers, the image charge of charged groups might have a non-trivial effect on the polymer conformation, especially for noble-metal NPs, which have a dielectric constant

of infinity. For a more detailed study involving explicit chains on NPs or cylinders, the effect of image charge should be considered.

At physiological conditions (100 mM monovalent salt), one can use the Debye-Huckel screened potential to capture the assembly behavior of charged proteins in salt solutions. Multi-scale modeling predicts the various MCP morphology formed when the number ratio of different shell components is varied, or when the interaction strength between the species is changed. These predictions qualitatively agree with experimental findings. However, our understanding of MCPs is still far from satisfactory. For example, due to computer time limitations, we only performed AA MD simulation of only the hexamer proteins. AA investigation of the other component proteins should enable a more accurate model of the whole MCP assembly. Secondly, the role of pseudo hexamer BMC-T in the multi-component shell is not fully understood. In this model, no significant pattern of BMC-T and BMC-H distribution on the MCP is found. We have not entirely reproduced the icosahedral arrangement of small MCPs [**31, 28**].

The level of resolution (nanometer scale) in this model causes the limitation that the conclusion from the CG model is mostly qualitative. The approximation of using Debye-Huckel potential instead of a direct $1/R$ Coulomb potential with explicit ions might lead to a change in the resulting assembly and cause a shift in energy in the phase diagram. For example, when electrostatic interactions are weakened by reducing the Bjerrium length $l_B$, the MCP assembly are slightly more elongated compared to $l_B = 0.7$nm (Fig. 4.17). Other errors from omitting atomic details of the protein can also cause a shift in the results. However, since the phase diagrams calculated using different parameters (Fig. 4.4 and 4.9) both agree with thermodynamic arguments, we expect the main conclusion

to hold in spite of these errors. Several improvements could be made following this work. Proteins are represented in the CG model by a rigid-body model whose shape is determined from their crystal structures. In reality, proteins are flexible and allow changes in their configurations to some extent, as can be seen from AAMD simulation trajectories. It would be interesting to study the problem of MCP assembly using a flexible CG model, such as the previously proposed elastic network model [**129**]. A more refined model with a better description of BMC-T and the cargo might be necessary to model the role of different components in Pdu MCPs.

Briefly, the work summarized in this dissertation shows three examples of modeling charged soft nanomaterials in moderate to high salt and surface charge conditions. These results contribute to a better understanding of monovalent salt, especially at high concentrations where classical mean-field theories fail. And we demonstrate that it is possible to understand the interactions and self-assembly of proteins by CG models and simple thermodynamic theories to a semi-quantitative level, and provide guidelines to experimental work and engineering applications. The methodology proposed here can also be applied to other charged materials such as viral capsids and other proteins with ordered structures (other than very flexible proteins).

# References


[1] C Maffeo, J Yoo, J Comer, DB Wells, B Luan, and A Aksimentiev. Close encounters with dna. *Journal of Physics: Condensed Matter*, 26(41):413101, 2014.

[2] Yufei Jing, Vikram Jadhao, Jos W. Zwanikken, and Monica Olvera De La Cruz. Ionic structure in liquids confined by dielectric interfaces. *Journal of Chemical Physics*, 143(19), 2015.

[3] Debarshee Bagchi, Trung Dac Nguyen, and Monica Olvera de la Cruz. Surface polarization effects in confined polyelectrolyte solutions. *Proceedings of the National Academy of Sciences*, 117(33):19677–19684, 2020.

[4] Mitchell Miller, Yihao Liang, Honghao Li, Miaoqi Chu, Sangjun Yoo, Wei Bu, Monica Olvera de la Cruz, and Pulak Dutta. Electrostatic origin of element selectivity during rare earth adsorption. *Physical review letters*, 122(5):058001, 2019.

[5] G Gouy. *J. Phys. Radium*, 1910.

[6] DL Chapman and B Oxon. London, edinburgh dublin philos. mag. *J. Sci*, 25(475-481):486, 1913.

[7] Itamar Borukhov, David Andelman, and Henri Orland. Steric effects in electrolytes: A modified poisson-boltzmann equation. *Physical review letters*, 79(3):435, 1997.

[8] Irina A Shkel, Oleg V Tsodikov, and M Thomas Record. Complete asymptotic solution of cylindrical and spherical poisson- boltzmann equations at experimental salt concentrations. *The Journal of Physical Chemistry B*, 104(21):5161–5170, 2000.

[9] GM Torrie and JP Valleau. Electrical double layers. i. monte carlo study of a uniformly charged surface. *The Journal of Chemical Physics*, 73(11):5807–5816, 1980.

[10] John G Kirkwood. On the theory of strong electrolyte solutions. *The Journal of Chemical Physics*, 2(11):767–781, 1934.

[11] SM Abdulkarim, AB Fatimah, and JG Anderson. Effect of salt concentrations on the growth of heat-stressed and unstressed escherichia coli. *Journal of Food Agriculture and Environment*, 7(3-4):51–54, 2009.

[12] M Thomas Record Jr, Elizabeth S Courtenay, D Scott Cayley, and Harry J Guttman. Responses of e. coli to osmotic stress: large changes in amounts of cytoplasmic solutes and water. *Trends in biochemical sciences*, 23(4):143–148, 1998.

[13] John P Valleau and L Kenneth Cohen. Primitive model electrolytes. i. grand canonical monte carlo computations. *The Journal of chemical physics*, 72(11):5935–5941, 1980.

[14] Guillermo Iván Guerrero-García, Enrique González-Tovar, and Mónica Olvera de la Cruz. Entropic effects in the electric double layer of model colloids with size-asymmetric monovalent ions. *The Journal of chemical physics*, 135(5):054701, 2011.

[15] Zareen Abbas, Elisabet Ahlberg, and Sture Nordholm. Monte carlo simulations of salt solutions: exploring the validity of primitive models. *The Journal of Physical Chemistry B*, 113(17):5905–5916, 2009.

[16] Markus Deserno and Christian Holm. How to mesh up ewald sums. i. a theoretical and numerical comparison of various particle mesh routines. *The Journal of chemical physics*, 109(18):7678–7693, 1998.

[17] Enrique Gonzales-Tovar, Marcelo Lozada-Cassou, and Douglas Henderson. Hypernetted chain approximation for the distribution of ions around a cylindrical electrode. ii. numerical solution for a model cylindrical polyelectrolyte. *The Journal of chemical physics*, 83(1):361–372, 1985.

[18] P González-Mozuelos, MS Yeom, and M Olvera de la Cruz. Molecular multivalent electrolytes: microstructure and screening lengths. *The European Physical Journal E*, 16(2):167–178, 2005.

[19] Felipe Jiménez-Ángeles, Gerardo Odriozola, and Marcelo Lozada-Cassou. Electrolyte distribution around two like-charged rods: Their effective attractive interaction and angular dependent charge reversal. *Journal of Chemical Physics*, 124(13), 2006.

[20] Gerhard Hummer, Dikeos M Soumpasis, and Martin Neumann. Pair correlations in an nacl-spc water model: Simulations versus extended rism computations. *Molecular Physics*, 77(4):769–785, 1992.

[21] Yaohua Li, Martin Girard, Meng Shen, Jaime Andres Millan, and Monica Olvera De La Cruz. Strong attractions and repulsions mediated by monovalent salts. *Proceedings of the National Academy of Sciences of the United States of America*, 114(45):11838–11843, 2017.

[22] Snehasis Chowdhuri and Amalendu Chandra. Molecular dynamics simulations of aqueous nacl and kcl solutions: Effects of ion concentration on the single-particle, pair, and collective dynamical properties of ions and water molecules. *The Journal of Chemical Physics*, 115(8):3732–3741, 2001.

[23] Alexander P Lyubartsev and Stjepan Marčelja. Evaluation of effective ion-ion potentials in aqueous electrolytes. *Physical Review E*, 65(4):041202, 2002.

[24] Sumit Kewalramani, Guillermo I Guerrero-García, Liane M Moreau, Jos W Zwanikken, Chad A Mirkin, Monica Olvera de la Cruz, and Michael J Bedzyk. Electrolyte-mediated assembly of charged nanoparticles. *ACS central science*, 2(4):219–224, 2016.

[25] David P Clark. *Molecular biology: academic cell update edition*. Academic Press, 2009.

[26] Cheryl A Kerfeld, Clement Aussignargues, Jan Zarzycki, Fei Cai, and Markus Sutter. Bacterial microcompartments. *Nature Reviews Microbiology*, 16(5):277, 2018.

[27] T.M. Nichols, N. W. Kennedy, and D. Tullman-Ercek. Cargo encapsulation in bacterial microcompartments. *Methods in Enzymology*, 617:155–186, 2019.

[28] Andrew R Hagen, Jefferson S Plegaria, Nancy Sloan, Bryan Ferlez, Clement Aussignargues, Rodney Burton, and Cheryl A Kerfeld. In vitro assembly of diverse bacterial microcompartment shell architectures. *Nano letters*, 18(11):7030–7037, 2018.

[29] JP Mahalik, Kirsten A Brown, Xiaolin Cheng, and Miguel Fuentes-Cabrera. Theoretical study of the initial stages of self-assembly of a carboxysome's facet. *ACS nano*, 10(6):5751–5758, 2016.

[30] Farzaneh Mohajerani and Michael F. Hagan. The role of the encapsulated cargo in microcompartment assembly. *PLoS Computational Biology*, 14(7):1–25, 2018.

[31] Markus Sutter, Basil Greber, Clement Aussignargues, and Cheryl A Kerfeld. Assembly principles and structure of a 6.5-mda bacterial microcompartment shell. *Science*, 356(6344):1293–1297, 2017.

[32] Yaohua Li, Martin Girard, Meng Shen, Jaime Andres Millan, and Monica Olvera De La Cruz. Strong attractions and repulsions mediated by monovalent salts. *Proceedings of the National Academy of Sciences*, 114(45):11838–11843, 2017.

[33] EB Sirota, HD Ou-Yang, SK Sinha, PM Chaikin, JD Axe, and Y Fujii. Complete phase diagram of a charged colloidal system: A synchro-tron x-ray scattering study. *Physical review letters*, 62(13):1524, 1989.

[34] Martin Neumann, Othmar Steinhauser, and G Stuart Pawley. Consistent calculation of the static and frequency-dependent dielectric constant in computer simulations. *Molecular Physics*, 52(1):97–113, 1984.

[35] Chad A Mirkin, Robert L Letsinger, Robert C Mucic, and James J Storhoff. A dna-based method for rationally assembling nanoparticles into macroscopic materials. *Nature*, 382(6592):607, 1996.

[36] Bartosz A Grzybowski, Bartlomiej Kowalczyk, István Lagzi, Dawei Wang, Konstantin V Tretiakov, and David A Walker. Inorganic salts direct the assembly of charged nanoparticles into composite nanoscopic spheres, plates, or needles. *Faraday Discussions*, 159(1):201–209, 2012.

[37] Alexander McPherson. Introduction to protein crystallization. *Methods*, 34(3):254–265, 2004.

[38] RA Curtis, JM Prausnitz, and HW Blanch. Protein-protein and protein-salt interactions in aqueous protein solutions containing concentrated electrolytes. *Biotechnology and bioengineering*, 57(1):11–21, 1998.

[39] André C Dumetz, Ann M Snellinger-O'Brien, Eric W Kaler, and Abraham M Lenhoff. Patterns of protein–protein interactions in salt solutions and implications for protein crystallization. *Protein Science*, 16(9):1867–1877, 2007.

[40] M Olvera de La Cruz, L Belloni, M Delsanti, JP Dalbiez, O Spalla, and M Drifford. Precipitation of highly charged polyelectrolyte solutions in the presence of multivalent salts. *The Journal of chemical physics*, 103(13):5781–5791, 1995.

[41] Francisco J Solis and Monica Olvera De La Cruz. Attractive interactions between rodlike polyelectrolytes: polarization, crystallization, and packing. *Physical Review E*, 60(4):4496, 1999.

[42] Ioulia Rouzina and Victor A Bloomfield. Macroion attraction due to electrostatic correlation between screening counterions. 1. mobile surface-adsorbed ions and diffuse ion cloud. *The Journal of Physical Chemistry*, 100(23):9977–9989, 1996.

[43] Eric Raspaud, M Olvera De La Cruz, J-L Sikorav, and Francoise Livolant. Precipitation of dna by polyamines: a polyelectrolyte behavior. *Biophysical journal*, 74(1):381–393, 1998.

[44] M Scott Shell. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *The Journal of chemical physics*, 129(14):144108, 2008.

[45] AK Soper. Empirical potential monte carlo simulation of fluid structure. *Chemical Physics*, 202(2-3):295–306, 1996.

[46] Linfeng Zhang, Jiequn Han, Han Wang, Roberto Car, and E Weinan. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Physical review letters*, 120(14):143001, 2018.

[47] Guillermo Iván Guerrero-García, Pedro González-Mozuelos, and Monica Olvera De La Cruz. Potential of mean force between identical charged nanoparticles immersed in a size-asymmetric monovalent electrolyte. *The Journal of chemical physics*, 135(16):164705, 2011.

[48] Honghao Li, Aykut Erbas, Jos Zwanikken, and Monica Olvera de la Cruz. Ionic conductivity in polyelectrolyte hydrogels. *Macromolecules*, 49(23):9239–9246, 2016.

[49] Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications*, volume 1. Elsevier, 2001.

[50] M Boström, DRM Williams, and BW Ninham. Specific ion effects: why dlvo theory fails for biology and colloid systems. *Physical review letters*, 87(16):168103, 2001.

[51] RM Pashley. Dlvo and hydration forces between mica surfaces in li+, na+, k+, and cs+ electrolyte solutions: A correlation of double-layer and hydration forces with surface cation exchange properties. *Journal of Colloid and Interface Science*, 83(2):531–546, 1981.

[52] Jonathan Ennis, Roland Kjellander, and D John Mitchell. Dressed ion theory for bulk symmetric electrolytes in the restricted primitive model. *The Journal of chemical physics*, 102(2):975–991, 1995.

[53] Jos W Zwanikken and Monica Olvera de la Cruz. Tunable soft structure in charged fluids confined by dielectric interfaces. *Proceedings of the National Academy of Sciences*, 110(14):5301–5308, 2013.

[54] Alexander A Rashin and Barry Honig. Reevaluation of the born model of ion hydration. *The journal of physical chemistry*, 89(26):5588–5593, 1985.

[55] SAJJA Suguna, DH Nandal, SURESH Kamble, AMBADASU Bharatha, and RAHUL Kunkulol. Genomic dna isolation from human whole blood samples by non enzymatic salting out method. *Int J pharm pharm sci*, 6(6):198–199, 2014.

[56] Philip Nelson. *Biological physics*. WH Freeman New York, 2004.

[57] Yun Luo and Benoît Roux. Simulation of osmotic pressure in concentrated aqueous salt solutions. *The journal of physical chemistry letters*, 1(1):183–189, 2010.

[58] William Humphrey, Andrew Dalke, Klaus Schulten, et al. Vmd: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38, 1996.

[59] Sho Asakura and Fumio Oosawa. Interaction between particles suspended in solutions of macromolecules. *Journal of Polymer Science Part A: Polymer Chemistry*, 33(126):183–192, 1958.

[60] Y Mao, ME Cates, and HNW Lekkerkerker. Depletion force in colloidal systems. *Physica a*, 222:10–24, 1995.

[61] Boran Ma, Trung Dac Nguyen, Victor A Pryamitsyn, and Monica Olvera de la Cruz. Ionic correlations in random ionomers. *ACS nano*, 12(3):2311–2318, 2018.

[62] Berk Hess, Christian Holm, and Nico van der Vegt. Modeling multibody effects in ionic solutions with a concentration dependent dielectric permittivity. *Physical review letters*, 96(14):147801, 2006.

[63] RC Ball, JF Marko, ST Milner, and TA Witten. Polymers grafted to a convex surface. *Macromolecules*, 24(3):693–703, 1991.

[64] Kurinji Krishnamoorthy, Sumit Kewalramani, Ali Ehlen, Liane M Moreau, Chad A Mirkin, Monica Olvera de la Cruz, and Michael J Bedzyk. Enzymatic degradation of dna probed by in situ x-ray scattering. *ACS nano*, 13(10):11382–11391, 2019.

[65] Soyoung E Seo, Martin Girard, Monica Olvera De La Cruz, and Chad A Mirkin. The importance of salt-enhanced electrostatic repulsion in colloidal crystal engineering with dna. *ACS central science*, 5(1):186–191, 2019.

[66] Ting ING Li, Rastko Sknepnek, Robert J Macfarlane, Chad A Mirkin, and Monica Olvera de la Cruz. Modeling the crystallization of spherical nucleic acid nanoparticle conjugates with molecular dynamics simulations. *Nano letters*, 12(5):2509–2514, 2012.

[67] Alexander M Smith, Alpha A Lee, and Susan Perkin. The electrostatic screening length in concentrated electrolytes increases with concentration. *The journal of physical chemistry letters*, 7(12):2157–2163, 2016.

[68] Carla S Perez-Martinez, Alexander M Smith, Susan Perkin, et al. Scaling analysis of the screening length in concentrated electrolytes. *Physical review letters*, 119(2):026002, 2017.

[69] Johannes Zeman, Svyatoslav Kondrat, and Christian Holm. Bulk ionic screening lengths from extremely large-scale molecular dynamics simulations. *Chemical Communications*, 2020.

[70] Zachary AH Goodwin and Alexei A Kornyshev. Underscreening, overscreening and double-layer capacitance. *Electrochemistry Communications*, 82:129–133, 2017.

[71] Barrett O'neill. *Elementary differential geometry*. Academic press, 2014.

[72] Hartmut Löwen. Interaction between charged rodlike colloidal particles. *Physical review letters*, 72(3):424, 1994.

[73] B-Y Ha and Andrea J Liu. Counterion-mediated attraction between two like-charged rods. *Physical Review Letters*, 79(7):1289, 1997.

[74] Niels Grønbech-Jensen, Robert J Mashl, Robijn F Bruinsma, and William M Gelbart. Counterion-induced attraction between rigid polyelectrolytes. *Physical Review Letters*, 78(12):2477, 1997.

[75] Roland Kjellander. Focus Article: Oscillatory and long-range monotonic exponential decays of electrostatic interactions in ionic liquids and other electrolytes: The significance of dielectric permittivity and renormalized charges. *Journal of Chemical Physics*, 148(19), 2018.

[76] Prudhvidhar Gaddam and William Ducker. Electrostatic screening length in concentrated salt solutions. *Langmuir*, 35(17):5719–5727, 2019.

[77] Vikram Jadhao, Francisco J Solis, and Monica Olvera De La Cruz. Simulation of charged systems in heterogeneous dielectric media via a true energy functional. *Physical review letters*, 109(22):223905, 2012.

[78] Teena Goel, Chandra N. Patra, Swapan K. Ghosh, and Tulsi Mukherjee. Molecular solvent model of cylindrical electric double layers: A systematic study by Monte Carlo simulations and density functional theory. *Journal of Chemical Physics*, 129(15), 2008.

[79] Dirk Stigter. The charged colloidal cylinder with a Gouy double layer. *Journal of Colloid and Interface Science*, 53(2):296–306, 1975.

[80] DA McQuarrie. Statistical mechanics. 2000. *Sausalito, Calif.: University Science Books*, 12:641, 2004.

[81] Jean-Pierre Hansen and Ian R McDonald. *Theory of simple liquids*. Elsevier, 1990.

[82] Erik Wernersson, Roland Kjellander, and Johannes Lyklema. Charge inversion and ion- ion correlation effects at the mercury/aqueous mgso4 interface: Toward the solution of a long-standing issue. *The Journal of Physical Chemistry C*, 114(4):1849–1866, 2010.

[83] Roya Zandi, Bogdan Dragnea, Alex Travesset, and Rudolf Podgornik. On virus growth and form. *Physics Reports*, 847:1–102, 2020.

[84] Rees F Garmann, Aaron M Goldfain, and Vinothan N Manoharan. Measurements of the self-assembly kinetics of individual viral capsids around their rna genome. *Proceedings of the National Academy of Sciences*, 116(45):22485–22490, 2019.

[85] Julien Jorda, David Lopez, Nicole M Wheatley, and Todd O Yeates. Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Science*, 22(2):179–195, 2013.

[86] Chiranjit Chowdhury, Sunny Chun, Allan Pang, Michael R Sawaya, Sharmistha Sinha, Todd O Yeates, and Thomas A Bobik. Selective molecular transport through the protein shell of a bacterial microcompartment organelle. *Proceedings of the National Academy of Sciences*, 112(10):2990–2995, 2015.

[87] Seth D Axen, Onur Erbilgin, and Cheryl A Kerfeld. A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Comput Biol*, 10(10):e1003898, 2014.

[88] Marilyn F Slininger Lee, Christopher M Jakobson, and Danielle Tullman-Ercek. Evidence for improved encapsulated pathway behavior in a bacterial microcompartment through shell protein engineering. *ACS synthetic biology*, 6(10):1880–1891, 2017.

[89] Sophia J Tsai and Todd O Yeates. Bacterial microcompartments: insights into the structure, mechanism, and engineering applications. In *Progress in molecular biology and translational science*, volume 103, pages 1–20. Elsevier, 2011.

[90] Shouqiang Cheng, Yu Liu, Christopher S Crowley, Todd O Yeates, and Thomas A Bobik. Bacterial microcompartments: their properties and paradoxes. *Bioessays*, 30(11-12):1084–1095, 2008.

[91] Leonhard Euler. Elementa doctrinae solidorum. *Novi commentarii academiae scientiarum Petropolitanae*, pages 109–140, 1758.

[92] Donald LD Caspar and Aaron Klug. Physical principles in the construction of regular viruses. In *Cold Spring Harbor symposia on quantitative biology*, volume 27, pages 1–24. Cold Spring Harbor Laboratory Press, 1962.

[93] Jack Lidmar, Leonid Mirny, and David R Nelson. Virus shapes and buckling transitions in spherical shells. *Physical Review E*, 68(5):051910, 2003.

[94] Chenguang Fan, Shouqiang Cheng, Yu Liu, Cristina M Escobar, Christopher S Crowley, Robert E Jefferson, Todd O Yeates, and Thomas A Bobik. Short n-terminal sequences package proteins into bacterial microcompartments. *Proceedings of the National Academy of Sciences*, 107(16):7509–7514, 2010.

[95] Nolan W Kennedy, Jasmine M Hershewe, Taylor M Nichols, Eric W Roth, Charlene D Wilke, Carolyn E Mills, Michael C Jewett, and Danielle Tullman-Ercek. Apparent size and morphology of bacterial microcompartments varies with technique. *PloS one*, 15(3):e0226395, 2020.

[96] Graziano Vernizzi, Rastko Sknepnek, and Monica Olvera de la Cruz. Platonic and archimedean geometries in multicomponent elastic membranes. *Proceedings of the National Academy of Sciences*, 108(11):4292–4296, 2011.

[97] Mark J Bowick and Rastko Sknepnek. Pathways to faceting of vesicles. *Soft Matter*, 9(34):8088–8095, 2013.

[98] Thomas A Bobik, Gregory D Havemann, Robert J Busch, Donna S Williams, and Henry C Aldrich. The Propanediol Utilization (pdu) Operon ofSalmonella enterica Serovar Typhimurium LT2 Includes Genes Necessary for Formation of Polyhedral Organelles Involved in Coenzyme B12-Dependent 1, 2-Propanediol Degradation. *Journal of bacteriology*, 181(19):5967–5975, 1999.

[99] Sharmistha Sinha, Shouqiang Cheng, Yea Won Sung, Dan E McNamara, Michael R Sawaya, Todd O Yeates, and Thomas A Bobik. Alanine scanning mutagenesis identifies an asparagine–arginine–lysine triad essential to assembly of the shell of the pdu microcompartment. *Journal of molecular biology*, 426(12):2328–2345, 2014.

[100] Christopher S Crowley, Duilio Cascio, Michael R Sawaya, Jeffery S Kopstein, Thomas A Bobik, and Todd O Yeates. Structural insight into the mechanisms of transport across the salmonella enterica pdu microcompartment shell. *Journal of Biological Chemistry*, 285(48):37838–37846, 2010.

[101] Allan Pang, Stefanie Frank, Ian Brown, Martin J Warren, and Richard W Pickersgill. Structural insights into higher order assembly and function of the bacterial microcompartment protein pdua. *Journal of Biological Chemistry*, 289(32):22377–22384, 2014.

[102] Yingssu Tsai, Michael R Sawaya, Gordon C Cannon, Fei Cai, Eric B Williams, Sabine Heinhorst, Cheryl A Kerfeld, and Todd O Yeates. Structural analysis of csos1a and the protein shell of the halothiobacillus neapolitanus carboxysome. *PLoS biology*, 5(6):e144, 2007.

[103] Gautam R Desiraju and Thomas Steiner. *The weak hydrogen bond: in structural chemistry and biology*, volume 9. International Union of Crystal, 2001.

[104] Allan Pang, Mingzhi Liang, Michael B Prentice, and Richard W Pickersgill. Substrate channels revealed in the trimeric Lactobacillus reuteri bacterial microcompartment shell protein PduB. *Acta Crystallographica Section D: Biological Crystallography*, 68(12):1642–1652, 2012.

[105] BV Venkataram Prasad, GJ Wang, John PM Clerx, and Wah Chiu. Three-dimensional structure of rotavirus. *Journal of molecular biology*, 199(2):269–275, 1988.

[106] Chenguang Fan, Shouqiang Cheng, Yu Liu, Cristina M. Escobar, Christopher S. Crowley, Robert E. Jefferson, Todd O. Yeates, and Thomas A. Bobik. Short N-terminal sequences package proteins into bacterial microcompartments. *Proceedings of the National Academy of Sciences of the United States of America*, 107(16):7509–7514, 2010.

[107] Andrew R. Hagen, Jefferson S. Plegaria, Nancy Sloan, Bryan Ferlez, Clement Aussignargues, Rodney Burton, and Cheryl A. Kerfeld. In Vitro Assembly of Diverse Bacterial Microcompartment Shell Architectures. *Nano Letters*, 18(11):7030–7037, 2018.

[108] Siyu Li, Polly Roy, Alex Travesset, and Roya Zandi. Why large icosahedral viruses need scaffolding proteins. *Proceedings of the National Academy of Sciences*, 115(43):10971–10976, 2018.

[109] Curt Waltmann, Roi Asor, Uri Raviv, and Monica Olvera de la Cruz. Assembly and Stability of Simian Virus 40 Polymorphs. *ACS Nano*, 2020.

[110] Nicole M Wheatley, Soheil D Gidaniyan, Yuxi Liu, Duilio Cascio, and Todd O Yeates. Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins. *Protein Science*, 22(5):660–665, 2013.

[111] Markus Sutter, Basil Greber, Clement Aussignargues, and Cheryl A Kerfeld. Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. *Science*, 356(6344):1293–1297, 2017.

[112] Shiho Tanaka, Michael R. Sawaya, Martin Phillips, and Todd O. Yeates. Insights from multiple structures of the shell proteins from the $\beta$-carboxysome. *Protein Science*, 18(1):108–120, 2009.

[113] G. Dean Price and Murray R. Badger. Evidence for the role of carboxysomes in the cyanobacterial CO 2 -concentrating mechanism . *Canadian Journal of Botany*, 69(5):963–973, 1991.

[114] Nolan W Kennedy, Jasmine M Hershewe, Taylor M Nichols, Eric W Roth, Charlene D Wilke, Carolyn E Mills, Michael C Jewett, and Danielle Tullman-Ercek. Apparent size and morphology of bacterial microcompartments varies with technique. *PloS one*, 15(3):e0226395, 2020.

[115] Allan Pang, Stefanie Frank, Ian Brown, Martin J Warren, and Richard W Pickersgill. Structural insights into higher order assembly and function of the bacterial microcompartment protein PduA. *Journal of Biological Chemistry*, 289(32):22377–22384, 2014.

[116] Gregory D Havemann, Edith M Sampson, and Thomas A Bobik. PduA is a shell protein of polyhedral organelles involved in coenzyme B12-dependent degradation of 1, 2-propanediol in Salmonella enterica serovar Typhimurium LT2. *Journal of Bacteriology*, 184(5):1253–1261, 2002.

[117] Sophia J Tsai and Todd O Yeates. Bacterial microcompartments: insights into the structure, mechanism, and engineering applications. In *Progress in molecular biology and translational science*, volume 103, pages 1–20. Elsevier, 2011.

[118] Joseph D Paulsen, Vincent Démery, Christian D Santangelo, Thomas P Russell, Benny Davidovitch, and Narayanan Menon. Optimal wrapping of liquid droplets with ultrathin sheets. *Nature materials*, 14(12):1206–1209, 2015.

[119] Nolan W Kennedy, Svetlana P Ikonomova, Marilyn Slininger Lee, Henry W Raeder, and Danielle Tullman-Ercek. Self-assembling shell proteins pdua and pduj have essential and redundant roles in bacterial microcompartment assembly. *Journal of Molecular Biology*, 433(2):166721, 2020.

[120] Berk Hess, Carsten Kutzner, David Van Der Spoel, and Erik Lindahl. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation*, 4(3):435–447, 2008.

[121] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E Mark, and Herman JC Berendsen. Gromacs: fast, flexible, and free. *Journal of computational chemistry*, 26(16):1701–1718, 2005.

[122] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L Klepeis, Ron O Dror, and David E Shaw. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins: Structure, Function, and Bioinformatics*, 78(8):1950–1958, 2010.

[123] Jochen S Hub, Bert L De Groot, and David Van Der Spoel. g_wham a free weighted histogram analysis implementation including robust error and autocorrelation estimates. *Journal of chemical theory and computation*, 6(12):3713–3720, 2010.

[124] Todd J Dolinsky, Jens E Nielsen, J Andrew McCammon, and Nathan A Baker. Pdb2pqr: an automated pipeline for the setup of poisson–boltzmann electrostatics calculations. *Nucleic acids research*, 32(suppl_2):W665–W667, 2004.

[125] Todd J Dolinsky, Paul Czodrowski, Hui Li, Jens E Nielsen, Jan H Jensen, Gerhard Klebe, and Nathan A Baker. Pdb2pqr: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic acids research*, 35(suppl_2):W522–W525, 2007.

[126] Felipe Jiménez-Ángeles, Katherine J Harmon, Trung Dac Nguyen, Paul Fenter, and Monica Olvera de la Cruz. Nonreciprocal interactions induced by water in confinement. *Physical Review Research*, 2(4):043244, 2020.

[127] Setare Mostajabi Sarhangi, Morteza M Waskasi, Seyed Majid Hashemianzadeh, and Dmitry V Matyushov. Effective dielectric constant of water at the interface with charged c60 fullerenes. *The Journal of Physical Chemistry B*, 123(14):3135–3143, 2019.

[128] Jiaxing Yuan, Hanne S. Antila, and Erik Luijten. Dielectric Effects on Ion Transport in Polyelectrolyte Brushes. *ACS Macro Letters*, 8(2):183–187, 2019.

[129] Lei Yang, Guang Song, and Robert L Jernigan. Protein elastic network models and the ranges of cooperativity. *Proceedings of the National Academy of Sciences*, 106(30):12347–12352, 2009.